



(12) 发明专利

(10) 授权公告号 CN 101395178 B

(45) 授权公告日 2013. 05. 01

(21) 申请号 200680053590. 8

(22) 申请日 2006. 12. 28

(30) 优先权数据

0502932-7 2005. 12. 30 SE

(85) PCT申请进入国家阶段日

2008. 08. 29

(86) PCT申请的申请数据

PCT/SE2006/001505 2006. 12. 28

(87) PCT申请的公布数据

W02007/078239 EN 2007. 07. 12

(73) 专利权人 思百博技术股份公司

地址 瑞典乌普萨拉

(72) 发明人 J·约翰松 G·耶尔姆 M·斯塔克

A·里辛 S·格里普 W·恩斯特伦

M·赫德哈马尔

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 罗菊华

(51) Int. Cl.

C07K 14/435 (2006. 01)

C12N 15/62 (2006. 01)

D01F 4/02 (2006. 01)

(56) 对比文件

SPONNER ET AL..The conserved C-termini contribute to the properties of spider silk fibroins. 《Biochemical and biophysical research communications》.2005, 第 338 卷 (第 2 期),

Natalia N. Pouchkina-Stantcheva et al.. Molecular studies of a novel dragline silk from a nursery web spider, Euprosthops sp.. 《Comparative biochemistry and physiology part b biochemistry & molecular biology》.2004, 第 138 卷 (第 4 期),

Anna Rising et al..N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriiform spider silk proteins. 《Biomacromolecules》.2006, 第 7 卷 (第 11 期),

LEWIS RANDOLPH V ET AL..Expression and Purification of a Spider Silk Protein A New Strategy for Producing Repetitive Proteins:A NEW strategy for producing repetitive proteins.《PROTEIN EXPRESSION AND PURIFICATION》.1996, 第 7 卷 (第 4 期),

审查员 张丽玮

权利要求书1页 说明书23页

序列表47页 附图7页

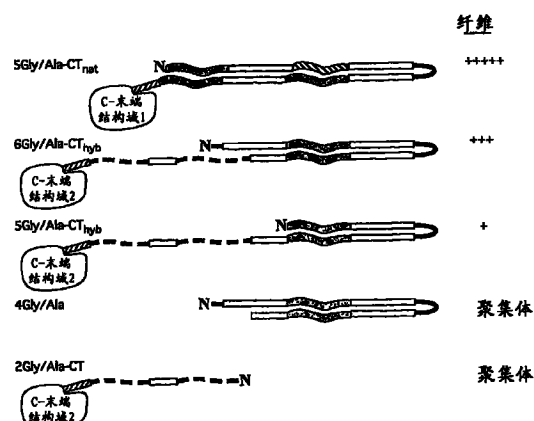
(54) 发明名称

蜘蛛丝蛋白和用于产生蜘蛛丝蛋白的方法

(57) 摘要

本发明提供了分离的大壶腹腺蛛丝蛋白,其由 150 至 420 个氨基酸残基组成并且由式 KEP-CP 定义。KEP 是重复的、具有 80 至 300 个氨基酸残基的 N-末端衍生的蛋白质片段。CT 是具有 70 至 120 个氨基酸残基的 C-末端衍生的蛋白质片段。本发明进一步提供了分离的融合蛋白,其由作为大壶腹腺蛛丝蛋白的第一蛋白质片段和包含融合伙伴及切割试剂识别位点的第二蛋白质片段组成。第一蛋白质片段通过所述切割试剂识别位点偶联至融合伙伴。本发明还提供了用于产生大壶腹腺蛛

丝蛋白及其多聚体的方法。



1. 分离的大壶腹腺蛛丝蛋白, 其为 SEQ ID NO :9。
2. 选自“X- 大壶腹腺蛛丝蛋白”和“大壶腹腺蛛丝蛋白 -X”的分离的融合蛋白, 其中所述大壶腹腺蛛丝蛋白是权利要求 1 的分离的大壶腹腺蛛丝蛋白 ; 和 X 是包含融合伙伴和切割试剂识别位点的蛋白质片段 ; 其中所述大壶腹腺蛛丝蛋白通过所述切割试剂识别位点偶联至所述融合伙伴。
3. 权利要求 1 的分离的大壶腹腺蛛丝蛋白的多聚体。
4. 一种分离的核酸分子, 其由选自下列的核酸序列组成 :SEQ IDNO :1 ; 编码 SEQ ID NO :9 的核酸序列, 和其互补核酸序列 ; 以及编码权利要求 2 的融合蛋白的核酸序列, 和其互补核酸序列。
5. 用于产生权利要求 2 的分离的融合蛋白的方法, 其包括下列步骤 :
 - (i) 在合适的宿主中表达编码权利要求 2 的融合蛋白的核酸分子 ; 和
 - (ii) 分离在步骤 (i) 中获得的融合蛋白。
6. 权利要求 5 的方法, 其中步骤 (ii) 包括除去 LPS 和其他致热原。
7. 用于产生权利要求 1 中定义的大壶腹腺蛛丝蛋白的多聚体的方法, 其包括下列步骤 :
 - (i) 提供权利要求 2 的融合蛋白在液体介质中的溶液,
 - (ii) 向所述液体介质中加入合适的切割试剂, 所述切割试剂用于在切割试剂识别位点处实现融合蛋白的切割, 从而获得大壶腹腺蛛丝蛋白 ;
 - (iii) 允许在步骤 (ii) 中获得的大壶腹腺蛛丝蛋白在液体介质中聚合。
8. 权利要求 7 的方法, 其进一步包括 : (iv) 从所述液体介质中分离在步骤 (iii) 中获得的多聚体。
9. 权利要求 8 的方法, 其中步骤 (iv) 包括除去 LPS 和其他致热原。
10. 权利要求 7 的用于产生权利要求 1 中定义的大壶腹腺蛛丝蛋白的多聚体的方法, 其中提供权利要求 2 的融合蛋白的溶液的步骤 (i) 包括 :
 - (a) 在合适的宿主中表达编码权利要求 2 的融合蛋白的核酸分子 ;
 - (b) 分离在步骤 (a) 中获得的融合蛋白 ; 和
 - (c) 提供在步骤 (b) 中获得的所述融合蛋白在液体介质中的溶液。
11. 权利要求 10 的方法, 其中步骤 (b) 包括除去 LPS 和其他致热原。
12. 权利要求 7-11 中任一项的方法, 其中允许在步骤 (ii) 中获得的大壶腹腺蛛丝蛋白在液体介质中聚合的所述步骤 (iii) 还包括, 提供所述液体介质与选自气相、液相和固相的另一相之间的界面, 其中所述聚合在所述界面处或在围绕所述界面的区域中开始。
13. 权利要求 12 的方法, 其中所述液体介质是水性介质, 并且所述另一相选自空气和与水不能混溶的有机溶剂。
14. 权利要求 4 的分离的核酸分子在制备编码蜘蛛丝蛋白的非天然基因中的用途。

蜘蛛丝蛋白和用于产生蜘蛛丝蛋白的方法

[0001] 本发明的技术领域

[0002] 本发明涉及蛋白质的重组产生的领域。更特别地,本发明涉及蜘蛛丝蛋白的重组产生。本发明提供了新的分离的大壶腹腺蛛丝蛋白 (major ampullate spidroin) 和大壶腹腺蛛丝蛋白融合蛋白,以及用于产生此类蛋白的方法和多聚核酸分子。还提供了大壶腹腺蛛丝蛋白的多聚体和用于产生此类多聚体的方法。

[0003] 发明背景

[0004] 蜘蛛丝是天然的高性能多聚体,由于强度和弹性的组合,其获得了极其高的韧性。在蜘蛛中存在高达 7 个专门化的腺体,所述腺体产生多种具有不同机械性质和功能的丝纤维类型。由大壶腹腺 (major ampullate gland) 产生的拖丝 (dragline silk) 是韧性最强的纤维,并且基于重量其性能优于人造材料例如高强度钢和芳纶 (Kevlar)。拖丝的性质在开发用于医疗或技术目的的新材料中是非常吸引人的。

[0005] 拖丝由两种主要的多肽 (通常称为大壶腹腺蛛丝蛋白 (MaSp) 1 和 2,但在十字园蛛 (*Araneus diadematus*) 中称为 ADF-3 和 ADF-4) 组成。取决于分析的取样年龄 (sample age) 和条件,这些蛋白质具有 200 至 720kDa 的表观分子量,但至今未报导过全长蜘蛛拖丝基因。在 Huebnerich, D. 等人, Novel assembly properties of recombinant spider dragline silk proteins. *Curr. Biol.* 14, 2070-2074 (2004) 中描述了拖丝多肽的性质。已知的拖丝蛛丝蛋白 (spidroin) 由交替出现的富含丙氨酸的区段 (其在纤维中形成结晶 β -折叠) 和富含甘氨酸的区段 (其更柔韧并且大部分缺乏有序结构) 的高度重复的嵌段 (block) 组成。C-末端区域是非重复的,是物种间高度保守的,并且采取 α -螺旋构象。拖丝蛋白的 N-末端区域直至最近才得到表征,其中揭示出了在不同蛛丝蛋白之间并且还在不同蜘蛛种类之间高度保守的 N-末端结构域 (Rising, A. 等人, N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindrical spider silk proteins. *Biomacromolecules* 7, 3120-3124 (2006))。

[0006] 拖丝的机械性质在物种之间是变化的; *Euprosthenops* sp 的拖丝比来自例如十字园蛛或棒络新妇蛛 (*Nephila clavipes*) 的拖丝更硬、更强 (需要更大的力才能扯断) 和可延伸性更低。来自 *Euprosthenops* sp 的拖丝看起来具有比来自十字园蛛的拖丝更大比例的结晶 β -折叠结构,这很可能是由于 *Euprosthenops* sp MaSp 在到目前为止所分析的所有种类中具有最高的聚丙氨酸含量 (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthenops* sp. (Pisauridae). *Comp Biochem Physiol B Biochem Mol Biol* 138, 371-376 (2004))。

[0007] 产生人造蜘蛛丝的尝试已使用了天然或合成的编码拖丝蛋白的基因片段,因为至今还未曾报导过全长基因。已在各种不同的系统 (包括细菌、酵母、哺乳动物细胞、植物、昆虫细胞、转基因蚕和转基因山羊) 中表达重组拖丝蛋白。参见,例如 Lewis, R. V. 等人, Expression and purification of a spider silk proteins: a new strategy for producing repetitive proteins. *Protein Expr. Purif.* 7, 400-406 (1996);

Fahnestock, S. R. & Irwin, S. L. Synthetic spiderdragline silk proteins and their production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 47, 23-32 (1997); Arcidiacono, S. 等人, Purification and characterization of recombinant spidersilk expressed in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 49, 31-38 (1998); Fahnestock, S. R. & Bedzyk, L. A. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 47, 33-39 (1997); 和 Lazaris, A. 等人, Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-476 (2002)。

[0008] W02004/016651 (The University of York) 公开了编码来自 *Euprostenopss* 的 MaSp1 蛋白的内部重复部分的核酸序列。没有表达蛋白质。

[0009] Huemmerich, D. 等人, Primary structure elements of spiderdragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004) 公开了合成基因“(AQ)₁₂NR3”, 其编码重复的富含 Ala 和富含 Gly/Gln 的片段以及非重复片段(所有片段都来源于来自园蛛属 (*Araneus*) 的 ADF3)。该基因被表达成可溶性蛋白质(59.8kD, >528 个氨基酸), 所述蛋白质聚集但不形成多聚体或纤维。该蛋白质的丙氨酸含量是 10-15%。

[0010] W003/057727 公开了可溶性重组丝多肽在哺乳动物细胞系和动物中的表达。一种经表达的丝多肽(ADF-3; 60kD, 652 个氨基酸)由重复单元和非重复亲水性结构域组成。另一种经表达的丝多肽(ADF-3His; 63kD, 677 个氨基酸)由重复单元、非重复亲水性结构域、c-myc 表位和六组氨酸标签组成。所述重复单元具有低含量的 Ala (10-20%)。所获得的丝多肽在水性介质中展现出差的溶解性和/或形成沉淀。因为所获得的丝多肽不会自发地聚合, 因而需要进行纺丝 (spinning) 以获得多聚体或纤维。

[0011] 几个因素使拖丝蛋白的表达变得复杂。由于基因的高度重复的性质以及相伴随的所述蛋白质的受限制的氨基酸组成, 因而发生转录和翻译错误。另一个原因可能是微生物表达系统中 tRNA 库的耗竭, 以及随后的不连续翻译, 从而导致蛋白质合成的过早终止。关于蛋白质合成的截短所讨论的其他原因是 mRNA 的二级结构的形成以及基因的重组。大于 2.5kb 的天然 MaSp 已显示在细菌宿主中是不稳定的。此外, 在以可溶形式保持重组丝蛋白方面存在许多困难, 因为天然来源的拖丝蛋白片段以及经设计的嵌段共聚物特别是来源于 MaSp1/ADF-4 的蛋白, 容易地自装配成无定形聚集体, 从而引起蛋白质的沉淀和损失。参见 Huemmerich, D. 等人, Primary structure elements of spiderdragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004); 和 Lazaris, A. 等人, Spidersilk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-476 (2002)。

[0012] 发明概述

[0013] 本发明的目的是提供新的蜘蛛丝蛋白, 其可提供蜘蛛丝纤维。

[0014] 本发明的另一个目的是提供水溶性蜘蛛丝蛋白, 其可容易地进行操作以便按希望地自体聚合成纤维。这允许独特的应用, 例如在纤维上培养真核细胞。此外, 该性质允许在生理条件下进行所有下列步骤, 这减少了毒性和蛋白质变性的风险。

[0015] 本发明的另外一个目的是提供新的蜘蛛丝蛋白的纤维。

[0016] 本发明的一个目的是以大规模提供蜘蛛丝蛋白,其可容易地进行操作以便按希望地自体聚合成纤维。

[0017] 本发明的目的还在于提供用于产生丝蛋白和蜘蛛丝蛋白的纤维的方法。

[0018] 关于根据下列公开内容将变得显而易见的这些和其他目的,根据一个方面,本发明提供了分离的大壶腹腺蛛丝蛋白或其衍生物,其中所述蛋白质由 150 至 420 个氨基酸残基组成并且由式 REP-CT 定义,其中 REP 是具有 80 至 300 个氨基酸残基的蛋白质片段,其中所述片段选自 $L(AG)_nL$ 、 $L(AG)_nAL$ 、 $L(GA)_nL$ 、 $L(GA)_nGL$,其中 n 是 4 至 8 的整数;每个独个 A 区段是 8 至 18 个氨基酸残基的氨基酸序列,其中所述氨基酸残基中的 0 至 3 个不是 Ala,而其余的氨基酸残基是 Ala;每个独个 G 区段是 12 至 30 个氨基酸残基的氨基酸序列,其中至少 40% 的氨基酸残基是 Gly;和每个独个 L 区段是 0 至 20 个氨基酸残基的连接体氨基酸序列;以及 CT 是具有 70 至 120 个氨基酸残基的蛋白质片段,该片段是源自大壶腹腺蛛丝蛋白的 C-末端片段。

[0019] 本发明基于鉴定足以形成丝样纤维的蛋白质基元以及所述基元用于构建重组 MaSp 蛋白的用途,所述重组 MaSp 蛋白可能在合适的宿主,例如细菌,优选大肠杆菌 (*E. coli*) 中产生。

[0020] 在本发明的某些实施方案中,每个独个 A 区段与选自下列的氨基酸序列具有至少 80% 的同一性:SEQ ID NO:3 的氨基酸残基 7-19、43-56、71-83、107-120、135-147、171-183、198-211、235-248、266-279、294-306、330-342、357-370、394-406、421-434、458-470、489-502、517-529、553-566、581-594、618-630、648-661、676-688、712-725、740-752、776-789、804-816、840-853、868-880、904-917、932-945、969-981、999-1013、1028-1042 和 1060-1073;SEQ ID NO:9 的氨基酸残基 31-42、61-75、90-104、122-135 和 153-171;SEQ ID NO:13 的氨基酸残基 12-25、46-60、75-88、112-119、150-158 和 173-180;SEQ ID NO:14 的氨基酸残基 31-42;和 SEQ ID NO:15 的氨基酸残基 122-135。在特定的实施方案中,每个独个 A 片段是选自该组氨基酸序列的氨基酸序列。

[0021] 在本发明的一些实施方案中,每个独个 G 区段与选自下列的氨基酸序列具有至少 80% 的同一性:SEQ ID NO:3 的氨基酸残基 20-42、57-70、84-106、121-134、148-170、184-197、212-234、249-265、280-293、307-329、343-356、371-393、407-420、435-457、471-488、503-516、530-552、567-580、595-617、631-647、662-675、689-711、726-739、753-775、790-803、817-839、854-867、881-903、918-931、946-968、982-998、1014-1027、1043-1059 和 1074-1092;SEQ ID NO:5;SEQ ID NO:6;SEQ ID NO:7;SEQ ID NO:9 的氨基酸残基 11-30、43-60、76-89、105-121 和 136-152;和 SEQ ID NO:13 的氨基酸残基 1-11、26-45、61-74、89-111、120-149 和 159-172。在特定的实施方案中,每个独个 G 片段与选自该组氨基酸序列的氨基酸序列相同。

[0022] 在本发明的某些实施方案中,所述 CT 片段与 SEQ ID NO:8 具有至少 50% 的同一性,或与选自 SEQ ID NO:4、SEQ ID NO:9 的氨基酸残基 172-269、SEQ ID NO:13 的氨基酸残基 181-276 和 SEQ ID NO:16 的氨基酸残基 172-269 以及图 3 的任何氨基酸序列(特别是图 3 的 MaSp1 序列)的氨基酸序列具有至少 80% 的同一性。在特定的实施方案中,所述 CT 片段是选自改组氨基酸序列的氨基酸序列。

[0023] 在本发明的某些实施方案中,在所述分离的大壶腹腺蛛丝蛋白中脂多糖 (LPS) 和

其他致热原的含量是 1 个内毒素单位 (EU)/mg 蛋白质或更低。

[0024] 根据另一个方面,本发明提供了分离的融合蛋白,所述融合蛋白由作为大壶腹腺蛛丝蛋白的第一蛋白质片段和第二蛋白质片段组成,其中所述第二蛋白质片段包含融合伙伴和切割试剂识别位点,其中所述第一蛋白质片段通过所述切割试剂识别位点偶联至所述融合伙伴。

