



(12) 发明专利申请

(10) 申请公布号 CN 112111504 A

(43) 申请公布日 2020.12.22

(21) 申请号 202010952359.4

G16B 30/00 (2019.01)

(22) 申请日 2020.09.11

C12R 1/19 (2006.01)

(71) 申请人 武汉海特生物制药股份有限公司
地址 430056 湖北省武汉市经济技术开发区海特科技园(珠山湖大道66号)

(72) 发明人 汤华东 李冠霖 柳余莉 欧阳聪
杨薇 高思偲 尹领鹏 谢倩
杨琴霞 张盼盼

(74) 专利代理机构 北京汇泽知识产权代理有限公司 11228

代理人 徐瑛

(51) Int. Cl.

C12N 15/62 (2006.01)

C12N 15/70 (2006.01)

C12N 1/21 (2006.01)

权利要求书2页 说明书10页
序列表7页 附图9页

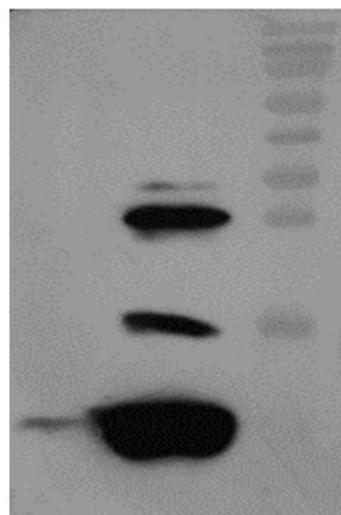
(54) 发明名称

一种筛选酶切适配融合蛋白的方法及IGF-I制备方法

(57) 摘要

本发明公开了一种筛选酶切适配融合蛋白的方法及IGF-I制备方法,该筛选方法是将目标蛋白序列编码基因分别插入到多个不同的含蛋白酶酶切位点的表达载体中,获得多个不同融合蛋白的氨基酸序列;所得融合蛋白的氨基酸序列进行二级结构预测;将不同融合蛋白氨基酸序列及其对应的二级结构预测结果作为输入文件,预测三维结构;三维结构预测结果与该融合蛋白相对应的蛋白酶采用Cluspro算法进行分子对接,选择对接结果正确比例超过阈值的酶切体系进行实验验证,筛选出蛋白酶最适配的融合蛋白。通过上述方法筛选出适用于IGF-I的凝血酶和HRV 3C蛋白酶可溶表达体系,具有稳定、可溶、高活性大量表达、酶切方便等优点。

1 2 M kDa



1. 一种筛选酶切适配融合蛋白的方法,其特征在于,包括以下步骤:

(1) 将目标蛋白序列编码基因分别插入到多个不同的含蛋白酶酶切位点的表达载体中,获得多个不同融合蛋白的氨基酸序列;所得融合蛋白的氨基酸序列通过二级结构预测算法PSIPRED进行二级结构预测;

(2) 将步骤(1)中的不同融合蛋白氨基酸序列及其对应的二级结构预测结果作为输入文件,通过结构预测算法I-TASSER预测三维结构;

(3) 将步骤(2)所得的三维结构预测结果与该融合蛋白相对应的蛋白酶采用Cluspro算法进行分子对接,选择对接结果正确比例超过阈值的酶切体系进行实验验证,筛选出蛋白酶最适配的融合蛋白。

2. 根据权利要求1所述的方法,其特征在于,还包括以下步骤:

(4) 对步骤(3)所得的最适配的融合蛋白编码基因进行密码子优化。

3. 根据权利要求1所述的方法,其特征在于,所述目标蛋白为人胰岛素样生长因子-I。

4. 根据权利要求3所述的方法,其特征在于,所述蛋白酶为肠激酶、凝血酶或HRV 3C蛋白酶,不同蛋白酶与其相对应的融合蛋白三维结构预测结果进行对接时,对接位点的选择如下:

肠激酶对应的融合蛋白选择氨基酸序列中所有赖氨酸(Lys)及其左右各4个氨基酸作为候选酶切位点;

凝血酶对应的融合蛋白选择精氨酸(Arg)及其左右各4个氨基酸作为候选酶切位点;

HRV 3C蛋白酶对应的融合蛋白选择谷氨酰胺(Gln)及其左右各4个氨基酸作为候选酶切位点。

5. 根据权利要求3所述的方法,其特征在于,步骤(3)中所述的阈值为50%。

6. 一种重组人胰岛素样生长因子-I的制备方法,其特征在于,使用氨基酸序列如SEQ ID NO.1所示的融合蛋白,使用pET-32a (+)载体在原核宿主中进行表达,表达所得的融合蛋白使用凝血酶进行酶切。

7. 根据权利要求6所述的制备方法,其特征在于,SEQ ID NO.1所示融合蛋白的编码基因为:

(1) 如SEQ ID NO.4所示的基因序列;或

(2) 与SEQ ID NO.4所示的基因序列同源性在90%至100%之间编码相同功能蛋白质的基因序列;或

(3) SEQ ID NO.4所示的基因序列经增加、缺失或替换一个或多个密码子,编码具有同等活性的蛋白质,由(1)衍生的基因序列。

8. 一种重组人胰岛素样生长因子-I的制备方法,其特征在于,使用氨基酸序列如SEQ ID NO.3所示的融合蛋白,使用pET-48b (+)载体在原核宿主中进行表达,表达所得的融合蛋白使用HRV 3C酶进行酶切。

9. 根据权利要求8所述的制备方法,其特征在于,SEQ ID NO.3所示融合蛋白对应的编码基因为:

(1) 如SEQ ID NO.5所示的基因序列;

(2) 与SEQ ID NO.5所示的基因序列同源性在90%至100%之间编码相同功能蛋白质的基因序列;或

(3) SEQ ID NO.5所示的基因序列经增加、缺失或替换一个或多个密码子,编码具有同等活性的蛋白质,由(1)衍生的基因序列。

10.根据权利要求6-9任一所述的制备方法,其特征在于,所述原核宿主为BL21 (DE3) 大肠杆菌菌株、Rosetta-gami B (DE3) 大肠杆菌菌株、Origami B (DE3) 大肠杆菌菌株或 Rosetta-gami 2 (DE3) 大肠杆菌菌株。

一种筛选酶切适配融合蛋白的方法及IGF-I制备方法

技术领域

[0001] 本发明涉及重组蛋白生产技术领域,具体涉及一种节省人力物力和时间的筛选酶切适配融合蛋白的方法,以及一种IGF-I的制备方法。

背景技术

[0002] 胰岛素样生长因子-I (IGF-I) 在细胞的生长、发育和增殖中具有重要作用,是一种多功能调控因子。目前,人IGF-I在临床上已用于治疗糖尿病、胰岛素抵抗综合征、侏儒症及神经系统疾病等多种疾病,并取得了良好的效果。若IGF-I在原核表达体系中直接表达,目标蛋白通常会形成包涵体,为了提高蛋白溶解度以及便于纯化,需要加入促溶标签和纯化标签进行融合表达;为得到目标蛋白,通常在促溶标签和纯化标签与目标蛋白之间插入一段蛋白酶酶切位点序列,以便于使用相应蛋白酶酶切除去标签,获得最终的目标蛋白。

[0003] 常用的蛋白酶有胰激酶、凝血酶、Xa因子、HRV 3C蛋白酶(人鼻病毒3C蛋白酶)等,每一种蛋白酶对其切割位点前后的氨基酸序列都有一定的限制(例如:凝血酶的最佳酶切位点序列为LVPR↓GS,↓标示该蛋白酶在此处切割开相应多肽链,下同),而蛋白药物不允许在成熟肽之前存在多余氨基酸,这就需要选出与融合蛋白相适应的蛋白酶,以保证切割后得到的目标蛋白具有天然未附加氨基酸残基的N-末端和/或C-末端。在实际生产当中,通常需要表达并纯化出多种含不同蛋白酶酶切位点序列的融合蛋白,比较不同蛋白酶的酶切效率才能选定最佳的表达系统;并且,这个过程中还会出现以下情况:(1)由于酶切位点包裹在融合蛋白内部,蛋白酶无法切割融合蛋白;(2)融合蛋白本身含有蛋白酶的多个酶切位点,蛋白酶会非特异性切割融合蛋白,甚至切割目标蛋白。无论上述哪一种情况,都会浪费人力物力成本,延长研发周期,甚至导致项目终止。

