(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114302890 A
(43)申请公布日 2022.04.08

(21)申请号 202080060466.4

(22)申请日 2020.07.21

(30)优先权数据
19187963.4 2019.07.23 EP

(85)PCT国际申请进入国家阶段日
2022.02.25

(86)PCT国际申请的申请数据
PCT/EP2020/070595 2020.07.21

(87)PCT国际申请的公布数据
WO2021/013854 EN 2021.01.28

(71)申请人 葛兰素史克生物有限公司
地址 比利时里克桑萨尔

(72)发明人 M•比亚吉尼 L•E•福塔纳
A•加利亚 N•诺赖斯
M•斯卡谢利

(74)专利代理机构 北京市金杜律师事务所 11256
代理人 邰红 徐志明

(51)Int.Cl.
C07K 14/31(2006.01)
C07K 14/33(2006.01)
C07K 14/34(2006.01)
C07K 14/245(2006.01)
C07K 14/22(2006.01)
C07K 14/21(2006.01)
C12N 15/31(2006.01)
C12N 1/21(2006.01)
C12N 15/70(2006.01)
C12N 15/54(2006.01)
A61K 39/02(2006.01)
A61K 39/085(2006.01)
A61K 39/09(2006.01)
A61K 39/095(2006.01)
A61K 39/102(2006.01)
A61K 39/385(2006.01)
A61P 31/04(2006.01)
G01N 33/68(2006.01)

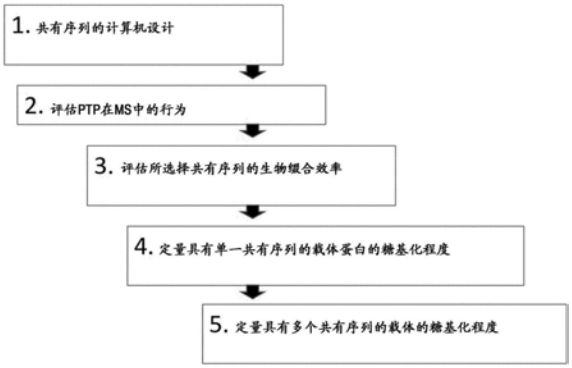
权利要求书4页 说明书58页
序列表74页 附图5页

(54)发明名称

生物缀合物糖基化的定量

(57)摘要

本发明提供了用于表征生物缀合物,特别是用于测量糖基化水平的分析工具,和用于绝对定量糖基化序列的方法,以及用于这类方法的序列和糖基化位点。



CN 114302890 A

1.一种经修饰的载体蛋白,其修饰以使得包含一个或多个共有序列,所述共有序列包含或由以下氨基酸序列组成:

K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R

其中X和Y独立地是除了脯氨酸之外的任何氨基酸,并且Z表示任何氨基酸;其中任选地X和Y独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸,Z表示除了赖氨酸或精氨酸之外的任何氨基酸,并且优选地Z表示除了半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸。

2.如权利要求1所述的经修饰的载体蛋白,其中所述共有序列是氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19),其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸,并且其中X_1和X_2以及Z_1和Z_2优选不是赖氨酸或精氨酸,其中任选地,其中X_1和X_2以及Z_1和Z_2不是半胱氨酸、天冬酰胺、谷氨酰胺、甲硫氨酸或精氨酸。

3.如权利要求1或2所述的经修饰的载体蛋白,其中所述共有序列包含或由SEQ ID NO:20,任选地SEQ ID No:42-45,任选地SEQ ID No:42-44中的任一个的氨基酸序列组成。

4.如权利要求1-3中任一项所述的经修饰的载体蛋白,其中所述共有序列(i)已经取代了所述载体蛋白序列的一个或多个氨基酸,或(ii)已经插入到所述载体蛋白序列中。

5.如权利要求1-4中任一项所述的经修饰的载体蛋白,其包含多于一个所述共有序列,任选地至少2、3、4或5个共有序列。

6.如权利要求5所述的经修饰的载体蛋白,其中所有所述共有序列具有不同的氨基酸序列。

7.如权利要求1-6中任一项所述的经修饰的载体蛋白,其中所述载体蛋白是CRM197、来自破伤风杆菌的TT、来自铜绿假单胞菌的EPA、来自铜绿假单胞菌的Hcp1、来自金黄色葡萄球菌的Hla、来自金黄色葡萄球菌的ClfA、来自大肠杆菌的MBP、来自大肠杆菌的PspA或来自淋病奈瑟菌的MtrE。

8.如权利要求7所述的经修饰的载体蛋白,其中所述载体蛋白包含SEQ ID No:1-16中任一个的氨基酸序列,或与SEQ ID NO:1-16中任一个具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列或由其组成。

9.如权利要求8所述的经修饰的载体蛋白,其中所述经修饰的载体蛋白包含与SEQ ID NO.1至16中的任一个具有至少80%、85%、90%、92%、95%、96%或97%同一性的氨基酸序列或由其组成。

10.如权利要求1-9中任一项所述的经修饰的载体蛋白,其中所述修饰的载体蛋白是糖基化的。

11.包含如权利要求1-10中任一项所述的经修饰的载体蛋白的缀合物(例如生物缀合物),其中所述经修饰的载体蛋白连接至多糖。

12.如权利要求11所述的缀合物(例如生物缀合物),其中所述多糖与所述修饰的载体蛋白上的氨基酸连接,所述氨基酸选自天冬酰胺、天冬氨酸、谷氨酸、赖氨酸、半胱氨酸、酪氨酸、组氨酸、精氨酸或色氨酸(例如天冬酰胺)。

13.如权利要求11或12所述的缀合物(例如生物缀合物),其中所述多糖是细菌荚膜多糖。

14.如权利要求13所述的缀合物(例如生物缀合物),其中所述荚膜多糖选自:金黄色葡

萄球菌5型荚膜糖、金黄色葡萄球菌8型荚膜糖、脑膜炎奈瑟菌血清群A荚膜糖(MenA)、脑膜炎奈瑟菌血清群C荚膜糖(MenC)、脑膜炎奈瑟菌血清群Y荚膜糖(MenY)、脑膜炎奈瑟菌血清群W荚膜糖(MenW)、流感嗜血杆菌b型荚膜糖(Hib)、B组链球菌I组荚膜糖、B组链球菌II组荚膜糖、B组链球菌III组荚膜糖、B组链球菌IV组荚膜糖、B组链球菌V组荚膜糖、来自伤寒沙门氏菌的Vi糖、脑膜炎奈瑟菌LPS(如L3和/或L2)、卡他拉莫菌(M.catarrhalis)LPS、流感嗜血杆菌LPS、志贺氏菌O-抗原、铜绿假单胞菌O-抗原、大肠杆菌O-抗原或肺炎链球菌荚膜多糖。

15.如权利要求13或14所述的缀合物(例如生物缀合物),其中所述荚膜多糖来自与所述载体蛋白相同的生物体。

16.如权利要求15所述的缀合物(例如生物缀合物),其中所述荚膜多糖来自与所述载体蛋白不同的生物体。

17.编码如权利要求1-10中任一项所述的经修饰的载体蛋白的多核苷酸。

18.包含权利要求17的多核苷酸的载体。

19.一种宿主细胞,其包含:

a.一种或多种编码糖基转移酶的核酸;

b.编码寡糖基转移酶的核酸;

c.编码如权利要求1-10中任一项所述的经修饰的载体蛋白的核酸;以及任选地

d.编码聚合酶(例如wzy)的核酸。

20.如权利要求19所述的宿主细胞,其中所述宿主细胞包含(a)将己糖单糖衍生物组装到十一异戊烯基焦磷酸(Und-PP)上的糖基转移酶和(b)能够将单糖添加到在Und-PP上组装的所述己糖单糖衍生物的一种或多种糖基转移酶。

21.如权利要求20所述的宿主细胞,其中将己糖单糖衍生物组装到Und-PP上的所述糖基转移酶对于所述宿主细胞是异源的和/或对于编码糖基转移酶的一种或多种基因是异源的,任选地其中将己糖单糖衍生物组装到Und-PP上的所述糖基转移酶来自埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特菌属物种或弯曲杆菌属物种,任选地wecA(例如来自大肠杆菌的wecA)。

22.如权利要求19-21中任一项所述的宿主细胞,其中所述己糖单糖衍生物是其中C-2位被乙酰胺基修饰的任何单糖,如N-乙酰葡糖胺(GlcNAc)、N-乙酰半乳糖胺(GalNAc)、2,4-二乙酰胺基-2,4,6-三脱氧己糖(DATDH)、N-乙酰基岩藻糖胺(FucNAc),或N-乙酰基异鼠李糖胺(QuiNAc)。

23.如权利要求19-22中任一项所述的宿主,其中所述一种或多种能够将单糖加到装配在Und-pp上的己糖单糖衍生物的糖基转移酶是来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)或来自大肠杆菌O167的呋喃半乳糖基转移酶(wfdK),或者是来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)和来自大肠杆菌O28的呋喃半乳糖基转移酶(wfdK)。

24.如权利要求19-23中任一项所述的宿主细胞,其中所述糖基转移酶包含能够将存在于供体寡糖或多糖的第一重复单元的还原末端的己糖单糖添加至所述己糖单糖衍生物的糖基转移酶,任选地其中所述一种或多种能够将单糖添加至所述己糖单糖衍生物的糖基转

移酶包含半乳糖基转移酶(wciP),任选地来自大肠杆菌O21,并且任选地包含能够将与存在于所述供体寡糖或多糖的第一重复单元的还原末端的所述己糖单糖相邻的单糖添加至存在于所述供体寡糖或多糖的第一重复单元的还原末端的己糖单糖的糖基转移酶,任选地葡糖基转移酶(wciQ),任选地来自大肠杆菌O21。

25.如权利要求19-24中任一项所述的宿主细胞,其中所述寡糖基转移酶源自空肠弯曲杆菌,任选地其中所述寡糖基转移酶是空肠弯曲杆菌的PglB,任选地其中空肠弯曲杆菌的PglB基因被整合到宿主基因组中,并且任选地其中所述宿主细胞的至少一个基因已功能性失活或缺失,任选地其中所述宿主细胞的waaL基因已功能性失活或缺失,任选地其中所述宿主细胞的waaL基因已被编码寡糖转移酶的核酸取代,任选地其中所述宿主细胞的waaL基因已被空肠弯曲杆菌PglB取代。

26.如权利要求19-25中任一项所述的宿主细胞,其中编码所述经修饰的载体蛋白的核酸在宿主细胞中的质粒中。

27.如权利要求19-26中任一项所述的宿主细胞,其中编码所述经修饰的载体蛋白的核酸被整合到所述宿主细胞的基因组中。

28.如权利要求19-27中任一项所述的宿主细胞,其中所述宿主细胞是大肠杆菌。

29.一种生产生物缀合物的方法,所述生物缀合物包含与糖连接的经修饰的载体蛋白,所述方法包括(i)在适于生产蛋白质的条件下培养如权利要求19-28中任一项所述的宿主细胞,和(ii)分离所述生物缀合物。

30.通过如权利要求29所述的方法获得的生物缀合物,其中所述生物缀合物包含与经修饰的载体蛋白连接的多糖。

31.可通过如权利要求29所述的方法获得的生物缀合物,其中所述生物缀合物包含连接至经修饰的载体蛋白的多糖。

32.免疫原性组合物,其包含如权利要求1-10中任一项所述的经修饰的载体蛋白,或如权利要求11-16中任一项所述的缀合物或生物缀合物。

33.一种制备如权利要求31所述的免疫原性组合物的方法,其包括将所述修饰的载体蛋白或所述缀合物或所述生物缀合物与药学上可接受的赋形剂或载体混合的步骤。

34.一种疫苗,其包含如权利要求31所述的免疫原性组合物和药学上可接受的赋形剂或载体。

35.一种用于在有需要的受试者中治疗或预防细菌感染的方法,其包括向所述受试者施用治疗有效量的如权利要求1-10中任一项所述的经修饰的载体蛋白,或如权利要求11-16中任一项所述的缀合物或生物缀合物。

36.一种免疫人类宿主以抵抗细菌感染的方法,其包括向所述宿主施用免疫保护剂量的如权利要求1到10中任一项所述的经修饰载体蛋白或如权利要求11到16中任一项所述的缀合物或生物缀合物。

37.一种在受试者中诱导针对细菌的免疫应答的方法,所述方法包括向受试者施用治疗或预防有效量的如权利要求1-10中任一项所述的经修饰的载体蛋白,或如权利要求11-16中任一项所述的缀合物或生物缀合物。

38.如权利要求1-10中任一项所述的经修饰的载体蛋白,或如权利要求11-16中任一项所述的缀合物或生物缀合物,用于治疗或预防由细菌感染引起的疾病。

39.如权利要求1-10中任一项所述的经修饰的载体蛋白或如权利要求11-16中任一项所述的缀合物或生物缀合物在制备用于治疗或预防由细菌感染引起的疾病的药物中的用途。

40.如权利要求34-36中任一项所述的方法，或用于如权利要求37所述用途的载体蛋白、缀合物或生物缀合物，或如权利要求38所述的用途，其中所述细菌或细菌感染选自金黄色葡萄球菌、脑膜炎奈瑟菌、流感嗜血杆菌、流感嗜血杆菌b型、B组链球菌、伤寒沙门氏菌、卡他拉莫菌、福氏志贺氏菌、铜绿假单胞菌、大肠杆菌或肺炎链球菌。

41.一种测量如权利要求1至10中任一项所述的载体蛋白的糖基化位点占有率水平的方法，所述方法包括：用蛋白酶消化所述糖基化载体蛋白；将经消化的蛋白质进行LC-MS；测定未经修饰的载体蛋白的浓度U；测定总载体蛋白的浓度T；以及根据以下等式计算糖基化位点占有率：

$$\text{位点占有率}(\%) = \frac{(\text{总} - \text{未经修饰的})\text{载体浓度}}{\text{总载体浓度}} \times 100$$

。

42.如权利要求40所述的方法，其中通过测定对应于共有序列的肽片段的浓度来测定未经修饰的载体蛋白的浓度U。

43.如权利要求40或权利要求41所述的方法，其中总载体蛋白的浓度T通过测定所述载体蛋白独特的一个或多个肽片段的浓度进行测定。

44.如权利要求40至42中任一项所述的方法，其中所述蛋白酶是胰蛋白酶。

生物缀合物糖基化的定量

技术领域

[0001] 本发明涉及用于表征生物缀合物,特别是测量糖基化水平的分析工具。

背景技术

[0002] 已经证明糖缀合物疫苗(glycoconjugate vaccine)在预防由包封的细菌引起的感染性疾病中是有效的和经济的。在过去十年中,已经采取了新的方法用于糖缀合物疫苗生产,包括利用细菌N-糖基化的技术。这些“生物缀合”技术中最成熟的是基于在大肠杆菌中产生糖蛋白,其中空肠弯曲杆菌糖基化机制PglB与参与病原体多糖链生物合成的酶和靶载体蛋白、多糖的受体共表达(Wacker等人,2002,Science298:1790-3),其被工程化以含有PglB的共有序列。

[0003] 生物缀合技术的主要优势是载体蛋白序列上糖基化位点的选择性。这是通过选择性地将特定的氨基酸引入到载体蛋白序列中,从而产生一个或多个有效的共有序列用于选择性地缀合到多糖链上而实现的。PglB的核心共有序列是D/E-X-N-Z-S/T,其中X和Z独立地是除了脯氨酸之外的任何氨基酸(参见Wacker等人,2002,Science298:1790-3),但是K-D/E-X-N-Z-S/T-K的延伸共有序列以更高的效率糖基化并且被更广泛地使用(参见例如WO2019/121924和WO2019/121926)。

[0004] 该技术对于疫苗开发特别有意义,尤其是当选择具有作为载体和抗原的双重作用的载体蛋白时,因为它可以保留关键的保护性表位。此外,与化学缀合相比,生物缀合更适于大规模生产疫苗,因为它减少了对病原体进行操作的需求,允许生产过程步骤的减少,并且时间和资源消耗更少。

[0005] 最近提出了生物缀合疫苗候选物用于预防革兰氏阴性(肠道沙门氏菌(Salmonella enterica),志贺氏菌属(Shigella spp),致病性大肠杆菌)和革兰氏阳性病原体感染(肺炎链球菌和金黄色葡萄球菌)(例如Wetter等人,2013,Glycoconj J.30:51 1-22.Engineering,conjugation,and immunogenicity assessment of Escherichia coli O121O antigen for its potential use as a typhoid vaccine component;Wacker等人,2014,J.Infect.Dis.209:1551-1561;Van den Dobbelsteen等人,2016,Vaccine 34:4152-60)。其中与金黄色葡萄球菌5型CP(Hla-CP5)生物缀合的金黄色葡萄球菌α毒素(Hla)显示出诱导识别聚糖和蛋白质部分两者的兔或小鼠保护性抗体,表明Hla蛋白质作为载体和作为保护性抗原的双重作用(Wacker等人,2014,J.Infect.Dis.209:1551-1561)。该数据与开发防止金黄色葡萄球菌的扩散的疫苗的特别相关,由于在全世界传播的多重耐药性菌株的增加(包括医院和社区相关的感染菌株),防止金黄色葡萄球菌的扩散变得越来越具有挑战性。

[0006] 尽管在疫苗开发领域中生物缀合物的关注增加,但仍缺乏评估载体糖基化效率所需的稳健分析工具(Micoli,F.等人,2018,Molecules23:1451)。特别地,糖基化程度的精确定量仍然是一个具有挑战性的任务,尽管该信息对于满足潜在的监管要求和监测抗原产生和表征是基本的。因此,在本领域中需要用于精确定量生物缀合物中糖基化位点占据的绝

对水平的稳健且可靠的方法。

发明内容

[0007] 本发明人已经设计了适于糖基化位点占有率的绝对定量的用于蛋白质N-糖基化的通用共有序列。具体地，这些共有序列的使用允许通过在液相色谱与串联质谱(LC-MS/MS)分析中使用重同位素标记的内标同时定量总体蛋白质浓度和蛋白质的未糖基化部分，并且精确地确定位点占据的程度(Zhu等人2015,J Am Soc Mass Spectrom.25:1012-7).

[0008] 基于Zhu等人的方法，本发明人设计了一种使用含有SEQ ID NO:40的糖基化共有位点KDQNRTK(描述于WO2019/121924)的Hla载体蛋白作为模型来定量糖基化的方法。该策略基于使用同位素标记的内标定量天然未糖基化形式的糖肽。简言之，在胰蛋白酶消化之前将两组重同位素标记的肽标准品掺入样品中，并通过LC-MS分析消化的样品。一组肽标准品用于测定总糖蛋白量，而另一标准品监测未糖基化的糖蛋白量。以这种方式，通过从总蛋白量中减去未糖基化的蛋白量来计算蛋白质糖基化部分的丰度，然后测定位点占有率。

[0009] 然而，发现KDQNRTK(SEQ ID NO:40)共有序列产生太短且太亲水而不能进行LC-MS定量的胰蛋白酶肽。其它常用的共有序列如KDQNATK(SEQ ID NO:41)也会遇到同样的问题。

[0010] 本发明人因此提出设计将与该方法相容的通用共有位点，即，其将以与先前使用的位点至少相同的效率糖基化，并且还将产生可通过LC-MS检测的胰蛋白酶肽。使用Hla作为原理验证，他们能够成功地设计适于通过质谱定量缀合程度的共有序列。

[0011] 本发明允许定量终产物中未糖基化的载体的量；过程中追踪生物缀合率；以及定量单个和多个共有位点上的糖基化程度。此外，检测的选择性降低了对于支持过程中和最终产物表征的大量样品纯化的必要性。

[0012] 因此，在第一方面，本发明提供包含以下氨基酸序列或由以下氨基酸序列组成的共有序列：

[0013] $K/R\text{-}Z_{0\text{-}9}\text{-}D/E\text{-}X\text{-}N\text{-}Y\text{-}S/T\text{-}Z_{0\text{-}9}\text{-}K/R$

[0014] 其中X和Y独立地是除了脯氨酸之外的任何氨基酸，并且Z表示任何氨基酸。在一个优选的实施方案中，X和Y独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸。在一个实施方案中，Z表示除赖氨酸或精氨酸以外的任何氨基酸。在一个实施方案中，X、Y和/或Z不是芳香族或疏水性氨基酸。在一个优选的实施方案中，Z表示除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸(例如SEQ ID NO:47)。

[0015] 在一个具体的实施方案中，本发明提供了包含氨基酸序列$K/R\text{-}D/E\text{-}X_1\text{-}N\text{-}X_2\text{-}S/T\text{-}Z_1\text{-}Z_2\text{-}K/R$(SEQ ID NO 19)或由其组成的共有序列，其中X_1和X_2独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸或半胱氨酸。在一个实施方案中，所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含SEQ ID No:42、SEQ ID No:43;SEQ ID No:44或SEQ ID No:45，优选SEQ ID No:42-44中的任何一个的氨基酸序列或由其组成。

[0016] 在一个方面，本发明提供了经修饰的载体蛋白，经修饰以使得其包含一个或多个本发明的共有序列。因此，本发明提供了经修饰的载体蛋白，其经修饰以使得其包含一个或多个共有序列，所述共有序列包含或由如上定义的氨基酸序列$K/R\text{-}Z_{0\text{-}9}\text{-}D/E\text{-}X\text{-}N\text{-}Y\text{-}S/T\text{-}Z_{0\text{-}9}\text{-}K/R$(SEQ ID NO:46)组成；例如氨基酸序列$K/R\text{-}D/E\text{-}X_1\text{-}N\text{-}X_2\text{-}S/T\text{-}Z_1\text{-}Z_2\text{-}K/R$(SEQ ID

NO 19)，其中X_1和X_2独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸或半胱氨酸。在一个实施方案中，共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42、SEQ ID No:43;SEQ ID No:44或SEQ ID No:45，优选SEQ ID No:42-44中的任何一个的氨基酸序列组成。

[0017] 在一个实施方案中，所述共有序列已经取代了载体蛋白序列的一个或多个氨基酸。在另一个实施方案中，所述共有序列已被插入到载体蛋白序列中。

[0018] 经修饰的载体蛋白可以包含多于一个所述共有序列，任选地至少2、3、4或5个共有序列。在一个优选的实施方案中，当经修饰的载体蛋白含有多于一个共有序列时，所有的所述共有序列是不同的，即具有不同的氨基酸序列。

[0019] 载体蛋白可以是任何蛋白，优选能够引发T-依赖性免疫应答的蛋白。在具体的实施方案中，载体蛋白是CRM197、来自破伤风杆菌的TT、来自铜绿假单胞菌的EPA、来自铜绿假单胞菌的Hcp1、来自金黄色葡萄球菌的Hla、来自金黄色葡萄球菌的ClfA、来自大肠杆菌的MBP、来自大肠杆菌的PspA或来自淋病奈瑟菌的MtrE。

[0020] 在一个实施方案中，载体蛋白包含或由SEQ ID No:1-16的任一个的氨基酸序列或与SEQ ID No.1-16的任一个具有至少80％，85％，90％，92％，95％，96％，97％，98％或99％同一性的氨基酸序列组成。

[0021] 在一个实施方案中，经修饰的载体蛋白包含或由与SEQ ID NO.1-16的任一个具有至少80％，85％，90％，92％，95％，96％或97％同一性的氨基酸序列组成。

[0022] 经修饰的载体蛋白可以是糖基化的。本发明还提供了本发明的糖基化载体蛋白以及包含与多糖连接的本发明经修饰的载体蛋白的缀合物(例如生物缀合物)。多糖与经修饰的载体蛋白上的氨基酸连接，所述氨基酸选自天冬酰胺、天冬氨酸、谷氨酸、赖氨酸、半胱氨酸、酪氨酸、组氨酸、精氨酸或色氨酸(优选天冬酰胺)。在一个实施方案中，荚膜多糖来自与载体蛋白相同的生物体。在一个实施方案中，荚膜多糖来自与载体蛋白不同的生物体。

[0023] 在一个实施方案中，所述多糖是细菌荚膜多糖，例如金黄色葡萄球菌5型荚膜糖、金黄色葡萄球菌8型荚膜糖、脑膜炎奈瑟菌血清群A荚膜糖(MenA)、脑膜炎奈瑟菌血清群C荚膜糖(MenC)、脑膜炎奈瑟菌血清群Y荚膜糖(MenY)、脑膜炎奈瑟菌血清群W荚膜糖(MenW)、流感嗜血杆菌b型荚膜糖(Hib)、B组链球菌I组荚膜糖、B组链球菌II组荚膜糖、B组链球菌III组荚膜糖、B组链球菌IV组荚膜糖、B组链球菌V组荚膜糖、来自伤寒沙门氏菌的Vi糖、脑膜炎奈瑟菌LPS(脂多糖，如L3和/或L2)、卡他拉莫菌(M.catarrhalis)LPS、流感嗜血杆菌LPS、志贺氏菌O-抗原、铜绿假单胞菌O-抗原、大肠杆菌O-抗原或肺炎链球菌或金黄色葡萄球菌荚膜多糖。

[0024] 根据本发明的另一方面，提供了编码本发明的经修饰的载体蛋白或生物缀合物的多核苷酸。

[0025] 根据本发明的另一方面，提供了包含编码本发明的经修饰的载体蛋白或生物缀合物的多核苷酸的载体。

[0026] 根据本发明的另一方面，提供了一种宿主细胞，其包含：

[0027] i)一种或多种编码糖基转移酶的核酸；

[0028] ii)编码寡糖基转移酶的核酸；

[0029] iii)编码本发明的经修饰的载体蛋白的核酸;和任选地

[0030] iv)编码聚合酶的核酸(例如wzy)。

[0031] 编码经修饰的载体蛋白的核酸可以携带在宿主细胞中的质粒上,或者可以整合到宿主细胞的基因组中。宿主细胞优选是大肠杆菌。

[0032] 根据本发明的另一方面,提供了用于产生生物缀合物的方法,所述生物缀合物包含连接至糖的经修饰的载体蛋白(或由其组成),所述方法包括:(i)在适于产生蛋白质的条件下培养本发明的宿主细胞,和(ii)分离由所述宿主细胞产生的生物缀合物。还提供了通过所述方法获得或可获得的生物缀合物,其中所述生物缀合物包含连接至经修饰的载体蛋白的多糖。

[0033] 根据本发明的另一方面,提供了免疫原性组合物,其包含本发明的经修饰的载体蛋白,或本发明的缀合物,或本发明的生物缀合物,和药学上可接受的赋形剂或载体。

[0034] 根据本发明的另一方面,提供了制备本发明的免疫原性组合物的方法,所述方法包括将本发明的经修饰的载体蛋白或缀合物或生物缀合物与药学上可接受的赋形剂或载体混合的步骤。

[0035] 根据本发明的另一方面,提供了包含本发明的经修饰的载体蛋白、缀合物或生物缀合物的免疫原性组合物。

[0036] 根据本发明的另一方面,提供了制备包含本发明的经修饰的载体蛋白、缀合物或生物缀合物的免疫原性组合物的方法,所述方法包括将本发明的经修饰的载体蛋白或缀合物或生物缀合物与药学上可接受的赋形剂或载体混合的步骤。

[0037] 根据本发明的另一方面,提供了包含本发明的免疫原性组合物和药学上可接受的赋形剂或载体的疫苗。

[0038] 根据本发明的另一方面,提供了用于治疗或预防有需要的受试者中的细菌感染的方法,所述方法包括向所述受试者施用治疗有效量的本发明的经修饰的载体蛋白、缀合物或生物缀合物。

[0039] 根据本发明的另一方面,提供了一种免疫人宿主以抵抗细菌感染的方法,包括向宿主使用免疫保护剂量的本发明的经修饰的载体蛋白、缀合物或生物缀合物。

[0040] 根据本发明的另一方面,提供了在受试者中诱导对细菌的免疫应答的方法,所述方法包括施用治疗或预防有效量的本发明的经修饰的载体蛋白、缀合物或生物缀合物。

[0041] 根据本发明的另一方面,提供了用于治疗或预防由细菌感染引起的疾病的本发明的经修饰的载体蛋白、缀合物或生物缀合物。

[0042] 根据本发明的另一方面,提供了本发明的经修饰的载体蛋白、缀合物或生物缀合物在制备用于治疗或预防由细菌感染引起的疾病的药物中的用途。

[0043] 在具体的实施方案中,所述细菌或细菌感染选自金黄色葡萄球菌、脑膜炎奈瑟菌、流感嗜血杆菌、流感嗜血杆菌b型、B组链球菌、伤寒沙门氏菌、卡他拉莫菌LPS、福氏志贺氏菌、铜绿假单胞菌、大肠杆菌或肺炎链球菌。

[0044] 根据本发明的另一方面,提供了测量本发明的载体蛋白的糖基化位点占有率水平的方法,所述方法包括用蛋白酶(例如胰蛋白酶)消化糖基化的载体蛋白;将经消化的蛋白质进行LC-MS;测定未修饰的载体蛋白的浓度U;测定总载体蛋白的浓度T;以及根据以下等式计算糖基化位点占有率:

[0045] $$位点占有率(\%)=\frac{(总-未经修饰的)载体浓度}{总载体浓度}\times 100$$

[0046] 通过测定对应于本发明共有序列的肽片段的浓度来测定未经修饰的载体蛋白的浓度U。总载体蛋白的浓度T可适当地通过测定所述载体蛋白独特的一个或多个肽片段的浓度来确定。

附图说明

[0047] 图1:所采取的策略的工作流程。

[0048] 图2A和2B:共有序列的计算机设计。(A)在32中天然空肠弯曲杆菌糖蛋白中发现的糖基化Asn残基的-6至+6区域中出现的氨基酸的统计分析。该分析报道于Kowarik等人,EMBO J.2006;25(9):1957-66。框的高度反映了天然存在的共有序列中氨基酸残基的频率。被证明对有效糖基化至关重要的Asn残基(在0位,糖基化位点)以及分别在-2和+2位的Asp和Thr残基在灰色框中以粗体表示。选择分别在位置-3、-1、+1、+3和+4的氨基酸残基(在灰色框中以粗体表示)用于设计四种共有序列(B)。

[0049] 图3A和3B:通过蛋白印迹评估新设计的载体的生物缀合效率。使用兔抗-Hla-CP5血清通过蛋白印迹分析从经工程化用于表达Hla-i-CP5、Hla-v-CP5和Hla-s-CP5(图3A,泳道1-3)的大肠杆菌制备的周质部分。将表达水平与携带与MS分析不相容的共有序列KDQNRTK的优化Hla(图3A,泳道4)进行比较。作为阴性对照,报道了从不表达Hla的相应菌株制备的周质部分的蛋白印迹分析(图3B,泳道1-4)。观察到的阳性信号可能与反应中间体十一异戊烯基连接的CP5分子(在该过程中产生并组装)有关。

[0050] 图4:剂量-反应线性。作为实例,报道了PTP-i的剂量-反应线性曲线。为了建立校准曲线,在y轴上绘制L/H面积比响应,其通过在胰蛋白酶消化之前,在50μg的大肠杆菌周质部分中掺入固定量的重形式PTPi(0.1pmol/μg)和标量浓度的轻PTPi(范围从0.0125至1.6pmol/μg,x轴)进行测定。根据国际协调会议(ICH)指南(www.ich.org/products/guidelines/quality/article/quality-guidelines.html),将每种肽的定量下限(LLOQ)设定为可定量检测的拟合曲线上的最低浓度点,并定义为10σ/S,其中σ=反应的标准偏差,且S=校准曲线的斜率,并计算为0.08pmol/μg周质蛋白。以相同的方式定义,对于PTP-s和PTP-v,LLOQ分别为0.11和0.06pmol/μg周质蛋白。

[0051] 图5:(A):构建体Hla-N,131、Hla-N,C和Hla-131,C的示意图,其各自携带用于生物缀合至金黄色葡萄球菌CP5的两个共有序列,可选地位于载体蛋白的N-末端、C-末端或131位,心脏它们各自在CP5生物缀合物中的计算糖基化程度。(B):在第131位插入的共有序列的糖基化%作为周质部分中蛋白质的总量的函数的曲线。

具体实施方式

[0052] 定义

[0053] 如本文所用,术语“载体蛋白”是指共价连接至多糖抗原(例如糖抗原)以产生缀合物(例如生物缀合物)的蛋白质。载体蛋白激活与其所缀合的多糖抗原相关的T细胞介导的免疫。

[0054] ClfA:来自葡萄球菌,特别是金黄色葡萄球菌的凝集因子A。

[0055] CRM197:白喉毒素的非毒性突变体。

[0056] EPA:铜绿假单胞菌的外毒素A。

[0057] HLA:溶血素A,也称为α毒素,来自葡萄球菌,特别是金黄色葡萄球菌。

[0058] Hcp1:来自铜绿假单胞菌的蛋白Hcp1

[0059] MBP:来自大肠杆菌的麦芽糖/麦芽糊精结合蛋白。

[0060] MtrE:来自淋病奈瑟菌的膜转运蛋白E。

[0061] PspA,来自大肠杆菌的噬菌体休克蛋白A。

[0062] CP:荚膜多糖。

[0063] 如本文所用术语“生物缀合物”是指在宿主细胞背景中制备的蛋白质(例如载体蛋白)和抗原(例如糖)之间的缀合物,其中宿主细胞机制将抗原连接至蛋白质(例如N-连接)。通常,在生物缀合物中,多糖通过N-乙酰葡糖胺连接至天冬酰胺。

[0064] 本文所用的术语“糖基化位点(glycosite)”是指由细菌寡糖转移酶例如空肠弯曲杆菌的PplB识别的氨基酸序列。PplB的最小共有序列是D/E-X-N-Z-S/T(SEQ ID NO:17),而延伸的共有序列K-D/E-X-N-Z-S/T-K(SEQ ID NO:18)也已被确定。本文描述了示例性的和替代的糖基化位点序列。

[0065] 除脯氨酸(pro,P)外的任何氨基酸:指选自丙氨酸(ala,A)、精氨酸(arg,R)、天冬酰胺(asn,N)、天冬氨酸(asp,D)、半胱氨酸(cys,C)、谷氨酰胺(gin,Q)、谷氨酸(glu,E)、甘氨酸(gly,G)、组氨酸(his,H)、异亮氨酸(ile,I)、亮氨酸(leu,L)、赖氨酸(lys,K)、甲硫氨酸(met,M)、苯丙氨酸(phe,F)、丝氨酸(ser,S)、苏氨酸(thr,T)、色氨酸(trp,W)、酪氨酸(tyr,Y)、缬氨酸(val,V)的氨基酸。

