

重组小分子胶原蛋白及其表达系统与制备方法

技术领域

本发明涉及重组小分子胶原蛋白及其表达系统与制备方法，属于合成生物学、基因工程与生物技术领域。

背景技术

胶原蛋白（collagen），有多种生物学功能、良好的生物相容性、生物活性和可降解性等独特功能特征，是最为理想的生物材料来源，可广泛应用于化工、医药、食品、化妆品等众多领域，具有广阔的应用前景。现在市场中以动物源性的胶原蛋白占主导地位，近几年基因工程制备的重组胶原蛋白开始兴起，但无论是动物源性胶原蛋白还是重组胶原蛋白，多为长氨基酸序列、大分子量的蛋白质，动物源性胶原蛋白分子量一般为 100~300KDa 及以上，重组胶原蛋白分子量一般也都在 20KDa 以上（200 个氨基酸以上），以 50KDa 居多。这样大分子量的胶原蛋白作为生物材料应用于组织再生修复、植入类的医疗器械、导入类的医美产品都是适宜的，但并不适合于一些需要透皮吸收的应用场景。

动物源性来源的分子量比较小的胶原蛋白多为小分子胶原多肽，在需要透皮吸收的应用场景中使用较多，如水产品来源的小分子胶原蛋白多肽，但这些产品多以酶解、水解、酸碱降解等方式制备获得，所获得的小分子胶原蛋白多肽并不是一种氨基酸序列、分子量明确、大小单一的小分子蛋白，而是一种统称，分子量从极小的几肽到一定氨基酸序列长度的小分子胶原多肽均有分布，几乎无法确知种类、数目时，氨基酸序列也是随机的（未知的），不同批次制备的小分子胶原多肽的质量也基本上随机的，其氨基酸序列、分子量大小均分布、比例均无法控制。综上，现有小分子胶原种类的产品，研究也较多，但多为动物源胶经过处理（酶解、水解）后获得的复合物，有用于透皮吸收相关的应用，但氨基酸序列、分子量、批次间差异等均不可有效控制。

申请号 CN202211579848.5 的专利，公开了一种小分子重组胶原蛋白肽及其制备方法，本质是仍是一种类似于动物源小胶原肽的制备方法，获得的小分子重组胶原蛋白肽是很多种氨基酸序列不同、分子量大小不同（只能计算其平均分子量）的胶原蛋白多肽的混合物，并不是一种氨基酸序列、分子量明确单一的小分子重组胶原蛋白（多肽）。而其它现有重组小分子胶原蛋白多肽，难以做到小分子量与生物学活性的平衡。合成生物学、基因工程制备的重组胶原蛋白多为中、大分子量的产品（有一部分研发成果称为多肽，但氨基酸数目已远大于 100 个，已不属于小分子的范畴），只有极少数重组小分子胶原多肽的研究成果，其重组胶原蛋白的氨基酸序列分子量比较小。胶原蛋白作为一种活性物质，需要一定的氨基酸序列和序列长度才能支撑其生物学活性，即小分子量与生物学活性需要达到一种平衡的状态方有意义，而现有小分子的重组

胶原多肽却没有相应的证据表明其有好的生物学活性。

其它的小分子蛋白质或多肽在进行重组表达时相同，重组小分子胶原蛋白整体氨基酸数目少、分子量小，在基因重组表达系统中，尤其是真核表达体系（比如毕赤酵母表达体系、哺乳动物细胞表达体系）中，其表达效率、产量受到极大的限制。有一些技术手段，比如串联表达提高基因拷贝数，但往往会在小分子蛋白质或多肽上引入非必要的氨基酸。胶原蛋白是一种特殊的蛋白质，其典型的特征是其氨基酸序列为 G-X-Y 三联体重复结构，引入非必要的氨基酸往往会破坏这种三联体重复结构，影响其功能活性等。

现有重组小分子胶原蛋白或多重组小分子胶原多肽的研究很少，明确的重组小分子胶原蛋白的产品基本处于空白，并且表达量的问题涉及更少。有研究使用融合前导肽的方法提高小分子蛋白质的表达量，如公开号为 CN108148114B 的专利，但需要表达的小分子蛋白质于体外（细胞）使用蛋白酶酶切才能去除前导肽，在表达产物中人为引入了外源性的蛋白，增加后续纯化工艺的步骤、成本，也会带来外源蛋白残留的风险。另外有一些研究使用小分子蛋白质串联表达的方法来提高表达效率以达到提高产量的目的，比如 CN110305890A，使用对 5 种抗菌肽串联表达重组，但使用的是甲酸切割串联的多肽，切割位点只能位 DP 两个氨基酸残基之间，设计空间受限，且切割后 DP 也无法有效去除，成为非自身氨基酸，同样属于人为引入了外源性蛋白。这些方案也未见其用于重组小分子胶原蛋白的表达生产。

所以，该技术领域关于重组小分子胶原蛋白的表达，需要做到小分子量、具有生物学活性、可规模化的量产、表达量高且其氨基序列与分子量需单一明确而非混合物等多方面的平衡，亟需一种有效的解决方式。

发明内容

本发明的目的在于，克服现有技术中存在的一些技术问题，提供一种重组小分子胶原蛋白、其表达系统及其制备方法。

为实现上述目的，本发明采用了以下技术方案：

本发明首先提供了一种重组小分子胶原蛋白，所述重组小分子胶原蛋白包括 SEQ ID NO.1、SEQ ID NO.3、SEQ ID NO.5、SEQ ID NO.7、SEQ ID NO.63 或 SEQ ID NO.67 所示的氨基酸序列、或与 SEQ ID NO.1、SEQ ID NO.3、SEQ ID NO.5、SEQ ID NO.7、SEQ ID NO.63 或 SEQ ID NO.67 具有 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、99%以上同一性的氨基酸序列，并维持胶原蛋白生物学活性。

进一步的，所述重组小分子胶原蛋白的羧基端有多个 His，可形成 6×His Tag 标签。

进一步的，所述重组小分子胶原蛋白包括 SEQ ID NO.2、SEQ ID NO.4、SEQ ID NO.6 或 SEQ ID NO.8 所示的氨基酸序列、或与 SEQ ID NO.2、SEQ ID NO.4、SEQ ID NO.6 或 SEQ ID NO.8

具有 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、99%以上同一性的氨基酸序列，并维持胶原蛋白生物学活性。

本发明还提供了一种编码所述重组小分子胶原蛋白的核酸。

进一步的，所述核酸包括 SEQ ID NO.9-16 所示的核苷酸序列、或其简并序列。

本发明还提供了一种包含所述核酸的重组载体，所述载体包括 pPICZaB、pFLDa、pPIC9K，连接位点为 Kex2 酶切位点序列与 NotI 酶切位点之间。

本发明还提供了一种包含所述核酸、或包含所述的重组载体、或表达所述重组小分子胶原蛋白的重组工程菌。

进一步的，所述工程菌的宿主菌包括毕赤酵母、酿酒酵母、汉逊酵母等，优选毕赤酵母。

根据本发明的实施例，所述工程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心，保藏编号为 CGMCC No.25811、CGMCC No.25823、CGMCC No.25812、CGMCC No.25824、CGMCC No.25825、CGMCC No.25813、CGMCC No.25826、CGMCC No. 25814。

本发明还提供重组小分子胶原蛋白的串联重复表达序列，所述串联重复表达序列以所述的重组小分子胶原蛋白、或人工设计的具有典型的 G-X-Y 三联体结构的胶原蛋白、或以人胶原蛋白序列的两个或多个区域组合的蛋白为基础单元串联重复；所述串联重复表达序列中每相邻的 2 个所述重组小分子胶原蛋白之间有 Kex2 酶、CPB 酶识别切割的位点，还可以同时有 Ste13 酶的识别、切割的位点。

所述串联重复表达序列包括 SEQ ID NO.30、SEQ ID NO.33、SEQ ID NO.36、SEQ ID NO.39、SEQ ID NO.42、SEQ ID NO.45、SEQ ID NO.48、SEQ ID NO.51、SEQ ID NO.54、SEQ ID NO.57、SEQ ID NO.60、SEQ ID NO.64、SEQ ID NO.68 所示序列、或与上述具有 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、99%以上同一性的氨基酸序列。

根据本发明的实施例，所述 Kex2 酶识别和切割的位点包括 KR 或 RR 双碱性氨基酸残基，双碱性氨基酸残基之后，是 EA、EAEA 或其它有助于 Kex2 酶或 Ste 13 酶识别和切割的氨基酸残基。

根据本发明的实施例，所述 CPB 酶识别和切割的位点包括蛋白羧基末端的碱性氨基酸残基，包括 K、R。

根据本发明的实施例，所述串联重复的小分子胶原蛋白单体（或片段）的氨基酸序列可以是任一序列，其氨基酸残基的具体排序、序列长度、单体（或片段）串联重复的次数不受限制，但需具有典型的 G-X-Y 三联体结构。

本发明还提供一种编码所述串联重复表达序列的核酸，进一步的，所述核酸包括 SEQ ID NO.71-81 所示序列、或其简并序列。

本发明还提供一种包含编码所述串联重复表达序列的核酸的重组载体。根据本发明的实施例，所述载体包括但不限于 pPICZαB、pFLDα、pPIC9K 等表达载体。

本发明还提供包含所述核酸、或重组载体的工程菌。

根据本发明的实施例，所述工程菌的宿主菌包括毕赤酵母、酿酒酵母、汉逊酵母等，优选毕赤酵母。根据本发明的实施例，所述工程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心，保藏编号为：CGMCC No.25819、CGMCC No.25821、CGMCC No.25827、CGMCC No.25829、CGMCC No.25828、CGMCC No.25820、CGMCC No.25822。

本发明还提供了一种获得重组小分子胶原蛋白的方法，所述方法包括：

构建本发明所述的重组小分子胶原蛋白的串联重复表达序列；

构建定位融合功能蛋白，将定位融合功能蛋白连接载体转化宿主菌得到底盘细胞或底盘工程菌；

将重组小分子胶原蛋白的串联重复表达序列连接到表达载体后转入底盘细胞或底盘工程菌中，得到包含或表达重组小分子胶原蛋白的重组工程菌，发酵诱导表达，得到重组小分子胶原蛋白。

根据本发明的实施例，所述定位融合功能蛋白包括 CPB 酶、有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域。

进一步的，所述定位融合功能蛋白还包括连接序列，所述连接序列用于 CPB 酶与有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列融合表达时的连接。优选的，所述连接序列为 Linker 序列，如 SEQ ID NO.21 所示：GGSGSGSGGS。

根据本发明的实施例，优选的，所述 CPB 酶来源于人或大鼠；所述 CPB 酶序列如 SEQ ID NO.17-18 所示。

所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列来源于酿酒酵母或毕赤酵母的 Kex2 酶、或其它有类似功能的蛋白质功能区域如 Ste13 蛋白酶。

进一步的，所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列如 SEQ ID NO.19-20 所示。

根据本发明的实施例，所述串联重复表达序列中每相邻的 2 个所述重组小分子胶原蛋白之间有 Kex2 酶、CPB 酶的识别、切割的位点，还可以包括 Ste13 酶的识别、切割的位点；所述小分子胶原蛋白优选本发明所述的重组小分子胶原蛋白。

根据本发明的实施例，所述 Kex2 酶识别和切割的位点包括 KR 或 RR 双碱性氨基酸残基，双碱性氨基酸残基之后，是 EA、EAEA 或其它有助于 Kex2 酶或 Ste 13 酶识别和切割的氨基酸残基。

根据本发明的实施例,所述 CPB 酶识别和切割的位点包括蛋白羧基末端的碱性氨基酸残基,包括 K、R。

进一步的,所述重组小分子胶原蛋白具有典型的 G-X-Y 三联体结构;优选的,所述小分子胶原蛋白为本发明所述的重组小分子胶原蛋白、或人工设计的具有典型的 G-X-Y 三联体结构的胶原蛋白、或以人胶原蛋白序列的两个或多个区域组合的胶原蛋白。。

根据本发明的实施例,所述载体包括但不限于 pPICZαB、pFLDα、pPIC9K 等表达载体。

根据本发明的实施例,所述包含或表达重组小分子胶原蛋白的重组工程菌的重组工程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心,保藏编号为:CGMCC No.25819、CGMCC No.25821、CGMCC No.25827、CGMCC No.25828、CGMCC No.25829、CGMCC No.25820、CGMCC No.25822。

根据本发明的实施例,上述的方法表达得到的重组小分子胶原蛋白,包括 SEQ ID NO.32、SEQ ID NO.35、SEQ ID NO.38、SEQ ID NO.41、SEQ ID NO.44、SEQ ID NO.47、SEQ ID NO.50、SEQ ID NO.53、SEQ ID NO.56、SEQ ID NO.59、SEQ ID NO.62、SEQ ID NO.66、或 SEQ ID NO.70 所示的氨基酸序列或与 SEQ ID NO.32、SEQ ID NO.35、SEQ ID NO.38、SEQ ID NO.41、SEQ ID NO.44、SEQ ID NO.47、SEQ ID NO.50、SEQ ID NO.53、SEQ ID NO.56、SEQ ID NO.59、SEQ ID NO.62、SEQ ID NO.66、或 SEQ ID NO.70 具有 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、99%以上同一性的氨基酸序列。

本发明还提供一种定位融合功能蛋白,所述定位融合功能蛋白用于串联重复的多个重组小分子胶原蛋白被 Kex2 酶切割为单体后其 C 端碱性氨基酸残基(K、R)的切割去除。

所述定位融合功能蛋白包括 CPB 酶成熟肽序列、有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列。

进一步的,所述定位融合功能蛋白还包括连接序列,所述连接序列用于 CPB 酶与有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列融合表达时的连接。

根据本发明的实施例,所述 CPB 酶来源于人或大鼠;进一步的,所述 CPB 酶优选 SEQ ID NO.17-18 所示的序列。

根据本发明的实施例,所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列来源于酿酒酵母或毕赤酵母的 Kex2 酶、或其它有类似功能的蛋白质功能区域如 Ste13 蛋白酶。

进一步的,所述 CPB 酶序列如 SEQ ID NO.17-18 所示;所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列如 SEQ ID NO.19-20 所示。

根据本发明的实施例,所述连接序列为 Linker 序列,如 SEQ ID NO.21 所示:GGSGSGSGGS。

根据本发明的实施例,所述定位融合功能蛋白包括如 SEQ ID NO.22、SEQ ID NO.24、SEQ ID NO.26 或 SEQ ID NO.28 所示的氨基酸序列或与 SEQ ID NO.22、SEQ ID NO.24、SEQ ID NO.26 或 SEQ ID NO.28 具有 80%以上、85%以上、90%以上、95%以上、96%以上、97%以上、98%以上、99%以上同一性的氨基酸序列。

本发明还提供编码本发明所述定位融合功能蛋白的核酸。

根据本发明的实施例,所述编码本发明所述定位融合功能蛋白的核酸包括 SEQ ID NO.23、SEQ ID NO.25、SEQ ID NO.27 或 SEQ ID NO.29 所示的核苷酸序列、或其简并序列。

本发明还提供一种包含编码本发明所述定位融合功能蛋白的核酸的重组载体。

根据本发明的实施例,所述载体包括但不限于 pPICZαB、pFLDα、pPIC9K 等表达载体,连接位置为 Kex2 酶切位点序列与载体上终止密码子“TGA”之间。

本发明还提供一种底盘细胞或底盘工程菌,所述底盘细胞或底盘工程菌包含所述编码本发明定位融合功能蛋白的核酸、或包含编码本发明所述定位融合功能蛋白的核酸的重组载体、或表达所述定位融合功能蛋白。

根据本发明的实施例,所述工程菌的宿主菌包括毕赤酵母、酿酒酵母、汉逊酵母等,优选毕赤酵母;所述构建的底盘细胞或工底盘程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心,保藏编号为:CGMCC No.25815、CGMCC No.25817、CGMCC No.25816、CGMCC No.25818。

本发明还提供一种重组小分子胶原蛋白的表达系统,所述表达系统包括所述定位融合功能蛋白、所述底盘细胞或底盘工程菌、所述重组小分子胶原蛋白的串联重复表达序列。

本发明还提供所述定位融合功能蛋白、所述底盘细胞或底盘工程菌、所述重组小分子胶原蛋白的串联重复表达序列在表达重组小分子胶原蛋白中的应用。

本发明所述的重组小分子胶原蛋白、或上述获得小分子胶原蛋白的方法表达得到的小分子胶原蛋白具有细胞粘附活性、促细胞增殖活性、良好的皮肤渗透性。

本发明还提供所述重组小分子胶原蛋白、或所述核酸、或所述重组载体、或所述的宿主细胞或重组工程菌、或所述的组合物、或所述的制品在制备药物、医疗器材、生物材料、组织工程产品、化妆品或保健品中的用途。

本发明的有益效果:

(1) 本发明首先提供了多种Ⅲ型、XVII 型胶原蛋白来源的重组小分子胶原蛋白,并表达了

多种Ⅲ型、XVII型胶原蛋白来源的重组小分子胶原蛋白，其氨基酸序列长度、分子量大小虽不相同，都可表达获得为条带单一、氨基酸序列明确的小分子胶原蛋白，同时都有很好的透皮吸收性、生物学活性，在小分子与透皮吸收及生物学活性上做到兼顾平衡。进一步的在500L级的发酵罐上实现了规模化量产，使重组小分子胶原蛋白的大量应用成为可能。解决了重组小分子胶原蛋白现有的产品缺失的问题，尤其是Ⅲ型、XVII型胶原蛋白来源的重组小分子胶原蛋白。

(2) 本发明建立了一个重组小分子胶原的表达体系和小分子胶原的制备方法，包括构建了专属的底盘细胞和胶原蛋白串联重复序列设计方法。本发明的技术方案使得小分子胶原蛋白的表达极大提升了产量；本发明的技术方案不引入外源蛋白质、不使用任何的体外酶切的方式，小分子胶原蛋白在细胞内分泌的过程中，就会去除重复串联设计时酶切位点的残留氨基酸残基，获得无非胶原序列或与天然胶原蛋白相应区域同源性100%的重组小分子胶原蛋白；整个表达体系避免了体外使用蛋白酶酶切带来的成本及外源蛋白残留风险，同时也能缩短后续纯化工艺流程的时间及成本。

本发明取得的小分子胶原蛋白进行了N、C端及全序列测序验证，证明整个表达体系，尤其是底盘细胞工程菌作为宿主菌表达重组小分子胶原的有效性；进一步以500L级发酵罐进行了具有大规模工业化生产可能性的发酵、纯化验证，结果表明重复串联表达策略构建的表达体系可以提升重组小分子胶原蛋白的产量4-5倍。

(3) 对本发明的重组小分子的细胞粘附活性、促细胞增殖活性及透皮吸收性进行了检测，结果表明，重组小分子胶原除去具有良好的透皮吸收性外，还有不弱于天然胶原蛋白及同源的大分子量重组胶原的细胞粘附活性，有促细胞增殖的活性。

附图说明

图1为8种重组小分子胶原蛋白3A5D1NT、3A5D1、3A5D2NT、3A5D2、17S1NNT、17S1N、17S3NT、17S3纯化后进行Tricine-PAGE检测的结果。