[0004] 所以,迫切需要一种快速高效的方法筛选出可以成功酶切的可溶表达体系,并顺利制备出重组人胰岛素样生长因子-I,以便进行大规模生产。

发明内容

[0005] 针对现有技术的以上缺陷或改进需求,本发明的一个目的是提供了一种筛选酶切适配融合蛋白的方法,采用虚拟筛选技术对蛋白酶酶切融合蛋白进行高效筛选,解决了以往用大量实验才能验证酶切效果的技术风险。

[0006] 本发明的另一个目的是提供了一种胰岛素样生长因子-I的制备方法,通过上述虚拟筛选方法获得了酶切效果好的胰岛素样生长因子-I融合蛋白,并实现了融合蛋白的可溶表达和成功酶切,解决了重组人胰岛素样生长因子-I难于在大肠杆菌表达系统中稳定、可溶及高活性表达的技术难题。

[0007] 下面详述本发明的技术方案:

[0008] 一种筛选酶切适配融合蛋白的方法,包括以下步骤:

[0009] (1)将目标蛋白序列编码基因分别插入到多个不同的含蛋白酶酶切位点的表达载体中,获得多个不同融合蛋白的氨基酸序列;所得融合蛋白的氨基酸序列通过二级结构预

测算法PSIPRED进行二级结构预测；

[0010] (2) 将步骤(1)中的不同融合蛋白氨基酸序列及其对应的二级结构预测结果作为输入文件,通过结构预测算法I-TASSER预测三维结构；

[0011] (3) 将步骤(2)所得的三维结构预测结果与该融合蛋白相对应的蛋白酶采用Cluspro算法进行分子对接,选择对接结果正确比例超过阈值的酶切体系进行实验验证,筛选出蛋白酶最适配的融合蛋白。

[0012] 上述筛选方法中,可以将候选蛋白酶均一一列出,分别得出与该蛋白酶对应的含目标蛋白及酶切位点的融合蛋白序列,根据蛋白酶数量,获得相应数量的融合蛋白。后续的二级结构预测算法PSIPRED、三级结构预测算法I-TASSER、分子对接算法Cluspro均为现有技术,本领域技术人员可以根据各算法的指导说明进行操作。

[0013] 优选的,上述方法中,还包括以下步骤:

[0014] (4) 对步骤(3)所得的最适配的融合蛋白编码基因进行密码子优化。经过优化以后能够提高表达量。

[0015] 优选的,上述方法中,所述目标蛋白为人胰岛素样生长因子-I。

[0016] 优选的,上述方法中,所述蛋白酶为肠激酶、凝血酶或HRV 3C蛋白酶,不同蛋白酶与其相对应的融合蛋白三维结构预测结果进行对接时,对接位点的选择如下:

[0017] 肠激酶对应的融合蛋白选择氨基酸序列中所有赖氨酸(Lys)及其左右各4个氨基酸作为候选酶切位点;

[0018] 凝血酶对应的融合蛋白选择精氨酸(Arg)及其左右各4个氨基酸作为候选酶切位点;

[0019] HRV 3C蛋白酶对应的融合蛋白选择谷氨酰胺(Gln)及其左右各4个氨基酸作为候选酶切位点。

[0020] 优选的,上述方法中,步骤(3)中所述的阈值为50%。

[0021] 一种重组人胰岛素样生长因子-I的制备方法,使用氨基酸序列如SEQ ID NO.1所示的融合蛋白,使用pET-32a(+)载体在原核宿主中进行表达,表达所得的融合蛋白使用凝血酶进行酶切。融合蛋白编码基因克隆入pET-32a(+)载体的MscI和EcoRV酶切位点之间。

[0022] 优选的,上述IGF-I制备方法中,SEQ ID NO.1所示融合蛋白的编码基因为:

[0023] (1) 如SEQ ID NO.4所示的基因序列;或

[0024] (2) 与SEQ ID NO.4所示的基因序列同源性在90%至100%之间编码相同功能蛋白质的基因序列;或

[0025] (3) SEQ ID NO.4所示的基因序列经增加、缺失或替换一个或多个密码子,编码具有同等活性的蛋白质,由(1)衍生的基因序列。

[0026] 一种重组人胰岛素样生长因子-I的制备方法,使用氨基酸序列如SEQ ID NO.3所示的融合蛋白,使用pET-48b(+)载体在原核宿主中进行表达,表达所得的融合蛋白使用HRV 3C酶进行酶切。融合蛋白编码基因克隆入pET-48b(+)载体的SacII和HindIII酶切位点之间。

[0027] 优选的,上述IGF-I制备方法中,SEQ ID NO.3所示融合蛋白对应的编码基因为:

[0028] (1) 如SEQ ID NO.5所示的基因序列;

[0029] (2) 与SEQ ID NO.5所示的基因序列同源性在90%至100%之间编码相同功能蛋白

质的基因序列;或

[0030] (3) SEQ ID NO.5所示的基因序列经增加、缺失或替换一个或多个密码子,编码具有同等活性的蛋白质,由(1)衍生的基因序列。

[0031] 优选的,以上任一所述的制备方法中,所述原核宿主为BL21 (DE3) 大肠杆菌菌株、Rosetta-gami B (DE3) 大肠杆菌菌株、Origami B (DE3) 大肠杆菌菌株或Rosetta-gami 2 (DE3) 大肠杆菌菌株。

[0032] 与现有技术相比,本发明具有如下有益效果:

[0033] 本发明提供的筛选酶切适配融合蛋白的方法,采用虚拟筛选手段,通过结构模拟和分子对接,可以虚拟比较各种蛋白酶的酶切效果,在实验之前就筛选出合适的融合蛋白表达体系和蛋白酶,并且经过实验验证后,与预测结果相符合,可以大大降低资金、人力和时间成本,相比于以往只能通过试验一个一个验证筛选的高投入方式更加方便高效。

[0034] 本发明通过上述筛选方法,虚拟筛选获得凝血酶、肠激酶、HRV 3C蛋白酶酶切效果,在实验验证后与预测结果相符,凝血酶、HRV 3C蛋白酶及其对应的融合蛋白更适合用于制备IGF-I,进而获得了适用于重组人胰岛素样生长因子-I的可以成功酶切的可溶表达体系。

[0035] 本发明提供的重组人胰岛素样生长因子-I融合蛋白的制备方法,具有可溶性表达,表达量大等优点,并可以使用蛋白酶酶切成功获得重组人胰岛素样生长因子-I,其解决了人胰岛素样生长因子-I不能直接获得的技术问题。

附图说明

[0036] 图1为PSIPRED在线预测服务器输入界面;

[0037] 图2为I-TASSER在线预测服务器输入界面;

[0038] 图3为Cluspro在线预测服务器输入界面;

[0039] 图4为大肠杆菌表达载体pET-32a (+) 图谱;

[0040] 图5为大肠杆菌表达载体pET-32a (+) 的多克隆位点区域图谱;

[0041] 图6为凝血酶融合蛋白的SDS-PAGE电泳分析图;

[0042] 图7为凝血酶融合蛋白的Western Blot免疫鉴别图谱;

[0043] 图8为大肠杆菌表达载体pET-48b (+) 图谱;

[0044] 图9为大肠杆菌表达载体pET-48b (+) 的多克隆位点区域图谱;

[0045] 图10为HRV 3C酶融合蛋白的Western Blot免疫鉴别图谱;

[0046] 图11为凝血酶融合蛋白经凝血酶酶切后的SDS-PAGE分析图;

[0047] 图12为HRV 3C酶融合蛋白经HRV 3C酶酶切后的WesternBlot免疫鉴别图谱;

[0048] 图13为肠激酶融合蛋白的Western Blot免疫鉴别图谱;

[0049] 图14为肠激酶融合蛋白经肠激酶酶切后的SDS-PAGE电泳分析图。

具体实施方式

[0050] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明的技术方案进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。此外,下面所描述的本发明各个实施方式中所涉及到的技

术特征只要彼此之间未构成冲突就可以相互组合。

[0051] 发明人前期通过对蛋白质结构以及分子模拟方法的深入研究,成功实现了用多种结构预测算法提高蛋白质的热稳定性(Li,et al.Appl Environ Microb,2018,84(2),e02129-17;Li,et al.RSCAdv,2018,8,1948)。本发明通过结构预测算法和分子对接算法,虚拟筛选出适合制备重组人胰岛素样生长因子-I的融合蛋白,并采用相应的蛋白酶成功酶切制备得到重组人胰岛素样生长因子-I。通过本发明的技术方案,能够实现IGF-I在大肠杆菌表达系统中稳定、可溶及高活性的表达,使其产业化生产成为可能,有利于新药开发和临床应用。

