



(12) 发明专利申请

(10) 申请公布号 CN 119213015 A

(43) 申请公布日 2024. 12. 27

(21) 申请号 202380040804.1

(22) 申请日 2023.03.30

(30) 优先权数据

63/326,615 2022.04.01 US

(85) PCT国际申请进入国家阶段日

2024.11.15

(86) PCT国际申请的申请数据

PCT/US2023/065162 2023.03.30

(87) PCT国际申请的公布数据

W02023/192953 EN 2023.10.05

(71) 申请人 丹尼斯科美国公司

地址 美国

(72) 发明人 C·邦吉奥尔尼 F·格德埃格伯

R·L·弗里希 H·穆尔德

(74) 专利代理机构 北京市中咨律师事务所

11247

专利代理师 高飞 黄革生

(51) Int.Cl.

C07K 14/32 (2006.01)

C12P 21/02 (2006.01)

C07K 14/195 (2006.01)

C12N 1/00 (2006.01)

权利要求书2页 说明书39页

序列表(电子公布) 附图4页

(54) 发明名称

增强革兰氏阳性细菌细胞中的蛋白质生产的前区突变

(57) 摘要

本公开总体上涉及包含新型前区DNA序列的重组多核苷酸。本公开的某些方面涉及重组革兰氏阳性细菌菌株,这些菌株包含一种或多种引入的多核苷酸,该一种或多种引入的多核苷酸包含可操作地连接到编码目的蛋白的DNA序列的新型前区DNA序列。

1. 一种变体前区序列,其包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代,其中所述变体前区的氨基酸位置根据SEQ ID NO:15编号。

2. 如权利要求1所述的变体前区,其衍生自与SEQ ID NO:15具有至少约70%、约75%、约80%、约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

3. 一种变体前区序列,其包含位置2处的甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入,以及选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,其中所述变体前区的氨基酸位置根据SEQ ID NO:14编号。

4. 如权利要求3所述的变体前区,其衍生自与SEQ ID NO:14具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

5. 一种变体前区序列,其包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)以及位置73处的异亮氨酸(I)至缬氨酸(V)取代(I72V),其中所述变体前区的氨基酸位置根据SEQ ID NO:15进行编号;或一种变体前区序列,其包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)以及位置80处的谷氨酸(E)至异亮氨酸(I)取代(E80I),其中所述变体前区的氨基酸位置根据SEQ ID NO:15编号。

6. 如权利要求5所述的变体前区,其衍生自与SEQ ID NO:15具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

7. 一种多核苷酸,其包含如权利要求1-6中任一项所述的变体前区核酸。

8. 一种多核苷酸,其包含编码如权利要求1-6中任一项所述的变体前区的上游核酸,所述上游核酸可操作地连接到编码目的蛋白(POI)的下游核酸序列。

9. 一种多核苷酸,其包含编码信号序列的上游核酸,所述上游核酸可操作地连接到编码如权利要求1-6中任一项所述的变体前区的下游核酸序列,所述编码如权利要求1-6中任一项所述的变体前区的下游核酸序列可操作地连接到编码目的蛋白(POI)的下游核酸序列。

10. 一种表达盒,其包含可操作地连接到如权利要求7-9中任一项所述的下游多核苷酸的上游启动子。

11. 一种革兰氏阳性宿主细胞,其包含引入的如权利要求10所述的盒。

12. 一种用于在革兰氏阳性细菌细胞中生产目的蛋白(POI)的方法,所述方法包括:

(a) 将表达盒引入革兰氏阳性细胞,所述表达盒包含可操作地连接到编码变体前区序列的下游核酸的上游启动子,所述变体前区序列包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代,所述编码变体前区序列的下游核酸可操作地连接到编码所述POI的下游核酸,其中所述变体前区的氨基酸位置根据SEQ ID NO:15进行编号,以及

(b) 在用于生产所述POI的条件下培养经修饰的细胞。

13. 一种用于在革兰氏阳性细菌细胞中生产目的蛋白(POI)的方法,所述方法包括:

(a) 将表达盒引入革兰氏阳性细胞,所述表达盒包含可操作地连接到编码变体前区序列的下游核酸的上游启动子序列,所述变体前区序列包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)的氨基酸插入和选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,所述编码变体前区序列的下游核酸可操作地连接到编码所述POI的下游核酸,其中所述变体前区的氨基酸位置根据SEQ ID NO:14进行编号,以及

(b) 在用于生产所述POI的条件下培养经修饰的细胞。

14. 如权利要求12或权利要求13所述的方法,其中所述盒进一步包含编码信号序列的核酸,所述核酸可操作地连接并位于上游启动子和下游变体前区之间。

15. 如权利要求12或权利要求13所述的方法,其中所述经修饰的细胞相对于具有引入的表达盒的对照革兰氏阳性细胞生产增加量的POI,所述表达盒包含可操作地连接到编码SEQ ID NO:15的前区序列的下游核酸的相同的上游启动子序列,所述编码前区序列的下游核酸可操作地连接到编码相同POI的下游核酸序列。

16. 如权利要求12或权利要求13所述的方法,其中所述增加量的POI为相对于所述对照细胞增加至少约0.1%。

17. 如权利要求12或权利要求13所述的方法,其中所述POI选自由以下组成的组:乙酰酯酶、氨肽酶、淀粉酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、碳酸酐酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、葡萄糖醛酸酶、糖基水解酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲基酯酶、果胶分解酶、过水解酶、多元醇氧化酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、肽酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶、及其组合。

增强革兰氏阳性细菌细胞中的蛋白质生产的前区突变

技术领域

[0001] 本公开总体上涉及微生物宿主细胞、分子生物学、蛋白质工程、发酵、蛋白质生产等领域。本公开的某些方面涉及新型前区核酸 (DNA) 序列、包含新型前区DNA序列的重组多核苷酸、包含一个或多个引入的多核苷酸的经遗传修饰 (重组) 的革兰氏阳性细菌菌株等, 该一个或多个引入的多核苷酸包含可操作地连接到编码目的蛋白的核酸 (DNA) 序列的新型前区DNA序列。

[0002] 相关申请的交叉引用

[0003] 本申请要求2022年4月01日提交的美国临时申请号63/326,615的权益, 该临时申请通过引用以其全文特此并入。

[0004] 序列表的引用

[0005] 命名为“NB41862W0PCT_SequenceListing.xml”的文本文件序列表的电子提交的内容创建于2023年3月29日, 并且大小为51KB, 将其通过引用以其全文特此并入。

背景技术

[0006] 革兰氏阳性微生物因其将其发酵产物分泌到其培养基中的能力而常用于大规模工业发酵。分泌的蛋白质通过细胞膜和细胞壁输出, 并且然后在随后释放到外部介质中。例如, 大规模工业发酵和分泌异源多肽是广泛应用于工业的技术, 其中用编码待表达的异源多肽的核酸转化微生物细胞。尽管在蛋白质生产方法中有多种进展, 但本领域仍需要提供更高效的蛋白表达方法, 其旨在增强用于多种工业用途的目的蛋白的生产。

发明内容

[0007] 如本文总体所述, 本公开尤其提供了用于在革兰氏阳性细菌 (宿主) 细胞中生产目的蛋白的组合物和方法。某些实施例涉及新型前区核酸 (DNA) 序列、包含新型前区DNA序列的重组多核苷酸 (例如, 载体、质粒、表达盒等)、包含可操作地连接到下游 (3') 基因编码序列的新型前区序列的重组多核苷酸、包含可操作地连接到编码前蛋白信号 (分泌) 序列的上游 (5') DNA序列的新型前区序列的重组多核苷酸等。

[0008] 在某些方面, 本公开提供了重组革兰氏阳性细菌菌株, 这些菌株表达一个或多个引入的编码目的蛋白的多核苷酸。在特定的实施例中, 一个或多个引入的多核苷酸包含可操作地连接到编码目的蛋白的DNA序列 (该编码目的蛋白的DNA序列可包含可操作地连接到其上的上游 (5') 蛋白信号序列) 的新型前区DNA序列。在某些其他方面, 本公开提供了用于设计/构建表达编码目的蛋白的一个或多个引入的新型多核苷酸构建体的重组革兰氏阳性细菌菌株的组合物和方法、用于培养表达目的蛋白的重组菌株的组合物和方法、用于增强目的蛋白的生产的组合物和方法等。

[0009] 例如, 在一个或多个实施例中, 本公开提供了如SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:32和SEQ ID NO:33中一个或多个氨基酸序列和/或其组合所示的新型变体前区序列。在其他一个或多个实施例中, 本公开提供了

如图1、图2、或图3和/或其组合所示的新型变体前区序列。在又其他一个或多个实施例中，本公开提供了表1-5之一和/或其组合所示或描述的新型变体前区序列。

[0010] 因此，某些一个或多个实施例涉及包含一个或多个氨基酸取代、一个或多个氨基酸插入等的变体(突变体)前区序列。在某些实施例中，变体前区序列包含位置30处的氨基酸取代，其中该变体前区的氨基酸位置根据SEQ ID NO:15编号。在其他实施例中，变体前区序列包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代，其中该变体前区的氨基酸位置根据SEQ ID NO:15编号。在一个或多个其他实施例中，变体前区序列衍生自亲本多肽或参比多肽，该亲本多肽或参比多肽与SEQ ID NO:15的位置1-84具有至少约70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%序列同一性。在某些其他实施例中，变体前区序列包含位置2处的至少甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入，其中该变体前区的氨基酸位置根据SEQ ID NO:14编号。在其他一个或多个实施例中，变体前区序列包含位置2处的至少甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入，以及选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代，其中该变体前区的氨基酸位置根据SEQ ID NO:14编号。在一些其他实施例中，变体前区衍生自亲本多肽或参比多肽，该亲本多肽或参比多肽与SEQ ID NO:14的位置1-86具有至少约70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%氨基酸序列同一性。在又其他实施例中，变体前区序列包含位置2处的至少甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丙氨酸(A)的氨基酸插入，其中该变体前区的氨基酸位置根据SEQ ID NO:30编号。在某些方面，变体前区衍生自亲本多肽或参比多肽，该亲本多肽或参比多肽与SEQ ID NO:30的位置1-87具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%氨基酸序列同一性。在某些其他一个或多个实施例中，变体前区序列包含位置2处的至少甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丝氨酸(S)的氨基酸插入，其中该变体前区的氨基酸位置根据SEQ ID NO:29编号。在一些其他实施例中，变体前区衍生自亲本多肽或参比多肽，该亲本多肽或参比多肽与SEQ ID NO:29的位置1-87具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%氨基酸序列同一性。在其他一个或多个实施例中，变体前区序列包含位置2处的至少甘氨酸(G)、位置3处的赖氨酸(K)、位置4处的丙氨酸(A)和位置5处的丙氨酸(A)的氨基酸插入，其中该变体前区的氨基酸位置根据SEQ ID NO:31编号。在某些一个或多个实施例或方面中，变体前区衍生自亲本多肽或参比多肽，该亲本多肽或参比多肽与SEQ ID NO:31的位置1-88具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%氨基酸序列同一性。在其他实施例中，变体前区序列包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)和位置73处的异亮氨酸(I)至缬氨酸(V)取代(I72V)，其中该变体前区的氨基酸位置根据SEQ ID NO:32编号。在其他一个或多个方面或实施例中，变体前区序列包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)和位置80处的谷氨酸(E)至异亮氨酸(I)取代(E80I)，其中该变体前区的氨基酸位置根据SEQ ID NO:33编号。在某些实施例中，变体前区衍生自亲本多肽或参

比多肽,该亲本多肽或参比多肽与SEQ ID NO:15、SEQ ID NO:32和/或SEQ ID NO:33的位置1-84具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。在其他方面,本公开的变体前区包含表1-5、图1、图2、图3、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:33中任一项及其组合所示的氨基酸修饰。在一个或多个实施例中,包含表1-5、图1、图2、图3、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:33中任一项及其组合所示的氨基酸修饰的变体前区衍生自亲本多肽或参比多肽,该亲本多肽或参比多肽与SEQ ID NO:15具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0011] 因此,某些其他一个或多个实施例提供了编码本公开的新型变体前区序列的重组核酸。因此,某些实施例涉及包含变体前区核酸等的多核苷酸。在一个或多个实施例中,本公开提供了包含编码本公开变体前区的上游(5')核酸(序列)的多核苷酸,该上游核酸(序列)可操作地连接到编码异源目的蛋白(POI)的下游(3')核酸序列。在某些其他方面,本公开提供了多核苷酸,这些多核苷酸包含编码前蛋白信号(分泌)序列的上游(5')核酸,该上游核酸可操作地连接到编码本公开变体前区的下游(3')核酸序列,该编码本公开变体前区的下游核酸序列可操作地连接到编码目的蛋白(POI)的下游核酸序列。某些其他实施例涉及表达盒,这些表达盒包含可操作地连接到本公开的下游(3')多核苷酸的上游(5')启动子区序列。

[0012] 在某些其他一个或多个实施例中,本公开涉及包含本公开的一个或多个引入的多核苷酸或表达盒的重组(经修饰的)革兰氏阳性细菌细胞/菌株。因此,本公开的其他实施例涉及用于在本文阐述和描述的重组革兰氏阳性细胞中生产异源目的蛋白的方法/工艺。

[0013] 在一些一个或多个实施例中,本公开提供了用于在革兰氏阳性细菌细胞中生产异源POI的方法,这些方法包括将表达盒引入革兰氏阳性细胞,该表达盒包含可操作地连接到编码变体前区的下游核酸序列的上游启动子区,该变体前区包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代,其中该变体前区核酸(序列)可操作地连接到编码该POI的下游核酸序列,其中该变体前区的氨基酸位置根据SEQ ID NO:15进行编号;以及在用于生产该POI的条件下使经修饰的细胞生长/培养/发酵。在某些其他实施例中,本公开提供了用于在革兰氏阳性细菌细胞中生产异源POI的方法,这些方法包括将表达盒引入革兰氏阳性细胞,该表达盒包含可操作地连接到编码变体前区的下游核酸(序列)的上游启动子区序列,该变体前区包含位置2处的甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入,以及选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,其中该变体前区核酸(序列)可操作地连接到编码POI的下游核酸序列,其中该变体前区的氨基酸位置根据SEQ ID NO:14进行编号;以及在用于生产该POI的条件下使经修饰的细胞生长/培养/发酵。在又其他实施例中,本公开涉及用于在革兰氏阳性细菌细胞中生产异源POI的方法,这些方法包括向革兰氏阳性细胞中引入本公开的表达盒,以及在用于生产该POI的条件下使经修饰的细胞生长/培养/发酵。在这些方法的某些一个或多个实施例中,该盒包含编码前蛋白信号(分泌)序列的核酸(序列),该核酸(序列)可操作地连接并且位于启动子与变体前区序列之间。在这些方

法的一些其他实施例中,相对于对照革兰氏阳性细胞,经修饰的细胞生产增加量的POI。例如,如下对照革兰氏阳性细胞,其包含引入的表达盒,该表达盒包含可操作地连接到编码包含SEQ ID NO:15的前区序列的下游核酸的相同的上游启动子序列(即,与经修饰的细胞相同),该编码前区序列的下游核酸可操作地连接到编码相同POI(即,与经修饰的细胞相同的POI)的下游核酸序列。在这些方法的某些实施例中,将引入的盒整合到该细胞的基因组中。在这些方法的某些实施例中,将至少两个盒引入该细胞。本文描述了本公开的某些其他一个或多个实施例并如下文阐述。

附图说明

[0014] 图1显示了野生型(参比)迟缓芽孢杆菌前区的氨基酸序列和本公开的某些SEL变体前区序列。如图1所示,参比迟缓芽孢杆菌前区序列包含SEQ ID NO:15所示的八十四(84)个氨基酸(残基)位置(其中位置30处的谷氨酸(E)残基加下划线),变体前区序列A包含SEQ ID NO:9所示的八十四(84)个氨基酸位置(其中位置30处的组氨酸(H)残基为粗体),变体前区序列B包含SEQ ID NO:11所示的八十四(84)个氨基酸位置(其中位置30处的甘氨酸(G)残基为粗体),变体前区序列C包含SEQ ID NO:14所示的八十六(86)个氨基酸位置(其中分别在位置2和3插入的甘氨酸(G)和赖氨酸(K)残基以粗体(GK)显示,并且位置32处的E残基加下划线),变体前区序列D包含SEQ ID NO:29所示的八十七(87)个氨基酸位置(其中分别在位置2-4插入的甘氨酸(G)、赖氨酸(K)和丝氨酸(S)残基以粗体(GKS)显示),变体前区序列E包含SEQ ID NO:30所示的八十七(87)个氨基酸位置(其中分别在位置2-4插入的甘氨酸(G)、赖氨酸(K)和丙氨酸(A)残基以粗体(GKA)显示),变体前区序列F包含SEQ ID NO:31所示的八十八(88)个氨基酸位置(其中分别在位置2-5插入的甘氨酸(G)、赖氨酸(K)、丙氨酸(A)和丙氨酸(A)残基以粗体(GKAA)显示),变体前区序列G包含SEQ ID NO:32所示的八十四(84)个氨基酸位置(其中位置30处的甘氨酸(G)残基以粗体显示,L68K和I72V取代加下划线)并且变体前区序列H包含SEQ ID NO:33所示的八十四(84)个氨基酸位置(其中位置30处的甘氨酸(G)残基为粗体,L68K和E80I取代加下划线)。

[0015] 图2显示了与野生型(WT)前区序列(SEQ 15)比对的变体前区序列A(SEQ 09)、变体前区序列B(SEQ 11)、变体前区序列C(SEQ 14)、变体前区序列D(SEQ 29)、变体前区序列E(SEQ 30)、变体前区序列F(SEQ 31)、变体前区序列G(SEQ 32)以及变体前区序列H(SEQ 33)的氨基酸序列。如图2所示,WT前区在位置30处包含谷氨酸(E)残基(E;SEQ 15),变体前区序列A在位置30处包含组氨酸(H)残基(H;SEQ 09),变体前区序列B在位置30处包含甘氨酸(G)残基(G;SEQ 11),变体前区序列C包含甘氨酸(G)和赖氨酸(K)两(2)个氨基酸插入(GK插入;SEQ 14),变体前区序列D包含甘氨酸(G)、赖氨酸(K)和丝氨酸(S)三(3)个氨基酸插入(GKS插入;SEQ 29),变体前区序列E包含甘氨酸(G)、赖氨酸(K)和丙氨酸(A)三(3)个氨基酸插入(GKA插入;SEQ 30),变体前区序列F包含甘氨酸(G)、赖氨酸(K)、丙氨酸(A)、丙氨酸(A)四(4)个氨基酸插入(GKAA插入;SEQ 31),变体前区序列G包含位置30处的甘氨酸(G)残基、位置68处的赖氨酸(K)残基和位置72处的缬氨酸(V)残基,并且变体前区序列H包含位置30处的甘氨酸(G)残基、位置68处的赖氨酸(K)残基和位置80处的异亮氨酸(I)残基。例如,如在WT前区序列(SEQ 15)和变体前区序列C(SEQ 14)的比对中所示,在WT前区(SEQ ID NO:15)位置1的丙氨酸(A)与位置2的谷氨酸(E)之间插入了两(2)个氨基酸残基GK,产生了SEQ ID