图2为本发明表达得到的重组小分子胶原蛋白3A5D1NT、3A5D1经LC-MS分析获取的其去卷积分子量结果。

图3为本发明表达得到的重组小分子胶原蛋白3A5D2经LC-MS分析获取的其去卷积分子量结果。

图4为本发明表达得到的重组小分子胶原蛋白17S1NNT、17S1N经LC-MS分析获取的其去卷积分子量结果。

图5为本发明表达得到的重组小分子胶原蛋白17S3NT、17S3经LC-MS分析获取的其去卷积分子量结果。

图6为3A5D1、3A5D2、17S1N、17S3羧基端6×His标签用抗6×His Tag的抗体进行Western

Blot 检测的结果：图中，+为阳性对照，-为阴性对照。

图 7 为本发明实施例 1 中重组小分子胶原蛋白 3A5D1、3A5D1NT 冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 8 为本发明实施例 1 中重组小分子胶原蛋白 3A5D2、3A5D2NT 冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 9 为本发明实施例 1 中重组小分子胶原蛋白 17S1N、17S1NNT 冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 10 为本发明实施例 1 中重组小分子胶原蛋白 17S3、17S3NT 冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 11 为本发明中重组表达载体 pPICZαB-RCPB-SCKEX2 的图谱。

图 12 为本发明中重组表达载体 pPICZαB-HCPB-SCKEX2 的图谱。

图 13 为本发明中重组表达载体 pPICZαB-RCPB-PPKEX2 的图谱。

图 14 为本发明中重组表达载体 pPICZαB-HCPB-PPKEX2 的图谱。

图 15 为 CPB 定位融合功能蛋白 HCPB-PPKEX2、HCPB-SCKEX2、RCPB-PPKEX2、RCPB-SCKEX2 在胞内裂解液和培养上清中表达的检测结果；

左图中泳道，1：HCPB-SCKEX2 菌体裂解液；2：HCPB-SCKEX2 菌培养上清；3、6：HCPB-PPKEX2 培养上清；4、5：HCPB-PPKEX2 菌体裂解液；(-)：阴性对照；(+)：阳性对照，重组人 CPB 酶；/：无样品孔。

右图中泳道，(-)：阴性对照；(+)：阳性对照，重组大鼠 CPB 酶；/：无样品孔；7：RCPB-SCKEX2 菌培养上清；8：RCPB-SCKEX2 菌体裂解液；9、10、12、13：RCPB-PPKEX2 培养上清；11、14：RCPB-PPKEX2 菌体裂解液。

图 16 为 3A5D29N-9 和 17S28-8 分别于 HCPB-PPKEX2 和 RCPB-PPKEX2 底盘细胞工程菌株表达上清 SDS-PAGE 检测结果；

图中各泳道为，1：3A5D29N-9 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；2：3A5D29N-9 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；3：17S28-8 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；4：17S28-8 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清。

图 17 为 17S1N6-6、17S1N7-7 和 17S1NK-7 分别于 HCPB-PPKEX2 和 RCPB-PPKEX2 底盘细胞工程菌株表达上清 SDS-PAGE 检测结果；

图中各泳道为，1：17S1N6-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；2：17S1N6-6 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；3：17S1N7-7 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；4：17S1N7-7 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；5：17S1NK-7

于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；6：17S1NK-7 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；(-)：阴性对照，pPIC9K 空载体转入 HCPB-PPKEX2 底盘细胞工程菌株中表达上清。

图 18 为 3A5D15D-5、3A5D15E-5、3A5D15G-6 和 3A5D15EKR-6 分别于 RCPB-PPKEX2 底盘细胞工程菌株表达上清，和 3A5D15R-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清 SDS-PAGE 检测结果；

图中各泳道为，1：3A5D15D-5 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；2：3A5D15E-5 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；3：3A5D15G-6 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；4：3A5D15EKR-6 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；5：3A5D15KR-6 于 RCPB-PPKEX2 底盘细胞工程菌株中表达上清；6：3A5D15R-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；(-)：阴性对照，pPIC9K 空载体转入 RCPB-PPKEX2 底盘细胞工程菌株中表达上清。

图 19 为 3A5D15D-5、3A5D15E-5、3A5D15G-6、3A5D15EKR-6、3A5D15KR-6、3A5D15R-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清 SDS-PAGE 检测结果；

图中各泳道为，1：3A5D15D-5 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；2：3A5D15E-5 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；3：3A5D15G-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；4：3A5D15EKR-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；5：3A5D15KR-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清；6：3A5D15R-6 于 HCPB-PPKEX2 底盘细胞工程菌株中表达上清。

图 20 为 3A5D29N、3A5D15D 的冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 21 为 17S28、17S1N6 的冻干品用胰蛋白酶酶解后的肽段经 Nano-HPLC-MS/MS 质谱检测后与原序列比对的结果。

图 22 为 3A5D29N、17S28 经 LC-MS 分析获取的其去卷积分子量结果。

图 23 为 HCPB-PPKEX2 和 RCPB-PPKEX2 底盘细胞工程菌株中表达的重组小分子胶原蛋白 3A5D29N 的 C 端肽段的二级质谱图。

图 24 为 HCPB-PPKEX2 和 RCPB-PPKEX2 底盘细胞工程菌株中表达的重组小分子胶原蛋白 17S28 的 C 端肽段的二级质谱图。

图 25 表达 3A5D2NT 和表达 3A5D29N 的工程菌株诱导一定时间后的发酵液上清 SDS-PAGE 检测结果。

图中各泳道为，泳道 1：表达 3A5D2NT 的工程菌株（CGMCC No.25812）诱导 48h 时发酵

液上清；泳道 2：表达 3A5D29N 的工程菌株（CGMCC No.25819）诱导 43h 时发酵液上清。

图 26 为重组小分子胶原蛋白 3A5D29N、3A5D2NT、3A5D1NT、17S1NNT、17S3NT、17S28、重组Ⅲ型胶原蛋白、重组 XVII 型胶原蛋白和天然人胶原蛋白的细胞黏附活性对比结果。

图 27 为重组小分子胶原蛋白 3A5D29N、17S28 促细胞增殖作用的验证结果。

图 28 为 1mg/mL 的重组小分子胶原蛋白 3A5D29N、重组Ⅲ型胶原蛋白、水解透明质酸、鱼皮胶原于涂抹 1h 时全层皮肤模型组织中的检测的荧光 IOD 平均值，其中*表示与 1mg/mL 重组小分子胶原蛋白 3A5D29N 相比的显著性（*表示差异显著 $P<0.05$ ，**表示差异极显著 $P<0.01$ ），#表示与重组Ⅲ型胶原蛋白相比的显著性（#表示差异显著 $P<0.05$ ，##表示差异极显著 $P<0.01$ ）。

图 29 为不同浓度重组小分子胶原蛋白 3A5D29N 及 1mg/mL 的水解透明质酸涂抹 12h 于渗透至培养液中的样品扩散百分比（即累积渗透率），其中 2#代表重组小分子胶原蛋白 3A5D29N，4#代表水解透明质酸。

具体实施方式

为了使本领域技术人员更好的理解本发明的技术方案，下面对本发明的较佳实施例进行详细的阐述，但是如下实施例并不限制本发明的保护范围。

本发明的实施例中，没有多作说明的都是采用常规实验方法完成，实施例中所涉及过程没有多作说明的都是本领域技术人员根据产品说明书或本领域基础知识可以理解并且容易实现的，因此不再详细描述。

实施例 1： 重组小分子胶原蛋白单序列重组表达

（1）重组小分子胶原蛋白单序列重组表达策略的氨基酸序列设计

选择人Ⅲ型胶原蛋白全长（包含 N 前肽、成熟肽链、C 前肽）序列（参照 Uniprot 数据库中 ID 为 P02461 的蛋白质序列，<https://www.uniprot.org/uniprotkb/P02461>）的第 990-1085 位氨基酸序列，将此段序列表达获得的重组小分子胶原蛋白命名为 3A5D1NT，共 96 个氨基酸，理论分子量 8596.08Da，其序列如 SEQ ID NO.1 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGP
AGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGH

在 3A5D1NT 序列的羧基端添加 5 个组氨酸（H），将此段序列表达获得的重组小分子胶原蛋白命名为 3A5D1，共 101 个氨基酸，理论分子量 9281.78Da，其序列如 SEQ ID NO.2 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGP
AGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHHHHH

选择人Ⅲ型胶原蛋白全长（包含 N 前肽、成熟肽链、C 前肽）序列（参照 Uniprot 数据库中 ID 为 P02461 的蛋白质序列，<https://www.uniprot.org/uniprotkb/P02461>）第 1036-1085 位氨基酸序

列，将此段序列表达获得的重组小分子胶原蛋白命名为 3A5D2NT，共 50 个氨基酸，理论分子量为 4532.83Da，其序列如 SEQ ID NO.3 所示：

GKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGH

在 3A5D2NT 序列的羧基端添加 5 个组氨酸 (H)，将此段序列表达获得的重组小分子胶原蛋白命名为 3A5D2，共 55 个氨基酸，理论分子量 5218.54Da，其序列如 SEQ ID NO.4 所示：

GKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHHHHH

选择人 XVII 胶原蛋白序列全长序列（参照 Uniprot 数据库中 ID 为 Q9UMD9-1 的蛋白质，<https://www.uniprot.org/uniprot/Q9UMD9>）中第 636-718 位氨基酸，将此段序列表达获得的重组小分子胶原蛋白命名为 17S1NNT，共 82 个氨基酸，理论分子量为 7483.28Da，其序列如 SEQ IDNO.5 所示：

GPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGP

在 17S1NNT 序列的羧基端添加 6 个组氨酸 (H)，将此段序列表达获得的重组小分子胶原蛋白命名为 17S1N，共 88 个氨基酸，理论分子量 8306.13Da，其序列如 SEQ ID NO.6 所示：

GPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGPHHHHH

选择人 XVII 胶原蛋白序列全长序列（参照 Uniprot 数据库中 ID 为 Q9UMD9-1 的蛋白质，<https://www.uniprot.org/uniprot/Q9UMD9>）中第 659-717 位氨基酸，将此段序列表达获得的重组小分子胶原蛋白命名为 17S3NT，共 59 个氨基酸，理论分子量为 5406.08Da，其序列如 SEQ IDNO.7 所示：

GVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGP

在 17S3NT 序列的羧基端添加 6 个组氨酸 (H)，将此段序列表达获得的重组小分子胶原蛋白命名为 17S3，共 65 个氨基酸，理论分子量 6228.93Da，其序列如 SEQ ID NO.8 所示：

GVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGP
HHHHHH

(2) 重组小分子胶原蛋白单序列重组表达策略的重组表达载体、工程菌株的构建

编码 3A5D1NT 的 DNA 序列以毕赤酵母为宿主进行密码子优化，优化后的序列如 SEQ ID NO.9 所示：

gacgggaaccaggagtgatggcctgccaggacgcgatggctccccgggagggaaggggacccgggcgaaatggttcgcccgga
gcccgggtgctcctgtgcacctggtccgccgggctgtcggaccagccggcaaatctggcgatcgaggagagagcgcccgactggtccc
gctggggcaccaggtccggctgggtcaaggggtgcacctggcctcaggagaccagaggagacaagggcgagaccggtgaacgtggagccg
cgggtattaaagtcactaa

编码 3A5D1 的 DNA 序列以毕赤酵母为宿主进行密码子优化，优化后的序列如 SEQ ID NO.10

所示:

gacgggaacccagggagtgatggcctgccaggacgcgatggctccccgggaggggaagggggaccggggcgaaaatggttcggccgga
gcggcgggtgctcctgtgcacctggctcgccgggctgtcggaccagccggcaaatctggcgatcgaggagagagcggcccgctggctcc
gctggggcaccaggtccggctgggtcaaggggtgcacctggccctcagggaccagaggagacaagggcgagaccggtgaacgtggagccg
cgggtattaaaggtcacatcatcatcaccactaa

编码 3A5D2NT 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.11 所示:

gaggcaggggaagtccggtgacaggggggagagcggacctgctggctcctgctggagccccgggtcctgctggctctcgggcgccag
gaccgcaagggccacgtggtgataaaggagaaacaggggaaagagggggccgaggtattaaaggtcactaa

编码 3A5D2 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.12 所示:

gaggcaggggaagtccggtgacaggggggagagcggacctgctggctcctgctggagccccgggtcctgctggctctcgggcgccag
gaccgcaagggccacgtggtgataaaggagaaacaggggaaagagggggccgaggtattaaaggtcacatcatcatcactaa

编码 17S1NNT 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.13 所示:

gaggccggccctccaggttcgggcgagaaaggggaaagagggggcgagcaggagaacccggaccccggtcctcctggagtcggggtt
ccgtaggcccaaaaggcagctctggaagtccggggccacaagggcctccagggccgttggtttacagggctacgcggcgaagtgggattgcc
aggggtaaagggcgacaaggtccgatgggacctccgggtcctaaggggtgatcaaggggagaaaggaccgtaa

编码 17S1N 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.14 所示:

gaggccggccctccaggttcgggcgagaaaggggaaagagggggcgagcaggagaacccggaccccggtcctcctggagtcggggtt
ccgtaggcccaaaaggcagctctggaagtccggggccacaagggcctccagggccgttggtttacagggctacgcggcgaagtgggattgcc
aggggtaaagggcgacaaggtccgatgggacctccgggtcctaaggggtgatcaaggggagaaaggaccgcatcaccatcatcattaa

编码 17S3NT 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.15 所示:

gaagccggtgtccctggctctgttgaccgaagggtagttcggaagcccagggcctcagggccgcccggacccgtaggtctgcaggg
gctcagaggagaggtgggcctacctggcgtcaaggcgacaaaggtccaatggggccgcctggaccaaaggggatcaaggtgaaaaaggcc
cctaa

编码 17S3 的 DNA 序列以毕赤酵母为宿主进行密码子优化, 优化后的序列如 SEQ ID NO.16 所示:

gaagccggtgtccctggctctgttgaccgaagggtagttcggaagcccagggcctcagggccgcccggacccgtaggtctgcaggg
gctcagaggagaggtgggcctacctggcgtcaaggcgacaaaggtccaatggggccgcctggaccaaaggggatcaaggtgaaaaaggcc
cccacatcatcaccatcactaa

DNA 序列 SEQ ID NO.9~SEQ ID NO.16 均由南京金斯瑞生物科技有限公司合成。将合成的基因克隆于 pPIC9K 表达载体(赛默飞世尔科技(中国)有限公司)上,克隆位置为 α -factor secretion signal (分泌信号肽) 序列中 Kex2 酶切位点序列 (Kex2 酶切位点末端 DNA 序列为 AAAGAAGAAGGGGTATCTCTCGAGAAAAGA) 与 NotI 酶切位点之间。获得重组表达载体 pPIC9K-3A5D1NT、pPIC9K-3A5D1、pPIC9K-3A5D2NT、pPIC9K-3A5D2、pPIC9K-17S1NNT、pPIC9K-17S1N、pPIC9K-17S3NT、pPIC9K-17S3。理论而言, 其它可用于毕赤酵母的表达载体如 pPICZ α B、pFLD α 也与 pPIC9K 有类似的效果, 相应重组表达载体均在本发明保护范围之内。

分别将上述重组表达载体质粒 (pPIC9K-3A5D1NT、pPIC9K-3A5D1、pPIC9K-3A5D2NT、pPIC9K-3A5D2、pPIC9K-17S1NNT、pPIC9K-17S1N、pPIC9K-17S3NT、pPIC9K-17S3) 10 μ g 用 SacI (购自大连 TaKaRa 公司, 具体操作按试剂盒说明书进行) 37°C 酶切消化过夜, 使其线性化, 再使用 PCR 产物纯化试剂盒 (购自生工生物工程(上海)股份有限公司) 回收线性化质粒, 使体积控制在 10 μ L 左右。

将线性化质粒电转化入空宿主菌种毕赤酵母 GS115 (购自中国工业微生物菌种保藏管理中心) 感受态细胞中, 将电转后的菌液涂布于 MD 平板上, 每 100 μ L~200 μ L 涂布一块平板, 室温静置 10min, 于 30°C 倒置培养 2-5 天, 直至有单菌落 (阳性转化子) 出现。

向 MD 平板表面加入 2mL 无菌双蒸水, 然后用无菌三角涂布器轻轻刮下平板表面的 His⁺转化子, 并转移到 50mL 离心管中。以无菌双蒸水稀释菌悬液, 105 个细胞涂布于含有 0.5mg/mL G418 的 YPD 平板上, 倒置, 30°C 培养 3~4d 后至单菌落出现。从 YPD 平板上挑取菌落至无菌 96 孔板中 (200 μ L YPD/孔), 混匀, 于 30°C 培养 48h; 混匀孔中菌液, 各取 10 μ L 接入至一块新的无菌 96 孔板, 于 30°C 培养 24h 后再重复一次此操作; 24h 后, 从第三块 96 孔板中取出 1 μ L 分别点在含有 1.0mg/mL 和 4mg/mL G418 的 YPD 平板上, 于 30°C 继续培养 96h~120h。毕赤酵母转化子若能在含高浓度 G418 的平板上生长, 说明该转化子能够高效率表达外源基因, 经过这一步筛选可得到高效表达的重组酵母工程菌种。

构建的 8 种工程菌株样本均送至中国微生物菌种保藏管理委员会普通微生物中心保藏, 菌种保藏编号分别对应的是:

表达重组小分子胶原蛋白 3A5D1NT 的菌株, 保藏编号为: CGMCC No.25823;

表达重组小分子胶原蛋白 3A5D1 的菌株, 保藏编号为: CGMCC No.25811;

表达重组小分子胶原蛋白 3A5D2NT 的菌株, 保藏编号为: CGMCC No.25824;

表达重组小分子胶原蛋白 3A5D2 的菌株, 保藏编号为: CGMCC No.25812;

表达重组小分子胶原蛋白 17S1NNT 的菌株, 保藏编号为: CGMCC No.25825;

表达重组小分子胶原蛋白 17S1N 的菌株, 保藏编号为: CGMCC No.25813;

表达重组小分子胶原蛋白 17S3NT 的菌株，保藏编号为：CGMCC No.25826；

表达重组小分子胶原蛋白 17S3 的菌株，保藏编号为：CGMCC No. 25814；

上述 8 种工程菌保藏地址均为：北京市朝阳区北辰西路 1 号院 3 号；保藏日期均为：2022 年 9 月 26 日。分类命名均为：巴斯德毕赤酵母 *Komagataella phaffii*。

(3) 重组小分子胶原蛋白单序列重组表达策略的表达、纯化与鉴定

使用步骤(2)中得到的 8 种高效表达的重组酵母工程菌株，采用分批补料培养方法，进行重组小分子胶原蛋白高密度发酵表达生产，获取含有重组胶原蛋白的发酵液，纯化获得高纯度的重组小分子胶原冻干海绵。具体操作如下：

分别挑选步骤(2)中得到的 8 种高效表达的重组酵母工程菌株接入含种子培养基 YPG 的 1L 摇瓶，220rpm、30℃，培养 18-20h，至 $OD_{600}=2\sim10$ ；将 200mL 种子液接入 5L 发酵罐（购自上海保兴生物设备工程有限公司），装液量 2L 发酵培养基，以甲醇、甘油混合碳源（甲醇：50%甘油=7：3）进行诱导培养。诱导 40~60h，UV 测量蛋白浓度增长幅度不明显或下降即可放罐。UV 测量蛋白浓度公式： $C(\text{mg/mL})=0.144*(A_{215}-A_{225})$ 。

