

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12N 15/62 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200880009677.4

[43] 公开日 2010 年 2 月 3 日

[11] 公开号 CN 101641441A

[22] 申请日 2008.1.30

[21] 申请号 200880009677.4

[30] 优先权

[32] 2007.1.31 [33] US [31] 60/887,486

[32] 2007.1.31 [33] US [31] 60/887,476

[86] 国际申请 PCT/US2008/052434 2008.1.30

[87] 国际公布 WO2008/094986 英 2008.8.7

[85] 进入国家阶段日期 2009.9.24

[71] 申请人 陶氏环球技术公司

地址 美国密歇根州

[72] 发明人 拉塞尔·J·科尔曼

黛安娜·雷塔拉克 简·C·施奈德

托马斯·M·拉姆西尔

查尔斯·D·赫什伯格

斯泰西·L·李 索尔·里斯尼克

[74] 专利代理机构 北京市柳沈律师事务所

代理人 吴培善

权利要求书 8 页 说明书 67 页 序列表 14 页
附图 11 页

[54] 发明名称

用于提高表达的细菌前导序列

[57] 摘要

提供了用于改善感兴趣蛋白质或多肽在宿主细胞中的表达和/或分泌的组合物和方法。提供了包含细菌分泌信号肽编码序列的组合物。可以在用于在宿主细胞中转化和表达感兴趣蛋白质或多肽的载体构建体或表达系统中使用该编码序列。本发明的组合物对于增加正确加工的蛋白质在宿主细胞的周质间隙中的积累，或者对于增加正确加工的蛋白质自宿主细胞的分泌是有用的。具体地，该分泌信号肽包含蛋白质二硫键异构酶 C (dsbC)、突变型磷酸结合蛋白 (pbp*)、蛋白质二硫键异构酶 A (dsbA)、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白 (ORF5550)、甲苯耐受蛋白 (Ttg2C)、或接纳甲基的趋化蛋白 (ORF8124) 分泌信号。

1. 一种分离的核酸分子，其包含选自下组的分泌多肽的分泌信号编码序列：蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、和接纳甲基的趋化蛋白(ORF8124)分泌多肽。
2. 权利要求1的核酸分子，其中所述核酸分子选自下组：
 - a) 包含SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列的核酸分子；
 - b) 包含与SEQ ID NO:5、3、7、9、11、13、15、17、21、或23的核苷酸序列具有至少90%序列同一性的核苷酸序列的核酸分子，其中所述核苷酸序列编码分泌多肽；
 - c) 编码包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列的聚多肽的核酸分子；
 - d) 包含编码与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
 - e) 包含编码与SEQ ID NO:20的氨基酸序列具有至少96%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
 - f) 在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列；和，
 - g) 在严格条件下在基本上全长里与编码选自下组的氨基酸序列的核苷酸序列杂交的核苷酸序列：SEQ ID NO:6、4、8、10、12、14、16、18、和22。
3. 权利要求2的核酸分子，其中所述杂交条件包括约60°C至约70°C的温度。
4. 权利要求2或3的核酸分子，其中所述杂交条件包括约68°C的温度。
5. 权利要求1的核酸分子，其中所述核酸分子已经得到调节以反映选择用来表达所述核酸分子的宿主生物体的密码子偏爱。
6. 一种载体，其包含蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、

NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、或接纳甲基的趋化蛋白(ORF8124)分泌多肽的分泌信号编码序列。

7. 权利要求6的载体，其中所述核酸分子选自下组：

- a) 包含SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列的核酸分子；
- b) 包含与SEQ ID NO:5、3、7、9、11、13、15、17、21、或23的核苷酸序列具有至少90%序列同一性的核苷酸序列的核酸分子，其中所述核苷酸序列编码分泌多肽；
- c) 编码包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列的聚多肽的核酸分子；
- d) 包含编码与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
- e) 包含编码与SEQ ID NO:20的氨基酸序列具有至少96%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
- f) 在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列；和，
- g) 在严格条件下在基本上全长里与编码选自下组的氨基酸序列的核苷酸序列杂交的核苷酸序列：SEQ ID NO:6、4、8、10、12、14、16、18、和22。

8. 权利要求7的载体，其中所述杂交条件包括约60°C至约70°C的温度。

9. 权利要求7或8的载体，其中所述杂交条件包括约68°C的温度。

10. 权利要求6的载体，其中所述核酸分子已经得到调节以反映选择用来表达所述核酸分子的宿主生物体的密码子偏爱。

11. 权利要求6的载体，其中所述分泌信号编码序列是可操作连接至编码感兴趣蛋白质或多肽的序列的。

12. 权利要求11的载体，其中所述感兴趣蛋白质或多肽对于表达所述感兴趣蛋白质或多肽的宿主生物体而言是天然的。

13. 权利要求6的载体，其中所述感兴趣蛋白质或多肽对于荧光假单胞菌而言是天然的。

14. 权利要求11的载体，其中所述感兴趣蛋白质或多肽衍生自对于表达所述

感兴趣蛋白质或多肽的宿主生物体而言不是天然的蛋白质或多肽。

15. 权利要求11的载体，其中所述感兴趣蛋白质或多肽来自不是假单胞菌的生物体。
16. 权利要求6的载体，其中所述感兴趣蛋白质或多肽衍生自真核生物体。
17. 权利要求16的载体，其中所述感兴趣蛋白质或多肽衍生自哺乳动物生物体。
18. 权利要求6的载体，其进一步包含所述信号多肽序列与所述感兴趣蛋白质或多肽序列之间的连接序列。
19. 权利要求18的载体，其中所述连接序列可被信号肽酶切割。
20. 权利要求6的载体，其中所述感兴趣蛋白质或多肽序列是可操作连接至第二信号序列的。
21. 权利要求20的载体，其中所述第二信号序列包含靶向至外膜分泌信号的序列。
22. 权利要求6的载体，其中所述载体进一步包含启动子。
23. 权利要求22的载体，其中所述启动子对于细菌宿主细胞是天然的。
24. 权利要求22的载体，其中所述启动子对于细菌宿主细胞不是天然的。
25. 权利要求23的载体，其中所述启动子对于大肠杆菌是天然的。
26. 权利要求22的载体，其中所述启动子是诱导型启动子。
27. 权利要求22的载体，其中所述启动子是*lac*启动子或*lac*启动子衍生物。
28. 一种重组细胞，其包含选自下组的分泌多肽的分泌信号编码序列：蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、和接纳甲基的趋化蛋白(ORF8124)分泌多肽。
29. 权利要求28的重组细胞，其中所述编码序列选自下组：
 - a) 包含SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列的核酸分子；
 - b) 包含与SEQ ID NO :5、3、7、9、11、13、15、17、21、或23的核苷酸序列具有至少90%序列同一性的核苷酸序列的核酸分子，其中所述核苷酸序列编码分泌多肽；
 - c) 编码包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或

- 24的氨基酸序列的聚多肽的核酸分子;
- d) 包含编码与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子, 其中所述聚多肽是分泌多肽;
 - e) 包含编码与SEQ ID NO:20的氨基酸序列具有至少96%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子, 其中所述聚多肽是分泌多肽;
 - f) 在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列; 和,
 - g) 在严格条件下在基本上全长里与编码选自下组的氨基酸序列的核苷酸序列杂交的核苷酸序列: SEQ ID NO:6、4、8、10、12、14、16、18、和22。
30. 权利要求28的细胞, 其中所述分泌信号编码序列是在表达载体中的。
31. 权利要求28的细胞, 其中所述分泌信号编码序列是可操作连接至编码感兴趣蛋白质或多肽的序列的。
32. 权利要求31的细胞, 其中所述细胞表达可操作连接至所述分泌信号多肽的所述感兴趣蛋白质或多肽。
33. 权利要求32的细胞, 其中所述蛋白质或多肽在所述细胞的周质区室中表达。
34. 权利要求32的细胞, 其中所述细胞中的酶自所述感兴趣蛋白质或多肽切割所述分泌信号多肽。
35. 权利要求28的细胞, 其中所述细胞衍生自细菌宿主。
36. 权利要求35的细胞, 其中所述宿主是假单胞菌。
37. 权利要求36的细胞, 其中所述宿主是荧光假单胞菌。
38. 权利要求35的细胞, 其中所述宿主是大肠杆菌。
39. 一种分离的多肽, 其包含选自下组的分泌多肽: 蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、和接纳甲基的趋化蛋白(ORF8124)分泌多肽。
40. 权利要求39的分离的多肽, 其中所述多肽选自下组:
- a) 包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列的多肽;

- b) 由SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列编码的多肽;
 - c) 包含与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%序列同一性的氨基酸序列的多肽, 其中所述多肽是分泌信号多肽;
 - d) 包含与SEQ ID NO:20的氨基酸序列具有至少96%序列同一性的氨基酸序列的多肽, 其中所述多肽是分泌信号多肽;
 - e) 由与SEQ ID NO:5、3、7、9、11、13、15、17、21、或23的核苷酸序列至少90%相同的核苷酸序列编码的多肽, 其中所述多肽是分泌信号多肽; 和,
 - f) 由在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列所编码的多肽。
41. 权利要求40的多肽, 其中所述杂交条件包括约60°C至约70°C的温度。
42. 权利要求39的多肽, 其中所述杂交条件包括约68°C的温度。
43. 权利要求39的多肽, 其中所述分泌信号多肽是可操作连接至感兴趣蛋白质或多肽的。
44. 权利要求43的多肽, 其中所述感兴趣蛋白质或多肽衍生自不是荧光假单胞菌生物体的生物体。
45. 用于表达感兴趣蛋白质或多肽的表达系统, 其包含:
- a) 宿主细胞; 和
 - b) 载体, 其包含编码可操作连接至选自下组的分泌信号多肽的所述感兴趣蛋白质或多肽的核酸分子: 蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、和接纳甲基的趋化蛋白(ORF8124)分泌多肽。
46. 权利要求45的表达系统, 其中所述分泌信号多肽由选自下组的核酸分子编码:
- a) 包含SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列的核酸分子;
 - b) 包含与SEQ ID NO:5、3、7、9、11、13、15、17、21、或23的核苷酸序列具有至少90%序列同一性的核苷酸序列的核酸分子, 其中所述核

- 苷酸序列编码分泌多肽;
- c) 编码包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列的聚多肽的核酸分子;
 - d) 包含编码与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子, 其中所述聚多肽是分泌多肽;
 - e) 包含编码与SEQ ID NO:20的氨基酸序列具有至少96%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子, 其中所述聚多肽是分泌多肽;
 - f) 在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列; 和,
 - g) 在严格条件下在基本上全长里与编码选自下组的氨基酸序列的核苷酸序列杂交的核苷酸序列: SEQ ID NO:6、4、8、10、12、14、16、18、和22。
47. 权利要求46的表达系统, 其中所述杂交条件包括约60°C至约70°C的温度。
48. 权利要求46的表达系统, 其中所述杂交条件包括约68°C的温度。
49. 权利要求45的表达系统, 其中所述宿主细胞表达可操作连接至所述分泌信号多肽的所述感兴趣蛋白质或多肽。
50. 权利要求49的表达系统, 其中所述感兴趣蛋白质或多肽在所述细胞的周质区室中表达。
51. 权利要求49的表达系统, 其中所述细胞中的酶自所述感兴趣蛋白质或多肽切割所述信号多肽。
52. 权利要求45的表达系统, 其中所述细胞衍生自细菌宿主。
53. 权利要求45的表达系统, 其中所述宿主是假单胞菌。
54. 权利要求53的表达系统, 其中所述宿主是荧光假单胞菌。
55. 权利要求52的表达系统, 其中所述宿主是大肠杆菌。
56. 权利要求45的表达系统, 其进一步包括发酵培养基。
57. 权利要求56的表达系统, 其中所述发酵培养基包含化学诱导物。
58. 用于在宿主细胞中表达重组蛋白的方法, 包括提供包含如下载体的宿主细胞, 所述载体编码可操作连接至选自下组的分泌信号多肽的感兴趣蛋白质或多肽: 蛋白质二硫键异构酶C (dsbC)、突变型磷酸结合蛋白(pbp*)、

蛋白质二硫键异构酶A (dsbA)、Bce、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(ORF5550)、甲苯耐受蛋白(Ttg2C)、和接纳甲基的趋化蛋白(ORF8124)分泌多肽。

59. 权利要求58的方法，其中所述分泌信号多肽由选自下组的核酸分子编码：

- a) 包含SEQ ID NO:5、1、3、7、9、11、13、15、17、19、21、或23的核苷酸序列的核酸分子；
- b) 包含与SEQ ID NO:5、3、7、9、11、13、15、17、21、或23的核苷酸序列具有至少90%序列同一性的核苷酸序列的核酸分子，其中所述核苷酸序列编码分泌多肽；
- c) 编码包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列的聚多肽的核酸分子；
- d) 包含编码与SEQ ID NO:6、4、8、10、12、14、16、18、22、或24的氨基酸序列具有至少90%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
- e) 包含编码与SEQ ID NO:20的氨基酸序列具有至少96%氨基酸序列同一性的聚多肽的核苷酸序列的核酸分子，其中所述聚多肽是分泌多肽；
- f) 在严格条件下在SEQ ID NO:5、3、7、9、11、13、15、17、或21的核苷酸序列的基本上全长里杂交的核苷酸序列；和，
- g) 在严格条件下在基本上全长里与编码选自下组的氨基酸序列的核苷酸序列杂交的核苷酸序列：SEQ ID NO:6、4、8、10、12、14、16、18、和22。

60. 权利要求59的方法，其中所述杂交条件包括约60℃至约70℃的温度。

61. 权利要求59的方法，其中所述杂交条件包括约68℃的温度。

62. 权利要求58的方法，其中所述细胞在矿物盐培养基中培养。

63. 权利要求58的方法，其中所述细胞以高细胞密度培养。

64. 权利要求63的方法，其中所述细胞以至少20g/L的细胞密度培养。

65. 权利要求58的方法，其进一步包括纯化所述重组蛋白。

66. 权利要求65的方法，其中通过亲和层析来纯化所述重组蛋白。

67. 权利要求58的方法，其中所述感兴趣蛋白质或多肽和所述分泌信号多肽的所述可操作连接可被对于所述宿主细胞而言天然的酶切割。

68. 权利要求67的方法, 其中所述分泌信号多肽是在表达期间中自所述感兴趣蛋白质或多肽切割的。
69. 权利要求58的方法, 其中所述感兴趣蛋白质或多肽对于衍生所述宿主细胞的生物体而言是天然的。
70. 权利要求58的方法, 其中所述感兴趣蛋白质或多肽对于荧光假单胞菌生物体而言是天然的。
71. 权利要求58的方法, 其中所述感兴趣蛋白质或多肽对于衍生所述宿主细胞的生物体而言不是天然的。
72. 权利要求58的方法, 其中所述感兴趣蛋白质或多肽衍生自不是假单胞菌的生物体。
73. 权利要求58的方法, 其中所述感兴趣蛋白质或多肽衍生自真核生物体。
74. 权利要求58的方法, 其中所述重组蛋白包含包括至少两个半胱氨酸残基的序列。
75. 权利要求58的方法, 其中在所述细胞中的所述重组蛋白中形成至少一个二硫键。
76. 权利要求58的方法, 进一步包括所述信号多肽序列和所述感兴趣蛋白质或多肽的序列之间的连接序列。
77. 权利要求69的方法, 其中至少50%的所述感兴趣蛋白质或多肽包含天然氨基末端。
78. 权利要求77的方法, 其中至少80%的所述感兴趣蛋白质或多肽包含天然氨基末端。
79. 权利要求78的方法, 其中至少90%的所述感兴趣蛋白质或多肽包含天然氨基末端。
80. 权利要求58的方法, 其中至少50%的所述重组蛋白是有活性的。
81. 权利要求80的方法, 其中至少80%的所述重组蛋白是有活性的。
82. 权利要求58的方法, 其中至少50%的所述重组蛋白在周质区室中表达。
83. 权利要求82的方法, 其中至少75%的所述重组蛋白在周质区室中表达。
84. 权利要求83的方法, 其中至少90%的所述重组蛋白在周质区室中表达。
85. 权利要求58的方法, 其中所述宿主细胞是假单胞菌细胞。
86. 权利要求85的方法, 其中所述细胞是荧光假单胞菌细胞。
87. 权利要求58的方法, 其中所述细胞是大肠杆菌细胞。

用于提高表达的细菌前导序列

发明领域

本发明在蛋白质生产领域内，具体针对靶向多肽用于生产正确加工的异源蛋白质的用途。

发明背景

超过150种重组生产的蛋白质和多肽已经获得美国食品和药品管理局(FDA)批准作为生物技术药物和疫苗使用，且另外的370种处于临床试验中。不像经由化学合成所生产的小分子治疗剂，蛋白质和多肽在活细胞中是最有效生成的。然而，当前在细菌中生产重组蛋白的方法通常生成不正确折叠的、聚集的或没有活性的蛋白质，而且许多类型的蛋白质要求使用已知方法低效率实现的次级修饰。

用已知方法的一个主要问题在于由聚集的蛋白质构成的包含体在细胞质中的形成，这在过量的蛋白质积累于细胞中时发生。重组蛋白生产中的另一个问题是为所表达的蛋白质建立正确的二级和三级构象。一种障碍是细菌细胞质活跃地抵抗二硫键形成，二硫键形成通常是正确蛋白质折叠的基础(Derman等 (1993) Science 262: 1744-7)。结果，许多重组蛋白（特别是那些具有真核起源的）在细菌中生产时是不正确折叠的，而且是没有活性的。

已经开发了许多尝试以增加正确折叠的蛋白质在重组系统中的生成。例如，研究者变更了发酵条件(Schein (1989) Bio/Technology, 7:1141-1149)，改变了启动子强度，或者使用了过量表达的伴侣蛋白(Hockney (1994) Trends Biotechnol. 12:456-463)，其能有助于防止包含体的形成。

增加正确折叠的蛋白质的收获的一种替代方法是自胞内环境分泌蛋白质。具有信号序列的多肽的最常见分泌形式牵涉Sec系统。Sec系统负责具有N端信号多肽的蛋白质穿过细胞质膜的输出（参见Agarraberes和Dice (2001) Biochim Biophys Acta. 1513:1-24; Muller等 (2001) Prog Nucleic Acid Res Mol. Biol. 66:107-157)。

已经开发了多种策略来将蛋白质从细胞分泌入上清液中。例如美国专利

No. 5,348,867; 美国专利No. 6,329,172; PCT公开文本号WO 96/17943; PCT公开文本号WO 02/40696; 及美国申请公开文本2003/0013150。用于提高表达的其它策略涉及将蛋白质靶向至周质。一些调查聚焦于非Sec型分泌(参见例如PCT公开文本号WO 03/079007; 美国公开文本No. 2003/0180937; 美国公开文本No. 2003/0064435; 及PCT公开文本号WO 00/59537)。然而, 大多数研究聚焦于用Sec型分泌系统分泌外源蛋白质。

已经记载了许多在表达重组多肽或蛋白质中使用的分泌信号。参见例如美国专利No. 5,914,254; 美国专利No. 4,963,495; 欧洲专利No. 0 177 343; 美国专利No. 5,082,783; PCT公开文本号WO 89/10971; 美国专利No. 6,156,552; 美国专利No. 6,495,357; 6,509,181; 6,524,827; 6,528,298; 6,558,939; 6,608,018; 6,617,143; 美国专利No. 5,595,898; 5,698,435; 及6,204,023; 美国专利No. 6,258,560; PCT公开文本号WO 01/21662, WO 02/068660和美国申请公开文本2003/0044906; 美国专利No. 5,641,671; 及欧洲专利No. EP 0,121, 352。

依赖于信号序列将蛋白质靶向出细胞质的策略通常产生不正确加工的蛋白质。这对于氨基端分泌信号(诸如那些导致经由Sec系统进行的分泌的)是尤其正确的。经由这种系统所加工的蛋白质通常或是保留分泌信号的一部分(这需要通常不正确地受到切割的连接元件), 或在末端处是截短的。

自上文所述技术显而易见的是, 已经开发了许多策略来将蛋白质靶向至宿主细胞的周质。然而, 已知的策略尚未导致正确加工的、有活性的重组蛋白(其可进行纯化以供治疗使用)的一致的高产率。先前策略的一个主要限制是具有差的分泌信号序列的蛋白质在不适当的细胞系统中的表达。

因此, 本领域中仍需要能够分泌和正确地加工重组多肽的改善的大规模表达系统以生产正确加工形式的转基因蛋白质。

发明概述

本发明提供了用于在细胞表达系统中生产高水平的正确加工的感兴趣蛋白质或多肽的改善的组合物和方法。具体地, 本发明提供了自细菌生物体衍生的分泌信号的新的氨基酸和核苷酸序列。在一个实施方案中, 本发明的分泌信号包括具有如下序列的分离的多肽, 所述序列是选自下组的荧光假单胞菌(*Pseudomonas fluorescens*, *P. fluorescens*)分泌多肽或与其基本上同源: 蛋白质二硫键异构酶C (protein disulfide isomerase C, dsbC)、突变型磷酸结合蛋

白(phosphate binding protein, pbp*)、蛋白质二硫键异构酶A (protein disulfide isomerase A, dsbA)、CupA2、CupB2、CupC2、NikA、FlgI、三十四肽重复家族蛋白(tetratricopeptide repeat family protein) (ORF5550)、甲苯耐受蛋白(toluene tolerance protein) (Ttg2C)、和接纳甲基的趋化蛋白(methyl accepting chemotaxis protein) (ORF8124)分泌多肽,以及具有生物学活性的变体、片段、和衍生物。在另一个实施方案中,本发明的分泌信号包括具有如下序列的分离的多肽,所述序列是凝结芽孢杆菌(*Bacillus coagulans*) Bce分泌信号序列或与其基本上同源。编码本发明信号序列的核苷酸序列在载体和表达系统中是有用的,用于促进所表达的感兴趣蛋白质或多肽向革兰氏阴性细菌的周质或向胞外环境中的靶向。

包含所述分泌信号序列的DNA构建体在宿主细胞中是有用的,用于表达重组蛋白。感兴趣蛋白质的核苷酸序列可操作连接至本文所述分泌信号。细胞可在周质区室中表达所述蛋白质。在某些实施方案中,细胞还可穿过细胞外壁向胞外分泌所表达的重组蛋白。宿主细胞包括真核细胞(包括酵母细胞、昆虫细胞、哺乳动物细胞、植物细胞等)和原核细胞(包括细菌细胞(诸如荧光假单胞菌、大肠杆菌)等)。可使用本发明的分泌多肽前导序列来表达任何感兴趣蛋白质,包括治疗用蛋白质、激素、生长因子、胞外受体或配体、蛋白酶、激酶、血液蛋白质、趋化因子、细胞因子、抗体等。

附图简述

图1描绘了dsbC SS-skp融合蛋白的表达构建体。

图2显示了约17kDa的Skp蛋白的表达(箭头)。标记2和3的条带与Skp蛋白一致。条带1表现出具有DNA结合蛋白(3691)和Skp两者。

图3是dsbC SS-skp在荧光假单胞菌中表达后在0和24小时后在标记2B-2(图3A)和2B-4(图3B)的样品的可溶性(S)和不溶性(I)级分中的分析。在图3A中,条带5、7和9是不溶性级分中的未加工dsbC-skp蛋白。条带6、8、和10是不溶性级分中的经加工dsbC-skp。条带1和3是可溶性级分中的经加工dsbC-skp。条带2和4是未知蛋白质。在图3B中,条带15、17、和19是不溶性级分中的未加工dsbC-skp蛋白。条带16、18、和20是不溶性级分中的经加工dsbC-skp。条带11和13是可溶性级分中的经加工dsbC-skp。条带12和14是未知蛋白质。

图4显示了表达DC694 (dsbA-PA83)后蛋白质积累的Western分析。通过Western分析评估0和24小时时可溶性(S)、不溶性(I)、和无细胞培养基(B)的积累。

图5显示了表达EP468-002.2 (dsbA)后蛋白质积累的Western分析。通过Western分析评估诱导后0和24小时时可溶性(S)和不溶性(I)蛋白质的积累。

图6展示了与pINS-008-5 (野生型pbp) 分泌信号相比pINS-008-3 (pbp*) 突变体的碱性磷酸酶活性。将细胞培养物调至1个OD₆₀₀单位, 然后通过添加4-甲基伞形酮(MUP)并在10分钟时测量荧光产物形成来测量PhoA活性。阴性对照含有MUP, 但没有细胞。

图7显示了表达pINS-008-3 (pbp*)和pINS0008-5 (野生型pbp) 后蛋白质积累的Western分析。通过Western分析评估胰岛素原-phoA在可溶性(Sol)、不溶性(Insol)、和胞外级分(Bro)中在I0、I16、和I40小时时的积累。将培养物的等分试样调至20个OD₆₀₀单位, 通过SDS-PAGE分开, 转移至滤纸, 并用针对胰岛素的抗体(鸡多克隆, Abcam产品目录编号ab14042)显现。

图8显示了EP484-003和EP484-004级分的SDS-PAGE分析。显示了SDS-PAGE分析的代表性结果。在中央处显示了分子量标志物(L)。标示了BSA标准品(BSA Stds.)。箭头指示了所诱导的条带。每道下方的是级分类型: 可溶性(Sol)、不溶性(Ins)、或无细胞培养基(CFB)。每道上方的是诱导时(I0)或诱导后24小时时(I24)的样品时间。在样品的每个分组下方显示了菌株号。EP484-004的I24可溶性级分中大的蛋白质条带对应于由Bce前导序列促进的增强的基因表达。

图9展示了Gal2 scFv表达的SDS-PAGE和Western分析。分析可溶性(S)和不溶性(I)级分。在每对泳道上方标示与Gal2融合的分选前导。在每个SDS-PAGE凝胶(顶部)或Western印迹(底部)的左侧描绘分子量标志物。箭头标示Gal2的迁移。

图10表示硫氧还蛋白(TrxA)表达的SDS-PAGE分析。分析可溶性级分。在每对泳道上方标示与TrxA融合的分选前导。在SDS-PAGE凝胶的左侧描绘分子量标志物。箭头标示未加工的(上方的箭头)和经加工的(下方的箭头)TrxA的迁移。

I. 综述

提供了用于在宿主细胞中生产高水平的正确加工的多肽的组合物和方法。具体地,提供了新的分泌信号,其促进可操作连接的感兴趣多肽向革兰氏阴性细菌的周质或向胞外环境中的靶向。出于本发明的目的,“分泌信号”、“分泌信号多肽”、“信号肽”或“前导序列”意指对于将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的肽序列(或编码所述肽序列的多核苷酸)。本发明的分泌信号序列包括选自下组的分泌多肽: pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、Ttg2C、和ORF8124分泌信号,及其片段和变体。在SEQ ID NO:2、4、6、8、10、12、14、16、18、20、22、和24中列出分泌信号的氨基酸序列。在SEQ ID NO:1、3、5、7、9、11、13、15、17、19、21、和23中分别提供了相应的核苷酸序列。本发明包含这些序列以及其片段和变体。

本发明的方法提供了在通常生成不正确折叠的、聚集的或没有活性的蛋白质的细菌中生产重组蛋白的当前方法的改良。另外,许多类型的蛋白质要求次级修饰,其使用已知方法仅低效率地实现。本文中的方法通过自胞内环境分泌蛋白质来增加正确折叠的蛋白质的收获。在革兰氏阴性细菌中,自细胞质分泌的蛋白质可终止于周质间隙中,附着至外膜,或者在胞外培养基中。该方法还避免由聚集的蛋白质构成的包含体。向周质间隙中的分泌还具有公知的效果,即促进正确的二硫键形成(Bardwell等 (1994) *Phosphate Microorg.* 270-5; Manoil (2000) *Methods in Enzymol.* 326: 35-47)。重组蛋白分泌的其它益处包括更有效的蛋白质分离;转基因蛋白质的正确折叠和二硫键形成,导致处于活性形式的蛋白质的百分比增加;包含体的形成减少和对宿主细胞的毒性降低;和处于可溶形式的重组蛋白的百分比增加。将感兴趣蛋白质分泌入培养物培养基中的潜力还可潜在地促进用于蛋白质生产的连续培养而非分批培养。

革兰氏阴性细菌已经发展出用于蛋白质穿过其双膜的主动输出的许多系统。这些分泌路径包括例如: ABC (I型) 途径、Path/Fla (III型) 途径、和Path/Vir (IV型) 途径,用于穿过质膜和外膜两者的1步转位; Sec (II型)、Tat、MscL和Holins途径,用于穿过质膜的转位; 和Sec-加上-菌毛引座孔蛋白(fimbrial usher porin, FUP)、Sec-加上-自运输物(autotransporter, AT)、Sec-加上-双配偶分泌(two partner secretion, TPS)、Sec-加上-主端分枝(main

terminal branch, MTB)、和Tat-加上-MTB途径,用于穿过质膜和外膜的2步转位。不是所有细菌都具有所有这些分泌途径。

三种蛋白质系统(类型I、III和IV)在单一能量偶联步骤中分泌蛋白质穿过两膜。四种系统(Sec、Tat、MscL和Holins)仅分泌穿过内膜,而其它四种系统(MTB、FUP、AT和TPS)仅分泌穿过外膜。

在一个实施方案中,本发明的信号序列利用Sec分泌系统。Sec系统负责具有N端信号多肽的蛋白质穿过细胞质膜的输出(参见Agarraberes和Dice (2001) *Biochim Biophys Acta*. 1513:1-24; Muller等 (2001) *Prog Nucleic Acid Res Mol. Biol.* 66:107-157)。在原核生物和真核生物中普遍地找到Sec家族的蛋白质复合物。细菌Sec系统由运输蛋白、伴侣蛋白(SecB)或信号识别颗粒(SRP)和信号肽酶(SP酶I和SP酶II)组成。大肠杆菌中的Sec运输复合物由三种整合内膜蛋白(SecY、SecE和SecG)和细胞质ATP酶SecA组成。SecA募集SecY/E/G复合物以形成有活性的转位通道。伴侣蛋白SecB结合初生的多肽链以阻止其折叠,并将其靶向至SecA。随后线性多肽链被运输穿过SecYEG通道,并且在切割信号多肽后,该蛋白质在周质中折叠。三种辅助蛋白(SecD、SecF和YajC)形成一种复合物,其对于分泌不是必需的,但是在许多条件(特别在低温时)下刺激分泌,高至10倍。

运输入周质中(即经由II型分泌系统)的蛋白质还可在进一步的步骤中输出入胞外培养基中。其机制一般是经由自运输物、双配偶分泌系统、主端分枝系统或菌毛引座孔蛋白。

革兰氏阴性细菌中的12种已知分泌系统中,已知8种利用作为所表达的蛋白质的一部分找到的靶向信号多肽。这些信号多肽与分泌系统的蛋白质相互作用,使得细胞正确地将蛋白质送往其合适的目的地。这8种基于信号多肽的分泌系统中的5种是那些牵涉Sec系统的。这5种被说成是牵涉Sec依赖性细胞质膜转位,并且它们在那里起作用的信号多肽可称为Sec依赖性分泌信号。开发合适的分泌信号中的难题之一是确保信号得到合适的表达,并自所表达的蛋白质切除。

用于sec途径的信号多肽一般由下列三种域组成:(i)带正电荷的n区,(ii)疏水性的h区,和(iii)不带电荷的但极性的c区。信号肽酶的切割位点位于c区中。然而,信号序列的程度保守和长度以及切割位点位置在不同蛋白质间可有所不同。

