(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117120618 A
(43)申请公布日 2023.11.24

(21)申请号 202280028013.2
(22)申请日 2022.02.11
(30)优先权数据
21156986.8 2021.02.12 EP
(85)PCT国际申请进入国家阶段日
2023.10.12
(86)PCT国际申请的申请数据
PCT/EP2022/053425 2022.02.11
(87)PCT国际申请的公布数据
WO2022/171827 EN 2022.08.18
(71)申请人 勃林格殷格翰 RCV 两合公司
地址 奥地利维也纳
申请人 瓦利多根股份有限公司
龙沙有限公司
(72)发明人 R·扎赫尔 O·阿塔·阿依库 D·马塔诺维奇 B·伽瑟
(74)专利代理机构 北京瑞恒信达知识产权代理事务所(普通合伙) 11382
专利代理师 丁磊
(51)Int.Cl.
C12N 15/62(2006.01)

权利要求书3页 说明书72页
序列表35页 附图1页

(54)发明名称

用于增加蛋白质分泌的信号肽

(57)摘要

本发明涉及一种编码融合蛋白的核酸分子，所述融合蛋白包含分泌信号和目的蛋白，所述分泌信号包含(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列；和任选地(ii)α-交配因子(MFα)前导序列。本发明进一步涉及一种如本文定义的分泌信号、包含所述核酸分子的表达盒以及包含所述核酸分子或表达盒的重组真核宿主细胞。还包括一种在真核宿主细胞中制备目的蛋白的方法和一种增加目的蛋白自真核宿主细胞中的分泌的方法。还提供了所述分泌信号用于增加重组目的蛋白自真核宿主细胞中的分泌的用途和重组宿主细胞用于制备重组目的蛋白的用途。

CN 117120618 A

1.一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

(a)分泌信号,所述分泌信号包含

(I)(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

(ii)α-交配因子(MFα)前导序列;

或

(II)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

(b)目的蛋白。

2.根据权利要求1所述的核酸分子,其中与表达权利要求1中定义的核酸分子,但该核酸分子包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非权利要求1中定义的分泌信号,的所述真核宿主细胞相比,所述分泌信号增加所述目的蛋白自真核宿主细胞中的分泌。

3.根据权利要求1或2所述的核酸分子,其中所述源自KRE1蛋白的信号肽序列包含SEQ ID NO:1或其功能同系物。

4.根据权利要求1或2所述的核酸分子,其中所述源自SWP1蛋白的信号肽序列包含SEQ ID NO:2或52或其功能同系物。

5.根据前述权利要求中任一项所述的核酸分子,其中所述MFα前导序列包含SEQ ID NO:3、53或74-80中的任一个或其功能同系物,优选地包含SEQ ID NO:3或53或其功能同系物。

6.根据前述权利要求中任一项所述的核酸分子,其中所述MFα前导序列在对应于SEQ ID NO:53的位置23的位置处包含Ser和/或在对应于SEQ ID NO:53的位置64的位置处包含Glu。

7.根据权利要求1至6中任一项所述的核酸分子,其中所述目的蛋白选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

8.根据权利要求1至7中任一项所定义的分泌信号。

9.一种表达盒或载体,其包含权利要求1至7中任一项所述的核酸分子和与其可操作地连接的启动子。

10.一种重组真核宿主细胞,其包含权利要求1至7中任一项所述的核酸分子,或权利要求9所述的表达盒或权利要求9所述的载体。

11.根据权利要求10所述的重组真核宿主细胞,其中所述宿主细胞是真菌或酵母宿主细胞。

12.根据权利要求11所述的重组真核宿主细胞,其中所述酵母宿主细胞选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、奇异酵母、真贝酵母、库德里阿兹威酵母、克鲁维酵母、葡萄汁酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母,或其中所述真菌宿主细胞选自

里氏木霉或黑曲霉。

13.根据权利要求10至12中任一项所述的重组真核宿主细胞,其中所述宿主细胞被工程化以过表达信号识别颗粒(SRP)的一种或多种组分。

14.一种在真核宿主细胞中制备目的蛋白的方法,所述方法包括

(i)用权利要求1至7中任一项所述的核酸分子或用权利要求9所述的表达盒或载体遗传工程化所述真核宿主细胞,并任选地遗传工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分;

(ii)在表达所述核酸分子和任选地过表达所述SRP的一种或多种组分,和在切割分泌信号后分泌目的蛋白的条件下培养遗传工程化的宿主细胞,

(iii)任选地从细胞培养物中分离目的蛋白,

(iv)任选地纯化所述目的蛋白,

(v)任选地修饰所述目的蛋白,和

(vi)任选地配制所述目的蛋白。

15.一种增加目的蛋白自真核宿主细胞中的分泌的方法,所述方法包括在所述真核宿主细胞中表达权利要求1至7中任一项所定义的核酸分子,和任选地工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分,从而与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非权利要求1至7中任一项所定义的所述分泌信号的权利要求1至7中任一项所述的核酸分子的所述宿主细胞相比增加所述目的蛋白的分泌。

16.根据权利要求14或15所述的方法,其包括

(i)工程化所述宿主细胞以引入表达权利要求1至7中任一项所述的核酸分子的表达构建体,和任选地遗传工程化所述宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分,

(ii)在适于表达所述核酸分子和任选地过表达所述SRP的一种或多种组分和在切割分泌信号后分泌目的蛋白的条件下培养所述宿主细胞,

(iii)任选地从细胞培养物中分离目的蛋白,

(iv)任选地纯化所述目的蛋白,

(v)任选地修饰所述目的蛋白,和

(vi)任选地配制所述目的蛋白。

17.根据权利要求14至16中任一项所述的方法,其中将所述核酸分子整合在所述宿主细胞的染色体中或包含在未整合到所述宿主细胞的基因组中的表达盒、载体或质粒中。

18.根据权利要求14至17中任一项所述的方法,其中所述真核宿主细胞是真菌或酵母宿主细胞。

19.根据权利要求14至18中任一项所述的方法,其中所述酵母宿主细胞选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母,或其中所述真菌宿主细胞选自里氏木霉或黑曲霉。

20.根据权利要求14至19中任一项所述的方法,其中所述目的蛋白选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽

的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

21.如权利要求1至8中任一项所定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途。

22.根据权利要求21所述的用途,其中与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非权利要求1中所定义的分泌信号的权利要求1中定义的融合蛋白的所述真核宿主细胞相比,所述分泌信号增加所述目的蛋白自真核宿主细胞中的分泌。

23.权利要求10至13中任一项所述的重组真核宿主细胞用于制备目的蛋白的用途。

24.根据权利要求1至7中任一项所述的融合蛋白,其中所述信号肽序列源自KRE1蛋白。

25.根据权利要求1至7中任一项所述的融合蛋白,其中所述信号肽序列源自SWP1蛋白。

26.一种分泌信号,所述分泌信号包含

(i)源自KRE1蛋白的信号肽序列;以及

(ii)α-交配因子(MFα)前导序列。

27.一种分泌信号,所述分泌信号包括

(i)源自SWP1蛋白的信号肽序列;以及

(ii)α-交配因子(MFα)前导序列。

28.一种通过在表达权利要求1至7中任一项所述的核酸分子和在切割分泌信号后分泌目的蛋白的条件下培养权利要求10至13中任一项所述的重组真核宿主细胞,并自宿主细胞培养物中分离目的蛋白,并任选地纯化和任选地修饰和任选地配制所述目的蛋白来生产目的蛋白的方法。

用于增加蛋白质分泌的信号肽

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年2月12日递交的欧洲专利申请21156986.8的优先权的权益，其内容在此通过引用整体并入以用于所有目的。

技术领域

[0003] 本发明涉及一种编码融合蛋白的核酸分子，所述融合蛋白包含分泌信号和目的蛋白，所述分泌信号包含(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列；和任选地(ii)α-交配因子(MFα)前导序列。本发明进一步涉及一种如本文所定义的分泌信号、包含所述核酸分子的表达盒以及包含所述核酸分子或表达盒的重组真核宿主细胞。本发明进一步涵盖一种在真核宿主细胞中制备目的蛋白的方法和一种增加目的蛋白自真核宿主细胞中的分泌的方法。本发明还提供了所述分泌信号用于增加重组目的蛋白自真核宿主细胞中的分泌的用途和重组宿主细胞用于制备重组目的蛋白的用途。

[0004] 发明背景

[0005] 酵母，特别是巴斯德毕赤酵母(P.pastoris，同物异名:Komagataella phaffii)是用于分泌重组蛋白的常用表达系统。分泌中的最初和关键步骤是重组蛋白向内质网(ER)中的易位。该过程由融合到重组蛋白的N-端分泌信号指导。信号序列指定了在常规分泌通路上到达ER的共翻译或翻译后靶向途径(Ng等，1996)。巴斯德毕赤酵母中最常用的分泌信号是酿酒酵母α-交配前原前导序列(MFα)(Lin-Cereghino等，2013)。该信号介导酿酒酵母中的翻译后易位，很可能在巴斯德毕赤酵母中也是如此(Fitzgerald和Glick，2014；Ng等，1996)。其它分泌信号不断地被加入到所有库中，并用不同的重组蛋白进行测试。

[0006] 由于许多哺乳动物蛋白的生物发生可能需要共翻译易位，MFα信号序列实际上可能是次优的，并且优选使用共翻译信号序列(Ng等，1996)。如今，哺乳动物抗体已经成为生物制药市场中的主要产品类别(Ecker等，2015)。已知抗体在其天然环境中共翻译易位(Feige等，2010)。发展更小抗原结合片段(如Fab、scFv和VHH)的趋势也是明显的(Nelson和Reichert，2009；Walsh，2014)。特别地，Fab片段有时分泌效率低，因此仅达到低生产滴度(titer)(Looser等，2015；Pfeffer等，2011)。这可能是由于已经被报道为引起易位瓶颈的翻译后信号序列MFα(Fitzgerald和Glick，2014；Zahrl等，2018)。WO2018165589A2和WO2018165594公开了一种重组分泌信号，其包含源自酿酒酵母的MFα前导序列和与源自酿酒酵母的MFα前导序列不同的信号肽。Fitzgerald等(Microb Cell Fact13，125(2014))公开了一种杂合分泌信号，其由后面是MFα前导序列的Ost1信号序列组成。

[0007] 因此，仍然需要增加各种蛋白(比如抗体)分泌的分泌信号。因此，技术问题是要满足这种需要。

发明内容

[0008] 上述技术问题通过权利要求中限定的主题来解决。本发明人惊奇地发现，包含源自KRE1蛋白(内部命名为SP14)或SWP1蛋白(内部命名为SP4)的信号肽序列、信号肽或前原

序列(所有术语可互换使用),任选与α-交配因子(MFα)前导序列组合的融合蛋白的分泌显著增加。换言之,包含本发明的分泌信号的蛋白将以较高速率分泌,同时分泌信号将被切除(参见实施例6-8)。

[0009] 因此,本发明涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0010] (a)分泌信号,所述分泌信号包含

[0011] (I)(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0012] (ii)α-交配因子(MFα)前导序列;

[0013] 或

[0014] (II)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0015] (b)目的蛋白。

[0016] 本发明涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0017] (a)分泌信号,所述分泌信号包含

[0018] (i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0019] (ii)α-交配因子(MFα)前导序列;以及

[0020] (b)目的蛋白。

[0021] 特别地,本发明提供编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0022] (a)分泌信号,所述分泌信号包含

[0023] (i)源自KRE1蛋白的信号肽序列;以及

[0024] (ii)α-交配因子(MFα)前导序列;以及

[0025] (b)目的蛋白。

[0026] 本发明还涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0027] (a)分泌信号,所述分泌信号包含

[0028] (i)源自SWP1蛋白的信号肽序列;以及

[0029] (ii)α-交配因子(MFα)前导序列;以及

[0030] (b)目的蛋白。

[0031] 本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0032] (a)分泌信号,所述分泌信号包含

[0033] (i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0034] (b)目的蛋白。

[0035] 特别地,本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0036] (a)分泌信号,所述分泌信号包含

[0037] (i)源自KRE1蛋白的信号肽序列;以及

[0038] (b)目的蛋白。

[0039] 本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0040] (a)分泌信号,所述分泌信号包含

[0041] (i)源自SWP1蛋白的信号肽序列;以及

[0042] (b)目的蛋白。

[0043] 本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0044] (a)分泌信号,所述分泌信号由

[0045] (i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列组成;以及

[0046] (b)目的蛋白。

[0047] 特别地,本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0048] (a)分泌信号,所述分泌信号由

[0049] (i)源自KRE1蛋白的信号肽序列组成;以及

[0050] (b)目的蛋白。

[0051] 本发明进一步涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0052] (a)分泌信号,所述分泌信号由

[0053] (i)源自SWP1蛋白的信号肽序列组成;以及

[0054] (b)目的蛋白。

[0055] 设想与表达本文定义的核酸分子、但该核酸分子包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本文定义的分泌信号,的真核宿主细胞相比,所述分泌信号增加目的蛋白自真核宿主细胞中的分泌。

[0056] 源自KRE1蛋白的信号肽序列可以包括SEQ ID NO:1或其功能同系物。源自KRE1蛋白的信号肽序列可由SEQ ID NO:1或其功能同系物组成。具体而言,SEQ ID NO:1的功能同系物与SEQ ID NO:1包含至少80%、或至少85%、或至少90%、或至少94%、或至少95%的序列同一性。具体而言,与SEQ ID NO:1相比,功能同系物包含一个、两个或三个点突变。具体而言,在例如真菌或酵母宿主细胞(比如Komagataella宿主细胞)的真核宿主细胞中功能同系物具有信号肽的功能。

[0057] 源自SWP1蛋白的信号肽序列可包括SEQ ID NO:2或52,或其功能同系物。源自SWP1蛋白的信号肽序列可由SEQ ID NO:2或52或其功能同系物组成。具体而言,SEQ ID NO:2或SEQ ID NO:52的功能同系物与相应的SEQ ID NO:2或SEQ ID NO:52包含至少80%、或至少85%、或至少90%、或至少94%、或至少95%的序列同一性,具体而言,与相应的SEQ ID NO:2或SEQ ID NO:52相比,功能同系物包含一个、两个或三个点突变。具体地,在例如真菌或酵母宿主细胞(比如Komagataella宿主细胞)的真核宿主细胞中功能同系物具有信号肽的功能。

[0058] MFα前导序列可以包含SEQ ID NO:3或53或74-80中的任何一个,或其功能同系物。MFα前导序列可以由SEQ ID NO:3或53或74-80中的任何一个或其功能同系物组成,优选由SEQ ID NO:3或53或其功能同系物组成。具体地,SEQ ID NO:3、53或74-80中任一个的功能同系物与相应的SEQ ID NO:3、53或74-80具有至少80%、或至少85%、或至少90%、或至少95%、或至少98%的序列同一性,具体地,与相应的SEQ ID NO:3、53或74-80相比,功能同系物包含一个、两个或三个点突变。具体地,在例如真菌或酵母宿主细胞(比如Komagataella

或Saccharomyces宿主细胞)的真核宿主细胞中功能同系物具有前导序列的功能。MFα前导序列优选在对应于SEQ ID NO:53的位置23的位置上包含Ser和/或在对应于SEQ ID NO:53的位置64的位置上包含Glu。

[0059] 目的蛋白选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

[0060] 在另一方面,本发明涉及一种如本文所定义的分泌信号。特别地,本发明涉及一种分泌信号,其包含(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;和(ii)α-交配因子(MFα)前导序列。更特别地,本发明涉及一种分泌信号,其包含(i)源自KRE1蛋白的信号肽序列和(ii)α-交配因子(MFα)前导序列。更特别地,本发明涉及一种分泌信号,其包含(i)源自SWP1蛋白的信号肽序列;和(ii)α-交配因子(MFα)前导序列。本发明还特别涉及一种包含源自KRE1蛋白的信号肽序列的分泌信号。进一步特别地,本发明涉及一种包含源自SWP1蛋白的信号肽序列的分泌信号。本发明还特别涉及一种由源自KRE1蛋白的信号肽序列组成的分泌信号。进一步特别地,本发明涉及一种由源自SWP1蛋白的信号肽序列组成的分泌信号。

[0061] 在另一方面,本发明还涉及一种包含本发明核酸分子和与其可操作连接的启动子的表达盒。该表达盒可以包含在载体(优选表达载体)中,或整合在染色体(特别是人工染色体)中。

[0062] 在另一方面,本发明还提供了一种包含本发明的核酸分子、本发明的载体或本发明的表达盒的重组真核宿主细胞。在本文中理解,用此类核酸分子或表达盒工程化的重组真核宿主细胞被遗传工程化以引入相应的核酸分子、载体或表达盒。可以将重组真核宿主细胞遗传工程化为在宿主细胞基因组内包含此类核酸分子、载体或表达盒。

[0063] 重组真核宿主细胞可以是真菌或酵母宿主细胞,优选为酵母宿主细胞,所述酵母宿主细胞选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母,或所述真菌宿主细胞选自里氏木霉或黑曲霉。

[0064] 宿主细胞可被工程化以过表达信号识别颗粒(SRP)的一种或多种组分。

[0065] 本发明还涉及一种通过在表达本发明的核酸分子和在切割分泌信号后分泌目的蛋白的条件下培养本发明的宿主细胞,并自宿主细胞培养物中分离目的蛋白,并任选地纯化和任选地修饰和任选地配制所述目的蛋白来生产目的蛋白的方法。

[0066] 在另一方面,本发明还涉及一种在真核宿主细胞中制备目的蛋白的方法,所述方法包括

[0067] (i)用本发明的核酸分子或用本发明的表达盒或载体遗传工程化所述真核宿主细胞,并任选地遗传工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分;

[0068] (ii)在表达所述核酸分子和任选地所述SRP的一种或多种组分,和在切割分泌信

号后分泌目的蛋白的条件下培养遗传工程化的宿主细胞，

[0069] (iii)任选地从细胞培养物中分离目的蛋白，

[0070] (iv)任选地纯化所述目的蛋白，

[0071] (v)任选地修饰所述目的蛋白，和

[0072] (vi)任选地配制所述目的蛋白。

[0073] 在另一方面，本发明还涉及一种增加目的蛋白自真核宿主细胞中的分泌的方法，所述方法包括在所述真核宿主细胞中表达本发明的核酸分子，和任选地工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分，从而与表达本发明的核酸分子、但该核酸分子包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本文所述的分泌信号，的宿主细胞相比增加所述目的蛋白的分泌。

[0074] 所述增加目的蛋白自真核宿主细胞中的分泌的方法可另外包括

[0075] (i)工程化所述宿主细胞以引入表达本发明的核酸分子的表达构建体，和任选地遗传工程化所述宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分，

[0076] (ii)在表达所述核酸分子和任选地过表达所述SRP的一种或多种组分和在切割分泌信号后分泌目的蛋白的条件下培养所述宿主细胞，

[0077] (iii)任选地从细胞培养物中分离目的蛋白，

[0078] (iv)任选地纯化所述目的蛋白，

[0079] (v)任选地修饰所述目的蛋白，和

[0080] (vi)任选地配制所述目的蛋白。

[0081] 可以将本发明的核酸分子整合到所述宿主细胞的染色体中，或者包含在未整合到所述宿主细胞的染色体中的表达盒、载体或质粒中。

[0082] 在另一方面，本发明还涉及本文所述的分泌信号(如，作为本发明核酸分子的一部分或在本发明核酸分子内)用于增加目的蛋白自真核宿主细胞中的分泌的用途。与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本发明所定义的分泌信号的本文所述的融合蛋白的所述真核宿主细胞相比，所述分泌信号进一步增加所述目的蛋白自真核宿主细胞中的分泌。

[0083] 在另一方面，本发明涉及本发明的重组宿主细胞用于制备目的蛋白的用途。

附图说明

[0084] 当分别结合非限制性实施例以及附图考虑时，参考详细描述能更好地理解本发明。附图示出了：

[0085] 图1:SPx-VHH(His6)克隆的免疫荧光抗-His染色。A:MFα分泌信号，B:SWP1(SP4)，C:KRE1(SP14):在荧光显微镜中用适当的滤片观察细胞。示出了荧光、DIC和合并图像。对图像进行了亮度和对比度调节。

具体实施方式

[0086] 下面将详细描述本发明，并通过所附的实施例和附图进行说明。

[0087] 本发明人惊奇地发现，包含KRE1(本文也称为SP14)或SWP1(本文也称为SP4)的信号肽(下文的信号肽序列)，(特别是当与诸如α-交配因子的前导序列的前导序列组合，从而

形成本发明的分泌信号时)的融合蛋白的分泌和产量显著增加(参见实施例6-8),而其它信号肽或组合不改善目的蛋白的分泌。因此,含有本发明分泌信号的融合蛋白被更有效地分泌。本发明人还发现,通过另外地过表达信号识别颗粒(SRP)的蛋白,目的蛋白的分泌和产量甚至进一步得到增加(参见实施例6-8)。

[0088] 如上文概述,与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本文定义的分泌信号的本文定义的融合蛋白的真核宿主细胞相比,重组宿主细胞更有效地分泌包含本发明分泌信号和目的蛋白的融合蛋白,即分泌本发明融合蛋白中包含的目的蛋白,同时在分泌过程中切除分泌信号。因此,本发明令人惊奇地证明,从N-端到C-端包含(a)分泌信号,所述分泌信号包含(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及(ii)任选地α-交配因子(MFα)前导序列;以及(b)目的蛋白的融合蛋白提供了优越的性能,如增加的目的蛋白的分泌(参见实施例6-8)。

[0089] 措辞“从N-端到C-端”不一定排除包括信号肽序列和α-交配因子(MFα)前导序列的分泌信号和目的蛋白被一个或多个氨基酸分隔开。这些一个或多个氨基酸可以是连接子或连接子序列。“连接子序列”(也称为“间隔序列”或“连接子”)是引入本文定义的分泌信号和本文定义的目的蛋白之间的氨基酸序列。“连接子序列”还可以是在信号肽序列和α-交配因子(MFα)前导序列之间和/或在α-交配因子(MFα)前导序列和目的蛋白之间引入的氨基酸序列。然而,优选地,在信号肽序列和α-交配因子(MFα)前导序列之间没有连接子。存在多种可能的连接子序列,并且基于例如本发明多肽的大小、序列和物理性质(如疏水性)选择合适的连接子序列在本领域技术人员的知识范围内。连接子序列可以由柔性残基(如甘氨酸和丝氨酸)或相对刚性的残基(如丙氨酸-脯氨酸)重复组成。为了确保最大的柔性,优选连接子序列不采用二级结构(例如α-螺旋结构或β-折叠)。连接子序列可以是,例如被特定蛋白酶识别,如被枯草杆菌蛋白酶/kexin样前蛋白转化酶(PC)家族的成员识别的蛋白酶切割位点。如本文所用的术语“连接子序列”或“连接子”是指不干扰所连接元件的功能的任何氨基酸序列。连接子可以连接如核苷酸序列或氨基酸序列。连接子可用于设计适当量的柔性。优选地,连接子是短的,如1-20个核苷酸或氨基酸或甚至更多,并且这些核苷酸或氨基酸通常是柔性的。通常使用的氨基酸连接子由许多甘氨酸、丝氨酸和任选的丙氨酸以任何顺序组成。根据需要,此类连接子通常具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15或20个氨基酸中的至少任何一个的长度。优选地,本文所用的连接子包含1至12个氨基酸残基,优选地,其为由至多5个氨基酸组成的短连接子。优选地,本文所用的连接子是GS、GGSGG、GSAGSAAGSG、(GS)n,其中“n”是1至10之间的任意数值,GSGSGSG、GSG或GGGGS(“G_4S”)连接子或它们的任意组合。在一些实施方式中,连接子包含基序(比如例如GS、GSG或G_4S)的一个或多个单元、重复或拷贝。

[0090] 可选地或另外地,融合蛋白可在MFα前导序列和目的蛋白之间包含切割位点。融合蛋白还可以包含一个或多个另外的切割位点,例如用于切割标签,如下所述。切割位点可以是,例如被特定蛋白酶识别的蛋白酶切割位点。因此,术语“蛋白酶切割位点”包括特定蛋白酶的识别位点,其是由蛋白酶特异性识别的氨基酸序列,和在蛋白的2个氨基酸之间的位点,在该位点发生切割。蛋白酶可选自TEV(烟草蚀纹病毒)蛋白酶、胰凝乳蛋白酶、肠激酶、胃蛋白酶、中性粒细胞弹性蛋白酶、蛋白酶K、嗜热菌蛋白酶、凝血酶和胰蛋白酶。通过特定蛋白酶发生切割的位点优选在如本文所定义的目的蛋白的N-端氨基酸和如本文所定义的

N-端融合的标签或N-端融合的α-交配因子(MFα)前导序列的C-端氨基酸之间,或在如本文所定义的目的蛋白的C-端氨基酸和C-端融合的标签或C-端融合的α-交配因子(MFα)前导序列的N-端氨基酸之间。识别位点可以与在本文定义的目的蛋白(POI)外部(在标签内)发生切割的相应位点相邻。更详细地,优选地,用于切割标签的蛋白酶选自肽链内切酶,如因子Xa、凝血酶、TEV(烟草蚀纹病毒蛋白酶)、天冬氨酸特异性半胱氨酸蛋白酶(半胱天冬酶)和肠激酶。因此,所用的识别位点可以选自因子Xa、凝血酶、TEV(烟草蚀纹病毒蛋白酶)、天冬氨酸特异性半胱氨酸蛋白酶(半胱天冬酶)和肠激酶识别位点,以及已知可用于从蛋白切割标签的蛋白酶的任何其它蛋白酶切割位点。半胱天冬酶-2的优选切割位点是VDVAD。

[0091] 如实施例中所示,与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本发明上下文中本文所述的分泌信号的融合蛋白的真核宿主细胞相比,分泌信号增加了融合蛋白的分泌,或更准确地说,增加了从中切除了分泌信号的目的蛋白从真核宿主细胞的分泌。因此,野生型酿酒酵母的MFα分泌信号(SEQ ID NO:4)可用作对照或参照以进行比较。

[0092] 根据本发明,由于作为由本发明核酸编码的融合蛋白的一部分的目的蛋白以及任选地SRP的一种或多种组分的(过)表达,即使当生物量保持低时,目的蛋白(POI,在分泌信号切割和分泌其自身之后)也可以以高产量获得。因此,在实验室、中试和工业规模中,以mg POI/g干生物量测量的高比产量可以在1至200的范围内,比如50至200,比如100至200的范围内。如本文所用的"增加的分泌"涉及与对照相比宿主细胞的上清液或培养基中更高量的可检测的目的蛋白;两者都在相同条件(如,宿主细胞物种、培养基、培养时间、培养温度、补料和诱导策略)下培养。对照可以是相同的宿主细胞,但在相同的宿主细胞中,目的蛋白表达为融合蛋白,该融合蛋白包含如SEQ ID NO:4所示的MFα分泌信号而非本发明的分泌信号。所述增加可以以分泌的倍数变化(FC)表示,例如增加至少1.1倍、至少1.2倍、至少1.3倍、至少1.4倍、至少1.5倍、至少1.6倍、至少1.7倍、至少1.8倍、至少1.9倍、至少2倍、至少2.5倍、至少3倍、至少5倍或至少10倍。宿主细胞的上清液或培养基中可检测的目的蛋白的量可以表示为以[目的蛋白g/细胞培养物L]计的体积滴度或以[目的蛋白mg/细胞干重g]计的产量等。分泌的倍数变化则是宿主细胞培养物的体积滴度或产率与对照培养物的体积滴度或产率的比率。

[0093] 融合蛋白或目的蛋白可进一步包含一个或多个(可检测的)标签、一个或多个蛋白酶切割位点和/或一个或多个连接子,它们例如在融合蛋白的一些元件(例如选自分泌信号、信号肽序列、α-交配因子(MFα)前导序列或目的蛋白的元件)之间和连接融合蛋白的一些元件,和/或作为任一元件的一部分,特别是作为目的蛋白的一部分。例如,连接子可以位于目的蛋白、标签和/或切割位点之间。因此,当本发明的融合蛋白从N-端到C-端由本文定义的分泌信号和本文定义的目的蛋白组成时,这种目的蛋白可任选地进一步包含一个或多个(可检测的)标签、一个或多个蛋白酶切割位点和/或一个或多个连接子。换言之,如果本发明的融合蛋白从N-端到C-端由本发明所述的分泌信号和目的蛋白组成,则本文定义的一个或多个标签、一个或多个切割位点和/或一个或多个连接子也可以N-端地或C-端地融合至此类目的蛋白,并由此也被所述目的蛋白包含。在此上下文中,N-端地或C-端地融合至此类目的蛋白的一种或多种标签、一种或多种切割位点和/或一种或多种连接子可以是该目的蛋白的一部分。此外,当本发明的融合蛋白从N-端到C-端由本文定义的分泌信号和同样

由本文定义的目的蛋白组成时,本文定义的分泌信号可任选地进一步包含一个或多个(可检测的)标签、一个或多个蛋白酶切割位点和/或一个或多个连接子。在此情况下,N-端地或C-端地融合至此类分泌信号、或融合至本文所定义的相应信号肽序列或融合至本文所定义的α-交配因子(MFα)前导序列的一个或多个标签、一个或多个蛋白酶切割位点和/或一个或多个连接子则可以是此类分泌信号的一部分。

[0094] 连接子是例如在[38]段所定义的。(可检测的)标签可以是用于纯化和/或增强表达和/或溶解度或目的蛋白检测的标签。有许多纯化、表达增强、溶解度增强标签并且本领域技术人员已知允许容易地检测和定量目的蛋白的标签。“纯化标签”(也称为“亲和标签”)是可用于例如纯化其所连接的蛋白(如在其N-端包括亲和标签的目的蛋白)的氨基酸序列。该标签对固体载体的适当配体如层析树脂或直接对树脂具有高亲和力。通过将包含纯化标签的目的蛋白选择性地结合到特定树脂上,可以仅通过一个层析步骤将目的蛋白高度有效地纯化。纯化标签是本领域技术人员已知的,并且可以是蛋白纯化标签,优选GST标签、FLAG标签、聚精氨酸标签、聚组氨酸标签(比如6His-标签)、MBP标签、S-标签、流感病毒HA标签、硫氧还蛋白标签或葡萄球菌蛋白A标签。更详细地,本文使用的纯化标签序列是以下中的任一种:组氨酸(His)标签,优选聚组氨酸标签(比如六组氨酸(6His)标签);精氨酸标签,优选聚精氨酸标签,抗体的肽底物,壳多糖结合结构域,RNAse S肽,蛋白A,S-半乳糖苷酶,FLAG标签,Strep II标签,链霉亲和素结合肽(SBP)标签,钙调蛋白结合肽(CBP),谷胱甘肽S-转移酶(GST),麦芽糖结合蛋白(MBP),S-标签,HA-标签,c-Myc标签或任何其它已知用于有效纯化与其融合的蛋白的标签。具体而言,可以通过IMAC,优选使用Ni-NTA层析材料捕获和纯化包括聚组氨酸或六组氨酸标签(His-标签)的融合蛋白(特别是目的蛋白)。在本发明的一个优选实施方案中,本文应用了包含在融合蛋白中或包含在如上文所定义的目的蛋白中的聚组氨酸或六组氨酸标签(His-标签),甚至更优选6-His标签。还存在许多蛋白酶切割位点,其可用于在从宿主细胞表达和/或分泌目的蛋白之后使用本领域技术人员已知的相应蛋白酶从目的蛋白切下标签。“表达和/或溶解度增强标签”可以C-或N-端地融合至本文定义的目的蛋白。与没有表达和/或溶解度增强标签的目的蛋白的表达相比,当在宿主细胞,例如原核或真核细胞、细菌细胞或酵母细胞,例如大肠杆菌,例如巴斯德毕赤酵母中表达时,例如在胞质、周质或自宿主细胞中分泌时,表达和/或溶解度增强标签可以显著增加目的蛋白的表达和/或滴度和/或溶解度和/或可溶性表达和/或可溶性滴度。本文所用的表达和/或溶解度增强标签序列可以选自钙调蛋白结合肽(CBP)、聚Arg、聚Lys、G B1结构域、蛋白D、葡萄球菌蛋白A的Z结构域和硫氧还蛋白或例如在宿主细胞中表达期间已知改善其所融合的蛋白的表达和/或溶解度的任何其它标签。表达和/或溶解度增强标签可以基于噬菌体基因的高度带电的肽,比如在US 8,535,908B2中列出的那些。优选地,所述溶解度增强标签选自T7C、T7B、T7B1、T7B2、T7B3、T7B3、T7B4、T7B5、T7B6、T7B6、T7B7、T7B8、T7B9、T7B10、T7B11、T7B12、T7B13、T7A、T7A1、T7A2、T7A3、T7A4、T7A5、T7AC T3、N1、N2、N3、N4、N5、N6、N7、钙调蛋白结合肽(CBP)、聚Arg、聚Lys、G B1结构域、蛋白D、葡萄球菌蛋白A的Z结构域、DsbA、DsbC和硫氧还蛋白及通过少数(比如1、2、3或甚至更多个)氨基酸置换获得的它们的变体。在本发明的一个优选实施方案中,本文应用了融合蛋白所包含的或上文定义的目的蛋白所包含的T7AC溶解度标签。“允许容易地检测和定量目的蛋白的标签”是可以C-或N-端地融合到如本文定义的目的蛋白的标签。它是一种可用于检测、定量、分析整个生产过程中的目的