N0:14的86个氨基酸的变体前区序列C。如图2所示,SEQ 15、SEQ 09、SEQ 11、SEQ 32和SEQ 33中显示的两个连字符(--)指示相对于SEQ 14所示的变体前区序列C的两(2)个氨基酸位置空位。

[0016] 图3显示了以下的氨基酸(残基)位置编号:参比(野生型)迟缓芽孢杆菌前区氨基酸序列(E₃₀;SEQ ID NO:15)、变体前区序列A(H₃₀;SEQ ID NO:09)、变体前区序列B(G₃₀;SEQ ID NO:11)、变体前区序列C(G₂K₃插入,SEQ ID NO:14)、变体前区序列D(G₂K₃S₄插入,SEQ ID NO:29)、变体前区序列E(G₂K₃A₄插入,SEQ ID NO:30)、变体前区序列F(G₂K₃A₄A₅插入,SEQ ID NO:31)、变体前区序列G(具有E30G、L68K和I72V取代;SEQ ID NO:32)和变体前区序列F(具有E30G、L68K和E80I取代;SEQ ID NO:33),其中氨基酸位置从5' (N-末端)到3' (C-末端)方向用下标数字指示,例如,SEQ ID NO:15的位置1处的丙氨酸(A₁)、SEQ ID NO:15的位置2处的谷氨酸(E₂)等。例如,本公开的突变前区序列可以是指野生型(参比)前区序列(编号SEQ ID NO:15,E30)、变体(参比)前区序列A(编号SEQ ID NO:9,H30)、变体(参比)前区序列B(编号SEQ ID NO:11,G30)、变体(参比)前区序列C(编号SEQ ID NO:14,GK插入)、变体(参比)前区序列D(编号SEQ ID NO:29,GKS插入)、变体(参比)前区序列E(编号SEQ ID NO:20,GKA插入)、变体(参比)前区序列F(编号SEQ ID NO:31,GKAA插入)、变体(参比)前区序列G(编号SEQ ID NO:32,L68K/I72V取代)、变体(参比)前区序列H(编号SEQ ID NO:33,L68K/E80I取代)或其组合,如图3所示。

[0017] 生物序列说明

[0018] SEQ ID NO:1是包含上游(5') aprE基因侧翼区、枯草芽孢杆菌rrnI-P2启动子变体和5'-aprE UTR区的核酸(DNA)序列。

[0019] SEQ ID NO:2是命名为“BG46”的野生型吉氏芽孢杆菌(*Bacillus gibsonii*)枯草杆菌蛋白酶的氨基酸序列。

[0020] SEQ ID NO:3是命名为“BG46_变体1”的变体吉氏芽孢杆菌(*B.gibsonii*)BG46枯草杆菌蛋白酶的氨基酸序列。

[0021] SEQ ID NO:4是编码野生型枯草芽孢杆菌AprE蛋白信号序列的DNA序列。

[0022] SEQ ID NO:5是编码变体前区序列A(30H,SEQ ID NO:9)的DNA序列。

[0023] SEQ ID NO:6是野生型解淀粉芽孢杆菌BPN'终止子的DNA序列。

[0024] SEQ ID NO:7是卡那霉素(kan)基因表达盒的DNA序列。

[0025] SEQ ID NO:8是命名为“BG46_变体2”的变体吉氏芽孢杆菌BG46枯草杆菌蛋白酶的氨基酸序列。

[0026] SEQ ID NO:9是由变体前区A DNA序列(SEQ ID NO:5)编码的氨基酸序列。

[0027] SEQ ID NO:10是编码变体前区序列B(30G,SEQ ID NO:11)的DNA序列。

[0028] SEQ ID NO:11是由变体前区B DNA序列(SEQ ID NO:10)编码的氨基酸序列。

[0029] SEQ ID NO:12是解淀粉芽孢杆菌BPN'前区序列的氨基酸序列。

[0030] SEQ ID NO:13是编码变体前区序列C(GK插入,SEQ ID NO:14)的DNA序列。

[0031] SEQ ID NO:14是由变体前区C DNA序列(SEQ ID NO:13)编码的氨基酸序列。

[0032] SEQ ID NO:15是野生型(WT)迟缓芽孢杆菌(*B.lentus*)前区序列的氨基酸序列。

[0033] SEQ ID NO:16是地衣芽孢杆菌(*B.licheniformis*)serA基因5' FR的DNA序列。

[0034] SEQ ID NO:17是包含rrnI-p3启动子区的DNA序列。

- [0035] SEQ ID NO:18是枯草芽孢杆菌aprE 5' -UTR的DNA序列。
- [0036] SEQ ID NO:19是地衣芽孢杆菌amyL终止子的DNA序列。
- [0037] SEQ ID NO:20是地衣芽孢杆菌serA基因3' FR的DNA序列。
- [0038] SEQ ID NO:21是地衣芽孢杆菌lysA基因5' FR的DNA序列。
- [0039] SEQ ID NO:22是地衣芽孢杆菌lysA基因3' FR的DNA序列。
- [0040] SEQ ID NO:23是编码变体前区序列C(GKA插入,SEQ ID NO:14)的DNA序列。
- [0041] SEQ ID NO:24是由变体前区C DNA序列(SEQ ID NO:23)编码的氨基酸序列。
- [0042] SEQ ID NO:25是编码变体前区序列C(GKAA插入,SEQ ID NO:14)的DNA序列。
- [0043] SEQ ID NO:26是由变体前区C DNA序列(SEQ ID NO:25)编码的氨基酸序列。
- [0044] SEQ ID NO:27是编码变体前区序列C(GKAS插入,SEQ ID NO:14)的DNA序列。
- [0045] SEQ ID NO:28是由变体前区C DNA序列(SEQ ID NO:27)编码的氨基酸序列。
- [0046] SEQ ID NO:29是变体前区序列D的氨基酸序列。
- [0047] SEQ ID NO:30是变体前区序列E的氨基酸序列。
- [0048] SEQ ID NO:31是变体前区序列F的氨基酸序列。
- [0049] SEQ ID NO:32是变体前区序列G的氨基酸序列。
- [0050] SEQ ID NO:33是变体前区序列H的氨基酸序列。

具体实施方式

[0051] 如上文简要阐述和下文详细描述,本公开尤其提供了新型前区核酸(DNA)序列、包含新型前区DNA序列的重组多核苷酸(例如,载体、质粒、表达盒等)、包含可操作地连接到下游(3')基因编码序列的新型前区序列的重组多核苷酸、包含可操作地连接到编码蛋白信号(分泌)序列的上游(5')DNA序列的新型前区序列的重组多核苷酸等。在某些方面,本公开提供了重组革兰氏阳性细菌菌株,这些菌株表达一个或多个引入的编码目的蛋白的多核苷酸。在特定的实施例中,一个或多个引入的多核苷酸包含可操作地连接到编码目的蛋白的DNA序列(该编码目的蛋白的DNA序列可包含可操作地连接到其上的上游(5')蛋白信号序列)的新型前区DNA序列。因此,本公开的某些其他方面尤其提供了用于设计/构建表达编码目的蛋白的一个或多个引入的新型多核苷酸构建体的重组革兰氏阳性细菌菌株的组合物和方法、用于培养表达目的蛋白的重组菌株的组合物和方法、用于增强目的蛋白的生产的组合物和方法等。

[0052] I. 定义

[0053] 鉴于本文所述的本公开的重组(经修饰的)菌株及其方法,定义了以下术语和短语。本文未定义的术语应当符合如本领域使用的其普通含义。

[0054] 除非另外定义,否则本文使用的所有技术和科学术语都具有与本发明组合物和方法应用的领域的普通技术人员通常所理解的相同的含义。尽管在本发明组合物和方法的实践或测试中也可以使用与本文所述的方法和材料类似或等同的任何方法和材料,但现在描述了代表性的说明性方法和材料。将本文引用的所有出版物和专利都通过引用以其全文并入。

[0055] 需进一步注意,可以撰写权利要求书以排除任何任选的要素。因此,此陈述旨在作为使用与权利要求要素的叙述有关的排他性术语如“单独”、“仅”、“排除”、“不包括”等或使

用“否定型”限定的前提基础(或其条件)。

[0056] 在阅读本公开后,如将对于本领域技术人员显而易见的,本文描述和说明的单独实施例中的每一个具有离散的组分和特征,这些组分和特征可以在不偏离本文所述的本发明组合物和方法的范围或精神的情况下容易地与其他几个实施例中的任一个的特征分离或组合。可以按照所叙述的事件的顺序或按照逻辑上可能的任何其他顺序来进行任何叙述的方法。

[0057] 如本文使用的,短语“革兰氏阳性细菌”、“革兰氏阳性细胞”、“革兰氏阳性细菌菌株”、和/或“革兰氏阳性细菌细胞”具有与本领域所用相同的含义。例如,革兰氏阳性细菌细胞包括放线菌门(Actinobacteria)和厚壁菌门(Firmicutes)的所有菌株。在某些实施例中,这种革兰氏阳性细菌属于杆菌、梭菌(Clostridia)和柔膜菌纲(Mollicutes)类。

[0058] 如本文使用的,“芽孢杆菌属”包括如本领域技术人员已知的“芽孢杆菌”属内的所有物种,包括但不限于枯草芽孢杆菌、地衣芽孢杆菌、迟缓芽孢杆菌(*B.lentus*)、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*)、解淀粉芽孢杆菌(*B.amyloliquefaciens*)、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(*B.thuringiensis*)。应认识到,芽孢杆菌属不断进行分类学重组。因此,该属旨在包括已重新分类的物种,包括但不限于诸如嗜热脂肪芽孢杆菌(其现在命名为“嗜热脂肪土芽孢杆菌(*Geobacillus stearothermophilus*)”)等的生物体。

[0059] 如本文使用的,术语“重组”或“非天然”是指具有至少一个工程化遗传改变或已通过引入异源核酸分子而修饰的生物体、微生物、细胞、核酸分子或载体,或者是指已经被改变使得可以控制异源或内源核酸分子或基因的表达的细胞(例如,微生物细胞)。重组还指细胞衍生自非天然细胞,或者是具有一个或多个这样的修饰的非天然细胞的后代。遗传改变包括例如引入编码蛋白质的可表达核酸分子的修饰,或其他核酸分子的添加、缺失、取代,或细胞遗传物质的其他功能性改变。例如,重组细胞可以表达未在天然(野生型)细胞内发现的相同或同源形式的基因或其他核酸分子(例如,融合蛋白或嵌合蛋白),或者可以提供改变的内源基因表达模式,诸如过表达、低表达、最低表达或根本不表达。“重组(recombination, recombining)”或生成“重组的(recombined)”核酸通常是两个或更多个核酸片段的组装,其中该组装产生嵌合基因。

[0060] 术语“衍生的”涵盖术语“起源的”、“获得的”、“可获得的”、和“产生的”,并且通常指示一种指定的材料或组合物在另一种指定的材料或组合物中找到它的起源,或者具有可以参考另一种指定的材料或组合物描述的特征。例如,本公开的重组革兰氏阳性细菌细胞可以衍生自/获得自任何已知的革兰氏阳性细菌菌株。

[0061] 如本文使用的,“核酸”是指核苷酸或多核苷酸序列及其片段或部分,以及基因组或合成来源的DNA、cDNA和RNA,其可以是双链或单链,无论代表有义链还是反义链。应理解,由于遗传密码的简并性,多种核苷酸序列可以编码给定的蛋白质。

[0062] 应理解,本文所述的多核苷酸(或核酸分子)包括“基因”、“载体”和“质粒”。

[0063] 相应地,术语“基因”是指编码氨基酸的特定序列的多核苷酸,其包含所有或部分蛋白质编码序列,并且可以包括调控(非转录的)DNA序列,诸如启动子序列,该启动子序列

确定例如表达基因的条件。基因的转录区可以包括非翻译区 (UTR) (包括5' -非翻译区 (UTR) 和3' -UTR) 以及编码序列。

[0064] 如本文使用的,“内源基因”是指位于生物体基因组的其天然位置中的基因。

[0065] 如本文使用的,“异源”基因、“非内源”基因或“外源”基因是指正常情况下未在宿主生物体中发现,但通过基因转移引入宿主生物体中的基因。术语一个或多个“外源”基因包含插入非天然生物体中的天然基因和/或插入天然或非天然生物体中的嵌合基因。

[0066] 如本文使用的,“异源控制序列”是指基因表达控制序列(例如,启动子、增强子、终止子等),其本质上不起作用以调控(控制)目的基因的表达。通常,异源核酸对于它们存在的细胞或基因组的一部分而言不是内源(天然)的,并且已通过感染、转染、转导、转化、显微注射、电穿孔等添加到细胞中。“异源”核酸构建体可以含有与在天然宿主细胞中发现的控制序列/DNA编码序列组合相同或不同的控制序列/DNA编码(ORF)序列组合。

[0067] 如本文使用的,术语“表达”是指衍生自本公开的核酸分子的有义(mRNA)或反义RNA的转录和稳定积累。表达也可以指将mRNA翻译成多肽。因此,术语“表达”包括多肽产生中涉及的任何步骤,这些步骤包括但不限于转录、转录后修饰、翻译、翻译后修饰、分泌等。

[0068] 如本文使用的,术语“编码序列”是指直接指定其(所编码的)蛋白质产物的氨基酸序列的核苷酸序列。编码序列的边界通常由常以ATG起始密码子开始的可读框(在下文中,“ORF”)确定。编码序列典型地包括DNA、cDNA和重组核苷酸序列。

[0069] 如本文使用的,术语“启动子”、“启动子元件”、“启动子序列”等是指能够控制基因编码序列(CDS)转录成信使RNA(mRNA)的核酸(DNA)序列,此时启动子区序列置于上游(5')并且可操作地连接到下游(3')基因CDS。如本领域技术人员通常理解的,启动子典型地提供用于由RNA聚合酶特异性结合和转录起始的位点。在某些方面,术语“启动子”是指启动转录所需的启动子核酸序列的最小部分(即,包含RNA聚合酶结合位点)。例如,启动子通常包含“-10”(共有序列)元件和“-35”(共有序列)元件,这两个元件位于上游(5')并且与待转录的基因CDS的+1转录起始位点(TSS)是相对的。核心启动子-10和-35元件在本领域中通常被分别称为“TATAAT”(Pribnow盒)共有区和“TTGACA”共有区。核心启动子(10和-35)区域的间隔通常被约十五至二十(15-20)个间隔碱基对(核苷酸)分开(间隔)。

[0070] 启动子可以全部衍生自天然基因,或者由衍生自在自然界中发现的不同启动子的不同元件组成,或者甚至包含合成核酸区段。本领域技术人员应理解,不同的启动子可以指导基因在不同细胞类型中、或在不同发育阶段、或响应于不同环境或生理条件而表达。启动子可以是组成型启动子、诱导型启动子、可调启动子、杂合启动子、合成启动子、串联启动子等。使基因在大多数细胞类型中在大多数时间表达的启动子通常被称为“组成型启动子”。应进一步认识到,由于在大多数情况下,尚未完全界定调控序列的确切边界,因此不同长度的DNA片段可以具有相同启动子活性。在某些方面,上游(5')启动子序列(pro)可操作地连接到编码蛋白信号序列(SS)的下游DNA序列,该编码蛋白信号序列(SS)的下游DNA序列可操作地连接到编码前区序列(PRO)的下游DNA序列,该编码前区序列(PRO)的下游DNA序列可操作地连接到编码成熟目的蛋白的下游(3')DNA序列(ORF),这可以示意性地呈现为5' - [pro] - [SS] - [PRO] - [ORF] - 3'。

[0071] 如本文使用的,“控制与目的基因的蛋白质编码序列连接的目的基因的表达的功能性启动子序列”是指控制所期望的革兰氏阳性宿主细胞中编码序列的转录和翻译的启动

子序列。例如,在某些实施例中,本公开提供了包含在革兰氏阳性细胞中有功能的上游(5')启动子(或5'启动子区、或串联5'启动子等)的多核苷酸,其中功能性启动子区可操作地连接到编码目的蛋白的核酸序列。

[0072] 如本文使用的,术语“前体蛋白”是指蛋白质的非活性形式。在某些方面,全长蛋白质以前序列和成熟蛋白(缩写为“前蛋白(pre-protein)”)的形式作为前体合成。在其他方面,全长蛋白质以信号肽序列、前序列和成熟蛋白(缩写为“pre-pro-蛋白”)的形式作为前体合成。例如,前序列(pre-sequence)通常作为转运的信号肽,并且前序列对于相关(成熟)蛋白的正确折叠典型地是必不可少的。

[0073] 如本文使用的,术语“成熟蛋白”是指与无活性前体(全长)蛋白质相对的蛋白质活性形式。

[0074] 如本文使用的,术语“信号序列”、“分泌信号”和“信号肽”可以互换使用并且是指可以参与前体蛋白的分泌或定向转运的氨基酸残基的序列。典型地,在易位过程中通过信号肽酶将信号(pre)序列从前体蛋白切割。信号(pre)序列典型地位于成熟蛋白序列的N-末端,或位于前区(pro)序列的N-末端,此时信号(pre)序列和前区(pro)序列以可操作的组合使用且位于成熟POI序列的上游(5')。

