



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2008131298/10, 28.12.2006

(24) Дата начала отсчета срока действия патента:
28.12.2006

Приоритет(ы):

(30) Конвенционный приоритет:
30.12.2005 SE 0502932-7

(43) Дата публикации заявки: 10.02.2010 Бюл. № 4

(45) Опубликовано: 10.03.2012 Бюл. № 7

(56) Список документов, цитированных в отчете о поиске: WO 03020916 A2, 13.03.2003. LEWIS R.V. et al. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins, Protein Expr. Purif., 1996, v.7, n.4, p.400-406. POUCHKINA-STANTCHEVA N.N. et al. Molecular studies of a novel dragline silk from a nursery web spider, Euprosthenops sp. (Pisauridae), Comp. Biochem. (см. прод.)

(85) Дата начала рассмотрения заявки РСТ на национальной фазе: 30.07.2008

(86) Заявка РСТ:
SE 2006/001505 (28.12.2006)(87) Публикация заявки РСТ:
WO 2007/078239 (12.07.2007)

Адрес для переписки:

129090, Москва, ул. Б.Спасская, 25, стр.3,
ООО "Юридическая фирма Городисский и
Партнеры", пат.пов. Е.Е.Назиной

(72) Автор(ы):

**ЙОХАНССОН Ян (SE),
ХЬЯЛМ Йеран (SE),
СТАРК Маргарета (SE),
РИСИНГ Анна (SE),
ГРИП Стефан (SE),
ЭНГСТРЕМ Вильгельм (SE),
ХЕДХАММАР Мю (SE)**

(73) Патентообладатель(и):

СПИБЕР ТЕКНОЛОДЖИЗ АБ (SE)

(54) РЕКОМБИНАНТНЫЙ СПИДРОИНОВЫЙ БЕЛОК ГЛАВНОЙ АМПУЛЯРНОЙ ЖЕЛЕЗЫ, СПОСОБНЫЙ К ПОЛИМЕРИЗАЦИИ, РЕКОМБИНАНТНЫЙ СЛИТЫЙ БЕЛОК, СПОСОБНЫЙ К УПРАВЛЯЕМОЙ ПОЛИМЕРИЗАЦИИ, ПОЛИМЕР СПИДРОИНОВОГО БЕЛКА ГЛАВНОЙ АМПУЛЯРНОЙ ЖЕЛЕЗЫ, КОМПОЗИЦИЯ ДЛЯ УПРАВЛЯЕМОЙ ПОЛИМЕРИЗАЦИИ, РЕКОМБИНАНТНАЯ МОЛЕКУЛА НУКЛЕИНОВОЙ КИСЛОТЫ (ВАРИАНТЫ), СПОСОБ ПОЛУЧЕНИЯ РАСТВОРИМОГО СЛИТОГО БЕЛКА, СПОСОБ ПОЛУЧЕНИЯ ПОЛИМЕРА СПИДРОИНОВОГО БЕЛКА ГЛАВНОЙ АМПУЛЯРНОЙ ЖЕЛЕЗЫ И СПОСОБ КУЛЬТИВИРОВАНИЯ ЭУКАРИОТИЧЕСКИХ КЛЕТОК

(57) Реферат:

Изобретение относится к области

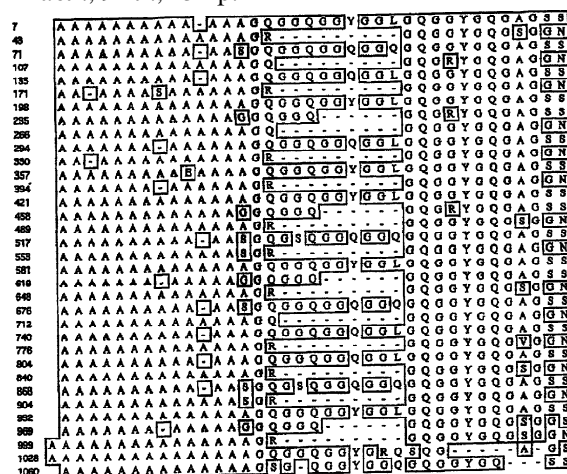
биотехнологии, конкретно к получению белков паука, и может быть использовано в

RU 2 4 4 4 5 2 7 C 2

RU 2 4 4 4 5 2 7 C 2

RU 2 4 4 4 5 2 7 C 2

RU 2444527 C2



ФИГ. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.

C07K 14/435 (2006.01)**C12N 15/62** (2006.01)**D01F 4/02** (2006.01)**(12) ABSTRACT OF INVENTION**(21)(22) Application: **2008131298/10, 28.12.2006**(24) Effective date for property rights:
28.12.2006

Priority:

(30) Priority:
30.12.2005 SE 0502932-7(43) Application published: **10.02.2010 Bull. 4**(45) Date of publication: **10.03.2012 Bull. 7**(85) Commencement of national phase: **30.07.2008**(86) PCT application:
SE 2006/001505 (28.12.2006)(87) PCT publication:
WO 2007/078239 (12.07.2007)

Mail address:

**129090, Moskva, ul. B.Spaskaja, 25, str.3, OOO
"Juridicheskaja firma Gorodisskij i Partnery",
pat.pov. E.E.Nazinoj**

(72) Inventor(s):

**JOKhANSSON Jan (SE),
Kh'JaLM Jeran (SE),
STARK Margareta (SE),
RISING Anna (SE),
GRIP Stefan (SE),
EhNGSTREM Vil'khel'm (SE),
KhEDKKhAMMAR Mju (SE)**

(73) Proprietor(s):

SPIBER TEKNOLODZhIZ AB (SE)

(54) POLYMERIZABLE RECOMBINANT SPIDROIN PROTEIN OF PRINCIPAL AMPULLARY GLAND, CONTROLLED POLYMERISABLE RECOMBINANT FUSED PROTEIN, SPIDROIN PROTEIN POLYMER OF PRINCIPAL AMPULLARY GLAND, CONTROLLED POLYMERISATION COMPOSITION, RECOMBINANT NUCLEIC ACID MOLECULE (VERSIONS), METHOD FOR PRODUCING SOLUBLE FUSED PROTEIN, METHOD FOR PRODUCING SPIDROIN PROTEIN POLYMER OF PRINCIPAL AMPULLARY GLAND AND METHOD FOR EUKARIOTIC CELL CULTURE

(57) Abstract:

FIELD: medicine.

SUBSTANCE: recombinant method is used to produce a spidroin protein of a principal ampullary gland which consists of 150-420 amino acid residues and is described by formula REP-CT where REP is a protein fragment consisting of 80-300 amino acid residues, selected from L(AG)nL, L(AG)nAL, L(GA)nL or L(GA)nGL where n is an integer from 4 to 8. Each individual segment A consists of 8-18 amino acid residues, where 0 to 3 amino acid residues are not Ala, while the rest are Ala. Each individual

segment G consists of 12-30 amino acid residues, where at least 40% of amino acid residues are Gly. Each individual segment L is a linker amino acid sequence of 0-20 amino acid residues. CN is a fragment consisting of 70 to 120 amino acid residues of a C-terminal spidroin protein fragment of the principal ampullary gland.

EFFECT: invention presents the polymerisable spidroin protein of the principal ampullary gland which is used as a basis to produce the controlled polymerisable fused protein.

15 cl, 7 dwg, 13 ex

Фиг. 1

RU 2 4 4 4 5 2 7 C 2

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

Настоящее изобретение относится к области рекомбинантного получения белков. Более подробно, настоящее изобретение относится к рекомбинантному получению белков шелка паука. В настоящем изобретении предоставляются новые выделенные спидроиновые белки главной ампулярной железы и слитые белки главной ампулярной железы, а также способы получения таких белков и молекулы полинуклеиновых кислот для их получения. Кроме того, предоставляются полимеры спидроиновых белков главной ампулярной железы и способы получения таких полимеров.

УРОВЕНЬ ТЕХНИКИ

Паучьи шелка представляют собой природные полимеры с высокими характеристиками, обладающие исключительной крепкостью благодаря комбинации прочности и эластичности. Пауки имеют до семи специализированных желез, которые вырабатывают множество видов шелковой нити с различными механическими свойствами и функциями. Канатный шелк, вырабатываемый главной ампулярной железой, представляет собой очень прочное волокно и в весовом соотношении оно превосходит созданные человеком материалы, такие как высокопрочная сталь и кевлар. Свойства канатного шелка являются привлекательными для создания новых материалов для медицинских или технических целей.

Канатный шелк состоит из двух основных полипептидов, обычно называемых спидроинами главной ампулярной железы (MaSp) 1 и 2, но полипептиды из *Araneus diadematus* называются ADF-3 и ADF-4. Указанные белки имеют кажущуюся молекулярную массу в диапазоне 200-720 кДа, в зависимости от возраста образца и условий анализа, но не было публикаций о полноразмерном гене канатного шелка паука. Свойства полипептидов канатного шелка описаны в Huemmerich, D. et al. Novel assembly properties of recombinant spider dragline silk proteins. Curr. Biol. 14, 2070-2074 (2004). Известные спидроины канатного шелка состоят из многократно повторяющихся блоков чередующихся сегментов, богатых аланином, которые образуют кристаллические β -листы в волокне, и сегментов, богатых глицином, которые являются более гибкими и в основном лишены упорядоченной структуры. С-концевой участок не имеет повторов, является высоко консервативным в разных видах и принимает конформацию α -спирали. N-концевой участок белков канатного шелка был охарактеризован совсем недавно, причем показано, что N-концевой участок является высоко консервативным в различных спидроинах и также у различных видов пауков (Rising, A. et al. N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. Biomacromolecules 7, 3120-3124 (2006)).

Механические свойства канатного шелка отличаются у различных видов пауков; канатный шелк *Euprosthénops sp.* является более жестким, прочным (требует большего усилия для разрыва) и менее гибким, чем канатный шелк, например *Araneus diadematus* или *Nephila clavipes*. Канатный шелк *Euprosthénops sp.*, по-видимому, содержит большую количественную долю кристаллической β -листовой структуры, чем канатный шелк *Araneus diadematus*, скорее всего по этой причине MaSp из *Euprosthénops sp.* имеет самое высокое содержание полиаланина среди всех исследованных до настоящего времени видов (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. 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)).

В опытах по получению искусственных волокон паутины использовались фрагменты натуральных или синтетических генов, кодирующих белки канатного

шелка, потому что полноразмерный ген еще не описан. Рекомбинантные белки канатного шелка экспрессированы в различных системах, включая бактерии, дрожжи, клетки млекопитающих, растения, клетки насекомых, трансгенного тутового шелкопряда и трансгенных коз. Для примера, см. Lewis, R. V. et al. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. *Protein Expr. Purif.* 7, 400-406 (1996); Fahnestock, S. R. & Irwin, S. L. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 47, 23-32 (1997); Arcidiacono, S. et al. Purification and characterization of recombinant spider silk 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); and Lazaris, A. et al. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-476 (2002).