[0066] 如本文所用,在向受试者施用疗法(例如本发明的免疫原性组合物或疫苗)的上下文中,术语“有效量”是指具有预防和/或治疗效果的疗法的量。

[0067] 如本文所用,术语“受试者”是指动物,特别是哺乳动物如灵长类动物(例如人)。

[0068] 如本文所用,提及两个氨基酸或核酸序列之间的百分序列同一性是指当比对时,在比较两个序列中相同的氨基酸或碱基的百分比。这种比对和百分同源性或序列同一性可以使用本领域已知的软件程序测定,例如描述于Current Protocols in Molecular Biology(F.M.Ausubel etal.,eds.,1987,Supplement 30)的7.7.18节中的那些。通过Smith-Waterman同源性搜索算法,使用空位开放罚分12和空位延伸罚分2,BLOSUM矩阵62的affine gap搜索来确定优选的比对。Smith-Waterman同源性搜索算法描述于Smith&Waterman(1981)Adv.Appl.Math.2:482-489。与任何特定序列(例如与特定SEQ ID)的百分同一性理想地在该序列的整个长度上计算。优选在较长序列的长度上计算不同长度的两个序列之间的百分序列同一性。可以使用全局或局部比对。优选地,使用全局比对。

[0069] 如本文所用,术语融合蛋白或目标蛋白或其缀合物(例如生物缀合物)的“纯化(purifying)”或“纯化(purification)”是指将其与一种或多种污染物分离。污染物是不同于所述融合蛋白或目标蛋白或其缀合物(例如生物缀合物)的任何物质。污染物可以是例如细胞碎片、核酸、脂质、除融合蛋白或目标蛋白以外的蛋白质、多糖和其它细胞组分。

[0070] “重组”多肽是在宿主细胞中产生的多肽,所述宿主细胞已经用编码所述多肽的核酸转化或转染,或者由于同源重组而产生所述多肽。

[0071] 本文所用的术语“保守氨基酸置换”包括用非天然残基取代天然氨基酸残基,使得

对该位置处的氨基酸残基的大小、极性、电荷、疏水性或亲水性影响很小或没有影响，并且不会导致降低的免疫原性。例如，这些可以是以下组内的置换：缬氨酸、甘氨酸；甘氨酸、丙氨酸；缬氨酸、异亮氨酸、亮氨酸；天冬氨酸、谷氨酸；天冬酰胺、谷氨酰胺；丝氨酸、苏氨酸；赖氨酸、精氨酸；和苯丙氨酸、酪氨酸。对多肽序列的保守氨基酸修饰(以及对编码核苷酸的相应修饰)可产生具有与亲本多肽相似的功能和化学特征的多肽。

[0072] 如本文所用，术语“缺失”是从蛋白质序列中去除一个或多个氨基酸残基。通常，在蛋白质分子内的任何一个位点处缺失不超过约1至6个残基(例如1至4个残基)。

[0073] 如本文所用，术语“插入”是在蛋白质序列中添加一个或多个非天然氨基酸残基。通常，在蛋白质分子内的任何一个位点处插入不超过约1至6个残基(例如1至4个残基)。

[0074] 如本文所用，术语“包含”表示可以存在除所提及的那些之外的其它组分，而术语“由…组成”表示其它组分不存在或不以可检测的量存在。术语“包含”自然包括术语“由…组成”。

[0075] 载体蛋白

[0076] 非T-依赖性抗原如糖与载体蛋白的缀合早已被确立为使T-细胞能够帮助成为对正常非T-依赖性抗原的免疫应答的一部分的方式。以这种方式，免疫应答可以通过允许免疫记忆的发展和应答的增强而增强。载体蛋白将非T-依赖性糖抗原转变为能够引发免疫记忆应答的T-依赖性抗原。通过将细菌荚膜糖与载体蛋白缀合而开发的成功的结合疫苗是本领域已知的；广泛用于商业化疫苗的载体蛋白包括破伤风类毒素、白喉类毒素、CRM197和流感嗜血杆菌的蛋白D。CRM197目前用于肺炎链球菌荚膜多糖结合疫苗PREVENAR^{TM}(Pfizer)且蛋白D、破伤风类毒素和白喉类毒素目前用作肺炎链球菌荚膜多糖结合疫苗SYNFLORIX^{TM}(GlaxoSmithKline)中荚膜多糖的载体。本领域已知的其它载体蛋白包括金黄色葡萄球菌血清型5和8荚膜多糖的EPA(铜绿假单胞菌的外毒素A)(Wacker等人，2014，J Infect.Dis.209:1551-1561)。

[0077] 为了提高缀合物的保护能力，也可以使用来自与缀合的多糖相同的生物体的蛋白抗原作为载体蛋白。例如，金黄色葡萄球菌蛋白抗原H1a已经成功地用作金黄色葡萄球菌荚膜多糖的载体蛋白。如WO2019/121924，WO2019/121926和PCT/EP2019/053463中所述，用H1a-CP5和ClfA-CP8生物缀合物接种能够在动物模型中诱导针对荚膜多糖和蛋白抗原两者的功能性抗体，并赋予针对金黄色葡萄球菌感染的保护。因此，任何蛋白抗原可以是用作多糖缀合物疫苗中的载体蛋白的候选物。优选地，所述蛋白抗原来自与多糖相同的生物体。然而，也可以使用来自不同生物体的蛋白抗原，例如以赋予针对多种病原体的保护。

[0078] 可用于本发明的示例性载体蛋白如下所述。

[0079] EPA：铜绿假单胞菌的外毒素A

[0080] 在一个实施方案中，载体蛋白是来自铜绿假单胞菌的外毒素A(EPA)。所述EPA可包含SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列。

[0081] 因此，在本发明的一个方面，提供了经修饰的EPA蛋白质，其包含SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其修饰使得包含一个或多个共有序列，所述共有序列包含如上定义的氨基酸序列K/R-$(Z)_{0-9}$-D/E-X-N-Y-S/T-$(Z)_{0-9}$-K/R(SEQ ID NO:46)或由其组成；例如具

有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO 19)的一个或多个共有序列,其中X_1和X_2独立地是除了脯氨酸、精氨酸和赖氨酸之外的任何氨基酸,并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中,所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中,所述共有序列包含或由SEQ ID No:42-45中任何一个的氨基酸序列组成。在一个优选的实施方案中,所述共有序列包含或由SEQ ID No:42-44中任何一个的氨基酸序列组成。

[0082] EPA蛋白质可被进一步修饰使得其包含脱毒突变,例如在对应于SEQ ID NO:1的L552位的氨基酸位置处的L至V置换,和/或在SEQ ID NO:1的E553或在与SEQ ID NO:1具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列(例如SEQ ID NO:2)内的等同位置处的缺失;和/或一个或多个氨基酸已被一个或多个共有序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO 19)取代。在一个实施方案中,所述取代是SEQ ID NO:1的A375、A376或K240的取代。因此,目标蛋白可包含SEQ ID NO:2的氨基酸序列或与SEQ ID NO:2具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其用具有SEQ ID NO:19、20或42-47的氨基酸序列的共有序列插入或取代一个或多个氨基酸。

[0083] 在一个实施方案中,所述经修饰的EPA蛋白包含多于一个所述共有序列,例如2、3、4或5个共有序列。在一个优选的实施方案中,其中存在多个共有序列,共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0084] 在一个实施方案中,本发明的经修饰的EPA蛋白可以衍生自与SEQ ID NO:1具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是SEQ ID NO:1的免疫原性片段和/或变体。在一个实施方案中,本发明的经修饰的EPA蛋白可衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:1或2的免疫原性片段,其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0085] 在一个实施方案中,本发明的经修饰的EPA蛋白可衍生自与SEQ ID NO:1具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是已通过缺失和/或添加和/或一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)的置换进行修饰的SEQ ID NO:1的变体。氨基酸置换可以是保守的或非保守的。在一个方面,氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合,只要该变体是免疫原性多肽。在一个实施方案中,本发明的经修饰的EPA蛋白可衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0086] 在一个实施方案中,本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测,例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK),和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0087] 术语“经修饰的EPA蛋白”是指已经通过添加、取代或缺失一个或多个氨基酸(例如,添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或通过用具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的EPA氨基酸序列(例如,具有SEQ ID NO:1的EPA氨基酸序列或与SEQ ID NO:1具有至少80%、85%、90%、91%、

92％、93％、94％、95％、96％、97％、98％或99％同一性的氨基酸序列)，该EPA氨基酸序列可以是野生型成熟EPA氨基酸序列(例如，SEQ ID NO:1的野生型氨基酸序列)。经修饰的EPA蛋白质还可以包含进一步的修饰(添加、置换、缺失)以及一个或多个共有序列的添加或取代。例如，可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如，在信号序列之后或肽标签之前，如果存在的话)。在一个实施方案中，本发明的经修饰的EPA蛋白可以是非天然存在的EPA蛋白。

[0088] 在本发明的一个实施方案中，经修饰的EPA氨基酸序列(例如，具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的EPA氨基酸序列，例如SEQ ID No 2)的一个或多个氨基酸(例如1-7个氨基酸，例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:46或SEQ ID NO:19，例如SEQ ID NO:20或42-45或47，特别是SEQ ID No:42-44的共有序列取代。例如，EPA氨基酸序列(例如SEQ ID NO:1)中的单个氨基酸可以被所述共有序列取代。或者，EPA氨基酸序列(例如SEQ ID NO:1或与SEQ ID NO:1具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的EPA氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列取代。

[0089] 引入共有序列能够使经修饰的EPA蛋白糖基化。因此，本发明还提供了本发明的经修饰的EPA蛋白，其中经修饰的EPA蛋白是糖基化的。在特定的实施方案中，将共有序列引入EPA氨基酸序列的特定区域中，例如蛋白质的表面结构、在蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0090] 本领域技术人员将理解，当EPA氨基酸序列是SEQ ID NO:2的氨基酸序列的变体和/或片段时，如与SEQ ID NO:2具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，提及“氨基酸之间……”是指如果该序列与SEQ ID NO:1的氨基酸序列对齐以最大化两个序列之间的序列同一性，则将等同于所定义位置的位置。序列比对工具描述如上。在SEQ ID NO:1的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异，然而，通过将突变序列与参考序列对齐，可以识别与参考序列中相应氨基酸的等同位置的氨基酸，并且因此可以确定用于共有序列添加或替换的适当位置。

[0091] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即，全部或部分添加糖基化位点)，或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即，不将氨基酸添加到蛋白质中，而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到，使用本领域已知的方法，例如包括修饰编码蛋白质的核酸序列的重组方法，可以容易地修饰蛋白质的氨基酸序列。

[0092] 在一个实施方案中，本发明的经修饰的EPA蛋白还包含“肽标签”或“标签”，即允许分离和/或鉴定经修饰的EPA蛋白的氨基酸序列。例如，向本发明的经修饰的EPA蛋白中添加标记可用于纯化该蛋白，并且因此可用于纯化包含标记的经修饰的EPA蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中，标签是六组氨酸标签。在某些实施方案中，一旦不再需要标签，例如在蛋白质已经被纯化之后，在此使用的标签是可去除的，例如通过化学试剂或通过酶方式去除。任选地，肽标签位于氨基酸序列的C-末端。任选地，肽标签包含在氨基酸序列的C-末端的6个组氨酸残基。肽标签可以包含

或在其前面有一个、两个或更多个另外的氨基酸残基，例如丙氨酸、丝氨酸和/或甘氨酸残基，例如GS。

[0093] Hla:金黄色葡萄球菌的溶血素A

[0094] 在一个实施方案中，载体蛋白是Hla(金黄色葡萄球菌溶血素A，也称为α毒素)。如上所述，HLA已经成功地用作金黄色葡萄球菌荚膜多糖的载体蛋白。Hla的成熟野生型氨基酸序列在SEQ ID NO 13中给出。

[0095] 因此，在本发明的一个方面，提供了经修饰的Hla蛋白，其包含SEQ ID NO:13的氨基酸序列或与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其修饰为包含一个或多个共有序列，所述共有序列包含或由如上定义的氨基酸序列K/R-$(Z)_{0-9}$-D/E-X-N-Y-S/T-$(Z)_{0-9}$-K/R(SEQ ID NO:46)组成；例如选自以下的一个或多个共有序列:K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO 19)，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中，共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选实施方案中，所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0096] 在一个实施方案中，所述经修饰的Hla蛋白包含多于一个所述共有序列，例如2、3、4或5个共有序列。在一个优选的实施方案中，其中存在多个共有序列，所述共有序列具有不同的序列以便可以区分每个单独位点的糖基化。在一个实施方案中，所述经修饰的Hla蛋白在Hla序列的N-末端和/或C-末端包含多个共有序列中的至少一个。

[0097] 因为Hla是毒素，所以在其可以体内施用之前需要将其脱毒(即当以适于保护的剂量提供时，使其对哺乳动物例如人无毒)。本发明的经修饰的Hla蛋白可以是遗传脱毒的(即通过突变)。遗传脱毒的序列可以去除不需要的活性，如形成脂双层穿透孔、膜渗透、细胞裂解的能力，和针对人红细胞和其它细胞的细胞溶解活性，以降低毒性，同时保留在施用于人后诱导抗-HIa保护性和/或中和抗体的能力。例如，如本文所述，可改变Hla蛋白，使其无生物活性，同时仍保持其免疫原性表位。本发明的经修饰的Hla蛋白可以通过一个或多个点突变遗传脱毒。例如，参与孔形成的残基与Hla的裂解活性有关。在一个方面，本发明的经修饰的Hla蛋白可以通过如Menzies和Kernodle(Menzies和Kernodle，1994，Infect Immun 62，1843-1847)中所述的氨基酸置换来脱毒，例如用另一氨基酸如赖氨酸置换H35、H48、H114和/或H259。例如，参照SEQ ID NO:13的氨基酸序列(或与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列中的等同位置)，本发明的经修饰的Hla蛋白可包含选自H35L、H114L或H259L的至少一个的氨基酸置换。优选地，经修饰的Hla蛋白包含置换H35L(例如SEQ ID NO:14)。

[0098] 因此，所述经修饰的Hla蛋白可被进一步修饰为所述氨基酸序列包含脱毒突变，例如在SEQ ID NO:13的位置H35处(例如H35L)或在与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列(例如SEQ ID NO:14和15)内的等同位置处的氨基酸置换。替代的脱毒突变是用PSGS取代Hla单体的茎区域，例如SEQ ID NO:16中。示例性的修饰序列是SEQ ID NO:31-34和36-39，特别是31-33和36-38的那些。

[0099] 在一个实施方案中，所述Hla序列可以可选地或另外地被修饰为氨基酸序列包含在SEQ ID NO:13的H48和G122位置或与SEQ ID NO:13具有至少80％、85％、90％、92％、

95％、96％、97％、98％或99％同一性的氨基酸序列内的等同位置处的氨基酸置换，其中所述置换分别为H置换为C和G置换为C(例如SEQ ID NO:15)。

[0100] 因此，提供了经修饰的H1a蛋白，其包含SEQ ID NO:15的氨基酸序列或与SEQ ID NO:15具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其中所述经修饰的H1a蛋白含有以下突变:H35L、H48C和G122C，其修饰为包含一个或多个选自SEQ ID No:19、20和42-45，特别是42-44的共有序列的氨基酸序列，其中所述经修饰的H1a蛋白含有以下突变:H35L，H48C和G122C。示例性序列是SEQ ID NO:31-34和36-39，特别是31-33和36-38中的那些。

[0101] 如本文所述，这些序列可以通过添加信号序列和任选地插入N-末端丝氨酸和/或丙氨酸进行修饰用于克隆目的。可以进一步修饰序列以包含脱毒突变，如本文所述的脱毒突变中的任何一个或全部。优选的脱毒突变是SEQ ID NO:14或15的H35L。

[0102] 在一个实施方案中，本发明的经修饰的H1a蛋白可以衍生自与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是SEQ ID NO:13的免疫原性片段和/或变体。在一个实施方案中，本发明的经修饰的H1a蛋白可以衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:13、14或15的免疫原性片段，其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0103] 在一个实施方案中，本发明的经修饰的H1a蛋白可衍生自与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)进行修饰的SEQ ID NO:13的变体。氨基酸置换可以是保守的或非保守的。在一个方面，氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合，只要该变体是免疫原性多肽。在一个实施方案中，本发明的经修饰的H1a蛋白可以衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。例如，本发明的经修饰的H1a蛋白可以衍生自是SEQ ID NO.13-16中任何一个的变体的氨基酸序列，其中它在N末端具有一个或两个另外的氨基酸，例如初始N末端SA(例如SEQ ID NO:36-39)。经修饰的H1a蛋白可以另外地或可选地在C末端具有一个或多个另外的氨基酸，例如1、2、3、4、5或6个氨基酸。此类另外的氨基酸可以包括肽标签以帮助纯化，并且包括例如GSHRHR(例如SEQ ID NO:36-39)。

[0104] 在一个实施方案中，本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测，例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK)，和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0105] 术语“经修饰的H1a蛋白”是指已经通过添加、取代或缺失一个或多个氨基酸(例如，用半胱氨酸置换SEQ ID NO:13的H48和G122，用赖氨酸置换SEQ ID NO:1的H35，添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或用具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的H1a氨基酸序列(例如，具有SEQ ID NO:13的H1a氨基酸序列或与SEQ ID NO:13具有至少80％、85％、90％、91％、92％、93％、94％、

95％、96％、97％、98％或99％同一性的氨基酸序列)，所述H1a氨基酸序列可以是野生型成熟H1a氨基酸序列(例如，SEQ ID NO:13的野生型氨基酸序列)。经修饰的H1a蛋白还可以包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如，可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如，在信号序列之后或肽标签之前，如果存在的话)。在一个实施方案中，本发明的经修饰的H1a蛋白可以是非天然存在的H1a蛋白。

[0106] 在本发明的一个实施方案中，经修饰的H1a氨基酸序列(例如，具有SEQ ID NO:13的氨基酸序列或与SEQ ID NO:13具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的H1a氨基酸序列，例如SEQ ID No:14-16)的一个或多个氨基酸(例如1-7个氨基酸，例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:46或SEQ ID NO:19，例如SEQ ID NO:20或42-45或47，特别是SEQ ID No:42-44的共有序列取代。例如，H1a氨基酸序列(例如SEQ ID NO:13)中的单个氨基酸可以被所述共有序列(例如SEQ ID NO:30-39)取代。在一个实施方案中，所述取代的氨基酸位于对应于SEQ ID NO:13的K131位的位置。或者，H1a氨基酸序列(例如SEQ ID NO:13或与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的H1a氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列(例如SEQ ID NO:51-53)取代。在一个实施方案中，所述取代的氨基酸是选自SEQ ID NO:13的N-末端、C-末端和131位的氨基酸中的2个或更多个氨基酸。

[0107] 引入共有序列能够使经修饰的H1a蛋白糖基化。因此，本发明还提供了本发明的经修饰的H1a蛋白，其中经修饰的H1a蛋白是糖基化的。在特定的实施方案中，将共有序列引入H1a氨基酸序列的特定区域中，例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。在本发明的一个方面，共有序列的位置提供了改善的糖基化，例如增加的产量。在一个实施方案中，共有序列已被添加或取代一个或多个氨基酸残基，或替代SEQ ID NO:13的氨基酸残基K131或在与SEQ ID NO:13具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列中的等同位置(例如在SEQ ID NO:14-16的氨基酸序列中的等同位置)，例如SEQ ID No:30-39。

[0108] 本领域技术人员将理解，当H1a氨基酸序列是SEQ ID NO:2的氨基酸序列的变体和/或片段时，例如与SEQ ID NO:2具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，提及“氨基酸之间……”是指如果该序列与SEQ ID NO:1的氨基酸序列对齐以最大化两个序列之间的序列同一性，将等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:13的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异，然而，通过将突变序列与参考序列对齐，可以识别与参考序列中相应氨基酸等同位置的氨基酸，并且因此可以确定用于共有序列添加或取代的适当位置。

[0109] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即，全部或部分添加糖基化位点)，或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即，不将氨基酸添加到蛋白质中，而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到，使用本领域已知的方法，例如包括修饰编码蛋白质的核酸序列的重组方法，可以容易地修饰蛋白质的氨基酸序列。因此，在一个实施方案中，本发明提供了具有氨基酸序列的经修饰的H1a蛋白，该氨基酸序列中对应于SEQ ID NO 13的H48和

G122的氨基酸或与SEQ ID NO:13具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的H1a氨基酸序列中的等同位置的氨基酸被半胱氨酸置换，并且其中糖基化位点被重组引入到SEQ ID NO:13的H1a氨基酸序列或与SEQ ID NO:13具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的H1a氨基酸序列中。

[0110] 在一个实施方案中，本发明的经修饰的H1a蛋白质还包含“肽标签”或“标签”，即允许分离和/或鉴定经修饰的H1a蛋白质的氨基酸序列。例如，向本发明的经修饰的H1a蛋白中添加标签可用于纯化该蛋白，并且因此可用于纯化包含标记的经修饰的H1a蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中，标签是六组氨酸标签。在另一实施方案中，标签是HR标签，例如HRHR标签。在某些实施方案中，一旦不再需要标签，例如在蛋白质已经被纯化之后，在此使用的标签是可去除的，例如通过化学试剂或通过酶方式去除。任选地，肽标签位于氨基酸序列的C-末端。任选地，肽标签包含在氨基酸序列的C-末端的6个组氨酸残基。任选地，肽标签包含在氨基酸序列C-末端的四个HR残基(HRHR)。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基，例如丙氨酸、丝氨酸和/或甘氨酸残基，例如GS。此类序列的示例是SEQ ID No:36-39。

[0111] 在一个实施方案中，本发明的经修饰的H1a蛋白包含能够将载体蛋白导向宿主细胞(例如细菌)的周质的信号序列。在特定实施方案中，信号序列来自福氏志贺氏菌鞭毛蛋白(Flgl)[MIKFLSALILLLVTTAAQA(SEQ ID NO:21)]。在其它实施方案中，信号序列来自大肠杆菌外膜孔蛋白A(OmpA)[MKKTAIAIAVALAGFATVAQA(SEQ ID NO:22)]、大肠杆菌麦芽糖结合蛋白(MalE)[MKIKTGARILALSALTTMMFSASALA(SEQ ID NO:23)]、胡萝卜软腐果胶杆菌(Pectobacterium carotovorum)果胶酸裂解酶(PelB)[MKYLLPTAAAGLLLLAAQPAMA(SEQ ID NO:24)]、不耐热的大肠杆菌肠毒素LTIIb[MSFKKIIKAFVIMAALVSVQAHA(SEQ ID NO:25)]、枯草芽孢杆菌木聚糖内切酶XynA[MFKFKKKFLVGLTAAFMSISMFSATASA(SEQ ID NO:26)]、大肠杆菌DsbA[MKKIWLALAGLVLAFSASA(SEQ ID NO:27)]、TolB[MKQALRVAFGFLILWASVLHA(SEQ ID NO:28)]或无乳链球菌SipA[MKMNKKVLLTSTMAASLLSVASVQAS(SEQ ID NO:29)]。在一个实施方案中，信号序列具有与SEQ ID NO:21-29至少80%、85%、90%、92%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。在一个方面，信号序列具有与大肠杆菌鞭毛蛋白信号序列(Flgl)[MIKFLSALILLLVTTAAQA(SEQ ID NO:21)]至少80%、85%、90%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。包含信号序列的示例性修饰的H1a序列是SEQ ID NO:35-39。

[0112] 在一个实施方案中，丝氨酸和/或丙氨酸残基被添加在信号序列和成熟蛋白质序列的起点之间，例如SA或S，优选S。这样的一个或多个残基具有导致前导序列更有效切割的优势。

[0113] ClfA:来自金黄色葡萄球菌的凝集因子A

[0114] 在一个实施方案中，载体蛋白是来自葡萄球菌，特别是金黄色葡萄球菌的凝集因子A(ClfA)。ClfA已经用作金黄色葡萄球菌荚膜多糖(CP8)的载体蛋白，并且ClfA-CP8缀合物能够诱导针对ClfA和CP8两者的功能性抗体，并且在动物模型中具有保护作用。ClfA含有520个氨基酸的N-末端A结构域(纤维蛋白原结合区)，其包含三个独立折叠的亚结构域N1、N2和N3。A结构域之后是丝氨酸-天冬氨酸二肽重复区域和跨细胞壁和膜区域，其含有用于分选酶催化锚定到细胞壁的LPDTG-基序。当用作抗原或载体蛋白时，仅使用N1-N3(SEQ ID

NO:10)或N2/N3(SEQ ID No:11)结构域。

[0115] 因此,所述ClfA可以包含SEQ ID NO:10或11的氨基酸序列或与SEQ ID NO:10或11具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列。

[0116] 可进一步修饰ClfA蛋白以降低其纤维蛋白原结合活性。因此,ClfA蛋白可进一步包含选自参照SEQ ID NO:11的氨基酸序列的P116至S和Y118至A(对应于SEQ ID NO:10的序列中的P336和Y338位)或与SEQ ID NO:11具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列中的等同位置的至少一个氨基酸置换。

[0117] 因此,在本发明的一个方面,提供了包含SEQ ID NO:10、11或12的氨基酸序列或与SEQ ID NO:10、11或12具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列的经修饰的ClfA蛋白,其修饰为包含一个或多个共有序列,所述共有序列包含或由如上定义的氨基酸序列K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R(SEQ ID NO:46)组成;例如具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO 19)的一个或多个共有序列,其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸,并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中,所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中,所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选的实施方案中,所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0118] 在一个实施方案中,所述经修饰的ClfA蛋白包含多于一个所述共有序列,例如2、3、4或5个共有序列。在一个优选的实施方案中,其中存在多个共有序列,共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0119] 在一个实施方案中,本发明的经修饰的ClfA蛋白可以衍生自与SEQ ID No:10、11或12具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是SEQ ID No:10、11或12的免疫原性片段和/或变体。在一个实施方案中,本发明的经修饰的ClfA蛋白可以衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:10、11或12的免疫原性片段,其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0120] 在一个实施方案中,本发明的经修饰的ClfA蛋白可衍生自与SEQ ID No:10、11或12具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)而进行修饰的SEQ ID NO:10、11或12的变体。氨基酸置换可以是保守的或非保守的。在一个方面,氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合,只要该变体是免疫原性多肽。在一个实施方案中,本发明的经修饰的ClfA蛋白可以衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0121] 在一个实施方案中,本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测,例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK),和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0122] 术语“经修饰的ClfA蛋白”是指ClfA氨基酸序列(例如,具有SEQ ID NO:10、11或12的ClfA氨基酸序列或与SEQ ID NO:10、11或12具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的氨基酸序列),该ClfA氨基酸序列已经通过添加、取代或缺失一个或多个氨基酸进行修饰(例如,添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)。经修饰的ClfA蛋白还可以包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如,可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如,在信号序列之后或肽标签之前,如果存在的话)。在一个实施方案中,本发明的经修饰的ClfA蛋白可以是非天然存在的ClfA蛋白。

[0123] 在本发明的一个实施方案中,经修饰的ClfA氨基酸序列(例如具有SEQ ID NO:10、11或12的氨基酸序列或与SEQ ID NO:10、11或12具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的ClfA氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸,例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:19或46,例如SEQ ID NO:20或42-45或47,特别是SEQ ID NO:42-44的共有序列取代。例如,ClfA氨基酸序列(例如SEQ ID NO:10、11或12)中的单个氨基酸可以被所述共有序列取代。或者,EPA氨基酸序列(例如SEQ ID NO:10、11或12或与SEQ ID NO:10、11或12具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的ClfA氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列取代。

[0124] 引入共有序列能够使经修饰的ClfA蛋白糖基化。因此,本发明还提供了本发明的经修饰的ClfA蛋白,其中经修饰的ClfA蛋白是糖基化的。在特定的实施方案中,将共有序列引入ClfA氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0125] 本领域技术人员将理解,当ClfA氨基酸序列是SEQ ID NO:10、11或12的氨基酸序列的变体和/或片段时,例如与SEQ ID NO:10、11或12具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列,提及“氨基酸之间……”是指如果该序列与SEQ ID NO:10、11或12的氨基酸序列对齐以最大化两个序列之间的序列同一性,将等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:10、11或12的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列添加或取代的适当位置。

[0126] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0127] 在一个实施方案中,本发明的经修饰的ClfA蛋白进一步包含“肽标签”或“标签”,即允许分离和/或鉴定经修饰的ClfA蛋白的氨基酸序列。例如,向本发明的经修饰的ClfA蛋白中添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的ClfA蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标

签是六组氨酸标签。在某些实施方案中，一旦不再需要标签，例如在蛋白质已经被纯化之后，在此使用的标签是可去除的，例如通过化学试剂或通过酶方式去除。任选地，肽标签位于氨基酸序列的C-末端。任选地，肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基，例如丙氨酸、丝氨酸和/或甘氨酸残基，例如GS。CRM197：白喉毒素的非毒性突变体。

[0128] 在一个实施方案中，载体蛋白是CRM197，与白喉毒素相比具有单点突变G52E的白喉毒素的遗传脱毒突变体。CRM197是广泛使用且经过充分测试的载体蛋白，其已用于若干商业化疫苗中。DT的氨基酸序列如SEQ ID NO:4所示，CRM197的氨基酸序列如SEQ ID NO:5所示。

[0129] 因此，在本发明的一个方面，提供了经修饰的CRM197蛋白，其包含SEQ ID NO:5的氨基酸序列或与SEQ ID NO:5具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其修饰为包含一个或多个共有序列，所述共有序列包含或由如上定义的氨基酸序列K/R-(Z)$_{0-9}$-D/E-X-N-Y-S/T-(Z)$_{0-9}$-K/R(SEQ ID NO:46)组成；例如，具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个共有序列，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中，所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42-45中任何一个的氨基酸序列组成。在一个优选的实施方案中，所述共有序列包含或由SEQ ID No:42-44中任何一个的氨基酸序列组成。

[0130] 在一个实施方案中，所述经修饰的CRM197蛋白包含多于一个所述共有序列，例如2、3、4或5个共有序列。在一个优选的实施方案中，其中存在多个共有序列，所述共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0131] 在一个实施方案中，本发明的经修饰的CRM197蛋白可衍生自与SEQ ID NO:5具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是SEQ ID NO:5的免疫原性片段和/或变体。在一个实施方案中，本发明的经修饰的CRM197蛋白可衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:5的免疫原性片段，其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0132] 在一个实施方案中，本发明的经修饰的CRM197蛋白可衍生自与SEQ ID NO:5具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)进行修饰的SEQ ID NO:5的变体。氨基酸置换可以是保守的或非保守的。在一个方面，氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合，只要该变体是免疫原性多肽。在一个实施方案中，本发明的经修饰的CRM197蛋白质可衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0133] 在一个实施方案中，本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测，例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK)，和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0134] 术语“经修饰的CRM197蛋白”是指已通过添加、取代或缺失一个或多个氨基酸(例

如,添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的CRM197氨基酸序列(例如,具有SEQ ID NO:5的CRM197氨基酸序列或与SEQ ID NO:5具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列)。经修饰的CRM197蛋白还可包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如,可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如,在信号序列之后或肽标签之前,如果存在的话)。

[0135] 在本发明的一个实施方案中,经修饰的CRM197氨基酸序列(例如,具有SEQ ID NO:5的氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸,例如一个氨基酸)被具有氨基酸序列SEQ ID NO:19或46,例如SEQ ID NO:20或42-45或47,特别是SEQ ID NO:42-44的共有序列取代。例如,CRM197氨基酸序列(例如SEQ ID NO:5)中的单个氨基酸可以被所述共有序列取代。或者,CRM197氨基酸序列(例如SEQ ID NO:5)中的2、3、4、5、6或7个氨基酸可被所述共有序列取代。

[0136] 引入共有序列能够使经修饰的CRM197蛋白糖基化。因此,本发明还提供了本发明的经修饰的CRM197蛋白,其中经修饰的CRM197蛋白是糖基化的。在具体的实施方案中,将共有序列引入CRM197氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫桥稳定的环中。

[0137] 本领域技术人员将理解,当CRM197氨基酸序列是SEQ ID NO:5的氨基酸序列的变体和/或片段时,如与SEQ ID NO:5具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,提及“氨基酸之间……”是指如果该序列与SEQ ID NO:5的氨基酸序列对齐以最大化两个序列之间的序列同一性,将等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:5的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列添加或取代的适当位置。

[0138] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0139] 在一个实施方案中,本发明的经修饰的CRM197蛋白进一步包含“肽标签”或“标签”,即允许分离和/或鉴定经修饰的CRM197蛋白的氨基酸序列。例如,向本发明的经修饰的CRM197蛋白添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的CRM197蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标签是六组氨酸标签。在某些实施方案中,一旦不再需要标签,例如在蛋白质已经被纯化之后,在此使用的标签是可去除的,例如通过化学试剂或通过酶方式去除。任选地,肽标签位于氨基酸序列的C-末端。任选地,肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基,例如丙氨酸、丝氨酸和/或甘氨酸残基,例如GS。

[0140] 破伤风毒素

[0141] 由破伤风杆菌培养物产生的破伤风毒素在甲醛灭活脱毒之后广泛用作载体。显示较低毒性的TT片段也已经通过重组方式产生。

[0142] 因此，在本发明的一个方面，提供了经修饰的TT蛋白，其包含SEQ ID NO:3的氨基酸序列或与SEQ ID NO:3具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其修饰为包含一个或多个共有序列，所述共有序列包含或由如上定义的氨基酸序列K/R-(Z)$_{0-9}$-D/E-X-N-Y-S/T-(Z)$_{0-9}$-K/R(SEQ ID NO:46)组成；例如，具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个共有序列，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中，所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42-45中任何一个的氨基酸组成。在一个优选的实施方案中，所述共有序列包含或由SEQ ID No:42-44中任一个的氨基酸序列组成。

[0143] 在一个实施方案中，所述经修饰的TT蛋白包含多于一个所述共有序列，例如2、3、4或5个共有序列。在一个优选的实施方案中，其中存在多个共有序列，所述共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0144] 在一个实施方案中，本发明的经修饰的TT蛋白可以衍生自与SEQ ID NO:3具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是SEQ ID NO:3的免疫原性片段和/或变体。在一个实施方案中，本发明的经修饰的TT蛋白可衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:3的免疫原性片段，其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0145] 在一个实施方案中，本发明的经修饰的TT蛋白可以衍生自与SEQ ID NO:3具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)进行修饰的SEQ ID NO:3的变体。氨基酸置换可以是保守的或非保守的。在一个方面，氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合，只要该变体是免疫原性多肽。在一个实施方案中，本发明的经修饰的TT蛋白可衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0146] 在一个实施方案中，本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测，例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK)，和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0147] 术语“经修饰的TT蛋白”是指已经通过添加、取代或缺失一个或多个氨基酸(例如，添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列；或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的TT氨基酸序列(例如，具有SEQ ID NO:3的TT氨基酸序列或与SEQ ID NO:3具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的氨基酸序列)，该TT氨基酸序列可以是野生型成熟TT氨基酸序列(例如，SEQ ID NO:3的野生型氨基酸序列)。经修饰的TT蛋白还可以包含进一步的修饰(添加、置换、缺失)以及一个或多个共有序列的添加或取代。例如，可以添加

