



(12)发明专利申请

(10)申请公布号 CN 108603214 A

(43)申请公布日 2018.09.28

(21)申请号 201780009483.3

(74)专利代理机构 北京市中咨律师事务所
11247

(22)申请日 2017.02.03

代理人 胡志君 黄革生

(30)优先权数据

16154372.3 2016.02.05 EP

(51)Int.Cl.

G12P 21/02(2006.01)

(85)PCT国际申请进入国家阶段日

C07C 215/90(2006.01)

2018.08.02

C07C 323/12(2006.01)

(86)PCT国际申请的申请数据

C07C 323/66(2006.01)

PCT/EP2017/052318 2017.02.03

(87)PCT国际申请的公布数据

W02017/134190 EN 2017.08.10

(71)申请人 豪夫迈·罗氏有限公司

地址 瑞士巴塞尔

(72)发明人 M·伯尼茨-杜拉特 M·沙特

权利要求书2页 说明书39页
序列表53页 附图1页

(54)发明名称

基于硫羟的深共晶溶剂

(57)摘要

本文中报导了一种由摩尔比1:2至1:3的氯化(2-羟乙基)三甲基铵和二硫苏糖醇和0%至10%的共溶剂组成的深共晶溶剂。

1. 深共晶溶剂,由摩尔比1:2至1:3的氯化(2-羟乙基)三甲基铵和二硫苏糖醇和多于0%至10%的共溶剂组成。

2. 深共晶溶剂,由以下组成

a) i) 氯化(2-羟乙基)三甲基铵,和

ii) 4-巯基-1-丁醇或1-巯基-2-丙醇,

i) 对ii)的摩尔比是约1:2,

和

b) 多于0%至约10%共溶剂。

3. 深共晶溶剂,由摩尔比约1:1的氯化(2-羟乙基)三甲基铵和2-巯基乙磺酸钠和多于0%至约10%的共溶剂组成。

4. 根据权利要求1至3中任一项所述的深共晶溶剂,其中共溶剂是含水共溶剂。

5. 根据权利要求4所述的深共晶溶剂,包含多于0%直至约5%(v/v)的含水共溶剂。

6. 用于酶促产生多肽的方法,包括以下步骤:

-在根据权利要求1至5中任一项所述的深共晶溶剂中温育

i) 包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)的第一多肽,

ii) i) 在其N端包含甘氨酸、丙氨酸或半胱氨酸化合物,或ii) 在其N端包含寡甘氨酸、或寡丙氨酸、或半胱氨酸氨基酸残基后接一个至三个甘氨酸或丙氨酸氨基酸残基,或iii) 在其5个N端氨基酸残基范围内包含赖氨酸氨基酸残基的第二多肽,和

iii) 具有分选酶A活性的第三多肽,

和由此产生多肽。

7. 根据权利要求6所述的方法,其中第二多肽在其N端具有氨基酸序列GGG、AAA、CGG、CAA、KGG或KAA (SEQ ID NO:29、162-166)。

8. 根据权利要求6或7所述的方法,其中第二多肽在其N端具有氨基酸序列GGG或AAA。

9. 根据权利要求6至8中任一项所述的方法,其中第一多肽在25个C端氨基酸残基范围内包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。

10. 根据权利要求6至9中任一项所述的方法,其中第一多肽在其C端包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。

11. 根据权利要求6至10中任一项所述的方法,其中第一多肽在其C端具有氨基酸序列LPETG (SEQ ID NO:04)或LPETA (SEQ ID NO:108)。

12. 根据权利要求6至11中任一项所述的方法,其中第一多肽和第二多肽彼此独立地选自抗体可变结构域、抗体重链Fab片段、抗体Fc区、标签和包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)的肽、接头和非分选酶基序部分。

13. 根据权利要求6至12中任一项所述的方法、其中第三多肽包含SEQ ID NO:05、SEQ ID NO:06、SEQ ID NO:27、SEQ ID NO:156、SEQ ID NO:159、SEQ ID:No:160或SEQ ID NO:161的氨基酸序列。

14. 根据权利要求6至13中任一项所述的方法,其中第三多肽具有SEQ ID NO:05或06的氨基酸序列。

基于硫羟的深共晶溶剂

[0001] 本文中报导了基于硫羟的深共晶溶剂及其用途。

[0002] 发明背景

[0003] 深共晶溶剂 (Deep eutectic solvents; DES) 是熔点低于 100°C 的无水溶液。DES 以氢键为基础。在 DES 中, 水和有机溶剂的特性组合。大部分 DES 在适于生物催化的温度为液体。DES 比水更疏水, 然而在一些情况下, 它们不导致疏水性诱导的生物催化剂失活 (Gorke, J. 等人, *Biotechnol. Bioprocess E* 15 (2010) 40-53)。另外, 可以无水产生深共晶溶剂并且因此对水解作用迟钝。离子液体具有与 DES 相似的特性, 但是更难以产生并更有害于环境。蛋白酶已经在 DES 中显示活性 (Zhao, H. 等人, *J. Mol. Catal. B-Enzym* 72 (2011) 163-167)。

[0004] 分选酶 A (SrtA) 是将蛋白质共价接合至细菌细胞壁的膜结合型酶。SrtA 底物上的特定识别基序是 LPXTG, 由此该酶在苏氨酸残基和甘氨酸残基之间切割。肽聚糖上的识别基序是五甘氨酸基序。已经显示, 在 N 端上的三甘氨酸和甚至二甘氨酸基序足以支持 SrtA 反应 (Clancy, K.W. 等人, *Peptide science* 94 (2010) 385-396)。该反应经硫酯酰基酶中间体推进, 通过攻击来自寡甘氨酸的胺亲核体、将肽聚糖共价连接至蛋白质底物并再生 SrtA 而消散。SrtA 可以用来将化学合成的肽共价缀合至重组表达的蛋白质。

[0005] 用于技术上生物缀合的适用分选酶是有限的。最广泛使用的金黄色葡萄球菌 (*Staphylococcus aureus*) 分选酶 A (St.au.SrtA) 显示适合技术应用的转化动力学, 但具有受限的底物谱, 仅识别 LPXTG 分选酶基序。缺少 N 端膜锚定基序的金黄色葡萄球菌 SrtA (Sa-SrtA) 已经用于细胞表面蛋白标记、共价性蛋白质固定和向蛋白质掺入新官能团。对于正交/双标记策略, 需要具有新底物谱的分选酶。这对标准分选酶介导的生物缀合方案同样如此, 其中产物中的 LPXTG 基序例如对其结构和/或功能具有不利作用。因此, 识别序列与 LPXTG 不同的分选酶其是有高度价值的。化脓性葡萄球菌 SrtA (St.py.SrtA) 识别 LPXTA 分选酶基序, 然而在技术规模上, 该酶的转化动力学参数令其成为不合适的分选酶。

[0006] 文献中报导了接受与 LPXTG 不同的分选酶基序的分选酶。其下有野生型, 例如来自酿脓链球菌 (*Streptococcus pyogenes*) 的分选酶 A (St.py.SrtA) 和来自艰难梭菌 (*Clostridium difficile*) 的分选酶 A (Cl.di.SrtA) 以及工程化分选酶。除了 St.py.SrtA 之外, 报导的分选酶均不识别 LPXTA 基序 (参见, 例如 van Leeuwen, H.C. 等人, *FEBS Lett.* 588 (2014) 4325-4333; Dorr, B.M. 等人, *Proc. Natl. Acad. Sci. USA* 111 (2014) 13343-13348; Bentley, M.L. 等人, *J. Biol. Chem.* 282 (2007) 6571-6581; Race, P.R. 等人, *J. Biol. Chem.* 284 (2009) 6924-33; Antos, J.M. 等人, *J. Am. Chem. Soc.* 131 (2009) 10800-10801)。

[0007] 分选酶反应在具有几个缺点的水溶液中进行。一个缺点是许多化合物在水中具有低溶解度-这对许多荧光团来说尤其如此。另外, 分选酶具有很低的 K_m , 从而使疏水性底物不可用于分选酶反应 (Chen, I. 等人, *Proc. Natl. Acad. Sci. USA* 108 (2011) 11399-11404)。但是, 与荧光团连接的抗体/结合性蛋白质可能是生物缀合过程中最重要的应用 (Shreve, P. 和 Aisen, A.M., *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine/Society of Magnetic Resonance in Medicine* 3 (1986) 336-340; Drapkin, R.L. 等人, *Am. J. Hematol.* 7 (1979) 163-172)。另一个问题是反应

期间形成的酶中间体。它可以遭水解,这导致产物损耗。这使得无水且有机溶剂成为水的有意义备选。但是,分选酶在有机溶剂中不稳定,例如超过20%的二甲基亚砷(DMSO)削弱分选酶活性(Pritz,S.(2008)Enzymatische Ligation von Peptiden, **Peptidnucleinsäuren** und Proteinen)。

[0008] 在WO 2010/087994中报导了用于连接的方法及其用途。针对IgG样双特异性抗体的重组方法由Marvin,J.S.等人(Acta Pharmacol.Sinica 26(2005)649-658)报导。在WO 2013/003555中,报导了使用分选酶安装用于蛋白质连接的点击化学手柄(click chemistry handles)。

[0009] Strijbis,K.等人(Traffic 13(2012)780-789)报导在活细胞中使用分选酶连接蛋白质。它们已经宣称Ca²⁺依赖性金黄色葡萄球菌分选酶A在胞内无功能,但是Ca²⁺非依赖性酿脓链球菌(S.pyogenes)分选酶A在酿酒酵母(Saccharomyces cerevisiae)和哺乳动物HEK293T细胞的细胞溶胶和内质网(ER)腔中均有功能。

[0010] Levary,D.A.等人报导由分选酶A催化的蛋白质-蛋白质融合(PLOS ONE 6(2011))。Madej,M.P.等人报导了将抗表皮生长因子受体抗体工程化成单链样式并通过分选酶A介导的蛋白质连接作用标记(Biotechnol.Bioeng.109(2012)1461-1470)。Ta,H.T.等人报导了酶促单链抗体加标签作为心血管疾病中靶向分子成像和细胞归巢的通用方法(Cir.Res.109(2011)365-373)。Popp,M.等人报导使用分选酶产生和破坏肽键-蛋白质工程(Angew.Chem.Int.Ed.Eng.50(2011)5024-5032)。在WO 2010/099536中报导了对DOTA螯合物具有高亲和力的工程化蛋白质。

[0011] 已经报导了阻断分选酶逆反应的不同尝试。Yamamura,Y.等人(Chem.Commun.47(2011)4742-4744)报导通过在底物识别位点周围引入β-发夹而在连接位点附近引入β-发夹结构,增强分选酶A介导的蛋白质连接。

[0012] Biswas,T.等人(Biochem.53(2014)2515-2524)报导了通过原型分选酶A分选LPXTG肽类(不变底物残基调节生产型底物的酶动力学和构象特征标识的作用)。

[0013] Li,Y.M.等人报导了通过使用分选酶对蛋白质的不可逆性位点特异性胍解(Angew.Chem.Int.Ed.Engl.53(2014)2198-2202)。Ling和合作者显示通过分选酶引入硫酯(Ling,J.J.J.等人,J.Am.Chem.Soc.134(2012)10749-10752)。Lindberg,D.等人(J.Biotechnol.147(2010)169-171)报导深共晶溶剂(DES)是酶催化环氧化物水解的可行共溶剂。Gorke,J.T.等人报导了深共晶溶剂中水解酶催化的生物转化(Chem.Commun.(2008)1235-1237)。

[0014] Bellucci,J.J.等人报导赖氨酸作为亲核体的用途(Angew.Chem.Int.Ed.Engl.53(2014)1-6)。

[0015] 在US 8,247,198中,报导了深共晶溶剂中的酶促过程。

[0016] 在WO 2002/26701中,报导了离子液体及其作为溶剂的用途。它报导了通过一种胺盐(如氯化胆碱)与能够与胺盐的离子形成氢键的有机化合物(如尿素)反应得到的凝固点直至100℃的离子化合物,其中所述反应中的摩尔比是1:1.5至1:2.5。

[0017] 在WO 2000/56700中报导了离子液体。它报导了熔点不超过60℃的离子液体,其由季铵化合物或其两种或更多种混合物与锌、锡或铁的卤化物或其两种或更多种混合物的反应形成。

- [0018] EP2 597 099报导了一种深共晶溶剂及其制备方法。
- [0019] Kareem, M.A. 等人 (J. Chem. Eng. Data 55 (2010) 4632-4637) 报导了基于磷鎓的离子液体类似物及其物理特性。
- [0020] 在US 7,763,768中,报导了在深共晶溶剂中制备活泼过氧化氢的方法。
- [0021] 在WO 2014/131906中,报导了使用深共晶溶剂 (DES) 稳定生物样品如细胞,组织,活检样品和血液中的生物分子如RNA, DNA和蛋白质是可能的。DES混合物可以用来固定和保存生物样品如组织块、癌活检样品和全血中的细胞形态。
- [0022] Smith, E.L. 等人描述了深共晶溶剂 (DES) 及其应用 (Chem. Rev. 114 (2014) 11060-11082)。
- [0023] Schmohl, L. 等人报导了分选酶介导的用于位点特异性修饰蛋白质的连接 (Curr. Opin. Chem. Biol. 22 (2014) 122-128)。
- [0024] Ling, J.J 等人报导了通过分选酶实现的蛋白质硫酯合成 (J. Am. Chem. Soc. 134 (2012) 10749-10752)。
- [0025] 发明概述
- [0026] 已经发现可以制备基于硫羰的深共晶溶剂。这类基于硫羰的深共晶溶剂包含 (a) 至少一种氢键接纳体和 (b) 含有至少 i) 一个硫羰基和 ii) 一个羟基或带负电荷的氧原子或其盐的至少一种氢键供体。
- [0027] 已经进一步发现,转肽反应可以在如本文报导的基于硫羰的深共晶溶剂中进行。已经进一步发现,水的减少/不存在并未有损于反应,反而阻抑水解副反应。
- [0028] 如本文中报告的一个方面是一种混合物/深共晶溶剂,其包含
- [0029] a) 至少一种氢键接纳体,和
- [0030] b) 含有 i) 至少一个硫羰基和 ii) 至少一个羟基或带负电荷的氧原子或其盐的至少一种脂族氢键供体,
- [0031] 其中至少一种氢键接纳体和至少一种脂族氢键供体的摩尔比是2:1至1:10。
- [0032] 在一个实施方案中,至少一种氢键接纳体和至少一种脂族氢键供体的摩尔比是1:0.5至1:4。在一个实施方案中,至少一种氢键接纳体和至少一种脂族氢键供体的摩尔比是1:1至1:3。
- [0033] 在一个实施方案中,混合物包含一种氢键接纳体和一种脂族氢键供体。
- [0034] 在一个实施方案中,脂族氢键供体包含1、2、3、4或5个硫羰基。在一个实施方案中,脂族氢键供体包含1或2个硫羰基。
- [0035] 在一个实施方案中,脂族氢键供体是包含 i) 至少一个硫羰基和 ii) 至少一个羟基或带负电荷的氧原子或其盐的链烷。在一个实施方案中,链烷是包含2至6个碳原子的直链或分枝链烷。在一个优选的实施方案中,链烷是包含2至4个碳原子的直链或分枝链烷。
- [0036] 在一个实施方案中,脂族氢键供体包含1至4个硫羰基。在一个优选的实施方案中,脂族氢键供体包含1或2个硫羰基。
- [0037] 在一个实施方案中,脂族氢键供体包含1至4个羟基。在一个优选的实施方案中,脂族氢键供体包含1或2个羟基。
- [0038] 在一个实施方案中,脂族氢键供体包含亚磺基或磺基或其盐。
- [0039] 在一个优选的实施方案中,脂族氢键供体是包含2至4个碳原子的直链或分枝链

烷,所述链烷包含i)一个或两个硫羟基和ii)一个或两个羟基或磺基或其盐。

[0040] 在一个实施方案中,脂族氢键供体额外包含1、2、3、4或5个官能团。在一个实施方案中,官能团彼此独立地选自羧基、氨基、醛基或醚基。

[0041] 在一个实施方案中,脂族氢键供体选自羟基-巯基化合物、硫代-亚磺基化合物和硫代磺基化合物及其盐。在一个实施方案中,氢键供体选自二硫苏糖醇、2-巯基乙醇、4-巯基-1-丁醇、1-巯基-2-丙醇和2-巯基乙磺酸钠。

[0042] 在一个实施方案中,至少一种氢键接纳体选自(2-羟乙基)三甲基铵盐、甲基三苯基磷鎓盐、亚乙基胺盐和三丁基胺盐。在一个实施方案中,盐是氯化物、溴化物和氢氧化物。

[0043] 在一个实施方案中,包含至少一个硫羟基的氢键供体具有200°C或更低的凝固点。在一个实施方案中,氢键供体具有85°C或更低的凝固点。

[0044] 亚乙基胺盐包括亚乙基二胺盐、二亚乙基三胺盐、三亚乙基四胺盐、四亚乙基五胺盐、五亚乙基六胺盐、2,2',2''-三氨基三乙胺、哌嗪、氨基乙基哌嗪。

[0045] 在一个优选的实施方案中,至少一种氢键接纳体和二硫苏糖醇的摩尔比是约1:2.5至1:3。在一个实施方案中,至少一种氢键接纳体是氯化胆碱。

[0046] 在一个优选的实施方案中,至少一种氢键接纳体和2-巯基乙醇的摩尔比是约1:2。在一个实施方案中,至少一种氢键接纳体是氯化胆碱。

[0047] 在一个优选的实施方案中,至少一种氢键接纳体和4-巯基-1-丁醇的摩尔比是约1:2。在一个实施方案中,至少一种氢键接纳体是氯化胆碱。

[0048] 在一个优选的实施方案中,至少一种氢键接纳体和1-巯基-2-丙醇的摩尔比是约1:2。在一个实施方案中,至少一种氢键接纳体是氯化胆碱。

[0049] 在一个优选的实施方案中,至少一种氢键接纳体和2-巯基乙磺酸钠的摩尔比是约1:1。在一个实施方案中,至少一种氢键接纳体是氯化胆碱。

[0050] 如本文中报导了的一个方面是一种用于制备深共晶溶剂的方法,所述方法包括步骤i)将至少一种氢键接纳体和包含至少一个硫羟基的至少一种氢键供体组合,和ii)加热该组合至60°C或更高的温度并冷却该加热的组合。

[0051] 在一个实施方案中,加热到60°C或更高且低于200°C的温度。在一个实施方案中,加热到100°C至150°C的温度。

[0052] 如本文中报导的一个方面是一种用于酶促产生多肽的方法,所述方法包括以下步骤:

[0053] -在如本文报告的深共晶溶剂中温育

[0054] i) (任选地在100个C端氨基酸残基范围内)包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)的第一多肽,

[0055] ii) i)在其N端包含甘氨酸、丙氨酸或半胱氨酸化合物,或ii)在其N端包含寡甘氨酸、或寡丙氨酸、或半胱氨酸氨基酸残基后接一个至三个甘氨酸或丙氨酸氨基酸残基,或iii)在其5个N端氨基酸残基范围内包含赖氨酸氨基酸残基的第二多肽,和

[0056] iii)具有分选酶A活性的第三多肽,

[0057] 和由此产生多肽。

[0058] 在一个实施方案中,具有分选酶A活性的第三多肽衍生自金黄色葡萄球菌分选酶

A,或衍生自酿脓链球菌分选酶A,或衍生自单核增生李斯特菌 (*Listeria monocytogenes*) 分选酶A。

[0059] 在一个实施方案中,第三多肽衍生自金黄色葡萄球菌分选酶A,或衍生自酿脓链球菌分选酶A,或衍生自单核增生李斯特菌分选酶A。

[0060] 在一个实施方案中,第三多肽包含SEQ ID NO:05、SEQ ID NO:06、SEQ ID NO:27、SEQ ID NO:156、SEQ ID NO:159、SEQ ID NO:160或SEQ ID NO:161的氨基酸序列。在一个优选的实施方案中,第三多肽包含SEQ ID NO:05或06的氨基酸序列。

[0061] 在一个实施方案中,第三多肽额外地在其N端或C端包含直接或通过间插接头缀合的标签。在一个实施方案中,第三多肽由SEQ ID NO:05的氨基酸序列组成并且C端标签由SEQ ID NO:32组成。在一个实施方案中,第三多肽由SEQ ID NO:05的氨基酸序列组成。

[0062] 在一个实施方案中,该方法用于两种多肽的酶促缀合。在一个实施方案中,该方法用于两种多肽的酶促转肽。

[0063] 在一个实施方案中,在30°C至40°C的温度温育。在一个实施方案中,在约37°C的温度温育。

[0064] 在一个实施方案中,第二多肽在其N端具有GGG (SEQ ID NO:29)、AAA (SEQ ID NO:162)、CGG (SEQ ID NO:163)、CAA (SEQ ID NO:164)、KGG (SEQ ID NO:165) 或KAA (SEQ ID NO:166) 的氨基酸序列。

[0065] 在一个实施方案中,第一多肽在其C端包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基) 或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。在一个实施方案中,第一多肽在其C端包含氨基酸序列LPETG (SEQ ID NO:04) 或LPETA (SEQ ID NO:108)。

[0066] 在一个实施方案中,第一多肽和第二多肽彼此独立地选自抗体可变结构域、抗体重链Fab片段、抗体Fc区、标签、抗体重链、抗体轻链和肽、接头和非分选酶基序部分,前者各自包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基) 或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。

[0067] 如本文中报告的一个方面是如本文报告的深共晶溶剂作为如溶剂在分选酶A催化的酶促转酰胺化反应中的用途。

[0068] 在如本文中报导的全部方面的一个实施方案中,深共晶溶剂包含多于0%和直至10% (v/v) 的含水共溶剂。在如本文中报导的全部方面的一个实施方案中,深共晶溶剂包含1至10% (v/v) 的含水共溶剂。在一个实施方案中,含水共溶剂是水质缓冲液。在一个实施方案中,含水共溶剂是水。

[0069] 发明的详细描述

[0070] 本发明至少部分地基于以下发现:深共晶溶剂可以用来实施转酰胺化反应,尤其分选酶催化的反应。

[0071] 本发明至少部分地基于以下发现:在深共晶溶剂中能用分选酶缀合几乎不水溶的底物。

[0072] I. 定义

[0073] 术语“衍生自”指相应的氨基酸序列包含亲本氨基酸序列的相同氨基酸序列,或含有其氨基酸序列变化或是其缩短的变体或融合变体。

[0074] 在本文中使用时,术语“包含”明确地包括术语“由……组成”。

[0075] 术语“甘氨酸、丙氨酸或半胱氨酸化合物”指包含甘氨酸或丙氨酸或半胱氨酸氨基酸残基的化合物,所述氨基酸残基具有游离 α 氨基(例如 NH_2 或 NH_3^+)和在位置1与另一个部分形成肽键的羧基,其中所述部分可以是含有任何氨基的部分(如分离的氨基酸残基、肽、多肽、蛋白质、小分子、染料等)。

[0076] 在本说明书和权利要求中,免疫球蛋白重链Fc区中残基的编号是Kabat的EU索引(Kabat,E.A.等人,Sequences of Proteins of Immunological Interest(目的免疫蛋白质的序列),第5版,Public Health Service,National Institutes of Health,Bethesda,MD(1991),NIH Publication 91-3242,所述文献通过引用的方式明确并入本文作为参考)。