[0025] 本发明提供了选自 X-REP-CT 和 REP-CT-X 的分离的融合蛋白,其中 REP 和 CT 是本发明的蛋白质片段;和 X 是包含融合伙伴和切割试剂识别位点的蛋白质片段;其中组合的蛋白质片段 REP-CT 通过所述切割试剂识别位点偶联至所述融合伙伴。

[0026] 在本发明的某些实施方案中,在所述分离的融合蛋白中 LPS 和其他致热原的含量是 1EU/mg 蛋白质或更低。

[0027] 根据另外一个方面,本发明提供了用于产生本发明的大壶腹腺蛛丝蛋白的方法,其包括下列步骤:(i) 提供本发明的融合蛋白在液体介质中的溶液,(ii) 向所述液体介质中加入合适的切割试剂,所述切割试剂用于在切割试剂识别位点处实现融合蛋白的切割,从而获得大壶腹腺蛛丝蛋白;和任选地,(iii) 从所述液体介质中分离在步骤(ii)中获得的大壶腹腺蛛丝蛋白。

[0028] 本发明还提供了用于产生本发明的大壶腹腺蛛丝蛋白的多聚体的方法,其包括下列步骤:(i) 提供本发明的融合蛋白在液体介质中的溶液,(ii) 向所述液体介质中加入合适的切割试剂,所述切割试剂用于在切割试剂识别位点处实现融合蛋白的切割,从而获得大壶腹腺蛛丝蛋白;(iii) 允许在步骤(ii)中获得的大壶腹腺蛛丝蛋白在液体介质中聚合;和任选地,(iv) 从所述液体介质中分离在步骤(iii)中获得的多聚体。在优选的方法中,所述步骤(iii)还包括提供所述液体介质与选自气相、液相和固相的另一相之间的界面,其中所述聚合在所述界面处或在围绕所述界面的区域中开始。在优选的方法中,所述液体介质是水性介质,并且所述其他相选自空气和与水不能混溶的有机溶剂。

[0029] 根据另一个方面,本发明提供了分离的多聚核酸分子,其包含编码本发明的大壶腹腺蛛丝蛋白的核酸序列或其互补核酸序列。

[0030] 根据另外一个方面,本发明提供了分离的多聚核酸分子,其包含编码本发明的融合蛋白的核酸序列或其互补核酸序列。

[0031] 本发明的另一个方面在于用于产生本发明的可溶性融合蛋白的方法,其包括下列步骤:(i) 在合适的宿主中表达编码本发明的可溶性融合蛋白的多聚核酸分子;和(ii) 分离在步骤(i)中获得的可溶性融合蛋白。任选地,分离可溶性融合蛋白的所述步骤(ii)包括除去 LPS 和其他致热原。

[0032] 本发明还提供了用于产生本发明的大壶腹腺蛛丝蛋白的方法,其包括下列步骤:(i) 在合适的宿主中表达编码本发明的可溶性融合蛋白的多聚核酸分子;(ii) 分离在步骤(i)中获得的可溶性融合蛋白;(iii) 提供在步骤(ii)中获得的所述可溶性融合蛋白在液体介质中的溶液;(iv) 向所述液体介质中加入合适的切割试剂,所述切割试剂用于在切割试剂识别位点处实现融合蛋白的切割,从而获得大壶腹腺蛛丝蛋白;和任选地,(v) 从所述液体介质中分离在步骤(iv)中获得的大壶腹腺蛛丝蛋白。此外,任选地,分离可溶性融合蛋白的所述步骤(ii),和任选地,分离大壶腹腺蛛丝蛋白的所述步骤(v),包括除去 LPS 和其他致热原。

[0033] 本发明还提供了用于产生本发明的大壶腹腺蛛丝蛋白的多聚体的方法,其包括下列步骤:(i) 在合适的宿主中表达编码本发明的可溶性融合蛋白的多聚核酸分子;(ii) 分离在步骤(i)中获得的可溶性融合蛋白;(iii) 提供在步骤(ii)中获得的所述可溶性融合蛋白在液体介质中的溶液;(iv) 向所述液体介质中加入合适的切割试剂,所述切割试剂用于在切割试剂识别位点处实现融合蛋白的切割,从而获得大壶腹腺蛛丝蛋白;(v) 允许在步骤(iv)中获得的大壶腹腺蛛丝蛋白在液体介质中聚合;和任选地,(vi) 从所述液体介质中分离在步骤(v)中获得的多聚体。在优选的方法中,所述步骤(v)还包括提供所述液体介质与选自气相、液相和固相的另一相之间的界面,其中所述聚合在所述界面处或在围绕所述界面的区域中开始。在优选的方法中,所述液体介质是水性介质,并且所述其他相选自空气和与水不能混溶的有机溶剂。

[0034] 根据另一个方面,本发明提供了本发明的大壶腹腺蛛丝蛋白的多聚体。本发明还提供了可通过本发明的方法获得的大壶腹腺蛛丝蛋白的多聚体。在优选的实施方案中,所述多聚体是纤维。在其他优选的实施方案中,所述多聚体形成选自泡沫、凝胶、网状物或薄膜的结构。

[0035] 根据另外一个方面,本发明提供了包含融合伙伴和切割试剂识别位点的蛋白质片段在制备融合蛋白中的新用途,所述融合蛋白包含通过所述切割试剂识别位点偶联至蜘蛛丝蛋白质片段的所述蛋白质片段。在优选的实施方案中,所述蜘蛛丝蛋白质片段由150至420个氨基酸残基组成。

[0036] 根据最后的方面,本发明提供了分离的多聚核酸分子,其包含选自SEQ ID NO:1和编码SEQ ID NO:2-16的核酸序列的核酸序列,或其互补核酸序列。本发明还提供了所述分离的多聚核酸分子在制备编码蜘蛛丝蛋白的非天然基因中的用途。

[0037] 附图简述

[0038] 图1是在Euprostenops australis MaSp1蛋白即SEQ ID NO:3的重复部分内的区段的比对。

[0039] 图2A图解说明了Euprostenops australis MaSp1蛋白(SEQ ID NO:3)的重复部分的示意性的、预测的结构组织。

[0040] 图2B图解说明了根据实施例5-8构建的蛛丝蛋白(SEQ ID NO:9-13)的示意性的、预测的结构组织。

[0041] 图3是MaSp1和MaSp2的C-末端区域的比对,其中图解说明了它们的保守性质。

[0042] 图4图解说明了从根据实施例5-8构建的蛛丝蛋白形成的纤维的肉眼可见的外观。(A):6Gly/Ala-CT_{hyb}蛋白(SEQ ID NO:13)纤维,标尺0.5cm;(B):5Gly/Ala-CT_{nat}蛋白(SEQ ID NO:9)纤维,标尺1cm;(C):5Gly/Ala-CT_{nat}蛋白(SEQ ID NO:9)纤维,标尺1cm。

[0043] 图5显示了从根据实施例5-8构建的蛛丝蛋白形成的纤维的扫描电子显微术(SEM)显微照片。来自6Gly/Ala-CT_{hyb}(SEQ ID NO:13)的单根纤维(a)和凝胶相(b、c)。在75%甲醇中进行拉伸的、经空气干燥的并放置在SEM-样品台(SEM-stubs)上的5Gly/Ala-CT_{nat}(SEQ ID NO:9)的纤维(d、e、f)。在空气干燥前扭曲的纤维(e),纤维的末端(f)。

[0044] 图6展示了6Gly/Ala-CT_{hyb}(SEQ ID NO:13)纤维的圆二色性(CD)光谱。

[0045] 图7图解说明了来自小鼠肥大细胞毒性研究的结果,所述结果显示了在体外产生的丝纤维存在或不存在的条件下培养3天后,活细胞和死细胞的数目。

[0046] 图 8 是在生物相容性研究中,在暴露于体外产生的丝纤维后 HEK293 细胞的照片。

[0047] 图 9 是应力 (stress)-应变 (strain) 曲线,其展示了来自 5Gly/Ala-CT_{nat} (SEQ ID NO:9) 的双拉伸纤维 (double drawn fiber) 的拉伸强度 (tensile strength)。

[0048] 图 10 显示了来自 5Gly/Ala-CT_{nat} (SEQ ID NO:9) 的重组纤维的 SEM 显微照片。a、b,自发形成的纤维。特写图像 (b) 显示了原纤维亚结构。凸出的原纤维 (箭头) 具有大约 300nm 的宽度。c-f,在 2 轮绷紧-松弛循环后的纤维。c 和 d 显示了不同放大倍数下的相同纤维。e 显示了经切割的纤维末端,和 f 显示了在拉伸试验后的断裂点。

[0049] 发明详述

[0050] 总的来说,本发明基于鉴定足以用于重组产生蜘蛛丝纤维的蛛丝蛋白基元。所述基元基于从来自 *Eupros thenopsaustralis* 的部分主要蛛丝蛋白 1 (major spidroin 1, MaSp1) 的 cDNA 的克隆和测序中推导出的氨基酸序列。由此断定,该分离的 MaSp1 cDNA 可用于构建新的蛛丝蛋白基因 (例如此处报导的那些基因) 的起始点。从由所述新的蛛丝蛋白 cDNA 产生的蛋白质形成的多聚体就它们的物理性质特别是高强度、弹性和轻量度的有用组合而言是有用的。就它们支持细胞粘着和生长的能力而言,它们也是有用的。拖丝的性质在开发用于医疗或技术目的的新材料方面是有吸引力的。特别地,本发明的蜘蛛丝可用于:医疗装置例如植入物,和医疗产品例如用作伤口闭合系统 (wound closure system)、急救绷带 (band-aid)、缝线、伤口敷料 (wound dressing) 以及用于组织工程和受指导的细胞再生的支架。本发明的蜘蛛丝还特别地可用于在例如降落伞、防弹服、座椅安全带等中用作织品或织物。

[0051] 此处所使用的术语“纤维”涉及厚度为至少 1 μ m 的多聚体,优选地涉及肉眼可见的 (macroscopic) 多聚体,所述肉眼可见的多聚体是人眼可见的,即厚度为至少 1 μ m,并且在长度上与其厚度相比具有相当大的延伸,优选大于 5mm。术语“纤维”不包括未组织的聚集体或沉淀。

[0052] 术语“大壶腹腺蛛丝蛋白”、“蛛丝蛋白”在整个本说明书中可互换使用,并且包括所有已知的“大壶腹腺蛛丝蛋白”,通常缩写为“MaSp”,或在十字园蛛的情况下缩写为“ADF”。这些大壶腹腺蛛丝蛋白通常为两种类型,即类型 1 和 2。此外,这些术语还包括所附权利要求书中定义的本发明的新型蛋白质,以及与已知的大壶腹腺蛛丝蛋白具有高度同一性和/或相似性的其他非天然蛋白质。

[0053] 本发明者已利用鉴定出的蛛丝蛋白基元来构建编码非天然蛛丝蛋白的新型基因构建体。已发现,由 150 至 420 个氨基酸残基,即大于或等于 150 个,优选大于或等于 220 个,优选大于或等于 250 个氨基酸残基,并且小于或等于 420 个,优选小于或等于 380 个,优选小于或等于 320,优选小于或等于 280 个氨基酸残基,例如 220-360 个氨基酸残基组成的大壶腹腺蛛丝蛋白可以例如在细菌或其他合适的生产生物中重组产生。所得到的蛛丝蛋白自发地形成本发明的肉眼可见的丝纤维。这是令人吃惊的结果,因为天然发生的蛛丝蛋白以及先前已知的、重组产生的、形成纤维的蛛丝蛋白比本发明的蛋白长得多。此外,天然发生的蛛丝蛋白以及先前已知的、重组产生的、形成纤维的蛛丝蛋白倾向于包含大量的内部重复序列并且需要使用纺丝和/或剧烈 (harsh) 溶剂来进行聚合。

[0054] 在此处第一次显示了,蛛丝蛋白可自发地在体外形成纤维。此处提供的数据还显示,只需要提供一部分的蛛丝蛋白序列来指导纤维形成。此外,包含 *Euprosthenops* 重复结

构域和新妇蛛 (Nephila) 非重复 C- 末端结构域的种间杂合体 (参见实施例 6C) 也形成纤维, 这表明该基元的形成纤维的潜能是很强的。

[0055] 在其总的方面, 本发明的大壶腹腺蛛丝蛋白由式 REP-CT 来定义。REP 蛋白片段和 CT 蛋白片段通常通过肽键而共价偶联。

[0056] 蛋白质片段 REP 具有重复的特征, 在富含丙氨酸的片段和富含甘氨酸的片段之间交替。REP 片段通常包含大于 80 个, 例如大于 140 个, 并且小于 300 个, 优选小于 240, 例如小于 200 个氨基酸残基, 并且其本身可分成几个 L (连接体) 区段、A (富含丙氨酸的) 区段和 G (富含甘氨酸的) 区段, 如将在下面更详细地解释的。通常, 所述连接体区段 (其为任选的) 位于 REP 片段末端, 而其余的区段依次为富含丙氨酸的和富含甘氨酸的区段。因此, REP 片段通常可具有下列结构中的任一种, 其中 n 是整数 :

[0057] $L(AG)_nL$, 例如 $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5L$;

[0058] $L(AG)_nAL$, 例如 $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5A_6L$;

[0059] $L(GA)_nL$, 例如 $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5L$; 或

[0060] $L(GA)_nGL$, 例如 $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5G_6L$ 。

[0061] 由此断定, 富含丙氨酸或富含甘氨酸的区段是否与 N- 末端或 C- 末端连接体区段相邻不是至关重要的。优选地, n 是 4 至 8 的整数, 更优选 4 至 6 的整数, 即 $n = 4$ 、 $n = 5$ 或 $n = 6$ 。

[0062] 在优选的实施方案中, 本发明的 REP 片段的丙氨酸含量高于 20%, 优选高于 25%, 更优选高于 30%, 并且低于 50%, 优选低于 40%, 更优选低于 35%。这是有利的, 因为可预见更高的丙氨酸含量提供更硬的和 / 或更强的和 / 或可延伸性更低的纤维。对此的原因可能是, 更高的丙氨酸含量与纤维中更高含量的 β - 折叠结构相关。因此, 在优选的实施方案中, 在本发明的大壶腹腺蛛丝蛋白的多聚体例如纤维中的 β - 折叠含量高于 50%, 即该蛋白质的超过 50% 的二级结构是 β - 折叠形式。

[0063] 在某些实施方案中, REP 片段缺乏脯氨酸残基, 即在 REP 片段中没有 Pro 残基。

[0064] 现在转向构成本发明的 REP 片段的区段, 将要强调的是, 每个区段是独个的 (individual), 即特定的 REP 片段的任何两个 A 区段、任何两个 G 区段或任何两个 L 区段可以是相同的或可以不是相同的。因此, 各类型的区段在特定 REP 片段内是相同的这一点不是本发明的一般特征。相反地, 下列公开内容为本领域技术人员提供了关于如何设计独个区段并将它们集合至 REP 片段中的指导方针, 所述 REP 片段是本发明的功能性蛛丝蛋白的部分。

[0065] 从此处提供的实验数据中得出结论, 每个独个 A 区段是具有 8 至 18 个氨基酸残基的氨基酸序列。优选地, 每个独个 A 区段包含 13 至 15 个氨基酸残基。还可能地, 大部分或超过两个的 A 区段包含 13 至 15 个氨基酸残基, 以及少数例如一个或两个 A 区段包含 8 至 18 个氨基酸残基, 例如 8-12 个或 16-18 个氨基酸残基。这些氨基酸残基的绝大部分是丙氨酸残基。更特别地, 0 至 3 个氨基酸残基不是丙氨酸残基, 并且其余的氨基酸残基是丙氨酸残基。因此, 无例外地或除了 1、2 或 3 个氨基酸残基可以是任意氨基酸外, 每个独个 A 区段中的所有氨基酸残基是丙氨酸残基。优选的是, 取代丙氨酸的氨基酸是天然氨基酸, 优选独个地选自丝氨酸、谷氨酸、半胱氨酸和甘氨酸, 更优选丝氨酸。当然, 可能的是, 一个或多个 A 区段是全丙氨酸区段, 而其余的 A 区段包含 1 至 3 个非丙氨酸残基例如丝氨酸、谷氨酸、半

胱氨酸或甘氨酸。

[0066] 在优选的实施方案中,每个A区段包含13-15个氨基酸残基,其中包括10-15个丙氨酸残基和0-3个上述的非丙氨酸残基。在更优选的实施方案中,每个A区段包含13-15个氨基酸残基,其中包括12-15个丙氨酸残基和0-1个上述的非丙氨酸残基。

[0067] 优选地,每个独个A区段与选自SEQ ID NO:3的氨基酸残基7-19、43-56、71-83、107-120、135-147、171-183、198-211、235-248、266-279、294-306、330-342、357-370、394-406、421-434、458-470、489-502、517-529、553-566、581-594、618-630、648-661、676-688、712-725、740-752、776-789、804-816、840-853、868-880、904-917、932-945、969-981、999-1013、1028-1042和1060-1073的氨基酸序列具有至少80%的同一性。该组的各序列相应于Euprosthopsa australis MaSp1蛋白的天然发生的序列的区段,所述序列是从相应的cDNA的克隆过程中推导出来的(参见实施例1-2和图1-2A)。备选地,每个独个A区段与选自SEQ ID NO:9的氨基酸残基31-42、61-75、90-104、122-135和153-171;SEQ ID NO:13的氨基酸残基12-25、46-60、75-88、112-119、150-158和173-180;SEQ ID NO:14的氨基酸残基31-42;和SEQ ID NO:15的氨基酸残基122-135的氨基酸序列具有至少80%的同一性。该组的各序列相应于经表达的本发明的非天然蛛丝蛋白的区段,所述蛋白具有在合适的条件下形成丝纤维的能力。参见实施例5-8、12,以及图2B。不希望受任何特定理论的束缚,设想了,本发明的A区段形成螺旋结构或 β -折叠。

[0068] 如下计算了在整个说明书和所附权利要求书中所使用的术语“同一性百分比”。使用CLUSTAL W算法(Thompson, J.D., Higgins, D.G. 和 Gibson, T.J., Nucleic Acids Research, 22:4673-4680(1994))将查询序列与靶序列进行比对。比较每个位置上的氨基酸残基,并将查询序列中具有在靶序列中的相同的相应残基的位置的百分比报道为同一性百分比。

