



(12) 发明专利

(10) 授权公告号 CN 111183221 B

(45) 授权公告日 2024. 04. 09

(21) 申请号 201880050251.7
(22) 申请日 2018.05.25
(65) 同一申请的已公布的文献号
 申请公布号 CN 111183221 A
(43) 申请公布日 2020.05.19
(30) 优先权数据
 1708476.5 2017.05.26 GB
 1708471.6 2017.05.26 GB
 1806655.5 2018.04.24 GB
(85) PCT国际申请进入国家阶段日
 2020.02.03
(86) PCT国际申请的申请数据
 PCT/EP2018/063832 2018.05.25
(87) PCT国际申请的公布数据
 W02018/215656 EN 2018.11.29
(73) 专利权人 杰诺维斯公司
 地址 瑞典隆德

(72) 发明人 弗雷德里克·里奥 罗尔夫·卢德
 斯蒂芬·比约克 马林·梅贾尔
 弗雷德里克·奥尔森

(74) 专利代理机构 华进联合专利商标代理有限公司 44224
 专利代理师 黄爱娇

(51) Int.Cl.
 C12N 9/24 (2006.01)
 C12Q 1/37 (2006.01)
 G01N 33/68 (2006.01)
 C12N 9/52 (2006.01)

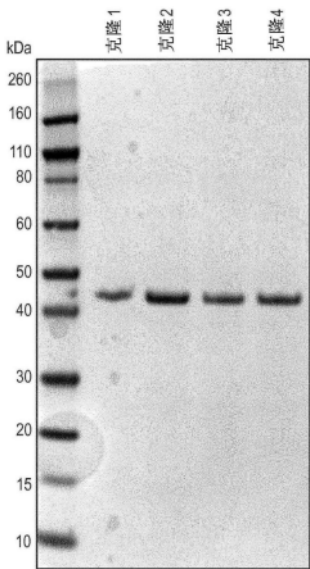
(56) 对比文件
 Lucas,S. et al..“hypothetical protein Amuc_1119 [Akkermansia muciniphila ATCC BAA-835],ACD04945.1.《GenBank》.2013,第1-2页.
 Lucas,S. et al..“hypothetical protein Amuc_1119 [Akkermansia muciniphila ATCC BAA-835],ACD04945.1.《GenBank》.2013,第1-2页.
 (续)

审查员 叶金金

权利要求书2页 说明书45页
序列表69页 附图40页

(54) 发明名称
 用于O-糖蛋白的蛋白酶和结合多肽

(57) 摘要
 本发明涉及一种新型内切蛋白酶,该内切蛋白酶具有结合但缺乏水解活性或具有降低的水解活性的突变体,以及在研究和分出O-连接糖蛋白的方法中的用途。



[转续页]

[接上页]

(56) 对比文件

NEIL D. RAWLINGS et al..Evolutionary Families of Metallopeptidases.《METHODS IN ENZYMOLOGY》.1995,第248卷第183页摘要、第184页第3段.

Gilles Vazeux et al..Identification of Glutamate Residues Essential for Catalytic Activity and Zinc Coordination in Aminopeptidase A.《THE JOURNAL OF BIOLOGICAL CHEMISTRY》.1996,第271卷(第15

期),第906页左栏第1段.

Kun Huang et al..Biochemical characterisation of the neuraminidase pool of the human gut symbiont *Akkermansia muciniphila*.《Carbohydrate Research》.2015,第415卷第60-65页.

杭锋等.细菌来源金属蛋白酶的研究进展.《工业微生物》.2015,第45卷(第6期),第53-59页.

1. 一种包含具有O-糖基化蛋白特异性内切蛋白酶活性的多肽和唾液酸酶Am1757或Am1757和Am0707的混合物的组合物、组合或试剂盒,其中所述多肽由以下组成:

(a) SEQ ID NO: 1、SEQ ID NO: 2或SEQ ID NO: 4的氨基酸序列,并且

其中Am1757为SEQ ID NO: 9至11中任何一个的多肽,并且其中Am0707是SEQ ID NOs: 12至14中任何一个的多肽。

2. 根据前述权利要求中任一项所述的组合物、组合或试剂盒,其特征在于,

(a) 所述多肽在N末端包含另外的甲硫氨酸和/或在C末端包含His标签,

(b) 所述多肽以溶液、冻干或固定的形式提供;和/或

(c) 所述Am1757是由SEQ ID NO: 11组成的多肽和/或其中Am0707是由SEQ ID NO: 14组成的多肽。

3. 一种水解O-糖蛋白的方法,其中,所述方法包括使包含O-糖蛋白的样品与根据权利要求1或2所述的组合物或组合接触。

4. 一种用于评估蛋白质的糖基化状态的方法,所述方法包括使包含所述蛋白质的样品与根据权利要求1或2所述的组合物或组合接触,以及检测和/或分析所产生的产物。

5. 根据权利要求3所述的方法,其特征在于,所述方法还包括对水解产物的检测或分析,并且其中所述分析或检测通过亲和色谱、SDS-PAGE、HPLC或质谱来实施。

6. 根据权利要求4所述的方法,其特征在于,

(a) 所述分析和/或检测通过亲和色谱、SDS-PAGE、HPLC或质谱来实施;

(b) 利用存在或不存在切割产物来确定所述样品中存在或不存在O-糖蛋白;

(c) 进行所述分析以鉴定O-糖链的类型;和/或

(d) 进行所述分析以鉴定O-糖链与O-糖蛋白的连接位置。

7. 一种包含多肽和唾液酸酶Am1757或Am1757和Am0707的混合物的组合物、组合或试剂盒,其中所述多肽能够与O-聚糖或O-糖蛋白结合并且缺乏O-糖基化蛋白特异性内切蛋白酶活性或具有降低的O-糖基化蛋白特异性内切蛋白酶活性,且由

SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 20或SEQ ID NO: 21的氨基酸序列组成,

并且其中Am1757为SEQ ID NO: 9至11中任何一个的多肽,并且Am0707是SEQ ID NOs: 12至14中任何一个的多肽。

8. 根据权利要求7所述的组合物、组合或试剂盒,其特征在于,

所述多肽以溶液、冻干或固定的形式提供。

9. 一种与O-聚糖、O-糖肽和/或O-糖蛋白结合的方法,其中所述方法包括使包含所述O-聚糖、O-糖肽和/或O-糖蛋白的样品与根据权利要求7或8所述的组合物或组合接触。

10. 根据权利要求9所述的方法,其特征在于,所述方法包括步骤:

(a) 确定是否已结合O-聚糖、O-糖肽或O-糖蛋白;和/或

(b) 从所得混合物中分离出O-聚糖和任何连接的糖蛋白、O-糖肽或O-糖蛋白;

其中所述方法用于从样品中分离出O-聚糖或连接的糖蛋白、O-糖肽或O-糖蛋白的目的。

11. 根据权利要求10所述的方法,其特征在于,所述确定和/或分离通过亲和色谱、SDS-PAGE、HPLC、凝集素印迹、ELISA或质谱来实施。

12. 根据权利要求10所述的方法,其特征在于,所述方法另外包括以下步骤:用缓冲液

从权利要求7或8所述的组合物或组合中的所述多肽洗脱结合的物质,所述缓冲液包括:

- (a) 高摩尔浓度的尿素;
- (b) 高浓度洗涤剂;或
- (c) 具有O-糖基化蛋白特异性内切蛋白酶活性的多肽,其中所述多肽由以下组成:
 - (i) SEQ ID NO: 1、SEQ ID NO: 2或SEQ ID NO: 4的氨基酸序列。

13. 一种用于评估蛋白质的糖基化状态的方法,其特征在于,所述方法包括使包含所述蛋白质的样品与根据权利要求7或8所述的组合物或组合接触,以及确定所述蛋白质是否与所述组合物或组合中的所述多肽结合。

14. 一种用于检测样品中O-糖肽和/或O-糖蛋白的方法,其特征在于,所述方法包括:

使所述样品与根据权利要求7或8所述的组合物或组合接触,从而使得O-糖肽和/或O-糖蛋白与所述组合物或组合中的所述多肽之间能够形成复合物。

15. 根据权利要求14所述的方法,其特征在于,所述方法还包括步骤:

- (a) 从接触的样品中分离出所述多肽;和
- (b) 确定分离的多肽是否与O-连接的糖肽或糖蛋白结合,从而确定样品中存在或者不存在O-连接的糖肽或糖蛋白。

16. 根据权利要求13至15中任一项所述的方法,其特征在于,所述确定和/或分离通过亲和色谱、SDS-PAGE、HPLC、凝集素印迹、ELISA或质谱来实施。

17. 根据权利要求13至15中任一项所述的方法,其特征在于,所述方法另外包括以下步骤:用缓冲液从权利要求7或8所述的组合物或组合中的所述多肽洗脱结合的物质,所述缓冲液包括:

- (a) 高摩尔浓度的尿素;
- (b) 高浓度洗涤剂;或
- (c) 具有O-糖基化蛋白特异性内切蛋白酶活性的多肽,其中所述多肽由以下组成:
 - (i) SEQ ID NO: 1、SEQ ID NO: 2或SEQ ID NO: 4的氨基酸序列。

用于O-糖蛋白的蛋白酶和结合多肽

技术领域

[0001] 本发明涉及一种具有结合活性但缺乏水解活性或具有降低的水解活性的新型内切蛋白酶、其突变体,以及在研究和分出O-连接糖蛋白的方法中的用途。

背景技术

[0002] 近来,糖基化对生物学功能的影响已引起越来越多的关注,特别是在O-连接的聚糖方面。然而,尽管已经恢复对这些重要的蛋白质修饰的兴趣,但仍缺乏有效研究聚糖和糖蛋白的工具。

[0003] 已经开发有对从天然蛋白中去除O-连接的聚糖和对聚糖测序都非常有用的几种外切糖苷酶和内切糖苷酶。这两种方法均可单独使用,以减少糖蛋白的异质性,从而有助于在质谱中分析蛋白质及其片段化的肽。还可以通过对受水解影响的功能进行下游分析来实施对聚糖的生物学效应的更有效的分析。然而,这样的工具例如在方便O-连接的糖蛋白的鉴定、糖基化位点的确定和O-连接的糖肽的纯化上不是有效的。

[0004] 1991/1992年报道了首个O-糖蛋白特异性内切蛋白酶,O-糖蛋白特异性内切蛋白酶与O-聚糖结合并主要水解接近聚糖的R-N键(Abdullah et al., J Bacteriol 173,5597-5603(1991); Abdullah et al., Infect Immun 60,56-62(1992))。然而,该酶在医学和生物技术中的用途是有限的,因为该酶仅对包含唾液酸的O-聚糖(大部分但远非所有O-连接的聚糖)具有特异性,并且要求特定的氨基酸,从而通常导致水解水平低,因此需要更好的用于研究O连接的糖蛋白的工具。

发明内容

[0005] 本发明人已经鉴定、纯化和表征了来自嗜粘蛋白艾克曼菌(*Akkermansia muciniphila*)的新型多肽,在本文中被称作LS。该多肽充当内切蛋白酶,该内切蛋白酶特别地切割/水解在O-连接的聚糖的N末端且与O-连接的聚糖接近的氨基酸键,而不显示出对特定氨基酸序列的任何特异性或限制。

[0006] 本发明人还修饰了LS的序列,并鉴定了能够结合O-连接的聚糖但缺乏水解糖蛋白能力或具有降低的水解糖蛋白能力的突变体。这些突变体可用于选择性去除、富集或纯化游离的O-聚糖、O-糖肽和/或O-糖蛋白。

[0007] 因此,在本发明的第一方面,提供一种对O-糖基化蛋白具有特异性内切蛋白酶活性的多肽,该多肽包含:

[0008] (a) SEQ ID NO:1的氨基酸序列;

[0009] (b) 与SEQ ID NO:1的氨基酸序列至少85%一致的氨基酸序列;或

[0010] (c) 这样的氨基酸序列:该氨基酸序列为SEQ ID NO:1的序列的片段或与SEQ ID NO:1的氨基酸序列85%一致的氨基酸序列的片段。

[0011] 本发明还提供一种水解O-糖蛋白的方法,其中该方法包括使包含该蛋白质的样品与本发明的多肽接触,并且任选地还包括对水解产物的检测或分析。

[0012] 另外,提供一种评估蛋白质的糖基化状态的方法,该方法包括使包含该蛋白质的样品与本发明的多肽接触,以及检测和/或分析产生的产物,任选地,其中使用切割产物存在或不存在来确定样品中O-糖蛋白的存在/不存在,和/或其中进行所述分析以鉴定O-糖链的类型和/或O-糖链与O-糖蛋白的连接位置。

[0013] 在本发明的第二方面,提供一种能够与O-聚糖、O-糖肽和/或O-糖蛋白结合并且对O-糖基化蛋白缺乏特异性内切蛋白酶活性或具有降低的特异性内切蛋白酶活性的多肽,该多肽包含:

[0014] (a) SEQ ID NO:5的氨基酸序列;

[0015] (b) 与SEQ ID NO:5或SEQ ID NO:20的氨基酸序列至少85%一致的氨基酸序列;或

[0016] (c) 这样的氨基酸序列:该氨基酸序列为SEQ ID NO:5或SEQ ID NO:20的序列的片段或与SEQ ID NO:5或SEQ ID NO:20的氨基酸序列85%一致的氨基酸序列的片段。

[0017] 本发明还提供一种与O-聚糖、O-糖肽和/或O-糖蛋白结合的方法,其中该方法包括使包含O-聚糖、O-糖肽和/或O-糖蛋白的样品与本发明的多肽接触,以及任选地确定是否已结合O-聚糖、O-糖肽或O-糖蛋白和/或从所得混合物中使O-聚糖与任何连接的糖蛋白、O-糖肽或O-糖蛋白分离。

[0018] 另外,提供一种用于评估蛋白质的糖基化状态的方法,该方法包括使包含蛋白质的样品与本发明的多肽接触,以及确定蛋白质是否与所述多肽结合。

[0019] 还提供了一种用于检测样品中O-糖肽和/或O-糖蛋白的方法,其中该方法包括:

[0020] (a) 使所述样品与本发明的多肽接触,从而使得在本发明的多肽与O-连接的糖肽和/或O-糖蛋白之间形成复合物(O-连接的糖肽/蛋白质-多肽复合物);

[0021] (b) 任选地从接触的样品中分离出所述多肽;和

[0022] (c) 确定分离的多肽是否与O-连接的糖肽或糖蛋白结合,从而确定样品中存在或不存在O-连接的糖肽或糖蛋白。

附图说明

[0023] 图1:LS的表达和纯化。LS在pET21 (a)+载体中表达为具有C末端His-tag的融合蛋白。转化进BL21 (DE3) Star后,表达了四个单独的克隆,并在His GravityFlow柱上纯化至均一。基于纯化样品中蛋白质的总量以及基于SDS-PAGE的纯度,所研究的所有四个克隆均同样表达良好。

[0024] 图2:LS特异性作用于含有O-聚糖的蛋白质-图中显示了通过SDS-PAGE分析的产物。LS与IgG或IgA孵育导致IgA特异性降解,但对IgG(赫赛汀/曲妥珠单抗)没有明显的活性。所有孵育均在37℃下在PBS中进行。在这些条件下,添加唾液酸酶(Am0707)不是LS的活性所需的。

[0025] 图3:最佳酶促条件。LS在很宽的pH范围内有活性(A),能很好地耐受NaCl(B),但对EDTA高度敏感(C,D),并且被Zn²⁺部分抑制(D)。所有实验(pH测定除外)均在PBS中在37℃下过夜进行。为了确定优化pH,将酶在20mM Tris-HCl (pH 6.8-8.8)或50mM乙酸(pH 5.6)中孵育。

[0026] 图4:聚糖组成调节LS的活性。(A)在用LS水解30分钟之前依次去除特定聚糖导致在唾液酸化蛋白中的活性非常低,在去唾液酸化蛋白中的活性高,而在除去半乳糖的样品

中则没有活性。(S) 唾液酸酶, (SG) 唾液酸酶和半乳糖苷酶, (LS) LS。(B) 长时间孵育(过夜) 经完全糖基化的糖蛋白(Enbre1) 或经唾液酸酶处理的糖蛋白(Enbre1(S)) 导致两个样品中完全水解。Enbre1在本文中也可称为依那西普(etanercept)。(C) 用唾液酸酶(“唾液酸酶”)、O-糖苷酶/唾液酸酶(“O-糖”) 或PNGaseF(“N-糖”) 预处理依那西普的TNF α 结合部分(TNF α R), 以分别除去唾液酸、O-聚糖和N-聚糖。将LS添加到样品中, 并在分析前继续过夜孵育。LS在除了用O-糖苷酶处理的那些样品之外的所有样品中外均具有活性。

[0027] 图5: 搜索结果显示LS水解了O-聚糖的糖蛋白N末端。对用LS水解成片段且随后用O-糖苷酶处理去糖基化的对依那西普进行质谱分析(LC/MS和MS/MS)。根据m/z值和MS/MS数据, 将鉴定出的肽(白色和阴影框) 拟合到依那西普序列上, y'和b'离子标被记为小灰色框。所有灰色和阴影框(例如肽) 直接以O-聚糖所连接的T或S开始。前面的氨基酸变化(P、S、H、T、G), 并且似乎可能不影响水解。(A) 使用偏倚方法(biased approach), 特别是搜索由S/T肽酶生成的肽, 进行分析。(B) 使用无偏方法(unbiased approach) 进行分析。

[0028] 图6: 灭活的LS特异性结合O-连接的糖蛋白。使金属蛋白酶活性位点突变以除去催化能力, 同时不影响底物亲和力或相互作用。具体而言, 这是通过将E换为A来完成的, 从而创建克隆“LS_{mut}”(也称为LS_{E206A})。(A) 虽然在唾液酸酶存在下LS能够水解Enbre1, 但失活的LS_{mut}在测试条件下不能水解Enbre1。SDS-PAGE证实活性丧失。(B) 尽管失去了水解活性, 但LS_{mut}仍然能够与O-糖蛋白结合。在具有固定LS_{E206A}的旋转柱上验证了特异性结合, 证明对O-连接的糖蛋白具有特异性亲和力。通过将LS_{mut}固定在琼脂糖凝胶(sepharose)上, 我们能够对IgA进行亲和纯化。缺乏O-聚糖的赫赛汀(曲妥珠单抗) 以及O-糖苷酶处理的IgA不与柱子结合, 但可以在流过液(FT) 中检测到。Neur=神经氨酸酶/唾液酸酶0707。

[0029] 图7: α 2-3键合的唾液酸限制了LS的效率。使用Enbre1作为糖蛋白底物, 将LS与一系列不同的唾液酸酶同时孵育30分钟至20小时, 显示在存在 α 2-3特异性唾液酸酶1757或混合液(0707+1757) 的情况下效率更高, 而广谱唾液酸酶0707并不是LS的似乎完全的活性所必需的, 因此表明 α 2-6(和 α 2-8) 键与LS活性无关。

[0030] 图8: LS活性的示意图。LS优先结合连接至O-连接的GalNAc的末端半乳糖, 导致连接有聚糖的丝氨酸或苏氨酸的N末端水解。唾液酸的存在会降低LS的效率, 但不会抑制LS。在N-连接的聚糖上不能观察到活性。

[0031] 图9: 促红细胞生成素用LS、用于去除N-聚糖的PNGaseF、用于去除唾液酸的唾液酸酶以及与用于去除O-聚糖的O-糖苷酶的不同组合切割的实验结果。通过SDS-PAGE、RPLC和ESI质谱来分析反应产物。(A) SDS-PAGE分析的结果: 泳道1=用PNGaseF和唾液酸酶处理的EPO; 泳道2=用PNGaseF和唾液酸酶+LS处理的EPO; 泳道3=用PNGaseF+LS处理的EPO; 泳道4=用PNGaseF、唾液酸酶、O-糖苷酶然后用LS处理的EPO; 泳道5=酶对照。条带X=未切割的EPO; 条带Y=LS消化的EPO的N末端片段; 条带Z=LS消化的EPO的C末端片段。泳道2和3显示在唾液酸已被去除的情况下以及在它们完好的情况下, LS切割EPO。泳道2-3显示LS也切割EPO, 其中N-聚糖被PNGaseF去除。泳道4显示, LS没有切割EPO, 其中O-聚糖已被去除。(B) UV色谱图显示了用PNGaseF和唾液酸酶+LS处理的EPO的RPLC分离结果。如图所示, 鉴定出两个主峰。峰1是被LS消化的EPO的C末端片段; 峰2是被LS消化的EPO的N末端片段; (C, D) 显示了质谱分析的结果。图C显示了具有仍然与(目前是N末端的) 丝氨酸连接的O聚糖的EPO的C末端片段的质量(正方形=GlcNAc, 圆圈=半乳糖)。质量差异是由于样品中某些部分的O-聚

糖的降解(损失末端半乳糖)的差异而引起的,这很可能是MS仪器中的电离能引起的。图D显示了缺少聚糖的EPO的N末端片段,加上仍连接有聚糖的未消化的EPO。

[0032] 图10:显示LS_{E206A}保留一定活性而LS_{H205A/E206A}(也可以称为LS_{HE206AA})完全失活的实验结果。针对包含依那西普的TNF α 结合部分(TNF α R2;在本文中也称为TNF α R)和依那西普本身(Etanercept)的去唾液酸化O糖基化底物,评估了LS突变体LS_{E206A}(A)和LS_{H205A/E206A}(B)对比野生型LS酶的活性。将不同浓度的LS突变体添加到1mg底物(1:1至15:1,酶:底物),在PBS中在37℃下孵育过夜,然后在SDS-PAGE上分析。

[0033] A) 泳道1:仅去唾液酸化底物;泳道2:仅LS,泳道3:0.5 μ g LS_{E206A},泳道4:5 μ g LS_{E206A},泳道5:TNF α R2+LS(比例为1:1),泳道6:TNF α R2+LS_{E206A}(比例为1:1)。

[0034] B) 泳道1:仅去唾液酸化底物;泳道2:LS_{H205A/E206A}+依那西普(比例为15:1),泳道3:LS_{H205A/E206A}+依那西普(比例为5:1),泳道4:LS_{H205A/E206A}+依那西普(比例为1:1),泳道5:LS+依那西普(比例为1:1),泳道6:LS_{H205A/E206A}。

[0035] 图11:显示固定在树脂上的LS_{H205A/E206A}与含有O-聚糖的蛋白质特异性结合的实验结果。图中显示了在每种情况下对起始/装载材料、流过液(FT)和洗脱液(E)的SDS-PAGE分析:(A)包含BSA(牛血清白蛋白)、依那西普、IgA和IgG的样品,该样品是天然的或用如所示的唾液酸酶混合物+/-O-糖苷酶预处理的;(B)包括O-糖基化蛋白(TNF α R和ApoE)、N-糖基化蛋白(阿柏西普(afibercept)、AGP(α -1-酸糖蛋白)、IgG Fc(IgG的Fc结构域)和非糖基化蛋白(BSA)的混合物的样品用唾液酸酶混合物预处理。(C)包括N-糖基化蛋白(西妥昔单抗、阿柏西普、AGP)和非糖基化蛋白(BSA、碳酸酐酶)和混合物的样品用唾液酸酶混合物预处理。

[0036] 图12:显示固定的LS_{H205A/E206A}具有结合O-糖蛋白的浓度依赖性能力的实验结果。去唾液酸化依那西普(50-250 μ g;100 μ l在PBS中)添加到50 μ l PBS平衡的具有不同的LS_{H205A/E206A}固定条件(5-15mg/mL)的LS_{H205A/E206A}树脂中。在室温下通过颠倒旋转使蛋白质与树脂结合2小时。将树脂用PBS(350 μ l)洗涤3次,然后通过添加8M尿素(50 μ l,5分钟孵育;重复两次)洗脱。为了研究尿素和盐酸胍(GHCl)在结合中的作用,将其与50 μ g去唾液酸化依那西普一起包含在结合缓冲液中,但处理方法相同。(A)所有样品在SDS-PAGE上分离,并使用GelDoc EZ和ImageLab软件通过光密度测定法确定条带强度。所示的百分比表示与对照相比的条带强度。(B),将根据条带强度确定的蛋白质结合能力相对固定的LS_{H205A/E206A}的量作图。

[0037] 图13:显示LS_{H205A/E206A}可以对约3mg依那西普/mL树脂进行亲和纯化的实验结果。去唾液酸化依那西普(10-200 μ g;在PBS中为100 μ l)添加到50 μ l PBS平衡的LS_{H205A/E206A}树脂中。在室温下通过颠倒旋转使蛋白质与树脂结合2小时。将树脂用PBS(350 μ l)洗涤3次,然后通过添加8M尿素(50 μ l,5分钟孵育;重复两次)洗脱。所有样品在SDS-PAGE上分离,并使用GelDoc EZ和ImageLab软件通过光密度测定法确定条带强度。所示的百分比表示与对照相比的条带强度。

[0038] 图14:显示LS_{H205A/E206A}-底物的相互作用对高离子强度和缓冲液体积/类型的差异不敏感且可在较宽的pH范围内工作的实验结果。(A)所示的去唾液酸化依那西普(50 μ g;在PBS中为100 μ l)与以不同浓度添加的NaCl;(B)如所示的去唾液酸化依那西普(50 μ g;在PBS中为100-300 μ l);(C)来自去唾液酸化依那西普(50 μ g)和BSA(50 μ g)的样品在不同缓冲液中

不同的pH下的流过液；(D)来自C样品的洗脱液。

[0039] 图15:显示变性或添加去污剂洗脱与LS_{H205A/E206A}结合的O-糖蛋白的实验结果。去唾液酸化依那西普(50 μ g;在PBS中为100 μ l)添加到50 μ l PBS平衡的LS_{H205A/E206A}树脂中。在室温下通过颠倒旋转使蛋白质与树脂结合2小时。将树脂用PBS(350 μ l)洗涤3次,然后通过添加1M至8M尿素或1.25%至10%SDS(50 μ l,5分钟孵育;重复两次)洗脱。所有样品在SDS-PAGE上分离以用于分析。

[0040] 图16A:使用LS酶促洗脱LS_{H205A/E206A}结合的O-糖蛋白的实验结果。去唾液酸化阿巴西普(10 μ g;在PBS中为100 μ l)和依那西普(50 μ g;在PBS中为100 μ l)添加到50 μ l PBS平衡的LS_{H205A/E206A}树脂中。在室温下通过颠倒旋转使蛋白质与树脂结合2小时。将树脂用结合缓冲液(350 μ l)洗涤3次,然后添加在总体积为100 μ l PBS中的50个单位的LS。将样品在摇动(450rpm)下于37 $^{\circ}$ C下再孵育6至24小时。通过离心(1000g,1分钟)收集LS释放的O糖蛋白/糖肽,然后通过添加8M尿素(50 μ l,5分钟孵育;重复两次)最后洗脱柱子。所有样品在SDS-PAGE上分离以用于分析。

[0041] 图16B:用LS洗脱的依那西普的质谱分析(LC/MS和MS/MS)结果。鉴定出的肽(图16B.1)与依那西普在LS消化中产生的肽(图16B.2)一致。根据m/z值和MS/MS数据,将鉴定出的肽(白色方框)拟合到依那西普的序列上,y'和b'离子标记为小灰方框。所有白框(例如肽)直接从与O-聚糖连接的T或S处开始。

[0042] 图17:表明O-糖基化血清蛋白的亲纯化和富集的结果。(A)去唾液酸化血清(20 μ l;在PBS中为100 μ l)添加到50 μ l PBS平衡的LS_{H205A/E206A}树脂中。在室温下通过颠倒旋转使蛋白质与树脂结合2小时。将树脂用结合缓冲液(350 μ l)洗涤3次,然后通过添加8M尿素洗脱。(B)为了研究聚糖对相互作用的影响,将样品用唾液酸酶混合物+/-O-糖苷酶预处理。如上所述进行下游纯化。(C)将血清(40 μ l)与PBS(高达100 μ l)和唾液酸酶混合物(50-500个单位)混合,并添加到PBS平衡的柱子,在室温下孵育2小时并颠倒旋转,此后如上所述洗涤并洗脱样品。所有样品在SDS-PAGE上分离以用于分析。

[0043] 图18:显示来自人血清中的O-糖蛋白富集的结果。将在PBS中稀释2.5倍至100 μ l的人血清涂在旋转柱中的50 μ l PBS平衡的LS_{H205A/E206A}树脂。添加50-500个单位的唾液酸酶混合物并在室温下在树脂上共孵育2小时。收集流过液,并用PBS洗涤树脂5至10次。结合的蛋白在8M尿素中洗脱,然后通过添加5mM DTT并在37 $^{\circ}$ C下孵育60分钟来变性和还原。将还原的半胱氨酸在室温下和黑暗中用15mM碘乙酰胺烷基化30分钟。在旋转脱盐柱上使样品进行缓冲液交换至50mM Tris pH 8.0中。将胰蛋白酶(2.5 μ g)添加到溶液中,并在37 $^{\circ}$ C下消化过夜。使用RP-LC MS/MS在C18柱上以0.1%FA的MQ溶液:0.1%FA的95%ACN的梯度溶液以45 $^{\circ}$ C及0.2ml/分钟的流量分离并分析肽。检测在ESI-Q-TOF上进行。数据转换为mgf格式文件,并在Swiss Prot数据库中进行搜索(A)鉴定出的肽来自标注为O-糖基化蛋白或非O-糖基化蛋白的蛋白。仅包括具有>6个匹配肽且MASCOT得分>200的蛋白。(B)不同的洗涤步骤导致所鉴定出的肽的变化,以及O-糖基化蛋白与非O-糖基化蛋白的比例改变(C)Sia=唾液酸酶处理的;Sia Pre=唾液酸酶预处理的。

[0044] 图19:显示固定的双突变体也与较短的O-糖肽结合的实验结果。A显示了与O-糖基化肽(糖基果蝇肽(glycodrosocin;GD))和几种非糖基化肽(H2686、H4062、H8390和胰岛素氧化的 β 链(I0B))的制备的混合物结合的LC/MS分析代表性结果。B显示了IgA的示意图,展

示胰蛋白酶消化将产生单个O-糖基化的肽。C显示了与IgA的胰蛋白酶消化物结合的LC/MS分析的代表性结果。

[0045] 图20:显示固定的双突变体比其他可商购的O-糖蛋白结合基质有优势。A显示了代表性的SDS-PAGE凝胶,比较了在与所示的不同的固定的凝集素或LS双突变体孵育后,流过液(FT)或洗脱液(E)中的依那西普或去唾液酸化依那西普(依那西普(S))的存在。B显示相对于1.5 μ g直接加样底物阳性对照的凝胶光密度分析。

[0046] 序列的简要说明

[0047] SEQ ID NO:1是具有O-糖蛋白特异性内切蛋白酶活性的多肽的氨基酸序列。

[0048] SEQ ID NO:2是具有O-糖蛋白特异性内切蛋白酶活性的本发明示例性多肽的氨基酸序列。相对于SEQ ID NO:1,SEQ ID NO:2在N末端包括另外的甲硫氨酸,并且在C末端包括接头+His₆标签。由该序列组成的多肽在本文中可称为LS。

[0049] SEQ ID NO:3是编码具有SEQ ID NO:2的序列的多肽的核苷酸序列。

[0050] SEQ ID NO:4是具有O-糖蛋白特异性内切蛋白酶活性的自嗜粘蛋白艾克曼菌分出的多肽的野生型氨基酸序列。相对于SEQ ID NO:1,SEQ ID NO:4包括N末端的信号基序。

[0051] SEQ ID NO:5是能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的多肽的氨基酸序列。

[0052] SEQ ID NO:6是能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明示例性多肽的氨基酸序列。相对于SEQ ID NO:5,SEQ ID NO:6包括另外的N末端甲硫氨酸和C末端接头+His₆标签。由该序列组成的多肽在本文中可称为LS_{E206A}。

[0053] SEQ ID NO:7是编码具有SEQ ID NO:6的序列的多肽的核苷酸序列。

[0054] SEQ ID NO:8是具有O-糖蛋白特异性内切蛋白酶活性的本发明示例性多肽的金属蛋白酶结构域基序。

[0055] SEQ ID NO:9是自嗜粘蛋白艾克曼菌分出的唾液酸酶Am1757的野生型氨基酸序列。SEQ ID NO:9包括N末端的信号基序。

[0056] SEQ ID NO:10是唾液酸酶Am1757的野生型氨基酸序列,相对于SEQ ID NO:9缺少N末端的信号基序。

[0057] SEQ ID NO:11是示例性唾液酸酶Am1757的氨基酸序列。相对于SEQ ID NO:10,SEQ ID NO:11包括另外的N末端甲硫氨酸和C末端接头+His₆标签。

[0058] SEQ ID NO:12是自嗜粘蛋白艾克曼菌分出的唾液酸酶Am0707的野生型氨基酸序列。SEQ ID NO:12包括N末端的信号基序。

[0059] SEQ ID NO:13是唾液酸酶Am0707的野生型氨基酸序列,相对于SEQ ID NO:12在N末端缺少信号基序。

[0060] SEQ ID NO:14是示例性唾液酸酶Am0707的氨基酸序列。相对于SEQ ID NO:13,SEQ ID NO:14包括另外的N末端甲硫氨酸和C末端接头+His₆标签。

[0061] SEQ ID NO:15是自口腔链球菌(*S.oralis*)分出的O-糖苷酶的氨基酸序列。

[0062] SEQ ID NO:16和17是引物序列。

[0063] SEQ ID NO:18显示了EP0的氨基酸序列。

[0064] SEQ ID NO:20是能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具

有降低的O-糖蛋白特异性内切蛋白酶活性的多肽的氨基酸序列。

[0065] SEQ ID NO:21是能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明示例性多肽的氨基酸序列。相对于SEQ ID NO:20,SEQ ID NO:21包括另外的N末端甲硫氨酸和C末端接头+His₆标签。由该序列组成的多肽在本文中可以称为LS_{HE206AA}或LS_{H205A/E206A}°。

[0066] SEQ ID NO:22是编码具有SEQ ID NO:21的序列的多肽的核苷酸序列。

[0067] SEQ ID NO:23、24和25是被破坏的金属蛋白酶结构域基序的序列,各自来自能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明的多肽。

[0068] SEQ ID NO:26、27和28是各自具有O-糖蛋白特异性内切蛋白酶活性的多肽的氨基酸序列。

[0069] SEQ ID NO:29、30和31是具有O-糖蛋白特异性内切蛋白酶活性的示例性多肽的氨基酸序列。相对于SEQ ID NO:26、27和28,SEQ ID NO:29、30和31各自包括另外的N末端甲硫氨酸和C末端接头+His₆标签。

[0070] SEQ ID NO:32、33和34分别是自铜绿假单胞菌 (*Pseudomonas aeruginosa*) PA01、多形拟杆菌 (*Bacteroides thetaiotaomicron*) VPI-5482和产气荚膜梭菌 (*Clostridium perfringens*) 分出的具有O-糖蛋白特异性内切蛋白酶活性的多肽的野生型氨基酸序列。分别地,相对于SEQ ID NO:26、27和28,各自包含N末端的信号基序。

[0071] SEQ ID NO:35、36和37是各自能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的多肽的氨基酸序列。

[0072] SEQ ID NO:38、39和40是各自能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明示例性多肽的氨基酸序列。分别地,相对于SEQ ID NO:35、36和37,SEQ ID NO:38、39和40各自包括另外的N末端甲硫氨酸和C末端接头+His₆标签。

[0073] SEQ ID NO:41至43是O-糖蛋白特异性内切蛋白酶的代表性金属蛋白酶基序的氨基酸序列。

[0074] SEQ ID NO:44至46是能够与O-聚糖结合但缺乏O-糖蛋白特异性内切蛋白酶活性或O-糖蛋白特异性内切蛋白酶活性降低的多肽的代表性破坏性金属蛋白酶基序的氨基酸序列。

[0075] SEQ ID NO:47是糖基果蝇肽的氨基酸序列。T残基上有O-糖基化位点。

[0076] SEQ ID NO:48至50是未被O-糖基化的肽的氨基酸序列。

具体实施方式

[0077] 应当理解,所公开的产品和方法的不同应用可以适合于本领域的特定需求。还应理解,本文所用的术语仅出于描述本发明的特定实施方式的目的,而并非旨在进行限制。本文引用的所有出版物、专利和专利申请,无论是上文还是下文中的,均通过引用全文并入本文。

[0078] 除非文中另外明确指出,如在本说明书和所附权利要求书中所使用的,单数形式“一个/一种(a/an)”和“该(the)”包括复数个指示物。因此,例如,提及“多肽(a

polypeptide)”包括“多个多肽(polypeptides)”等。

[0079] 一般的多肽特征

[0080] “多肽”在本文以最广泛的意义使用,是指两个或更多个亚基氨基酸、氨基酸类似物或其他拟肽的化合物。因此,术语“多肽”包括短肽序列以及更长的多肽和蛋白质。术语“蛋白质”、“肽”和“多肽”可以互换使用。如本文所用,术语“氨基酸”是指天然和/或非天然或合成的氨基酸,包括D或L旋光异构体,以及氨基酸类似物和拟肽。

[0081] 多肽可以通过合适的方法产生,包括重组或合成方法。例如,多肽可以使用本领域已知的标准技术直接合成,例如Fmoc固相化学、Boc固相化学或通过溶液相肽合成。可供选择地,可以通过用编码所述多肽的核酸分子或载体转化细胞,通常转化细菌细胞来产生多肽。下文描述了并且在实施例举例说明了通过在细菌宿主细胞中表达来产生多肽的方法。本发明提供了编码本发明多肽的核酸分子和载体。本发明还提供了包含这种核酸或载体的宿主细胞。编码本文公开的多肽的示例性多核苷酸分子以SEQ ID NO:3和7提供。这些序列中的每一条在5'端包括N末端甲硫氨酸(ATG)的密码子,并且在3'端在终止密码子(TAA)之前包括Gly-Ser-Gly接头和6x His标签的密码子,可以任选不包括所述GSGLE接头和6x His标签的密码子。下文将更详细地讨论另外的甲硫氨酸和标签的任性包含。

[0082] 术语“核酸分子”和“多核苷酸”在本文可互换使用,并且是指任何长度的核苷酸(无论是脱氧核糖核苷酸还是核糖核苷酸)的聚合形式或其类似物。多核苷酸的非限制性实例包括基因、基因片段、信使RNA(mRNA)、cDNA、重组多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针和引物。本发明的多核苷酸编码本发明的多肽,并且可以以分离的或基本上分离的形式提供。基本上分离是指可以从任何周围介质中基本上但不完全分离多肽。可以将多核苷酸与不会干扰其预期用途的运载体或稀释剂混合,并且仍然被认为是基本上分离的。“编码”所选多肽的核酸序列是这样一种核酸分子,所述核酸分子在受适当调控序列控制时,在活体内转录(对于DNA而言)并翻译(对于mRNA而言)成多肽,例如在表达载体中。编码序列的边界由5'(氨基)末端的起始密码子和3'(羧基)末端的翻译终止密码子确定。出于本发明的目的,这样的核酸序列可以包括但不限于来自病毒的cDNA、原核或真核mRNA、来自病毒的基因组序列、或原核DNA或RNA,甚至合成的DNA序列。转录终止序列可以位于编码序列的3'处。

[0083] 多核苷酸可以根据本领域众所周知的方法来合成,如例如Sambrook et al.(1989,分子克隆-实验指南(Molecular Cloning-a laboratory manual);冷泉港出版社)中所描述。本发明的核酸分子可以以表达盒的形式提供,所述表达盒包括可操作地连接至插入序列的控制序列,从而允许本发明的多肽在活体内(例如,在原核或真核表达系统中)表达。这些表达盒又通常在载体(例如质粒或重组病毒载体)内提供。可以将这样的表达盒直接给药至宿主对象。可供选择地,可以将包含本发明的多核苷酸的载体给药至宿主对象。优选地,使用遗传载体进行多核苷酸制备和/或给药。合适的载体可以是能够携带足量的遗传信息并允许表达本发明的多肽的任何载体。

[0084] 因此,本发明包括包含这样的多核苷酸序列的表达载体。这样的表达载体在分子生物学领域中是常规构建的,并且例如可以涉及使用质粒DNA和适当的起始子、启动子、增强子和其他元件,例如,为了允许表达本发明的肽而可能是必要的并且以正确的取向定位的聚腺苷酸化信号。其他合适的载体对本领域技术人员来说将是显而易见的。作为在这方

面的进一步的实例,我们参考Sambrook et al.。

[0085] 本发明还包括已经修饰以表达本发明的多肽的细胞。这样的细胞通常包括原核细胞,例如细菌细胞,例如大肠杆菌。可以使用常规方法培养这样的细胞以产生本发明的多肽。

[0086] 多肽可以衍生化或经修饰以有助于该多肽的产生、分离或纯化。例如,在通过细菌宿主细胞中的重组表达来产生本发明的多肽的情况下,该多肽的序列可以在N末端包含另外的甲硫氨酸(M)残基以改善表达。作为另一个实例,本发明的多肽可以通过添加能够直接且特异性地与分开手段结合的配体来衍生化或进行修饰。可供选择地,多肽可以通过添加结合对中的一个成员来衍生化或进行修饰,并且分开手段包括通过添加结合对中的其他成员而衍生化或进行修饰的试剂。可以使用任何合适的结合对。在用于本发明的多肽通过添加结合对中的一个成员而衍生化或进行修饰的优选实施方式中,该多肽优选是组氨酸标记的或生物素标记的。通常,在基因水平上包括组氨酸或生物素标签的氨基酸编码序列,并且该多肽在大肠杆菌中进行重组表达。组氨酸或生物素标签通常存在于多肽的任一端,优选地在C末端。组氨酸或生物素标签可以与多肽直接连接或通过任何合适的接头序列(例如3、4或5个甘氨酸残基或甘氨酸和丝氨酸残基的混合物)间接连接。组氨酸标签通常由六个组氨酸残基组成,尽管组氨酸标签可以比这更长,通常多达7、8、9、10或20个氨基酸,或更短,例如5、4、3、2或1个氨基酸。

[0087] 多肽可以以基本上分离的或纯化的形式提供。也就是说,从存在于表达多肽的细胞的细胞提取物中的其他大部分成分中分离的。将理解的是,基本上纯化是指多肽被纯化至至少50%、60%、70%、80%或优选至少90%同质性。纯度水平可通过任何合适的方法进行评估,但通常涉及样品的SDS-PAGE分析,然后进行考马斯蓝检测。多肽可以与不会干扰多肽预期目的的运载体、稀释剂或防腐剂混合,并且仍被为是基本上分离的或纯化的。在多肽提供于具有另外的活性成分例如另一种多肽的组合物中的情况下,则将每种所述多肽分别纯化至高水平的同质性,然后以适当的比例混合以用于各自的预期目的。例如,可以将两种多肽各自纯化至至少90%的同质性,然后以1:1的比例组合。

[0088] 多肽(或其混合物)可以以适于在使用前在水溶液中重构的冻干形式提供。冻干的组合物具有改善的稳定性,使得能够更长久地储存多肽。本文提供了一种制备冻干形式的多肽(或其混合物)的方法,所述方法包括在合适的缓冲液(例如Tris-缓冲盐水(TBS))中冷冻干燥所述多肽(或混合物)。多肽在冷冻干燥之前通常是基本上纯化的。还提供了冻干形式的所得多肽(或混合物)。还提供了一种制备多肽(或混合物)的溶液的方法,该方法包括提供冻干形式的多肽(或混合物)并用合适的运载体或稀释剂(例如水)重构。

[0089] 多肽可以使用本领域已知的方法固定,例如Datta S et al., Enzyme immobilization: an overview on techniques and support materials, 3Biotech, 3(1): 1-9 (2013) 中所述。例如,可以通过吸附、共价结合、亲和力固定或包封来固定多肽。可用作支持物的材料包括但不限于,例如天然支持物(例如琼脂糖(agarose)、胶原、明胶、纤维素、果胶、琼脂糖凝胶)、无机材料(例如陶瓷、二氧化硅、玻璃、活性炭或木炭)或合成聚合物。例如,多肽可以固定在任选地作为树脂提供的琼脂糖凝胶、或琼脂糖上。

[0090] 具有内切蛋白酶活性的多肽

[0091] 具有内切蛋白酶活性的多肽的功能特征

[0092] 在一种实施方式中,本发明涉及一种对O-糖基化蛋白具有特异性内切蛋白酶活性的多肽。换句话说,该多肽具有O-糖蛋白特异性内切蛋白酶活性。该多肽切割任何O-连接的糖蛋白,优选任何O-连接的人糖蛋白。O-连接的糖蛋白的实例包括包含免疫球蛋白的全部或部分的或由免疫球蛋白的全部或部分组成的任何蛋白质,包括全长抗体、Fc片段和Fc融合蛋白,特别是IgA、IgD和IgG3同种型的那些。O-连接糖蛋白的另一个实例是依那西普,依那西普具有许多O-糖基化位点,是人TNF α 受体2的配体结合域与IgG1的Fc部分连接的融合蛋白。O-连接糖蛋白的其他实例包括促红细胞生成素(EPO)、TNF α 受体、胎球蛋白和纤溶酶原。

[0093] 底物糖蛋白的水解(即切割)通常以高特异性发生在O-糖基化丝氨酸或苏氨酸的N末端和紧密接近O-糖基化丝氨酸或苏氨酸的肽键处,并且是O-聚糖依赖性的。本发明的多肽优选能够切割这种与底物糖蛋白中的每个O-糖基化位点紧密接近的肽键。该反应优选不显示任何氨基酸特异性或限制性,并且特别地,不需要任何特定的氨基酸存在于O-糖基化的丝氨酸或苏氨酸的N末端。当使用标准质谱参数评估时,通常观察到切割位点在与每个O-糖基化残基的N末端紧邻的肽键处。

[0094] 给定多肽的内切蛋白酶活性和特异性可以通过合适的测定手段来确定。例如,可以将标准的O-糖蛋白底物(例如IgA分子或促红细胞生成素(EPO))与测试多肽一起孵育。然后通过SDS-PAGE和/或质谱来分析起始材料和反应产物,以确定切割产物(如果有的话)的存在,并且如果需要还可以进一步表征那些产物。未O-糖基化的糖蛋白底物(例如IgG1分子)可以用作阴性对照。可以将结果与在底物与本发明的示例性多肽(例如由SEQ ID NO:2的氨基酸序列组成的多肽)接触时在相同测定中获得的结果进行比较。一个单位的SEQ ID NO:2的多肽定义为根据SDS-PAGE的监测,与一个单位唾液酸酶混合物组合在20mM Tris缓冲液(pH 6.8)中37℃下过夜消化大约90%的1 μ g促红细胞生成素(EPO)所需的量(优选的唾液酸酶混合物在下文中进一步描述)。测试多肽优选在以相同的量存在时达到相似的活性水平。示例性测定也在实施例中进行描述。

[0095] 具有内切蛋白酶活性的多肽的结构特征

[0096] 本部分列出了根据该实施方式的多肽的结构特征,除了在前一部分中概述的功能特征外,这些结构特征也适用。

[0097] 多肽的长度通常至少为150、200、250、275、280、290、300、310、320、330、340、350或360个氨基酸。多肽的长度通常不大于400、395、390、385、380、375、370或365个氨基酸。应当理解,以上列出的任何下限可以与以上列出的任何上限组合以提供多肽长度的范围。例如,该多肽的长度可以为150至400个氨基酸,或长度为280至380个氨基酸。多肽的长度优选为340至380个氨基酸,多肽的长度最优选为360至375个氨基酸。

[0098] 多肽的一级结构(氨基酸序列)基于嗜粘蛋白艾克曼菌的Amuc1119基因所编码的多肽的一级结构。该多肽的完整序列示于SEQ ID NO:4,该序列包括第1至24位上的信号基序。去除了信号基序的序列示于SEQ ID NO:1。

[0099] 本发明的多肽可以包含SEQ ID NO:1的序列、基本上由SEQ ID NO:1的序列组成或由SEQ ID NO:1的序列组成。

[0100] 可供选择地,本发明的多肽可以包含以下、基本上由以下组成或由以下组成:与SEQ ID NO:1的氨基酸序列至少50%一致的SEQ ID NO:1的氨基酸序列的变体。变体序列可

以与SEQ ID NO:1的氨基酸序列至少60%、至少70%、至少80%、至少85%、至少90%、至少95%、至少98%或至少99%一致。一致性水平优选为至少85%或更高。相对于SEQ ID NO:1的序列的一致性可以在SEQ ID NO:1所示的序列的至少100、至少200、至少300或至少350或更多个连续氨基酸的区域上测量,或更优选在SEQ ID NO:1的全长上测量。变体的长度通常比参考序列长或短不超过50个氨基酸,并且优选具有与参考序列大约(或完全)相同的长度。

[0101] 可以使用任何合适的算法来计算氨基酸一致性。例如,PILEUP和BLAST算法可用于计算一致性或将序列对齐(例如识别等效序列或相应序列(通常在其默认设置上)),如在Altschul S.F.(1993)J Mol Evol 36:290-300;Altschul,S,F et al(1990)J Mol Biol 215:403-10中所描述。进行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information;<http://www.ncbi.nlm.nih.gov/>)公开获得。该算法涉及首先通过在与数据库序列中相同长度的字比对时在匹配或满足一些正值阈值得分T的查询序列中鉴定长度为W的短字来鉴定高评分序列对(HSP)。T被称为邻近字得分阈值(Altschul et al,同上)。这些最初的邻近字命中充当启动搜索以查找包含它们的HSP的种子。字命中沿着每个序列在两个方向上延伸,直到可以增加累积比分为止。在以下情况下,停止在每个方向上的字命中扩展:累积比得分从其最大实现值下降了数量X;由于一个或多个负得分残基比对的累积,累积得分变为零或更低;或到达任一序列的末尾。BLAST算法参数W、T和X确定比对的灵敏度和速度。BLAST程序使用默认值字长(W)为11,BLOSUM62评分矩阵(请参见Henikoff and Henikoff(1992)Proc.Natl.Acad.Sci.USA 89:10915-10919)比对(B)为50,期望(E)为10,M=5,N=4,以及对两条链的比较。

[0102] BLAST算法对两个序列之间的相似性进行统计分析。参见例如Karlin and Altschul(1993)Proc.Natl.Acad.Sci.USA 90:5873-5787。BLAST算法提供了一种相似性度量是最小总和概率(P(N)),最小总和概率(P(N))提供偶然会发生两个多核苷酸或氨基酸序列匹配的概率的指示。例如,如果第一序列与第二序列相比的最小总和概率小于约1,优选小于约0.1,更优选小于约0.01,并且最优选小于约0.001,则认为该序列与另一个序列相似。可供选择地,UWGCG软件包提供BESTFIT程序,该程序可用于计算一致性(例如以其默认设置使用)(Devereux et al(1984)Nucleic Acids Research 12,387-395)。

[0103] 本发明的多肽的序列可以包含相对于SEQ ID NO:1氨基酸序列进行了修饰(例如氨基酸添加、缺失或取代)的SEQ ID NO:1序列的变体。除非另有说明,修饰优选是保守氨基酸取代。保守取代用具有相似化学结构、相似化学性质或相似侧链体积的其他氨基酸来替代氨基酸。所引入的氨基酸可以具有与它们所替代的氨基酸相似的极性、亲水性、疏水性、碱性、酸性、中性或电荷。可供选择地,保守取代可以引入另一种芳族或脂族氨基酸替代现有的芳族或脂族氨基酸。保守的氨基酸变化是本领域熟知的,并且可以根据以下表A1中定义的20种主要氨基酸的性质进行选择。在氨基酸具有相似的极性的情况下,可以通过参考表A2中氨基酸侧链的亲水性等级来确定。本发明的多肽序列可以包含SEQ ID NO:1的氨基酸序列进行了多达10、20、30、40、50或60个保守取代的变体。

[0104] 表A1-氨基酸的化学性质

[0105]

Ala (A)	脂族的、疏水的、中性的	Met (M)	1 疏水的、中性的
Cys (C)	极性的、疏水的、中性的	Asn (N)	极性的、亲水的、中性的
Asp (D)	极性的、亲水的、带负电荷的	Pro (P)	疏水的、中性的
Glu (E)	极性的、亲水的、带负电荷的	Gln (Q)	极性的、亲水的、中性的
Phe (F)	芳族的、疏水的、中性的	Arg (R)	极性的、亲水的、带正电荷的
Gly (G)	脂族的、中性的	Ser (S)	极性的、亲水的、中性的
His (H)	芳族的、极性的、亲水的、带正电荷的	Thr (T)	极性的、亲水的、中性的
Ile (I)	脂族的、疏水的、中性的	Val (V)	脂族的、疏水的、中性的
Lys (K)	极性的、亲水的、带正电荷的	Trp (W)	芳族的、疏水的、中性的
Leu (L)	脂族的、疏水的、中性的	Tyr (Y)	芳族的、极性的、疏水的

[0106] 表A2-亲水等级

	侧链	亲水性
	Ile	4.5
	Val	4.2
	Leu	3.8
	Phe	2.8
	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
[0107]	Ser	-0.8
	Trp	-0.9
	Tyr	-1.3
	Pro	-1.6
	His	-3.2
	Glu	-3.5
	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

[0108] 本发明的多肽的氨基酸序列可以包含如上所述的SEQ ID NO:1的氨基酸序列的变体。然而,SEQ ID NO:1的氨基酸序列中的某些残基优选保留在所述变体序列内。例如,所述变体序列通常保留已知内切蛋白酶活性所需的某些残基。因此,优选将SEQ ID NO:1的第182位(与SEQ ID NO:4的第206位对应)的谷氨酸保留在本发明多肽的氨基酸序列中。该残基被认为是在活性部位进行电子转移所需的。因此,本发明的多肽通常包括SEQ ID NO:1的氨基酸序列的变体,所述变体在所述变体序列的与SEQ ID NO:1的第182位对应的位置处具有谷氨酸(E)。类似地,SEQ ID NO:1的第181位(与SEQ ID NO:4的第205位对应)上的组氨酸优选保留在本发明多肽的氨基酸序列中。该残基被认为是与锌离子辅因子结合所需的。

[0109] 所述谷氨酸和所述组氨酸残基通常都包含在具有基序HEbbH的金属蛋白酶结构域中,其中b是不带电荷的氨基酸,例如氨基酸A、C、F、G、I、L、M、N、P、Q、S、T、V或W。这样的结构域的优选实例有序列HELGH(SEQ ID NO:41),该序列与SEQ ID NO:1的第181至185位对应

(与SEQ ID NO:4的第205至209位对应)。从而,本发明的多肽通常包含SEQ ID NO:1的氨基酸序列的变体,该变体在与SEQ ID NO:1的第181至185位对应的位置处包含基序HEbbH(例如HEIGH(SEQ ID NO:42)或HELGH,优选HELGH)。本发明的多肽通常包含位于金属蛋白酶结构域C末端的O-聚糖特异性结合域。

[0110] 基序HEbbH可在具有基序abxHEbbHbc的较大的金属蛋白酶结构域内受损,其中a是氨基酸V、T或G,b是不带电荷的氨基酸,例如氨基酸A、C、F、G、I、L、M、N、P、Q、S、T、V或W,x是任何氨基酸,c是疏水性氨基酸,例如A、C、F、I、L、M、P、V、W或Y。这样的结构域的优选实例具有序列GMAHELGHGL(SEQ ID NO:8),该序列与SEQ ID NO:1的第178至187位对应(SEQ ID NO:4的第202至211位)。其它实例包括GVAHELGHNF(SEQ ID NO:43)。因此,本发明的多肽优选包含SEQ ID NO:1的氨基酸序列的变体,该变体包含在与SEQ ID NO:1的第178至187位对应的位置处的基序abxHEbbHbc(例如GMAHELGHGL或GVAHELGHNF,优选GMAHELGHGL)。本发明的多肽通常包含位于金属蛋白酶结构域C末端的O-聚糖特异性结合域。

[0111] 可供选择地,本发明的多肽可包含以下,基本上由以下组成或由以下组成:SEQ ID NO:1的较短片段或其如上所述的变体。所述片段可以描述为保留O-糖蛋白特异性内切蛋白酶活性的SEQ ID NO:1的截短形式。这样的片段比SEQ ID NO:1短,并且长度通常为至少100、150或200个氨基酸。所述片段通常包括在与SEQ ID NO:1的第178至187位对应的位置处的金属蛋白酶结构域和位于金属蛋白酶结构域的C末端的O-聚糖特异性结合域,该金属蛋白酶结构域包括与SEQ ID NO:1的第182位对应的位置处的谷氨酸残基(E)和与SEQ ID NO:1的第181位对应的位置处的组氨酸残基(H)。

[0112] 包含SEQ ID NO:1的本发明的任何多肽或其变体或其任一片段可任选地在N末端包括另外的甲硫氨酸和在C末端包括组氨酸或其他标签。这样的另外的序列可有助于表达和/或纯化。组氨酸标签优选由六个组氨酸残基组成。组氨酸标签优选通过通常是短氨基酸序列(例如3至5个氨基酸)的接头连接至C末端。接头通常主要由甘氨酸和丝氨酸残基组成,并且可以优选地包括序列GSG。例如,GSG和GSGLE是合适的接头。

[0113] 因此,总之,本发明的多肽是一种具有O-糖蛋白特异性内切蛋白酶活性的多肽,所述多肽包含:

[0114] (a)SEQ ID NO:1的氨基酸序列;

[0115] (b)与SEQ ID NO:1的氨基酸序列至少85%一致的氨基酸序列;或

[0116] (c)这样的氨基酸序列:所述氨基酸序列为SEQ ID NO:1的序列的片段或与SEQ ID NO:1的氨基酸序列85%一致的氨基酸的片段;

[0117] 任选地,其中所述多肽在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸标签,该标签可以通过接头与C末端连接。

[0118] 本发明的示例性多肽的序列以SEQ ID NO:2提供。该多肽可以包含SEQ ID NO:2的氨基酸序列或由SEQ ID NO:2的氨基酸序列组成。编码该多肽的示例性多核苷酸序列示于SEQ ID NO:3。

[0119] 具有O-糖蛋白特异性内切蛋白酶活性的可供选择的多肽已在铜绿假单胞菌PA01、多形拟杆菌VPI-5482和产气荚膜梭菌中鉴定出来(参见Noach et al;PNAS 2017,pE679-688和支持附录,特别是用于克隆、蛋白表达和纯化的材料和方法(Materials and Methods for Cloning,Protein Expression and Purification)中描述的三种肽酶)。这些多肽的

全长序列以SEQ ID NO:32、33和34提供。这些序列中的每种都包括如上所述的具有基序HEbbH的金属蛋白酶结构域。产气荚膜梭菌序列也具有如上所述的具有基序abxHEbbHbc的较长的金属蛋白酶结构域。这些序列中的每种可以任选地进行修饰,以去除可能存在的任何信号序列或酶原序列和/或包括在N末端的另外的甲硫氨酸和/或包括在C末端的组氨酸或其他标签。这样的另外的序列可以帮助表达(例如在大肠杆菌中)和/或纯化。去除信号和其他未成熟序列的相应序列以SEQ ID NO:26、27和28提供。针对在大肠杆菌中表达和随后纯化(通过在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸标签)而优化的这些序列的版本以SEQ ID NO:29、30和31提供。在本文所述的具有O-糖蛋白特异性内切蛋白酶活性的本发明多肽的使用方法中,本发明的多肽可以任选地被这些多肽之一替代。因此,用于这样的方法的优选多肽包含SEQ ID NO:26至31中的任一种,基本上由SEQ ID NO:26至31中的任一种组成或由SEQ ID NO:26至31中的任一种组成。

[0120] 利用多肽内切蛋白酶活性的方法

[0121] 本发明还提供了一种水解O-糖蛋白的方法,其中所述方法包括使所述蛋白的样品与具有O-糖蛋白特异性内切蛋白酶活性的本发明的多肽接触,并且任选地还包括检测水解产物。

[0122] 本发明还可以包括一种用于评估蛋白质的糖基化状态的方法,该方法包括使所述样品与具有O-糖蛋白特异性内切蛋白酶活性的本发明的多肽接触和分析所产生的产物。切割产物的存在表明所述样品中的蛋白质是O-糖基化的,因此该方法也可用于检测O-糖蛋白。可以任选地进一步分析切割产物以鉴定聚糖链及其与蛋白质的连接位置。

[0123] 在这样的方法中,在适于多肽与样品中的任何蛋白质相互作用并且适于发生水解/切割反应(内切蛋白酶活性)的条件下,使样品与本发明的多肽接触。合适的条件包括与本发明的多肽孵育至少20分钟、30分钟、40分钟、50分钟、60分钟、70分钟、80分钟、90分钟或120分钟、3小时、5小时、10小时或过夜。孵育优选在室温下进行,更优选在约20℃、25℃、30℃、35℃、40℃或45℃下,最优选在约37℃下进行。上述方法可以在任何合适的pH下进行。合适的pH值包括例如约3.0、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9或9.5的pH。本发明多肽的活性的优选pH在5.6至6.8的范围内。该方法可以在任何合适的缓冲液中进行,例如tris缓冲盐水(TBS)或磷酸盐缓冲盐水(PBS)。本发明的多肽与样品的蛋白质含量(酶:底物)的近似比例可以是1:1、2:1、4:1、6:1、10:1、15:1、20:1、1:2、1:4或1:6、1:10、1:15、1:20、1:40、1:100、1:200或1:400。优选的比例是1:20。如果需要更短的反应时间,或者如果O-糖蛋白大量唾液酸化,则较高的酶与底物的比例可能是有益的。可供选择地,如下文中更详细地讨论的,可以使用较早或同时的唾液酸酶孵育步骤来降低唾液酸含量。底物通常以0.1mg/ml至10mg/ml,优选为约0.1mg/ml至2mg/ml的浓度存在。

[0124] 可以通过任何合适的分析方法来评估所产生产物的检测或分析,所述分析方法例如但不限于质谱、HPLC、亲和色谱、凝胶电泳、SDS-PAGE、ELISA、凝集素印迹、光谱测定、毛细管电泳和其他标准的实验室蛋白质分析技术。

[0125] 上述任何方法中的样品可以是取自患者,优选人类患者的样品。获得的结果可以用于诊断目的,例如用以检测涉及O-连接的糖基化的癌症的存在。这样的用途可以涉及将从患者样品获得的结果与使用从健康对照获得的样品得到的结果进行比较。

[0126] 在本发明的任何方法中,该多肽可以与另一种酶(例如蛋白酶或糖苷酶)组合使

用。另外的蛋白酶或糖苷酶通常将进一步消化底物蛋白,这可增强本发明的多肽的活性和/或允许更容易或更详细地分析产物。

[0127] 例如,本发明人已经确定,如果首先修饰底物蛋白的O-聚糖以去除唾液酸,则本发明的多肽表现出改善的内切蛋白酶活性。因此,在本发明的优选方法中,使样品与试剂接触以去除唾液酸。所述试剂优选可以是唾液酸酶或这样的酶的混合物,所述唾液酸酶或这样的酶的混合物可以存在于合适的缓冲液例如TBS或PBS中。缓冲液优选包含低浓度的NaCl,通常最高至300mM、250mM、200mM或150mM。NaCl浓度优选为约150mM,例如125mM至175mM。唾液酸酶(或神经氨酸酶)催化末端唾液酸从糖蛋白上的复杂碳水化合物中被切割,并显示出高度的特异性。这些酶靶向O-糖蛋白中常见的三个独特的唾液酸键,即 α 2-3、 α 2-6和 α 2-8键。适用于所述方法的唾液酸酶包括靶向所有 α 2-3、 α 2-6或 α 2-8键的广谱唾液酸酶,以及通常仅靶向一种类型的键的窄谱唾液酸酶。 α 2-3键是人类糖蛋白中最常见的键,因此,如果使用窄谱唾液酸酶,则该窄谱唾液酸酶优选靶向该键。合适的唾液酸酶可以包括病毒或哺乳动物唾液酸酶,但是优选是自细菌分出的唾液酸酶,所述细菌包括但不限于产气荚膜梭菌,产脲节杆菌(*Arthrobacter ureafaciens*),肺炎链球菌(*Streptococcus pneumoniae*),霍乱弧菌(*Vibrio cholera*)和嗜粘蛋白艾克曼菌的菌株。

