(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107406514 B
(45)授权公告日 2021.11.30

(21)申请号 201580059731.6
(22)申请日 2015.10.23
(65)同一申请的已公布的文献号
申请公布号 CN 107406514 A
(43)申请公布日 2017.11.28
(30)优先权数据
62/074,494 2014.11.03 US
62/209,010 2015.08.24 US
(85)PCT国际申请进入国家阶段日
2017.05.03
(86)PCT国际申请的申请数据
PCT/US2015/057125 2015.10.23
(87)PCT国际申请的公布数据
WO2016/073228 EN 2016.05.12
(73)专利权人 默克专利有限公司
地址 德国达姆施塔特
(72)发明人 M·齐尔曼 J·奥兰多
(74)专利代理机构 北京高文律师事务所 11359
代理人 徐江华
(51)Int.Cl.
C07K 19/00(2006.01)
C12N 9/50(2006.01)
C12N 9/00(2006.01)
B01D 15/38(2006.01)
C07K 1/22(2006.01)
(56)对比文件
CN 104053779 A,2014.09.28
Zhanga A等.Construction of a mini-intein fusion system to allow both direct monitoring of soluble protein expression and rapid purification of target proteins.《Gene》.2001,第275卷(第2期),第241-252页.
Garbe D等.Protein trans-splicing on an M13 bacteriophage: towards directed evolution of a semisynthetic split intein by phage display.《Journal of Peptide ence》.2010,第16卷(第10期),第575-581页.
审查员 陈晋
权利要求书3页 说明书29页
序列表24页 附图7页

(54)发明名称

可溶内含肽融合蛋白以及用于纯化生物分子的方法

(57)摘要

本发明涉及包括N-内含肽多肽和N-内含肽溶解配偶体的融合蛋白,和包括此类融合蛋白的亲和色谱基质,及其使用方法。

CN 107406514 B

1.一种可溶融合蛋白,包括由肽键连接的N-内含肽多肽和N-内含肽溶解配偶体,所述N-内含肽多肽能够通过与包括C-内含肽多肽的融合蛋白相缔合来形成有活性的内含肽复合体,其中所述N-内含肽多肽包括选自SEQ ID NO:1-8,29-35和42-53的氨基酸序列,所述C-内含肽多肽包括选自SEQ ID NO:9-10,36-41和54-60的氨基酸序列,所述N-内含肽溶解配偶体选自溶解配偶体46、206和246中的至少一种,并且所述N-内含肽多肽共价附接至固体支持物;

其中所述溶解配偶体46的氨基酸序列为:

MREYPNGEKTHLTVMAAGFPSLTGDHKVIYVAADRHVTSEEILEAAIRLLS

所述溶解配偶体206的氨基酸序列为:

MSHLDEVIARVDAAIEESVIAHMNELLIALSDDAELSREDRYTQQQRLRTAIAHHGRKHKEDMEARHEQLTKGGTIL

所述溶解配偶体246的氨基酸序列为:

MNKETQPIDRETLLKEANKIIREHEDTLAGIEATGVTQRNGVLVFTGDYFLDEQGLPTAKSTAVFNMFKHLAHVLSEKYHLVD。

2.如权利要求1所述的融合蛋白,其中该N-内含肽多肽是氨基酸序列SEQ ID NO:1的GP41-1 N-内含肽、或GP41-1 N-内含肽的变体,该GP41-1 N-内含肽的变体为选自由氨基酸序列SEQ ID NO:2的GP41-1 NINTΔA_CC N-内含肽变体、氨基酸序列SEQ ID NO:3的GP41-1 NINTΔA_AC N-内含肽变体、氨基酸序列SEQ ID NO:4的GP41-1NINTΔA_CK N-内含肽变体、氨基酸序列SEQ ID NO:5的GP41-1 NINTΔA_AM N-内含肽变体、氨基酸序列SEQ ID NO:6的GP41-1 NINTΔA_TM N-内含肽变体、氨基酸序列SEQ ID NO:7的GP41-1 NINTΔA_AK N-内含肽变体、氨基酸序列SEQ ID NO:8的GP41-1 NINTΔA_TK N-内含肽变体、氨基酸序列SEQ ID NO:9的GP41-1 C-内含肽或氨基酸序列SEQ ID NO:10的GP41-1 C-内含肽-硫氧还蛋白融合蛋白组成的组中的一种。

3.如权利要求1所述的融合蛋白,其中该N-内含肽多肽缺少用于将该融合蛋白附接至固体支持物的氨基酸。

4.如权利要求3所述的融合蛋白,其中用于将该融合蛋白附接至固体支持物的氨基酸是半胱氨酸。

5.如权利要求1-4中任一项所述的融合蛋白,其中N-内含肽溶解配偶体被连接至该N-内含肽多肽的N末端。

6.如权利要求1-4中任一项所述的融合蛋白,其中N-内含肽溶解配偶体被连接至该N-内含肽多肽的C末端。

7.如权利要求1所述的融合蛋白,其中在按质量计小于25%的所产生的融合蛋白存在于包涵体中的条件下,该融合蛋白在大肠杆菌中产生。

8.如权利要求1所述的融合蛋白,其中修饰该融合蛋白以包括一个可检测标签。

9.如权利要求8所述的融合蛋白,其中该可检测标签是荧光染料。

10.如权利要求1所述的融合蛋白,其中该融合蛋白在大肠杆菌中进行表达。

11.一种编码如权利要求1-10所述的融合蛋白的核酸。

12.如权利要求11所述的核酸,其中该核酸包括在表达载体中。

13.如权利要求11所述的核酸,其中该核酸包括在质粒中。

14.一种宿主细胞,包括如权利要求11、12、或13所述的核酸。

15.如权利要求14所述的宿主细胞,其中该宿主细胞是大肠杆菌。

16.一种亲和色谱基质,包括附接至固体支持物的如权利要求1-10中任一项所述的融合蛋白。

17.如权利要求16所述的亲和色谱基质,其中该固体支持物是色谱树脂。

18.如权利要求17所述的亲和色谱基质,其中该色谱树脂包括亲水性聚乙烯醚基。

19.如权利要求16所述的亲和色谱基质,其中该固体支持物是珠粒、中空纤维、实心纤维、衬垫、凝胶、膜、盒、柱、薄片、载片或平板。

20.如权利要求19所述的亲和色谱基质,其中该固体支持物是磁珠。

21.如权利要求16所述的亲和色谱基质,其中该固体支持物包括可控孔度玻璃、硅石、氧化锆、氧化钛、琼脂糖、聚甲基丙烯酸酯、聚丙烯酸酯、聚丙烯酰胺、聚乙烯醇、聚苯乙烯或其衍生物。

22.如权利要求16所述的亲和色谱基质,其中该基质进一步包括在该融合蛋白和固体支持物之间的间隔分子。

23.如权利要求16所述的亲和色谱基质,其中该融合蛋白在N-内含肽溶解配偶体中的单个位点处附接至该固体支持物。

24.如权利要求16所述的亲和色谱基质,其中该融合蛋白在N-内含肽溶解配偶体中的多于一个位点处附接至该固体支持物。

25.如权利要求16所述的亲和色谱基质,其中当该融合蛋白附接至该固体支持物时,在该融合蛋白中的N-内含肽多肽保持活性。

26.如权利要求16-25中任一项所述的亲和色谱基质,其中当该融合蛋白附接至该固体支持物时,在该融合蛋白中的N-内含肽多肽取向为远离该固体支持物。

27.一种亲和纯化样品中靶分子的方法,该方法包括:

a)提供包含第一融合蛋白的样品,该第一融合蛋白包括通过肽键连接至靶分子的C-内含肽多肽;

b)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成一种无活性的内含肽复合体的条件下,将该样品与包括第二融合蛋白的亲和色谱基质接触,该第二融合蛋白为前述权利要求1-10中任意所述的融合蛋白;

c)洗涤该包含无活性的内含肽复合体的亲和色谱基质,以去除未结合的污染物;

d)将该内含肽复合体暴露在该内含肽复合体具有活性并且从C-内含肽多肽裂解靶分子的条件下;并且

e)回收该经裂解的靶分子。

28.如权利要求27所述的方法,其中该样品是粗蛋白制剂。

29.如权利要求27或28所述的方法,其中该靶分子是治疗靶标的单克隆抗体。

30.如权利要求27或28所述的方法,其中该靶分子是治疗靶标。

31.如权利要求28所述的方法,其中该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成一种无催化活性的内含肽复合体的条件包括:

a)温度处于4℃-25℃的范围内,并且缓冲液包括100mM Tris-HCl、25mM NaCl、0.1mM氯

化锌、pH＝9；

b)温度处于4℃-25℃的范围内，并且缓冲液包括50mM NaAc、0.5MNaCl、pH＝5；或

c)温度处于4℃-25℃的范围内，并且缓冲液包括0.5M NaCl、10mM Tris-HCl、pH＝8。

32.如权利要求28所述的方法，其中促进内含肽复合体的催化活性的这些条件包括：

a)缓冲液包括50mM Tris-HCl、pH＝7.0、300mM NaCl、1mM EDTA；

b)缓冲液包括0.3M L-精氨酸、5mM EDTA、50mM磷酸缓冲液、pH＝6.5；或

c)缓冲液包括0.5M NaCl、10mM Tris-HCl、50mM DTT、pH＝8.0。

33.如权利要求28所述的方法，其中该N-内含肽多肽是氨基酸序列SEQ ID NO:1的GP41-1 N-内含肽或GP41-1 N-内含肽的变体，该GP41-1 N-内含肽的变体为选自由氨基酸序列SEQ ID NO:2的GP41-1 NINTΔA_CC N-内含肽变体、氨基酸序列SEQ ID NO:3的GP41-1 NINTΔA_AC N-内含肽变体、氨基酸序列SEQ ID NO:4的GP41-1NINTΔA_CK N-内含肽变体、氨基酸序列SEQ ID NO:5的GP41-1 NINTΔA_AM N-内含肽变体、氨基酸序列SEQ ID NO:6的GP41-1 NINTΔA_TM N-内含肽变体、氨基酸序列SEQ ID NO:7的GP41-1 NINTΔA_AK N-内含肽变体、氨基酸序列SEQ ID NO:8的GP41-1 NINTΔA_TK N-内含肽变体、氨基酸序列SEQ ID NO:9的GP41-1 C-内含肽或氨基酸序列SEQ ID NO:10的GP41-1 C-内含肽-硫氧还蛋白融合蛋白组成的组中的一种。

34.如权利要求28所述的方法，进一步包括清洁、再生或存储该亲和色谱基质以用于随后使用。

35.一种筛选适合用于亲和纯化的内含肽复合体的方法，该方法包括：

a)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成内含肽复合体的条件下，将包括通过肽键连接至靶分子的C-内含肽多肽的第一融合蛋白与第二融合蛋白接触，该第二融合蛋白为前述权利要求1-10中任意所述的融合蛋白；并且

b)在支持内含肽活性的条件下，确定该靶分子是否从该C-内含肽多肽裂解下来，其中该经裂解的靶分子的存在表明有催化活性的内含肽复合体存在。

可溶内含肽融合蛋白以及用于纯化生物分子的方法

[0001] 相关申请

[0002] 本申请要求于2014年11月3日提交的美国临时申请号62/074,494,以及于2015年8月24日提交的美国临时申请号62/209,010的权益。上述申请的全部教示内容通过引用结合在此。

发明领域

[0003] 本发明涉及包括N-内含肽多肽和N-内含肽溶解配偶体的融合蛋白,包括此类融合蛋白的亲和色谱基质,和在蛋白质纯化和肽连接过程中,特别是在工业规模下使用此类融合蛋白的方法。

[0004] 发明背景

[0005] 涉及用亲和标记物来标记感兴趣的蛋白质的蛋白质纯化方法广泛用于针对R&D应用的实验室条件,但已经证明对于大规模生产操作是不切实际的。在生物加工工业中,仅使用可裂解的亲和标记物来确保终产物不包含标记物,该标记物必须在生产过程中通常使用位点特异性蛋白酶来除去。去除该亲和标记物需要另外的加工步骤,其很大程度地增加了成本和时间,特别在工业规模下。此外,低效和位点外裂解导致最终蛋白质产物被分别保留标记物和截短的蛋白质片段的蛋白质污染,这在生物加工应用中是不可接受的。

[0006] 因此,需要开发允许在工业条件下大规模纯化蛋白质的改进的亲和色谱试剂和方法。

[0007] 发明概述

[0008] 内含肽是包含蛋白酶和连接酶活性的一类自动催化的酶。称为"断裂内含肽"的一类内含肽涉及,包括两种互补的半内含肽,称为N-内含肽和C-内含肽,其选择性地并且非常紧密地缔合以形成活性的内含肽酶(沙阿N.H.(Shah N.H.)等人,美国化学会会志(J.Amer.Chem.Soc.)135:18673-18681;达萨B.(Dassa B.)等人,核酸研究(Nucl.Acids Res.),37:2560-2573(2009))。

[0009] 先前在现有技术中已经描述了内含肽(包括断裂内含肽)在大规模的蛋白质纯化过程中的用途(参见,例如,WO 2013/045632)。先前还描述了断裂内含肽用于从粗混合物中色谱分离感兴趣的蛋白质的用途(参见,例如,中国公开号CN 101884910;关D.(Guan D.),等人,生物技术与生物工程(Biotech.Bioeng.)110:2471-2481(2013);刘W.(Lu W.),等人,染色体杂志(J.Chrom.A),1218:2553-2560(2011))。

[0010] 然而,当在常见的表达系统(如大肠杆菌)中表达时,内含肽在大规模蛋白质纯化过程中的使用受其差溶解度阻碍。此外,尚未描述包括共价附接至固体支持物的基于内含肽的亲和配体的色谱基质,其对于有效的工业规模的蛋白质纯化过程是关键的。

[0011] 本发明提供了包括能够通过与包括C-内含肽多肽的第二融合蛋白相缔合来形成有活性的内含肽复合体的N-内含肽多肽的可溶融合蛋白。可以将包括N-内含肽多肽的融合蛋白共价附接至固体支持物,以产生适合用于大规模生物加工应用的亲和色谱基质。

[0012] 因此,在一个实施例中,本发明涉及包括由肽键连接的N-内含肽多肽和N-内含肽

溶解配偶体的融合蛋白。在本实施例的一个具体的方面，该N-内含肽溶解配偶体具有小于约15kDa的分子量、小于约60的脂肪族氨基酸指数值、和小于-1的总平均亲水性值，并且增强(例如，增加和/或促进)了N-内含肽多肽的溶解度。在本实施例的另外方面，该N-内含肽溶解配偶体包括SEQ ID NO:15。在本实施例的又另一个方面中，该N-内含肽多肽是GP41-1N-内含肽(SEQ ID NO:1)，或其变体。

[0013] 在另一个实施例中，本发明涉及包括融合蛋白的亲和色谱基质，该融合蛋白包括N-内含肽多肽和N-内含肽溶解配偶体，并且附接至固体支持物。在本实施例的一个具体的方面，该固体支持物是包括亲水性聚乙烯醚基的色谱树脂。

[0014] 在另外的实施例中，本发明涉及将样品中的靶分子亲和纯化的方法。在本实施例的一个方面，该方法包括a)提供一种包含第一融合蛋白的样品，该第一融合蛋白包括通过肽键连接至靶分子的C-内含肽多肽；b)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成一种无活性的内含肽复合体的条件下，将该样品与包括第二融合蛋白的亲和色谱基质接触，其中该第二融合蛋白包括通过肽键结合至N-内含肽溶解配偶体的N-内含肽多肽，该N-内含肽溶解配偶体促进该N-内含肽多肽的溶解度；c)洗涤该包含无活性的内含肽复合体的亲和色谱基质，以去除未结合的污染物；d)将该内含肽复合体暴露在该内含肽复合体具有活性并且从C-内含肽多肽裂解靶分子的条件下；并且e)回收该经裂解的靶分子。

[0015] 在又另一个实施例中，本发明涉及筛选适合在亲和纯化中使用的有催化活性的内含肽复合体的方法。在本实施例的一个方面，该方法包括a)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成内含肽复合体的条件下，将包括连接至靶分子(例如，通过肽键或连接分子)的C-内含肽多肽的第一融合蛋白与包括连接至N-内含肽溶解配偶体(例如，通过肽键或连接分子)的N-内含肽多肽的第二融合蛋白接触；和b)在支持内含肽活性的条件下，确定该靶分子是否从该C-内含肽多肽裂解下来，其中该经裂解的靶分子的存在表明有催化活性的内含肽复合体存在。

[0016] 本发明的N-内含肽融合蛋白具有改进的溶解度和增强的催化活性，并且作为试剂当与相应的C-内含肽配合时，用于进行大规模蛋白质纯化(例如，亲和色谱法)和修饰过程(例如，肽裂解和连接反应)是有用的。

[0017] 附图简述

[0018] 该专利或申请文件包含至少一张彩色附图。在请求并支付必要的费用后，官方将会提供带有彩色附图的本专利或专利申请公开物的副本。

[0019] 图1是描述本发明的示例性亲和纯化方法的原理图。该方法采用本发明的包括融合蛋白的示例性亲和色谱基质，该融合蛋白具有融合至附接至固体支持物(表面)的N-内含肽溶解配偶体的N-内含肽多肽。将包括与在亲和色谱基质中的N-内含肽互补的C-内含肽的第二融合蛋白融合至待纯化的靶蛋白(感兴趣的蛋白质)，以及表达所需的任何其他的因子，如分泌信号。图1A示出了在C-内含肽融合蛋白结合至N-内含肽亲和色谱基质之前的各种组分。图1B示出了在内含肽缔合的适当的条件下(例如，pH、盐、氧化/还原)，与N-内含肽亲和色谱基质结合的C-内含肽融合蛋白。图1C示出了在内含肽复合体的催化活性的适当的条件下，在N-和C-外显肽已经从它们各自的融合蛋白中裂解后的这些组分。

[0020] 图2A是描绘融合极性对与GP41-1N-内含肽的N末端(SOLP-NINT)或C末端(NINT-

SOLP)融合的三种候选N-内含肽溶解配偶体(46、206、246;参见表2)的催化活性(裂解速率)的影响的图。

[0021] 图2B是描绘融合极性对与GP41-1N-内含肽的N末端(SOLP-NINT)或C末端(NINT-SOLP)融合的三种候选N-内含肽溶解配偶体(46、206、246)的大肠杆菌中的蛋白质表达的影响的图。

[0022] 图3是描绘与GP41-1N-内含肽的C末端融合的七种候选溶解配偶体(46、206、246、51、138、342、368)的底物裂解速率和可溶表达效价的图。

[0023] 图4是描绘候选溶解配偶体的经计算的物理特性和包含与N-内含肽的C末端融合的溶解配偶体的融合蛋白在大肠杆菌中的总(效价)或可溶表达(可溶性效价)表达之间的相关性的图。mw:分子量;pI:等电点;AI:脂肪族氨基酸指数;GRAVY:总平均亲水性。

[0024] 图5A是显示在相应于GP41-1内含肽的位置65处的半胱氨酸的残基处在大约一百个GP41-1同系物中发现的特定氨基酸的频率的图表。

[0025] 图5B是显示在相应于GP41-1内含肽的位置89处的半胱氨酸的残基处在大约一百个GP41-1同系物中发现的特定氨基酸的频率的图表。

[0026] 图6是描绘溶解配偶体138与包含在位置65和89处的两个中心分布的、天然存在的半胱氨酸残基的野生型GP41-1N-内含肽,或包含针对在位置65和89处的一个或两个半胱氨酸残基的氨基酸取代的GP41-1N-内含肽变体的融合蛋白的催化活性(裂解率)图。

[0027] 图7描绘了针对溶解配偶体138(蛋白质数据库结构1RYK)的NMR溶液结构。该蛋白质包含四个α螺旋结构域,是球状的,具有长的非结构化螺旋,该螺旋形成与N-内含肽的羧基端的连接(圆形区域;N-内含肽未显示)。由黄色突出指示的环形区域GKL和GYQ被靶向用于半胱氨酸残基插入,以制造溶解配偶体138(138_GKL22GCKL、138_GYQ48GCYQ、和138_GYQ48GCGYQ)的新形式(GCKL(SEQ ID NO:61)、GCYQ(SEQ ID NO:62)、和GCGYQ(SEQ ID NO:63))。

[0028] 发明详细描述

[0029] 本发明的示例实施例的描述如下。

[0030] I.定义

[0031] 为了更容易地理解本披露,首先定义某些术语。另外的定义贯穿详细说明而阐述。除非另外指明,本文所用的所有科学技术术语具有如本发明所属领域的普通技术人员通常理解的相同意义。