Sec依赖性蛋白质输出的特征(signature)是在所输出的蛋白质中存在短的(约30个氨基酸)、主要疏水的氨基端信号序列。信号序列有助于蛋白质输出,并在所输出的蛋白质到达周质时被周质信号肽酶切去。典型的N端Sec信号多肽含有具有至少一个精氨酸或赖氨酸残基的N域,接着是含有一段疏水残基的域,和含有信号肽酶的切割位点的C域。

已经开发了细菌蛋白质生产系统,其中转基因蛋白质构建体被改造为含有感兴趣蛋白质和分泌信号两者的融合蛋白,试图将蛋白质靶向出细胞质。

荧光假单胞菌已经证明是用于生产多种蛋白质的改良平台,并且已经自此生物体鉴定出数种有效的分泌信号(参见美国申请公开文本号20060008877,通过提及而完整收入本文)。荧光假单胞菌以比通常在其它细菌表达系统中所看到的水平要高的水平生成正确加工形式的外源蛋白质,并以较高的水平将这些蛋白质运输至细胞周质,导致完全加工的重组蛋白的回收增加。因此,在一个实施方案中,本发明提供了用于通过表达与分泌信号连接的靶蛋白来在荧光假单胞菌细胞中生产外源蛋白质的方法。

本发明的分泌信号序列在假单胞菌中是有用的。与其它细菌表达系统相比,假单胞菌系统为多肽和酶的商业表达提供优势。具体地,荧光假单胞菌已经鉴定为有利的表达系统。荧光假单胞菌涵盖一组常见的、非病原性腐生生物,其拓殖土壤、水和植物表面环境。自荧光假单胞菌衍生的商品化酶已经用于降低环境污染,用作去污添加剂,及用于立体选择性水解。荧光假单胞菌还在农业上用于控制病原体。美国专利号4,695,462记载了重组细菌蛋白在荧光假单胞菌中的表达。在1985年和2004年之间,许多公司利用荧光假单胞菌的农业用途来生产杀虫的(pesticidal)、杀昆虫的(insecticidal)、和杀线虫的毒素,以及利用特定毒性序列和遗传操作来增强这些的表达。参见例如PCT公开文本号WO 03/068926和WO 03/068948; PCT公开文本号WO 03/089455; PCT申请号WO 04/005221; 及美国专利公开文本号20060008877。

II. 组合物

A. 分离的多肽

在本发明的一个实施方案中,提供了分离的多肽,其中所述分离的多肽是对将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的新的分泌信号。在一个实施方案中,所述多肽具有

如下氨基酸序列，其是pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、Ttg2C、或ORF8124分泌信号、或其片段或变体，或与其基本上同源。在另一个实施方案中，此分离的多肽是分泌信号和感兴趣蛋白质或多肽的融合蛋白。

在另一个实施方案中，多肽序列是SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24中所列的分泌信号多肽、或者由SEQ ID NO:1、3、5、7、9、11、13、15、17、19、21、或23中所列的多核苷酸序列所编码的分泌信号多肽，或与其基本上同源。在另一个实施方案中，所述多肽序列包含SEQ ID NO:2的至少第2位-第24位氨基酸、SEQ ID NO:4的至少第2位-第22位氨基酸、SEQ ID NO:6的至少第2位-第21位氨基酸、SEQ ID NO:8的至少第2位-第33位氨基酸、SEQ ID NO:10的至少第2位-第25位氨基酸、SEQ ID NO:12的至少第2位-第24位氨基酸、SEQ ID NO:14的至少第2位-第23位氨基酸、SEQ ID NO:16的至少第2位-第21位氨基酸、SEQ ID NO:18的至少第2位-第21位氨基酸、SEQ ID NO:20的至少第2位-第21位氨基酸、SEQ ID NO:22的至少第2位-第33位氨基酸、或SEQ ID NO:24的至少第2位-第39位氨基酸。在又一个实施方案中，所述多肽序列包含SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的片段，其自氨基端截短1个、2个、3个、4个、5个、6个、7个、8个、9个、或10个氨基酸，但保留生物学活性，即分泌信号活性。

在一个实施方案中，同源多肽的氨基酸序列是给定最初多肽的变体，其中变体序列可通过用其它氨基酸残基替换最初多肽的高至或约30%的氨基酸残基来获得，包括高至约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、或30%，前提是变体保留最初多肽的想要的功能。基本上同源的变体氨基酸会是与给定多肽至少约70%、至少约75%、至少约80%、约85%、约90%、约95%、约96%、约97%、约98%、或至少约99%同源的。变体氨基酸可以以多种方式获得，包括SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的一个或多个氨基酸的氨基酸替代、删除、截短和插入，包括多至约1处、约2处、约3处、约4处、约5处、约6处、约7处、约8处、约9处、约10处、约15处、约20处、约25处、或更多处氨基酸替代、删除或插入。

通过“基本上同源的”或“基本上类似的”意指使用本文中所描述的比

对程序之一使用标准参数与参照序列相比具有至少约60%或65%序列同一性、约70%或75%序列同一性、约80%或85%序列同一性、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%或更大序列同一性的氨基酸或核苷酸序列。本领域技术人员会认识到，通过考虑密码子简并性、氨基酸相似性、阅读框定位等，可适当地调整这些数值以确定由两种核苷酸序列所编码的蛋白质的相应同一性。

例如，优选的是，可在一个或多个预测的、优选非必需的氨基酸残基处进行保守的氨基酸替代。“非必需的”氨基酸残基是可自分泌信号多肽的野生型序列改变而不改变生物学活性的残基，而“必需的”氨基酸残基是生物学活性所需要的。“保守氨基酸替代”是其中氨基酸残基用具有类似侧链的氨基酸残基替换的。表1中列出了保守和半保守氨基酸残基的家族。

表1: 类似氨基酸替代组	
保守组(8)	半保守组(7)
Arg, Lys	Arg, Lys, His
Asp, Gln	Asn, Asp, Gly, Gln
Asn, Glu	
Ile, Leu, Val	Ile, Leu, Val, Met, Phe
Ala, Gly	Ala, Gly, Pro, Ser, Thr
Ser, Thr	Ser, Thr, Tyr
Phe, Tyr	Phe, Trp, Tyr
Cys (非半胱氨酸), Ser	Cys (非半胱氨酸), Ser, Thr

本发明所涵盖的变异型蛋白质是有生物学活性的，即它们继续拥有天然蛋白质的想要的生物学活性，即保留分泌信号活性。通过“保留活性”意指变体会具有天然蛋白质的至少约30%、至少约50%、至少约70%、至少约80%、约90%、约95%、约100%、约110%、约125%、约150%、至少约200%或更大分泌信号活性。

B. 分离的多核苷酸

本发明还包括具有如下序列的分离的核酸，所述序列编码对于将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的新的分泌信号。在一个实施方案中，所述分离的多核苷酸编码与pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、

Ttg2C、或ORF8124分泌信号多肽基本上同源的多肽序列。在另一个实施方案中，本发明提供了如下核酸，其编码与SEQ ID NO:2的至少第2位-第24位氨基酸、SEQ ID NO:4的至少第2位-第22位氨基酸、SEQ ID NO:6的至少第2位-第21位氨基酸、SEQ ID NO:8的至少第2位-第33位氨基酸、SEQ ID NO:10的至少第2位-第25位氨基酸、SEQ ID NO:12的至少第2位-第24位氨基酸、SEQ ID NO:14的至少第2位-第23位氨基酸、SEQ ID NO:16的至少第2位-第21位氨基酸、SEQ ID NO:18的至少第2位-第21位氨基酸、SEQ ID NO:20的至少第2位-第21位氨基酸、SEQ ID NO:22的至少第2位-第33位氨基酸、或SEQ ID NO:24的至少第2位-第39位氨基酸基本上同源的多肽序列，或者提供了与SEQ ID NO:1、3、5、7、9、11、13、15、17、19、20、21、或23基本上同源的核酸，包括其生物学活性变体和片段。在另一个实施方案中，所述核酸序列与SEQ ID NO:1、3、5、7、9、11、13、15、17、19、20、21、或23的序列至少约60%、至少约65%、至少约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%、或至少约99%相同。在另一个实施方案中，所述核酸编码如下多肽，其与SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的氨基酸序列至少约70%、至少约75%、至少约80%、约85%、约90%、约95%、约96%、约97%、约98%、或至少约99%相同。

本发明优选的分泌信号多肽由与SEQ ID NO:1或3的核苷酸序列基本上同源的核苷酸序列编码。使用诸如PCR、杂交等方法，可鉴定相应的分泌信号多肽序列，此类序列与本发明的序列基本上相同。参见例如Sambrook J. 和Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)及Innis等 (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, NY)。变异型核苷酸序列还包括合成方式衍生的核苷酸序列，其已经例如通过使用定点诱变生成，但其仍编码本发明中所公开的分泌信号多肽，如下文所讨论的。本发明所涵盖的变异型分泌信号多肽是有生物学活性的，即它们继续拥有天然蛋白质的想要的生物学活性，即保留分泌信号活性。通过“保留活性”意指变体会具有至少约30%、至少约50%、至少70%、至少约80%、至少约85%、至少约90%、至少约95%、约96%、约97%、约98%、至少约99%或更大的天然分泌信号多肽活性。用于测量分泌信号多肽活性的方法在本文中的别处有所讨论。

熟练技术人员会进一步理解的是,变化可通过突变而导入本发明的核苷酸序列中,由此导致所编码分泌信号多肽的氨基酸序列的变化,而不改变分泌信号多肽的生物学活性。如此,可如下创建变异型分离的核酸分子,即将一处或多处核苷酸替代、添加、或删除导入本文中所公开的相应核苷酸序列中,使得将一处或多处氨基酸替代、添加或删除导入所编码的蛋白质中。可通过标准技术(诸如定点诱变和PCR介导的诱变)来导入突变。本发明也涵盖此类变异型核苷酸序列。

C. 核酸和氨基酸同源性

依照本领域中公知的多种方法之任一种来测定核酸和氨基酸序列同源性。有用的序列比对和同源性测定方法的例子包括那些下文所述的。

可使用美国国立生物技术信息中心(NCBI)程序MegaBLAST(目前可在<http://www.ncbi.nlm.nih.gov/BLAST/>获得的)来实施类似序列的比对和搜索。以百分比同一性选项设置在例如氨基酸序列的70%或设置在例如核苷酸序列的90%使用此程序会鉴定出那些与查询序列具有70%或90%或更大序列同一性的序列。本领域中已知的其它软件也可用于比对和/或搜索类似的序列,例如与含有依照本发明的分泌信号序列的信息串至少70%或90%相同的序列。例如,用于比较以鉴定与查询序列至少70%或90%相同的序列的序列比对可如下进行,即使用例如可在GCG序列分析软件包(可获自Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705)中获得的GAP、BESTFIT、BLAST、FASTA、和TFASTA程序,其中缺省参数如在那里所指定的,加上设置在想要的百分比的序列同一性程度的参数。还有,例如可使用CLUSTAL程序(可在来自Intelligenetics, Mountain View, Cal.的PC/Gene软件包中获得的)。

这些和其它序列比对方法在本领域中是公知的,并且可通过手动比对、通过目视检查、或通过序列比对算法(诸如由上文所述程序所体现的那些中的任一种)的手动或自动应用来进行。多种有用的算法包括例如W. R. Pearson和D. J. Lipman, Proc. Natl. Acad. Sci. USA 85:2444-48(1988年4月)中所记载的相似性搜索方法; T. F. Smith和M. S. Waterman, 于Adv. Appl. Math. 2:482-89(1981)和于J. Molec. Biol. 147:195-97(1981)中所记载的局部同源性方法; S. B. Needleman和C. D. Wunsch, J. Molec. Biol. 48(3):443-53(1970年3月)中所记载的同源性比对方法; 和例如由W. R. Pearson, 于Genomics

11(3):635-50 (1991年11月); 由W. R. Pearson, 于Methods Molec. Biol. 24:307-31和25:365-89 (1994); 及由D. G. Higgins和P. M. Sharp, 于Comp. Appl'ns in Biosci. 5:151-53 (1989)和于Gene 73(1):237-44 (1988年12月15日)所记载的多种方法。

除非另有说明, 会使用GAP第10版(其使用Needleman和Wunsch (1970)见上文的算法)来测定序列同一性或相似性, 其使用如下参数进行: 核苷酸序列的%同一性和%相似性使用GAP权重(Weight) 50和长度权重(Length Weight) 3, 及nwsgapdna.cmp得分矩阵; 氨基酸序列的%同一性或%相似性使用GAP权重8和长度权重2, 及BLOSUM62得分程序。也可使用等同的程序。通过“等同的程序”意指与由GAP第10版所产生的相应比对相比时, 对任何两种所讨论的序列产生具有相同核苷酸残基匹配和相同百分比序列同一性的比对的任何序列比较程序。在多个实施方案中, 序列比较贯穿整个查询序列或目标序列或两者进行。

D. 杂交条件

在本发明的另一方面, 提供了与具有如下序列的分离的核酸杂交的核酸, 所述序列编码具有与pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、Ttg2C、或ORF8124分泌信号多肽基本上类似的多肽。在某些实施方案中, 杂交的核酸会在高严格条件下结合。在多个实施方案中, 杂交贯穿编码分泌信号多肽的核苷酸序列的基本上全长发生, 例如, 贯穿SEQ ID NO:1、3、5、7、9、11、13、15、17、19、21、或23中的一项或多项的基本上全长发生。若核酸分子在SEQ ID NO:1、3、5、7、9、11、13、15、17、19、21、或23中的一项或多项的全长的至少80%、全长的至少85%、至少90%、或至少95%里杂交, 则该核酸分子与本文中所公开的编码分泌信号的核苷酸序列的“基本上全长”杂交。除非另有说明, “基本上全长”指编码分泌信号的核苷酸序列的全长的至少80%, 其中长度以连续的核苷酸测量(例如杂交至SEQ ID NO:3的至少53个连续的核苷酸、SEQ ID NO:5的至少51个连续的核苷酸、SEQ ID NO:7的至少80个连续的核苷酸等)。

在一种杂交方法中, 可使用整个或部分的编码分泌信号多肽的核苷酸序列来筛选cDNA或基因组文库。用于构建此类cDNA和基因组文库的方法是本领域中普遍知道的, 并披露于Sambrook和Russell, 2001。所谓杂交探针可以是基因组DNA片段、cDNA片段、RNA片段或其它寡核苷酸, 而且可以用可

检测的基团（诸如 ^{32}P ）、或任何其它可检测的标志物（诸如其它放射性同位素、荧光化合物、酶、或酶辅因子）标记。可通过标记合成的寡核苷酸来制备供杂交用的探针，所述合成的寡核苷酸基于本文中所公开的已知的编码分泌信号多肽的核苷酸序列。另外还可使用基于核苷酸序列或所编码的氨基酸序列中的保守核苷酸或氨基酸残基而设计的简并引物。探针通常包含在严格条件下与本发明的编码分泌信号多肽的核苷酸序列或其片段或变体的至少约10个、至少约15个、至少约16个、17个、18个、19个、20个、或更多个连续核苷酸杂交的核苷酸序列区域。用于制备供杂交用的探针的方法是本领域中普通知道的，并披露于Sambrook和Russell, 2001，通过提及而收入本文。

在杂交技术中，整个或部分的已知核苷酸序列作为探针使用，所述探针与自选定生物体所克隆的基因组DNA片段或cDNA片段的群体（即基因组或cDNA文库）中所存在的其它相应核苷酸序列选择性杂交。杂交探针可以是基因组DNA片段、cDNA片段、RNA片段或其它寡核苷酸，而且可以用可检测的基团（诸如 ^{32}P ）、或任何其它可检测的标志物标记。如此，例如可通过标记合成的寡核苷酸来制备供杂交用的探针，所述合成的寡核苷酸基于本文中编码分泌信号多肽的核苷酸序列。用于制备供杂交用的探针和用于构建cDNA和基因组文库的方法是本领域中普遍知道的，并且披露于Sambrook等 (1989) *Molecular Cloning: A Laboratory Manual*（第2版, Cold Spring Harbor Laboratory Press, Plainview, New York）。

例如，本文中所公开的整个编码分泌信号多肽的核苷酸序列或其一个或多个部分可作为如下探针使用，所述探针能够与编码分泌信号多肽的相应核苷酸序列和信使RNA特异性杂交。为了在多种条件下实现特异性杂交，此类探针包括独特的且优选长至少约10个核苷酸或长至少约15个核苷酸的序列。此类探针可用于通过PCR来自选定生物体扩增相应的编码分泌信号多肽的核苷酸序列。此技术可用于自想要的生物体分离别的编码序列或用作诊断测定法以测定编码序列在生物体中的存在。杂交技术包括对铺板的DNA文库（或是噬菌斑或菌落）的杂交筛选；参见例如Sambrook等 (1989) *Molecular Cloning: A Laboratory Manual*（第2版, Cold Spring Harbor Laboratory Press, Plainview, New York）。

此类序列的杂交可在严格条件下实施。通过“严格条件”或“严格杂交条件”意指探针会以比其它序列大的可检测程度（例如超过背景的至少2

倍)与其靶序列杂交的条件。严格条件是序列依赖性的,而且在不同情况中会是不同的。通过控制杂交和/或清洗条件的严格性,可鉴定出与探针100%互补的靶序列(同源探查)。或者,可调整严格条件以容许序列中的一些错配,使得检测出较低程度的相似性(异源探查)。一般而言,探针的长度小于约1000个核苷酸,优选长度小于500个核苷酸。

通常,严格条件会是那些如下的,其中盐浓度小于约1.5M Na离子,通常约0.01至1.0M Na离子浓度(或其它盐类),于pH7.0至8.3,且温度是至少约60°C,优选约68°C。还可通过添加去稳定剂(诸如甲酰胺)来实现严格条件。例示性的低严格条件包括用30至35% 甲酰胺、1M NaCl、1% SDS(十二烷基硫酸钠)的缓冲溶液于37°C进行的杂交,和在1X至2X SSC(20X SSC = 3.0M NaCl/0.3M 柠檬酸三钠)中于50至55°C进行的清洗。例示性的中等严格条件包括在40至45% 甲酰胺、1.0M NaCl、1% SDS中于37°C进行的杂交,和在0.5X至1X SSC中于55至60°C进行的清洗。例示性的高严格条件包括在50% 甲酰胺、1M NaCl、1% SDS中于37°C进行的杂交,和在0.1X SSC中于60至68°C进行的清洗。任选地,清洗缓冲液可包含约0.1%至约1% SDS。杂交的持续时间一般小于约24小时,通常约4至约12小时。

特异性通常是杂交后清洗的函数,至关重要的因素是最终清洗溶液的离子强度和温度。对于DNA-DNA杂合物, T_m 可自Meinkoth和Wahl(1984) Anal. Biochem. 138:267-284的方程式估算: $T_m = 81.5^{\circ}\text{C} + 16.6 (\log M) + 0.41 (\% \text{GC}) - 0.61 (\% \text{ form}) - 500/L$; 其中M是单价阳离子的摩尔浓度,%GC是DNA中鸟嘌呤和胞嘧啶核苷酸的百分比,% form是杂交溶液中甲酰胺的百分比,而L是以碱基对计的杂合物长度。 T_m 是50%的互补靶序列杂交至完全匹配的探针的温度(在规定的离子强度和pH下)。每1%的错配使 T_m 下降约1°C; 如此,可调整 T_m 、杂交和/或清洗条件以杂交至具有想要的同一性的序列。例如,若寻找具有大于等于90%同一性的序列,则可使 T_m 下降10°C。一般而言,严格条件选择为比特定序列及其互补物在规定离子强度和pH的热熔点(T_m)低约5°C。然而,极严格条件可利用比热熔点(T_m)低1、2、3或4°C的杂交和/或清洗; 中等严格条件可利用比热熔点(T_m)低6、7、8、9或10°C的杂交和/或清洗; 低严格条件可利用比热熔点(T_m)低11、12、13、14、15、或20°C的杂交和/或清洗。使用所述方程式、杂交和清洗组合物、和想要的 T_m ,普通技术人员会理解,内在描述了杂交和/或清洗溶液的严格性的变化。若想要的错配

程度导致 T_m 小于45°C（水溶液）或32°C（甲酰胺溶液），则优选提高SSC浓度以使得可使用更高的温度。对核酸杂交的广泛指导记载于Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes*, 部分I, 章2 (Elsevier, New York); 及Ausubel等编 (1995) *Current Protocols in Molecular Biology*, 章2 (Greene Publishing and Wiley-Interscience, New York)。参见Sambrook等 (1989) *Molecular Cloning: A Laboratory Manual* (第2版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)。

E. 密码子选择

可根据宿主生物体的密码子选择来调整本文中所公开的核酸序列。密码子选择或密码子偏爱在本领域中是公知的。选定的编码序列可通过改变其遗传密码来改良以匹配由细菌宿主细胞所采用的，并可增强其密码子序列来更好地接近宿主所采用的。可依照本领域普通技术人员已知的多种方法之任一种（例如寡核苷酸指导的诱变）来实施遗传密码选择和密码子频率增强。帮助此方法的有用的在线因特网资源包括例如：(1) Kazusa DNA研究所(2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818 Japan)的密码子选择数据库，并可在 www.kazusa.or.jp/codon 获得；和(2)可自NCBI分类学数据库在 www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c 获得的遗传密码表。例如，假单胞菌物种报告为利用NCBI分类学位置的遗传密码翻译表11，而在Kazusa位置报告为展现出www.kazusa.or.jp/codon/cgi-bin处所显示表格的密码子用法频率。认识到可关于密码子选择来调整分泌信号多肽、本文中别处所描述的感兴趣多肽或两者的编码序列。

F. 表达载体

本发明的另一个实施方案包括表达载体，其包括编码对于将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的新的分泌多肽的核酸。在一个实施方案中，载体包含可操作连接至启动子的，编码与本文中所公开的分泌信号多肽基本上类似的多肽的多核苷酸序列。可表达的编码序列会可操作附着至能够在选定的宿主细胞中运行的转录启动子，以及所有其它所需要的转录和翻译调控元件。

术语“可操作连接的”指任何如下的构造，其中转录和任何翻译调控元件以相对于编码序列的这样一种或多种布置共价附着于编码序列以使得在

宿主细胞中和通过宿主细胞的作用，调控元件可指导编码序列的表达。

载体通常会包含一种或多种表型选择标志和复制起点以确保载体的维持，并在想要时提供宿主内的扩增。依照本公开内容的适于转化的宿主包括假单胞菌属内的多种物种，并且特别优选的是荧光假单胞菌的宿主细胞株。

在一个实施方案中，载体进一步包含用于表达可操作连接至本文中所公开的分泌信号的感兴趣蛋白质或多肽的编码序列。可自多核苷酸表达重组蛋白和多肽，在所述多核苷酸中靶多肽编码序列可操作连接至前导序列和转录和翻译调控元件以形成功能性基因，宿主细胞可自该功能性基因表达蛋白质或多肽。编码序列可以是靶多肽的天然编码序列（若可获得的话），但是会更优选已经为了在选定的表达宿主细胞中的应用进行选择、改善或优化的编码序列：例如，通过合成基因以反映宿主物种的密码子使用偏爱。在本发明的一个实施方案中，宿主物种是荧光假单胞菌，并在设计信号序列和/或蛋白质或多肽序列两者时考虑荧光假单胞菌的密码子偏爱。将一种或多种基因在一种或多种载体内构建或插入一种或多种载体中，然后可将所述载体转化入表达宿主细胞中。

其它调控元件可包括在载体（也称为“表达构建体”）中。此类元件包括但不限于例如转录增强子序列、翻译增强子序列、其它启动子、激活子、翻译起始和终止信号、转录终止子、顺反子调节物、多顺反子调节物、标签序列（诸如核苷酸序列“标签”和“标签”多肽编码序列），其促进所表达的多肽的鉴定、分开、纯化和/或分离。

在另一个实施方案中，表达载体进一步包含与分泌信号的编码序列或与感兴趣蛋白质或多肽的编码序列邻近的标签序列。在一个实施方案中，此标签序列容许纯化蛋白质。标签序列可以是亲和标签，诸如六组氨酸亲和标签。在另一个实施方案中，亲和标签可以是谷胱甘肽-S-转移酶分子。标签还可以是荧光分子，诸如YFP或GFP，或此类荧光蛋白的类似物。标签还可以是对纯化有用的已知结合配偶体的已知抗原或配体、或抗体分子的一部分。

在蛋白质编码序列之外，依照本发明的蛋白质编码基因还可包括与其可操作连接的下列调控元件：启动子、核糖体结合位点(RBS)、转录终止子、翻译起始和终止信号。有用的RBS可自在依照本发明的表达系统中可用作宿主细胞的任何物种获得，优选自选定的宿主细胞获得。许多特有的和多种共有的RBS是已知的，例如那些记载于D. Frishman等，Starts of bacterial genes:

estimating the reliability of computer predictions, *Gene* 234(2):257-65 (1999年7月8日); 及B. E. Suzek等, A probabilistic method for identifying start codons in bacterial genomes, *Bioinformatics* 17(12):1123-30 (2001年12月)的和由它们提及的。另外, 可使用天然的或合成的RBS, 例如那些记载于EP 0207459 (合成的RBS); O. Ikehata等, Primary structure of nitrile hydratase deduced from the nucleotide sequence of a *Rhodococcus* species and its expression in *Escherichia coli*, *Eur. J. Biochem.* 181(3):563-70 (1989)的(天然RBS序列AAGGAAG)。本发明中有用的方法、载体、和翻译和转录元件及其它元件的进一步的例子记载于例如Gilroy的美国专利No. 5,055,294和Gilroy等的美国专利No. 5,128,130; Rammner等的美国专利No. 5,281,532; Barnes等的美国专利No. 4,695,455和4,861,595; Gray等的美国专利No. 4,755,465; 及Wilcox的美国专利No. 5,169,760。

通过将增强子序列插入载体或质粒中来增加编码本发明蛋白质的DNA的转录。典型的增强子是大小通常约10-300bp的DNA顺式作用元件, 其作用于启动子以增加其转录。例子包括多种假单胞菌增强子。

一般而言, 重组表达载体会包括复制起点和容许宿主细胞转化的选择标志和自高度表达的基因衍生以指导下游结构序列转录的启动子。此类启动子可以自编码酶(诸如3-磷酸甘油酸激酶(PGK)、酸性磷酸酶或热休克蛋白等)的操纵子衍生。异源结构序列以合适的状态与翻译起始和终止序列和优选能够指导所翻译的多肽分泌的分泌序列装配在一起。任选地, 异源序列可编码包含赋予想要的特性(例如所表达重组产物的稳定化或简化的纯化)的N端鉴定多肽的融合多肽。

用于在宿主细胞中表达重组蛋白的载体在本领域中是已知的, 并且可使用这些中的任一种来表达依照本发明的基因。此类载体包括例如质粒、粘粒、和噬菌体表达载体。有用的质粒载体的例子包括但不限于表达质粒pBBR1MCS、pDSK519、pKT240、pML122、pPS10、RK2、RK6、pRO1600、和RSF1010。此类有用的载体的其它例子包括那些记载于例如: N. Hayase, 于*Appl. Envir. Microbiol.* 60(9):3336-42 (1994年9月); A. A. Lushnikov等, 于*Basic Life Sci.* 30:657-62 (1985); S. Graupner和W. Wackemagel, 于*Biomolec. Eng.* 17(1):11-16. (2000年10月); H. P. Schweizer, 于*Curr. Opin. Biotech.* 12(5):439-45 (2001年10月); M. Bagdasarian和K. N. Timmis, 于*Curr. Topics*

Microbiol. Immunol. 96:47-67 (1982); T. Ishii等, 于FEMS Microbiol. Lett. 116(3):307-13 (1994年3月1日); I. N. Olekhovich和Y. K. Fomichev, 于Gene 140(1):63-65 (1994年3月11日); M. Tsuda和T. Nakazawa, 于Gene 136(1-2):257-62 (1993年12月22日); C. Nieto等, 于Gene 87(1):145-49 (1990年3月1日); J. D. Jones和N. Gutterson, 于Gene 61(3):299-306 (1987); M. Bagdasarian等, 于Gene 16(1-3):237-47 (1981年12月); H. P. Schweizer等, 于Genet. Eng. (NY) 23:69-81 (2001); P. Mukhopadhyay等, 于J. Bact. 172(1):477-80 (1990年1月); D. O. Wood等, 于J. Bact. 145(3):1448-51 (1981年3月); 及R. Holtwick等, 于Microbiology 147(Pt 2):337-44 (2001年2月)的。

可用于宿主细胞的、包含本发明的分泌信号构建体的表达载体的进一步例子包括那些列于表2中的、自指示的复制子衍生的。

表2: 有用的表达载体的例子	
复制子	载体
PPS10	PCN39, PCN51
RSF1010	PKT261-3
	PMMB66EH
	PEB8
	PPLGN1
	PMYC1050
RK2/RP1	PRK415
	PJB653
PRO1600	PUCP
	PBSP

表达质粒RSF1010记载于例如F. Heffron等, 于Proc. Nat'l Acad. Sci. USA 72(9):3623-27 (1975年9月)及K. Nagahari和K. Sakaguchi, 于J. Bact. 133(3):1527-29 (1978年3月)。质粒RSF1010及其衍生物是本发明中特别有用的载体。例示性的、有用的RSF1010衍生物(其在本领域中是已知的)包括例如pKT212、pKT214、pKT231及相关质粒, 和pMYC1050及相关质粒(参见例如Thompson等的美国专利No. 5,527,883和5,840,554), 诸如例如pMYC1803。质粒pMYC1803衍生自基于RSF1010的质粒pTJS260(参见Wilcox的美国专利No. 5,169,760), 其携带来自RSF1010质粒的可调型四环素抗性标

志及复制和转移基因座(replication and mobilization loci)。其它例示性的有用的载体包括包括那些记载于Puhler等的美国专利No. 4,680,264的。

在一个实施方案中,表达质粒作为表达载体使用。在另一个实施方案中,RSF1010或其衍生物作为表达载体使用。在又一个实施方案中, pMYC1050或其衍生物, 或pMYC4803或其衍生物作为表达载体使用。

可通过包含于质粒中的选择标志基因来将质粒维持在宿主细胞中。这可以是抗生素抗性基因(其中将相应的抗生素添加至发酵培养基), 或本领域中已知的任何其它类型的选择标志基因, 例如原养型恢复基因, 其中在相应的性状(例如生物催化性状, 诸如氨基酸生物合成或核苷酸生物合成性状, 或碳源利用性状)营养缺陷的宿主细胞中使用所述质粒。

依照本发明使用的启动子可以是组成型启动子或可调型启动子。有用的可调型启动子的常见例子包括那些自lac启动子(即lacZ启动子)衍生的家族的, 特别是记载于DeBoer的美国专利No. 4,551,433的*tac*和*trc*启动子, 以及Ptac16、Ptac17、PtacII、PlacUV5、和T7lac启动子。在一个实施方案中, 启动子不是自宿主细胞生物体衍生的。在某些实施方案中, 启动子是自大肠杆菌生物体衍生的。

在依照本发明的表达系统中有用的非lac型启动子的常见例子包括例如那些列于表3的。

表3: 非 <i>lac</i> 启动子的例子	
启动子	诱导物
P _R	高温
P _L	高温
P _m	烃基或卤代苯甲酸盐
P _u	烃基或卤代甲苯
P _{sal}	水杨酸盐

