



(12) 发明专利

(10) 授权公告号 CN 106068273 B

(45) 授权公告日 2021. 01. 08

(21) 申请号 201480076332.6  
(22) 申请日 2014.12.08  
(65) 同一申请的已公布的文献号  
    申请公布号 CN 106068273 A  
(43) 申请公布日 2016.11.02  
(30) 优先权数据  
    61/922,613 2013.12.31 US  
(85) PCT国际申请进入国家阶段日  
    2016.08.25  
(86) PCT国际申请的申请数据  
    PCT/US2014/069019 2014.12.08  
(87) PCT国际申请的公布数据  
    W02015/102814 EN 2015.07.09  
(73) 专利权人 丹尼斯科美国公司  
    地址 美国加利福尼亚州  
(72) 发明人 C·邦焦尔尼 B·F·施密特  
    R·奥楚卡 A·范齐蒙耐德

(74) 专利代理机构 北京市中咨律师事务所  
    11247  
    代理人 柴云峰 黄草生  
(51) Int.Cl.  
    C07K 14/32 (2006.01)  
    C12N 15/75 (2006.01)  
(56) 对比文件  
    WO 2008066931 A8,2009.07.16  
    ANA AMOROSO ET AL.A peptidoglycan  
fragment triggers[beta]-lactam resistance  
in bacillus licheniformis.《PLOS  
PATHOGENS》.2012,e1002571.  
    ANA AMOROSO ET AL.A peptidoglycan  
fragment triggers[beta]-lactam resistance  
in bacillus licheniformis.《PLOS  
PATHOGENS》.2012,e1002571.  
    Dervyn E et.al.A vector for  
systematic gene inactivation in Bacillus  
subtilis.《Microbiology》.1998,3097-3104页.  
    审查员 黄蕊

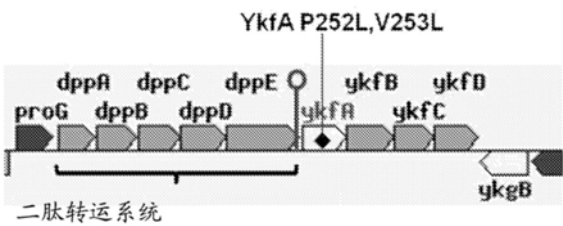
权利要求书2页 说明书27页  
序列表8页 附图5页

(54) 发明名称

增强的蛋白质表达

(57) 摘要

本发明一般涉及具有导致感兴趣的蛋白质的表达提高的基因改变的细菌细胞以及制备和使用该细胞的方法。本发明的方面包括革兰氏阳性微生物如芽孢杆菌菌种,其具有改变ykf操纵子编码的蛋白质的活性并导致感兴趣的蛋白质的表达增强的基因改变。



1. 一种用于使来自芽孢杆菌属细菌细胞的感兴趣的蛋白质的表达提高的方法,所述方法包括:

a) 获取能够产生感兴趣的蛋白质的改变的芽孢杆菌属细菌细胞,其中所述改变的细胞包含编码枯草芽孢杆菌的YkfA蛋白质的基因的基因改变,其中所述基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L的取代或与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L的取代;以及

b) 在使得所述改变的芽孢杆菌属细菌细胞表达所述感兴趣的蛋白质的条件下培养所述改变的芽孢杆菌属细菌细胞,

其中与在相同的培养条件下生长的相应未改变的芽孢杆菌属细菌细胞中所述感兴趣的蛋白质的表达相比,所述感兴趣的蛋白质在所述改变的芽孢杆菌属细菌细胞中的表达提高。

2. 根据权利要求1的方法,其中YkfA蛋白质由SEQ ID NO:2的氨基酸序列组成。

3. 根据权利要求1或2所述的方法,其中所述编码枯草芽孢杆菌的YkfA蛋白质的基因由SEQ ID NO:1的核苷酸序列组成。

4. 根据权利要求1或2所述的方法,其中所述基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L的取代以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L的取代。

5. 根据权利要求1或2所述的方法,其中所述感兴趣的蛋白质是酶。

6. 根据权利要求1或2所述的方法,其中所述感兴趣的蛋白质是蛋白酶。

7. 根据权利要求1或2所述的方法,所述方法还包括回收所述感兴趣的蛋白质。

8. 一种改变的芽孢杆菌属细菌细胞,其中所述改变的芽孢杆菌属细菌细胞包含编码枯草芽孢杆菌的YkfA蛋白质的基因的基因改变,其中所述基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L的取代或与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L的取代,其中所述改变的芽孢杆菌属细胞,与在相同的培养条件下生长的未改变的芽孢杆菌属细胞相比,产生增加量的感兴趣的内源或异源蛋白质。

9. 根据权利要求8的改变的细胞,其中所述YkfA蛋白质由SEQ ID NO:2的氨基酸序列组成。

10. 根据权利要求8或9所述的改变的细胞,其中所述感兴趣的蛋白质是酶。

11. 根据权利要求8或9所述的改变的细胞,其中所述感兴趣的蛋白质是蛋白酶。

12. 一种包含来源于枯草芽孢杆菌的ykfA基因的变体序列的多核苷酸,其中所述变体序列编码蛋白质,所述蛋白质与枯草芽孢杆菌的YkfA蛋白质在下述氨基酸取代上不同:SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L的取代或与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L的取代。

13. 权利要求12的多核苷酸,其中所述枯草芽孢杆菌的YkfA蛋白质由SEQ ID NO:2的氨基酸序列组成。

14. 权利要求12或13的多核苷酸,其中所述变体序列长度为至少960个核苷酸,并且与由SEQ ID NO:1组成的ykfA基因在至少一个基因改变处不同,所述基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L的取代或与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L的取代。

15. 根据权利要求14所述的多核苷酸,其中所述基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L。

16. 一种包含权利要求12至15中任一项所述的多核苷酸序列的载体。

17. 根据权利要求16所述的载体,其中所述载体是靶向载体,所述靶向载体被设计成在转化到所述芽孢杆菌属细菌细胞中时通过同源重组将所述多核苷酸序列中的至少一个突变引入到所述芽孢杆菌属细菌细胞的ykf操纵子中的对应位置。

18. 一种用于增强芽孢杆菌属细菌细胞中感兴趣的蛋白质表达的方法,所述方法包括:

- a) 用权利要求17所述的载体转化亲本芽孢杆菌属细菌细胞;
- b) 允许所述载体与所述亲本芽孢杆菌属细菌细胞的ykf操纵子中相应区域的同源重组,以产生改变的芽孢杆菌属细菌细胞;以及
- c) 使所述改变的芽孢杆菌属细菌细胞在适于表达所述感兴趣的蛋白质的条件下生长,其中与所述转化步骤之前的所述芽孢杆菌属细菌细胞相比,所述感兴趣的蛋白质在所述改变的芽孢杆菌属细菌细胞中的产量提高。

## 增强的蛋白质表达

[0001] 相关申请的交叉引用

[0002] 本申请要求于2013年12月31日提交的美国临时专利申请USSN 61/922,613的优先权的权益,该申请的全部内容以引用方式并入本文。

### 技术领域

[0003] 本发明一般涉及具有导致感兴趣的蛋白质的表达提高的基因改变的细菌细胞以及制备和使用该细胞的方法。本发明的方面包括革兰氏阳性微生物如芽孢杆菌 (*Bacillus*) 菌种,其具有改变ykf操纵子编码的蛋白质的活性并导致感兴趣的蛋白质的表达增强的基因改变。

### 背景技术

[0004] 基因工程允许改良微生物以用作工业生物反应器、细胞工厂并用于食品发酵。包括多个芽孢杆菌菌种的革兰氏阳性生物用于产生大量有用的蛋白质和代谢物(参见例如 Zukowski, "Production of commercially valuable products," In: Doi和McGlouglin(编辑) *Biology of Bacilli: Applications to Industry*, Butterworth-Heinemann, Stoneham, Mass第311-337页[1992])。用于工业的常见芽孢杆菌菌种包括地衣芽孢杆菌 (*B. licheniformis*)、解淀粉芽孢杆菌 (*B. amyloliquefaciens*) 和枯草芽孢杆菌 (*B. subtilis*)。因为其GRAS(一般被认为是安全的)状态,所以这些芽孢杆菌菌种的菌株是产生用于食品和制药工业的蛋白质的天然候选。革兰氏阳性生物所产生的蛋白质的示例包括酶例如 $\alpha$ -淀粉酶、中性蛋白酶和碱性(或丝氨酸)蛋白酶。

[0005] 尽管对于在细菌宿主细胞中产生蛋白质的理解已有进展,但仍需要开发表达水平提高的感兴趣的蛋白质的新重组菌株。

### 发明内容

[0006] 本发明提供表达水平提高的感兴趣的蛋白质的重组革兰氏阳性细胞以及制备和使用其的方法。具体地,本发明涉及具有这样一种基因改变的细菌细胞,其与不具有基因改变的细菌细胞相比导致感兴趣的蛋白质的表达提高。本发明的方面包括革兰氏阳性微生物如芽孢杆菌菌种,其具有改变ykf操纵子编码的一个或多个蛋白质的活性并导致感兴趣的蛋白质的表达增强的基因改变。还提供了制备和使用该重组细菌细胞的方法。

[0007] 本发明的方面包括一种用于使来自革兰氏阳性细菌细胞的感兴趣的蛋白质的表达提高的方法,该方法包括:a) 获取能够产生感兴趣的蛋白质的改变的革兰氏阳性细菌细胞,其中所述改变的革兰氏阳性细菌细胞包含至少一个改变ykf操纵子所编码的一个或多个蛋白质的活性的基因改变。以及b) 在使得所述改变的革兰氏阳性细菌细胞表达所述感兴趣的蛋白质的条件下培养所述改变的革兰氏阳性细菌细胞,其中与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中所述感兴趣的蛋白质的表达相比,所述感兴趣的蛋白质在所述改变的革兰氏阳性细菌细胞中的表达提高。

[0008] 在某些实施方案中,改变的革兰氏阳性细菌细胞为芽孢杆菌属菌种的菌株(例如芽孢杆菌属菌种的菌株选自:地衣芽孢杆菌、迟缓芽孢杆菌(*B.lentus*)、枯草芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌(*B.brevis*)、嗜热脂肪芽孢杆菌(*B.stearothermophilus*)、嗜碱芽孢杆菌(*B.alkalophilus*)、凝结芽孢杆菌(*B.coagulans*)、环状芽孢杆菌(*B.circulans*)、短小芽孢杆菌(*B.pumilus*)、灿烂芽孢杆菌(*B.lautus*)、克劳氏芽孢杆菌(*B.clausii*)、巨大芽孢杆菌(*B.megaterium*)和苏云金芽孢杆菌(*B.thuringiensis*))。在某些实施方案中,芽孢杆菌属菌种的菌株为枯草芽孢杆菌菌株。在某些实施方案中,改变的革兰氏阳性细菌细胞进一步在选自以下的基因中包含突变:degU、degQ、degS、scoC4、spoIIE和oppA。在某些实施方案中,突变为degU (Hy) 32。

[0009] 在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有降低的YkfA蛋白质活性。在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有提高的YkfA蛋白质活性。

[0010] 在某些实施方案中,基因改变在所述ykf操纵子的ykfA基因中。在一些实施方案中,基因改变在ykf操纵子的内源性ykfA基因中。在某些实施方案中,ykfA基因与SEQ ID NO:1有至少60%同一性。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。

[0011] 在某些实施方案中,感兴趣的蛋白质是同源蛋白质。在某些实施方案中,感兴趣的蛋白质是异源蛋白质。在某些实施方案中,感兴趣的蛋白质是酶。在某些实施方案中,该酶选自:蛋白酶、纤维素酶、支链淀粉酶、淀粉酶、糖酶、脂肪酶、异构酶、转移酶、激酶和磷酸酶。在某些实施方案中,感兴趣的蛋白质是蛋白酶。在某些实施方案中,蛋白酶为枯草杆菌蛋白酶。在某些实施方案中,枯草杆菌蛋白酶选自:枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN'、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147、枯草杆菌蛋白酶309、以及它们的变体。

[0012] 在某些实施方案中,该方法还包括回收所述感兴趣的蛋白质。

[0013] 本发明的方面包括改变的革兰氏阳性细菌细胞,其中与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞相比,所述改变的革兰氏阳性细菌细胞包含至少一个基因改变,所述基因改变使ykf操纵子所编码的一个或多个蛋白质的活性改变。在某些实施方案中,改变的革兰氏阳性细菌细胞是芽孢杆菌属菌种的菌株。在某些实施方案中,芽孢杆菌属菌种的菌株选自:地衣芽孢杆菌、迟缓芽孢杆菌、枯草芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、灿烂芽孢杆菌、克劳氏芽孢杆菌、巨大芽孢杆菌和苏云金芽孢杆菌。在某些实施方案中,

芽孢杆菌属菌种的菌株为枯草芽孢杆菌菌株。在某些实施方案中,改变的革兰氏阳性细菌细胞进一步在选自以下的基因中包含突变:degU、degQ、degS、scoC4、spoIIE和oppA。在某些实施方案中,突变为degU (Hy) 32。

[0014] 在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有降低的YkfA蛋白质活性。在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有提高的YkfA蛋白质活性。

[0015] 在某些实施方案中,基因改变在所述ykf操纵子的ykfA基因中。在一些实施方案中,基因改变在ykf操纵子的内源性ykfA基因中。在某些实施方案中,ykfA基因与SEQ ID NO:1有至少60%同一性。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。在某些实施方案中,改变的细胞表达感兴趣的蛋白质。在某些实施方案中,感兴趣的蛋白质是同源蛋白质。在某些实施方案中,感兴趣的蛋白质是异源蛋白质。在某些实施方案中,感兴趣的蛋白质是酶。

[0016] 在某些实施方案中,该酶选自:蛋白酶、纤维素酶、支链淀粉酶、淀粉酶、糖酶、脂肪酶、异构酶、转移酶、激酶和磷酸酶。在某些实施方案中,感兴趣的蛋白质是蛋白酶。在某些实施方案中,蛋白酶为枯草杆菌蛋白酶。在某些实施方案中,枯草杆菌蛋白酶选自:枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN'、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147、枯草杆菌蛋白酶309、以及它们的变体。

[0017] 本发明的方面包括用于获取具有改善的蛋白质产生能力的改变的革兰氏阳性细菌细胞的方法,该方法包括将至少一个基因改变引入到亲本革兰氏阳性细菌细胞中,从而改变ykf操纵子所编码的一个或多个蛋白质的活性。在某些实施方案中,改变的革兰氏阳性细菌细胞是芽孢杆菌属菌种的菌株。在某些实施方案中,芽孢杆菌属菌种的菌株选自:地衣芽孢杆菌、迟缓芽孢杆菌、枯草芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、灿烂芽孢杆菌、克劳氏芽孢杆菌、巨大芽孢杆菌和苏云金芽孢杆菌。在某些实施方案中,芽孢杆菌属菌种的菌株为枯草芽孢杆菌菌株。在某些实施方案中,改变的革兰氏阳性细菌细胞进一步在选自以下的基因中包含突变:degU、degQ、degS、scoC4、spoIIE和oppA。在某些实施方案中,突变为degU (Hy) 32。

[0018] 在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有降低的YkfA蛋白质活性。在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有提高的YkfA蛋白质活性。

[0019] 在某些实施方案中,基因改变在所述ykf操纵子的ykfA基因中。在一些实施方案

中,基因改变在ykf操纵子的内源性ykfA基因中。在某些实施方案中,ykfA基因与SEQ ID NO:1有至少60%同一性。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。

[0020] 在某些实施方案中,所述改变的革兰氏阳性细菌细胞表达感兴趣的蛋白质。在某些实施方案中,所述方法还包括将编码所述感兴趣的蛋白质的表达盒引入到所述亲本革兰氏阳性细菌细胞中。在某些实施方案中,所述方法还包括将编码所述感兴趣的蛋白质的表达盒引入到所述改变的革兰氏阳性细菌细胞中。在某些实施方案中,感兴趣的蛋白质是同源蛋白质。在某些实施方案中,感兴趣的蛋白质是异源蛋白质。在某些实施方案中,感兴趣的蛋白质是酶。在某些实施方案中,该酶选自:蛋白酶、纤维素酶、支链淀粉酶、淀粉酶、糖酶、脂肪酶、异构酶、转移酶、激酶和磷酸酶。在某些实施方案中,感兴趣的蛋白质是蛋白酶。在某些实施方案中,蛋白酶为枯草杆菌蛋白酶。在某些实施方案中,枯草杆菌蛋白酶选自:枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN<sup>′</sup>、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147、枯草杆菌蛋白酶309、以及它们的变体。

[0021] 在某些实施方案中,所述方法还包括在使得所述改变的革兰氏阳性细菌细胞表达所述感兴趣的蛋白质的条件下培养所述改变的革兰氏阳性细菌细胞。在某些实施方案中,该方法还包括回收所述感兴趣的蛋白质。

[0022] 本发明的方面包括由上述方法产生的改变的革兰氏阳性细菌细胞。

[0023] 本发明的方面包括一种包含来源于ykfA基因的变体序列的多核苷酸,其中所述变体序列:

[0024] 长度为至少15个核苷酸,

[0025] 与全部或部分SEQ ID NO:1有至少60%同一性,并且

[0026] 包含至少一个在ykfA基因中的核苷酸位置处的基因改变,当所述至少一个突变存在于革兰氏阳性细菌细胞的内源性ykfA基因中时,所述基因改变导致YkfA蛋白质的活性改变。