В международном патенте WO 2004/016651 (Университет Йорка) раскрыты последовательности нуклеиновых кислот, кодирующих внутренние повторяющиеся части белков MaSp1 из *Euprosthenops sp.* Белок не экспрессирован.

В публикации Huemmerich, D. et al. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004) раскрывается искусственный ген, "(AQ)₁₂NR3", кодирующий повторяющиеся фрагменты, богатые Ala, и фрагменты, богатые Gly/Gln, и неповторяющийся фрагмент, которые получены из ADF3 *Araneus*. Ген экспрессирован с образованием растворимого белка (59,8 кДа, >528 aa), который агрегирует, но не образует полимеров или волокон. Содержание аланина в белке составляет 10-15%.

В международном патенте WO 03/057727 раскрыта экспрессия растворимых рекомбинантных полипептидов шелка в клеточных линиях млекопитающих и животных. Один экспрессированный полипептид шелка (ADF-3; 60 кДа, 652 aa) состоит из повторяющейся единицы и неповторяющегося гидрофильного домена. Другой экспрессированный полипептид шелка (ADF-3 His; 63 кДа, 677 aa) состоит из повторяющейся единицы, неповторяющегося гидрофильного домена, эпитопа с-тус и тага из шести гистидинов. Повторяющаяся единица имеет низкое содержание Ala (10-20%). Полученные полипептиды шелка имеют низкую растворимость в водных средах и/или образуют преципитаты. Поскольку полученные полипептиды шелка не полимеризуются самопроизвольно, то для получения полимеров или волокон необходимо формирование.

Несколько факторов затрудняют экспрессию белков канатного шелка. Из-за многократной повторяемости генов, которая сопровождается ограниченным аминокислотным составом белков, происходят ошибки транскрипции и трансляции. Другой причиной может быть истощение пулов тРНК в микробных системах экспрессии с последующей разобщенной трансляцией, приводящей к преждевременной терминации синтеза белка. В качестве других причин прерывания синтеза белка обсуждаются строение вторичной структуры мРНК и рекомбинация генов. Показано, что нативные гены MaSp длиной больше чем 2,5 т.п.н. являются нестабильными в бактериальных хозяевах. Кроме того, существуют сложности в поддержании рекомбинантных белков шелка в растворимой форме, поскольку как фрагменты канатного шелка природного происхождения, так и созданные блочные сополимеры, особенно белки, полученные из MaSp/ADF-4, легко образуют аморфные агрегаты, вызывающие осаждение и потерю белка. См. Huemmerich, D. et al. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004) and Lazaris, A. et al. Spider silk fibers spun from soluble recombinant silk

produced in mammalian cells. Science 295, 472-476 (2002).

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Целью настоящего изобретения является предоставление нового белка шелка паука, из которого могут быть получены волокна шелка паука.

Другой целью настоящего изобретения является предоставление водорастворимого белка шелка паука, который при простом воздействии самополимеризуется в волокна. Это дает возможность для уникальных применений, таких как культивирование эукариотических клеток на волокнах. Кроме того, данное свойство делает возможным проведение всех последующих стадий при физиологических условиях, что снижает риск токсичности и денатурации белка.

Целью настоящего изобретения также является предоставление волокон нового белка шелка паука.

Одной целью настоящего изобретения является предоставление белков шелка паука в большом количестве, которые после простой обработки могут легко самополимеризоваться с образованием волокна.

Также целью изобретения является предоставление способов получения белков шелка и волокон белков шелка паука.

Для данных и других целей, которые становятся очевидными из последующего раскрытия, в настоящем изобретении по одному аспекту предоставляется выделенный спидроиновый белок главной ампулярной железы, где белок состоит из 150-420 аминокислотных остатков и описывается формулой **REP-CT**, где **REP** представляет собой фрагмент белка, имеющий 80-300 аминокислотных остатков, где указанный фрагмент выбран из группы **L(AG)_nL**, **L(AG)_nAL**, **L(GA)_nL**, **L(GA)_nGL**, где **n** представляет собой целое число от 4 до 8; каждый индивидуальный сегмент **A** представляет собой аминокислотную последовательность из 8-18 аминокислотных остатков, где от 0 до 3 аминокислотных остатков не являются Ala, и остальные аминокислотные остатки представляют собой Ala; каждый индивидуальный сегмент **G** является аминокислотной последовательностью из 12-30 аминокислотных остатков, где по меньшей мере 40% аминокислотных остатков являются Gly; и каждый индивидуальный сегмент **L** является линкерной аминокислотной последовательностью из 0-20 аминокислотных остатков; и **CT** представляет собой фрагмент белка, имеющий от 70 до 120 аминокислотных остатков, фрагмент которого является С-концевым фрагментом, полученным из спидроинового белка главной ампулярной железы или из его производных.

Настоящее изобретение основано на идентификации белкового мотива, которого достаточно для формирования волокон, подобных шелку, и на применении указанного мотива для построения рекомбинантных белков MaSp, которые возможно вырабатывать в подходящих хозяевах, таких как бактерии, предпочтительно *E. coli*.

В определенных воплощениях согласно изобретению каждый индивидуальный сегмент **A** имеет, по меньшей мере, 80% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 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 и 1060-1073 SEQ ID NO: 3; аминокислотных остатков 31-42, 61-75, 90-104, 122-135 и 153-171 SEQ ID NO: 9; аминокислотных остатков 12-25, 46-60, 75-88, 112-119, 150-158 и 173-180 SEQ ID NO: 13; аминокислотных остатков 31-42 SEQ ID NO: 14; и аминокислотных остатков 122-135 SEQ ID NO: 15. В специфических воплощениях каждый

индивидуальный сегмент **A** является аминокислотной последовательностью, выбранной из данной группы аминокислотных последовательностей.

В некоторых воплощениях согласно изобретению каждый индивидуальный сегмент **G** имеет, по меньшей мере, 80% идентичности с аминокислотной

последовательностью, выбранной из группы аминокислотных остатков 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 и 1074-1092 SEQ ID NO: 3; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; аминокислотных остатков 11-30, 43-60, 76-89, 105-121 и 136-152 SEQ ID NO: 9; и аминокислотных остатков 1-11, 26-45, 61-74, 89-111, 120-149 и 159-172 SEQ ID NO: 13. В специфических воплощениях каждый индивидуальный сегмент **G** является идентичным аминокислотной последовательности, выбранной из данной группы аминокислотных последовательностей.

В определенных воплощениях согласно изобретению указанный фрагмент **СТ** имеет по меньшей мере 50% идентичности с SEQ ID NO: 8 или по меньшей мере 80% идентичности с аминокислотной последовательностью, выбранной из группы, состоящей из SEQ ID NO: 4, аминокислотных остатков 172-269 SEQ ID NO: 9, аминокислотных остатков 181-276 SEQ ID NO: 13, и аминокислотных остатков 172-269 SEQ ID NO: 16, а также с любой аминокислотной последовательностью фигуры 3, особенно с последовательностями MaSp1 фигуры 3. В определенных воплощениях указанный фрагмент **СТ** представляет собой аминокислотную последовательность, выбранную из данной группы аминокислотных последовательностей.

В определенных воплощениях согласно настоящему изобретению содержание липополисахаридов (LPS) и других пирогенов в выделенном спидроиновом белке главной ампулярной железы составляет 1 единицу эндотоксина (ЭЕ)/мг белка или меньше.

Согласно другому аспекту в настоящем изобретении предоставляется выделенный слитый белок, состоящий из первого белкового фрагмента, который представляет собой спидроиновый белок главной ампулярной железы и второго белкового фрагмента, отличающегося тем, что указанный второй белковый фрагмент содержит партнер по слитию и участок узнавания расщепляющего агента, где указанный первый белковый фрагмент соединяется через указанный участок узнавания расщепляющего агента с указанным партнером по слитию.

В настоящем изобретении предоставляется выделенный слитый белок, выбранный из группы **X-REP-СТ**, и **REP-СТ-X**, где **REP** и **СТ** являются белковыми фрагментами согласно изобретению; и **X** является белковым фрагментом, содержащим партнер по слитию и участок узнавания расщепляющего агента; где комбинированный белковый фрагмент **REP-СТ** соединяется через указанный участок узнавания расщепляющего агента с указанным партнером по слитию.

В определенных воплощениях согласно изобретению содержание LPS и других пирогенов в выделенном спидроиновом белке составляет 1 ЭЕ/мг белка или меньше.

Согласно еще одному аспекту в настоящем изобретении предоставляется способ получения спидроинового белка главной ампулярной железы согласно изобретению, включающий в себя стадии: (i) предоставления раствора слитого белка согласно изобретению в жидкой среде, (ii) добавления к указанной жидкой среде подходящего расщепляющего агента для расщепления слитого белка в участке узнавания расщепляющего агента и получения благодаря этому спидроинового белка главной

ампулярной железы; и, если надо, (iii) выделения спидроинового белка главной ампулярной железы, полученного на стадии (ii) из указанной жидкой среды.

Настоящее изобретение также предоставляет способ получения полимера спидроинового белка главной ампулярной железы согласно изобретению, включающий в себя стадии: (i) предоставления раствора слитого белка согласно изобретению в жидкой среде, (ii) добавления к указанной жидкой среде подходящего расщепляющего агента для расщепления слитого белка в участке узнавания расщепляющего агента, и получения благодаря этому спидроинового белка главной ампулярной железы; (iii) вызывания полимеризации спидроинового белка главной ампулярной железы, полученного на стадии (ii) в жидкой среде; и, если надо, (iv) выделения полимера, полученного на стадии (iii) из указанной жидкой среды. В предпочтительном способе указанная стадия (iii) далее включает в себя создание границы раздела между указанной жидкой средой и другой фазой, выбранной из группы, состоящей из газовой фазы, жидкой фазы и твердой фазы, где указанная полимеризация вызывается в указанной границе раздела или в области, окружающей границу раздела. В предпочтительном способе указанная жидкая среда является водной средой и указанную другую фазу выбирают из группы, состоящей из органических растворителей, несмешиваемых с воздухом и водой.

Согласно другому аспекту в настоящем изобретении предоставляется выделенная молекула полинуклеиновой кислоты, включающая в себя последовательность нуклеиновой кислоты, которая кодирует спидроиновый белок главной ампулярной железы согласно изобретению или комплементарную ей последовательность нуклеиновой кислоты.

Согласно еще одному аспекту в настоящем изобретении предоставляется выделенная молекула полинуклеиновой кислоты, включающая в себя последовательность нуклеиновой кислоты, которая кодирует слитый белок согласно изобретению или комплементарную ей последовательность нуклеиновой кислоты.

Другой аспект изобретения относится к способу получения растворимого слитого белка согласно изобретению, который включает в себя стадии: (i) экспрессии молекулы полинуклеиновой кислоты, кодирующей растворимый слитый белок согласно изобретению, в подходящем хозяине; и (ii) выделения растворимого слитого белка, полученного на стадии (i). Если надо, указанная стадия (ii) выделения растворимого слитого белка включает в себя удаление LPS и других пирогенов.