[0052] 关于实施例,应该指出,以下具体说明都是例示性的,旨在对本发明提供进一步的说明。除非特别指明,否则所用技术术语为本领域中的普通技术人员常用术语;未注明具体条件的实验方法是按常规实验方法;所用的试验材料如无特别说明均为市售购买产品,各种试剂及培养基的成分和配制方法可参见常规实验手册中的操作。

[0053] 实施例1:融合蛋白氨基酸序列的获得

[0054] 经文献报道,胰岛素样生长因子-I(IGF-I)在大肠杆菌中直接表达会形成包涵体(Rosano GL,et al.Front Microbiol,2014,5:172),而添加促溶蛋白标签进行融合表达可以实现蛋白的可溶表达。目前,使用最为广泛的大肠杆菌表达载体为pET系统,其载体添加了Thioredoxin(硫氧还蛋白)等促溶标签以及His-tag纯化标签,可以促进融合蛋白的可溶表达及纯化。

[0055] 为了实现蛋白酶酶切融合蛋白后IGF-I的N端无其他氨基酸,候选蛋白酶为凝血酶、FactorXa、肠激酶、TEV酶、HRV 3C酶;考虑到用酶成本和获取难度,采用凝血酶、肠激酶、HRV 3C酶进行模拟酶切实验。在不同的pET表达载体中,含有凝血酶、肠激酶和HRV 3C酶的酶切位点,可以通过蛋白酶酶切获得目标蛋白。最终,确定了pET-32a(+)-凝血酶酶切位点-IGF-I、pET-32a(+)-肠激酶酶切位点-IGF-I、pET-48b(+)-HRV 3C酶酶切位点-IGF-I三种融合蛋白进行后续实验。

[0056] IGF-I的氨基酸序列如SEQ ID No.6所示:

[0057] GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSRRAPQTGIVDECCFRSCDLRRL~~EMYCAPL~~KPAKSA。

[0058] 将IGF-I的氨基酸序列分别插入至pET-32a(+)载体的凝血酶酶切位点序列LVPR,肠激酶酶切位点DDDDK序列,以及pET-48b(+)载体的HRV 3C酶酶切位点LEVL~~FQ~~序列之后,得到相应的融合蛋白氨基酸序列,即

[0059] 凝血酶融合蛋白氨基酸序列SEQ ID No.1:

[0060] MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNP~~GTA~~PKYGI~~RGIPT~~LLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGMH~~HHHHH~~HSSGLVPRGPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSRRAPQTGIVDECCFRSCDLRRL~~EMYCAPL~~KPAKSA。

[0061] 肠激酶融合蛋白氨基酸序列SEQ ID No.2:

[0062] MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNP~~GTA~~PKYGI~~RGIPT~~LLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGMH~~HHHHH~~HSSGLVPRGSGMKETA~~AAK~~FERQHMDSPDLGTDDDDKGPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSRRAPQTGIVDECCFRSCDLRRL~~EMYCAPL~~KPAKSA。

[0078] 凝血酶融合蛋白选择Arg及其左右各4个氨基酸作为候选酶切位点,肠激酶融合蛋白选择氨基酸序列中所有Lys及其左右各4个氨基酸作为候选酶切位点,HRV 3C酶融合蛋白选择Gln及其左右各4个氨基酸作为候选酶切位点;优选地,凝血酶融合蛋白进行对接的区域为SEQ ID No.1序列中第49-57,66-78,93-101,125-133,146-154位;肠激酶融合蛋白进行对接的区域为SEQ ID No.2序列中第1-8,15-23,33-41,49-62,66-75,79-105,130-144,154-162,181-189,219-228位;HRV 3C酶融合蛋白进行对接的区域为SEQ ID No.3序列中第47-55,59-67,95-103,147-155,162-170,187-195位。

[0079] 凝血酶(PDB ID:1ETR)的氨基酸序列如SEQ ID No.7所示:

[0080] TFGAGEADCGLRPLFEKKQVQDQTEKELFESYIEGRIVEGQDAEVGLSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPWDKNFTVDDLLVRIGKHSRTRYERKVEKISMLDKIYIHPRYNWKENLDRDIALLLKLRPIELSDYIHPVCLPDKQTAAKLLHAGFKGRVTGWGNRRETWTTSVAEVQPSVLQVNNLPLVERPVCKASTRIRITDNMFCAGYKPGEGKRGDACEGDSGGPFVMKSPYNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLLKQKVIDRLGS

[0081] 其对接区域为第51-53,55-56,58-65,79-80,83,85-86,88,90,101,104-105,131-132,135,175-179,182-190,192,194,214-215,224-225,234-241,261-267,269,273-276位;

[0082] 肠激酶(PDB ID:1EKB)的氨基酸序列如SEQ ID No.8所示:

[0083] CGKKLVTEVSPKIVGGSDSREGAWPVVVALYFDDQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGLHMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIPICLPEENQVFPPGRICSIAGWGLIYQGSTADVLQEADVPLLSNEKCCQQMPEYNITENMVCAGYEAGGVDSQCQDGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWISFLH

[0084] 其对接区域为第33,36-39,54-55,57-59,97,99-100,102,105,195-200,218-221,231-232位;

[0085] HRV 3C酶(PDB ID:2IN2)的氨基酸序列如SEQ ID No.9所示:

[0086] GPNTEFALSLLRKNIMTITTSKGFTGLGIHDRVCVIPHTAQPGDDVLVNGQKIRVKDKYKLVDPENINLELTVLTLDRNEKFRDIRGFISEDLEGVDATLVVHSSNFTNTILEVGPVTMAGLINLSSTPTNRMIRYDYATKTGQCGGVLCATGKIFGIHVGGNGRQGFSAQLKKQYFVEKQ

[0087] 其对接区域为第26-30,44-48,64-68,74,76-78,111-112,131-137,139,147-152,166-170,174位。

[0088] 输入上述对接文件和对接位点后,点击“Dock”按钮进行对接预测,Cluspro会输出20个最可能的对接复合物结构集合及其pdb结构文件,并给出各集合含有的结构数量。观察这20个集合的复合物结构文件,得出各复合物的对接位置。

[0089] 凝血酶与融合蛋白对接输出结果如下:

[0090] 1)最大的结构集合共包含220个复合物结构,酶切位置为PR↓GP;2) IR↓GI (161个);3) 无明确对接位置;4) PR↓GP (64个);5) 无明确对接位置;6) IR↓GI (49个);7) PR↓GP (46个);8) PR↓GP (43个);9) PR↓GP (29个);10) DR↓GF (26个);11) 无明确对接位置;12) 无明确对接位置;13) 无明确对接位置;14) DR↓GF (15个);15) IR↓GI (14个);16) PR↓GP (12个);17) PR↓GP (10个);18) 无明确对接位置;19) IR↓GI (4个);20) DR↓GF (1个)。

[0091] 将各个对接位置及其数量进行统计,可得出凝血酶活性位点在正确酶切位置(PR↓GP)进行对接所占百分比为61.1%:

[0092] MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPQGTAPKYGIR (32.8%) GIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHSSGLVPR (61.1%) GPETLCGAELVDALQFVCGDR (6.1%) GFYFNKPTGYGSSRRAPQGTIVDECCFRSCDLRRLEMYCAPLKPAKSA

[0093] 同理, 以下是肠激酶各酶切位置及其百分比, 可知正确酶切位置(DK↓GP)的比例为14.7%:

[0094] MSDK (17.1%) I IHLTDDSFDTDVLKADGAILVDFWAEWCGPCK (1.8%) MIAPILDEIADEYQGLTVAKLNIDQNPQGTAPK (30.1%) YGIRGIPTLLLFKNGEVAATK (10.4%) VGALSK (2.3%) GQLKEFLDANLAGSGSGHMHSSGLVPRGSGMK (14.4%) ETAAAK (3.0%) FERQHMDSPDLGTDDDDK (14.7%) GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSRRAPQGTIVDECCFRSCDLRRLEMYCAPLK (6.1%) PAKSA