[0069] 如对于“同一性百分比”所描述的,计算在整个说明书和所附权利要求书中所使用的术语“相似性百分比”,除了疏水残基Ala、Val、Phe、Pro、Leu、Ile、Trp、Met和Cys是相似的;碱性残基Lys、Arg和His是相似的;酸性残基Glu和Asp是相似的;以及亲水性的、不带电荷的残基Gln、Asn、Ser、Thr和Tyr是相似的外。在该情形下,剩下的天然氨基酸Gly与任何其他氨基酸都不相似。

[0070] 在整个本说明书中,本发明的备选的实施方案并不满足指定的同一性百分比,而是满足相应的相似性百分比。其他备选的实施方案满足指定的同一性百分比,以及另一个更高的相似性百分比,其选自对于各序列的优选同一性百分比。例如,序列可以与另一个序列70%相似;或者其可以与另一个序列70%同一;或者其可以与另一个序列70%同一和90%相似。

[0071] 在本发明的优选实施方案中,每个独个A区段与选自下列的氨基酸序列具有至少90%,更优选95%,最优选100%的同一性:SEQ ID NO:3的氨基酸残基7-19、43-56、71-83、107-120、135-147、171-183、198-211、235-248、266-279、294-306、330-342、357-370、394-406、421-434、458-470、489-502、517-529、553-566、581-594、618-630、648-661、676-688、712-725、740-752、776-789、804-816、840-853、868-880、904-917、932-945、969-981、999-1013、1028-1042和1060-1073;SEQ ID NO:9的氨基酸残基31-42、61-75、90-104、122-135和153-171;SEQ ID NO:13的氨基酸残基12-25、46-60、75-88、112-119、

150-158 和 173-180 ;SEQ ID NO :14 的氨基酸残基 31-42 ;和 SEQ ID NO :15 的氨基酸残基 122-135。因此,在本发明的某些实施方案中,每个独个 A 区段与选自上述氨基酸区段的氨基酸序列相同。

[0072] 此外,从此处提供的实验数据中得出结论,每个独个 G 区段是 12 至 30 个氨基酸残基的氨基酸序列。优选地,每个独个 G 区段由 14 至 23 个氨基酸残基组成。每个 G 区段的至少 40% 的氨基酸残基是甘氨酸残基。通常,每个独个 G 区段的甘氨酸含量在 40% 至 60% 的范围内。

[0073] 优选地,每个独个 G 区段与选自 SEQ ID NO :3 的氨基酸残基 20-42、57-70、84-106、121-134、148-170、184-197、212-234、249-265、280-293、307-329、343-356、371-393、407-420、435-457、471-488、503-516、530-552、567-580、595-617、631-647、662-675、689-711、726-739、753-775、790-803、817-839、854-867、881-903、918-931、946-968、982-998、1014-1027、1043-1059 和 1074-1092 的氨基酸序列具有至少 80% 的同一性。该组的各序列相应于 *Euprosthopsa australis* MaSp1 蛋白的天然发生的序列的区段,所述序列是从相应的 cDNA 的克隆过程中推导出来的(参见实施例 1-2 和图 1-2A)。备选地,每个独个 G 区段与选自 SEQ ID NO :9 的氨基酸残基 11-30、43-60、76-89、105-121 和 136-152 ;和 SEQ ID NO :13 的氨基酸残基 1-11、26-45、61-74、89-111、120-149 和 159-172 的氨基酸序列具有至少 80% 的同一性。该组的各序列相应于经表达的本发明的非天然蛛丝蛋白的区段,所述蛋白具有在合适的条件下形成丝纤维的能力。参见实施例 5-8、12,以及图 2B。

[0074] 在本发明的优选实施方案,每个独个 G 区段与选自下列的氨基酸序列具有至少 90%,更优选 95%,最优选 100% 的同一性:SEQ ID NO :3 的氨基酸残基 20-42、57-70、84-106、121-134、148-170、184-197、212-234、249-265、280-293、307-329、343-356、371-393、407-420、435-457、471-488、503-516、530-552、567-580、595-617、631-647、662-675、689-711、726-739、753-775、790-803、817-839、854-867、881-903、918-931、946-968、982-998、1014-1027、1043-1059 和 1074-1092 ;SEQ ID NO :9 的氨基酸残基 11-30、43-60、76-89、105-121 和 136-152 ;和 SEQ ID NO :13 的氨基酸残基 1-11、26-45、61-74、89-111、120-149 和 159-172。因此,在本发明的某些实施方案中,每个独个 G 区段与选自上述氨基酸区段的氨基酸序列相同。

[0075] 在某些实施方案中,本发明的每个 G 区段的前两个氨基酸残基不是 -Gln-Gln-。

[0076] 在某些实施方案中,相应于保守的 Tyr 残基(即相应于 SEQ ID NO :5 中的位置 16, SEQ ID NO :6 中的位置 10,和 SEQ ID NO :7 中的位置 7)的位置在本发明的任何 G 区段中不是 Phe。

[0077] 在某些实施方案中,相应于保守的 Tyr 残基(即相应于 SEQ ID NO :5 中的位置 16, SEQ ID NO :6 中的位置 10,和 SEQ ID NO :7 中的位置 7)的位置在本发明的每个 G 区段中是 Tyr。

[0078] 由此断定,本发明蛋白质的某些实施方案展示了上述限制条件的组合,即本发明的每个 G 区段的前两个氨基酸残基不是 -Gln-Gln-,并且保守的 Tyr 残基(即相应于 SEQ ID NO :5 的位置 16, SEQ ID NO :6 中的位置 10,和 SEQ ID NO :7 中的位置 7)在本发明的每个 G 区段中是 Tyr。在某些实施方案中,上述限制条件(分开地或以任何可能的组合形式采用)可进一步与 REP 片段缺乏脯氨酸残基这一上文所述的限制条件相组合。

[0079] 参考图 1-2 以及实施例 3-4, 存在 3 个亚型的本发明的 G 区段。该分类基于 *Euprostenops australis* MaSp1 蛋白质序列 (图 1-2A) 的仔细分析, 并且已在构建新的非天然蜘蛛丝蛋白中采用和验证了所述信息 (图 2B)。

[0080] 本发明的 G 区段的第一种亚型由氨基酸单字母共有序列 GQG(G/S)QGG(Q/Y)GG(L/Q)GQGGYGQGA GSS (如图 2A 和 SEQ ID NO:5 中所示的) 表示。该第一种并且通常最长的 G 区段亚型通常包含 23 个氨基酸残基, 但可包含少至 17 个氨基酸残基, 并且不存在带电荷的残基或包含一个带电荷的残基。因此, 优选的是, 该第一种 G 区段亚型包含 17-23 个氨基酸残基, 但可预见其可包含少至 12 个或多至 30 个氨基酸残基。不希望受任何特定理论的束缚, 设想了, 该亚型形成卷曲结构 (coil structure) 或 3_1 -螺旋结构。该第一种亚型的代表性 G 区段是 SEQ ID NO:3 的氨基酸残基 20-42、84-106、148-170、212-234、307-329、371-393、435-457、530-552、595-617、689-711、753-775、817-839、881-903、946-968、1043-1059 和 1074-1092; SEQ ID NO:9 的氨基酸残基 11-30、105-121 和 136-152; 和 SEQ ID NO:13 的氨基酸残基 26-45 和 89-111。该第一种亚型的备选的 G 区段是 SEQ ID NO:13 的氨基酸残基 120-149 和 159-172。在某些实施方案中, 本发明的该第一种亚型的每个 G 区段的前两个氨基酸残基不是 -Gln-Gln-。

[0081] 本发明的 G 区段的第二种亚型由氨基酸单字母共有序列 GQGGQGGQ(G/R)Y GQG(A/S)G(S/G)S (如图 2A 和 SEQ ID NO:6 中所示的) 表示。该第二种并且通常中等大小的 G 区段亚型通常包含 17 个氨基酸残基, 并且不存在带电荷的残基或包含一个带电荷的残基。优选的是, 该第二种 G 区段亚型包含 14-20 个氨基酸残基, 但可预见其可包含少至 12 个或多至 30 个氨基酸残基。不希望受任何特定理论的束缚, 设想了, 该亚型形成卷曲结构。该第二种亚型的代表性 G 区段是 SEQ ID NO:3 的氨基酸残基 249-265、471-488、631-647 和 982-998; 和 SEQ ID NO:9 的氨基酸残基 43-60。

[0082] 本发明的 G 区段的第三种亚型由氨基酸单字母共有序列 G(R/Q)GQG(G/R)YGQG(A/S/V)GGN (如图 2A 和 SEQ ID NO:7 中所示的) 表示。该第三种 G 区段亚型通常包含 14 个氨基酸残基, 并且通常是最短的本发明的 G 区段亚型。优选的是, 该第三种 G 区段亚型包含 12-17 个氨基酸残基, 但可预见其可包含多至 23 个氨基酸残基。不希望受任何特定理论的束缚, 设想了, 该亚型形成转角结构。该第三种亚型的代表性 G 区段是 SEQ ID NO:3 的氨基酸残基 57-70、121-134、184-197、280-293、343-356、407-420、503-516、567-580、662-675、726-739、790-803、854-867、918-931、1014-1027; SEQ ID NO:9 的氨基酸残基 76-89; 和 SEQ ID NO:13 的氨基酸残基 61-74。该第三种亚型的备选的 G 区段是 SEQ ID NO:13 的氨基酸残基 1-11。

[0083] 因此, 在优选的实施方案中, 每个独个 G 区段与选自 SEQ ID NO:5、SEQ ID NO:6 和 SEQ ID NO:7 的氨基酸序列具有至少 80%、优选 90%、更优选 95% 的同一性。

[0084] 在 REP 片段的 A 和 G 区段的交替排列性序列的优选实施方案中, 每第二个 G 区段是第一种亚型, 而其余的 G 区段是第三种亚型, 例如 ... A₁G_短A₂G_长A₃G_短A₄G_长A₅G_短...。在 REP 片段的另一个优选实施方案中, 一个第二种亚型的 G 区段通过插入而打断了 G 区段的规律性 (regularity), 例如 ... A₁G_短A₂G_长A₃G_{中等}A₄G_短A₅G_长...。

[0085] 每个独个 L 区段代表可有可无的连接体氨基酸序列, 其可包含 0 至 20 个氨基酸残基, 例如 0 至 10 个氨基酸残基。虽然该区段是可有可无的, 并且对于蛛丝蛋白在功能上不

是至关重要的,但是其的存在仍然允许获得完全功能性的蛛丝蛋白,从而形成本发明的蜘蛛丝纤维。在来自 *Euprosthénops australis* 的 MaSp1 蛋白的经推导的氨基酸序列的重复部分 (SEQ ID NO :3) 中也存在连接体氨基酸序列。特别地,连接体区段的氨基酸序列可与所描述的 A 或 G 区段中任一个相似,但通常不足以满足它们的在此处定义的标准。

[0086] 如图 2A 中所示的,排列在 REP 片段的 C- 末端部分的连接体区段可由富含丙氨酸的氨基酸单字母共有序列 ASASAAASAA STVANSVS 和 ASAASAAA 代表。事实上,可以认为第二种序列是本发明的 A 区段,而第一种序列与本发明的 A 区段具有高度的相似性。本发明的连接体区段的另一个实例具有单字母氨基酸序列 GSAMGQGS,该序列富含甘氨酸并且与本发明的 G 区段具有高度的相似性。

[0087] 代表性的 L 区段是 SEQ ID NO :3 的氨基酸残基 1-6 和 1093-1110 ;SEQ ID NO :9 的氨基酸残基 1-10 和 153-171 ;和 SEQ ID NO :13 的氨基酸残基 173-180,但本领域技术人员将容易地认识到,对于这些区段存在许多合适的备选的氨基酸序列。在本发明的 REP 片段的一个实施方案中,所述 L 区段之一包含 0 个氨基酸,即所述 L 区段之一是没有的。在本发明的 REP 片段的另一个实施方案中,两个 L 区段都包含 0 个氨基酸,即两个 L 区段都是没有的。因此,本发明的 REP 片段的这些实施方案可如下示意性地表示 : $(AG)_nL$ 、 $(AG)_nAL$ 、 $(GA)_nL$ 、 $(GA)_nGL$; $L(AG)_n$ 、 $L(AG)_nA$ 、 $L(GA)_n$ 、 $L(GA)_nG$;和 $(AG)_n$ 、 $(AG)_nA$ 、 $(GA)_n$ 、 $(GA)_nG$ 。这些 REP 片段中的任一片段适合于与下面定义的任何 CT 片段一起使用。

[0088] 本发明的蛛丝蛋白的 C- 末端 (CT) 片段与蛛丝蛋白的 C- 末端氨基酸序列具有高度的相似性。如图 3 中所示的,该氨基酸序列在各种不同的物种以及蛛丝蛋白 (包括 MaSp1 和 MaSp2) 中是非常保守的。在下列实施例中证明了,哪一个特定的 CT 片段存在于本发明的蛛丝蛋白中不是至关重要的,只要 CT 片段不完全丢失。因此,本发明的 CT 片段可选自图 3 中显示的氨基酸序列中的任一序列或具有高度相似性的序列。值得注意的是,SEQ ID NO :13 的 CT_{hyb} 片段与共有氨基酸序列 SEQ ID NO :8 具有 96% 的同一性,而 SEQ ID NO :9 的 CT_{nat} 片段只展示出与共有氨基酸序列 SEQ ID NO :8 具有 59% 的同一性。这说明,各种各样的 C- 末端序列可以用于本发明的蛛丝蛋白。

[0089] 本发明的 CT 片段的序列与共有氨基酸序列 SEQ ID NO :8 (其基于图 3 的氨基酸序列) 具有至少 50% 的同一性,优选至少 60% 的同一性。在优选的实施方案中,本发明的 CT 片段的序列与共有氨基酸序列 SEQ ID NO :8 的氨基酸残基 1-71 具有至少 65% 的同一性,优选至少 70% 的同一性。在优选的实施方案中,本发明的 CT 片段进一步地与共有氨基酸序列 SEQ ID NO :8 或其氨基酸残基 1-71 具有 70%,优选 80% 的相似性。

[0090] 本发明的代表性 CT 片段是 *Euprosthénops australis* 序列 SEQ ID NO :4、*Euprosthénops australis* 来源的 SEQ ID NO :9 的氨基酸残基 172-269 和 SEQ ID NO :13 的氨基酸残基 181-276,其据称来源于 *Euprosthénops* sp (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. Molecular studies of a novel dragline silk from a nurseryweb spider, *Euprosthénops* sp. (Pisauridae). Comp Biochem Physiol B Biochem Mol Biol 138, 371-376 (2004)),但与来自棒络新妇蛛和塞内加尔络新妇蛛 (*Nephila senegalensis*) 的 MaSp1 具有高度的相似性。因此,根据本发明的优选方面,CT 片段与 SEQ ID NO :4 ;SEQ ID NO :9 的氨基酸残基 172-269 ;SEQ ID NO :13 的氨基酸残基 181-276 ;SEQ ID NO :16 的氨基酸残基 172-269 ;或图 3 和实施例 4 的任何独个 MaSp1/ADF-4 氨基酸序列具有至少 80% 的同

一性。在本发明的优选方面,CT 片段与 SEQ ID NO:4;SEQ ID NO:9 的氨基酸残基 172-269;SEQ ID NO:13 的氨基酸残基 181-276;SEQ ID NO:16 的氨基酸残基 172-269;或图 3 和实施例 4 中的任何独个 MaSp1/ADF-4 氨基酸序列具有至少 90%,例如至少 95%的同一性。在本发明的优选方面,CT 片段与 SEQ ID NO:4;SEQ ID NO:9 的氨基酸残基 172-269;SEQ ID NO:13 的氨基酸残基 181-276;SEQ ID NO:16 的氨基酸残基 172-269;或图 13 和实施例 4 的任何独个 MaSp1/ADF-4 氨基酸序列相同。

[0091] CT 片段通常由 70 至 120 个氨基酸残基组成。优选的是,CT 片段包含至少 70 个,或大于 80 个,优选大于 90 个氨基酸残基。也优选的是,CT 片段包含至多 120 个,或小于 110 个氨基酸残基。典型的 CT 片段包含大约 100 个氨基酸残基。

[0092] 根据另一个方面,本发明提供了分离的融合蛋白,其由作为大壶腹腺蛛丝蛋白的第一蛋白质片段(优选地由 150 至 420 个氨基酸残基组成)和包含融合伙伴及切割试剂识别位点的第二蛋白质片段组成。第一蛋白质片段通过切割试剂识别位点偶联至融合伙伴,即可通过用合适的切割试剂在合适的条件下处理融合蛋白来切掉融合伙伴,从而提供大壶腹腺蛛丝蛋白,其优选地由 150 至 420 个氨基酸残基组成。该融合蛋白的优点是,当在合适的宿主,例如细菌,优选大肠杆菌中产生时,可在溶液中,优选在生理学介质中,通常在缓冲水性介质例如 10-100mM Tris-HCl 缓冲液,pH6-9 中产生大量的融合蛋白,而不引起沉淀和其他生产问题。溶液中的融合蛋白在长时间内,通常在数天或数周内是溶解的,这有利于大规模生产且降低了蛋白质聚集的风险。术语“溶解的”和“在溶液中”是指,所述蛋白质不以可见的方式聚集,并且在 60000xg 下不会从溶剂中沉淀出来。根据所希望的,溶液中的融合蛋白可经历使用合适的切割试剂进行的切割,从而提供自发形成丝纤维的大壶腹腺蛛丝蛋白。

[0093] 在优选的方面,本发明提供了选自 X-REP-CT 和 REP-CT-X 的分离的融合蛋白,优选 X-REP-CT。REP 和 CT 是本发明的蛋白质片段,这暗示着所得到的 REP-CT 这种形式的 MaSp 蛋白是本发明的 MaSp 蛋白。X 是此处定义的包含融合伙伴和切割试剂识别位点的蛋白质片段。该组合的蛋白质片段 REP-CT 通过切割试剂识别位点与融合伙伴偶联。

