



(12) 发明专利申请

(10) 申请公布号 CN 115698296 A

(43) 申请公布日 2023. 02. 03

(21) 申请号 202180037500.0

(22) 申请日 2021.03.31

(30) 优先权数据

2020-065065 2020.03.31 JP

(85) PCT国际申请进入国家阶段日

2022.11.23

(86) PCT国际申请的申请数据

PCT/JP2021/014096 2021.03.31

(87) PCT国际申请的公布数据

W02021/201198 JA 2021.10.07

(71) 申请人 日商基因编辑力股份有限公司

地址 日本福冈县

(72) 发明人 一濑瑞穗 B·古曼 八木祐介

赤岩友美 岛尻恭香 中村崇裕

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

专利代理师 于洁 褚瑶杨

(51) Int.Cl.

C12N 15/52 (2006.01)

C07K 14/415 (2006.01)

C12N 1/15 (2006.01)

C12N 1/19 (2006.01)

C12N 1/21 (2006.01)

C12N 5/10 (2006.01)

C12N 15/29 (2006.01)

权利要求书1页 说明书31页

序列表69页 附图12页

(54) 发明名称

编辑靶标RNA的方法

(57) 摘要

本发明提供一种将靶标RNA中包含的编辑靶标C转换成U、或者将编辑靶标U转换成C的方法。一种编辑靶标RNA的方法,其将包含由下述a、b、c和bc中的任一多肽构成的DYW结构域的人工DYW蛋白应用于靶标RNA:a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;bc.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽(序列中,x表示任意的氨基酸,⋯表示任意的多肽片段)。

CN 115698296 A

1. 一种编辑靶标RNA的方法, 其将包含由下述a、b、c和bc中的任一多肽构成的DYW结构域的人工DYW蛋白应用于靶标RNA,

a. 具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

b. 具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

c. 具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

bc. 具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽,

序列中, x表示任意的氨基酸, \cdots 表示任意的多肽片段。

2. 一种DYW结构域, 其由下述a、b、c和bc中的任一多肽构成,

a. 具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

b. 具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

c. 具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

bc. 具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽。

3. 一种DYW蛋白, 其包含: 包含至少1个PPR基序、能够与靶标RNA序列特异性结合的RNA结合结构域; 以及权利要求2所述的DYW结构域。

4. 如权利要求2所述的DYW蛋白, 其中, RNA结合结构域为PLS型。

5. 一种编辑靶标RNA的方法, 其包括将由下述c或bc的多肽构成的DYW结构域应用于靶标RNA从而将编辑靶标U转换成C的步骤,

c. 具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽;

bc. 具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽。

6. 如权利要求5所述的方法, 其中, DYW结构域与RNA结合结构域融合, 该RNA结合结构域包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA序列特异性地结合。

7. 一种用于在真核细胞中编辑RNA的组合物, 其包含权利要求2所述的DYW结构域。

8. 一种核酸, 其编码权利要求2所述的DYW结构域、或者权利要求3或4所述的DYW蛋白。

9. 一种载体, 其包含权利要求8所述的核酸。

10. 一种细胞, 其包含权利要求9所述的载体, 所述细胞不包括人个体的细胞。

编辑靶标RNA的方法

技术领域

[0001] 本发明涉及使用能够与靶标RNA结合的蛋白质的RNA编辑技术。本发明在医疗(创新药物开发、治疗)、农业(农水畜产品生产、育种)、化学(生物学物质生产)等广泛的领域中有用。

背景技术

[0002] 在植物的线粒体和叶绿体中,基因组中的特定碱基频繁地发生在RNA水平上进行置换的RNA编辑,已知该现象是由作为RNA结合蛋白的三角状五肽重复(PPR)蛋白介导的。

[0003] PPR蛋白根据构成该蛋白质的PPR基序的结构被分类成P和PLS这2个家族(非专利文献1)。P类PPR蛋白由标准的35个氨基酸的PPR基序(P)的单纯重复构成,另一方面,PLS蛋白除了包含P以外,还包含与其类似的被称为L和S的2个基序。PLS蛋白的PPR阵列(PPR基序的排列)中,P1(约35个氨基酸)、L1(约35个氨基酸)以及S1(约31个氨基酸)这3个PPR基序构成P-L-S重复单元,在该P1L1S1的C末端侧接续至少序列不同的P-L-S、即P2(35个氨基酸)、L2(36个氨基酸)以及S2(32个氨基酸)基序。另外,与由P-L-S重复单元构成的蛋白不同,有时也存在SS(31个氨基酸)重复的情况。此外,有时在该最后的P2L2S2基序的C末端侧接续被称为E1和E2的2个PPR样基序、以及具有136个氨基酸的胞苷脱氨酶结构域样序列的DYW结构域(非专利文献2)。

[0004] PPR蛋白与RNA的相互作用由PPR密码来规定,该PPR密码通过各PPR基序中的几处氨基酸的组合来指定结合RNA碱基,这些氨基酸藉由氢键与对应的核苷酸结合(专利文献1、专利文献2、非专利文献3、非专利文献4、非专利文献5、非专利文献6、非专利文献7、非专利文献8)。

[0005] 现有技术文献

[0006] 专利文献

[0007] 专利文献1:国际公开W02013/058404

[0008] 专利文献2:国际公开W02014/175284

[0009] 非专利文献

[0010] 非专利文献1:Lurin,C.,Andres,C.,Aubourg,S.,Bellaoui,M.,Bitton,F.,Bruyere,C.,Caboche,M.,Debast,C.,Gualberto,J.,Hoffmann,B.,et al.(2004).Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis.Plant Cell 16:2089-2103.

[0011] 非专利文献2:Cheng,S.,Gutmann,B.,Zhong,X.,Ye,Y.,Fisher,M.F.,Bai,F.,Castleden,I.,Song,Y.,Song,B.,Huang,J.,et al.(2016).Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants.Plant J.85:532-547.

[0012] 非专利文献3:Barkan,A.,Rojas,M.,Fujii,S.,Yap,A.,Chong,Y.S.,Bond,C.S.,and Small,I.(2012).A combinatorial amino acid code for RNA recognition by

pentatricopeptide repeat proteins. *PLoS Genet.* 8:e1002910.

[0013] 非专利文献4: Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., Wang, X., Wang, Q., Zhang, Q., Fan, S., et al. (2016). Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun.* 7:11285.

[0014] 非专利文献5: Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8:e57286.

[0015] 非专利文献6: Kobayashi, T., Yagi, Y., and Nakamura, T. (2019). Comprehensive Prediction of Target RNA Editing Sites for PLS-Class PPR Proteins in *Arabidopsis thaliana*. *Plant Cell Physiol.* 60:862-874.

[0016] 非专利文献7: Yan, J., Yao, Y., Hong, S., Yang, Y., Shen, C., Zhang, Q., Zhang, D., Zou, T., and Yin, P. (2019). Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res.* 47:3728-3738.

[0017] 非专利文献8: Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8:e65343.

[0018] 非专利文献9: Knie, N., Grewe, F., Fischer, S., and Knoop, V. (2016). Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns—a monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. *BMC Evol. Biol.* 16:134.

[0019] 非专利文献10: Gerke, P., Szovenyi, P., Neubauer, A., Lenz, H., Gutmann, B., McDowell, R., Small, I., Schallenberg-Rudinger, M., and Knoop, V. (2020). Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort *Anthoceros agrestis*. *New Phytologist* 225:1974-1992.

[0020] 非专利文献11: Gutmann B., Royan S., Schallenberg-Rudinger M., Lenz H., Castleden I.R., McDowell R., Vacher M.A., Tonti-Filippini J., Bond C.S., Knoop V., and Small I.D. (2020). The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant* 13:215-230

[0021] 非专利文献12: Oldenkott, B., Yang, Y., Lesch, E., Knoop, V., and Schallenberg-Rudinger, M. (2019). Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Communications Biology* 2:85.

发明内容

[0022] 发明所要解决的课题

[0023] 在陆生植物的细胞器中, RNA碱基通常发生从胞苷替换成尿苷的C-U RNA编辑。但是, 在一部分角苔类(hornworts)、石松类(lycophytes)和蕨类植物(leptosporangiate

monilophytes)中,也观察到尿昔置换成胞昔的U-C RNA编辑(非专利文献9)。根据两种不同的生物信息学研究,在具有U-C RNA编辑的植物中,发现了序列与标准的DYW结构域不同的独特的DYW结构域(非专利文献10、非专利文献11)。

[0024] 据报告,来自小立碗藓(*Physcomitrella patens*)的2种DYW:PG蛋白在大肠杆菌内显示出从C到U的RNA编辑活性(非专利文献12)。在陆生植物的进化初期分支出的植物组中,DYW结构域大致分类成2组。第1组被称为DYW:PG/WW组,其中包括作为标准DYW结构域 DYW:PG型、以及在PG盒上添加有1个色氨酸(W)的DYW:WW型。第2组被称为DYW:KP,该DYW结构域具有PG盒、以及C末端的3个氨基酸序列与PG/WW组不同的序列,此外仅存在于具有U-C RNA编辑的植物中。将任意RNA序列内的1个碱基变为特定碱基的技术在基因治疗、产业利用中的基因突变导入技术中是有用的。迄今为止开发出了各种RNA结合分子,但尚未确立通过使用DYW结构域和人工RNA结合蛋白将任意的靶标RNA的胞昔(C)转换成尿昔(U)、或者反向的将尿昔(U)转换成胞昔(C)的分子的设计方法。

[0025] 用于解决课题的手段

[0026] 本研究中,发明人揭示出,PPR-DYW蛋白的包含C末端DYW结构域的部分可用作RNA编辑的模块结构域(modular editing domain)。将本研究中所设计的3种DYW结构域分别与PPR-P或PLS阵列融合的情况下,DYW:PG和DYW:WW结构域显示出C-U的RNA编辑活性、DYW:KP显示出U-C的RNA编辑活性。

[0027] 本发明提供下述方案。

[0028] [1]一种编辑靶标RNA的方法,其将包含由下述a、b、c和bc中的任一多肽构成的DYW结构域的人工DYW蛋白应用于靶标RNA。

[0029] a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0030] b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0031] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0032] bc.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0033] (序列中,x表示任意的氨基酸,⋯表示任意的多肽片段。)

[0034] [2]一种DYW结构域,其由下述a、b、c和bc中的任一多肽构成。

[0035] a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0036] b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0037] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0038] bc.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0039] [3]一种DYW蛋白,其包含:包含至少1个PPR基序、能够与靶标RNA序列特异性结合

的RNA结合结构域;以及2所述的DYW结构域。

[0040] [4]如2所述的DYW蛋白,其中,RNA结合结构域为PLS型。

[0041] [5]一种编辑靶标RNA的方法,其包括将由下述c或多肽构成的DYW结构域应用于靶标RNA从而将编辑靶标U转换成C的步骤。

[0042] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0043] bc.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0044] [6]如5所述的方法,其中,DYW结构域与RNA结合结构域融合,该RNA结合结构域包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA序列特异性地结合。

[0045] [7]一种用于在真核细胞中编辑RNA的组合物,其包含2所述的DYW结构域。

[0046] [8]一种核酸,其编码2所述的DYW结构域、或者3或4所述的DYW蛋白。

[0047] [9]一种载体,其包含8所述的核酸。

[0048] [10]一种细胞(不包括人个体细胞),其包含9所述的载体。

[0049] [1]一种编辑靶标RNA的方法,其将包含由下述a~c中的任一多肽构成的DYW结构域的人工DYW蛋白应用于靶标RNA。

[0050] a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots HSE\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0051] b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots HSE\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0052] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots HAE\cdots Cx_{c4}x_{c5}CH\cdots Dx_{c6}x_{c7}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0053] (序列中,x表示任意的氨基酸,⋯表示任意的多肽片段。)

[0054] [2]一种DYW结构域,其由下述a~c中的任一多肽构成。

[0055] a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots HSE\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0056] b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots HSE\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0057] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots HAE\cdots Cx_{c4}x_{c5}CH\cdots Dx_{c6}x_{c7}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0058] [3]一种DYW蛋白,其包含:包含至少1个PPR基序、能够与靶标RNA序列特异性结合的RNA结合结构域;以及2所述的DYW结构域。

[0059] [4]如2所述的DYW蛋白,其中,RNA结合结构域为PLS型。

[0060] [5]一种编辑靶标RNA的方法,其包括将由下述c的多肽构成的DYW结构域应用于靶标RNA从而将编辑靶标U转换成C的步骤。

[0061] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots HAE\cdots Cx_{c4}x_{c5}CH\cdots Dx_{c6}x_{c7}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0062] [6]如5所述的方法,其中,DYW结构域与RNA结合结构域融合,该RNA结合结构域包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA序列特异性地结合。

- [0063] [7]一种用于在真核细胞中编辑RNA的组合物,其包含2所述的DYW结构域。
- [0064] [8]一种核酸,其编码2所述的DYW结构域、或者3或4所述的DYW蛋白。
- [0065] [9]一种载体,其包含8所述的核酸。
- [0066] [10]一种细胞(不包括人个体细胞),其包含9所述的载体。
- [0067] 发明的效果
- [0068] 根据本发明,能够将靶标RNA中包含的编辑靶标C转换成U、或者将编辑靶标U转换成C。

附图说明

[0069] 图1是DYW蛋白中的C末端结构域的近似的最大似然系统发育树。使用FastTree制作(a)DYW:PG型、(b)DYW:WW型以及(c)DYW:KP型结构域的系统发育树。DYW:PG结构域基于来源于石松类的DYW:PG蛋白进行设计。将为了设计DYW:WW以及DYW:KP结构域而选择的蛋白质的系统群(进化枝)以黑色的线表示。系统发育树使用iTOL (Letunic, I. and Bork, P. (2016) *Nucleic Acids Res.* 44W242-245) 进行可视化,红色表示来源于角苔类的蛋白质,绿色表示来源于石松类的蛋白质,青色表示来源于蕨类植物的蛋白质。

[0070] 图2是本研究中设计的PPR蛋白的构成。(a) PPR蛋白包含:N末端侧的硫氧还蛋白、His-标签和TEV位点、与其接续的由P或PLS基序构成的PPR阵列、最后的DYW结构域(DYW:PG、DYW:WW或DYW:KP)。PPR蛋白的靶标序列(包含RNA编辑位点)被插入在终止密码子的下游。(b) P、P1、L1、S1和P2基序中的与RNA识别相关的第4位和第*i*位氨基酸。L2、S2、E1和E2的第4位和第*i*位氨基酸与CLB19中的相同(Chateigner-Boutin, A.L., et al. (2008) *Plant J.* 56 590-602.)。

[0071] 图3是使用大肠杆菌的RNA编辑分析。(a) 各PPR-DYW的DNA上的靶标碱基(C或T)。(b) 表示PPR-DYW:PG和WW的C-U RNA编辑活性、(c) 表示PPR-DYW:KP的U-C RNA编辑活性。示出了3次独立进行的实验结果中的1次的结果作为示例。(d) 3次测定的C-U和(e) U-C RNA编辑效率的平均值(其中,P-DYW:KP为2次)。误差条表示标准偏差。

[0072] 图4-1是所选出的各DYW结构域系统群中的氨基酸的出现频率。以通过WebLogo制作的序列标识进行了可视化。图4-2和图4-3中同样。(a) DYW:PG

[0073] [图4-2] (b) DYW:WW

[0074] [图4-3] (c) DYW:KP

[0075] 图5是HEK293T细胞中的RNA编辑活性(C-U或U-C)。是在PLS型中融合PG1或WW1结构域(a)、或者KP1结构域(b)时的靶标位点的测序结果。为3次独立实验中的RNA编辑活性(c)。条状图表示平均,各点表示3次实验中的RNA编辑活性。左侧的条状图表示C-U编辑活性,右侧的条状图表示U-C编辑活性。

[0076] 图6是由结构域交换带来的KP结构域性能的提高。保留KP1的包含活性位点(HxEx_nCxxCH)的中央部分,对于前后的结构域,与WW1结构域进行交换,调查所得到的嵌合KP1a的RNA编辑活性。结构域交换的示意图(a)、实际的RNA编辑活性(b、c)。

[0077] 图7是KP结构域突变体的RNA编辑活性测定。KP1、KP2、KP3、KP4在大肠杆菌中的RNA编辑活性(a)和在HEK293T中的RNA编辑活性(b)、KP5~KP23在HEK293T中的RNA编辑活性(c)。

[0078] 图8是PG结构域突变体的RNA编辑活性测定。PG1、PG2在大肠杆菌中的RNA编辑活性(a)和在HEK293T中的RNA编辑活性(b)。PG3~PG13在HEK293T中的RNA编辑活性

[0079] 图9是WW结构域突变体的RNA编辑活性测定。WW1、WW2在大肠杆菌中的RNA编辑活性(a)和在HEK293T中的RNA编辑活性(b)。WW3~WW14在HEK293T中的RNA编辑活性

[0080] 图10是人线粒体RNA编辑。靶标序列信息(a)。RNA编辑活性(b,c)

具体实施方式

[0081] 本发明涉及一种编辑靶标RNA的方法,其将包含由下述a、b、c和bc中的任一多肽构成的DYW结构域的DYW蛋白应用于靶标RNA。

[0082] a.具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、与序列编号1的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0083] b.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 、与序列编号2的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0084] c.具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0085] bc.具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽

[0086] 关于本发明,C-U/U-C编辑活性是指将对象多肽连结至能够与靶标RNA序列特异性结合的RNA结合结构域的C末端侧而进行编辑分析时,能够将靶标RNA中包含的编辑靶标C转换成U、或者能够将编辑靶标U转换成C的活性。关于转换,只要在适当的条件下编辑靶标碱基的至少约3%、优选约5%被置换成目标碱基即可。

[0087] [DYW结构域]

[0088] $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 、 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 和 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 均表示氨基酸序列。序列中,x各自独立地表示任意的氨基酸, \cdots 各自独立地表示由任意长度的任意的氨基酸序列构成的多肽片段。关于本发明,DYW结构域可以表示这3个氨基酸序列中的任一者。特别地,有时将由 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\cdots Hx_{aa}E\cdots Cx_{a17}x_{a18}CH\cdots DYW$ 构成的DYW结构域表示为DYW:PG、将由 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{bb}E\cdots Cx_{b17}x_{b18}CH\cdots DYW$ 构成的DYW结构域表示为DYW:WW、并且将由 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 构成的DYW结构域表示为DYW:KP。

[0089] DYW结构域具有N末端的包含由约15个氨基酸构成的PG盒的区域、中心部的zinc结合结构域(HxE_xC_xCH 、 x_n 为任意氨基酸的任意数n个的连结)、以及C末端的DYW这三个区域。zinc结合结构域可以进一步分成HxE区域和CxxCH区域。各DYW结构域的这些区域可以如下表所示。

[0090] [表1]

[0091]

	包含PG盒的区域	HxE区域	CxxCH区域	DYW
DYW:PG	$x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP$	$Hx_{aa}E$	$Cx_{a17}x_{a18}CH$	DYW
DYW:WW	$x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP$	$Hx_{bb}E$	$Cx_{b17}x_{b18}CH$	DYW
DYW:KP	$KPAx_{c1}Ax_{c2}IEx_{c3}$	$Hx_{cc}E$	$Cx_{c4}x_{c5}CH$	$x_{c6}x_{c7}x_{c8}$

[0092] (DYW:PG)

[0093] DYW:PG是由 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\dots Hx_{aa}E\dots Cx_{a17}x_{a18}CH\dots$ DYW构成的多肽。优选为具有 $x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP\dots Hx_{aa}E\dots Cx_{a17}x_{a18}CH\dots$ DYW、与序列编号1的序列具有序列一致性(在[术语]项中详细说明)、且具有C-U/U-C编辑活性的多肽。DYW:PG具有将编辑靶标C转换成U的活性(C-U编辑活性)。序列编号1中示出了本说明书的实施例项所示的实验中使用的、由全长136个氨基酸构成的DYW:PG的序列。该序列在本申请中首次公开,为新序列。

[0094] DYW:PG的全长只要能够发挥出C-U编辑活性就没有特别限定,例如为110~160个氨基酸的长度、优选为124~148个氨基酸的长度、更优选为128~144个氨基酸的长度、进一步优选为132~140个氨基酸的长度。

[0095] 在DYW:PG的包含PG盒的区域($x_{a1}PGx_{a2}SWIEx_{a3}-x_{a16}HP$)中:

[0096] 关于 x_{a1} ,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选为E(谷氨酸)或与其性质相似的氨基酸,更优选为G。

[0097] 关于 x_{a2} ,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选为C(半胱氨酸)或与其性质相似的氨基酸,更优选为C。

[0098] 关于 $x_{a3}-x_{a16}$ 的各氨基酸,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选与序列编号1的序列的位置9~22的对应氨基酸相同、或者为与对应氨基酸的性质相似的氨基酸,更优选为与序列编号1的序列的位置9~22的对应氨基酸相同的氨基酸。

[0099] 在优选的一个方式中,DYW:PG的HxE区域无论在其他区域如何的情况下均为HSE。

[0100] 在DYW:PG的CxxCH区域、即 $Cx_{a17}x_{a18}CH$ 中:

[0101] 关于 x_{a17} ,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选为G(甘氨酸)或与其性质相似的氨基酸,更优选为G。

[0102] 关于 x_{a18} ,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选为D(天冬氨酸)或与其性质相似的氨基酸,更优选为D。

[0103] DYW:PG中,将包含PG盒的区域与 $Hx_{aa}E$ 结合的…部分、将 $Hx_{aa}E$ 区域与CxxCH区域结合的…部分、以及将CxxCH区域与DYW连结的…部分依次称为第一连结部、第二连结部、以及第三连结部(在其他DYW结构域中也同样)。

[0104] 关于DYW:PG的第一连结部的全长,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,例如为39~47个氨基酸的长度、优选为40~46个氨基酸的长度、更优选为41~45个氨基酸的长度、进一步优选为42~44个氨基酸的长度。关于第一连结部的氨基酸序列,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选与序列编号1的序列的位置25~67的部分相同、或者为在该部分序列中有1~22个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0105] DYW:PG的第一连结部的优选方式之一是无无论DYW结构域的其他部分的序列如何的情况下均为43个氨基酸长度的下式所表示的多肽。

[0106] $N_{a25}-N_{a26}-N_{a27}-\dots-N_{a65}-N_{a66}-N_{a67}$

[0107] 上述多肽优选与序列编号1的序列的位置25~67的部分相同、或者为该部分序列中有复数个氨基酸发生了置换且能够作为DYW:PG发挥出C-U编辑活性的序列。此时氨基酸的置换优选按照下述方式进行:在图4的对应位置中比特(bits)值大的氨基酸(例如 N_{a29} 、 N_{a30} 、 N_{a32} 、 N_{a33} 、 N_{a35} 、 N_{a36} 、 N_{a40} 、 N_{a44} 、 N_{a45} 、 N_{a47} 、 N_{a48} 、 N_{a52} 、 N_{a53} 、 N_{a54} 、 N_{a55} 、 N_{a58} 、 N_{a61} 、 N_{a65} 、 N_{a67})与图4

相同,除此以外的氨基酸被置换。

[0108] 关于DYW:PG的第二连结部的全长,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,例如为21~29个氨基酸的长度、优选为22~28个氨基酸的长度、更优选为23~27个氨基酸的长度、进一步优选为24~26个氨基酸的长度。关于第二连结部的氨基酸序列,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选与序列编号1的序列的位置71~95的部分相同、或者为在该部分序列中有1~13个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0109] DYW:PG的第二连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为25个氨基酸长度的下式所表示的多肽。

[0110] $N_{a71}-N_{a72}-N_{a73}-\cdots-N_{a93}-N_{a94}-N_{a95}$

[0111] 上述多肽优选与序列编号1的序列的位置71~95的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:PG发挥出C-U编辑活性的序列。此时氨基酸的置换优选以下述方式进行:在图4的对应位置中比特值大的氨基酸(例如 N_{a71} 、 N_{a72} 、 N_{a73} 、 N_{a76} 、 N_{a77} 、 N_{a78} 、 N_{a79} 、 N_{a81} 、 N_{a82} 、 N_{a86} 、 N_{a88} 、 N_{a89} 、 N_{a91} 、 N_{a92} 、 N_{a93} 、 N_{a94})与图4相同,除此以外的氨基酸被置换。

[0112] 关于DYW:PG的第三连结部的全长,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,例如为29~37个氨基酸的长度、优选为30~36个氨基酸的长度、更优选为31~35个氨基酸的长度、进一步优选为32~34个氨基酸的长度。关于第三连结部的氨基酸序列,只要能够作为DYW:PG发挥出C-U编辑活性就没有特别限定,优选与序列编号1的序列的位置101~133的部分相同、或者为在该部分序列中有1~17个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0113] DYW:PG的第三连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为33个氨基酸长度的下式所表示的多肽。

[0114] $N_{a101}-N_{a102}-N_{a103}-\cdots-N_{a131}-N_{a132}-N_{a133}$

[0115] 上述多肽优选与序列编号1的序列的位置101~133的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:PG发挥出C-U编辑活性的序列。此时氨基酸的置换优选以下述方式进行:在图4的对应位置中比特值大的氨基酸(例如 N_{a102} 、 N_{a104} 、 N_{a107} 、 N_{a112} 、 N_{a114} 、 N_{a117} 、 N_{a118} 、 N_{a121} 、 N_{a122} 、 N_{a123} 、 N_{a124} 、 N_{a125} 、 N_{a128} 、 N_{a130} 、 N_{a131} 、 N_{a132})与图4相同,除此以外的氨基酸被置换。

[0116] 通过在由序列编号1的序列构成的PG结构域中导入突变,能够提高RNA编辑活性。这样的导入了突变的结构域的优选例为本说明书的实施例项中所示的PG11(由序列编号50的氨基酸序列构成的多肽)。

[0117] (DYW:WW)

[0118] DYW:WW是由 x_{b1} PG x_{b2} SWWTD x_{b3} - x_{b16} HP \cdots H x_{bb} E \cdots C x_{b17} x_{b18} CH \cdots DYW构成的多肽。优选为具有 x_{b1} PG x_{b2} SWWTD x_{b3} - x_{b16} HP \cdots H x_{bb} E \cdots C x_{b17} x_{b18} CH \cdots DYW、与序列编号2的序列具有序列一致性、且具有C-U/U-C编辑活性的多肽。DYW:WW具有将编辑靶标C转换成U的活性(C-U编辑活性)。序列编号2中示出了本说明书的实施例项所示的实验中使用的、由全长137个氨基酸构成的DYW:WW的序列。该序列在本申请中首次公开,为新序列。

[0119] 关于DYW:WW的全长,只要能够发挥出C-U编辑活性就没有特别限定,例如为110~

160个氨基酸的长度、优选为125~149个氨基酸的长度、更优选为129~145个氨基酸的长度、进一步优选为133~141个氨基酸的长度。

[0120] 在DYW:WW的包含PG盒的区域、即 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP$ 中,由WTD构成的部分可以为WSD。

[0121] 在DYW:WW的包含PG盒的区域中:

[0122] 关于 x_{b1} ,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选为K(赖氨酸)或与其性质相似的氨基酸,更优选为K。

[0123] 关于 x_{b2} ,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选为Q(谷氨酰胺)或与其性质相似的氨基酸,更优选为Q。

[0124] 关于 $x_{b3}-x_{b16}$ 的各氨基酸,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选与序列编号2的序列的位置10~23的对应氨基酸相同、或者为与对应氨基酸的性质相似的氨基酸,更优选为与序列编号2的序列的位置10~23的对应氨基酸相同的氨基酸。

[0125] 在优选的一个方式中,DYW:WW的HxE区域无论在其他部分的序列如何的情况下均为HSE。

[0126] 在DYW:WW的CxxCH区域、即 $Cx_{b17}x_{b18}CH$ 中:

[0127] 关于 x_{b17} ,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选为D(天冬氨酸)或与其性质相似的氨基酸,更优选为D。

[0128] 关于 x_{b18} ,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选为D或与其性质相似的氨基酸,更优选为D。