[0077] 如本文所用,术语“脂族”指未取代的或取代的直链或分枝烃。这种烃可以是完全饱和的,即它仅含有碳-碳单键(也称作链烷),或它可以含有一个或多个碳-碳双键(也称作烯)或叁键(也称作炔),但任何情况下绝不包含芳族部分。例如,脂族化合物包括未取代的和取代的直链或分枝链烷、烯和炔以及其杂合体。在某些实施方案中,直链或分枝的碳原子链在其主链中具有约6个或更少的碳原子。示例性脂族化合物基团包括乙烷、正丙烷、异丙烷、正丁烷、仲丁烷、异丁烷和叔丁烷。术语“包含i)至少一个硫羟基和ii)至少一个羟基或带负电荷的氧原子或其盐的脂族氢键供体”指用至少一个硫羟基(HS^-)和至少一个羟基(HO^-)取代的脂族化合物。术语“改变”指(例如至少包含Fc区的FcRn结合部分的抗体或融合多肽的)亲本氨基酸序列中突变、添加或缺失一个或多个氨基酸残基以获得变异抗体或融合多肽。

[0078] 术语“氨基酸突变”指在亲本氨基酸序列的氨基酸序列中的修饰。示例性修饰包括氨基酸置换、插入和/或缺失。在一个实施方案中,氨基酸突变是置换。术语“在该位置的氨基酸突变”指置换或缺失指定的残基或相邻于指定的残基插入至少一个氨基酸残基。术语“相邻于指定的残基插入”指在一个至两个残基内部插入该残基。插入可以相对于指定的残基是N端或C端。

[0079] 术语“氨基酸置换”指将预先确定的亲本氨基酸序列中的至少一个氨基酸残基替换为不同的“替换”氨基酸残基。该替换残基或多个残基可以是“天然存在的氨基酸残基”(即由遗传密码编码)并且选自:丙氨酸(Ala);精氨酸(Arg);天冬酰胺(Asn);天冬氨酸(Asp);半胱氨酸(Cys);谷氨酰胺(Gln);谷氨酸(Glu);甘氨酸(Gly);组氨酸(His);异亮氨酸(Ile);亮氨酸(Leu);赖氨酸(Lys);甲硫氨酸(Met);苯丙氨酸(Phe);脯氨酸(Pro);丝氨酸(Ser);苏氨酸(Thr);色氨酸(Trp);酪氨酸(Tyr);和缬氨酸(Val)。在一个实施方案中,替换残基不是半胱氨酸。本文的氨基酸置换定义中还涵盖用一种或多种非天然存在的氨基酸残基置换。“非天然存在的氨基酸残基”指除上文所列的那些天然存在氨基酸残基之外,能够共价结合多肽链中相邻氨基酸残基的残基。非天然存在的氨基酸残基的例子包括正亮氨酸、鸟氨酸、正缬氨酸、高丝氨酸、 α -氨基异丁酸(aib)和其他氨基酸残基类似物,如在Ellman等人,Meth.Enzym.202(1991)301-336中描述的那些。为了产生这类非天然存在的氨基酸残基,可以使用Noren等人(Science 244(1989)182)和/或Ellman等人(上文)的方法。简而言之,这些方法涉及用非天然存在的氨基酸残基化学活化阻抑tRNA,随后体外转录并翻译RNA。也可以通过化学地合成肽并随后使这些肽与重组产生的多肽如抗体或抗体片段融合,将非天然存在的氨基酸掺入肽中。

[0080] 术语“氨基酸插入”指将至少一个额外的氨基酸残基掺入预先确定的亲本氨基酸序列中。尽管插入通常将由插入一个或两个氨基酸残基组成,但是本申请构思了更大的“肽插入”,例如插入约三个至约五个或甚至多达约十个氨基酸残基。插入的残基可以是如上文定义的天然存在或非天然存在的。

[0081] 术语“氨基酸缺失”指在氨基酸序列中的预定位置处移除至少一个氨基酸残基。

[0082] 在本申请范围内无论何时提到氨基酸改变,它总是人为的氨基酸改变并且不是随机氨基酸修饰。

[0083] 术语“标签”指通过具有特异性结合特性的肽键彼此连接的氨基酸残基的序列。在一个实施方案中,标签是亲和力标签或纯化标签。在一个实施方案中,标签选自Arg标签、His标签、Flag标签、3xFlag标签、Strep标签、HA标签、Nano标签、SBP标签、c-myc标签、S标签、SNUT标签、NusA、T7、硫氧还蛋白、钙调蛋白结合肽、纤维素结合结构域、壳多糖结合结构域、GST标签或MBP标签(参见,例如Amau, J. 等人, *Prot. Expr. Purif.* 48 (2006) 1-13)。

[0084] 在一个实施方案中,标签选自SEQ ID NO:07(RRRRRR)、或SEQ ID NO:08(RRRRRR)、或SEQ ID NO:09(HHHHHH)、或SEQ ID NO:10(KDHLIHNVHKEFHAAHANK)、或SEQ ID NO:11(DYKDDDDK)、或SEQ ID NO:12(DYKDHDGDYKDHDIDYKDDDDK)、或SEQ ID NO:13(AWRHPQFGG)、或SEQ ID NO:14(WSHPQFEK)、或SEQ ID NO:15(MDVEAWLGAR)、或SEQ ID NO:16(MDVEAWLGARVPLVET)、或SEQ ID NO:17(MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP)、或SEQ ID NO:18(EQKLISEEDL)、或SEQ ID NO:19(KETAAAKFERQHMS)、或SEQ ID NO:20(KRRWKNFIAVSAANRFKKISSSGAL)、或SEQ ID NO:21(纤维素结合结构域)、或SEQ ID NO:22(纤维素结合结构域)、或SEQ ID NO:23(TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEP SNVPALWQLQ)、或SEQ ID NO:24(GST标签)、或SEQ ID NO:25(MBP标签)、或SEQ ID NO:32(MRGSHHHHHHGS)。

[0085] 药剂(例如,药物制剂)的“有效量”指有效实现所需治疗性或预防性结果的量、以实现前述结果的剂量和持续实现前述结果所需要的时间段。

[0086] 术语“个体”或“受试者”指哺乳动物。哺乳动物包括但不限于驯化动物(例如,奶牛、绵羊、猫、犬和马)、灵长类(例如,人和非人灵长类如猴)、兔和啮齿类(例如,小鼠、大鼠和仓鼠)。在某些实施方案中,个体或受试者是人。

[0087] 术语“药物制剂”指一种制备物,所述制备物处于这类形式从而允许其中所含的有效成分的生物活性有效,并且不含有对于将会施用该制剂的受试者不可接受地有毒的额外组分。

[0088] “可药用载体”指药物制剂中除有效成分之外的对受试者无毒的成分。可药用载体包括但不限于缓冲剂、赋形剂、稳定剂或防腐剂。

[0089] 术语“位置”指氨基酸残基在多肽的氨基酸序列中的位置。可以将位置依次或根据建立的格式(例如用于抗体编号的Kabat的EU索引)编号。

[0090] 如本文所用,名词“治疗”(和其语法变型如动词“治疗”或分词“治疗”)指意欲改变正在接受治疗的个体的自然过程的临床介入,并且可以旨在预防或在临床病理学过程期间实施。想要的治疗效果包括但不限于防止疾病出现或复发、减轻症状、减小疾病的任何直接或间接病理学后果、防止转移、降低病情进展速率、改善或缓和疾病状态,以及缓解或预后改善。在一些实施方案中,本发明的抗体用来延缓疾病发展或用来减慢疾病的进展。

[0091] II. 深共晶溶剂

[0092] 如本文所用的术语“深共晶溶剂 (DES)”指由形成共晶体系的化合物的混合物形成的离子溶剂,所述共晶体系的熔点显著低于其每种独立组分的熔点。DES通常是包含氮盐或金属盐作为氢键接纳体(第一组分)和氢键供体(第二组分)的、在特定应用比率形成共晶点的混合物。

[0093] 深共晶溶剂是包含当合并时熔化温度降低的极性组分(液体或固体)的混合物。可以通过将氯化胆碱(ChCl)与有机氢供体例如胺、酰胺、醇或羧酸混合,产生DES(参见,例如Abbott, A.P.等人, *J. Am. Chem. Soc.* 126 (2004) 9142-9147; 同上, *Chem. Commun.* (2003) 70-71)。DES包含阴离子、阳离子和中性组分,从而DES具有离子特性和有机溶剂特性(Zhang等人, *Angew. Chem. Int. Ed.* 48 (2009) 3486-3490)。

[0094] 蛋白酶已经在DES中显示活性(Zhao, H.等人, *J. Mol. Catal. B-Enzym* 72 (2011) 163-167)。同样,脂肪酶已经在50°C含有1% (v/v) 水的DES中(Zhao, Z.等人, *J. Mol. Catal. B-Enzymatic* 85-86 (2013) 243-247)或在37°C包含0.135mL 2-丙醇、1.215mL Tris-HCl缓冲液(50mmol L^{-1} , pH 8.0)和0-1.2mol L^{-1} DES的溶液中或在不额外添加水的情况下在50°C(Huang, Z.-L.等人, *J. Chem. Technol. Biotechnol.* 89 (2014) 1975-1981)或在40°C至60°C(US 8,247,198)合并0.4-0.5mL油、2.1mL DES、0.9mL甲醇和300mg酶粉末时显示活性。

[0095] 通常,可以通过单纯地加温并搅拌各组分一段时间,任选地伴以后续干燥步骤,制备DES。例如,可以通过热处理事先混合的各组分(参见Abbott (2003) 上文; Abbott, A.P.等人, *ChemPhysChem* 7 (2006) 803)或通过冷冻干燥法(Gutierrez, M.C.等人, *Langmuir* 25 (2009) 5509; 同上, *Angew. Chem. Int. Ed.* 49 (2010) 2158),制备DES。

[0096] 例如,将氯化胆碱与包含至少一个硫羟基和至少一个羟基或其盐的脂族氢键供体(例如,二硫苏糖醇、2-巯基乙醇、4-巯基-1-丁醇、1-巯基-2-丙醇或2-巯基乙磺酸钠)在升高的温度(例如在120°C)按需要的摩尔比搅拌下混合半小时,并且此后(在DES已经形成后)冷却加热的组合。为了除去痕量水,可以此后进行五氧化二磷干燥(例如在45°C持续两周)。

[0097] 还例如,将氯化胆碱与包含至少一个硫羟基和至少一个氨基或其盐的脂族氢键供体(二硫苏糖醇、2-巯基乙醇、4-巯基-1-丁醇、1-巯基-2-丙醇或2-巯基乙磺酸钠)按预期的比率(即分别为1:22至1:3或1:1)混合并且例如使用磁力搅拌器,在火焰上混合。在无色清亮液体已经形成后,将混合物冷却至室温。此后,可能在干燥器中经五氧化二磷干燥(例如在室温,2周)。

[0098] US 8,247,198中报导了另一个示例性方法:将氮盐(0.05mol)和氢键供体(对于氯化胆碱混合物,0.1mol;对于氯化乙基铵混合物,0.075mol)添加至20mL小瓶并且在80°C加热,直至形成清亮、均一液体,一般一小时。

[0099] 根据US 8,247,198,DES内部诸组分对酶的反应性比预期令人惊讶和显著地低20倍至多于600倍。其中报导了一种包括在包含深共晶溶剂的溶液中酶促催化化学反应的方法。该反应可以例如是酯交换、氨解、水解、过水解(perhydrolysis)和/或醇脱氢酶活性。该反应可以是聚合反应。聚合反应可以由酶催化,所述酶例如是由催化酯交换、氨解、水解、过水解、醇脱氢、氧化还原或脱氢的酶组成的群组的成员。该反应可以产生例如加合产物或缩合产物。聚合反应可以例如产生聚酯或聚酰胺。催化酶催化化学反应的酶可以例如是由转

酯酶、水解酶、脂肪酶、酰胺酶和脱氢酶(第2列,第4至第19行)组成的群组的成员。

[0100] 深共晶溶剂可以包含有机组分或水质组分作为共溶剂。例如,在一个实施方案中,如本文报导的深共晶溶剂包含在10%和99% (v/v) 之间的深共晶溶剂并且剩余部分是非深共晶溶剂。

[0101] 在一个实施方案中,氢键接纳体选自有机酸、尿素、硫脲、酰胺、羟基、二醇、甘油、氯化胆碱和它们的组合。在一个实施方案中,氢键接纳体选自氯化胆碱、氯化乙基铵、溴化胆碱,甘油,氯化叔丁基铵、氯化三乙基苄基铵、氯化锌和氯化乙酰胆碱。在一个优选的实施方案中,深共晶溶剂包含氯化胆碱作为第一组分及二硫苏糖醇或2-巯基乙磺酸钠作为第二组分。

[0102] 在一个实施方案中,作为氢键接纳体,氮盐选自具有带正电荷的氮原子的化合物。在一个实施方案中,氮盐选自伯、仲、叔和季氮组分。在一个实施方案中,氮盐是铵盐或取代铵化合物。在一个实施方案中,氮盐是卤化-氮盐。在一个实施方案中,金属盐是过渡金属盐。在一个实施方案中,金属盐是卤化金属盐。在一个实施方案中,氢键供体选自羟基、酰胺、胺、醛和羧酸。过渡金属是来自化学元素周期表的具有序数21至30、39至48、71至80和103至112的元素。卤化物是氟化物(F⁻)、氯化物(Cl⁻)、溴化物(Br⁻)、碘化物(I⁻)和砷化物(At⁻)。在一个实施方案中,DES的诸组分基于它们所用的体积具有特定比率。在一个实施方案中,该比率是约1:3份至约3:1份的氮盐或金属盐对包含至少一个硫羟基的氢键供体。

[0103] 在酶促反应中使用,DES中还将存在其他组分,例如酶和其底物。归因于某些酶不以固体形式(与固相支持物缀合或作为粉末)可获得的事实,酶促反应中使用的DES将包含共溶剂(在大多数情况下水或水质缓冲液)到某个百分数。因此,DES占据酶促反应混合物体积的至少95%、至少90%、至少85%、或至少80%并且其余由水或水质缓冲液组成。在一个优选的实施方案中,酶促反应混合物包含至少90%的DES,更优选地约95%的DES。

[0104] 许多反应可以在DES中进行,所述DES例如包含酶和DES和任选地共溶剂。酶和其底物以有效浓度并且以可用来产生预期产物的比率存在。有效浓度不包括不适于以有意义的量及在合理时间内产生预期产物的浓度。在一个实施方案中,酶以至少约0.1、至少约1、至少约5、至少约10或至少约20mg酶/ml溶剂(即约0.1至约10mg/ml)的量/浓度使用/存在。类似地,在一个实施方案中,酶底物以至少约1%、至少约5%、至少约10%、至少约20%、或至少约30%底物wt/溶剂wt(即约1%至约25%)的浓度存在。

[0105] III. 使用分选酶A的酶促缀合

[0106] 可以通过使用酶分选酶、尤其分选酶A在体外获得包含体内不共价缔合的二种实体的共价缀合物(即融合多肽)。

[0107] 通常,转酰胺酶催化肽键(酰胺键)在酰基供体和亲核性酰基接纳体之间形成。为了形成肽键,酰基接纳体必需含有NH₂-CH₂-部分。革兰氏阳性菌包括以下属:放线菌属(Actinomyces)、芽孢杆菌属(Bacillus)、双歧杆菌属(Bifidobacterium)、纤维单胞菌属(Cellulomonas)、梭菌属(Clostridium)、棒状杆菌属(Corynebacterium)、微球菌属(Micrococcus)、分枝杆菌属(Mycobacterium)、诺卡菌属(Nocardia)、葡萄球菌属(Staphylococcus)、链球菌属(Streptococcus)和链霉菌属(Streptomyces)。

[0108] 基于61种来自革兰氏阳性细菌基因组的分选酶的序列比对和系统进化分析,已经将分选酶划分成4类,命名为A、B、C和D(Drams, S. 等人, Res. Microbiol. 156 (2005) 289-

297)。这些分类对应于以下亚家族,其中Comfort和Clubb也 已经将分选酶划分成所述亚家族(Comfort,D.和Clubb,R.T.,*Infect. Immun.* 72 (2004) 2710-2722):A类(亚家族1)、B类(亚家族2)、C类(亚家族3)、D类(亚家族4和5)。前述参考文献公开了许多分选酶和识别基序(也参见Pallen,M.J.等人,*Trends Microbiol.* 9 (2001) 97-101)。采用这条信息,本领域技术人员可以将分选酶基于其序列和/或其他特征(如Dramsì(上文)中描述的那些)归属至正确的分类。

[0109] 分选酶A(SrtA)是具有转酰胺酶活性的膜结合型酶。已经从革兰氏阳性细菌鉴定并分离它。体内分选酶A将蛋白质共价接合至细菌细胞壁。SrtA底物上的特定识别基序是LPXTG(SEQ ID NO:01),由此该酶在苏氨酸残基和甘氨酸残基之间切割。肽聚糖上的识别基序是五甘氨酸基序。已经显示,在N端上的三甘氨酸和甚至二甘氨酸基序足以支持SrtA反应(Clancy,K.W.等人,*Peptide science* 94 (2010) 385-396)。该反应经硫酯酰基酶中间体推进,通过攻击来自寡甘氨酸的胺亲核体、将肽聚糖共价连接至蛋白质底物并再生SrtA而消散。SrtA可以用来将化学合成的肽共价缀合至重组表达的蛋白质。

[0110] 许多革兰氏阳性细菌使用分选酶以将多种表面蛋白(包括毒力因子)共价地锚定至它们的细胞壁(肽聚糖)上。分选酶是膜结合的酶。野生型金黄色葡萄球菌分选酶A(SrtA)是具有N端跨膜区的206个氨基酸的多肽。在第一步骤中,分选酶A识别含有LPXTG(SEQ ID NO:01)氨基酸序列基序的底物蛋白并且借助活性部位Cys切割Thr和Gly之间的酰胺键。这种肽切割反应导致分选酶A-底物硫酯中间体。在第二步骤中,通过亲核性攻击含有寡甘氨酸的第二底物多肽(对应于金黄色葡萄球菌中肽聚糖的五甘氨酸单元)的氨基,消除硫酯酰基-酶中间体,导致共价连接的细胞壁蛋白和分选酶A的再生。在不存在寡甘氨酸亲核体的情况下,酰基-酶中间体可以被水分子水解。

[0111] 分选酶介导的连接/缀合已经开始用于多种蛋白质工程和生物缀合目的。这项技术能够将天然和合成性官能团引入加LPXTG标签的重组或化学合成的多肽。例子包括共价连接寡甘氨酸衍生化聚合物(例如PEG)、荧光团、维生素(例如生物素和叶酸)、脂质、糖、核酸、合成肽和蛋白质(例如GFP)(例如参见,Tsukiji,S.和Nagamune,T.,*ChemBioChem* 10 (2009) 787-798;Popp,M.W.L.和Ploegh,H.L.,*Angew. Chem. Int. Ed. Engl.* 50 (2011) 5024-5032)。

[0112]

年/引用文献	内容
1990 Fischetti 等人 Mol. Microbiol. 4 (1990) 1603-1605	LPETG (SEQ ID NO: 04)分选酶基序
1999 Mazmanian 等人 Science 285 (1999) 760-763	<p>-金黄色葡萄球菌菌株 OS2</p> <p>-保守的 Leu-Pro-X-Thr-Gly (LPXTG)基序</p> <p>-srtA 基因指定一种含 206 个氨基酸的蛋白质, 其具有潜在 NH₂ 末端信号肽/膜锚序列和在位置 184 处假定的活性部位半胱氨酸</p> <p>-srtA 同源物存在于内氏放线菌(Actinomyces naeslundii)、枯草芽孢杆菌(Bacillus subtilis)、粪肠球菌(Enterococcus faecalis)、金黄色葡萄球菌、变异链球菌(Streptococcus mutans)、肺炎链球菌(Streptococcus pneumoniae)和酿脓链球菌中</p> <p>-AF162687 公开了编码序列(可读框 1)</p> <p>MKKWTNRLMT IAGVVLILVA AYLFAKPHID NYLHDKDKDE KIEQYDKNVK EQASKDKKQQ AKPQIPKDKS KVAGYIEIPD ADIKEPVYPG PATPEQLNRG VSFAEENESL DDQNISIAGH TFIDRPNYQF TNLKAAKKS MUYFKVGNET RKYKMTSIRD VKPTDVGVLDEQKGGKDKQLT LITCDDYNEK TGVWEKRKIF VATEVK (SEQ ID NO: 167)</p>

[0113]

年/引用文献	内容
1999 Ton-That 等人 Proc. Natl. Acad. Sci. USA 96 (1999) 12424-12429	-金黄色葡萄球菌菌株 OS2 -金黄色葡萄球菌分选酶 A -残基 2-25 缺失(N 端信号序列) - MW: 22139 Da -突变 C184S 消除催化活性 -来自酿脓链球菌(Spyo)、粪肠球菌、内氏放线菌放线菌(Anei)、变异链球菌(Smut)、枯草芽孢杆菌(Bsub)和肺炎链球菌(SpnA、SpnB 和 SpnC)的分选酶同源物 -水可以消除酰基-酶中间体
2000 Dhar 等人 Biochem. 39 (2000) 3725-3733	-单核增生李斯特菌具有化学上异于葡萄球菌五甘氨酸交联桥并且较短得多的肽聚糖交联桥(m-Dpm) -单核增生李斯特菌分选酶 A 具有与金黄色葡萄球菌分选酶 A 相同的分选酶基序: LPXTG (LPTTG; SEQ ID NO: 36)
2000 Ton-That 等人 J. Biol. Chem. 275 (2000) 9876-9881	-纯化的重组金黄色葡萄球菌分选酶 A -水解具有 LPXTG 基序的肽 -三重 G-基序 -在 H ₂ N-GGG 排他性转肽存在下 -在氨基供体存在下, 分选酶介导的 LPXTG 基序剪切作用速率增加 -分选酶是一个含 206 个氨基酸的带 N 端信号序列/终止转移结构域的多肽, 锚定在葡萄球菌的胞质膜中 -残基 1-25 对应于 N 端信号序列 -反应条件: - 5 mM 氨基亲核体 - 4.71 μM SrtADN -150 mM NaCl, 5 mM CaCl ₂ , 50 mM Tris-HCl pH 7.5) -体积 520 μl -反应条件: -10 μM 荧光肽 - 5 mM 氨基亲核体 H ₂ NGGG - 15 μM SrtADN -150 mM NaCl, 5 mM CaCl ₂ , 50 mM Tris-HCl pH 7.5) -体积 520 μl - 37 °C, 16 h

表 III
SrtA_{ΔN} 的动力学分析

使用 Lineweaver-Burk 图从 Michaelis-Menten 方程的曲线拟合计算动力学常数 K_m 、 V_{max} 和 k_{cat} 。反应条件在图 4 的说明中描述。

亲核体	K_m	V_{max}	K_{cat}	K_m/K_{cat}
	μM	$\mu M/s$	$1/s$	$1/\mu M \cdot s$
H ₂ O	10.88	5.08×10^{-5}	1.06×10^{-5}	9.77×10^{-7}
NH ₂ -Gly ₃	16.48	1.08×10^{-4}	2.27×10^{-5}	1.38×10^{-6}