收集发酵液，2000g、30min、4℃离心分离菌体和发酵上清，取发酵上清进行阳离子交换层析纯化（层析填料为苏州纳微产 UniGel-80sp 装载于利穗科技产 GCC-50-400 层析柱，使用 GE AKTA Pure 蛋白质分离层析纯化系统），再进行超滤（使用山东博纳生物科技集团有限公司的 1KDa 有机膜过滤系统，型号 BNUF402-2-A）、脱盐、浓缩，随后冷冻干燥，收集重组小分子胶原蛋白冻干海绵。

取纯化后冻干海绵溶于超纯水，加入 2×小分子蛋白上样缓冲液，置于 100℃金属浴加热 10min，进行 Tricine-PAGE 检测（小分子蛋白适用的 Tricine-PAGE 电泳所需全套试剂均购自上海万生昊天生物技术有限公司，电泳方法参照说明书执行）。检测结果如图 1 所示，图中可见，8 种重组小分子胶原蛋白，有效分泌表达于上清中，且电泳条带单一，无水解小分子胶原蛋白大小无法控制的问题（胶原蛋白于电泳时会有一定的电泳迁移延迟，所以在电泳时其表观分子量会偏大）。

取纯度高的 8 种重组小分子胶原蛋白冻干品进行 LC-MS 分析获取其去卷积分子量（委托北京百泰派克生物科技股份有限公司完成）。结果如图 2-图 5 所示，其中部分结果图中去卷积分子量值显示值保留小数点后 1 位，有四舍五入处理，图中可见，3A5D1NT 理论分子量 8596.08Da，实测去卷积分子量为 8796.68Da；3A5D1 理论分子量 9281.78Da，实测去卷积分子量为 9333.96Da；3A5D2NT 理论分子量为 4532.83Da，实测去卷积分子量为 4591.46Da；3A5D2 理论分子量 5218.54Da，实测去卷积分子量为 5173.04Da；17S1NNT 理论分子量为 7483.28Da，实测去卷积分子量为 7648.41Da；17S1N 理论分子量 8306.13Da，实测去卷积分子量为 8306.00Da；17S3NT 理论分子量为 5406.08Da，实测去卷积分子量为 5499.72Da；17S3 理论分子量 6228.93Da，实测

去卷积分子量为 6503.79Da; 误差的产生是由于蛋白质在表达过程可能产生的糖基化修饰以及检测误差产生的, 所以, 本发明的重组小分子胶原蛋白 3A5D1NT、3A5D1、3A5D2NT、3A5D2、17S1NNT、17S1N、17S3NT、17S3 的实测去卷积分子量与理论值基本一致。

因 3A5D1、3A5D2、17S1N、17S3 羧基端有 6×His 标签, 可以用抗 6×His Tag 的抗体(购自南京金斯瑞生物科技股份有限公司)进行 Western Blot 检测, 使用 ECL 化学发光显色, 全自动化学发光图像分析系统(Tanon 5200)检测。检测结果如图 6 所示, 3A5D1、3A5D2、17S1N、17S3 成功检测到 6×His 标签(阴性对照(-)为 GS115 空菌株的培养上清、阳性对照(+)为带 His 标签的重组人 MIF, 购于生工生物工程(上海)股份有限公司, 均正常), 其条带相应位置也与预期大小相符。

将冻干品用胰蛋白酶将其酶解, Nano-HPLC-MS/MS 质谱检测重组胶原的胰蛋白酶解后的肽段(委托苏州普泰生物技术有限公司完成), 并将检测到的肽段与 Uniprot 数据库中各天然蛋白序列进行比对。结果如图 7-图 10, 质谱检测结果表明 3A5D1、3A5D2、17S1N、17S3、3A5D1NT、3A5D2NT、17S1NNT、17S3NT 这几种重组小分子胶原蛋白酶解后得到的肽段及其所覆盖序列均属于人胶原蛋白序列的相关区域, 说明本发明的重组小分子胶原蛋白成功表达。

实施例 2: 高表达系统或高表达底盘细胞或底盘工程菌的构建

(1) CPB 定位融合功能蛋白的设计

本发明中以毕赤酵母为出发细胞, 配合特殊的表达序列设计, 首先构建一种专属的底盘细胞或底盘工程菌, 所述的底盘细胞或底盘工程菌包括定位融合功能蛋白, 所述定位融合功能蛋白用于串联重复的多个所述的重组小分子胶原蛋白被 Kex2 酶切割为单体后其 C 端碱性氨基酸残基(K、R)的切割去除。

所述定位融合功能蛋白包括 CPB 酶成熟肽序列、有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列和连接序列, 所述连接序列用于 CPB 酶与有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列融合表达时的连接。

其中所述 CPB 酶可以是来源于任何物种的, 本发明优选来源于人或大鼠。

所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列优选来源于酿酒酵母或毕赤酵母, 所述功能区域优选 Kex2 酶、或其它有类似功能的蛋白质功能区域, 如 Ste13 蛋白酶。

所述的连接序列即 Linker 序列不限, 只要能起到连接且不影响 CPB 酶与有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域的连接作用即可。

本实施例中, 所述 CPB 酶优选人 CPB 酶全长(包含信号肽、前肽、成熟肽)序列(Uniprot 数据库 ID 为 P15086, <https://www.uniprot.org/uniprotkb/P15086>)中第 111-417 位氨基酸序列(即

人 CPB 酶成熟肽序列) 为本发明中优选的人 CPB 酶氨基酸序列, 如 SEQ ID NO.17 所示:

ATGHSYKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKVGKAGQNKPAIFMDC
GFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSFRWRT
RSTHTGSSCIGTDPNRNFDAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAY
LTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKELASLHGTYKYTYGPGATTIYPAAGGSD
DWAYDQGIRYSFTFELRDTGRYGFLLPESQIRATCEETFLAIKYVASYVLEHLY

本实施例中, 所述 CPB 酶还优选大鼠 CPB 酶全长(包含信号肽、前肽、成熟肽)序列(Uniprot 数据库 ID 为 P19223, <https://www.uniprot.org/uniprotkb/P19223>) 中第 109-415 位氨基酸序列(即大鼠 CPB 酶成熟肽序列) 为本发明中优选的大鼠 CPB 酶氨基酸序列, 如 SEQ ID NO.18 所示:

ASGHSYTKYNKWETIEAWIQVATDNPDLVTQSVIGTTFEGRNMYVLKIGKTRPNKPAIFI
DCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRM
WRKTRSTMAGSSCLGVRPNRNFNAGWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNL
STIKAYLTIHSYSQMMLYPYSYDYKLPEYEEELNALVKGAAKELATLHGTYKYTYGPGATTIYP
AAGGSDDWSYDQGIKYSFTFELRDTGFFGFLLPESQIRQTCEETMLAVKYIANVYVREHLY

本实施例中, 所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域优选酿酒酵母(*Saccharomyces cerevisiae*)Kex2 酶全长氨基酸序列(Uniprot 数据库 ID 为 P13134 蛋白质, <https://www.uniprot.org/uniprotkb/P13134>) 中第 679-814 位氨基酸序列, 即位于羧基端有内膜定位及各细胞器之间的转换与运输作用的区域, 序列如 SEQ ID NO.19 所示:

YFLTIFLIGATFLVLYFMFFMKSRRRIRRSRAETYEFDIIDTDSEYDSTLDNGTSGITEPEEVE
DFDFDLSDHLASLSSSENGDAEHTIDSVLTNENPFSDPIKQKFPNDANAESASNKLQELQPD
VPPSSGRS

本实施例中, 所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域还优选毕赤酵母(*Komagataella phaffii*, 即 *Pichia pastoris*) Kex2 酶全长氨基酸序列(Uniprot 数据库 ID 为 C4R095 蛋白质, <https://www.uniprot.org/uniprotkb/C4R095>) 中第 681-777 位氨基酸序列, 即位于羧基端有内膜定位及各细胞器之间的转换与运输作用的区域, 如 SEQ ID NO.20 所示:

YLAFLGLGLFLICIIFLFTNRNKLERRQRRNRDEYFDLIPADDDFDTEEDQEANSQFTLD
SDAELMFEDTSQREASPHEYQDSLGSNEHPKRAAL

作为 CPB 酶与 Kex2 酶羧基端序列融合表达时的连接, 使用 Linker 序列, 如序列 SEQ ID NO.21 所示: GSGSGSGGS

本发明构建的 CPB 定位融合功能蛋白如下:

本发明构建的融合表达的人 CPB 酶与毕赤酵母 Kex2 酶羧基端序列的 CPB 定位融合功能蛋白, 命名为 HCPB-PPKEX2, 其序列如 SEQ ID NO.22 所示:

ATGHSYKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKVGKAGQNKPAIFM

acactgggtcctcttgcacggaacagacccaaccgtaatttcgatgcaggctggcgagataggggctagtcgtaatccgtgcgacgagacgt
actgcgggcccgcgctgaatcagaaaaggaaactaaggcgcttgcggatttcattagaaataactttccagtataaagcgatttgaccatacac
tcgtatagccaaatgatgatctaccctactcctacgcgtataaacttggtagaaacaacgccgaactgaatgccctcgccaaggcaactgtgaagga
gctagcaagcttgcacggaactaaatacacttacgggccttggccacaaccatataccggctgccgggggtagtgacgattgggcttatgatcag
gggtataagggtatagctttacgttcgagctgcgtgatacgggccgttatggattcttgcctcctgagtcctcaattagggtacgtgaagaaacattct
ggctattaaatatgttgcaccttacgtattagaacatttatacggaggctcgggatctgggtcgggtggctcatattcctcacaatattcttaataggcgc
aacgttttttagtgcgtgtttttatgttttcatgaaatcaaggcggagaatccggcggttcacgcgggagacctacgagttcgacatcatagatactgact
cggataacgattcaacctggataacgggacctcgggtattactgaaccggaggaggtagaggatttcgatttcgacctctctgacgaagatcatcta
gccagtctatcgctccagtgaatggagatgctgaacacacaatcgacagcgctcttacgaatgagaacctgtttccgacatctattaaacaaaagttt
cccaacgacgcgaatgcagagagcgccagtaacaagttgcaggaattgcagccagacgttccgccatcctctggctgatcttaa

融合表达的大鼠 CPB 酶与毕赤酵母 Kex2 酶羧基端序列的 CPB 定位融合功能蛋白，命名为 RCPB-PPKEX2，其序列如 SEQ ID NO.26 所示：

ASGHSYTKYNKWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKIGKTRPNKPAIFI
DCGFHAREWISPAFCQWVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRM
WRKTRSTMAGSSCLGVRPNRNFNAGWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNL
STIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGA AKELATLHGTYKYTYGPGATTIYP
AAGGSDDWSYDQGIKYSFTFELRDTGFFGFLLPESQIRQTCEETMLAVKYIANVYVREHLYGGS
GSGSGGSYLAFLGLGLFLICIFLFTNRNKLERRQRRNRDEYEFDLIPADDDFDTEEDQEANSQ
FTLSDAELMFEDTSQREASPHEYQDSLGSNEHPKRAAL

编码 SEQ ID NO.26 的 DNA 序列，如 SEQ ID NO.27 所示：

gcaagcggctcactcgtacacaaagtataacaaatgggaaactatcgaggcctggattcaacaagtagccacggataacctgatttagtcacg
caatcgggtcatagggacaaccttcgagggcagaacatgtacgtattaaagataggaaagactcgtccaataaacctgctatttttatcgactgtggc
ttccacgctagagaatggattagccctgcctctgccaatgttcgtgagggagcagtagcgacgtacaaccaggaaattcatatgaagcaactgct
cgacgaattagatttctatgtttaccggtagttaacatcgtggctacgtttacacatggacaaaagaccgtatgtggcgggaagactagatcgacgatg
gcaggctcatcctgtctgggtgtgcgtccgaataggaactcaatcggggttgggtgtgaagtggggcgctccgcgtctcctgttcagagacatactg
cggacctgccccagaatccgaaaaagagacgaagctcttgcgtattttattaggaacaacctaaagtaccattaaagcatacttaacctcatagtta
cagccagatgatgctgtatccctattctacgactataagctaccggagaactatgaagaactcaatgctttagttaaaggagcggccaaggagttgg
cgacattgcatggaacgaaglacacttatggccaggggccactaccatctatccggcagctgggtgatcagacgactggagttatgaccaaggtat
aaagtactctttcaccttgaattacgtgatactgggtttttggcttctgctccagagtcacagatccgccagacttgcaggaaacctgcttgcggt
gaagtatatagctaattacgtccgtgaacatctatacggcggtaccgggtcggggagtggtgggagttatctagcattttgctagggctcggcttct
gatctgcataatattctgtttaccaaccgcaataaattggaacgaaggcagcgacgaaatcgccgagacgaatatgatttgacctcattccggcgg
atgatgattttgatacggaggaggaccaggaggccaatagccaattcacattggactctgacgcggaacttatgttcgaagatagtcacagagaga
ggcctcgccccacgagatcaagattctctgggatctaataacacccaaaacgggctgcactttaa

融合表达的大鼠 CPB 酶与酿酒酵母 Kex2 酶羧基端序列的 CPB 定位融合功能蛋白，命名为

RCPB-SCKEX2, 其序列如 SEQ ID NO.28 所示:

ASGHSYTKYNKWETIEAWIQVATDNPDLVTQSVIGTTFEGRNMYVLKIGKTRPNKPAIFI
DCGFHAREWISPAFCQWVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRM
WRKTRSTMAGSSCLGVRPNRNFNAGWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNL
STIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGA AKELATLHGTYKYTYGPGATTIYP
AAGGSDDWSYDQGIKYSFTFELRDTGFFGFLLPESQIRQTCEETMLAVKYIANIYVREHLYGGS
GSGSGGSYFLTIFLIGATFLVLYFMFFMKSRRRIRRSRAETYEFDIIDTDSEYDSTLDNGTSGITEP
EEVEDFDLSDHLASLSSSENGDAEHTIDSVLTNENPFSDPIKQKFPNDANAESASNKLQE
LQPDVPPSSGRS

编码 SEQ ID NO.28 的 DNA 序列, 如 SEQ ID NO.29 所示:

gcaagcggctcactcgtacacaaagtataacaaatgggaaactatcgaggcctggattcaacaagtagccacggataaccctgatttagtcacg
caatcggctcatagggacaaccttcgagggcagaacatgtacgtattaaagataggaaagactcgtcccaataaacctgctattttatcgactgtggc
ttccacgctagagaatggattagccctgccttctgccaatggctcgtgagggagcagtagcgacgtacaaccaggaaattcatatgaagcaactgct
cgacgaattagattctatgtttaccgtagttaacatcgatggctacgtttacacatggacaaaagaccgtatgtggcggaagactagatcgacgatg
gcaggctcatcctgtctgggtgtgcgtccgaataggaacttcaatgcgggttggtgtgaagtggggcgctcccgctctccctgttcagagacatactg
cggacctgccccagaatccgaaaagagacgaaagctcttctgctgattttattaggaacaacctaagtaccattaaagcatacttaaccatccatagtta
cagccagatgatgctgtatccctattctacgactataagctaccggagaactatgaagaactcaatgctttagttaaaggagcggccaaggagttgg
cgacattgcatggaacgaagtacacttatggccaggggcccactaccatctatccggcagctggtggatcagacgactggagttatgaccaaggtat
aaagtactctttcaccttgaattacgtgatactggtttttggctttctgctccagagtcacagatccgccagacttgcgaggaaacatgcttgcggt
gaagtatatagctaattacgtccgtgaacatctatacggcgatccgggtcggggagtggtgggagttatttctcacaatattcttaataggcgcaac
gtttttagtgctgtattttatgttttcatgaaatcaaggcgggagaatccggcgttcacgcgcggagacctacgagttcgacatcatagatactgactcgg
aatacattcaaccctggataacgggacctcgggtattactgaacccgaggagtagaggatttcgatttcgacctctctgacgaagatcatctagcc
agtctatcgtccagtgaatggagatgctgaacacacaatcgacagcgtccttacgaatgagaacccgtttccgacatctattaacaaaagtgtccc
aacgacgcgaatgcagagagcgccagtaacaagttgcaggaattgcagccagacgttcgccatcctctgtgctgatcttaa

在本发明的定位融合功能蛋白设计中, 优选的 Kex2 酶, 是酵母类微生物(自然包括毕赤酵母)自身表达的一种钙离子依赖型的丝氨酸蛋白酶, 能特异性识别和切割氨基酸序列中的 RR、KR 等双碱性氨基酸羧基端肽键, 在酵母的蛋白质分泌途径中起关键作用。同时, 酵母类微生物有 STE 13 基因, 于胞内会表达 Ste 13 蛋白酶(严格来说, 属于 Dipeptidyl aminopeptidase 二肽氨肽酶), 可以切割蛋白质氨基末端的 EA、EAEA 的氨基酸序列。

Kex2 酶在基因工程中应用最多的是在酵母外源分泌途径中, 切割外源蛋白前体中的信号肽或前肽序列, 释放成熟态的分泌蛋白, Kex2 酶会在外源蛋白 KR 或 RR 双碱性氨基酸羧基端将肽键切割, 同时为了提高切割效率, KR 或 RR 之后往往会添加 EA、EAEA 等氨基酸序列(非必须, 也可以是外源蛋白的序列, 但切割效率会根据 KR、RR 之后 1-4 个氨基酸残基种类有所不同)。另外 Kex2 酶还可作为融合蛋白之间的酶切位点将表达的一整条肽段切割为两条或多条

肽段，这在表达不同亚基的蛋白质情形下使用比较多。相应的，这样的应用也可以迁移到蛋白质多序列重复串联表达时，于其中添加 Kex2 酶切位点，可达到间接增长拷贝数的目的，以实现外源蛋白表达量的提高。但是，如上的 Kex2 酶作用时，KR、RR 等双碱基氨基酸是有效切割的前提条件，但是切割后 KR、RR 无法去除，为了切割效率的提高，KR、RR 之前往往也要添加 1-2 个氨基酸残基（非必须，可以根据实际情况进行调节、设计），对于重组表达的蛋白质而言，引入的非自身氨基酸会更多。而胶原蛋白，其典型的序列结构是 G-X-Y 不断重复的三联体结构，非自身氨基酸的引入往往会破坏这样的结构特征。

在本发明的定位融合蛋白设计过程中，需要考虑 KR、RR 的去除。

重组羧肽酶 B (CPB 酶，一种外分泌蛋白酶) 能特异性切割蛋白质羧基末端的碱性氨基酸（尤其是 K、R），直至 C 端的碱性氨基酸切完，其它非碱性氨基酸裸露于蛋白质的 C 端。在生物工程或生物制药领域一般使用 CPB 酶，基本上都是将重组蛋白质或蛋白质药物表达、纯化后，额外添加 CPB 酶以去除蛋白质的 C 端的 K、R 等碱性氨基酸，这相当于在重组蛋白的表达纯化工作中又加了一种外源性物质，除去购买 CPB 酶的成本增加外，还要添加 CPB 去除、残留量检测的工艺流程，必要的情况下还要对 CPB 酶残留进行相应的评估，整个工艺的难度、流程、成本、时间也会在增加，还会有可能 CPB 酶去除不净而残留的隐患。