信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如,在信号序列之后或肽标签之前,如果存在的话)。在一个实施方案中,本发明的经修饰的TT蛋白可以是非天然存在的TT蛋白。

[0148] 在本发明的一个实施方案中,经修饰的TT氨基酸序列(例如,具有SEQ ID NO:3的氨基酸序列或与SEQ ID NO:3具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的TT氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸,例如一个氨基酸)被具有氨基酸序列SEQ ID NO:46或SEQ ID NO:19,例如SEQ ID NO:20或42-45或47,特别是SEQ ID NO:42-44的共有序列取代。例如,TT氨基酸序列(例如SEQ ID NO:3)中的单个氨基酸可以被所述共有序列替代。或者,TT氨基酸序列(例如SEQ ID NO:3或与SEQ ID NO:3具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的TT氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列替代。

[0149] 引入共有序列能够使经修饰的TT蛋白糖基化。因此,本发明还提供了本发明的经修饰的TT蛋白,其中经修饰的TT蛋白是糖基化的。在特定的实施方案中,将共有序列引入TT氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0150] 本领域技术人员将理解,当TT氨基酸序列是SEQ ID NO:3的氨基酸序列的变体和/或片段时,例如与SEQ ID NO:3具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,提及“氨基酸之间……”是指,如果该序列与SEQ ID NO:3的氨基酸序列对齐以最大化两个序列之间的序列同一性,等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:3的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列的添加或取代的适当位置。

[0151] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0152] 在一个实施方案中,本发明的经修饰的TT蛋白进一步包含“肽标签”或“标签”,即允许分离和/或鉴定经修饰的TT蛋白质的氨基酸序列。例如,向本发明的经修饰的TT蛋白中添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的TT蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标签是六组氨酸标签。在某些实施方案中,一旦不再需要标签,例如在蛋白质已经被纯化之后,在此使用的标签是可去除的,例如通过化学试剂或通过酶方式去除。任选地,肽标签位于氨基酸序列的C-末端。任选地,肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基,例如丙氨酸、丝氨酸和/或甘氨酸残基,例如GS。HCP1:来自铜绿假单胞菌的HCP1蛋白

[0153] 在一个实施方案中,载体蛋白是来自铜绿假单胞菌的Hcp1(Hcp1)。所述Hcp1可包含SEQ ID NO:6的氨基酸序列或与SEQ ID NO:6具有至少80%、85%、90%、92%、95%、

96%、97%、98%或99%同一性的氨基酸序列。

[0154] 因此，在本发明的一个方面，提供了经修饰的Hcp1蛋白，其包含SEQ ID NO:6的氨基酸序列或与SEQ ID NO:6具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列的经修饰的ClfA蛋白，其修饰为包含一个或多个共有序列，所述共有序列包含或由如上定义的氨基酸序列K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R(SEQ ID NO:46)组成；例如具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个共有序列，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中，共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选的实施方案中，所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0155] 在一个实施方案中，所述经修饰的Hcp1蛋白包含多于一个所述共有序列，例如2、3、4或5个共有序列。在一个优选的实施方案中，其中存在多个共有序列，共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0156] 在一个实施方案中，本发明的经修饰的Hcp1蛋白可以衍生自与SEQ ID NO:6具有至少至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列，其是SEQ ID No:6的免疫原性片段和/或变体。在一个实施方案中，本发明的经修饰的Hcp1蛋白可以衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:6的免疫原性片段，其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0157] 在一个实施方案中，本发明的经修饰的Hcp1蛋白可以衍生自与SEQ ID NO:6具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列，其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)而进行修饰的SEQ ID NO:6的变体。氨基酸置换可以是保守的或非保守的。在一个方面，氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合，只要该变体是免疫原性多肽。在一个实施方案中，本发明的经修饰的Hcp1蛋白可以衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0158] 在一个实施方案中，本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测，例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK)和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0159] 术语“经修饰的Hcp1蛋白”是指已经通过添加、取代或缺失一个或多个氨基酸(例如，添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列；或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的Hcp1氨基酸序列(例如，具有SEQ ID NO:6的Hcp1氨基酸序列或与SEQ ID NO:6具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列)。经修饰的Hcp1蛋白还可以包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如，可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如，在信

号序列之后或肽标签之前，如果存在的话)。在一个实施方案中，本发明的经修饰的Hcp1蛋白可以是非天然存在的Hcp1蛋白。

[0160] 在本发明的一个实施方案中，经修饰的Hcp1氨基酸序列(例如具有SEQ ID NO:6的氨基酸序列或与SEQ ID NO:6具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的Hcp1氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸，例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:19或46，例如SEQ ID NO:20或42-45或47，特别是SEQ ID NO:42-44的共有序列取代。例如，Hcp1氨基酸序列(例如SEQ ID NO:6)中的单个氨基酸可以被所述共有序列替代。或者，Hcp1氨基酸序列(例如SEQ ID NO:6或与SEQ ID NO:6具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的Hcp1氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列取代。

[0161] 引入共有序列使经修饰的Hcp1蛋白能够糖基化。因此，本发明还提供了本发明的经修饰的Hcp1蛋白，其中经修饰的Hcp1蛋白是糖基化的。在特定的实施方案中，将共有序列引入Hcp1氨基酸序列的特定区域，例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0162] 本领域技术人员将理解，当Hcp1氨基酸序列是SEQ ID NO:6的氨基酸序列的变体和/或片段时，例如与SEQ ID NO:6具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，提及“氨基酸之间……”是指，如果该序列与SEQ ID NO:6的氨基酸序列对齐以最大化两个序列之间的序列同一性，等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:6的变体和/或片段中添加或缺失氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异，然而，通过将突变序列与参考序列对齐，可以确定与参考序列中相应氨基酸等同位置的氨基酸，并且因此可以确定用于共有序列添加或取代的适当位置。

[0163] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即，全部或部分添加糖基化位点)，或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即，不将氨基酸添加到蛋白质中，而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到，使用本领域已知的方法，例如包括修饰编码蛋白质的核酸序列的重组方法，可以容易地修饰蛋白质的氨基酸序列。

[0164] 在一个实施方案中，本发明的经修饰的Hcp1蛋白还包含“肽标签”或“标签”，即允许分离和/或鉴定经修饰的Hcp1蛋白质的氨基酸序列。例如，向本发明的经修饰的Hcp1蛋白中添加标签可用于纯化该蛋白，并且因此可用于纯化包含标记的经修饰的Hcp1蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中，标签是六组氨酸标签。在某些实施方案中，一旦不再需要标签，例如在蛋白质已经被纯化之后，在此使用的标签是可去除的，例如通过化学试剂或通过酶方式去除。任选地，肽标签位于氨基酸序列的C-末端。任选地，肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基，例如丙氨酸、丝氨酸和/或甘氨酸残基，例如GS。MBP:来自大肠杆菌的麦芽糖/麦芽糊精结合蛋白。

[0165] 在一个实施方案中，载体蛋白是来自铜绿假单胞菌的外毒素A(MBP)。所述MBP可包含SEQ ID NO:8的氨基酸序列或与SEQ ID NO:8具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列。

[0166] 因此,在本发明的一个方面,提供了经修饰的MBP蛋白,其包含SEQ ID NO:8的氨基酸序列或与SEQ ID NO:8具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其修饰为包含一个或多个共有序列,所述共有序列包含或由如上定义的氨基酸序列K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R(SEQ ID NO:46)组成;例如具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个共有序列,其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸,并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中,共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中,所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选的实施方案中,所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0167] 在一个实施方案中,所述经修饰的MBP蛋白包含一个以上所述共有序列,例如2、3、4或5个共有序列。在一个优选的实施方案中,其中存在多个共有序列,共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0168] 在一个实施方案中,本发明的经修饰的MBP蛋白可衍生自与SEQ ID NO:8具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是SEQ ID NO:8的免疫原性片段和/或变体。在一个实施方案中,本发明的经修饰的MBP蛋白可衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:8的免疫原性片段,其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0169] 在一个实施方案中,本发明的经修饰的MBP蛋白可衍生自与SEQ ID NO:8具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)进行修饰的SEQ ID NO:8的变体。氨基酸置换可以是保守的或非保守的。在一个方面,氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合,只要该变体是免疫原性多肽。在一个实施方案中,本发明的经修饰的MBP蛋白质可衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0170] 在一个实施方案中,本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测,例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK),和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0171] 术语“经修饰的MBP蛋白”是指通过添加、取代或缺失一个或多个氨基酸(例如,添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的MBP氨基酸序列(例如,具有SEQ ID NO:8的MBP氨基酸序列或与SEQ ID NO:8具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列)。经修饰的MBP蛋白还可包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如,可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如,在信号序列之后或肽标签之前,如果存在的话)。在一个实施方案中,本发明的经修饰的MBP蛋白可以是非天然存在的MBP蛋白。

[0172] 在本发明的一个实施方案中,经修饰的MBP氨基酸序列(例如,具有SEQ ID NO:8的

氨基酸序列或与SEQ ID NO:8具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的MBP氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸,例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:46或SEQ ID NO:19,例如SEQ ID NO:20或42-45或47,特别是SEQ ID No:42-44的共有序列取代。例如,MBP氨基酸序列(例如SEQ ID NO:8)中的单个氨基酸可被所述共有序列替代。或者,MBP氨基酸序列(例如SEQ ID NO:8或与SEQ ID NO:8具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的MBP氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列替代。

[0173] 引入共有序列使经修饰的MBP蛋白能够糖基化。因此,本发明还提供了本发明的经修饰的MBP蛋白,其中经修饰的MBP蛋白是糖基化的。在特定的实施方案中,将共有序列导入MBP氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0174] 本领域技术人员将理解,当MBP氨基酸序列是SEQ ID NO:8的氨基酸序列的变体和/或片段时,如与SEQ ID NO:8具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列,提及“氨基酸之间……”是指,如果该序列与SEQ ID NO:8的氨基酸序列对齐以最大化两个序列之间的序列同一性,等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:8的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列添加或取代的适当位置。

[0175] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0176] 在一个实施方案中,本发明的经修饰的MBP蛋白还包含“肽标签”或“标签”,即允许分离和/或鉴定经修饰的MBP蛋白质的氨基酸序列。例如,向本发明的经修饰的MBP蛋白中添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的MBP蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标签是六组氨酸标签。在某些实施方案中,一旦不再需要标签,例如在蛋白质已经被纯化之后,在此使用的标签是可去除的,例如通过化学试剂或通过酶方式去除。任选地,肽标签位于氨基酸序列的C-末端。任选地,肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基,例如丙氨酸、丝氨酸和/或甘氨酸残基,例如GS。MtrE:来自淋病奈瑟菌的膜转运蛋白E

[0177] 在一个实施方案中,载体蛋白是来自淋病奈瑟菌的膜转运蛋白E(MtrE)。所述MtrE可包含SEQ ID NO:9的氨基酸序列或与SEQ ID NO:9具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列。

[0178] 因此,在本发明的一个方面,提供了包含SEQ ID NO:9的氨基酸序列或与SEQ ID NO:9具有至80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列的经修饰的MtrE蛋白,其修饰为包含一个或多个共有序列,所述共有序列包含或由如上定义的

氨基酸序列K/R-(Z)$_{0-9}$-D/E-X-N-Y-S/T-(Z)$_{0-9}$-K/R(SEQ ID NO:46)组成;例如具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个共有序列,其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸,并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中,所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中,所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选的实施方案中,所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0179] 在一个实施方案中,所述经修饰的MtrE蛋白包含一个以上所述共有序列,例如2、3、4或5个共有序列。在一个优选的实施方案中,其中存在多个共有序列,共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0180] 在一个实施方案中,本发明的经修饰的MtrE蛋白可以衍生自与SEQ ID NO:9具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是SEQ ID NO:9的免疫原性片段和/或变体。在一个实施方案中,本发明的经修饰的MtrE蛋白可衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:9的免疫原性片段,其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0181] 在一个实施方案中,本发明的经修饰的MtrE蛋白可以衍生自与SEQ ID NO:9具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)进行修饰的SEQ ID NO:9的变体。氨基酸置换可以是保守的或非保守的。在一个方面,氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合,只要该变体是免疫原性多肽。在一个实施方案中,本发明的经修饰的MtrE蛋白质可衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0182] 在一个实施方案中,本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测,例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK),和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0183] 术语“经修饰的MtrE蛋白”是指通过添加、取代或缺失一个或多个氨基酸(例如,添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列;或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的MtrE氨基酸序列(例如,具有SEQ ID NO:9的MtrE氨基酸序列或与SEQ ID NO:9具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列),该MtrE氨基酸序列可以是野生型成熟MtrE氨基酸序列(例如,SEQ ID NO:9的野生型氨基酸序列)。经修饰的MtrE蛋白还可以包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如,可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如,在信号序列之后或肽标签之前,如果存在的话)。在一个实施方案中,本发明的经修饰的MtrE蛋白可以是非天然存在的MtrE蛋白。

[0184] 在本发明的一个实施方案中,经修饰的MtrE氨基酸序列(例如,具有SEQ ID NO:9的氨基酸序列或与SEQ ID NO:9具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的MtrE氨基酸序列)的一个或多个氨基酸(例如1-7个氨基

酸,例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:46或SEQ ID NO:19,例如SEQ ID NO:20或42-45或47,特别是SEQ ID NO:42-44的共有序列取代。例如,MtrE氨基酸序列(例如SEQ ID NO:9)中的单个氨基酸可以被所述共有序列替代。或者,MtrE氨基酸序列(例如SEQ ID NO:9或与SEQ ID NO:9具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的MtrE氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列替代。

[0185] 引入共有序列使经修饰的MtrE蛋白能够糖基化。因此,本发明还提供了本发明的经修饰的MtrE蛋白,其中经修饰的MtrE蛋白是糖基化的。在特定的实施方案中,将共有序列引入MtrE氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0186] 本领域技术人员将理解,当MtrE氨基酸序列是SEQ ID NO:9的氨基酸序列的变体和/或片段时,例如与SEQ ID NO:9具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,提及"氨基酸之间……"是指,如果该序列与SEQ ID NO:9的氨基酸序列对齐以最大化两个序列之间的序列同一性,等同于定义位置的位置。序列比对工具描述如上。从SEQ ID NO:9的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列添加或取代的适当位置。

[0187] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0188] 在一个实施方案中,本发明的经修饰的MtrE蛋白还包含"肽标签"或"标签",即允许分离和/或鉴定经修饰的MtrE蛋白质的氨基酸序列。例如,向本发明的经修饰的MtrE蛋白中添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的MtrE蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标签是六组氨酸标签。在某些实施方案中,一旦不再需要标签,例如在蛋白质已经被纯化之后,在此使用的标签是可去除的,例如通过化学试剂或通过酶方式去除。任选地,肽标签位于氨基酸序列的C-末端。任选地,肽标签包含在氨基酸序列C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基,例如丙氨酸、丝氨酸和/或甘氨酸残基,例如GS。PspA,来自大肠杆菌的噬菌体休克蛋白A

[0189] 在一个实施方案中,载体蛋白是来自铜绿假单胞菌的噬菌体休克蛋白A(PspA)。所述PspA可包含SEQ ID NO:7的氨基酸序列或与SEQ ID NO:7具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列。

[0190] 因此,在本发明的一个方面,提供了经修饰的PspA蛋白,其包含SEQ ID NO:7的氨基酸序列或与SEQ ID NO:7具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列的经修饰的PspA蛋白,其经修饰为包含一个或多个共有序列,所述共有序列包含或由如上定义的氨基酸序列K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R(SEQ ID NO:46)组成;例如具有氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)的一个或多个

共有序列，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中Z_1和Z_2不是赖氨酸或精氨酸。在一个实施方案中，共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中，所述共有序列包含或由SEQ ID No:42-45的任何一个的氨基酸序列组成。在一个优选的实施方案中，所述共有序列包含或由SEQ ID No:42-44的任何一个的氨基酸序列组成。

[0191] 在一个实施方案中，所述经修饰的PspA蛋白包含一个以上所述共有序列，例如2、3、4或5个共有序列。在一个优选的实施方案中，其中存在多个共有序列，共有序列具有不同的序列以便可以区分每个单独位点的糖基化。

[0192] 在一个实施方案中，本发明的经修饰的PspA蛋白可以衍生自与SEQ ID NO:7具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是SEQ ID No:7的免疫原性片段和/或变体。在一个实施方案中，本发明的经修饰的PspA蛋白可以衍生自包含全长序列的至少约15、至少约20、至少约40或至少约60个连续氨基酸残基的SEQ ID NO:7的免疫原性片段，其中所述多肽能够引发对所述氨基酸序列特异性的免疫应答。

[0193] 在一个实施方案中，本发明的经修饰的PspA蛋白可衍生自与SEQ ID NO:7具有至少80％、85％、90％、92％、95％、96％、97％、98％或99％同一性的氨基酸序列，其是已通过缺失和/或添加和/或置换一个或多个氨基酸(例如1、2、3、4、5、6、7、8、9、10、11或12个氨基酸)而进行修饰的SEQ ID NO:7的变体。氨基酸置换可以是保守的或非保守的。在一个方面，氨基酸置换是保守的。置换、缺失、添加或其任何组合可以在单个变体中组合，只要该变体是免疫原性多肽。在一个实施方案中，本发明的经修饰的PspA蛋白可以衍生自其中1至10、5至10、1至5、1至3、1至2或1个氨基酸以任何组合被置换、缺失或添加的变体。

[0194] 在一个实施方案中，本发明包括含有B细胞或T细胞表位的片段和/或变体。这样的表位可以使用2D结构预测，例如使用PSIPRED程序(来自David Jones,Brunei Bioinformatics Group,Dept.Biological Sciences,Brunei University,Uxbridge UB8 3PH,UK)，和基于Jameson和Wolf描述的方法(CABIOS 4:181-186[1988])计算的抗原指数的组合进行预测。

[0195] 术语“经修饰的PspA蛋白”是指通过添加、取代或缺失一个或多个氨基酸(例如，添加(插入)具有氨基酸序列SEQ ID NO:19或46的共有序列；或通过具有氨基酸序列SEQ ID NO:19或46的共有序列取代一个或多个氨基酸)进行修饰的PspA氨基酸序列(例如，具有SEQ ID NO:7的PspA氨基酸序列或与SEQ ID NO:7具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的氨基酸序列)。经修饰的PspA蛋白还可以包含进一步的修饰(添加、置换、缺失)以及添加或取代一个或多个共有序列。例如，可以添加信号序列和/或肽标签。可以包括N和/或C末端处的另外的氨基酸以帮助克隆(例如，在信号序列之后或肽标签之前，如果存在的话)。在一个实施方案中，本发明的经修饰的PspA蛋白可以是非天然存在的PspA蛋白。

[0196] 在本发明的一个实施方案中，经修饰的PspA氨基酸序列(例如，具有SEQ ID NO:7的氨基酸序列或与SEQ ID NO:7具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％同一性的PspA氨基酸序列)的一个或多个氨基酸(例如1-7个氨基酸，例如一个氨基酸)已经被具有氨基酸序列SEQ ID NO:19或46，例如SEQ ID NO:20或42-45或47，特别是SEQ ID NO:42-44的共有序列取代。例如，PspA氨基酸序列(例如SEQ ID NO:

7)中的单个氨基酸可以被所述共有序列替代。或者,PspA氨基酸序列(例如SEQ ID NO:7或与SEQ ID NO:7具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的PspA氨基酸序列)中的2、3、4、5、6或7个氨基酸可被所述共有序列取代。

[0197] 引入共有序列使经修饰的PspA蛋白能够糖基化。因此,本发明还提供了本发明的经修饰的PspA蛋白,其中经修饰的PspA蛋白是糖基化的。在特定的实施方案中,将共有序列引入PspA氨基酸序列的特定区域,例如蛋白质的表面结构、蛋白质的N或C末端和/或通过二硫键稳定的环中。

[0198] 本领域技术人员将理解,当PspA氨基酸序列是SEQ ID NO:7的氨基酸序列的变体和/或片段时,如与SEQ ID NO:7具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,提及"氨基酸之间……"是指,如果该序列与SEQ ID NO:7的氨基酸序列对齐以最大化两个序列之间的序列同一性,等同于定义位置的位置。序列比对工具描述如上。在SEQ ID NO:7的变体和/或片段中添加或删除氨基酸可导致突变序列中共有序列的实际氨基酸位置的差异,然而,通过将突变序列与参考序列对齐,可以确定与参考序列中相应氨基酸等同位置的氨基酸,并且因此可以确定用于共有序列添加或取代的适当位置。

[0199] 这种糖基化位点的引入可以通过例如向蛋白质的一级结构中添加新的氨基酸(即,全部或部分添加糖基化位点),或通过突变蛋白质中存在的氨基酸以产生糖基化位点(即,不将氨基酸添加到蛋白质中,而是将蛋白质的选定氨基酸突变以形成糖基化位点)来实现。本领域技术人员将认识到,使用本领域已知的方法,例如包括修饰编码蛋白质的核酸序列的重组方法,可以容易地修饰蛋白质的氨基酸序列。

[0200] 在一个实施方案中,本发明的经修饰的PspA蛋白还包含肽标签"或"标签",即允许分离和/或鉴定经修饰的PspA蛋白质的氨基酸序列。例如,向本发明的经修饰的PspA蛋白中添加标签可用于纯化该蛋白,并且因此可用于纯化包含标记的经修饰的TT蛋白的结合疫苗。可用于本文的示例性标签包括但不限于组氨酸(HIS)标签。在一个实施方案中,标签是六组氨酸标签。在某些实施方案中,一旦不再需要标签,例如在蛋白质已经被纯化之后,在此使用的标签是可去除的,例如通过化学试剂或通过酶方式去除。任选地,肽标签位于氨基酸序列的C-末端。任选地,肽标签包含在氨基酸序列的C-末端的6个组氨酸残基。肽标签可以包含或在其前面有一个、两个或更多个另外的氨基酸残基,例如丙氨酸、丝氨酸和/或甘氨酸残基,例如GS。

[0201] 信号序列和其它修饰

[0202] 在一个实施方案中,本发明的经修饰的载体蛋白包含能够将所述蛋白导向宿主细胞(例如细菌)的周质的信号序列。在特定实施方案中,信号序列来自福氏志贺氏菌鞭毛蛋白(Flgl)[MIKFLSALILLLVTTAAQA(SEQ ID NO:21)]。在其它实施方案中,信号序列来自大肠杆菌外膜孔蛋白A(OmpA)[MKKTA1A1AVALAGFATVAQA(SEQ ID NO:22)]、大肠杆菌麦芽糖结合蛋白(MalE)[MKIKTGARILALSALTTMMFSASALA(SEQ ID NO:23)]、胡萝卜软腐果胶杆菌果胶酸裂解酶(PelB)[MKYLLPTAAAGLLLLAAQPAMA(SEQ ID NO:24]、不耐热的大肠杆菌肠毒素LTIIb[MSFKKIIKAFVIMAALVSVQAHA(SEQ ID NO:25)]、枯草芽孢杆菌木聚糖内切酶XynA[MFKFKKKFLVGLTAAFMSISMFSATASA(SEQ ID NO:26)]、大肠杆菌DsbA[MKKI WLALAG LVLAFSASA(SEQ ID NO:27)]、TolB[MKQALRVAFGFLILWASVLHA(SEQ ID NO:28)]或无乳链球

菌SipA[MKMNKKVLLTSTMAASLLSVASVQAS(SEQ ID NO:29)]。在一个实施方案中,信号序列具有与SEQ ID NO:21-29至少80%、85%、90%、92%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。在一个方面,信号序列具有与大肠杆菌鞭毛蛋白信号序列(Flg1)[MIKFLSALILLLVTTAAQA(SEQ ID NO:21)]至少80%、85%、90%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。

[0203] 在一个实施方案中,丝氨酸和/或丙氨酸残基被添加在信号序列和成熟蛋白序列的起点之间,例如SA或S,优选S。这样的一个或多个残基具有导致更有效的前导序列切割的优势。

[0204] 糖基化位点

[0205] 本发明为糖基化位点提供了的新的通用Pgl B特异性共有序列,其与通过LC-MS对糖基化位点占有率的定量相容。本发明人确定了这些序列以及本发明的共有序列所共有的几个特征:

[0206] 产生长度为8和16个之间,例如8、9、10、11、12、13、14、15或16个氨基酸的胰蛋白酶肽。

[0207] 在质谱分析中显示强且可再现的信号(母离子和过渡离子);

[0208] 开始和终止于精氨酸或赖氨酸(用于胰蛋白酶切割)且不含半胱氨酸(以增加胰蛋白酶肽的离子化能力);

[0209] 优选不含对修饰敏感的氨基酸(对脱氨反应敏感的天冬酰胺和谷氨酰胺氨基酸残基,或对氧化敏感的甲硫氨酸、半胱氨酸和色氨酸氨基酸残基,或疏水或芳族氨基酸);

[0210] 定位在蛋白质表面的充分暴露的环上,以便也在至少部分折叠的分子中寡糖转移酶(Pgl B)可及,并且不干扰正常的折叠过程。

[0211] 因此,本发明提供了包含或由以下氨基酸序列组成的共有序列:

[0212] $K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R$

[0213] 其中X和Y独立地是除了脯氨酸之外的任何氨基酸,并且Z表示任何氨基酸。在一个优选的实施方案中,X和Y独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸。在一个实施方案中,Z表示除赖氨酸或精氨酸以外的任何氨基酸。在一个实施方案中,X、Y和/或Z不是芳香族或疏水性氨基酸。在一个优选的实施方案中,Z表示除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸(例如SEQ ID NO:47)。

[0214] 优选地,所述共有序列的总长度为16个或更少的氨基酸,例如8、10、11、12、13、14、15或16个氨基酸。优选地,序列的总长度为8个或更多个氨基酸,例如8、10、11、12、13、14、15或16个氨基酸。

[0215] 在一个具体的实施方案中,本发明提供了包含氨基酸序列$K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R$(SEQ ID NO:19)或由其组成的共有序列,其中X_1和X_2独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸,并且其中Z_1和Z_2不是赖氨酸或精氨酸或半胱氨酸。在一个实施方案中,所述共有序列包含或由SEQ ID NO:20的氨基酸序列组成。在一个实施方案中,所述共有序列包含或由SEQ ID No:42、SEQ ID No:43;SEQ ID No:44或SEQ ID No:45,优选SEQ ID No:42-44中的任何一个的氨基酸序列组成。

[0216] 多糖

[0217] 在一个实施方案中,本发明的缀合物(例如生物缀合物)中的抗原之一是糖,如细

菌荚膜糖、细菌脂多糖或细菌寡糖。在一个实施方案中，抗原是细菌荚膜糖。

[0218] 糖可以选自：金黄色葡萄球菌5型荚膜糖、金黄色葡萄球菌8型荚膜糖、脑膜炎奈瑟菌血清群A荚膜糖(MenA)、脑膜炎奈瑟菌血清群C荚膜糖(MenC)、脑膜炎奈瑟菌血清群Y荚膜糖(MenY)、脑膜炎奈瑟菌血清群W荚膜糖(MenW)、流感嗜血杆菌b型荚膜糖(Hib)、B组链球菌I组荚膜糖、B组链球菌II组荚膜糖、B组链球菌III组荚膜糖、B组链球菌IV组荚膜糖、B组链球菌V组荚膜糖、来自伤寒沙门氏菌的Vi糖、脑膜炎奈瑟菌LPS(如L3和/或L2)、卡他拉莫菌LPS、流感嗜血杆菌LPS、志贺氏菌O-抗原、铜绿假单胞菌O-抗原、大肠杆菌O-抗原或肺炎链球菌荚膜多糖。

[0219] 在一个实施方案中，抗原是多糖或寡糖。在一个实施方案中，抗原包含两个或更多个单糖，例如2、3、4、5、6、7、8、9、10或更多个单糖。在一个实施方案中，抗原是含有不超过20、15、12、10、9或8个单糖的寡糖。在一个实施方案中，抗原是含有不超过500、400、300、200、100、90、80、70、60、50、40、30、20、10或5个单糖的寡糖。

[0220] 宿主细胞

[0221] 本发明还提供了宿主细胞，其包含：

[0222] i)一种或多种编码糖基转移酶的核酸；

[0223] ii)编码寡糖基转移酶的核酸；

[0224] iii)编码本发明的经修饰的载体蛋白的核酸；和任选地

[0225] iv)编码聚合酶(例如wzy)的核酸。

[0226] 可用于产生本发明的生物缀合物的宿主细胞包括古细菌、原核宿主细胞和真核宿主细胞。用于生产本发明的生物缀合物的示例性原核宿主细胞是但不限于埃希氏菌属(Escherichia)物种、志贺氏菌属(Shigella)物种、克雷伯氏菌属(Klebsiella)物种、黄单胞菌属(Xhantomonas)物种、沙门氏菌属(Salmonella)物种、耶尔森菌属(Yersinia)物种、乳球菌属(Lactococcus)物种、乳杆菌属(Lactobacillus)物种、假单胞菌属(Pseudomonas)物种、棒杆菌属(Corynebacterium)物种、链霉菌属(Streptomyces)物种、链球菌属(Streptococcus)物种、葡萄球菌属(Staphylococcus)物种、芽孢杆菌属(Bacillus)物种和梭菌属(Clostridium)物种。在一个具体的实施方案中，所述宿主细胞是大肠杆菌。

[0227] 在一个实施方案中，用于产生本发明的生物缀合物的宿主细胞被工程化以包含异源核酸，例如编码一种或多种载体蛋白的异源核酸和/或编码一种或多种蛋白质的异源核酸，例如编码一种或多种蛋白质的基因。在特定的实施方案中，可以将编码参与糖基化途径(例如原核和/或真核糖基化途径)的蛋白质的异源核酸引入本发明的宿主细胞中。这样的核酸可以编码蛋白质，包括但不限于寡糖基转移酶、差向异构酶、翻转酶、聚合酶和/或糖基转移酶。异源核酸(例如编码载体蛋白的核酸和/或编码其它蛋白(例如参与糖基化的蛋白)的核酸)可使用诸如电穿孔、通过热休克的化学转化、自然转化、噬菌体转导和缀合的方法引入本发明的宿主细胞中。在具体的实施方案中，使用质粒将异源核酸引入本发明的宿主细胞中，例如异源核酸通过质粒(例如表达载体)在宿主细胞中表达。在另一个具体的实施方案中，使用国际专利申请号PCT/EP2013/068737(公开为WO14/037585)中描述的插入方法将异源核酸引入本发明的宿主细胞中。

[0228] 因此，本发明还提供了一种宿主细胞，其包含：

[0229] i)一种或多种编码糖基转移酶的核酸；

[0230] ii)编码寡糖基转移酶的核酸；

[0231] iii)编码本发明的经修饰的载体蛋白的核酸；

[0232] iv)编码聚合酶(例如wzy)的核酸和编码翻转酶(例如wxy)的核酸。

[0233] 在一个实施方案中，可以将另外的修饰引入(例如使用重组技术)到本发明的宿主细胞中。例如，编码形成可能竞争或干扰糖基化途径的部分的蛋白质(例如，竞争或干扰重组引入宿主细胞的参与糖基化的一个或多个异源基因)的宿主细胞核酸(例如，基因)可以在宿主细胞背景(基因组)中以使其失活/功能失调的方式删除或修饰(即缺失/修饰的宿主细胞核酸不编码功能性蛋白质或不编码任何蛋白质)。在一个实施方案中，当核酸从本发明的宿主细胞的基因组中删除时，它们被所需的序列，例如用于糖蛋白生产的序列所取代。

[0234] 可以在宿主细胞中缺失(并且，在一些情况下，用其他所需的核酸序列取代)的示例性基因包括参与糖脂生物合成的宿主细胞基因，例如waaL(参见，例如Feldman等人，2005,PNAS USA 102:3016-3021)、脂质A核心生物合成簇(waa)、半乳糖簇(gal)、阿拉伯糖簇(ara)、荚膜异多糖酸(colonic acid)簇(wc)、荚膜多糖簇、十一异戊烯醇-焦磷酸生物合成基因(例如uppS(十一异戊烯基焦磷酸合酶)、uppP(十一异戊烯基二磷酸酶))、Und-p再循环基因、参与核苷酸活化的糖生物合成的代谢酶、肠细菌共同抗原簇和原噬菌体O抗原修饰簇如gtrABS簇。

[0235] 这种修饰的原核宿主细胞包含编码酶的核酸，所述酶能够产生包含抗原的生物缀合物，例如与本发明的经修饰的Hla载体蛋白连接的糖抗原。这样的宿主细胞可以天然表达对糖抗原的产生特异性的核酸，或者可以使宿主细胞表达这样的核酸，即在某些实施方案中，所述核酸对于宿主细胞是异源的。在某些实施方案中，一种或多种对糖抗原产生特异性的所述核酸对宿主细胞是异源的，并整合到宿主细胞的基因组中。在某些实施方案中，本发明的宿主细胞包含编码在蛋白质的N-糖基化中有活性的其它酶的核酸，例如本发明的宿主细胞还包含编码寡糖基转移酶的核酸和/或编码其它糖基转移酶的一种或多种核酸。

[0236] 可以将包含荚膜多糖基因簇的核酸序列插入到本发明的宿主细胞中。在一个具体的实施方案中，插入本发明宿主细胞中的荚膜多糖基因簇是来自大肠杆菌菌株、葡萄球菌(Staphylococcus)菌株(例如金黄色葡萄球菌)、链球菌(Streptococcus)菌株(例如肺炎链球菌、化脓性链球菌(S.pyogenes)、无乳链球菌(S.agalacticae)或伯克霍尔德氏菌(Burkholderia)菌株(例如鼻疽伯克霍尔德菌(B.mallei)、假鼻疽伯克霍尔德菌(B.pseudomallei)、B.thailandensis)的荚膜多糖基因簇。用于制备能够产生生物缀合物的此类宿主细胞的方法公开于WO06/119987、WO09/104074、WO11/62615、WO11/138361、WO14/57109、WO14/72405和WO16/20499中。