[0128] 优选的窄谱唾液酸酶是从嗜粘蛋白艾克曼菌分出的Am1757。Am1757对 α 2-3键具有特异性活性。Am1757的野生型序列以SEQ ID NO:9提供,SEQ ID NO:9包括信号序列。缺少信号序列的Am1757的野生型序列以SEQ ID NO:10提供。可以任选地修饰这些序列,以在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸或其他标签。这样的另外的序列可以帮助表达(例如在大肠杆菌中)和/或纯化。组氨酸标签优选由六个组氨酸残基组成。组氨酸标签优选通过通常是短氨基酸序列(例如3至5个氨基酸)的接头连接至C末端。接头通常主要由甘氨酸和丝氨酸残基组成,并且可以优选地包括序列GSG。例如,GSG和GSGLE是合适的接头。在N末端具有另外的甲硫氨酸且在C末端具有GSGLE接头和His₆标签的示例性的Am1757序列以SEQ ID NO:11提供。本发明中任何对Am1757的提及都可以表示SEQ ID NO:9、10或11中的任一种,但优选地是指包含SEQ ID NO:10的氨基酸序列或由SEQ ID NO:10的氨基酸序列组成的多肽。最优选地是由SEQ ID NO:11的氨基酸序列组成的多肽。

[0129] 优选的广谱唾液酸酶是从嗜粘蛋白艾克曼菌分出的Am0707。Am0707对 α 2-3、 α 2-6和 α 2-8键具有活性。Am0707的野生型序列以SEQ ID NO:12提供,SEQ ID NO:12包括信号序列。缺少信号序列的Am0707的野生型序列以SEQ ID NO:13提供。可以任选地修饰这些序列,以在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸或其他标签。这样的另外的序列可以帮助表达和/或纯化。组氨酸标签优选由六个组氨酸残基组成。组氨酸标签优选通过通常是短氨基酸序列(例如3至5个氨基酸)的接头连接至C末端。接头通常主要由甘氨酸和丝氨酸残基组成,并且可以优选地包括序列GSG。例如,GSG和GSGLE是合适的接头。在N末端具有另外的甲硫氨酸且在C末端具有GSGLE接头和His₆标签的示例性的Am0707序列以SEQ ID NO:14提供。本发明中任何对Am0707的提及都可以表示SEQ ID NO:12、13或14中的任一种,但优选地是指包含SEQ ID NO:13的氨基酸序列或由SEQ ID NO:13组成的多肽。最优选地,是由SEQ ID NO:14的氨基酸序列组成的多肽。

[0130] 能够水解所有唾液酸键的优选的唾液酸酶混合物包含从嗜粘蛋白艾克曼菌分出的Am1757和Am0707。Am1757和Am0707的比例通常为1:1。特别优选的混合物可以包含由SEQ

ID NO:11的氨基酸序列组成的多肽和由SEQ ID NO:14的氨基酸序列组成的多肽。

[0131] 本发明的方法可优选包括在适合唾液酸酶活性的条件下,在本发明的多肽之前或与本发明的多肽同时,将样品与Am1757或与Am1757和Am0707的混合物一起孵育。本发明还提供了包含本发明的多肽以及Am1757或Am1757和Am0707的混合物的组合物(冻干或溶液形式)。这样的组合物可以优选在pH可约为7.6的tris缓冲盐水中冻干。在这样的组合物中,Am1757和Am0707将优选彼此以1:1的比例存在,并且总唾液酸酶含量(Am1757+Am0707)相对于本发明的多肽也以1:1的比例存在。例如,如果组合物包含2000个单位的本发明的多肽,则该组合物还将包含2000个单位的唾液酸酶,其中所述2000个单位的唾液酸酶包含1000个单位的Am1757和1000个单位的Am0707。一个单位唾液酸酶混合物通常是当在20mM Tris pH 6.8中在37℃下孵育2小时时,根据SDS-PAGE所监测的,从大约90%的1μg糖蛋白(胎球蛋白)中水解唾液酸所需的量。一个单位的本发明的多肽通常是当在20mM Tris缓冲液(pH 6.8)中孵育时,根据SDS-PAGE所监测的,与一个单位唾液酸酶混合物在37℃下过夜消化大约90%的1μg促红细胞生成素(EPO)所需的量。

[0132] 本发明还提供了一种试剂盒,该试剂盒在单独的容器中包含来自Am1757或Am1757和Am0707的混合物的本发明的多肽,以及关于不同酶的组合使用的说明。

[0133] 作为另一个实例,在本文所述的任何方法中,在将样品与本发明的多肽接触之前、同时或之后,可以将样品与N-糖苷酶一起孵育,以从靶蛋白去除N-聚糖。示例性的N-糖苷酶是PNGaseF。当样品包括免疫球蛋白时,可以使用的其他N-糖苷酶是EndoS(参见W02008071418的SEQ ID NO:1)或EndoS2(可以称为EndoS49,参见W02013037824的SEQ ID NO:1)。这些酶中的每一种均从IgG1的Asn-297去除N-连接的糖蛋白。在特别的实施方式中,除了本发明的多肽之外,样品还可以与N-糖苷酶和唾液酸酶(或其混合物)接触。在这样的方法中,可以在同时添加N-糖苷酶和本发明的多肽之前首先应用唾液酸酶(或混合物)。这种方法特别适合于随后的O-糖基化位点的评估,通常通过分离产物(例如使用RPLC)和对不同的级分的后续分析(例如使用质谱)来实现。

[0134] 作为另一个实例,在本文所述的任何方法中,在将样品与本发明的多肽接触之前、同时或之后,可以将样品与蛋白酶一起孵育以进一步消化靶蛋白。合适的一般蛋白酶包括胰蛋白酶、胰凝乳蛋白酶、Lys-C、Asp-N、Glu-C、Arg-C或类似的内切蛋白酶、或牙龈卟啉单胞菌(*Porphyromonas gingivalis*)的Arg-牙龈蛋白酶(RgpB)。

[0135] 如果样品包括免疫球蛋白,则可以使用免疫球蛋白蛋白酶,例如SpeB(参见W02015040125中的序列),化脓性链球菌(*S. pyogenes*)的免疫球蛋白G降解酶(IdeS-参见W02015040125中的序列),马链球菌兽疫亚种的免疫球蛋白G降解酶(IdeZ),牙龈卟啉单胞菌的Lys-牙龈蛋白酶(Kgp)和无乳链球菌(*S. agalactiae*)的免疫球蛋白G降解酶(IgdE_{无乳}-参见PCT/EP2017/052463的SEQ ID NO:3)。在本发明的方法中使用这些蛋白酶的任何组合可有助于例如使用质谱(自中向下(middle down)法)确定单克隆抗体及其亚基上的O-糖基化位点。

[0136] 作为另一个实例,在本文描述的任何方法中,在使样品与本发明的多肽接触之后,可以将样品与O-糖苷酶一起孵育。例如,为了简化对产生的产物的分析,在通过任何合适的方法进一步分析之前,通过O-糖苷酶使产物进行消化以除去O-聚糖。合适的O-糖苷酶可从以下细菌获得:粪肠球菌(*Enterococcus faecalis*)、口腔链球菌(*Streptococcus*

oralis)、两歧双歧杆菌(*Bifidobacterium bifidum*),优选粪肠球菌或口腔链球菌,最优选口腔链球菌。来自口腔链球菌的示例性O-糖苷酶的序列以SEQ ID NO:15提供。

[0137] 与O-糖蛋白结合但缺乏内切蛋白酶活性或具有降低的内切蛋白酶活性的多肽

[0138] 缺乏内切蛋白酶活性的多肽的功能特征

[0139] 在一种实施方式中,本发明涉及一种在保持结合O-聚糖的能力的同时缺乏内切蛋白酶活性或具有降低的内切蛋白酶活性的多肽。换句话说,该多肽可以被描述为不显著水解连接有所述聚糖的糖蛋白的O-聚糖特异性结合剂。

[0140] O-糖蛋白内切蛋白酶活性可以使用任何合适的方法来确定,但是通常可以采用与上文针对具有这样的活性的本发明多肽所述相同的测定法。与O-糖蛋白底物一起孵育后,不存在切割产物将表明测试多肽中缺乏活性。包含SEQ ID NO:1的氨基酸序列的多肽对相同底物的切割可以用作阳性对照。可以通过与相同对照比较来确定测试多肽的活性的降低。本发明的多肽通常具有相对于包含SEQ ID NO:1的氨基酸序列的多肽的活性降低的O-糖蛋白内切蛋白酶活性。与SEQ ID NO:1氨基酸序列的多肽的活性相比,本发明的多肽通常具有小于95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%或5%的O-糖蛋白内切蛋白酶活性。

[0141] 多肽与O-聚糖或O-糖蛋白结合的能力也可以通过任何合适的方法来评估。一种这样的方法涉及将测试多肽固定在例如旋转柱中的琼脂糖凝胶上,然后与含有O-糖蛋白和/或O-聚糖的样品一起孵育。如果测试多肽具有O-聚糖和/或O-糖蛋白结合能力,则可检测到O-糖蛋白和/或O-聚糖结合到色谱柱或在随后的洗脱液中。优选地,该多肽能够结合可被包含SEQ ID NO:1的氨基酸序列的多肽水解的所有O-糖蛋白。

[0142] 在实施例中描述了该类型的示例性测定。

[0143] 缺乏内切蛋白酶活性的多肽的结构特征

[0144] 本部分列出了根据该实施方式的多肽的结构特征,除了在前面部分中概述的功能特征外,这些结构特征也适用。除了通过一个或多个氨基酸的添加、缺失或取代来修饰氨基酸序列使得所述活性降低或消除之外,根据本发明该实施方式的多肽可具有与具有内切蛋白酶活性的多肽相同的上述结构特征。通常,根据本发明该实施方式的多肽将不包括完整的金属蛋白酶基序HEbbH或abxHEbbHbc。所述基序可以通过添加、缺失或取代而被破坏,优选通过至少一个氨基酸取代而被破坏。优选地,取代涉及用可供选择的氨基酸取代所述基序中的谷氨酸(E)残基和/或替代与较短的基序的第1位(较长的基序的第4位)对应的位置中的组氨酸(H)残基和/或替代与较短的基序的第5位(较长的基序的第8位)对应的位置中的组氨酸(H)残基。优选地,三种所述取代中的一种或两种或全部是非保守的。E残基的取代应减少或消除电子转移。H残基中任一个的取代应减少或消除锌离子辅因子结合。因此,E残基优选被非极性或不带电荷的氨基酸取代,例如A、C、F、G、I、L、M、N、P、Q、S、T、V或W,但是大多数优选被丙氨酸(A)或甘氨酸(G)取代。H残基可各自独立地被任何非H氨基酸取代,但是非极性氨基酸(例如A和G)仍然是优选的。

[0145] 因此,本发明的多肽可以包含SEQ ID NO:1的氨基酸序列、基本上由SEQ ID NO:1的氨基酸序列组成或由SEQ ID NO:1的氨基酸序列组成,在SEQ ID NO:1的氨基酸序列中,金属蛋白酶基序HEbbH或abxHEbbHbc优选通过用可供选择的氨基酸替代在与SEQ ID NO:1的182位对应的位置处的谷氨酸残基和/或替代在与SEQ ID NO:1的181位对应的位置处的

位置处的组氨酸残基和/或用可供选择的氨基酸替代在SEQ ID NO:1的185位对应的位置处的组氨酸残基而被破坏。换句话说,该多肽可被描述为不包含金属蛋白酶基序HEbbH,并且优选包含所述基序的破坏形式,使得:

[0146] (a) 第一位的H被可供选择的氨基酸替代,所述可供选择的氨基酸优选A或G;和/或

[0147] (b) 第二位的E被不带电荷的氨基酸替代,所述不带电荷的氨基酸任选地为A、C、F、G、I、L、M、N、P、Q、S、T、V或W,优选为A或G;和/或

[0148] (c) 第五位的H被可供选择的氨基酸,优选A或G替代,

[0149] 其中所述基序中的b是不带电荷的氨基酸,任选地为A、C、F、G、I、L、M、N、P、Q、S、T、V或W。

[0150] 因此,所述多肽可以描述为包含基序xbbbx,其中:

[0151] (a) x优选为除H以外的任何氨基酸,并且优选为A或G;和/或

[0152] (b) b是不带电荷的氨基酸,任选地为A、C、F、G、I、L、M、N、P、Q、S、T、V或W,优选A或G;

[0153] 任选地,其中所述基序存在于所述多肽中与SEQ ID NO:1的第181至185位相对应的位置。

[0154] 因此,所述多肽可以包含破坏的金属蛋白酶基序,例如具有以下序列中的任何一个:HALGH(SEQ ID NO:44)、AELGH(SEQ ID NO:45)或最优选AALGH(SEQ ID NO:46)。包含这种类型的对SEQ ID NO:1的特异性改变的序列示于SEQ ID NO:5和SEQ ID NO:20。换句话说,因此,本发明该实施方式的多肽可以包含SEQ ID NO:5或SEQ ID NO:20的氨基酸序列、基本上由SEQ ID NO:5或SEQ ID NO:20的氨基酸序列组成或由SEQ ID NO:5或SEQ ID NO:20的氨基酸序列组成。

[0155] 所述多肽可替代地被描述为包含基序abxxbbbxbc,其中:

[0156] (a) a为氨基酸V、T或G;

[0157] (b) b是不带电荷的氨基酸,任选地为A、C、F、G、I、L、M、N、P、Q、S、T、V或W,优选A或G;

[0158] (c) x是任何氨基酸,除了基序的第4位和/或第8位的氨基酸优选不是N,并且优选是A或G;和

[0159] (d) c为疏水性氨基酸,任选地为A、C、F、I、L、M、P、V、W或Y;任选地,其中所述基序存在于所述多肽中在与SEQ ID NO:1的第178至187位对应的位置处。

[0160] 因此,所述多肽可以包含破坏的金属蛋白酶基序,例如具有以下序列中的任一种:GMAHALGHGL(SEQ ID NO:23),GMAAELGHGL(SEQ ID NO:24)或最优选GMAAALGHGL(SEQ ID NO:25)。包含这种类型的对SEQ ID NO:1的特异性改变的序列示于SEQ ID NO:5和SEQ ID NO:20。因此,换句话说,本发明该实施方式的多肽可以包含SEQ ID NO:5或SEQ ID NO:20的氨基酸序列、基本上由SEQ ID NO:5或SEQ ID NO:20的氨基酸序列组成或由SEQ ID NO:5或SEQ ID NO:20的氨基酸序列组成。

[0161] 可供选择地,本发明的多肽可以包含以下、基本上由以下组成或由以下组成:与SEQ ID NO:5的氨基酸序列至少50%一致的SEQ ID NO:5或SEQ ID NO:20的氨基酸序列的变体,条件是在与SEQ ID NO:1的第182位对应的位置处不引入谷氨酸残基和/或在与SEQ ID NO:1的第181位对应的位置处不引入组氨酸残基和/或在与SEQ ID NO:1的第185位对应的位置处不引入组氨酸残基。

[0162] 变体序列可以与SEQ ID NO:5的序列至少60%、至少70%、至少80%、至少85%、至

少90%、至少95%、至少98%或至少99%一致。一致性水平优选为至少85%或更高。相对于SEQ ID NO:5或SEQ ID NO:20的序列的一致性可以在SEQ ID NO:5或SEQ ID NO:20所示的序列的至少100、至少200、至少300或至少350或更多个连续氨基酸的区域上测量,或更优选在SEQ ID NO:5或SEQ ID NO:20的全长上测量。本发明的多肽的序列可以包含SEQ ID NO:5或SEQ ID NO:20的进行了多达10、20、30、40、50或60个保守取代的氨基酸序列的变体。序列一致性的确定以及对保守和非保守取代的解释在与具有内切蛋白酶活性的多肽有关的部分中提供,并且在此同样适用。

[0163] 可供选择地,本发明的多肽可以包含以下、基本上由以下组成或由以下组成:SEQ ID NO:5或SEQ ID NO:20的较短片段或其如上所述的变体。所述片段可以描述为SEQ ID NO:5或SEQ ID NO:20的保留O-糖蛋白结合活性的截短形式。这样的片段比SEQ ID NO:1短,并且长度通常为至少100、150或200个氨基酸。

[0164] 包含SEQ ID NO:5或SEQ ID NO:20或其变体或其任一片段的本发明的任何多肽可任选地在N末端包括另外的甲硫氨酸和在C末端包括组氨酸或其他标签。这样的另外的序列可有助于表达和/或纯化。组氨酸标签优选由六个组氨酸残基组成。组氨酸标签优选通过通常是短氨基酸序列(例如3至5个氨基酸)的接头连接至C末端。接头通常主要由甘氨酸和丝氨酸残基组成,并且可以优选地包括序列GSG。例如,GSG和GSGLE是合适的接头。

[0165] 因此,总而言之,本发明的多肽是具有O-糖蛋白结合活性但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的多肽,该多肽包括:

[0166] (a) SEQ ID NO:5或SEQ ID NO:20的氨基酸序列;

[0167] (b) 与SEQ ID NO:5或SEQ ID NO:20的氨基酸序列至少85%一致的氨基酸序列;

[0168] (c) 这样的氨基酸序列:所述氨基酸序列为SEQ ID NO:5或SEQ ID NO:20的序列的片段或与SEQ ID NO:5或SEQ ID NO:20的氨基酸序列85%一致的氨基酸的片段;

[0169] 任选地,其中所述多肽在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸标签,该标签可以通过接头与C末端连接。

[0170] 本发明的示例性多肽的序列以SEQ ID NO:6提供。该多肽可以包含SEQ ID NO:6的氨基酸序列或由SEQ ID NO:6的氨基酸序列组成。编码该多肽的示例性多核苷酸序列示于SEQ ID NO:7。本发明的另一个示例性多肽的序列以SEQ ID NO:21提供。该多肽可以包含SEQ ID NO:21的氨基酸序列或由SEQ ID NO:21的氨基酸序列组成。编码该多肽的示例性多核苷酸序列示于SEQ ID NO:22。

[0171] 所述多肽优选以固定形式提供,例如在琼脂糖或琼脂糖凝胶上,任选地以树脂形式提供。

[0172] 具有O-糖蛋白结合活性但缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明的另外的多肽可以通过破坏包含金属蛋白酶结构域基序HEbbH或abxHEbbHbc的具有O-糖蛋白特异性内切蛋白酶活性的任何其他多肽中的金属蛋白酶结构域基序HEbbH或abxHEbbHbc来产生。在本文所述的缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明多肽的使用方法中,对本发明多肽的提及包括这样的多肽。优选如上所述地实现对所述基序的破坏,使得:

[0173] (a) 第一位的H被可供选择的氨基酸替代,所述可供选择的氨基酸优选A或G;和/或

[0174] (b) 第二位的E被不带电荷的氨基酸替代,所述不带电荷的氨基酸任选地为A、C、F、

G、I、L、M、N、P、Q、S、T、V或W,优选为A或G;和/或

[0175] (c) 第五位的H被可供选择的氨基酸,优选A或G替代,

[0176] 其中所述基序中的b是不带电荷的氨基酸,任选地是A、C、F、G、I、L、M、N、P、Q、S、T、V或W。

[0177] 因此,所述多肽可以描述为包含基序xbbbx,其中:

[0178] (a) x优选为除H以外的任何氨基酸,并且优选为A或G;和/或

[0179] (b) b是不带电荷的氨基酸,任选地为A、C、F、G、I、L、M、N、P、Q、S、T、V或W,优选A或G;

[0180] 上文描述了具有O-糖蛋白特异性内切蛋白酶活性并可能以这种方式被破坏的其他多肽,这些多肽已在铜绿假单胞菌PA01、多形拟杆菌VPI-5482和产气荚膜梭菌中被鉴定出(参见Noach et al;PNAS 2017,pE679-688和支持附录,特别是用于克隆、蛋白表达和纯化的材料和方法(Materials and Methods for Cloning,Protein Expression and Purification)中描述的三种肽酶)。这些多肽的全长序列以SEQ ID NO:32、33和34提供。相应的成熟序列(例如,去除信号和其他序列)以SEQ ID NO:26、27和28提供。针对在大肠杆菌中表达和随后纯化(通过在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸标签)而优化的这些序列的变体以SEQ ID NO:29、30和31提供。因此,SEQ ID NO:26至34中每个都包含具有基序HEbbH的金属蛋白酶结构域以产生本发明的另外的多肽,所述基序HEbbH可以被破坏以产生如上所述的基序xbbbx。HEbbH基序已被如此破坏的SEQ ID NO:26、27和28的变体以SEQ ID NO:35、36和37提供。针对在大肠杆菌中表达和随后纯化(通过在N末端包括另外的甲硫氨酸和/或在C末端包括组氨酸标签)而优化的这些序列的变体以SEQ ID NO:38、39和40提供。缺乏O-糖蛋白特异性内切蛋白酶活性或具有降低的O-糖蛋白特异性内切蛋白酶活性的本发明的多肽可包含以下,基本上由以下组成或由以下组成:SEQ ID NO:35、36、37、38、39或40中的任一种。

[0181] 使用缺少内切蛋白酶活性或具有降低的内切蛋白酶活性的LS突变体的方法

[0182] 本发明还提供了与O-聚糖结合的方法,其中该方法包括使包含O-聚糖的样品与本发明的能够与O-聚糖结合并且缺乏O-糖基化蛋白特异性内切蛋白酶活性或具有降低的O-糖基化蛋白特异性内切蛋白酶活性的多肽接触。该方法任选地进一步包括确定是否已经结合O-聚糖和/或从所得混合物中分离出O-聚糖和任何连接的糖蛋白。

[0183] 本发明还可以包括一种用于评估蛋白质的糖基化状态的方法,该方法包括使所述样品与本发明的能够与O-聚糖结合并且缺乏O-糖基化蛋白特异性内切蛋白酶活性或具有降低的O-糖基化蛋白特异性内切蛋白酶活性的多肽接触,以及确定蛋白质是否与所述多肽结合。

[0184] 本发明还可以包括用于检测样品中的O-连接的糖蛋白的方法,其中该方法包括使所述样品与本发明的能够与O-聚糖结合并且缺乏O-糖基化蛋白特异性内切蛋白酶活性或具有降低的O-糖基化蛋白特异性内切蛋白酶活性的多肽接触,从而允许形成O-连接的糖蛋白-多肽复合物。该方法可以任选地包括从接触的样品中分离所述多肽,以及确定分离的多肽是否与O-连接的糖蛋白结合,从而确定样品中O-连接的糖蛋白存在或不存在。该方法还可用于从包含O-聚糖或O-连接的糖蛋白的样品中分出O-聚糖或O-连接的糖蛋白。

[0185] 在这样的方法中,在适于使多肽与样品中的任何O-聚糖或蛋白质相互作用并发生结合的条件下,使样品与本发明的多肽接触。合适的条件包括与本发明的多肽孵育至少20

分钟、30分钟、40分钟、50分钟、60分钟、70分钟、80分钟、90分钟或120分钟、3小时、5小时、10小时或过夜,通常伴随混合,例如颠倒混合。孵育优选在室温下进行,更优选在约20℃、25℃、30℃、35℃、40℃或45℃下,最优选在约37℃下进行。上述方法可以在任何合适的pH下实施。合适的pH值包括例如约3.0、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9或9.5的pH。对于本发明多肽的活性,优选pH为5.6至6.8。该方法可以在任何合适的缓冲液中进行,例如tris缓冲盐水(TBS)或磷酸盐缓冲盐水(PBS)。本发明的多肽与样品的蛋白质含量的近似比例可以是1:1、2:1、4:1、6:1、10:1、15:1、20:1、1:2、1:4或1:6、1:10、1:15、1:20、1:40、1:100、1:200或1:400(wt:wt)。优选的比例是1:1(wt:wt)。如果需要更短的反应时间,或者如果O-糖蛋白大量唾液酸化,则较高的多肽与底物的比例可能是有益的。可供选择地,如下文中更详细地讨论的,可以使用较早或同时的唾液酸酶孵育步骤来降低唾液酸含量。底物的浓度通常为约0.01mg/ml至10mg/ml,优选为约0.1mg/ml至10mg/ml、约0.01mg/ml至2mg/ml或约0.1mg/ml至2mg/ml。

[0186] 可以通过任何合适的分析方法来评估样品的检测或分析以确定是否已经结合O-聚糖或O-连接的糖蛋白,所述分析方法例如但不限于质谱、HPLC、亲和色谱、凝胶电泳、SDS-PAGE、ELISA、凝集素印迹、光谱测定、毛细管电泳和其他标准的实验室蛋白质分析技术。例如,可以分析多肽的分子量。与未结合O-聚糖或O-连接的糖蛋白的多肽相比,与O-聚糖或O-连接的糖蛋白结合的本发明的多肽将具有更高的分子量。

[0187] 结合的O-聚糖或O-连接的糖蛋白和本发明的多肽的分离可以通过任何合适的分离手段实施。例如,分离手段可包括磁性纳米颗粒群。这些可以使用磁场分离,优选高梯度磁场分离来从样品中分离。试剂或分离手段的实例是能够结合本发明多肽的磁性颗粒群。例如,在用组氨酸标签使多肽衍生化的情况下,磁性颗粒在其表面上包含带有镍、铜或锌离子的螯合基团。可供选择地,在用生物素标签使多肽衍生化的情况下,磁性颗粒在其表面上包含链霉亲和素。

[0188] 分离手段还可以包括固定有本发明的多肽的固体支持物。固体支持物的实例包括在先前部分中描述的那些,并且可以包括琼脂糖或琼脂糖凝胶树脂、交联的琼脂糖珠粒或类似物。支持物可以用作亲和色谱柱中的基质。可供选择地,固体支持物可包含可直接吸附本发明的多肽的合适的基于二氧化硅的材料或聚苯乙烯、或塑料容器,例如微量滴定板或等同物。

[0189] 替代的分离手段包括包含可以通过本领域标准方法产生的对本发明的多肽的具有特异性的抗体的试剂。在这种意义上的抗体包括单克隆抗体、多克隆抗体、单链抗体、嵌合抗体、CDR移植抗体或人源化抗体。该抗体可以是完整的免疫球蛋白分子或其片段,例如Fab、F(ab')₂或Fv片段。如果存在一种以上的抗体,则这些抗体优选具有不同的非重叠决定簇,使得这些抗体优选可以同时结合本发明的多肽。抗体可以结合到固体支持物上,或者可以标记或缀合到另一个化学基团或分子上,以帮助抗体分离或分出。例如,典型的化学基团包括荧光标记(例如荧光素(FITC)或藻红蛋白(PE))或标签(例如生物素)。

[0190] 其他合适的分离手段包括通过使来自接触的样品的多肽与合适的洗脱缓冲液接触以从(通常是固定的)多肽洗脱蛋白质。洗脱缓冲液的选择可能取决于蛋白质的酸敏感性。优选的洗脱缓冲液可包含高摩尔浓度的尿素(通常至少5M、6M、7M或最优选至少8M)或高浓度的去污剂(通常至少约1%,5%或10%)。合适的去污剂包括Nonidet P40、Triton X-

100、Tween 20、CHAPS、脱氧胆酸钠和RapiGest SF表面活性剂,但是十二烷基硫酸钠(SDS)是优选的。高摩尔尿素比去污剂更优选,因为下游程序更可能对去污剂的存在是敏感的。

[0191] 另一种优选的洗脱缓冲液包含本发明的具有O-糖蛋白内切蛋白酶活性的多肽,例如SEQ ID NO:1的多肽。该多肽在O-聚糖处的切割将释放结合的O-糖蛋白,从而消除对基于尿素或去污剂的洗脱液的需求。

[0192] 实施例中证明了从本发明的固定的多肽洗脱O-糖蛋白的优选方法。

[0193] 上述任何方法中的样品可以是取自患者(优选人类患者)的样品。获得的结果可以用于诊断目的,例如用于检测涉及O-连接的糖基化的癌症的存在。这样的用途可以涉及将获得自患者样品的结果与使用自健康对照获得的样品得到的结果进行比较。

[0194] 在本发明的任何方法中,该多肽可以与另一种酶(例如蛋白酶或糖苷酶)组合使用。另外的蛋白酶或糖苷酶通常将进一步消化底物蛋白质或聚糖,这可以使得对产物进行更容易或更详细的分析。

[0195] 例如,本发明的多肽可以与试剂结合使用以去除唾液酸。所述试剂可以优选是唾液酸酶或以上部分中所述的这些酶的混合物。本发明还提供了包含本发明的多肽和Am1757或Am1757和Am0707的混合物的组合物(冻干或溶液形式)。本发明还提供一种试剂盒,该试剂盒在分离的容器中包含来自Am1757或Am1757和Am0707的混合物的本发明的多肽,以及关于不同酶组合使用的说明。

[0196] 作为另一个实例,在本文所述的任何方法中,在将样品与本发明的多肽接触之前、同时或之后,可以将样品与N-糖苷酶一起孵育,以从靶蛋白去除N-聚糖。示例性的N-糖苷酶是PNGaseF。当样品包括免疫球蛋白时可以使用其他N-糖苷酶是EndoS(参见W02008071418的SEQ ID NO:1)或EndoS2(可以称为EndoS49-参见W02013037824的SEQ ID NO:1)。这些酶中的每一种均从IgG1的Asn-297去除N-连接的糖蛋白。除了本发明的多肽之外,样品还可以与N-糖苷酶和唾液酸酶(或其混合物)接触。在这样的方法中,可以在同时添加N-糖苷酶和本发明的多肽之前首先应用唾液酸酶(或混合物)。

[0197] 作为另一个实例,在本文所述的任何方法中,在将样品与本发明的多肽接触之前、同时或之后,可以将样品与蛋白酶一起孵育以进一步消化靶蛋白。合适的一般蛋白酶包括胰蛋白酶、胰凝乳蛋白酶、Lys-C、Asp-N、Glu-C、Arg-C或类似的内切蛋白酶、或牙龈卞啉单胞菌的Arg-牙龈蛋白酶(RgpB)。

[0198] 如果样品包括免疫球蛋白,则可以使用免疫球蛋白蛋白酶,例如SpeB(参见W02015040125中的序列),化脓性链球菌(*S. pyogenes*)的免疫球蛋白G降解酶(IdeS-参见W02015040125中的序列),马链球菌兽疫亚种的免疫球蛋白G降解酶(IdeZ),牙龈卞啉单胞菌的Lys-牙龈蛋白酶(Kgp)和无乳链球菌(*S. agalactiae*)的免疫球蛋白G降解酶(IgDE_{无乳}-参见PCT/EP2017/052463的SEQ ID NO:3)。在本发明的方法中使用这些蛋白酶的任何组合可有助于例如使用质谱来分析底物蛋白质或聚糖。

[0199] 作为另一个实例,在本文描述的任何方法中,在通过任何合适的方法进一步分析之前,可以将分出的O-连接的糖蛋白与O-糖苷酶一起孵育以去除O-聚糖。合适的O-糖苷酶可从以下细菌获得:粪肠球菌、口腔链球菌、或两歧双歧杆菌,优选粪肠球菌或口腔链球菌,最优选口腔链球菌。来自口腔链球菌的示例性O-糖苷酶的序列以SEQ ID NO:15提供。

[0200] 以下实施例说明了本发明:

[0201] 实施例1

[0202] 材料和方法

[0203] LS的诱变

[0204] 使用引物E206A_fwd 5'-ATGGCGCACGC GCTGGGCCACG-3' 和5'-GCCACCGTAC CATTTCGTC-3' (rev), 根据制造商的说明(退火温度68℃, 延伸3分钟), 使用Q5 (NEB) 进行定点诱变; 因此, 将来自嗜粘蛋白艾克曼菌的Amuc1119基因中的谷氨酸变为丙氨酸, 从而产生突变体Amuc1119_{E206A} (LS_{E206A})。将构建体转化到DH5α大肠杆菌中, 分离并使用测序 (GATC Biotech) 来验证。

[0205] LS和LS_{E206A}的重组表达

[0206] 将来自嗜粘蛋白艾克曼菌ATCC BAA-835的基因Amuc1119和突变体Amuc1119_{E206A} (Amuc1119-LS; Amuc1119_{E206A}-LS_{E206A}) 针对在大肠杆菌 (DNA 2.0) 中表达进行优化, 并克隆到具有作为融合蛋白的一部分的C末端6xHis标签的表达载体中。

[0207] 将经密码子优化的基因转化到BL21 (DE3) Star细胞中。在LB中在37℃下在180rpm下常规地培养大肠杆菌。在质粒存在下, 添加50μg/mL卡那霉素。过夜孵育后, 将培养物以1:20稀释在新鲜的LB (kana) 中, 并使培养物生长至OD₆₂₀ 约为0.7至0.8, 之后通过添加1mM IPTG诱导重组蛋白表达, 并且表达持续6小时, 然后收集细胞并冷冻。将冷冻的细胞解冻并溶解在His结合缓冲液中 (20mM NaP pH 7.4, 500mM NaCl, 20mM咪唑) 中, 进行超声处理以释放细胞内蛋白质。通过离心去除细胞碎片。在镍柱上对无菌的过滤的上清液进行亲和纯化, 并在PD-25柱上重新缓冲至20mM Tris-HCl pH 8.0中。使用Nanodrop确定蛋白质的浓度, 并通过SDS-PAGE估算纯度。

[0208] 使用蛋白质底物进行活性评估

[0209] 将TNFα与LS以2:1的比例混合, 并在37℃下孵育15至60分钟, 然后在4%至20% Novex梯度SDS-PAGE上分离蛋白质。研究了NaCl (0至1M)、二价阳离子、EDTA和pH对LS活性的影响, 并使用Gel Doc EZ (BioRad) 通过光密度分析法测量了生成的水解片段的差异。

[0210] 活性的时间和剂量依赖性

[0211] 将TNFαR (0.5μg) 与不同剂量的LS在PBS中于37℃孵育15或60分钟, 然后在4% - 20% Novex梯度SDS-PAGE上分离蛋白质。使用生成的片段的强度 (光密度测定) 确定有效孵育条件的最佳剂量和时间。

[0212] 底物特异性

[0213] LS与各种N-连接和O-连接的底物在37℃下以2:1的比例 (底物:酶) 孵育过夜。LS与EPO (0.3mg/ml) 以50:1的比例 (底物:酶) 孵育。将蛋白质分离, 并在4%至20% Novex梯度SDS-PAGE凝胶上进行分析。

[0214] LS_{E206A}的固定

[0215] 将LS_{E206A}重悬于偶联缓冲液 (0.2M NaHCO₃, 0.5M NaCl pH 8.3) 中, 并浓缩至20mg/ml。根据制造商的说明, 准备了NHS活化的琼脂糖4Fast Flow (GE Healthcare) 以进行偶联 (例如, 在偶联缓冲液中进行HCl洗涤和平衡)。LS_{E206A}通过用琼脂糖在4℃下过夜孵育而固定, 缓慢摇动以保持恒定混合。通过添加0.1M Tris pH 8.5封闭琼脂糖, 用0.1M Tris pH 8.5/0.1M NaAc, 0.5M NaCl pH 5.0重复洗涤3次, 并保存在EtOH中直至使用。

[0216] LS_{E206A}的结合亲和力

[0217] PBS平衡的具有50 μ l固定的LS_{E206A}的旋转柱与10 μ g糖蛋白一起孵育,所述糖蛋白用唾液酸混合物(Am0707:Am1757)或用唾液酸和口腔链球菌内- α -N-乙酰半乳糖苷酶(例如O-糖苷酶)的组合进行了预处理。将样品在37 $^{\circ}$ C下孵育2小时,此后用PBS(10体积;100g,30s)洗涤色谱柱,并用0.1M甘氨酸pH 3.0洗脱。在SDS-PAGE上分析级分。

[0218] 质谱分析

[0219] 依那西普(Enbrel[®])是经临床批准的与TNF α 结合的Fc融合蛋白。依那西普含有几种O-聚糖。为了测试酶促切割特异性,将内切蛋白酶与依那西普一起在37 $^{\circ}$ C下孵育过夜。为了简化质谱分析,使用唾液酸酶和O-糖苷酶(过夜,在PBS中,所有单种酶的比例为1:40)进行了第二轮酶促处理,以去除残留的O-聚糖。在通过C18反相液相色谱使肽分离之后,通过MS/MS分析产生的肽。

[0220] 结果

[0221] LS是推定的金属蛋白酶

[0222] 基于序列和结构域的相似性,LS与几种金属蛋白酶具有同源性,包含推定的活性位点序列GMAHELGHGL,从而与一般的金属蛋白酶序列abxHEbbHbc(a=V/T,b=不带电荷的,c=疏水性的)共有相似性。组氨酸通常参与底物结合和Zn²⁺亲和,而谷氨酸与组氨酸一起介导电子转移,从而介导水解作用。为了能够进一步表征该酶,我们构建了LS_{E206A}突变体,该突变体能够结合底物,但是通过将E改变为A而缺乏水解能力或具有降低的水解能力。进一步修饰(例如,将H改变为A)对于完全失活可能是必需的。两种构建体均表达良好,并且容易地基于其His标签使用亲和色谱法进行纯化(图1)。

[0223] LS特异性水解带有O-聚糖的糖蛋白

[0224] 为了研究LS的底物特异性,将蛋白酶与多种蛋白质一起孵育。如图2所示,将LS与IgA和赫赛汀(曲妥珠单抗)一起孵育。LS仅能作用于具有O-连接的聚糖的蛋白质,例如IgA。尽管末端唾液酸的存在似乎部分地抑制LS的活性,但唾液酸的存在并不是水解的先决条件(图4)。

[0225] LS可以在多种条件下作用于O-连接的糖蛋白

[0226] 实施SDS-PAGE凝胶的光密度分析,以评估LS的酶促性质。LS在大多数条件下是有活性的,优选略酸性的pH值和低NaCl浓度(图3A-B)。尽管Mg²⁺和Ca²⁺离子均积极影响LS的水解活性,但Zn²⁺的存在显著降低了该活性,而EDTA则完全消除该活性(图3C-D)。

[0227] O-连接的半乳糖苷酶残基对LS的活性至关重要

[0228] 尽管在不存在末端唾液酸的情况下具有增加的活性,但是O-聚糖中其他碳水化合物对LS活性的重要性还没有被完全理解。尽管在不存在末端唾液酸的情况下LS的活性显著增加,但是去除半乳糖完全抑制了LS的活性(图4A)。此外,如过夜孵育后的完全水解所证明,LS对唾液酸化蛋白的活性较低不是由于在唾液酸存在下不能使键水解(图4B)。LS的活性完全依赖于O-聚糖,因为去除N-聚糖不影响LS的水解(图4C)。

[0229] O-连接的聚糖引导LS的切割位点

[0230] 已经证明了O-聚糖对于活性至关重要,我们接下来试图研究LS的特定切割位点。使用质谱,我们能够证明LS水解O-糖基化的Ser/Thr及其N末端氨基酸之间的氨基键,无论氨基酸的类型如何(例如似乎不能抑制脯氨酸水解)(图5)。

[0231] 依那西普因其O-连接的聚糖丰度高而用作模型蛋白,将依那西普用LS处理,此后

用O-糖苷酶进行后续处理以便于质谱分析。将质谱分析产生的m/z值与MS/MS数据相结合,拟合至依那西普。所有鉴定出的肽具有N末端丝氨酸或苏氨酸,这与LS仅切割O-聚糖的N末端一致(图5)。该分析以定向搜索(在参数中定义S/T水解;图5A)和无偏方法(图5B)鉴定了肽。

[0232] LS的水解失活变体特异性结合含O-聚糖的蛋白质

[0233] 由于LS结合O聚糖并特异性水解紧邻聚糖(例如,紧邻Ser/Thr)的氨基酸键的能力,我们假设LS的E_{206A}突变体将缺乏水解活性,但保留结合能力。这样的工具对于以下以及其他将是具有价值的:a)鉴定O-连接的糖蛋白,b)用于去除或用于研究的O-连接的糖肽的亲亲和纯化,以及c)对O-聚糖进行亲和纯化。

[0234] 图6A显示突变体LS不具有任何可检测的水解活性。尽管在唾液酸酶存在下LS能够水解依那西普,但LSmut不能水解依那西普,这证实了在测试条件下遗传改变确实使O-糖基蛋白酶失活。

[0235] 将LS_{E206A}固定在琼脂糖凝胶上,并添加到旋转柱上,以便于处理。重要的是,LS_{E206A}与不同底物的结合与LS的水解活性完全相关(图6B)。LS_{E206A}(标记为LS_{mut})显示出对O连接的糖蛋白的特异性亲和力。通过将LS_{mut}固定在琼脂糖凝胶上,我们能够对IgA进行亲和纯化。然而,由于亲和力强,我们无法洗脱蛋白质。缺乏O-聚糖以及O-糖苷酶处理的IgA的赫赛汀(曲妥珠单抗)没有结合到色谱柱上,但可以在流过液(FT)中检测到。

[0236] 2-3唾液酸键对于去除LS的全部活性是重要的

[0237] 我们最近确定,内切蛋白酶活性依赖于特定的唾液酸键,从而为了完整的作用,必须去除2-3和2-6连接的唾液酸。为了确定特定唾液酸键对LS活性的单独作用,我们将Enbre1与不同的唾液酸酶和LS组合一起孵育了30分钟至20小时。对LS水解来说,去除2-3键似乎是足够的(图7)。

[0238] LS切割促红细胞生成素(EPO)

[0239] 将EPO用PNGaseF、唾液酸酶(Smix,包含Am0707和Am1757)和/或O-糖苷酶处理,并与LS一起孵育。

[0240] 然后通过SDS-PAGE和考马斯蓝染色以及RPLC和质谱来分析所得产物。SDS-PAGE的结果示于图9A,该图显示了当唾液酸被去除和当它们完好时,LS均会切割EPO。此外,LS还消化了已用PNGaseF去除N-聚糖的EPO,从而证实LS活性不受N-聚糖去除的影响。然而,当用O-糖苷酶去除O-聚糖时,LS不切割EPO,这表明LS切割蛋白需要O-聚糖。以10:1、5:1和2:1(底物:酶)的比例观察到了等效的结果(数据未显示)。

[0241] 与PNGaseF、Smix和LS孵育后的样品混合物通过反相液相色谱来进行分离,并通过ESI质谱进行分析以鉴定酶促处理后的反应产物。

[0242] 图9B显示了来自RPLC的UV色谱图。如所预期的,假设EPO仅具有一个建议的O-聚糖位置(参见下文的SEQ ID NO:14的预测位置),则色谱图显示与由LS切割产生的2个片段对应的2个峰。

[0243] 通过MS对这些片段进行了进一步分析(参见图9C和D),并鉴定如下:

[0244] SAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGD(质量=4900.5868Da-与序列C末端至切割位点对应,因此包括仍连接至N末端丝氨酸的O-聚糖);和APPRLICDSRVLERYLLEAK EAEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVD

KAVSGLRSLTTLRLALGAQKEAISPPDAA (质量 = 13714.1199Da, 与序列N末端至切割点对应)。

[0245] 因此, 结合使用PNGaseF、唾液酸酶混合物和LS允许分出和精确鉴定EPO中带有O-聚糖的丝氨酸。这种类型的方法可适用于任何O-糖蛋白, 并允许快速识别O-聚糖连接位置。

[0246] 实施例2

[0247] 介绍

[0248] 实施例1中描述的LSE206A突变体掺入了LS的活性位点的定点突变(abxHEbbHbc至abxHAbbHbc), 从而去除了酶促切割的电子转移能力。如以下进一步解释的, 在进一步的压力测试上, 发现尽管这种改变相对于野生型序列降低了O-糖基蛋白酶的活性, 但是并未完全消除该活性。因此, 本发明人已经开发并表征了在酶促切割中掺入了另外的取代基的另一种突变体。具体而言, 将对辅因子锌离子的取向重要的His残基替代为Ala。所得的双突变体被称为H205A/E206A(abxHEbbHbc变为abxAAbbHbc)。

[0249] 2.1双突变体的产生

[0250] 相对于嗜粘蛋白艾克曼菌的Amuc1119基因, 将使用标准方案(例如如实施例1中的标准方案)进行的定点诱变用于将组氨酸和谷氨酸改变为丙氨酸, 以产生双突变体Amuc1119H205A/E206A(LS_{H205A/E206A})。如实施例1中所述, 将构建体转化到大肠杆菌中, 分出和使用测序进行验证。如实施例1中所述在大肠杆菌中进行表达。表达的蛋白质的序列以SEQ ID NO:21提供。

[0251] 2.2双突变体的表征

[0252] 2.2.1双突变体使LS的活性完全失活

[0253] 如实施例1所示, 鉴于单突变体LSE206A无法在2小时内水解O-糖蛋白, 因此单突变体LSE206A被认为是失活的。然而, 在压力测试中, 发现O-糖基蛋白酶的活性没有完全被消除, 而只是降低了, 因为在更高的酶:O-糖蛋白比例和更长孵育时间下观察到了一些活性。

[0254] LSE206A: 去唾液酸化的O-糖蛋白以1:1(wt:wt)的比例孵育24小时会导致底物的显著水解, 尽管程度不如野生型LS(图10A)。相比之下, 双突变LSH205A/E206A在过夜孵育后, 甚至在酶:O-糖蛋白的比例为15:1(wt:wt)时也没有产生任何水解迹象(图10B), 表明该酶在添加第二种突变时是完全失活的。

[0255] 2.2.2双突变体特异性结合O-糖蛋白

[0256] 为了评估与不同蛋白的结合, 固定的LS_{H205A/E206A} (50μl树脂) (使用与实施例1相同的方案制备) 在PBS中平衡, 此后以0.5mg/mL的浓度添加50μg不同的蛋白样品, 在室温下颠倒旋转孵育2小时。流过液通过离心(200g, 1分钟)来收集, 树脂用350μl PBS洗涤3次。结合的蛋白质通过与50μl 8M尿素孵育5分钟, 相继两次, 然后离心(1000g, 1分钟)来洗脱。所有样品均以相同体积上样。通过SDS-PAGE评估起始/上样材料、流过液和洗脱液。

[0257] 在第一个实验中(参见图11A), 用唾液酸酶混合物(Am0707:Am1757)或唾液酸酶混合物与口腔链球菌内-α-N-乙酰半乳糖苷酶(例如O-糖苷酶)的组合预处理糖基化蛋白或非糖基化蛋白, 然后与树脂一起孵育、洗涤并洗脱。根据制造商的说明, 对样品进行预处理(唾液酸酶混合物+/-O-糖苷酶)。在不存在唾液酸的情况下, 仅具有O-聚糖的蛋白质与树脂结合, 并具有增加的亲和力。如用O-糖苷酶处理后缺乏相互作用所表明, O-聚糖的存在对于任何结合都是必不可少的。

[0258] 在第二个实验中(参见图11B), 将N-糖基化蛋白、O-糖基化蛋白和非糖基化蛋白的

混合物与LS双突变体树脂一起孵育。仅O-糖基化蛋白(TNF α R和ApoE)与基质结合,并用8M尿素洗脱。N-糖基化(阿柏西普、AGP(α -1-酸糖蛋白)、IgG的Fc结构域(IgG Fc))和非糖基化(BSA)未结合至LS双突变体树脂,并在流过液中被发现。因此,当该样品包含N-糖基化蛋白、O-糖基化蛋白和非糖基化蛋白的混合物时,双突变体树脂仅与O糖基化蛋白特异性结合。

[0259] 在第三个实验中(参见图11C),N-糖基化和非糖基化蛋白的混合物与LS双突变体树脂一起孵育。即使在不存在可能与O-糖蛋白竞争的情况下,也没有非特异性结合(均不存在)。在洗脱液中未发现蛋白质。因此,双突变体树脂不与缺乏O-聚糖的蛋白质结合。

[0260] 2.2.3双突变体可以固定在不同浓度的树脂上以增强容量

[0261] 为了研究提高固定的双突变体树脂结合更多O-糖基化蛋白的能力,在固定到树脂上的过程中使用了不同浓度的双突变体(5mg/mL至15mg/mL)。代表性凝胶示于图12A。所示的%是相对于阳性对照的结合水平,并通过凝胶的光密度分析来确定。结果显示在图12B的图表中。当在固定过程中使用较高浓度的双突变体时,观察到容量随较高的O-糖蛋白结合能力而呈剂量依赖性增加。使用15mg/mL的固定的双突变体继续进行进一步的实验。另外,即使在存在1M尿素和1M G HCl的情况下,也维持高度的O-糖蛋白结合,即便1M尿素和1M G HCl显著降低了结合效率。

[0262] 2.2.4双突变体的亲和纯化能力为约3mg糖蛋白/mL树脂

[0263] 为了专门研究双突变体树脂对O-糖蛋白的亲和纯化能力,以及样品浓度对该能力的影响,向树脂中添加了不同量和不同浓度的去唾液酸化依那西普。单个柱子(包含50 μ l双突变体树脂)具有结合约150 μ g O-糖蛋白的容量(即3mg O-糖蛋白/mL树脂)。图13显示了代表性的凝胶。

[0264] 2.2.5O-糖蛋白与双突变体的结合不受离子强度或缓冲液体积/类型的显著影响,并且在宽pH范围内起作用

[0265] 在室温下,在一系列不同条件下,通过颠倒旋转,使样品蛋白与双突变体树脂结合2小时,以测试离子强度、缓冲液体积/类型和pH对树脂的结合能力的影响。在每种情况下,然后将树脂用其各自的结合缓冲液(350 μ l)洗涤三次,然后通过添加8M尿素洗脱(50 μ l,孵育5分钟;重复2次)。然后通过SDS-PAGE分析所有样品。

[0266] 在第一个实验中(参见图14A),为了研究在具有不同离子强度的缓冲液中相互作用的稳定性,样品由去唾液酸化依那西普组成,该样品与双突变体树脂在0至4M NaCl中孵育,以及用相应的NaCl浓度执行所有洗涤步骤。添加NaCl不会显著影响去唾液酸化依那西普的结合。

[0267] 在第二个实验中(参见图14B),样品由一系列不同体积的PBS中的去唾液酸化依那西普组成。洗涤步骤使用PBS。底物体积在100 μ l至300 μ l之间变化不会显著影响效率。

[0268] 在第三个实验中(参见图14C和D),样品由不同pH(pH 4至9)下的不同缓冲液(100mM乙酸钠、50mM磷酸钠和50mM Tris)中的去唾液酸化依那西普和BSA组成。洗涤步骤使用匹配的缓冲液。发现pH 6-8作用最佳,而pH 4完全不起作用,pH 9的效率略低于pH8。不包含任何O-聚糖的BSA在任何结合条件下均不与树脂结合。

[0269] 2.2.6尿素和SDS可以洗脱亲和结合的O-糖蛋白

[0270] 基于双突变体与其O-糖蛋白底物之间的高亲和力,本发明人研究了不基于离子强度从树脂上洗脱结合蛋白的不同方法。尿素呈剂量依赖性洗脱,使用8M尿素时洗脱接近

100% (图15A)。高浓度的SDS (例如5%-10%) 也洗脱了大多数的结合蛋白 (图15B)。然而, 由于许多下游应用对去污剂的存在敏感, 因此使用高水平尿素对于结合蛋白/肽的非酶促释放可能是更具实用性的。

[0271] 2.2.7野生型LS可用于洗脱结合双突变体的O-糖蛋白

[0272] 本发明人推测, 向结合双突变体的蛋白中添加LS可以导致蛋白释放, 因此不必须为洗脱而添加尿素。阿巴西普和依那西普均可在6小时内被LS水解并从双突变体树脂上洗脱下来, 但在24h后洗脱略微更完全 (图16A)。此后添加尿素表明非常少的O-糖蛋白保持附着在亲和基质上, 这表明LS洗脱策略是高效的。

[0273] 对LS洗脱的依那西普也进行质谱分析 (LC/MS和MS/MS)。鉴定出的肽 (图16B.1) 与依那西普的LS消化过程中产生的肽 (图16B.2) 一致。下表显示了该实验的其他MS数据:

[0274]

行	OK	Cmpd.	m/z meas.	Mr calc.	z	$\Delta m/z$ [ppm]	RMS90 [ppm]
1	TRUE	1901	392,18578	782,3545447	2	3,139355727	39,5628229
2	TRUE	1240	460,24964	459,2441464	1	-3,873732048	6,970140396
3	TRUE	2395	593,93513	1778,79071	3	-4,012614897	37,85883109
4	TRUE	1019	598,28693	2986,418959	5	-6,916951725	23,18052633
5	TRUE	1388	642,65448	1924,956428	3	-7,685331295	12,36799202
6	TRUE	2169	672,31437	1342,621401	2	-5,364932116	7,267732308
7	TRUE	2169	672,31437	1342,621401	2	-5,364932116	7,267732308
8	TRUE	1666	695,85597	1389,70747	2	-7,245075789	13,08730439
9	TRUE	1887	745,4213	744,4170254	1	-4,027105599	40,39538537
10	TRUE	2308	846,8813	1691,758682	2	-6,278826744	10,77131727

[0275]

11	TRUE	1636	878,42107	1754,839667	2	-6,875687278	9,154478723
12	TRUE	2298	895,05953	2682,161633	3	-1,814552481	8,770132859
13	TRUE	2048	898,39499	1794,785625	2	-5,675814522	10,69492018
14	TRUE	2319	924,06736	2769,193661	3	-4,837599076	9,79089835
15	TRUE	2299	1072,96596	2143,922906	2	-2,58127791	10,42173435
16	TRUE	1315	1189,55981	2377,120652	2	-6,550738429	9,551870744
17	TRUE	1325	1189,55981	2377,120652	2	-6,550738429	9,851027886
18	TRUE	2377	1219,18407	3654,552682	3	-6,097374291	9,225746116
行	Int.	分数					范围
1	100030	20.3 (D.metascore:20.3,D.btScore:4.0,D.fragCov:50.0,D.intCov:8.2)					457 - 459
2	29654	23.0 (D.metascore:23.0,D.btScore:0.0,D.fragCov:12.5,D.intCov:42.3)					182 - 185
3	30476	31.2 (D.metascore:31.2,D.btScore:127.0,D.fragCov:40.0,D.intCov:24.3)					216 - 225
4	219206	26.5 (D.metascore:26.5,D.btScore:265.0,D.fragCov:33.3,D.intCov:21.1)					184 - 204
5	655550	23.9 (D.metascore:23.9,D.btScore:136.0,D.fragCov:46.7,D.intCov:12.3)					184 - 198
6	374698	36.2 (D.metascore:36.2,D.btScore:61.0,D.fragCov:33.3,D.intCov:39.2)					217 - 225
7	374698	36.2 (D.metascore:36.2,D.btScore:61.0,D.fragCov:33.3,D.intCov:39.2)					217 - 225
8	39488	26.9 (D.metascore:26.9,D.btScore:12.0,D.fragCov:35.7,D.intCov:20.3)					186 - 199
9	18900	66.8 (D.metascore:66.8,D.btScore:191.0,D.fragCov:92.9,D.intCov:48.1)					1 - 7
10	108010	22.4 (D.metascore:22.4,D.btScore:13.0,D.fragCov:22.2,D.intCov:22.5)					217 - 225
11	275854	29.6 (D.metascore:29.6,D.btScore:260.0,D.fragCov:39.3,D.intCov:22.3)					186 - 199
12	30586	24.2 (D.metascore:24.2,D.btScore:509.0,D.fragCov:28.1,D.intCov:20.8)					216 - 231
13	48978	35.0 (D.metascore:35.0,D.btScore:125.0,D.fragCov:35.0,D.intCov:35.0)					216 - 225
14	38966	23.7 (D.metascore:23.7,D.btScore:253.0,D.fragCov:23.5,D.intCov:23.8)					216 - 232
15	29176	20.1 (D.metascore:20.1,D.btScore:29.0,D.fragCov:25.0,D.intCov:16.2)					216 - 225
16	2495538	23.1 (D.metascore:23.1,D.btScore:535.0,D.fragCov:59.4,D.intCov:9.0)					184 - 199
17	182236	22.9 (D.metascore:22.9,D.btScore:4096.0,D.fragCov:43.8,D.intCov:12.0)					184 - 199
18	360318	21.1 (D.metascore:21.1,D.btScore:33.0,D.fragCov:22.2,D.intCov:20.0)					208 - 225
行	#Cmps.	P	序列	修饰	蛋白质		
1	1	0	Y.TQK.S	乙酰基: 1; Hex(1)HexNAc(1): 1	依那西普		
2	1	1	T.SPTR.S		依那西普		
3	5	2	P.STSFLLPMG.P.S	Hex(1)HexNAc(1): 1, 2	依那西普		
4	13	4	P.TRSMAPGAVHLPQPVSTRSQH.T	Hex(1)HexNAc(1): 16, 17	依那西普		
5	6	1	P.TRSMAPGAVHLPQPV.S	Hex(1)HexNAc(1): 3	依那西普		
6	4	1	S.TSFLLPMG.P.S	氧化: 7; Hex(1)HexNAc(1): 2	依那西普		
7	4	1	S.TSFLLPMG.P.S	氧化: 7; Hex(1)HexNAc(1): 1	依那西普		
8	2	1	R.SMAPGAVHLPQPV.S.T		依那西普		
9	5	0	-.LPAQVAF.T		依那西普		
10	2	1	S.TSFLLPMG.P.S	Hex(1)HexNAc(1): 1, 2	依那西普		
11	5	1	R.SMAPGAVHLPQPV.S.T	Hex(1)HexNAc(1): 14	依那西普		
12	2	3	P.STSFLLPMGSPPAEG.S	Hex(1)HexNAc(1): 1, 2, 3	依那西普		
13	10	2	P.STSFLLPMG.P.S	Oxidation: 8; Hex(1)HexNAc(1): 1, 2	依那西普		
14	3	4	P.STSFLLPMGSPPAEG.S.T	Hex(1)HexNAc(1): 1, 2, 3	依那西普		
15	5	2	P.STSFLLPMG.P.S	Hex(1)HexNAc(1): 1, 2, 3	依那西普		
16	19	2	P.TRSMAPGAVHLPQPV.S.T	Hex(1)HexNAc(1): 3, 16	依那西普		
17	19	2	P.TRSMAPGAVHLPQPV.S.T	Hex(1)HexNAc(1): 1, 3	依那西普		
18	8	5	P.TPEPSTAPSTSFLLPMG.P.S	Hex(1)HexNAc(1): 5, 6, 9, 10, 11	依那西普		

[0276]

[0277]

2.2.8双突变体可用于对来自复杂样品中的O-糖蛋白进行亲和纯化

[0278] 作为对该系统不仅可以在简化的系统中而且可以在复杂的培养基中充当O-糖基化蛋白的一般亲和基质的观点的证明,本发明人研究了双突变体从人血清中纯化O-糖蛋白的能力。人血清主要由非糖基化(BSA)和N-糖基化(IgG)蛋白质组成,而总血清蛋白质组中只有一小部分是O-糖基化的。

[0279] 向50 μ l固定的双突变体树脂柱施加20 μ l唾液酸酶处理的血清(约1.2mg蛋白质),使得去除几乎所有的非糖基化蛋白和N-糖基化蛋白,而洗脱几种选定的蛋白质(图17A)。通过添加更高量的经过或未经过唾液酸酶和O-糖苷酶预处理的血清(例如2.5mg蛋白),证明了相互作用取决于O-聚糖以及末端唾液酸的去除(图17B)。此外,得出结论,与未经唾液酸酶处理的样品相比,唾液酸酶的预处理显著增加了结合的O-糖蛋白的量。添加50U唾液酸酶混合物(Am0707:Am1757)足以改善亲和纯化的O-糖蛋白的量(图17C)。

[0280] 通过质谱分析,可以将绝大多数亲和纯化的血清蛋白标注为O-糖基化蛋白(参见图18A,以及下表中的粗体斜体字的名称)。相对于非O-糖蛋白肽,已鉴定的O-糖蛋白肽的数量可能会受到洗涤步骤中不同严格性的影响,无论是在鉴定出的肽的总数方面(图18B),还是在O-糖蛋白肽对比非O糖蛋白肽的比例上(图18C)。因此清楚的是,亲和树脂在O-糖蛋白的特异性和选择性亲和纯化和富集的能力方面是高效的。下表显示了该实验的其他MS数据。

[0281]

行	OK	蛋白质	MW [kDa]
1	TRUE	<i>Ig α-1 链 C ϵ</i> OS=智人 GN=IGHA1 PE=1 SV=2	37,6
2	TRUE	<i>载脂蛋白 B-100</i> OS=智人 GN=APOB PE=1 SV=2	515,3
3	TRUE	<i>激肽原-1</i> OS=智人 GN=KNG1 PE=1 SV=2	71,9
4	TRUE	<i>补体 C4-B</i> OS=智人 GN=C4B PE=1 SV=2	192,6
5	TRUE	<i>补体 C4-A</i> OS=智人 GN=C4A PE=1 SV=2	192,7
6	TRUE	补体 C3 OS=智人 GN=C3 PE=1 SV=2	187
7	TRUE	<i>内-α-胰蛋白酶抑制剂重链 H4</i> OS=智人 GN=ITIH4 PE=1 SV=4	103,3
8	TRUE	<i>内-α-胰蛋白酶抑制剂重链 H2</i> OS=智人 GN=ITIH2 PE=1 SV=2	106,4
9	TRUE	<i>血纤维蛋白溶酶原</i> OS=智人 GN=PLG PE=1 SV=2	90,5

[0282]

10	TRUE	内- α -胰蛋白酶抑制剂重链 <i>IH1</i> OS=智人 GN=ITIH1 PE=1 SV=3			101,3
11	TRUE	血清白蛋白 OS=智人 GN=ALB PE=1 SV=2			69,3
12	TRUE	α -2-巨球蛋白 OS=智人 GN=A2M PE=1 SV=3			163,2
13	TRUE	纤连蛋白 OS=智人 GN=FN1 PE=1 SV=4			262,5
14	TRUE	补体 C1s 亚组分 OS=智人 GN=C1S PE=1 SV=1			76,6
15	TRUE	N-乙酰胞壁酰基-L-丙氨酸酰胺酶 OS=智人 GN=PGLYRP2 PE=1 SV=1			62,2
16	TRUE	血浆蛋白酶 <i>C1 抑制剂</i> OS=智人 GN=SERPING1 PE=1 SV=2			55,1
17	TRUE	凝血酶原 OS=智人 GN=F2 PE=1 SV=2			70
18	TRUE	C4b-结合蛋白 α 链 OS=智人 GN=C4BPA PE=1 SV=2			67
19	TRUE	凝血因子 <i>XII</i> OS=智人 GN=F12 PE=1 SV=3			67,7
20	TRUE	补体 C1r 亚组分 OS=智人 GN=C1R PE=1 SV=2			80,1
21	TRUE	<i>Ig δ 链 C 区</i> OS=智人 GN=IGHD PE=1 SV=2			42,2
22	TRUE	载脂蛋白 <i>F</i> OS=智人 GN=APOF PE=1 SV=2			35,4
23	TRUE	补体组分 <i>C7</i> OS=智人 GN=C7 PE=1 SV=2			93,5
24	TRUE	Ig μ 链 C 区 OS=智人 GN=IGHM PE=1 SV=3			49,3
25	TRUE	内- α -胰蛋白酶抑制剂重链 <i>H3</i> OS=智人 GN=ITIH3 PE=1 SV=2			99,8
26	TRUE	组氨酸富集的糖蛋白 OS=智人 GN=HRG PE=1 SV=1			59,5
27	TRUE	补体组分 <i>C9</i> OS=智人 GN=C9 PE=1 SV=2			63,1
28	TRUE	α -2-HS-糖蛋白 OS=智人 GN=AHSG PE=1 SV=1			39,3
29	TRUE	蛋白 <i>AMBP</i> OS=智人 GN=AMBP PE=1 SV=1			39
30	TRUE	<i>Ig γ-1 链 C 区</i> OS=智人 GN=IGHG1 PE=1 SV=1			36,1
31	TRUE	α -1-抗胰蛋白酶 OS=智人 GN=SERPINA1 PE=1 SV=3			46,7
32	TRUE	血浆激肽释放酶 OS=智人 GN=KLKB1 PE=1 SV=1			71,3
33	TRUE	玻连蛋白 OS=智人 GN=VTN PE=1 SV=1			54,3
34	TRUE	载脂蛋白 <i>A-I</i> OS=智人 GN=APOA1 PE=1 SV=1			30,8
35	TRUE	维生素 K-依赖性蛋白 <i>S</i> OS=智人 GN=PROS1 PE=1 SV=1			75,1
36	TRUE	血清铁传递蛋白 OS=智人 GN=TF PE=1 SV=3			77
37	TRUE	载脂蛋白 <i>E</i> OS=智人 GN=APOE PE=1 SV=1			36,1
38	TRUE	血管假性血友病因子 OS=智人 GN=VWF PE=1 SV=4			309,1
39	TRUE	蛋白聚糖 <i>4</i> OS=智人 GN=PRG4 PE=1 SV=2			151
行	pI	#肽	SC [%]	分数	RMS90 [ppm]
1	6,1	149	75,1	1239.4 (M.期望:0.0,M.分数:1239.4,M.siglimit:38.0)	5,46
2	6,6	137	41,2	7043.5 (M.期望:0.0,M.分数:7043.5,M.siglimit:38.0)	6,02
3	6,3	86	45,7	1638.5 (M.期望:0.0,M.分数:1638.5,M.siglimit:38.0)	5,49
4	6,9	78	63,3	4951.3 (M.期望:0.0,M.分数:4951.3,M.siglimit:38.0)	6,46
5	6,7	76	61,3	4468.7 (M.期望:0.0,M.分数:4468.7,M.siglimit:38.0)	6,12
6	6	75	64,1	4596.8 (M.期望:0.0,M.分数:4596.8,M.siglimit:38.0)	6,13
7	6,5	52	56,9	2389.7 (M.期望:0.0,M.分数:2389.7,M.siglimit:38.0)	6,29
8	6,4	49	50,4	2175.2 (M.期望:0.0,M.分数:2175.2,M.siglimit:38.0)	7,02