[0095] 以下是HRV 3C酶各酶切位置及其百分比, 可知正确酶切位置(FQ↓GP)的比例为56.6%:

[0096] MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPQGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHTSGGGSNNNPPTPTPSSGSGHHHHHSAALEVLFQ (56.6%) GPETLCGAELVDALQ (43.4%) FVCGDRGFYFNKPTGYGSSRRAPQGTIVDECCFRSCDLRRLEMYCAPLKPAKSA

[0097] 我们认为, 预测结果正确酶切位置的比例超过50%意味着蛋白酶可以顺利与底物进行接触, 并特异性地针对正确的酶切位点进行切割, 所以将50%作为阈值进行筛选。随后, 选择凝血酶酶切体系和HRV 3C酶酶切体系作为实施例进行表达实验, 而肠激酶酶切体系作为对比例进行表达实验。

[0098] 实施例5融合蛋白的获得及酶切

[0099] 5.1融合蛋白基因序列的获得

[0100] 为了实现各融合蛋白在大肠杆菌中大量、高效表达, 发明人参照大肠杆菌密码子偏好性, 并考虑密码子简并性、GC含量、合适的限制性内切酶等因素, 最终设计出合适的基因序列用于表达。

[0101] 具体地, 凝血酶融合蛋白基因序列如SEQ ID No.4所示, 其5'端带有MscI限制性内切酶位点(TGGCCA), 3'端带有连续的终止密码子(TAATGA), 以及EcoRV的识别位点(GATATC):

[0102] TGGCCATATGCACCATCATCATCATCTTCTTCTGGTCTGGTGCCACGCGGTCCGGAGACCCTGTGCGGTGCGGAACTGGTGGACGCGCTGCAATTTGTTTTCGGTGATCGTGGCTTCTACTTTAACAAGCCGACCGGTTATGGTAGCAGCAGCCGTCGTGCGCCGAGACCGGTATCGTTGACGAGTGCTGCTTCCGTAGCTGCGATCTGCGTCTGTTGAAATGTATTGCGCGCCGCTGAAGCCGGCGAAAAGCGCGTAATGAGATATC

[0103] HRV 3C酶融合蛋白基因序列如SEQ ID No.5所示, 其5'端带有SacI限制性内切酶位点(CCGCGG), 3'端带有终止密码子(TAA), 以及HindIII的识别位点(AAGCTT):

[0104] CCGCGGCTCTGGAAGTGCTGTTTCAAGGTCCGGAGACCCTGTGCGGTGCGGAACTGGTGGACGCGCTGCAATTTGTTTTCGGTGATCGTGGCTTCTACTTTAACAAGCCGACCGGTTATGGTAGCAGCAGCCGTCGTGCGCCGAGACCGGTATCGTTGACGAGTGCTGCTTCCGTAGCTGCGATCTGCGTCTGTTGAAATGTATTGCGCGCCGCTGAAGCCGGCGAAAAGCGCGTAAAAGCTT

[0105] 通过化学合成方式获取以上序列。

[0106] 5.2构建凝血酶融合蛋白的重组表达载体及融合表达工程菌

[0107] pET-32a (+) 表达载体图谱及其多克隆位点序列如图4和图5所示。将人工合成的凝血酶融合蛋白的cDNA序列 (SEQ ID No.4), 用MscI和EcoRV双酶切后, 通过T4 DNA连接酶连接到同样经双酶切的pET-32a (+) 表达载体上, 构建重组表达载体, 记作pET-32a-Thrombin-IGF-I; 转化大肠杆菌克隆菌株Top10, 在含有50 μ g/ml氨苄青霉素的LB培养基上37 $^{\circ}$ C培养筛选重组子, 经PCR与酶切鉴定正确后, 测序验证其序列正确性。

[0108] 提取鉴定正确的重组子, 转化到大肠杆菌表达菌株Origami B (DE3) 中, 在含有50 μ g/ml氨苄青霉素 (Amp)、15 μ g/ml卡那霉素 (Kan) 以及12.5 μ g/ml四环素 (Tet) 的LB培养基上, 37 $^{\circ}$ C培养筛选重组子, 获得的重组子即为Trx融合表达的IGF-I工程菌, 命名为Thrombin-IGF。将其保存于15%甘油中, 冻存于-80 $^{\circ}$ C冰箱。其中, Thrombin-IGF工程菌表达产生的融合蛋白应含有硫氧还蛋白 (Trx) 标签、His-tag纯化标签、凝血酶 (Thrombin) 酶切位点以及成熟肽IGF-I序列, 理论分子量约为21.6kDa。

[0109] 取甘油管冻存的Thrombin-IGF工程菌, 按0.2%接种于含有新鲜LB抗性培养基 (含50 μ g/ml Amp, 15 μ g/ml Kan以及12.5 μ g/ml Tet) 的试管中, 振荡培养过夜, 活化菌株; 再转接至300mL的LB培养基中, 37 $^{\circ}$ C振荡培养约3h, 测其OD₆₀₀值在0.6~0.8时, 加入终浓度为0.05mM异丙基- β -D-硫代半乳糖苷 (IPTG), 于25 $^{\circ}$ C诱导表达24h, 离心收集菌体; 称量菌体湿重, 按菌体和破菌液 (50mM Tris-HCl, 0.5M NaCl, pH8.0) 重量比为1:15的比例加入破菌液, 超声破菌后, 进行SDS-PAGE检测。结果如图6所示, 其中M为蛋白Marker; 1为诱导前全菌液; 2为诱导后全菌液; 3为诱导后破菌上清。结果说明破菌上清中有分子量接近26kDa的融合蛋白表达, 与理论分子量接近。

[0110] 将上述融合蛋白采用IGF-I单抗 (Abcam公司产品, 货号: ab9572, 下同) 进行蛋白质免疫印迹 (Western Blot) 鉴定, 结果如图7所示, 其中M为蛋白Marker; 1为诱导前全菌液; 2为诱导后全菌液; 3为诱导后的破菌上清。结果表明破菌上清中的融合蛋白具有IGF-I的免疫活性, 并且通过图中可以看到, 融合蛋白还会形成少量二聚体结构。

[0111] 5.3构建HRV 3C酶融合蛋白的重组表达载体及融合表达工程菌

[0112] pET-48b (+) 表达载体图谱及其多克隆位点序列如图8和图9所示。将人工合成的HRV 3C酶融合蛋白的cDNA序列 (SEQ ID No.5), 用SacI和HindIII双酶切后, 通过T4 DNA连接酶连接到同样经双酶切的pET-48b (+) 表达载体上, 构建重组表达载体, 记作pET-48b-HRV 3C-IGF-I; 转化大肠杆菌克隆菌株Top10, 在含有50 μ g/ml卡那霉素的LB培养基上37 $^{\circ}$ C培养筛选重组子, 经PCR与酶切鉴定正确后, 测序验证其序列正确性。

[0113] 提取鉴定正确的重组子, 转化到大肠杆菌表达菌株Rosetta-gami2 (DE3) 中, 在含有50 μ g/ml Kan, 34 μ g/ml Chl, 50 μ g/ml链霉素 (Str) 以及12.5 μ g/ml Tet的LB培养基上, 37 $^{\circ}$ C培养筛选重组子, 获得的重组子即为Trx融合表达的IGF-I工程菌, 命名为HRV 3C-IGF。将其保存于15%甘油中, 冻存于-80 $^{\circ}$ C冰箱。其中, HRV 3C-IGF工程菌表达产生的融合蛋白应含有硫氧还蛋白 (Trx) 标签、His-tag纯化标签、HRV 3C酶酶切位点以及成熟肽IGF-I序列, 理论分子量约为23.5kDa。

[0114] 取甘油管冻存的HRV 3C-IGF工程菌, 按0.2%接种于含有新鲜LB抗性培养基 (含50 μ g/ml Kan, 50 μ g/ml Str, 34 μ g/ml Chl以及12.5 μ g/ml Tet) 的试管中, 振荡培养过夜, 活化菌株; 再转接至300mL的LB培养基中, 37 $^{\circ}$ C振荡培养约3h, 测其OD₆₀₀值在0.6~0.8时, 加入终浓度为0.5mM异丙基- β -D-硫代半乳糖苷 (IPTG), 于25 $^{\circ}$ C诱导表达25h, 离心收集菌体; 称量