[0075] 如本文使用的,术语“前序列”(pro sequence)、“前序列”(pro-sequence)和“前区序列”(pro-region sequence)可互换使用并且缩写为“PRO”序列、“Pro”序列、“pro”序列等。本文使用的术语前序列具有与本领域所理解的相同含义。例如,枯草芽孢杆菌碱性丝氨酸蛋白酶“枯草杆菌蛋白酶”首先作为pre-pro-枯草杆菌蛋白酶产生,其由以下组成:用于蛋白质分泌的信号(pre)序列,其后是七十七(77)个氨基酸的前区(pro)序列,其后是编码成熟枯草杆菌蛋白酶(例如,pre-pro-枯草杆菌蛋白酶)的氨基酸序列。前序列充当分子内伴侣(例如,直接催化蛋白折叠反应),对于相关(成熟)蛋白的正确折叠通常是必不可少的。同样,成熟目的蛋白的折叠和细胞内转运(或分泌)可能都需要前序列,这表明这两种功能密切相关。在某些方面,本公开的前区序列包含衍生自SEQ ID NO:15的野生型(WT,参比)迟缓芽孢杆菌前区序列的氨基酸序列。在某些实施例中,前区序列衍生自与SEQ ID NO:15、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32和/或SEQ ID NO:33包含同源性的氨基酸序列。因此,在某些方面,本文所述的一个或多个前区变体的氨基酸修饰通过参比SEQ ID NO:15、SEQ ID NO:9、SEQ ID NO:11或SEQ ID NO:14的前区氨基酸序列进行编号。

[0076] 例如,使用比对算法可以将本文所述的一个或多个前区变体的氨基酸序列与SEQ ID NO:15的氨基酸序列进行比对,并且与SEQ ID NO:15的氨基酸残基比对(优选地最佳比对)的给定氨基酸序列中的每个氨基酸残基通过参比相应氨基酸残基的数字位置而方便地编号。当与查询序列(有时也称为“参比序列”)相比时,序列比对算法将鉴定在主题序列中发生插入或缺失的一个或多个位置。与其他前区氨基酸序列的序列比对可以通过使用氨基酸比对算法来确定。

[0077] 在某些方面,提及前区氨基酸序列“位置”可以呈现为单字母氨基酸(残基),其后是位置编号(例如,SEQ ID NO:15,位置1处的丙氨酸呈现为“A1”,位置2处的谷氨酸呈现为“E2”,等)。在相关实施例中,提及变体(突变)前区氨基酸序列可以呈现为单字母氨基酸(残基),其中将亲本(参比)序列的氨基酸位置编号,其后是位于相同位置的突变氨基酸(残基)

(例如,SEQ ID NO:15,位置1处的丙氨酸(A)被甘氨酸(G)取代呈现为“A1G”,位置30处的谷氨酸(E)被组氨酸(H)取代呈现为“E30H”等)。还可以在前区氨基酸序列的相同位置处取代多个氨基酸残基,其中将亲本(参比)序列的氨基酸位置编号,其后是由斜线分开的相同位置处的一个或多个突变氨基酸残基(例如,SEQ ID NO:15,位置1处的丙氨酸(A)被甘氨酸(G)、谷氨酸(E)或赖氨酸(K)取代可以呈现为“A1G/E/K”等)。

[0078] 本文提供了某些范围,其中数值前面是术语“约”。术语“约”在本文中用于为其后面的确切数字以及与该术语后面的数字接近或近似的数字提供文字支持。在判定数字是否接近或近似于特定叙述的数字时,接近或近似的未叙述的数字可以是在呈现其的上下文中提供特定叙述的数字的实质性等效物的数字。例如,关于数值,术语“约”是指数值的-10%至+10%的范围,除非术语在上下文中另外特别定义。

[0079] 如本文使用的,编码“PRO”序列的多核苷酸、编码“前区”序列的多核苷酸和编码“前序列”的多核苷酸可以互换使用,并且是指位于紧邻上游(5')并可操作地连接到编码成熟目的蛋白(POI)的下游(3')基因编码序列(CDS)的DNA序列。在某些方面,编码前区序列的多核苷酸是指如下DNA序列,其位于并可操作地连接到上游(5')蛋白信号序列(SS)和编码成熟POI的下游(3')基因CDS之间。例如,前区DNA序列(PRO)可操作地连接编码蛋白信号序列的上游(5')DNA序列(SS)并可操作地连接到编码成熟POI的下游(3')DNA基因CDS(例如,5'-[SS]-[PRO]-[ORF]-3';N-[pre]-[pro]-[蛋白]-C)。

[0080] 如本文使用的,“野生型(WT)迟缓芽孢杆菌前区”序列包含SEQ ID NO:15所示的氨基酸序列。在某些方面,WT(参比)前区序列(SEQ ID NO:15)可以指“E30前序列”或“E₃₀前序列”。

[0081] 如本文使用的,命名为“变体前区A”的突变前区序列包含SEQ ID NO:9所示的氨基酸序列。在某些实施例中,变体前区A序列(SEQ ID NO:9)可称为“H30前序列”或“H₃₀前序列”。

[0082] 如本文使用的,命名为“变体前区B”的突变前区序列包含SEQ ID NO:11所示的氨基酸序列。在某些实施例中,变体前区B序列(SEQ ID NO:11)可称为“G30前序列”或“G₃₀前序列”。

[0083] 如本文使用的,命名为“变体前区C”的突变前区序列包含SEQ ID NO:14所示的氨基酸序列。在某些实施例中,变体前区C序列(SEQ ID NO:14)可称为“GK插入前序列”或“G₂K₃插入前序列”。

[0084] 如本文使用的,命名为“变体前区D”的突变前区序列包含SEQ ID NO:29所示的氨基酸序列。在某些实施例中,变体前区D序列(SEQ ID NO:29)可称为“GKS插入前序列”或“G₂K₃S₄插入前序列”。

[0085] 如本文使用的,命名为“变体前区E”的突变前区序列包含SEQ ID NO:30所示的氨基酸序列。在某些实施例中,变体前区E序列(SEQ ID NO:30)可称为“GKA插入前序列”或“G₂K₃A₄插入前序列”。

[0086] 如本文使用的,命名为“变体前区F”的突变前区序列包含SEQ ID NO:31所示的氨基酸序列。在某些实施例中,变体前区F序列(SEQ ID NO:31)可称为“GKAA插入前序列”或“G₂K₃A₄A₅插入前序列”。

[0087] 如本文使用的,命名为“变体前区G”的突变前区序列包含SEQ ID NO:32所示的氨

氨基酸序列。在某些实施例中,变体前区G序列(SEQ ID NO:32)可称为L68K/I72V前序列。

[0088] 如本文使用的,命名为“变体前区H”的突变前区序列包含SEQ IDNO:32所示的氨基酸序列。在某些实施例中,变体前区H序列(SEQ ID NO:33)可称为L68K/E80I前序列。

[0089] 如图3总体所示,氨基酸(残基)位置编号可参照WT前区氨基酸序列(E30,SEQ ID NO:15)、变体前区序列A(H30;SEQ ID NO:09)、变体前区序列B(G₃₀;SEQ ID NO:11)、变体前区序列C(GK插入,SEQ ID NO:14)、变体前区序列D(GKS插入,SEQ ID NO:29)、变体前区序列E(GKA插入,SEQ ID NO:30)、变体前区序列F(GKAA插入,SEQ ID NO:31)、变体前区序列G(L68K/I72V,SEQ ID NO:32)和/或变体前区序列H(L68K/E80I,SEQ ID NO:33)。

[0090] 如本文使用的,短语“编码全长蛋白质的多核苷酸”是指编码“前体”蛋白的DNA序列;并且短语“编码成熟蛋白的多核苷酸”是指编码如本文定义的“成熟”蛋白的DNA序列。在某些方面,编码前体蛋白质的多核苷酸包含至少编码前区氨基酸序列的上游(5')DNA序列,该上游序列可操作地连接到编码成熟目的蛋白(POI)的氨基酸序列的下游(3')DNA序列(例如,可读框,ORF)。在其他方面,编码前体蛋白的多核苷酸包含至少编码蛋白信号序列的上游(5')DNA序列,该上游序列可操作地连接到编码前区氨基酸序列的下游(3')DNA序列,该编码前区氨基酸序列的下游DNA序列可操作地连接到编码成熟POI的氨基酸序列的下游(3')ORF。

[0091] 如本文使用的,术语“非翻译区”可以缩写为“UTR”。

[0092] 如本文使用的,短语“5端(5')非翻译区”、“5'非翻译区”和/或“5'转录前导序列”可互换使用并缩写为“5'-UTR”。如本领域通常所理解的,已知5'-UTR是信使RNA(mRNA)的区域,直接位于起始密码子的上游(5')。

[0093] 当核酸与另一核酸序列处于功能关系时,该核酸是“可操作地连接的”。例如,如果编码分泌性前导序列(即,信号序列)的DNA表达为参与多肽分泌的前蛋白,那么该编码分泌性前导序列的DNA可操作地连接到编码该多肽的DNA;如果启动子或增强子影响编码序列的转录,则该启动子或增强子可操作地连接到该序列(CDS,ORF);或者如果核糖体结合位点被定位以促进翻译,则该核糖体结合位点可操作地连接到编码序列。通常,“可操作地连接”意指被连接的DNA序列是连续的,并且在分泌性前导序列的情况下,是连续的并且处于阅读相(reading phase)中。然而,增强子不必是连续的。通过在方便的限制位点处连接来实现连接。如果这样的位点不存在,则按照常规实践使用合成的寡核苷酸衔接子或接头。因此,术语可操作地连接通常是指核酸序列在单个核酸片段上的缔合(并置),这样使得一个核酸片段的功能受到另一个的影响。例如,如果启动子(pro)控制基因编码序列(基因CDS)的转录,则该启动子可操作地连接到基因CDS(例如,5'-[pro]-[基因CDS]3')。

[0094] 如本文使用的,短语“变体rrnI-P2启动子和5'-UTR区”(缩写为“P2启动子/5'-UTR区”)是指SEQ ID NO:1所示的变体枯草芽孢杆菌rrnI-P2启动子/5'-aprE UTR区DNA序列。

[0095] 如本文使用的,编码“野生型枯草芽孢杆菌aprE信号肽序列”的DNA可缩写为“aprESS”,并且包含SEQ ID NO:4的核苷酸序列。

[0096] 如本文使用的,“野生型解淀粉芽孢杆菌BPN'终止子(BPN'term)”可缩写为“term”,并且包含SEQ ID NO:6的核苷酸序列。

[0097] 如本文使用的,包含SEQ ID NO:3的氨基酸序列的变体吉氏芽孢杆菌(BG46)枯草杆菌蛋白酶缩写为“BG46_变体1”。

[0098] 如本文使用的,包含SEQ ID NO:8的氨基酸序列的变体吉氏芽孢杆菌(BG46)枯草杆菌蛋白酶缩写为“BG46_变体2”,该BG46_变体2衍生自野生型吉氏芽孢杆菌(BG46)枯草杆菌蛋白酶报告蛋白(SEQ ID NO:2)。

[0099] 如本文使用的,“合适的调控序列”是指位于编码序列上游(5'非编码序列)、内部或下游(3'非编码序列),并且影响相关编码序列的转录、RNA加工或稳定性或者翻译的核苷酸序列。调控序列可以包括启动子、转录前导序列、RNA加工位点、效应子结合位点和茎环结构。

[0100] 如本文使用的,“宿主细胞”是指具有作为新引入的DNA序列的宿主或表达媒介物的能力的细胞。因此,在本公开的某些实施例中,宿主细胞是革兰氏阳性细胞(例如,芽孢杆菌属物种)和/或革兰氏阴性细胞(例如,大肠杆菌)。

[0101] 如本文使用的,“经修饰的细胞”是指包含至少一个遗传修饰的重组细胞,该遗传修饰不存在于衍生出经修饰的细胞的亲本细胞、参比细胞、或对照细胞中。

[0102] 如本文使用的,当将重组(经修饰的)细胞中目的蛋白(POI)的表达和/或生产与未经修饰的(对照)细胞中相同POI的表达和/或生产相比较时,应理解,在相同条件(例如,相同条件如培养基、温度、pH等)下生长/培养/发酵经修饰的和未经修饰的细胞。

[0103] 如本文使用的,当在短语如“相对于未经修饰的(对照)细胞,重组细胞‘表达/生产增加量’的目的蛋白”中使用,“增加量”特别是指在重组细胞中表达/生产的“增加量”目的蛋白(POI),该“增加量”总是相对于表达/生产相同POI的未经修饰的(对照)细胞,其中经修饰的和未经修饰的细胞在相同条件下生长/培养/发酵。

[0104] 如本文使用的,“增加”蛋白质生产或“增加的”蛋白质生产意指生产的蛋白质(例如,目的蛋白)的量增加。蛋白质可以在宿主细胞内产生,或被分泌(或转运)到培养基中。在某些实施例中,目的蛋白产生(被分泌)到培养基中。增加的蛋白质生产可以被检测为例如,如与亲本宿主细胞相比更高的最大水平的蛋白质或酶活性(例如像淀粉酶活性),或生产的总的细胞外蛋白质。

[0105] 如本文使用的,术语“修饰”和“遗传修饰”可互换使用,并且包括:(a)引入、取代或去除基因(或其ORF)中的一个或多个核苷酸,或者引入、取代或去除基因或其ORF的转录或翻译所需的调控元件中的一个或多个核苷酸,(b)基因破坏,(c)基因转换,(d)基因缺失,(e)基因下调,(f)本文公开的任何一个或多个基因的特异性诱变和/或(g)随机诱变。

[0106] 在某些方面,遗传修饰特别是指在编码本公开的前区(氨基酸)序列的核酸(DNA)序列中引入、取代或去除一个或多个核苷酸。例如,在某些实施例中,如本文所述对编码SEQ ID NO:15所示的天然前区氨基酸序列的DNA序列进行遗传修饰。

[0107] 如本文使用的,如在诸如“向革兰氏阳性细菌细胞中引入‘基因’、‘多核苷酸’、‘可读框(ORF)’、‘基因编码序列’、‘载体’、‘表达盒’”等的短语中使用的,术语“引入”包括本领域中已知用于将多核苷酸(DNA)引入细胞中的方法,包括但不限于原生质体融合、自然或人工转化(例如,氯化钙、电穿孔)、转导、转染、缀合等。

[0108] 如本文使用的,“转化的”或“转化”意指通过使用重组DNA技术转化细胞。转化典型地通过将一个或多个核苷酸序列(例如,多核苷酸、ORF、或基因)插入细胞中而发生。插入的核苷酸序列可以是异源核苷酸序列(即,在待转化的细胞中非天然存在的序列)。因此,转化通常是指将外源DNA引入宿主细胞中,使得DNA保持为染色体整合体或自我复制的染色体外

载体。

[0109] 如本文使用的,“转化DNA”、“转化序列”和“DNA构建体”是指用于将序列引入宿主细胞或生物体中的DNA。转化DNA是用于将序列引入宿主细胞或生物体中的DNA。可以通过PCR或任何其他合适的技术在体外产生DNA。在一些实施例中,转化DNA包含输入序列,而在其他实施例中,它进一步包含侧翼为同源盒的输入序列。在另外的实施例中,转化DNA包含添加到末端的其他非同源序列(即,填充序列或侧翼)。末端可以闭合,使得转化DNA形成闭环,例如像插入载体中。

[0110] 如本文使用的,“基因的破坏”或“基因破坏”可以互换使用,并且广泛地是指基本上防止宿主细胞生产功能性基因产物(例如,蛋白质)的任何遗传修饰。因此,如本文使用的,基因破坏包括但不限于移码突变、提前终止密码子(即,使得不生产功能性蛋白质)、消除或降低活性的蛋白质内部缺失的取代(使得不生产功能性蛋白质)、破坏编码序列的插入,去除转录所需的天然启动子与可读框之间的可操作连接的突变等。

[0111] 如本文使用的,“输入序列”是指引入细菌细胞染色体中的DNA序列。在一些实施例中,输入序列是DNA构建体的一部分。在其他实施例中,输入序列编码一种或多种目的蛋白。在一些实施例中,输入序列包含可以已经或可以不存在于待转化的细胞的基因组中的序列(即,它可以是同源或异源序列)。在一些实施例中,输入序列编码一种或多种目的蛋白、基因和/或突变的或修饰的基因。在可替代的实施例中,输入序列编码功能性野生型基因或操纵子、功能性突变基因或操纵子、或者非功能性基因或操纵子。在一些实施例中,可以将非功能性序列插入基因中以破坏基因的功能。在另一个实施例中,输入序列包括选择性标记。在另外的实施例中,输入序列包括两个同源盒。

[0112] 如本文使用的,“同源盒”是指与细菌细胞染色体中的序列同源的核酸序列。更特别地,根据本发明,同源盒是上游或下游区,该上游或下游区与待缺失、破坏、失活、下调等的基因或基因的一部分的直接侧翼编码区具有约80%与100%之间的序列同一性、约90%与100%之间的序列同一性、或约95%与100%之间的序列同一性。这些序列指导在细菌细胞染色体中DNA构建体的整合位置,并且指导染色体的哪部分被输入序列替代。虽然不意在限制本公开,但同源盒可以包括约1个碱基对(bp)至200千碱基(kb)之间。优选地,同源盒包括约1bp与10.0kb之间;1bp与5.0kb之间;1bp与2.5kb之间;1bp与1.0kb之间;以及0.25kb与2.5kb之间。同源盒还可以包括约10.0kb、5.0kb、2.5kb、2.0kb、1.5kb、1.0kb、0.5kb、0.25kb和0.1kb。在一些实施例中,选择性标记的5'和3'端的侧翼为同源盒,其中该同源盒包含与基因的编码区紧密侧接的核酸序列。

[0113] 如本文使用的,宿主细胞“基因组”、细菌(宿主)细胞“基因组”、或芽孢杆菌属物种(宿主)细胞“基因组”包括染色体基因和染色体外基因。

[0114] 如本文使用的,术语“质粒”、“载体”和“盒”是指染色体外元件,其通常携带典型地不是细胞的中心代谢的一部分的基因,并且通常呈环状双链DNA分子的形式。这样的元件可以是衍生自任何来源的单链或双链DNA或RNA的线性或环状自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其中多种核苷酸序列已连接或重组到单一构造中,该单一构造能够将用于选定基因产物的启动子片段和DNA序列连同适当3'非翻译序列一起引入细胞中。

[0115] 如本文使用的,术语“质粒”是指用作克隆载体,并且在许多细菌和一些真核生物中形成染色体外自我复制遗传元件的环状双链(ds)DNA构建体。在一些实施例中,质粒被掺