所以本发明也需要在细胞内尤其是蛋白质表达分泌途径中，建立 CPB 酶的合成代谢途径，使 CPB 酶定位于胞内（内质网腔、高尔基体（trans-Golgi network, TGN）等构成的蛋白质表达分泌途径）中，在 Kex2 酶将经过设计的长序列重组胶原蛋白氨基酸切割为小片段小分子量的蛋白质后，CPB 酶可以和 Ste 13 蛋白酶协同分别把 C 端、N 端的 KR、EA 等非胶原氨基酸切除，保持胶原蛋白 G-X-Y 三联体重复结构。

综上，本发明需要解决 CPB 酶不仅仅是在细胞内表达，更要设法让它能够进入分泌表达途径（其方式与分泌表达完全一致）却仍可以保留于由内质网腔、高尔基体等蛋白质分泌途径之中而不分泌于胞外，且能与 Kex2 酶的活化时间保持时间、空间上的持续性。

所以，本发明将 CPB 酶与一段有内膜定位或细胞中分泌细胞器之间转换与运输作用的序列融合表达，形成一个全新的 CPB 定位融合功能蛋白，能分泌、能定位于细胞内膜上、也能伴随 Kex2 酶与重组胶原蛋白在胞内分泌细胞器之间转换与运输、不分泌（定位细胞内膜上无法脱离）。

Kex2 酶整体上包括信号肽、前体肽、催化结构域、P 结构域、Ser/Thr 富集结构域、跨膜区以及胞外结构域七部分，其酶活性主要由催化结构域、P 结构域、Ser/Thr 富集结构域实现，保留这三部分仍能保持 Kex2 酶的活性，而 C 端的膜区以及胞外结构域主要负责 Kex2 蛋白酶在细胞内各细胞器之间的转换与运输，本发明选择了 Kex2 酶 C 端的相关功能区域与 CPB 酶融合表达，为了不影响融合两部分的结构与活性，于中间添加了具有柔性的 Linker 序列（非必须）。

而且相同的 C 端序列, 会将 Kex2 酶与的 CPB 定位融合功能蛋白集中定位于同一区域, Kex2 酶的作用产物则是 CPB 定位融合功能蛋白的作用底物, 集中于同一区域更有利于两种酶的协同作用。

(2) CPB 定位融合功能蛋白的表达载体及底盘细胞的构建与鉴定

CPB 底盘细胞需要将 CPB 定位融合功能蛋白表达且定位于内质网、高尔基体等构成的细胞蛋白质分泌途径之中, 与酵母自身的 Kex2、Ste13 等蛋白酶进行协同作用, 因此构建 CPB 定位融合功能蛋白的表达载体时, 需要于其前添加信号肽序列以使其得以分泌进行内质网及后续的蛋白质分泌途径中, 而能进入内质网腔的蛋白质所使用的信号肽都有类似的作用, 本发明以常用的 α -factor secretion signal 作为信号肽序列进行示例。

毕赤酵母能使用的表达载体众多, 载体中含有的启动子作用很重要, 常用的启动子有: (1) AOX1 启动子、FLD1 启动子、FDH1 启动子、DAS1 启动子 和 DAS2 启动子等诱导型启动子, 使用甲醇或甲胺才可诱导启动蛋白质的表达, 可以采用诱导剂的添加调节 CPB 定位融合功能蛋白转录表达的时机; (2) GAP、GCW14 等不依赖诱导剂的组成型启动子, 只要细胞处于生存状态, 它们就会启动蛋白质的转录表达, 即 CPB 定位融合功能蛋白是一直转录表达的; (3) THI4 启动子、THI11 等不需要特定的诱导剂的启动子, 只要减少部分培养物质 (如减少硫胺素) 即可启动蛋白质表达, 在一定程度上也可调节 CPB 定位融合功能蛋白转录表达的时机; (4) 调节培养环境其它条件的启动子, 如 HSP82 启动子, 是一种热诱导启动子, 在菌株正常生长时, 更改生长温度, 进行热刺激, 便可以诱导 HSP82 启动子进行转录启动, 从调节 CPB 定位融合功能蛋白转录表达。但无论哪种启动子, 最终的作用是一样的, 即启动 CPB 定位融合功能蛋白转录表达, 最终作用并无区别, 均在本发明保护的范围内, 本发明以常见的 AOX1 启动子作为示例。毕赤酵母或酵母类的表达载体众多, 最终目的只要是完成 CPB 定位融合功能蛋白的表达, 其作用都是相同的, 本发明使用常用的表达载体构建了多种表达载体。

DNA 序列 SEQ ID NO.23、SEQ ID NO.25、SEQ ID NO.27、SEQ ID NO.29 均由南京金斯瑞生物科技有限公司合成。将合成的基因分别克隆于 pPICZ α B、pFLD α 、pPIC9K 等表达载体 (均购自赛默飞世尔科技 (中国) 有限公司) 上, 克隆位置为 α -factor secretion signal 序列中 kex2 酶切位点序列 (其末端 DNA 序列为 AAAGAAGAAGGGGTATCTCTCGAGAAAAGA) 与载体上终止密码子 “TGA” 之间。可获得 3 组构建底盘微生物所需要重组表达载体:

(1) pPICZ α B-HCPB-PPKEX2、pPICZ α B-HCPB-SCKEX2、pPICZ α B-RCPB-PPKEX2、pPICZ α B-RCPB-SCKEX2, 使用 AOX1 启动子和 α -factor secretion signal 信号肽序列;

(2) pFLD α -HCPB-PPKEX2、pFLD α -HCPB-SCKEX2、pFLD α -RCPB-PPKEX2、pFLD α -RCPB-SCKEX2, 用 FLD 启动子和 α -factor secretion signal 信号肽序列;

(3) pPIC9K-HCPB-PPKEX2、pPIC9K-HCPB-SCKEX2、pPIC9K-RCPB-PPKEX2、pPIC9K-RCPB-SCKEX2, 使用 AOX1 启动子和 α -factor secretion signal 信号肽序列。

这三组重组表达载体中每个均可单独使用, 且均使用 α -factor secretion signal 信号肽序列, 虽然启动子有所不同, 但均可使用甲醇进行诱导 (FLD 启动子也可使用甲胺), 其最终效果是相同的, 本发明示例中使用第一组作为后续工程菌株构建的示例, 如图 11-图 14 所示, 为该实施例构建的其中 4 种重组表达载体。

底盘细胞使用的空白出发菌株可以是原始毕赤酵母菌株或商品化的 X-33、SMD1168、GS115 等菌株, 或者其它的酵母菌, 最终达的效果是相同的, 即于细胞蛋白质分泌途径中构建了 CPB 定位融合功能蛋白的表达, 并与酵母自身的 Kex2、Ste13 等蛋白酶进行协同作用, 将长的蛋白质进行表达并分割, 本发明使用了商品化空白出发菌株 GS115 作为示例。

将 pPICZ α B-HCPB-PPKEX2、pPICZ α B-HCPB-SCKEX2、pPICZ α B-RCPB-PPKEX2、pPICZ α B-RCPB-SCKEX2 载体质粒用 Pme I (购自大连 TaKaRa 公司) 37°C 酶切消化过夜, 使其线性化, 再使用 PCR 产物纯化试剂盒 (购自生工生物工程(上海)股份有限公司) 回收线性化质粒, 使体积控制在 10 μ L 左右。将线性化质粒电转化入空宿主菌种毕赤酵母 GS115 (购自中国工业微生物菌种保藏管理中心) 感受态细胞中, 涂至 Zeocin (购自赛默飞世尔科技(中国)有限公司) 为 0.1mg/mL 的 YPD 平板中, 待单菌落长出后, 用牙签挑取单克隆接种至 96 孔板中, 于 30°C 培养 48h 后, 分别点板于含 0.5mg/mL、1mg/mL Zeocin 的 YPD 平板, 挑选于高浓度 (1mg/mL) Zeocin 平板上生长的菌株接入 BMGY 培养基摇瓶中, 30°C、220rpm 过夜培养后, 更换为 BMMY 培养基进行诱导表达, 每 24h 补充 100% 甲醇至终浓度为 1.0%, 培养 36h 以上, 收集培养上清和菌体, 用 8M 尿素裂解菌体, 制备样品后进行 SDS-PAGE 检测, 使用抗人或鼠 CPB 酶的抗体 (购自生工生物工程(上海)股份有限公司), 以重组人或大鼠 CPB 酶蛋白 (购自生工生物工程(上海)股份有限公司) 作为阳性对照, 空宿主菌的培养上清为阴性对照, 使用 ECL 化学发光显色, 全自动化学发光图像分析系统 (Tanon 5200) 进行结果检测合成图像。

结果如图 15 所示, 图 15 为 CPB 定位融合功能蛋白 HCPB-PPKEX2、HCPB-SCKEX2、RCPB-PPKEX2、RCPB-SCKEX2 在胞内裂解液和培养上清中表达的检测结果; 左图中泳道, 1: HCPB-SCKEX2 菌体裂解液; 2: HCPB-SCKEX2 菌培养上清; 3、6: HCPB-PPKEX2 培养上清; 4、5: HCPB-PPKEX2 菌体裂解液; (-): 阴性对照; (+): 阳性对照, 重组人 CPB 酶; /: 无样品孔。右图中泳道, (-): 阴性对照; (+): 阳性对照, 重组大鼠 CPB 酶; /: 无样品孔; 7: RCPB-SCKEX2 菌培养上清; 8: RCPB-SCKEX2 菌体裂解液; 9、10、12、13: RCPB-PPKEX2 培养上清; 11、14: RCPB-PPKEX2 菌体裂解液。

作为阳性对照的人 CPB 酶成功检测, 而 HCPB-PPKEX2、HCPB-SCKEX2、RCPB-PPKEX2、

RCPB-SCKEX2, 均只于胞内裂解液中检测出, 且条带均大于阳性对照, 条带大小符合融合蛋白的预期 (融合其它蛋白, 电泳表观分子量自然就更大)。最主要的是, 培养上清中没检测出相应的条带, 这说明作为融合蛋白表达的 HCPB-PPKEX2、HCPB-SCKEX2、RCPB-PPKEX2、RCPB-SCKEX2 均未分泌于胞外, 而是都位于胞内, 这与 4 种 CPB 定位融合功能蛋白设计目的相符, 即人或大鼠的 CPB 酶与来源于 Kex2 酶羧基端有内膜定位及各细胞器之间的转换与运输作用区域序列融合表达, 在信号肽的作用下, 它们能够进入内质网腔及高尔基体等构成的细胞内蛋白质分泌途径之中, 但与 Kex2 酶一样, 也只能存于细胞内, 而无法分泌出细胞外。

参照 Invitrogen 的操作手册以不同浓度的 Zeocin 抗生素的 YPD 平板梯度, 点板筛选的方法, 结合胞内 WB 检测结果, 陆续筛选出 HCPB-PPKEX2、HCPB-SCKEX2、RCPB-PPKEX2、RCPB-SCKEX2 这 4 种融合功能蛋白合成代谢途径的底盘细胞的工程菌株, 作为后续进行重组串联表达策略的宿主菌。

构建的 CPB 定位融合功能蛋白功能途径的 4 种底盘细胞的工程菌株样本均送至中国微生物菌种保藏管理委员会普通微生物中心保藏, 菌种保藏编号分别对应的是:

表达 HCPB-PPKEX2 的菌株, 保藏编号为: CGMCC No.25815;

表达 RCPB-PPKEX2 的菌株, 保藏编号为: CGMCC No.25817;

表达 HCPB-SCKEX2 的菌株, 保藏编号为: CGMCC No.25816;

表达 RCPB-SCKEX2 的菌株, 保藏编号为: CGMCC No.25818;

上述工程菌保藏地址均为: 北京市朝阳区北辰西路 1 号院 3 号; 保藏日期均为: 2022 年 9 月 26 日。分类命名均为: 巴斯德毕赤酵母 *Komagataella phaffii*。

实施例 3: 小分子胶原蛋白氨基酸序列设计与串联重复表达的小分子胶原蛋白的表征

为达到本发明所述的小分子胶原蛋白的高效表达的目的, 本发明中对目标小分子胶原蛋白的氨基酸序列进行了设计, 设计要求:

(1) 重复串联的小分子胶原蛋白单体 (或片段) 之间设计酵母 Kex2 酶能够有效识别和切割的位点, 其突出特征是有 KR、RR 等双碱性氨基酸。

(2) 于双碱性氨基酸残基之后, 可以是 EA、EAEA 或其它有助于 Kex2 酶识别和切割的氨基酸序列, 比如 D 或 A 等氨基酸序列, 或直接是小分子胶原蛋白蛋白单体 (或片段) 自身的氨基酸 (发明中有相关的示例)。

(3) 重复串联的小分子胶原蛋白单体 (或片段) 的氨基酸序列可以是任一序列, 其氨基酸残基的具体排序、序列长度、单体 (或片段) 串联重复的次数都是可以自由安排、不受限制的, 当然, 作为小分子胶原蛋白, 一般需要有典型的 G-X-Y 三联体结构, 且氨基酸序列长度控制在 100 个氨基酸以内。

所示:

EAGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除,羧基端的 KR 两个氨基酸会被 CPB 酶切割去除,最终分泌获得的是一段 63 个(记为 17S28),且氨基酸序列均为典型的 G-X-Y 三联体胶原蛋白序列,其序列如 SEQ ID NO.35 所示:

GVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLE

(3) 参考 17S1NNT 的氨基酸序列,设计重复串联序列,重复 6 次,命名为 17S1N6-6,共 540 个氨基酸,其序列如 SEQ ID NO. 36 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKR

当 17S1N6-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时,Kex2 酶会在每个 KREA 中的 KR 与 EA 之间,将其分割为相同的小分子蛋白序列,其序列如 SEQ ID NO.37 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLEKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除,羧基端的 KR 两个氨基酸会被 CPB 酶切割去除,最终分泌获得的是一段 86 个(记为 17S1N6),且氨基酸序列均为典型的 G-X-Y 三联体胶原蛋白序列,其序列如 SEQ ID NO.38 所示:

GPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLE

(4) 参考 17S1NNT 的氨基酸序列,设计重复串联序列,重复 7 次,命名为 17S1N7-7,共 528 个氨基酸,其序列如 SEQ ID NO.39 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGKREAAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKG

SSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGKREAGPPGSG
EKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPM
GPPGPKGDQGEKGPRGKREAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQG
PPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGKREAGPPGSGEKGERGAAG
EPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQ
GEKGPRGKREAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQG
LRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGKR

当 17S1N7-7, 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KREA 中的 KR 与 EA 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.40 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGPRGKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除, 羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 84 个 (记为 17S1N7), 其序列如 SEQ ID NO.41 所示:

GPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGPRG

(5) 参考 17S1NNT 的氨基酸序列, 设计重复串联序列, 重复 7 次, 命名为 17S1NK-7, 共 595 个氨基酸, 其序列如 SEQ ID NO.42 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGKREAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSG
SPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGKREAGPPGSGEKGERG
AAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPK
GDQGEKGKREAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQG
LRGEVGLPGVKGDKGPMGPPGPKGDQGEKGKREAGPPGSGEKGERGAAGEPGPHGPPGVPG
SVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGKREAGPP
GSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPGVKGDK
GPMGPPGPKGDQGEKGKREAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQG
PPGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGKR

当 17S1NK-7 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KREA 中的 KR 与 EA 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.43 所示:

EAGPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSSGSPGPQGPPGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除, 羧基端的 KR 两个氨基酸会被

CPB 酶切割去除，最终分泌获得的是一段 84 个氨基酸的重组小分子胶原蛋白（记为 17S1NK），其序列如 SEQ ID NO.44 所示：

GPPGSGEKGERGAAGEPGPHGPPGVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPG
VKGDKGPMGPPGPKGDQGEKGPRG

（6）参考 3A5D1 的氨基酸序列，设计重复串联序列，重复 5 次，命名为 3A5D15D-5，共 510 个氨基酸，其序列如 SEQ ID NO.45 所示：

DGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLEKRDDGNPGSDGLPGRDGSPGGKGD
RGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRGDKG
ETGERGAAGIKGHRGLEKRDDGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPV
GPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLEKRDGN
PGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGP
AGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLEKRDDGNPGSDGLPGRDGSPGGKGDRGEN
GSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRGDKGETGER
GAAGIKGHRGLEKR

当 3A5D15D-5 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时，Kex2 酶会在每个 KRD 中的 KR 与 D 之间，将其分割为相同的小分子蛋白序列，其序列如 SEQ ID NO.46 所示：

DGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLEKR

羧基端的 KR 两个氨基酸会被 CPB 酶切割去除，最终分泌获得的是一段 100 个氨基酸的重组小分子胶原蛋白（记为 3A5D15D），为典型的 G-X-Y 三联体胶原蛋白序列，其序列如 SEQ ID NO.47 所示：

DGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAG
PAGAPGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLE

（7）同样的，使用其它的酶切位点也可以达到与 3A5D15D 相同的作用，将酶切位点区域序列设计为 KREA，设计重复串联序列，重复 5 次，命名为 3A5D15E-5，共 520 个氨基酸，其序列如 SEQ ID NO.48 所示：

EADGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGP
AGAPGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRGLEKREADGNPGSDGLPGRDGSPG
GKGDRGENSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGR
GDKGETGERGAAGIKGHRGLEKREADGNPGSDGLPGRDGSPGGKGDRGENSPGAPGAPGH
PGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRGDKGETGERGAAGIKGHRG

LEKREADGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESG
PAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLEKREADGNPGSDGLPGRD
GSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGP
QGPRGDKGETGERGAAGIKGHRGLEKR

当 3A5D15E-5 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KRD 中的 KR 与 EA 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.49 所示:

EADGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGP
AGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLEKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除, 羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 100 个氨基酸的重组小分子胶原蛋白(与 3A5D15D 表达切割后获得的序列相同, 这里记为 3A5D15E), 为典型的 G-X-Y 三联体胶原蛋白序列, 其序列如 SEQ ID NO.50 所示:

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLE

(8) 参考 3A5D1NT 的氨基酸序列, 设计重复串联序列, 重复 6 次, 命名为 3A5D15R-6, 共 570 个氨基酸, 其序列如 SEQ ID NO.51 所示:

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIKRDGNPGSDGLPGRDGSPGGKGDRGENGSP
GAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGA
AGIKRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGP
AGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKRDGNPGSDGLPGRDGSPGGKGDRG
ENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGET
GERGAAGIKRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDR
GESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKRDGNPGSDGLPGRDGSPGGK
GDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGD
KGETGERGAAGIKR

当 3A5D15R-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KRD 中的 KR 与 D 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.52 所示:

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAG
PAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKR

羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 93 个氨基酸的重组

小分子胶原蛋白（记为 3A5D15R），其序列如 SEQ ID NO.53 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGI

（9）参考 3A5D1NT 的氨基酸序列，设计重复串联序列，重复 6 次，命名为 3A5D15KR-6，共 582 个氨基酸，其序列如 SEQ ID NO.54 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIKGKRDDGNPGSDGLPGRDGSPGGKGDRGENG
SPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERG
AAGIKGKRDDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRG
ESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGKRDDGNPGSDGLPGRDGSPGG
KGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRG
DKGETGERGAAGIKGKRDDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGP
AGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGKRDDGNPGSDGLP
GRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGA
PGPQGPRGDKGETGERGAAGIKGKR