[0032] 术语“感兴趣的生物分子”和“靶分子”在此可互换地使用,来指生物分子(例如,蛋白质)、材料、或大分子集合,该集合有待进行,例如,纯化或从混合物(例如,粗蛋白混合物)中去除。感兴趣的示例性生物分子包括,例如,重组肽和蛋白质,包括抗体(例如,单克隆抗体)、疫苗、病毒、和其他大分子集合,如可以掺入生物分子和合成的组分的病毒样颗粒和纳米颗粒。通过举例,感兴趣的生物分子可以包括蛋白质和生物分子集合(例如,由重组DNA技术产生的),例如像,激素(例如胰岛素、人生长激素、红细胞生成素、干扰素、粒细胞集落刺激因子、组织纤溶酶原活化物),单克隆抗体(mAb)和mAb-衍生物(例如,双特异性mAb、Fab、scFv、鲨鱼和骆驼抗体),支架衍生的治疗法(例如,DARPin、亲合体、抗体模拟物(anticalin)),治疗性酶(例如,α半乳糖苷酶A、α-L-艾杜糖醛酸酶、N-乙酰半乳糖胺-4-硫酸酯酶、葡糖脑苷脂酶),毒素(例如,肉毒杆菌、CRM 197、蓖麻毒素),重组疫苗(例如,炭疽、

白喉、破伤风、肺炎、乙型肝炎病毒、人乳头状瘤病毒)，病毒样颗粒(例如，乙型肝炎、人乳头状瘤、流行性感冒、细小病毒、诺瓦克病毒)，以及工业酶(例如，木瓜蛋白酶、菠萝蛋白酶、胰蛋白酶、蛋白酶K、BENZONASE™酶、DENERASE™酶、脲酶、胃蛋白酶等)，以及诊断试剂(例如，葡萄糖和乳酸脱氢酶、DNA聚合酶、碱性磷酸酶、辣根过氧化物酶、限制性酶、杂交瘤衍生的抗体等)。在一个具体实施例中，该靶分子是治疗靶标的抗体(例如，单克隆抗体)。

[0033] 术语“融合蛋白”是指天然存在的、合成的、半合成的或重组单个蛋白分子，其包括通过肽键连接的两个或更多个异源多肽的全部或部分。

[0034] 如在此使用的术语“肽的”是指还可以掺入非氨基酸分子(例如，生色团、药物、毒素、成像造影剂等)的、长度大于两个氨基酸的肽和蛋白质。

[0035] 术语“多肽”是指氨基酸的聚合物，并且不是指特定长度；因此，肽、寡肽、和蛋白质包括在多肽的定义内。

[0036] 如在此使用的术语“断裂内含肽”，是指从自然界分离或通过重组DNA技术产生的蛋白质，该蛋白质具有以下特性：(1)蛋白质以高亲和性和高选择性相互作用的两半存在；(2)这两半必须包含催化活性所需要的所有内含肽序列，并且还可以包含附带的非内含肽肽序列；(3)仅当这两半紧密缔合时，该蛋白质具有酶活性；并且(4)该酶活性是位点选择性肽裂解或连接，其可用于从非内含肽肽序列中分离内含肽序列或将该非内含肽肽序列连接至连续的线性或环状蛋白质。

[0037] 在此使用的术语“互补的内含肽”是指断裂内含肽对的N-内含肽和C-内含肽部分。

[0038] 如在此使用的术语“N-内含肽”是指与单个内含肽多肽的N末端部分具有同源性，并且与互补的C-内含肽缔合以形成活性内含肽酶的内含肽多肽。

[0039] 如在此使用的术语“C-内含肽”是指与单个内含肽多肽的C末端部分具有同源性，并且与互补的N-内含肽缔合以形成活性内含肽酶的内含肽多肽。

[0040] 如在此使用的术语“外显肽”是指自然界中与N-和C-内含肽融合的并且通过断裂内含肽的酶促作用来操纵(例如，裂解或连接)的N末端和C末端的肽序列。

[0041] 如在此使用的术语“配体”是指能够与另一种分子强烈和选择性相互作用的分子，特别是当附接至一种表面，如色谱树脂时。在本发明的一些实施例中，该配体可以是在此描述的N-内含肽融合蛋白。

[0042] 如在此使用的术语“溶解配偶体”是指相对于在溶解配偶体不存在情况下所表达的可溶N-内含肽的量，当融合至N-内含肽时，增强(例如，增加、促进、或维持)大肠杆菌中所表达的可溶N-内含肽的量的一种蛋白。例如，在各种实施例中，相对于当没有与溶解配偶体一起表达时的内含肽的溶解度，将N-内含肽表达为与溶解配偶体的融合蛋白可以增加N-内含肽的溶解度至少约10％(例如，约20％、约30％、约40％、约50％、约60％、约70％、约80％、约90％，或更多)。

[0043] 在一个实施例中，将溶解配偶体E(SEQ ID NO:25)融合至N-内含肽，并且将生成的融合蛋白的溶解度用于提供实验基线。当单独的N-内含肽不可溶或不稳定时，这是特别有用的。

[0044] 如在此使用的术语“亲本分子”或”野生型(wt)对应物”或“wt蛋白质”或“wt结构域”旨在指处于其基本天然的形式的相应的蛋白质(例如，N-内含肽、N-内含肽溶解配偶体)，或蛋白质的结构域，其通常用作本文的对照。

[0045] 术语“序列一致性”意指两种核苷酸或氨基酸序列，当最佳比对时，如通过程序GAP或BESTFIT使用默认空位权重，共有至少70％序列一致性、或至少80％序列一致性、或至少85％序列一致性、或至少90％序列一致性、或至少95％序列一致性，或更多序列一致性。对于序列比较，通常一个序列用作参考序列(例如，亲本序列)，与其比较测试序列。当使用序列比较算法时，测试和参考序列被输入到计算机中，指定子序列坐标(如果需要)，并且指定序列算法程序参数。基于指定的程序参数，序列比较算法然后计算一种或多种测试序列相对于参考序列的序列一致性百分比。

[0046] 可以进行用于比较的序列的最佳比对，例如，通过史密斯(Smith)和沃特曼(Waterman)，高等应用数学(Adv.Appl.Math.)2:482(1981)的局部同源性算法，通过尼的曼(Needleman)和文施(Wunsch)，分子生物学杂志(J.Mol.Biol.)48:443(1970)的同源比对算法，通过皮尔森(Pearson)和利普曼(Lipman)，美国国家科学院院刊(Proc.Nat’l.Acad.Sci.USA)85:2444(1988)的相似性搜索方法，通过这些算法的计算机化执行(在威斯康星遗传软件包，遗传计算组，575Science Dr.，麦迪逊，威斯康星州中的GAP、BESTFIT、FASTA、和TFASTA)，或者通过目测检查(通常参见奥斯贝(Ausubel)等人，现代分子生物学实验技术(Current Protocols in Molecular Biology)。适合于确定序列一致性百分比和序列相似性的算法的一个实例是BLAST算法，其描述于阿尔丘尔(Altschul)等人，分子生物学杂志(J.Mol.Biol.)215:403(1990)。用于进行BLAST分析的软件可通过国家生物技术信息中心公开获得(可通过国家卫生研究院NCBI互联网服务器公开访问)。通常，默认程序参数可以用于进行序列比较，尽管也可以使用自定义的参数。对于氨基酸序列，BLASTP程序使用词长(W)为3、期望值(E)为10、以及BLOSUM62得分矩阵作为默认值(参见赫尼科夫(Henikoff)和赫尼科夫(Henikoff)，美国国家科学院院刊(Proc.Natl.Acad.Sci.USA)89:10915(1989))。

[0047] 如在此使用的术语“色谱法”是指动态分离技术，其将感兴趣的靶分子与在混合物中的其他分子分开，并允许其被分离。通常，在色谱法中，流动相(液体或气体)将包含感兴趣的靶分子的样品运输跨越或穿过固定相(通常为固体)培养基。对固定相的分配或亲和力的差异使不同分子分离，而流动相在不同时间携带走不同的分子。

[0048] 如在此使用的术语“亲和色谱法”是指一种色谱模式，其中待分离的靶分子是通过其与一种分子(例如，根据本发明包括N-内含肽和N-内含肽溶解因子的亲和色谱配体)相互作用而分离的，该分子特异性地与靶分子相互作用。在一个实施例中，亲和色谱法涉及向在其上带有如在此描述的基于N-内含肽的配体的固体支持物添加包含靶分子(例如，免疫球蛋白或包含Fc的蛋白质)的样品。

[0049] 如在此可互换使用的术语“亲和基质”或“亲和色谱基质”是指亲和色谱配体(例如，N-内含肽融合蛋白或其结构域)附接其上的一种色谱支持物。待从混合物中纯化或去除的配体能够通过亲和相互作用(例如，互补的C-内含肽融合蛋白)结合至感兴趣的分子。

[0050] 术语“免疫球蛋白”、“Ig”或“抗体”(在此可互换地使用)是指具有由两条重链和两条轻链组成的碱性四多肽链结构的蛋白质，所述链例如通过链间二硫键稳定的，该蛋白质具有特异性结合抗原的能力。术语“单链免疫球蛋白”或“单链抗体”(在此可互换地使用)是指具有由一条重链和一条轻链组成的二多肽链结构的蛋白质，所述链例如通过链间肽接头被稳定，该蛋白质具有特异性结合抗原的能力。术语“结构域”是指包括肽环(例如，包括3个

至4个肽环)的重链或轻链多肽的球状区域,这些肽环通过例如β折叠片和/或键内二硫键而稳定。基于在“恒定”结构域的情况下在各类成员的结构域内的序列变异的相对缺乏,或基于在“可变”结构域的情况下在各类成员的结构域内的显着变化,结构域在此进一步称为“恒定的”或“可变的”。抗体或多肽“结构域”通常在本领域可互换地称为抗体或多肽“区域”。抗体轻链的“恒定”结构域可互换地称为“轻链恒定区域”、“轻链恒定结构域”、“CL”区域、或“CL”结构域。抗体重链的“恒定”结构域可互换地称为“重链恒定区域”、“重链恒定结构域”、“CH”区域、或“CH”结构域。抗体轻链的“可变”结构域可互换地称为“轻链可变区域”、“轻链可变结构域”、“VL”区域、或“VL”结构域。抗体重链的“可变”结构域可互换地称为“重链可变区域”、“重链可变结构域”、“VH”区域、或“VH”结构域。

[0051] “抗体”或“免疫球蛋白”可以是单克隆的或多克隆的,并且可以按单体或聚合体形式存在,例如,以五聚体形式存在的IgM抗体和/或以单体、二聚体、或多聚体形式存在的IgA抗体。术语“片段”是指相比完整的或完全的抗体或抗体链,包括较少氨基酸残基的一段或一部分抗体或抗体链。可以通过化学或酶处理完整或完全的抗体或抗体链来获得片段。片段还可以通过重组手段来获得。示例性片段包括Fab、Fab’、F(ab’)2、Fc、和/或Fv片段。

[0052] 术语“多核苷酸”和”核酸分子”在此可互换地使用,是指任何长度的核苷酸的聚合体形式(核糖核苷酸或脱氧核糖核苷酸)。这些术语包括单链、双链、或三链的DNA,基因组DNA,cDNA,RNA,DNA-RNA杂合体,或包括嘌呤碱和嘧啶碱,或其他天然的、经化学或生物化学修饰的、非天然的、或衍生的核苷酸碱的聚合物。多核苷酸的主链可包括糖和磷酸基团(正如通常可在RNA或DNA中所见的)或者经修饰或取代的糖或磷酸基团。此外,可以通过合成互补链,并且在适当的条件下使链退火,或通过使用DNA多聚酶用合适引物从头合成互补链,从化学合成的单链多核苷酸产物而获得双链多核苷酸。核酸分子可以采取许多不同的形式,例如基因或基因片段、一个或多个外显子、一个或多个内含子、mRNA、cDNA、重组多核苷酸、分枝多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针和引物。多核苷酸可以包括经修饰的核苷酸,如经甲基化的核苷酸和核苷酸类似物,尿嘧啶基,其他糖,和连接基团(如氟代核糖和硫代酸酯),以及核苷酸分枝。如在此使用,“DNA”或“核苷酸序列”不仅包括碱基A、T、C、和G,而且还包括任何它们的类似物,或这些碱基的经修饰的形式,如经甲基化的核苷酸,核苷酸间修饰如不带电的键联和硫代酸酯,糖类似物,以及经修饰的和/或可替代的主链结构,如聚酰胺的使用。在一个具体实施例中,核酸分子包括如在此描述的编码N-内含肽融合蛋白或其变体的核苷酸序列。

[0053] II.基于内含肽的融合蛋白

[0054] 内含肽是在1990年发现的一类自动催化的酶,其包含在这些分子的天然生活周期中起作用的蛋白酶和连接酶活性。已经证明内含肽试剂具有用于裂解、连接、和环化肽底物的效用。1998年,发现了称为“断裂内含肽”的一类新内含肽,其中酶天然地存在于两部分中,称为N-内含肽和C-内含肽(互补的半内含肽)。尽管已经在多种较低等的原核生物中发现断裂内含肽(赛特勒J.(Zettler J.)等人,欧洲生化学会联合会快报(FEBS Letters),553:909-914(2009);达萨B(Dassa B.)等人,生物化学(Biochemistry),46:322-330(2007);蔡J.(Choi J.)等人,分子生物学杂志(J Mol Biol.)556:1093-1106(2006);卡斯皮(Caspi)等人,分子微生物学(Mol Microbiol),50:1569-1577(2003);刘X(Liu X.)和杨J.(Yang J.),生物化学杂志(J Biol Chem.),275:26315-26318(2003);吴H(Wu H.),等人,

美国国家科学院院刊(Proc Natl Acad Sci USA.)5:9226-9231(1998)),在真核生物中没有鉴定出断裂内含肽(参见由新英格兰生物实验室维护的内含肽数据库(http://tools.neb.com/inbase/list.php))。两种断裂内含肽最近被表征为它们二者对于相邻的外显肽序列都非常快并且相当多样。一类是Npu DnaE内含肽(艾维I.(Iwai I.)等人,欧洲生化学会联合会快报(FEBS Letters)550:1853-1858(2006);赛特勒J.(Zettler J.),等人,欧洲生化学会联合会快报(FEBS Letters),553:909-914(2009))以及其他,从宏基因组数据鉴定出的GP41断裂内含肽(卡瓦杰-维拉乔斯P(Carvajal-Vallejos P.),等人,生物化学杂志(J.Biol.Chem.)287:28686-28696(2012);国际PCT公开号WO 2013045632)。

[0055] 具有附接的外显肽(通过内含肽活性连接的两个半蛋白)的N-和C-内含肽通过多个结构域间相互作用极其特异地和紧密地连接,以形成有活性的内含肽酶(沙阿N.H.(Shah N.H.)等人,美国化学会会志(J.Amer.Chem.Soc.)135:18673-18681;达萨B.(Dassa B.)等人,核酸研究(Nucl.Acids Res.),37:2560-2573(2009))。除了存在于第一类内含肽中的连接酶和蛋白酶活性外,由于N-和C-内含肽结构域的紧密的和选择性的相互作用,断裂内含肽在亲和分离中具有效用。

[0056] 本发明部分基于如下发现:将N-内含肽多肽表达为与某些异源蛋白质(在此称为溶解配偶体)的融合蛋白增加了该内含肽的溶解度,从而使内含肽适合作为用于可以在小规模或大规模实践的亲和色谱以及其他蛋白质纯化和修饰应用的试剂。更具体地,本发明提供高度可溶的融合蛋白,这些融合蛋白包括N-内含肽多肽和能够通过与互补的C-内含肽多肽缔合而形成有活性的内含肽复合体的N-内含肽溶解配偶体。本发明还提供包括C-内含肽多肽和靶分子的融合蛋白,其中该融合蛋白能够与包括互补的N-内含肽多肽和N-内含肽溶解配偶体的另一种融合蛋白缔合。

[0057] 因此,在一个实施例中,本发明涉及包括全部的,或部分的N-内含肽多肽和N-内含肽溶解配偶体的融合蛋白。各种N-内含肽多肽在本领域是已知的。示例性N-内含肽包括在表1中示出的N-内含肽,以及在本文的其他地方被描述的其他N-内含肽。在此披露的N-内含肽,和本领域熟知的其他N-内含肽,以及此类N-内含肽的变体(这些变体与野生型N-内含肽具有至少约75%序列一致性(例如,至少约80%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%序列一致性)),可以包括在在此描述的融合蛋白中。

[0058] 内含肽N末端结构域中的第一个氨基酸通常是高度保守的,并且对于蛋白质剪接反应可以是重要的。然而,在一些实施例中,在内含肽N末端结构域中的第一个氨基酸(例如,半胱氨酸、丝氨酸)可以被氨基酸(例如,除了半胱氨酸或丝氨酸以外的氨基酸)取代,其防止或减少内含肽和异源多肽之间的裂解。在一个具体实施例中,在内含肽N末端结构域中的第一个氨基酸被丙氨酸取代。

[0059] 在一个具体实施例中,在此描述的N-内含肽融合蛋白包括野生型GP41-1N-内含肽(SEQ ID NO:1或SEQ ID NO:29)或其变体。适合的变体GP41-1N-内含肽可以与野生型GP41-1N-内含肽(SEQ ID NO:1)具有至少约75%序列一致性(例如,至少约80%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%序列一致性)。包含在本发明的融合蛋白中的变体GP41-1N-内含肽的具体实例包括在此指定为SEQ ID NO:2-8的GP41-1变体。在某些实施例中,该GP41-1N-内含肽变体缺乏半胱氨酸残基。在一个具体实施例中,删除在GP41-1N-内含肽中天然存在的一个或多个半胱氨酸残基。在另一个实施例中,在

GP41-1N-内含肽(SEQ ID NO:1的位置7、65、和89)中天然存在的一个或多个半胱氨酸残基被另一个氨基酸残基(例如,苏氨酸、赖氨酸、或天冬酰胺)取代。在一个实施例中,在SEQ ID NO:1的位置65处的GP41-1N-内含肽中天然存在的半胱氨酸残基被另一个氨基酸残基(例如,丝氨酸、苏氨酸)取代。在一个具体实施例中,在SEQ ID NO:1的位置65处的半胱氨酸残基被苏氨酸取代。在又另一个实施例中,在SEQ ID NO:1的位置89处的GP41-1N-内含肽中天然存在的半胱氨酸残基被另一个氨基酸残基(例如,甲硫氨酸、酪氨酸)取代。在一个具体实施例中,在SEQ ID NO:1的位置89处的半胱氨酸残基被甲硫氨酸取代。在一些实施例中,该GP41-1变体是GP41-1NINTΔA_TM N-内含肽变体(SEQ ID NO:6)或GP41-1NINTΔA_TK N-内含肽变体(SEQ ID NO:8)。

[0060] 在一些实施例中,缺少一些或所有半胱氨酸残基的GP41-1N-内含肽变体比连接或裂解反应中的天然GP41-1N-内含肽多出至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、或至少10倍活性。通常可以在还原条件下使用SDS凝胶电泳来分析裂解或连接的内含肽活性(例如,赛特勒J.(Zettler J.),舒兹V.(Schütz V.),穆兹H.D.(Mootz H.D.),欧洲生化学会联合会快报(FEBS Letters)583:909-914,2009;阿润克A.S.(Aranko A.S.),苏格S(Züger S),布青格E(Buchinger E),伊娃H(Iwai H),公共科学图书馆期刊(PLoS ONE)4:e5185,2009)。简言之,通过添加包含还原剂(例如,二硫苏糖醇或β-巯基乙醇)的SDS凝胶加样缓冲液来停止内含肽反应(通常是时间过程),将样品煮沸至使其完全变性,然后将它们加载到包含SDS连同合适的蛋白质尺寸标记物的聚丙烯酰胺凝胶上。电泳完成后,反应中的蛋白质根据其分子量被分离,并且可以通过用传统染料或荧光染料进行染色而可视化。作为时间函数的各种中间体和产物的量可以通过光密度测定法和通过应用曲线拟合程序转化为将作为酶速率(kobs)的作为时间函数的强度来定量。

[0061] 表1.示例性GP41-1断裂内含肽及其变体

[0062]

SEQ ID NO:	名称	序列
1	GP41-1 N-内含肽	mtrsgyCLDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLF PTQTGEMNISGGLKEGMCLYVKEgg
2	GP41-1 NINTΔA_CC N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLF PTQTGEMNISGGLKEGMCLYVKEgg
3	GP41-1 NINTΔA_AC N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIIASEEHLF PTQTGEMNISGGLKEGMCLYVKEgg
4	GP41-1 NINTΔA_CK N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLF PTQTGEMNISGGLKEGMKLYVKEgg
5	GP41-1 NINTΔA_AM N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIIASEEHLF PTQTGEMNISGGLKEGMMLYVKEgg
6	GP41-1 NINTΔA_TM N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIITSEEHLF PTQTGEMNISGGLKEGMMLYVKEgg
7	GP41-1 NINTΔA_AK N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIIASEEHLF PTQTGEMNISGGLKEGMKLYVKEgg
8	GP41-1 NINTΔA_TK N-内含肽变体	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNT GYNEVLNVFPKSKKKSYKITLEDGKEIITSEEHLF PTQTGEMNISGGLKEGMKLYVKEgg
9	GP41-1 C-内含肽(CINT)	MGKNSMMLKKILKIEELDERELIDIEVSGNHLFYA NDILTHN
10	GP41-1 C-内含肽-硫氧还蛋白融合蛋白(CINT_TRX) (硫氧还蛋白部分加下划线)	MGKNSMMLKKILKIEELDERELIDIEVSGNHLFYA NDILTHNMSDKIIHLTDDSFDTDVLKADGAILVDF WAEWCGPCKMIAPILDEIADEYQGKLTVAKLNID QNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSK GQLKEFLDANLAHHHHHH