入宿主细胞的基因组中。在一些实施例中,质粒存在于亲本细胞中并且在子细胞中丢失。

[0116] 如本文使用的,“转化盒”是指包含基因(或其ORF)并且除了外源基因之外还具有促进特定宿主细胞的转化的元件的特定载体。

[0117] 如本文使用的,术语“载体”是指可以在细胞中复制(传播)并且可以携带新基因或DNA区段到细胞中的任何核酸。因此,该术语是指设计用于在不同宿主细胞之间转移的核酸构建体。载体包括为“附加体”(即,其自主复制或可以整合到宿主生物体的染色体中)的病毒、噬菌体、前病毒、质粒、噬菌粒、转座子、和人工染色体诸如YAC(酵母人工染色体)、BAC(细菌人工染色体)、PLAC(植物人工染色体)等。

[0118] “表达载体”是指具有掺入在细胞中并在细胞中表达异源DNA的能力的载体。许多原核和真核表达载体是可商购获得的并且是本领域技术人员已知的。适当的表达载体的选择在本领域技术人员的知识范围内。

[0119] 如本文使用的,术语“表达盒”和“表达载体”是指重组或合成生产的具有允许特定核酸在靶细胞中转录的一系列指定核酸元件(即,这些是载体或载体元件,如上所述)的核酸构建体。可以将重组表达盒掺入质粒、染色体、线粒体DNA、质体DNA、病毒、或核酸片段中。典型地,表达载体的重组表达盒部分包括(除了其他序列之外)待转录的核酸序列和启动子。在一些实施例中,DNA构建体还包括允许特定核酸在靶细胞中转录的一系列指定核酸元件。在某些实施例中,本公开的DNA构建体包含如本文定义的选择性标记和失活染色体、或基因、或DNA区段。

[0120] 如本文使用的,“靶向载体”是如下载体,该载体包括与该靶向载体转化至其中的宿主细胞的染色体中的区同源并且可以驱动在该区处的同源重组的多核苷酸序列。例如,靶向载体可用于通过同源重组将突变引入宿主细胞的染色体中。在一些实施例中,靶向载体包含例如添加到末端的其他非同源序列(即,填充序列或侧翼序列)。末端可以闭合,使得靶向载体形成闭环,例如像插入载体中。例如,在某些实施例中,通过在亲本地衣芽孢杆菌(宿主)细胞中引入一个或多个“靶向载体”来修饰(例如转化)该细胞。

[0121] 如本文使用的,术语“目的蛋白”或“POI”是指期望在经修饰的(重组)革兰氏阳性宿主细胞中表达的目的多肽,其中该POI优选地以增加的水平表达(即,相对于“未经修饰的”(亲本或对照)细胞)。因此,如本文使用的,POI可以是酶、底物结合蛋白、表面活性蛋白、结构蛋白、受体蛋白等。在某些实施例中,相对于对照细胞,本公开的经修饰的细胞产生增加量的异源目的蛋白。在特定的实施例中,由本公开的经修饰的细胞产生的目的蛋白的增加量是相对于对照细胞至少0.5%增加、至少1.0%增加、至少5.0%增加、或超过5.0%增加。

[0122] 类似地,如本文定义的,“目的基因”或“GOI”是指编码POI的核酸序列(例如,多核苷酸、基因、或ORF)。编码“目的蛋白”的“目的基因”可以是天然存在的基因、突变的基因或合成基因。

[0123] 如本文使用的,术语“多肽”和“蛋白质”可互换使用,并且是指包含通过肽键连接的氨基酸残基的任何长度的聚合物。在本文中使用氨基酸残基的常规一(1)字母或三(3)字母代码。多肽可以是线性的或支链的,它可以包含经修饰的氨基酸,并且它可以被非氨基酸中断。术语多肽还涵盖如下氨基酸聚合物,该氨基酸聚合物已经天然地或通过干预修饰;例如,二硫键形成、糖基化、脂化、乙酰化、磷酸化或任何其他操作或修饰,诸如与标记组分缀

合。该定义内还包括例如含有一种或多种氨基酸类似物(包括例如非天然氨基酸等)以及本领域已知的其他修饰的多肽。

[0124] 在某些实施例中,本公开的基因编码商业上相关的工业目的蛋白,诸如酶(例如,乙酰酯酶、氨肽酶、淀粉酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、碳酸酐酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡糖苷酶、 β -葡糖苷酶、葡糖醛酸酶、糖基水解酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲基酯酶、果胶分解酶、过水解酶、多元醇氧化酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、肽酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶、及其组合)。

[0125] 如本文使用的,“变体”多肽是指典型地通过重组DNA技术,通过取代、添加、或缺失一个或多个氨基酸,衍生自亲本(或参比)多肽的多肽。变体多肽可能与亲本多肽有少量氨基酸残基的差异,并且可以通过它们与亲本(参比)多肽的一级氨基酸序列同源性/同一性的水平来定义。

[0126] 在某些一个或多个实施例中,变体多肽与亲本(参比)多肽序列具有至少约40%至约70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或甚至至少99%的氨基酸序列同一性。

[0127] 如本文使用的,“变体”多核苷酸是指与亲本(或参比)多核苷酸具有指定程度的序列同源性/同一性的多核苷酸,或者与亲本多核苷酸(或其互补序列)在严格杂交条件下杂交的多核苷酸。在某些一个或多个实施例中,本公开的变体多核苷酸与亲本(参比)多核苷酸序列具有至少约40%到至少约100%的核苷酸序列同一性。在某些其他实施例中,变体多核苷酸与亲本(参比)多核苷酸序列具有至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、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91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或至少99%的核苷酸序列同一性。

[0128] 如本文使用的,“突变”是指核酸序列中的任何变化或改变。存在几种类型的突变,包括点突变、缺失突变、沉默突变、移码突变、剪接突变等。突变可以特异性地(例如,经由定点诱变)或随机地(例如,经由化学试剂、通过修复减去细菌菌株传代)进行。

[0129] 如本文使用的,在多肽或其序列的上下文中,术语“取代”意指一个氨基酸被另一个氨基酸替代(即,取代)。

[0130] 如本文使用的,术语“同源性”涉及同源多核苷酸或多肽。如果两个或更多个多核苷酸或者两个或更多个多肽是同源的,则这意味着同源多核苷酸或多肽具有至少60%、更优选至少70%、甚至更优选至少85%、仍更优选至少90%、更优选至少95%、以及最优选至少98%的“同一性程度”。两个多核苷酸或多肽序列是否具有如本文定义的同源程度足够高的同一性,可以通过使用本领域已知的计算机程序比对两个序列来适当地研究,该计算机程序如GCG程序包中提供的“GAP”(威斯康星程序包手册(Program Manual for the Wisconsin Package),第8版,1994年8月,遗传学计算机集团(Genetics Computer Group),科学大道575号(575 Science Drive),麦迪逊,威斯康星州,美国53711)(Needleman和

Wunsch, (1970))。使用具有以下设置的GAP进行DNA序列比较:GAP产生罚分5.0和GAP延伸罚分0.3。

[0131] 如本文使用的,术语“同一性百分比(%)”是指当使用序列比对程序进行比对时,编码多肽的核酸序列或多肽的氨基酸序列之间的核酸或氨基酸序列同一性的水平。

[0132] 如本文使用的,“比生产力”是在给定时间段内生产的蛋白质的总量/细胞/时间。

[0133] 如本文使用的,术语“纯化的”、“分离的”或“富集的”意指生物分子(例如,多肽或多核苷酸)通过将其与它在自然界中相关联的天然存在的成分中的一些或所有分离而从其天然状态改变。这样的分离或纯化可以通过本领域公认分离技术诸如离子交换色谱法、亲和色谱法、疏水分离、透析、蛋白酶处理、硫酸铵沉淀或其他蛋白质盐沉淀、离心、尺寸排阻色谱法、过滤、微量过滤、凝胶电泳、或梯度分离完成,以去除最终组合物中所不期望的全细胞、细胞碎片、杂质、外来蛋白质、或酶。然后可以进一步向纯化的或分离的生物分子组合物中添加提供另外益处的成分,例如激活剂、抗抑制剂、期望的离子、控制pH的化合物、或者其他酶或化学品。

[0134] II. 用于增强蛋白质生产的突变PRO序列

[0135] 如本文总体所述和下文实例中阐述,申请人设计并构建了位点评估文库(SEL)来测试/筛选某些遗传修饰(突变)以提高重组蛋白的生产力。特别地,如实例中阐述,利用重组蛋白(酶)作为报告蛋白来监测本公开的重组芽孢杆菌属菌株中的蛋白表达。已经描述了适合用于异源(重组)成熟蛋白的表达/生产的某些前区序列。例如,PCT公开号WO 2008/112258和PCT公开号2010/123754描述了克劳氏芽孢杆菌碱性前体蛋白酶和迟缓芽孢杆菌枯草杆菌蛋白酶前体蛋白酶的前序列,其中某些突变前序列促使成熟蛋白酶的生产力增强。PCT公开号WO 1993/20214描述了用于在枯草芽孢杆菌中表达革兰氏阴性脂肪酶的某些AprE信号(pre)和前区序列。

[0136] 更特别地,如下文实例所述,对编码示例性(枯草杆菌蛋白酶)报告蛋白的表达构建体的前区进行遗传修饰(SEL)。特别地,如实例中总体阐述(参见图1-图3)对衍生自野生型(参比)前序列(SEQ ID NO:15)的前区序列进行SEL,以筛选用于增强成熟枯草杆菌蛋白酶报告蛋白生产力的突变。

[0137] 如实例1所述,以编码变体前区序列A(30H,SEQ ID NO:9;图1)的氨基酸序列为模板建立位点评估文库(SEL),并开发成4.4kb的片段,其中表达盒的线性DNA用于转化感受态芽孢杆菌属细胞。如实例1所述,进行序列分析以确定SEL中独特的前区序列A变体和/或前区序列B变体,其中变体前区序列A(SEQ ID NO:9)的84个氨基酸位置中的每一个都被改变(取代)为所有其他十九(19)个天然存在的氨基酸残基。特别地,表1(实例1)显示七十二(72)小时后相比于使用WT PRO序列(SEQ ID NO:15)表达的报告蛋白的参比构建体,报告蛋白生产力(性能指数(PI)值)的结果。例如,如表1显示,前区序列A(E30H,SEQ ID NO:9)的PI相对于WT PRO序列(30E;SEQ ID NO:15)的PI增加约50%。同样,如表1所示,氨基酸位置1、2、3、4、6、14、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、83或84中的任一个与位点30组合的突变在72小时相对于参比前区序列(SEQ ID NO:15)显示出增强的报告蛋白生产力。此外,约80%的选择突变体组合具有额外的正电荷(见表1,Δ电荷)。

[0138] 如实例3进一步所述,对编码示例性(枯草杆菌蛋白酶)报告蛋白(BG46_变体2)的表达构建体前区进行遗传修饰(SEL)。更特别地,以编码变体前区序列B(30G,SEQ ID NO:

11;图1)的氨基酸序列为模板来开发包含本公开的一个或多个前区的线性DNA表达盒。特别地,七十八(78)个氨基酸的野生型AprE前肽序列(SEQ ID NO:12)的5'序列用于交换前三(3)个氨基酸残基(位置1-3,“AGK”;SEQ ID NO:12)与作为丙氨酸(A)的第一氨基酸残基,产生变体前区序列C(SEQ ID NO:14;GK插入,图2)。例如,如表1(实例1)显示,前区变体A序列(SEQ ID NO:9;表1,突变体“E30G”)的突变体显示相比于野生型前区(表1,PI 1.0)和前区变体A序列(SEQ ID NO:11;表1,突变体“E30H”,PI 1.5)在生长72小时后报告蛋白的高表达(PI 2.3)。

[0139] 本公开的实例4进一步描述了基于参比前区序列(SEQ ID NO:15)的组合文库的制备,其中插入氨基酸甘氨酸(G)和赖氨酸(K) (“GK”)。更特别地,如实例3所述,将氨基酸残基GK插入野生型前区序列(SEQ ID NO:15)的N-末端,从而产生SEQ ID NO:14所示的前区变体C氨基酸序列,该序列包含86个氨基酸位置,如图2比对中所示。例如,表2(实例4)呈现了显示BG46_变体2报告蛋白生产力(PI)增加的前区突变组合,其中电荷的变化(Δ 电荷)总体上在+1和+4之间(表2)。在某些方面,申请人注意到这样的正电荷适合作为前区序列的环区中的组合(例如像参比SEQ ID NO:15的残基位置66-73)。

[0140] 如表2显示,将SEL前区变体C序列(衍生自SEQ ID NO:14)的PI值与野生型(WT)前区序列(SEQ ID NO:15)的PI值进行比较,其中氨基酸位置编号(表2,第1列)与WT前区序列(即,没有“AGK”插入)进行比较。因此,如实例4所述,与野生型前区(表2,PI 1.0)和前区变体B序列(突变体“E30G”,PI 1.7)相比,在生长72小时后,表2所示的前区变体C序列显示报告蛋白的高表达(PI值介于约1.9与2.3之间)。

[0141] 如实例5总体上所述,BG46_变体2枯草杆菌蛋白酶用作报告蛋白来监测在地衣芽孢杆菌菌株中的表达。更特别地,编码本公开的变体前区序列的DNA可操作地连接到编码成熟BG46_变体2蛋白的DNA序列,其中菌株构建如实例5总体上所述。例如,表3显示包含变体前区序列的地衣芽孢杆菌菌株相对于包含E30G突变的对照前区序列B(SEQ ID NO:11)的PI值,其中在生长72小时后变体前区序列相比于对照前区序列(E30G)表现出报告蛋白的高表达。

[0142] 如实例6进一步所述,对变体前区序列C(SEQ ID NO:14)进行另外的N-末端修饰。例如,进一步修饰变体前区序列C(有86个氨基酸位置),以便在位置三(3)中赖氨酸(K)之后引入(插入)另外的氨基酸,如丙氨酸(A)或丝氨酸(S)。特别地,如实例6所述(且如图1-图3所示),构建变体前区序列D(SEQ ID NO:29,GKS插入)、E(SEQ ID NO:30,GKA插入)和F(GKAA插入;SEQ ID NO:31),其中将前区变体序列D、E和F的性能指数(PI)值(表4)与变体前区序列C(SEQ ID NO:14,“GK”插入)进行比较。如表4呈现,在生长72小时后,三种N-末端前区突变体(GKA、GKS、GKAA)的PI指数显示与前区序列C(GK;SEQ ID NO:14)相比报告蛋白BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)的表达升高。

[0143] 同样,如实例7所述对前区序列进行某些C-末端修饰。特别地,对变体前区序列B(SEQ ID NO:11)进行进一步工程化,以将位置六十八(68)处的亮氨酸(L)残基取代为赖氨酸(K),并将位置七十二(72)处的异亮氨酸(I)取代为缬氨酸(V),从而产生变体前区序列G(SEQ ID NO:32),或者对其进行工程化以将位置六十八(68)处的亮氨酸(L)取代为赖氨酸(K),并将位置七十二(80)处的谷氨酸(E)取代为异亮氨酸(I),从而产生变体前区序列F(SEQ ID NO:33),如图1-图3所示。如表5呈现,与参比(对照)变体前区B序列(SEQ ID NO:

11)相比,变体前区序列G(SEQ ID NO:32)和H(SEQ ID NO:33)的PI值增加。

[0144] 因此,本公开的某些实施例尤其提供了编码新型前区序列的重组多核苷酸、包含可操作地连接到编码成熟目的蛋白的DNA序列的编码新型前区序列的DNA序列的表达盒、表达编码包含新型前区序列的前体蛋白的多核苷酸的经修饰的革兰氏阳性细菌细胞/菌株等。在某些方面,本公开的新型变体(突变体)前区序列包含衍生自SEQ ID NO:15的野生型迟缓芽孢杆菌前序列的氨基酸序列。

[0145] 如本文使用的,短语“变体前区”是指典型地通过重组DNA技术通过取代、添加、或缺失一个或多个氨基酸,衍生自参比前区序列的多肽序列。变体前区(氨基酸)序列可能与参比(亲本)前区序列有少量氨基酸残基的差异,并且可以通过它们与参比(亲本)前区序列的一级氨基酸序列同源性/同一性水平来定义。

[0146] 如上文所阐述,在两个多核苷酸或多肽序列的上下文中,术语“同一”是指在两个序列中的核苷酸或氨基酸在比对最大对应关系时相同,如使用以下描述的和本领域已知的序列比较或分析算法测量的。短语“同一性百分比(%)”(缩写为“PID”)是指多核苷酸(核酸)或多肽(氨基酸)序列同一性。同一性百分比可以使用本领域已知的标准技术来确定。在某些方面,可以通过比对序列以直接比较序列信息来确定目的序列共有的氨基酸同一性百分比,例如通过使用例如BLAST、MUSCLE、或CLUSTAL等比对程序/算法。例如,Altschul等人(1990)和Karlin等人(1993)描述了BLAST算法。特别地,氨基酸序列同一性百分比(%)值由匹配相同残基的数目除以“参比”序列的残基总数(包括由程序为最佳/最大比对创建的任何空位)来确定。BLAST算法将“参比”序列称为“查询”序列。

[0147] 如本文使用的,“同源前区”是指在一级、二级和/或三级结构中具有不同相似性的前区。当比对蛋白质时,蛋白质同源性可指线性氨基酸序列的相似性。同源性可以通过例如使用诸如BLAST、MUSCLE或CLUSTAL的程序进行的氨基酸序列比对来确定。蛋白序列的同源搜索可以使用来自NCBI BLAST的BLASTP和PSI-BLAST使用阈值(E值截止值)0.001进行(例如,参见Altschul等人,1997)。

[0148] BLAST程序使用几个搜索参数,其中大部分被设置为默认值。NCBI BLAST算法找到在生物相似性方面最相关的序列,但是不推荐用于少于20个残基的查询序列(Altschul等人,1997和Schaffer等人,2001)。用于核酸序列搜索的示例性默认BLAST参数包括:相邻字长阈值=11;E值截止值=10;评分矩阵(Scoring Matrix)=NUC.3.1(匹配=1,错配=-3);空位开放=5;以及空位延伸=2。用于氨基酸序列搜索的示例性默认BLAST参数包括:字长=3;E值截止值=10;评分矩阵=BLOSUM62;空位开放=11;以及空位延伸=1。使用该信息,可以将蛋白序列分组和/或从中构建系统发生树。可以在例如Vector NTI Advance套件等程序中输入氨基酸序列,并且可以使用邻接(Neighbor Joining(NJ))法创建引导树(Saitou和Nei,1987)。可以使用针对序列距离的Kimura校正并且忽略具有空位的位置来计算树结构。程序例如AlignX可以在系统发生树上显示的分子名称之后的括号内显示计算的距离值。