[0114]

年/引用文献	内容												
	<p>表 IV 不同亲核试剂对分选酶(<i>SrtA_{AN}</i>)的 LPXTG 肽切割速率的影响</p> <table border="1"> <thead> <tr> <th>亲核体</th> <th>M (s⁻¹)^a</th> </tr> </thead> <tbody> <tr> <td>H₂O</td> <td>1.84 (±0.11)</td> </tr> <tr> <td>NH₂OH</td> <td>1.91 (±0.07)</td> </tr> <tr> <td>NH₂-Gly</td> <td>1.95 (±0.05)</td> </tr> <tr> <td>NH₂-Gly₂</td> <td>2.03 (±0.13)</td> </tr> <tr> <td>NH₂-Gly₃</td> <td>2.91 (±0.03)</td> </tr> </tbody> </table> <p>^a 如图 4 所示的动力学曲线的斜率。将底物肽 <i>d</i>-QALPETGEE-e 与 <i>SrtA_{AN}</i> 和多种亲核试剂一起温育。测量苏氨酸和甘氨酸之间的底物切割为荧光的增加。除水外,所有亲核试剂均以 5 mM 的浓度加入。从三次独立实验计算平均值并报告标准偏差(括号)。</p>	亲核体	M (s ⁻¹) ^a	H ₂ O	1.84 (±0.11)	NH ₂ OH	1.91 (±0.07)	NH ₂ -Gly	1.95 (±0.05)	NH ₂ -Gly ₂	2.03 (±0.13)	NH ₂ -Gly ₃	2.91 (±0.03)
亲核体	M (s ⁻¹) ^a												
H ₂ O	1.84 (±0.11)												
NH ₂ OH	1.91 (±0.07)												
NH ₂ -Gly	1.95 (±0.05)												
NH ₂ -Gly ₂	2.03 (±0.13)												
NH ₂ -Gly ₃	2.91 (±0.03)												
2001 Bolken 等人 Infect. Immun. 69 (2001) 75-80	<ul style="list-style-type: none"> -来自戈登链球菌(<i>Streptococcus gordonii</i>)的分选酶 A -252 个氨基酸残基,带 N 端信号序列 -位置 210 处的半胱氨酸 -与金黄色葡萄球菌相比,戈登链球菌蛋白的羧基端处延长 12 个氨基酸 												
2001 Glaser 等人 Science 294 (2001) 849-852	<ul style="list-style-type: none"> -单核增生李斯特菌基因组含有 41 种含有 LPXTG 的蛋白质 -Lm 基因组比任何其他革兰氏阳性细菌含有更多的 LPXTG 蛋白(酿脓链球菌: 13 种;金黄色葡萄球菌: 18 种) 												
2001 Ilangoan 等人 Proc. Natl. Acad. Sci. USA 98 (2001) 6056-6061	<ul style="list-style-type: none"> -金黄色葡萄球菌分选酶 A 的 NMR 结构 -包含二个短螺旋和几个环的独特 b-桶结构 -半胱氨酸-184 的活性部位硫氢基由组氨酸-120 保持恒定供电离,假定这促使所得到的硫醇盐攻击 LPXTG 肽 -H120 和 C184 保守 -活性部位附近的钙结合作用刺激催化 -约 2 mM 钙离子进行刺激 -镁离子和锰离子可以替代钙离子 -残基 1-29 缺失的分选酶 -残基 1-59 缺失的分选酶 -观测的平均质量 16,595.12 Da -反应条件: -2 mM H₂N-GGG -5 μM 分选酶 N 端缺失变体 -150 mM NaCl, 5 mM CaCl₂, 50 mM Tris*HCl pH 7.5 -反应体积 520 μL -金黄色葡萄球菌 <i>SrtA</i> 残基 26-59 显示无氨基酸保守;核心 <i>SrtA</i> 残基 60-206 存在于检查的全部分选酶同源物中 												

[0115]

年/引用文献	内容
<p>2001 Mazmanian 等人 Mol. Microbiol. 40 (2001) 1049-1057</p>	<p>- 金黄色葡萄球菌分选基序: LPETG、LPDTG、LPKTG、LPNTG、PLAAG、LPKAG、LPQTG (SEQ ID NO: 04、SEQ ID NO: 37-42) - 内氏放线菌放线菌, 炭疽芽孢杆菌(Bacillus anthracis)、枯草芽孢杆菌、丙酮丁醇梭菌(Clostridium acetabutylicum)、白喉棒状杆菌(Corynebacterium diphtheria)、粪肠球菌、单核增生李斯特菌、变异链球菌、肺炎链球菌、酿脓链球菌</p>
<p>2001 Pallen 等人 TRENDS Microbiol. 9 (2001) 97-101</p>	<p>- LPXTG 样基序, 后面随即是一个跨膜疏水结构域和带电荷的羧基端尾</p> <pre> SC7A8.19 VASRYVDA-----IQPAVYFRLGALIKSSAIEIDRNDQGV- SCH19.23 VVQRYVDA-----QIPAVYFRLGALIKSSAIEVDRDQRT- SC5C11.07 IAGRYVDT-----TRAVYFARLDQLDDEKRFQVDRADGR- SCX19.15c WYQRYVDT-----TRPAVYFRLSTLEPQQTIRVARDQEV- BB3596_bactd LAGRRDT-----VFRRKGLLIGSDLVVEMPTGS- yhc8_bacu LAGRRDT-----VFRRKGLLIGSDLVVEMPTGS- BB4019_bactd LAGRRDT-----VFRRKGLLIGSDLVVEMPTGS- slp2_bacan LAGRRDT-----VFRRKGLLIGSDLVVEMPTGS- BB2127_bactd IAGRRR-----ITRRYFRRLEVEVGVVIVTVMNRR- slp_cloab LAGRRY-----TFRRYFRRLEVEVGVVIVTVMNRR- SCH49.10c LAGRR-----TRRRYFRRLEVEVGVVIVTVMNRR- SCH49.19c LARRD-----GGRRYFRRLEVEVGVVIVTVMNRR- fap1_actna ITARRGL-----ALATWYFRRLEVEVGVVIVTVMNRR- slp4_cordi ITARRGL-----ALATWYFRRLEVEVGVVIVTVMNRR- slp2_strps ITARRGL-----PTARRYFRRLEVEVGVVIVTVMNRR- slp4_strps ITARRGL-----PTARRYFRRLEVEVGVVIVTVMNRR- slp2_enfae IARRGL-----PARRYFRRLEVEVGVVIVTVMNRR- slp2_strq IARRGL-----PARRYFRRLEVEVGVVIVTVMNRR- slp2_strpy IARRGL-----PARRYFRRLEVEVGVVIVTVMNRR- slp1_cordi ITARRGL-----ANATLFRRLLEVEVGVVIVTVMNRR- slp2_cordi ITARRGL-----ANATLFRRLLEVEVGVVIVTVMNRR- slp3_cordi IARRGL-----GRRYFRRLEVEVGVVIVTVMNRR- slp5_cordi IARRGL-----GRRYFRRLEVEVGVVIVTVMNRR- BB2015_bactd IARRGYRGR-----RFRRLFRRLLEVEVGVVIVTVMNRR- BB0162_bactd IARRGYRGR-----RFRRLFRRLLEVEVGVVIVTVMNRR- slp_strmu LARRYFRRY-----SGRRYFRRLEVEVGVVIVTVMNRR- slp1_strpy LARRYFRRY-----SGRRYFRRLEVEVGVVIVTVMNRR- slp1_strps LARRYFRRY-----SGRRYFRRLEVEVGVVIVTVMNRR- slp1_enfae LARRYFRRY-----LFRRLFRRLLEVEVGVVIVTVMNRR- BB2294_bactd VDRRYFRRY-----LFRRLFRRLLEVEVGVVIVTVMNRR- slp3_bacan FRRRKLYYDT-LFRRYVDRYSAVTTT--SFRRYETFRS-I slp3_staa YRRKLEFRD-KYRYQLQVSAKTTT--RFRRYETFRS-I slp3_strpy FRRRKLEFRD-KYRYQLQVSAKTTT--RFRRYETFRS-I slp4_strpy FRRRKLEFRD-KYRYQLQVSAKTTT--RFRRYETFRS-I slp1_strps IARRR-----FRRYFRRLEVEVGVVIVTVMNRR- slp1_bacan IARRR-----FRRYFRRLEVEVGVVIVTVMNRR- yape_bacu LARRR-----GRRLFRRLLEVEVGVVIVTVMNRR- slp3_enfae LARRR-----FRRLFRRLLEVEVGVVIVTVMNRR- sortase_staa LARRYD-----RFRYFRRLEVEVGVVIVTVMNRR- slp_shep IARRDT-----RFRYFRRLEVEVGVVIVTVMNRR- SCH22A.15c VLRRYV-----RFRYFRRLEVEVGVVIVTVMNRR- slp_cloab IRRRR-----RFRYFRRLEVEVGVVIVTVMNRR- WTH1829_meth IRRRT-----YRFRYFRRLEVEVGVVIVTVMNRR- </pre>

[0117]

年/引用文献	内容
	肺炎链球菌: LPETG、VPDTG、IPQTG、YPRTG (SEQ ID NO: 04、53-55) 白喉棒状杆菌: LPMTG、LALTG、LPKTG、LGNTG、LPLTG、LAFTG (SEQ ID NO: 43、46、38、57、51) 腐败希瓦氏菌(<i>S. putrefasciens</i>): LPQTS (SEQ ID NO: 58)
2002 Garandeau 等人 Infect. Immun. 70 (2002) 1382-1390	-计算机模拟鉴定来自单核增生李斯特菌的分选酶 -222个氨基酸残基 -分选酶活性部位的 TLXTC (SEQ ID NO: 59)共有基序 MLKKTIAAAA LAAGLLLIIFS PFIKNGIVKY MSGHETIEQY KASDIKKNNE KDATFDVESV QLPSMTSVIK GAANYDKDAV VGSIAVPSVD VNLLVFKGTN TANLLAGATT MRSDQVMGKG NYPLAGHHMR DESMLFGPIM KVKKGDKIYL TDLENLYEYT VTETKTIDET EVSVIDNTKD ARITLITCDK PTETTKRFVA VGELEKTEKL TKELENKYFP SK (SEQ ID NO: 168)
2002 Bierne 等人 Mol. Microbiol. 43 (2002) 869-881	-单核增生李斯特菌 -BLAST 分析: 一个与金黄色葡萄球菌 SrtA 28%同一并编码 222 个氨基酸的预期分子量 24.7 kDa 的蛋白质的序列 -含有推定的信号肽/跨膜区, 预期的 TLXTC 序列和含 13 个和 31 个氨基酸的二个序列段不存在于 SrtA 中
2004 Mao 等人 J. Am. Chem. Soc. 126 (2004) 2670-2671	-具有单一氨基糖苷作为亲核体的肽 -攻击比水快 50 倍 -产率不受甘氨酸残基数目影响 -30 分钟后 30%产率, 6 小时后 50%产率, 24 小时后 90%产率 -缀合含有 D-氨基酸的肽; 速率对 L-氨基酸肽减半 -缀合与三甘氨酸偶联的小分子(叶酸) -分枝肽与线形肽同样有效地偶联
2004 Kruger 等人 Biochem. 43 (2004) 1541-1551	-金黄色葡萄球菌、戈登链球菌、单核增生李斯特菌、天蓝色链霉菌 (<i>Streptomyces coelicolor</i>)、酿脓链球菌、猪链球菌(<i>Streptococcus suis</i>) 中的 SrtA -二种金黄色葡萄球菌分选酶同工型: SrtA、SrtB (MPQTN 基序) -LPXTG 基序在全部革兰氏阳性细菌之间均高度保守 -NPQTN (SEQ ID NO: 60)基序似乎仅在含有血红素铁获得 <i>isd</i> 基因座的至少三种细菌(炭疽芽孢杆菌、耐盐芽孢杆菌(<i>Bacillus halodurans</i>) 和金黄色葡萄球菌)之间保守 -金黄色葡萄球菌分选酶 A -残基 2-24 缺失 -转肽酶反应条件: -100 μ L 反应体积 -150 mM NaCl, 300 mM Tris, 5 mM CaCl ₂ (pH 7.5), 五甘氨酸 (2 mM), SrtA Δ N24 (840 nM), 和 0 至 10 mM Abz-LPETG-Dap(Dnp)-NH ₂ -37°C; 30 分钟 -转肽酶反应条件: -100 μ L 反应体积 -150 mM NaCl, 300 mM Tris, 5 mM CaCl ₂ (pH 7.5)

[0118]

年/引用文献	内容
	<ul style="list-style-type: none"> - SrtAAN24 (15 μM), Gly5 (2 mM), 肽(300 μM) - 37°C, 30 分钟 -用 1 N HCl (50 μL)猝灭 -转肽酶反应条件: - 100 μL 反应体积 -150 mM NaCl, 300 mM Tris, 5 mM CaCl₂ (pH 7.5) - SrtAAN24 (60 μM), Gly5 (2 mM), 肽(300 μM) - 37°C, 360 分钟 -分选酶基序或 SrtB 基序倒转时无反应 -初始速度基序: LPXTG (X= 除 P、C 之外的任一者; X = M 最快) -端点基序: L/M-P-X-A/L/S/T/V-G (X=除 P、C 之外的任一者; L 优于 M; T 和 A 可比, 随后 S 和 V 和 L 可比; SEQ ID NO: 61) - 表 3 分选信号: LM: IPKTG、IPALG、LAASS、LPATG、LPKAG、LPISS、LPKTS (SEQ ID NO: 62、63、64、65、41、66、67)
<p>2007 Parthasarathy 等人 Bioconjug. Chem. 18 (2007) 469-476</p>	<ul style="list-style-type: none"> -缀合至非蛋白质种类的表面(聚苯乙烯珠, PEG) -烷基胺作为亲核体 -产生环状肽 -使用 159 个氨基酸的分选酶 -转肽酶反应条件: - 50 mM Tris, 150 mM NaCl, pH 8, 0.1% Tween-20, 6 mM CaCl₂, 3 mM β-巯基乙醇 - 37°C, 3 小时 - 15 μM eGFP, 10 μM 分选酶 -添加 10 mM EDTA 以终止反应 -转肽酶反应条件: - 50 mM Tris, 150 mM NaCl, pH 8 - 37°C, 3 小时 - 12 μM eGFP-LPETG, 40 μM 分选酶, 36 μM GGG-eGFP -采用珠时, 必须使用更高浓度的分选酶以取得可比产率 -使用可切割的上游标签保护亲核体(GGG) -酶显示在第一位置强烈偏好甘氨酸;显然可以将在第二位置的丙氨酸和缬氨酸置换为甘氨酸, 不过反应不那么高效
<p>2007 Chan 等人 PlosOne 11 (2007) e1164</p>	<ul style="list-style-type: none"> -缀合至固相支持物 - 四甘氨酸珠比二甘氨酸珠反应更快, 二甘氨酸珠比单甘氨酸珠反应更快 -反应条件: - 85 μM eGFP-LPETGG-His₆ - 40 nM His₆-分选酶 - 50 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl₂, pH 7.5 -对金黄色葡萄球菌基因组扩增的分选酶 -移除 N 端膜导引序列 - SDS 凝胶上 30 kDa

[0119]

年/引用文献	内容
2008 Samantaray 等人 J. Am. Chem. Soc. 130 (2008) 2132-2133	- 肽-糖缀合 - 6-氨基己糖 - 肽抗生素连接(氨基糖苷类) - 在抗生素和肽之间缀合, 对卡那霉素类的产率为 35%至 70%, 对核糖霉素类为约 18-30%: - YALPET-糖加合物(SEQ ID NO: 68) - YALPMTGK-糖加合物 (SEQ ID NO: 69) - 金黄色葡萄球菌分选酶和肽的 LPNTG 基序(SEQ ID NO: 39)

[0120] 为了酶促缀合, 可以使用缺少跨膜区的可溶性截短分选酶A (SrtA; 金黄色葡萄球菌 SrtA 的氨基酸残基 60-206) (SEQ ID NO: 05; 也参见 Ton-That, H. 等人, Proc. Natl. Acad. Sci. USA 96 (1999) 12424-12429; Ilangovan, H. 等人, Proc. Natl. Acad. Sci. USA 98 (2001) 6056-6061)。

[0121] 分选酶A介导的反应导致含有分选酶基序(序列)的种类与携带一个或多个N端甘氨酸残基的那些种类连接。分选酶基序可以是氨基酸序列LPXTG (SEQ ID NO: 01), 但还可以与之不同(参见下文)。但是, 使用这类序列作为酰基供体的缺点是LPXT (SEQ ID NO: 35) 单元转移至亲核性酰基接纳体释放了化学计量数量的含有至少一个N端甘氨酸残基的相应片段。释放的含有甘氨酸的片段与预期的酰基接纳体竞争酶促中间体并且对抗酶促连接反应的进展。另外, 虽然是相对缓慢的过程, 但酶促中间体以及含有LPXTG的底物的水解剪切与该反应竞争。在使用分选酶介导的反应伊始, 仅使用浓度至少5mM包含分选酶基序的酰基供体才能获得有用水平的连接。

[0122] 一般分选酶基序具有氨基酸序列LPXT, 其中X可以是任何氨基酸残基, 即天然存在的氨基酸残基或非天然存在的氨基酸残基。在一些实施方案中, X选自包含D、E、A、N、Q、K和R或其组成(以单字母代码)的氨基酸残基群组。在一些实施方案中, 分选酶基序选自包含氨基酸序列LPXT、LPXA、SPXT、LAXT、LSXT、NPXT、VPXT、IPXT、LGXT和YPXR (SEQ ID NO: 35、70-78) 或其组成的群组。在一些实施方案中, 分选酶基序选自由LPST、LPKT、LPIT、LPDT、SPKT、LAET、LAAT、LAST、LPLT、LSRT、LPET、VPDT、IPQT、YPRR、LPMT、LAFT和LPQT (SEQ ID NO: 79-95) 组成的氨基酸序列群组。在其中使用分选酶A的某些实施方案中, 分选酶基序包含氨基酸序列X1PX2X3 (SEQ ID NO: 96), 其中i) X1选自氨基酸残基亮氨酸、异亮氨酸、缬氨酸和甲硫氨酸, ii) X2是任何氨基酸, 并且iii) X3选自苏氨酸、丝氨酸和丙氨酸。在具体实施方案中, 如上文所示, X1是亮氨酸并且X3是苏氨酸。在某些实施方案中, X2选自天冬氨酸、谷氨酸、丙氨酸、谷氨酰胺、赖氨酸和甲硫氨酸。

[0123] 在一些实施方案中, 分选酶基序选自包含LPKTG、LPITG、LPDTA、SPKTG、LAETG、LAATG、LAHTG、LASTG、LPLTG、LSRTG、LPETG、VPDTG、IPQTG、YPRRG、LPMTG、LAFTG和LPQTS (SEQ ID NO: 38、44、45、46、47、48、49、50、51、52、35、53、54、97、43、98、58) 或其组成的氨基酸序列群组。在本发明的一些实施方案中, 分选酶是分选酶A (SrtA)。SrtA识别具有氨基酸序列LPXTG (SEQ ID NO: 01) 的分选酶基序。常见的分选酶基序氨基酸序列例如是LPKTG、LPATG、LPETG和LPNTG (SEQ ID NO: 38、65、04和39)。在一些实施方案中, 使用LPETG (SEQ ID NO: 04)。但是, 也可以识别与这种共有分选酶基序氨基酸序列不一致的分选酶基序。例如, 在一

些实施方案中,分选酶基序在位置4包含氨基酸残基A而非氨基酸残基T,例如LPXAG或LPNAG (SEQ ID NO:99、100)。在一些实施方案中,分选酶基序在位置5包含氨基酸残基A而非氨基酸残基G,例如LPXTA或LPNTA (SEQ ID NO:101、102)。在一些实施方案中,分选酶基序在位置2包含氨基酸残基G而非氨基酸残基P,例如LGXTG或LGATG。在一些实施方案中,分选酶基序在位置1包含氨基酸残基I而非氨基酸残基L,例如,IPXTG或IPNTG或IPETG (SEQ ID NO:105、106、107)。

[0124] 在一些实施方案中,若分选酶基序是LPXTG或LPXT (SEQ ID NO:01、35),则X选自D、E、A、N、Q、K和R。在一些实施方案中,X选自由LPXTG或LPXT基序中的K、E、N、Q和A组成的氨基酸残基,其中分选酶是分选酶A。在一个实施方案中,分选酶基序是LPET或LPETG或LPETA (SEQ ID NO:89、04和108)。

[0125] 在某些实施方案中,若使用金黄色葡萄球菌的分选酶A (St.au.SrtA),则分选酶基序具有氨基酸序列LPX1TX2 (SEQ ID NO:109),其中i) X1选自由D、E、A、N、Q、K和R组成的氨基酸残基群组,并且ii) X2选自由丙氨酸和甘氨酸组成的氨基酸残基群组。在某些实施方案中,金黄色葡萄球菌SrtA的分选酶基序是LPX1TA (SEQ ID NO:110)。在其他实施方案中,金黄色葡萄球菌SrtA的分选酶基序是LPX1TG (SEQ ID NO:111)。X1具有如前文略述的含义。

[0126] 酿脓链球菌分选酶A (St.py.SrtA) 将接受基于双丙氨酸的亲核体。这种分选酶将在苏氨酸和丙氨酸残基之间高效地切割分选酶基序氨基酸序列LPXTA (SEQ ID NO:101) 并安装修饰的基于丙氨酸的亲核体。酿脓链球菌SrtA也将识别并且切割LPXTG (SEQ ID NO:01) 基序,尽管效率降低。

[0127] 金黄色葡萄球菌分选酶A (St.au.SrtA) 将不显著地切割LPXTA (SEQ ID NO:101) 基序或接受基于丙氨酸的亲核体。

[0128] 在一个实施方案中,多肽与Strep.SrtA和含有丙氨酸的亲核体接触。包含分选酶基序氨基酸序列的多肽可以由Strep.SrtA在其C端或C端附近识别,并且亲核体在其N端或N端附近包含能够充当金黄色葡萄球菌SrtA介导的反应的亲核体的一个或多个氨基酸(例如,(G)_n,其中n在1和10之间,例如,在1和5之间)。这导致反应部位处形成LPXTA (SEQ ID NO:101),一种抵抗金黄色葡萄球菌SrtA剪切的基序。这种允许例如金黄色葡萄球菌SrtA作用于N端,而不影响Strep.SrtA安置的C端修饰。

[0129] 具有分选酶转酰胺化活性的分选酶片段可以用于如本文中报导的方法中。可以通过产生分选酶片段,例如,通过重组技术或蛋白酶解消化全长分选酶并且确定肽键形成(即连接)速率,鉴定分选酶片段。该片段可以包含全长分选酶的约80%氨基酸序列、全长分选酶(如金黄色葡萄球菌分选酶A (GenBank登录号AAD48437))的约70%、约60%、约50%、约40%或约30%氨基酸序列。在一些实施方案中,该片段缺少全长分选酶氨基酸序列的对分选酶催化活性并非必需的N端部分,例如,该片段缺少延伸至膜锚定物序列末端的N端部分。在一些实施方案中,该片段包含全长分选酶氨基酸序列的C端。在一些实施方案中,该片段包含分选酶的催化核心区。在一个实施方案中,核心区是从SrtA(例如,金黄色葡萄球菌SrtA)的约位置60至约位置206,或约从Strep.SrtA的位置82至约位置249。

