



## (12)发明专利申请

(10)申请公布号 CN 107532176 A

(43)申请公布日 2018.01.02

(21)申请号 201680024069.5

(74)专利代理机构 北京市柳沈律师事务所  
11105

(22)申请日 2016.04.26

代理人 张文辉

(30)优先权数据

1507104.6 2015.04.27 GB

(51)Int.Cl.

*C12N 15/62*(2006.01)

(85)PCT国际申请进入国家阶段日

*C12N 15/85*(2006.01)

2017.10.25

(86)PCT国际申请的申请数据

PCT/GB2016/051164 2016.04.26

(87)PCT国际申请的公布数据

W02016/174408 EN 2016.11.03

(71)申请人 UCL商务股份有限公司

地址 英国伦敦

(72)发明人 M.普莱 S.科多巴

权利要求书2页 说明书24页  
序列表48页 附图9页

(54)发明名称

核酸构建体

(57)摘要

本发明提供核酸构建体,其包含以下结构:  
A-X-B,其中A和B是编码第一和第二感兴趣的多  
肽(POI)的核酸序列;和X是编码切割位点的核酸  
序列,其中第一或者第二POI是包含细胞内保留  
信号的跨膜蛋白。

1. 核酸构建体,其包含以下结构:  
A-X-B  
其中  
A和B是编码第一和第二感兴趣的多肽 (POI) 的核酸序列;和  
X是编码切割位点的核酸序列,  
其中所述第一或者第二POI是包含细胞内保留信号的跨膜蛋白。
2. 根据权利要求1的核酸构建体,其中所述跨膜蛋白包含胞内域,其包含所述细胞内保留信号。
3. 根据权利要求1或2的核酸构建体,其中所述细胞内保留信号将所述跨膜蛋白引导为远离分泌途径。
4. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号将所述跨膜蛋白引导至膜结合的细胞内区室。
5. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号将所述跨膜蛋白引导至溶酶体、内体或高尔基体区室。
6. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号选自下组:内吞信号;高尔基体保留信号;反面高尔基网络 (trans-Golgi network, TGN) 再循环信号;内质网 (ER) 保留信号;和溶酶体分选信号。
7. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号是基于酪氨酸的分选信号、基于二亮氨酸的分选信号、酸性簇信号、溶酶体回避信号、NPFX' (1,2) D型信号、KDEL、KKX' X' 或KX' KX' X信号 (其中X' 是任意氨基酸)。
8. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号包含选自下组的序列:NPX' Y、YX' X' Z、[DE]X' X' X' L[LI]、DX' X' LL、DP[FW]、FX' DX' F、NPF、LZX' Z[DE]、LLDLL、PVDLW、KDEL、HDEL、KKX' X' 或KX' KX' X';其中X' 是任意氨基酸且Z' 是具有大体积的疏水性侧链的氨基酸。
9. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号包含表2至表6中所示的任意序列。
10. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号包含酪氨酸酶相关蛋白 (TYRP)-1细胞内保留信号或腺病毒E3/19K保留信号。
11. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号包含序列NQPLLTD (SEQ ID No.35), KYKSRRSFIDEKKMP (SEQ ID No.36);或DEKKMP (SEQ ID No.37)。
12. 根据前述权利要求中任一项的核酸构建体,其中所述细胞内保留信号靠近所述跨膜蛋白的跨膜结构域。
13. 根据权利要求1至11中任一项的核酸构建体,其中所述细胞内保留信号远离所述跨膜蛋白的跨膜结构域。
14. 根据权利要求1至11中任一项的核酸构建体,其中X是编码自切割肽、弗林蛋白酶切割位点或烟草蚀刻病毒切割位点的核酸序列。
15. 根据权利要求14的核酸构建体,其中X编码来自口疮病毒或心脏病毒的2A自切割肽或2A样肽。
16. 根据权利要求15的核酸构建体,其中所述跨膜蛋白是嵌合抗原受体 (CAR)。

17. 根据前述权利要求中任一项的核酸构建体,其中所述第一和第二POI各自是跨膜蛋白;其中所述跨膜蛋白的至少一个包含如权利要求3至11任一项中定义的细胞内保留信号。

18. 根据前述权利要求中任一项的核酸构建体,其中所述第一和第二POI之一或两者是单通跨膜蛋白。

19. 根据权利要求18的核酸构建体,其中所述第一和第二POI之一或两者是I型跨膜蛋白。

20. 核酸构建体,其包含以下结构:

A-X-B-Y-C

其中

A、B和C是编码第一、第二和第三感兴趣的多肽(POI)的核酸序列;和

X和Y是可以为相同或不同的核酸序列,其各自编码切割位点,

其中所述POI中的至少两个是包含细胞内保留信号的跨膜蛋白。

21. 根据权利要求20的核酸序列,其中是跨膜蛋白并包含细胞内保留信号的至少两个:

(a) 包含不同的细胞内保留信号;和/或

(b) 具有位于所述POI中的不同位置处的细胞内保留信号,

使得当所述核酸在细胞中表达时,在所述细胞表面存在所述至少两个POI的差异相对表达。

22. 载体,其包含根据前述权利要求中任一项的核酸构建体。

23. 根据权利要求22的逆转录病毒载体或慢病毒载体或转座子。

24. 细胞,其包含根据权利要求1至21中任一项的核酸构建体或根据权利要求22或23的载体。

25. 用于调节从单个核酸构建体与第二蛋白一起表达的跨膜蛋白的相对细胞表面表达的方法,其通过在编码所述跨膜蛋白的核酸序列中包括细胞内保留信号进行。

26. 根据权利要求25的方法,其中所述核酸序列是如权利要求1至21任一项中所定义的。

## 核酸构建体

### 发明领域

[0001] 本发明涉及用于调节从单个载体共表达的多肽的相对表达的构建体和方法。特别地,本发明涉及调节从单个载体与第二多肽共表达的跨膜蛋白的表达。

### [0002] 发明背景

[0003] 通常希望从同一载体表达不同的蛋白质,因为同一细胞的多重转导是困难的,昂贵的和不可预测的。因此已经开发了不同的方法以允许从单个载体共表达两种蛋白(参见图1)。

[0004] 最初的尝试在同一盒中使用两种不同的启动子。这产生两个单独的转录物,其中每个编码单独的蛋白质。因为一些原因,这是困难的方法。一个关键问题是“启动子干扰”,其中一个启动子占据优势并引起第二个启动子的沉默。此外,不同的启动子在不同的细胞环境中不同地起作用,这使得难以实现每个转基因的相对表达的一致的“调整(tuning)”。

[0005] 另一种方法是使用内部核糖体进入序列(IRES)。这里,生成单个转录物。该转录物中的IRES序列置于两个转基因的开放阅读框之间,并模拟mRNA帽结构。因此,核糖体在5'帽或者IRES处起始翻译,导致两个单独的蛋白的表达。这种方法的关键限制是不能控制相对表达。3'转录物通常表达少于5'转录物,但表达比例难以预测和调整。

[0006] 在口蹄疫病毒(FMDV) 2A肽在允许FMDV(和相关病毒)从单个开放阅读框(ORF)表达多个蛋白质中的作用表征(Donnelly et al; J.Gen.Virol.; 82,1027-1041 (2001))之后提供了进一步的方法。在ORF的翻译后,2A肽(和同源物)立即以非常高的效率切割,使得能够从单个ORF表达多个肽。使用2A肽在同一ORF中的不同肽之间切割的问题是表达被限制为1:1的比例。

[0007] 因此,需要与上述缺点不相关的用于从单个载体表达多于一个蛋白的替代方法。

### [0008] 发明概述

[0009] 本发明是基于确定当跨膜蛋白与第二蛋白共表达为翻译后随后被切割以分开这两种蛋白的多聚蛋白时,在跨膜蛋白中掺入细胞内保留信号允许相对于第二蛋白调节跨膜蛋白的细胞表面表达,通过降低其运输到细胞表面和/或降低其在细胞表面的半衰期进行。这不需要限于转基因对,而是可以用于允许控制最初翻译为多聚蛋白的多种蛋白的相对表达。

[0010] 如本文所使用的,“多聚蛋白”是指下述多肽序列,所述多肽序列从单个核酸构建体以单个实体翻译,但包含随后分离并作为单独的实体(例如单独的蛋白)发挥功能的多肽序列。

[0011] 因此,本发明的第一个方面提供包含以下结构的核酸构建体:

[0012] A-X-B

[0013] 其中

[0014] A和B是编码第一和第二感兴趣的蛋白质(POI)的核酸序列;和

[0015] X是编码切割位点的核酸序列,

[0016] 其中所述第一或者第二POI是包含细胞内保留信号的跨膜蛋白。

[0017] 细胞内保留信号对于跨膜蛋白是异源的,即其不存在于跨膜蛋白(或其部分)的“野生型”形式中,并通过重组方法引入。细胞内保留信号可以是合成的或自另一种生物体或另一种蛋白可衍生。

[0018] 跨膜蛋白的胞内域(endodomain)可以包含细胞内保留信号。

[0019] 细胞内保留信号可以将跨膜蛋白引导为远离分泌途径和/或引导至膜结合的细胞内区室,如溶酶体,内体或高尔基体区室。

[0020] 细胞内保留信号可以是例如基于酪氨酸的分选信号,基于二亮氨酸的分选信号,酸性簇信号,溶酶体回避信号,NPF $X'$ (1,2)D型信号,KDEL,KK $X'$  $X'$ 或KK $X'$ K $X'$ X信号(其中 $X'$ 是任意氨基酸)。

[0021] 细胞内保留信号可以包含选自下组的序列:NP $X'$ Y,Y $X'$  $X'$ Z,[DE] $X'$  $X'$  $X'$ L[LI],DX' $X'$ LL,DP[FW],FX'D $X'$ F,NPF,LZ $X'$ Z[DE],LLDLL,PWDLW,KDEL,HDEL,KK $X'$  $X'$ 或KK $X'$ K $X'$  $X'$ ;

[0022] 其中 $X'$ 是任意氨基酸且 $Z'$ 是具有大体积的疏水性侧链的氨基酸。

[0023] 细胞内保留信号可以包含表1至5中示出的任意序列。

[0024] 细胞内保留信号可以包含酪氨酸酶相关蛋白(TYRP)-1细胞内保留信号。细胞内保留信号可以包含TYRP-1细胞内结构域。细胞内保留信号可以包含序列NQPLLTD(SEQ ID No.35)。

[0025] 细胞内保留信号可以包含腺病毒E3/19K细胞内保留信号。细胞内保留信号可以包含E3/19K胞质结构域。细胞内保留信号可以包含序列KYKSRRSFIDEKKMP(SEQ ID No.36);或DEKKMP(SEQ ID No.37)。

[0026] 细胞内保留信号可以靠近或远离跨膜蛋白的跨膜结构域。

[0027] X可以是编码自切割肽,弗林蛋白酶切割位点或烟草蚀刻病毒切割位点的核酸序列。

[0028] X可以是编码来自口疮病毒或心脏病毒的2A自切割肽或2A样肽的核酸序列。

[0029] 跨膜蛋白可以是任何转基因表达的跨膜蛋白。

[0030] 跨膜蛋白可以选自以下列表:兴奋性受体如41BB,OX40,CD27,CD28和相关分子;或抑制性受体如PD1,CTLA4,LAIR1,CD22和相关分子;或细胞因子受体分子如IL1R,IL2R,IL7R,IL15R和相关分子;或归巢分子如N-CAM,V-CAM,L1-CAM,LFA-1,CDH1-3,选择蛋白或整合蛋白;

[0031] 跨膜蛋白可以是合成的蛋白,例如自杀基因或标志物基因。

[0032] 跨膜蛋白可以是或包含T细胞受体的 $\alpha$ 和/或 $\beta$ 链。

[0033] 跨膜蛋白可以是嵌合抗原受体(CAR)。

[0034] 第一和第二POI中的任一个或两者可以是跨膜蛋白;且跨膜蛋白中的任一个或两者可以包含如本文所定义的细胞内保留信号。

[0035] 第一和第二POI中的任一个或两者可以是单通跨膜蛋白,例如I型跨膜蛋白。

[0036] 与从相同的核酸构建体表达的不包含细胞内保留信号的跨膜蛋白相比,在细胞表面表达的包含细胞内保留信号的跨膜蛋白的量可以例如小于90%,70%,50%或30%。

[0037] 本发明还提供编码多于两个感兴趣的蛋白的核酸构建体。例如,在一个实施方案中,本发明提供包含以下结构的核酸构建体:

[0038] A-X-B-Y-C

[0039] 其中

[0040] A,B和C是编码第一,第二和第三感兴趣的多肽(POI)的核酸序列;和

[0041] X和Y是可以相同或不同的核酸序列,其各自编码切割位点,

[0042] 其中所述POI的至少两个是包含细胞内保留信号的跨膜蛋白。

[0043] 在该实施方案中,所述至少两个POI可以:

[0044] (a) 包含不同的细胞内保留信号;和/或

[0045] (b) 具有位于所述POI中的不同位置的细胞内保留信号,

[0046] 使得当核酸在细胞中表达时,在所述细胞表面存在所述至少两个POI的差异相对表达。

[0047] 在第二个方面,本发明提供包含根据本发明的第一个方面的核酸构建体的载体。

[0048] 载体可以是逆转录病毒载体或慢病毒载体或转座子。

[0049] 在第三个方面,本发明提供包含根据本发明的第一个方面的核酸构建体或根据本发明的第二个方面的载体的细胞。

[0050] 本发明还涉及用于调节从单个核酸构建体与第二蛋白一起以多聚蛋白表达的跨膜蛋白的相对细胞表面表达的方法,其通过在编码所述跨膜蛋白的核酸序列中包括细胞内保留信号进行。

[0051] 在跨膜蛋白中包括细胞内保留信号降低在细胞表面表达的跨膜蛋白的量。因此,可以调节自包括第二多肽的多聚蛋白衍生的跨膜蛋白的相对表达水平。在跨膜蛋白仅在细胞表面有活性(或主要在细胞表面上有活性)时,降低蛋白的相对细胞表面表达也降低其相对活性。

[0052] 可以延伸本发明以调节由切割位点分开的作为联接(concatenated)多肽表达的三种或更多种蛋白质的相对表达,并通过不同活性的保留信号决定相对表面表达。

[0053] 附图简述

[0054] 图1:用于从同一载体表达不同的蛋白的方法

[0055] (a) 同一盒内的两个不同的启动子导致两个不同的转录物,其各自产生单独的蛋白。(b) 使用内部核糖体进入序列(IRES)导致单个转录物,其翻译成两个单独的蛋白。(c) FMDV 2A肽的使用导致单个转录物和单个多聚蛋白,其快速切割成两个单独的蛋白。

[0056] 图2:TYRP1胞内域能够指导具有复合胞内域的跨膜蛋白的保留

[0057] Tyrp1是I型跨膜蛋白,长537aa。将蛋白保留在细胞内区室中的二亮氨酸基序在胞质结构域上表示为黑色矩形。(A) Tyrp1(wt)。野生型Tyrp1由肽信号,腔性结构域(luminal domain),跨膜结构域和胞质结构域组成。胞质结构域含有二亮氨酸保留信号。(B) Tyrp1(wt)-SG接头-eGFP。该构建体含有通过丝氨酸-甘氨酸-甘氨酸-甘氨酸-丝氨酸接头与eGFP简单融合的野生型Tyrp1。Tyrp1-L-eGFP代表细胞质近端Tyrp1。(C) Tyrp1腔性(LM)-跨膜(TM)-SG接头-eGFP-Tyrp1胞质(CP)。该构建体构成胞质-远端Tyrp1,因为SG接头-eGFP插入跨膜和胞质结构域之间。D: Tyrp1腔性(LM)-跨膜(TM)-SG接头-eGFP。该构建体充当阳性对照,因为含有保留信号的胞质结构域已经排除。所有构建体与IRES.CD34共表达。显示了转染的SupT1细胞的染色,其中细胞内和表面染色分别显示于底部左/右。

[0058] 图3:原代细胞中TYRP1保留信号的功能性

[0059] 使用FMD-2A样肽生成了共表达抗CD19和抗CD33 CAR的构建体。还产生了该构建体

的两个变体：在第一个变体中，将来自TYRP1的二亮氨酸基序插入抗CD19 CAR胞内域，就在TM结构域近端；在第二个变体中，将相同的TYRP1二亮氨酸基序附接到抗CD19 CAR胞内域的羧基末端。从血液中分离PBMC，并使用PHA和IL-2刺激。两天后，使用含有不同CD19:CD33 CAR构建体的逆转录病毒在纤维连接蛋白包被的板上转导细胞。在第5天，通过使用重组CD19-Fc和CD33-Fc融合物的流式细胞术评估了由构建体翻译的两种CAR的表达水平。A. 显示构建为允许共表达的合成基因的图；B. 显示由三种构建体产生的随后的蛋白对的图；C. 显示通过流式细胞术的两种受体的表达。在原始的构建体中，相等地表达这两个CAR。在抗CD19 CAR的胞内域中远端掺入二亮氨酸基序的情况下，CD33 CAR表达保持恒定，但是CD19表达下降到中等水平。在抗CD19 CAR的胞内域中近端掺入二亮氨酸基序的情况下，CD33 CAR表达保持恒定，但是CD19表达降低到低水平。

[0060] 图4：来自E3/19K胞质尾的保留信号

[0061] 使用FMD-2A样肽生成了共表达抗CD19和抗CD33 CAR的构建体。还产生了该构建体的两个变体：在第一个变体中，将来自E3/19K的最后6个氨基酸(DEKKMP) (发现对于其高尔基体/ER保留能力至关重要的) 附接到抗CD33 CAR胞内域的羧基末端；在第二个变体中，将腺病毒E3/19K蛋白的整个胞质尾部附接到抗CD33 CAR胞内域的羧基末端。

[0062] 图5：E3/19K保持信号的功能性

[0063] 将图4中显示的构建体转染到293T细胞中，并通过使用重组CD19-Fc和CD33-Fc融合物的流式细胞术评估由构建体翻译的两种CAR的表达水平。当全长腺病毒E3/19K蛋白或DEKKMP基序置于抗CD33受体上时，观察到明显的保留。抗CD19受体表达水平不受影响。

[0064] 发明详述

[0065] 本发明提供包含以下结构的核酸构建体：

[0066] A-X-B

[0067] 其中

[0068] A和B是编码第一和第二感兴趣的多肽(POI)的核酸序列；和

[0069] X是编码切割位点的核酸序列，

[0070] 其中所述第一或者第二POI是包含细胞内保留信号的跨膜蛋白。

[0071] 跨膜蛋白

[0072] 本发明使得能够调节跨膜表面蛋白的相对表达。跨膜表面蛋白是在不存在细胞内保留信号的情况下在细胞表面表达的蛋白。当在细胞表面表达时，跨膜蛋白的至少一个结构域是胞质外的(即在细胞的外部)。