В настоящем изобретении также предоставляется способ производства спидроинового белка главной ампулярной железы согласно изобретению, включающий в себя стадии: (i) экспрессии молекулы полинуклеиновой кислоты, кодирующей растворимый слитый белок согласно изобретению, в подходящем хозяине; и (ii) выделения растворимого слитого белка, полученного на стадии (i); (iii) предоставления раствора указанного растворимого слитого белка, полученного на стадии (ii) в жидкой среде, (iv) добавления к указанной жидкой среде подходящего расщепляющего агента для расщепления слитого белка в сайте узнавания расщепляющего агента, и получения благодаря этому спидроинового белка главной ампулярной железы; и, если надо, (v) выделения спидроинового белка главной ампулярной железы, полученного на стадии (iv) из указанной жидкой среды. Далее по желанию указанная стадия (ii) выделения растворимого слитого белка и по желанию указанная стадия (v) выделения спидроинового белка главной ампулярной железы включают в себя удаление LPS и других пирогенов.

Далее в настоящем изобретении предоставляется способ получения полимера

спидроинового белка главной ампулярной железы согласно изобретению, включающий в себя стадии: (i) экспрессии молекулы полинуклеиновой кислоты, кодирующей растворимый слитый белок согласно изобретению, в подходящем хозяине; и (ii) выделения растворимого слитого белка, полученного на стадии (i); (iii) предоставления раствора указанного растворимого слитого белка, полученного на стадии (ii) в жидкой среде, (iv) добавления к указанной жидкой среде подходящего расщепляющего агента для расщепления слитого белка в сайте узнавания расщепляющего агента, и получения благодаря этому спидроинового белка главной ампулярной железы; (v) вызывание полимеризации спидроинового белка главной ампулярной железы, полученного на стадии (iv) в жидкой среде; и, если надо, (vi) выделение полимера, полученного на стадии (v) из указанной жидкой среды. В предпочтительном способе указанная стадия (v) далее включает в себя создание границы раздела между указанной жидкой средой и другой фазой, выбранной из группы, состоящей из газовой фазы, жидкой фазы и твердой фазы, где указанная полимеризация вызывается в указанной границе раздела или в области, окружающей границу раздела. В предпочтительном способе указанная жидкая среда является водной средой, и указанная другая фаза выбрана из группы, состоящей из органических растворителей, несмешиваемых с воздухом и с водой.

Согласно другому аспекту в настоящем изобретении предоставляется полимер спидроинового белка главной ампулярной железы согласно изобретению. В настоящем изобретении также предоставляется полимер спидроинового белка главной ампулярной железы, полученный по способу согласно изобретению. В предпочтительном воплощении указанный полимер является волокном. В других предпочтительных воплощениях указанный полимер образует структуру, выбранную из группы, состоящей из пены, геля, сетки или пленки.

Согласно еще одному аспекту в настоящем изобретении предоставляется новое применение белкового фрагмента, включающего в себя партнера по слитию и участок узнавания расщепляющего агента для производства слитого белка, включающего в себя указанный белковый фрагмент, соединенный через указанный участок узнавания расщепляющего агента с фрагментом белка шелка паука. В предпочтительных воплощениях указанный фрагмент белка шелка паука состоит из 150-420 аминокислотных остатков.

Согласно последнему аспекту в настоящем изобретении предоставляется выделенная молекула полинуклеиновой кислоты, включающая в себя последовательность нуклеиновой кислоты, выбранную из группы, состоящей из SEQ ID NO: 1 и последовательностей нуклеиновых кислот, кодирующих SEQ ID NO: 2-16, или комплементарных им последовательностей нуклеиновых кислот. В настоящем изобретении предоставляется применение выделенной молекулы полинуклеиновой кислоты для производства искусственного гена, кодирующего белок шелка паука.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

На фигуре 1 показано выравнивание сегментов внутри повторяющейся части белка MaSp1 *Euprosthenops australis*, т.е. SEQ ID NO: 3.

На фигуре 2А проиллюстрирована в виде схемы предсказанная структурная организация повторяющейся части белка MaSp1 *Euprosthenops australis* (SEQ ID NO: 3).

На фигуре 2В проиллюстрированы в виде схем предсказанные структурные организации спидроиновых белков, построенных согласно примерам 5-8 (SEQ ID NO: 9-13).

На фигуре 3 показано выравнивание С-концевых участков MaSp1 и MaSp2, которое

иллюстрирует их консервативную природу.

На фигуре 4 проиллюстрирован макроскопический внешний вид волокон, сформированных из спидроиновых белков, которые построены согласно примерам 5-8. (A): волокна белка 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13), полоска 0,5 см; (B): волокна белка 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), полоска 1 см; (C): волокна белка 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), полоска 1 см.

На фигуре 5 представлены полученные с помощью сканирующего электронного микроскопа (SEM) микрофотографии волокон, сформированных из спидроиновых белков, которые построены согласно примерам 5-8. Одиночные волокна (a) и гелевая фаза (b, c) из 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13). Волокна 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), вытянутые в 75% метаноле, высушенные на воздухе и нанесенные на целики SEM (d, e, f). Волокно, скрученное перед высушиванием на воздухе (e); конец волокна (f).

На фигуре 6 показан спектр кругового дихроизма (CD) волокна 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13).

На фигуре 7 проиллюстрированы результаты цитотоксического исследования на тучных клетках мышей, которые показывают число живых и мертвых клеток через три дня культивирования в присутствии и в отсутствии шелковых волокон, полученных *in vitro*.

На фигуре 9 представлена кривая зависимости между напряжением и деформацией, которая показывает прочность при растяжении дважды вытянутых волокон из 5Gly/Ala-CT_{nat} (SEQ ID NO: 9).

На фигуре 10 представлены SEM-микрофотографии рекомбинантных волокон из 5Gly/Ala-CT_{nat} (SEQ ID NO: 9). (a, b) Спонтанно образованные волокна. На изображении (b) крупным планом показана фибриллярная субструктура. Тонкая выступающая фибрилла (отмечена стрелкой) имеет ширину около 300 нм. (c-f) Волокна после двух циклов растяжения-расслабления. На (c) и (d) показано одно и то же волокно в различных увеличениях. На (e) показан срезанный конец волокна, и на (f) показана точка разрыва после теста на растяжение.

ПОДРОБНОЕ РАСКРЫТИЕ ИЗОБРЕТЕНИЯ

Настоящее изобретение в целом основано на идентификации мотива спидроинового белка, который является достаточным для рекомбинантного получения волокон шелка паука. Мотив основан на расшифрованной аминокислотной последовательности в результате клонирования и секвенирования частичной кДНК спидроина 1 (MaSp1) главной ампулярной железы *Euprosthenops australis*. Из этого следует, что выделенная кДНК MaSp1 является полезной в качестве точки старта для построения новых генов спидроинов, таких как гены, представленные в настоящем описании. Полимеры, которые созданы из белков, полученных из новых спидроиновых кДНК, являются полезными благодаря своим физическим свойствам, и особенно полезной является комбинация высокой прочности, эластичности и малого веса. Они также являются полезными благодаря своей способности поддерживать клеточную адгезию и рост. Свойства канатного шелка являются привлекательными для разработки новых материалов для медицинских и технических целей. В частности, паучьи шелка согласно изобретению могут быть полезны в медицинских устройствах, таких как импланты, и в медицинских продуктах, таких как системы для ушивания ран, пластыри, шовный материал, повязки на рану, и в каркасных структурах для инженеринга тканей и для направленной клеточной регенерации. Паучьи шелка согласно изобретению также являются особенно полезными для применения в качестве ткани или материала, такого как материал для парашютов, для

пуленепробиваемой одежды, для ремней безопасности и т.п.

Термин "волокно", использованный в настоящем описании, относится к полимерам, имеющим толщину, по меньшей мере, 1 мкм, предпочтительными являются макроскопические полимеры, которые видимы человеческим глазом, например, 5 имеющие толщину, по меньшей мере, 1 мкм, и которые имеют значительное растяжение в длину, по сравнению с их толщиной, предпочтительно, приблизительно 5 мм. Термин "волокно" не включает в себя неструктурированные агрегаты или преципитаты.

10 Термины "спидроиновые белки главной ампулярной железы", "спидроиновые белки" используются взаимозаменяемо в описании и включают в себя все известные спидроиновые белки главной ампулярной железы, обычно обозначаемые аббревиатурой "MaSp" или "ADF" в случае *Araneus diadematus*. Данные спидроиновые белки главной ампулярной железы обычно бывают двух типов, 1 и 2. Указанные 15 термины, кроме того, включают в себя новые белки согласно изобретению, как описано в приложенной формуле изобретения, и другие искусственные белки с высокой степенью идентичности и/или сходства с известными спидроиновыми белками главной ампулярной железы.

20 Авторы настоящего изобретения использовали идентифицированный мотив спидроинового белка для создания новых генных конструкций, кодирующих искусственные спидроиновые белки. Было обнаружено, что спидроиновый белок главной ампулярной железы, состоящий из 150-420 аминокислотных остатков, т.е. больше или равно 150, предпочтительно больше или равно 220, предпочтительно 25 больше или равно 250, и меньше или равно 420, предпочтительно меньше или равно 380 аминокислотных остатков, предпочтительно меньше или равно 320 аминокислотных остатков, предпочтительно меньше или равно 280 аминокислотных остатков, а именно 220-360 аминокислотных остатков, может быть получен 30 рекомбинантно, например, в бактериях или в других подходящих продукционных организмах. Полученные спидроиновые белки спонтанно образуют макроскопические шелковые волокна согласно изобретению. Это удивительный результат, природные спидроиновые белки и ранее известные рекомбинантные, образующие волокна спидроиновые белки являются значительно более длинными, чем белки согласно 35 изобретению. Кроме того, природные спидроиновые белки и ранее известные рекомбинантные, образующие волокна спидроиновые белки, как правило, содержат большое число внутренних повторов и требуют центрифугирования и/или жестких растворителей для полимеризации.

40 В настоящем описании впервые показано, что спидроиновые белки могут спонтанно образовывать волокна *in vitro*. Данные, представленные в настоящем описании, также показывают, что для запуска образования волокон необходимо только присутствие фракции последовательности спидроина. Кроме того, видовой гибрид, содержащий повторяющийся домен *Euprosthénops* и неповторяющийся С-концевой домен *Nephila* (перенесенный в пример 6С) также образует волокна, 45 проявляя сильный потенциал к образованию волокон у данного мотива.

В общем виде спидроиновый белок главной ампулярной железы согласно изобретению описывается формулой **REP-CT**. Белковый фрагмент **REP** и белковый 50 фрагмент **CT** соединены ковалентно, обычно с помощью пептидной связи.