[0063] 对于SEQ ID NO:1-8,使用小写字母文本指示非内含肽序列,并且由大写字母文本指示针对SEQ ID NO:1-8的内含肽序列。

[0064] 通常,当N-内含肽多肽在常见的表达系统,如大肠杆菌中表达时,其具有较差的溶解度。本发明通过例如将N-内含肽表达为与N-内含肽溶解配偶体的融合蛋白来避免这一问题,其增加了该N-内含肽的溶解度(例如,当在大肠杆菌中表达时)。优选地,该N-内含肽溶解配偶体增加N-内含肽多肽的溶解度,这样使得在表达系统(例如大肠杆菌)中产生后,按质量计小于约25%的生成的融合蛋白存在于包涵体中。在表达系统中产生后存在于包涵体中的所表达的蛋白质的按质量计的百分比可以由本领域的普通技术人员使用标准技术和试剂容易地确定。

[0065] 本领域普通技术人员可以使用本领域已知的和本文描述的技术容易地选择可以增加给定N-内含肽的溶解度的潜在的溶解配偶体。例如,在表达系统(例如大肠杆菌)中过量表达时产生可溶产物的概率可以使用威尔金森(Wilkinson)和哈里森(Harrison)的算法(威尔金森DL(Wilkinson DL)和哈里森RG(Harrison RG),生物/技术(Bio/Technology),9:443,1991)来计算。对蛋白质是否包含功能性分泌信号的预测可以使用可从丹麦技术大学的生物序列分析中心获得的SignalP 4.1算法来进行(http://genome.cbs.dtu.dk/services/SignalP/)。同样参见在此披露的实例1-3中所描述的方法。最终,必须通过实验筛选从候选溶解配偶体中选择提供溶解度最佳增强同时允许最大内含肽催化活性的溶解配偶体。

[0066] 具有某些物理特性的N-内含肽溶解配偶体特别适合于包含在本发明的融合蛋白中。此类物理特性包括但不限于，小于约15kDa的分子量、小于约60的脂肪族氨基酸指数(AI)值、以及小于-1的GRAVY值。本领域普通技术人员使用标准测定和技术，例如使用是生物信息学工具的SwissProt ExPASy程序组的一部分的在线ProtParam工具(http://web.expasy.org/tools/protparam/)，可以针对给定的溶解配偶体来确定这些特性中的每一个。

[0067] 将线性多肽序列的总平均亲水性(GRAVY)(凯特J(Kyte J)和杜利特尔RF.(Doolittle RF.)，分子生物学杂志(J.Mol.Biol.)157:105,1982)计算为所有氨基酸的亲水性值的总和，除以该序列中的残基数目。增加的正评分表明更高的疏水性。该计算是基于凯特-杜利特尔(Kyte-Doolittle)指数。GRAVY是用于显示蛋白质的亲水性状的简单方法。

[0068]

丙氨酸	1.8	亮氨酸	3.8
精氨酸	-4.5	赖氨酸	-3.9
天冬酰胺	-3.5	甲硫氨酸	1.9
天冬氨酸	-3.5	苯丙氨酸	2.8
半胱氨酸	2.5	脯氨酸	-1.6
谷氨酰胺	-3.5	丝氨酸	-0.8
谷氨酸	-3.5	苏氨酸	-0.7
甘氨酸	-0.4	色氨酸	-0.9
组氨酸	-3.2	酪氨酸	-1.3
异亮氨酸	4.5	缬氨酸	4.2

[0069] 在各种实施例中，在此描述的N-内含肽融合蛋白具有小于-1的GRAVY值。

[0070] 蛋白质脂肪族氨基酸指数(伊凯AJ.(Ikai,AJ.)，生物化学杂志(J.Biochem.)88:1895,1980)被定义为由脂肪族侧链(丙氨酸、缬氨酸、异亮氨酸、和亮氨酸)占据的相对体积。它可以被认为是球状蛋白的热稳定性增加的正因子。根据以下公式计算蛋白质的脂肪族氨基酸指数：脂肪族氨基酸指数＝X(Ala)+a*X(Val)+b*(X(Ile)+X(Leu))。*系数a和b是缬氨酸侧链(a＝2.9)和Leu/Ile侧链(b＝3.9)相对于丙氨酸侧链的相对体积。在大肠杆菌中过量表达时产生可溶产物的概率还可以使用威尔金森(Wilkinson)和哈里森(Harrison)的算法(威尔金森DL(Wilkinson DL)和哈里森RG(Harrison RG)，生物/技术(Bio/Technology)，9:443,1991)来计算。其他可用的算法不一定给出类似的结果。在各种实施例中，在此描述的N-内含肽融合蛋白具有小于约60的脂肪族氨基酸指数(AI)值，以及小于-1的GRAVY值。

[0071] 优选地，该N-内含肽溶解配偶体具有小于约15kDa的分子量、小于约60的脂肪族氨基酸指数值。

[0072] 表2中披露了具体的N-内含肽溶解配偶体的实例。

[0073] 表2.示例性N-内含肽溶解配偶体

[0074]

SEQ ID NO:	名称	GID	序列
11	溶解配偶体 46：Qin 前噬菌体：冷休克诱导的蛋白质	170081219	MREYPNGEKTHLTVMAAGFPSLTGD HKVIYVAADRHVTSEEILEAAIRLLS
12	溶解配偶体 206：假定蛋白 ECDH10B_1576[大肠杆菌品系 K-12 亚系 DH10B]	170081120	MSHLDEVIARVDAAIEESVIAHMNEL LIALSDDAELSREDRYTQQQRLRTAIA HHGRKHKEDMEARHEQLTKGGTIL
13	溶解配偶体 246：假定蛋白 ECDH10B_1388[大肠杆菌品系 K-12 亚系 DH10B]	170080950	MNKETQPIDRETLLKEANKIIREHEDT LAGIEATGVTQRNGVLVFTGDYFLDE QGLPTAKSTAVFNMFKHLAHVLSEKY HLVD
14	溶解配偶体 51：假定蛋白 ECDH10B_0422[大肠杆菌品系 K-12 亚系 DH10B]	170080051	MSLENAPDDVKLAVDLIVLLEENQIPA STVLRALDIVKRDYEKKLTRDDEAEK
15	溶解配偶体 138：推定的胁迫应答蛋白[大肠杆菌品系 K-12 亚系 DH10B]	170083502	MNKDEAGGNWKQFKGKVKEQWGK LTDDDMTIIEGKRDQLVGKIQERYGY QKDQAEKEVVDWETRNEYRW
16	溶解配偶体 138 GKL22GCKL：推定的胁迫应答蛋白[大肠杆菌品系 K-12 亚系 DH10B]	NA	MNKDEAGGNWKQFKGKVKEQWG**C** KLTDDDMTIIEGKRDQLVGKIQERYG YQKDQAEKEVVDWETRNEYRW
17	溶解配偶体 138 GYQ48GCYQ：推定的胁迫应答蛋白[大肠杆菌品系 K-12 亚系 DH10B]	NA	MNKDEAGGNWKQFKGKVKEQWGK LTDDDMTIIEGKRDQLVGKIQERYG**C** YQKDQAEKEVVDWETRNEYRW
18	溶解配偶体 138 GYQ48GCGY：推定的胁迫应答蛋白[大肠杆菌品系 K-12 亚系 DH10B]	NA	MNKDEAGGNWKQFKGKVKEQWGK LTDDDMTIIEGKRDQLVGKIQERYG**C** **G**YQKDQAEKEVVDWETRNEYRW
19	溶解配偶体 342：假定蛋白 ECDH10B_2487 [大肠杆菌品系 K-12 亚系 DH10B]	170081941	MIAEFESRILALIDGMVDHASDDELFA SGYLRGHLTLAIAELESGDDHSAQAV HTTVSQSLEKAIGAGELSPRDQALVT DMWENLFQQASQQ

[0075]

20	溶解配偶体 368：推定的 σ(54)调节蛋白[大肠杆菌品系 K-12 亚系 DH10B]	170082737	MQLNITGNNVEITEALREFVTAKFAK LEQYFDRINQVYVVLKVEKVTHTSDA TLHVNGGEIHASAEGQDMYAAIDGLI DKLARQLTKHKDKLKQH
21	溶解配偶体 A：3 型分泌系统的 EspA 组分	9626250	MDTSNATSVVNVSASSSTSTIYDLGN MSKDEVVKLFEELGVFQAAILMFSY MYQAQSNLSIAKFADMNEASKASTTA QKMANLVDAKIADVQSSTDKNAKAK LPQDVIDYINDPRNDISVTGISDLSGD LSAGDLQTVKAAISAKANNLTTVVNN SQLEIQQMSNTLNLLTSARSDVQSLQ YRTISAISLGK
22	溶解配偶体 B：来自肝片吸虫(HiTag)的 fh8 蛋白酶	147611	MPSVEVEKLLHVLDRNGDGKVSAEE LKAFADDSKYPLDSNKIKAFIKEHDK NKDGKLDLKELVSILSS
23	溶解配偶体 C：来自肝片吸虫(HiTag)的 fh8 蛋白酶	387618410	MPSVEVEKLLH
24	溶解配偶体 D：谷胱甘肽 S-转移酶	251787291	MGQLIDGVWHDTWYDTKSTGGKFQ RSASAFRNWLTADGAPGPTGKGGFAA EKDRYHLYVSLACPWAHRTLIMRKLK GLEPFISVSVVNPLMLENGWTFDDSF PGATGDTLYQHEFLYQLYLHADPHYS GRVTVPVLWDKKNHTIVSNESAEIIR MFNTAFDALGAKAGDYYPPALQPKID ELNGWIYDTVNNGVYKAGFATSQQA YDEAVAKVFESLARLEQILGQHRYLT GNQLTEADIRLWTTLVRFDPVYVTHF KCDKHRISDYLNLYGFLRDIYQMPGI AETVNFDHIRNHYFRSHKTINPTGIISI GPWQDLDEPHGRDVRFG
25	溶解配偶体 E：头部-DNA 稳定蛋白[肠杆菌噬菌体 λ]	410480759	MASWSHPQFEKASKETFTHYQPQGN SDPAHTATAPGGLSAKAPAMTPLMLD TSSRKLVAWDGTTDGAAVGILAVAAD QTSTTLTFYKSGTFRYEDVLWPEAAS DETKKRTAFAGTAISIV
26	溶解配偶体 F：麦芽糖结合蛋白	218465276	MKIKTGARILALSALTTMMFSASALA KIEEGKLVIWINGDKGYNGLAEVGKK FEKDTGIKVTVEHPDKLEEKFPQVAAT GDGPDIIFWAHDRFGGYAQSGLLAEIT PDKAFQDKLYPFTWDAVRYNGKLIAY PIAVEALSLIYNKDLLPNPPKTWEEIPA LDKELKAKGKSALMFNLQEPYFTWP LIAADGGYAFKYENGKYDIKDVGVD NAGAKAGLTFLVDLIKNKHMNADTD YSIAEAAFNKGETAMTINGPWAWSNI DTSKVNYGVTVLPTFKGQPSKPFVGV LSAGINAASPNKELAKEFLENYLLTDE GLEAVNKDKPLGAVALKSYEEELAKD PRIAATMENAQKGEIMPNIPQMSAFW YAVRTAVINAASGRQTVDEALKDAQT RITK
27	溶解配偶体 G：硫氧还蛋白	218465276	MSDKIIHLTDDSFDTDVLKADGAILVD FWAEWCGPCKMIAPILDEIADEYQGK LTVAKLNIDQNPGTAPKYGIRGIPTLLL FKNGEVAATKVGALSKGQLKEFLDA NLA

[0076]

28	溶解配偶体 H：转录终止因子 NusA	387509083	MNKEILAVVEAVSNEKALPREKIFEAL ESALATATKKKYEQEIDVRVQIDRKSG DFDTFRRWLVVDEVTQPTKEITLEAA RYEDESLNLGDYVEDQIESVTFDRITT QTAKQVIVQKVREAERAMVVDQFRE HEGEIITGVVKKVNRDNISLDLGNNA EAVILREDMLPRENFRPGDRVRGVLY SVRPEARGAQLFVTRSKPEMLIELFRI EVPEIGEEVIEIKAAARDPGSRAKIAV KTNDKRIDPVGACVGMRGARVQAVS TELGGERIDIVLWDDNPAQFVINAMA PADVASIVVDEDKHTMDIAVEAGNLA QAIGRNGQNVRLASQLSGWELNVMT VDDLQAKHQAEAHAAIDTFTKYLDID EDFATVLVEEGFSTLEELAYVPMKELL EIEGLDEPTVEALRERAKNALATIAQA QEESLGDNKPADDLLNLEGVDRDLAF KLAARGVCTLEDLAEQGIDDLADIEG LTDEKAGALIMAARNICWFGDEA

[0077] “GID”是指基因库ID(http://www.ncbi.nlm.nih.gov/genbank)。溶解配偶体A-H是已知的溶解配偶体,它们中的许多已经掺入市售的融合系统中,以增加在大肠杆菌中产生的重组蛋白的产量和溶解度。已经被人工取代或插入到亲本序列中的氨基酸以粗体和下划线突出显示。

[0078] 在一个具体实施例中,该N-内含肽溶解配偶体是,或包括全部或部分的溶解配偶体138(SEQ ID NO:15),或其变体(例如,溶解配偶体138GKL22GCKL(SEQ ID NO:16);溶解配偶体138GYQ48GCYQ(SEQ ID NO:17);溶解配偶体138GYQ48GCGY(SEQ ID NO:18))。

[0079] 制备融合蛋白或嵌合蛋白的方法是本领域熟知的,包括但不限于标准重组DNA技术。例如,根据常规技术将编码不同蛋白质序列(例如,N-内含肽和N-内含肽溶解配偶体;C-内含肽和靶分子)的DNA片段在框架内连接在一起。在另一个实施例中,该融合基因可以通过包括DNA自动合成仪的常规技术合成。可替代地,可以使用两个连续核酸片段之间的产生互补突出端的锚定引物进行核酸片段的PCR扩增,所述互补突出端随后可以退火并且再扩增以产生嵌合核酸序列(参见奥斯贝(Ausubel)等人,现代分子生物学实验技术(Current Protocols in Molecular Biology),1992)。此外,许多已经编码融合部分(例如,GST部分、Fc部分)的表达载体是市售的。

[0080] 优选地,该融合蛋白由在瞬时或稳定转染或转化的原核或真核宿主细胞或生物体中的编码核酸表达。用于表达重组蛋白质的常见的宿主细胞或生物体包括,例如,大肠杆菌、谷氨酸棒状杆菌、荧光假单胞菌、乳酸乳球菌、毕赤酵母、酿酒酵母、玉蜀黍、烟草、野胡萝卜、SF9细胞、CHO细胞(例如,CHO DG44细胞、CHO DXB11细胞)、NSO细胞、HEK 293细胞,以及所有的动物,如奶牛和山羊。在一个实施例中,该N-内含肽融合蛋白在大肠杆菌中表达。然后可以使用常规分离和色谱方法,如通过深层过滤进行澄清,通过阴离子和阳离子交换色谱法纯化,以及通过超滤浓缩,将表达的N-内含肽融合蛋白从污染的细胞蛋白质中纯化出来。

[0081] 可以将该异源蛋白质(例如,N-内含肽溶解配偶体、靶分子)融合至内含肽多肽的任一端。在一个实施例中,将N-内含肽溶解配偶体连接至N-内含肽多肽的N末端。在另一个实施例中,将N-内含肽溶解配偶体连接至N-内含肽多肽的C末端。

[0082] 在一些实施例中,该内含肽多肽(例如,N-内含肽、C-内含肽)和异源蛋白质(例如,

N-内含肽溶解配偶体、靶分子)直接通过肽键来连接。在另一个实施例中,该融合蛋白包括间隔子,或接头,内含肽多肽(例如,N-内含肽、C-内含肽)和异源蛋白质(例如,N-内含肽溶解配偶体、靶分子)之间的分子。适合的间隔子/连接分子是本领域已知的。

[0083] 在此处描述的融合蛋白中,该内含肽N末端结构域可以直接地(例如,通过肽键)或间接地(例如,通过接头氨基酸序列)融合至异源多肽。因此,在一些实施例中,异源多肽直接地或间接地融合至内含肽N末端结构域的N末端。在某些实施例中,异源多肽的第一个氨基酸选自下组,该组由以下各项组成:Met、Cys、Thr、Arg、Lys、Ser、GIn、His、Ala、Tyr、Phe、Asn、Trp、Val、Leu、Asp、He、Gly、Glu、和Pro。

[0084] 在一些实施例中,该融合蛋白包括异源多肽和内含肽序列之间的接头。例如,该融合蛋白可以包括异源蛋白质的C末端和内含肽的N末端结构域的N末端之间的接头。该接头的长度可以是,例如,从约1个至约10个氨基酸。在一些实施例中,该接头的长度可以是约1个至约5个氨基酸。例如,该接头可以包含1个、2个、3个、4个、或5个氨基酸。在一些实施例中,接触异源多肽和内含肽的N末端结构域的N末端的接头的最后一个氨基酸选自下组,该组由以下各项组成:Met、Cys、Thr、Arg、Lys、Ser、GIn、His、Ala、Tyr、Phe、Asn、Trp、Val、Leu、Asp、Ile、Gly、Glu、和Pro。

[0085] 在一些实施例中,该接头可以包括外显肽序列。在一些实施例中,该接头可以包括天然的外显肽序列。在一些实施例中,该外显肽包括选自下组的序列,该组由以下各项组成:来自WO 201345632的SEQ ID NO:4、8、13、17、21、25、35、和39。在一些实施例中,包括外显肽的氨基酸的接头包括,例如,选自下组的序列的最初的(即,N末端的)约1个至约5个氨基酸,该组由以下各项组成:SEQ ID NO:4、8、13、17、21、25、35、和39。在一些实施例中,该接头包括选自下组的序列的约1个、2个、3个、4个、或5个氨基酸,该组由以下各项组成:SEQ ID NO:4、8、13、17、21、25、35、和39。在一些实施例中,融合蛋白包括在自然中一起发现的内含肽结构域和外显肽结构域(例如,GP41-1N-内含肽和GP41-1C-内含肽)。在其他实施例中,融合蛋白包括在自然中没有与特别的内含肽结构域一起发现的内含肽结构域和外显肽结构域,在此还称为“异源外显肽结构域”。通过举例,融合蛋白可以包括GP41-1内含肽结构域和IMPDH外显肽结构域。

[0086] 本发明的融合蛋白可以任选地进一步包括一个或多个可检测标签。根据本发明适合使用的标签是本领域已知的,并且通常包括通过其化学性质,以及无论通过直接或间接手段提供允许检测蛋白质的可识别信号的任何分子。因此,例如融合蛋白可以按常规方式,如用特异性报道分子、荧光团、放射性材料、或酶(例如过氧化物酶,磷酸酶)进行标记。在一个具体实施例中,这些融合蛋白包括作为可检测标签的一种或多种荧光染料。用于修饰蛋白质以包括可检测标签的标准方法是本领域已知的。

[0087] 在各种实施例中,本发明进一步涉及包括编码本发明的融合蛋白的核苷酸序列的一种分离的核酸,包括此类核酸的表达载体,以及携带此类表达载体的宿主细胞。

[0088] III.包括N-内含肽融合蛋白的亲和色谱基质

[0089] 在此描述的包含N-内含肽多肽和N-内含肽溶解配偶体的融合蛋白尤其是作为亲和色谱应用的配体具有效用。因此,在某些实施例中,本发明提供包括融合蛋白的亲和色谱基质,该融合蛋白包括N-内含肽多肽和N-内含肽溶解配偶体,并且附接至固体支持物。

[0090] 在一个具体实施例中,该固体支持物是色谱树脂。在某些实施例中,该色谱树脂包

括亲水性聚乙烯醚基。具有亲水性聚乙烯醚基的适合的色谱树脂包括但不限于，ESHMUNO®树脂(EMD密理博公司(Millipore Corporation))。

[0091] 在另一个实施例中，该色谱树脂是合成的基于甲基丙烯酸酯的聚合体培养基(例如，具有粒度在约20-40μm或约40-90μm范围内的珠粒)。在一些实施例中，该色谱树脂具有羧酸功能性。具有羧酸功能性的适合的色谱树脂包括但不限于，FRACTOGEL®COO树脂(EMD密理博公司(Millipore Corporation))。

[0092] 用于本发明的亲和色谱基质的其他适合的固体支持物可以包括，例如，可控孔度玻璃、硅石、氧化锆、氧化钛、琼脂糖、聚甲基丙烯酸酯、聚丙烯酸酯、聚丙烯酰胺、聚乙烯醇、和聚苯乙烯，以及其衍生物(例如，其合金)。

[0093] 用作固体支持物的多孔材料可以由亲水性化合物、疏水性化合物、疏油性化合物、亲油性化合物、或其任何组合组成。该多孔材料可以由聚合物或共聚物组成。适合的多孔材料的实例，包括但不限于聚醚砜、聚酰胺(例如尼龙)、多糖(例如像，琼脂糖和纤维素)、聚丙烯酸酯、聚甲基丙烯酸酯、聚丙烯酰胺、聚甲基丙烯酰胺、聚四氟乙烯、聚砜、聚酯、聚偏二氟乙烯、聚丙烯、聚乙烯、聚乙烯醇、聚碳酸酯、碳氟化合物的聚合物(例如聚(四氟乙烯-共全氟(烷基乙烯醚)))、玻璃、硅石、氧化锆、二氧化钛、陶瓷、金属及其合金。