[0073] 跨膜蛋白可以是单次跨膜蛋白，即其可以包含单个跨膜结构域，或者其可以包含多个跨膜结构域。

[0074] 跨膜蛋白可以通过拓扑学分类，即参照N-和C-末端结构域的位置。I型，II型和III型跨膜蛋白是单通分子，而IV型跨膜蛋白是多通分子。I型跨膜蛋白以停止转移锚定序列(stop-transfer anchor sequence)锚定到脂质膜上，并且在合成过程中其N末端结构域靶向至ER腔(以及细胞外空间，当成熟形式位于质膜上时)。II型和III型以信号-锚定序列锚定，其中II型以其C末端结构域靶向至ER腔，而III型使其N末端结构域靶向至ER腔。IV型细分为IV-A，其中其N末端结构域靶向胞质溶胶，和IV-B，其中N端结构域靶向腔。

[0075] 本发明的跨膜蛋白可以是I-IV型中的任一种。

[0076] 跨膜结构域可以是在膜中热力学稳定的任何蛋白结构。这通常是包含几个疏水残基的 $\alpha$ 螺旋。任何跨膜蛋白的跨膜结构域可用于提供跨膜部分。蛋白的跨膜结构域的存在和跨度可以由本领域技术人员使用TMHMM算法 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) 确定。此外, 鉴于蛋白的跨膜结构域是相对简单的结构, 即预测形成足够长度以跨越膜的疏水性 $\alpha$ 螺旋的多肽序列, 也可以使用人工设计的TM结构域 (US7052906 B1描述了合成的跨膜组分)。

[0077] 跨膜结构域可以衍生自CD28, 其给出良好的稳定性。

[0078] I型跨膜蛋白的结构和加工是本领域公知的。此类蛋白通常包含细胞外结构域, 跨膜结构域和细胞内结构域, 并且是具有穿过细胞膜的单个 $\alpha$ -螺旋的单通分子。

[0079] I型跨膜蛋白通常具有信号肽, 其被内质网 (ER) 快速识别, 因此翻译中的蛋白被快速再导向至ER。疏水螺旋将蛋白锁定然后锚定在ER的膜中。

[0080] 如上文所述, I型跨膜蛋白通过停止转移锚定序列锚定到脂质膜上。停止转移序列停止多肽的进一步移位并充当跨膜锚。

[0081] 如本文所使用的, 术语I型跨膜蛋白涵盖包含I型跨膜结构域和停止转移锚定序列的任何蛋白, 并且在不存在外源性细胞内保留信号的情况下靶向用于在细胞表面上表达。

[0082] 适用于本发明的各种I型跨膜蛋白是本领域已知的。此类蛋白包括但不限于抑制性受体, 刺激受体, 细胞因子受体和G蛋白。

[0083] 跨膜蛋白可以是T细胞受体 $\alpha$ 或 $\beta$ 链。

[0084] 跨膜蛋白可以是嵌合抗原受体 (CAR)。

[0085] CAR是将抗原结合结构域移植到T细胞的效应功能的蛋白。它们通常的形式是具有全部连接到传输T细胞存活和激活信号的复合胞内域的抗原识别氨基末端、间隔区、跨膜结构域的I型跨膜结构域蛋白的形式。

[0086] 抗原结合结构域可以衍生自抗体或抗体模拟物, 或者其可以是特异性结合抗原的另一实体, 例如配体。

[0087] 这些分子的最常见形式是通过间隔区和跨膜结构域与信号传导胞内域融合的自识别靶抗原的单克隆抗体衍生的单链可变片段 (scFv) 的融合物。此类分子响应于scFv对其靶标的识别而导致T细胞的活化。当T细胞表达此类CAR时, 它们识别并杀死表达靶抗原的靶细胞。已经开发了几种针对肿瘤相关抗原的CAR, 并且使用此类表达CAR的T细胞的过继转移方法目前正在临床试验中用于治疗各种癌症。

[0088] 信号传导胞内域也可能存在于单独的分子上。与本发明有关的术语“CAR”还涵盖包含与跨膜结构域连接的抗原结合结构域的分子。此类CAR可以能够与细胞内信号传导结构域相互作用以便刺激T细胞活化。

[0089] 在本发明中, 核酸序列A或B可以是编码包含细胞内保留信号的跨膜蛋白的核酸序列。

[0090] 大多数感兴趣的跨膜蛋白仅当在细胞表面时是有活性的, 或者主要当在细胞表面时是有活性的。因此, 使一定比例的蛋白在胞内保留降低蛋白质在细胞表面的相对表达, 因此降低蛋白质的相对活性。

[0091] 信号序列

[0092] 跨膜蛋白还可以包含信号序列, 使得当跨膜蛋白在细胞内表达时, 新生的蛋白质

被引导至内质网(ER)。

[0093] 信号肽的核心可以含有具有形成单个 $\alpha$ -螺旋的倾向的疏水性氨基酸的长段。信号肽可以以氨基酸的带正电荷的短区段开始,所述短区段有助于在移位期间执行多肽的适当拓扑。在信号肽的末端,通常有被信号肽酶识别和切割的一段氨基酸。信号肽酶可以在移位期间或完成之后切割,以产生游离信号肽和成熟蛋白。然后,游离的信号肽被特异性蛋白酶消化。

[0094] 信号肽可以在分子的氨基末端。

[0095] 切割位点

[0096] 本核酸构建体包含编码切割位点的序列,其位于编码第一和第二多肽的核酸序列之间,使得第一和第二多肽可以作为单独的实体表达。

[0097] 切割位点可以是能够使包含第一和第二POI的多肽分离的任何序列。

[0098] 术语“切割”在本文中为了方便而使用,但是切割位点可以通过除经典切割以外的机制使第一和第二POI分离成个别的实体。例如,对于口蹄疫病毒(FMDV) 2A自切割肽(见下文),已经提出了各种模型来解释“切割”活性:通过宿主细胞蛋白酶的蛋白水解、自体蛋白水解或翻译效应(Donnelly et al (2001) J.Gen.Virol.82:1027-1041)。此类“切割”的确切机制对于本发明的目的而言并不重要,只要切割位点在位于编码第一和第二多肽的核酸序列之间时导致第一和第二多肽表达为单独的实体。

[0099] 切割位点可以是弗林蛋白酶切割位点。

[0100] 弗林蛋白酶是属于枯草杆菌蛋白酶样前蛋白转化酶家族的酶。该家族的成员是将潜伏的前体蛋白加工成其生物学活性产物的前蛋白转化酶。弗林蛋白酶是钙依赖性丝氨酸内切蛋白酶,其可以在它们配对的碱性氨基酸加工位点处有效切割前体蛋白。弗林蛋白酶底物的实例包括甲状旁腺激素、转化生长因子 $\beta$ 1前体、白蛋白原、 $\beta$ -分泌酶原、膜1型基质金属蛋白酶、神经生长因子原的 $\beta$ 亚基和Willebrand因子。弗林蛋白酶刚好在碱性氨基酸靶序列(典型地为Arg-X-(Arg/Lys)-Arg')的下游切割蛋白,并且在高尔基体中富集。

[0101] 切割位点可以是烟草蚀刻病毒(TEV)切割位点。

[0102] TEV蛋白酶是高度序列特异性的半胱氨酸蛋白酶,其是糜蛋白酶样蛋白酶。它对其靶切割位点是非常特异性的,因此经常用于既在体外又在体内对融合蛋白的受控切割。共有TEV切割位点为ENLYFQ\S(其中“\”表示切割的肽键)。哺乳动物细胞(如人类细胞)不表达TEV蛋白酶。因此,在本核酸构建体包含TEV切割位点并且在哺乳动物细胞中表达的实施方案中,也必须在哺乳动物细胞中表达外源TEV蛋白酶。

[0103] 切割位点可以编码自切割肽。

[0104] “自切割肽”是指下述肽,该肽发挥功能而使得当包含第一和第二POI和自切割肽的多肽产生时,其立即被“切割”或分离成独立的且离散的第一和第二多肽,而不需要任何外部切割活性。

[0105] 自切割肽可以是来自口疮病毒或心脏病毒的2A自切割肽。口疮病毒和心脏病毒的主要2A/2B切割由在其自身C末端的2A“切割”介导。在口疮病毒例如口蹄疫病毒(FMDV)和马鼻肺炎A型病毒中,2A区域是约18个氨基酸的短区,其与蛋白2B的N末端残基(保守的脯氨酸残基)一起代表了能够在其自身C末端介导“切割”的自主元件。

[0106] 较长的心脏病毒蛋白的C端19个氨基酸与2B的N末端脯氨酸一起以大致等于口疮

病毒FMDV 2a序列的效率介导“切割”。心脏病毒包括脑心肌炎病毒 (EMCV) 和泰勒氏 (Theiler's) 鼠脑炎病毒 (TMEV)。

[0107] EMCV和FMDV 2A的突变分析揭示了基序DxExNPGP密切参与“切割”活性 (Donnelly et al (2001), 如上文)。

[0108] 本发明的切割位点可以包含以下氨基酸序列:

[0109] Dx<sub>1</sub>Ex<sub>2</sub>NPGP, 其中x<sub>1</sub>和x<sub>2</sub>是任意氨基酸。X<sub>1</sub>可以选自下组: I, V, M和S。X<sub>2</sub>可以选自下组: T, M, S, L, E, Q和F。

[0110] 例如, 切割位点可以包含表1所示的氨基酸序列之一。

[0111] 表1

| 基序                       | 存在于:                                                                      |
|--------------------------|---------------------------------------------------------------------------|
| DIETNPGP (SEQ ID No. 1)  | 小核糖核酸病毒 (picornaviruses) EMCB, EMCD, EMCPV21                              |
| DVETNPGP (SEQ ID No. 2)  | 小核糖核酸病毒 MENGO 和TMEBEAN; 昆虫病毒 DCV, ABPV                                    |
| DVEMNPGP (SEQ ID No. 3)  | 小核糖核酸病毒 TMEGD7 和TMEBEAN                                                   |
| DVESNPGP (SEQ ID No. 4)  | 小核糖核酸病毒 FMDA10, FMDA12, FMDC1, FMD01K, FMDSAT3, FMDVSAT2, ERAV; 昆虫病毒 CrPV |
| DMESNPGP (SEQ ID No. 5)  | 小核糖核酸病毒 FMDV01G                                                           |
| DVELNPGP (SEQ ID No. 6)  | 小核糖核酸病毒ERBV; 猪轮状病毒                                                        |
| DVEENPGP (SEQ ID No. 7)  | 猪轮状病毒PTV-1; 昆虫病毒TaV; 锥虫TSR1                                               |
| DIELNPGP (SEQ ID No. 8)  | 牛轮状病毒, 人轮状病毒                                                              |
| DIEQNPGP (SEQ ID No. 9)  | 锥虫AP内切核酸酶                                                                 |
| DSEFNPGP (SEQ ID No. 10) | 细菌序列海栖热袍菌( <i>T. maritima</i> )                                           |

[0114] 基于2A序列的切割位点可以是例如长度为15-22个氨基酸。该序列可以包含2A蛋

白的C末端,随后是脯氨酸残基(其对应于2B的N末端脯氨酸)。

[0115] 突变研究还显示了,除天然存在的2A序列之外,一些变体也是活性的。切割位点可以对应于来自天然存在的2A多肽的变体序列,具有一个,两个或三个氨基酸取代,其保留了诱导将多聚蛋白序列“切割”成两个或更多个单独的蛋白的能力。

[0116] 切割序列可以选自以下各项,其都已经显示在一定程度上是有活性的 (Donnelly et al (2001), 如上文):

[0117] LLNFDLLKLAGDVESNPGP (SEQ ID No.11)

[0118] LLNFDLLKLAGDVQSNPGP (SEQ ID No.12)

[0119] LLNFDLLKLAGDVEINPGP (SEQ ID No.13)

[0120] LLNFDLLKLAGDVEFNPGP (SEQ ID No.14)

[0121] LLNFDLLKLAGDVESHPPGP (SEQ ID No.15)

[0122] LLNFDLLKLAGDVESEPPGP (SEQ ID No.16)

[0123] LLNFDLLKLAGDVESQPGP (SEQ ID No.17)

[0124] LLNFDLLKLAGDVESNPGG (SEQ ID No.18)

[0125] 基于DxExNPGP基序的序列,已经在除口疮病毒或心脏病毒外的小核糖核酸病毒,“小核糖核酸病毒样”昆虫病毒,C型轮状病毒,锥虫属物种中的重复序列和细菌序列中发现了“2A样”序列 (Donnelly et al (2001), 如上文)。切割位点可以包含这些2A样序列之一,例如:

[0126] YHADYYKQRLIHDVEMNPGP (SEQ ID No.19)

[0127] HYAGYFADLLIHDITNPGP (SEQ ID No.20)

[0128] QCTNYALLKLAGDVESNPGP (SEQ ID No.21)

[0129] ATNFSLLKQAGDVEENPGP (SEQ ID No.22)

[0130] AARQMLLLLSGDVETNPGP (SEQ ID No.23)

[0131] RAEGRGSLTTCGDVEENPGP (SEQ ID No.24)

[0132] TRAEIEDELIRAGIESNPGP (SEQ ID No.25)

[0133] TRAEIEDELIRADIESNPGP (SEQ ID No.26)

[0134] AKFQIDKILISGDVELNPGP (SEQ ID No.27)

[0135] SSIIRTKMLVSGDVEENPGP (SEQ ID No.28)

[0136] CDAQRQKLLLSGDIEQNPGP (SEQ ID No.29)

[0137] YPIDFGGFLVKADSEFNPGP (SEQ ID No.30)

[0138] 切割位点可以包含如SEQ ID No.24所示的2A样序列 (RAEGRGSLTTCGDVEENPGP)。

[0139] 已经表明,包括5至39个氨基酸的N末端“延伸”可以增加活性 (Donnelly et al (2001), 如上文)。特别地,切割序列可以包含以下序列之一或其变体,所述变体具有例如多达5个氨基酸变化,且保留了切割位点活性:

[0140] VTELLYRMKRAETYCPRLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No.31)

[0141] LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No.32)

[0142] EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No.33)

[0143] APVKQTLNFDLLKLAGDVESNPGP (SEQ ID No.34)

[0144] 细胞内保留信号

[0145] 本发明的核酸构建体包含编码跨膜蛋白的序列,所述跨膜蛋白包含细胞内保留信号。

[0146] 蛋白靶向或蛋白分选是将蛋白运送到细胞中或其外部的适当目的地的生物学机制。可以将蛋白质靶向细胞器的内部空间,不同的细胞内膜,质膜或经由分泌到细胞外部。该递送过程基于蛋白本身中包含的序列信息进行。

[0147] 在真核细胞的粗面内质网(ER)中合成的蛋白使用胞吐途径(exocytic pathway)运输到其最终目的地。缺乏特殊分选信号的蛋白经由高尔基体和反面高尔基体网络(TGN)从ER矢量运输(vectorially transport)到质膜。其它蛋白具有用于整合到胞吐途径的特定细胞器,例如内体和溶酶体中的靶向信号。

[0148] 溶酶体是酸性细胞器,其中内源性和内化的大分子被腔内水解酶(luminal hydrolases)降解。内源性大分子通过在TGN中分选到达溶酶体,它们从TGN转运到内体,然后是溶酶体。

[0149] 本发明可以利用细胞使用的可以将蛋白分选到正确的细胞内位置的靶向信号。该信号可以大致分为以下类型:

[0150] i) 内吞信号

[0151] ii) 高尔基体保留信号

[0152] iii) TGN再循环信号

[0153] iv) ER保留信号

[0154] v) 溶酶体分选信号

[0155] “细胞内保留信号”是指下述氨基酸序列,所述氨基酸序列将包含它的蛋白质引导到细胞区室而非细胞表面膜或细胞的外部。

[0156] 与不包含细胞内保留信号的等同对照跨膜蛋白相比,细胞内保留信号导致在细胞表面上表达的跨膜蛋白的量减少。

[0157] 换句话说,在细胞表面上表达的包含细胞内保留信号的翻译的跨膜蛋白的比例小于不包含细胞内保留信号的等同的、翻译的对照跨膜蛋白的比例。

[0158] 例如,在细胞表面上表达的包含细胞内保留信号的跨膜蛋白的量可以是不包含细胞内保留信号的等同对照跨膜蛋白的量的小于75%,小于50%,小于25%或小于10%。

[0159] 表达随后由蛋白酶切割的多聚蛋白的构建体通常受到来自多聚蛋白的肽的表达限于1:1比例的事实的限制。然而,在本发明中,在跨膜蛋白中包含细胞内保留信号意味着可以调节其在细胞表面上的表达(例如与不包含细胞内保留信号的等同对照跨膜蛋白相比减少)。因此,与在多聚蛋白中表达的第二蛋白的表达相比,细胞表面上表达的包含细胞内保留信号的跨膜蛋白的比例可以为例如约1:1.5,1:1.5-1:2,1:2-1:3,1:3-1:4,1:4-1:5或大于1:5。

[0160] 可以使用本领域已知的方法,例如流式细胞术或荧光显微镜术来测定在细胞表面上表达的跨膜蛋白的量。

[0161] 在从ER移位期间,细胞内保留信号可以引导跨膜蛋白远离分泌途径。

[0162] 细胞内保留信号可以将跨膜蛋白引导到细胞内区室或复合物。细胞内保留信号可以将跨膜蛋白引导到膜结合的细胞内区室。

[0163] 例如,细胞内保留信号可以将蛋白质引导至溶酶体,内体或高尔基体区室(反面高尔基体网络,“TGN”)。

[0164] 在正常细胞内,从生物发生或内吞途径产生的蛋白按照顺序的分选决定集合分选到适当的细胞内区室中。在质膜处,蛋白可以保留在细胞表面,也可以内化到内体。在TGN,选择是去往质膜或被转移到内体。在内体中,蛋白可以再循环到质膜或去往溶酶体。这些决定由蛋白本身上的分选信号决定。

[0165] 溶酶体是含有分解废物材料和细胞碎片的酸性水解酶的细胞器。溶酶体周围的膜允许消化酶在它们所需的pH下工作。溶酶体与自噬泡(吞噬体)融合,并将其酶分配到自噬泡中,消化其内容物。

[0166] 内体是真核细胞内的膜边界区室。它是从质膜到溶酶体的内吞膜运输途径的区室,并且在到达降解性溶酶体之前提供待分选物质的环境。内体可以根据内吞的材料到达它们花费的时间而分类为早期内体,晚期内体或再循环内体。本发明中使用的细胞内保留信号可将蛋白引导至晚期内体区室。