当 3A5D15KR-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时，Kex2 酶会在每个 KRD 中的 KR 与 D 之间，将其分割为相同的小分子蛋白序列，其序列如 SEQ ID NO.55 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIKGKR

羧基端的 KR 两个氨基酸会被 CPB 酶切割去除，最终分泌获得的是一段 95 个氨基酸的重组小分子胶原蛋白（记为 3A5D15KR），其序列如 SEQ ID NO.56 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIKG

（10）参考 3A5D1NT 的氨基酸序列，设计重复串联序列，重复 6 次，命名为 3A5D15EKR-6，共 576 个氨基酸，其序列如 SEQ ID NO.57 所示：

DGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPRGDKGETGERGAAGIEEKRDDGNPGSDGLPGRDGSPGGKGDRGENGSP
GAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGA
AGIEEKRDDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESG
PAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIEEKRDDGNPGSDGLPGRDGSPGGKGDR
GENGSPGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGE
TGERGAAGIEEKRDDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSG
DRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIEEKRDDGNPGSDGLPGRDGSP

GGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAGPAGPAGSRGAPGPQGP
RGDKGETGERGAAGIEKR

当 3A5D15EKR-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KRD 中的 KR 与 D 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.58 所示:

DGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRDKGETGERGAAGIEKR

羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 94 个氨基酸的重组小分子胶原蛋白 (记为 3A5D15EKR), 为典型的 G-X-Y 三联体胶原蛋白序列, 其序列如 SEQ ID NO.59 所示:

DGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRDKGETGERGAAGIE

(11) 参考 3A5D1NT 的氨基酸序列, 设计重复串联序列, 重复 6 次, 命名为 3A5D15G-6, 共 576 个氨基酸, 其序列如 SEQ ID NO.60 所示:

DGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRDKGETGERGAAGIGKRDDGNPGSDGLPGRDGSPGGKGDRGENSG
PGAPGAPGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRDKGETGERG
AAGIGKRDDGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGES
GPAGPAGAPGPAGSRGAPGPQGPGRDKGETGERGAAGIGKRDDGNPGSDGLPGRDGSPGGKG
DRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRDK
GETGERGAAGIGKRDDGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGK
SGDRGESGPAGPAGAPGPAGSRGAPGPQGPGRDKGETGERGAAGIGKRDDGNPGSDGLPGRDG
SPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQ
GPGRDKGETGERGAAGIGKR

当 3A5D15G-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 KRD 中的 KR 与 D 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.61 所示:

DGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAG
APGPAGSRGAPGPQGPGRDKGETGERGAAGIGKR

羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 94 个氨基酸的重组小分子胶原蛋白 (记为 3A5D15G), 为典型的 G-X-Y 三联体胶原蛋白序列, 其序列如 SEQ ID NO.62 所示:

DGNPGSDGLPGRDGSPGGKGDRGENSGPAGPAGHPGPPGPVGPAGKSGDRGESGPAGPAG

APGPAGSRGAPGPQGPRGDKGETGERGAAGIG

除去以上所描述的以人III型胶原蛋白序列或人 XVII 胶原蛋白序列中一个区域为基础进行重复串联序列的设计外，以人胶原蛋白序列的两个或多个区域的组合拼接的序列作为单体，进行串联重复，也可以应用于本发明所述重复串联表达体系（包括重复串联表达序列设计方法与底盘细胞的使用）和单体表达体系。

(12) 参考人III型胶原蛋白序列（参照 Uniprot 数据库中 ID 为 P02461 的蛋白质序列，<https://www.uniprot.org/uniprotkb/P02461>）的几个不同功能区域，第 880-915 位氨基酸序列、第 949-966 位氨基酸序列、第 1012-1038 位氨基酸序列、第 1060-1074 位氨基酸序列组合设计重组小分子胶原蛋白，共 96 个氨基酸，可用于串联的单体，其序列如 SEQ ID NO.63 所示，
GPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGET

以此为基础设计重复串联序列，重复 6 次，命名为 3A5D16M-6，共 500 个氨基酸，其序列如 SEQ ID NO.64 所示：

EAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKR

当 3A5D16M-6 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时，Kex2 酶会在每个 **KREA** 中的 KR 与 EA 之间，将其分割为相同的小分子蛋白序列，其序列如 SEQ ID NO.65 所示：

EAGPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除，羧基端的 KR 两个氨基酸会被 CPB 酶切割去除，最终分泌获得的是一段 96 个氨基酸的重组小分子胶原蛋白，命名为 3A5D16M，其序列如 SEQ ID NO.66 所示：

GPPGPAGNTGAPGSPGVSGPKGDAGQPGEKGGSPAQGPQGVKGESGKPGANGLSGENGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLE

(13) 参考人III型胶原蛋白序列（参照 Uniprot 数据库中 ID 为 P02461 的蛋白质序列，

<https://www.uniprot.org/uniprotkb/P02461>) 两个功能区域, 即第 1060-1074 位氨基酸序列与 1015-1038 位氨基酸序列组合, 共 39 个氨基酸, 可用于单体表达, 其序列如 SEQ ID NO.67 所示,

GSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGET

以此为基础设计重复串联序列, 重复 12 次, 命名为 3A5D15M-12, 共 516 个氨基酸, 其序列如 SEQ ID NO.68 所示:

EAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGP
PGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGP
RDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAP
GHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGKSGAPG
PQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPG
APGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGK
SGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKRE
AGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKREAGSPGAPGAPGHPGPPGPV
GPAGKSGAPGPQGPRGDKGLEKR

当 3A5D15M-12 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会在每个 **KREA** 中的 KR 与 EA 之间, 将其分割为相同的小分子蛋白序列, 其序列如 SEQ ID NO.69 所示:

EAGSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLEKR

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除, 羧基端的 KR 两个氨基酸会被 CPB 酶切割去除, 最终分泌获得的是一段 39 个氨基酸的重组小分子胶原蛋白, 命名为 3A5D15M, 其序列如 SEQ ID NO.70 所示:

GSPGAPGAPGHPGPPGPVGPAGKSGAPGPQGPRGDKGLE

人工设计的非天然胶原蛋白序列理论上也可以应用本发明的重复串联表达体系, 原理如上, 在此不再赘述。

重复串联表达体系重组表达载体的构建与表达效果的验证:

本实施例中基于底盘细胞和上述氨基酸序列的设计, 重组小分子胶原蛋白的使用的表达载体并不受限制, 只要所选用的表达载体有信号肽序列, 能将翻译表达的蛋白质导入内质网即可, 常用的表达载体如 pPICZαB、pFLDα、pPIC9K 均可, 其它基于有类似作用的表达载体的构建也在本发明的保护范围内。同样, 编码相应重组小分子胶原蛋白的 DNA 序列可以进行优化, 只要最终编码氨基酸序列相同, 则于本发明保护范围内。

编码 3A5D29N-9 的 DNA 序列, 如 SEQ ID NO.71 所示:

gaagctggaagtcggcgaccggggagagtcaggcgctgcaggcgctgctggtgccccaggaccggcgccagtagaggtgctcctg
gtcctcaaggaccacggggcgacaaagtgaaacaggtgaacgaggtgcggcaggtataaaagggcacagagggtagaaaagcggaagc

aggaaaaagtggcgatcgcggggaatcaggtcccgcggggccggcgggtgcgccggaccggcggggtccagagggtcccctggaccgca
gggacctcggggtgacaaagggtgaaactggggaaaagggggctgcgggcataaaggacaccgcggactcgagaagagagaggcaggtaaa
agtggggatagaggcgaatctggccggcaggaccagcgggcgccccggctctgctggtacacgaggtgccccgggtcgcgaaggctctagg
ggtgacaaaggcgagacgggcgaacgaggggcagcaggaattaaaggccaccgagggctggagaagcgggaagcaggtaaatcaggagat
aggggggaaagcggaccggccgacccgctggtgcgcggggacctgctggtacgagaggtgcaccgggtcccaagggcctcgtggagata
aaggggagaccggggagcgcggggctgcagggattaaggacaccgcggctggagaagcgaagggctggtaaaaggcgaccgcggcg
aatcgggtcggcggggtcccgcgggcgcacctggggcctgctgggtctaggggagcggcaggcctcaaggctcctcgcggtgacaaggagaa
acaggcgagcgtggggcagcgggcattaaaggctcatcgtgggttagaaaagcgtgaggtcgcaaatctggcgatcgcggggagagcggcca
gcaggctccggctggtgctccaggctccagctggtagccgtggcgccccaggctcctcaaggctcctggtgcgacaaggggagaccggcgagcgt
ggagcggcaggcatcaaaggctcatcgtggacttgaagaaggagggaggggaaagtcgggcgatcgggtgaatccggcccagcgggtcctgct
ggagccccagccccctggtatcgcggggcgacctggcccccaaggaccgctggggataaaggggagactggcgagagagcggcgcc
ggaatcaaaggtcaccggggattagaaaagaggagggcagggaaagtcgggtgatcaggcgcaatcaggctcctggtgacctggtgcgcgc
gggtcctgctggttagcagaggagccccctgggcctcaaggcccagggggtgataagggtgaaccggagaacgagggcggcgtggtataaagg
gacatcgtggactcgagaagcgagaagcggggaaaagtggggacagaggagagtcgtgctcctgccggaccgccgggtgctcctggtccagca
ggctccgcggcgccctggtccgaaggtcctaggggagacaaaggagaacgggcgagcgagggcgccgggatcaaggggcatcgg
gggttgagaagaggtaa

编码 17S28-8 的 DNA 序列, 如 SEQ ID NO.72 所示:

gaggccggtgtaccgggaagcgtggggccgaagggcagctcgggttcaccaggcccacaggggccaccaggccccggtaggcctacaa
ggtttgcgcggtgaggtggggttaccggagtaaaagggtgataaggggcctatggggcctcccggtcctaaaggcgaccagggtgaaaagggtc
cgcggggactggagaaaagagaagcgggtgttcagggtcgggtggtcctaagggaagtgcgggttccggggcgcaaggctcctcagggtcc
cgtcggtttacaggccctacgtggagaagtcggactccccggagttaggcgataagggacccatgggtcctcctggtccaaagggtgaccaag
gcgagaaaaggcccaaggggactcgagaagcgggaagcgggagtcctcgtgctcgttgccctaagggaagtccgggaagtcgggaccgcaa
ggctcctcgtggcctgtcggcctcaagggttgaggggagaggttgactccccggcgtaaggcgacaaaggccctatggggccacctgggc
ctaagggcgaccaaggcgagaaaggccctcgcgggcttgagaagcgtgaagctggagtgctggtcagtagggccaaaaggatcgtccgat
caccgggacctcagggggccaccgggtcctgtaggtctacaggccctcaggggcgaggtagggttacctggagtcagggaagataaggacat
ggggccgccaggtccgaaaggcgatcaaggggaaaaagggcctcgtggactagaaaaacgtgaggcaggcgctcctggatcgggtggtcctaa
aggttcttcaggtagtcgggtccccaggacctccggcgccgtagggctacagggttacgagggaggtgggcctccagggggttaaaggg
gataaaggctcctatgggtcctccgggtccaaaagggtgatcagggtgagaagggtcctcgtgggttggaagagagaagcaggtgttcagggtc
tgttgaccgaagggtcttctggaagccctggctcctcaaggctcctcgtgctcgggtcggacttcagggttcgagggcaggtaggtctgcccgg
gggttaagggtgacaaaggccaatgggaccgccaggggcgaaaggcgaccaagggtgaaaaaggccctcgaggtctggaagaaagagaagctg
gagtaccaggggagcgtcgccccaaagggaagtcggggtccctggaccgcaagggccaccagggtcccgtaggggttacaaggctcgcggggg
gaggttgggttcccggagttaaagggtgataagggaccaatgggtcctcccggtcctcaagggtgatcaaggagagaaaggaccgcggcggtg
aaaagagggaaggcgtgtccaggtagcgttggtcctaaaggcgagctcaggtccccaggaccccaaggggccggggcctgtgggtgtgc
aagggttgcgaggagaagtggtcctccgggtgaagggggacaagggacctatgggaccacccgggtcccaagggtgaccagggtgaaaag
ggtcctcgtggacttgaagaaagataa

编码 17S1N6-6 的 DNA 序列, 如 SEQ ID NO.73 所示:

gaagcgggacctcccgagtggtgaaaagggcgaaacgaggcgagccggagagcctggacctcatggcctccggcgctccggga
agtgttggtcctaaaggtagctccgatctccaggtccacaggaccacctggcccgtaggccttcaaggactgagaggtgaggtcggcctcca
ggcgtgaagggcgataaaggtcctatgggtccacctggaccaagggagatcagggggagaaagggccgaggggcttagaaaaacgcgaagc
agggcctcctgggagcggagaaaaagggcgagcggagccgagcgcagccaggtcctatggtcctccaggagtgcctgggtccgtcggcc
caaaggggtcttctggttcgctggcctcaaggaccaccaggtcccgttggttgcaggggtctgctggagaagtgggctccgggagttaagg
gcgacaaggggtcctatggggcctccggggccgaagggcgatcaggcgaaaaagggcccgagggttagaaaaagagagggcaggaccacc
cggctcaggtgagaaggggtgaaagaggagccgctggggagccaggaccacacggccccccggagtcggggatcgttgaccctaaaggtt
cttctgggagcccggtccccaggaccgcccaggtcaggtctacaaggactcagagggaggtgggtctaccgggggtcaaaggagata
agggctctatgggtcctcccggtccaaaaggggaccaagggagagaagggacctcggggttgaaaaacgggaagctggtcctcccgtagcg
gtgagaaagggcgagcgaggtgctgctgggaaccggggccacacgggtcctcctggcgtgcccggtcggtcggacctaaggttcttctgttct
ccgggtcctcagggggccgctggtcctgtcgggttcaaggtctccggtgaggttgggtccccggtgtcaaggagacaaagggccgatgg
ggccgccaggltccaaaggagaccagggagaaaaaggtcctcgtgggttagaaaaacgggagggcgggaccgcctggttccggcgagaaggg
ggaacgagggcgggccgggaaccaggaccccatggaccgcccaggtgttccgggggtccgtaggtccgaaaggctcatcggttcgcctggccc
acaaggaccaccagggccagtggcctacaaggtctgaggggggaagttgggtcctcctggcgtaaaggagacaaagggccgatgggtcctcc
tggaccctaaaggtgaccaaggagaaaaaggtccgaggggcttagagaaacgtgaagcaggtccgccaggttcaggtgagaagggggaacgtg
gagctgccggggagcctggcccgacggccctccggcgttccgggtcagttggaccaaaggatcagtggaagtcctggtcctcaaggccc
acctgttcagtaggtctacagggcttgaggggggaggttaggtcgcggcgtaaaaggcgataaggggtccatgggtcctccagggcctaaa
ggtgatcaggggggagaaggggtccccgaggggttgagaagcggtaa

编码 17S1N7-7 的 DNA 序列, 如 SEQ ID NO.74 所示:

gaggcaggtcctccgggttctggtgagaagggcgaaaagggagctcggggagaacctggcccatggtcctcctggtgtaccggggtcg
gttggtcctaaaggttctcaggtcctcccggtcccgaggaccaccgggtcctgttggtcctccaggcctcgaggggaggtgggacttccggg
agtcaaaaggagacaaaggtcctatgggtcctcctggtcctaaaggcgatcagggtgaaaagggaccgcgagggcaagagagaggtggtcctccc
ggaagcgggtgagaaggggtgagcgaggggctgctggagagccaggtccacacggacctcccggtgttccggcagcgtaggcccaaaagggtc
gtctgggtccccggggccacaaggtccccctggaccgtaggggtgcaaggcctacggggagaaagtaggtaccgggtgtaagggcgacaa
aggtcctatgggaccccccggtcccaaggtgatcagggtgagaaggggtccgaggggtaaacgcgaagctggtccaccaggtatggcgaaaa
gggagaacgtggtgctgcgggagagcctggggccacacgggtccccgggagtagctggaggttggccgaagggcagtagtggtcaccgg
ggccgcaagggccccccggaccggtaggactacaaggcttacgggggaggttaggtcttctgggtgaaaggggataaaaggtcctatgggac
caccggccccaaaaagggaccagggagagaaaagggccacgcggaaaaagggagggcagggccaccggggtcgggtgaaaaaggtgaacgc
ggagcggcgggggaaccggggcctcatggggccaccggcggtccaggttcggttggtcctaaaggttcaagcggtatccaggccccgaggg
acctcctggaccggttgggttcagggtcgtggtgaggtcgttctcccggcgtaaggagacaaaggtcctatgggtcctccgggtccaa
agggagatcaggggtgagaagggaccagaggggaaagagaagctggccctccggggtccggggagaagggggaacgtggggcagccggtg
agccgggacctcacggccaccagcggtgcctgggtccgttggtccaaaggggtcctcgggtcgcagggccctcagggaccaccgggtccag
tcggattacaaggttgcggggagaggttgggtcctcggaggttaaggtgacaaaggccctatgggtccccgggtcctaaaggcgaccaagggc
gaaaaagggaccgagaggggaaacgtgaagcaggtccacctgggtcaggtgagaaaggcgagcgggggccgggggaaccggggccacac

ggctcctctggcgtccccggcagcgtcgccccgaaaggaagtctggaagtccgggaccacagggtccgcccggctctgtagggtctacaaggttt
aagaggcgaagtgggactccctggcgtgaaaggtgacaaagggcccatgggaccaccagggtcctaaaggggatcaggagaaaaagggcccta
ggggtaaacgagaggcaggtcccccgaggagcggagaaaaggggtgagcgaggggcccggcgggtgaaccaggccccgatgtccaccaggcg
tccccgatcagtcggacaaaaggcagttctggctcacctggaccgcaagggtccgcctggacctgtaggacttcaaggcctgcgtggtgaagttg
ggttcctggggtaagggggataagggtcctatgggtccccctgggcaaaaaggggatcaaggcgaaaaggggtccccggggaaaaaggtta