[0129] 关于DYW:WW的第一连结部的全长,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,例如为39~47个氨基酸的长度、优选为40~46个氨基酸的长度、更优选为41~45个氨基酸的长度、进一步优选为42~44个氨基酸的长度。关于第一连结部的氨基酸序列,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选与序列编号2的序列的位置25~67的部分相同、或者为在该部分序列中有1~22个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0130] DYW:WW的第一连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为43个氨基酸长度的下式所表示的多肽。

[0131] $N_{b26}-N_{b27}-N_{b28}-\dots-N_{b66}-N_{b67}-N_{b68}$

[0132] 上述多肽优选与序列编号2的序列的位置26~68的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:PG发挥出C-U编辑活性的序列。此时氨基酸的置换优选以下述方式进行:在图4的对应位置中比特值大的氨基酸(例如 N_{b26} 、 N_{b30} 、 N_{b33} 、 N_{b34} 、 N_{b37} 、 N_{b41} 、 N_{b45} 、 N_{b46} 、 N_{b48} 、 N_{b49} 、 N_{b51} 、 N_{b52} 、 N_{b53} 、 N_{b55} 、 N_{b56} 、 N_{b57} 、 N_{b59} 、 N_{b61} 、 N_{b62} 、 N_{b63} 、 N_{b64} 、 N_{b66} 、 N_{b67} 、 N_{b68})与图4相同,除此以外的氨基酸被置换。

[0133] 关于DYW:WW的第二连结部的全长,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,例如为21~29个氨基酸的长度、优选为22~28个氨基酸的长度、更优选为23~27个氨基酸的长度、进一步优选为24~26个氨基酸的长度。关于第二连结部的氨基酸序列,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选与序列编号2的序列的位置71~95的部分相同、或者为在该部分序列中有1~13个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0134] DYW:WW的第二连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为25个氨基酸长度的下式所表示的多肽。

[0135] $N_{b72}-N_{b73}-N_{b74}-\cdots-N_{b94}-N_{b95}-N_{b96}$

[0136] 上述多肽优选与序列编号2的序列的位置72~96的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:WW发挥出C-U编辑活性的序列。此时氨基酸的置换优选以下述方式进行：在图4的对应位置中比特值大的氨基酸(例如 N_{b72} 、 N_{b73} 、 N_{b74} 、 N_{b75} 、 N_{b77} 、 N_{b78} 、 N_{b79} 、 N_{b81} 、 N_{b82} 、 N_{b84} 、 N_{b88} 、 N_{b89} 、 N_{b90} 、 N_{b91} 、 N_{b92} 、 N_{b93} 、 N_{b94} 、 N_{b95} 、 N_{b96})与图4相同,除此以外的氨基酸被置换。

[0137] 关于DYW:WW的第三连结部的全长,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,例如为29~37个氨基酸的长度、优选为30~36个氨基酸的长度、更优选为31~35个氨基酸的长度、进一步优选为32~34个氨基酸的长度。关于第三连结部的氨基酸序列,只要能够作为DYW:WW发挥出C-U编辑活性就没有特别限定,优选与序列编号2的序列的位置101~133的部分相同、或者为在该部分序列中有1~17个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0138] DYW:WW的第三连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为33个氨基酸长度的下式所表示的多肽。

[0139] $N_{b102}-N_{b103}-N_{b104}-\cdots-N_{b132}-N_{b133}-N_{b134}$

[0140] 上述多肽优选与序列编号2的序列的位置102~134的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:WW发挥出C-U编辑活性的序列。此时氨基酸的置换优选以下述方式进行：在图4的对应位置中比特值大的氨基酸(例如 N_{b104} 、 N_{b105} 、 N_{b107} 、 N_{b108} 、 N_{b109} 、 N_{b110} 、 N_{b111} 、 N_{b113} 、 N_{b115} 、 N_{b116} 、 N_{b117} 、 N_{b118} 、 N_{b119} 、 N_{b121} 、 N_{b122} 、 N_{b123} 、 N_{b124} 、 N_{b126} 、 N_{b129} 、 N_{b131} 、 N_{b132} 、 N_{b133} 、 N_{b134})与图4相同,除此以外的氨基酸进行置换。

[0141] 通过在由序列编号2的序列构成的WW结构域中导入突变,能够提高RNA编辑活性。这样的导入了突变的结构域的优选例为本说明书的实施例项中所示的WW2~11和WW13,特别是编辑活性高的为WW11(由序列编号63的氨基酸序列构成的多肽)。

[0142] (DYW:KP)

[0143] DYW:KP是由 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 构成的多肽。优选为具有 $KPAx_{c1}Ax_{c2}IEx_{c3}\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots x_{c6}x_{c7}x_{c8}$ 、与序列编号3的序列具有序列一致性、且具有C-U/U-C编辑活性的多肽。DYW:KP具有将编辑靶标U转换成C的活性(U-C编辑活性)。序列编号3中示出了本说明书的实施例项所示的实验中使用的、由全长133个氨基酸构成的DYW:KP的序列。该序列在本申请中首次公开,为新序列。

[0144] 关于DYW:KP的全长,只要能够发挥出U-C编辑活性就没有特别限定,例如为110~160个氨基酸的长度、优选为121~145个氨基酸的长度、更优选为125~141个氨基酸的长度、进一步优选为129~137个氨基酸的长度。

[0145] 在DYW:KP的包含PG盒的区域、即 $KPAx_{c1}Ax_{c2}IEx_{c3}$ 中:

[0146] 关于 x_{c1} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为S(丝氨酸)或与其性质相似的氨基酸,更优选为S。

[0147] 关于 x_{c2} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为L(亮氨酸)或与其性质相似的氨基酸,更优选为L。

[0148] 关于 x_{c3} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为V(缬氨酸)或与其性质相似的氨基酸,更优选为V。

[0149] 在优选的一个方式中,DYW:KP的HxE区域无论在其他部分的序列如何的情况下均为HAE。

[0150] 在DYW:KP的CxxCH区域、即 $Cx_{c4}x_{c5}CH$ 中:

[0151] 关于 x_{c4} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为N(天冬酰胺)或与其性质相似的氨基酸,更优选为N。

[0152] 关于 x_{b5} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为D(天冬氨酸)或与其性质相似的氨基酸,更优选为D。

[0153] 在DYW:KP的与DYW相当的部分、即 $x_{c6}x_{c7}x_{c8}$ 中:

[0154] 关于 x_{c6} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为D(天冬氨酸)或与其性质相似的氨基酸,更优选为D。

[0155] 关于 x_{c7} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为M(蛋氨酸)或与其性质相似的氨基酸,更优选为M。

[0156] 关于 x_{b8} ,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选为F(苯丙氨酸)或与其性质相似的氨基酸,更优选为F。

[0157] 在优选的一个方式中, $x_{c6}x_{c7}x_{c8}$ 无论在其他部分的序列如何的情况下均为 $Dx_{c7}x_{c8}$ 。

[0158] 在另一优选方式中, $x_{c6}x_{c7}x_{c8}$ 无论在其他部分的序列如何的情况下均为GRP。

[0159] 关于DYW:KP的第一连结部的全长,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,例如为51~59个氨基酸的长度、优选为52~58个氨基酸的长度、更优选为53~57个氨基酸的长度、进一步优选为54~56个氨基酸的长度。关于第一连结部的氨基酸序列,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选与序列编号3的序列的位置10~64的部分相同、或者为在该部分序列中有1~28个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0160] DYW:KP的第一连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为55个氨基酸长度的下式所表示的多肽。

[0161] $N_{c10}-N_{c11}-N_{c12}-\dots-N_{c62}-N_{c63}-N_{c64}$

[0162] 上述多肽优选与序列编号3的序列的位置10~64的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:KP发挥出U-C编辑活性的序列。此时氨基酸的置换优选以下述方式进行:在图4的对应位置中比特值大的氨基酸(例如 N_{c10} 、 N_{c13} 、 N_{c14} 、 N_{c15} 、 N_{c16} 、 N_{c17} 、 N_{c18} 、 N_{c19} 、 N_{c25} 、 N_{c26} 、 N_{c29} 、 N_{c30} 、 N_{c33} 、 N_{c34} 、 N_{c36} 、 N_{c38} 、 N_{c41} 、 N_{c42} 、 N_{c44} 、 N_{c45} 、 N_{c47} 、 N_{c49} 、 N_{c55} 、 N_{c58} 、 N_{c59} 、 N_{c62} 、 N_{c63} 、 N_{c64})与图4相同,除此以外的氨基酸被置换。

[0163] 关于DYW:KP的第二连结部的全长,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,例如为21~29个氨基酸的长度、优选为22~28个氨基酸的长度、更优选为23~27个氨基酸的长度、进一步优选为24~26个氨基酸的长度。关于第二连结部的氨基酸序列,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选与序列编号3的序列的位置68~92的部分相同、或者为在该部分序列中有1~13个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0164] DYW:KP的第二连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何

的情况下均为25个氨基酸长度的下式所表示的多肽。

[0165] $N_{c68}-N_{c69}-N_{c70}-\cdots-N_{c90}-N_{c91}-N_{c92}$

[0166] 上述多肽优选与序列编号3的序列的位置68~92的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:KP发挥出U-C编辑活性的序列。此时氨基酸的置换优选以下述方式进行：在图4的对应位置中比特值大的氨基酸(例如 N_{c68} 、 N_{c70} 、 N_{c71} 、 N_{c72} 、 N_{c73} 、 N_{c74} 、 N_{c75} 、 N_{c76} 、 N_{c77} 、 N_{c78} 、 N_{c79} 、 N_{c80} 、 N_{c81} 、 N_{c83} 、 N_{c84} 、 N_{c85} 、 N_{c86} 、 N_{c87} 、 N_{c88} 、 N_{c89} 、 N_{c90} 、 N_{c91} 、 N_{c92})与图4相同,除此以外的氨基酸被置换。

[0167] 关于DYW:KP的第三连结部的全长,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,例如为29~37个氨基酸的长度、优选为30~36个氨基酸的长度、更优选为31~35个氨基酸的长度、进一步优选为32~34个氨基酸的长度。关于第三连结部的氨基酸序列,只要能够作为DYW:KP发挥出U-C编辑活性就没有特别限定,优选与序列编号3的序列的位置98~130的部分相同、或者为在该部分序列中有1~17个氨基酸发生了置换、缺失或添加的序列、或者为与该部分序列具有序列一致性的序列,更优选为与该部分序列相同的序列。

[0168] DYW:KP的第三连结部的优选方式之一是在无论在DYW结构域的其他部分的序列如何的情况下均为33个氨基酸长度的下式所表示的多肽。

[0169] $N_{c98}-N_{c99}-N_{c100}-\cdots-N_{c128}-N_{c129}-N_{c130}$

[0170] 上述多肽优选与序列编号3的序列的位置98~130的部分相同、或者为在该部分序列中有复数个氨基酸发生了置换且能够作为DYW:KP发挥出U-C编辑活性的序列。此时氨基酸的置换优选以下述方式进行：在图4的对应位置中比特值大的氨基酸(例如 N_{c98} 、 N_{c100} 、 N_{c101} 、 N_{c103} 、 N_{c104} 、 N_{c105} 、 N_{c107} 、 N_{c109} 、 N_{c110} 、 N_{c111} 、 N_{c112} 、 N_{c114} 、 N_{c115} 、 N_{c117} 、 N_{c118} 、 N_{c119} 、 N_{c120} 、 N_{c121} 、 N_{c122} 、 N_{c123} 、 N_{c124} 、 N_{c125} 、 N_{c127} 、 N_{c129} 、 N_{c130})与图4相同,除此以外的氨基酸被置换。

[0171] 通过在由序列编号3的序列构成的KP结构域中导入突变,能够提高编辑活性。这样的导入了突变的结构域的优选例为本说明书的实施例项中所示的KP2~23(序列编号68~89)。KP22(序列编号88)的U-C编辑活性最高、且C-U编辑活性最低,与由序列编号3的序列构成的KP结构域相比, RNA编辑活性得到了改善。

[0172] (嵌合DYW)

[0173] DYW结构域可以根据氨基酸序列的保守性划分成若干个区域。通过对这些各区域进行交换而得到的嵌合DYW,可提高C-U编辑活性或U-C编辑活性。优选的一个方式是将DYW:WW的包含PG盒的区域、即 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP$ 和DYW与DYW:KP的其他区域($\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots$)融合而成的嵌合DYW。此时的 x_{b1} 、 x_{b2} 、 x_{b3} 、 x_{b16} 如关于DYW:WW的上述说明所述。另外,第一连结部、第二连结部、第三连结部如关于DYW:KP的上述说明所述。DYW部分可以为 $Dx_{bc1}x_{bc2}$ 。关于全长,只要能够发挥出U-C编辑活性就没有特别限定,例如为110~160个氨基酸的长度、优选为125~149个氨基酸的长度、更优选为129~145个氨基酸的长度、进一步优选为133~141个氨基酸的长度。

[0174] 优选的嵌合结构域之一由具有 $x_{b1}PGx_{b2}SWWTDx_{b3}-x_{b16}HP\cdots Hx_{cc}E\cdots Cx_{c4}x_{c5}CH\cdots Dx_{bc1}x_{bc2}$ 、与序列编号2的90的序列具有至少40%的序列一致性、且具有C-U/U-C编辑活性的多肽构成。

[0175] 特别优选的嵌合结构域之一为与序列编号90的序列具有序列一致性、且具有U-C编辑活性的多肽。序列编号90中示出了本说明书的实施例项所示的实验中使用的嵌合结构

域的序列。关于该结构域，确认到比由序列编号3的序列构成的DYW:KP更高的U-C编辑活性，而且几乎没有C-U编辑活性。

[0176] (与公知的序列的比较)

[0177] 如上所述，由序列编号1、2、3的序列构成的DYW结构域是新序列。以下示出使用PPR数据库(<https://ppr.plantenergy.uwa.edu.au/onekp/>; 上述非专利文献11)进行调查得到的结果。

[0178] [表2-1]

DYW:PG(序列编号:1)

序列名称	来源	一致性
- PYHZ-2068577-R	水韭属(Isoetes), 小叶类	78%
- ENQF-2003735-R	多穗石松(Lycopodium annotinum), 小叶类	

DYW:WW(序列编号:2)

[0179] - ANON-2000164-R	平孢角苔(Leiosporoceros Dussii), 角苔类	84%
- ANON-2000163-R		
- ANON-2004532-F		
- ANON-2001227-F		

DYW:KP(序列编号:3)

- UWOD-2140463-R	华东瘤足蕨(Plagiogyria japonica), 蕨类植物	86%
------------------	-----------------------------------	-----

[0180] 因此，本发明还提供下述d~i中的任一多肽：

[0181] d. 与序列编号1的序列具有大于78%的序列一致性、优选具有80%以上的序列一致性、更优选具有85%以上的序列一致性、进一步优选具有90%的序列一致性、进一步优选具有95%的序列一致性、进一步优选具有97%的序列一致性、且具有C-U编辑活性的多肽；

[0182] e. 与序列编号2的序列具有大于84%的序列一致性、优选具有85%以上的序列一致性、更优选具有90%以上的序列一致性、进一步优选具有95%的序列一致性、进一步优选具有97%的序列一致性、且具有C-U编辑活性的多肽；

[0183] f. 与序列编号3的序列具有大于86%的序列一致性、优选具有87%以上的序列一致性、更优选具有90%以上的序列一致性、进一步优选具有95%的序列一致性、进一步优选具有97%的序列一致性、且具有U-C编辑活性的多肽。

[0184] g. 具有在序列编号1的序列中1~29个、优选1~25个、更优选1~21个、进一步优选1~17个、进一步优选1~13个、进一步优选1~9个、进一步优选1~5个氨基酸发生了置换、缺失或添加而成的序列、且具有C-U编辑活性的多肽；

[0185] h. 具有在序列编号2的序列中1~21个、优选1~18个、更优选1~15个、进一步优选1~12个、进一步优选1~9个、进一步优选1~6个氨基酸发生了置换、缺失或添加而成的序列、且具有C-U编辑活性的多肽；

[0186] i. 具有在序列编号3的序列中1~18个、优选1~16个、更优选1~14个、进一步优选

[0189] [RNA结合结构域]

[0190] 本发明中,在使用DYW结构域对编辑靶标进行转换时,为了靶向编辑靶标所包含的RNA,使用RNA结合蛋白作为RNA结合结构域。

[0191] 作为RNA结合结构域使用的RNA结合蛋白的优选例为由PPR基序构成的PPR蛋白。

[0192] (PPR基序)

[0193] 关于PPR基序,除特别记载的情况以外,是指具有下述氨基酸序列的由30~38个氨基酸构成的多肽,所述氨基酸序列在利用网络上的蛋白质结构域检索程序对氨基酸序列进行分析时,利用Pfam中的PF01535、Prosite中的PS51375得到的E值为规定值以下(优选为E-03)。构成本发明中定义的PPR基序的氨基酸的位置编号与PF01535中大致同义,另一方面,其相当于从PS51375的氨基酸的位置减去2而得到的数(例如,本发明的1位→PS51375的3位)。其中,在提到“ii”(-2)位的氨基酸时,是从构成PPR基序的氨基酸的末尾(C末端侧)起第2位的氨基酸;或者是相对于下一个PPR基序的1位氨基酸向N末端侧数2个、即第-2位的氨基酸。在未明确鉴定出下一个PPR基序的情况下,将相对于下一个螺旋结构的第1位氨基酸向前数2个的氨基酸作为“ii”。关于Pfam可以参考<http://pfam.sanger.ac.uk/>,关于Prosite可以参考<http://www.expasy.org/prosite/>。

[0194] PPR基序的保守氨基酸序列在氨基酸水平上的保守性低,但在二级结构上2个 α 螺旋是非常保守的。代表性的PPR基序由35个氨基酸构成,但其长度在30~38个氨基酸的范围内可变。

[0195] 更具体地说,PPR基序由式1所表示的长度30~38个氨基酸的多肽构成。

[0196] [化1]

[0197] (螺旋A)-X-(螺旋B)-L (式1)

[0198] 式中:

[0199] 螺旋A为长度12个氨基酸的能够形成 α 螺旋结构的部分,由式2表示,

[0200] [化2]

[0201] $A_1-A_2-A_3-A_4-A_5-A_6-A_7-A_8-A_9-A_{10}-A_{11}-A_{12}$ (式2)

[0202] 式2中, $A_1\sim A_{12}$ 各自独立地表示氨基酸;

[0203] X不存在或者为由长度1~9个氨基酸构成的部分;

[0204] 螺旋B为由长度11~13个氨基酸构成的能够形成 α 螺旋结构的部分;

[0205] L为长度2~7个氨基酸的由式3所表示的部分;

[0206] [化3]

[0207] $L_{v11}-L_{vi}-L_v-L_{iv}-L_{iii}-L_{ii}-L_i$ (式3)

[0208] 式3中,各氨基酸从C末端侧起被编号为“i”(-1)、“ii”(-2),

[0209] 其中, $L_{iii}\sim L_{vii}$ 有不存在的情况。

[0210] (PPR密码)

[0211] PPR基序中,第1、4、ii位的3个氨基酸的组合对于与碱基的特异性结合很重要,可以通过它们的组合来决定所结合的碱基是哪一种碱基。第1、4、ii位的3个氨基酸的组合与能够结合的碱基的关系作为PPR密码为人所知(上述专利文献2),如下文所述。

[0212] (1) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为缬氨酸、天冬酰胺和天冬氨酸的情况下,该PPR基序具有与U强结合、其次与C、再次与A或G结合这样的选择性RNA碱基结合能力。

[0213] (2) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为缬氨酸、苏氨酸、天冬酰胺的情况下，该PPR基序具有与A强结合、其次与G、再次与C结合、但不与U结合这样的选择性RNA碱基结合能力。

[0214] (3) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为缬氨酸、天冬酰胺、天冬酰胺的情况下，该PPR基序具有与C强结合、其次与A或U结合、但不与G结合这样的选择性RNA碱基结合能力。

[0215] (4) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为谷氨酸、甘氨酸、天冬氨酸的情况下，该PPR基序具有与G强结合、但不与A、U和C结合这样的选择性RNA碱基结合能力。

[0216] (5) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为异亮氨酸、天冬酰胺、天冬酰胺的情况下，该PPR基序具有与C强结合、其次与U、再次与A结合、但不与G结合这样的选择性RNA碱基结合能力。

[0217] (6) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为缬氨酸、苏氨酸、天冬氨酸的情况下，该PPR基序具有与G强结合、其次与U结合、但不与A和C结合这样的选择性RNA碱基结合能力。

[0218] (7) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为赖氨酸、苏氨酸、天冬氨酸的情况下，该PPR基序具有与G强结合、其次与A结合、但不与U和C结合这样的选择性RNA碱基结合能力。

[0219] (8) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为苯丙氨酸、丝氨酸、天冬酰胺的情况下，该PPR基序具有与A强结合、其次与C、再次与G和U结合这样的选择性RNA碱基结合能力。

[0220] (9) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为缬氨酸、天冬酰胺、丝氨酸的情况下，该PPR基序具有与C强结合、其次与U结合、但不与A和G结合这样的选择性RNA碱基结合能力。

[0221] (10) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为苯丙氨酸、苏氨酸、天冬酰胺的情况下，该PPR基序具有与A强结合、但不与G、U和C结合这样的选择性RNA碱基结合能力。

[0222] (11) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为异亮氨酸、天冬酰胺、天冬氨酸的情况下，该PPR基序具有与U强结合、其次与A结合、但不与G和C结合这样的选择性RNA碱基结合能力。

[0223] (12) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为苏氨酸、苏氨酸、天冬酰胺的情况下，该PPR基序具有与A强结合、但不与G、U和C结合这样的选择性RNA碱基结合能力。

[0224] (13) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为异亮氨酸、蛋氨酸、天冬氨酸的情况下，该PPR基序具有与U强结合、其次与C结合、但不与A和G结合这样的选择性RNA碱基结合能力。

[0225] (14) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为苯丙氨酸、脯氨酸、天冬氨酸的情况下，该PPR基序具有与U强结合、其次与C结合、但不与A和G结合这样的选择性RNA碱基结合能力。

[0226] (15) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为酪氨酸、脯氨酸、天冬氨酸的情况下，该PPR基序具有与U强结合、但不与A、G和C结合这样的选择性RNA碱基结合能力。

[0227] (16) A_1 、 A_4 以及 L_{ii} 这3个氨基酸的组合依次为亮氨酸、苏氨酸、天冬氨酸的情况下，该PPR基序具有与G强结合、但不与A、U和C结合这样的选择性RNA碱基结合能力。

[0228] (P阵列)

[0229] PPR蛋白根据构成的PPR基序的结构被分类成P和PLS这2个家族。

[0230] P型PPR蛋白由标准的35个氨基酸的PPR基序(P)的单纯重复(P阵列)构成。本发明

的DYW结构域可以与P型PPR蛋白连结来使用。

[0231] (PLS阵列)

[0232] PLS型PPR蛋白的PPR基序的排列以P1、L1以及S1这3个PPR基序的重复单元的形式构成,在该重复的C末端侧接续P2、L2以及S2基序。此外有时在该最后的P2L2S2基序的C末端侧接续被称为E1和E2的2个PPR样基序、以及DYW结构域。

[0233] 关于P1的全长,只要能够与靶标碱基结合就没有特别限定,例如为33~37个氨基酸的长度、优选为34~36个氨基酸的长度、更优选为35个氨基酸的长度。关于L1的全长,只要能够与靶标碱基结合就没有特别限定,例如为33~37个氨基酸的长度、优选为34~36个氨基酸的长度、更优选为35个氨基酸的长度。关于S1的全长,只要能够与靶标碱基结合就没有特别限定,例如为30~33个氨基酸的长度、优选为30~32个氨基酸的长度、更优选为31个氨基酸的长度。

[0234] 关于P2的全长,只要能够与靶标碱基结合就没有特别限定,例如为33~37个氨基酸的长度、优选为34~36个氨基酸的长度、更优选为35个氨基酸的长度。关于L2的全长,只要能够与靶标碱基结合就没有特别限定,例如为34~38个氨基酸的长度、优选为35~37个氨基酸的长度、更优选为36个氨基酸的长度。关于S2的全长,只要能够与靶标碱基结合就没有特别限定,例如为30~34个氨基酸的长度、优选为31~33个氨基酸的长度、更优选为32个氨基酸的长度。在序列编号17、22和27中示出了本说明书的实施例项中使用的P2的序列。在序列编号18、23和28中示出了本说明书的实施例项中使用的L2的序列。在序列编号:19、24和29中示出了本说明书的实施例项中使用的S2的序列。

[0235] 关于E1的全长,只要能够与靶标碱基结合就没有特别限定,例如为32~36个氨基酸的长度、优选为33~35个氨基酸的长度、更优选为34个氨基酸的长度。在序列编号20、25、30中示出了本说明书的实施例项中使用的E1的序列。

[0236] 关于E2的全长,只要能够与靶标碱基结合就没有特别限定,例如为30~34个氨基酸的长度、优选为31~33个氨基酸的长度、更优选为33个氨基酸的长度。序列编号21、26、31中示出了本说明书的实施例项中使用的E2的序列。

[0237] PLS型PPR蛋白中,P1L1S1的重复部分和P2之前的部分可以按照上述PPR密码的规则根据靶标RNA的序列进行设计。

[0238] S2基序与ii的氨基酸(序列编号19的第31位的N)所对应的核苷酸具有相关性(上述非专利文献8)。另外,要注意S2基序的靶标碱基向右4个的C或U是DYW结构域的编辑靶标碱基,可以将其并入到PLS型PPR蛋白中。

[0239] 在E1基序中,仅在第4位(序列编号20的第4位的G)的氨基酸处观察到与核苷酸的相关性(Ruwe et al. (2019) *New Phytol.* 222 218-229.)。

[0240] E2基序中的第4位(序列编号21的第4位的V)以及最后(序列编号21的第33位的K)的氨基酸具有高度保守性,不参与特定的PPR-RNA识别(上述非专利文献2)。

[0241] 关于P1L1S1的重复数,只要能够与靶标碱基序列结合就没有特别限定,例如为1~5、优选为2~4、更优选为3。原理上,即使是1个单元(3个)也可使用。已知由5个PPR基序构成的MEF8(L1-S1-P2-L2-S2-E-DYW)参与约60个部位的编辑。

[0242] 在天然的PPR蛋白中,位于最开始和最后的P1L1S1与位于特定位置的氨基酸残基相比观察到明确的差异,与内部的P1L1S1不同。从设计出尽可能接近天然存在的PPR基序的

人工PLS阵列的方面出发,可分成位于第一位(N末端侧)的P1L1S1、位于内部的P1L1S1、以及位于紧靠P2L2S2之前的最后(C末端侧)的P1L1S1,设计出与PPR基序的位置对应的3种P1L1S1。需要说明的是,与天然地由P-L-S的重复单元构成的PPR基序不同,也存在SS(31个氨基酸)重复的情况,这样的PPR基序也可用于本发明中。

[0243] [DYW蛋白]

[0244] 本发明提供一种用于编辑靶标RNA的DYW蛋白,其包含:包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA序列特异性地结合的RNA结合结构域;以及作为上述DYW:PG、DYW:WW或DYW:KP中的任一者的DYW结构域。

[0245] 这样的DYW蛋白可以是人工物。人工是指并非天然物、而是人为合成的物质。人工例如对应于下述情况:具有非天然存在的序列的DYW结构域的情况;具有非天然存在的序列的PPR结合结构域的情况;具有RNA结合结构域和DYW结构域的非天然存在的组合的情况;动物细胞中的RNA编辑用蛋白被进一步添加了植物来源的天然DYW蛋白中不存在的部分、例如核转移信号、人线粒体转移信号的情况。作为核转移信号序列,例如可以举出SV40大T抗原来源的PKKKRKV(SEQ ID NO:32)、作为核质蛋白的NLS的KRPAATKKAGQAKKKK(SEQ ID NO:33)等。