[0027] 在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。在某些实施方案中,变体序列与全部或部分SEQ ID NO:3有至少90%同一性。在某些实施方案中,变体序列与全部或部分SEQ ID NO:3相同。在某些实施方案中,变体序列的长

度为至少20个核苷酸。在某些实施方案中,变体序列的长度为至少50个核苷酸。在某些实施方案中,变体序列的长度为至少200个核苷酸。

[0028] 本发明的方面包括一种包含来源于野生型YkfA多肽序列(示于SEQ ID NO:2中)的变体序列的分离的多肽,其中所述变体序列:

[0029] 长度为至少5个氨基酸,

[0030] 与全部或部分SEQ ID NO:2有至少60%同一性,并且

[0031] 在YkfA多肽序列基因中包含至少一个氨基酸改变,导致YkfA蛋白质的活性改变。

[0032] 在某些实施方案中,改变在与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸中。在某些实施方案中,基因改变在与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸中。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。在某些实施方案中,变体多肽序列与全部或部分SEQ ID NO:4有至少90%同一性。在某些实施方案中,变体多肽序列与全部或部分SEQ ID NO:4相同。在某些实施方案中,变体多肽序列的长度为至少5个氨基酸。在某些实施方案中,变体多肽序列的长度为至少20个氨基酸。在某些实施方案中,变体多肽序列的长度为至少50个氨基酸。

[0033] 本发明的方面包括包含如上所述的多核苷酸序列的载体。在某些实施方案中,载体是靶向载体,所述靶向载体被设计成在转化到所述革兰氏阳性细菌细胞中时通过同源重组将所述多核苷酸序列中的至少一个突变引入到革兰氏阳性细菌细胞的ykf操纵子中的对应位置。

[0034] 本发明的方面包括一种用于增强革兰氏阳性细菌细胞中感兴趣的蛋白质表达的方法,所述方法包括:

[0035] a) 用以上载体转化亲本革兰氏阳性细菌细胞;

[0036] b) 允许所述载体与所述亲本革兰氏阳性细菌细胞的ykf操纵子中相应区域的同源重组,以产生改变的革兰氏阳性细菌细胞;以及

[0037] c) 使所述改变的革兰氏阳性细菌细胞在适于表达所述感兴趣的蛋白质的条件下生长,其中与所述转化步骤之前的所述革兰氏阳性细菌细胞相比,所述感兴趣的蛋白质在所述改变的革兰氏阳性细菌细胞中的产量提高。

[0038] 在某些实施方案中,亲本革兰氏阳性细菌细胞是芽孢杆菌属菌种的菌株。在某些实施方案中,芽孢杆菌属菌种的菌株选自:地衣芽孢杆菌、迟缓芽孢杆菌、枯草芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、灿烂芽孢杆菌、克劳氏芽孢杆菌、巨大芽孢杆菌和苏云金芽孢杆菌。在某些实施方案中,芽孢杆菌属菌种的菌株为枯草芽孢杆菌菌株。在某些实施方案中,改变的革兰氏阳性细菌细胞进一步在选自以下的基因中包含突变:degU、degQ、degS、scoC4、spoIIE和oppA。在某些实施方案中,突变为degU(Hy) 32。

[0039] 在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有降低的YkfA蛋白质活



性。在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中YkfA蛋白质活性相比,改变的革兰氏阳性细菌细胞具有提高的YkfA蛋白质活性。

[0040] 在某些实施方案中,突变在所述ykf操纵子的ykfA基因中。在一些实施方案中,基因改变在ykf操纵子的内源性ykfA基因中。在某些实施方案中,ykfA基因与SEQ ID NO:1有至少60%同一性。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(示于SEQ ID NO:4)。在某些实施方案中,感兴趣的蛋白质是同源蛋白质。在某些实施方案中,感兴趣的蛋白质是异源蛋白质。在某些实施方案中,感兴趣的蛋白质是酶。在某些实施方案中,该酶选自:蛋白酶、纤维素酶、支链淀粉酶、淀粉酶、糖酶、脂肪酶、异构酶、转移酶、激酶和磷酸酶。在某些实施方案中,感兴趣的蛋白质是蛋白酶。在某些实施方案中,蛋白酶为枯草杆菌蛋白酶。在某些实施方案中,枯草杆菌蛋白酶选自:枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN<sup>′</sup>、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147、枯草杆菌蛋白酶309、以及它们的变体。

[0041] 在某些实施方案中,该方法还包括回收所述感兴趣的蛋白质。

## 附图说明

[0042] 图1示出得自枯草芽孢杆菌的ykf操纵子的示意图。示出了实施例所述的沉默突变的位置。示出了ykfA、ykfB、ykfC和ykfD基因。示出了导致P252L和V253L的基因改变。

[0043] 图2A示出表达蛋白酶的CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)的细胞密度图。图2B示出CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)中FNA产量图。

[0044] 图3A示出产生GFP的CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)的细胞密度图。在进入稳定期(生长4至6小时之间)时,ykfA突变株的细胞生长的下降相较于对照菌株有延缓,这表明改善的细胞生存力归因于ykfA突变。图3B示出CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)中GFP产量图。该图示出GFP的产量从生长6小时提高,这是由于ykfA突变。

[0045] 图4A示出产生Bg1C的CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)的细胞密度图。包含ykfA突变的细胞具有更高的细胞生长,这表明由于存在ykfA突变而使细胞生存力改善。图4B示出CB15-14衍生株:CB15-14#1和#2(对照菌株)以及CB15-14ykfA#1和#2(包含ykfA突变的菌株)中Bg1C产量图。该图示出Bg1C的产量在所有时间点提高,这是由于存在ykfA突变。

## 具体实施方式

[0046] 本发明一般涉及具有导致感兴趣的蛋白质的表达提高的基因改变的细菌细胞以及制备和使用该细胞的方法。本发明的方面包括革兰氏阳性微生物如芽孢杆菌菌种的细胞,其具有改变ykf操纵子所编码的蛋白质的活性的基因改变,导致感兴趣的蛋白质的表达增强。

[0047] 在更详细地描述本发明的组合物和方法之前,应理解本发明的组合物和方法并不限于所描述的具体实施方案,因为这些实施方案当然是可变的。还应理解,本文所用的术语仅出于描述具体的实施方案的目的,并不意在具有限制意义,因为本发明的组合物和方法的范围将仅受所附权利要求的限定。

[0048] 在提供数值范围的情况中,应当理解,该范围和在该所述范围内的任意其它所述的值或中间值的上限和下限之间的每个中间数值(至下限的个位的十分之一,除非上下文另有清楚规定)也涵盖在本发明的组合物和方法之中。这些较小范围的上限和下限可独立地被包括在较小范围中,并且还被涵盖在本发明的组合物和方法中,但依据该规定的范围中的任何被具体排除的界限而定。在规定的范围包括界限中的一个或两个的情况中,排除这些被包括的界限中的任一个或两个的范围,也被包括在本发明的组合物和方法中。

[0049] 某些范围在本文中通过前面带有术语“约”的数值表述。术语“约”在本文中用于为它之后的确切数字以及接近或近似该术语之后的数字的数字提供字面支持。在确定数字是否接近或近似于明确所记载的数字方面,接近或近似的未记载数字可以是在其中陈述它的上下文中提供明确记载数字的基本上的等同物的数字。例如,与数值结合,术语“约”是指数值的-10%至+10%范围,除非术语在上下文中另外明确定义。又如,短语“约6的pH值”是指5.4至6.6的pH值,除非pH值另有明确定义。

[0050] 本文提供的标题并非对本发明的组合物和方法的各个方面或实施方案的限制,这些方面或实施方案可通过将说明书作为一个整体来参考而得到。相应地,下面就要定义的术语通过将说明书作为一个整体来参考会得到更完全的定义。

[0051] 本文件被组织成多个章节以便于阅读;然而,读者应理解,在一个章节中作出的陈述可适用于其它章节。这样,用于本公开的不同章节的标题不应理解为限制性的。

[0052] 除非另有定义,否则本文所用的所有技术和科学术语都具有本发明的组合物和方法所属领域普通技术人员通常理解的含义。尽管与本文所述的那些类似或等价的任何方法和材料还可用于本发明的组合物和方法的实践或测试,但现在将描述代表性的例示性方法和材料。

[0053] 本说明书中引用的所有出版物和专利以引用方式并入本文,如同各单独出版物或专利具体和单独地指示为通过引用并入一样,并且以引用方式并入本文以公开和描述与引用出版物有关的方法和/或材料。任何出版物的引用均指其申请日之前的公开,并不能被理解为承认由于先前发明而允许本发明组合物和方法不享受早于这些公布的权利。此外,提出的公开日期可以不同于实际的公开日期,这可能需要单独地确认。

[0054] 根据该具体实施方式,使用下面的缩写和定义。需注意,单数形式“一个”、“一种”和“所述”包括多个指代物,除非上下文明确指明不是这样。因此,例如,提到“酶”包括多种此类酶,而提到“剂量”包括提到一种或多种剂量以及本领域技术人员知道的它们的等同物,以此类推。

[0055] 还应该注意,权利要求书可能撰写成排除任何可选元素。因此,此陈述旨在充当使用与权利要求要素的引用相关的如“仅”、“只”等此类排他性术语或使用“否定性”限制的前提基础。

[0056] 还应当注意,如本文所用术语“基本上由…组成”是指组合物,其中术语之后的一个或多个组分是在总量为按所述组合物总体的重量计少于30%并不会影响或干扰一个或多个组分的作用或活性的一个或多个其它已知组分的存在下的。

[0057] 还应当注意,如本文所用术语“包括”意指包括但不限于术语“包括”之后的一个或多个组分。术语“包括”之后的一个或多个组分是必要的或强制性的,但是包含一个或多个组分的组合物还可包括一个或多个其它非强制性或任选组分。

[0058] 还应当注意,如本文所用术语“由…组成”意指包括并限于术语“由…组成”之后的一个或多个组分。术语“由…组成”之后的一个或多个组分因此是必要的或强制性的,并且没有一个或多个其它组分存在于组合物中。

[0059] 对本领域普通技术人员而言,阅读了本公开内容后将显而易见的是,本文所述和示意的各个单独实施方案都具有独立的部件和特征,在不背离本文所述的本发明的组合物和方法的范围或精神的前提下,这些部件和特征可容易地与其它多个实施方案中任意的特征分离或合并。任何叙述的方法可以按照叙述事件的顺序或以任何其它逻辑上可能的顺序进行。

#### [0060] 定义

[0061] 如本文所用,“宿主细胞”是指具有能够充当用于新导入的DNA序列的宿主或表达载体的细胞。

[0062] 在本发明的某些实施方案中,宿主细胞为细菌细胞,例如革兰氏阳性宿主细胞芽孢杆菌属菌种。

[0063] 如本文所用,“芽孢杆菌属”或“芽孢杆菌属菌种”包括“芽孢杆菌”属中的所有种,如本领域技术人员所知的,包括但不限于枯草芽孢杆菌、地衣芽孢杆菌、迟缓芽孢杆菌、短芽孢杆菌 (*B. brevis*)、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、解淀粉芽孢杆菌、克劳氏芽孢杆菌 (*B. clausii*)、耐盐芽孢杆菌 (*B. halodurans*)、巨大芽孢杆菌 (*B. megaterium*)、凝结芽孢杆菌 (*B. coagulans*)、环状芽孢杆菌 (*B. circulans*)、灿烂芽孢杆菌 (*B. lautus*) 和苏云金芽孢杆菌。已经认识到,芽孢杆菌属继续经历着分类学上的重构。因此,预期的是该属包括已被重新归类的种,包括但不限于诸如嗜热脂肪芽孢杆菌(其现在被命名为嗜热脂肪地芽孢杆菌 (*Geobacillus stearothermophilus*))之类的生物体。在氧气的存在下产生有抵抗力的内生孢子被认为是芽孢杆菌属的限定性特征,尽管该特征也适用于最新命名的脂环酸芽孢杆菌 (*Alicyclobacillus*)、双芽孢杆菌 (*Amphibacillus*)、解硫胺素芽孢杆菌 (*Aneurinibacillus*)、厌氧芽孢杆菌 (*Anoxybacillus*)、短芽孢杆菌 (*Brevibacillus*)、线状芽孢杆菌 (*Filobacillus*)、薄壁芽孢杆菌 (*Gracilibacillus*)、喜盐芽孢杆菌 (*Halobacillus*)、类芽孢杆菌 (*Paenibacillus*)、土壤芽孢杆菌 (*Salibacillus*)、耐热芽孢杆菌 (*Thermobacillus*)、解脲芽孢杆菌 (*Ureibacillus*) 和枝芽孢杆菌 (*Virgibacillus*)。

[0064] 如本文所用,“核酸”是指核苷酸或多核苷酸序列及其片段或部分,以及指可为双链或单链的基因组来源的或合成来源的DNA、cDNA和RNA,不论其代表有义链还是反义链。应当理解,由于遗传密码的简并性,多个核苷酸序列可编码给定的蛋白质。

[0065] 如本文所用,术语“载体”是指能在细胞中被复制并且可将新基因或DNA片段携带入细胞的任何核酸。因此,该术语是指设计成在不同宿主细胞之间进行转移的核酸构建体。“表达载体”是指能够将异源DNA片段掺入外源细胞中并使其表达的载体。许多原核和真核表达载体可商购获得。“靶向载体”是包含多核苷酸序列的载体,所述多核苷酸序列与多核苷酸序列转化到其中的宿主细胞的染色体区域同源并且可在该区域促使同源重组。靶向载体可用于通过同源重组将突变引入细胞的染色体中。在一些实施方案中,靶向载体包含例如添加到末端的其它非同源序列(即填充序列或旁侧序列)。可使端部闭合,由此使得靶向载体形成封闭的环,例如插入载体中。适当载体的选择和/或构建在本领域技术人员知识范围内。

[0066] 如本文所用,术语“质粒”是指用作克隆载体的环状双链(ds) DNA构建体,并且其在许多细菌和一些真核生物中形成染色体外自我复制的遗传元件。在一些实施方案中,质粒被掺入宿主细胞的基因组中。

[0067] “纯化的”或“分离的”或“富集的”意指生物分子(例如多肽或多核苷酸)由其天然状态进行改变,这是通过将其与部分或所有与天然状态相关的天然存在的组分进行分离。此类分离或纯化可以通过本领域公知的分离技术来实现,例如离子交换层析、亲和层析、疏水分离、透析、蛋白酶处理、硫酸铵沉淀或者其它蛋白质盐沉淀、离心、大小排阻层析、过滤、微过滤、凝胶电泳或者梯度分离来除去全部细胞、细胞碎屑、杂质、外来蛋白质、或在最终组合物中不期望的酶。然后还可向纯化或分离的生物分子组合物中加入提供附加有益效果的组分,例如活化剂、抗抑制剂、期望的离子、控制pH的化合物或者其它酶或化学物质。

[0068] 如本文所用,术语“增强的”、“改善的”和“提高的”在涉及感兴趣的生物分子(例如感兴趣的蛋白质)的表达时在本文互换使用,其意指生物分子的表达在未根据本文教导改变但在基本相同的生长条件下生长的相应宿主菌株(例如野生型和/或亲本株)中的表达水平之上。

[0069] 如本文所用,术语“表达”在用于蛋白质时是指这样一种方法:通过该方法基于基因的核酸序列产生蛋白质并因此包括转录和翻译两者。

[0070] 如本文所用,在将多核苷酸引入细胞中的语境中,术语“引入”是指任何适于将多核苷酸转移进细胞中的方法。用于引入的此类方法包括但不限于原生质体融合、转染、转化、缀合和转导(参见例如Ferrari等人,“Genetics,” in Hardwood等人(编辑),*Bacillus*, Plenum Publishing Corp.,第57-72页,[1989])。

[0071] 如本文所用,术语“转化的”和“稳定转化的”是指通过人工干预将多核苷酸序列引入其中的细胞。可将多核苷酸整合进细胞的基因组中或作为被维持至少两代的附加体质粒存在。

[0072] 如本文所用,术语“可选标记”或“选择性标记”是指能够在宿主细胞中表达的核酸(例如基因),其使得能容易地选择含有该核酸的那些宿主。此类可选标记的示例包括但不限于抗微生物剂。因此,术语“可选标记”是指提供这样指示的基因:宿主细胞已吸收引入感兴趣的DNA或已发生一些其它反应。通常,可选标记是赋予宿主细胞抗微生物抗性或代谢优势的基因以允许将含有外源DNA的细胞与在转化过程中未接受任何外源序列的细胞区分开。根据本发明可用的其它标记包括但不限于营养缺陷型标记,例如色氨酸;和检测标记,例如 $\beta$ -半乳糖苷酶。

[0073] 如本文所用,术语“启动子”是指起到引导下游基因的转录的作用的核酸序列。在实施方案中,启动子适于在其中表达靶基因的宿主细胞。启动子以及其他转录和翻译调控核酸序列(也称为“控制序列”)是表达给定基因所必需的。通常,转录和翻译调控序列包括但不限于启动子序列、核糖体结合位点、转录起始和终止序列、翻译起始和终止序列以及增强子或激活子序列。