[0094] 本发明的融合伙伴包括提高其伙伴蛋白质片段(此处为本发明的 MaSp 蛋白)的溶解性和/或稳定性的任何蛋白质片段。融合伙伴也提供了用于亲和纯化的合适的手段。不局限于此,本发明的融合伙伴的实例包括硫氧还蛋白、麦芽糖结合蛋白、谷胱甘肽 S-转移酶(GST)、MTB32-C、Gb1、ZZ 和 Nus A。本领域技术人员熟知备选的合适的融合伙伴。在本发明的优选实施方案中,融合伙伴是与 His 标签以及 S 标签相组合的硫氧还蛋白部分(ThrX)。在本发明的一个优选实施方案中,融合伙伴是与 2 个 His 标签相组合的 ThrX 部分,即 His 标签/ThrX/His 标签。在本发明的另一个优选实施方案中,融合伙伴是硫氧还蛋白部分(ThrX)。

[0095] 切割试剂识别位点位于偶联至 MaSp 蛋白片段的该 X 蛋白片段末端,从而在该识别位点处的切割导致产生 MaSp 蛋白和融合伙伴。不局限于此,本发明的切割试剂识别位点的实例包括具有氨基酸序列 LVPRGS(在 R 和 G 之间切割)的凝血酶识别位点;具有氨基酸序列 DDDK(在 K 后切割)的胰激酶识别位点;具有氨基酸序列 NG(在 N 和 G 之间切割)的羟胺识别位点;具有氨基酸序列 LGVLFQGP(在 Q 和 G 之间切割)的 HRV3C 蛋白酶识别位点;具有氨基酸序列 I(E/D)GR(在 R 后切割)的凝血因子 Xa 识别位点;具有氨基酸序列 EXXYXQ(G/

S),通常 ENLYFQG(在 Q 和 G/S 之间切割)的 TEV 识别位点;具有氨基酸序列 EDNLYFQG(在 Q 和 G 之间切割)的 Ac-TEV 识别位点;和具有氨基酸序列 LEVLFQGP(在 Q 和 G 之间切割)的 PreScission 识别位点。其他合适的识别位点是关于胰蛋白酶、内切蛋白酶、V8 蛋白酶、胃蛋白酶和 CNBr 的切割位点。合适的切割识别位点的其他实例完全在本领域技术人员能力所达的范围内。在本发明的优选实施方案中,切割试剂识别位点是凝血酶识别位点。

[0096] 在优选的实施方案中,本发明的 X 片段具有结构 ThrX/His- 标签 /S- 标签 / 凝血酶切割位点,并且 X 片段偶联至本发明的 REP-CT 蛋白质片段的 N- 末端。

[0097] 在一个优选的实施方案中,本发明的 X 片段具有结构 His- 标签 /ThrX/His- 标签 / 凝血酶切割位点,并且 X 片段偶联至本发明的 REP-CT 蛋白质片段的 N- 末端。

[0098] 根据另一个方面,本发明提供了用于产生本发明的大壶腹腺蛛丝蛋白的方法。在第一个步骤中,提供本发明的融合蛋白在液体介质中的溶液。优选地,融合蛋白不聚集,从而不需要再溶解过程。可重组产生融合蛋白,并使用在融合蛋白上的亲和手段例如 His- 标签或在融合蛋白中的任何合适表位来纯化所述融合蛋白。液体介质可以是任何合适的介质,优选生理学介质,通常为缓冲水性介质,例如 10-100mM Tris-HCl 缓冲液, pH6-9。在第二个步骤中,为了在切割试剂识别位点处获得融合蛋白的切割,向液体介质中加入本发明的切割试剂。如上面所公开的,由此获得本发明的大壶腹腺蛛丝蛋白。在第三个任意的步骤中,使用合适的分离技术例如层析和 / 或过滤从液体介质中分离如此获得的大壶腹腺蛛丝蛋白。

[0099] 根据另一个方面,本发明提供了用于产生本发明的大壶腹腺蛛丝蛋白的多聚体的方法。在第一个步骤中,提供本发明的融合蛋白在液体介质中的溶液。优选地,融合蛋白不聚集,从而不需要再溶解过程。可重组产生融合蛋白,并使用在融合蛋白上的亲和手段例如 His- 标签或在融合蛋白中的任何合适表位来纯化所述融合蛋白。液体介质可以是任何合适的介质,优选生理学介质,通常为缓冲水性介质,例如 10-100mM Tris-HCl 缓冲液, pH6-9。在第二个步骤中,为了在切割试剂识别位点处获得融合蛋白的切割,向液体介质中加入本发明的切割试剂。如上面所公开的,由此获得本发明的大壶腹腺蛛丝蛋白。在第三个步骤中,使如此获得的大壶腹腺蛛丝蛋白在液体介质中聚合。聚合通常在两种不同相之间的界面例如液 / 空气、液 / 固和水 / 油界面处开始。因此,该第三个步骤还可进一步包括提供液体介质和另一相之间的界面。所述另一相选自气相、液相和固相。如上所述,液体介质通常是水性介质,并且合适的其他相例如为空气和与水不能混溶的有机溶剂,例如油,例如适用于 PCR 反应的矿物油。所得到的界面的存在刺激在该界面处或在围绕该界面的区域中的聚合,所述区域延伸入液体介质中,从而所述聚合在所述界面处或在所述界面区域中开始。优选的界面包括水 / 空气和水 / 油界面。当在室温下温育时,聚合通常在数分钟或数小时内,例如在 1 分钟至 5 小时内自发发生。在第四个任意的步骤中,使用合适的分离技术从液体介质中分离如此获得的大壶腹腺蛛丝蛋白的多聚体。

[0100] 如上所述,通过从融合蛋白中蛋白水解释放小型的蛛丝蛋白来引起纤维形成。如果在从一侧至另一侧轻轻地摇动的试管中进行切割反应,则纤维沿着试管在空气 - 水界面处形成。试管可由任何合适的材料例如塑料或玻璃制成。如果允许切割混合物静置,则在空气 - 水界面处形成薄膜。如果在水性切割混合物之上加入油,则在油 - 水界面处形成薄膜,其中让其静置或摇动。如果例如通过空气起泡或搅打来使切割混合物发泡,则泡沫是稳

定的,并且如果允许干燥则发生凝固。

[0101] 通过使用本发明的方法,可能重组产生大量的本发明的融合蛋白,所述融合蛋白可被切割,并且被允许按所希望的进行聚合。这提供了聚合过程的更好的控制,并允许优化参数以获得具有所需性质的丝纤维。

[0102] 通常使用各种合适的宿主来重组产生本发明的大壶腹腺蛛丝蛋白。根据另一个方面,本发明因此提供了分离的多聚核酸分子,其包含编码本发明的大壶腹腺蛛丝蛋白的核酸序列或其互补核酸序列,例如 SEQ ID NO :9-13,优选 SEQ ID NO :9、12 和 13。这些多聚核酸分子以及编码此处公开的各种不同蛋白质的多聚核酸分子 (SEQ ID NO :2-16) 还可用于进一步开发非天然蛛丝蛋白或其生产系统。

[0103] 通常使用各种合适的宿主例如细菌、酵母、哺乳动物细胞、植物、昆虫细胞和转基因动物来重组产生本发明的融合蛋白。优选的是,在细菌中产生本发明的融合蛋白。

[0104] 根据另一个方面,本发明因此提供了分离的多聚核酸分子,其包含编码本发明的融合蛋白的核酸序列或其互补核酸序列。所述多聚核酸分子还可用于进一步开发非天然蛛丝蛋白或其生产系统。

[0105] 本发明的多聚核酸分子可以是 DNA 分子,包括 cDNA 分子或 RNA 分子。正如本领域技术人员所充分意识到的,核酸序列同样可通过其互补核酸序列来描述。因此,与本发明的核酸序列互补的核酸序列也包括在本发明的保护范围内。

[0106] 根据一个方面,本发明提供了用于产生本发明的可溶性融合蛋白的方法。在第一个步骤中,在合适的宿主中表达编码本发明的融合蛋白的多聚核酸分子。在第二个步骤中,例如使用层析和 / 或过滤来分离在前述步骤中如此获得的可溶性融合蛋白。任选地,分离可溶性融合蛋白的所述第二个步骤包括除去 LPS 和其他致热原。

[0107] 本发明还提供了用于产生本发明的大壶腹腺蛛丝蛋白的方法。在第一个步骤中,在合适的宿主中表达编码本发明的融合蛋白的多聚核酸分子。在第二个步骤中,例如使用层析和 / 或过滤来分离如此获得的可溶性融合蛋白。在第三个步骤中,提供所述分离的融合蛋白的溶液,并在第四个步骤中,向液体介质中加入合适的切割试剂。这实现了在切割试剂识别位点处切割融合蛋白,并从而提供了大壶腹腺蛛丝蛋白。在任选的第五个步骤中,从液体介质中分离如此获得的大壶腹腺蛛丝蛋白。进一步任选地,分离可溶性融合蛋白的所述第二个步骤,以及任选地,分离大壶腹腺蛛丝蛋白的所述第五个步骤,包括除去 LPS 和其他致热原。

[0108] 本发明还提供了用于产生本发明的大壶腹腺蛛丝蛋白的多聚体的方法。在第一个步骤中,在合适的宿主中表达编码本发明的融合蛋白的多聚核酸分子。在第二个步骤中,例如使用层析和 / 或过滤来分离如此获得的可溶性融合蛋白。在第三个步骤中,提供所述分离的融合蛋白的溶液,并在第四个步骤中,向液体介质中加入合适的切割试剂。这实现了在切割试剂识别位点处切割融合蛋白,并从而提供了大壶腹腺蛛丝蛋白。在第五个步骤中,使如此获得的大壶腹腺蛛丝蛋白在液体介质中聚合。聚合通常在两种不同相之间的界面例如液 / 空气、液 / 固和水 / 油界面处开始。因此,该第五个步骤还可进一步包括提供液体介质和另一相之间的界面。所述另一相选自气相、液相和固相。如上所述,液体介质通常是水性介质,并且合适的其他相例如为空气和与水不能混溶的有机溶剂,例如油,例如适合用于 PCR 反应的矿物油。所得到的界面的存在刺激在该界面处或在围绕该界面的区域中的聚合,

所述区域延伸入液体介质中,从而所述聚合在所述界面处或在所述界面区域中开始。优选的界面包括水/空气和水/油界面。当在室温下温育时,聚合通常在数分钟或数小时内,例如在 1 分钟至 5 小时内自发发生。在任选的第六个步骤中,从液体介质中分离如此获得的多聚体。

[0109] 为了获得具有低致热原含量的蛋白质(这对于用作体内生物材料来说是强制性的),已开发了对于除去脂多糖(LPS)来说经优化的纯化方案。为避免被所释放的 LPS 污染,将生产性细菌细胞经历使用改变的 CaCl_2 和 EDTA 的洗涤步骤。在细胞裂解后,在低导电性缓冲液中进行所有随后的纯化步骤,以便使靶蛋白和 LPS 之间的疏水相互作用减少至最低。通过将蛋白质溶液通过 Endotrap 柱来使 LPS 的含量进一步降至最低,所述 Endotrap 柱具有特异性吸附 LPS 的配体。为了确保恒定的低含量的 LPS 和其他致热原,使用体外致热原测试(IPT)和/或鲎变形细胞溶解物(LAL)动力学测定法来分析所有批次。尽管在革兰氏阴性细菌宿主中产生,但可如此纯化出重组蛛丝蛋白,即 LPS 和其他致热原的残留水平低于动物测试所要求的限度,即低于 25EU/植入物。在本发明的某些实施方案中,在分离的融合蛋白中 LPS 和其他致热原的含量是 1EU/mg 蛋白质或更低。在本发明的某些实施方案中,在分离的大壶腹腺蛛丝蛋白中 LPS 和其他致热原的含量是 1EU/mg 蛋白质或更低,优选 0.25EU/mg 蛋白质或更低。

[0110] 根据另一个方面,本发明提供了本发明的大壶腹腺蛛丝蛋白的多聚体。在优选的实施方案中,该蛋白质的多聚体可通过本发明的用于此的任一种方法获得。

[0111] 在优选的实施方案中,本发明的大壶腹腺蛛丝蛋白的多聚体的 β -折叠含量高于 50%,即该蛋白的多聚体的超过 50%的二级结构是以 β -折叠形式存在的。这是有利的,因为可以预见,更高含量的 β -折叠结构提供了更硬和/或更强和/或可延伸性更低的纤维。

[0112] 优选的是,本发明的蛛丝蛋白的多聚体是具有肉眼可见的大小,即具有大于 $1\mu\text{m}$,优选大于 $10\mu\text{m}$ 的直径和大于 5mm 的长度的纤维。优选的是,所述纤维的直径在 $10\text{--}400\mu\text{m}$,优选 $60\text{--}120\mu\text{m}$ 的范围内,以及长度在 0.5–300cm,优选 1–100cm 的范围内。其他优选的范围是 0.5–30cm 和 1–20cm。也优选的是,本发明的蛛丝蛋白的多聚体具有大于 1MPa,优选大于 2MPa,更优选 10MPa 或更高的拉伸强度。优选的是,本发明的蛛丝蛋白的多聚体具有大于 100MPa,更优选 200MPa 或更高的拉伸强度。所述纤维具有在物理操作过程中保持完整的能力,即可用于纺丝、织造、加捻、钩织和相似的操作。

[0113] 在其他优选的实施方案中,本发明的蛛丝蛋白的多聚体形成泡沫、凝胶、网状物或薄膜。

[0114] 根据另一个方面,本发明提供了包含融合伙伴和切割试剂识别位点的蛋白质片段用于制备融合蛋白的新用途。所述融合蛋白包含所述蛋白质片段和本发明的蜘蛛丝蛋白片段,并且这两个片段通过所述切割试剂识别位点偶联。所述蜘蛛丝蛋白片段优选由 150 至 420 个氨基酸残基组成。

[0115] 将在下面通过下列非限定性实施例来进一步举例说明本发明。

[0116] 实施例

[0117] 实施例 1. Euprosthénops australis MaSp1 cDNA 的克隆和测序

[0118] 将来自在南非收集的大约 100 个成年雌性 Euprosthénopsaustralis 蜘蛛的大壶腹腺用于构建定制的基于 pDONR222 的 CloneMiner cDNA 文库 (Invitrogen, Paisley, UK)。

通过用 cDNA 探针筛选该文库来获得编码 MaSp1 蛋白的 cDNA 克隆,所述 cDNA 探针编码来源于未知亚种的 *Euprosthénops* 蜘蛛的富含丙氨酸和富含甘氨酸的片段。按照厂商说明书,使用 ECL 直接标记和检测系统 (Amersham Biosciences, Uppsala, Sweden) 来进行菌落印迹和检测。

[0119] 选择一个单个的克隆以进行进一步的表征。为了从该克隆中获得 cDNA 插入片段的全长序列,使用 Erase-a-Base 系统 (Promega, Southampton, UK) 来产生嵌套缺失,并在 MegaBase 1000 设备 (Amersham Biosciences) 上进行测序。

[0120] 所得到的 3.8kb cDNA (SEQ ID NO:1) 编码 1207 个氨基酸残基的 MaSp1 蛋白 (SEQ ID NO:2),该蛋白包含 34 个富含丙氨酸和富含甘氨酸的区段的重复片段 (SEQ ID NO:3),以及 97 个氨基酸残基的 C-末端非重复片段 (SEQ ID NO:4)。

[0121] 实施例 2. *Euprosthénops australis* MaSp1 蛋白的重复片段的序列分析

[0122] 通过该片段的重复区段的比对来进一步分析实施例 1 的 *Euprosthénops australis* MaSp1 蛋白序列的重复片段 (SEQ ID NO:3),参见图 1。仔细查看该比对,并得出下列结构信息。

[0123] *Euprosthénops australis* MaSp1 蛋白的富含丙氨酸的区段的长度为 13-15 个氨基酸残基,并且只由丙氨酸残基组成或除了一个残基(该残基是丝氨酸、谷氨酸或甘氨酸残基)外全部由丙氨酸残基组成。

[0124] *Euprosthénops australis* MaSp1 蛋白的重复片段还包含 3 个相关的、但不同的类型的富含甘氨酸的区段,参见图 2A。富含甘氨酸的区段中的两个区段几乎只在长度和出现率上不同;最常见的富含甘氨酸的区段包含 23 个氨基酸残基,而丰度较低的变体包含 17 个氨基酸残基。这两种富含甘氨酸的区段通常都不包含带电荷的残基或包含一个带电荷的残基。相反地,包含 14 个氨基酸残基的最短的富含甘氨酸的片段独特地在 N-末端包含序列 GRGQG 或 GQGQG,和在 C-末端包含 GN。

[0125] 最长的富含甘氨酸的区段由氨基酸单字母共有序列 GQG(G/S)QGG(Q/Y)GG(L/Q)GQGGYGQGA GSS (SEQ ID NO:5) 表示,并且不包含带电荷的残基。预测该区段形成卷曲结构或 3_1 -螺旋结构。中等大小的富含甘氨酸的区段由氨基酸单字母共有序列 GQGGQGGQ(G/R)YGQG(A/S)G(S/G)S (SEQ ID NO:6) 表示,并且不包含带电荷的残基或包含一个带电荷的残基。预测该区段形成卷曲结构。最短的富含甘氨酸的区段由氨基酸单字母共有序列 G(R/Q)GQG(G/R)YGQG(A/S/V)GGN (SEQ ID NO:7) 表示。预测该区段形成转角结构。

[0126] 从交替的富含丙氨酸和富含甘氨酸的区段(例如... $A_1G_1A_2G_2A_3G_3A_4G_4A_5G_5$...)建立 *Euprosthénops australis* MaSp1 蛋白的重复片段。观察到,上面鉴定的最短和最长的富含甘氨酸的区段中的每一个通常以每隔一富含甘氨酸的区段的形式出现,例如... $A_1G_{短}A_2G_{长}A_3G_{短}A_4G_{长}A_5G_{短}$...。相反地,较不丰富的、中等大小的、富含甘氨酸的片段通常在更长类型的富含甘氨酸的区段和更短类型的富含甘氨酸的区段之间出现,例如... $A_1G_{短}A_2G_{长}A_3G_{中}$ 等 $A_4G_{短}A_5G_{长}$...。

[0127] 实施例 3. *Euprosthénops australis* MaSp1 蛋白的重复片段的二级和三级结构的预测

[0128] 溶液中的蛛丝蛋白多肽通常通过形成发夹结构而折叠,所述发夹结构预示了成熟纤维的反平行 β -折叠结构。为了辨别实施例 1-2 的 *Euprosthénops australis* MaSp1 蛋