参见例如: J. Sanchez-Romero和V. De Lorenzo (1999) Genetic Engineering of Nonpathogenic *Pseudomonas* strains as Biocatalysts for Industrial and Environmental Processes, 于 Manual of Industrial Microbiology and Biotechnology (A. Demain和J. Davies编) 页460-74 (ASM Press, Washington, D.C.); H. Schweizer (2001) Vectors to express foreign genes and techniques to monitor gene expression for *Pseudomonads*, Current Opinion in Biotechnology,

12:439-445; 及R. Slater和R. Williams (2000) *The Expression of Foreign DNA in Bacteria*, 于Molecular Biology and Biotechnology (J. Walker和R. Rapley编) 页125-54 (The Royal Society of Chemistry, Cambridge, UK)。还可使用具有对于选定的细菌宿主细胞而言天然的启动子的核苷酸序列的启动子来控制编码靶多肽的转基因的表达, 例如假单胞菌邻氨基苯甲酸盐或苯甲酸盐操纵子启动子(Pant, Pben)。还可使用串联启动子, 其中超过一种启动子共价衔接至另一种, 无论序列是相同的还是不同的(例如Pant-Pben串联启动子(启动子间的杂合物)或Plac-Plac串联启动子), 或者无论是自相同的还是不同的生物体衍生的。

可调型启动子利用启动子调控蛋白以控制启动子是其一部分的基因的转录。若在本文中使用可调型启动子, 则相应的启动子调控蛋白也会是依照本发明的表达系统的一部分。启动子调控蛋白的例子包括: 激活蛋白, 例如大肠杆菌分解代谢物激活蛋白、MalT蛋白; AraC家族转录激活物; 阻抑蛋白, 例如大肠杆菌LacI蛋白; 和双功能调控蛋白, 例如大肠杆菌NagC蛋白。许多可调型启动子/启动子调控蛋白对在本领域中是已知的。

启动子调控蛋白与效应器化合物相互作用, 所述效应器化合物即如下化合物, 其可逆地或不可逆地与调控蛋白相结合以使得所述蛋白质能够释放或结合启动子控制下的基因中的至少一个DNA转录调控区, 由此容许或阻断转录酶在启动基因转录中的作用。效应器化合物分类为诱导物或协阻抑物, 并且这些化合物包括天然效应器化合物和安慰诱导物化合物。许多可调型启动子/启动子调控蛋白/效应器化合物三件一套在本领域中是已知的。虽然效应器化合物可贯穿细胞培养或发酵使用, 但是在使用可调型启动子的一个优选的实施方案中, 在宿主细胞生物量生长到想要的数量或密度后, 将合适的效应器化合物添加至培养物以直接地或间接地导致编码所述感兴趣蛋白质或多肽的想要基因的表达。

举例而言, 在利用*lac*家族启动子的情况中, *lacI*基因也可存在于系统中。*lacI*基因(其(通常)是组成型表达的基因)编码Lac阻抑蛋白(LacD蛋白), 其结合这些启动子的*lac*操纵基因。如此, 在利用*lac*家族启动子的情况中, *lacI*基因也可包括在表达系统中, 并在表达系统中表达。在*lac*启动子家族成员(例如*tac*启动子)的情况中, 效应器化合物是诱导物, 优选安慰诱导物, 诸如IPTG(异丙基- β -D-硫代吡喃半乳糖苷, 也称作“异丙基硫代半乳糖苷”)。

对于感兴趣蛋白质或多肽的表达,也可使用任何植物启动子。启动子可以是植物RNA聚合酶II启动子。植物启动子中包括的元件可以是通常位于转录起始位点上游(5')约25-35个碱基对的TATA框或Goldberg-Hogness框,和位于上游70和100个碱基对之间的CCAAT框。在植物中,CCAAT框可具有不同于哺乳动物启动子的功能类似序列的共有序列(Messing等 (1983) 于: Genetic Engineering of Plants, Kosuge等编, 页211-227)。另外,几乎所有启动子都包括自转录起始位点上游约-100bp至-1,000bp或更多延伸的别的上游激活序列或增强子(Benoist和Chambon (1981) Nature 290:304-310; Gruss等 (1981) Proc. Nat. Acad. Sci. 78:943-947; 及Khoury和Gruss (1983) Cell 27:313-314)。

G. 表达系统

本发明进一步提供了对于将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的改良表达系统。在一个实施方案中,该系统包括宿主细胞和上文所述载体,其包含编码可操作连接至选自下组的分泌信号的感兴趣蛋白质或多肽的核苷酸序列: pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、Ttg2C、和ORF8124分泌信号序列,或与本文中所公开的分泌信号序列SEQ ID NO:1、3、5、7、9、11、13、15、17、19、20、21、或23基本上同源的序列,或编码SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的核苷酸序列。在一些实施方案中,在信号序列和感兴趣蛋白质或多肽之间没有进行修饰。然而,在某些实施方案中,掺入别的切割信号以促进多肽氨基端的正确加工。

分泌系统还可包括发酵培养基,诸如下文所述的。在一个实施方案中,该系统包括矿物盐培养基。在另一个实施方案中,该系统包括培养基中的化学诱导物。

CHAMPION™ pET表达系统提供了高水平的蛋白质生产。自T7lac强启动子诱导表达。此系统利用噬菌体T7 RNA聚合酶的高活性和特异性来高水平地转录感兴趣基因。位于启动子区中的lac操纵基因提供了比传统的基于T7的载体更紧密的调控,这改善了质粒稳定性和细胞生存力(Studier和Moffatt (1986) J Molecular Biology 189(1): 113-30; Rosenberg等 (1987) Gene 56(1): 125-35)。T7表达系统使用T7启动子和T7 RNA聚合酶(T7 RNAP)来高水平地转录感兴趣基因。在T7表达系统中实现高水平表达,因为T7 RNAP比天然的

大肠杆菌RNAP更具有进行性(processive), 而且致力于转录感兴趣基因。通过在宿主细胞中提供T7 RNAP的来源来诱导所鉴定基因的表达。这通过使用含有T7 RNAP基因的染色体拷贝的BL21大肠杆菌宿主来实现。T7 RNAP基因在可由IPTG诱导的lacUV5启动子的控制下。T7 RNAP在诱导后进行表达, 并转录感兴趣基因。

pBAD表达系统经由特定碳源(诸如葡萄糖、甘油和阿拉伯糖)的存在而容许感兴趣蛋白质或多肽的紧密控制的、可滴定的表达(Guzman等(1995) J Bacteriology 177(14): 4121-30)。独特地设计pBAD载体以对表达水平给予精确的控制。在araBAD启动子处启动来自pBAD载体的异源基因表达。启动子受到araC基因产物的正向和负向调控。AraC是与L-阿拉伯糖形成复合物的转录调节物。在没有L-阿拉伯糖的情况中, AraC二聚体阻断转录。为了最大限度的转录激活, 需要两种事件: (i.) L-阿拉伯糖结合至AraC, 这容许转录开始; (ii.) cAMP激活蛋白(CAP)-cAMP复合物结合至DNA, 并刺激AraC结合至启动子区的正确位置。

*trc*表达系统容许大肠杆菌中自*trc*启动子的高水平、可调型表达。*trc*表达载体已经得到了优化以在大肠杆菌中表达真核基因。*trc*启动子是自色氨酸(*trp*)和乳糖(*lac*)启动子衍生的杂合强启动子。它受到lacO操纵基因和lacIQ基因产物的调控(Brosius, J. (1984) Gene 27(2): 161-72)。

可使用本领域中已知的任何转化方法来实施本文中所公开的载体对宿主细胞的转化, 而且可以以完整细胞或以原生质体(即包括细胞质)来转化细菌宿主细胞。例示性的转化方法包括穿孔法例如电穿孔、原生质体融合、细菌接合、和二价阳离子处理例如氯化钙处理或CaCl/Mg²⁺处理、或本领域中其它公知方法。参见例如Morrison, J. Bact., 132:349-351 (1977); Clark-Curtiss和Curtiss, Methods in Enzymology, 101:347-362 (Wu等编, 1983); Sambrook等, Molecular Cloning, A Laboratory Manual (第2版, 1989); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); 及 Current Protocols in Molecular Biology (Ausubel等编, 1994)。

H. 宿主细胞

在一个实施方案中, 本发明提供了对于将可操作连接的感兴趣蛋白质或多肽靶向至革兰氏阴性细菌的周质或靶向入胞外空间中有用的表达系统。在一个实施方案中, 此系统利用分泌信号肽。在另一个实施方案中, 该表达系

统是用于表达包含本文中所公开的分泌信号的蛋白质的荧光假单胞菌表达系统。本发明的这个方面基于如下令人惊讶的发现，即荧光假单胞菌能够正确地加工和靶向来自荧光假单胞菌和非荧光假单胞菌系统两者的分泌信号。

在此实施方案中，宿主细胞可选自“革兰氏阴性变形菌门(Proteobacteria)亚组18”。“革兰氏阴性变形菌门亚组18”定义为物种荧光假单胞菌的所有亚种、变种、菌株和其它亚物种单位的组，包括那些属于例如下列各项的（圆括号中显示了例示性菌株的ATCC或其它保藏号）：荧光假单胞菌生物型A，也称为生物变型1或生物变型I (ATCC 13525)；荧光假单胞菌生物型B，也称为生物变型2或生物变型II (ATCC 17816)；荧光假单胞菌生物型C，也称为生物变型3或生物变型III (ATCC 17400)；荧光假单胞菌生物型F，也称为生物变型4或生物变型IV (ATCC 12983)；荧光假单胞菌生物型G，也称为生物变型5或生物变型V (ATCC 17518)；荧光假单胞菌生物变型VI；荧光假单胞菌Pf0-1；荧光假单胞菌Pf-5 (ATCC BAA-477)；荧光假单胞菌SBW25；及荧光假单胞菌纤维素亚种(NCIMB 10462)。

宿主细胞可选自“革兰氏阴性变形菌门亚组19”。“革兰氏阴性变形菌门亚组19”定义为荧光假单胞菌生物型A的所有菌株的组。此生物型的特别优选的菌株是荧光假单胞菌菌株MB101（参见Wilcox的美国专利5,169,760）及其衍生物。它优选的衍生物的例子是荧光假单胞菌菌株MB214，它是通过将天然大肠杆菌PlacI-lacI-lacZYA构建体（即其中删除了PlacZ）插入MB101染色体*asd*（天冬氨酸脱氢酶基因）基因座中而构建的。

可在本发明中使用的别的荧光假单胞菌菌株包括荧光假单胞菌Migula和荧光假单胞菌Loitokitok，其具有下列ATCC名称：[NCIB 8286]；NRRL B-1244；NCIB 8865菌株CO1；NCIB 8866菌株CO₂；1291 [ATCC 17458；IFO 15837；NCIB 8917；LA；NRRL B-1864；吡咯烷；PW2 [ICMP 3966；NCPPB 967；NRRL B-899]；13475；NCTC 10038；NRRL B-1603 [6；IFO 15840]；52-1C；CCEB 488-A [BU 140]；CCEB 553 [EM 15/47]；IAM 1008 [AHH-27]；IAM 1055 [AHH-23]；1 [IFO 15842]；12 [ATCC 25323；NIH 11；den Dooren de Jong 216]；18 [IFO 15833；WRRL P-7]；93 [TR-10]；108 [52-22；IFO 15832]；143 [IFO 15836；PL]；149 [2-40-40；IFO 15838]；182 [IFO 3081；PJ 73]；184 [IFO 15830]；185 [W2 L-1]；186 [IFO 15829；PJ 79]；187 [NCPPB 263]；188 [NCPPB 316]；189 [PJ227；1208]；191 [IFO 15834；PJ 236；22/1]；194

[Klinge R-60; PJ 253]; 196 [PJ 288]; 197 [PJ 290]; 198 [PJ 302]; 201 [PJ 368]; 202 [PJ 372]; 203 [PJ 376]; 204 [IFO 15835; PJ 682]; 205 [PJ 686]; 206 [PJ 692]; 207 [PJ 693]; 208 [PJ 722]; 212. [PJ 832]; 215 [PJ 849]; 216 [PJ 885]; 267 [B-9]; 271 [B-1612]; 401 [C71A; IFO 15831; PJ 187]; NRRL B-3178 [4; IFO. 15841]; KY 8521; 3081; 30-21; [IFO 3081]; N; PYR; PW; D946-B83 [BU 2183; FERM-P 3328]; P-2563 [FERM-P 2894; IFO 13658]; IAM-1126 [43F]; M-1; A506 [A5-06]; A505 [A5-05-1]; A526 [A5-26]; B69; 72; NRRL B-4290; PMW6 [NCIB 11615]; SC 12936; A1 [IFO 15839]; F 1847 [CDC-EB]; F 1848 [CDC 93]; NCIB 10586; P17; F-12; AmMS 257; PRA25; 6133D02; 6519E01; Ni; SC15208; BNL-WVC; NCTC 2583 [NCIB 8194]; H13; 1013 [ATCC 11251; CCEB 295]; IFO 3903; 1062; 或Pf-5。

在一个实施方案中, 宿主细胞可以是能够生成感兴趣蛋白质或多肽的任何细胞, 包括上文所描述的荧光假单胞菌细胞。由于它们在大型分批培养中相对便宜的生长要求和生产蛋白质的潜在能力, 用于生成感兴趣蛋白质或多肽的最常用系统包括某些细菌细胞 (特别是大肠杆菌)。酵母也用于表达生物学有关蛋白质和多肽, 特别是出于研究目的。系统包括酿酒酵母 (*Saccharomyces cerevisiae*) 或巴斯德毕赤氏酵母 (*Pichia pastoris*)。这些系统是充分表征的, 一般提供可接受水平的总蛋白质表达, 而且是比较快速且便宜的。昆虫细胞表达系统也已经显现为用于表达生物学活性形式的重组蛋白的备选。在一些情况中, 可生成经翻译后修饰的正确折叠的蛋白质。哺乳动物细胞表达系统 (诸如中国仓鼠卵巢细胞) 也已经用于表达感兴趣蛋白质或多肽。这些表达系统在小规模上通常是有效的。某些生物制品可自蛋白质衍生, 特别是在动物或人保健应用中。在另一个实施方案中, 宿主细胞是植物细胞, 包括但不限于烟草细胞、玉米、来自拟南芥 (*Arabidopsis*) 物种的细胞、马铃薯或稻细胞。在另一个实施方案中, 在过程中分析或修饰多细胞生物体, 包括但不限于转基因生物体。用于分析和/或修饰多细胞生物体的技术一般基于记载用于修饰下文所述细胞的技术。

在另一个实施方案中, 宿主细胞可以是原核生物, 诸如细菌细胞, 包括但不限于埃希氏菌属 (*Escherichia*) 或假单胞菌属 (*Pseudomonas*) 物种。典型的细菌细胞记载于例如“Biological Diversity: Bacteria and Archaeans”, 即由 M J Farabee 博士 (Estrella Mountain Community College, Arizona, USA) 在网站

www.emc.maricotpa.edu/faculty/farabee/BIOBK/BioBookDiversity提供的在线生物学书籍的一章。在某些实施方案中，宿主细胞可以是假单胞菌细胞，而且通常可以是荧光假单胞菌细胞。在其它实施方案中，宿主细胞也可以是大肠杆菌细胞。在另一个实施方案中，宿主细胞可以是真核细胞，例如昆虫细胞，包括但不限于来自灰翅夜蛾属(*Spodoptera*)、粉纹夜蛾属(*Trichoplusia*)、果蝇属(*Drosophila*)或顶灯蛾属(*Estigmene*)物种的细胞，或哺乳动物细胞，包括但不限于鼠细胞、仓鼠细胞、猴、灵长类或人细胞。

在一个实施方案中，宿主细胞可以是任何细菌分类单位的成员。例如，细胞可以是真细菌的任何物种的成员。宿主可以是下组分类单位之任一种的成员：酸杆菌门(*Acidobacteria*)、放线杆菌门(*Actinobacteria*)、产液菌门(*Aquificae*)、拟杆菌门(*Bacteroidetes*)、绿菌门(*Chlorobi*)、衣原体门(*Chlamydiae*)、绿屈挠菌门(*Chloroflexi*)、金矿菌门(*Chrysiogenetes*)、蓝细菌门(*Cyanobacteria*)、脱铁杆菌门(*Deferribacteres*)、异常球菌(*Deinococcus*)、网团菌门(*Dictyoglomi*)、丝状杆菌门(*Fibrobacteres*)、厚壁菌门(*Firmicutes*)、梭杆菌门(*Fusobacteria*)、芽单胞菌门(*Gemmatimonadetes*)、黏胶球形菌门(*Lentisphaerae*)、硝化螺旋菌门(*Nitrospirae*)、浮霉状菌门(*Planctomycetes*)、变形菌门(*Proteobacteria*)、螺旋体门(*Spirochaetes*)、热脱硫杆菌门(*Thermodesulfobacteria*)、热微菌门(*Thermomicrobia*)、栖热袍菌门(*Thermotogae*)、栖热菌属(*Thermus*) (栖热菌目(*Thermales*))、或疣微菌门(*Verrucomicrobia*)。在真细菌宿主细胞的一个实施方案中，细胞可以是真细菌(除蓝细菌门之外)的任何物种的成员。

细菌宿主还可以是变形菌门的任何物种的成员。变形菌门宿主细胞可以是分类单位 α -变形菌纲、 β -变形菌纲、 γ -变形菌纲、 δ -变形菌纲或 ε -变形菌纲中任一种的成员。另外，宿主可以是分类单位 α -变形菌纲、 β -变形菌纲或 γ -变形菌纲中任一种的成员，和 γ -变形菌纲的任何物种的成员。

在 γ -变形菌纲宿主的一个实施方案中，宿主会是分类单位气单胞菌目(*Aeromonadales*)、交替单胞菌目(*Alteromonadales*)、肠杆菌目(*Enterobacteriales*)、假单胞菌目(*Pseudomonadales*)、或黄单胞菌目(*Xanthomonadales*)中任一种的成员；或肠杆菌目或假单胞菌目的任何物种的成员。在一个实施方案中，宿主细胞可以是肠杆菌目的，宿主细胞会是肠杆菌科的成员，或可以是欧文氏菌属(*Erwinia*)、埃希氏菌属(*Escherichia*)、或沙

雷氏菌属(*Serratia*)中任一种的成员; 或埃希氏菌属的成员。若宿主细胞是假单胞菌目的, 则宿主细胞可以是假单胞菌科的成员, 包括假单胞菌属。 γ -变形菌纲宿主包括物种大肠杆菌的成员和物种荧光假单胞菌的成员。

其它假单胞菌生物体也可能是有用的。假单胞菌和紧密相关的物种包括革兰氏阴性变形菌门亚组1, 其包括属于由R. E. Buchanan和N.E. Gibbons (编), *Bergey's Manual of Determinative Bacteriology*, 页217-289 (第8版, 1974) (The Williams & Wilkins Co., Baltimore, Md., USA) (下文中的“Bergey (1974)”)以“革兰氏阴性好氧杆菌和球菌”描述的科和/或属的变形菌门组。表4呈现了这些科和属的生物体。

表4: “革兰氏阴性好氧杆菌和球菌” 部分(Bergey (1974))中所列的科和属	
科I. 假单胞菌科(<i>Pseudomonaceae</i>)	葡糖杆菌属(<i>Gluconobacter</i>) 假单胞菌属(<i>Pseudomonas</i>) 黄单胞菌属(<i>Xanthomonas</i>) 动胶菌属(<i>Zoogloea</i>)
科II. 固氮菌科(<i>Azotobacteraceae</i>)	氮单胞菌属(<i>Azomonas</i>) 固氮菌属(<i>Azotobacter</i>) 拜叶林克氏菌属(<i>Beijerinckia</i>) 德克斯氏菌属(<i>Derxia</i>)
科III. 根瘤菌科(<i>Rhizobiaceae</i>)	土壤杆菌属(<i>Agrobacterium</i>) 根瘤菌属(<i>Rhizobium</i>)
科IV. 甲基单胞菌科(<i>Methylobacteriaceae</i>)	甲基球菌属(<i>Methylococcus</i>) 甲基单胞菌属(<i>Methylobacter</i>)
科V. 盐杆菌科(<i>Halobacteriaceae</i>)	盐杆菌属(<i>Halobacterium</i>) 盐球菌属(<i>Halococcus</i>)
其它属	醋杆菌属(<i>Acetobacter</i>) 产碱菌属(<i>Alcaligenes</i>) 博德特氏菌属(<i>Bordetella</i>) 布鲁氏菌属(<i>Brucella</i>) 弗朗西丝氏菌属(<i>Francisella</i>) 栖热菌属(<i>Thermus</i>)

“革兰氏阴性变形菌门亚组1” 还包括依照分类中所使用的标准会分类

在此标题中的变形菌门。该标题还包括先前分类在此部分中但不再在此部分中的组，诸如食酸菌属(*Acidovorax*)、短波单胞菌属(*Brevundimonas*)、伯克霍尔德氏菌属(*Burkholderia*)、氢噬胞菌属(*Hydrogenophaga*)、海洋单胞菌属(*Oceanimonas*)、罗尔斯通氏菌属(*Ralstonia*)、和寡养单胞菌属(*Stenotrophomonas*)，鞘氨醇单胞菌属(*Sphingomonas*) (及其自其衍生的属芽单胞菌属(*Blastomonas*)) (其通过对属于黄单胞菌属 (且先前称为黄单胞菌属的种) 的生物体重新分组而创建)、酸单胞菌属(*Acidomonas*) (其通过对属于Bergey (1974)中定义的酸杆菌属(*Acetobacter*)的生物体重新分组而创建)。另外，宿主可包括来自下组的细胞：假单胞菌属、海洋假单胞菌(*Pseudomonas enalia*) (ATCC 14393)、生黑假单胞菌(*Pseudomonas nigrifaciensi*) (ATCC 19375)、和腐败假单胞菌(*Pseudomonas putrefaciens*) (ATCC 8071)，它们已经分别被重新分类为河豚毒素交替单胞菌(*Alteromonas haloplanktis*)、产黑交替单胞菌(*Alteromonas nigrifaciens*)、和腐败交替单胞菌(*Alteromonas putrefaciens*)。类似地，例如后来食酸假单胞菌(*Pseudomonas acidovorans*) (ATCC 15668)和睾丸酮假单胞菌(*Pseudomonas testosterone*) (ATCC 11996)已经被分别重新分类为食酸丛毛单胞菌(*Comamonas acidovorans*)和睾丸酮丛毛单胞菌(*Comamonas testosterone*)；而生黑假单胞菌(ATCC 19375)和杀鱼假单胞菌(*Pseudomonas piscicida*) (ATCC 15057)已经被分别重新分类为生黑假交替单胞菌(*Pseudoalteromonas nigrifaciens*)和杀鱼假交替单胞菌(*Pseudoalteromonas piscicida*)。“革兰氏阴性变形菌门亚组1”还包括变形菌门中分类为属于任何如下科的：假单胞菌科、固氮菌科 (现在通常以同义词即假单胞菌科的“固氮菌组”称呼)、根瘤菌科、和甲基单胞菌科 (现在通常以同义词即“甲基球菌科”称呼)。因此，在那些于本文中其它部分所述的属之外，变形菌门中落入“革兰氏阴性变形菌门亚组1”内的其它属包括：1)嗜氮根瘤菌属(*Azorhizophilus*)的固氮菌组细菌；2)纤维弧菌属(*Cellvibrio*)、寡源杆菌属(*Oligella*)、和Teredinibacter的假单胞菌科细菌；3)螯合杆菌属(*Chelatobacter*)、剑菌属(*Ensifer*)、Liberibacter (也称为“*Candidatus Liberibacter*”)、和中华根瘤菌属(*Sinorhizobium*)的根瘤菌科细菌；及4)甲基细菌属(*Methylobacter*)、甲基暖菌属(*Methylocaldum*)、甲基微菌属(*Methylomicrobium*)、甲基八叠球菌属(*Methylosarcina*)、和甲基球形菌属(*Methylosphaera*)的甲基球菌科细菌。

在另一个实施方案中，宿主细胞选自“革兰氏阴性变形菌门亚组2”。“革兰氏阴性变形菌门亚组2”定义为下列属（在圆括号中指明产品目录列出的、公众可获得的、已保藏的其菌株的总数，均保藏于ATCC，除非另有指明）的变形菌门组：酸单胞菌属(2)；酸杆菌属(93)；葡糖杆菌属(37)；短波单胞菌属(23)；拜叶林克氏菌属(*Beyerinckia*) (13)；德克斯氏菌属(2)；布鲁氏菌属(4)；土壤杆菌属(79)；螯合杆菌属(2)；剑菌属(3)；根瘤菌属(144)；中华根瘤菌属(24)；芽单胞菌属(1)；鞘氨醇单胞菌属(27)；产碱菌属(88)；博德特氏菌属(43)；伯克霍尔德氏菌属(73)；罗尔斯通氏菌属(33)；食酸菌属(20)；氢噬胞菌属(9)；动胶菌属(9)；甲基细菌属(2)；甲基暖菌属(在NCIMB的1种)；甲基球菌属(2)；甲基微菌属(2)；甲基单胞菌属(9)；甲基八叠球菌属(1)；甲基球形菌属；氮单胞菌属(9)；嗜氮根瘤菌属(5)；固氮菌属(64)；纤维弧菌属(3)；寡源杆菌属(5)；假单胞菌属(1139)；弗朗西丝氏菌属(4)；黄单胞菌属(229)；寡养单胞菌属(50)；和海洋单胞菌属(4)。

“革兰氏阴性变形菌门亚组2”的例示性宿主细胞物种包括但不限于下列细菌（在圆括号中显示其例示性菌株的ATCC或其它保藏号）：甲醇酸单胞菌(*Acidomonas methanolica*) (ATCC 43581)；醋化醋杆菌(*Acetobacter aceti*) (ATCC 15973)；氧化葡糖杆菌(*Gluconobacter oxydans*) (ATCC 19357)；缺陷短波单胞菌(*Brevundimonas diminuta*) (ATCC 11568)；印度拜叶林克氏菌(*Beijerinckia indica*) (ATCC 9039和ATCC 19361)；胶德克斯氏菌(*Derxia gummosa*) (ATCC 15994)；马耳他布鲁氏菌(*Brucella melitensis*) (ATCC 23456)；流产布鲁氏菌(*Brucella abortus*) (ATCC 23448)；根癌土壤杆菌(*Agrobacterium tumefaciens*) (ATCC 23308)、放射形土壤杆菌(*Agrobacterium radiobacter*) (ATCC 19358)、发根土壤杆菌(*Agrobacterium rhizogenes*) (ATCC 11325)；何氏螯合杆菌(*Chelatobacter heintzii*) (ATCC 29600)；粘着剑菌(*Ensifer adhaerens*) (ATCC 33212)；豌豆根瘤菌(*Rhizobium leguminosarum*) (ATCC 10004)；费氏中华根瘤菌(*Sinorhizobium fredii*) (ATCC 35423)；*Blastomonas natatoria* (ATCC 35951)；少动鞘氨醇单胞菌(*Sphingomonas paucimobilis*) (ATCC 29837)；粪产碱菌(*Alcaligenes faecalis*) (ATCC 8750)；百日咳博德特氏菌(*Bordetella pertussis*) (ATCC 9797)；洋葱伯克霍尔德氏菌(*Burkholderia cepacia*) (ATCC 25416)；皮氏罗尔斯通氏菌属(*Ralstonia pickettii*) (ATCC 27511)；敏捷食酸菌(*Acidovorax facilis*) (ATCC 11228)；黄色氢噬胞菌

(*Hydrogenophaga flava*) (ATCC 33667); 生枝动胶菌(*Zoogloea ramigera*) (ATCC 19544); 藤黄甲基细菌(*Methylobacter luteus*) (ATCC 49878); *Methylocaldum gracile* (NCIMB 11912); 荚膜甲基球菌(*Methylococcus capsulatus*) (ATCC 19069); 活泼甲基微菌(*Methylomicrobium agile*) (ATCC 35068); 甲烷甲基单胞菌(*Methylomonas methanica*) (ATCC 35067); *Methylosarcina fibrata* (ATCC 700909); *Methylosphaera hansonii* (ACAM 549); 敏捷氮单胞菌(*Azomonas agilis*) (ATCC 7494); 雀稗嗜氮根瘤菌(*Azorhizophilus paspali*) (ATCC 23833); 褐球固氮菌(*Azotobacter chroococcum*) (ATCC 9043); 混合纤维弧菌(*Cellvibrio mixtus*) (UQM 2601); 尿道寡源杆菌(*Oligella urethralis*) (ATCC 17960); 铜绿假单胞菌(*Pseudomonas aeruginosa*) (ATCC 10145); 荧光假单胞菌(*Pseudomonas fluorescens*) (ATCC 35858); 土拉热弗朗西丝氏菌(*Francisella tularensis*) (ATCC 6223); 嗜麦芽糖寡养单胞菌(*Stenotrophomonas maltophilia*) (ATCC 13637); 野油菜黄单胞菌(*Xanthomonas campestris*) (ATCC 33913); 和 *Oceanimonas doudoroffii* (ATCC 27123)。

在另一个实施方案中, 宿主细胞选自“革兰氏阴性变形菌门亚组3”。“革兰氏阴性变形菌门亚组3”定义为下列属的变形菌门组: 短波单胞菌属; 土壤杆菌属; 根瘤菌; 中华根瘤菌属; 芽单胞菌; 鞘氨醇单胞菌属; 产碱菌属; 伯克霍尔德氏菌属; 罗尔斯通氏菌属; 食酸菌属; 氢噬胞菌属; 甲基细菌属; 甲基暖菌属; 甲基球菌属; 甲基微菌属; 甲基单胞菌属; 甲基八叠球菌属; 甲基球形菌属; 氮单胞菌属; 嗜氮根瘤菌属; 固氮菌属; 纤维弧菌属; 寡源杆菌属; 假单胞菌属; *Teredinibacter*; 弗朗西丝氏菌属; 寡养单胞菌属; 黄单胞菌属; 和海洋单胞菌属。

在另一个实施方案中, 宿主细胞选自“革兰氏阴性变形菌门亚组4”。“革兰氏阴性变形菌门亚组4”定义为下列属的变形菌门组: 短波单胞菌属; 芽单胞菌; 鞘氨醇单胞菌属; 伯克霍尔德氏菌属; 罗尔斯通氏菌属; 食酸菌属; 氢噬胞菌属; 甲基细菌属; 甲基暖菌属; 甲基球菌属; 甲基微菌属; 甲基单胞菌属; 甲基八叠球菌属; 甲基球形菌属; 氮单胞菌属; 嗜氮根瘤菌属; 固氮菌属; 纤维弧菌属; 寡源杆菌属; 假单胞菌属; *Teredinibacter*; 弗朗西丝氏菌属; 寡养单胞菌属; 黄单胞菌属; 和海洋单胞菌属。