编码 17S1NK-7 的 DNA 序列, 如 SEQ ID NO.75 所示:

gaagcaggtccgccaggagcgggtgaaaaaggagaacgcggagcggcgaggccaggccccgcacggcgcccgggtgttccccg
ctctgttgaccgaaaaggtagttctggttaccggaccctcaagggtcctctgggcctgttggacttcaggggcttagaggtaggtcggactccctg
gtgtgaaaagggataaaggcctatgggtcctctgggcaaaaaggggaccaggagagaaggggaagaggaggtgtgtcctccaggatcg
ggggaaaaaggagaacgtggagctccggcgagccggggccacacgggtccaccggcgctccccgggtccgtgggaccaagggtcatcag
gatcgccaggaccgagggtcctccaggtcctgttgggttcaggggttacgtggagaagtgggtctccccgggtgttaagggtgacaagggtccca
tgggacccccgggtcctaaaggagaccaaggagagaaaaggttaaacgagaggccgggacacgtgggtccggcgagaaggcgcaacgtgtgtgc
cgctggcgagccccgggcccacgggtccaccgggagtgccgggtcagtagggccgaaaggaagttccggaagtccaggctcctcaagggtcccc
gggacctgtaggcctgcaagggttcgtggtgaggtcgggtcctccctggggtaagggtgacaaaggtcctatgggtcctccggacctaaagggg
atcaaggggaaaaaggaaaacgagaggctggaccaccaggttctggcgaaaaaggtgagcgaggcgccggcggaacccggcccgcatgg
ccccccaggcgtaccaggctctgtgggtccgaagggtcgcagcggttcacctggacctcaaggggccacctggaccggttgggtctacaaggccta
cggggggagggtgggaclaccggcgtaaaaggtgacaaaggcccgatgggaccaccaggccctaaggagatcaggcgagaaagggtlaaga
gagaagcaggaccgccgggaagcggcgaaaaaggtgagcgcgagcagcagggggaacccggcgcatggtccccctggcgtaacctgtgta
gcgttggctctaaagggtcgtcgggttccccagggtcctcagggtccccctgggtcctgtcggtctgcaagggttaaggggggaagtaggattaccgg
gcgtgaaaggagataaagggtcctatggggccacctgggtcctaagggcgaccagggtgagaagggtlaagcgcaagctggccccggggtagtg
gtgagaaaggcgagcggggggcagcggcggaaccagggtcctcatggtcctctggtgtgccagggtccgtcgccctaagggatctagcggaa
gccccgggtccgcagggaccaccagggtcctgttggcctcaagggttgagagggtgaagtcgggttaccagggggtcaagggtataaaggccgatg
ggtccacctgggcctaagggggatcagggggagaagggaagcgggaagccgggtcccccggaagtggagaaaagggggagaggggggct
gcaggtgaacctgggcctcacggaccaccgggtgttccaggatcagttgacccaagggtagttcgggatctcctggccctcaaggccccctgg
tctgtaggacttcaagggtcgcgggggaggtaggtcctcctggtgtcaaggagacaaaggaccgatgggtcctcctgtcctaaggagatc
aaggagaaaaagggaagcgctaa

编码 3A5D15D-5 的 DNA 序列, 如 SEQ ID NO.76 所示:

gatgggaacccgggtcggacggcctgccgggtcgcgacggctcaccagggtggcaaggcgacctggtgagaatggatctccgggag
caccgggtccccctggacatcctgttcccccgggcctgtcgggtcctgtgtggcaagtcgggtgataaggagaatccgggctgccggtcctgt
ggggcaccaggggccagcaggaagcgtggggctccgggtcctcaagggtccccgggggacaagggtgagaccggagaacgcggagctcgg
ggaatcaaggggcatcgcggcttagagaaacgtgatggcaacccgggctctacggggttcccggtcgggacggcagccccgggtgggaaaggt
gatcgaggggaaaacggctcgcctggggctccagggtcccccgggccacctggaccgccagggtcctgtcgccctgccggtlaagagcggagat
cgagggtaatcaggacctgccggcccagccgggtcctcctgggtccagctggaagccgtggtgtcctccggccccgagggtcctcgtgtgcgataag
gcgagacgggtgaacggggcgagccggaattaaaggacacagaggtcttgaagaggggatggaacccagggtcagatggtctaccaggc
agggacgggtcaccggagggaagggtgatcggggcgaaaacgggtccccgggagcgccagggtgccctgggcatccccgggccaccggac

cggtgggtcccgcagggaaatcgggggatagggcgagcttgaccagcggggcctgcgggagcacctgggccagcaggtagtaggggggc
gcccgggtccgcaaggggccacggggcgacaaaggggagactggcgagaggggggctgccggcatcaaggggccaccgcggcctagaaaaga
gagacggcaatcctggttctgacggattgcctggtcgagatggttctcctggtggaaaggggtagaggggagaatggatctcctggcgcccca
ggtgcgcggggcatccaggccccggcgccggttgaccagcgggtaaatcgggcgacaggggggaaagcggctctgcaggaccagcg
ggcgccaccggggcctgctgggagtagaggcgtccaggctccgcaaggtccccgaggggacaaggcgaaactggggagcgtggtgcggccc
gcataaagggacaccgaggactcgaaaaacgcgatggaatcccggtccgacggactccctggcggggatggcagtcgggaggtaaaggga
gacagaggcgagaatggaagtcctggggcgctggcgctccggacatccaggggccaccggaccggtagaccggcaggaatccggtg
accgcggcgagtcaggctccgcaggaccagcgggtgcacctggcgccggcggttagtctggtgggtcccgggccgcagggtccgagaggtgat
aagggagagacaggagaacgcggtgcccgccggaataaagggcaccgaggcctggagaaacgataa

编码 3A5D15E-5 的 DNA 序列, 如 SEQ ID NO.77 所示:

gaagccgatggtaccaggaagtgatggcctcccggggcgcgacggttctctggcgaaaaggcgatcgtggagaaaatgggtcgc
aggtgccccgggagcgcggggcatcctggtccaccaggccccggttggtctgcagggaagagtggggaccgcggagagtcgggtccagcg
gacctgcaggcgctcctggcccagccggcagtcgcggtgccccctggggcacagggtccagaggagacaagggggaaaccggggaaagg
agctgctggaataaagggacaccgggggtagagaaaaggagggcgacgggaaccaggtatcagacgggtctccgggtcgggatgggagc
cctggcggaaggggatagaggcgaacggtagtccggggcgccctggagcaccggccaccaggtcctctggaccagttggtccggc
aggtaaatccgggatcgcgcgagagcggccccgctggacctgcggggcgccaggccccggcggtatcgcgaggtccccctggtccgcag
ggggcgctggcgataagggtgagacaggtgagcggggagccgcgggcatcaaaggatcaggggcttagagaagcgtgaagcagatggaaa
tcccggtatctgatggactcggggacgagatggctccccctgggggcaaggcgacagaggagaaaatggttcaccgggcgcaccaggtgctcc
aggccatccaggtcccccgcccagttggtcctgctggcaaaagtggagacagaggtgagtcgggacccgctgggcctgcgggagcaccagg
tccggccggctcgcggtgctccaggccccagggtccgcggggcgataaggcgagacaggagagcggggagccgcaggtatgaaagg
catcgaggtttgagaaacgtgaagcagacggaaaccgggttcagatggactaccaggccgtgacggcagccccgggggaaagggtgaccg
aggtgaaaacggttctcctggagccccgggtgccccggggcatcctgggctccaggccccgtaggacccgcaggcaaatctggggaccgagg
ggaaaaggccccgctggtcctgcgggcgccccgggtccgcgggttctctggtgacactggccccaaagtcctagggtgacaaggggga
aacgggggagcgtggggcagctgggataaaaggccaccggggtcttgaaaacgcgagggcgatgggaatccgggaagcgacggctaccg
ggacgggacggctacccggggggaagggtgatcgcggggagaatggtccccgggtcctcctggggctcctgggcaccctggtccgcctggt
cccgtcgaccggcagggaaatcaggggacagaggggaatccggacctgctggtcctgcggggctcccggtcctgctggttctctgtaggcc
ctggtcctcaaggccaagggtgataagggtgaaactggcgagaggggagctgcagggatcaaggccaccgaggactcgaaaagaggttaa

编码 3A5D15R-6 的 DNA 序列, 如 SEQ ID NO.78 所示:

gacggaaatcccgaagcgacggtcttcccgaagggtatggtccccctggcgggaaaaggatcgcgcgaaaacgggtctcccgcg
gccccggccccggccaccctggggccacctggtccggttggtcctgctggcaaatcaggtgacctgggggaatcggggccagctggtcctgct
ggggcgccaggaccgcaggcagctggtggcgccccgggctcaaggccacgaggtgataaggagagaccggcgagagggggcgccg
tggaatcaaacgggacggtaatccccggttcagacggcctccccgggacgcgacggctccccgggggtaaggggatcgcggtgaaaacggatc
gccaggtgcgccaggggcgccccgacacctggtcctccggggcctgtcgccccggcaggcaagtctggcgacagggcgagctggtcctg
caggccccgaggtgctcctggcccagcaggatcacgtggcggccccggaccacaggggccgagaggggataaggcgaaactggtgaacg
ggggcgcggtggaattaagagagatggtaatcccggtagtgtgggtaccgggacgagacggttccccggggtaagggaagatcgaggtga

gaatggcagtcaccggcgcccctgggtcctcctgggcatcccggcccaccgggtccgggtggacccggcgggcaaatccggggaccggggggaga
gtggcccggcggtcctgcgggtgccccgggaccagcggggaagtaggggggctccggggccacaaggctcctgtggagataaggcgaaac
aggtgagagaggagctgcagggaattaagaggacggaaatcccgggtcggatgggttcctggacgggatggtagccccgggtgtaaggaga
ccgaggcgaaaacggatctcccgtgcctggagccccaggtacccggggccacctggctctgtaggacctgcaggtaagtcgggtgaccg
cggagagagcgggaccggcggtcctgcaggagccccctggacctgcgggtcccggggagcaccaggcccccaaggaccgagaggtgacaag
ggggagacaggtgagcgtggggctgctgggataaaaagagatggaacccgggtcctgatgggtcctggaaggatgggagcccaggtgg
aaaaggagatcggggtgagaacggttctcctggagccccgggagccccggccatccggggccgcccgggacctgtcggtcctgtgaaaatc
aggcgtatcgtgggaatcaggaccagctgggctcctggagcaccggggccagcaggagcagaggtgtcctggacctcagggaccgcga
ggcgacaaaaggagacggggcaacgcggtgcggcaggtataaaacgagacgggaaccaggctctgatggtttaccaggctgtgacggctc
gcctggaggcaaggagatcgcggtgaaaatggtccccgggtgctcctggagcgctgggcatccggggcctccaggtccggtaggtccagca
ggcaagagcggggacaggggagaatccggaccagctggtcctgcggagcgccaggccccgcaggaaagtcgaggggctcctggtcccaag
gtcctcgtggcgacaaaaggcgagaccggtgaacgcggcgctgcaggcatcaagagataa

编码 3A5D15KR-6 的 DNA 序列, 如 SEQ ID NO.79 所示:

gacggtaaccctggaagtacggattgccggggcgagacgggtcaccggaggaaaggcgatcgcgcgaaaatgggtcaccgggcg
ctcctggggcaccaggggcatccgggtcccccggtcctgttgggcccggcggttaagtctggtgatcaggcgagcttgccctgcgggtcctgct
ggtgcacctggacctgcggatcgagaggggtccaggggcgaaggacctagggggtgacaaggcgaaacgggtgagcgaggagccggccg
gaattaaggggaaaagggaaggcaacccgggtccgatggcctgccaggacgcgatgggtcgcccggaggcaagggtgacagaggggagaa
tggctctccaggtgcaccaggtgccccggggcaccccggtccgctggaccagttggaccggcgggaaagtctggagatcgcggggagagtg
accggctggtcctgcaggggcacgggtcccgtgggtctcgcggagcgctggcccccaaggaccagaggggataaggcgaaaactggag
agcgtggcgcggcaggaataaaggcgaaacgtgatggcaatccagggtcggacggacttctggccgggacggttcccagggggaaagggt
gacagaggcgaaaacggctcggcggtgcgccaggggcgctggtcatccggggcctccgggtccagtgggaccagccggaaaaagcgggg
atcgggggggaatctggaccggctggaccagccggcgaccaggccccggcgatcgagggggggcgccggaccgcagggtcctagagggg
ataaagggtgagaccggcgagagggcgagctgggataaagggttaagcgggacggtaacccggctccgacctcaccaggacgagacgg
gtcaccgggttgtaaggagaccgagggcaaatgggtcacctggtgcaccggcgccccctggacacccggctcctcctggaccagttgggacc
cgaggggaagagcggtgacagaggggagagtggtcctgcaggaccagctggcgctccgggaccggcggaagtcgggggtgtccccggccc
ccaagggcctagaggggataaagggtgagaccgggggaacgtggcgcgccgggtatcaagggcaaacgcgacggcaaccagggtcggatgga
ctaccagggcgagatggtcacctggcggtaaaggcgatcggggggaaaatggcagccccgggtgctcctggagcaccgggtcatcctgtgtccg
ctggtcctgttgaccggctgggaaatccggtgaccgcggagaatccggccggcggtccggctggagcccctggcgccggtggcagtcgtg
gcgcgctggcccacagggtccgctggagataaaggggaaacaggtgagaggggcgcagcgggaatcaaggggaaacgggatggcaacc
cggggagcgacgggttaccgggcgagacggtccccaggaggcaaggcgatcgtggcgagaatggtagccccgggtcctcaggagcacct
ggccaccagggccaccaggaccggtagaccgcctggtaaatcaggcgatcgggggtgagagtggccctgcgggaccgcgggcgccccg
gacctgccgggagcccggtgctcctggaccacaaggctcctagggcgacaaagggtgaaacaggggaacgtggcgctcgggggattaaaggga
aaaaggtaa

编码 3A5D15EKR-6 的 DNA 序列, 如 SEQ ID NO.80 所示:

gacggaaaccctggtcgtgatggtcttccaggacgggatggcagtcgccggcgaaaaggcgaccggggagaaaacggttaccgggtgc

tcctggagcaccaggacatcctgggccacccggccccgttggtccggcggggaaatcaggcgacaggggcgaaagtgggcctgccgggcctg
cgggagccccgggtccggcaggtagccgaggggcaccaggccctcaaggggccacgaggtgacaaggagaaacaggcgaaagggtgctg
ctggaatcgagaaacgtgatgggaatccccgggtcagacggcctaccgggacgggatgggtccccggcgcaaaggggaccgaggggagaa
ggttcgcttggtgccccaggcgcgccgggtcatcctggtccccgggacccgtcggtcctgctggcaaaagtggggatcgtggagaaagtggac
ctgccggacctgccgggggtcctggtcctgctggatcgcgcgccgcaccaggacctcaaggtcctagaggagacaaaggggagacggggagaa
agaggagctgcagggataaaaaacgggatgggaatcctggaagcgatgggtaccaggcagggacggaagccccggcggggaagggggaca
gggggggagaatgggtcaccagggtgcggccggagccccgggacatccccgggccctggtccggtagggcctgcaggtaaatccggtgatagag
gagaatccggacctgcggggccagccggagcgccggggccggcgggtagcagggcgccaccgggacctcaaggggccacggggtgataagg
gagaaaccggtgagagggggcggggtggaattgagaagcgtgatggtaacctgggtccgatggactccccggcgagacggctcgctggcg
gcaaggggcgaccgcggcgagaatgggtctcccggtgctcctggtgctcctggacacctggtcctccccgggtccggtgggtcctgccggcaagagt
ggggacagaggggaatcgggggcggtggaccggcggtgcaccgggtcccgctggctcacgaggtgcggggggccacaaggcccacgc
ggcgataagggggagactggtgagcgtggagctgcaggaatcgagaagcgcatggcaaccaggtagcgacggattgccaggtcgtgacgg
atctctgggggtaagggtgatcgaggagaaaatgggtcaccgggtgctcctggggcgccccggccacccgggacctcctggtcctgtatgtcccc
ccggcaaatcgggcgaccgggggtgagtctgggcctgcaggccccggcggtgccccagggtccagcagggtctcgtggcgcgcccggtccccag
ggctcctggtgggataaggcgagacggggagaaagaggtcgggcaggcatagaaaagcgcgatgggaacctggtatctgatgggtccccagg
gagagacggaagccccggcggaaggtgatcggggcgaaaacggaagtccaggtgcggcgagcaccggccaccaggtccaccggg
accggtggtccccgtgggaaatctggtgacagaggggagtcggggccagcagggcgccggcgccaccaggaccagctggttctcgtggtgc
ccctggacctcaaggtcctaggggggacaaaaggcgagaccggggaacgcggtgcggctgggattgagaagcgctaa

编码 3A5D15G-6 的 DNA 序列，如 SEQ ID NO.81 所示：

gacggcaatcctggatctgacgggctgcctggccgtgacgggagcccagggggtaaaaggcgaccgcggagaaaatggtagccctggcg
ctcctggtgctccccgtcatcctggtccccgggaccagtggggccggctggaaaaagcgggtgaccgtggtgaatctggcccagcgggacctgc
gggagccccgtgctcctgccggttctcgtggcgacccggacctcaaggtccgggggtgataaaggcgagacaggggagcgaggtgccgcag
gcattgaaaaaggacggaatccgggttcagacggcttgcagggaagggtggtcgcgggtggttaaggagacaggggagaaaacgggt
ccccagggtgctcctggcgacctggccacctggtcctccggggcctgtaggaccagctggcaagagtggagatcgcgagagatctgggcctgc
gggtccggcaggtgctccccgacccgccggatcacgcggtgctcctggaccgcaaggtccgcgggtgacaaggagagactggggagagg
ggtgcggcggtattgtaagcgtgatggcaatccaggcagcgaggtctaccggggcgggatgggagtcgggtgggaagggggatagggg
tgagaacggctctccaggggtccccggtgccccgggcatcctggtcctcctgggcctggtggctcctgccggaagtcaggggatcgaggcgag
agtggctcctgctggtcctgctggggcgccaggaccagcggggtctcaggcgccccctggcccgagggtccaagaggtgataagggggagacc
ggggaacgaggagccgctgggatagggaagcgcgacggaatccccgggagtgatgggcttcaggccgtgatggaagtccccgagggaagg
gagatcgggcgaaaacggttctcccggtgctcctggagcgccggacacccggcccaccaggccctgttggggcgccgggaaaagcggc
gatcgtggcgaatctggtcctgcgggtccagctggtgctcctggggcgcgaggttctcgtggtgcacccggaccgagggaccagaggagata
aaggggaaacgggtgagcgaggggagcaggtatcgggaaaaggacggtaacctgggtcctgatggcttaccggggcgagacggatcgcc
gggggggtaaggagaccggggagagaacggtcaccaggcgacccggcgctccagggcacatcctggggcaccggccccagttggtcctgctg
gtaaatcgggggaccgagcggaatcaggacctgcaggcctcggggtgccccggccccgctgggagtcgtggtgcacctggggccacaaggc
ccgcgcggcgacaagggggagacgggagagcgcggtgccgtggaatcggaagagagatggaacccgggagcgacggactcccaggc

agagatgggtccccggcgcaaggagacagaggtgaaatggatcacctggtgcgcctggtgctcctggacacccaggaccgccaggtcca
gtcggctctgcgggaaatcgggtgatcgtggtgaatccggcctgcaggtcccgtggagcgccgggtcccgagggtcgagggcgcgcca
ggtcacaaaggtcctcggggcgataagggcgaaacaggtgaaagaggcgcgccggcatagaaaaagataa

以上 DNA 序列 SEQ ID NO.71~SEQ ID NO.81 均由南京金斯瑞生物科技有限公司合成。将合成的基因克隆于 pPIC9K 载体上, 克隆位置为第 1203bp (即 α -factor secretion signal/ cleavage site 1203, 查找序列为 AAAGAAGAAGGGGTATCTCTCGAGAAAAGA) 与酶切位点 Not I 之间。获得重组表达载体 pPIC9K-3A5D29N-9、pPIC9K-17S28-8、pPIC9K-17S1N6-6、pPIC9K-17S1N7-7、pPIC9K-17S1NK-7、pPIC9K-3A5D15D-5、pPIC9K-3A5D15E-5、pPIC9K-3A5D15G-6、pPIC9K-3A5D15EKR-6、pPIC9K-3A5D15KR-6、pPIC9K-3A5D15R-6。其它可用于毕赤酵母的表达载体如 pPICZ α B、pFLD α 也与 pPIC9K 有类似的效果。