[0246] PPR基序的数目可以根据作为靶标的RNA的序列适当设定。PPR基序的数目至少为1个即可,也可以为2个以上。已知PPR基序为2个时,能够与RNA结合(Nucleic Acids Research, 2012, Vol. 40, No. 6, 2712-2723)。

[0247] DYW蛋白的优选方式之一如下所述。

[0248] 一种用于编辑靶标RNA的人工DYW蛋白,其包含:包含至少1个、优选2~25个、更优选5~20个、进一步优选10~18个PPR基序,能够按照PPR密码的规则与靶标RNA序列特异性结合的RNA结合结构域;以及作为上述DYW:PG、DYW:WW或DYW:KP中的任一者的DYW结构域。

[0249] 其他DYW蛋白的优选方式之一如下所述。

[0250] 一种用于编辑靶标RNA的DYW蛋白,其包含:包含至少1个、优选2~25个、更优选5~20个、进一步优选10~18个PPR基序,能够按照PPR密码的规则与动物的靶标RNA序列特异性结合的RNA结合结构域(优选作为PLS型PPR蛋白的RNA结合结构域);以及作为上述DYW:PG、DYW:WW或DYW:KP中的任一者的DYW结构域。

[0251] 本发明的作为DYW结构域、RNA结合结构域的PPR蛋白以及DYW蛋白可以利用本领域技术人员熟知的方法比较大量地制备。这样的方法可以包括:根据目标结构域或蛋白质所具有的氨基酸序列来决定编码其的核酸序列,进行克隆,制作生产目标结构域或蛋白质的转化体。

[0252] [DYW蛋白的利用]

[0253] (编码DYW蛋白等的核酸、载体、细胞)

[0254] 本发明还提供编码上述PPR基序、DYW蛋白或DYW蛋白的核酸、包含核酸的载体(例如用于扩增的载体、表达载体)。载体中也包括病毒载体。用于扩增的载体可以使用大肠杆菌或酵母作为宿主。本说明书中,表达载体是指例如从上游起包含具有启动子序列的DNA、编码所期望的蛋白质的DNA、以及具有终止子序列的DNA的载体,但只要发挥出所期望的功能,不一定必须按照该顺序排列。本发明中,可以将本领域技术人员通常可使用的各种载体重组来使用。

[0255] 具体地说,本发明提供一种编码DYW蛋白的核苷酸序列,该DYW蛋白包含:包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA(优选动物的靶标RNA)序列特异性结合的RNA结合结构域(优选作为PLS型PPR蛋白的RNA结合结构域);以及作为上述DYW:PG、DYW:WW或DYW:KP中的任一者的DYW结构域。

[0256] 另外,具体地说,本发明提供一种用于编辑靶标RNA的载体,其包含编码DYW蛋白的核苷酸序列,该DYW蛋白包含:包含至少1个PPR基序、能够按照PPR密码的规则与靶标RNA(优选动物的靶标RNA)序列特异性结合的RNA结合结构域(优选作为PLS型PPR蛋白的RNA结合结构域);以及作为上述DYW:PG、DYW:WW或者DYW:KP中的任一者的DYW结构域。

[0257] 本发明的DYW蛋白可以在真核生物(例如动物、植物、微生物(酵母等)、原生生物)的细胞中发挥功能。本发明的DYW蛋白特别是可以在动物细胞内(体外或体内)发挥功能。作为可导入本发明的DYW蛋白、或者表达DYW蛋白的载体的动物细胞,例如可以举出来源于人、猴、猪、牛、马、狗、猫、小鼠、大鼠的细胞。另外,作为可导入本发明的DYW蛋白、或者表达DYW蛋白的载体的培养细胞,例如可以举出中国仓鼠卵巢(CHO)细胞、COS-1细胞、COS-7细胞、VERO(ATCC CCL-81)细胞、BHK细胞、犬肾来源的MDCK细胞、仓鼠AV-12-664细胞、HeLa细胞、WI38细胞、HEK293细胞、HEK293T细胞、PER.C6细胞,但并不限于这些。

[0258] (用途)

[0259] 利用本发明的DYW蛋白,能够将靶标RNA中包含的编辑靶标C转换成U、或者将编辑靶标U转换成C。RNA结合性的PPR蛋白参与了在细胞器中观察到的全部的RNA加工的步骤、切割、RNA编辑、翻译、剪接、RNA稳定化。

[0260] 另外,利用本发明的DYW蛋白,能够进行线粒体的RNA1碱基编辑。线粒体具有单独的基因组,编码与呼吸或ATP生产相关的重要复合体的构成蛋白。已知由于它们的突变会导致各种疾病的发病。通过使用了本发明的突变修复,能够期待对各种疾病的处置。

[0261] 另一方面,在CRISPR-Cas系统中,通过使Cas蛋白与经改造的ADAR结构域融合,开发出了细胞质中的C-U RNA编辑工具(Abudayyeh et al.,2019年)。但是,CRISPR-Cas系统由蛋白质和引导RNA构成,难以对引导RNA进行有效的线粒体运输。另一方面,PPR蛋白能够以1分子进行RNA编辑,通常蛋白质通过在N端侧融合线粒体定位信号序列,能够送达至线粒体中。因此,为了确认本技术是否能够用于线粒体的RNA编辑,设计了以MT-ND2和MT-ND5为靶标的PPR,制作出融合了PG或WW结构域的基因(图10a)。

[0262] 由线粒体靶标序列(MTS)和PPR-P序列构成的这些蛋白质以MT-ND2和MT-ND5的第178位和第301位密码子的第3位的位置作为靶标,以使其不会对HEK293T细胞带来不良影响(图10a)。用质粒导入至HEK293T细胞中后,在mRNA内确认到MT-ND2和MT-ND5的编辑(图10的b、c)。这4种蛋白质对靶标检测到最高为70%的编辑活性,对于相同的mRNA分子未检测到脱靶突变(图10的bc)。

[0263] 因此,由本发明提供的RNA碱基的编辑方法可在各种领域中期待以下的利用。

[0264] (1) 医疗

[0265] • 识别并编辑与特定疾病相关的特定RNA。通过利用本发明,可处置由1个碱基突变所致的遗传病。遗传病中大部分的突变方向是从C到U的突变。因此,特别是能够将U转换成C的本发明的方法是有用的。

[0266] • 制作对RNA的抑制/表达进行了调控的细胞。这样的细胞包括对分化/未分化状

态进行了监控的干细胞(例如iPS细胞);化妆品评价用模型细胞;以阐明药物创制的机理或药理试验为目的、能够进行功能性RNA表达的开/关的细胞。

[0267] (2) 农林水产业

[0268] • 在农作物、林产物、水产物等中,改善产量、品质。

[0269] • 培育抗病性提高、环境耐性提高、具有提高的功能性或新的功能性的生物。

[0270] 例如,关于杂交第一代(F1)作物,利用基于DYW蛋白的线粒体RNA的编辑人工制作出F1作物,具有能够改善产率、品质的可能性。基于DYW蛋白的RNA编辑能够比现有技术更准确且迅速地进行生物的品种改良、育种(对生物进行遗传性改良)。另外,基于DYW蛋白的RNA编辑不是像基因重组那样利用外源基因进行性状转化,因此可以说与突变体的筛选、回交之类的以往的育种方法相近。因此,对于全球规模的粮食问题、环境问题也可确实且迅速地应对。

[0271] (3) 化学

[0272] • 在利用微生物、培养细胞、植物体、动物体(例如昆虫体)的有用物质生产中,通过RNA的操作来调控蛋白表达量。由此可以提高有用物质的生产率。有用物质的示例除了抗体、疫苗、酶等蛋白质性的物质以外,还有药品的中间体、香料、色素等较低分子量的化合物。

[0273] • 通过改变藻类、微生物的代谢途径来改善生物燃料的产生效率。

[0274] [术语]

[0275] 关于数值范围 $x\sim y$,除特别记载的情况以外,包括两端的值 x 和 y 。

[0276] 关于蛋白质、多肽的氨基酸序列,有时将氨基酸残基简称为氨基酸。

[0277] 关于碱基序列(有时也称为核苷酸序列)或氨基酸序列的“一致性”,除特别记载的情况以外,是指将2个序列以最佳方式比对时2个序列间共有的一致性的碱基或氨基酸的个数的百分率。即,可以通过一致性=(一致的位置的数目/位置的总数) $\times 100$ 来算出,可以使用市售的算法进行计算。另外,这样的算法并入在Altschul et al., J.Mol.Biol.215(1990)403-410中记载的NBLAST和XBLAST程序中。更详细地说,碱基序列或氨基酸序列的一致性的相关检索/分析可由本领域技术人员通过公知的算法或程序(例如BLASTN、BLASTP、BLASTX、ClustalW)来进行。使用程序时的参数可以由本领域技术人员适当地设定,另外也可以使用各程序的默认参数。这些分析方法的具体方法也是本领域技术人员公知的。

[0278] 关于碱基序列或氨基酸序列的序列一致性,除特别记载的情况以外,越高越优选。具体地说,优选为40%以上、更优选为45%以上、进一步优选为50%以上、进一步优选为55%以上、进一步优选为60%以上、进一步优选为65%以上。另外,优选为70%以上、更优选为80%以上、进一步优选为85%以上、进一步优选为90%以上、进一步优选为95%以上、进一步优选为97.5%以上。

[0279] 在关于多肽或蛋白质提到“发生了置换、缺失或添加的序列”时,关于被置换等的氨基酸的个数,除特别记载的情况以外,在任一基序或蛋白质中,只要由该氨基酸序列构成的基序或蛋白质具有所期望的功能就均没有特别限定,为1~9个或1~4个左右,或者,若置换为性质相似的氨基酸,则可以具有更多个数的置换等。用于制备这样的氨基酸序列所涉及的多核苷酸或蛋白质的手段是本领域技术人员熟知的。

[0280] 性质相似的氨基酸是指疏/亲水性、荷电、pKa、溶解性等物性相似的氨基酸,例如

是指下述氨基酸。

[0281] 疏水性(非极性)氨基酸:丙氨酸、缬氨酸、甘氨酸、异亮氨酸、亮氨酸、苯丙氨酸、脯氨酸、色氨酸、酪氨酸

[0282] 非疏水性氨基酸:精氨酸、天冬酰胺、天冬氨酸、谷氨酸、谷氨酰胺、赖氨酸、丝氨酸、苏氨酸、半胱氨酸、组氨酸、蛋氨酸;

[0283] 亲水性氨基酸:精氨酸、天冬酰胺、天冬氨酸、谷氨酸、谷氨酰胺、赖氨酸、丝氨酸、苏氨酸;

[0284] 酸性氨基酸:天冬氨酸、谷氨酸;

[0285] 碱性氨基酸:赖氨酸、精氨酸、组氨酸;

[0286] 中性氨基酸:丙氨酸、天冬酰胺、半胱氨酸、谷氨酰胺、甘氨酸、异亮氨酸、亮氨酸、蛋氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸;

[0287] 含硫氨基酸:蛋氨酸、半胱氨酸;

[0288] 含芳香环氨基酸:酪氨酸、色氨酸、苯丙氨酸。

[0289] 实施例

[0290] [1.DYW结构域的设计和大肠杆菌内的RNA编辑活性测定]

[0291] [结果]

[0292] (DYW结构域的设计)

[0293] 目前可利用的植物的基因组信息大部分为被子植物的基因组信息,因此人工DYW蛋白的开发限于DYW:PG型。另一方面,转录组数据尽管为部分序列,但也包含具有C-U和U-C RNA编辑这两者的初期陆生植物种的信息。因此,发明人使用基于由千种植物转录组(1KP)国际联盟制作的转录组数据集构建的PPR蛋白的数据库(上述非专利文献11),设计出了完全的人工DYW结构域。

[0294] 在设计人工DYW结构域时,除了胞苷脱氨酶样DYW结构域以外,还在设计中加入了PPR(P2和L2)基序以及PPR样(S2、E1和E2)基序。这是由于,这些基序对于RNA编辑活性的重要性在现阶段还不清楚。发明人制作了角苔类、石松类和蕨类植物的DYW结构域的系统发育树,分类成DYW:PG、DYW:WW和DYW:KP这3个组。接下来,对于这些各组的DYW结构域重新制作系统发育树,筛选出可用于人工DYW结构域的设计的蛋白质的系统群(图1)。DYW:PG组的序列间的差异大,因此难以选择用于设计的蛋白质的系统群(图1a)。石松类的蛋白质的序列富于多样性、数目也多,因此发明人决定专注于该蛋白质(PG1)。在DYW:WW结构域的设计中,选择了仅在角苔类中观察到的进化枝短的蛋白质的系统群(图1b、WW1)。该系统群中,由于进化枝短,因此可推定蛋白质间的基因的突变小。并且,在DYW:KP结构域的设计中着眼于在蕨类植物中具有特异性的DYW结构域的系统群。该系统群中包含的蛋白质的序列的氨基酸的突变大,但长度具有保守性(图1c、KP1)。

[0295] (RNA结合结构域的设计)

[0296] 发明人将各DYW结构域与基于由植物的基因组信息鉴定的PPR基序序列(上述非专利文献2)设计的人工P或PLS阵列进行了融合(图2a)。P阵列的PPR基序是基于由35个氨基酸的P基序的比对得到的共有序列构建的、为了提高RNA识别而对若干氨基酸进行了置换的PPR基序。在PLS阵列的设计中,选择了标准长度的35个氨基酸的基序(P1和L1)和31个氨基酸的基序(S1)。在天然的PPR蛋白中,位于最开始和最后的P1L1S1与位于特定位置的氨基酸

残基相比观察到明确的差异,与内部的P1L1S1不同。为了设计出尽可能接近天然存在的PPR基序的人工PLS阵列,分成位于第一位(N末端侧)的P1L1S1、位于内部的P1L1S1、以及位于紧靠P2L2S2之前的最后(C末端侧)的P1L1S1,设计出与PPR基序的位置对应的3种P1L1S1。下文中将这些蛋白质基于其结构进行命名。例如,在P阵列上融合DYW:WW结构域而得到的蛋白质称为P-DYW:WW。

[0297] (人工DYW蛋白能够特异性地编辑靶标序列)

[0298] 在拟南芥中,PPR蛋白CLB19识别位于叶绿体rpoA和clpPRNA上的RNA编辑位点。发明人决定设计以rpoA编辑位点作为靶标的PPR蛋白。P1、L1、S1和P2基序中的第4位和第ii位氨基酸按照PPR密码来设计。另一方面,由于与C末端PPR样基序(L2、S2、E1、E2)相关的PPR密码是未知的,因此关于L2、S2、E1和E2基序,使用了CLB19的第4位和第ii位的氨基酸(图2b)。

[0299] 将编码重组人工DYW蛋白的基因区域克隆至表达载体中,在其终止密码子的下游添加靶标序列。基于过去进行的关于小立碗藓的两种PPR蛋白(PPR56、PPR65)的研究(上述非专利文献12),开发出了在大肠杆菌内对所设计的PPR蛋白的RNA编辑活性进行试验的方法。所设计的PLS-DYW:PG1和PLS-DYW:WW1未观察到对于DNA的编辑活性,另一方面,对于RNA,以高于90%的编辑效率将胞苷置换成尿苷(图3的a、b、d)。在使用P阵列代替PLS结构域的情况下,编辑效率降低了10~40%。另一方面,在P-DYW:KP和PLS-DYW:KP这两者中观察到了U-C RNA编辑活性(图3的c、e)。但是,与DYW:PG和DYW:WW相比,其编辑活性降低。

[0300] [方法]

[0301] (系统发育树)

[0302] 从PPR数据库(<https://ppr.plantenergy.uwa.edu.au/onekp/>;上述非专利文献11)中提取包含DYW结构域(最短132个氨基酸)的P2-L2-S2-E1-E2-DYW区域。使用MAFFT L-INS-i(v7.407 automatic mode)(K.Katoh,D.M.Standley(2013).Mol Biol Evol.30(4):772-780.)制作所得到的序列的比对,之后使用trimAl(v.1.4.rev15)(Salvador Capella-Gutierrez,et al.(2009).Bioinformatics.25(15):1972-1973.)进行修剪。修剪中,将参数设为gt 0.2cons 20。在活性位点(HxEx_nCxxCH)具有突变的序列从比对中排除(图4)。

[0303] 使用FastTree(v2.1.10)(Price,M.N.,et al.(2010).PLoS One 5:e9490.)由余下的序列制作系统发育树,对DYW:PG、DYW:WW和DYW:KP进行鉴定。此时的参数为wag和cat 8。

[0304] (Trx-PPR-DYW蛋白和靶标序列的克隆)

[0305] 使用EMBOSS:cons(v.6.6.0.0)设计各DYW结构域(包含P2、L2、S2、E1、E2)的共有序列。关于蛋白质表达用载体,对pET21b+PA进行改造,去除原来的Esp3I和Bpi I限制酶位点,添加2个Esp3I位点作为克隆化位点。基因分成4个片段(Trx、PPR阵列、DYW结构域和RNA编辑位点),使用2阶段的金门(Golden Gate)法来构建。首先,在改造pET21b载体的Esp3I位点进行1)硫氧还蛋白-6×His-TEV基因区域(在3'包含Bpi I限制酶位点)、2)P2-L2-S2-E1-E2-DYW基因区域(在5'包含Bpi I限制酶位点)、以及3)RNA编辑位点的编码序列区域这3部分的克隆。接下来,在Bpi I位点克隆全长PLS结构域或P结构域,制作出PLS:DIW(序列编号35~37)或P:DIW(序列编号38~40)蛋白。

[0306] (在大肠杆菌内的RNA编辑活性测定)

[0307] 为了在大肠杆菌内分析重组蛋白的RNA编辑活性,发明人对于由Oldenkott等人开

发的方案(上述非专利文献12)进行了改良。将上述制作的质粒DNA导入至大肠杆菌Rosetta 2株中,利用1mL的LB培养基(含有羧苄西林50 μ g/mL和氯霉素17 μ g/mL)在37 $^{\circ}$ C培养一夜。在深底24孔板中准备5mL的包含适当的抗生素的LB培养基,在其中接种预培养液100 μ L。使该培养液在37 $^{\circ}$ C、200rpm的条件下生长至吸光度(OD₆₀₀)达到0.4~0.6,之后将培养板在4 $^{\circ}$ C冷却10~15分钟。接着,加入0.4mM的ZnSO₄和0.4mM的IPTG,进一步在16 $^{\circ}$ C、180rpm的条件下培养18小时。采集培养液750 μ L进行离心分离后,将菌体颗粒在液氮中冷冻,保存于-80 $^{\circ}$ C。

[0308] 将冷冻的细胞块重悬于200 μ L的1-硫代甘油/匀浆液中,以40W进行10秒超声波处理,将细胞分离后,加入200 μ L的溶解缓冲液。使用Maxwell(注册商标)RSC simplyRNA组织试剂盒(Promega)提取RNA。将RNA用DNaseI(Takara Bio)进行处理,使用SuperScript(注册商标)III逆转录酶(Invitrogen),利用处理后的RNA1 μ g和1.25 μ M的随机引物(6mer)进行cDNA合成。使用NEBNext超保真2xPCR预混液(New England Biolabs)、1 μ L的cDNA以及硫氧还蛋白和T7终止子序列的引物,对包含编辑位点的区域进行扩增。利用NucleoSpin(注册商标)凝胶和PCR产物纯化试剂盒(Takara Bio)对PCR产物进行纯化,使用DYW结构域序列特异性的正向引物进行测序分析,确定RNA编辑位点的碱基。在RNA编辑效率的测定中,使用编辑位点的C和U的波形峰的高度比,通过 $U/(C+U) \times 100$ 计算出C-U RNA编辑效率,通过 $C/(C+U) \times 100$ 计算出U-C编辑效率。重复进行3次独立的实验。

[0309] [所引用的序列的一览]

[0310] [表3-1]

PLS阵列 (3 P1L1S1三联体)

SEQ ID NO:

P1	FSWNSMIRGYARSGQPEEALSLFSQMRRSGIKPDS	4
L1	YTFPFVLKACASLSSLEEGKQIHAAHVIKSGFESDV	5
S1	YVQSSLIDMYAKCGSLEDARKVFDEMPERNV	6
P1	VSWNAMISGYAQNGQSEEALELFREMQKEGKIPSE	7
L1	FTFCSVLKACASLGSLEMKGQIHGYVIKSGFESIV	8
S1	FVGNALIDMYAKCGSLEDARKVFDEMPERTV	9
P1	VSWTAMISGYAQNGQSEEALELFREMQREGVKPDE	10
L1	VTLPSVLSACANLGALEQGGKQIHAYVIKNGFESDV	11
S1	FVGSALIDMYAKCGSIEEARKVFDEMPEKDV	12

[0311]

P 阵列 (10 PPR 基序)

SEQ ID NO:

P	VTYTTLIDGLCKAGKVDEALELFKEMRSKGVKPNV	13
P	VTYNTLIDGLCKAGRLDEAEELLEEMEEKGIKPDV	14
P	VTYNTLIDGLCKAGRLDEAEELLEEMEEKGIKPDV	14
P	VTYTTLIDGLCKAGKVDEALELFKEMRSKGVKPNV	13
P	VTYNTLIDGLCKSGKIEEALKLFKEMEEKGITPSV	15
P	VTYTTLIDGLCKAGKVDEALELFKEMRSKGVKPNV	13
P	VTYNTLIDGLCKSGKIEEALKLFKEMEEKGITPSV	15
P	VTYTTLIDGLCKAGKVDEALELFDKEMKERGKIPDV	16
P	VTYNTLIDGLCKAGRLDEAEELLEEMEEKGIKPDV	14
P	VTYTTLIDGLCKAGKVDEALELFDKEMKERGKIPDV	16

[0312] [表3-2]

DYW:PG		SEQ ID NO:
P2	VSWNAMIAAYAQHGHGKEALQLFQQMQQEGVKPSE	17
L2	VTFTSILSACSHAGLVDEGHYFESMSPDYGITPRV	18
S2	EHYGCMVDLLGRAGRLDEAEDLIKSMFPQPNV	19
E1	VVWGTLGACRVHGDVERGERAAERILELDPESA	20
E2	APYVLLSNIYAAAGRWDEAAKVRKLMKERGKVK	21
DYW	EPGCSWIEVNNKVHEFVAGDKSHPQTKEIYAELERLS KQMKEAGYVPDTKFVLHDVEEEKEQLLCYHSEKLAI AFGLISTPPGTPLRI IKNLRVCGDCHTATKFIKIVG REIVVRDANRFHHFKDGVCSGDIW	1

[0313]

DYW:WW		SEQ ID NO:
P2	VTWNALIAGYARQGESDLVFHLLERMROEGIQPSG	22
L2	VTFTSVLTVCSHAGLVDEGQKYFDAMSEDYGITPRI	23
S2	EHYGCMVDLLGRAGQMDEAVAMVEKMPFQPNL	24
E1	VTWGTLGACRKNWVVEIGRHAFECVRLDEKSA	25
E2	AAYVLMNSIYADAHMWEERDKIQAMRNARAWK	26
DYW	KPGQSWWTDGIVHTFVVGDTKHPQSQDIYAKLKDL YVKMKEEGYVPHLDCVLWDISDDEKEDALCGHSEKLA IACALINTPPGTPIRIVKNLRVCDCHKAIALISKIE GRNIICRDASRFHNYKDGKCSGDIW	2

DYW:KP		SEQ ID NO:
P2	VAWNSMIKGYGMNHEGKMAVQCFEDMQKQGVKPSA	27
L2	VTFTCLLTACSHASLVSEGQYFKMMREEYGIAPRV	28
S2	EHYGCMVDLLARSGHLYEAEKFLEMLCPPNE	29
E1	GTWGALLSACKTYGEVELGLRCFQQLVQLNPESA	30
E2	AWYVLMADIYAGAGRWDDAYRIEELRKHAGAKK	31
DYW	KPASALIEVNKKVHEFVVGNNQSEEISAMLKSLNSRM KEEGHVPNLDLVLKPVSDDEEKEAALCEHAEKLAIAFG LLNTPQGQTLRVTKNLRMCNDCHNASKIISKIERREI ILRDDCCIHHFKDGLCSCGDMF	3

[0314] [表3-3]

[0315]

<p><PLS 阵列 (3 P1L1S1 三联体) + DYW:PG-lyc> VFSWNSMIRGYARNGQPEEALSLSYQMRSGIKPDNYTFFPVLKACASLSLKEGKQIHGHVIXSGFESDVYVQSALIDMYAKCGE LEDARKVFDEMPERNVSVWNAMISGYAQNQSQSEEALELFREMQQEGIKPSEFTFCSVLSACASLGSLEMGGKQIHGYVIXSGFESIV FVGNALIDMYAKCGSIEDARKVFDEMPERTVVSWTAMISGYAQNQSQSEEALELFREMQRREGVKPDEVTLPSVLSACANLGALEQGGK QIHAYVIXSGFESDVYVQSALIDMYAKCGSIEDARKVFDKMPERDVVSWNAMIAAYAQHGHGKEALQLFQQMQQEGVKPSEVTFTS ILSACSHAGLVDEGHHYFESMSPDYGITPRVEHYGCMVDLLGRAGRLDEAEDLIKSMFPQPNVVVWGTLLGACRVHGDVERGERAA ERILELDPEAAFPYVLLSNIYAAAGRWDEAAKVRKLMKERGVKKEPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMK AGYVPDTKFLVLDHVEEEKEQLLCYHSEKLAIAFGLISTPPGTPLRRIKKNLRVCGDCHTATKFIISKIVGREIVVRDANRFHFKD VCSCGDYW (SEQ ID NO:35)</p>
<p><PLS 阵列 (3 P1L1S1 三联体) + DYW:WW-WW5> VFSWNSMIRGYARNGQPEEALSLSYQMRSGIKPDNYTFFPVLKACASLSLKEGKQIHGHVIXSGFESDVYVQSALIDMYAKCGE LEDARKVFDEMPERNVSVWNAMISGYAQNQSQSEEALELFREMQQEGIKPSEFTFCSVLSACASLGSLEMGGKQIHGYVIXSGFESIV FVGNALIDMYAKCGSIEDARKVFDEMPERTVVSWTAMISGYAQNQSQSEEALELFREMQRREGVKPDEVTLPSVLSACANLGALEQGGK QIHAYVIXSGFESDVYVQSALIDMYAKCGSIEDARKVFDKMPERDVVWNALIAAGYARQGESDLVHLLERMQRREGIQPSGVTFTS VLTVCSHAGLVDEGQKYFDAMSEDYGITPRIEHYGCMVDLLGRAGQMQDEAVAMVEKMPFQPNLVTWGTLLGACRVHGDVERGERAA ECAVRLDEKSAAYVLMNSIYADAHMWEERDKIQAMRKNARAWKPKGQSWWTDGIVHTFVVGDTKHPQSQDIYAKLKDLYVKMK EEGYVPHLDCVLDISDDEKEDALCGHSEKLAIAACALINTPPGTPRIRIVKNLRVCGDCHKAIALISKIEGRNIIICRDASRFHNYK GKCSGDYW (SEQ ID NO:36)</p>
<p><PLS 阵列 (3 P1L1S1 三联体) + DYW:KP-lepto> VFSWNSMIRGYARNGQPEEALSLSYQMRSGIKPDNYTFFPVLKACASLSLKEGKQIHGHVIXSGFESDVYVQSALIDMYAKCGE LEDARKVFDEMPERNVSVWNAMISGYAQNQSQSEEALELFREMQQEGIKPSEFTFCSVLSACASLGSLEMGGKQIHGYVIXSGFESIV FVGNALIDMYAKCGSIEDARKVFDEMPERTVVSWTAMISGYAQNQSQSEEALELFREMQRREGVKPDEVTLPSVLSACANLGALEQGGK QIHAYVIXSGFESDVYVQSALIDMYAKCGSIEDARKVFDKMPERDVVAWNSMIKGYGMNHEGKMAVQC FEDMQQGVKPSAVTFTC LLTACSHASLVSEGQYFKMMREEYGIAPRVEHYGCMVDLLARSGLHYEAEKFLMELCPPNEGTVGALLSACKTYGEVELGLRCFQ QLVQLNPESAAYVLMADIYAGAGRWDAYRIEELRKHAGAKKPPASALIEVNNKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGH PNLDLVLKPVSDKEEAALCEHAEKLAIAFGLLNTPQQQLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHFFKDLGCLSC GDMF (SEQ ID NO:37)</p>
<p><P 阵列 (10 PPR 基序) + DYW:PG-lyc> VVVYTTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGR LDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGI TPSVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGITPSVVYTYNTLLIDGLCK AGKVDLEALELFDEMKEGKIPDVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFDEMKE RGIKPDVVSWNAMIAAYAQHGHGKEALQLFQQMQQEGVKPSEVTFTSILSACSHAGLVDEGHHYFESMSPDYGITPRVEHYGCMVD LLGRAGRLDEAEDLIKSMFPQPNVVVWGTLLGACRVHGDVERGERAAERILELDPEAAFPYVLLSNIYAAAGRWDEAAKVRKLMKE RGVKKEPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKKEAGYVPDTKFLVLDHVEEEKEQLLCYHSEKLAIAFGLIST PPGTPLRRIKKNLRVCGDCHTATKFIISKIVGREIVVRDANRFHFKDGVCSGDYW (SEQ ID NO:38)</p>
<p><P 阵列 (10 PPR 基序) + DYW:WW-WW5> VVVYTTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGR LDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGI TPSVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGITPSVVYTYNTLLIDGLCK AGKVDLEALELFDEMKEGKIPDVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFDEMKE RGIKPDVVTWNALIAAGYARQGESDLVHLLERMQRREGIQPSGVTFTSVLTVCSHAGLVDEGQKYFDAMSEDYGITPRIEHYGCMVD LLGRAGQMQDEAVAMVEKMPFQPNLVTWGTLLGACRVHGDVERGERAAERILELDPEAAFPYVLLSNIYADAHMWEERDKIQAMRKN ARAWKPKGQSWWTDGIVHTFVVGDTKHPQSQDIYAKLKDLYVKMKKEEGYVPHLDCVLDISDDEKEDALCGHSEKLAIAACALIN TPPGTPRIRIVKNLRVCGDCHKAIALISKIEGRNIIICRDASRFHNYKDGKCSGDYW (SEQ ID NO:39)</p>
<p><P 阵列 (10 PPR 基序) + DYW:KP-lepto> VVVYTTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGR LDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGI TPSVVYTYNTLLIDGLCKAGKVDLEALELFKEMRSKGVKPNVVYTYNTLLIDGLCKSGKIEEALKLFKEMEEKGITPSVVYTYNTLLIDGLCK AGKVDLEALELFDEMKEGKIPDVVYTYNTLLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVYTYNTLLIDGLCKAGKVDLEALELFDEMKE RGIKPDVVAWNSMIKGYGMNHEGKMAVQC FEDMQQGVKPSAVTFTCLLACSHASLVSEGQYFKMMREEYGIAPRVEHYGCMVD LLARSGLHYEAEKFLMELCPPNEGTVGALLSACKTYGEVELGLRCFQQLVQLNPESAAYVLMADIYAGAGRWDAYRIEELRKH GAKKPPASALIEVNNKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSDKEEAALCEHAEKLAIAFGLLNTPQQ TLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHFFKDLGCLSCGDMF (SEQ ID NO:40)</p>