[0074] 如本文所用,“功能性地连接”或“可操作地连接”意指具有已知或期望活性的调控区或功能结构域例如启动子、终止子、信号序列或增强子区域以使得调控区或功能结构域根据其已知或期望活性控制靶标(例如基因或多肽)的表达、分泌或功能的方式连接至或间接至该靶标。

[0075] 术语“基因改变”或“遗传改变”在用于描述重组细胞时意指细胞相较于亲本细胞具有至少一个基因差异。一个或多个遗传差异可为染色体突变(例如一个染色体区被另一个染色体区插入、缺失、取代、倒位、替换(例如用异源启动子替换染色体启动子)等)和/或引入染色体外多核苷酸(例如质粒)。在一些实施方案中,可将染色体外多核苷酸整合进宿主细胞的染色体中以产生稳定的转染体/转化体。本公开的实施方案包括改变由ykf操纵子(ykfA、ykfB、ykfC和ykfD)中基因编码的一个或多个蛋白质的活性的基因改变。如本文详述,此类改变改善感兴趣的蛋白质的表达。

[0076] 如本文所述,改变蛋白质活性可以以任何便利方式实现。例如,可通过提高或降低蛋白质的活性来改变蛋白质活性。也可通过例如改变蛋白质稳定性、蛋白质与蛋白质间的相互作用或结合、或改变蛋白质的底物特异性来改变蛋白质活性。活性的改变可处于转录、mRNA稳定化、翻译的水平,或者可能是由于在一个或多个由ykf操纵子产生的多肽中存在降低其活性的变异(即其为基于多肽活性的表达的“功能性”减少)。由此,并不意欲限制基因改变的类型或通过其使ykf操纵子所编码的至少一个蛋白质的表达或活性改变的方式。例如,在一些实施方案中,革兰氏阳性细胞中的基因改变为改变一个或多个启动子的基因改变,导致转录活性降低。在某些实施方案中,改变导致mRNA转录物的水平降低。或者,革兰氏阳性细胞中的基因改变可为改变核苷酸的基因改变,导致转录物在细胞中的稳定性降低。在某些实施方案中,减少一个或多个基因的表达的多于一个基因改变可存在于经基因改变的革兰氏阳性细胞中。

[0077] 基因“失活”意指基因的表达或其编码的生物分子的活性被阻断或以其它方式不能发挥其已知功能。失活可通过任何适宜的方式发生,例如通过如上所述的基因改变。在一个实施方案中,失活基因的表达产物是具有相应的蛋白质生物活性改变的截短的蛋白质。在一些实施方案中,改变的革兰氏阳性细菌菌株包括一个或多个基因的失活,其优选导致稳定的且不可逆的失活。

[0078] 在一些实施方案中,失活通过缺失来实现。在一些实施方案中,通过同源重组缺失靶向缺失的区域(例如基因)。例如,使用包含具有选择性标记的引入序列的DNA构建体(其中该序列在本文可称为“同源盒”),该选择性标记在每一侧侧接有与靶向缺失的区域同源的序列。将DNA构建体与宿主染色体的同源序列进行比对并在双交换事件中从宿主染色体中切除靶向缺失的区域。

[0079] “插入”或“添加”是核苷酸或氨基酸序列的变化,相较于天然存在的或亲本序列,其分别导致添加了一个或多个核苷酸或氨基酸残基。

[0080] 如本文所用,“取代”是分别用不同核苷酸或氨基酸替换一个或多个核苷酸或氨基酸的结果。

[0081] 使基因突变的方法在本领域中是公知的并且包括但不限于定点突变、产生随机突变以及缺口双链方法(参见例如美国专利4,760,025;Moring等人,Biotech.2:646[1984];和Kramer等人,Nucleic Acids Res.,12:9441[1984])。

[0082] 如本文所用,“同源基因”是指来自于不同的但通常相关的物种的一对基因,它们彼此对应,并且彼此相同或非常相似。该术语包括通过物种形成(即新物种的进化)分离的基因(例如直系同源基因),以及通过遗传复制已分离的基因(例如旁系同源基因)。

[0083] 如本文所用,“直系同源物”和“直系同源基因”是指不同物种中通过物种形成而从共同的先祖基因(即同源基因)进化而来的基因。通常,直系同源物在进化的过程中保留了相同的功能。对直系同源物的鉴别可用于可靠地预测新测序的基因组中的基因功能。

[0084] 如本文所用,“旁系同源物”和“旁系同源基因”是指因基因组内的复制而相关联的基因。直系同源物在进化的过程中保留相同的功能,而旁系同源物进化出新的功能,虽然一些功能通常与原始的功能相关。旁系同源基因的示例包括但不限于编码胰蛋白酶、胰凝乳蛋白酶、弹性蛋白酶和凝血酶的基因,它们都是丝氨酸蛋白酶并一起出现在相同的物种中。

[0085] 如本文所用,“同源性”是指序列相似性或同一性,同一性是优选的。使用本领域中已知的标准技术来确定此种同源性(参见例如Smith and Waterman,Adv.Appl.Math.,2:482[1981];Needleman和Wunsch,J.Mol.Biol.,48:443[1970];Pearson和Lipman,Proc.Natl.Acad.Sci.USA 85:2444[1988];Wisconsin Genetics软件包(Genetics Computer Group,Madison,WI)中的程序,诸如GAP、BESTHT、FASTA和TFASTA;和Devereux等人,Nucl.Acid Res.,12:387-395[1984])。

[0086] 如本文所用,“相似序列”其中的基因功能与枯草芽孢杆菌菌株168表示的基因基本上相同。另外,相似基因与枯草芽孢杆菌菌株168基因的序列有至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%序列同一性。作为另外一种选择,相似序列具有枯草芽孢杆菌168区域所见的基因的70至100%的比对和/或具有与枯草芽孢杆菌168染色体中的基因所比对的区域中所见的至少5-10个基因。在另外的实施方案中,将多于一个以上特性应用于序列。通过已知的序列比对方法来测定相似序列。常用的比对方法为BLAST,但如上文和下文所指出的那样,还存在还可用于比对序列的其它方法。

[0087] 可用的算法的一个示例是PILEUP。PILEUP利用渐进性双序列比对从一组相关序列产生多序列比对。其还可以绘出关系树,该关系树示出了用于产生比对的聚类关系。PILEUP使用Feng和Doolittle的渐进比对方法的简化形式(Feng and Doolittle,J.Mol.Evol.,35:351-360[1987])。该方法类似于Higgins和Sharp所描述的方法(Higgins和Sharp,CABIOS 5:5:151-153[1989])。可用的PILEUP参数包括:默认空位权重=3.00,默认空位长度权重=0.10,以及加权的末端空位。

[0088] 可用算法的另一个示例为BLAST算法,Altschul等人(Altschul等人,J.Mol.Biol.,215:403-410,[1990];和Karlin等人,Proc.Natl.Acad.Sci.USA 90:5873-5877[1993])对其进行了描述。特别有用的BLAST程序是WU-BLAST-2程序(参见Altschul等人,Meth.Enzymol.,266:460-480[1996])。WU-BLAST-2使用数个搜索参数,其中大多数被设置为默认值。可调整的参数用下列值来设定:重叠跨度(overlap span)=1,重叠分数

(overlapfraction)=0.125,字阈值(word threshold)(T)=11。HSP S和HSP S2参数是动态的值,并且取决于具体序列的组成和据以对所关注序列进行检索的具体数据库的组成而由程序本身确立。然而,可调节所述值以提高灵敏度。氨基酸序列同一性百分数值是由匹配的相同残基的数目除以比对区域中“较长”序列的残基总数来确定。所述“较长”序列是在比对区域中具有最多实际残基的序列(忽略由WU-Blast-2引入以使比对得分最大化的空位)。

[0089] 如本文所用,相对于本文所鉴定的氨基酸或核苷酸序列的“序列同一性百分比(%)”定义为,在比对序列和引入缺口(如果需要的话)以获得最大序列同一性百分比之后,与Ma13A序列中氨基酸残基或核苷酸相同的候选序列中的氨基酸残基或核苷酸的百分比,而不考虑任何保守替代作为序列同一性的一部分。

[0090] “同源物”(或“同系物”)将意指与主题氨基酸序列和主题核苷酸序列具有指定程度的同一性的实体。同源序列可包括使用常规序列对比工具(例如Clustal、BLAST等)与主题序列有至少60%、65%、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或甚至99%同一性的氨基酸序列。通常,除非另外指明,同源物将包括与例如主题氨基酸序列相同的活性位点残基。

[0091] 用于进行序列比对和测定序列同一性的方法对技术人员是已知的,可以在无需过多实验的情况下进行,并且可以明确得到同一性的计算值。参见例如Ausubel等人编辑(1995)Current Protocols in Molecular Biology,第19章(Greene Publishing and Wiley-Interscience,New York);和ALIGN程序(Dayhoff(1978),Atlas of Protein Sequence and Structure 5:增刊3(National Biomedical Research Foundation,Washington,D.C.))。多个算法可用于比对序列和测定序列同一性,并且包括例如Needleman等人(1970)J.Mol.Biol.48:443的同源比对算法;Smith等人(1981)Adv.Appl.Math.2:482的局部同源性算法;Pearson等人(1988)Proc.Natl.Acad.Sci.85:2444的相似性搜索方法;Smith-Waterman算法(Meth.Mol.Biol.70:173-187(1997));以及BLASTP、BLASTN和BLASTX算法(参见Altschul等人(1990)J.Mol.Biol.215:403-410)。

[0092] 使用这些算法的计算机化程序也是可用的,并且包括但不限于:ALIGN或Megalign(DNASTAR)软件、或WU-BLAST-2(Altschul等人,Meth.Enzym.,266:460-480(1996));或GAP、BESTFIT、BLAST、FASTA和TFASTA,在遗传学计算组(GCG)软件包(版本8,Madison,Wisconsin,USA)可得到;和Intelligenetics(Mountain View,California)的PC/Gene程序中的CLUSTAL。本领域的技术人员可确定用于测量比对的适当参数,包括在要比较的序列的长度上实现最大比对所需的算法。优选地,序列同一性用程序所确定的预设参数来测定。具体地,序列同一性可通过使用Clustal W(Thompson J.D.等人(1994)Nucleic Acids Res.22:4673-4680)采用预设参数来确定。

	空位开放罚分:	10.0
	空位延伸罚分:	0.05
	蛋白质权重矩阵:	BLOSUM 系列
	DNA 权重矩阵:	IUB
	延迟发散序列%:	40
[0093]	空位间隔距离:	8
	DNA 转换权重:	0.50
	亲水性残基列表:	GPSNDQEK R
	使用负矩阵:	关 (OFF)
	切换残基特定罚分:	开 (ON)
	切换亲水性罚分:	开 (ON)
	切换末端空位间隔罚分	关 (OFF)

[0094] 如本文所用,术语“杂交”是指用于使一条核酸链通过本领域已知的碱基配对与互补链接合的方法。

[0095] 如果一条核酸序列与参考核酸序列在中等至高严格性杂交和洗涤条件下彼此特异性杂交,则该核酸序列被认为是与该参考核酸序列“可选择性杂交”。杂交条件是基于核酸结合复合物或探针的解链温度( $T_m$ )。例如,“最大严格性”通常发生在约 $T_m-5^{\circ}\text{C}$ (低于探针的 $T_m5^{\circ}\text{C}$ )下;“高严格性”在低于该 $T_m$ 约 $5-10^{\circ}\text{C}$ 下;“中度严格性”在低于探针的 $T_m$ 约 $10-20^{\circ}\text{C}$ 下;而“低严格性”在低于该 $T_m$ 约 $20-25^{\circ}\text{C}$ 下。功能上,最大严格性条件可用于鉴别与杂交探针有严格同一性或接近严格同一性的序列;而中度或低严格性杂交可用于鉴别或检测多核苷酸序列同源物。

[0096] 中等和高严格性杂交条件是本领域所熟知的。高严格性条件的示例包括在约 $42^{\circ}\text{C}$ 下,在50%甲酰胺、5X SSC、5X邓哈特溶液(Denhardt's solution)、0.5%SDS和100 $\mu\text{g/ml}$ 变性载体DNA中杂交,然后在室温下,在2X SSC和0.5%SDS中洗涤两次,并且在 $42^{\circ}\text{C}$ 下,在0.1X SSC和0.5%SDS中再洗涤两次。中度严格性条件的示例包括在 $37^{\circ}\text{C}$ 下,在包含20%甲酰胺、5x SSC(150mM NaCl、15mM柠檬酸三钠)、50mM磷酸钠(pH 7.6)、5x邓哈特溶液、10%硫酸葡聚糖和20mg/ml变性剪切鲑鱼精子DNA的溶液中过夜温育,然后在约 $37-50^{\circ}\text{C}$ 下,在1x SSC中洗涤过滤器。本领域的技术人员了解如何根据需要来调整温度、离子强度等以适应诸如探针长度等的因素。

[0097] 术语“重组的”在用于生物组分或组合物(例如细胞、核酸、多肽/酶、载体等)时是指生物组分或组合物处于非天然存在的状态。换言之,生物组分或组合物已通过人工干预由其自然状态进行修饰。例如,重组细胞涵盖表达一个或多个在其天然亲本(非重组)细胞中未发现的基因的细胞,表达的一个或多个天然基因的量不同于其天然亲本细胞的细胞,和/或在不同于其天然亲本细胞的条件下表达一个或多个天然基因的细胞。重组核酸可与天然序列有一个或多个核苷酸的差异,可被可操作地连接至异源序列(例如异源启动子、编码非天然或变体信号序列的序列等),可不含内含子序列,和/或可为分离形式。重组多肽/酶可与天然序列有一个或多个氨基酸的差异,可与异源序列融合,可为截短的或具有内部缺失的氨基酸,可以以在天然细胞中未发现的方式表达(例如来自过表达多肽的重组细胞,这是由于细胞中存在编码多肽的表达载体),和/或可为分离形式。应当强调在一些实施方案中,重组多核苷酸或多肽/酶具有与其野生型对应物相同但为非天然形式(例如为分离或



富集形式)的序列。

[0098] 如本文所用,术语“靶序列”是指宿主细胞中的DNA序列,其所编码的序列是引入序列插入到宿主细胞基因组中的期望位置。在一些实施方案中,靶序列编码功能性野生型基因或操纵子,而在其它实施方案中,靶序列编码功能性突变基因或操纵子,或非功能性基因或操纵子。

[0099] 如本文所用,“旁侧序列”是指所讨论序列的上游或下游的任何序列(例如对于基因A-B-C,基因B侧接有A和C基因序列)。在一个实施方案中,引入序列在各侧侧接有同源盒。在另一个实施方案中,引入序列和同源盒包含在各侧侧接有填充序列的单元。在一些实施方案中,旁侧序列仅存在于单侧(3'或5'),而在一些实施方案中,其存在于被侧接序列的各侧。各个同源盒的序列与芽孢杆菌染色体的序列同源。这些序列指导新构建体在芽孢杆菌染色体中得以整合的地方以及芽孢杆菌染色体将被引入序列所替代的部分。在一个实施方案中,选择性标记的5'和3'端侧接有包含失活染色体片段的片段的多核苷酸序列。在一些实施方案中,旁侧序列仅存在于单侧(3'或5'),而在一些实施方案中,其存在于被侧接序列的各侧。

[0100] 如本文所用,术语“可扩增标记”、“可扩增基因”和“扩增载体”是指在适当生长条件下允许该基因扩增的基因或编码基因的载体。

[0101] “模板特异性”在大多数扩增技术中是通过酶的选择来实现的。扩增酶是这样的酶:在它们被使用的条件下,仅加工异源核酸混合物中核酸的特异性序列。例如,就QB复制酶而言,MDV-1RNA是复制酶的特异性模板(参见例如Kacian等人,Proc.Natl.Acad.Sci.USA 69:3038[1972])。该扩增酶不对其它核酸进行复制。相似地,就T7RNA聚合酶而言,该扩增酶对其自身启动子具有严格的特异性(参见Chamberlin等人,Nature 228:227[1970])。就T4DNA连接酶而言,当寡核苷酸或多核苷酸底物与模板之间在连接结合点处存在错配时,该酶不会连接两个寡核苷酸或多核苷酸(参见Wu和Wallace,Genomics 4:560[1989])。最后,Taq和Pfu聚合酶由于其能够在高温下发挥作用而发现对于被引物结合并由此被限定的序列展现出高特异性;高温导致了有利于引物与靶序列杂交而不与非靶序列杂交的热力学条件。

[0102] 如本文所用,术语“可扩增核酸”是指可通过任何扩增方法进行扩增的核酸。预期“可扩增核酸”将通常包括“样品模板”。

[0103] 如本文所用,术语“样品模板”是指源自被用于分析“靶”(在下面给予定义)的存在的样品的核酸。相比之下,“背景模板”被用来指代除样品模板之外的核酸,其可能存在于或可能不存在于样品中。背景模板通常是最容易被疏忽的。这可为物质遗留的结果,或者可能是由于试图从样品中纯化去除的核酸污染物的存在。例如,来自生物体而非那些待检测的核酸可在测试样品中作为背景而存在。

[0104] 如本文所用,术语“引物”是指天然存在的寡核苷酸(如经纯化限制性消化物中)或合成产生的寡核苷酸,当置于与核酸链互补的引物延伸产物的合成被诱导的条件下(即存在核苷酸和诱导剂如DNA聚合酶的情况下,以及在合适的温度和pH下)时,它能够作为合成的起始点而发挥作用。引物优选为单链以达到最大的扩增效率,但也可为双链。如果是双链的,在用于制备延伸产物之前,首先对该引物进行处理以使其双链分开。优选地,引物为寡聚脱氧核苷酸。该引物必须足够长,以在诱导剂存在的情况下引发延伸产物的合成。引物