分别将上述重组表达载体质粒 (pPIC9K-3A5D29N-9、pPIC9K-17S28-8、pPIC9K-17S1N6-6、pPIC9K-17S1N7-7、pPIC9K-17S1NK-7、pPIC9K-3A5D15D-5、pPIC9K-3A5D15E-5、pPIC9K-3A5D15G-6、pPIC9K-3A5D15EKR-6、pPIC9K-3A5D15KR-6、pPIC9K-3A5D15R-6) 10 μ g 用 SacI (购自大连 TaKaRa 公司) 37°C 酶切消化过夜, 使其线性化, 再使用 PCR 产物纯化试剂盒 (购自生工生物工程(上海)股份有限公司) 回收线性化质粒, 使体积控制在 10 μ L 左右。

将线性化质粒电转化入本发明实施例 2 中构建的 CPB 定位融合功能蛋白功能途径的 4 种底盘工程菌株 (HCPB-PPKEX2 菌株, 保藏编号 CGMCC No.25815; RCPB-PPKEX2 菌株, 保藏编号 CGMCC No.25817; HCPB-SCKEX2 菌株, 保藏编号 CGMCC No.25816; RCPB-SCKEX2 菌株, 保藏编号 CGMCC No.25818) 的感受态细胞中, 将电转后的菌液涂布于 MD 平板上, 每 100 μ L~200 μ L 涂布一块平板, 室温静置 10min, 于 30°C 倒置培养 2-5 天, 直至有单菌落 (阳性转化子) 出现。

向 MD 平板表面加入 2mL 无菌双蒸水, 然后用无菌三角涂布器轻轻刮下平板表面的 His⁺转化子, 并转移到 50mL 离心管中。以无菌双蒸水稀释菌悬液, 105 个细胞涂布于含有 0.5mg/mL G418 的 YPD 平板上, 倒置, 30°C 培养 3~4d 后至单菌落出现。从 YPD 平板上挑取菌落至无菌 96 孔板中 (200 μ L YPD/孔), 混匀, 于 30°C 培养 48h; 混匀孔中菌液, 各取 10 μ L 接入至一块新的无菌 96 孔板, 于 30°C 培养 24h 后再重复一次此操作; 24h 后, 从第三块 96 孔板中取出 1 μ L 分别点在含有 1.0mg/mL 和 4mg/mL G418 的 YPD 平板上, 于 30°C 继续培养 96h~120h。毕赤酵母转化子若能在含高浓度 G418 的平板上生长, 说明该转化子能够高效率表达外源基因, 经过这一步筛选可得到高效表达的重组酵母工程菌种。

后续以表达 HCPB-PPKEX2 的底盘细胞工程菌株 (保藏编号 CGMCC No.25815) 和表达 RCPB-PPKEX2 的底盘细胞工程菌株 (保藏编号 CGMCC No.25817) 为表达宿主细胞来进行说

明。

本发明的重复串联表达体系构建了多种表达重组小分子胶原蛋白的工程菌株，其以 5 种表达 HCPB-PPKEX2 的底盘细胞工程菌株为表达宿主细胞构建的工程菌株送至中国微生物菌种保藏管理委员会普通微生物中心保藏，菌种保藏编号分别对应的是：

表达重组小分子胶原蛋白 3A5D29N 的菌株，保藏编号为：CGMCC No.25819；

表达重组小分子胶原蛋白 17S28 的菌株，保藏编号为：CGMCC No.25821；

表达重组小分子胶原蛋白 3A5D15D 的菌株，保藏编号为：CGMCC No.25827；

表达重组小分子胶原蛋白 17S1N6 的菌株，保藏编号为：CGMCC No.25829；

表达重组小分子胶原蛋白 3A5D15E 的菌株，保藏编号为：CGMCC No.25828。

其以 2 种表达 RCPB-PPKEX2 的底盘细胞工程菌株为表达宿主细胞构建的工程菌株送至中国微生物菌种保藏管理委员会普通微生物中心保藏，菌种保藏编号分别对应的是：

表达重组小分子胶原蛋白 3A5D29N 的菌株，保藏编号为：CGMCC No.25820；

表达重组小分子胶原蛋白 17S28 的菌株，保藏编号为：CGMCC No.25822；

以上工程菌株的保藏地址均为：北京市朝阳区北辰西路 1 号院 3 号；保藏日期均为：2022 年 9 月 26 日。分类命名均为：巴斯德毕赤酵母 *Komagataella phaffii*。

分别挑选高浓度 G418 的平板上生长的单菌落，置于装有 10mL BMGY 培养基的 100mL 三角瓶中，于 28-30°C、220rpm 培养至 OD600 为 2~6(16-18 h)。室温下 1500~3000g 离心 5min，收集菌体，用 BMMY 培养基重悬菌体，使 OD600 为 2 左右，放置于 28-30°C、220 rpm 的摇床上继续生长 3 天，每 24h 向培养基中添加 100% 甲醇至终浓度为 1.0%。按时间点（开始诱导后每 24 小时取样一次）分别取菌液样品，取样量为 1mL，置于 1.5mL EP 管中，12000rpm 离心 2~3min，收集上清，加入 5×上样缓冲液（250mM Tris-HCl、pH 6.8，10% SDS，0.5% 溴酚蓝，50% 甘油，5% β-巯基乙醇），置于 100°C 金属浴加热 10min，进行 SDS-PAGE 检测。

SDS-PAGE 检测结果如图 16-19 中所示结果，pPIC9K-3A5D29N-9、pPIC9K-17S28-8、pPIC9K-17S1N6-6、pPIC9K-17S1N7-7、pPIC9K-17S1NK-7、pPIC9K-3A5D15D-5、pPIC9K-3A5D15E-5、pPIC9K-3A5D15G-6、pPIC9K-3A5D15EKR-6、pPIC9K-3A5D15KR-6、pPIC9K-3A5D15R-6 均于两种底盘细胞工程菌株中，成功有效表达相应大小的重组小分子胶原蛋白，且有效分泌表达于上清中，且电泳条带单一，说明本发明小分子胶原蛋白串联重复氨基酸序列设计的目的实现：即表达多达几百个以上氨基酸序列的大分子蛋白（小分子胶原蛋白单体串联重复结构），在底盘细胞工程菌株中，最终可表达获得序列相同且条带单一的重组小分子胶原蛋白（胶原蛋白于电泳时会有一定的电泳迁移延迟，所以在电泳时其表观分子量会偏大）。

对得到的小分子胶原蛋白 3A5D29N、17S28、3A5D15D、17S1N6 的冻干品用胰蛋白酶将其

酶解, Nano-HPLC-MS/MS 质谱检测重组胶原的胰蛋白酶解后肽段(交由苏州普泰生物技术有限公司完成), 并将检测到的肽段进行序列比对(Uniprot 数据库)。如图 20 和 21 所示, 质谱检测结果表明 3A5D29N、17S28、3A5D15D、17S1N6 这几种重组小分子胶原蛋白酶解后检测到的肽段及其所覆盖序列均属于氨基酸序列设计时选择的人胶原蛋白序列的相关区域, 说明其成功表达。

取 3A5D29N、17S28 的冻干品进行 LC-MS 分析(毛细管高效液相色谱仪 Thermo Fisher Scientific Ultimate 3000、电喷雾-四极杆飞行时间质谱仪 AB SCIEX TripleTOF 5600 Mass Spectrometer, 色谱柱 ACQUITY UPLC Protein BEH C4 Column)获取其去卷积分子量(委托北京百泰派克生物科技有限公司完成)。结果如图 22 所示部分结果(图中去卷积分子量值显示值保留小数点后 1 位, 有四舍五入处理), 3A5D29N 理论分子量 4988.35Da, 实测去卷积分子量为 4987.5Da; 17S28 理论分子量 5861.60Da, 实测去卷积分子量为 5833.92Da; 蛋白质在表达过程可能产生的糖基化修饰以及检测产生的误差, 重组小分子胶原蛋白 3A5D29N、17S28 实测去卷积分子量与理论值基本一致。

底盘细胞工程菌株效果及表达体系的验证:

本发明构建的底盘细胞工程菌除去有将序列长的蛋白质序列切割为单一且短的小分子胶原蛋白外, 还有一个重要的作用就是将切割后单一且短的小分子胶原蛋白单体的氨基、羧基两端的非胶原蛋白氨基酸序列进行切割, 以 3A5D29N-9、17S28-8 于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株的表达示例, 将表达的重组小分子胶原蛋白 3A5D29N、17S28 进行氨基、羧基两端的检测, 以验证底盘细胞工程菌株中 Kex2、Ste 13 及 CPB 酶的协同作用是否建立。委托北京百泰派克生物科技有限公司对 3A5D29N-9、17S28-8 分别于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株的表达重组小分子胶原蛋白冻干品进行 N 端(氨基端)、C 端(羧基端)的测序验证以及基于 LC-MS/MS 的蛋白全序列分析。

N 端测序: 通过岛津全自动蛋白质多肽测序仪(PPSQ-33A)对样品的 N 端序列进行分析(Edman 降解法): 取适量重组小分子胶原冻干品样品溶解, 将样品溶液滴在 PVDF 膜上, 放置到反应器中, 组装好反应器后将其放置于仪器固定位置, 通过软件 PPSQ-30 Analysis 设置: 样品名称、样品号、测试循环数、选择方法文件, 设置完成后开始测试, PPSQ-33A 产生的原始数据及图谱由 PPSQ-30 DataProcessing 软件识别标峰并导出对应图谱, 数据分析后, 确定蛋白质 N 端序列。

结果检测到 3A5D29N-9 于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株的表达重组小分子胶原蛋白的 N 端序列均为: NH₂-Gly-Lys-Ser-Gly-Asp-Arg-Gly, 即

GKSGDRG, 与理论序列

(GKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLE)的N端氨基酸序列一致。检测到 17S28-8 于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株的表达重组小分子胶原蛋白的 N 端序列均为: NH₂-Gly-Val-Pro-Gly-Ser-Val-Gly, 即 GVPGSVG, 与理论序列

(GVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLE)的N端氨基酸序列一致。

C 端测序: 取适量重组小分子胶原冻干品样品溶解后进行胰蛋白酶 (Trypsin)、胃蛋白酶 (Pepsin) 酶解处理, 之后将处理好的样品通过液质联用 (LC-MS/MS) 分析, 得到质谱原始结果的 raw 文件, 经过软件 Byonic 分析, 匹配数据, 质谱数据经过数据库检索, 得到鉴定结果, 其检测 C 端肽段的二级质谱图如下图 23 和 24 所示。

结果表明: (1) 检测 3A5D29N-9 于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株中表达得到的重组小分子胶原蛋白的检测 C 端序列均列为: GHRGLE, 与理论序列

(GKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLE)的C端一致; (2) 检测 17S28-8 于 HCPB-PPKEX2 底盘细胞工程菌株中表达重组小分子胶原蛋白的检测 C 端序列为 VGLPGVKGDKGPMGPPGPKGDQGEKGPRGLE, 检测 17S28-8 于 RCPB-PPKEX2 底盘细胞工程菌株中表达重组小分子胶原蛋白的检测 C 端序列为 KGPMGPPGPKGDQGEKGPRGLE, 均与理论序列

(GVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGPRGLE)的C端一致。

基于 LC-MS/MS 的蛋白全序列分析: 进一步的, 取重组小分子胶原冻干品样品进行胰蛋白酶 (Trypsin)、糜蛋白酶 (Chymotrypsin)、胃蛋白酶 (Pepsin)、胰蛋白酶 (Trypsin) & Glu-C 蛋白酶、胰蛋白酶 (Trypsin) & Asp-N 蛋白酶酶解处理, 之后将处理好的样品通过液质联用 (LC-MS/MS) 分析, 得到质谱原始结果的 raw 文件, 经过软件 Byonic 分析, 匹配数据, 得到全序列测序验证的结果。综合检测结果分析, 3A5D29N-9、17S28-8 分别于 HCPB-PPKEX2 底盘细胞工程菌株、RCPB-PPKEX2 底盘细胞工程菌株中表达重组小分子胶原蛋白冻干品样品的氨基酸序列总覆盖率均为 100%, 本发明表达得到的小分子胶原蛋白的氨基酸序列与目标小分子胶原蛋白氨基酸序列一致。

在本实施例中, 3A5D29N-9 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时, Kex2 酶会将其分割为相同的小分子蛋白序列单体:

EAGKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLEKR。

其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除，羧基端的 KR 两个氨基酸会被 CPB 酶切割去除，最终分泌获得的是一段 54 个氨基酸的重组小分子胶原蛋白 3A5D29N：

GKSGDRGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGLE。

在本实施例中，当 17S28-8 经转录翻译进入内质网、高尔基体构成的蛋白质分泌途径中时，Kex2 酶会将其分割为相同的小分子蛋白序列单体：

EAGVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEK
GPRGLEKR。其中氨基端的 EA 两个氨基酸会被 Ste 13 蛋白酶切割去除，羧基端的 KR 两个氨基酸会被 CPB 酶切割去除，最终分泌获得的是一段 63 个氨基酸的重组小分子胶原蛋白 17S28：

GVPGSVGPKGSSGSPGPQGPVGLQGLRGEVGLPGVKGDKGPMGPPGPKGDQGEKGP
RGLE。

N、C 端的氨基酸序列是否正确直接说明了底盘细胞工程菌株中 CPB 酶的合成代谢途径是否有效建立、建立后其是否有生物学活性及与其酵母自身 Kex2 酶、Ste13 蛋白酶的协同作用如何。本实施例的结果表明，底盘细胞工程菌株中 CPB 酶（无论是人 CPB 酶还是大鼠的 CPB 酶），均成功的在毕赤酵母中建立了合成代谢途径并正常能发挥其生物学酶活性，并与酵母自身 Kex2 酶、Ste13 蛋白酶协同作用，将经过特定氨基酸设计的长序列的大分子重组胶原蛋白在表达后成功的切割后并去除两端非胶原序列，分泌获得重组小分子胶原蛋白。

实施例 4：小分子胶原蛋白的 500L 级的生产验证

小型发酵罐或摇瓶难以体现大规模工业化生产的实际情况，500L 级的发酵罐才具有放大的可能性及表征大规模工业化生产的实际情况。本发明通过对高表达工程菌株进行发酵工艺的优化，在 500L 级的发酵罐上对各工程菌株的产量进行验证。

3A5D2NT、3A5D29N 两种重组小分子胶原蛋白长度高度相似、序列高度同源且纯化方法一致，本发明具体以这 2 种菌进行小分子胶原蛋白表达的对比。

如下表所示，两种工程菌（表达 3A5D29N 工程菌株为 CGMCC No.25819，表达 3A5D2NT 工程菌株为 CGMCC No.25812）的发酵参数优化后相应的各项指标基本能达到一致，但发酵上清中胶原蛋白表达量（UV 法，UV 蛋白定量经验公式： $C(\text{mg/mL})=0.144*(A_{215}-A_{225})$ 测定蛋白浓度）及最终纯化冻干品的产量有较大的差异，3A5D29N 的发酵上清表达量一般是 3A5D2NT 的 2 倍左右（UV 法受发酵液中色素干扰大，倍比值偏低），最终得到的纯化冻干品产量是 3A5D2NT 的 4~5 倍（直接称重）；最佳批次发酵时，发酵上清中蛋白产量（UV 法）3A5D29N 是 3A5D2NT 的 2 倍，其发酵上清液的 SDS-PAGE 如图 25 所示，相同体积发酵液得到的纯化冻干品产量比较，3A5D29N 的产量接近 3A5D2NT 的 5 倍（直接称重），产量得到巨大的提升。其它

重组小分子胶原蛋白的重复串联表达体系中的重组小分子胶原蛋白产量也是小分子胶原蛋白单体序列表达的 4~6 倍。

重组小分子 胶原蛋白	多批次发酵、纯化指标					最佳批次（冻干产量最高）		
	诱导 时间	菌浓度 OD ₆₀₀	菌体湿重	发酵上清表达 量（UV 法）	纯化冻干 品产量	发酵上 清体积	上清蛋白产 量（UV 法）	纯化冻干 品产量
3A5D2NT	40-48h	290-350	320-350g/L	3.6-4.7 g/L	50~60g	190 L	888.3g	60g
3A5D29N	40-43h	300-364	320-351g/L	8.1-9.0 g/L	257~295g	190 L	1682.1g	295g

实施例 5：重组小分子胶原蛋白生物学活性与透皮吸收性检测

（1）细胞黏附活检测

重组小分子胶原蛋白的细胞黏附活检测方法参考文献 Juming Yao,Satoshi Yanagisawa,Tetsuo Asakura. Design,Expression and Characterization of Collagen-Like Proteins Based on the Cell Adhesive and Crosslinking Sequences Derived from Native Collagens ,J Biochem .136,643-649(2004)。

具体实施方法：正常培养 NIH/3T3 细胞（购自中国科学院细胞库，货号 GNM6，培养、传代方法参照细胞说明书执行）。取重组小分子胶原蛋白 3A5D29N、3A5D2NT、3A5D1NT、17S1NNT、17S3NT、17S28 冻干品，对照本公司（江苏创健医疗生物科技股份有限公司）生产的重组Ⅲ型胶原蛋白(专利申请号 CN201310033299.6, 公开日 2013 年 5 月 15 日,分子量 43.6KDa)和重组 XVII 型胶原蛋白冻干海绵(专利申请号 CN202110520499.9, 公开日 2021 年 7 月 30 日，分子量 23.8KDa)、对照天然人胶原蛋白 (Sigma，货号 C7774)及牛血清白蛋白（BSA，购自生工生物(上海)股份有限公司)溶解(超纯水或 1M HCl 溶液),以 UV 蛋白定量经验公式： $C(mg/mL)=0.144*(A_{215}-A_{225})$ 测定蛋白浓度，再以 PBS（pH 7.4）稀释至 0.5mg/mL。

向 96 孔细胞培养板中加入 100μL 各种蛋白溶液和空白 PBS 溶液对照，室温静置 60min；再向每孔中加入 105 个培养状态良好的 NIH/3T3 细胞，37℃、5%CO₂ 孵育 60min。以 PBS 清洗 4 次孔中细胞。使用 LDH 检测试剂盒(Roche，04744926001)检测 OD492nm 的吸光度值，分析数据并进行显著性差异分析（SPSS 22 软件，Ducan 法，P<0.05）。

OD492nm 的吸光度相应的表征出胶原蛋白样品的细胞粘附活性：黏附活性越高，说明蛋白粘附的细胞越多，胶原蛋白能在短时间内帮助细胞贴壁或粘附于细胞外基质之上，更利于构建最佳的细胞外环境。如图 26 所示，重组小分子胶原蛋白 3A5D29N、3A5D2NT、3A5D1NT、17S1NNT、17S3NT、17S28 的细胞粘附活性均不低或优于天然人胶原蛋白；同样浓度下 3A5D29N、3A5D2NT、3A5D1NT 三种重组小分子胶原蛋白的细胞粘附活显著优于更大分子量的重组Ⅲ型胶原蛋白，同样浓度下 17S1NNT、17S3NT、17S28 的细胞粘附活与更大分子量的重组 XVII 型胶原蛋白无显著差异。这说明，虽然重组小分子胶原蛋白的氨基酸残基数目更少、序