[0316] [2. 在动物培养细胞中的实施例]

[0317] [结果]

[0318] 将PLS型PPR与各DYW结构域 (PG1 (序列编号1)、WW1 (序列编号2)、KP1 (序列编号3)) 进行融合,将包含所得到的基因和靶标序列的质粒转染到HEK293T细胞中,培养后回收RNA。通过桑格测序来分析靶标位点的从胞苷 (C) 到尿苷 (U) 的转换效率、或者从尿苷 (U) 到胞苷 (C) 的转换效率 (图5)。在融合了PG1或WW1结构域的情况下,具有90%以上的C向U的活性,另一方面,未检测到U向C的活性 (图5的a、c)。在融合了KP1结构域的情况下,检测到25%的U向C的活性,还检测到10%左右的C向U的活性 (图5b、c)。由这些可知,在动物培养细胞中,编辑

酶也发挥功能。

[0319] [方法]

[0320] (动物培养细胞试验用PPR表达质粒的制作)

[0321] 通过PCR由图3中使用的质粒扩增包含6xHis-PPR-DYW蛋白(与大肠杆菌中的实验所用的蛋白相同的蛋白(序列编号35~37))基因序列和编辑位点的区域(序列编号34),通过金门组装法克隆到动物细胞表达载体中。载体通过包含CMV启动子和人 β -珠蛋白嵌合内含子的启动子的控制而表达PPR,通过SV40多聚腺苷酸化信号被赋予多聚A信号。

[0322] (HEK293T细胞培养)

[0323] HEK293T细胞使用在包含高葡萄糖、谷氨酰胺、酚红、丙酮酸钠(富士胶片和光纯药株式会社)的杜氏改良伊格尔培养(DMEM)中加入有10%胎牛血清(Capricorn)和1%青霉素-链霉素(富士胶片和光纯药株式会社)的培养基在37℃、5%CO₂的条件下进行培养。在细胞达到80-90%融合的时刻每隔2-3天进行传代。

[0324] (转染)

[0325] 在RNA编辑分析中,在24孔平底细胞培养板的各孔中加入约 8.0×10^4 个HEK293T细胞,在37℃、5%CO₂的条件下培养24小时。在各孔中加入500ng的质粒、18.5 μ l的Opti-MEM(注册商标)I减血清培养基(ThermoFisher)和1.5 μ l的FuGENE(注册商标)HD转染试剂(Promega),使最终容量为25 μ l。混合液在添加至细胞中之前在室温下温育10分钟。转染24小时后回收细胞。

[0326] (RNA提取、逆转录、测序)

[0327] 与上述大肠杆菌中的分析同样地进行。需要说明的是,以下的实施例中,除特别记载的情况以外,利用与大肠杆菌中的分析和本实施例相同的方法进行实验。

[0328] [3. 结构域交换对RNA编辑活性的影响]

[0329] DYW结构域根据氨基酸序列的保守性可以被划分成若干个区域,但它们与RNA编辑活性的关联性是未知的。此处,将KP1与WW1结构域的一部分进行交换,调查在HEK293T细胞中对RNA编辑活性的影响(图6)。可知在将WW1结构域的PG盒和DYW与KP1结构域的包含活性位点的中央部位融合而得到的序列(chimKP1a序列编号90)中,确认到高于KP1结构域的接近50%的U向C活性,此外C向U活性几乎消失。通过结构域交换,成功地改善了KP结构域的U-C编辑性能。

[0330] [4. 突变导入对KP结构域的RNA编辑活性的改善]

[0331] 在KP结构域中导入各种突变,旨在提高KP结构域的RNA编辑活性。设计KP2~KP23(序列编号68~89),利用大肠杆菌(图7a)和HEK293T细胞(图7的b、c)调查C-U或U-C的RNA编辑活性。KP22(序列编号88)对U-C编辑活性最高、对C-U的编辑活性最低,与KP1相比成功地改善了RNA编辑活性。

[0332] [5. 突变导入对PG结构域的RNA编辑活性的改善]

[0333] 在PG结构域中导入各种突变,旨在提高PG结构域的RNA编辑活性。设计PG2~PG13(序列编号41~53),利用大肠杆菌(图8a)和HEK293T细胞(图8的b、c)调查C-U的RNA编辑活性。PG11(序列编号50)对C-U的编辑活性最高,成功地改善了RNA编辑活性。

[0334] [6. 突变导入对WW结构域的RNA编辑活性的改善]

[0335] 在WW结构域中导入各种突变,旨在提高WW结构域的RNA编辑活性。设计WW2~WW14,

利用大肠杆菌(图9a)和HEK293T细胞(图9的b、c)调查C-U的RNA编辑活性。WW11(序列编号63)对C-U的编辑活性最高,与WW1相比成功地改善了RNA编辑活性。

[0336] [7.使用了PPR蛋白的人线粒体RNA编辑]

[0337] [结果]

[0338] 线粒体具有单独的基因组,编码与呼吸、ATP生产相关的重要复合体的构成蛋白。已知由于它们的突变会导致各种疾病的发病,正在寻求突变修复方法。

[0339] 在CRISPR-Cas系统中,通过使Cas蛋白与经改造的ADAR结构域融合,开发出了细胞质中的C-U RNA编辑工具(Abudayyeh et al.2019Science Vol.365,Issue 6451,pp.382-386)。但是,CRISPR-Cas系统由蛋白质和引导RNA构成,难以对引导RNA进行有效的线粒体传输。另一方面,PPR蛋白能够以1分子进行RNA编辑,通常蛋白质通过在N端侧融合线粒体定位信号序列,能够送达至线粒体中。因此,为了确认本技术是否能够用于线粒体的RNA编辑,设计了以MT-ND2和MT-ND5为靶标的PPR,制作出融合了PG1或WW1结构域的基因(图10a)。

[0340] 由线粒体靶标序列(MTS)和P-DYW序列构成的这些蛋白质以MT-ND2和MT-ND5的第178位和第301位密码子的第3位的位置作为靶标,以使其不会对HEK293T细胞带来不良影响(图10a)。用质粒导入至HEK293T细胞中后,在mRNA内确认到MT-ND2和MT-ND5的编辑(图10的b、c)。这4种蛋白质对于靶标检测到最高为70%的编辑活性,对于相同的mRNA分子未检测到脱靶突变(图10的bc)。

[0341] [方法]

[0342] (用于线粒体编辑的克隆化)

[0343] 将来自玉米的LOC100282174蛋白的线粒体靶标序列(Chin et al.2018)、10个PPR-P和PPR样基序、以及DYW结构域部分(DYW结构域使用PG1和WW1)通过金门组装克隆到CMV启动子控制下的表达质粒中(序列编号91~94)。

[0344] [引用的序列一览]

[0345] [表4-1]

[0346]

	序列	SEQ ID NO:
PG1	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	1
PG2	QPGCSSIEVDNKVHEFVIGDRSHPTKEIYAELERLNKMKEAGYVVDTRLVLDVDEEEKEHLLCHHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	41
PG3	KPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	42
PG4	EPACSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	43
PG5	EPGCAWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	44
PG6	EPGCSTIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	45
PG7	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLNKMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	46
PG8	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	47
PG9	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHFDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	48
PG10	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRVTKNLRVCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	49
PG11	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRMCGDCHTATKFI SKIVGREIVVRDANRFHHFKDGVCSGDIY	50
PG12	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVKREIVVRDANRFHHFKDGVCSGDIY	51
PG13	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDDCRFHHFKDGVCSGDIY	52
PG14	EPGCSWIEVNNKVHEFVAGDKSHPTKEIYAELERLSKQMKEAGYVVDTKFVLHDVEEEKEQLLCYHSE KLAIAFGLISTPPGTPLRIKKNLRVCGDCHTATKFI SKIVGREIVVRDANCVHHFKDGVCSGDIY	53

[0347]

[表4-2]

	序列	SEQ ID NO:
WW1	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	2
WW2	QPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAAYALLKTPPEGTTIRVVKNLRVCGDCHTATALISKIEQRTIICRDASRFHNYKDGKCSGDIY	54
WW3	EPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	55
WW4	KPAQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	56
WW5	KPGQAWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	57
WW6	KPGQSTWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	58
WW7	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS KLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	59
WW8	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	60
WW9	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	61
WW10	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	62
WW11	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVTKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	63
WW12	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRMCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	64
WW13	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	65
WW14	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASRFHNYKDGKCSGDIY	66
WW15	KPGQSWWTDGTVVHTFVVGDTKHPQSQDIYAKLKDLYVKMKEEGYVPHLDCVLWDI SDDEKEDALCGHS EKLAIAACALINTPPGTPIRIVKNLRVCGDCHKAIALISKIEGRNIICRDASCVHNYKDGKCSGDIY	67

[0348]

[0349]

[表4-3]

[0350]

	序列	SEQ ID NO:
KP1	KPASALIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	3
KP2	KPAVAGIEVDRHVHTFVVGEEGRDISCKVRSMNTRLKEEGGHVPHTLVLKPLSEKEKEDEL CGHAEKLA LAYGLLNT PDGTPLLVTKNLRMCRDCHSSTEIMSRIEKREIVVRDAHRVHRFVDGSCSCGGRP	68
KP3	KPAVATIEVNKQVHRFLVGERREDIALKVESTNARLKQEGGHVPHTKLVLPCEKEKEDALCGHAEKLA LAFGLLNTPSGTPLVVTKNLRMCSDCHSSTEIMSRLEQRDIVVRDGYRVHRFAAGCCSCKGRP	69
KP4	KPAVATIEVNKQVHRFLVGERREDIALKVESTNARLKQEGGHVPHTKLVLPCEKEKEDVLCGHAEKLA LAFGLLNTPSGTPLVVTKNLRMCSDCHSSTEIMSRLEQRDIVVRDGYRVHRFANGCCSCKGRP	70
KP5	KPGSALIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	71
KP6	KPAQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	72
KP7	KPASSLIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	73
KP8	KPGQAWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	74
KP9	KPASAWIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	75
KP10	KPGQSTWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	76
KP11	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	77
KP12	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	78
KP13	KPASALIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGYVNPDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	79
KP14	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGYVNPDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	80
KP15	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKDVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	81
KP16	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRIVKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	82
KP17	KPASALIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIEGREIILRDDCCIHHFKDGLCSCGDMF	83
KP18	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIEGREIILRDDCCIHHFKDGLCSCGDY W	84
KP19	KPASALIEVNKKVHEFVVGNNQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDASCIHHFKDGLCSCGDMF	85
KP20	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDASCIHHFKDGLCSCGDY W	86
KP21	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	87
KP22	KPGQSWWTD TGGIVHTFVVGDTKHPQSQDIYAKLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDMF	88
KP23	KPGQSWWTD TGGIVHTFVVGDTKHPQSQDIYAKLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	89
chim	KPGQSWWTD TGGIVHTFVVGDTKHPQSEESI SAMLKSLNSRMKEEGHVPNLDLVLKPVSD EEEKAAALCEHA	90
KP1a	EKLA IAFGLLNT PQGQTLRVTKNLRMCNDCHNASKIISKIERREIILRDDCCIHHFKDGLCSCGDY W	

[0351] [表4-4]

[0352]

	序列	SEQ ID NO:
MTS-P-PGI_ND 2	MALLRAAVSELRRRGRGALTPLPALSSLLSSLSRSPASTRPEPNNPHADRRHVIALRRCPPLPASAVLAP ELLHARGLLPRHWSHASPLSTSSSSSRPADKAQLTWVDKWIPEAARPYSHGSGGSGGLHMGSGGLSSGDY KDHDGDYKDHDIDYKDDDDKSGSGSGSGNSVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYN TLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYTT LIDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKAGVDEALELFDEMKERGIKPDVVTYNTL IDGLCKSGKIEEALKLFKEMEKEGITPSVVTYNTLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYTTLI DGLCKAGVDEALELFDEMKERGIKPDVVSWNAMIAAAYAQHGHGKEALQLFQQMQQEGVVKPSVVTFTSILSA CSHAGLVDEGHHYFESMSPDYGITPRVEHYGCMVDLLGRAGRLDEAEDLIKSMFPQPNVVVWGTLLGACRV HGDVERGERAAERILELDPESAAPYVLLSNIYAAAGRWEAAKVRKLMKERGVKKEPGCSWIEVNNKVHEF VAGDKSHPQTKEIYAELERLSKQMKAEAGYVDPDKFVLHDVVEEKEQLLCYHSEKLAIAFGLISTPPGTP RIKKNLRVCGDCHTATKFISKIVGREIVVRDANRFHFKDGVCSGDIY	93
MTS-P-PGI_ND 2	MALLRAAVSELRRRGRGALTPLPALSSLLSSLSRSPASTRPEPNNPHADRRHVIALRRCPPLPASAVLAP ELLHARGLLPRHWSHASPLSTSSSSSRPADKAQLTWVDKWIPEAARPYSHGSGGSGGLHMGSGGLSSGDY KDHDGDYKDHDIDYKDDDDKSGSGSGSGNSVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYN TLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYTT LIDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKAGVDEALELFDEMKERGIKPDVVTYNTL IDGLCKSGKIEEALKLFKEMEKEGITPSVVTYNTLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYTTLI DGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYTTLI GLCKAGVDEALELFDEMKERGIKPDVVTWNALAGYARQGESDLVHLLERMRQEGIQPSVVTFTSILSA CSHAGLVDEGQKYFDAMSEDYGITPRIEHYGCMVDLLGRAGQDEAVAMVEKMPFQPNLVWGTLLGACRV WNNVEIGRHAFCACAVRLDEKSAAYVLLSNIYADAHMWEERDKIQAMRKNARAWKPKQSWWTDGIVHT FVVGDTKHPQSDIYAKLDLYVKMKEEGYVPHLDCVLWDISDDEKEDALCGHSEKLAIAACALINTPPGTP IRIVKNLRVCGDCHKAIALISKIEGRNICRDSRFRHNYKDGKCSGDIY	94
MTS-P-PGI_ND 5	MALLRAAVSELRRRGRGALTPLPALSSLLSSLSRSPASTRPEPNNPHADRRHVIALRRCPPLPASAVLAP ELLHARGLLPRHWSHASPLSTSSSSSRPADKAQLTWVDKWIPEAARPYSHGSGGSGGLHMGSGGLSSGDY KDHDGDYKDHDIDYKDDDDKSGSGSGSGNSVVTYNTLIDGLCKAGVDEALELFKEMRSKGVKPNVVTYN TLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYTTLIDGLCKAGVDEALELFDEMKERGIKPDVVTYTT LIDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYTT IDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYNTLI DGLCKSGKIEEALKLFKEMEKEGITPSVVTYTTLIDGLCKAGVDEALELFKEMRSKGVKPNVVTYTTLI GLCKAGVDEALELFKEMRSKGVKPNVSWTAMIAAAYAQHGHGKEALQLFQQMQQEGVKNVVTFTSILSA CSHAGLVDEGHHYFESMSPDYGITPRVEHYGCMVDLLGRAGRLDEAEDLIKSMFPQPNVVVWGTLLGACRV HGDVERGERAAERILELDPESAAPYVLLSNIYAAAGRWEAAKVRKLMKERGVKKEPGCSWIEVNNKVHEF VAGDKSHPQTKEIYAELERLSKQMKAEAGYVDPDKFVLHDVVEEKEQLLCYHSEKLAIAFGLISTPPGTP RIKKNLRVCGDCHTATKFISKIVGREIVVRDANRFHFKDGVCSGDIY	95
MTS-P-WWI_ND 5	MALLRAAVSELRRRGRGALTPLPALSSLLSSLSRSPASTRPEPNNPHADRRHVIALRRCPPLPASAVLAP ELLHARGLLPRHWSHASPLSTSSSSSRPADKAQLTWVDKWIPEAARPYSHGSGGSGGLHMGSGGLSSGDY KDHDGDYKDHDIDYKDDDDKSGSGSGSGNSVVTYNTLIDGLCKAGVDEALELFKEMRSKGVKPNVVTYN TLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYTTLIDGLCKAGVDEALELFDEMKERGIKPDVVTYTT LIDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKSGKIEEALKLFKEMEKEGITPSVVTYTT IDGLCKAGVDEALELFKEMRSKGVKPNVVTYNTLIDGLCKAGRLDEAEELLEEMEEKGKIPDVVTYNTLI DGLCKSGKIEEALKLFKEMEKEGITPSVVTYTTLIDGLCKAGVDEALELFKEMRSKGVKPNVVTYTTLI GLCKAGVDEALELFKEMRSKGVKPNVVTWTALAGYARQGESDLVHLLERMRQEGIQPNVVTFTSVLTV CSHAGLVDEGQKYFDAMSEDYGITPRIEHYGCMVDLLGRAGQDEAVAMVEKMPFQPNLVWGTLLGACRV WNNVEIGRHAFCACAVRLDEKSAAYVLLSNIYADAHMWEERDKIQAMRKNARAWKPKQSWWTDGIVHT FVVGDTKHPQSDIYAKLDLYVKMKEEGYVPHLDCVLWDISDDEKEDALCGHSEKLAIAACALINTPPGTP IRIVKNLRVCGDCHKAIALISKIEGRNICRDSRFRHNYKDGKCSGDIY	96

<211> 137
 <212> PRT
 <213> 人工序列
 <220>
 <223> DYW:WW
 <400> 2
 Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala
 20 25 30
 Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro
 35 40 45
 His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60
 Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile
 65 70 75 80
 Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val
 85 90 95
 Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly
 100 105 110
 Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp
 115 120 125
 Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 3
 <211> 133
 <212> PRT
 <213> 人工序列
 <220>
 <223> DYW:KP
 <400> 3
 Lys Pro Ala Ser Ala Leu Ile Glu Val Asn Lys Lys Val His Glu Phe
 1 5 10 15
 Val Val Gly Asn Asn Gln Ser Glu Glu Ile Ser Ala Met Leu Lys Ser
 20 25 30
 Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro Asn Leu Asp Leu
 35 40 45
 Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala Ala Leu Cys Glu
 50 55 60

<213> 人工序列

<220>

<223> S1基序

<400> 6

Tyr Val Gln Ser Ser Leu Ile Asp Met Tyr Ala Lys Cys Gly Ser Leu

1 5 10 15

Glu Asp Ala Arg Lys Val Phe Asp Glu Met Pro Glu Arg Asn Val

 20 25 30

<210> 7

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P1基序

<400> 7

Val Ser Trp Asn Ala Met Ile Ser Gly Tyr Ala Gln Asn Gly Gln Ser

1 5 10 15

Glu Glu Ala Leu Glu Leu Phe Arg Glu Met Gln Lys Glu Gly Ile Lys

 20 25 30

Pro Ser Glu

35

<210> 8

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> L1基序

<400> 8

Phe Thr Phe Cys Ser Val Leu Lys Ala Cys Ala Ser Leu Gly Ser Leu

1 5 10 15

Glu Met Gly Lys Gln Ile His Gly Tyr Val Ile Lys Ser Gly Phe Glu

 20 25 30

Ser Ile Val

35

<210> 9

<211> 31

<212> PRT

<213> 人工序列

<220>

<223> S1基序

<400> 9

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

<210> 10

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P1基序

<400> 10

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

Pro Asp Glu

35

<210> 11

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> L1基序

<400> 11

Val Thr Leu Pro Ser Val Leu Ser Ala Cys Ala Asn Leu Gly Ala Leu
 1 5 10 15
 Glu Gln Gly Lys Gln Ile His Ala Tyr Val Ile Lys Asn Gly Phe Glu
 20 25 30

Ser Asp Val

35

<210> 12

<211> 31

<212> PRT

<213> 人工序列

<220>

<223> S1基序

<400> 12

Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr
 20 25 30

Pro Ser Val
 35

<210> 16

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P基序

<400> 16

Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val
 1 5 10 15

Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys
 20 25 30

Pro Asp Val
 35

<210> 17

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P2基序

<400> 17

Val Ser Trp Asn Ala Met Ile Ala Ala Tyr Ala Gln His Gly His Gly
 1 5 10 15

Lys Glu Ala Leu Gln Leu Phe Gln Gln Met Gln Gln Glu Gly Val Lys
 20 25 30

Pro Ser Glu
 35

<210> 18

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> L2基序

<400> 18

Val Thr Phe Thr Ser Ile Leu Ser Ala Cys Ser His Ala Gly Leu Val
 1 5 10 15

Asp Glu Gly His His Tyr Phe Glu Ser Met Ser Pro Asp Tyr Gly Ile
 20 25 30

Thr Pro Arg Val
 35

<210> 19

<211> 32

<212> PRT

<213> 人工序列

<220>

<223> S2基序

<400> 19

Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala Gly Arg Leu
 1 5 10 15

Asp Glu Ala Glu Asp Leu Ile Lys Ser Met Pro Phe Gln Pro Asn Val
 20 25 30

<210> 20

<211> 34

<212> PRT

<213> 人工序列

<220>

<223> E1基序

<400> 20

Val Val Trp Gly Thr Leu Leu Gly Ala Cys Arg Val His Gly Asp Val
 1 5 10 15

Glu Arg Gly Glu Arg Ala Ala Glu Arg Ile Leu Glu Leu Asp Pro Glu
 20 25 30

Ser Ala

<210> 21

<211> 33

<212> PRT

<213> 人工序列

<220>

<223> E2基序

<400> 21

Ala Pro Tyr Val Leu Leu Ser Asn Ile Tyr Ala Ala Ala Gly Arg Trp
 1 5 10 15

Asp Glu Ala Ala Lys Val Arg Lys Leu Met Lys Glu Arg Gly Val Lys
 20 25 30

Lys

<210> 22

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P2基序

<400> 22

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

<210> 23

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> L2基序

<400> 23

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

<210> 24

<211> 32

<212> PRT

<213> 人工序列

<220>

<223> S2基序

<400> 24

Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala Gly Gln Met
 1 5 10 15
 Asp Glu Ala Val Ala Met Val Glu Lys Met Pro Phe Gln Pro Asn Leu
 20 25 30

<210> 25

<211> 34

<212> PRT

<213> 人工序列

<220>

<223> E1基序

<400> 25

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

Ser Ala

<210> 26

<211> 33

<212> PRT

<213> 人工序列

<220>

<223> E2基序

<400> 26

Ala	Ala	Tyr	Val	Leu	Met	Ser	Asn	Ile	Tyr	Ala	Asp	Ala	His	Met	Trp
1				5					10					15	
Glu	Glu	Arg	Asp	Lys	Ile	Gln	Ala	Met	Arg	Lys	Asn	Ala	Arg	Ala	Trp
			20					25					30		

Lys

<210> 27

<211> 35

<212> PRT

<213> 人工序列

<220>

<223> P2基序

<400> 27

Val	Ala	Trp	Asn	Ser	Met	Ile	Lys	Gly	Tyr	Gly	Met	Asn	His	Glu	Gly
1				5					10					15	
Lys	Met	Ala	Val	Gln	Cys	Phe	Glu	Asp	Met	Gln	Lys	Gln	Gly	Val	Lys
			20					25					30		

Pro Ser Ala

35

<210> 28

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> L2基序

<400> 28

Val Thr Phe Thr Cys Leu Leu Thr Ala Cys Ser His Ala Ser Leu Val
 1 5 10 15
 Ser Glu Gly Gln Glu Tyr Phe Lys Met Met Arg Glu Glu Tyr Gly Ile
 20 25 30
 Ala Pro Arg Val
 35

<210> 29

<211> 31

<212> PRT

<213> 人工序列

<220>

<223> S2基序

<400> 29

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

<210> 30

<211> 34

<212> PRT

<213> 人工序列

<220>

<223> E1基序

<400> 30

Gly Thr Trp Gly Ala Leu Leu Ser Ala Cys Lys Thr Tyr Gly Glu Val
 1 5 10 15
 Glu Leu Gly Leu Arg Cys Phe Gln Gln Leu Val Gln Leu Asn Pro Glu
 20 25 30

Ser Ala

<210> 31

<211> 33

<212> PRT

<213> 人工序列

<220>

<223> E2基序

<400> 31

Ala Trp Tyr Val Leu Met Ala Asp Ile Tyr Ala Gly Ala Gly Arg Trp
 1 5 10 15
 Asp Asp Ala Tyr Arg Ile Glu Glu Leu Arg Lys His Ala Gly Ala Lys
 20 25 30