[0283]

9	7	45	67,3	2752.5 (M.期望:0.0,M.分数:2752.5,M.siglimit:38.0)	5,54
10	6,3	39	43,8	1838.3 (M.期望:0.0,M.分数:1838.3,M.siglimit:38.0)	5,16
11	5,9	37	63,7	2350.1 (M.期望:0.0,M.分数:2350.1,M.siglimit:38.0)	5,59
12	6	36	38,1	1970.3 (M.期望:0.0,M.分数:1970.3,M.siglimit:38.0)	5,67
13	5,5	33	25,3	1632.8 (M.期望:0.0,M.分数:1632.8,M.siglimit:38.0)	5,58
14	4,8	23	46,1	1454.6 (M.期望:0.0,M.分数:1454.6,M.siglimit:38.0)	6,01
15	7,3	22	56,4	1288.8 (M.期望:0.0,M.分数:1288.8,M.siglimit:38.0)	5,72
16	6,1	22	46,2	1166.8 (M.期望:0.0,M.分数:1166.8,M.siglimit:38.0)	6,15
17	5,6	20	43,4	1158.1 (M.期望:0.0,M.分数:1158.1,M.siglimit:38.0)	6,13
18	7,2	20	51,4	1111.9 (M.期望:0.0,M.分数:1111.9,M.siglimit:38.0)	5,13
19	8	20	39,7	691.9 (M.期望:0.0,M.分数:691.9,M.siglimit:38.0)	5,03
20	5,8	19	39,6	1097.7 (M.期望:0.0,M.分数:1097.7,M.siglimit:38.0)	6,07
21	8,1	19	46,9	687.5 (M.期望:0.0,M.分数:687.5,M.siglimit:38.0)	4,87
22	5,4	19	43,3	344.0 (M.期望:0.0,M.分数:344.0,M.siglimit:38.0)	7,71
23	6,1	18	38,3	1100.7 (M.期望:0.0,M.分数:1100.7,M.siglimit:38.0)	5,73
24	6,3	17	50,9	1159.0 (M.期望:0.0,M.分数:1159.0,M.siglimit:38.0)	5,25
25	5,5	16	28,3	747.2 (M.期望:0.0,M.分数:747.2,M.siglimit:38.0)	5,8
26	7,1	15	35,4	964.5 (M.期望:0.0,M.分数:964.5,M.siglimit:38.0)	5,33
27	5,4	15	34,3	876.5 (M.期望:0.0,M.分数:876.5,M.siglimit:38.0)	5,86
28	5,4	14	46	941.9 (M.期望:0.0,M.分数:941.9,M.siglimit:38.0)	5,78
29	5,9	14	55,4	886.3 (M.期望:0.0,M.分数:886.3,M.siglimit:38.0)	6,06
30	8,5	14	60,3	825.9 (M.期望:0.0,M.分数:825.9,M.siglimit:38.0)	6,04
31	5,4	14	39	740.9 (M.期望:0.0,M.分数:740.9,M.siglimit:38.0)	6,02
32	8,6	14	26	647.7 (M.期望:0.0,M.分数:647.7,M.siglimit:38.0)	5,62
33	5,6	13	37	752.2 (M.期望:0.0,M.分数:752.2,M.siglimit:38.0)	5,1
34	5,6	13	46,8	723.0 (M.期望:0.0,M.分数:723.0,M.siglimit:38.0)	5,57
35	5,5	13	27,7	676.0 (M.期望:0.0,M.分数:676.0,M.siglimit:38.0)	6,18
36	6,8	13	22,9	544.9 (M.期望:0.0,M.分数:544.9,M.siglimit:38.0)	6,02
37	5,6	13	45,1	535.3 (M.期望:0.0,M.分数:535.3,M.siglimit:38.0)	5,73
38	5,3	13	6,3	505.5 (M.期望:0.0,M.分数:505.5,M.siglimit:38.0)	5,52
39	9,5	13	5,6	214.2 (M.期望:0.0,M.分数:214.2,M.siglimit:38.0)	4,7

[0284] 2.2.9固定的双突变体还与较短的O-糖肽结合

[0285] 为了证明LS双突变体对O-糖肽的特异性,进行了一系列实验。在第一个实验中,将O-糖基化肽(糖基果蝇肽(GD)=在苏氨酸上具有核心10-聚糖的GKPRPYSPRPTSHPRPIRV(SEQ ID NO:47))和几种非糖基化肽(H2686、H4062 H8390和胰岛素氧化β链(IOB))的混合物与LS双突变体树脂一起孵育。(H2686=YIYGSKF(SEQ ID NO:48),H4062=KKLVFFA(SEQ ID NO:49),H8390=FLPLILGKLVKGLL(SEQ ID NO:50))。

[0286] 在室温下通过颠倒旋转使肽混合物与50μl固定的双突变体树脂结合2小时。用结合缓冲液(300μl)洗涤树脂五次,然后添加8M尿素洗脱。用LC/MS分析加样液、流过液和洗脱液中的肽。在RP-LC C18柱上(Advance BioPeptide Map 2.1x100 2.7μm,来自Agilent)进行分离,并用ESI-Q-TOF Bruker Impact II检测。结果示于图19A。糖基果蝇肽是混合物中唯一包含O-GalNAcGal的肽,主要存在于洗脱级分中,而非糖基化肽则存在于流过级分中。

[0287] 在第二个实验中,研究了LS双突变体能否从胰蛋白酶蛋白质消化物(例如不同类型的肽混合物)中富集O-糖基化肽。

[0288] 选择IgA作为消化的靶标。根据IgA中的胰蛋白酶位点和已报道的O-糖基化位点,胰蛋白酶消化应仅产生与IgA第89至126位对应的单个O-糖基化肽(请参见图19B中的示意

图)。

[0289] 为了产生胰蛋白酶消化物,将IgA与尿素混合至6M,与DTT混合至5mM,然后在37℃下孵育1小时。将IAM添加至15mM,然后在黑暗中在室温下孵育30分钟。然后将样品在Zebaspin 7000K色谱柱上缓冲液交换至pH 8.0的50mM Tris中。然后以1:20的比例添加胰蛋白酶,接着在37℃下孵育过夜。添加胰蛋白酶抑制剂1mg/mg,然后在室温下孵育20分钟。将唾液酸酶混合物和NaCl添加到所得的胰蛋白酶消化物中。在室温下通过颠倒旋转使混合物与树脂结合2小时。将树脂用PBS缓冲液(300μl)洗涤10次,然后通过添加8M尿素(50μl,2分钟,重复2次)来洗脱。

[0290] 使用RP-LC MSMS在C18色谱柱(Advance BioPeptide Plus 2.1x150mm 2.7μm,来自Agilent Technologies)上,在0.1%FA的MQ溶液:0.1%FA的95%CAN梯度溶液中,在45℃下,以0.2ml/分钟的流速分离和分析加样液、流过液和洗脱液中的肽。在ESI-Q-TOF Bruker Impact II仪器上进行检测。结果示于图19C。O-糖基化肽89-126在洗脱液中显著富集,鉴定出完整的特异性O-糖肽89-126。

[0291] 2.2.10固定的双突变体优于其他O-糖蛋白结合基质

[0292] 本发明人评估了双突变体对比其他可商购的O-糖蛋白结合基质(特别是凝集素花生凝集素(PNA)和野豌豆凝集素(VVA))对O-糖蛋白的亲合纯化能力。使用依那西普和去唾液酸化依那西普用作模型底物。

[0293] 向50μl体积的在各自的缓冲液(总共100μl)中平衡的不同的固定的凝集素或LS双突变体树脂添加50μg在PBS(PNA和LS双突变体)或凝集素结合缓冲液(VVA)中的底物。(凝集素结合缓冲液是20mM Tris-HCl pH 7.4,150mM NaCl,MgCl₂,CaCl₂,ZnCl₂和MnCl₂各1mM)。在室温下,使底物与树脂通过颠倒混合相互作用2小时。未结合的蛋白质分别用PBS或凝集素结合缓冲液洗掉(100g,1分钟;3x)。通过离心(1000g,1分钟)干燥树脂。根据生产商通过添加8M尿素(对于PNA和LS双突变体树脂)或VVA洗脱缓冲液(针对VVA树脂)洗脱结合蛋白(50μl,5分钟处理然后1000g离心1分钟;2x),将流过液(FT)和洗脱液(E)在SDS-PAGE上分析。将1.5μg底物添加到每种凝胶(例如3μl)中作为阳性对照,并且进行光密度分析,从而相对于假设为100%效率的1.5μg负载底物,评估树脂的效率。依那西普和去唾液酸化依那西普(Etanercept S)的代表性凝胶示于图20A。密度分析的结果示于图20B。LSH205A/E206A双突变体在纯化去唾液酸化底物的效率上至少与性能最佳的商业凝集素一样好。

[0294] 序列

[0295] SEQ ID NO:1

[0296] EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLD
IKSPGRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGMG
RNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHELGHGLNLPNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP
ASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSFS
RRLGKKSGKFSMKIDKKELEGLNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS

[0297] SEQ ID NO:2

[0298] MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGL
DIKSPGRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGMG
GRNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHELGHGLNLPNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP

PASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSF
SRRLGKKS~~GKFSMKIDKKELEGLNNEFRISLMFILANGLHMQKHFTFHW~~DALQDYRDGSKSGSGHHHHHH

[0299] SEQ ID NO:3

[0300] ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAAT
ATCGTCTACTTCTGGGTTCTGATACCGAACCGGTTCCGGACTACGAGCGCGTCTGAGCGAGCTGCTGTTGTATCT
GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCACCGG
GTCGTGTGAACATTATCGAGTACAAAGCGAAGAACCCGGCAGCGCATTACCCGTATGAGAATGGTGGCGGCTGGAAA
GCTGCACAAGAACTGGACGAATTTTTCAAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCTGATCATCATGCC
TACCTGGAATGATGAGAAAAATGGTCTGACAATCCGGGTGGCGTTCGGTTCTATGGTATGGGTCGTAATTGTTTTG
CGTTGGACTACCCGGCGTTTGATATCAAGCACCTGGGTCAGAAAACGCGTGAGGGTCGTCTGCTGACGAAATGGTAC
GGTGGCATGGCGCACGAAGTGGGCCACGGCTGAATCTGCCGCACAATCACCAGACCGCGAGCGATGGCAAGAAATA
TGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGTACCAGCCCGACGTTCCCTGACCCCGCGAGCTGTGCGC
TGCTGGATGCCTGCGAAGTGTTTCAGCGTTACCCCGAGCCAACAGTTTTATGAGGGTAAGCCAGAAGTCGAGGTTGGT
GATGTTGCAATTTCTTTCAAGGGTGATCAAACTCTGGTCAGCGGTAACATAAGAGCCCGCAAACCGTGAAAGCTCT
GAACGTTTACATTTCAGGATCCGCCGTACGCCGTGAACCAAGACTACGATGCAGTGAGCTTTAGCCGTCGTCTGGGCA
AAAAGTCCGGTAAGTTTAGCATGAAGATTGACAAAAAAGAACTGGAAGGCCTGAATAACAACGAATTCGTATTTCC
TTGATGTTTATTCTGGCAAACGGCTTACACATGCAGAAGCACTTTACGTTTACTGGGATGCGCTGCAAGACTACCG
TGACGGTAGCAAATCTGGTTCGGGTCATCATCACCACCATCACTGA

[0301] SEQ ID NO:4

[0302] MLKRLLSAFFSLFFLGAASGTSFAEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSEL
LLYLQQFYGKEMQRHGYGARSFGLDIKSPGRVNIIEYKAKNPAAHYPYENGGGWKAAQELDEFFKAHPDRKKSQHT
LIIMPTWNDEKNGPDNPGGVPFYGMGRNCFALDYPAFDIKHLGQKTREGRLTKWYGGMAHELGHGLNLPNHQTA
SDGKKYGTALMGSGNYTFGTSPFTLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKS
PQTVKALNVYIQDPPYAVNQDYDAVSFSRRLGKKS~~GKFSMKIDKKELEGLNNEFRISLMFILANGLHMQKHFTFHW~~
WDALQDYRDGSKS (信号序列加有下划线)

[0303] SEQ ID NO:5

[0304] EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLYLQQFYGKEMQRHGYGARSFGLD
IKSPGRVNIIEYKAKNPAAHYPYENGGGWKAAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGM
RNCFALDYPAFDIKHLGQKTREGRLTKWYGGMAHALGHGLNLPNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP
ASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSFS
RRLGKKS~~GKFSMKIDKKELEGLNNEFRISLMFILANGLHMQKHFTFHW~~DALQDYRDGSKS

[0305] SEQ ID NO:6

[0306] MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLYLQQFYGKEMQRHGYGARSFGL
DIKSPGRVNIIEYKAKNPAAHYPYENGGGWKAAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGM
GRNCFALDYPAFDIKHLGQKTREGRLTKWYGGMAHALGHGLNLPNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP
PASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSF
SRRLGKKS~~GKFSMKIDKKELEGLNNEFRISLMFILANGLHMQKHFTFHW~~DALQDYRDGSKSGSGHHHHHH

[0307] SEQ ID NO:7

[0308] ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAAT

ATCGTCTACTTCTGGGTTCTGATACCGAACCGGTTCCGGACTACGAGCGCCGTCTGAGCGAGCTGCTGTTGTATCT
GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCACCGG
GTCGTGTGAACATTATCGAGTACAAAGCGAAGAACCCGGCAGCGCATTACCCGTATGAGAATGGTGGCGGCTGGAAA
GCTGCACAAGAACTGGACGAATTTTTCAAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCTGATCATCATGCC
TACCTGGAATGATGAGAAAAATGGTCCTGACAATCCGGGTGGCGTTCCGTTCTATGGTATGGGTCGTAATTGTTTTG
CGTTGGACTACCCGGCGTTTGATATCAAGCACCTGGGTCAGAAAACGCGTGAGGGTCGTCTGCTGACGAAATGGTAC
GGTGGCATGGCGCACGCGTGGGCCACGGCTGAATCTGCCGCACAATCACCAGACCGCGAGCGATGGCAAGAAATA
TGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGTACCAGCCCGACGTTCCCTGACCCCGCGAGCTGTGCGC
TGCTGGATGCCTGCGAAGTGTTTCAGCGTTACCCCGAGCCAACAGTTTTATGAGGGTAAGCCAGAAGTCGAGGTTGGT
GATGTTGCAATTTCTTTCAAGGGTGATCAAACTCTGGTCAGCGGTAACATAAGAGCCCGCAAACCGTGAAAGCTCT
GAACGTTTACATTTCAGGATCCGCCGTACGCCGTGAACCAAGACTACGATGCAGTGAGCTTTAGCCGTCGTCTGGGCA
AAAAGTCCGGTAAGTTTAGCATGAAGATTGACAAAAAAGAACTGGAAGGCCTGAATAACAACGAATTCCGTATTTCC
TTGATGTTTATTCTGGCAAACGGCTTACACATGCAGAAGCACTTTACGTTTACTGGGATGCGCTGCAAGACTACCG
TGACGGTAGCAAATCTGGTTCGGGTCATCATCACCACCATCACTGA

[0309] SEQ ID NO:8

[0310] GMAHELGHGL (金属蛋白酶基序)

[0311] SEQ ID NO:9

[0312] MKNLLFALLTGSFCCCYAQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDN
GKTWEMKAAPAGYPPNYGGIPKESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKR
KNLKKLGGIMRTPVFNKGRRVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEAT
VLEMKDGTWALARTSQDQAWQAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENATAGNG
TWEDVFTNRDSHHIAMSGDEGKTWYGFREIILDEHRNHPGYATLDGPEDRGKHQSEMVQLDKNRILISLGQHKHNR
RLVIVDRRWVGAKTATQTGKDLSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEK
SNVDYRNGGATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLG
MKKVPFTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANSPNGASYIHFISTGSQPDAGILLDTVNARVK (信号
序列加有下划线)

[0313] SEQ ID NO:10

[0314] QQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDNGKTWEMKAAPAGYPPNYGG
IPKESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFNKG
RVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEATVLEMKDGTWALARTSQDQAW
QAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSHHIAMSGDEGK
TWYGFREIILDEHRNHPGYATLDGPEDRGKHQSEMVQLDKNRILISLGQHKHNRRLVIVDRRWVGAKTATQTGKDL
DSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNGGATWNFPNGTTGLVKF
RFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLGKKVPFTPGAWHEISLLWQGGQAVV
SLDGKKAGTLKMANSPNGASYIHFISTGSQPDAGILLDTVNARVK

[0315] SEQ ID NO:11

[0316] MQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDNGKTWEMKAAPAGYPPNYG
GIPKESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFNKG
RRVIVPFHNMGGGTFKHISDDGGLTWHVSRNGVTSRHEARPPHQGVRFNNAVEATVLEMKDGTWALARTSQDQA

WQAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSHHIAMSDEG
KTYWGFREIILDEHRNHPGYATLDGPEDRGKHQSEMVQLDKNRILISLGQHKHNRRLVIVDRRWVGAKTTRATQTGKD
LDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNGGATWNFPNGTTGLVK
FRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLGMKKVPFTPGAWHEISLLWQGGQAV
VSLDGKKAGTLKMANKSPNGASYIHFISTGSQPDAGILLDTVNARVKSGGLEHHHHHH

[0317] SEQ ID NO:12

[0318] MTWLLCGRGKWNKVKRMNSVFKCLMSAVCAVALPAFGQEEKTGFPDRAVTVFSAGEGNPYASIRIP
ALLSIGKGQLLAFAEGRYKNTDQGENDIIMSVSKNGGKTWSRPRAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYP
AGVKERQPNIPDGWDDEKCIRNFMIQSRNGGSSWTKPQEITKTTKRPSGVDIMASGPNAGTQLKSGAHKGRLVIPM
NEGPFQKVVISCIYSDDGGKSWKLGQPTANMKGMVNETSIAETDNGGVMMVARHWGAGNCRRIAWSQDGGGETWGQV
EDAPELFCDDSTQNSLMTYSLSDQPAYGGKSRIILFSGPSAGRIKQVAMSVDNGKTWPVKLLGEGGFAYSSLAMV
EPGIVGVLYEENQEHIKKLKFPITMEWLTGEGDTGLAPGKKAPVLK (信号序列加有下划线)

[0319] SEQ ID NO:13

[0320] QEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAEGRYKNTDQGENDIIMSVSKNGGKT
WSRPRAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYPAGVKERQPNIPDGWDDEKCIRNFMIQSRNGGSSWTKPQEI
TKTTKRPSGVDIMASGPNAGTQLKSGAHKGRLVIPMNEGPFQKVVISCIYSDDGGKSWKLGQPTANMKGMVNETSIA
ETDNGGVMMVARHWGAGNCRRIAWSQDGGGETWGQVEDAPELFCDDSTQNSLMTYSLSDQPAYGGKSRIILFSGPSAGR
IKQVAMSVDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEENQEHIKKLKFPITMEWLTGEGDTGLAPGKKA
PVLK

[0321] SEQ ID NO:14

[0322] MQEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAEGRYKNTDQGENDIIMSVSKNGGK
TWSRPRAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYPAGVKERQPNIPDGWDDEKCIRNFMIQSRNGGSSWTKPQEI
TKTTKRPSGVDIMASGPNAGTQLKSGAHKGRLVIPMNEGPFQKVVISCIYSDDGGKSWKLGQPTANMKGMVNETSIA
AETDNGGVMMVARHWGAGNCRRIAWSQDGGGETWGQVEDAPELFCDDSTQNSLMTYSLSDQPAYGGKSRIILFSGPSAGR
RIKQVAMSVDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEENQEHIKKLKFPITMEWLTGEGDTGLAPGKKA
APVLKSGLEHHHHHH

[0323] SEQ ID NO:15-来自口腔链球菌的O-糖苷键

[0324] MDKRFEFKCKFSIRKFTLGVASVMIGATFFAASPVADQARVGSTDNLPSSELADLDKKASDEGHDFDK
EAAQNPQSAETTEGPQTEEELLAQEKEKSEKPSNLPKELEDKLEKAEDNGREVDKQDLAQDTGKLVDPEDVAKTTNG
ELNYGATVGIKTPSGEGSGIVVAKDLVLTVSHNFIKDSQEGNIRKVVNDNDQGDGDIYSISYPLPDVKFSKKDIIHW
DREGYLKGFKNLALVRLRTVLENTPEVTKKPVVKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEGVQGIQYQG
GKPGASGGGIFDTEGKLVGVHQNQGVVGKRSGLFSPAQLKWIQDHMQGISSVKPADLEEKEKPAEEKPKEDKPAAA
KPETPKAVTPEWQTVANKEQQGTVTIREEKGVRYNQLSSTAQNNDNGKPALEKQGLTVDANGNATVDLTFKDDSEK
GKSRLFVFLKFKDTKNNVFGYDQGGWFWEYKTPGNSTWYKGNRVAPEPGSVNRLSITLKSQGLNASNNDVNLF
TVTLPGAVNENLKNEKKILLKAGTYSNDRTVSVKTDNQEGVKADDTAQAQKETGPAVDDSKVTDYDTIQSKVLKAVID
QAFPRVKEYTLNGHTLPGQVQQFNQVFINNHRITPEVTYKKINETTAEYLMKLRRDAHLINAEMTVRLQVVDNQLHF
DVTKIVNHNQVTPGQKIDDERKLLSTISFLGNALVSVDSDQAGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYM
YGFVSTDKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSWEQWEKAYKGIVFPEYTKELPSAKVVIT
EDANADNKVDWQDGAIAYSIMNPNQGWKEVKDITAYRIAMNFGSQAQNPFLMTLDGIKKINLHTDGLGQGVLLKGY

GSEGHDSGHLNYADIGKRIGGVEDFKTLIEKAKKYGAHLGIHVNASETYESPKNENILRKNPDGSYSYGNWLDQ
GINIDAAYDLAHGRLARWEDLKKKLGEGLDFIYVDVWNGGQSGDNGAWATHVLAKEINKQGWRFAIEWGHGGEYDST
FQHWAADLTYYGYTNKGINSATIRFIRNHQKDSWVGDIYSYGGGAANYPLLGGYSMKDFEGWQGRSDYNGYVTNLFAH
DVMTKYFQHFTVSKWENGTPVTMTDNGSTYKWTPEMKVELVDAAGNKVVVTRKSNDVNSPQYRERTVTNLGRVIQDG
SAYLTPWNWDANGKKLPTEKEKMYFNTQAGATTWTLPSDWANSKVLYKLTDDQKTEEQELTVDGKITLDDLQ
PYVLYRSKQTNPEMSWSEGMHIYDQGFNSGTLKHWITISGDASKAEIVKSQGANEMLRISQGNKSKVSLTQKLTGLKPN
TKYAVYVGVDNRNNAKASITVNTGEKEVTYTNKSLALNYIKAYAHNNRRENATVDDTSYFQNMAYFDTTGSVDVSNV
TLTSLREAGDEATYFDEIRTFENNSSMYGDKHDTGQGTFFKQDFENVAQGFVVGVEGVEDNRTHLSEKHDPYTQ
RGWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTIPTQNRFEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRRGTQ
ASNLEMHELPNTWTDSKKAKKVTFVLVTGAETGDTWVGIIYSTGNASNTRGDAGGNANFRGYNDFMMDNLQIEEITLTG
KMLTENALKNYLPTVAMTNYTKESMDALKEAVFNLSQADDDISVEEARAEIAKIEALKNALVQKKTALVAEDFESLD
APAQPGEGLNFAFDGNVSSLWHTSWNGGDVGKPMVLKEPTEITGLRYVPRASDSNGNLRDVKLVVTDESGKEHTF
NVTDWPNNNPKDIDFGKTIKAKKIVLTGTYGDKYQSAELIFTRPQVAETPLDLGYEALAKAQKLTDKD
NQEEVASVQASMKYATDNHLLTERMVAYFADYLNQLKDSATKPDAPTSSKGEEQPPVLDVPEFKGGVNATEAAVHEV
PEFKGGVNAVQALVHELPEYKGGANAVLAAANEVPEYKGGANAVEALVNEKPAYTGVLATAGDQAAPTVEKPEYPLT
PSPVADTKTPGAKDEEKLPAATGEHSSEVALFLASVSIALSAVLATKRKEEGSGLEHHHHHH

[0325] SEQ ID NO:16-E206A_正向引物

[0326] ATGGCGCACGC GCTGGGCCACG

[0327] SEQ ID NO:17-E206A_反向引物

[0328] GCCACCGTAC CATTTGCTC

[0329] SEQ ID NO:18-EPO

[0330] APPRLICDSRVLERYLLEAKEAEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAVEVWQGLAL
LSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYS
NFLRGKCLKYTGEACRTGDR

[0331] (注意:预测的带有丝氨酸的O-聚糖加有下划线;C末端精氨酸在表达过程中通常是截短的)

[0332] SEQ ID NO:20

[0333] EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLD
IKSPGRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGMG
RNCFALDYPAFDIKHLGQKTREGRLTKWYGGMAAALGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP
ASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSFS
RRLGKKSFGKSMKIDKKEGLENNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS

[0334] SEQ ID NO:21

[0335] MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGL
DIKSPGRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVPFYGM
GRNCFALDYPAFDIKHLGQKTREGRLTKWYGGMAAALGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSPFTLTP
PASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSF
SRRLGKKSFGKSMKIDKKEGLENNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKSGSGHHHHHH

[0336] SEQ ID NO:22

[0337] ATGGAAGTCACTGTGCCGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAAT
ATCGTCTACTTCTGGGTTCTGATACCGAACCGGTTCCGGACTACGAGCGCGTCTGAGCGAGCTGCTGTTGTATCT
GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCACCGG
GTCGTGTGAACATTATCGAGTACAAAGCGAAGAACCCGGCAGCGCATTACCCGTATGAGAATGGTGGCGGCTGGAAA
GCTGCACAAGAACTGGACGAATTTTTCAAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCTGATCATCATGCC
TACCTGGAATGATGAGAAAAATGGTCTGACAAATCCGGGTGGCGTTCGGTTCTATGGTATGGGTCGTAATTGTTTTG
CGTTGGACTACCCGGCGTTTGATATCAAGCACCTGGGTGAGAAAACGCGTGAGGGTCGTCTGCTGACGAAATGGTAC
GGTGGCATGGCGGCCGCGCTGGGCCACGGCCTGAATCTGCCGCACAATCACCAGACCGCGAGCGATGGCAAGAAATA
TGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGGTACCAGCCCCGACGTTCCCTGACCCCGCGAGCTGTGCGC
TGCTGGATGCCTGCGAAGTGTTTACGCGTTACCCCGAGCCAACAGTTTTATGAGGGTAAGCCAGAAGTCGAGGTTGGT
GATGTTGCAATTTCTTCAAGGGTGATCAAACTCTGGTACGCGGTAAGTACAAGAGCCCGCAAACCGTGAAAGCTCT
GAACGTTTACATTCAGGATCCGCCGTACGCCGTGAACCAAGACTACGATGCAGTGAGCTTTAGCCGTCTGCTGGGCA
AAAAGTCCGGTAAGTTTAGCATGAAGATTGACAAAAAAGAACTGGAAGGCCTGAATAACAACGAATTCGTATTTCC
TTGATGTTTATTCTGGCAAACGGCTTACACATGCAGAAGCACTTTACGTTTACTGGGATGCGCTGCAAGACTACCG
TGACGGTAGCAAATCTGGTTTCGGGTCATCATCACCACCATCACTGA

[0338] SEQ ID NO:23

[0339] GMAHALGHGL (被破坏的金属蛋白酶基序)

[0340] SEQ ID NO:24

[0341] GMAAELGHGL (被破坏的金属蛋白酶基序)

[0342] SEQ ID NO:25

[0343] GMAAALGHGL (被破坏的金属蛋白酶基序)

[0344] SEQ ID 26

[0345] 铜绿假单胞菌PA01 (去除了信号序列的天然序列)

[0346] ATQEEILDALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLQAYDPGAASMWVLPANP
DNTLPFLIGDKRVLASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLVNGDPAATAKDFKVSVVGVDK
TAALNGLKSAGLQPADAACNALTDASCASTSKLLVLGNGASAASLSATVRARLQAGLPILFVHTNGWNQSSTGQQIL
AGLGLQEGPYGGNYWDKDRVPSSRTRRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRTTCDDVPGLSD
FSKRVDVLKGALDAYNQKAQNLFALPGTTSRLWLLWADAVRQNIIRYPMDKAADTARFQETVADAIVGYVREAGAA
QKELGSYAGQRQQSMPVSGSEETLTLTLPQAQGFTAIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNTR
QYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPFLDIQPGEDSSQAIADFIQALD
ADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGWD
CDSETLHKLPQTQHINVDQYAQCGGGCSGNPYDQTWGLNPRGWGESHELGHNLQVNRLKVYGGRSGEISNQIFPLHK
DWRVLREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKNLPLQG
WDIWTLLYLHQKQVDKSDWDANKAALGYGTYAQRPGNSGDASSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAAA
QAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGSPAWPFP

[0347] 金属蛋白酶基序加有下划线。

[0348] SEQ ID 27

[0349] 多形拟杆菌VPI-5482 (去除了信号序列的天然序列)

[0350] DKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDLNPTGISVKAGDDIIVLVGDTYGGNISMQCIWETGT

EYKQTASSGDVYMLNPGVNKLTMKGEGQLFVMYNTELTSTAKPIKIHPLGSGTVNGFFDLKEHKTDEKYAELLKK
STHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNNHLFAISPEGSYMWASDY
QIGFVYTYLGNILLEDNVMAAEDNAWGAHEIGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRGNLGSVATA
RYANGQAWYNMGDATHQNEDETETHMRMNWQLWIIYHRCEYKTDWFQTLFKLMREVMTEGEDPGKKQLEFAKMASKA
ANQNLTDFFEMWGFEPVNTTIEQYGYTKYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRKKSEFPSNDYRYSAVGD
VGYYTQFKENQKITKAITAELAGRKVS IQNGDEAVAFELRENDENGKLLYFSTFTTTEIPSSILMVNAKLYAVQADG
KRILL

[0351] 金属蛋白酶基序加有下划线。

[0352] SEQ ID NO:28

[0353] 产气荚膜梭菌(去除了信号序列的天然序列)

[0354] VLELEMRGDSISEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQ
LKPGKNEITIP EINYENSGIPKDV IQGGDLFFTNYKSDSQKRAPKVRIEGASKYPVFI LGKSDENEVMKELEAYVEK
IKAEPKTT PNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRIMPMV
KNLSGGAFMNAGNGVIGIRPGNQDAILAANKGWGVAHELGHNFDTGGRTIVEVTNNMPLFFESKYKTKTRITDQNI
WENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYL DNTFYGKFERQFRERDFGNKNREDIYKSWVVAASDAM
ELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSGSNGNIKLSFSVDDENKD
NILGYEIRRDGKYVGFTSNDSFVDTKSNLDEDGVYVVTPYDRKLNTLNPIEVN

[0355] 金属蛋白酶基序加有下划线。

[0356] SED ID NO:29

[0357] 铜绿假单胞菌PA01(N末端Met,C末端接头/标签)

[0358] MATQEEILDAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLQAYDPGAASMWVLPAN
PDNTLPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTLQSGTNAAHAPLLKRAVQWL VNGDPGAATAKDFKVS VVGVD
KTAALNGLKSAGLQPADAACNALTDASCASTSKLLVLGNGASAASLSATVRARLQAGLPILFVHTNGWNQSSTGQQI
LAGLGLQEGPYGGNYWDKDRVPSSRTRTRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRITCDDVPGLS
DFSKRVDVLKGALDAYNQKAQNL FALPGTTSRLWLLWADAVRQNI RYPMDKAADTARFQETVADAIVGYVREAGA
AQKELGSYAGQRQQSMPVSGSEETLTLTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNT
RQYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPFLDIQPGEDSSQAIADFIQAL
DADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGW
DCDSETLHKLPQTQHINVDQYACGGGCSGNPYDQTWGLNPRGWGESHELGHNLQVNRLKVYGGRSGEISNQIFPLH
KDWRVLREFGQNLDDTRVNYRNAYNLI VAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKN DPLQ
GWDIWTLLYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGDASSTDGNDNLLGLSWLTQRDQRPTFALWGIRTSAA
AQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGSPA WPFPGSGHHHHHH

[0359] 金属蛋白酶基序加有下划线。

[0360] SEQ ID NO:30

[0361] 多形拟杆菌VPI-5482(N末端Met,C末端接头/标签)

[0362] MDKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDLDNPTGISVKAGDDIIVLVGDTYGNISMQCIWETG
TEYKQTASSGDVYMLNPGVNKLTMKGEGQLFVMYNTELTSTAKPIKIHPLGSGTVNGFFDLKEHKTDEKYAELLK
KSTHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNNHLFAISPEGSYMWASD
YQIGFVYTYLGNILLEDNVMAAEDNAWGAHEIGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRGNLGSVAT

ARYANGQAWYNMGDATHQNETTETHMRMNWQLWIYYHRCEYKTDWFQTLFKLMREVMNTEGEDPGKKQLEFAKMASK
AANQNLTDFFEMWGFPEPVNTTIEQYGTYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRKSEFPSNDYRYSAVG
DVGYYTQFKENQKITKAITAELAGRKVS IQNGDEAVAFELRENDENGKLLYFSTFTTTEIPSSILMVNAKLYAVQAD
GKRILLGSGHHHHHH

[0363] 金属蛋白酶基序加有下划线。

[0364] SEQ ID NO:31

[0365] 产气荚膜梭菌(N末端Met,C末端接头/标签)

[0366] MVLELEMRGDSISEAKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSF
QLKPGKNEITPEINYESNGIPKDVIQGGDLFFTNYSKSDSKRAPKVRIEGASKYPVFILGKSDENEVMKELEAYVE
KIKAEPKTTPNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRIMPM
VKNLSGGAFMNAGNGVIGIRPGNQDAILAANKGWGVAHELGHNFDTGGRTIVEVTNNMPLFFESKYKTKTRITDQN
IWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLNDTFYGKFERQFRERDFGNKNREDIYKSWVVAASDA
MELDLTEFFARHGIRVDDKVKEDLAKYPPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSGSNGNIKLSFSVDDENK
DNILGYEIRRDGKYVGFTSNDSSFVDTKSNLDEGYYVVTPTYDRKLNLTNP IEVNGSGHHHHHH

[0367] 金属蛋白酶基序加有下划线。

[0368] SEQ ID NO:32铜绿假单胞菌PA01(包含信号序列的完整天然序列)

[0369] MSLSTTAFPSLQGENMSRSPIPRHRALLAGFCLAGALSAQAATQEEILDAALVSGDSSQLTDSHLVALR
LQQQVERIRQTRTQLLDGLYQNLQAYDPAASMWVLPANPDNTLPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQ
LSGTNAAHAPLLKRAVQWL VNGDPAATAKDFKVS VGVVDKTAALNGLKSAGLQPADAACNALTDASCASTSKLLVL
GNGASAASLSATVRARLQAGLPILFVHTNGWNQSSTGQQILAGLGLQEGPYGGNYWDKDRVPSSRTRRSVELGGAY
GQDPALVQQIVDGSWRDQYDWSKCTSYVGRITCDDVPGLSDFSKRVDVLKALDAYNQKAQNL FALPGTTSRLWLL
WADAVRQNI RYPMDKAADTARFQET FVADAIVGYVREAGAAQKELGSYAGRQQSMPVSGSEETLTTLPSAQGFTA
IGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYS
GATPGQTVTVKVTGAASQPFLDIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSDKDYGGDVQ
RFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGWDCDSETLHKLPQTQHINVDQYACGGGCSGNPYDQTW
GLNPRGWGESHELGHNLQVNRLKVYGGRSGEISNQIFPLHKDWRVLRFEFGQNLDDTRVNYRNAYNLIVAGRAEADPL
AGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKNLPLQGWDIWTLTYLHQRQVDKSDWDANKAALGYGTYAQRPG
NSGDASSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAAAQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQG
SPAWPFP

[0370] Uniprot登录号:Q9I5W4.1

[0371] 金属蛋白酶基序加有下划线。

[0372] 信号序列加粗且加有下划线。

[0373] SEQ ID NO:33

[0374] 多形拟杆菌VPI-5482(包含信号序列和其他序列的完整天然序列)

[0375] MTKRFITNLLALFTLFTVSLACKDTEKSIINSSFSISEEYLIQNLDKSSTSVQIPINTSMELAQWSVS
YEANWLQCSKQKTAAEGTFLRITVNENTGETKRTANIKVTSTTATYITVNQYAKGEVIVEGDIKVTPGTGKASEHQ
EGQDIENTYDGKFSTDGAAPFHTPWGQSAKFPVTLEYFYGKDEIDYLIYYTRSGNGNFGKVKYTTTTNPDRSDYTL
QGEYDFKEQNAPSKVSFSEGIKATGIKFEVLSGLGDFVSCDEMEFYKTNTDKTLTKLLTVFTDITCTEIKNNVTNE
QIQALPDYFVRIAEAVRDNTYDKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDLNPTGISVKAGDDIIVLVGDTYG

QNISMQCIWETGTEYKQTASSGDVYMLNPGVKNLTMKGEGQLFVMYNTELTSNTAKPIKIHPLGSGTVNGFFDLKE
HKTDEKYAELLKKSTHKYFCIRGEKIMFYFHRNKLLEYVPPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNNHLFA
ISPEGSYMWASDYQIGFVYTYLGNILLEDNVMAAEDNAWGPAHEIGHVHQAAINWASSTESSNNLFSNFIIYKLGKY
KSRGNGLGSVATARYANGQAWYNMGDATHQNETTETHMRMNWQLWIIYHRCEYKTDWFQTLFKLMREVMTEGEDPG
KKQLEFAKMASKAANQNLTDFFEMWGFPEPVNTTIEQYGTYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRKKE
FPSNDYRYSAGVDVGYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTTTEIPSSIL
MVNAKLYAVQADGKRILL

[0376] Uniprot登录号:Q89ZX7.1

[0377] 金属蛋白酶基序加有下划线。

[0378] 信号序列加粗且加有下划线。

[0379] 成熟蛋白质中去除的其他序列加粗且倾斜。

[0380] SEQ ID NO:34

[0381] 产气荚膜梭菌(包含信号和其他序列的完整天然序列)

[0382] MNKRKIAAII LATMITNLSATTIDVLAQELNTKNSKVEVSHDDESHQARVSKFDLYNSDKLDAYNQEF
QVSRNIIKSINNNGGKYNSSITDKAIDGNLETHWETGKPNANFTNEVVVTFNEITNIDRIVYSARRDSARGKGFAG
EFEIYASLKDEGDDFNLVSSGEYTESTRDLVEIKFNPTDFKRLKFKFKKADQNWASAAEFMFYKEDKLNKFNGLFT
DSSMNKVSEEFNTLEKLNAFENELKDHPIDYDLYKEGLNNARAILTETSENPTKATLGQITYNLNDDYNNQYRMPYKN
IKAIKNNGRHYAAQNIKAIDNDVNTYWETGTLNSSSFNNEVEVEFNDLVTLDRIVYGSRQSDLKGFEEVYIYASR
TSKGDYTKLVATGAHEATKGLVEAKFEPTFKRVKFKFKSKQNSATLNELMFYKPDEVYSSIPKLFTDGTMSELSE
EFNSLEKINAFKEKAKNHPLYNDFNETIELAESLISNPRKEDVLELEMRGDSISEAKKRKVWNFQDWQITGLSARAG
DKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQLKPGKNEITIPEINYESNGIPKDVIIQGGDLFTNYKSDSQKRA
PKVRIEGASKYPVFILGKSDENEVMKELEAYVEKIKAEPKTTPNIFAVSSNKSLEFVQATYALDWYKKNNKTPKYTA
EQWDQYIADAMGFWGFDNSKDVNSDFNFRIMPMVKNLSSGAFMNAGNGVIGIRPGNQDAILAANKGWGVAHELGHNF
DTGGRTIVEVTNNMPLFFESKYKTKTRITDQNIWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNT
FYGKFERQFRERDFGKNREDIYKSWVVAASDAMELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNY
KGDGFTENAKVSSTSGSNGNIKLSFSVDDENKDNILGYEIRRDGKYVGFTSNDSFVDTKSNLDEDGVYVVTPTYDRK
LNTLNPIEVNALQPTLSVNPVITLALGEEFNEEYIVAKDIKGNLSSESVKVKSSNVNTSKVGEYEVLYSLEDKGN
EYTKTSKVNVSRRKEYMSDLTPKQSSNGWGTVRKDKSISGGVIGLTRDGDVFDYNGKGLGLHSNAEYVYDLEGKDYDY
FESYVGVDKAMSSRPASSVIFKVLVDGEEKFNSGVMRSTTPQKYVKVDVKNKELKLIVNDAGDGDSSDHASFGDAK
LATLSSKPIIKGENLAYSMDEKVDLMKGITATDIEDGNITSKVQIKSSDFVEGKSGIFTVVYSVTDSDGLTSECRT
IAVTDKETQLSDLNWKSATIGSGSVRKDRVSGNQIRLLNEDNSVETFAKGIGTHSYSEIVYNSEGYDIFDTWVGID
RHVADKKVSSVKFKVYVDGELKAETDVMRIDTPKKRLVVDVRNSKEIKLVVDVADNGNNWDHADWADAKFRNLAEYD
ASELNKAIEEAKKLDLNNYTEESSEALKNAISKGEEALLSKDKETINSALEELNKEMNSLVKVDLNAVINIPDKYLL
KSIQNQLNKTGDLTLGDMYSLTTLTSGVEDLTGLENAKNLETNMDYNEVKDLRPLSKLKKLNTLNAQEQAAGE
LKPSNGKVIKDSKVYNREGKNVAKTIRVVDKNGNTILEQDAKDEFTINTKDLSSGLYGVHVLFEDEGFSGVMFYLFN
V

[0383] Uniprot登录号:A0A0H2YN38.1

[0384] 金属蛋白酶基序加有下划线。

[0385] 信号序列加粗且加有下划线。

[0386] 成熟蛋白质中去除的其他序列加粗且倾斜。

[0387] SEQ ID NO:35

[0388] 铜绿假单胞菌PA01 (去除信号序列的双突变体)

[0389] ATQEEILDAAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLSQAYDPGAASMWVLPANP
DNTLPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLVNGDPAATAKDFKVSVVGVDK
TAALNGLKSAGLQPADAACNALTDASCASTSKLLVLGNGASAAASLSATVRARLQAGLPILFVHTNGWNQSSTGQQIL
AGLGLQEGPYGGNYWDKDRVPSSRTRTRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRITTCDDVPGLSD
FSKRVDVLKGALDAYNQKAQNLFALPGTTSRLWLLWADAVRQNIIRYPMDKAADTARFQETFVADAIVGYVREAGAA
QKELGSYAGQRQQSMPVSGSEETLTLTPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNTR
QYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPFLDIQPGEDSSQAIADFIQALD
ADKADWLEIRSGSVEVHAKVEKVRGSDIKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGWD
CDSETLHKLPGTQHINVDQYAQCGGGCSGNPYDQTWGLNPRGWGESAAALGHNLQVNRLKVYGGRSGEISNQIFPLHK
DWRVLREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQVWHYWADLKNLPLQG
WDIWTLLYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGDASSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAAA
QAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGSAPWFPF

[0390] 被破坏的金属蛋白酶基序加有下划线

[0391] SEQ ID NO:36

[0392] 多形拟杆菌VPI-5482 (去除信号序列和其他未成熟序列的双突变体)

[0393] DKWEKEFRIRSYPYSNIAEWADKLMTKKYSDDLNP TGISVKAGDDIIVLVGDTYGGNISMQCIWETGT
EYKQTASSGDVYMLNPGVNKLTMKGEGQLFVMYNTELTSTAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAELLKK
STHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNNHLFAISPEGSYMWASDY
QIGFVYTYLGNILLEDNVMAAEDNAWGPA^{AAA}IGHVHQ^{AA}INWASSTESSNNLFSNFIYKLGKYKSRGNLGSVATA
RYANGQAWYNMGDATHQ^{NED}TETHMRMNWQLWIIYHRCEYKTD^{FW}QTLFKLMREVMTEGEDPGKKQLEFAKMASKA
ANQNLTDFFEMWGFPEPVNTTIEQYGYTKYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRKKSEFPSNDYRYSVAGD
VGYYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTTFEIPSSILMVNAKLYAVQADG
KRILL

[0394] 被破坏的金属蛋白酶基序加有下划线

[0395] SEQ ID NO:37

[0396] 产气荚膜梭菌 (去除信号序列和其他未成熟序列的双突变体)

[0397] VLELEMRGDSISEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQ
LKPGKNEITIP^EINYESNGIPKDVIQGGDLFFTNYKSDSQKRAPKVRIEGASKYPVFILGKSDENEVMKELEAYVEK
IKAEPKTT^{PN}IFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRIMPMV
KNLSGGAFMNAGNGVIGIRPGNQDAILAANKGWGVAAALGHNFDTGGRTIVEVTNNMPLFFESKYKTKTRITDQNI
WENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYL^{YD}NTFYGKFERQFRERDFGNKNREDIYKSWVVAASDAM
ELDLTEFFARHGIRVDDKV^{KED}LAKYPKPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSGSNGNIKLSFSVDDENKD
NILGYEIRRDGKYVGFTSNDSFVDTKSNLDEDGVYVVTPYDRKLNTLNPIEVN

[0398] 被破坏的金属蛋白酶基序加有下划线

[0399] SEQ ID NO:38

[0400] 铜绿假单胞菌PA01 (去除信号序列并且具有N末端Met、C末端接头/标签的双突变

体)

[0401] MATQEEILDAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLQAYDPGAASMWVLPAN
PDNTLPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWL VNGDPGAATAKDFKVS VVGVD
KTAALNGLKSAGLQPADAACNALTDASCASTSKLLVLNGGASAAASLSATVRARLQAGLPILFVHTNGWNQSSTGQQI
LAGLGLQEGPYGGNYWDKDRVPSSRTRTRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRITCDDVPGLS
DFSKRVDVLKGALDAYNQKAQNLFALPGTTSRLWLLWADAVRQNI RYPMDKAADTARFQETVADAIVGYVREAGA
AQKELGSYAGQRQQSMPVSGSEETLTLTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNT
RQYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPFLDIQPGEDSSQAIADFIQAL
DADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGW
DCDSETLHKLPGTQHINVDQYACGGGCSGNPYDQTWGLNPRGWGESAAALGHNLQVNRLKVYGGRSGEISNQIFPLH
KDWRVLRREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKN DPLQ
GWDIWTLLYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGDASSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAA
AQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGSPA WPFPGSGHHHHHH

[0402] 被破坏的金属蛋白酶基序加有下划线

[0403] SEQ ID NO:39

[0404] 多形拟杆菌VPI-5482 (去除信号序列和来自不成熟蛋白的其他序列并且具有N末端Met、C末端接头/标签的双突变体)

[0405] MDKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDDL NPTGISVKAGDDIIVLVGDTYGNISMQCIWETG
TEYKQTASSGDVYMLNPGVNKLTMKGEGQLFVMYNTELTSTNAKPIKIHIPLSGTVNGFFDLKEHKTDEKYAELLK
KSTHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNNHLFAISPEGSYMWASD
YQIGFVYTYLGNILLEDNVMAAEDNAWGPA³AAIGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRGNGLGSVAT
ARYANGQAWYNMGDATHQNEDETHMRMNWLWIIYHRCEYKTD³FWQTLFKLMREVMTEGEDPGKKQLEFAKMASK
AANQNLTDFFEMWGFFEPVNTTIEQYGTYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRKSEFPSNDYRYSAVG
DVGYYTQFKENQKITKAITAELAGRKVS IQNGDEAVAFELRENDENGKLLYFSTFTT³FEIPSSILMVNAKLYAVQAD
GKRILLGSGHHHHHH

[0406] 被破坏的金属蛋白酶基序加有下划线

[0407] SEQ ID NO:40

[0408] 产气荚膜梭菌 (去除信号序列和来自不成熟蛋白的其他序列并且具有N末端Met、C末端接头/标签的双突变体)

[0409] MVLELEMRGDSISEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSF
QLKPGKNEITIPEINYESNGIPKDVIQGGDLFFTNYSKSDSQKRAPKVRIEGASKYPVFILGKSDENEVMKELEAYVE
KIKAEPKTTPNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRIMPM
VKNLSGGAFMNAGNGVIGIRPGNQDAILAANKGWGVAAALGHNFD³TGGRTIVEVTNNMPLFFESKYKTKTRITDQN
IWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNREDIYKSWVVAASDA
MELDLTEFFARHGIRVDDKVKEDLAKYPKPKDKIYYLNDLAMNYKGDGFTENAKVSVSTSGSNGNIKLSFSVDDENK
DNILGYEIRRDGKYVGFTSND³SFVDTKSNLDEDGVYVVTPTYDRKLNTLNPIEVNGSGHHHHHH

[0410] 被破坏的金属蛋白酶基序加有下划线

[0411] SEQ ID NO:41

[0412] HELGH (金属蛋白酶基序)

- [0413] SEQ ID NO:42
- [0414] HEIGH(金属蛋白酶基序)
- [0415] SEQ ID NO:43
- [0416] GVAHELGHNF(金属蛋白酶基序)
- [0417] SEQ ID NO:44
- [0418] HALGH(被破坏的金属蛋白酶基序)
- [0419] SEQ ID NO:45
- [0420] AELGH(被破坏的金属蛋白酶基序)
- [0421] SEQ ID NO:46
- [0422] AALGH(被破坏的金属蛋白酶基序)
- [0423] SEQ ID NO:47
- [0424] GKPRPYSPRPTSHPRPIRV(在T上具有O糖基化位点的糖基果蝇肽)
- [0425] SEQ ID NO:48
- [0426] YIYGSFK(非O-糖基化肽)
- [0427] SEQ ID NO:49
- [0428] KKLVFFA(非O-糖基化肽)
- [0429] SEQ ID NO:50
- [0430] FLPLILGKLVKGLL(非O-糖基化肽)

[0001]	序列表									
[0002]	<110> 杰诺维斯公司									
[0003]	<120> 用于O-糖蛋白的蛋白酶和结合多肽									
[0004]	<130> N410795W0									
[0005]	<150> GB1708471.6									
[0006]	<151> 2017-05-26									
[0007]	<150> GB1708476.5									
[0008]	<151> 2017-05-26									
[0009]	<150> GB1806655.5									
[0010]	<151> 2018-04-24									
[0011]	<160> 50									
[0012]	<170> PatentIn version 3.5									
[0013]	<210> 1									
[0014]	<211> 361									
[0015]	<212> PRT									
[0016]	<213> Akkermansia muciniphila									
[0017]	<400> 1									
[0018]	Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys Lys									
[0019]	1 5 10 15									
[0020]	Thr Ala Arg Gln Leu Asn Ile Val Tyr Phe Leu Gly Ser Asp Thr Glu									
[0021]	20 25 30									
[0022]	Pro Val Pro Asp Tyr Glu Arg Arg Leu Ser Glu Leu Leu Tyr Leu									
[0023]	35 40 45									
[0024]	Gln Gln Phe Tyr Gly Lys Glu Met Gln Arg His Gly Tyr Gly Ala Arg									
[0025]	50 55 60									
[0026]	Ser Phe Gly Leu Asp Ile Lys Ser Pro Gly Arg Val Asn Ile Ile Glu									
[0027]	65 70 75 80									
[0028]	Tyr Lys Ala Lys Asn Pro Ala Ala His Tyr Pro Tyr Glu Asn Gly Gly									
[0029]	85 90 95									
[0030]	Gly Trp Lys Ala Ala Gln Glu Leu Asp Glu Phe Phe Lys Ala His Pro									
[0031]	100 105 110									
[0032]	Asp Arg Lys Lys Ser Gln His Thr Leu Ile Ile Met Pro Thr Trp Asn									
[0033]	115 120 125									
[0034]	Asp Glu Lys Asn Gly Pro Asp Asn Pro Gly Gly Val Pro Phe Tyr Gly									
[0035]	130 135 140									
[0036]	Met Gly Arg Asn Cys Phe Ala Leu Asp Tyr Pro Ala Phe Asp Ile Lys									
[0037]	145 150 155 160									
[0038]	His Leu Gly Gln Lys Thr Arg Glu Gly Arg Leu Leu Thr Lys Trp Tyr									
[0039]	165 170 175									
[0040]	Gly Gly Met Ala His Glu Leu Gly His Gly Leu Asn Leu Pro His Asn									
[0041]	180 185 190									

[0042]	His Gln Thr Ala Ser Asp Gly Lys Lys Tyr Gly Thr Ala Leu Met Gly
[0043]	195 200 205
[0044]	Ser Gly Asn Tyr Thr Phe Gly Thr Ser Pro Thr Phe Leu Thr Pro Ala
[0045]	210 215 220
[0046]	Ser Cys Ala Leu Leu Asp Ala Cys Glu Val Phe Ser Val Thr Pro Ser
[0047]	225 230 235 240
[0048]	Gln Gln Phe Tyr Glu Gly Lys Pro Glu Val Glu Val Gly Asp Val Ala
[0049]	245 250 255
[0050]	Ile Ser Phe Lys Gly Asp Gln Ile Leu Val Ser Gly Asn Tyr Lys Ser
[0051]	260 265 270
[0052]	Pro Gln Thr Val Lys Ala Leu Asn Val Tyr Ile Gln Asp Pro Pro Tyr
[0053]	275 280 285
[0054]	Ala Val Asn Gln Asp Tyr Asp Ala Val Ser Phe Ser Arg Arg Leu Gly
[0055]	290 295 300
[0056]	Lys Lys Ser Gly Lys Phe Ser Met Lys Ile Asp Lys Lys Glu Leu Glu
[0057]	305 310 315 320
[0058]	Gly Leu Asn Asn Asn Glu Phe Arg Ile Ser Leu Met Phe Ile Leu Ala
[0059]	325 330 335
[0060]	Asn Gly Leu His Met Gln Lys His Phe Thr Phe His Trp Asp Ala Leu
[0061]	340 345 350
[0062]	Gln Asp Tyr Arg Asp Gly Ser Lys Ser
[0063]	355 360
[0064]	<210> 2
[0065]	<211> 371
[0066]	<212> PRT
[0067]	<213> Artificial Sequence
[0068]	<220>
[0069]	<223> LS (N terminal methionine and a C-terminal linker + His6 tag)
[0070]	<400> 2
[0071]	Met Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys
[0072]	1 5 10 15
[0073]	Lys Thr Ala Arg Gln Leu Asn Ile Val Tyr Phe Leu Gly Ser Asp Thr
[0074]	20 25 30
[0075]	Glu Pro Val Pro Asp Tyr Glu Arg Arg Leu Ser Glu Leu Leu Leu Tyr
[0076]	35 40 45
[0077]	Leu Gln Gln Phe Tyr Gly Lys Glu Met Gln Arg His Gly Tyr Gly Ala
[0078]	50 55 60
[0079]	Arg Ser Phe Gly Leu Asp Ile Lys Ser Pro Gly Arg Val Asn Ile Ile
[0080]	65 70 75 80
[0081]	Glu Tyr Lys Ala Lys Asn Pro Ala Ala His Tyr Pro Tyr Glu Asn Gly
[0082]	85 90 95
[0083]	Gly Gly Trp Lys Ala Ala Gln Glu Leu Asp Glu Phe Phe Lys Ala His

[0084]	100	105	110
[0085]	Pro Asp Arg Lys Lys Ser Gln His Thr Leu Ile Ile Met Pro Thr Trp		
[0086]	115	120	125
[0087]	Asn Asp Glu Lys Asn Gly Pro Asp Asn Pro Gly Gly Val Pro Phe Tyr		
[0088]	130	135	140
[0089]	Gly Met Gly Arg Asn Cys Phe Ala Leu Asp Tyr Pro Ala Phe Asp Ile		
[0090]	145	150	155
[0091]	Lys His Leu Gly Gln Lys Thr Arg Glu Gly Arg Leu Leu Thr Lys Trp		
[0092]	165	170	175
[0093]	Tyr Gly Gly Met Ala His Glu Leu Gly His Gly Leu Asn Leu Pro His		
[0094]	180	185	190
[0095]	Asn His Gln Thr Ala Ser Asp Gly Lys Lys Tyr Gly Thr Ala Leu Met		
[0096]	195	200	205
[0097]	Gly Ser Gly Asn Tyr Thr Phe Gly Thr Ser Pro Thr Phe Leu Thr Pro		
[0098]	210	215	220
[0099]	Ala Ser Cys Ala Leu Leu Asp Ala Cys Glu Val Phe Ser Val Thr Pro		
[0100]	225	230	235
[0101]	Ser Gln Gln Phe Tyr Glu Gly Lys Pro Glu Val Glu Val Gly Asp Val		
[0102]	245	250	255
[0103]	Ala Ile Ser Phe Lys Gly Asp Gln Ile Leu Val Ser Gly Asn Tyr Lys		
[0104]	260	265	270
[0105]	Ser Pro Gln Thr Val Lys Ala Leu Asn Val Tyr Ile Gln Asp Pro Pro		
[0106]	275	280	285
[0107]	Tyr Ala Val Asn Gln Asp Tyr Asp Ala Val Ser Phe Ser Arg Arg Leu		
[0108]	290	295	300
[0109]	Gly Lys Lys Ser Gly Lys Phe Ser Met Lys Ile Asp Lys Lys Glu Leu		
[0110]	305	310	315
[0111]	Glu Gly Leu Asn Asn Asn Glu Phe Arg Ile Ser Leu Met Phe Ile Leu		
[0112]	325	330	335
[0113]	Ala Asn Gly Leu His Met Gln Lys His Phe Thr Phe His Trp Asp Ala		
[0114]	340	345	350
[0115]	Leu Gln Asp Tyr Arg Asp Gly Ser Lys Ser Gly Ser Gly His His His		
[0116]	355	360	365
[0117]	His His His		
[0118]	370		
[0119]	<210> 3		
[0120]	<211> 1116		
[0121]	<212> DNA		
[0122]	<213> Artificial Sequence		
[0123]	<220>		
[0124]	<223> Nucleotide sequence of LS		
[0125]	<400> 3		

[0126]	atggaagtca ctgtgccgga cgccctgaaa gatcgcatcg cgctgaagaa aaccgctcgt	60
[0127]	cagctgaata tcgtctactt cctgggttct gataccgaac cggttccgga ctacgagcgc	120
[0128]	cgtctgagcg agctgtgtt gtatctgcag caattctatg gtaaagaaat gcagcgccat	180
[0129]	ggctatggcg cacgcagctt tggctctggac attaagtcac cgggtcgtgt gaacattatc	240
[0130]	gagtacaaag cgaagaaccc ggcagcgcat taccctgatg agaatggtgg cggctggaaa	300
[0131]	gctgcacaag aactggacga atttttcaag gcccatccag accgcaagaa aagccagcac	360
[0132]	acctgatca tcatgcctac ctggaatgat gagaaaaatg gtcttgacaa tccgggtggc	420
[0133]	gttccgttct atggtatggg tcgtaattgt tttgcgttgg actaccggc gtttgatc	480
[0134]	aagcacctgg gtcagaaaac gcgtgagggt cgtctgctga cgaaatggtg cgggtggcatg	540
[0135]	gcgcacgaac tgggccacgg cctgaatctg ccgcacaatc accagaccgc gagcgatggc	600
[0136]	aagaaatg gcaccgccct gatgggtagc ggcaactaca cgttcggtag cagcccgacg	660
[0137]	ttctgaccc cggcgagctg tgcgtgctg gatgcctgcg aagtgttcag cgttacccc	720
[0138]	agccaacagt tttatgagg taagccagaa gtcgaggtg gtgatgttc aatttccttc	780
[0139]	aagggtgatc aaatcttgg cagcggtaac tacaagagcc cgcaaaccgt gaaagctctg	840
[0140]	aacgtttaca ttcaggatcc gccgtacgcc gtgaaccaag actacgatgc agtgagcttt	900
[0141]	agccgtcgtc tgggcaaaaa gtccggtgaag tttagcatga agattgacaa aaaagaactg	960
[0142]	gaaggcctga ataacaacga attccgtatt tccttgatgt tcattctggc aaacggctta	1020
[0143]	cacatgcaga agcactttac gtttactgg gatgcgtgc aagactaccg tgacggtagc	1080
[0144]	aaatctggtt cgggtcatca tcaccaccat cactga	1116
[0145]	<210> 4	
[0146]	<211> 385	
[0147]	<212> PRT	
[0148]	<213> Akkermansia muciniphila	
[0149]	<400> 4	
[0150]	Met Leu Lys Arg Leu Leu Ser Ala Phe Phe Ser Leu Phe Phe Leu Gly	
[0151]	1 5 10 15	
[0152]	Ala Ala Ser Gly Thr Ser Phe Ala Glu Val Thr Val Pro Asp Ala Leu	
[0153]	20 25 30	
[0154]	Lys Asp Arg Ile Ala Leu Lys Lys Thr Ala Arg Gln Leu Asn Ile Val	
[0155]	35 40 45	
[0156]	Tyr Phe Leu Gly Ser Asp Thr Glu Pro Val Pro Asp Tyr Glu Arg Arg	
[0157]	50 55 60	
[0158]	Leu Ser Glu Leu Leu Leu Tyr Leu Gln Gln Phe Tyr Gly Lys Glu Met	
[0159]	65 70 75 80	
[0160]	Gln Arg His Gly Tyr Gly Ala Arg Ser Phe Gly Leu Asp Ile Lys Ser	
[0161]	85 90 95	
[0162]	Pro Gly Arg Val Asn Ile Ile Glu Tyr Lys Ala Lys Asn Pro Ala Ala	
[0163]	100 105 110	
[0164]	His Tyr Pro Tyr Glu Asn Gly Gly Gly Trp Lys Ala Ala Gln Glu Leu	
[0165]	115 120 125	
[0166]	Asp Glu Phe Phe Lys Ala His Pro Asp Arg Lys Lys Ser Gln His Thr	
[0167]	130 135 140	

[0168]	Leu Ile Ile Met Pro Thr Trp Asn Asp Glu Lys Asn Gly Pro Asp Asn
[0169]	145 150 155 160
[0170]	Pro Gly Gly Val Pro Phe Tyr Gly Met Gly Arg Asn Cys Phe Ala Leu
[0171]	165 170 175
[0172]	Asp Tyr Pro Ala Phe Asp Ile Lys His Leu Gly Gln Lys Thr Arg Glu
[0173]	180 185 190
[0174]	Gly Arg Leu Leu Thr Lys Trp Tyr Gly Gly Met Ala His Glu Leu Gly
[0175]	195 200 205
[0176]	His Gly Leu Asn Leu Pro His Asn His Gln Thr Ala Ser Asp Gly Lys
[0177]	210 215 220
[0178]	Lys Tyr Gly Thr Ala Leu Met Gly Ser Gly Asn Tyr Thr Phe Gly Thr
[0179]	225 230 235 240
[0180]	Ser Pro Thr Phe Leu Thr Pro Ala Ser Cys Ala Leu Leu Asp Ala Cys
[0181]	245 250 255
[0182]	Glu Val Phe Ser Val Thr Pro Ser Gln Gln Phe Tyr Glu Gly Lys Pro
[0183]	260 265 270
[0184]	Glu Val Glu Val Gly Asp Val Ala Ile Ser Phe Lys Gly Asp Gln Ile
[0185]	275 280 285
[0186]	Leu Val Ser Gly Asn Tyr Lys Ser Pro Gln Thr Val Lys Ala Leu Asn
[0187]	290 295 300
[0188]	Val Tyr Ile Gln Asp Pro Pro Tyr Ala Val Asn Gln Asp Tyr Asp Ala
[0189]	305 310 315 320
[0190]	Val Ser Phe Ser Arg Arg Leu Gly Lys Lys Ser Gly Lys Phe Ser Met
[0191]	325 330 335
[0192]	Lys Ile Asp Lys Lys Glu Leu Glu Gly Leu Asn Asn Asn Glu Phe Arg
[0193]	340 345 350
[0194]	Ile Ser Leu Met Phe Ile Leu Ala Asn Gly Leu His Met Gln Lys His
[0195]	355 360 365
[0196]	Phe Thr Phe His Trp Asp Ala Leu Gln Asp Tyr Arg Asp Gly Ser Lys
[0197]	370 375 380
[0198]	Ser
[0199]	385
[0200]	<210> 5
[0201]	<211> 361
[0202]	<212> PRT
[0203]	<213> Artificial Sequence
[0204]	<220>
[0205]	<223> Polypeptide capable of binding to O-glycans but lacks or has
[0206]	reduced O-glycoprotein-specific endoprotease activity
[0207]	<400> 5
[0208]	Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys Lys
[0209]	1 5 10 15