菌体湿重,按菌体和破菌液(50mM Tris-HCl,0.5MNaCl,pH8.0)重量比为1:15的比例加入破菌液,超声破菌后,将诱导前后的菌液以及破菌上清采用IGF-I单抗进行Western Blot检测,结果如图10所示,其中M为蛋白Marker;1为诱导后全菌液;2为诱导后的破菌上清。结果表明破菌上清中的融合蛋白具有IGF-I免疫活性,蛋白大小在26kDa附近,与理论分子量接近。

[0115] 5.4融合蛋白的纯化

[0116] 凝血酶融合蛋白和HRV 3C酶融合蛋白的表达是采用实施例5.2和5.3中构建的Thrombin-IGF工程菌和HRV 3C-IGF工程菌来完成。发酵液离心收集菌体,超声破菌,收集上清,使用Ni-Sepharose 6FF柱进行亲和层析,使用20mM Tris-HCl (pH7.8)-0.5MNaCl进行平衡,使用50mM→500mM咪唑梯度洗脱收集目的蛋白。为了减少高浓度盐溶液对蛋白酶活性的影响,使用G-15脱盐柱除去多余盐分。

[0117] 5.5融合蛋白的酶切

[0118] IGF-I成熟肽的大小约为10kDa,若正确酶切,则可以在10kDa附近检测到条带,并具有IGF-I的免疫活性。

[0119] 凝血酶切割融合蛋白的体系采用1U凝血酶(Solarbio公司产品,货号:T8021)切割100 μ g融合蛋白,酶切缓冲体系终浓度为20mM Tris-HCl (pH 8.0),0.15M NaCl,0/2.5/10/20mM CaCl₂,使用37 $^{\circ}$ C摇床酶切24h。酶切结果如图11所示,其中M为蛋白Marker,1为无CaCl₂,2为2.5mM CaCl₂,3为10mM CaCl₂,4为20mM CaCl₂。如图所示,CaCl₂终浓度为10mM时,酶切效果最好,凝血酶几乎可以完全切割融合蛋白,而CaCl₂浓度再提高后会降低酶活力,降低切割效果。

[0120] HRV 3C酶切割融合蛋白的体系采用1U/10U HRV 3C酶(Takara公司产品,货号:7360)切割100 μ g融合蛋白,酶切缓冲体系终浓度为50mMTris-HCl (pH7.5),0.15M NaCl,使用4 $^{\circ}$ C酶切24h。酶切产物的IGF-I单抗WB结果如图12所示,其中M为蛋白Marker,1为加入1U酶,2为加入10U酶。由图可知,加入1U酶能够切割融合蛋白,获得少量IGF-I成熟肽;而加入10U酶酶切效果较好,可以获得大量IGF-I成熟肽。

[0121] 对比例1肠激酶融合蛋白的获得及酶切

[0122] 1.1融合蛋白基因序列的获得

[0123] 肠激酶融合蛋白基因序列如SEQ ID No.10所示,其5'端带有KpnI限制性内切酶位点(GGTACC),3'端带有连续的终止密码子(TAATGA),以及HindIII的识别位点(AAGCTT):

[0124] GGTACCGACGACGACGACAAGGGTCCGAGACCCTGTGCGGTGCGGAAGTGGTGGACGCGCTGCAATTTGTTTGCAGTATCGTGGCTTCTACTTTAACAAGCCGACCGTTATGGTAGCAGCAGCCGTCGTGCGCCGACACCGGTATCGTTGACGAGTGCTGCTTCCGTAGCTGCGATCTGCGTCGTCTGGAAATGTATTGCGCGCCGCTGAAGCCGGCGAAAGCGCGTAATGAAAGCTT

[0125] 1.2构建肠激酶融合蛋白的重组表达载体及融合表达工程菌

[0126] 将人工合成的肠激酶融合蛋白的cDNA序列(SEQ ID No.10),用KpnI和HindIII双酶切后,通过T4 DNA连接酶连接到同样经双酶切的pET-32a (+)表达载体上,构建重组表达载体,记作pET-32a-Enterokinase-IGF-I;转化大肠杆菌克隆菌株Top10,在含有50 μ g/ml氨苄青霉素的LB培养基上37 $^{\circ}$ C培养筛选重组子,经PCR与酶切鉴定正确后,测序验证其序列正确性。

[0127] 提取鉴定正确的重组子,转化到大肠杆菌感受态表达菌株Origami B(DE3)中,在含有50 μ g/ml氨苄青霉素(Amp)、15 μ g/ml卡那霉素(Kan)以及12.5 μ g/ml四环素(Tet)的LB培养基上,37 $^{\circ}$ C培养筛选重组子,获得的重组子即为Trx融合表达的IGF-I工程菌,命名为Ek-IGF。将其保存于15%甘油中,冻存于-80 $^{\circ}$ C冰箱。其中,Ek-IGF工程菌表达产生的融合蛋白应含有硫氧还蛋白(Trx)标签、His-tag纯化标签、肠激酶(Enterokinase)酶切位点以及成熟肽IGF-I序列,理论分子量约为24.7kDa。

[0128] 取甘油管冻存的Ek-IGF工程菌,按0.2%接种于含有新鲜LB抗性培养基(含50 μ g/ml Amp,15 μ g/ml Kan以及12.5 μ g/ml Tet)的试管中,振荡培养过夜,活化菌株;再转接至300mL的LB培养基中,37 $^{\circ}$ C振荡培养约3h,测其OD₆₀₀值在0.6~0.8时,加入终浓度为0.5mM IPTG,于25 $^{\circ}$ C诱导表达24h,离心收集菌体;称量菌体湿重,按菌体和破菌液(50mM Tris-HCl,0.5M NaCl,pH8.0)重量比为1:15的比例加入破菌液,超声破菌后,将破菌上清和沉淀进行IGF-I单抗的Western Blot鉴定,结果如图13所示,其中M为蛋白Marker;1为破菌上清;2为破菌沉淀。结果表明破菌上清中的融合蛋白具有IGF-I免疫活性,蛋白大小在26-34kDa之间。

[0129] 1.3肠激酶融合蛋白的纯化

[0130] 肠激酶融合蛋白的表达是采用对比例1.1中构建的Ek-IGF工程菌来完成。发酵液离心收集菌体,超声破菌,收集上清,使用Ni-Sepharose 6FF柱进行亲和层析,使用20mM Tris-HCl(pH7.8)-0.5M NaCl进行平衡,使用50mM \rightarrow 500mM咪唑梯度洗脱收集目的蛋白。为了减少高浓度盐溶液对蛋白酶活性的影响,使用G-15脱盐柱除去多余盐分。

[0131] 1.4肠激酶融合蛋白的酶切

[0132] 肠激酶切割肠激酶融合蛋白的体系采用不同量的重组牛肠激酶(耀海生物公司产品,货号:ez00001)切割1mg融合蛋白,酶切缓冲体系终浓度为20mM Tris-HCl(pH 8.0),50mM NaCl,2mM CaCl₂,使用16 $^{\circ}$ C酶切24h。酶切产物的SDS-PAGE结果如图14所示,其中M为蛋白Marker,1为加入1IU酶,2为加入5IU酶,3为加入10IU酶,4为加入50IU酶,5为加入100IU酶,6为肠激酶切割阳性对照融合蛋白,7为未加酶的阳性对照融合蛋白。由图可知,肠激酶可以切割相应阳性对照蛋白,但无法切割肠激酶融合蛋白。

[0133] 以上所述实施例仅是为充分说明本发明而所举的较佳的实施例,本发明的保护范围不限于此。本技术领域的技术人员在本发明基础上所作的等同替代或变换,均在本发明的保护范围之内。本发明的保护范围以权利要求书为准。

序列表

<110> 武汉海特生物制药股份有限公司

<120> 一种筛选酶切适配融合蛋白的方法及IGF-I制备方法

<130> WH2008252-1

<160> 10

<170> SIPOSequenceListing 1.0

<210> 1

<211> 199

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 1

```

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

```

<210> 2

<211> 228

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 2

```

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

```

225

<210> 3

<211> 221

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 3

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

<210> 4

<211> 271

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 4

tggccatatg caccatcacc atcatcattc ttctggtctg gtgccacgcg gtccggagac 60
 cctgtgcggt gcggaactgg tggacgcgct gcaatttggt tgcggtgata gtggcttcta 120
 ctttaacaag ccgaccggtt atggtagcag cagccgtcgt gcgccgcaga ccggtatcgt 180
 tgacgagtgc tgcttccgta getgcgatct gcgtcgtctg gaaatgtatt gcgcgccgct 240
 gaagccggcg aaaagcgcgt aatgagatat c 271