Lys

<210> 32

<211> 7

<212> PRT

<213> 猴病毒40

<400> 32

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 33

<211> 16

<212> PRT

<213> 人工序列

<220>

<223> 核质蛋白NLS

<400> 33

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 34

<211> 17

<212> RNA

<213> 人工序列

<220>

<223> 靶标序列

<400> 34

auuacacgug caaaauy

<210> 35

<211> 610

<212> PRT

<213> 人工序列

<220>

<223> PLS阵列(3 P1L1S1三联体) + DYW:PG-lyc

<400> 35

Val Phe Ser Trp Asn Ser Met Ile Arg Gly Tyr Ala Arg Asn Gly Gln

1 5 10 15

Pro Glu Glu Ala Leu Ser Leu Tyr Ser Gln Met Arg Arg Ser Gly Ile

17

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

Pro Ser Glu Val Thr Phe Thr Ser Ile Leu Ser Ala Cys Ser His Ala
 340 345 350
 Gly Leu Val Asp Glu Gly His His Tyr Phe Glu Ser Met Ser Pro Asp
 355 360 365
 Tyr Gly Ile Thr Pro Arg Val Glu His Tyr Gly Cys Met Val Asp Leu
 370 375 380
 Leu Gly Arg Ala Gly Arg Leu Asp Glu Ala Glu Asp Leu Ile Lys Ser
 385 390 395 400
 Met Pro Phe Gln Pro Asn Val Val Val Trp Gly Thr Leu Leu Gly Ala
 405 410 415
 Cys Arg Val His Gly Asp Val Glu Arg Gly Glu Arg Ala Ala Glu Arg
 420 425 430
 Ile Leu Glu Leu Asp Pro Glu Ser Ala Ala Pro Tyr Val Leu Leu Ser
 435 440 445
 Asn Ile Tyr Ala Ala Ala Gly Arg Trp Asp Glu Ala Ala Lys Val Arg
 450 455 460
 Lys Leu Met Lys Glu Arg Gly Val Lys Lys Glu Pro Gly Cys Ser Trp
 465 470 475 480
 Ile Glu Val Asn Asn Lys Val His Glu Phe Val Ala Gly Asp Lys Ser
 485 490 495
 His Pro Gln Thr Lys Glu Ile Tyr Ala Glu Leu Glu Arg Leu Ser Lys
 500 505 510
 Gln Met Lys Glu Ala Gly Tyr Val Pro Asp Thr Lys Phe Val Leu His
 515 520 525
 Asp Val Glu Glu Glu Glu Lys Glu Gln Leu Leu Cys Tyr His Ser Glu
 530 535 540
 Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser Thr Pro Pro Gly Thr Pro
 545 550 555 560
 Leu Arg Ile Ile Lys Asn Leu Arg Val Cys Gly Asp Cys His Thr Ala
 565 570 575
 Thr Lys Phe Ile Ser Lys Ile Val Gly Arg Glu Ile Val Val Arg Asp
 580 585 590
 Ala Asn Arg Phe His His Phe Lys Asp Gly Val Cys Ser Cys Gly Asp
 595 600 605
 Tyr Trp
 610
 <210> 36
 <211> 611
 <212> PRT

<213> 人工序列

<220>

<223> PLS阵列(3 P1L1S1三联体) + DYW:WW-WW5

<400> 36

Val	Phe	Ser	Trp	Asn	Ser	Met	Ile	Arg	Gly	Tyr	Ala	Arg	Asn	Gly	Gln
1				5					10					15	
Pro	Glu	Glu	Ala	Leu	Ser	Leu	Tyr	Ser	Gln	Met	Arg	Arg	Ser	Gly	Ile
			20					25					30		
Lys	Pro	Asp	Asn	Tyr	Thr	Phe	Pro	Phe	Val	Leu	Lys	Ala	Cys	Ala	Ser
		35					40					45			
Leu	Ser	Ser	Leu	Lys	Glu	Gly	Lys	Gln	Ile	His	Gly	His	Val	Ile	Lys
	50					55					60				
Ser	Gly	Phe	Glu	Ser	Asp	Val	Tyr	Val	Gln	Ser	Ala	Leu	Ile	Asp	Met
65					70					75					80
Tyr	Ala	Lys	Cys	Gly	Glu	Leu	Glu	Asp	Ala	Arg	Lys	Val	Phe	Asp	Glu
				85					90					95	
Met	Pro	Glu	Arg	Asn	Val	Val	Ser	Trp	Asn	Ala	Met	Ile	Ser	Gly	Tyr
			100						105					110	
Ala	Gln	Asn	Gly	Gln	Ser	Glu	Glu	Ala	Leu	Glu	Leu	Phe	Arg	Glu	Met
			115						120				125		
Gln	Gln	Glu	Gly	Ile	Lys	Pro	Ser	Glu	Phe	Thr	Phe	Cys	Ser	Val	Leu
						135						140			
Ser	Ala	Cys	Ala	Ser	Leu	Gly	Ser	Leu	Glu	Met	Gly	Lys	Gln	Ile	His
145					150						155				160
Gly	Tyr	Val	Ile	Lys	Ser	Gly	Phe	Glu	Ser	Ile	Val	Phe	Val	Gly	Asn
				165						170				175	
Ala	Leu	Ile	Asp	Met	Tyr	Ala	Lys	Cys	Gly	Ser	Ile	Glu	Asp	Ala	Arg
			180						185				190		
Lys	Val	Phe	Asp	Glu	Met	Pro	Glu	Arg	Thr	Val	Val	Ser	Trp	Thr	Ala
			195					200					205		
Met	Ile	Ser	Gly	Tyr	Ala	Gln	Asn	Gly	Gln	Ser	Glu	Glu	Ala	Leu	Glu
			210				215					220			
Leu	Phe	Arg	Glu	Met	Gln	Arg	Glu	Gly	Val	Lys	Pro	Asp	Glu	Val	Thr
225					230						235				240
Leu	Pro	Ser	Val	Leu	Ser	Ala	Cys	Ala	Asn	Leu	Gly	Ala	Leu	Glu	Gln
				245							250			255	
Gly	Lys	Gln	Ile	His	Ala	Tyr	Val	Ile	Lys	Ser	Gly	Phe	Glu	Ser	Asp
			260						265				270		
Val	Phe	Val	Gly	Ser	Ala	Leu	Ile	Asp	Met	Tyr	Ala	Lys	Cys	Gly	Ser

275	280	285
Ile Glu Asp Ala Arg Lys Val Phe Asp Lys Met Pro Glu Arg Asp Val		
290	295	300
Val Thr Trp Asn Ala Leu Ile Ala Gly Tyr Ala Arg Gln Gly Glu Ser		
305	310	315
Asp Leu Val Phe His Leu Leu Glu Arg Met Arg Gln Glu Gly Ile Gln		
325	330	335
Pro Ser Gly Val Thr Phe Thr Ser Val Leu Thr Val Cys Ser His Ala		
340	345	350
Gly Leu Val Asp Glu Gly Gln Lys Tyr Phe Asp Ala Met Ser Glu Asp		
355	360	365
Tyr Gly Ile Thr Pro Arg Ile Glu His Tyr Gly Cys Met Val Asp Leu		
370	375	380
Leu Gly Arg Ala Gly Gln Met Asp Glu Ala Val Ala Met Val Glu Lys		
385	390	395
Met Pro Phe Gln Pro Asn Leu Val Thr Trp Gly Thr Leu Leu Gly Ala		
405	410	415
Cys Arg Lys Trp Asn Asn Val Glu Ile Gly Arg His Ala Phe Glu Cys		
420	425	430
Ala Val Arg Leu Asp Glu Lys Ser Ala Ala Ala Tyr Val Leu Met Ser		
435	440	445
Asn Ile Tyr Ala Asp Ala His Met Trp Glu Glu Arg Asp Lys Ile Gln		
450	455	460
Ala Met Arg Lys Asn Ala Arg Ala Trp Lys Lys Pro Gly Gln Ser Trp		
465	470	475
Trp Thr Asp Thr Gly Gly Ile Val His Thr Phe Val Val Gly Asp Thr		
485	490	495
Lys His Pro Gln Ser Gln Asp Ile Tyr Ala Lys Leu Lys Asp Leu Tyr		
500	505	510
Val Lys Met Lys Glu Glu Gly Tyr Val Pro His Leu Asp Cys Val Leu		
515	520	525
Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp Ala Leu Cys Gly His Ser		
530	535	540
Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile Asn Thr Pro Pro Gly Thr		
545	550	555
Pro Ile Arg Ile Val Lys Asn Leu Arg Val Cys Asp Asp Cys His Lys		
565	570	575
Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly Arg Asn Ile Ile Cys Arg		
580	585	590

Asp Ala Ser Arg Phe His Asn Tyr Lys Asp Gly Lys Cys Ser Cys Gly
 595 600 605
 Asp Tyr Trp
 610
 <210> 37
 <211> 606
 <212> PRT
 <213> 人工序列
 <220>
 <223> PLS阵列(P1L1S1三联体) + DYW:KP-lepto
 <400> 37
 Val Phe Ser Trp Asn Ser Met Ile Arg Gly Tyr Ala Arg Asn Gly Gln
 1 5 10 15
 Pro Glu Glu Ala Leu Ser Leu Tyr Ser Gln Met Arg Arg Ser Gly Ile
 20 25 30
 Lys Pro Asp Asn Tyr Thr Phe Pro Phe Val Leu Lys Ala Cys Ala Ser
 35 40 45
 Leu Ser Ser Leu Lys Glu Gly Lys Gln Ile His Gly His Val Ile Lys
 50 55 60
 Ser Gly Phe Glu Ser Asp Val Tyr Val Gln Ser Ala Leu Ile Asp Met
 65 70 75 80
 Tyr Ala Lys Cys Gly Glu Leu Glu Asp Ala Arg Lys Val Phe Asp Glu
 85 90 95
 Met Pro Glu Arg Asn Val Val Ser Trp Asn Ala Met Ile Ser Gly Tyr
 100 105 110
 Ala Gln Asn Gly Gln Ser Glu Glu Ala Leu Glu Leu Phe Arg Glu Met
 115 120 125
 Gln Gln Glu Gly Ile Lys Pro Ser Glu Phe Thr Phe Cys Ser Val Leu
 130 135 140
 Ser Ala Cys Ala Ser Leu Gly Ser Leu Glu Met Gly Lys Gln Ile His
 145 150 155 160
 Gly Tyr Val Ile Lys Ser Gly Phe Glu Ser Ile Val Phe Val Gly Asn
 165 170 175
 Ala Leu Ile Asp Met Tyr Ala Lys Cys Gly Ser Ile Glu Asp Ala Arg
 180 185 190
 Lys Val Phe Asp Glu Met Pro Glu Arg Thr Val Val Ser Trp Thr Ala
 195 200 205
 Met Ile Ser Gly Tyr Ala Gln Asn Gly Gln Ser Glu Glu Ala Leu Glu
 210 215 220

Leu Phe Arg Glu Met Gln Arg Glu Gly Val Lys Pro Asp Glu Val Thr
 225 230 235 240
 Leu Pro Ser Val Leu Ser Ala Cys Ala Asn Leu Gly Ala Leu Glu Gln
 245 250 255
 Gly Lys Gln Ile His Ala Tyr Val Ile Lys Ser Gly Phe Glu Ser Asp
 260 265 270
 Val Phe Val Gly Ser Ala Leu Ile Asp Met Tyr Ala Lys Cys Gly Ser
 275 280 285
 Ile Glu Asp Ala Arg Lys Val Phe Asp Lys Met Pro Glu Arg Asp Val
 290 295 300
 Val Ala Trp Asn Ser Met Ile Lys Gly Tyr Gly Met Asn His Glu Gly
 305 310 315 320
 Lys Met Ala Val Gln Cys Phe Glu Asp Met Gln Lys Gln Gly Val Lys
 325 330 335
 Pro Ser Ala Val Thr Phe Thr Cys Leu Leu Thr Ala Cys Ser His Ala
 340 345 350
 Ser Leu Val Ser Glu Gly Gln Glu Tyr Phe Lys Met Met Arg Glu Glu
 355 360 365
 Tyr Gly Ile Ala Pro Arg Val Glu His Tyr Gly Cys Met Val Asp Leu
 370 375 380
 Leu Ala Arg Ser Gly His Leu Tyr Glu Ala Glu Lys Phe Leu Glu Met
 385 390 395 400
 Leu Cys Pro Pro Asn Glu Gly Thr Trp Gly Ala Leu Leu Ser Ala Cys
 405 410 415
 Lys Thr Tyr Gly Glu Val Glu Leu Gly Leu Arg Cys Phe Gln Gln Leu
 420 425 430
 Val Gln Leu Asn Pro Glu Ser Ala Ala Trp Tyr Val Leu Met Ala Asp
 435 440 445
 Ile Tyr Ala Gly Ala Gly Arg Trp Asp Asp Ala Tyr Arg Ile Glu Glu
 450 455 460
 Leu Arg Lys His Ala Gly Ala Lys Lys Lys Pro Ala Ser Ala Leu Ile
 465 470 475 480
 Glu Val Asn Lys Lys Val His Glu Phe Val Val Gly Asn Asn Gln Ser
 485 490 495
 Glu Glu Ile Ser Ala Met Leu Lys Ser Leu Asn Ser Arg Met Lys Glu
 500 505 510
 Glu Gly His Val Pro Asn Leu Asp Leu Val Leu Lys Pro Val Ser Asp
 515 520 525
 Glu Glu Lys Glu Ala Ala Leu Cys Glu His Ala Glu Lys Leu Ala Ile

530	535	540																	
Ala Phe Gly Leu Leu Asn Thr Pro Gln Gly Gln Thr Leu Arg Val Thr																			
545	550	555	560																
Lys Asn Leu Arg Met Cys Asn Asp Cys His Asn Ala Ser Lys Ile Ile																			
	565	570	575																
Ser Lys Ile Glu Arg Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile																			
	580	585	590																
His His Phe Lys Asp Gly Leu Cys Ser Cys Gly Asp Met Phe																			
	595	600	605																
<210>	38																		
<211>	657																		
<212>	PRT																		
<213>	人工序列																		
<220>																			
<223>	P阵列(10 PPR基序) + DYW:PG-lyc																		
<400>	38																		
Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys																			
1	5	10	15																
Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val																			
	20	25	30																
Lys Pro Asn Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys																			
	35	40	45																
Ala Gly Arg Leu Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu																			
	50	55	60																
Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly																			
65	70	75	80																
Leu Cys Lys Ala Gly Arg Leu Asp Glu Ala Glu Glu Leu Leu Glu Glu																			
	85	90	95																
Met Glu Glu Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Thr Thr Leu																			
	100	105	110																
Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu																			
	115	120	125																
Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr																			
	130	135	140																
Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala																			
145	150	155	160																
Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val																			
	165	170	175																
Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val																			

	180		185		190										
Asp	Glu	Ala	Leu	Glu	Leu	Phe	Lys	Glu	Met	Arg	Ser	Lys	Gly	Val	Lys
	195		200		205										
Pro	Asn	Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ser
	210		215		220										
Gly	Lys	Ile	Glu	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Glu	Met	Glu	Glu	Lys
225			230		235										240
Gly	Ile	Thr	Pro	Ser	Val	Val	Thr	Tyr	Thr	Thr	Leu	Ile	Asp	Gly	Leu
			245		250										255
Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu	Ala	Leu	Glu	Leu	Phe	Asp	Glu	Met
			260		265										270
Lys	Glu	Arg	Gly	Ile	Lys	Pro	Asp	Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile
	275		280		285										
Asp	Gly	Leu	Cys	Lys	Ala	Gly	Arg	Leu	Asp	Glu	Ala	Glu	Glu	Leu	Leu
	290		295		300										
Glu	Glu	Met	Glu	Glu	Lys	Gly	Ile	Lys	Pro	Asp	Val	Val	Thr	Tyr	Thr
305			310		315										320
Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu	Ala	Leu
			325		330										335
Glu	Leu	Phe	Asp	Glu	Met	Lys	Glu	Arg	Gly	Ile	Lys	Pro	Asp	Val	Val
			340		345										350
Ser	Trp	Asn	Ala	Met	Ile	Ala	Ala	Tyr	Ala	Gln	His	Gly	His	Gly	Lys
	355		360		365										
Glu	Ala	Leu	Gln	Leu	Phe	Gln	Gln	Met	Gln	Gln	Glu	Gly	Val	Lys	Pro
	370		375		380										
Ser	Glu	Val	Thr	Phe	Thr	Ser	Ile	Leu	Ser	Ala	Cys	Ser	His	Ala	Gly
385			390		395										400
Leu	Val	Asp	Glu	Gly	His	His	Tyr	Phe	Glu	Ser	Met	Ser	Pro	Asp	Tyr
			405		410										415
Gly	Ile	Thr	Pro	Arg	Val	Glu	His	Tyr	Gly	Cys	Met	Val	Asp	Leu	Leu
			420		425										430
Gly	Arg	Ala	Gly	Arg	Leu	Asp	Glu	Ala	Glu	Asp	Leu	Ile	Lys	Ser	Met
	435		440		445										
Pro	Phe	Gln	Pro	Asn	Val	Val	Val	Trp	Gly	Thr	Leu	Leu	Gly	Ala	Cys
	450		455		460										
Arg	Val	His	Gly	Asp	Val	Glu	Arg	Gly	Glu	Arg	Ala	Ala	Glu	Arg	Ile
465			470		475										480
Leu	Glu	Leu	Asp	Pro	Glu	Ser	Ala	Ala	Pro	Tyr	Val	Leu	Leu	Ser	Asn
			485		490										495

Ile Tyr Ala Ala Ala Gly Arg Trp Asp Glu Ala Ala Lys Val Arg Lys
 500 505 510
 Leu Met Lys Glu Arg Gly Val Lys Lys Glu Pro Gly Cys Ser Trp Ile
 515 520 525
 Glu Val Asn Asn Lys Val His Glu Phe Val Ala Gly Asp Lys Ser His
 530 535 540
 Pro Gln Thr Lys Glu Ile Tyr Ala Glu Leu Glu Arg Leu Ser Lys Gln
 545 550 555 560
 Met Lys Glu Ala Gly Tyr Val Pro Asp Thr Lys Phe Val Leu His Asp
 565 570 575
 Val Glu Glu Glu Glu Lys Glu Gln Leu Leu Cys Tyr His Ser Glu Lys
 580 585 590
 Leu Ala Ile Ala Phe Gly Leu Ile Ser Thr Pro Pro Gly Thr Pro Leu
 595 600 605
 Arg Ile Ile Lys Asn Leu Arg Val Cys Gly Asp Cys His Thr Ala Thr
 610 615 620
 Lys Phe Ile Ser Lys Ile Val Gly Arg Glu Ile Val Val Arg Asp Ala
 625 630 635 640
 Asn Arg Phe His His Phe Lys Asp Gly Val Cys Ser Cys Gly Asp Tyr
 645 650 655
 Trp
 <210> 39
 <211> 658
 <212> PRT
 <213> 人工序列
 <220>
 <223> P阵列(10 PPR基序) + DYW:WW-WW5
 <400> 39
 Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys
 1 5 10 15
 Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val
 20 25 30
 Lys Pro Asn Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys
 35 40 45
 Ala Gly Arg Leu Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu
 50 55 60
 Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly
 65 70 75 80
 Leu Cys Lys Ala Gly Arg Leu Asp Glu Ala Glu Glu Leu Leu Glu Glu

	85	90	95
Met Glu Glu Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Thr Thr Leu			
	100	105	110
Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu			
	115	120	125
Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr			
	130	135	140
Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala			
145	150	155	160
Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val			
	165	170	175
Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val			
	180	185	190
Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys			
	195	200	205
Pro Asn Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser			
	210	215	220
Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys			
225	230	235	240
Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu			
	245	250	255
Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu Met			
	260	265	270
Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu Ile			
	275	280	285
Asp Gly Leu Cys Lys Ala Gly Arg Leu Asp Glu Ala Glu Glu Leu Leu			
	290	295	300
Glu Glu Met Glu Glu Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Thr			
305	310	315	320
Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu			
	325	330	335
Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val			
	340	345	350
Thr Trp Asn Ala Leu Ile Ala Gly Tyr Ala Arg Gln Gly Glu Ser Asp			
	355	360	365
Leu Val Phe His Leu Leu Glu Arg Met Arg Gln Glu Gly Ile Gln Pro			
	370	375	380
Ser Gly Val Thr Phe Thr Ser Val Leu Thr Val Cys Ser His Ala Gly			
385	390	395	400

Leu Val Asp Glu Gly Gln Lys Tyr Phe Asp Ala Met Ser Glu Asp Tyr
 405 410 415
 Gly Ile Thr Pro Arg Ile Glu His Tyr Gly Cys Met Val Asp Leu Leu
 420 425 430
 Gly Arg Ala Gly Gln Met Asp Glu Ala Val Ala Met Val Glu Lys Met
 435 440 445
 Pro Phe Gln Pro Asn Leu Val Thr Trp Gly Thr Leu Leu Gly Ala Cys
 450 455 460
 Arg Lys Trp Asn Asn Val Glu Ile Gly Arg His Ala Phe Glu Cys Ala
 465 470 475 480
 Val Arg Leu Asp Glu Lys Ser Ala Ala Ala Tyr Val Leu Met Ser Asn
 485 490 495
 Ile Tyr Ala Asp Ala His Met Trp Glu Glu Arg Asp Lys Ile Gln Ala
 500 505 510
 Met Arg Lys Asn Ala Arg Ala Trp Lys Lys Pro Gly Gln Ser Trp Trp
 515 520 525
 Thr Asp Thr Gly Gly Ile Val His Thr Phe Val Val Gly Asp Thr Lys
 530 535 540
 His Pro Gln Ser Gln Asp Ile Tyr Ala Lys Leu Lys Asp Leu Tyr Val
 545 550 555 560
 Lys Met Lys Glu Glu Gly Tyr Val Pro His Leu Asp Cys Val Leu Trp
 565 570 575
 Asp Ile Ser Asp Asp Glu Lys Glu Asp Ala Leu Cys Gly His Ser Glu
 580 585 590
 Lys Leu Ala Ile Ala Cys Ala Leu Ile Asn Thr Pro Pro Gly Thr Pro
 595 600 605
 Ile Arg Ile Val Lys Asn Leu Arg Val Cys Asp Asp Cys His Lys Ala
 610 615 620
 Ile Ala Leu Ile Ser Lys Ile Glu Gly Arg Asn Ile Ile Cys Arg Asp
 625 630 635 640
 Ala Ser Arg Phe His Asn Tyr Lys Asp Gly Lys Cys Ser Cys Gly Asp
 645 650 655

Tyr Trp

<210> 40

<211> 653

<212> PRT

<213> 人工序列

<220>

<223> P阵列(10 PPR基序) + DYW:KP-lepto

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

Glu Glu Met Glu Glu Lys Gly Ile Lys Pro Asp Val Val Thr Tyr Thr
 305 310 315 320
 Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu
 325 330 335
 Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val
 340 345 350
 Ala Trp Asn Ser Met Ile Lys Gly Tyr Gly Met Asn His Glu Gly Lys
 355 360 365
 Met Ala Val Gln Cys Phe Glu Asp Met Gln Lys Gln Gly Val Lys Pro
 370 375 380
 Ser Ala Val Thr Phe Thr Cys Leu Leu Thr Ala Cys Ser His Ala Ser
 385 390 395 400
 Leu Val Ser Glu Gly Gln Glu Tyr Phe Lys Met Met Arg Glu Glu Tyr
 405 410 415
 Gly Ile Ala Pro Arg Val Glu His Tyr Gly Cys Met Val Asp Leu Leu
 420 425 430
 Ala Arg Ser Gly His Leu Tyr Glu Ala Glu Lys Phe Leu Glu Met Leu
 435 440 445
 Cys Pro Pro Asn Glu Gly Thr Trp Gly Ala Leu Leu Ser Ala Cys Lys
 450 455 460
 Thr Tyr Gly Glu Val Glu Leu Gly Leu Arg Cys Phe Gln Gln Leu Val
 465 470 475 480
 Gln Leu Asn Pro Glu Ser Ala Ala Trp Tyr Val Leu Met Ala Asp Ile
 485 490 495
 Tyr Ala Gly Ala Gly Arg Trp Asp Asp Ala Tyr Arg Ile Glu Glu Leu
 500 505 510
 Arg Lys His Ala Gly Ala Lys Lys Lys Pro Ala Ser Ala Leu Ile Glu
 515 520 525
 Val Asn Lys Lys Val His Glu Phe Val Val Gly Asn Asn Gln Ser Glu
 530 535 540
 Glu Ile Ser Ala Met Leu Lys Ser Leu Asn Ser Arg Met Lys Glu Glu
 545 550 555 560
 Gly His Val Pro Asn Leu Asp Leu Val Leu Lys Pro Val Ser Asp Glu
 565 570 575
 Glu Lys Glu Ala Ala Leu Cys Glu His Ala Glu Lys Leu Ala Ile Ala
 580 585 590
 Phe Gly Leu Leu Asn Thr Pro Gln Gly Gln Thr Leu Arg Val Thr Lys
 595 600 605
 Asn Leu Arg Met Cys Asn Asp Cys His Asn Ala Ser Lys Ile Ile Ser

610	615	620
Lys Ile Glu Arg Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile His		
625	630	635
His Phe Lys Asp Gly Leu Cys Ser Cys Gly Asp Met Phe		
	645	650

<210> 41

<211> 136

<212> PRT

<213> 人工序列

<220>

<223> PG2

<400> 41

Gln Pro Gly Cys Ser Ser Ile Glu Val Asp Asn Lys Val His Glu Phe		
1	5	10
Ile Val Gly Asp Arg Ser His Pro Gln Thr Glu Glu Ile Tyr Ala Glu		
	20	25
Leu Glu Arg Leu Asn Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp		
	35	40
Thr Arg Leu Val Leu His Asp Val Asp Glu Glu Glu Lys Glu His Leu		
	50	55
Leu Cys His His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser		
65	70	75
Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys		
	85	90
Gly Asp Cys His Thr Ala Thr Lys Tyr Ile Ser Lys Ile Val Gly Arg		
	100	105
Glu Ile Ile Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly		
	115	120
Val Cys Ser Cys Gly Asp Tyr Trp		
	130	135

<210> 42

<211> 136

<212> PRT

<213> 人工序列

<220>

<223> PG3

<400> 42

Lys Pro Gly Cys Ser Trp Ile Glu Val Asn Asn Lys Val His Glu Phe		
1	5	10
		15

Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp
 35 40 45
 Thr Lys Phe Val Leu His Asp Val Glu Glu Glu Glu Lys Glu Gln Leu
 50 55 60
 Leu Cys Tyr His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser
 65 70 75 80
 Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys
 85 90 95
 Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Gly Arg
 100 105 110
 Glu Ile Val Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly
 115 120 125
 Val Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 43
 <211> 136
 <212> PRT
 <213> 人工序列
 <220>
 <223> PG4
 <400> 43
 Glu Pro Ala Cys Ser Trp Ile Glu Val Asn Asn Lys Val His Glu Phe
 1 5 10 15
 Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp
 35 40 45
 Thr Lys Phe Val Leu His Asp Val Glu Glu Glu Glu Lys Glu Gln Leu
 50 55 60
 Leu Cys Tyr His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser
 65 70 75 80
 Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys
 85 90 95
 Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Gly Arg
 100 105 110
 Glu Ile Val Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly
 115 120 125

Val Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 44
 <211> 136
 <212> PRT
 <213> 人工序列
 <220>
 <223> PG5
 <400> 44
 Glu Pro Gly Cys Ala Trp Ile Glu Val Asn Asn Lys Val His Glu Phe
 1 5 10 15
 Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp
 35 40 45
 Thr Lys Phe Val Leu His Asp Val Glu Glu Glu Glu Lys Glu Gln Leu
 50 55 60
 Leu Cys Tyr His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser
 65 70 75 80
 Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys
 85 90 95
 Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Gly Arg
 100 105 110
 Glu Ile Val Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly
 115 120 125
 Val Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 45
 <211> 136
 <212> PRT
 <213> 人工序列
 <220>
 <223> PG6
 <400> 45
 Glu Pro Gly Cys Ser Thr Ile Glu Val Asn Asn Lys Val His Glu Phe
 1 5 10 15
 Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp

35	40	45
Thr Lys Phe Val Leu His Asp	Val Glu Glu Glu Glu Lys	Glu Gln Leu
50	55	60
Leu Cys Tyr His Ser Glu Lys	Leu Ala Ile Ala Phe Gly	Leu Ile Ser
65	70	80
Thr Pro Pro Gly Thr Pro Leu Arg	Ile Ile Lys Asn Leu Arg	Val Cys
85	90	95
Gly Asp Cys His Thr Ala Thr Lys	Phe Ile Ser Lys Ile Val Gly	Arg
100	105	110
Glu Ile Val Val Arg Asp Ala Asn Arg	Phe His His Phe Lys Asp	Gly
115	120	125
Val Cys Ser Cys Gly Asp Tyr Trp		
130	135	
<210>	46	
<211>	136	
<212>	PRT	
<213>	人工序列	
<220>		
<223>	PG7	
<400>	46	
Glu Pro Gly Cys Ser Trp Ile Glu Val	Asn Asn Lys Val His Glu Phe	
1	5	10
Val Ala Gly Asp Lys Ser His Pro Gln	Thr Lys Glu Ile Tyr Ala Glu	
20	25	30
Leu Glu Arg Leu Asn Lys Gln Met Lys	Glu Ala Gly Tyr Val Pro Asp	
35	40	45
Thr Lys Phe Val Leu His Asp	Val Glu Glu Glu Glu Lys	Glu Gln Leu
50	55	60
Leu Cys Tyr His Ser Glu Lys	Leu Ala Ile Ala Phe Gly	Leu Ile Ser
65	70	80
Thr Pro Pro Gly Thr Pro Leu Arg	Ile Ile Lys Asn Leu Arg	Val Cys
85	90	95
Gly Asp Cys His Thr Ala Thr Lys	Phe Ile Ser Lys Ile Val Gly	Arg
100	105	110
Glu Ile Val Val Arg Asp Ala Asn Arg	Phe His His Phe Lys Asp	Gly
115	120	125
Val Cys Ser Cys Gly Asp Tyr Trp		
130	135	
<210>	47	

<211> 136
 <212> PRT
 <213> 人工序列
 <220>
 <223> PG8
 <400> 47
 Glu Pro Gly Cys Ser Trp Ile Glu Val Asn Asn Lys Val His Glu Phe
 1 5 10 15
 Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly His Val Pro Asp
 35 40 45
 Thr Lys Phe Val Leu His Asp Val Glu Glu Glu Glu Lys Glu Gln Leu
 50 55 60
 Leu Cys Tyr His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser
 65 70 75 80
 Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys
 85 90 95
 Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Gly Arg
 100 105 110
 Glu Ile Val Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly
 115 120 125
 Val Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 48
 <211> 136
 <212> PRT
 <213> 人工序列
 <220>
 <223> PG9
 <400> 48
 Glu Pro Gly Cys Ser Trp Ile Glu Val Asn Asn Lys Val His Glu Phe
 1 5 10 15
 Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu
 20 25 30
 Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp
 35 40 45
 Thr Lys Phe Val Leu His Pro Val Glu Glu Glu Glu Lys Glu Gln Leu
 50 55 60

<220>

<223> PG11

<400> 50

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

<210> 51

<211> 136

<212> PRT

<213> 人工序列

<220>

<223> PG12

<400> 51

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

	85		90		95
Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Lys Arg					
	100		105		110
Glu Ile Val Val Arg Asp Ala Asn Arg Phe His His Phe Lys Asp Gly					
	115		120		125
Val Cys Ser Cys Gly Asp Tyr Trp					
	130		135		
<210>	52				
<211>	136				
<212>	PRT				
<213>	人工序列				
<220>					
<223>	PG13				
<400>	52				
Glu Pro Gly Cys Ser Trp Ile Glu Val Asn Asn Lys Val His Glu Phe					
1	5		10		15
Val Ala Gly Asp Lys Ser His Pro Gln Thr Lys Glu Ile Tyr Ala Glu					
	20		25		30
Leu Glu Arg Leu Ser Lys Gln Met Lys Glu Ala Gly Tyr Val Pro Asp					
	35		40		45
Thr Lys Phe Val Leu His Asp Val Glu Glu Glu Glu Lys Glu Gln Leu					
	50		55		60
Leu Cys Tyr His Ser Glu Lys Leu Ala Ile Ala Phe Gly Leu Ile Ser					
65	70		75		80
Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile Lys Asn Leu Arg Val Cys					
	85		90		95
Gly Asp Cys His Thr Ala Thr Lys Phe Ile Ser Lys Ile Val Gly Arg					
	100		105		110
Glu Ile Val Val Arg Asp Asp Cys Arg Phe His His Phe Lys Asp Gly					
	115		120		125
Val Cys Ser Cys Gly Asp Tyr Trp					
	130		135		
<210>	53				
<211>	136				
<212>	PRT				
<213>	人工序列				
<220>					
<223>	PG14				
<400>	53				

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

<210> 54

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW2

<400> 54

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

Arg Thr Ile Ile Cys Arg Asp Ala Ser Arg Phe His Val Tyr Lys Asp
 115 120 125

Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 55

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW3

<400> 55

Glu Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15

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

Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro
 35 40 45

His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60

Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile
 65 70 75 80

Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val
 85 90 95

Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly
 100 105 110

Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp
 115 120 125

Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 56

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW4

<400> 56

Lys Pro Ala Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15

Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala

	20		25		30										
Lys	Leu	Lys	Asp	Leu	Tyr	Val	Lys	Met	Lys	Glu	Glu	Gly	Tyr	Val	Pro
	35						40					45			
His	Leu	Asp	Cys	Val	Leu	Trp	Asp	Ile	Ser	Asp	Asp	Glu	Lys	Glu	Asp
	50						55					60			
Ala	Leu	Cys	Gly	His	Ser	Glu	Lys	Leu	Ala	Ile	Ala	Cys	Ala	Leu	Ile
65						70				75				80	
Asn	Thr	Pro	Pro	Gly	Thr	Pro	Ile	Arg	Ile	Val	Lys	Asn	Leu	Arg	Val
				85				90					95		
Cys	Asp	Asp	Cys	His	Lys	Ala	Ile	Ala	Leu	Ile	Ser	Lys	Ile	Glu	Gly
			100					105					110		
Arg	Asn	Ile	Ile	Cys	Arg	Asp	Ala	Ser	Arg	Phe	His	Asn	Tyr	Lys	Asp
	115						120						125		
Gly	Lys	Cys	Ser	Cys	Gly	Asp	Tyr	Trp							
	130						135								
<210>	57														
<211>	137														
<212>	PRT														
<213>	人工序列														
<220>															
<223>	WW5														
<400>	57														
Lys	Pro	Gly	Gln	Ala	Trp	Trp	Thr	Asp	Thr	Gly	Gly	Ile	Val	His	Thr
1				5					10					15	
Phe	Val	Val	Gly	Asp	Thr	Lys	His	Pro	Gln	Ser	Gln	Asp	Ile	Tyr	Ala
			20					25				30			
Lys	Leu	Lys	Asp	Leu	Tyr	Val	Lys	Met	Lys	Glu	Glu	Gly	Tyr	Val	Pro
	35						40					45			
His	Leu	Asp	Cys	Val	Leu	Trp	Asp	Ile	Ser	Asp	Asp	Glu	Lys	Glu	Asp
	50						55					60			
Ala	Leu	Cys	Gly	His	Ser	Glu	Lys	Leu	Ala	Ile	Ala	Cys	Ala	Leu	Ile
65						70				75				80	
Asn	Thr	Pro	Pro	Gly	Thr	Pro	Ile	Arg	Ile	Val	Lys	Asn	Leu	Arg	Val
				85				90					95		
Cys	Asp	Asp	Cys	His	Lys	Ala	Ile	Ala	Leu	Ile	Ser	Lys	Ile	Glu	Gly
			100					105					110		
Arg	Asn	Ile	Ile	Cys	Arg	Asp	Ala	Ser	Arg	Phe	His	Asn	Tyr	Lys	Asp
	115						120						125		
Gly	Lys	Cys	Ser	Cys	Gly	Asp	Tyr	Trp							

130	135
<210> 58	
<211> 137	
<212> PRT	
<213> 人工序列	
<220>	
<223> WW6	
<400> 58	
Lys Pro Gly Gln Ser Thr Trp Thr Asp Thr Gly Gly Ile Val His Thr	
1	5 10 15
Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala	
	20 25 30
Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro	
	35 40 45
His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp	
	50 55 60
Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile	
65	70 75 80
Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val	
	85 90 95
Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly	
	100 105 110
Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp	
	115 120 125
Gly Lys Cys Ser Cys Gly Asp Tyr Trp	
130	135
<210> 59	
<211> 136	
<212> PRT	
<213> 人工序列	
<220>	
<223> WW7	
<400> 59	
Lys Pro Gly Gln Ser Trp Thr Asp Thr Gly Gly Ile Val His Thr Phe	
1	5 10 15
Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala Lys	
	20 25 30
Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro His	
	35 40 45

Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp Ala
 50 55 60
 Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile Asn
 65 70 75 80
 Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val Cys
 85 90 95
 Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly Arg
 100 105 110
 Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp Gly
 115 120 125
 Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 60
 <211> 137
 <212> PRT
 <213> 人工序列
 <220>
 <223> WW8
 <400> 60
 Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala
 20 25 30
 Lys Leu Lys Asp Leu Asn Val Lys Met Lys Glu Glu Gly Tyr Val Pro
 35 40 45
 His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60
 Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile
 65 70 75 80
 Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val
 85 90 95
 Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly
 100 105 110
 Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp
 115 120 125
 Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 61
 <211> 137

<212> PRT
 <213> 人工序列
 <220>
 <223> WW9
 <400> 61
 Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala
 20 25 30
 Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly His Val Pro
 35 40 45
 His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60
 Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile
 65 70 75 80
 Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val
 85 90 95
 Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly
 100 105 110
 Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp
 115 120 125
 Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 62
 <211> 137
 <212> PRT
 <213> 人工序列
 <220>
 <223> WW10
 <400> 62
 Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala
 20 25 30
 Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro
 35 40 45
 His Leu Asp Cys Val Leu Trp Pro Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60
 Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile

65		70		75		80
Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val						
		85		90		95
Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly						
		100		105		110
Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp						
		115		120		125
Gly Lys Cys Ser Cys Gly Asp Tyr Trp						
		130		135		
<210>	63					
<211>	137					
<212>	PRT					
<213>	人工序列					
<220>						
<223>	WW11					
<400>	63					
Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr						
1		5		10		15
Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala						
		20		25		30
Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro						
		35		40		45
His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp						
		50		55		60
Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile						
65		70		75		80
Asn Thr Pro Pro Gly Thr Pro Ile Arg Val Thr Lys Asn Leu Arg Val						
		85		90		95
Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly						
		100		105		110
Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp						
		115		120		125
Gly Lys Cys Ser Cys Gly Asp Tyr Trp						
		130		135		
<210>	64					
<211>	137					
<212>	PRT					
<213>	人工序列					
<220>						

<223> WW12

<400> 64

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

<210> 65

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW13

<400> 65

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

Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Lys
 100 105 110

Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg Phe His Asn Tyr Lys Asp
 115 120 125

Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 66

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW14

<400> 66

Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15

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

Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro
 35 40 45

His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp
 50 55 60

Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile
 65 70 75 80

Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val
 85 90 95

Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly
 100 105 110

Arg Asn Ile Ile Cys Arg Asp Asp Cys Arg Phe His Asn Tyr Lys Asp
 115 120 125

Gly Lys Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 67

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> WW15

<400> 67

Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr

1	5	10	15
Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Gln Asp Ile Tyr Ala			
	20	25	30
Lys Leu Lys Asp Leu Tyr Val Lys Met Lys Glu Glu Gly Tyr Val Pro			
	35	40	45
His Leu Asp Cys Val Leu Trp Asp Ile Ser Asp Asp Glu Lys Glu Asp			
	50	55	60
Ala Leu Cys Gly His Ser Glu Lys Leu Ala Ile Ala Cys Ala Leu Ile			
65	70	75	80
Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile Val Lys Asn Leu Arg Val			
	85	90	95
Cys Asp Asp Cys His Lys Ala Ile Ala Leu Ile Ser Lys Ile Glu Gly			
	100	105	110
Arg Asn Ile Ile Cys Arg Asp Ala Ser Cys Val His Asn Tyr Lys Asp			
	115	120	125
Gly Lys Cys Ser Cys Gly Asp Tyr Trp			
	130	135	
<210>	68		
<211>	133		
<212>	PRT		
<213>	人工序列		
<220>			
<223>	KP2		
<400>	68		
Lys Pro Ala Val Ala Gly Ile Glu Val Asp Arg His Val His Thr Phe			
1	5	10	15
Val Val Gly Glu Glu Arg Gly Asp Ile Ser Cys Lys Val Arg Ser Met			
	20	25	30
Asn Thr Arg Leu Lys Glu Glu Gly Gly His Val Pro His Thr Glu Leu			
	35	40	45
Val Leu Lys Pro Leu Ser Glu Lys Glu Lys Glu Asp Glu Leu Cys Gly			
	50	55	60
His Ala Glu Lys Leu Ala Leu Ala Tyr Gly Leu Leu Asn Thr Pro Asp			
65	70	75	80
Gly Thr Pro Leu Leu Val Thr Lys Asn Leu Arg Met Cys Arg Asp Cys			
	85	90	95
His Ser Ser Thr Glu Ile Met Ser Arg Ile Glu Lys Arg Glu Ile Val			
	100	105	110
Val Arg Asp Ala His Arg Val His Arg Phe Val Asp Gly Ser Cys Ser			

<210> 72
 <211> 137
 <212> PRT
 <213> 人工序列
 <220>
 <223> KP6
 <400> 72
 Lys Pro Ala Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Glu Glu Ile Ser Ala
 20 25 30
 Met Leu Lys Ser Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro
 35 40 45
 Asn Leu Asp Leu Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala
 50 55 60
 Ala Leu Cys Glu His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu
 65 70 75 80
 Asn Thr Pro Gln Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met
 85 90 95
 Cys Asn Asp Cys His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg
 100 105 110
 Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp
 115 120 125
 Gly Leu Cys Ser Cys Gly Asp Tyr Trp
 130 135

<210> 73
 <211> 133
 <212> PRT
 <213> 人工序列
 <220>
 <223> KP7
 <400> 73
 Lys Pro Ala Ser Ser Leu Ile Glu Val Asn Lys Lys Val His Glu Phe
 1 5 10 15
 Val Val Gly Asn Asn Gln Ser Glu Glu Ile Ser Ala Met Leu Lys Ser
 20 25 30
 Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro Asn Leu Asp Leu
 35 40 45
 Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala Ala Leu Cys Glu

50	55	60													
His Ala Glu Lys Leu	Ala Ile Ala Phe Gly Leu Leu	Asn Thr Pro Gln													
65	70	75	80												
Gly Gln Thr Leu Arg	Val Thr Lys Asn Leu Arg Met Cys	Asn Asp Cys													
	85	90	95												
His Asn Ala Ser Lys	Ile Ile Ser Lys Ile Glu Arg Arg	Glu Ile Ile													
	100	105	110												
Leu Arg Asp Asp Cys	Cys Ile His His Phe Lys Asp Gly Leu Cys	Ser													
	115	120	125												
Cys Gly Asp Met Phe															
130															
<210>	74														
<211>	137														
<212>	PRT														
<213>	人工序列														
<220>															
<223>	KP8														
<400>	74														
Lys Pro Gly Gln Ala	Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr														
1	5	10	15												
Phe Val Val Gly Asp	Thr Lys His Pro Gln Ser Glu Glu Ile Ser Ala														
	20	25	30												
Met Leu Lys Ser Leu	Asn Ser Arg Met Lys Glu Glu Gly His Val Pro														
	35	40	45												
Asn Leu Asp Leu Val	Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala														
	50	55	60												
Ala Leu Cys Glu His	Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu														
65	70	75	80												
Asn Thr Pro Gln Gly	Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met														
	85	90	95												
Cys Asn Asp Cys His	Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg														
	100	105	110												
Arg Glu Ile Ile Leu	Arg Asp Asp Cys Cys Ile His His Phe Lys Asp														
	115	120	125												
Gly Leu Cys Ser Cys	Gly Asp Tyr Trp														
	130	135													
<210>	75														
<211>	133														
<212>	PRT														

<213> 人工序列

<220>

<223> KP9

<400> 75

Lys Pro Ala Ser Ala Trp Ile Glu Val Asn Lys Lys Val His Glu Phe
 1 5 10 15
 Val Val Gly Asn Asn Gln Ser Glu Glu Ile Ser Ala Met Leu Lys Ser
 20 25 30
 Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro Asn Leu Asp Leu
 35 40 45
 Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala Ala Leu Cys Glu
 50 55 60
 His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu Asn Thr Pro Gln
 65 70 75 80
 Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met Cys Asn Asp Cys
 85 90 95
 His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg Arg Glu Ile Ile
 100 105 110
 Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp Gly Leu Cys Ser
 115 120 125
 Cys Gly Asp Met Phe
 130

<210> 76

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> KP10

<400> 76

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

<400> 78

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

<210> 79

<211> 133

<212> PRT

<213> 人工序列

<220>

<223> KP13

<400> 79

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

	100		105		110
Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp Gly Leu Cys Ser					
	115		120		125
Cys Gly Asp Met Phe					
	130				
<210>	80				
<211>	137				
<212>	PRT				
<213>	人工序列				
<220>					
<223>	KP14				
<400>	80				
Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr					
1	5		10		15
Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Glu Glu Ile Ser Ala					
	20		25		30
Met Leu Lys Ser Leu Asn Ser Arg Met Lys Glu Glu Gly Tyr Val Pro					
	35		40		45
Asn Leu Asp Leu Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala					
	50		55		60
Ala Leu Cys Glu His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu					
65	70		75		80
Asn Thr Pro Gln Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met					
	85		90		95
Cys Asn Asp Cys His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg					
	100		105		110
Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp					
	115		120		125
Gly Leu Cys Ser Cys Gly Asp Tyr Trp					
	130		135		
<210>	81				
<211>	137				
<212>	PRT				
<213>	人工序列				
<220>					
<223>	KP15				
<400>	81				
Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr					
1	5		10		15

Gly Leu Cys Ser Cys Gly Asp Tyr Trp
 130 135
 <210> 83
 <211> 133
 <212> PRT
 <213> 人工序列
 <220>
 <223> KP17
 <400> 83
 Lys Pro Ala Ser Ala Leu Ile Glu Val Asn Lys Lys Val His Glu Phe
 1 5 10 15
 Val Val Gly Asn Asn Gln Ser Glu Glu Ile Ser Ala Met Leu Lys Ser
 20 25 30
 Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro Asn Leu Asp Leu
 35 40 45
 Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala Ala Leu Cys Glu
 50 55 60
 His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu Asn Thr Pro Gln
 65 70 75 80
 Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met Cys Asn Asp Cys
 85 90 95
 His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Gly Arg Glu Ile Ile
 100 105 110
 Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp Gly Leu Cys Ser
 115 120 125
 Cys Gly Asp Met Phe
 130
 <210> 84
 <211> 137
 <212> PRT
 <213> 人工序列
 <220>
 <223> KP18
 <400> 84
 Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr Gly Gly Ile Val His Thr
 1 5 10 15
 Phe Val Val Gly Asp Thr Lys His Pro Gln Ser Glu Glu Ile Ser Ala
 20 25 30
 Met Leu Lys Ser Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro

35	40	45
Asn Leu Asp Leu Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala		
50	55	60
Ala Leu Cys Glu His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu		
65	70	75
Asn Thr Pro Gln Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met		
	85	90
Cys Asn Asp Cys His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Gly		
	100	110
Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp		
	115	120
Gly Leu Cys Ser Cys Gly Asp Tyr Trp		
	130	135
<210>	85	
<211>	133	
<212>	PRT	
<213>	人工序列	
<220>		
<223>	KP19	
<400>	85	
Lys Pro Ala Ser Ala Leu Ile Glu Val Asn Lys Lys Val His Glu Phe		
1	5	10
Val Val Gly Asn Asn Gln Ser Glu Glu Ile Ser Ala Met Leu Lys Ser		
	20	25
Leu Asn Ser Arg Met Lys Glu Glu Gly His Val Pro Asn Leu Asp Leu		
	35	40
Val Leu Lys Pro Val Ser Asp Glu Glu Lys Glu Ala Ala Leu Cys Glu		
	50	55
His Ala Glu Lys Leu Ala Ile Ala Phe Gly Leu Leu Asn Thr Pro Gln		
65	70	75
Gly Gln Thr Leu Arg Val Thr Lys Asn Leu Arg Met Cys Asn Asp Cys		
	85	90
His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg Arg Glu Ile Ile		
	100	110
Leu Arg Asp Ala Ser Cys Ile His His Phe Lys Asp Gly Leu Cys Ser		
	115	120
Cys Gly Asp Met Phe		
	130	
<210>	86	

<220>

<223> KP23

<400> 89

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

<210> 90

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> 嵌合KP1a

<400> 90

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

	85	90	95	
Cys Asn Asp	Cys His Asn Ala Ser Lys Ile Ile Ser Lys Ile Glu Arg			
	100	105	110	
Arg Glu Ile Ile Leu Arg Asp Asp Cys Cys Ile His His Phe Lys Asp				
	115	120	125	
Gly Leu Cys Ser Cys Gly Asp Tyr Trp				
	130	135		
<210>	91			
<211>	22			
<212>	RNA			
<213>	人工序列			
<220>				
<223>	ND2			
<400>	91			
cccagcuacg caaaaucuuu gc				22
<210>	92			
<211>	22			
<212>	RNA			
<213>	人工序列			
<220>				
<223>	ND5			
<400>	92			
augacaucaa aaaaaucgua gc				22
<210>	93			
<211>	830			
<212>	PRT			
<213>	人工序列			
<220>				
<223>	MTS-P-PG1_ND2			
<400>	93			
Met Ala Leu Leu Arg Ala Ala Val Ser Glu Leu Arg Arg Arg Gly Arg				
1	5	10	15	
Gly Ala Leu Thr Pro Leu Pro Ala Leu Ser Ser Leu Leu Ser Ser Leu				
	20	25	30	
Ser Pro Arg Ser Pro Ala Ser Thr Arg Pro Glu Pro Asn Asn Pro His				
	35	40	45	
Ala Asp Arg Arg His Val Ile Ala Leu Arg Arg Cys Pro Pro Leu Pro				
	50	55	60	
Ala Ser Ala Val Leu Ala Pro Glu Leu Leu His Ala Arg Gly Leu Leu				

65	70	75	80
Pro Arg His Trp Ser	His Ala Ser Pro	Leu Ser Thr Ser	Ser Ser Ser Ser
	85	90	95
Ser Arg Pro Ala Asp	Lys Ala Gln Leu	Thr Trp Val Asp	Lys Trp Ile
	100	105	110
Pro Glu Ala Ala Arg	Pro Tyr Ser His	Gly Ser Gly Gly	Ser Gly Gly
	115	120	125
Gly Leu His Met Gly	Ser Gly Gly Leu	Ser Ser Gly Asp	Tyr Lys Asp
	130	135	140
His Asp Gly Asp Tyr	Lys Asp His Asp	Ile Asp Tyr Lys	Asp Asp Asp
145	150	155	160
Asp Lys Gly Ser Gly	Gly Ser Gly Gly	Ser Gly Asn Ser	Val Val Thr
	165	170	175
Tyr Asn Thr Leu Ile	Asp Gly Leu Cys	Lys Ser Gly Lys	Ile Glu Glu
	180	185	190
Ala Leu Lys Leu Phe	Lys Glu Met Glu	Glu Lys Gly Ile	Thr Pro Ser
	195	200	205
Val Val Thr Tyr Asn	Thr Leu Ile Asp	Gly Leu Cys Lys	Ser Gly Lys
	210	215	220
Ile Glu Glu Ala Leu	Lys Leu Phe Lys	Glu Met Glu Glu	Lys Gly Ile
225	230	235	240
Thr Pro Ser Val Val	Thr Tyr Asn Thr	Leu Ile Asp Gly	Leu Cys Lys
	245	250	255
Ser Gly Lys Ile Glu	Glu Ala Leu Lys	Leu Phe Lys Glu	Met Glu Glu
	260	265	270
Lys Gly Ile Thr Pro	Ser Val Val Thr	Tyr Thr Thr Leu	Ile Asp Gly
	275	280	285
Leu Cys Lys Ala Gly	Lys Val Asp Glu	Ala Leu Glu Leu	Phe Lys Glu
	290	295	300
Met Arg Ser Lys Gly	Val Lys Pro Asn	Val Val Thr Tyr	Thr Thr Leu
305	310	315	320
Ile Asp Gly Leu Cys	Lys Ala Gly Lys	Val Asp Glu Ala	Leu Glu Leu
	325	330	335
Phe Asp Glu Met Lys	Glu Arg Gly Ile	Lys Pro Asp Val	Val Thr Tyr
	340	345	350
Asn Thr Leu Ile Asp	Gly Leu Cys Lys	Ser Gly Lys Ile	Glu Glu Ala
	355	360	365
Leu Lys Leu Phe Lys	Glu Met Glu Glu	Lys Gly Ile Thr	Pro Ser Val
	370	375	380

Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Arg Leu
 385 390 395 400
 Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu Lys Gly Ile Lys
 405 410 415
 Pro Asp Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala
 420 425 430
 Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys
 435 440 445
 Gly Val Lys Pro Asn Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu
 450 455 460
 Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met
 465 470 475 480
 Glu Glu Lys Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile
 485 490 495
 Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe
 500 505 510
 Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Ser Trp Asn
 515 520 525
 Ala Met Ile Ala Ala Tyr Ala Gln His Gly His Gly Lys Glu Ala Leu
 530 535 540
 Gln Leu Phe Gln Gln Met Gln Gln Glu Gly Val Lys Pro Ser Glu Val
 545 550 555 560
 Thr Phe Thr Ser Ile Leu Ser Ala Cys Ser His Ala Gly Leu Val Asp
 565 570 575
 Glu Gly His His Tyr Phe Glu Ser Met Ser Pro Asp Tyr Gly Ile Thr
 580 585 590
 Pro Arg Val Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala
 595 600 605
 Gly Arg Leu Asp Glu Ala Glu Asp Leu Ile Lys Ser Met Pro Phe Gln
 610 615 620
 Pro Asn Val Val Val Trp Gly Thr Leu Leu Gly Ala Cys Arg Val His
 625 630 635 640
 Gly Asp Val Glu Arg Gly Glu Arg Ala Ala Glu Arg Ile Leu Glu Leu
 645 650 655
 Asp Pro Glu Ser Ala Ala Pro Tyr Val Leu Leu Ser Asn Ile Tyr Ala
 660 665 670
 Ala Ala Gly Arg Trp Asp Glu Ala Ala Lys Val Arg Lys Leu Met Lys
 675 680 685
 Glu Arg Gly Val Lys Lys Glu Pro Gly Cys Ser Trp Ile Glu Val Asn

690	695	700
Asn Lys Val His Glu Phe	Val Ala Gly Asp Lys Ser His Pro Gln Thr	
705	710	715
Lys Glu Ile Tyr Ala Glu Leu Glu Arg Leu Ser Lys Gln Met Lys Glu		
	725	730
Ala Gly Tyr Val Pro Asp Thr Lys Phe Val Leu His Asp Val Glu Glu		
	740	745
Glu Glu Lys Glu Gln Leu Leu Cys Tyr His Ser Glu Lys Leu Ala Ile		
	755	760
Ala Phe Gly Leu Ile Ser Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile		
	770	780
Lys Asn Leu Arg Val Cys Gly Asp Cys His Thr Ala Thr Lys Phe Ile		
785	790	795
Ser Lys Ile Val Gly Arg Glu Ile Val Val Arg Asp Ala Asn Arg Phe		
	805	810
His His Phe Lys Asp Gly Val Cys Ser Cys Gly Asp Tyr Trp		
	820	825
		830
<210> 94		
<211> 831		
<212> PRT		
<213> 人工序列		
<220>		
<223> MTS-P-WW1_ND2		
<400> 94		
Met Ala Leu Leu Arg Ala Ala Val Ser Glu Leu Arg Arg Arg Gly Arg		
1	5	10
Gly Ala Leu Thr Pro Leu Pro Ala Leu Ser Ser Leu Leu Ser Ser Leu		
	20	25
Ser Pro Arg Ser Pro Ala Ser Thr Arg Pro Glu Pro Asn Asn Pro His		
	35	40
Ala Asp Arg Arg His Val Ile Ala Leu Arg Arg Cys Pro Pro Leu Pro		
	50	55
Ala Ser Ala Val Leu Ala Pro Glu Leu Leu His Ala Arg Gly Leu Leu		
65	70	75
Pro Arg His Trp Ser His Ala Ser Pro Leu Ser Thr Ser Ser Ser Ser		
	85	90
Ser Arg Pro Ala Asp Lys Ala Gln Leu Thr Trp Val Asp Lys Trp Ile		
	100	105
Pro Glu Ala Ala Arg Pro Tyr Ser His Gly Ser Gly Gly Ser Gly Gly		