[0237] 在一个实施方案中，宿主细胞包含在宿主细胞的质粒中的编码本发明的修饰的载体蛋白的核酸。

[0238] 糖基化机制

[0239] 本发明的宿主细胞包含和/或可以被修饰以包含编码能够产生杂合寡糖和/或多糖的遗传机制(例如糖基转移酶、翻转酶、聚合酶和/或寡糖转移酶)以及能够将抗原与本发明的经修饰的载体蛋白连接的遗传机制的核酸。

[0240] 荚膜多糖通过保守途径装配在细菌膜载体脂质十一异戊烯基焦磷酸上，所述保守途径与革兰氏阴性细菌中O型多糖合成的聚合酶依赖性途径具有同源性。通过将磷酸糖从

DP-供体转移到十一异戊烯基磷酸来启动抗原组装。脂质连接的O抗原通过不同糖基转移酶的顺序作用在内膜的细胞质侧组装。然后将糖脂翻转到周质空间并聚合。通过用寡糖基转移酶PglB替代O抗原连接酶WaaL,聚合的O抗原可以被转移到蛋白质载体而不是脂质A核心。

[0241] 糖基转移酶

[0242] 本发明的宿主细胞包含编码产生寡糖或多糖重复单位的糖基转移酶的核酸。在一个实施方案中,所述重复单元不包含还原末端的己糖,并且所述寡糖或多糖重复单元衍生自在还原末端包含己糖的供体寡糖或多糖重复单元。

[0243] 在一个实施方案中,本发明的宿主细胞可以包含编码将己糖单糖衍生物组装到十一异戊烯基焦磷酸(Und-PP)上的糖基转移酶的核酸。在一个方面,将己糖单糖衍生物组装到Und-PP上的糖基转移酶对于宿主细胞是异源的和/或对于编码糖基转移酶的一个或多个基因是异源的。所述糖基转移酶可以源自例如埃希氏菌属物种,志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森菌属物种、气单胞菌属(Aeromonas)物种、弗朗西斯菌属(Francisella)物种、螺杆菌属(Helicobacter)物种、变形杆菌属(Proteus)物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特氏菌属(Listeria)物种或弯曲杆菌属物种。在一个具体的实施方案中,将己糖单糖衍生物组装到Und-PP上的糖基转移酶是wecA,任选地来自大肠杆菌(wecA可以将GlcNAc从UDP-GlcNAc组装到UndP上)。在一个实施方案中,己糖单糖选自葡萄糖、半乳糖、鼠李糖、阿拉伯糖醇、岩藻糖和甘露糖(例如半乳糖)。

[0244] 在一个实施方案中,本发明的宿主细胞可以包含编码一种或多种能够将单糖添加到组装在Und-PP上的己糖单糖衍生物的糖基转移酶的核酸。在一个具体的实施方案中,所述一种或多种能够将单糖添加到己糖单糖衍生物的糖基转移酶是来自鲍氏志贺氏菌(Shigella boyedii)的半乳糖基转移酶(wfeD)。在另一个具体的实施方案中,所述一种或多种能够将单糖添加至己糖单糖衍生物的糖基转移酶是来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)。在另一个具体的实施方案中,所述一种或多种能够将单糖添加到己糖单糖衍生物的糖基转移酶是来自大肠杆菌O167的呋喃半乳糖基转移酶(wfdK)。Galf-转移酶,诸如wfdK和wbeY,可以将Galf(呋喃半乳糖)从UDP-Galf转移至-GlcNAc-P-P-十一异戊烯基。在另一个具体的实施方案中,所述一种或多种能够将单糖添加到己糖单糖衍生物的糖基转移酶是来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)和来自大肠杆菌O167的呋喃半乳糖基转移酶(wfdK)。

[0245] 在一个实施方案中,本发明的宿主细胞包含编码将供体寡糖或多糖重复单元组装到己糖单糖衍生物上的糖基转移酶的核酸。

[0246] 在一个实施方案中,将供体寡糖或多糖重复单元组装到己糖单糖衍生物上的糖基转移酶包括能够将存在于供体寡糖或多糖的第一重复单元的还原末端的己糖单糖添加到己糖单糖衍生物的糖基转移酶。示例性的糖基转移酶包括半乳糖基转移酶(wciP),例如来自大肠杆菌O21的wciP。

[0247] 在一个实施方案中,将供体寡糖或多糖重复单元组装到己糖单糖衍生物上的糖基转移酶包括能够将与存在于供体寡糖或多糖的第一重复单元的还原末端的己糖单糖相邻的单糖添加到存在于供体寡糖或多糖的第一重复单元的还原末端的己糖单糖的糖基转移

酶。示例性的糖基转移酶包括葡糖基转移酶(wciQ),例如来自大肠杆菌O21的wciQ。

[0248] 在一个实施方案中,本发明的宿主包括选自金黄色葡萄球菌CP5或CP8基因簇的用于合成寡糖或多糖的重复单元的糖基转移酶。在一个具体的实施方案中,用于合成寡糖或多糖的重复单元的糖基转移酶来自金黄色葡萄球菌CP5基因簇。金黄色葡萄球菌CP5和CP8具有与铜绿假单胞菌O11抗原合成基因类似的结构,因此这些基因可与大肠杆菌单糖合成基因组合以合成由重复三糖单元组成的十一异戊烯基焦磷酸连接的CP5或CP8聚合物。

[0249] 在一个实施方案中,本发明的宿主包含将供体寡糖或多糖重复单元组装到己糖单糖衍生物上的糖基转移酶,包括能够将存在于供体寡糖或多糖的第一重复单元的还原末端的己糖单糖添加到己糖单糖衍生物的糖基转移酶。

[0250] 寡糖基转移酶

[0251] N-连接蛋白糖基化-将碳水化合物分子添加到靶蛋白多肽链的天冬酰胺残基上-是真核生物内质网中发生的翻译后修饰的最常见类型。该过程通过酶促寡糖转移酶复合物(OST)完成,该复合物负责将预组装的寡糖从脂质载体(多萜醇磷酸酯)转移到内质网中保守序列Asn-X-Ser/Thr(其中X是除脯氨酸外的任何氨基酸)内的新生蛋白的天冬酰胺残基。

[0252] 已经显示,细菌(食物携带的病原体空肠弯曲杆菌)由于其拥有其自身的糖基化机制,也可以使其蛋白质N-糖基化(Wacker等人,Science.2002;298(5599):1790-3)。负责该反应的机制由称为“pgl”(用于蛋白质糖基化)的簇编码。

[0253] 空肠杆菌糖基化机制可转移至大肠杆菌以允许由大肠杆菌细胞表达的重组蛋白的糖基化。先前的研究已经表明如何产生可以进行N-糖基化的大肠杆菌菌株(参见,例如Wacker等人,Science.2002;298(5599):1790-3;Nita-Lazar等人.Glycobiology.2005;15(4):361-7;Feldman等人.Proc Natl Acad Sci U S A.2005;102(8):3016-21;Kowarik等人.EMBO J.2006;25(9):1957-66;Wacker等人.Proc Natl Acad Sci U S A.2006;103(18)7088-93;国际专利申请公开No.WO2003/074687、WO2006/119987、WO2009/104074和WO/2011/06261和WO2011/138361)。在WO2016/107818中描述了具有最佳性质的PglB突变体。优选的突变体是$PglB_{cuo\ N311V-K482R-D483H-A669V}$。

[0254] 寡糖基转移酶将脂质连接的寡糖转移至包含N-糖基化共有基序的新生多肽链的天冬酰胺残基上,例如Asn-X-Ser(Thr),其中X可为除Pro以外的任何氨基酸;或Asp(Glu)-X-Asn-Z-Ser(Thr),其中X和Z独立地选自除了Pro之外的任何天然氨基酸(参见WO2006/119987)。参见例如WO2003/074687和WO2006/19987,其公开内容通过引用全文并入本文。

[0255] 在一个实施方案中,本发明的宿主细胞包含编码寡糖基转移酶的核酸。如上所述,编码寡糖基转移酶的核酸可以是宿主细胞天然的或可以使用遗传方法引入宿主细胞中。在一个具体的实施方案中,寡糖基转移酶是来自弯曲杆菌的寡糖基转移酶。在另一个具体的实施方案中,寡糖基转移酶是来自空肠弯曲杆菌的寡糖基转移酶(即pglB;参见例如Wacker等人.2002,Science 298:1790-1793;还参见,例如NCBI Gene ID:3231775,UniProt Accession No.O86154)。在另一个具体的实施方案中,寡糖基转移酶是来自红嘴鸥弯曲杆菌(Campylobacter lari)的寡糖基转移酶(参见,例如NCBI Gene ID:7410986)。

[0256] 在一个具体的实施方案中,本发明的宿主细胞包含编码寡糖基转移酶的核酸序列,其中所述编码寡糖基转移酶(例如来自空肠弯曲杆菌的pglB)的核酸序列整合到宿主细胞的基因组中。

[0257] 在一个具体的实施方案中，本发明的宿主细胞包含编码寡糖基转移酶的核酸序列，其中所述编码寡糖基转移酶(例如来自空肠弯曲杆菌的pglB)的核酸序列是质粒携带的。

[0258] 在另一个具体的实施方案中，本文提供了经修饰的原核宿主细胞，其包含(i)衍生自来自金黄色葡萄球菌的荚膜多糖簇的糖基转移酶，其中所述糖基转移酶整合到所述宿主细胞的基因组中；(ii)编码寡糖基转移酶(例如来自空肠弯曲杆菌的pglB)的核酸，其中所述编码寡糖基转移酶的核酸是质粒携带的和/或整合到宿主细胞的基因组中；和(iii)本发明的修饰的载体蛋白，其中所述修饰的载体蛋白是质粒携带的或整合到宿主细胞的基因组中。还提供了制备经修饰的原核宿主细胞的方法，其包括(i)将衍生自荚膜多糖簇的糖基转移酶整合到所述宿主细胞的基因组中；(ii)将一种或多种编码寡糖基转移酶(例如来自空肠弯曲杆菌的pglB)的核酸整合到宿主细胞中，其是质粒携带的和/或整合到宿主细胞的基因组中；和(iii)将本发明的修饰的载体蛋白整合到宿主细胞中，其是质粒携带的或整合到宿主细胞的基因组中。

[0259] 在本发明的宿主细胞的具体实施方案中，其中所述宿主细胞的至少一个基因已功能失活或缺失，任选地其中所述宿主细胞的waaL基因已功能失活或缺失，任选地其中所述宿主细胞的waaL基因已被编码寡糖转移酶的核酸取代，任选地其中所述宿主细胞的waaL基因已被空肠杆菌pglB取代。

[0260] 聚合酶

[0261] 在一个实施方案中，将聚合酶(例如wzy)引入本发明的宿主细胞中(即聚合酶对宿主细胞是异源的)。在一个实施方案中，聚合酶是细菌聚合酶。在一个实施方案中，聚合酶是荚膜多糖聚合酶(例如wzy)或O抗原聚合酶(例如wzy)。在一个实施方案中，聚合酶是荚膜多糖聚合酶(例如wzy)。

[0262] 在一个实施方案中，将荚膜多糖生物合成途径的聚合酶引入本发明的宿主细胞中。

[0263] 翻转酶

[0264] 在一个实施方案中，将翻转酶(wzx或同源物)引入本发明的宿主细胞中(即翻转酶与宿主细胞异源)。因此，本发明的宿主细胞可以进一步包含翻转酶。在一个实施方案中，翻转酶是细菌翻转酶。翻转酶将野生型重复单元和/或它们相应的工程化(杂化)重复单元从细胞质移位到宿主细胞(例如大肠杆菌)的周质(periplam)中。因此，本发明的宿主细胞可以包含编码翻转酶(wzx)的核酸。

[0265] 在一个具体的实施方案中，将荚膜多糖生物合成途径的翻转酶引入本发明的宿主细胞中。

[0266] 遗传背景

[0267] 可用于产生本发明宿主细胞的示例性宿主细胞包括但不限于大肠杆菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森菌属物种、乳球菌属物种、乳杆菌属物种，假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、葡萄球菌属物种、芽孢杆菌属物种和梭菌属物种。在一个具体的实施方案中，本文使用的宿主细胞是大肠杆菌。

[0268] 在一个实施方案中，宿主细胞遗传背景通过例如缺失一个或多个基因改变。可以

在宿主细胞中缺失(并且,在一些情况下,用其他所需的核酸序列替代)的示例性基因包括参与糖脂生物合成的宿主细胞基因,例如waaL(参见,例如Feldman等人,2005,PNAS USA102:3016-3021)、O抗原簇(rfb或wb)、肠细菌共同抗原簇(wec)、脂质A核心生物合成簇(waa)和原噬菌体O抗原修饰簇如gtrABS簇。在一个具体的实施方案中,waaL基因、gtrA基因、gtrB基因、gtrS基因的一个或多个或者来自wec簇的一个或多个基因或来自rfb基因簇的一个或多个基因从本发明的原核宿主细胞的基因组中缺失或功能失活。在一个实施方案中,本文所用的宿主细胞是大肠杆菌,其中waaL基因、gtrA基因、gtrB基因、gtrS基因从宿主细胞的基因组中缺失或功能失活。在另一个实施方案中,本文所用的宿主细胞是大肠杆菌,其中waaL基因和gtrS基因从宿主细胞的基因组中缺失或功能失活。在另一个实施方案中,本文所用的宿主细胞是大肠杆菌,其中waaL基因和来自wec簇的基因从宿主细胞的基因组中缺失或功能失活。

[0269] 生物缀合物

[0270] 本发明的宿主细胞可用于产生包含糖抗原的生物缀合物,例如与本发明的修饰的载体蛋白连接的细菌荚膜多糖抗原。在一个实施方案中,多糖与修饰的载体蛋白中的天冬酰胺连接,例如通过N-乙酰葡糖胺。使用宿主细胞产生生物缀合物的方法描述于例如WO2003/074687、WO2006/19987和WO2011/138361中。如本文所述的生物缀合物与抗原-载体蛋白的化学缀合物相比具有有利的性质,因为它们在制造中需要更少的化学品并且就产生的最终产物而言更一致。

[0271] 在一个实施方案中,本文提供了包含与多糖(特别是多糖抗原)连接的本发明的修饰的载体蛋白的生物缀合物。在一个实施方案中,本文提供了包含本发明的修饰的载体蛋白和选自金黄色葡萄球菌5型荚膜糖、金黄色葡萄球菌8型荚膜糖、脑膜炎奈瑟菌血清群A荚膜糖(MenA)、脑膜炎奈瑟菌血清群C荚膜糖(MenC)、脑膜炎奈瑟菌血清群Y荚膜糖(MenY)、脑膜炎奈瑟菌血清群W荚膜糖(MenW)、流感嗜血杆菌b型荚膜糖(Hib)、B组链球菌I组荚膜糖、B组链球菌II组荚膜糖、B组链球菌III组荚膜糖、B组链球菌IV组荚膜糖、B组链球菌V组荚膜糖、来自伤寒沙门氏菌的Vi糖、脑膜炎奈瑟菌LPS(如L3和/或L2)、卡他拉莫菌LPS、流感嗜血杆菌LPS、志贺氏菌O-抗原、铜绿假单胞菌O-抗原、大肠杆菌O-抗原或肺炎链球菌荚膜多糖的抗原的生物缀合物。

[0272] 本发明的生物缀合物可例如通过色谱法(例如离子交换、阳离子交换、阴离子交换、亲和和分级柱色谱法)、离心、差异溶解度或通过用于纯化蛋白质的任何其它标准技术来纯化。参见例如Saraswat等人,2013,Biomed.Res.Int.ID#312709(p.1-18);还参见WO2009/104074中描述的方法。此外,生物缀合物可与本文所述或本领域另外已知的异源多肽序列融合以促进纯化。用于纯化特定生物缀合物的实际条件将部分取决于合成策略和诸如生物缀合物的净电荷、疏水性和/或亲水性的因素,并且对于本领域技术人员而言是显而易见的。

[0273] 本发明的另一个方面是生产生物缀合物的方法,所述生物缀合物包含与多糖连接的修饰的载体蛋白(或由其组成),所述方法包括(i)在适于产生蛋白质的条件下(和任选地在适于产生糖类的条件下)培养本发明的宿主细胞,和(ii)分离由所述宿主细胞产生的生物缀合物。

[0274] 本发明的另一个方面是通过本发明的方法产生的生物缀合物,其中所述生物缀合

物包含与修饰的载体蛋白连接的糖。

[0275] 质谱方法

[0276] 本发明通过分析驱动的设计方法提供了载体蛋白，其允许通过与质谱联用的液相色谱(LC-MS)测量糖基化位点占有率。这与(i)定量终产物中的未糖基化载体，(ii)跟踪过程中的生物缀合速率和(iii)在设计有多个糖基化位点的载体蛋白以增加糖基化速率的情况下，定量单一位点上的糖基化程度特别相关。

[0277] 该策略基于使用同位素标记的内标定量天然未糖基化形式的糖肽。特别地，两组重同位素标记的肽标准品在蛋白水解之前掺入样品中，并通过LC-MS分析消化的样品。一组肽标准品用于测定总糖蛋白量，而另一标准品监测未糖基化的糖蛋白量。以这种方式，通过从总蛋白量中减去未糖基化的蛋白量来计算蛋白糖基化部分的丰度，然后测定位点占有率。

[0278] 免疫原性组合物

[0279] 本发明的修饰的载体蛋白和缀合物(例如生物缀合物)特别适于包含在免疫原性组合物和疫苗中。本发明提供了包含本发明的修饰的载体蛋白或本发明的缀合物或本发明的生物缀合物的免疫原性组合物。

[0280] 还提供了制备本发明的免疫原性组合物的方法，其包括将本发明的修饰的载体蛋白或缀合物(例如生物缀合物)与药学上可接受的赋形剂或载体混合的步骤。

[0281] 免疫原性组合物包含免疫有效量的本发明的修饰的载体蛋白或缀合物(例如生物缀合物)以及任何其它组分。“免疫有效量”是指将该量作为单剂量或作为系列的一部分给予个体对于治疗或预防是有效的。该量根据待治疗个体的健康和身体状况、年龄、所需的保护程度、疫苗的制剂和其它相关因素而变化。预期该量将落在可通过常规试验确定的相对宽的范围内。

[0282] 本发明的免疫原性组合物还可以含有稀释剂，例如水、盐水、甘油等。另外，可以存在辅助物质，例如润湿剂或乳化剂、pH缓冲物质、多元醇等。

[0283] 包含本发明的修饰的载体蛋白或缀合物(或生物缀合物)的免疫原性组合物可以包含适用于药物施用的任何另外的组分。在具体实施方案中，本发明的免疫原性组合物是单价制剂。在其它实施方案中，本发明的免疫原性组合物是多价制剂，例如二价、三价和四价制剂。例如，多价制剂包含多于一种抗原，例如多于一种缀合物。

[0284] 疫苗

[0285] 本发明还提供包含本发明的免疫原性组合物和药学上可接受的赋形剂或载体的疫苗。

[0286] 药学上可接受的赋形剂和载体可由本领域技术人员选择。例如，药学上可接受的赋形剂或载体可以包括缓冲剂，例如Tris(三羟甲基氨基甲烷)、磷酸盐(例如磷酸钠)、乙酸盐、硼酸盐(例如硼酸钠)、柠檬酸盐、甘氨酸、组氨酸和琥珀酸盐(例如琥珀酸钠)，合适地氯化钠、组氨酸、磷酸钠或琥珀酸钠。药学上可接受的赋形剂可以包括盐，例如氯化钠、氯化钾或氯化镁。任选地，药学上可接受的赋形剂含有至少一种稳定溶解度和/或稳定性的组分。增溶剂/稳定剂的实例包括去污剂，例如月桂酰肌氨酸(laurel sarcosine)和/或聚山梨醇酯(例如TWEEN™ 80)。稳定剂的实例还包括泊洛沙姆(例如泊洛沙姆124、泊洛沙姆188、泊洛沙姆237、泊洛沙姆338和泊洛沙姆407)。药学上可接受的赋形剂可以包括非离子表面活性

剂,例如聚氧乙烯脱水山梨糖醇脂肪酸酯、聚山梨醇酯-80(TWEEN™ 80)、聚山梨醇酯-60(TWEEN™ 60)、聚山梨醇酯-40(TWEEN™ 40)和聚山梨醇酯-20(TWEEN™ 20),或聚氧乙烯烷基醚(合适地聚山梨醇酯-80)。可选的增溶剂/稳定剂包括精氨酸和玻璃形成多元醇(例如蔗糖、海藻糖等)。药物赋形剂可以是防腐剂,例如苯酚、2-苯氧乙醇或硫柳汞。其它药学上可接受的赋形剂包括糖(例如乳糖、蔗糖)和蛋白质(例如明胶和白蛋白)。药学上可接受的载体包括水、盐水溶液、葡萄糖水溶液和甘油溶液。许多药学上可接受的赋形剂和载体描述于例如Remington's pharmaceutical sciences,by E.W.Martin,Mack Publishing Co.Easton,PA,第5版(975)。

[0287] 在一个实施方案中,本发明的免疫原性组合物或疫苗另外包含一种或多种缓冲剂,例如磷酸盐缓冲剂和/或蔗糖磷酸盐谷氨酸盐缓冲剂。在其它实施方案中,本发明的免疫原性组合物或疫苗不包含缓冲剂。

[0288] 在一个实施方案中,本发明的免疫原性组合物或疫苗另外包含一种或多种盐,例如氯化钠、氯化钙、磷酸钠、谷氨酸一钠和铝盐(例如氢氧化铝,磷酸铝,明矾(硫酸铝钾)或这些铝盐的混合物)。在其它实施方案中,本发明的免疫原性组合物或疫苗不包含盐。

[0289] 本发明的免疫原性组合物或疫苗可另外包含防腐剂,例如汞衍生物硫柳汞。在一个具体实施方案中,本发明的免疫原性组合物或疫苗包含0.001%-0.01%硫柳汞。在其它实施方案中,本发明的免疫原性组合物或疫苗不包含防腐剂。

[0290] 通常当以多剂量形式包装时,本发明的疫苗或免疫原性组合物还可包含抗微生物剂。例如,本发明的免疫原性组合物或疫苗可包含2-苯氧乙醇。

[0291] 本发明的疫苗或免疫原性组合物还可以包含去污剂,例如聚山梨醇酯,如TWEEN™ 80。去污剂通常以低水平存在,例如<0.01%,但是已经提出更高水平用于稳定抗原制剂,例如高达10%。

[0292] 本发明的免疫原性组合物可以与施用说明书一起包括在容器、包装或分配器中。

[0293] 本发明的免疫原性组合物或疫苗可在使用前储存,例如组合物可冷冻储存(例如在约-20℃或约-70℃下);在冷藏条件下(例如在约4℃下)储存;或在室温下储存。

[0294] 本发明的免疫原性组合物或疫苗可储存于溶液中或冻干。在一个实施方案中,溶液在糖如蔗糖、海藻糖或乳糖存在下冻干。在另一个实施方案中,本发明的疫苗被冻干并在使用前临时重构。

[0295] 疫苗制剂通常描述于Vaccine Design("The subunit and adjuvant approach"(Powell M.F.&NewmanM.J.编辑)(1995)Plenum Press New York)。脂质体内的包封描述于Fullerton,美国专利4235877。

[0296] 本发明还提供包含本发明的免疫原性组合物和药学上可接受的赋形剂或载体的疫苗。

[0297] 佐剂

[0298] 在一个实施方案中,本发明的免疫原性组合物或疫苗包含佐剂或与佐剂组合施用。用于与本发明的免疫原性组合物或疫苗组合施用的佐剂可以在施用所述免疫原性组合物或疫苗之前、同时或之后施用。在一些实施方案中,术语"佐剂"是指当与本发明的免疫原性组合物或疫苗联合或作为本发明的免疫原性组合物或疫苗的一部分施用时增强、提高和/或加强对生物缀合物的免疫应答的化合物,但当单独施用所述化合物时,不产生对修饰

的载体蛋白/缀合物/生物缀合物的免疫应答。在一些实施方案中，佐剂对修饰的载体蛋白、缀合物或生物缀合物产生免疫应答，而不产生变态反应或其它不良反应。

[0299] 在一个实施方案中，本发明的免疫原性组合物或疫苗是佐剂化的。佐剂可以通过几种机制增强免疫应答，包括例如淋巴细胞募集、刺激B和/或T细胞，和刺激巨噬细胞。佐剂的具体示例包括但不限于铝盐(明矾)(如氢氧化铝、磷酸铝和硫酸铝)、3-脱O-酰化单磷酰脂质A(MPL)(参见英国专利GB 2220211)、MF59(Novartis)、AS03(GlaxoSmithKline)、AS04(GlaxoSmithKline)、聚山梨醇酯80(TWEEN™ 80;ICL Americas,Inc.)、咪唑并吡啶化合物(参见国际申请号PCT/US2007/064857,以国际公开号WO2007/109812公开)、咪唑并喹喔啉化合物(参见国际申请号PCT/US2007/064858,以国际公布号WO2007/109813公开)和皂苷，如QS21(参见Kensil等人,Vaccine Design:The Subunit and Adjuvant Approach(eds.Powell&Newman,Plenum Press,NY,1995);美国专利5057540号)。在一些实施方案中，佐剂是弗氏佐剂(完全或不完全)。其它佐剂为水包油乳剂(如角鲨烯或花生油)，任选地与免疫刺激剂如单磷酯酰脂质A组合(参见Stoute等人N.Engl.J.Med.336,86-91(1997))。另一种佐剂是CpG(Bioworld Today,Nov.15,1998)。

[0300] 在本发明的一个方面，佐剂是铝盐如氢氧化铝凝胶(明矾)或磷酸铝。

[0301] 在本发明的另一方面，选择佐剂作为TH1或TH2型反应的优先诱导剂。高水平的Th1型细胞因子倾向于诱导细胞介导的针对给定抗原的免疫反应，而高水平的Th2型细胞因子倾向于有利诱导针对该抗原的体液免疫反应。重要的是记住，Th1和Th2型免疫反应的区别不是绝对的。实际上，个体将支持被描述为主要是Th1或主要是Th2的免疫反应。然而，通常方便地考虑根据Mosmann和Coffman(Mosmann,T.R.和Coffman,R.L.(1989)TH1and TH2cells:different patterns of lymphokine secretion lead to different functional properties.Annual Review of Immunology,7,p145-173)在鼠CD4+ve T细胞克隆中所描述的细胞因子家族。传统上，Th1型反应与T淋巴细胞产生INF-γ和IL-2细胞因子有关。通常与Th1型免疫反应的诱导直接相关的其它细胞因子不是由T细胞产生的，例如IL-12。相反，Th2型反应与IL-4、IL-5、IL-6、IL-10的分泌有关。促进主要Th1反应的合适佐剂系统包括:单磷酰脂质A或其衍生物，特别是3-脱O-酰化单磷酰脂质A(3D-MPL)(其制备参见GB 2220211A);在脂质体中的MPL，例如3D-MPL，和皂苷QS21，例如包含胆固醇和DPOC的脂质体;和单磷酰脂质A(例如3-脱-O-酰化单磷酰脂质A)与铝盐(例如磷酸铝或氢氧化铝)或水包油乳液的组合。在这种组合中，抗原和3D-MPL可以包含在相同的颗粒结构中，允许更有效地递送抗原性和免疫刺激性信号。研究表明，3D-MPL能够进一步增强明矾吸附的抗原的免疫原性(Thoelen等人.Vaccine(1998)16:708-14;EP689454-B1)。含有未甲基化CpG的寡核苷酸(WO96/02555)也是TH1反应的优先诱导剂，并且适用于本发明。

[0302] 本发明的疫苗或免疫原性组合物可含有水包油乳液，因为已建议这些可用作佐剂组合物(EP399843;WO95/17210)。可以使用水包油乳剂，例如WO95/17210(其公开了包含2-10％角鲨烯、2-10％α生育酚和0.3-3％吐温80的水包油乳剂，及其单独或与QS21和/或3D-MPL组合的用途)、WO99/12565(其公开了包含可代谢油、皂苷和甾醇和MPL的水包油乳剂组合物)或WO99/11241中所描述的那些。其它水包油乳液如WO09/127676和WO09/127677中公开的那些也是合适的。在WO95/17210中描述了在水包油乳剂中包含QS21、3D-MPL和生育酚的特别有效的佐剂制剂。在一个具体的实施方案中，免疫原性组合物或疫苗另外包含皂苷，

例如QS21。免疫原性组合物或疫苗还可以包含水包油乳剂和生育酚(WO95/17210)。

[0303] 预防和治疗用途

[0304] 本发明还提供了治疗和/或预防受试者的细菌感染的方法,包括向受试者施用本发明的经修饰的载体蛋白、缀合物或生物缀合物。修饰的载体蛋白、缀合物或生物缀合物可以是免疫原性组合物或疫苗的形式。在一个具体实施方案中,本发明的免疫原性组合物或疫苗用于预防受试者(例如人类受试者)的细菌感染。可以使用本发明的经修饰的载体蛋白、缀合物或生物缀合物治疗和/或预防的细菌感染包括由葡萄球菌属物种、埃希氏菌属物种、志贺氏菌属物种、奈瑟菌属物种、莫拉氏菌属物种、嗜血杆菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特菌属物种或弯曲菌属物种、金黄色葡萄球菌、脑膜炎奈瑟菌、流感嗜血杆菌、流感嗜血杆菌b型、B组链球菌、伤寒沙门氏菌、卡他拉莫菌LPS、福氏志贺氏菌、铜绿假单胞菌、大肠杆菌或肺炎链球菌引起的感染。

[0305] 本文还提供了在受试者中诱导针对细菌的免疫应答的方法,包括向受试者施用本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)。在一个实施方案中,所述受试者在施用时患有细菌感染。在另一个实施方案中,所述受试者在施用时没有细菌感染。本发明的经修饰的载体蛋白、缀合物或生物缀合物可用于诱导针对葡萄球菌属物种、埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特氏菌属物种或弯曲杆菌属物种的免疫应答。在一个具体的实施方案中,本发明的经修饰的Hla蛋白或缀合物或生物缀合物用于诱导针对葡萄球菌属物种(例如金黄色葡萄球菌)的免疫应答。

[0306] 本文还提供了在受试者中诱导针对细菌的调理吞噬抗体的产生的方法,其包括向所述受试者施用本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)。在一个实施方案中,所述受试者在施用时患有细菌感染。在另一个实施方案中,所述受试者在施用时没有细菌感染。本文提供的本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)可用于诱导产生针对葡萄球菌属物种,埃希氏菌属物种、志贺氏菌属物种,克雷伯氏菌属物种,黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特氏菌属物种或弯曲杆菌属物种的调理吞噬抗体。在一个具体的实施方案中,本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)用于诱导产生针对葡萄球菌属物种(例如金黄色葡萄球菌)的调理吞噬抗体。

[0307] 本发明还提供了治疗和/或预防受试者的细菌感染的方法,包括向受试者施用本发明的经修饰的载体蛋白、缀合物或生物缀合物。修饰的载体蛋白、缀合物或生物缀合物可以是免疫原性组合物或疫苗的形式。在一个具体实施方案中,本发明的免疫原性组合物或

疫苗用于预防受试者(例如人类受试者)的细菌感染。可以使用本发明的经修饰的载体蛋白、缀合物或生物缀合物治疗和/或预防的细菌感染包括由埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特氏菌属物种或弯曲菌属物种引起的感染。在一个具体实施方案中,本发明的免疫原性组合物或疫苗用于治疗或预防链球菌属物种(例如肺炎链球菌)的感染。

[0308] 本文还提供了在受试者中诱导针对细菌的免疫应答的方法,包括向受试者施用本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)。在一个实施方案中,所述受试者在施用时患有细菌感染。在另一个实施方案中,所述受试者在施用时没有细菌感染。本发明的经修饰的载体蛋白、缀合物或生物缀合物可用于诱导针对埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特氏菌属物种或弯曲杆菌属物种的免疫应答。在一个具体的实施方案中,本发明的经修饰的载体蛋白或缀合物或生物缀合物用于诱导针对链球菌属物种(例如肺炎链球菌)的免疫应答。

[0309] 本文还提供了在受试者中诱导针对细菌的调理吞噬抗体的产生的方法,其包括向所述受试者施用本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)。在一个实施方案中,所述受试者在施用时具有细菌感染。在另一个实施方案中,所述受试者在施用时没有细菌感染。本文提供的本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)可用于诱导产生针对埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森氏菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特菌属物种或弯曲菌属物种的调理吞噬抗体。在一个具体的实施方案中,本发明的经修饰的载体蛋白或缀合物或生物缀合物(或免疫原性组合物或疫苗)用于诱导产生针对链球菌属物种(例如肺炎链球菌)的调理吞噬抗体。

[0310] 在一个实施方案中,本发明是通过施用治疗有效量的本发明的免疫原性组合物或疫苗(儿科疫苗)在婴儿(在本发明的上下文中定义为0-2岁)中引发免疫应答的改进的方法。在一个实施方案中,所述疫苗是儿科疫苗。

[0311] 在一个实施方案中,本发明是通过施用治疗有效量的本发明的免疫原性组合物或疫苗在老年人群(在本发明的上下文中,如果患者是50岁或50岁以上,通常55岁以上,更通常60岁以上,则认为患者是老年人)中引发免疫应答的改进方法。在一个实施方案中,所述疫苗是用于老年人的疫苗。

[0312] 本专利说明书中引用的所有参考文献或专利申请在此通过引用并如本文。

[0313] 为了更好地理解本发明,给出以下实施例。这些实施例仅用于说明的目的,不应解释为以任何方式限制本发明的范围。

[0314] 实施例

[0315] 材料和方法

[0316] 菌株、克隆和生长条件。

[0317] 应用聚合酶不完全引物延伸(PIPE)方法(Klock HE等人,2009,Methods Mol Biol.,498:91-103)进行诱变和克隆实验以获得携带新设计的载体Hla-i、Hla-s、Hla-v和Hla-a载体的序列(SEQ ID NO:36-39)的质粒Bcp218、Bcp233、Bcp234和Bcp235,其中重组形式的Hla的第131位的生物缀合共有序列(专利PCT/EP2018/085854,公开为WO2019/121924中的SEQ ID NO:9;质粒:pGVXN2872;还参见Wacker,M.等人,J.Infect.Dis.2014,209,1551-1561)分别被序列KDSNITSAR(SEQ ID NO:42)、KDSNSTSAR(SEQ ID NO:43)、KDSNVTSAR(SEQ ID NO:44)和KDSNATSAR(SEQ ID NO:45)取代。作为131位的替代方案,或在131位之外,这些糖基化位点序列被添加到Hla蛋白的N-末端和/或C-末端以获得另外三个新设计的Hla载体,Hla-N,131、Hla-131,C和Hla-N,C,其中在N-末端的糖基化位点之前和之后引入一些氨基酸(GSGGG,SEQ ID NO:50),从而在Flg1信号序列和蛋白质序列的起始之间产生间隔区,而仅在C-末端的糖基化位点之前具有间隔区(SEQ ID No:51-53)。编码用于设计的双位点构建体的序列的DNA质粒获自GeneArt®供应商,并且Hla构建体的DNA插入片段通过上述PIPE方法进行克隆。通过使用这样的诱变PCR反应产物转化化学感受态大肠杆菌topTEN(Thermo®)允许获得携带特定质粒构建体的菌落(Klock HE等人,2009,Methods Mol Biol.,498:91-103)。纯化来自所选择克隆的质粒并测序以确认和选择目标Hla构建体。