Белковый фрагмент **REP** имеет повторяющийся характер, с чередованием участков, богатых аланином и богатых глицином. Фрагмент **REP** обычно содержит более чем 80, а именно более чем 140 и менее чем 300, предпочтительно менее чем 240, а

именно менее чем 200, аминокислотных остатков, и может быть разделен на несколько сегментов **L** (линкерных), сегментов **A** (богатых аланином) и сегментов **G** (богатых глицином), как будет подробно объяснено далее. Обычно указанные линкерные сегменты, которые являются произвольными, расположены на концах фрагмента **REP**, в то время как остальные сегменты представляют собой чередующиеся богатые аланином и богатые глицином сегменты. Таким образом, фрагмент **REP** может иметь в целом любую из следующих структур, где n является целым числом:

$L(AG)_nL$, а именно $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5L$;

$L(AG)_nAL$, а именно $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5A_6L$;

$L(GA)_nL$, а именно $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5L$; или

$L(GA)_nGL$, а именно $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5G_6L$.

Из этого следует, что не имеет особой важности, какой из двух сегментов, богатый аланином или богатый глицином, граничит с N-концевыми или C-концевыми линкерными сегментами. Предпочтительно n является целым числом от 4 до 8, более предпочтительно от 4 до 6, например $n=4$, $n=5$ или $n=6$.

В предпочтительных воплощениях содержание аланина в фрагменте **REP** согласно изобретению составляет выше 20%, предпочтительно выше 25%, наиболее предпочтительно выше 30%, и ниже 50%, предпочтительно ниже 40%, более предпочтительно ниже 35%. Это является преимуществом, поскольку предполагается, что более высокое содержание аланина дает более жесткое, и/или более крепкое, и/или менее растяжимое волокно. Причиной этого явления по-видимому является то, что более высокое содержание аланина ассоциировано с более высоким содержанием β -листных структур в волокне. Таким образом, в предпочтительном воплощении, содержание β -листных структур в полимере, таком как волокно, спидроинового белка главной ампулярной железы согласно изобретению составляет более 50%, т.е. более чем 50% вторичной структуры белка имеет β -листовую форму.

В определенных воплощениях фрагмент **REP** лишен остатков пролина, т.е. во фрагменте **REP** нет остатков Pro.

Далее, обращаясь к сегментам, из которых построен фрагмент **REP** согласно изобретению, следует подчеркнуть, что каждый сегмент является индивидуальным, т.е. любые два сегмента **A**, любые два сегмента **G** или любые два сегмента **L** специфического фрагмента **REP** могут быть идентичными или могут не быть идентичными. Таким образом, общей особенностью изобретения является то, что каждый тип сегмента не является одинаковым внутри специфического фрагмента **REP**. Скорее, последующее раскрытие обеспечивает специалиста в данной области техники рекомендациями, как создать индивидуальные сегменты и собрать их во фрагмент **REP**, который является частью функционального спидроинового белка согласно изобретению.

Из представленных в настоящем описании экспериментальных данных было сделано заключение, что каждый индивидуальный сегмент **A** представляет собой аминокислотную последовательность, имеющую от 8 до 18 аминокислотных остатков. Предпочтительно, каждый индивидуальный сегмент **A** содержит от 13 до 15 аминокислотных остатков. Также возможно, что большинство или более чем два сегмента **A** содержат от 13 до 15 аминокислотных остатков, и что меньшинство, а именно один или два сегмента **A**, содержат от 8 до 18 аминокислотных остатков, а именно 8-12 или 16-18 аминокислотных остатков. Значительное большинство аминокислотных остатков являются остатками аланина. Более подробно, от 0 до 3

аминокислотных остатков являются неаланиновыми остатками и остальные аминокислотные остатки являются остатками аланина. Таким образом, все аминокислотные остатки в каждом индивидуальном сегменте **A** являются остатками аланина, без исключения или за исключением одного, двух или трех аминокислотных остатков, которые могут быть любой аминокислотой. Предпочтительно, аланин-замещающая аминокислота (аминокислоты) является (являются) природными аминокислотами, предпочтительно индивидуально выбранными из группы из серина, глутаминовой кислоты, цистеина и глицина, более предпочтителен серин. Конечно, возможно, что один или более сегментов **A** являются полностью аланиновыми сегментами, в то время как остальные сегменты **A** содержат 1-3 неаланиновых остатка, таких как серин, глутаминовая кислота, цистеин или глицин.

В предпочтительном воплощении каждый сегмент **A** содержит 13-15 аминокислотных остатков, включая 10-15 остатков аланина и 0-3 неаланиновых остатка, как описано выше. В более предпочтительном воплощении каждый сегмент **A** содержит 13-15 аминокислотных остатков, включая 12-15 остатков аланина и 0-1 неаланиновых остатков, как описано выше.

Предпочтительно, каждый индивидуальный сегмент **A** имеет, по меньшей мере, 80% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 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 и 1060-1073 SEQ ID NO: 3. Каждая последовательность данной группы соответствует сегменту природной последовательности белка MaSp1 *Euprosthenoops australis*, которая расшифрована с помощью клонирования соответствующей кДНК, см. примеры 1-2 и фигуры 1-2A. Альтернативно каждый индивидуальный сегмент **A** имеет, по меньшей мере, 80% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 31-42, 61-75, 90-104, 122-135 и 153-171 SEQ ID NO: 9, аминокислотных остатков 12-25, 46-60, 75-88, 112-119, 150-158 и 173-180 SEQ ID NO: 13, аминокислотных остатков 31-42 SEQ ID NO: 14 и аминокислотных остатков 122-135 SEQ ID NO: 15. Каждая последовательность данной группы соответствует сегменту экспрессированных, искусственных спидроиновых белков согласно изобретению, белки которого обладают способностью образовывать шелковые волокна при определенных условиях. См. примеры 5-8, 12 и фигуру 2B. Вне связи с какой-либо теорией, авторы изобретения предусматривают, что сегменты **A** согласно изобретению образуют спиральные структуры или бета-листы.

Термин "% идентичности", применяемый в настоящем описании и в прилагаемой формуле изобретения, вычисляется следующим образом. Тестируемую последовательность выравнивают с целевой последовательностью с помощью алгоритма CLUSTAL W (Thompson, J. D., Higgins, D. G. and Gibson, T.J., Nucleic Acids Research, 22: 4673-4680 (1994)). Сравнивают аминокислотные остатки в каждом положении и процентное содержание положений в тестируемой последовательности, которые имеют идентичные соответствия в исследуемой последовательности, описывают как % идентичности.

Термин "% сходства", применяемый в настоящем описании и в прилагаемой формуле изобретения, рассчитывается так же, как описано для "% идентичности", за исключением того, что сходными являются гидрофобные остатки Ala, Val, Phe, Pro, Leu, Ile, Trp, Met и Cys; сходными являются основные остатки Lys, Arg и His; сходными

являются кислые остатки Glu и Asp; и сходными являются гидрофильные, незаряженные остатки Gln, Asn, Ser, Thr и Tyr. Оставшаяся природная аминокислота Gly является не сходной ни с одной аминокислотой в данном контексте.

В настоящем описании альтернативные воплощения согласно изобретению соответствуют вместо указанного процентного значения идентичности, соответствующему процентному значению сходства. Другие альтернативные воплощения отвечают соответствующему процентному значению идентичности, а также другому, более высокому процентному значению сходства, выбранному из группы предпочтительных процентных значений идентичности для каждой последовательности. Например, последовательность может быть на 70% сходной с другой последовательностью; или она может быть на 70% идентична другой последовательности; или она может быть на 70% идентичной и на 90% сходной с другой последовательностью.

В предпочтительных воплощениях согласно изобретению каждый индивидуальный сегмент **A** имеет, по меньшей мере, 90%, более предпочтительно 95%, наиболее предпочтительно 100% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 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 и 1060-1073 SEQ ID NO: 3; аминокислотных остатков 31-42, 61-75, 90-104, 122-135 и 153-171 SEQ ID NO: 9; аминокислотных остатков 12-25, 46-60, 75-88, 112-119, 150-158 и 173-180 SEQ ID NO: 13; аминокислотных остатков 31-42 SEQ ID NO: 14; и аминокислотных остатков 122-135 SEQ ID NO: 15. Таким образом, в определенных воплощениях согласно изобретению каждый индивидуальный сегмент **A** является идентичным аминокислотной последовательности, выбранной из упомянутых выше аминокислотных сегментов.

Кроме того, из представленных в настоящем описании экспериментальных данных было сделано заключение, что каждый индивидуальный **G** сегмент представляет собой аминокислотную последовательность из 12-30 аминокислотных остатков. Предпочтительно каждый индивидуальный сегмент **G** состоит из 14-23 аминокислотных остатков. По меньшей мере, 40% аминокислотных остатков каждого сегмента **G** представляют собой остатки глицина. Обычно содержание глицина в каждом индивидуальном сегменте **G** находится в диапазоне 40-60%.

Предпочтительно каждый индивидуальный сегмент **G** имеет, по меньшей мере, 80% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 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 и 1074-1092 SEQ ID NO: 3. Каждая последовательность данной группы соответствует сегменту природной последовательности белка MaSp1 *Euprosthonops australis*, которая установлена с помощью клонирования соответствующей кДНК, см. примеры 1-2 и фигуру 1-2А. Альтернативно каждый индивидуальный сегмент **G** имеет, по меньшей мере, 80% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 11-30, 43-60, 76-89, 105-121 и 136-152 SEQ ID NO: 9 и аминокислотных остатков 1-11, 26-45, 61-74, 89-111, 120-149 и 159-172 SEQ ID NO: 13. Каждая последовательность данной группы соответствует сегменту

экспрессированных искусственных спидроиновых белков согласно изобретению, белки которого обладают способностью образовывать шелковые волокна при определенных условиях. См. примеры 5-8, 12 и фигуру 2В.

В предпочтительных воплощениях согласно изобретению каждый индивидуальный сегмент **G** имеет, по меньшей мере, 90%, более предпочтительно 95%, наиболее предпочтительно 100% идентичности с аминокислотной последовательностью, выбранной из группы аминокислотных остатков 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 и 1074-1092 SEQ ID NO: 3; аминокислотных остатков 11-30, 43-60, 76-89, 105-121 и 136-152 SEQ ID NO: 9; и аминокислотных остатков 1-11, 26-45, 61-74, 89-111, 120-149 и 159-172 SEQ ID NO: 13. Таким образом, в определенных воплощениях согласно изобретению, каждый индивидуальный сегмент **G** является идентичным аминокислотной последовательности, выбранной из упомянутых выше аминокислотных сегментов.

В определенных воплощениях первые два аминокислотных остатка каждого сегмента **G** согласно изобретению не являются -Gln-Gln-.

В определенных воплощениях положение, соответствующее консервативному остатку Tyr (а именно соответствующее положению 16 в SEQ ID NO: 5, положению 10 в SEQ ID NO: 6 и положению 7 в SEQ ID NO: 7) не является Phe в любом сегменте **G** согласно изобретению.