[0167] 高尔基体是细胞内膜系统的一部分,高尔基体将细胞内的蛋白质包裹起来,之后将它们送至其目的地;其在加工蛋白以进行分泌中特别重要。

[0168] 调查已知蛋白中存在的分选信号,以及改变其序列和/或在分子内的位置的效果的研究产生了相当多的知识(Bonifacino and Traub (2003) *Ann.Rev.Biochem.* 72:395-447;Braulke and Bonifacino (2009) *Biochimica and Biophysica Acta* 1793:605-614;Griffith (2001) *Current Biology* 11:R226-R228;Mellman and Nelson (2008) *Nat Rev Mol Cell Biol.* 9:833-845;Dell’Angelica and Payne (2001) *Cell* 106:395-398;Schafer et al (1995) *EMBO J.* 14:2424-2435;Trejo (2005) *Mol.Pharmacol.* 67:1388-1390)。许多研究已经表明,为了改变感兴趣的蛋白的细胞内位置,可以将一个或多个分选信号插入感兴趣的蛋白质(Pelham (2000) *Meth.Enzymol.* 327:279-283)。

[0169] 因此,完全有可能选择具有所需定位性质的分选信号并将其包括在感兴趣的蛋白内,以引导该蛋白的细胞内位置。结合本申请,因此可以选择在质膜上具有所需的表达降低量的分选信号。

[0170] 分选信号在感兴趣的新生蛋白中的最佳位置可以取决于跨膜蛋白的类型(即I-IV型)以及C末端是否在膜的腔体或细胞质侧(Goder and Spiess (2001) *FEBS Lett* 504:87-93)。这可以通过考虑分选蛋白在其天然蛋白中的位置来容易地确定。

[0171] 内吞信号的实例包括来自转铁蛋白受体和无唾液酸糖蛋白受体的那些。

[0172] 引起TGN-内体再循环的信号的实例包括来自蛋白如CI-和CD-MPRs,sortilin,LDL-受体相关蛋白LRP3和LRP10和 $\beta$ -分泌酶,GGA1-3,LIMP-II,NCP1,粘蛋白-1,sialin, GLUT8和不变链的那些。

[0173] TGN保留信号的实例包括来自定位于TGN的以下蛋白的那些:激素原加工酶弗林蛋白酶,PC7,CPD和PAM;疱疹病毒3的糖蛋白E和TGN38。

[0174] ER保留信号的实例包括C末端信号,例如KDEL, KKXX或KKKXX和钾通道的RXR(R)基序。已知的ER蛋白包括腺病毒E19蛋白和ERGIC53。

[0175] 溶酶体分选信号的实例包括在溶酶体膜蛋白中发现的那些,例如LAMP-1和LAMP-2,CD63,CD68,endolyn,DC-LAMP,cystinosin,糖磷酸盐交换子2(sugar phosphate

exchanger 2) 和酸性磷酸酶。

[0176] 细胞内保留信号可以来自腺病毒E19蛋白。细胞内保留信号可以来自蛋白E3/19K, 其也称为E3gp 19kDa;E19或GP19K。细胞内保留信号可以包含E3/19K的完全胞质尾部, 其如SEQ ID No.36所示;或该尾部的最后6个氨基酸, 其如SEQ ID No.37所示。本发明人已经表明, 该最后6个氨基酸对于保留特别重要(实施例3和图5)。

[0177] SEQ ID No.36:KYKSRRSFIDEKKMP

[0178] SEQ ID No.37:DEKKMP

[0179] 可调节性

[0180] 可以通过各种方法使用本发明的方法精确调节一个或多个蛋白的相对表达, 例如:

[0181] a) 改变蛋白分子中细胞内保留信号的位置;和/或

[0182] b) 选择特定的细胞内保留信号。

[0183] 选项a) 在下文更详细的讨论。

[0184] 关于选项b), 可以从大量天然存在的蛋白获得一系列细胞内保留信号, 这些蛋白被分选到真核细胞内的不同细胞位置。还可能使用“合成的”细胞内保留信号, 其包含在天然存在的蛋白中发现的一个或多个基序(参见下一小节)并且具有相似的分选信号功能。

[0185] 取决于分选信号将相关蛋白发送到的细胞内位置, 可获得信号强度的级联。广义上说, 分选信号引导的位置越为“细胞内”, 该信号在降低蛋白的相对表达方面就越“强”。

[0186] 当分选信号将蛋白引导至溶酶体区室时, 蛋白被细胞内化和降解, 导致到细胞表面的相对较少逃逸。一旦蛋白进入溶酶体, 其就会被降解和从系统中丢失。因此, 在降低细胞表面的相对表达方面, 溶酶体分选信号例如LAMP1是“最强的”。

[0187] 当分选信号引导蛋白保留在ER中时, 只有非常小比例的蛋白到达细胞表面。因此, 在降低细胞表面的相对表达方面, ER保留或再循环信号, 如ER-GIC-53和KKFF信号是接下来最强的。

[0188] 当分选信号将蛋白引导到内体, 高尔基体或TGN区室时, 蛋白然后可能在一定程度上在TGN, 内体区室和质膜之间循环。这些信号提供了更有限水平的表达降低, 因为相当大比例的蛋白仍将达到质膜。

[0189] 一般来说, 用已知的分选信号看到的表达的减少可以总结如下:

[0190] 溶酶体分选信号 > ER保留/再循环信号 > TGN保留/再循环信号 > 内吞信号

[0191] 当考虑多种蛋白(每种蛋白具有它们自己的相对表达)的表达时, 使用不同分选信号和/或分选信号在蛋白内的不同位置的可调节性特别有用。例如, 考虑具有以下结构的核酸构建体:

[0192] A-X-B-Y-C

[0193] 其中

[0194] A, B和C是编码多肽的核酸序列;和

[0195] X和Y是编码切割位点的核酸序列。

[0196] 核酸构建体将编码三种蛋白A, B和C, 其中任一种或全部可以是跨膜蛋白。例如, B和C可以是包含细胞内保留信号的跨膜蛋白。如果希望A, B和C表达使得相对水平为A>B>C, 则核酸序列A可以不具有细胞内保留信号, 核酸序列B可具有细胞内保留信号, 其导致小部

分蛋白B保留在细胞中(即不在细胞表面表达),并且核酸序列C可具有细胞内保留信号,其导致大部分蛋白C保留在细胞中。

[0197] 如下所述,导致不同量的细胞表面表达的细胞内保留的差异量可以通过以下方式实现:

[0198] (a) 对蛋白使用不同的细胞内保留信号;和/或

[0199] (b) 使细胞内保留信号位于蛋白中的不同位置。

[0200] 信号类型

[0201] 包括细胞内保留信号并被引导至细胞内区室的许多蛋白是本领域已知的。

[0202] 细胞内保留信号可以是来自存在于溶酶体,内体或高尔基体区室中的蛋白的保留信号。

[0203] 细胞内保留信号是本领域公知的(参见例如Bonifacino&Traub; Annu.Rev.Biochem.;2003;72;395-447)。

[0204] 细胞内保留信号可以是基于酪氨酸的分选信号,基于二亮氨酸的分选信号,酸性聚簇信号,溶酶体回避信号,NPF $X'$ (1,2) D型信号,KDEL,KK $X'$  $X'$ 或K $X'$ K $X'$  $X'$ 信号(其中 $X'$ 是任意氨基酸)。

[0205] 基于酪氨酸的分选信号介导跨膜蛋白从质膜的快速内化并将蛋白靶向至溶酶体(Bonifacino&Traub;如上文)。由NPX $'Y$ 和YX $'X'Z'$ 共有基序(其中 $Z'$ 是具有大体积疏水侧链的氨基酸)表示两种类型的基于酪氨酸的分选信号。

[0206] 已经显示NPX $'Y$ 信号介导I型跨膜蛋白的快速内化,它们发生在诸如LDL受体,整联蛋白 $\beta$ 和 $\beta$ -淀粉样蛋白前体蛋白家族的成员等的家族中。

[0207] 在表2中提供NPX $'Y$ 信号的实例。

[0208] 表2-NPX $'Y$ 信号

| 蛋白          | 物种 | 序列                              |
|-------------|----|---------------------------------|
| LDL受体       | 人类 | Tm-10-INF <b>DNPVY</b> QKTT-29  |
| LRP1(1)     | 人类 | Tm-21-VEIG <b>NPTY</b> KMYE-64  |
| LRP1(2)     | 人类 | Tm-55-TNFT <b>NPVY</b> ATLY-33  |
| LRP1        | 果蝇 | Tm-43-GNF <b>ANPVY</b> ESMY-38  |
| LRP1(1)     | 线虫 | Tm-54-TTFT <b>NPVY</b> ELED-91  |
| LRP1(2)     | 线虫 | Tm-140-LRVD <b>NPLY</b> DPDS-4  |
| Megalin(1)  | 人类 | Tm-70-IIF <b>ENPMY</b> SARD-125 |
| Megalin(2)  | 人类 | Tm-144-TNF <b>ENPIYA</b> QME-53 |
| [0209]      |    |                                 |
| 整联蛋白13-1(1) | 人类 | Tm-18-DTGE <b>NPIY</b> KS       |
| 整联蛋白13-1(2) | 人类 | Tm-30-TTV <b>VNP</b> KYEG       |
| 整联蛋白13(1)   | 果蝇 | Tm-26-WDT <b>ENPIY</b> KQ       |
| 整联蛋白13(2)   | 果蝇 | Tm-35-STFK <b>NPMY</b> AG       |
| APLP1       | 人类 | Tm-33-HGY <b>ENPTY</b> RF       |
| APP         | 人类 | Tm-32-NGY <b>ENPTY</b> KF       |
| APP-样       | 果蝇 | Tm-38-NGY <b>ENPTY</b> KY       |
| 胰岛素受体       | 人类 | Tm-36-YASS <b>NPEY</b> LS       |
| EGR受体(1)    | 人类 | Tm-434-GSVQ <b>NPVY</b> HN      |
| EGR受体(2)    | 人类 | Tm-462-TAVG <b>NPEY</b> LN      |
| EGR受体(3)    | 人类 | Tm-496-ISLD <b>NPDY</b> QQ      |

[0210] 括号中的数字表示同一蛋白中存在多于一个拷贝的基序。该表格和其他表格中的信号应该被视为实例。关键残基以粗体显示。示出了信号之前(即氨基末端)和之后(即羧基末端)的氨基酸数量。缩写:Tm,跨膜;LDL,低密度脂蛋白;LRP1,LDL受体相关蛋白1;APP,13-淀粉样蛋白前体蛋白;APLP1,APP样蛋白1。

[0211] YX'X'Z'型信号存在于在内吞受体如转铁蛋白受体和无唾液酸糖蛋白受体,细胞内分选受体如CI-和CD-MPRs,溶酶体膜蛋白如LAMP-1和LAMP-2中,和TGN蛋白如TGN38和弗林蛋白酶,以及定位于特化的内体-溶酶体细胞器例如蛋白抗原加工区室(例如HLA-DM)和细胞毒性颗粒(例如,GMP-17)的蛋白中。YX'X'Z'型信号参与蛋白从质膜的快速内化。然而,它们的功能不限于胞吞作用,因为相同的基序已经涉及将跨膜蛋白靶向至溶酶体和溶酶体相关的细胞器。

[0212] 在表3中提供YX'X'Z'型信号的实例。

[0213] 表3-YX'X'Z'型信号

|        | 蛋白         | 物种            | 序列                    |
|--------|------------|---------------|-----------------------|
|        | LAMP-1     | 人类            | Tm-RKRSHAGYQTI        |
|        | LAMP-2a    | 人类            | Tm-KHHHAGYEQF         |
|        | LAMP-2a    | 鸡             | Tm-KKHHNTGYEQF        |
|        | LAMP-2b    | 鸡             | Tm-RRKSRTGYQSV        |
|        | LAMP-2c    | 鸡             | Tm-RRKSYAGYQTL        |
|        | LAMP       | 果蝇            | Tm-RRRSTSRGYMSF       |
|        | LAMP       | 蚯蚓            | Tm-RKRSRRGYESV        |
|        | CD63       | 人类            | Tm-KSIRSGYEV          |
|        | GMP-17     | 人类            | Tm-HCGGPRPGYETL       |
|        | GMP-17     | 小鼠            | Tm-HCRTRRAEYETL       |
|        | CD68       | 人类            | Tm-RRRPSAYQAL         |
|        | CD1b       | 人类            | Tm-RRRSYQNIP          |
| [0214] | CD1c       | 人类            | Tm-KKHCSYQDIL         |
|        | CD1d       | 小鼠            | Tm-RRRSAYQDIR         |
|        | CD1        | 大鼠            | Tm-RKRRRSYQDIM        |
|        | Endolyn    | 大鼠            | Tm-KFCKSKERNYHTL      |
|        | Endolyn    | 果蝇            | Tm-KFYKARNERNYHTL     |
|        | TSC403     | 人类            | Tm-KIRLRCQSSGYQRI     |
|        | TSC403     | 小鼠            | Tm-KIRQRHQSSAYQRI     |
|        | Cystinosin | 人类            | Tm-HFCLYRKRPDYDQLN    |
|        | 推定的溶质载体    | 人类            | Tm-12-SLSRSGSYKEI     |
|        | TRP-2      | 人类            | Tm-RRLRKGYTPLMET-1    |
|        | HLA-DM     | 人类            | Tm-RRAGHSSYTPLPGS-9   |
|        | LmpA       | Dictyostelium | Tm-KKLRQQKQQGYQAI     |
|        | 推定的溶酶体都拿白  | Dictyostelium | Tm-RSKSNQNQSYNLIQL    |
|        | LIMP-II    | Dictyostelium | Tm-RKTFYNNNQYNGYN     |
|        | 转铁蛋白受体     | 人类            | 16-PLSYTRFSLA-35-Tm   |
|        | 脱唾液酸糖蛋白受体  | 人类            | MTKEYQDLQHL-29-Tm     |
|        | CI-MPR     | 人类            | Tm-22-SYKYSKVNKE-132  |
|        | CD-MPR     | 人类            | Tm-40-PAAVRGVGDD-16   |
| [0215] | CTLA-4     | 人类            | Tm-10-TGVYVKMPPT-16   |
|        | 弗林蛋白酶      | 人类            | Tm-17-LISYKGLPPE-29   |
|        | TGN38      | 大鼠            | Tm-23-ASDYQRLNLKL     |
|        | gp41       | HIV-1         | Tm-13-RQGYSPLSFQT-144 |
|        | 酸性磷酸酶      | 人类            | Tm-RMQAQQPPGYRHVAD    |
|        |            | GEDHA         |                       |

[0216] 信号形式的解释参见对表1的说明

[0217] 基于二亮氨酸的分选信号 ([DE]X' X' X' LL[LI]) 在许多I型、II型和多跨跨膜蛋白的分选中发挥关键作用。基于二亮氨酸的分选信号参与跨膜蛋白的快速内化和溶酶体降解, 和将蛋白靶向至晚期内体-溶酶体区室。含有该信号的组成型活性形式的跨膜蛋白主要定位于晚期内体和溶酶体。

[0218] 在表4中提供[DE]X' X' X' LL[LI]分选信号的实例。

[0219] 表4-[DE]X'X'X'LL[LI]分选信号

|        | 蛋白            | 物种         | 信号                            |
|--------|---------------|------------|-------------------------------|
|        | CD3- $\gamma$ | 人类         | Tm-8-SDKQTL LPN-26            |
|        | LIMP-II       | 大鼠         | Tm-11-DERAPLIRT               |
|        | Nimb          | 人类         | Tm-37-QEKDPLLKN-7             |
|        | QNR-71        | 鹌鹑         | Tm-37-TERNPLLKS-5             |
|        | Pmel17        | 人类         | Tm-33-GENSPLL SG-3            |
|        | 酪氨酸酶          | 人类         | Tm-8-EEKQPLLME-12             |
|        | 酪氨酸酶          | Medakafish | Tm-16-GERQPLLQS-13            |
|        | 酪氨酸酶          | 鸡          | Tm-8-PEIQPLLTE-13             |
|        | TRP-1         | 金鱼         | Tm-7-EGRQPLLGD-15             |
|        | TRP-1         | 人类         | Tm-7-EANQPLLTD-20             |
|        | TRP-1         | 鸡          | Tm-7-ELHQPLLTD-20             |
|        | TRP-2         | 斑马鱼        | Tm-5-REFEPLLNA-11             |
| [0220] | VMAT2         | 人类         | Tm-6-EEKMAILMD-29             |
|        | VMAT1         | 人类         | Tm-6-EEKLA IL SQ-32           |
|        | VAchT         | 小鼠         | Tm-10-SERDVLLDE-42            |
|        | VAMP4         | 人类         | 19-SERRN LLED-88-Tm           |
|        | NeonatalFcR   | 大鼠         | Tm-16-DDSGDLLPG-19            |
|        | CD4           | 人类         | Tm-12-SQIKRLLSE-17            |
|        | CD4           | 猫          | Tm-12-SHIKRLLE-17             |
|        | GLUT4         | 小鼠         | Tm-17-RRTPSLLEQ-17            |
|        | GLUT4         | 人类         | Tm-17-HRTPSLLEQ-17            |
|        | IRAP          | 大鼠         | 46-EPRGSRL LV R-53-Tm         |
|        | Ii            | 人类         | MDDQRDLISNNEQLPMLGR-11-T      |
|        | Ii            | 小鼠         | MDDQRDLISNHEQLPILGN-10-Tm     |
|        | Ii            | 鸡          | MAEEQRDLISSDGSSGVLP I-12-T    |
|        | Ii-1          | 斑马鱼        | MEPDHQNESLIQ RVPSAETILGR-     |
|        | Ii-2          | 斑马鱼        | MSSEGNETPLISDQSSVNMGPQP-      |
|        | Lamp          | 锥体虫        | Tm-RPRRRTEEDEL LP EEAEGLIDPQN |
|        | Menkesprotein | 人类         | Tm-74-PDKHSLLV GDFREDDDTAL    |
|        | NPC1          | 人类         | Tm-13-TERERLLNF               |
| [0221] | AQP4          | 人类         | Tm-32-VETDDLIL-29             |
|        | RME-2         | 线虫         | Tm-104-FENDSLL                |
|        | Vam3p         | 酿酒酵母       | 153-NEQSPLLHN-121-Tm          |
|        | ALP           | 酿酒酵母       | 7-SEQTRLVP-18-Tm              |
|        | Gap1p         | 酿酒酵母       | Tm-23-EVDLDLLK-24             |

[0222] 信号形式的解释参见对表1的说明

[0223] DX'X'LL信号构成了独特类型的基于二亮氨酸的分类信号。这些信号存在于几种跨膜受体和在TGN和内体之间循环的其它蛋白中,例如CI-和CD-MPRs,sortilin,LDL受体相关蛋白LRP3和LRP10,以及 $\beta$ -分泌酶。