[0149] 了解分子之间的同源性可以揭示分子的进化历史以及它们的功能信息;如果新测序的蛋白质与已经表征的蛋白质同源,则存在新蛋白质的生化功能的强烈指示。如果两个分子衍生自共同的祖先,则它们被称为是同源的。同源分子或同源物可以分为两类:旁系同源物和直系同源物。旁系同源物是存在于一个物种内的同源物。旁系同源物在它们的详细

生化功能上往往不同。直系同源物是存在于不同物种内并且具有非常相似或相同功能的同源物。蛋白质超家族是可以推断出共同祖先的蛋白质的最大分组(进化枝)。通常这个共同祖先是基于序列比对和机械相似性。

[0150] CLUSTAL W算法是序列比对算法的另一个实例(Thompson等人,1994)。CLUSTAL W算法的默认参数包括: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;空位分隔距离=8;DNA转换权重=0.50;列表亲水残基=GPSNDQEK;使用负性矩阵=关;切换特殊残基罚分=开;切换亲水罚分=开;以及切换结束空位分隔罚分=关。在CLUSTAL算法中,包括在任一末端发生的缺失。例如,在500个氨基酸的多肽的任一末端(或多肽内)具有五个氨基酸缺失的变体相对于“参比”多肽将具有99%(495/500个相同的残基 $\times$ 100)的序列同一性百分比。这样的变体将由与该多肽具有“至少99%序列同一性”的变体所涵盖。

[0151] 在某些实施例中,变体前区序列与本公开的参比(亲本)前区序列具有至少约40%至约70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或甚至至少99%氨基酸序列同一性。在某些实施例中,前区序列衍生自与SEQ ID NO:15、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32和/或SEQ ID NO:33包含同源性的氨基酸序列。因此,在某些方面,通过参比SEQ ID NO:15、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32和/或SEQ ID NO:33的前区氨基酸序列对本文所述的一个或多个前区变体的氨基酸修饰进行编号。

[0152] 在某些方面,新型(变体)前区序列衍生自亲本(参比)前区序列,该亲本(参比)前区序列与SEQ ID NO:15的氨基酸序列具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。在某些实施例中,根据SEQ ID NO:15氨基酸(位置)编号,变体前区序列在氨基酸(残基)位置30以及选自由以下位置组成的组的一个或多个氨基酸残基位置处包含突变:1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84。

[0153] 在其他实施例中,本公开的变体前区序列包含位置2处的甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入,以及选自位置1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,其中氨基酸位置根据SEQ ID NO:14编号。在某些实施例中,新型(变体)前区序列衍生自参比前区序列,该参比前区序列与SEQ ID NO:14的氨基酸序列具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0154] 在其他实施例中,本公开的变体前区序列包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丝氨酸(S)的氨基酸插入,其中氨基酸位置根据SEQ ID NO:29编号。在某些实施例中,新型(变体)前区序列衍生自参比前区序列,该参比前区序列与SEQ ID NO:29的氨基酸序列具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0155] 在其他实施例中,本公开的变体前区序列包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丙氨酸(A)的氨基酸插入,其中氨基酸位置根据SEQ ID NO:30编号。在某些实施例中,变体前区序列衍生自参比前区序列,该参比前区序列与SEQ ID NO:30的氨

氨基酸序列具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0156] 在其他实施例中,本公开的变体前区序列包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)、位置4处的丙氨酸(A)和位置5处的丙氨酸(A)的氨基酸插入,其中氨基酸位置根据SEQ ID NO:31编号。在某些实施例中,变体前区序列衍生自参比前区序列,该参比前区序列与SEQ ID NO:31的氨基酸序列具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0157] 在某些其他实施例中,变体前区序列衍生自亲本(参比)前区序列,该亲本(参比)前区序列与SEQ ID NO:11的氨基酸序列(G30;前区B序列)具有至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性。

[0158] 在某些实施例中,变体前区序列包含选自位置68、位置72和位置80的位置处的至少一个氨基酸取代,其中氨基酸位置根据SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:15、SEQ ID NO:32、或SEQ ID NO:33进行编号。

[0159] 在某些其他实施例中,变体前区序列在氨基酸(残基)位置30以及选自由以下位置组成的组的一个或多个氨基酸残基位置处包含突变:1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84,其中根据SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:15、SEQ ID NO:32、或SEQ ID NO:33氨基酸(位置)编号。

[0160] 在某些实施例中,对变体前区序列进行设计、工程化、构建等,使得该前区序列的净电荷相对于参比(亲本)前区序列(如SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32和SEQ ID NO:33的参比(亲本)序列)是正电荷(例如,+1净电荷)。在某些实施例中,对本公开的变体前区序列进行设计、工程化、构建等,使得前区序列的净电荷是至少正1(+1)至大约正4(+4)。在其他一个或多个实施例中,本公开的变体前区序列包含表1、表2、表3、表4和/或表5所示的一个或多个氨基酸修饰。

[0161] III. 重组多核苷酸和分子生物学

[0162] 如上文总体所述,本公开的某些实施例涉及新型突变前区序列。在相关方面,本公开提供了重组多核苷酸,其包含一个或多个突变前区核酸(DNA)序列。因此,某些实施例涉及重组多核苷酸(例如,载体、质粒、表达盒等)、表达目的蛋白的重组革兰氏阳性细菌细胞/菌株等。在某些方面,本公开提供了适合引入重组革兰氏阳性细菌细胞(菌株)中用于增强目的蛋白生产的多核苷酸构建体。在某些方面,本公开的多核苷酸构建体被称为表达盒,其中该盒在5'至3'方向上并且在可操作组合中包含至少上游(5')前区DNA序列,该前区DNA序列连接到编码成熟目的蛋白(POI)的下游(3')基因CDS。

[0163] 例如,本文所述的一种或多种核酸序列可以通过使用任何合适的合成、操作和/或分离技术或其组合来产生。例如,本文所述的一种或多种多核苷酸可以使用本领域技术人员熟知的标准核酸合成技术如固相合成技术来产生。在此类技术中,典型地合成多达五十(50)个或更多个核苷酸碱基的片段,然后连接(例如通过酶或化学连接方法)以基本上形成任何希望的连续核酸序列。还可以通过本领域已知的任何合适的方法来促进本文描述的一种或多种多核苷酸的合成,包括但不限于使用以下方法的化学合成:经典的亚磷酰胺方法

(例如,Beaucage和Caruthers,1981),或如典型地在自动合成方法中所实践的Matthes等人(1984)描述的方法。本文所述的一种或多种多核苷酸也可以通过使用自动DNA合成仪产生。可以从各种商业来源(例如,ATUM(DNA 2.0),美国加利福尼亚州纽瓦克(Newark,CA,USA);生命技术公司(Life Tech)(GeneArt),美国加利福尼亚州卡尔斯巴德(Carlsbad,CA,USA);金斯瑞公司(GenScript),加拿大安大略省(Ontario,Canada);Base Clear B.V.,荷兰莱顿市(Leiden,Netherlands);集成DNA技术公司(Integrated DNA Technologies),美国伊利诺伊州斯科基(Skokie,IL,USA);银杏生物工作室(Ginkgo Bioworks)(Gen9),美国马萨诸塞州波士顿(Boston,MA,USA);以及特威斯特生物科学公司(Twist Bioscience),美国加利福尼亚州旧金山(San Francisco,CA,USA))订购定制核酸。用于合成核酸的其他技术和相关原理是本领域中描述和已知的。

[0164] 用于修饰核酸的重组DNA技术是本领域熟知的,例如像限制性内切核酸酶消化、连接、逆转录和cDNA产生、以及聚合酶链式反应(例如,PCR)。本文所述的一种或多种多核苷酸还可以通过使用一个或多个寡核苷酸探针来筛选cDNA文库而获得,探针可以与编码本文所述的一种或多种变体的多核苷酸杂交或对其进行PCR扩增。用于筛选并分离cDNA克隆的程序以及PCR扩增程序是本领域技术人员熟知的并且描述于本领域技术人员已知的标准参考文献中。可以通过例如已知的诱变程序(例如,定点诱变、位点饱和诱变、以及体外重组)改变天然存在的多核苷酸主链(例如,编码本文所述的一种或多种变体前区序列)来获得本文所述的一种或多种多核苷酸。适合于产生编码本文所述的一种或多种变体的经修饰的本文所述的多核苷酸的多种方法是本领域已知的,这些方法包括但不限于例如位点饱和诱变、扫描诱变、插入诱变、缺失诱变、随机诱变、定点诱变和定向进化、以及各种其他重组方法。

[0165] 如上文总体上所述和下文实例中进一步描述的,本公开的某些实施例涉及能够生产增加量的异源目的蛋白的重组(经修饰的)革兰氏阳性细胞。因此,某些实施例涉及用于构建此类具有增加的蛋白质生产能力的重组革兰氏阳性细胞的方法。在某些实施例中,将编码目的蛋白的一种或多种表达盒引入本公开的革兰氏阳性细胞中。在示例性实施例中,将这些盒整合到细胞的基因组中。因此,某些实施例涉及适用于构建重组(经修饰的)革兰氏阳性宿主细胞的核酸分子、多核苷酸(例如,载体、质粒、表达盒)、调控元件等。

[0166] 因此,如实例中所示和本文总体上所述,可以由本领域技术人员使用本领域熟知的标准和常规重组DNA和分子克隆技术来构建本公开的重组细胞。用于遗传修饰的方法包括但不限于(a)在基因中引入、取代、或去除一个或多个核苷酸,或者在基因的转录或翻译所需的调控元件中引入、取代、或去除一个或多个核苷酸,(b)基因破坏,(c)基因转换,(d)基因缺失,(e)基因下调,(f)位点特异性诱变,和/或(g)随机诱变。

[0167] 在某些实施例中,可以通过使用本领域熟知的方法(例如,插入、破坏、替代或缺失)减少或消除基因的表达来构建本公开的经修饰的细胞。待修饰或失活的基因的部分可以是例如编码区或编码区表达所需的调控元件。

[0168] 此类调控或控制序列的实例可以是启动子序列或其功能性部分(即,足以影响核酸序列表达的部分)。用于修饰的其他控制序列包括但不限于前导序列、前肽序列、信号序列、转录终止子、转录激活子等。

[0169] 在某些其他实施例中,通过基因缺失构建经修饰的细胞以消除或减少基因的表达。基因缺失技术使得能够部分或完全去除一个或多个基因,从而消除它们的表达,或者表

达非功能性(或活性降低的)蛋白质产物。在此类方法中,一个或多个基因的缺失可以通过使用已构建为连续含有与该基因侧接的5'和3'区的质粒进行同源重组来完成。可以将连续的5'和3'区引入细胞中,例如,在温度敏感的质粒上,在允许温度下与第二可选择标记结合以允许质粒在细胞中建立。然后将细胞转移到不允许温度,以选择将质粒整合到染色体同源侧翼区之一的细胞。通过选择第二可选择标记来影响质粒整合的选择。在整合后,通过将细胞移至允许温度持续几代而不进行选择来刺激第二同源侧翼区处的重组事件。将细胞铺板以获得单菌落,并检查菌落是否损失两种可选择标记。因此,本领域的技术人员可容易地鉴定在基因的编码序列和/或基因的非编码序列中的核苷酸区(适于完全或部分缺失)。

[0170] 在其他实施例中,通过在基因或者其转录或翻译所需的调控元件中引入、取代或去除一个或多个核苷酸来构建经修饰的细胞。例如,可以插入或去除核苷酸,以便导致终止密码子的引入、起始密码子的去除或可读框的移码。这样的修饰可以根据本领域已知的方法通过定点诱变或PCR产生的诱变来完成。因此,在某些实施例中,通过完全或部分缺失使本公开的基因失活。

[0171] 在另一个实施例中,通过基因转换工艺构建经修饰的细胞。例如,在基因转换方法中,对应于一个或多个基因的核酸序列在体外诱变以生产缺陷核酸序列,然后将该缺陷核酸序列转化到亲本细胞中以产生缺陷基因。通过同源重组,缺陷核酸序列替代内源基因。可能期望的是,缺陷基因或基因片段也编码可以用于选择含有缺陷基因的转化体的标记。例如,可以将缺陷基因与可选择标记缔和引入非复制或温度敏感的质粒上。通过在不允许质粒复制的条件下选择标记来影响质粒整合的选择。通过检查菌落是否丢失可选择标记和是否获得经突变的基因来影响导致基因替代的第二重组事件的选择。可替代地,缺陷核酸序列可以含有基因的一个或多个核苷酸的插入、取代或缺失,如下文所述。

[0172] 在其他实施例中,通过建立的反义技术使用与基因的核酸序列互补的核苷酸序列来构建经修饰的细胞。更特别地,可以通过引入与基因的核酸序列互补的核苷酸序列来减少(下调)或消除革兰氏阳性细胞的基因的表达,该核苷酸序列可以在细胞中转录并且能够与细胞中产生的mRNA杂交。在允许互补反义核苷酸序列与mRNA杂交的条件下,翻译的蛋白质的量因此减少或消除。此类反义方法包括但不限于RNA干扰(RNAi)、小干扰RNA(siRNA)、微小RNA(miRNA)、反义寡核苷酸等,所有这些都是熟练技术人员熟知的。

[0173] 在其他实施例中,经由CRISPR-Cas9编辑产生/构建经修饰的细胞。例如,编码目的蛋白的基因可以通过核酸指导的内切核酸酶的方式编辑或破坏(或缺失或下调),该方式通过将指导RNA(例如,Cas9)和Cpf1或指导DNA(例如,NgAgo)结合发现其靶DNA,这将内切核酸酶募集到DNA上的靶序列上,其中该内切核酸酶可以在DNA中产生单链或双链断裂。此靶向DNA断裂成为DNA修复的底物,并且可以与所提供的编辑模板重组以使基因破坏或缺失。例如,编码核酸指导的内切核酸酶(出于此目的,来自化脓链球菌的Cas9)的基因或编码Cas9核酸酶的密码子优化的基因可操作地连接到在革兰氏阳性细胞中有活性的启动子和在革兰氏阳性细胞中有活性的终止子,从而产生革兰氏阳性细胞Cas9表达盒。同样,本领域技术人员容易地鉴定目的基因特有的一个或多个靶位点。例如,为了构造编码gRNA的DNA构建体-指向目的基因内的靶位点,可变的靶向结构域(VT)将包含靶位点的核苷酸,这些核苷酸位于(PAM)前间区序列邻近基序(TGG)的5',这些核苷酸与编码化脓链球菌Cas9的Cas9内切核酸酶识别结构域(CER)的DNA融合。将编码VT结构域的DNA和编码CER结构域的DNA组合,从

而产生编码gRNA的DNA。因此,通过将编码gRNA的DNA可操作地连接到革兰氏阳性细胞中有活性的启动子和革兰氏阳性细胞中有活性的终止子来产生gRNA的革兰氏阳性表达盒。

[0174] 在某些实施例中,将由内切核酸酶诱导的DNA断裂用输入序列修复/替代。例如,为了精确修复由上述Cas9表达盒和gRNA表达盒产生的DNA断裂,提供了核苷酸编辑模板,使得细胞的DNA修复机构可以利用该编辑模板。例如,靶基因的约500bp 5'可以与靶基因的约500bp3'融合以产生编辑模板,该模板由革兰氏阳性宿主机构用于修复由RGEN产生的DNA断裂。

[0175] 可以使用许多不同的方法(例如,原生质体融合、电穿孔、自然感受态或诱导感受态)将Cas9表达盒、gRNA表达盒、和编辑模板共同递送至丝状真菌细胞。通过用正向和反向引物扩增基因座,通过PCR扩增靶基因的基因座来筛选经转化的细胞。这些引物可以扩增野生型基因座或已由RGEN编辑的经修饰的基因座。然后使用测序引物对这些片段进行测序以鉴定经编辑的菌落。

[0176] 在又其他实施例中,使用本领域熟知的方法(包括但不限于化学诱变和转座),通过随机或特异性诱变构建经修饰的细胞。可以通过使亲本细胞经受诱变并且筛选基因表达已减少或消除的突变型细胞来进行基因的修饰。可以例如通过使用合适的物理或化学诱变剂、使用合适的寡核苷酸、或使DNA序列经受PCR产生的诱变来进行可以是特异性的或随机的诱变。此外,诱变可以通过使用这些诱变方法的任何组合来进行。

[0177] 适合用于本发明目的的物理或化学诱变剂的实例包括紫外线(UV)辐照、羟胺、N-甲基-N'-硝基-N-亚硝基胍(MNNG)、N-甲基-N'-亚硝基胍(NTG)、O-甲基羟胺、亚硝酸、甲烷磺酸乙酯(EMS)、亚硫酸氢钠、甲酸、和核苷酸类似物。当使用这样的试剂时,典型地通过如下方式来进行诱变:在合适的条件下在选择的诱变剂的存在下孵育待诱变的亲本细胞,并且选择展现出基因的表达降低或无表达的突变型细胞。

[0178] PCT公开号W0 2003/083125公开了用于修饰革兰氏阳性(芽孢杆菌属)细胞的方法,如使用PCR融合来产生芽孢杆菌属缺失菌株和DNA构建体以绕过大肠杆菌。PCT公开号W0 2002/14490公开了用于修饰芽孢杆菌属细胞的方法,这些方法包括(1)构建和转化整合的质粒(pComK), (2)随机诱变编码序列、信号序列和前肽序列, (3)同源重组, (4)通过向转化DNA添加非同源侧翼提高转化效率, (5)优化双交叉整合, (6)定点诱变,和(7)无标记缺失。

[0179] 本领域技术人员熟知用于将多核苷酸序列引入细菌细胞(例如,革兰氏阴性细胞、革兰氏阳性细胞)中的合适方法。实际上,包括原生质体转化和中板集合、转导和原生质体融合在内的诸如转化等的方法是已知的并且适合用于本公开。转化方法特别优选地用于将本公开的DNA构建体引入宿主细胞中。