[0318] 通过电穿孔用编码金黄色葡萄球菌Hla-i、Hla-v、Hla-s、Hla-N,131、Hla-131,C或Hla-N,C异形体,以及空肠弯曲杆菌寡糖转移酶PglBN311V-K482R-D483H-A669V的质粒(pGVXN1221)共转化大肠杆菌W3110StGVXN9268细胞。

[0319] 转化的细菌在补充有两种抗生素卡那霉素[50μg/ml]和壮观霉素[80μg/ml]的选择性琼脂平板上生长过夜,以分别维持编码Hla和PglB的质粒。将细菌接种在含有抗生素的50ml Lysogeny肉汤(LB)中,并在锥形瓶中在37℃,180rpm下振荡过夜。将补充有卡那霉素[50μg/ml]和壮观霉素[80μg/ml]的50ml HTMC培养基的培养物接种至600nm下0.1光密度(OD_{600nm})的稀释度,在37℃下在摇床中以180rpm孵育,直至平均OD_{600nm}为0.8-1.0,然后通过使用用于诱导Hla表达的0.2%阿拉伯糖和用于诱导PglB表达的1mM IPTG诱导过夜。将培养物在37℃,180rpm下振荡过夜,并且在诱导20小时之后收获200D_{600nm}的细菌。弃去上清液,并且沉淀物立即用于周质提取。

[0320] 周质提取

[0321] 对20分钟离心(4℃下4000rcf)之后回收的200D_{600nm}的细菌沉淀进行周质提取(PPE),重悬在600uL裂解缓冲液(3 0mM Tris HCl pH 8.5,1mM EDTA,20%(重量/体积)蔗糖)中,然后在4℃下在旋转振荡器中用1mg/mL最终溶菌酶处理20分钟。在16000rcf 4℃下离心20分钟之后,立即收集上清液并在-20℃下储存直至其使用。总蛋白质含量通过二喹啉甲酸测定法(试剂盒Reducing Agent Compatible,Thermo Fischer Scientific)进行评估。

[0322] 蛋白质印迹分析

[0323] 通过SDS-PAGE(MES 1×中4-12%Bis Tris,150V下1小时)和免疫印迹Hla-CP5抗血清(1:1000)分析周质Hla的糖基化状态。以1:5000稀释度使用市售的山羊抗兔HRP-缀合

的二抗(DAKO Ab,Agilent,CA,USA)。

[0324] 用于Hla总蛋白量定量的PTP的选择。

[0325] 将20μg的重组Hla H35L在含有5mM DTT和0.1%(重量/体积)RapiGestSF(Waters,USA)的50mM碳酸氢铵中在95℃下煮沸5分钟,并用胰蛋白酶[1/25(重量/重量),酶/底物比](Promega)在37℃下消化过夜。通过在与配备有微电喷雾ESI源的Thermo Scientific Q Exactive plus质谱仪(Thermo)偶联的Acquity UPLC系统(Waters®)上进行的LC-MS/MS分析肽混合物。样品使用全回路注射以40uL/min的流速在流动相A(0.1%甲酸FA)中加载。然后在nano Acquity UPLC Peptide BEH C18柱75μm×100mm(Waters®)上使用60分钟梯度3-98%流动相B(98%(体积/体积)乙腈,0.1%(体积/体积)FA)以40μl/min的流速分离肽。

[0326] 使用Xcalibur软件,在前10个m/z上对洗脱的肽进行自动数据依赖性采集(DDA)。使用Peaks X软件在从NCBInr数据库下载的大肠杆菌K-12数据库上进行肽识别,其中插入了Hla H35L序列。作为变量修改的检索参数为:甲硫氨酸氧化、谷氨酰胺和天冬酰胺脱酰胺、胰蛋白酶切割(切割K或R的C-末端侧,除非下一个残基是P),肽质量容差为0.15Da,肽MS/MS容差为0.15Da,漏切(missed cleavage)=2,离子电荷状态:+2、+3、+4)。

[0327] 基于以下标准选择合适的PTP:(i)对Hla特异性的肽,(ii)对亲本或片段离子显示强MS信号强度的肽,(iii)不含甲硫氨酸和色氨酸残基(其易于氧化)或N-末端谷氨酰胺(以避免环化)的肽。

[0328] SRM-MS方法设置

[0329] 使用与结合UPLC的Xevo TQ三重四极质谱仪(Waters,USA)偶联的反相柱(ACQUITY UPLC HSS T3柱,100Å,1.8μm,2.1mm×30mm,Waters,USA),用在0.1%(体积/体积)FA中的轻和重形式的混合溶液中的1pmol合成肽进行肽的检测和色谱洗脱优化,所述反相柱。用流动相A 0.1%(体积/体积)FA水溶液和B 0.1%(体积/体积)FA乙腈溶液建立洗脱梯度。合成肽用于优化从通过使用PinPoint软件在计算机上计算的理论值开始的碰撞能量(CE)值,使用公式CE=0.034×(母离子m/z)+1.314计算(MacLean B.等人,Anal.Chem.2010;82:10116-10124),并验证50μg基质中的转换。通过使用优化的CE和选择的转换,在SRM采集模式下,在纯的和基质背景中进行色谱分离的优化(参见表3)。

[0330] 用于细菌周质提取物的LC-SRM分析的样品制备。

[0331] 使用由PreOmics(Martinsried,Germany)提供的in-stage-tips(iST)样品制备试剂盒消化50μg PPE部分。它是在柱上进行的3步方案:1)裂解、变性、还原和烷基化;2)通过LysC和胰蛋白酶的蛋白水解消化;3)根据供应商的建议进行的肽脱盐。将回收的肽在45℃下真空干燥,并重悬于0.1%(体积/体积)FA中至终浓度为1μg/μL,并在-20℃储存直至MS分析。

[0332] SRM-MS分析。

[0333] 通过每次运行将用IsT样品制备试剂盒消化的10μg周质部分注入柱中进行SRM,每个样品一式三份进行分析。使用以下参数:Q1分离窗口1.0m/z,Q3分离窗口0.7m/z,从MS到MS/MS的转换时间(停留时间)0.03s,以及设置为30V的碰撞室出口和入口电位。1700V的喷雾电压与用于去溶剂化的270℃的加热的离子转移设置一起使用。使用MassLynx软件

(2.1.0版;Waters)获取数据。停留时间设置为30ms,且扫描宽度设置为0.02m/z。在证实每种肽的所有转换的共洗脱之后并遵循Carr等人,Mol Cell Proteomics 13(3):907-917,2014中报道的最佳实践用TargetLynx软件(1.0.0.1版;Waters)测定峰面积定量。

[0334] PTP剂量-范围线性响应和HlA定量

[0335] 在从用作参考背景以考虑基质效应的大肠杆菌糖感受态细胞(用PglB质粒转化的stGVXN9268)制备的周质细菌样品中评估所选PTP的剂量-范围线性响应。在用IsT样品制备试剂盒消化之前,将标记的PTP(终浓度0.1pmol/μL)和未标记的PTP(终浓度1.6pmol/μL至0.0125pmol/μL)掺入50μg的周质部分中。

[0336] 对于每种PTP,将浓度绘制为峰面积轻(可变)/峰面积重(恒定)的比率,并使用拟合曲线获得所选择的内生PTP的浓度。根据国际协调会议(ICH)指南(http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html),将每种肽的定量下限(LLOQ)设定为可定量检测的拟合曲线上的最低浓度点,并定义为10σ/S,其中σ=响应的标准偏差,S=校准曲线的斜率。

[0337] 考虑到Hla-i(34093.07Da)、Hla-s(34066.99Da)、Hla-v(34079.05Da),Hla-N,131(36962.06Da)、Hla-131,C(36518.60Da)和Hla-N,C(37390.51Da)异形体的分子量平均值,以皮克/微克总周质蛋白提取物报告Hla浓度。通过在相关的剂量-反应线性曲线中插入每个肽-反应值获得定量(图4)。

[0338] 结果

[0339] 图1报告了设计用于生物缀合的新载体蛋白的工作流程。

[0340] 第一步是共有序列的计算机设计,预计是PglB酶的底物(Kowarik等人,2006,EMBO J.25:1957-1966)并且能够产生适于通过MS定量糖基化程度的胰蛋白酶肽(称为蛋白特征性肽(proteotypic peptide)PTP)(Zhu等人,2014,J Am Soc Mass Spectrom.25:1012-7)。

[0341] 通过在精氨酸残基中掺入13C-15N以天然或重标记形式化学合成设计的PTP,并使用三重四极仪器研究它们在MS/MS分析中的行为。

[0342] 一旦鉴定出适于通过MS定量的PTP,将相应的序列引入载体蛋白中,并评价PglB酶识别和糖基化新载体的效率。

[0343] 然后通过使用同位素标记的内标和SRM方法,从糖肽的非糖基化形式的绝对定量测定每个共有序列的位点占有率。两组重同位素标记的肽标准品在蛋白水解和LC-SRM MS之前掺入样品中。一组肽标准品用于测定总载体浓度,而另一组标准品监测载体的非糖基化部分。以这种方式,通过从总蛋白质浓度中减去非糖基化蛋白质丰度来计算糖基化蛋白质部分的丰度,然后测定位点占有率。已经证明该方法对于天然糖基化的真核生物蛋白质的定量是成功的(Zhu等人,2014,J Am Soc Mass Spectrom.25:1012-7).

[0344] 作为概念的验证,将新设计的共有序列引入金黄色葡萄球菌Hla的重组形式,这是一种用作金黄色葡萄球菌CP5的生物缀合的载体蛋白的底物(参见PCT/EP2018/085854,公开号为WO2019/121924)。据报道,该载体通过在131位插入共有序列(-3)KDQNRTK(+3)被有效地糖基化,其中Asn残基(在0位)是聚糖受体。不幸的是,发现该抗原设计不适应于糖基化程度的定量,因为胰蛋白酶对载体的消化产生了未修饰的肽(-2)DQNR(+1),其太短和是亲水性的而不能通过LC-MS/MS监测。测试了不同的树脂和梯度而没有成功。

[0345] 共有序列的计算机设计

[0346] 空肠弯曲杆菌PglB的共有序列底物已被充分表征(Kowarik等人,2006,EMBO J.25:1957-1966)。该序列特征在于在糖的asn受体的-2位中的带负电侧链氨基酸残基(asp或glu)以及+2位中的ser或thr的存在。此外,有效的生物缀合还需要共有位点在载体蛋白的可及的和柔性的环中(Silverman等人,2016,J.Biol.Chem.291,22001-22010)。

[0347] 在32个天然空肠弯曲杆菌糖蛋白中发现的糖基化的asn残基的-6至+6区域中氨基酸出现的统计分析报道于Kowari等人,EMBO J.2006;25(9):1957-66,如图2A所示。分别选择在位置-3、-1、+1、+3和+4的氨基酸残基(在灰色框中以粗线表示)用于设计四种共有序列(图2B)。这些残基因为时常发现并对应于设置的标准而选择。+5位中的氨基酸arg(未在统计学分析中报道)和-3位中的氨基酸lys是产生PTP所需的胰蛋白酶底物。PTP彼此区别仅在于+1位的氨基酸残基。

[0348] 考虑到这些最低要求,设计了预测为PglB的底物并能够产生胰蛋白酶肽的4个共有序列(图2B)。具体地,考虑以下标准:(i)为了避免对载体结构的可能干扰,插入的共有序列不超过9个氨基酸残基,并且限制疏水性和芳香族氨基酸残基的插入,(ii)为了避免低估的定量,限制易于翻译后修饰如氧化(met、cys、trp)和脱氨基(asn和gln)的氨基酸残基,并且将共有序列设计为胰蛋白酶的底物,根据其高特异性、效力和产生C-末端带正电荷的肽的能力进行选择,(iii)考虑了从含有32个活性空肠弯曲杆菌N-糖基化位点的数据集的比较证明的围绕asn(糖的受体)的优先氨基酸残基(4),和(iv)新设计的共有序列对于新设计的载体异形体是独特的。

[0349] 四个设计的共有序列是(-3)KDSNXTSAR(+5),其中X是Ile、Ser、Val或Ala氨基酸残基。LysC/胰蛋白酶消化之后,产生PTP(-2)DSNXTSAR(+5),根据+1位中存在的氨基酸残基命名为PTP-i(SEQ ID NO:42)、PTP-s(SEQ ID NO:43)、PTP-v(SEQ ID NO:44)和PTP-a(SEQ ID NO:45)(图2B)。

[0350] 用于定量糖基化程度的SRM测定

[0351] 在SRM测定中评价化学合成的PTP-i-s-v-a的行为。在与三相四极质谱仪(具有范围为1.65-5.74分钟的明确区别的保留时间(在肽PTP-a和PTP-s之间观察到0.5分钟的最小差异)和具有强MS信号)偶联的反相C18上在线分离四种PTP。

[0352] 对于每种PTP,通过PinPoint软件计算4个转换(含有对每个糖基化位点特异性的选择性氨基酸残基的前体/产物对b4、y5和y6,以及所有肽所共有的y4),并首先通过严整(in-neat)注射进行优化(表1)。

[0353] 然后在用LysC/胰蛋白酶消化的大肠杆菌周质部分验证转换,以评估基质的作用(Lange等人,2008,Mol.Syst.Biol.4:222)。虽然基质对PTP-s、PTP-v和PTP-i性能具有较小的影响(表1),但它对PTP-a具有有害的影响,因为其在重复分析中保留时间和转换都不稳定。因此,没有进一步研究生物缀合共有序列(-5)KDSNATSAR(+3)。

[0354] 对于PTP-s、PTP-v和PTP-i,优化碰撞能量并建立剂量-反应线性曲线,在胰蛋白酶消化之前,向50μg大肠杆菌周质部分添加固定量的重形式PTP(0.1pmol/μg)和标量浓度的轻PTP(0.0125至1.6pmol/μg)。根据国际协调会议(ICH)指南(http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html),将每种肽的LLOQ设定为可定量检测的拟合曲线上的最低浓度点,并定义为10σ/S,其中σ=响应的标准偏差,和S=校准曲线的斜率。此外,基于ICH指南定义了LOD或检测极限(LOD=3.3σ/S)(图4)。

[0355] 选择的共有序列是PglB的底物。将三种新设计的共有序列PTP-i(SEQ ID NO:42)、PTP-s(SEQ ID NO:43)和PTP-v(SEQ ID NO:44)插入优化的Hla抗原的131位(参见PCT/EP2018/085854,公开号为WO2019/121924),以产生生物缀合物Hla-i-CP5、Hla-v-CP5和Hla-s-CP5,一旦用LysC/胰蛋白酶消化,它们分别产生PTP-i、PTP-v、PTP-s肽。分离携带生物缀合所需机制的糖感受态大肠杆菌菌株的周质,并使用识别Hla-CP5生物缀合物的鼠血清通过蛋白质印迹分析评估CP5与Hla的缀合(图3)。载体通过在三种不同构建体中相当的部分糖基化谱进行表征,尽管糖基化程度不能从蛋白印迹进行定量。Hla-CP5糖基化程度的定量

[0356] 通过SRM评估糖基化程度。具体地,通过以下等式确定位点占有率(Zhu等人,2015,J Am Soc Mass Spectrom.25:1012-7):

[0357] $$\text{位点占有率}(\%) = \frac{(\text{总} - \text{未经修饰的})\text{载体浓度}}{\text{总载体浓度}} \times 100$$

[0358] 其中未经修饰的载体浓度通过定量内源PTP-i、PTP-v、PTP-s测定,和总载体浓度通过对Hla特异性的肽进行定量。

[0359] 为了鉴定合适的Hla特异性肽,用胰蛋白酶消化重组Hla并通过LC-MS/MS分析产生的肽。选择分别命名为$PTP-_2$和$PTP-_1$的两种肽,42T-50K(TGDLVTYK,SEQ ID NO:48)和225A-234K(AADNFLDPNK,SEQ ID NO:49)。关于两种肽的所有信息(转换、优化的CE、保留时间和在基质中的LLOQ)报告于表3中。

[0360] 此外,SRM测定需要载体的有效蛋白酶消化以确保每种选择的PTP的定量的一致性。通过SDS/PAGE检查消化效率,并通过LC-MS/MS评估漏切数量低于总识别肽的2%(Biagini M等人,2016,Proc Natl Acad Sci USA.113:2714-9)。

[0361] PTP-i-v-s和PTP_{-1-2}的互换性通过在胰蛋白酶消化之前,在大肠杆菌周质部分中掺入重组Hla和已知量的每种同位素标记的PTP来证明。

[0362] 表达不同的新设计的Hla载体与CP5的生物缀合物(Hla-i-CP5、Hla-v-CP5和Hla-s-CP5)的大肠杆菌糖感受态菌株分离周质部分,和在LysC/胰蛋白酶消化并通过LC-SRM分析之前添加已知量的每种同位素标记的PTP。实验由一式三份运行的三个独立的消化进行。PTP的定量在运行之间是高度再现的,几乎所有变异系数(CV)都低于10%,而与消化和生物缀合程度定量的再现性相关的CV低于14%(表1)。这些值符合文献中报告的CV和分析的预期误差(Hüttenhain等人,2009,Curr.Opin.Chem.Biol.13 518-525)。总和未糖基化载体量的浓度使得确定Hla-i-CP5、Hla-v-CP5和Hla-s-CP5分别占表达的载体总量的41.47%、45.14%和42.73%(表4)。糖基化的类似程度与蛋白质印迹分析中观察到的模式一致(图3)。在共有序列中引入可变氨基酸残基(ser、val或ile)不显著影响PglB酶对生物缀合物Hla载体的效力。如下所述,三种不同共有序列的鉴定允许设计带有多个可定量缀合位点的Hla载体。

[0363] 携带两个糖基化位点的Hla构建体的设计

[0364] 通过在Hla蛋白的N-末端、C-末端和/或131位以可选的组合插入PTP-i和PTP-v来设计具有适于糖基化程度定量的多个共有序列底物的Hla载体。根据它们各自的共有序列插入位置,在此将它们命名为Hla-N,131(SEQ ID NO:50)、Hla-131,C(SEQ ID NO:51)和Hla-N,C(SEQ ID NO:52)。

[0365] 三种载体的定量和计算的糖基化程度报告于下表5中并总结于图5A中。

[0366] 在周质提取物中定量低量的载体Hla-N,131(2.90ng/μg总周质蛋白),并且得到的含有糖-序列的两种肽(PTP-i和PTP-v)不能用在分析的LLOQ下检测的值定量。为此,两个共有序列被认为是完全缀合的。对于Hla-N,C异形体,定量结果为10.66ng/μg周质提取物。在N-末端结构域携带糖-序列的肽(PTP-v)导致完全糖基化,而对于C-末端位置的糖-序列(PTP-i)仅达到约19%的糖基化程度。Hla-131,C异形体是表达最多的异形体,具有38.71ng/μg总周质蛋白,在131位(PTP-i)以及在C-末端(PTP-v)具有约19%的糖基化程度。

[0367] 这些数据表明,当几个位点同时插入载体蛋白中时,通过本发明方法对糖基化程度的测量也可以在单个生物缀合位点上进行评估。此外,位于载体蛋白N-末端结构域的糖基化位点在所分析的所有异形体中完全糖基化。相反,当在131位(其是Hla柔性且暴露于溶剂的环)插入糖基化位点时,糖基化程度与载体蛋白量成反比,而与是否存在第二糖基化位点无关(图5B)。最后,当在更高结构化的C-末端区域插入糖基化位点时,所测试的两种载体蛋白Hla-131,C和Hla-N,C的糖基化程度相似,表明PglB也能够在一定程度上对折叠的蛋白部分进行N-糖基化(Fisher A.C.等人,Applied and Environmental Microbiology,77(3)871-881)。

[0368] 表1:所选PTP的优化SRM转换的列表。对于每种PTP,报告了PTP名称、肽序列、分子量以及SRM分析中的优化转换和稳定的色谱条件。对于每个选择的片段,监测TIC信号评估检测的再现性。

[0369]

肽名称	序列	分子量（Da）	电荷状态	Q1 (m/z) 前体离子	CE	Q3 (m/z) 碎片离子	碎片离子	RT	1 pmol 加载在柱上 TIC	通道强度	RT	基质中 1pmol 加载在柱上 TIC	通道强度
PTP-i	DSNITSAR (SEQ ID NO:42)	862.89	2	432.21	16	434.21	y4	5.74	4.4E+04	30700	4.97	4.49E+04	33800
					14	547.15	y5			8030			5140
					14	661.47	y6			7480			5700
					12	430.20	b4			3980			2610
PTP-s	DSNSTSAR (SEQ ID NO:43)	836.81	2	419.19	16	434.27	y4	1.65	3.32E+04	12000	1.66	3.3E+03	1070
					14	521.39	y5			15100			1640
					14	634.96	y6			8010			1180
					16	404.14	b4			2480			145
PTP-v	DSNVTSAR (SEQ ID NO:44)	848.97	2	425.21	16	434.23	y4	2.02	8.00E+04	46000	2.05	3.47E+04	18600
					18	533.30	y5			12400			5400
					18	647.34	y6			9670			6070
					12	416.38	b4			12000			6690
PTP-a	DSNATSAR (SEQ ID NO:45)	820.81	2	411.19	15	434.24	y4	1.70	3.68E+04	11400	---	---	---
					17	505.27	y5			10300			---
					17	619.32	y6			12400			---
					13	388.27	b4			4710			---

[0370] 表2:优化的色谱梯度

[0371]

时间	mL/min	%A	%B

0.00	0.080	97.0	3.0
1.00	0.080	97.0	3.0
3.00	0.080	95.0	5.0
10.00	0.080	65.0	35.0
13.00	0.100	10.0	90.0
15.00	0.100	10.0	90.0
15.01	0.100	10.0	90.0
17.00	0.100	93.0	7.0
17.01	0.100	93.0	7.0

[0372] 表3关于作为42T-50K(TGDLVTYK,SEQ ID NO:48)和225A-234K(AADNFLDPNK,SEQ ID NO:49)的两个肽PTP-$_2$和PTP-$_1$的信息:转换、优化的CE、保留时间和基质中的LLOQ

[0373]

肽名称	序列	Q1 前体离子(m/Z)	Q3碎片离子(m/Z)	碎片离子	优化的 CE	电荷状态	RT (min)	基质中的 LLOQ (pmol/ug)
PTP-2	TGDLVTDK (SEQ ID NO:48)	506.2532	526.25	y4	18	2	7.42	0.053
			625.32	y5	18	2		
			853.43	y7	18	2		
PTP-2*	TGDLVTDK [HeavyK] (SEQ ID NO: 48)	510.2603	534.26	y4	18	2	7.42	0.0053
			633.33	y5	18	2		
			861.44	y7	18	2		
PTP-1	AADNFLDPNK (SEQ ID NO: 49)	552.7696	473.23	y4	18	2	8.30	0.050
			586.32	y5	18	2		
			962.46	y8	18	2		
			733.39	y6	18	2		
			358.21	y3	18	2		
PTP-1*	AADNFLDPNK [HeavyK] (SEQ ID NO: 49)	556.7767	481.25	y4	18	2	8.30	0.050
			594.33	y5	18	2		
			970.47	y8	18	2		
			741.40	y6	18	2		
			366.22	y3	18	2		

[0374] 表4:用于定义%位点占有率的PTP的定量。对于每种Hla异形体(Hla-i-v-s),进行3次蛋白水解消化,分别使用2组PTP PTP-1-2和PTP-i-v-s定量总HLA和未糖基化形式。对于每次消化,PTP量以pmol/μg总周质蛋白提取物报告,并用于从每次消化计算生物缀合的程度。报告运行和消化之间的CV。获得的每种异形体的%位点占有率报告为从每次消化推导的%位点占有率的平均值,此外,Hla异形体的总量报告为使用每种Hla构建体的ng/μg总周质蛋白提取物,平均MW报告于下表(灰色框)中。

[0375]

Hla-i (平均 MW: 34093.07 Da)									
	消化 1			消化 2			消化 3		
	PTP-1	PTP-2	PTI-i	PTP-1	PTP-2	PTI-i	PTP-1	PTP-2	PTI-i
运行1	0.732	0.669	0.434	0.604	0.516	0.286	0.885	0.818	0.526
运行2	0.719	0.665	0.418	0.603	0.512	0.297	0.879	0.812	0.545
运行3	0.729	0.670	0.410	0.604	0.512	0.295	0.885	0.828	0.534
CV(%)运行	0.94	0.40	2.90	0.10	0.45	2.00	0.39	1.00	1.78
平均PTP-1 &2或PTP-i	0.697		0.421	0.559		0.293	0.851		0.535
%位点占有率			39.67			47.60			37.14
载体量: 23.94 ng/μg 周质蛋白									
%位点占有率 : 41.47 - CV (%): 13.16									

Hla-v (平均 MW: 34079.05 Da)									
	消化 1			消化 2			消化 3		
	PTP-1	PTP-2	PTI-v	PTP-1	PTP-2	PTI-v	PTP-1	PTP-2	PTI-v
运行1	0.561	0.522	0.288	0.475	0.434	0.231	0.845	0.805	0.497
运行2	0.561	0.519	0.293	0.477	0.435	0.243	0.85	0.794	0.45
运行3	0.54	0.523	0.291	0.467	0.434	0.249	0.865	0.796	0.475
CV(%)运行	2.19	0.40	0.87	1.12	0.13	3.80	1.22	0.73	4.96
平均PTP-1 &2或PTP-v	0.538		0.291	0.454		0.241	0.826		0.474
%位点占有率			45.94			46.88			42.60
载体量 20.64 ng/μg周质蛋白									
%位点占有率 : 45.14 - CV (%): 4.98									

Hla-s (平均 MW: 34066.99 Da)									
	消化 1			消化 2			消化 3		
	PTP-1	PTP-2	PTI-s	PTP-1	PTP-2	PTI-s	PTP-1	PTP-2	PTI-s
运行1	0.692	0.623	0.336	0.546	0.502	0.286	0.787	0.727	0.451
运行2	0.692	0.627	0.372	0.578	0.499	0.294	0.789	0.741	0.429
运行3	0.687	0.628	0.387	0.545	0.498	0.321	0.777	0.733	0.474
CV(%)运行	0.42	0.42	7.18	3.37	0.42	6.11	0.82	0.96	4.99
平均PTP-1 &2或PTP-s	0.658		0.365	0.528		0.300	0.759		0.451
%位点占有率			44.54			43.12			40.54
载体量: 22.09 ng/μg周质蛋白									
位点占有率: 42.73 % - CV (%): 4.75									

[0376] 表5用于定义%位点占有率的PTP的定量。对于每种Hla异形体Hla-N,131、Hla-N,C和Hla-131,C,进行3次蛋白水解消化Dig1、Dig2和Dig3,分别使用2组PTP PTP-1-2和PTP-i-v-s定量总HLA和未糖基化形式。对于每次消化,PTP量以pmol/μg总周质蛋白提取物报告,并用于从每次消化计算生物缀合的程度。报告运行和消化之间的CV。获得的每种异形体的%位点占有率报告为从每次消化推导的%位点占有率的平均值,此外,Hla异形体的总量报告为使用每种Hla构建体的ng/μg总周质蛋白提取物,平均MW报告于下表(灰色框)中。

[0377]

Hla-N,131 (平均 MW: 336962.06 Da)												
	消化 1				消化 2				消化 3			
	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v
运行1	0.094	0.086	<LLOQ	<LLOQ	0.075	0.065	<LLOQ	<LLOQ	0.079	0.072	<LLOQ	<LLOQ
运行2	0.095	0.086			0.074	0.066			0.079	0.071		
运行3	0.095	0.087			0.073	0.065			0.08	0.071		
CV(%) 运行	0.61	0.67	-	-	1.35	0.88	-	-	0.73	0.81	-	-
平均 PTP-1 &2 或 PTP-i 或 PTP-v	0.091		-	-	0.070		-	-	0.075		-	-
%位点占有率			>99%	>99%			>99%	>99%			>99%	>99%
载体量 2.90 ng/μg 周质蛋白												
位点占有率 PTP-i: >99%												
位点占有率 PTP-v: >99%												

Hla-131,C (平均 MW: 36518.60 Da)												
	消化 1				消化 2				消化 3			
	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v
运行1	1.119	1.079	0.888	0.955	0.956	0.922	0.748	0.759	1.152	1.123	0.955	0.94
运行2	1.125	1.068	0.885	0.949	0.938	0.987	0.742	0.771	1.167	1.114	0.962	0.979
运行3	1.129	1.075	0.904	0.952	0.947	0.896	0.737	0.732	1.163	1.121	0.93	0.987
CV(%)运行	0.45	0.52	1.14	0.32	0.95	5.01	0.74	2.65	0.67	0.42	1.77	2.60
平均 PTP-1 &2 或 PTP-i 或 PTP-v	1.099		0.892	0.952	0.941		0.742	0.754	1.140		0.949	0.969
% SITE OCCUPANCY			18.817	16.492			21.112	25.191			16.754	18.054
载体量: 38.71 ng/μg 周质蛋白												
位点占有率 PTP-i: 18.98 % - CV (%): 11.54												
位点占有率 PTP-v: 19.12% - CV (%): 23.29												

Hla-N,C (平均 MW: 37390.51 Da)												
	消化 1				消化 2				消化 3			
	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v	PTP-1	PTP-2	PTP-i	PTP-v
运行1	0.26	0.24	0.205		0.325	0.299	0.242		0.308	0.284	0.238	
运行2	0.257	0.241	0.2	<LLOQ	0.325	0.294	0.246	<LLOQ	0.315	0.281	0.241	
运行3	0.258	0.24	0.201		0.319	0.296	0.252		0.312	0.279	0.251	
CV(%)运行	0.59	0.24	1.31	-	1.07	0.85	2.04	-	1.13	0.89	2.80	-
平均 PTP-1 &2 或 PTP-i 或 PTP-v	0.249		0.202	-	0.310		0.247	-	0.297		0.243	
% SITE OCCUPANCY			18.98	>99%			20.34	>99%			17.93	>99%
载体量: 10.66 ng/μg 周质蛋白												
位点占有率 PTP-i: 18.98% - CV (%): 6.34												
位点占有率 PTP-v: >99%												

[0378] 本发明的各方面概述于以下编号的段落中：

[0379] 1.一种经修饰的载体蛋白，其修饰为包含一个或多个共有序列，所述共有序列包含或由以下氨基酸序列组成：

[0380] K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R

[0381] 其中X和Y独立地是除了脯氨酸之外的任何氨基酸，并且Z表示任何氨基酸，其中任选地X和Y独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸，Z表示除赖氨酸或精氨酸以外的任何氨基酸，且优选地Z表示除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸(例如SEQ ID NO:47)。

[0382] 2.如段落1所述的经修饰的载体蛋白，其中所述共有序列是氨基酸序列K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-K/R(SEQ ID NO:19)，其中X_1和X_2独立地是除了脯氨酸之外的任何氨基酸，并且其中X_1和X_2以及Z_1和Z_2优选不是赖氨酸或精氨酸，其中任选地，其中X_1和X_2以及Z_1和Z_2不是半胱氨酸、天冬酰胺、谷氨酰胺、甲硫氨酸或精氨酸。

[0383] 3.如段落1或段落2所述的经修饰的载体蛋白，其中所述共有序列包含或由SEQ ID NO:20，任选地SEQ ID No:42-45，任选地SEQ ID No:42-44中的任一个的氨基酸序列组成。

[0384] 4.如段落1-3中任一所述的经修饰的载体蛋白，其中所述共有序列(i)已经取代载体蛋白序列中的一个或多个氨基酸，或(ii)已经插入到载体蛋白序列中。

[0385] 5.如段落1-4中任一所述的经修饰的载体蛋白,其包含多于一个所述共有序列,任选地至少2、3、4或5个共有序列。

[0386] 6.如段落5所述的经修饰的载体蛋白,其中所有所述共有序列具有不同的氨基酸序列。

[0387] 7.如段落1-6中任一所述的经修饰的载体蛋白,其中所述载体蛋白是CRM197、来自破伤风杆菌的TT、来自铜绿假单胞菌的EPA、来自铜绿假单胞菌的Hcp1、来自金黄色葡萄球菌的Hla、来自金黄色葡萄球菌的ClfA、来自大肠杆菌的MBP、来自大肠杆菌的PspA或来自淋病奈瑟菌的MtrE。

[0388] 8.如段落7所述的经修饰的载体蛋白,其中所述载体蛋白包含SEQ ID NO:1至16的任一个的氨基酸序列,或与SEQ ID NO:1至16中任一个具有至少80%、85%、90%、92%、95%、96%、97%、98%或99%同一性的氨基酸序列,或由其组成。

[0389] 9.如段落8所述的经修饰的载体蛋白,其中所述经修饰的载体蛋白包含或由与SEQ ID NO:1至16中的任一个具有至少80%、85%、90%、92%、95%、96%或97%同一性的氨基酸序列组成。

[0390] 10.如段落1-9中任一所述的经修饰的载体蛋白,其中所述经修饰的载体蛋白是糖基化的。

[0391] 11.包含如段落1-10中任一所述的经修饰的载体蛋白的缀合物(例如生物缀合物),其中所述经修饰的载体蛋白连接至多糖。

[0392] 12.如段落11所述的缀合物(例如生物缀合物),其中所述多糖连接至所述经修饰的载体蛋白上的氨基酸,所述氨基酸选自天冬酰胺、天冬氨酸、谷氨酸、赖氨酸、半胱氨酸、酪氨酸、组氨酸、精氨酸或色氨酸(例如天冬酰胺)。