[0224] 在表5中提供DX'X'LL分选信号的实例。

[0225] 表5-DX'X'LL分选信号

|        | 蛋白                | 物种        | 序列                           |
|--------|-------------------|-----------|------------------------------|
|        | CI-MPR            | 人类        | Tm-151-SFHDDS <b>DE</b> DLL  |
|        | CI-MPR            | 牛         | Tm-150-TFHDDS <b>DE</b> DLL  |
|        | CI-MPR            | 兔         | Tm-151-SFHDDS <b>DE</b> DLL  |
|        | CI-MPR            | 鸡         | Tm-148-SFHDDS <b>DE</b> DLL  |
|        | CD-MPR            | 人类        | Tm-54-EESEER <b>DD</b> HLL   |
|        | CD-MPR            | 鸡         | Tm-54-DESEER <b>DD</b> HLL   |
|        | Sortilin          | 人类        | Tm-41-GYHDDS <b>DE</b> DLL   |
|        | SorLA             | 人类        | Tm-41-ITGFSD <b>DD</b> VPMVI |
|        | Head-activator BP | 水螅        | Tm-41-INRFSD <b>DE</b> PLVV  |
|        | LRP3              | 人类        | Tm-237-MLEASD <b>DE</b> AL   |
| [0226] | ST7               | 人类        | Tm-330-KNETSD <b>DE</b> AL   |
|        | LRP10             | 小鼠        | Tm-235-WVVEAE <b>DE</b> PL   |
|        | LRP10             | 人类        | Tm-237-WVAEAE <b>DE</b> PL   |
|        | Beta-分泌酶          | 人类        | Tm-9-HDDFAD <b>DIS</b> LL    |
|        | Mucolipin-1       | 小鼠        | Tm-43-GRDSPED <b>HS</b> LL   |
|        | 非经典MHC-I          | Deermouse | Tm-6-VRCHPE <b>DD</b> RLL    |
|        | FLJ30532          | 人类        | Tm-83-HRVSQ <b>DD</b> LDLL   |
|        | GGA1              | 人类        | 350-ASVSLL <b>DD</b> ELMS    |
|        | GGA1              | 人类        | 415-ASSGL <b>DD</b> LDLLG    |
|        | GGA2              | 人类        | 408-VQNPSAD <b>RN</b> LLD    |
|        | GGA3              | 人类        | 384-NALSWL <b>DE</b> ELLC    |
|        | GGA               | 果蝇        | 447-TVDS <b>IS</b> DDVPLLS   |

[0227] 信号形式的解释参见对表1的说明。丝氨酸和苏氨酸残基加下划线。

[0228] 通过含有CKII磷酸化位点的酸性残基簇提供了另一家族的分选基序。这种类型的基序通常存在于在稳定状态下定位于TGN的跨膜蛋白中,包括激素原加工酶弗林蛋白酶,PC6B,PC7,CPD和PAM,以及疱疹病毒3的糖蛋白E。

[0229] 在表6中提供酸性簇信号的实例。

[0230] 表6-酸性簇分选信号

[0231]

| 蛋白                   | 物种              | 序列                                         |
|----------------------|-----------------|--------------------------------------------|
| 弗林蛋白酶                | 小鼠              | Tm-31-QEECP <u>SD</u> SEEDEG-14            |
| PC6B(1) <sup>a</sup> | 小鼠              | Tm-39-RDRDYDEDD <u>ED</u> DI-36            |
| PC6B(2) PC7          | 小鼠              | Tm-69-LDE <u>TE</u> DDDELEYDD <u>ES</u> -4 |
| CPD                  | 人类              | Tm-38-KDPDEVET <u>ES</u> -47               |
| PAM                  | 人类              | Tm-36-HEFQDETDTEET-6                       |
| VMAT2                | 人类              | Tm-59-QEKEDDG <u>SE</u> SEEEY-12           |
| VMAT1                | 人类              | Tm-35-GEDE <u>SE</u> SD                    |
| VAMP4                | 人类              | Tm-35-GEDSDDEEPDHEE                        |
| 糖蛋白B                 | 人类              | 25-LEDDSDDEEDF-81-Tm                       |
|                      | 人类HCMV          | Tm-125-KDSDEEENV                           |
| 糖蛋白E                 | 疱疹病毒3           | Tm-28-FEDSE <u>ST</u> DEEEF-21             |
| Nef                  | HIV-1(AAL65476) | 55-LEAQEEEEV-139                           |
| Kex1p(1)             | 酿酒酵母            | Tm-29-ADDLESGLGAEDDLEQDEOLEG-40            |
| Kex1p(2)             | 酿酒酵母            | Tm-79-TEID <u>SE</u> FEMIDF                |
| Kex2p                | 酿酒酵母            | Tm-36-TEPEEVEDFD <u>FD</u> LSDEDH-61       |
| Vps10p               | 酿酒酵母            | Tm-112-FEIEEDDVPTLEEEH-37                  |
|                      | 酿酒酵母            |                                            |

[0232] 信号形式的解释参见对表1的说明。丝氨酸和苏氨酸残基加下划线。

[0233] <sup>a</sup>括号中的编号是基序编号。

[0234] KDEL受体结合ER-高尔基体中间区室或早期高尔基体中的蛋白并将其返回ER。虽然常见的哺乳动物信号是KDEL,但是已经显示KDEL受体更紧密地结合HDEL序列(Scheel et al;J.Biol.Chem.268;7465(1993))。细胞内保留信号可以是HDEL。

[0235] KKX'X'和KX'KX'X'信号是可以置于I型膜蛋白的细胞质侧的检索信号(retrieval signals)。这些信号的序列要求由Teasdale&Jackson详细提供(Annu.Rev.Cell Dev.Biol.;12;27(1996))。

[0236] 细胞内保留信号可以选自下组:NPX'Y,YX'X'Z,[DE]X'X'X'L[LI],DX'X'LL,DP[FW],FX'DX'F,NPF,LZX'Z[DE],LLDLL,PWDLW,KDEL,HDEL,KKX'X'或KX'KX'X';其中X'是任意氨基酸,Z'是具有大体积的疏水侧链的氨基酸。

[0237] 细胞内保留信号可以是表2至6中所示的任何序列。

[0238] 细胞内保留信号可以包含酪氨酸酶相关蛋白(TYRP)-1细胞内保留信号。细胞内保留信号可以包含TYRP-1细胞内结构域。细胞内保留信号可以包含序列NQPL LTD (SEQ ID No.35)。

[0239] TYRP1是一种良好表征的黑素体蛋白,其以>99%的效率保留在黑素体(一种特化的溶酶体)中。TYRP1是具有腔性结构域(1-477aa),跨膜结构域(478-501)和胞质结构域(502-537)的537个氨基酸的跨膜蛋白。存在于胞质结构域上的二亮氨酸信号引起蛋白的保留。该二亮氨酸信号具有SEQ ID No.35(NQPL LTD)所示的序列。

[0240] 细胞内保留信号可以在跨膜蛋白的胞内域中。换句话说,细胞内保留信号可以在跨膜蛋白的结构域中,如果该蛋白在细胞表面正确表达,则所述结构域将在细胞膜的细胞内侧。

[0241] 跨膜蛋白的胞内域可以包含至少100个,至少150个,至少200个,至少300个或至少500个氨基酸。

[0242] 包含细胞内保留信号的胞内域可以位于跨膜蛋白的羧基末端。特别地,在跨膜蛋白在肽的氨基末端包含信号序列的情况下,包含细胞内保留信号的胞内域可以位于羧基末端。

[0243] 细胞内保留信号可以在跨膜结构域的近端,例如紧接与其连接。细胞内保留信号可以在跨膜结构域的远端-例如在胞内域的羧基末端。保留信号的定位调节其活性,允许“调节”两个蛋白的相对表达。例如在TYRP1二亮氨酸基序的情况下,近端放置导致低水平的表面表达,而远端放置导致中等表面表达,如实施例中所示。

[0244] 感兴趣的多肽

[0245] 本文中定义的构建体中的核酸序列A或B,或A,B或C中的任意或全部可以编码包含细胞内保留信号的跨膜蛋白。

[0246] 另一核酸序列可以编码任何感兴趣的多肽(POI)。例如,另一POI可以是细胞内蛋白,例如核蛋白,胞质蛋白或定位于膜边界区室的蛋白;分泌蛋白或跨膜蛋白。

[0247] 本文中定义的构建体中的核酸序列的A或B,或A,B或C中的任意或全部可以编码嵌合抗原受体(CAR)。实施例中描述的核酸构建体编码两种嵌合抗原受体。

[0248] 实施例中描述的核酸构建体编码以下多聚蛋白质,其包含按照它们列出的顺序的各种组分:

[0249] 1. 包含抗CD19和抗CD33 CAR的多聚蛋白,其在抗CD19 CAR上具有近端tyrp-1保留

[0250] 来源于人CD8a的信号序列:

[0251] MSLPVTALLLPLALLLHAARP (SEQ ID No.38)

[0252] scFv aCD19:

[0253]

DIQMTQTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISN  
LEQEDIATYFCQQGNTLPYTFGGGKLEITKAGGGSGGGSGGGSGGGGSEVKLQESGPGLVAPSQSLSVTCTVS  
GVSLPDYGVSWIRQPPrKGLEWLGVWGSETTYNSALKSRLTIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYG  
GSYAMDYWGQTSVTVS (SEQ ID No.39)

[0254] 接头:

[0255] SD

[0256] 人CD8aSTK:

[0257] PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No.40)

[0258] 人CD28TM:

[0259] FWVLVVVGVLACYSLLVTVAFIIFWV (SEQ ID No.41)

[0260] 人Typr-1细胞内结构域(保留信号):

[0261] RARRSMDEANQPLLTDQYQCYAEEYEKLQNPNQSVV (SEQ ID No.42)

[0262] 人CD3zeta细胞内结构域:

[0263]

RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPGEGKPRRKNPQEGLYNELQKDKMAEAYSEIGM  
KGERRRGKHDGLYQGLSTATKDTYDALHMQALPPR (SEQ ID No.43)

[0264] 2A肽:

[0265] RAEGRGSLLTCGDVEENPGP (SEQ ID No.24)

[0266] 来源于小鼠 Ig kappa 的信号肽:

[0267] MAVPTQVLGLLLLWLTD (SEQ ID No.44)

[0268] scFv aCD33:

[0269]

RCDIQMTQSPSSLSASVGDRVITITCRASEDIYFNLVWYQQKPKGKAPKLLIYDTNRLADGVPSRFSGSGSGTQYTLTI  
SSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGGSGGGGSGGGGSGGGGSRSEVQLVESGGGLVQPGGSLRLS  
CAASGFTLSNYGMHWIRQAPGKGLEWVSSISLNGGSTYYRDSVKGRFTISRDNASTLYLQMNSLRAEDTAVYYCAA  
QDAYTGGYFDYWGQGTLLTVSSM (SEQ ID No.45)

[0270] 接头:

[0271] DPA

[0272] 来源于人 IgG1 的铰链和 Fc, 其具有突变以防止 FcRg 结合 (HCH2CH3pvaa):

[0273]

EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE  
EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK  
GFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVMHEALHNHYTQKSLSLSPGK  
(SEQ ID No.46)

[0274] 接头:

[0275] KDPK (SEQ ID No.47)

[0276] 人 CD148TM:

[0277] AVFGCIFGALVIVTVGGFIFW (SEQ ID No.48)

[0278] 人 CD148 细胞内结构域:

[0279]

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKYAAELAENRGKNRYNN  
VLPHYDISRVKLSVQTHSTDDYINANYMPGYHSHKKDFIATQGPLPNTLKDFWRMVWEKNVYAIIMLTKEVEQGRKCE  
EYWPSKQAQDYGDITVAMTSEIVLPEWTIRDFTVKNIQTSESHPLRQFHFTSWPDHGVDPDTDLLINFRYLVRDYMK  
QSPPEPILVHCSAGVGRTGTFFIAIDRLIYQIENENTVDVYGIYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKD  
SKVDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No.49)

[0280] 2. 包含抗 CD19 CAR 和抗 CD33 CAR 的多聚蛋白, 其在抗 CD19 CAR 上具有远端 tyrp-1 保留

[0281] 来源于 Human CD8a 的信号肽:

[0282] MSLPVTALLLPLALLLHAARP (SEQ ID No.38)

[0283] scFv aCD19:

[0284]

DIQMTQTSSLSASLGDRVITISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISN  
LEQEDIATYFCQQGNTLPYTFGGGKLEITKAGGGGSGGGGSGGGGSGGGGSEVKLQESGPGLVAPSQSLSVTCTVS  
GVSLPDYGVSWIRQPPRKGLEWLGVWGETTYNSALKSRLTIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYG  
GSYAMDYWGQTSVTVS (SEQ ID No.39)

[0285] 接头:

[0286] SD

- [0287] 人CD8aSTK:
- [0288] PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No.40)
- [0289] 人CD28TM:
- [0290] FWVLVVVGGVLACYSLLVTVAFIIFWV (SEQ ID No.41)
- [0291] 人CD3zeta细胞内结构域:
- [0292] RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGM  
KGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ ID No.43)
- [0293] 人Typr-1细胞内结构域(保留信号):
- [0294] RARRSMDEANQPLLTDQYQCYAEEYEKLQNPNSVV (SEQ ID No.42)
- [0295] 2A肽:
- [0296] RAEGRGSLLTCGDVEENPGP (SEQ ID No.24)
- [0297] 来源于小鼠Ig kappa的信号肽:
- [0298] MAVPTQVLGLLLLWLTD (SEQ ID No.44)
- [0299] scFv aCD33:
- [0300] RCDIQMTQSPSSLSASVGDRVITTCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPSRFSGSGSGTQYTLTI  
SSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGGSGGGGSGGGGSRSEVQLVESGGGLVQPGGSLRLS  
CAASGFTLSNYGMHWIRQAPGKGLEWVSSISLNGGSTYYRDSVKGRFTISRDNASTLYLQMNSLRAEDTAVYYCAA  
QDAYTGGYFDYWGQGLVTVSSM (SEQ ID No.45)
- [0301] 接头:
- [0302] DPA
- [0303] 来源于人IgG1的铰链和Fc,其具有突变以防止FcRg结合 (HCH2CH3pvaa):
- [0304] EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE  
EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK  
GFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVVFSCVMHEALHNHYTQKSLSLSPGK  
(SEQ ID No.46)
- [0305] 接头:
- [0306] KDPK (SEQ ID No.47)
- [0307] 人CD148TM:
- [0308] AVFGCIFGALVIVTVGGFIFW (SEQ ID No.48)
- [0309] 人CD148细胞内结构域:
- [0310] RKKRKDAKNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKYAAELAENRGKNRYNN  
VLPYDISRVKLSVQTHSTDDYINANYMPGYHKKDFIATQGPLPNTLKDFWRMVWEKNVYAIIMLTKEVEQGRKCE  
EYWPSKQAQDYGDITVAMTSEIVLPEWTIRDFTVKNIQTSESHPLRQFHFTSWPDHGVPTDILLINFRYLVRDYM  
KQSPPEPILVHCSAGVGRTGTFFIAIDRLIYQIENENTVDVYGIYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKD  
SKVDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No.49)

- [0311] 3.包含抗CD19 CAR和抗CD33 CAR的多聚蛋白,其在抗CD33 CAR上具有E3/19K保留
- [0312] 来源于人CD8a的信号肽:
- [0313] MSLPVTALLLPLALLLHAARP (SEQ ID No.38)
- [0314] scFv aCD19:
- [0315]
- DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRFSGSGSGTDYSLTISN  
LEQEDIATYFCQQGNTLPYTFGGGTKLEITKAGGGSGGGSGGGSGGGGSEVKLQESGPGLVAPSQSLSVTCTVS  
GVSLPDYGVSWIRQPPRKGLEWLGVWGSETTYNSALKSRLTI IKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYG  
GSYAMDYWGQTSVTVS (SEQ ID No.39)
- [0316] 接头:
- [0317] SD
- [0318] 人CD8aSTK:
- [0319] PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No.40)
- [0320] 人CD28TM:
- [0321] FWVLVVVGVLACYSLLVTVAFIIFWV (SEQ ID No.41)
- [0322] 人CD3zeta细胞内结构域:
- [0323]
- RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGM  
KGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ ID No.43)
- [0324] 2A肽:
- [0325] RAEGRGSLLTCGDVEENPGP (SEQ ID No.24)
- [0326] 来源于小鼠Ig kappa的信号肽:
- [0327] MAVPTQVLGLLLLWLTDA (SEQ ID No.44)
- [0328] scFv aCD33:
- [0329]
- RCDIQMTQSPSSLSASVGDRVTITCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPSRFSGSGSGTQYTLTI  
SSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGSGGGSGGGSGGGGSRSEVQLVESGGGLVQPGGSLRLS  
CAASGFTLSNYGMHWIRQAPGKGLEWVSSISLNGGSTYYRDSVKGRFTISRDNASTLYLQMNSLRAEDTAVYYCAA  
QDAYTGGYFDYWGQGLVTVSSM (SEQ ID No.45)
- [0330] 接头:
- [0331] DPA
- [0332] 来源于人IgG1的铰链和Fc,其具有突变以防止FcRg结合 (HCH2CH3pvaa):
- [0333]
- EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPRE  
EQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK  
GFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFSVSMHEALHNHYTQKSLSLSPGK  
(SEQ ID No.46)
- [0334] 接头:
- [0335] KDPK (SEQ ID No.47)

[0336] 人CD148TM:

[0337] AVFGCIFGALVIVTVGGFIFW (SEQ ID No.48)

[0338] 人CD148细胞内结构域:

[0339]

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKYAAELAENRGKNRYNN  
VLPYDISRVKLSVQTHSTDDYINANYMPGYHKKDFIATQGPLPNTLKDFWRMVWEKNVYAIIMLTKEVEQGRKCE  
EYWPSKQAQDYGDITVAMTSEIVLPEWTIRDFTVKNIQTSESHPLRQFHFTSWPDHGVDPDITDILLINFRYLVRDYM  
KQSPPEPILVHCSAGVGRTGTFFAIDRLIYQIENENTVDVYGIYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKD  
SKVDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No.49)

[0340] 腺病毒E3/19K胞质尾:

[0341] KYKSRRSFIDEKKMP (SEQ ID No.36)

[0342] 在上述多聚蛋白3中,可以使用具有以下序列的截短型替代E3/19K胞质尾:  
DEKKMP。

[0343] 如实施例所示,发现对于tyrp-1保留信号,通过将该保留信号置于上文给出的序  
列中的“人CD28TM”和“人CD3 zeta细胞内结构域”之间,可以实现低水平的表达,而通过将  
保留信号置于“人CD3 zeta细胞内结构域”和“2A肽”之间可以实现中等水平的表达。