[0130] 也可以在如本文报导的方法中使用来自其他生物的分选酶。这类分选酶通常基本上与编码SrtA的核苷酸序列相同或相似的核苷酸序列编码。相似或基本上相同的核苷酸序列可以包含对天然序列的修饰,如置换、缺失、或插入一个或多个核苷酸。包括与天然核

核苷酸序列至少55%、60%、65%、70%、75%、80%、或85%或更多同一并且通常与天然核苷酸序列90%或95%或更多同一(每个同一性百分数可以包含1%、2%、3%或4%变异)的核苷酸序列。一个确定二种核酸是否基本上同一的检验是确定二种核酸之间共有的相同核苷酸位置的百分数。

[0131] 可以使用SrtA核苷酸序列作为“查询序列”对公共数据库进行检索,以鉴定相关的序列。例如,可以使用Altschul等人(J.Mol.Biol.215(1990)403-410)的NBLAST及XBLAST程序(版本2.0)执行此类检索。BLAST核苷酸检索可以用NBLAST程序,评分=100、字长度=12执行,以获得同源核苷酸序列。为了出于比较目的获得带空位的比对结果,可以如Altschul等人(Nuc.Acids Res.25(1997)3389-3402)中所述那样使用空位BLAST。当使用BLAST和空位BLAST程序时,可以使用相应程序(例如XBLAST和NBLAST)的默认参数(例如参见www.ncbi.nlm.nih.gov)。

[0132] 变体氨基酸序列异于天然氨基酸序列。可以基于极性、电荷、溶解度、疏水性、亲水性、螺旋形成特性和/或两亲特性方面的相似性,进行氨基酸置换,并且用合适测定法,如欧洲专利申请EP14198535中报导的那种,对所产生的变体筛选酶活性。例如,带负电荷的氨基酸包括天冬氨酸和谷氨酸;带正电荷的氨基酸包括赖氨酸和精氨酸;并且具有相似亲水性值的具有不带电荷极性首基的氨基酸包括亮氨酸、异亮氨酸、缬氨酸、甘氨酸、丙氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、苯丙氨酸和酪氨酸。在某些实施方案中,可以根据下表进行保守性置换。第二列中处于相同区块内和第三列中处于相同行里的氨基酸可以在保守性置换中彼此互相置换。某些保守性置换是将与第二列中某个区块相对应的第三列的一行的氨基酸置换为来自第二列同一个区块内部的第三列另一行的氨基酸。

[0133]	脂族氨基酸残基	非极性	G、A、P
			I、L、V
	极性, 不带电荷	C、S、T、M	
		N、Q	
	极性, 带电荷	D、E	
		K、R	
芳族		H、F、W、Y	

[0134] 在某些实施方案中,可以进行同源置换,这是置换或替换类似的氨基酸,如碱性氨基酸置换碱性氨基酸、酸性氨基酸置换酸性氨基酸,极性氨基酸置换极性氨基酸和疏水性氨基酸置换疏水性氨基酸,例如。可以将非同源置换引入到天然序列,如来自一个分类的残基引入到另一类(例如非疏水性氨基酸引入到疏水性氨基酸),或将天然存在的氨基酸用非天然氨基酸置换或非经典氨基酸替换。

[0135] 在如本文报导的方法中,将分选酶、包含分选酶基序的多肽(即酰基供体)和亲核体(即酰基接纳体)在适合引起包含分选酶基序的多肽的N端部分和亲核体之间肽键形成的条件下一起温育。如本文所用,术语“温育”或其语法等同物指引起该方法的诸组分彼此紧

邻以允许分子之间接触。温育可以例如通过添加它们到一个反应容器来进行。可以按多种方式,例如,通过振荡容器、使容器置于涡旋混合生成装置或用一个或多个移液器反复混合,将系统中的组分混合。组分可以按任何顺序添加到系统。

[0136] 分选酶反应可以在任何便利的容器(例如,管如微量离心管、烧瓶、皿)、微量滴定板(例如,96孔或384孔平板)、载玻片、硅片、滤器或具有分子在其上面固定(任选地包覆)和任选以阵列方式排布的表面的任何固相或半固相支持物(例如参见,US 6,261,776和Fodor, Nature 364 (1993) 555-556)和微流体装置(例如参见,US 6,440,722;US 6,429,025;US 6,379,974和US 6,316,781)中进行。

[0137] 反应混合物通常是无细胞的并且还不包含细菌细胞壁组分或完整细菌细胞壁。在一些实施方案中,包含分选酶基序的多肽和/或亲核体在细胞中由一个或多个重组核苷酸序列表达,所述核苷酸序列整合入细胞基因组或未整合(例如,在质粒中)。

[0138] 将反应混合物维持在可以进行分选酶反应的任何方便的温度处。在一些实施方案中,分选酶反应在约15°C和约50°C之间和包括前者的温度进行。在一些实施方案中,分选酶反应在约23°C和约37°C之间和包括前者的温度进行。在某些实施方案中,温度是室温(即约20°C至25°C)。可以通过在不同温度反复进行相同的分选酶反应并测定连接速率,优化温度。

[0139] 可以使用任何便利的体积和组分比率。

[0140] 在某些实施方案中,使用1:1000或更大的分选酶对包含分选酶基序的多肽的(摩尔)比率,或使用1:1000或更大的分选酶对亲核体的(摩尔)比率。在具体实施方案中,分选酶对包含分选酶基序的多肽或酶对亲核体的比率是约1:1,包括1:2或更大、1:3或更大、1:4或更大、1:5或更大、1:6或更大、1:7或更大、1:8或更大和1:9或更大。

[0141] 在一些实施方案中,包含分选酶基序的多肽以范围约10 μ M至约10mM的浓度存在。在一些实施方案中,包含分选酶基序的多肽以范围约100 μ M至约1mM的浓度存在。在一些实施方案中,包含分选酶基序的多肽以范围约100 μ M至约50mM的浓度存在。在一些实施方案中,包含分选酶基序的多肽以范围约200 μ M至约1mM的浓度存在。在一些实施方案中,包含分选酶基序的多肽以范围约200 μ M至约800 μ M的浓度存在。在一些实施方案中,包含分选酶基序的多肽以范围约400 μ M至约600 μ M的浓度存在。

[0142] 在某些实施方案中,亲核体相对于包含分选酶基序的多肽过量存在。在某些实施方案中,亲核体相对于包含分选酶基序的多肽以10倍过量存在。在某些实施方案中,亲核体相对于包含分选酶基序的多肽以25倍过量存在。在某些实施方案中,亲核体相对于包含分选酶基序的多肽以50倍过量存在。在某些实施方案中,亲核体相对于包含分选酶基序的多肽以100倍过量存在。在某些实施方案中,亲核体相对于包含分选酶基序的多肽以250倍过量存在。

[0143] 在某些实施方案中,亲核体以范围约1 μ M至约50mM的浓度存在。在某些实施方案中,亲核体以范围约15 μ M至约1500 μ M的浓度存在。在某些实施方案中,亲核体以范围约25 μ M至约1000 μ M的浓度存在。在某些实施方案中,亲核体以范围约40 μ M至约250 μ M的浓度存在。

[0144] 在某些实施方案中,分选酶以范围约1 μ M至约500 μ M的浓度存在。在某些实施方案中,分选酶以范围约15 μ M至约150 μ M的浓度存在。在某些实施方案中,分选酶以范围约25 μ M至约100 μ M的浓度存在。在某些实施方案中,分选酶以范围约40 μ M至约60 μ M的浓度存在。

[0145] 在某些实施方案中,该方法在包含水质环境的反应混合物中进行。通常可以使用具有适宜缓冲剂和/或盐含量的水。醇或有机溶剂可以纳入某些实施方案中。有机溶剂的量通常在连接过程中不明显酯化蛋白质或肽(例如,当添加醇或有机溶剂时,酯化的蛋白质或肽通常仅增加5%或更少)。醇和/或有机溶剂含量有时是20%或更少、15%或更少、10%或更少或5%或更少,并且在使用较大的醇或有机溶剂的实施方案中,存在30%或更少、40%或更少、50%或更少、60%或更少、70%或更少或80%或更少的醇或有机溶剂。在某些实施方案中,反应混合物仅包含醇或有机溶剂,其中如果水存在的话,其量仅为有限。

[0146] 在一些实施方案中,反应混合物包含缓冲液。本领域技术人员将熟悉可以根据如本文中报导的方法使用的多种缓冲液。在一些实施方案中,缓冲溶液包含钙离子。在某些实施方案中,缓冲溶液不含有使钙离子沉淀的物质。在一些实施方案中,缓冲溶液不包含磷酸盐离子。在一些实施方案中,缓冲溶液不含有螯合剂。

[0147] 在一些实施方案中,该方法在范围6至8.5的pH值进行。在一些实施方案中,该方法在范围6至8的pH值进行。在一些实施方案中,该方法在范围6至7.5的pH值进行。在一些实施方案中,该方法在范围6.5至8.5的pH值进行。在一些实施方案中,该方法在范围7至8.5的pH值进行。在一些实施方案中,该方法在范围7.5至8.5的pH值进行。在一些实施方案中,该方法在范围7.0至8.5的pH值进行。在一些实施方案中,该方法在范围7.3至7.8的pH值进行。

[0148] 反应混合物或产物的一种或多种组分可以固定至固相支持物。反应混合物组分和固相支持物之间的接合可以是共价或非共价的(对非共价接合而言,例如参见US 6,022,688)。固相支持物可以是系统的一个或多个表面,例如,微量滴定板的每个孔中的一个或多个表面、载玻片、硅片(silicon wafer)、BIAcore芯片的表面、任选与另一种固相支持物连接的粒子(例如,珠)的表面(例如参见,Lam,Nature 354(1991)82-84)或微流体装置中的通道。已知固相支持物的类型、用于共价和非共价接合于固相支持物的接头分子和将分子至固相支持物固定的方法(例如参见,US 6,261,776;US 5,900,481;US 6,133,436;US 6,022,688;WO 2001/18234)。可以使用任何材料,例如,塑料(例如,聚苯乙烯)、金属、玻璃、纤维素、凝胶(例如,至少部分地从有机聚合物(如PDMS)形成)等。在一些实施方案中,固相支持物是半固态和/或凝胶样的、可变形、柔软等。

[0149] 在引入分选酶基序或寡甘氨酸或寡丙氨酸后,任何多肽最终可以作为包含分选酶基序的多肽或亲核体用于如本文报导的方法中。

[0150] 综上,第一底物,也称作供体,包含分选酶识别基序。它由分选酶在识别基序中的苏氨酸残基后剪切。因而生成C端活化的羧基(酰基中间体)。第二底物,也称作接纳体或亲核体,提供(游离N端)氨基。在游离氨基和活化的羧基之间,肽键在分选酶催化的转肽反应中形成。

[0151] 因此,对于分选酶介导的酶促转肽反应,仅需要将包含分选酶识别基序的供体和包含N端游离甘氨酸、丙氨酸、半胱氨酸或等同官能团的接纳体与具有分选酶A催化活性的多肽温育。剩余的供体以及剩余的接纳体不干扰反应。

[0152] 因此,分选酶介导的转肽反应可以实际上用彼此独立作为供体或接纳体的任何蛋白质或小分子进行,只要它们包含分选酶识别序列和亲核体对子。

[0153] 这由现有技术证实。

[0154] 例如,Marraffini等人(Microbiol.Mol.Biol.Rev.70(2006)192-221)报导了分选

酶A可以用来将含有甘氨酸残基的带游离氨基的化学物质掺入重组蛋白(即,无蛋白质的限制)的LPXTG基序。展示的例子是用(GFP或Cre或p27)-LPETG-His6高效率缀合三甘氨酸-赖氨酸-叶酸、分枝的肽AT-P-022掺入多肽、和嵌合体His6-分选酶-LPETG-靶蛋白的自剪切作用(一旦已经通过添加钙和三甘氨酸活化酶,则融合物自我剪切)。

[0155] 另外,Antos等人(J.Am.Chem.Soc.131(2009)10800-10801)报导,分选酶A催化的转肽反应允许实际上用任何类型的功能性材料位点特异地使蛋白质衍生化。靶蛋白经工程化以在其C端附近含有识别位点(LPXTG),由此允许其中将苏氨酸C端的残基交换为合成性寡甘氨酸肽的酰基转移反应。据报导,在不引起该反应的情况下,分选酶识别基序的末端G残基可以被甲基酯替换。在本文件中,包含荧光标记物或蛋白质的亲核体用于缀合至霍乱毒素B亚基。

[0156] 另外,Popp等人(Proc.Natl.Acad.Sci.USA 108(2011)3169-3174)报导了分选酶用于多肽环化和PEG化的用途。该方法有通用性并适用于多种类型的蛋白质。分选酶转肽酶反应允许对多种不同蛋白质轻易地位点特异性PEG化,如使用干扰素 α 2、GCSF和促红细胞生成素所例举。在检验的全部情形中,位点特异性C端PEG化高效地推进。

[0157] 在EP 2 990 423中,报导了自我剪切性分选酶构建体。在这种构建体中,分选酶识别序列LPETG和分选酶催化结构域已经合并于同一个分子中。将包含分选酶识别序列的任何蛋白质作为蛋白质,例如,选自包含聚合物蛋白质、糖蛋白、细胞因子、生长因子、血液制品、疫苗、激素、酶、抗体及其部分或片段(分离的轻链或重链)的蛋白质。

[0158] IV.分选酶

[0159] 全长酿脓链球菌分选酶A(Uniprot Q1J6K9;催化核心加下划线;保守的组氨酸加下划线)具有以下氨基酸序列:

[0160]

MVKKQKRRKI	KSMSWARKLL	IAVLLILGLA	LLFNKPIRNT
LIARNSNKYQ	VTKVSKKQIK	KNKEAKSTFD	FQAVEPVSTE
SVLQAQMAAQ	QLPVIGGIAI	PELGINLPIF	KGLGNTTELIY
GAGTMKEEQV	MGGENNYSLA	<u>SHHIFGITGS</u>	SQMLFSPLER
AQNGMSIYLT	DKEKIY EYII	KDVFTVAPER	VDVIDDTAGL

[0161]

KEVTLV<u>VTCTD</u>	IEATERIIVK	GELKTEYDFD	KAPADV LKAF
---------------------------	-------------------	-------------------	--------------------

NHSYNQVST (SEQ ID NO: 159).

[0162] 衍生自酿脓链球菌的成熟可溶性分选酶A的氨基酸序列是

[0163]

VLQAQMAAQQ LPVIGGIAIP ELGINLPIFK GLGNTELIYG
AGTMKEEQVM GGENNYSLAS HHIFGITGSS QMLFSPLERA
QNGMSIYLTD KEKIYEYIHK DVFTVAPERV DVIDDTAGLK
EVTLVCTDI EATERIIVKG ELKTEYDFDK APADVLKAFN

HSYNQVST (SEQ ID NO: 156).

[0164] 全长金黄色葡萄球菌分选酶A (参见Mazmani等人;催化核心加下划线;保守的组氨酸加下划线) 具有以下氨基酸序列:

[0165]

MKKWTNRLMT IAGVVLILVA AYLFKPHID NYLHDKDKDE
KIEQYDKNVK EQASKDKKQQ AKPQIPKDKS KVAGYIEIPD
ADIKEPVYPG PATPEQLNRG VSFAEENESL DDQNISIAGH
TFIDRPNYQF TNLKAACKGS MVEYFKVGNET RKYKMTSIRD

VKPTDVGVLDEQKGGKDKQLT LITCDDYNEK TGVWEKRKIF

VATEVK (SEQ ID NO: 26).

[0166] 无N端28个氨基酸残基 (N (2-29) 跨膜结构域) 的金黄色葡萄球菌分选酶A具有以下氨基酸序列:

[0167]

MDNYLHDKDK DEKIEQYDKN VKEQASKDKK QQAKPQIPKD
KSKVAGYIEI PDADIKEPVY PGPATPEQLN RGVVSFAEENE
SLDDQNISIA GHTFIDRPNY QFTNLKAACK GSMVEYFKVGN
ETRYKMTSI RDVKPTDVGVLDEQKGGKDKQ LTLITCDDYN

EKTGVWEKRKIFVATEVK (SEQ ID NO: 160).

[0168] 无N端59个氨基酸残基 (跨膜结构域) 的金黄色葡萄球菌分选酶A具有以下氨基酸序列:

QAKPQIPKDK SKVAGYIEIP DADIKEPVYP GPATPEQLNR

[0169]

GVSFAEENES LDDQNISIAG HTFIDRPNYQ FTNLKAACKG
SMVEYFKVGN TRKYKMTSIR DVKPTDVGVL DEQKGGKDKQL

[0170]

TLITCDDYNE KTGWEKRKI FVATEVK (SEQ ID NO: 05).

[0171] 来自单核增生李斯特菌的全长分选酶A具有以下氨基酸序列 (催化中心加下划线;保守的组氨酸加下划线):

[0172]

MLKKTIAAAA LAAGLLLIFS PFIKNGIVKY MSGHETIEQY
KASDIKKNNE KDATFDFESV QLPSMTSVIK GAANYDKDAV
VGSIAVPSVD VNLLVFKGTN TANLLAGATT MRSDQVMGKG
NYPLAGHHMR DESMLFGPIM KVKKGDKIYL TDLENLYEYT
VTETKTIDET EVSVIDDTKD ARITLITCDK PTETTKRFVA
VGELEKTEKL TKELENKYFP SK (SEQ ID NO: 161).

[0173] 衍生自单核增生李斯特菌分选酶A的成熟可溶性分选酶A的氨基酸序列具有以下氨基酸序列:

[0174]

ANYDKDAVVG SIAVPSVDVN LLVFKGTNTA NLLAGATTMR
SDQVMGKGN Y PLAGHHMRDE SMLFGPIMKV KKGDKIYLT
LENLYEYTVT ETKTIDETE V SVIDDTKDAR ITLITCDKPT
ETTKRFVAVG ELEKTEKLT K ELENKYFPSK

(SEQ ID NO: 06)

[0175] V. 如本文报导的方法和用途

[0176] 如本文中报导的一个方面是一种用于酶促产生多肽的方法,所述方法包括以下步骤:

[0177] -在如本文报告的深共晶溶剂中温育

[0178] i) (任选地在100个C端氨基酸残基范围内)包含氨基酸序列LPXTG (SEQ ID NO:01, 其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101, 其中X可以是任何氨基酸残基)的第一多肽,

[0179] ii) i) 在其N端具有甘氨酸、丙氨酸或半胱氨酸化合物,或 ii) 在其N端具有寡甘氨酸、或寡丙氨酸、或半胱氨酸氨基酸残基后接一个至三个甘氨酸或丙氨酸氨基酸残基,或 iii) 在其5个N端氨基酸残基范围内具有赖氨酸氨基酸残基的第二多肽,和

[0180] iii) 具有分选酶A活性的第三多肽,

[0181] 和由此产生多肽。

[0182] 在一个实施方案中,具有分选酶A活性的第三多肽衍生自金黄色葡萄球菌分选酶A,或衍生自酿脓链球菌分选酶A,或衍生自单核增生李斯特菌分选酶A。在一个实施方案中,第三多肽包含SEQ ID NO:05、SEQ ID NO:06、SEQ ID NO:27、SEQ ID NO:156、SEQ ID NO:159或SEQ ID NO:161的氨基酸序列。在一个优选的实施方案中,第三多肽包含SEQ ID NO:05或06的氨基酸序列。

[0183] 在一个实施方案中,第三多肽额外地在其N端或C端包含直接或通过间插接头缀合的标签。在一个实施方案中,第三多肽由SEQ ID NO:05的氨基酸序列组成并且C端标签由SEQ ID NO:32组成。在一个实施方案中,第三多肽由SEQ ID NO:05的氨基酸序列组成。

[0184] 在一个实施方案中,该方法用于两种多肽的酶促缀合。

[0185] 在一个实施方案中,深共晶溶剂包含氯化胆碱。在一个实施方案中,深共晶溶剂是氯化胆碱与二硫苏糖醇按摩尔比1:2.5至1:3或与2-巯基乙磺酸钠按摩尔比1:1的混合物。在一个实施方案中,深共晶溶剂包含含水共溶剂。在一个实施方案中,深共晶溶剂包含多于0%和直至约10% (v/v) 的共溶剂。在一个实施方案中,深共晶溶剂包含多于0%和多达至约5% (v/v) 的共溶剂。在一个实施方案中,深共晶溶剂包含从约1%、2%、3%或4%和直至约6%、7%、8%、9%或10% (v/v) 的含水共溶剂。下限值和上限值的任何组合是本文中的实施方案。在一个优选的实施方案中,深共晶溶剂是氯化胆碱与二硫苏糖醇甘油按摩尔比1:2.5至1:3或与2-巯基乙磺酸钠按摩尔比1:1的混合物,包含多于0%和直至约5% (v/v) 的含水共溶剂。

[0186] 在一个实施方案中,第二多肽在其N端具有氨基酸序列GGG、AAA、CGG、CAA、KGG或KAA (SEQ ID NO:29、162-166)。

[0187] 在一个实施方案中,第一多肽(任选地在250个C端氨基酸残基范围内)包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。在一个实施方案中,第一多肽(任选地在100个C端氨基酸残基范围内)包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。在一个实施方案中,第一多肽在25个C端氨基酸残基范围内包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。在一个实施方案中,第一多肽在10个C端氨基酸残基范围内包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。

[0188] 在一个实施方案中,第一多肽在其C端包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)。在一个实施方案中,第一多肽在其C端包含氨基酸序列LPETG (SEQ ID NO:04)或LPETA (SEQ ID NO:108)。

[0189] 在一个实施方案中,第一多肽和第二多肽彼此独立地选自抗体可变结构域、抗体重链Fab片段、抗体Fc区、标签和包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)的肽、接头和非分选酶基序部分。

[0190] 如本文中报导了的一个方面是深共晶溶剂作为溶剂在分选酶A催化的酶促转酰胺化反应中的用途,所述深共晶溶剂按摩尔比1:2.5至1:3包含氯化胆碱和二硫苏糖醇或按摩尔比1:1包含氯化胆碱和2-巯基乙磺酸钠,并且还包含直至约5% (v/v) 的含水共溶剂。

[0191] 如本文中报告的一个方面是一种分选酶A催化产生多肽的方法,所述方法包括以下步骤

[0192] -在基于硫羟的深共晶溶剂中温育

[0193] i) (任选地在20个C端氨基酸残基范围内)包含氨基酸序列LPXTG (SEQ ID NO:01,其中X可以是任何氨基酸残基)或LPXTA (SEQ ID NO:101,其中X可以是任何氨基酸残基)的第一多肽,

[0194] ii) i) 在其N端具有甘氨酰、丙氨酰基或半胱氨酰化合物,或 ii) 在其N端具有寡甘

氨酸、或寡丙氨酸、或半胱氨酸氨基酸残基后接一个至三个甘氨酸或丙氨酸氨基酸残基，或 iii) 在其5个N端氨基酸范围内具有赖氨酸氨基酸残基的第二多肽，和

[0195] iii) 具有SEQ ID NO:05、SEQ ID NO:06、SEQ ID NO:27、SEQ ID NO:156、SEQ ID NO:159或SEQ ID NO:161的氨基酸序列并且任选地包含SEQ ID NO:32的C端标签的可溶性分选酶，