白的重复片段 (SEQ ID NO :3) 的可能的折叠模式, 鉴定了与发夹或转角结构的形成相容的蛋白质区域。富含丙氨酸的区段不可能是转角形成的候选者, 因为它们预期形成螺旋结构, 并且更重要地, 这些区段在纤维中通常被保持用于构成 β -折叠。

[0129] 通过使用最近描述的关于转角预测的算法 (Fuchs, PF & Alix, AJ, High accuracy prediction of beta-turns and their types using propensity and multiple alignments. *Proteins* 59, 828-839 (2005)), 最短的富含甘氨酸的区段显示出关于形成 II 型 β -转角的高度可能性, 而两个更长的富含甘氨酸的区段预测形成卷曲结构。在更长的富含甘氨酸的区段中高含量的 GGX 三联体暗示了, 它们可形成 3_1 -螺旋结构。

[0130] 蛛丝蛋白氨基酸序列的重复性质暗示了折叠模式的同样地重复的性质。综合起来, 这些观察结果导致图 2A 中所示的 *Euprosthénops australis* MaSp1 蛋白的重复片段的折叠。值得注意的是, 带正电荷的残基几乎总是位于所提及的转角结构中。

[0131] 根据 *Euprosthénops australis* MaSp1 蛋白的重复区段的折叠模式, 可辨别由富含丙氨酸的区段 / (更长的) 富含甘氨酸的卷曲区段 / 富含丙氨酸的区段 / (更短的) 富含甘氨酸的转角区段 / 富含丙氨酸的区段 / (更长的) 富含甘氨酸的卷曲区段 / 富含丙氨酸的区段组成的基元 (图 2A 中示意性地图解说明的)。

[0132] 实施例 4. MaSp1 蛋白的非重复 C-末端片段的序列分析

[0133] 将实施例 1 中获得的来自 *Euprosthénops australis* 的 MaSp1 蛋白的 C-末端非重复片段 (SEQ ID NO :4) 的一级结构与许多已知的 MaSp1 和 MaSp2 蛋白的 C-末端片段进行比对, 所述 MaSp1 和 MaSp2 蛋白尤其来自 *Euprosthénops* sp. (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthénops* sp. (Pisauridae). *Comp Biochem Physiol B Biochem Mol Biol* 138, 371-376 (2004))、棒络新妇蛛 P19837-5 (Xu, M & Lewis, RV, Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci USA* 87, 7120-7124 (1990)) 和其他种类。

[0134] 根据在图 3 中显示的从重复片段中的最后 Ser 开始的比对, 显而易见的是, MaSp1 和 MaSp2 的 C-末端区域是十分保守的。*Euprosthénops* sp 和棒络新妇蛛具有 95% 的相同残基; *Euprosthénops australis* 和棒络新妇蛛具有 54% 的相同残基; 以及 *Euprosthénops australis* 和 *Euprosthénops* sp 具有 55% 的相同残基。作为 SEQ ID NO :8 提供了 MaSp1 和 MaSp2 的 C-末端区域的共有序列。在图 3 中, 比对了下列 MaSp 蛋白, 它们用 GenBank 登录号 (使用这些登录号可获得这些蛋白) 进行了标注:

[0135] <u>物种和蛛丝蛋白</u>	<u>登录号</u>
[0136] <i>Euprosthénops</i> sp MaSp1 (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, 同前)	Cthyb_Esp
[0138] <i>Euprosthénops australis</i> MaSp1 (SEQ ID NO :4)	CTnat_Eau
[0139] 三带金蛛 (<i>Argiope trifasciata</i>) MaSp1	AF350266_At1
[0140] 皿云斑蛛 (<i>Cyrtophora moluccensis</i>) Sp1	AY666062_Cm1
[0141] 棕寡妇蛛 (<i>Latrodectus geometricus</i>) MaSp1	AF350273_Lg1
[0142] 黑寡妇蛛 (<i>Latrodectus hesperus</i>) MaSp1	AY953074_Lh1
[0143] 赫氏粒突蛛 (<i>Macrothele holsti</i>) Sp1	AY666068_Mh1

[0144]	棒络新妇蛛 MaSp1	U20329_Nc1
[0145]	毛络新妇蛛 (<i>Nephila pilipes</i>)MaSp1	AY666076_Np1
[0146]	马达加斯加络新妇蛛 (<i>Nephila madagascariensis</i>)MaSp1	AF350277_Nm1
[0147]	塞内加尔络新妇蛛 MaSp1	AF350279_Ns1
[0148]	变异涡蛛 (<i>Octonoba varians</i>)Sp1	AY666057_Ov1
[0149]	中华络网蛛 (<i>Psechrus sinensis</i>)Sp1	AY666064_Ps1
[0150]	考艾岛肖蛸 (<i>Tetragnatha kauaiensis</i>)MaSp1	AF350285_Tk1
[0151]	花斑肖蛸 (<i>Tetragnatha versicolor</i>)MaSp1	AF350286_Tv1
[0152]	大腹园蛛 (<i>Araneus bicentenarius</i>)Sp2	ABU20328_Ab2
[0153]	悦目金蛛 (<i>Argiope amoena</i>)MaSp2	AY365016_Aam2
[0154]	橙色金蛛 (<i>Argiope aurantia</i>)MaSp2	AF350263_Aau2
[0155]	三带金蛛 MaSp2	AF350267_At2
[0156]	乳头棘腹蛛 (<i>Gasteracantha mammosa</i>)MaSp2	AF350272_Gm2
[0157]	棕寡妇蛛 MaSp2	AF350275_Lg2
[0158]	黑寡妇蛛 MaSp2	AY953075_Lh2
[0159]	棒络新妇蛛 MaSp2	AY654293_Nc2
[0160]	马达加斯加络新妇蛛 MaSp2	AF350278_Nm2
[0161]	塞内加尔络新妇蛛 MaSp2	AF350280_Ns2
[0162]	黑狡蛛 (<i>Dolomedes tenebrosus</i>)Fb1	AF350269_DtFb1
[0163]	黑狡蛛 Fb2	AF350270_DtFb2
[0164]	十字园蛛 ADF-1	U47853_ADF1
[0165]	十字园蛛 ADF-2C	U47854_ADF2
[0166]	十字园蛛 ADF-3	U47855_ADF3
[0167]	十字园蛛 ADF-4	U47856_ADF4

[0168] 实施例 5. MaSp1 基因的构建

[0169] 使用来自实施例 1 的 cDNA 文库的 MaSp1 克隆作为模板,利用 Advantage GC2 试剂盒 (BD Biosciences, San Jose, CA, USA),通过 PCR 来扩增编码 *Euprosthenops australis* 来源的蛋白质 5Gly/Ala-CT_{nat} (SEQ ID NO:9) 的 DNA 序列。在 5'-和 3'-末端分别导入限制性内切酶识别位点 BamHI 和 HindIII,并通过使用经设计的引物在 HindIII 位点的上游引入终止密码子。然后将 BamHI-5Gly/Ala-CT_{nat}-HindIII 构建体亚克隆入按下面实施例 6(C)中所述进行制备的经修饰的 pET32 载体 (Merck Biosciences, Darmstadt, Germany) 中。

[0170] 实施例 6. 嵌合 MaSp1 基因的构建

[0171] (A) REP 基因片段

[0172] 使用编码 *Euprosthenops* sp MaSp1 蛋白的重复区域的部分 cDNA 克隆 (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthenops* sp. (Pisauridae). Comp Biochem Physiol B Biochem Mol Biol 138,371-376 (2004)) (GenBank 登录号 CQ974358 或 CQ816656) 作为模板,在甜菜碱 (Henke W 等人, Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 25,3957-3958 (1997)) 存在的情况下,利用

LA Taq(TaKaRa Bio;Saint Germain-en-laye,France),通过 PCR 来扩增编码标示为 3Gly/Ala 和 4Gly/Ala 的部分重复片段 (REP) 的 DNA 序列。在 5'-和 3'-末端引入限制性内切酶识别位点,从而得到下列构建体:待与 CT 片段(见下面)进行连接的 NcoI-3Gly/Ala-NheI 和 NcoI-4Gly/Ala-NheI;和待单独地表达的 NcoI-4Gly/Ala-XhoI 克隆,其中直接在 XhoI 位点的上游插入终止密码子。

[0173] (B)CT 基因片段

[0174] 使用编码 C-末端 MaSp1 结构域的基因组 DNA 克隆,通过 PCR 来扩增编码来自 *Euprostenops* sp 的非重复 C-末端结构域(但与来自棒络新妇蛛和塞内加尔络新妇蛛的 MaSp1 具有高度的相似性)的 DNA 序列(Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, Molecular studies of a novel dragline silk from a nursery web spider, *Euprostenops* sp. (Pisauridae). Comp Biochem Physiol B Biochem Mol Biol 138, 371-376(2004))。在 5'-和 3'-末端引入限制性内切酶识别位点,从而得到待与 3Gly/Ala 和 4Gly/Ala 部分 REP 克隆(见上面)进行连接的 NheI-2Gly/Ala-CT_{hyb}-XhoI,和待单独地表达的 NcoI-2Gly/Ala-CT_{hyb}-XhoI。

[0175] (C)REP-CT 杂合 MaSp1 基因的构建

[0176] 使用 pCR[®] 2.1-**TOP0**[®]载体 (Invitrogen),将 3Gly/Ala 和 4Gly/Ala REP 克隆与 CT 克隆连接在一起。然后,用 NcoI 和 XhoI 切出所得的融合的 5Gly/Ala-CT_{hyb} 和 6Gly/Ala-CT_{hyb} 克隆,并将其亚克隆入经修饰的 pET32 载体 (Novagen) 中,在该载体中除去了原始凝血酶切割位点,并在肠激酶切割位点下游引入了新的凝血酶位点。

[0177] 实施例 7. MaSp1 融合蛋白的表达

[0178] 使用经修饰的 pET32 载体:编码硫氧还蛋白/His₆/S-标签/凝血酶切割位点/MaSp1 蛋白的硫氧还蛋白-标签/His-标签/S-标签/凝血酶切割位点/MaSp1 基因和氨苄青霉素抗性基因(在 T7 启动子控制下),如下将由在实施例 5-6 中构建的基因编码的 MaSp1 蛋白表达为融合蛋白(X-REP-CT 类型)。

[0179] 将 pET32 表达载体中的不同的 MaSp1 构建体转化入大肠杆菌 BL21(DE3) 细胞 (Merck Biosciences) 中。使细胞在 30℃ 下在含有氨苄青霉素的 Luria-Bertani 培养基中生长至 1.0-1.5 的 OD₆₀₀,用 IPTG 进行诱导,并在室温下再温育 4 小时。通过离心收获细胞,和用在 20mM Tris-HCl (pH8.0)、20mM 咪唑(含 0.5M NaCl) 中的 DNA 酶 I 和溶菌酶来进行裂解,并通过在 Ni-NTA 琼脂糖 (Qiagen, West Sussex, UK) 上的 His-标签亲和层析来进一步纯化。使用 200mM 在 20mM Tris-HCl (pH8.0) (含 0.5M NaCl) 中的咪唑从 Ni-NTA 柱上洗脱结合的融合蛋白,并将其针对 20mM Tris-HCl (pH8.0) 进行透析。如通过经考马斯染色的 SDS 聚丙烯酰胺凝胶所判定的,所得到的融合蛋白的纯度 >90%,并且溶解在 20mM Tris-HCl (pH8.0) 中。该方法产生大约 40mg 融合蛋白/1 培养物,该融合蛋白在数周内是稳定的而无明显的沉淀。

[0180] 在另一个实验中,从包含有在 T7 启动子控制之下的相应基因和卡那霉素抗性基因的质粒中,将融合蛋白表达为 His₆/硫氧还蛋白/His₆/凝血酶切割位点/MaSp1 蛋白。

[0181] 实施例 8. 从 MaSp1 蛋白形成纤维

[0182] 在室温下在非常温和的摇动下,在 20mM Tris-HCl (pH8) 中,以 1:1000 (w/w) 的凝血酶:融合蛋白比率,将标签从得自实施例 7 的融合蛋白中切除。如通过 SDS-PAGE 所判定

的,在 30 至 60 分钟内完成了凝血酶切割。所得到的 MaSp1 蛋白(图 2B,SEQ ID NO :9-13)自发地聚合成不同程度的肉眼可见的纤维,参见表 1。所述纤维最初在水/空气界面处形成。从温育大约 1 小时后开始可通过肉眼观察到这种形成(参见图 4A,4B),并且在大约 5 小时后没有出现进一步的纤维生长。6Gly/Ala-CT_{hyb} 纤维的长度直到大约 2cm,并且 5Gly/Ala-CT_{nat} 纤维的长度 $\geq 10\text{cm}$ 。重复的实验产生长度 $\geq 20\text{cm}$ (参见图 4C),甚至 $\geq 2\text{m}$ 的 5Gly/Ala-CT_{nat} 纤维。从温育大约 10 分钟后开始可通过肉眼观察到纤维形成。

[0183] 分离纤维,并使用缓冲液进行洗涤,然后使其经历 N-末端氨基酸序列分析,所述分析只显示了 MaSp1 蛋白的序列。这表明,在所述纤维中不存在经切割的标签。

[0184] 实施例 9. MaSp1 蛋白纤维的分析

[0185] A. 拉伸强度测量

[0186] 如下测定实施例 8 的 6Gly/Ala-CT_{hyb}(SEQ ID NO :13)和 5Gly/Ala-CT_{nat}(SEQ ID NO :9)纤维的拉伸强度。为了操作较短(1-2cm)的 6Gly/Ala-CT_{hyb} 纤维以进行拉伸强度测量,在将它们进行空气干燥之前,将它们在 15%的甘油水溶液中短时间温育。较长(10cm)的 5Gly/Ala-CT_{nat} 纤维在空气干燥之前不进行处理,在 15%的甘油中短时间温育,或在 75%的甲醇中用手进行拉伸。通过在 Zwick MaterialTester 中以 10mm/分钟的速率牵拉所述纤维来测量经空气干燥的纤维的拉伸强度。参见表 1。

[0187] 经甘油处理的、经空气干燥的、1-2cm 长的来自 6Gly/Ala-CT_{hyb}(SEQ ID NO :13)的纤维的拉伸强度为大约 2MPa,来自 5Gly/Ala-CT_{nat}(SEQ ID NO :9)的 10cm 纤维的强度为 4-5MPa。在空气干燥之前在脱水溶剂甲醇中进行拉伸的 10cm 长的 5Gly/Ala-CT_{nat} 纤维展示出 2-3MPa 的拉伸强度,该拉伸强度略低于经甘油处理的相同类型的纤维。目前测得的最高拉伸强度为 10MPa,该强度是对于无进一步处理的经空气干燥的 10cm 长的 5Gly/Ala-CT_{nat} 纤维而发现的。

[0188] 拉伸强度的范围(2-10MPa)与对于再生的蜘蛛丝纤维而报导的较低值(2-320MPa)相当。最长的自发形成的纤维来源于 5Gly/Ala-CT_{nat} 构建本,并且此类经空气干燥的纤维也显示出最大的拉伸强度。可能地,这可以是由于其 12-15 个残基长的多聚 Ala 区段(相对于在 6Gly/Ala-CT_{hyb} 中的 8-14 个残基的 Ala 区段),该区段可在先前的蛋白质中产生更大比例的结晶 β -折叠构象。

[0189]

表 1. MaSp1 蛋白的纤维形成能力

蛋白质	SEQ ID NO	纤维形成能力	纤维长度 (cm)	纤维拉伸强度 (Mpa)
5Gly/Ala-CT _{nat}	9	++++	> 10	2-10
5Gly/Ala-CT _{nat} (实施例 11)	9	+++++	> 20	180-230
4Gly/Ala	10	聚集体	n. a.	n. a.
2Gly/Ala-CT _{hyb}	11	聚集体	n. a.	n. a.
5Gly/Ala-CT _{hyb}	12	+	< 1	n. d.
6Gly/Ala-CT _{hyb}	13	+++	1-2	< 2
n. a. = 不可应用				
n. d. = 未测定				

[0190] B. 扫描电子显微术

[0191] 使用扫描电子显微术 (SEM) 来分析 6Gly/Ala-CT_{hyb} 和 5Gly/Ala-CT_{nat} 纤维的显微结构 (图 5)。简而言之,将样品放置在 SEM- 样品台上,并用 6nm 的金和铂层真空覆盖所述样品。使用 10kV 的加速电压,在 LEO 1550 FEG SEM 中观察试样并进行照相。

[0192] 该分析揭示出单根纤维的直径为 10-30 μm , 其中单个纤维展示出相当均一的直径 (图 5a, 其显示了 6Gly/Ala-CT_{hyb}, SEQ ID NO:13)。除了肉眼可见的纤维外,还发现了凝胶样颗粒。于直接在 SEM- 样品台上对 6Gly/Ala-CT_{hyb} 的此类颗粒进行空气干燥后,看到直径大约 10-15 μm 的纤维 (图 5b, c)。在 75% 甲醇中进行拉伸并经空气干燥的 5Gly/Ala-CT_{nat} (SEQ ID NO:9) 的肉眼可见的纤维的直径为 60-120 μm , 并且它们似乎包含几种经排列的纤维 (图 5d-f)。在空气干燥前扭曲的纤维 (图 5e), 纤维的末端 (图 5f)。

[0193] C. 圆二色性光谱学

[0194] 在 20mM 磷酸盐缓冲液 (pH7) 中洗涤在实施例 8 中制备的由 6Gly/Ala-CT_{hyb} 蛋白 (SEQ ID NO:13) 或 5Gly/Ala-CT_{nat} (SEQ ID NO:9) 组成的纤维,并将所述纤维悬浮于在相同缓冲液中的 2% SDS 之中。使用 Jasco J-810 旋光分光计,在 22°C 下在 0.1cm 光程长度的石英杯中记录从 250 至 190nm 的圆二色性光谱。扫描速度为 50nm/ 分钟,响应时间为 2 秒,采集间隔为 0.1nm,和带宽为 1nm。

[0195] 图 6 中显示的光谱是 6Gly/Ala-CT_{hyb} 蛋白 (SEQ ID NO:13) 的纤维的三次扫描的累积。其在 220nm 处展示出最小值和在 195nm 处展示出最大值,这些特征是反平行 β - 折叠结构所特有的。对于 5Gly/Ala-CT_{nat} 的纤维获得了的高度相似的光谱 (未显示)。因此,所述自发形成的纤维展示了与天然的和再生的蜘蛛丝纤维相似的形态学和结构。