列更短，但它们的氨基酸序列设计实现了分子量小（即序列短）与保持生物学活性（即保持一定的生物学活性位点序列）之间的平衡。

（2）细胞增殖检测

重组小分子胶原蛋白的细胞增殖活检测方法参考 2020 版《中华人民共和国药典》“三部通则 3528 人表皮生长因子生物学活性测定法”。

具体实施方法简述：

取重组小分子胶原蛋白 3A5D29N、17S28 冻干品溶于纯水，调节 pH 值于 7.2-7.4，0.22 μ m 滤膜过滤除菌，以无菌的维持培养液（0.5%FBS+95.5%DMEM(H)）稀释至 1mg/mL、0.5mg/mL、0.25mg/mL、0.125mg/mL；以 hEGF（购自 Lonza 公司）为阳性对照，设置浓度为 1ng/mL、0.5ng/mL、0.25ng/mL、0.125ng/mL，以维持培养液为阴性对照。取 HaCat 细胞株（购自中国科学院细胞库，货号 SCSP-5091）用完全培养液（88%DMEM+10%FBS+1%谷氨酰胺+1%丙酮酸钠）于 37℃、5%二氧化碳条件下培养，控制细胞浓度为每 1mL 含 $1.0\times 10^5\sim 5.0\times 10^5$ 个细胞，传代后 24~36 小时用于生物学活性测定。弃去培养瓶中的培养液，消化和收集细胞，用完全培养液配成每 1mL 含 5.0×10^4 个细胞的细胞悬液，接种于 96 孔细胞培养板中，每孔 100 μ L，于 37℃、5%CO₂ 条件下培养。24 小时后换成维持培养液，于 37℃、5%二氧化碳条件下培养 24 小时。制备的细胞培养板弃去维持培养液，加入对照样品溶液和供试品溶液，每孔 100 μ L，于 37℃、5%CO₂ 条件下培养 24~72 小时。取出 96 孔板，在显微镜下观察细胞形态，然后去除液体，每孔加入 50 μ L MTT 溶液（购自上海碧云天生物技术有限公司，以 DPBS 配置，浓度为 1mg/mL），置 37℃，5% CO₂ 培养箱中培养。4 小时后去除上清，每孔加入 100 μ L 异丙醇溶解结晶，在酶标仪上以 650nm 为参比波长，在波长 570nm 处测定吸光度值，记录测定结果。

如图 27 所示，作为阳性对照的 hEGF 有促 HaCat 细胞增殖的作用，重组小分子胶原蛋白 3A5D29N、17S28 在 0.125~1mg/mL 的浓度下也具有促 HaCat 细胞增殖的作用，尤其是 17S28 促 HaCat 细胞增殖活性呈现随样品浓度增长而增长的趋势，较高浓度的 17S28 与 0.5ng/mL hEGF 的促增殖效果基本相当。

（3）透皮吸收性检测

以全层皮肤模型作为载体，对检测样品以 FITC 荧光标记后，采用表面给药方式，将样品均匀涂抹于模型表面，通过统计荧光切片荧光强度、荧光酶标仪检测接收液计算扩散百分比，观察样品的皮肤渗透行为。委托陕西博溪通用检测科技有限公司完成检测。

具体实施方法简述：

以 FITC（异硫氰酸荧光素）标记重组小分子胶原蛋白 3A5D29N、重组Ⅲ型胶原蛋白（江苏创健医疗科技股份有限公司，专利申请号 CN201310033299.6，公开日 2013 年 5 月 15 日，分子

量 43.6kDa)、水解透明质酸(分子量 ≤ 5000 Da)、鱼皮胶原(Nitta Gelatin Inc.Maringen SP03(PF)), 分子筛纯化去除未偶联的 FITC。将全层皮肤模型组织 FulKutis®(广东博溪生物科技有限公司) 组织转移到 6 孔板中, 每孔加入 2mL 的 3D 全层皮肤模型培养液。

根据测试分组, 进行不同浓度的样品配置。样品组采用表面给药, 给药体积为 20 μ L。培养至合适时间时, 可用无菌 PBS 溶液的洗瓶清洗模型表面残留的受试物, 用无菌棉签轻轻拭去模型内、外残留液体, 环切模型后浸于 4%多聚甲醛溶液中固定(固定时间 ≥ 24 h), 冷冻切片后荧光拍照, 使用 IPP 软件对图片中的目的物质进行累积光密度值(IOD)的数据统计, 组间比较采用 t-test 统计分析(双尾)。收集培养液, 进行荧光 IOD 值检测, 依据不同样品的荧光标准曲线, 计算样品 12h 时渗透至培养液中的样品量, $\text{扩散百分比} = \text{样品量} / \text{样品的理论含量} \times 100\%$, 即累积透皮率。

如图 28 所示, 1h 时 1mg/mL 的重组小分子胶原蛋白 3A5D29N 和重组 III 型胶原蛋白已于全皮肤模型中检测到荧光信号, 而相同浓度的水解透明质酸、鱼皮胶原则未检测到荧光信号, 此时 FITC 标记的重组小分子胶原蛋白样品已渗透进入表皮, 进入皮肤但尚未完全穿透全层皮肤模型组织, 荧光信号强度分析显示, 重组小分子胶原蛋白 3A5D29N 的荧光信号极显著强于其余三种样品, 说明其透皮速度极显著优于重组 III 型胶原蛋白、水解透明质酸与鱼皮胶原。同时, 如图 29 所示, 12h 时, 重组小分子胶原蛋白样品已完全穿透全层皮肤模型组织, 随着给药浓度的增加, 重组小分子胶原蛋白 3A5D29N 的扩散百分率(即累积透皮率)不断增加, 且在 1mg/mL 时与相同浓度水解透明质酸的扩散百分率(即累积透皮率)基本一致, 均超过 40%。以上结果说明: 重组小分子胶原蛋白可在短时间内实现快速透皮且较长时间内有较高的累积透皮率。本实施例中全层皮肤模型组织实验结果验证了本发明的小分子胶原蛋白具有良好的皮肤渗透性能, 这也有助于重组小分子胶原在相关的产品开发, 如在化妆品方面的应用有着极为广阔的前景。

PCT

(原件为电子形式)

0-1	PCT/RO/134表 有关保藏的微生物或其他生物材料的说明 (PCT细则第13条之二)	
0-1-1	软件版本	i-system Version 1.0.26 20240201 MT/FOP 20140331/0.20.5.21
0-2	国际申请号	PCT/CN2024/082594
0-3	申请人或代理人的档案号	ZZLPCT230098
1	下面的说明与本申请说明书中此处提到的 保藏的微生物或其他生物材料相关:	
1-1	页码	3
1-2	行号:	11
1-3	保藏事项	
1-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
1-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物 研究所 100101
1-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
1-3-4	保藏号	CGMCC No.25823
1-4	补充说明	
1-5	本说明是对下列指定国	所有指定国
1-6	单独提交的说明 这些说明将随后提交给国际局	
2	下面的说明与本申请说明书中此处提到的 保藏的微生物或其他生物材料相关:	
2-1	页码	3
2-2	行号:	11
2-3	保藏事项	
2-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
2-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物 研究所 100101
2-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
2-3-4	保藏号	CGMCC No.25811
2-4	补充说明	
2-5	本说明是对下列指定国	所有指定国
2-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

3	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
3-1	页码	3
3-2	行号:	11
3-3	保藏事项	
3-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
3-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
3-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
3-3-4	保藏号	CGMCC No.25824
3-4	补充说明	
3-5	本说明是对下列指定国	所有指定国
3-6	单独提交的说明 这些说明将随后提交给国际局	
4	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
4-1	页码	3
4-2	行号:	11
4-3	保藏事项	
4-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
4-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
4-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
4-3-4	保藏号	CGMCC No.25812
4-4	补充说明	
4-5	本说明是对下列指定国	所有指定国
4-6	单独提交的说明 这些说明将随后提交给国际局	
5	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
5-1	页码	3
5-2	行号:	12
5-3	保藏事项	
5-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
5-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
5-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
5-3-4	保藏号	CGMCC No.25825
5-4	补充说明	
5-5	本说明是对下列指定国	所有指定国
5-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

6	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
6-1	页码	3
6-2	行号:	12
6-3	保藏事项	
6-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
6-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
6-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
6-3-4	保藏号	CGMCC No.25813
6-4	补充说明	
6-5	本说明是对下列指定国	所有指定国
6-6	单独提交的说明 这些说明将随后提交给国际局	
7	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
7-1	页码	3
7-2	行号:	12
7-3	保藏事项	
7-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
7-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
7-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
7-3-4	保藏号	CGMCC No.25826
7-4	补充说明	
7-5	本说明是对下列指定国	所有指定国
7-6	单独提交的说明 这些说明将随后提交给国际局	
8	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
8-1	页码	3
8-2	行号:	12
8-3	保藏事项	
8-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
8-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
8-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
8-3-4	保藏号	CGMCC No. 25814
8-4	补充说明	
8-5	本说明是对下列指定国	所有指定国
8-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

9	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
9-1	页码	6
9-2	行号:	17
9-3	保藏事项	
9-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
9-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
9-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
9-3-4	保藏号	CGMCC No.25815
9-4	补充说明	
9-5	本说明是对下列指定国	所有指定国
9-6	单独提交的说明 这些说明将随后提交给国际局	
10	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
10-1	页码	6
10-2	行号:	17
10-3	保藏事项	
10-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
10-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
10-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
10-3-4	保藏号	CGMCC No.25816
10-4	补充说明	
10-5	本说明是对下列指定国	所有指定国
10-6	单独提交的说明 这些说明将随后提交给国际局	
11	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
11-1	页码	6
11-2	行号:	17
11-3	保藏事项	
11-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
11-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
11-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
11-3-4	保藏号	CGMCC No.25817
11-4	补充说明	
11-5	本说明是对下列指定国	所有指定国
11-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

12	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
12-1	页码	6
12-2	行号:	18
12-3	保藏事项	
12-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
12-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
12-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
12-3-4	保藏号	CGMCC No.25818
12-4	补充说明	
12-5	本说明是对下列指定国	所有指定国
12-6	单独提交的说明 这些说明将随后提交给国际局	
13	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
13-1	页码	4
13-2	行号:	6
13-3	保藏事项	
13-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
13-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
13-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
13-3-4	保藏号	CGMCC No.25819
13-4	补充说明	
13-5	本说明是对下列指定国	所有指定国
13-6	单独提交的说明 这些说明将随后提交给国际局	
14	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
14-1	页码	4
14-2	行号:	6
14-3	保藏事项	
14-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
14-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
14-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
14-3-4	保藏号	CGMCC No.25821
14-4	补充说明	
14-5	本说明是对下列指定国	所有指定国
14-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

15	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
15-1	页码	4
15-2	行号:	6
15-3	保藏事项	
15-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
15-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
15-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
15-3-4	保藏号	CGMCC No.25827
15-4	补充说明	
15-5	本说明是对下列指定国	所有指定国
15-6	单独提交的说明 这些说明将随后提交给国际局	
16	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
16-1	页码	4
16-2	行号:	7
16-3	保藏事项	
16-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
16-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
16-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
16-3-4	保藏号	CGMCC No.25829
16-4	补充说明	
16-5	本说明是对下列指定国	所有指定国
16-6	单独提交的说明 这些说明将随后提交给国际局	
17	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
17-1	页码	4
17-2	行号:	7
17-3	保藏事项	
17-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
17-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
17-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
17-3-4	保藏号	CGMCC No.25828
17-4	补充说明	
17-5	本说明是对下列指定国	所有指定国
17-6	单独提交的说明 这些说明将随后提交给国际局	

PCT

(原件为电子形式)

18	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
18-1	页码	4
18-2	行号:	7
18-3	保藏事项	
18-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
18-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
18-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
18-3-4	保藏号	CGMCC No.25820
18-4	补充说明	
18-5	本说明是对下列指定国	所有指定国
18-6	单独提交的说明 这些说明将随后提交给国际局	
19	下面的说明与本申请说明书中此处提到的保藏的微生物或其他生物材料相关:	
19-1	页码	4
19-2	行号:	7
19-3	保藏事项	
19-3-1	保藏单位名称	CGMCC 中国普通微生物菌种保藏管理中心
19-3-2	保藏单位地址	中国北京市朝阳区北辰西路1号院3号中国科学院微生物研究所 100101
19-3-3	保藏日期	2022年 9月 26日 (26.09.2022)
19-3-4	保藏号	CGMCC No.25822
19-4	补充说明	
19-5	本说明是对下列指定国	所有指定国
19-6	单独提交的说明 这些说明将随后提交给国际局	

由受理局填写

0-4	本表格与国际申请一起收到: (是或否)	
0-4-1	授权官员	

由国际局填写

0-5	国际局收到本表格日期:	
0-5-1	授权官员	

权利要求书

1. 一种重组小分子胶原蛋白表达系统，其特征在于，所述表达系统包括定位融合功能蛋白连接载体转化宿主菌得到的底盘细胞或底盘工程菌和重组小分子胶原蛋白的串联重复表达序列；

其中所述串联重复表达序列以重组小分子胶原蛋白、或人工设计的具有典型的 G-X-Y 三联体结构的胶原蛋白、或以人胶原蛋白序列的两个或多个区域组合的胶原蛋白为基础单元串联重复；所述串联重复表达序列中每相邻的 2 个基础单元之间有 Kex2 酶、CPB 酶的识别、切割的位点；

所述定位融合功能蛋白包括 CPB 酶、有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域；所述定位融合功能蛋白还包括连接序列，所述连接序列用于 CPB 酶与有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列融合表达时的连接；

所述定位融合功能蛋白为如 SEQ ID NO.22、SEQ ID NO.24、SEQ ID NO.26 或 SEQ ID NO.28 所示的氨基酸序列；

所述底盘细胞或底盘工程菌选自酵母。

2. 根据权利要求 1 所述的表达系统，其特征在于，所述串联重复表达序列中每相邻的 2 个基础单元之间还可包括 Ste13 酶的识别、切割的位点。

3. 根据权利要求 2 所述的表达系统，其特征在于，所述 Kex2 酶识别切割的位点包括 KR 或 RR 双碱性氨基酸残基，双碱性氨基酸残基之后，是 EA、EAEA 或其它有助于 Kex2 酶或 Ste 13 酶识别和切割的氨基酸残基。

4. 根据权利要求 1 所述的表达系统，其特征在于，所述 CPB 酶识别和切割的位点包括蛋白羧基末端的碱性氨基酸残基，包括 K、R。

5. 根据权利要求 1~4 任一项所述的表达系统，其特征在于，所述重组小分子胶原蛋白为 SEQ ID NO.1、SEQ ID NO.3、SEQ ID NO.5、SEQ ID NO.7、SEQ ID NO.63 或 SEQ ID NO.67 所示的氨基酸序列。

6. 根据权利要求 5 所述的表达系统，其特征在于，所述重组小分子胶原蛋白的羧基端有多个 His。

7. 根据权利要求 6 所述的表达系统，其特征在于，所述重组小分子胶原蛋白为 SEQ ID NO.2、SEQ ID NO.4、SEQ ID NO.6、SEQ ID NO.8 所示的氨基酸序列。

8. 根据权利要求 1 所述的表达系统，其特征在于，所述串联重复表达序列为 SEQ ID NO.30、SEQ ID NO.33、SEQ ID NO.36、SEQ ID NO.39、SEQ ID NO.42、SEQ ID NO.45、SEQ ID NO.48、SEQ ID NO.51、SEQ ID NO.54、SEQ ID NO.57、SEQ ID NO.60、SEQ ID NO.64 或 SEQ ID NO.68 所示序列。

9. 根据权利要求 8 所述的表达系统，其特征在于，编码所述串联重复表达序列的核酸为 SEQ ID NO.71-81 所示序列、或其简并序列。

10. 根据权利要求 1 所述的表达系统，其特征在于，所述 CPB 酶来源于人或大鼠。

11. 根据权利要求 10 所述的表达系统，其特征在于，所述 CPB 酶序列如 SEQ ID NO.17-18 所示。

12. 根据权利要求 1 所述的表达系统，其特征在于，所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列来源于酿酒酵母或毕赤酵母的 Kex2 酶。

13. 根据权利要求 12 所述的表达系统，其特征在于，所述有细胞内膜定位或各细胞器之间的转换与运输作用的功能区域序列如 SEQ ID NO.19-20 所示。

14. 根据权利要求 1 所述的表达系统，其特征在于，所述连接序列为 Linker 序列，如 SEQ ID NO.21 所示：GGSGSGSGGS。

15 根据权利要求 1 所述的表达系统，其特征在于，编码所述定位融合功能蛋白的核酸为 SEQ ID NO.23、SEQ ID NO.25、SEQ ID NO.27 或 SEQ ID NO.29 所示的核苷酸序列、或其简并序列。

16. 根据权利要求 1 所述的表达系统，其特征在于，所述底盘细胞或底盘工程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心，保藏编号为：CGMCC No.25815、CGMCC No.25817、CGMCC No.25816、CGMCC No.25818。

17. 权利要求 1-16 任一项所述的表达系统在获取重组小分子胶原蛋白中的应用。

18. 一种制备重组小分子胶原蛋白的方法，其特征在于，所述方法采用权利要求 1-16 任一项所述的表达系统，所述方法包括：

构建重组小分子胶原蛋白的串联重复表达序列；所述串联重复表达序列以重组小分子胶原蛋白、或人工设计的具有典型的 G-X-Y 三联体结构的胶原蛋白、或以人胶原蛋白序列的两个或多个区域组合的胶原蛋白为基础单元串联重复；所述串联重复表达序列中每相邻的 2 个基础单元之间有 Kex2 酶、CPB 酶的识别、切割的位点；

构建定位融合功能蛋白，将定位融合功能蛋白连接载体转化宿主菌得到底盘细胞或底盘工程菌；

将重组小分子胶原蛋白的串联重复表达序列连接到表达载体后转入底盘细胞或底盘工程菌中，得到包含或表达重组小分子胶原蛋白的重组工程菌，发酵诱导表达，得到重组小分子胶原蛋白。

19. 根据权利要求 18 所述的方法，其特征在于，所述包含或表达重组小分子胶原蛋白的重组工程菌保藏于中国微生物菌种保藏管理委员会普通微生物中心，保藏编号为：CGMCC

No.25819、CGMCC No.25821、CGMCC No.25827、CGMCC No.25829、CGMCC No.25828、CGMCC No.25820、CGMCC No.25822。

20. 根据权利要求 18 所述的方法，其特征在于，所述方法得到的重组小分子胶原蛋白为 SEQ ID NO.32、SEQ ID NO.35、SEQ ID NO.38、SEQ ID NO.41、SEQ ID NO.44、SEQ ID NO.47、SEQ ID NO.50、SEQ ID NO.53、SEQ ID NO.56、SEQ ID NO.59、SEQ ID NO.62、SEQ ID NO.66 或 SEQ ID NO.70 所示的氨基酸序列。

21. 根据权利要求 18-20 任一项所述的方法得到的重组小分子胶原蛋白。

22. 权利要求 18-20 任一项所述的方法得到的重组小分子胶原蛋白、或权利要求 21 所述重组小分子胶原蛋白在制备生物材料、组织工程产品、化妆品中的用途。