[0344] 当置于抗CD33 CAR的C末端时,E3/19K保留信号导致抗CD33 CAR的表达降低。

[0345] 载体

[0346] 本发明还提供了包含根据本发明第一个方面的核酸构建体的载体。

[0347] 可以使用此类载体将核酸构建体引入宿主细胞,使得其表达第一和第二多肽。

[0348] 载体可以是例如质粒或病毒载体,例如逆转录病毒载体或慢病毒载体,或基于转  
座子的载体或合成的mRNA。

[0349] 载体可以能够转染或转导哺乳动物细胞,例如T细胞。

[0350] 细胞

[0351] 本发明还提供了包含本发明的核酸构建体或载体的细胞,其表达由核酸序列编码  
的第一和第二多肽。

[0352] 细胞可以是能够在细胞表面表达跨膜蛋白的任何真核细胞,例如免疫细胞。

[0353] 蛋白

[0354] 本发明还提供了包含细胞外结构域,跨膜结构域和胞内域的跨膜蛋白,其中所述  
胞内域包含如本文所定义的外源细胞内保留信号。

[0355] “外源”是指细胞内保留信号不是跨膜蛋白的野生型序列的一部分。野生型序列是  
指通常在自然界存在的蛋白的氨基酸序列。

[0356] 本发明人已经证明在跨膜蛋白的胞内域中包括外源细胞内保留信号导致跨膜蛋  
白被引导至细胞内区室。因此,与在跨膜结构域中不包括细胞内保留信号的等同跨膜蛋白  
相比,在细胞表面表达的该跨膜蛋白的量减少。

[0357] 方法

[0358] 在进一步的方面,本发明涉及用于调节从单个核酸构建体与第二蛋白一起作为多  
聚蛋白表达的跨膜蛋白的相对细胞表面表达的方法;其通过在编码该跨膜蛋白的核酸序列  
中包括细胞内保留信号进行。

[0359] 现在将通过实施例进一步描述本发明,这些实施例旨在帮助本领域普通技术人员实施本发明,而并不意图以任何方式限制本发明的范围。

## 实施例

### [0360] 实施例1-TYRP1溶酶体保留信号的剖析

[0361] 使用多个构建体确定酪氨酸酶相关蛋白1 (TYRP1) 保留信号当在更复杂的胞内域的情况下时引起多肽的保留的能力(图2)。将野生型构建体与其中增强型绿色荧光蛋白(eGFP)添加至TYRP1胞内域或替换TYRP1胞内域的构建体比较。在添加eGFP的情况下,将其置于天然的胞内域之后或之前,因此保留信号位于其天然位置(位于膜正下方),或位于其远端。

[0362] 所有构建体与IRES.CD34共表达。图2中显示了转导的SupT1细胞染色及细胞内和表面染色。

[0363] 发现胞内域的替换导致非常明亮的表面表达,在保留信号之后引入eGFP导致几乎没有表面表达,而在保留信号之前引入导致中等表面表达(图2)。

### [0364] 实施例2-调节与不同的蛋白一起从单个表达盒表达的跨膜蛋白的相对表达

[0365] 修饰编码两种CAR跨膜蛋白的表达盒,使得CAR蛋白之一具有在膜的近端或远端引入的来自TYRP1的溶酶体保留信号。将这两种新的变体中每种在细胞表面的表达与原始的未修饰CAR蛋白的表达比较。

[0366] 从血液中分离PBMC,并使用PHA和IL-2刺激。两天后,使用含有CD19:CD33 CAR构建体的逆转录病毒在纤维连接蛋白包被的板上转导细胞。在第5天,通过流式细胞术评估由构建体翻译的两种CAR的表达水平,并将细胞耗竭CD56+细胞(主要是NK细胞)。在第6天,将PBMC与靶细胞以1:2的效应细胞与靶细胞比率共培养。在第8天,收集上清液并通过ELISA分析IFN- $\gamma$ 分泌。

[0367] 使用Tyrrp1-eGFP融合物观察到的模式,观察到具有远端保留信号的修饰的跨膜蛋白的表达的一些降低和在近端保留信号的情况下的显著降低。如预期的,从盒中表达第二跨膜蛋白没有改变(图3)。

### [0368] 实施例3-使用来自腺病毒E3/19K蛋白的保留信号调节表达

[0369] 人腺病毒E3/19K蛋白是内质网/高尔基体的I型跨膜糖蛋白,其消除主要组织相容性复合物I类(MHC-1)和MHC-1相关链A和B(MICA/B)分子的细胞表面转运。保留基序鉴定为依赖于腺病毒E3/19K蛋白的胞质尾。更具体地,发现最后6aa DEKKMP对于保留是最重要的。发现最佳定位位于蛋白的c末端。

[0370] 修饰如实施例2中所述的编码两种CAR跨膜蛋白的表达盒,使得CAR蛋白之一具有来自腺病毒E3/19K蛋白的保留基序。在该实验中,保留基序在表达盒中的第二CAR(抗CD33抑制性CAR)上。

[0371] 生成了构建体,其包含腺病毒E3/19K蛋白的整个胞质尾或仅来自E3/19K的最后6aa(DEKKMP),其发现对于其高尔基体/ER保留能力是至关重要的(图4)。将这些构建体转染到293T细胞中,并首先使用嵌合可溶性CD19-兔Fc和嵌合可溶性CD33-小鼠Fc蛋白染色。然后用抗兔Fc-FITC和抗小鼠Fc-APC染色这些细胞(图5)。当全长腺病毒E3/19K蛋白或DEKKMP基序置于抗CD33受体上时,这些细胞呈现明确的保留,但对抗CD19受体表达水平没有影响。

[0372] 上述说明书中提到的所有出版物通过引用并入本文。在不脱离本发明的范围和精神的情况下,本发明所述的方法和系统的各种修改和变化对于本领域技术人员将是显而易见的。虽然已经结合具体的优选实施方案描述了本发明,但是应当理解,所要求保护的本发明不应被过度限于这些具体实施方案。实际上,对于分子生物学,细胞生物学或相关领域的技术人员显而易见的用于实施本发明的所述模式的各种修改旨在在所附权利要求书的范围内。

## 序列表

&lt;110&gt; UCL商务股份有限公司

&lt;120&gt; 核酸构建体

&lt;130&gt; P106298PCT

&lt;150&gt; GB1507104.6

&lt;151&gt; 2015-04-27

&lt;160&gt; 197

&lt;170&gt; PatentIn version 3.5

&lt;210&gt; 1

&lt;211&gt; 8

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 切割位点

&lt;400&gt; 1

Asp Ile Glu Thr Asn Pro Gly Pro

1 5

&lt;210&gt; 2

&lt;211&gt; 8

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 切割位点

&lt;400&gt; 2

Asp Val Glu Thr Asn Pro Gly Pro

1 5

&lt;210&gt; 3

&lt;211&gt; 8

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 切割位点

&lt;400&gt; 3

Asp Val Glu Met Asn Pro Gly Pro

1 5

&lt;210&gt; 4

&lt;211&gt; 8

&lt;212&gt; PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 4

Asp Val Glu Ser Asn Pro Gly Pro

1 5

<210> 5

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 5

Asp Met Glu Ser Asn Pro Gly Pro

1 5

<210> 6

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 6

Asp Val Glu Leu Asn Pro Gly Pro

1 5

<210> 7

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 7

Asp Val Glu Glu Asn Pro Gly Pro

1 5

<210> 8

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 8

Asp Ile Glu Leu Asn Pro Gly Pro

1 5

<210> 9

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 9

Asp Ile Glu Gln Asn Pro Gly Pro

1 5

<210> 10

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 切割位点

<400> 10

Asp Ser Glu Phe Asn Pro Gly Pro

1 5

<210> 11

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 切割序列

<400> 11

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn

1 5 10 15

Pro Gly Pro

<210> 12

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 切割序列

<400> 12

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

|   |   |    |    |
|---|---|----|----|
| 1 | 5 | 10 | 15 |
|---|---|----|----|

Pro Gly Pro  
<210> 13  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 13  
Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ile Asn

|   |   |    |    |
|---|---|----|----|
| 1 | 5 | 10 | 15 |
|---|---|----|----|

Pro Gly Pro  
<210> 14  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 14  
Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Phe Asn

|   |   |    |    |
|---|---|----|----|
| 1 | 5 | 10 | 15 |
|---|---|----|----|

Pro Gly Pro  
<210> 15  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 15  
Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser His

|   |   |    |    |
|---|---|----|----|
| 1 | 5 | 10 | 15 |
|---|---|----|----|

Pro Gly Pro  
<210> 16  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 16

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Glu  
1                      5                      10                      15  
Pro Gly Pro  
<210> 17  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 17  
Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Gln  
1                      5                      10                      15  
Pro Gly Pro  
<210> 18  
<211> 19  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割序列  
<400> 18  
Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn  
1                      5                      10                      15  
Pro Gly Gly  
<210> 19  
<211> 20  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割位点, 2A样序列  
<400> 19  
Tyr His Ala Asp Tyr Tyr Lys Gln Arg Leu Ile His Asp Val Glu Met  
1                      5                      10                      15  
Asn Pro Gly Pro  
                    20  
<210> 20  
<211> 20  
<212> PRT  
<213> 人工序列  
<220>

<223> 切割位点, 2A样序列

<400> 20

His Tyr Ala Gly Tyr Phe Ala Asp Leu Leu Ile His Asp Ile Glu Thr

1 5 10 15

Asn Pro Gly Pro

20

<210> 21

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 21

Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser

1 5 10 15

Asn Pro Gly Pro

20

<210> 22

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 22

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

Pro Gly Pro

<210> 23

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 23

Ala Ala Arg Gln Met Leu Leu Leu Leu Ser Gly Asp Val Glu Thr Asn

1 5 10 15

Pro Gly Pro

<210> 24

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 24

Arg Ala Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu

1 5 10 15

Asn Pro Gly Pro

20

<210> 25

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 25

Thr Arg Ala Glu Ile Glu Asp Glu Leu Ile Arg Ala Gly Ile Glu Ser

1 5 10 15

Asn Pro Gly Pro

20

<210> 26

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 26

Thr Arg Ala Glu Ile Glu Asp Glu Leu Ile Arg Ala Asp Ile Glu Ser

1 5 10 15

Asn Pro Gly Pro

20

<210> 27

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 切割位点, 2A样序列

<400> 27

Ala Lys Phe Gln Ile Asp Lys Ile Leu Ile Ser Gly Asp Val Glu Leu

|                                                                 |   |    |    |
|-----------------------------------------------------------------|---|----|----|
| 1                                                               | 5 | 10 | 15 |
| Asn Pro Gly Pro                                                 |   |    |    |
| 20                                                              |   |    |    |
| <210> 28                                                        |   |    |    |
| <211> 20                                                        |   |    |    |
| <212> PRT                                                       |   |    |    |
| <213> 人工序列                                                      |   |    |    |
| <220>                                                           |   |    |    |
| <223> 切割位点, 2A样序列                                               |   |    |    |
| <400> 28                                                        |   |    |    |
| Ser Ser Ile Ile Arg Thr Lys Met Leu Val Ser Gly Asp Val Glu Glu |   |    |    |
| 1                                                               | 5 | 10 | 15 |
| Asn Pro Gly Pro                                                 |   |    |    |
| 20                                                              |   |    |    |
| <210> 29                                                        |   |    |    |
| <211> 20                                                        |   |    |    |
| <212> PRT                                                       |   |    |    |
| <213> 人工序列                                                      |   |    |    |
| <220>                                                           |   |    |    |
| <223> 切割位点, 2A样序列                                               |   |    |    |
| <400> 29                                                        |   |    |    |
| Cys Asp Ala Gln Arg Gln Lys Leu Leu Leu Ser Gly Asp Ile Glu Gln |   |    |    |
| 1                                                               | 5 | 10 | 15 |
| Asn Pro Gly Pro                                                 |   |    |    |
| 20                                                              |   |    |    |
| <210> 30                                                        |   |    |    |
| <211> 20                                                        |   |    |    |
| <212> PRT                                                       |   |    |    |
| <213> 人工序列                                                      |   |    |    |
| <220>                                                           |   |    |    |
| <223> 切割位点, 2A样序列                                               |   |    |    |
| <400> 30                                                        |   |    |    |
| Tyr Pro Ile Asp Phe Gly Gly Phe Leu Val Lys Ala Asp Ser Glu Phe |   |    |    |
| 1                                                               | 5 | 10 | 15 |
| Asn Pro Gly Pro                                                 |   |    |    |
| 20                                                              |   |    |    |
| <210> 31                                                        |   |    |    |
| <211> 57                                                        |   |    |    |
| <212> PRT                                                       |   |    |    |

<213> 人工序列

<220>

<223> 切割序列

<400> 31

Val Thr Glu Leu Leu Tyr Arg Met Lys Arg Ala Glu Thr Tyr Cys Pro

1 5 10 15

Arg Pro Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile

20 25 30

Val Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala

35 40 45

Gly Asp Val Glu Ser Asn Pro Gly Pro

50 55

<210> 32

<211> 40

<212> PRT

<213> 人工序列

<220>

<223> 切割序列

<400> 32

Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile Val

1 5 10 15

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

20 25 30

Asp Val Glu Ser Asn Pro Gly Pro

35 40

<210> 33

<211> 33

<212> PRT

<213> 人工序列

<220>

<223> 切割序列

<400> 33

Glu Ala Arg His Lys Gln Lys Ile Val Ala Pro Val Lys Gln Thr Leu

1 5 10 15

Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly

20 25 30

Pro

<210> 34

<211> 24

<212> PRT

<213> 人工序列

<220>

<223> 切割序列

<400> 34

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

1 5 10 15

Asp Val Glu Ser Asn Pro Gly Pro

20

<210> 35

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<400> 35

Asn Gln Pro Leu Leu Thr Asp

1 5

<210> 36

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<400> 36

Lys Tyr Lys Ser Arg Arg Ser Phe Ile Asp Glu Lys Lys Met Pro

1 5 10 15

<210> 37

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<400> 37

Asp Glu Lys Lys Met Pro

1 5

<210> 38

<211> 21

<212> PRT

<213> 人工序列

<220>

<223> 来源于人CD8a的信号肽

<400> 38

```
Met Ser Leu Pro Val Thr Ala Leu Leu Leu Pro Leu Ala Leu Leu Leu
1           5           10           15
His Ala Ala Arg Pro
                20
```

<210> 39

<211> 248

<212> PRT

<213> 人工序列

<220>

<223> scFv aCD19

<400> 39

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

Lys Ser Arg Leu Thr Ile Ile Lys Asp Asn Ser Lys Ser Gln Val Phe  
 195 200 205  
 Leu Lys Met Asn Ser Leu Gln Thr Asp Asp Thr Ala Ile Tyr Tyr Cys  
 210 215 220  
 Ala Lys His Tyr Tyr Tyr Gly Gly Ser Tyr Ala Met Asp Tyr Trp Gly  
 225 230 235 240  
 Gln Gly Thr Ser Val Thr Val Ser  
 245

<210> 40

<211> 47

<212> PRT

<213> 人工序列

<220>

<223> 人CD8aSTK

<400> 40

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

<210> 41

<211> 27

<212> PRT

<213> 人工序列

<220>

<223> 人CD28TM

<400> 41

Phe Trp Val Leu Val Val Val Gly Gly Val Leu Ala Cys Tyr Ser Leu  
 1 5 10 15  
 Leu Val Thr Val Ala Phe Ile Ile Phe Trp Val  
 20 25

<210> 42

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> Human Typr-1 intracellular domain (retention signal)

<400> 42

Arg Ala Arg Arg Ser Met Asp Glu Ala Asn Gln Pro Leu Leu Thr Asp  
 1                      5                      10                      15  
 Gln Tyr Gln Cys Tyr Ala Glu Glu Tyr Glu Lys Leu Gln Asn Pro Asn  
                     20                      25                      30  
 Gln Ser Val Val  
                     35

<210> 43

<211> 113

<212> PRT

<213> 人工序列

<220>

<223> 人CD3zeta细胞内结构域

<400> 43

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

Arg

<210> 44

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> 来源于小鼠Ig kappa的信号肽

<400> 44

Met Ala Val Pro Thr Gln Val Leu Gly Leu Leu Leu Leu Trp Leu Thr  
 1                      5                      10                      15

Asp Ala

<210> 45

&lt;211&gt; 254

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; scFv aCD33

&lt;400&gt; 45

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

&lt;210&gt; 46

&lt;211&gt; 231

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

<223> 来源于人IgG1的铰链和Fc,具有突变以防止  
FcRg结合 (HCH2CH3pvaa)

&lt;400&gt; 46

```

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

```

&lt;210&gt; 47

&lt;211&gt; 4

<212> PRT

<213> 人工序列

<220>

<223> 接头

<400> 47

Lys Asp Pro Lys

1

<210> 48

<211> 21

<212> PRT

<213> 人工序列

<220>

<223> 人CD148TM

<400> 48

Ala Val Phe Gly Cys Ile Phe Gly Ala Leu Val Ile Val Thr Val Gly

1

5

10

15

Gly Phe Ile Phe Trp

20

<210> 49

<211> 341

<212> PRT

<213> 人工序列

<220>

<223> 人CD148细胞内结构域

<400> 49

Arg Lys Lys Arg Lys Asp Ala Lys Asn Asn Glu Val Ser Phe Ser Gln

1

5

10

15

Ile Lys Pro Lys Lys Ser Lys Leu Ile Arg Val Glu Asn Phe Glu Ala

20

25

30

Tyr Phe Lys Lys Gln Gln Ala Asp Ser Asn Cys Gly Phe Ala Glu Glu

35

40

45

Tyr Glu Asp Leu Lys Leu Val Gly Ile Ser Gln Pro Lys Tyr Ala Ala

50

55

60

Glu Leu Ala Glu Asn Arg Gly Lys Asn Arg Tyr Asn Asn Val Leu Pro

65

70

75

80

Tyr Asp Ile Ser Arg Val Lys Leu Ser Val Gln Thr His Ser Thr Asp

85

90

95

Asp Tyr Ile Asn Ala Asn Tyr Met Pro Gly Tyr His Ser Lys Lys Asp

100

105

110

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

&lt;210&gt; 50

&lt;211&gt; 4

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 细胞内保留信号

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (4) .. (4)

<223> Xaa可以是任何天然发生的氨基酸

<400> 50

Asn Pro Phe Xaa

1

<210> 51

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<400> 51

Lys Asp Glu Leu

1

<210> 52

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<220>

<221> misc\_feature

<222> (3) .. (4)

<223> Xaa可以是任何天然发生的氨基酸

<400> 52

Lys Lys Xaa Xaa

1

<210> 53

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<220>

<221> misc\_feature

<222> (2) .. (2)

<223> Xaa 可以是任何天然发生的氨基酸

<220>

<221> misc\_feature

<222> (4) .. (5)