[0094] 多孔材料可以由有机或无机分子或有机和无机分子的组合组成，并且可以由一个或多个官能团组成，例如羟基基团、硫醇基团、氨基基团、羰基基团、或羧酸基团，其适合用于反应，例如形成共价键用于进一步的化学修饰，以便共价结合至蛋白质。在另一个实施例中，该多孔材料可以不具有官能团，但是可以用带有官能团(如羟基基团、硫醇基团、氨基酸基团、羰基基团、或羧酸基团)的材料层进行涂覆。

[0095] 在一些实施例中，例如，使用有机性质和基于聚合物的常规亲和分离基质，该常规亲和分离基质将亲水性表面暴露于所使用的水介质，例如，暴露在它们的外部以及(如果存在的话)还有在内表面上的羟基(-OH)、羧基(-COOH)、羰基(-CHO，或RCO-R')、甲酰氨(-$CONH_2$，可能以N-取代的形式)、氨基(-NH_2，可能以取代的形式)、寡-或聚乙烯氧基基团。在一个实施例中，该聚合物可以，例如基于多糖，如右旋糖酐、淀粉、纤维素、支链淀粉、琼脂糖等，其有利地交联，例如与双环氧化物、表卤代醇、烯丙基溴、烯丙基缩水甘油醚、1,2,3-三卤取代的低级烃，以提供合适的孔隙率和刚度。在另一个实施例中，该固体支持物包括多孔的琼脂糖珠粒。在本发明中使用的各种支持物可以根据本领域已知的标准方法容易地制备，例如像，描述的反向悬浮凝胶，例如，在杰藤(Hjerten)，生物化学和生物物理学报(Biochim Biophys Acta)79(2)，393-398(1964)。可替代地，这些碱性基质可以是市售的产品，如SEPHAROSE™快速流动(FastFlow)介质(GE医疗集团(GE Healthcare)，瑞典乌普萨拉市)。在一些实施例中，对于大规模分离特别有利，该支持物适于增加其刚度，并且因此使得基质更适合于高流速。

[0096] 可替代地，该固体支持物可以基于合成的聚合物，如聚乙烯醇、聚羟烷基丙烯酸酯、聚羟烷基甲基丙烯酸酯、聚丙烯酰胺、聚甲基丙烯酰胺等。在疏水性聚合物如基于二乙烯基和单乙烯基取代的苯的基质的情况下，经常将基质的表面亲水化以将如上定义的亲水基团暴露于周围的水性液体。此类聚合物可以根据标准方法容易地产生，参见例如，阿莎迪(Arshady)，化学工业(Chimica e L'Industria)70(9)，70-75(1988)。可替代地，可以使用市售的产品，如SOURCE™(GE医疗集团(GE Healthcare)，瑞典乌普萨拉市)和POROS树脂(应

用生物系统,福斯特市,加利福尼亚州)。

[0097] 在又其他实施例中,该固体支持物包括无机性质的支持物,例如,硅石、氧化锆、氧化钛、及其合金。通常修饰无机基质的表面来包括适合的反应基团。实例包括CM氧化锆(赛弗吉(Ciphergen)-拜尔赛普拉(BioSepra)(蓬图瓦兹(Cergypontoise),法国))和CPG®支持物(密理博公司(Millipore Corporation))。

[0098] 在一些实施例中,该固体支持物可以,例如,基于处于可控孔度玻璃形式的氧化锆、二氧化钛、或硅石,其可以被修饰为包含反应基团和/或维持腐蚀性浸泡,以与配体偶联。

[0099] 示例性固体支持物形式包括但不限于,珠粒(球形的或不规则的)、中空纤维、实心纤维、衬垫、凝胶、膜、盒、柱、薄片、载片、平板、或整段材料。

[0100] 在一个实施例中,关于基质的形式,它是处于多孔的整段材料的形式。在可替代的实施例中,该基质可以是多孔的或无孔的珠粒或颗粒形式。可以将珠粒状或颗粒形式的基质用作填充床或悬浮形式。悬浮形式包括称为膨胀床和纯悬浮液的那些,其中颗粒或珠粒可自由移动。在整段材料、填充床、和膨胀床的情况下,分离过程通常跟随浓度梯度的常规色谱。在纯悬浮液的情况下,将使用分批模式。此外,可以使用处于如表面、薄片、毛细管、或过滤器形式的固体支持物。

[0101] 该基质还可以处于在药筒中的膜的形式。该膜可以处于平板、螺旋、或中空纤维的形式。

[0102] 在某些实施例中,该固体支持物可以是可溶支持物,例如,可溶聚合物或水溶性聚合物。示例性可溶支持物包括但不限于,生物聚合物,如像蛋白质或核酸。该聚合物还可以是合成的可溶聚合物,例如像,包括但不限于含有带负电的基团(羧基或磺酸),带正电的基团(季铵、叔胺、仲或伯基团),疏水性基团(苯基或丁基基团),亲水性基团(羟基、或氨基基团)或上述各项的组合的聚合物。示例性合成的可溶聚合物可以在国际PCT公开号WO 2008091740和美国公开号US 20080255027中发现,将其中每个的全部教示内容通过引用并入本文。

[0103] 在一些实施例中,该固体支持物可以包括亲和素分子(例如,链霉亲和素),并且该N-内含肽融合蛋白可以包括生物素标记(例如,共价地附接至融合蛋白中溶解配偶体的生物素分子),这样使得通过亲和素和生物素分子的相互作用实现融合蛋白与固体支持物的结合。

[0104] 本发明的N-内含肽融合蛋白可以附接至在融合蛋白中的仅一个位点处的(单点附接)或融合蛋白中多于一个位点处的(多点接触)固体支持物。优选地,当融合蛋白附接至固体支持物时,在该融合蛋白中的N-内含肽多肽取向为远离该固体支持物。例如,可以将独特的反应氨基酸基团(例如,半胱氨酸残基)定位于在远离N-内含肽结构域的活性区域的位置处的溶解配偶体中,以确保N-内含肽被引导远离固体支持物。

[0105] 优选地,在融合蛋白中参与附接至固体支持物的一个或多个位点(例如,独特的反应氨基酸基团)仅位于N-内含肽溶解配偶体中。因此,为了实现这一点,可能需要修饰N-内含肽多肽以去除提供独特反应位点(例如,半胱氨酸)的氨基酸,例如通过在N-内含肽各处存在的此类氨基酸的缺失或取代。缺失或取代蛋白质中的氨基酸的方法是本领域熟知的。

[0106] 固定的N-内含肽融合蛋白可以适用于柱或多孔色谱分离,或者可以是顺磁性的,

这样使得其可以通过应用磁场从溶液中捕获。

[0107] 可以使用任何合适的技术将在此所述的融合蛋白附接到支持物上，例如，包括本领域已知的和在此所述的那些的固体支持物。例如，在一些实施例中，可以将该融合蛋白通过使用例如融合蛋白中存在的硫醇、氨基、和/或羧基基团的常规偶联技术附接至支持物上。例如，双环氧化物、表氯醇、CNBr、N-羟基琥珀酰亚胺(NHS)等是熟知的偶联剂。在一些实施例中，在支持物和融合蛋白之间引入间隔子，这改进了融合蛋白的可用性并促进融合蛋白与支持物的化学偶联。

[0108] 将N-内含肽融合蛋白附接至固体支持物可以通过许多不同的方法来实现，大多数的方法是本领域熟知的，以及在此描述的那些。参见，例如，霍曼森(Hermanson)等人，固定的亲和配体技术(Immobilized Affinity Ligand Techniques)，学术出版社(Academic Press)，第51-136页(1992)。例如，蛋白质配体可以通过固体支持物或蛋白质配体表面上的活性基团(例如像，羟基、硫醇、环氧化物、氨基、羰基、环氧化物、或羧酸基团)偶联到固体支持物上。附接可以使用已知的化学物质来实现，包括但不限于使用溴化氰(CNBr)、N-羟基琥珀酰亚胺酯、环氧(双环氧乙烷)活化，以及还原胺化。

[0109] 在一个具体实施例中，使用具有羧酸(-COOH)或氨基(-NH_2)基团的色谱树脂(例如，珠粒)。在另外实施例中，该色谱树脂还具有可以转化为-COOH或-NH_2或-OH的羟基(-OH)基团和/或其他的官能团。

[0110] 在一些实施例中，可以使用硫醇引导的蛋白质偶联将本发明的N-内含肽融合蛋白附接至固体支持物。硫醇引导的蛋白质偶联已经在文献中进行了描述。参见，例如，金奎斯(Ljungquist)等人，欧洲生物化学杂志(Eur.J.Biochem.)第186卷，第558-561页(1989)。已知马来酰亚胺在pH 7.0-7.5下与硫醇基团选择性反应。在pH>8情况下，它们还可以与胺基团反应，并且另外，趋于水解(格雷格T.(Greg T.)霍曼森(Hermanson)，生物偶联技术(Bioconjugation Techniques)，学术出版社(Academic Press)，2008;伊恩约翰逊(Ian Johnson)，米歇尔T.Z.(Michelle T.Z.)斯彭思(Spence)，分子探针手册(Molecular Probes Handbook)，荧光探针和标记技术指南(A Guide to Fluorescent Probes and Labeling Technologies)，2010)。在低于pH 8情况下，碘乙酰胺对硫醇基团还具有高度选择性(格雷格T.(Greg T.)霍曼森(Hermanson)，生物偶联技术(Bioconjugation Techniques)，学术出版社(Academic Press)，2008;伊恩约翰逊(Ian Johnson)，米歇尔T.Z.(Michelle T.Z.)斯彭思(Spence)，分子探针手册(Molecular Probes Handbook)，荧光探针和标记技术指南(A Guide to Fluorescent Probes and Labeling Technologies)，2010)。然而，碘乙酰胺在光下本质上是不稳定的，并且大多数市售的接头不是水溶性的和/或非常昂贵的。因为碘乙酰胺对硫醇基团的选择性不优于马来酰亚胺，马来酰亚胺通常是大规模制造的更好选择。

[0111] 在一些实施例中，N-内含肽配体可以通过溶解结构域中的单个可用的巯基基团偶联至AMP或碘乙酰胺活化的FG-COO。衍生的树脂的配体密度可以通过测量带有来自溶液的C-内含肽的融合蛋白的耗尽来计算。到目前为止，已经实现了1gm/升的FG-COO的未优化的N-内含肽配体密度。

[0112] 许多蛋白质也已成功地偶联至环氧活化的树脂上，如FRACTOGEL®环氧。环氧化物与伯氨基基团、羟基、和巯基基团反应，并产生非常稳定的亲和基质(PV库兹涅佐夫(PV

Kuznetsov)1993,药物化学杂志(Pharmaceutical Chemistry Journal)27:439-52)。

[0113] 在一些实施例中,这些蛋白质配体可以通过中介接头偶联至固体支持物。该接头可以包括至少一个偶联至连接部分的官能团。该连接部分可以包括能够偶联至官能团的任何分子。例如,该连接部分可以包括任何烷基、烯基、或炔基基团。该连接部分可以包括范围从1个至30个碳原子的碳链。在一些实施例中,该接头可以由超过30个碳原子组成。该连接部分可以包括至少一个杂原子,如氮、氧、和硫。该连接部分可以由支链、非支链、或环链组成。该连接部分可以被两个或更多个官能团取代。

[0114] 选择用于将蛋白质配体偶联至固体支持物的合适的缓冲条件完全在本领域技术人员的能力范围内。合适的缓冲液包括任何不含胺的缓冲液,如碳酸盐、碳酸氢盐、硫酸盐、磷酸盐、和乙酸盐缓冲液。当使用缔合化学时,缓冲液的盐浓度将取决于所使用的缔合基团。例如,该盐浓度可以在5nM-100mM的范围内。在使用带电物质时,该盐浓度可以是至少5nM,但不小于0.1M,至少5nM,但不小于0.01M,至少5nM,但不小于0.001M。在某些实施例中,该盐浓度可以是0.01M。在使用疏水性物质时,通常希望高盐浓度。因此,盐浓度可以大于0.001M、大于0.01M、或大于0.1M。

[0115] 在一些实施例中,当使用缔合化学时,在范围从0℃至99℃的温度下进行该反应。在某些实施例中,将该反应方法在低于60℃、低于40℃、低于20℃、或低于10℃的温度下实施。在一些实施例中,在约4℃的温度下实施本发明的方法。在其他实施例中,在20℃的温度下实施本发明的方法。

[0116] 在其他实施例中,该N-内含肽融合蛋白可以与各种调节物(膜、聚合体表面、荧光或其他检测标签)组合,与合适的交联或缩合化学物质组合,以形成包括N-内含肽融合蛋白和调节物的共价加和物。

[0117] IV.使用本发明的基于内含肽的融合蛋白的方法

[0118] 在此描述的包含N-内含肽多肽和N-内含肽溶解配偶体的融合蛋白,和包括此类融合蛋白的亲和色谱基质尤其在如下各项中具有效用:如在此进一步描述的亲和纯化方法、筛选适合用于亲和纯化方法的活性断裂内含肽复合体的方法、以及肽裂解和连接方法。

[0119] 因此,在某些实施例中,本发明涉及亲和纯化样品中的靶分子的方法。在本实施例的一个方面,该方法包括a)提供一种包含第一融合蛋白的样品,该第一融合蛋白包括通过肽键连接至靶分子的C-内含肽多肽;b)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成一种无活性的内含肽复合体的条件下,将该样品与包括第二融合蛋白的亲和色谱基质接触,其中该第二融合蛋白包括通过肽键结合至N-内含肽溶解配偶体的N-内含肽多肽,该N-内含肽溶解配偶体促进该N-内含肽多肽的溶解度;c)洗涤该包含无活性的内含肽复合体的亲和色谱基质,以去除未结合的污染物;d)将该内含肽复合体暴露在该内含肽复合体具有活性并且从C-内含肽多肽裂解靶分子的条件下;并且e)回收该经裂解的靶分子。

[0120] 包括连接至N-内含肽溶解配偶体的N-内含肽多肽的融合蛋白可以是在本文的其他地方被描述的任何N-内含肽融合蛋白。

[0121] 包含该融合蛋白的样品可以是任何适合的样品(例如,生物学样品),该融合蛋白连接至靶分子的C-内含肽多肽。在一个实施例中,该样品是粗蛋白制剂或混合物(例如,细胞提取物)。

[0122] 该靶分子可以是任何感兴趣的生物分子。通过举例，感兴趣的生物分子可以包括蛋白质和生物分子集合(例如，由重组DNA技术产生的)，例如像，激素(例如胰岛素、人生长激素、红细胞生成素、干扰素、粒细胞集落刺激因子、组织纤溶酶原活化物)，单克隆抗体(mAb)和mAb-衍生物(例如，双特异性mAb、Fab、scFv、鲨鱼和骆驼抗体)，支架衍生的治疗法(例如，DARPin、亲合体、抗体模拟物(anticalin))，治疗性酶(例如，α半乳糖苷酶A、α-L-艾杜糖醛酸酶、N-乙酰半乳糖胺-4-硫酸酯酶、葡糖脑苷脂酶)，毒素(例如，肉毒杆菌、CRM197、蓖麻毒素)，重组疫苗(例如，炭疽、白喉、破伤风、肺炎、乙型肝炎病毒、人乳头状瘤病毒)，病毒样颗粒(例如，乙型肝炎、人乳头状瘤、流行性感冒、细小病毒、诺瓦克病毒)，以及工业酶(例如，木瓜蛋白酶、菠萝蛋白酶、胰蛋白酶、蛋白酶K、BENZONASE™酶、DENERASE™酶、脲酶、胃蛋白酶等)，以及诊断试剂(例如，葡萄糖和乳酸脱氢酶、DNA聚合酶、碱性磷酸酶、辣根过氧化物酶、限制性酶、杂交瘤衍生的抗体等)。在一个具体实施例中，该靶分子是治疗靶标的抗体(例如，单克隆抗体)。

[0123] 取决于所使用的特定内含肽(例如，集胞藻属物种(Ssp)DnaB、点状念珠藻(Npu)DnaE、GP41-1)，加载、洗涤、裂解、和洗脱条件显著不同。尽管如此，针对特定内含肽的合适的加载、洗涤、裂解、和洗脱条件可以由本领域的普通技术人员容易地确定。针对特定内含肽的合适的条件(例如离液剂和还原剂/氧化剂的浓度，金属离子(例如锌、钙、锶、镁、锰)，体积排除剂(例如PEG、PVP、右旋糖酐)，洗涤剂，盐，温度，和pH)包括但不限于下文所述的条件。特别是对于GP41-1内含肽，已知活性相对不受6-10范围内的pH的影响(卡瓦杰-维拉乔斯P(Carvajal-Vallejos P.)，等人，生物化学杂志(J.Biol.Chem.)287:28686-28696(2012))。

[0124] 第一融合蛋白中的C-内含肽多肽选择性地结合至第二融合蛋白中的N-内含肽多肽以形成无催化活性的内含肽复合体的条件可以由本领域的普通技术人员针对给定的内含肽而确定。通常，对于工业规模的方法，在色谱分离期间改变温度被认为是不切实际的，因为改变将导致长时间的平衡步骤，以确保柱和填充温度在整个过程中是均匀的。示例性结合条件包括a)在约4℃-25℃的范围内的温度，和包括50mM Tris/HCl、300mM NaCl、1mM EDTA、10％(v/v)甘油、2mM DTT的缓冲液，pH＝7(例如，对于GP41-1，参见卡瓦杰-维拉乔斯P(Carvajal-Vallejos P.)等人，生物化学杂志(J.Biol.Chem.)287:28686-28696(2012))；b)在约4℃-25℃的范围内的温度，和包括50mM NaAc、0.5M NaCl的缓冲液，pH＝5(例如，对于DnaB内含肽，参见刘W.(Lu W.)等人，染色体杂志(J.Chrom.)A,1218:2553-2560(2011))；以及c)在约4℃-25℃的范围内的温度，和包括0.5M NaCl、10mM Tris-HCl、0.5mm氯化锌的缓冲液，pH＝8(例如，对于Npu DnaE，参见甘D.(Guan D.)等人，生物技术与生物工程(Biotech.Bioeng.)110:2471-2481(2013))。

[0125] 类似地，促进内含肽复合体的催化活性的条件可以根据所使用的内含肽而变化，并且可以由本领域普通技术人员确定。用于促进催化内含肽活性的示例性条件包括a)包括50mM Tris-HCl(pH＝7.0)、300mM NaCl、1mM EDTA的缓冲液；b)包括0.3M L-精氨酸、5mM EDTA、50mM磷酸缓冲液的缓冲液，pH＝6.5；以及c)包括0.5M NaCl、10mM Tris-HCl、50mM DTT的缓冲液，pH＝8.0。

[0126] 在本实施例的一个方面，该方法可以进一步包括清洁、再生和/或存储本发明的亲和色谱基质。通常，在碱性或酸性条件下(取决于基质的组成)可以洗涤亲和色谱基质。用于

清洁、再生、恢复、和/或存储亲和基质的适当的条件可以通过本领域的普通技术人员确定。

[0127] 在此披露的图1和实例10中提供了本发明的示例性亲和纯化方法。

[0128] 在又另一个实施例中，本发明涉及筛选适合在亲和纯化中使用的有催化活性的内含肽复合体的方法。在本实施例的一个方面，该方法包括a)在该第一融合蛋白中的C-内含肽多肽选择性地结合至该第二融合蛋白中的N-内含肽多肽、从而形成内含肽复合体的条件下，将包括通过肽键连接至靶分子的C-内含肽多肽的第一融合蛋白与包括通过肽键连接至N-内含肽溶解配偶体的N-内含肽多肽的第二融合蛋白接触；并且b)在支持内含肽活性的条件下，确定该靶分子是否从该C-内含肽多肽裂解下来，其中该经裂解的靶分子的存在表明有催化活性的内含肽复合体存在。

[0129] 在本实施例的方法中使用的N-和C-内含肽可以是任何对互补的断裂内含肽，例如像，在此披露的断裂内含肽对(例如，GP41-1N-内含肽和C-内含肽)。

[0130] 第一融合蛋白中的C-内含肽多肽选择性地结合至第二融合蛋白中的N-内含肽多肽以形成无催化活性的内含肽复合体的条件可以取决于使用的内含肽而变化，并且可以由本领域的普通技术人员来确定。示例性结合条件包括a)在约4℃-25℃范围内的温度，和包括100mM Tris-HCl、25mM NaCl、0.1mM氯化锌的缓冲液，pH＝9；b)在约4℃-25℃范围内的温度，和包括50mM NaAc、0.5M NaCl的缓冲液，pH＝5；以及c)在约4℃-25℃范围内的温度，和包括0.5M NaCl、10mM Tris-HCl的缓冲液，pH＝8。