[0180] 除了常用方法之外,在一些实施例中,直接转化宿主细胞(即,在引入宿主细胞之前,不使用中间细胞来扩增DNA构建体或以其他方式处理DNA构建体)。将DNA构建体引入宿主细胞中包括本领域已知的将DNA引入宿主细胞中而不插入质粒或载体中的那些物理和化学方法。此类方法包括但不限于氯化钙沉淀、电穿孔、裸DNA、脂质体等。在另外的实施例中,将DNA构建体与质粒一起共转化而不插入该质粒中。在另外的实施例中,通过本领域已知的方法,将选择性标记从经修饰的芽孢杆菌属菌株中缺失或基本上切除。在一些实施例中,载体从宿主染色体上分解下来,将侧翼区留在染色体上,而将固有的染色体区去除。

[0181] 用于在革兰氏阳性细胞中表达基因、其编码序列(CDS)、可读框(ORF)和/或变体序

列的启动子和启动子序列区通常是本领域技术人员已知的。通常选择本公开的启动子序列,使得它们在革兰氏阳性细胞中起作用。例如,可用于驱动芽孢杆菌属细胞中基因表达的启动子包括但不限于枯草芽孢杆菌碱性蛋白酶(aprE)启动子、枯草芽孢杆菌的 α -淀粉酶启动子(amyE)、地衣芽孢杆菌的 α -淀粉酶启动子(amyL)、解淀粉芽孢杆菌的 α -淀粉酶启动子、来自枯草芽孢杆菌的中性蛋白酶(nprE)启动子、突变aprE启动子,或者来自地衣芽孢杆菌或其他相关芽孢杆菌的任何其他启动子。在公开号W0 2002/14490中描述了用于在芽孢杆菌属细胞中筛选和产生具有一系列活性(启动子强度)的启动子文库的方法。

[0182] IV. 用于生产蛋白质的发酵革兰氏阳性细胞

[0183] 如上文总体上所述,某些实施例涉及用于构建和获得具有增加的蛋白质生产表型的革兰氏阳性细胞的组合物和方法。因此,某些实施例涉及通过在合适的培养基中发酵细胞而在革兰氏阳性细胞中生产目的蛋白的方法。本领域熟知的发酵方法可用于发酵本公开的革兰氏阳性细胞。

[0184] 在一些实施例中,将细胞在分批或连续发酵条件下培养。经典的分批发酵是封闭的系统,其中在发酵开始时设定培养基的组成,并且该组成在发酵期间不改变。在发酵开始时,向培养基接种一种或多种期望的生物体。在该方法中,使发酵发生而不向系统中添加任何组分。典型地,分批发酵符合关于添加碳源的“分批”的资格,并且经常对控制因素(诸如pH和氧浓度)进行尝试。分批系统的代谢物和生物质组成不断变化直到发酵停止时。在典型分批培养中,细胞可以通过静态停滞期进展到高生长对数期,并且最后进入生长速率减小或停止的静止期。如果不经处理,处于静止期的细胞最终死亡。通常,对数期的细胞负责产物的大量生产。

[0185] 标准分批系统的合适变化是“补料分批”发酵系统。在典型分批系统的该变型中,随着发酵的进展,以增量添加底物。当分解代谢物阻遏可能抑制细胞的代谢时并且在培养基中期望具有有限量的底物的情况下,补料分批系统是有用的。在补料分批系统中实际底物浓度的测量是困难的并且因此基于可测量因素(如pH、溶解氧和废气(如CO₂)的分压)的变化对其进行估计。分批和补料分批发酵在本领域中是常用并且已知的。

[0186] 连续发酵是开放的系统,其中将限定的发酵培养基连续添加到生物反应器中,并同时去除等量的条件培养基以用于加工。连续发酵通常将培养物维持在恒定的高密度,其中细胞主要处于对数期生长。连续发酵允许对影响细胞生长和/或产物浓度的一种或多种因素进行调节。例如,在一个实施例中,将限制性营养物质(诸如碳源或氮源)维持在固定的速率,并且允许调节所有其他参数。在其他系统中,影响生长的许多因素可以不断改变,而通过培养基浊度测量的细胞浓度保持不变。连续系统努力维持稳态生长条件。因此,由于抽出培养基而引起的细胞损失应当与发酵中的细胞生长速率相平衡。调节用于连续发酵工艺的营养物质和生长因子的方法以及用于最大化产物形成速率的技术是工业微生物学领域中熟知的。

[0187] 在某些实施例中,可以通过常规程序从培养基中回收由本公开的革兰氏阳性细胞表达/生产的蛋白,这些常规程序包括通过离心或过滤从培养基中分离宿主细胞,或者如果需要,破坏细胞并从细胞部分和碎片中去除上清液。典型地,在澄清后,将上清液或滤液的蛋白质组分借助盐(例如,硫酸铵)沉淀。然后将沉淀出的蛋白质溶解,并且可以通过多种色谱程序(例如,离子交换色谱法、凝胶过滤)进行纯化。

[0188] 在一些实施例中,将细胞在分批或连续发酵条件下培养。经典的分批发酵是封闭的系统,其中在发酵开始时设定培养基的组成,并且该组成在发酵期间不改变。在发酵开始时,向培养基接种一种或多种期望的生物体。在该方法中,使发酵发生而不向系统中添加任何组分。典型地,分批发酵符合关于添加碳源的“分批”的资格,并且经常对控制因素(诸如pH和氧浓度)进行尝试。分批系统的代谢物和生物质组成不断变化直到发酵停止时。在典型分批培养中,细胞可以通过静态停滞期进展到高生长对数期,并且最后进入生长速率减小或停止的静止期。如果不经处理,处于静止期的细胞最终死亡。通常,对数期的细胞负责产物的大量生产。

[0189] 标准分批系统的合适变化是“补料分批”发酵系统。在典型分批系统的该变型中,随着发酵的进展,以增量添加底物。当分解代谢物阻遏可能抑制细胞的代谢时并且在培养基中期望具有有限量的底物的情况下,补料分批系统是有用的。在补料分批系统中实际底物浓度的测量是困难的并且因此基于可测量因素(如pH、溶解氧和废气(如CO₂)的分压)的变化对其进行估计。分批和补料分批发酵在本领域中是常用并且已知的。

[0190] 连续发酵是开放的系统,其中将限定的发酵培养基连续添加到生物反应器中,并同时去除等量的条件培养基以用于加工。连续发酵通常将培养物维持在恒定的高密度,其中细胞主要处于对数期生长。连续发酵允许对影响细胞生长和/或产物浓度的一种或多种因素进行调节。例如,在一个实施例中,将限制性营养物质(诸如碳源或氮源)维持在固定的速率,并且允许调节所有其他参数。在其他系统中,影响生长的许多因素可以不断改变,而通过培养基浊度测量的细胞浓度保持不变。连续系统努力维持稳态生长条件。因此,由于抽出培养基而引起的细胞损失应当与发酵中的细胞生长速率相平衡。调节用于连续发酵工艺的营养物质和生长因子的方法以及用于最大化产物形成速率的技术是工业微生物学领域中熟知的。

[0191] 在某些实施例中,可以通过常规程序从培养基中回收由本公开的革兰氏阳性细胞表达/生产的蛋白,这些常规程序包括通过离心或过滤从培养基中分离宿主细胞,或者如果需要,破坏细胞并从细胞部分和碎片中去除上清液。典型地,在澄清后,将上清液或滤液的蛋白质组分借助盐(例如,硫酸铵)沉淀。然后将沉淀出的蛋白质溶解,并且可以通过多种色谱程序(例如,离子交换色谱法、凝胶过滤)进行纯化。

[0192] V. 目的蛋白

[0193] 本公开的目的蛋白(P0I)可以是任何内源或异源蛋白质,并且它可以是这样的P0I的变体。蛋白质可以含有一个或多个二硫桥,或者是其功能形式为单体或多聚体的蛋白质,即蛋白质具有四级结构并且由多个相同的(同源)或不相同的(异源)亚基组成,其中P0I或其变体P0I优选是具有目的特性的P0I。

[0194] 例如,在某些实施例中,相对于其未经修饰的(参比或对照)细胞,本公开的经修饰的革兰氏阳性细胞生产多至少约0.1%、多至少约0.5%、多至少约1%、多至少约5%、多至少约6%、多至少约7%、多至少约8%、多至少约9%、或多至少约10%或更多的P0I。

[0195] 在某些实施例中,相对于对照细胞,本公开的经修饰的革兰氏阳性细胞表现出P0I的比生产力(Q_p)增加。例如,比生产力(Q_p)的检测是用于评估蛋白质生产的合适方法。比生产力(Q_p)可以使用以下等式确定:

[0196] “ $Q_p = gP / gDCW \cdot hr$ ”

[0197] 其中,“gP”是罐中产生的蛋白质的克数;“gDCW”是罐中的干细胞重量(DCW)的克数;并且“hr”是从接种时间开始的以小时计的发酵时间,其包括生产时间以及生长时间。

[0198] 因此,在某些其他实施例中,相对于未经修饰的(亲本)细胞,本公开的经修饰的革兰氏阳性细胞包含至少约0.1%、至少约1%、至少约5%、至少约6%、至少约7%、至少约8%、至少约9%、或至少约10%或更多的比生产力(Q_p)增加。

[0199] 在某些实施例中,P0I或其变体P0I选自由以下组成的组:乙酰酯酶、氨肽酶、淀粉酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、碳酸酐酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡萄糖淀粉酶、葡萄糖氧化酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、葡萄糖醛酸酶、糖基水解酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲基酯酶、果胶分解酶、过水解酶、多元醇氧化酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、肽酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶、及其组合。

[0200] 因此,在某些实施例中,P0I或其变体P0I是选自酶学委员会(EC)编号EC 1、EC 2、EC 3、EC 4、EC 5、或EC 6的酶。

[0201] 存在本领域普通技术人员已知的用于检测和测量细胞内和细胞外表达的蛋白质的活性的各种测定。

[0202] VI. 示例性实施例

[0203] 本文公开的组合物和方法的非限制性实施例如下:

[0204] 1. 一种变体前区序列,其包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代,其中所述变体前区的氨基酸位置根据SEQ ID NO:15编号。

[0205] 2. 如实施例1所述的变体前区,其衍生自与SEQ ID NO:15的位置1-84具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0206] 3. 一种变体前区序列,其包含位置2处的甘氨酸(G)和位置3处的赖氨酸(K)的氨基酸插入,以及选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,其中所述变体前区的氨基酸位置根据SEQ ID NO:14编号。

[0207] 4. 如实施例3所述的变体前区,其衍生自与SEQ ID NO:14的位置1-86具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0208] 5. 一种变体前区序列,其包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丙氨酸(A)的氨基酸插入,其中该变体前区的氨基酸位置根据SEQ ID NO:30编号。

[0209] 6. 如实施例5所述的变体前区,其衍生自与SEQ ID NO:30的位置1-87具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0210] 7. 一种变体前区序列,其包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)和位置4处的丝氨酸(S)的氨基酸插入,其中该变体前区的氨基酸位置根据SEQ ID NO:29编号。

[0211] 8.如实施例7所述的变体前区,其衍生自与SEQ ID NO:29的位置1-87具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0212] 9.一种变体前区序列,其包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)、位置4处的丙氨酸(A)和位置5处的丙氨酸(A)的氨基酸插入,其中该变体前区的氨基酸位置根据SEQ ID NO:31编号。

[0213] 10.如实施例9所述的变体前区,其衍生自与SEQ ID NO:31的位置1-88具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0214] 11.一种变体前区序列,其包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)和位置73处的异亮氨酸(I)至缬氨酸(V)取代(I72V),其中该变体前区的氨基酸位置根据SEQ ID NO:32编号。

[0215] 12.如实施例11所述的变体前区,其衍生自与SEQ ID NO:15或SEQ ID NO:32的位置1-84具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0216] 13.一种变体前区序列,其包含位置30处的谷氨酸(E)至甘氨酸(G)取代(E30G)、位置68处的亮氨酸(L)至赖氨酸(K)取代(L68K)和位置80处的谷氨酸(E)至异亮氨酸(I)取代(E80I),其中该变体前区的氨基酸位置根据SEQ ID NO:33编号。

[0217] 14.如实施例13所述的变体前区,其衍生自与SEQ ID NO:15或SEQ ID NO:33的位置1-84具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。

[0218] 15.一种核酸,其编码如实施例1-14中任一项所述的变体前区。

[0219] 16.一种多核苷酸,其包含如实施例15所述的变体前区核酸。

[0220] 17.一种多核苷酸,其包含编码如实施例1-14中任一项所述的变体前区的上游(5')核酸,该上游核酸可操作地连接到编码目的蛋白(POI)的下游(3')核酸序列。

[0221] 18.一种多核苷酸,其包含编码蛋白信号(分泌)序列的上游(5')核酸,该上游核酸可操作地连接到编码如实施例1-14中任一项所述的变体前区的下游(3')核酸序列,该编码如实施例1-14中任一项所述的变体前区的下游核酸序列可操作地连接到编码目的蛋白(POI)的下游核酸序列。

[0222] 19.一种表达盒,其包含可操作地连接到如实施例16-18中任一项所述的下游(3')多核苷酸的上游(5')启动子区序列。

[0223] 20.一种革兰氏阳性宿主细胞,其包含引入的如实施例16-18中任一项所述的多核苷酸或引入的如实施例19所述的盒。

[0224] 21.如实施例1所述的变体前区,其中该一个或多个氨基酸取代使得该变体前区序列相对于SEQ ID NO:15的参比前区净电荷增加。

[0225] 22.如实施例3所述的变体前区,其中该一个或多个氨基酸取代使得该变体前区序列相对于SEQ ID NO:14的参比前区净电荷增加。

[0226] 23.如实施例21或实施例22所述的变体前区,其中该净电荷的增加为至少1(+1)到4(+4)。

- [0227] 24. 如实施例1所述的变体前区,其包含表1所示的氨基酸修饰。
- [0228] 25. 如实施例1所述的变体前区,其中位置30处的氨基酸取代是甘氨酸(G)、组氨酸(H)或天冬酰胺(N)。
- [0229] 26. 如实施例1所述的变体前区,其中位置1处的氨基酸取代是苏氨酸(T)。
- [0230] 27. 如实施例1所述的变体前区,其中位置2处的氨基酸取代是丙氨酸(A)或脯氨酸(P)。
- [0231] 28. 如实施例1所述的变体前区,其中位置3处的氨基酸取代是丙氨酸(A)、胱氨酸(C)、甘氨酸(G)、精氨酸(R)、丝氨酸(S)或缬氨酸(V)。
- [0232] 29. 如实施例1所述的变体前区,其中位置4处的氨基酸取代是谷氨酰胺(Q)。
- [0233] 30. 如实施例1所述的变体前区,其中位置6处的氨基酸取代是精氨酸(R)。
- [0234] 31. 如实施例1所述的变体前区,其中位置14处的氨基酸取代是苏氨酸(T)。
- [0235] 32. 如实施例1所述的变体前区,其中位置19处的氨基酸取代是精氨酸(R)。
- [0236] 33. 如实施例1所述的变体前区,其中位置20处的氨基酸取代是赖氨酸(K)。
- [0237] 34. 如实施例1所述的变体前区,其中位置23处的氨基酸取代是赖氨酸(K)、天冬酰胺(N)或谷氨酰胺(Q)。
- [0238] 35. 如实施例1所述的变体前区,其中位置36处的氨基酸取代是甘氨酸(G)。
- [0239] 36. 如实施例1所述的变体前区,其中位置37处的氨基酸取代是苏氨酸(T)。
- [0240] 37. 如实施例1所述的变体前区,其中位置38处的氨基酸取代是组氨酸(H)。
- [0241] 38. 如实施例1所述的变体前区,其中位置39处的氨基酸取代是赖氨酸(K)或谷氨酰胺(Q)。
- [0242] 39. 如实施例1所述的变体前区,其中位置44处的氨基酸取代是异亮氨酸(I)。
- [0243] 40. 如实施例1所述的变体前区,其中位置49处的氨基酸取代是苏氨酸(T)或谷氨酰胺(Q)。
- [0244] 41. 如实施例1所述的变体前区,其中位置50处的氨基酸取代是精氨酸(R)。
- [0245] 42. 如实施例1所述的变体前区,其中位置64处的氨基酸取代是赖氨酸(K)。
- [0246] 43. 如实施例1所述的变体前区,其中位置65处的氨基酸取代是赖氨酸(K)。
- [0247] 44. 如实施例1所述的变体前区,其中位置67处的氨基酸取代是丝氨酸(S)。
- [0248] 45. 如实施例1所述的变体前区,其中位置68处的氨基酸取代是赖氨酸(K)或甲硫氨酸(M)。
- [0249] 46. 如实施例1所述的变体前区,其中位置79处的氨基酸取代是谷氨酰胺(Q)。
- [0250] 47. 如实施例1所述的变体前区,其中位置83处的氨基酸取代是异亮氨酸(I)。
- [0251] 48. 如实施例1所述的变体前区,其中位置84处的氨基酸取代是亮氨酸(L)。
- [0252] 49. 如实施例3所述的变体前区,其包含表2、表3、表4、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31及其组合所示的氨基酸修饰。
- [0253] 50. 如实施例3所述的变体前区,其中位置1处的氨基酸取代是苏氨酸(T)。
- [0254] 51. 如实施例3所述的变体前区,其中位置32处的氨基酸取代是甘氨酸(G)。
- [0255] 52. 如实施例3所述的变体前区,其中位置38处的氨基酸取代是甘氨酸(G)。
- [0256] 53. 如实施例3所述的变体前区,其中位置46处的氨基酸取代是异亮氨酸(I)。
- [0257] 54. 如实施例3所述的变体前区,其中位置66处的氨基酸取代是赖氨酸(K)。