[0196] 其中基于巯基的深共晶溶剂包含

[0197] a) 摩尔比1:2.5至1:3的氯化胆碱和二硫苏糖醇，或

[0198] b) 摩尔比约1:2的氯化胆碱和2-巯基乙醇，或

[0199] c) 摩尔比约1:2的氯化胆碱和4-巯基-1-丁醇，或

[0200] d) 摩尔比约1:2的氯化胆碱和1-巯基-2-丙醇，或

[0201] e) 摩尔比约1:1的氯化胆碱和2-巯基乙磺酸钠，

[0202] 和

[0203] 多于0%和直至约10% (v/v) 的含水共溶剂

[0204] 和由此产生多肽。

[0205] 第一或第二多肽

[0206] 如果分选酶基序(氨基酸序列)未直接包含于以下这些分子之一中，则分选酶基序可以与之缀合或并入其中：治疗剂(药物)、细胞毒性剂(例如毒素如多柔比星或百日咳毒素)、荧光团如荧光染料如荧光素或罗丹明、用于成像性金属或放疗性金属的螯合剂、肽酰基或非肽酰基标记物、标签、或清除调节剂如聚乙二醇的多种异构体、与第三组分、另一种糖或亲脂物质结合的肽、或小分子，例如合成的小分子(例如乙酰水杨酸)。如果基序借助缀合过程并入，这种缀合过程可以直接进行或借助间接接头进行。另外，第一和/或第二多肽可以重组地产生或可以是合成的或半合成的，即重组产生并此后经化学修饰。

[0207] a) 治疗剂

[0208] 治疗剂可以是具有治疗效果的任何化合物、部分或基团，例如抗体、细胞毒性或抑制细胞性化合物。抗体可以是全长或完整的抗体或其抗原结合片段。

[0209] 已知针对细胞表面分子及其配体的许多治疗性抗体，如Rituxan/MabThera/利妥昔单抗、2H7/奥瑞珠单抗(Ocrelizumab)、Zevalin/替伊莫单抗(Ibrizumomab)、Arzerra/奥法木单抗(Ofatumumab) (CD20)、HLL2/依帕珠单抗、奥英妥珠单抗(Inotuzomab) (CD22)、Zenapax/达克珠单抗、Simulect/巴利昔单抗(Basiliximab) (CD25)、赫赛汀/曲妥珠单抗、培妥珠单抗(Her2/ERBB2)、Mylotarg/吉妥珠单抗(CD33)、Raptiva/依法珠单抗(Cd11a)、爱必妥(Erbitux)/西妥昔单抗(EGFR, 表皮生长因子受体)、IMC-1121B(VEGF受体2)、Tysabri/那他珠单抗($\alpha 4\beta 1$ 和 $\alpha 4\beta 7$ 整联蛋白的 $\alpha 4$ 亚基)、ReoPro/阿昔单抗(abciximab) (gpIIb-gpIIa和 $\alpha v\beta 3$ 整联蛋白)、Orthoclone OKT3/莫罗单抗CD3(CD3)、Benlysta/贝利单抗(Belimumab) (BAFF)、Tolerx/Oteliximab(CD3)、Soliris/依库珠单抗(Eculizumab) (C5补体蛋白)、Actemra/托珠单抗(Tocilizumab) (IL-6R)、Panorex/依决洛单抗(EpCAM, 上皮细胞黏附分子)、CEA-CAM5/拉贝珠单抗(Labetuzumab) (CD66/CEA, 癌胚抗原)、CT-11(PD-1, 程序性死亡-1T细胞抑制性受体, CD-d279)、H224G11(c-Met受体)、SAR3419(CD19)、IMC-A12/西妥木单抗(Cixutumumab) (IGF-1R、胰岛素样生长因子1受体)、MEDI-575(PDGF-R、血小板衍生长因子受体)、CP-675、206/曲美木单抗(Tremelimumab) (细胞毒T淋巴细胞抗原4)、

R05323441 (胎盘生长因子或PGF)、HGS1012/马帕木单抗 (Mapatumumab) (TRAIL-R1)、SGN-70 (CD70)、维多汀 (SGN-35)/贝伦妥单抗 (Brentuximab) (CD30) 和ARH460-16-2 (CD44)。

[0210] 用如本文报告的方法所获得的缀合物可以用于制备药物,所述药物用于治疗例如肿瘤学疾病、心血管疾病、传染病、炎性疾病、自身免疫疾病、代谢(例如,内分泌)疾病或神经学(例如神经变性)疾病。这些疾病的示例性非限制例子是阿尔茨海默病、非霍奇金淋巴瘤、B细胞急性和慢性淋巴样白血病、Burkitt淋巴瘤、霍奇金淋巴瘤、多毛细胞白血病、急性和慢性髓样白血病、T细胞淋巴瘤和白血病、多发性骨髓瘤、胶质瘤、Waldenstrom巨球蛋白血症、癌(如口腔癌、胃肠道癌、结肠癌、胃癌、肺气道(pulmonary tract)癌、肺癌、乳腺癌、卵巢癌、前列腺癌、子宫癌、子宫内膜癌、子宫颈癌、膀胱癌、胰癌、骨癌、肝脏癌、胆囊癌、肾癌、皮肤癌和睾丸癌)、黑素瘤、肉瘤、胶质瘤、和皮肤癌、急性特发性血小板减少性紫癜、慢性特发性血小板减少性紫癜、皮炎、Sydenham舞蹈病、重症肌无力、系统性红斑狼疮、狼疮肾炎、风湿热、多内分泌腺综合征、大疱性类天疱疮、糖尿病、Henoch-Schonlein紫癜、链球菌感染后肾炎、结节性红斑、Takayasu动脉炎、阿狄森病、类风湿性关节炎、多发性硬化、肉样瘤病、溃疡性结肠炎、多形性红斑、IgA肾病、结节性多发性动脉炎、强直性脊柱炎、肺出血肾炎综合征、闭塞性血栓性脉管炎、Sjogren综合征、原发性胆汁性肝硬化、桥本氏甲状腺炎 (Hashimoto's thyroiditis)、甲状腺毒症、硬皮病、慢性活动性肝炎、多发性肌炎/皮炎、多软骨炎、寻常天疱疮、韦格纳肉芽肿病 (Wegener's granulomatosis)、膜性肾病、肌萎缩侧索硬化、脊髓痨 (tabes dorsalis)、巨细胞动脉炎/多肌痛、恶性贫血、快速进展性肾小球肾炎、银屑病或纤维化肺泡炎。

[0211] 许多细胞表面标志物及其配体是已知的。例如已经报导癌细胞表达至少一种以下细胞表面标志物和/或配体,包括但不限于碳酸酐酶IX、甲胎蛋白、 α -辅肌动蛋白-4、A3(对A33抗体特异的抗原)、ART-4、B7、Ba-733、BAGE、BrE3抗原、CA125、CAMEL、CAP-1、CASP-8/m、CCCL19、CCCL21、CD1、CD1a、CD2、CD3、CD4、CDS、CD8、CD1-1A、CD14、CD15、CD16、CD18、CD19、CD20、CD21、CD22、CD23、CD25、CD29、CD30、CD32b、CD33、CD37、CD38、CD40、CD40L、CD45、CD46、CD54、CD55、CD59、CD64、CD66a-e、CD67、CD70、CD74、CD79a、CD80、CD83、CD95、CD126、CD133、CD138、CD147、CD154、CDC27、CDK-4/m、CDKN2A、CXCR4、CXCR7、CXCL12、HIF-1- α 、结肠特异性抗原-p (CSAp)、CEA (CEACAM5)、CEACAM6、c-met、DAM、EGFR、EGFRvIII、EGP-1、EGP-2、ELF2-M、Ep-CAM、Flt-1、Flt-3、叶酸受体、G250抗原、GAGE、GROB、HLA-DR、HM1.24、人绒毛膜促性腺素释放激素 (HCG) 及其亚基、HER2/neu、HMGB-1、低氧诱导因子 (HIF-1)、HSP70-2M、HST-2或1a、IGF-1R、IFN- γ 、IFN- α 、IFN- β 、IL-2、IL-4R、IL-6R、IL-13R、IL-15R、IL-17R、IL-18R、IL-6、IL-8、IL-12、IL-15、IL-17、IL-18、IL-25、胰岛素样生长因子-1 (IGF-1)、KC4抗原、KS-1抗原、KS1-4、Le-Y、LDR/FUT、巨噬细胞移行抑制因子 (MIF)、MAGE、MAGE-3、MART-1、MART-2、NY-ESO-1、TRAG-3、mCRP、MCP-1、MIP-1A、MIP-1B、MIF、MUC1、MUC2、MUC3、MUC4、MUC5、MUM-1/2、MUM-3、NCA66、NCA95、NCA90、胰腺癌黏蛋白、胎盘生长因子、p53、PLAGL2、前列腺酸性磷酸酶、PSA、PRAME、PSMA、P1GF、ILGF、ILGF-1R、IL-6、IL-25、RS5、RANTES、T101、SAGE、S100、存活素、存活素-2B、TAC、TAG-72、生腱蛋白、TRAIL受体、TNF- α 、Tn抗原、Thomson-Friedenreich抗原、肿瘤坏死抗原、VEGFR、ED-B纤连蛋白、WT-1、17-1A抗原、补体因子C3、C3a、C3b、C5a、C5、血管生成标志物、bc1-2、bc1-6、Kras、cMET、癌基因标志物和癌基因产物(例如参见, Sensi等人, Clin. Cancer Res. 12 (2006) 5023-5032; Parmiani等人, J. Immunol. 178 (2007)

1975-1979;Novellino等人,Cancer Immunol.Immunother.54(2005)187-207)。

[0212] 因此,识别特定细胞表面受体(包括其配体)的抗体可以用于特异性和选择性靶向并结合至许多/多种与疾病相关的细胞表面标志物。细胞表面标志物是位于例如与信号传导事件或配体结合相关的细胞(例如疾病相关细胞)的表面上的多肽。

[0213] 在一个实施方案中,为了治疗癌症/肿瘤,产生靶向肿瘤相关抗原的多特异性结合分子/双特异性抗体,如在Herberman,“Immunodiagnosis of Cancer”,引自Fleisher(编著),“The Clinical Biochemistry of Cancer”,第347页(American Association of Clinical Chemists(1979))中和在US4,150,149;US 4,361,544和US 4,444,744中报导的那些。

[0214] 关于肿瘤相关抗原(TAA)的报告包括Mizukami等人(Nature Med.11(2005)992-997);Hatfield等人(Curr.Cancer Drug Targets 5(2005)229-248);Vallbohmer等人(J Clin.Oncol.23(2005)3536-3544)和Ren等人(Ann.Surg.242(2005)55-63),每篇文献就确定的TAA而言通过引用方式并入本文作为参考。

[0215] 在疾病涉及淋巴瘤、白血病或自身免疫疾病的情况下,靶向的抗原可以选自CD4、CD5、CD8、CD14、CD15、CD19、CD20、CD21、CD22、CD23、CD25、CD33、CD37、CD38、CD40、CD40L、CD46、CD54、CD67、CD74、CD79a、CD80、CD126、CD138、CD154、CXCR4、B7、MUC1或1a、HM1.24、HLA-DR、生腱蛋白、VEGF、PIGF、ED-B纤连蛋白、癌基因、癌基因产物(例如,c-met或PLAGL2)、CD66a-d、坏死抗原、IL-2、T101、TAG、IL-6、MIF、TRAIL-R1(DR4)和TRAIL-R2(DR5)。

[0216] 已知针对两种不同靶如BCMA/CD3、组合的HER家族不同抗原(EGFR、HER2、HER3)、CD19/CD3、IL17RA/IL7R、IL-6/IL-23、IL-1-β/IL-8、IL-6或IL-6R/IL-21或IL-21R的许多双特异性抗体,第一特异性针对选自Lewis x-结构、Lewis b-结构和Lewis y-结构、Globo H-结构、KH1、Tn抗原、TF抗原的抗原的糖表位、和黏蛋白、CD44、糖脂和糖鞘脂如Gg3、Gb3、GD3、GD2、Gb5、Gm1、Gm2、唾液酰四糖基神经酰胺的糖结构;并且第二特异性针对选自EGFR、HER2、HER3和HER4、GD2的ErbB受体酪氨酸激酶,所述第二特异性与下述第二抗原结合位点组合,所述第二抗原结合位点与选自T淋巴细胞NK细胞、B-淋巴细胞、树突细胞、单核细胞、巨噬细胞、嗜中性粒细胞、间充质干细胞、神经干细胞的免疫细胞、ANG2/VEGF、VEGF/PDGFR-β、血管内皮生长因子(VEGF)接纳体2/CD3、PSMA/CD3、EPCAM/CD3、选自VEGFR-1、VEGFR-2、VEGFR-3、FLT3、c FMS/CSF1R、RET、c-Met、EGFR、Her2/neu、HER3、HER4、IGFR、PDGFR、c-KIT、BCR、整联蛋白和MMP的抗原组合、与选自VEGF、EGF、PIGF、PDGF、HGF和血管生成素、ERBB-3/C-MET、ERBB-2/C-MET、EGF受体1/CD3、EGFR/HER3、PSCA/CD3、C-MET/CD3、内皮唾液酸蛋白(Endosialin)/CD3、EPCAM/CD3、IGF-1R/CD3、FAPALPHA/CD3、EGFR/IGF-1R、IL 17A/F、EGF受体1/CD3和CD19/CD16的水溶性配体缔合。

[0217] 毒性药物部分包括:(i)可以充当微管蛋白抑制剂、有丝分裂抑制剂、拓扑异构酶抑制剂或DNA嵌入剂的化疗药;(ii)可以以酶促方式发挥作用的蛋白质毒素;和(iii)放射性同位素。

[0218] 示例性毒性药物部分包括但不限于美坦生类、澳瑞司他汀(auristatin)、多拉司他汀、单端孢霉烯族化合物(trichothecene)、CC1065、刺孢霉素和其他烯二炔类抗生素、紫杉烷、蒽环类及其立体异构体、电子等排体、类似物或衍生物。

[0219] 蛋白质毒素包括白喉毒素A链、白喉毒素的非结合活性片段、外毒素A链(来自铜绿

假单胞菌 (*Pseudomonas aeruginosa*)、蓖麻毒蛋白A链 (Vitetta等人 (1987) *Science*, 238: 1098)、相思豆毒蛋白A链、蒴莲根毒蛋白A链、 α -帚曲菌素、油桐 (*Aleurites fordii*) 蛋白、香石竹毒蛋白、垂序商陆 (*Phytolaca americana*) 蛋白 (PAPI、PAPII和PAP-5)、苦瓜 (*Momordica charantia*) 抑制剂、麻疯树毒蛋白、巴豆毒蛋白、肥皂草 (*Saponaire officinalis*) 抑制剂、多花白树毒蛋白、丝林霉素 (mitogellin)、局限曲菌素、酚霉素、伊诺霉素和单端孢霉烯类 (WO 93/21232)。

[0220] 治疗性放射性同位素包括 ^{32}P 、 ^{33}P 、 ^{90}Y 、 ^{125}I 、 ^{131}I 、 ^{131}In 、 ^{153}Sm 、 ^{186}Re 、 ^{188}Re 、 ^{211}At 、 ^{212}B 、 ^{212}Pb 和Lu的放射性同位素。

[0221] 放射性同位素或其他标记物可以按已知方式掺入 (Fraker等人 (1978) *Biochem Biophys Res Commun.* 80:49-57; "Monoclonal Antibodies in Immunoscintigraphy" Chatal, CRC Press 1989)。碳-14-标记的1-异硫氰酸根合苜蓿基-3-甲基二乙三胺五乙酸 (MX-DTPA) 是将放射性核素与复合物缀合的示例性螯合剂 (WO 94/11026)。

[0222] b) 标记物

[0223] 非分选酶基序部分可以是标记物。可以使用能与分选酶氨基酸序列共价结合的任何标记物部分 (参见例如Singh等人 (2002) *Anal. Biochem.* 304:147-15; Harlow E. 和 Lane, D. (1999) *Using Antibodies: A Laboratory Manual*, Cold Springs Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Lundblad R.L. (1991) *Chemical Reagents for Protein Modification*, 第2版, CRC Press, Boca Raton, Fla.)。标记物可以发挥以下作用: (i) 提供可检测信号; (ii) 与第二标记物相互作用以调整第一或第二标记物提供的可检测信号, 例如, 以产生FRET (荧光共振能量转移); (iii) 通过电荷、疏水性、形状或其他物理参数影响迁移率例如电泳迁移率或细胞通透性, 或 (iv) 提供捕获部分例如以调节离子复合过程。

[0224] 包含如本文报告的半抗原化标记物的缀合物可以用于诊断测定法中, 例如, 用于检测特定细胞、组织或血清中目的抗原的表达。对于诊断性应用, 使用其中第一结合特异性与靶结合并且第二结合特异性与半抗原化标记物结合的双特异性抗体。半抗原一般用可检测部分标记。许多标记物是可用的, 它们通常可以划分成以下类别:

[0225] (a) 放射性同位素 (放射性核素), 如 ^3H 、 ^{11}C 、 ^{14}C 、 ^{18}F 、 ^{32}P 、 ^{35}S 、 ^{64}Cu 、 ^{68}Gn 、 ^{86}Y 、 ^{89}Zr 、 $^{99\text{Tc}}$ 、 ^{111}In 、 ^{123}I 、 ^{124}I 、 ^{125}I 、 ^{131}I 、 ^{133}Xe 、 ^{177}Lu 、 ^{211}At 或 ^{212}Bi 。放射性同位素标记的缀合物用于受体定向的成像实验中。使用 *Current Protocols in Immunology*, (1991) 第1和第2卷, Coligen等人编著, Wiley-Interscience, New York, N.Y., Pubs中所述的技术, 抗原 (半抗原) 可以用结合、螯合或络合放射性同位素金属的配体试剂标记。可以络合金属离子的螯合配体包括DOTA、DOTP、DOTMA、DTPA和TETA (Macrocyclics, Dallas, Tex.)。放射性核素可以通过用如本文报告的复合物复合进行靶向 (Wu等人, *Nature Biotechnology* 23 (9) (2005) 1137-1146)。采用放射性核素标记的复合物的受体靶成像法可以通过检测和定量肿瘤组织中复合物或相应治疗性抗体进展性蓄积提供途径活化标志物 (Albert等人 (1998) *Bioorg. Med. Chem. Lett.* 8:1207-1210)。

[0226] 合适作为成像实验用标记物的金属-螯合复合物 (US 2010/0111856; US 5,342,606; US 5,428,155; US 5,316,757; US 5,480,990; US 5,462,725; US 5,428,139; US 5,385,893; US 5,739,294; US 5,750,660; US 5,834,456; Hnatowich等人,

J. Immunol. Methods 65 (1983) 147-157; Meares 等人, Anal. Biochem. 142 (1984) 68-78; Mirzadeh 等人, Bioconjugate Chem. 1 (1990) 59-65; Meares 等人, J. Cancer (1990), Suppl. 10: 21-26; Izard 等人, Bioconjugate Chem. 3 (1992) 346-350; Nikula 等人, Nucl. Med. Biol. 22 (1995) 387-90; Camera 等人, Nucl. Med. Biol. 20 (1993) 955-62; Kukis 等人, J. Nucl. Med. 39 (1998) 2105-2110; Verel 等人, J. Nucl. Med. 44 (2003) 1663-1670; Camera 等人, J. Nucl. Med. 21 (1994) 640-646; Ruegg 等人, Cancer Res. 50 (1990) 4221-4226; Verel 等人, J. Nucl. Med. 44 (2003) 1663-1670; Lee 等人, Cancer Res. 61 (2001) 4474-4482; Mitchell 等人, J. Nucl. Med. 44 (2003) 1105-1112; Kobayashi 等人, Bioconjugate Chem. 10 (1999) 103-111; Miederer 等人, J. Nucl. Med. 45 (2004) 129-137; DeNardo 等人, Clinical Cancer Research 4 (1998) 2483-90; Blend 等人, Cancer Biotherapy & Radiopharmaceuticals 18 (2003) 355-363; Nikula 等人, J. Nucl. Med. 40 (1999) 166-76; Kobayashi 等人, J. Nucl. Med. 39 (1998) 829-36; Mardirossian 等人, Nucl. Med. Biol. 20 (1993) 65-74; Roselli 等人, Cancer Biotherapy & Radiopharmaceuticals, 14 (1999) 209-20)。

[0227] (b) 荧光标记物如稀土元素螯合物(钆螯合物)、荧光素型荧光标记物,包括FITC、5-羧基荧光素、6-羧基荧光素;罗丹明型荧光标记物,包括TAMRA;丹磺酰;丽丝胺;花菁;藻红蛋白;得克萨斯红;及其类似物。荧光标记物可以使用在例如Current Protocols in Immunology,上文中公开的技术与抗原(半抗原)缀合。荧光染料和荧光标记物试剂包括从Invitrogen/Molecular Probes (Eugene, Oregon, 美国)和Pierce Biotechnology, Inc (Rockford, Ill.)市售的那些。

[0228] 检测标记物如荧光染料和化学发光染料 (Briggs 等人“Synthesis of Functionalised Fluorescent Dyes and Their Coupling to Amines and Amino Acids”, J. Chem. Soc., Perkin-Trans. 1 (1997) 1051-1058) 提供了可检测信号,并且通常适用于标记,特别具有以下特性:(i) 标记的缀合物应当产生极高信号且具低背景,从而可以在无细胞和基于细胞的测定法中灵敏地检测到少量缀合物;和(ii) 标记的缀合物应当是光稳定的,从而可以观察、监测并记录荧光信号,而没有明显的光漂白过程。对于涉及标记的缀合物与膜或细胞表面(尤其活细胞)的细胞表面结合作用的应用,标记物应当(iii) 具有良好的水溶性以实现有效的缀合物浓度和检测灵敏度并且(iv) 对活细胞无毒,从而不破坏细胞的正常代谢过程或造成过早细胞死亡。

[0229] (c) 可获得或公开了多种酶-底物标记物(参见,例如US 4,275,149)。酶通常催化生色底物的化学改变,其中可以使用多种技术测量所述化学改变。例如,酶可以催化底物中能以分光光度形式测量的颜色变化。备选地,酶可以改变底物的荧光或化学发光。化学发光底物因化学反应被电子激发并且可以随后发射(例如使用化学发光计)可测量的光或馈赠能量至荧光受体。酶标记物的例子包括萤光素酶(例如,萤火虫萤光素酶和细菌萤光素酶;US4,737,456)、萤光素、2,3-二氢酞嗪二酮、苹果酸脱氢酶、脲酶、过氧化物酶如辣根过氧化物酶(HRP)、碱性磷酸酶(AP)、 β -半乳糖苷酶、葡糖淀粉酶、溶菌酶、糖氧化酶(例如,葡萄糖氧化酶、半乳糖氧化酶和葡萄糖-6-磷酸脱氢酶)、杂环氧化酶(如尿酸酶和黄嘌呤氧化酶)、乳过氧化物酶、微过氧化物酶等。用于酶与多肽缀合的技术在O'Sullivan等人“Methods for the Preparation of Enzyme-Antibody Conjugates for use in Enzyme