[0210]	Thr	Ala	Arg	Gln	Leu	Asn	Ile	Val	Tyr	Phe	Leu	Gly	Ser	Asp	Thr	Glu
[0211]				20				25					30			
[0212]	Pro	Val	Pro	Asp	Tyr	Glu	Arg	Arg	Leu	Ser	Glu	Leu	Leu	Leu	Tyr	Leu
[0213]			35					40					45			
[0214]	Gln	Gln	Phe	Tyr	Gly	Lys	Glu	Met	Gln	Arg	His	Gly	Tyr	Gly	Ala	Arg
[0215]		50					55					60				
[0216]	Ser	Phe	Gly	Leu	Asp	Ile	Lys	Ser	Pro	Gly	Arg	Val	Asn	Ile	Ile	Glu
[0217]	65					70					75				80	
[0218]	Tyr	Lys	Ala	Lys	Asn	Pro	Ala	Ala	His	Tyr	Pro	Tyr	Glu	Asn	Gly	Gly
[0219]					85					90					95	
[0220]	Gly	Trp	Lys	Ala	Ala	Gln	Glu	Leu	Asp	Glu	Phe	Phe	Lys	Ala	His	Pro
[0221]				100					105						110	
[0222]	Asp	Arg	Lys	Lys	Ser	Gln	His	Thr	Leu	Ile	Ile	Met	Pro	Thr	Trp	Asn
[0223]				115				120						125		
[0224]	Asp	Glu	Lys	Asn	Gly	Pro	Asp	Asn	Pro	Gly	Gly	Val	Pro	Phe	Tyr	Gly
[0225]		130					135					140				
[0226]	Met	Gly	Arg	Asn	Cys	Phe	Ala	Leu	Asp	Tyr	Pro	Ala	Phe	Asp	Ile	Lys
[0227]	145					150					155				160	
[0228]	His	Leu	Gly	Gln	Lys	Thr	Arg	Glu	Gly	Arg	Leu	Leu	Thr	Lys	Trp	Tyr
[0229]					165					170					175	
[0230]	Gly	Gly	Met	Ala	His	Ala	Leu	Gly	His	Gly	Leu	Asn	Leu	Pro	His	Asn
[0231]				180					185						190	
[0232]	His	Gln	Thr	Ala	Ser	Asp	Gly	Lys	Lys	Tyr	Gly	Thr	Ala	Leu	Met	Gly
[0233]			195						200					205		
[0234]	Ser	Gly	Asn	Tyr	Thr	Phe	Gly	Thr	Ser	Pro	Thr	Phe	Leu	Thr	Pro	Ala
[0235]		210					215						220			
[0236]	Ser	Cys	Ala	Leu	Leu	Asp	Ala	Cys	Glu	Val	Phe	Ser	Val	Thr	Pro	Ser
[0237]	225					230					235				240	
[0238]	Gln	Gln	Phe	Tyr	Glu	Gly	Lys	Pro	Glu	Val	Glu	Val	Gly	Asp	Val	Ala
[0239]					245					250					255	
[0240]	Ile	Ser	Phe	Lys	Gly	Asp	Gln	Ile	Leu	Val	Ser	Gly	Asn	Tyr	Lys	Ser
[0241]				260					265					270		
[0242]	Pro	Gln	Thr	Val	Lys	Ala	Leu	Asn	Val	Tyr	Ile	Gln	Asp	Pro	Pro	Tyr
[0243]				275					280					285		
[0244]	Ala	Val	Asn	Gln	Asp	Tyr	Asp	Ala	Val	Ser	Phe	Ser	Arg	Arg	Leu	Gly
[0245]		290						295					300			
[0246]	Lys	Lys	Ser	Gly	Lys	Phe	Ser	Met	Lys	Ile	Asp	Lys	Lys	Glu	Leu	Glu
[0247]	305					310					315				320	
[0248]	Gly	Leu	Asn	Asn	Asn	Glu	Phe	Arg	Ile	Ser	Leu	Met	Phe	Ile	Leu	Ala
[0249]					325					330					335	
[0250]	Asn	Gly	Leu	His	Met	Gln	Lys	His	Phe	Thr	Phe	His	Trp	Asp	Ala	Leu
[0251]				340					345					350		

[0252]	Gln Asp Tyr Arg Asp Gly Ser Lys Ser
[0253]	355 360
[0254]	<210> 6
[0255]	<211> 371
[0256]	<212> PRT
[0257]	<213> Artificial Sequence
[0258]	<220>
[0259]	<223> LS E206A (N-terminal methionine and a C-terminal linker + His6
[0260]	tag)
[0261]	<400> 6
[0262]	Met Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys
[0263]	1 5 10 15
[0264]	Lys Thr Ala Arg Gln Leu Asn Ile Val Tyr Phe Leu Gly Ser Asp Thr
[0265]	20 25 30
[0266]	Glu Pro Val Pro Asp Tyr Glu Arg Arg Leu Ser Glu Leu Leu Leu Tyr
[0267]	35 40 45
[0268]	Leu Gln Gln Phe Tyr Gly Lys Glu Met Gln Arg His Gly Tyr Gly Ala
[0269]	50 55 60
[0270]	Arg Ser Phe Gly Leu Asp Ile Lys Ser Pro Gly Arg Val Asn Ile Ile
[0271]	65 70 75 80
[0272]	Glu Tyr Lys Ala Lys Asn Pro Ala Ala His Tyr Pro Tyr Glu Asn Gly
[0273]	85 90 95
[0274]	Gly Gly Trp Lys Ala Ala Gln Glu Leu Asp Glu Phe Phe Lys Ala His
[0275]	100 105 110
[0276]	Pro Asp Arg Lys Lys Ser Gln His Thr Leu Ile Ile Met Pro Thr Trp
[0277]	115 120 125
[0278]	Asn Asp Glu Lys Asn Gly Pro Asp Asn Pro Gly Gly Val Pro Phe Tyr
[0279]	130 135 140
[0280]	Gly Met Gly Arg Asn Cys Phe Ala Leu Asp Tyr Pro Ala Phe Asp Ile
[0281]	145 150 155 160
[0282]	Lys His Leu Gly Gln Lys Thr Arg Glu Gly Arg Leu Leu Thr Lys Trp
[0283]	165 170 175
[0284]	Tyr Gly Gly Met Ala His Ala Leu Gly His Gly Leu Asn Leu Pro His
[0285]	180 185 190
[0286]	Asn His Gln Thr Ala Ser Asp Gly Lys Lys Tyr Gly Thr Ala Leu Met
[0287]	195 200 205
[0288]	Gly Ser Gly Asn Tyr Thr Phe Gly Thr Ser Pro Thr Phe Leu Thr Pro
[0289]	210 215 220
[0290]	Ala Ser Cys Ala Leu Leu Asp Ala Cys Glu Val Phe Ser Val Thr Pro
[0291]	225 230 235 240
[0292]	Ser Gln Gln Phe Tyr Glu Gly Lys Pro Glu Val Glu Val Gly Asp Val
[0293]	245 250 255

[0294]	Ala Ile Ser Phe Lys Gly Asp Gln Ile Leu Val Ser Gly Asn Tyr Lys	
[0295]	260	265 270
[0296]	Ser Pro Gln Thr Val Lys Ala Leu Asn Val Tyr Ile Gln Asp Pro Pro	
[0297]	275	280 285
[0298]	Tyr Ala Val Asn Gln Asp Tyr Asp Ala Val Ser Phe Ser Arg Arg Leu	
[0299]	290	295 300
[0300]	Gly Lys Lys Ser Gly Lys Phe Ser Met Lys Ile Asp Lys Lys Glu Leu	
[0301]	305	310 315 320
[0302]	Glu Gly Leu Asn Asn Asn Glu Phe Arg Ile Ser Leu Met Phe Ile Leu	
[0303]	325	330 335
[0304]	Ala Asn Gly Leu His Met Gln Lys His Phe Thr Phe His Trp Asp Ala	
[0305]	340	345 350
[0306]	Leu Gln Asp Tyr Arg Asp Gly Ser Lys Ser Gly Ser Gly His His His	
[0307]	355	360 365
[0308]	His His His	
[0309]	370	
[0310]	<210> 7	
[0311]	<211> 1116	
[0312]	<212> DNA	
[0313]	<213> Artificial Sequence	
[0314]	<220>	
[0315]	<223> Nucleotide sequence encoding LS E206A	
[0316]	<400> 7	
[0317]	atggaagtca ctgtgccgga cgccctgaaa gatcgcatcg cgctgaagaa aaccgctcgt	60
[0318]	cagctgaata tcgtctactt cctgggttct gataccgaac cggttccgga ctacgagcgc	120
[0319]	cgtctgagcg agctgtgttt gtatctgcag caattctatg gtaaagaaat gcagcgccat	180
[0320]	ggctatggcg cacgcagctt tggctctggac attaagtcac cgggtcgtgt gaacattatc	240
[0321]	gagtacaaag cgaagaaccc ggcagcgcac taccgctatg agaatggtgg cggctggaaa	300
[0322]	gctgcacaag aactggacga atttttcaag gcccatccag accgcaagaa aagccagcac	360
[0323]	accctgatca tcatgcctac ctggaatgat gagaaaaatg gtcttgacaa tccgggtggc	420
[0324]	gttcggttct atggtatggg tcgtaattgt tttgcgttgg actaccggcg gtttgatatt	480
[0325]	aagcacctgg gtcagaaaac gcgtgagggt cgtctgctga cgaaatggta cgggtggcatg	540
[0326]	gcgcacgcgc tgggccacgg cctgaatctg ccgcacaatc accagaccgc gagcgatggc	600
[0327]	aagaaatatg gcaccgccct gatgggttagc ggcaactaca cgttcggtac cagcccgacg	660
[0328]	ttcctgaccc cggcgagctg tgcgctgctg gatgcctgcg aagtgttcag cgttaccccg	720
[0329]	agccaacagt tttatgaggg taagccagaa gtcgaggttg gtgatgttgc aatttccttc	780
[0330]	aagggtgatc aaatcttggc cagcggtaac tacaagagcc cgcaaaccgt gaaagctctg	840
[0331]	aacgtttaca ttcaggatcc gccgtacgcc gtgaaccaag actacgatgc agtgagcttt	900
[0332]	agccgtcgtc tgggcaaaaa gtccggttaag tttagcatga agattgacaa aaaagaactg	960
[0333]	gaaggcctga ataacaacga attccgtatt tccttgatgt tcattctggc aaacggctta	1020
[0334]	cacatgcaga agcactttac gtttacttgg gatgcgctgc aagactaccg tgacggtagc	1080
[0335]	aaatctgggt cgggtcatca tcaccacat cactga	1116

[0336]	<210>	8
[0337]	<211>	10
[0338]	<212>	PRT
[0339]	<213>	Artificial Sequence
[0340]	<220>	
[0341]	<223>	Metalloprotease domain motif
[0342]	<400>	8
[0343]	Gly Met Ala His Glu Leu Gly His Gly Leu	
[0344]	1 5 10	
[0345]	<210>	9
[0346]	<211>	595
[0347]	<212>	PRT
[0348]	<213>	Akkermansia muciniphila
[0349]	<400>	9
[0350]	Met Lys Asn Leu Leu Phe Ala Leu Leu Thr Gly Ser Phe Cys Cys Cys	
[0351]	1 5 10 15	
[0352]	Tyr Ala Gln Gln Lys Ala Ala Pro Val Pro Glu Pro Glu Val Val Ala	
[0353]	20 25 30	
[0354]	Thr Pro Pro Ala Asp Ala Gly Arg Gly Leu Ile Arg Val Asp Ser Arg	
[0355]	35 40 45	
[0356]	Glu Ile Arg His Tyr Ser Gly Thr Arg Lys Glu Pro Asp Tyr Leu Val	
[0357]	50 55 60	
[0358]	Ser Arg Asp Asn Gly Lys Thr Trp Glu Met Lys Ala Ala Pro Ala Gly	
[0359]	65 70 75 80	
[0360]	Tyr Pro Pro Asn Tyr Gly Gly Ile Pro Lys Glu Ser Pro Ala Ile Val	
[0361]	85 90 95	
[0362]	Arg Asn Pro Leu Thr Arg Glu Phe Ile Arg Val Gln Pro Ile Gly Gly	
[0363]	100 105 110	
[0364]	Phe Val Phe Leu Ser Arg Gly Gly Leu Asp Gly Lys Trp Leu Ala Val	
[0365]	115 120 125	
[0366]	Thr Asn Asp Gly Lys Leu Glu Glu Asp Trp Lys Asp Pro Glu Lys Arg	
[0367]	130 135 140	
[0368]	Lys Asn Leu Lys Lys Leu Gly Gly Ile Met Arg Thr Pro Val Phe Val	
[0369]	145 150 155 160	
[0370]	Asn Lys Gly Arg Arg Val Ile Val Pro Phe His Asn Met Gly Gly Gly	
[0371]	165 170 175	
[0372]	Thr Lys Phe His Ile Ser Asp Asp Gly Gly Leu Thr Trp His Val Ser	
[0373]	180 185 190	
[0374]	Arg Asn Gly Val Thr Ser Pro Arg His Glu Ala Arg Pro Pro His Gln	
[0375]	195 200 205	
[0376]	GlyVal Arg Trp Phe Asn Asn Ala Val Glu Ala Thr Val Leu Glu Met	
[0377]	210 215 220	

[0378]	Lys Asp Gly Thr Leu Trp Ala Leu Ala Arg Thr Ser Gln Asp Gln Ala		
[0379]	225	230	235 240
[0380]	Trp Gln Ala Phe Ser Lys Asp Tyr Gly Glu Thr Trp Ser Lys Pro Glu		
[0381]		245	250 255
[0382]	Pro Ser Arg Phe Phe Gly Thr Leu Thr Met Asn Thr Leu Gly Arg Leu		
[0383]		260	265 270
[0384]	Asp Asp Gly Thr Ile Val Ser Leu Trp Thr Asn Thr Met Ala Leu Pro		
[0385]		275	280 285
[0386]	Glu Asn Ala Thr Ala Gly Asn Gly Thr Trp Glu Asp Val Phe Thr Asn		
[0387]	290	295	300
[0388]	Arg Asp Ser His His Ile Ala Met Ser Gly Asp Glu Gly Lys Thr Trp		
[0389]	305	310	315 320
[0390]	Tyr Gly Phe Arg Glu Ile Ile Leu Asp Glu His Arg Asn His Pro Gly		
[0391]		325	330 335
[0392]	Tyr Ala Thr Leu Asp Gly Pro Glu Asp Arg Gly Lys His Gln Ser Glu		
[0393]		340	345 350
[0394]	Met Val Gln Leu Asp Lys Asn Arg Ile Leu Ile Ser Leu Gly Gln His		
[0395]		355	360 365
[0396]	Lys Asn His Arg Arg Leu Val Ile Val Asp Arg Arg Trp Val Gly Ala		
[0397]		370	375 380
[0398]	Lys Thr Arg Ala Thr Gln Thr Gly Lys Asp Leu Asp Ser Gln Trp Thr		
[0399]	385	390	395 400
[0400]	Ile His Thr Tyr Ile Pro Gln Lys Lys Gly His Cys Ser Tyr Asn Arg		
[0401]		405	410 415
[0402]	Lys Pro Ser Ala Glu Leu Val Gln Asp Pro Ser Gly Gly Thr Lys Lys		
[0403]		420	425 430
[0404]	Val Leu Gln Ile Lys Arg Leu Asp Asp Pro Glu Leu Val Asn Glu Lys		
[0405]		435	440 445
[0406]	Ser Asn Val Asp Tyr Arg Asn Gly Gly Ala Thr Trp Asn Phe Pro Asn		
[0407]		450	455 460
[0408]	Gly Thr Thr Gly Leu Val Lys Phe Arg Phe Arg Val Val Asp Gly Glu		
[0409]	465	470	475 480
[0410]	Gln Ala Asp Asp Ser Gly Leu Gln Val Ser Leu Thr Asp Arg Leu Phe		
[0411]		485	490 495
[0412]	Asn Ala Cys Asp Ser Thr Thr Lys Asp Tyr Ala Leu Phe Thr Phe Pro		
[0413]		500	505 510
[0414]	Ile Arg Leu Lys Pro Ala Pro His Leu Leu Leu Gly Met Lys Lys Val		
[0415]		515	520 525
[0416]	Pro Phe Thr Pro Gly Ala Trp His Glu Ile Ser Leu Leu Trp Gln Gly		
[0417]		530	535 540
[0418]	Gly Gln Ala Val Val Ser Leu Asp Gly Lys Lys Ala Gly Thr Leu Lys		
[0419]	545	550	555 560

[0420]	Met Ala Asn Lys Ser Pro Asn Gly Ala Ser Tyr Ile His Phe Ile Ser
[0421]	565 570 575
[0422]	Thr Gly Ser Gln Pro Asp Ala Gly Ile Leu Leu Asp Thr Val Asn Ala
[0423]	580 585 590
[0424]	Arg Val Lys
[0425]	595
[0426]	<210> 10
[0427]	<211> 577
[0428]	<212> PRT
[0429]	<213> Akkermansia muciniphila
[0430]	<400> 10
[0431]	Gln Gln Lys Ala Ala Pro Val Pro Glu Pro Glu Val Val Ala Thr Pro
[0432]	1 5 10 15
[0433]	Pro Ala Asp Ala Gly Arg Gly Leu Ile Arg Val Asp Ser Arg Glu Ile
[0434]	20 25 30
[0435]	Arg His Tyr Ser Gly Thr Arg Lys Glu Pro Asp Tyr Leu Val Ser Arg
[0436]	35 40 45
[0437]	Asp Asn Gly Lys Thr Trp Glu Met Lys Ala Ala Pro Ala Gly Tyr Pro
[0438]	50 55 60
[0439]	Pro Asn Tyr Gly Gly Ile Pro Lys Glu Ser Pro Ala Ile Val Arg Asn
[0440]	65 70 75 80
[0441]	Pro Leu Thr Arg Glu Phe Ile Arg Val Gln Pro Ile Gly Gly Phe Val
[0442]	85 90 95
[0443]	Phe Leu Ser Arg Gly Gly Leu Asp Gly Lys Trp Leu Ala Val Thr Asn
[0444]	100 105 110
[0445]	Asp Gly Lys Leu Glu Glu Asp Trp Lys Asp Pro Glu Lys Arg Lys Asn
[0446]	115 120 125
[0447]	Leu Lys Lys Leu Gly Gly Ile Met Arg Thr Pro Val Phe Val Asn Lys
[0448]	130 135 140
[0449]	Gly Arg Arg Val Ile Val Pro Phe His Asn Met Gly Gly Gly Thr Lys
[0450]	145 150 155 160
[0451]	Phe His Ile Ser Asp Asp Gly Gly Leu Thr Trp His Val Ser Arg Asn
[0452]	165 170 175
[0453]	Gly Val Thr Ser Pro Arg His Glu Ala Arg Pro Pro His Gln Gly Val
[0454]	180 185 190
[0455]	Arg Trp Phe Asn Asn Ala Val Glu Ala Thr Val Leu Glu Met Lys Asp
[0456]	195 200 205
[0457]	Gly Thr Leu Trp Ala Leu Ala Arg Thr Ser Gln Asp Gln Ala Trp Gln
[0458]	210 215 220
[0459]	Ala Phe Ser Lys Asp Tyr Gly Glu Thr Trp Ser Lys Pro Glu Pro Ser
[0460]	225 230 235 240
[0461]	Arg Phe Phe Gly Thr Leu Thr Met Asn Thr Leu Gly Arg Leu Asp Asp

[0462]		245		250		255
[0463]	Gly Thr Ile Val Ser Leu Trp Thr Asn Thr Met Ala Leu Pro Glu Asn					
[0464]		260		265		270
[0465]	Ala Thr Ala Gly Asn Gly Thr Trp Glu Asp Val Phe Thr Asn Arg Asp					
[0466]		275		280		285
[0467]	Ser His His Ile Ala Met Ser Gly Asp Glu Gly Lys Thr Trp Tyr Gly					
[0468]		290		295		300
[0469]	Phe Arg Glu Ile Ile Leu Asp Glu His Arg Asn His Pro Gly Tyr Ala					
[0470]	305		310		315	320
[0471]	Thr Leu Asp Gly Pro Glu Asp Arg Gly Lys His Gln Ser Glu Met Val					
[0472]		325		330		335
[0473]	Gln Leu Asp Lys Asn Arg Ile Leu Ile Ser Leu Gly Gln His Lys Asn					
[0474]		340		345		350
[0475]	His Arg Arg Leu Val Ile Val Asp Arg Arg Trp Val Gly Ala Lys Thr					
[0476]		355		360		365
[0477]	Arg Ala Thr Gln Thr Gly Lys Asp Leu Asp Ser Gln Trp Thr Ile His					
[0478]		370		375		380
[0479]	Thr Tyr Ile Pro Gln Lys Lys Gly His Cys Ser Tyr Asn Arg Lys Pro					
[0480]	385		390		395	400
[0481]	Ser Ala Glu Leu Val Gln Asp Pro Ser Gly Gly Thr Lys Lys Val Leu					
[0482]		405		410		415
[0483]	Gln Ile Lys Arg Leu Asp Asp Pro Glu Leu Val Asn Glu Lys Ser Asn					
[0484]		420		425		430
[0485]	Val Asp Tyr Arg Asn Gly Gly Ala Thr Trp Asn Phe Pro Asn Gly Thr					
[0486]		435		440		445
[0487]	Thr Gly Leu Val Lys Phe Arg Phe Arg Val Val Asp Gly Glu Gln Ala					
[0488]		450		455		460
[0489]	Asp Asp Ser Gly Leu Gln Val Ser Leu Thr Asp Arg Leu Phe Asn Ala					
[0490]	465		470		475	480
[0491]	Cys Asp Ser Thr Thr Lys Asp Tyr Ala Leu Phe Thr Phe Pro Ile Arg					
[0492]		485		490		495
[0493]	Leu Lys Pro Ala Pro His Leu Leu Leu Gly Met Lys Lys Val Pro Phe					
[0494]		500		505		510
[0495]	Thr Pro Gly Ala Trp His Glu Ile Ser Leu Leu Trp Gln Gly Gly Gln					
[0496]		515		520		525
[0497]	Ala Val Val Ser Leu Asp Gly Lys Lys Ala Gly Thr Leu Lys Met Ala					
[0498]		530		535		540
[0499]	Asn Lys Ser Pro Asn Gly Ala Ser Tyr Ile His Phe Ile Ser Thr Gly					
[0500]	545		550		555	560
[0501]	Ser Gln Pro Asp Ala Gly Ile Leu Leu Asp Thr Val Asn Ala Arg Val					
[0502]		565		570		575
[0503]	Lys					

[0504]	<210>	11
[0505]	<211>	589
[0506]	<212>	PRT
[0507]	<213>	Artificial Sequence
[0508]	<220>	
[0509]	<223>	Am1757 (N-terminal methionine and a C-terminal linker + His6 tag)
[0510]	<400>	11
[0511]	Met Gln Gln Lys Ala Ala Pro Val Pro Glu Pro Glu Val Val Ala Thr	
[0512]	1	5 10 15
[0513]	Pro Pro Ala Asp Ala Gly Arg Gly Leu Ile Arg Val Asp Ser Arg Glu	
[0514]	20 25 30	
[0515]	Ile Arg His Tyr Ser Gly Thr Arg Lys Glu Pro Asp Tyr Leu Val Ser	
[0516]	35 40 45	
[0517]	Arg Asp Asn Gly Lys Thr Trp Glu Met Lys Ala Ala Pro Ala Gly Tyr	
[0518]	50 55 60	
[0519]	Pro Pro Asn Tyr Gly Gly Ile Pro Lys Glu Ser Pro Ala Ile Val Arg	
[0520]	65 70 75 80	
[0521]	Asn Pro Leu Thr Arg Glu Phe Ile Arg Val Gln Pro Ile Gly Gly Phe	
[0522]	85 90 95	
[0523]	Val Phe Leu Ser Arg Gly Gly Leu Asp Gly Lys Trp Leu Ala Val Thr	
[0524]	100 105 110	
[0525]	Asn Asp Gly Lys Leu Glu Glu Asp Trp Lys Asp Pro Glu Lys Arg Lys	
[0526]	115 120 125	
[0527]	Asn Leu Lys Lys Leu Gly Gly Ile Met Arg Thr Pro Val Phe Val Asn	
[0528]	130 135 140	
[0529]	Lys Gly Arg Arg Val Ile Val Pro Phe His Asn Met Gly Gly Gly Thr	
[0530]	145 150 155 160	
[0531]	Lys Phe His Ile Ser Asp Asp Gly Gly Leu Thr Trp His Val Ser Arg	
[0532]	165 170 175	
[0533]	Asn Gly Val Thr Ser Pro Arg His Glu Ala Arg Pro Pro His Gln Gly	
[0534]	180 185 190	
[0535]	Val Arg Trp Phe Asn Asn Ala Val Glu Ala Thr Val Leu Glu Met Lys	
[0536]	195 200 205	
[0537]	Asp Gly Thr Leu Trp Ala Leu Ala Arg Thr Ser Gln Asp Gln Ala Trp	
[0538]	210 215 220	
[0539]	Gln Ala Phe Ser Lys Asp Tyr Gly Glu Thr Trp Ser Lys Pro Glu Pro	
[0540]	225 230 235 240	
[0541]	Ser Arg Phe Phe Gly Thr Leu Thr Met Asn Thr Leu Gly Arg Leu Asp	
[0542]	245 250 255	
[0543]	Asp Gly Thr Ile Val Ser Leu Trp Thr Asn Thr Met Ala Leu Pro Glu	
[0544]	260 265 270	
[0545]	Asn Ala Thr Ala Gly Asn Gly Thr Trp Glu Asp Val Phe Thr Asn Arg	

[0546]	275	280	285
[0547]	Asp Ser His His Ile Ala Met Ser Gly Asp Glu Gly Lys Thr Trp Tyr		
[0548]	290	295	300
[0549]	Gly Phe Arg Glu Ile Ile Leu Asp Glu His Arg Asn His Pro Gly Tyr		
[0550]	305	310	315
[0551]	Ala Thr Leu Asp Gly Pro Glu Asp Arg Gly Lys His Gln Ser Glu Met		
[0552]	325	330	335
[0553]	Val Gln Leu Asp Lys Asn Arg Ile Leu Ile Ser Leu Gly Gln His Lys		
[0554]	340	345	350
[0555]	Asn His Arg Arg Leu Val Ile Val Asp Arg Arg Trp Val Gly Ala Lys		
[0556]	355	360	365
[0557]	Thr Arg Ala Thr Gln Thr Gly Lys Asp Leu Asp Ser Gln Trp Thr Ile		
[0558]	370	375	380
[0559]	His Thr Tyr Ile Pro Gln Lys Lys Gly His Cys Ser Tyr Asn Arg Lys		
[0560]	385	390	395
[0561]	Pro Ser Ala Glu Leu Val Gln Asp Pro Ser Gly Gly Thr Lys Lys Val		
[0562]	405	410	415
[0563]	Leu Gln Ile Lys Arg Leu Asp Asp Pro Glu Leu Val Asn Glu Lys Ser		
[0564]	420	425	430
[0565]	Asn Val Asp Tyr Arg Asn Gly Gly Ala Thr Trp Asn Phe Pro Asn Gly		
[0566]	435	440	445
[0567]	Thr Thr Gly Leu Val Lys Phe Arg Phe Arg Val Val Asp Gly Glu Gln		
[0568]	450	455	460
[0569]	Ala Asp Asp Ser Gly Leu Gln Val Ser Leu Thr Asp Arg Leu Phe Asn		
[0570]	465	470	475
[0571]	Ala Cys Asp Ser Thr Thr Lys Asp Tyr Ala Leu Phe Thr Phe Pro Ile		
[0572]	485	490	495
[0573]	Arg Leu Lys Pro Ala Pro His Leu Leu Leu Gly Met Lys Lys Val Pro		
[0574]	500	505	510
[0575]	Phe Thr Pro Gly Ala Trp His Glu Ile Ser Leu Leu Trp Gln Gly Gly		
[0576]	515	520	525
[0577]	Gln Ala Val Val Ser Leu Asp Gly Lys Lys Ala Gly Thr Leu Lys Met		
[0578]	530	535	540
[0579]	Ala Asn Lys Ser Pro Asn Gly Ala Ser Tyr Ile His Phe Ile Ser Thr		
[0580]	545	550	555
[0581]	Gly Ser Gln Pro Asp Ala Gly Ile Leu Leu Asp Thr Val Asn Ala Arg		
[0582]	565	570	575
[0583]	Val Lys Gly Ser Gly Leu Glu His His His His His His		
[0584]	580	585	
[0585]	<210> 12		
[0586]	<211> 419		
[0587]	<212> PRT		

[0588]	<213> Akkermansia muciniphila															
[0589]	<400> 12															
[0590]	Met	Thr	Trp	Leu	Leu	Cys	Gly	Arg	Gly	Lys	Trp	Asn	Lys	Val	Lys	Arg
[0591]	1			5						10					15	
[0592]	Met	Met	Asn	Ser	Val	Phe	Lys	Cys	Leu	Met	Ser	Ala	Val	Cys	Ala	Val
[0593]				20					25					30		
[0594]	Ala	Leu	Pro	Ala	Phe	Gly	Gln	Glu	Glu	Lys	Thr	Gly	Phe	Pro	Thr	Asp
[0595]			35					40					45			
[0596]	Arg	Ala	Val	Thr	Val	Phe	Ser	Ala	Gly	Glu	Gly	Asn	Pro	Tyr	Ala	Ser
[0597]		50					55					60				
[0598]	Ile	Arg	Ile	Pro	Ala	Leu	Leu	Ser	Ile	Gly	Lys	Gly	Gln	Leu	Leu	Ala
[0599]	65				70					75					80	
[0600]	Phe	Ala	Glu	Gly	Arg	Tyr	Lys	Asn	Thr	Asp	Gln	Gly	Glu	Asn	Asp	Ile
[0601]				85					90					95		
[0602]	Ile	Met	Ser	Val	Ser	Lys	Asn	Gly	Gly	Lys	Thr	Trp	Ser	Arg	Pro	Arg
[0603]				100					105					110		
[0604]	Ala	Ile	Ala	Lys	Ala	His	Gly	Ala	Thr	Phe	Asn	Asn	Pro	Cys	Pro	Val
[0605]			115				120						125			
[0606]	Tyr	Asp	Ala	Lys	Thr	Arg	Thr	Val	Thr	Val	Val	Phe	Gln	Arg	Tyr	Pro
[0607]		130					135					140				
[0608]	Ala	Gly	Val	Lys	Glu	Arg	Gln	Pro	Asn	Ile	Pro	Asp	Gly	Trp	Asp	Asp
[0609]	145				150				155						160	
[0610]	Glu	Lys	Cys	Ile	Arg	Asn	Phe	Met	Ile	Gln	Ser	Arg	Asn	Gly	Gly	Ser
[0611]				165					170					175		
[0612]	Ser	Trp	Thr	Lys	Pro	Gln	Glu	Ile	Thr	Lys	Thr	Thr	Lys	Arg	Pro	Ser
[0613]				180					185					190		
[0614]	Gly	Val	Asp	Ile	Met	Ala	Ser	Gly	Pro	Asn	Ala	Gly	Thr	Gln	Leu	Lys
[0615]		195					200						205			
[0616]	Ser	Gly	Ala	His	Lys	Gly	Arg	Leu	Val	Ile	Pro	Met	Asn	Glu	Gly	Pro
[0617]		210					215						220			
[0618]	Phe	Gly	Lys	Trp	Val	Ile	Ser	Cys	Ile	Tyr	Ser	Asp	Asp	Gly	Gly	Lys
[0619]	225				230					235					240	
[0620]	Ser	Trp	Lys	Leu	Gly	Gln	Pro	Thr	Ala	Asn	Met	Lys	Gly	Met	Val	Asn
[0621]				245					250					255		
[0622]	Glu	Thr	Ser	Ile	Ala	Glu	Thr	Asp	Asn	Gly	Gly	Val	Val	Met	Val	Ala
[0623]				260					265					270		
[0624]	Arg	His	Trp	Gly	Ala	Gly	Asn	Cys	Arg	Arg	Ile	Ala	Trp	Ser	Gln	Asp
[0625]		275						280					285			
[0626]	Gly	Gly	Glu	Thr	Trp	Gly	Gln	Val	Glu	Asp	Ala	Pro	Glu	Leu	Phe	Cys
[0627]		290					295					300				
[0628]	Asp	Ser	Thr	Gln	Asn	Ser	Leu	Met	Thr	Tyr	Ser	Leu	Ser	Asp	Gln	Pro
[0629]	305				310					315					320	

[0630]	Ala Tyr Gly Gly Lys Ser Arg Ile Leu Phe Ser Gly Pro Ser Ala Gly		
[0631]		325	330 335
[0632]	Arg Arg Ile Lys Gly Gln Val Ala Met Ser Tyr Asp Asn Gly Lys Thr		
[0633]		340	345 350
[0634]	Trp Pro Val Lys Lys Leu Leu Gly Glu Gly Gly Phe Ala Tyr Ser Ser		
[0635]		355	360 365
[0636]	Leu Ala Met Val Glu Pro Gly Ile Val Gly Val Leu Tyr Glu Glu Asn		
[0637]		370	375 380
[0638]	Gln Glu His Ile Lys Lys Leu Lys Phe Val Pro Ile Thr Met Glu Trp		
[0639]		385	390 395 400
[0640]	Leu Thr Asp Gly Glu Asp Thr Gly Leu Ala Pro Gly Lys Lys Ala Pro		
[0641]		405	410 415
[0642]	Val Leu Lys		
[0643]	<210> 13		
[0644]	<211> 381		
[0645]	<212> PRT		
[0646]	<213> Akkermansia muciniphila		
[0647]	<400> 13		
[0648]	Gln Glu Glu Lys Thr Gly Phe Pro Thr Asp Arg Ala Val Thr Val Phe		
[0649]	1	5	10 15
[0650]	Ser Ala Gly Glu Gly Asn Pro Tyr Ala Ser Ile Arg Ile Pro Ala Leu		
[0651]		20	25 30
[0652]	Leu Ser Ile Gly Lys Gly Gln Leu Leu Ala Phe Ala Glu Gly Arg Tyr		
[0653]		35	40 45
[0654]	Lys Asn Thr Asp Gln Gly Glu Asn Asp Ile Ile Met Ser Val Ser Lys		
[0655]		50	55 60
[0656]	Asn Gly Gly Lys Thr Trp Ser Arg Pro Arg Ala Ile Ala Lys Ala His		
[0657]	65	70	75 80
[0658]	Gly Ala Thr Phe Asn Asn Pro Cys Pro Val Tyr Asp Ala Lys Thr Arg		
[0659]		85	90 95
[0660]	Thr Val Thr Val Val Phe Gln Arg Tyr Pro Ala Gly Val Lys Glu Arg		
[0661]		100	105 110
[0662]	Gln Pro Asn Ile Pro Asp Gly Trp Asp Asp Glu Lys Cys Ile Arg Asn		
[0663]		115	120 125
[0664]	Phe Met Ile Gln Ser Arg Asn Gly Gly Ser Ser Trp Thr Lys Pro Gln		
[0665]		130	135 140
[0666]	Glu Ile Thr Lys Thr Thr Lys Arg Pro Ser Gly Val Asp Ile Met Ala		
[0667]	145	150	155 160
[0668]	Ser Gly Pro Asn Ala Gly Thr Gln Leu Lys Ser Gly Ala His Lys Gly		
[0669]		165	170 175
[0670]	Arg Leu Val Ile Pro Met Asn Glu Gly Pro Phe Gly Lys Trp Val Ile		
[0671]		180	185 190

[0672]	Ser Cys Ile Tyr Ser Asp Asp Gly Gly Lys Ser Trp Lys Leu Gly Gln		
[0673]	195	200	205
[0674]	Pro Thr Ala Asn Met Lys Gly Met Val Asn Glu Thr Ser Ile Ala Glu		
[0675]	210	215	220
[0676]	Thr Asp Asn Gly Gly Val Val Met Val Ala Arg His Trp Gly Ala Gly		
[0677]	225	230	235 240
[0678]	Asn Cys Arg Arg Ile Ala Trp Ser Gln Asp Gly Gly Glu Thr Trp Gly		
[0679]	245	250	255
[0680]	Gln Val Glu Asp Ala Pro Glu Leu Phe Cys Asp Ser Thr Gln Asn Ser		
[0681]	260	265	270
[0682]	Leu Met Thr Tyr Ser Leu Ser Asp Gln Pro Ala Tyr Gly Gly Lys Ser		
[0683]	275	280	285
[0684]	Arg Ile Leu Phe Ser Gly Pro Ser Ala Gly Arg Arg Ile Lys Gly Gln		
[0685]	290	295	300
[0686]	Val Ala Met Ser Tyr Asp Asn Gly Lys Thr Trp Pro Val Lys Lys Leu		
[0687]	305	310	315 320
[0688]	Leu Gly Glu Gly Gly Phe Ala Tyr Ser Ser Leu Ala Met Val Glu Pro		
[0689]	325	330	335
[0690]	Gly Ile Val Gly Val Leu Tyr Glu Glu Asn Gln Glu His Ile Lys Lys		
[0691]	340	345	350
[0692]	Leu Lys Phe Val Pro Ile Thr Met Glu Trp Leu Thr Asp Gly Glu Asp		
[0693]	355	360	365
[0694]	Thr Gly Leu Ala Pro Gly Lys Lys Ala Pro Val Leu Lys		
[0695]	370	375	380
[0696]	<210> 14		
[0697]	<211> 393		
[0698]	<212> PRT		
[0699]	<213> Artificial Sequence		
[0700]	<220>		
[0701]	<223> Am0707 (including N-terminal methionine and a C-terminal linker +		
[0702]	His6 tag)		
[0703]	<400> 14		
[0704]	Met Gln Glu Glu Lys Thr Gly Phe Pro Thr Asp Arg Ala Val Thr Val		
[0705]	1	5	10 15
[0706]	Phe Ser Ala Gly Glu Gly Asn Pro Tyr Ala Ser Ile Arg Ile Pro Ala		
[0707]	20	25	30
[0708]	Leu Leu Ser Ile Gly Lys Gly Gln Leu Leu Ala Phe Ala Glu Gly Arg		
[0709]	35	40	45
[0710]	Tyr Lys Asn Thr Asp Gln Gly Glu Asn Asp Ile Ile Met Ser Val Ser		
[0711]	50	55	60
[0712]	Lys Asn Gly Gly Lys Thr Trp Ser Arg Pro Arg Ala Ile Ala Lys Ala		
[0713]	65	70	75 80

[0714]	His Gly Ala Thr Phe Asn Asn Pro Cys Pro Val Tyr Asp Ala Lys Thr		
[0715]		85	90 95
[0716]	Arg Thr Val Thr Val Val Phe Gln Arg Tyr Pro Ala Gly Val Lys Glu		
[0717]		100	105 110
[0718]	Arg Gln Pro Asn Ile Pro Asp Gly Trp Asp Asp Glu Lys Cys Ile Arg		
[0719]		115	120 125
[0720]	Asn Phe Met Ile Gln Ser Arg Asn Gly Gly Ser Ser Trp Thr Lys Pro		
[0721]		130	135 140
[0722]	Gln Glu Ile Thr Lys Thr Thr Lys Arg Pro Ser Gly Val Asp Ile Met		
[0723]		145	150 155 160
[0724]	Ala Ser Gly Pro Asn Ala Gly Thr Gln Leu Lys Ser Gly Ala His Lys		
[0725]		165	170 175
[0726]	Gly Arg Leu Val Ile Pro Met Asn Glu Gly Pro Phe Gly Lys Trp Val		
[0727]		180	185 190
[0728]	Ile Ser Cys Ile Tyr Ser Asp Asp Gly Gly Lys Ser Trp Lys Leu Gly		
[0729]		195	200 205
[0730]	Gln Pro Thr Ala Asn Met Lys Gly Met Val Asn Glu Thr Ser Ile Ala		
[0731]		210	215 220
[0732]	Glu Thr Asp Asn Gly Gly Val Val Met Val Ala Arg His Trp Gly Ala		
[0733]		225	230 235 240
[0734]	Gly Asn Cys Arg Arg Ile Ala Trp Ser Gln Asp Gly Gly Glu Thr Trp		
[0735]		245	250 255
[0736]	Gly Gln Val Glu Asp Ala Pro Glu Leu Phe Cys Asp Ser Thr Gln Asn		
[0737]		260	265 270
[0738]	Ser Leu Met Thr Tyr Ser Leu Ser Asp Gln Pro Ala Tyr Gly Gly Lys		
[0739]		275	280 285
[0740]	Ser Arg Ile Leu Phe Ser Gly Pro Ser Ala Gly Arg Arg Ile Lys Gly		
[0741]		290	295 300
[0742]	Gln Val Ala Met Ser Tyr Asp Asn Gly Lys Thr Trp Pro Val Lys Lys		
[0743]		305	310 315 320
[0744]	Leu Leu Gly Glu Gly Gly Phe Ala Tyr Ser Ser Leu Ala Met Val Glu		
[0745]		325	330 335
[0746]	Pro Gly Ile Val Gly Val Leu Tyr Glu Glu Asn Gln Glu His Ile Lys		
[0747]		340	345 350
[0748]	Lys Leu Lys Phe Val Pro Ile Thr Met Glu Trp Leu Thr Asp Gly Glu		
[0749]		355	360 365
[0750]	Asp Thr Gly Leu Ala Pro Gly Lys Lys Ala Pro Val Leu Lys Gly Ser		
[0751]		370	375 380
[0752]	Gly Leu Glu His His His His His His		
[0753]		385	390
[0754]	<210> 15		
[0755]	<211> 2133		

[0756] <212> PRT
 [0757] <213> Streptococcus oralis
 [0758] <400> 15
 [0759] Met Asp Lys Arg Phe Phe Glu Lys Arg Cys Lys Phe Ser Ile Arg Lys
 [0760] 1 5 10 15
 [0761] Phe Thr Leu Gly Val Ala Ser Val Met Ile Gly Ala Thr Phe Phe Ala
 [0762] 20 25 30
 [0763] Ala Ser Pro Val Leu Ala Asp Gln Ala Arg Val Gly Ser Thr Asp Asn
 [0764] 35 40 45
 [0765] Leu Pro Ser Glu Leu Ala Asp Leu Asp Lys Lys Ala Ser Asp Glu Gly
 [0766] 50 55 60
 [0767] His Asp Phe Asp Lys Glu Ala Ala Ala Gln Asn Pro Gly Ser Ala Glu
 [0768] 65 70 75 80
 [0769] Thr Thr Glu Gly Pro Gln Thr Glu Glu Glu Leu Leu Ala Gln Glu Lys
 [0770] 85 90 95
 [0771] Glu Lys Ser Glu Lys Pro Ser Asn Leu Pro Lys Glu Leu Glu Asp Lys
 [0772] 100 105 110
 [0773] Leu Glu Lys Ala Glu Asp Asn Gly Arg Glu Val Asp Lys Asp Gln Leu
 [0774] 115 120 125
 [0775] Ala Gln Asp Thr Gly Lys Leu Val Pro Glu Asp Val Ala Lys Thr Thr
 [0776] 130 135 140
 [0777] Asn Gly Glu Leu Asn Tyr Gly Ala Thr Val Lys Ile Lys Thr Pro Ser
 [0778] 145 150 155 160
 [0779] Gly Glu Gly Ser Gly Ile Val Val Ala Lys Asp Leu Val Leu Thr Val
 [0780] 165 170 175
 [0781] Ser His Asn Phe Ile Lys Asp Ser Gln Glu Gly Asn Ile Arg Lys Val
 [0782] 180 185 190
 [0783] Val Asp Asn Asp Gln Gly Asp Gly Asp Ile Tyr Ser Ile Ser Tyr Pro
 [0784] 195 200 205
 [0785] Gly Leu Pro Asp Val Lys Phe Ser Lys Lys Asp Ile Ile His Trp Asp
 [0786] 210 215 220
 [0787] Arg Glu Gly Tyr Leu Lys Gly Phe Lys Asn Asp Leu Ala Leu Val Arg
 [0788] 225 230 235 240
 [0789] Leu Arg Thr Val Leu Glu Asn Thr Pro Val Glu Val Thr Lys Lys Pro
 [0790] 245 250 255
 [0791] Val Val Lys Lys Ile Gly Asp Lys Leu His Val Phe Gly Tyr Pro Glu
 [0792] 260 265 270
 [0793] Gly Lys Leu Asn Pro Ile Val Asn Thr Thr Val Asp Phe Ala Glu Pro
 [0794] 275 280 285
 [0795] Tyr Gly Glu Gly Val Gln Gly Ile Gly Tyr Gln Gly Gly Lys Pro Gly
 [0796] 290 295 300
 [0797] Ala Ser Gly Gly Gly Ile Phe Asp Thr Glu Gly Lys Leu Val Gly Val

[0798]	305	310	315	320
[0799]	His Gln Asn Gly Val Val Gly Lys Arg Ser Gly Gly Ile Leu Phe Ser			
[0800]		325	330	335
[0801]	Pro Ala Gln Leu Lys Trp Ile Gln Asp His Met Gln Gly Ile Ser Ser			
[0802]		340	345	350
[0803]	Val Lys Pro Ala Asp Leu Glu Glu Lys Glu Lys Pro Ala Glu Glu Lys			
[0804]		355	360	365
[0805]	Pro Lys Glu Asp Lys Pro Ala Ala Ala Lys Pro Glu Thr Pro Lys Ala			
[0806]		370	375	380
[0807]	Val Thr Pro Glu Trp Gln Thr Val Ala Asn Lys Glu Gln Gln Gly Thr			
[0808]	385	390	395	400
[0809]	Val Thr Ile Arg Glu Glu Lys Gly Val Arg Tyr Asn Gln Leu Ser Ser			
[0810]		405	410	415
[0811]	Thr Ala Gln Asn Asp Asn Asp Gly Lys Pro Ala Leu Phe Glu Lys Gln			
[0812]		420	425	430
[0813]	Gly Leu Thr Val Asp Ala Asn Gly Asn Ala Thr Val Asp Leu Thr Phe			
[0814]		435	440	445
[0815]	Lys Asp Asp Ser Glu Lys Gly Lys Ser Arg Phe Gly Val Phe Leu Lys			
[0816]		450	455	460
[0817]	Phe Lys Asp Thr Lys Asn Asn Val Phe Val Gly Tyr Asp Gln Gly Gly			
[0818]	465	470	475	480
[0819]	Trp Phe Trp Glu Tyr Lys Thr Pro Gly Asn Ser Thr Trp Tyr Lys Gly			
[0820]		485	490	495
[0821]	Asn Arg Val Ala Ala Pro Glu Pro Gly Ser Val Asn Arg Leu Ser Ile			
[0822]		500	505	510
[0823]	Thr Leu Lys Ser Asp Gly Gln Leu Asn Ala Ser Asn Asn Asp Val Asn			
[0824]		515	520	525
[0825]	Leu Phe Asp Thr Val Thr Leu Pro Gly Ala Val Asn Glu Asn Leu Lys			
[0826]		530	535	540
[0827]	Asn Glu Lys Lys Ile Leu Leu Lys Ala Gly Thr Tyr Ser Asn Asp Arg			
[0828]	545	550	555	560
[0829]	Thr Val Val Ser Val Lys Thr Asp Asn Gln Glu Gly Val Lys Ala Asp			
[0830]		565	570	575
[0831]	Asp Thr Pro Ala Gln Lys Glu Thr Gly Pro Ala Val Asp Asp Ser Lys			
[0832]		580	585	590
[0833]	Val Thr Tyr Asp Thr Ile Gln Ser Lys Val Leu Lys Ala Val Ile Asp			
[0834]		595	600	605
[0835]	Gln Ala Phe Pro Arg Val Lys Glu Tyr Thr Leu Asn Gly His Thr Leu			
[0836]		610	615	620
[0837]	Pro Gly Gln Val Gln Gln Phe Asn Gln Val Phe Ile Asn Asn His Arg			
[0838]	625	630	635	640
[0839]	Ile Thr Pro Glu Val Thr Tyr Lys Lys Ile Asn Glu Thr Thr Ala Glu			

[0840]	645							650							655			
[0841]	Tyr	Leu	Met	Lys	Leu	Arg	Asp	Asp	Ala	His	Leu	Ile	Asn	Ala	Glu	Met		
[0842]	660							665							670			
[0843]	Thr	Val	Arg	Leu	Gln	Val	Val	Asp	Asn	Gln	Leu	His	Phe	Asp	Val	Thr		
[0844]	675							680							685			
[0845]	Lys	Ile	Val	Asn	His	Asn	Gln	Val	Thr	Pro	Gly	Gln	Lys	Ile	Asp	Asp		
[0846]	690							695							700			
[0847]	Glu	Arg	Lys	Leu	Leu	Ser	Thr	Ile	Ser	Phe	Leu	Gly	Asn	Ala	Leu	Val		
[0848]	705	710							715							720		
[0849]	Ser	Val	Ser	Ser	Asp	Gln	Ala	Gly	Ala	Lys	Phe	Asp	Gly	Ala	Thr	Met		
[0850]	725							730							735			
[0851]	Ser	Asn	Asn	Thr	His	Val	Ser	Gly	Asp	Asp	His	Ile	Asp	Val	Thr	Asn		
[0852]	740							745							750			
[0853]	Pro	Met	Lys	Asp	Leu	Ala	Lys	Gly	Tyr	Met	Tyr	Gly	Phe	Val	Ser	Thr		
[0854]	755							760							765			
[0855]	Asp	Lys	Leu	Ala	Ala	Gly	Val	Trp	Ser	Asn	Ser	Gln	Asn	Ser	Tyr	Gly		
[0856]	770							775							780			
[0857]	Gly	Gly	Ser	Asn	Asp	Trp	Thr	Arg	Leu	Thr	Ala	Tyr	Lys	Glu	Thr	Val		
[0858]	785	790							795							800		
[0859]	Gly	Asn	Ala	Asn	Tyr	Val	Gly	Ile	His	Ser	Ser	Glu	Trp	Gln	Trp	Glu		
[0860]	805							810							815			
[0861]	Lys	Ala	Tyr	Lys	Gly	Ile	Val	Phe	Pro	Glu	Tyr	Thr	Lys	Glu	Leu	Pro		
[0862]	820							825							830			
[0863]	Ser	Ala	Lys	Val	Val	Ile	Thr	Glu	Asp	Ala	Asn	Ala	Asp	Asn	Lys	Val		
[0864]	835							840							845			
[0865]	Asp	Trp	Gln	Asp	Gly	Ala	Ile	Ala	Tyr	Arg	Ser	Ile	Met	Asn	Asn	Pro		
[0866]	850							855							860			
[0867]	Gln	Gly	Trp	Glu	Lys	Val	Lys	Asp	Ile	Thr	Ala	Tyr	Arg	Ile	Ala	Met		
[0868]	865	870							875							880		
[0869]	Asn	Phe	Gly	Ser	Gln	Ala	Gln	Asn	Pro	Phe	Leu	Met	Thr	Leu	Asp	Gly		
[0870]	885							890							895			
[0871]	Ile	Lys	Lys	Ile	Asn	Leu	His	Thr	Asp	Gly	Leu	Gly	Gln	Gly	Val	Leu		
[0872]	900							905							910			
[0873]	Leu	Lys	Gly	Tyr	Gly	Ser	Glu	Gly	His	Asp	Ser	Gly	His	Leu	Asn	Tyr		
[0874]	915							920							925			
[0875]	Ala	Asp	Ile	Gly	Lys	Arg	Ile	Gly	Gly	Val	Glu	Asp	Phe	Lys	Thr	Leu		
[0876]	930							935							940			
[0877]	Ile	Glu	Lys	Ala	Lys	Lys	Tyr	Gly	Ala	His	Leu	Gly	Ile	His	Val	Asn		
[0878]	945	950							955							960		
[0879]	Ala	Ser	Glu	Thr	Tyr	Pro	Glu	Ser	Lys	Tyr	Phe	Asn	Glu	Asn	Ile	Leu		
[0880]	965							970							975			
[0881]	Arg	Lys	Asn	Pro	Asp	Gly	Ser	Tyr	Ser	Tyr	Gly	Trp	Asn	Trp	Leu	Asn		

[0882]	980	985	990
[0883]	Gln Gly Ile Asn Ile Asp Ala Ala Tyr Asp Leu Ala His Gly Arg Leu		
[0884]	995	1000	1005
[0885]	Ala Arg Trp Glu Asp Leu Lys Lys Lys Leu Gly Glu Gly Leu Asp		
[0886]	1010	1015	1020
[0887]	Phe Ile Tyr Val Asp Val Trp Gly Asn Gly Gln Ser Gly Asp Asn		
[0888]	1025	1030	1035
[0889]	Gly Ala Trp Ala Thr His Val Leu Ala Lys Glu Ile Asn Lys Gln		
[0890]	1040	1045	1050
[0891]	Gly Trp Arg Phe Ala Ile Glu Trp Gly His Gly Gly Glu Tyr Asp		
[0892]	1055	1060	1065
[0893]	Ser Thr Phe Gln His Trp Ala Ala Asp Leu Thr Tyr Gly Gly Tyr		
[0894]	1070	1075	1080
[0895]	Thr Asn Lys Gly Ile Asn Ser Ala Ile Thr Arg Phe Ile Arg Asn		
[0896]	1085	1090	1095
[0897]	His Gln Lys Asp Ser Trp Val Gly Asp Tyr Arg Ser Tyr Gly Gly		
[0898]	1100	1105	1110
[0899]	Ala Ala Asn Tyr Pro Leu Leu Gly Gly Tyr Ser Met Lys Asp Phe		
[0900]	1115	1120	1125
[0901]	Glu Gly Trp Gln Gly Arg Ser Asp Tyr Asn Gly Tyr Val Thr Asn		
[0902]	1130	1135	1140
[0903]	Leu Phe Ala His Asp Val Met Thr Lys Tyr Phe Gln His Phe Thr		
[0904]	1145	1150	1155
[0905]	Val Ser Lys Trp Glu Asn Gly Thr Pro Val Thr Met Thr Asp Asn		
[0906]	1160	1165	1170
[0907]	Gly Ser Thr Tyr Lys Trp Thr Pro Glu Met Lys Val Glu Leu Val		
[0908]	1175	1180	1185
[0909]	Asp Ala Ala Gly Asn Lys Val Val Val Thr Arg Lys Ser Asn Asp		
[0910]	1190	1195	1200
[0911]	Val Asn Ser Pro Gln Tyr Arg Glu Arg Thr Val Thr Leu Asn Gly		
[0912]	1205	1210	1215
[0913]	Arg Val Ile Gln Asp Gly Ser Ala Tyr Leu Thr Pro Trp Asn Trp		
[0914]	1220	1225	1230
[0915]	Asp Ala Asn Gly Lys Lys Leu Pro Thr Glu Lys Glu Lys Met Tyr		
[0916]	1235	1240	1245
[0917]	Tyr Phe Asn Thr Gln Ala Gly Ala Thr Thr Trp Thr Leu Pro Ser		
[0918]	1250	1255	1260
[0919]	Asp Trp Ala Asn Ser Lys Val Tyr Leu Tyr Lys Leu Thr Asp Gln		
[0920]	1265	1270	1275
[0921]	Gly Lys Thr Glu Glu Gln Glu Leu Thr Val Thr Asp Gly Lys Ile		
[0922]	1280	1285	1290
[0923]	Thr Leu Asp Leu Leu Ala Asn Gln Pro Tyr Val Leu Tyr Arg Ser		

[0924]	1295	1300	1305
[0925]	Lys Gln Thr Asn Pro Glu Met Ser Trp Ser Glu Gly Met His Ile		
[0926]	1310	1315	1320
[0927]	Tyr Asp Gln Gly Phe Asn Ser Gly Thr Leu Lys His Trp Thr Ile		
[0928]	1325	1330	1335
[0929]	Ser Gly Asp Ala Ser Lys Ala Glu Ile Val Lys Ser Gln Gly Ala		
[0930]	1340	1345	1350
[0931]	Asn Glu Met Leu Arg Ile Gln Gly Asn Lys Ser Lys Val Ser Leu		
[0932]	1355	1360	1365
[0933]	Thr Gln Lys Leu Thr Gly Leu Lys Pro Asn Thr Lys Tyr Ala Val		
[0934]	1370	1375	1380
[0935]	Tyr Val Gly Val Asp Asn Arg Ser Asn Ala Lys Ala Ser Ile Thr		
[0936]	1385	1390	1395
[0937]	Val Asn Thr Gly Glu Lys Glu Val Thr Thr Tyr Thr Asn Lys Ser		
[0938]	1400	1405	1410
[0939]	Leu Ala Leu Asn Tyr Ile Lys Ala Tyr Ala His Asn Asn Arg Arg		
[0940]	1415	1420	1425
[0941]	Glu Asn Ala Thr Val Asp Asp Thr Ser Tyr Phe Gln Asn Met Tyr		
[0942]	1430	1435	1440
[0943]	Ala Phe Phe Thr Thr Gly Ser Asp Val Ser Asn Val Thr Leu Thr		
[0944]	1445	1450	1455
[0945]	Leu Ser Arg Glu Ala Gly Asp Glu Ala Thr Tyr Phe Asp Glu Ile		
[0946]	1460	1465	1470
[0947]	Arg Thr Phe Glu Asn Asn Ser Ser Met Tyr Gly Asp Lys His Asp		
[0948]	1475	1480	1485
[0949]	Thr Gly Gln Gly Thr Phe Lys Gln Asp Phe Glu Asn Val Ala Gln		
[0950]	1490	1495	1500
[0951]	Gly Ile Phe Pro Phe Val Val Gly Gly Val Glu Gly Val Glu Asp		
[0952]	1505	1510	1515
[0953]	Asn Arg Thr His Leu Ser Glu Lys His Asp Pro Tyr Thr Gln Arg		
[0954]	1520	1525	1530
[0955]	Gly Trp Asn Gly Lys Lys Val Asp Asp Val Ile Glu Gly Asn Trp		
[0956]	1535	1540	1545
[0957]	Ser Leu Lys Thr Asn Gly Leu Val Ser Arg Arg Asn Leu Val Tyr		
[0958]	1550	1555	1560
[0959]	Gln Thr Ile Pro Gln Asn Phe Arg Phe Glu Ala Gly Lys Thr Tyr		
[0960]	1565	1570	1575
[0961]	Arg Val Thr Phe Glu Tyr Glu Ala Gly Ser Asp Asn Thr Tyr Ala		
[0962]	1580	1585	1590
[0963]	Phe Val Val Gly Lys Gly Glu Phe Gln Ser Gly Arg Arg Gly Thr		
[0964]	1595	1600	1605
[0965]	Gln Ala Ser Asn Leu Glu Met His Glu Leu Pro Asn Thr Trp Thr		

[0966]	1610	1615	1620
[0967]	Asp Ser Lys Lys Ala Lys Lys Val Thr Phe Leu Val Thr Gly Ala		
[0968]	1625	1630	1635
[0969]	Glu Thr Gly Asp Thr Trp Val Gly Ile Tyr Ser Thr Gly Asn Ala		
[0970]	1640	1645	1650
[0971]	Ser Asn Thr Arg Gly Asp Ala Gly Gly Asn Ala Asn Phe Arg Gly		
[0972]	1655	1660	1665
[0973]	Tyr Asn Asp Phe Met Met Asp Asn Leu Gln Ile Glu Glu Ile Thr		
[0974]	1670	1675	1680
[0975]	Leu Thr Gly Lys Met Leu Thr Glu Asn Ala Leu Lys Asn Tyr Leu		
[0976]	1685	1690	1695
[0977]	Pro Thr Val Ala Met Thr Asn Tyr Thr Lys Glu Ser Met Asp Ala		
[0978]	1700	1705	1710
[0979]	Leu Lys Glu Ala Val Phe Asn Leu Ser Gln Ala Asp Asp Asp Ile		
[0980]	1715	1720	1725
[0981]	Ser Val Glu Glu Ala Arg Ala Glu Ile Ala Lys Ile Glu Ala Leu		
[0982]	1730	1735	1740
[0983]	Lys Asn Ala Leu Val Gln Lys Lys Thr Ala Leu Val Ala Glu Asp		
[0984]	1745	1750	1755
[0985]	Phe Glu Ser Leu Asp Ala Pro Ala Gln Pro Gly Glu Gly Leu Glu		
[0986]	1760	1765	1770
[0987]	Asn Ala Phe Asp Gly Asn Val Ser Ser Leu Trp His Thr Ser Trp		
[0988]	1775	1780	1785
[0989]	Asn Gly Gly Asp Val Gly Lys Pro Ala Thr Met Val Leu Lys Glu		
[0990]	1790	1795	1800
[0991]	Pro Thr Glu Ile Thr Gly Leu Arg Tyr Val Pro Arg Ala Ser Asp		
[0992]	1805	1810	1815
[0993]	Ser Asn Gly Asn Leu Arg Asp Val Lys Leu Val Val Thr Asp Glu		
[0994]	1820	1825	1830
[0995]	Ser Gly Lys Glu His Thr Phe Asn Val Thr Asp Trp Pro Asn Asn		
[0996]	1835	1840	1845
[0997]	Asn Lys Pro Lys Asp Ile Asp Phe Gly Lys Thr Ile Lys Ala Lys		
[0998]	1850	1855	1860
[0999]	Lys Ile Val Leu Thr Gly Thr Lys Thr Tyr Gly Asp Gly Gly Asp		
[1000]	1865	1870	1875
[1001]	Lys Tyr Gln Ser Ala Ala Glu Leu Ile Phe Thr Arg Pro Gln Val		
[1002]	1880	1885	1890
[1003]	Ala Glu Thr Pro Leu Asp Leu Ser Gly Tyr Glu Ala Ala Leu Ala		
[1004]	1895	1900	1905
[1005]	Lys Ala Gln Lys Leu Thr Asp Lys Asp Asn Gln Glu Glu Val Ala		
[1006]	1910	1915	1920
[1007]	Ser Val Gln Ala Ser Met Lys Tyr Ala Thr Asp Asn His Leu Leu		

[1008]	1925	1930	1935
[1009]	Thr Glu Arg Met Val Ala Tyr Phe Ala Asp Tyr Leu Asn Gln Leu		
[1010]	1940	1945	1950
[1011]	Lys Asp Ser Ala Thr Lys Pro Asp Ala Pro Thr Ser Ser Lys Gly		
[1012]	1955	1960	1965
[1013]	Glu Glu Gln Pro Pro Val Leu Asp Val Pro Glu Phe Lys Gly Gly		
[1014]	1970	1975	1980
[1015]	Val Asn Ala Thr Glu Ala Ala Val His Glu Val Pro Glu Phe Lys		
[1016]	1985	1990	1995
[1017]	Gly Gly Val Asn Ala Val Gln Ala Leu Val His Glu Leu Pro Glu		
[1018]	2000	2005	2010
[1019]	Tyr Lys Gly Gly Ala Asn Ala Val Leu Ala Ala Ala Asn Glu Val		
[1020]	2015	2020	2025
[1021]	Pro Glu Tyr Lys Gly Gly Ala Asn Ala Val Glu Ala Leu Val Asn		
[1022]	2030	2035	2040
[1023]	Glu Lys Pro Ala Tyr Thr Gly Val Leu Ala Thr Ala Gly Asp Gln		
[1024]	2045	2050	2055
[1025]	Ala Ala Pro Thr Val Glu Lys Pro Glu Tyr Pro Leu Thr Pro Ser		
[1026]	2060	2065	2070
[1027]	Pro Val Ala Asp Thr Lys Thr Pro Gly Ala Lys Asp Glu Glu Lys		
[1028]	2075	2080	2085
[1029]	Leu Pro Ala Thr Gly Glu His Ser Ser Glu Val Ala Leu Phe Leu		
[1030]	2090	2095	2100
[1031]	Ala Ser Val Ser Ile Ala Leu Ser Ala Ala Val Leu Ala Thr Lys		
[1032]	2105	2110	2115
[1033]	Arg Lys Glu Glu Gly Ser Gly Leu Glu His His His His His His		
[1034]	2120	2125	2130
[1035]	<210> 16		
[1036]	<211> 22		
[1037]	<212> DNA		
[1038]	<213> Artificial Sequence		
[1039]	<220>		
[1040]	<223> E206A forward primer		
[1041]	<400> 16		
[1042]	atggcgcacg cgctgggccca cg 22		
[1043]	<210> 17		
[1044]	<211> 19		
[1045]	<212> DNA		
[1046]	<213> Artificial Sequence		
[1047]	<220>		
[1048]	<223> E206A reverse primer		
[1049]	<400> 17		