- [0258] 55.如实施例3所述的变体前区,其中位置67处的氨基酸取代是赖氨酸(K)。
- [0259] 56.如实施例3所述的变体前区,其中位置70处的氨基酸取代是赖氨酸(K)。
- [0260] 57.如实施例3所述的变体前区,其中位置73处的氨基酸取代是精氨酸(R)。
- [0261] 58.一种变体前区,其包含表1-5、图1、图2、图3、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:33中任一项及其组合所示的氨基酸修饰。
- [0262] 59.如实施例58所述的变体前区,其衍生自与SEQ ID NO:15的位置1-84具有至少70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%氨基酸序列同一性的亲本多肽或参比多肽。
- [0263] 60.一种用于在革兰氏阳性细菌细胞中生产异源目的蛋白(POI)的方法,该方法包括(a)将表达盒引入革兰氏阳性细胞,该表达盒包含可操作地连接到编码变体前区序列的下游(3')核酸的上游(5')启动子,该变体前区序列包含位置30和选自1、2、3、4、6、14、16、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、79、83和84的一个或多个位置处的氨基酸取代,该编码变体前区序列的下游核酸可操作地连接到编码该POI的下游(3')核酸序列,其中该变体前区的氨基酸位置根据SEQ ID NO:15进行编号;以及(b)在用于生产该POI的合适条件下使经修饰的细胞生长/培养/发酵。
- [0264] 61.一种用于在革兰氏阳性细菌细胞中生产异源目的蛋白(POI)的方法,该方法包括(a)将表达盒引入革兰氏阳性细胞,该表达盒包含可操作地连接到编码变体前区序列的下游(3')核酸的上游(5')启动子序列,该变体前区序列包含位置2处的甘氨酸(G)、位置3处的赖氨酸(K)的氨基酸插入和选自1、32、38、46、66、67、70和73的一个或多个位置处的氨基酸取代,该编码变体前区序列的下游核酸可操作地连接到编码该POI的下游(3')核酸序列,其中该变体前区的氨基酸位置根据SEQ ID NO:14进行编号;以及(b)在用于生产该POI的合适条件下使经修饰的细胞生长/培养/发酵。
- [0265] 62.一种用于在革兰氏阳性细菌细胞中生产异源目的蛋白(POI)的方法,该方法包括(a)向革兰氏阳性细胞中引入如实施例19所述的表达盒,以及(b)在用于生产该POI产生的合适条件下生长/培养/发酵经修饰的细胞。
- [0266] 63.如实施例60-62中任一项所述的方法,其中该盒进一步包含编码前蛋白信号(分泌)序列的核酸,该核酸(序列)可操作地连接并且位于启动子与变体前区序列之间。
- [0267] 64.如实施例60-62中任一项所述的方法,其中该经修饰的细胞相对于具有引入的表达盒的对照革兰氏阳性细胞生产增加量的POI,该表达盒包含可操作地连接到编码SEQ ID NO:15的前区序列的下游(3')核酸的相同的上游(5')启动子序列,该编码前区序列的下游核酸可操作地连接到编码相同POI的下游核酸序列。
- [0268] 65.如实施例64所述的方法,其中相对于对照细胞,该POI的增加量是多至少约0.1%、多至少约0.5%、多至少约1%、多至少约5%、多至少约6%、多至少约7%、多至少约8%、多至少约9%、或多至少约10%或更多的POI。
- [0269] 66.如实施例60-63中任一项所述的方法,其中该经修饰的细胞相对于具有引入的表达盒的对照革兰氏阳性细胞的性能指数(PI)值为至少约1.1,该表达盒包含可操作地连接到编码SEQ ID NO:15的前区序列的下游核酸的相同的上游启动子序列,该编码前区序列的下游核酸可操作地连接到与该经修饰的细胞编码相同POI的下游核酸序列,其中该经修

饰的细胞和对照细胞在相同条件下培养。

[0270] 67. 如实施例60-63中任一项所述的方法, 其中将引入的盒整合到细胞的基因组中。

[0271] 68. 如实施例60-63中任一项所述的方法, 该方法包括至少两个引入的盒。

[0272] 69. 如实施例60-63中任一项所述的方法, 其中该POI选自由以下组成的组: 酶、抗体、受体、和调控蛋白。

[0273] 70. 如实施例60-63中任一项所述的方法, 其中该POI选自由以下组成的组: 乙酰酯酶、氨肽酶、淀粉酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、碳酸酐酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、葡萄糖醛酸酶、糖基水解酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲基酯酶、果胶分解酶、过水解酶、多元醇氧化酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、肽酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶及其组合。

[0274] 71. 如权利要求70所述的方法, 其中该蛋白酶是天然的或变体的枯草杆菌蛋白酶。

[0275] 72. 如权利要求60-71中任一项所述的方法, 其中该革兰氏阳性细菌细胞是芽孢杆菌属物种细胞, 任选地其中该芽孢杆菌属物种细胞选自由以下组成的组: 枯草芽孢杆菌、地衣芽孢杆菌、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、解淀粉芽孢杆菌、克劳氏芽孢杆菌、耐盐芽孢杆菌、巨大芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、灿烂芽孢杆菌、苏云金芽孢杆菌和嗜热脂肪土芽孢杆菌。

[0276] 实例

[0277] 根据以下实例可以进一步理解本公开的某些方面, 这些实例不应被解释为限制性的。对材料和方法的修改对于本领域技术人员而言是显而易见的。本文使用的标准重组DNA和分子克隆技术是本领域熟知的 (Ausubel等人, 1987; Sambrook等人, 1989)。

[0278] 实例1

[0279] 通过位点评估文库对GG36 PRO序列进行修饰

[0280] 在本实例中, 将变体枯草杆菌蛋白酶蛋白 (BG46_变体1; SEQ ID NO:2) 用作报告蛋白来监测如本文所述的蛋白表达。更特别地, 包含上游 (5') aprE基因侧翼区的DNA片段可操作地连接到包含变体枯草芽孢杆菌rrnI-P2启动子/5'-aprE UTR区DNA序列的多核苷酸构建体 (例如, 表达盒), 从而产生SEQ ID NO:1, 该序列可操作地连接到编码AprE信号序列的DNA序列 (SEQ ID NO:4), 该DNA序列可操作地连接到编码变体前区序列A (SEQ ID NO:9) 的DNA序列 (SEQ ID NO:5), 该编码变体前区序列A的DNA序列可操作地连接到编码成熟BG46_变体1 (SEQ ID NO:2) 的DNA序列, 该编码成熟BG46_变体1的DNA序列可操作地连接到解淀粉芽孢杆菌BPN' 终止子DNA序列 (SEQ ID NO:6), 该多核苷酸构建体可操作地连接到包含卡那霉素 (kan) 基因表达盒 (SEQ ID NO:7) 的下游 (3') aprE基因侧翼区序列。更特别地, 该DNA片段使用标准分子生物学技术组装, 并用作模板来开发包含本文所述的一个或多个前区修饰 (突变) 的线性DNA表达盒。

[0281] 前区序列A的位点评估文库 (SEL)

[0282] 编码变体前区序列A(30H,SEQ ID NO:9)的氨基酸序列用作模板创建位点评估文库(SEL),并已由特威斯特生物科学公司总部(南旧金山)开发为4.4kb片段。将表达盒的线性DNA用于转化感受态枯草芽孢杆菌细胞,其中将转化混合物铺板到含有1.8ppm卡那霉素的LA板上并在37°C下孵育过夜。挑取单菌落并使其在抗生素选择下在Luria肉汤中于37°C生长。

[0283] 进行序列分析以确定独特的前区序列A变体,将这些变体精心挑选到96孔微量滴定板(MTP)中。例如,变体前区序列A(SEQ ID NO:9)的第一个(1st)氨基酸位置是丙氨酸(“Ala”或“A”),它可以改变(取代)为所有其他十九(19)个天然存在的氨基酸残基,并且变体前区序列A(SEQ ID NO:9)的第二个(2nd)氨基酸位置是谷氨酸(“Glu”或“E”),它可以被改变(取代)为所有其他十九(19)个天然存在的氨基酸残基,以此类推,直到前区序列A(SEQ ID NO:9)的所有八十四(84)个位置已被评估(例如,见图1)。

[0284] 在报告蛋白表达实验中,将转化后的细胞置于96孔MTP中,在培养基(基于MOP缓冲液的富集半限定培养基)中在32°C下以300rpm、湿度80%在振荡培养箱中生长3天,并离心和过滤。使用澄清的培养上清液来测量(测定)报告蛋白酶活性以确定生产力水平,其中样品在七十二(72)小时之后采集。在以下实例2中进一步描述报告蛋白酶活性测定。

[0285] 影响生产力的位置

[0286] 如下表所述,表1显示相比于其中报告蛋白基因用天然GG36前区序列(SEQ ID NO:15)表达的参比构建体的相对报告蛋白生产力的结果,其中通过以下实例2中所述的活性测定来测量表达。例如,如表1所示,给出了七十二(72)小时后采集的样品的性能指数(PI)值(PI计算如实例2所述),其中结合位置30影响生产力的氨基酸(残基)位置为1、2、3、4、6、14、19、20、23、36、37、38、39、42、43、44、49、50、64、65、67、68、71、83、84。5'氨基酸的变化,以及3'氨基酸修饰影响生产力,此外80%的选择组合具有另外的正电荷。

[0287] 表1前区变体A序列与WT PRO的性能指数比较

[0288]

变体相对于 WT	PI	Δ 电荷
无	1.0	0
E30H	1.5	1
E30G	2.3	1
E6R-E30H	2.5	3
S19R-E30H	2.3	2
E30H-E39Q	2.4	2
E23Q-E30H	2.3	2
E3G-E30H	2.7	2
E30H-L44I	2.2	1
E30H-T83I	2.0	1
E30H-E38H	2.5	2
E23N-E30H	2.1	2
E3A-E30N	2.5	2
E30H-E67S	2.1	2
E30H-E39K	2.7	3
E30H-A79Q	2.2	1
E3V-E30H	2.5	2
E3C-E30H	2.4	2
A1T-E30H	2.5	1
A4Q-E30H	2.6	1
E30H-T50R	2.3	2
E30H-L68K	1.5	2

[0289]

E30H-E37T	2.4	2
E30H-E36G	2.1	2
E23K-E30H	2.2	3
E14T-E30H	2.1	2
E2P-E30H	2.4	2
E2A-E30H	2.1	2
E3R-E30H	2.7	3
E30H-E49T	2.3	2
E3S-E30H	2.2	2
E30H-E49Q	2.3	2
E30H-M84L	2.2	2
E30H-A65K	2.3	2
E20K-E30H	2.2	3
E30H-D64K	2.3	3
E30H-L68M	1.5	1
E30H-E43D	2.2	1
E16S-E30H	2.4	2
E30H-I42V	2.4	1
E30H-A71R	2.5	2
A4N-E30H	2.2	1
E30H-D64A	2.0	2
E3P-E30H	2.4	2
E6Q-E30H	2.4	2

[0290] 实例2

[0291] 蛋白酶活性测定

[0292] 通过测量合成的suc-AAPF-pNA肽底物的水解来确定BG46_变体1枯草杆菌蛋白酶的蛋白酶活性。对于AAPF测定,使用的试剂溶液是:100mM Tris pH 8.6,10mM CaCl₂, 0.005% **Tween®**-80 (Tris/Ca缓冲液)和在DMSO中的160mM suc-AAPF-pNA (suc-AAPF-

pNA原液;西格玛公司:S-7388)。为了制备工作溶液,将一(1)mL suc-AAPF-pNA原液添加到100mL Tris/Ca缓冲液中并混合。将酶样品添加到含有一(1)mg/mL suc-AAPF-pNA工作溶液的微量滴定板(MTP)中,并且使用SpectraMax读板器以动力学方式在室温下经三到五(3-5)分钟在405nm处测定活性。将蛋白酶活性表示为mOD/分钟。测量每个构建的变体的活性,并将其与在同一板中生长的参比构建体进行比较。将参比样品值除以变体样品值,即性能指数(PI)。

[0293] 实例3

[0294] 5' GG36 PRO序列的修饰

[0295] 在本实例中,将BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)用作报告蛋白来监测如本文所述的表达。更特别地,包含上游(5') aprE基因侧翼区的DNA片段可操作地连接到包含变体(5') 枯草芽孢杆菌rrnI-P2启动子/5'-UTR aprE区DNA序列的多核苷酸构建体,从而产生SEQ ID NO:1,该序列可操作地连接到编码AprE信号肽序列的DNA序列(SEQ ID NO:4),该DNA序列可操作地连接到编码变体前区序列B(SEQ ID NO:9,30H)的DNA序列(SEQ ID NO:10),该编码变体前区序列B的DNA序列可操作地连接到编码成熟BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)的DNA序列,该编码成熟BG46_变体2枯草杆菌蛋白酶的DNA序列可操作地连接到解淀粉芽孢杆菌BPN' 终止子DNA序列(SEQ ID NO:6),该多核苷酸构建体可操作地连接到包含卡那霉素(kan)基因表达盒(SEQ ID NO:7)的下游(3') aprE基因侧翼区序列。更特别地,该DNA片段使用标准分子生物学技术组装,并用作模板来开发包含本文所述的一个或多个启动子区修饰的线性DNA表达盒。

[0296] 引入5' 前区修饰

[0297] 使用七十八(78)个氨基酸的AprE前区序列(SEQ ID NO:12)的5' (N-末端)序列将前三(3)个氨基酸残基与变体前区序列C(SEQ ID NO:14)的第一个氨基酸残基(A)交换,产生SEQ ID NO:13。如上表1所示,前区变体A序列(即,表1,“E030G”;SEQ ID NO:11)的突变体显示出报告蛋白的高表达。采用标准分子生物学技术制备表达构建体并与参比前区进行比较。将变体(突变体)开发为4.4kb的线性DNA片段,并且用于转化感受态枯草芽孢杆菌细胞,其中将转化混合物铺板到含有1.8ppm卡那霉素的LA板上并在37°C下孵育过夜。挑取单菌落并使其在抗生素选择下在Luria肉汤中于37°C生长。

[0298] 在报告蛋白表达实验中,将转化后的细胞置于96孔MTP中,在培养基(基于MOP缓冲液的富集半限定培养基)中在32°C下以300rpm、湿度80%在振荡培养箱中生长3天,并离心和过滤。使用澄清的培养上清液来测量(测定)报告蛋白酶活性以确定生产力水平,其中样品在72小时之后采集。报告蛋白酶活性测定如以上实例2中所述进行,其中变体前区C的蛋白质生产力是E030G-GG36-Pro(SEQ ID NO:11)的约1.1倍。

[0299] 实例4

[0300] GG36 PRO序列的进一步修饰

[0301] 在本实例中,将BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)用作报告蛋白来监测如本文所述的表达。根据上述实例3中的描述进行构建。更特别地,在参比前区序列(SEQ ID NO:15)上制备组合文库,其中氨基酸甘氨酸(G)和赖氨酸(K)在5' (N-末端)序列处插入(“GK”) (例如,见图2,SEQ ID NO:14)。

[0302] 变体前区C中的组合文库

[0303] 从表1数据进行选择以组合前区突变,其中使用标准分子生物学技术来制备表达构建体并且将该构建体与参比前区进行比较。将线性DNA表达盒用于转化感受态枯草芽孢杆菌细胞,其中将转化混合物铺板到含有1.8ppm卡那霉素的LA板上并在37°C下孵育过夜。挑取单菌落并使其在抗生素选择下在Luria肉汤中于37°C生长。在报告蛋白表达实验中,将转化后的细胞置于96孔MTP中,在培养基(基于MOP缓冲液的富集半限定培养基)中在32°C下以300rpm、湿度80%在振荡培养箱中生长3天,并离心和过滤。使用澄清的培养上清液来测量(测定)报告蛋白酶活性以确定生产力水平,其中样品在72小时之后采集。如上述实例2中所述进行报告蛋白酶活性测定。

[0304] 提高生产力的前区中组合

[0305] 如下文呈现,表2揭示了显示所表达的报告蛋白(SEQ ID NO:8)的生产力增加的前区突变组合的序列结果,其中 δ (delta) (Δ) 电荷在+1与+4之间变化。更特别地,正电荷作为环区中的组合(即,残基位置66-73)。给出了在72小时后采集的样品的性能指数(PI)值,其如上文在实例2中所述的进行计算。例如,将这些前区变体的PI值与WT前区序列(表2,SEQ ID NO:15)进行比较,其中氨基酸位置编号(表2,第1列)是与SEQ ID NO:14比较的,没有“GK”插入。如图3所示,变体前区序列可以根据参比变体前区序列C(SEQ ID NO:14;包含86个氨基酸位置)或野生型(参比)前区序列(SEQ ID NO:15;包含84个氨基酸位置)的位置编号进行编号。

[0306] 表2

[0307] 前区变体C(SEQ ID NO:14)序列与没有GK插入的WT PRO(SEQ ID

[0308] NO:15)的性能指数比较

[0309]

变体相对于 WT	PI	Δ 电荷
无	1.0	0
E32G	1.7	1
L46I-D66K-A73R	1.9	3
A67K-L70K-A73R	1.9	3
L70K-A73R	2.0	2
E38G-A67K-L70K	2.0	3
D66K-A67K-L70K	2.0	4
E32G-L70K	2.0	2
A1T-L46I-A73R	2.0	1

[0310]

L70K-A73R	2.0	2
E38G-L46I-D66K-A67K	2.0	4
E38G-D66K-A67K	2.0	4
L46I-D66K-A67K-L70K	2.0	4
E38G-L46I-A67K	2.0	2
L46I-L70K	2.1	1
L46I-L70K-A73R	2.1	2
E38G-L46I-A67K	2.1	2
L46I-D66K-A67K-L70K	2.2	4
L46I-A67K-L70K	2.3	2

[0311] 实例5

[0312] 地衣芽孢杆菌5' GG36 PRO序列的修饰

[0313] 在本实例中,将BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)用作报告蛋白来监测如本文所述的表达。通常如下进行构建。将包含包括地衣芽孢杆菌serA可选择标记表达盒(SEQ ID NO:16)的(5')serA基因侧翼区(5'serA基因FR)的第一DNA片段可操作地连接到多核苷酸构建体,该多核苷酸构建体包含可操作地连接到枯草芽孢杆菌aprE 5'非翻译区(5'-UTR;SEQ ID NO:18)的DNA序列的上游(5')枯草芽孢杆菌rrnI-p3启动子区DNA序列(SEQ ID NO:17),该非翻译区的DNA序列可操作地连接到编码枯草芽孢杆菌AprE信号序列的DNA(SEQ ID NO:4),该DNA可操作地连接到编码变体前区C序列(SEQ ID NO:14)的DNA,该编码变体前区C序列的DNA可操作地连接到编码成熟BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)的DNA序列,该编码成熟BG46_变体2枯草杆菌蛋白酶的DNA序列可操作地连接到地衣芽孢杆菌amyL终止子(SEQ ID NO:19),其进而可操作地连接到(3')serA基因侧翼区(3'serA基因FR)(SEQ ID NO:20)。