<223> Xaa 可以是任何天然发生的氨基酸  
<400> 53  
Lys Xaa Lys Xaa Xaa  
1 5  
<210> 54  
<211> 4  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<220>  
<221> misc\_feature  
<222> (3) .. (3)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 54  
Asn Pro Xaa Tyr  
1  
<210> 55  
<211> 4  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<220>  
<221> MISC\_FEATURE  
<222> (2) .. (2)  
<223> Xaa 可以是任何天然发生的氨基酸.  
<220>  
<221> MISC\_FEATURE  
<222> (3) .. (3)  
<223> Xaa 可以是任何天然发生的氨基酸.  
<220>  
<221> MISC\_FEATURE  
<222> (4) .. (4)  
<223> Xaa 是具有大疏水侧链的氨基酸.  
<400> 55  
Tyr Xaa Xaa Xaa  
1  
<210> 56

<211> 8  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<220>  
<221> misc\_feature  
<222> (3) .. (5)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 56  
Asp Glu Xaa Xaa Xaa Leu Leu Ile  
1 5  
<210> 57  
<211> 5  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<220>  
<221> misc\_feature  
<222> (2) .. (3)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 57  
Asp Xaa Xaa Leu Leu  
1 5  
<210> 58  
<211> 4  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<400> 58  
Asp Pro Phe Trp  
1  
<210> 59  
<211> 5  
<212> PRT  
<213> 人工序列  
<220>

<223> 细胞内保留信号  
<220>  
<221> misc\_feature  
<222> (2) .. (2)  
<223> Xaa 可以是任何天然发生的氨基酸  
<220>  
<221> misc\_feature  
<222> (4) .. (4)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 59  
Phe Xaa Asp Xaa Phe  
1 5  
<210> 60  
<211> 6  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<220>  
<221> MISC\_FEATURE  
<222> (2) .. (2)  
<223> Xaa 是具有大疏水侧链的氨基酸.  
<220>  
<221> MISC\_FEATURE  
<222> (3) .. (3)  
<223> Xaa 可以是任何天然发生的氨基酸.  
<220>  
<221> MISC\_FEATURE  
<222> (4) .. (4)  
<223> Xaa 是具有大疏水侧链的氨基酸.  
<400> 60  
Leu Xaa Xaa Xaa Asp Glu  
1 5  
<210> 61  
<211> 5  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号

<400> 61  
Leu Leu Asp Leu Leu  
1 5  
<210> 62  
<211> 5  
<212> PRT  
<213> 人工序列  
<220>  
<223> 细胞内保留信号  
<400> 62  
Pro Trp Asp Leu Trp  
1 5  
<210> 63  
<211> 4  
<212> PRT  
<213> 人工序列  
<220>  
<223> Furin 切割位点  
<220>  
<221> MISC\_FEATURE  
<222> (2) .. (2)  
<223> Xaa 可以是任何天然发生的氨基酸.  
<220>  
<221> MISC\_FEATURE  
<222> (3) .. (3)  
<223> Xaa可以是ARG或LYS  
<400> 63  
Arg Xaa Xaa Arg  
1  
<210> 64  
<211> 7  
<212> PRT  
<213> 人工序列  
<220>  
<223> 烟草蚀刻病毒 (TEV) 切割位点  
<400> 64  
Glu Asn Leu Tyr Phe Gln Ser  
1 5  
<210> 65

<211> 8  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割位点  
<220>  
<221> misc\_feature  
<222> (2) .. (2)  
<223> Xaa 可以是任何天然发生的氨基酸  
<220>  
<221> misc\_feature  
<222> (4) .. (4)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 65  
Asp Xaa Glu Xaa Asn Pro Gly Pro  
1 5  
<210> 66  
<211> 8  
<212> PRT  
<213> 人工序列  
<220>  
<223> 切割位点  
<220>  
<221> MISC\_FEATURE  
<222> (2) .. (2)  
<223> Xaa 可以选自下组: ILE, VAL, MET和  
SER  
<220>  
<221> MISC\_FEATURE  
<222> (4) .. (4)  
<223> Xaa 可以选自下组: THR, MET, SER, LEU,  
GLU, GLN和PHE.  
<400> 66  
Asp Xaa Glu Xaa Asn Pro Gly Pro  
1 5  
<210> 67  
<211> 4  
<212> PRT  
<213> 人工序列

<220>

<223> 内质网 (ER) 保留信号

<220>

<221> misc\_feature

<222> (2) .. (2)

<223> Xaa 可以是任何天然发生的氨基酸

<400> 67

Arg Xaa Arg Arg

1

<210> 68

<211> 12

<212> PRT

<213> 人类

<400> 68

Ile Asn Phe Asp Asn Pro Val Tyr Gln Lys Thr Thr

1

5

10

<210> 69

<211> 12

<212> PRT

<213> 人类

<400> 69

Val Glu Ile Gly Asn Pro Thr Tyr Lys Met Tyr Glu

1

5

10

<210> 70

<211> 12

<212> PRT

<213> 人类

<400> 70

Thr Asn Phe Thr Asn Pro Val Tyr Ala Thr Leu Tyr

1

5

10

<210> 71

<211> 12

<212> PRT

<213> 果蝇

<400> 71

Gly Asn Phe Ala Asn Pro Val Tyr Glu Ser Met Tyr

1

5

10

<210> 72

<211> 12

<212> PRT

<213> 秀丽隐杆线虫

<400> 72

Thr Thr Phe Thr Asn Pro Val Tyr Glu Leu Glu Asp

1 5 10

<210> 73

<211> 12

<212> PRT

<213> 秀丽隐杆线虫

<400> 73

Leu Arg Val Asp Asn Pro Leu Tyr Asp Pro Asp Ser

1 5 10

<210> 74

<211> 12

<212> PRT

<213> 人类

<400> 74

Ile Ile Phe Glu Asn Pro Met Tyr Ser Ala Arg Asp

1 5 10

<210> 75

<211> 12

<212> PRT

<213> 人类

<400> 75

Thr Asn Phe Glu Asn Pro Ile Tyr Ala Gln Met Glu

1 5 10

<210> 76

<211> 12

<212> PRT

<213> 人类

<400> 76

Asp Thr Gly Glu Asn Pro Ile Tyr Lys Ser Ala Val

1 5 10

<210> 77

<211> 11

<212> PRT

<213> 人类

<400> 77

Thr Thr Val Val Asn Pro Lys Tyr Glu Gly Lys

|                                                 |   |    |
|-------------------------------------------------|---|----|
| 1                                               | 5 | 10 |
| <210> 78                                        |   |    |
| <211> 12                                        |   |    |
| <212> PRT                                       |   |    |
| <213> 果蝇                                        |   |    |
| <400> 78                                        |   |    |
| Trp Asp Thr Glu Asn Pro Ile Tyr Lys Gln Ala Thr |   |    |
| 1                                               | 5 | 10 |
| <210> 79                                        |   |    |
| <211> 11                                        |   |    |
| <212> PRT                                       |   |    |
| <213> 果蝇                                        |   |    |
| <400> 79                                        |   |    |
| Ser Thr Phe Lys Asn Pro Met Tyr Ala Gly Lys     |   |    |
| 1                                               | 5 | 10 |
| <210> 80                                        |   |    |
| <211> 12                                        |   |    |
| <212> PRT                                       |   |    |
| <213> 人类                                        |   |    |
| <400> 80                                        |   |    |
| His Gly Tyr Glu Asn Pro Thr Tyr Arg Phe Leu Glu |   |    |
| 1                                               | 5 | 10 |
| <210> 81                                        |   |    |
| <211> 12                                        |   |    |
| <212> PRT                                       |   |    |
| <213> 人类                                        |   |    |
| <400> 81                                        |   |    |
| Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu |   |    |
| 1                                               | 5 | 10 |
| <210> 82                                        |   |    |
| <211> 12                                        |   |    |
| <212> PRT                                       |   |    |
| <213> 果蝇                                        |   |    |
| <400> 82                                        |   |    |
| Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Tyr Phe Glu |   |    |
| 1                                               | 5 | 10 |
| <210> 83                                        |   |    |
| <211> 12                                        |   |    |
| <212> PRT                                       |   |    |

<213> 人类

<400> 83

Tyr Ala Ser Ser Asn Pro Glu Tyr Leu Ser Ala Ser

1 5 10

<210> 84

<211> 12

<212> PRT

<213> 人类

<400> 84

Gly Ser Val Gln Asn Pro Val Tyr His Asn Gln Pro

1 5 10

<210> 85

<211> 12

<212> PRT

<213> 人类

<400> 85

Thr Ala Val Gly Asn Pro Glu Tyr Leu Asn Thr Val

1 5 10

<210> 86

<211> 12

<212> PRT

<213> 人类

<400> 86

Ile Ser Leu Asp Asn Pro Asp Tyr Gln Gln Asp Phe

1 5 10

<210> 87

<211> 11

<212> PRT

<213> 人类

<400> 87

Arg Lys Arg Ser His Ala Gly Tyr Gln Thr Ile

1 5 10

<210> 88

<211> 10

<212> PRT

<213> 人类

<400> 88

Lys His His His Ala Gly Tyr Glu Gln Phe

1 5 10

<210> 89

<211> 11

<212> PRT

<213> 原鸡

<400> 89

Lys Lys His His Asn Thr Gly Tyr Glu Gln Phe

1 5 10

<210> 90

<211> 11

<212> PRT

<213> 原鸡

<400> 90

Arg Arg Lys Ser Arg Thr Gly Tyr Gln Ser Val

1 5 10

<210> 91

<211> 11

<212> PRT

<213> 原鸡

<400> 91

Arg Arg Lys Ser Tyr Ala Gly Tyr Gln Thr Leu

1 5 10

<210> 92

<211> 12

<212> PRT

<213> 果蝇

<400> 92

Arg Arg Arg Ser Thr Ser Arg Gly Tyr Met Ser Phe

1 5 10

<210> 93

<211> 11

<212> PRT

<213> *Lumbricus terrestris*

<400> 93

Arg Lys Arg Ser Arg Arg Gly Tyr Glu Ser Val

1 5 10

<210> 94

<211> 10

<212> PRT

<213> 人类

<400> 94

Lys Ser Ile Arg Ser Gly Tyr Glu Val Met

1 5 10

<210> 95

<211> 12

<212> PRT

<213> 人类

<400> 95

His Cys Gly Gly Pro Arg Pro Gly Tyr Glu Thr Leu

1 5 10

<210> 96

<211> 12

<212> PRT

<213> 小鼠

<400> 96

His Cys Arg Thr Arg Arg Ala Glu Tyr Glu Thr Leu

1 5 10

<210> 97

<211> 10

<212> PRT

<213> 人类

<400> 97

Arg Arg Arg Pro Ser Ala Tyr Gln Ala Leu

1 5 10

<210> 98

<211> 9

<212> PRT

<213> 人类

<400> 98

Arg Arg Arg Ser Tyr Gln Asn Ile Pro

1 5

<210> 99

<211> 10

<212> PRT

<213> 人类

<400> 99

Lys Lys His Cys Ser Tyr Gln Asp Ile Leu

1 5 10

<210> 100

<211> 10

<212> PRT

<213> 小鼠

<400> 100

Arg Arg Arg Ser Ala Tyr Gln Asp Ile Arg

1 5 10

<210> 101

<211> 11

<212> PRT

<213> 大鼠

<400> 101

Arg Lys Arg Arg Arg Ser Tyr Gln Asp Ile Met

1 5 10

<210> 102

<211> 13

<212> PRT

<213> 大鼠

<400> 102

Lys Phe Cys Lys Ser Lys Glu Arg Asn Tyr His Thr Leu

1 5 10

<210> 103

<211> 14

<212> PRT

<213> 果蝇

<400> 103

Lys Phe Tyr Lys Ala Arg Asn Glu Arg Asn Tyr His Thr Leu

1 5 10

<210> 104

<211> 14

<212> PRT

<213> 人类

<400> 104

Lys Ile Arg Leu Arg Cys Gln Ser Ser Gly Tyr Gln Arg Ile

1 5 10

<210> 105

<211> 14

<212> PRT

<213> 小鼠

<400> 105

Lys Ile Arg Gln Arg His Gln Ser Ser Ala Tyr Gln Arg Ile

1 5 10

<210> 106

<211> 15

<212> PRT

<213> 人类

<400> 106

His Phe Cys Leu Tyr Arg Lys Arg Pro Gly Tyr Asp Gln Leu Asn

1 5 10 15

<210> 107

<211> 11

<212> PRT

<213> 人类

<400> 107

Ser Leu Ser Arg Gly Ser Gly Tyr Lys Glu Ile

1 5 10

<210> 108

<211> 13

<212> PRT

<213> 人类

<400> 108

Arg Arg Leu Arg Lys Gly Tyr Thr Pro Leu Met Glu Thr

1 5 10

<210> 109

<211> 14

<212> PRT

<213> 人类

<400> 109

Arg Arg Ala Gly His Ser Ser Tyr Thr Pro Leu Pro Gly Ser

1 5 10

<210> 110

<211> 18

<212> PRT

<213> Dictyostelium discoideum

<400> 110

Lys Lys Leu Arg Gln Gln Lys Gln Gln Gly Tyr Gln Ala Ile Ile Asn

1 5 10 15

Asn Glu

<210> 111

<211> 15

<212> PRT

<213> Dictyostelium discoideum

<400> 111

Arg Ser Lys Ser Asn Gln Asn Gln Ser Tyr Asn Leu Ile Gln Leu

1 5 10 15

<210> 112

<211> 17

<212> PRT

<213> Dictyostelium discoideum

<400> 112

Arg Lys Thr Phe Tyr Asn Asn Asn Gln Tyr Asn Gly Tyr Asn Ile Ile

1 5 10 15

Asn

<210> 113

<211> 10

<212> PRT

<213> 人类

<400> 113

Pro Leu Ser Tyr Thr Arg Phe Ser Leu Ala

1 5 10

<210> 114

<211> 11

<212> PRT

<213> 人类

<400> 114

Met Thr Lys Glu Tyr Gln Asp Leu Gln His Leu

1 5 10

<210> 115

<211> 10

<212> PRT

<213> 人类

<400> 115

Ser Tyr Lys Tyr Ser Lys Val Asn Lys Glu

1 5 10

<210> 116

<211> 10

<212> PRT

<213> 人类

<400> 116

Pro Ala Ala Tyr Arg Gly Val Gly Asp Asp

1 5 10

<210> 117

<211> 10

<212> PRT

<213> 人类

<400> 117

Thr Gly Val Tyr Val Lys Met Pro Pro Thr

1 5 10

<210> 118

<211> 10

<212> PRT

<213> 人类

<400> 118

Leu Ile Ser Tyr Lys Gly Leu Pro Pro Glu

1 5 10

<210> 119

<211> 11

<212> PRT

<213> 大鼠

<400> 119

Ala Ser Asp Tyr Gln Arg Leu Asn Leu Lys Leu

1 5 10

<210> 120

<211> 11

<212> PRT

<213> 人免疫缺陷病毒1型

<400> 120

Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr

1 5 10

<210> 121

<211> 19

<212> PRT

<213> 人类

<400> 121

Arg Met Gln Ala Gln Pro Pro Gly Tyr Arg His Val Ala Asp Gly Glu

1 5 10 15

Asp His Ala

<210> 122  
<211> 9  
<212> PRT  
<213> 人工序列  
<220>  
<223> 基于二亮氨酸的分选信号  
<220>  
<221> misc\_feature  
<222> (3) .. (5)  
<223> Xaa 可以是任何天然发生的氨基酸  
<400> 122  
Asp Glu Xaa Xaa Xaa Leu Leu Leu Ile  
1 5  
<210> 123  
<211> 9  
<212> PRT  
<213> 人类  
<400> 123  
Ser Asp Lys Gln Thr Leu Leu Pro Asn  
1 5  
<210> 124  
<211> 9  
<212> PRT  
<213> 大鼠  
<400> 124  
Asp Glu Arg Ala Pro Leu Ile Arg Thr  
1 5  
<210> 125  
<211> 9  
<212> PRT  
<213> 人类  
<400> 125  
Gln Glu Lys Asp Pro Leu Leu Lys Asn  
1 5  
<210> 126  
<211> 9  
<212> PRT  
<213> 鹌鹑  
<400> 126

Thr Glu Arg Asn Pro Leu Leu Lys Ser

1 5

<210> 127

<211> 9

<212> PRT

<213> 人类

<400> 127

Gly Glu Asn Ser Pro Leu Leu Ser Gly

1 5

<210> 128

<211> 9

<212> PRT

<213> 人类

<400> 128

Glu Glu Lys Gln Pro Leu Leu Met Glu

1 5

<210> 129

<211> 9

<212> PRT

<213> 青鳉

<400> 129

Gly Glu Arg Gln Pro Leu Leu Gln Ser

1 5

<210> 130

<211> 9

<212> PRT

<213> 原鸡

<400> 130

Pro Glu Ile Gln Pro Leu Leu Thr Glu

1 5

<210> 131

<211> 9

<212> PRT

<213> 鲫鱼

<400> 131

Glu Gly Arg Gln Pro Leu Leu Gly Asp

1 5

<210> 132

<211> 9

<212> PRT

<213> 人类

<400> 132

Glu Ala Asn Gln Pro Leu Leu Thr Asp

1 5

<210> 133

<211> 9

<212> PRT

<213> 原鸡

<400> 133

Glu Leu His Gln Pro Leu Leu Thr Asp

1 5

<210> 134

<211> 9

<212> PRT

<213> 斑马鱼

<400> 134

Arg Glu Phe Glu Pro Leu Leu Asn Ala

1 5

<210> 135

<211> 9

<212> PRT

<213> 人类

<400> 135

Glu Glu Lys Met Ala Ile Leu Met Asp

1 5

<210> 136

<211> 9

<212> PRT

<213> 人类

<400> 136

Glu Glu Lys Leu Ala Ile Leu Ser Gln

1 5

<210> 137

<211> 9

<212> PRT

<213> 小鼠

<400> 137

Ser Glu Arg Asp Val Leu Leu Asp Glu

---

1 5  
<210> 138  
<211> 9  
<212> PRT  
<213> 人类  
<400> 138  
Ser Glu Arg Arg Asn Leu Leu Glu Asp  
1 5  
<210> 139  
<211> 9  
<212> PRT  
<213> 大鼠  
<400> 139  
Asp Asp Ser Gly Asp Leu Leu Pro Gly  
1 5  
<210> 140  
<211> 9  
<212> PRT  
<213> 人类  
<400> 140  
Ser Gln Ile Lys Arg Leu Leu Ser Glu  
1 5  
<210> 141  
<211> 9  
<212> PRT  
<213> 猫  
<400> 141  
Ser His Ile Lys Arg Leu Leu Ser Glu  
1 5  
<210> 142  
<211> 9  
<212> PRT  
<213> 小鼠  
<400> 142  
Arg Arg Thr Pro Ser Leu Leu Glu Gln  
1 5  
<210> 143  
<211> 9  
<212> PRT