Immunoassay”,引自Methods in Enzym. (J.Langone和IT Van Vunakis编著),Academic Press,New York,73 (1981) 147-166中描述。

[0230] 酶-底物组合的例子(US 4,275,149;US 4,318,980)例如包括:

[0231] (i) 以过氧化氢作为底物的辣根过氧化物酶(HRP),其中所述过氧化氢氧化染料前体(例如,邻苯二胺(OPD)或3,3',5,5'-四甲基联苯胺盐酸盐(TMB));

[0232] (ii) 以对硝基苯磷酸酯作为生色底物的碱性磷酸酶(AP);和

[0233] (iii) β -D-半乳糖苷酶(β -D-Gal)和生色底物(例如,对硝基苯- β -D-半乳糖苷酶)或荧光底物4-甲基伞形基- β -D-半乳糖苷酶。

[0234] 如本文报告的标记缀合物可以用于任何已知的测定法中,如ELISA、竞争性结合测定法、直接和间接夹心测定法和免疫沉淀测定法(Zola, Monoclonal Antibodies: A Manual of Techniques (1987) 第147-158页, CRC Press, Inc.)。

[0235] 如本文报告的标记缀合物由生物学成像和分子成像的多种方法和技术用作成像性生物标志物和探针,如:(i) MRI(磁共振成像术);(ii) MicroCT(计算机断层成像);(iii) SPECT(单光子发射计算机断层成像);(iv) PET(正电子发射断层摄影术) Tinianow, J. 等人, Nuclear Medicine and Biology, 37 (3) (2010) 289-297; Chen等人, Bioconjugate Chem. 15 (2004) 41-49; US 2010/0111856; (v) 生物发光法; (vi) 荧光法; 和 (vii) 超声法。免疫闪烁照相术是一种成像方法,其中将放射性物质标记的缀合物施用至动物或人患者并且拍摄身体内缀合物存在的部位的画面(US 6,528,624)。成像性生物标志物可以客观地测量并且作为正常生物学过程、发病过程或治疗性介入的药理学反应的指示物进行评价。生物标志物可以是几种类型的:0型标志物是疾病的自然历史标志物并且与已知的临床指标纵向相关,例如类风湿性关节炎中滑膜炎的MRI评价;I型标志物根据作用机制捕获介入治疗的效果,即便该机制可能与临床结局不相关;II型标志物充当替代终点,其中生物标志物的变化或来自生物标志物的信号预测出“确认”靶向反应的临床益处,如类风湿性关节炎中通过CT测量的骨侵蚀。成像性生物标志物因而可以提供关于以下方面的药物动力学(PD)治疗信息:(i) 靶蛋白的表达、(ii) 治疗药与靶蛋白的结合(即选择性)、和(iii) 清除率和半寿期药物代谢动力学数据。相对于基于实验室的生物标志物而言,体内成像性生物标志物的优点包括:非侵入性处理、可定量、全身评估、反复给药和评估(即多个时间点)和从临床前结果(小动物)至临床结果(人类)的潜在可转移效果。对于一些应用,生物成像法取代了临床前研究中的动物实验或使其数目最小化。

[0236] 肽标记方法是熟知的。参见Haugland (2003) Molecular Probes Handbook of Fluorescent Probes and Research Chemicals, Molecular Probes, Inc.; Brinkley (1992) Bioconjugate Chem. 3:2; Garman, (1997) Non-Radioactive Labeling: A Practical Approach, Academic Press, London; Means (1990) Bioconjugate Chem. 1:2; Glazer等人 Chemical Modification of Proteins. Laboratory Techniques in Biochemistry and Molecular Biology (T.S. Work和E. Work编著) American Elsevier Publishing Co., New York; Lundblad, R.L. 和 Noyes, C.M. (1984) Chemical Reagents for Protein Modification, 第I卷和第II卷, CRC Press, New York; Pfleiderer, G. (1985) "Chemical Modification of Proteins", Modern Methods in Protein Chemistry, H. Tschesche 编著, Walter DeGruyter, Berlin and New York; 以及 Wong (1991) Chemistry of Protein

Conjugation and Cross-linking, CRC Press, Boca Raton, Fla.); DeLeon-Rodriguez 等人, Chem. Eur. J. 10 (2004) 1149-1155; Lewis 等人, Bioconjugate Chem. 12 (2001) 320-324; Li 等人, Bioconjugate Chem. 13 (2002) 110-115; Mier 等人 Bioconjugate Chem. 16 (2005) 240-237。

[0237] c) 接头

[0238] 术语“接头”指可以用来将第一部分与第二部分缀合(连接)的双官能或多官能部分。可以使用具有二个反应性官能团的接头,便利地制备连接的缀合物。

[0239] 在一个实施方案中,接头具有反应部位,所述反应部位具有与分选酶氨基酸序列中存在的亲核基团反应的亲电基团。有用的亲电基团包括,但不限于另一个巯基、马来酰亚胺和卤代乙酰胺基团。(参见例如在 Klussman 等人, Bioconjugate Chemistry 15 (4) (2004) 765-773 的第 766 页的缀合方法)。

[0240] 巯基反应官能团的例子包括但不限于巯基、马来酰亚胺、 α -卤代乙酰基、活化酯如琥珀酰亚胺酯、4-硝基苯基酯、全氟苯基酯、四氟苯基酯、酞、酰氯、磺酰氯、异氰酸酯和异硫氰酸酯。

[0241] 接头可以包含将分选酶氨基酸序列连接至非分选酶基序部分的氨基酸残基。氨基酸残基可以形成二肽、三肽、四肽、五肽、六肽、七肽、八肽、九肽、十肽、十一肽或十二肽单元。氨基酸残基包括天然存在的那些以及非天然存在的氨基酸类似物,例如瓜氨酸或 β -氨基酸,例如 β -丙氨酸,或 ω -氨基酸如 4-氨基-丁酸。

[0242] 在另一个实施方案中,接头具有反应性官能团,所述反应性官能团具有与分选酶氨基酸序列中存在的亲电基团反应的亲核基团。有用的亲电基团包括但不限于醛基和酮羰基。接头的亲核基团的杂原子可以与分选酶氨基酸序列中的亲电基团反应并与分选酶氨基酸序列形成共价键。接头上有用的亲核基团包括但不限于酰肼、肟、氨基、肼、硫代半卡巴腓、肼羧化物和芳酰肼。抗原(半抗原)上的亲电基团为连接至接头提供了便利位点。

[0243] 一般,肽型接头可以通过在两个或更多个氨基酸和/或肽片段之间形成肽键来制备。此类肽键可以例如根据肽化学领域熟知的液相合成方法制备(E. Schroder 和 K. Lubke "The Peptides", 第 1 卷 (1965) 76-136, Academic Press)。

[0244] 在另一个实施方案中,接头可以用调节溶解度或反应性的基团取代。例如,带电取代基如磺酸盐($-\text{SO}_3^-$)或铵或聚合物如 PEG,可以增加试剂的水溶解度并且促进接头试剂与抗原(半抗原)或药物部分的偶联反应,或根据所用的合成途径,促进偶联反应。

[0245] 明确地构思了包含如本文报导的非分选酶基序部分的缀合物,但它们不限于,用以下接头试剂制备的复合物: BMPEO、BMPS、EMCS、GMBS、HBVS、LC-SMCC、MBS、MPBH、SBAP、SIA、SIAB、SMCC、SMPB、SMPH、硫代-EMCS、硫代-GMBS、硫代-KMUS、硫代-MBS、硫代-SIAB、硫代-SMCC 和硫代-SMPB 及 SVSB(琥珀酰亚胺-(4-乙烯基砵)苯甲酸酯)制备并且包含以下双马来酰亚胺试剂: DTME、BMB、BMDB、BMH、BMOE、BM(PEO)₃ 和 BM(PEO)₄, 所述接头试剂从 Pierce Biotechnology, Inc. 可商业获得。双马来酰亚胺试剂允许例如巯基以依次或同时方式与含巯基的药物部分、标记物或接头中间体接合。除马来酰亚胺之外,与例如巯基反应的其他官能团还包括碘乙酰胺、溴乙酰胺、乙烯基吡啶、二硫化物、吡啶基二硫化物、异氰酸酯和异硫氰酸酯。

[0246] 示例性接头包括具有马来酰亚胺伸张物和对氨基苄基氨甲酰基(PAB)自毁性间隔

团的缬氨酸-瓜氨酸(val-cit或vc)二肽接头试剂和具有马来酰亚胺伸张物单元和对氨基苄基自毁性间隔团的phe-lys(Mtr)二肽接头试剂。

[0247] 半胱氨酸硫羟基是亲核的,并且通过硫化物交换,能够与接头试剂和非分选酶基序部分或分选酶氨基酸序列上的亲电基团反应以形成共价键,其中亲电基团包括:(i)活性酯如NHS酯、HOBt酯、卤代甲酸酯和酰卤;(ii)烷基卤和苄基卤,如卤乙酰胺;(iii)醛、酮、羧基和马来酰亚胺基;和(iv)二硫化物,包括吡啶基二硫化物。半抗原化合物上的亲核基团包括但不限于能够与接头部分和接头试剂上的亲电基团反应以形成共价键的胺、巯基、羟基、酰肼、胍、胍、硫代半卡巴脲、胍羧化物和芳酰胍基。

[0248] V. 重组方法

[0249] 可以从真核细胞(例如HEK293细胞、CHO细胞)表达并且从其上清液纯化在其N端包含亲核性氨基酸序列(例如寡甘氨酸基序(GG(SEQ ID NO:28)、GGG(SEQ ID NO:29)、GGGG(SEQ ID NO:30)、GGGGG(SEQ ID NO:31))的任何多肽结构域(例如,单链抗原结合多肽如scFv、scFab或darpin或者多链抗原结合多肽如dsFv或Fab)。多肽是否为分离的多肽或包含于多聚体实体或异聚体实体并不重要。

[0250] 用于克隆或表达/分泌编码多肽的载体的合适宿主细胞包括本文所述的原核或真核细胞。例如,可以在细菌中产生多肽,尤其在不需要糖基化时如此(参见,例如,US 5,648,237、US 5,789,199和US 5,840,523,Charlton,Methods in Molecular Biology 248(2003)245-254(B.K.C.Lo(编著),Humana Press,Totowa,NJ),其描述在大肠杆菌中表达抗体片段)。在表达后,多肽可以从细菌细胞糊分离在可溶性级分中或可以自不溶性级分分离所谓的包涵体,其可以溶解并再折叠成生物学活性形式。此后,可以进一步纯化多肽。

[0251] 除原核生物之外,真核微生物如丝状真菌或酵母是编码多肽的载体的合适克隆宿主或表达宿主,包括其糖基化途径已经“人源化”的真菌和酵母菌株,导致产生具有部分或完全人类糖基化模式的多肽(参见例如Gerngross,Nat.Biotech.22(2004)1409-1414和Li等人,Nat.Biotech.24(2006)210-215)。

[0252] 用于表达糖基化多肽的合适宿主细胞也源自多细胞生物(无脊椎动物和脊椎动物)。无脊椎动物细胞的例子包括植物细胞和昆虫细胞。已经鉴定了可以与昆虫细胞一起使用,尤其用于转染草地贪夜蛾(Spodoptera frugiperda)细胞的众多杆状病毒毒株。

[0253] 也可以利用植物细胞培养物作为宿主(例如,见US 5,959,177、US 6,040,498、US 6,420,548、US 7,125,978和US 6,417,429(描述了在转基因植物中产生抗体的PLANTIBODIES™技术))。

[0254] 也可以使用脊椎动物细胞作为宿主。例如,适应于悬浮培养的哺乳动物细胞系可以有有用。可用的哺乳动物宿主细胞系的其他例子是COS-7细胞系(由SV40转化的猴肾CV1细胞);HEK293细胞系(人胚肾);BHK细胞系(幼仓鼠肾);TM4小鼠支持细胞系(在例如Mather,Biol.Reprod.23(1980)243-251中所述的TM4细胞);CV1细胞系(猴肾细胞);VERO-76细胞系(非洲绿猴肾细胞);HELA细胞系(人宫颈癌细胞);MDCK细胞系(犬肾细胞);BRL-3A细胞系(布法罗大鼠肝细胞);W138细胞系(人肺细胞);HepG2细胞系(人肝细胞);MMT 060562细胞系(小鼠乳腺癌细胞);TRI细胞系(例如,Mather等人,Anal.N.Y.Acad.Sci.383(1982)44-68中所述);MRC5细胞系;和FS4细胞系。其他可用的哺乳动物宿主细胞系包括CHO细胞系(中国仓鼠卵巢细胞),包括DHFR阴性CHO细胞系(例如参见Urlaub等人,Proc.Natl.Acad.Sci.USA

77 (1980) 4216); 和骨髓瘤细胞系如Y0、NS0和Sp2/0细胞系。关于适于产生多肽的某些哺乳动物宿主细胞系的综述, 参见, 例如, Yazaki和Wu, Methods in Molecular Biology, Antibody Engineering 248 (2004) 255-268 (B.K.C.Lo, (编著), Humana Press, Totowa, NJ)。

[0255] 附图简述

[0256] 图1实施例4的反应产物的色谱图; 8.433min=GGG-生物素, 13.350min=Sa-SrtA, 20.942=LCR640-LPETG (离析物), 21.987min=LCR640-LPETG-生物素 (产物)。

[0257] 提供以下实施例、附图和序列以辅助理解本发明, 本发明的真实范围在所附权利要求书中阐述。可以理解, 能对所述方法作出修改而不脱离本发明的精神。

实施例

[0258] 重组DNA技术

[0259] 如Sambrook, J. 等人, Molecular cloning: A laboratory manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989) 中所述的标准方法用来操作DNA。根据制造商的说明使用分子生物学试剂。

[0260] 基因和寡核苷酸合成

[0261] 通过化学合成在Geneart GmbH (雷根斯堡, 德国) 制备目的基因区段。将合成的基因片段克隆入大肠杆菌质粒用于增殖/扩增。通过DNA测序验证亚克隆基因片段的DNA序列。备选地, 通过复性化学合成的寡核苷酸或通过PCR装配短的合成性DNA片段。相应的寡核苷酸由metabion GmbH (Planegg-Martinsried, 德国) 制备。

[0262] 基本/标准哺乳动物表达质粒的描述

[0263] 为了表达目的基因/蛋白质 (例如全长抗体重链、全长抗体轻链或在其N端含有寡甘氨酸的Fc-链), 使用包含以下功能性元件的转录单位:

[0264] -来自人巨细胞病毒的立即早期增强子和启动子 (P-CMV), 包含内含子A,

[0265] -人重链免疫球蛋白5'非翻译区 (5' UTR),

[0266] -鼠免疫球蛋白重链信号序列,

[0267] -待表达的基因/蛋白质 (例如缩短的金黄色葡萄球菌分选酶A), 和

[0268] -牛生长激素多聚腺苷化序列 (BGH pA)。

[0269] 除了包含待表达目的基因的表达单元/盒之外, 基本/标准哺乳动物表达质粒还含有

[0270] 来自载体pUC18的允许这个质粒在大肠杆菌中复制的复制起点, 和

[0271] 在大肠杆菌中赋予氨苄青霉素抗性的 β -内酰胺酶基因。

[0272] 蛋白质测定

[0273] 通过使用基于多肽的氨基酸序列计算的摩尔消光系数在280nm测定光密度 (OD), 确定纯化多肽的蛋白质浓度。

[0274] 实施例1

[0275] 产生可溶性分选酶A的表达质粒

[0276] 金黄色葡萄球菌衍生的分选酶A

[0277] 分选酶基因编码N端截短的金黄色葡萄球菌分选酶A (60-206) 分子 (SEQ ID NO:05

的氨基酸序列)。

[0278] 除可溶性分选酶表达盒之外,在HEK293细胞中瞬时表达可溶性分选酶的表达质粒还包含来自载体pUC18的允许这个质粒在大肠杆菌中复制的复制起点和在大肠杆菌中赋予氨苄青霉素耐药性的 β -内酰胺酶基因。

[0279] 可溶性分选酶的转录单位包含以下功能性元件:

[0280] -来自人巨细胞病毒的立即早期增强子和启动子(P-CMV),包含内含子A,

[0281] -人重链免疫球蛋白5'非翻译区(5' UTR),

[0282] -鼠免疫球蛋白重链信号序列,

[0283] -编码纯化标签的核酸,

[0284] -编码N端截短型金黄色葡萄球菌分选酶A的核酸,和

[0285] -牛生长激素多聚腺苷化序列(BGH pA)。

[0286] 成熟可溶性分选酶的氨基酸序列是

[0287]

QAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSF AEENESLDDQNI SIAGHTFIDRPNYQFTNLKAA
KKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQK GKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK (SEQ ID
NO:05)。

[0288] 纯化标签具有氨基酸序列MRGSHHHHHHGS (SEQ ID NO:32)。

[0289] 化脓性葡萄球菌衍生的分选酶A

[0290] 分选酶基因编码N端截短的化脓性葡萄球菌分选酶A分子(SEQ ID NO:156的氨基酸序列)。

[0291] 除可溶性分选酶表达盒之外,在HEK293细胞中瞬时表达可溶性分选酶的表达质粒还包含来自载体pUC18的允许这个质粒在大肠杆菌中复制的复制起点和在大肠杆菌中赋予氨苄青霉素耐药性的 β -内酰胺酶基因。

[0292] 可溶性分选酶的转录单位包含以下功能性元件:

[0293] -来自人巨细胞病毒的立即早期增强子和启动子(P-CMV),包含内含子A,

[0294] -人重链免疫球蛋白5'非翻译区(5' UTR),

[0295] -鼠免疫球蛋白重链信号序列,

[0296] -编码纯化标签的核酸,

[0297] -编码N端截短型酿脓链球菌分选酶A的核酸,和

[0298] -牛生长激素多聚腺苷化序列(BGH pA)。

[0299] 成熟可溶性分选酶的氨基酸序列是

[0300]

VLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAGTMKEEQVMGGENNYS LASHHIFGITGSSQMLFSPLE
ERAQNGMSIYLT DKEKIYEYIIKDVFTVAPERVDV IDDTAGLKEVTLVTCTDIEATERIIVKGELKTEYDFDKAPAD
VLKAFNHSYNQVST (SEQ ID NO:156)。

[0301] 纯化标签具有氨基酸序列MRGSHHHHHHGS (SEQ ID NO:32)。

[0302] 单核增生李斯特菌衍生的分选酶A

[0303] 分选酶基因编码N端截短的单核增生李斯特菌分选酶A(73-222)分子(SEQ ID NO:06的氨基酸序列)。

[0304] 除可溶性分选酶表达盒之外,在HEK293细胞中瞬时表达可溶性分选酶的表达质粒还包含来自载体pUC18的允许这个质粒在大肠杆菌中复制的复制起点和在大肠杆菌中赋予氨苄青霉素耐药性的 β -内酰胺酶基因。

[0305] 可溶性分选酶的转录单位包含以下功能性元件:

[0306] -来自人巨细胞病毒的立即早期增强子和启动子(P-CMV),包含内含子A,

[0307] -人重链免疫球蛋白5'非翻译区(5' UTR),

[0308] -鼠免疫球蛋白重链信号序列,

[0309] -编码纯化标签的核酸,

[0310] -编码N端截短型单核增生李斯特菌分选酶A的核酸,和

[0311] -牛生长激素多聚腺苷化序列(BGH pA)。

[0312] 纯化标签具有氨基酸序列MRGSHHHHHHGS(SEQ ID NO:32)。

[0313] 实施例2

[0314] 瞬时表达和分析性表征

[0315] 通过瞬时转染在F17培养基(Invitrogen Corp.)中培养的(人胚肾细胞系293衍生的)HEK293细胞,进行重组生产。为了转染,使用“293-Fectin”转染试剂(Invitrogen)。如制造商的说明书中所述进行转染。转染后三至七(3-7)天收获细胞培养上清液。在降低的温度(例如-80°C)储存上清液。

[0316] 在Meissner,P.等人,Biotechnol.Bioeng.75(2001)197-203中给出关于例如HEK293细胞中重组表达人免疫球蛋白的一般信息。

[0317] 通过使用基于氨基酸序列计算的摩尔消光系数,在280nm测量光密度(OD)确定蛋白质浓度。在还原剂(5mM 1,4-二硫苏糖醇)存在和不存在下通过SDS-PAGE和考马斯亮兰染色,分析纯度。

[0318] 实施例3

[0319] 产生基于硫羟的深共晶溶剂

[0320] 二硫苏糖醇和氯化胆碱

[0321] 氯化胆碱和二硫苏糖醇按1:2.5摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.25mol二硫苏糖醇。将混合物在火焰上缓慢加热并振摇直至已经形成清亮、均一的溶液。

[0322] 使液体冷却至室温。产率是定量的并且产物具有低于室温的熔点。将约10g合成的深共晶溶剂置于封闭的小瓶中并在干燥的空间保存。

[0323] β -巯基乙醇和氯化胆碱

[0324] 氯化胆碱和 β -巯基乙醇按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol β -巯基乙醇。将混合物在火焰上缓慢加热并振摇直至已经形成均一和清亮的溶液。

[0325] 使液体冷却至室温。产率是定量的并且产物具有低于室温的熔点。将约10g合成的深共晶溶剂置于封闭的小瓶中并在干燥的空间保存。

[0326] 4-巯基-1-丁醇和氯化胆碱

[0327] 氯化胆碱和4-巯基-1-丁醇按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol 4-巯基-1-丁醇。将混合物在火焰上缓慢加热并振摇直至形成均一的

溶液。

[0328] 使液体冷却至室温。产率是定量的并且产物具有低于室温的熔点。将约10g合成的深共晶溶剂置于封闭的小瓶中并在干燥的空间保存。

[0329] 1-巯基-2-丙醇和氯化胆碱

[0330] 氯化胆碱和1-巯基-2-丙醇按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol 1-巯基-2-丙醇。将混合物在火焰上缓慢加热并振摇直至形成均一和清亮的溶液。

[0331] 使液体冷却至室温。产率是定量的并且产物具有低于室温的熔点。将约10g合成的深共晶溶剂置于封闭的小瓶中并在干燥的空间保存。

[0332] 2-巯基乙磺酸钠和氯化胆碱

[0333] 氯化胆碱和2-巯基乙磺酸钠按1:1摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.1mol 2-巯基乙磺酸钠。将混合物在火焰上缓慢加热并振摇直至已经形成均一的溶液。

[0334] 随后将液体离心并弃去沉淀物。将约2g合成的深共晶溶剂置于封闭的小瓶中并在干燥的空间保存。

[0335] 4-巯基苯乙酸和氯化胆碱

[0336] 氯化胆碱和4-巯基苯乙酸按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol 4-巯基苯乙酸。将混合物在火焰上缓慢加热并振摇。无共晶形成。

[0337] 2-甲基-2-丙硫醇和氯化胆碱

[0338] 氯化胆碱和2-甲基-2-丙硫醇按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol 2-甲基-2-丙硫醇。将混合物在火焰上缓慢加热并振摇。无共晶形成。

[0339] 2-,3-,10-巯基蒎烷和氯化胆碱

[0340] 氯化胆碱和2-,3-,10-巯基蒎烷按1:2摩尔比混合。将氯化胆碱(0.1mol)作为干燥粉末使用并添加至0.2mol 2-,3-,10-巯基蒎烷。将混合物在火焰上缓慢加热并振摇。无共晶形成。