[0393] 13.如段落11或段落12所述的缀合物(例如生物缀合物),其中所述多糖是细菌荚膜多糖。

[0394] 14.如段落13所述的缀合物(例如生物缀合物),其中所述荚膜多糖选自:金黄色葡萄球菌5型荚膜糖、金黄色葡萄球菌8型荚膜糖、脑膜炎奈瑟菌血清群A荚膜糖(MenA)、脑膜炎奈瑟菌血清群C荚膜糖(MenC)、脑膜炎奈瑟菌血清群Y荚膜糖(MenY)、脑膜炎奈瑟菌血清群W荚膜糖(MenW)、流感嗜血杆菌b型荚膜糖(Hib)、B组链球菌I组荚膜糖、B组链球菌II组荚膜糖、B组链球菌III组荚膜糖、B组链球菌IV组荚膜糖、B组链球菌V组荚膜糖、来自伤寒沙门氏菌的Vi糖、脑膜炎奈瑟菌LPS(如L3和/或L2)、卡他拉莫菌LPS、流感嗜血杆菌LPS、志贺氏菌O-抗原、铜绿假单胞菌O-抗原、大肠杆菌O-抗原或肺炎链球菌荚膜多糖。

[0395] 15.如段落13或段落14的缀合物(例如生物缀合物),其中所述荚膜多糖来自与所述载体蛋白相同的生物体。

[0396] 16.如段落15所述的缀合物(例如生物缀合物),其中所述荚膜多糖来自与所述载体蛋白不同的生物体。

[0397] 17.编码段落1-10中任一所述的经修饰的载体蛋白的多核苷酸。

[0398] 18.包含段落17所述的多核苷酸的载体。

[0399] 19.一种宿主细胞,其包含:

[0400] a.一种或多种编码糖基转移酶的核酸;

[0401] b.编码寡糖基转移酶的核酸;

[0402] c.编码如段落1-10中任一所述的经修饰的载体蛋白的核酸;以及任选地

[0403] d.编码聚合酶(例如wzy)的核酸。

[0404] 20.如段落19所述的宿主细胞,其中所述宿主细胞包含(a)将己糖单糖衍生物组装到十一异戊烯基焦磷酸(Und-PP)上的糖基转移酶和(b)能够将单糖添加到在Und-PP上组装的己糖单糖衍生物的一种或多种糖基转移酶。

[0405] 21.如段落20所述的宿主细胞,其中所述将己糖单糖衍生物组装到Und-PP上的糖基转移酶对于所述宿主细胞是异源的和/或对于编码糖基转移酶的一种或多种基因是异源的,任选地其中所述将己糖单糖衍生物组装到Und-PP上的糖基转移酶来自埃希氏菌属物种、志贺氏菌属物种、克雷伯氏菌属物种、黄单胞菌属物种、沙门氏菌属物种、耶尔森菌属物种、气单胞菌属物种、弗朗西斯菌属物种、螺杆菌属物种、变形杆菌属物种、乳球菌属物种、乳杆菌属物种、假单胞菌属物种、棒杆菌属物种、链霉菌属物种、链球菌属物种、肠球菌属物种、葡萄球菌属物种、芽孢杆菌属物种、梭菌属物种、李斯特菌属物种或弯曲杆菌属物种,任选地wecA(例如来自大肠杆菌的wecA)。

[0406] 22.如段落19-21中任一所述的宿主细胞,其中所述己糖单糖衍生物是其中C-2位被乙酰胺基修饰的任何单糖,如N-乙酰葡糖胺(GlcNAc)、N-乙酰半乳糖胺(GalNAc)、2,4-二乙酰胺基-2,4,6-三脱氧己糖(DATDH)、N-乙酰基岩藻糖胺(FucNAc),或N-乙酰基异鼠李糖胺(QuiNAc)。

[0407] 23.如段落19-22中任一项所述的宿主细胞,其中所述一种或多种能够将单糖添加至在Und-pp上组装的己糖单糖衍生物的糖基转移酶是来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)或来自大肠杆菌O167的呋喃半乳糖基转移酶(wfdK),或者来自大肠杆菌O28的呋喃半乳糖基转移酶(wbeY)和来自大肠杆菌O167的呋喃半乳糖基转移酶(wfdK)。

[0408] 24.如段落19-23中任一项所述的宿主细胞,其中所述糖基转移酶包含能够将存在于供体寡糖或多糖的第一重复单元的还原端的己糖单糖添加至己糖单糖衍生物的糖基转移酶,任选地其中所述一种或多种能够将单糖添加至己糖单糖衍生物的糖基转移酶包含半乳糖基转移酶(wciP),任选地来自大肠杆菌O21,并且任选地包含能够将与存在于供体寡糖或多糖的第一重复单元的还原端处的己糖单糖相邻的单糖添加至存在于供体寡糖或多糖的第一重复单元的还原端处的己糖单糖的糖基转移酶,任选地葡糖基转移酶(wciQ),任选地来自大肠杆菌O21。

[0409] 25.如段落19-24中任一项所述的宿主细胞,其中所述寡糖基转移酶源自空肠弯曲杆菌,任选地其中所述寡糖基转移酶是空肠弯曲杆菌的PglB,任选地其中空肠弯曲杆菌的PglB基因被整合到宿主细胞基因组中,并且任选地其中宿主细胞的至少一个基因功能失活或缺失,任选地其中宿主细胞的waaL基因功能失活或缺失,任选地其中宿主细胞的waaL基因被编码寡糖其转移酶的核酸取代,任选地其中宿主细胞的waaL基因被空肠弯曲杆菌PglB取代。

[0410] 26.如段落19-25中任一项所述的宿主细胞,其中所述编码经修饰的载体蛋白的核酸在所述宿主细胞的质粒中。

[0411] 27.如段落19-26中任一项所述的宿主细胞,其中所述编码该修饰的载体蛋白的核酸被整合到所述宿主细胞的基因组中。

[0412] 28.如段落19-27中任一项所述的宿主细胞,其中所述宿主细胞是大肠杆菌。

[0413] 29.一种生产生物缀合物的方法,所述生物缀合物包含与糖连接的经修饰的载体蛋白,所述方法包括(i)在适于生产蛋白质的条件下培养如段落19-28中任一所述的宿主细胞,和(ii)分离所述生物缀合物。

[0414] 30.通过段落29的方法产生的生物缀合物,其中所述生物缀合物包含连接至经修饰的载体蛋白的多糖。

[0415] 31.免疫原性组合物,其包含段落1-10中任一所述的经修饰的载体蛋白,或段落11-16中任一所述的缀合物或生物缀合物。

[0416] 32.一种制备段落31的免疫原性组合物的方法,其包括将经修饰的载体蛋白或缀合物或生物缀合物与药学上可接受的赋形剂或载体混合的步骤。

[0417] 33.一种疫苗,其包含段落31所述的免疫原性组合物和药学上可接受的赋形剂或载体。

[0418] 34.一种用于治疗或预防有需要的受试者中的细菌感染的方法,所述方法包括向所述受试者施用治疗有效量的如段落1-10中任一项所述的经修饰的载体蛋白,或段落11-16中任一项所述的缀合物或生物缀合物。

[0419] 35.一种免疫人宿主以抵抗细菌感染的方法,其包括向所述宿主施用免疫保护剂量的如段落1-10中任一项所述的经修饰的载体蛋白,或如段落11-16中任一项所述的缀合物或生物缀合物。

[0420] 36.一种在受试者中诱导针对细菌的免疫应答的方法,所述方法包括向受试者施用治疗或预防有效量的如段落1-10中任一项所述的经修饰的载体蛋白,或如段落11-16中任一项所述的缀合物或生物缀合物。

[0421] 37.如段落1-10中任一所述的经修饰的载体蛋白,或如段落11-16中任一所述的缀合物或生物缀合物在用于治疗或预防由细菌感染引起的疾病中的用途。

[0422] 38.如段落1-10中任一所述的经修饰的载体蛋白或如段落11-16中任一所述的缀合物或生物缀合物在制备用于治疗或预防由细菌感染引起的疾病的药物中的用途。

[0423] 39.如段落34-36中任一所述的方法,或用于如段落37所述的用途的载体蛋白、缀合物或生物缀合物,或如段落38所述的用途,其中所述细菌或细菌感染选自金黄色葡萄球菌、脑膜炎奈瑟菌、流感嗜血杆菌、流感嗜血杆菌b型、B组链球菌、伤寒沙门氏菌、卡他拉莫菌、福氏志贺氏菌、铜绿假单胞菌、大肠杆菌或肺炎链球菌。

[0424] 40.一种测量如段落1至10中任一所述的载体蛋白的糖基化位点占有率水平的方法,所述方法包括:用蛋白酶消化糖基化的载体蛋白;将经消化的蛋白质进行LC-MS;测定未经修饰的载体蛋白的浓度U;测定总载体蛋白的浓度T;以及根据以下等式计算糖基化位点占有率:

[0425] $$\text{位点占有率}(\%) = \frac{(\text{总} - \text{未经修饰的})\text{载体浓度}}{\text{总载体浓度}} \times 100$$

[0426] 41.如段落40所述的方法,其中通过测定对应于共有序列的肽片段的浓度来测定未经修饰的载体蛋白的浓度U。

[0427] 42.如段落40或段落41所述的方法,其中所述总载体蛋白的浓度T通过测定所述载体蛋白独特的一个或多个肽片段的浓度进行测定。

[0428] 43.如段落40-42任一项所述的方法,其中所述蛋白酶是胰蛋白酶。

[0429] 序列表

[0430] SEQ ID NO:1成熟野生型EPA的氨基酸序列。粗体和下划线为置换/去除用于脱毒的残基。生物体:铜绿假单胞菌。

[0431]
AEEAFDLWNECAKACVLDLKDGVRSSRMSVDPAIADTNGQGVLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
ATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRWSEWASGKVLCLLDPLDGVYNYLAQQRCNLDDTWEG
KIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGYPVQRLVA
LYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTLAAAESERFVRQGTGNDEAGAASADVVSL
TCPVAAGECAGPADSGDALLERNYPTGAEFLGDGGDVSFSTRGTQNWTVERLLQAHRQLEERGYVFVGYHGTFL
EAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTL
AAPEAAGEVERLIGHPLPLRLDAITGPEEEGGR**LE**TILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQ
AISALPDYASQPGKPPREDLK

[0432] SEQ ID NO:2具有L552V/ΔE553脱毒突变(粗体,下划线)的EPA的氨基酸序列。人工序列。

[0433]
AEEAFDLWNECAKACVLDLKDGVRSSRMSVDPAIADTNGQGVLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
ATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRWSEWASGKVLCLLDPLDGVYNYLAQQRCNLDDTWEG
KIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGYPVQRLVA
LYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTLAAAESERFVRQGTGNDEAGAASADVVSL
TCPVAAGECAGPADSGDALLERNYPTGAEFLGDGGDVSFSTRGTQNWTVERLLQAHRQLEERGYVFVGYHGTFL
EAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTL
AAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQA
ISALPDYASQPGKPPREDLK

[0434] SEQ ID NO:3:不含起始甲硫氨酸的破伤风毒素前体TT(AA1-1315)。生物体:破伤风杆菌

[0435]
MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFNPPSSLIEGASEYYDPNY
LRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGNSYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAML
TNLIIFGPGPVLNKNEVRGIVLRVDNKNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLL
MHELIHVLHGLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDYKAIANKLSQ
VTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFTEIELGKKFNIKTRLSYFSMNHDPVKI
PNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVNTNAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDL
GGELCIKIKNEDLTFIAEKNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTKGIPYA
PEYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYFPSVISKVNQGAQGILFLQWV
RDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQGYEGNFIGALETTGVVLLLEYIPEITLPVIAALSIAES
STQKEKIIKTIDNFLEKRYEKWIEVYKLVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQ
IADEINNLKNKLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANSKFIGITELKKL
ESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVILKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIH
LVNNESSEVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLI
WTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIREDNNITLKLDRCN
NNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAP
SYTNGKLNIYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAP
GIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQIGNDPNRDILIASNWYFNHLKDKILGCDWYFVPTD
EGWTND

[0436] SEQ ID NO:4白喉毒素(DT)。生物体:白喉杆菌

[0437]
GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLS
GKAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVE
YINNWEQAKALSVELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKE
HGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADN
LEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQV
VHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVN
KSKTHISVNGRKIRMRCRAIDGDVTFCRPKSPVYVGNGVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTV
DHTKVNSKLSLFFEIKS

[0438] SEQ ID NO:5:CRM197,白喉毒素的无毒突变体。人工序列。

[0439] GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKEFYSTDNKYDAAGYSVDNENPLS
GKAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVE
YINNWEQAKALSVELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKE
HGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADN
LEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAAYNFVESIINLFQV
VHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVN
KSKTHISVNGRKIRMRCRAIDGDVTFCRPKSPVYVGNGVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTV
DHTKVNSKLSLFFEIKS

[0440] SEQ ID NO:6:Hcp1。生物体:铜绿假单胞菌

MAVDMFIKIGDVKGESKDKTHAEEIDVLAWSWGMSQSGSMHMGGGGGAGKVNVQDLSFTKYIDKSTPNLMMACS
SGKHYPQAKLTIRKAGGENQVEYLIITLKEVLVSSVSTGGSGGEDRLTENVTLNFAQVQVDYQPQKADGAKDGG
PIKYGWNIRQNVQA

[0441] SEQ ID NO:7:PspA,不含起始甲硫氨酸的噬菌体休克蛋白A。

[0442] 生物体:大肠杆菌

[0443] GIFSRFADIVNANINALLEKAEDPQKLVRLMIQEMEDTLVEVRSTSARALAEKKQLTRRIEQASAREVEWQEKA
ELALLKEREDLARAALIEKQKLTDLIKSLEHEVTLVDDTLARMKKEIGELENKLSETRARQQALMLRHQAANSS
RDVRRQLDSGKLDEAMARFESFERRIDQMEAEAESHSFGKQKSLDDQFAELKADDAISEQLAQLKAKMKQDNQ

[0444] SEQ ID NO:8:MBP,麦芽糖/麦芽糊精结合蛋白。生物体:大肠杆菌

[0445] MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFP
QVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLL
PNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLV
DLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAAS
PNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVR
TAVINAASGRQTVDEALKDAQTRITK

[0446] SEQ ID NO:9:成熟mtrE,膜转运蛋白E。生物体:淋病奈瑟菌

[0447] MIPQYEQPKVEVAETFQNDTSVSSIRAVDLGWHDYFADPRLQKLIDIALERNTSLRTAVLNSEIYRKQYMIERN
NLLPTLAANANGSRQGSLSGGNVSSSYNVGLGAASYELDLFGRVRSSSEAALQGYFASVANRDAAHLSLIATVA
KAYFNERYAEEAMSLAQRVLKTREETYNAVRIAVQGRRDFRRRPAPAEALIESAKADYAHAARSREQARNALAT
LINRPIPEDLPAGLPLDKQFFVEKLPAGLSSEVLLDRPDIRAAEHALKQANANIGAARAAFFPSIRLTGSVGTG
SVELGGLFKSGTGVWAFAPSITLPIFTWGTNKANLDVAKLRQQAQIVAYESAVQSAFQDVANALAAREQLDKAY
DALSKQSRASKEALRLVGLRYKHGVSGALDLLDAERSSYSAEGAALSAQLTRAENLADLYKALGGGLKRDTQTG
K

[0448] SEQ ID NO:10-野生型成熟ClfAN1N2N3。生物体:金黄色葡萄球菌。

[0449] ASENSVTQSDSASNESKSNDSSSVSAAPKTDDTNVSDTKTSSNTNNGETSVAQNPAQQETTQSSSTNATTEETP
VTGEATTTTTNQANTPATTQSSNTNAEELVNQTSNETTFNDTNTVSSVNSPQNSTNAENVSTTQDTSTEATPSN
NESAPQSTDASNKDVVNQAVNTSAPRMRAFSLAAVAADAPAAGTDITNQLTNVTVGIDSGTTVYPHQAGYVKLN
YGFSVPNSAVKGDTFKITVPKELNLNGVTSTAKVPPIMAGDQVLANGVIDSDGNVIYTFTDYVNTKDDVKATLT
MPAYIDPENVKKTGNVTLATGIGSTTANKTVLVDYEKYGKFYNLSIKGTIDQIDKTNNTYRQTIYVNPSGDNVI
APVLTGNLKPNTDSNALIDQQNTSIKVYKVDNAADLSESYFVNPENFEDVTNSVNITFPNPNQYKVEFNTPDDQ
ITTPYIVVVNGHIDPNSKGDLALRSTLYGYNSNIIWRSMSWDNEVAFNNGSGSGDGIDKPVVPEQPDEPGEIEP
IPED

[0450] SEQ ID NO:11-野生型成熟ClfAN2N3。生物体:金黄色葡萄球菌。

[0451] VAADAPAAGTDITNQLTNVTVGIDSGTTVYPHQAGYVKLNYGFSVPNSAVKGDTFKITVPKELNLNGVTSTAKV
PPIMAGDQVLANGVIDSDGNVIYTFTDYVNTKDDVKATLTMPAYIDPENVKKTGNVTLATGIGSTTANKTVLVD
YEKYGKFYNLSIKGTIDQIDKTNNTYRQTIYVNPSGDNVIAPVLTGNLKPNTDSNALIDQQNTSIKVYKVDNAA
DLSESYFVNPENFEDVTNSVNITFPNPNQYKVEFNTPDDQITTPYIVVVNGHIDPNSKGDLALRSTLYGYNSNI
IWRSMSWDNEVAFNNGSGSGDGIDKPVVPEQPDEPGEIEPIPED

[0452] SEQ ID NO:12-ClfAN2N3P116S/Y118A。人工序列。

[0453] VAADAPAAGTDITNQLTNVTVGIDSGTTVYPHQAGYVKLNYGFSVPNSAVKGDTFKITVPKELNLNGVTSTAKV
PPIMAGDQVLANGVIDSDGNVIYTFTDYVNTKDDVKATLTMSAAIDPENVKKTGNVTLATGIGSTTANKTVLVD
YEKYGKFYNLSIKGTIDQIDKTNNTYRQTIYVNPSGDNVIAPVLTGNLKPNTDSNALIDQQNTSIKVYKVDNAA
DLSESYFVNPENFEDVTNSVNITFPNPNQYKVEFNTPDDQITTPYIVVVNGHIDPNSKGDLALRSTLYGYNSNI
IWRSMSWDNEVAFNNGSGSGDGIDKPVVPEQPDEPGEIEP***I***PED

[0454] SEQ ID NO:13:成熟野生型H1a的氨基酸序列。生物体:金黄色葡萄球菌。

[0455] ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMHKKVFYSFIDDKNHNKKLLVIRTKGTIAGQYRVYSE EGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNGNVTGDDTGKIGGLIGANV SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNF LDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSE RYKIDWEKEEMTN

[0456] SEQ ID NO:14-成熟H1aH35L的氨基酸序列。人工序列。

[0457] ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNHNKKLLVIRTKGTIAGQYRVYSE EGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNGNVTGDDTGKIGGLIGANV SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNF LDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSE RYKIDWEKEEMTN

[0458] SEQ ID NO:15-成熟H1a H35L/H48C/G122C的氨基酸序列,人工序列。

[0459] ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKIGGLIGANVSIGHTLKY VQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSS GFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEKEEMTN

[0460] SEQ ID NO:16:H1aPSGS。人工序列。

[0461] ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMHKKVFYSFIDDKNHNKKLLVIRTKGTIAGQYRVYSEEGAN KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTPSGSVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYD RDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQ LHWTSTNWKGTNTKDKWIDRSSERYKIDWEKEEMTN

[0462] SEQ ID NO:17:最小PglB糖基化位点共有序列。人工序列。

[0463] D/E-X_1-N-X_2-S/T

[0464] 其中X_1和X_2是除脯氨酸之外的任何氨基酸。

[0465] SEQ ID NO:18:全PglB糖基化位点共有序列。人工序列

[0466] K-D/E-X_1-N-X_2-S/T-K

[0467] 其中X_1和X_2是除脯氨酸之外的任何氨基酸。

[0468] SEQ ID NO:19:MS定量相容的PglB糖基化位点共有序列。人工序列。

[0469] K/R-D/E-X_1-N-X_2-S/T-Z_1-Z_2-R/K

[0470] 其中X_1和X_2是除脯氨酸之外的任何氨基酸,并且Z_1和Z_2不是赖氨酸或精氨酸。

[0471] SEQ ID NO:20:MS定量相容的PglB糖基化位点共有序列。人工序列。

[0472] K-D/E-X_1-N-X_2-S/T-S-A-R

[0473] 其中X_1和X_2是除脯氨酸之外的任何氨基酸。

[0474] SEQ ID NO:21-Flgl信号序列。生物体:福氏志贺氏菌

[0475] MIKFLSALILLLVTTAAQA

[0476] SEQ ID NO:22-OmpA信号序列。生物体:大肠杆菌

[0477] MKKTAIAIAVALAGFATVAQA

[0478] SEQ ID NO:23-MalE信号序列。生物体:大肠杆菌

[0479] MKIKTGARILALSALTTMMFSASALA

[0480] SEQ ID NO:24-PelB信号序列。生物体:胡萝卜软腐果胶杆菌(胡萝卜软腐欧文氏菌(Erwinia carotovora))。

[0481] MKYLLPTAAAGLLLLAAQPAMA

[0482] SEQ ID NO:25-LTllb信号序列。生物体:大肠杆菌

[0483] MSFKKIIKAFVIMAALVSVQAHA

[0484] SEQ ID NO:26-XynA信号序列。生物体:枯草芽孢杆菌

[0485] MFKFKKKFLVGLTAAFMSISMFSATASA

[0486] SEQ ID NO:27-DsbA信号序列。生物体:大肠杆菌

[0487] MKKIWLALAGLVLAFSASA

[0488] SEQ ID NO:28-TolB信号序列。生物体:大肠杆菌

[0489] MKQALRVAFGFLILWASVLHA

[0490] SEQ ID NO:29-SipA信号序列。生物体:无乳链球菌

[0491] MKMNKKVLLTSTMAASLLSVASVQAS

[0492] SEQ ID NO:30:具有N末端S、Flg1信号序列、C-末端GSHRHR和取代残基K131的KDSNITSAR的Hla H35L/H48C/G122C的氨基酸序列。人工序列。

[0493] MIKFLSALILLLVTTAAQASADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLV IRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDD TGKDQNRTKIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLF MKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNT KDKWIDRSSERYKIDWEKEEMTNGSHRHR

[0494] SEQ ID NO:31:具有N-末端S、Flg1信号序列、C-末端GSHRHR和取代残基K131的KDSNITSAR的成熟Hla H35L/G122C/H48C的氨基酸序列,人工序列。

[0495] MIKFLSALILLLVTTAAQASADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLV IRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDD TGKDSNITSARIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQ LFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGT NTKDKWIDRSSERYKIDWEKEEMTNGSHRHR

[0496] SEQ ID NO:32:具有N-末端S、Flg1信号序列、C-末端GSHRHR和取代残基K131的KDSNSTSAR的成熟Hla H35L/G122C/H48C的氨基酸序列,人工序列。

[0497] MIKFLSALILLLVTTAAQASADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLV IRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDD TGKDSNSTSARIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQ LFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGT NTKDKWIDRSSERYKIDWEKEEMTNGSHRHR

[0498] SEQ ID NO:33:具有N-末端S、Flg1信号序列、C-末端GSHRHR和取代残基K131的KDSNVTSAR的成熟Hla H35L/G122C/H48C的氨基酸序列,人工序列。

[0499] MIKFLSALILLLVTTAAQASADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLV IRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDD TGKDSNVTSARIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQ LFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGT NTKDKWIDRSSERYKIDWEKEEMTNGSHRHR

[0500] SEQ ID NO:34:具有N-末端S、Flg1信号序列、C-末端GSHRHR和取代残基K131的KDSNATSAR的成熟Hla H35L/G122C/H48C的氨基酸序列。人工序列。

[0501] MIKFLSALILLLVTTAAQASADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLV IRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDD TGKDSNVTSARIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQ LFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGT NTKDKWIDRSSERYKIDWEKEEMTNGSHRHR

[0502] SEQ ID NO:35:用KDQNRTK取代残基K131的成熟Hla H35L/H48C/G122C的氨基酸序列。人工序列。

[0503] ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDQNRTKIGGLIGANVSI GHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKA SSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEKEE MTN

[0504] SEQ ID NO:36:用KDSNITSAR取代残基K131的成熟Hla H35L/G122C/H48C的氨基酸序列。人工序列。

[0505]
ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN
KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDSNITSARIGGLIGANV
SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPN
KASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEK
EEMTN

[0506] SEQ ID NO:37:用KDSNSTSAR取代残基K131的成熟Hla H35L/G122C/H48C的氨基酸序列。人工序列。

[0507]
ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN
KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDSNSTSARIGGLIGANV
SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPN
KASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEK
EEMTN

[0508] SEQ ID NO:38:用KDSNVTSAR取代残基K131的成熟Hla H35L/G122C/H48C的氨基酸序列。人工序列。

[0509]
ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN
KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDSNVTSARIGGLIGANV
SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPN
KASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEK
EEMTN

[0510] SEQ ID NO:39:用KDSNATSAR取代残基K131的成熟Hla H35L/G122C/H48C的氨基酸序列。人工序列。

[0511]
ADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGAN
KSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDSNVTSARIGGLIGANV
SIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPN
KASSLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEK
EEMTN

[0512] SEQ ID NO:40:KDQNRTK糖基化位点。人工序列。

[0513] KDQNRTK

[0514] SEQ ID NO:41:KDQNATK糖基化位点。人工序列。

[0515] KDQNATK

[0516] SEQ ID NO:42:KDSNITSAR糖基化位点。人工序列。

[0517] KDSNITSAR

[0518] SEQ ID NO:43:KDSNSTSAR糖基化位点。人工序列。

[0519] KDSNSTSAR

[0520] SEQ ID NO:44:KDSNVTSAR糖基化位点。人工序列。

[0521] KDSNVTSAR

[0522] SEQ ID NO:45:KDSNATSAR糖基化位点。人工序列。

[0523] KDSNATSAR

[0524] SEQ ID NO:46:MS定量相容的Pg1B糖基化位点共有序列。人工序列。

[0525] K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R

[0526] 其中X和Y独立地是除了脯氨酸、赖氨酸或精氨酸之外的任何氨基酸，并且Z表示除了半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸。

[0527] SEQ ID NO:47:MS定量相容的Pg1B糖基化位点共有序列。人工序列。

[0528] K/R-Z_{0-9}-D/E-X-N-Y-S/T-Z_{0-9}-K/R

[0529] 其中X和Y独立地是除了脯氨酸、半胱氨酸、甲硫氨酸、天冬酰胺或谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸，并且Z表示除了半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸。

[0530] SEQ ID NO:48:命名为PTP-2的肽42T-50K。生物体:金黄色葡萄球菌。

[0531] TGDLVTYK

[0532] SEQ ID NO:49:命名为PTP-3的肽225A-234K。生物体:金黄色葡萄球菌。

[0533] AADNFLDPNK

[0534] SEQ IN NO:50:间隔区

[0535] GSGGG

[0536] SEQ IN NO:51:具有N-末端S、取代残基K131的KDSNITSAR糖基化位点;在N-末端处糖基化位点KDSNVTSAR及在此糖基化位点之前和之后的GSGGG间隔区;Flg1信号序列;和C-末端的His标签的成熟Hla H35L/G122C/H48C(以Ala-21起始,粗体)的氨基酸序列。人工序列。

[0537] MIKFLSALILLLVTTAAQASAGSGGGKDSNVTSARGSGGGKLADSDINIKTGTTDIGSNTTVKTGDLV TYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDN EVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGKDSNITSARIGGLIGANVSIGHTLKYVQPDF KTILESPTDKKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKAS SLLSSGFSPDFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKI DWEKEEMTNGSHHHHHH

[0538] SEQ ID NO:52:具有N-末端S、取代残基K131的KDSNITSAR糖基化位点;在C-末端处的糖基化位点KDSNVTSAR与在此糖基化位点之前的GSGGG间隔区;Flg1信号序列;和C-末端处的His标签的成熟Hla H35L/G122C/H48C(以Ala-21起始,粗体)的氨基酸序列。人工序列。

[0539] MIKFLSALILLLVTTAAQASAADSDINIKTGTTDIGSNTTVKTGDLVTYDKENGMLKKVFYSFIDDKNC NKKLLVIRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDNEVAQISDYYPRNSIDTKEYMST LTYGFNCNVTGDDTGKDSNITSARIGGLIGANVSIGHTLKYVQPDFKTILESPTDKKVGWKVIFNNMV NQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSPDFATVITMDRKAS KQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEKEEMTNLGSGGGKDSNV TSARGSHHHHHH

[0540] SEQ ID NO:53:具有N-末端S、C-末端处的KDSNITSAR糖基化位点(其之前有GSGGG间隔区);在N-末端处的糖基化位点KDSNVTSAR与在此糖基化位点之前和之后的GSGGG间隔区;Flg1信号序列;和C-末端处的His标签的成熟Hla H35L/G122C/H48C(以Ala-21起始,粗体)的氨基酸序列。人工序列。

[0541] MIKFLSALILLLVTTAAQASAGSGGGKDSNVTSARGSGGGKLADSDINIKTGTTDIGSNTTVKTGDLV TYDKENGMLKKVFYSFIDDKNCNKKLLVIRTKGTIAGQYRVYSEEGANKSGLAWPSAFKVQLQLPDN EVAQISDYYPRNSIDTKEYMSTLTYGFNCNVTGDDTGIGGLIGANVSIGHTLKYVQPDFKTILESPTD KKVGWKVIFNNMVNQNWGPYDRDSWNPVYGNQLFMKTRNGSMKAADNFLDPNKASSLLSSGFSP DFATVITMDRKASKQQTNIDVIYERVRDDYQLHWTSTNWKGTNTKDKWIDRSSERYKIDWEKEEMT NLGSGGGKDSNITSARGSHHHHHH

序列表

<110> 葛兰素史克生物有限公司

<120> 生物缀合物糖基化的定量

<130> VB66747 WO

<150> 19187963.4
<151> 2019-07-23

<160> 53

<170> PatentIn version 3.5

<210> 1
<211> 613
<212> PRT
<213> 铜绿假单胞菌

<400> 1

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1 5 10 15

[0001]

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
20 25 30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
35 40 45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
50 55 60

Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu
65 70 75 80

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85 90 95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100 105 110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His

115 120 125

Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
130 135 140

Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
145 150 155 160

Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
165 170 175

Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
180 185 190

Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr
195 200 205

Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile
[0002] 210 215 220

Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys
225 230 235 240

Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu
245 250 255

Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Ala Phe
260 265 270

Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys Gly
275 280 285

Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu Ser
290 295 300

Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro Gly
305 310 315 320

Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala
325 330 335

Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg
340 345 350

Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val
355 360 365

Ser Leu Thr Cys Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp
370 375 380

Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe
385 390 395 400

Leu Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn
405 410 415

[0003]

Trp Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg
420 425 430

Gly Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln
435 440 445

Ser Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala
450 455 460

Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly
465 470 475 480

Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly
485 490 495

Ala Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr
500 505 510

Arg Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu
515 520 525

```
Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly
    530                 535                 540

Pro Glu Glu Glu Gly Gly Arg Leu Glu Thr Ile Leu Gly Trp Pro Leu
545                 550                 555                 560

Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg
                565                 570                 575

Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln
            580                 585                 590

Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro
        595                 600                 605

Arg Glu Asp Leu Lys
    610
```

[0004]

<210> 2
<211> 612
<212> PRT
<213> 人工序列

<220>
<223> 具有 L552V/ΔE553 脱毒突变的 EPA 的氨基酸序列

<400> 2

```
Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1               5                   10                  15

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
            20                  25                  30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
        35                  40                  45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
    50                  55                  60
```

Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu
65 70 75 80

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85 90 95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100 105 110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
115 120 125

Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
130 135 140

Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
145 150 155 160

[0005]

Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
165 170 175

Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
180 185 190

Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr
195 200 205

Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile
210 215 220

Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys
225 230 235 240

Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu
245 250 255

Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Ala Phe
260 265 270

Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys Gly
275 280 285

Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu Ser
290 295 300

Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro Gly
305 310 315 320

Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala
325 330 335

Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg
340 345 350

Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val
355 360 365

[0006]

Ser Leu Thr Cys Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp
370 375 380

Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe
385 390 395 400

Leu Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn
405 410 415

Trp Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg
420 425 430

Gly Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln
435 440 445

Ser Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala
450 455 460

Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly
465 470 475 480

Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly
485 490 495

Ala Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr
500 505 510

Arg Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu
515 520 525

Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly
530 535 540

Pro Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala
545 550 555 560

Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn
565 570 575

[0007] Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala
580 585 590

Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg
595 600 605

Glu Asp Leu Lys
610

<210> 3
<211> 1315
<212> PRT
<213> 破伤风杆菌

<400> 3

Met Pro Ile Thr Ile Asn Asn Phe Arg Tyr Ser Asp Pro Val Asn Asn
1 5 10 15

Asp Thr Ile Ile Met Met Glu Pro Pro Tyr Cys Lys Gly Leu Asp Ile
20 25 30

```
Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Val Pro Glu
        35                  40                  45

Arg Tyr Glu Phe Gly Thr Lys Pro Glu Asp Phe Asn Pro Pro Ser Ser
    50                  55                  60

Leu Ile Glu Gly Ala Ser Glu Tyr Tyr Asp Pro Asn Tyr Leu Arg Thr
65                  70                  75                  80

Asp Ser Asp Lys Asp Arg Phe Leu Gln Thr Met Val Lys Leu Phe Asn
                85                  90                  95

Arg Ile Lys Asn Asn Val Ala Gly Glu Ala Leu Leu Asp Lys Ile Ile
            100                 105                 110

Asn Ala Ile Pro Tyr Leu Gly Asn Ser Tyr Ser Leu Leu Asp Lys Phe
        115                 120                 125
```

[0008]

```
Asp Thr Asn Ser Asn Ser Val Ser Phe Asn Leu Leu Glu Gln Asp Pro
    130                 135                 140

Ser Gly Ala Thr Thr Lys Ser Ala Met Leu Thr Asn Leu Ile Ile Phe
145                 150                 155                 160

Gly Pro Gly Pro Val Leu Asn Lys Asn Glu Val Arg Gly Ile Val Leu
                165                 170                 175

Arg Val Asp Asn Lys Asn Tyr Phe Pro Cys Arg Asp Gly Phe Gly Ser
            180                 185                 190

Ile Met Gln Met Ala Phe Cys Pro Glu Tyr Val Pro Thr Phe Asp Asn
        195                 200                 205

Val Ile Glu Asn Ile Thr Ser Leu Thr Ile Gly Lys Ser Lys Tyr Phe
    210                 215                 220

Gln Asp Pro Ala Leu Leu Leu Met His Glu Leu Ile His Val Leu His
```