[1050] gccaccgtac catttcgtc 19
 [1051] <210> 18
 [1052] <211> 166
 [1053] <212> PRT
 [1054] <213> Artificial Sequence
 [1055] <220>
 [1056] <223> EP0
 [1057] <400> 18
 [1058] Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val Leu Glu Arg Tyr Leu
 [1059] 1 5 10 15
 [1060] Leu Glu Ala Lys Glu Ala Glu Asp Ile Thr Thr Gly Cys Ala Glu His
 [1061] 20 25 30
 [1062] Cys Ser Leu Asp Glu Asn Ile Thr Val Pro Asp Thr Lys Val Asp Phe
 [1063] 35 40 45
 [1064] Tyr Ala Trp Lys Arg Met Glu Val Gly Gln Gln Ala Val Glu Val Trp
 [1065] 50 55 60
 [1066] Gln Gly Leu Ala Leu Leu Ser Glu Ala Val Leu Arg Gly Gln Ala Leu
 [1067] 65 70 75 80
 [1068] Leu Val Asn Ser Ser Gln Pro Trp Glu Pro Leu Gln Leu His Val Asp
 [1069] 85 90 95
 [1070] Lys Ala Val Ser Gly Leu Arg Ser Leu Thr Thr Leu Leu Arg Ala Leu
 [1071] 100 105 110
 [1072] Gly Ala Gln Lys Glu Ala Ile Ser Pro Pro Asp Ala Ala Ser Ala Ala
 [1073] 115 120 125
 [1074] Pro Leu Arg Thr Ile Thr Ala Asp Thr Phe Arg Lys Leu Phe Arg Val
 [1075] 130 135 140
 [1076] Tyr Ser Asn Phe Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala
 [1077] 145 150 155 160
 [1078] Cys Arg Thr Gly Asp Arg
 [1079] 165
 [1080] <210> 19
 [1081] <400> 19
 [1082] 000
 [1083] <210> 20
 [1084] <211> 361
 [1085] <212> PRT
 [1086] <213> Artificial Sequence
 [1087] <220>
 [1088] <223> Polypeptide capable of binding to O-glycans but lacks or has
 [1089] reduced O-glycoprotein-specific endoprotease activity
 [1090] <400> 20
 [1091] Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys Lys

[1092]	1	5	10	15
[1093]	Thr Ala Arg Gln Leu Asn Ile Val Tyr Phe Leu Gly Ser Asp Thr Glu			
[1094]	20	25	30	
[1095]	Pro Val Pro Asp Tyr Glu Arg Arg Leu Ser Glu Leu Leu Tyr Leu			
[1096]	35	40	45	
[1097]	Gln Gln Phe Tyr Gly Lys Glu Met Gln Arg His Gly Tyr Gly Ala Arg			
[1098]	50	55	60	
[1099]	Ser Phe Gly Leu Asp Ile Lys Ser Pro Gly Arg Val Asn Ile Ile Glu			
[1100]	65	70	75	80
[1101]	Tyr Lys Ala Lys Asn Pro Ala Ala His Tyr Pro Tyr Glu Asn Gly Gly			
[1102]	85	90	95	
[1103]	Gly Trp Lys Ala Ala Gln Glu Leu Asp Glu Phe Phe Lys Ala His Pro			
[1104]	100	105	110	
[1105]	Asp Arg Lys Lys Ser Gln His Thr Leu Ile Ile Met Pro Thr Trp Asn			
[1106]	115	120	125	
[1107]	Asp Glu Lys Asn Gly Pro Asp Asn Pro Gly Gly Val Pro Phe Tyr Gly			
[1108]	130	135	140	
[1109]	Met Gly Arg Asn Cys Phe Ala Leu Asp Tyr Pro Ala Phe Asp Ile Lys			
[1110]	145	150	155	160
[1111]	His Leu Gly Gln Lys Thr Arg Glu Gly Arg Leu Leu Thr Lys Trp Tyr			
[1112]	165	170	175	
[1113]	Gly Gly Met Ala Ala Ala Leu Gly His Gly Leu Asn Leu Pro His Asn			
[1114]	180	185	190	
[1115]	His Gln Thr Ala Ser Asp Gly Lys Lys Tyr Gly Thr Ala Leu Met Gly			
[1116]	195	200	205	
[1117]	Ser Gly Asn Tyr Thr Phe Gly Thr Ser Pro Thr Phe Leu Thr Pro Ala			
[1118]	210	215	220	
[1119]	Ser Cys Ala Leu Leu Asp Ala Cys Glu Val Phe Ser Val Thr Pro Ser			
[1120]	225	230	235	240
[1121]	Gln Gln Phe Tyr Glu Gly Lys Pro Glu Val Glu Val Gly Asp Val Ala			
[1122]	245	250	255	
[1123]	Ile Ser Phe Lys Gly Asp Gln Ile Leu Val Ser Gly Asn Tyr Lys Ser			
[1124]	260	265	270	
[1125]	Pro Gln Thr Val Lys Ala Leu Asn Val Tyr Ile Gln Asp Pro Pro Tyr			
[1126]	275	280	285	
[1127]	Ala Val Asn Gln Asp Tyr Asp Ala Val Ser Phe Ser Arg Arg Leu Gly			
[1128]	290	295	300	
[1129]	Lys Lys Ser Gly Lys Phe Ser Met Lys Ile Asp Lys Lys Glu Leu Glu			
[1130]	305	310	315	320
[1131]	Gly Leu Asn Asn Asn Glu Phe Arg Ile Ser Leu Met Phe Ile Leu Ala			
[1132]	325	330	335	
[1133]	Asn Gly Leu His Met Gln Lys His Phe Thr Phe His Trp Asp Ala Leu			

[1134]	340	345	350
[1135]	Gln Asp Tyr Arg Asp Gly Ser Lys Ser		
[1136]	355	360	
[1137]	<210> 21		
[1138]	<211> 371		
[1139]	<212> PRT		
[1140]	<213> Artificial Sequence		
[1141]	<220>		
[1142]	<223> LS HE206AA or LS H205A/E206A (including N-terminal methionine		
[1143]	and C-terminal linker)		
[1144]	<400> 21		
[1145]	Met Glu Val Thr Val Pro Asp Ala Leu Lys Asp Arg Ile Ala Leu Lys		
[1146]	1 5 10 15		
[1147]	Lys Thr Ala Arg Gln Leu Asn Ile Val Tyr Phe Leu Gly Ser Asp Thr		
[1148]	20 25 30		
[1149]	Glu Pro Val Pro Asp Tyr Glu Arg Arg Leu Ser Glu Leu Leu Leu Tyr		
[1150]	35 40 45		
[1151]	Leu Gln Gln Phe Tyr Gly Lys Glu Met Gln Arg His Gly Tyr Gly Ala		
[1152]	50 55 60		
[1153]	Arg Ser Phe Gly Leu Asp Ile Lys Ser Pro Gly Arg Val Asn Ile Ile		
[1154]	65 70 75 80		
[1155]	Glu Tyr Lys Ala Lys Asn Pro Ala Ala His Tyr Pro Tyr Glu Asn Gly		
[1156]	85 90 95		
[1157]	Gly Gly Trp Lys Ala Ala Gln Glu Leu Asp Glu Phe Phe Lys Ala His		
[1158]	100 105 110		
[1159]	Pro Asp Arg Lys Lys Ser Gln His Thr Leu Ile Ile Met Pro Thr Trp		
[1160]	115 120 125		
[1161]	Asn Asp Glu Lys Asn Gly Pro Asp Asn Pro Gly Gly Val Pro Phe Tyr		
[1162]	130 135 140		
[1163]	Gly Met Gly Arg Asn Cys Phe Ala Leu Asp Tyr Pro Ala Phe Asp Ile		
[1164]	145 150 155 160		
[1165]	Lys His Leu Gly Gln Lys Thr Arg Glu Gly Arg Leu Leu Thr Lys Trp		
[1166]	165 170 175		
[1167]	Tyr Gly Gly Met Ala Ala Ala Leu Gly His Gly Leu Asn Leu Pro His		
[1168]	180 185 190		
[1169]	Asn His Gln Thr Ala Ser Asp Gly Lys Lys Tyr Gly Thr Ala Leu Met		
[1170]	195 200 205		
[1171]	Gly Ser Gly Asn Tyr Thr Phe Gly Thr Ser Pro Thr Phe Leu Thr Pro		
[1172]	210 215 220		
[1173]	Ala Ser Cys Ala Leu Leu Asp Ala Cys Glu Val Phe Ser Val Thr Pro		
[1174]	225 230 235 240		
[1175]	Ser Gln Gln Phe Tyr Glu Gly Lys Pro Glu Val Glu Val Gly Asp Val		

[1176]	245	250	255
[1177]	Ala Ile Ser Phe Lys Gly Asp Gln Ile Leu Val Ser Gly Asn Tyr Lys		
[1178]	260	265	270
[1179]	Ser Pro Gln Thr Val Lys Ala Leu Asn Val Tyr Ile Gln Asp Pro Pro		
[1180]	275	280	285
[1181]	Tyr Ala Val Asn Gln Asp Tyr Asp Ala Val Ser Phe Ser Arg Arg Leu		
[1182]	290	295	300
[1183]	Gly Lys Lys Ser Gly Lys Phe Ser Met Lys Ile Asp Lys Lys Glu Leu		
[1184]	305	310	315
[1185]	Glu Gly Leu Asn Asn Asn Glu Phe Arg Ile Ser Leu Met Phe Ile Leu		
[1186]	325	330	335
[1187]	Ala Asn Gly Leu His Met Gln Lys His Phe Thr Phe His Trp Asp Ala		
[1188]	340	345	350
[1189]	Leu Gln Asp Tyr Arg Asp Gly Ser Lys Ser Gly Ser Gly His His His		
[1190]	355	360	365
[1191]	His His His		
[1192]	370		
[1193]	<210> 22		
[1194]	<211> 1116		
[1195]	<212> DNA		
[1196]	<213> Artificial Sequence		
[1197]	<220>		
[1198]	<223> Nucleotide sequence encoding LS HE206AA or LS H205A/E206A		
[1199]	<400> 22		
[1200]	atggaagtca ctgtgccgga cgccctgaaa gatcgcatcg cgctgaagaa aaccgctcgt 60		
[1201]	cagctgaata tcgtctactt cctgggttct gataccgaac cggttccgga ctacgagcgc 120		
[1202]	cgtctgagcg agctgctgtt gtatctgcag caattctatg gtaaagaaat gcagcgccat 180		
[1203]	ggctatggcg cagcgagctt tggctctggac attaagtcac cgggtcgtgt gaacattatc 240		
[1204]	gagtacaaag cgaagaaccc ggcagcgcat taccgatatg agaatggtgg cggttggaag 300		
[1205]	gctgcacaag aactggacga atttttcaag gcccatccag accgcaagaa aagccagcac 360		
[1206]	accctgatca tcatgcctac ctggaatgat gagaaaaatg gtcttgacaa tccgggtggc 420		
[1207]	gttccgttct atggtatggg tcgtaattgt tttgcgttgg actaccggcg gtttgatatc 480		
[1208]	aagcacctgg gtcagaaaac gcgtgagggt cgtctgctga cgaaatggta cggtggcatg 540		
[1209]	gcggccgcgc tgggccacgg cctgaatctg ccgcacaatc accagaccgc gagcgatggc 600		
[1210]	aagaaatatg gcaccgccct gatgggtagc ggcaactaca cgttcggtac cagcccgacg 660		
[1211]	ttctgacccc cggcgagctg tgcgctgctg gatgcctgcg aagtgttcag cgttaccccc 720		
[1212]	agccaacagt tttatgaggg taagccagaa gtcgaggttg gtgatgttgc aatttccttc 780		
[1213]	aagggtgata aaatcttggc cagcggtaac tacaagagcc cgcaaaccgt gaaagctctg 840		
[1214]	aacgtttaca ttcaggatcc gccgtacgcc gtgaaccaag actacgatgc agtgagcttt 900		
[1215]	agccgtcgtc tgggcaaaaa gtccggtgaag tttagcatga agattgacaa aaaagaactg 960		
[1216]	gaaggcctga ataacaacga attccgtatt tccttgatgt tcattctggc aaacggctta 1020		
[1217]	cacatgcaga agcactttac gtttcaactgg gatgcgctgc aagactaccg tgacggtagc 1080		

[1218]	aaatctgggtt cgggtcatca tcaccacat cactga 1116			
[1219]	<210> 23			
[1220]	<211> 10			
[1221]	<212> PRT			
[1222]	<213> Artificial Sequence			
[1223]	<220>			
[1224]	<223> disrupted metalloprotease motif			
[1225]	<400> 23			
[1226]	Gly Met Ala His Ala Leu Gly His Gly Leu			
[1227]	1	5	10	
[1228]	<210> 24			
[1229]	<211> 10			
[1230]	<212> PRT			
[1231]	<213> Artificial Sequence			
[1232]	<220>			
[1233]	<223> disrupted metalloprotease motif			
[1234]	<400> 24			
[1235]	Gly Met Ala Ala Glu Leu Gly His Gly Leu			
[1236]	1	5	10	
[1237]	<210> 25			
[1238]	<211> 10			
[1239]	<212> PRT			
[1240]	<213> Artificial Sequence			
[1241]	<220>			
[1242]	<223> disrupted metalloprotease motif			
[1243]	<400> 25			
[1244]	Gly Met Ala Ala Ala Leu Gly His Gly Leu			
[1245]	1	5	10	
[1246]	<210> 26			
[1247]	<211> 882			
[1248]	<212> PRT			
[1249]	<213> Pseudomonas aeruginosa			
[1250]	<400> 26			
[1251]	Ala Thr Gln Glu Glu Ile Leu Asp Ala Ala Leu Val Ser Gly Asp Ser			
[1252]	1	5	10	15
[1253]	Ser Gln Leu Thr Asp Ser His Leu Val Ala Leu Arg Leu Gln Gln Gln			
[1254]		20	25	30
[1255]	Val Glu Arg Ile Arg Gln Thr Arg Thr Gln Leu Leu Asp Gly Leu Tyr			
[1256]		35	40	45
[1257]	Gln Asn Leu Ser Gln Ala Tyr Asp Pro Gly Ala Ala Ser Met Trp Val			
[1258]	50	55	60	
[1259]	Leu Pro Ala Asn Pro Asp Asn Thr Leu Pro Phe Leu Ile Gly Asp Lys			

[1260]	65	70	75	80
[1261]	Gly Arg Val Leu Ala Ser Leu Ser Leu Glu Ala Gly Gly Arg Gly Leu			
[1262]		85	90	95
[1263]	Ala Tyr Gly Thr Asn Val Leu Thr Gln Leu Ser Gly Thr Asn Ala Ala			
[1264]		100	105	110
[1265]	His Ala Pro Leu Leu Lys Arg Ala Val Gln Trp Leu Val Asn Gly Asp			
[1266]		115	120	125
[1267]	Pro Gly Ala Ala Thr Ala Lys Asp Phe Lys Val Ser Val Val Gly Val			
[1268]		130	135	140
[1269]	Asp Lys Thr Ala Ala Leu Asn Gly Leu Lys Ser Ala Gly Leu Gln Pro			
[1270]	145	150	155	160
[1271]	Ala Asp Ala Ala Cys Asn Ala Leu Thr Asp Ala Ser Cys Ala Ser Thr			
[1272]		165	170	175
[1273]	Ser Lys Leu Leu Val Leu Gly Asn Gly Ala Ser Ala Ala Ser Leu Ser			
[1274]		180	185	190
[1275]	Ala Thr Val Arg Ala Arg Leu Gln Ala Gly Leu Pro Ile Leu Phe Val			
[1276]		195	200	205
[1277]	His Thr Asn Gly Trp Asn Gln Ser Ser Thr Gly Gln Gln Ile Leu Ala			
[1278]		210	215	220
[1279]	Gly Leu Gly Leu Gln Glu Gly Pro Tyr Gly Gly Asn Tyr Trp Asp Lys			
[1280]	225	230	235	240
[1281]	Asp Arg Val Pro Ser Ser Arg Thr Arg Thr Arg Ser Val Glu Leu Gly			
[1282]		245	250	255
[1283]	Gly Ala Tyr Gly Gln Asp Pro Ala Leu Val Gln Gln Ile Val Asp Gly			
[1284]		260	265	270
[1285]	Ser Trp Arg Thr Asp Tyr Asp Trp Ser Lys Cys Thr Ser Tyr Val Gly			
[1286]		275	280	285
[1287]	Arg Thr Thr Cys Asp Asp Val Pro Gly Leu Ser Asp Phe Ser Lys Arg			
[1288]		290	295	300
[1289]	Val Asp Val Leu Lys Gly Ala Leu Asp Ala Tyr Asn Gln Lys Ala Gln			
[1290]	305	310	315	320
[1291]	Asn Leu Phe Ala Leu Pro Gly Thr Thr Ser Leu Arg Leu Trp Leu Leu			
[1292]		325	330	335
[1293]	Trp Ala Asp Ala Val Arg Gln Asn Ile Arg Tyr Pro Met Asp Lys Ala			
[1294]		340	345	350
[1295]	Ala Asp Thr Ala Arg Phe Gln Glu Thr Phe Val Ala Asp Ala Ile Val			
[1296]		355	360	365
[1297]	Gly Tyr Val Arg Glu Ala Gly Ala Ala Gln Lys Glu Leu Gly Ser Tyr			
[1298]		370	375	380
[1299]	Ala Gly Gln Arg Gln Gln Ser Met Pro Val Ser Gly Ser Glu Glu Thr			
[1300]	385	390	395	400
[1301]	Leu Thr Leu Thr Leu Pro Ser Ala Gln Gly Phe Thr Ala Ile Gly Arg			

[1302]	405							410							415				
[1303]	Met	Ala	Ala	Pro	Gly	Lys	Arg	Leu	Ser	Ile	Arg	Ile	Glu	Asp	Ala	Gly			
[1304]	420							425							430				
[1305]	Gln	Ala	Ser	Leu	Ala	Val	Gly	Leu	Asn	Thr	Gln	Arg	Ile	Gly	Ser	Thr			
[1306]	435							440							445				
[1307]	Arg	Leu	Trp	Asn	Thr	Arg	Gln	Tyr	Asp	Arg	Pro	Arg	Phe	Leu	Lys	Ser			
[1308]	450							455							460				
[1309]	Pro	Asp	Ile	Lys	Leu	Gln	Ala	Asn	Gln	Ser	Val	Ala	Leu	Val	Ser	Pro			
[1310]	465							470							475			480	
[1311]	Tyr	Gly	Gly	Leu	Leu	Gln	Leu	Val	Tyr	Ser	Gly	Ala	Thr	Pro	Gly	Gln			
[1312]	485							490							495				
[1313]	Thr	Val	Thr	Val	Lys	Val	Thr	Gly	Ala	Ala	Ser	Gln	Pro	Phe	Leu	Asp			
[1314]	500							505							510				
[1315]	Ile	Gln	Pro	Gly	Glu	Asp	Ser	Ser	Gln	Ala	Ile	Ala	Asp	Phe	Ile	Gln			
[1316]	515							520							525				
[1317]	Ala	Leu	Asp	Ala	Asp	Lys	Ala	Asp	Trp	Leu	Glu	Ile	Arg	Ser	Gly	Ser			
[1318]	530							535							540				
[1319]	Val	Glu	Val	His	Ala	Lys	Val	Glu	Lys	Val	Arg	Gly	Ser	Ile	Asp	Lys			
[1320]	545							550							555			560	
[1321]	Asp	Tyr	Gly	Gly	Asp	Val	Gln	Arg	Phe	Ile	Arg	Glu	Leu	Asn	Glu	Val			
[1322]	565							570							575				
[1323]	Phe	Ile	Asp	Asp	Ala	Tyr	Thr	Leu	Ala	Gly	Phe	Ala	Ile	Pro	Asn	Gln			
[1324]	580							585							590				
[1325]	Ala	Lys	Thr	Pro	Ala	Ile	Gln	Gln	Glu	Cys	Ala	Ala	Arg	Gly	Trp	Asp			
[1326]	595							600							605				
[1327]	Cys	Asp	Ser	Glu	Thr	Leu	His	Lys	Leu	Pro	Gly	Thr	Gln	His	Ile	Asn			
[1328]	610							615							620				
[1329]	Val	Asp	Gln	Tyr	Ala	Gln	Cys	Gly	Gly	Gly	Cys	Ser	Gly	Asn	Pro	Tyr			
[1330]	625							630							635			640	
[1331]	Asp	Gln	Thr	Trp	Gly	Leu	Asn	Pro	Arg	Gly	Trp	Gly	Glu	Ser	His	Glu			
[1332]	645							650							655				
[1333]	Leu	Gly	His	Asn	Leu	Gln	Val	Asn	Arg	Leu	Lys	Val	Tyr	Gly	Gly	Arg			
[1334]	660							665							670				
[1335]	Ser	Gly	Glu	Ile	Ser	Asn	Gln	Ile	Phe	Pro	Leu	His	Lys	Asp	Trp	Arg			
[1336]	675							680							685				
[1337]	Val	Leu	Arg	Glu	Phe	Gly	Gln	Asn	Leu	Asp	Asp	Thr	Arg	Val	Asn	Tyr			
[1338]	690							695							700				
[1339]	Arg	Asn	Ala	Tyr	Asn	Leu	Ile	Val	Ala	Gly	Arg	Ala	Glu	Ala	Asp	Pro			
[1340]	705							710							715			720	
[1341]	Leu	Ala	Gly	Val	Tyr	Lys	Arg	Leu	Trp	Glu	Asp	Pro	Gly	Thr	Tyr	Ala			
[1342]	725							730							735				
[1343]	Leu	Asn	Gly	Glu	Arg	Met	Ala	Phe	Tyr	Thr	Gln	Trp	Val	His	Tyr	Trp			

[1344]	740	745	750
[1345]	Ala Asp Leu Lys Asn Asp Pro Leu Gln Gly Trp Asp Ile Trp Thr Leu		
[1346]	755	760	765
[1347]	Leu Tyr Leu His Gln Arg Gln Val Asp Lys Ser Asp Trp Asp Ala Asn		
[1348]	770	775	780
[1349]	Lys Ala Ala Leu Gly Tyr Gly Thr Tyr Ala Gln Arg Pro Gly Asn Ser		
[1350]	785	790	795
[1351]	Gly Asp Ala Ser Ser Thr Asp Gly Asn Asp Asn Leu Leu Leu Gly Leu		
[1352]	805	810	815
[1353]	Ser Trp Leu Thr Gln Arg Asp Gln Arg Pro Thr Phe Ala Leu Trp Gly		
[1354]	820	825	830
[1355]	Ile Arg Thr Ser Ala Ala Ala Gln Ala Gln Val Ala Ala Tyr Gly Phe		
[1356]	835	840	845
[1357]	Ala Glu Gln Pro Ala Phe Phe Tyr Ala Asn Asn Arg Thr Asn Glu Tyr		
[1358]	850	855	860
[1359]	Ser Thr Val Lys Leu Leu Asp Met Ser Gln Gly Ser Pro Ala Trp Pro		
[1360]	865	870	875
[1361]	Phe Pro		
[1362]	<210> 27		
[1363]	<211> 536		
[1364]	<212> PRT		
[1365]	<213> Bacteroides thetaiotaomicron		
[1366]	<400> 27		
[1367]	Asp Lys Trp Glu Lys Glu Phe Arg Ile Arg Ser Tyr Glu Pro Tyr Ser		
[1368]	1	5	10
[1369]	Asn Ile Ala Glu Trp Ala Asp Lys Leu Met Thr Lys Lys Tyr Ser Asp		
[1370]	20	25	30
[1371]	Leu Asp Asn Pro Thr Gly Ile Ser Val Lys Ala Gly Asp Asp Ile Ile		
[1372]	35	40	45
[1373]	Val Leu Val Gly Asp Thr Tyr Gly Gln Asn Ile Ser Met Gln Cys Ile		
[1374]	50	55	60
[1375]	Trp Glu Thr Gly Thr Glu Tyr Lys Gln Thr Ala Ser Ser Gly Asp Val		
[1376]	65	70	75
[1377]	Tyr Met Leu Asn Pro Gly Val Asn Lys Leu Thr Met Lys Gly Glu Gly		
[1378]	85	90	95
[1379]	Gln Leu Phe Val Met Tyr Asn Thr Glu Leu Thr Ser Asn Thr Ala Lys		
[1380]	100	105	110
[1381]	Pro Ile Lys Ile His Ile Pro Leu Gly Ser Gly Thr Val Asn Gly Phe		
[1382]	115	120	125
[1383]	Phe Asp Leu Lys Glu His Lys Thr Asp Glu Lys Tyr Ala Glu Leu Leu		
[1384]	130	135	140
[1385]	Lys Lys Ser Thr His Lys Tyr Phe Cys Ile Arg Gly Glu Lys Ile Met		

[1386]	145	150	155	160
[1387]	Phe Tyr Phe His Arg Asn Lys Leu Leu Glu Tyr Val Pro Asn Asn Ile			
[1388]		165	170	175
[1389]	Leu Ser Ala Ile His Leu Trp Asp Asn Ile Val Gly Trp Gln Gln Glu			
[1390]		180	185	190
[1391]	Leu Met Gly Ile Asp Asp Val Arg Pro Ser Gln Val Asn Asn His Leu			
[1392]		195	200	205
[1393]	Phe Ala Ile Ser Pro Glu Gly Ser Tyr Met Trp Ala Ser Asp Tyr Gln			
[1394]	210	215	220	
[1395]	Ile Gly Phe Val Tyr Thr Tyr Leu Gly Asn Ile Leu Leu Glu Asp Asn			
[1396]	225	230	235	240
[1397]	Val Met Ala Ala Glu Asp Asn Ala Trp Gly Pro Ala His Glu Ile Gly			
[1398]		245	250	255
[1399]	His Val His Gln Ala Ala Ile Asn Trp Ala Ser Ser Thr Glu Ser Ser			
[1400]		260	265	270
[1401]	Asn Asn Leu Phe Ser Asn Phe Ile Ile Tyr Lys Leu Gly Lys Tyr Lys			
[1402]		275	280	285
[1403]	Ser Arg Gly Asn Gly Leu Gly Ser Val Ala Thr Ala Arg Tyr Ala Asn			
[1404]	290	295	300	
[1405]	Gly Gln Ala Trp Tyr Asn Met Gly Asp Ala Thr His Gln Asn Glu Asp			
[1406]	305	310	315	320
[1407]	Thr Glu Thr His Met Arg Met Asn Trp Gln Leu Trp Ile Tyr Tyr His			
[1408]		325	330	335
[1409]	Arg Cys Glu Tyr Lys Thr Asp Phe Trp Gln Thr Leu Phe Lys Leu Met			
[1410]		340	345	350
[1411]	Arg Glu Val Asn Met Thr Glu Gly Glu Asp Pro Gly Lys Lys Gln Leu			
[1412]		355	360	365
[1413]	Glu Phe Ala Lys Met Ala Ser Lys Ala Ala Asn Gln Asn Leu Thr Asp			
[1414]	370	375	380	
[1415]	Phe Phe Glu Met Trp Gly Phe Phe Glu Pro Val Asn Thr Thr Ile Glu			
[1416]	385	390	395	400
[1417]	Gln Tyr Gly Thr Tyr Lys Tyr Tyr Val Ser Asp Ala Met Ile Arg Glu			
[1418]		405	410	415
[1419]	Ala Lys Glu Tyr Met Ala Gln Phe Pro Ala Pro Lys His Ala Phe Gln			
[1420]		420	425	430
[1421]	Tyr Ile Glu Asp Arg Lys Lys Ser Glu Phe Pro Ser Asn Asp Tyr Arg			
[1422]		435	440	445
[1423]	Tyr Ser Ala Val Gly Asp Val Gly Tyr Tyr Thr Gln Phe Lys Glu Asn			
[1424]	450	455	460	
[1425]	Gln Lys Ile Thr Lys Ala Ile Thr Ala Glu Leu Ala Gly Arg Lys Val			
[1426]	465	470	475	480
[1427]	Ser Ile Gln Asn Gly Asp Glu Ala Val Ala Phe Glu Leu Arg Glu Asn			

[1428]	485								490				495				
[1429]	Asp	Glu	Asn	Gly	Lys	Leu	Leu	Tyr	Phe	Ser	Thr	Phe	Thr	Thr	Phe	Glu	
[1430]	500								505				510				
[1431]	Ile	Pro	Ser	Ser	Ile	Leu	Met	Val	Asn	Ala	Lys	Leu	Tyr	Ala	Val	Gln	
[1432]	515								520				525				
[1433]	Ala	Asp	Gly	Lys	Arg	Ile	Leu	Leu									
[1434]	530								535								
[1435]	<210> 28																
[1436]	<211> 507																
[1437]	<212> PRT																
[1438]	<213> Clostridium perfringens																
[1439]	<400> 28																
[1440]	Val	Leu	Glu	Leu	Glu	Met	Arg	Gly	Asp	Ser	Ile	Ser	Glu	Ala	Lys	Lys	
[1441]	1	5								10				15			
[1442]	Arg	Lys	Val	Trp	Asn	Phe	Gln	Asp	Trp	Gln	Ile	Thr	Gly	Leu	Ser	Ala	
[1443]	20								25				30				
[1444]	Arg	Ala	Gly	Asp	Lys	Ile	Thr	Val	Tyr	Val	Asp	Val	Ala	Glu	Gly	Asp	
[1445]	35								40				45				
[1446]	Pro	Thr	Pro	Thr	Leu	Leu	Tyr	Lys	Gln	Ser	Leu	Thr	Gln	His	Gly	Gly	
[1447]	50								55				60				
[1448]	Ala	Thr	Ser	Phe	Gln	Leu	Lys	Pro	Gly	Lys	Asn	Glu	Ile	Thr	Ile	Pro	
[1449]	65	70								75				80			
[1450]	Glu	Ile	Asn	Tyr	Glu	Ser	Asn	Gly	Ile	Pro	Lys	Asp	Val	Ile	Gln	Gly	
[1451]	85								90				95				
[1452]	Gly	Asp	Leu	Phe	Phe	Thr	Asn	Tyr	Lys	Ser	Asp	Ser	Gln	Lys	Arg	Ala	
[1453]	100								105				110				
[1454]	Pro	Lys	Val	Arg	Ile	Glu	Gly	Ala	Ser	Lys	Tyr	Pro	Val	Phe	Ile	Leu	
[1455]	115								120				125				
[1456]	Gly	Lys	Ser	Asp	Glu	Asn	Glu	Val	Met	Lys	Glu	Leu	Glu	Ala	Tyr	Val	
[1457]	130								135				140				
[1458]	Glu	Lys	Ile	Lys	Ala	Glu	Pro	Lys	Thr	Thr	Pro	Asn	Ile	Phe	Ala	Val	
[1459]	145	150								155				160			
[1460]	Ser	Ser	Asn	Lys	Ser	Leu	Glu	Phe	Val	Gln	Ala	Thr	Tyr	Ala	Leu	Asp	
[1461]	165								170				175				
[1462]	Trp	Tyr	Lys	Lys	Asn	Asn	Lys	Thr	Pro	Lys	Tyr	Thr	Ala	Glu	Gln	Trp	
[1463]	180								185				190				
[1464]	Asp	Gln	Tyr	Ile	Ala	Asp	Ala	Met	Gly	Phe	Trp	Gly	Phe	Asp	Asn	Ser	
[1465]	195								200				205				
[1466]	Lys	Asp	Val	Asn	Ser	Asp	Phe	Asn	Phe	Arg	Ile	Met	Pro	Met	Val	Lys	
[1467]	210								215				220				
[1468]	Asn	Leu	Ser	Gly	Gly	Ala	Phe	Met	Asn	Ala	Gly	Asn	Gly	Val	Ile	Gly	
[1469]	225	230								235				240			

[1470]	Ile Arg Pro Gly Asn Gln Asp Ala Ile Leu Ala Ala Asn Lys Gly Trp
[1471]	245 250 255
[1472]	Gly Val Ala His Glu Leu Gly His Asn Phe Asp Thr Gly Gly Arg Thr
[1473]	260 265 270
[1474]	Ile Val Glu Val Thr Asn Asn Met Met Pro Leu Phe Phe Glu Ser Lys
[1475]	275 280 285
[1476]	Tyr Lys Thr Lys Thr Arg Ile Thr Asp Gln Asn Ile Trp Glu Asn Asn
[1477]	290 295 300
[1478]	Thr Tyr Pro Lys Val Gly Leu Asp Asp Tyr Ser Asn Asn Glu Leu Tyr
[1479]	305 310 315 320
[1480]	Asn Lys Ala Asp Ser Thr His Leu Ala Gln Leu Ala Pro Leu Trp Gln
[1481]	325 330 335
[1482]	Leu Tyr Leu Tyr Asp Asn Thr Phe Tyr Gly Lys Phe Glu Arg Gln Phe
[1483]	340 345 350
[1484]	Arg Glu Arg Asp Phe Gly Asn Lys Asn Arg Glu Asp Ile Tyr Lys Ser
[1485]	355 360 365
[1486]	Trp Val Val Ala Ala Ser Asp Ala Met Glu Leu Asp Leu Thr Glu Phe
[1487]	370 375 380
[1488]	Phe Ala Arg His Gly Ile Arg Val Asp Asp Lys Val Lys Glu Asp Leu
[1489]	385 390 395 400
[1490]	Ala Lys Tyr Pro Lys Pro Asp Lys Lys Ile Tyr Tyr Leu Asn Asp Leu
[1491]	405 410 415
[1492]	Ala Met Asn Tyr Lys Gly Asp Gly Phe Thr Glu Asn Ala Lys Val Ser
[1493]	420 425 430
[1494]	Val Ser Thr Ser Gly Ser Asn Gly Asn Ile Lys Leu Ser Phe Ser Val
[1495]	435 440 445
[1496]	Asp Asp Glu Asn Lys Asp Asn Ile Leu Gly Tyr Glu Ile Arg Arg Asp
[1497]	450 455 460
[1498]	Gly Lys Tyr Val Gly Phe Thr Ser Asn Asp Ser Phe Val Asp Thr Lys
[1499]	465 470 475 480
[1500]	Ser Asn Leu Asp Glu Asp Gly Val Tyr Val Val Thr Pro Tyr Asp Arg
[1501]	485 490 495
[1502]	Lys Leu Asn Thr Leu Asn Pro Ile Glu Val Asn
[1503]	500 505
[1504]	<210> 29
[1505]	<211> 892
[1506]	<212> PRT
[1507]	<213> Artificial Sequence
[1508]	<220>
[1509]	<223> polypeptide having O-glycoprotein-specific endoprotease activity
[1510]	(including N-terminal methionine and a C-terminal linker + His6
[1511]	tag)

[1512]	<400> 29															
[1513]	Met	Ala	Thr	Gln	Glu	Glu	Ile	Leu	Asp	Ala	Ala	Leu	Val	Ser	Gly	Asp
[1514]	1			5			10			15						
[1515]	Ser	Ser	Gln	Leu	Thr	Asp	Ser	His	Leu	Val	Ala	Leu	Arg	Leu	Gln	Gln
[1516]	20					25					30					
[1517]	Gln	Val	Glu	Arg	Ile	Arg	Gln	Thr	Arg	Thr	Gln	Leu	Leu	Asp	Gly	Leu
[1518]	35				40				45							
[1519]	Tyr	Gln	Asn	Leu	Ser	Gln	Ala	Tyr	Asp	Pro	Gly	Ala	Ala	Ser	Met	Trp
[1520]	50			55			60									
[1521]	Val	Leu	Pro	Ala	Asn	Pro	Asp	Asn	Thr	Leu	Pro	Phe	Leu	Ile	Gly	Asp
[1522]	65		70		75		80									
[1523]	Lys	Gly	Arg	Val	Leu	Ala	Ser	Leu	Ser	Leu	Glu	Ala	Gly	Gly	Arg	Gly
[1524]	85					90					95					
[1525]	Leu	Ala	Tyr	Gly	Thr	Asn	Val	Leu	Thr	Gln	Leu	Ser	Gly	Thr	Asn	Ala
[1526]	100				105				110							
[1527]	Ala	His	Ala	Pro	Leu	Leu	Lys	Arg	Ala	Val	Gln	Trp	Leu	Val	Asn	Gly
[1528]	115			120			125									
[1529]	Asp	Pro	Gly	Ala	Ala	Thr	Ala	Lys	Asp	Phe	Lys	Val	Ser	Val	Val	Gly
[1530]	130		135		140											
[1531]	Val	Asp	Lys	Thr	Ala	Ala	Leu	Asn	Gly	Leu	Lys	Ser	Ala	Gly	Leu	Gln
[1532]	145		150		155		160									
[1533]	Pro	Ala	Asp	Ala	Ala	Cys	Asn	Ala	Leu	Thr	Asp	Ala	Ser	Cys	Ala	Ser
[1534]	165					170					175					
[1535]	Thr	Ser	Lys	Leu	Leu	Val	Leu	Gly	Asn	Gly	Ala	Ser	Ala	Ala	Ser	Leu
[1536]	180				185				190							
[1537]	Ser	Ala	Thr	Val	Arg	Ala	Arg	Leu	Gln	Ala	Gly	Leu	Pro	Ile	Leu	Phe
[1538]	195			200			205									
[1539]	Val	His	Thr	Asn	Gly	Trp	Asn	Gln	Ser	Ser	Thr	Gly	Gln	Gln	Ile	Leu
[1540]	210		215		220											
[1541]	Ala	Gly	Leu	Gly	Leu	Gln	Glu	Gly	Pro	Tyr	Gly	Gly	Asn	Tyr	Trp	Asp
[1542]	225		230		235		240									
[1543]	Lys	Asp	Arg	Val	Pro	Ser	Ser	Arg	Thr	Arg	Thr	Arg	Ser	Val	Glu	Leu
[1544]	245					250					255					
[1545]	Gly	Gly	Ala	Tyr	Gly	Gln	Asp	Pro	Ala	Leu	Val	Gln	Gln	Ile	Val	Asp
[1546]	260				265				270							
[1547]	Gly	Ser	Trp	Arg	Thr	Asp	Tyr	Asp	Trp	Ser	Lys	Cys	Thr	Ser	Tyr	Val
[1548]	275			280			285									
[1549]	Gly	Arg	Thr	Thr	Cys	Asp	Asp	Val	Pro	Gly	Leu	Ser	Asp	Phe	Ser	Lys
[1550]	290		295		300											
[1551]	Arg	Val	Asp	Val	Leu	Lys	Gly	Ala	Leu	Asp	Ala	Tyr	Asn	Gln	Lys	Ala
[1552]	305		310		315		320									
[1553]	Gln	Asn	Leu	Phe	Ala	Leu	Pro	Gly	Thr	Thr	Ser	Leu	Arg	Leu	Trp	Leu

[1554]		325		330		335
[1555]	Leu Trp Ala Asp Ala Val Arg Gln Asn Ile Arg Tyr Pro Met Asp Lys					
[1556]		340		345		350
[1557]	Ala Ala Asp Thr Ala Arg Phe Gln Glu Thr Phe Val Ala Asp Ala Ile					
[1558]		355		360		365
[1559]	Val Gly Tyr Val Arg Glu Ala Gly Ala Ala Gln Lys Glu Leu Gly Ser					
[1560]		370		375		380
[1561]	Tyr Ala Gly Gln Arg Gln Gln Ser Met Pro Val Ser Gly Ser Glu Glu					
[1562]	385		390		395	400
[1563]	Thr Leu Thr Leu Thr Leu Pro Ser Ala Gln Gly Phe Thr Ala Ile Gly					
[1564]		405		410		415
[1565]	Arg Met Ala Ala Pro Gly Lys Arg Leu Ser Ile Arg Ile Glu Asp Ala					
[1566]		420		425		430
[1567]	Gly Gln Ala Ser Leu Ala Val Gly Leu Asn Thr Gln Arg Ile Gly Ser					
[1568]		435		440		445
[1569]	Thr Arg Leu Trp Asn Thr Arg Gln Tyr Asp Arg Pro Arg Phe Leu Lys					
[1570]		450		455		460
[1571]	Ser Pro Asp Ile Lys Leu Gln Ala Asn Gln Ser Val Ala Leu Val Ser					
[1572]	465		470		475	480
[1573]	Pro Tyr Gly Gly Leu Leu Gln Leu Val Tyr Ser Gly Ala Thr Pro Gly					
[1574]		485		490		495
[1575]	Gln Thr Val Thr Val Lys Val Thr Gly Ala Ala Ser Gln Pro Phe Leu					
[1576]		500		505		510
[1577]	Asp Ile Gln Pro Gly Glu Asp Ser Ser Gln Ala Ile Ala Asp Phe Ile					
[1578]		515		520		525
[1579]	Gln Ala Leu Asp Ala Asp Lys Ala Asp Trp Leu Glu Ile Arg Ser Gly					
[1580]		530		535		540
[1581]	Ser Val Glu Val His Ala Lys Val Glu Lys Val Arg Gly Ser Ile Asp					
[1582]	545		550		555	560
[1583]	Lys Asp Tyr Gly Gly Asp Val Gln Arg Phe Ile Arg Glu Leu Asn Glu					
[1584]		565		570		575
[1585]	Val Phe Ile Asp Asp Ala Tyr Thr Leu Ala Gly Phe Ala Ile Pro Asn					
[1586]		580		585		590
[1587]	Gln Ala Lys Thr Pro Ala Ile Gln Gln Glu Cys Ala Ala Arg Gly Trp					
[1588]		595		600		605
[1589]	Asp Cys Asp Ser Glu Thr Leu His Lys Leu Pro Gly Thr Gln His Ile					
[1590]		610		615		620
[1591]	Asn Val Asp Gln Tyr Ala Gln Cys Gly Gly Gly Cys Ser Gly Asn Pro					
[1592]	625		630		635	640
[1593]	Tyr Asp Gln Thr Trp Gly Leu Asn Pro Arg Gly Trp Gly Glu Ser His					
[1594]		645		650		655
[1595]	Glu Leu Gly His Asn Leu Gln Val Asn Arg Leu Lys Val Tyr Gly Gly					

[1596]	660	665	670
[1597]	Arg Ser Gly Glu Ile Ser Asn Gln Ile Phe Pro Leu His Lys Asp Trp		
[1598]	675	680	685
[1599]	Arg Val Leu Arg Glu Phe Gly Gln Asn Leu Asp Asp Thr Arg Val Asn		
[1600]	690	695	700
[1601]	Tyr Arg Asn Ala Tyr Asn Leu Ile Val Ala Gly Arg Ala Glu Ala Asp		
[1602]	705	710	715
[1603]	Pro Leu Ala Gly Val Tyr Lys Arg Leu Trp Glu Asp Pro Gly Thr Tyr		
[1604]	725	730	735
[1605]	Ala Leu Asn Gly Glu Arg Met Ala Phe Tyr Thr Gln Trp Val His Tyr		
[1606]	740	745	750
[1607]	Trp Ala Asp Leu Lys Asn Asp Pro Leu Gln Gly Trp Asp Ile Trp Thr		
[1608]	755	760	765
[1609]	Leu Leu Tyr Leu His Gln Arg Gln Val Asp Lys Ser Asp Trp Asp Ala		
[1610]	770	775	780
[1611]	Asn Lys Ala Ala Leu Gly Tyr Gly Thr Tyr Ala Gln Arg Pro Gly Asn		
[1612]	785	790	795
[1613]	Ser Gly Asp Ala Ser Ser Thr Asp Gly Asn Asp Asn Leu Leu Leu Gly		
[1614]	805	810	815
[1615]	Leu Ser Trp Leu Thr Gln Arg Asp Gln Arg Pro Thr Phe Ala Leu Trp		
[1616]	820	825	830
[1617]	Gly Ile Arg Thr Ser Ala Ala Ala Gln Ala Gln Val Ala Ala Tyr Gly		
[1618]	835	840	845
[1619]	Phe Ala Glu Gln Pro Ala Phe Phe Tyr Ala Asn Asn Arg Thr Asn Glu		
[1620]	850	855	860
[1621]	Tyr Ser Thr Val Lys Leu Leu Asp Met Ser Gln Gly Ser Pro Ala Trp		
[1622]	865	870	875
[1623]	Pro Phe Pro Gly Ser Gly His His His His His His		
[1624]	885	890	
[1625]	<210> 30		
[1626]	<211> 546		
[1627]	<212> PRT		
[1628]	<213> Artificial Sequence		
[1629]	<220>		
[1630]	<223> polypeptide having O-glycoprotein-specific endoprotease activity		
[1631]	(including N-terminal methionine and a C-terminal linker + His6		
[1632]	tag)		
[1633]	<400> 30		
[1634]	Met Asp Lys Trp Glu Lys Glu Phe Arg Ile Arg Ser Tyr Glu Pro Tyr		
[1635]	1	5	10
[1636]	Ser Asn Ile Ala Glu Trp Ala Asp Lys Leu Met Thr Lys Lys Tyr Ser		
[1637]	20	25	30

[1638]	Asp	Leu	Asp	Asn	Pro	Thr	Gly	Ile	Ser	Val	Lys	Ala	Gly	Asp	Asp	Ile
[1639]		35						40					45			
[1640]	Ile	Val	Leu	Val	Gly	Asp	Thr	Tyr	Gly	Gln	Asn	Ile	Ser	Met	Gln	Cys
[1641]		50						55					60			
[1642]	Ile	Trp	Glu	Thr	Gly	Thr	Glu	Tyr	Lys	Gln	Thr	Ala	Ser	Ser	Gly	Asp
[1643]	65					70					75					80
[1644]	Val	Tyr	Met	Leu	Asn	Pro	Gly	Val	Asn	Lys	Leu	Thr	Met	Lys	Gly	Glu
[1645]					85					90					95	
[1646]	Gly	Gln	Leu	Phe	Val	Met	Tyr	Asn	Thr	Glu	Leu	Thr	Ser	Asn	Thr	Ala
[1647]					100					105					110	
[1648]	Lys	Pro	Ile	Lys	Ile	His	Ile	Pro	Leu	Gly	Ser	Gly	Thr	Val	Asn	Gly
[1649]					115					120					125	
[1650]	Phe	Phe	Asp	Leu	Lys	Glu	His	Lys	Thr	Asp	Glu	Lys	Tyr	Ala	Glu	Leu
[1651]		130						135					140			
[1652]	Leu	Lys	Lys	Ser	Thr	His	Lys	Tyr	Phe	Cys	Ile	Arg	Gly	Glu	Lys	Ile
[1653]	145					150					155					160
[1654]	Met	Phe	Tyr	Phe	His	Arg	Asn	Lys	Leu	Leu	Glu	Tyr	Val	Pro	Asn	Asn
[1655]					165					170					175	
[1656]	Ile	Leu	Ser	Ala	Ile	His	Leu	Trp	Asp	Asn	Ile	Val	Gly	Trp	Gln	Gln
[1657]					180					185					190	
[1658]	Glu	Leu	Met	Gly	Ile	Asp	Asp	Val	Arg	Pro	Ser	Gln	Val	Asn	Asn	His
[1659]					195					200					205	
[1660]	Leu	Phe	Ala	Ile	Ser	Pro	Glu	Gly	Ser	Tyr	Met	Trp	Ala	Ser	Asp	Tyr
[1661]		210						215					220			
[1662]	Gln	Ile	Gly	Phe	Val	Tyr	Thr	Tyr	Leu	Gly	Asn	Ile	Leu	Leu	Glu	Asp
[1663]	225					230					235					240
[1664]	Asn	Val	Met	Ala	Ala	Glu	Asp	Asn	Ala	Trp	Gly	Pro	Ala	His	Glu	Ile
[1665]					245					250					255	
[1666]	Gly	His	Val	His	Gln	Ala	Ala	Ile	Asn	Trp	Ala	Ser	Ser	Thr	Glu	Ser
[1667]					260					265					270	
[1668]	Ser	Asn	Asn	Leu	Phe	Ser	Asn	Phe	Ile	Ile	Tyr	Lys	Leu	Gly	Lys	Tyr
[1669]					275					280				285		
[1670]	Lys	Ser	Arg	Gly	Asn	Gly	Leu	Gly	Ser	Val	Ala	Thr	Ala	Arg	Tyr	Ala
[1671]		290						295					300			
[1672]	Asn	Gly	Gln	Ala	Trp	Tyr	Asn	Met	Gly	Asp	Ala	Thr	His	Gln	Asn	Glu
[1673]	305						310				315					320
[1674]	Asp	Thr	Glu	Thr	His	Met	Arg	Met	Asn	Trp	Gln	Leu	Trp	Ile	Tyr	Tyr
[1675]					325					330					335	
[1676]	His	Arg	Cys	Glu	Tyr	Lys	Thr	Asp	Phe	Trp	Gln	Thr	Leu	Phe	Lys	Leu
[1677]					340					345				350		
[1678]	Met	Arg	Glu	Val	Asn	Met	Thr	Glu	Gly	Glu	Asp	Pro	Gly	Lys	Lys	Gln
[1679]		355						360					365			

[1680]	Leu Glu Phe Ala Lys Met Ala Ser Lys Ala Ala Asn Gln Asn Leu Thr
[1681]	370 375 380
[1682]	Asp Phe Phe Glu Met Trp Gly Phe Phe Glu Pro Val Asn Thr Thr Ile
[1683]	385 390 395 400
[1684]	Glu Gln Tyr Gly Thr Tyr Lys Tyr Tyr Val Ser Asp Ala Met Ile Arg
[1685]	405 410 415
[1686]	Glu Ala Lys Glu Tyr Met Ala Gln Phe Pro Ala Pro Lys His Ala Phe
[1687]	420 425 430
[1688]	Gln Tyr Ile Glu Asp Arg Lys Lys Ser Glu Phe Pro Ser Asn Asp Tyr
[1689]	435 440 445
[1690]	Arg Tyr Ser Ala Val Gly Asp Val Gly Tyr Tyr Thr Gln Phe Lys Glu
[1691]	450 455 460
[1692]	Asn Gln Lys Ile Thr Lys Ala Ile Thr Ala Glu Leu Ala Gly Arg Lys
[1693]	465 470 475 480
[1694]	Val Ser Ile Gln Asn Gly Asp Glu Ala Val Ala Phe Glu Leu Arg Glu
[1695]	485 490 495
[1696]	Asn Asp Glu Asn Gly Lys Leu Leu Tyr Phe Ser Thr Phe Thr Thr Phe
[1697]	500 505 510
[1698]	Glu Ile Pro Ser Ser Ile Leu Met Val Asn Ala Lys Leu Tyr Ala Val
[1699]	515 520 525
[1700]	Gln Ala Asp Gly Lys Arg Ile Leu Leu Gly Ser Gly His His His His
[1701]	530 535 540
[1702]	His His
[1703]	545
[1704]	<210> 31
[1705]	<211> 517
[1706]	<212> PRT
[1707]	<213> Artificial Sequence
[1708]	<220>
[1709]	<223> polypeptide having O-glycoprotein-specific endoprotease activity
[1710]	(including N-terminal methionine and a C-terminal linker + His6
[1711]	tag)
[1712]	<400> 31
[1713]	Met Val Leu Glu Leu Glu Met Arg Gly Asp Ser Ile Ser Glu Ala Lys
[1714]	1 5 10 15
[1715]	Lys Arg Lys Val Trp Asn Phe Gln Asp Trp Gln Ile Thr Gly Leu Ser
[1716]	20 25 30
[1717]	Ala Arg Ala Gly Asp Lys Ile Thr Val Tyr Val Asp Val Ala Glu Gly
[1718]	35 40 45
[1719]	Asp Pro Thr Pro Thr Leu Leu Tyr Lys Gln Ser Leu Thr Gln His Gly
[1720]	50 55 60
[1721]	Gly Ala Thr Ser Phe Gln Leu Lys Pro Gly Lys Asn Glu Ile Thr Ile

[1722]	65	70	75	80
[1723]	Pro Glu Ile Asn Tyr Glu Ser Asn Gly Ile Pro Lys Asp Val Ile Gln			
[1724]		85	90	95
[1725]	Gly Gly Asp Leu Phe Phe Thr Asn Tyr Lys Ser Asp Ser Gln Lys Arg			
[1726]		100	105	110
[1727]	Ala Pro Lys Val Arg Ile Glu Gly Ala Ser Lys Tyr Pro Val Phe Ile			
[1728]		115	120	125
[1729]	Leu Gly Lys Ser Asp Glu Asn Glu Val Met Lys Glu Leu Glu Ala Tyr			
[1730]		130	135	140
[1731]	Val Glu Lys Ile Lys Ala Glu Pro Lys Thr Thr Pro Asn Ile Phe Ala			
[1732]	145	150	155	160
[1733]	Val Ser Ser Asn Lys Ser Leu Glu Phe Val Gln Ala Thr Tyr Ala Leu			
[1734]		165	170	175
[1735]	Asp Trp Tyr Lys Lys Asn Asn Lys Thr Pro Lys Tyr Thr Ala Glu Gln			
[1736]		180	185	190
[1737]	Trp Asp Gln Tyr Ile Ala Asp Ala Met Gly Phe Trp Gly Phe Asp Asn			
[1738]		195	200	205
[1739]	Ser Lys Asp Val Asn Ser Asp Phe Asn Phe Arg Ile Met Pro Met Val			
[1740]		210	215	220
[1741]	Lys Asn Leu Ser Gly Gly Ala Phe Met Asn Ala Gly Asn Gly Val Ile			
[1742]	225	230	235	240
[1743]	Gly Ile Arg Pro Gly Asn Gln Asp Ala Ile Leu Ala Ala Asn Lys Gly			
[1744]		245	250	255
[1745]	Trp Gly Val Ala His Glu Leu Gly His Asn Phe Asp Thr Gly Gly Arg			
[1746]		260	265	270
[1747]	Thr Ile Val Glu Val Thr Asn Asn Met Met Pro Leu Phe Phe Glu Ser			
[1748]		275	280	285
[1749]	Lys Tyr Lys Thr Lys Thr Arg Ile Thr Asp Gln Asn Ile Trp Glu Asn			
[1750]		290	295	300
[1751]	Asn Thr Tyr Pro Lys Val Gly Leu Asp Asp Tyr Ser Asn Asn Glu Leu			
[1752]	305	310	315	320
[1753]	Tyr Asn Lys Ala Asp Ser Thr His Leu Ala Gln Leu Ala Pro Leu Trp			
[1754]		325	330	335
[1755]	Gln Leu Tyr Leu Tyr Asp Asn Thr Phe Tyr Gly Lys Phe Glu Arg Gln			
[1756]		340	345	350
[1757]	Phe Arg Glu Arg Asp Phe Gly Asn Lys Asn Arg Glu Asp Ile Tyr Lys			
[1758]		355	360	365
[1759]	Ser Trp Val Val Ala Ala Ser Asp Ala Met Glu Leu Asp Leu Thr Glu			
[1760]		370	375	380
[1761]	Phe Phe Ala Arg His Gly Ile Arg Val Asp Asp Lys Val Lys Glu Asp			
[1762]	385	390	395	400
[1763]	Leu Ala Lys Tyr Pro Lys Pro Asp Lys Lys Ile Tyr Tyr Leu Asn Asp			

[1764]		405		410		415
[1765]	Leu	Ala	Met	Asn	Tyr	Val
[1766]		420		425		430
[1767]	Ser	Val	Ser	Thr	Ser	Gly
[1768]		435		440		445
[1769]	Val	Asp	Asp	Glu	Asn	Lys
[1770]		450		455		460
[1771]	Asp	Gly	Lys	Tyr	Val	Gly
[1772]	465			470		475
[1773]	Lys	Ser	Asn	Leu	Asp	Glu
[1774]		485		490		495
[1775]	Arg	Lys	Leu	Asn	Thr	Leu
[1776]		500		505		510
[1777]	His	His	His	His	His	
[1778]		515				
[1779]	<210>	32				
[1780]	<211>	923				
[1781]	<212>	PRT				
[1782]	<213>	Pseudomonas aeruginosa				
[1783]	<400>	32				
[1784]	Met	Ser	Leu	Ser	Thr	Thr
[1785]	1		5		10	
[1786]	Ser	Arg	Ser	Pro	Ile	Pro
[1787]		20		25		30
[1788]	Leu	Ala	Gly	Ala	Leu	Ser
[1789]		35		40		45
[1790]	Asp	Ala	Ala	Leu	Val	Ser
[1791]		50		55		60
[1792]	Leu	Val	Ala	Leu	Arg	Leu
[1793]	65			70		75
[1794]	Arg	Thr	Gln	Leu	Leu	Asp
[1795]		85		90		95
[1796]	Asp	Pro	Gly	Ala	Ala	Ser
[1797]		100		105		110
[1798]	Thr	Leu	Pro	Phe	Leu	Ile
[1799]		115		120		125
[1800]	Ser	Leu	Glu	Ala	Gly	Gly
[1801]		130		135		140
[1802]	Thr	Gln	Leu	Ser	Gly	Thr
[1803]	145			150		155
[1804]	Ala	Val	Gln	Trp	Leu	Val
[1805]		165		170		175

[1806]	Asp Phe Lys Val Ser Val Val Gly Val Asp Lys Thr Ala Ala Leu Asn		
[1807]		180	185
[1808]	Gly Leu Lys Ser Ala Gly Leu Gln Pro Ala Asp Ala Ala Cys Asn Ala		190
[1809]		195	200
[1810]	Leu Thr Asp Ala Ser Cys Ala Ser Thr Ser Lys Leu Leu Val Leu Gly		205
[1811]		210	215
[1812]	Asn Gly Ala Ser Ala Ala Ser Leu Ser Ala Thr Val Arg Ala Arg Leu		220
[1813]		225	230
[1814]	Gln Ala Gly Leu Pro Ile Leu Phe Val His Thr Asn Gly Trp Asn Gln		235
[1815]		245	250
[1816]	Ser Ser Thr Gly Gln Gln Ile Leu Ala Gly Leu Gly Leu Gln Glu Gly		255
[1817]		260	265
[1818]	Pro Tyr Gly Gly Asn Tyr Trp Asp Lys Asp Arg Val Pro Ser Ser Arg		270
[1819]		275	280
[1820]	Thr Arg Thr Arg Ser Val Glu Leu Gly Gly Ala Tyr Gly Gln Asp Pro		285
[1821]		290	295
[1822]	Ala Leu Val Gln Gln Ile Val Asp Gly Ser Trp Arg Thr Asp Tyr Asp		300
[1823]		305	310
[1824]	Trp Ser Lys Cys Thr Ser Tyr Val Gly Arg Thr Thr Cys Asp Asp Val		315
[1825]		325	330
[1826]	Pro Gly Leu Ser Asp Phe Ser Lys Arg Val Asp Val Leu Lys Gly Ala		335
[1827]		340	345
[1828]	Leu Asp Ala Tyr Asn Gln Lys Ala Gln Asn Leu Phe Ala Leu Pro Gly		350
[1829]		355	360
[1830]	Thr Thr Ser Leu Arg Leu Trp Leu Leu Trp Ala Asp Ala Val Arg Gln		365
[1831]		370	375
[1832]	Asn Ile Arg Tyr Pro Met Asp Lys Ala Ala Asp Thr Ala Arg Phe Gln		380
[1833]		385	390
[1834]	Glu Thr Phe Val Ala Asp Ala Ile Val Gly Tyr Val Arg Glu Ala Gly		395
[1835]		405	410
[1836]	Ala Ala Gln Lys Glu Leu Gly Ser Tyr Ala Gly Gln Arg Gln Gln Ser		415
[1837]		420	425
[1838]	Met Pro Val Ser Gly Ser Glu Glu Thr Leu Thr Leu Thr Leu Pro Ser		430
[1839]		435	440
[1840]	Ala Gln Gly Phe Thr Ala Ile Gly Arg Met Ala Ala Pro Gly Lys Arg		445
[1841]		450	455
[1842]	Leu Ser Ile Arg Ile Glu Asp Ala Gly Gln Ala Ser Leu Ala Val Gly		460
[1843]		465	470
[1844]	Leu Asn Thr Gln Arg Ile Gly Ser Thr Arg Leu Trp Asn Thr Arg Gln		475
[1845]		485	490
[1846]	Tyr Asp Arg Pro Arg Phe Leu Lys Ser Pro Asp Ile Lys Leu Gln Ala		495
[1847]		500	505
			510

[1848]	Asn Gln Ser Val Ala Leu Val Ser Pro Tyr Gly Gly Leu Leu Gln Leu
[1849]	515 520 525
[1850]	Val Tyr Ser Gly Ala Thr Pro Gly Gln Thr Val Thr Val Lys Val Thr
[1851]	530 535 540
[1852]	Gly Ala Ala Ser Gln Pro Phe Leu Asp Ile Gln Pro Gly Glu Asp Ser
[1853]	545 550 555 560
[1854]	Ser Gln Ala Ile Ala Asp Phe Ile Gln Ala Leu Asp Ala Asp Lys Ala
[1855]	565 570 575
[1856]	Asp Trp Leu Glu Ile Arg Ser Gly Ser Val Glu Val His Ala Lys Val
[1857]	580 585 590
[1858]	Glu Lys Val Arg Gly Ser Ile Asp Lys Asp Tyr Gly Gly Asp Val Gln
[1859]	595 600 605
[1860]	Arg Phe Ile Arg Glu Leu Asn Glu Val Phe Ile Asp Asp Ala Tyr Thr
[1861]	610 615 620
[1862]	Leu Ala Gly Phe Ala Ile Pro Asn Gln Ala Lys Thr Pro Ala Ile Gln
[1863]	625 630 635 640
[1864]	Gln Glu Cys Ala Ala Arg Gly Trp Asp Cys Asp Ser Glu Thr Leu His
[1865]	645 650 655
[1866]	Lys Leu Pro Gly Thr Gln His Ile Asn Val Asp Gln Tyr Ala Gln Cys
[1867]	660 665 670
[1868]	Gly Gly Gly Cys Ser Gly Asn Pro Tyr Asp Gln Thr Trp Gly Leu Asn
[1869]	675 680 685
[1870]	Pro Arg Gly Trp Gly Glu Ser His Glu Leu Gly His Asn Leu Gln Val
[1871]	690 695 700
[1872]	Asn Arg Leu Lys Val Tyr Gly Gly Arg Ser Gly Glu Ile Ser Asn Gln
[1873]	705 710 715 720
[1874]	Ile Phe Pro Leu His Lys Asp Trp Arg Val Leu Arg Glu Phe Gly Gln
[1875]	725 730 735
[1876]	Asn Leu Asp Asp Thr Arg Val Asn Tyr Arg Asn Ala Tyr Asn Leu Ile
[1877]	740 745 750
[1878]	Val Ala Gly Arg Ala Glu Ala Asp Pro Leu Ala Gly Val Tyr Lys Arg
[1879]	755 760 765
[1880]	Leu Trp Glu Asp Pro Gly Thr Tyr Ala Leu Asn Gly Glu Arg Met Ala
[1881]	770 775 780
[1882]	Phe Tyr Thr Gln Trp Val His Tyr Trp Ala Asp Leu Lys Asn Asp Pro
[1883]	785 790 795 800
[1884]	Leu Gln Gly Trp Asp Ile Trp Thr Leu Leu Tyr Leu His Gln Arg Gln
[1885]	805 810 815
[1886]	Val Asp Lys Ser Asp Trp Asp Ala Asn Lys Ala Ala Leu Gly Tyr Gly
[1887]	820 825 830
[1888]	Thr Tyr Ala Gln Arg Pro Gly Asn Ser Gly Asp Ala Ser Ser Thr Asp
[1889]	835 840 845