蛋白的标签(例如,通过例如光谱或荧光或其它方法直接在线、原位、在线(on-line或atline)或图谱测定,或通过现有技术方法离线,测量发酵液中目的蛋白的滴度或整个生产过程中不同溶液(例如,在色谱洗脱液、细胞匀浆、过滤滞留物或滤液等)中POI的含量)。因此,标签可以例如赋予POI如荧光、UV-VIS吸光度、以及对于用于本领域已知的在线(on-line,in-line,atline)或原位定量方法的其它光谱或荧光方法有用的其它吸光度的特征。所述标签还可以是亲和标签或其它标签,其可用于定量亲和层析,例如作为离线测量的亲和HPLC或免疫测定(如ELISA)。例如,作为允许容易地检测和定量目的蛋白的标签的一个实例的监测标签是以下中的任一种:m-Cherry、GFP或f-肌动蛋白或用于在目的蛋白的生产期间(包括发酵、分离和纯化)通过简单的原位、在线或在线检测器(如UV、IR、拉曼、荧光等)检测或定量目的蛋白的任何其它标签。作为允许容易地检测和定量目的蛋白的标签的另一个实例,检测标签还可以是蛋白A标签、S-半乳糖苷酶标签、FLAG标签、Strep标签或链霉亲和素结合肽(SBP)标签或用在定量HPLC或ELISA中的Strep II标签。

[0095] 在本发明的一个实施方案中,包含本文定义的分泌信号(比如包含源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列,任选包含α-交配因子(MFα)前导序列)以及包含本文定义的目的蛋白的融合蛋白还包含溶解度增强标签(甚至更优选地为T7AC),和/或纯化标签(甚至更优选地为6His标签),和/或半胱天冬酶(优选半胱天冬酶-2)的蛋白酶切割位点(比如VDVAD),优选地其中本文定义的溶解度增强标签和/或本文定义的纯化标签和/或本文定义的蛋白酶切割位点N-端地融合至本文定义的目的蛋白。详细地,在本发明的一个实施方案中,包含分泌信号以及包含本文定义的目的蛋白的融合蛋白还包含溶解度增强标签(甚至更优选地为T7AC),和/或纯化标签(甚至更优选地为6His标签),和/或半胱天冬酶(优选半胱天冬酶-2)的蛋白酶切割位点(比如VDVAD),所述分泌信号包含源自KRE1蛋白的信号肽序列和α-交配因子(MFα)前导序列,优选其中本文定义的溶解度增强标签和/或本文定义的纯化标签和/或本文定义的蛋白酶切割位点N-端地融合至本文定义的目的蛋白。详细地,在本发明的另一个实施方案中,包含分泌信号以及包含本文定义的目的蛋白的融合蛋白还包含溶解度增强标签(甚至更优选地为T7AC),和/或纯化标签(甚至更优选地为6His标签),和/或半胱天冬酶(caspase)(优选半胱天冬酶-2)的蛋白酶切割位点(比如VDVAD),所述分泌信号包含源自SWP1蛋白的信号肽序列和α-交配因子(MFα)前导序列,优选其中本文定义的溶解度增强标签和/或本文定义的纯化标签和/或本文定义的蛋白酶切割位点N-端地融合至本文定义的目的蛋白。在本发明的优选实施方案中,包含本文定义的分泌信号(例如其包含源自KRE1蛋白或源自SWP1蛋白的信号肽序列并包含α-交配因子(MFα)前导序列)以及包含本文定义的目的蛋白的融合蛋白还包含溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD,甚至更优选地其中溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD N-端地融合到本文定义的目的蛋白。当本发明的融合蛋白从N-端至C-端由本文定义的分泌信号(例如其包含源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列,任选低包含α-交配因子(MFα)前导序列)和本文也定义的目的蛋白组成时,此类目的蛋白还可任选地包含溶解度增强标签(甚至更优选地为T7AC),和/或纯化标签(甚至更优选地为6His标签),和/或半胱天冬酶(优选半胱天冬酶-2)的蛋白酶切割位点(比如VDVAD),优选地其中本文定义的溶解度增强标签和/或本文定义的纯化标签和/或本文定义的蛋白酶切割位点N-端地融合至本文定义的目的蛋白。在此上下文中,此类标签和/或切割

位点是目的蛋白的一部分。在一个优选的实施方案中,当本发明的融合蛋白从N-端至C-端由包含源自KRE1蛋白的信号肽序列和α-交配因子(MFα)前导序列的分泌信号和本文也定义的目的蛋白组成时,此类目的蛋白还包含溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD,甚至更优选地其中溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD N-端地融合至本文定义的目的蛋白。在另一个优选实施方案中,当本发明的融合蛋白从N-端到C-端由包含源自SWP1蛋白的信号肽序列和α-交配因子(MFα)前导序列的分泌信号和本文也定义的目的蛋白组成时,此类目的蛋白还包含溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD,甚至更优选地其中溶解度增强标签T7AC和6His纯化标签和蛋白酶切割位点VDVAD N-端地融合至本文定义的目的蛋白。同样,在此情况下,所述T7AC和6His标签和所述VDVAD切割位点是目的蛋白的一部分。

[0096] 分泌信号

[0097] 为了分泌,蛋白必须通过产生它的细胞的细胞内分泌途径。该通过N-端分泌信号将蛋白导向该途径,而非导向备选的细胞目的地。至少,分泌信号包含信号肽序列。信号肽序列通常由13至36个主要是疏水性氨基酸组成,其侧翼为N-端碱性氨基酸和C-端极性氨基酸。信号肽序列可与信号识别颗粒(SRP)或其它转运蛋白(如SND、GET)相互作用,所述信号识别颗粒或其它转运蛋白介导新生蛋白从胞质向ER腔中的共翻译或翻译后转运。在ER中,通常将信号肽序列切除,蛋白折叠并经历翻译后修饰。然后将蛋白从ER递送到高尔基体,然后递送到分泌囊泡和细胞外部。除了信号肽序列之外,天然指定用于分泌的新生蛋白的亚集携带分泌信号,所述分泌信号还包含前导肽,比如α-交配因子前导序列。前导肽通常由被带电或极性氨基酸中断的疏水性氨基酸组成。不希望受理论的束缚,认为前导肽减慢了转运并确保了蛋白的正确折叠,和/或促进了蛋白从ER向高尔基体的转运,在高尔基体中前导肽通常被切除。本文所用的“源自KRE1或SWP1蛋白的信号肽序列”描述了一种氨基酸序列,即信号肽序列,其存在于本文所定义的KRE1蛋白或本文所定义的SWP1蛋白中。由于包括信号肽序列的分泌信号通常在分泌过程中被切除,因此本文所述的信号肽序列在分泌前和/或在信号肽序列被切除前源自KRE1蛋白或SWP1。“源自”可以与“衍生自”互换使用。

[0098] KRE1,也称为杀伤毒素抗性蛋白1,是一种酵母分泌的蛋白。KRE1可能涉及细胞壁1,6-β-葡聚糖合成和装配的晚期阶段。它在细胞壁1,6-β-葡聚糖装配和结构中具有结构性而非酶促功能,这种功能可能是通过参与将1,6-β-葡聚糖与诸如1,3-β-葡聚糖、壳多糖和某些甘露糖蛋白的其他细胞壁成分共价交联而实现的。KRE1还作为酵母K1病毒毒素的质膜受体。因此,KRE1携带信号肽序列。KRE1可以来自任何真核物种,特别是来自任何酵母。示例性的酵母包括但不限于Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、奇异酵母、真贝酵母、库德里阿兹威酵母、克鲁维酵母、葡萄汁酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母。KRE1也可来自里氏木霉或黑曲霉。优选地,KRE1来自K.phaffii。源自KRE1的信号肽序列可包括全长KRE1蛋白(即包含从编码KRE1蛋白的mRNA翻译的信号肽的蛋白)的前18个氨基酸或其功能同系物,或由其组成,所述全长KRE1蛋白例如是来自K.phaffii的KRE1蛋白。在优选的实施方案中,KRE1蛋白对应于UniProt数据库条目F2QWV3,2011年5月31日的序列版本1(染色体位置PP7435_Chr3-0933)或其功能同系物,其中信号肽序列优选对应于也在SEQ ID NO:1中描述的所述数据库条目的氨基酸1-18。因此,源自KRE1蛋白的信号肽序列可包含SEQ ID

NO:1或其功能同系物或由其组成。源自KRE1蛋白的信号肽序列可由SEQ ID NO:1或其功能同系物组成。源自KRE1蛋白的信号肽序列可与SEQ ID NO:1具有至少80％、85％、90％、94％或95％序列同一性中的任一种,其可指如本文定义的其功能同系物。源自KRE1蛋白的信号肽序列可与SEQ ID NO:1具有至少90％的序列同一性。源自KRE1蛋白的信号肽序列可与SEQ ID NO:1具有至少94％的序列同一性。源自KRE1蛋白的信号肽序列可与SEQ ID NO:1具有至少95％的序列同一性。

[0099] SWP1,也称为多萜基二磷酸寡聚糖蛋白葡糖基转移酶亚基SWP1,其是寡聚糖基转移酶(OST)复合物的一种亚基,其催化确定的聚糖(真核生物中的$Glc_3Man_9GlcNAc_2$)从脂质载体多萜醇-焦磷酸盐至新生多肽链中Asn-X-Ser/Thr共有基序内的天冬酰胺残基的初始转移,这是蛋白N-糖基化中的第一步。SWP1也携带信号肽序列。SWP1可以来自任何真核物种,特别是来自任何酵母。示例性的酵母包括但不限于Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、奇异酵母、真贝酵母、库德里阿兹威酵母、克鲁维酵母、葡萄汁酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.、Komagataella pastoris和粟酒裂殖酵母。SWP1也可来自里氏木霉或黑曲霉。优选地,SWP1来自Komagataella phaffii。源自SWP1的信号肽序列可以包括全长SWP1(即包括从编码SWP1蛋白的mRNA翻译的信号肽的蛋白)蛋白的前18个氨基酸或由其组成,所述全长SWP1例如是来自K.phaffii或其功能同系物的SWP1蛋白。在优选的实施方案中,SWP1蛋白对应于UniProt数据库条目F2QNI3,2011年5月31日的序列版本1(基因PP7435_Chr1-0255)或其功能同系物,其中信号肽序列优选对应于也在SEQ ID NO:2中描述的所述数据库条目的氨基酸1-18。因此,源自SWP1蛋白的信号肽序列可包含SEQ ID NO:2或其功能同系物或由其组成。源自SWP1蛋白的信号肽序列可由SEQ ID NO:2或其功能同系物组成。源自SWP1蛋白的信号肽序列可与SEQ ID NO:2具有至少80％、85％、90％、94％或95％序列同一性中的任一种,其可指如本文定义的其功能同系物。源自SWP1蛋白的信号肽序列可与SEQ ID NO:2具有至少90％的序列同一性。源自SWP1蛋白的信号肽序列可与SEQ ID NO:2具有至少94％的序列同一性。源自SWP1蛋白的信号肽序列可与SEQ ID NO:2具有至少95％的序列同一性。优选地,SWP1来自巴斯德毕赤酵母。因此,源自SWP1蛋白的信号肽序列可包含SEQ ID NO:52或其功能同系物。源自SWP1蛋白的信号肽序列可由SEQ ID NO:52或其功能同系物组成。源自SWP1蛋白的信号肽序列可与SEQ ID NO:52具有至少80％、85％、90％、94％或95％序列同一性中的任一种,其可指如本文定义的其功能同系物。源自SWP1蛋白的信号肽序列与SEQ ID NO:52具有至少90％的序列同一性。源自SWP1蛋白的信号肽序列与SEQ ID NO:52具有至少95％的序列同一性。

[0100] α-交配因子(MFα),也称为交配因子α-1、α-1交配信息素或交配因子α,是一种激素,其中活性因子(无分泌信号的MFα)由α交配型的单倍体细胞分泌到培养基中,并作用于相反交配型(a型)的细胞。MFα通过抑制a型细胞中DNA合成的启动并使它们与α型同步来介导两种类型之间的缀合过程。MFα携带包含信号肽序列(前原序列)和前导序列的分泌信号。MFα可以来自任何真核物种,特别是来自任何酵母,优选来自酵母属的任何酵母,比如,奇异酵母(Saccharomyces paradoxus)、真贝酵母、酿酒酵母、克鲁维酵母、葡萄汁酵母或库德里阿兹威酵母。示例性的酵母包括但不限于Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、奇异酵母、真贝酵母、库德里阿兹威酵母、克鲁维酵母、葡萄汁酵母、乳

酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母。MFα也可以来自里氏木霉或黑曲霉。优选地,来源是酿酒酵母。前导序列可包含全长MFα蛋白(即包括由编码MFα蛋白的mRNA翻译的信号肽的蛋白和前导序列)的氨基酸20-89或由其组成,所述全长MFα蛋白例如来自酿酒酵母的MFα蛋白或其功能同系物。在优选的实施方案中,全长MFα蛋白对应于UniProt数据库条目P01149,1988年4月1日的序列版本1或其功能同系物,其中优选地,α交配因子(MFα)前导序列对应于所述数据库条目的氨基酸20-89,更优选地对应于SEQ ID NO:53所示的氨基酸20-85。MFα前导序列可以包含SEQ ID NO:3或其功能同系物。MFα前导序列可由SEQ ID NO:3或其功能同系物组成。MFα前导序列可与SEQ ID NO:3具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指如本文定义的其功能同系物。MFα前导序列可与SEQ ID NO:3具有至少为90%序列同一性。MFα前导序列可与SEQ ID NO:3具有至少为95%序列同一性。MFα前导序列可与SEQ ID NO:3具有至少为98%序列同一性。MFα前导序列可包含SEQ ID NO:53或其功能同系物。MFα前导序列可由SEQ ID NO:53或其功能同系物组成。MFα原序列可与SEQ ID NO:53具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指如本文定义的其功能同系物。MFα前导序列可与SEQ ID NO:53具有至少为90%序列同一性。MFα前导序列可与SEQ ID NO:53具有至少为95%序列同一性。MFα前导序列可与SEQ ID NO:53具有至少为98%序列同一性。

[0101] MFα前导序列可源自奇异酵母。MFα前导序列可以包含SEQ ID NO:74或其功能同系物。MFα前导序列可由SEQ ID NO:74或其功能同系物组成。MFα前导序列可与SEQ ID NO:74具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:74具有至少90%的序列同一性。MFα前导序列可与SEQ ID NO:74具有至少95%的序列同一性。MFα前导序列可与SEQ ID NO:74具有至少98%的序列同一性。MFα前导序列可以包含SEQ ID NO:75或其功能同系物。MFα前导序列可由SEQ ID NO:75或其功能同系物组成。MFα前导序列可与SEQ ID NO:75具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:75具有至少90%的序列同一性。MFα前导序列可与SEQ ID NO:75具有至少95%的序列同一性。MFα前导序列可与SEQ ID NO:75具有至少98%的序列同一性。MFα前导序列可以包含SEQ ID NO:76或其功能同系物。MFα前导序列可由SEQ ID NO:76或其功能同系物组成。MFα前导序列可与SEQ ID NO:76具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:76具有至少90%的序列同一性。MFα前导序列可与SEQ ID NO:76具有至少95%的序列同一性。MFα前导序列可与SEQ ID NO:76具有至少98%的序列同一性。MFα前导序列可以包含SEQ ID NO:77或其功能同系物。MFα前导序列可由SEQ ID NO:77或其功能同系物组成。MFα前导序列可与SEQ ID NO:77具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:77具有至少90%的序列同一性。MFα前导序列可与SEQ ID NO:77具有至少95%的序列同一性。MFα前导序列可与SEQ ID NO:77具有至少98%的序列同一性。MFα前导序列可以包含SEQ ID NO:78或其功能同系物。MFα前导序列可由SEQ ID NO:78或其功能同系物组成。MFα前导序列可与SEQ ID NO:78具有至少80%、85%、90%、95%或98%序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:78具有至少90%的序列同一性。MFα前导序列可与SEQ ID NO:

78具有至少95％的序列同一性。MFα前导序列可与SEQ ID NO:78具有至少98％的序列同一性。

[0102] MFα前导序列可源自真贝酵母。MFα前导序列可以包含SEQ ID NO:79或其功能同系物。MFα前导序列可由SEQ ID NO:79或其功能同系物组成。MFα前导序列可与SEQ ID NO:79具有至少80％、85％、90％、95％或98％序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:79具有至少90％的序列同一性。MFα前导序列可与SEQ ID NO:79具有至少95％的序列同一性。MFα前导序列可与SEQ ID NO:79具有至少98％的序列同一性。

[0103] MFα前导序列可源自库德里阿兹威酵母。MFα前导序列可以包含SEQ ID NO:80或其功能同系物。MFα前导序列可由SEQ ID NO:80或其功能同系物组成。MFα前导序列可与SEQ ID NO:80具有至少80％、85％、90％、95％或98％序列同一性中的任一种,其可指本文所定义的其功能同系物。MFα前导序列可与SEQ ID NO:80具有至少90％的序列同一性。MFα前导序列可与SEQ ID NO:80具有至少95％的序列同一性。MFα前导序列可与SEQ ID NO:80具有至少98％的序列同一性。

[0104] 优选地,MFα前导序列在对应于SEQ ID NO:53的位置23的位置处具有Ser和/或在对应于SEQ ID NO:53的位置64的位置处具有Glu,优选在两个位置都具有。这可以进一步增加分泌。SEQ ID NO:3已经含有这些突变。

[0105] 功能同系物是本文所述核酸序列或肽、多肽或蛋白的功能等价物。功能同系物可以是与给定的多肽序列,比如SEQ ID NO:1、2、3、52或53的序列具有至少约70％、至少约80％、至少约90％或至少约95％氨基酸序列同一性的生物活性序列。在一些实施方案中,功能同系物是与天然多肽序列具有至少约70％、至少约80％、至少约90％或至少约95％氨基酸序列同一性的生物活性序列。对于核酸序列,遗传密码的简并性允许用限定相同氨基酸的其它密码子置换一些密码子并由此产生相同的蛋白。核酸序列可以有很大变化,因为除了甲硫氨酸和色氨酸之外,已知的氨基酸可以由一个以上的密码子编码。因此,可以合成本文所述的部分或全部核酸序列,以得到与其所示序列明显不同的核酸序列。然而,其编码的氨基酸序列将被保留。

[0106] 功能同系物也可以描述本文中描述的氨基酸序列或肽、多肽或蛋白的功能等价物,其具有至多5个保守突变,即功能同系物可以具有1、2、3、4或5个保守突变。在一些实施方案中,特别是对于MFα前导序列的功能同系物,该功能同系物也可以具有1、2、3、4、5、6、7、8、9、10、11、12、13或14个保守突变。本文所用的“保守突变”优选是导致点突变的突变,所述点突变为如一个氨基酸的置换、插入或缺失,特别是其中所述置换是用化学上相似的氨基酸残基对氨基酸残基的替代。保守置换的实例为以下组的成员之间的替代:1)丙氨酸、丝氨酸和苏氨酸;2)天冬氨酸和谷氨酸;3)天冬酰胺和谷氨酰胺;4)精氨酸和赖氨酸;5)异亮氨酸、亮氨酸、甲硫氨酸和缬氨酸;和6)苯丙氨酸、酪氨酸和色氨酸。保守突变也可以是不影响本文所述肽、多肽或蛋白的氨基酸序列的生物活性的一个氨基酸的任何置换、插入或缺失。功能同系物可具有至多14个保守突变。功能同系物可具有至多13个保守突变。功能同系物可具有至多12个保守突变。功能同系物可以具有至多11个保守突变。功能同系物可具有10个保守突变。功能同系物可以具有至多9个保守突变。功能同系物可具有至多8个保守突变。功能同系物可以具有至多7个保守突变。功能同系物可具有6个保守突变。功能同系物可具

有至多5个保守突变。功能同系物可以具有至多4个保守突变。功能同系物可以具有至多3个保守突变。功能同系物可以具有至多2个保守突变。功能同系物可具有1个保守突变。

[0107] 表1:部分分泌信号的概览。

[0108]

信号	核苷酸序列（5'-3'）	氨基酸序列
SP4（*K. phaffii* 的 SWP1 的信号肽序列）	ATGAAGCTGATCTCCGTGGGTA TAGTGACGACATTACTGACTTT GGCCAGTTGC (SEQ ID NO: 18)	MKLISVGIVTTLLTLASC (SEQ ID NO: 2)
K. pastoris 的 SWP1 的信号肽序列		MKLFFVGIVTTLLTLVSC (SEQ ID NO: 52)
SP14（*K. phaffii* 的 KRE1	ATGTTAAACAAGCTGTTCATTG	MLNKLFIAILIVITAVIG (SEQ

[0109]

的信号肽序列）	CAATACTCATAGTCATCACTGC TGTCATAGGC (SEQ ID NO: 19)	ID NO: 1)
酿酒酵母的 WT MFα 前导序列		APVNTTTEDETAQIPAEAVI GYLDLEGDFDVAVLPFSNST NNGLLFINTTIASIAAKEEGV SLDKR (SEQ ID NO: 53)
基于酿酒酵母的 WT MFα 前导序列的 MFα 前导序列，包括两个突变：L23S, D64E	GCCCCTGTTAACACTACCACTG AAGACGAGACTGCTCAAATTCC AGCTGAAGCAGTTATCGGTTAC TCTGACCTTGAGGGTGATTTCG ACGTCGCTGTTTTGCCTTTCTCT AACTCCACTAACAACGGTTTGT TGTTCATTAACACCACTATCGC TTCCATTGCTGCTAAGGAAGAG GGTGTCTCTCTCGAGAAGAGA (SEQ ID NO: 20)	APVNTTTEDETAQIPAEAVI GYSDLEGDFDVAVLPFSNST NNGLLFINTTIASIAAKEEGV SLEKR (SEQ ID NO: 3)
MFα 分泌信号（前原前导序列），其可用作基于酿酒酵母的 WT MFα 分泌信号的比较的参照，包括两	ATGAGATTCCCATCTATTTTCA CCGCTGTCTTGTTCGCTGCCTCC TCTGCATTGGCTGCCCCTGTTA ACACTACCACTGAAGACGAGA	MRFPSIFTAVLFAASSALAAP VNTTTEDETAQIPAEAVIGY SDLEGDFDVAVLPFSNSTNN GLLFINTTIASIAAKEEGVSL

[0110]

个突变:L23S，D64E	CTGCTCAAATTCCAGCTGAAGC AGTTATCGGTTACTCTGACCTT GAGGGTGATTTCGACGTCGCTG TTTTGCCTTTCTCTAACTCCACT AACAACGGTTTGTTGTTCATTA ACACCACTATCGCTTCCATTGC TGCTAAGGAAGAGGGTGTCTCT CTCGAGAAGAGA (SEQ ID NO: 21)	EKR (SEQ ID NO: 4)

[0111] 本发明还涉及一种如本文定义的分泌信号。本文还公开了一种分泌信号,所述分泌信号包含(i)源自KRE1蛋白的信号肽序列;和(ii)α-交配因子(MFα)前导序列。信号肽序列可包含SEQ ID NO:1或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:3或其功能同系物。信号肽序列可包含SEQ ID NO:1或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:53或其功能同系物。本发明还涉及一种分泌信号,所述分泌信号包括(i)源自SWP1蛋白的信号肽序列;和(ii)α-交配因子(MFα)前导序列。信号肽序列可包含SEQ ID NO:2或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:3或其功能同系物。信号肽序列可包含SEQ ID NO:2或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:53或其功能同系物。信号肽序列可包含SEQ ID NO:52或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:3或其功能同系物。信号肽序列可包含SEQ ID NO:52或其功能同系物,和α-交配因子(MFα)前导序列可包含SEQ ID NO:53或其功能同系物。

[0112] 目的蛋白

[0113] 本文所用的术语“目的蛋白”(POI)是指通过重组技术在宿主细胞中产生的蛋白,更具体而言,该蛋白可以是在宿主细胞中非天然存在的多肽,即一种异源蛋白,如人工蛋白,比如野生型细胞非天然产生的蛋白,或者可以是宿主细胞天然的蛋白,即宿主细胞的同源蛋白,但是例如通过用含有编码该POI的核酸序列的自复制载体转化,或者通过重组技术将编码该POI的核酸序列的一个或多个拷贝整合到宿主细胞的基因组中,或者通过重组修饰控制编码该POI的基因的表达的一个或多个调控序列(例如启动子序列)而产生。一般而言,本文所指的目的蛋白可以通过本领域技术人员熟知的重组表达方法产生。目的蛋白可以是重组蛋白。

[0114] 本文对目的蛋白(POI)没有限制。通常,POI是真核或原核多肽,其变体或衍生物,或人工多肽,比如野生型细胞非天然产生的蛋白。POI可以是任何真核或原核蛋白。所述蛋

白可以是天然分泌蛋白,也可以是细胞内蛋白(即非天然分泌的蛋白)。本发明还包括蛋白的生物活性片段。在另一个实施方案中,POI可以是氨基酸链或以复合物如二聚体、三聚体、异-二聚体、多聚体或寡聚体存在。POI与本发明分泌信号的融合可以使任何POI得到分泌。POI可以是需要共翻译易位的蛋白。

[0115] 目的蛋白可以是用作营养的、膳食的(dietary)、消化的、补充剂的蛋白,如用在食品、饲料或化妆品中。所述食品可以是,例如,肉汤、甜点、谷物棒、糕饼(confectionery)、运动饮料、膳食产品或其他营养产品。优选地,所述目的蛋白是食品添加剂。

[0116] 在另一个实施方案中,目的蛋白可用在动物饲料中。

[0117] POI的进一步实例包括抗微生物蛋白,如乳铁蛋白、溶菌酶、乳铁素(lactoferricin)、lactohedrin、κ-酪蛋白、咕啉结合蛋白(haptocorrin)、乳过氧化物酶、牛奶蛋白、急性期蛋白(例如在生产动物中响应感染而正常产生的蛋白)和如溶菌酶和乳铁蛋白的小抗微生物蛋白。其它实例包括抗菌蛋白、抗病毒蛋白、急性期蛋白(在生产动物中响应感染而诱导的)、益生菌蛋白、抑菌蛋白和阳离子抗微生物蛋白。

[0118] "饲料"是指旨在或适合用于非人动物食用、摄入、消化的任何天然或人工饮食、膳食等或这些膳食的组份。"饲料添加剂"通常是指添加在饲料中的物质。饲料添加剂通常包括如维生素、矿物质、酶和合适的载体和/或赋形剂的一种或多种化合物。对于本发明而言,食品添加剂可以是酶或其它蛋白。可用作饲料添加剂的酶的实例包括植酸酶、木聚糖酶和β-葡聚糖酶。"食品"是指旨在或适合用于人食用、摄入、消化的任何天然或人工饮食膳食等或这些膳食的组份。

[0119] "食品添加剂"通常是指添加在食物中的物质。食品添加剂通常包括如维生素、矿物质、酶和合适的载体和/或赋形剂的一种或多种化合物。对于本发明而言,食品添加剂可以是酶或其他蛋白质。可用作食品添加剂的酶的实例包括蛋白酶、脂酶、乳糖酶、果胶甲基酯酶(pectin methyl esterase)、果胶酶、谷氨酰胺转移酶、淀粉酶、β-葡聚糖酶、乙酰乳酸脱羧酶和漆酶。

[0120] 在一些实施方案中,所述食品添加剂是抗微生物蛋白,所述抗微生物蛋白包括:例如,(i)抗微生物牛奶蛋白(人或非人)乳铁蛋白、溶菌酶、乳铁素(lactoferricin)、lactohedrin、κ-酪蛋白、咕啉结合蛋白(haptocorrin)、乳过氧化物酶、α-1-抗胰蛋白酶和如IgA的免疫球蛋白;(ii)急性期蛋白,如C-反应蛋白(CRP)、乳铁蛋白、溶菌酶、血清淀粉样蛋白A(SAA)、铁蛋白、结合珠蛋白(Hp)、补体(complement)2-9(特别是补体-3)、血清粘蛋白、铜蓝血浆蛋白(Cp)、15-酮-13,14-二氢-前列腺素F2α(PGFM)、纤维素原(Fb)、α(1)-酸性糖蛋白(AGP)、α(1)-抗胰蛋白酶、甘露糖结合蛋白、脂多糖结合蛋白(lipoplysaccharide binding protein)、α-2巨球蛋白和各种防御素;(iii)抗微生物肽,如杀菌肽、爪蟾抗菌肽(magainin)、防御素、速普肽(tachyplesin)、parasin 1.buforin I、PMAP-23、moronecidin、anoplin、gambicin和SAMP-29;和(iv)其它抗微生物蛋白,包括CAP37、粒溶蛋白、分泌型白细胞蛋白酶抑制剂、CAP18、ubiquicidin、牛抗微生物蛋白-1、Ace-AMP1、速普肽(tachyplesin)、大防御素(big defensin)、Ac-AMP2、Ah-AMP1和CAP18。

[0121] POI可以是酶。优选的酶是可用于如制备洗涤剂、淀粉、燃料、纺织品、纸浆和纸、油、个人护理产品,或烘烤、有机合成等工业应用的那些。这些酶的实例包括用于污渍去除和清洁的蛋白酶、淀粉酶、脂酶、甘露聚糖酶和纤维素酶(cellulose);用于淀粉液化和糖化

的支链淀粉酶淀粉酶和淀粉葡糖苷酶;用于葡萄糖向果糖转化的葡萄糖异构酶;用于环糊精生产的环糊精-糖基转移酶;用于燃料和淀粉中粘度降低(xiscosity reduction)的木聚糖酶;用于烘烤中面团的稳定性和调整的淀粉酶、木聚糖酶、脂酶、磷脂酶、葡萄糖酶(glucose)、氧化酶、脂氧合酶、谷氨酰胺转移酶;在纺织品制备中用于牛仔布整理和棉花软化的纤维素酶;用于纺织品的脱浆的淀粉酶;用于精练的果胶裂解酶;用于漂白终止的过氧化氢酶;用于漂白的漆酶;用于过量染料去除的过氧化物酶;用在纸浆和纸制造中的脂酶、蛋白酶、淀粉酶、木聚糖酶、纤维素酶(cellulose);用于酯交换反应的脂酶和用于脂肪和油的脂肪加工时脱胶的磷脂酶;用于有机合成中手性醇和酰胺的分辨的脂酶;用于半合成青霉素的合成的酰基转移酶,用于对应异构体羧酸的合成的腈水解酶;用于皮革生产的蛋白酶和脂酶;用于制备个人护肤品的淀粉葡萄糖苷酶、葡萄糖氧化酶和过氧化物酶(参见Kirk等,Current Opinion in Biotechnology(2002)13:345-351)。

[0122] POI可以是治疗性蛋白。POI可以是但不限于,适合作为如抗体或抗体片段、生长因子、激素、酶或疫苗等的生物制药物质的蛋白。

[0123] 所述POI可以是天然分泌的蛋白或细胞内蛋白(即非天然分泌的蛋白)。本发明还提供天然分泌或非天然分泌的蛋白的功能同系物、功能等价变体、衍生物和生物活性片段的重组生产。优选地,功能同系物是与序列相同或相应的并具有序列的功能特征。

[0124] 所述POI可以与天然蛋白结构类似,和可以通过在C-和/或N-端,或者天然蛋白的侧链上添加一个或多个氨基酸,在天然氨基酸序列上的一个或数个不同位点置换一个或多个氨基酸,在天然蛋白的任一个或两个末端或者在氨基酸序列的一个或若干位点上缺失一个或多个氨基酸,或在天然氨基酸序列的一个或多个位点上插入一个或多个氨基酸,而衍生自天然蛋白质。上述若干蛋白的这些修饰是熟知的。

[0125] 优选地,所述目的蛋白是哺乳动物多肽,或甚至更优选地是人多肽。优选地,所述目的蛋白是治疗性或生物制药蛋白。特别优选的治疗性蛋白是指可施用于哺乳动物,甚至更优选地施用于人的任何多肽、蛋白、蛋白质变体、融合蛋白和/或其片段。目的蛋白也可以是人工蛋白,或一种或多种天然蛋白或人工蛋白或融合蛋白的一部分。设想但不是必须的,根据本发明的治疗性蛋白是细胞异源的。可通过本发明的细胞生产的蛋白的实例是但不限于,酶、调控蛋白、受体、肽激素、生长因子、细胞因子、支架结合蛋白(如基于脂质运载蛋白家族的突变蛋白)、结构蛋白、淋巴因子、粘附分子、受体、膜或转运蛋白,和可作为激动剂或拮抗剂和/或具有治疗或诊断用途的任何其他多肽。此外,所述目的蛋白可以是用于接种的抗原、疫苗、抗原结合蛋白、免疫刺激蛋白。其还可以是可包括本领域已知的任何合适的抗原结合抗体片段的抗体的抗原结合片段。例如,抗体片段可包括但不限于:Fv(一种包含VL和VH的分子)、单链Fv(scFV)(一种包含由肽连接子连接的VL和VH的分子)、Fab、Fab'、F(ab')$_2$、单结构域抗体(sdAb)(包含单可变结构域和3CDR的分子)和他们的多价存在形式(presentation)。所述抗体或其片段可以是鼠、人、人源化或嵌合的抗体或其片段。治疗性蛋白的实例包括:抗体、多克隆抗体、单克隆抗体、重组抗体、抗体片段,如Fab'、F(ab')$_2$、Fv、scFv、di-scFv、bi-scFv、串联scFv、双特异性串联scFv、sdAb、VHH、V_H和V_L,或人抗体、人源化抗体、嵌合抗体、IgA抗体、IgD抗体、IgE抗体、IgG抗体、IgM抗体、胞内抗体、微抗体或使用纤连蛋白III型结构域(FN3)作为分子支架构建的合成结合蛋白(也称为单抗体(monobody))。