<213> 人类

<400> 143

His Arg Thr Pro Ser Leu Leu Glu Gln

1 5

<210> 144

<211> 10

<212> PRT

<213> 大鼠

<400> 144

Glu Pro Arg Gly Ser Arg Leu Leu Val Arg

1 5 10

<210> 145

<211> 19

<212> PRT

<213> 人类

<400> 145

Met Asp Asp Gln Arg Asp Leu Ile Ser Asn Asn Glu Gln Leu Pro Met

1 5 10 15

Leu Gly Arg

<210> 146

<211> 19

<212> PRT

<213> 小鼠

<400> 146

Met Asp Asp Gln Arg Asp Leu Ile Ser Asn His Glu Gln Leu Pro Ile

1 5 10 15

Leu Gly Asn

<210> 147

<211> 20

<212> PRT

<213> 原鸡

<400> 147

Met Ala Glu Glu Gln Arg Asp Leu Ile Ser Ser Asp Gly Ser Ser Gly

1 5 10 15

Val Leu Pro Ile

20

<210> 148

<211> 23

<212> PRT

<213> 斑马鱼

<400> 148

Met Glu Pro Asp His Gln Asn Glu Ser Leu Ile Gln Arg Val Pro Ser

1 5 10 15

Ala Glu Thr Ile Leu Gly Arg

20

<210> 149

<211> 23

<212> PRT

<213> 斑马鱼

<400> 149

Met Ser Ser Glu Gly Asn Glu Thr Pro Leu Ile Ser Asp Gln Ser Ser

1 5 10 15

Val Asn Met Gly Pro Gln Pro

20

<210> 150

<211> 24

<212> PRT

<213> 布氏锥虫

<400> 150

Arg Pro Arg Arg Arg Thr Glu Glu Asp Glu Leu Leu Pro Glu Glu Ala

1 5 10 15

Glu Gly Leu Ile Asp Pro Gln Asn

20

<210> 151

<211> 19

<212> PRT

<213> 人类

<400> 151

Pro Asp Lys His Ser Leu Leu Val Gly Asp Phe Arg Glu Asp Asp Asp

1 5 10 15

Thr Ala Leu

<210> 152

<211> 9

<212> PRT

<213> 人类

<400> 152

Thr Glu Arg Glu Arg Leu Leu Asn Phe

1 5

<210> 153

<211> 8

<212> PRT

<213> 人类

<400> 153

Val Glu Thr Asp Asp Leu Ile Leu

1 5

<210> 154

<211> 7

<212> PRT

<213> 秀丽隐杆线虫

<400> 154

Phe Glu Asn Asp Ser Leu Leu

1 5

<210> 155

<211> 9

<212> PRT

<213> 酿酒酵母

<400> 155

Asn Glu Gln Ser Pro Leu Leu His Asn

1 5

<210> 156

<211> 8

<212> PRT

<213> 酿酒酵母

<400> 156

Ser Glu Gln Thr Arg Leu Val Pro

1 5

<210> 157

<211> 8

<212> PRT

<213> 酿酒酵母

<400> 157

Glu Val Asp Leu Asp Leu Leu Lys

1 5

<210> 158

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 基于二亮氨酸的分选信号

<220>

<221> misc\_feature

<222> (2) .. (3)

<223> Xaa 可以是任何天然发生的氨基酸

<400> 158

Asp Xaa Xaa Leu Leu

1 5

<210> 159

<211> 13

<212> PRT

<213> 人类

<400> 159

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu His Ile

1 5 10

<210> 160

<211> 13

<212> PRT

<213> 牛

<400> 160

Thr Phe His Asp Asp Ser Asp Glu Asp Leu Leu His Val

1 5 10

<210> 161

<211> 13

<212> PRT

<213> 穴兔

<400> 161

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu Asn Ile

1 5 10

<210> 162

<211> 13

<212> PRT

<213> 原鸡

<400> 162

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu Asn Val

1 5 10

<210> 163

<211> 13

<212> PRT

<213> 人类

<400> 163

Glu Glu Ser Glu Glu Arg Asp Asp His Leu Leu Pro Met

1 5 10

<210> 164

<211> 13

<212> PRT

<213> 原鸡

<400> 164

Asp Glu Ser Glu Glu Arg Asp Asp His Leu Leu Pro Met

1 5 10

<210> 165

<211> 12

<212> PRT

<213> 人类

<400> 165

Gly Tyr His Asp Asp Ser Asp Glu Asp Leu Leu Glu

1 5 10

<210> 166

<211> 13

<212> PRT

<213> 人类

<400> 166

Ile Thr Gly Phe Ser Asp Asp Val Pro Met Val Ile Ala

1 5 10

<210> 167

<211> 13

<212> PRT

<213> *Hydra vulgaris*

<400> 167

Ile Asn Arg Phe Ser Asp Asp Glu Pro Leu Val Val Ala

1 5 10

<210> 168

<211> 13

<212> PRT

<213> 人类

<400> 168

Met Leu Glu Ala Ser Asp Asp Glu Ala Leu Leu Val Cys

---

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 169  
<211> 13  
<212> PRT  
<213> 人类  
<400> 169  
Lys Asn Glu Thr Ser Asp Asp Glu Ala Leu Leu Leu Cys

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 170  
<211> 12  
<212> PRT  
<213> 小鼠  
<400> 170  
Trp Val Val Glu Ala Glu Asp Glu Pro Leu Leu Ala

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 171  
<211> 12  
<212> PRT  
<213> 人类  
<400> 171  
Trp Val Ala Glu Ala Glu Asp Glu Pro Leu Leu Thr

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 172  
<211> 12  
<212> PRT  
<213> 人类  
<400> 172  
His Asp Asp Phe Ala Asp Asp Ile Ser Leu Leu Lys

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 173  
<211> 13  
<212> PRT  
<213> 小鼠  
<400> 173  
Gly Arg Asp Ser Pro Glu Asp His Ser Leu Leu Val Asn

|   |   |    |
|---|---|----|
| 1 | 5 | 10 |
|---|---|----|

<210> 174  
<211> 12  
<212> PRT

<213> *Peromyscus maniculatus*

<400> 174

Val Arg Cys His Pro Glu Asp Asp Arg Leu Leu Gly

1 5 10

<210> 175

<211> 13

<212> PRT

<213> 人类

<400> 175

His Arg Val Ser Gln Asp Asp Leu Asp Leu Leu Thr Ser

1 5 10

<210> 176

<211> 13

<212> PRT

<213> 人类

<400> 176

Ala Ser Val Ser Leu Leu Asp Asp Glu Leu Met Ser Leu

1 5 10

<210> 177

<211> 13

<212> PRT

<213> 人类

<400> 177

Ala Ser Ser Gly Leu Asp Asp Leu Asp Leu Leu Gly Lys

1 5 10

<210> 178

<211> 13

<212> PRT

<213> 人类

<400> 178

Val Gln Asn Pro Ser Ala Asp Arg Asn Leu Leu Asp Leu

1 5 10

<210> 179

<211> 13

<212> PRT

<213> 人类

<400> 179

Asn Ala Leu Ser Trp Leu Asp Glu Glu Leu Leu Cys Leu

1 5 10

<210> 180

<211> 13

<212> PRT

<213> 果蝇

<400> 180

Thr Val Asp Ser Ile Asp Asp Val Pro Leu Leu Ser Asp

1 5 10

<210> 181

<211> 13

<212> PRT

<213> 小鼠

<400> 181

Gln Glu Glu Cys Pro Ser Asp Ser Glu Glu Asp Glu Gly

1 5 10

<210> 182

<211> 13

<212> PRT

<213> 小鼠

<400> 182

Arg Asp Arg Asp Tyr Asp Glu Asp Asp Glu Asp Asp Ile

1 5 10

<210> 183

<211> 15

<212> PRT

<213> 小鼠

<400> 183

Leu Asp Glu Thr Glu Asp Asp Glu Leu Glu Tyr Asp Asp Glu Ser

1 5 10 15

<210> 184

<211> 10

<212> PRT

<213> 人类

<400> 184

Lys Asp Pro Asp Glu Val Glu Thr Glu Ser

1 5 10

<210> 185

<211> 13

<212> PRT

<213> 人类

<400> 185

His Glu Phe Gln Asp Glu Thr Asp Thr Glu Glu Glu Thr

1 5 10

<210> 186

<211> 14

<212> PRT

<213> 人类

<400> 186

Gln Glu Lys Glu Asp Asp Gly Ser Glu Ser Glu Glu Glu Tyr

1 5 10

<210> 187

<211> 9

<212> PRT

<213> 人类

<400> 187

Gly Glu Asp Glu Glu Ser Glu Ser Asp

1 5

<210> 188

<211> 12

<212> PRT

<213> 人类

<400> 188

Gly Glu Asp Ser Asp Glu Glu Pro Asp His Glu Glu

1 5 10

<210> 189

<211> 11

<212> PRT

<213> 人类

<400> 189

Leu Glu Asp Asp Ser Asp Glu Glu Glu Asp Phe

1 5 10

<210> 190

<211> 9

<212> PRT

<213> 人巨细胞病毒

<400> 190

Lys Asp Ser Asp Glu Glu Glu Asn Val

1 5

<210> 191

<211> 13

<212> PRT

<213> 人疱疹病毒3

<400> 191

Phe Glu Asp Ser Glu Ser Thr Asp Thr Glu Glu Glu Phe

1 5 10

<210> 192

<211> 9

<212> PRT

<213> 人免疫缺陷病毒1型

<400> 192

Leu Glu Ala Gln Glu Glu Glu Val

1 5

<210> 193

<211> 22

<212> PRT

<213> 酿酒酵母

<400> 193

Ala Asp Asp Leu Glu Ser Gly Leu Gly Ala Glu Asp Asp Leu Glu Gln

1 5 10 15

Asp Glu Gln Leu Glu Gly

20

<210> 194

<211> 12

<212> PRT

<213> 酿酒酵母

<400> 194

Thr Glu Ile Asp Glu Ser Phe Glu Met Thr Asp Phe

1 5 10

<210> 195

<211> 18

<212> PRT

<213> 酿酒酵母

<400> 195

Thr Glu Pro Glu Glu Val Glu Asp Phe Asp Phe Asp Leu Ser Asp Glu

1 5 10 15

Asp His

<210> 196

<211> 15

<212> PRT

<213> 酿酒酵母

<400> 196

Phe Glu Ile Glu Glu Asp Asp Val Pro Thr Leu Glu Glu Glu His

1

5

10

15

<210> 197

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 细胞内保留信号

<400> 197

His Asp Glu Leu

1

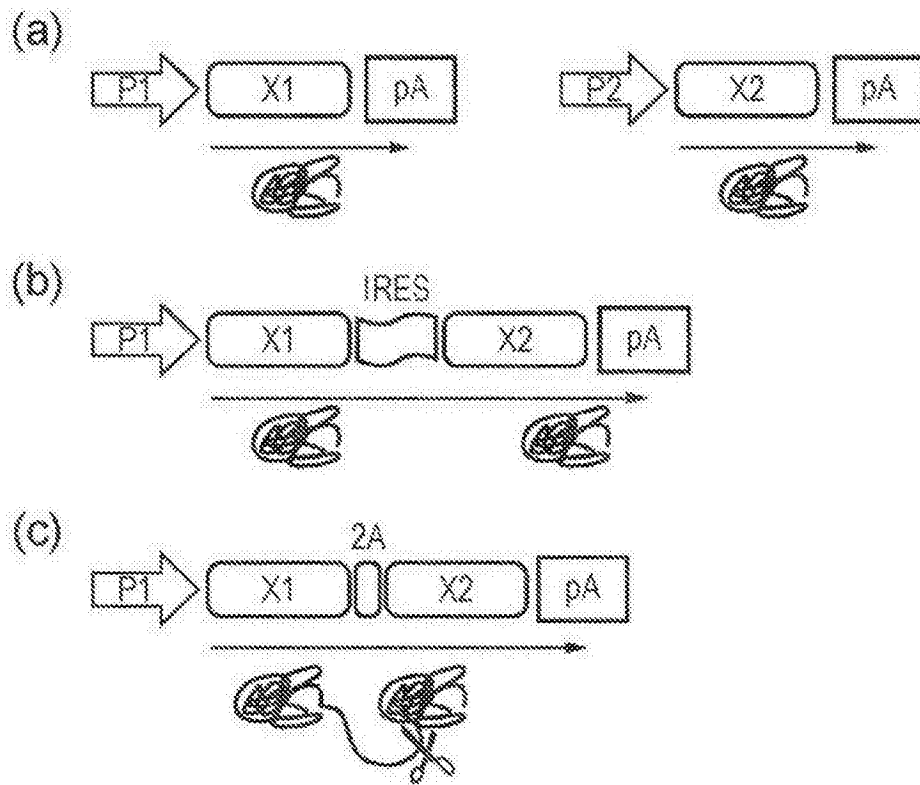


图1

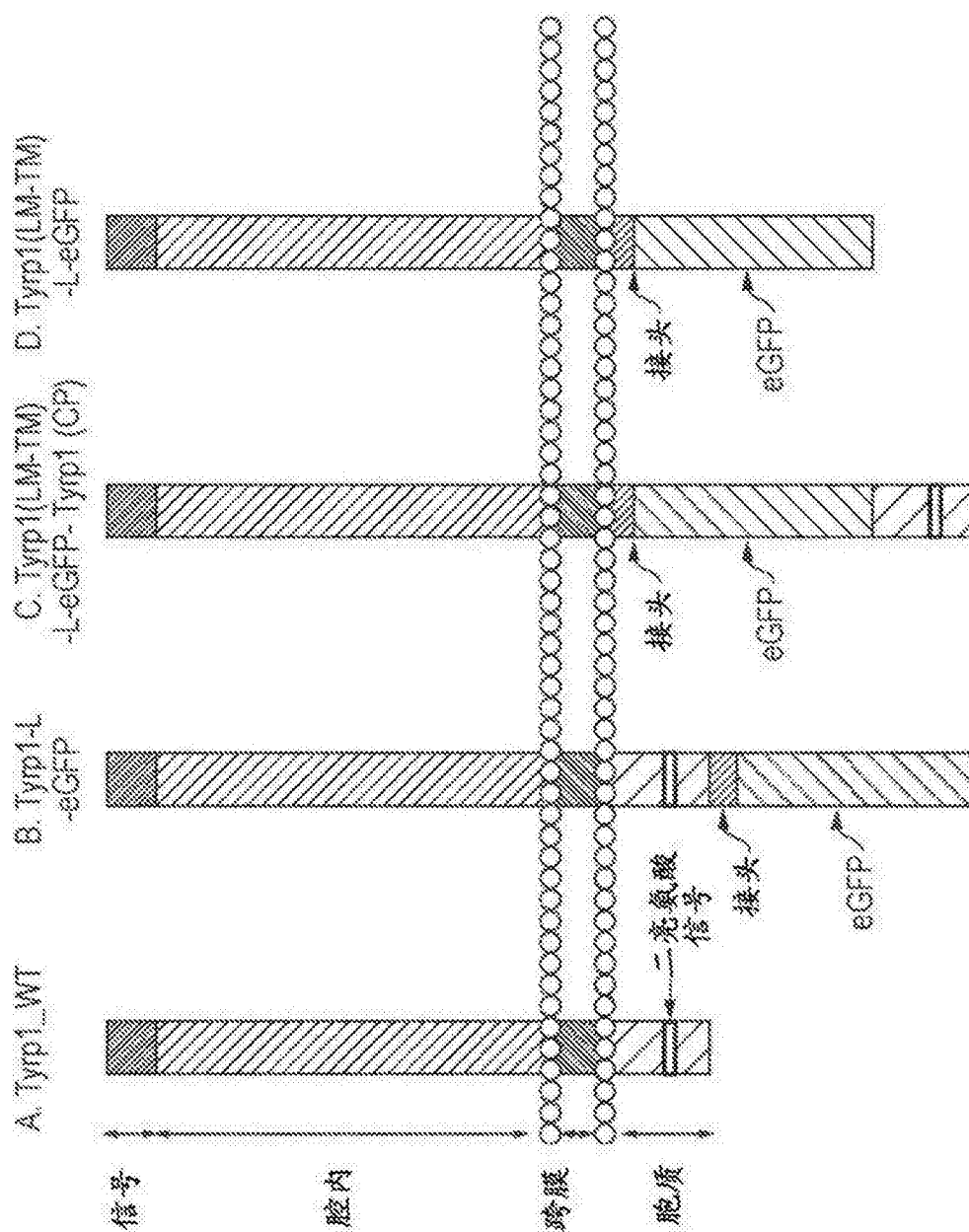


图2

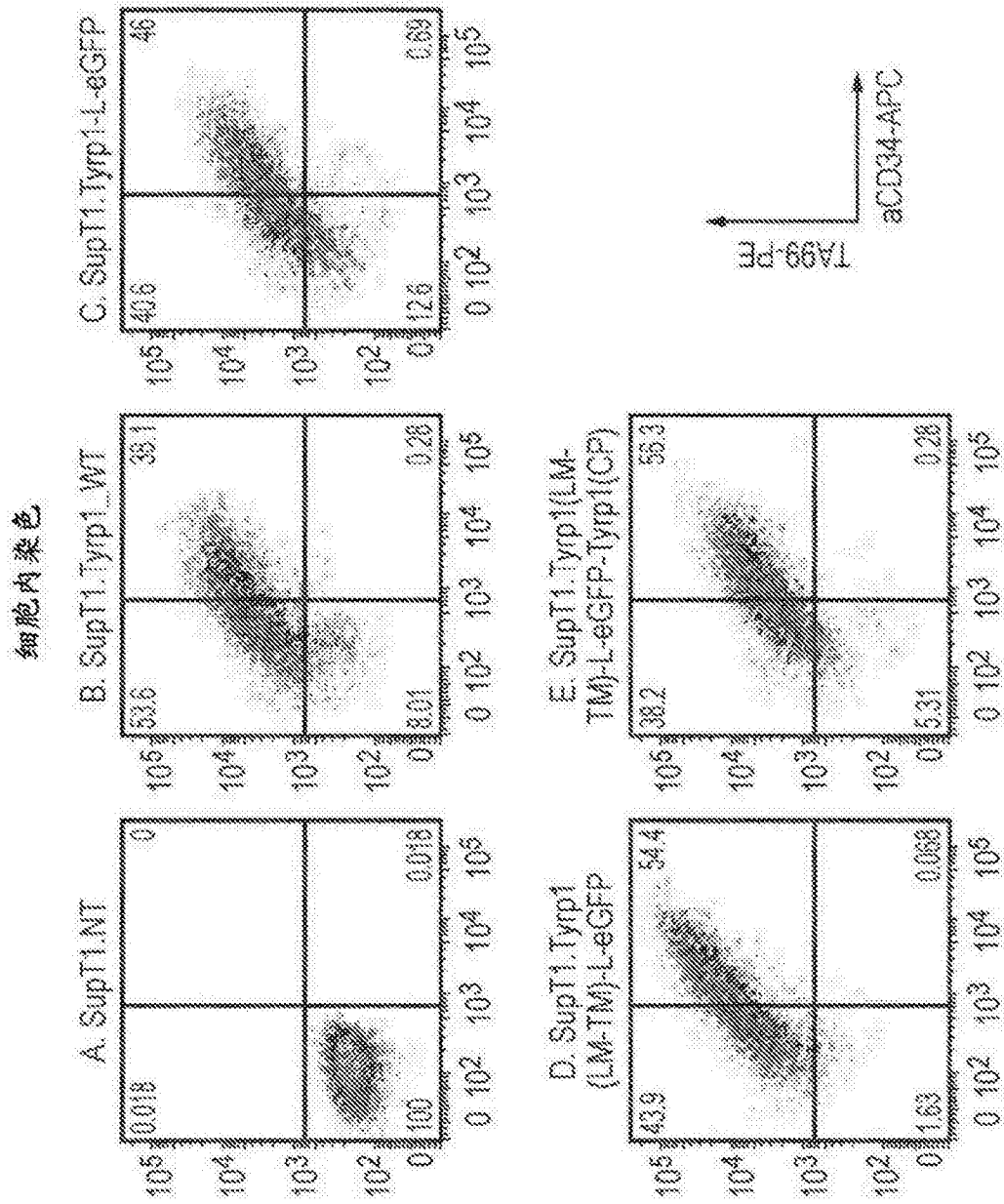


图2(续)