的确切长度将依赖于许多因素,包括温度、引物的来源以及使用的方法。

[0105] 如本文所用,术语“探针”是指能够与另一种感兴趣的寡核苷酸杂交的一种寡核苷酸(即核苷酸的序列),无论其是作为经纯化限制性酶消化物而天然存在还是以合成、重组或通过PCR扩增方式而产生。探针可为单链或双链的。探针可用于检测、鉴定和分离特定基因序列。预期本发明使用的任何探针都可以用任何“报告分子”标记,以便在任何检测系统中可检测到,包括但不限于酶(例如ELISA,以及基于酶的组织化学测定法)、荧光、放射活性和发光系统。无意于将本发明限制为任何特定的检测系统或标签。

[0106] 如本文所用,术语“靶”在用于聚合酶链反应中时,是指由用于聚合酶链反应的引物所结合的核酸区域。因此,试图将“靶”从其它核酸序列中挑选出来。“片段”定义为在靶序列内的核酸区域。

[0107] 如本文所用,术语“聚合酶链反应”(“PCR”)是指美国专利4,683,195 4,683,202和4,965,188的方法,其以引用方式并入本文,这些专利包括不通过克隆或纯化而提高靶序列的片段在基因组DNA混合物中的浓度的方法。用于扩增靶序列的该过程包括将大大过量的两种寡核苷酸引物引入到含有期望靶序列的DNA混合物中,之后在存在DNA聚合酶的情况下进行精确顺序的热循环。这两种引物与其双链靶序列中的对应链互补。为了实现扩增,使混合物变性,然后使引物退火到其在靶分子内的互补序列上。在退火之后,采用聚合酶使引物延伸,以便形成一对新的互补链。变性、引物退火和聚合酶延伸的步骤可重复多次(即变性、退火和延伸构成一个“循环”;可以有多个“循环”),以获得高浓度的期望靶序列的扩增片段。期望靶序列的扩增片段的长度由引物相对于彼此的相对位置决定,因此该长度为可控制的参数。由于该过程的重复方面,该方法被称为“聚合酶链反应”(下文为“PCR”)。由于靶序列的期望扩增片段成为混合物中的优势序列(就浓度而言),所以它们被称为“PCR扩增的”。

[0108] 如本文所用,术语“扩增试剂”是指除引物、核酸模板和扩增酶之外扩增所需的那些试剂(脱氧核糖核苷酸三磷酸、缓冲剂等)。通常,扩增试剂以及其它反应组分被放置于和包含于反应容器中(试管、微孔等)。

[0109] 借助PCR,可以将基因组DNA中的特定靶序列的单一拷贝扩增到通过几种不同的方法可检测的水平(例如与标记探针杂交;引入生物素酰化引物,之后进行抗生物素蛋白-酶偶联物检测;将<sup>32</sup>P标记的脱氧核糖核苷三磷酸例如dCTP或者dATP引入到扩增片段中)。除了基因组DNA之外,任何寡核苷酸或多核苷酸序列也可用适当组合的引物分子来扩增。具体地,经由PCR过程自身而产生的扩增片段,其本身就是后续PCR扩增的有效模板。

[0110] 如本文所用,术语“PCR产物”、“PCR片段”和“扩增产物”指在变性、退火和延伸的PCR步骤的两个或更多个循环完成后所得的化合物的混合物。这些术语涵盖这样的情况,即一个或多个靶序列的一个或多个片段已被扩增。

[0111] 如本文所用,术语“RT-PCR”是指RNA序列的复制和扩增。在该方法中,逆转录与PCR偶联,最经常使用其中采用热稳定聚合酶的一步酶法(one enzyme procedure),如在美国专利5,322,770中有所描述,其以引用方式并入本文。在RT-PCR中,RNA模板由于聚合酶的逆转录酶活性而转化为cDNA,然后利用聚合酶的聚合活性进行扩增(即如在其它PCR方法中)。

[0112] 如本文所用,“基因改变的宿主菌株”(例如基因改变的芽孢杆菌菌株)是指遗传工程化的宿主细胞,也称为重组宿主细胞。在一些实施方案中,与在基本相同的生长条件下生长的相应未改变的宿主细胞中的相同感兴趣的蛋白质的表达和/或产量相比,经基因改变

的宿主细胞的感兴趣的蛋白质的表达增强(提高)。在一些实施方案中,增强水平的表达由 ykf 操纵子所编码的一个或多个蛋白质的活性改变引起。在一些实施方案中,改变的菌株为具有一个或多个缺失的固有染色体区域或其片段的经遗传工程化的芽孢杆菌属菌种,其中与在基本相同的生长条件下生长的相应未改变的芽孢杆菌宿主菌株相比,感兴趣的蛋白质具有增强水平的表达或产量。

[0113] 如本文所用,“相应未改变的芽孢杆菌菌株”等为不具有指示的基因改变的宿主菌株(例如,起始(亲本)和/或野生型菌株)。

[0114] 本文所用的术语“染色体整合”是指这样一种方法,通过该方法将引入序列引入到宿主细胞(例如芽孢杆菌)的染色体中。对转化的DNA的同源区域与染色体的同源区域进行比对。随后,在双交换(即同源重组)中,同源盒之间的序列被引入序列替换。在本发明的一些实施方案中,将DNA构建体的失活染色体片段的同源部分与芽孢杆菌染色体的固有染色体区域的旁侧同源区进行比对。随后,在双交换(即同源重组)中通过DNA构建体缺失掉固有染色体区域。

[0115] “同源重组”意指在相同或几乎相同的核苷酸序列的位点处DNA片段在两个DNA分子或成对染色体之间进行交换。在一个实施方案中,染色体整合是同源重组。

[0116] 如本文所用“同源序列”意指在最佳比对进行比较时,核酸或多肽序列与另一个核酸或多肽序列有100%、99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、88%、85%、80%、75%或70%序列同一性。在一些实施方案中,同源序列具有85%和100%之间的序列同一性,而在其它实施方案中,存在90%和100%之间的序列同一性,并且在更多实施方案中,存在95%和100%的序列同一性。

[0117] 如本文所用,“氨基酸”是指肽或蛋白质序列或其部分。术语“蛋白质”、“肽”和“多肽”可互换使用。

[0118] 如本文所用,“感兴趣的蛋白质”和“感兴趣的多肽”是指期望的和/或待评估的蛋白质/多肽。在一些实施方案中,感兴趣的蛋白质在细胞内,而在其它实施方案中,其为分泌性多肽。具体地讲,多肽包括酶,包括但不限于选自淀粉分解酶、蛋白水解酶、溶细胞酶、氧化还原酶和植物细胞壁降解酶的那些。更具体地,这些酶包括但不限于淀粉酶、蛋白酶、木聚糖酶、脂肪酶、漆酶、酚氧化酶、氧化酶、角质酶、纤维素酶、半纤维素酶、酯酶、过氧化物酶、过氧化氢酶、葡萄糖氧化酶、植酸酶、果胶酶、葡萄糖苷酶、异构酶、转移酶、半乳糖苷酶和壳多糖酶。在本发明的特定实施方案中,感兴趣的多肽为蛋白酶。在一些实施方案中,感兴趣的蛋白质是融合至信号肽的分泌性多肽(即在待分泌蛋白质上氨基端的延伸)。几乎所有的分泌性蛋白质采用氨基末端蛋白质延伸,这在靶向前体蛋白质和前体蛋白质跨膜运输中起至关重要的作用。该延伸在膜传输期间或紧接膜传输之后通过信号肽酶蛋白水解切除。

[0119] 在本发明的一些实施方案中,感兴趣的多肽选自激素、抗体、生长因子、受体等。本发明所涵盖的激素包括但不限于卵泡刺激激素、黄体化激素、促肾上腺皮质激素释放因子、促生长素抑制素、促性腺激素、后叶加压素、催产素、促红细胞生成素、胰岛素等。生长因子包括但不限于血小板衍生生长因子、胰岛素样生长因子、表皮生长因子、神经生长因子、成纤维细胞生长因子、转化生长因子、细胞因子例如白介素(例如IL-1至IL-13)、干扰素、集落刺激因子等。抗体包括但不限于直接从期望由其产生抗体的任何物种获得的免疫球蛋白。

此外,本发明涵盖修饰的抗体。本发明也涵盖多克隆和单克隆抗体。在特定实施方案中,抗体为人抗体。

[0120] 如本文所用,多肽的“衍生物”或“变体”是指这样的多肽:通过在C末端或N末端任一端或这两个末端添加一个或多个氨基酸、在氨基酸序列中的一个或多个不同位点置换一个或多个氨基酸、在多肽的任一端或两个末端或在氨基酸序列中的一个或多个位点缺失一个或多个氨基酸、在氨基酸序列中的一个或多个位点插入一个或多个氨基酸、以及它们的任意组合而来源于前体多肽(例如,天然多肽)。多肽衍生物/变体的制备可以以任意方便的方式实现,例如通过修饰编码天然多肽的DNA序列、将该DNA序列转化进合适的宿主中并使该经修饰的DNA序列表达而形成衍生多肽/变体多肽。衍生物或变体还包括经化学修饰的多肽。

[0121] 如本文所用,术语“异源蛋白质”是指不在宿主细胞中天然出现的蛋白质或多肽。异源蛋白质的示例包括酶,例如水解酶,包括蛋白酶、纤维素酶、淀粉酶、糖酶和脂肪酶;异构酶例如消旋酶、表异构酶、互变异构酶或变位酶;转移酶、激酶和磷酸酶。在一些实施方案中,蛋白质是治疗上有意义的蛋白质或肽,包括但不限于生长因子、细胞因子、配体、受体和抑制剂、以及疫苗和抗体。在另外的实施方案中,蛋白质是商业上重要的工业蛋白质/肽(例如蛋白酶、糖酶例如淀粉酶和葡糖淀粉酶、纤维素酶、氧化酶和脂肪酶)。在一些实施方案中,编码蛋白质的基因为天然存在的基因,而在其它实施方案中,使用突变的和/或合成的基因。

[0122] 如本文所用,“同源蛋白质”是指细胞中天然的或天然存在的蛋白质或多肽。在一些实施方案中,细胞为革兰氏阳性细胞,而在特定实施方案中,细胞为芽孢杆菌宿主细胞。在另选的实施方案中,同源蛋白质是其它生物体产生的天然蛋白质,包括但不限于大肠杆菌(E.coli)。本发明涵盖经由重组DNA技术产生同源蛋白质的宿主细胞。

[0123] 如本文所用,“操纵子”包括可从一个共同的启动子作为一个转录单位被转录并由此经受共调节的连续基因组。在一些实施方案中,操纵子可包括驱动多个不同mRNA转录的多个启动子。

[0124] 本发明一般涉及具有导致感兴趣的蛋白质的表达提高的基因改变的细菌细胞以及制备和使用该细胞的方法。本发明的方面包括革兰氏阳性微生物如芽孢杆菌菌种,其具有改变ykf操纵子所编码的蛋白质的活性并导致感兴趣的蛋白质的表达增强的基因改变。

[0125] 如上所概述,本发明的方面包括用于使来自革兰氏阳性细菌细胞的感兴趣的蛋白质的表达提高的方法,并且基于如下观察:在经基因改变的革兰氏阳性细胞中产生的感兴趣的蛋白质提高,从而与相应的未经基因改变的革兰氏阳性细胞(例如野生型和/或亲本细胞)中相同的感兴趣的蛋白质的表达水平相比,ykf操纵子(例如ykfA)所编码的一个或多个蛋白质的活性改变。在一些实施方案中,对革兰氏阳性细胞进行基因改变,以降低由ykf操纵子(例如ykfA)所编码的一个或多个蛋白质的活性。在一些实施方案中,对革兰氏阳性细胞进行基因改变,以提高由ykf操纵子(例如ykfA)所编码的一个或多个蛋白质的活性。遗传改变意指宿主细胞中的任何改变,例如通过附加体添加和/或染色体插入、缺失、倒位、碱基改变等改变宿主细胞基因组成。并非意欲对这一点进行限制。

[0126] 在某些实施方案中,该方法涉及制备或获取包含至少一个基因改变的改变的革兰氏阳性细菌细胞,该基因改变使ykf操纵子所编码的一个或多个蛋白质的活性改变并能够

产生感兴趣的蛋白质,并在使得改变的革兰氏阳性细菌细胞表达感兴趣的蛋白质的条件下培养改变的革兰氏阳性细菌细胞。与基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中感兴趣的蛋白质的表达相比,感兴趣的蛋白质在改变的革兰氏阳性细菌细胞中的表达由此提高。

[0127] 根据某些实施方案,经基因改变的革兰氏阳性细菌细胞(或产生经基因改变的革兰氏阳性细菌细胞的亲本细胞)可为芽孢杆菌菌株。在一些实施方案中,感兴趣的芽孢杆菌菌株是嗜碱性的。已知多种嗜碱芽孢杆菌菌株(参见例如美国专利5,217,878;和Aunstrup等人,Proc IV IFS:Ferment.Technol.Today,299-305[1972])。在一些实施方案中,感兴趣的芽孢杆菌菌株是工业芽孢杆菌菌株。工业芽孢杆菌菌株的示例包括但不限于地衣芽孢杆菌、迟缓芽孢杆菌(*B.lentus*)、枯草芽孢杆菌和解淀粉芽孢杆菌。在另外的实施方案中,芽孢杆菌宿主菌株选自迟缓芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、苏云金芽孢杆菌、克劳氏芽孢杆菌和巨大芽孢杆菌(*B.megaterium*)以及如上所述的芽孢杆菌属内的其它生物体。在特定实施方案中,使用枯草芽孢杆菌。例如,美国专利5,264,366和4,760,025(RE 34,606)描述了在本发明可使用的各种芽孢杆菌宿主菌株,但其它适宜的菌株也预期可用于本发明。

[0128] 如本文所述的经基因改变的细胞的亲本株(例如亲本芽孢杆菌菌株)可为工业用菌株,其包括非重组菌株、天然存在的菌株的突变变株、或重组菌株。在某些实施方案中,亲本株为重组宿主菌株,其中已将编码感兴趣的多肽的多核苷酸引入宿主中。虽然编码感兴趣的多肽的多核苷酸的引入可在亲本株中进行时,但该步骤还可在已经基因改变以提高多肽产量(如本文详述)的菌株中进行。在一些实施方案中,宿主菌株为枯草芽孢杆菌宿主菌株,例如重组枯草芽孢杆菌宿主菌株。

[0129] 已知多种枯草芽孢杆菌菌株可用于本发明的方面,包括但不限于1A6(ATCC 39085)、168(1A01)、SB19、W23、Ts85、B637、PB1753至PB1758、PB3360、JH642、1A243(ATCC 39,087)、ATCC 21332、ATCC 6051、MI113、DE100(ATCC 39,094)、GX4931、PBT 110和PEP 211菌株(参见例如Hoch等人,Genetics,73:215-228[1973];美国专利4,450,235;美国专利4,302,544;和EP 0134048)。Palva等人和其他人(参见Palva等人,Gene 19:81-87[1982];还参见Fahnestock和Fischer,J.Bacteriol.,165:796-804[1986];和Wang等人,Gene 69:39-47[1988])进一步描述了枯草芽孢杆菌作为表达宿主的用途。

[0130] 在某些实施方案中,产生芽孢杆菌菌株的工业用蛋白酶可用作亲本表达宿主。在一些实施方案中,在本发明使用的这些菌株提供效率和蛋白酶产量的进一步增强。两种通用类型的蛋白酶通常由芽孢杆菌属菌种分泌,即中性(或“金属蛋白酶”)和碱性(或“丝氨酸”)蛋白酶。丝氨酸蛋白酶是催化肽键水解的酶,其中在活性位点存在必需的丝氨酸残基。丝氨酸蛋白酶具有范围为25,000至30,000的分子量(参见Priest,Bacteriol.Rev.,41:711-753[1977])。枯草杆菌蛋白酶为用于本发明的丝氨酸蛋白酶。已鉴定并测序了各种各样的枯草芽孢杆菌,例如枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN<sup>+</sup>、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147和枯草杆菌蛋白酶309(参见例如EP 414279 B;WO 89/06279;和Stahl等人,J.Bacteriol.,159:811-818[1984])。在本发明的一些实施方案中,芽孢杆菌宿主菌株产生突变(例如变体)蛋白酶。多个参考文献提供变体蛋白酶和基准的示例(参见例如WO 99/20770;WO 99/20726;WO 99/20769;WO 89/06279;RE 34,606;美国专利4,

914,031;美国专利4,980,288;美国专利5,208,158;美国专利5,310,675;美国专利5,336,611;美国专利5,399,283;美国专利5,441,882;美国专利5,482,849;美国专利5,631,217;美国专利5,665,587;美国专利5,700,676;美国专利5,741,694;美国专利5,858,757;美国专利5,880,080;美国专利6,197,567;和美国专利6,218,165)。

[0131] 此处应注意,本发明不限于作为感兴趣的蛋白质的蛋白酶。实际上,本公开涵盖各种各样的感兴趣的蛋白质,期望该感兴趣的蛋白质在革兰氏阳性细胞中的表达提高(以下详述)。