115	120	125
Gly Leu His Met Gly Ser	Gly Gly Leu Ser Ser	Gly Asp Tyr Lys Asp
130	135	140
His Asp Gly Asp Tyr Lys Asp	His Asp Ile Asp Tyr Lys Asp Asp Asp	
145	150	155
Asp Lys Gly Ser Gly Gly Ser	Gly Gly Ser Gly Asn Ser Val Val Thr	
165	170	175
Tyr Asn Thr Leu Ile Asp Gly	Leu Cys Lys Ser Gly Lys Ile Glu Glu	
180	185	190
Ala Leu Lys Leu Phe Lys Glu	Met Glu Glu Lys Gly Ile Thr Pro Ser	
195	200	205
Val Val Thr Tyr Asn Thr Leu	Ile Asp Gly Leu Cys Lys Ser Gly Lys	
210	215	220
Ile Glu Glu Ala Leu Lys Leu	Phe Lys Glu Met Glu Glu Lys Gly Ile	
225	230	235
Thr Pro Ser Val Val Thr Tyr	Asn Thr Leu Ile Asp Gly Leu Cys Lys	
245	250	255
Ser Gly Lys Ile Glu Glu Ala	Leu Lys Leu Phe Lys Glu Met Glu Glu	
260	265	270
Lys Gly Ile Thr Pro Ser Val	Val Thr Tyr Thr Thr Leu Ile Asp Gly	
275	280	285
Leu Cys Lys Ala Gly Lys Val	Asp Glu Ala Leu Glu Leu Phe Lys Glu	
290	295	300
Met Arg Ser Lys Gly Val Lys	Pro Asn Val Val Thr Tyr Thr Thr Leu	
305	310	315
Ile Asp Gly Leu Cys Lys Ala	Gly Lys Val Asp Glu Ala Leu Glu Leu	
325	330	335
Phe Asp Glu Met Lys Glu Arg	Gly Ile Lys Pro Asp Val Val Thr Tyr	
340	345	350
Asn Thr Leu Ile Asp Gly Leu	Cys Lys Ser Gly Lys Ile Glu Glu Ala	
355	360	365
Leu Lys Leu Phe Lys Glu Met	Glu Glu Lys Gly Ile Thr Pro Ser Val	
370	375	380
Val Thr Tyr Asn Thr Leu Ile	Asp Gly Leu Cys Lys Ala Gly Arg Leu	
385	390	395
Asp Glu Ala Glu Glu Leu Leu	Glu Glu Met Glu Glu Lys Gly Ile Lys	
405	410	415
Pro Asp Val Val Thr Tyr Thr	Thr Leu Ile Asp Gly Leu Cys Lys Ala	
420	425	430

Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys
 435 440 445
 Gly Val Lys Pro Asn Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu
 450 455 460
 Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met
 465 470 475 480
 Glu Glu Lys Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile
 485 490 495
 Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe
 500 505 510
 Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr Trp Asn
 515 520 525
 Ala Leu Ile Ala Gly Tyr Ala Arg Gln Gly Glu Ser Asp Leu Val Phe
 530 535 540
 His Leu Leu Glu Arg Met Arg Gln Glu Gly Ile Gln Pro Ser Gly Val
 545 550 555 560
 Thr Phe Thr Ser Val Leu Thr Val Cys Ser His Ala Gly Leu Val Asp
 565 570 575
 Glu Gly Gln Lys Tyr Phe Asp Ala Met Ser Glu Asp Tyr Gly Ile Thr
 580 585 590
 Pro Arg Ile Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala
 595 600 605
 Gly Gln Met Asp Glu Ala Val Ala Met Val Glu Lys Met Pro Phe Gln
 610 615 620
 Pro Asn Leu Val Thr Trp Gly Thr Leu Leu Gly Ala Cys Arg Lys Trp
 625 630 635 640
 Asn Asn Val Glu Ile Gly Arg His Ala Phe Glu Cys Ala Val Arg Leu
 645 650 655
 Asp Glu Lys Ser Ala Ala Ala Tyr Val Leu Met Ser Asn Ile Tyr Ala
 660 665 670
 Asp Ala His Met Trp Glu Glu Arg Asp Lys Ile Gln Ala Met Arg Lys
 675 680 685
 Asn Ala Arg Ala Trp Lys Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr
 690 695 700
 Gly Gly Ile Val His Thr Phe Val Val Gly Asp Thr Lys His Pro Gln
 705 710 715 720
 Ser Gln Asp Ile Tyr Ala Lys Leu Lys Asp Leu Tyr Val Lys Met Lys
 725 730 735
 Glu Glu Gly Tyr Val Pro His Leu Asp Cys Val Leu Trp Asp Ile Ser

	740		745		750														
Asp	Asp	Glu	Lys	Glu	Asp	Ala	Leu	Cys	Gly	His	Ser	Glu	Lys	Leu	Ala				
	755						760					765							
Ile	Ala	Cys	Ala	Leu	Ile	Asn	Thr	Pro	Pro	Gly	Thr	Pro	Ile	Arg	Ile				
	770						775					780							
Val	Lys	Asn	Leu	Arg	Val	Cys	Asp	Asp	Cys	His	Lys	Ala	Ile	Ala	Leu				
785						790				795					800				
Ile	Ser	Lys	Ile	Glu	Gly	Arg	Asn	Ile	Ile	Cys	Arg	Asp	Ala	Ser	Arg				
				805						810					815				
Phe	His	Asn	Tyr	Lys	Asp	Gly	Lys	Cys	Ser	Cys	Gly	Asp	Tyr	Trp					
			820					825						830					
<210>	95																		
<211>	830																		
<212>	PRT																		
<213>	人工序列																		
<220>																			
<223>	MTS-P-PG1_ND5																		
<400>	95																		
Met	Ala	Leu	Leu	Arg	Ala	Ala	Val	Ser	Glu	Leu	Arg	Arg	Arg	Gly	Arg				
1				5					10					15					
Gly	Ala	Leu	Thr	Pro	Leu	Pro	Ala	Leu	Ser	Ser	Leu	Leu	Ser	Ser	Leu				
			20					25					30						
Ser	Pro	Arg	Ser	Pro	Ala	Ser	Thr	Arg	Pro	Glu	Pro	Asn	Asn	Pro	His				
		35					40					45							
Ala	Asp	Arg	Arg	His	Val	Ile	Ala	Leu	Arg	Arg	Cys	Pro	Pro	Leu	Pro				
	50					55					60								
Ala	Ser	Ala	Val	Leu	Ala	Pro	Glu	Leu	Leu	His	Ala	Arg	Gly	Leu	Leu				
65					70					75					80				
Pro	Arg	His	Trp	Ser	His	Ala	Ser	Pro	Leu	Ser	Thr	Ser	Ser	Ser	Ser				
				85					90					95					
Ser	Arg	Pro	Ala	Asp	Lys	Ala	Gln	Leu	Thr	Trp	Val	Asp	Lys	Trp	Ile				
			100						105				110						
Pro	Glu	Ala	Ala	Arg	Pro	Tyr	Ser	His	Gly	Ser	Gly	Gly	Ser	Gly	Gly				
	115						120						125						
Gly	Leu	His	Met	Gly	Ser	Gly	Gly	Leu	Ser	Ser	Gly	Asp	Tyr	Lys	Asp				
	130					135					140								
His	Asp	Gly	Asp	Tyr	Lys	Asp	His	Asp	Ile	Asp	Tyr	Lys	Asp	Asp	Asp				
145					150				155						160				
Asp	Lys	Gly	Ser	Gly	Gly	Ser	Gly	Gly	Ser	Gly	Asn	Ser	Val	Val	Thr				

				165						170					175
Tyr	Thr	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu
				180						185					190
Ala	Leu	Glu	Leu	Phe	Lys	Glu	Met	Arg	Ser	Lys	Gly	Val	Lys	Pro	Asn
				195						200					205
Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ala	Gly	Arg
				210						215					220
Leu	Asp	Glu	Ala	Glu	Glu	Leu	Leu	Glu	Glu	Met	Glu	Glu	Lys	Gly	Ile
225										230					235
Lys	Pro	Asp	Val	Val	Thr	Tyr	Thr	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys
										245					250
Ala	Gly	Lys	Val	Asp	Glu	Ala	Leu	Glu	Leu	Phe	Asp	Glu	Met	Lys	Glu
										260					265
Arg	Gly	Ile	Lys	Pro	Asp	Val	Val	Thr	Tyr	Thr	Thr	Leu	Ile	Asp	Gly
										275					280
Leu	Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu	Ala	Leu	Glu	Leu	Phe	Lys	Glu
										290					295
Met	Arg	Ser	Lys	Gly	Val	Lys	Pro	Asn	Val	Val	Thr	Tyr	Asn	Thr	Leu
305															310
Ile	Asp	Gly	Leu	Cys	Lys	Ser	Gly	Lys	Ile	Glu	Glu	Ala	Leu	Lys	Leu
										315					320
															325
															330
Phe	Lys	Glu	Met	Glu	Glu	Lys	Gly	Ile	Thr	Pro	Ser	Val	Val	Thr	Tyr
										335					340
															345
Thr	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu	Ala
										350					355
															360
Leu	Glu	Leu	Phe	Lys	Glu	Met	Arg	Ser	Lys	Gly	Val	Lys	Pro	Asn	Val
										365					370
															375
Val	Thr	Tyr	Asn	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ala	Gly	Arg	Leu
385															390
															395
Asp	Glu	Ala	Glu	Glu	Leu	Leu	Glu	Glu	Met	Glu	Glu	Lys	Gly	Ile	Lys
															400
															405
															410
Pro	Asp	Val	Val	Thr	Tyr	Asn	Thr	Leu	Ile	Asp	Gly	Leu	Cys	Lys	Ser
										415					420
															425
Gly	Lys	Ile	Glu	Glu	Ala	Leu	Lys	Leu	Phe	Lys	Glu	Met	Glu	Glu	Lys
															430
															435
															440
Gly	Ile	Thr	Pro	Ser	Val	Val	Thr	Tyr	Thr	Thr	Leu	Ile	Asp	Gly	Leu
															445
															450
															455
Cys	Lys	Ala	Gly	Lys	Val	Asp	Glu	Ala	Leu	Glu	Leu	Phe	Lys	Glu	Met
465															470
															475
															480

Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile
 485 490 495
 Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe
 500 505 510
 Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Ser Trp Thr
 515 520 525
 Ala Met Ile Ala Ala Tyr Ala Gln His Gly His Gly Lys Glu Ala Leu
 530 535 540
 Gln Leu Phe Gln Gln Met Gln Gln Glu Gly Val Lys Pro Asn Glu Val
 545 550 555 560
 Thr Phe Thr Ser Ile Leu Ser Ala Cys Ser His Ala Gly Leu Val Asp
 565 570 575
 Glu Gly His His Tyr Phe Glu Ser Met Ser Pro Asp Tyr Gly Ile Thr
 580 585 590
 Pro Arg Val Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala
 595 600 605
 Gly Arg Leu Asp Glu Ala Glu Asp Leu Ile Lys Ser Met Pro Phe Gln
 610 615 620
 Pro Asn Val Val Val Trp Gly Thr Leu Leu Gly Ala Cys Arg Val His
 625 630 635 640
 Gly Asp Val Glu Arg Gly Glu Arg Ala Ala Glu Arg Ile Leu Glu Leu
 645 650 655
 Asp Pro Glu Ser Ala Ala Pro Tyr Val Leu Leu Ser Asn Ile Tyr Ala
 660 665 670
 Ala Ala Gly Arg Trp Asp Glu Ala Ala Lys Val Arg Lys Leu Met Lys
 675 680 685
 Glu Arg Gly Val Lys Lys Glu Pro Gly Cys Ser Trp Ile Glu Val Asn
 690 695 700
 Asn Lys Val His Glu Phe Val Ala Gly Asp Lys Ser His Pro Gln Thr
 705 710 715 720
 Lys Glu Ile Tyr Ala Glu Leu Glu Arg Leu Ser Lys Gln Met Lys Glu
 725 730 735
 Ala Gly Tyr Val Pro Asp Thr Lys Phe Val Leu His Asp Val Glu Glu
 740 745 750
 Glu Glu Lys Glu Gln Leu Leu Cys Tyr His Ser Glu Lys Leu Ala Ile
 755 760 765
 Ala Phe Gly Leu Ile Ser Thr Pro Pro Gly Thr Pro Leu Arg Ile Ile
 770 775 780
 Lys Asn Leu Arg Val Cys Gly Asp Cys His Thr Ala Thr Lys Phe Ile

785	790	795	800
Ser Lys Ile Val Gly Arg Glu Ile Val Val Arg Asp Ala Asn Arg Phe			
	805	810	815
His His Phe Lys Asp Gly Val Cys Ser Cys Gly Asp Tyr Trp			
	820	825	830
<210> 96			
<211> 831			
<212> PRT			
<213> 人工序列			
<220>			
<223> MTS-P-WW1_ND5			
<400> 96			
Met Ala Leu Leu Arg Ala Ala Val Ser Glu Leu Arg Arg Arg Gly Arg			
1	5	10	15
Gly Ala Leu Thr Pro Leu Pro Ala Leu Ser Ser Leu Leu Ser Ser Leu			
	20	25	30
Ser Pro Arg Ser Pro Ala Ser Thr Arg Pro Glu Pro Asn Asn Pro His			
	35	40	45
Ala Asp Arg Arg His Val Ile Ala Leu Arg Arg Cys Pro Pro Leu Pro			
	50	55	60
Ala Ser Ala Val Leu Ala Pro Glu Leu Leu His Ala Arg Gly Leu Leu			
65	70	75	80
Pro Arg His Trp Ser His Ala Ser Pro Leu Ser Thr Ser Ser Ser Ser			
	85	90	95
Ser Arg Pro Ala Asp Lys Ala Gln Leu Thr Trp Val Asp Lys Trp Ile			
	100	105	110
Pro Glu Ala Ala Arg Pro Tyr Ser His Gly Ser Gly Gly Ser Gly Gly			
	115	120	125
Gly Leu His Met Gly Ser Gly Gly Leu Ser Ser Gly Asp Tyr Lys Asp			
	130	135	140
His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp Tyr Lys Asp Asp Asp			
145	150	155	160
Asp Lys Gly Ser Gly Gly Ser Gly Gly Ser Gly Asn Ser Val Val Thr			
	165	170	175
Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu			
	180	185	190
Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn			
	195	200	205
Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Arg			

210	215	220
Leu Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu Lys Gly Ile		
225	230	235
Lys Pro Asp Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys		240
	245	250
Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu		255
	260	265
Arg Gly Ile Lys Pro Asp Val Val Thr Tyr Thr Thr Leu Ile Asp Gly		270
	275	280
Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Lys Glu		285
	290	300
Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr Asn Thr Leu		
305	310	315
Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu		
	325	330
Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val Val Thr Tyr		335
	340	345
Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala		350
	355	360
Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val		365
	370	375
Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Arg Leu		380
385	390	395
Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu Lys Gly Ile Lys		400
	405	410
Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser		415
	420	425
Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys		430
	435	440
Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu		445
	450	455
Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met		460
465	470	475
Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile		480
	485	490
Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe		495
	500	505
Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Trp Thr		510
	515	520
		525