[1890]	Gly Asn Asp Asn Leu Leu Leu Gly Leu Ser Trp Leu Thr Gln Arg Asp
[1891]	850 855 860
[1892]	Gln Arg Pro Thr Phe Ala Leu Trp Gly Ile Arg Thr Ser Ala Ala Ala
[1893]	865 870 875 880
[1894]	Gln Ala Gln Val Ala Ala Tyr Gly Phe Ala Glu Gln Pro Ala Phe Phe
[1895]	885 890 895
[1896]	Tyr Ala Asn Asn Arg Thr Asn Glu Tyr Ser Thr Val Lys Leu Leu Asp
[1897]	900 905 910
[1898]	Met Ser Gln Gly Ser Pro Ala Trp Pro Phe Pro
[1899]	915 920
[1900]	<210> 33
[1901]	<211> 857
[1902]	<212> PRT
[1903]	<213> Bacteroides thetaiotaomicron
[1904]	<400> 33
[1905]	Met Thr Ile Lys Arg Phe Ile Thr Asn Leu Leu Ala Leu Phe Thr Leu
[1906]	1 5 10 15
[1907]	Phe Thr Val Ser Leu Ala Cys Lys Asp Thr Glu Lys Ser Ile Ile Asn
[1908]	20 25 30
[1909]	Ser Ser Phe Ser Ile Ser Glu Glu Tyr Leu Ile Gln Asn Leu Asp Lys
[1910]	35 40 45
[1911]	Ser Ser Thr Ser Val Gln Ile Pro Ile Asn Thr Ser Met Glu Leu Ala
[1912]	50 55 60
[1913]	Gln Trp Ser Val Ser Tyr Glu Ala Asn Trp Leu Gln Cys Ser Lys Gln
[1914]	65 70 75 80
[1915]	Lys Thr Ala Ala Glu Gly Thr Phe Leu Arg Ile Thr Val Asn Glu Asn
[1916]	85 90 95
[1917]	Thr Gly Glu Thr Lys Arg Thr Ala Asn Ile Lys Val Thr Ser Thr Thr
[1918]	100 105 110
[1919]	Ala Thr Tyr Thr Ile Thr Val Asn Gln Tyr Ala Lys Gly Glu Val Ile
[1920]	115 120 125
[1921]	Val Glu Gly Asp Ile Lys Val Thr Pro Thr Gly Gly Lys Ala Ser Glu
[1922]	130 135 140
[1923]	His Gln Glu Gly Gln Asp Ile Glu Asn Thr Tyr Asp Gly Lys Phe Ser
[1924]	145 150 155 160
[1925]	Thr Asp Gly Ala Ala Pro Phe His Thr Pro Trp Gly Gln Ser Ala Lys
[1926]	165 170 175
[1927]	Phe Pro Val Thr Leu Glu Tyr Tyr Phe Lys Gly Asp Thr Glu Ile Asp
[1928]	180 185 190
[1929]	Tyr Leu Ile Tyr Tyr Thr Arg Ser Gly Asn Gly Asn Phe Gly Lys Val
[1930]	195 200 205
[1931]	Lys Val Tyr Thr Thr Thr Asn Pro Asp Arg Ser Asp Tyr Thr Leu Gln

[1932]	210	215	220
[1933]	Gly Glu Tyr Asp Phe Lys Glu Gln Asn Ala Pro Ser Lys Val Ser Phe		
[1934]	225	230	235 240
[1935]	Ser Glu Gly Ile Lys Ala Thr Gly Ile Lys Phe Glu Val Leu Ser Gly		
[1936]	245	250	255
[1937]	Leu Gly Asp Phe Val Ser Cys Asp Glu Met Glu Phe Tyr Lys Thr Asn		
[1938]	260	265	270
[1939]	Thr Asp Lys Thr Leu Asp Lys Gln Leu Leu Thr Val Phe Thr Asp Ile		
[1940]	275	280	285
[1941]	Thr Cys Thr Glu Ile Lys Asn Asn Val Thr Asn Glu Gln Ile Gln Ala		
[1942]	290	295	300
[1943]	Leu Pro Asp Tyr Phe Val Arg Ile Ala Glu Ala Val Arg Asp Asn Thr		
[1944]	305	310	315 320
[1945]	Tyr Asp Lys Trp Glu Lys Glu Phe Arg Ile Arg Ser Tyr Glu Pro Tyr		
[1946]	325	330	335
[1947]	Ser Asn Ile Ala Glu Trp Ala Asp Lys Leu Met Thr Lys Lys Tyr Ser		
[1948]	340	345	350
[1949]	Asp Leu Asp Asn Pro Thr Gly Ile Ser Val Lys Ala Gly Asp Asp Ile		
[1950]	355	360	365
[1951]	Ile Val Leu Val Gly Asp Thr Tyr Gly Gln Asn Ile Ser Met Gln Cys		
[1952]	370	375	380
[1953]	Ile Trp Glu Thr Gly Thr Glu Tyr Lys Gln Thr Ala Ser Ser Gly Asp		
[1954]	385	390	395 400
[1955]	Val Tyr Met Leu Asn Pro Gly Val Asn Lys Leu Thr Met Lys Gly Glu		
[1956]	405	410	415
[1957]	Gly Gln Leu Phe Val Met Tyr Asn Thr Glu Leu Thr Ser Asn Thr Ala		
[1958]	420	425	430
[1959]	Lys Pro Ile Lys Ile His Ile Pro Leu Gly Ser Gly Thr Val Asn Gly		
[1960]	435	440	445
[1961]	Phe Phe Asp Leu Lys Glu His Lys Thr Asp Glu Lys Tyr Ala Glu Leu		
[1962]	450	455	460
[1963]	Leu Lys Lys Ser Thr His Lys Tyr Phe Cys Ile Arg Gly Glu Lys Ile		
[1964]	465	470	475 480
[1965]	Met Phe Tyr Phe His Arg Asn Lys Leu Leu Glu Tyr Val Pro Asn Asn		
[1966]	485	490	495
[1967]	Ile Leu Ser Ala Ile His Leu Trp Asp Asn Ile Val Gly Trp Gln Gln		
[1968]	500	505	510
[1969]	Glu Leu Met Gly Ile Asp Asp Val Arg Pro Ser Gln Val Asn Asn His		
[1970]	515	520	525
[1971]	Leu Phe Ala Ile Ser Pro Glu Gly Ser Tyr Met Trp Ala Ser Asp Tyr		
[1972]	530	535	540
[1973]	Gln Ile Gly Phe Val Tyr Thr Tyr Leu Gly Asn Ile Leu Leu Glu Asp		

[1974]	545	550	555	560
[1975]	Asn Val Met Ala Ala Glu Asp Asn Ala Trp Gly Pro Ala His Glu Ile			
[1976]		565	570	575
[1977]	Gly His Val His Gln Ala Ala Ile Asn Trp Ala Ser Ser Thr Glu Ser			
[1978]		580	585	590
[1979]	Ser Asn Asn Leu Phe Ser Asn Phe Ile Ile Tyr Lys Leu Gly Lys Tyr			
[1980]		595	600	605
[1981]	Lys Ser Arg Gly Asn Gly Leu Gly Ser Val Ala Thr Ala Arg Tyr Ala			
[1982]	610	615	620	
[1983]	Asn Gly Gln Ala Trp Tyr Asn Met Gly Asp Ala Thr His Gln Asn Glu			
[1984]	625	630	635	640
[1985]	Asp Thr Glu Thr His Met Arg Met Asn Trp Gln Leu Trp Ile Tyr Tyr			
[1986]		645	650	655
[1987]	His Arg Cys Glu Tyr Lys Thr Asp Phe Trp Gln Thr Leu Phe Lys Leu			
[1988]		660	665	670
[1989]	Met Arg Glu Val Asn Met Thr Glu Gly Glu Asp Pro Gly Lys Lys Gln			
[1990]		675	680	685
[1991]	Leu Glu Phe Ala Lys Met Ala Ser Lys Ala Ala Asn Gln Asn Leu Thr			
[1992]	690	695	700	
[1993]	Asp Phe Phe Glu Met Trp Gly Phe Phe Glu Pro Val Asn Thr Thr Ile			
[1994]	705	710	715	720
[1995]	Glu Gln Tyr Gly Thr Tyr Lys Tyr Tyr Val Ser Asp Ala Met Ile Arg			
[1996]		725	730	735
[1997]	Glu Ala Lys Glu Tyr Met Ala Gln Phe Pro Ala Pro Lys His Ala Phe			
[1998]		740	745	750
[1999]	Gln Tyr Ile Glu Asp Arg Lys Lys Ser Glu Phe Pro Ser Asn Asp Tyr			
[2000]		755	760	765
[2001]	Arg Tyr Ser Ala Val Gly Asp Val Gly Tyr Tyr Thr Gln Phe Lys Glu			
[2002]		770	775	780
[2003]	Asn Gln Lys Ile Thr Lys Ala Ile Thr Ala Glu Leu Ala Gly Arg Lys			
[2004]	785	790	795	800
[2005]	Val Ser Ile Gln Asn Gly Asp Glu Ala Val Ala Phe Glu Leu Arg Glu			
[2006]		805	810	815
[2007]	Asn Asp Glu Asn Gly Lys Leu Leu Tyr Phe Ser Thr Phe Thr Thr Phe			
[2008]		820	825	830
[2009]	Glu Ile Pro Ser Ser Ile Leu Met Val Asn Ala Lys Leu Tyr Ala Val			
[2010]		835	840	845
[2011]	Gln Ala Asp Gly Lys Arg Ile Leu Leu			
[2012]		850	855	
[2013]	<210> 34			
[2014]	<211> 1687			
[2015]	<212> PRT			

[2016]	<213> Clostridium perfringens															
[2017]	<400> 34															
[2018]	Met	Asn	Lys	Arg	Lys	Ile	Ala	Ala	Ile	Ile	Leu	Ala	Thr	Met	Ile	Thr
[2019]	1				5					10					15	
[2020]	Asn	Leu	Ser	Ala	Thr	Thr	Ile	Asp	Val	Leu	Ala	Gln	Glu	Leu	Asn	Thr
[2021]					20				25					30		
[2022]	Lys	Asn	Asn	Ser	Lys	Val	Glu	Val	Ser	His	Asp	Asp	Glu	Ser	His	Gln
[2023]					35				40				45			
[2024]	Ala	Arg	Val	Ser	Lys	Phe	Asp	Leu	Tyr	Asn	Ser	Asp	Lys	Leu	Asp	Ala
[2025]		50					55					60				
[2026]	Tyr	Asn	Gln	Glu	Phe	Gln	Val	Ser	Arg	Ser	Asn	Ile	Lys	Ser	Ile	Asn
[2027]	65					70					75				80	
[2028]	Asn	Asn	Gly	Gly	Lys	Tyr	Asn	Ser	Ser	Thr	Ile	Asp	Lys	Ala	Ile	Asp
[2029]					85					90					95	
[2030]	Gly	Asn	Leu	Glu	Thr	His	Trp	Glu	Thr	Gly	Lys	Pro	Asn	Asp	Ala	Asn
[2031]					100					105					110	
[2032]	Phe	Thr	Asn	Glu	Val	Val	Val	Thr	Phe	Asn	Glu	Ile	Thr	Asn	Ile	Asp
[2033]					115					120					125	
[2034]	Arg	Ile	Val	Tyr	Ser	Ala	Arg	Arg	Asp	Ser	Ala	Arg	Gly	Lys	Gly	Phe
[2035]		130					135					140				
[2036]	Ala	Lys	Glu	Phe	Glu	Ile	Tyr	Ala	Ser	Leu	Lys	Asp	Glu	Gly	Asp	Asp
[2037]	145					150					155				160	
[2038]	Phe	Asn	Leu	Val	Ser	Ser	Gly	Glu	Tyr	Thr	Glu	Ser	Thr	Arg	Asp	Leu
[2039]					165					170					175	
[2040]	Val	Glu	Ile	Lys	Phe	Asn	Pro	Thr	Asp	Phe	Lys	Arg	Leu	Lys	Phe	Lys
[2041]					180					185					190	
[2042]	Phe	Lys	Lys	Ala	Asp	Gln	Asn	Trp	Ala	Ser	Ala	Ala	Glu	Phe	Met	Phe
[2043]			195					200					205			
[2044]	Tyr	Lys	Glu	Asp	Lys	Leu	Asn	Glu	Lys	Phe	Asn	Gly	Leu	Phe	Thr	Asp
[2045]		210					215					220				
[2046]	Ser	Ser	Met	Asn	Lys	Val	Ser	Glu	Glu	Phe	Asn	Thr	Leu	Glu	Lys	Leu
[2047]	225					230					235				240	
[2048]	Asn	Ala	Phe	Glu	Asn	Glu	Leu	Lys	Asp	His	Pro	Ile	Tyr	Asp	Leu	Tyr
[2049]					245					250					255	
[2050]	Lys	Glu	Gly	Leu	Asn	Asn	Ala	Arg	Ala	Ile	Leu	Thr	Glu	Thr	Ser	Glu
[2051]					260					265					270	
[2052]	Asn	Pro	Thr	Lys	Ala	Thr	Leu	Gly	Gln	Ile	Thr	Tyr	Asn	Leu	Asn	Asp
[2053]					275					280				285		
[2054]	Asp	Tyr	Asn	Asn	Gln	Tyr	Arg	Met	Pro	Tyr	Lys	Asn	Ile	Lys	Ala	Ile
[2055]		290					295					300				
[2056]	Lys	Asn	Asn	Gly	Arg	His	Tyr	Ala	Ala	Gln	Asn	Ile	Glu	Lys	Ala	Ile
[2057]	305					310					315				320	

[2058]	Asp Asn Asp Val Asn Thr Tyr Trp Glu Thr Gly Thr Leu Asn Ser Ser		
[2059]		325	330 335
[2060]	Ser Phe Asn Asn Glu Val Glu Val Glu Phe Asn Asp Leu Val Thr Leu		
[2061]		340	345 350
[2062]	Asp Arg Ile Val Tyr Gly Ser Arg Gln Ser Asp Leu Lys Gly Phe Ala		
[2063]		355	360 365
[2064]	Glu Glu Val Tyr Ile Tyr Ala Ser Arg Thr Ser Lys Gly Asp Thr Tyr		
[2065]		370	375 380
[2066]	Lys Leu Val Ala Thr Gly Ala His Glu Ala Thr Lys Gly Leu Val Glu		
[2067]		385	390 395 400
[2068]	Ala Lys Phe Glu Pro Thr Glu Phe Lys Arg Val Lys Phe Lys Phe Lys		
[2069]		405	410 415
[2070]	Lys Ser Lys Gln Asn Ser Ala Thr Leu Asn Glu Leu Met Phe Tyr Lys		
[2071]		420	425 430
[2072]	Pro Asp Glu Val Tyr Ser Ser Ile Pro Lys Leu Phe Thr Asp Gly Thr		
[2073]		435	440 445
[2074]	Met Ser Glu Leu Ser Glu Glu Phe Asn Ser Leu Glu Lys Ile Asn Ala		
[2075]		450	455 460
[2076]	Phe Lys Glu Lys Ala Lys Asn His Pro Leu Tyr Asn Asp Phe Asn Glu		
[2077]		465	470 475 480
[2078]	Thr Ile Glu Leu Ala Glu Ser Leu Ile Ser Asn Pro Arg Lys Glu Asp		
[2079]		485	490 495
[2080]	Val Leu Glu Leu Glu Met Arg Gly Asp Ser Ile Ser Glu Ala Lys Lys		
[2081]		500	505 510
[2082]	Arg Lys Val Trp Asn Phe Gln Asp Trp Gln Ile Thr Gly Leu Ser Ala		
[2083]		515	520 525
[2084]	Arg Ala Gly Asp Lys Ile Thr Val Tyr Val Asp Val Ala Glu Gly Asp		
[2085]		530	535 540
[2086]	Pro Thr Pro Thr Leu Leu Tyr Lys Gln Ser Leu Thr Gln His Gly Gly		
[2087]		545	550 555 560
[2088]	Ala Thr Ser Phe Gln Leu Lys Pro Gly Lys Asn Glu Ile Thr Ile Pro		
[2089]		565	570 575
[2090]	Glu Ile Asn Tyr Glu Ser Asn Gly Ile Pro Lys Asp Val Ile Gln Gly		
[2091]		580	585 590
[2092]	Gly Asp Leu Phe Phe Thr Asn Tyr Lys Ser Asp Ser Gln Lys Arg Ala		
[2093]		595	600 605
[2094]	Pro Lys Val Arg Ile Glu Gly Ala Ser Lys Tyr Pro Val Phe Ile Leu		
[2095]		610	615 620
[2096]	Gly Lys Ser Asp Glu Asn Glu Val Met Lys Glu Leu Glu Ala Tyr Val		
[2097]		625	630 635 640
[2098]	Glu Lys Ile Lys Ala Glu Pro Lys Thr Thr Pro Asn Ile Phe Ala Val		
[2099]		645	650 655

[2100]	Ser Ser Asn Lys Ser Leu Glu Phe Val Gln Ala Thr Tyr Ala Leu Asp
[2101]	660 665 670
[2102]	Trp Tyr Lys Lys Asn Asn Lys Thr Pro Lys Tyr Thr Ala Glu Gln Trp
[2103]	675 680 685
[2104]	Asp Gln Tyr Ile Ala Asp Ala Met Gly Phe Trp Gly Phe Asp Asn Ser
[2105]	690 695 700
[2106]	Lys Asp Val Asn Ser Asp Phe Asn Phe Arg Ile Met Pro Met Val Lys
[2107]	705 710 715 720
[2108]	Asn Leu Ser Gly Gly Ala Phe Met Asn Ala Gly Asn Gly Val Ile Gly
[2109]	725 730 735
[2110]	Ile Arg Pro Gly Asn Gln Asp Ala Ile Leu Ala Ala Asn Lys Gly Trp
[2111]	740 745 750
[2112]	Gly Val Ala His Glu Leu Gly His Asn Phe Asp Thr Gly Gly Arg Thr
[2113]	755 760 765
[2114]	Ile Val Glu Val Thr Asn Asn Met Met Pro Leu Phe Phe Glu Ser Lys
[2115]	770 775 780
[2116]	Tyr Lys Thr Lys Thr Arg Ile Thr Asp Gln Asn Ile Trp Glu Asn Asn
[2117]	785 790 795 800
[2118]	Thr Tyr Pro Lys Val Gly Leu Asp Asp Tyr Ser Asn Asn Glu Leu Tyr
[2119]	805 810 815
[2120]	Asn Lys Ala Asp Ser Thr His Leu Ala Gln Leu Ala Pro Leu Trp Gln
[2121]	820 825 830
[2122]	Leu Tyr Leu Tyr Asp Asn Thr Phe Tyr Gly Lys Phe Glu Arg Gln Phe
[2123]	835 840 845
[2124]	Arg Glu Arg Asp Phe Gly Asn Lys Asn Arg Glu Asp Ile Tyr Lys Ser
[2125]	850 855 860
[2126]	Trp Val Val Ala Ala Ser Asp Ala Met Glu Leu Asp Leu Thr Glu Phe
[2127]	865 870 875 880
[2128]	Phe Ala Arg His Gly Ile Arg Val Asp Asp Lys Val Lys Glu Asp Leu
[2129]	885 890 895
[2130]	Ala Lys Tyr Pro Lys Pro Asp Lys Lys Ile Tyr Tyr Leu Asn Asp Leu
[2131]	900 905 910
[2132]	Ala Met Asn Tyr Lys Gly Asp Gly Phe Thr Glu Asn Ala Lys Val Ser
[2133]	915 920 925
[2134]	Val Ser Thr Ser Gly Ser Asn Gly Asn Ile Lys Leu Ser Phe Ser Val
[2135]	930 935 940
[2136]	Asp Asp Glu Asn Lys Asp Asn Ile Leu Gly Tyr Glu Ile Arg Arg Asp
[2137]	945 950 955 960
[2138]	Gly Lys Tyr Val Gly Phe Thr Ser Asn Asp Ser Phe Val Asp Thr Lys
[2139]	965 970 975
[2140]	Ser Asn Leu Asp Glu Asp Gly Val Tyr Val Val Thr Pro Tyr Asp Arg
[2141]	980 985 990

[2142]	Lys Leu Asn Thr Leu Asn Pro Ile Glu Val Asn Ala Leu Gln Pro Thr		
[2143]	995	1000	1005
[2144]	Leu Ser Val Asn Pro Val Ile Thr Leu Ala Leu Gly Glu Glu Phe		
[2145]	1010	1015	1020
[2146]	Asn Glu Glu Glu Tyr Ile Val Ala Lys Asp Ile Lys Gly Asn Ser		
[2147]	1025	1030	1035
[2148]	Leu Ser Glu Ser Val Lys Val Lys Ser Ser Asn Val Asn Thr Ser		
[2149]	1040	1045	1050
[2150]	Lys Val Gly Glu Tyr Glu Val Leu Tyr Ser Leu Glu Asp Ser Lys		
[2151]	1055	1060	1065
[2152]	Gly Asn Glu Tyr Thr Lys Thr Ser Lys Val Asn Val Val Ser Arg		
[2153]	1070	1075	1080
[2154]	Lys Glu Tyr Met Ser Asp Leu Thr Pro Lys Gln Ser Ser Asn Gly		
[2155]	1085	1090	1095
[2156]	Trp Gly Thr Val Arg Lys Asp Lys Ser Ile Ser Gly Gly Val Ile		
[2157]	1100	1105	1110
[2158]	Gly Leu Thr Arg Asp Gly Asp Phe Val Asp Tyr Asn Lys Gly Leu		
[2159]	1115	1120	1125
[2160]	Gly Leu His Ser Asn Ala Glu Tyr Val Tyr Asp Leu Glu Gly Lys		
[2161]	1130	1135	1140
[2162]	Asp Tyr Asp Tyr Phe Glu Ser Tyr Val Gly Val Asp Lys Ala Met		
[2163]	1145	1150	1155
[2164]	Ser Ser Arg Pro Ala Ser Ser Val Ile Phe Lys Val Leu Val Asp		
[2165]	1160	1165	1170
[2166]	Gly Glu Glu Lys Phe Asn Ser Gly Val Met Arg Ser Thr Thr Pro		
[2167]	1175	1180	1185
[2168]	Gln Lys Tyr Val Lys Val Asp Val Lys Asn Ala Lys Glu Leu Lys		
[2169]	1190	1195	1200
[2170]	Leu Ile Val Asn Asp Ala Gly Asp Gly Asp Ser Ser Asp His Ala		
[2171]	1205	1210	1215
[2172]	Ser Phe Gly Asp Ala Lys Leu Ala Thr Leu Ser Ser Lys Pro Ile		
[2173]	1220	1225	1230
[2174]	Ile Lys Gly Glu Asn Leu Ala Tyr Ser Met Asp Glu Lys Val Asp		
[2175]	1235	1240	1245
[2176]	Leu Met Lys Gly Ile Thr Ala Thr Asp Ile Glu Asp Gly Asn Ile		
[2177]	1250	1255	1260
[2178]	Thr Ser Lys Val Gln Ile Lys Ser Ser Asp Phe Val Glu Gly Lys		
[2179]	1265	1270	1275
[2180]	Ser Gly Ile Phe Thr Val Val Tyr Ser Val Thr Asp Ser Asp Gly		
[2181]	1280	1285	1290
[2182]	Leu Thr Ser Glu Cys Ser Arg Thr Ile Ala Val Thr Asp Lys Glu		
[2183]	1295	1300	1305

[2184]	Thr Gln Leu Ser Asp Leu Asn Trp Lys Ser Ala Thr Ile Gly Ser
[2185]	1310 1315 1320
[2186]	Gly Ser Val Arg Lys Asp Arg Ala Val Ser Gly Asn Gln Ile Arg
[2187]	1325 1330 1335
[2188]	Leu Leu Asn Glu Asp Asn Ser Val Glu Thr Phe Ala Lys Gly Ile
[2189]	1340 1345 1350
[2190]	Gly Thr His Ser Tyr Ser Glu Ile Val Tyr Asn Ser Glu Gly Tyr
[2191]	1355 1360 1365
[2192]	Asp Ile Phe Asp Thr Trp Val Gly Ile Asp Arg His Val Ala Asp
[2193]	1370 1375 1380
[2194]	Lys Lys Val Ser Ser Val Lys Phe Lys Val Tyr Val Asp Gly Glu
[2195]	1385 1390 1395
[2196]	Leu Lys Ala Glu Thr Asp Val Met Arg Ile Asp Thr Pro Lys Lys
[2197]	1400 1405 1410
[2198]	Arg Leu Val Val Asp Val Arg Asn Ser Lys Glu Ile Lys Leu Val
[2199]	1415 1420 1425
[2200]	Val Asp Val Ala Asp Asn Gly Asn Asn Trp Asp His Ala Asp Trp
[2201]	1430 1435 1440
[2202]	Ala Asp Ala Lys Phe Arg Asn Leu Ala Glu Tyr Asp Ala Ser Glu
[2203]	1445 1450 1455
[2204]	Leu Asn Lys Ala Ile Glu Glu Ala Lys Lys Leu Asp Leu Asn Asn
[2205]	1460 1465 1470
[2206]	Tyr Thr Glu Glu Ser Ser Glu Ala Leu Lys Asn Ala Ile Ser Lys
[2207]	1475 1480 1485
[2208]	Gly Glu Glu Ala Leu Leu Ser Lys Asp Lys Glu Thr Ile Asn Ser
[2209]	1490 1495 1500
[2210]	Ala Leu Glu Glu Leu Asn Lys Glu Met Asn Ser Leu Val Lys Val
[2211]	1505 1510 1515
[2212]	Asp Leu Asn Ala Val Ile Asn Ile Pro Asp Lys Tyr Leu Leu Lys
[2213]	1520 1525 1530
[2214]	Ser Ile Gln Asn Gln Leu Asn Lys Thr Gly Asp Ile Thr Leu Gly
[2215]	1535 1540 1545
[2216]	Asp Met Tyr Ser Leu Thr Thr Leu Thr Leu Ser Gly Val Glu Asp
[2217]	1550 1555 1560
[2218]	Leu Thr Gly Leu Glu Asn Ala Lys Asn Leu Glu Thr Leu Asn Met
[2219]	1565 1570 1575
[2220]	Asp Tyr Asn Glu Val Lys Asp Leu Arg Pro Leu Ser Lys Leu Lys
[2221]	1580 1585 1590
[2222]	Lys Leu Asn Thr Leu Asn Ala Gln Glu Gln Phe Ile Ala Ala Gly
[2223]	1595 1600 1605
[2224]	Glu Leu Lys Pro Ser Asn Gly Lys Val Ile Gly Asp Ser Lys Val
[2225]	1610 1615 1620

[2226]	Tyr Asn Arg Glu Gly Lys Asn Val Ala Lys Thr Ile Arg Val Val
[2227]	1625 1630 1635
[2228]	Asp Lys Asn Gly Asn Thr Ile Leu Glu Gln Asp Ala Lys Asp Glu
[2229]	1640 1645 1650
[2230]	Phe Thr Ile Asn Thr Lys Asp Leu Ser Ser Gly Leu Tyr Gly Val
[2231]	1655 1660 1665
[2232]	His Val Leu Phe Glu Asp Glu Gly Phe Ser Gly Val Met Phe Tyr
[2233]	1670 1675 1680
[2234]	Leu Phe Asn Val
[2235]	1685
[2236]	<210> 35
[2237]	<211> 882
[2238]	<212> PRT
[2239]	<213> Artificial Sequence
[2240]	<220>
[2241]	<223> Polypeptide capable of binding to O-glycans but lack or has
[2242]	reduced O-glycoprotein-specific endoprotease activity -
[2243]	Pseudomonas aeruginosa PA01 (double mutant with removed signal
[2244]	sequence)
[2245]	<400> 35
[2246]	Ala Thr Gln Glu Glu Ile Leu Asp Ala Ala Leu Val Ser Gly Asp Ser
[2247]	1 5 10 15
[2248]	Ser Gln Leu Thr Asp Ser His Leu Val Ala Leu Arg Leu Gln Gln Gln
[2249]	20 25 30
[2250]	Val Glu Arg Ile Arg Gln Thr Arg Thr Gln Leu Leu Asp Gly Leu Tyr
[2251]	35 40 45
[2252]	Gln Asn Leu Ser Gln Ala Tyr Asp Pro Gly Ala Ala Ser Met Trp Val
[2253]	50 55 60
[2254]	Leu Pro Ala Asn Pro Asp Asn Thr Leu Pro Phe Leu Ile Gly Asp Lys
[2255]	65 70 75 80
[2256]	Gly Arg Val Leu Ala Ser Leu Ser Leu Glu Ala Gly Gly Arg Gly Leu
[2257]	85 90 95
[2258]	Ala Tyr Gly Thr Asn Val Leu Thr Gln Leu Ser Gly Thr Asn Ala Ala
[2259]	100 105 110
[2260]	His Ala Pro Leu Leu Lys Arg Ala Val Gln Trp Leu Val Asn Gly Asp
[2261]	115 120 125
[2262]	Pro Gly Ala Ala Thr Ala Lys Asp Phe Lys Val Ser Val Val Gly Val
[2263]	130 135 140
[2264]	Asp Lys Thr Ala Ala Leu Asn Gly Leu Lys Ser Ala Gly Leu Gln Pro
[2265]	145 150 155 160
[2266]	Ala Asp Ala Ala Cys Asn Ala Leu Thr Asp Ala Ser Cys Ala Ser Thr
[2267]	165 170 175

[2268]	Ser Lys Leu Leu Val Leu Gly Asn Gly Ala Ser Ala Ala Ser Leu Ser		
[2269]	180	185	190
[2270]	Ala Thr Val Arg Ala Arg Leu Gln Ala Gly Leu Pro Ile Leu Phe Val		
[2271]	195	200	205
[2272]	His Thr Asn Gly Trp Asn Gln Ser Ser Thr Gly Gln Gln Ile Leu Ala		
[2273]	210	215	220
[2274]	Gly Leu Gly Leu Gln Glu Gly Pro Tyr Gly Gly Asn Tyr Trp Asp Lys		
[2275]	225	230	235
[2276]	Asp Arg Val Pro Ser Ser Arg Thr Arg Thr Arg Ser Val Glu Leu Gly		
[2277]	245	250	255
[2278]	Gly Ala Tyr Gly Gln Asp Pro Ala Leu Val Gln Gln Ile Val Asp Gly		
[2279]	260	265	270
[2280]	Ser Trp Arg Thr Asp Tyr Asp Trp Ser Lys Cys Thr Ser Tyr Val Gly		
[2281]	275	280	285
[2282]	Arg Thr Thr Cys Asp Asp Val Pro Gly Leu Ser Asp Phe Ser Lys Arg		
[2283]	290	295	300
[2284]	Val Asp Val Leu Lys Gly Ala Leu Asp Ala Tyr Asn Gln Lys Ala Gln		
[2285]	305	310	315
[2286]	Asn Leu Phe Ala Leu Pro Gly Thr Thr Ser Leu Arg Leu Trp Leu Leu		
[2287]	325	330	335
[2288]	Trp Ala Asp Ala Val Arg Gln Asn Ile Arg Tyr Pro Met Asp Lys Ala		
[2289]	340	345	350
[2290]	Ala Asp Thr Ala Arg Phe Gln Glu Thr Phe Val Ala Asp Ala Ile Val		
[2291]	355	360	365
[2292]	Gly Tyr Val Arg Glu Ala Gly Ala Ala Gln Lys Glu Leu Gly Ser Tyr		
[2293]	370	375	380
[2294]	Ala Gly Gln Arg Gln Gln Ser Met Pro Val Ser Gly Ser Glu Glu Thr		
[2295]	385	390	395
[2296]	Leu Thr Leu Thr Leu Pro Ser Ala Gln Gly Phe Thr Ala Ile Gly Arg		
[2297]	405	410	415
[2298]	Met Ala Ala Pro Gly Lys Arg Leu Ser Ile Arg Ile Glu Asp Ala Gly		
[2299]	420	425	430
[2300]	Gln Ala Ser Leu Ala Val Gly Leu Asn Thr Gln Arg Ile Gly Ser Thr		
[2301]	435	440	445
[2302]	Arg Leu Trp Asn Thr Arg Gln Tyr Asp Arg Pro Arg Phe Leu Lys Ser		
[2303]	450	455	460
[2304]	Pro Asp Ile Lys Leu Gln Ala Asn Gln Ser Val Ala Leu Val Ser Pro		
[2305]	465	470	475
[2306]	Tyr Gly Gly Leu Leu Gln Leu Val Tyr Ser Gly Ala Thr Pro Gly Gln		
[2307]	485	490	495
[2308]	Thr Val Thr Val Lys Val Thr Gly Ala Ala Ser Gln Pro Phe Leu Asp		
[2309]	500	505	510

[2310]	Ile	Gln	Pro	Gly	Glu	Asp	Ser	Ser	Gln	Ala	Ile	Ala	Asp	Phe	Ile	Gln
[2311]			515					520					525			
[2312]	Ala	Leu	Asp	Ala	Asp	Lys	Ala	Asp	Trp	Leu	Glu	Ile	Arg	Ser	Gly	Ser
[2313]		530					535					540				
[2314]	Val	Glu	Val	His	Ala	Lys	Val	Glu	Lys	Val	Arg	Gly	Ser	Ile	Asp	Lys
[2315]	545					550					555					560
[2316]	Asp	Tyr	Gly	Gly	Asp	Val	Gln	Arg	Phe	Ile	Arg	Glu	Leu	Asn	Glu	Val
[2317]				565					570						575	
[2318]	Phe	Ile	Asp	Asp	Ala	Tyr	Thr	Leu	Ala	Gly	Phe	Ala	Ile	Pro	Asn	Gln
[2319]			580					585						590		
[2320]	Ala	Lys	Thr	Pro	Ala	Ile	Gln	Gln	Glu	Cys	Ala	Ala	Arg	Gly	Trp	Asp
[2321]		595					600						605			
[2322]	Cys	Asp	Ser	Glu	Thr	Leu	His	Lys	Leu	Pro	Gly	Thr	Gln	His	Ile	Asn
[2323]	610					615					620					
[2324]	Val	Asp	Gln	Tyr	Ala	Gln	Cys	Gly	Gly	Gly	Cys	Ser	Gly	Asn	Pro	Tyr
[2325]	625					630				635						640
[2326]	Asp	Gln	Thr	Trp	Gly	Leu	Asn	Pro	Arg	Gly	Trp	Gly	Glu	Ser	Ala	Ala
[2327]				645					650						655	
[2328]	Leu	Gly	His	Asn	Leu	Gln	Val	Asn	Arg	Leu	Lys	Val	Tyr	Gly	Gly	Arg
[2329]			660						665						670	
[2330]	Ser	Gly	Glu	Ile	Ser	Asn	Gln	Ile	Phe	Pro	Leu	His	Lys	Asp	Trp	Arg
[2331]		675					680						685			
[2332]	Val	Leu	Arg	Glu	Phe	Gly	Gln	Asn	Leu	Asp	Asp	Thr	Arg	Val	Asn	Tyr
[2333]	690					695						700				
[2334]	Arg	Asn	Ala	Tyr	Asn	Leu	Ile	Val	Ala	Gly	Arg	Ala	Glu	Ala	Asp	Pro
[2335]	705					710				715						720
[2336]	Leu	Ala	Gly	Val	Tyr	Lys	Arg	Leu	Trp	Glu	Asp	Pro	Gly	Thr	Tyr	Ala
[2337]				725					730						735	
[2338]	Leu	Asn	Gly	Glu	Arg	Met	Ala	Phe	Tyr	Thr	Gln	Trp	Val	His	Tyr	Trp
[2339]			740						745						750	
[2340]	Ala	Asp	Leu	Lys	Asn	Asp	Pro	Leu	Gln	Gly	Trp	Asp	Ile	Trp	Thr	Leu
[2341]		755					760						765			
[2342]	Leu	Tyr	Leu	His	Gln	Arg	Gln	Val	Asp	Lys	Ser	Asp	Trp	Asp	Ala	Asn
[2343]	770					775						780				
[2344]	Lys	Ala	Ala	Leu	Gly	Tyr	Gly	Thr	Tyr	Ala	Gln	Arg	Pro	Gly	Asn	Ser
[2345]	785					790					795					800
[2346]	Gly	Asp	Ala	Ser	Ser	Thr	Asp	Gly	Asn	Asp	Asn	Leu	Leu	Leu	Gly	Leu
[2347]				805					810						815	
[2348]	Ser	Trp	Leu	Thr	Gln	Arg	Asp	Gln	Arg	Pro	Thr	Phe	Ala	Leu	Trp	Gly
[2349]			820						825					830		
[2350]	Ile	Arg	Thr	Ser	Ala	Ala	Ala	Gln	Ala	Gln	Val	Ala	Ala	Tyr	Gly	Phe
[2351]		835						840						845		

[2352]	Ala Glu Gln Pro Ala Phe Phe Tyr Ala Asn Asn Arg Thr Asn Glu Tyr
[2353]	850 855 860
[2354]	Ser Thr Val Lys Leu Leu Asp Met Ser Gln Gly Ser Pro Ala Trp Pro
[2355]	865 870 875 880
[2356]	Phe Pro
[2357]	<210> 36
[2358]	<211> 536
[2359]	<212> PRT
[2360]	<213> Artificial Sequence
[2361]	<220>
[2362]	<223> Polypeptide capable of binding to O-glycans but lack or has
[2363]	reduced O-glycoprotein-specific endoprotease activity -
[2364]	Bacteroides thetaiotaomicron VPI-5482 (double mutant with removed
[2365]	signal and other immature sequences)
[2366]	<400> 36
[2367]	Asp Lys Trp Glu Lys Glu Phe Arg Ile Arg Ser Tyr Glu Pro Tyr Ser
[2368]	1 5 10 15
[2369]	Asn Ile Ala Glu Trp Ala Asp Lys Leu Met Thr Lys Lys Tyr Ser Asp
[2370]	20 25 30
[2371]	Leu Asp Asn Pro Thr Gly Ile Ser Val Lys Ala Gly Asp Asp Ile Ile
[2372]	35 40 45
[2373]	Val Leu Val Gly Asp Thr Tyr Gly Gln Asn Ile Ser Met Gln Cys Ile
[2374]	50 55 60
[2375]	Trp Glu Thr Gly Thr Glu Tyr Lys Gln Thr Ala Ser Ser Gly Asp Val
[2376]	65 70 75 80
[2377]	Tyr Met Leu Asn Pro Gly Val Asn Lys Leu Thr Met Lys Gly Glu Gly
[2378]	85 90 95
[2379]	Gln Leu Phe Val Met Tyr Asn Thr Glu Leu Thr Ser Asn Thr Ala Lys
[2380]	100 105 110
[2381]	Pro Ile Lys Ile His Ile Pro Leu Gly Ser Gly Thr Val Asn Gly Phe
[2382]	115 120 125
[2383]	Phe Asp Leu Lys Glu His Lys Thr Asp Glu Lys Tyr Ala Glu Leu Leu
[2384]	130 135 140
[2385]	Lys Lys Ser Thr His Lys Tyr Phe Cys Ile Arg Gly Glu Lys Ile Met
[2386]	145 150 155 160
[2387]	Phe Tyr Phe His Arg Asn Lys Leu Leu Glu Tyr Val Pro Asn Asn Ile
[2388]	165 170 175
[2389]	Leu Ser Ala Ile His Leu Trp Asp Asn Ile Val Gly Trp Gln Gln Glu
[2390]	180 185 190
[2391]	Leu Met Gly Ile Asp Asp Val Arg Pro Ser Gln Val Asn Asn His Leu
[2392]	195 200 205
[2393]	Phe Ala Ile Ser Pro Glu Gly Ser Tyr Met Trp Ala Ser Asp Tyr Gln

[2394]	210	215	220
[2395]	Ile Gly Phe Val Tyr Thr Tyr Leu Gly Asn Ile Leu Leu Glu Asp Asn		
[2396]	225	230	235 240
[2397]	Val Met Ala Ala Glu Asp Asn Ala Trp Gly Pro Ala Ala Ala Ile Gly		
[2398]		245	250 255
[2399]	His Val His Gln Ala Ala Ile Asn Trp Ala Ser Ser Thr Glu Ser Ser		
[2400]		260	265 270
[2401]	Asn Asn Leu Phe Ser Asn Phe Ile Ile Tyr Lys Leu Gly Lys Tyr Lys		
[2402]		275	280 285
[2403]	Ser Arg Gly Asn Gly Leu Gly Ser Val Ala Thr Ala Arg Tyr Ala Asn		
[2404]	290	295	300
[2405]	Gly Gln Ala Trp Tyr Asn Met Gly Asp Ala Thr His Gln Asn Glu Asp		
[2406]	305	310	315 320
[2407]	Thr Glu Thr His Met Arg Met Asn Trp Gln Leu Trp Ile Tyr Tyr His		
[2408]		325	330 335
[2409]	Arg Cys Glu Tyr Lys Thr Asp Phe Trp Gln Thr Leu Phe Lys Leu Met		
[2410]		340	345 350
[2411]	Arg Glu Val Asn Met Thr Glu Gly Glu Asp Pro Gly Lys Lys Gln Leu		
[2412]		355	360 365
[2413]	Glu Phe Ala Lys Met Ala Ser Lys Ala Ala Asn Gln Asn Leu Thr Asp		
[2414]	370	375	380
[2415]	Phe Phe Glu Met Trp Gly Phe Phe Glu Pro Val Asn Thr Thr Ile Glu		
[2416]	385	390	395 400
[2417]	Gln Tyr Gly Thr Tyr Lys Tyr Tyr Val Ser Asp Ala Met Ile Arg Glu		
[2418]		405	410 415
[2419]	Ala Lys Glu Tyr Met Ala Gln Phe Pro Ala Pro Lys His Ala Phe Gln		
[2420]		420	425 430
[2421]	Tyr Ile Glu Asp Arg Lys Lys Ser Glu Phe Pro Ser Asn Asp Tyr Arg		
[2422]		435	440 445
[2423]	Tyr Ser Ala Val Gly Asp Val Gly Tyr Tyr Thr Gln Phe Lys Glu Asn		
[2424]	450	455	460
[2425]	Gln Lys Ile Thr Lys Ala Ile Thr Ala Glu Leu Ala Gly Arg Lys Val		
[2426]	465	470	475 480
[2427]	Ser Ile Gln Asn Gly Asp Glu Ala Val Ala Phe Glu Leu Arg Glu Asn		
[2428]		485	490 495
[2429]	Asp Glu Asn Gly Lys Leu Leu Tyr Phe Ser Thr Phe Thr Thr Phe Glu		
[2430]		500	505 510
[2431]	Ile Pro Ser Ser Ile Leu Met Val Asn Ala Lys Leu Tyr Ala Val Gln		
[2432]		515	520 525
[2433]	Ala Asp Gly Lys Arg Ile Leu Leu		
[2434]	530	535	
[2435]	<210> 37		

[2436]	<211>	507
[2437]	<212>	PRT
[2438]	<213>	Artificial Sequence
[2439]	<220>	
[2440]	<223>	Polypeptide capable of binding to O-glycans but lack or has
[2441]		reduced O-glycoprotein-specific endoprotease activity -
[2442]		Clostridium perfringens (double mutant with removed signal and
[2443]		other immature sequences)
[2444]	<400>	37
[2445]	Val Leu Glu Leu Glu Met Arg Gly Asp Ser Ile Ser Glu Ala Lys Lys	
[2446]	1	5 10 15
[2447]	Arg Lys Val Trp Asn Phe Gln Asp Trp Gln Ile Thr Gly Leu Ser Ala	
[2448]		20 25 30
[2449]	Arg Ala Gly Asp Lys Ile Thr Val Tyr Val Asp Val Ala Glu Gly Asp	
[2450]		35 40 45
[2451]	Pro Thr Pro Thr Leu Leu Tyr Lys Gln Ser Leu Thr Gln His Gly Gly	
[2452]		50 55 60
[2453]	Ala Thr Ser Phe Gln Leu Lys Pro Gly Lys Asn Glu Ile Thr Ile Pro	
[2454]	65	70 75 80
[2455]	Glu Ile Asn Tyr Glu Ser Asn Gly Ile Pro Lys Asp Val Ile Gln Gly	
[2456]		85 90 95
[2457]	Gly Asp Leu Phe Phe Thr Asn Tyr Lys Ser Asp Ser Gln Lys Arg Ala	
[2458]		100 105 110
[2459]	Pro Lys Val Arg Ile Glu Gly Ala Ser Lys Tyr Pro Val Phe Ile Leu	
[2460]		115 120 125
[2461]	Gly Lys Ser Asp Glu Asn Glu Val Met Lys Glu Leu Glu Ala Tyr Val	
[2462]		130 135 140
[2463]	Glu Lys Ile Lys Ala Glu Pro Lys Thr Thr Pro Asn Ile Phe Ala Val	
[2464]	145	150 155 160
[2465]	Ser Ser Asn Lys Ser Leu Glu Phe Val Gln Ala Thr Tyr Ala Leu Asp	
[2466]		165 170 175
[2467]	Trp Tyr Lys Lys Asn Asn Lys Thr Pro Lys Tyr Thr Ala Glu Gln Trp	
[2468]		180 185 190
[2469]	Asp Gln Tyr Ile Ala Asp Ala Met Gly Phe Trp Gly Phe Asp Asn Ser	
[2470]		195 200 205
[2471]	Lys Asp Val Asn Ser Asp Phe Asn Phe Arg Ile Met Pro Met Val Lys	
[2472]		210 215 220
[2473]	Asn Leu Ser Gly Gly Ala Phe Met Asn Ala Gly Asn Gly Val Ile Gly	
[2474]	225	230 235 240
[2475]	Ile Arg Pro Gly Asn Gln Asp Ala Ile Leu Ala Ala Asn Lys Gly Trp	
[2476]		245 250 255
[2477]	Gly Val Ala Ala Ala Leu Gly His Asn Phe Asp Thr Gly Gly Arg Thr	

[2478]	260	265	270
[2479]	Ile Val Glu Val Thr Asn Asn Met Met Pro Leu Phe Phe Glu Ser Lys		
[2480]	275	280	285
[2481]	Tyr Lys Thr Lys Thr Arg Ile Thr Asp Gln Asn Ile Trp Glu Asn Asn		
[2482]	290	295	300
[2483]	Thr Tyr Pro Lys Val Gly Leu Asp Asp Tyr Ser Asn Asn Glu Leu Tyr		
[2484]	305	310	315 320
[2485]	Asn Lys Ala Asp Ser Thr His Leu Ala Gln Leu Ala Pro Leu Trp Gln		
[2486]	325	330	335
[2487]	Leu Tyr Leu Tyr Asp Asn Thr Phe Tyr Gly Lys Phe Glu Arg Gln Phe		
[2488]	340	345	350
[2489]	Arg Glu Arg Asp Phe Gly Asn Lys Asn Arg Glu Asp Ile Tyr Lys Ser		
[2490]	355	360	365
[2491]	Trp Val Val Ala Ala Ser Asp Ala Met Glu Leu Asp Leu Thr Glu Phe		
[2492]	370	375	380
[2493]	Phe Ala Arg His Gly Ile Arg Val Asp Asp Lys Val Lys Glu Asp Leu		
[2494]	385	390	395 400
[2495]	Ala Lys Tyr Pro Lys Pro Asp Lys Lys Ile Tyr Tyr Leu Asn Asp Leu		
[2496]	405	410	415
[2497]	Ala Met Asn Tyr Lys Gly Asp Gly Phe Thr Glu Asn Ala Lys Val Ser		
[2498]	420	425	430
[2499]	Val Ser Thr Ser Gly Ser Asn Gly Asn Ile Lys Leu Ser Phe Ser Val		
[2500]	435	440	445
[2501]	Asp Asp Glu Asn Lys Asp Asn Ile Leu Gly Tyr Glu Ile Arg Arg Asp		
[2502]	450	455	460
[2503]	Gly Lys Tyr Val Gly Phe Thr Ser Asn Asp Ser Phe Val Asp Thr Lys		
[2504]	465	470	475 480
[2505]	Ser Asn Leu Asp Glu Asp Gly Val Tyr Val Val Thr Pro Tyr Asp Arg		
[2506]	485	490	495
[2507]	Lys Leu Asn Thr Leu Asn Pro Ile Glu Val Asn		
[2508]	500	505	
[2509]	<210> 38		
[2510]	<211> 892		
[2511]	<212> PRT		
[2512]	<213> Artificial Sequence		
[2513]	<220>		
[2514]	<223> Polypeptide capable of binding to O-glycans but lack or has		
[2515]	reduced O-glycoprotein-specific endoprotease activity -		
[2516]	Pseudomonas aeruginosa PA01 (double mutant with removed signal		
[2517]	sequence, with N-term Met, C-term linker/tag)		
[2518]	<400> 38		
[2519]	Met Ala Thr Gln Glu Glu Ile Leu Asp Ala Ala Leu Val Ser Gly Asp		

[2520]	1	5	10	15
[2521]	Ser Ser Gln Leu Thr Asp Ser His Leu Val Ala Leu Arg Leu Gln Gln			
[2522]	20	25	30	
[2523]	Gln Val Glu Arg Ile Arg Gln Thr Arg Thr Gln Leu Leu Asp Gly Leu			
[2524]	35	40	45	
[2525]	Tyr Gln Asn Leu Ser Gln Ala Tyr Asp Pro Gly Ala Ala Ser Met Trp			
[2526]	50	55	60	
[2527]	Val Leu Pro Ala Asn Pro Asp Asn Thr Leu Pro Phe Leu Ile Gly Asp			
[2528]	65	70	75	80
[2529]	Lys Gly Arg Val Leu Ala Ser Leu Ser Leu Glu Ala Gly Gly Arg Gly			
[2530]	85	90	95	
[2531]	Leu Ala Tyr Gly Thr Asn Val Leu Thr Gln Leu Ser Gly Thr Asn Ala			
[2532]	100	105	110	
[2533]	Ala His Ala Pro Leu Leu Lys Arg Ala Val Gln Trp Leu Val Asn Gly			
[2534]	115	120	125	
[2535]	Asp Pro Gly Ala Ala Thr Ala Lys Asp Phe Lys Val Ser Val Val Gly			
[2536]	130	135	140	
[2537]	Val Asp Lys Thr Ala Ala Leu Asn Gly Leu Lys Ser Ala Gly Leu Gln			
[2538]	145	150	155	160
[2539]	Pro Ala Asp Ala Ala Cys Asn Ala Leu Thr Asp Ala Ser Cys Ala Ser			
[2540]	165	170	175	
[2541]	Thr Ser Lys Leu Leu Val Leu Gly Asn Gly Ala Ser Ala Ala Ser Leu			
[2542]	180	185	190	
[2543]	Ser Ala Thr Val Arg Ala Arg Leu Gln Ala Gly Leu Pro Ile Leu Phe			
[2544]	195	200	205	
[2545]	Val His Thr Asn Gly Trp Asn Gln Ser Ser Thr Gly Gln Gln Ile Leu			
[2546]	210	215	220	
[2547]	Ala Gly Leu Gly Leu Gln Glu Gly Pro Tyr Gly Gly Asn Tyr Trp Asp			
[2548]	225	230	235	240
[2549]	Lys Asp Arg Val Pro Ser Ser Arg Thr Arg Thr Arg Ser Val Glu Leu			
[2550]	245	250	255	
[2551]	Gly Gly Ala Tyr Gly Gln Asp Pro Ala Leu Val Gln Gln Ile Val Asp			
[2552]	260	265	270	
[2553]	Gly Ser Trp Arg Thr Asp Tyr Asp Trp Ser Lys Cys Thr Ser Tyr Val			
[2554]	275	280	285	
[2555]	Gly Arg Thr Thr Cys Asp Asp Val Pro Gly Leu Ser Asp Phe Ser Lys			
[2556]	290	295	300	
[2557]	Arg Val Asp Val Leu Lys Gly Ala Leu Asp Ala Tyr Asn Gln Lys Ala			
[2558]	305	310	315	320
[2559]	Gln Asn Leu Phe Ala Leu Pro Gly Thr Thr Ser Leu Arg Leu Trp Leu			
[2560]	325	330	335	
[2561]	Leu Trp Ala Asp Ala Val Arg Gln Asn Ile Arg Tyr Pro Met Asp Lys			

[2562]	340	345	350
[2563]	Ala Ala Asp Thr Ala Arg Phe Gln Glu Thr Phe Val Ala Asp Ala Ile		
[2564]	355	360	365
[2565]	Val Gly Tyr Val Arg Glu Ala Gly Ala Ala Gln Lys Glu Leu Gly Ser		
[2566]	370	375	380
[2567]	Tyr Ala Gly Gln Arg Gln Gln Ser Met Pro Val Ser Gly Ser Glu Glu		
[2568]	385	390	395
[2569]	Thr Leu Thr Leu Thr Leu Pro Ser Ala Gln Gly Phe Thr Ala Ile Gly		
[2570]	405	410	415
[2571]	Arg Met Ala Ala Pro Gly Lys Arg Leu Ser Ile Arg Ile Glu Asp Ala		
[2572]	420	425	430
[2573]	Gly Gln Ala Ser Leu Ala Val Gly Leu Asn Thr Gln Arg Ile Gly Ser		
[2574]	435	440	445
[2575]	Thr Arg Leu Trp Asn Thr Arg Gln Tyr Asp Arg Pro Arg Phe Leu Lys		
[2576]	450	455	460
[2577]	Ser Pro Asp Ile Lys Leu Gln Ala Asn Gln Ser Val Ala Leu Val Ser		
[2578]	465	470	475
[2579]	Pro Tyr Gly Gly Leu Leu Gln Leu Val Tyr Ser Gly Ala Thr Pro Gly		
[2580]	485	490	495
[2581]	Gln Thr Val Thr Val Lys Val Thr Gly Ala Ala Ser Gln Pro Phe Leu		
[2582]	500	505	510
[2583]	Asp Ile Gln Pro Gly Glu Asp Ser Ser Gln Ala Ile Ala Asp Phe Ile		
[2584]	515	520	525
[2585]	Gln Ala Leu Asp Ala Asp Lys Ala Asp Trp Leu Glu Ile Arg Ser Gly		
[2586]	530	535	540
[2587]	Ser Val Glu Val His Ala Lys Val Glu Lys Val Arg Gly Ser Ile Asp		
[2588]	545	550	555
[2589]	Lys Asp Tyr Gly Gly Asp Val Gln Arg Phe Ile Arg Glu Leu Asn Glu		
[2590]	565	570	575
[2591]	Val Phe Ile Asp Asp Ala Tyr Thr Leu Ala Gly Phe Ala Ile Pro Asn		
[2592]	580	585	590
[2593]	Gln Ala Lys Thr Pro Ala Ile Gln Gln Glu Cys Ala Ala Arg Gly Trp		
[2594]	595	600	605
[2595]	Asp Cys Asp Ser Glu Thr Leu His Lys Leu Pro Gly Thr Gln His Ile		
[2596]	610	615	620
[2597]	Asn Val Asp Gln Tyr Ala Gln Cys Gly Gly Gly Cys Ser Gly Asn Pro		
[2598]	625	630	635
[2599]	Tyr Asp Gln Thr Trp Gly Leu Asn Pro Arg Gly Trp Gly Glu Ser Ala		
[2600]	645	650	655
[2601]	Ala Leu Gly His Asn Leu Gln Val Asn Arg Leu Lys Val Tyr Gly Gly		
[2602]	660	665	670
[2603]	Arg Ser Gly Glu Ile Ser Asn Gln Ile Phe Pro Leu His Lys Asp Trp		

[2604]	675	680	685
[2605]	Arg Val Leu Arg Glu Phe Gly Gln Asn Leu Asp Asp Thr Arg Val Asn		
[2606]	690	695	700
[2607]	Tyr Arg Asn Ala Tyr Asn Leu Ile Val Ala Gly Arg Ala Glu Ala Asp		
[2608]	705	710	715
[2609]	Pro Leu Ala Gly Val Tyr Lys Arg Leu Trp Glu Asp Pro Gly Thr Tyr		
[2610]	725	730	735
[2611]	Ala Leu Asn Gly Glu Arg Met Ala Phe Tyr Thr Gln Trp Val His Tyr		
[2612]	740	745	750
[2613]	Trp Ala Asp Leu Lys Asn Asp Pro Leu Gln Gly Trp Asp Ile Trp Thr		
[2614]	755	760	765
[2615]	Leu Leu Tyr Leu His Gln Arg Gln Val Asp Lys Ser Asp Trp Asp Ala		
[2616]	770	775	780
[2617]	Asn Lys Ala Ala Leu Gly Tyr Gly Thr Tyr Ala Gln Arg Pro Gly Asn		
[2618]	785	790	795
[2619]	Ser Gly Asp Ala Ser Ser Thr Asp Gly Asn Asp Asn Leu Leu Leu Gly		
[2620]	805	810	815
[2621]	Leu Ser Trp Leu Thr Gln Arg Asp Gln Arg Pro Thr Phe Ala Leu Trp		
[2622]	820	825	830
[2623]	Gly Ile Arg Thr Ser Ala Ala Ala Gln Ala Gln Val Ala Ala Tyr Gly		
[2624]	835	840	845
[2625]	Phe Ala Glu Gln Pro Ala Phe Phe Tyr Ala Asn Asn Arg Thr Asn Glu		
[2626]	850	855	860
[2627]	Tyr Ser Thr Val Lys Leu Leu Asp Met Ser Gln Gly Ser Pro Ala Trp		
[2628]	865	870	875
[2629]	Pro Phe Pro Gly Ser Gly His His His His His His		
[2630]	885	890	
[2631]	<210> 39		
[2632]	<211> 546		
[2633]	<212> PRT		
[2634]	<213> Artificial Sequence		
[2635]	<220>		
[2636]	<223> Polypeptide capable of binding to O-glycans but lack or has		
[2637]	reduced O-glycoprotein-specific endoprotease activity -		
[2638]	Bacteroides thetaiotaomicron VPI-5482 (double mutant with removed		
[2639]	signal and other sequences from immature protein, with N-term		
[2640]	<400> 39		
[2641]	Met Asp Lys Trp Glu Lys Glu Phe Arg Ile Arg Ser Tyr Glu Pro Tyr		
[2642]	1	5	10
[2643]	Ser Asn Ile Ala Glu Trp Ala Asp Lys Leu Met Thr Lys Lys Tyr Ser		
[2644]	20	25	30
[2645]	Asp Leu Asp Asn Pro Thr Gly Ile Ser Val Lys Ala Gly Asp Asp Ile		

[2646]	35	40	45
[2647]	Ile Val Leu Val Gly Asp Thr Tyr Gly Gln Asn Ile Ser Met Gln Cys		
[2648]	50	55	60
[2649]	Ile Trp Glu Thr Gly Thr Glu Tyr Lys Gln Thr Ala Ser Ser Gly Asp		
[2650]	65	70	75
[2651]	Val Tyr Met Leu Asn Pro Gly Val Asn Lys Leu Thr Met Lys Gly Glu		
[2652]	85	90	95
[2653]	Gly Gln Leu Phe Val Met Tyr Asn Thr Glu Leu Thr Ser Asn Thr Ala		
[2654]	100	105	110
[2655]	Lys Pro Ile Lys Ile His Ile Pro Leu Gly Ser Gly Thr Val Asn Gly		
[2656]	115	120	125
[2657]	Phe Phe Asp Leu Lys Glu His Lys Thr Asp Glu Lys Tyr Ala Glu Leu		
[2658]	130	135	140
[2659]	Leu Lys Lys Ser Thr His Lys Tyr Phe Cys Ile Arg Gly Glu Lys Ile		
[2660]	145	150	155
[2661]	Met Phe Tyr Phe His Arg Asn Lys Leu Leu Glu Tyr Val Pro Asn Asn		
[2662]	165	170	175
[2663]	Ile Leu Ser Ala Ile His Leu Trp Asp Asn Ile Val Gly Trp Gln Gln		
[2664]	180	185	190
[2665]	Glu Leu Met Gly Ile Asp Asp Val Arg Pro Ser Gln Val Asn Asn His		
[2666]	195	200	205
[2667]	Leu Phe Ala Ile Ser Pro Glu Gly Ser Tyr Met Trp Ala Ser Asp Tyr		
[2668]	210	215	220
[2669]	Gln Ile Gly Phe Val Tyr Thr Tyr Leu Gly Asn Ile Leu Leu Glu Asp		
[2670]	225	230	235
[2671]	Asn Val Met Ala Ala Glu Asp Asn Ala Trp Gly Pro Ala Ala Ala Ile		
[2672]	245	250	255
[2673]	Gly His Val His Gln Ala Ala Ile Asn Trp Ala Ser Ser Thr Glu Ser		
[2674]	260	265	270
[2675]	Ser Asn Asn Leu Phe Ser Asn Phe Ile Ile Tyr Lys Leu Gly Lys Tyr		
[2676]	275	280	285
[2677]	Lys Ser Arg Gly Asn Gly Leu Gly Ser Val Ala Thr Ala Arg Tyr Ala		
[2678]	290	295	300
[2679]	Asn Gly Gln Ala Trp Tyr Asn Met Gly Asp Ala Thr His Gln Asn Glu		
[2680]	305	310	315
[2681]	Asp Thr Glu Thr His Met Arg Met Asn Trp Gln Leu Trp Ile Tyr Tyr		
[2682]	325	330	335
[2683]	His Arg Cys Glu Tyr Lys Thr Asp Phe Trp Gln Thr Leu Phe Lys Leu		
[2684]	340	345	350
[2685]	Met Arg Glu Val Asn Met Thr Glu Gly Glu Asp Pro Gly Lys Lys Gln		
[2686]	355	360	365
[2687]	Leu Glu Phe Ala Lys Met Ala Ser Lys Ala Ala Asn Gln Asn Leu Thr		

[2688]	370	375	380
[2689]	Asp Phe Phe Glu Met Trp Gly Phe Phe Glu Pro Val Asn Thr Thr Ile		
[2690]	385	390	395 400
[2691]	Glu Gln Tyr Gly Thr Tyr Lys Tyr Tyr Val Ser Asp Ala Met Ile Arg		
[2692]		405	410 415
[2693]	Glu Ala Lys Glu Tyr Met Ala Gln Phe Pro Ala Pro Lys His Ala Phe		
[2694]		420	425 430
[2695]	Gln Tyr Ile Glu Asp Arg Lys Lys Ser Glu Phe Pro Ser Asn Asp Tyr		
[2696]		435	440 445
[2697]	Arg Tyr Ser Ala Val Gly Asp Val Gly Tyr Tyr Thr Gln Phe Lys Glu		
[2698]		450	455 460
[2699]	Asn Gln Lys Ile Thr Lys Ala Ile Thr Ala Glu Leu Ala Gly Arg Lys		
[2700]		465	470 475 480
[2701]	Val Ser Ile Gln Asn Gly Asp Glu Ala Val Ala Phe Glu Leu Arg Glu		
[2702]		485	490 495
[2703]	Asn Asp Glu Asn Gly Lys Leu Leu Tyr Phe Ser Thr Phe Thr Thr Phe		
[2704]		500	505 510
[2705]	Glu Ile Pro Ser Ser Ile Leu Met Val Asn Ala Lys Leu Tyr Ala Val		
[2706]		515	520 525
[2707]	Gln Ala Asp Gly Lys Arg Ile Leu Leu Gly Ser Gly His His His His		
[2708]		530	535 540
[2709]	His His		
[2710]	545		
[2711]	<210> 40		
[2712]	<211> 517		
[2713]	<212> PRT		
[2714]	<213> Artificial Sequence		
[2715]	<220>		
[2716]	<223> Polypeptide capable of binding to O-glycans but lack or has		
[2717]	reduced O-glycoprotein-specific endoprotease activity -		
[2718]	Clostridium perfringens (double mutant with removed signal and		
[2719]	other sequences from immature protein, with N-term Met, C-term		
[2720]	<400> 40		
[2721]	Met Val Leu Glu Leu Glu Met Arg Gly Asp Ser Ile Ser Glu Ala Lys		
[2722]	1	5	10 15
[2723]	Lys Arg Lys Val Trp Asn Phe Gln Asp Trp Gln Ile Thr Gly Leu Ser		
[2724]		20	25 30
[2725]	Ala Arg Ala Gly Asp Lys Ile Thr Val Tyr Val Asp Val Ala Glu Gly		
[2726]		35	40 45
[2727]	Asp Pro Thr Pro Thr Leu Leu Tyr Lys Gln Ser Leu Thr Gln His Gly		
[2728]		50	55 60
[2729]	Gly Ala Thr Ser Phe Gln Leu Lys Pro Gly Lys Asn Glu Ile Thr Ile		