<210> 5

<211> 245

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 5

ccgcggctct ggaagtgctg tttcaaggct cggagaccct gtgcggtgcg gaactggtgg 60
 acgcgctgca atttgtttgc ggtgatcgtg gcttctactt taacaagccg accggttatg 120
 gtagcagcag ccgtcgtgcg ccgcagaccg gtatcgttga cgagtgctgc ttccgtagct 180
 gcgatctgcg tcgtctggaa atgtattgcg cgccgctgaa gccggcgaaa agcgcgtaaa 240
 agctt 245

<210> 6

<211> 70

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 6

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

<210> 7

<211> 295

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 7

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

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

<210> 9

<211> 182

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 9

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

Popular Analyses

PSIPRED 4.0 (Predict Secondary Structure) DISOPRED3 (Disopred Prediction)

MEMSAT-SVM (Membrane Helix Prediction) pGenTHREADER (Profile Based Fold Recognition)

Contact Analysis

DeepMetaPSICOV 1.0 (Structural Contact Prediction) MEMPACK (TM Topology and Helix Packing)

Fold Recognition

GenTHREADER (Rapid Fold Recognition) pDomTHREADER (Protein Domain Fold Recognition)

Structure Modelling

Bioserf 2.0 (Automated Homology Modelling) Domserf 2.1 (Automated Domain Homology Modelling)

DMPfold 1.0 Fast Mode (Protein Structure Prediction)

Domain Prediction

DomPred (Protein Domain Prediction)

Function Prediction

FFPred 3 (Eukaryotic Function Prediction)

[Help...](#)

Submission details

Protein Sequence

Protein Sequence

[Help...](#)
If you wish to test these services follow this link to retrieve a test fasta sequence.

Job name

Job name

Email (optional)

Email (optional)

图1

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence below ([10, 1500] residues in FASTA format) [Click here for a sample input](#).

Or upload the sequence from your local computer:
 未选择任何文件

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click [here](#) if you do not have a password)

ID: (optional, your given name of the protein)

▶ **Option I:** Assign additional restraints & templates to guide I-TASSER modeling.

▶ **Option II:** Exclude some templates from I-TASSER template library.

▼ **Option III:** Specify secondary structure for specific residues.

- Specify secondary structure 未选择任何文件
- Upload a file listing secondary structure [Explanation](#)

Keep my results public (uncheck this box if you want to keep your job private, and a key will be assigned for you to access the results. We received numerous requests from users who loss their key to access result. To save your time, please keep results public, or ensure you remember the key if you choose to keep job private)

(Please submit a new job only after your old job is completed)

图2

Dock

Note: Due to increased server usage, please submit no more than 15 jobs at a time.

Job

Name:

Server:

Accepted PDB Input:
20 standard amino acids and RNA (as
receptor only), ref: [RNA](#) Select Heparin Mode
to use Heparin as Ligand.

Receptor	Ligand
PDB <input type="checkbox"/>	PDB <input type="checkbox"/>
ID: Upload PDB	ID: Upload PDB
Chains: <input type="text"/>	Chains: <input type="text"/>

Whitespace separate desired chains. Leave
chains blank to use all chains.

Advanced Options**Attraction and Repulsion**

Enter attraction and repulsion of residues as
whitespace separated "chain-residue" entries.
eg. a-23 a-25 a-26 a-27

Attraction:	<input type="text"/>	Attraction:	<input type="text"/>
-------------	----------------------	-------------	----------------------

Structure Modification**Multimer Docking
(Beta Version)****Antibody Mode****Others Mode****Heparin Ligand****Saxs Profile****Restraints**