Ala Leu Ile Ala Gly Tyr Ala Arg Gln Gly Glu Ser Asp Leu Val Phe
530 535 540

His Leu Leu Glu Arg Met Arg Gln Glu Gly Ile Gln Pro Asn Gly Val
545 550 555 560

Thr Phe Thr Ser Val Leu Thr Val Cys Ser His Ala Gly Leu Val Asp
565 570 575

Glu Gly Gln Lys Tyr Phe Asp Ala Met Ser Glu Asp Tyr Gly Ile Thr
580 585 590

Pro Arg Ile Glu His Tyr Gly Cys Met Val Asp Leu Leu Gly Arg Ala
595 600 605

Gly Gln Met Asp Glu Ala Val Ala Met Val Glu Lys Met Pro Phe Gln
610 615 620

Pro Asn Leu Val Thr Trp Gly Thr Leu Leu Gly Ala Cys Arg Lys Trp
625 630 635 640

Asn Asn Val Glu Ile Gly Arg His Ala Phe Glu Cys Ala Val Arg Leu
645 650 655

Asp Glu Lys Ser Ala Ala Ala Tyr Val Leu Met Ser Asn Ile Tyr Ala
660 665 670

Asp Ala His Met Trp Glu Glu Arg Asp Lys Ile Gln Ala Met Arg Lys
675 680 685

Asn Ala Arg Ala Trp Lys Lys Pro Gly Gln Ser Trp Trp Thr Asp Thr
690 695 700

Gly Gly Ile Val His Thr Phe Val Val Gly Asp Thr Lys His Pro Gln
705 710 715 720

Ser Gln Asp Ile Tyr Ala Lys Leu Lys Asp Leu Tyr Val Lys Met Lys
725 730 735

Glu Glu Gly Tyr Val Pro His Leu Asp Cys Val Leu Trp Asp Ile Ser
740 745 750

Asp Asp Glu Lys Glu Asp Ala Leu Cys Gly His Ser Glu Lys Leu Ala
755 760 765

Ile Ala Cys Ala Leu Ile Asn Thr Pro Pro Gly Thr Pro Ile Arg Ile
770 775 780

Val Lys Asn Leu Arg Val Cys Asp Asp Cys His Lys Ala Ile Ala Leu
785 790 795 800

Ile Ser Lys Ile Glu Gly Arg Asn Ile Ile Cys Arg Asp Ala Ser Arg
805 810 815

Phe His Asn Tyr Lys Asp Gly Lys Cys Ser Cys Gly Asp Tyr Trp
820 825 830

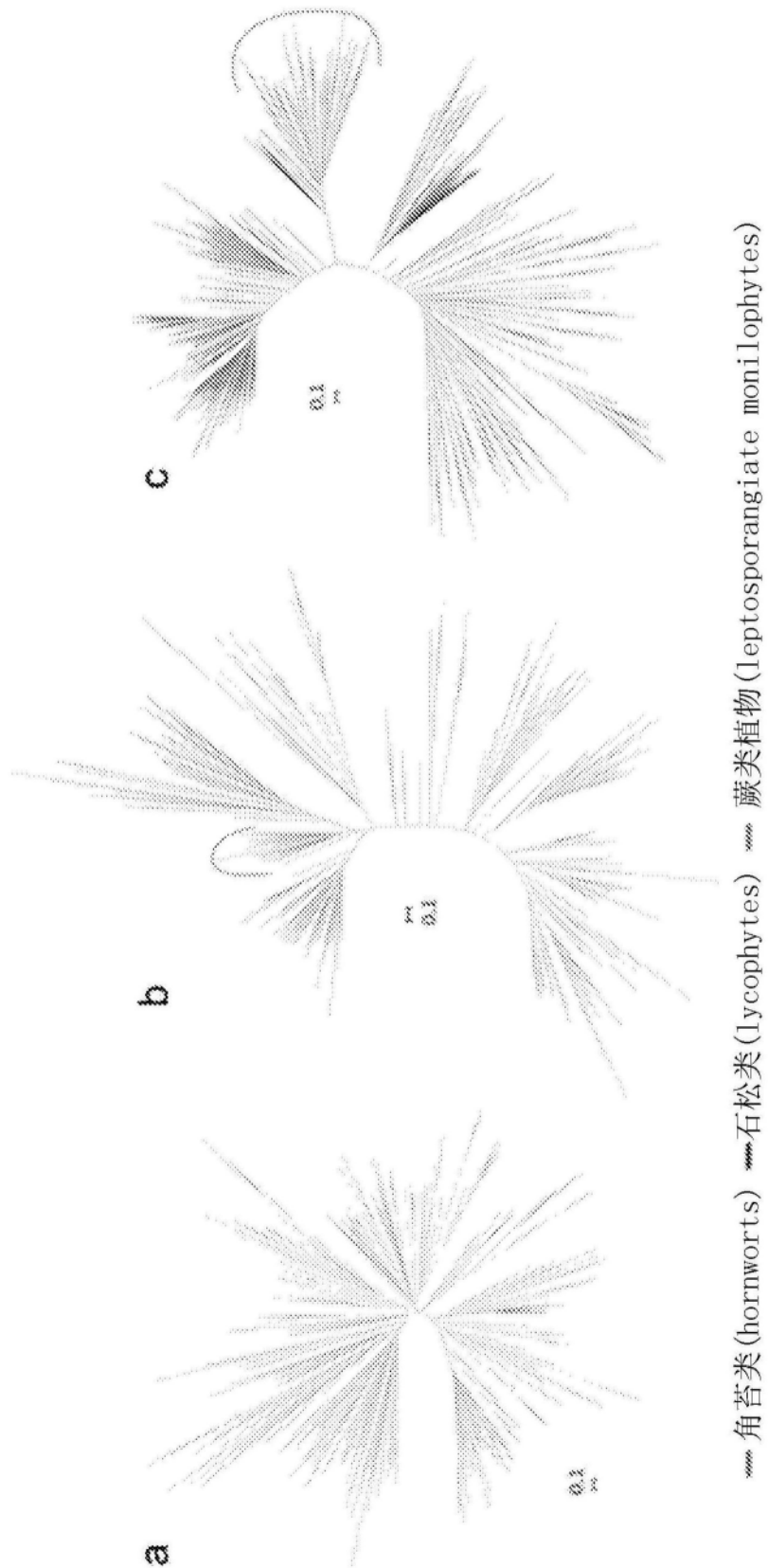


图1

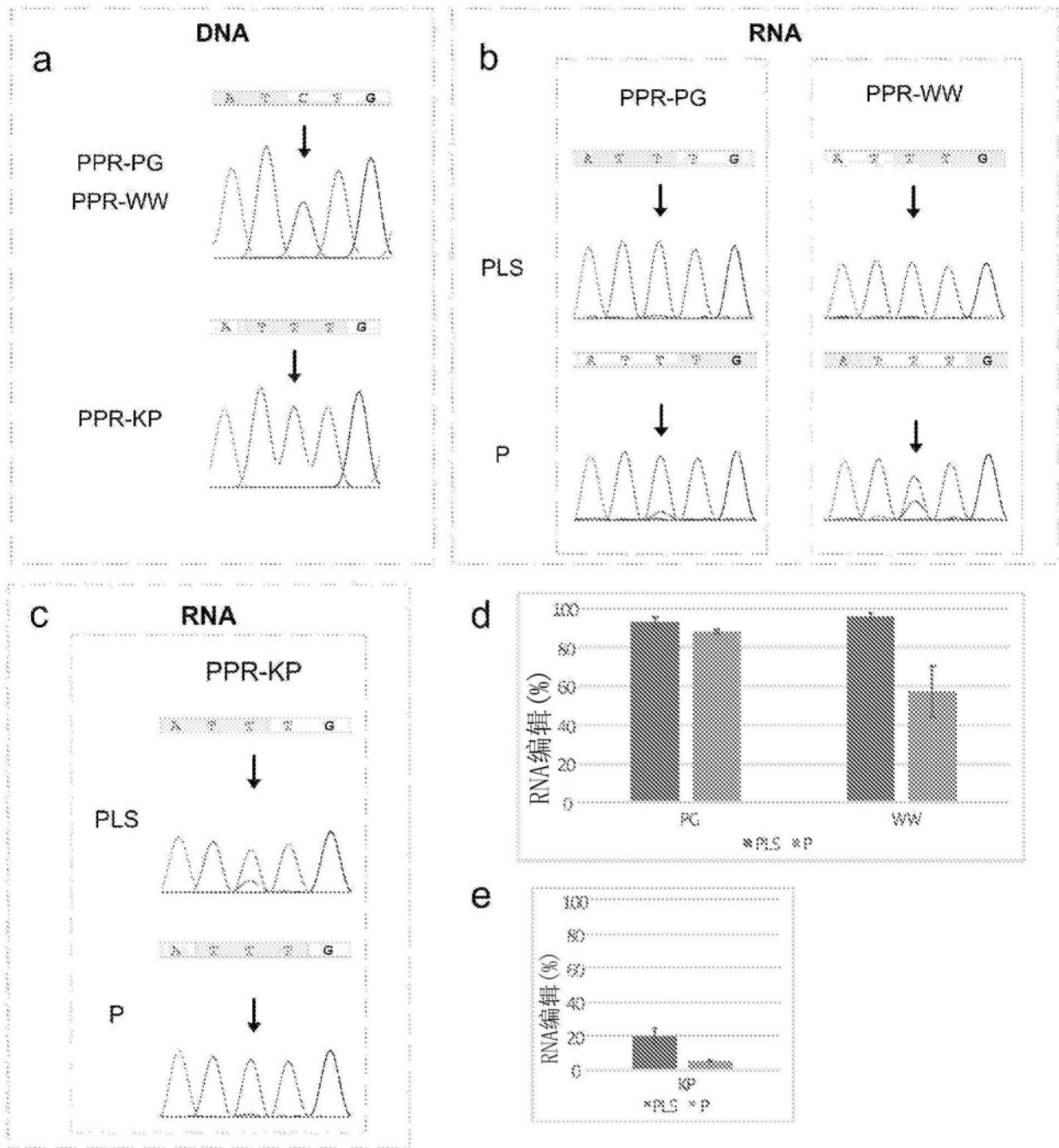


图3

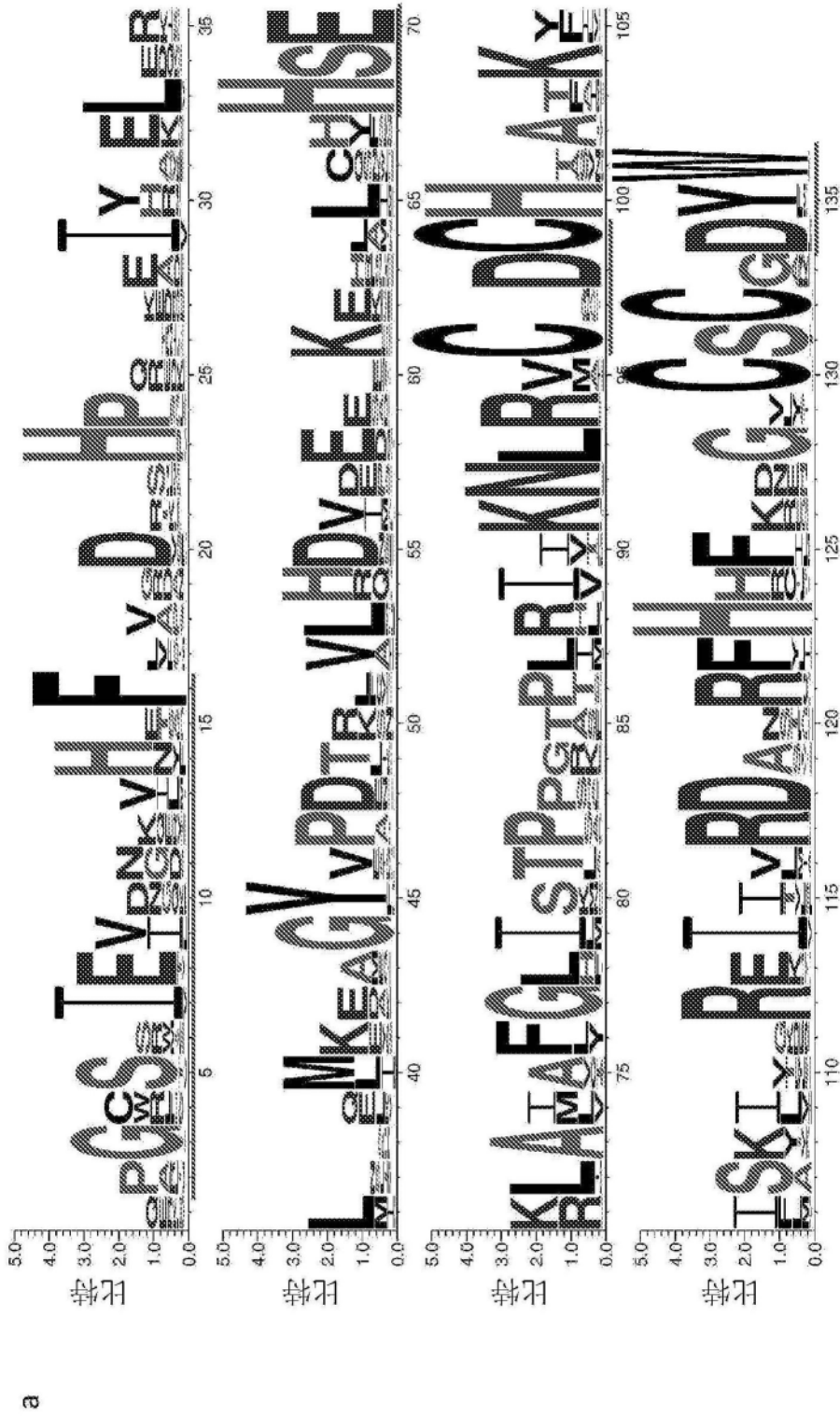


图4-1

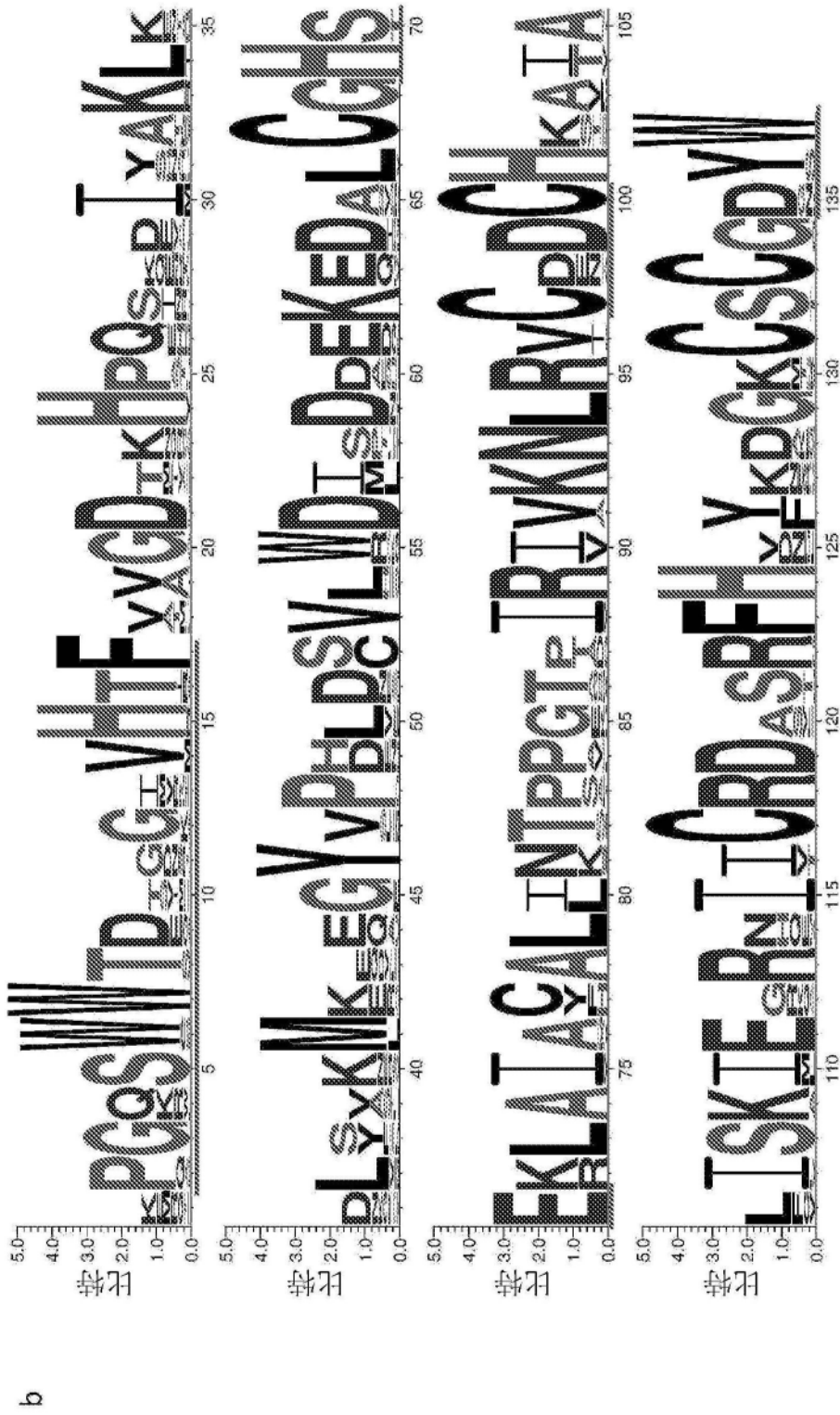


图4-2

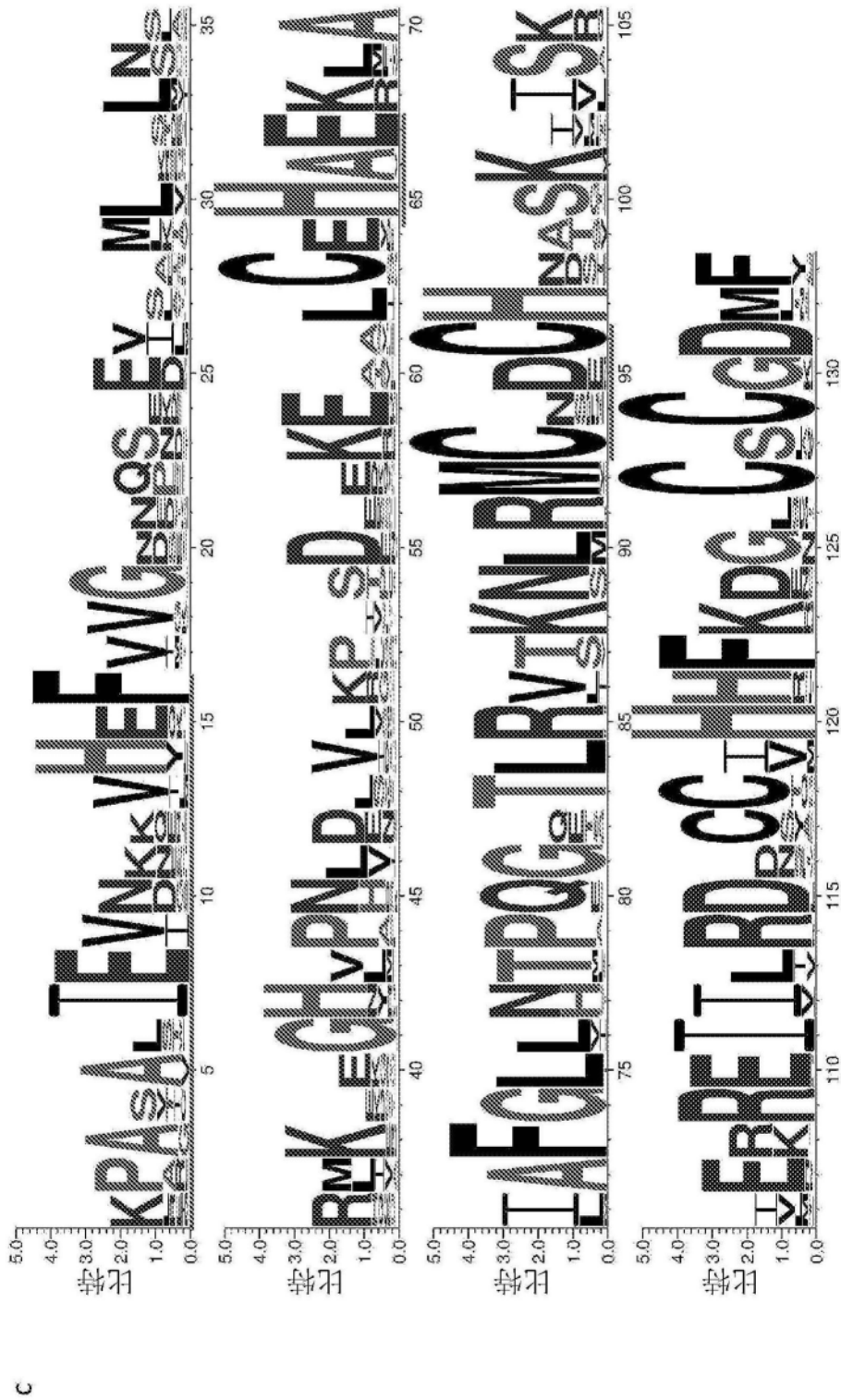


图4-3

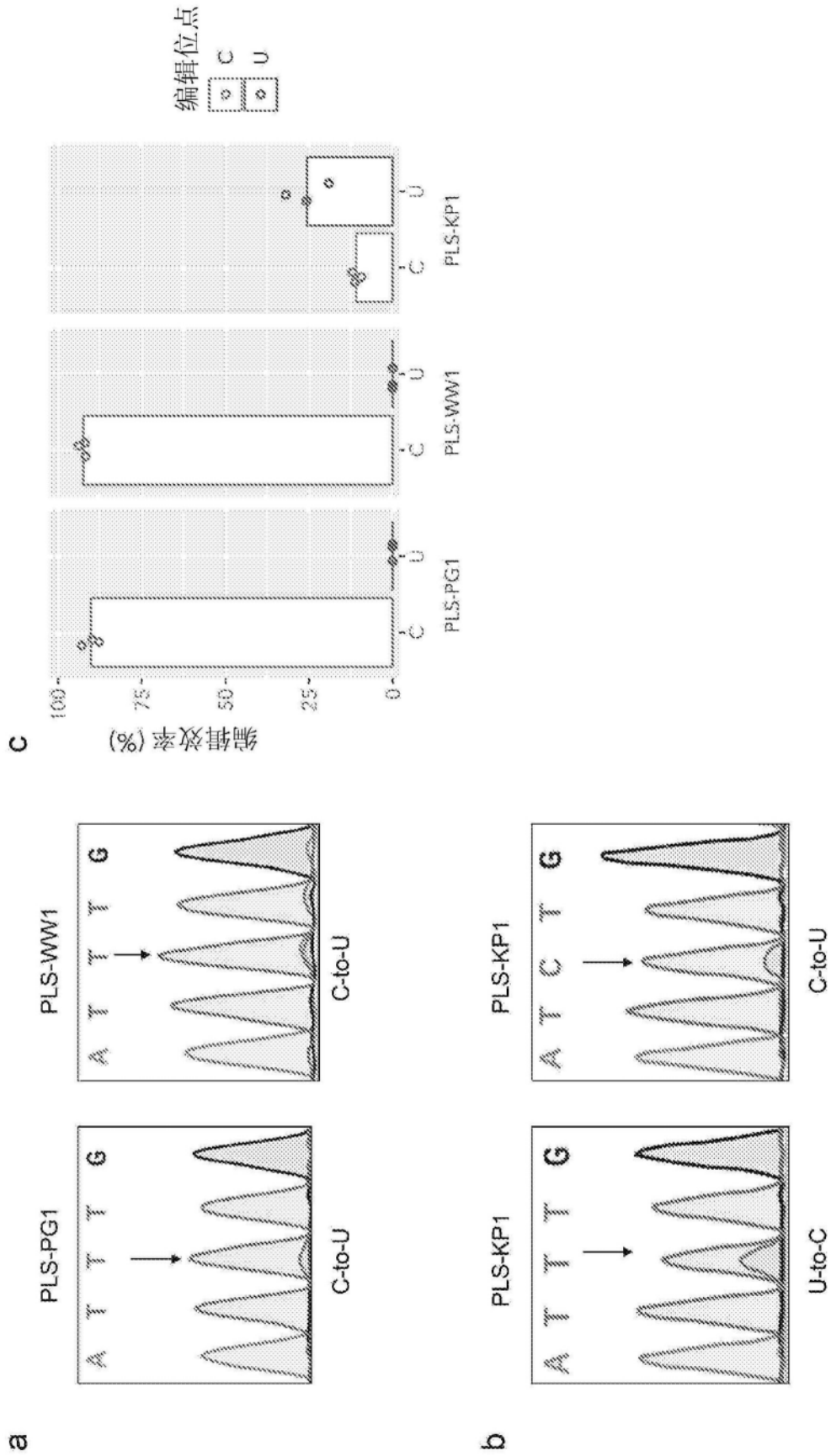


图5

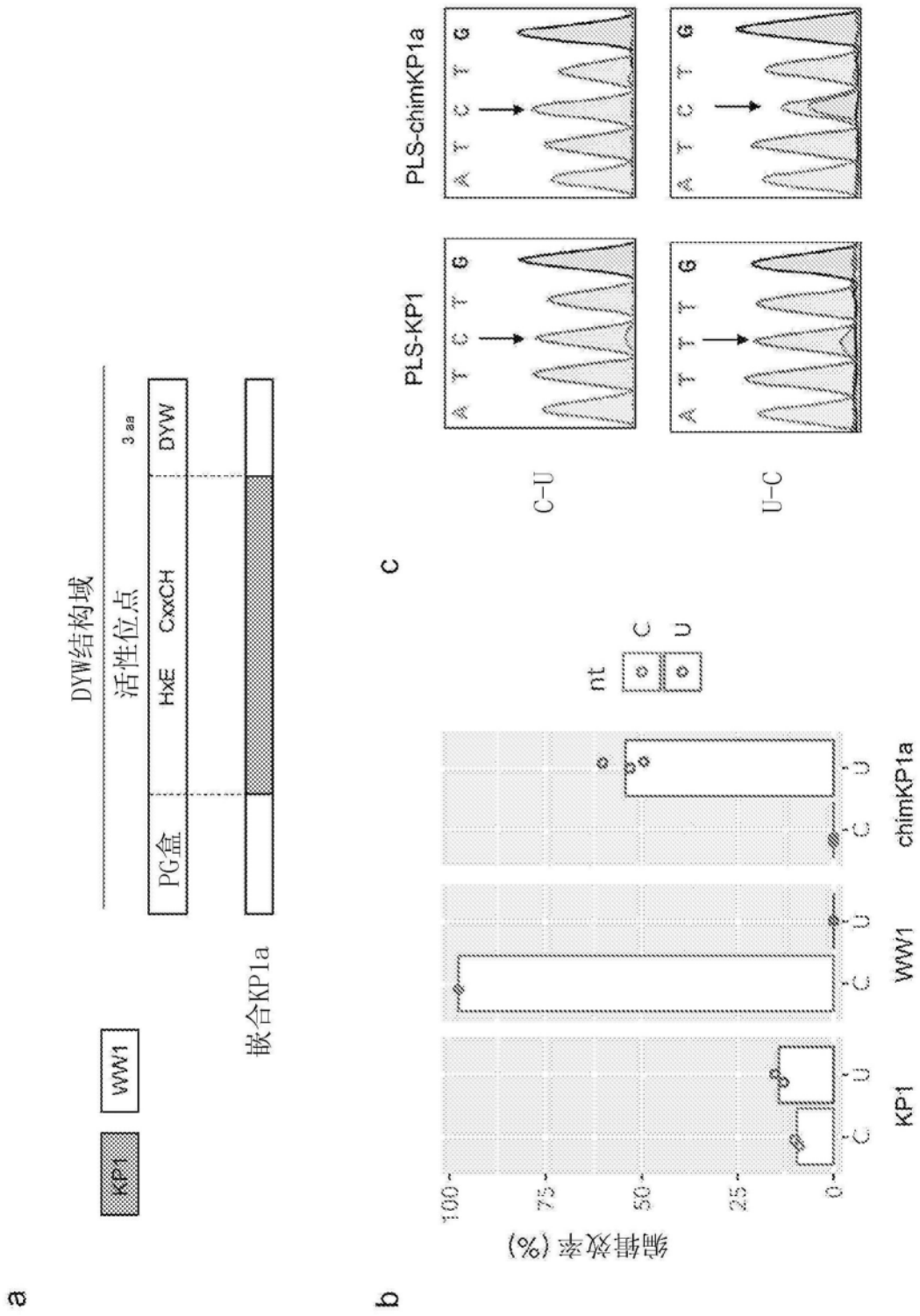


图6

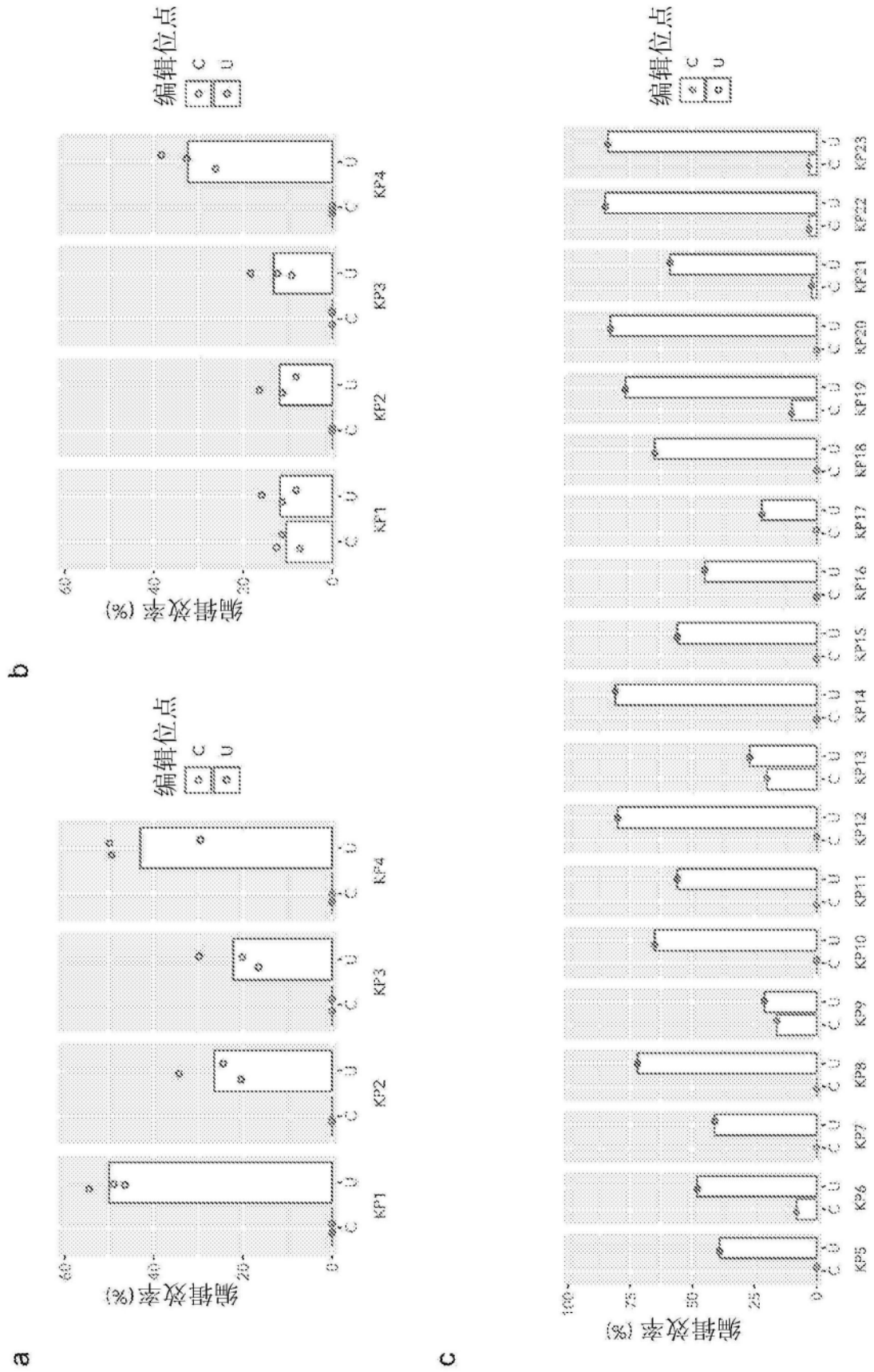


图7

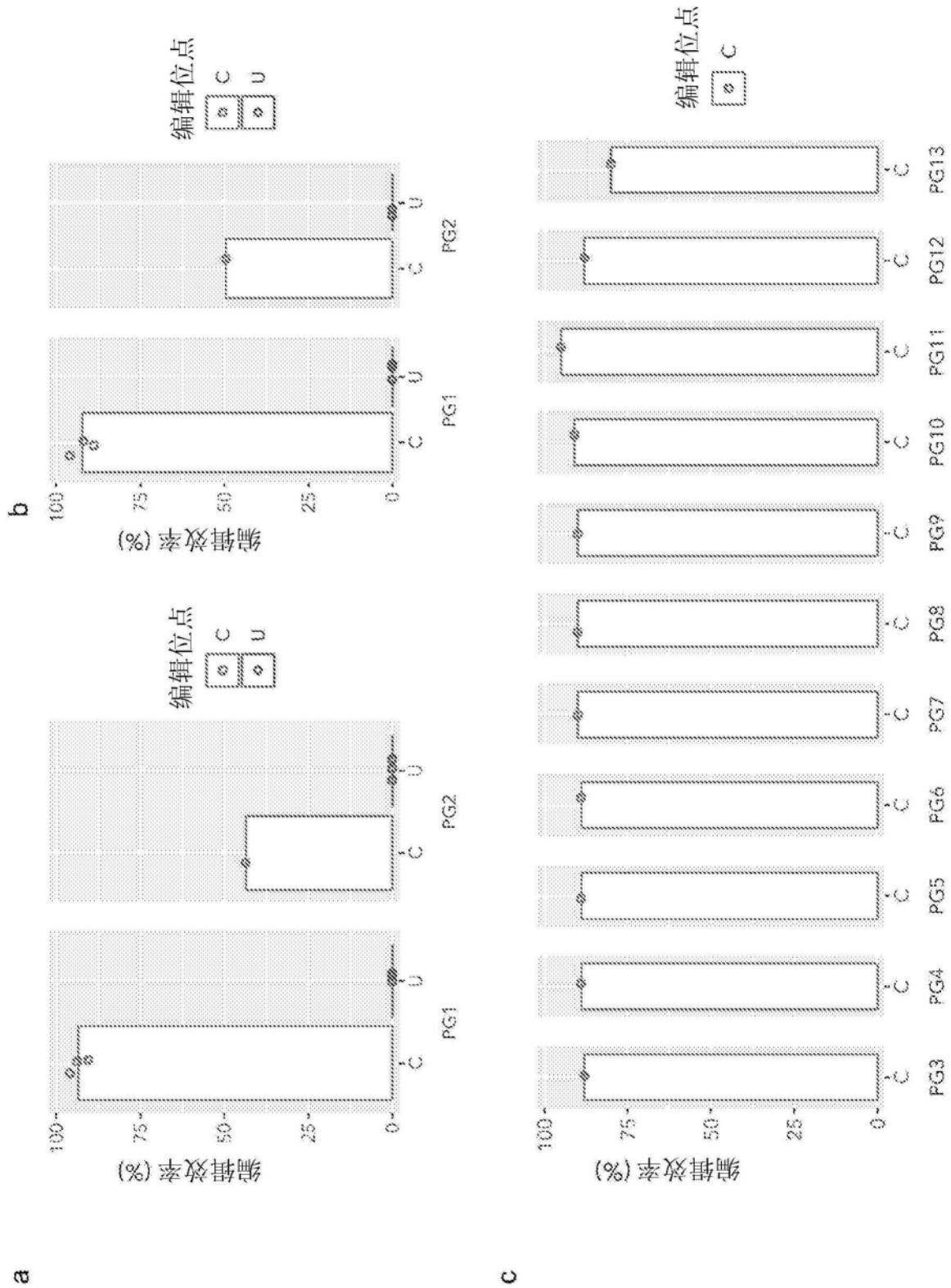


图8

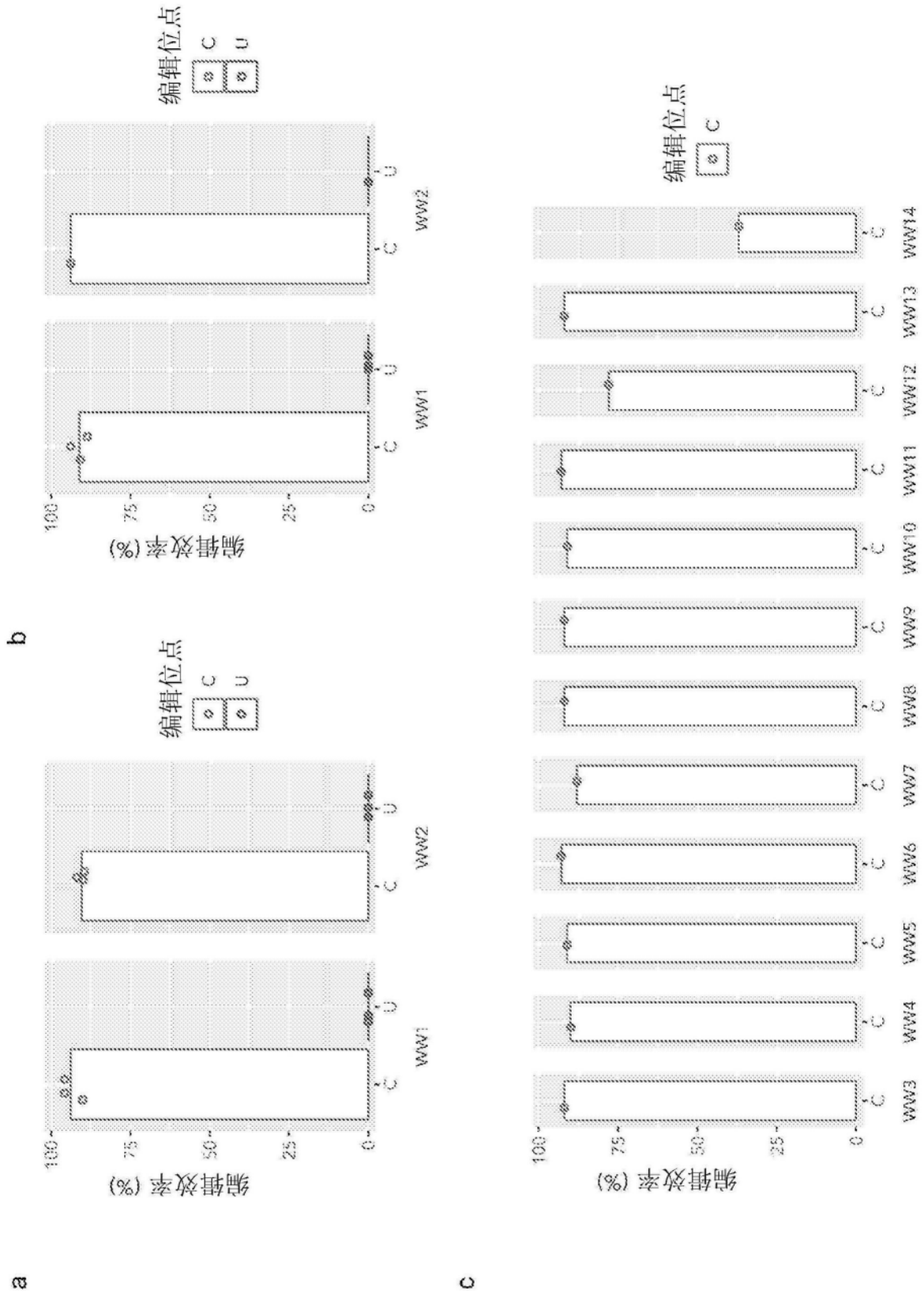


图9