[0341] 实施例4

[0342] 基于巯基的深共晶溶剂中分选酶介导的转酰胺化

[0343] 将分子LCR640-ULPETGGRRRC(与 β 丙氨酸(U)缀合的LCR640荧光团;SEQ ID NO:33)和GGG-生物素在包含10% (v/v) 水的基于巯基的DES(氯化胆碱:二硫苏糖醇摩尔比1:3)中溶解至终浓度2mM。在纯水中,LCR640-ULPETGGRRRC溶解至小于0.1mM的浓度。

[0344] 将这种溶液与包含1mM Sa-SrtA(50mM Tris-HCl pH 7.5,150mM NaCl,10mM CaCl₂)的溶液按体积比19:1混合。将反应混合物在37°C温育5小时。

[0345] 图1中显示反应产物的色谱图。下表中显示化合物的保留时间。

保留时间 [分钟]	化合物
8.433	GGG-生物素
13.350	Sa-SrtA
20.942	LCR640-LPETG
21.987	LCR640-LPETG-生物素

[0347] 21.987分钟保留时间处的峰对应于反应产物。

Leu Pro Glu Thr Gly
 1 5
 <210> 5
 <211> 147
 <212> PRT
 <213> 人工序列
 <220>
 <223> 金黄色葡萄球菌分选酶 A 截短形式
 <400> 5
 Gln Ala Lys Pro Gln Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr
 1 5 10 15
 Ile Glu Ile Pro Asp Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro
 20 25 30
 Ala Thr Pro Glu Gln Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn
 35 40 45
 Glu Ser Leu Asp Asp Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile
 50 55 60
 Asp Arg Pro Asn Tyr Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly
 65 70 75 80
 Ser Met Val Tyr Phe Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met
 85 90 95
 Thr Ser Ile Arg Asp Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu
 [0002] 100 105 110
 Gln Lys Gly Lys Asp Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr
 115 120 125
 Asn Glu Lys Thr Gly Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr
 130 135 140
 Glu Val Lys
 145
 <210> 6
 <211> 150
 <212> PRT
 <213> 人工序列
 <220>
 <223> 单核增生李斯特菌分选酶 A 截短形式
 <400> 6
 Ala Asn Tyr Asp Lys Asp Ala Val Val Gly Ser Ile Ala Val Pro Ser
 1 5 10 15
 Val Asp Val Asn Leu Leu Val Phe Lys Gly Thr Asn Thr Ala Asn Leu
 20 25 30
 Leu Ala Gly Ala Thr Thr Met Arg Ser Asp Gln Val Met Gly Lys Gly
 35 40 45
 Asn Tyr Pro Leu Ala Gly His His Met Arg Asp Glu Ser Met Leu Phe
 50 55 60

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

Lys Tyr Phe Pro Ser Lys
 145 150

<210> 7

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> Arg 标签

<400> 7

Arg Arg Arg Arg Arg
 1 5

[0003]

<210> 8

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> Arg 标签 2

<400> 8

Arg Arg Arg Arg Arg Arg
 1 5

<210> 9

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> His 标签

<400> 9

His His His His His His
 1 5

<210> 10

<211> 19

<212> PRT

<213> 人工序列

<220>

[0005]

<210> 15
 <211> 10
 <212> PRT
 <213> 人工序列
 <220>
 <223> 氨基酸标签
 <400> 15
 Met Asp Val Glu Ala Trp Leu Gly Ala Arg
 1 5 10
 <210> 16
 <211> 16
 <212> PRT
 <213> 人工序列
 <220>
 <223> 氨基酸标签
 <400> 16
 Met Asp Val Glu Ala Trp Leu Gly Ala Arg Val Pro Leu Val Glu Thr
 1 5 10 15
 <210> 17
 <211> 38
 <212> PRT
 <213> 人工序列
 <220>
 <223> 氨基酸标签
 <400> 17
 Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
 1 5 10 15
 Leu Ala Gly Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro
 20 25 30
 Gln Gly Gln Arg Glu Pro
 35
 <210> 18
 <211> 10
 <212> PRT
 <213> 人工序列
 <220>
 <223> 氨基酸标签
 <400> 18
 Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
 1 5 10
 <210> 19
 <211> 15
 <212> PRT
 <213> 人工序列

[0006]

<220>

<223> 氨基酸标签

<400> 19

Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser
 1 5 10 15

<210> 20

<211> 26

<212> PRT

<213> 人工序列

<220>

<223> 氨基酸标签

<400> 20

Lys Arg Arg Trp Lys Lys Asn Phe Ile Ala Val Ser Ala Ala Asn Arg
 1 5 10 15
 Phe Lys Lys Ile Ser Ser Ser Gly Ala Leu
 20 25

<210> 21

<211> 47

<212> PRT

<213> 人工序列

<220>

<223> 氨基酸标签

<400> 21

Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Ser His Tyr
 1 5 10 15
 Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys Ala Ser
 20 25 30
 Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys Leu
 35 40 45

<210> 22

<211> 32

<212> PRT

<213> 溶纤维丁酸弧菌 (*Butyrivibrio fibrisolvens*)

<400> 22

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

<210> 23

<211> 51

<212> PRT

<213> 人工序列

<220>

<223> 壳多糖结合结构域

<400> 23
 Thr Asn Pro Gly Val Ser Ala Trp Gln Val Asn Thr Ala Tyr Thr Ala
 1 5 10 15
 Gly Gln Leu Val Thr Tyr Asn Gly Lys Thr Tyr Lys Cys Leu Gln Pro
 20 25 30
 His Thr Ser Leu Ala Gly Trp Glu Pro Ser Asn Val Pro Ala Leu Trp
 35 40 45
 Gln Leu Gln
 50
 <210> 24
 <211> 209
 <212> PRT
 <213> 皱波角叉菜(*Chondrus crispus*)
 <400> 24
 Met Pro Glu Ile Lys Leu Thr Tyr Phe Asp Met Arg Gly Arg Ala Glu
 1 5 10 15
 Ala Ser Arg Leu Ala Leu Val Val Gly Glu Ile Pro Phe Glu Asp Glu
 20 25 30
 Arg Val Val Phe Asp His Trp Lys Glu Ala Lys Pro Lys Thr Pro Tyr
 35 40 45
 Ala Ala Leu Pro Met Leu Thr Val Asp Gly Met Gln Val Ala Gln Ser
 50 55 60
 Asp Ala Ile Leu Arg Tyr Cys Gly Lys Leu Ala Gly Leu Tyr Pro Ser
 65 70 75 80
 Asp Pro Leu Glu Ala Ala Lys Val Asp Glu Val Gly Gly Val Ile Asp
 85 90 95
 Asp Val Thr His Ala Met Tyr Arg Tyr Arg Gly Asp Asp Lys Asp Lys
 100 105 110
 Leu Arg Glu Glu Arg Asp Lys Phe Ser Lys Val Asp Val Pro Arg Tyr
 115 120 125
 Val Gly Ala Leu Glu Lys Arg Leu Glu Ala Phe Gly Asp Gly Pro Trp
 130 135 140
 Ala Val Gly Gly Asn Met Thr Ile Ala Asp Leu His Ile Cys His Leu
 145 150 155 160
 Val Thr Asn Ile Arg Cys Gly Met Leu Asp Phe Val Asp Lys Asp Leu
 165 170 175
 Leu Glu Gly Tyr Val Arg Ile Val Lys Ser Tyr Ser Ala Val Met Glu
 180 185 190
 His Pro Lys Val Thr Glu Trp Tyr Glu Lys Lys Pro Val Lys Met Phe
 195 200 205
 Ser
 <210> 25
 <211> 396

[0007]

<212> PRT
 <213> 大肠杆菌
 <400> 25
 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
 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

[0008]

	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> 26		
	<211> 206		
	<212> PRT		
	<213> 金黄色葡萄球菌		
	<400> 26		
	Met Lys Lys Trp Thr Asn Arg Leu Met Thr Ile Ala Gly Val Val Leu		
	1 5 10 15		
	Ile Leu Val Ala Ala Tyr Leu Phe Ala Lys Pro His Ile Asp Asn Tyr		
	20 25 30		
	Leu His Asp Lys Asp Lys Asp Glu Lys Ile Glu Gln Tyr Asp Lys Asn		
	35 40 45		
	Val Lys Glu Gln Ala Ser Lys Asp Lys Lys Gln Gln Ala Lys Pro Gln		
	50 55 60		
[0009]	Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr Ile Glu Ile Pro Asp		
	65 70 75 80		
	Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro Ala Thr Pro Glu Gln		
	85 90 95		
	Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn Glu Ser Leu Asp Asp		
	100 105 110		
	Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile Asp Arg Pro Asn Tyr		
	115 120 125		
	Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly Ser Met Val Tyr Phe		
	130 135 140		
	Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met Thr Ser Ile Arg Asp		
	145 150 155 160		
	Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu Gln Lys Gly Lys Asp		
	165 170 175		
	Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr Asn Glu Lys Thr Gly		
	180 185 190		
	Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr Glu Val Lys		
	195 200 205		
	<210> 27		
	<211> 222		
	<212> PRT		
	<213> 单核增生李斯特菌		

<400> 27
 Met Leu Lys Lys Thr Ile Ala Ala Ala Ala Leu Ala Ala Gly Leu Leu
 1 5 10 15
 Leu Ile Phe Ser Pro Phe Ile Lys Asn Gly Ile Val Lys Tyr Met Ser
 20 25 30
 Gly His Glu Thr Ile Glu Gln Tyr Lys Ala Ser Asp Ile Lys Lys Asn
 35 40 45
 Asn Glu Lys Asp Ala Thr Phe Asp Phe Glu Ser Val Gln Leu Pro Ser
 50 55 60
 Met Thr Ser Val Ile Lys Gly Ala Ala Asn Tyr Asp Lys Asp Ala Val
 65 70 75 80
 Val Gly Ser Ile Ala Val Pro Ser Val Asp Val Asn Leu Leu Val Phe
 85 90 95
 Lys Gly Thr Asn Thr Ala Asn Leu Leu Ala Gly Ala Thr Thr Met Arg
 100 105 110
 Ser Asp Gln Val Met Gly Lys Gly Asn Tyr Pro Leu Ala Gly His His
 115 120 125
 Met Arg Asp Glu Ser Met Leu Phe Gly Pro Ile Met Lys Val Lys Lys
 130 135 140
 Gly Asp Lys Ile Tyr Leu Thr Asp Leu Glu Asn Leu Tyr Glu Tyr Thr
 145 150 155 160
 Val Thr Glu Thr Lys Thr Ile Asp Glu Thr Glu Val Ser Val Ile Asp
 [0010] 165 170 175
 Asn Thr Lys Asp Ala Arg Ile Thr Leu Ile Thr Cys Asp Lys Pro Thr
 180 185 190
 Glu Thr Thr Lys Arg Phe Val Ala Val Gly Glu Leu Glu Lys Thr Glu
 195 200 205
 Lys Leu Thr Lys Glu Leu Glu Asn Lys Tyr Phe Pro Ser Lys
 210 215 220
 <210> 28
 <211> 2
 <212> PRT
 <213> 人工序列
 <220>
 <223> 寡甘氨酸
 <400> 28
 Gly Gly
 1
 <210> 29
 <211> 3
 <212> PRT
 <213> 人工序列
 <220>
 <223> 寡甘氨酸

<400> 29
 Gly Gly Gly
 1
 <210> 30
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 寡甘氨酸
 <400> 30
 Gly Gly Gly Gly
 1
 <210> 31
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 寡甘氨酸
 <400> 31
 Gly Gly Gly Gly Gly
 1 5
 [0011] <210> 32
 <211> 12
 <212> PRT
 <213> 人工序列
 <220>
 <223> 具有接头的六组氨酸标签
 <400> 32
 Met Arg Gly Ser His His His His His His Gly Ser
 1 5 10
 <210> 33
 <211> 11
 <212> PRT
 <213> 人工序列
 <220>
 <223> LCR640-ULPETGGGRRRC
 <220>
 <221> MISC_FEATURE
 <222> (1)..(1)
 <223> 与 LCR640-U 缀合的 β -丙氨酸
 <400> 33
 Xaa Leu Pro Glu Thr Gly Gly Gly Arg Arg Cys
 1 5 10
 <210> 34

<211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了P之外的任何氨基酸残基
 <400> 34
 Leu Pro Xaa Thr Gly
 1 5
 <210> 35
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了P之外的任何氨基酸残基
 <400> 35
 Leu Pro Xaa Thr
 1
 <210> 36
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 36
 Leu Pro Thr Thr Gly
 1 5
 <210> 37
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 37
 Leu Pro Asp Thr Gly
 1 5
 <210> 38

[0012]

<211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 38
 Leu Pro Lys Thr Gly
 1 5
 <210> 39
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 39
 Leu Pro Asn Thr Gly
 1 5
 <210> 40
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 [0013] <223> 分选酶识别基序
 <400> 40
 Pro Leu Ala Ala Gly
 1 5
 <210> 41
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 41
 Leu Pro Lys Ala Gly
 1 5
 <210> 42
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 42
 Leu Pro Gln Thr Gly
 1 5

[0014]

<210> 43
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 43
Leu Pro Met Thr Gly
1 5
<210> 44
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 44
Leu Pro Ile Thr Gly
1 5
<210> 45
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 45
Leu Pro Asp Thr Ala
1 5
<210> 46
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 46
Ser Pro Lys Thr Gly
1 5
<210> 47
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 47
Leu Ala Glu Thr Gly

[0015]

1 5
 <210> 48
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 48
 Leu Ala Ala Thr Gly
 1 5
 <210> 49
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 49
 Leu Ala His Thr Gly
 1 5
 <210> 50
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 50
 Lcu Ala Ser Thr Gly
 1 5
 <210> 51
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 51
 Leu Pro Leu Thr Gly
 1 5
 <210> 52
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 52

[0016]

Leu Ser Arg Thr Gly
 1 5
 <210> 53
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 53
 Val Pro Asp Thr Gly
 1 5
 <210> 54
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 54
 Ile Pro Gln Thr Gly
 1 5
 <210> 55
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 55
 Tyr Pro Arg Thr Gly
 1 5
 <210> 56
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 56
 Leu Ala Leu Thr Gly
 1 5
 <210> 57
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序

<400> 57
 Leu Gly Asn Thr Gly
 1 5
 <210> 58
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 58
 Leu Pro Gln Thr Ser
 1 5
 <210> 59
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X-除了P之外的任何氨基酸残基
 [0017] <400> 59
 Thr Leu Xaa Thr Cys
 1 5
 <210> 60
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 60
 Asn Pro Gln Thr Asn
 1 5
 <210> 61
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (1)..(1)
 <223> X-L 或 M

[0018]

<220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <220>
 <221> MISC_FEATURE
 <222> (4)..(4)
 <223> X=A 或 L 或 S 或 T 或 V
 <400> 61
 Xaa Pro Xaa Xaa Gly
 1 5
 <210> 62
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 62
 Ile Pro Lys Thr Gly
 1 5
 <210> 63
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 63
 Ile Pro Ala Leu Gly
 1 5
 <210> 64
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 64
 Leu Ala Ala Ser Ser
 1 5
 <210> 65
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序

<400> 65
 Leu Ala Pro Thr Gly
 1 5
 <210> 66
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 66
 Leu Pro Ile Ser Ser
 1 5
 <210> 67
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 67
 Leu Pro Lys Thr Ser
 1 5
 [0019] <210> 68
 <211> 6
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 68
 Tyr Ala Leu Pro Glu Thr
 1 5
 <210> 69
 <211> 8
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 69
 Tyr Ala Leu Pro Met Thr Gly Lys
 1 5
 <210> 70
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>

[0020]

<223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 70
 Leu Pro Xaa Ala
 1
 <210> 71
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 71
 Ser Pro Xaa Thr
 1
 <210> 72
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 72
 Leu Ala Xaa Thr
 1
 <210> 73
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基

<400> 73
Leu Ser Xaa Thr
1
<210> 74
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X=除了P之外的任何氨基酸残基
<400> 74
Asn Pro Xaa Thr
1
<210> 75
<211> 4
<212> PRT
<213> 人工序列
<220>
[0021] <223> 分选酶识别基序
<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X=除了P之外的任何氨基酸残基
<400> 75
Val Pro Xaa Thr
1
<210> 76
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X=除了P之外的任何氨基酸残基
<400> 76
Ile Pro Xaa Thr
1
<210> 77
<211> 4

<212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 77
 Leu Gly Xaa Thr
 I
 <210> 78
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 78
 [0022] Tyr Pro Xaa Arg
 I
 <210> 79
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 79
 Leu Pro Ser Thr
 I
 <210> 80
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 80
 Leu Pro Lys Thr
 I
 <210> 81
 <211> 4

<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 81
Leu Pro Ile Thr
1
<210> 82
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 82
Leu Pro Asp Thr
1
<210> 83
<211> 4
<212> PRT
<213> 人工序列
<220>
[0023] <223> 分选酶识别基序
<400> 83
Ser Pro Lys Thr
1
<210> 84
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 84
Leu Ala Glu Thr
1
<210> 85
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 85
Leu Ala Ala Thr
1
<210> 86

<211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 86
 Leu Ala Ser Thr
 1
 <210> 87
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 87
 Leu Pro Leu Thr
 1
 <210> 88
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 [0024] <223> 分选酶识别基序
 <400> 88
 Leu Ser Arg Thr
 1
 <210> 89
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 89
 Leu Pro Glu Thr
 1
 <210> 90
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 90
 Val Pro Asp Thr
 1

[0025]

<210> 91
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 91
Ile Pro Gln Thr
1
<210> 92
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 92
Tyr Pro Arg Arg
1
<210> 93
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 93
Leu Pro Met Thr
1
<210> 94
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 94
Leu Ala Phe Thr
1
<210> 95
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 分选酶识别基序
<400> 95
Leu Pro Gln Thr

I
 <210> 96
 <211> 4
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (1)..(1)
 <223> X= L 或 I 或 V 或 M
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了P之外的任何氨基酸残基
 <220>
 <221> MISC_FEATURE
 <222> (4)..(4)
 <223> X=T 或 S 或 A
 <400> 96
 Xaa Pro Xaa Xaa
 [0026]
 I
 <210> 97
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 97
 Tyr Pro Pro Arg Gly
 I 5
 <210> 98
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 98
 Leu Ala Phe Thr Gly
 I 5
 <210> 99
 <211> 5
 <212> PRT
 <213> 人工序列

[0027]

<220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 99
 Leu Pro Xaa Ala Gly
 1 5
 <210> 100
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 100
 Leu Pro Asn Ala Gly
 1 5
 <210> 101
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了 P 之外的任何氨基酸残基
 <400> 101
 Leu Pro Xaa Thr Ala
 1 5
 <210> 102
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 102
 Leu Pro Asn Thr Ala
 1 5
 <210> 103
 <211> 5
 <212> PRT
 <213> 人工序列

<220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了P之外的任何氨基酸残基
 <400> 103
 Leu Gly Xaa Thr Gly
 1 5
 <210> 104
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 104
 Leu Gly Ala Thr Gly
 1 5
 <210> 105
 <211> 5
 <212> PRT
 <213> 人工序列
 [0028] <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=除了P之外的任何氨基酸残基
 <400> 105
 Ile Pro Xaa Thr Gly
 1 5
 <210> 106
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 106
 Ile Pro Asn Thr Gly
 1 5
 <210> 107
 <211> 5
 <212> PRT
 <213> 人工序列

<220>
 <223> 分选酶识别基序
 <400> 107
 Ile Pro Glu Thr Gly
 1 5
 <210> 108
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <400> 108
 Leu Pro Glu Thr Ala
 1 5
 <210> 109
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 [0029] <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=D, E, A, N, Q, K, R
 <220>
 <221> MISC_FEATURE
 <222> (5)..(5)
 <223> X=A, G
 <400> 109
 Leu Pro Xaa Thr Xaa
 1 5
 <210> 110
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=D, E, A, N, Q, K, R
 <400> 110
 Leu Pro Xaa Thr Ala
 1 5

<210> 111
 <211> 5
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶识别基序
 <220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> X=D, E, A, N, Q, K, R
 <400> 111
 Leu Pro Xaa Thr Gly
 1 5
 <210> 112
 <211> 85
 <212> PRT
 <213> 天蓝色链霉菌(*Streptomyces coelicolor*)
 <400> 112
 Val Ala Gly His Val Asp Asn Ala Glu Gly Pro Ala Val Phe Tyr Arg
 1 5 10 15
 Leu Gly Ala Leu Glu Lys Gly Ser Ala Ile Glu Ile Asp Arg Arg Asp
 20 25 30
 Gly Gly Val Ala Val Phe Thr Val Asp Ala Val Glu Val Tyr Ala Ala
 35 40 45
 Asp Ala Phe Pro Asp Glu Lys Val Tyr Gly Ala Ala Asp Arg Pro Glu
 50 55 60
 Leu Arg Val Ile Thr Cys Gly Gly Pro Tyr Ser Arg Ser Thr Gly Tyr
 65 70 75 80
 Gln Gly Asn Val Val
 85
 <210> 113
 <211> 85
 <212> PRT
 <213> 天蓝色链霉菌
 <400> 113
 Val Val Gly His Val Asp Asn Gln Gln Gly Pro Ala Val Phe Tyr Gly
 1 5 10 15
 Leu Gly Ala Leu Lys Lys Gly Asn Lys Val Glu Val His Arg Gln Asp
 20 25 30
 Gly Lys Thr Ala Val Phe Glu Ile Tyr Gly Ile Glu Val Phe Glu Lys
 35 40 45
 Asn Asn Phe Pro Gly Asp Arg Val Tyr Gly Ser Lys Gly Ser Pro Glu
 50 55 60
 Leu Arg Val Ile Thr Cys Gly Gly Gly Phe Thr Lys Gln Asn Gly Tyr

	65	70	75	80
	Asp Gly Asn Val Val			
		85		
	<210> 114			
	<211> 86			
	<212> PRT			
	<213> 天蓝色链霉菌			
	<400> 114			
	Ile Ala Gly His Val Asp Thr Lys Thr Ser Ala Ala Val Phe Ala Arg			
	1	5	10	15
	Leu Asp Gln Leu Asp Lys Gly Asp Lys Phe Gln Val Arg Arg Ala Asp			
		20	25	30
	Gly Arg Ser Ala Thr Phe Val Val Asp Gly Leu Glu Thr Phe Ala Lys			
		35	40	45
	Asp Glu Phe Pro Ser Asp Arg Val Tyr Gly Asp Ala Asp Arg Pro Glu			
		50	55	60
	Val Arg Leu Ile Thr Cys Ala Gly Asp Tyr Asp His Lys Val Lys Asp			
	65	70	75	80
	Tyr Thr Asp Asn Leu Val			
		85		
	<210> 115			
	<211> 88			
	<212> PRT			
	<213> 天蓝色链霉菌			
	<400> 115			
	Met Val Gly His Val Asp Thr Glu Thr Arg Pro Ala Val Phe Tyr Gln			
	1	5	10	15
	Leu Ser Thr Leu Glu Pro Gly Gln Thr Ile Arg Val Ala Arg Asp Asp			
		20	25	30
	Asp Glu Val Ala Glu Phe Thr Val Asp Asp Val Gln Val Leu Thr Arg			
		35	40	45
	Asp Gly Phe Asp Ala Gln Gln Ala Tyr Gly Pro Arg Asp Thr Gly Arg			
		50	55	60
	Ser Glu Leu Arg Leu Ile Thr Cys Gly Gly Thr Phe Asp Gln Thr Thr			
	65	70	75	80
	Asp Ser Tyr Thr Ala Asn Val Val			
		85		
	<210> 116			
	<211> 78			
	<212> PRT			
	<213> 耐盐芽孢杆菌(Bacillus halodurans)			
	<400> 116			
	Leu Ser Gly His Arg Asp Thr Val Phe Arg Asp Met Gly Lys Leu Glu			
	1	5	10	15