[0132] 在其它实施方案中,用于本发明的方面的菌株可在提供有益表型的其它基因中具有另外的基因改变。例如,可采用在至少一个以下基因degU、degS、degR和degQ中包含突变或缺失的芽孢杆菌属菌种。在一些实施方案中,突变在degU基因中,例如degU (Hy) 32突变。(参见Msadek等人,J.Bacteriol.,172:824-834[1990];和Olmos等人,Mol.Gen.Genet.,253:562-567[1997])。因此,可用于本发明的方面的亲本/基因改变的革兰氏阳性菌株的一个示例为携带degU32 (Hy) 突变的枯草芽孢杆菌细胞。在另一个实施方案中,芽孢杆菌宿主可包含scoC4(参见Caldwell等人,J.Bacteriol.,183:7329-7340[2001]);spoIIE(参见Arigoni等人,Mol.Microbiol.,31:1407-1415[1999]);oppA或opp操纵子的其它基因(参见Perego等人,Mol.Microbiol.,5:173-185[1991])的突变或缺失。实际上,预期opp操纵子中的任何突变(导致与oppA基因中的突变相同的表型)可用于本发明的改变的芽孢杆菌菌株的一些实施方案中。在一些实施方案中,这些突变可单独发生,而在其它实施方案中,存在突变的组合。在一些实施方案中,本发明的改变的芽孢杆菌获自亲本芽孢杆菌宿主菌株,该菌株已包含一个或多个以上所提及基因的突变。在另选实施方案中,使本发明的先前经基因改变的芽孢杆菌进一步工程化以包含一个或多个以上提及基因的突变。

[0133] 在某些实施方案中,在经基因改变的革兰氏阳性细胞中,ykf操纵子编码的一个或多个蛋白质的活性降低至在基本相同培养条件下培养的野生型和/或亲本细胞中的活性水平的约3%,包括约4%、约5%、约6%、约7%、约8%、约9%、约10%、约11%、约12%、约13%、约14%、约15%、约16%、约17%、约18%、约19%、约20%、约21%、约22%、约23%、约24%、约25%、约26%、约27%、约28%、约29%、约30%、约35%、约40%、约45%、约50%、约55%、约60%、约65%、约70%、约75%、或约80%。因此,ykf操纵子中一个或多个基因的表达减少的范围可为约3%至约80%、约4%至约75%、约5%至约70%、约6%至约65%、约7%至约60%、约8%至约50%、约9%至约45%、约10%至约40%、约11%至约35%、约12%至约30%、约13%至约25%、约14%至约20%等。可以考虑如上所述范围内的表达的任何子范围。

[0134] 在某些实施方案中,在经基因改变的革兰氏阳性细胞中,ykf操纵子编码的一个或多个蛋白质的活性提高至在基本相同培养条件下培养的野生型和/或亲本细胞中的活性水平的约3%,包括约4%、约5%、约6%、约7%、约8%、约9%、约10%、约11%、约12%、约13%、约14%、约15%、约16%、约17%、约18%、约19%、约20%、约21%、约22%、约23%、约24%、约25%、约26%、约27%、约28%、约29%、约30%、约35%、约40%、约45%、约50%、约55%、约60%、约65%、约70%、约75%、或约80%。因此,ykf操纵子中一个或多个基因的表达减少的范围可为约3%至约80%、约4%至约75%、约5%至约70%、约6%至约65%、约7%至约60%、约8%至约50%、约9%至约45%、约10%至约40%、约11%至约35%、约12%至约

30%、约13%至约25%、约14%至约20%等。可以考虑如上所述范围内的表达的任何子范围。

[0135] 在某些实施方案中,与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞中的这些蛋白质活性相比,改变的革兰氏阳性细菌细胞具有改变的YkfA蛋白质、和/或YkfB蛋白质、和/或YkfC蛋白质、和/或YkfD蛋白质、或它们的任何组合的活性。

[0136] 在一个实施方案中,突变的基因属于肽酶S66超家族。在某些实施方案中,突变的基因是属于肽酶S66超家族的羧肽酶。在某些实施方案中,突变的羧肽酶与由ykf操纵子编码的基因至少60%同源。在某些实施方案中,突变的羧肽酶与由ykfA(示于SEQ ID NO:1)编码的基因至少60%同源。

[0137] 在某些实施方案中,基因改变在ykf操纵子的ykfA基因中。在亲本革兰氏阳性细胞中的ykfA基因(即如本文所述,在被基因改变之前)为与SEQ ID NO:1有至少60%同一性,包括与SEQ ID NO:1有至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%同一性的基因。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(如示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(如示于SEQ ID NO:4)。在某些实施方案中,基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(如示于SEQ ID NO:4)。

[0138] 如所指出的那样,许多不同蛋白质可在革兰氏阳性细胞中用作感兴趣的蛋白质(即其表达在经基因改变的细胞中提高的蛋白质)。感兴趣的蛋白质可为同源蛋白质或异源蛋白质并且可为野生型蛋白质或天然或重组变体。在某些实施方案中,感兴趣的蛋白质是酶,其中在某些例子中,该酶选自蛋白酶、纤维素酶、支链淀粉酶、淀粉酶、糖酶、脂肪酶、异构酶、转移酶、激酶和磷酸酶。在某些实施方案中,感兴趣的蛋白质是蛋白酶,其中蛋白酶可为枯草杆菌蛋白酶,例如,该枯草杆菌蛋白酶选自:枯草杆菌蛋白酶168、枯草杆菌蛋白酶BPN'、枯草菌溶素、枯草杆菌蛋白酶DY、枯草杆菌蛋白酶147、枯草杆菌蛋白酶309、以及它们的变体。在某些实施方案中,感兴趣的蛋白质是荧光蛋白,例如绿荧光蛋白(GFP)。

[0139] 在某些实施方案中,该方法还包括回收该感兴趣的蛋白质。因为感兴趣的蛋白质在经基因改变的革兰氏阳性细胞(相较于野生型或亲本细胞)中的表达/产量的水平提高,所以与在基本相同的培养条件下(并在相同规模下)培养的相应野生型和/或亲本细胞相比,回收的感兴趣的蛋白质的量提高。存在本领域的普通技术人员已知的各种测定法来检测和测定细胞内和细胞外表达的多肽的表达水平/产量。此类测定法由本发明的用户决定并可取决于感兴趣的蛋白质的同一性和/或活性(例如酶活性)。例如,对于蛋白酶,存在基于从酪蛋白或血红蛋白释放的酸溶性肽而在280nm处测为吸光度或通过比色使用Folin方法的测定法(参见例如Bergmeyer等人,“Methods of Enzymatic Analysis”第5卷,Peptidases,Proteinases and their Inhibitors,Verlag Chemie,Weinheim[1984])。其它测定法涉及生色底物的增溶(参见例如Ward,“Proteinases,”in Fogarty(编辑).,

Microbial Enzymes and Biotechnology, Applied Science, London, [1983], 第251-317页)。测定法的其它示例包括琥珀酰-Ala-Ala-Pro-Phe-对硝基酰苯胺测定法(SAAPpNA)和2,4,6-三硝基苯磺酸钠盐测定法(TNBS测定法)。本领域技术人员已知的多个其它参考文献提供适当的方法(参见例如Wells等人, Nucleic Acids Res. 11:7911-7925[1983]; Christianson等人, Anal. Biochem., 223:119-129[1994]; 和Hsia等人, Anal. Biochem., 242:221-227[1999])。

[0140] 也如以上所指出的那样,用于测定感兴趣的蛋白质在宿主细胞中的分泌水平并检测表达的蛋白质的方法包括使用采用对感兴趣的蛋白质特异的多克隆或单克隆抗体的免疫测定法。示例包括酶联免疫吸附测定法(ELISA)、放射免疫测定法(RIA)、荧光免疫测定法(FIA)和荧光激活细胞分选(FACS)。然而,其它方法对于本领域技术人员是已知的并可用于评估感兴趣的蛋白质(参见例如Hampton等人, Serological Methods, A Laboratory Manual, APS Press, St. Paul, MN[1990]; 和Maddox等人, J. Exp. Med., 158:1211[1983])。如在本领域中已知,用本发明产生的改变的芽孢杆菌细胞在适用于在表达并从细胞培养物回收感兴趣的多肽的条件下维持和生长(参见例如Hardwood和Cutting(编辑) Molecular Biological Methods for Bacillus, John Wiley&Sons[1990])。还应注意,如本文所述的经基因改变的细胞可表达多于一种感兴趣的蛋白质,包括两种或更多种、三种或更多种、四种或更多种、五种或更多种、六种或更多种、七种或更多种、八种或更多种、九种或更多种、十种或更多种等。在一些实施方案中,期望细菌分泌蛋白质组中的蛋白质表达提高,其包括由细胞分泌的多种不同蛋白质。

[0141] 本发明的方面包括一种用于获取具有改善的蛋白质产生能力的改变的革兰氏阳性细菌细胞的方法。一般来讲,该方法包括对亲本革兰氏阳性细胞进行基因改变以产生经基因改变的菌株,其中ykf操纵子编码的一个或多个蛋白质的活性改变(如上定义)。

[0142] 在某些实施方案中,该方法包括将多核苷酸序列引入亲本革兰氏阳性细菌细胞中,在整合进染色体或作为附加体遗传因子维持时,产生经基因改变的革兰氏阳性细胞,其中ykf操纵子编码的一个或多个蛋白质的活性改变。

[0143] 已知转化芽孢杆菌菌种的各种方法,从而采用多核苷酸载体(例如质粒构建体)改变芽孢杆菌的染色体或维持芽孢杆菌中的附加体遗传因子,这是公知的。用于将多核苷酸序列引入到芽孢杆菌细胞中的适宜方法见于例如Ferrari等人,“Genetics,” in Harwood等人(编辑), Bacillus, Plenum Publishing Corp. [1989], 第57-72页;还可参见Saunders等人, J. Bacteriol., 157:718-726[1984]; Hoch等人, J. Bacteriol., 93:1925-1937[1967]; Mann等人, Current Microbiol., 13:131-135[1986]; 和Holubova, Folia Microbiol., 30:97[1985]; 对于枯草芽孢杆菌, Chang等人, Mol. Gen. Genet., 168:11-115[1979]; 对于巨大芽孢杆菌, Vorobjeva等人, FEMS Microbiol. Lett., 7:261-263[1980]; 对于解淀粉芽孢杆菌, Smith等人, Appl. Env. Microbiol., 51:634(1986); 对于苏云金芽孢杆菌, Fisher等人, Arch. Microbiol., 139:213-217[1981]; 并且对于球形芽孢杆菌(*B. sphaericus*), McDonald, J. Gen. Microbiol., 130:203[1984]。实际上,诸如转化(包括原生质体转化和中板集合、转导和原生质体融合)之类的方法是已知的并且适用于本发明。转化的方法具体地而言是将本发明提供的DNA构建体引入到宿主细胞中。

[0144] 此外, DNA构建体向宿主细胞的引入包括本领域已知的物理和化学方法来将DNA引



入到宿主细胞中而无需将靶向DNA构建体插入到质粒或载体中。此类方法包括但不限于氯化钙沉淀、电穿孔、裸DNA、脂质体等。在另外的实施方案中，DNA构建体可与质粒共转化，而无需插入质粒中。

[0145] 在其中可选标记基因用于选择稳定的转化体的一些实施方案中，可期望使用任何常规方法从经基因改变的革兰氏阳性菌株中除去选择性标记，其中许多方法在本领域中是已知的（参见Stahl等人，*J. Bacteriol.*, 158:411-418[1984]；和Palmeros等人，*Gene* 247: 255-264[2000]）。

[0146] 在一些实施方案中，将两个或更多个DNA构建体（即各自设计用于基因改变宿主细胞的DNA构建体）引入到亲本革兰氏阳性细胞中，从而导致将两个或更多个基因改变引入细胞中，例如在两个或更多个染色体区域的改变。在一些实施方案中，这些区域是连续的（例如在ykf操纵子内或在ykf操纵子和相邻基因或操纵子内的两个区域），而在其它实施方案中，区域是分开的。在一些实施方案中，一个或多个基因改变通过添加附加体遗传因子实现。

[0147] 在一些实施方案中，用一个或多个根据本发明的DNA构建体转化宿主细胞以产生改变的芽孢杆菌菌株，其中两个或更多个基因在宿主细胞中已失活。在一些实施方案中，宿主细胞染色体中缺失两个或更多个基因。在另选的实施方案中，通过插入DNA构建体而使两个或更多个基因失活。在一些实施方案中，失活基因是连续的（无论是通过缺失和/或插入而失活），而在其它实施方案中，它们不是连续的基因。

[0148] 一旦产生经基因改变的宿主细胞，就可在使得感兴趣的蛋白质表达的条件下对其进行培养，而在某些实施方案中，对感兴趣的蛋白质进行回收。

[0149] 本发明的方面包括改变的革兰氏阳性细菌细胞，其中与在基本相同的培养条件下生长的相应未改变的革兰氏阳性细菌细胞相比，改变的革兰氏阳性细菌细胞包含至少一个基因改变，所述基因改变使ykf操纵子所编码的一个或多个蛋白质的活性改变。在一些实施方案中，经基因改变的革兰氏阳性细胞如上文所述产生。进一步由上可见，改变的革兰氏阳性细菌细胞可为芽孢杆菌属菌种的菌株，例如地衣芽孢杆菌、迟缓芽孢杆菌、枯草芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、灿烂芽孢杆菌、克劳氏芽孢杆菌、巨大芽孢杆菌或苏云金芽孢杆菌菌株。在某些实施方案中，芽孢杆菌属菌种的菌株为枯草芽孢杆菌菌株。在一些方面，改变的革兰氏阳性细菌细胞还包含改善细胞表型的附加突变，例如在选自degU、degQ、degS、scoC4、spoIIE和oppA的基因中的突变。在某些实施方案中，突变为degU (Hy) 32。

[0150] 在一些实施方案中，本发明包括包含引入序列的DNA构建体，该引入序列在稳定掺入宿主细胞时会使细胞发生基因改变，使得ykf操纵子所编码的一个或多个蛋白质的活性改变（如以下所详述）。在一些实施方案中，DNA构建体在体外组装，之后将构建体直接克隆到感受态革兰氏阳性（例如芽孢杆菌）宿主中，使得DNA构建体整合进宿主细胞染色体中。例如，PCR融合和/或连接可用于在体外组装DNA构建体。在一些实施方案中，DNA构建体为非质粒构建体，而在其它实施方案中，将其掺入载体（例如质粒）中。在一些实施方案中，使用环状质粒。在一些实施方案中，环状质粒设计成使用适当的限制性酶（即不破坏DNA构建体的限制性酶）。因此，线性质粒可用于本发明。然而，其它方法适用于本发明，如本领域技术人员所知的（参见例如，Perego, “Integrational Vectors for Genetic Manipulation in

*Bacillus subtilis*,”in (Sonenshein等人(编辑), *Bacillus subtilis and Other Gram-Positive Bacteria*, American Society for Microbiology, Washington, DC[1993])。

[0151] 在某些实施方案中, DNA靶向载体的引入序列包括包含来源于ykfA基因的变体序列的多核苷酸。在这些实施方案的一些中, 变体序列的长度为至少约15个核苷酸, 与全部或部分SEQ ID NO:1有至少60%同一性, 并且在ykfA基因的核苷酸位置具有至少一个突变, 从而当突变存在于革兰氏阳性细菌细胞的内源性ykfA基因中时导致ykf操纵子编码的蛋白质的活性改变。变体序列可为至少约20个核苷酸、约30个核苷酸、约40个核苷酸、约50个核苷酸、约60个核苷酸、约80个核苷酸、约90个核苷酸、约100个核苷酸、约200个核苷酸、约300个核苷酸、约400个核苷酸、约500个核苷酸、约600个核苷酸、约700个核苷酸、约800个核苷酸、约900个核苷酸、约1000个核苷酸、约1100个核苷酸、约1200个核苷酸、约1300个核苷酸、约1400个或更多个核苷酸。如以上进一步指出, 变体序列可与SEQ ID NO:1有至少60%同一性, 包括与SEQ ID NO:1有至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%同一性。在某些实施方案中, 基因改变导致与SEQ ID NO:2的第252或253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中, 基因改变导致与SEQ ID NO:2的第252和253位氨基酸对应的位置处的氨基酸改变。在某些实施方案中, 基因改变导致与SEQ ID NO:2的第252位氨基酸对应的位置处的氨基酸由P变为L(如示于SEQ ID NO:4)。在某些实施方案中, 基因改变导致与SEQ ID NO:2的第253位氨基酸对应的位置处的氨基酸由V变为L(如示于SEQ ID NO:4)。在某些实施方案中, 基因改变导致与SEQ ID NO:2的第252位氨基酸对应的氨基酸位置处P变为L以及与SEQ ID NO:2的第253位氨基酸对应的氨基酸位置处V变为L(如示于SEQ ID NO:4)。

[0152] 本发明的方面包括包含如上所述的多核苷酸序列的载体。在某些实施方案中, 载体是靶向载体, 所述靶向载体被设计成在转化到革兰氏阳性细菌细胞中时通过同源重组将多核苷酸序列中的至少一个突变引入到革兰氏阳性细菌细胞的ykf操纵子中的对应位置。在一些实施方案中, 引入序列/载体包含选择性标记。在一些实施方案中, 选择性标记位于两个loxP位点之间(参见Kuhn和Torres, *Meth. Mol. Biol.*, 180:175-204[2002]), 并且然后抗微生物基因通过Cre蛋白质的作用而缺失。