В определенных воплощениях положение, соответствующее консервативному остатку Tyr (а именно соответствующее положению 16 в SEQ ID NO: 5, положению 10 в SEQ ID NO: 6 и положению 7 в SEQ ID NO: 7) является Tyr в каждом сегменте **G** согласно изобретению.

Из этого следует, что определенные воплощения белков согласно изобретению демонстрируют комбинацию упомянутых выше ограничений, т.е. первые два аминокислотных остатка каждого сегмента **G** согласно изобретению не являются -Gln-Gln- и консервативный остаток Tyr (а именно соответствующий положению 16 в SEQ ID NO: 5, положению 10 в SEQ ID NO: 6 и положению 7 в SEQ ID NO: 7) является Tyr в каждом сегменте **G** согласно изобретению. В определенных воплощениях упомянутые выше ограничения, взятые отдельно или в любой возможной комбинации, далее могут сочетаться с тем ограничением, что фрагмент **REP** не имеет остатков пролина, как обсуждалось выше.

На фигурах 1-2 и в примерах 3-4 представлены три подтипа сегмента **G** согласно изобретению. Данная классификация основана на тщательном анализе последовательности белка MaSp1 *Euprosthops australis* (фигуры 1-2А), и полученная информация была использована и проверена при построении новых искусственных белков шелка паука (фигура 2В).

Первый подтип сегмента **G** согласно изобретению представлен однобуквенной аминокислотной консенсусной последовательностью GQG(G/S)QGG(Q/Y)GG(L/Q)GQGGYGQGA GSS, как показано на фигуре 2А и SEQ ID NO: 5. Первый, и, в целом, самый длинный подтип сегмента **G** обычно содержит 23 аминокислотных остатка, но может содержать только 17 аминокислотных остатков, и не содержит заряженных остатков или содержит один заряженный остаток. Таким образом, предпочтительно указанный первый подтип сегмента **G** содержит 17-23 аминокислотных остатков, но предполагается, что он может содержать минимально 12 или до 30 аминокислотных остатков. Вне связи с какой-либо теорией

предусматривается, что данный подтип образует спиральные структуры или 3₁-спиральные структуры. Типичные сегменты **G** указанного первого подтипа представляют собой аминокислотные остатки 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 и 1074-1092 SEQ ID NO: 3; аминокислотные остатки 11-30, 105-121 и 136-152 SEQ ID NO: 9; и аминокислотные остатки 26-45 и 89-111 SEQ ID NO: 13. Альтернативно сегменты **G** указанного первого подтипа представляют собой аминокислотные остатки 120-149 и 159-172 SEQ ID NO: 13. В определенных воплощениях первые два аминокислотных остатка каждого сегмента **G** данного первого подтипа согласно изобретению не являются -Gln-Gln-.

Второй подтип сегмента **G** согласно изобретению представлен однобуквенной аминокислотной консенсусной последовательностью GQGGQGQG (G/R)YGQG(A/S)G (S/G)S, как показано на фигуре 2A и SEQ ID NO: 6. Указанный второй подтип сегмента **G**, как правило, имеет средний размер, обычно содержит 17 аминокислотных остатков и не содержит заряженных остатков или содержит один заряженный остаток. Предпочтительно указанный второй подтип сегмента **G** содержит 14-20 аминокислотных остатков, но предусматривается, что он может содержать минимально 12 или до 30 аминокислотных остатков. Вне связи с какой-либо теорией предусматривается, что данный подтип образует спиральные структуры. Типичными сегментами **G** указанного второго подтипа являются аминокислотные остатки 249-265, 471-488, 631-647 и 982-998 SEQ ID NO: 3; и аминокислотные остатки 43-60 SEQ ID NO: 9.

Третий подтип сегмента **G** согласно изобретению представлен однобуквенной аминокислотной консенсусной последовательностью G(R/Q)GQG(G/R)YGQG (A/S/V)GGN, как показано на фигуре 2A и SEQ ID NO: 7. Указанный третий подтип сегмента **G** обычно содержит 14 аминокислотных остатков, и, как правило, является самым коротким среди подтипов сегмента **G** согласно изобретению. Предпочтительно указанный третий подтип сегмента **G** содержит 12-17 аминокислотных остатков, но предусматривается, что он может содержать до 23 аминокислотных остатков. Вне связи с какой-либо теорией предусматривается, что данный подтип образует изогнутые структуры. Показательными сегментами **G** указанного третьего подтипа являются аминокислотные остатки 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: 3; аминокислотные остатки 76-89 SEQ ID NO: 9; и аминокислотные остатки 61-74 SEQ ID NO: 13. Альтернативно сегмент **G** указанного третьего подтипа представляет собой аминокислотные остатки 1-11 SEQ ID NO: 13.

Таким образом, в предпочтительных воплощениях, каждый индивидуальный сегмент **G** имеет, по меньшей мере, 80%, предпочтительно 90%, более предпочтительно 95%, идентичности с аминокислотной последовательностью, выбранной из SEQ ID NO: 5, SEQ ID NO: 6 и SEQ ID NO: 7.

В предпочтительном воплощении последовательности из сменяющих друг друга сегментов **A** и **G** фрагмента **REP** каждый второй сегмент **G** относится к первому подтипу, в то время как остальные сегменты **G** относятся к третьему подтипу, например, ...**A**₁**G**_{short}**A**₂**G**_{long}**A**₃**G**_{short}**A**₄**G**_{long}**A**₅**G**_{short}... В другом предпочтительном воплощении фрагмента **REP** регулярность сегментов **G** прерывается вставкой одного сегмента **G** второго подтипа, например, ...**A**₁**G**_{short}**A**₂**G**_{long}**A**₃**G**_{mid}**A**₄**G**_{short}**A**₅**G**_{long}...

Каждый индивидуальный сегмент **L** представляет собой произвольную линкерную аминокислотную последовательность, которая может содержать от 0 до 20

аминокислотных остатков, как, например, от 0 до 10 аминокислотных остатков.

Несмотря на то, что данный сегмент является произвольным и не является функционально критическим в спидроиновом белке, тем не менее его присутствие предусматривается в полностью функциональных спидроиновых белках, образующих
 5 волокна шелка паука согласно изобретению. Линкерные аминокислотные последовательности также представлены в повторяющейся части (SEQ ID NO: 3) в расшифрованной аминокислотной последовательности белка MaSp1 *Euprostheno*
australis. В частности, аминокислотная последовательность линкерного сегмента
 10 может иметь сходство с любым из описанных сегментов **A** или **G**, но обычно указанная последовательность недостаточно соответствует их критериям, описанным в настоящем документе.

Как показано на рисунке 2А, линкерный сегмент, расположенный в С-концевой части фрагмента **REP** может быть представлен однобуквенными аминокислотными
 15 консенсусными последовательностями ASASAAASAA STVANSVS и ASAASAAA, богатыми аланином. В действительности, вторая последовательность может рассматриваться как сегмент **A** согласно изобретению, тогда как первая последовательность имеет высокую степень сходства с сегментами **A** согласно
 20 изобретению. Другой пример линкерного сегмента согласно изобретению имеет однобуквенную аминокислотную последовательность GSAMGQGS, которая богата глицином и имеет высокую степень сходства с сегментами **G** согласно изобретению.

Показательные сегменты **L** представляют собой аминокислотные остатки 1-6 и 1093-1110 SEQ ID NO: 3; аминокислотные остатки 1-10 и 153-171 SEQ ID NO: 9; и
 25 аминокислотные остатки 173-180 SEQ ID NO: 13, однако специалист в данной области техники легко обнаружит, что существует много подходящих альтернативных аминокислотных последовательностей для данных сегментов. В одном воплощении фрагмента **REP** согласно изобретению один из сегментов **L** содержит 0 аминокислот,
 30 т.е. один из сегментов **L** ликвидируется. В другом воплощении фрагмента **REP** согласно изобретению оба сегмента **L** содержат 0 аминокислот, т.е. оба сегмента **L** ликвидируются. Таким образом, данные воплощения фрагментов **REP** согласно изобретению могут быть схематично представлены следующим образом: $(AG)_nL$, $(AG)_nAL$, $(GA)_nL$, $(GA)_nGL$; $L(AG)_n$, $L(AG)_nA$, $L(GA)_n$, $L(GA)_nG$; и $(AG)_n$, $(AG)_nA$,
 35 $(GA)_n$, $(GA)_nG$. Любой из указанных фрагментов **REP** является подходящим для применения с любым фрагментом **CT**, описанным далее.

С-концевой (**CT**) фрагмент спидроинового белка согласно изобретению имеет высокую степень сходства с С-концевой аминокислотной последовательностью
 40 спидроиновых белков. Как показано на фигуре 3, данная аминокислотная последовательность хорошо сохранилась в различных видах и в спидроиновых белках, включая MaSp1 и MaSp2. В следующих примерах показано, что не важно, какой конкретно фрагмент **CT** представлен в спидроиновых белках согласно изобретению, если только фрагмент **CT** не отсутствует полностью. Таким образом,
 45 фрагмент **CT** согласно изобретению может быть выбран из любой из аминокислотных последовательностей, показанных на фигуре 3 или из последовательностей с высокой степенью сходства. Примечательно, что фрагмент CT_{hyb} SEQ ID NO: 13 имеет 96% идентичности с консенсусной аминокислотной последовательностью SEQ ID NO: 8, в
 50 то время как фрагмент CT_{nat} SEQ ID NO: 9 демонстрирует только 59% идентичности с консенсусной аминокислотной последовательностью SEQ ID NO: 8. Эти данные показывают, что в спидроиновом белке согласно изобретению может быть применен широкий ряд С-концевых последовательностей.

Последовательность фрагмента **СТ** согласно изобретению имеет по меньшей мере 50% идентичности, предпочтительно по меньшей мере 60% идентичности, с консенсусной аминокислотной последовательностью SEQ ID NO: 8, которая основана на аминокислотных последовательностях фигуры 3. В предпочтительном воплощении последовательность фрагмента **СТ** согласно изобретению имеет по меньшей мере 65% идентичности, предпочтительно по меньшей мере 70% идентичности с аминокислотными остатками 1-71 консенсусной аминокислотной последовательности SEQ ID NO: 8. В предпочтительных воплощениях фрагмент **СТ** согласно изобретению имеет, кроме того, 70%, предпочтительно 80%, сходства с консенсусной аминокислотной последовательностью SEQ ID NO: 8, или ее аминокислотными остатками 1-71.