在另一个实施方案中, 宿主细胞选自“革兰氏阴性变形菌门亚组5”。“革兰氏阴性变形菌门亚组5”定义为下列属的变形菌门组: 甲基细菌属; 甲基

暖菌属；甲基球菌属；甲基微菌属；甲基单胞菌属；甲基八叠球菌属；甲基球形菌属；氮单胞菌属；嗜氮根瘤菌属；固氮菌属；纤维弧菌属；寡源杆菌属；假单胞菌属；*Teredinibacter*；弗朗西丝氏菌属；寡养单胞菌属；黄单胞菌属；和海洋单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组6”。“革兰氏阴性变形菌门亚组6”定义为下列属的变形菌门组：短波单胞菌属；芽单胞菌；鞘氨醇单胞菌属；伯克霍尔德氏菌属；罗尔斯通氏菌属；食酸菌属；氢噬胞菌属；氮单胞菌属；嗜氮根瘤菌属；固氮菌属；纤维弧菌属；寡源杆菌属；假单胞菌属；*Teredinibacter*；寡养单胞菌属；黄单胞菌属；和海洋单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组7”。“革兰氏阴性变形菌门亚组7”定义为下列属的变形菌门组：氮单胞菌属；嗜氮根瘤菌属；固氮菌属；纤维弧菌属；寡源杆菌属；假单胞菌属；*Teredinibacter*；寡养单胞菌属；黄单胞菌属；和海洋单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组8”。“革兰氏阴性变形菌门亚组8”定义为下列属的变形菌门组：短波单胞菌属；芽单胞菌；鞘氨醇单胞菌属；伯克霍尔德氏菌属；罗尔斯通氏菌属；食酸菌属；氢噬胞菌属；假单胞菌属；寡养单胞菌属；黄单胞菌属；和海洋单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组9”。“革兰氏阴性变形菌门亚组9”定义为下列属的变形菌门组：短波单胞菌属；伯克霍尔德氏菌属；罗尔斯通氏菌属；食酸菌属；氢噬胞菌属；假单胞菌属；寡养单胞菌属；和海洋单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组10”。“革兰氏阴性变形菌门亚组10”定义为下列属的变形菌门组：伯克霍尔德氏菌属；罗尔斯通氏菌属；假单胞菌属；寡养单胞菌属；和黄单胞菌属。

宿主细胞可选自“革兰氏阴性变形菌门亚组11”。“革兰氏阴性变形菌门亚组11”定义为下列属的变形菌门组：假单胞菌属；寡养单胞菌属；和黄单胞菌属。宿主细胞可选自“革兰氏阴性变形菌门亚组12”。“革兰氏阴性变形菌门亚组12”定义为下列属的变形菌门组：伯克霍尔德氏菌属；罗尔斯通氏菌属；假单胞菌属。宿主细胞可选自“革兰氏阴性变形菌门亚组13”。“革兰氏阴性变形菌门亚组13”定义为下列属的变形菌门组：伯克霍尔德氏菌属；罗尔斯通氏菌属；假单胞菌属；和黄单胞菌属。宿主细胞可选自“革兰氏阴

性变形菌门亚组14”。“革兰氏阴性变形菌门亚组14”定义为下列属的变形菌门组：假单胞菌属和黄单胞菌属。宿主细胞可选自“革兰氏阴性变形菌门亚组15”。“革兰氏阴性变形菌门亚组15”定义为假单胞菌属的变形菌门组。

宿主细胞可选自“革兰氏阴性变形菌门亚组16”。“革兰氏阴性变形菌门亚组16”定义为下列假单胞菌物种（在圆括号中显示了例示性菌株的ATCC或其他保藏号）的变形菌门组：*Pseudomonas abietaniphila* (ATCC 700689)；铜绿假单胞菌(*Pseudomonas aeruginosa*) (ATCC 10145)；产碱假单胞菌(*Pseudomonas alcaligenes*) (ATCC 14909)；病鳝假单胞菌(*Pseudomonas anguilliseptica*) (ATCC 33660)；香茅醇假单胞菌(*Pseudomonas citronellolis*) (ATCC 13674)；变黄假单胞菌(*Pseudomonas flavescens*) (ATCC 51555)；门多萨假单胞菌(*Pseudomonas mendocina*) (ATCC 25411)；硝基还原假单胞菌(*Pseudomonas nitroreducens*) (ATCC 33634)；食油假单胞菌(*Pseudomonas oleovorans*) (ATCC 8062)；类产碱假单胞菌(*Pseudomonas pseudoalcaligenes*) (ATCC 17440)；食树脂假单胞菌(*Pseudomonas resinovorans*) (ATCC 14235)；稻草假单胞菌(*Pseudomonas straminea*) (ATCC 33636)；伞菌假单胞菌(*Pseudomonas agarici*) (ATCC 25941)；*Pseudomonas alcaliphila*；*Pseudomonas alginovora*；*Pseudomonas andersonii*；铁角蕨假单胞菌(*Pseudomonas asplenii*) (ATCC 23835)；壬二酸假单胞菌(*Pseudomonas azelaica*) (ATCC 27162)；拜氏假单胞菌(*Pseudomonas beyerinckii*) (ATCC 19372)；*Pseudomonas borealis*；北城假单胞菌(*Pseudomonas boreopolis*) (ATCC 33662)；*Pseudomonas brassicacearum*；食丁酸假单胞菌(*Pseudomonas butanovora*) (ATCC 43655)；*Pseudomonas cellulosa* (ATCC 55703)；桔黄假单胞菌(*Pseudomonas aurantiaca*) (ATCC 33663)；绿针假单胞菌(*Pseudomonas chlororaphis*) (ATCC 9446, ATCC 13985, ATCC 17418, ATCC 17461)；莓实假单胞菌(*Pseudomonas fragi*) (ATCC 4973)；隆德假单胞菌(*Pseudomonas lundensis*) (ATCC 49968)；腐臭假单胞菌(*Pseudomonas taetrolens*) (ATCC 4683)；青紫葛假单胞菌(*Pseudomonas cissicola*) (ATCC 33616)；晕斑假单胞菌(*Pseudomonas coronafaciens*)；*Pseudomonas diterpeniphila*；伸长假单胞菌(*Pseudomonas elongata*) (ATCC 10144)；弯曲假单胞菌(*Pseudomonas flectens*) (ATCC 12775)；产氮假单胞菌(*Pseudomonas azotoformans*)；*Pseudomonas brenneri*；*Pseudomonas cedrella*；皱纹假单胞菌(*Pseudomonas corrugate*) (ATCC 29736)；*Pseudomonas*

extremorientalis; 荧光假单胞菌(*Pseudomonas fluorescens*) (ATCC 35858); *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii* (ATCC 700871); 边缘假单胞菌(*Pseudomonas marginalis*) (ATCC 10844); *Pseudomonas migulae*; 霉味假单胞菌(*Pseudomonas mucidolens*) (ATCC 4685); 东方假单胞菌(*Pseudomonas orientalis*); *Pseudomonas rhodesiae*; 类黄假单胞菌(*Pseudomonas synxantha*) (ATCC 9890); 托拉氏假单胞菌(*Pseudomonas tolaasii*) (ATCC 33618); *Pseudomonas veronii* (ATCC 700474); *Pseudomonas frederiksbergensis*; 弯曲假单胞菌(*Pseudomonas geniculata*) (ATCC 19374); *Pseudomonas gingeri*; *Pseudomonas graminis*; 格氏假单胞菌(*Pseudomonas grimontii*); 盐脱氮假单胞菌(*Pseudomonas halodenitrificans*); 嗜盐假单胞菌(*Pseudomonas halophila*); 栖木槿假单胞菌(*Pseudomonas hibiscicola*) (ATCC 19867); 哈特假单胞菌(*Pseudomonas huttiensis*) (ATCC 14670); 噬氢假单胞菌(*Pseudomonas hydrogenovora*); *Pseudomonas jessenii* (ATCC 700870); *Pseudomonas kilonensis*; *Pseudomonas lanceolata* (ATCC 14669); *Pseudomonas lini*; 划界假单胞菌(*Pseudomonas marginata*) (ATCC 25417); 臭味假单胞菌(*Pseudomonas mephitica*) (ATCC 33665); 脱氮假单胞菌(*Pseudomonas denitrificans*) (ATCC 19244); 百日咳假单胞菌(*Pseudomonas pertucinogena*) (ATCC 190); 皮克特假单胞菌(*Pseudomonas pictorum*) (ATCC 23328); *Pseudomonas psychrophila*; 黄褐假单胞菌(*Pseudomonas filva*) (ATCC 31418); 蒙氏假单胞菌(*Pseudomonas monteilii*) (ATCC 700476); *Pseudomonas mosselii*; 栖稻假单胞菌(*Pseudomonas oryzihabitans*) (ATCC 43272); 变形假单胞菌(*Pseudomonas plecoglossicida*) (ATCC 700383); 恶臭假单胞菌(*Pseudomonas putida*) (ATCC 12633); *Pseudomonas reactans*; 多刺假单胞菌(*Pseudomonas spinosa*) (ATCC 14606); *Pseudomonas balearica*; 浅黄假单胞菌(*Pseudomonas luteola*) (ATCC 43273); 施氏假单胞菌(*Pseudomonas stutzeri*) (ATCC 17588); 扁桃假单胞菌(*Pseudomonas amygdali*) (ATCC 33614); *Pseudomonas avellanae* (ATCC 700331); 番木瓜假单胞菌(*Pseudomonas caricapapayae*) (ATCC 33615); 菊苣假单胞菌(*Pseudomonas cichorii*) (ATCC 10857); 天仙果假单胞菌(*Pseudomonas ficuserectae*) (ATCC 35104); 褐鞘假单胞菌(*Pseudomonas fuscovaginae*); 苦楝假单胞菌(*Pseudomonas meliae*) (ATCC 33050); 丁香假单胞菌(*Pseudomonas syringae*) (ATCC 19310); 绿黄假

单胞菌 (*Pseudomonas viridiflava*) (ATCC 13223); 热食碳酸假单胞菌 (*Pseudomonas thermocarboxydovorans*) (ATCC 35961); 耐热假单胞菌 (*Pseudomonas thermotolerans*); *Pseudomonas thivervalensis*; 温哥华假单胞菌 (*Pseudomonas Vancouverensis*) (ATCC 700688); 威斯康辛假单胞菌 (*Pseudomonas wisconsinensis*); 和厦门假单胞菌 (*Pseudomonas xiamenensis*)。

宿主细胞可选自“革兰氏阴性变形菌门亚组17”。“革兰氏阴性变形菌门亚组17”定义为本领域中称为“荧光假单胞菌”的变形菌门组,所述荧光假单胞菌包括那些属于例如下列假单胞菌物种的:产氮假单胞菌; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata*; *Pseudomonas extremorientalis*; 荧光假单胞菌; *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii*; 边缘假单胞菌; *Pseudomonas migulae*; 霉味假单胞菌; 东方假单胞菌; *Pseudomonas rhodesiae*; 类黄假单胞菌; 托拉氏假单胞菌; 和 *Pseudomonas veronii*。

其它合适的宿主包括那些在参考文献的其它部分中进行分类的,诸如革兰氏(+)变形菌门。在一个实施方案中,宿主细胞是大肠杆菌。已经为大肠杆菌MG1655建立了大肠杆菌的基因组序列(Blattner等 (1997) The complete genome sequence of *Escherichia coli* K-12, *Science* 277(5331): 1453-74), 而且对于大肠杆菌K12, DNA微阵列是商品化的(MWG Inc, High Point, N.C.)。大肠杆菌可在丰富培养基诸如Luria-Bertani (LB) (10g/L胰蛋白胨、5g/L NaCl、5g/L酵母提取物)或具有合适的碳源(诸如1%葡萄糖)的限定基本培养基诸如M9 (6g/L Na₂HPO₄、3g/L KH₂PO₄、1g/L NH₄Cl、0.5g/L NaCl、pH7.4)中培养。常规的是,将大肠杆菌细胞的过夜培养物稀释,并接种入摇瓶或发酵罐中的新鲜的丰富培养基或基本培养基中,并于37°C培养。

宿主也可以是哺乳动物起源的,诸如自包括任何人或非人哺乳动物的哺乳动物衍生的细胞。哺乳动物可以包括但不限于灵长类、猴、猪(porcine)、绵羊(ovine)、牛、啮齿类、有蹄类、猪(pig)、猪(swine)、绵羊(sheep)、羔羊、山羊、牛、鹿、骡、马、猴、猿、犬、猫、大鼠、和小鼠。

宿主细胞也可以是植物起源的。可选择任何植物来鉴定基因和调控序列。供基因和调控序列的分离用的合适的植物靶物的例子会包括但不限于苜蓿(alfalfa)、苹果(apple)、杏(apricot)、拟南芥(*Arabidopsis*)、朝鲜蓟(artichoke)、芝麻菜(arugula)、芦笋(asparagus)、鳄梨(avocado)、香蕉(banana)、大麦(barley)、

豆类(bean)、甜菜(beet)、黑莓(blackberry)、蓝莓(blueberry)、花椰菜(broccoli)、孢子甘蓝(brussels sprouts)、甘蓝(cabbage)、芸苔(canola)、网纹瓜(cantaloupe)、胡萝卜(carrot)、木薯(cassaya)、蓖麻(castorbean)、花椰菜(cauliflower)、芹菜(celery)、樱桃(cherry)、菊苣(chicory)、芫荽(cilantro)、柑橘(citrus)、克莱门氏小柑橘(clementines)、车轴草(clover)、椰子(coconut)、咖啡(coffee)、玉米(corn)、棉(cotton)、酸果蔓(cranberry)、黄瓜(cucumber)、黄杉(Douglas fir)、茄子(eggplant)、苣荬菜(endive)、菊苣(escarole)、桉(eucalyptus)、茴香(fennel)、无花果(figs)、蒜(garlic)、葫芦(gourd)、葡萄(grape)、葡萄柚(grapefruit)、蜜瓜(honey dew)、豆薯(jicama)、猕猴桃(kiwifruit)、莴苣(lettuce)、葱(leeks)、柠檬(lemon)、来檬(lime)、火炬松(Loblolly pine)、亚麻(linseed)、芒果(mango)、甜瓜(melon)、蘑菇(mushroom)、油桃(nectarine)、坚果(nut)、燕麦(oat)、油棕(oil palm)、油菜(oil seed rape)、秋葵(okra)、橄榄(olive)、葱(onion)、柑橘(orange)、观赏植物、棕榈(palm)、番木瓜(papaya)、欧芹(parsley)、欧防风(parsnip)、豌豆(pea)、桃(peach)、花生(peanut)、梨(pear)、胡椒(pepper)、柿子(persimmon)、松(pine)、凤梨(pineapple)、车前(plantain)、梅(plum)、石榴(pomegranate)、杨(poplar)、马铃薯(potato)、南瓜(pumpkin)、榲桲(quince)、辐射松(radiata pine)、菊苣(radischio)、萝卜(radish)、油菜籽(rapeseed)、悬钩子(raspberry)、稻(rice)、黑麦(rye)、高粱(sorghum)、南方松(Southern pine)、大豆(soybean)、菠菜(spinach)、西葫芦(squash)、草莓(strawberry)、甜菜(sugarbeet)、甘蔗(sugarcane)、向日葵(sunflower)、甘薯(sweet potato)、枫香树(sweetgum)、橘(tangerine)、茶(tea)、烟草(tobacco)、番茄(tomato)、黑小麦(triticale)、草皮(turf)、芜菁(turnip)、藤(vine)、西瓜(watermelon)、小麦(wheat)、薯蓣(yam)、和绿皮西葫芦(zucchini)。在一些实施方案中,所述方法中有用的植物是拟南芥、玉米、小麦、大豆、和棉。

III. 方法

本发明的方法提供了包含选自下组的分泌信号多肽的融合蛋白的表达: pbp*、dsbA、dsbC、Bce、CupA2、CupB2、CupC2、NikA、FlgI、ORF5550、Ttg2C、或ORF812分泌信号。在一个实施方案中,该方法包括表达连接至本发明的分泌信号的感兴趣蛋白质的宿主细胞。该方法包括提供如下宿主细胞(优选荧光假单胞菌宿主细胞),其包含编码包含可操作连接至本文中所公

开的分泌信号序列的感兴趣蛋白质或多肽的重组蛋白的载体，并在导致所述蛋白质或多肽表达的条件下培养所述细胞。或者，使用所鉴定的分泌信号来表达蛋白质或多肽的方法可以在任何给定的宿主系统（包括真核或原核起源的宿主细胞）中使用。所述载体可具有上文所述特征之任一项。在一个实施方案中，载体包含编码本文中公开为SEQ ID NO:6、2、4、8、10、12、14、16、18、20、22、或24的分泌信号多肽或其变体和片段的核苷酸序列。在另一个实施方案中，载体包含包含SEQ ID NO:1、3、5、7、9、11、13、15、17、19、20、21、或23的核苷酸序列。

在另一个实施方案中，宿主细胞具有周质，并且分泌信号多肽的表达导致基本上所有的感兴趣蛋白质或多肽靶向至细胞周质。认识到周质中的小部分所表达蛋白质实际上可穿过细胞膜而泄露入胞外空间中；然而，大部分的靶向多肽会保留在周质间隙内。

表达可以进一步导致胞外蛋白质的生成。该方法还可包括从周质或从胞外培养基中纯化感兴趣蛋白质或多肽的步骤。分泌信号可以以如下方式表达，其中它是连接至蛋白质的，且可自细胞纯化与信号连接的蛋白质。因此，在一个实施方案中，这种分离的多肽是分泌信号与感兴趣蛋白质或多肽的融合蛋白。然而，在将蛋白质靶向至周质后，也可自蛋白质切除分泌信号。在一个实施方案中，对分泌信号和蛋白质或多肽之间的连接进行修饰以提高对分泌信号的切割。

本发明的方法还可导致宿主细胞内感兴趣蛋白质或多肽的产量(production)增加。或者，产量增加可以是每克所生成蛋白质或每克宿主蛋白质中正确加工的蛋白质或多肽的水平升高。产量增加还可以是每克重组物或每克宿主细胞蛋白质中所生产的可回收蛋白质或多肽的水平升高。产量增加还可以是总蛋白质的水平升高、正确加工的蛋白质的水平升高、或有活性的或可溶性的蛋白质的水平升高的任意组合。在此实施方案中，术语“升高的”是相对于没有本发明的分泌信号多肽的情况中在细胞中表达感兴趣蛋白质或多肽时生成的、正确加工的、可溶性的、和/或可回收的蛋白质或多肽的水平而言的。

感兴趣蛋白质或多肽的表达改善也可指蛋白质溶解度的增加。可生成感兴趣蛋白质或多肽，并自宿主细胞的细胞质、周质或胞外培养基回收。蛋白质或多肽可以是不溶的或可溶的。蛋白质或多肽可包括一种或多种靶向序列

以帮助纯化，如上文所讨论的。

如本文中所使用的，术语“可溶的”指通过以约5,000和20,000x重力之间的离心在旋转10-30分钟后在生理学条件下的缓冲液中蛋白质不被沉淀。可溶性蛋白质不是包含体或其它沉淀块的一部分。类似地，“不溶的”指通过以5,000和20,000x重力之间的离心在旋转10-30分钟后在生理学条件下的缓冲液中蛋白质或多肽可被沉淀。不溶性蛋白质或多肽可以是包含体或其它沉淀块的一部分。术语“包含体”意图包括细胞内含有的任何胞内体，其中已经隔离蛋白质或多肽的聚集体。

本发明的方法可生成定位于宿主细胞周质的蛋白质。在一个实施方案中，该方法在细胞中生成正确加工的感兴趣蛋白质或多肽。在另一个实施方案中，分泌信号多肽的表达可以在细胞中生成有活性的感兴趣蛋白质或多肽。与没有本发明的分泌信号的情况中表达蛋白质时相比，本发明的方法还可导致感兴趣蛋白质或多肽的产率(yield)提高。

在一个实施方案中，该方法在周质区室中生成至少0.1g/L蛋白质。在另一个实施方案中，该方法在细胞中生成0.1-10g/L周质蛋白质，或至少约0.2、约0.3、约0.4、约0.5、约0.6、约0.7、约0.8、约0.9或至少约1.0g/L周质蛋白质。在一个实施方案中，所生成的总体感兴趣蛋白质或多肽是至少1.0g/L、至少约2g/L、至少约3g/L、约4g/L、约5g/L、约6g/L、约7g/L、约8g/L、约10g/L、约15g/L、约20g/L、至少约25g/L、或更多。在一些实施方案中，所生成的周质蛋白质的量是所生成的总体感兴趣蛋白质或多肽的至少约5%、约10%、约15%、约20%、约25%、约30%、约40%、约50%、约60%、约70%、约80%、约90%、约95%、约96%、约97%、约98%、约99%、或更多。

在一个实施方案中，该方法生成至少0.1g/L正确加工的蛋白质。正确加工的蛋白质具有天然蛋白质的氨基端。在一些实施方案中，至少50%的感兴趣蛋白质或多肽包含天然氨基端。在另一个实施方案中，至少60%、至少70%、至少80%、至少90%、或更多的蛋白质具有天然蛋白质的氨基端。在多个实施方案中，该方法在细胞中生成0.1-10g/L正确加工的蛋白质，包括至少约0.2、约0.3、约0.4、约0.5、约0.6、约0.7、约0.8、约0.9或至少约1.0g/L正确加工的蛋白质。在另一个实施方案中，所生成的总体、正确加工的感兴趣蛋白质或多肽是至少1.0g/L、至少约2g/L、至少约3g/L、约4g/L、约5g/L、约6g/L、约7g/L、约8g/L、约10g/L、约15g/L、约20g/L、约25g/L、约30g/L、约35g/L、

约40g/l、约45g/l、至少约50g/L、或更大。在一些实施方案中，所生成的正确加工的蛋白质的量是正确加工形式的总体重组蛋白的至少约5%、约10%、约15%、约20%、约25%、约30%、约40%、约50%、约60%、约70%、约80%、约90%、约95%、约96%、约97%、约98%、至少约99%、或更多。

本发明的方法还可导致感兴趣蛋白质或多肽的产率提高。在一个实施方案中，该方法生成占总体细胞蛋白质(tcp)的至少约5%、至少约10%、约15%、约20%、约25%、约30%、约40%、约45%、约50%、约55%、约60%、约65%、约70%、约75%、或更大的感兴趣蛋白质或多肽。“百分比总体细胞蛋白质”是宿主细胞中蛋白质或多肽的量占聚集细胞蛋白质的百分比。百分比总体细胞蛋白质的测定在本领域中是公知的。

在一个具体的实施方案中，在矿物盐培养基中培养（即于约4°C-约55°C的温度范围内，包括约10°C、约15°C、约20°C、约25°C、约30°C、约35°C、约40°C、约45°C、和约50°C）时，宿主细胞可以具有至少1% tcp的重组多肽、多肽、蛋白质或其片段表达水平和至少40g/L的细胞密度。在一个特别优选的实施方案中，在矿物盐培养基中以至少约10L发酵规模培养（即于约4°C-约55°C的温度范围内，包括端点）时，表达系统会具有至少5% tcp的蛋白质或多肽表达水平和至少40g/L的细胞密度。

实际上，通常在培养基中找到靶向至周质的异源蛋白质（参见欧洲专利号EP 0 288 451），这可能是由于对细胞外膜的破坏或细胞外膜流动性的增加。这种“被动”分泌的速率可以通过使用透化细胞外膜的多种机制来提高：大肠菌素(Miksch等 (1997) Arch. Microbiol. 167: 143-150); 生长速率(Shokri等 (2002) App Microbiol Biotechnol 58:386-392); TolIII过表达(Wan和Baneyx (1998) Protein Expression Purif. 14: 13-22); 细菌素释放蛋白(Hsiung等 (1989) Bio/Technology 7: 267-71); 大肠菌素A溶解蛋白(Lloubes等 (1993) Biochimie 75: 451-8) 突变体，其泄露周质蛋白质(Furlong和Sundstrom (1989) Developments in Indus. Microbio. 30: 141-8); 融合配偶体(Jeong和Lee (2002) Appl. Environ. Microbio. 68: 4979-4985); 通过渗透压震扰进行的回收(Taguchi等 (1990) Biochimica Biophysica Acta 1049: 278-85)。工程化蛋白质向周质间隙的转运（随后定位于培养基中）已经用于在大肠杆菌中生成正确折叠的且有活性的蛋白质(Wan和Baneyx (1998) Protein Expression Purif. 14: 13-22; Simmons等 (2002) J. Immun. Meth. 263: 133-147; Lundell等 (1990) J. Indust.

Microbio. 5: 215-27)。

A. 活性蛋白质的生成

在一些实施方案中，蛋白质也可以以活性形式生成。术语“有活性的”指生物学活性的存在，其中所述生物学活性与相应的天然蛋白质或多肽的生物学活性相当或基本上对应。在蛋白质的背景中，这通常意味着多核苷酸或多肽包含在使用标准参数衡量时与相应的天然蛋白质或多肽相比具有至少约20%、约50%、优选至少约60-80%、最优选至少约90-95%活性的生物学功能或效应。可以利用对特定的蛋白质或多肽相应的标准靶向比较生物学测定法来实施蛋白质或多肽活性测定。感兴趣蛋白质或多肽维持生物学活性的一种指征是多肽与天然多肽有免疫学交叉反应性。

本发明还能改善有活性的感兴趣蛋白质或多肽的回收。活性蛋白质可具有衍生该序列的天然蛋白质或多肽的比活的至少约20%、至少约30%、至少约40%、约50%、约60%、至少约70%、约80%、约90%、或至少约95%的比活。此外，任选地，底物特异性(k_{cat}/K_m)基本上类似于天然蛋白质或多肽。通常， k_{cat}/K_m 会是至少约30%、约40%、约50%、约60%、约70%、约80%、至少约90%、至少约95%、或更大。测定和量化蛋白质和多肽活性和底物特异性(k_{cat}/K_m)的度量的方法对于本领域技术人员是公知的。

感兴趣蛋白质或多肽的活性也可以与先前建立的天然蛋白质或多肽标准活性进行比较。或者，感兴趣蛋白质或多肽的活性可以在同时的、或基本上同时的比较测定法中与天然蛋白质或多肽一起测定。例如，可以使用体外测定法来测定感兴趣蛋白质或多肽与靶物之间（例如所表达的酶与底物之间，所表达的激素与激素受体之间，所表达的抗体与抗原之间等）的任何可检测的相互作用。此检测可包括量热变化、增殖变化、细胞死亡、细胞排斥、放射性变化、溶解度变化、通过凝胶电泳和/或凝胶排阻方法所测量的分子量变化、磷酸化能力、抗体特异性测定法（诸如ELISA测定法）等的测量。另外，体内测定法包括但不限于用于检测与天然蛋白质或多肽的生理学效应相比假单胞菌所生成的蛋白质或多肽的生理学效应（例如重量增加、电解质平衡的变化、血液凝固时间的变化、凝块溶解的变化和抗原应答的诱导）的测定法。一般而言，可以使用任何体外或体内测定法来测定容许与天然蛋白质或多肽进行比较分析的感兴趣蛋白质或多肽的活性性质，只要此类活性是可测定的。或者，可以对本发明中所生成的蛋白质或多肽测定刺激或抑制蛋白

质或多肽和通常与所述蛋白质或多肽相互作用的分子（例如信号途径中通常与天然蛋白质相互作用的成分或底物）之间的相互作用的能力。此类测定法通常可以包括如下步骤，即在容许蛋白质或多肽与靶分子相互作用的条件下组合蛋白质与底物分子，并检测蛋白质与靶分子相互作用的生化后果。

可用于测定蛋白质或多肽活性的测定法记载于例如Ralph, P. J.等 (1984) J. Immunol. 132:1858或Saiki等 (1981) J. Immunol. 127:1044; Steward, W. E. II (1980) The Interferon Systems. Springer-Verlag, Vienna和New York; Broxmeyer, H. E.等 (1982) Blood 60:595; Molecular Cloning: A Laboratory Manual, 第2版, Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch和T. Maniatis编, 1989; 及Methods in Enzymology: Guide to Molecular Cloning Techniques, Academic Press, Berger, S. L.和A. R. Kimmel编, 1987; A K Patra等, Protein Expr Purif, 18(2): p/182-92 (2000); Kodama等, J. Biochem. 99: 1465-1472 (1986); Stewart等, Proc. Nat'l Acad. Sci. USA 90: 5209-5213 (1993); Lombillo等, J. Cell Biol. 128:107-115 (1995); Vale等, Cell 42:39-50 (1985)。

B. 细胞生长条件

本文中所描述的宿主细胞的细胞生长条件可以包括促进感兴趣蛋白质的表达的，和/或促进所表达的感兴趣蛋白质的发酵的。如本文中所使用的，术语“发酵”包括采用字面上的发酵的实施方案和采用其它非发酵培养模式的实施方案两者。发酵可以以任何规模实施。在一个实施方案中，发酵培养基可以自丰富培养基、基本培养基、和矿物盐培养基中选择；可使用丰富培养基，但优选避免丰富培养基。在另一个实施方案中，选择基本培养基或矿物盐培养基。在又一个实施方案中，选择基本培养基。在又一个实施方案中，选择矿物盐培养基。矿物盐培养基是特别优选的。

矿物盐培养基由矿物盐和碳源（诸如例如葡萄糖、蔗糖、或甘油）组成。矿物盐培养基的例子包括例如M9培养基、假单胞菌培养基(ATCC 179)、Davis和Mingioli培养基（参见BD Davis和ES Mingioli (1950) J. Bact. 60:17-28）。用于构成矿物盐培养基的矿物盐类包括那些自例如磷酸钾、硫酸铵或氯化铵、硫酸镁或氯化镁、和痕量矿物（诸如氯化钙、硼酸盐、及锌、锰、铜和铁的硫酸盐）之中选择的。矿物盐培养基中不包括有机氮源，诸如蛋白胨、胰蛋白胨、氨基酸或酵母提取物。使用无机氮源代替，并且这可以自例如铵盐、氨水、和氨气中选择。一种优选的矿物盐培养基会含有作为碳源的葡萄糖。

与矿物盐培养基相比，基本培养基也可以含有矿物盐和碳源，但是可以补充有例如低水平的氨基酸、维生素、蛋白胨、或其它成分，虽然这些以极低水平添加。

在一个实施方案中，可以使用下文表5中所列的成分来制备培养基。可以以如下顺序添加各成分：首先可将(NH₄)HPO₄、KH₂PO₄和柠檬酸溶解于约30L蒸馏水中；然后可添加痕量元素的溶液，接着添加消泡剂，诸如Ucolub N 115。然后，热灭菌（诸如于约121°C）后，可以添加葡萄糖MgSO₄和盐酸硫胺素的无菌溶液。可以使用氨水来将pH控制在约6.8。然后可添加无菌蒸馏水来将起始体积调至371减去甘油原液(123mL)。化学品可以购自多个供应商，诸如Merck。对于假单胞菌物种和相关细菌的生长，这种培养基能容许高细胞密度培养(HCDC)。HCDC可以以分批过程开始，接着是两阶段补料-分批培养。在分批部分中不受限制的生长后，可以在3个倍增时间的期间里将生长控制在降低的比生长速率，其中生物质浓度可增加数倍。此类培养规程的进一步详情记载于Riesenberg, D.; Schulz, V.; Knorre, W. A.; Pohl, H. D.; Korz, D.; Sanders, E. A.; Ross, A.; Deckwer, W. D. (1991) “High cell density cultivation of. Escherichia coli, at controlled specific growth rate” J Biotechnol: 20(1) 17-27。

表5: 培养基组成	
成分	起始浓度
KH ₂ PO ₄	13.3g l ⁻¹
(NH ₄) ₂ HPO ₄	4.0g l ⁻¹
柠檬酸	1.7g l ⁻¹
MgSO ₄ -7H ₂ O	1.2g l ⁻¹
痕量金属溶液	10ml l ⁻¹
盐酸硫胺素	4.5mg l ⁻¹
葡萄糖-H ₂ O	27.3g l ⁻¹
消泡剂Ucolub N115	0.1ml l ⁻¹