[0126] 目的蛋白可进一步选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

[0127] 治疗性蛋白包括但不限于:胰岛素、胰岛素样生长因子、hGH、tPA、细胞因子,例如如IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、IL-13、IL-14、IL-15、IL-16、IL-17、IL-18的白细胞介素、干扰素(IFN)α、IFNβ、IFNγ、IFNω或IFNτ、肿瘤坏死因子(TNF)TNFα和TNFβ、TRAIL;G-CSF、GM-CSF、M-CSF、MCP-1和VEGF。

[0128] 在优选的实施方案中,所述蛋白是抗体。术语"抗体"旨在包括含有具有适合并识别表位的特异性形状的分子结构的任何多肽链,其中一个或多个非共价结合相互作用稳定该分子结构与所述表位之间的复合物。原型的抗体分子是免疫球蛋白,且来自所有来源(如人、啮齿动物、兔、牛、羊、猪、狗、其它哺乳动物、鸡、其它禽类等)的所有类型的免疫球蛋白(IgG、IgM、IgA、IgE、IgD等)都被认为是"抗体"。已经描述了多个编码抗体的序列;其它抗体可通过本领域熟知的方法提出。

[0129] 例如,可通过本领域已知的方法生产抗体或抗原结合抗体片段。通常抗体生产细胞对所需抗原或免疫原敏感。从抗体生产细胞中分离的信使RNA用作采用PCR扩增制备cDNA的模板。通过将扩增的免疫球蛋白cDNA的合适区段插入表达载体中来生产载体文库,其中每一个载体含有保留了原始抗原特异性的一个重链基因和一个轻链基因。通过将重链基因文库与轻链基因文库组合构建组合文库。这得到了共表达重链和轻链(类似于抗体分子的Fab片段或抗原结合片段)的克隆文库。将携带这些基因的载体共转染至宿主细胞。当在转染的宿主中诱导抗体基因合成时,所述重链和轻链蛋白自组装以生产能够通过用抗原或免疫原筛选而检测的活性抗体。

[0130] 编码抗体的目标序列包括由天然序列编码的那些,以及借助于遗传密码的简并性在序列上与已公开的核酸不同的核酸和它们的变体。变体多肽可包括氨基酸(aa)置换、添加或缺失。所述氨基酸置换可以是保守性氨基酸置换或消除非必须氨基酸的置换,例如以改变糖基化位点,或通过置换或缺失一个或多个对功能而言非必需的半胱氨酸残基来使错误折叠最少。可对变体进行设计从而使其保留或具有蛋白特定区域(如,功能结构域、起催化作用的氨基酸残基等)的增强的生物活性。变体还包括本文所公开的多肽的片段,特别是生物活性片段和/或对应于功能结构域的片段。克隆基因的体外诱变技术是已知的。本发明的主题还包括采用一般分子生物技术修饰从而改善了它们对蛋白质降解的抵抗的或优化了溶解性能的或使它们更适合作为治疗剂的多肽。

[0131] 可通过将可变轻链和重链区(VK和VH)组合的重组方式来制备嵌合的抗体,所述可变轻链和重链区从一个物种的抗体生产细胞中获得,而恒定轻链和重链区来自另一物种。通常,为了生产主要是人结构域的抗体,嵌合的抗体利用啮齿动物或兔可变区和人恒定区。这些嵌合的抗体的生产是本领域熟知的,并且可通过标准手段实现(如,例如在美国专利No.5,624,659中所描述的)。

[0132] 人源化抗体被工程化为含有甚至更多人样免疫球蛋白结构域,并仅结合

(incorporate)动物来源抗体的互补决定区。这一过程通过仔细检查单克隆抗体可变区的高变环序列,并使这些高变环序列与人抗体链的结构相装配来完成。尽管从表面上看是复杂的,但是这一过程已被直接实践。参见例如美国专利No.6,187,287。

[0133] 除了完整的免疫球蛋白(或它们的重组对应物),可合成包含表位结合位点(如Fab'、F(ab')$_2$或其他片段)的免疫球蛋白片段。可利用重组免疫球蛋白技术设计"片段"或最小免疫球蛋白。比如可通过合成可变轻链区和可变重链区生产本发明所用的"Fv"免疫球蛋白。还对抗体的组合感兴趣,如包含两个不同Fv特异性的双抗体。

[0134] 可在翻译后修饰免疫球蛋白,如添加化学连接子、可检测部分(例如荧光染料)、酶、作用物(substrate)、化学发光部分等,或在本发明的方法和组合物中使用特异性结合部分,例如链霉亲和素、抗生物素蛋白或生物素等。

[0135] 治疗性蛋白的进一步实例包括凝血因子(VII、VIII、IX)、镰刀菌属碱性蛋白酶、降钙素、CD4受体达贝泊汀(darbepoetin)、DNA酶(囊性纤维化)、促红细胞生成素、eutropin(人生长激素衍生物)、促卵泡激素(促滤泡素)、明胶、胰高血糖素、葡糖脑苷脂酶(戈谢病)、源自黑曲霉(A.niger)的葡糖淀粉酶、源自黑曲霉的葡萄糖氧化酶、促性腺激素、生长因子(GCSF、GMCSF)、生长激素(生长激素(somatotropine))、乙肝疫苗、水蛭素、人抗体片段、人载脂蛋白AI、人降钙素前体、人胶原酶IV、人表皮生长因子、人胰岛素样生长因子、人白细胞介素6、人层粘连蛋白、人前脱脂蛋白AI、人血清白蛋白、胰岛素和突变蛋白、胰岛素、干扰素α和突变蛋白、干扰素β、干扰素γ(突变蛋白)、白细胞介素2、促黄体生成激素、单克隆抗体5T4、小鼠胶原蛋白、OP-1(成骨的,神经保护因子)、奥普瑞白介素(白细胞介素11-激动剂)、有机磷酸水解酶(organophosphohydrolase)、PDGF-激动剂、肌醇六磷酸酶、血小板衍化生长因子(PDGF)、重组血纤溶酶原激活剂G、葡萄球菌激酶、干细胞因子、破伤风毒素片段C、组织型纤溶酶原激活物和肿瘤坏死因子(参见Schmidt,Appl Microbiol Biotechnol(2004)65:363-372)。

[0136] 目的蛋白可以包含SEQ ID NO:26描述的氨基酸序列或由其组成。目的蛋白可以包含SEQ ID NO:27描述的氨基酸序列或由其组成。目的蛋白可以包含SEQ ID NO:28描述的氨基酸序列或由其组成。目的蛋白可以包含SEQ ID NO:29描述的氨基酸序列或由其组成。当本发明的融合蛋白从N-端到C-端由如本文定义的分泌信号和如SEQ ID NO:26、27、28或29描述的目的蛋白组成时,此类目的蛋白可以任选地进一步包含如本文其它地方定义的一个或多个(可检测的)标签、一个或多个蛋白酶切割位点和/或一个或多个连接子。换言之,如果本发明的融合蛋白从N-端至C-端由本上下文中定义的分泌信号和此类目的蛋白组成,则本文中定义的一个或多个标签、一个或多个切割位点和/或一个或多个连接子也可以N-或C-端地融合至SEQ ID NO:26、27、28或29描述的此类目的蛋白,从而同样包含在所述目的蛋白中。在本上下文中,如本文其它定义的,N-或C-端融合至此类目的蛋白的所述一个或多个标签、一个或多个切割位点和/或一个或多个接头则可以是该目的蛋白的一部分。同样,如本文已经定义的,当本发明的融合蛋白从N-端到C-端由本文定义分泌信号和SEQ ID NO:26、27、28或29描述的目的蛋白组成时,本文定义的分泌信号可任选地进一步包含一个或多个(可检测的)标签、一个或多个蛋白酶切割位点和/或一个或多个连接子。

[0137] 表2:待表达和分泌的示例性蛋白的序列。

[0138]

目的基因	编码 DNA 序列(5'-3')	蛋白序列（* 终止密码子）
VHH（可以在 pAOX1 下表	CAGGTTCAGCTGCAGGAGT	QVQLQESGGGLVQAGGSLRLS

[0139]

达，CYC1 可以用作终止子），（Prielhofer 等，2017）	CCGGTGGTGGTCTGGTTCA AGCCGGTGGTTCATTAAGA TTGTCCTGTGCTGCCTCTGG TAGAACTTTCACTTCTTTCG CAATGGGTTGGTTTAGACA AGCACCTGGAAAAGAGAG AGAGTTTGTTGCTTCTATCT CCAGATCCGGTACTTTAAC TAGATACGCTGACTCTGCC AAGGGTAGATTCACTATTT CTGTTGACAACGCCAAGAA CACTGTTTCTTTGCAAATG GACAACCTTAACCCAGATG ACACCGCAGTCTATTACTG TGCCGCTGACTTGCACAGA CCATACGGTCCAGGAACCC AAAGATCCGATGAGTACGA TTCTTGGGGTCAGGGAACT CAAGTCACTGTCTCTTCAG GTGGTGGATCTGGTGGTGG AGGTTCAGGTGGTGGAGGA	CAASGRTFTSFAMGWFRQAPG KEREFVASISRSGTLTRYADSA KGRFTISVDNAKNTVSLQMDN LNPDDTAVYYCAADLHRPYGP GTQRSDEYDSWGQGTQVTVSS GGGSGGGGSGGGGSGGGGSGG GGSGGGEVQLVESGGALVQPG GSLRLSCAASGFPVNRYSMRW YRQAPGKEREWVAGMSSAGD RSSYEDSVKGRFTISRDDARNT VYLQMNSLKPEDTAVYYCNVN VGFEYWGQGTQVTVSSGGHHH HHH** (SEQ ID NO: 26)

[0140]

	TCCGGTGGTGGTGGTTCTG	
	GTGGTGGTGGATCTGGTGG	
	AGGTGAAGTTCAACTTGTC	
	GAATCCGGTGGTGCACTTG	
	TCCAACCTGGTGGATCTCT	
	TAGACTTTCTTGTGCCGCCT	
	CCGGTTTTCCTGTTAACCGT	
	TACTCTATGCGTTGGTACA	
	GACAAGCCCCTGGAAAAG	
	AACGTGAATGGGTTGCCGG	
	AATGTCCTCAGCTGGTGAC	
	AGATCCTCCTACGAAGATT	
	CTGTGAAGGGACGTTTCAC	
	CATCTCCAGAGATGACGCC	
	CGTAACACCGTTTACCTTC	
	AAATGAACTCCCTTAAGCC	
	TGAGGATACTGCCGTCTAC	
	TATTGTAACGTGAATGTCG	
	GATTTGAATACTGGGGACA	
	GGGAACCCAAGTTACTGTC	
	TCTTCCGGTGGACATCACC	

[0141]

	ACCACCATCACTAATAG （SEQ ID NO: 22）	
scR（可以在 pAOX1 下表达，CYC1 可以用作终止子）（Prielhofer 等， 2017）	CAGGAACAACTAATGGAGT CTGGGGGTGGTTTGGTTAC CCTGGGTGGTTCTCTTAAG CTTTCATGTAAGGCCTCTG GTATTGATTTTTCGCACTAC GGTATCTCCTGGGTTAGAC AAGCTCCTGGAAAAGGTCT GGAATGGATCGCTTACATT TACCCAAATTACGGTTCTG TTGACTATGCCTCCTGGGT CAATGGTAGGTTCACTATT TCCCTTGACAACGCTCAGA ACACGGTATTCCTACAGAT GATCTCCCTAACCGCTGCT GATACTGCAACCTACTTCT GTGCTCGTGACAGAGGTTA CTACTCTGGCTCTCGTGGA ACTAGACTTGACTTATGGG GACAAGGTACTCTCGTTAC	QEQLMESGGGLVTLGGSLKLS CKASGIDFSHYGISWVRQAPGK GLEWIAYIYPNYGSVDYASWV NGRFTISLDNAQNTVFLQMISL TAADTATYFCARDRGYYSGSR GTRLDLWGQGTLVTISSGGGGS GGGGSGGGGSELVMTQTPPSLS ASVGETVRIRCLASEFLFNGVS WYQQKPGKPPKFLISGASNLES GVPPRFSGSGSGTDYTLTIGGV QAEDVATYYCLGGYSGSSGLT FGAGTNVEIKGGHHHHHH** (SEQ ID NO: 27)

[0142]

	CATCTCTAGTGGTGGAGGT	
	GGTTCTGGAGGAGGAGGTT	
	CCGGCGGAGGTGGTAGCG	
	AGCTGGTCATGACTCAAAC	
	CCCTCCATCCCTATCTGCAT	
	CAGTCGGTGAAACCGTTAG	
	AATTAGATGCCTTGCATCT	
	GAGTTCTTGTTCAACGGTG	
	TGTCCTGGTATCAACAAAA	
	GCCTGGTAAGCCTCCAAAG	
	TTTCTCATTTCTGGTGCCTC	
	AAACCTCGAATCTGGAGTG	
	CCACCAAGATTTTCCGGAT	
	CTGGCTCTGGTACTGACTA	
	CACTCTGACAATTGGTGGT	
	GTTCAAGCTGAGGATGTTG	
	CTACCTACTATTGTCTCGGT	
	GGTTACTCAGGATCTTCCG	
	GCCTAACTTTCGGTGCCGG	
	TACAAACGTCGAGATCAAA	
	GGTGGACATCACCACCACC	

[0143]

	ATCACTAATAG (SEQ ID NO: 23)	
SDZ-Fab- HC（可以在pAOX1 下表达，CYC1 可以用作终止子）（Prielhofer 等，2017）	GAGGTCCAATTGGTCCAAT CTGGTGGAGGATTGGTTCA ACCAGGTGGATCTCTGAGA TTGTCTTGTGCTGCTTCTGG TTTCACCTTCTCTCACTACT GGATGTCATGGGTTAGACA AGCTCCTGGTAAGGGTTTG GAATGGGTTGCTAACATCG AGCAAGATGGATCAGAGA AGTACTACGTTGACTCTGT TAAGGGAAGATTCACTATT TCCCGTGATAACGCCAAGA ACTCCTTGTACCTGCAAAT GAACTCCCTTAGAGCTGAG GATACTGCTGTCTACTTCT GTGCTAGAGACTTGGAAGG TTTGCATGGTGATGGTTAC TTCGACTTATGGGGTAGAG GTACTCTTGTCACCGTTTCA	EVQLVQSGGGLVQPGGSLRLSC AASGFTFSHYWMSWVRQAPG KGLEWVANIEQDGSEKYYVDS VKGRFTISRDNAKNSLYLQMNS LRAEDTAVYFCARDLEGLHGD GYFDLWGRGTLVTVSSASTKG PSVFPLAPCSRSTSESTAALGCL VKDYFPEPVTVSWNSGALTSG VHTFPAVLQSSGLYSLSSVVTV PSSSLGTKTYTCNVDHKPSNTK VDKRVESKYGPP** (SEQ ID NO: 28)

[0144]

	TCTGCCTCTACCAAAGGAC CTTCTGTGTTCCCATTAGCT CCATGTTCCAGATCCACCT CCGAATCTACTGCAGCTTT GGGTTGTTTGGTGAAGGAC TACTTTCCTGAACCAGTGA CTGTCTCTTGGAACTCTGG TGCTTTGACTTCTGGTGTTC ACACCTTTCCTGCAGTTTTG CAGTCATCTGGTCTGTACT CTCTGTCCTCAGTTGTCACT GTTCCTTCCTCATCTCTTGG TACCAAGACCTACACTTGC AACGTTGACCATAAGCCAT CCAATACCAAGGTTGACAA GAGAGTTGAGTCCAAGTAT GGTCCACCTTAATAG (SEQ ID NO: 24)	
SDZ-Fab- LC（可以在 pDAS1 下表达，TDH1 可以用作终止子）（Prielhofer	GCTATCCAGTTGACTCAAT CACCATCCTCTTTGTCTGCT TCTGTTGGTGATAGAGTCA	AIQLTQSPSSLSASVGDRVILTC RASQGVSSALAWYQQKPGKAP KLLIYDASSLESGVPSRFSGSGS

[0145]

等，2017）	TCCTGACTTGTCGTGCATCT CAAGGTGTTTCCTCAGCTT TAGCTTGGTACCAACAAAA GCCAGGTAAAGCTCCAAAG TTGCTGATCTACGACGCTT CATCCCTTGAATCTGGTGT TCCTTCACGTTTCTCTGGAT CTGGATCAGGTCCTGATTT CACTCTGACTATCTCATCC CTTCAACCAGAGGACTTTG CTACCTACTTCTGTCAACA GTTCAACTCTTACCCTTTGA CCTTTGGAGGTGGAACTAA GTTGGAGATCAAGAGAACT GTTGCTGCACCATCAGTGT TCATCTTTCCTCCATCTGAT GAGCAACTGAAGTCTGGTA CTGCATCTGTTGTCTGCTTA CTGAACAACTTCTACCCAA GAGAAGCTAAGGTCCAATG GAAGGTTGACAATGCCTTG	GPDFTLTISSLQPEDFATYFCQQ FNSYPLTFGGGTKLEIKRTVAA PSVFIFPPSDEQLKSGTASVVCL LNNFYPREAKVQWKVDNALQS GNSQESVTEQDSKDSTYSLSST LTLSKADYEKHKVYACEVTHQ GLSSPVTKSFNRGEC** (SEQ ID NO: 29)

[0146]

	CAATCTGGTAACTCTCAAG AGTCTGTTACTGAGCAAGA CTCTAAGGACTCTACTTAC TCCCTTTCTTCCACCTTGAC TTTGTCTAAGGCTGATTAC GAGAAGCACAAGGTTTACG CTTGTGAGGTTACTCACCA AGGTTTGTCCTCTCCTGTTA CCAAGTCTTTCAACAGAGG TGAATGCTAATAG (SEQ ID NO: 25)	
包含FLAG标签的mCherry，其以灰色和粗体突出显示（可以在pAOX1下表达，CYC1可以用作终止子）	GTGAGCAAGGGCGAGGAGG ATAACATGGCCATCATCAAG GAGTTCATGCGCTTCAAGGT GCACATGGAGGGCTCCGTGA ACGGCCACGAGTTCGAGATC GAGGGCGAGGGCGAGGGCC GCCCCTACGAGGGCACCCAG ACCGCCAAGCTGAAGGTGAC CAAGGGTGGCCCCCTGCCCT TCGCCTGGGACATCCTGTCC CCTCAGTTCATGTACGGCTCC AAGGCCTACGTGAAGCACCC CGCCGACATCCCCGACTACT TGAAGCTGTCCTTCCCCGAG GGCTTCAAGTGGGAGCGCGT	VSKGEEDNMAIIKEFMRFKVH MEGSVNGHEFEIEGEGEGRPYE GTQTAKLKVTKGGPLPFAWDIL VSKGEEDNMAIIKEFMRFKVH MEGSVNGHEFEIEGEGEGRPYE GTQTAKLKVTKGGPLPFAWDIL SPQFMYGSKAYVKHPADIPDYL KLSFPEGFKWERVMNFEDGGV VTVTQDSSLQDGEFIYKVKLRG TNFPSDGPVMQKKTMGWEASS

[0147]

	GATGAACTTCGAGGACGGCG GCGTGGTGACCGTGACCCAG GACTCCTCCCTGCAGGACGG CGAGTTCATCTACAAGGTGA AGCTGCGCGGCACCAACTTC CCCTCCGACGGCCCCGTAAT GCAGAAAAAGACCATGGGCT GGGAGGCCTCCTCCGAGCGG ATGTACCCCGAGGACGGCGC CCTGAAGGGCGAGATCAAGC AGAGGCTGAAGCTGAAGGA CGGCGGCCACTACGACGCTG AGGTCAAGACCACCTACAAG GCCAAGAAGCCCGTGCAGCT GCCCGGCGCCTACAACGTCA ACATCAAGTTGGACATCACC TCCCACAACGAGGACTACAC CATCGTGGAACAGTACGAAC GCGCCGAGGGCCGCCACTCC ACCGGCGGCATGGACGAGCT GTACAAG**GGTGGAGATTAC** **AAGGATGACGATGATAAGT** **AATAG** (SEQ ID NO: 54)	ERMYPEDGALKGEIKQRLKLK DGGHYDAEVKTTYKAKKPVQL PGAYNVNIKLDITSHNEDYTIVE QYERAEGRHSTGGMDELYK**G** **GDYKDDDDK** (SEQ ID NO: 55)

[0148] 本发明还包括,本文所述的融合蛋白包含分泌信号的元件,如本文进一步所述,其可操作地连接至本文所定义的目的蛋白。

[0149] 根据一个具体的方面,本发明涉及一种编码融合蛋白的核酸分子,所述融合蛋白包含:

[0150] (a)分泌信号,所述分泌信号包含

[0151] (I)(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0152] (ii)α-交配因子(MFα)前导序列;

[0153] 或

[0154] (II)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0155] (b)目的蛋白,

[0156] 其中所述分泌信号可操作地连接至目的蛋白。

[0157] 本发明的核酸分子

[0158] 为了利用本发明的分泌信号,可将它们与目的蛋白融合。如本文所述的此类融合蛋白可以由核酸分子编码。本发明的核酸分子可以例如转化到宿主细胞中。因此,本发明涉及一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含(a)分泌信号,所述分泌信号包含(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;和(ii)α-交配因子(MFα)前导序列;和(b)目的蛋白。

[0159] 如本文所用,“编码”意指当表达编码蛋白的核酸或多核苷酸时,其导致所述蛋白的产生。

[0160] 本文所用的术语“核酸分子”是指DNA或RNA。“核酸分子”、“核酸”、“多核苷酸”或简单地“核苷酸”都可以互换使用,它们是指从5’至3’端读出的脱氧核糖核苷酸或核糖核苷酸碱基的单链或双链聚合物。其包括自复制质粒、DNA或RNA的感染性聚合物和非功能DNA或RNA。

[0161] 本文所用的“表达盒”涉及(载体)DNA的独特组分,其包含诸如本发明的核酸分子的基因和待由转染细胞表达的调节序列(比如启动子)。表达盒可以指导宿主细胞的机器(machinery)表达目的蛋白。通常,表达盒由一个或多个基因和控制其表达的序列组成。因此,本发明还涉及一种包含本发明的核酸分子和与其可操作连接的启动子的表达盒。表达盒可以是载体的形式。表达盒可以包含在载体中。

[0162] 在其它实施方案中,可以将本发明的核酸分子和/或编码SRP的一种或多种组分的多核苷酸整合在质粒或载体中。因此,本发明还涉及一种包含本发明的核酸的载体。术语“质粒”可以涉及用作载体以人工携带外源遗传物质进入另一细胞的DNA分子,在另一细胞中质粒可以被复制和/或表达(例如质粒、粘粒、λ噬菌体)。含有外源DNA的载体称为重组DNA。载体包括但不限于,质粒、病毒载体、粘粒和人工染色体,优选为质粒。载体本身通常可以是由插入物(转基因)和作为载体“骨架”的较大序列组成的DNA序列。所有载体都可以用于克隆,因此可以看作是克隆载体,但是也有专门为克隆设计的载体,而其它载体可以专门为其它目的设计,所述其它目的例如转录和蛋白表达。特别设计用于在靶细胞中表达转基因的载体称为表达载体,并且通常具有驱动转基因表达的启动子序列。技术人员能够根据所使用的宿主细胞采用合适的质粒或载体。

[0163] 优选地,载体是真核表达载体,优选为酵母表达载体。

[0164] 优选地,载体是真核表达载体,优选为酵母表达载体。使用酵母作为宿主的载体的实例包括YIp型载体、YEp型载体、YRp型载体、YCp型载体、pGPD-2、pAO815、pGAPZ、pGAPZα、pHIL-D2、pHIL-S1、pPIC3.5K、pPIC9K、pPICZ、pPICZα、pPIC3K、pHWO10、pPUZZLE和2μm质粒。此类载体是已知的,并例如在Cregg等,Mol Biotechnol.(2000)16(1):23-52。优选地,载体是pPM2dZ30载体(描述于WO2008/128701A2)。或者,载体是基于Golden Gate的GoldenPiCS,其由骨架BB1、BB2和BB3aK/BB3eH/BB3rN组成(Prielhofer等,2017)。

[0165] 载体可用于在适宜的宿主生物体中克隆的重组核苷酸序列(即重组基因)的转录和它们mRNA的翻译。载体还可用于通过本领域已知的方法,如J.Sambrook等,Molecular Cloning:A Laboratory Manual(3rd edition),Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory Press,New York(2001)或Stearns等,(1990),Methods in Enzymology,185:280-297描述的方法,将靶多核苷酸整合至宿主细胞基因组中。“载体”通

常包括用于在宿主细胞中自主复制的起点(优选本发明宿主细胞的细菌起点和真核起点)、选择标记、多个限制性酶切割位点、适宜的启动子序列和转录终止子,所述组份可操作地连接在一起。将编码多肽的目的序列可操作地连接至在宿主细胞中提供多肽表达的转录和翻译调控序列。

[0166] 本领域技术人员已知大量适宜的质粒或载体,并且很多质粒或载体是商购可获得的。Sambrook等编辑,Molecular Cloning:A Laboratory Manual(2nd Ed.),Vols.1-3,Cold Spring Harbor Laboratory(1989),和Ausubel等编辑,Current Protocols in Molecular Biology,John Wiley&Sons,Inc.,New York(1997)提供了适宜的载体的实例。

[0167] 本发明的载体或质粒涵盖包含端粒(telomeric)、着丝粒(centromeric)和复制的起点(复制起点)序列的酵母人工染色体,所述酵母人工染色体是指可以被遗传修饰为含有异源DNA序列(如,大到3000kb的DNA序列)的DNA构建体。

[0168] 编码本发明融合蛋白或分泌信号的序列可以可操作地连接至启动子。本文所用术语“启动子”是指有助于特定基因的转录的区域。与无启动子存在时所表达的重组产物的量相比,启动子通常使从核苷酸序列表达的重组产物的量增加。可利用源自一个生物体的启动子来增强从源自另一生物体的序列的重组产物表达。可以通过采用本领域已知的方法的同源重组(如,Datsenko等,Proc.Natl.Acad.Sci.U.S.A.,97(12):6640-6645(2000))来将启动子整合至宿主细胞染色体中。此外,一个启动子元件能够增加串联连接的多条序列所表达的产物的量。因此,一个启动子元件能够增强一种或多种重组产物的表达。

[0169] 启动子可以是“诱导型启动子”或“组成型启动子”。“诱导型启动子”是指可以被某些因子的存在或缺少诱导的启动子,“组成型启动子”是指允许启动子的一种或多种相关基因连续转录的不受调控的启动子。

[0170] 在优选的实施方案中,编码本发明融合蛋白或本发明分泌信号的核苷酸序列由诱导型启动子驱动。

[0171] 本领域已知很多诱导型启动子。在Gatz,Curr.Op.Biotech.,7:168(1996)(还参见Gatz,Ann.Rev.Plant.Physiol.Plant Mol.Biol.,48:89(1997))的综述中描述了许多诱导型启动子。实例包括四环素抑制子系统(tetracycline repressor system)、Lac抑制子系统、铜诱导系统、水杨酸诱导系统(如PR1系统)、糖皮质激素诱导型(Aoyama等,1997)、醇诱导系统(如AOX启动子)和蜕皮激素诱导系统(ecdysome-inducible system)。还包括苯磺酰胺诱导型(美国专利No.5,364,780)和醇诱导型(WO97/06269和WO 97/06268)诱导系统以及谷胱甘肽S-转移酶启动子。

[0172] 在Mattanovich等,Methods Mol.Biol.(2012)824:329-58中描述了与酵母宿主细胞一起使用的适宜的启动子序列,包括如磷酸丙糖异构酶(TPI)、磷酸甘油酸激酶(PGK)、甘油醛-3-磷酸脱氢酶(GAPDH或GAP)的糖酵解酶及其变体、乳糖酶(LAC)和半乳糖苷酶(GAL)、巴斯德酵母(P.pastoris)葡萄糖-6-磷酸异构酶启动子(PPGI)、3-磷酸甘油酸激酶启动子(PPGK)、甘油醛磷酸脱氢酶启动子(PGAP)、翻译延伸因子启动子(PTEF)和巴斯德酵母烯醇酶1启动子(PENO1)、丙糖磷酸异构酶(PTPI)、核糖体亚基蛋白(PRPS2、PRPS7、PRPS31、PRPL1)、醇氧化酶启动子(PAOX)或其具有修饰特性的变体、甲醛脱氢酶启动子(PFLD)、异柠檬酸裂合酶启动子(PICL)、α-酮异己酸脱羧酶启动子(PTHI)、热休克蛋白家族成员的启动子(PSSA1、PHSP90、PKAR2)、6-磷酸葡糖酸脱氢酶(PGND1)、磷酸甘油酸变位酶(PGPM1)、转酮

酶(PTKL1)、磷脂酰肌醇合成酶(PPIS1)、铁-02-氧化还原酶(ferro-02-oxidoreductase)(PFET3)、高亲和力铁通透酶(PFTR1)、抑制性碱性磷酸酶(PPHO8)、N-肉豆蔻酰基转移酶(PNMT1)、信息素应答转录因子(PMCM1)、泛素(PUBI4)、单链DNA核酸内切酶(PRAD2)、线粒体内膜的主要ADF/ATP载体的启动子(PPET9)(WO2008/128701)和甲酸脱氢酶(FMD)启动子。GAP启动子、AOX启动子或源自GAP或AOX启动子的启动子是特别优选的。AOX启动子可被甲醇诱导和被葡萄糖抑制。WO 2006/089329、EP 1851312B1和EP 2199389B1中描述了用于甲醇诱导的以及无甲醇产生的源自AOX启动子的启动子。可使用碳源可调控启动子,例如,如WO2013050551(例如,pG1-pG8,pG1的片段,命名为pG1a-pG1f)和WO2017021541(例如,pG1-D1240或pG1-D1427)中所述的去阻抑(de-repressible)启动子。其它实例是组成型启动子,比如MDH3、POR1、PDC1、FBA1-1或GPM1(Prielhofer等,2017,BMC Sys Biol.11:123),或如WO2014139608中所公开的(如pCS1)。

[0173] 适宜的启动子的进一步实例包括酿酒酵母烯醇化酶(ENO-1)、酿酒酵母半乳糖激酶(GAL1)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH1、ADH2/GAP)、酿酒酵母丙糖磷酸异构酶(TPI)、酿酒酵母金属硫蛋白(CUP1)、和酿酒酵母3-磷酸甘油酸激酶(PGK)、和麦芽糖酶基因启动子(MAL)。

[0174] Romanos等,1992,Yeast 8:423-488描述了酵母宿主细胞的其它有用启动子。

[0175] 本发明的宿主细胞

[0176] 如本文所用,“宿主细胞”是指一种能够表达蛋白和分泌蛋白的细胞。这种宿主细胞用于本发明的方法。针对使宿主细胞表达一种多肽这一目的,使编码该融合蛋白的核苷酸序列存在于该细胞或将其引入该细胞。本发明提供的宿主细胞可以是真核生物。如本领域技术人员所理解的,原核细胞无核膜,而真核细胞具有核膜。真核细胞的实例包括但不限于:脊椎动物细胞、哺乳动物细胞、人细胞、动物细胞、无脊椎动物细胞、植物细胞、线虫细胞、昆虫细胞、干细胞、真菌细胞或酵母细胞。优选地,所述宿主细胞为酵母细胞。

[0177] 因此,本发明涉及一种包含本发明核酸分子的宿主细胞。另外,本发明涉及一种包含本发明的表达盒的宿主细胞。本发明还涉及一种包含本发明载体的宿主细胞。

[0178] 酵母细胞的实例包括但不限于:酵母属(如酿酒酵母、克鲁维酵母(Saccharomyces kluyveri)、葡萄汁酵母(Saccharomyces uvarum)、奇异酵母(Saccharomyces paradoxus)、真贝酵母(Saccharomyces eubayanus)、库德里阿兹威酵母(Saccharomyces kudriavzevii))、Komagataella属(Komagataella pastoris、Komagataella pseudopastoris或Komagataella phaffii)、克鲁维酵母属(如,乳酸克鲁维酵母、马克思克鲁维酵母(Kluyveromyces marxianus))、假丝酵母属(如产朊假丝酵母(Candina utilis)、可可假丝酵母(Candida cacaoi))、地霉属(如,发酵地霉(Geotrichum fermentans))、以及多形汉森酵母和解脂耶氏酵母。因此,本发明的真核宿主细胞或用在本发明的方法和用途中的真核宿主细胞可以是真菌或酵母宿主细胞,优选为选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母的酵母宿主细胞,或诸如里氏木霉或黑曲霉的真菌宿主细胞。