[0131] 该靶分子可以是任何适合的靶分子，包括但不限于在此披露的任何靶分子。

[0132] 实例

[0133] 实例1：选择GP41-1N-内含肽候选溶解配偶体的表征

[0134] 使用4000+种已知的埃希氏菌属蛋白质的组作为起始点，使用以下标准选择七种增溶配偶体(参见表2，SEQ ID NO:11-15、19、20)用于测试：

[0135] (1)所选择的蛋白质缺乏半胱氨酸残基；

[0136] (2)当在大肠杆菌中过量表达时，计算机模拟预测经选择的蛋白质是可溶的；

[0137] (3)所选择的蛋白质具有小于11kDa的分子量；

[0138] (4)计算机模拟预测或已知不分泌所选择的蛋白质；

[0139] (5)当有关于蛋白质相互作用的信息可用时，所选择的蛋白质是单体的而不是多聚体的；

[0140] (6)在有关于蛋白质功能的信息可用时，所选择的蛋白质本质上不是调节性或毒性的，意味着它们不参与主要细胞途径的控制或可能导致过量表达它们(例如核酸酶，聚合酶等)的大肠杆菌的死亡；和

[0141] (7)具有已知的NMR或X射线晶体结构的蛋白质是有利的。

[0142] 表3提供了针对在本研究中评估的内含肽和溶解配偶体的物理特性(分子量(mw)、等电点pH(pI)、在大肠杆菌中的可溶表达的概率，蛋白质是否被预测在大肠杆菌中分泌、总平均疏水性(GRAVY)、和脂肪族氨基酸指数(AI))，其使用公开可用的算法来计算。使用生物信息学工具的SwissProt ExPASy程序组的一部分的在线ProtParam工具(http://web.expasy.org/tools/protparam/)来计算除了在大肠杆菌中过量表达时的溶解度概率和分泌可能性的预测以外的所有物理参数。该分子量以道尔顿给出。针对每种蛋白质的pI是在蛋白质没有净电荷时的pH值。等电点(pI)是蛋白质的净电荷是0时的pH。蛋白质的情况

下,等电点主要取决于七种带电荷的氨基酸:谷氨酸(δ-羧基基团)、天冬氨酸(β-羧基基团)、半胱氨酸(硫醇基团)、酪氨酸(苯酚基团)、组氨酸(咪唑侧链)、赖氨酸(ε-铵基基团)以及精氨酸(胍基基团)。另外,人们应该考虑蛋白质端基团的电荷(NH2i COOH)。它们各自具有称为pK的独特的酸性解离常数。此外,蛋白质的净电荷与溶液(缓冲液)pH紧密相关。牢记这一点,我们可以使用亨德森-哈萨巴奇(Henderson-Hasselbalch)方程计算在某些pH下的蛋白质电荷:

[0143]

氨基酸	NH2	COOH	C	D	E	H	K	R	Y
pKa(维基百科)	8.2	3.65	8.18	3.9	4.07	6.04	10.54	12.48	10.46

[0144] 将线性多肽序列的总平均亲水性(GRAVY)(凯特J(Kyte J)和杜利特尔RF.(Doolittle RF.),分子生物学杂志(J.Mol.Biol.)157:105,1982)计算为所有氨基酸的亲水性值的总和,除以该序列中的残基数目。增加的正评分表明更高的疏水性。该计算是基于凯特-杜利特尔(Kyte-Doolittle)指数。GRAVY是用于显示蛋白质的亲水性状的简单方法。

[0145]

丙氨酸	1.8	亮氨酸	3.8
精氨酸	-4.5	赖氨酸	-3.9
天冬酰胺	-3.5	甲硫氨酸	1.9
天冬氨酸	-3.5	苯丙氨酸	2.8
半胱氨酸	2.5	脯氨酸	-1.6
谷氨酰胺	-3.5	丝氨酸	-0.8
谷氨酸	-3.5	苏氨酸	-0.7
甘氨酸	-0.4	色氨酸	-0.9
组氨酸	-3.2	酪氨酸	-1.3
异亮氨酸	4.5	缬氨酸	4.2

[0146] 蛋白质脂肪族氨基酸指数(伊凯AJ.(Ikai,AJ.),生物化学杂志(J.Biochem.)88:1895,1980)被定义为由脂肪族侧链(丙氨酸、缬氨酸、异亮氨酸、和亮氨酸)占据的相对体积。它可以被认为是球状蛋白的热稳定性增加的正因子。根据以下公式计算蛋白质的脂肪族氨基酸指数:脂肪族氨基酸指数=X(Ala)+a*X(Val)+b*(X(Ile)+X(Leu))。*系数a和b是缬氨酸侧链(a=2.9)和Leu/Ile侧链(b=3.9)相对于丙氨酸侧链的相对体积。在大肠杆菌中过量表达时产生可溶产物的概率还可以使用威尔金森(Wilkinson)和哈里森(Harrison)的算法(威尔金森DL(Wilkinson DL)和哈里森RG(Harrison RG),生物/技术(Bio/Technology),9:443,1991)来计算。其他可用的算法不一定给出类似的结果。

[0147] 对蛋白质是否包含功能性分泌信号的预测使用可从丹麦技术大学的生物序列分析中心获得的SignalP 4.1算法来进行(http://genome.cbs.dtu.dk/services/SignalP/)。

[0148] 表3.针对本研究中使用的内含肽和N-内含肽溶解配偶体所测量的物理参数

[0149]

蛋白质	分子量(D)	PI	在大肠杆菌中的概率溶解度	分泌的	GRAVY	脂肪族氨基酸指数
溶解配偶体 46	5639.4	5.8	74	无	-0.086	95.7
溶解配偶体 206	8799.8	5.6	97	无	-0.636	93.9
溶解配偶体 246	9385.6	5.4	84	无	-0.38	89.3
溶解配偶体 51	6024.8	4.5	98	无	-0.391	117.7
溶解配偶体 138	8325.2	5.4	69	无	-1.526	48.0
溶解配偶体 138_GKL22GCKL	8428.3	5.4	60	无	-1.469	47.3
溶解配偶体 138_GYQ48GCYQ	8428.3	5.4	60	无	-1.469	47.3
溶解配偶体 138_GYQ48GCGYQ	8485.4	5.4	60	无	-1.454	46.6
溶解配偶体 342	10000	4.3	98	无	-0.148	93.4
溶解配偶体 368	10750.2	6.5	72	无	-0.398	95.5
GP41-1 N-内含肽变体 NINTΔA_CC	10589.1	5.8	21	无	-0.44	80.1
GP41-1 C-内含肽(CINT)	4916.7	5.1	86	无	-0.312	111.4
具有硫氧还蛋白的 GP41-1 C-内含肽融合物(CINT_TRX)	17528.1	5.5	78	无	-0.189	101.35
溶解配偶体 A	20504.9	4.6	34	无	-0.14	91.5
溶解配偶体 B	7597.6	5.6	51	无	-0.606	96.0
溶解配偶 C	1281.5	5.4	94	无	0.127	123.6
溶解配偶体 D	37389.1	6.3	42	无	-0.388	77.9
溶解配偶体 E	12740.2	5.8	41	无	-0.299	60.4
溶解配偶体 F	43387.6	5.5	40	无	-0.251	85.4
溶解配偶体 G	12629.4	5.7	74	无	-0.143	97.6
溶解配偶体 H	54870.9	4.5	95	无	-0.278	98.0

[0150] 实例2:大肠杆菌蛋白表达构建体的建立

[0151] 带有针对潜在溶解配偶体46、206、和246的编码序列的质粒构建体通过NINTΔA_CC的氨基或羧基端的氨基酸融合至针对NINTΔA_CC的编码序列,并从DNA2.0插入到pJ414的形式中。使用常规的方法将这些构建体转化为感受态的BL21De3大肠杆菌细胞,并且分离氨苄青霉素抗性菌落。使用SDS聚丙烯酰胺电泳(SDS PAGE)确认预期大小的蛋白质的产生。

[0152] 将6种构建体中的每一种的转化体在包含来自用相应构建体转化的BL21DE3大肠杆菌细胞的甘油储液的100ug氨苄青霉素/mL(LB+Amp)的2mL LB中进行培养。该预接种物在37℃和250rpm下生长过夜,并用于接种200mL的LB+Amp(1%接种物)。将培养物在37℃和250rpm下孵育至OD_{600}介于0.5-0.6之间。通过添加0.4mM IPTG来诱导蛋白质表达。温度降低至30℃,并且将该培养物在这一温度和250rpm下孵育5小时。在这段时间后,通过离心(4500g,25min,4℃)收集细胞,弃去上清液,并且将细胞沉淀保持在-80℃下用于进一步的蛋白质纯化。

[0153] 将针对测试底物蛋白CINT_TRX的编码区域克隆到pSABAD92A(基因库登录号HM070247)中,并且转化到感受态BL21DE3细胞中。在卢里亚肉汤培养基加50ug/ml羧苄青霉素(LB+C)上分离成功的转化体。使用SDS PAGE确认预期大小的蛋白质的产生。将三个BL21克隆/构建体中的每一个的甘油储液存储在-80℃下。

[0154] 在37℃,250rpm下,使用少量的冷冻BL21甘油储液来将5ml的培养物接种在LB+C上。第二天,将0.1ml过夜生长的培养物用于接种10ml LB+C,并将该培养物在37℃,250rpm下生长至0.6-0.9的OD_{600}。在28℃、250rpm下用0.02%的阿拉伯糖诱导培养物5小时。诱导后,通过离心(4500g,25min,4℃)收集细胞,弃去上清液,并且将细胞沉淀保持在-80℃下用于进一步的蛋白质纯化。

[0155] 实例3:可溶与不溶的比率和表达的蛋白质总量的确定

[0156] 为了确定每种构建体的表达产量和可溶:不溶的比率,将对应于如上所述培养的等同生物量的生长培养物的等分试样在5000g,4℃下离心15min。弃去培养物上清液后,将细胞重悬浮在200uL的溶解缓冲液中,该溶解缓冲液由50mM Tris(pH 8)、300mM NaCl、0.5%Triton X-100组成。通过超声处理破碎细胞(10次破裂×3,布兰森250超声波仪(Branson 250Sonifier),每个系列之间的时间以允许样品冷却)。为了分离可溶和不可溶级分,将样品在16000g和4℃下离心10min。将可溶级分移至单独的管中,同时通过超声处理(使用与先前超声处理中相同的参数)将不溶级分重悬浮于200μL相同的溶解缓冲液中。

[0157] 在SDS PAGE凝胶后考马斯染色后,使用定量BSA曲线作为参考的光密度分析,定量细胞裂解物中的重组蛋白。将每个克隆的三个不同的样品体积沿着BSA标准曲线(从0.2至1.2ug的6个点)进行加载。使用“Quantity One”(伯乐公司(Biorad))软件通过光密度测定法确定蛋白带的强度。考虑BSA/重组蛋白质分子量比率,针对每种蛋白质应用校正因子。

[0158] 使用计算的消光系数和它们在280nm处计算的吸光度来确定纯化的蛋白质的浓度。

[0159] 实例4:表达的蛋白质的纯化

[0160] 为了纯化用作贯穿本文的裂解底物的C-内含肽融合蛋白CINT_TRX,将表达蛋白质的大肠杆菌细胞重悬浮于包含50mM Tris-HCl(pH=8.0)、300mM NaCl、0.5X CelLytic B(西格玛-奥德里奇公司(Sigma-Aldrich))、以及20mM咪唑的缓冲液中。将细胞在冰上用30%脉冲活性循环(布兰森250超声波仪(Branson 250Sonifier)超声处理20min,并在4℃下,以34500g离心30min。遵循制造商的说明书,将可溶C-内含肽融合物在His-Trap HP(GE医疗集团(GE Healthcare))柱上,从上清液中纯化。将包含经纯化的C-内含肽融合蛋白的经洗脱的级分合并,在2mM DTT的存在下,在裂解缓冲液(50mM Tris-HCl(pH=7.0)、300mM NaCl、1mM EDTA、10%(v/v)甘油;CB)中进行透析,并且以等分试样存储在-80℃下。

[0161] 在天然条件下从具有表达构建体的大肠杆菌细胞中纯化N-内含肽融合物。将细胞沉淀重悬浮在包含100mM Tris-HCl(pH=8.0)、150mM NaCl、和1mM EDTA的缓冲液中。然后将细胞在冰上用30%脉冲活性循环(布兰森250超声波仪(Branson 250Sonifier)超声处理20min,并在4℃下,以34500g离心30min。如上所述,在His-Trap HP柱上通过色谱法纯化N-内含肽融合物的可溶级分。将包含经纯化蛋白质的经洗脱的级分合并,在2mM DTT的存在下,在CB中透析,并且以等分试样存储在-80℃下。

[0162] 实例5:确定所表达蛋白质的裂解动力学

[0163] 如先前描述进行体外反应(卡瓦杰-维拉乔斯P(Carvajal-Vallejos P.)等人,生物化学杂志(J.Biol.Chem.)287:28686,2012)。简言之,将经纯化的N-和C-融合蛋白在相应的测试条件下分别地短暂进行预孵育。通过在裂解缓冲液中,以5μM等摩尔浓度混合互补的N-和C-内含肽融合蛋白开始裂解反应,并且在25℃和37℃下孵育。对于与本发明相关的实

验，裂解配偶体通常是CINT_TRX。以特定的时间间隔取出等分试样，并且通过添加包含8％SDS(w/v)和20％β-巯基乙醇(v/v)的SDS PAGE缓冲液，随后煮沸5分钟终止反应。通过SDS PAGE(来自诺维克斯(Novex)，英杰公司(Invitrogen)，卡尔斯巴德，美国的4％-12％Bis-Tris凝胶)，随后考马斯亮蓝(西格玛(Sigma))染色来定量反应产物。使用Quantity One(伯乐公司(BioRad))程序，经光密度法确定蛋白带的相对强度。不同的裂解产物根据它们的相应分子量进行标准化。从裂解产物和内含肽标记的前体CINT_TRX的比率来计算蛋白质裂解的百分比。使用GraFit软件(艾瑞赛克斯(Erithacus)，素里(Surrey)，英国)通过将数据拟合到方程$P=P_0(1-e^{-kt})$来确定恒定速率(kobs)，其中P是在时间t形成的裂解的C-内含肽融合产物的量，P_0是可以获得的裂解产物的最大量(产量)，e是欧拉常数，并且k是观察到的速率。在如下的假定下，将所有反应作为不可逆，预稳定状态和一级工艺处理：两个互补的内含肽片段快速缔合之后，C-内含肽融合蛋白的裂解像单分子反应一样进行。

[0164] 实例6：确定NINT溶解配偶体的最佳放置和特性

[0165] 针对所有的溶解配偶体，在两个可能的取向(即与NINTΔA_CC的N末端或C末端融合)中产生与潜在溶解配偶体的NINTΔA_CC融合物。将六种所得的构建体在大肠杆菌中表达，并如前所述，关于产生的总量和溶解度，分析产生的蛋白质。另外，将来自每个构建体的蛋白质纯化，并且使用经纯化的CINT_TRX作为底物来表征裂解速率。这个分析的结果显示在图2A和2B中。虽然溶解配偶体与NINTΔA_CC的N末端的融合，针对融合了溶解配偶体与C末端的所有测试的构建体在大肠杆菌中产生了更多量的蛋白质，但当测量裂解率时，观察到相反的趋势。使用不同的断裂内含肽系统所公开的工作，而正在进行中的这些研究表明溶解配偶体和N-内含肽的位置与N-内含肽活性之间的类似关系，其通过参考已知的结构信息来解释，表明当以相反的极性进行融合时，外显肽结构域之间的空间干扰的可能性更大(甘D(Guan D)，拉米雷斯M(Ramirez M)，陈Z(Chen Z.)，生物技术和生物工程(Biotechnol Bioeng.)110:2471,2013)。

[0166] 这项工作扩展到包括在大小和等电点(pI)方面不同于先前表征的溶解配偶体的另外的溶解配偶体51、138、342、和368(参见表2和3)。如先前用溶解配偶体46、206、和246所示，将所有这些融合至NINTΔA_CC的羧基端，以产生具有最高催化活性的融合物。这些构建体如前所述在大肠杆菌中表达、纯化、并针对裂解速率进行分析。将这些分析的结果呈现在图3中。虽然溶解配偶体246明显具有最高的活性，但是对于溶解配偶体138观察到催化活性和可溶表达之间的最佳折中。

[0167] 为了理解在大肠杆菌中表达期间使其有效溶解N-内含肽的溶解配偶体138的特性，表3中针对每个候选溶解的经计算的蛋白质参数与图4中的可溶性效价相关。虽然这些参数没有一个与整体表达强烈相关，但AI和GRAVY值二者显示出与可溶性效价负相关。

[0168] 实例7：选择氨基酸替代NINTΔA_CC中的半胱氨酸残基

[0169] 从天然来源分离的GP41-1N-内含肽包含三个半胱氨酸残基，但其中之一先前已经被替代以给出NINTΔA_CC，本发明的亲本构建体。NINTΔA_CC中包含的剩余的两个半胱氨酸残基被靶向替代，这样使得可以将唯一、反应半胱氨酸残基引入溶解结构域中用于随后的固定或其他修饰。

[0170] 为了鉴定可以取代NINTΔA_CC中的两个半胱氨酸残基，并且仍然产生稳定的和功能性的内含肽蛋白质的氨基酸，进行了若干种系统发育分析，其中将蛋白质序列进行比对，

并且将SEQ ID NO:1的位置65和89处天然存在的氨基酸变体进行了检测。预期用在相似内含肽中的这些位置处存在的其他氨基酸替代GP41-1中天然存在的内部半胱氨酸，将产生功能性的和/或稳定的GP41-1变体蛋白质，因为天然选择已经允许这些变体在自然界中持续存在。当用GP41内含肽类(1、2、3、4、5、6;达萨B.(Dassa B.)等人，核酸研究(Nucl.Acids Res.)，37:2560-2573(2009))的N-内含肽来进行此类分析时，发现在SEQ ID NO:1的位置65和89处的两个半胱氨酸残基是高度保守的，表明取代这些半胱氨酸将不利地影响GP41-1内含肽的活性和/或稳定性。然而，如果扩展分析以包括可能具有或可能不具有内含肽功能的略微更不同的蛋白质，则鉴定出与大肠杆菌磷酸盐调节子的phoH基因具有同源性的许多蛋白质。使用BLAST搜索工具从GenBank获得大约一百种同系物，并使用免费软件工具BioEdit用CLUSTAL算法进行比对(霍尔TA.(Hall TA.)，核酸专题论文集系列(Nucl.Acids.Symp.Ser.)41:95,1999)。将这一分析的结果在图5中示出，其中位置编号是基于NINTΔA_CC(SEQ ID NO:2)。从该分析可以清楚地看出，苏氨酸和丙氨酸经常在位置65处存在，并且赖氨酸、甲硫氨酸和天冬酰胺经常在位置89处存在，表明天然半胱氨酸被这些天然存在的氨基酸取代将产生稳定的蛋白质。

[0171] 实例8:针对最佳特性筛选NINTΔA_CC氨基酸变体

[0172] 如先前所述的，将基于NINTΔA_CC(SEQ ID NO:2)的含有表4中示出的氨基酸取代的构建体进行创建，表达，纯化和针对催化活性来表征。

[0173] 由每种构建体制备的N-内含肽融合蛋白的裂解速率测量在图6中给出。将NINTΔA_CC亲本(+cnt)在左边示出，用于比较。在位置65和89处的氨基酸显示在图的底部。在位置65处的苏氨酸残基产生比亲本具有显著更高活性的N-内含肽融合物。在测试的构建体中，在位置65处具有苏氨酸并且在位置89处具有甲硫氨酸的N-内含肽融合物给出具有比亲本构建体快约三倍的催化速率的构建体。

[0174] 表4.GP41-1N-内含肽变体

[0175]

GP41-1 N-内含肽变体	序列
NINTΔA_CC(SEQ ID NO:2)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**C**SEEHLFPTQTGEMNISGGLKEGM**C**LYVKEgg
NINTΔA_AC (SEQ ID NO:3)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**A**SEEHLFPTQTGEMNISGGLKEGM**C**LYVKEgg
NINTΔA_CK (SEQ ID NO:4)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**C**SEEHLFPTQTGEMNISGGLKEGM**K**LYVKEgg
NINTΔA_AM (SEQ ID NO:5)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**A**SEEHLFPTQTGEMNISGGLKEGM**M**LYVKEgg
NINTΔA_TM (SEQ ID NO:6)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**T**SEEHLFPTQTGEMNISGGLKEGM**M**LYVKEgg
NINTΔA_AK (SEQ ID NO:7)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**A**SEEHLFPTQTGEMNISGGLKEGM**K**LYVKEgg
NINTΔA_TK (SEQ ID NO:8)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYK ITLEDGKEII**T**SEEHLFPTQTGEMNISGGLKEGM**K**LYVKEgg

[0176] 对于SEQ ID NO:2-8，使用小写字母文本指示非内含肽序列，并且由大写字母指示内含肽序列。

[0177] 实例9:用于将独特的半胱氨酸残基引入溶解配偶体138的策略

[0178] 为了允许N-内含肽融合蛋白的化学修饰而不减少其催化活性，最终修饰的位点应该尽可能远离N-内含肽的活性位点。在溶解配偶体138的结构信息不存在的情况下，合理的