[0153] 本发明的方面包括一种用于增强革兰氏阳性细菌细胞中感兴趣的蛋白质的表达的方法, 该方法包括用以上所述DNA构建体或载体(即包含引入序列的DNA构建体或载体, 该引入序列在稳定掺入宿主细胞时会使细胞发生基因改变, 使得ykf操纵子所编码的一个或多个蛋白质的活性改变, 例如在如上所述ykfA基因中包含突变的DNA构建体或载体)转化亲本革兰氏阳性细菌细胞, 从而允许载体和亲本革兰氏阳性细菌细胞的ykf操纵子中的相应区域发生同源重组以产生改变的革兰氏阳性细菌细胞; 并使改变的革兰氏阳性细菌细胞在适于表达感兴趣的蛋白质的条件下生长, 其中与在转化步骤之前的革兰氏阳性细菌细胞相比, 在改变的革兰氏阳性细菌细胞中产生的感兴趣的蛋白质提高。可用于本发明的该方面的革兰氏阳性菌株、突变及其它特征的示例如上所详述。

[0154] 无论DNA构建体掺入载体中或在不存在质粒DNA的情况下使用, 其均用于转化微生物。预期用于转化的任何适宜方法可用于本发明。在一些实施方案中, 将DNA构建体的至少一个拷贝整合进宿主芽孢杆菌染色体中。在一些实施方案中, 使用本发明的一个或多个DNA

构建体来转化宿主细胞。

[0155] 本领域技术人员参照以下实施例,可以更充分理解实现本发明的方式和方法,这些实施例并非意在以任何方式限制本发明或其权利要求书指定的范围。

#### [0156] 实验

[0157] 提供下面的实施例,从而证明和进一步阐述本发明的某些实施方案和方面,但不应该理解为限制其范围。

[0158] 在随后的实验公开中,应用以下缩写中的某些:°C (摄氏度);rpm (转/分钟);μg (微克);mg (毫克);μl (微升);ml (毫升);mM (毫摩尔);μM (微摩尔);sec (秒);min (s) (分钟);hr (s) (小时);OD<sub>280</sub> (280nm处光密度);OD<sub>600</sub> (600nm处光密度);PCR (聚合酶链反应);RT-PCR (逆转录PCR);SDS (十二烷基硫酸钠)。

#### [0159] 实施例1

##### [0160] ykfA基因的突变对芽孢杆菌菌种中蛋白质表达的影响

[0161] 枯草芽孢杆菌的ykfA基因是参与肽聚糖的循环利用的ykf操纵子(图1)的第一编码序列。YkfA是切割L-和D-氨基酸之间酰胺键的LD-羧肽酶,其天然存在于细菌肽聚糖中。

[0162] 已鉴定出三种单核苷酸多态性,其导致芽孢杆菌菌株的ykfA基因的两种非同义突变(P252L和V253L)。使用Janes和Stibitz (Infection and Immunity, 74 (3):1949, 2006)所述的方法将ykfA突变引入适宜的芽孢杆菌菌株CB15-14 (amyE::xylRPxylAcomK-ermC, ΔoppA, ΔspoIIE, ΔaprE, ΔnprE, degUHy32, ΔscoC) 中。

[0163] 为了测试ykfA突变对FNA蛋白酶(包含Y217L取代的枯草杆菌蛋白酶BPN')表达的影响,将PaprE-FNA catR构建体引入CB15-14和CB15-14ykfA突变株的aprE基因座中,并在含有25μg/ml氯霉素的Luria琼脂平板上对构建体进行扩增。使ykfA突变的菌株和野生型菌株在5mL的Luria肉汤培养基(Luria broth medium)中生长过夜。在250rpm下用1ml的预培养物在Thompson烧瓶中接种25ml的2XNB (2X营养肉汤, 1XSNB盐, 如在W02010/14483中有所描述), 以测试蛋白酶表达。使用SpectraMax分光光度计(Molecular Devices, Downington, PA, USA) 以每小时间隔在600nm处测定稀释20X的整个液体培养基的细胞密度。使600nm处的吸光度作为时间的函数进行作图, 并且结果示于图2A。ykfA突变的存在导致2XSNB培养基中细胞的生长更高。

[0164] 用N-suc-AAPF-pNA底物(得自Sigma Chemical Co.)来监测蛋白酶表达, 如在W02010/144283中有所描述。简而言之, 整个液体培养基用测定缓冲液(100mM Tris, 0.005% Tween 80, pH 8.6) 稀释40X, 并使10μl的经稀释样品排列在微量滴定板中。对AAPF原液进行稀释并将测定缓冲液(用DMSO将100mg/ml AAPF原液稀释100X) 和190μl的该溶液添加至微量滴定板中, 并使用SpectraMax分光光度计(Molecular Devices, Downington, PA, USA) 在405nm处测定溶液的吸光度。使405nm处的吸光度作为时间的函数进行作图, 并且结果示于图2B。FNA的产量在较后的时间点(8小时)提高, 这是由于ykfA突变。

#### [0165] 实施例2

##### [0166] ykfA突变对芽孢杆菌菌种中绿荧光蛋白表达的影响

[0167] 为了测试ykfA突变对其它蛋白质表达的影响, 将PaprE-GFP catR构建体引入CB15-14和CB15-14ykfA突变株的aprE基因座中, 并在含有5μg/ml氯霉素的Luria琼脂平板上选择转化体。使ykfA突变的菌株和野生型菌株在5mL的Luria肉汤中生长过夜。在37°C和

250rpm下用1ml的预培养物在摇瓶中接种25ml的2XNB培养基(2X营养肉汤,1X SNB盐),以测试绿色荧光蛋白(GFP)的表达。使用SpectraMax分光光度计(Molecular Devices, Downington,PA,USA)以每小时间隔在600nm处测定稀释20X的整个液体培养基的细胞密度。使600nm处的吸光度作为时间的函数进行作图,并且结果示于图3A。在进入稳定期(在2XNB中生长的4至6小时之间)时,ykfA突变株的细胞生长的下降相较于对照菌株有延缓,这表明改善的细胞生存力归因于ykfA突变。

[0168] 为了测定GFP表达,将100 $\mu$ l的培养物转移至96孔微量滴定板并在荧光读板机(采用485nm的激发波长,508nm的发射波长,具有495nm发射截止滤波器)中测定GFP表达。使485/508nm处的相对荧光单位(RFU)作为时间的函数进行作图,并且结果示于图3B。GFP的产量从生长6小时提高,这是由于ykfA突变。

### [0169] 实施例3

#### [0170] ykfA突变对芽孢杆菌菌种中 $\beta$ -D-葡萄糖苷酶表达的影响

[0171] 为了测试ykfA突变对 $\beta$ -D-葡萄糖苷酶表达的影响,将PaprE-BglC catR构建体引入CB15-14和CB15-14ykfA突变株的aprE基因座中,并在含有5 $\mu$ g/ml氯霉素的Luria琼脂平板上选择转化体。使ykfA突变的菌株和野生型菌株在5mL的Luria肉汤中生长过夜。在37 $^{\circ}$ C和250rpm下用1ml的预培养物在摇瓶中接种25ml的2XNB培养基(2X营养肉汤,1X SNB盐),以测试分泌的 $\beta$ -D-葡萄糖苷酶的表达。使用SpectraMax分光光度计(Molecular Devices, Downington,PA,USA)以每小时间隔在600nm处测定稀释20X的整个液体培养基的细胞密度。使600nm处的吸光度作为时间的函数进行作图,并且结果示于图4A。包含ykfA突变的细胞具有更高的细胞生长,这表明由于存在ykfA突变而使细胞生存力改善。

[0172] 使用4-硝基苯基- $\beta$ -D-纤维二糖底物(Sigma Chemicals,St.Louis,MO,USA,Cat.# N57590)来监测 $\beta$ -D-葡萄糖苷酶表达。将底物溶解于1ml的DMSO中以制成100mg/ml的原液。通过用10ml的测定缓冲液(100mM Tris,0.005%Tween 80,pH 8.6)稀释35 $\mu$ l的原液来制备底物工作溶液。将40微升的各培养物转移至96孔微量滴定板中并向各个孔添加180 $\mu$ l的底物工作溶液。在室温下将微量滴定板温育5小时,并且在温育期结束时,用SpectraMax分光光度计(Molecular Devices,Downington,PA,USA)在405nm处测定溶液的吸光度。使405nm处的吸光度作为时间的函数进行作图,并且结果示于图4B。BglC的产量在所有时间点提高,这是由于存在ykfA突变。

[0173] 根据上述数据,显而易见的是,当在相同或基本相同的培养条件下培养时,革兰氏阳性细菌细胞(即相较于亲本细胞)中ykf操纵子(例如ykfA基因)所编码的蛋白质的活性的改变导致感兴趣的蛋白质的表达相较于亲本细胞而言提高。

[0174] 虽然出于清楚地理解的目的通过说明和实施例对前述组合物和方法进行了相当详细的描述,但根据本文的教导,对于本领域的普通技术人员而言显而易见的是,在不脱离所附权利要求书的精神或范围的情况下,可以对前述组合物和方法进行某些改变和修改。

[0175] 因此,以上仅仅示例了本发明组合物和方法的原理。应当理解,本领域技术人员能够设计出各种布置,尽管这些布置在本文中未明确说明或示出,但它们仍体现了本发明组合物和方法的原理并且包括在其精神和范围内。此外,本文引用的所有实施例和条件语言原则上旨在辅助读者理解本发明的组合物和方法的原理及发明者为改进现有技术所贡献的概念,并且应当理解为不局限于这些具体引用的实施例和条件。此外,本文阐述的本发明

的组合物和方法的原理、方面和实施方案以及其具体实施例的所有表述旨在涵盖其结构和功能的等同物。此外,希望此类等同物包括当前已知的等同物以及未来开发的等同物,即所开发的任何执行相同功能的要素,无论结构如何。因此本发明的组合物和方法的范围并非旨在限于本文所示和所述的实施方案。

[0176] 序列

[0177] SEQ ID NO:1—ykfA野生型核苷酸序列

[0178] atgaaaggagtgttttcgttgaattacaagccgaaagcgttgaacaagggtgatacagtcggagtgat  
cgcgcccgaagtccgccggatccaaaaagccttgacaccgcgctttttatttttagaagagctcggtcttcaggtg  
aagttgggcaaggcgctgaaaaaccagcacggctatttagcgggacaggatgatgagcggctggcggatctccatg  
agatgttcagagacgatgaggtaaaagcagtgttgtgcgcatgcgggggttttgggacaggacgtatcgccgctggg  
cattgatttcagcttgatccgcaaacacctaaaatcttttggggatacagcgatattacgtttttacatactgcc  
attcatcaaaacacaggtcttgtcactttccatggcccgatgctcagcacggatattggccttgacgacgttcacc  
cgctgacaaaagcgtcatataagcagctcttccaggagacggaattcacctatacagaagagctttctccgctgac  
cgagcttgttcttgaaaagcggaaggcgagcttgtcgggggaatctgtctttgctgacgtctacactgggcacg  
ccatttgaaattgatacagagaggaaagcttctgtttattgaagatattgacgaggagccttatcaaatcgaccgga  
tgctgaatcagctgaaaatgggggggaagctgacggacgcggcggggaattctagtttgtgattttcacaattgtgt  
cccgtgaaagcgagagaagctctctctcgcttgagcaggtgctggaagactatattttctgcgggcaggcctgct  
ctgagaggatttaaaatcgccactgctcgccaagtattgccgttccgatcggtgcgaaagctgctatgaatacag  
cagaaaaaacagccgtaatagaggcgggcgtttcagaaggggcgctgaagacatga

[0179] SEQ ID NO:2—YkfA野生型蛋白质序列

[0180] MKGVFSLNYKPKALNKGDTVGVIPASPPDPKKLDTALLFLEELGLQVKLGKALKNQHGYLEAGQDDERL  
ADLHEMFRDDEVKAVLCACGGFGTGRIAAGIDFSLIRKHPKIFWGYSDITFLHTAIHQNTGLVTFHGPMPLSTDIGLD  
DVHPLTKASYKQLFQETEFYTEELSPLTELVPKAEGLVGGNLSLLTSTLGTPFEIDTRGKLLFIEDIDEEPYQI  
DRMLNQLKMGGKLTDAAGILVCDFHNCVPVKREKSLSEQVLEDYIIISAGRPALRGFKIGHCSPSIAVPIGAKAAMN  
TAEKTAVIEAGVSEGALKT

[0181] SEQ ID NO:3—ykfA突变核苷酸序列

[0182] atgaaaggagtgttttcgttgaattacaagccgaaagcgttgaacaagggtgatacagtcggagtgat  
cgcgcccgaagtccgccggatccaaaaagccttgacaccgcgctttttatttttagaagagctcggtcttcaggtg  
aagttgggcaaggcgctgaaaaaccagcacggctatttagcgggacaggatgatgagcggctggcggatctccatg  
agatgttcagagacgatgaggtaaaagcagtgttgtgcgcatgcgggggttttgggacaggacgtatcgccgctggg  
cattgatttcagcttgatccgcaaacacctaaaatcttttggggatacagcgatattacgtttttacatactgcc  
attcatcaaaacacaggtcttgtcactttccatggcccgatgctcagcacggatattggccttgacgacgttcacc  
cgctgacaaaagcgtcatataagcagctcttccaggagacggaattcacctatacagaagagctttctccgctgac  
cgagcttgttcttgaaaagcggaaggcgagcttgtcgggggaatctgtctttgctgacgtctacactgggcacg  
ccatttgaaattgatacagagaggaaagcttctgtttattgaagatattgacgaggagccttatcaaatcgaccgga  
tgctgaatcagctgaaaatgggggggaagctgacggacgcggcggggaattctagtttgtgattttcacaattgtgt  
cctgctcaagcgagagaagctctctctcgcttgagcaggtgctggaagactatattttctgcgggcaggcctgct  
ctgagaggatttaaaatcgccactgctcgccaagtattgccgttccgatcggtgcgaaagctgctatgaatacag  
cagaaaaaacagccgtaatagaggcgggcgtttcagaaggggcgctgaagacatga

[0183] SEQ ID NO:4:YkfA突变蛋白质序列(具有P252L和V253L改变)

[0184] MKGVFSLNYKPKALNKGDTVGVIPASPPDPKKLDTALLFLEELGLQVKLGKALKNQHGYPAGQDDERL  
ADLHEMFRDDEVKAVLCACGGFGTGRIAAGIDFSLIRKHPKIFWGYSDITFLHTAIHQNTGLVTFHGPMSTDIGLD  
DVHPLTKASYKQLFQETEFYTEELSPLTELVPKAEGLVGGNLSLLTSTLGTPFEIDTRGKLLFIEDIDEEPYQI  
DRMLNQLKMGKLTDAAGILVCFHNCVLLKREKSLSLEQVLEDYIISAGRPALRGFKIGHCSPSIAPVIGAKAAMN  
TAEKTAVIEAGVSEGALKT

[0185] SEQ ID NO:5:FNA蛋白质序列(具有原结构域)

[0186] A Q S V P Y G V S Q I K A P A L H S Q G Y T G S N V K V A V I D S G I  
D S S H P D L K V A G G A S M V P S E T N P F Q D N N S H G T H V A G T V A A  
L N N S I G V L G V A P S A S L Y A V K V L G A D G S G Q Y S W I I N G I E W  
A I A N N M D V I N M S L G G P S G S A A L K A A V D K A V A S G V V V V A A  
A G N E G T S G S S S T V G Y P G K Y P S V I A V G A V D S S N Q R A S F S S  
V G P E L D V M A P G V S I Q S T L P G N K Y G A L N G T S M A S P H V A G A  
A A L I L S K H P N W T N T Q V R S S L E N T T T K L G D S F Y Y G K G L I N  
V Q A A A Q

[0187] SEQ ID NO:6:GFP蛋白质序列

[0188] V N R N V L K N T G L K E I M S A K A S V E G I V N N H V F S M E G F  
G K G N V L F G N Q L M Q I R V T K G G P L P F A F D I V S I A F Q Y G N R T  
F T K Y P D D I A D Y F V Q S F P A G F F Y E R N L R F E D G A I V D I R S D  
I S L E D D K F H Y K V E Y R G N G F P S N G P V M Q K A I L G M E P S F E V  
V Y M N S G V L V G E V D L V Y K L E S G N Y Y S C H M K T F Y R S K G G V K  
E F P E Y H F I H H R L E K T Y V E E G S F V E Q H E T A I A Q L T T I G K P  
L G S L H E W V