[0179] 毕赤酵母属是特别感兴趣的。毕赤酵母属包含许多种,包括巴斯德毕赤酵母(Pichia pastoris)种、甲醇毕赤酵母(Pichia methanolica)种、克鲁维毕赤酵母(Pichia

kluyveri)种和安格斯毕赤酵母(Pichia angusta)种。最优选的是巴斯德毕赤酵母种。

[0180] 已将前种巴斯德毕赤酵母分类并重新命名为Komagataella pastoris和Komagataella phaffii。因此巴斯德毕赤酵母与Komagataella pastoris和Komagataella phaffii是同义词,优选地与Komagataella phaffii是同义词。

[0181] 用于本发明的巴斯德酵母菌株的实例是X33及其亚型GS115、KM71、KM71H;CBS7435(mut+)及其亚型CBS7435 mut^s、CBS7435 mut^sΔArg、CBS7435 mut^sΔHis、CBS7435 mut^sΔArg、ΔHis、CBS7435 mut^s PDI$^+$、CBS 704(=NRRL Y-1603=DSMZ 70382)、CBS2612(=NRRL Y-7556)、CBS 9173-9189和DSMZ 70877及其突变体。优选地,所述宿主细胞为巴斯德毕赤酵母CBS7435 mut^S或其亚型,更优选地为巴斯德毕赤酵母CBS7435mut^s。

[0182] 根据另一个优选的实施方案,宿主细胞为巴斯德毕赤酵母、多形汉森酵母、里氏木霉、酿酒酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母和Komagataella,和粟酒裂殖酵母。所述宿主细胞也可以是来自玉米黑粉菌(Ustilago maydis)的宿主细胞。

[0183] 如本文所用,“重组(的)”是指通过人为干预的遗传材料的改变。通常,重组是指通过分子生物(重组DNA技术)方法(包括克隆和重组)对病毒、细胞、质粒或载体中DNA或RNA的操作。借助参考重组细胞、多肽或核酸与天然存在的对应物(“野生型”)的区别来描述重组细胞、多肽或核酸。“重组细胞”或“重组宿主细胞”是指这样一种细胞或宿主细胞,所述细胞或宿主细胞已经被遗传改变从而包含了对所述细胞而言为非天然的核酸序列。

[0184] 本文所用术语“制备”是指目的蛋白表达的过程。“用于制备目的蛋白的宿主细胞”是指可引入编码目的蛋白的核酸序列的宿主细胞。本发明中的重组宿主细胞并不一定包含编码目的蛋白的核酸序列。本领域技术人员理解,可以例如在试剂盒中提供所述宿主细胞,用于将所需核苷酸序列插入该宿主细胞中。

[0185] 术语“多肽”与“蛋白”互换使用。术语“多肽”是指一种含有两个或更多个氨基酸,典型地至少3个,优选地至少20个,更优选地至少30个,诸如至少50个氨基酸的蛋白或肽。因此,多肽包含氨基酸序列,和因此,有时包含氨基酸序列的多肽在本文中称为“包含多肽序列的多肽”。因此,本文中术语“多肽序列”与术语“氨基酸序列”互换使用。

[0186] 术语“氨基酸”是指天然存在的和合成的氨基酸,以及以与天然存在的氨基酸类似的方式发挥作用的氨基酸类似物(analog)和氨基酸模拟物(mimetic)。天然存在的氨基酸是由遗传密码编码的那些,以及之后修饰的那些氨基酸,如羟脯氨酸、γ-羧基谷氨酸和O-磷酸丝氨酸。氨基酸类似物是指与天然存在的氨基酸具有相同的基础化学结构的化合物,即,将碳连接至氢、羰基、氨基和R基,如高丝氨酸、正亮氨酸、蛋氨酸亚砜、蛋氨酸的甲基锍。这些类似物具有修饰的R基(如正亮氨酸)或修饰的肽骨架,但是保留与天然存在的氨基酸相同的基础化学结构。表达包含合成氨基酸或氨基酸类似物的目的蛋白的系统是本领域技术人员已知的,包括但不限于,使用扩展的遗传密码。扩展的遗传密码的关键先决条件是:待编码的非标准氨基酸,待采用的未使用的密码子,识别该密码子的tRNA,以及仅识别该tRNA和所述非标准氨基酸的tRNA合成酶。氨基酸模拟物是指具有与氨基酸的一般化学结构不同的结构,但是以与天然存在的氨基酸类似的方式发挥功能的化学化合物。

[0187] 为了进一步增加融合蛋白或目的蛋白在本文所述的分泌信号切割后的分泌(参见实施例4-9),可另外地将(重组)宿主细胞工程化以过表达信号识别颗粒(SRP)的一种或多

种组分。

[0188] 本文所用的信号识别颗粒(SRP)涉及丰富的、胞质的、普遍保守的核糖核蛋白(蛋白-RNA复合物),其识别并靶向特定蛋白至真核生物内质网和原核生物质膜。在真核生物中,当SRP从核糖体中出现时,其结合新合成的肽的信号肽序列-例如源自SWP1或KRE1蛋白的信号肽序列。这种结合导致称为"延伸停滞"的蛋白合成减慢,这是一种促进蛋白翻译和蛋白易位过程的偶联的SRP的保守功能。然后SRP将该整个复合物(核糖体-新生链复合物)靶向至ER(内质网)膜中的蛋白传导通道(也称为易位子)。这通过SRP与其同源SRP受体的相互作用和锚定(docking)而发生,所述同源SRP受体位于紧邻易位子的位置。

[0189] 在真核生物中,SRP和其受体之间存在三个在鸟苷三磷酸(GTP)结合和水解中起作用的结构域。这些结构域位于SRP受体(SRα和SRβ)和SRP蛋白SRP54中的两个相关亚基中。锚定后,新生肽链被插入到易位子通道中,在此进入ER。当SRP从核糖体释放时,蛋白合成重新开始。SRP-SRP受体复合物通过GTP水解解离,SRP介导的蛋白易位的循环继续。因此,在一些情况下,也可以将SRP依赖性易位看作共翻译易位模式。在一个具体的实施方案中,本发明的融合蛋白向ER的易位是共翻译的。

[0190] 一旦进入ER内,可通过信号肽酶将信号肽序列从核心蛋白切除。因此,信号肽序列不是成熟蛋白(比如在自本发明的融合蛋白中切割分泌信号后分泌的目的蛋白)的一部分。

[0191] SRP可包含SRP68、SRP72、SRP9-21、SRP54、SRP14、Sec65和7SL RNA。因此,"SRP的一种或多种组分"可涉及选自SRP68、SRP72、SRP9-21、SRP54、SRP14、Sec65和7SL RNA中的至少一种。在优选的实施方案中,在宿主细胞中过表达SRP的所有组分,即SRP68、SRP72、SRP9-21、SRP54、SRP14、Sec65和7SL RNA。有利地,在宿主细胞中过表达的SRP的一种或多种组分源自与宿主细胞相同的物种。然而,过表达异源的SRP的一个或多个组分也涵盖在本发明中。此外,还涵盖过表达SRP的一种或多种组分的功能同系物,包括但不限于(优选Komagataella phaffii的)SRP68、SRP72、SRP9-21、SRP54、SRP14、Sec65和7SL RNA中一种或多种的功能同系物;或(优选Komagataella phaffii的)SRP68、SRP72、SRP9-21、SRP54、SRP14、Sec65和7SL RNA的功能同系物。

[0192] 源自Komagataella phaffii的SRP组分中的每一种的示例性序列列于下表3中。

[0193] 表3:可在宿主细胞中过表达的SRP组分中的每一种的示例性序列。

[0194]

SRP 亚单元	序列(5'-3')	蛋白序列

[0195]

SRP68 （在 pADH2 下表达，RPL2att 用作终止子）（Prielhofer 等，2017）	ATGGAATCGCCCTTGCAATCTA	MESPLQSTYGERAERYLDSA
	CATACGGAGAAAGAGCCGAAA	DAFHKQRHRLNRRLHKLRK
	GGTATTTAGATAGTGCTGATGC	SLDIHVTDTKNYREKEQISKI
	TTTTCATAAACAAAGACACAGA	DLESYNRDKRYGDIILFTAE
	TTGAATCGAAGGCTGCACAAGT	RDHMYSDEVKEIMKVHHSK
	TACGTAAGAGCTTGGATATTCA	SREKFIVSRLKRSLDHGRKL
	TGTTACTGATACTAAGAACTAT	LILVGDEPDEMRKLEVFVYV
	AGAGAGAAAGAGCAGATTTCC	ALIQGKLSIANKNWTNAQY
	AAAATTGATCTAGAGTCGTACA	ALSVARCGLQFLDKYGTET
	ACAGGGATAAGCGATATGGTG	QTDLYNGIIDTHIDQMLKFVI
	ACATTATACTGTTCACTGCAGA	YQATKNNSPILDTECRHQIR
	GAGGGATCACATGTATAGTGAT	TDTLGYLDQARQIIESKDPEF
	GAGGTCAAGGAGATCATGAAG	LNVGVVETQLIWWDYDISIH
	GTCCATCATAGTAAATCGAGAG	SEEVAKLISDANEKLQLIED
	AAAAGTTTATTGTTTCTAGATT	GNVSSYDPALLTLQEALDA
	GAAGAGATCACTGGACCACGG	HQLLMARNVDNFADDDQN
	TAGAAAATTACTGATCCTAGTT	NHVLLSYIRYLLLITTLRRDI
	GGAGACGAGCCTGATGAGATG	TLIDQVRNRSVVNSSLAVAL
	AGAAAATTGGAAGTATTTGTTT	ERAKDVGRIFDNIVKKVNEL
	ATGTTGCATTGATTCAGGGTAA	KDVPGVYNKQEEWNSLQAL
	ACTTTCCATTGCAAACAAGAAT	DAYFQASKIQHLASTHLLFN

[0196]

	TGGACCAATGCTCAGTATGCTC	RSKESLALLIKAKSLVKGHTI
	TCAGTGTGGCGAGATGTGGGCT	AGEYPTNFPTNKDLSSILEQI
	CCAGTTTTTGGACAAATATGGT	NQDILKAYVLAKYKQESSL
	ACTGAAACACAAACTGACCTCT	GGVSEYDFIADNRNKVPSNP
	ATAATGGCATAATTGACACTCA	SLHKIASVSYKNVKPVNVKP
	CATAGATCAAATGTTGAAATTT	VLFDIAFNYVSQPNQIFEEPI
	GTGATCTACCAAGCTACTAAAA	ESSNKQERQADSESPSPEKK
	ATAACAGTCCTATTTTGGATAC	KKGLFGLFR** (SEQ ID
	AGAGTGCAGACATCAAATTAG	NO: 5)
	GACGGACACCCTAGGGTATTTG	
	GATCAGGCAAGGCAAATAATA	
	GAATCAAAAGATCCCGAGTTTC	
	TGAATGTTGGAGTTGTTGAAAC	
	TCAGTTGATTTGGTGGGACTAC	
	GATATCTCTATTCATTCAGAGG	
	AGGTAGCAAAGCTGATTTCAGA	
	TGCGAACGAAAAGCTGCAACTT	
	ATCGAGGATGGAAACGTCTCCT	
	CATATGATCCGGCTCTACTAAC	
	TCTTCAAGAAGCGCTGGATGCT	
	CATCAGTTGTTGATGGCCAGAA	

[0197]

	ATGTTGACAACTTCGCAGACGA	
	CGATCAAAACAATCATGTTTTA	
	CTGTCGTACATCAGATATTTGT	
	TACTTATCACCACTTTGAGAAG	
	GGACATTACTTTGATAGACCAA	
	GTTAGAAACAGATCTGTGGTTA	
	ATTCTTCCCTAGCTGTGGCTCT	
	GGAACGTGCTAAAGACGTTGGT	
	AGAATTTTCGACAATATCGTCA	
	AGAAAGTCAATGAGTTGAAAG	
	ACGTTCCAGGTGTTTACAACAA	
	GCAAGAGGAGTGGAATTCGTT	
	GCAGGCATTGGATGCTTATTTC	
	CAAGCATCCAAGATCCAACATT	
	TGGCATCTACCCACCTTTTATTC	
	AACAGATCCAAGGAATCATTG	
	GCGTTATTAATAAAGGCAAAGT	
	CCTTGGTAAAGGGGCACACTAT	
	CGCCGGAGAATATCCCACTAAT	
	TTCCCTACGAATAAAGATTTGA	
	GTTCGATCTTAGAACAAATTAA	

[0198]

	TCAAGACATCCTTAAGGCTTAT GTTTTGGCCAAGTATAAGCAAG AGTCCTCTTTAGGTGGTGTATC GGAGTATGATTTCATTGCTGAC AATCGCAACAAGGTTCCGTCGA ATCCCAGTCTGCACAAGATTGC CTCTGTATCCTACAAGAATGTC AAACCTGTCAATGTTAAGCCTG TATTGTTTGACATAGCTTTCAA CTACGTGAGTCAACCAAACCAG ATATTTGAGGAACCTATCGAGA GTTCAAACAAGCAAGAGAGAC AAGCGGATTCTGAATCTCCGTC ACCAGAGAAGAAAAAAAAGGG ATTGTTTGGATTGTTCCGCTAG TAG (SEQ ID NO: 12)	
SRP72（在 pPOR1 下表达，RPP1btt 用作终止子）（Prielhofer 等，2017）	ATGTCGTCTCTTTCAGAGCTGG TATCGGAATTGGCAATCCATTC TGAGAAGAGGCAGTACAAAGA AGCATATGAGAAAGCAAAGCG CATTATAGATTTGGGCCACCCT	MSSLSELVSELAIHSEKRQY KEAYEKAKRIIDLGHPLDLD TLKLGLVASINLDQYHNAG RLISKSKDHIVYDGMKEFLL LIGYVYYKNGDSKNFETLLK

[0199]

	CTTGACCTTGACACATTGAAGC	DSAFQGRAFEHLKAQYYYK
	TAGGTTTGGTGGCTTCGATCAA	IGENEKALKIYRELSKNPLDE
	CCTGGACCAATATCACAACGCA	VVDLSVNERAVISQLLELDG
	GGTCGCCTCATATCAAAAAGTA	VVEQPVSRPIDDSYDCKFND
	AGGATCATATCGTATATGATGG	ALYQVKIGDYESALDLLEEA
	AATGAAGGAATTCTTGCTATTA	KAICEENTKDLPLDTREAEIV
	ATTGGATATGTGTACTACAAGA	PILLQIAYVKQLKGKKEESL
	ACGGAGACTCGAAGAATTTTGA	TALRSLSKPKDSLLDLIYRN
	AACTCTACTAAAGGATTCAGCT	NLLSLRIDEYGRNDTNFHIL
	TTCCAAGGAAGAGCATTTGAAC	YRELGFPNSIDINKDKLTVSQ
	ACCTCAAAGCCCAATACTATTA	RVALTRNESLLALELGKIPS
	CAAGATCGGGGAGAACGAAAA	QSDLKLFYDATSEFLDLNTK
	GGCACTCAAGATTTACCGAGAG	LEASMIYNYFMRRPGQQEV
	CTATCTAAGAACCCATTGGATG	PNALLTAQLAINVGNINNAR
	AAGTTGTAGACTTGAGTGTCAA	TVLETVVSNDEKNLLEPSIV
	TGAAAGAGCTGTCATTAGCCAG	VSLYLIYDKLQSGRLQVELL
	CTTTTGGAATTGGATGGTGTCG	KKVADLLLESEISSTQQRKFF
	TTGAACAGCCTGTATCTCGACC	KDIAFKTLNHDAVLANRLFE
	AATAGACGACTCATACGATTGC	KLHSIYPNDELVSTYLNSSSN
	AAATTCAACGATGCCCTTTACC	ASNNNTTTTNFSELDDLVLG
	AGGTAAAGATTGGTGACTATGA	IDTDKLISEGFDTFESSKRPT

[0200]

	ATCTGCATTGGATCTCTTGGAA	TIISSTNKKRRTRLKPKHEAK
	GAAGCCAAAGCCATATGTGAA	EKYKRLDEERWLPLKDRSY
	GAAAATACAAAAGATCTGCCTT	YRPKKGKKIRNTTQGTVTSN
	TGGACACACGAGAAGCAGAGA	TSEISGLKKTLPKKSSKKKG
	TTGTTCCTATTCTGTTACAAATT	RK** (SEQ ID NO: 6)
	GCTTACGTCAAACAACTTAAGG	
	GTAAAAAGGAAGAATCCTTGA	
	CTGCGCTGAGAAGTCTTTCTAA	
	ACCAAAGGACTCTCTTTTAGAT	
	CTTATTTACAGAAATAACTTAC	
	TGTCATTAAGGATTGATGAATA	
	CGGAAGAAATGATACCAACTTT	
	CATATTCTTTATCGTGAGTTAG	
	GATTCCCTAATTCGATAGACAT	
	TAATAAAGACAAGTTGACAGT	
	GTCCCAAAGGGTTGCGTTGACC	
	AGAAATGAATCATTATTGGCAC	
	TCGAGCTTGGAAAGATCCCATC	
	TCAAAGTGATCTCAAGCTCTTT	
	TATGATGCTACTTCGGAATTTT	
	TAGATTTGAACACCAAGCTAGA	

[0201]

	AGCCTCAATGATTTATAATTAT	
	TTCATGAGACGCCCTGGCCAGC	
	AAGAGGTTCCAAATGCTCTTTT	
	AACTGCACAGCTGGCTATCAAC	
	GTTGGTAACATCAATAATGCAA	
	GAACTGTTCTTGAAACTGTGGT	
	GTCGAACGACGAAAAGAATTT	
	ACTAGAACCATCTATTGTGGTA	
	TCTTTGTATTTGATTTACGATAA	
	GCTTCAAAGCGGAAGATTGCA	
	AGTTGAACTCCTGAAGAAAGTA	
	GCGGATCTTTTGCTAGAATCAG	
	AGATTTCCAGCACTCAACAACG	
	TAAGTTTTTCAAAGATATTGCC	
	TTCAAAACCCTCAATCACGATG	
	CAGTTTTGGCCAATCGACTATT	
	TGAGAAACTGCATAGTATTTAC	
	CCTAATGATGAGTTGGTATCCA	
	CGTACTTGAACTCTTCATCTAA	
	TGCATCCAACAATAACACTACC	
	ACGACCAACTTCTCCGAATTGG	

[0202]

	ACGACTTGGTCCTAGGCATAGA CACGGATAAGCTTATTAGTGAA GGATTTGATACTTTTGAGTCCA GCAAAAGACCGACCACGATTA TCAGCTCAACTAACAAGAAAC GTCGTACTAGATTGAAGCCCAA ACATGAAGCCAAGGAAAAGTA TAAGCGTCTGGATGAGGAAAG ATGGTTACCTTTAAAGGACCGC AGTTATTACAGACCCAAAAAG GGAAAGAAAATTAGAAATACC ACTCAGGGTACTGTCACTAGTA ATACTAGTGAAATAAGTGGCTT GAAAAAGACTCTGCCAAAGAA AAGTTCCAAAAAGAAAGGAAG AAAATGATAG (SEQ ID NO: 13)	
SRP9-21（在 pPDC1 下表达，RPS25att 用作终止子）（Prielhofer 等，2017）	ATGCCTCCTGTGAAATCTCTGG ACATCTTTTTCAACCGCACAGA GAAGCTCTTAGAAGCCAACCCC ACAACGACAAAAGTTTCCATCA AATTGGGCGTAAATTTCAATGA	MPPVKSLDIFFNRTEKLLEA NPTTTKVSIKLGVNFNDHEN PQSKHNVITVRVSDPVSGSN FKFKVTNKTDMLKIFSFLGP HGIELPISGQQSQIKSNDQTQ

[0203]

	TCACGAGAATCCTCAAAGCAA GCACAACGTCATAACGGTGAG AGTATCTGATCCAGTGAGCGGG TCCAATTTCAAATTCAAAGTGA CCAATAAAACTGATATGCTGAA AATATTCAGTTTCTTAGGTCCT CATGGCATTGAGTTACCAATTT CTGGCCAGCAAAGCCAGATAA AGAGTAATGATCAGACTCAGA GTGACAATACTGAAGTGCCTAC CACATTTCATAGGGGAGCCACC AGTATTTTGGCTAATAAGGCAT TTGAGAAGAAACCACTGATTAT TAAGGATTCAAGTACCGCAAA GAAAGGTGGTAAAGGTGGTAA GAAGAAGGGTAAGAAATTTTA A (SEQ ID NO: 14)	SDNTEVPTTFHRGATSILAN KAFEKKPLIIKDSSTAKKGG KGGKKKGKKF* (SEQ ID NO: 7)
SEC65（在 pRPP1b 下表达，RPS17btt 用作终止子）（Prielhofer 等，2017）	ATGCCATTACTAGAGGAAATAA GTGATGCAGAGGACATAGACA ACTTGGAGATGGATTTAGCCGA GTTTGATCCTACTTTAAGGACT	MPLLEEISDAEDIDNLEMDL AEFDPTLRTPIAEQRPAPQV VRSQDAESGQTPLVPNQDQI SQYIEQFKEGGTINKDQVIRP

[0204]

	CCGATAGCTGAGCAAAGACCA	DEMMEKEMAELKSFQILYP
	GCTCCTCAGGTTGTCAGATCAC	CYFDKNRSVKEGRRCQKEY
	AAGATGCCGAAAGTGGACAGA	GVENPLAKTILDACRYLDIP
	CTCCTTTGGTTCCTAACCAGGA	CILEPEKTHPQDFGNPGRVR
	TCAAATAAGTCAGTATATTGAA	VAIKESGKYLDEQYKTKRK
	CAATTCAAAGAAGGTGGCACC	LIQLVGQFLVEHPTTLQKVQ
	ATAAACAAGGATCAAGTGATT	ELPGPPELQQGGYIPERVPR
	AGACCCGACGAAATGATGGAA	VKGLKMNEIVPLHSPFTIKH
	AAAGAAATGGCAGAGTTGAAA	PSTKSVYEREPEPAPPAAVP
	AGCTTCCAAATTTTGTACCCAT	KAPKQKKIMVRR** (SEQ ID
	GTTACTTTGATAAAAATAGAAG	NO: 8)
	TGTTAAAGAAGGAAGAAGATG	
	CCAAAAGGAGTATGGTGTGGA	
	GAACCCCCTGGCAAAGACAAT	
	ATTAGATGCTTGCAGGTACTTG	
	GATATACCTTGCATCCTGGAGC	
	CTGAAAAGACTCATCCTCAAGA	
	TTTTGGTAATCCAGGAAGAGTG	
	AGAGTGGCTATCAAGGAGAGT	
	GGGAAGTATCTCGATGAACAAT	
	ATAAGACCAAAAGGAAACTAA	

[0205]

	TACAGTTGGTAGGACAATTTCT GGTTGAACATCCAACAACGTTA CAGAAAGTTCAAGAATTGCCCG GTCCACCTGAGTTGCAACAGGG CGGGTACATTCCAGAACGTGTA CCCCGAGTGAAAGGGTTAAAG ATGAACGAAATTGTTCCTTTGC ATTCGCCATTCACTATTAAGCA TCCAAGTACTAAATCTGTTTAT GAAAGGGAACCTGAGCCCGCA CCACCCGCCGCCGTGCCCAAAG CTCCGAAACAGAAGAAAATAA TGGTGAGAAGATAATAG (SEQ ID NO: 15)	
SRP54（在 pFBA1-1 下表达，RPS2tt 用作终止子）（Prielhofer 等，2017）	ATGGTATTGGCAGATCTTGGAA GGCGTATCAATAACGCCGTTGG AAATGTCACCAAGTCCAATGTT GTTGACGCTGACGTCATCAGCA ACATGTTAAAGGAGATTTGTAA CGCCCTATTGGAGTCCGATGTG AACATTAAACTAGTTGCCCAAT	MVLADLGRRINNAVGNVTK SNVVDADVISNMLKEICNAL LESDVNIKLVAQLREKIRKQI DAEDKPGINKKKLIQKVVFD ELVKLVDCNEAELFKPKKK QTNVIMMVGLQGAGKTTTC TKLAVYYQRRGFKVGMVC

[0206]

TGAGAGAGAAAATACGAAAAC	GDTFRAGAFDQLKQNATKA
AGATCGACGCAGAGGATAAAC	KIPYYGSYTETDPVKVTFDG
CAGGAATTAATAAGAAGAAGC	VEEFRKEKFEIIIVDTSGRHR
TGATCCAGAAGGTCGTTTTTGA	QEEDLFEEMVQIGKAIKPNQ
TGAGCTGGTGAAACTTGTTGAT	TIMVLDASIGQSAESQSKAF
TGCAACGAAGCTGAGCTGTTCA	KESSDFGAIIITKMDSNSKGG
AGCCAAAGAAAAAACAGACGA	GALSAIAATNTPVAFIATGE
ATGTGATCATGATGGTCGGTTT	HIQNFEKFSGRGFISKLLGIG
ACAAGGTGCTGGTAAGACAAC	DIEGLMEHVQSMNLDQGDT
AACCTGTACTAAACTGGCAGTG	IKNFKEGKFTLQDFQTQLNN
TATTACCAGAGAAGAGGATTCA	IMKMGPLSKLAQMLPGGMG
AAGTGGGAATGGTCTGTGGTGA	QLMGQVGEEEASKRLKRMI
CACTTTCCGAGCTGGTGCGTTT	YIMDSMTKQELSSDGRLFID
GACCAGCTGAAACAAAACGCT	QPSRMVRVARGSGTSVTEV
ACCAAGGCTAAGATTCCCTACT	ELVLLQQKMMARMALQSK
ATGGTTCATATACAGAAACTGA	NMMSGAGGPAGMASKMNP
CCCTGTGAAAGTGACCTTTGAT	ANMRRAMQQMQSNPGMM
GGTGTGGAAGAATTCAGGAAG	DNMMNMFGGAGGAGGAG
GAAAAGTTTGAAATAATAATTG	MPDMQEMMKQMSSGQMK
TGGATACTTCTGGTAGACACAG	MPSQQEMMSMMKQFGMG*
GCAGGAGGAAGATTTATTCGA	* (SEQ ID NO: 9)

[0207]

	AGAGATGGTACAAATTGGAAA	
	AGCTATCAAGCCTAATCAAACA	
	ATCATGGTACTGGATGCTTCCA	
	TAGGTCAATCTGCCGAATCTCA	
	ATCTAAAGCATTTAAGGAATCA	
	TCCGATTTTGGTGCCATTATCA	
	TAACTAAAATGGATTCCAATTC	
	CAAGGGAGGAGGTGCCCTTTCA	
	GCTATAGCTGCCACCAACACTC	
	CAGTAGCGTTTATTGCCACCGG	
	AGAGCACATTCAGAATTTCGAA	
	AAGTTTTCAGGAAGAGGATTTA	
	TCTCAAAACTTTTAGGAATTGG	
	TGATATAGAGGGTCTTATGGAA	
	CATGTTCAGTCGATGAACTTGG	
	ATCAAGGTGATACTATCAAGAA	
	TTTCAAGGAAGGAAAGTTTACT	
	TTACAGGATTTTCAAACGCAAT	
	TGAACAACATCATGAAGATGG	
	GGCCACTGTCCAAACTCGCTCA	
	AATGTTGCCTGGTGGAATGGGA	

[0208]

	CAATTGATGGGACAGGTTGGTG	
	AAGAGGAGGCTTCAAAGAGAT	
	TGAAGCGAATGATTTATATAAT	
	GGATTCAATGACGAAGCAAGA	
	GTTGTCAAGTGACGGTAGATTG	
	TTTATTGATCAGCCTTCAAGGA	
	TGGTAAGAGTTGCTAGAGGCTC	
	TGGTACCTCTGTAACTGAGGTG	
	GAGCTTGTTCTTTTACAGCAAA	
	AGATGATGGCTCGTATGGCATT	
	ACAATCTAAGAATATGATGAGT	
	GGGGCCGGCGGTCCAGCAGGG	
	ATGGCTTCCAAAATGAATCCAG	
	CTAATATGAGAAGAGCTATGCA	
	ACAAATGCAATCAAACCCAGG	
	AATGATGGATAACATGATGAAT	
	ATGTTTGGTGGAGCTGGAGGAG	
	CTGGAGGAGCTGGAATGCCGG	
	ATATGCAAGAAATGATGAAGC	
	AAATGTCCAGTGGCCAAATGA	
	AAATGCCCAGTCAACAGGAAA	

[0209]

	TGATGAGCATGATGAAACAGTT TGGTATGGGCTAATAG (SEQ ID NO: 16)	
SRP14（在 pGPM1 下表达，RPS3tt 用作终止子）（Prielhofer 等，2017）	ATGTCCACAACTACTAAGAAAA ACAAGAACAGGATCTTGATAG AGAATCACAAACAGTTCCTGGA AGAAGTTTCCAAAACAGCCACT TTATCAGTTTGGAACTCGAAAT TTTCAATCAAACGACTGTCTCT AGAAGCAGATCCCGTGGAAGG GACGCCTGAAGGAATCAGAGA TATCCCACAAGGAGTAGAGAC AAACTCTATAATAGGAAACAG CGTAGAGAATGATTCAAAGTCA CACCCCATTTTATTCAGATATA CAGCTAGACACGCAAAGGAGA AAATACCAGAGGTGCGAATTTC AACCACCGTTGATTCAGAGCAG CTAAGCACCTTCTGGAGAGATT ATGTGGACATATTGAAGGGAA GCTCCCAATTGAAACTGCAGTC	MSTTTKKNKNRILIENHKQF LEEVSKTATLSVWNSKFSIK RLSLEADPVEGTPEGIRDIPQ GVETNSIIGNSVENDSKSHPI LFRYTARHAKEKIPEVRISTT VDSEQLSTFWRDYVDILKGS SQLKLQSETKKVSSKKSKAK KKRGKGAW* (SEQ ID NO: 10)

[0210]

	AGAAACCAAGAAAGTCAGTAG TAAGAAGAGCAAGGCGAAGAA GAAGAGAGGAAAGGGTGCATG GTAA (SEQ ID NO: 17)	
非编码 RNA(在 pTEF2 下表达，IDPtt 用作终止子)（Prielhofer 等，2017）。用锤头状和 HDV 核酶过表达非编码 RNA(下划线)，以移除RNA聚合酶II转录后的 mRNA 特征	ATGCAGCCTCTGATGAGTCCGT GAGGACGAAACGAGTAAGCTC GTCAGGCTGTTATGGCGCATCC GGGGGAGGTAGTTACTTGACCT TGATTCCTAATAGCTTACAACT GAGGTGTCTCGTTCGATCCTGG CGGTCCGCAATATTTTCCATAC GAGTAATCTGTGGGGGAAGGC GAGCAATAAGACGTGCCACCG CCCAAGGGGAGCAATCCAGCA GGGAACACGTCCCGCAAGGAG GCGGGTGAGATAGCATCTCGTT GGTAATGGGCTGTTGGTGAACA AAGTTTGACTATGTGAACCGGC TATTTACATTTTTGCTTTTT (SEQ ID NO: 11)	

[0211] 因此,本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:5-11中的一种或多种或其功能同系物。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:5-11中的一种或多种。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,也可以被工程化以过表达SEQ ID NO:5-11或其功能同系物。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,也可以被工程化以过表达SEQ ID

NO:5-11。可选地,本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:11-17中的一种或多种或其功能同系物。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:11-17中的一种或多种。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:11-17或其功能同系物。本发明的宿主细胞,优选巴斯德毕赤酵母宿主细胞,可以被工程化以过表达SEQ ID NO:11-17。

[0212] 如本文所用,“工程化的”宿主细胞为已经用遗传工程(即,人为干预)操作了的宿主细胞。当宿主细胞被“工程化为过表达”给定蛋白时,操作该宿主细胞使得与非工程化的宿主细胞相比,该宿主细胞具有增加的表达蛋白或功能同系物的能力,从而与操作之前的相同条件下的宿主细胞相比,给定蛋白,例如本发明的融合蛋白,或另外信号识别颗粒(SRP)的一种或多种组分的表达增加。

[0213] 当将“工程化之前”用于本发明的宿主细胞的上下文中时,是指这种宿主细胞没有被工程化使得编码蛋白(比如本发明融合蛋白和/或SRP的一种或多种组分)或其功能同系物的多核苷酸过表达。

[0214] 本文所用术语“重组”是指用编码目的蛋白的异源多核苷酸装备本发明的宿主细胞,即,本发明的宿主细胞被工程化为包含编码目的蛋白的异源多核苷酸。这可以通过如转化或转染或用于将多核苷酸引入宿主细胞的任何其它本领域已知的适宜技术来实现。