方法是将进行系统发育分析，如以上针对GP41-1N-内含肽所述(参见实例7)，确定显示高变异性的蛋白质区域，通过插入半胱氨酸修饰这些，并且然后测试所有得到的构建体。然而，存在可用于溶解配偶体138(蛋白质数据库结构1RYK)的NMR溶液结构，其显示在图7中。该蛋白质包含四个α螺旋结构域，是球状的，具有长的非结构化螺旋，该螺旋形成与N-内含肽的羧基端的连接(圆形区域;N-内含肽未显示)。由黄色突出指示的环形区域GKL和GYQ被靶向用于半胱氨酸残基的插入，以制造溶解配偶体138(138_GKL22GCKL(SEQ ID NO:16)、138_GYQ48GCYQ(SEQ ID NO:17)、和138_GYQ48GCGYQ(SEQ ID NO:18))的新形式(GCKL(SEQ ID NO:61)、GCYQ(SEQ ID NO:62)、和GCGYQ(SEQ ID NO:63))。

[0179] 实例10:N-内含肽融合蛋白(配体)与色谱树脂的偶联

[0180] 包含融合至GP41-1变体NINTΔA_TM(SEQ ID NO:6)的羧基端的溶解配偶体138_GYQ48GCGYQ(SEQ ID NO:18)的可溶融合蛋白从大肠杆菌中的编码核酸表达，随后使用常规的分离方法从污染的细胞蛋白质中将其分离。

[0181] 然后使用标准的技术，通过在融合蛋白的溶解配偶体结构域中的独特的反应半胱氨酸位点，将该经纯化的N-内含肽融合蛋白偶联至FRACTOGEL®或ESHMUNO®色谱树脂(EMD密理博公司(Millipore Corporation))。

[0182] 活化准备中，在pH＝6.5下，在布氏(Buechner)漏斗中，将5ml的湿FRACTOGEL® COO(FG-COO)树脂用DI水洗涤一次，并用150mM 2-(N-吗啉代)乙磺酸(MES缓冲液)洗涤三次，并且转移至肖特(Schott)玻璃瓶中。将0.1035gm 1-乙基-3-(3-二甲氨基-丙基)碳二亚胺(EDC)溶解于3ml的MES缓冲液中，并且添加至FG-COO。将该混合物在室温下孵育2min。添加0.1372gmN-(3-氨基丙基)马来酰亚胺三氟乙酸(APM)在4ml的MES缓冲液中的溶液，并且将该混合物保持在室温下搅拌过夜。通过用1M NaOH滴定，将pH维持在6.5。为了存储，将该活化的树脂重悬浮于包含20％乙醇的150mM NaCl中，并且存储在冰箱中。为了分析功能化，在pH＝7.2下(PO缓冲液)，在包含150mM NaCl的100mM磷酸缓冲液中，制备活化树脂的50％v/v溶液。将0.5ml的活化的FG-COO与在PO缓冲液中的1mL的204uM半胱氨酸盐酸盐溶液混合，并且孵育一小时。将未用AMP活化的FG-COO样品作为阴性对照来平行处理。然后用PO缓冲液，0.5M NaCl将该树脂进行充分洗涤，并且重悬浮于0.5M NaCl中，使用埃尔曼氏试剂(5,5'-二硫-双-(2-硝基苯甲酸))和已知的方法来分析游离的巯基基团。使用该分析确定高达400μmol/克干燥树脂的配体密度。

[0183] 实例11:使用内含肽融合蛋白的硫氧还蛋白的亲和纯化

[0184] 将包含根据在此的实例10制备的固定的N-内含肽融合蛋白的树脂装入标准的色谱柱中，并且在4℃-25℃范围内的温度下，并且使用包含100mM Tris-HCl、25mM NaCl、0.1mM氯化锌(pH＝9)的加样缓冲液，将包含CINT_TRX融合蛋白(SEQ ID NO:10)的粗蛋白混合物(该粗蛋白混合物包括融合至GP41-1C-内含肽的羧基端的靶分子硫氧还蛋白)添加至包含固定的N-内含肽融合蛋白的柱中，以允许GP41-1N-和C-内含肽结构域之间强烈的相互作用，而不允许内含肽催化。

[0185] 然后使用包含洗涤剂(例如Triton X100，ND40)或盐(例如乙酸盐，磷酸盐，氯化物，钠、铵、或钾的硫酸盐)的洗涤缓冲液来洗涤加载的柱，以去除未结合的和弱结合的污染物。

[0186] 通过添加裂解缓冲液(50mM Tris-HCl(pH＝7.0)、300mM NaCl、1mM EDTA)，完成了

C-内含肽融合蛋白的硫氧还蛋白部分的裂解和洗脱。然后将该裂解的硫氧还蛋白在洗脱物中回收。

[0187] 表5.示例性断裂内含肽以及它们的序列

[0188]

内含肽	SEQ ID NO:	序列
GP41.1的N末端结构域	29	CLDLKTQVQT PQGMKEISNI QVGDLVLSNT GYNEVLNVFP KSKKKSYKIT LEDGKEIICS EEHLFPTQTG EMNISGGLKE GMCLYVKE
GP41.8的N末端结构域	30	CLSLDTMVVT NGKAIEIRDV KVGDWLESEC GPVQVTEVLP IIKQPVFEIV LKSGKKIRVS ANHKFPTKDG LKTINSGLKV GDFLRSRA
NrdJ1的N末端结构域	31	CLVGSSEIIT RNYGKTTIKE VVEIFDNDKN IQVLAFNTHT DNIEWAPIKA AQLTRPNAEL VELEINTLHG VKTIRCTPDH PVYTKNRDYV RADELTDDDE LVVAI
IMPDH1的N末端结构域	32	CFVPGTLVNT ENGLKKIEEI KVGDKVFSHT GKLQEVVDTL IFDRDEEIIS INGIDCTKNH EFYVIDKENA NRVNEDNIHL FARWVHAEEL DMKKHLLIEL E
NrdA-2的N末端结构域	33	CLTGDAKIDV LIDNIPISQI SLEEVVNLFN EGKEIYVLSY NIDTKEVEYK EISDAGLISE SAEVLEIIDE ETGQKIVCTP DHKVYTLNRG YVSAKDLKED DELVFS
Npu DNA-E的N末端结构域(基因库登录号ZP_00111398)	34	CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR GEQEVFEYCL EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL PN
Ssp DNA-B的N末端结构域(基因库登录号Q55418)	35	CISGDSLISL ASTGKRVSIK DLLDEKDFEI WAINEQTMKL ESAKVSRVFC TGKKLVYILK TRLGRTIKAT ANHRFLTIDG WKRLDELSLK EHIALPRKLE SSSLQ
GP41.1的C末端结构域	9	MMLKKILKIE ELDERELIDI EVSGNHLFYA NDILTHN
GP41.8的C末端结构域	36	MCEIFENEID WDEIASIEYV GVEETIDINV TNDRLFFANG ILTHN
NrdJ1的C末端结构域	37	MEAKTYIGKL KSRKIVSNED TYDIQTSTHN FFANDILVHN
IMPDH1的C末端结构域	38	MKFKLKEITS IETKHYKGKV HDLTVNQDHS YNVRGTVVHN
NrdA-2的C末端结构域	39	MGLKIIKRES KEPVFDITVK DNSNFFANNI LVHN

[0189]

Npu DNA-E的C末端结构域(基因库登录号ZP_00111398)	40	MIKIATRKYL GKQNVYDIGV ERDHNFALKN GFIASN
Ssp DNA-B的C末端结构域(基因库登录号Q55418)	41	SPEIEKLSQS DIYWDSIVSI TETGVEEVF DLTVPGPHNF VANDIIVHN

[0190] 下划线序列对应该内含肽N末端结构域的N1框。双下划线序列对应该内含肽C末端结构域的C1框(例如,缺少外显肽的第一个氨基酸)。

[0191] 另外的示例性N-内含肽序列

[0192] gp 41-2

[0193] CLDLKTQVQTQQGLKDISNIQVGDLVL(SEQ ID NO:42)

[0194] gp 41-3

[0195] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKS(SEQ ID NO:43)

[0196] gp 41-4

[0197] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLFPTQTGEMNISGGLKEGMCLYVKE(SEQ ID NO:44)

[0198] gp 41-5

[0199] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYKITLEDGKEIICSEEHLFPTQTGEMNISGGLKEGMCLYVKE(SEQ ID NO:45)

[0200] gp 41-6

[0201] SYKITLEDGKEIICSEEHLFPTQNGEVNIKGGLKEGMCLYVKE(SEQ ID NO:46)

[0202] gp 41-7

[0203] MMLKKILKIEELDERELIDIEVSGNH(SEQ ID NO:47)

[0204] NrdA-1

[0205] CVAGDTKIKIKYPESVGDQYGTWYWNVLEKEIQIEDLEDYIIMRECEIYDSNAPQIEVLSYNIETGEQEWKPITAFAQTSPKAKVMKITDEESGKSIVVTPEHQVFTKNRGYVMAKDLIETDEPIIVNKDMNF(SEQ ID NO:48)

[0206] NrdA-4

[0207] CLAGDTTVTVLEGDIVFEMTLENLVSLYKNVFSVSVLSFNPETQKQEFKPVTNAALMNPESKVLKITDSDTGKSIVCTPDHKVFTKNRGYVIASELNAEDILEIK(SEQID NO:49)

[0208] NrdA-5

[0209] HTETVRRVGTITAFAQTSPKSKVMKITDEESGNSIVVTPEHKVFTKNRGYVMAKNLVETDELVIN(SEQIDNO:50)

[0210] NrdA-6

[0211] YVCSRDDTTGFKLICTPDHMIYTKNRGYIMAKYLKEDDELLINEIHLPT(SEQID NO:51)

[0212] NrdJ-1

[0213] CLVGSSEIITRNYGKTTIKEVVEIFDNDKNIQVLAFNTHTDNIEWAPIKAAQLTRPNAELVELEIDTLHGVKTIRCTPDHPVYTKNRGYVRADELTDDDELVVAI(SEQID NO:52)

[0214] NrdJ-2

[0215] CLVGSSEIITRNYGKTTIKEVVEIFDNDKNIQVLAFNTHTDNIEWAPIKAAQLTRPNAELVELEINTLHGVKTIRCTPDHPVYTKNRDYVRADELTDDDELVVAI(SEQID NO:53)

[0216] 另外的示例性C-内含肽序列

[0217] gp 41-9

[0218] MIMKNRERFITEKILNIEEIDDDLTVDIGMDNEDHYFVANDILTHNT(SEQ ID NO:54)

[0219] IMPDH-2

[0220] MKFTLEPITKIDSYEVTAEPVYDIEVENDHSFCVeNGFVVHNS(SEQ ID NO:55)

[0221] IMPDH-3

[0222] MKFKLVEITSKETFNYSGQ-VHDLTVEDDHSYSI-NNIVVHNS(SEQ ID NO:56),

[0223] NrdA-3

[0224] MLKIEYLEEEIPVYDITVEETHNFFANDILIHNC(SEQ ID NO:57),

[0225] NrdA-5

[0226] MLKIEYLEEEIPVYDITVEGTHNLAYSL(SEQ ID NO:58)，

[0227] NrdA-6

[0228] MGIKIRKLEQNRVYDIKVEKIIIFCNNILVHNC(SEQ ID NO:59)，和

[0229] NrdJ-1

[0230] MEAKTYIGKLKSRKIVSNEDTYDIQTSTHNFFANDILVHNS(SEQ ID NO:60)。

[0231] 所有专利、公开的申请以及在此引用的参考文献的相关教示内容都通过引用以其全文进行结合。

[0232] 除非另作说明，本说明书(包括权利要求书)中使用的表达成分的量、表达条件、处理条件等等的所有数值应被理解为在所有情况中被术语“约”修饰。因此，除非另外相反地指明，这些数值参数是近似值，并且可以取决于本发明所寻求获得的所希望的特性而变化。除非另外指明，在一系列元素前面的术语“至少”应被理解为指系列中的每一个元素。本领域的技术人员仅使用常规实验就将认识到或能够确定在此描述的本发明这些具体实施例的许多相等形式。此类等效物旨在由以下权利要求书涵盖。

[0233] 虽然本发明参考其示例实施例已经进行了具体显示和描述，本领域的技术人员应当理解的是，在不偏离由所附权利要求书所涵盖的本发明的范围的情况下，可以在其中做出在形式和细节方面的多种改变。

SEQUENCE LISTING

<110> 默克专利有限公司

<120> 可溶内含肽融合蛋白以及用于纯化生物分子的方法

<130> 0046.2053-002

<150> US 62/074,494
<151> 2014-11-03

<150> US 62/209,010
<151> 2015-08-24

<160> 63

<170> 用于Windows版本4.0的FastSEQ

<210> 1
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽

<400> 1

[0001]

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

<210> 2
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 2

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

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
50 55 60
Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65 70 75 80
Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly
85 90 95

<210> 3
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 3
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
1 5 10 15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
20 25 30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
35 40 45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
50 55 60
Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65 70 75 80
Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly
85 90 95

[0002]

<210> 4
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 4
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
1 5 10 15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
20 25 30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
35 40 45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
50 55 60
Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65 70 75 80
Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
85 90 95

<210> 5
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 5
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
1 5 10 15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
20 25 30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
35 40 45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
50 55 60
Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65 70 75 80
Ser Gly Gly Leu Lys Glu Gly Met Met Leu Tyr Val Lys Glu Gly Gly
85 90 95

<210> 6
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 6
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
[0003] 1 5 10 15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
20 25 30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
35 40 45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
50 55 60
Thr Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65 70 75 80
Ser Gly Gly Leu Lys Glu Gly Met Met Leu Tyr Val Lys Glu Gly Gly
85 90 95

<210> 7
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 7
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
1 5 10 15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
20 25 30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
35 40 45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile

```
    50                  55                  60
Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65                  70                  75                  80
Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
                85                  90                  95
```

<210> 8
<211> 96
<212> PRT
<213> 人工序列

<220>
<223> 具有侧翼非内含肽序列的GP41-1 N内含肽变体

<400> 8

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

[0004]

<210> 9
<211> 42
<212> PRT
<213> 噬藻体

<400> 9

```
Met Gly Lys Asn Ser Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu
 1               5                  10                  15
Leu Asp Glu Arg Glu Leu Ile Asp Ile Glu Val Ser Gly Asn His Leu
            20                  25                  30
Phe Tyr Ala Asn Asp Ile Leu Thr His Asn
        35                  40
```

<210> 10
<211> 157
<212> PRT
<213> 人工序列

<220>
<223> GP41-1 C-内含肽-硫氧还蛋白融合蛋白

<400> 10

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

```
          35                  40                  45
His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
      50                  55                  60
Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
65                  70                  75                  80
Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
                85                  90                  95
Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
            100                 105                 110
Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
          115                 120                 125
Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
      130                 135                 140
Phe Leu Asp Ala Asn Leu Ala His His His His His His
145                 150                 155
```

<210> 11
<211> 51
<212> PRT
<213> 大肠杆菌

<400> 11

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

[0005]

<210> 12
<211> 77
<212> PRT
<213> 大肠杆菌

<400> 12

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

<210> 13
<211> 83
<212> PRT
<213> 大肠杆菌

<400> 13

```
Met Asn Lys Glu Thr Gln Pro Ile Asp Arg Glu Thr Leu Leu Lys Glu
```

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

<210> 14
<211> 53
<212> PRT
<213> 大肠杆菌

<400> 14

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

[0006]

<210> 15
<211> 69
<212> PRT
<213> 大肠杆菌

<400> 15

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

<210> 16
<211> 70
<212> PRT
<213> 大肠杆菌

<400> 16

```
Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1               5                  10                  15
Val Lys Glu Gln Trp Gly Cys Lys Leu Thr Asp Asp Asp Met Thr Ile
            20                  25                  30
Ile Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr
```

```
        35                  40                  45
Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr
    50                  55                  60
Arg Asn Glu Tyr Arg Trp
65                  70
```

<210> 17
<211> 70
<212> PRT
<213> 大肠杆菌

<400> 17

```
Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1               5                  10                  15
Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
            20                  25                  30
Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
        35                  40                  45
Cys Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr
    50                  55                  60
Arg Asn Glu Tyr Arg Trp
65                  70
```

<210> 18
<211> 71
<212> PRT
<213> 大肠杆菌

[0007]

<400> 18

```
Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1               5                  10                  15
Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
            20                  25                  30
Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
        35                  40                  45
Cys Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu
    50                  55                  60
Thr Arg Asn Glu Tyr Arg Trp
65                  70
```

<210> 19
<211> 92
<212> PRT
<213> 大肠杆菌

<400> 19

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

65 70 75 80
Met Trp Glu Asn Leu Phe Gln Gln Ala Ser Gln Gln
85 90

<210> 20
<211> 95
<212> PRT
<213> 大肠杆菌

<400> 20
Met Gln Leu Asn Ile Thr Gly Asn Asn Val Glu Ile Thr Glu Ala Leu
1 5 10 15
Arg Glu Phe Val Thr Ala Lys Phe Ala Lys Leu Glu Gln Tyr Phe Asp
20 25 30
Arg Ile Asn Gln Val Tyr Val Val Leu Lys Val Glu Lys Val Thr His
35 40 45
Thr Ser Asp Ala Thr Leu His Val Asn Gly Gly Glu Ile His Ala Ser
50 55 60
Ala Glu Gly Gln Asp Met Tyr Ala Ala Ile Asp Gly Leu Ile Asp Lys
65 70 75 80
Leu Ala Arg Gln Leu Thr Lys His Lys Asp Lys Leu Lys Gln His
85 90 95

<210> 21
<211> 192
<212> PRT
[0008] <213> 大肠杆菌

<400> 21
Met Asp Thr Ser Asn Ala Thr Ser Val Val Asn Val Ser Ala Ser Ser
1 5 10 15
Ser Thr Ser Thr Ile Tyr Asp Leu Gly Asn Met Ser Lys Asp Glu Val
20 25 30
Val Lys Leu Phe Glu Glu Leu Gly Val Phe Gln Ala Ala Ile Leu Met
35 40 45
Phe Ser Tyr Met Tyr Gln Ala Gln Ser Asn Leu Ser Ile Ala Lys Phe
50 55 60
Ala Asp Met Asn Glu Ala Ser Lys Ala Ser Thr Thr Ala Gln Lys Met
65 70 75 80
Ala Asn Leu Val Asp Ala Lys Ile Ala Asp Val Gln Ser Ser Thr Asp
85 90 95
Lys Asn Ala Lys Ala Lys Leu Pro Gln Asp Val Ile Asp Tyr Ile Asn
100 105 110
Asp Pro Arg Asn Asp Ile Ser Val Thr Gly Ile Ser Asp Leu Ser Gly
115 120 125
Asp Leu Ser Ala Gly Asp Leu Gln Thr Val Lys Ala Ala Ile Ser Ala
130 135 140
Lys Ala Asn Asn Leu Thr Thr Val Val Asn Asn Ser Gln Leu Glu Ile
145 150 155 160
Gln Gln Met Ser Asn Thr Leu Asn Leu Leu Thr Ser Ala Arg Ser Asp
165 170 175
Val Gln Ser Leu Gln Tyr Arg Thr Ile Ser Ala Ile Ser Leu Gly Lys
180 185 190

<210> 22

<211> 68
<212> PRT
<213> 肝片吸虫

<400> 22
Met Pro Ser Val Glu Val Glu Lys Leu Leu His Val Leu Asp Arg Asn
1 5 10 15
Gly Asp Gly Lys Val Ser Ala Glu Glu Leu Lys Ala Phe Ala Asp Asp
20 25 30
Ser Lys Tyr Pro Leu Asp Ser Asn Lys Ile Lys Ala Phe Ile Lys Glu
35 40 45
His Asp Lys Asn Lys Asp Gly Lys Leu Asp Leu Lys Glu Leu Val Ser
50 55 60
Ile Leu Ser Ser
65