[0189] SEQ ID NO:7:Bgl1C蛋白质序列

[0190] A A G T K T P V A K N G Q L S I K G T Q L V N R D G K A V Q L K G I S  
S H G L Q W Y G E Y V N K D S L K W L R D D W G I T V F R A A M Y T A D G G Y  
I D N P S V K N K V K E A V E A A K E L G I Y V I I D W H I L N D G N P N Q N  
K E K A K E F F K E M S S L Y G N T P N V I Y E I A N E P N G D V N W K R D I  
K P Y A E E V I S V I R K N D P D N I I I V G T G T W S Q D V N D A A D D Q L  
K D A N V M Y A L H F Y A G T H G Q F L R D K A N Y A L S K G A P I F V T E W  
G T S D A S G N G G V F L D Q S R E W L K Y L D S K T I S W V N W N L S D K Q  
E S S S A L K P G A S K T G G W R L S D L S A S G T F V R E N I L G T K D S T  
K D I P E T P S K D K P T Q E N G I S V Q Y R A G D G S M N S N Q I R P Q L Q  
I K N N G N T T V D L K D V T A R Y W Y K A K N K G Q N F D C D Y A Q I G C G  
N V T H K F V T L H K P K Q G A D T Y L E L G F K N G T L A P G A S T G N I Q  
L R L H N D D W S N Y A Q S G D Y S F F K S N T F K T T K K I T L Y D Q G K L  
I W G T E P N

## 序列表

<110> 丹尼斯科美国公司  
 <120> 增强的蛋白质表达  
 <130> 40354-WO-PCT  
 <140> PCT/US14/69019  
 <141> 2014-12-08  
 <150> US 61/922,613  
 <151> 2013-12-31  
 <160> 7  
 <170> PatentIn版本3.5  
 <210> 1  
 <211> 960  
 <212> DNA  
 <213> 枯草芽孢杆菌 (Bacillus subtilis)

<220>  
 <221> 尚未归类的特征  
 <223> ykfA野生型核苷酸序列

[0001]

<400> 1  
 atgaaaggag tgttttcgtt gaattacaag ccgaaagcgt tgaacaaggg tgatacagtc 60  
 ggagtgatcg cgcccgcaag tccgccgat ccaaaaaagc ttgacaccgc gcttttattt 120  
 ttagaagagc tcggtcttca ggtgaagttg ggcaaggcgc tgaanaacca gcacggctat 180  
 ttagcgggac aggatgatga gcggctggcg gatctccatg agatgttcag agacgatgag 240  
 gtaaaagcag tgttgtgcgc atgcgggggt ttggggacag gacgtatcgc cgcgggcatt 300  
 gatttcagct tgatccgcaa acaccctaaa atcttttggg gatacagcga tattacgttt 360  
 ttacatactg ccattcatca aaacacaggt cttgtcactt tccatggccc gatgctcage 420  
 acggatatcg gccttgacga cgttcacccc ctgacaaaag cgtcatataa gcagctcttc 480  
 caggagacgg aattcaccta tacagaagag ctttctccgc tgaccgagct tgttcctgga 540  
 aaagcggaag gcgagcttgt cgggggaaat ctgtctttgc tgacgtctac actgggcacg 600  
 ccatttgaaa ttgatacgag aggaaagcct ctgtttattg aagatattga cgaggagcct 660  
 tatcaaatcg accggatgct gaatcagctg aaaatggggg ggaagctgac ggacgcggcg 720  
 ggaattctag ttgtgatatt tcacaattgt gtcccgtgta agcgagagaa gtctctctcg 780  
 cttgagcagg tgctggaaga ctatattatt tctgcgggca ggcctgctct gagaggattt 840  
 aaaatcgccc actgctcgcc aagtattgcc gttccgatcg gtgcgaaagc tgctatgaat 900  
 acagcagaaa aaacagccgt aatagaggcg ggcgtttcag aaggggcgct gaagacatga 960

<210> 2  
 <211> 319  
 <212> PRT  
 <213> 枯草芽孢杆菌 (Bacillus subtilis)

<220>  
 <221> 尚未归类的特征  
 <223> YkfA野生型蛋白质序列

<400> 2

Met Lys Gly Val Phe Ser Leu Asn Tyr Lys Pro Lys Ala Leu Asn Lys  
 1 5 10 15

Gly	Asp	Thr	Val	Gly	Val	Ile	Ala	Pro	Ala	Ser	Pro	Pro	Asp	Pro	Lys
			20					25					30		
Lys	Leu	Asp	Thr	Ala	Leu	Leu	Phe	Leu	Glu	Glu	Leu	Gly	Leu	Gln	Val
		35					40					45			
Lys	Leu	Gly	Lys	Ala	Leu	Lys	Asn	Gln	His	Gly	Tyr	Leu	Ala	Gly	Gln
	50					55					60				
Asp	Asp	Glu	Arg	Leu	Ala	Asp	Leu	His	Glu	Met	Phe	Arg	Asp	Asp	Glu
65					70					75					80
Val	Lys	Ala	Val	Leu	Cys	Ala	Cys	Gly	Gly	Phe	Gly	Thr	Gly	Arg	Ile
				85					90					95	
Ala	Ala	Gly	Ile	Asp	Phe	Ser	Leu	Ile	Arg	Lys	His	Pro	Lys	Ile	Phe
			100					105					110		
Trp	Gly	Tyr	Ser	Asp	Ile	Thr	Phe	Leu	His	Thr	Ala	Ile	His	Gln	Asn
		115					120					125			
Thr	Gly	Leu	Val	Thr	Phe	His	Gly	Pro	Met	Leu	Ser	Thr	Asp	Ile	Gly
	130					135					140				
Leu	Asp	Asp	Val	His	Pro	Leu	Thr	Lys	Ala	Ser	Tyr	Lys	Gln	Leu	Phe
145					150					155					160
[0002]	Gln	Glu	Thr	Glu	Phe	Thr	Tyr	Thr	Glu	Glu	Leu	Ser	Pro	Leu	Thr
				165					170						175
Leu	Val	Pro	Gly	Lys	Ala	Glu	Gly	Glu	Leu	Val	Gly	Gly	Asn	Leu	Ser
			180					185					190		
Leu	Leu	Thr	Ser	Thr	Leu	Gly	Thr	Pro	Phe	Glu	Ile	Asp	Thr	Arg	Gly
		195					200					205			
Lys	Leu	Leu	Phe	Ile	Glu	Asp	Ile	Asp	Glu	Glu	Pro	Tyr	Gln	Ile	Asp
	210					215					220				
Arg	Met	Leu	Asn	Gln	Leu	Lys	Met	Gly	Gly	Lys	Leu	Thr	Asp	Ala	Ala
225					230					235					240
Gly	Ile	Leu	Val	Cys	Asp	Phe	His	Asn	Cys	Val	Pro	Val	Lys	Arg	Glu
				245					250					255	
Lys	Ser	Leu	Ser	Leu	Glu	Gln	Val	Leu	Glu	Asp	Tyr	Ile	Ile	Ser	Ala
			260					265					270		
Gly	Arg	Pro	Ala	Leu	Arg	Gly	Phe	Lys	Ile	Gly	His	Cys	Ser	Pro	Ser
		275					280					285			
Ile	Ala	Val	Pro	Ile	Gly	Ala	Lys	Ala	Ala	Met	Asn	Thr	Ala	Glu	Lys
	290					295					300				
Thr	Ala	Val	Ile	Glu	Ala	Gly	Val	Ser	Glu	Gly	Ala	Leu	Lys	Thr	
305					310					315					



<210>	3	
<211>	960	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成: ykfA突变核苷酸序列	
<400>	3	
	atgaaaggag tgttttcgtt gaattacaag ccgaaagcgt tgaacaaggg tgatacagtc	60
	ggagtgatcg cgcccgcaag tccgccgat ccaaaaaagc ttgacaccgc gcttttattt	120
	ttagaagagc tcggtcttca ggtgaagttg ggcaaggcgc tgaaaaacca gcacggctat	180
	ttagcgggac aggatgatga gcggctggcg gatctccatg agatgttcag agacgatgag	240
	gtaaaagcag tgttgtgcgc atgcgggggt tttgggacag gacgtatcgc cgcgggcatt	300
	gatttcagct tgatccgcaa acaccctaaa atcttttggg gatacagcga tattacgttt	360
	ttacatactg ccattcatca aaacacaggt ctgttcactt tccatggccc gatgctcagc	420
	acggatatitg gccttgacga cgttcacccg ctgacaaaag cgtcatataa gcagctcttc	480
	caggagacgg aattcaccta tacagaagag ctttctccgc tgaccgagct tgttctctga	540
	aaagcggaag gcgagcttgt cgggggaaat ctgtctttgc tgacgtctac actgggcacg	600
	ccatttgaaa ttgatacagag aggaaagctt ctgtttattg aagatattga cgaggagcct	660
	tatcaaatcg accggatgct gaatcagctg aaaatggggg ggaagctgac ggacgcggcg	720
	ggaattctag tttgtgattt tcacaattgt gtcctgctca agcgagagaa gtctctctcg	780
	cttgagcagg tgctggaaga ctatattatt tctgcgggca ggcctgctct gagaggattt	840
[0003]	aaaatcggcc actgctcgcc aagtattgcc gttccgatcg gtgcgaaagc tgctatgaat	900
	acagcagaaa aaacagccgt aatagaggcg ggcgtttcag aaggggcgct gaagacatga	960
<210>	4	
<211>	319	
<212>	PRT	
<213>	人工序列	
<220>		
<223>	合成: YkfA突变蛋白质序列 (具有P252L和V253L改变)	
<400>	4	
	Met Lys Gly Val Phe Ser Leu Asn Tyr Lys Pro Lys Ala Leu Asn Lys	
	1 5 10 15	
	Gly Asp Thr Val Gly Val Ile Ala Pro Ala Ser Pro Pro Asp Pro Lys	
	20 25 30	
	Lys Leu Asp Thr Ala Leu Leu Phe Leu Glu Glu Leu Gly Leu Gln Val	
	35 40 45	
	Lys Leu Gly Lys Ala Leu Lys Asn Gln His Gly Tyr Leu Ala Gly Gln	
	50 55 60	
	Asp Asp Glu Arg Leu Ala Asp Leu His Glu Met Phe Arg Asp Asp Glu	
	65 70 75 80	
	Val Lys Ala Val Leu Cys Ala Cys Gly Gly Phe Gly Thr Gly Arg Ile	
	85 90 95	

Ala Ala Gly Ile Asp Phe Ser Leu Ile Arg Lys His Pro Lys Ile Phe  
100 105 110

Trp Gly Tyr Ser Asp Ile Thr Phe Leu His Thr Ala Ile His Gln Asn  
115 120 125

Thr Gly Leu Val Thr Phe His Gly Pro Met Leu Ser Thr Asp Ile Gly  
130 135 140

Leu Asp Asp Val His Pro Leu Thr Lys Ala Ser Tyr Lys Gln Leu Phe  
145 150 155 160

Gln Glu Thr Glu Phe Thr Tyr Thr Glu Glu Leu Ser Pro Leu Thr Glu  
165 170 175

Leu Val Pro Gly Lys Ala Glu Gly Glu Leu Val Gly Gly Asn Leu Ser  
180 185 190

Leu Leu Thr Ser Thr Leu Gly Thr Pro Phe Glu Ile Asp Thr Arg Gly  
195 200 205

Lys Leu Leu Phe Ile Glu Asp Ile Asp Glu Glu Pro Tyr Gln Ile Asp  
210 215 220

Arg Met Leu Asn Gln Leu Lys Met Gly Gly Lys Leu Thr Asp Ala Ala  
225 230 235 240

[0004] Gly Ile Leu Val Cys Asp Phe His Asn Cys Val Leu Leu Lys Arg Glu  
245 250 255

Lys Ser Leu Ser Leu Glu Gln Val Leu Glu Asp Tyr Ile Ile Ser Ala  
260 265 270

Gly Arg Pro Ala Leu Arg Gly Phe Lys Ile Gly His Cys Ser Pro Ser  
275 280 285

Ile Ala Val Pro Ile Gly Ala Lys Ala Ala Met Asn Thr Ala Glu Lys  
290 295 300

Thr Ala Val Ile Glu Ala Gly Val Ser Glu Gly Ala Leu Lys Thr  
305 310 315

<210> 5  
<211> 275  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成: FNA蛋白质序列 (具有原结构域)

<400> 5

Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu  
1 5 10 15

His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp  
20 25 30

[0005]

Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala  
 35 40 45  
 Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn Asn Ser His  
 50 55 60  
 Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Asn Asn Ser Ile Gly  
 65 70 75 80  
 Val Leu Gly Val Ala Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu  
 85 90 95  
 Gly Ala Asp Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu  
 100 105 110  
 Trp Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met Ser Leu Gly Gly  
 115 120 125  
 Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala  
 130 135 140  
 Ser Gly Val Val Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly  
 145 150 155 160  
 Ser Ser Ser Thr Val Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala  
 165 170 175  
 Val Gly Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val  
 180 185 190  
 Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr  
 195 200 205  
 Leu Pro Gly Asn Lys Tyr Gly Ala Leu Asn Gly Thr Ser Met Ala Ser  
 210 215 220  
 Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn  
 225 230 235 240  
 Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Thr Lys  
 245 250 255  
 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala  
 260 265 270  
 Ala Ala Gln  
 275  
 <210> 6  
 <211> 238  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> 合成: GFP蛋白质序列  
 <400> 6  
 Val Asn Arg Asn Val Leu Lys Asn Thr Gly Leu Lys Glu Ile Met Ser  
 1 5 10 15

Ala Lys Ala Ser Val Glu Gly Ile Val Asn Asn His Val Phe Ser Met  
 20 25 30  
 Glu Gly Phe Gly Lys Gly Asn Val Leu Phe Gly Asn Gln Leu Met Gln  
 35 40 45  
 Ile Arg Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Val  
 50 55 60  
 Ser Ile Ala Phe Gln Tyr Gly Asn Arg Thr Phe Thr Lys Tyr Pro Asp  
 65 70 75 80  
 Asp Ile Ala Asp Tyr Phe Val Gln Ser Phe Pro Ala Gly Phe Phe Tyr  
 85 90 95  
 Glu Arg Asn Leu Arg Phe Glu Asp Gly Ala Ile Val Asp Ile Arg Ser  
 100 105 110  
 Asp Ile Ser Leu Glu Asp Asp Lys Phe His Tyr Lys Val Glu Tyr Arg  
 115 120 125  
 Gly Asn Gly Phe Pro Ser Asn Gly Pro Val Met Gln Lys Ala Ile Leu  
 130 135 140  
 Gly Met Glu Pro Ser Phe Glu Val Val Tyr Met Asn Ser Gly Val Leu  
 145 150 155 160  
 [0006] Val Gly Glu Val Asp Leu Val Tyr Lys Leu Glu Ser Gly Asn Tyr Tyr  
 165 170 175  
 Ser Cys His Met Lys Thr Phe Tyr Arg Ser Lys Gly Gly Val Lys Glu  
 180 185 190  
 Phe Pro Glu Tyr His Phe Ile His His Arg Leu Glu Lys Thr Tyr Val  
 195 200 205  
 Glu Glu Gly Ser Phe Val Glu Gln His Glu Thr Ala Ile Ala Gln Leu  
 210 215 220  
 Thr Thr Ile Gly Lys Pro Leu Gly Ser Leu His Glu Trp Val  
 225 230 235  
 <210> 7  
 <211> 471  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> 合成: BglC蛋白质序列  
 <400> 7  
 Ala Ala Gly Thr Lys Thr Pro Val Ala Lys Asn Gly Gln Leu Ser Ile  
 1 5 10 15  
 Lys Gly Thr Gln Leu Val Asn Arg Asp Gly Lys Ala Val Gln Leu Lys  
 20 25 30

	Gly	Ile	Ser	Ser	His	Gly	Leu	Gln	Trp	Tyr	Gly	Glu	Tyr	Val	Asn	Lys
			35					40					45			
	Asp	Ser	Leu	Lys	Trp	Leu	Arg	Asp	Asp	Trp	Gly	Ile	Thr	Val	Phe	Arg
		50					55					60				
	Ala	Ala	Met	Tyr	Thr	Ala	Asp	Gly	Gly	Tyr	Ile	Asp	Asn	Pro	Ser	Val
	65					70					75					80
	Lys	Asn	Lys	Val	Lys	Glu	Ala	Val	Glu	Ala	Ala	Lys	Glu	Leu	Gly	Ile
					85					90					95	
	Tyr	Val	Ile	Ile	Asp	Trp	His	Ile	Leu	Asn	Asp	Gly	Asn	Pro	Asn	Gln
				100					105						110	
	Asn	Lys	Glu	Lys	Ala	Lys	Glu	Phe	Phe	Lys	Glu	Met	Ser	Ser	Leu	Tyr
			115					120					125			
	Gly	Asn	Thr	Pro	Asn	Val	Ile	Tyr	Glu	Ile	Ala	Asn	Glu	Pro	Asn	Gly
		130					135					140				
	Asp	Val	Asn	Trp	Lys	Arg	Asp	Ile	Lys	Pro	Tyr	Ala	Glu	Glu	Val	Ile
	145					150					155					160
	Ser	Val	Ile	Arg	Lys	Asn	Asp	Pro	Asp	Asn	Ile	Ile	Ile	Val	Gly	Thr
					165					170					175	
[0007]	Gly	Thr	Trp	Ser	Gln	Asp	Val	Asn	Asp	Ala	Ala	Asp	Asp	Gln	Leu	Lys
				180					185						190	
	Asp	Ala	Asn	Val	Met	Tyr	Ala	Leu	His	Phe	Tyr	Ala	Gly	Thr	His	Gly
			195					200					205			
	Gln	Phe	Leu	Arg	Asp	Lys	Ala	Asn	Tyr	Ala	Leu	Ser	Lys	Gly	Ala	Pro
		210					215					220				
	Ile	Phe	Val	Thr	Glu	Trp	Gly	Thr	Ser	Asp	Ala	Ser	Gly	Asn	Gly	Gly
	225					230					235					240
	Val	Phe	Leu	Asp	Gln	Ser	Arg	Glu	Trp	Leu	Lys	Tyr	Leu	Asp	Ser	Lys
					245					250					255	
	Thr	Ile	Ser	Trp	Val	Asn	Trp	Asn	Leu	Ser	Asp	Lys	Gln	Glu	Ser	Ser
				260					265						270	
	Ser	Ala	Leu	Lys	Pro	Gly	Ala	Ser	Lys	Thr	Gly	Gly	Trp	Arg	Leu	Ser
			275					280					285			
	Asp	Leu	Ser	Ala	Ser	Gly	Thr	Phe	Val	Arg	Glu	Asn	Ile	Leu	Gly	Thr
		290					295					300				
	Lys	Asp	Ser	Thr	Lys	Asp	Ile	Pro	Glu	Thr	Pro	Ser	Lys	Asp	Lys	Pro
	305					310					315					320
	Thr	Gln	Glu	Asn	Gly	Ile	Ser	Val	Gln	Tyr	Arg	Ala	Gly	Asp	Gly	Ser
					325					330					335	