[0215] 如将在下面详细描述的,可通过本领域技术人员已知的任何方法实现过表达。一般而言,可通过提高基因的转录和/或翻译,如通过提高所述基因的拷贝数或者改变或修饰与基因表达相关的调控序列或位点来实现过表达。例如,可以通过将一个或多个拷贝的编码蛋白(如:与调控序列(如启动子)可操作地连接的目的蛋白和/或信号识别颗粒(SRP)的一种或多种组分)的多核苷酸通过转化引入宿主细胞或甚至宿主细胞的基因组相应(resp.)染色体中来实现过表达。例如,为了达到高表达水平,可将所述基因可操作地连接至组成型启动子和/或强的广谱表达型(ubiquitous)启动子。这种启动子可以是内源性启动子或重组启动子。可选地,可以将调控序列移除从而使表达变为组成型的和/或当负调控序列移除时使表达增加。人们可以用增加基因表达的或导致基因组成型表达的宿主细胞的基因组相应(resp.)染色体中的异源启动子替换启动子。例如,启动子可以是强的组成型启动子和/或强的广谱表达型(ubiquitous)启动子或强的诱导型启动子,以便达到比天然启动子更高的表达水平。在表达对宿主细胞而言外来的POI和/或信号识别颗粒(SRP)的一种或多种组分的情况下,术语“表达”和“过表达”和“(过)表达”可互换使用。例如,与工程化之前和在相同条件下培养的宿主细胞相比,宿主细胞(过)表达大于1%、2%、3%、4%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、200%或大于300%的目的蛋白和/或信号识别颗粒(SRP)的一种或多种组分。因此,如本文所用,“过表达”可涉及与天然表达蛋白但未工程化以过表达所述蛋白的宿主细胞(对照)相比,宿主细胞的蛋白(如信号识别颗粒(SRP)的一种或多种组分或目的蛋白)表达增加,并且蛋白表达可增加大于1%、2%、3%、4%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、200%或300%中的任一个。此外,本文所用的“过表达”可以涉及与与未被工程化为表达该蛋白质时不表达该蛋白质的宿主细胞相比,蛋白(如目的蛋白)的表达。额外地使用诱导型启动子可使宿主细胞培养过程中的表达增加。此外,还可通过以下方式实现过表达,例如修饰特定基因的染

色体位置;改变与特定基因相邻的核酸序列,所述特定基因如核糖体结合位点或转录终止子,修饰与所述基因转录和/或与所述基因产物的翻译相关的蛋白(如,调控蛋白、抑制子、增强子、转录激活因子等等);或任何其它本领域常规的使特定基因表达失调的常规技术(包括但不限于,反义核苷酸分子的应用,例如以阻断阻遏蛋白的表达,或者缺失或突变通常抑制需要过表达的基因的表达的转录因子的基因。延长mRNA的寿命也可提高表达水平。例如,某些终止子区可用于延长mRNA的半衰期(Yamanishi等,Biosci.Biotechnol.Biochem.(2011)75:2234和US 2013/0244243)。若包括多个拷贝的基因,这些基因既可以位于可变拷贝数的质粒中,也可以是在染色体中整合的并扩增的。若所述宿主细胞不包含基因产物,则可以将基因产物引入宿主细胞用于表达。此时,"过表达"是指采用本领域技术人员已知的任何方法表达所述基因产物。可以通过本领域技术人员已知的方法,如通过SDS Page、SDS Page/蛋白质印记、ELISA或具有后续质谱(蛋白质组学)的肽图谱来测量与如上所述的对照相比的过表达,例如信号识别颗粒(SRP)的一种或多种组分或其功能同源物的过表达。例如,可以测量细胞内信号识别颗粒(SRP)的一种或多种组分的含量,并将其与对照中的含量进行比较。因此,必须分解细胞,然后测量细胞裂解物中的含量。目的蛋白的过表达相应的表达可以通过例如测量相应宿主细胞和对照的上清液中目的蛋白的含量,然后比较这些含量来测量。上清液中目的蛋白的含量可以通过本领域技术人员已知的方法来测量,例如通过SDS Page、SDS Page/蛋白质印记、ELISA、RP(反相)HPLC、离子交换HPLC等。

[0216] 本领域技术人员将在Martin等(Bio/Technology 5,137-146(1987))、Guerrero等(Gene 138,35-41(1994))、Tsuchiya和Morinaga(Bio/Technology6,428-430(1988))、Eikmanns等(Gene 102,93-98(1991))、EP 0 472 869、US 4,601,893、Schwarzer和Pöhler(Bio/Technology 9,84-87(1991))、Reinscheid等(Applied and Environmental Microbiology 60,126-132(1994))、LaBarre等(Journal of Bacteriology 175,1001-1007(1993))、WO 96/15246、Malumbres等(Gene 134,15-24(1993))、JP-A-10-229891、Jensen和Hammer(Biotechnology and Bioengineering 58,191-195(1998))和Makrides(Microbiological Reviews 60,512-538(1996))中,特别是在公知的遗传和分子生物教科书中发现相关说明。

[0217] 优选地将编码本发明融合蛋白的核酸或载体和/或编码SRP的一种或多种组分的多核苷酸整合到基因组中,更优选地整合到宿主细胞的染色体或非染色体基因组位点中。术语"基因组"通常是指以DNA(或对于某些病毒物种以RNA)编码的生物体的全部遗传信息。基因组可存在于染色体中、质粒或载体中或两者兼有。优选地,将编码SRP的一种或多种组分的载体或核酸整合到宿主细胞的基因组中。

[0218] 可通过将相关基因各自连接到一个载体中而将编码本发明融合蛋白的多核苷酸和/或编码SRP的一种或多种组分的多核苷酸重组在宿主细胞中。可以构建携带基因的单个载体,或两个分开的载体,一个携带本发明的融合蛋白,另一个携带SRP的一种或多种组分。通过使用诸如本发明的核酸的核酸或本文所述的载体转化宿主细胞来将这些基因整合到宿主细胞基因组中。在一些实施方案中,将编码本发明融合蛋白的基因整合到基因组中,和将编码SRP的一种或多种组分的基因整合到质粒或载体中。在一些实施方案中,将编码SRP的一种或多种组分的基因整合到基因组中,和将编码融合蛋白的基因整合到质粒或载体

中。在一些实施方案中,将编码本发明融合蛋白的基因和编码SRP的一种或多种组分的多核苷酸整合到基因组中。在一些实施方案中,将编码本发明融合蛋白的基因和/或编码SRP的一种或多种组分的多核苷酸整合到质粒或载体中。如果使用多个编码融合蛋白的基因,则可以将一些编码融合蛋白的基因整合到基因组中,而将将另一些整合到相同或不同的质粒或载体中。如果使用多个编码SRP的一种或多种组分的基因,则可以将一些编码SRP的一种或多种组分的基因可整合到基因组中,而另一些整合到相同或不同的质粒或载体中。

[0219] 还可通过用导致宿主细胞内源SRP蛋白更高转录的调控元件交换宿主细胞内源SRP蛋白的调控元件来实现SRP的一种或多种组分的过表达。

[0220] 本发明的方法

[0221] 本发明还涉及一种在真核宿主细胞中制备目的蛋白的方法,所述方法包括

[0222] (i)用本发明的核酸分子或用本发明的表达盒或载体遗传工程化重组宿主细胞,并任选地遗传工程化所述重组宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分;

[0223] (ii)在表达所述核酸分子和任选地所述SRP的一种或多种组分,和在切割分泌信号后分泌目的蛋白的条件下培养遗传工程化的宿主细胞,

[0224] (iii)任选地从细胞培养物中分离目的蛋白,

[0225] (iv)任选地纯化所述目的蛋白,

[0226] (v)任选地修饰所述目的蛋白,和

[0227] (vi)任选地配制所述目的蛋白。

[0228] 也可将制备方法的步骤(i)理解为遗传工程化重组宿主细胞以过表达本发明的融合蛋白和任选地SRP的一种或多种组分(也参见“本发明的宿主细胞”部分中的相关公开)。当宿主细胞被“工程化以过表达”给定蛋白时,操作宿主细胞以使该宿主细胞具有表达,优选过表达本发明的核酸分子和任选SRP的一种或多种组分的能力,从而与操作前的相同条件下的宿主细胞相比,给定蛋白(如本发明的融合蛋白和任选地SRP的一种或多种组分)的表达增加。在一个实施方案中,“工程化以过表达”意指对宿主细胞进行遗传改变以过表达蛋白或增加蛋白的表达,即(有意地)将细胞遗传工程化为过表达此类蛋白。工程化以过表达可包括用导致更高转录的启动子交换内源SRP的一种或多种组分的启动子。工程化以过表达可包括用编码本文所述融合蛋白和/或SRP的一种或多种组分的核酸分子、表达盒或载体遗传工程化宿主细胞。

[0229] 本发明还涉及一种通过在表达本发明的核酸分子和在切割分泌信号后分泌目的蛋白的条件下培养本发明的重组真核宿主细胞,并在切割分泌信号后自宿主细胞培养物中分离目的蛋白,并任选地纯化和任选地修饰和任选地配制所述目的蛋白来生产目的蛋白的方法。

[0230] 当将“工程化之前”或“操作之前”用于本发明的宿主细胞的上下文中时,是指这种宿主细胞没有用编码本发明的融合蛋白的核酸工程化。因此,所述术语还指宿主细胞不过表达编码本发明融合蛋白的核酸和/或未被工程化以过表达SRP的一种或多种组分。

[0231] 本领域技术人员熟知用于操作如编码本发明的融合蛋白和/或SRP的一种或多种组分的多核苷酸序列、启动子、增强子、前导区(leader)等的程序,如J.Sambrook等,Molecular Cloning:ALaboratory Manual(3rd edition),Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory Press,New York(2001)描述的。

[0232] 可通过多种手段,如通过同源重组或通过采用特异性靶向在整合位点的序列的混合重组酶(hybrid recombinase),将外源或靶多核苷酸(如本发明的核酸分子)插入染色体中。典型地,所述外源或靶多核苷酸存在于载体(“插入载体”)中。典型地,在用于同源重组之前,这些载体是环状的和线性的。作为一种选择,所述外源或靶多核苷酸可以是稍后被重组至宿主细胞中的通过融合PCR连接的DNA片段或合成构建的DNA片段。除了同源臂之外,载体还含有适于选择或筛选的标记、复制起点和其它元件。还可以使用实现随机或非靶向整合的异源重组。异源重组是指DNA分子与显著不同的序列之间的重组。本领域已知重组的方法,例如在Boer等,Appl Microbiol Biotechnol(2007)77:513-523中描述的。对于酵母细胞的遗传操作还可参考Primrose和Twyman的Principles of Gene Manipulation and Genomics(第七版,Blackwell Publishing 2006)。

[0233] 本发明的核酸分子和/或SRP的一种或多种组分也可存在于载体(如表达载体)上。这样的载体是本领域已知的。在表达载体中,启动子置于编码异源蛋白的基因的上游,并调控所述基因的表达。由于多克隆载体的多克隆位点,其是特别有用的。对于表达,通常将启动子置于多克隆位点的上游。既可通过首先制备含有编码本发明的目的蛋白或SRP的一种或多种组分的完整DNA序列的DNA构建体,并随后将这一构建体插入适宜的表达载体中,也可以通过顺序地插入含有个体元件(如,DNA结合结构域,活化结构域)遗传信息的DNA片段,然后连接,来构建用于编码本发明融合蛋白或SRP的一种或多种组分的核酸分子的整合的载体。作为片段的限制和连接的替代,可把基于附着(att)位点的重组方法和重组酶用于将DNA序列插入载体中。这些方法由例如Landy(1989)Ann.Rev.Biochem.58:913-949描述,并是本领域技术人员已知的。

[0234] 可通过将包含靶多核苷酸序列(比如本发明的核酸分子)的载体或质粒引入细胞来获得根据本发明的宿主细胞。本领域熟知用于转染或转化真核细胞或者转化原核细胞的技术。这些技术可以包括脂囊泡介导的摄取、热休克介导的摄取、电穿孔、磷酸钙介导的转染(磷酸钙/DNA共沉淀)、病毒感染(特别是使用经修饰的病毒诸如例如经修饰的腺病毒)、显微注射和电穿孔。对于原核转化,技术可以包括热休克介导的摄取、与完整细胞的细菌原生质体融合、显微注射和电穿孔。用于植物转化的技术包括土壤杆菌(Agrobacterium)介导的转移(诸如通过根瘤土壤杆菌(A.tumefaciens)实现的)、被快速推进的钨或金微弹、电穿孔、显微注射和聚乙二醇介导的摄取。DNA可以是单链的或双链的、线性的或环状的、松弛的或超螺旋的DNA。对于用于转染哺乳动物细胞的多种技术,参见例如Keown等(1990)Processes in Enzymology185:527-537。

[0235] 短语“在表达本发明的核酸分子和任选地过表达“SRP的一种或多种组分”的条件下培养(遗传工程化)宿主细胞”指在适于或足以获得所需目的蛋白的生产或过表达SRP的一种或多种组分的条件(如温度、压力、pH、诱导、生长速率、培养基、持续时间、补料等)下维持和/或生长真核宿主细胞。

[0236] 优选地,通过用本发明的核酸分子或用本发明的表达盒工程化宿主细胞,和任选地遗传工程化宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分而获得的本发明的宿主细胞可首先在有效地生长至大细胞数目而没有表达重组蛋白的负担的条件下培养。当制备用于融合蛋白表达的细胞时,选择并优化合适的培养条件以生产融合蛋白。

[0237] 例如,对融合蛋白和SRP的一种或多种组分使用不同的启动子和/或拷贝和/或整

合位点,可相对于SRP的一种或多种组分的表达,在时间点和诱导强度方面控制融合蛋白的表达。例如,在诱导融合蛋白之前,SRP的一种或多种组分可首先表达。这具有的优点是在于融合蛋白翻译的开始时,SRP的一种或多种组分已经存在。或者,可同时诱导融合蛋白和SRP的一种或多种组分。

[0238] 可以使用诱导型启动子,其在施加诱导刺激时立即活化,以指导在其控制下的基因的转录。在具有诱导刺激的生长条件下,细胞通常比在正常条件下生长得更慢,但是由于培养物在前一阶段已经生长至高细胞数目,因此培养系统作为整体产生大量的重组蛋白。优选地,诱导刺激是适当试剂(例如用于AOX-启动子的甲醇)的加入或适当营养物(例如用于MET3-启动子的甲硫氨酸)的消耗。此外,乙醇、甲胺、镉或铜以及加热或渗透压增强剂的加入可诱导表达,这取决于与融合蛋白和SRP的一种或多种组分可操作连接的启动子。

[0239] 优选将根据本发明的宿主在生物反应器中在优化的生长条件下培养以获得至少1g/L、优选至少10g/L细胞干重、更优选至少50g/L细胞干重的细胞密度。不仅在实验室规模中,而且在中试或工业规模中,实现这样的生物分子产量是有益的。

[0240] 通过使用标准试验,如ELISA、活性检验、HPLC、表面等离子体共振(Biacore)、蛋白质印迹、毛细管电泳(Caliper)或SDS-Page测量细胞培养物上清液或细胞匀浆后细胞的细胞匀浆中目的蛋白的滴度来测试本发明的宿主细胞的表达/分泌能力或产量。

[0241] 优选将细胞培养在含有合适碳源的矿物质培养基中,由此进一步显著地简化分离过程。作为示例,所述矿物质培养基含有可利用碳源(例如葡萄糖、甘油、乙醇或甲醇)、含有常量元素(钾、镁、钙、铵、氯化物、硫酸盐、磷酸盐)和痕量元素(铜、碘、锰、钼酸盐、钴、锌和铁盐以及硼酸)的盐。

[0242] 在酵母细胞的情况下,可以用上述一种或多种表达载体转化所述细胞,并在被修饰为适合诱导启动子、选择转化子或扩增编码所需序列的基因的常规营养培养基中培养。本领域已知许多适于酵母菌生长的矿物质培养基。根据需要,可向这些培养基的任一种中补充盐(如氯化钠、钙、镁和磷酸盐)、缓冲液(如HEPES、柠檬酸和磷酸盐缓冲液)、核苷(如腺嘌呤和胸腺嘧啶)、痕量元素、维生素和葡萄糖或等效能源。还可包括以适当浓度的任何其它必须的补充物,这是本领域技术人员已知的。所述培养条件(如温度、pH等)是之前选择用于宿主细胞表达的条件,这是普通技术人员已知的。针对其它类型宿主细胞的细胞培养条件也是已知的并可由技术人员容易地确定。对各种微生物的培养基的描述例如涵盖在美国微生物学会(American Society for Bacteriology)的“一般细菌学方法手册”的指南(华盛顿D.C,USA,1981)中。

[0243] 可以通过常规的培养方法,例如试管培养、振荡培养(例如旋转振荡培养,摇瓶培养等)、掺气旋动培养(aeration spinner culture)或发酵,优选地连续或间歇地在液体培养基中培养(例如维持和/或生长)细胞。在一些实施方案中,在摇瓶或深孔板中培养细胞。在另一些实施方案中,在生物反应器中(例如在生物反应器培养过程中)培养细胞。培养过程包括、但不限于,分批培养、进料分批培养和连续方法培养。术语“分批过程”和“分批培养”是指一个封闭的系统,该系统中的培养基组成、营养物、补充添加剂等在培养开始时已设定,并在培养期间不经受改变;但是,可尝试控制如pH和氧浓度的因素来阻止过度培养基酸化和/或细胞死亡。术语“进料分批过程”和“进料分批培养”是指具有以下例外的一种分批培养:随着培养的进行,添加(例如,以增量或连续方式添加)一种或多种底物或补充物。

术语“连续过程”和“连续培养”是指一种系统,在该系统中将限定的培养基连续地添加至生物反应器,并同时将等量的所采用的或“条件化”的培养质移除,例如用于回收所需产物。已经开发了多种这样的过程并且这些过程是本领域熟知的。

[0244] 在一些实施方案中,将细胞培养约12至24小时,在另一些实施方案中,将细胞培养约24至36小时,约36至48小时,约48至72小时,约72至96小时,约96至120小时,约120至144小时,或培养持续时间大于144小时。在又一些实施方案中,培养持续一段足以达到POI的所需生产产量的时间。

[0245] 本发明的方法(如制备目的蛋白的方法或增加目的蛋白的分泌的方法)可进一步包括分离表达的POI的步骤。若所述POI是从细胞中分泌出来的,则该POI可采用现有技术从培养基中分离并纯化。在分泌过程中,信号肽被切除。POI从细胞中分泌出来通常是优选的,因为可以从培养上清液中回收产物,而不是从当细胞被破坏以释放胞内蛋白时产生的复杂蛋白混合物中回收。蛋白酶抑制剂在抑制纯化过程中的蛋白质降解可能是有益的。可采用本领域已知的方法将组合物浓缩、过滤、透析等。发酵/培养后的细胞培养物可以使用分离器或管式离心机离心以从培养物上清液中分离细胞。然后通过使用切向流动过滤将上清液过滤或浓缩。

[0246] 对于获得POI的分离和纯化方法,可以使用利用溶解度差异的方法,比如盐析、溶剂沉淀、热沉淀;利用分子量差异的方法,比如尺寸排阻色谱、超滤和凝胶电泳;利用电荷差异的方法,比如离子交换层析;利用特异亲和性的方法,比如亲和层析;利用疏水性差异的方法,比如疏水性互作用色谱和反相高效液相层析;以及利用等电点差异的方法,比如等电点聚集;和利用某些氨基酸的方法,比如IMAC(固定化金属离子亲和色谱)。如果POI表达为无活性和可溶的包涵体,则该可溶的包涵体需要重折叠。

[0247] 分离和纯化的POI可以通过常规方法,比如蛋白质印迹或对POI活性的特异性检验来鉴定。纯化POI的结构可以通过氨基酸分析、氨基-末端分析、一级结构分析等(例如质谱法、RP-HPLC、离子交换-HPLC、ELISA等)来确定。优选可以大量和高纯度水平获得POI,这样即可满足作为药物组合物中的活性成分或者作为饲料或食品添加剂的必要要求。

[0248] 如本文所用的术语“分离的”是指在自然界中不存在的形式或环境中的物质。分离的物质的非限制性实例包括(1)任何非天然存在的物质,(2)任何物质,包括但不限于任何酶、变体、核酸、蛋白、肽或辅助因子,其至少部分地从与其天然相关的一种或多种或全部天然存在的成分中移除;(3)相对于在自然界发现的物质,通过人工修饰的任何物质,例如从mRNA制备的cDNA;或(4)通过增加该物质相对于与其天然相关的其它组分的量而修饰的任何物质(例如,在宿主细胞中重组生产;多拷贝的编码该物质的基因;和使用比与编码该物质的基因天然相关的启动子更强的启动子)。

[0249] 在本上下文中,术语“修饰所述目的蛋白”意指该POI是经化学或酶促修饰的。本领域已知许多修饰蛋白的方法。可以将蛋白质与碳水化合物或脂质偶联。可以将POIPEG化(将POI与聚乙二醇化学偶联)或HES化(将POI与羟乙基淀粉化学偶联)以延长半衰期。还可将POI与其它部分,比如针对如人血清白蛋白的亲和结构域偶联,以延长半衰期。还可以通过蛋白酶或在水解条件下处理POI以进行剪切(cleavage),以从预序列形成活性成分或切下标签(比如用于纯化的亲和标签)。还可以将POI与其他部分,例如毒素、放射性部分或任何其它部分偶联。可以在形成二聚体、三聚体等的条件下进一步处理POI。

[0250] 此外,术语“配制所述目的蛋白”是指使POI处于这样的条件下,在该条件下POI可以长时间存储。本领域已知许多不同方法均可用于稳定蛋白质。通过在纯化和/或修饰后交换其中存在POI的缓冲液,可以将POI置于它在其中更稳定的条件下。可以使用本领域已知的不同缓冲物质和添加剂,如蔗糖、温和的洗涤剂、稳定剂等。也可以通过冻干从而使POI稳定。对于一些POI,可以通过形成POI与脂质或脂蛋白的复合物(比如人工合成多聚物(polyplex)等)来进行配制。也可以将一些蛋白质与其它蛋白质共同配制。

[0251] 本发明的分泌信号的使用可增加目的蛋白的分泌。因此,本发明还涉及一种增加目的蛋白自真核宿主细胞分泌的方法,所述方法包括在所述真核宿主细胞中表达本发明的核酸分子,并任选地遗传工程化所述宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分,从而与表达本文定义的融合蛋白,但该融合蛋白在本发明的核酸分子中包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非所定义的分泌信号,的所述宿主细胞相比增加所述目的蛋白的分泌。

[0252] 如本文所用的“分泌”涉及将形成本发明融合蛋白的一部分的目的蛋白转移出(重组)宿主细胞。因此,仅目的蛋白而不是分泌信号被分泌。信号肽序列在内质网中被切除,MFα前导序列在高尔基体中被切除。因此,如果分泌增加,则仅目的蛋白的分泌增加。细胞培养物上清液中目的蛋白的滴度可以使用标准试验来测定,例如如ELISA、活性检验、HPLC、表面等离子体共振(Biacore)、蛋白质印迹、毛细管电泳(Caliper)或SDS-Page。

[0253] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括工程化所述宿主细胞以引入表达构建体来表达本发明的核酸分子,和任选地遗传工程化真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分。

[0254] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括在表达本发明的核酸分子,和任选地遗传工程化宿主细胞以过表达SRP的一种或多种组分,和在切割分泌信号后分泌目的蛋白的条件下培养所述宿主细胞。

[0255] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括从细胞培养物中分离目的蛋白。

[0256] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括纯化目的蛋白。

[0257] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括修饰目的蛋白。

[0258] 增加目的蛋白自真核宿主细胞中的分泌的方法可进一步包括配制目的蛋白。

[0259] 在增加目的蛋白自真核宿主细胞中的分泌的方法中,可以进一步包括,将本发明的核酸分子整合到所述宿主细胞的染色体中或包含在未整合到所述宿主细胞的染色体中的载体或质粒中。

[0260] 本发明的用途

[0261] 本领域技术人员将容易地理解本文所述的分泌信号可用于增加重组目的蛋白的分泌。因此,本发明还涉及如本文定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途。如本文已经描述的,分泌信号从N-端到C-端可包含源自KRE1蛋白的信号肽序列,任选地随后是α-交配因子(MFα)前导序列。因此,本发明还涉及如本文定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途,其中所述分泌信号从N-端至C-端包含源自KRE1蛋白的信号肽序列,随后是α-交配因子(MFα)前导序列。本发明还涉及如本文定义的分泌信号用于增加目的蛋白从真核宿主细胞分泌的用途,其中所述分泌信号包含源自

KRE1蛋白的信号肽序列。如本文已经描述的,分泌信号从N-端到C-端可包含源自SWP1蛋白的信号肽序列,任选地随后是α-交配因子(MFα)前导序列。因此,本发明还涉及如本文定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途,其中所述分泌信号从N-端到C-端包含源自SWP1蛋白的信号肽序列,随后是α-交配因子(MFα)前导序列。本发明还涉及如本文定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途,其中所述分泌信号包含源自SWP1蛋白的信号肽序列。

[0262] 与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非本文定义的分泌信号的本发明融合蛋白的真核宿主细胞相比,所述分泌信号可增加目的蛋白自真核宿主细胞中的分泌。

[0263] 如本领域技术人员将容易认识到的,本发明的重组宿主细胞可用于生产各种目的蛋白,因为这些目的蛋白被有效地分泌到上清液中,从而避免宿主细胞的裂解。因此,本发明进一步涉及本发明的重组宿主细胞用于制备目的蛋白的用途。

[0264] 项目

[0265] 1.一种编码融合蛋白的核酸分子,所述融合蛋白从N-端到C-端包含

[0266] (a)分泌信号,所述分泌信号包含

[0267] (I)(i)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;以及

[0268] (ii)α-交配因子(MFα)前导序列;

[0269] 或

[0270] (II)源自KRE1蛋白的信号肽序列或源自SWP1蛋白的信号肽序列;

[0271] 以及

[0272] (b)目的蛋白。

[0273] 2.根据项目1所述的核酸分子,其中与表达项目1的核酸分子,但该核酸分子包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非项目1中定义的分泌信号,的所述真核宿主细胞相比,所述分泌信号增加所述目的蛋白自真核宿主细胞中的分泌。

[0274] 3.根据项目1或2所述的核酸分子,其中所述源自KRE1蛋白的信号肽序列包含SEQ ID NO:1或其功能同系物。

[0275] 4.根据项目1或2所述的核酸分子,其中所述源自SWP1蛋白的信号肽序列包含SEQ ID NO:2或52或其功能同系物。

[0276] 5.根据前述项目中任一项所述的核酸分子,其中所述MFα前导序列包含SEQ ID NO:3、53或其功能同系物,和/或其中所述MFα前导序列在对应于SEQ ID NO:53的位置23的位置处包含Ser和/或在对应于SEQ ID NO:53的位置64的位置处包含Glu。

[0277] 6.根据项目1至5中任一项所述的核酸分子,其中所述目的蛋白选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

[0278] 7.根据项目1至6中任一项所定义的分泌信号。

[0279] 8.一种表达盒或载体,其包含项目1至6中任一项所述的核酸分子和与其可操作地连接的启动子。

[0280] 9.一种重组真核宿主细胞,其包含项目1至6中任一项所述的核酸分子,或项目8所述的表达盒或载体,优选地

[0281] (a)其中所述宿主细胞是真菌或酵母宿主细胞,优选为酵母宿主细胞,其选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母,或真菌宿主细胞,比如里氏木霉或黑曲霉;和/或

[0282] (b)其中所述宿主细胞被工程化以过表达信号识别颗粒(SRP)的一种或多种组分。

[0283] 10.一种在真核宿主细胞中制备目的蛋白的方法,所述方法包括

[0284] (i)用项目1至6中任一项所述的核酸分子或用项目8所述的表达盒或载体遗传工程化所述真核宿主细胞,并任选地遗传工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分;

[0285] (ii)在表达所述核酸分子和任选地过表达所述SRP的一种或多种组分,和在切割分泌信号后分泌目的蛋白的条件下培养遗传工程化的宿主细胞,

[0286] (iii)任选地从细胞培养物中分离目的蛋白,

[0287] (iv)任选地纯化所述目的蛋白,

[0288] (v)任选地修饰所述目的蛋白,和

[0289] (vi)任选地配制所述目的蛋白。

[0290] 11.一种增加目的蛋白自真核宿主细胞中的分泌的方法,所述方法包括在所述真核宿主细胞中表达项目1至6中任一项所定义的核酸分子,和任选地工程化所述真核宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分,从而与表达除了包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非项目1至6中任一项所定义的所述分泌信号的项目1至6的核酸分子的所述宿主细胞相比增加所述目的蛋白的分泌。

[0291] 12.根据项目10或11所述的方法,

[0292] (a)其中,所述方法包括

[0293] (i)工程化所述宿主细胞以引入表达项目1至6中任一项所述的核酸分子的表达构建体,和任选地遗传工程化所述宿主细胞以过表达信号识别颗粒(SRP)的一种或多种组分,

[0294] (ii)在表达所述核酸分子和任选地过表达所述SRP的一种或多种组分和在切割分泌信号后分泌目的蛋白的条件下培养所述宿主细胞,

[0295] (iii)任选地从细胞培养物中分离目的蛋白,

[0296] (iv)任选地纯化所述目的蛋白,

[0297] (v)任选地修饰所述目的蛋白,和

[0298] (vi)任选地配制所述目的蛋白;和/或

[0299] (b)其中将所述核酸分子整合在所述宿主细胞的染色体中或包含在未整合到所述宿主细胞的染色体中的表达盒、载体或质粒中;和/或

[0300] (c)其中所述真核宿主细胞是真菌或酵母宿主细胞,优选为酵母宿主细胞,其选自Komagataella phaffii(巴斯德毕赤酵母)、多形汉森酵母、酿酒酵母、奇异酵母、真贝酵母、库德里阿兹威酵母、克鲁维酵母、葡萄汁酵母、乳酸克鲁维酵母、解脂耶氏酵母、甲醇毕赤酵

母、博伊丁假丝酵母、Komagataella spp.和粟酒裂殖酵母,或真菌宿主细胞,比如里氏木霉或黑曲霉;和/或

[0301] (d)其中所述目的蛋白选自抗体,比如嵌合抗体、人源化抗体或人抗体,或双特异性抗体,或诸如Fab或$F(ab)_2$的抗原结合抗体片段,诸如scFv的单链抗体,诸如骆驼的VHH片段的单结构域抗体或重链抗体或结构域抗体(dAbs),人工抗原结合分子,比如DARPIN、ibody、affibody、humabody或基于脂质运载蛋白家族多肽的突变蛋白,酶,比如加工酶,细胞因子,生长因子,激素,蛋白抗生素,诸如毒素融合蛋白的融合蛋白,结构蛋白,调节蛋白和疫苗抗原,优选其中所述目的蛋白为治疗性蛋白、食品添加剂或饲料添加剂。

[0302] 13.如项目1至7中任一项所定义的分泌信号用于增加目的蛋白自真核宿主细胞中的分泌的用途,优选地,其中与表达包含野生型酿酒酵母α-交配因子分泌信号(比如SEQ ID NO:4)而非项目1中所定义的分泌信号的项目1中定义的融合蛋白的所述真核宿主细胞相比,所述分泌信号增加所述目的蛋白自真核宿主细胞中的分泌。

[0303] 14.项目9所述的重组真核宿主细胞用于制备目的蛋白的用途。

[0304] 15.一种通过在表达项目1至6中任一项所述的核酸分子和在切割分泌信号后分泌目的蛋白的条件下培养项目9所述的重组真核宿主细胞,并自宿主细胞培养物中分离目的蛋白,并任选地纯化和任选地修饰和任选地配制所述目的蛋白来生产目的蛋白的方法。

[0305] ****

[0306] 注意,如本文所用,除非上下文另有明确指示,本文所用单数形式“一(a)”、“一种(an)”和“所述/该(the)”包括所提及的复数因此,例如,提及“一种试剂”包括一种或多种这样的不同试剂,并且提及“该方法”包括本领域普通技术人员已知的可修饰或置换本发明的方法的等效步骤和方法。

[0307] 除非另有说明,一系列元素之前的术语“至少”将被理解为是指该系列中的每一元素。本领域技术人员将承认或采用常规实验能够确定,本文所述的本发明的具体实施方案的许多等价方案。这种等价方案意图被本发明所涵盖。

[0308] 术语“和/或”无论在本文何处使用均包括“和”、“或”和“由所述术语连接的元素的全部或任意其它组合”的含义。

[0309] 术语“约”是指正或负20%,优选正或负10%,更优选正或负5%,最优选正或负1%。

[0310] 术语“小于”或“大于”不包括具体的数字。

[0311] 例如,小于20意味着小于所指示的数字。类似地,大于或多于意味着大于或多于指示的数字,例如大于80%意味着大于或多于指示的数字80%。

[0312] 本说明书及随后的权利要求全文中,除非上下文有其它要求,词语“包括/包含(comprise)”以及变形词语诸如“包括/包含(comprises)”和“包括/包含(comprising)”将被理解为意指包括陈述的整数或步骤、或整数或步骤的集合但并不排除任何其它的整数或步骤、或整数或步骤的集合。当本文使用时,术语“包括/包含(comprising)”能够以术语“含有(containing)”、“包含(including)”、“具有(having)”替代。当本文使用时,“由…组成”排除了未规定的任何元素、步骤或成分。

[0313] 术语“包括”是指“包括但不限于”。“包括”和“包括但不限于”可互换使用。

[0314] 应当理解,本发明并不限于本文所述的具体的方法学、方案、材料、试剂和物质等,

它们可以改变。本文所使用的术语只用于描述特定实施方案的目的,其并不意在限定本发明的范围,本发明的范围仅由权利要求限定。

[0315] 本说明书全文所引用的所有的出版物(包括所有的专利、专利申请、科学出版物、指导书等),无论在前还是在后,均通过引用以其全文并入本文。本文无任何内容能被解释为承认由于在先发明,本发明不早于这样的公开内容。就通过引入并入的材料与本说明书矛盾或不一致来说,本说明书将取代任何这样的材料。