补料溶液	
MgSO ₄ -7H ₂ O	19.7g l ⁻¹
葡萄糖-H ₂ O	770g l ⁻¹
NH ₃	23g
痕量金属溶液	
6g l ⁻¹ 柠檬酸铁(III) 1.5 g l ⁻¹	
MnCl ₂ -4H ₂ O	
0.8g l ⁻¹ ZnCH ₂ COOI ₂ -2H ₂ O 0.3 g l ⁻¹	
H ₃ BO ₃	
0.25g l ⁻¹ Na ₂ MoO ₄ -2H ₂ O 0.25 g l ⁻¹	
CoCl ₂ 6H ₂ O	
0.15g l ⁻¹ CuCl ₂ 2H ₂ O 0.84 g l ⁻¹	
ethylene	
Dinitrilo-tetracetic acid Na ₂ sah	
2H ₂ O	
(Tritriplex III, Merck)	

依照本发明的表达系统可以以任何发酵形式培养。例如，本文中可采用分批、补料-分批、半连续、和连续发酵模式。若蛋白质被分泌入胞外培养基中，则优选连续发酵。

依照本发明的表达系统对于任何发酵规模（即体积）的转基因表达是有用的。如此，例如可以使用微升规模、厘升规模、和分升规模发酵体积；并且可以使用1升规模和更大的发酵体积。在一个实施方案中，发酵体积会处于或超过1升。在另一个实施方案中，发酵体积会处于或超过5升、10升、15升、20升、25升、50升、75升、100升、200升、500升、1,000升、2,000升、5,000升、10,000升、或50,000升。

在本发明中，在容许宿主细胞生存的温度范围内，优选于约4°C-约55°C的范围内（包括端点）的温度实施受转化宿主细胞的生长、培养和/或发酵。

如此，例如如本文中关于本发明宿主细胞所使用的，术语“生长”、“培养”和“发酵”内在地指约4°C-约55°C的温度范围内（包括端点）的“生长”、“培养”和“发酵”。另外，“生长”用于指明活跃的细胞分裂和/或扩大的生物学状态以及非分裂和/或非扩大的细胞在新陈代谢方面得到维持的生物学状态两者，术语“生长”的后者应用与术语“维持”同义。

在表达分泌型蛋白质中使用荧光假单胞菌的另一项优点包括与一些其它细菌表达系统相比荧光假单胞菌以高细胞密度培养的能力。为此，依照本发明的荧光假单胞菌表达系统可以提供约20g/L或更多的细胞密度。依照本发明的荧光假单胞菌表达系统同样也能提供至少约70g/L的细胞密度，就每体积生物质而言，生物质以细胞干重测量。

在一个实施方案中，细胞密度会是至少约20g/L。在另一个实施方案中，细胞密度会是至少约25g/L、约30g/L、约35g/L、约40g/L、约45g/L、约50g/L、约60g/L、约70g/L、约80g/L、约90g/L、约100g/L、约110g/L、约120g/L、约130g/L、约140g/L、约或至少约150g/L。

在另一个实施方案中，诱导时的细胞密度会在约20g/L和约150g/L之间；在约20g/L和约120g/L之间；在约20g/L和约80g/L之间；在约25g/L和约80g/L之间；在约30g/L和约80g/L之间；在约35g/L和约80g/L之间；在约40g/L和约80g/L之间；在约45g/L和约80g/L之间；在约50g/L和约80g/L之间；在约50g/L和约75g/L之间；在约50g/L和约70g/L之间；在约40g/L和约80g/L之间。

C. 感兴趣蛋白质或多肽的分离

为了测量感兴趣蛋白质的产量(yield)、溶解度、构象、和/或活性，自宿主细胞和/或胞外培养基分离蛋白质可能是想要的。分离可以是粗略的、半粗略的、或纯净的分离，这取决于用于进行合适的测量的测定法的要求。蛋白质可以在细胞质中生成，靶向至周质，或可以分泌入培养物或发酵培养基中。为了自周质释放靶向蛋白质，使用了牵涉如下化学品的处理，诸如氯仿(Ames等 (1984) *J. Bacteriol.*, 160: 1181-1183)、盐酸胍、和Triton X-100 (Naglak和Wang (1990) *Enzyme Microb. Technol.*, 12: 603-611)。然而，这些化学品不是惰性的，并且可能对许多重组蛋白产物或随后的纯化规程具有有害的影响。也有报告，对大肠杆菌细胞的甘氨酸处理（引起外膜的透化）释放周质内容物(Arigo等 (1989) *J. Ferm. Bioeng.*, 68: 243-246)。重组蛋白的周质释放的最广泛使用的方法是渗透压震扰(Nosal和Heppel (1966) *J. Biol. Chem.*, 241:

3055-3062; Neu和Heppel (1965) J. Biol. Chem., 240: 3685-3692)、鸡蛋清(HEW)溶菌酶/乙二胺四乙酸(EDTA)处理(Neu和Heppel (1964) J. Biol. Chem., 239: 3893-3900; Witholt等 (1976) Biochim. Biophys. Acta, 443: 534-544; Pierce等 (1995) ICheme Research. Event, 2: 995-997)、和HEW溶菌酶/渗透压震荡组合处理(French等 (1996) Enzyme and Microb. Tech., 19: 332-338)。弗氏法牵涉细胞在分级缓冲液中的重悬,接着是周质级分的回收,其中渗透压震荡就在溶菌酶处理后。

通常,这些规程包括渗透压稳定化培养基中的起始破坏,接着是非稳定化培养基中的选择性释放。这些培养基的组成(pH、保护剂)和所使用的破坏方法(氯仿、HEW溶菌酶、EDTA、超声处理)在所报告的具体规程中有所不同。使用双极性离子型去垢剂替换EDTA的一种HEW溶菌酶/EDTA处理变化形式记载于Stabel等 (1994) Veterinary Microbiol., 38: 307-314。关于使用胞内容胞酶系统来破坏大肠杆菌的一般性综述,参见Dabora和Cooney (1990)于Advances in Biochemical Engineering/Biotechnology, 卷43, A. Fiechter,编(Springer-Verlag: Berlin), 页11-30。

用于自细胞质以可溶性蛋白质或折射颗粒回收感兴趣蛋白质或多肽的常规方法牵涉通过机械破坏来碎裂细菌细胞。机械破坏通常牵涉在液体悬浮液中产生局部空化、用刚性珠子进行的快速搅动、超声处理、或细胞悬液的研磨(Bacterial Cell Surface Techniques, Hancock和Poxton (John Wiley & Sons Ltd, 1988), 章3, 页55)。

HEW溶菌酶以生化方式起水解细胞壁肽聚糖骨架的作用。该方法最初由Zinder和Arndt (1956) Proc. Natl. Acad. Sci. USA, 42: 586-590开发,他们用卵清蛋白(其含有HEW溶菌酶)处理大肠杆菌以产生称为原生质球的圆形细胞球。这些结构保留了一些细胞壁成分但是具有大的表面积,其中细胞质膜是暴露的。美国专利No. 5,169,772披露了用于自细菌纯化肝素酶的方法,包括使用例如EDTA、溶菌酶、或有机化合物在渗透压稳定化培养基(例如20%蔗糖溶液)中破坏细菌的包膜,通过将细菌暴露于低离子强度缓冲液来自被破坏细菌的周质间隙释放非肝素酶样蛋白质,和通过将经低离子强度清洗的细菌暴露于缓冲盐溶液来释放肝素酶样蛋白质。

已经对广泛的表达系统使用了这些方法的许多不同改良,获得了不同程度的成功(Joseph-Liaun等 (1990) Gene, 86: 291-295; Carter等 (1992)

Bio/Technology, 10: 163-167)。诱导重组细胞培养物生成溶菌酶的努力已经有报告。EP 0,155,189披露了用于诱导重组细胞培养物生成溶菌酶的手段,通常预期所述手段会通过破坏或溶解细胞壁结构的手段来杀死此类宿主细胞。

美国专利No. 4,595,658披露了用于促进转运至大肠杆菌周质间隙的蛋白质外化(externalization)的方法。这种方法容许在不需要对细胞进行溶菌酶处理、机械研磨、或渗透压震扰处理的情况中选择性分离位于周质中的蛋白质。美国专利No. 4,637,980披露了如下生成细菌产物,即用编码(直接地或间接地)产物的DNA分子转化温度敏感性溶源体,在允许条件下培养转化体以在胞内表达基因产物,并通过提高温度来诱导噬菌体编码的功能而使产物外化。Asami等(1997) J. Ferment. and Bioeng., 83: 511-516披露了通过T4噬菌体感染来同步破坏大肠杆菌细胞,而Tanji等(1998) J. Ferment. and Bioeng., 85: 74-78披露了T4噬菌体中所编码溶解基因的受控表达以温和破坏大肠杆菌细胞。

细胞溶解后,基因组DNA泄露出细胞质而进入培养基中,并导致流体粘度的显著升高,这可阻止固体在离心场中的沉降。在没有剪切力(诸如那些在机械破坏过程中所施加的用以打碎DNA聚合物的)的情况下,固体穿过粘性流体的较慢沉降速率导致离心期间固体与流体的分开不足。除机械剪切力之外,还存在有降解DNA聚合物的溶核酶。在大肠杆菌中,内源基因endA编码内切核酸酶(成熟蛋白质的分子量为约24.5kD),其通常分泌至周质,并以内切溶核方式将DNA切割成寡脱氧核糖核苷酸。已经有提示, endA由大肠杆菌相对较弱地表达(Wackemagel等(1995) Gene 154: 55-59)。

在一个实施方案中,不需要额外的促进二硫键的条件或试剂以便自宿主细胞回收有活性的、可溶性形式的含有二硫键的所鉴定的多肽。在一个实施方案中,转基因多肽、多肽、蛋白质、或其片段在其活性状态中具有折叠的分子内构象。在一个实施方案中,转基因多肽、多肽、蛋白质、或其片段在其活性状态中含有至少一个分子内二硫键;并且或许有多至2个、4个、6个、8个、10个、12个、14个、16个、18个、或20个或更多个二硫键。

可以通过本领域中公知的标准技术分离本发明的蛋白质并纯化至基本上纯净,包括但不限于硫酸铵或乙醇沉淀、酸提取、阴离子或阳离子交换层析、磷酸纤维素层析、疏水相互作用层析、亲和层析、镍层析、羟磷灰石层析、反相层析、凝集素层析、制备电泳、去污剂增溶、用诸如柱层析的物质

进行的选择性沉淀、免疫纯化法等。例如，可以将具有已建立的分子粘附特性的蛋白质与配体可逆地融合。凭借合适的配体，可以将蛋白质选择性吸附至纯化柱，然后以相对纯净的形式从柱中释放。然后通过酶活性切出所融合的蛋白质。另外，可以使用免疫亲和柱或Ni-NTA柱来纯化蛋白质。一般性技术进一步记载于例如R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag: N.Y. (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990); 美国专利No. 4,511,503; S. Roe, *Protein Purification Techniques: A Practical Approach* (Practical Approach Series), Oxford Press (2001); D. Bollag等, *Protein Methods*, Wiley-Lisa, Inc. (1996); AK Patra等, *Protein Expr Purif*, 18(2): p/182-92 (2000); 及R. Mukhija等, *Gene* 165(2): 页303-6 (1995)。还可参见例如Ausubel等 (1987和定期补充); Deutscher (1990) “Guide to Protein Purification”, *Methods in Enzymology* 卷182, 及此丛书中的其它卷; Coligan等 (1996和定期补充) *Current Protocols in Protein Science* Wiley/Greene, NY; 及制造商关于蛋白质纯化产品的文献, 例如Pharmacia, Piscataway, N.J.或Bio-Rad, Richmond, Calif. 与重组技术的组合容许与合适的区段融合, 例如与FLAG序列或等同物, 它们可经由蛋白酶可切除的序列来融合。还可参见例如Hochuli (1989) *Chemische Industrie* 12:69-70; Hochuli (1990) “Purification of Recombinant Proteins with Metal Chelate Absorbent” 于 Setlow (编) *Genetic Engineering, Principle and Methods* 12:87-98, Plenum Press, NY; 及Crowe等 (1992) *QIAexpress: The High Level Expression & Protein Purification System* QUIAGEN, Inc., Chatsworth, Calif.

通过本领域中已知的方法来实现所表达蛋白质的检测, 包括例如放射性免疫测定法、Western印迹技术或免疫沉淀。

本发明中所表达的某些蛋白质可以形成不溶性聚集体 (“包含体”)。数种方案适于自包含体纯化蛋白质。例如, 包含体的纯化通常牵涉通过破坏宿主细胞 (例如通过在50mM TRIS/HCL pH7.5、50mM NaCl、5mM MgCl₂、1mM DTT、0.1mM ATP、和1mM PMSF的缓冲液中温育) 对包含体的提取、分离和/或纯化。通常通过2-3次通过弗氏压碎器来溶解细胞悬浮液。细胞悬浮液也可以使用Polytron (Brinkman Instruments)进行匀浆化或在冰上进行超声处理。溶解细菌的备选方法对于本领域技术人员是显而易见的 (参见例如 Sambrook等, 见上文; Ausubel等, 见上文)。

在必要时，可以溶解包含体，并且通常可以对溶解的细胞悬浮液离心，以便除去不想要的不可溶性物质。可以通过用相容的缓冲液进行的稀释或透析来使形成包含体的蛋白质复性。合适的溶剂包括但不限于尿素（约4M-约8M）、甲酰胺（至少约80%，体积/体积计）、和盐酸胍（约4M-约8M）。虽然盐酸胍和类似的试剂是变性剂，但是这种变性不是不可逆的，并且复性可以在除去（例如通过透析）或稀释变性剂后发生，这容许免疫学和/或生物学活性蛋白质的重新形成。其它合适的缓冲液对于本领域技术人员是已知的。

通过本领域技术人员公知的标准分离技术能将上清液中存在的异源表达的蛋白质与宿主蛋白质分开。例如，起始的盐分级可将许多不想要的宿主细胞蛋白质（或自细胞培养基衍生的蛋白质）与感兴趣蛋白质或多肽分开。一个这样的例子可以是硫酸铵。硫酸铵通过有效地降低蛋白质混合物中的水量而使蛋白质沉淀。然后蛋白质根据它们的溶解度而沉淀。蛋白质越疏水，其越有可能在较低的硫酸铵浓度沉淀。一个典型方案包括将饱和硫酸铵添加至蛋白质溶液，使得所得硫酸铵浓度在20-30%之间。此浓度会沉淀大多数疏水性蛋白质。然后弃去沉淀物（除非感兴趣蛋白质是疏水的），并将硫酸铵添加至上清液至使感兴趣蛋白质沉淀的已知浓度。然后将沉淀物溶解于缓冲液中，并在必要时除去过量的盐，其经由透析或渗滤实现。依赖于蛋白质溶解度的其它方法（诸如冷乙醇沉淀）对于本领域技术人员是公知的，并可用于对复杂的蛋白质混合物进行分级。

可以利用感兴趣蛋白质或多肽的分子量来将其与较大和较小尺寸的蛋白质分离，其使用穿过不同孔径的膜（例如Amicon或Millipore膜）的超滤来实现。作为第1步，蛋白质混合物可以穿过孔径的分子量截留比感兴趣蛋白质的分子量低的膜而进行超滤。然后超滤的保留物可以针对分子截留比感兴趣蛋白质的分子量大的膜进行超滤。感兴趣蛋白质或多肽会穿过膜而进入滤液中。然后可以对滤液进行层析，如下文所述的。

分泌型感兴趣蛋白质或多肽也可以根据其大小、表面净电荷、疏水性、和对配体的亲和力而与其它蛋白质分开。另外，可以将针对蛋白质所产生的抗体偶联至柱基质，并对蛋白质免疫纯化。所有的这些方法在本领域中是公知的。对于技术人员会显而易见的是，可以以任何规模和使用来自许多不同制造商（例如Pharmacia Biotech）的仪器进行层析技术。

D. 复性和重折叠

在本发明的一些实施方案中，所生成的超过50%的所表达的转基因多肽、多肽、蛋白质或其片段可以在宿主细胞中以可复性形式生成。在另一个实施方案中，约60%、70%、75%、80%、85%、90%、95%的所表达的蛋白质以活性形式获得或可以复性成活性形式。

不溶性蛋白质可以复性或重折叠以产生二级和三级蛋白质结构构象。必要时，可以在完成重组产物的构造中使用蛋白质重折叠步骤。可以使用本领域中已知的用于促进蛋白质解离/结合的试剂来实现重折叠和复性。例如，可将蛋白质与二硫苏糖醇一起温育，接着与氧化型谷胱甘肽二钠盐一起温育，接着与含有重折叠试剂（诸如尿素）的缓冲液一起温育。

也可以例如通过使其针对磷酸盐缓冲盐水(PBS)或50mM醋酸钠，pH6缓冲液加200mM NaCl的透析来使感兴趣蛋白质或多肽复性。或者，可以在固定化于柱（诸如Ni NTA柱）上时，通过使用500mM NaCl、20% 甘油、20mM Tris/HCl pH7.4（含有蛋白酶抑制剂）中的6M-1M尿素线性梯度来使蛋白质重折叠。可以在1.5小时或更长的一段时间里实施复性。复性后，可以通过添加250mM咪唑来洗脱蛋白质。可以通过针对PBS或50mM醋酸钠pH 6缓冲液加200mM NaCl的最终透析步骤来除去咪唑。纯化的蛋白质可以于4°C贮存或于-80°C冷冻。

其它方法包括例如那些可记载于M H Lee等, Protein Expr. Purif., 25(1): 页166-73 (2002); W. K. Cho等, J. Biotechnology, 77(2-3): 页169-78 (2000); Ausubel等(1987和定期补充); Deutscher (1990) "Guide to Protein Purification," Methods in Enzymology 卷182, 及此丛书中的其它卷; Coligan等(1996和定期补充) Current Protocols in Protein Science Wiley/Greene, NY; S. Roe, Protein Purification Techniques: A Practical Approach (Practical Approach Series), Oxford Press (2001); D. Bollag等, Protein Methods, Wiley-Lisa, Inc. (1996)。

E. 感兴趣蛋白质

本发明的方法和组合物对于在细胞表达系统中生成高水平的正确加工的感兴趣蛋白质或多肽是有用的。感兴趣蛋白质或多肽（在本文中也称为“靶蛋白质”或“靶多肽”）可以是任何物种的和任何大小的。然而，在某些实施方案中，感兴趣蛋白质或多肽是在治疗方面有用的蛋白质或多肽。在一些实施方案中，蛋白质可以是哺乳动物蛋白质，例如人蛋白质，而且可以是例

如生长因子、细胞因子、趋化因子或血液蛋白质。感兴趣蛋白质或多肽可以以与天然蛋白质或多肽类似的方式加工。在某些实施方案中，蛋白质或多肽不包括编码序列中的分泌信号。在某些实施方案中，感兴趣蛋白质或多肽是小于100kD、小于50kD、或小于30kD大小的。在某些实施方案中，感兴趣蛋白质或多肽是具有至少约5个、10个、15个、20个、30个、40个、50个或100个氨基酸的多肽。

分子遗传学和遗传工程技术所要求的大量序列信息是可广泛地公开获得的。可自GenBank于网站//www.ncbi.nlm.nih.gov/Entrez获得哺乳动物以及人的基因、cDNA序列、氨基酸序列和基因组的全部核苷酸序列的访问(access)。别的信息也可自GeneCards (来自魏茨曼科学基因组和生物信息学研究所(Weizmann Institute of Science Genome and Bioinformatics)的整合关于基因及其产物和生物医学应用的信息的电子百科全书(bioinformatics.weizmann.ac.il/cards))获得,核苷酸序列信息也可自EMBL核苷酸序列数据库(www.ebi.ac.uk/embl/)或日本DNA数据库(DDBJ, www.ddbi.nig.ac.jp)获得;关于氨基酸序列的信息的别的位置包括Georgetown的蛋白质信息资源网站(www-nbrf.Georgetown.edu/pirl)和Swiss-Prot (au.expasy.org/sprot/sprot-top.html)。

可以在本发明中表达的蛋白质的例子包括如下分子,诸如例如,肾素、生长激素,包括人生长激素;牛生长激素;生长激素释放因子;甲状旁腺激素;促甲状旁腺激素;脂蛋白; α -1-抗胰蛋白酶;胰岛素A链;胰岛素B链;胰岛素原;血小板生成素;促卵泡激素;降钙素;黄体生成素;胰高血糖素;凝血因子,诸如因子VIII、因子IX、组织因子、和von Willebrands因子;抗凝血因子,诸如蛋白C;心房钠尿肽;肺表面活性剂;纤溶酶原激活物,诸如尿激酶或人尿或组织型纤溶酶原激活物(t-PA);铃蟾肽;凝血酶;造血生长因子;肿瘤坏死因子- α 和- β ;脑啡肽酶;血清清蛋白,诸如人血清清蛋白;缪勒抑制物质(mullerian-inhibiting substance);松弛素A链;松弛素B链;松弛素原;小鼠促性腺激素相关多肽;微生物蛋白质,诸如 β -内酰胺酶;DNA酶;抑制素;活化素(activin);血管内皮生长因子(VEGF);激素的或生长因子的受体;整联蛋白;蛋白A或D;类风湿因子;神经营养因子,诸如脑衍生神经营养因子(BDNF)、神经营养因子-3、-4、-5、或-6 (NT-3、NT-4、NT-5、或NT-6)、或神经生长因子,诸如NGF- β ;心肌营养蛋白(心脏肥大因子),

诸如心肌营养蛋白-1 (CT-1); 血小板衍生生长因子(PDGF); 成纤维细胞生长因子, 诸如aFGF和bFGF; 表皮生长因子(EGF); 转化生长因子(TGF), 诸如TGF- α 和TGF- β , 包括TGF- β 1、TGF- β 2、TGF- β 3、TGF- β 4、或TGF- β 5; 胰岛素样生长因子-I和-II (IGF-I和IGF-II); des(1-3)-IGF-I (脑IGF-I)、胰岛素样生长因子结合蛋白; CD蛋白, 诸如CD-3、CD-4、CD-8、和CD-19; 红细胞生成素; 骨诱导因子; 免疫毒素; 骨形态发生蛋白(BMP); 干扰素, 诸如干扰素- α 、- β 、和- γ ; 集落刺激因子(CSF), 例如M-CSF、GM-CSF、和G-CSF; 白介素(IL), 例如IL-1至IL-10; 抗HER-2抗体; 超氧化物歧化酶; T细胞受体; 表面膜蛋白; 衰变加速因子; 病毒抗原, 诸如例如AIDS被膜的一部分; 转运蛋白; 归巢受体; 地址素; 调控蛋白; 抗体; 及上文所列多肽中任一种的片段。

在某些实施方案中, 蛋白质或多肽可选自IL-1、IL-1a、IL-1b、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、IL-12elasti、IL-13、IL-15、IL-16、IL-18、IL-18BP α 、IL-23、IL-24、VIP、红细胞生成素、GM-CSF、G-CSF、M-CSF、血小板衍生生长因子(PDGF)、MSF、FLT-3配体、EGF、成纤维细胞生长因子 (FGF; 例如 α -FGF (FGF-1)、 β -FGF (FGF-2)、FGF-3、FGF-4、FGF-5、FGF-6、或FGF-7)、胰岛素样生长因子 (例如IGF-1、IGF-2); 肿瘤坏死因子 (例如TNF、淋巴毒素)、神经生长因子 (例如NGF)、血管内皮生长因子(VEGF); 干扰素 (例如IFN- α 、IFN- β 、IFN- γ); 白血病抑制因子(LIF); 睫状节神经营养因子(CNTF); 制瘤素M; 干细胞因子(SCF); 转化生长因子 (例如TGF- α 、TGF- β 1、TGF- β 2、TGF- β 3); TNF超家族 (例如LIGHT/TNFSF14、STALL-1/TNFSF13B (BLy5、BAFF、THANK)、TNF α /TNFSF2和TWEAK/TNFSF12); 或趋化因子 (BCA-1/BLC-1、BRAK/Kec、CXCL16、CXCR3、ENA-78/LIX、Eotaxin-1、Eotaxin-2/MPIF-2、Exodus-2/SLC、Fractalkine/Neurotactin、GRO α /MGSA、HCC-1、I-TAC、Lymphotactin/ATAC/SCM、MCP-1/MCAF、MCP-3、MCP-4、MDC/STCP-1/ABCD-1、MIP-1.quadrature、MIP-1.quadrature、MIP-2.quadrature./GRO.quadrature、MIP-3.quadrature./Exodus/LARC、MIP-3/Exodus-3/ELC、MIP-4/PARC/DC-CK1、PF-4、RANTES、SDF1、TARC、或TECK)。

在本发明的一个实施方案中, 感兴趣蛋白质可以是多亚基蛋白质或多

肽。可以表达的多亚基蛋白质包括同聚体和异聚体蛋白质。多亚基蛋白质可以包括两个或更多个亚基，它们可以是相同的或不同的。例如，蛋白质可以是包含2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个或更多个亚基的同聚体蛋白质。蛋白质也可以是包含2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、或更多个亚基的异聚体蛋白质。例示性的多亚基蛋白质包括：受体，包括离子通道受体；胞外基质蛋白质，包括软骨素；胶原；免疫调节剂，包括MHC蛋白、全链抗体、和抗体片段；酶，包括RNA聚合酶、和DNA聚合酶；及膜蛋白。

在另一个实施方案中，感兴趣蛋白质可以是血液蛋白质。在此实施方案中所表达的血液蛋白质包括但不限于载体蛋白，诸如清蛋白（包括人的和牛的清蛋白），运铁蛋白、重组运铁蛋白半分子、触珠蛋白、血纤蛋白原和其它凝结因子、补体成分、免疫球蛋白、酶抑制剂、物质（诸如血管紧张素和缓激肽）的前体、胰岛素、内皮缩血管肽、和球蛋白（包括 α 、 β 、 γ -球蛋白），及主要在哺乳动物的血液中的其它类型的蛋白质、多肽及其片段。已经报告了许多血液蛋白质的氨基酸序列（参见S. S. Baldwin (1993) *Comp. Biochem Physiol.* 106b:203-218），包括人血清清蛋白的(Lawn, L. M.等 (1981) *Nucleic Acids Research*, 9:6103-6114)和人血清运铁蛋白的氨基酸序列(Yang, F.等 (1984) *Proc. Natl. Acad. Sci. USA* 81:2752-2756)。

在另一个实施方案中，感兴趣蛋白质可以是重组体的酶或辅因子。在此实施方案中所表达的酶和辅因子包括但不限于醛缩酶、胺氧化酶、氨基酸氧化酶、天冬氨酸酶、B12依赖性酶、羧肽酶、羧酸酯酶(carboxyesterase)、羧酸分解酶(carboxylyase)、化学胰蛋白酶(chemotrypsin)、需要CoA的酶(CoA requiring enzyme)、羧腈合成酶、胱硫醚合酶(cystathione synthase)、脱羧酶、脱氢酶、醇脱氢酶、脱水酶、心肌黄酶(diaphorase)、加双氧酶、烯酸还原酶(enoate reductase)、环氧化物水化酶(epoxide hydase)、延胡索酸酶(fumerase)、半乳糖氧化酶、葡萄糖异构酶、葡萄糖氧化酶、糖基转移酶、甲基转移酶、腈水化酶(nitrile hydase)、核苷磷酸化酶、氧化还原酶、氧腈酶(oxynitilase)、肽酶、糖基转移酶、过氧化物酶、与治疗活性多肽融合的酶、组织纤溶酶原激活物；尿激酶、蛇毒凝血酶(reptilase)、链激酶；过氧化氢酶、超氧化物歧化酶；DNA酶、氨基酸水解酶（例如天冬酰胺酶、酰胺水解酶）；羧肽酶；蛋白酶、胰蛋白酶、胃蛋白酶、胰凝乳蛋白酶、木瓜蛋白酶、菠萝蛋白酶、

胶原酶；神经氨酸酶；乳糖酶、麦芽糖酶、蔗糖酶、和阿拉伯呋喃糖苷酶。

在另一个实施方案中，感兴趣蛋白质可以是单链、Fab片段和/或全链抗体或其片段或部分。单链抗体可以包括抗体中在稳定折叠的多肽单链上的抗原结合区。Fab片段可以是特定抗体中的一段。Fab片段可以含有抗原结合位点。Fab片段可以含有2条链：轻链和重链片段。这些片段可经由接头或二硫键连接。

感兴趣蛋白质或多肽的编码序列可以是靶多肽的天然编码序列（若可获得的话），但是会更优选的是如下编码序列，其已经经过选择、改良或优化以在选定的表达宿主细胞中使用：例如，通过合成反映假单胞菌物种（诸如荧光假单胞菌）或其它合适的生物体的密码子使用偏爱的基因来实现。所得的基因会在一种或多种载体内得到构建或会被插入一种或多种载体中，然后所述载体会被转化入表达宿主细胞中。要以“可表达形式”提供的核酸或多核苷酸指含有能由选定的表达宿主细胞表达的至少一种基因的核酸或多核苷酸。

在某些实施方案中，感兴趣蛋白质是天然蛋白质（诸如天然哺乳动物或人蛋白质）或与其基本上同源。在这些实施方案中，蛋白质没有以多联体形式找到，而是仅连接至分泌信号和任选的供纯化和/或识别用的标签序列。

在其它实施方案中，感兴趣蛋白质是在约20至约42°C的温度有活性的蛋白质。在一个实施方案中，蛋白质在生理温度是有活性的，而在加热至高的或极端的温度（诸如超过65°C的温度）时被灭活。

在一个实施方案中，感兴趣蛋白质是如下蛋白质，其在约20至约42°C的温度是有活性的，和/或在加热至高的或极端的温度（诸如超过65°C的温度）时被灭活；是天然蛋白质（诸如天然哺乳动物或人蛋白质）或与其基本上同源，并且不是自多联体形式的核酸表达的；并且启动子不是荧光假单胞菌中的天然启动子，而是衍生自另一种生物体，诸如大肠杆菌。

在其它实施方案中，蛋白质在生成时还包括别的靶向序列，例如将蛋白质靶向至胞外培养基的序列。在一个实施方案中，别的靶向序列可操作连接至蛋白质的羧基端。在另一个实施方案中，蛋白质包括自运输物、双配偶体分泌系统、主端分枝系统或菌毛引座孔蛋白的分泌信号。

为了例示而非为了限制，提供了下列实施例。

实施例

实施例1: *dsbC*前导序列的鉴定

I. 材料和方法

A. pDOW2258表达质粒的构建

在用于表达DsbC前导肽-Skp融合蛋白的质粒pDOW2258的构建中使用标准重组DNA技术(图1)。

使用Herculase Master混合物(Stratagene, #600610-51)、引物RC-322 (5'-AATTACTAGTAGGAGGTACATTATGCGCTT-3', SEQ ID NO: 25)和RC-323 (5'-TATACTCGAGTTATTAACTGTTTCAGTA-3', SEQ ID NO: 26)、和模板质粒pDOW3001(已经含有通过SOE PCR所产生的克隆的*dsbC*前导-*skp*编码序列融合物)使用制造商的方案来实施PCR扩增反应以扩增521bp *dsbC-skp*编码序列。使用QIAQUICK®凝胶提取试剂盒(Qiagen, #28704)来纯化PCR片段,用*SpeI*和*XhoI*限制性核酸酶(New England Biolabs, R0133和R0146)来消化,然后使用T4 DNA连接酶(New England Biolabs, M0202)连接至表达质粒pDOW1169(已经用*SpeI*和*XhoI*消化的),其依照制造商的方案实现。通过电穿孔将连接反应转化入荧光假单胞菌DC454 (*lsc::lacI^{Q1} ΔpyrF*)中,在含大豆的SOC培养基(SOC-with-Soy medium)中恢复,并铺板于选择培养基(M9葡萄糖琼脂)上。通过对质粒DNA的限制性消化(Qiagen, 产品目录编号27106)来分析菌落。10个含有插入物的克隆经测序证实了无错*dsbC-skp*编码序列的存在。将来自经序列证实的分离群的质粒称为pDOW2258。