	Met	Asn	Ser	Asn	Gln	Ile	Arg	Pro	Gln	Leu	Gln	Ile	Lys	Asn	Asn	Gly	
				340					345					350			
	Asn	Thr	Thr	Val	Asp	Leu	Lys	Asp	Val	Thr	Ala	Arg	Tyr	Trp	Tyr	Lys	
			355					360					365				
	Ala	Lys	Asn	Lys	Gly	Gln	Asn	Phe	Asp	Cys	Asp	Tyr	Ala	Gln	Ile	Gly	
		370					375					380					
	Cys	Gly	Asn	Val	Thr	His	Lys	Phe	Val	Thr	Leu	His	Lys	Pro	Lys	Gln	
	385					390					395					400	
[0008]	Gly	Ala	Asp	Thr	Tyr	Leu	Glu	Leu	Gly	Phe	Lys	Asn	Gly	Thr	Leu	Ala	
					405					410					415		
	Pro	Gly	Ala	Ser	Thr	Gly	Asn	Ile	Gln	Leu	Arg	Leu	His	Asn	Asp	Asp	
				420					425					430			
	Trp	Ser	Asn	Tyr	Ala	Gln	Ser	Gly	Asp	Tyr	Ser	Phe	Phe	Lys	Ser	Asn	
			435					440					445				
	Thr	Phe	Lys	Thr	Thr	Lys	Lys	Ile	Thr	Leu	Tyr	Asp	Gln	Gly	Lys	Leu	
		450					455					460					
	Ile	Trp	Gly	Thr	Glu	Pro	Asn										
	465					470											

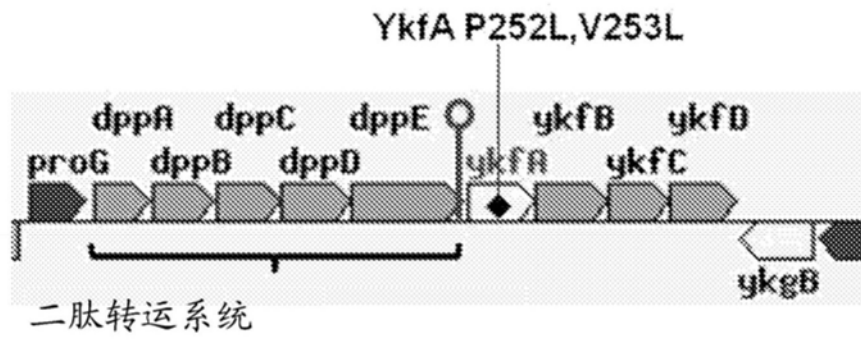


图1

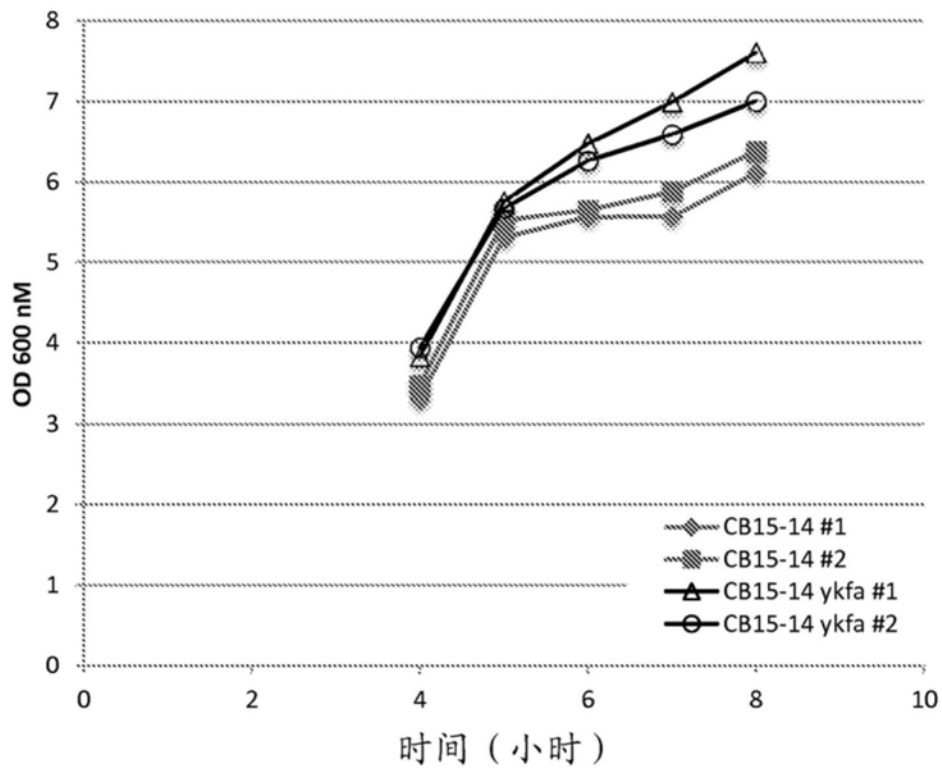


图2A

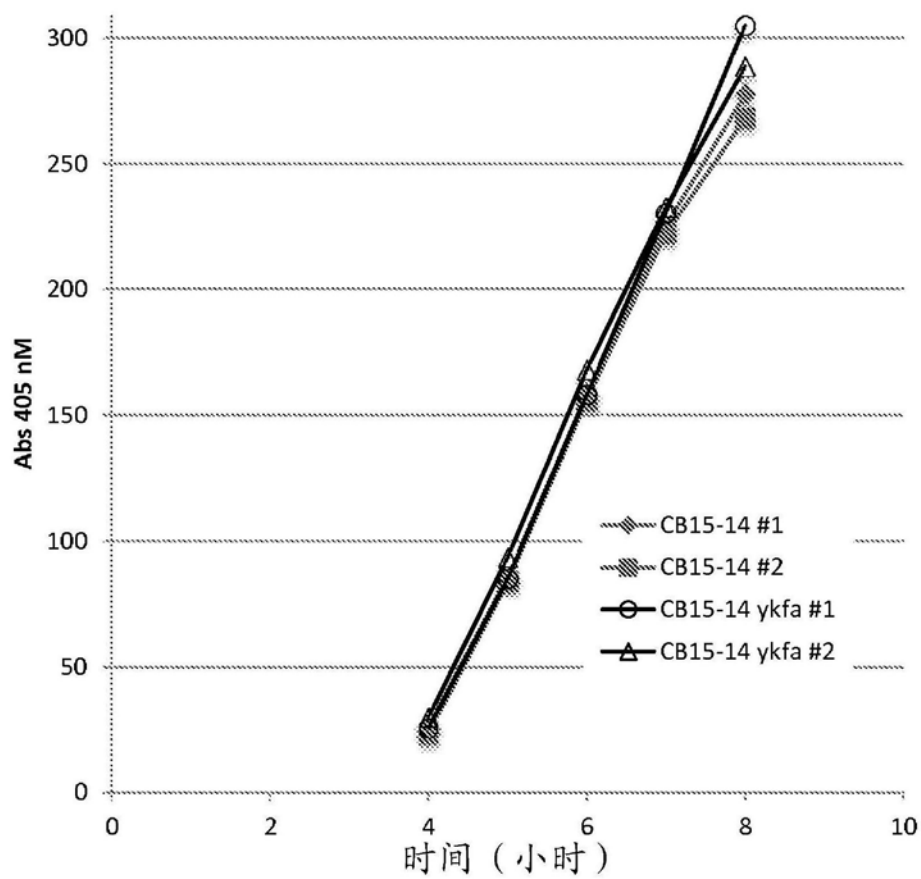


图2B



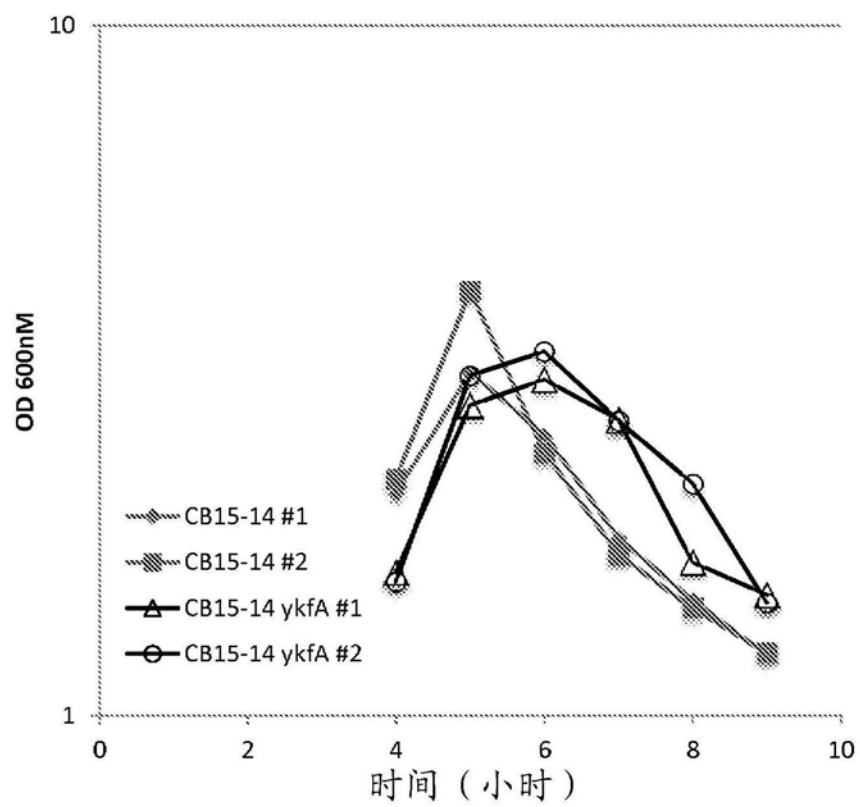


图3A

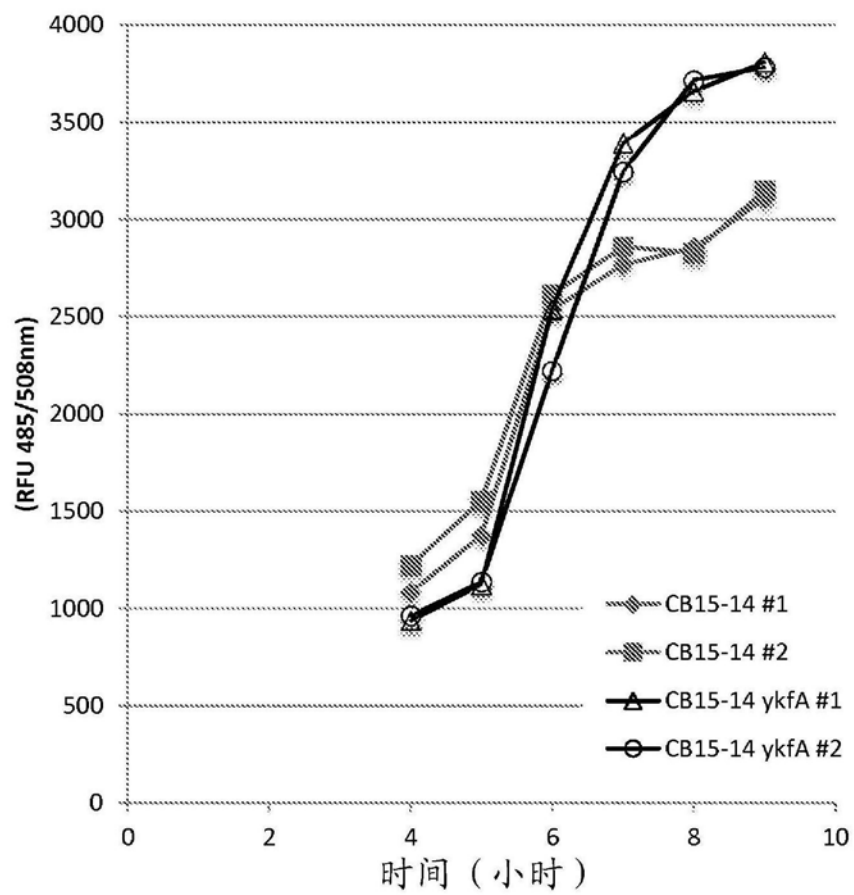


图3B

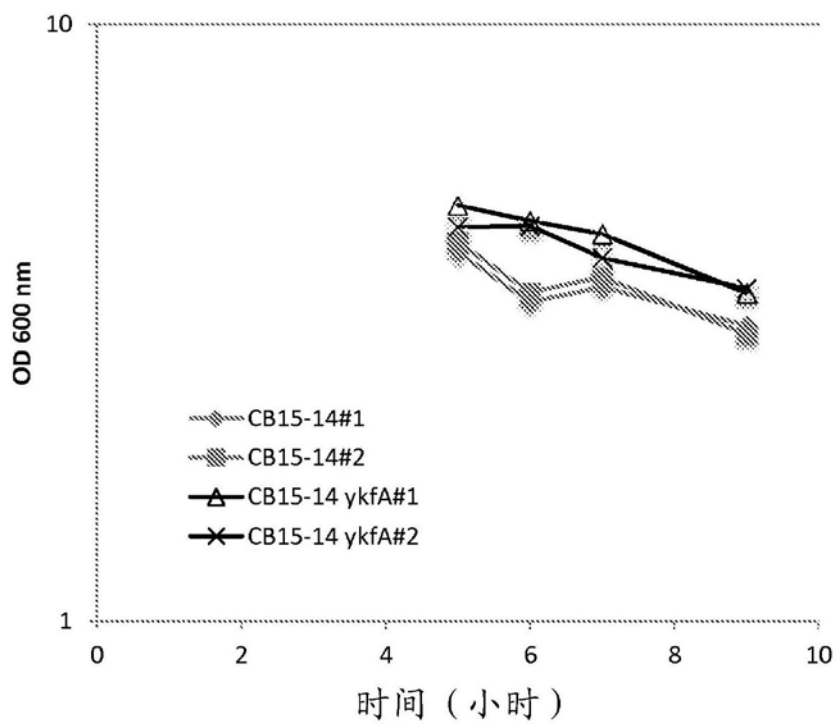


图4A

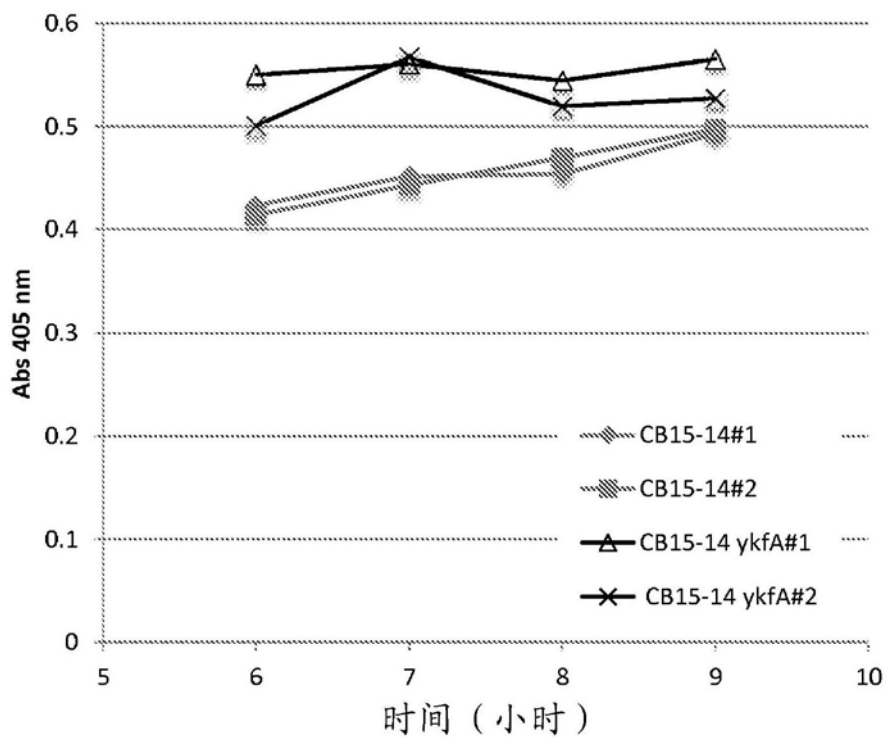


图4B