[0316] 在此引用的所有文献和专利文献的内容通过引用以其全文并入本文。

[0317] 实施例

[0318] 从以下提出的仅用于说明目的实施例将清楚地了解本发明及其优点。这些实施例不以任何方式限制本发明的范围。已经尝试保证所使用的数字(例如量、温度、浓度等)的准确性,但应该允许一些实验误差和偏差。除非另外说明,份是重量份,分子量是平均分子量,温度是摄氏度,并且压力是大气压或接近大气压。

[0319] 以下实施例将证明,与已知的分泌前导序列(其包括常用的酿酒酵母α-交配因子前原前导序列)相比,与酿酒酵母α-交配因子的前导序列组合的新鉴定的信号肽增加了所分泌的重组蛋白的效价(以mg/L表示的每体积的产物)和产量(以mg/g生物质表示的每生物质的产物,生物质以湿细胞重量来度量)。作为实例,新型信号肽的使用导致不同的重组蛋白和抗体衍生物在巴斯德毕赤酵母中的产量增加,所述不同的重组蛋白和抗体衍生物包括单链可变区片段、单结构域抗体、抗原结合片段和易于检测的蛋白(scR、VHH、SDZ-Fab、m-cherry)。在振荡培养(在24深孔板中进行)和进料-分批培养中显示出了积极作用。

[0320] 实施例1:新型信号肽的选择

[0321] 包括巴斯德毕赤酵母(同物异名:Komagataella spp.)的酵母中最常用的分泌信号是酿酒酵母α-交配因子(MFα)前原前导序列,即MFα分泌信号。一些重组蛋白与MFα分泌信号一起分泌效率低,因此仅达到低生产滴度。因此,迫切需要新型信号肽和分泌信号以增加分泌效率。

[0322] 分泌中的最初和关键步骤是重组蛋白向内质网(ER)中的易位。该过程由融合到重组蛋白的N-端可切割分泌信号指导。对于重组蛋白分泌,至少需要N-端可切割的信号肽序列。这些可以采用最广泛使用的程序SignalP(Nielsen,2017)从氨基酸序列中预测。基于巴斯德毕赤酵母序列数据,SignalP 4.1预测了241个具有可切割信号肽(SP)的蛋白(Valli等,2016)。然而,未对共翻译SP或翻译后SP进行区分,即预测的SP可以在共翻译时或翻译后起作用,并且氨基酸序列本身不能给出预测的信号肽序列是否能够分泌重组融合蛋白(即与通过该信号肽序列分泌的天然分泌蛋白不同的蛋白)的信息。迄今为止,在主要的物理化学性质,比如信号肽序列的疏水性、推定的结合基序或其它序列模式方面,没有发现足够明显的特征来描述用于重组蛋白的共翻译或翻译后易位的有效信号肽(Janda等,2010;Pechmann等,2014;Massahi等,(2016),.J Theor Biol.,408:22-33)。

[0323] 我们使用计算机方法来选择新型有效的SP。首先,使用对信号肽中氨基酸数目标准化的亲水指数计算疏水平均值(Kyte和Doolittle,1982)。接下来,我们在SP中搜索8个氨基酸的最疏水的区段,并赋予最大值为1的得分,其代表整个集合中的最疏水的区段。此外,我们评估了天然相关蛋白的密码子43至48的相对适应度(adaptiveness)的平均值(Sharp和Li,1987),并赋予最大值为1的得分,代表在该区段中具有“最慢”密码子的平均值。换

yanzhi,当在该区段中仅存在最佳密码子时,赋予0分。将两者产生的得分相加,选择代表不同等级的候选者(表4:)。在评分的候选者中,SP4(SWP1、PP7435_Chr1-0255的前18个氨基酸)和SP14(KRE1、PP7435_Chr3-0933的前18个氨基酸)分别代表具有最高和最低总分的信号肽。

[0324] 表4:选择的来自Komagataella phaffii的信号肽序列候选者和来自酿酒酵母的MFα(MFα1)的物理化学性质的排序

[0325]

编号	基因名称	CBS7435 ORF	氨基酸数目	密码子偏好得分	亲水性区段(8aa)得分	综合得分
SP4*	SWP1	PP7435_Chr1-0255	18	***0,01***	**0,21**	***0,22***
SP10	SLP1	PP7435_Chr1-1234	16	**0,06**	**0,57**	**0,64**
SP22		PP7435_Chr3-1213	20	**0,09**	**0,63**	**0,72**
ScMFα	MFα1	--	20	**0,41**	**0,38**	**0,79**
SP8	GDT1	PP7435_Chr2-0066	20	**0,35**	**0,60**	**0,95**
SP9	NCR1	PP7435_Chr1-0478	18	**0,25**	**0,80**	**1,05**
SP7*	OST3	PP7435_Chr4-0360	18	***0,85***	**0,30**	***1,15***
SP2	GET1	PP7435_Chr4-0582	16	**0,81**	**0,54**	***1,35***
SP15		PP7435_Chr2-0965	24	**0,63**	**0,75**	**1,38**
SP5*	WBP1	PP7435_Chr2-0876	20	**0,97**	**0,44**	***1,41***
SP11	MSB2	PP7435_Chr1-0283	20	**0,95**	**0,51**	**1,45**
SP6*	OST1	PP7435_Chr3-0451	16	**0,74**	**0,90**	***1,64***
SP16		PP7435_Chr4-0694	18	**0,84**	**0,81**	**1,65**
SP13	PEP1	PP7435_Chr2-0657	20	**0,98**	**0,72**	**1,70**
SP18		PP7435_Chr4-0664	25	**0,88**	**0,85**	**1,73**
SP3	FET3	PP7435_Chr2-0482	20	**0,87**	**0,93**	***1,80***
SP14	KRE1	PP7435_Chr3-0933	18	**0,92**	**1,00**	**1,92**

[0326] 实施例2:使用新型信号肽序列分泌重组分泌蛋白的巴斯德毕赤酵母菌株的构建和选择

[0327] 携带目的基因的质粒的构建

[0328] 将巴斯德毕赤酵母CBS7435 mut^{S}变体(由Sturmberger等,2016对其基因组测序)用作宿主菌株。通过Geneart或DNA2.0(现在为ATUM)对编码POI的基因(如SDZ-Fab-LC,SDZ-Fab-HC,scR,VHH和mCherry)进行密码子优化,并以合成DNA获得。His6标签C-端地融合至scR和VHH基因以进行检测,而FLAG标签在mCherry的C-端添加。在说明书的表2示出了这些蛋白的序列。

[0329] 为了构建具有新型信号肽序列的表达载体(参见表4和表7),通过PCR(Q5® High-

Fidelity DNA Polymerase,New England Biolabs)扩增选择用于克隆的片段。将来自巴斯德毕赤酵母CBS7435 mut^{S}的基因组DNA、合成基因或gBlocks(整合DNA技术)用作PCR模板。将扩增的编码序列克隆到基于pPUZZLE的表达质粒pPM2dZ30(描述在WO2008/128701A2)中,或克隆到GoldenPiCS载体(Prielhofer等,2017)中。信号肽(SP)与启动子、终止子和产物编码序列被直接装配到表达质粒中。

[0330] GoldenPiCS系统(Prielhofer等2017,BMC Systems Biol.doi:10.1186/s12918-017-0492-3)需要在一些编码序列中引入沉默突变。或者,从商业供应商(包括Integrated DNA Technology IDT,Geneart和ATUM)获得gBlocks或合成密码子优化的基因。将扩增的编码序列克隆到基于pPUZZLE的表达质粒pPM2aK21或pPM2eH21,或GoldenPiCS系统(由骨架BB1,BB2和BB3aZ/BB3aK/BB3eH/BB3rN组成)中。通过使用限制酶将表1和表2中所列的基因片段导入表达质粒。在Prielhofer等,2017(BMC Systems Biol.doi:10.1186/s12918-017-0492-3)中描述了用于在BB2或BB3骨架中装配表达盒的所有启动子和终止子。pPM2aK21和BB3aK允许整合到3'-AOX1基因组区域中,并且包含KanMX选择标记盒以用于在大肠杆菌和酵母中进行选择。pPM2eH21和BB3eH含有5'-ENO1基因组整合区域和用于在潮霉素上选择的HphMX选择标记盒。BB3rN含有5'-RGI1基因组整合区和用于在诺尔丝菌素上选择的NatMX选择标记盒。BB3aZ质粒含有博来霉素选择标记盒,并用AscI线性化以允许整合到3'-AOX1基因组区中。所有质粒都含有大肠杆菌(pUC19)的复制起点。

[0331] 电穿孔(实施例3)后,在30℃温育48小时后,在含有整合的选择标记所需抗生素(用于具有Sh ble基因的博来霉素抗性标记的50μg/mL博来霉素,用于KanMX的500μg/mL G418,用于NatMX的100μG/mL诺尔丝菌素)的YPD(酵母提取物、蛋白胨、右旋糖)平板上选择转化体,并在相同条件下挑选(singled)。

[0332] pPM2d_pGAP和pPM2d_pAOX表达质粒是WO2008/128701A2中所述的pPuzzle_ZeoR质粒骨架的衍生物,其由pUC19细菌复制起点和博来霉素(Zeocin)抗生素抗性盒组成。异源基因的表达分别由巴斯德毕赤酵母甘油醛-3-磷酸脱氢酶(GAP)启动子或醇氧化酶(AOX)启动子和酿酒酵母CYC1转录终止子介导。一些质粒已经含有N-端酿酒酵母α交配因子前原前导序列。在用XhoI和BamHI(用于scR)或EcoRV(用于VHH)限制性消化后,将每个基因连接到用XhoI和BamHI或EcoRV消化的质粒pPM2d_pGAP和pPM2d_pAOX中。

[0333] 在电穿孔(采用描述于Gasser等,2013.Future Microbiol.8(2):191-208中的标准的转化方案-参见实施例3)至巴斯德毕赤酵母中之前,分别采用AvrII限制性酶(针对pPM2d_pGAP)或PmeI限制性酶(针对pPM2d_pAOX)将质粒线性化。在含有50μg/mL博来霉素的YPD板(每升:10g酵母提取物、20g蛋白胨、20g葡萄糖、20g琼脂-琼脂)上实施阳性转化体的筛选。

[0334] 在以下实施例中,将不同的异源蛋白用作报道子:二价骆驼抗体的可变vH区(VHH)、信号链可变片段抗体(scFv,命名为scR)、人IgG的抗原结合片段(SDZ-Fab)和荧光蛋白mCherry。异源POI的表达由合适的巴斯德毕赤酵母启动子(如由巴斯德毕赤酵母醇氧化酶(AOX1)启动子(VHH,scR,SDZ-Fab-HC,mCherry)或二羟基丙酮合酶(DAS1)启动子(SDZ-Fab-LC))和酿酒酵母CYC1(VHH,scR,mCherry和SDZ-Fab-HC)或巴斯德毕赤酵母TDH1(SDZ-Fab-LC)转录终止子介导。通过使用如Prielhofer等(2017)中所述的采用Golden Gate组装法的GoldenPiCS试剂盒来完成质粒的构建。通过PCR,使用含有SP序列的引物从载体DNA模

板(携带目的基因)扩增POI基因,并使用限制酶BsaI将每个POI基因连接到BB1质粒中,从而产生携带测试的信号肽的表达载体。对于多聚体蛋白(如Fab),然后通过使用限制酶BpiI将各表达载体BB1_SDZ-Fab-HC和BB1_SDZ-Fab-LC装配到BB2质粒中,得到BB2_ab_pAOX1_SDz-Fab-HC-CYC1tt和BB2_bc_pDAS1_SDZ-Fab-LC-TDH1tt质粒。最后将这些BB2质粒组合,产生BB3aZ_SDZ-Fab质粒,其包含Fab片段的HC和LC的表达盒。对于VHH和scR,BB1质粒直接与BB3aZ质粒装配,分别产生BB3aZ_VHH和BB3aZ_scR。新型信号肽通过使用融合PCR作为5'引物序列的一部分添加,或在通过PCR扩增它们之后通过Golden Gate组装法添加。在序列验证之后,通过使用相容的限制酶Bpi I和BsaI将所有蛋白的表达盒组合到一个载体上。在表2中给出了重组蛋白的编码DNA序列。对于分泌,使用单独的新鉴定的信号肽(序列参见表7)、新鉴定的信号肽与酿酒酵母α交配因子(MFα)分泌前导序列(SEQ ID NO:3)的组合、或使用酿酒酵母MFα的完整分泌信号(SEQ ID NO:4)。

[0335] 实施例3:生产具有新型信号肽序列的重组分泌蛋白的巴斯德毕赤酵母菌株的产生

[0336] 在电穿孔至巴斯德毕赤酵母之前,通过AscI将POI表达质粒(高达3μg)线性化。为此目的,将巴斯德毕赤酵母菌株制成电感受态。将菌株接种到100mL YPD培养基(主培养物)中达16-20小时(25℃;180rpm),并通过在两个50mL离心管中离心(5分钟;1500g;4℃)在光密度(OD 600)为1.8-3时收获。将细胞沉淀重悬于10mL YPD+20mM HEPES+25mM DTT中,并温育(30分钟;25℃;180rpm)。在温育一段时间之后,向离心管中加入40mL冰冷的无菌蒸馏水,并离心(5分钟;1500g;4℃)(Eppendorf AG,Germany)。将细胞沉淀重悬于冰冷的无菌1mM HEPES缓冲液(pH 8)中,离心(30分钟;25℃;180rpm)。将细胞沉淀重悬于45mL冰冷的1M山梨醇中,离心(30分钟;25℃;180rpm)。将沉淀重悬于500μL冰冷的1M山梨醇中,并以80μL等分至冰冷的1.5mL Eppendorf管中。将等分的电感受态细胞保持在-80℃下直到使用。

[0337] 电穿孔在2kV下进行4毫秒(Gene Pulser,Bio-Rad Laboratories,Inc,USA)。转化后,将电穿孔的细胞悬浮于1mL YPD培养基中,在30℃下以650rpm的转速在thermoshaker(Eppendorf AG,Germany)上振荡再生1.5小时至3小时。之后,将20μL和200μL细胞悬浮液铺板于含有50μg/mL博来霉素(CBS7435 mutS背景)或50μg/mL博来霉素和100μg/mL诺尔丝菌素(CBS7435 mutS共表达SRP,参见实施例5)的YPD平板(每升:10g酵母提取物,20g蛋白胨,20g葡萄糖,20g琼脂-琼脂)上用于选择,并在30℃下温育48小时。将出现的菌落重新划线接种到含有适当抗生素的新鲜YPD平板上。

[0338] 实施例4:生产具有新型信号肽序列的重组分泌蛋白的巴斯德毕赤酵母菌株的小规模培养(筛选)

[0339] 培养

[0340] 将用空气渗透膜密封的24深孔板(DWP)用于筛选。从转化平板挑取单菌落(来自每次转化的20个菌落)用于预培养,并用于接种2mL YPD,基于用于筛选的抗生素抗性所述YPD具有合适的抗生素。使预培养物在24-DWP中在25℃和280rpm下生长约24小时,并且随后用于接种2mL含有25g·L^{-1}多糖和0.35%葡萄糖释放酶溶液(Enpresso)的合成筛选培养基ASMv6(每升:22.0g柠檬酸一水合物、6.3g$(NH_4)_2HPO_4$、0.49g$MgSO_4*7H_2O$、2.64g KCl、0.054g $CaCl_2*2H_2O$、1.47mL PTM0痕量盐储备液(每升:6.0g $CuSO_4\cdot 5H_2O$、0.08g NaI、3.36g $MnSO_4\cdot H_2O$、0.2g$Na_2MoO_4\cdot 2H_2O$、0.02g H_3BO_3、0.82g $CoCl_2*2H_2O$、20.0g $ZnCl_2$、

65.0gFeSO$_4$·7H$_2$O和5.0mL H$_2$SO$_4$(95%-98%));4mg生物素;用0.75g KOH(固体)将pH设定为6.695)至起始OD$_{600}$为8(t=0小时)。主培养持续48小时,并四次添加甲醇用于诱导(在t=4h时0.5%,在t=20h、t=28h和t=44h时1%各一次)。

[0341] 收获和分析

[0342] 48小时后,取出1mL的每种培养物并在预称重的Eppendorf管中离心。通过称重具有细胞沉淀物的Eppendorf管来测定湿细胞重量(WCW),并如下计算:重量(满)-重量(空)=湿池重量(WCW)(g/L)。使用如下所述的微流体毛细管电泳(mCE)、ELISA或荧光光谱分析,将上清液用于定量重组分泌蛋白的浓度。从该数据计算产率:产量(μg/mg)=蛋白浓度/湿细胞重量。体积滴度(也称为"滴度")是如实施例4和6中所述的培养上清液中目的蛋白(POI)的含量,以g/L或mg/L等计。在该分析中仅包含具有一个拷贝的目的基因的克隆。通过qPCR测定选择用于生物反应器培养的克隆的基因拷贝数(实施例4)。分别给出了相对于参照克隆的效价和产量倍数改变,所述参照克隆具有1拷贝的目的基因,其通过使用酿酒酵母MFα分泌信号(SEQ ID NO:4)来分泌。

[0343] 通过微流体毛细管电泳(mCE)的定量

[0344] 将'LabChip GX/GXII System"(PerkinElmer)用于定量分析培养物上清液中的分泌蛋白滴度。使用耗材'Protein Express Lab Chip'(760499,PerkinElmer)和'Protein Express Reagent Kit'(CLS960008,PerkinElmer)。简言之,将6μL培养上清液与21μL非还原性样品缓冲液混合。将该混合物在100℃下变性5分钟,短暂离心,并进一步与105μL水(Milli-Q®或等同物)混合。然后将样品在1200g下离心2分钟并施加到仪器上。使用基于微流体的电泳系统,根据蛋白大小在仪器中分析荧光标记的样品。内标物使得以kDa计的大小和检测信号的近似浓度的近似分配成为可能。

[0345] 通过ELISA的Fab的定量

[0346] 采用抗-人IgG抗体(ab7497,Abcam)作为包被抗体和山羊抗-人IgG(Fc特异性)-碱性磷酸酶缀合的抗体(Sigma A8542)作为检测抗体,通过ELISA进行完整Fab的定量。人Fab/κ、IgG片段(Bethyl P80-115)用作标准品,起始浓度为100ng/mL,相应地稀释上清液样品。用pNPP(Sigma S0942)进行检测。包被-、稀释-和洗涤缓冲液是基于PBS(2mM KH$_2$PO$_4$、10mM Na$_2$HPO$_4$.2 H$_2$O、2.7mM g KCl、8mM NaCl,pH 7.4)的,并相应地用BSA(1%(w/v))和/或吐温20(0.1%(v/v))来完成。

[0347] 通过荧光光谱法的定量

[0348] 通过将100μL的每种筛选上清液转移到FluoroNunc™/LumiNunc™96孔板中直接测量分泌的mCherry荧光(参考10366281,Thermo Fischer Scientific,Waltham,USA)。将mCherry标准品(参考TP790040,OriGene Technologies,Herford,Germany)在相同的板中用PBSG(10%甘油中的1×PBS)稀释,得到0-8080ng.μL^{-1}的标准曲线。将板装入Infinite M200装置中,并通过i-控制v.1.6软件来控制。在测量之前,将样品以线性模式振动5秒,振幅为1mm。以荧光顶部读取模式进行的测量包括在587nm(9nm带宽)下25次每次20μs的闪光,并且在640nm(20nm带宽)下无滞后时间地读取发射。

[0349] 通过定量实时PCR(qPCR)的基因拷贝数的测定

[0350] 首先制备来自1mL培养物的细胞沉淀以用于用溶细胞酶(5U/μL,其中加入了50mM

2-巯基乙醇)裂解,然后用DNeasy® Blood&Tissue Kit(Qiagen)进一步提取基因组DNA。qPCR反应物由0.25μL正向引物和0.25μL反向引物(表5)、5μL 2x SensiMix SYBR Hi-ROX Kit(Bioline)、3.5μL无核酸酶的水和1μL样品组成。在以下条件下进行qPCR:在95℃下热启动10分钟,接着在Rotorgene 6000(Qiagen)上进行45个循环,95℃下15s,60℃下20s和72℃下15s。通过使用Rotor-Gene软件比较定量方法,将各个样品的荧光信号相对于所选对照样品标准化来确定GCN。该比值进一步被标准化为同一样品的ACT1信号以补偿初始浓度差异。

[0351] 表5:定量PCR分析所需的引物列表。DNA序列的符号从5'到3'

[0352]

基因名称	正向引物	反向引物
SRP14	AGACACGCAAAGGAGAAAA (SEQ ID NO: 30)	GTCCACATAATCTCTCCAGAAG (SEQ ID NO: 31)
SRP21	GGGGAGCCACCAGTATTTT (SEQ ID NO: 32)	CCACCTTTACCACCTTTCTTT (SEQ ID NO: 33)
SEC65	TTCCTTTGCATTCGCCATT (SEQ ID NO: 34)	TTCTTCTGTTTCGGAGCTTTG (SEQ ID NO: 35)
SPR54	AAGATGATGGCTCGTATGG (SEQ ID NO: 36)	TCCTGGGTTTGATTGCATTT (SEQ ID NO: 37)
SRP68	CAACTACGTGAGTCAACCAA (SEQ ID NO: 38)	AGCGGAACAATCCAAACAA (SEQ ID NO: 39)
SRP72	TGCCTTCAAAACCCTCAATC (SEQ ID NO: 40)	GGTCGTGGTAGTGTTATTGT (SEQ ID NO: 41)
scpRNA	GGGAAGGCGAGCAATAAG (SEQ ID NO: 42)	ACCAACAGCCCATTACCA (SEQ ID NO: 43)
scR	AAGCCTGGTAAGCCTCCAAAGT (SEQ ID NO: 44)	TCCTCAGCTTGAACACCACCAAT (SEQ ID NO: 45)
VHH	TGTAACGTGAATGTCGGATTTG (SEQ ID NO: 46)	TAGTGATGGTGGTGGTGATG (SEQ ID NO: 47)
SDZ-Fab-HC	TACTGCTGCTTTGGGTTGTTTGGT (SEQ ID NO: 48)	AAGGGACAGTAACAACAGAGGACA (SEQ ID NO: 49)
SDZ-Fab-LC	GATGAACAATTGAAGTCTGGTAC (SEQ ID NO: 50)	GAGTAACTTCACAAGCGTAAACC (SEQ ID NO: 51)
mCherry	CATCAAGTTGGACATCACCTCC (SEQ ID NO: 67)	CACCCTTGTACAGCTCGTCC (SEQ ID NO: 68)

[0353]

[0354] 实施例5:信号识别颗粒(SRP)过表达巴斯德毕赤酵母菌株的产生

[0355] 本发明人进一步决定在SRP过表达菌株中测试SP,以测试分泌途径是否准备好应对与新型SP融合的重组蛋白的高负荷。

[0356] 为了产生SRP表达质粒236_BB3rN,使用Prielhofer等(2017)所述的基于Golden Gate的克隆将SRP的所有七个亚基(6个蛋白亚基和1个非编码RNA)的表达盒装配在一个单一质粒中,以确保相等的基因拷贝数。通过PCR从CBS7435的基因组中扩增出所有亚基的基因,并将它们克隆到BB1质粒中,然后将这些BB1质粒与BB2质粒中的各自启动子和终止子组装。表6中给出了SRP亚基表达所用的启动子和转录终止子。用锤头状和HDV核酶过表达非编码RNA,以移除RNA聚合酶II转录后的mRNA特征。

[0357] 表6:用于SRP背景菌株的SRP亚基的过表达的启动子和终止子。

[0358]

SRP亚基	染色体位置	启动子*	终止子*
SRP68	PP7435_Chr1-0901	pADH2	RPL2aTT
SRP72	PP7435_Chr1-0988	pPOR1	RPP1bTT
SRP9-21	PP7435_Chr3-0697	pPDC1	RPS25aTT
SEC65	PP7435_Chr3-0671	pRPP1b	RPS17bTT
SRP54	PP7435_Chr4-0671	pFBA1-1	RPS2TT
SRP14	PP7435_Chr4-0320	pGPM1	RPS3TT
非编码RNA	PP7435_Chr1-2610	pTEF2	IDP1TT

[0359] *启动子和终止子序列的参考:Prielhofer,R.,Barrero,J.J.,Steuer,S.等,GoldenPiCS:a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris.BMC Syst Biol 11,123(2017).https://doi.org/10.1186/s12918-017-0492-3

[0360] 为了产生SRP过表达的背景菌株,用携带所有七个SRP亚基的质粒236_BB3rN转化CBS7435 mutS(实施例3)。根据24深孔板筛选方法(实施例4)培养20个转化体和4个对照菌株。转化的与4个未转化的克隆的湿细胞重量之间没有显示任何显著差异。接着,对三个随机选择的克隆(克隆#3、#7和#10)进行7个SRP蛋白组分的基因拷贝数(GCN)分析(实施例4)。对于#3和#10,SRP基因的GCN测定为2,表明1个过表达盒整合到基因组中。克隆#7(命名为CBS7435 mutS SRP#7)具有3个拷贝的基因,表明整合了2个过表达盒。将该克隆选择为用于异源蛋白表达的背景SRP过表达巴斯德毕赤酵母菌株。

[0361] 实施例6:生产具有新型信号肽序列的重组分泌蛋白的巴斯德毕赤酵母菌株的生物反应器培养

[0362] 在小规模筛选培养后(实施例4),选择通过使用新型信号肽序列表达目的蛋白(POI,例如VHH,scR,SDZ-Fab,mCherry)的克隆以及使用全长MFα分泌信号(SEQ ID NO:4)的对照菌株(实施例2)。通过进料分批生物反应器培养,在较大培养体积中进一步评估所选择的克隆。

[0363] 预培养

[0364] 将相应的克隆接种到装有50mL YPhyG(每升:20.0g植物蛋白胨,10.0g细菌酵母提取物,20.0g甘油)的宽颈、带挡板(baffled)、有盖的300mL振荡摇瓶中,并在28℃下以110rpm振荡过夜(预培养物1)。从预培养物1中以OD_{600}(在600nm处测量的光密度)在下午晚些时候达到约20(针对YPhyG培养基测量)的方式接种预培养物2(宽颈、带挡板、有盖的2000mL振荡摇瓶中的200mL YPhyG)。预培养物2的温育也在28℃,110rpm下进行。

[0365] 生产-分批阶段

[0366] 发酵(生物反应器培养;所有阶段)在完全仪器化和可控的Infors Multifors 1L-反应器(750mL工作体积)中进行。将所有发酵罐从预培养2分别接种至OD_{600}为2.0,所述发酵罐装有400mL pH约为5.5的BSM-培养基(每升:13.5mL H_3PO_4(85%)、0.5g $CaCl \cdot 2H_2O$、7.5g $MgSO_4 \cdot 7H_2O$、9.0g K_2SO_4、2.0g KOH、40g甘油、0.25g NaCl、0.1mL消泡剂(Glanapon 2000)、4.35mL PTM1(每升:0.2g生物素、6.0g $CuSO_4 \cdot 5H_2O$、0.09g KI、3.0g$MnSO_4 \cdot H_2O$、0.2g $Na_2MoO_4 \cdot 2H_2O$、0.02g H_3BO_3、0.5g $CoCl_2$、42.2g$ZnSO_4 \cdot 7H_2O$、65.0g Fe(II)$SO_4 \cdot 7H_2O$、5mL H_2SO_4)。巴斯德毕赤酵母在甘油上生长以产生生物质,并且所述培养物随后经受甘油进料(60%w/w+12ml/L PTM1),随后是混合的甲醇/甘油进料,然后是甲醇进料。氨溶液(25%)用于pH控制。

[0367] 在初始分批阶段,将温度设定为28℃。在开始生产阶段之前的最后一小时期间,将其降至24℃并在整个剩余过程中保持在该水平,同时pH降至5.0并保持在该水平。将整个过程(级联控制:搅拌器,流量,氧补充)的氧饱和度设定为30%。以700至1200rpm施加搅拌并选择1.0-2.0L min^{-1}的流量范围(空气)。

[0368] 在分批阶段,生成生物质($\mu \sim 0.30h^{-1}$)直至约110-120g/L的湿细胞重量(WCW)。经典的分批阶段(生物质产生)将持续约14小时。

[0369] 生产-进料分批阶段

[0370] 甘油以由方程2.6+0.3*t(g甘油(60%)/h)定义的速率进料,其中t=以小时计的时间,因此在8小时内补充总共30g甘油(60%)。将第一取样点选择为20小时。在接下来的18个小时(从过程时间20至38小时),应用甘油/甲醇的混合进料:

[0371] -甘油进料速率由以下方程定义:2.5+0.13*t(甘油(60%)/h),供给66g甘油(60%)

[0372] -甲醇进料速率由以下方程定义:0.72+0.05*t(g甲醇(100%)/h),加入21g甲醇

[0373] 在接下来的72.5个小时(从过程时间38到约110.5小时),甲醇通过方程2.2+0.016*t(g甲醇(100%)/h)进料,从而供给约223g甲醇。

[0374] 取样和分析

[0375] 采用以下程序,在指定的时间点取样:将取样的发酵液(fermentation broth)的第一个3mL(用注射器)弃去。将1x 1mL的新鲜采集样品(3-5mL)转移至1.5mL离心管中,并在13200rpm(16100g)下旋转5分钟。将上清液迅速转移至单独的小瓶中,并与相应的湿沉淀部分一起立即冷冻。为了确定WCW,将1mL的发酵液在称量去皮的Eppendorf小瓶中以13200rpm(16100g)离心5分钟,然后将所得的上清液准确地移除。将小瓶称重(精确度0.1mg),并减去空瓶的皮重以获得湿细胞重量。如实施例4中所述进行蛋白质的定量。

[0376] 实施例7:生产具有新型信号肽序列的重组分泌蛋白的巴斯德毕赤酵母菌株的筛选结果

[0377] 如实施例3所述进行巴斯德毕赤酵母菌株的转化。通过滴度和产量倍数变化值(也称为FC滴度或滴度FC相应的FC产量或产量FC)测量用新型SP获得的分泌改善(实施例4),所述滴度和产量倍数变化值与在相同菌株背景(CBS7435 mutS WT或CBS7435 mutS SRP#7,实施例5)中使用MFα分泌信号(SEQ ID NO 4)分泌POI的相应对照克隆(实施例1)相关。将滴度倍数变化理解为为相应发酵或小规模培养的滴度除以对照的滴度的商。将产量倍数变化理解为相应发酵或小规模培养的产量除以对照产量的商。

[0378] 如实施例4所述,进行与MFα分泌信号相比新型信号肽(SP)生产目的蛋白(POI)的性能的筛选。

[0379] 表7中示出了使用VHH作为分泌的报道蛋白,表4和表7中所示的SP候选物的筛选结果。

[0380] 表7:与在SRP背景(CBS7435 mutS SRP#7)中使用MFα分泌信号相比,通过使用不含MFα-前导序列的单独的候选SP作为分泌信号来分泌作为报告子POI的VHH或mCherry而获得的筛选结果。显示了每种构建体高达20个克隆的滴度的平均倍数变化。

[0381]

名称	POI	FC 滴度	序列
SP4*	VHH	1.34	MKLISVGIVTTLLTLASC (SEQ ID NO: 2)
SP10	VHH	0.86	MLVAWFLLLLVSSCIC (SEQ ID NO: 56)
SP22	VHH	0.59	MKFAISTLLIILQAAAVFA (SEQ ID NO: 57)
SP8	VHH	0.59	MKFGLGSLGLAVALIPIASA (SEQ ID NO: 58)
SP9	VHH	0.61	MIILLPLLFLFVAGLVQA (SEQ ID NO: 59)
SP2	VHH	不正确的切割	MDPFSILLTLTLIILA (SEQ ID NO: 60)
SP15	VHH	0.55	MRLSYECLFSVFLVLAYHLKGTKA (SEQ ID NO: 61)
SP11	VHH	0.65	MINLNSFLILTVTLLSPALA (SEQ ID NO: 62)
SP16	VHH	0.53	MQLQYLAVLCALLLNVQS (SEQ ID NO: 63)
SP13	VHH	0.47	MWIERNLIASILLFSTSAYA (SEQ ID NO: 64)
SP18	VHH	不正确的切割	MNISTASKISRLLQLVIALISLVLT (SEQ ID NO: 65)
SP3	VHH	0.22	MFVFEPVLLAVLVASTCVTA (SEQ ID NO: 65)
SP14	VHH	1.04	MLNKLFIAILIVITAVIG (SEQ ID NO: 1)
SP14	mCherry	1.40	MLNKLFIAILIVITAVIG (SEQ ID NO: 1)

[0382] 如表7所示,当用作POI分泌的唯一分泌信号时,SP4和SP14代表了高效SP,与MFα分泌信号相比,超过分泌产物滴度约1.3-1.4倍。

[0383] 接下来,发明人决定通过将新型SP融合到前导序列(如MFα前导序列)来测试是否存在进一步的影响。

[0384] 当将信号肽与MFα-前导序列融合时,观察到了主要的影响(见表8)。从这些筛选中我们得出结论:诸如MFα-前导序列的前导序列的添加对重组蛋白生产甚至更有利。