B. 摇瓶中的生长和表达分析

通过标准的Dow 1L规模摇瓶表达方案来分析含有pDOW2258的荧光假单胞菌菌株DC454 (*lsc::lacI^{Q1} ΔpyrF*)分离群。简言之,使用在补充有1%葡萄糖和痕量元素的M9培养基中所培养的种子培养物来接种200mL含有作为碳源的5%甘油的限定极限盐培养基(defined minimal salts medium)。起始的24小时生长阶段后,用0.3mM异丙基-β-D-1-硫代吡喃半乳糖苷(IPTG)诱导经由*Ptac*启动子的表达。

在诱导时(I0)、诱导后24小时(I24)、和诱导后48小时(I48)对培养物取样。通过600nm的光密度(OD₆₀₀)来测量细胞密度。将细胞密度调节至OD₆₀₀ = 20,并将100μL的等分试样以14,000 x rpm离心5分钟,并除去上清液。

使用EASYLYSE™ (Epicentre Technologies)自摇瓶样品产生可溶性和不

溶性级分。重悬细胞沉淀物，在溶解缓冲液中以1:4稀释，并在摇动的情况中于室温温育30分钟。将溶胞物以14,000rpm离心20分钟(4°C)，并除去上清液。将上清液作为可溶性级分保存。将样品与含有β-巯基乙醇(BioRad产品目录编号161-0737)的2X Laemmli样品缓冲液以1:1混合，并煮沸5分钟，之后在12% Bis-Tris凝胶(BioRad产品目录编号45-0112 批号cx090505C2)上加载20μL，并在1X MES缓冲液(产品目录编号161-0788 批号210001188)中电泳。用SIMPLYBLUE™ SafeStain (Invitrogen产品目录编号LC6060)依照制造商的方案对凝胶染色，并使用Alpha Innotech成像系统进行成像。

C. N端测序分析

使用10mM CAPS (2.21g/L) pH11 (含NaOH) 和10%甲醇作为转移缓冲液将通过SDS-PAGE分开的可溶性和不溶性级分以40V 1.5小时转移至测序级PVDF膜(Bio-Rad, 产品目录编号162-0236)。将印迹在染色溶液(0.2%考马斯亮蓝R-250、40%甲醇、10%乙酸)中染色10秒，然后立即脱色3次，每次10秒。自印迹切出感兴趣蛋白质条带，并使用来自Applied Biosystems (Foster City, CA)的PROCISE®蛋白质测序仪(494型)上实施的Edman降解来测序。

II. 结果

SDS-PAGE分析证实在诱导后24小时(I24)和48小时(I48)可溶性和不溶性级分两者中Skp的预测分子量(约17kDa)处的蛋白质的显著积累(图2)。

N端测序分析证实I24时Skp蛋白预期大小的所诱导可溶性条带产生经加工形式的DsbC-Skp的预测蛋白质序列的前5个氨基酸(ADKIA, SEQ ID NO:27)。N端分析还显示I24时不溶性级分中所积累的两条条带产生经加工形式的和未加工形式的DsbC-Skp两者。较高的分子量条带产生未加工形式的DsbC-Skp的预测蛋白质序列的前10个氨基酸(MRLTQIIAAA, SEQ ID NO:28)，而较低的分子量条带产生经加工形式的DsbC-Skp蛋白质的预测蛋白质序列的前10个氨基酸(ADKIAIVNMG, SEQ ID NO:29)。参见图3A和图3B。

实施例2: pbp*分泌信号的鉴定

I. 材料和方法

A. 菌株

如下构建DC206 (*ΔpyrF, lsc::lacI^{Q1}*)，即使用掺入*lacI^{Q1}*启动子突变(Calos

等 1980)的引物自pCN51*lacI* (Schneider等 2005b) PCR扩增大肠杆菌*lacI*基因, 并通过等位基因交换将该基因重组入MB101 Δ *pyrF* (Schneider, Jenings等 2005b)的*lsc* (左聚糖蔗糖酶) 基因座中。

B. 构建转座体(transposome)以筛选荧光假单胞菌信号序列

通过将*kanR*基因(编码对卡那霉素的抗性)和*phoA*报告基因(其缺少起始密码子和N端信号序列)插入转座载体pMOD-2<MCS> (Epicentre Technologies)中的载体编码的转座酶结合位点(“镶嵌末端”)之间来改造转座载体。通过用*XhoI*进行的限制性消化来从pUC4-KIXX (Pharmacia)中纯化1.6kB *kanR*基因, 然后连接入经*SalI*消化的pMOD2<MCS>中以形成pDOW1245。自大肠杆菌K12 (ATCC) PCR扩增出缺少信号序列的*phoA*基因, 其中通过引物添加*BamHI*和*XbaI*位点。限制性消化后, 将基因连接入经*BamHI*和*XbaI*消化的pDOW1245中以制备pDOW1208。通过用*PshA1*限制性消化pDOW1208并使用Ultrafree DA (Amicon)对侧翼为镶嵌末端的3.3kb片段进行凝胶纯化来制备线性转座体。流过MicroBioSpin 6柱(Biorad)后, 将30ng与4个单位的转座酶(Epicentre)混合, 并将等分试样电穿孔入荧光假单胞菌MB101中。

C. 改良pbp信号序列的鉴定

将*pbp*-胰岛素原-*phoA*表达质粒设计成将*pbp*-胰岛素原蛋白融合至成熟的PhoA酶, 从而能测量胰岛素原-*phoA*在周质中的积累, 并能通过测定PhoA活性来选择积累改善的菌株。通过SOE PCR (Horton等 1990)使用与分泌前导和胰岛素原的编码序列及胰岛素原和成熟形式PhoA (即不含天然分泌前导)的编码序列交叠的引物在pINS-008中构建*pbp*信号序列、人胰岛素原、和*phoA*之间的融合物。在用*SpeI*和*XhoI*限制性消化并用虾碱性磷酸酶处理的pDOW1169 (Schneider等 2005a; Schneider, Jenings等 2005b)中在*tac*启动子的控制下克隆融合物, 然后连接, 并电穿孔入DC206中以形成pINS-008。为了在荧光假单胞菌中表达, 优化胰岛素原基因模板的密码子, 并合成(DNA 2.0)。自大肠杆菌MG1655基因组DNA扩增出*phoA*基因。在含有BCIP (碱性磷酸酶活性的一种比色指示剂)的琼脂平板上筛选菌落, 其中用IPTG来诱导*pbp*-胰岛素原-*phoA*基因的表达。在展现出BCIP水解的菌落中, 一个菌落比其它菌落生长得大得多。发现此分离群在编码分泌肽的区域中具有C至T突变, 引起第20位氨基酸从丙氨酸变成缬氨酸 (A20V, SEQ ID NO:2; 参见表

6)。

通过标准的摇瓶方案来评估这两种菌株的表达。添加IPTG诱导物后不久，两者的生长便变得平坦。突变型pINS-008-3菌株中的碱性磷酸酶活性高3-4倍（图6），并且（可溶性）蛋白质的积累较高（图7）。

表6：荧光假单胞菌中所鉴定的Sec分泌信号

安排的蛋白质功能	缩写	预测的信号序列 (SignalP-HMM)	SEQ ID NO:	DNA 序列	SEQ ID NO:
pbp (信号序列突变体)	pbp*	MKLKRLMAAMTFV AAGVATVNAVA	2	atgaaactgaaacgtttgatggcg gcaatgacttttgcgctgctggcg ttgcgaccgtcaacgcggtggcc	1
孔蛋白E1前体	PO	MKKSTLAVAVTLGA IAQQAGA	31	atgaagaagtcacacttggctgtg gctgtaacgttggcgcaatgcg ccagcaagcaggcgct	30
外膜孔蛋白F	OP	MKLKNTLGLAIGSLI AATSFGVLA	33	atgaaactgaaaaacaccttggg cttggccattggttcttattgccg ctacttcttgcgcttctggca	32
周质磷酸结合蛋白 (pbp)	PB	MKLKRLMAAMTFV AAGVATANAVA	35	atgaaactgaaacgtttgatggcg gcaatgacttttgcgctgctggcg ttgcgaccgccaacgcggtggcc	34
天青蛋白	AZ	MFAKLVAVSLTLA SGQLLA	37	atgtttgccaaactcgttgcgtttc cctgtgactctggcgagcgggcc agtgtcttgc	36
罕见的脂蛋白B前体	L	MIKRNLLVMGLAVL LSA	39	atgatcaaacgcaatctgctggtta tggcccttgcgctgctgttgagcg ct	38
赖氨酸-精氨酸-鸟氨酸结合蛋白	LAO	MQNYKKFLLAAAV SMAFSATAMA	41	atgcagaactataaaaattccttc tggccgcggcgctctcgatggcg ttcagcgccacggccatggca	40
铁(III)结合蛋白	IB	MIRDNRLKTSLLRG LTLTLLSLTLLSPAA HS	43	atgatccgtgacaaccgactcaa gacatcccttctgcgcggcctgac cctcaccctactcagctgaccct gctctgcccgcggccattct	42

D. 基因组测序

通过DNA Easy试剂盒(Invitrogen)来纯化基因组DNA，并将10μg作为模板用于测序，其使用2X ABI PRISM BigDye终止物v3.0循环试剂盒(Applied Biosystems)用转座子特异性引物进行。纯化反应物，并在ABI PRISM 3100遗传分析仪(Applied Biosystems)上依照制造商的指导进行加载。

E. 信号序列编码区的克隆

通过SPScan软件测定信号序列，或如(De等 1995)中那样测定。这些实验的结果已经披露于2004年11月22日提交的共同悬而未决的美国专利申请号20060008877。通过聚合酶链式反应(PCR)扩增外膜孔蛋白F (*oprF*)、磷酸结合蛋白(pbp)、孔蛋白E1 (*porE*)、天青蛋白、脂蛋白B和铁结合蛋白分泌前导。将所得的PCR产物克隆入pCRII Blunt TOPO载体中，并转化入大肠杆菌

Top10细胞(Invitrogen)中,其依照制造商的方案进行。通过用M13正向引物和M13反向引物测序来对所得转化体筛选正确的插入物。阳性克隆命名如下:
oprF (pDOW1112)、pbp (pDOW1113)、孔蛋白E1 (pDOW1183)、天青蛋白 (pDOW1180)、脂蛋白B (pDOW1182)、铁结合蛋白(pDOW1181)。

F. 用于在荧光假单胞菌中分泌的gal2 scFv克隆的构建。

使用pDOW1112或pDOW1113作为模板来扩增OprF和pbp信号序列以于+2位处融合至Gal2编码序列。使用pGal2 (Martineau等 1998)作为模板来扩增gal2编码序列。将837bp SOE-PCR产物克隆入pCR BLUNT II TOPO载体中,并证实序列。用XbaI和SalI限制酶自TOPO载体切割出与信号序列融合的scFv基因,并使用标准的克隆技术(Sambrook等 2001)克隆入pMYC1803的SpeI和XhoI位点中以产生pDOW1122 (oprF:gal2)和pDOW1123 (pbp:gal2)。将所得的质粒转化入在补充有30μg/mL四环素和50μg/mL卡那霉素的LB琼脂上选定的DC191中。

通过SOE-PCR将来自pDOW1183的porE信号序列融合至自pDOW1123扩增的gal2并通过凝胶提取来纯化PCR产物。将所得的PCR产物克隆入PCR II Blunt TOPO中,随后依照制造商的指示(Invitrogen)转化入大肠杆菌Top10细胞中。对所得的克隆测序,并选择阳性克隆(pDOW1185)进行亚克隆。用SpeI和SalI限制性消化pDOW1185,并对porE-gal2片段进行凝胶纯化。使用T4 DNA连接酶(NEB)将纯化的片段连接至经SpeI-SalI消化的pDOW1169。将连接混合物转化入电感受态DC454中,并在M9 1%葡萄糖琼脂平板上选择。通过使用SpeI和SalI限制性消化质粒DNA来筛选转化体。分离阳性克隆,并以pDOW1186贮存。

分别自克隆pDOW1180、pDOW1181和pDOW1182扩增出天青蛋白、铁结合蛋白和lipB的信号序列。使用合适的引物自pDOW1123扩增出gal2基因以融合至每种分泌前导,分离所得的PCR产物,并通过SOE-PCR来融合,如上文所述的。将SOE-PCR产物克隆入pCR-BLUNT II TOPO中,对所得的克隆测序,并将每种融合物的阳性克隆亚克隆入pDOW1169中,如上文所述的。

G. 有C端组氨酸标签的荧光假单胞菌分泌载体的构建

通过DNA 2.0 (pJ5:G03478)合成含有如下插入物的克隆,所述插入物具有pbp分泌前导、MCS及C端His标签和rrnT1T2转录终止子。通过用SpeI和NdeI进行的限制性消化来分离450bp分泌盒,并进行凝胶纯化。将该片段连

接至用相同酶消化的pDOW1219 (自pMYC1803 (Shao等 2006)衍生的)。将连接产物转化入化学感受态大肠杆菌JM109中。制备质粒DNA, 并通过使用载体特异性引物的PCR筛选插入物。所得的质粒进行序列证实, 并命名为pDOW3718。然后用该质粒转化电感受态荧光假单胞菌DC454, 并在补充有250 μ g/mL尿嘧啶和30 μ g/mL四环素的LB琼脂上选择。

使用来自人ORF组(ORFeome)集合(Invitrogen)的模板来扩增编码人蛋白质的可读框。用*NheI*和*XhoI*限制性消化PCR产物, 然后连接至经*NheI*-*XhoI*消化的pDOW3718。随后将连接产物转化入电感受态荧光假单胞菌DC454中, 并在补充有250 μ g/mL尿嘧啶和30 μ g/mL四环素的LB琼脂上选择转化体。对阳性克隆测序以证实插入物序列。

H. 大肠杆菌分泌克隆的构建

如上文那样扩增人ORF, 只是设计了如下引物, 其中5'引物上有*NcoI*位点且3'引物上有*XhoI*。用*NcoI*和*XhoI* (NEB)限制性消化PCR产物, 然后使用Qiaquick提取试剂盒(Qiagen)来纯化。使用T4 DNA连接酶(NEB)将经消化的产物连接至经*NcoI*-*XhoI*消化的pET22b (Novagen), 并将连接产物转化入化学感受态大肠杆菌Top10细胞中。在LB琼脂氨苄青霉素平板(Teknova)中选择转化体。制备质粒DNA (Qiagen), 并对阳性克隆测序以证实插入物序列。随后将每种的一个经证实的所克隆质粒转化入BL21(DE3) (Invitrogen)中进行表达分析。

I. DNA测序

使用G-50 (Sigma)来纯化测序反应(Big染料第3.1版(Applied Biosystems)), 并加载入ABI3100测序仪中。

J. 高通量(HTP)表达分析

使用标准的Dow HTP表达方案来分析荧光假单胞菌菌株。简言之, 使用在补充有1%葡萄糖和痕量元素的M9培养基中所培养的种子培养物来接种2.0mL 96孔深孔板中的0.5mL具有作为碳源的5%甘油的限定极限盐培养基。在30°C的起始生长阶段后, 用0.3mM异丙基- β -D-1-硫代吡喃半乳糖苷(IPTG)诱导经由*Ptac*启动子的表达。通过600nm的光密度(OD₆₀₀)来测量细胞密度。

K. 供SDS-PAGE分析用的HTP样品的制备

使用EASY LYSE™ (Epicentre Technologies产品目录编号RP03750)自培养物样品产生可溶性和不溶性级分。如下溶解25 μ L完全培养基样品, 即添加

175mL EASY LYSE™缓冲液，在温和摇动的情况中于室温温育30分钟。将溶胞物以14,000rpm离心20分钟(4°C)，并除去上清液。将上清液作为可溶性级分保存。然后将沉淀物（不溶性级分）重悬于等体积的溶解缓冲液中，并通过上下吹吸来重悬。对于选定的克隆，融化无细胞的培养基样品，并不进行稀释的情况中分析。

L. 感兴趣蛋白质或多肽的表达和分泌分析

使用在补充有1%葡萄糖(Teknova)、补充有痕量元素溶液的1X M9中所培养的种子培养物来以2%接种物接种50mL Dow限定极限盐培养基，并于30°C在摇动的情况中温育。用0.3mM IPTG（异丙基-β-D-1-硫代吡喃半乳糖苷）诱导细胞约24小时耗用发酵时间(EFT, elapsed fermentation time)。在诱导时(I0)和诱导后16小时(I16)、24小时(I24)、或40小时(I40)采集样品以进行分析。通过600nm的光密度(OD₆₀₀)来测量细胞密度。将细胞密度调节至OD₆₀₀ = 20，并将1mL以14000 x g离心5分钟。将上清液（无细胞培养基）移液入新的微量离心管中，然后将细胞沉淀物和无细胞培养基样品冷冻于-80°C，用于后面的加工。

M. SDS-PAGE分析

使用EASY LYSE™缓冲液(Epicentre Technologies)自摇瓶样品产生可溶性和不溶性级分。将冷冻的沉淀物重悬于1mL溶解缓冲液中。将50uL添加至另外的150uL EASY LYSE™缓冲液，并在摇动的情况中于室温温育30分钟。将溶胞物以14,000rpm离心20分钟(4°C)，并除去上清液。将上清液作为可溶性级分保存。然后将沉淀物重悬于等体积(200μL)的溶解缓冲液中，并通过上下吹吸来重悬；这作为不溶性级分保存。融化无细胞培养基样品，并足额使用。

N. Western分析

使用依照制造商的方案所制备的1X转移缓冲液(Invitrogen)，将所制备的并通过SDS-PAGE分开的可溶性和不溶性级分以100V 1小时转移至硝酸纤维素(BioRad)。转移后，用POLY-HRP稀释剂(Research Diagnostics, Inc.)封闭印迹，并用1:5,000稀释的抗His标签抗体(Sigma或US Biologicals)探查。用1X PBS-Tween清洗印迹，随后使用Immunopure金属增强DAB底物试剂盒(Pierce)来显现。

O. 20L发酵

如下产生每一种发酵罐培养物的接种物，即用冷冻的培养物原种来接种装有600mL补充有酵母提取物和甘油的化学限定培养基的摇瓶。在摇动的情况中于32°C温育16-24小时后，然后在无菌条件下将摇瓶培养物转移至20L装有设计用于支持高生物量的培养基的发酵罐。通过调节喷射气流和搅动速率来在液体培养物中将溶解氧维持在正水平(positive level)。经由添加氨水来将pH控制在想要的设定点。补料-分批高密度发酵过程分成约24小时的起始生长阶段和添加IPTG来启动重组基因表达的基因表达阶段。然后让发酵的表达阶段进行24小时。

P. N端氨基酸序列分析

如上文在SDS-PAGE分析中所描述的那样运行样品，并转移至Criterion Sequi-Blot PVDF膜(Biorad)。用GelCode Blue染色剂试剂(Pierce)对膜染色，随后用50%甲醇 1%乙酸脱色，用10%甲醇接着用去离子水漂洗，然后干燥。自膜切出感兴趣的条带，提取，并在Procise蛋白质测序系统，494型(Applied Biosystems, Foster City, CA)上进行8轮Edman降解。P. Edman, Acta Chem. Scand. 4, 283 (1950); 综述R. A. Laursen等, Methods Biochem. Anal. 26, 201-284 (1980)。

II. 结果

A. 通过转座子诱变鉴定天然分泌信号序列

为了鉴定会将异源蛋白质分泌至周质或培养基的荧光假单胞菌信号序列，将分泌报告基因克隆入转座体。所使用的分泌报告基因是不含起始密码子或N端信号序列的大肠杆菌碱性磷酸酶基因(*phoA*)。PhoA在周质（而非细胞溶胶）中是有活性的，这是由于容许二聚化成活性形式的分子内二硫键的形成(Derman等 1991)。使用一种称为“基因组扫描”的类似方法在大肠杆菌中寻找分泌蛋白质(Bailey等 2002)。*phoA*基因也已经用于分析大肠杆菌中(Manoil等 1985)和其它细菌中(Gicquel等 1996)周质的、膜的、和输出的蛋白质中的分泌信号。电穿孔和在指示培养基上铺板后，分离了8个蓝色菌落。对转座体在基因组中的插入位点测序，并用于搜索荧光假单胞菌MB101的私有基因组数据库。表6中显示了鉴定为能够表达活性PhoA的8种基因融合物。

B. 信号序列-gal2融合物的克隆

使用SignalP程序(J.D. Bendtsen 2004)来预测上文所鉴定的分泌蛋白质，外膜孔蛋白F (OP)、磷酸结合蛋白*porE* (PB)、铁结合蛋白(IB)、天青蛋白(AZ)、

脂蛋白B (L)和赖氨酸-鸟氨酸-精氨酸结合蛋白(LOA)的信号序列。OP、PE和AZ的信号序列先前已经在其它系统中鉴定[Arvidsson, 1989 #25; De, 1995 #24; Yamano, 1993 #23]。还平行分析了另一项研究中所鉴定的别的分泌前导pbpA20V (Schneider等 2006)的活性。在此研究中,通过材料和方法中所描述的重叠延伸PCR (SOE-PCR)使用剪接来将6种天然荧光假单胞菌信号序列和荧光假单胞菌磷酸结合蛋白信号序列的突变体的编码区(参见表6)各自融合至gal2 scFv基因,使得切割信号肽后Gal2的N端的4个氨基酸会是AQVQ。扩增LAO信号序列的反复尝试失败了,并且从进一步分析中去掉此信号序列。将基因融合物克隆入荧光假单胞菌表达载体pDOW1169中,并转化入DC454宿主菌株($\Delta pyrF\ lsc::lacIQ1$)中。随后对所得的菌株评估Gal2 scFv表达和分泌前导的正确加工。

C. 分泌的Gal2 scFv的表达

在摇瓶规模, PB、OP、PO、AZ、IB、和L与gal2 scFv的融合物达到预期的OD₆₀₀,只是L-gal2 scFv在传代培养入生产培养基中后未能生长(数据未显示)。Western印迹分析证实,自Gal2 scFv融合物切出PB、OP、PO、AZ和IB信号序列。然而, Western印迹显示未加工的PB-Gal2和OP-Gal2的存在。在无细胞培养基中找到自AZ和IB融合物表达的一些可溶性Gal2 scFv,这指明可溶性蛋白质表达了,并自周质间隙泄漏。实施了氨基端序列分析以证实对信号序列的切割。自天青蛋白(pDOW1191)融合物表达的不溶性Gal2蛋白显示具有经加工的和未加工的分泌信号的蛋白质的混合物。然而,观察到信号序列是自IB-Gal2融合物完全加工的。

使用标准发酵条件,在20L发酵规模评估与7种前导之每一种融合的Gal2 scFv的表达。所有菌株如预期的那样生长,18-24小时时达到诱导OD₆₀₀(约180个单位)。lipB-Gal2菌株比其它菌株生长得略微慢些。这不是完全出乎意料的,因为lipB-Gal2菌株在以小规模发酵接种摇瓶培养基后不生长。通过SDS-PAGE和Western印迹来评估Gal2 scFv的表达和加工。SDS-PAGE分析显示,在融合至OP或PB分泌信号时表达高水平的Gal2。然而,仅一部分(约50%)的OP-Gal2融合蛋白表现出分泌至周质,且信号序列被切除。如在小规模时所观察到的, Gal2主要以不溶性级分表达,虽然通过Western印迹检测出可溶性蛋白质。还在培养物上清液中检测出少量的蛋白质,指明自周质的泄漏(图7)。N端序列分析证实, ibp和天青蛋白前导得到如预期的那样的加

工, 导致N端氨基酸序列AQVQL (SEQ ID NO:44)。同样地, 根据Western分析, PorE分泌前导表现为得到加工, 并通过N端分析得到证实。不溶性PorE-Gal2的表达水平比不溶性ibp-Gal2和天青蛋白-Gal2的表达水平略低。LipB-Gal2以与PorE-Gal2的水平类似的水平显示经加工的Gal2的表达。自表达pbp-Gal2和pbpA20V-Gal2的菌株观察到最大量的蛋白质。自pbpA20V-Gal2菌株表达的Gal2量表现为甚至高于由pbp-Gal2菌株所生成的(图6)。通过Western分析检测出可溶性的、经加工的Gal2, 其是未加工的和经加工的不溶性蛋白质的混合物(图7)。不溶性蛋白质的N端序列分析证实未加工的和正确加工的Gal2的混合物。

实施例3: bce前导序列的鉴定

I. 材料和方法

BceL是一种分泌前导, 其鉴定为由含有来自凝结芽孢杆菌(*Bacillus coagulans*) CMC 104017的水解酶的基因的DNA插入物的一部分所编码。此菌株凝结芽孢杆菌在多个商业培养物保藏中心中也称为NCIMB 8041、ATCC 10545和DSMZ 2311, 并且其起源为NRS784。NRS 784来自形成孢子的细菌的NR Smith收集(Smith等 Aerobic spore forming bacteria US. Dep. Agr. Monogra. 16:1-148 (1952))。由NCIMB所引用的、关于此菌株的其它原始参考文献是Cambell, L.L.和Sniff E.E. (1959. J. Bacteriol. 78:267 An investigation of Folic acid requirements of *Bacillus coagulans*)。

序列和生物信息学分析

对来自凝结芽孢杆菌CMC 104017的4,127bp DNA插入物测序并分析, 以定位潜在编码水解酶的编码序列。在5'端的*lac*启动子后面鉴定出一种1,314bp的编码序列, 称为CDS1。CDS1的DNA和预测蛋白质序列分别在SEQ ID NO:45和46中列出。根据预测蛋白质序列的BLASTP分析, 确定CDS1最有可能编码水解酶。CDS1序列显示与来自沼泽红假单胞菌(*Rhodopseudomonas palustris*) HaA2的 β -内酰胺酶的同源性(E值: $2e-36$)。CDS1的SignalP 3.0隐藏Markov模型分析(Bendtsen JD, Nielson G, von Heijne G, Brunak S: Improved prediction of signal peptides: Signal 3.0. J. Mol. Biol 2004, 340:783.)预测生物体类别革兰氏阳性细菌的信号序列的存在, 且在SEQ ID NO:46的残基33/34间有信号肽酶切割位点。

蛋白质表达质粒的构建

在表达质粒的构建中使用标准的克隆方法(Sambrook J, Russell D: Molecular Cloning a Laboratory Manual, 第3版 Cold Spring Harbor: Cold Spring Harbor Press; 2001)。使用SOE-PCR方法来实施DNA序列融合(Horton, R. M., Z. Cai, S. N. Ho和L. R. Pease (1990). “Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction.” BioTechniques 8(5): 528-30, 532, 534-5))。对所有的PCR反应使用Phusion DNA聚合酶(New England Biolabs产品目录编号F531S)。

设计质粒以表达和定位进入荧光假单胞菌的细胞质或周质间隙中的来自凝结芽孢杆菌CMC104017的酯酶蛋白。用*SpeI*和*XhoI*限制性内切核酸酶(New England Biolabs产品目录编号R0133和R0146)消化最终的PCR产物，然后使用T4 DNA连接酶(New England Biolabs产品目录编号M0202S)连接入表达载体pDOW1169（也用*SpeI*和*XhoI*消化的）中以生成细胞质CMC104641 CDS-1表达载体 p484-001和天然 Bce 前导 CMC104641 CDS-1表达载体 p484-002。然后通过电穿孔将连接反应混合物转化入荧光假单胞菌菌株DC454 ($\Delta pyrF$, $lacI^{O1}$), 在含大豆的SOC培养基(Teknova产品目录编号2S2699)中恢复，并铺板于选择培养基（M9葡萄糖琼脂，Teknova产品目录编号2M1200）上。通过限制性消化微量制备的质粒DNA (Qiagen, 产品目录编号27106)来分析菌落。对来自每次转化的10个克隆测序以证实正确的插入物。

表达分析

在装有200mL具有作为碳源的5%甘油的限定极限盐培养基（“Dow培养基”）的摇瓶中检查携带每种克隆的荧光假单胞菌菌株DC454。起始生长阶段后, 用0.3mM异丙基- β -D-硫代吡喃半乳糖苷(IPTG)诱导经由*tac*启动子的表达。在诱导时(I0)、和诱导后24小时(I24)对培养物取样。通过600nm处的光密度(OD₆₀₀)来测量细胞密度。表7中显示了显示摇瓶编号方案的表格。

表7

宿主菌株	质粒号（前导-基因）	烧瓶号		
DC454	P484-001 (细胞质484)	EP484-001	EP484-002	EP484-003
DC454	P484-002 (天然前导484)	EP484-004	EP484-005	EP484-006

在每个取样时间，将样品的细胞密度调节至 $OD_{600} = 20$ ，并将1mL等分试样以 $14000 \times g$ 离心5分钟。将上清液（无细胞培养基）移液入新的微量离心管中，然后于 -20°C 冷冻细胞沉淀物和无细胞培养基样品。

细胞溶解和SDS-PAGE分析

使用Easy Lyse (Epicentre Technologies)自摇瓶样品产生可溶性和不溶性级分。重悬冷冻的沉淀物，在溶解缓冲液中以1:4溶解，并在摇动的情况中于室温温育30分钟。将溶胞物以 $14,000\text{rpm}$ 离心20分钟(4°C)，并除去上清液。将上清液作为可溶性级分保存。然后将沉淀物（不溶性级分）重悬于等体积的溶解缓冲液中，并通过上下吹吸来重悬。融化无细胞培养基样品，并足额使用。将样品与含有 β -巯基乙醇(BioRad产品目录编号161-0737)的2X Laemmli样品缓冲液以1:1混合，并煮沸5分钟，之后在Bio-Rad Criterion 10% Criterion XT凝胶(BioRad产品目录编号45-0112)上加载 $20\mu\text{L}$ ，并通过在推荐的1X MOPS缓冲液(产品目录编号161-0788 批号210001188)中电泳来分开。用SIMPLYBLUE™ SafeStain (Invitrogen产品目录编号LC6060)依照制造商的方案对凝胶染色，并使用Alpha Innotech成像系统进行成像。通过与加载于相同凝胶上的BSA蛋白质标准品比较来评估感兴趣的凝胶条带的蛋白质数量。

II. 结果

使用了总共6个摇瓶（每种菌株3个烧瓶）来评估水解酶表达。周质的和细胞质的设计菌株的生长与荧光假单胞菌菌株的正常生长相一致，在诱导后24小时达到 OD_{600} 约15。实施了SDS-PAGE分析以评估诱导时和诱导后24小时的水解酶（CDS1蛋白）表达。通过SDS-PAGE分析了可溶性级分、不溶性级分、和无细胞培养基级分。对于细胞质的CDS-1菌株(p484-001)，具有细胞质水解酶的预期大小(44.1kDa)的蛋白质在I24（IPTG诱导后24小时）时在所有3种分离群中以 0.1mg/mL 的估计产率几乎完全积累于可溶性级分中。图8显示了以EP484-003评估的细胞质菌株的代表性结果。在不溶性级分中能检测到具有预期大小的可忽略条带，而在无细胞培养基中没检测到CDS1蛋白。对于表达天然Bce前导-CDS1的周质菌株(p484-002)，具有天然酯酶预期大小的蛋白质在I24时在所有3种分离群中以 0.8mg/mL 的估计产率几乎完全积累于可溶性级分中。图8显示了以EP484-004评估的含有Bce前导融合物的细胞质菌株的代表性结果。不清楚所表达的天然酯酶是否是完全加工的，因为所使用的凝胶加载使其难以辨别预测的 47.6kDa 的未加工大小和 44.1kDa 的经加