[0314] 将包含包括地衣芽孢杆菌lysA可选择标记表达盒(SEQ ID NO:21)的(5')lysA基因侧翼区(5'lysA基因FR)的第二DNA片段可操作地连接到多核苷酸构建体,该多核苷酸构建体包含可操作地连接到枯草芽孢杆菌aprE 5'-UTR(SEQ ID NO:18)的DNA序列的上游(5')枯草芽孢杆菌rrnI-p3启动子区DNA序列(SEQ ID NO:17),该非翻译区的DNA序列可操作地连接到编码枯草芽孢杆菌AprE信号序列的DNA(SEQ ID NO:4),该DNA可操作地连接到编码变体前区C序列(SEQ ID NO:14)的DNA,该编码变体前区C序列的DNA可操作地连接到编码成熟BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)的DNA序列,该编码成熟BG46_变体2枯草杆菌蛋白酶的DNA序列可操作地连接到地衣芽孢杆菌amyL终止子(SEQ ID NO:10),其进而可操作地连接到(3')lysA基因侧翼区(3'lysA基因FR;SEQ ID NO:22)。

[0315] 更特别地,这些DNA片段使用标准分子生物学技术组装,并用作模板来开发包含本文所述的一个或多个前区序列修饰的线性DNA表达盒。例如,通过在基因组中整合第一和第

二DNA片段(如上文所述)构建了包含变体前区序列的地衣芽孢杆菌菌株,其中第一和第二片段含有下文表3所示的前区序列变体。如表3所示,如先前所述使用标准方法(例如,参见WO 2019/055261)确定报告蛋白的生产并且相对于含有E30G变体的对照前区序列归一化。

[0316] 表3

[0317] 地衣芽孢杆菌中前区变体C(GK,SEQ:14)序列与对照前区序列B

[0318] (E32G;SEQ:11)的性能指数比较

[0319]

变体相对于WT	PI	Δ 电荷
E32G	1.0	1
E32G-L70K	1.1	2
E32G-A73R	1.3	2
E32G-L70K-A73R	1.2	3

[0320] 实例6

[0321] 5' GG36 PRO序列的另外修饰

[0322] 在本实例中,将BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)用作报告蛋白来监测如本文所述的表达。特别地,由86个氨基酸组成的突变的GG36 PRO序列(SEQ ID NO:14;变体前区序列C)的5(N-末端)序列经进一步修饰以在位置三(3)的赖氨酸之后引入另外的氨基酸,例如丙氨酸(A)或丝氨酸(S)(例如,以潜在地促进成熟蛋白酶释放到培养基中)。例如,如图1-3所示这些N-末端修饰产生三(3)个变体前区序列D-F(SEQ ID NO:29-31)。

[0323] 将这些N-末端修饰的GG36 PRO序列(SEQ ID NO:29、SEQ ID NO:30和SEQ ID NO:31)组装在表达盒中,该表达盒包含上游(5')aprE基因侧翼区,其可操作地连接到包含变体(5')枯草芽孢杆菌rrnI-P2启动子/5'-UTR aprE区DNA序列的多核苷酸构建体,从而产生SEQ ID NO:1,该序列可操作地连接到编码AprE信号肽序列的DNA序列(SEQ ID NO:4),该DNA序列可操作地连接到编码变体前区序列的DNA序列,该编码变体前区序列的DNA序列可操作地连接到编码成熟BG46_变体2枯草杆菌蛋白酶(SEQ ID NO:8)的DNA序列,该编码成熟BG46_变体2枯草杆菌蛋白酶的DNA序列可操作地连接到解淀粉芽孢杆菌BPN'终止子DNA序列(SEQ ID NO:6),该多核苷酸构建体可操作地连接到包含下游卡那霉素(kan)基因表达盒(SEQ ID NO:7)的下游(3')aprE基因侧翼区(3'aprE基因FR)。使用标准分子生物学技术将包含突变/变体PRO序列的盒转化到枯草芽孢杆菌细胞中。

[0324] 将转化后的细胞置于96孔MTP中,在培养基(基于MOP缓冲液的富集半限定培养基)中在32°C下以300rpm、湿度80%在振荡培养箱中生长3天,并离心和过滤。使用澄清的培养上清液来测量(测定)报告蛋白酶活性以确定生产力水平,其中样品在七十二(72)小时之后采集。特别地,表4显示72小时后相比于使用WT PRO序列(SEQ ID NO:14)表达报告蛋白的参比构建体,报告蛋白生产力(性能指数(PI)值)的结果。

[0325] 表4

[0326] 前区变体D-F(SEQ ID NO:29-31)与变体C前区序列(SEQ ID NO:14)的性能指数比较

[0327]

名称	变体	PI	SEQ ID
变体C	GK	1.00	14
变体E	GKA	1.3	30

变体D	GKS	1.2	29
变体F	GKAA	1.4	31

[0328] 如上表4所示,将前区变体的PI值与前区序列变体前区序列C (SEQ ID NO:14, “G₂K₃”)进行比较,其中氨基酸位置编号是与SEQ ID NO:14 (即,具有“GK”插入)比较的。在生长72小时后,与变体C (AGK;SEQ ID NO:14)相比,三(3)种N-末端前区突变体 (AGKA、AGKS、AGKAA)的PI指数显示报告蛋白BG46_变体2枯草杆菌蛋白酶 (SEQ ID NO:8)的高表达。

[0329] 实例7

[0330] GG36 PRO序列的C-末端修饰

[0331] 在本实例中,GG36变体前区序列B (SEQ ID NO:11)被进一步工程化以用赖氨酸(K)取代位置六十八(68)处的亮氨酸(L)并且用缬氨酸(V)取代位置七十二(72)处的异亮氨酸(I),或用异亮氨酸(I)取代位置八十(80)处的谷氨酸(E),产生两(2)个工程化的前区序列变体序列G (SEQ ID NO:32)和H (SEQ ID NO:33),例如,见图1-图3。特别地,转化后的细胞如上文所述生长,并且用澄清的培养上清液测量(测定)报告蛋白酶活性以确定生产力水平,其中样品在72小时之后采集。

[0332] 例如,下表5显示72小时后相比于使用变体B前区序列 (SEQ ID NO:11)表达报告蛋白的参比构建体,报告蛋白生产力(PI值)的结果。更特别地,如表5所示,与参比/对照变体B前区序列相比,变体前区序列G和H的PI值增加,其中变体前区序列G (包含L68K和I72V突变)相对于前区序列B (E30G)产生1.08的PI,并且变体前区序列H (包含L68K和E80I突变)相对于前区序列B (E30G)产生1.18的PI。

[0333] 表5前区变体G和H与前区变体C的性能指数比较

[0334]

名称	变体	PI	Δ电荷相对于WT
变体B	E30G	1.0	1
变体G	E30G L68K I72V	1.4	2
变体F	E30G L68K E80I	1.5	3

[0335] 参考文献

[0336] PCT Publication No.W01993/20214

[0337] PCT Publication No.W02002/14490

[0338] PCT Publication No.W02003/083125

[0339] PCT Publication No.W02008/112258

[0340] PCT Publication No.W02010/123754

[0341] PCT Publication No.W02016/205710

[0342] Altschul et al.,“Basic local alignment search tool”,J.Mol.Biol.,215 (13):403-410,1990.

[0343] Altschul et al.,“Gapped BLAST and PSI BLAST a new generation of protein database search programs”,Nucleic Acids Res,Set 1;25(17):3389-402, 1997.

[0344] Karlin et al.,“Applications and statistics for multiple high-scoring segments in molecular sequences”,PNAS USA,90(12):5873-5787,1993.

[0345] Saitou and Nei,“The neighbor-joining method:a new method for

reconstructing phylogenetic trees”, *Mol. Biol. Evol.*, Vol. 4, issue 4, pages 406-425, 1987.

[0346] Schaffer et al., “Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements”, *Nucleic Acids Res.*, 29(14):2994-3005, 2001.

[0347] Thompson et al., “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”, *Nucleic Acids Res.*, 22(22):4673-4680, 1994.

[0348] Beaucage and Caruthers, “Deoxynucleoside phosphoramidites-A new class of key intermediates for deoxypolynucleotide synthesis”, *Tetrahedron Lett.*, 22, 1859-1862, 1981.

[0349] Matthes et al., “Simultaneous rapid chemical synthesis of over one hundred oligonucleotides on a microscale”, *EMBO J.*, 3:801-805, 1984.

野生型和变体前区序列

野生型迟缓芽孢杆菌前区序列 (SEQ ID NO: 15)

AEAAKEKYLIGFNEQEA^VSEFVEQVEANDEVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 A (SEQ ID NO: 9)

AEAAKEKYLIGFNEQEA^VSEFVEQVEAND**H**VAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 B (SEQ ID NO: 11)

AEAAKEKYLIGFNEQEA^VSEFVEQVEAND**G**VAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 C (SEQ ID NO: 14)

A**G**KSEAAKEKYLIGFNEQEA^VSEFVEQVEANDEVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 D (SEQ ID NO: 29)

A**G**KSEAAKEKYLIGFNEQEA^VSEFVEQVEANDEVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 E (SEQ ID NO: 30)

A**G**KAAEEAAKEKYLIGFNEQEA^VSEFVEQVEANDEVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 F (SEQ ID NO: 31)

A**G**KAAEEAAKEKYLIGFNEQEA^VSEFVEQVEANDEVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPAISYIEEDAEVTM

变体前区序列 G (SEQ ID NO: 32)

AEAAKEKYLIGFNEQEA^VSEFVEQVEAND**G**VAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPA**V**SYIEEDAEVTM

变体前区序列 H (SEQ ID NO: 33)

AEAAKEKYLIGFNEQEA^VSEFVEQVEAND**G**VAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDALDPA**K**DPAISYIEEDA**I**VTM

图1

比对的区序列

A--EEAKEKYLIGFNEQEA	VSEFVEQVEAND	EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 15)
A--EEAKEKYLIGFNEQEA	VSEFVEQVEAND	H EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 09)
A--EEAKEKYLIGFNEQEA	VSEFVEQVEAND	G EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 11)
AGKEE AKEKYLIGFNEQEA	VSEFVEQVEAND	EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 14)
AGKSE EEAKEKYLIGFNEQEA	VSEFVEQVEAND	EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 29)
AGKAA EEAKEKYLIGFNEQEA	VSEFVEQVEAND	EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 30)
AGKAA EEAKEKYLIGFNEQEA	VSEFVEQVEAND	EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	PAISYIEEDA	AEVTTM	(SEQ 31)
A--EEAKEKYLIGFNEQEA	VSEFVEQVEAND	G EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	VS YIEEDA	AEVTTM	(SEQ 32)
A--EEAKEKYLIGFNEQEA	VSEFVEQVEAND	G EVAILSEEEEEVEIELLHEFETIPVLSVELSPEDVDA	LELDP	VS YIEEDA	I AEVTTM	(SEQ 33)

图2

前区氨基酸位置编号

野生型 (参比) 迟缓芽孢杆菌前区序列 (SEQ ID NO: 15, E30)

A₁E₂E₃A₄K₅E₆K₇Y₈L₉I₁₀G₁₁F₁₂N₁₃E₁₄Q₁₅E₁₆A₁₇V₁₈S₁₉E₂₀F₂₁V₂₂E₂₃Q₂₄V₂₅E₂₆A₂₇N₂₈D₂₉**E**₃₀V₃₁A₃₂I₃₃L₃₄S₃₅E₃₆E₃₇E₃₈E₃₉V₄₀
 E₄₁I₄₂E₄₃L₄₄L₄₅H₄₆E₄₇F₄₈E₄₉T₅₀I₅₁P₅₂V₅₃L₅₄S₅₅V₅₆E₅₇L₅₈S₅₉P₆₀E₆₁D₆₂V₆₃D₆₄A₆₅L₆₆E₆₇L₆₈D₆₉P₇₀A₇₁I₇₂S₇₃Y₇₄I₇₅E₇₆E₇₇D₇₈
 A₇₉E₈₀V₈₁T₈₂T₈₃M₈₄

变体 (参比) 前区A序列(SEQ ID NO: 9, H30)

A₁E₂E₃A₄K₅E₆K₇Y₈L₉I₁₀G₁₁F₁₂N₁₃E₁₄Q₁₅E₁₆A₁₇V₁₈S₁₉E₂₀F₂₁V₂₂E₂₃Q₂₄V₂₅E₂₆A₂₇N₂₈D₂₉**H**₃₀V₃₁A₃₂I₃₃L₃₄S₃₅E₃₆E₃₇E₃₈E₃₉V₄₀
 E₄₁I₄₂E₄₃L₄₄L₄₅H₄₆E₄₇F₄₈E₄₉T₅₀I₅₁P₅₂V₅₃L₅₄S₅₅V₅₆E₅₇L₅₈S₅₉P₆₀E₆₁D₆₂V₆₃D₆₄A₆₅L₆₆E₆₇L₆₈D₆₉P₇₀A₇₁I₇₂S₇₃Y₇₄I₇₅E₇₆E₇₇D₇₈
 A₇₉E₈₀V₈₁T₈₂T₈₃M₈₄

变体 (参比) 前区B序列(SEQ ID NO: 11, G30)

A₁E₂E₃A₄K₅E₆K₇Y₈L₉I₁₀G₁₁F₁₂N₁₃E₁₄Q₁₅E₁₆A₁₇V₁₈S₁₉E₂₀F₂₁V₂₂E₂₃Q₂₄V₂₅E₂₆A₂₇N₂₈D₂₉**G**₃₀V₃₁A₃₂I₃₃L₃₄S₃₅E₃₆E₃₇E₃₈E₃₉V₄₀
 E₄₁I₄₂E₄₃L₄₄L₄₅H₄₆E₄₇F₄₈E₄₉T₅₀I₅₁P₅₂V₅₃L₅₄S₅₅V₅₆E₅₇L₅₈S₅₉P₆₀E₆₁D₆₂V₆₃D₆₄A₆₅L₆₆E₆₇L₆₈D₆₉P₇₀A₇₁I₇₂S₇₃Y₇₄I₇₅E₇₆E₇₇D₇₈
 A₇₉E₈₀V₈₁T₈₂T₈₃M₈₄

变体 (参比) 前区C序列 (SEQ ID NO: 14, GK 插入)

A₁**G**₂**K**₃E₄E₅A₆K₇E₈K₉Y₁₀L₁₁I₁₂G₁₃F₁₄N₁₅E₁₆Q₁₇E₁₈A₁₉V₂₀S₂₁E₂₂F₂₃V₂₄E₂₅Q₂₆V₂₇E₂₈A₂₉N₃₀D₃₁E₃₂V₃₃A₃₄I₃₅L₃₆S₃₇E₃₈E₃₉E₄₀
 E₄₁V₄₂E₄₃I₄₄E₄₅L₄₆L₄₇H₄₈E₄₉F₅₀E₅₁T₅₂I₅₃P₅₄V₅₅L₅₆S₅₇V₅₈E₅₉L₆₀S₆₁P₆₂E₆₃D₆₄V₆₅D₆₆A₆₇L₆₈E₆₉L₇₀D₇₁P₇₂A₇₃I₇₄S₇₅Y₇₆I₇₇E₇₈
 E₇₉D₈₀A₈₁E₈₂V₈₃T₈₄T₈₅M₈₆

变体 (参比) 前区D序列 (SEQ ID NO: 29; GKS 插入)

A₁**G**₂**K**₃**S**₄E₅E₆A₇K₈E₉K₁₀Y₁₁L₁₂I₁₃G₁₄F₁₅N₁₆E₁₇Q₁₈E₁₉A₂₀V₂₁S₂₂E₂₃F₂₄V₂₅E₂₆Q₂₇V₂₈E₂₉A₃₀N₃₁D₃₂E₃₃V₃₄A₃₅I₃₆L₃₇S₃₈E₃₉E₄₀E₄₁
 E₄₂V₄₃E₄₄I₄₅E₄₆L₄₇L₄₈H₄₉E₅₀F₅₁E₅₂T₅₃I₅₄P₅₅V₅₆L₅₇S₅₈V₅₉E₆₀L₆₁S₆₂P₆₃E₆₄D₆₅V₆₆D₆₇A₆₈L₆₉E₇₀L₇₁D₇₂P₇₃A₇₄I₇₅S₇₆Y₇₇I₇₈E₇₉
 E₈₀D₈₁A₈₂E₈₃V₈₄T₈₅T₈₆M₈₇

图3

前区氨基酸位置编号 (续)

变体 (参比) 前区E序列 (SEQ ID NO: 30; GKA 插入)

A1**G2K3A4**E5E6A7K8E9K10Y11L12L13G14F15N16E17Q18E19A20V21S22E23F24V25E26Q27V28E29A30N31D32E33V34A35I36L37S38E39E40E41E42V43E44I45E46L47L48H49E50F51E52T53I54P55V56L57S58V59E60L61S62P63E64D65V66D67A68L69E70L71D72P73A74I75S76Y77I78E79E80D81A82E83V84T85T86M87

变体 (参比) 前区F序列 (SEQ ID NO: 31; GKAA 插入)

A1**G2K3A4A5**E6E7A8K9E10K11Y12L13I14G15F16N17E18Q19E20A21V22S23E24F25V26E27Q28V29E30A31N32D33E34V35A36I37L38S39E40E41E42E43V44E45I46E47L48L49H50E51F52E53T54I55P56V57L58S59V60E61L62S63P64E65D66V67D68A69L70E71L72D73P74A75I76S77Y78I79E80E81D82A83E84V85T86T87M88

变体 (参比) 前区G序列 (SEQ ID NO: 32; L68K + I72V 取代)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29G30V31A32I33L34S35E36E37E38E39V40E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67**K68**D69P70A71**V72**S73Y74I75E76E77D78A79E80V81T82T83M84

变体 (参比) 前区H序列 (SEQ ID NO: 33; L68K + E80I 取代)

A1E2E3A4K5E6K7Y8L9I10G11F12N13E14Q15E16A17V18S19E20F21V22E23Q24V25E26A27N28D29G30V31A32I33L34S35E36E37E38E39V40E41I42E43L44L45H46E47F48E49T50I51P52V53L54S55V56E57L58S59P60E61D62V63D64A65L66E67**K68**D69P70A71I72S73Y74I75E76E77D78A79**I80**V81T82T83M84

图3(续)