图3

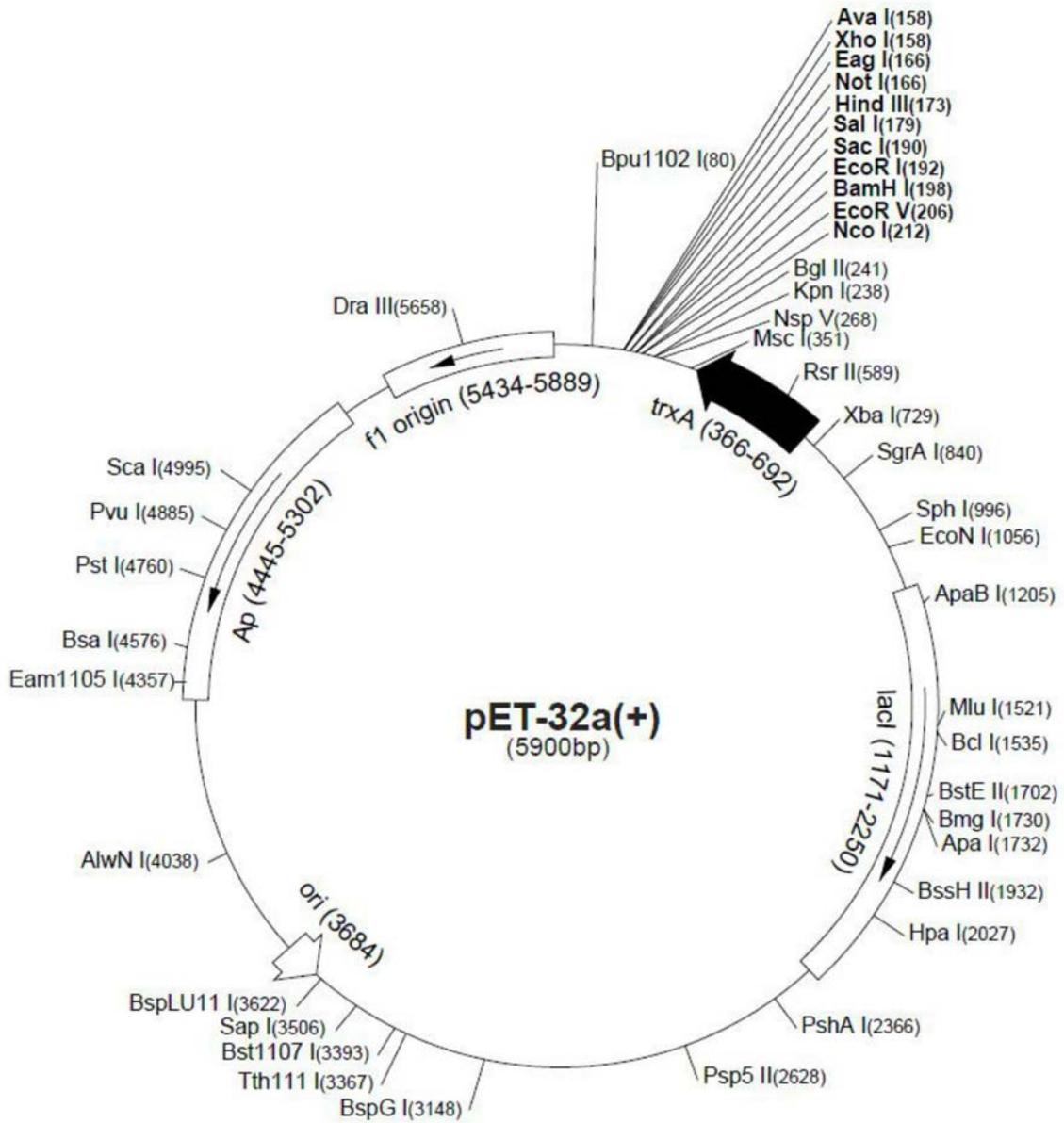


图4

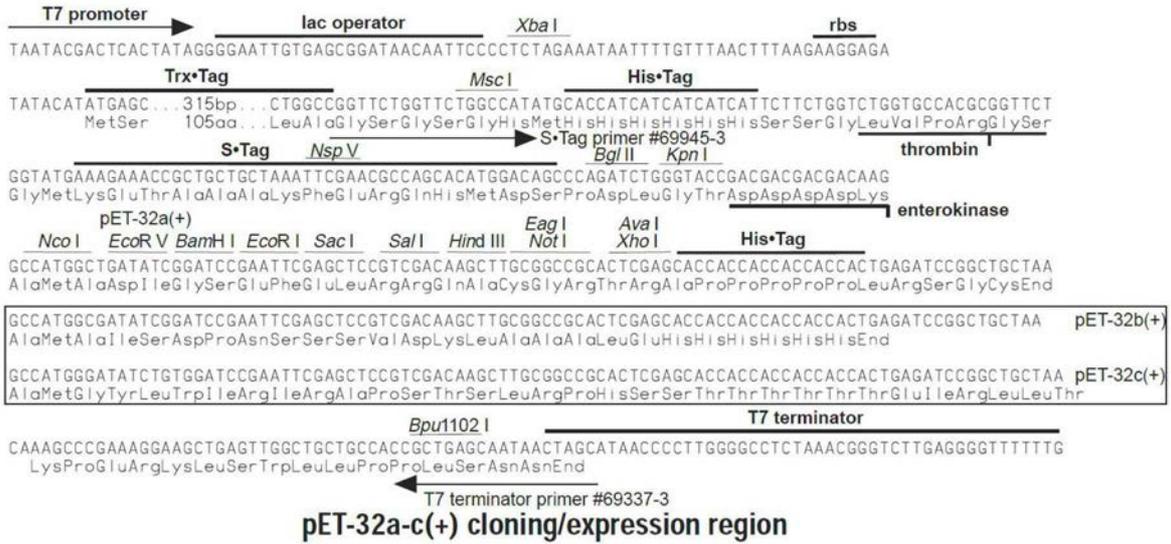


图5

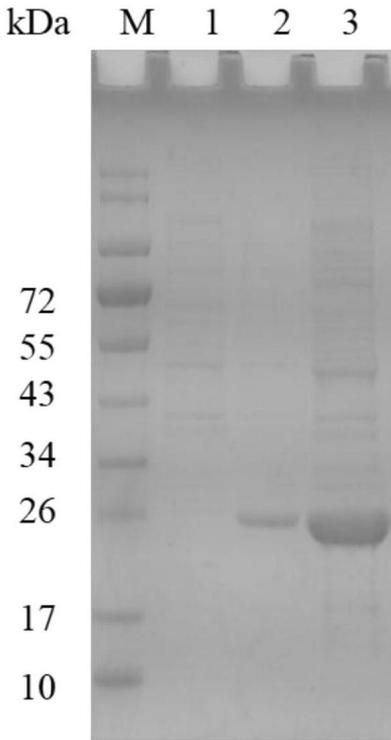


图6

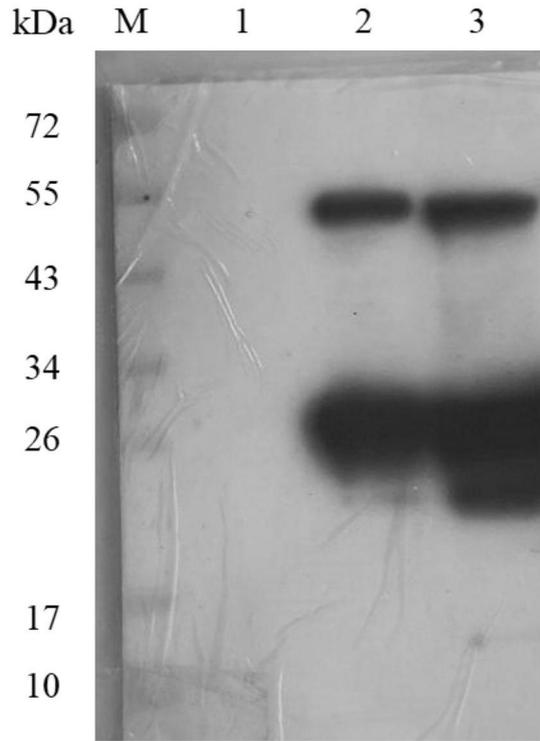


图7

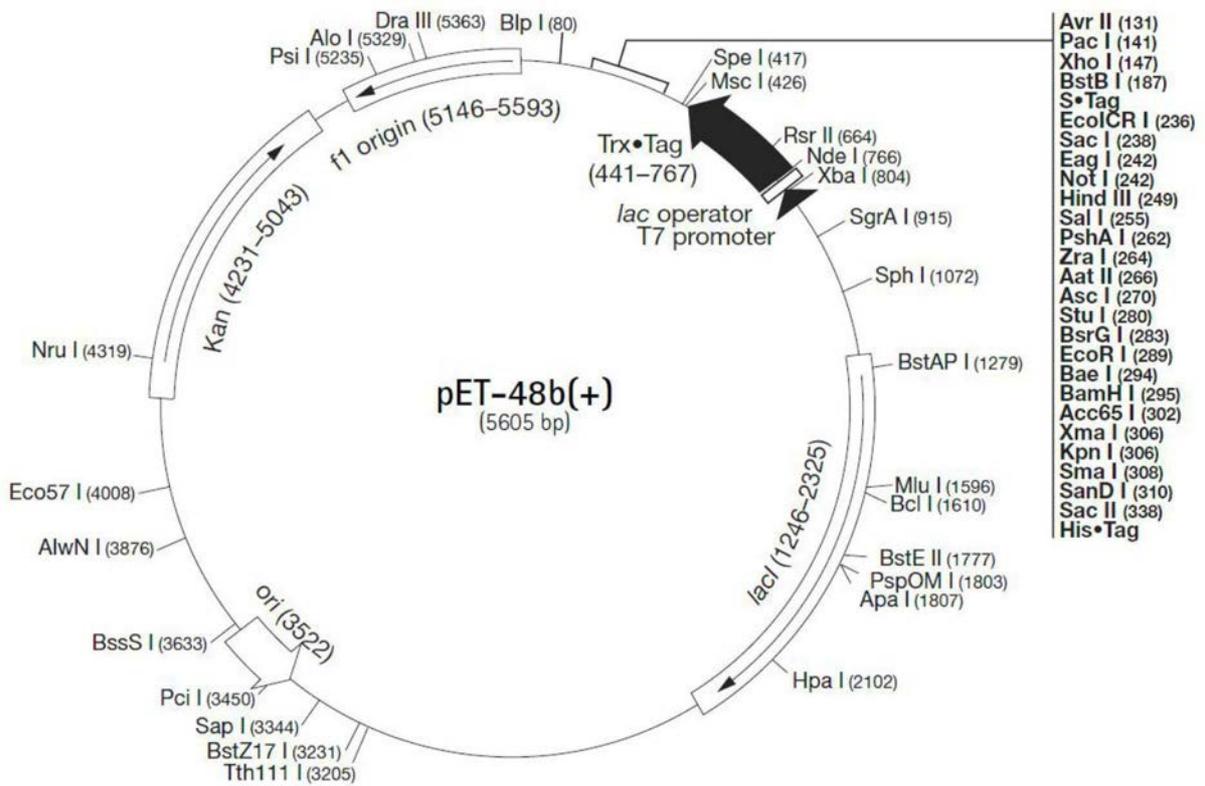
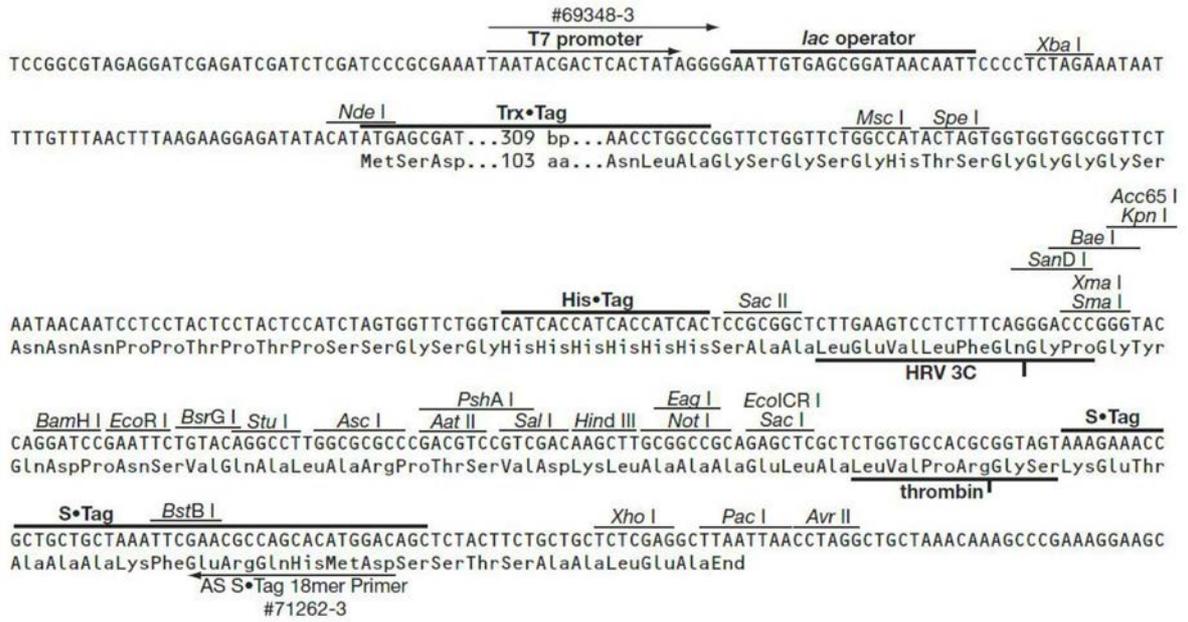