[2730]	65	70	75	80
[2731]	Pro Glu Ile Asn Tyr Glu Ser Asn Gly Ile Pro Lys Asp Val Ile Gln			
[2732]		85	90	95
[2733]	Gly Gly Asp Leu Phe Phe Thr Asn Tyr Lys Ser Asp Ser Gln Lys Arg			
[2734]		100	105	110
[2735]	Ala Pro Lys Val Arg Ile Glu Gly Ala Ser Lys Tyr Pro Val Phe Ile			
[2736]		115	120	125
[2737]	Leu Gly Lys Ser Asp Glu Asn Glu Val Met Lys Glu Leu Glu Ala Tyr			
[2738]		130	135	140
[2739]	Val Glu Lys Ile Lys Ala Glu Pro Lys Thr Thr Pro Asn Ile Phe Ala			
[2740]	145	150	155	160
[2741]	Val Ser Ser Asn Lys Ser Leu Glu Phe Val Gln Ala Thr Tyr Ala Leu			
[2742]		165	170	175
[2743]	Asp Trp Tyr Lys Lys Asn Asn Lys Thr Pro Lys Tyr Thr Ala Glu Gln			
[2744]		180	185	190
[2745]	Trp Asp Gln Tyr Ile Ala Asp Ala Met Gly Phe Trp Gly Phe Asp Asn			
[2746]		195	200	205
[2747]	Ser Lys Asp Val Asn Ser Asp Phe Asn Phe Arg Ile Met Pro Met Val			
[2748]		210	215	220
[2749]	Lys Asn Leu Ser Gly Gly Ala Phe Met Asn Ala Gly Asn Gly Val Ile			
[2750]	225	230	235	240
[2751]	Gly Ile Arg Pro Gly Asn Gln Asp Ala Ile Leu Ala Ala Asn Lys Gly			
[2752]		245	250	255
[2753]	Trp Gly Val Ala Ala Ala Leu Gly His Asn Phe Asp Thr Gly Gly Arg			
[2754]		260	265	270
[2755]	Thr Ile Val Glu Val Thr Asn Asn Met Met Pro Leu Phe Phe Glu Ser			
[2756]		275	280	285
[2757]	Lys Tyr Lys Thr Lys Thr Arg Ile Thr Asp Gln Asn Ile Trp Glu Asn			
[2758]		290	295	300
[2759]	Asn Thr Tyr Pro Lys Val Gly Leu Asp Asp Tyr Ser Asn Asn Glu Leu			
[2760]	305	310	315	320
[2761]	Tyr Asn Lys Ala Asp Ser Thr His Leu Ala Gln Leu Ala Pro Leu Trp			
[2762]		325	330	335
[2763]	Gln Leu Tyr Leu Tyr Asp Asn Thr Phe Tyr Gly Lys Phe Glu Arg Gln			
[2764]		340	345	350
[2765]	Phe Arg Glu Arg Asp Phe Gly Asn Lys Asn Arg Glu Asp Ile Tyr Lys			
[2766]		355	360	365
[2767]	Ser Trp Val Val Ala Ala Ser Asp Ala Met Glu Leu Asp Leu Thr Glu			
[2768]		370	375	380
[2769]	Phe Phe Ala Arg His Gly Ile Arg Val Asp Asp Lys Val Lys Glu Asp			
[2770]	385	390	395	400
[2771]	Leu Ala Lys Tyr Pro Lys Pro Asp Lys Lys Ile Tyr Tyr Leu Asn Asp			

[2772]	405	410	415
[2773]	Leu Ala Met Asn Tyr Lys Gly Asp Gly Phe Thr Glu Asn Ala Lys Val		
[2774]	420	425	430
[2775]	Ser Val Ser Thr Ser Gly Ser Asn Gly Asn Ile Lys Leu Ser Phe Ser		
[2776]	435	440	445
[2777]	Val Asp Asp Glu Asn Lys Asp Asn Ile Leu Gly Tyr Glu Ile Arg Arg		
[2778]	450	455	460
[2779]	Asp Gly Lys Tyr Val Gly Phe Thr Ser Asn Asp Ser Phe Val Asp Thr		
[2780]	465	470	475
[2781]	Lys Ser Asn Leu Asp Glu Asp Gly Val Tyr Val Val Thr Pro Tyr Asp		
[2782]	485	490	495
[2783]	Arg Lys Leu Asn Thr Leu Asn Pro Ile Glu Val Asn Gly Ser Gly His		
[2784]	500	505	510
[2785]	His His His His His		
[2786]	515		
[2787]	<210> 41		
[2788]	<211> 5		
[2789]	<212> PRT		
[2790]	<213> Artificial Sequence		
[2791]	<220>		
[2792]	<223> Metalloprotease motif		
[2793]	<400> 41		
[2794]	His Glu Leu Gly His		
[2795]	1	5	
[2796]	<210> 42		
[2797]	<211> 5		
[2798]	<212> PRT		
[2799]	<213> Artificial Sequence		
[2800]	<220>		
[2801]	<223> Metalloprotease motif		
[2802]	<400> 42		
[2803]	His Glu Ile Gly His		
[2804]	1	5	
[2805]	<210> 43		
[2806]	<211> 10		
[2807]	<212> PRT		
[2808]	<213> Artificial Sequence		
[2809]	<220>		
[2810]	<223> Metalloprotease motif		
[2811]	<400> 43		
[2812]	Gly Val Ala His Glu Leu Gly His Asn Phe		
[2813]	1	5	10

[2814]	<210> 44
[2815]	<211> 5
[2816]	<212> PRT
[2817]	<213> Artificial Sequence
[2818]	<220>
[2819]	<223> disrupted metalloprotease motif
[2820]	<400> 44
[2821]	His Ala Leu Gly His
[2822]	1 5
[2823]	<210> 45
[2824]	<211> 5
[2825]	<212> PRT
[2826]	<213> Artificial Sequence
[2827]	<220>
[2828]	<223> disrupted metalloprotease motif
[2829]	<400> 45
[2830]	Ala Glu Leu Gly His
[2831]	1 5
[2832]	<210> 46
[2833]	<211> 5
[2834]	<212> PRT
[2835]	<213> Artificial Sequence
[2836]	<220>
[2837]	<223> disrupted metalloprotease motif
[2838]	<400> 46
[2839]	Ala Ala Leu Gly His
[2840]	1 5
[2841]	<210> 47
[2842]	<211> 19
[2843]	<212> PRT
[2844]	<213> Artificial Sequence
[2845]	<220>
[2846]	<223> glycodrosocin peptide with O-gly site on the T
[2847]	<400> 47
[2848]	Gly Lys Pro Arg Pro Tyr Ser Pro Arg Pro Thr Ser His Pro Arg Pro
[2849]	1 5 10 15
[2850]	Ile Arg Val
[2851]	<210> 48
[2852]	<211> 7
[2853]	<212> PRT
[2854]	<213> Artificial Sequence
[2855]	<220>

[2856]	<223> Non-0-glycosylated peptide
[2857]	<400> 48
[2858]	Tyr Ile Tyr Gly Ser Phe Lys
[2859]	1 5
[2860]	<210> 49
[2861]	<211> 7
[2862]	<212> PRT
[2863]	<213> Artificial Sequence
[2864]	<220>
[2865]	<223> Non-0-glycosylated peptide
[2866]	<400> 49
[2867]	Lys Lys Leu Val Phe Phe Ala
[2868]	1 5
[2869]	<210> 50
[2870]	<211> 14
[2871]	<212> PRT
[2872]	<213> Artificial Sequence
[2873]	<220>
[2874]	<223> Non-0-glycosylated peptide
[2875]	<400> 50
[2876]	Phe Leu Pro Leu Ile Leu Gly Lys Leu Val Lys Gly Leu Leu
[2877]	1 5 10

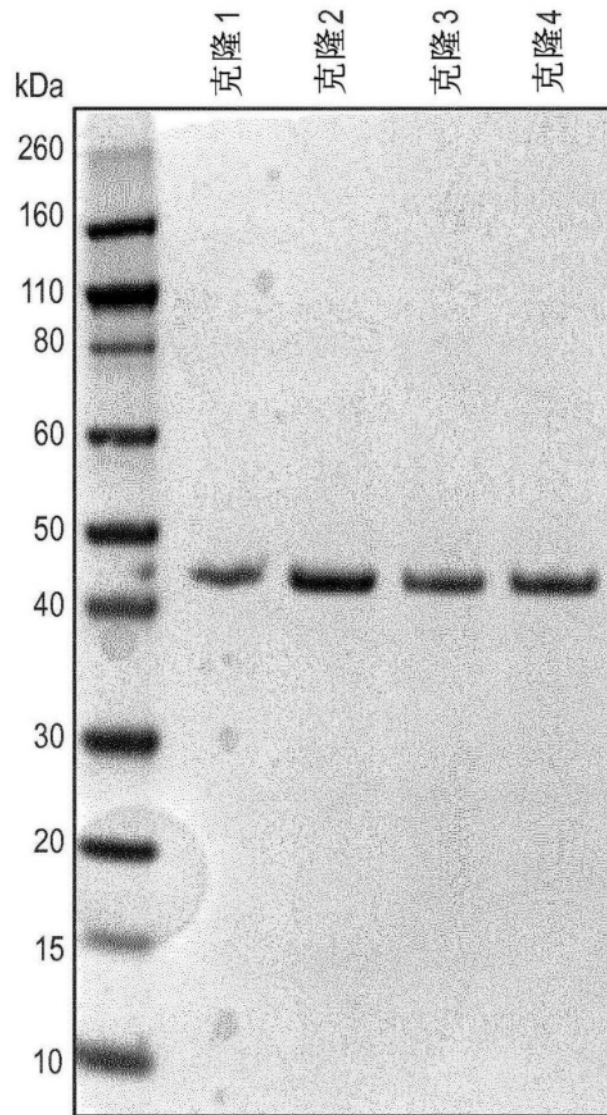


图1

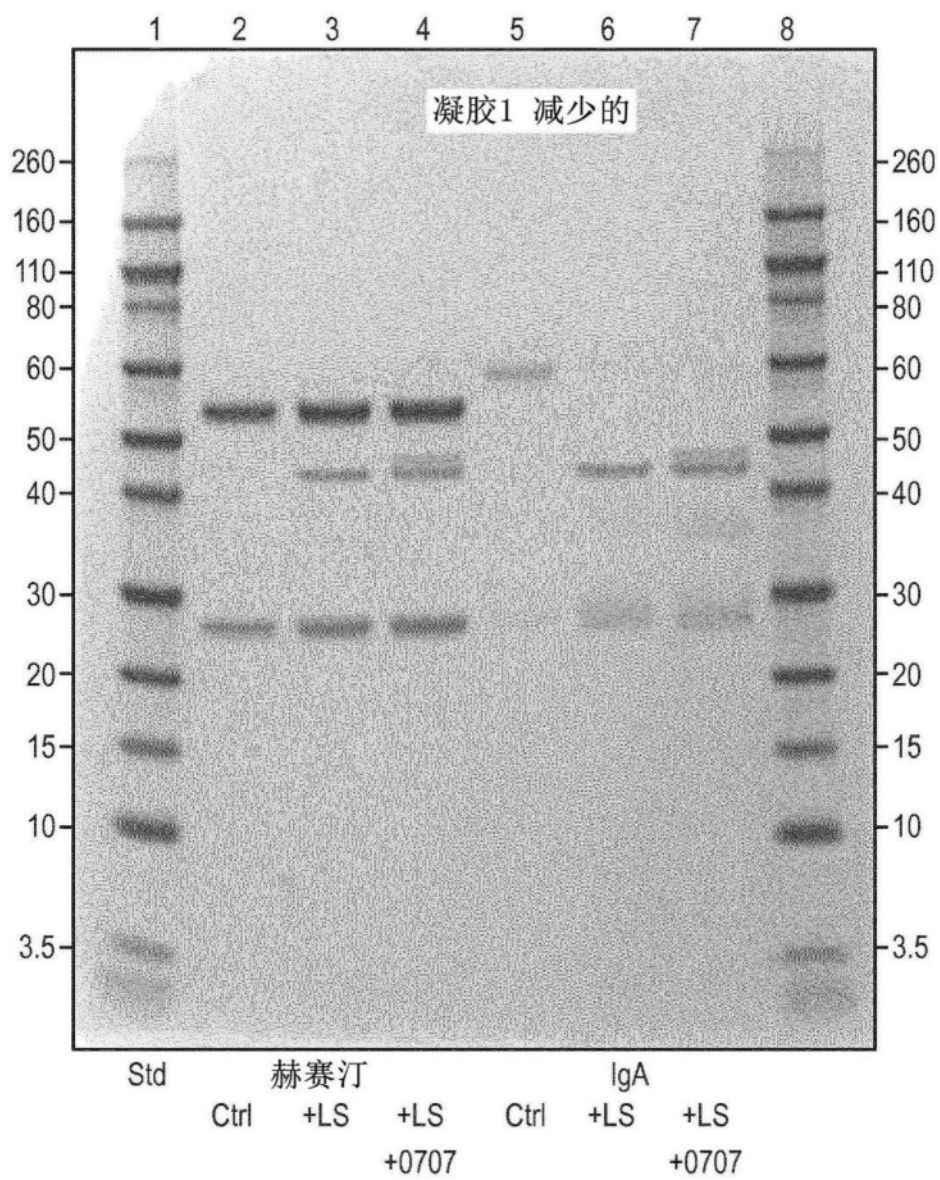


图2

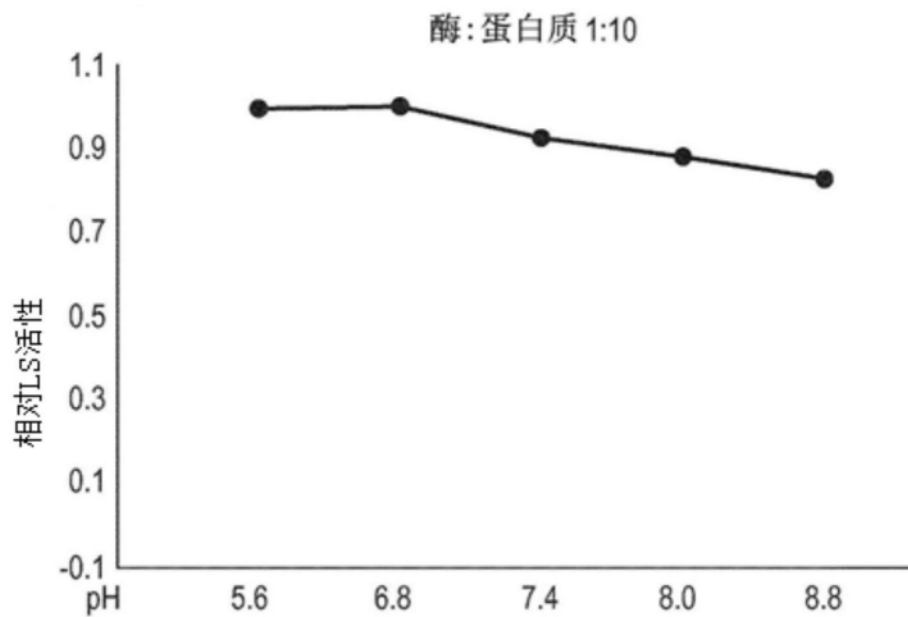


图3A

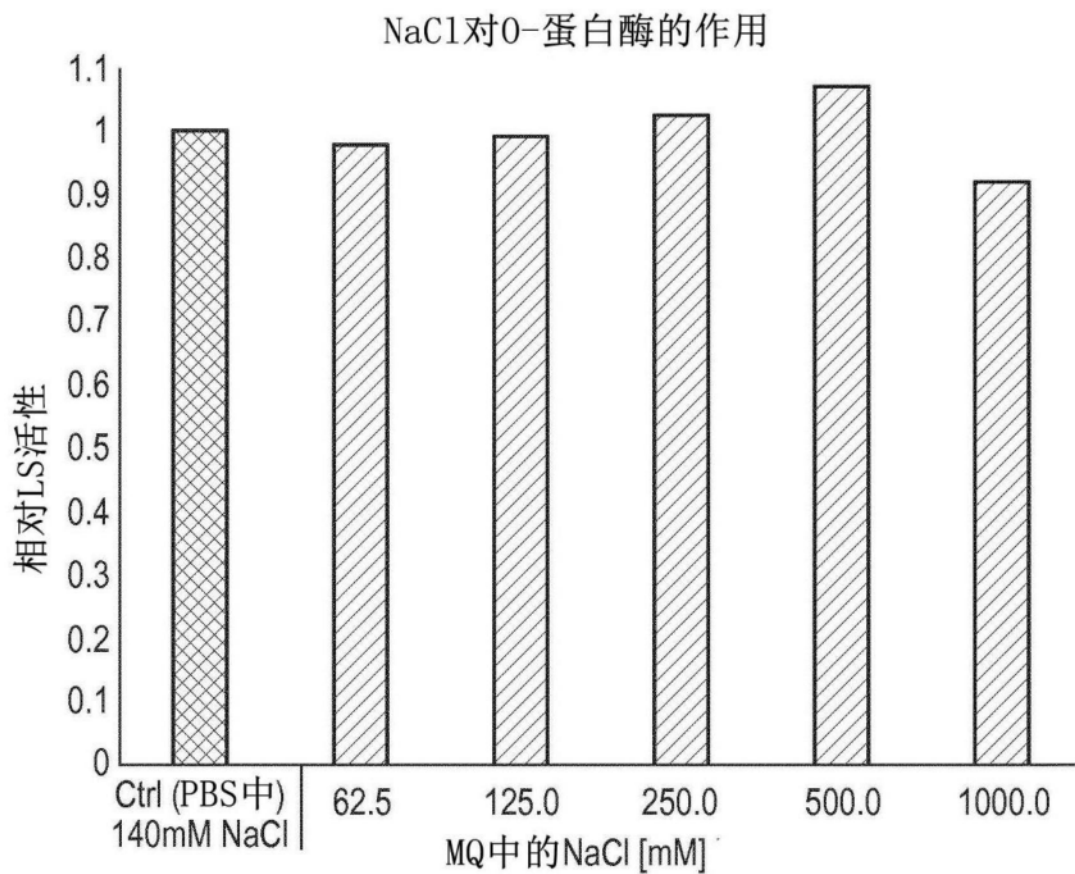


图3B

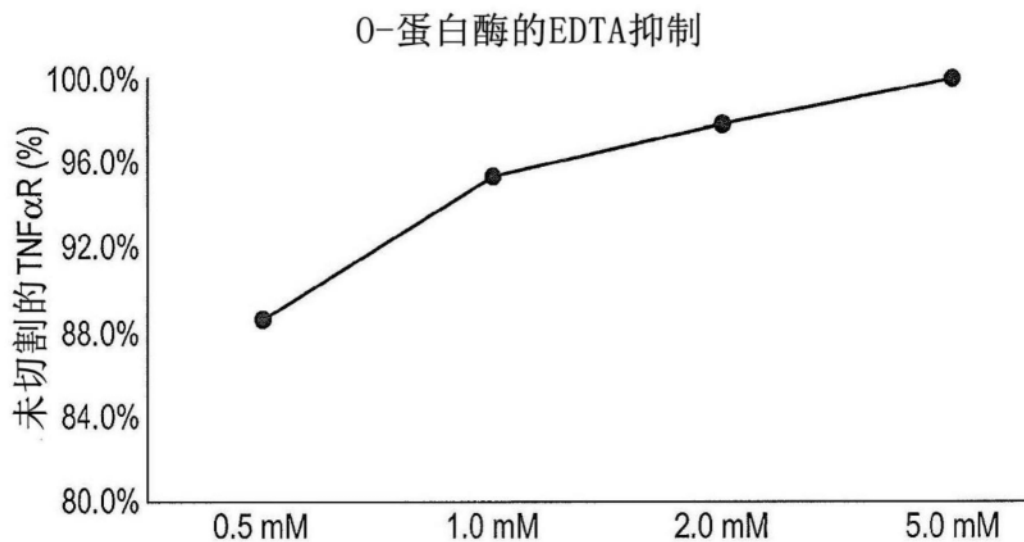


图3C

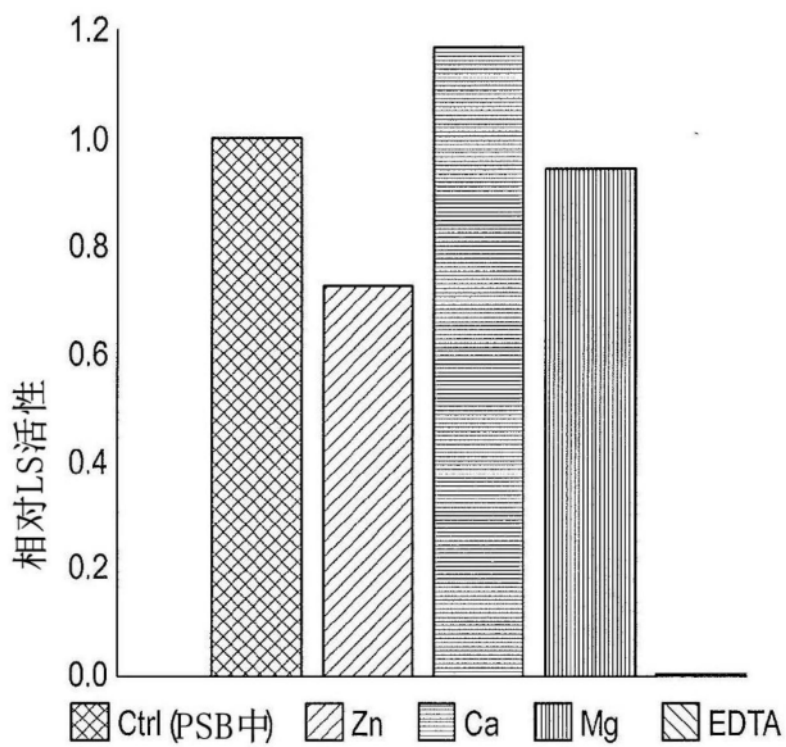


图3D

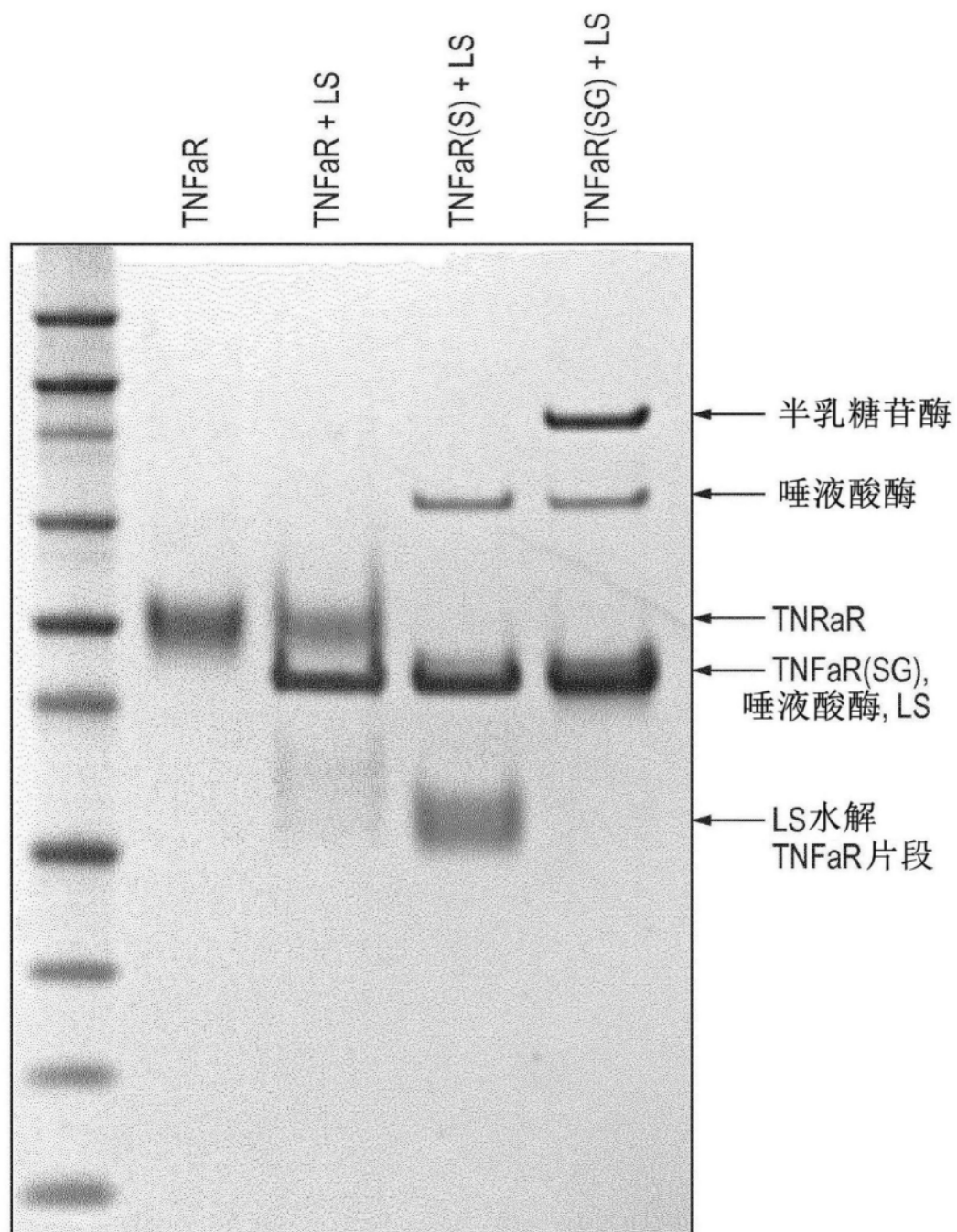


图4A

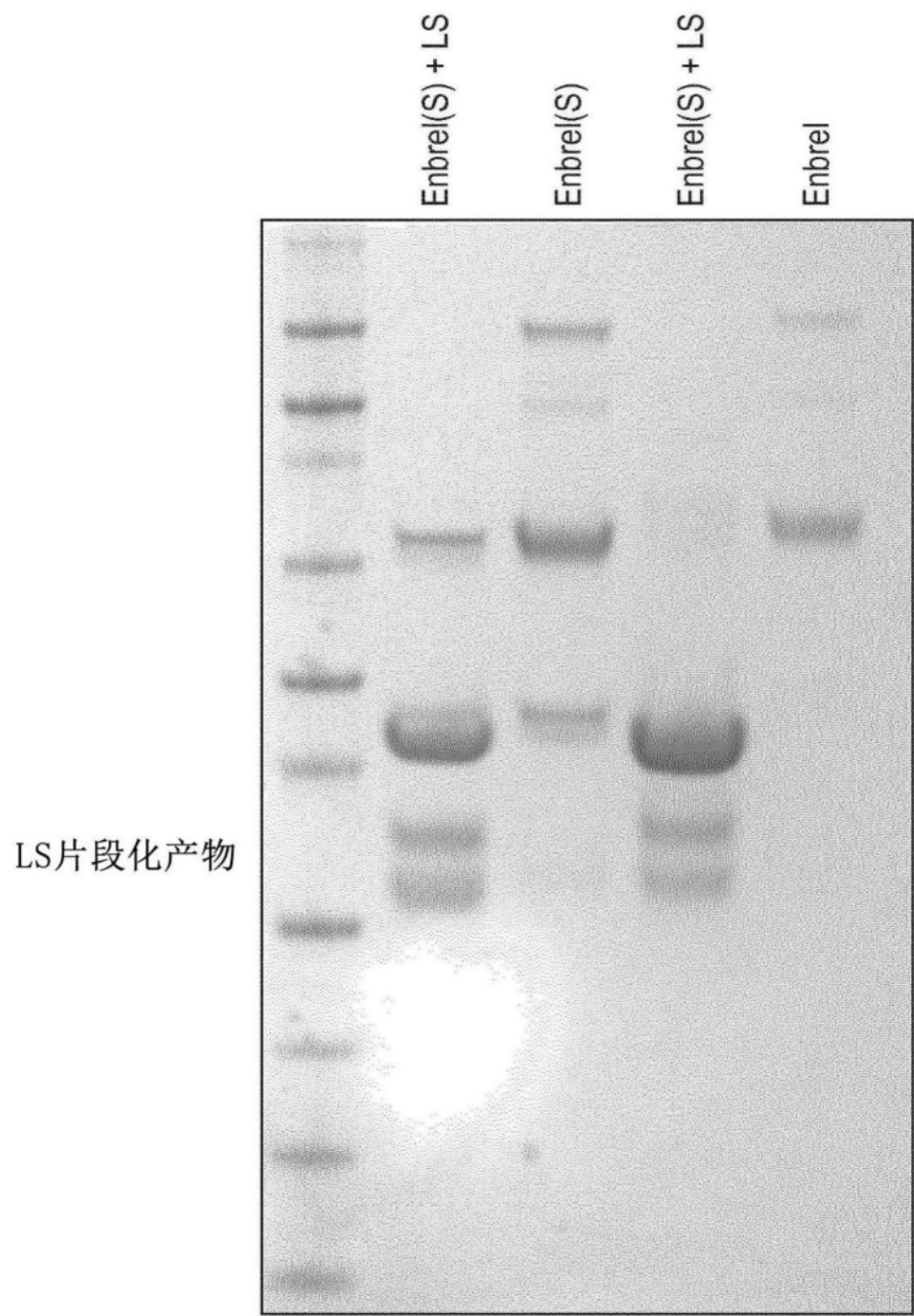


图4B

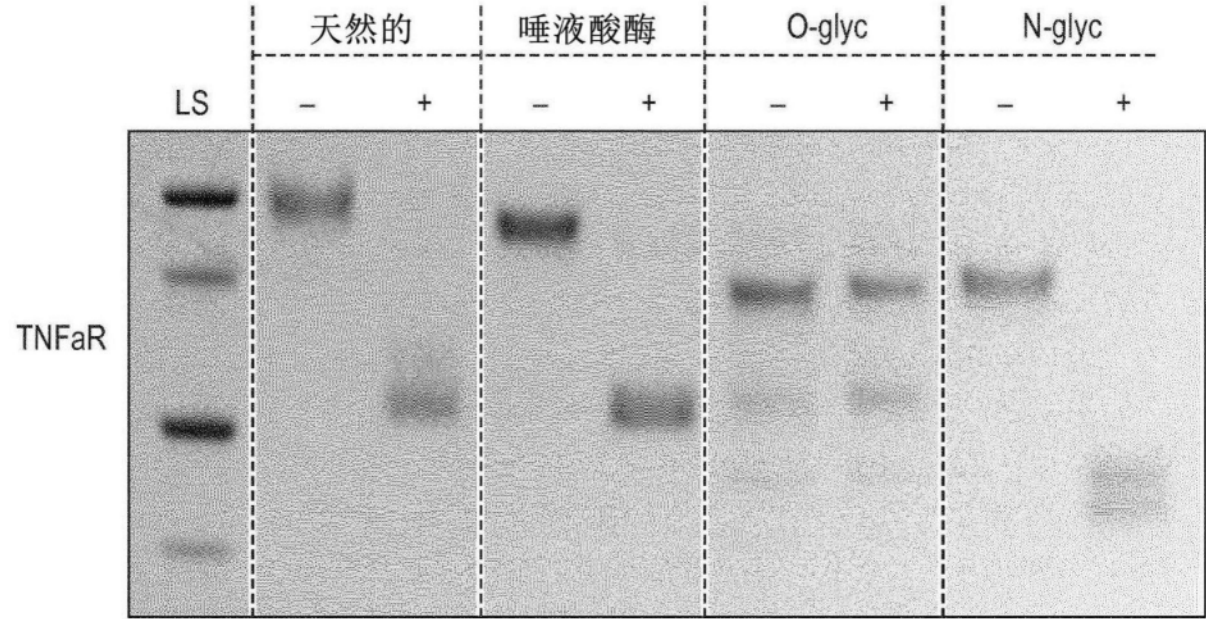


图4C

10	20	30	40	50	60	70	80	90
LPAQVAFTPY	APEPGSTCRL	REYYDQTAQM	CCSKCSPGOH	AKVFCIKTSD	TVCDSCEDST	YTQJLNNWVPE	CLSCGSRCSS	DQVETQACTR
100	110	120	130	140	150	160	170	180
EQRNICTCRP	GWYCALSKE	GCRLCAPLRK	CRPGFGVARP	GTETSDVVCK	PCAPGTFST	TSSTDICRPH	QICNVVAIPG	NASMDAVCTS
190	200	210	220	230	240	250	260	270
TSPTRSMAPG	AVHLPQPVST	RSQHTOPTPE	PSTAPSTSL	LPMGPSPPAE	GSTGDEPKSC	DKHTCPCPP	APELLGGPSV	FLFPPKPKDT
280	290	300	310	320	330	340	350	360
LMISRTPEVT	CVVVDVSHED	PEVKFNWYVD	GVEVHNAKTK	PREEQYNSTY	RVVSVLTVLH	QDWLNGKEYK	CKVSNKALPA	PIEKTISKAK
370	380	390	400	410	420	430	440	450
GQPREPQVYT	LPPSREEMTK	NOVSLTCLVK	GFYPDSIAVE	WESNGQPENN	YKTPPVLDS	DGSFTLYSKL	TVDKSRWQOG	NVFSCSVMHE
460	470							
ALHNHYTQKS	LSLSPGK							