[0385] 表8:与在CBS7435 mutS WT(WT)或CBS7435 mutS SRP#7(SRP)背景中使用MFα分泌信号相比,通过使用与MFα-前导序列融合的SP、SP4和SP14分泌作为报告子POI的VHH、scR或SDZ-Fab而获得的筛选结果。

[0386] 显示了每种构建体高达20个克隆的滴度的平均倍数变化。

[0387]

背景	蛋白	前原	前导	滴度FC	产量FC
SRP	VHH	SP4	MFα-前导	1.9	1.9
SRP	VHH	SP14	MFα-前导	1.7	1.8
WT	VHH	SP14	MFα-前导	1.5	1.5
SRP	scR	SP4	MFα-前导	1.4	1.5
SRP	scR	SP14	MFα-前导	1.3	1.3
WT	scR	SP4	MFα-前导	1.2	1.2
WT	SDZ-Fab	SP14	MFα-前导	1.7	1.7

[0388] 与和MFα-前导序列的组合相比,不同信号肽(SP)与其它毕赤酵母(Komagataella phaffii)衍生的前导-序列的融合导致POI分泌较少或甚至不分泌(参见表9)。

[0389] 表9:与在CBS7435 mutS WT(WT)或CBS7435 mutS SRP#7(SRP)背景中使用MFα分泌信号(SEQ ID NO:4)相比,通过使用融合到不同的巴斯德毕赤酵母(Komagataella phaffii)前导序列的SP、SP4和SP14分泌作为报告子POI的VHH、scR或SDZ-Fab而获得的筛选结果。显示了每种构建体高达20个克隆的滴度的平均倍数变化。

[0390]

背景	蛋白	前原	前导	滴度 FC	产量 FC
WT	VHH	SP4	SEQ ID NO: 69	0.0	0.0
WT	VHH	SP14	SEQ ID NO: 69	0.0	0.0
WT	VHH	SP4	SEQ ID NO: 70	0.7	0.7
WT	VHH	SP14	SEQ ID NO: 70	0.4	0.5
WT	VHH	SP4	SEQ ID NO: 71	0.6	0.6
WT	VHH	SP14	SEQ ID NO: 71	0.8	0.8
WT	VHH	SP4	SEQ ID NO: 72	0.2	0.2
WT	VHH	SP14	SEQ ID NO: 72	0.0	0.0

[0391]

WT	VHH	SP4	SEQ ID NO: 73	0.4	0.4
WT	VHH	SP14	SEQ ID NO: 73	0.8	0.8
WT	scR	SP4	SEQ ID NO: 71	0.6	0.6
WT	SDZ-Fab	SP14	SEQ ID NO: 71	0.4	0.4
WT	SDZ-Fab	SP14	SEQ ID NO: 73	0.0	0.0

[0392] 实施例8:生产具有新型信号肽序列的重组分泌蛋白的巴斯德毕赤酵母菌株的进料分批培养

[0393] 如实施例6中所述进行生物反应器培养。首先,如实施例3和4)中所述获得和筛选分泌POI的单拷贝克隆(参见表2)并将它们用于生物反应器培养以进一步证实与作为对照的MFα分泌信号相比,不含(表7)或含有(表8)MFα前导序列的SP4和SP14的生产性能。

[0394] 为了生产POI和测试本发明的信号肽(SP4(SEQ ID NO 2)和SP14(SEQ ID NO 1)的

分泌增加效果,选择在筛选POI滴度、POI产量和生长方面表现最好的菌株用于生物反应器培养,这些菌株也可以是多拷贝菌株(即它们含有多拷贝的POI表达盒)。因此,为了通过单独使用没有任何前导序列的SP4或SP14或通过使用具有MFα前导序列的SP4或SP14来表达和分泌POI,如实施例2、3、4和5中所述产生和选择载体和菌株,并将它们如实施例6中所述培养,除了在如实施例4中所述筛选后,如实施例6中所述选择表现最佳的一个或多个菌株用于发酵。这也适用于通过使用MFα分泌信号(SEQ ID NO:4)表达相应POI的相应对照菌株。计算与通过使用MFα分泌信号(SEQ ID NO:4)的表达/分泌相关的通过使用具有或不具有MFα前导序列的SP4或SP14的POI的表达/分泌的滴度和产量倍数变化。

[0395] 采用生物反应器培养,我们证实了,与MFα分泌信号(SEQ ID NO:4)相比,与MFα前导序列组合的SP4和SP14明显提高了分泌能力,这可以从POI滴度和产量看出(表10)。同样,在WT和SRP过表达菌株背景之间没有观察到强烈的差异,虽然与在WT中相比,在SRP过表达菌株中采用SP4的scR分泌水平更高,这表明SRP过表达对一些POI产生益处。

[0396] 表10:与在CBS7435 mutS WT或SRP背景(CBS7435 mutS SRP#7)中使用MFα分泌信号相比,通过使用融合到MFα前导序列的候选SP分泌作为报告子POI的VHH、scR或SDZ-Fab而获得的生物反应器结果。通过使用最后的取样点,与MFα分泌信号对照相比,计算滴度和产量FC。

[0397]

背景	蛋白	前原	前导	滴度FC	产量FC
SRP	VHH	SP4	MFα-前导	1.3	1.5
WT	VHH	SP14	MFα-前导	1.4	1.5
SRP	scR	SP4	MFα-前导	1.4	1.4
SRP	scR	SP14	MFα-前导	1.2	1.2
WT	scR	SP4	MFα-前导	1.1	1.1
WT	SDZ-Fab	SP14	MFα-前导	1.8	1.9

[0398] 实施例9:当使用新型信号肽序列时易位的共翻译模式的确认

[0399] 使用免疫荧光显微术来确认SP4和SP14的共翻译易位模式。将用SP4或SP14生产VHH的巴斯德毕赤酵母菌株用于显微术分析(菌株的产生参见实施例2)。将小鼠抗6xHis抗体(abcam,ab18184)和适当的荧光标记的二抗用于免疫荧光显微术。尽管使用标准MFα分泌信号的克隆显示了表明翻译后易位(图1A)的普遍的胞质模式,但SP4-VHH和SP14-VHH都导致典型的ER模式(核环),这是共翻译易位模式(图1B-C)的指示。这证实了它们优先被共翻译机制易位。

[0400] 参考文献

[0401] Aw&Polizzi.Microb Cell Fact.2013.12,128.

[0402] Ecker等,MAbs.2015.7,9-14.

[0403] Feige等,Trends Biochem Sci.2010.35,189-89

[0404] Fitzgerald&Glick.Microb Cell Fact.2014.13,125.

[0405] Gasser等,Future Microbiol.2013.8(2),191-208.

[0406] Janda等,Nature.2010.465,507-510.

[0407] Kyte&Doolittle.J Mol Biol.1982.157(1),105-32.

[0408] Lin-Cereghino等,Gene.2013.519,311-7.

[0409] Looser等,Biotechnol Adv.2015.33,1177-93.
[0410] Nelson&Reichert.Nat Biotechnol.2009.27,331-7.
[0411] Ng等,The Journal of cell biology.1996.134(2),269-78.
[0412] Nielsen.Methods Mol Biol.2017.1611,59-73.
[0413] Pechmann等,Nat Struct Mol Biol.2014.21,1100-5.
[0414] Pfeffer等,Microb Cell Fact.2011.10,47.
[0415] Prielhofer等,BMC Syst Biol.2017.11,123.
[0416] Prielhofer等,BMC Genomics.2015.16,167.
[0417] Rapoport.Nature.2007.450,663-669.
[0418] Schwarzhans等,Microb Cell Fact.2016.15,84.
[0419] Sharp和Li.Nucleic Acids Res.1987.15,1281-95.
[0420] Sturmberger等,Journal of biotechnology.2016.235,121-31.
[0421] Valli等,FEMS Yeast Research.2016.16(6)2016.
[0422] Walsh.Nat Biotechnol.2014.32,992-1000.
[0423] Zahrl等,Microbiology.2018.

序列表

<110> 勃林格殷格翰 RCV 两合公司
瓦利多根股份有限公司
龙沙有限公司

<120> 用于增加蛋白质分泌的信号肽

<130> LC23310019P

<150> EP 21 156 986.8

<151> 2021-02-12

<160> 80

<170> PatentIn version 3.5

<210> 1

<211> 18

<212> PRT

<213> Komagataella phaffii

<400> 1

Met Leu Asn Lys Leu Phe Ile Ala Ile Leu Ile Val Ile Thr Ala Val
1 5 10 15
Ile Gly

<210> 2

<211> 18

<212> PRT

<213> Komagataella phaffii

<400> 2

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

<210> 3

<211> 66

<212> PRT

<213> 人工序列

<220>

<223> 包含L23S, D64E的α-MF的前导序列

<400> 3

Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala
1 5 10 15
Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala
20 25 30
Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn

35 40 45

Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Glu
50 55 60

Lys Arg
65

<210> 4
<211> 85
<212> PRT
<213> 人工序列
<220>
<223> L23S, D64E的α-MF(前原前导序列)
<400> 4

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15
Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
20 25 30
Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe
35 40 45
Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
50 55 60
Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65 70 75 80
Ser Leu Glu Lys Arg
85

<210> 5
<211> 567
<212> PRT
<213> Komagataella phaffii
<400> 5

Met Glu Ser Pro Leu Gln Ser Thr Tyr Gly Glu Arg Ala Glu Arg Tyr
1 5 10 15
Leu Asp Ser Ala Asp Ala Phe His Lys Gln Arg His Arg Leu Asn Arg
20 25 30
Arg Leu His Lys Leu Arg Lys Ser Leu Asp Ile His Val Thr Asp Thr
35 40 45
Lys Asn Tyr Arg Glu Lys Glu Gln Ile Ser Lys Ile Asp Leu Glu Ser
50 55 60
Tyr Asn Arg Asp Lys Arg Tyr Gly Asp Ile Ile Leu Phe Thr Ala Glu
65 70 75 80

Arg Asp His Met Tyr Ser Asp Glu Val Lys Glu Ile Met Lys Val His
85 90 95
His Ser Lys Ser Arg Glu Lys Phe Ile Val Ser Arg Leu Lys Arg Ser
100 105 110
Leu Asp His Gly Arg Lys Leu Leu Ile Leu Val Gly Asp Glu Pro Asp
115 120 125
Glu Met Arg Lys Leu Glu Val Phe Val Tyr Val Ala Leu Ile Gln Gly
130 135 140
Lys Leu Ser Ile Ala Asn Lys Asn Trp Thr Asn Ala Gln Tyr Ala Leu
145 150 155 160
Ser Val Ala Arg Cys Gly Leu Gln Phe Leu Asp Lys Tyr Gly Thr Glu
165 170 175
Thr Gln Thr Asp Leu Tyr Asn Gly Ile Ile Asp Thr His Ile Asp Gln
180 185 190
Met Leu Lys Phe Val Ile Tyr Gln Ala Thr Lys Asn Asn Ser Pro Ile
195 200 205
Leu Asp Thr Glu Cys Arg His Gln Ile Arg Thr Asp Thr Leu Gly Tyr
210 215 220
Leu Asp Gln Ala Arg Gln Ile Ile Glu Ser Lys Asp Pro Glu Phe Leu
225 230 235 240
Asn Val Gly Val Val Glu Thr Gln Leu Ile Trp Trp Asp Tyr Asp Ile
245 250 255
Ser Ile His Ser Glu Glu Val Ala Lys Leu Ile Ser Asp Ala Asn Glu
260 265 270
Lys Leu Gln Leu Ile Glu Asp Gly Asn Val Ser Ser Tyr Asp Pro Ala
275 280 285
Leu Leu Thr Leu Gln Glu Ala Leu Asp Ala His Gln Leu Leu Met Ala
290 295 300
Arg Asn Val Asp Asn Phe Ala Asp Asp Asp Gln Asn Asn His Val Leu
305 310 315 320
Leu Ser Tyr Ile Arg Tyr Leu Leu Leu Ile Thr Thr Leu Arg Arg Asp
325 330 335
Ile Thr Leu Ile Asp Gln Val Arg Asn Arg Ser Val Val Asn Ser Ser
340 345 350
Leu Ala Val Ala Leu Glu Arg Ala Lys Asp Val Gly Arg Ile Phe Asp
355 360 365
Asn Ile Val Lys Lys Val Asn Glu Leu Lys Asp Val Pro Gly Val Tyr
370 375 380
Asn Lys Gln Glu Glu Trp Asn Ser Leu Gln Ala Leu Asp Ala Tyr Phe

385 390 395 400
Gln Ala Ser Lys Ile Gln His Leu Ala Ser Thr His Leu Leu Phe Asn
405 410 415
Arg Ser Lys Glu Ser Leu Ala Leu Leu Ile Lys Ala Lys Ser Leu Val
420 425 430
Lys Gly His Thr Ile Ala Gly Glu Tyr Pro Thr Asn Phe Pro Thr Asn
435 440 445
Lys Asp Leu Ser Ser Ile Leu Glu Gln Ile Asn Gln Asp Ile Leu Lys
450 455 460
Ala Tyr Val Leu Ala Lys Tyr Lys Gln Glu Ser Ser Leu Gly Gly Val
465 470 475 480
Ser Glu Tyr Asp Phe Ile Ala Asp Asn Arg Asn Lys Val Pro Ser Asn
485 490 495
Pro Ser Leu His Lys Ile Ala Ser Val Ser Tyr Lys Asn Val Lys Pro
500 505 510
Val Asn Val Lys Pro Val Leu Phe Asp Ile Ala Phe Asn Tyr Val Ser
515 520 525
Gln Pro Asn Gln Ile Phe Glu Glu Pro Ile Glu Ser Ser Asn Lys Gln
530 535 540
Glu Arg Gln Ala Asp Ser Glu Ser Pro Ser Pro Glu Lys Lys Lys Lys
545 550 555 560
Gly Leu Phe Gly Leu Phe Arg
565

<210> 6
<211> 605
<212> PRT
<213> Komagataella phaffii
<400> 6

Met Ser Ser Leu Ser Glu Leu Val Ser Glu Leu Ala Ile His Ser Glu
1 5 10 15
Lys Arg Gln Tyr Lys Glu Ala Tyr Glu Lys Ala Lys Arg Ile Ile Asp
20 25 30
Leu Gly His Pro Leu Asp Leu Asp Thr Leu Lys Leu Gly Leu Val Ala
35 40 45
Ser Ile Asn Leu Asp Gln Tyr His Asn Ala Gly Arg Leu Ile Ser Lys
50 55 60
Ser Lys Asp His Ile Val Tyr Asp Gly Met Lys Glu Phe Leu Leu Leu
65 70 75 80
Ile Gly Tyr Val Tyr Tyr Lys Asn Gly Asp Ser Lys Asn Phe Glu Thr

85 90 95

Leu Leu Lys Asp Ser Ala Phe Gln Gly Arg Ala Phe Glu His Leu Lys
100 105 110

Ala Gln Tyr Tyr Tyr Lys Ile Gly Glu Asn Glu Lys Ala Leu Lys Ile
115 120 125

Tyr Arg Glu Leu Ser Lys Asn Pro Leu Asp Glu Val Val Asp Leu Ser
130 135 140

Val Asn Glu Arg Ala Val Ile Ser Gln Leu Leu Glu Leu Asp Gly Val
145 150 155 160

Val Glu Gln Pro Val Ser Arg Pro Ile Asp Asp Ser Tyr Asp Cys Lys
165 170 175

Phe Asn Asp Ala Leu Tyr Gln Val Lys Ile Gly Asp Tyr Glu Ser Ala
180 185 190

Leu Asp Leu Leu Glu Glu Ala Lys Ala Ile Cys Glu Glu Asn Thr Lys
195 200 205

Asp Leu Pro Leu Asp Thr Arg Glu Ala Glu Ile Val Pro Ile Leu Leu
210 215 220

Gln Ile Ala Tyr Val Lys Gln Leu Lys Gly Lys Lys Glu Glu Ser Leu
225 230 235 240

Thr Ala Leu Arg Ser Leu Ser Lys Pro Lys Asp Ser Leu Leu Asp Leu
245 250 255

Ile Tyr Arg Asn Asn Leu Leu Ser Leu Arg Ile Asp Glu Tyr Gly Arg
260 265 270

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

Ser Ile Asp Ile Asn Lys Asp Lys Leu Thr Val Ser Gln Arg Val Ala
290 295 300

Leu Thr Arg Asn Glu Ser Leu Leu Ala Leu Glu Leu Gly Lys Ile Pro
305 310 315 320

Ser Gln Ser Asp Leu Lys Leu Phe Tyr Asp Ala Thr Ser Glu Phe Leu
325 330 335

Asp Leu Asn Thr Lys Leu Glu Ala Ser Met Ile Tyr Asn Tyr Phe Met
340 345 350

Arg Arg Pro Gly Gln Gln Glu Val Pro Asn Ala Leu Leu Thr Ala Gln
355 360 365

Leu Ala Ile Asn Val Gly Asn Ile Asn Asn Ala Arg Thr Val Leu Glu
370 375 380

Thr Val Val Ser Asn Asp Glu Lys Asn Leu Leu Glu Pro Ser Ile Val
385 390 395 400

Val Ser Leu Tyr Leu Ile Tyr Asp Lys Leu Gln Ser Gly Arg Leu Gln
405 410 415
Val Glu Leu Leu Lys Lys Val Ala Asp Leu Leu Leu Glu Ser Glu Ile
420 425 430
Ser Ser Thr Gln Gln Arg Lys Phe Phe Lys Asp Ile Ala Phe Lys Thr
435 440 445
Leu Asn His Asp Ala Val Leu Ala Asn Arg Leu Phe Glu Lys Leu His
450 455 460
Ser Ile Tyr Pro Asn Asp Glu Leu Val Ser Thr Tyr Leu Asn Ser Ser
465 470 475 480
Ser Asn Ala Ser Asn Asn Asn Thr Thr Thr Thr Asn Phe Ser Glu Leu
485 490 495
Asp Asp Leu Val Leu Gly Ile Asp Thr Asp Lys Leu Ile Ser Glu Gly
500 505 510
Phe Asp Thr Phe Glu Ser Ser Lys Arg Pro Thr Thr Ile Ile Ser Ser
515 520 525
Thr Asn Lys Lys Arg Arg Thr Arg Leu Lys Pro Lys His Glu Ala Lys
530 535 540
Glu Lys Tyr Lys Arg Leu Asp Glu Glu Arg Trp Leu Pro Leu Lys Asp
545 550 555 560
Arg Ser Tyr Tyr Arg Pro Lys Lys Gly Lys Lys Ile Arg Asn Thr Thr
565 570 575
Gln Gly Thr Val Thr Ser Asn Thr Ser Glu Ile Ser Gly Leu Lys Lys
580 585 590
Thr Leu Pro Lys Lys Ser Ser Lys Lys Lys Gly Arg Lys
595 600 605

<210> 7
<211> 151
<212> PRT
<213> Komagataella phaffii
<400> 7

Met Pro Pro Val Lys Ser Leu Asp Ile Phe Phe Asn Arg Thr Glu Lys
1 5 10 15
Leu Leu Glu Ala Asn Pro Thr Thr Thr Lys Val Ser Ile Lys Leu Gly
20 25 30
Val Asn Phe Asn Asp His Glu Asn Pro Gln Ser Lys His Asn Val Ile
35 40 45
Thr Val Arg Val Ser Asp Pro Val Ser Gly Ser Asn Phe Lys Phe Lys
50 55 60

Val Thr Asn Lys Thr Asp Met Leu Lys Ile Phe Ser Phe Leu Gly Pro
65 70 75 80
His Gly Ile Glu Leu Pro Ile Ser Gly Gln Gln Ser Gln Ile Lys Ser
85 90 95
Asn Asp Gln Thr Gln Ser Asp Asn Thr Glu Val Pro Thr Thr Phe His
100 105 110
Arg Gly Ala Thr Ser Ile Leu Ala Asn Lys Ala Phe Glu Lys Lys Pro
115 120 125
Leu Ile Ile Lys Asp Ser Ser Thr Ala Lys Lys Gly Gly Lys Gly Gly
130 135 140
Lys Lys Lys Gly Lys Lys Phe
145 150

<210> 8
<211> 270
<212> PRT
<213> Komagataella phaffii
<400> 8

Met Pro Leu Leu Glu Glu Ile Ser Asp Ala Glu Asp Ile Asp Asn Leu
1 5 10 15
Glu Met Asp Leu Ala Glu Phe Asp Pro Thr Leu Arg Thr Pro Ile Ala
20 25 30
Glu Gln Arg Pro Ala Pro Gln Val Val Arg Ser Gln Asp Ala Glu Ser
35 40 45
Gly Gln Thr Pro Leu Val Pro Asn Gln Asp Gln Ile Ser Gln Tyr Ile
50 55 60
Glu Gln Phe Lys Glu Gly Gly Thr Ile Asn Lys Asp Gln Val Ile Arg
65 70 75 80
Pro Asp Glu Met Met Glu Lys Glu Met Ala Glu Leu Lys Ser Phe Gln
85 90 95
Ile Leu Tyr Pro Cys Tyr Phe Asp Lys Asn Arg Ser Val Lys Glu Gly
100 105 110
Arg Arg Cys Gln Lys Glu Tyr Gly Val Glu Asn Pro Leu Ala Lys Thr
115 120 125
Ile Leu Asp Ala Cys Arg Tyr Leu Asp Ile Pro Cys Ile Leu Glu Pro
130 135 140
Glu Lys Thr His Pro Gln Asp Phe Gly Asn Pro Gly Arg Val Arg Val
145 150 155 160
Ala Ile Lys Glu Ser Gly Lys Tyr Leu Asp Glu Gln Tyr Lys Thr Lys
165 170 175

Arg Lys Leu Ile Gln Leu Val Gly Gln Phe Leu Val Glu His Pro Thr
180 185 190
Thr Leu Gln Lys Val Gln Glu Leu Pro Gly Pro Pro Glu Leu Gln Gln
195 200 205
Gly Gly Tyr Ile Pro Glu Arg Val Pro Arg Val Lys Gly Leu Lys Met
210 215 220
Asn Glu Ile Val Pro Leu His Ser Pro Phe Thr Ile Lys His Pro Ser
225 230 235 240
Thr Lys Ser Val Tyr Glu Arg Glu Pro Glu Pro Ala Pro Pro Ala Ala
245 250 255
Val Pro Lys Ala Pro Lys Gln Lys Lys Ile Met Val Arg Arg
260 265 270

<210> 9
<211> 518
<212> PRT
<213> Komagataella phaffii
<400> 9

Met Val Leu Ala Asp Leu Gly Arg Arg Ile Asn Asn Ala Val Gly Asn
1 5 10 15
Val Thr Lys Ser Asn Val Val Asp Ala Asp Val Ile Ser Asn Met Leu
20 25 30
Lys Glu Ile Cys Asn Ala Leu Leu Glu Ser Asp Val Asn Ile Lys Leu
35 40 45
Val Ala Gln Leu Arg Glu Lys Ile Arg Lys Gln Ile Asp Ala Glu Asp
50 55 60
Lys Pro Gly Ile Asn Lys Lys Lys Leu Ile Gln Lys Val Val Phe Asp
65 70 75 80
Glu Leu Val Lys Leu Val Asp Cys Asn Glu Ala Glu Leu Phe Lys Pro
85 90 95
Lys Lys Lys Gln Thr Asn Val Ile Met Met Val Gly Leu Gln Gly Ala
100 105 110
Gly Lys Thr Thr Thr Cys Thr Lys Leu Ala Val Tyr Tyr Gln Arg Arg
115 120 125
Gly Phe Lys Val Gly Met Val Cys Gly Asp Thr Phe Arg Ala Gly Ala
130 135 140
Phe Asp Gln Leu Lys Gln Asn Ala Thr Lys Ala Lys Ile Pro Tyr Tyr
145 150 155 160
Gly Ser Tyr Thr Glu Thr Asp Pro Val Lys Val Thr Phe Asp Gly Val
165 170 175

Glu Glu Phe Arg Lys Glu Lys Phe Glu Ile Ile Ile Val Asp Thr Ser
180 185 190
Gly Arg His Arg Gln Glu Glu Asp Leu Phe Glu Glu Met Val Gln Ile
195 200 205
Gly Lys Ala Ile Lys Pro Asn Gln Thr Ile Met Val Leu Asp Ala Ser
210 215 220
Ile Gly Gln Ser Ala Glu Ser Gln Ser Lys Ala Phe Lys Glu Ser Ser
225 230 235 240
Asp Phe Gly Ala Ile Ile Ile Thr Lys Met Asp Ser Asn Ser Lys Gly
245 250 255
Gly Gly Ala Leu Ser Ala Ile Ala Ala Thr Asn Thr Pro Val Ala Phe
260 265 270
Ile Ala Thr Gly Glu His Ile Gln Asn Phe Glu Lys Phe Ser Gly Arg
275 280 285
Gly Phe Ile Ser Lys Leu Leu Gly Ile Gly Asp Ile Glu Gly Leu Met
290 295 300
Glu His Val Gln Ser Met Asn Leu Asp Gln Gly Asp Thr Ile Lys Asn
305 310 315 320
Phe Lys Glu Gly Lys Phe Thr Leu Gln Asp Phe Gln Thr Gln Leu Asn
325 330 335
Asn Ile Met Lys Met Gly Pro Leu Ser Lys Leu Ala Gln Met Leu Pro
340 345 350
Gly Gly Met Gly Gln Leu Met Gly Gln Val Gly Glu Glu Glu Ala Ser
355 360 365
Lys Arg Leu Lys Arg Met Ile Tyr Ile Met Asp Ser Met Thr Lys Gln
370 375 380
Glu Leu Ser Ser Asp Gly Arg Leu Phe Ile Asp Gln Pro Ser Arg Met
385 390 395 400
Val Arg Val Ala Arg Gly Ser Gly Thr Ser Val Thr Glu Val Glu Leu
405 410 415
Val Leu Leu Gln Gln Lys Met Met Ala Arg Met Ala Leu Gln Ser Lys
420 425 430
Asn Met Met Ser Gly Ala Gly Gly Pro Ala Gly Met Ala Ser Lys Met
435 440 445
Asn Pro Ala Asn Met Arg Arg Ala Met Gln Gln Met Gln Ser Asn Pro
450 455 460
Gly Met Met Asp Asn Met Met Asn Met Phe Gly Gly Ala Gly Gly Ala
465 470 475 480
Gly Gly Ala Gly Met Pro Asp Met Gln Glu Met Met Lys Gln Met Ser

485 490 495
Ser Gly Gln Met Lys Met Pro Ser Gln Gln Glu Met Met Ser Met Met
500 505 510
Lys Gln Phe Gly Met Gly
515

<210> 10
<211> 151
<212> PRT
<213> Komagataella phaffii
<400> 10
Met Ser Thr Thr Thr Lys Lys Asn Lys Asn Arg Ile Leu Ile Glu Asn
1 5 10 15
His Lys Gln Phe Leu Glu Glu Val Ser Lys Thr Ala Thr Leu Ser Val
20 25 30
Trp Asn Ser Lys Phe Ser Ile Lys Arg Leu Ser Leu Glu Ala Asp Pro
35 40 45
Val Glu Gly Thr Pro Glu Gly Ile Arg Asp Ile Pro Gln Gly Val Glu
50 55 60
Thr Asn Ser Ile Ile Gly Asn Ser Val Glu Asn Asp Ser Lys Ser His
65 70 75 80
Pro Ile Leu Phe Arg Tyr Thr Ala Arg His Ala Lys Glu Lys Ile Pro
85 90 95
Glu Val Arg Ile Ser Thr Thr Val Asp Ser Glu Gln Leu Ser Thr Phe
100 105 110
Trp Arg Asp Tyr Val Asp Ile Leu Lys Gly Ser Ser Gln Leu Lys Leu
115 120 125
Gln Ser Glu Thr Lys Lys Val Ser Ser Lys Lys Ser Lys Ala Lys Lys
130 135 140
Lys Arg Gly Lys Gly Ala Trp
145 150

<210> 11
<211> 323
<212> DNA
<213> Komagataella phaffii
<400> 11
atgcagcctc tgatgagtcc gtgaggacga aacgagtaag ctcgtcaggc tgttatggcg 60
catccggggg aggtagttac ttgaccttga ttcctaatag cttacaactg aggtgtctcg 120
ttcgatcctg gcggtccgca atattttcca tacgagtaat ctgtggggga aggcgagcaa 180
taagacgtgc caccgcccaa ggggagcaat ccagcaggga acacgtcccg caaggaggcg 240

ggtgagatag catctcgttg gtaatgggct gttggtgaac aaagtttgac tatgtgaacc 300
ggctatttac atttttgctt ttt 323

<210> 12
<211> 1707
<212> DNA
<213> Komagataella phaffii
<400> 12

atggaatcgc ccttgcaatc tacatacgga gaaagagccg aaaggtattt agatagtgct 60
gatgcttttc ataaacaaag acacagattg aatcgaaggc tgcacaagtt acgtaagagc 120
ttggatattc atgttactga tactaagaac tatagagaga aagagcagat ttccaaaatt 180
gatctagagt cgtacaacag ggataagcga tatggtgaca ttatactgtt cactgcagag 240
agggatcaca tgtatagtga tgaggtcaag gagatcatga aggtccatca tagtaaatcg 300
agagaaaagt ttattgtttc tagattgaag agatcactgg accacggtag aaaattactg 360
atcctagttg gagacgagcc tgatgagatg agaaaattgg aagtatttgt ttatgttgca 420
ttgattcagg gtaaactttc cattgcaaac aagaattgga ccaatgctca gtatgctctc 480
agtgtggcga gatgtgggct ccagtttttg gacaaatatg gtactgaaac acaaactgac 540
ctctataatg gcataattga cactcacata gatcaaatgt tgaaatttgt gatctaccaa 600
gctactaaaa ataacagtcc tattttggat acagagtgca gacatcaaat taggacggac 660
accctagggt atttggatca ggcaaggcaa ataatagaat caaaagatcc cgagtttctg 720
aatgttggag ttgttgaaac tcagttgatt tggtgggact acgatatctc tattcattca 780
gaggaggtag caaagctgat ttcagatgcg aacgaaaagc tgcaacttat cgaggatgga 840
aacgtctcct catatgatcc ggctctacta actcttcaag aagcgctgga tgctcatcag 900
ttgttgatgg ccagaaatgt tgacaacttc gcagacgacg atcaaaacaa tcatgtttta 960
ctgtcgtaca tcagatattt gttacttatc accactttga gaagggacat tactttgata 1020
gaccaagtta gaaacagatc tgtggttaat tcttccctag ctgtggctct ggaacgtgct 1080
aaagacgttg gtagaatttt cgacaatatc gtcaagaaag tcaatgagtt gaaagacgtt 1140
ccaggtgttt acaacaagca agaggagtgg aattcgttgc aggcattgga tgcttatttc 1200
caagcatcca agatccaaca tttggcatct acccaccttt tattcaacag atccaaggaa 1260
tcattggcgt tattaataaa ggcaaagtcc ttggtaaagg ggcacactat cgccggagaa 1320
tatcccacta atttccctac gaataaagat ttgagttcga tcttagaaca aattaatcaa 1380
gacatcctta aggcttatgt tttggccaag tataagcaag agtcctcttt aggtggtgta 1440
tcggagtatg atttcattgc tgacaatcgc aacaaggttc cgtcgaatcc cagtctgcac 1500
aagattgcct ctgtatccta caagaatgtc aaacctgtca atgttaagcc tgtattgttt 1560
gacatagctt tcaactacgt gagtcaacca aaccagatat ttgaggaacc tatcgagagt 1620
tcaaacaagc aagagagaca agcggattct gaatctccgt caccagagaa gaaaaaaaag 1680
ggattgtttg gattgttccg ctagtag 1707