图8



pET-48b(+) cloning/expression region

图9

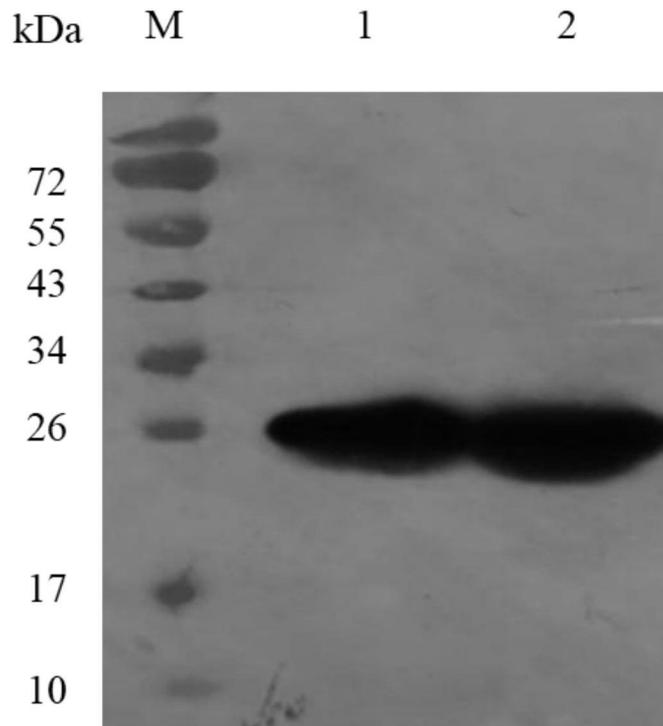


图10

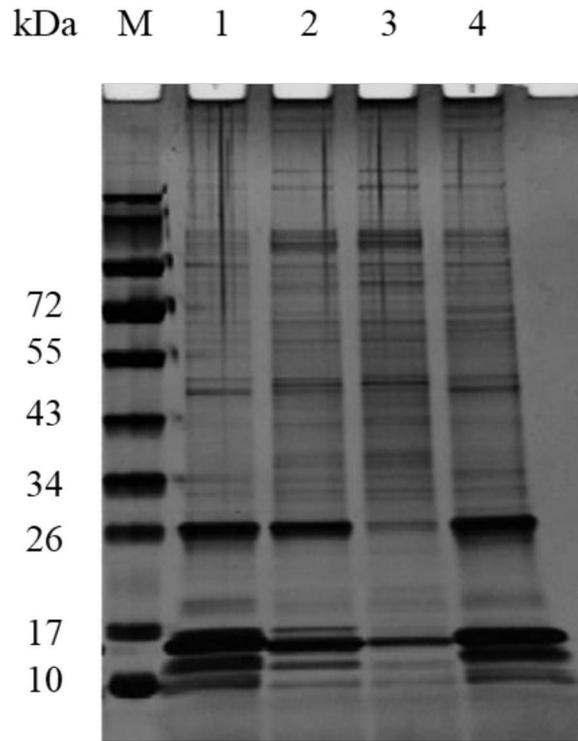


图11

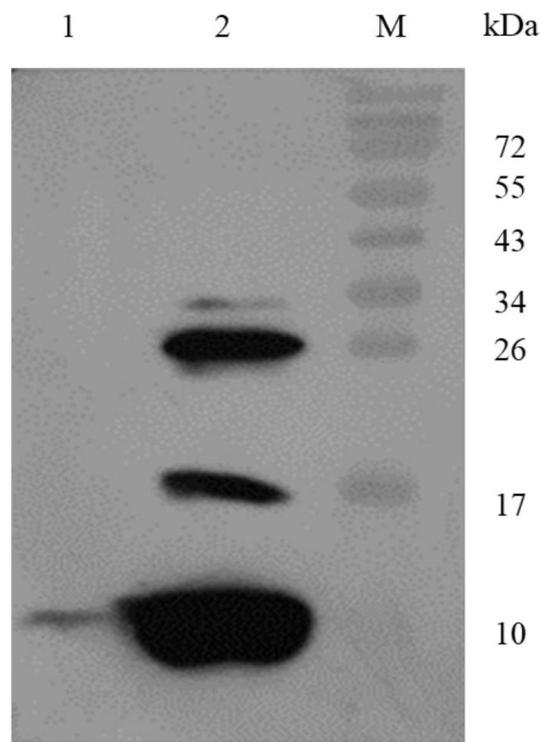


图12

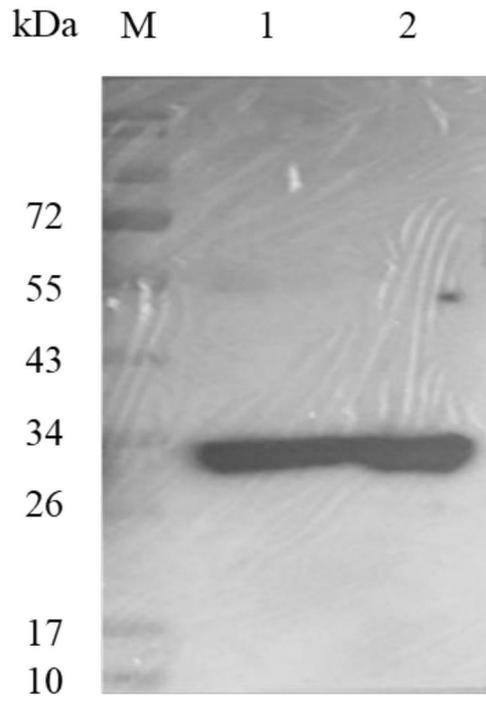


图13

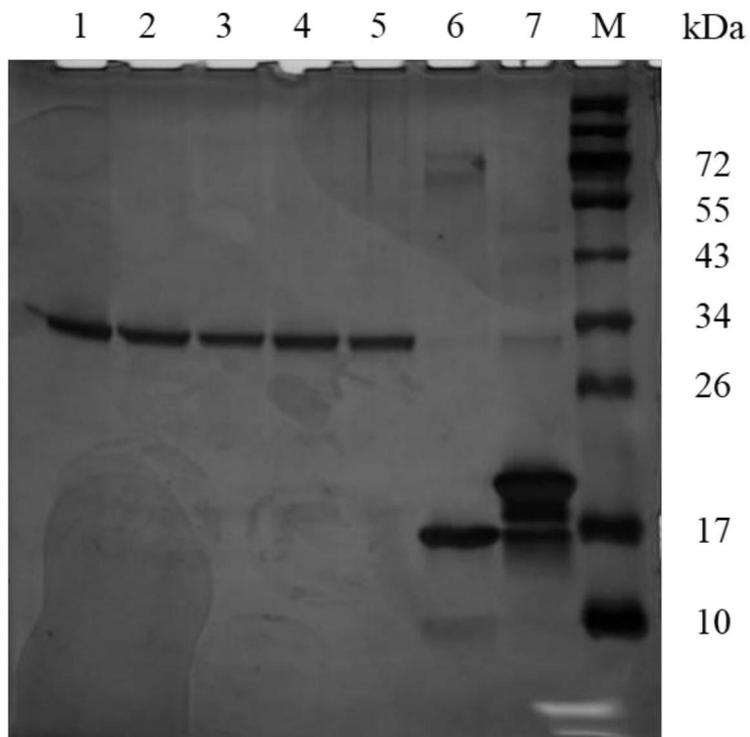


图14