[0196] 实施例 10. 重组蜘蛛丝的生物相容性

[0197] 因为希望将蜘蛛丝纤维用于生物医学应用,因此通过使用两种不同细胞类型研究重组产生的丝的效应来评估纤维的生物相容性。

[0198] 如实施例 7-8 中所述,在细菌中表达 MaSp1 蛋白 5Gly/Ala-CT_{nat} (SEQ ID NO:9)。将纯化的蛋白质用于产生具有 >10cm, 甚至 >20-200cm 的长度和大约 100 μm 的直径的人造丝纤维。

[0199] A. 胚胎小鼠肥大细胞

[0200] 以两种不同的细胞密度接种胚胎（第 12.5 天）小鼠肥大细胞（使用 IL-8 和肥大干细胞因子体外增殖 8 周），较高的密度为较低密度的大约 4 倍。这些细胞不粘着至塑料表面，但以悬浮形式生长。向小孔中加入各自大约 0.5cm 长的丝纤维小段。在丝纤维存在或不存在的条件下，将肥大细胞温育 3 天，并在这之后在用台盼蓝染色后计数活细胞和死细胞（图 7）。这些柱条显示了具有标准误差的平均值， $n = 2$ ，每个样品重复计数 3 次。

[0201] 丝纤维的存在不影响肥大细胞。在 3 天的生长后，与在没有丝纤维的情况下生长的负对照相比，细胞的死亡或增殖没有显著的差异。

[0202] B. 人胚胎肾 (HEK) 293 细胞

[0203] 通过让它们从小体积的缓冲液中干燥来将大约 0.5cm 长的丝纤维小段吸附至 6 孔微量滴定板的底部。当加入细胞生长培养基时，纤维不脱离。然后，以不同的细胞密度将人胚胎肾 (HEK) 293 细胞进行铺板，并让其生长总共 6 天。HEK-293 细胞粘着并生长附着至塑料小室的表面。研究 HEK-293 细胞在纤维附近生长的能力，和细胞对纤维的物理附着。

[0204] HEK-293 细胞在包含丝纤维的小孔中正常附着和生长（如在光学显微镜下观察到的）。细胞非常紧密地沿着纤维边缘生长，并且很明显地，甚至在部分脱离的纤维下生长（图 8）。7 天后，小心地从塑料表面上分离纤维，并且清楚地看到细胞群与纤维发生物理附着。纤维覆盖了该图的右上一半部分。看到 HEK293 细胞附着至纤维的边缘，并且也在纤维下面生长。

[0205] 重组丝纤维的存在不影响所研究的两种不同的细胞类型（肥大细胞、HEK-293），即使在相当高的数量的丝存在下亦如此。这表明，受试人造丝纤维在非毒性和生物相容性方面与 *Euprosthopsa australis* 的野生型拖丝相似。因此，所述人造丝纤维看起来适合用于生物医学应用。

[0206] 实施例 11. MaSp1 蛋白纤维的机械性质和结构

[0207] 使用用于产生应力 - 应变曲线而实施的拉伸试验来检查来自 5Gly/Ala-CT_{nat} (SEQ ID NO :9) 的纤维的机械性质（图 9）。使用 ZwickRoell Z2.5 材料测试器 (Zwick, Ulm, Germany) 来表征拉伸性能。在环境条件 (20°C 和 52% 相对湿度) 下，使用 10mm/ 分钟的负载速度在空气中进行所述试验。直接从缓冲液中转移出纤维小段，安放，和经历 2 轮绷紧 - 松弛循环。为了产生适合于拉伸试验的均一的丝线，使用绷紧 - 松弛循环来伸长纤维。首先，通过牵拉直至 0.1N 的力来使纤维伸长。在松弛后，再次拉伸它们直至施加了 0.25N 的力。

[0208] 该处理产生具有大约 80 μ m 的直径（通过使用 Mitutoyo IDC-112B 设备 (Mitutoyo Corp, Tokyo, Japan) 进行高度测量所确定的，并如下通过扫描电子显微术 (SEM) 所确认的) 的伸长的均一纤维。在绷紧 - 松弛循环之前和之后，将纤维小段施加在 SEM 样品台上并在空气干燥过夜。使用 6nm 的金和铂层真空覆盖所述样品。使用 10kV 的加速电压，用 LEO 1550 FEG 显微镜 (Carl Zeiss, Oberkochen, Germany) 观察试样并进行照相。

[0209] 将经拉抽的纤维切成小段，用胶 (Loctite 420, Loctite, **Göteborg**, Sweden) 将所述小段的末端固定在卡纸板纸 (cardboard paper) 之间。然后将纤维样品固定在材料测试仪的夹具 (grip) 上，并绷紧直至它们断裂。使用拉伸前的纤维的初始横截面积（假定为圆形横截面）来构建应力 - 应变曲线。将应力值对纤维的初始横截面积进行标准化。应变

相应于 dL/L_0 , 其中 L_0 是纤维的初始长度, 和 dL 是纤维长度的变化。在图 9 中, 显示了 5Gly/Ala-CT_{nat} (SEQ ID NO :9) 的双拉伸纤维的三个不同样品的应力-应变曲线, 和经测量为大约 0.2GPa 的它们的拉伸强度。

[0210] 通过 SEM 来分析纤维的显微构造 (图 10)。自发形成的纤维具有均一的扁平的外观和直至数百微米的宽度, 而高度经测量为大约 10 微米 (图 10a, b)。

[0211] 在纤维经历绷紧-松弛循环后, 它们的横截面采取更圆的形状, 该性状具有紧密排列的原纤维的致密的亚结构 (图 10c-f)。切面或断裂面的外观 (图 10e, f) 进一步证明了所产生的纤维的致密性。

[0212] 总之, 自发形成的纤维显示出与天然的或再生的蜘蛛丝纤维相似的形态学和机械性质, 即使在不进行纺丝的情况下。

[0213] 实施例 12. 蛛丝蛋白变体

[0214] 强的分子间相互作用被认为促成了蜘蛛丝的令人印象深刻的拉伸强度。因此, 产生了小型蛛丝蛋白的变体, 其允许纤维中的分子间共价交联。通过定向诱变来构建两种不同的突变型蛛丝蛋白, 所述定向诱变分别在第一 (SEQ ID NO :14, 位置 36 和 37) 和第四 (SEQ ID NO :15, 位置 128 和 129) 丙氨酸嵌段 (alanine block) 中引入 2 个半胱氨酸残基。表达这些变体, 并使用与在实施例 7-8 中对于在实施例 5-6 中构建的基因所描述的方案相同的方案来进行分离。这些变体 (SEQ ID NO :14-15) 以与 5Gly/Ala-CT_{nat} (SEQ ID NO :9) 相同的方式形成纤维。

[0215] 为了阐明 C-末端结构域的二聚化的重要性, 已构建了一种变体, 其中将 C-末端结构域中的半胱氨酸残基交换为丝氨酸残基 (SEQ ID NO :16, 位置 222)。然而, 该变体 (SEQ ID NO :16) 以与 5Gly/Ala-CT_{nat} (SEQ ID NO :9) 相同的方式形成纤维。

[0216] 实施 13. 从表达的蛛丝蛋白中除去 LPS 和其他致热原

[0217] 使用下列缓冲液来洗涤表达所需蛛丝蛋白融合蛋白的大肠杆菌细胞:

[0218] A: 100mM Tris, pH8,

[0219] B: 5mM CaCl₂, 100mM Tris, pH8,

[0220] C: 10mM EDTA, 100mM Tris, pH8,

[0221] D: 100mM Tris, pH8, 和

[0222] E: 100mM Tris, pH8。

[0223] 然后, 在补充有溶菌酶和 DNA 酶 I 的 20mM Tris (pH8) 中裂解细胞。然后将蛋白质样品加载至 Ni-sepharose 基质上, 并用 20mM Tris, 10-100mM 咪唑, pH8 进行洗涤, 随后用 20mM Tris, 100-300mM 咪唑, pH8 进行洗脱。汇集相关级分, 并针对 20mM Tris (pH8) 透析过夜。然后用 100 μ M CaCl₂ 补充蛋白质样品, 最后使其通过事先用 20mM Tris, 100 μ M CaCl₂, pH8 平衡的 EndoTrap Blue 柱。以这种方式, 可获得致热原含量为 1EU/mg 蛋白质的蛋白质样品, 如通过 IPT 和 LAL 动力学测定法所判定的。

[0224] 然后, 以 1:1000 (w/w) 凝血酶:融合蛋白比例, 用凝血酶通过蛋白水解来切割融合蛋白, 这引起纤维形成 (如上所述的)。将纤维在 20mM Tris (pH8) 中洗涤 3 次, 最后在水中洗涤 3 次。这产生了致热原含量为 0.25EU/mg 纤维的纤维。

[0225] 在 125°C 和 1.5 巴下高压灭菌 10 分钟后, 纤维的结构特征未受影响, 这使得能够对所述材料进行有效的灭菌。所述纤维在化学上是稳定的, 并且不能溶解于 8M 尿素、6M GuaHCl 或纯 HAc 中。然而, 所述纤维可溶解在纯 HFIP 或甲酸中。

[0001]

序列表

- <110> Johansson, Jan
Grip, Stefan
Hjältn, Göran
Rising, Anna
Stark, Margareta
Engström, Wilhelm
Hedhammar, My
- <120> 蜘蛛丝蛋白和用于产生蜘蛛丝蛋白的方法
- <130> PC-21028983
- <150> SE 0502932-7
<151> 2005-12-30
- <160> 16
- <170> PatentIn version 3.3
- <210> 1
<211> 3766
<212> DNA
<213> Euprosthénops australis
- <400> 1
gtcaaggtgc tggaggtaat gccgctgcag cagccgcagc agcagcagca gcagcagctg 60
gacagggcgg tcaaggtgga tatggtggac taggtcaagg aggatatgga cagggctgcag 120
gaagttctgc agccgccgcc gccgcagcag cagcagcagc tgcagcagct ggacgaggtc 180
aaggaggata tggtaaggt tctggaggta atgccgctgc agcagccgca gcagctgcag 240
cagcagcatc tggacaagga ggtcaaggag gacaaggtgg acaaggtcaa ggtggatatg 300
gacaaggtgc aggaagttct gcagccgccg ccgccgcagc agcagcagcc gccgcagcag 360
ctggacaagg tcaaggacga tatggtcaag gtgctggagg laatgccgct gcagcagccg 420
cagcagctgc agcagcagca gctggacaag gaggtcaagg aggacaaggt ggactaggtc 480
aaggaggata tggacaaggt gcaggaagtt ctgcagccgc cggccgcagca tcagcagccg 540

[0002]

cgcagcagc tggacgaggt caaggaggat atggtcaagg tgcaggaggt aatgccgctg	600
cagcagccgc agcagctgcc gccgccgcag cagctggaca gggtaggtcaa ggtggatatg	660
gtggactagg tcaaggagga tatggacaag gtgcaggaag ttctgcagcc gctgccgccg	720
cagcagcagc agccgccgcc gcaggtggac aaggtggaca aggtcaagga agatatggac	780
aaggtgcagg aagtcttgc gccgctgccg ccgcagcagc agcagccgcc gcagcagctg	840
gacaaggtca aggaggatat ggtcaagggt ctggaggtaa tgccgctgca gcagccgcag	900
cagctgcagc agcagcagct ggacaaggag gtcaaggagg acaaggtgga ctaggtcaag	960
gaggatatgg acaaggtgca ggaagttctg ccgccgccgc cgcagcagca gcagccgccg	1020
cagcagctgg acgagggtcaa ggaggatatg gtcaagggtc tggaggtaat gccgctgcag	1080
cagccgcagc agctgccgaa gccgcagcag ctggacaggg tggtaagggt ggatatgggtg	1140
gactaggta aggaggatat ggacaagggt caggaagttc tgcagccgcc gccgcagcag	1200
cagcagccgc cgcagcagct ggacgaggtc aaggaggata tggtaagggt gctggaggta	1260
atgccgctgc agcagccgca gcagctgccg ccgccgcagc agctggacag ggtggtcaag	1320
gtggatatgg tggactaggt caaggaggat atggacaagg tgcaggaagt tctgcagccg	1380
ctgccgccgc agcagcagca gccgccgccg caggtggaca aggtggacaa ggtcaaggaa	1440
gatatggaca aggtgcagga agttctgcag ccgctgccgc cgcagcagca gcagcagccg	1500
cagcagctgg acgagggtcaa ggaggatatg gtcaagggtc tggaggtaat gccgctgcag	1560
cagccgcagc agctgcagca gcagcatctg gacaaggaag tcaaggagga caaggtggac	1620
aaggtcaagg tggatatgga caaggtgcag gaagttctgc agccgccgcc gccgccgcag	1680
cagcagccgc cgcagcatct ggacgaggtc aaggaggata tggtaagggt gctggaggta	1740
atgccgctgc tgcagccgca gcagctgccg ccgccgcagc agctggacag ggcggtcaag	1800
giggatatgg tggactaggt caaggaggat atggacaagg tgcaggaagt tctgcagccg	1860

[0003]

ctgccgccgc cgcagcagcc gccgcagcag gtggacaagg tggacaaggt caaggaggat	1920
atggacaagg tgcaggaagt tctgcagccg ccgccgcagc agcagcagca gcagccgcag	1980
cagctggacg aggtcaagga ggatatggtc aaggttcttg aggtaatgcc gctgcagcag	2040
ccgcagcagc tgcagcagca gcctctggac aaggaggtca aggaggacaa ggtggacaag	2100
gtcaaggtgg rtatggacaa ggtgcaggaa gttctgcagc cgcgccgcc gcagcagcag	2160
cagccgccgc agcagctgga caaggtcaag gaggatatgg tcaagtgct ggaggtaatg	2220
ccgctgcagc agccgcagca gctgcagcag cagcagctgg acaaggaggt caaggaggac	2280
aaggtggact aggtcaagga ggatatggac aaggtgcagg aagttctgca gccgccgccg	2340
cagcmgcmgc agcagccgcc gcagcagctg gacgaggtca aggaggatat ggtcaaggtg	2400
ttggaggtaa tgccgctgca gcagccgcag cagctgcagc agcagcagct ggacaaggag	2460
gtcaaggagg acaaggtgga ctaggtcaag gaggatatgg acaaggtgca ggtagttctg	2520
cagccgccgc cgcgccgca gcagcagccg ccgcagcagc tggacgaggt caaggaggat	2580
atgttcaagg ttctggaggt aatgccgctg cagcagccgc agcagctgca gcagcagcat	2640
ctggacaagg aagtcaagga ggacaaggtg gacaaggtca aggtggatat ggacaaggtg	2700
caggaagttc tgcagccgcc gccgccgcag cagcagcagc cgcgcagca tctggacgag	2760
gtcaaggagg atatggtcaa ggtgctggag gtaatgccgc tgctgcagcc gcagcagctg	2820
ccgccgccgc agcagctgga cagggcggtc aaggtggata tggaggacta ggtcaaggag	2880
gatatggaca aggtgcagga agttctgcag ccgctgccgc cgcgcagca gccgccgcag	2940
caggtggaca aggtggacaa ggtcaaggag gatattggaca aggttcagga ggttctgcag	3000
ccgccgccgc cgcgcagca gcagcagcag ctgcagcagc tggacgaggt caaggaggat	3060
atgttcaagg ttctggaggt aatgtgctg ccgcagccgc tgccgccgcc gccgccgctg	3120
cagcagccgc acaggaggt caaggtggat atggtagaca aagccaaggt gctggttccg	3180

[0004]

ctgctgctgc tgctgctgct gctgccgctg ctgctgctgc aggatctgga caaggtggat 3240
 acggtggaca aggtcaagga ggttatggtc agagtagtgc ttctgcttca gctgctgcgt 3300
 cagctgctag tactgtagct aattcggtga gtcgcctctc atcgccctcc gcagtatctc 3360
 gagttttcttc agcagtttct agcttggttt caaatgggtca agtgaatatg gcagcggttac 3420
 ctaatatcat ttccaacatt tctttctctg tcagtcgcatc tgctcctggt gcttctggat 3480
 gigaggicac agtgcaagct ctactcgaag tcatcactgc tcttggtcaa atcgttagtt 3540
 cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat gttgttgcta 3600
 atgccatggc tcaagtaatg ggctgaggtt tttaatagta aaaggtgtga tattctctcaa 3660
 tgttttgaaa attattaatc gaatttttac ctigtgtgct atcagatata aattgaagta 3720
 taataaataa atatttgcac ttcaaaaaa aaaaaaaaaa aaaaaa 3766

<210> 2

<211> 1207

<212> PRT

<213> Euprostenops australis

<400> 2

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
 1 5 10 15

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
 20 25 30

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala
 35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly
 50 55 60

[0005]

Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ser Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln
85 90 95

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
100 105 110

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Arg Tyr Gly
115 120 125

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
130 135 140

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln
145 150 155 160

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
165 170 175

Ser Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln
180 185 190

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
210 215 220

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
225 230 235 240

[0006]

Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly
245 250 255

Arg Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
260 265 270

Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln
275 280 285

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
290 295 300

Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly
305 310 315 320

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
325 330 335

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
340 345 350

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Glu Ala Ala
355 360 365

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
370 375 380

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
385 390 395 400

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
405 410 415

[0007]

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
420 425 430

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
435 440 445

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
450 455 460

Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Arg
465 470 475 480

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
485 490 495

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
500 505 510

Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
515 520 525

Ser Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly
530 535 540

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
545 550 555 560

Ala Ala Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
565 570 575

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
580 585 590

[0008]

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
595 600 605

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
610 615 620

Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
625 630 635 640

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
645 650 655

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
660 665 670

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
675 680 685

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
690 695 700

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
705 710 715 720

Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala
725 730 735

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
740 745 750

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
755 760 765

[0009]

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
770 775 780

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Val
785 790 795 800

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
805 810 815

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
820 825 830

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
835 840 845

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
850 855 860

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
865 870 875 880

Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
885 890 895

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
900 905 910

Ala Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala
915 920 925

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
930 935 940

[0010]

Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly
945 950 955 960

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
965 970 975

Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly
980 985 990

Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
995 1000 1005

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
1010 1015 1020

Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
1025 1030 1035

Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln
1040 1045 1050

Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
1055 1060 1065

Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln
1070 1075 1080

Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
1085 1090 1095

Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser
1100 1105 1110

[0011]

Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu
1115 1120 1125

Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile
1130 1135 1140

Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser
1145 1150 1155

Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
1160 1165 1170

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro
1175 1180 1185

Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala
1190 1195 1200

Gln Val Met Gly
1205

<210> 3

<211> 1110

<212> PRT

<213> Euprosthénops australis

<220>

<221> REPEAT

<222> (7)..(19)

<220>

<221> REPEAT

<222> (20)..(42)

[0012]

<220>	
<221>	REPEAT
<222>	(43).. (56)
<220>	
<221>	REPEAT
<222>	(57).. (70)
<220>	
<221>	REPEAT
<222>	(71).. (83)
<220>	
<221>	REPEAT
<222>	(84).. (106)
<220>	
<221>	REPEAT
<222>	(107).. (120)
<220>	
<221>	REPEAT
<222>	(121).. (134)
<220>	
<221>	REPEAT
<222>	(135).. (147)
<220>	
<221>	REPEAT
<222>	(148).. (170)
<220>	
<221>	REPEAT
<222>	(171).. (183)
<220>	
<221>	REPEAT
<222>	(184).. (197)
<220>	
<221>	REPEAT
<222>	(198).. (211)

[0013]

<220>
<221> REPEAT
<222> (212).. (234)

<220>
<221> REPEAT
<222> (235).. (248)

<220>
<221> REPEAT
<222> (249).. (265)

<220>
<221> REPEAT
<222> (266).. (279)

<220>
<221> REPEAT
<222> (280).. (293)

<220>
<221> REPEAT
<222> (294).. (306)

<220>
<221> REPEAT
<222> (307).. (329)

<220>
<221> REPEAT
<222> (330).. (342)

<220>
<221> REPEAT
<222> (343).. (356)

<220>
<221> REPEAT
<222> (357).. (370)

<220>
<221> REPEAT
<222> (371).. (393)

[0014]

<220>
<221> REPEAT
<222> (394).. (406)

<220>
<221> REPEAT
<222> (407).. (420)

<220>
<221> REPEAT
<222> (421).. (434)

<220>
<221> REPEAT
<222> (435).. (457)

<220>
<221> REPEAT
<222> (458).. (470)

<220>
<221> REPEAT
<222> (471).. (488)

<220>
<221> REPEAT
<222> (489).. (502)

<220>
<221> REPEAT
<222> (503).. (516)

<220>
<221> REPEAT
<222> (517).. (529)

<220>
<221> REPEAT
<222> (530).. (552)

<220>
<221> REPEAT
<222> (553).. (566)

[0015]

<220>
<221> REPEAT
<222> (567).. (580)

<220>
<221> REPEAT
<222> (581).. (594)

<220>
<221> REPEAT
<222> (595).. (617)

<220>
<221> REPEAT
<222> (618).. (630)

<220>
<221> REPEAT
<222> (631).. (647)

<220>
<221> REPEAT
<222> (648).. (661)

<220>
<221> REPEAT
<222> (662).. (675)

<220>
<221> REPEAT
<222> (676).. (688)

<220>
<221> REPEAT
<222> (689).. (711)

<220>
<221> REPEAT
<222> (712).. (725)

<220>
<221> REPEAT
<222> (726).. (739)

[0016]

<220>
<221> REPEAT
<222> (740).. (752)

<220>
<221> REPEAT
<222> (753).. (775)

<220>
<221> REPEAT
<222> (776).. (789)

<220>
<221> REPEAT
<222> (790).. (803)

<220>
<221> REPEAT
<222> (804).. (816)

<220>
<221> REPEAT
<222> (817).. (839)

<220>
<221> REPEAT
<222> (840).. (853)

<220>
<221> REPEAT
<222> (854).. (867)

<220>
<221> REPEAT
<222> (868).. (880)

<220>
<221> REPEAT
<222> (881).. (903)

<220>
<221> REPEAT
<222> (904).. (917)

[0017]

<220>
<221> REPEAT
<222> (918).. (931)

<220>
<221> REPEAT
<222> (932).. (945)

<220>

<221> REPEAT
<222> (946).. (968)

<220>
<221> REPEAT
<222> (969).. (981)

<220>
<221> REPEAT
<222> (982).. (998)

<220>
<221> REPEAT
<222> (999).. (1013)

<220>
<221> REPEAT
<222> (1014).. (1027)

<220>
<221> REPEAT
<222> (1028).. (1042)

<220>
<221> REPEAT
<222> (1043).. (1059)

<220>
<221> REPEAT
<222> (1060).. (1073)

<220>
<221> REPEAT

[0018]

<222> (1074)..(1092)

<400> 3

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
1 5 10 15

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
20 25 30

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala
35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly
50 55 60

Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ser Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln
85 90 95

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
100 105 110

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Arg Tyr Gly
115 120 125

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
130 135 140

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln
145 150 155 160

[0019]

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
165 170 175

Ser Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln
180 185 190

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
210 215 220

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
225 230 235 240

Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly
245 250 255

Arg Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
260 265 270

Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln
275 280 285

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
290 295 300

Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly
305 310 315 320

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
325 330 335

[0020]

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
340 345 350

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Glu Ala Ala
355 360 365

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
370 375 380

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
385 390 395 400

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
405 410 415

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
420 425 430

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
435 440 445

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
450 455 460

Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Arg
465 470 475 480

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
485 490 495

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
500 505 510

[0021]

Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
515 520 525

Ser Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly
530 535 540

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
545 550 555 560

Ala Ala Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
565 570 575

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
580 585 590

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
595 600 605

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
610 615 620

Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
625 630 635 640

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
645 650 655

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
660 665 670

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
675 680 685

[0022]

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
690 695 700

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
705 710 715 720

Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala
725 730 735

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
740 745 750

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
755 760 765

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
770 775 780

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Val
785 790 795 800

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
805 810 815

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
820 825 830

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
835 840 845

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
850 855 860

[0023]

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
865 870 875 880

Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
885 890 895

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
900 905 910

Ala Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala
915 920 925

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
930 935 940

Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly
945 950 955 960

Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
965 970 975

Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly
980 985 990

Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
995 1000 1005

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
1010 1015 1020

Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
1025 1030 1035

[0024]

Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln
1040 1045 1050

Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
1055 1060 1065

Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln
1070 1075 1080

Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
1085 1090 1095

Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser
1100 1105 1110

<210> 4
<211> 97
<212> PRT
<213> *Euprosthénops australis*

<400> 4

Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser
1 5 10 15

Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile
20 25 30

Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser
35 40 45

Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu
50 55 60

[0025]

Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala
65 70 75 80

Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met
85 90 95

Gly

<210> 5

<211> 23

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthénops australis* MaSp1 的内部重复序列的共有序列

<220>

<221> VARIANT

<222> (4)..(4)

<223> Ser

<220>

<221> VARIANT

<222> (8)..(8)

<223> Tyr

<220>

<221> VARIANT

<222> (11)..(11)

<223> Gln

<400> 5

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
1 5 10 15

Gly Gln Gly Ala Gly Ser Ser

[0026]

20

<210> 6
<211> 17
<212> PRT
<213> 人工的

<220>
<223> 来源于 *Euprosthonops australis* MaSp1 的内部重复序列的共有序列

<220>
<221> VARIANT
<222> (9)..(9)
<223> Arg

<220>
<221> VARIANT
<222> (14)..(14)
<223> Ser

<220>
<221> VARIANT
<222> (16)..(16)
<223> Gly

<400> 6

Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser
1 5 10 15

Ser

<210> 7
<211> 14
<212> PRT
<213> 人工的

<220>
<223> 来源于 *Euprosthonops australis* MaSp1 的内部重复序列的共有序列

[0027]

<220>

<221> VARIANT

<222> (2).. (2)

<223> Gln

<220>

<221> VARIANT

<222> (6).. (6)

<223> Arg

<220>

<221> VARIANT

<222> (11).. (11)

<223> Ser

<220>

<221> VARIANT

<222> (11).. (11)

<223> Val

<400> 7

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn

1

5

10

<210> 8

<211> 100

<212> PRT

<213> 人工的

<220>

<223> 来源于已知的 MaSp1 和 MaSp2 蛋白的共有序列

<220>

<221> MISC_FEATURE

<222> (1).. (71)

<223> 在已知物种变体中存在的序列长度

<220>

<221> VARIANT

<222> (7).. (7)

[0028]

<223> Glu

<400> 8

Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Asn
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Val His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val Gly Gln Ser Val Ala Gln Ala
85 90 95

Leu Gly Glu Phe
100

<210> 9

<211> 269

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthonops australis* MaSp1 的序列

<220>

<221> DOMAIN

<222> (1)..(171)

[0029]

<223> 重复片段

<220>

<221> REPEAT

<222> (11).. (30)

<220>

<221> REPEAT

<222> (31).. (42)

<220>

<221> REPEAT

<222> (43).. (60)

<220>

<221> REPEAT

<222> (61).. (75)

<220>

<221> REPEAT

<222> (76).. (89)

<220>

<221> REPEAT

<222> (90).. (104)

<220>

<221> REPEAT

<222> (105).. (121)

<220>

<221> REPEAT

<222> (122).. (135)

<220>

<221> REPEAT

<222> (136).. (152)

<220>

<221> REPEAT

<222> (153).. (171)

<220>

<221> DOMAIN

[0030]

<222> (172)..(269)

<223> C-末端片段

<400> 9

Gly Ser Ala Met Gly Tyr Leu Trp Ile Gln Gly Gln Gly Tyr Gly
1 5 10 15

Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln
35 40 45

Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly
65 70 75 80

Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
100 105 110

Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala
115 120 125

Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly
130 135 140

Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
145 150 155 160

[0031]

Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser
165 170 175

Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser
180 185 190

Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile
195 200 205

Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val
210 215 220

Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val
225 230 235 240

Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile
245 250 255

Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
260 265

<210> 10
<211> 126
<212> PRT
<213> Euprosthénops australis

<400> 10

Gly Ser Ala Met Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala
1 5 10 15

Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Gln Gly Gly Gln Gly Gly
20 25 30

[0032]

Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ile Ser Ala Ala Ala
35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Gly Ala Ala Ala Gly Arg Gly Gln
50 55 60

Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
85 90 95

Gly Gly Gln Gly Leu Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala
100 105 110

Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Ser Val Tyr
115 120 125

<210> 11

<211> 152

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthénops australis* 和 *Euprosthénops* sp MaSpl 的杂合序列

<400> 11

Gly Ser Ala Met Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly
1 5 10 15

Gly Gln Gly Ala Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala
20 25 30

Ala Ala Gly Gly Ser Gly Gln Gly Gly Tyr Gly Gly Val Gly Ser Gly
35 40 45

[0033]

Ala Ser Ala Ala Ser Ala Ala Ala Ser Arg Leu Ser Ser Pro Glu Ala
50 55 60

Ser Ser Arg Val Ser Ser Ala Val Ser Asn Leu Val Ser Ser Gly Pro
65 70 75 80

Thr Asn Ser Ala Ala Leu Ser Ser Thr Ile Ser Asn Val Val Ser Gln
85 90 95

Ile Gly Ala Ser Asn Pro Gly Leu Ser Gly Cys Asp Val Leu Val Gln
100 105 110

Ala Leu Leu Glu Val Val Ser Ala Leu Ile His Ile Leu Gly Ser Ser
115 120 125

Ser Ile Gly Gln Val Asn Tyr Gly Ser Ala Gly Gln Ala Thr Gln Leu
130 135 140

Val Gly Gln Ser Val Tyr Gln Ala
145 150

<210> 12

<211> 257

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthénops australis* 和 *Euprosthénops* sp MaSpl 的杂合序列

<400> 12

Gly Ser Ala Met Ala Ser Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln
1 5 10 15

[0034]

Gly Gly Tyr Gly Gln Gly Ala Gly Ile Ser Ala Ala Ala Ala Ala Ala
20 25 30

Ala Ala Ala Ala Ala Gly Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
35 40 45

Gly Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Gln
65 70 75 80

Gly Leu Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Gly Arg Gly Gln Ser Val Tyr Ala Ser Gly Gly Ala
100 105 110

Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gly Gln Gly Ala Gly Arg Gly
115 120 125

Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ser Gly Gln
130 135 140

Gly Gly Tyr Gly Gly Val Gly Ser Gly Ala Ser Ala Ala Ser Ala Ala
145 150 155 160

Ala Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala
165 170 175

Val Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser
180 185 190

[0035]

Ser Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser Asn Pro Gly
195 200 205

Leu Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser
210 215 220

Ala Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr
225 230 235 240

Gly Ser Ala Gly Gln Ala Thr Gln Leu Val Gly Gln Ser Val Tyr Gln
245 250 255

Ala

<210> 13

<211> 276

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthénops australis* 和 *Euprosthénops* sp MaSp1 的杂合序列

<220>

<221> REPEAT

<222> (1)..(11)

<220>

<221> DOMAIN

<222> (1)..(180)

<223> 重复片段

<220>

<221> REPEAT

<222> (12)..(25)

<220>

<221> REPEAT

[0036]

<222> (26).. (45)

<220>

<221> REPEAT

<222> (46).. (60)

<220>

<221> REPEAT

<222> (61).. (74)

<220>

<221> REPEAT

<222> (75).. (88)

<220>

<221> REPEAT

<222> (89).. (111)

<220>

<221> REPEAT

<222> (112).. (119)

<220>

<221> REPEAT

<222> (120).. (149)

<220>

<221> REPEAT

<222> (150).. (158)

<220>

<221> REPEAT

<222> (159).. (172)

<220>

<221> REPEAT

<222> (173).. (180)

<220>

<221> DOMAIN

<222> (181).. (276)

<223> C-末端片段

<400> 13

[0037]

Gly Ser Ala Met Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala
1 5 10 15

Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Gln Gly Gly Gln Gly Gly
20 25 30

Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ile Ser Ala Ala Ala
35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Gly Ala Ala Ala Gly Arg Gly Gln
50 55 60

Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
85 90 95

Gly Gly Gln Gly Leu Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala
100 105 110

Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Ser Val Tyr Ala Ser
115 120 125

Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gly Gln Gly Ala
130 135 140

Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly
145 150 155 160

Ser Gly Gln Gly Gly Tyr Gly Gly Val Gly Ser Gly Ala Ser Ala Ala
165 170 175

[0038]

Ser Ala Ala Ala Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val
180 185 190

Ser Ser Ala Val Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala
195 200 205

Ala Leu Ser Ser Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser
210 215 220

Asn Pro Gly Leu Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu
225 230 235 240

Val Val Ser Ala Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln
245 250 255

Val Asn Tyr Gly Ser Ala Gly Gln Ala Thr Gln Leu Val Gly Gln Ser
260 265 270

Val Tyr Gln Ala
275

<210> 14

<211> 269

<212> PRT

<213> 人工的

<220>

<223> 来源于 *Euprosthennops australis* MaSp1 的序列

<220>

<221> DOMAIN

<222> (1)..(171)

<223> 重复片段

[0039]

<220>
<221> REPEAT
<222> (11).. (30)

<220>
<221> REPEAT
<222> (31).. (42)

<220>
<221> MUTAGEN
<222> (36).. (37)

<220>
<221> REPEAT
<222> (43).. (60)

<220>
<221> REPEAT
<222> (61).. (75)

<220>
<221> REPEAT
<222> (76).. (89)

<220>
<221> REPEAT
<222> (90).. (104)

<220>
<221> REPEAT
<222> (105).. (121)

<220>
<221> REPEAT
<222> (122).. (135)

<220>
<221> REPEAT
<222> (136).. (152)

<220>
<221> REPEAT
<222> (153).. (171)

[0040]

<220>
<221> DOMAIN
<222> (172)..(269)
<223> C-末端片段

<220>
<221> DOMAIN
<222> (172)..(269)
<223> C-末端片段

<400> 14

Gly Ser Ala Met Gly Tyr Leu Trp Ile Gln Gly Gln Gly Gly Tyr Gly
1 5 10 15

Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala
20 25 30

Ala Ala Ala Cys Cys Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln
35 40 45

Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly
65 70 75 80

Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
100 105 110

Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala
115 120 125

[0041]

Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly
130 135 140

Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
145 150 155 160

Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser
165 170 175

Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser
180 185 190

Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile
195 200 205

Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val
210 215 220

Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val
225 230 235 240

Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile
245 250 255

Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
260 265

<210> 15
<211> 269
<212> PRT
<213> 人工的

<220>
<223> 来源于 *Euprosthrops australis* MaSp1 的序列

[0042]

<220>
<221> DOMAIN
<222> (1).. (171)
<223> 重复片段

<220>
<221> REPEAT
<222> (11).. (30)

<220>
<221> REPEAT
<222> (31).. (42)

<220>
<221> REPEAT
<222> (43).. (60)

<220>
<221> REPEAT
<222> (61).. (75)

<220>
<221> REPEAT
<222> (76).. (89)

<220>
<221> REPEAT
<222> (90).. (104)

<220>
<221> REPEAT
<222> (105).. (121)

<220>
<221> REPEAT
<222> (122).. (135)

<220>
<221> MUTAGEN
<222> (128).. (129)

<220>

[0043]

<221> REPEAT

<222> (136)..(152)

<220>

<221> REPEAT

<222> (153)..(171)

<220>

<221> DOMAIN

<222> (172)..(269)

<223> C-末端片段

<400> 15

Gly Ser Ala Met Gly Tyr Leu Trp Ile Gln Gly Gln Gly Gly Tyr Gly
1 5 10 15

Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln
35 40 45

Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly
65 70 75 80

Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
100 105 110

Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Cys
115 120 125

[0044]

Cys Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly
130 135 140

Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
145 150 155 160

Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser
165 170 175

Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser
180 185 190

Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile
195 200 205

Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val
210 215 220

Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val
225 230 235 240

Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile
245 250 255

Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
260 265

<210> 16
<211> 269
<212> PRT
<213> 人工的

<220>

[0045]

<223> 来源于 *Euprosthénops australis* MaSp1 的序列

<220>

<221> DOMAIN

<222> (1)..(171)

<223> 重复片段

<220>

<221> REPEAT

<222> (11)..(30)

<220>

<221> REPEAT

<222> (31)..(42)

<220>

<221> REPEAT

<222> (43)..(60)

<220>

<221> REPEAT

<222> (61)..(75)

<220>

<221> REPEAT

<222> (76)..(89)

<220>

<221> REPEAT

<222> (90)..(104)

<220>

<221> REPEAT

<222> (105)..(121)

<220>

<221> REPEAT

<222> (122)..(135)

<220>

<221> REPEAT

<222> (136)..(152)

[0046]