<210> 23
<211> 11
<212> PRT
<213> 肝片吸虫

<400> 23
Met Pro Ser Val Glu Val Glu Lys Leu Leu His
1 5 10

<210> 24
[0009] <211> 328
<212> PRT
<213> 大肠杆菌

<400> 24
Met Gly Gln Leu Ile Asp Gly Val Trp His Asp Thr Trp Tyr Asp Thr
1 5 10 15
Lys Ser Thr Gly Gly Lys Phe Gln Arg Ser Ala Ser Ala Phe Arg Asn
20 25 30
Trp Leu Thr Ala Asp Gly Ala Pro Gly Pro Thr Gly Lys Gly Gly Phe
35 40 45
Ala Ala Glu Lys Asp Arg Tyr His Leu Tyr Val Ser Leu Ala Cys Pro
50 55 60
Trp Ala His Arg Thr Leu Ile Met Arg Lys Leu Lys Gly Leu Glu Pro
65 70 75 80
Phe Ile Ser Val Ser Val Val Asn Pro Leu Met Leu Glu Asn Gly Trp
85 90 95
Thr Phe Asp Asp Ser Phe Pro Gly Ala Thr Gly Asp Thr Leu Tyr Gln
100 105 110
His Glu Phe Leu Tyr Gln Leu Tyr Leu His Ala Asp Pro His Tyr Ser
115 120 125
Gly Arg Val Thr Val Pro Val Leu Trp Asp Lys Lys Asn His Thr Ile
130 135 140
Val Ser Asn Glu Ser Ala Glu Ile Ile Arg Met Phe Asn Thr Ala Phe
145 150 155 160
Asp Ala Leu Gly Ala Lys Ala Gly Asp Tyr Tyr Pro Pro Ala Leu Gln
165 170 175
Pro Lys Ile Asp Glu Leu Asn Gly Trp Ile Tyr Asp Thr Val Asn Asn
180 185 190
Gly Val Tyr Lys Ala Gly Phe Ala Thr Ser Gln Gln Ala Tyr Asp Glu

```
        195                 200                 205
Ala Val Ala Lys Val Phe Glu Ser Leu Ala Arg Leu Glu Gln Ile Leu
    210                 215                 220
Gly Gln His Arg Tyr Leu Thr Gly Asn Gln Leu Thr Glu Ala Asp Ile
225                 230                 235                 240
Arg Leu Trp Thr Thr Leu Val Arg Phe Asp Pro Val Tyr Val Thr His
                245                 250                 255
Phe Lys Cys Asp Lys His Arg Ile Ser Asp Tyr Leu Asn Leu Tyr Gly
            260                 265                 270
Phe Leu Arg Asp Ile Tyr Gln Met Pro Gly Ile Ala Glu Thr Val Asn
        275                 280                 285
Phe Asp His Ile Arg Asn His Tyr Phe Arg Ser His Lys Thr Ile Asn
    290                 295                 300
Pro Thr Gly Ile Ile Ser Ile Gly Pro Trp Gln Asp Leu Asp Glu Pro
305                 310                 315                 320
His Gly Arg Asp Val Arg Phe Gly
                325
```

<210> 25
<211> 120
<212> PRT
<213> 肠杆菌噬菌体λ

<400> 25

[0010]

```
Met Ala Ser Trp Ser His Pro Gln Phe Glu Lys Ala Ser Lys Glu Thr
 1               5                  10                  15
Phe Thr His Tyr Gln Pro Gln Gly Asn Ser Asp Pro Ala His Thr Ala
            20                  25                  30
Thr Ala Pro Gly Gly Leu Ser Ala Lys Ala Pro Ala Met Thr Pro Leu
        35                  40                  45
Met Leu Asp Thr Ser Ser Arg Lys Leu Val Ala Trp Asp Gly Thr Thr
    50                  55                  60
Asp Gly Ala Ala Val Gly Ile Leu Ala Val Ala Ala Asp Gln Thr Ser
65                  70                  75                  80
Thr Thr Leu Thr Phe Tyr Lys Ser Gly Thr Phe Arg Tyr Glu Asp Val
                85                  90                  95
Leu Trp Pro Glu Ala Ala Ser Asp Glu Thr Lys Lys Arg Thr Ala Phe
            100                 105                 110
Ala Gly Thr Ala Ile Ser Ile Val
        115                 120
```

<210> 26
<211> 396
<212> PRT
<213> 大肠杆菌

<400> 26

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

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

[0011]

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
340 345 350
Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
355 360 365
Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
370 375 380
Glu Ala Leu Lys Asp Ala Gln Thr Arg Ile Thr Lys
385 390 395

<210> 27
<211> 109
<212> PRT
<213> 大肠杆菌

<400> 27
Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
1 5 10 15
Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20 25 30
Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
35 40 45
Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
50 55 60
Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu

65 70 75 80
Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
85 90 95
Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala
100 105

<210> 28
<211> 495
<212> PRT
<213> 大肠杆菌

<400> 28
Met Asn Lys Glu Ile Leu Ala Val Val Glu Ala Val Ser Asn Glu Lys
1 5 10 15
Ala Leu Pro Arg Glu Lys Ile Phe Glu Ala Leu Glu Ser Ala Leu Ala
20 25 30
Thr Ala Thr Lys Lys Lys Tyr Glu Gln Glu Ile Asp Val Arg Val Gln
35 40 45
Ile Asp Arg Lys Ser Gly Asp Phe Asp Thr Phe Arg Arg Trp Leu Val
50 55 60
Val Asp Glu Val Thr Gln Pro Thr Lys Glu Ile Thr Leu Glu Ala Ala
65 70 75 80
Arg Tyr Glu Asp Glu Ser Leu Asn Leu Gly Asp Tyr Val Glu Asp Gln
85 90 95
Ile Glu Ser Val Thr Phe Asp Arg Ile Thr Thr Gln Thr Ala Lys Gln
100 105 110
Val Ile Val Gln Lys Val Arg Glu Ala Glu Arg Ala Met Val Val Asp
115 120 125
[0012] Gln Phe Arg Glu His Glu Gly Glu Ile Ile Thr Gly Val Val Lys Lys
130 135 140
Val Asn Arg Asp Asn Ile Ser Leu Asp Leu Gly Asn Asn Ala Glu Ala
145 150 155 160
Val Ile Leu Arg Glu Asp Met Leu Pro Arg Glu Asn Phe Arg Pro Gly
165 170 175
Asp Arg Val Arg Gly Val Leu Tyr Ser Val Arg Pro Glu Ala Arg Gly
180 185 190
Ala Gln Leu Phe Val Thr Arg Ser Lys Pro Glu Met Leu Ile Glu Leu
195 200 205
Phe Arg Ile Glu Val Pro Glu Ile Gly Glu Glu Val Ile Glu Ile Lys
210 215 220
Ala Ala Ala Arg Asp Pro Gly Ser Arg Ala Lys Ile Ala Val Lys Thr
225 230 235 240
Asn Asp Lys Arg Ile Asp Pro Val Gly Ala Cys Val Gly Met Arg Gly
245 250 255
Ala Arg Val Gln Ala Val Ser Thr Glu Leu Gly Gly Glu Arg Ile Asp
260 265 270
Ile Val Leu Trp Asp Asp Asn Pro Ala Gln Phe Val Ile Asn Ala Met
275 280 285
Ala Pro Ala Asp Val Ala Ser Ile Val Val Asp Glu Asp Lys His Thr
290 295 300
Met Asp Ile Ala Val Glu Ala Gly Asn Leu Ala Gln Ala Ile Gly Arg
305 310 315 320
Asn Gly Gln Asn Val Arg Leu Ala Ser Gln Leu Ser Gly Trp Glu Leu
325 330 335
Asn Val Met Thr Val Asp Asp Leu Gln Ala Lys His Gln Ala Glu Ala
340 345 350
His Ala Ala Ile Asp Thr Phe Thr Lys Tyr Leu Asp Ile Asp Glu Asp

355 360 365
Phe Ala Thr Val Leu Val Glu Glu Gly Phe Ser Thr Leu Glu Glu Leu
370 375 380
Ala Tyr Val Pro Met Lys Glu Leu Leu Glu Ile Glu Gly Leu Asp Glu
385 390 395 400
Pro Thr Val Glu Ala Leu Arg Glu Arg Ala Lys Asn Ala Leu Ala Thr
405 410 415
Ile Ala Gln Ala Gln Glu Glu Ser Leu Gly Asp Asn Lys Pro Ala Asp
420 425 430
Asp Leu Leu Asn Leu Glu Gly Val Asp Arg Asp Leu Ala Phe Lys Leu
435 440 445
Ala Ala Arg Gly Val Cys Thr Leu Glu Asp Leu Ala Glu Gln Gly Ile
450 455 460
Asp Asp Leu Ala Asp Ile Glu Gly Leu Thr Asp Glu Lys Ala Gly Ala
465 470 475 480
Leu Ile Met Ala Ala Arg Asn Ile Cys Trp Phe Gly Asp Glu Ala
485 490 495

<210> 29
<211> 88
<212> PRT
<213> GP41-1 N-内含肽（噬藻体）

<400> 29
Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
1 5 10 15
Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
20 25 30
[0013] Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys
35 40 45
Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu
50 55 60
Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
65 70 75 80
Gly Met Cys Leu Tyr Val Lys Glu
85

<210> 30
<211> 88
<212> PRT
<213> Unknown

<220>
<223> GP41.8的N端结构域

<400> 30
Cys Leu Ser Leu Asp Thr Met Val Val Thr Asn Gly Lys Ala Ile Glu
1 5 10 15
Ile Arg Asp Val Lys Val Gly Asp Trp Leu Glu Ser Glu Cys Gly Pro
20 25 30
Val Gln Val Thr Glu Val Leu Pro Ile Ile Lys Gln Pro Val Phe Glu
35 40 45
Ile Val Leu Lys Ser Gly Lys Lys Ile Arg Val Ser Ala Asn His Lys
50 55 60
Phe Pro Thr Lys Asp Gly Leu Lys Thr Ile Asn Ser Gly Leu Lys Val
65 70 75 80

Gly Asp Phe Leu Arg Ser Arg Ala
85

<210> 31
<211> 105
<212> PRT
<213> Unknown

<220>
<223> NrdJ1的N端结构域

<400> 31
Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
1 5 10 15
Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
20 25 30
Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
35 40 45
Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
50 55 60
Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
65 70 75 80
Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr
85 90 95
Asp Asp Asp Glu Leu Val Val Ala Ile
100 105

[0014]

<210> 32
<211> 101
<212> PRT
<213> Unknown

<220>
<223> IMPDH1的N端结构域

<400> 32
Cys Phe Val Pro Gly Thr Leu Val Asn Thr Glu Asn Gly Leu Lys Lys
1 5 10 15
Ile Glu Glu Ile Lys Val Gly Asp Lys Val Phe Ser His Thr Gly Lys
20 25 30
Leu Gln Glu Val Val Asp Thr Leu Ile Phe Asp Arg Asp Glu Glu Ile
35 40 45
Ile Ser Ile Asn Gly Ile Asp Cys Thr Lys Asn His Glu Phe Tyr Val
50 55 60
Ile Asp Lys Glu Asn Ala Asn Arg Val Asn Glu Asp Asn Ile His Leu
65 70 75 80
Phe Ala Arg Trp Val His Ala Glu Glu Leu Asp Met Lys Lys His Leu
85 90 95
Leu Ile Glu Leu Glu
100

<210> 33
<211> 106
<212> PRT
<213> Unknown

<220>
<223> NrdA-2的N端结构域

<400> 33
Cys Leu Thr Gly Asp Ala Lys Ile Asp Val Leu Ile Asp Asn Ile Pro
1 5 10 15
Ile Ser Gln Ile Ser Leu Glu Glu Val Val Asn Leu Phe Asn Glu Gly
20 25 30
Lys Glu Ile Tyr Val Leu Ser Tyr Asn Ile Asp Thr Lys Glu Val Glu
35 40 45
Tyr Lys Glu Ile Ser Asp Ala Gly Leu Ile Ser Glu Ser Ala Glu Val
50 55 60
Leu Glu Ile Ile Asp Glu Glu Thr Gly Gln Lys Ile Val Cys Thr Pro
65 70 75 80
Asp His Lys Val Tyr Thr Leu Asn Arg Gly Tyr Val Ser Ala Lys Asp
85 90 95
Leu Lys Glu Asp Asp Glu Leu Val Phe Ser
100 105

<210> 34
<211> 102
<212> PRT
<213> 点状念珠藻

<400> 34
Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu
1 5 10 15
[0015] Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
20 25 30
Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
35 40 45
Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60
Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
65 70 75 80
Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
85 90 95
Val Asp Asn Leu Pro Asn
100

<210> 35
<211> 105
<212> PRT
<213> 集胞藻

<400> 35
Cys Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser Thr Gly Lys Arg
1 5 10 15
Val Ser Ile Lys Asp Leu Leu Asp Glu Lys Asp Phe Glu Ile Trp Ala
20 25 30
Ile Asn Glu Gln Thr Met Lys Leu Glu Ser Ala Lys Val Ser Arg Val
35 40 45
Phe Cys Thr Gly Lys Lys Leu Val Tyr Ile Leu Lys Thr Arg Leu Gly
50 55 60
Arg Thr Ile Lys Ala Thr Ala Asn His Arg Phe Leu Thr Ile Asp Gly

65 70 75 80
Trp Lys Arg Leu Asp Glu Leu Ser Leu Lys Glu His Ile Ala Leu Pro
85 90 95
Arg Lys Leu Glu Ser Ser Ser Leu Gln
100 105

<210> 36
<211> 45
<212> PRT
<213> Unknown

<220>
<223> GP41.8的C端结构域

<400> 36
Met Cys Glu Ile Phe Glu Asn Glu Ile Asp Trp Asp Glu Ile Ala Ser
1 5 10 15
Ile Glu Tyr Val Gly Val Glu Glu Thr Ile Asp Ile Asn Val Thr Asn
20 25 30
Asp Arg Leu Phe Phe Ala Asn Gly Ile Leu Thr His Asn
35 40 45

<210> 37
<211> 40
<212> PRT
<213> Unknown

[0016]

<220>
<223> NrdJ1的C端结构域

<400> 37
Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val
1 5 10 15
Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe
20 25 30
Ala Asn Asp Ile Leu Val His Asn
35 40

<210> 38
<211> 40
<212> PRT
<213> Unknown

<220>
<223> IMPDH1的C端结构域

<400> 38
Met Lys Phe Lys Leu Lys Glu Ile Thr Ser Ile Glu Thr Lys His Tyr
1 5 10 15
Lys Gly Lys Val His Asp Leu Thr Val Asn Gln Asp His Ser Tyr Asn
20 25 30
Val Arg Gly Thr Val Val His Asn
35 40

<210> 39
<211> 34
<212> PRT
<213> Unknown

<220>
<223> NrdA-2的C端结构域

<400> 39
Met Gly Leu Lys Ile Ile Lys Arg Glu Ser Lys Glu Pro Val Phe Asp
1 5 10 15
Ile Thr Val Lys Asp Asn Ser Asn Phe Phe Ala Asn Asn Ile Leu Val
20 25 30
His Asn

<210> 40
<211> 36
<212> PRT
<213> 点状念珠藻

<400> 40
Met Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr
1 5 10 15
Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe
20 25 30
Ile Ala Ser Asn
35

[0017]

<210> 41
<211> 48
<212> PRT
<213> 集胞藻

<400> 41
Ser Pro Glu Ile Glu Lys Leu Ser Gln Ser Asp Ile Tyr Trp Asp Ser
1 5 10 15
Ile Val Ser Ile Thr Glu Thr Gly Val Glu Glu Val Phe Asp Leu Thr
20 25 30
Val Pro Gly Pro His Asn Phe Val Ala Asn Asp Ile Ile Val His Asn
35 40 45

<210> 42
<211> 27
<212> PRT
<213> Unknown

<220>
<223> GP41-2 N-内含肽序列

<400> 42
Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Gln Gln Gly Leu Lys Asp
1 5 10 15
Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu
20 25

<210> 43
<211> 46
<212> PRT
<213> Unknown

<220>
<223> GP41-3 N-内含肽序列

<400> 43
Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
1 5 10 15
Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
20 25 30
Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser
35 40 45

<210> 44
<211> 88
<212> PRT
<213> Unknown

<220>
<223> GP41-4 N-内含肽序列

<400> 44
Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
1 5 10 15
[0018] Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
20 25 30
Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys
35 40 45
Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu
50 55 60
Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
65 70 75 80
Gly Met Cys Leu Tyr Val Lys Glu
85

<210> 45
<211> 88
<212> PRT
<213> Unknown

<220>
<223> GP41-5 N-内含肽序列

<400> 45
Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
1 5 10 15
Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
20 25 30
Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys
35 40 45
Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu

50 55 60
Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
65 70 75 80
Gly Met Cys Leu Tyr Val Lys Glu
85

<210> 46
<211> 43
<212> PRT
<213> Unknown

<220>
<223> GP41-6 N-内含肽序列

<400> 46
Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu
1 5 10 15
Glu His Leu Phe Pro Thr Gln Asn Gly Glu Val Asn Ile Lys Gly Gly
20 25 30
Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu
35 40

<210> 47
<211> 26
<212> PRT
<213> Unknown

[0019]

<220>
<223> GP41-7 N-内含肽序列

<400> 47
Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu Leu Asp Glu Arg Glu
1 5 10 15
Leu Ile Asp Ile Glu Val Ser Gly Asn His
20 25

<210> 48
<211> 133
<212> PRT
<213> Unknown

<220>
<223> NrdA-1 N-内含肽序列

<400> 48
Cys Val Ala Gly Asp Thr Lys Ile Lys Ile Lys Tyr Pro Glu Ser Val
1 5 10 15
Gly Asp Gln Tyr Gly Thr Trp Tyr Trp Asn Val Leu Glu Lys Glu Ile
20 25 30
Gln Ile Glu Asp Leu Glu Asp Tyr Ile Ile Met Arg Glu Cys Glu Ile
35 40 45
Tyr Asp Ser Asn Ala Pro Gln Ile Glu Val Leu Ser Tyr Asn Ile Glu
50 55 60
Thr Gly Glu Gln Glu Trp Lys Pro Ile Thr Ala Phe Ala Gln Thr Ser
65 70 75 80

```
Pro Lys Ala Lys Val Met Lys Ile Thr Asp Glu Glu Ser Gly Lys Ser
                85                  90                  95
Ile Val Val Thr Pro Glu His Gln Val Phe Thr Lys Asn Arg Gly Tyr
            100                 105                 110
Val Met Ala Lys Asp Leu Ile Glu Thr Asp Glu Pro Ile Ile Val Asn
        115                 120                 125
Lys Asp Met Asn Phe
    130
```

<210> 49
<211> 105
<212> PRT
<213> Unknown

<220>
<223> NrdA-4 N-内含肽序列

<400> 49

```
Cys Leu Ala Gly Asp Thr Thr Val Thr Val Leu Glu Gly Asp Ile Val
 1               5                  10                  15
Phe Glu Met Thr Leu Glu Asn Leu Val Ser Leu Tyr Lys Asn Val Phe
            20                  25                  30
Ser Val Ser Val Leu Ser Phe Asn Pro Glu Thr Gln Lys Gln Glu Phe
        35                  40                  45
Lys Pro Val Thr Asn Ala Ala Leu Met Asn Pro Glu Ser Lys Val Leu
    50                  55                  60
Lys Ile Thr Asp Ser Asp Thr Gly Lys Ser Ile Val Cys Thr Pro Asp
65                  70                  75                  80
His Lys Val Phe Thr Lys Asn Arg Gly Tyr Val Ile Ala Ser Glu Leu
                85                  90                  95
Asn Ala Glu Asp Ile Leu Glu Ile Lys
            100                 105
```

[0020]

<210> 50
<211> 65
<212> PRT
<213> Unknown

<220>
<223> NrdA-5 N-内含肽序列

<400> 50

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

<210> 51
<211> 49

<212> PRT
<213> Unknown

<220>
<223> NrdA-6 N-内含肽序列

<400> 51
Tyr Val Cys Ser Arg Asp Asp Thr Thr Gly Phe Lys Leu Ile Cys Thr
1 5 10 15
Pro Asp His Met Ile Tyr Thr Lys Asn Arg Gly Tyr Ile Met Ala Lys
20 25 30
Tyr Leu Lys Glu Asp Asp Glu Leu Leu Ile Asn Glu Ile His Leu Pro
35 40 45
Thr

<210> 52
<211> 105
<212> PRT
<213> Unknown

<220>
<223> NrdJ-1 N-内含肽序列

<400> 52
Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
1 5 10 15
[0021] Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
20 25 30
Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
35 40 45
Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
50 55 60
Ile Asp Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
65 70 75 80
Pro Val Tyr Thr Lys Asn Arg Gly Tyr Val Arg Ala Asp Glu Leu Thr
85 90 95
Asp Asp Asp Glu Leu Val Val Ala Ile
100 105

<210> 53
<211> 105
<212> PRT
<213> Unknown

<220>
<223> NrdJ2 N-内含肽序列

<400> 53
Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
1 5 10 15
Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
20 25 30
Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
35 40 45
Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu

```
    50                  55                  60
Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
65                  70                  75                  80
Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr
                85                  90                  95
Asp Asp Asp Glu Leu Val Val Ala Ile
            100                 105
```

<210> 54
<211> 47
<212> PRT
<213> Unknown

<220>
<223> GP41-9 C-内含肽序列

<400> 54
Met Ile Met Lys Asn Arg Glu Arg Phe Ile Thr Glu Lys Ile Leu Asn
 1   5   10   15
Ile Glu Glu Ile Asp Asp Asp Leu Thr Val Asp Ile Gly Met Asp Asn
            20   25   30
Glu Asp His Tyr Phe Val Ala Asn Asp Ile Leu Thr His Asn Thr
        35   40   45

[0022]

<210> 55
<211> 42
<212> PRT
<213> Unknown

<220>
<223> IMPDH-2 C-内含肽序列

<400> 55
Met Lys Phe Thr Leu Glu Pro Ile Thr Lys Ile Asp Ser Tyr Glu Val
 1   5   10   15
Thr Ala Glu Pro Val Tyr Asp Ile Glu Val Glu Asn Asp His Ser Phe
            20   25   30
Cys Val Asn Gly Phe Val Val His Asn Ser
        35   40

<210> 56
<211> 41
<212> PRT
<213> Unknown

<220>
<223> IMPDH-3 C-内含肽序列

<400> 56
Met Lys Phe Lys Leu Val Glu Ile Thr Ser Lys Glu Thr Phe Asn Tyr
 1   5   10   15
Ser Gly Gln Val His Asp Leu Thr Val Glu Asp Asp His Ser Tyr Ser
            20   25   30
Ile Asn Asn Ile Val Val His Asn Ser
        35   40