<210> 13
<211> 1821
<212> DNA

<213> Komagataella phaffii

<400> 13

atgtcgtctc tttcagagct ggtatcggaa ttggcaatcc attctgagaa gaggcagtac 60
aaagaagcat atgagaaagc aaagcgcatt atagatttgg gccaccctct tgaccttgac 120
acattgaagc taggtttggt ggcttcgatc aacctggacc aatatcacaa cgcaggtcgc 180
ctcatatcaa aaagtaagga tcatatcgta tatgatggaa tgaaggaatt cttgctatta 240
attggatatg tgtactacaa gaacggagac tcgaagaatt ttgaaactct actaaaggat 300
tcagctttcc aaggaagagc atttgaacac ctcaaagccc aatactatta caagatcggg 360
gagaacgaaa aggcactcaa gatttaccga gagctatcta agaacccatt ggatgaagtt 420
gtagacttga gtgtcaatga aagagctgtc attagccagc ttttggaatt ggatggtgtc 480
gttgaacagc ctgtatctcg accaatagac gactcatacg attgcaaatt caacgatgcc 540
ctttaccagg taaagattgg tgactatgaa tctgcattgg atctcttgga agaagccaaa 600
gccatatgtg aagaaaatac aaaagatctg cctttggaca cacgagaagc agagattgtt 660
cctattctgt tacaaattgc ttacgtcaaa caacttaagg gtaaaaagga agaatccttg 720
actgcgctga gaagtctttc taaaccaaag gactctcttt tagatcttat ttacagaaat 780
aacttactgt cattaaggat tgatgaatac ggaagaaatg ataccaactt tcatattctt 840
tatcgtgagt taggattccc taattcgata gacattaata aagacaagtt gacagtgtcc 900
caaagggttg cgttgaccag aaatgaatca ttattggcac tcgagcttgg aaagatccca 960
tctcaaagtg atctcaagct cttttatgat gctacttcgg aatttttaga tttgaacacc 1020
aagctagaag cctcaatgat ttataattat ttcatgagac gccctggcca gcaagaggtt 1080
ccaaatgctc ttttaactgc acagctggct atcaacgttg gtaacatcaa taatgcaaga 1140
actgttcttg aaactgtggt gtcgaacgac gaaaagaatt tactagaacc atctattgtg 1200
gtatctttgt atttgattta cgataagctt caaagcggaa gattgcaagt tgaactcctg 1260
aagaaagtag cggatctttt gctagaatca gagatttcca gcactcaaca acgtaagttt 1320
ttcaaagata ttgccttcaa aaccctcaat cacgatgcag ttttggccaa tcgactattt 1380
gagaaactgc atagtattta ccctaatgat gagttggtat ccacgtactt gaactcttca 1440
tctaatgcat ccaacaataa cactaccacg accaacttct ccgaattgga cgacttggtc 1500
ctaggcatag acacggataa gcttattagt gaaggatttg atacttttga gtccagcaaa 1560
agaccgacca cgattatcag ctcaactaac aagaaacgtc gtactagatt gaagcccaaa 1620
catgaagcca aggaaaagta taagcgtctg gatgaggaaa gatggttacc tttaaaggac 1680
cgcagttatt acagacccaa aaagggaaag aaaattagaa ataccactca gggtactgtc 1740
actagtaata ctagtgaaat aagtggcttg aaaaagactc tgccaaagaa aagttccaaa 1800
aagaaaggaa gaaaatgata g 1821

<210> 14

<211> 456

<212> DNA

<213> Komagataella phaffii

<400> 14

atgcctcctg tgaaatctct ggacatcttt ttcaaccgca cagagaagct cttagaagcc 60

aaccccacaa cgacaaaagt ttccatcaaa ttgggcgtaa atttcaatga tcacgagaat 120
cctcaaagca agcacaacgt cataacggtg agagtatctg atccagtgag cgggtccaat 180
ttcaaattca aagtgaccaa taaaactgat atgctgaaaa tattcagttt cttaggtcct 240
catggcattg agttaccaat ttctggccag caaagccaga taaagagtaa tgatcagact 300
cagagtgaca atactgaagt gcctaccaca tttcataggg gagccaccag tattttggct 360
aataaggcat ttgagaagaa accactgatt attaaggatt caagtaccgc aaagaaaggt 420
ggtaaaggtg gtaagaagaa gggtaagaaa ttttaa 456

<210> 15
<211> 816
<212> DNA
<213> Komagataella phaffii
<400> 15

atgccattac tagaggaaat aagtgatgca gaggacatag acaacttgga gatggattta 60
gccgagtttg atcctacttt aaggactccg atagctgagc aaagaccagc tcctcaggtt 120
gtcagatcac aagatgccga aagtggacag actcctttgg ttcctaacca ggatcaaata 180
agtcagtata ttgaacaatt caaagaaggt ggcaccataa acaaggatca agtgattaga 240
cccgacgaaa tgatggaaaa agaaatggca gagttgaaaa gcttccaaat tttgtaccca 300
tgttactttg ataaaaatag aagtgttaaa gaaggaagaa gatgccaaaa ggagtatggt 360
gtggagaacc ccctggcaaa gacaatatta gatgcttgca ggtacttgga tataccttgc 420
atcctggagc ctgaaaagac tcatcctcaa gattttggta atccaggaag agtgagagtg 480
gctatcaagg agagtgggaa gtatctcgat gaacaatata agaccaaaag gaaactaata 540
cagttggtag gacaatttct ggttgaacat ccaacaacgt tacagaaagt tcaagaattg 600
cccggtccac ctgagttgca acagggcggg tacattccag aacgtgtacc ccgagtgaaa 660
gggttaaaga tgaacgaaat tgttcctttg cattcgccat tcactattaa gcatccaagt 720
actaaatctg tttatgaaag ggaacctgag cccgcaccac ccgccgccgt gcccaaagct 780
ccgaaacaga agaaaataat ggtgagaaga taatag 816

<210> 16
<211> 1560
<212> DNA
<213> Komagataella phaffii
<400> 16

atggtattgg cagatcttgg aaggcgtatc aataacgccg ttggaaatgt caccaagtcc 60
aatgttgttg acgctgacgt catcagcaac atgttaaagg agatttgtaa cgccctattg 120
gagtccgatg tgaacattaa actagttgcc caattgagag agaaaatacg aaaacagatc 180
gacgcagagg ataaaccagg aattaataag aagaagctga tccagaaggt cgtttttgat 240
gagctggtga aacttgttga ttgcaacgaa gctgagctgt tcaagccaaa gaaaaaacag 300
acgaatgtga tcatgatggt cggtttacaa ggtgctggta agacaacaac ctgtactaaa 360
ctggcagtgt attaccagag aagaggattc aaagtgggaa tggtctgtgg tgacactttc 420
cgagctggtg cgtttgacca gctgaaacaa aacgctacca aggctaagat tccctactat 480

ggttcatata cagaaactga ccctgtgaaa gtgacctttg atggtgtgga agaattcagg 540
aaggaaaagt ttgaaataat aattgtggat acttctggta gacacaggca ggaggaagat 600
ttattcgaag agatggtaca aattggaaaa gctatcaagc ctaatcaaac aatcatggta 660
ctggatgctt ccataggtca atctgccgaa tctcaatcta aagcatttaa ggaatcatcc 720
gattttggtg ccattatcat aactaaaatg gattccaatt ccaagggagg aggtgccctt 780
tcagctatag ctgccaccaa cactccagta gcgtttattg ccaccggaga gcacattcag 840
aatttcgaaa agttttcagg aagaggattt atctcaaaac ttttaggaat tggtgatata 900
gagggtctta tggaacatgt tcagtcgatg aacttggatc aaggtgatac tatcaagaat 960
ttcaaggaag gaaagtttac tttacaggat tttcaaacgc aattgaacaa catcatgaag 1020
atggggccac tgtccaaact cgctcaaatg ttgcctggtg gaatgggaca attgatggga 1080
caggttggtg aagaggaggc ttcaaagaga ttgaagcgaa tgatttatat aatggattca 1140
atgacgaagc aagagttgtc aagtgacggt agattgttta ttgatcagcc ttcaaggatg 1200
gtaagagttg ctagaggctc tggtacctct gtaactgagg tggagcttgt tcttttacag 1260
caaaagatga tggctcgtat ggcattacaa tctaagaata tgatgagtgg ggccggcggt 1320
ccagcaggga tggcttccaa aatgaatcca gctaatatga gaagagctat gcaacaaatg 1380
caatcaaacc caggaatgat ggataacatg atgaatatgt ttggtggagc tggaggagct 1440
ggaggagctg gaatgccgga tatgcaagaa atgatgaagc aaatgtccag tggccaaatg 1500
aaaatgccca gtcaacagga aatgatgagc atgatgaaac agtttggtat gggctaatag 1560

<210> 17
<211> 456
<212> DNA
<213> Komagataella phaffii
<400> 17

atgtccacaa ctactaagaa aaacaagaac aggatcttga tagagaatca caaacagttc 60
ctggaagaag tttccaaaac agccacttta tcagtttgga actcgaaatt ttcaatcaaa 120
cgactgtctc tagaagcaga tcccgtggaa gggacgcctg aaggaatcag agatatccca 180
caaggagtag agacaaactc tataatagga aacagcgtag agaatgattc aaagtcacac 240
cccattttat tcagatatac agctagacac gcaaaggaga aaataccaga ggtgcgaatt 300
tcaaccaccg ttgattcaga gcagctaagc accttctgga gagattatgt ggacatattg 360
aagggaagct cccaattgaa actgcagtca gaaaccaaga aagtcagtag taagaagagc 420
aaggcgaaga agaagagagg aaagggtgca tggtaa 456

<210> 18
<211> 54
<212> DNA
<213> Komagataella phaffii
<400> 18

atgaagctga tctccgtggg tatagtgacg acattactga ctttggccag ttgc 54

<210> 19
<211> 54

<212> DNA
<213> Komagataella phaffii
<400> 19
atgttaaaca agctgttcat tgcaatactc atagtcatca ctgctgtcat aggc 54
<210> 20
<211> 198
<212> DNA
<213> 人工
<220>
<223> 包含 L23S, D64E的α-MF的前导序列
<400> 20
gcccctgtta acactaccac tgaagacgag actgctcaaa ttccagctga agcagttatc 60
ggttactctg accttgaggg tgatttcgac gtcgctgttt tgcctttctc taactccact 120
aacaacggtt tgttgttcat taacaccact atcgcttcca ttgctgctaa ggaagagggt 180
gtctctctcg agaagaga 198
<210> 21
<211> 255
<212> DNA
<213> 人工
<220>
<223> 包含L23S, D64E的α-MF (前原前导序列)
<400> 21
atgagattcc catctatttt caccgctgtc ttgttcgctg cctcctctgc attggctgcc 60
cctgttaaca ctaccactga agacgagact gctcaaattc agctgaagc agttatcggt 120
tactctgacc ttgagggtga tttcgacgtc gctgttttgc ctttctctaa ctccactaac 180
aacggtttgt tgttcattaa caccactatc gcttccattg ctgctaagga agagggtgtc 240
tctctcgaga agaga 255
<210> 22
<211> 837
<212> DNA
<213> 人工
<220>
<223> vHH蛋白
<400> 22
caggttcagc tgcaggagtc cggtggtggt ctggttcaag ccggtggttc attaagattg 60
tcctgtgctg cctctggtag aactttcact tctttcgcaa tgggttggtt tagacaagca 120
cctggaaaag agagagagtt tgttgcttct atctccagat ccggtacttt aactagatac 180
gctgactctg ccaagggtag attcactatt tctgttgaca acgccaagaa cactgtttct 240
ttgcaaatgg acaaccttaa cccagatgac accgcagtct attactgtgc cgctgacttg 300

cacagaccat acggtccagg aacccaaaga tccgatgagt acgattcttg gggtcaggga 360
actcaagtca ctgtctcttc aggtggtgga tctggtggtg gaggttcagg tggtggagga 420
tccggtggtg gtggttctgg tggtggtgga tctggtggag gtgaagttca acttgtcgaa 480
tccggtggtg cacttgtcca acctggtgga tctcttagac tttcttgtgc cgcctccggt 540
tttcctgtta accgttactc tatgcgttgg tacagacaag cccctggaaa agaacgtgaa 600
tgggttgccg gaatgtcctc agctggtgac agatcctcct acgaagattc tgtgaaggga 660
cgtttcacca tctccagaga tgacgcccgt aacaccgttt accttcaaat gaactccctt 720
aagcctgagg atactgccgt ctactattgt aacgtgaatg tcggatttga atactgggga 780
cagggaaccc aagttactgt ctcttccggt ggacatcacc accaccatca ctaatag 837

<210> 23
<211> 774
<212> DNA
<213> 人工
<220>
<223> scR蛋白
<400> 23

caggaacaac taatggagtc tgggggtggt ttggttaccc tgggtggttc tcttaagctt 60
tcatgtaagg cctctggtat tgatttttcg cactacggta tctcctgggt tagacaagct 120
cctggaaaag gtctggaatg gatcgcttac atttacccaa attacggttc tgttgactat 180
gcctcctggg tcaatggtag gttcactatt tcccttgaca acgctcagaa cacggtattc 240
ctacagatga tctccctaac cgctgctgat actgcaacct acttctgtgc tcgtgacaga 300
ggttactact ctggctctcg tggaactaga cttgacttat ggggacaagg tactctcgtt 360
accatctcta gtggtggagg tggttctgga ggaggaggtt ccggcggagg tggtagcgag 420
ctggtcatga ctcaaacccc tccatcccta tctgcatcag tcggtgaaac cgttagaatt 480
agatgccttg catctgagtt cttgttcaac ggtgtgtcct ggtatcaaca aaagcctggt 540
aagcctccaa agtttctcat ttctggtgcc tcaaacctcg aatctggagt gccaccaaga 600
ttttccggat ctggctctgg tactgactac actctgacaa ttggtggtgt tcaagctgag 660
gatgttgcta cctactattg tctcggtggt tactcaggat cttccggcct aactttcggt 720
gccggtacaa acgtcgagat caaaggtgga catcaccacc accatcacta atag 774

<210> 24
<211> 687
<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-HC
<400> 24

gaggtccaat tggtccaatc tggtggagga ttggttcaac caggtggatc tctgagattg 60
tcttgtgctg cttctggttt caccttctct cactactgga tgtcatgggt tagacaagct 120
cctggtaagg gtttggaatg ggttgctaac atcgagcaag atggatcaga gaagtactac 180

gttgactctg ttaagggaag attcactatt tcccgtgata acgccaagaa ctccttgtac 240
ctgcaaatga actcccttag agctgaggat actgctgtct acttctgtgc tagagacttg 300
gaaggtttgc atggtgatgg ttacttcgac ttatggggta gaggtactct tgtcaccgtt 360
tcatctgcct ctaccaaagg accttctgtg ttcccattag ctccatgttc cagatccacc 420
tccgaatcta ctgcagcttt gggttgtttg gtgaaggact actttcctga accagtgact 480
gtctcttgga actctggtgc tttgacttct ggtgttcaca cctttcctgc agttttgcag 540
tcatctggtc tgtactctct gtcctcagtt gtcactgttc cttcctcatc tcttggtacc 600
aagacctaca cttgcaacgt tgaccataag ccatccaata ccaaggttga caagagagtt 660
gagtccaagt atggtccacc ttaatag 687

<210> 25
<211> 648
<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-LC
<400> 25

gctatccagt tgactcaatc accatcctct ttgtctgctt ctgttggtga tagagtcatc 60
ctgacttgtc gtgcatctca aggtgtttcc tcagctttag cttggtacca acaaaagcca 120
ggtaaagctc caaagttgct gatctacgac gcttcatccc ttgaatctgg tgttccttca 180
cgtttctctg gatctggatc aggtcctgat ttcactctga ctatctcatc ccttcaacca 240
gaggactttg ctacctactt ctgtcaacag ttcaactctt acccctttgac ctttggaggt 300
ggaactaagt tggagatcaa gagaactgtt gctgcaccat cagtgttcat ctttcctcca 360
tctgatgagc aactgaagtc tggtactgca tctgttgtct gcttactgaa caacttctac 420
ccaagagaag ctaaggtcca atggaaggtt gacaatgcct tgcaatctgg taactctcaa 480
gagtctgtta ctgagcaaga ctctaaggac tctacttact ccctttcttc caccttgact 540
ttgtctaagg ctgattacga gaagcacaag gtttacgctt gtgaggttac tcaccaaggt 600
ttgtcctctc ctgttaccaa gtctttcaac agaggtgaat gctaatag 648

<210> 26
<211> 277
<212> PRT
<213> 人工
<220>
<223> vHH蛋白
<400> 26

Gln Val Gln Leu Gln Glu Ser Gly Gly Gly Leu Val Gln Ala Gly Gly
1 5 10 15
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Arg Thr Phe Thr Ser Phe
20 25 30
Ala Met Gly Trp Phe Arg Gln Ala Pro Gly Lys Glu Arg Glu Phe Val

35 40 45

Ala Ser Ile Ser Arg Ser Gly Thr Leu Thr Arg Tyr Ala Asp Ser Ala
50 55 60
Lys Gly Arg Phe Thr Ile Ser Val Asp Asn Ala Lys Asn Thr Val Ser
65 70 75 80
Leu Gln Met Asp Asn Leu Asn Pro Asp Asp Thr Ala Val Tyr Tyr Cys
85 90 95
Ala Ala Asp Leu His Arg Pro Tyr Gly Pro Gly Thr Gln Arg Ser Asp
100 105 110
Glu Tyr Asp Ser Trp Gly Gln Gly Thr Gln Val Thr Val Ser Ser Gly
115 120 125
Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
130 135 140
Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Glu Val Gln Leu Val Glu
145 150 155 160
Ser Gly Gly Ala Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys
165 170 175
Ala Ala Ser Gly Phe Pro Val Asn Arg Tyr Ser Met Arg Trp Tyr Arg
180 185 190
Gln Ala Pro Gly Lys Glu Arg Glu Trp Val Ala Gly Met Ser Ser Ala
195 200 205
Gly Asp Arg Ser Ser Tyr Glu Asp Ser Val Lys Gly Arg Phe Thr Ile
210 215 220
Ser Arg Asp Asp Ala Arg Asn Thr Val Tyr Leu Gln Met Asn Ser Leu
225 230 235 240
Lys Pro Glu Asp Thr Ala Val Tyr Tyr Cys Asn Val Asn Val Gly Phe
245 250 255
Glu Tyr Trp Gly Gln Gly Thr Gln Val Thr Val Ser Ser Gly Gly His
260 265 270
His His His His His
275

<210> 27
<211> 256
<212> PRT
<213> 人工
<220>
<223> scR蛋白
<400> 27
Gln Glu Gln Leu Met Glu Ser Gly Gly Gly Leu Val Thr Leu Gly Gly

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

<210> 28
<211> 227
<212> PRT
<213> 人工
<220>
<223> SDZ-Fab-HC
<400> 28
Glu Val Gln Leu Val Gln Ser Gly Gly Gly Leu Val Gln Pro Gly Gly

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

<210> 29
<211> 214
<212> PRT
<213> 人工
<220>
<223> SDZ-Fab-LC
<400> 29

Ala Ile Gln Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15
Asp Arg Val Ile Leu Thr Cys Arg Ala Ser Gln Gly Val Ser Ser Ala

20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

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

Ser Gly Ser Gly Pro Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Phe Cys Gln Gln Phe Asn Ser Tyr Pro Leu
85 90 95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys
210

<210> 30
<211> 19
<212> DNA
<213> 人工
<220>
<223> SRP14 fwd
<400> 30
agacacgcaa aggagaaaa 19
<210> 31
<211> 22
<212> DNA
<213> 人工
<220>
<223> SRP14 rev

<400> 31
gtccacataa tctctccaga ag 22
<210> 32
<211> 19
<212> DNA
<213> 人工
<220>
<223> SRP21 fwd
<400> 32
ggggagccac cagtatttt 19
<210> 33
<211> 21
<212> DNA
<213> 人工
<220>
<223> SRP21 rev
<400> 33
ccacctttac cacctttctt t 21
<210> 34
<211> 19
<212> DNA
<213> 人工
<220>
<223> Sec65 fwd
<400> 34
ttcctttgca ttcgccatt 19
<210> 35
<211> 21
<212> DNA
<213> 人工
<220>
<223> SEC65 rev
<400> 35
ttcttctgtt tcggagcttt g 21
<210> 36
<211> 19
<212> DNA
<213> 人工
<220>

<223> SRP54 fwd
<400> 36
aagatgatgg ctcgtatgg 19
<210> 37
<211> 20
<212> DNA
<213> 人工
<220>
<223> SRP54 rev
<400> 37
tcctgggttt gattgcattt 20
<210> 38
<211> 20
<212> DNA
<213> 人工
<220>
<223> SRP68 fwd
<400> 38
caactacgtg agtcaaccaa 20
<210> 39
<211> 19
<212> DNA
<213> 人工
<220>
<223> SRP68 rev
<400> 39
agcggaacaa tccaaacaa 19
<210> 40
<211> 20
<212> DNA
<213> 人工
<220>
<223> SRP72 fwd
<400> 40
tgccttcaaa accctcaatc 20
<210> 41
<211> 20
<212> DNA
<213> 人工

<220>
<223> SRP72 rev
<400> 41
ggtcgtggta gtgttattgt 20
<210> 42
<211> 18
<212> DNA
<213> 人工
<220>
<223> scpRNA fwd
<400> 42
gggaaggcga gcaataag 18
<210> 43
<211> 18
<212> DNA
<213> 人工
<220>
<223> scpRNA rev
<400> 43
accaacagcc cattacca 18
<210> 44
<211> 22
<212> DNA
<213> 人工
<220>
<223> scR fwd
<400> 44
aagcctggta agcctccaaa gt 22
<210> 45
<211> 23
<212> DNA
<213> 人工
<220>
<223> scR rev
<400> 45
tcctcagctt gaacaccacc aat 23
<210> 46
<211> 22
<212> DNA

<213> 人工
<220>
<223> vHH fwd
<400> 46
tgtaacgtga atgtcggatt tg 22
<210> 47
<211> 20
<212> DNA
<213> 人工
<220>
<223> vHH rev
<400> 47
tagtgatggt ggtggtgatg 20
<210> 48
<211> 24
<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-HC fwd
<400> 48
tactgctgct ttgggttgtt tggt 24
<210> 49
<211> 24
<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-HC rev
<400> 49
aagggacagt aacaacagag gaca 24
<210> 50
<211> 23
<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-LC fwd
<400> 50
gatgaacaat tgaagtctgg tac 23
<210> 51
<211> 23

<212> DNA
<213> 人工
<220>
<223> SDZ-Fab-LC rev
<400> 51
gagtaacttc acaagcgtaa acc 23
<210> 52
<211> 18
<212> PRT
<213> Komagataella pastoris
<400> 52
Met Lys Leu Phe Phe Val Gly Ile Val Thr Thr Leu Leu Thr Leu Val
1 5 10 15
Ser Cys
<210> 53
<211> 66
<212> PRT
<213> 酿酒酵母
<400> 53
Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala
1 5 10 15
Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala
20 25 30
Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn
35 40 45
Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Asp
50 55 60
Lys Arg
65
<210> 54
<211> 741
<212> DNA
<213> 人工
<220>
<223> mCherry
<400> 54
gtgagcaagg gcgaggagga taacatggcc atcatcaagg agttcatgcg cttcaaggtg 60
cacatggagg gctccgtgaa cggccacgag ttcgagatcg agggcgaggg cgagggccgc 120
ccctacgagg gcacccagac cgccaagctg aaggtgacca agggtggccc cctgcccttc 180

gcctgggaca tcctgtcccc tcagttcatg tacggctcca aggcctacgt gaagcacccc 240
gccgacatcc ccgactactt gaagctgtcc ttccccgagg gcttcaagtg ggagcgcgtg 300
atgaacttcg aggacggcgg cgtggtgacc gtgacccagg actcctccct gcaggacggc 360
gagttcatct acaaggtgaa gctgcgcggc accaacttcc cctccgacgg ccccgtaatg 420
cagaaaaaga ccatgggctg ggaggcctcc tccgagcgga tgtaccccga ggacggcgcc 480
ctgaagggcg agatcaagca gaggctgaag ctgaaggacg gcggccacta cgacgctgag 540
gtcaagacca cctacaaggc caagaagccc gtgcagctgc ccggcgccta caacgtcaac 600
atcaagttgg acatcacctc ccacaacgag gactacacca tcgtggaaca gtacgaacgc 660
gccgagggcc gccactccac cggcggcatg gacgagctgt acaagggtgg agattacaag 720
gatgacgatg ataagtaata g 741

<210> 55
<211> 245
<212> PRT
<213> 人工
<220>
<223> mCherry
<400> 55

Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe Met
1 5 10 15
Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe Glu
20 25 30
Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala
35 40 45
Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp Ile
50 55 60
Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His Pro
65 70 75 80
Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe Lys
85 90 95
Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr
100 105 110
Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys Leu
115 120 125
Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys Thr
130 135 140
Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly Ala
145 150 155 160
Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly His
165 170 175

Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val Gln
180 185 190
Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser His
195 200 205
Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly Arg
210 215 220
His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Gly Asp Tyr Lys
225 230 235 240
Asp Asp Asp Asp Lys
245

<210> 56
<211> 16
<212> PRT
<213> Komagataella phaffii
<400> 56
Met Leu Val Ala Trp Phe Leu Leu Leu Leu Val Ser Ser Cys Ile Cys
1 5 10 15

<210> 57
<211> 19
<212> PRT
<213> Komagataella phaffii
<400> 57
Met Lys Phe Ala Ile Ser Thr Leu Leu Ile Ile Leu Gln Ala Ala Ala
1 5 10 15
Val Phe Ala

<210> 58
<211> 20
<212> PRT
<213> Komagataella phaffii
<400> 58
Met Lys Phe Gly Leu Gly Ser Leu Gly Leu Ala Val Ala Leu Ile Pro
1 5 10 15
Ile Ala Ser Ala
20

<210> 59
<211> 18
<212> PRT
<213> Komagataella phaffii
<400> 59

Met Ile Ile Leu Leu Pro Leu Leu Phe Leu Phe Val Ala Gly Leu Val
1 5 10 15
Gln Ala
<210> 60
<211> 16
<212> PRT
<213> Komagataella phaffii
<400> 60
Met Asp Pro Phe Ser Ile Leu Leu Thr Leu Thr Leu Ile Ile Leu Ala
1 5 10 15
<210> 61
<211> 24
<212> PRT
<213> Komagataella phaffii
<400> 61
Met Arg Leu Ser Tyr Glu Cys Leu Phe Ser Val Phe Leu Val Leu Ala
1 5 10 15
Tyr His Leu Lys Gly Thr Lys Ala
20
<210> 62
<211> 20
<212> PRT
<213> Komagataella phaffii
<400> 62
Met Ile Asn Leu Asn Ser Phe Leu Ile Leu Thr Val Thr Leu Leu Ser
1 5 10 15
Pro Ala Leu Ala
20
<210> 63
<211> 18
<212> PRT
<213> Komagataella phaffii
<400> 63
Met Gln Leu Gln Tyr Leu Ala Val Leu Cys Ala Leu Leu Leu Asn Val
1 5 10 15
Gln Ser
<210> 64
<211> 20
<212> PRT

<213> Komagataella phaffii
<400> 64
Met Trp Ile Glu Arg Asn Leu Ile Ala Ser Ile Leu Leu Phe Ser Thr
1 5 10 15
Ser Ala Tyr Ala
20
<210> 65
<211> 25
<212> PRT
<213> Komagataella phaffii
<400> 65
Met Asn Ile Ser Thr Ala Ser Lys Ile Ser Arg Leu Leu Gln Leu Val
1 5 10 15
Ile Ala Leu Ile Ser Leu Val Leu Thr
20 25
<210> 66
<211> 20
<212> PRT
<213> Komagataella phaffii
<400> 66
Met Phe Val Phe Glu Pro Val Leu Leu Ala Val Leu Val Ala Ser Thr
1 5 10 15
Cys Val Thr Ala
20
<210> 67
<211> 22
<212> DNA
<213> 人工
<220>
<223> mCherry 正向
<400> 67
catcaagttg gacatcacct cc 22
<210> 68
<211> 20
<212> DNA
<213> 人工
<220>
<223> mCherrey 反向
<400> 68

caccettgta cagctcgtcc 20

<210> 69
<211> 100
<212> PRT
<213> Komagataella phaffii
<400> 69

Leu Leu Ile Pro Ser Leu Asp Gln Leu Asn Ile Gln Leu Pro Phe Ser
1 5 10 15
Leu Pro His His Thr Glu Ser Pro Ser Leu Lys Leu Gln Gly Ser Asn
20 25 30
Pro Phe Glu Ser Ser Thr Val Arg Pro Asp Pro Ile Gln Ile Tyr Ser
35 40 45
Thr Gly Tyr Lys Val Ile Glu Asn Ser Tyr Ile Val Thr Val Asp Ser
50 55 60
Ser Ile Thr Asp Ser Glu Leu Gln Gln Leu Tyr Asp Tyr Ile Lys Gly
65 70 75 80
Gly Tyr Glu Phe Met Leu Asn Asn Glu Asp Pro Phe Phe Val Ala Met
85 90 95
Gly Ile Lys Arg
100

<210> 70
<211> 62
<212> PRT
<213> Komagataella phaffii
<400> 70

Cys Leu Gln Leu Leu Thr Thr Ser Ile Pro Pro Ser Phe Leu Thr Met
1 5 10 15
Val Pro Glu His Tyr Ile Gly Ser Lys Ser Val Asp Glu Val Pro Thr
20 25 30
Ser Glu Asp Pro Arg Val Asn Ala Cys Pro Tyr Ile Cys Asp Ile Gly
35 40 45
Asp Cys Ser Arg Gly Tyr Ser Arg Ala Ser His Leu Lys Arg
50 55 60

<210> 71
<211> 13
<212> PRT
<213> Komagataella phaffii
<400> 71

Arg Pro Leu Glu His Ala His His Gln His Asp Lys Arg

1 5 10

<210> 72

<211> 8

<212> PRT

<213> Komagataella phaffii

<400> 72

Tyr Pro Leu Val Ile Lys Lys Arg

1 5

<210> 73

<211> 2

<212> PRT

<213> Komagataella phaffii

<400> 73

Lys Arg

1

<210> 74

<211> 66

<212> PRT

<213> 奇异酵母

<400> 74

Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala
1 5 10 15

Glu Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala
20 25 30

Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn
35 40 45

Thr Thr Ile Ala Asn Ile Ala Ala Glu Glu Glu Gly Val Thr Leu Asn
50 55 60

Lys Arg
65

<210> 75

<211> 66

<212> PRT

<213> 奇异酵母

<400> 75

Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala
1 5 10 15

Glu Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Ile Ala
20 25 30

Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn
35 40 45
Thr Thr Ile Ala Asn Ile Ala Ala Glu Glu Glu Gly Val Thr Leu Asn
50 55 60
Lys Arg
65

<210> 76
<211> 65
<212> PRT
<213> 奇异酵母
<400> 76

Pro Val Asn Thr Thr Thr Glu Asp Glu Met Ala Gln Ile Pro Ala Glu
1 5 10 15
Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala Val
20 25 30
Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn Thr
35 40 45
Thr Ile Ala Asn Ile Ala Ala Glu Glu Glu Gly Val Thr Leu Asn Lys
50 55 60
Arg
65

<210> 77
<211> 65
<212> PRT
<213> 奇异酵母
<400> 77

Pro Val Asn Thr Thr Thr Glu Asp Glu Met Ala Gln Ile Pro Ala Glu
1 5 10 15
Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala Val
20 25 30
Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn Thr
35 40 45
Thr Ile Ala Asn Ile Ala Ala Glu Glu Glu Gly Val Thr Leu Asn Lys
50 55 60
Arg
65

<210> 78
<211> 65
<212> PRT

<213> 奇异酵母
<400> 78
Pro Val Asn Thr Thr Thr Glu Asp Glu Met Ala Arg Ile Pro Ala Glu
1 5 10 15
Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala Val
20 25 30
Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn Thr
35 40 45
Thr Ile Ala Asn Ile Ala Ala Glu Glu Glu Gly Val Thr Leu Asn Lys
50 55 60
Arg
65
<210> 79
<211> 66
<212> PRT
<213> 真贝酵母
<400> 79
Ala Pro Val Asn Thr Thr Thr Glu Asn Glu Thr Ala Gln Ile Pro Ala
1 5 10 15
Glu Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Ser Asp Val Ala
20 25 30
Val Leu Pro Phe Ala Asn Ser Thr Asn Thr Gly Leu Leu Phe Ile Asn
35 40 45
Thr Thr Ile Ser Asn Leu Ala Ala Lys Glu Glu Gly Val Ser Leu Ser
50 55 60
Lys Arg
65
<210> 80
<211> 66
<212> PRT
<213> 库德里阿兹威酵母
<400> 80
Ala Pro Val Asn Thr Thr Ser Glu Ser Glu Thr Val Gln Ile Pro Ala
1 5 10 15
Glu Ala Ile Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe Asp Val Ala
20 25 30
Val Leu Pro Phe Ala Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn
35 40 45
Thr Thr Ile Ala Asn Leu Ala Thr Lys Glu Glu Ser Val Pro Leu Ser

50 55 60

Lys Arg

65

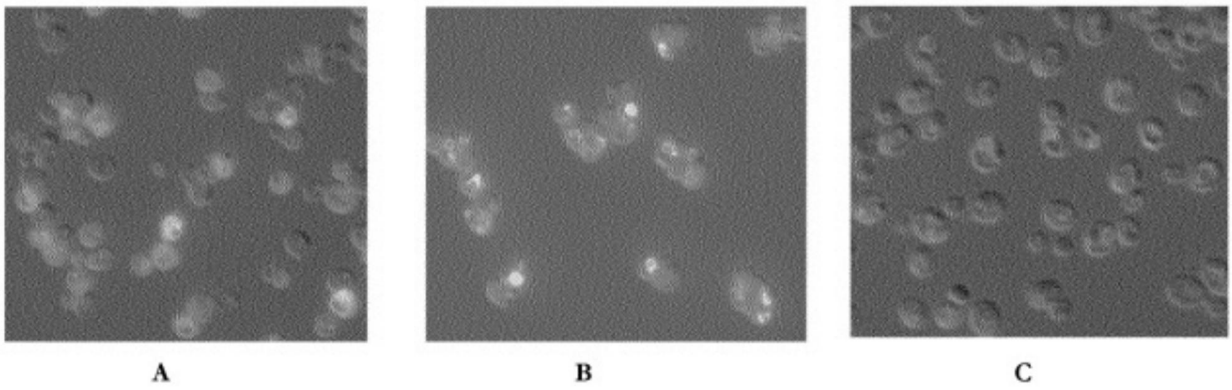


图1