225 230 235 240

Gly Leu Tyr Gly Met Gln Val Ser Ser His Glu Ile Ile Pro Ser Lys
245 250 255

Gln Glu Ile Tyr Met Gln His Thr Tyr Pro Ile Ser Ala Glu Glu Leu
260 265 270

Phe Thr Phe Gly Gly Gln Asp Ala Asn Leu Ile Ser Ile Asp Ile Lys
275 280 285

Asn Asp Leu Tyr Glu Lys Thr Leu Asn Asp Tyr Lys Ala Ile Ala Asn
290 295 300

Lys Leu Ser Gln Val Thr Ser Cys Asn Asp Pro Asn Ile Asp Ile Asp
305 310 315 320

Ser Tyr Lys Gln Ile Tyr Gln Gln Lys Tyr Gln Phe Asp Lys Asp Ser
[0009] 325 330 335

Asn Gly Gln Tyr Ile Val Asn Glu Asp Lys Phe Gln Ile Leu Tyr Asn
340 345 350

Ser Ile Met Tyr Gly Phe Thr Glu Ile Glu Leu Gly Lys Lys Phe Asn
355 360 365

Ile Lys Thr Arg Leu Ser Tyr Phe Ser Met Asn His Asp Pro Val Lys
370 375 380

Ile Pro Asn Leu Leu Asp Asp Thr Ile Tyr Asn Asp Thr Glu Gly Phe
385 390 395 400

Asn Ile Glu Ser Lys Asp Leu Lys Ser Glu Tyr Lys Gly Gln Asn Met
405 410 415

Arg Val Asn Thr Asn Ala Phe Arg Asn Val Asp Gly Ser Gly Leu Val
420 425 430

Ser Lys Leu Ile Gly Leu Cys Lys Lys Ile Ile Pro Pro Thr Asn Ile
435 440 445

Arg Glu Asn Leu Tyr Asn Arg Thr Ala Ser Leu Thr Asp Leu Gly Gly
450 455 460

Glu Leu Cys Ile Lys Ile Lys Asn Glu Asp Leu Thr Phe Ile Ala Glu
465 470 475 480

Lys Asn Ser Phe Ser Glu Glu Pro Phe Gln Asp Glu Ile Val Ser Tyr
485 490 495

Asn Thr Lys Asn Lys Pro Leu Asn Phe Asn Tyr Ser Leu Asp Lys Ile
500 505 510

Ile Val Asp Tyr Asn Leu Gln Ser Lys Ile Thr Leu Pro Asn Asp Arg
515 520 525

[0010]

Thr Thr Pro Val Thr Lys Gly Ile Pro Tyr Ala Pro Glu Tyr Lys Ser
530 535 540

Asn Ala Ala Ser Thr Ile Glu Ile His Asn Ile Asp Asp Asn Thr Ile
545 550 555 560

Tyr Gln Tyr Leu Tyr Ala Gln Lys Ser Pro Thr Thr Leu Gln Arg Ile
565 570 575

Thr Met Thr Asn Ser Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile
580 585 590

Tyr Ser Tyr Phe Pro Ser Val Ile Ser Lys Val Asn Gln Gly Ala Gln
595 600 605

Gly Ile Leu Phe Leu Gln Trp Val Arg Asp Ile Ile Asp Asp Phe Thr
610 615 620

Asn Glu Ser Ser Gln Lys Thr Thr Ile Asp Lys Ile Ser Asp Val Ser
625 630 635 640

Thr Ile Val Pro Tyr Ile Gly Pro Ala Leu Asn Ile Val Lys Gln Gly
645 650 655

Tyr Glu Gly Asn Phe Ile Gly Ala Leu Glu Thr Thr Gly Val Val Leu
660 665 670

Leu Leu Glu Tyr Ile Pro Glu Ile Thr Leu Pro Val Ile Ala Ala Leu
675 680 685

Ser Ile Ala Glu Ser Ser Thr Gln Lys Glu Lys Ile Ile Lys Thr Ile
690 695 700

Asp Asn Phe Leu Glu Lys Arg Tyr Glu Lys Trp Ile Glu Val Tyr Lys
705 710 715 720

Leu Val Lys Ala Lys Trp Leu Gly Thr Val Asn Thr Gln Phe Gln Lys
725 730 735

[0011] Arg Ser Tyr Gln Met Tyr Arg Ser Leu Glu Tyr Gln Val Asp Ala Ile
740 745 750

Lys Lys Ile Ile Asp Tyr Glu Tyr Lys Ile Tyr Ser Gly Pro Asp Lys
755 760 765

Glu Gln Ile Ala Asp Glu Ile Asn Asn Leu Lys Asn Lys Leu Glu Glu
770 775 780

Lys Ala Asn Lys Ala Met Ile Asn Ile Asn Ile Phe Met Arg Glu Ser
785 790 795 800

Ser Arg Ser Phe Leu Val Asn Gln Met Ile Asn Glu Ala Lys Lys Gln
805 810 815

Leu Leu Glu Phe Asp Thr Gln Ser Lys Asn Ile Leu Met Gln Tyr Ile
820 825 830

Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu Lys Lys Leu Glu
835 840 845

```
Ser Lys Ile Asn Lys Val Phe Ser Thr Pro Ile Pro Phe Ser Tyr Ser
    850                 855                 860

Lys Asn Leu Asp Cys Trp Val Asp Asn Glu Glu Asp Ile Asp Val Ile
865                 870                 875                 880

Leu Lys Lys Ser Thr Ile Leu Asn Leu Asp Ile Asn Asn Asp Ile Ile
                885                 890                 895

Ser Asp Ile Ser Gly Phe Asn Ser Ser Val Ile Thr Tyr Pro Asp Ala
            900                 905                 910

Gln Leu Val Pro Gly Ile Asn Gly Lys Ala Ile His Leu Val Asn Asn
        915                 920                 925

Glu Ser Ser Glu Val Ile Val His Lys Ala Met Asp Ile Glu Tyr Asn
    930                 935                 940
```

[0012]

```
Asp Met Phe Asn Asn Phe Thr Val Ser Phe Trp Leu Arg Val Pro Lys
945                 950                 955                 960

Val Ser Ala Ser His Leu Glu Gln Tyr Gly Thr Asn Glu Tyr Ser Ile
                965                 970                 975

Ile Ser Ser Met Lys Lys His Ser Leu Ser Ile Gly Ser Gly Trp Ser
            980                 985                 990

Val Ser Leu Lys Gly Asn Asn Leu  Ile Trp Thr Leu Lys  Asp Ser Ala
        995                 1000                1005

Gly Glu  Val Arg Gln Ile Thr  Phe Arg Asp Leu Pro  Asp Lys Phe
    1010                1015                1020

Asn Ala  Tyr Leu Ala Asn Lys  Trp Val Phe Ile Thr  Ile Thr Asn
    1025                1030                1035

Asp Arg  Leu Ser Ser Ala Asn  Leu Tyr Ile Asn Gly  Val Leu Met
```

1040 1045 1050

Gly Ser Ala Glu Ile Thr Gly Leu Gly Ala Ile Arg Glu Asp Asn
1055 1060 1065

Asn Ile Thr Leu Lys Leu Asp Arg Cys Asn Asn Asn Asn Gln Tyr
1070 1075 1080

Val Ser Ile Asp Lys Phe Arg Ile Phe Cys Lys Ala Leu Asn Pro
1085 1090 1095

Lys Glu Ile Glu Lys Leu Tyr Thr Ser Tyr Leu Ser Ile Thr Phe
1100 1105 1110

Leu Arg Asp Phe Trp Gly Asn Pro Leu Arg Tyr Asp Thr Glu Tyr
1115 1120 1125

Tyr Leu Ile Pro Val Ala Ser Ser Ser Lys Asp Val Gln Leu Lys
[0013] 1130 1135 1140

Asn Ile Thr Asp Tyr Met Tyr Leu Thr Asn Ala Pro Ser Tyr Thr
1145 1150 1155

Asn Gly Lys Leu Asn Ile Tyr Tyr Arg Arg Leu Tyr Asn Gly Leu
1160 1165 1170

Lys Phe Ile Ile Lys Arg Tyr Thr Pro Asn Asn Glu Ile Asp Ser
1175 1180 1185

Phe Val Lys Ser Gly Asp Phe Ile Lys Leu Tyr Val Ser Tyr Asn
1190 1195 1200

Asn Asn Glu His Ile Val Gly Tyr Pro Lys Asp Gly Asn Ala Phe
1205 1210 1215

Asn Asn Leu Asp Arg Ile Leu Arg Val Gly Tyr Asn Ala Pro Gly
1220 1225 1230

Ile Pro Leu Tyr Lys Lys Met Glu Ala Val Lys Leu Arg Asp Leu
1235 1240 1245

Lys Thr Tyr Ser Val Gln Leu Lys Leu Tyr Asp Asp Lys Asn Ala
1250 1255 1260

Ser Leu Gly Leu Val Gly Thr His Asn Gly Gln Ile Gly Asn Asp
1265 1270 1275

Pro Asn Arg Asp Ile Leu Ile Ala Ser Asn Trp Tyr Phe Asn His
1280 1285 1290

Leu Lys Asp Lys Ile Leu Gly Cys Asp Trp Tyr Phe Val Pro Thr
1295 1300 1305

Asp Glu Gly Trp Thr Asn Asp
1310 1315

[0014] <210> 4
<211> 535
<212> PRT
<213> 白喉棒状杆菌

<400> 4

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45

Asp Trp Lys Gly Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60

Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80

Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95

Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110

Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140

Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160

Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175

[0015]

Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180 185 190

Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195 200 205

Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210 215 220

Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225 230 235 240

Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245 250 255

His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
260 265 270

Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val
275 280 285

Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu
290 295 300

Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala
305 310 315 320

Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser
325 330 335

Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp
340 345 350

Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe
355 360 365

Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His
370 375 380

[0016]

Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr
385 390 395 400

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His
405 410 415

Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val
420 425 430

Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr
435 440 445

His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile
450 455 460

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly
465 470 475 480

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu
500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser
515 520 525

Leu Phe Phe Glu Ile Lys Ser
530 535

<210> 5
<211> 535
<212> PRT
<213> 人工序列

<220>
<223> CRM197, 白喉毒素的非毒性突变体

<400> 5

[0017]

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45

Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60

Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80

Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95

Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110

Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140

Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160

Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175

Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180 185 190

Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195 200 205

[0018]

Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210 215 220

Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225 230 235 240

Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245 250 255

His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
260 265 270

Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val
275 280 285

Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu
290 295 300

Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala
305 310 315 320

Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser
325 330 335

Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp
340 345 350

Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe
355 360 365

Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His
370 375 380

Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr
385 390 395 400

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His
405 410 415

[0019] Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val
420 425 430

Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr
435 440 445

His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile
450 455 460

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly
465 470 475 480

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu
500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser

515 520 525

Leu Phe Phe Glu Ile Lys Ser
530 535

<210> 6
<211> 162
<212> PRT
<213> 铜绿假单胞菌

<400> 6

Met Ala Val Asp Met Phe Ile Lys Ile Gly Asp Val Lys Gly Glu Ser
1 5 10 15

Lys Asp Lys Thr His Ala Glu Glu Ile Asp Val Leu Ala Trp Ser Trp
20 25 30

Gly Met Ser Gln Ser Gly Ser Met His Met Gly Gly Gly Gly Gly Ala
35 40 45

[0020]

Gly Lys Val Asn Val Gln Asp Leu Ser Phe Thr Lys Tyr Ile Asp Lys
50 55 60

Ser Thr Pro Asn Leu Met Met Ala Cys Ser Ser Gly Lys His Tyr Pro
65 70 75 80

Gln Ala Lys Leu Thr Ile Arg Lys Ala Gly Gly Glu Asn Gln Val Glu
85 90 95

Tyr Leu Ile Ile Thr Leu Lys Glu Val Leu Val Ser Ser Val Ser Thr
100 105 110

Gly Gly Ser Gly Gly Glu Asp Arg Leu Thr Glu Asn Val Thr Leu Asn
115 120 125

Phe Ala Gln Val Gln Val Asp Tyr Gln Pro Gln Lys Ala Asp Gly Ala
130 135 140

Lys Asp Gly Gly Pro Ile Lys Tyr Gly Trp Asn Ile Arg Gln Asn Val

145 150 155 160

Gln Ala

<210> 7
<211> 221
<212> PRT
<213> 大肠杆菌

<400> 7

Gly Ile Phe Ser Arg Phe Ala Asp Ile Val Asn Ala Asn Ile Asn Ala
1 5 10 15

Leu Leu Glu Lys Ala Glu Asp Pro Gln Lys Leu Val Arg Leu Met Ile
20 25 30

Gln Glu Met Glu Asp Thr Leu Val Glu Val Arg Ser Thr Ser Ala Arg
35 40 45

[0021]

Ala Leu Ala Glu Lys Lys Gln Leu Thr Arg Arg Ile Glu Gln Ala Ser
50 55 60

Ala Arg Glu Val Glu Trp Gln Glu Lys Ala Glu Leu Ala Leu Leu Lys
65 70 75 80

Glu Arg Glu Asp Leu Ala Arg Ala Ala Leu Ile Glu Lys Gln Lys Leu
85 90 95

Thr Asp Leu Ile Lys Ser Leu Glu His Glu Val Thr Leu Val Asp Asp
100 105 110

Thr Leu Ala Arg Met Lys Lys Glu Ile Gly Glu Leu Glu Asn Lys Leu
115 120 125

Ser Glu Thr Arg Ala Arg Gln Gln Ala Leu Met Leu Arg His Gln Ala
130 135 140

Ala Asn Ser Ser Arg Asp Val Arg Arg Gln Leu Asp Ser Gly Lys Leu

145 150 155 160

Asp Glu Ala Met Ala Arg Phe Glu Ser Phe Glu Arg Arg Ile Asp Gln
165 170 175

Met Glu Ala Glu Ala Glu Ser His Ser Phe Gly Lys Gln Lys Ser Leu
180 185 190

Asp Asp Gln Phe Ala Glu Leu Lys Ala Asp Asp Ala Ile Ser Glu Gln
195 200 205

Leu Ala Gln Leu Lys Ala Lys Met Lys Gln Asp Asn Gln
210 215 220

<210> 8
<211> 396
<212> PRT
<213> 大肠杆菌

<400> 8

[0022]

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
[0023] 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
                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
```

[0024]

```
<210>  9
<211>  445
<212>  PRT
<213>  淋病奈瑟菌

<400>  9

Met Ile Pro Gln Tyr Glu Gln Pro Lys Val Glu Val Ala Glu Thr Phe
1               5                   10                  15

Gln Asn Asp Thr Ser Val Ser Ser Ile Arg Ala Val Asp Leu Gly Trp
            20                  25                  30

His Asp Tyr Phe Ala Asp Pro Arg Leu Gln Lys Leu Ile Asp Ile Ala
        35                  40                  45

Leu Glu Arg Asn Thr Ser Leu Arg Thr Ala Val Leu Asn Ser Glu Ile
    50                  55                  60

Tyr Arg Lys Gln Tyr Met Ile Glu Arg Asn Asn Leu Leu Pro Thr Leu
65                  70                  75                  80
```

Ala Ala Asn Ala Asn Gly Ser Arg Gln Gly Ser Leu Ser Gly Gly Asn
85 90 95

Val Ser Ser Ser Tyr Asn Val Gly Leu Gly Ala Ala Ser Tyr Glu Leu
100 105 110

Asp Leu Phe Gly Arg Val Arg Ser Ser Ser Glu Ala Ala Leu Gln Gly
115 120 125

Tyr Phe Ala Ser Val Ala Asn Arg Asp Ala Ala His Leu Ser Leu Ile
130 135 140

Ala Thr Val Ala Lys Ala Tyr Phe Asn Glu Arg Tyr Ala Glu Glu Ala
145 150 155 160

Met Ser Leu Ala Gln Arg Val Leu Lys Thr Arg Glu Glu Thr Tyr Asn
165 170 175

[0025]

Ala Val Arg Ile Ala Val Gln Gly Arg Arg Asp Phe Arg Arg Arg Pro
180 185 190

Ala Pro Ala Glu Ala Leu Ile Glu Ser Ala Lys Ala Asp Tyr Ala His
195 200 205

Ala Ala Arg Ser Arg Glu Gln Ala Arg Asn Ala Leu Ala Thr Leu Ile
210 215 220

Asn Arg Pro Ile Pro Glu Asp Leu Pro Ala Gly Leu Pro Leu Asp Lys
225 230 235 240

Gln Phe Phe Val Glu Lys Leu Pro Ala Gly Leu Ser Ser Glu Val Leu
245 250 255

Leu Asp Arg Pro Asp Ile Arg Ala Ala Glu His Ala Leu Lys Gln Ala
260 265 270

Asn Ala Asn Ile Gly Ala Ala Arg Ala Ala Phe Phe Pro Ser Ile Arg
275 280 285

Leu Thr Gly Ser Val Gly Thr Gly Ser Val Glu Leu Gly Gly Leu Phe
290 295 300

Lys Ser Gly Thr Gly Val Trp Ala Phe Ala Pro Ser Ile Thr Leu Pro
305 310 315 320

Ile Phe Thr Trp Gly Thr Asn Lys Ala Asn Leu Asp Val Ala Lys Leu
325 330 335

Arg Gln Gln Ala Gln Ile Val Ala Tyr Glu Ser Ala Val Gln Ser Ala
340 345 350

Phe Gln Asp Val Ala Asn Ala Leu Ala Ala Arg Glu Gln Leu Asp Lys
355 360 365

Ala Tyr Asp Ala Leu Ser Lys Gln Ser Arg Ala Ser Lys Glu Ala Leu
370 375 380

[0026] Arg Leu Val Gly Leu Arg Tyr Lys His Gly Val Ser Gly Ala Leu Asp
385 390 395 400

Leu Leu Asp Ala Glu Arg Ser Ser Tyr Ser Ala Glu Gly Ala Ala Leu
405 410 415

Ser Ala Gln Leu Thr Arg Ala Glu Asn Leu Ala Asp Leu Tyr Lys Ala
420 425 430

Leu Gly Gly Gly Leu Lys Arg Asp Thr Gln Thr Gly Lys
435 440 445

<210> 10
<211> 522
<212> PRT
<213> 金黄色葡萄球菌

<400> 10

Ala Ser Glu Asn Ser Val Thr Gln Ser Asp Ser Ala Ser Asn Glu Ser
1 5 10 15

```
Lys Ser Asn Asp Ser Ser Ser Val Ser Ala Ala Pro Lys Thr Asp Asp
            20                  25                  30

Thr Asn Val Ser Asp Thr Lys Thr Ser Ser Asn Thr Asn Asn Gly Glu
        35                  40                  45

Thr Ser Val Ala Gln Asn Pro Ala Gln Gln Glu Thr Thr Gln Ser Ser
    50                  55                  60

Ser Thr Asn Ala Thr Thr Glu Glu Thr Pro Val Thr Gly Glu Ala Thr
65                  70                  75                  80

Thr Thr Thr Thr Asn Gln Ala Asn Thr Pro Ala Thr Thr Gln Ser Ser
                85                  90                  95

Asn Thr Asn Ala Glu Glu Leu Val Asn Gln Thr Ser Asn Glu Thr Thr
            100                 105                 110
```

[0027]

```
Phe Asn Asp Thr Asn Thr Val Ser Ser Val Asn Ser Pro Gln Asn Ser
        115                 120                 125

Thr Asn Ala Glu Asn Val Ser Thr Thr Gln Asp Thr Ser Thr Glu Ala
    130                 135                 140

Thr Pro Ser Asn Asn Glu Ser Ala Pro Gln Ser Thr Asp Ala Ser Asn
145                 150                 155                 160

Lys Asp Val Val Asn Gln Ala Val Asn Thr Ser Ala Pro Arg Met Arg
                165                 170                 175

Ala Phe Ser Leu Ala Ala Val Ala Ala Asp Ala Pro Ala Ala Gly Thr
            180                 185                 190

Asp Ile Thr Asn Gln Leu Thr Asn Val Thr Val Gly Ile Asp Ser Gly
        195                 200                 205

Thr Thr Val Tyr Pro His Gln Ala Gly Tyr Val Lys Leu Asn Tyr Gly
    210                 215                 220
```

Phe Ser Val Pro Asn Ser Ala Val Lys Gly Asp Thr Phe Lys Ile Thr
225 230 235 240

Val Pro Lys Glu Leu Asn Leu Asn Gly Val Thr Ser Thr Ala Lys Val
245 250 255

Pro Pro Ile Met Ala Gly Asp Gln Val Leu Ala Asn Gly Val Ile Asp
260 265 270

Ser Asp Gly Asn Val Ile Tyr Thr Phe Thr Asp Tyr Val Asn Thr Lys
275 280 285

Asp Asp Val Lys Ala Thr Leu Thr Met Pro Ala Tyr Ile Asp Pro Glu
290 295 300

Asn Val Lys Lys Thr Gly Asn Val Thr Leu Ala Thr Gly Ile Gly Ser
305 310 315 320

[0028] Thr Thr Ala Asn Lys Thr Val Leu Val Asp Tyr Glu Lys Tyr Gly Lys
325 330 335

Phe Tyr Asn Leu Ser Ile Lys Gly Thr Ile Asp Gln Ile Asp Lys Thr
340 345 350

Asn Asn Thr Tyr Arg Gln Thr Ile Tyr Val Asn Pro Ser Gly Asp Asn
355 360 365

Val Ile Ala Pro Val Leu Thr Gly Asn Leu Lys Pro Asn Thr Asp Ser
370 375 380

Asn Ala Leu Ile Asp Gln Gln Asn Thr Ser Ile Lys Val Tyr Lys Val
385 390 395 400

Asp Asn Ala Ala Asp Leu Ser Glu Ser Tyr Phe Val Asn Pro Glu Asn
405 410 415

Phe Glu Asp Val Thr Asn Ser Val Asn Ile Thr Phe Pro Asn Pro Asn

420 425 430

Gln Tyr Lys Val Glu Phe Asn Thr Pro Asp Asp Gln Ile Thr Thr Pro
435 440 445

Tyr Ile Val Val Val Asn Gly His Ile Asp Pro Asn Ser Lys Gly Asp
450 455 460

Leu Ala Leu Arg Ser Thr Leu Tyr Gly Tyr Asn Ser Asn Ile Ile Trp
465 470 475 480

Arg Ser Met Ser Trp Asp Asn Glu Val Ala Phe Asn Asn Gly Ser Gly
485 490 495

Ser Gly Asp Gly Ile Asp Lys Pro Val Val Pro Glu Gln Pro Asp Glu
500 505 510

Pro Gly Glu Ile Glu Pro Ile Pro Glu Asp
515 520

[0029]

<210> 11
<211> 340
<212> PRT
<213> 金黄色葡萄球菌

<400> 11

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

Thr Asn Val Thr Val Gly Ile Asp Ser Gly Thr Thr Val Tyr Pro His
20 25 30

Gln Ala Gly Tyr Val Lys Leu Asn Tyr Gly Phe Ser Val Pro Asn Ser
35 40 45

Ala Val Lys Gly Asp Thr Phe Lys Ile Thr Val Pro Lys Glu Leu Asn
50 55 60

Leu Asn Gly Val Thr Ser Thr Ala Lys Val Pro Pro Ile Met Ala Gly

65 70 75 80

Asp Gln Val Leu Ala Asn Gly Val Ile Asp Ser Asp Gly Asn Val Ile
85 90 95

Tyr Thr Phe Thr Asp Tyr Val Asn Thr Lys Asp Asp Val Lys Ala Thr
100 105 110

Leu Thr Met Pro Ala Tyr Ile Asp Pro Glu Asn Val Lys Lys Thr Gly
115 120 125

Asn Val Thr Leu Ala Thr Gly Ile Gly Ser Thr Thr Ala Asn Lys Thr
130 135 140

Val Leu Val Asp Tyr Glu Lys Tyr Gly Lys Phe Tyr Asn Leu Ser Ile
145 150 155 160

Lys Gly Thr Ile Asp Gln Ile Asp Lys Thr Asn Asn Thr Tyr Arg Gln
[0030] 165 170 175

Thr Ile Tyr Val Asn Pro Ser Gly Asp Asn Val Ile Ala Pro Val Leu
180 185 190

Thr Gly Asn Leu Lys Pro Asn Thr Asp Ser Asn Ala Leu Ile Asp Gln
195 200 205

Gln Asn Thr Ser Ile Lys Val Tyr Lys Val Asp Asn Ala Ala Asp Leu
210 215 220

Ser Glu Ser Tyr Phe Val Asn Pro Glu Asn Phe Glu Asp Val Thr Asn
225 230 235 240

Ser Val Asn Ile Thr Phe Pro Asn Pro Asn Gln Tyr Lys Val Glu Phe
245 250 255

Asn Thr Pro Asp Asp Gln Ile Thr Thr Pro Tyr Ile Val Val Val Asn
260 265 270

Gly His Ile Asp Pro Asn Ser Lys Gly Asp Leu Ala Leu Arg Ser Thr
275 280 285

Leu Tyr Gly Tyr Asn Ser Asn Ile Ile Trp Arg Ser Met Ser Trp Asp
290 295 300

Asn Glu Val Ala Phe Asn Asn Gly Ser Gly Ser Gly Asp Gly Ile Asp
305 310 315 320

Lys Pro Val Val Pro Glu Gln Pro Asp Glu Pro Gly Glu Ile Glu Pro
325 330 335

Ile Pro Glu Asp
340

<210> 12
<211> 340
<212> PRT
<213> 人工序列

[0031] <220>
<223> ClfAN2N3P116S/Y118A

<400> 12

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

Thr Asn Val Thr Val Gly Ile Asp Ser Gly Thr Thr Val Tyr Pro His
20 25 30

Gln Ala Gly Tyr Val Lys Leu Asn Tyr Gly Phe Ser Val Pro Asn Ser
35 40 45

Ala Val Lys Gly Asp Thr Phe Lys Ile Thr Val Pro Lys Glu Leu Asn
50 55 60

Leu Asn Gly Val Thr Ser Thr Ala Lys Val Pro Pro Ile Met Ala Gly
65 70 75 80

Asp Gln Val Leu Ala Asn Gly Val Ile Asp Ser Asp Gly Asn Val Ile

85 90 95

Tyr Thr Phe Thr Asp Tyr Val Asn Thr Lys Asp Asp Val Lys Ala Thr
100 105 110

Leu Thr Met Ser Ala Ala Ile Asp Pro Glu Asn Val Lys Lys Thr Gly
115 120 125

Asn Val Thr Leu Ala Thr Gly Ile Gly Ser Thr Thr Ala Asn Lys Thr
130 135 140

Val Leu Val Asp Tyr Glu Lys Tyr Gly Lys Phe Tyr Asn Leu Ser Ile
145 150 155 160

Lys Gly Thr Ile Asp Gln Ile Asp Lys Thr Asn Asn Thr Tyr Arg Gln
165 170 175

Thr Ile Tyr Val Asn Pro Ser Gly Asp Asn Val Ile Ala Pro Val Leu
[0032] 180 185 190

Thr Gly Asn Leu Lys Pro Asn Thr Asp Ser Asn Ala Leu Ile Asp Gln
195 200 205

Gln Asn Thr Ser Ile Lys Val Tyr Lys Val Asp Asn Ala Ala Asp Leu
210 215 220

Ser Glu Ser Tyr Phe Val Asn Pro Glu Asn Phe Glu Asp Val Thr Asn
225 230 235 240

Ser Val Asn Ile Thr Phe Pro Asn Pro Asn Gln Tyr Lys Val Glu Phe
245 250 255

Asn Thr Pro Asp Asp Gln Ile Thr Thr Pro Tyr Ile Val Val Val Asn
260 265 270

Gly His Ile Asp Pro Asn Ser Lys Gly Asp Leu Ala Leu Arg Ser Thr
275 280 285

Leu Tyr Gly Tyr Asn Ser Asn Ile Ile Trp Arg Ser Met Ser Trp Asp
290 295 300

Asn Glu Val Ala Phe Asn Asn Gly Ser Gly Ser Gly Asp Gly Ile Asp
305 310 315 320

Lys Pro Val Val Pro Glu Gln Pro Asp Glu Pro Gly Glu Ile Glu Pro
325 330 335

Ile Pro Glu Asp
340

<210> 13
<211> 293
<212> PRT
<213> 金黄色葡萄球菌

<400> 13

[0033]
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp
115 120 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
195 200 205

[0034] Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
260 265 270

Lys Trp Ile Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
275 280 285

Glu Glu Met Thr Asn
290

<210> 14
<211> 293
<212> PRT

〈213〉 人工序列

〈220〉

〈223〉 成熟 HlaH35L 的氨基酸序列

〈400〉 14

```
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1               5                  10                  15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
            20                  25                  30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
        35                  40                  45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
    50                  55                  60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65                  70                  75                  80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
                85                  90                  95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
            100                 105                 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp
        115                 120                 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
    130                 135                 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
145                 150                 155                 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
                165                 170                 175
```

[0035]

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
260 265 270

Lys Trp Ile Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
[0036] 275 280 285

Glu Glu Met Thr Asn
290

<210> 15
<211> 293
<212> PRT
<213> 人工序列

<220>
<223> 成熟 Hla H35L/H48C/G122C 的氨基酸序列

<400> 15

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys

35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
115 120 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
[0037] 130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
260 265 270

Lys Trp Ile Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
275 280 285

Glu Glu Met Thr Asn
290

<210> 16
<211> 258
<212> PRT
<213> 人工序列

<220>
<223> HlaPSGS

<400> 16

[0038]

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Pro Ser Gly

100 105 110

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

Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp
130 135 140

Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu
145 150 155 160

Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu
165 170 175

Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp
180 185 190

Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr
195 200 205

[0039]

Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His
210 215 220

Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile
225 230 235 240

Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met
245 250 255

Thr Asn

<210> 17
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
<221> 杂项特征
<222> (1)..(1)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (2)..(2)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (4)..(4)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (5)..(5)
<223> Xaa 可以是 Ser 或 Thr

<400> 17

Xaa Xaa Asn Xaa Xaa
1 5

[0040]

<210> 18
<211> 7
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
<221> 杂项特征
<222> (2)..(2)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (3)..(3)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (5)..(5)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (6)..(6)
<223> Xaa 可以是 Ser 或 Thr

<400> 18

Lys Xaa Xaa Asn Xaa Xaa Lys
1 5

<210> 19
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
<221> 杂项特征
<222> (1)..(1)
<223> Xaa 可以是 Lys 或 Arg

[0041] <220>
<221> 杂项特征
<222> (2)..(2)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (3)..(3)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (5)..(5)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (6)..(6)
<223> Xaa 可以是 Ser 或 Thr

<220>
<221> 杂项特征
<222> (7)..(7)
<223> Xaa 是赖氨酸或精氨酸之外的任何氨基酸

<220>

<221> 杂项特征
<222> (8)..(8)
<223> Xaa 是赖氨酸或精氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (9)..(9)
<223> Xaa 可以是 Lys 或 Arg

<400> 19

Xaa Xaa Xaa Asn Xaa Xaa Xaa Xaa Xaa
1 5

<210> 20
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
[0042] <221> 杂项特征
<222> (2)..(2)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (3)..(3)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (5)..(5)
<223> Xaa 是脯氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (6)..(6)
<223> Xaa 可以是 Ser 或 Thr

<400> 20

Lys Xaa Xaa Asn Xaa Xaa Ser Ala Arg
1 5

<210> 21

<210> 19
<211> PRT
<212> 大肠杆菌

<400> 21

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

Ala Gln Ala

<210> 22
<211> 21
<212> PRT
<213> 大肠杆菌

<400> 22

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

[0043]

Thr Val Ala Gln Ala
20

<210> 23
<211> 26
<212> PRT
<213> 大肠杆菌

<400> 23

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
20 25

<210> 24
<211> 22
<212> PRT
<213> 胡萝卜软腐欧文氏菌

<400> 24

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15

Ala Gln Pro Ala Met Ala
20

<210> 25
<211> 23
<212> PRT
<213> 大肠杆菌

<400> 25

Met Ser Phe Lys Lys Ile Ile Lys Ala Phe Val Ile Met Ala Ala Leu
1 5 10 15

Val Ser Val Gln Ala His Ala
20

[0044]

<210> 26
<211> 28
<212> PRT
<213> 大肠杆菌

<400> 26

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

Met Ser Ile Ser Met Phe Ser Ala Thr Ala Ser Ala
20 25

<210> 27
<211> 19
<212> PRT
<213> 大肠杆菌

<400> 27

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1 5 10 15

Ala Ser Ala

<210> 28
<211> 21
<212> PRT
<213> 大肠杆菌

<400> 28

Met Lys Gln Ala Leu Arg Val Ala Phe Gly Phe Leu Ile Leu Trp Ala
1 5 10 15

Ser Val Leu His Ala
20

<210> 29
<211> 26
<212> PRT
<213> 人工序列

<220>
<223> SipA 信号序列

<400> 29

[0045]

Met Lys Met Asn Lys Lys Val Leu Leu Thr Ser Thr Met Ala Ala Ser
1 5 10 15

Leu Leu Ser Val Ala Ser Val Gln Ala Ser
20 25

<210> 30
<211> 325
<212> PRT
<213> 人工序列

<220>
<223> N-末端 S、Flgl 信号序列、C-末端 GSHRHR 和取代残基 K131 的 KDQNRTK 的 Hla H35L/H48C/G122C 的氨基酸序列

<400> 30

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

Ala Gln Ala Ser Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr

20 25 30

Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr
35 40 45

Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp
50 55 60

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

Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser
85 90 95

Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp
100 105 110

Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp
[0046] 115 120 125

Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val
130 135 140

Thr Gly Asp Asp Thr Gly Lys Asp Gln Asn Arg Thr Lys Ile Gly Gly
145 150 155 160

Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln
165 170 175

Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly
180 185 190

Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr
195 200 205

Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys
210 215 220

Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn
225 230 235 240

Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr
245 250 255

Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp
260 265 270

Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser
275 280 285

Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser
290 295 300

Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn Gly
305 310 315 320

[0047] Ser His Arg His Arg
325

<210> 31
<211> 327
<212> PRT
<213> 人工序列

<220>
<223> 具有 N-末端 S、Flgl 信号序列、C-末端 GSHRHR 和取代残基 K131 的 KDSNITSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 31

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

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

Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr
35 40 45

Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp
50 55 60

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

Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser
85 90 95

Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp
100 105 110

Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp
115 120 125

Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val
130 135 140

[0048]

Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn Ile Thr Ser Ala Arg Ile
145 150 155 160

Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr
165 170 175

Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys
180 185 190

Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly
195 200 205

Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe
210 215 220

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

Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe
245 250 255