Показательные фрагменты **СТ** согласно изобретению представляют собой последовательность SEQ ID NO: 4 *Euprosthénops australis*, полученные из *Euprosthénops australis* аминокислотные остатки 172-269 SEQ ID NO: 9 и аминокислотные остатки 181-276 SEQ ID NO: 13, которые, как считают, происходят из *Euprosthénops sp.* (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. 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)), но которые имеют высокую степень сходства с MaSp1 из *Nephila clavipes* и из *Nephila senegalensis*. Таким образом, согласно предпочтительному аспекту изобретения фрагмент **СТ** имеет, по меньшей мере, 80% идентичности с SEQ ID NO: 4, с аминокислотными остатками 172-269 SEQ ID NO: 9, с аминокислотными остатками 181-276 SEQ ID NO: 13, с аминокислотными остатками 172-269 SEQ ID NO: 16 или с любой индивидуальной аминокислотной последовательностью MaSp1/ADF-4 фигуры 3 и примера 4. В предпочтительных аспектах изобретения фрагмент **СТ** имеет по меньшей мере 90%, а именно по меньшей мере 95% идентичности с SEQ ID NO: 4, с аминокислотными остатками 172-269 SEQ ID NO: 9, аминокислотными остатками 181-276 SEQ ID NO: 13, аминокислотными остатками 172-269 SEQ ID NO: 16 или с любой индивидуальной аминокислотной последовательностью MaSp1/ADF-4 фигуры 3 и примера 4. В предпочтительных аспектах изобретения фрагмент **СТ** является идентичным SEQ ID NO: 4, аминокислотным остаткам 172-269 SEQ ID NO: 9, аминокислотным остаткам 181-276 SEQ ID NO: 13, аминокислотным остаткам 172-269 SEQ ID NO: 16 или любой индивидуальной аминокислотной последовательности MaSp1/ADF-4 фигуры 3 и примера 4.

Фрагмент **СТ** обычно состоит из 70-120 аминокислотных остатков. Предпочтительно фрагмент **СТ** содержит, по меньшей мере, 70 или более чем 80, предпочтительно более чем 90 аминокислотных остатков. Также предпочтительно фрагмент **СТ** содержит, самое большее, 120 или менее чем 110 аминокислотных остатков. Типичный фрагмент **СТ** содержит приблизительно 100 аминокислотных остатков.

Согласно другому аспекту в настоящем изобретении предоставляется выделенный слитый белок, состоящий из первого белкового фрагмента, который является спидпроиновым белком главной ампулярной железы, предпочтительно состоящим из 150-420 аминокислотных остатков, и второго белкового фрагмента, который включает в себя партнера по слитию и участок узнавания расщепляющего агента. Первый белковый фрагмент соединяется через участок узнавания расщепляющего агента с партнером по слитию, т.е. партнер по слитию может быть удален при обработке слитого белка подходящим расщепляющим агентом при подходящих условиях, с получением спидпроинового белка главной ампулярной железы,

предпочтительно состоящего из 150-420 аминокислотных остатков. Преимуществом данного слитого белка является то, что он может быть получен в больших количествах в растворе, предпочтительно в физиологической среде, обычно в буферной водной среде, такой как 10-100 мМ Tris-HCl буфер, pH 6-9, без преципитации и других проблем получения, если его получают в подходящих хозяевах, таких как бактерии, предпочтительно *E. coli*. Растворимые слитые белки могут находиться в растворе в течение долгого периода времени, обычно в течение дней или недель, что способствует их массовому производству и снижает риск белковой агрегации.

Термины "растворимый" и "в растворе" означают, что нет видимой агрегации белка, и белок не осаждается из растворителя при 60000 x g. По желанию слитые белки в растворе могут быть подвергнуты расщеплению подходящим расщепляющим агентом, с получением спидроинового белка главной ампулярной железы, который спонтанно образует шелковые волокна.

В предпочтительном аспекте в настоящем изобретении предоставляется выделенный слитый белок, выбранный из группы **X-REP-CT** и **REP-CT-X**, предпочтительно **X-REP-CT**. **REP** и **CT** являются белковыми фрагментами согласно изобретению, что означает, что итоговый белок MaSp формы **REP-CT** является белком MaSp согласно изобретению. **X** является белковым фрагментом, включающим в себя партнера по слитию и участок узнавания расщепляющего агента, как описано выше. Комбинированный белковый фрагмент **REP-CT** соединяется через участок узнавания расщепляющего агента с партнером по слитию.

Партнеры по слитию согласно изобретению включают в себя любой белковый фрагмент, который улучшает растворимость и/или стабильность его белкового фрагмента - партнера, в настоящем описании белка MaSp согласно изобретению. Партнер по слитию также включает в себя подходящую основу для аффинной очистки. Примеры партнеров по слитию согласно изобретению включают в себя без ограничения тиоредоксин, мальтоза-связывающий белок, глутатион-S-трансфераза (GST), MTB32-C, Gb1, ZZ и Nus A. Специалисту в данной области техники хорошо известны альтернативные подходящие партнеры по слитию. В предпочтительном воплощении изобретения партнер по слитию представляет собой тиоредоксиновую часть (ThrX) в комбинации с His-тагом или с S-тагом. В одном предпочтительном воплощении изобретения партнер по слитию представляет собой часть ThrX в комбинации с двумя His-тагами, а именно His-tag/ThrX/His-tag. В другом предпочтительном воплощении изобретения партнер по слитию представляет собой тиоредоксиновую часть (ThrX).

Участок узнавания расщепляющего агента располагается в том конце белкового фрагмента **X**, который соединяется с белковым фрагментом MaSp, так что расщепление в участке узнавания дает в результате белок MaSp и партнер по слитию. Примеры участка узнавания расщепляющего агента согласно изобретению включают в себя без ограничения участок узнавания тромбина, имеющий аминокислотную последовательность LVPRGS (расщепляется между R и G); участок узнавания энтерокиназы, имеющий аминокислотную последовательность DDDK (расщепляется после K); участок узнавания гидроксиламина, имеющий аминокислотную последовательность NG (расщепляется между N и G); участок узнавания протеазы HRV 3C, имеющий аминокислотную последовательность LGVLFQGP (расщепляется между Q и G); участок узнавания фактора Ха, имеющий аминокислотную последовательность I(E/D)GR (расщепляется после R); участок узнавания TEV, имеющий аминокислотную последовательность EXXYXQ (G/S), общий ENLYFQG

(расщепляется между Q и G/S), участок узнавания Ac-TEV, имеющий аминокислотную последовательность EDNLYFQG (расщепляется между Q и G); и участок узнавания PreScission имеющий аминокислотную последовательность LEVLFQGP (расщепляется между Q и G). Другие подходящие участки узнавания представляют собой участки расщепления для трипсина, эндопротеиназы, протеазы V8, пепсина и CNBr. Другие примеры подходящих участков узнавания для расщепления являются доступными специалисту в данной области техники. В предпочтительном воплощении изобретения участок узнавания расщепляющего агента является участком узнавания тромбина.

В предпочтительном воплощении фрагмент **X** согласно изобретению имеет структуру ThrX/His-tag/S-tag/тромбин-расщепляемый участок, и фрагмент **X** соединяется с N-концом белкового фрагмента **REP-CT** согласно изобретению.

В одном предпочтительном воплощении фрагмент **X** согласно изобретению имеет структуру His-tag/ThrX/His-tag/тромбин-расщепляемый участок, и фрагмент **X** соединяется с N-концом белкового фрагмента **REP-CT** согласно изобретению.

Согласно другому аспекту в настоящем изобретении предоставляется способ получения спидроинового белка главной ампулярной железы согласно изобретению.

На первой стадии получают раствор слитого белка согласно изобретению в жидкой среде. Предпочтительно слитый белок не агрегирует и, следовательно, не требуется процедуры повторной солюбилизации. Слитый белок может быть получен рекомбинантным путем и очищен с помощью аффинной основы в слитом белке, такой как His-таг или любой подходящий эпитоп в слитом белке. Жидкая среда может быть любой подходящей средой, предпочтительно физиологической средой, обычно буферной водной средой, такой как 10-100 mM Tris-HCl буфер, pH 6-9. На второй стадии расщепляющий агент согласно изобретению добавляют к жидкой среде для расщепления слитого белка в участке узнавания расщепляющего агента. Как было раскрыто выше, при этом получают спидроиновый белок главной ампулярной железы согласно изобретению. На третьей необязательной стадии, полученный таким образом спидроиновый белок главной ампулярной железы выделяют из жидкой среды с помощью подходящих техник выделения, таких как хроматография и/или фильтрация.

Согласно еще одному аспекту в настоящем изобретении предоставляется способ получения полимера спидроинового белка главной ампулярной железы согласно изобретению. На первой стадии получают раствор слитого белка согласно изобретению в жидкой среде. Предпочтительно, слитый белок не агрегирует и, следовательно, не требуются процедуры повторной солюбилизации. Слитый белок может быть получен рекомбинантным путем и очищен с помощью аффинной основы в слитом белке, такой как His-таг или любой подходящий эпитоп в слитом белке. Жидкая среда может быть любой подходящей средой, предпочтительно физиологической средой, обычно буферной водной средой, такой как 10-100 mM Tris-HCl буфер, pH 6-9. На второй стадии расщепляющий агент согласно изобретению добавляют к жидкой среде для расщепления слитого белка в участке узнавания расщепляющего агента. Как было раскрыто выше, при этом получают спидроиновый белок главной ампулярной железы согласно изобретению. На третьей стадии проводят полимеризацию полученного таким образом спидроинового белка главной ампулярной железы в жидкой среде. Полимеризация обычно инициируется на границе раздела между двумя различными фазами, такими как границы раздела жидкость/воздух, жидкость/твердая фаза, и вода/масло. Таким образом, данная третья стадия также может далее включать в себя создание границы раздела между жидкой

средой и другой фазой. Другая фаза выбирается из группы, состоящей из газовой фазы, жидкой фазы и твердой фазы. Как подробно описано выше, жидкая среда обычно является водной средой, и другие подходящие фазы представляют собой, например, органические растворители, не смешиваемые с воздухом и водой, такие как
 5 масло, например минеральное масло, подходящее для реакций ПЦР. Присутствие полученной границы раздела стимулирует полимеризацию в самой границе раздела или в окружающей ее области, и данная область распространяется в жидкую среду, так что указанная полимеризация начинается в указанной границе раздела или в
 10 указанной области. Предпочтительные границы раздела включают в себя границы раздела вода/воздух и вода/масло. Полимеризация обычно происходит спонтанно в течение минут или нескольких часов, а именно от 1 мин до 5 часов, при инкубации при комнатной температуре. На необязательной четвертой стадии полученный таким образом полимер спидроинового белка главной ампулярной железы выделяют из
 15 жидкой среды с помощью подходящих техник выделения.

Как обсуждалось выше, образование волокон вызывается протеолитическим высвобождением спидроина небольшого размера из слитого белка. Если реакцию расщепления проводят в пробирке, которую осторожно покачивают из стороны в
 20 сторону, то волокно образуется на границе раздела воздух/вода вдоль по пробирке. Пробирка может быть сделана из любого подходящего материала, такого как пластик или стекло. Если смесь, в которой происходит реакция расщепления, не перемешивают, то, на границе раздела воздух/вода образуется пленка. Если поверх водной смеси для реакции расщепления добавляют масло, то на границе раздела
 25 масло/вода образуется пленка и при перемешивании, и без него. Если смесь, в которой проходит реакция расщепления, вспенивают, например, барботированием воздуха или вибрацией, то получают устойчивую пену, которая затвердевает при высыхании.