工大小。与细胞质表达菌株的结果类似，在不溶性级分中能检测到具有预期大小的可忽略条带，而在无细胞培养基中没检测到CDS1蛋白。感兴趣Bce前导的翻译序列在SEQ ID NO:8中列出。

实施例4：荧光假单胞菌分泌前导的鉴定和分析

用信号肽预测程序SignalP 2.0 (Nielsen, H.等 Protein Eng, 1997. 10(1): 1-6)分析6,433种来自MB214基因组的翻译性ORF。通过HMM模型预测1326种含有信号肽。用PsortB 2.0 (Gardy, J.L.等 Bioinformatics, 2005. 21(5): 617-23)分析这些蛋白质，除去所有那些PsortB最终定位鉴定为细胞质的或细胞质膜的，留下891种。除去其中含有信号肽的SignalP HMM概率低于0.79的82种蛋白质，产生809种。选择0.79的截留，因为那是不排除aprA (RXF04304, 已知为胞外蛋白质)的最高数值。使用CLUSTALX 1.81 (Thompson, J.D.等 Nucleic Acids Res, 1997. 25(24): 4876-82)比对这809种翻译性ORF的氨基端序列，其含有通过SignalP神经网络算法所预测的信号肽加上经加工蛋白质的前7个氨基酸。

Huber等提示，高度疏水的信号序列更有可能在共翻译的情况中被分泌 (Huber, D.等 J Bacteriol, 2005. 187(9): 2983-91)。出于鉴定在共翻译的情况中所分泌的蛋白质的目的，发现Wertz-Scheraga (WS)的氨基酸指数(Wertz, D.H.和H.A. Scheraga, Macromolecules, 1978. 11(1): 9-15)是最好的。对于此研究，在万维网上于 www.genome.jp/dbget-bin/www_bget?aa1:WERD780101 自AAindex获得这些指数。改良由Boyd报告的算法(Boyd, D., C. Schierle和J. Beckwith, Protein Sci, 1998. 7(1): 201-5)，并用于根据疏水性对809种蛋白质分级。该算法扫描每个序列，在12的窗口内计算WS得分的平均值。使用最疏水的区域来为整个蛋白质指派WS得分。这产生WS得分大于0.69的142种信号序列，0.69是Huber等中限定的截留。这种较小的列表与来自由印地安那应用蛋白质科学中心(Indiana Centers for Applied Protein Sciences, INCAPS)实施的2D-LC全蛋白质组实验的数据相互参照。这些实验试图鉴定和量化多种生长条件下在MB214 (源于荧光假单胞菌MB101的)中所表达的所有蛋白质。在此列表中出现的具有高的最大限度表达水平的蛋白质是有可能高度表达的。在这些数据中，1或3的优先级(priority)得分指明蛋白质鉴定的高置信度。表8中以它们最大限度表达水平的顺序列出了来自INCAPS实验中所鉴定的142

种的列表的、优先级为1或3的的蛋白质。

表8: 以最大限度表达水平的顺序列出的、来自INCAPS实验过程中找到的142种的列表的、优先级为1或3（指明鉴定的高度置信）的7种独特蛋白质。

优先级	蛋白质ID	安排的功能	最大值
1	RXF05550.1	三十四肽重复家族蛋白	377264.2
1	RXF08124.1	接纳甲基的趋化蛋白	134887.4
1	RXF07256.1	TolB蛋白	88429.16
3	RXF07256.1_a1	TolB蛋白	84020.51
3	RXF04046.2_a1	细胞色素c氧化酶，单血红素亚基，膜结合的(ec 1.9.3.1)	79275.3
3	RXF03895.1_a1	asma	50164.08
3	RXF07256.1_pn	TolB蛋白	49215.09
1	RXF06792.1	保守的假设蛋白质	47485.35
3	RXF02291.1	甲苯耐受蛋白ttg2C	45703.08

已经鉴定出数种在大肠杆菌中在共翻译的情况中分泌的蛋白质。使用这些中的数种的序列来对MB214基因组搜寻同源物。大肠杆菌基因是：DsbA、TorT、SfmC、FocC、CcmH、YraI、TolB、NikA、FlgI。使用BLASTP算法(Altschul, S.F.等, J Mol Biol, 1990. 215(3): 403-10)来对MB214的数据库搜寻翻译性ORF。根据它们与它们的大肠杆菌对应物所显示的同源性程度，将MB214蛋白放入两种种类中。高同源性蛋白质与 $2e^{-84}$ 或更好的预期得分匹配。低同源性蛋白质具有 $8e^{-17}$ 和 $5e^{-32}$ 之间的预期得分。这种方法产生11种独特的潜在同源物，其中一些与上文所获得的7种靶物交叠。

使用SignalP来分析具有18种独特蛋白质的组合列表，并选择预测具有单一可能信号肽酶切割位点的9种最终的靶物进行表达研究。

分泌前导的分离和序列分析

自DC454（源自荧光假单胞菌MB101的）基因组DNA扩增所鉴定的荧光假单胞菌分泌前导，并克隆入pCRBLUNTII-TOPO (Invitrogen)中以进行DNA序列验证。表9中提及了所分离的每种荧光假单胞菌分泌前导的DNA和推导氨基酸序列。

表9: 荧光假单胞菌分泌前导序列

前导	DNA SEQ ID NO:	氨基酸SEQ ID NO:
----	----------------	---------------

CupA2	9	10
CupB2	11	12
CupC2	13	14
TolB	49	50
NikA	15	16
FlgI	17	18
ORF5550	19	20
Ttg2C	21	22
ORF8124	23	24

分泌前导与Gal2 scFv和大肠杆菌硫氧还蛋白的融合和表达分析

通过交叠延伸PCR (Horton R.M.等 1990 Biotechniques 8:528), 使用剪接, 将每种分泌前导 (表9) 以符合读码框的方式融合至Gal2 scFv序列 (Martineau, P.等 1998 J. Mol Bio. 280:117)和/或大肠杆菌硫氧还蛋白(TrxA)序列 (SEQ ID NO:46)。纯化所得的片段, 随后作为模板用于第二轮PCR以将NikA分泌前导编码序列融合至*trxA*序列。然后将融合物克隆入荧光假单胞菌表达载体pDOW1169中, 在*tac*启动子的控制下。将每种构建体转化入荧光假单胞菌DC454中, 并以高通量形式评估表达。将培养物在补充有5%甘油的限定矿物盐培养基中在2mL深孔板中以0.5mL的培养体积培养。24小时生长阶段后, 用0.3mM IPTG诱导重组蛋白, 并容许表达24小时。通过超声处理来使培养物分级, 并通过SDS-CGE和Western印迹来评估蛋白质表达和分泌前导加工 (图9)。发现所测试的每种前导 (除Bce前导之外) 是自Gal2 scFv蛋白质序列部分或完全加工的。与编码细胞质Gal2 scFv的表达菌株 (无) 相比, 每种也极大地改善Gal2 scFv的表达, 指明在指导亚细胞定位之外, 这些分泌前导还能改善总体表达。预料之中的是, 还观察到不同水平的Gal2 scFv表达和溶解。Western分析证实, 在融合至CupA2、CupC2、NikA、FlgI和ORF 5550时生成了一些可溶性Gal2 (图10)。虽然与Gal2融合的TolB前导的表达低于用其它前导所观察到的, 但是Western分析显示所表达的所有蛋白质是可溶的。N端分析显示TolB、CupA2、CupC2、FlgI、NikA和ORF5550前导自Gal2 scFv切割, 如预期的 (数据未显示)。

虽然Bce前导没有自Gal2 scFv加工, 但是发现其自TrxA加工。硫氧还蛋白已经被记载为用于鉴定共翻译分泌前导的模式蛋白质, 因为其在细胞质中

快速地折叠(Huber等 2005 J. Bacteriol. 187:2983)。利用Bce前导的可溶性TrxA的成功分泌可以指明此前导以共翻译方式起作用以促进周质分泌。

说明书中所提及的所有出版物和专利申请指明本发明所属技术领域中的技术人员的水平。本文通过提及而收录所有出版物和专利申请，其程度就像明确且单独指明通过提及而收录每篇单独的出版物或专利申请一样。

虽然出于理解清晰的目的，已经通过例示和实施例较为详细地描述了上述发明，但是显而易见的是，可以在所附权利要求书的范围内实施某些变化和改良。

<110> 陶氏环球技术公司 (DOW GLOBAL TECHNOLOGIES INC.)

<120> 用于提高表达的细菌前导序列

<130> 43292/339376

<160> 50

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 72

<212> DNA

<213> 人工序列

<220>

<223> 突变型磷酸结合蛋白 (mutant phosphate binding protein) (pbp*) 前导序列

<221> CDS

<222> (1)... (72)

<400> 1

```
atg aaa ctg aaa cgt ttg atg gcg gca atg act ttt gtc gct gct ggc 48
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
1 5 10 15
```

```
gtt gcg acc gtc aac gcg gtg gcc 72
Val Ala Thr Val Asn Ala Val Ala
20
```

<210> 2

<211> 24

<212> PRT

<213> 人工序列

<220>

<223> 突变型磷酸结合蛋白 (pbp*) 前导序列

<400> 2

```
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
1 5 10 15
```

```
Val Ala Thr Val Asn Ala Val Ala
20
```

<210> 3

<211> 66

<212> DNA

<213> 荧光假单胞菌 (*Pseudomonas fluorescens*)

<220>

<221> CDS

<222> (1)... (66)

<400> 3

```
atg cgt aat ctg atc ctc agc gcc gct ctc gtc act gcc agc ctc ttc 48
Met Arg Asn Leu Ile Leu Ser Ala Ala Leu Val Thr Ala Ser Leu Phe
1 5 10 15
```

```
ggc atg acc gca caa gct 66
Gly Met Thr Ala Gln Ala
20
```

<210> 4

<211> 22

<212> PRT

<213> 荧光假单胞菌

<400> 4
 Met Arg Asn Leu Ile Leu Ser Ala Ala Leu Val Thr Ala Ser Leu Phe
 1 5 10 15
 Gly Met Thr Ala Gln Ala
 20

<210> 5
 <211> 63
 <212> DNA
 <213> 荧光假单胞菌

<220>
 <221> CDS
 <222> (1)... (63)

<400> 5
 atg cgc ttg acc cag att att gcc gcc gca gcc att gcg ttg gtt tcc 48
 Met Arg Leu Thr Gln Ile Ile Ala Ala Ala Ala Ile Ala Leu Val Ser
 1 5 10 15
 acc ttt gcg ctg gcc 63
 Thr Phe Ala Leu Ala
 20

<210> 6
 <211> 21
 <212> PRT
 <213> 荧光假单胞菌

<400> 6
 Met Arg Leu Thr Gln Ile Ile Ala Ala Ala Ala Ile Ala Leu Val Ser
 1 5 10 15
 Thr Phe Ala Leu Ala
 20

<210> 7
 <211> 99
 <212> DNA
 <213> 凝结芽孢杆菌 (*Bacillus coagulans*)

<220>
 <221> CDS
 <222> (1)... (99)

<400> 7
 atg agc aca cga atc ccc cgc cga caa tgg ctg aaa ggc gcc tcg ggc 48
 Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1 5 10 15
 ctg ctg gcc gcc gcg agc ctg ggc cgg ttg gcc aac cgc gag gcg cgc 96
 Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
 20 25 30
 gcc 99
 Ala

<210> 8
 <211> 33
 <212> PRT
 <213> 凝结芽孢杆菌

<400> 8
 Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1 5 10 15
 Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
 20 25 30

Ala

<210> 9
 <211> 75
 <212> DNA
 <213> 荧光假单胞菌

<220>
 <221> CDS
 <222> (1)... (75)

<400> 9
 atg tcg tgc aca cgt gca ttc aaa cca ctg ctg ctg atc ggc ctg gcc 48
 Met Ser Cys Thr Arg Ala Phe Lys Pro Leu Leu Leu Ile Gly Leu Ala
 1 5 10 15
 aca ctg atg tgt tcc cat gca ttc gct 75
 Thr Leu Met Cys Ser His Ala Phe Ala
 20 25

<210> 10
 <211> 25
 <212> PRT
 <213> 荧光假单胞菌

<400> 10
 Met Ser Cys Thr Arg Ala Phe Lys Pro Leu Leu Leu Ile Gly Leu Ala
 1 5 10 15
 Thr Leu Met Cys Ser His Ala Phe Ala
 20 25

<210> 11
 <211> 72
 <212> DNA
 <213> 荧光假单胞菌

<220>
 <221> CDS
 <222> (1)... (72)

<400> 11
 atg ctt ttt cgc aca tta ctg gcg agc ctt acc ttt gct gtc atc gcc 48
 Met Leu Phe Arg Thr Leu Leu Ala Ser Leu Thr Phe Ala Val Ile Ala
 1 5 10 15
 ggc tta ccg tcc acg gcc cac gcg 72
 Gly Leu Pro Ser Thr Ala His Ala
 20

<210> 12
 <211> 24
 <212> PRT
 <213> 荧光假单胞菌

<400> 12
 Met Leu Phe Arg Thr Leu Leu Ala Ser Leu Thr Phe Ala Val Ile Ala
 1 5 10 15
 Gly Leu Pro Ser Thr Ala His Ala
 20

<210> 13
 <211> 69
 <212> DNA
 <213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (69)

<400> 13

```

atg ccg cct cgt tct atc gcc gca tgt ctg ggg ctg ctg ggc ttg ctc 48
Met Pro Pro Arg Ser Ile Ala Ala Cys Leu Gly Leu Leu Gly Leu Leu
1           5           10           15

```

```

atg gct acc cag gcc gcc gcc 69
Met Ala Thr Gln Ala Ala Ala
           20

```

<210> 14

<211> 23

<212> PRT

<213> 荧光假单胞菌

<400> 14

```

Met Pro Pro Arg Ser Ile Ala Ala Cys Leu Gly Leu Leu Gly Leu Leu
1           5           10           15
Met Ala Thr Gln Ala Ala Ala
           20

```

<210> 15

<211> 63

<212> DNA

<213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (63)

<400> 15

```

atg cgc ctc gct gcc cta ccg cta ttg ctt gcc cct ctc ttt att gcg 48
Met Arg Leu Ala Ala Leu Pro Leu Leu Leu Ala Pro Leu Phe Ile Ala
1           5           10           15

```

```

ccg atg gcc gtt gcg 63
Pro Met Ala Val Ala
           20

```

<210> 16

<211> 21

<212> PRT

<213> 荧光假单胞菌

<400> 16

```

Met Arg Leu Ala Ala Leu Pro Leu Leu Leu Ala Pro Leu Phe Ile Ala
1           5           10           15
Pro Met Ala Val Ala
           20

```

<210> 17

<211> 63

<212> DNA

<213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (63)

<400> 17

```

atg aag ttc aaa cag ctg atg gcg atg gcg ctt ttg ttg gcc ttg agc 48
Met Lys Phe Lys Gln Leu Met Ala Met Ala Leu Leu Leu Ala Leu Ser
1           5           10           15

```

```

gct gtg gcc cag gcc 63
Ala Val Ala Gln Ala

```


20

<210> 18
 <211> 21
 <212> PRT
 <213> 荧光假单胞菌

<400> 18
 Met Lys Phe Lys Gln Leu Met Ala Met Ala Leu Leu Leu Ala Leu Ser
 1 5 10 15
 Ala Val Ala Gln Ala
 20

<210> 19
 <211> 63
 <212> DNA
 <213> 荧光假单胞菌

<220>
 <221> CDS
 <222> (1)... (63)

<400> 19
 atg aat aga tct tcc gcg ttg ctc ctc gct ttt gtc ttc ctc agc ggc 48
 Met Asn Arg Ser Ser Ala Leu Leu Leu Ala Phe Val Phe Leu Ser Gly
 1 5 10 15
 tgc cag gcc atg gcc 63
 Cys Gln Ala Met Ala
 20

<210> 20
 <211> 21
 <212> PRT
 <213> 荧光假单胞菌

<400> 20
 Met Asn Arg Ser Ser Ala Leu Leu Leu Ala Phe Val Phe Leu Ser Gly
 1 5 10 15
 Cys Gln Ala Met Ala
 20

<210> 21
 <211> 99
 <212> DNA
 <213> 荧光假单胞菌

<220>
 <221> CDS
 <222> (1)... (99)

<400> 21
 atg caa aac cgc act gtg gaa atc ggt gtc ggc ctt ttc ttg ctg gct 48
 Met Gln Asn Arg Thr Val Glu Ile Gly Val Gly Leu Phe Leu Leu Ala
 1 5 10 15
 ggc atc ctg gct tta ctg ttg ttg gcc ctg cga gtc agc ggc ctt tcg 96
 Gly Ile Leu Ala Leu Leu Leu Leu Ala Leu Arg Val Ser Gly Leu Ser
 20 25 30
 gcc 99
 Ala

<210> 22
 <211> 33
 <212> PRT

<213> 荧光假单胞菌

<400> 22

Met Gln Asn Arg Thr Val Glu Ile Gly Val Gly Leu Phe Leu Leu Ala
 1 5 10 15
 Gly Ile Leu Ala Leu Leu Leu Leu Ala Leu Arg Val Ser Gly Leu Ser
 20 25 30
 Ala

<210> 23

<211> 117

<212> DNA

<213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (117)

<400> 23

atg tct ctt cgt aat atg aat atc gcc ccg agg gcc ttc ctc ggc ttc 48
 Met Ser Leu Arg Asn Met Asn Ile Ala Pro Arg Ala Phe Leu Gly Phe
 1 5 10 15
 gcg ttt att ggc gcc ttg atg ttg ttg ctc ggt gtg ttc gcg ctg aac 96
 Ala Phe Ile Gly Ala Leu Met Leu Leu Leu Gly Val Phe Ala Leu Asn
 20 25 30
 cag atg agc aaa att cgt gcg 117
 Gln Met Ser Lys Ile Arg Ala
 35

<210> 24

<211> 39

<212> PRT

<213> 荧光假单胞菌

<400> 24

Met Ser Leu Arg Asn Met Asn Ile Ala Pro Arg Ala Phe Leu Gly Phe
 1 5 10 15
 Ala Phe Ile Gly Ala Leu Met Leu Leu Leu Gly Val Phe Ala Leu Asn
 20 25 30
 Gln Met Ser Lys Ile Arg Ala
 35

<210> 25

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 寡核苷酸引物

<400> 25

aattactagt aggaggtaca ttatgcgctt 30

<210> 26

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 寡核苷酸引物

<400> 26

tatactcgag ttatttaacc tgtttcagta 30

<210> 27

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 经加工形式的 dsbC-Skp 的预测蛋白质序列的前 5 个氨基酸

<400> 27

Ala Asp Lys Ile Ala
1 5

<210> 28

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 未经加工形式的 dsbC-Skp 的预测蛋白质序列的前 10 个氨基酸

<400> 28

Met Arg Leu Thr Gln Ile Ile Ala Ala Ala
1 5 10

<210> 29

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 经加工形式的 dsbC-Skp 的预测蛋白质序列的前 10 个氨基酸

<400> 29

Ala Asp Lys Ile Ala Ile Val Asn Met Gly
1 5 10

<210> 30

<211> 63

<212> DNA

<213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (63)

<400> 30

atg aag aag tcc acc ttg gct gtg gct gta acg ttg ggc gca atc gcc 48
Met Lys Lys Ser Thr Leu Ala Val Ala Val Thr Leu Gly Ala Ile Ala
1 5 10 15cag caa gca ggc gct 63
Gln Gln Ala Gly Ala
20

<210> 31

<211> 21

<212> PRT

<213> 荧光假单胞菌

<400> 31

Met Lys Lys Ser Thr Leu Ala Val Ala Val Thr Leu Gly Ala Ile Ala
1 5 10 15
Gln Gln Ala Gly Ala
20

<210> 32

<211> 72

<212> DNA
 <213> 荧光假单胞菌

 <220>
 <221> CDS
 <222> (1)... (72)

 <400> 32
 atg aaa ctg aaa aac acc ttg ggc ttg gcc att ggt tct ctt att gcc 48
 Met Lys Leu Lys Asn Thr Leu Gly Leu Ala Ile Gly Ser Leu Ile Ala
 1 5 10 15

 gct act tct ttc ggc gtt ctg gca 72
 Ala Thr Ser Phe Gly Val Leu Ala
 20

<210> 33
 <211> 24
 <212> PRT
 <213> 荧光假单胞菌

 <400> 33
 Met Lys Leu Lys Asn Thr Leu Gly Leu Ala Ile Gly Ser Leu Ile Ala
 1 5 10 15
 Ala Thr Ser Phe Gly Val Leu Ala
 20

<210> 34
 <211> 72
 <212> DNA
 <213> 荧光假单胞菌

 <220>
 <221> CDS
 <222> (1)... (72)

 <400> 34
 atg aaa ctg aaa cgt ttg atg gcg gca atg act ttt gtc gct gct gcc 48
 Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
 1 5 10 15

 gtt gcg acc gcc aac gcg gtg gcc 72
 Val Ala Thr Ala Asn Ala Val Ala
 20

<210> 35
 <211> 24
 <212> PRT
 <213> 荧光假单胞菌

 <400> 35
 Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
 1 5 10 15
 Val Ala Thr Ala Asn Ala Val Ala
 20

<210> 36
 <211> 60
 <212> DNA
 <213> 荧光假单胞菌

 <220>
 <221> CDS
 <222> (1)... (60)

 <400> 36
 atg ttt gcc aaa ctc gtt gct gtt tcc ctg ctg act ctg gcg agc gcc 48
 Met Phe Ala Lys Leu Val Ala Val Ser Leu Leu Thr Leu Ala Ser Gly

<210>	41
<211>	23
<212>	PRT

<213> 荧光假单胞菌

<400> 41

```
Met Gln Asn Tyr Lys Lys Phe Leu Leu Ala Ala Ala Val Ser Met Ala
 1             5             10             15
Phe Ser Ala Thr Ala Met Ala
                20
```

<210> 42

<211> 93

<212> DNA

<213> 荧光假单胞菌

<220>

<221> CDS

<222> (1)... (93)

<400> 42

```
atg atc cgt gac aac cga ctc aag aca tcc ctt ctg cgc ggc ctg acc 48
Met Ile Arg Asp Asn Arg Leu Lys Thr Ser Leu Leu Arg Gly Leu Thr
 1             5             10             15

ctc acc cta ctc agc ctg acc ctg ctc tcg ccc gcg gcc cat tct 93
Leu Thr Leu Leu Ser Leu Thr Leu Leu Ser Pro Ala Ala His Ser
                20             25             30
```

<210> 43

<211> 31

<212> PRT

<213> 荧光假单胞菌

<400> 43

```
Met Ile Arg Asp Asn Arg Leu Lys Thr Ser Leu Leu Arg Gly Leu Thr
 1             5             10             15
Leu Thr Leu Leu Ser Leu Thr Leu Leu Ser Pro Ala Ala His Ser
                20             25             30
```

<210> 44

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 经加工的天青蛋白和 ibp 的 N-端氨基酸序列

<400> 44

```
Ala Gln Val Gln Leu
 1             5
```

<210> 45

<211> 1317

<212> DNA

<213> 凝结芽孢杆菌

<220>

<221> CDS

<222> (1)... (1317)

<400> 45

```
atg agc aca cga atc ccc cgc cga caa tgg ctg aaa ggc gcc tcg ggc 48
Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1             5             10             15

ctg ctg gcc gcc gcg agc ctg ggc cgg ttg gcc aac cgc gag gcg cgc 96
Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
                20             25             30
```

gcc gcc gaa gcg agc gcc gcc gcg ccg ctc gac act ggc tcg ctg ggc Ala Ala Glu Ala Ser Ala Ala Ala Pro Leu Asp Thr Gly Ser Leu Gly 35 40 45	144
gcc tcg ccg cgc gcg acg ctc gac gcc tgc ctg caa aaa gcc gtc gac Ala Ser Pro Arg Ala Thr Leu Asp Ala Cys Leu Gln Lys Ala Val Asp 50 55 60	192
gac ggc acg ctc aag agc gtg gtg gcg atg gcc gcc acc gag cgc ggc Asp Gly Thr Leu Lys Ser Val Val Ala Met Ala Ala Thr Glu Arg Gly 65 70 75 80	240
ctc gcc tac cag ggc gcg cgc ggc ccg gcc aac gcg gcc ggc gag ccg Leu Ala Tyr Gln Gly Ala Arg Gly Pro Ala Asn Ala Ala Gly Glu Pro 85 90 95	288
atc ggc ccc gat acg gtg ttc tgg atg ctg tcg atg acc aag gcg atc Ile Gly Pro Asp Thr Val Phe Trp Met Leu Ser Met Thr Lys Ala Ile 100 105 110	336
acc gcc acc gcc tgc atg cag ctg atc gag cag ggc cgg ctc ggc ctc Thr Ala Thr Ala Cys Met Gln Leu Ile Glu Gln Gly Arg Leu Gly Leu 115 120 125	384
gac cag ccc gcc gcc gag atc ctg ccg caa ctg aag gcg ccg cag gtg Asp Gln Pro Ala Ala Glu Ile Leu Pro Gln Leu Lys Ala Pro Gln Val 130 135 140	432
ctg gag ggc ttc gac gcc gcc ggc cag ccc agg ctg cgc ccg gcg cgc Leu Glu Gly Phe Asp Ala Ala Gly Gln Pro Arg Leu Arg Pro Ala Arg 145 150 155 160	480
cgc gcg atc acg gtg cgc cac ctg ctc acg cat acc tcg ggc tat acc Arg Ala Ile Thr Val Arg His Leu Leu Thr His Thr Ser Gly Tyr Thr 165 170 175	528
tac agc atc tgg agc gag gcg ctg ggc cgc tac gaa cag gtc acg ggc Tyr Ser Ile Trp Ser Glu Ala Leu Gly Arg Tyr Glu Gln Val Thr Gly 180 185 190	576
atg ccc gac atc ggc tac tcg ctg aac ggc gcc ttc gcg gcc ccg ctc Met Pro Asp Ile Gly Tyr Ser Leu Asn Gly Ala Phe Ala Ala Pro Leu 195 200 205	624
gaa ttc gag ccc ggc gag cgc tgg caa tac ggc atc ggc atg gat tgg Glu Phe Glu Pro Gly Glu Arg Trp Gln Tyr Gly Ile Gly Met Asp Trp 210 215 220	672
gtg ggc aag ctg gtg gag gcg gtg acc gac cag tcg ctg gaa gtg gcg Val Gly Lys Leu Val Glu Ala Val Thr Asp Gln Ser Leu Glu Val Ala 225 230 235 240	720
ttc cgc gag cgg atc ttc gcg ccg ctc ggc atg cac gat acg ggc ttc Phe Arg Glu Arg Ile Phe Ala Pro Leu Gly Met His Asp Thr Gly Phe 245 250 255	768
ctg atc ggc agc gcg caa aag cgc cgc gtc gcc acg ctg cat cgg cgc Leu Ile Gly Ser Ala Gln Lys Arg Arg Val Ala Thr Leu His Arg Arg 260 265 270	816
cag gcc gat ggc tcg ctg acg ccg gaa ccc ttc gag acc aac cag cgg Gln Ala Asp Gly Ser Leu Thr Pro Glu Pro Phe Glu Thr Asn Gln Arg 275 280 285	864
ccc gag ttc ttc atg ggc ggc ggc ggg ctg ttc agc acc ccg cgc gac Pro Glu Phe Phe Met Gly Gly Gly Gly Leu Phe Ser Thr Pro Arg Asp 290 295 300	912
tac ctc gcc ttc ctg cag atg ctg ctg aac ggc ggc gcc tgg cgc ggc Tyr Leu Ala Phe Leu Gln Met Leu Leu Asn Gly Gly Ala Trp Arg Gly 305 310 315 320	960

```

gag cgg ctg ctg cgg ccc gac acc gtg gcg agc atg ttc cgc aac cag 1008
Glu Arg Leu Leu Arg Pro Asp Thr Val Ala Ser Met Phe Arg Asn Gln
325 330 335

atc ggc gat ctt cag gtt cgc gaa atg aag acc gcc cag ccg gcc tgg 1056
Ile Gly Asp Leu Gln Val Arg Glu Met Lys Thr Ala Gln Pro Ala Trp
340 345 350

tcg aac agc ttc gac caa ttc ccc ggc gcg acg cac aag tgg ggg ctg 1104
Ser Asn Ser Phe Asp Gln Phe Pro Gly Ala Thr His Lys Trp Gly Leu
355 360 365

tcc ttc gat ctc aac agc gag ccg ggg ccg cac ggg cgc ggc gcc ggc 1152
Ser Phe Asp Leu Asn Ser Glu Pro Gly Pro His Gly Arg Gly Ala Gly
370 375 380

tcg ggt agc tgg gcc ggc ctg ctg aac acc tac ttc tgg atc gat ccc 1200
Ser Gly Ser Trp Ala Gly Leu Leu Asn Thr Tyr Phe Trp Ile Asp Pro
385 390 395

gcc aag cgc gtg acg ggg gcg ctg ttc acg cag atg ctg ccg ttc tac 1248
Ala Lys Arg Val Thr Gly Ala Leu Phe Thr Gln Met Leu Pro Phe Tyr
405 410 415

gac gcg cgc gtg gtc gat ctc tac ggg cgc ttc gag cgc ggg ctc tac 1296
Asp Ala Arg Val Val Asp Leu Tyr Gly Arg Phe Glu Arg Gly Leu Tyr
420 425 430

gac ggg ctg ggc cgc gcc tga 1317
Asp Gly Leu Gly Arg Ala *
435

```

<210> 46
 <211> 438
 <212> PRT
 <213> 凝结芽孢杆菌

```

<400> 46
Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
1 5 10 15
Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
20 25 30
Ala Ala Glu Ala Ser Ala Ala Ala Pro Leu Asp Thr Gly Ser Leu Gly
35 40 45
Ala Ser Pro Arg Ala Thr Leu Asp Ala Cys Leu Gln Lys Ala Val Asp
50 55 60
Asp Gly Thr Leu Lys Ser Val Val Ala Met Ala Ala Thr Glu Arg Gly
65 70 75 80
Leu Ala Tyr Gln Gly Ala Arg Gly Pro Ala Asn Ala Ala Gly Glu Pro
85 90 95
Ile Gly Pro Asp Thr Val Phe Trp Met Leu Ser Met Thr Lys Ala Ile
100 105 110
Thr Ala Thr Ala Cys Met Gln Leu Ile Glu Gln Gly Arg Leu Gly Leu
115 120 125
Asp Gln Pro Ala Ala Glu Ile Leu Pro Gln Leu Lys Ala Pro Gln Val
130 135 140
Leu Glu Gly Phe Asp Ala Ala Gly Gln Pro Arg Leu Arg Pro Ala Arg
145 150 155 160
Arg Ala Ile Thr Val Arg His Leu Leu Thr His Thr Ser Gly Tyr Thr
165 170 175
Tyr Ser Ile Trp Ser Glu Ala Leu Gly Arg Tyr Glu Gln Val Thr Gly
180 185 190
Met Pro Asp Ile Gly Tyr Ser Leu Asn Gly Ala Phe Ala Ala Pro Leu
195 200 205
Glu Phe Glu Pro Gly Glu Arg Trp Gln Tyr Gly Ile Gly Met Asp Trp
210 215 220
Val Gly Lys Leu Val Glu Ala Val Thr Asp Gln Ser Leu Glu Val Ala
225 230 235 240
Phe Arg Glu Arg Ile Phe Ala Pro Leu Gly Met His Asp Thr Gly Phe
245 250 255
Leu Ile Gly Ser Ala Gln Lys Arg Arg Val Ala Thr Leu His Arg Arg