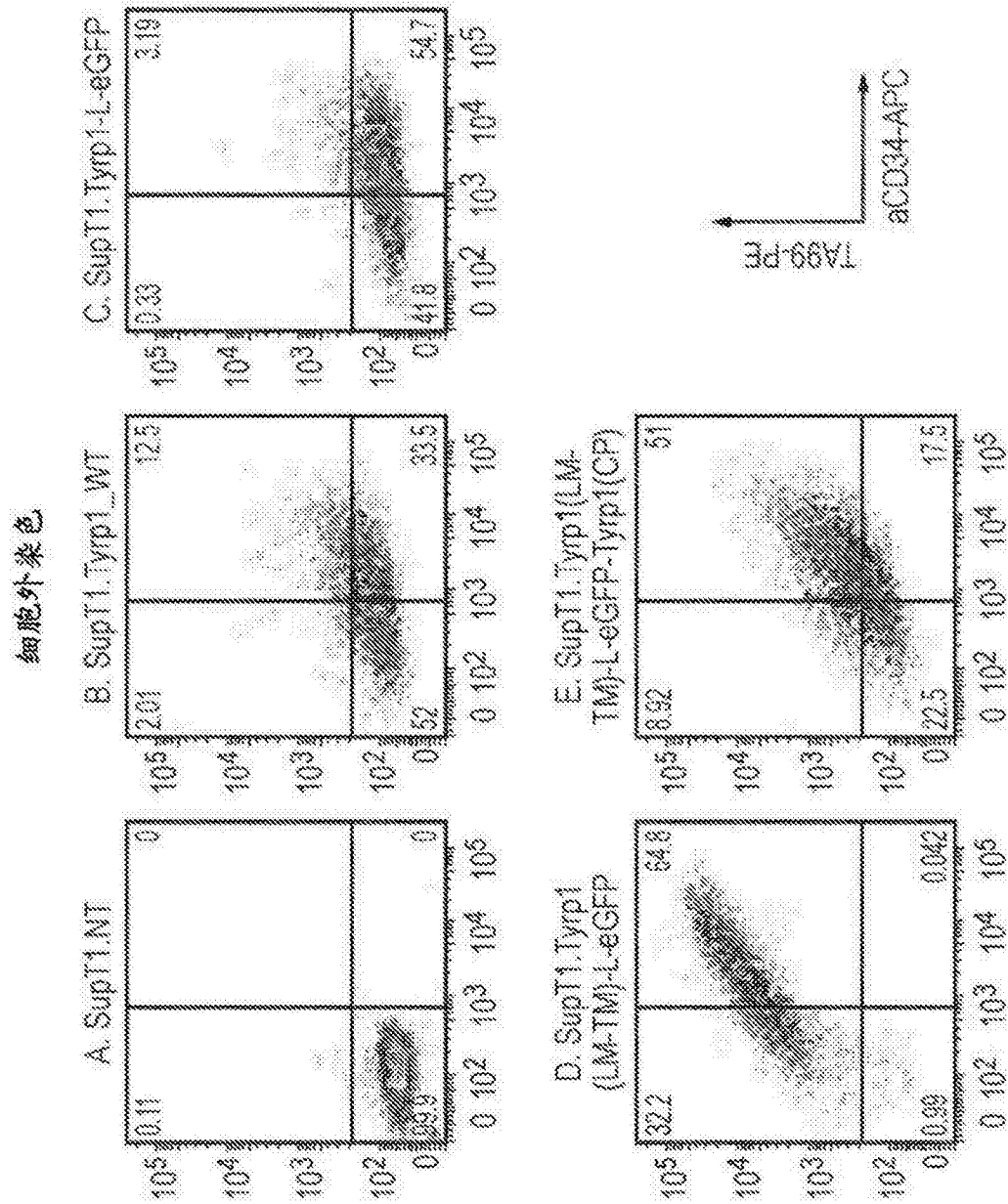


图2(续)

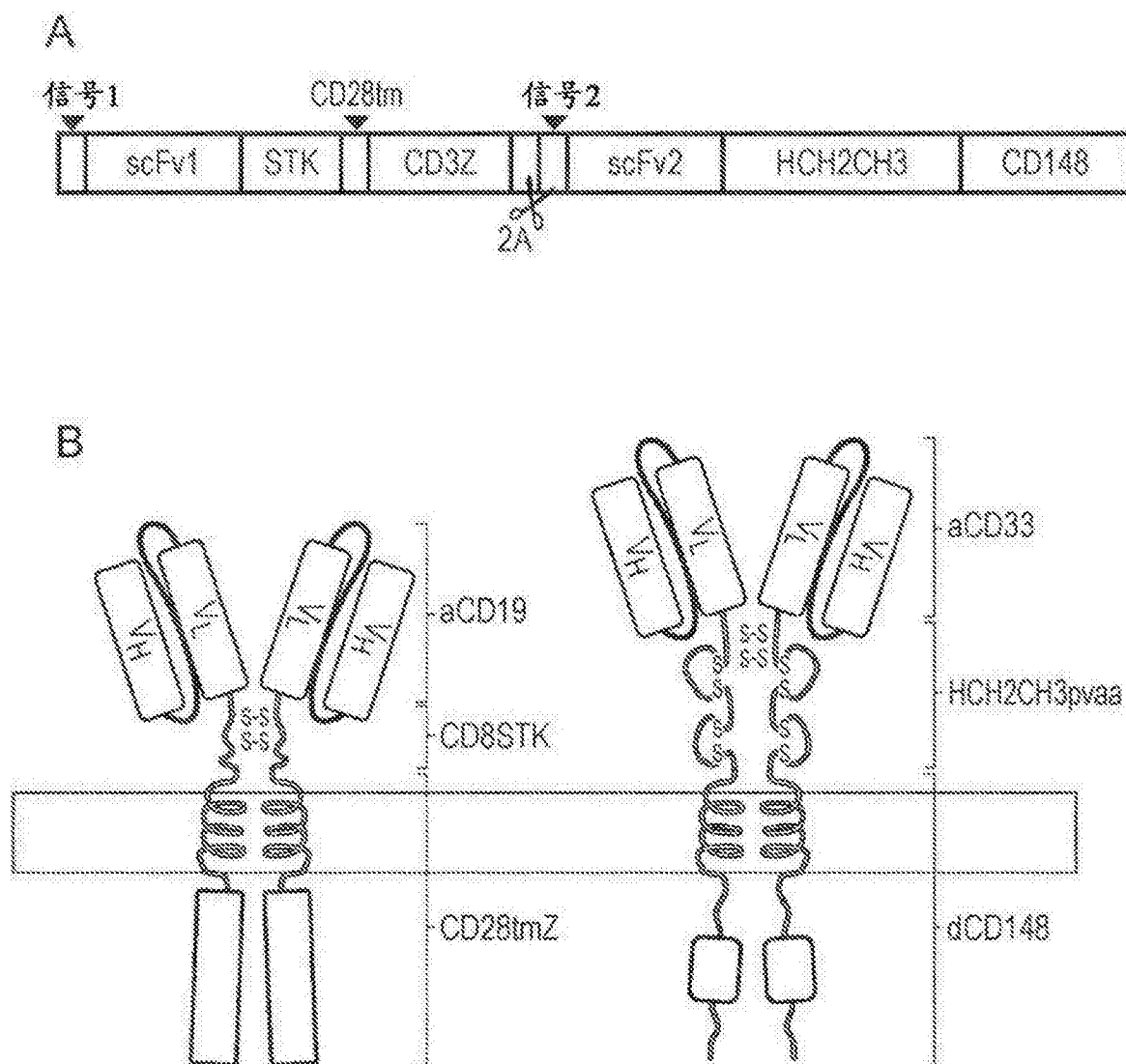


图3

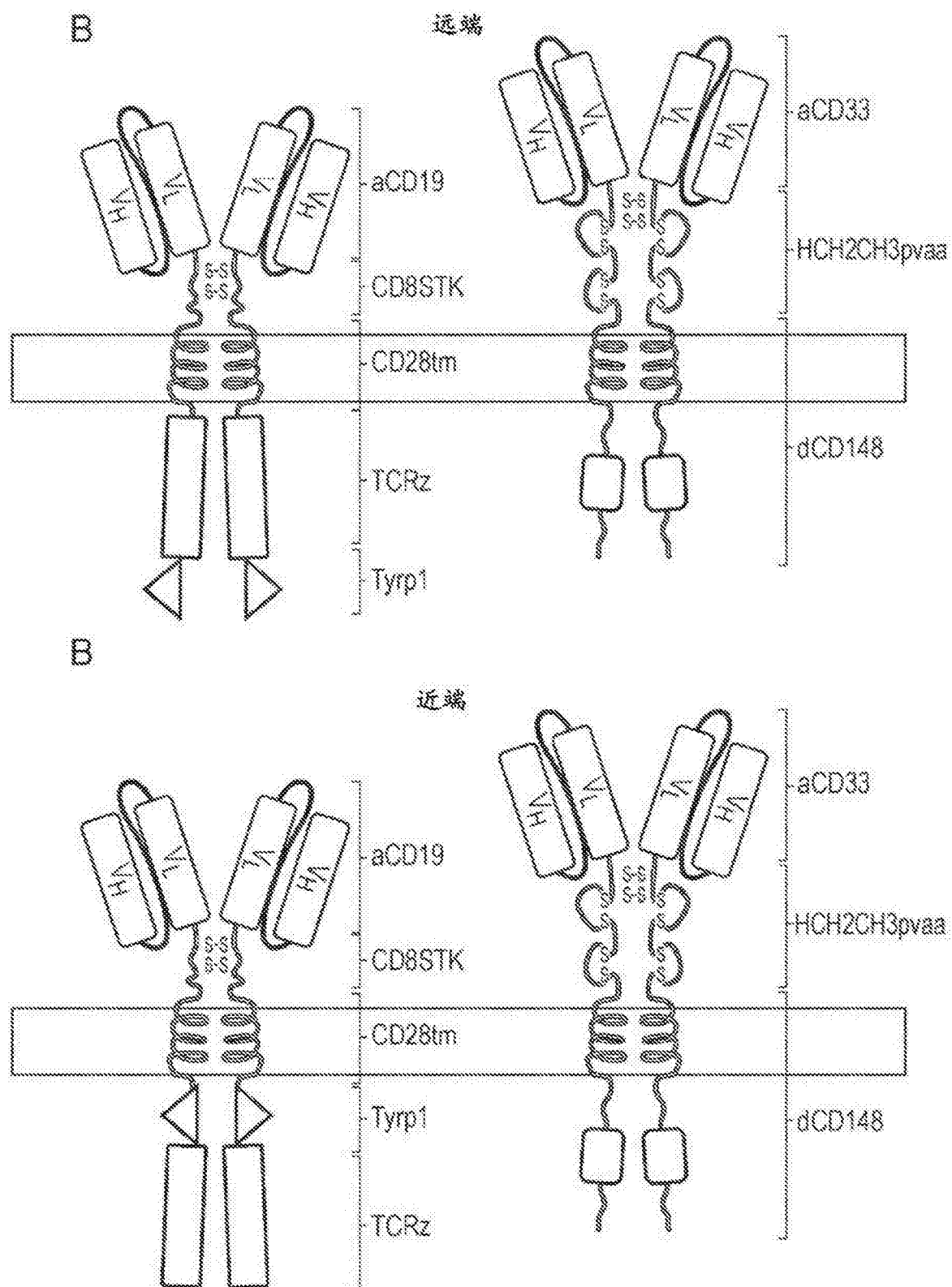


图3(续)

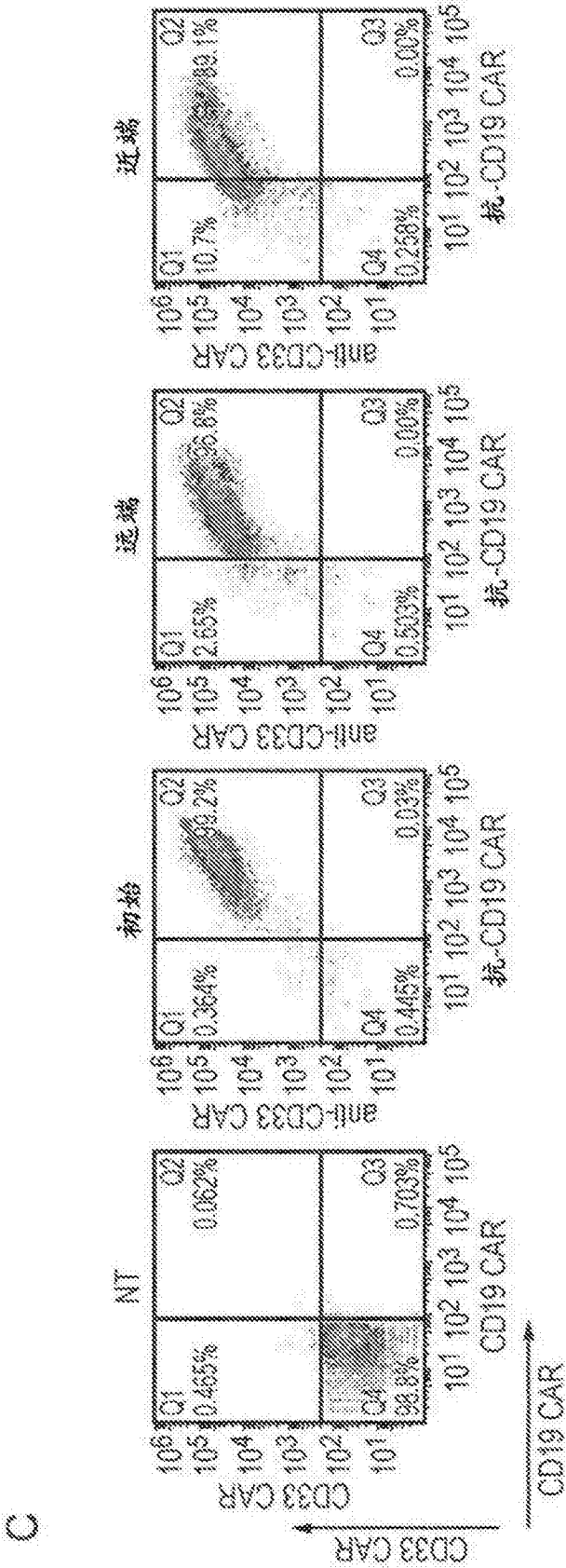


图3(续)

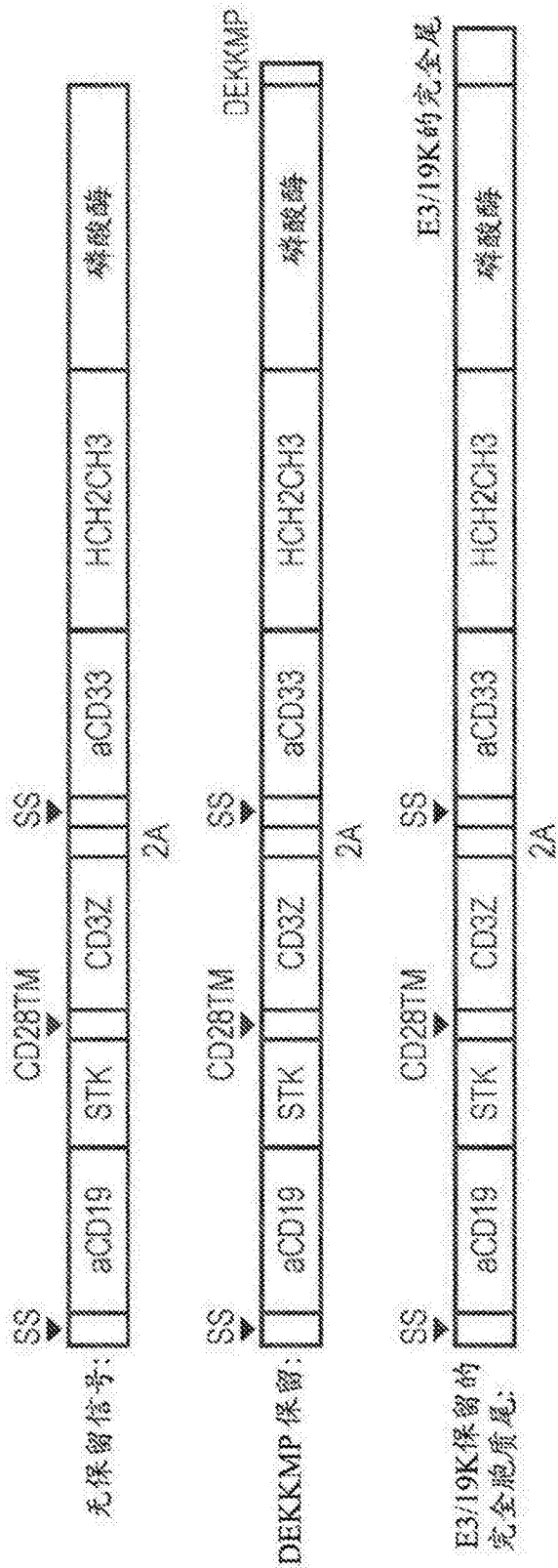


图4