Применяя способ(ы) настоящего изобретения, можно получить рекомбинантным
 30 путем большие количества слитых белков согласно изобретению, которые можно расщеплять и проводить их полимеризацию по желанию. Это обеспечивает лучший контроль над процессом полимеризации и позволяет оптимизировать параметры получения шелковых волокон с желаемыми свойствами.

Спидроиновый белок главной ампулярной железы согласно изобретению обычно
 35 получают рекомбинантным путем, используя ряд подходящих хозяев. Согласно другому аспекту в настоящем изобретении предоставляется выделенная полинуклеотидная молекула, включающая в себя последовательность нуклеиновой кислоты, которая кодирует спидроиновый белок главной ампулярной железы
 40 согласно изобретению, или комплементарную ей последовательность нуклеиновой кислоты, а именно SEQ ID NO: 9-13, предпочтительно SEQ ID NO: 9, 12 и 13. Данные молекулы полинуклеиновых кислот, а также молекулы полинуклеиновых кислот, кодирующие различные белки, раскрытые в настоящем документе (SEQ ID NO: 2-16) также могут быть полезны в дальнейших разработках искусственных спидроиновых
 45 белков или систем для их получения.

Слитый белок согласно изобретению обычно получают рекомбинантным путем, с помощью ряда подходящих хозяев, таких как бактерии, дрожжи, клетки млекопитающих, растения, клетки насекомых и трансгенные животные.

Предпочтительно слитый белок согласно изобретению получают в бактериях.

Согласно другому аспекту в настоящем изобретении в силу вышесказанного предоставляется выделенная молекула полинуклеотидной кислоты, включающая в себя последовательность нуклеиновой кислоты, которая кодирует слитый белок

согласно изобретению, или комплементарную ему последовательность нуклеиновой кислоты. Молекула полинуклеиновой кислоты также может быть применена в создании других искусственных спидроиновых белков или систем для их получения.

Молекулы полинуклеиновых кислот согласно изобретению могут быть молекулами ДНК, включая молекулы кДНК, или молекулами РНК. Специалисту в данной области техники хорошо известно, что последовательность нуклеиновой кислоты также может быть описана с помощью комплементарной ей последовательности нуклеиновой кислоты. Поэтому последовательности нуклеиновых кислот, которые комплементарны последовательностям нуклеиновых кислот согласно изобретению, также входят в охраняемый объем изобретения.

Согласно одному аспекту в настоящем изобретении предоставляется способ получения растворимого слитого белка согласно изобретению. На первой стадии молекулу полинуклеиновой кислоты, которая кодирует слитый белок согласно изобретению, экспрессируют в подходящем хозяине. На второй стадии полученный таким образом растворимый слитый белок выделяют, например, с помощью хроматографии и/или фильтрации. По желанию указанная вторая стадия выделения растворимого слитого белка включает в себя удаление LPS и других пирогенов.

В настоящем изобретении далее предоставляется способ получения спидроинового белка главной ампулярной железы согласно изобретению. На первой стадии молекулу полинуклеиновой кислоты, которая кодирует слитый белок согласно изобретению, экспрессируют в подходящем хозяине. На второй стадии полученный таким образом растворимый слитый белок выделяют, например, с помощью хроматографии и/или фильтрации. На третьей стадии получают раствор выделенного слитого белка, на четвертой стадии добавляют подходящий расщепляющий агент в жидкую среду. При этом происходит расщепление слитого белка в участке узнавания расщепляющего агента, и таким образом получают спидроиновый белок главной ампулярной железы. На необязательной пятой стадии полученный таким образом спидроиновый белок главной ампулярной железы выделяют из жидкой среды. Далее по желанию указанная вторая стадия выделения растворимого слитого белка и по желанию пятая стадия выделения спидроинового белка главной ампулярной железы включают в себя удаление LPS и других пирогенов.

В настоящем изобретении также предоставляется способ получения полимера спидроинового белка главной ампулярной железы согласно изобретению. На первой стадии молекулу полинуклеиновой кислоты, которая кодирует слитый белок согласно изобретению, экспрессируют в подходящем хозяине. На второй стадии полученный таким образом растворимый слитый белок выделяют, например, с помощью хроматографии и/или фильтрации. На третьей стадии получают раствор выделенного слитого белка, на четвертой стадии добавляют подходящий расщепляющий агент в жидкую среду. При этом происходит расщепление слитого белка в участке узнавания расщепляющего агента, и таким образом получают спидроиновый белок главной ампулярной железы. На пятой стадии проводят полимеризацию полученного таким образом спидроинового белка главной ампулярной железы в жидкой среде.

Полимеризация обычно инициируется на границе раздела между двумя различными фазами, такими как границы раздела жидкость/воздух, жидкость/твердая фаза и вода/масло. Таким образом, данная пятая стадия также может далее включать в себя создание границы раздела между жидкой средой и другой фазой. Другая фаза выбирается из группы, состоящей из газовой фазы, жидкой фазы и твердой фазы. Как подробно описано выше, жидкая среда обычно является водной средой, и другие

подходящие фазы представляют собой, например, органические растворители, не смешиваемые с воздухом и водой, такие как масло, например минеральное масло, подходящее для реакций ПЦР. Присутствие полученной границы раздела стимулирует полимеризацию в самой границе раздела или в окружающей ее области, и данная область распространяется в жидкую среду, так что указанная полимеризация начинается в указанной границе раздела или в указанной области. Предпочтительные границы раздела включают в себя границы раздела вода/воздух и вода/масло. Полимеризация обычно происходит спонтанно в течение минут или нескольких часов, а именно от 1 мин до 5 часов, при инкубации при комнатной температуре. На необязательной шестой стадии полученный таким образом полимер железа выделяют из жидкой среды.

Для получения белка с низким содержанием пирогенов, что является обязательным при применении биоматериала *in vivo*, был разработан оптимизированный протокол очистки для удаления липополисахаридов (LPS). Чтобы избежать контаминации высвобождаемыми LPS, продуцирующие бактериальные клетки подвергали стадиям отмывки с заменой CaCl_2 и EDTA. После лизиса клеток все последующие стадии очистки проводят в буферах с низкой электропроводностью, чтобы минимизировать гидрофобные взаимодействия между получаемым белком и LPS. Затем содержание LPS минимизируют путем пропускания раствора белка через колонку Endotrap, которая имеет лиганд, который специфически адсорбирует LPS. Чтобы гарантировать постоянное, низкое содержание LPS и других пирогенов, все партии проверяют с помощью теста на пирогенность *in vitro* (IPT) и/или с помощью кинетического теста с лизатом *Limulus amoebocyte* (LAL). Хотя произведенные в Грам-отрицательном бактериальном хозяине рекомбинантные спидроиновые белки могут быть очищены так, чтобы остаточные уровни LPS и других пирогенов были ниже предельных значений, требуемых в тестах на животных, т.е. ниже 25 ЭЕ/имплант. В определенных воплощениях согласно изобретению содержание LPS и других пирогенов в выделенном слитом белке составляет 1 ЭЕ/мг белка или меньше. В определенных воплощениях согласно изобретению содержание LPS и других пирогенов в выделенном спидроиновом белке главной ампулярной железы составляет 1 ЭЕ/мг белка или меньше, предпочтительно 0,25 ЭЕ/мг белка или меньше.

Согласно другому аспекту в настоящем изобретении предоставляется полимер спидроинового белка главной ампулярной железы согласно изобретению. В предпочтительном воплощении полимер данного белка представляет собой полимер, получаемый любым из способов согласно изобретению.

В предпочтительных воплощениях содержание β -листовой формы полимера спидроинового белка главной ампулярной железы согласно изобретению составляет свыше 50%, т.е. более чем 50% вторичной структуры полимера данного белка находится в β -листовой форме. Это является преимуществом, поскольку предполагается, что более высокое содержание β -листовых структур дает более жесткое и/или более крепкое и/или менее растяжимое волокно.

Предпочтительно полимер спидроинового белка согласно изобретению представляет собой волокно макроскопического размера, т.е. с диаметром больше 1 мкм, предпочтительно больше 10 мкм и длиной более 5 мм. Предпочтительно волокно имеет диаметр в диапазоне 10-400 мкм, предпочтительно 60-120 мкм, и длину в диапазоне 0,5-300 см, предпочтительно 1-100 см. Другими предпочтительными диапазонами являются 0,5-30 см и 1-20 см. Также предпочтительно полимер спидроинового белка согласно изобретению имеет предел прочности больше 1 МПа,

предпочтительно больше 2 МПа, более предпочтительно 10 МПа или больше. Предпочтительно полимер спидроинового белка согласно изобретению имеет предел прочности больше 100 МПа, более предпочтительно 200 МПа или больше. Волокно обладает способностью оставаться неповрежденным при физическом воздействии, т.е. его можно применять в прядении, ткачестве, кручении, вязании и в подобных процедурах.

В других предпочтительных воплощениях полимер спидроинового белка согласно изобретению образует пену, гель, сетку или пленку.

Согласно еще одному аспекту в настоящем изобретении предоставляется новое применение белкового фрагмента, включающего в себя партнера по слитию и участок узнавания расщепляющего агента, для производства слитого белка. Слитый белок состоит из указанного белкового фрагмента и из белка шелка паука согласно изобретению, и два фрагмента соединяются через указанный участок узнавания расщепляющего агента. Белковый фрагмент шелка паука предпочтительно состоит из 150-420 аминокислотных остатков.

Далее настоящее изобретение проиллюстрировано в следующих неограничивающих примерах.

ПРИМЕРЫ

Пример 1. Клонирование и секвенирование кДНК MaSp1 *Euprosthrops australis*

Главные ампулярные железы приблизительно из 100 взрослых самок пауков *Euprosthrops australis*, собранных в Южной Африке, использовали для создания по особому заказу библиотеки кДНК на основе pDONR222 (Invitrogen, Paisley, UK). Клоны кДНК, кодирующие белок MaSp1, получили путем скрининга библиотеки с помощью зонда кДНК, кодирующего богатый аланином и глицином фрагмент, полученный из неизвестных подвидов пауков вида *Euprosthrops*. Блоттинг колоний и обнаружение осуществляли с помощью ЭХЛ-направленного мечения и системы обнаружения (Amersham Biosciences, Uppsala, Sweden) согласно инструкции производителя.

Для дальнейшего исследования выбирали один единичный клон. Для получения полноразмерной последовательности вставки кДНК из данного клона делали внутригенные делеции с помощью системы Erase-a-Base System (Promega, Southampton, UK) и осуществляли секвенирование на MegaBase 1000 instrument (Amersham Biosciences).