<210> 57
<211> 34
<212> PRT
<213> Unknown

<220>
<223> NrdA-3 C-内含肽序列

<400> 57
Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
1 5 10 15
Thr Val Glu Glu Thr His Asn Phe Phe Ala Asn Asp Ile Leu Ile His
20 25 30
Asn Cys

<210> 58
<211> 28
<212> PRT
<213> Unknown

<220>
<223> NrdA-5 C-内含肽序列

<400> 58
Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
1 5 10 15
[0023] Thr Val Glu Gly Thr His Asn Leu Ala Tyr Ser Leu
20 25

<210> 59
<211> 33
<212> PRT
<213> Unknown

<220>
<223> NrdA-6 C-内含肽序列

<400> 59
Met Gly Ile Lys Ile Arg Lys Leu Glu Gln Asn Arg Val Tyr Asp Ile
1 5 10 15
Lys Val Glu Lys Ile Ile Ile Phe Cys Asn Asn Ile Leu Val His Asn
20 25 30
Cys

<210> 60
<211> 41
<212> PRT
<213> Unknown

<220>
<223> NrdJ-1 C-内含肽序列

<400> 60
Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val
1 5 10 15
Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe
20 25 30
Ala Asn Asp Ile Leu Val His Asn Ser
35 40

<210> 61
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 大肠杆菌的环形区域

<400> 61
Gly Cys Lys Leu
1

[0024]

<210> 62
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 大肠杆菌的环形区域

<400> 62
Gly Cys Tyr Gln
1

<210> 63
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 大肠杆菌的环形区域

<400> 63
Gly Cys Gly Tyr Gln
1 5

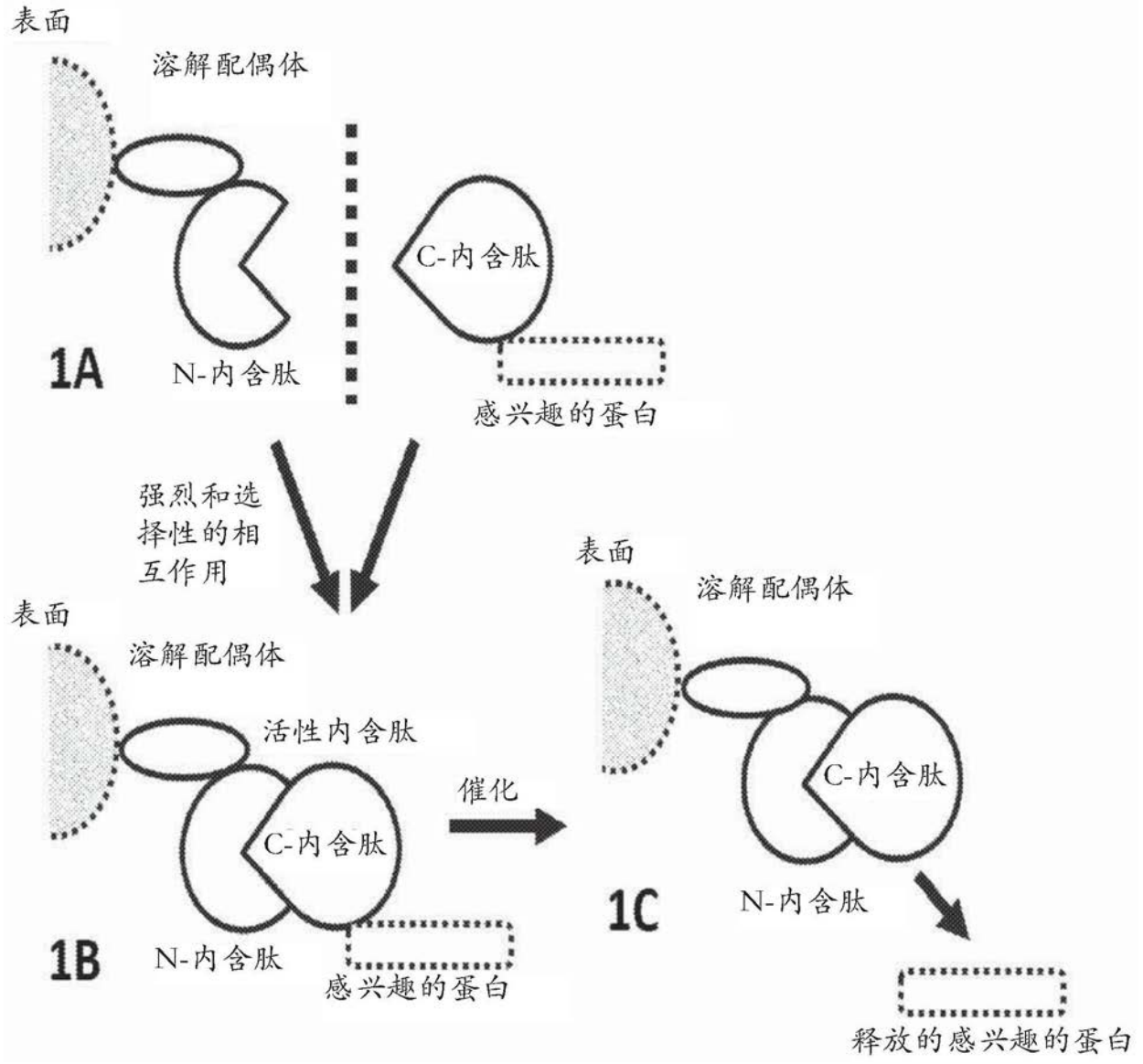

表面
溶解配偶体
C-内含肽
1A
N-内含肽
感兴趣的蛋白
强烈和选择性的相互作用
表面
溶解配偶体
活性内含肽
C-内含肽
1B
N-内含肽
感兴趣的蛋白
催化
表面
溶解配偶体
C-内含肽
1C
N-内含肽
释放的感兴趣的蛋白

图1

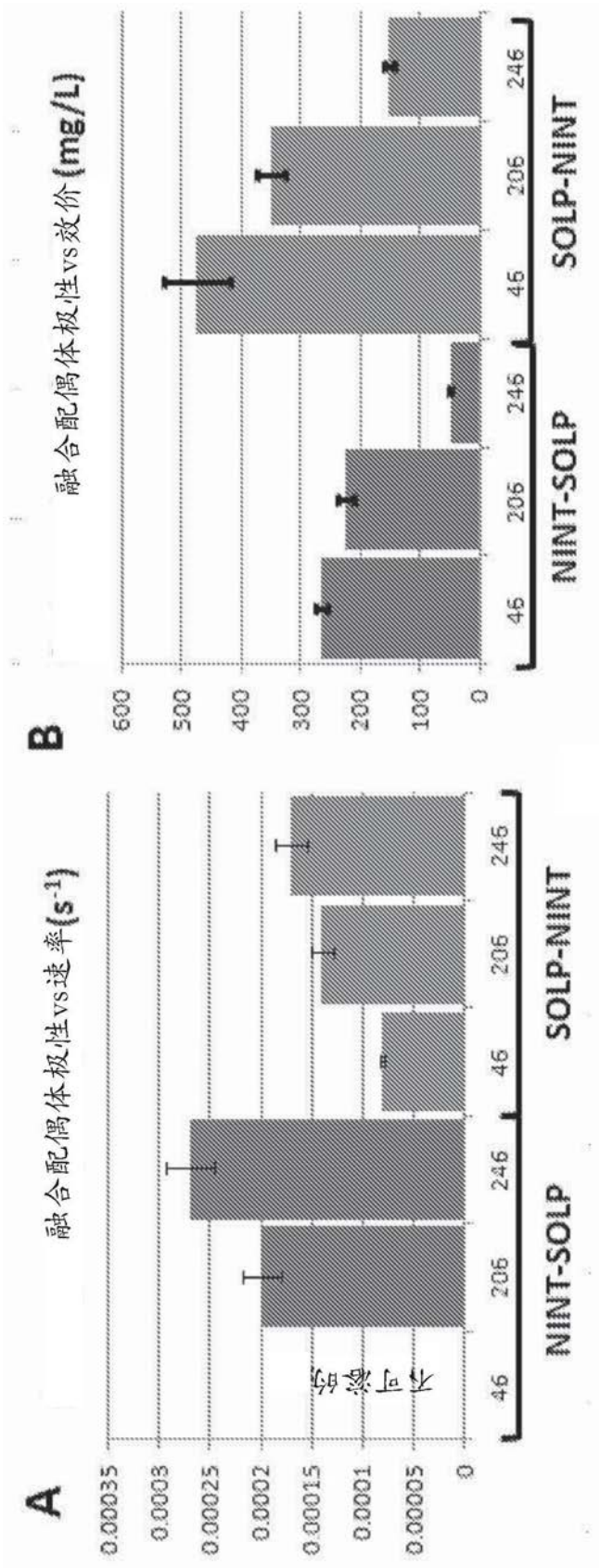

A
融合配偶体极性vs速率(s^{-1})
0.00035
0.0003
0.00025
0.0002
0.00015
0.0001
0.00005
0
46
206
246
NINT-SOLP
SOLP-NINT
B
融合配偶体极性vs效价(mg/L)
600
500
400
300
200
100
0
46
206
246
NINT-SOLP
SOLP-NINT

图2

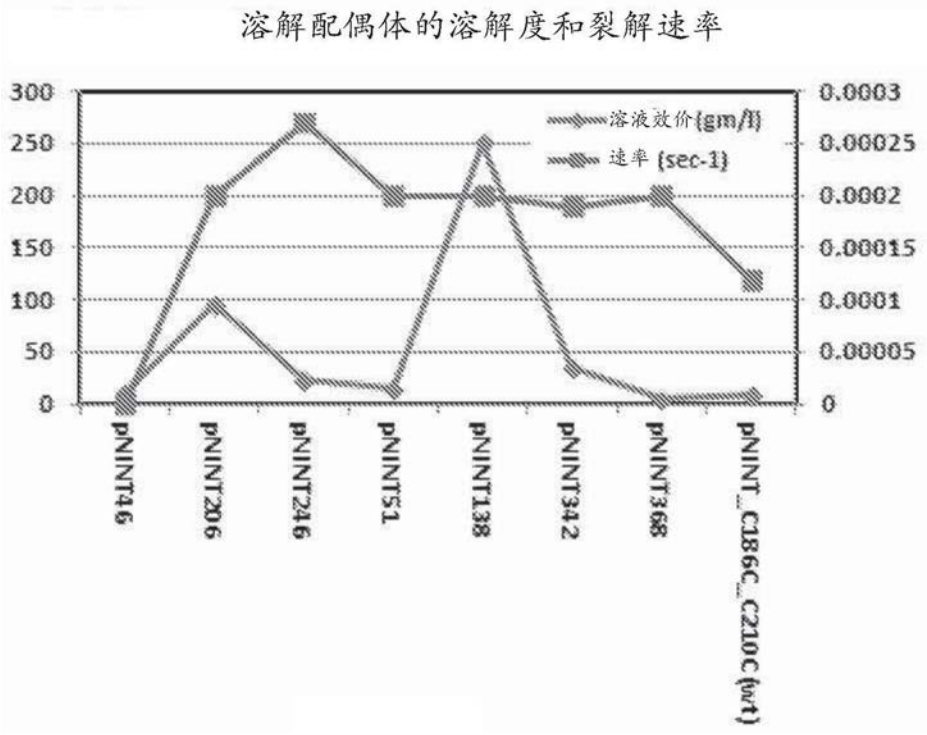

溶解配偶体的溶解度和裂解速率
溶液效价(gm/l)
速率 (sec-1)
300
250
200
150
100
50
0
0.0003
0.00025
0.0002
0.00015
0.0001
0.00005
0
pNINT46
pNINT206
pNINT246
pNINT51
pNINT138
pNINT342
pNINT368
pNINT_C186C_C210C (wt)

图3

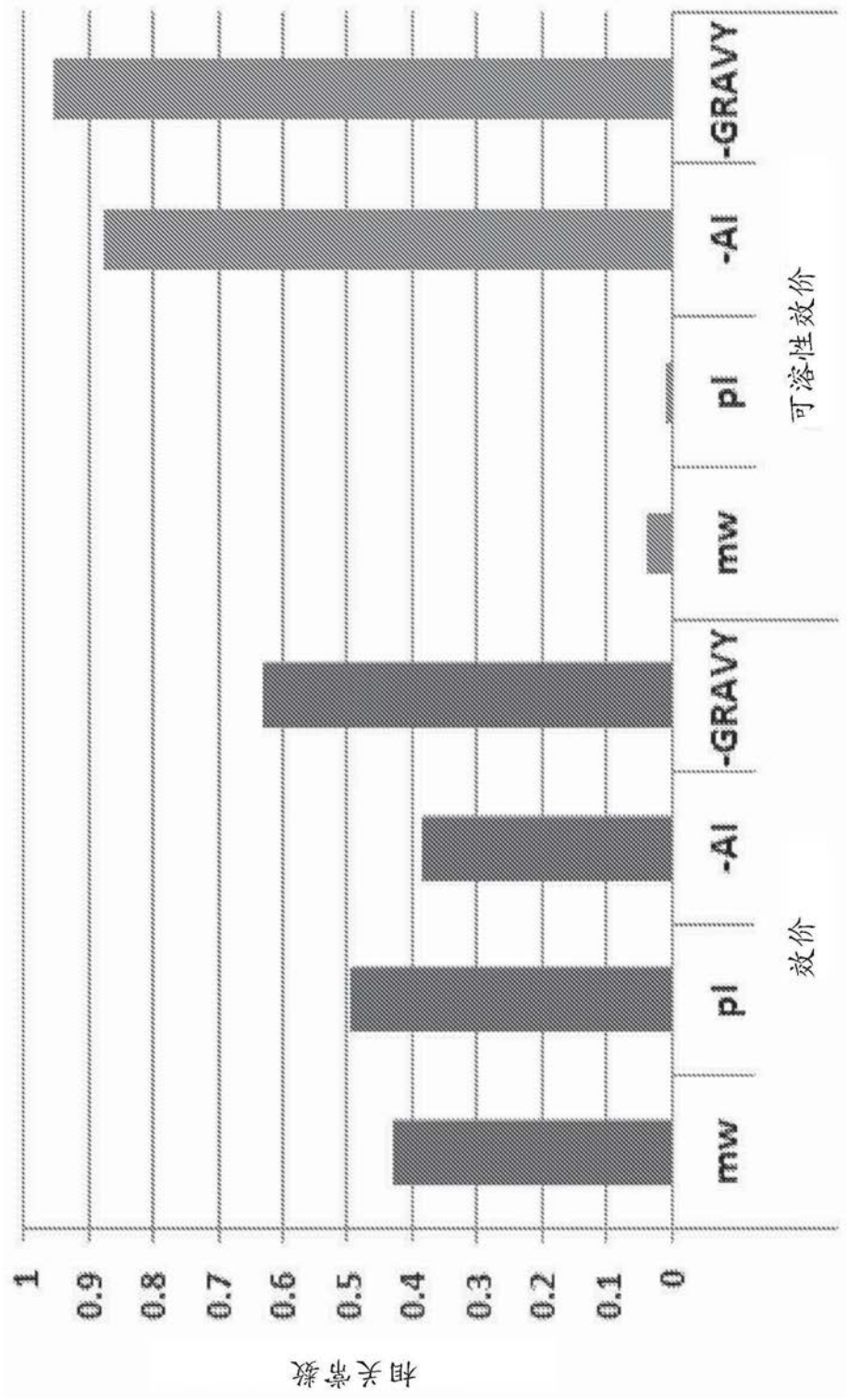

相关常数
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
mw
pI
-AI
-GRAVY
效价
mw
pI
-AI
-GRAVY
可溶性效价

图4

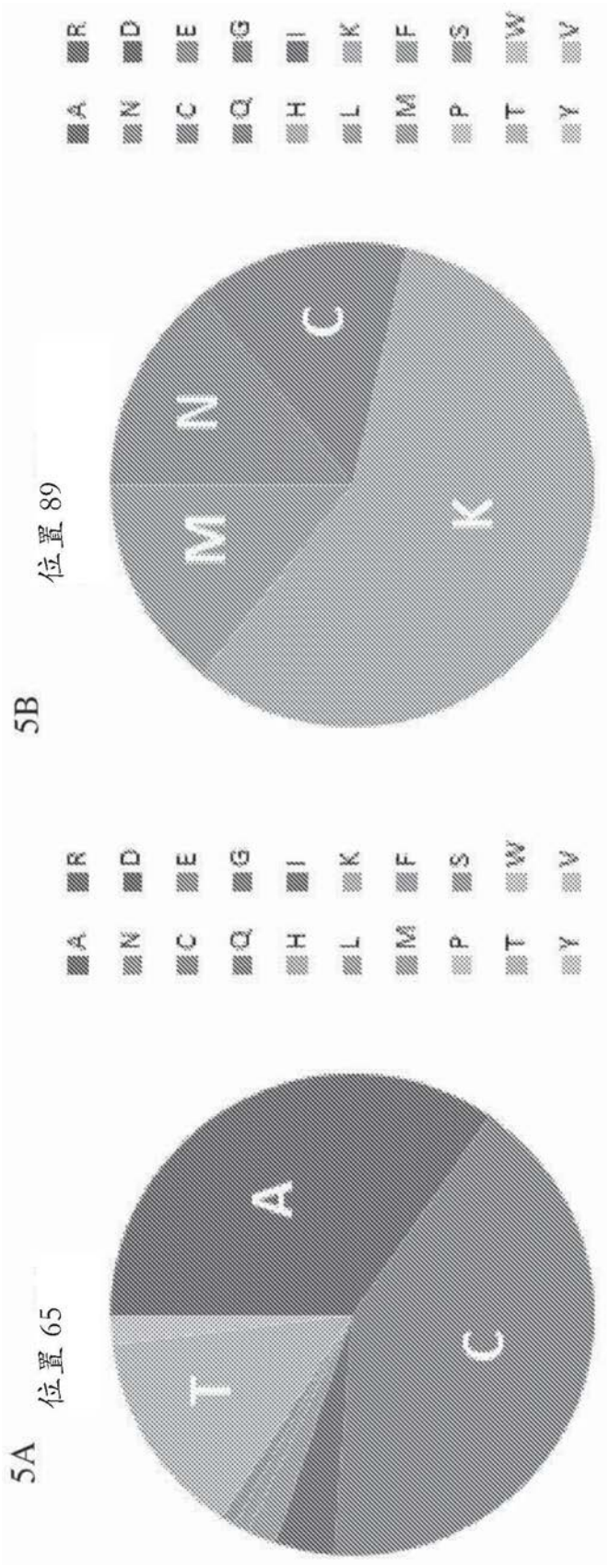


图5A-5B

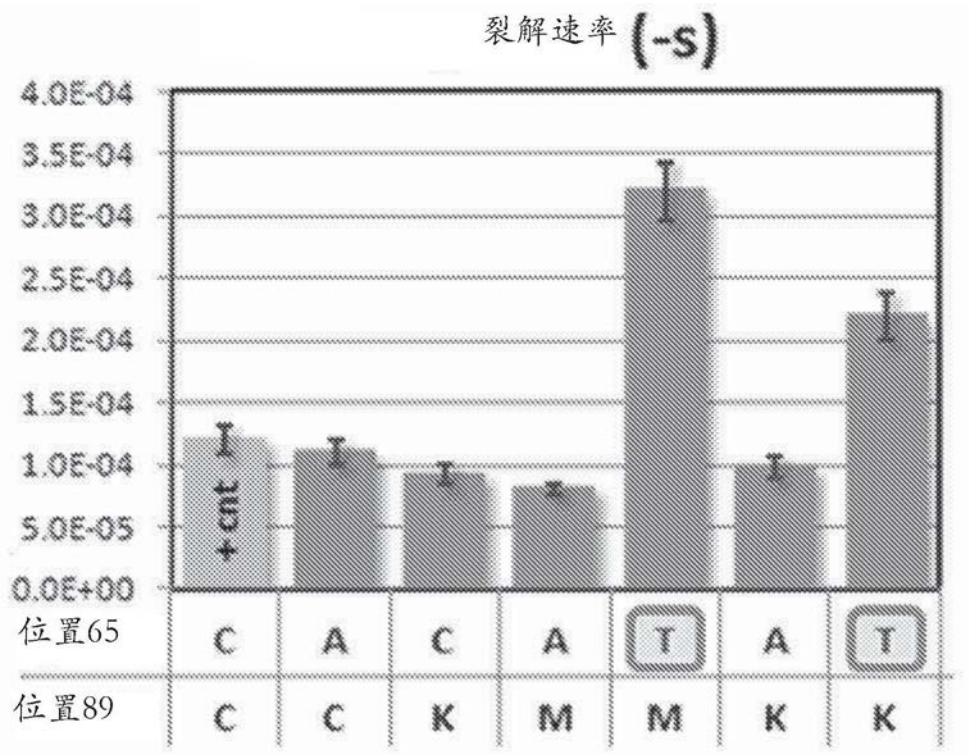

裂解速率 (-S)
4.0E-04
3.5E-04
3.0E-04
2.5E-04
2.0E-04
1.5E-04
1.0E-04
5.0E-05
0.0E+00
+tus
位置65
C A C A T A T
位置89
C C K M M K K

图6



图7