```



```

      260      265      270
Gln Ala Asp Gly Ser Leu Thr Pro Glu Pro Phe Glu Thr Asn Gln Arg
      275      280      285
Pro Glu Phe Phe Met Gly Gly Gly Leu Phe Ser Thr Pro Arg Asp
      290      295      300
Tyr Leu Ala Phe Leu Gln Met Leu Leu Asn Gly Gly Ala Trp Arg Gly
305      310      315      320
Glu Arg Leu Leu Arg Pro Asp Thr Val Ala Ser Met Phe Arg Asn Gln
      325      330      335
Ile Gly Asp Leu Gln Val Arg Glu Met Lys Thr Ala Gln Pro Ala Trp
      340      345      350
Ser Asn Ser Phe Asp Gln Phe Pro Gly Ala Thr His Lys Trp Gly Leu
      355      360      365
Ser Phe Asp Leu Asn Ser Glu Pro Gly Pro His Gly Arg Gly Ala Gly
      370      375      380
Ser Gly Ser Trp Ala Gly Leu Leu Asn Thr Tyr Phe Trp Ile Asp Pro
385      390      395      400
Ala Lys Arg Val Thr Gly Ala Leu Phe Thr Gln Met Leu Pro Phe Tyr
      405      410      415
Asp Ala Arg Val Val Asp Leu Tyr Gly Arg Phe Glu Arg Gly Leu Tyr
      420      425      430
Asp Gly Leu Gly Arg Ala
      435

```

<210> 47
 <211> 324
 <212> DNA
 <213> 大肠杆菌 (Escherichia coli)

<220>
 <221> CDS
 <222> (1)... (324)

```

<400> 47
agc gat aaa att att cac ctg act gac gac agt ttt gac acg gat gta 48
Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val
  1          5          10          15

ctc aaa gcg gac ggg gcg atc ctc gtc gat ttc tgg gca gag tgg tgc 96
Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys
      20          25          30

ggt ccg tgc aaa atg atc gcc ccg att ctg gat gaa atc gct gac gaa 144
Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu
      35          40          45

tat cag ggc aaa ctg acc gtt gca aaa ctg aac atc gat caa aac cct 192
Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro
      50          55          60

ggc act gcg ccg aaa tat ggc atc cgt ggt atc ccg act ctg ctg ctg 240
Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu
      65          70          75          80

ttc aaa aac ggt gaa gtg gcg gca acc aaa gtg ggt gca ctg tct aaa 288
Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys
      85          90          95

ggt cag ttg aaa gag ttc ctc gac gct aac ctg gcg 324
Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala
      100          105

```

<210> 48
 <211> 108
 <212> PRT
 <213> 大肠杆菌

```

<400> 48
Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val
  1          5          10          15

```

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

<210> 49
<211> 63
<212> DNA
<213> 荧光假单胞菌

<220>
<221> CDS
<222> (1)... (63)

<400> 49
atg aga aac ctt ctt cga gga atg ctt gtc gtt att tgc tgt atg gca 48
Met Arg Asn Leu Leu Arg Gly Met Leu Val Val Ile Cys Cys Met Ala
1 5 10 15

ggg ata gcg gcg gcg 63
Gly Ile Ala Ala Ala
20

<210> 50
<211> 21
<212> PRT
<213> 荧光假单胞菌

<400> 50
Met Arg Asn Leu Leu Arg Gly Met Leu Val Val Ile Cys Cys Met Ala
1 5 10 15
Gly Ile Ala Ala Ala
20

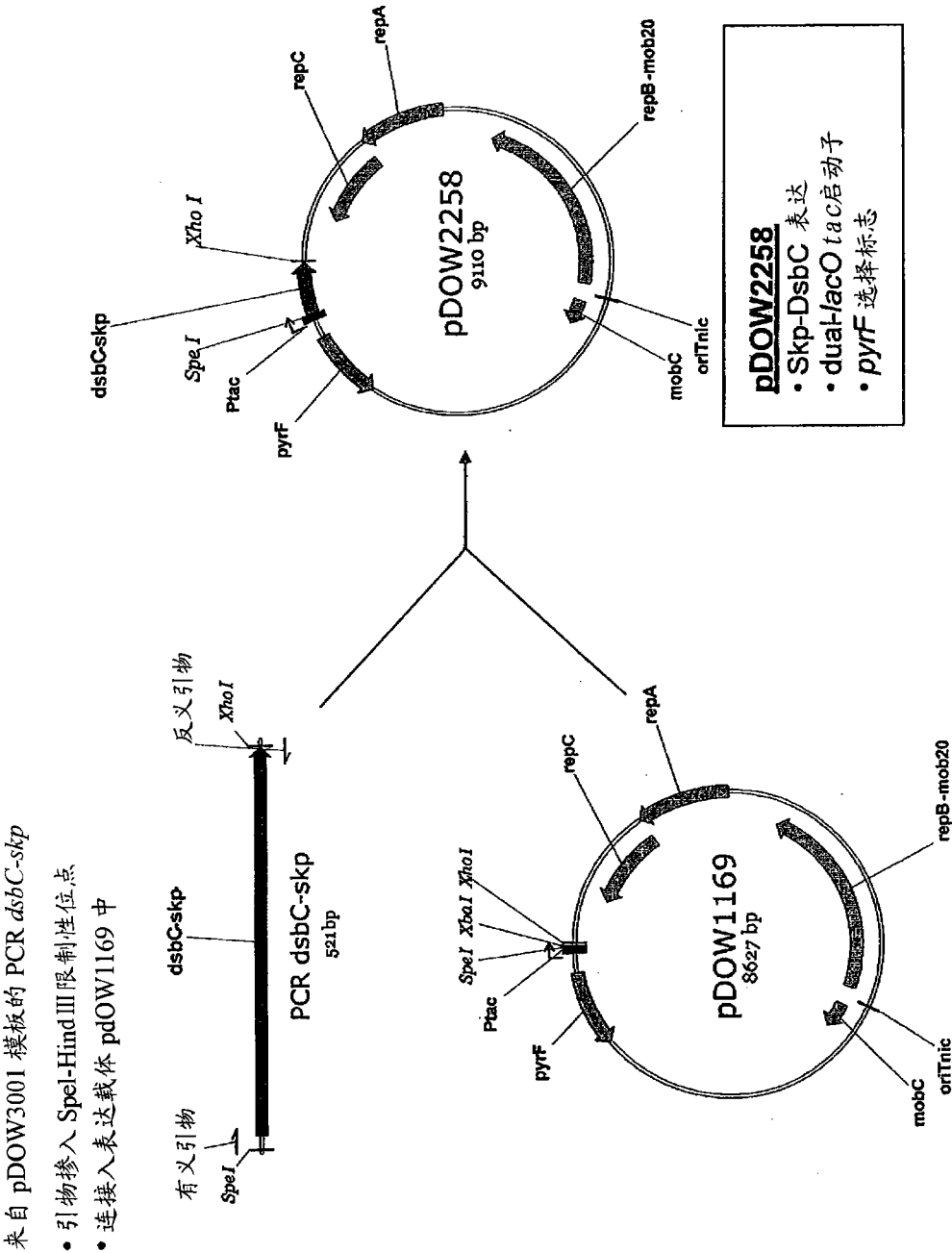


图 1

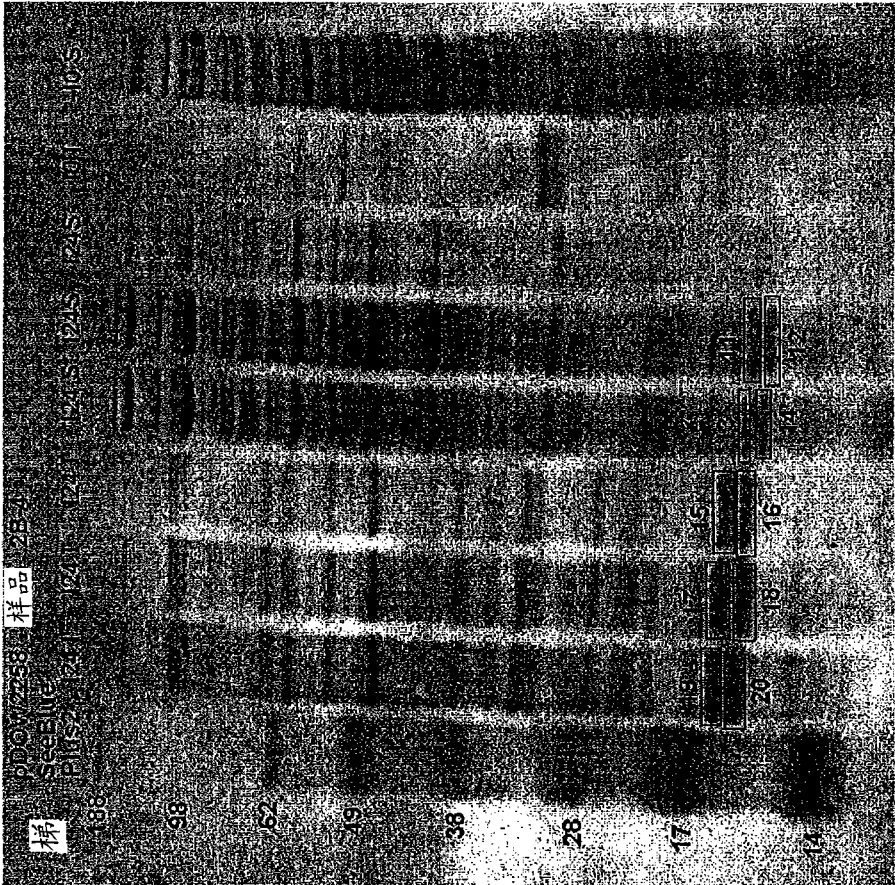


图 3 B

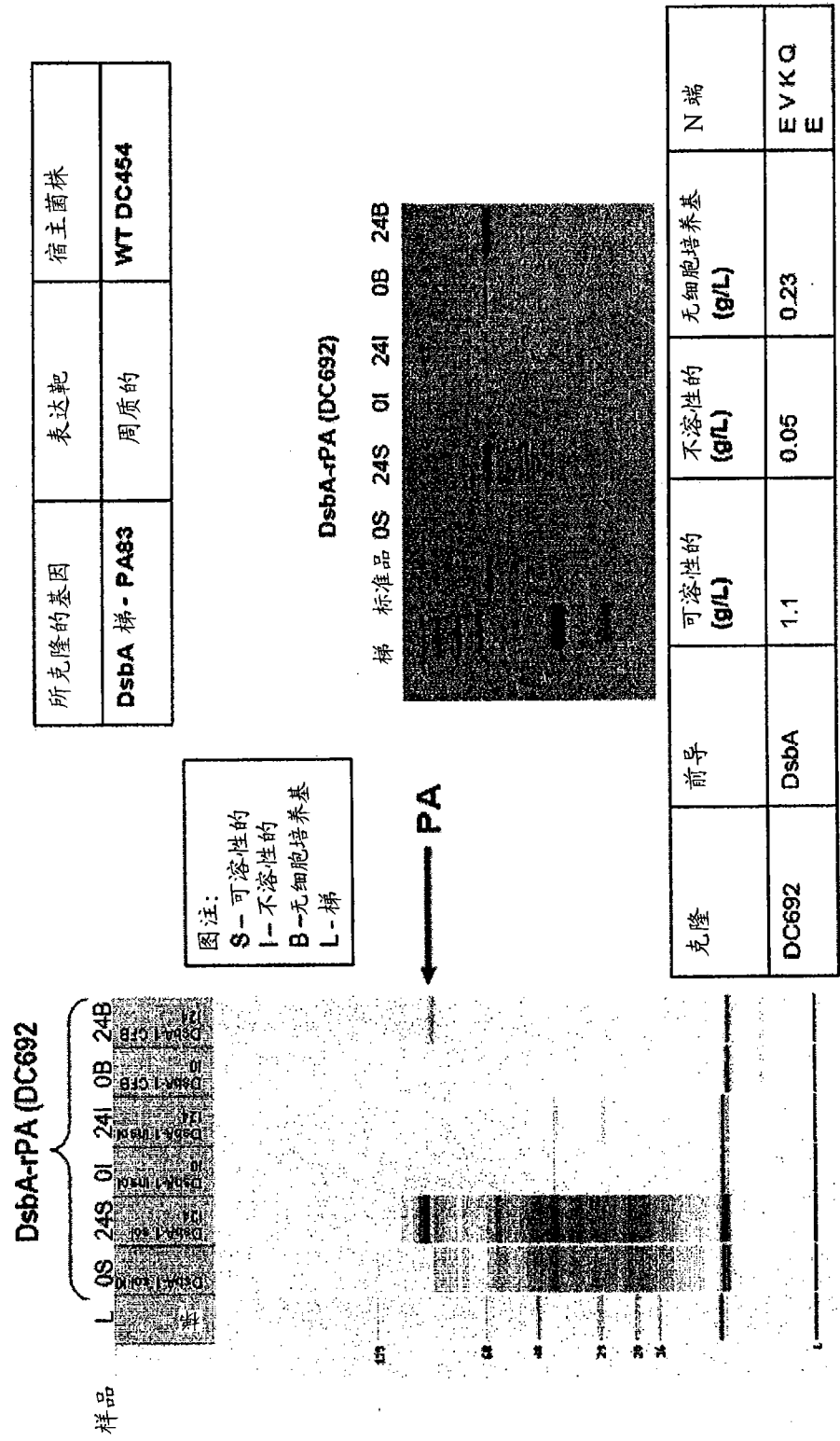


图 4

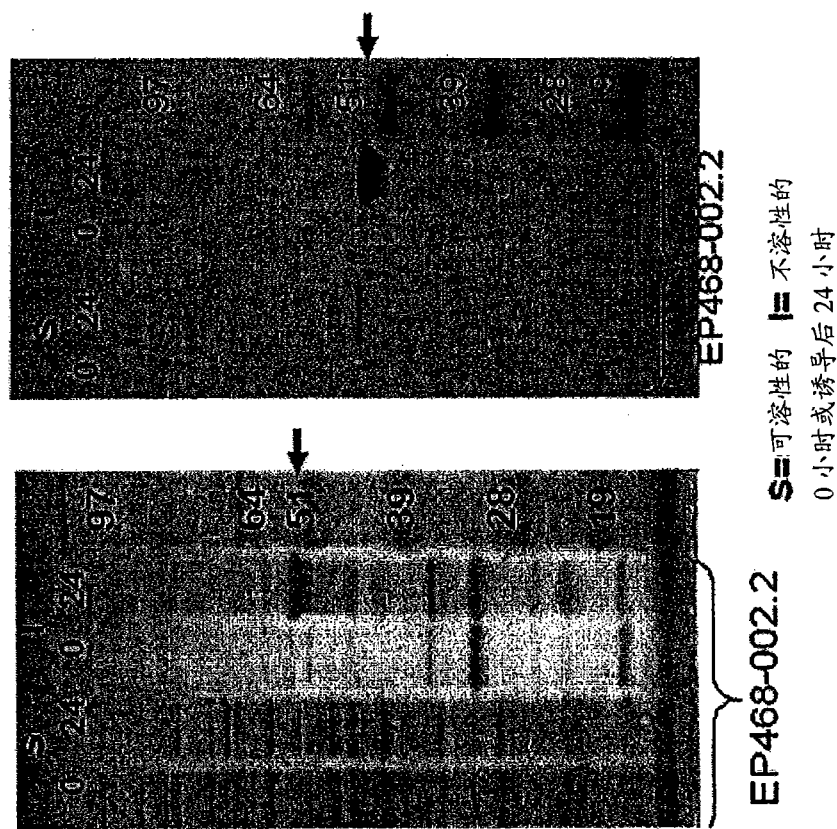


图 5

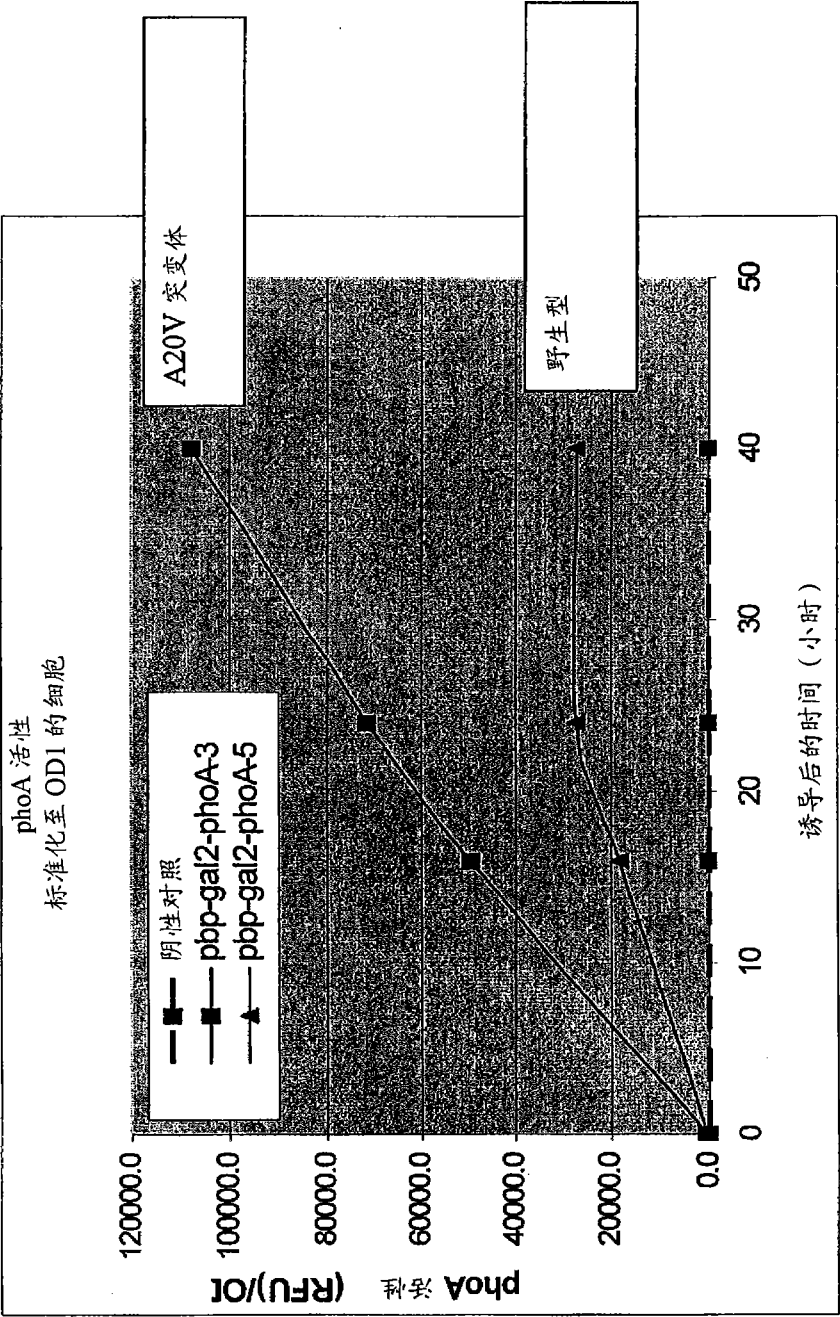


图 6

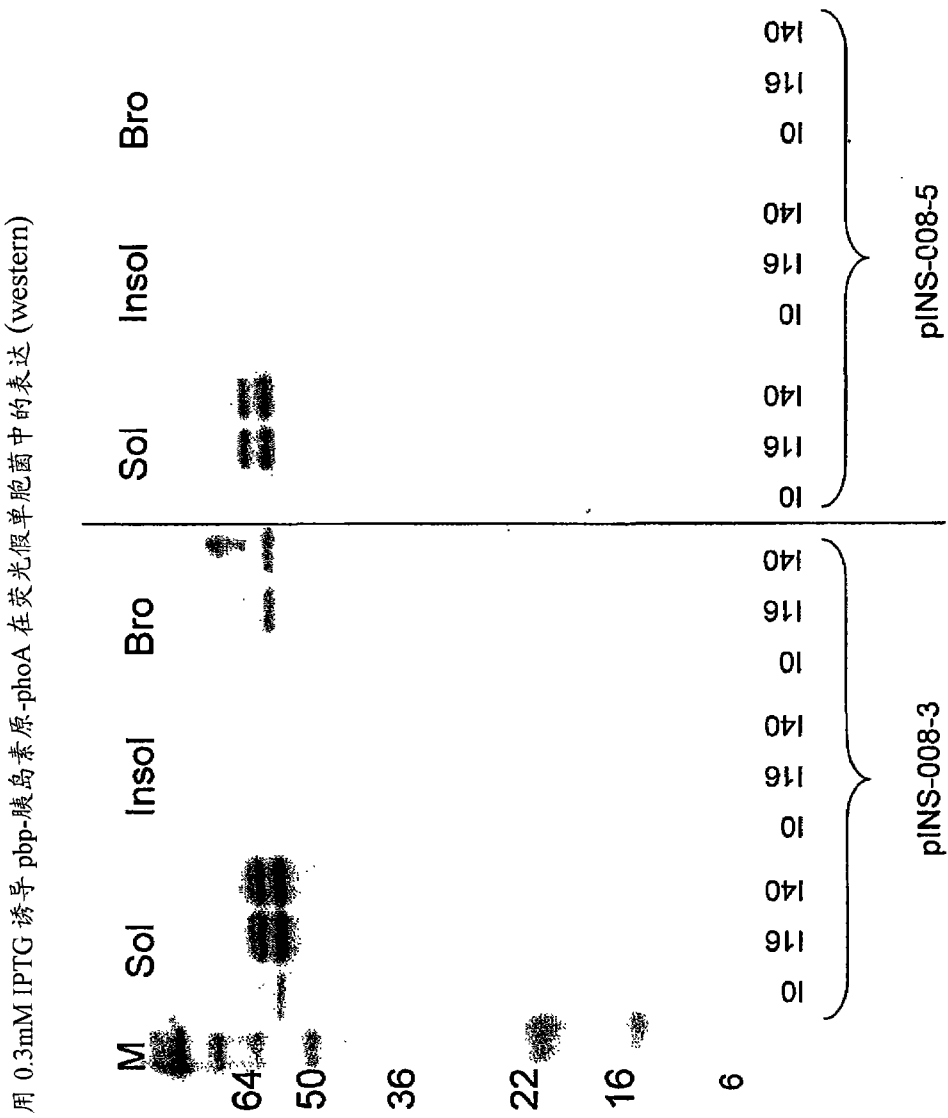


图 7

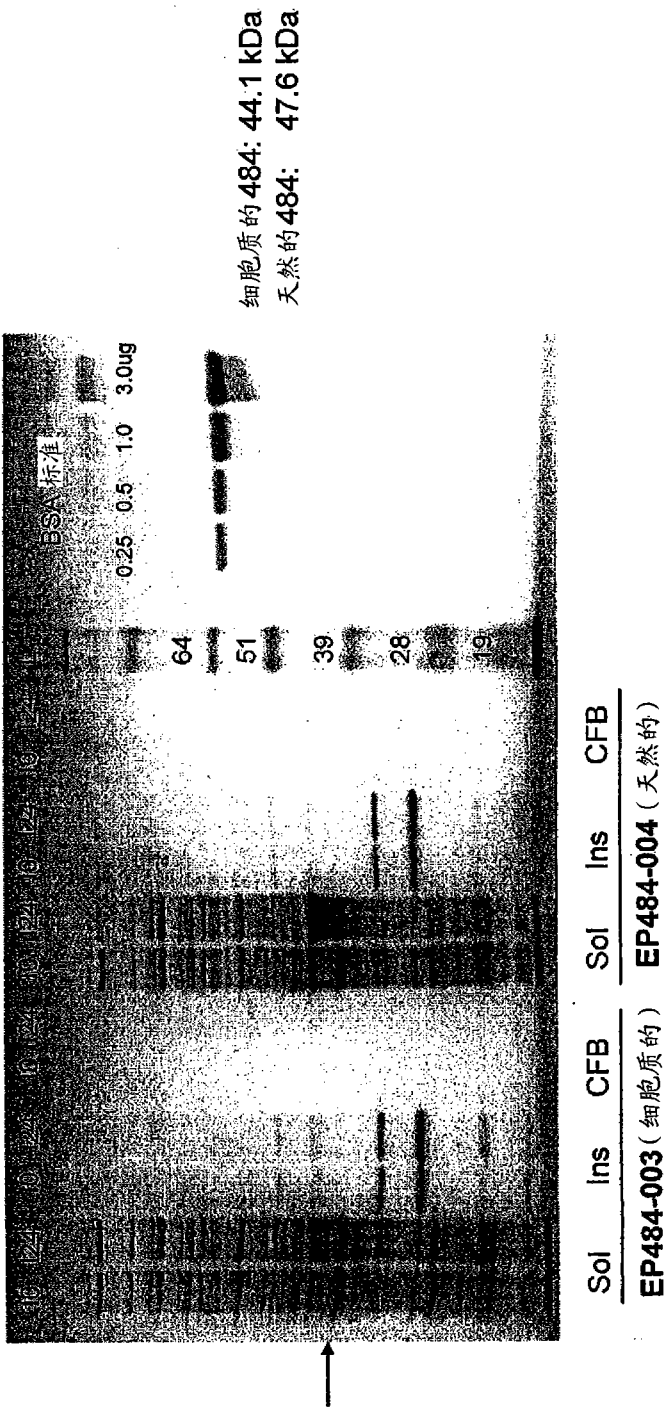


图 8

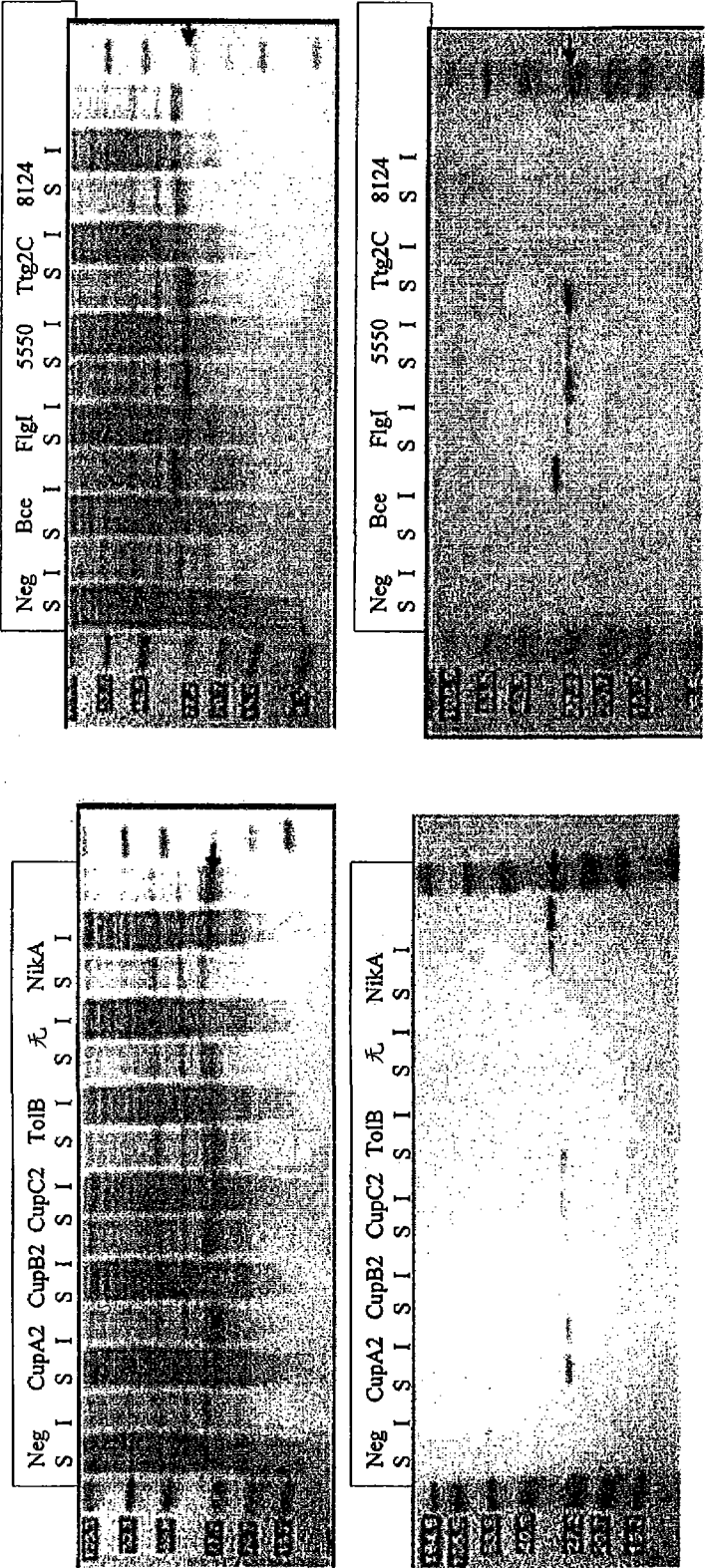


图9

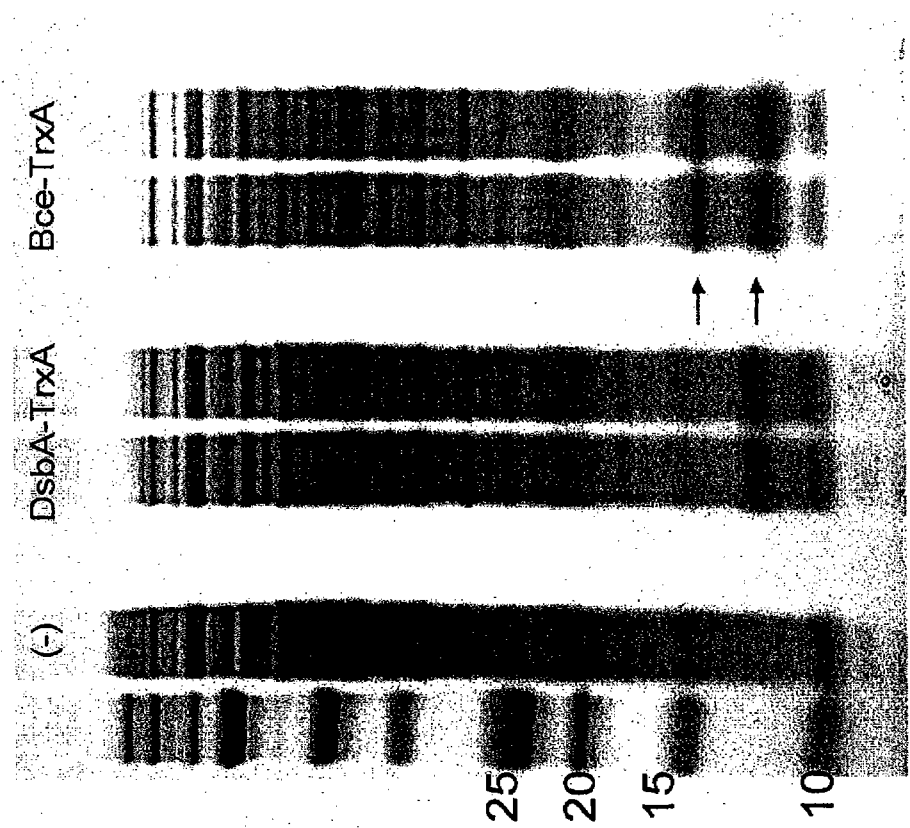


图 10