```
Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn
            260                 265                 270

Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp
        275                 280                 285

Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp
    290                 295                 300

Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr
305                 310                 315                 320

Asn Gly Ser His Arg His Arg
                325
```

<210> 32
<211> 327
<212> PRT
<213> 人工序列

[0049]

<220>
<223> 具有 N-末端 S、Flgl 信号序列、C-末端 GSHRHR 和取代残基 K131 的 KDSNSTSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 32

```
Met Ile Lys Phe Leu Ser Ala Leu Ile Leu Leu Leu Val Thr Thr Ala
1               5                   10                  15

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

Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr
        35                  40                  45

Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp
    50                  55                  60

Asp Lys Asn Cys Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr
65                  70                  75                  80
```

Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser
85 90 95

Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp
100 105 110

Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp
115 120 125

Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val
130 135 140

Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn Ser Thr Ser Ala Arg Ile
145 150 155 160

Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr
165 170 175

[0050]

Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys
180 185 190

Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly
195 200 205

Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe
210 215 220

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

Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe
245 250 255

Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn
260 265 270

Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp
275 280 285

Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp
290 295 300

Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr
305 310 315 320

Asn Gly Ser His Arg His Arg
325

<210> 33
<211> 327
<212> PRT
<213> 人工序列

<220>
<223> 具有 N-末端 S、Flg1 信号序列、C-末端 GSHRHR 和取代残基 K131 的 KDSNVTSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 33

[0051]

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

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

Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr
35 40 45

Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp
50 55 60

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

Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser
85 90 95

Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp
100 105 110

Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp
115 120 125

Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val
130 135 140

Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn Val Thr Ser Ala Arg Ile
145 150 155 160

Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr
165 170 175

Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys
180 185 190

Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly
195 200 205

[0052] Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe
210 215 220

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

Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe
245 250 255

Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn
260 265 270

Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp
275 280 285

Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp
290 295 300

Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr

305 310 315 320

Asn Gly Ser His Arg His Arg
325

<210> 34
<211> 327
<212> PRT
<213> 人工序列

<220>
<223> 具有 N-末端 S、Flgl 信号序列、C-末端 GSHRHR 和取代残基 K131 的 KDSNATSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 34

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

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

[0053]

Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr
35 40 45

Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp
50 55 60

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

Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser
85 90 95

Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp
100 105 110

Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp
115 120 125

Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val

130 135 140

Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn Val Thr Ser Ala Arg Ile
145 150 155 160

Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr
165 170 175

Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys
180 185 190

Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly
195 200 205

Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe
210 215 220

Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp
[0054] 225 230 235 240

Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe
245 250 255

Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn
260 265 270

Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp
275 280 285

Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp
290 295 300

Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr
305 310 315 320

Asn Gly Ser His Arg His Arg
325

<210> 35
<211> 299
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDQNRTK 的成熟 Hla H35L/H48C/G122C 的氨基酸序列

<400> 35

```
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1               5                  10                  15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
            20                  25                  30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys
        35                  40                  45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
    50                  55                  60
```

[0055]

```
Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65                  70                  75                  80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
                85                  90                  95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
            100                 105                 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
        115                 120                 125

Thr Gly Lys Asp Gln Asn Arg Thr Lys Ile Gly Gly Leu Ile Gly Ala
    130                 135                 140

Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys
145                 150                 155                 160

Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp Lys Val Ile
```

165 170 175

Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser
180 185 190

Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg Asn Gly
195 200 205

Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser
210 215 220

Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val Ile Thr Met
225 230 235 240

Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu
245 250 255

Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr Asn Trp Lys
260 265 270

[0056]

Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu Arg Tyr
275 280 285

Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn
290 295

<210> 36
<211> 301
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDSNITSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 36

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
115 120 125

[0057] Thr Gly Lys Asp Ser Asn Ile Thr Ser Ala Arg Ile Gly Gly Leu Ile
130 135 140

Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro Asp
145 150 155 160

Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp Lys
165 170 175

Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg
180 185 190

Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg
195 200 205

Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala
210 215 220

Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val Ile

225 230 235 240

Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val Ile
245 250 255

Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr Asn
260 265 270

Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu
275 280 285

Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn
290 295 300

<210> 37
<211> 301
<212> PRT
<213> 人工序列

<220>
[0058] <223> 具有取代残基 K131 的 KDSNSTSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 37

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

```
Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
            100                 105                 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
        115                 120                 125

Thr Gly Lys Asp Ser Asn Ser Thr Ser Ala Arg Ile Gly Gly Leu Ile
    130                 135                 140

Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro Asp
145                 150                 155                 160

Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp Lys
                165                 170                 175

Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg
            180                 185                 190
```

[0059]

```
Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg
        195                 200                 205

Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala
    210                 215                 220

Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val Ile
225                 230                 235                 240

Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val Ile
                245                 250                 255

Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr Asn
            260                 265                 270

Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu
        275                 280                 285

Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn
```

290 295 300

<210> 38
<211> 301
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDSNVTSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 38

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys
35 40 45

[0060] Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
115 120 125

Thr Gly Lys Asp Ser Asn Val Thr Ser Ala Arg Ile Gly Gly Leu Ile
130 135 140

Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro Asp
145 150 155 160

```
Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp Lys
                165                 170                 175

Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg
            180                 185                 190

Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg
        195                 200                 205

Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala
    210                 215                 220

Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val Ile
225                 230                 235                 240

Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val Ile
                245                 250                 255
```

[0061]

```
Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr Asn
            260                 265                 270

Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu
        275                 280                 285

Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn
    290                 295                 300
```

<210> 39
<211> 301
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDSNATSAR 的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 39

```
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1               5                   10                  15
```

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
20 25 30

Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
100 105 110

[0062]

Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn Val Thr Gly Asp Asp
115 120 125

Thr Gly Lys Asp Ser Asn Val Thr Ser Ala Arg Ile Gly Gly Leu Ile
130 135 140

Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro Asp
145 150 155 160

Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp Lys
165 170 175

Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg
180 185 190

Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg
195 200 205

Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala
210 215 220

Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val Ile
225 230 235 240

Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val Ile
245 250 255

Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr Asn
260 265 270

Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu
275 280 285

Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn
290 295 300

<210> 40
<211> 7
<212> PRT
<213> 人工序列

[0063]

<220>
<223> KDQNRTK 糖基化位点

<400> 40

Lys Asp Gln Asn Arg Thr Lys
1 5

<210> 41
<211> 7
<212> PRT
<213> 人工序列

<220>
<223> KDQNATK 糖基化位点

<400> 41

Lys Asp Gln Asn Ala Thr Lys
1 5

<210> 42

<211> 9
<212> PRT
<213> 人工序列

<220>
<223> KDSNITSAR 糖基化位点

<400> 42

Lys Asp Ser Asn Ile Thr Ser Ala Arg
1 5

<210> 43
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> KDSNSTSAR 糖基化位点

<400> 43

Lys Asp Ser Asn Ser Thr Ser Ala Arg
1 5

[0064]

<210> 44
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> KDSNVTSAR 糖基化位点

<400> 44

Lys Asp Ser Asn Val Thr Ser Ala Arg
1 5

<210> 45
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> KDSNATSAR 糖基化位点

<400> 45

Lys Asp Ser Asn Ala Thr Ser Ala Arg
1 5

<210> 46
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
<221> 杂项特征
<222> (1)..(1)
<223> Xaa 可以是 Lys 或 Arg

<220>
<221> 杂项特征
<222> (2)..(2)
<223> Zaa 表示 0-9 个除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸的氨基酸

<220>
[0065] <221> 杂项特征
<222> (3)..(3)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (4)..(4)
<223> Xaa 是除脯氨酸、赖氨酸或精氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (6)..(6)
<223> Xaa 是除脯氨酸、赖氨酸或精氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (7)..(7)
<223> Xaa 可以是 Ser 或 Thr

<220>
<221> 杂项特征
<222> (8)..(8)
<223> Zaa 表示 0-9 个除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸的氨基酸

<220>

<221> 杂项特征
<222> (9)..(9)
<223> Xaa 可以是 Lys 或 Arg

<400> 46

Xaa Xaa Xaa Xaa Asn Xaa Xaa Xaa Xaa
1 5

<210> 47
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 共有序列

<220>
<221> 杂项特征
<222> (1)..(1)
<223> Xaa 可以是 Lys 或 Arg

<220>
[0066] <221> 杂项特征
<222> (2)..(2)
<223> Xaa 表示 0-9 个除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸的氨基酸

<220>
<221> 杂项特征
<222> (3)..(3)
<223> Xaa 可以是 Asp 或 Glu

<220>
<221> 杂项特征
<222> (4)..(4)
<223> Xaa 是除脯氨酸、半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (6)..(6)
<223> Xaa 是除脯氨酸、半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸

<220>
<221> 杂项特征
<222> (7)..(7)
<223> Xaa 可以是 Ser 或 Thr

<220>
<221> 杂项特征
<222> (8)..(8)
<223> Xaa 表示 0-9 个除半胱氨酸、甲硫氨酸、天冬酰胺、谷氨酰胺、赖氨酸或精氨酸之外的任何氨基酸的氨基酸

<220>
<221> 杂项特征
<222> (9)..(9)
<223> Xaa 可以是 Lys 或 Arg

<400> 47

Xaa Xaa Xaa Xaa Asn Xaa Xaa Xaa Xaa
1 5

<210> 48
<211> 8
<212> PRT
<213> 人工序列

<220>
<223> 命名为 PTP-2 的肽 42T-50K

[0067]

<400> 48

Thr Gly Asp Leu Val Thr Tyr Lys
1 5

<210> 49
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 命名为 PTP-3 的肽 225A-234K

<400> 49

Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys
1 5 10

<210> 50
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 间隔区

<400> 50

Gly Ser Gly Gly Gly
1 5

<210> 51
<211> 351
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDSNITSAR 糖基化位点；N-末端的糖基化位点 KDSNVTSAR 与该糖基化位点之前和之后的 GSGGG 间隔区；Flgl 信号序列；和 C-末端的 His 标签的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 51

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

[0068] Ala Gln Ala Ser Ala Gly Ser Gly Gly Gly Lys Asp Ser Asn Val Thr
20 25 30

Ser Ala Arg Gly Ser Gly Gly Gly Lys Leu Ala Asp Ser Asp Ile Asn
35 40 45

Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr
50 55 60

Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met Leu Lys Lys Val
65 70 75 80

Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys Asn Lys Lys Leu Leu Val
85 90 95

Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu
100 105 110

Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val
115 120 125

Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr
130 135 140

Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr
145 150 155 160

Gly Phe Asn Cys Asn Val Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn
165 170 175

Ile Thr Ser Ala Arg Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile
180 185 190

Gly His Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu
195 200 205

Ser Pro Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met
210 215 220

[0069] Val Asn Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val
225 230 235 240

Tyr Gly Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala
245 250 255

Ala Asp Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser
260 265 270

Gly Phe Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala
275 280 285

Ser Lys Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp
290 295 300

Asp Tyr Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr
305 310 315 320

Lys Asp Lys Trp Ile Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp

325 330 335

Glu Lys Glu Glu Met Thr Asn Gly Ser His His His His His His
340 345 350

<210> 52
<211> 345
<212> PRT
<213> 人工序列

<220>
<223> 具有取代残基 K131 的 KDSNITSAR 糖基化位点；C-末端糖基化位点 KDSNVTSAR 与该糖基化位点之前的 GSGGG 间隔区；Flgl 信号序列；和 C-末端的 His 标签的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 52

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

Ala Gln Ala Ser Ala Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr
[0070] 20 25 30

Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr
35 40 45

Tyr Asp Lys Glu Asn Gly Met Leu Lys Lys Val Phe Tyr Ser Phe Ile
50 55 60

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

Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys
85 90 95

Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro
100 105 110

Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile
115 120 125

Asp Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr Gly Phe Asn Cys Asn
130 135 140

Val Thr Gly Asp Asp Thr Gly Lys Asp Ser Asn Ile Thr Ser Ala Arg
145 150 155 160

Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys
165 170 175

Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys
180 185 190

Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp
195 200 205

Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu
210 215 220

[0071] Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu
225 230 235 240

Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp
245 250 255

Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr
260 265 270

Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His
275 280 285

Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile
290 295 300

Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met
305 310 315 320

Thr Asn Leu Gly Ser Gly Gly Gly Lys Asp Ser Asn Val Thr Ser Ala
325 330 335

Arg Gly Ser His His His His His His
340 345

<210> 53
<211> 357
<212> PRT
<213> 人工序列

<220>
<223> 具有之前为 GSGGG 间隔区的 C-末端的 KDSNITSAR 糖基化位点；N-末端糖基化位点 KDSNVTSAR 与该糖基化位点之前和之后的 GSGGG 间隔区；Flgl 信号序列；和 C-末端的 His 标签的成熟 Hla H35L/G122C/H48C 的氨基酸序列

<400> 53

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

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

[0072]

Ser Ala Arg Gly Ser Gly Gly Gly Lys Leu Ala Asp Ser Asp Ile Asn
35 40 45

Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr
50 55 60

Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met Leu Lys Lys Val
65 70 75 80

Phe Tyr Ser Phe Ile Asp Asp Lys Asn Cys Asn Lys Lys Leu Leu Val
85 90 95

Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu
100 105 110

Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val
115 120 125

Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr
130 135 140

Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Met Ser Thr Leu Thr Tyr
145 150 155 160

Gly Phe Asn Cys Asn Val Thr Gly Asp Asp Thr Gly Ile Gly Gly Leu
165 170 175

Ile Gly Ala Asn Val Ser Ile Gly His Thr Leu Lys Tyr Val Gln Pro
180 185 190

Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly Trp
195 200 205

Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr Asp
210 215 220

Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys Thr
225 230 235 240

[0073] Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn Lys
245 250 255

Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr Val
260 265 270

Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp Val
275 280 285

Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser Thr
290 295 300

Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Ile Asp Arg Ser Ser
305 310 315 320

Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn Leu Gly
325 330 335

Ser Gly Gly Gly Lys Asp Ser Asn Ile Thr Ser Ala Arg Gly Ser His

340 345 350

[0074]

His His His His His
355

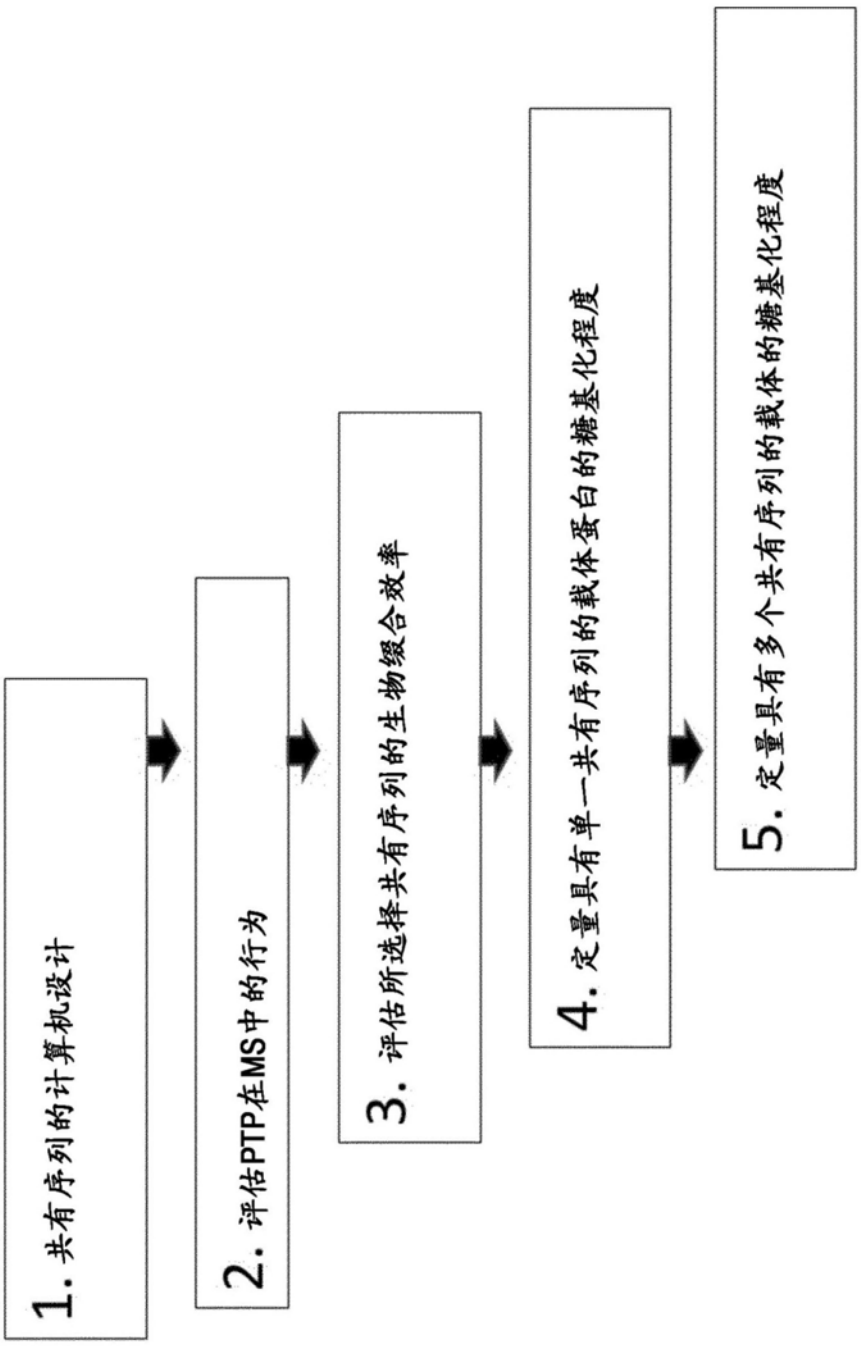

1. 共有序列的计算机设计
2. 评估PTP在MS中的行为
3. 评估所选择共有序列的生物缀合效率
4. 定量具有单一共有序列的载体蛋白的糖基化程度
5. 定量具有多个共有序列的载体的糖基化程度

图1

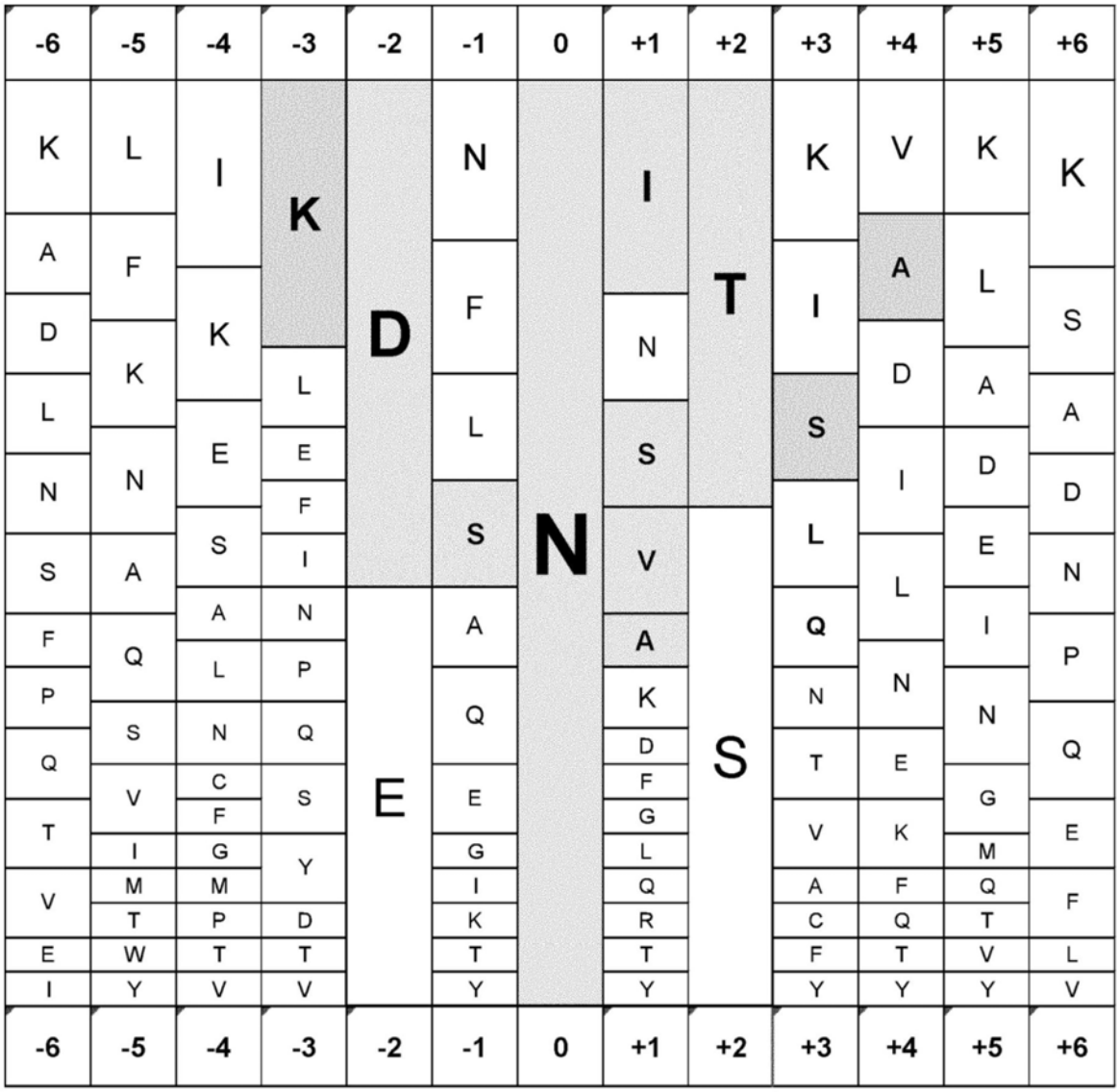

-6 -5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5 +6
A
PTP-i K D S N I T S A R
PTP-s K D S N S T S A R
PTP-n K D S N V T S A R
PTP-a K D S N A T S A R
B

图2

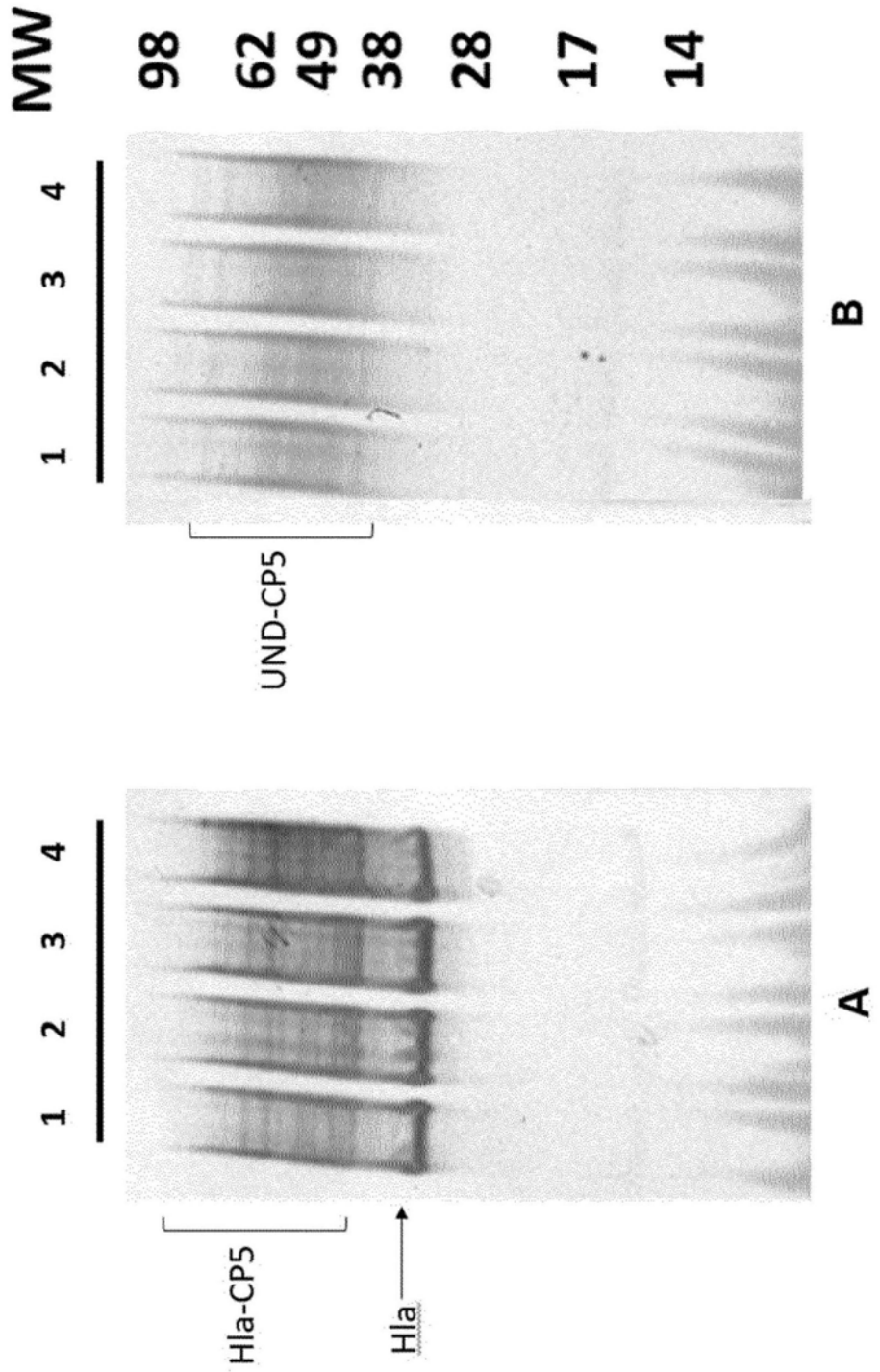


图3

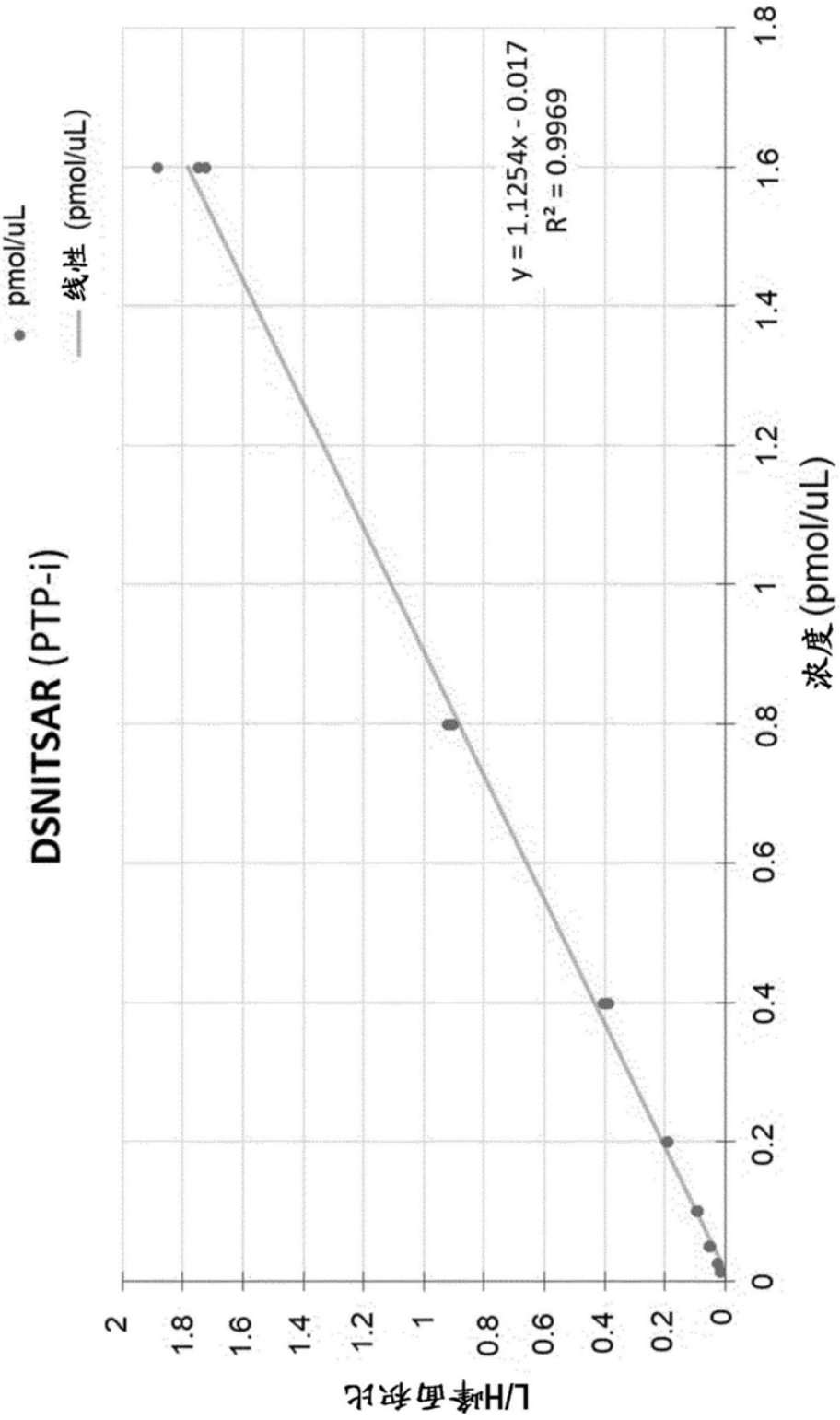
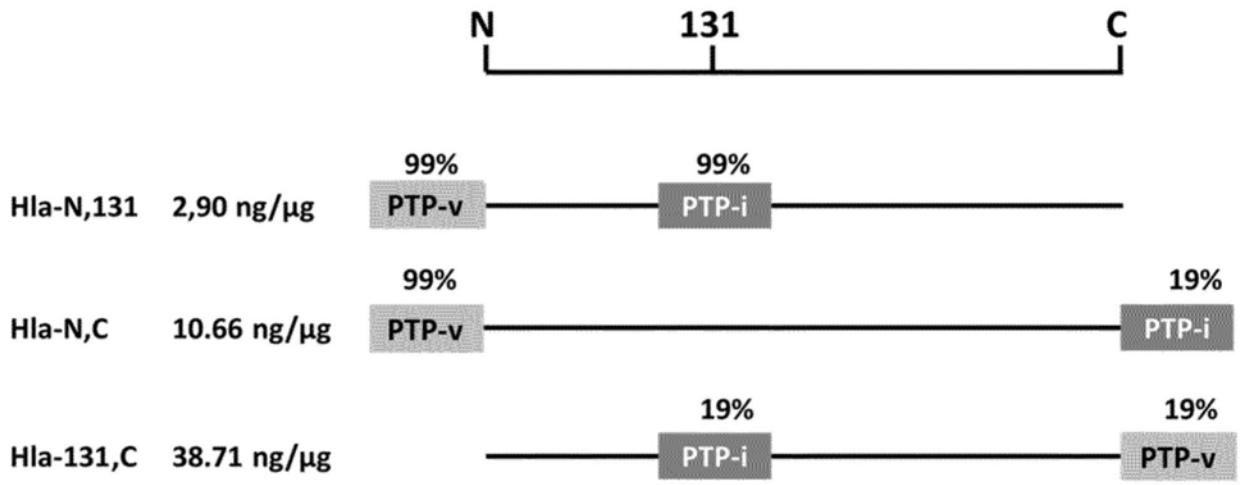
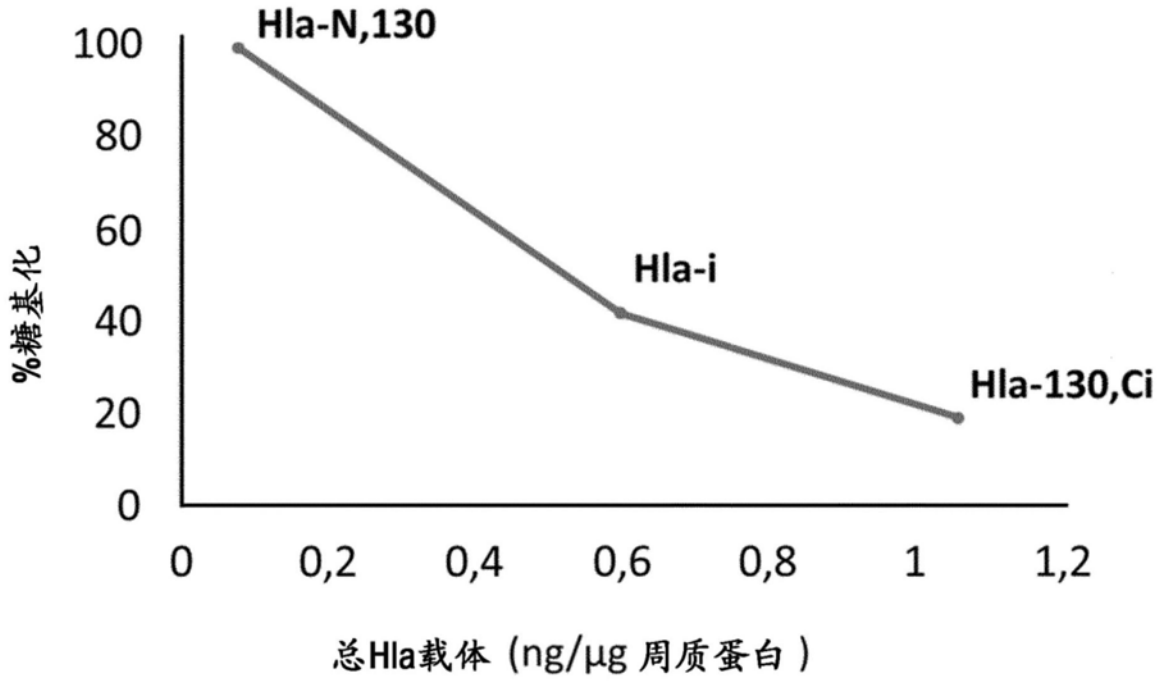

DSNITSAR (PTP-i)
pmol/uL
线性 (pmol/uL)
y = 1.1254x - 0.017
R² = 0.9969
L/H峰面积比
浓度 (pmol/uL)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

图4


N
131
C
99%
99%
Hla-N,131
2,90 ng/µg
PTP-v
PTP-i
99%
19%
Hla-N,C
10.66 ng/µg
PTP-v
PTP-i
19%
19%
Hla-131,C
38.71 ng/µg
PTP-i
PTP-v

A


Hla-N,130
Hla-i
Hla-130,Ci
100
80
60
40
20
0
%糖基化
0
0,2
0,4
0,6
0,8
1
1,2
总Hla载体（ng/µg 周质蛋白）

B

图5