<220>

<221> REPEAT

<222> (153)..(171)

<220>

<221> DOMAIN

<222> (172)..(269)

<223> C-末端片段

<220>

<221> MUTAGEN

<222> (222)..(222)

<400> 16

Gly Ser Ala Met Gly Tyr Leu Trp Ile Gln Gly Gln Gly Gly Tyr Gly
1 5 10 15

Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln
35 40 45

Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly
65 70 75 80

Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr
100 105 110

Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala

[0047]

115	120	125
Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly		
130	135	140
Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala		
145	150	155 160
Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser		
165	170	175
Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser		
180	185	190
Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile		
195	200	205
Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Ser Glu Val		
210	215	220
Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val		
225	230	235 240
Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile		
245	250	255
Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly		
260	265	

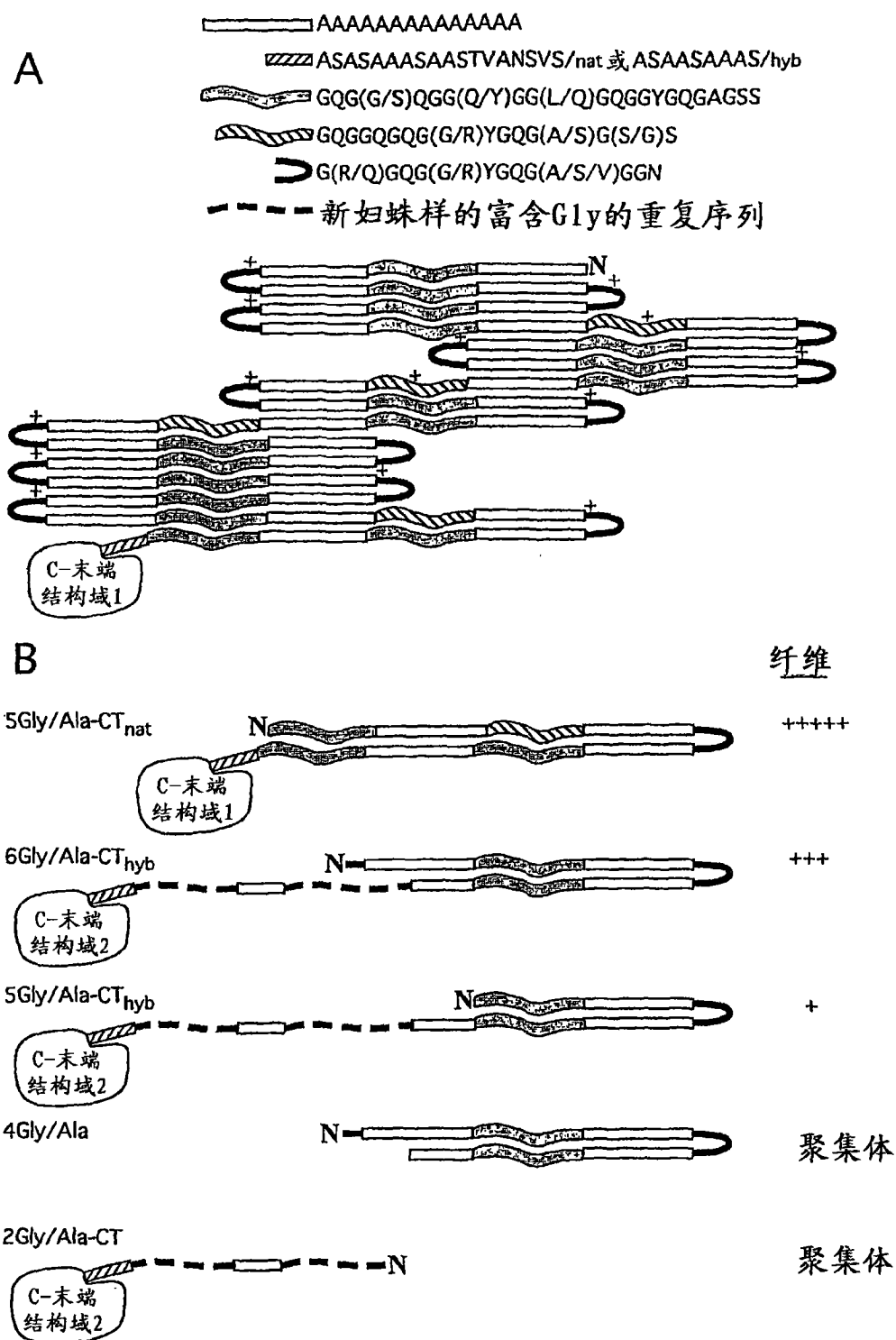


图 2

[illegible]

3
冬

共有序列

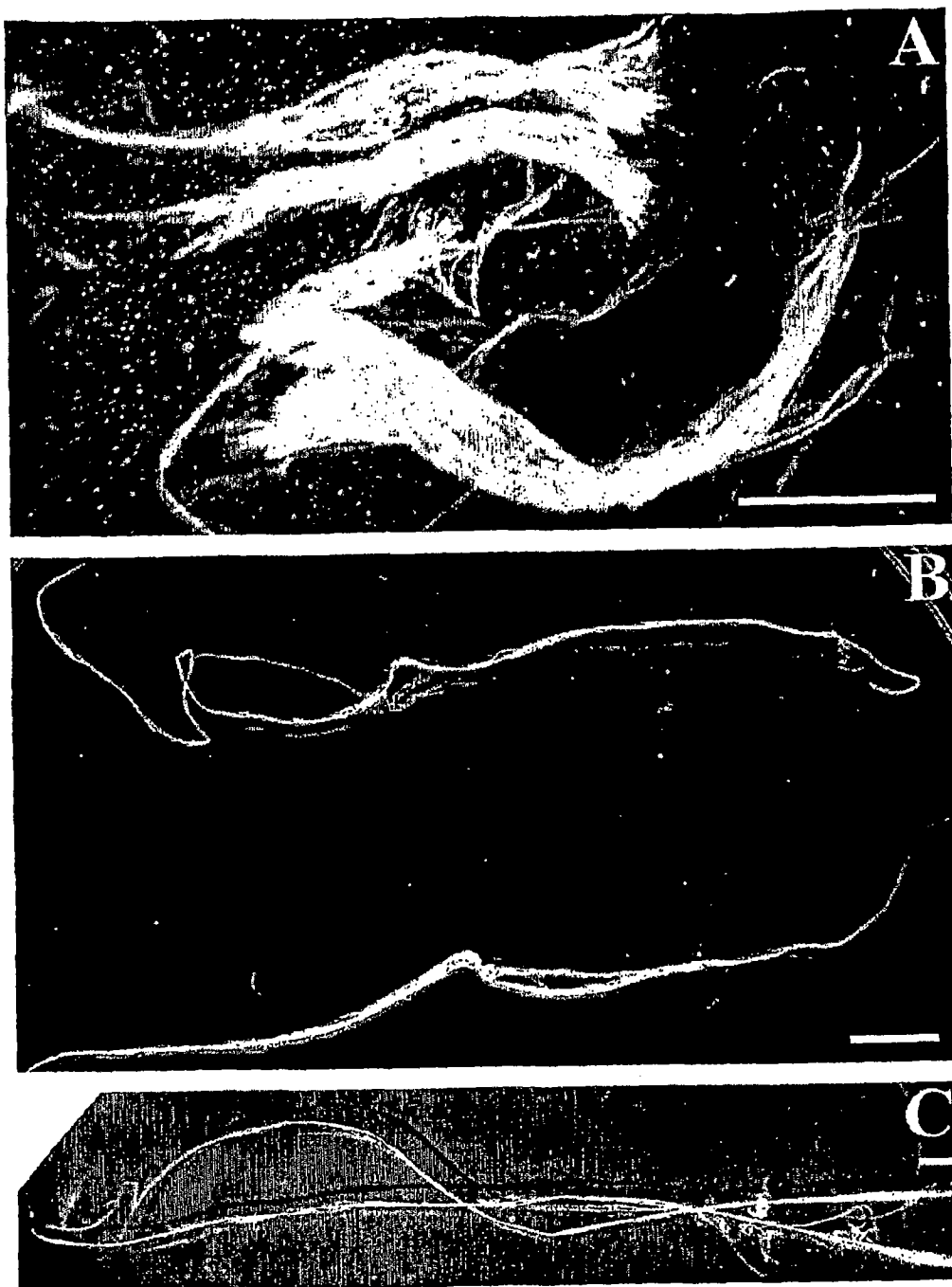


图 4

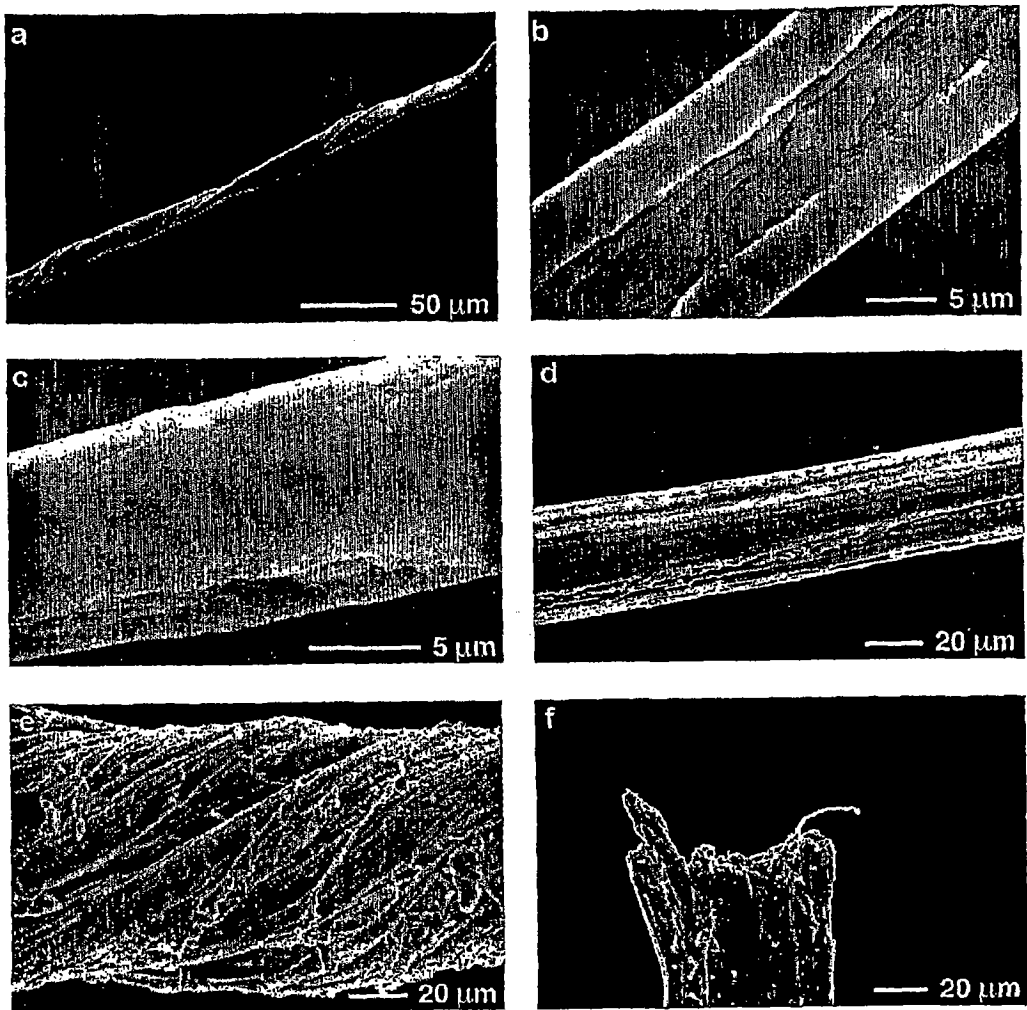


图 5

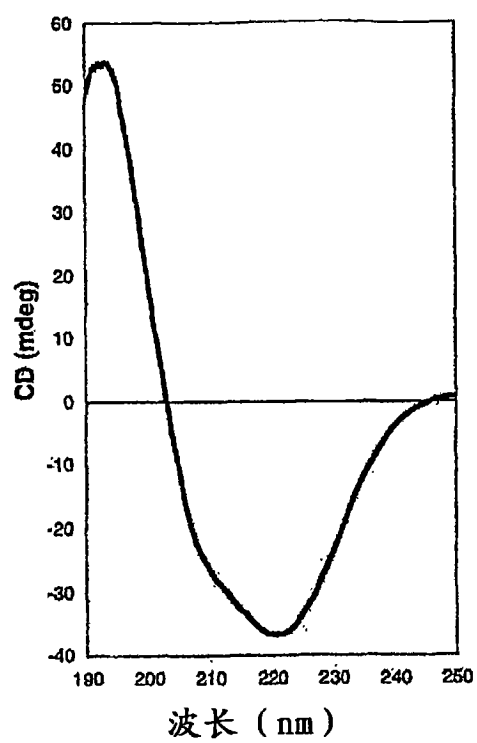


图 6

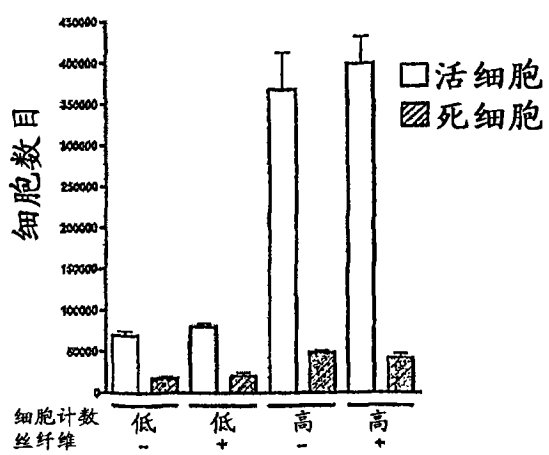


图 7



图 8

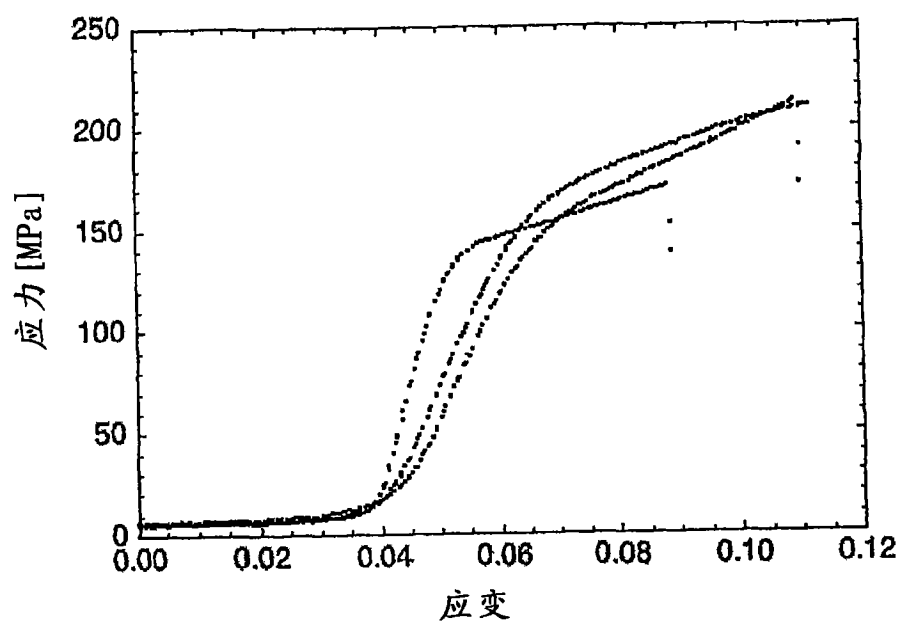


图 9

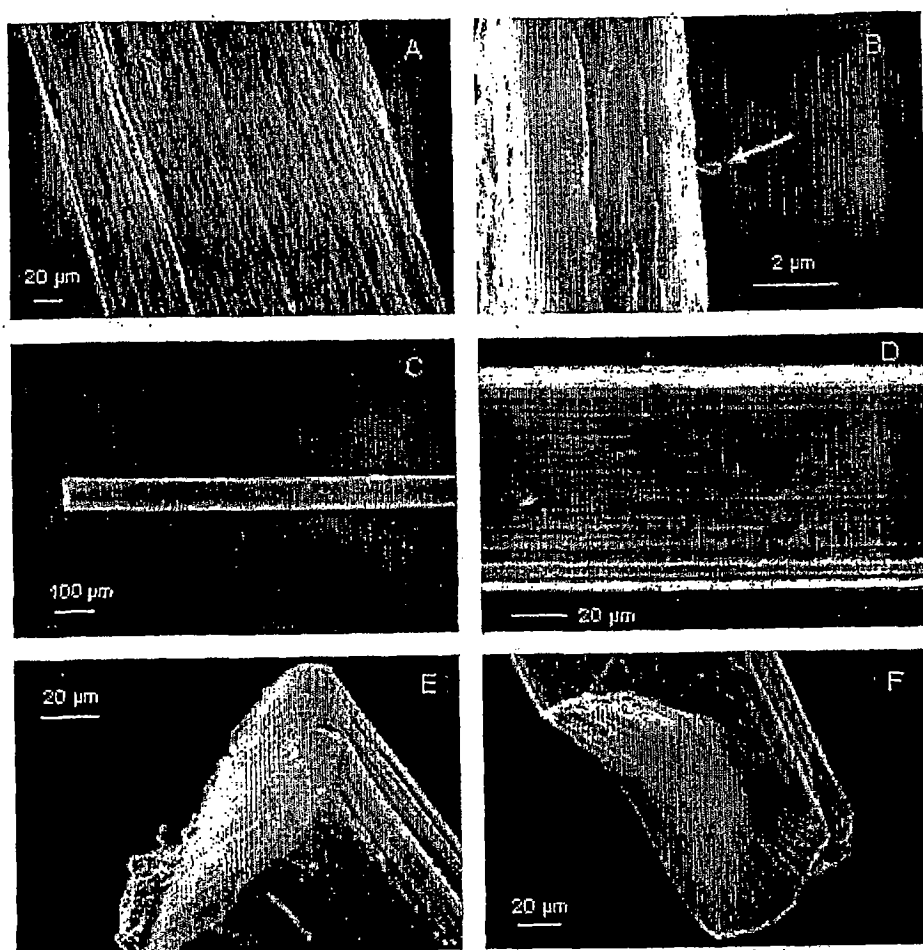


图 10