[0031]

Ile Gly Asp Asp Leu Thr Val His Met Pro Tyr Gly Ser Tyr Thr Tyr
 20 25 30
 Arg Ile Val Asp Thr Glu Ile Val Asp Ala Asn Asp Thr Ser Val Ile
 35 40 45
 Arg Ser Thr Ala Pro Asp Glu Val Leu Thr Leu Ser Thr Cys Tyr Pro
 50 55 60
 Phe Asn Phe Ile Gly Ser Ala Pro Glu Arg Tyr Ile Ile Tyr
 65 70 75

<210> 117

<211> 78

<212> PRT

<213> 枯草芽孢杆菌 (Bacillus subtilis)

<400> 117

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

[0032]

<210> 118

<211> 77

<212> PRT

<213> 耐盐芽孢杆菌

<400> 118

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

<210> 119

<211> 77

<212> PRT

<213> 炭疽芽孢杆菌 (Bacillus anthracis)

<400> 119

Leu Ser Gly His Arg Asp Thr Val Phe Thr Asp Leu Gly Gln Leu Lys

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

<210> 120
 <211> 81
 <212> PRT
 <213> 耐盐芽孢杆菌
 <400> 120

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

[0033]

Met

<210> 121
 <211> 81
 <212> PRT
 <213> 丙酮丁醇梭菌 (Clostridium acetobutylicum)
 <400> 121

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

<210> 122

<211> 89
 <212> PRT
 <213> 天蓝色链霉菌
 <400> 122
 Leu Ala Gly His Arg Asn Thr His Gly Glu Pro Phe Arg Tyr Ile Asn
 1 5 10 15
 Lys Leu Glu Pro Gly Asp Pro Ile Val Val Glu Thr Gln Asp Lys Tyr
 20 25 30
 Phe Val Tyr Lys Met Ala Ser Ile Leu Pro Val Thr Ser Pro Ser Asn
 35 40 45
 Val Ser Val Leu Asp Pro Val Pro Lys Gln Ser Gly Phe Lys Gly Pro
 50 55 60
 Gly Arg Tyr Ile Thr Leu Thr Thr Cys Thr Pro Glu Phe Thr Ser Lys
 65 70 75 80
 Tyr Arg Met Ile Val Asn Gly Lys Met
 85

<210> 123
 <211> 89
 <212> PRT
 <213> 天蓝色链霉菌
 <400> 123

[0034]

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

<210> 124
 <211> 81
 <212> PRT
 <213> 内氏放线菌 (Actinomyces naeslundii)
 <400> 124

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

Glu Ala Leu Arg Val Glu Glu Gly Lys Asp Leu Leu Thr Leu Val Thr
 50 55 60
 Cys Ile Pro Leu Gly Ile Asn Thr His Arg Ile Leu Leu Thr Gly Glu
 65 70 75 80
 Arg

<210> 125
 <211> 81
 <212> PRT
 <213> 白喉棒杆菌 (Corynebacterium diphtheriae)
 <400> 125

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

[0035]

<210> 126
 <211> 82
 <212> PRT
 <213> 肺炎链球菌 (Streptococcus pneumoniae)
 <400> 126

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

<210> 127
 <211> 82
 <212> PRT
 <213> 肺炎链球菌

<400> 127
 Ile Thr His Ala Arg Gly Leu Pro Thr Ala Glu Leu Phe Ser Gln Leu
 1 5 10 15
 Asp Lys Met Lys Lys Gly Asp Ile Phe Tyr Leu His Val Leu Asp Gln
 20 25 30
 Val Leu Ala Tyr Gln Val Asp Gln Ile Val Thr Val Glu Pro Asn Asp
 35 40 45
 Phe Glu Val Leu Ile Gln His Gly Glu Asp Ala Tyr Ala Thr Leu Leu
 50 55 60
 Thr Cys Thr Pro Tyr Met Ile Asn Ser His Arg Leu Leu Val Arg Gly
 65 70 75 80
 Lys Arg

<210> 128

<211> 83

<212> PRT

<213> 粪肠球菌 (Enterococcus faecalis)

<400> 128

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

[0036]

<210> 129

<211> 80

<212> PRT

<213> 马肠链球菌 (Streptococcus equinus)

<400> 129

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

Thr Glu Asp Leu Arg Pro Glu Gln Gly Lys Asp Tyr Ile Leu Leu Ile
 50 55 60
 Thr Cys Thr Pro Tyr Gly Ile Asn Thr His Arg Leu Met Val Arg Gly
 65 70 75 80
 His Gln

<210> 133

<211> 80

<212> PRT

<213> 白喉棒杆菌

<400> 133

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

[0038]

<210> 134

<211> 81

<212> PRT

<213> 白喉棒杆菌

<400> 134

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

<210> 135

<211> 74

<212> PRT

<213> 耐盐芽孢杆菌

<400> 135

Ile Ala Gly His Arg Gly Tyr Arg Gly Asn Arg His Phe Ser Arg Leu

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

65 70

<210> 136

<211> 80

<212> PRT

<213> 耐盐芽孢杆菌

<400> 136

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

[0039]

Thr Cys Thr Arg Ser Gly Lys Gln Arg Val Ala Val Arg Gly Val Leu
 65 70 75 80

<210> 137

<211> 85

<212> PRT

<213> 变异链球菌 (Streptococcus mutans)

<400> 137

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

Thr Leu Val Thr Cys Thr Asp Ala Gly Ala Thr Ala Arg Thr Ile Val
 65 70 75 80

His Gly Thr Tyr Lys

85

<210> 138

<211> 88

<212> PRT

Thr Cys Gly Asp Leu Gln Ala Thr Thr Arg Ile Ala Val Gln Gly Thr
65 70 75 80

Leu Ala

<210> 141

<211> 90

<212> PRT

<213> 耐盐芽孢杆菌

<400> 141

Asp His His Glu Gly Phe Tyr Tyr Asp Thr Leu Tyr Asn Arg Tyr Asp
1 5 10 15

Val Glu Val Phe Ser Ala Tyr Val Thr Thr Thr Asp Phe Tyr Tyr Ile
20 25 30

Glu Thr Glu Phe Pro Ser Lys Asp Asp Tyr Lys Ala Phe Leu Asn Glu
35 40 45

Leu Lys Lys Arg Ser Val Val Gln Thr Asn Val Glu Val Gly Glu Glu
50 55 60

Asp Gln Ile Ile Thr Leu Ser Thr Cys Asp Tyr Arg Leu Asp Arg Asp
65 70 75 80

Arg Gly Arg Leu Val Val His Gly Lys Leu
85 90

[0041]

<210> 142

<211> 91

<212> PRT

<213> 炭疽芽孢杆菌

<400> 142

Phe Met Ser His Arg Lys Leu Tyr Tyr Asp Thr Leu Phe Glu Gly Tyr
1 5 10 15

Asp Leu Glu Val Phe Ser Val Tyr Thr Thr Thr Thr Asp Phe Tyr Tyr
20 25 30

Ile Glu Thr Asp Phe Ser Ser Asp Thr Glu Tyr Thr Ser Phe Leu Glu
35 40 45

Lys Ile Gln Glu Lys Ser Leu Tyr Lys Thr Asp Thr Thr Val Thr Ala
50 55 60

Gly Asp Gln Ile Val Thr Leu Ser Thr Cys Asp Tyr Ala Leu Asp Pro
65 70 75 80

Glu Ala Gly Arg Leu Val Val His Ala Lys Leu
85 90

<210> 143

<211> 91

<212> PRT

<213> 金黄色葡萄球菌

<400> 143

Tyr Glu Lys His Lys Ile Ile Glu Pro Asp Asn Lys Tyr Gly Lys Tyr

Ala Asp Val Val Glu Glu Thr Asn Lys Lys Glu Ile Thr Leu Ile Thr
 50 55 60
 Cys Asp Lys Ala Val Lys Thr Glu Gly Arg Leu Val Val Lys Gly Glu
 65 70 75 80
 Leu

<210> 149

<211> 81

<212> PRT

<213> 粪肠球菌

<400> 149

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

[0044]

<210> 150

<211> 84

<212> PRT

<213> 金黄色葡萄球菌

<400> 150

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

<210> 151

<211> 79

<212> PRT

<213> 腐败交替单胞菌 (Shewanella putrefaciens)

<400> 151
 Ile Ala Gly His Arg Asp Thr His Phe Ala Ile Leu Lys Gly Met Thr
 1 5 10 15
 Val Gly Arg Arg Leu Ala Leu Gln Thr Ala Ala Gly Lys Glu Ile Val
 20 25 30
 Tyr Gln Val Val Ala Thr Lys Glu Ser Gln Thr Glu Leu Met Ala Pro
 35 40 45
 Ser Asp Asp Asn Arg Leu Thr Leu Ile Thr Cys Tyr Pro Pro Asp Ala
 50 55 60
 Leu Gln Gly Val Ala Glu Leu Arg Phe Val Val Gln Ala Val Pro
 65 70 75

<210> 152

<211> 91

<212> PRT

<213> 天蓝色链霉菌

<400> 152

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

[0045]

<210> 153

<211> 90

<212> PRT

<213> 艰难梭菌 (Clostridium difficile)

<400> 153

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

	85	90
<210>	154	
<211>	86	
<212>	PRT	
<213>	热自养甲烷热细菌 (Methanothermobacter thermautotrophicus)	
<400>	154	
Ile Leu Gly His Arg Thr Thr Tyr Ser Gly Pro Phe Arg Lys Ile Gly		
1	5	10
Ala Leu Arg Lys Gly Asp Arg Val Ile Ile Glu Asp Ala Ser Ser Ser		
	20	25
Ile Arg Tyr Ile Tyr Thr Val Thr Ser Asn Gly Asp Asp Ile Arg Trp		
	35	40
Asp Tyr Arg Thr Asn Pro Val Arg Phe Ser Gln Ser Gly Asp Ala Arg		
	50	55
Leu Met Leu Ile Thr Cys Tyr Pro Pro Gly Gln Lys Lys Ala Ala Trp		
65	70	75
Ile Thr His Cys Lys Leu		
	85	
<210>	155	
<211>	147	
<212>	PRT	
<213>	人工序列	
<220>		
<223>	金黄色葡萄球菌分选酶A截短形式 N(1-59)	
<400>	155	
Gln Ala Lys Pro Gln Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr		
1	5	10
Ile Glu Ile Pro Asp Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro		
	20	25
Ala Thr Pro Glu Gln Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn		
	35	40
Glu Ser Leu Asp Asp Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile		
	50	55
Asp Arg Pro Asn Tyr Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly		
65	70	75
Ser Met Val Tyr Phe Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met		
	85	90
Thr Ser Ile Arg Asp Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu		
	100	105
Gln Lys Gly Lys Asp Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr		
	115	120
Asn Glu Lys Thr Gly Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr		
	130	135
Glu Val Lys		140

[0046]

Asn Thr Lys Asp Ala Arg Ile Thr Leu Ile Thr Cys Asp Lys Pro Thr
 180 185 190
 Glu Thr Thr Lys Arg Phe Val Ala Val Gly Glu Leu Glu Lys Thr Glu
 195 200 205
 Lys Leu Thr Lys Glu Leu Glu Asn Lys Tyr Phe Pro Ser Lys
 210 215 220
 <210> 159
 <211> 249
 <212> PRT
 <213> 酿脓链球菌
 <400> 159
 Met Val Lys Lys Gln Lys Arg Arg Lys Ile Lys Ser Met Ser Trp Ala
 I 5 10 15
 Arg Lys Leu Leu Ile Ala Val Leu Leu Ile Leu Gly Leu Ala Leu Leu
 20 25 30
 Phe Asn Lys Pro Ile Arg Asn Thr Leu Ile Ala Arg Asn Ser Asn Lys
 35 40 45
 Tyr Gln Val Thr Lys Val Ser Lys Lys Gln Ile Lys Lys Asn Lys Glu
 50 55 60
 Ala Lys Ser Thr Phe Asp Phe Gln Ala Val Glu Pro Val Ser Thr Glu
 65 70 75 80
 [0049] Ser Val Leu Gln Ala Gln Met Ala Ala Gln Gln Leu Pro Val Ile Gly
 85 90 95
 Gly Ile Ala Ile Pro Glu Leu Gly Ile Asn Leu Pro Ile Phe Lys Gly
 100 105 110
 Leu Gly Asn Thr Glu Leu Ile Tyr Gly Ala Gly Thr Met Lys Glu Glu
 115 120 125
 Gln Val Met Gly Gly Glu Asn Asn Tyr Ser Leu Ala Ser His His Ile
 130 135 140
 Phe Gly Ile Thr Gly Ser Ser Gln Met Leu Phe Ser Pro Leu Glu Arg
 145 150 155 160
 Ala Gln Asn Gly Met Ser Ile Tyr Leu Thr Asp Lys Glu Lys Ile Tyr
 165 170 175
 Glu Tyr Ile Ile Lys Asp Val Phe Thr Val Ala Pro Glu Arg Val Asp
 180 185 190
 Val Ile Asp Asp Thr Ala Gly Leu Lys Glu Val Thr Leu Val Thr Cys
 195 200 205
 Thr Asp Ile Glu Ala Thr Glu Arg Ile Ile Val Lys Gly Glu Leu Lys
 210 215 220
 Thr Glu Tyr Asp Phe Asp Lys Ala Pro Ala Asp Val Leu Lys Ala Phe
 225 230 235 240
 Asn His Ser Tyr Asn Gln Val Ser Thr
 245
 <210> 160

<211> 178
 <212> PRT
 <213> 人工序列
 <220>
 <223> 金黄色葡萄球菌分选酶 A 截短形式 N(2-29)
 <400> 160
 Met Asp Asn Tyr Leu His Asp Lys Asp Lys Asp Glu Lys Ile Glu Gln
 1 5 10 15
 Tyr Asp Lys Asn Val Lys Glu Gln Ala Ser Lys Asp Lys Lys Gln Gln
 20 25 30
 Ala Lys Pro Gln Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr Ile
 35 40 45
 Glu Ile Pro Asp Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro Ala
 50 55 60
 Thr Pro Glu Gln Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn Glu
 65 70 75 80
 Ser Leu Asp Asp Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile Asp
 85 90 95
 Arg Pro Asn Tyr Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly Ser
 100 105 110
 Met Val Tyr Phe Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met Thr
 115 120 125
 Ser Ile Arg Asp Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu Gln
 130 135 140
 Lys Gly Lys Asp Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr Asn
 145 150 155 160
 Glu Lys Thr Gly Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr Glu
 165 170 175
 Val Lys

[0050]

<210> 161
 <211> 222
 <212> PRT
 <213> 单核增生李斯特菌
 <400> 161
 Met Leu Lys Lys Thr Ile Ala Ala Ala Ala Leu Ala Ala Gly Leu Leu
 1 5 10 15
 Leu Ile Phe Ser Pro Phe Ile Lys Asn Gly Ile Val Lys Tyr Met Ser
 20 25 30
 Gly His Glu Thr Ile Glu Gln Tyr Lys Ala Ser Asp Ile Lys Lys Asn
 35 40 45
 Asn Glu Lys Asp Ala Thr Phe Asp Phe Glu Ser Val Gln Leu Pro Ser
 50 55 60
 Met Thr Ser Val Ile Lys Gly Ala Ala Asn Tyr Asp Lys Asp Ala Val

[0052]

Cys Ala Ala
 1
 <210> 165
 <211> 3
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶标签
 <400> 165
 Lys Gly Gly
 1
 <210> 166
 <211> 3
 <212> PRT
 <213> 人工序列
 <220>
 <223> 分选酶标签
 <400> 166
 Lys Ala Ala
 1
 <210> 167
 <211> 206
 <212> PRT
 <213> 金黄色葡萄球菌
 <400> 167
 Met Lys Lys Trp Thr Asn Arg Leu Met Thr Ile Ala Gly Val Val Leu
 1 5 10 15
 Ile Leu Val Ala Ala Tyr Leu Phe Ala Lys Pro His Ile Asp Asn Tyr
 20 25 30
 Leu His Asp Lys Asp Lys Asp Glu Lys Ile Glu Gln Tyr Asp Lys Asn
 35 40 45
 Val Lys Glu Gln Ala Ser Lys Asp Lys Lys Gln Gln Ala Lys Pro Gln
 50 55 60
 Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr Ile Glu Ile Pro Asp
 65 70 75 80
 Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro Ala Thr Pro Glu Gln
 85 90 95
 Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn Glu Ser Leu Asp Asp
 100 105 110
 Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile Asp Arg Pro Asn Tyr
 115 120 125
 Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly Ser Met Val Tyr Phe
 130 135 140
 Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met Thr Ser Ile Arg Asp

145 150 155 160
 Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu Gln Lys Gly Lys Asp
 165 170 175
 Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr Asn Glu Lys Thr Gly
 180 185 190
 Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr Glu Val Lys
 195 200 205
 <210> 168
 <211> 222
 <212> PRT
 <213> 单核增生李斯特菌
 <400> 168
 Met Leu Lys Lys Thr Ile Ala Ala Ala Ala Leu Ala Ala Gly Leu Leu
 1 5 10 15
 Leu Ile Phe Ser Pro Phe Ile Lys Asn Gly Ile Val Lys Tyr Met Ser
 20 25 30
 Gly His Glu Thr Ile Glu Gln Tyr Lys Ala Ser Asp Ile Lys Lys Asn
 35 40 45
 Asn Glu Lys Asp Ala Thr Phe Asp Phe Glu Ser Val Gln Leu Pro Ser
 50 55 60
 [0053] Met Thr Ser Val Ile Lys Gly Ala Ala Asn Tyr Asp Lys Asp Ala Val
 65 70 75 80
 Val Gly Ser Ile Ala Val Pro Ser Val Asp Val Asn Leu Leu Val Phe
 85 90 95
 Lys Gly Thr Asn Thr Ala Asn Leu Leu Ala Gly Ala Thr Thr Met Arg
 100 105 110
 Ser Asp Gln Val Met Gly Lys Gly Asn Tyr Pro Leu Ala Gly His His
 115 120 125
 Met Arg Asp Glu Ser Met Leu Phe Gly Pro Ile Met Lys Val Lys Lys
 130 135 140
 Gly Asp Lys Ile Tyr Leu Thr Asp Leu Glu Asn Leu Tyr Glu Tyr Thr
 145 150 155 160
 Val Thr Glu Thr Lys Thr Ile Asp Glu Thr Glu Val Ser Val Ile Asp
 165 170 175
 Asn Thr Lys Asp Ala Arg Ile Thr Leu Ile Thr Cys Asp Lys Pro Thr
 180 185 190
 Glu Thr Thr Lys Arg Phe Val Ala Val Gly Glu Leu Glu Lys Thr Glu
 195 200 205
 Lys Leu Thr Lys Glu Leu Glu Asn Lys Tyr Phe Pro Ser Lys
 210 215 220

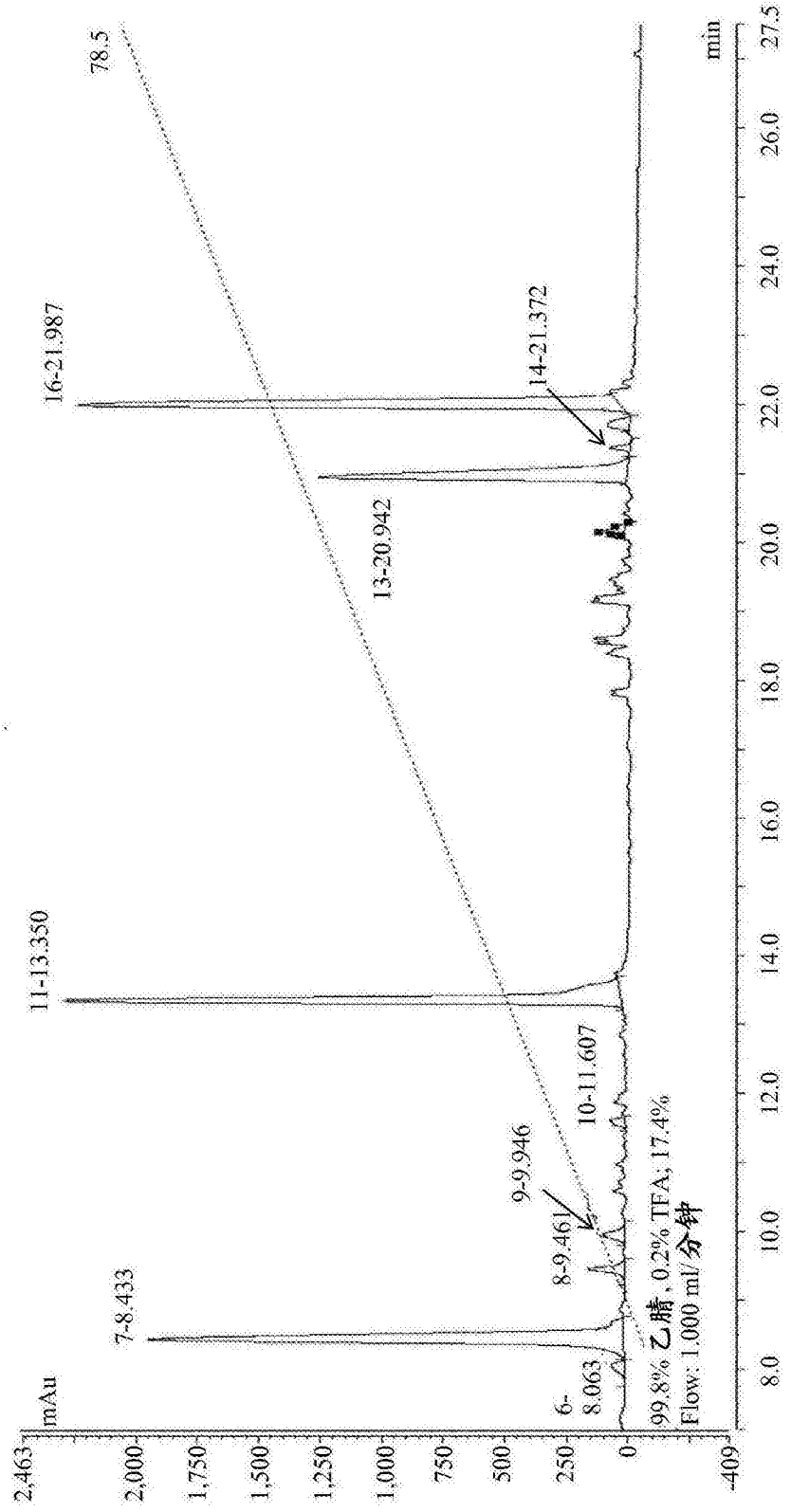


图1