图5A

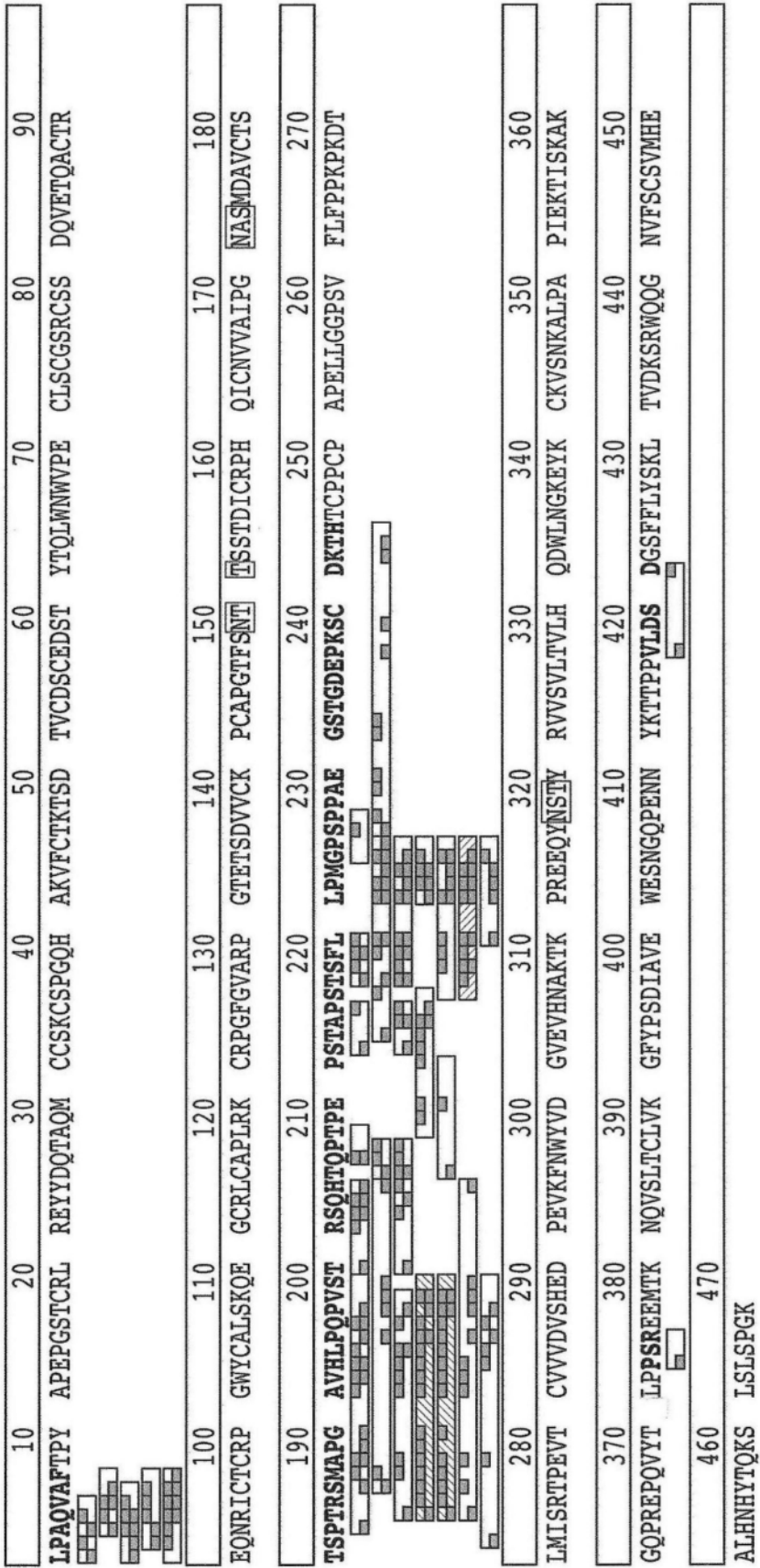


图5B

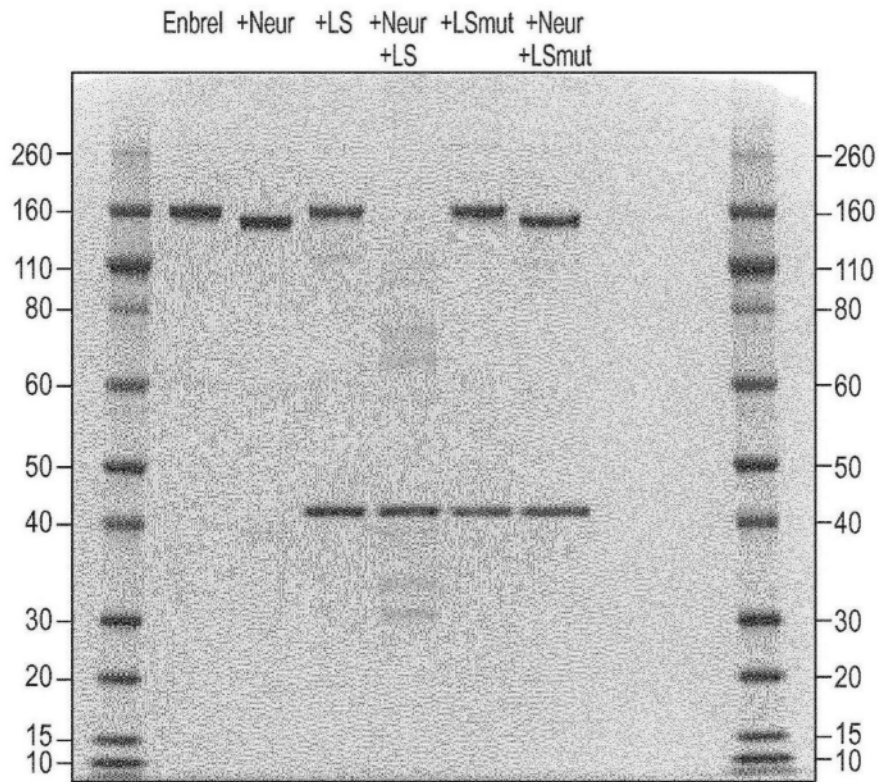


图6A

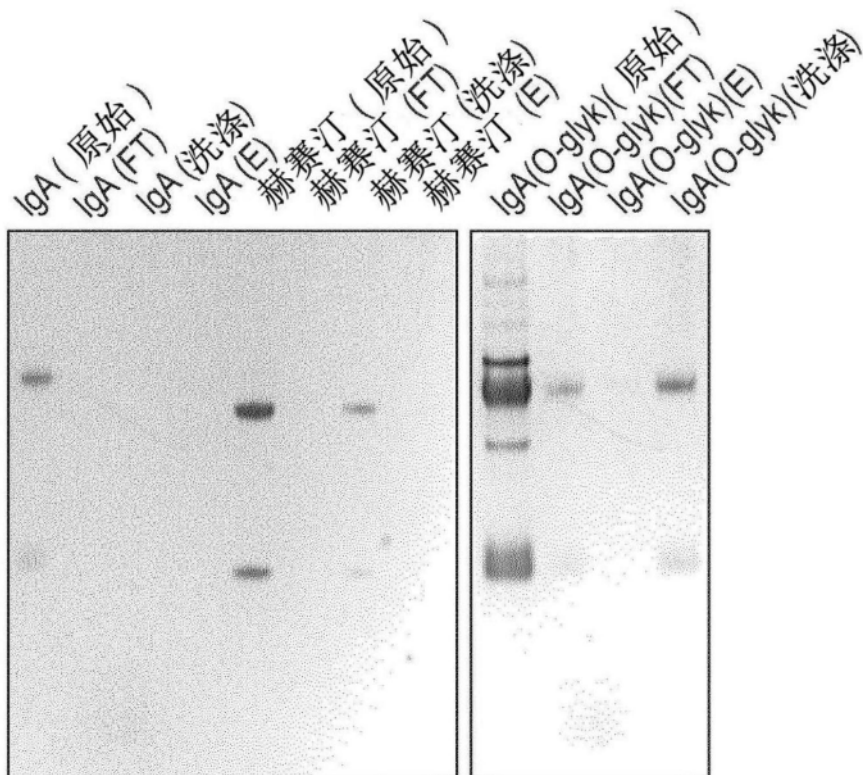


图6B

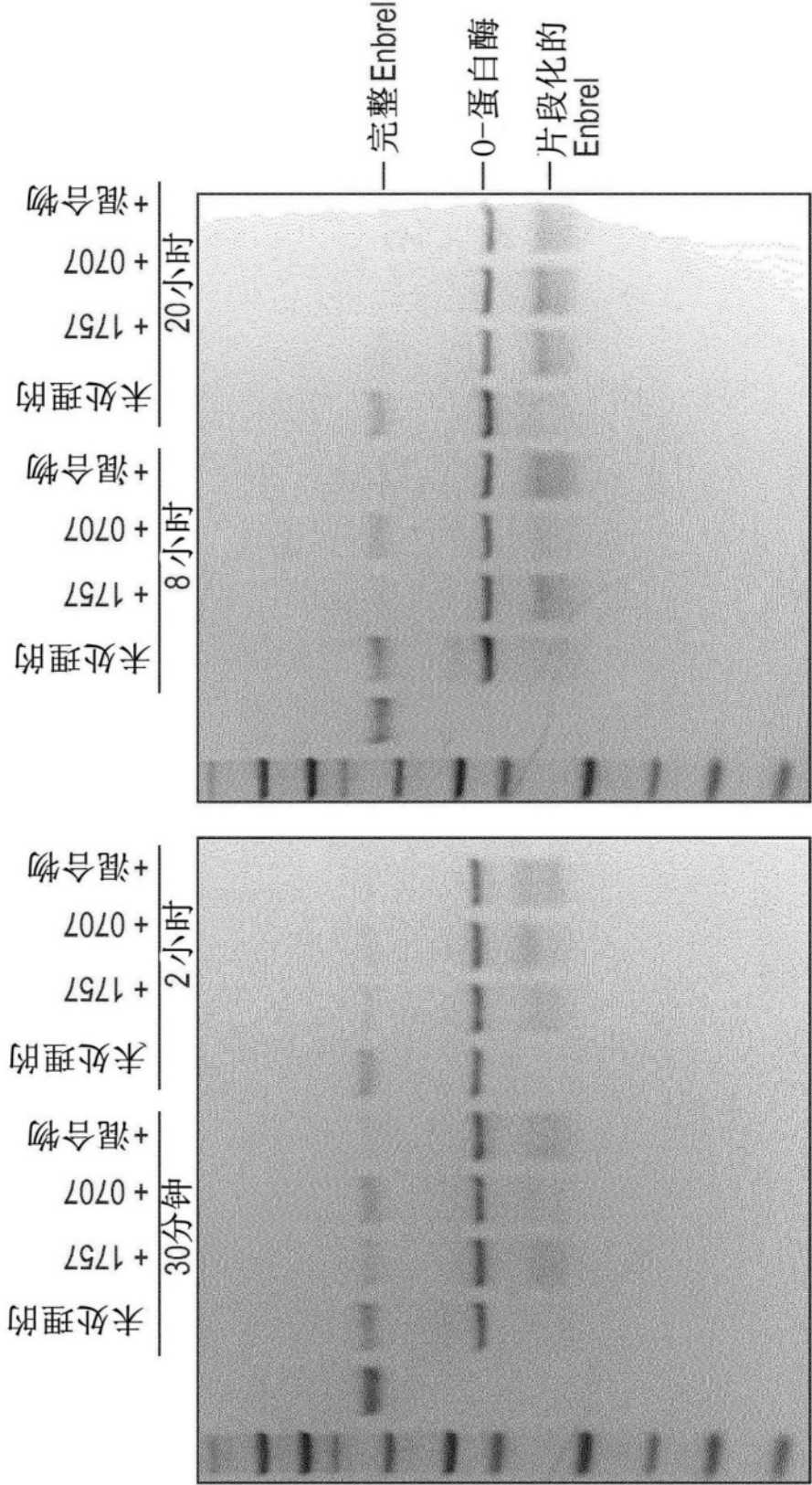


图7

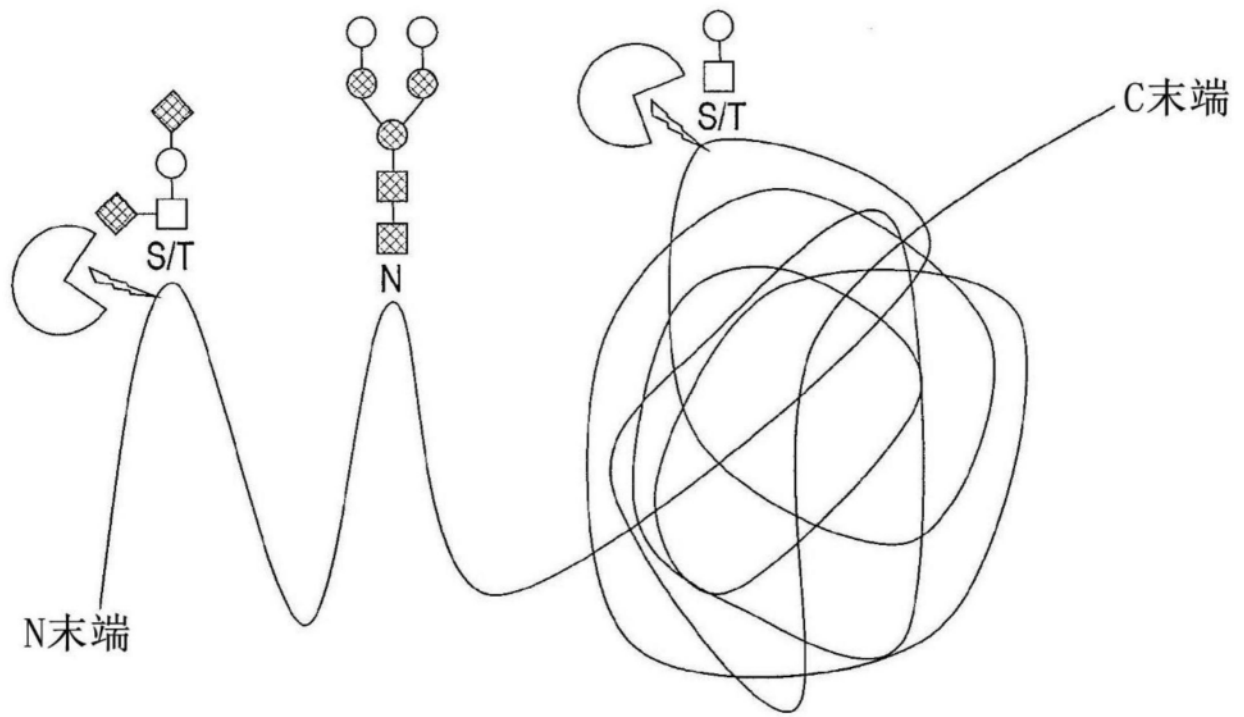


图8

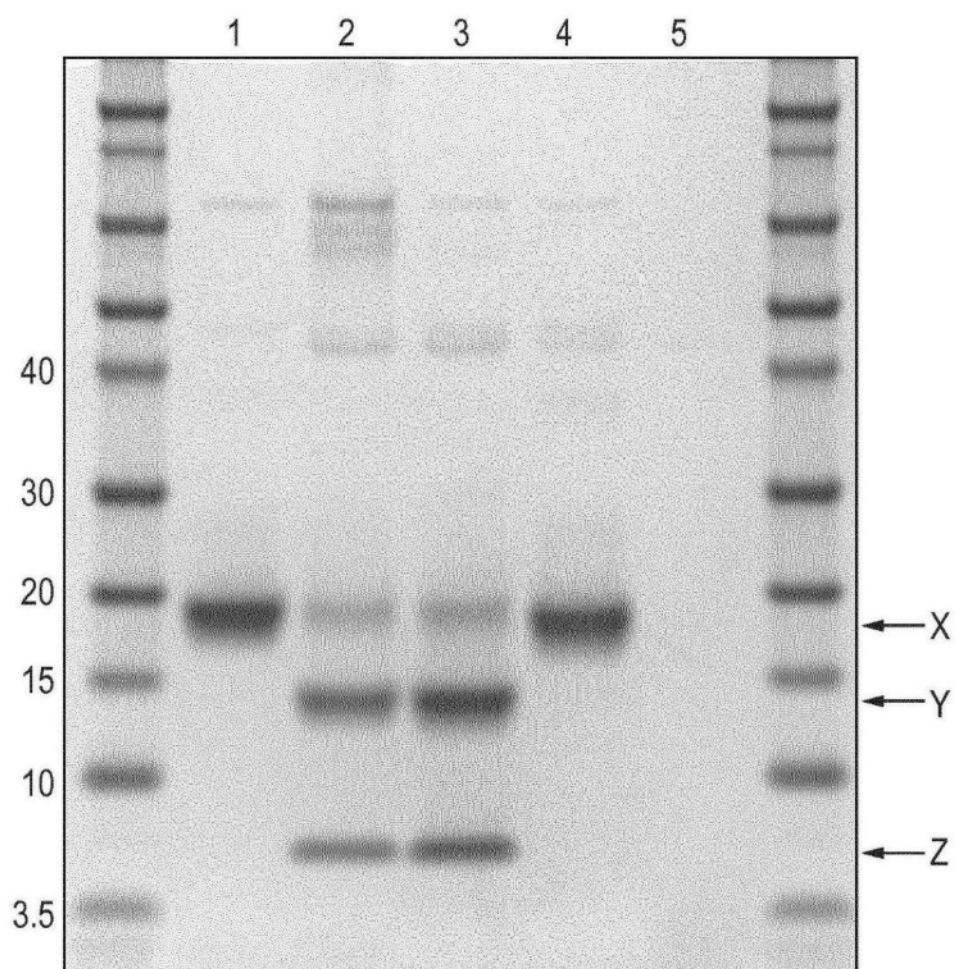


图9A

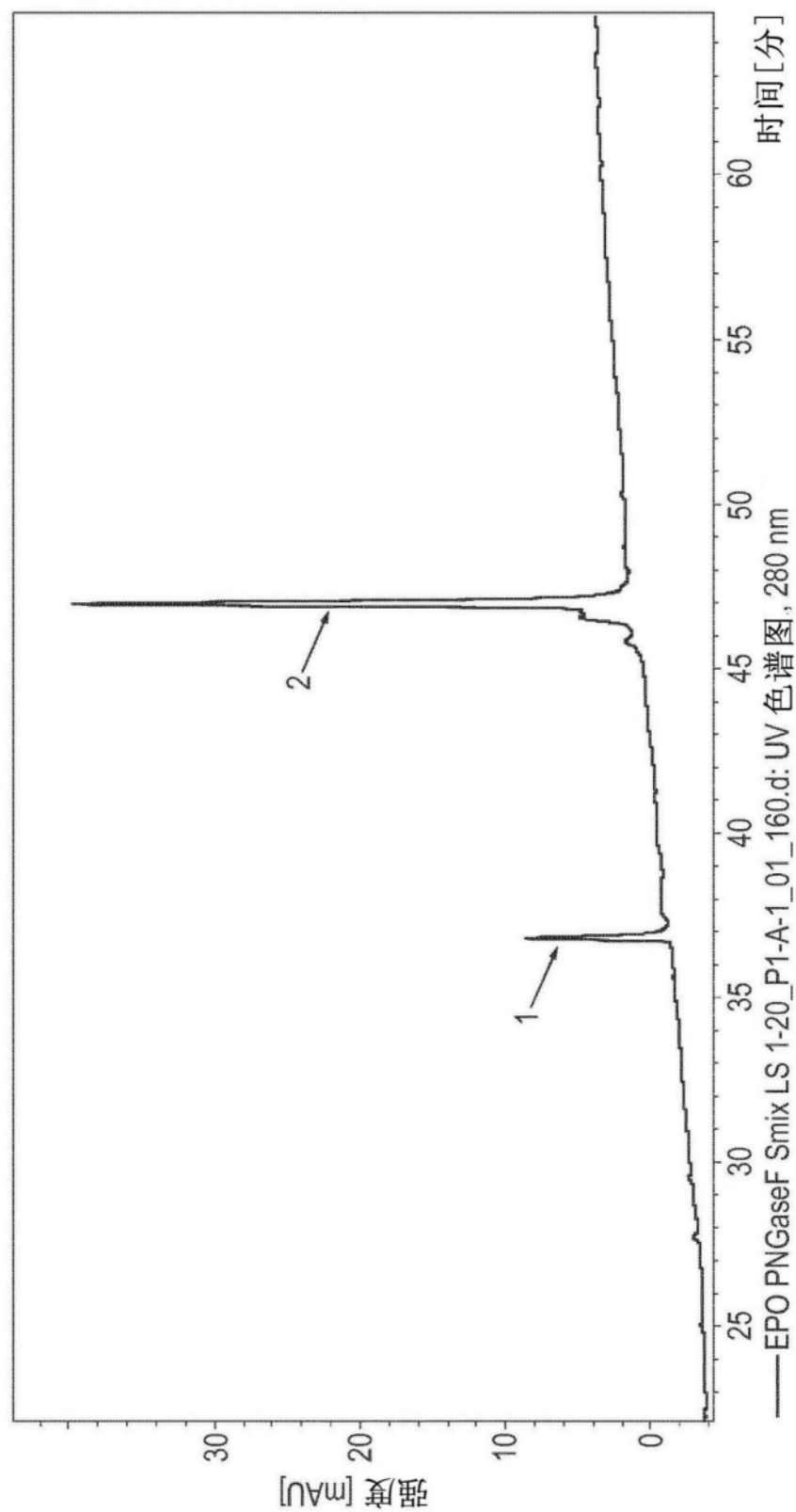


图9B

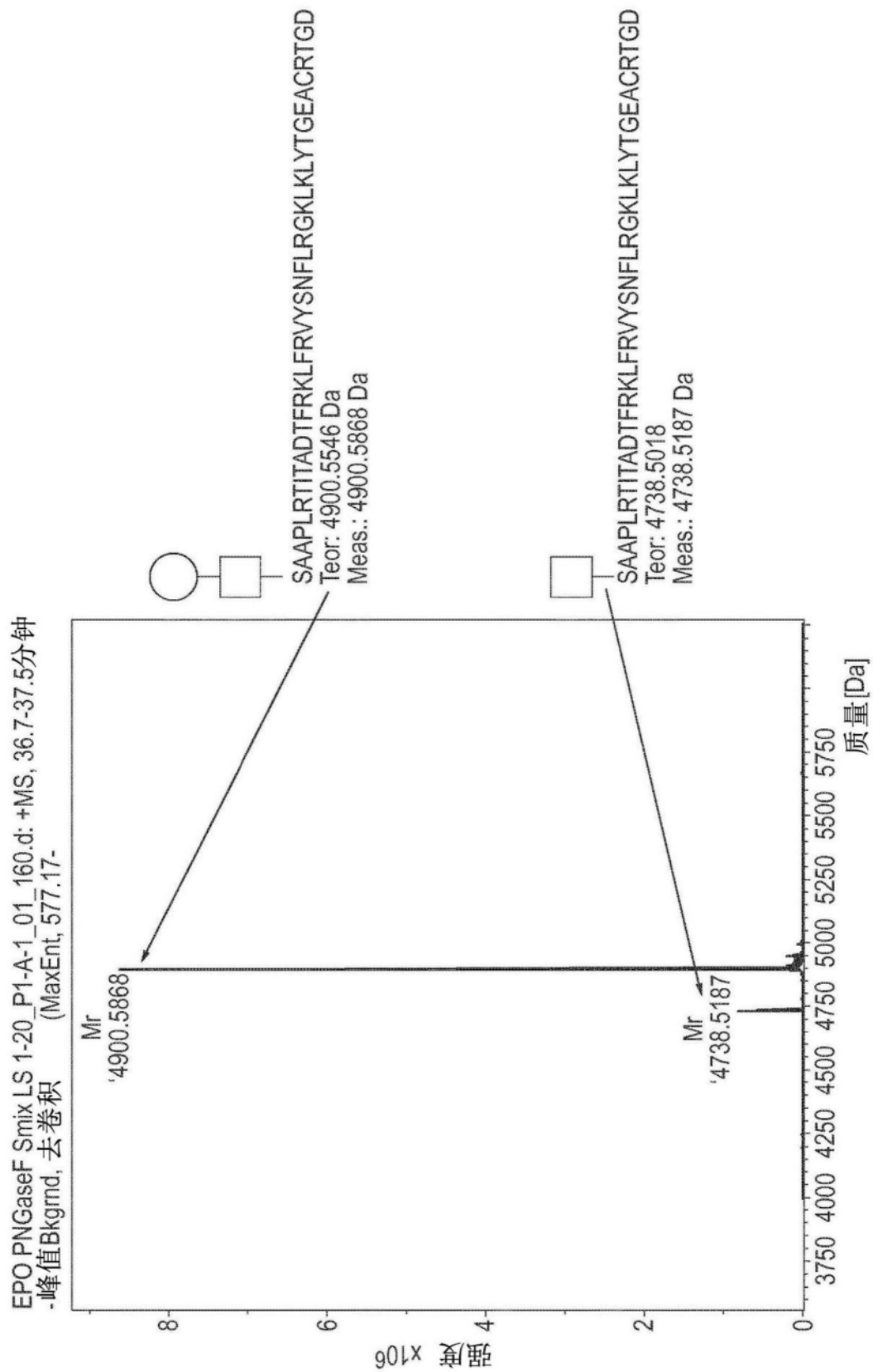


图9C

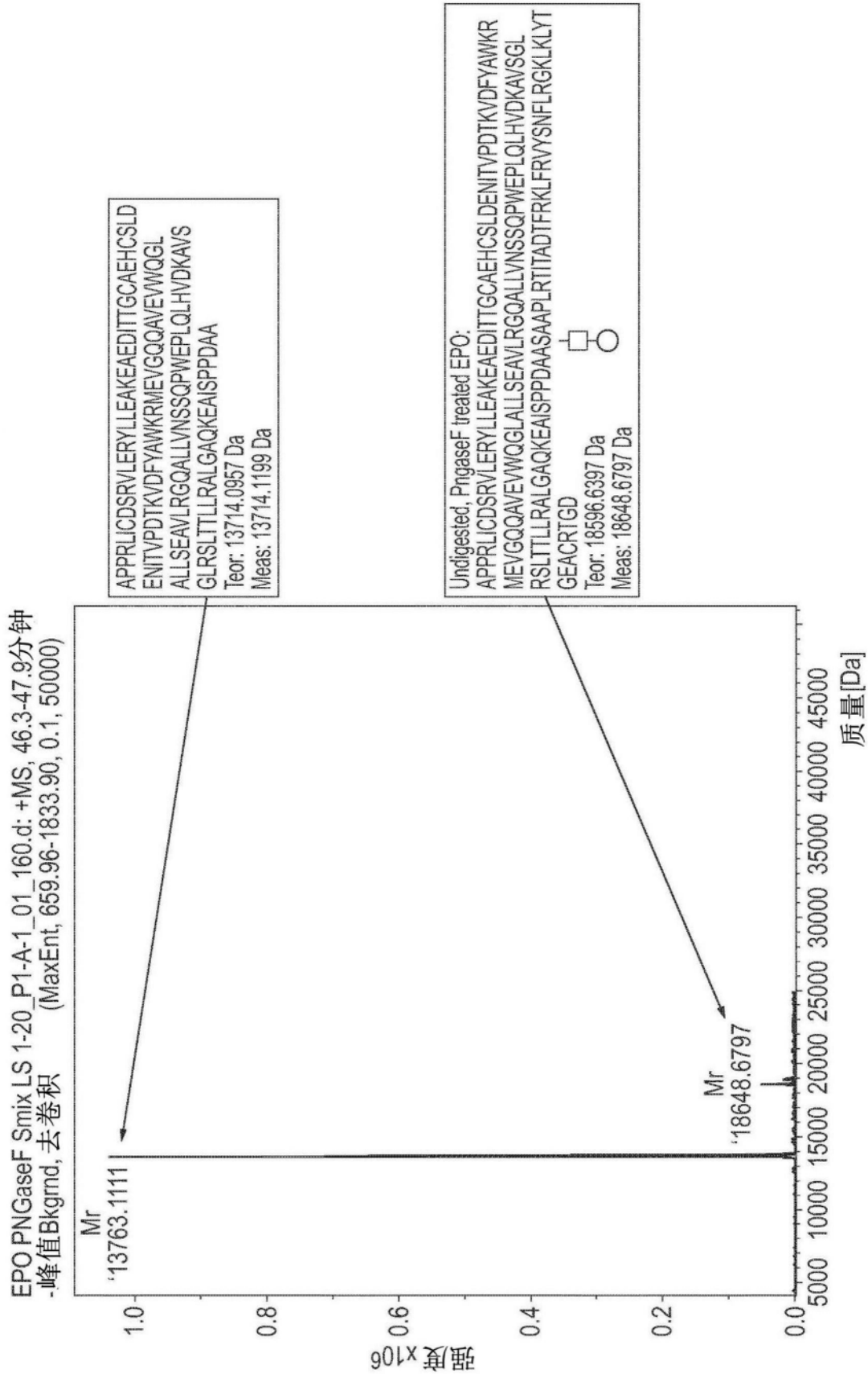


图9D

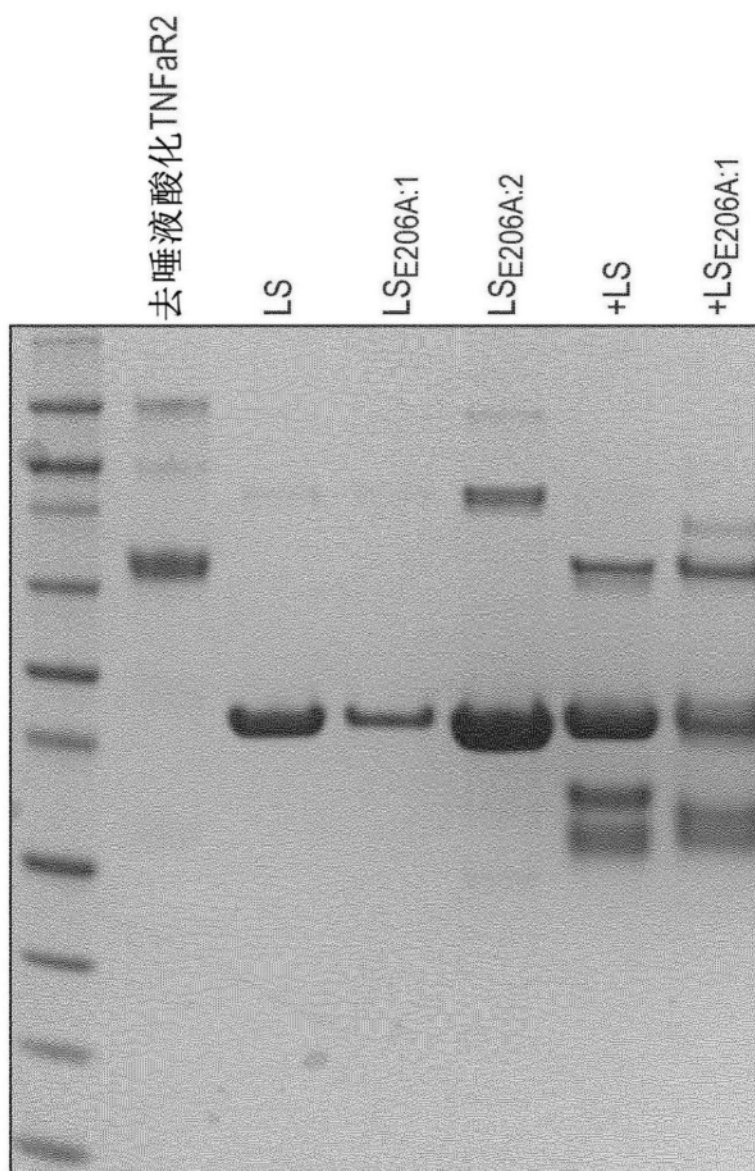


图10A

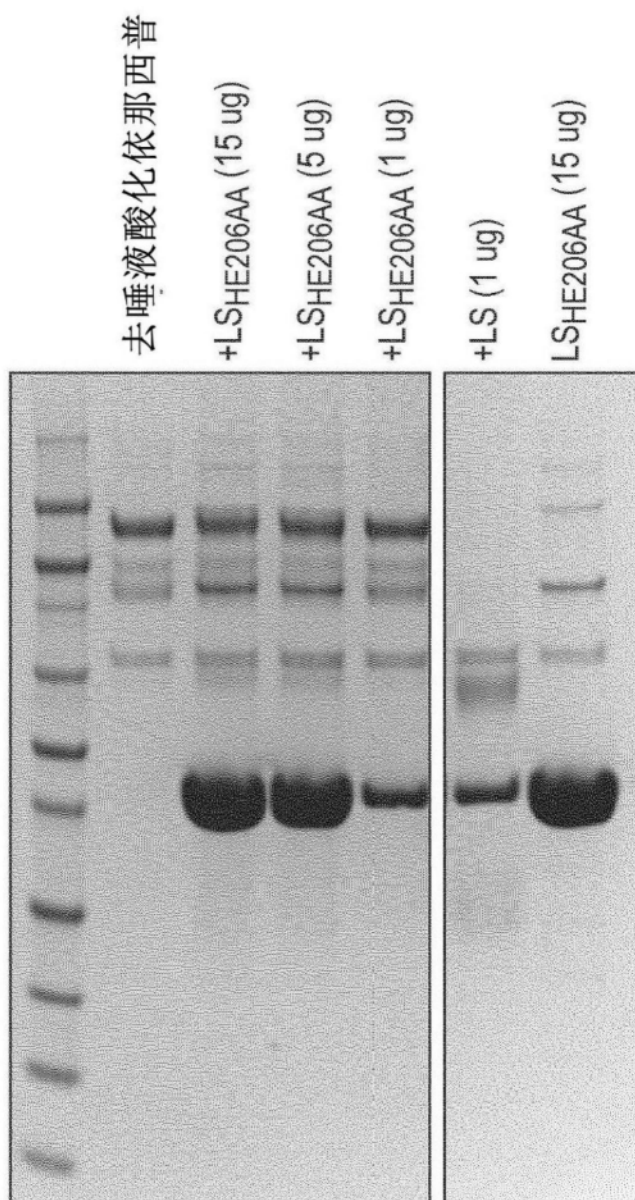


图10B

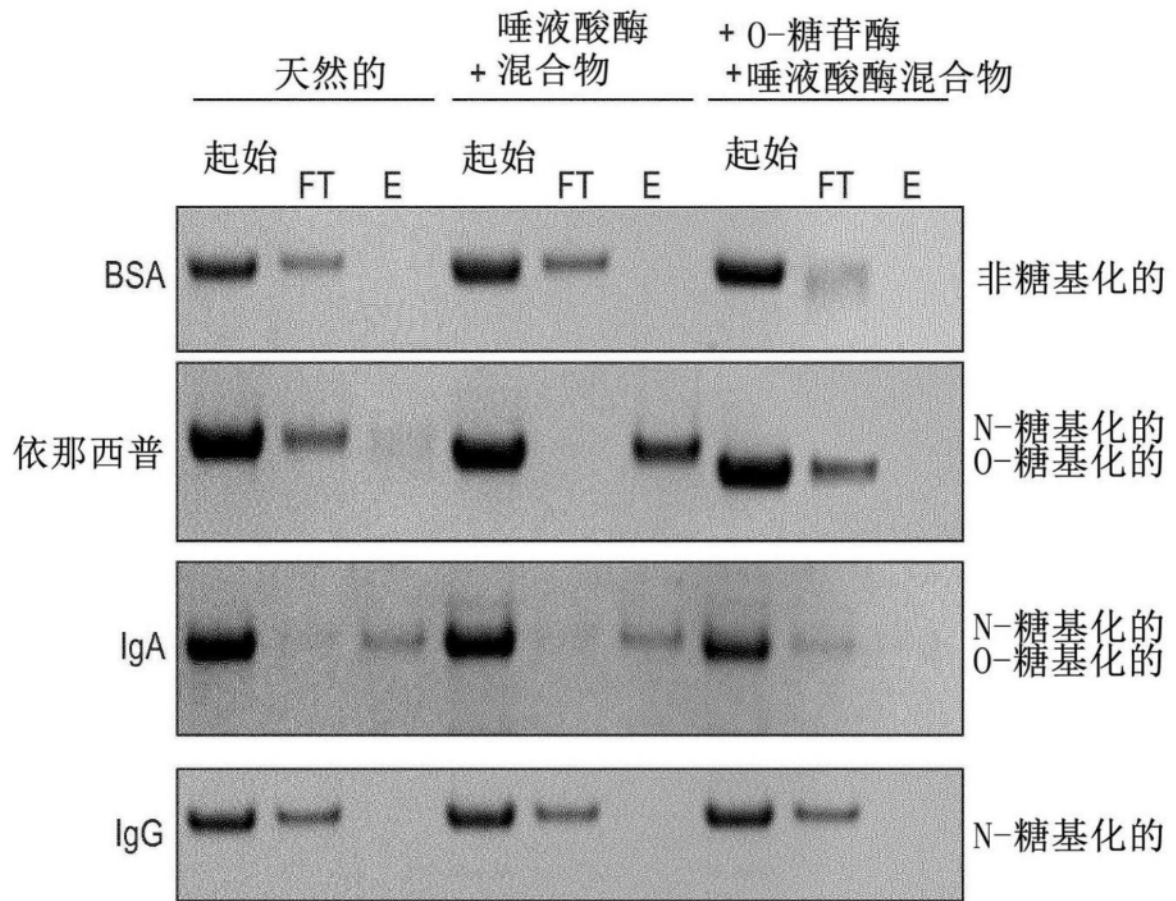


图11A

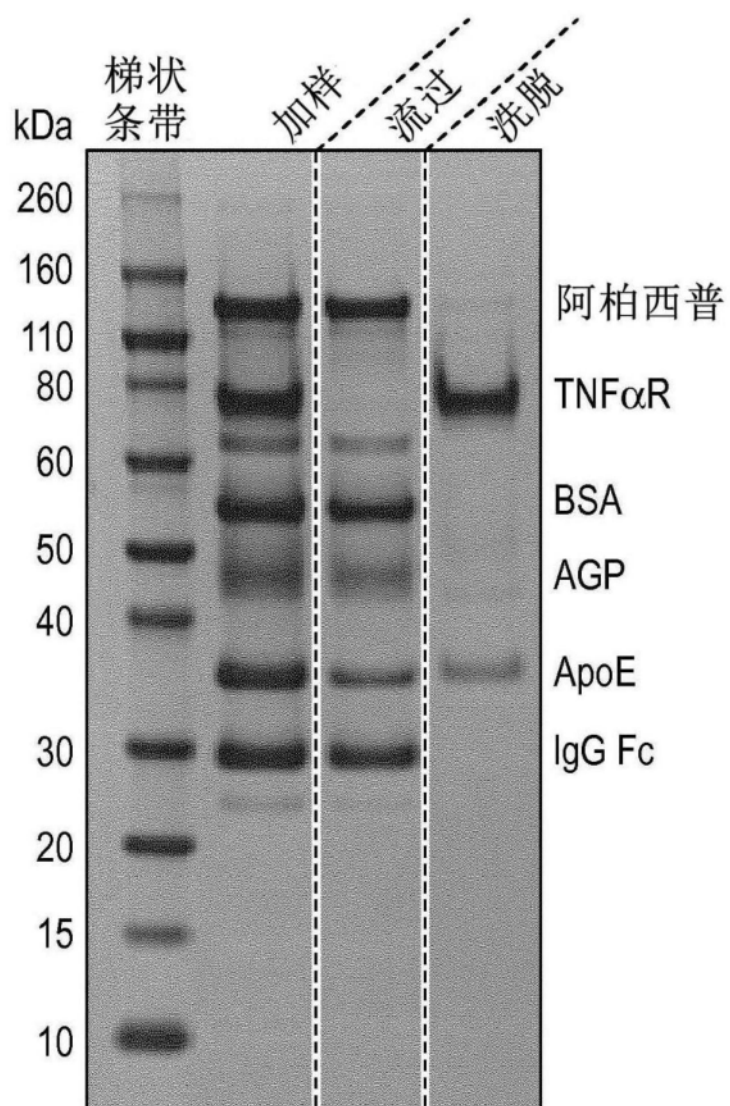


图11B

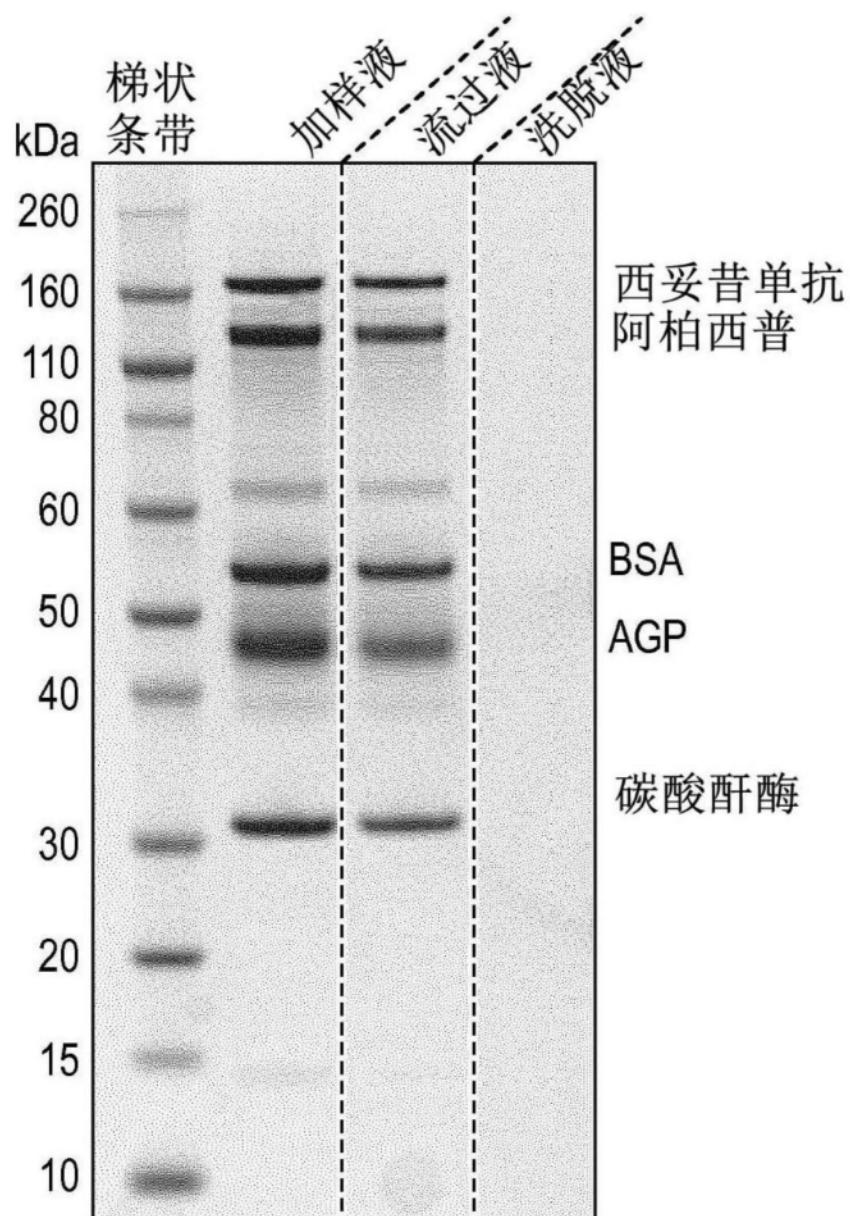


图11C

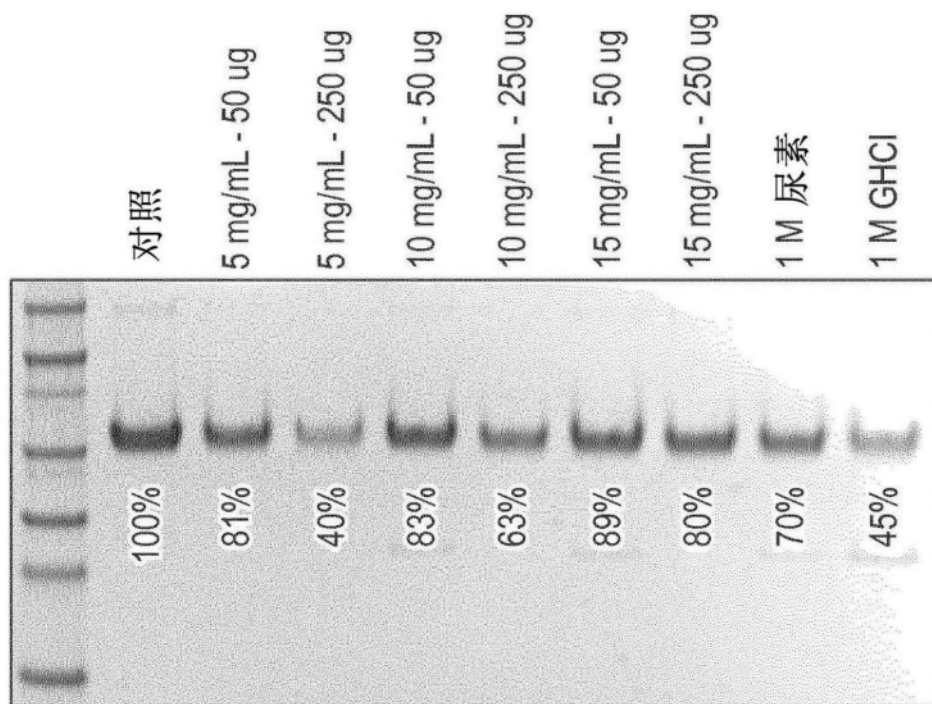


图12A

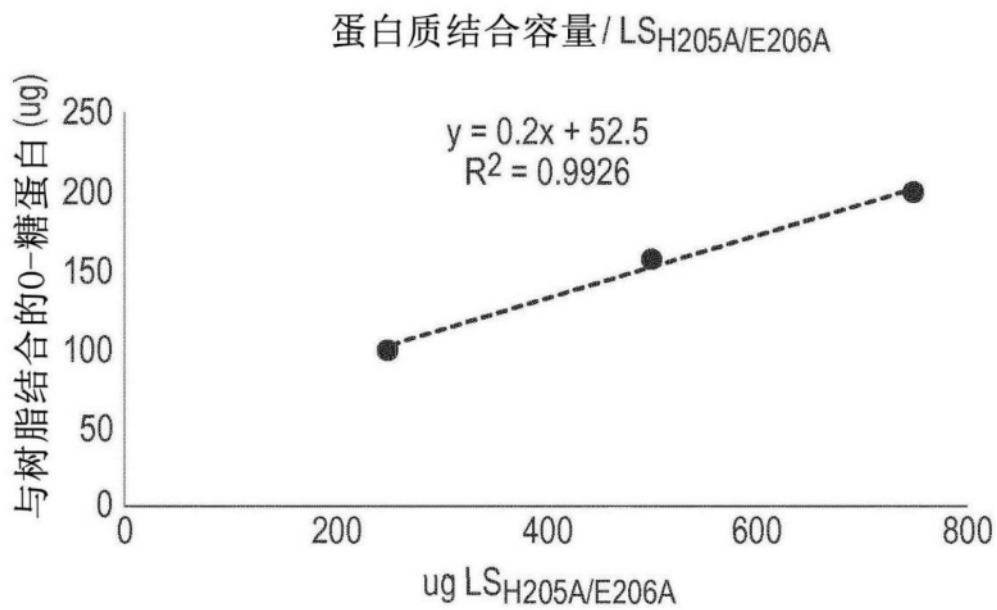


图12B

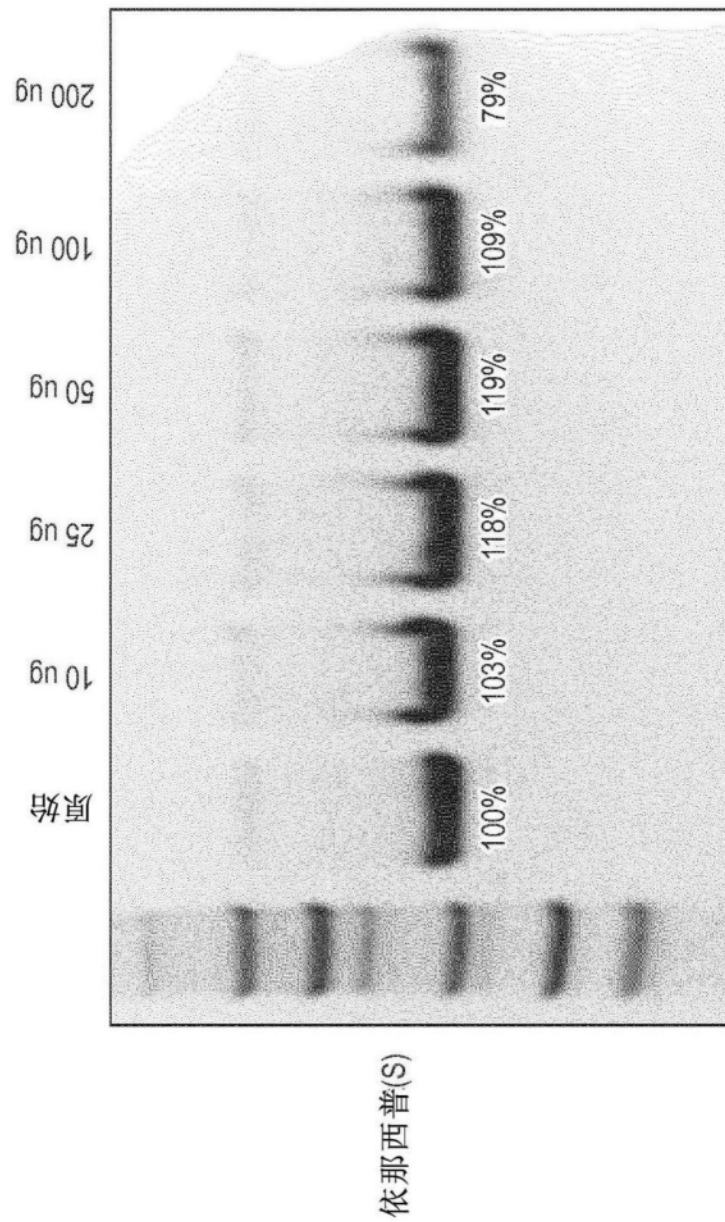


图13

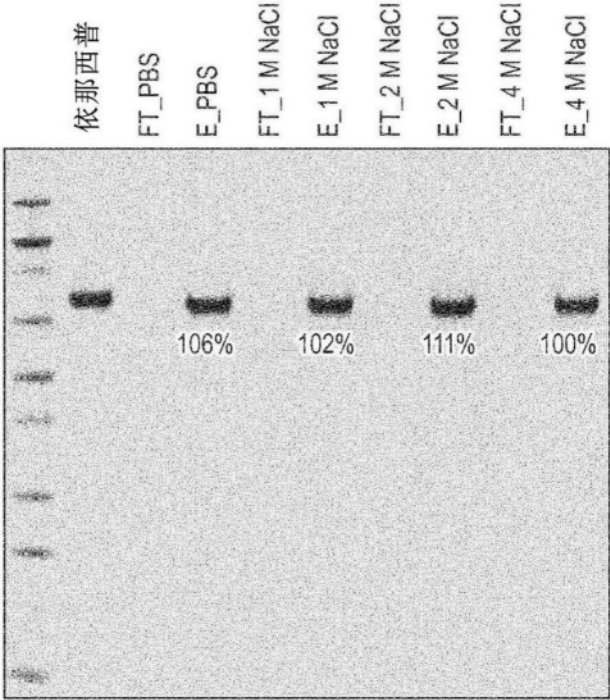


图14A

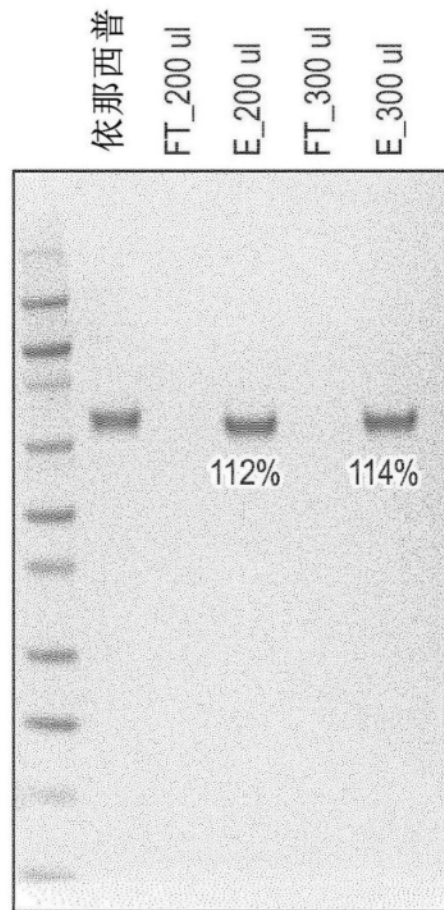


图14B

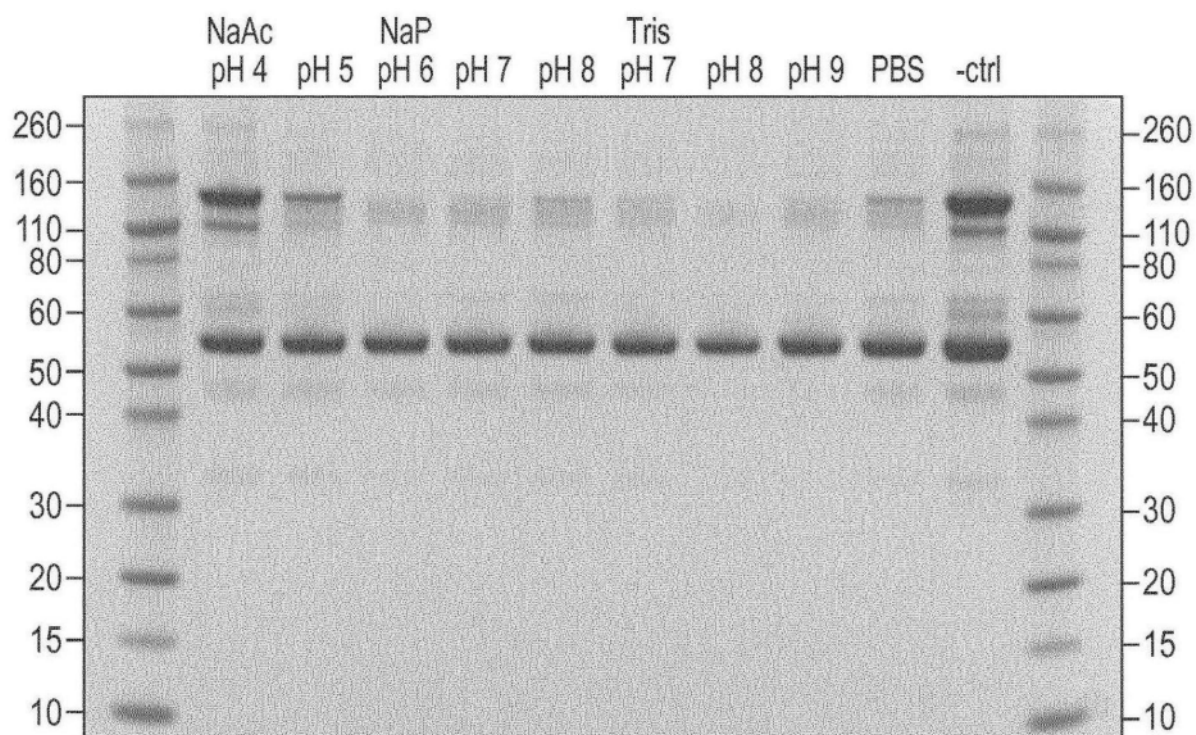


图14C

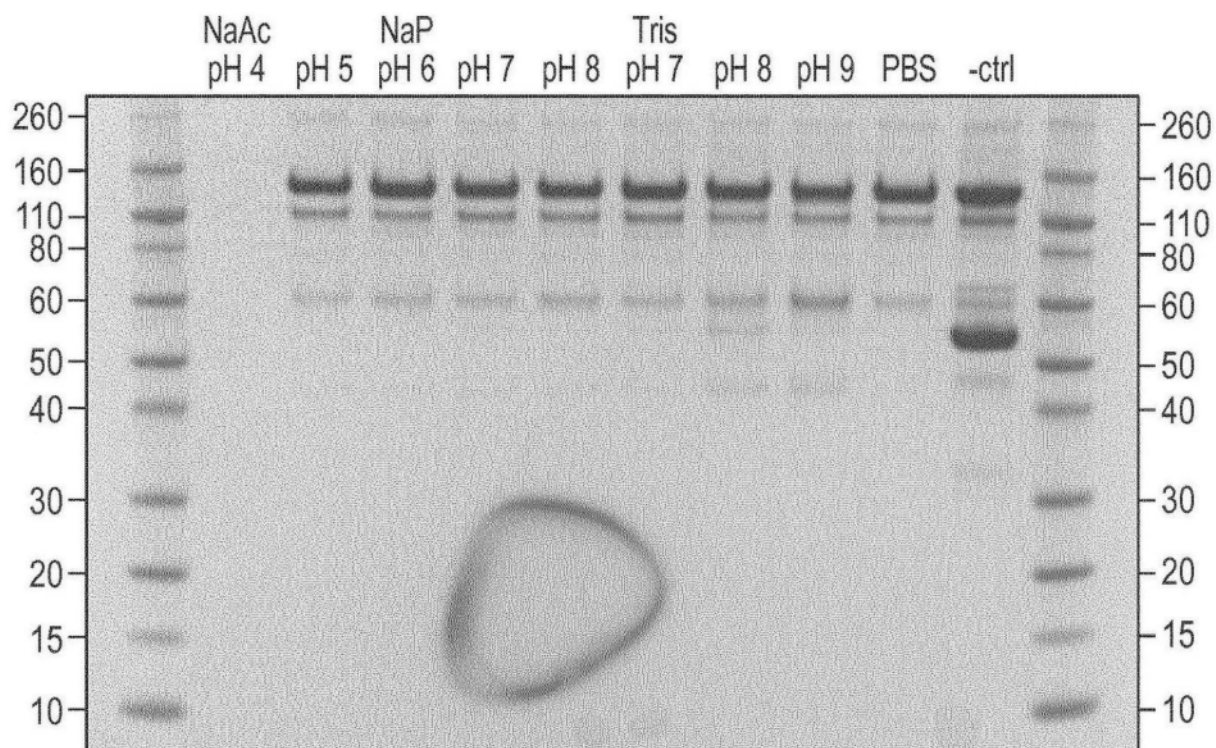


图14D

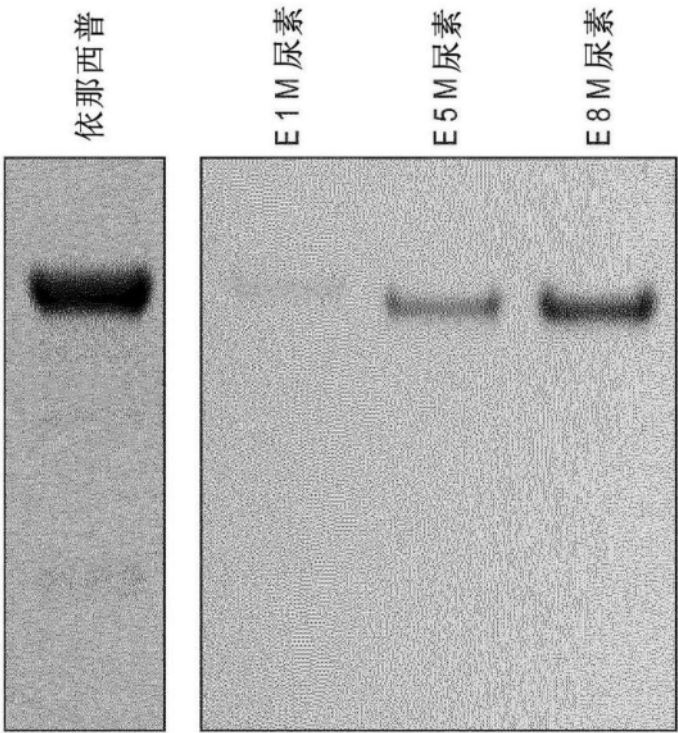


图15A

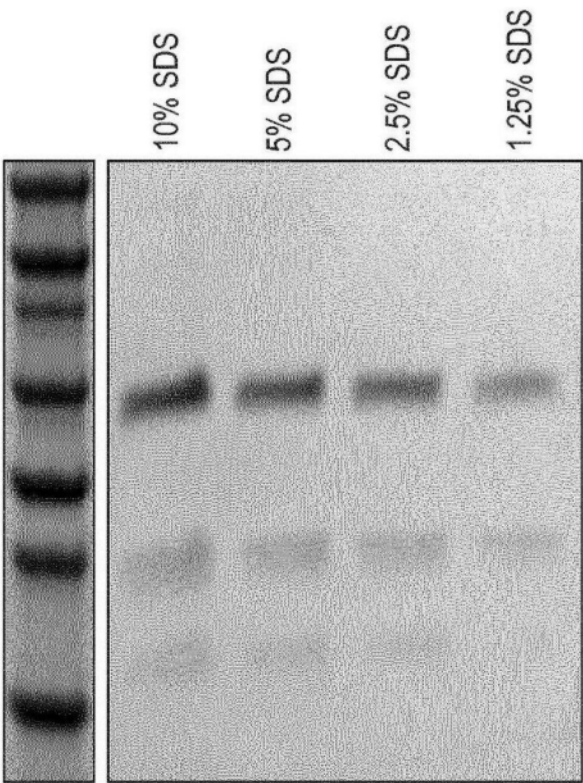


图15B

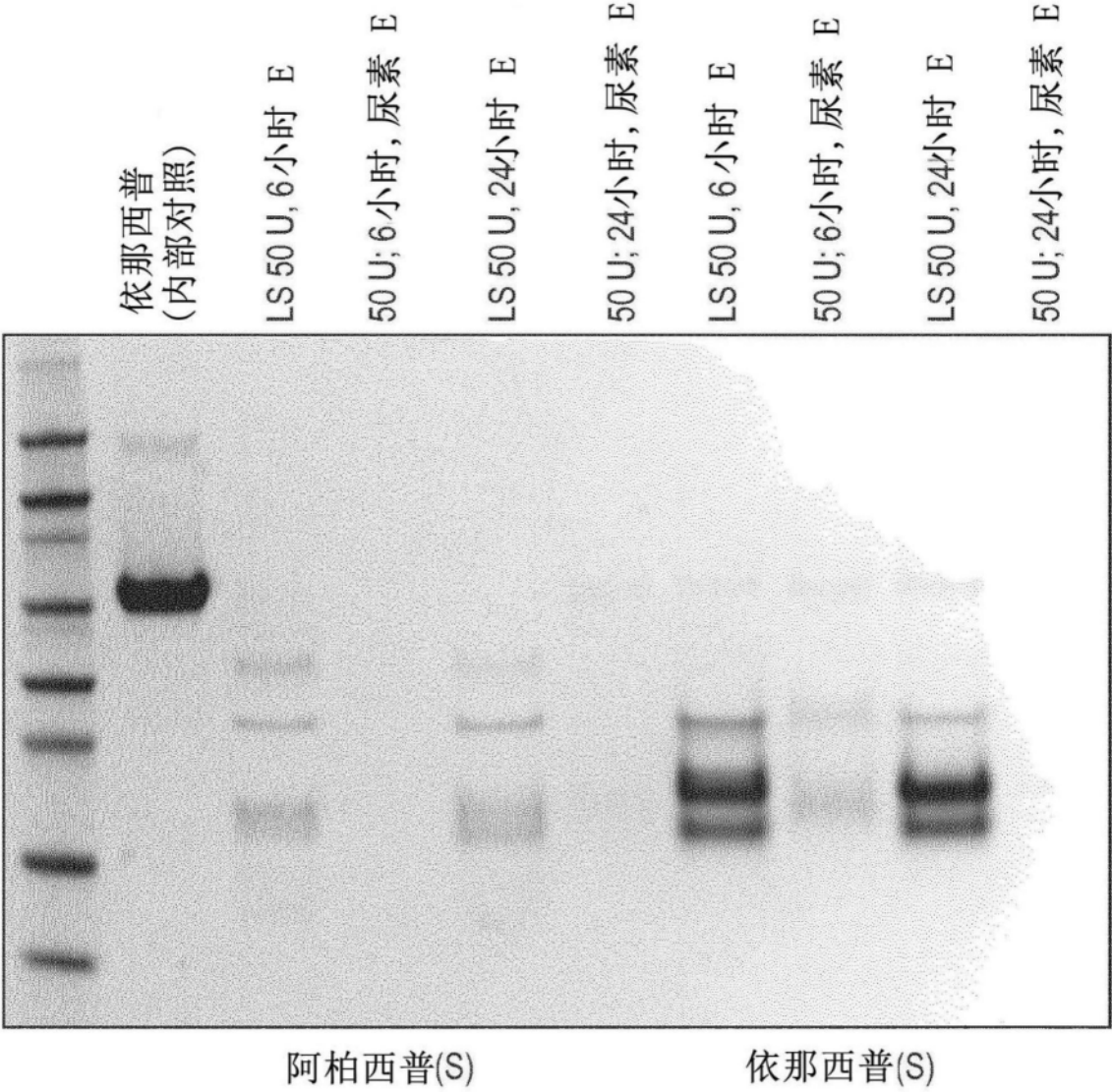


图16A

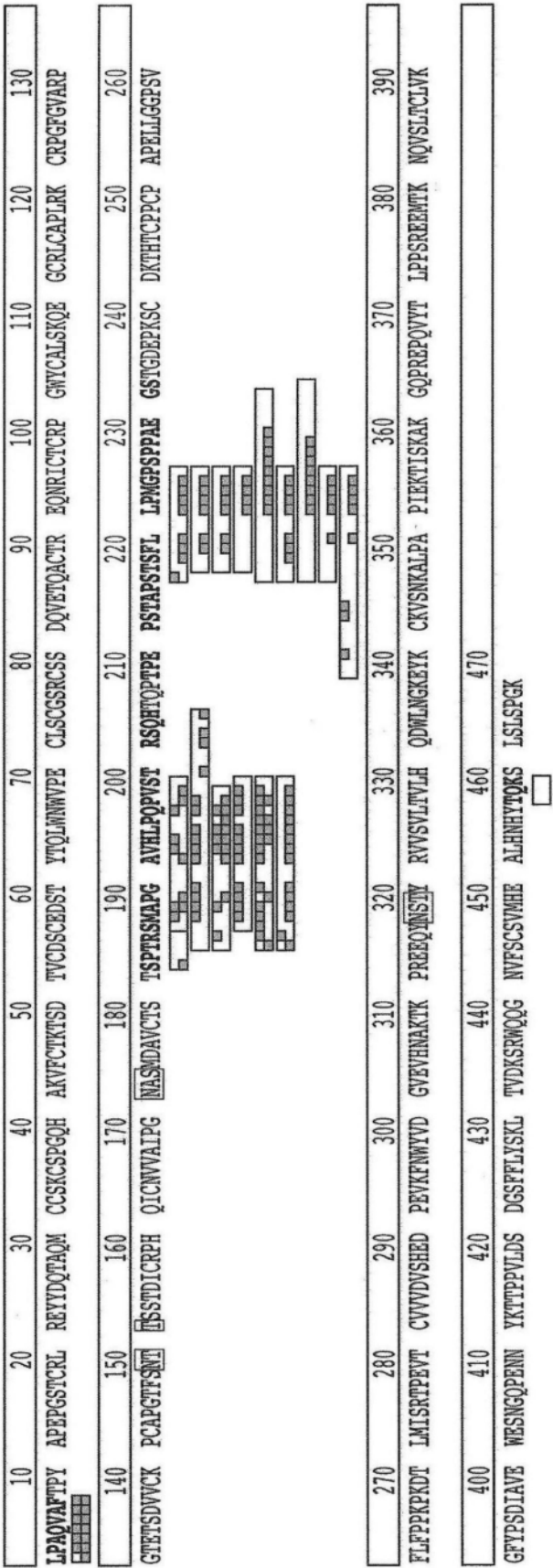


图16B.1

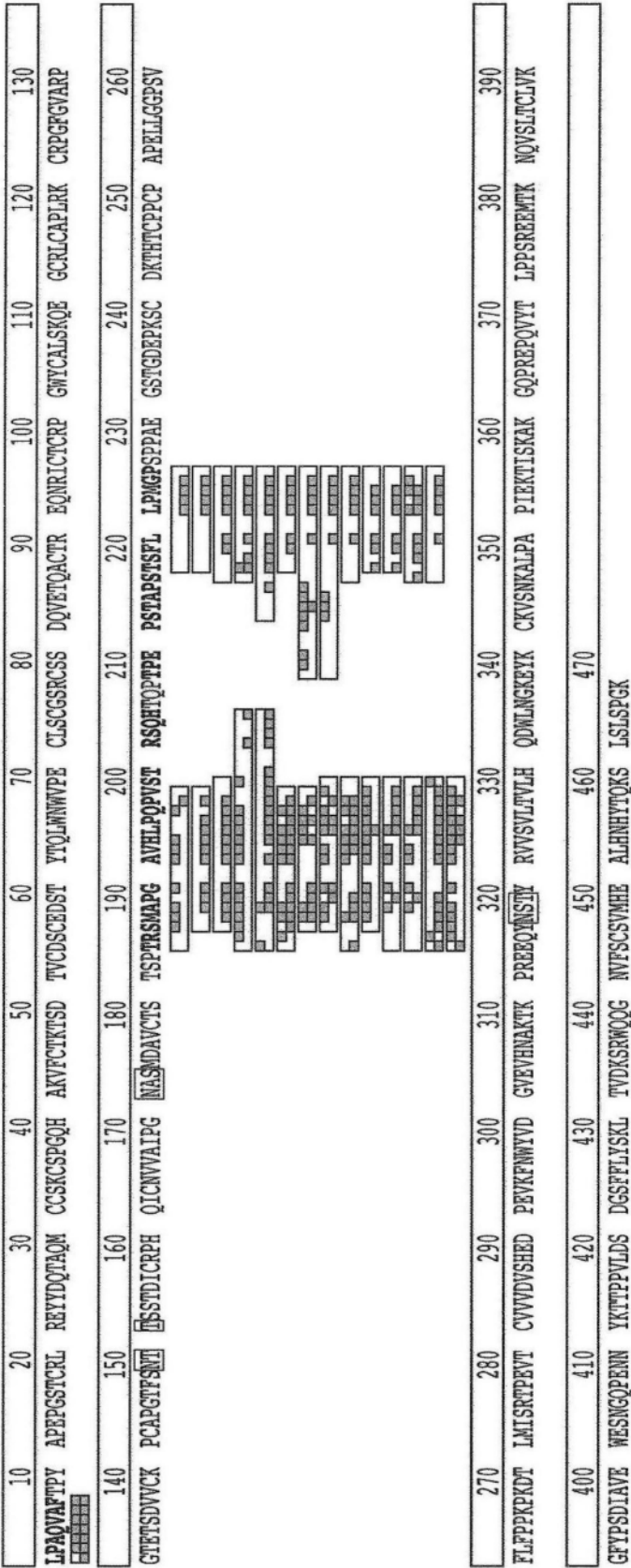


图16B.2

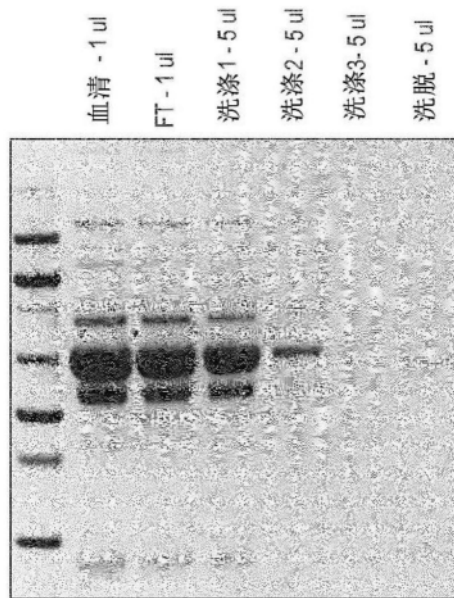


图17A

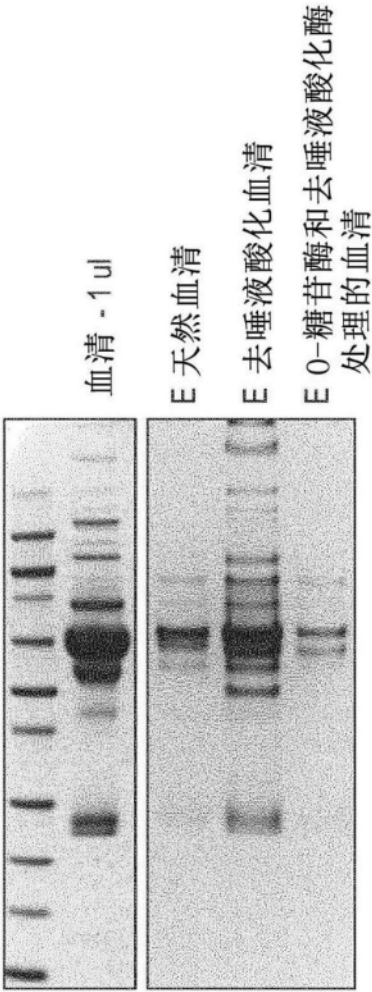


图17B

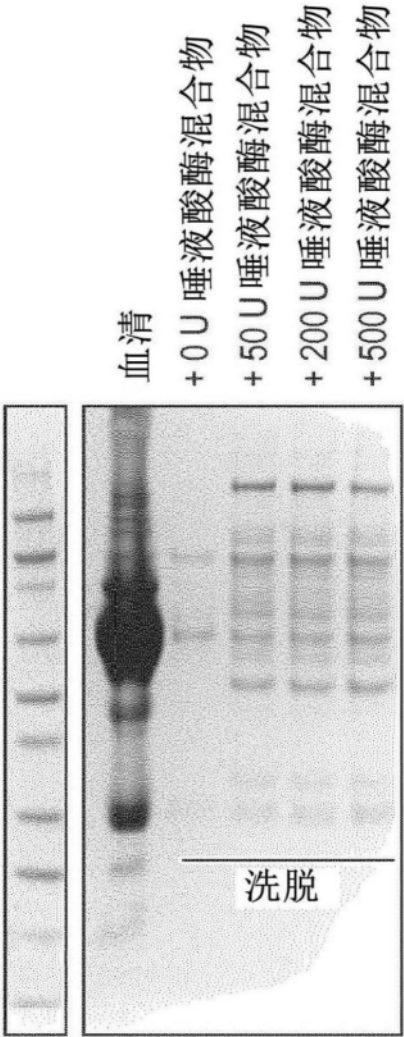


图17C

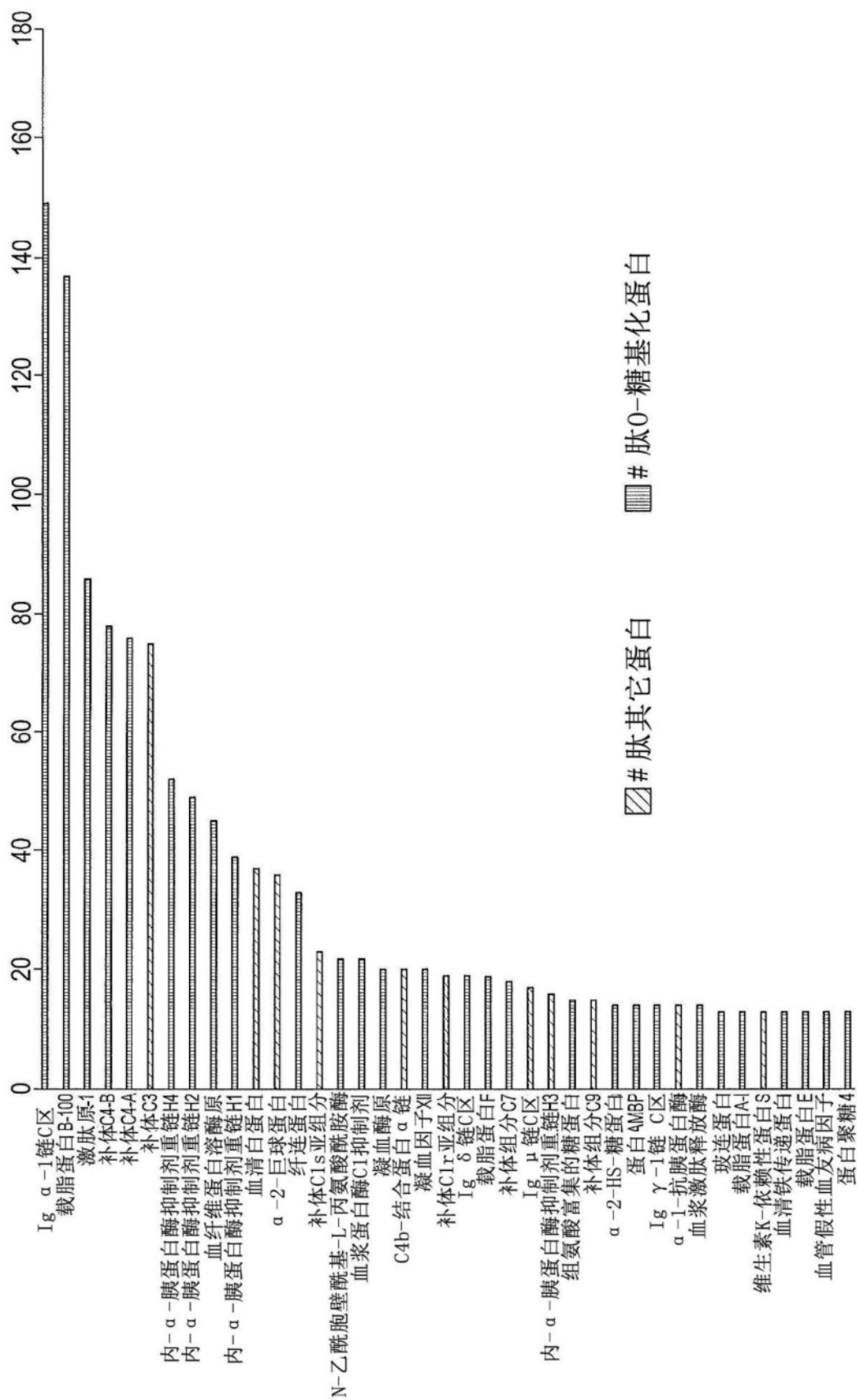


图18A

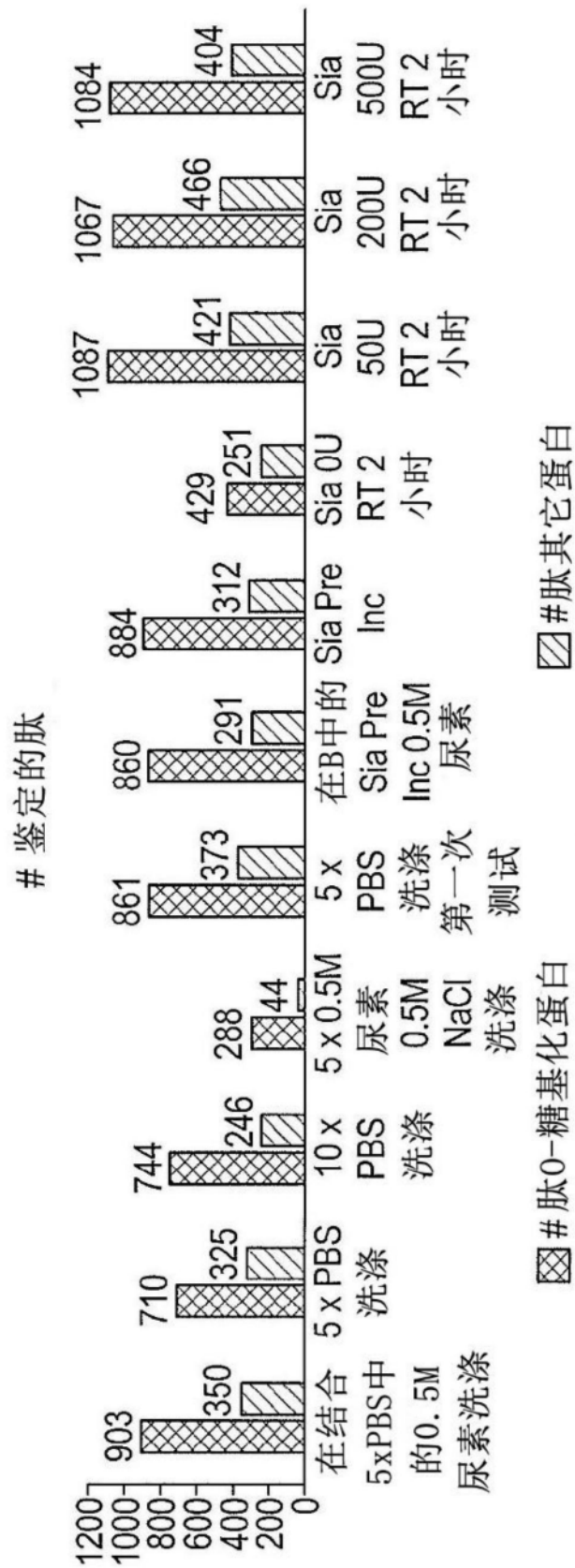


图18B

比例# 鉴定的O-糖基化蛋白肽/其它蛋白肽

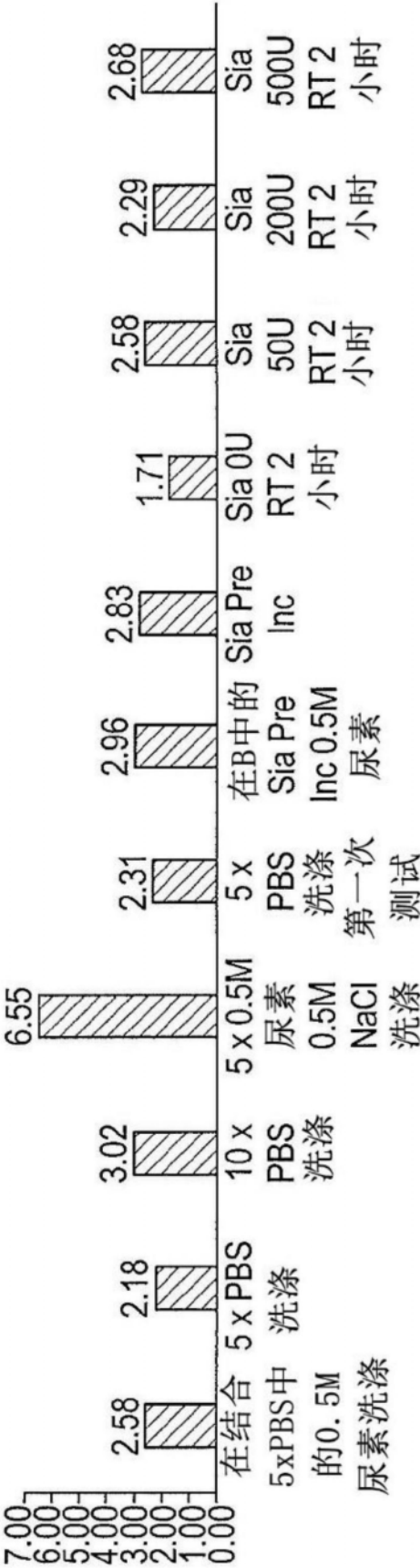


图18C

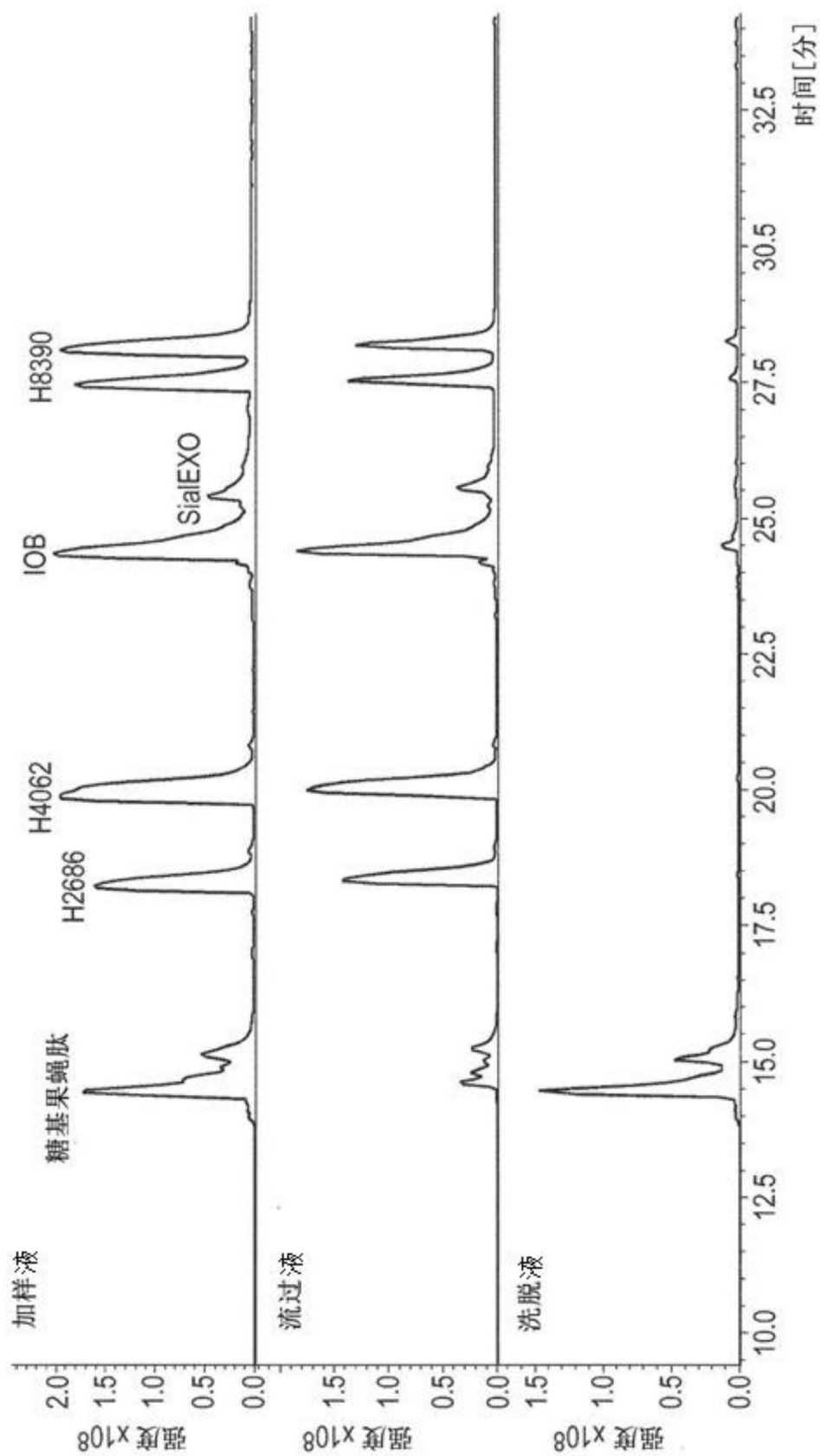


图19A

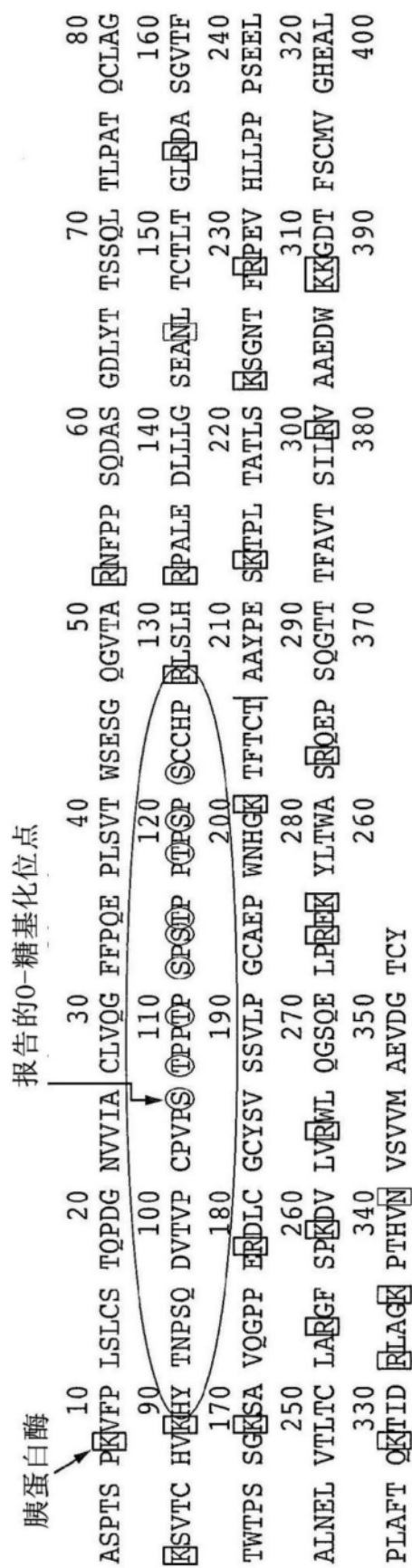


图19B

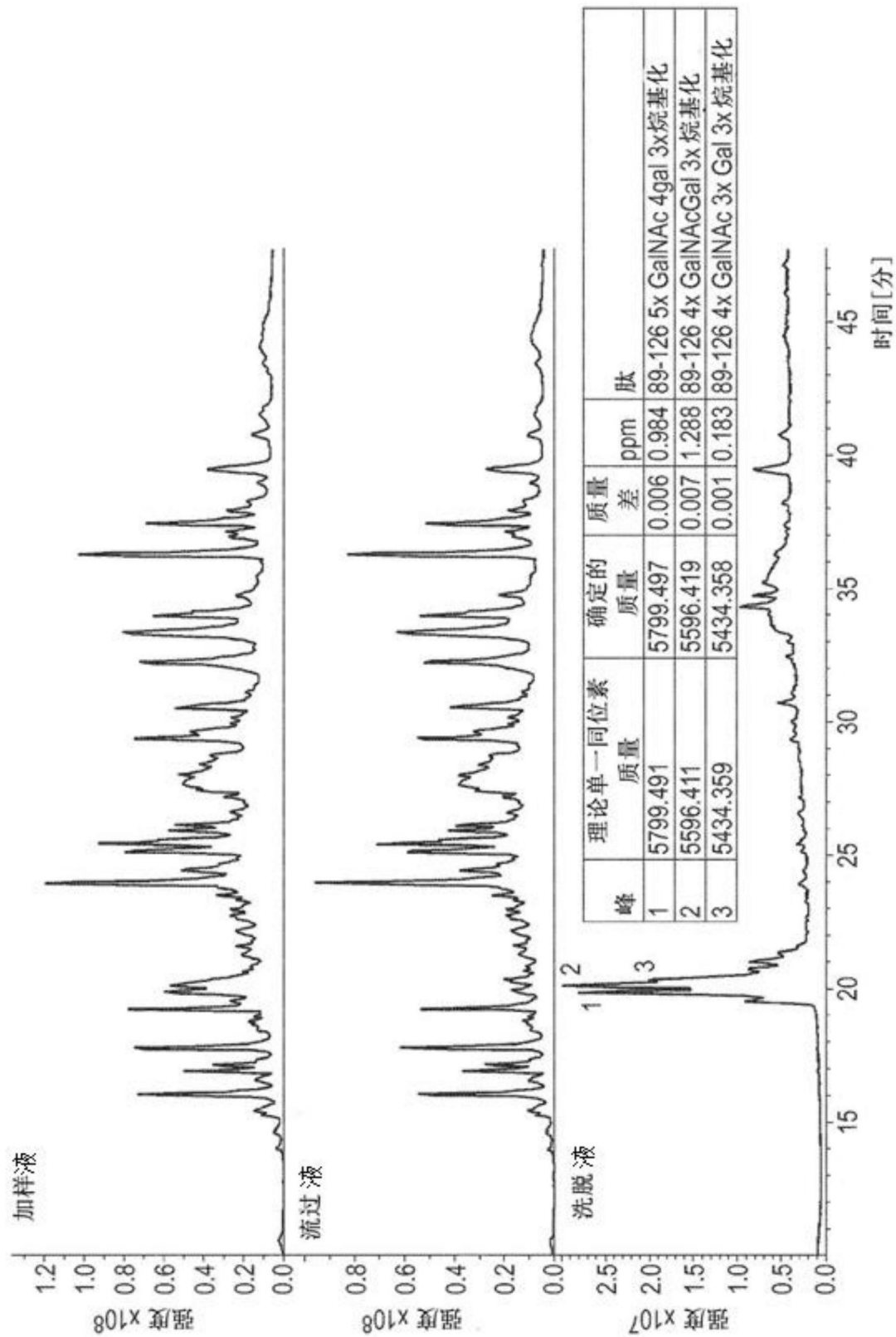


图19C

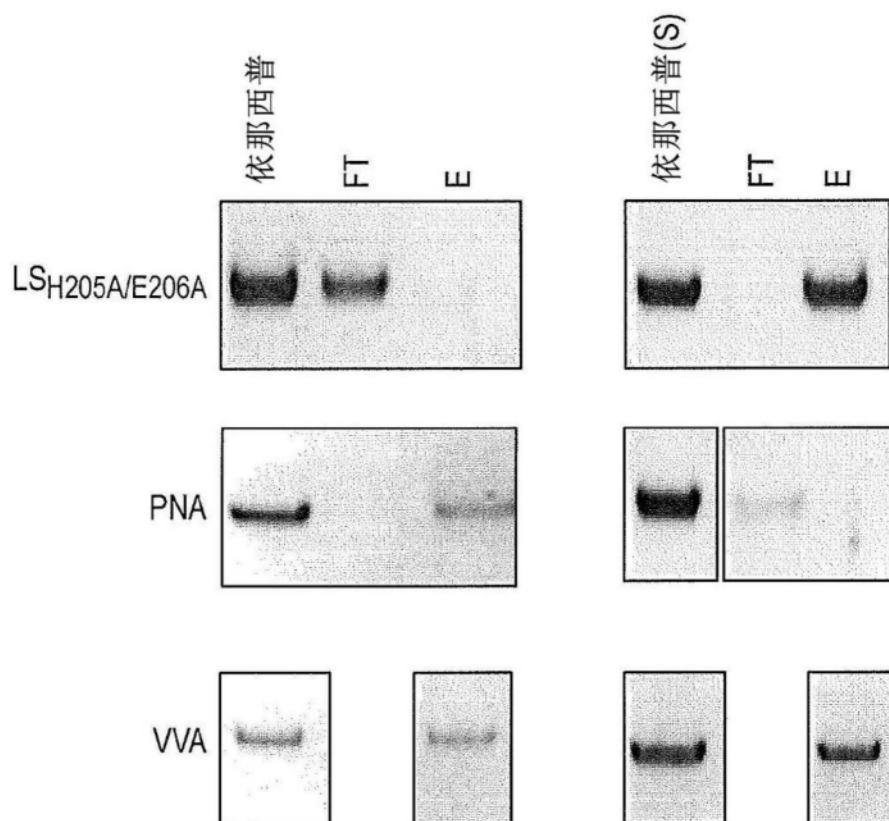


图20A

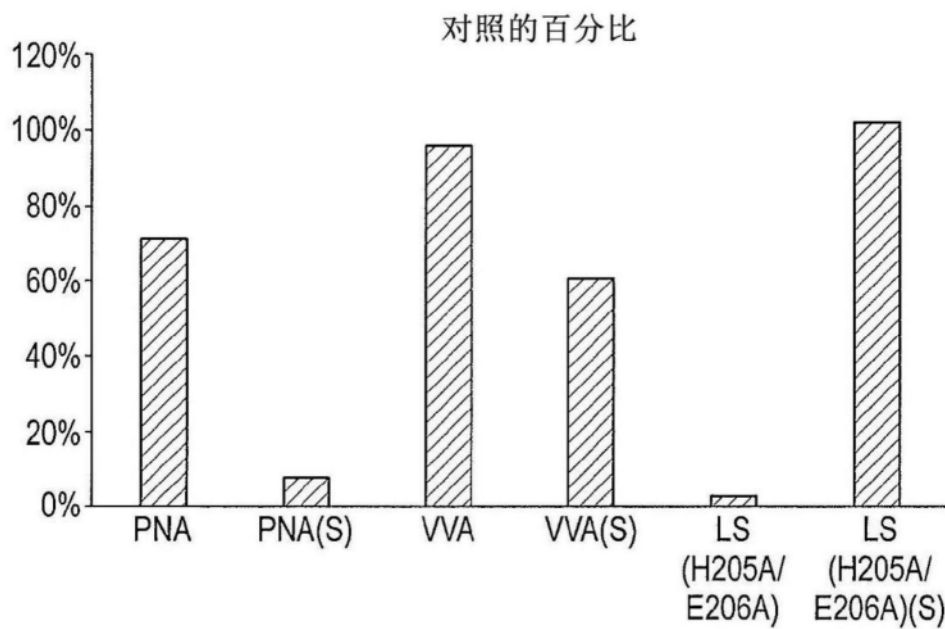


图20B