Полученная кДНК из 3,8 т.п.н. (SEQ ID NO: 1) кодирует белок MaSp1 (SEQ ID NO: 2) из 1207 аминокислотных остатков, содержащий повторяющийся фрагмент из 34 аланин- и глицин-богатых сегментов (SEQ ID NO: 3) и С-концевой неповторяющийся фрагмент из 97 аминокислотных остатков (SEQ ID NO: 4).

Пример 2. Анализ последовательности повторяющегося фрагмента белка MaSp1 *Euprosthrops australis*

Белковую последовательность повторяющегося фрагмента белка MaSp1 *Euprosthrops australis* примера 1 (SEQ ID NO: 3) далее анализировали путем выравнивания повторяющихся сегментов фрагмента, см. фигуру 1. Результаты выравнивания были тщательно исследованы, и были сделаны следующие выводы о структуре.

Аланин-богатые сегменты белка MaSp1 *Euprosthrops australis* имеют длину 13-15 аминокислотных остатков и состоят только из остатков аланина или из остатков аланина и одного остатка, который представляет собой остаток серина, глутамата или глицина.

Повторяющийся фрагмент белка MaSp1 *Euprosthénops australis* содержит три родственных, но различных типа глицин-богатых сегментов, перенесенных на фигуру 2А. Два глицин-богатых сегмента отличаются только по длине и встречаемости; самый распространенный глицин-богатый сегмент содержит 23 аминокислотных остатка, в то время как наименее распространенный вариант содержит 17 аминокислотных остатков. Оба указанных глицин-богатых сегмента обычно не имеют заряженных остатков или содержат один заряженный остаток. В отличие от этого самый короткий глицин-богатый сегмент, состоящий из 14 аминокислотных остатков, единственный содержит последовательность GRGQG или GQGQG в N-концевой области, и GN в C-концевой области.

Самый длинный глицин-богатый сегмент представлен однобуквенной аминокислотной консенсусной последовательностью

GQG(G/S)QGG(Q/Y)GG (L/Q)GQGGYGQGA GSS (SEQ ID NO: 5), и не имеет заряженных остатков. Предсказывается, что данный сегмент образует спиральные структуры или 3_1 -спиральные структуры. Богатый глицином сегмент среднего размера представлен однобуквенной аминокислотной консенсусной последовательностью GQGGQGQG(G/R)Y GQG(A/S)G(S/G)S (SEQ ID NO: 6), и не имеет заряженных остатков или содержит один заряженный остаток. Как предсказывают, данный сегмент образует спиральные структуры. Самый короткий глицин-богатый сегмент представлен однобуквенной аминокислотной консенсусной последовательностью G(R/Q)GQG(G/R)YGQG (A/S/V)GGN (SEQ ID NO: 7).

Предсказывается, что данный сегмент образует изогнутые структуры.

Повторяющийся фрагмент белка MaSp1 *Euprosthénops australis* строится из чередующихся богатых аланином и богатых глицином сегментов, например,

...A₁G₁A₂G₂A₃G₃A₄G₄A₅G₅...

Отмечается, что каждый из идентифицированных выше самый короткий и самый длинный глицин-богатые сегменты обычно встречаются в качестве каждого второго глицин-богатого сегмента, например,

...A₁G_{short}A₂G_{long}A₃G_{short}A₄G_{long}A₅G_{short}...

И напротив, менее распространенный, богатый глицином фрагмент среднего размера обычно находится между глицин-богатым сегментом более длинного типа и глицин-богатым сегментом более короткого типа, например,

...A₁G_{short}A₂G_{long}A₃G_{mid}A₄G_{short}A₅G_{long}...

Пример 3. Предсказание вторичной и третичной структуры повторяющегося фрагмента белка MaSp1 *Euprosthénops australis*

Спидроиновые полипептиды в растворе обычно сворачиваются с образованием структуры шпильки, которая служит прообразом антипараллельной β -складчатой структуры готового волокна. Чтобы определить возможные типы сворачивания повторяющегося фрагмента (SEQ ID NO: 3) белка MaSp1 *Euprosthénops australis* из примеров 1-2, идентифицируют участки белка, которые являются сходными с участками, образующими шпильку или изогнутые структуры. Аланин-богатые сегменты являются маловероятными кандидатами для изогнутого образования, поскольку, согласно расчетам, они образуют спиральные структуры, и, что более важно, указанные сегменты обычно образуют β -слои в волокне.

При применении недавно описанного алгоритма предсказания формы (Fuchs, PF & Alix, AJ, High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 59, 828-839 (2005)), самые короткие глицин-богатые сегменты показывают высокую вероятность формирования β -изгибов типа II, в то

время как для двух более длинных глицин-богатых сегментов предсказывают образование спиральных структур. Высокое содержание триплетов GGX в длинных Gly-богатых сегментах предполагает, что они могут образовывать 3₁-спиральные структуры.

Повторяющийся характер аминокислотных последовательностей спидроина предполагает таким же образом повторяющийся характер типа укладки. Данные наблюдения, взятые вместе, дают в результате укладку повторяющегося фрагмента белка MaSp1 *Euprosthénops australis*, которая показана на фигуре 2А. Примечательно, что положительно заряженные остатки почти всегда расположены в предполагаемых изогнутых структурах.

Из типа укладки повторяющегося фрагмента белка MaSp1 *Euprosthénops australis* может быть выделен мотив следующего состава: аланин-богатый сегмент/(длинный) глицин-богатый спиральный сегмент/аланин-богатый сегмент/(короткий) глицин-богатый изогнутый сегмент/аланин-богатый сегмент/(длинный) глицин-богатый спиральный сегмент/аланин-богатый сегмент, (проиллюстрировано в виде схемы на фигуре 2А).

Пример 4. Анализ последовательности неповторяющегося С-концевого фрагмента белков MaSp1

Производили выравнивание первичной структуры С-концевого неповторяющегося фрагмента (SEQ ID NO: 4) белка MaSp1 *Euprosthénops australis*, полученного в примере 1, с рядом известных С-концевых фрагментов белков MaSp1 и MaSp2, в частности из вида *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)), *Nephila clavipes* P19837-5 (Xu, M & Lewis, RV, Structure of a protein superfiber: spider dragline silk. Proc Natl Acad Sci USA 87, 7120-7124 (1990)) и из других видов.

Из показанного на фигуре 3 выравнивания, которое начинается с последнего Ser в повторяющемся фрагменте, становится очевидным, что С-концевые участки MaSp1 и MaSp2 являются высоко консервативными. Виды *Euprosthénops* sp и *Nephila clavipes* имеют 95% идентичных остатков; *Euprosthénops australis* и *Nephila clavipes* имеют 54% идентичных остатков; и *Euprosthénops australis* и *Euprosthénops* sp имеют 55% идентичных остатков. Консенсусная последовательность С-концевых участков MaSp1 и MaSp2 предоставляется в виде SEQ ID NO: 8. На фигуре 3 показано выравнивание следующих белков MaSp, обозначенных номерами доступа GenBank:

Виды пауков и спидроиновый белок

Код

<i>Euprosthénops</i> sp MaSp1 (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, <i>ibid</i>)	Cthyb_Esp
<i>Euprosthénops australis</i> MaSp1 (SEQ ID NO: 4)	CTnat_Eau
<i>Argiope trifasciata</i> MaSp1	AF350266_At1
<i>Cyrtophora moluccensis</i> Sp1	AY666062_Cm1
<i>Latrodectus geometricus</i> MaSp1	AF350273_Lg1
<i>Latrodectus hesperus</i> MaSp1	AY953074_Lh1
<i>Macrothele holsti</i> Sp1	AY666068_Mh1
<i>Nephila clavipes</i> MaSp1	U20329_Nc1

	<i>Nephila pilipes</i> MaSp1	AY666076_Np1
	<i>Nephila madagascariensis</i> MaSp1	AF350277_Nm1
	<i>Nephila senegalensis</i> MaSp1	AF350279_Ns1
5	<i>Octonoba varians</i> Sp1	AY666057_Ov1
	<i>Psechrus sinensis</i> Sp1	AY666064_Ps1
	<i>Tetragnatha kauaiensis</i> MaSp1	AF350285_Tk1
	<i>Tetragnatha versicolor</i> MaSp1	AF350286_Tv1
10	<i>Araneus bicentenarius</i> Sp2	ABU20328_Ab2
	<i>Argiope amoena</i> MaSp2	AY365016_Aam2
	<i>Argiope aurantia</i> MaSp2	AF350263_Aau2
15	<i>Argiope trifasciata</i> MaSp2	AF350267_At2
	<i>Gasteracantha mammosa</i> MaSp2	AF350272_Gm2
	<i>Latrodectus geometricus</i> MaSp2	AF350275_Lg2
	<i>Latrodectus hesperus</i> MaSp2	AY953075_Lh2
20	<i>Nephila clavipes</i> MaSp2	AY654293_Nc2
	<i>Nephila madagascariensis</i> MaSp2	AF350278_Nm2
	<i>Nephila senegalensis</i> MaSp2	AF350280_Ns2
25	<i>Dolomedes tenebrosus</i> Fb1	AF350269_DtFb1
	<i>Dolomedes tenebrosus</i> Fb2	AF350270_DtFb2
	<i>Araneus diadematus</i> ADF-1	U47853_ADF1
	<i>Araneus diadematus</i> ADF-2	U47854_ADF2
30	<i>Araneus diadematus</i> ADF-3	U47855_ADF3
	<i>Araneus diadematus</i> ADF-4	U47856_ADF4

Пример 5. Построение генов MaSp1

35 Последовательность ДНК, кодирующую белок 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), полученный из *Euprosthenops australis*, амплифицировали методом ПЦР с набором Advantage GC2 kit (BD Biosciences, San Jose, CA, USA) с применением в качестве матрицы клона MaSp1 из библиотеки кДНК примера 1. Участки узнавания ферментов рестрикции *Bam*HI и *Hind*III вводили в 5'- и 3'-концы, соответственно, и стоп-
40 кодон вводили перед участком *Hind*III, с применением желаемых праймеров. Затем конструкцию *Bam*HI-5Gly/Ala-CT_{nat}-*Hind*III субклонировали в модифицированный вектор pET32 (Merck Biosciences, Darmstadt, Germany), приготовленный, как описано в примере 6(С) ниже.

Пример 6. Построение химерных генов MaSp1

(А) фрагменты гена REP

45 Последовательности ДНК, кодирующие неполные повторяющиеся фрагменты (REP), обозначаемые 3Gly/Ala и 4Gly/Ala, были амплифицированы методом ПЦР с *LA Taq* (TaKaRa Bio; Saint Germain-en-laye, France) в присутствии бетаина (Henke W et al, 50 Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 25, 3957-3958 (1997)), с применением в качестве матрицы клона неполной кДНК, кодирующей повторяющийся участок белка MaSp1 вида *Euprosthenops* 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)) (номер Генбанка CQ974358 или CQ816656). Участки узнавания фермента рестрикции вводили в 5'- и 3'-концы с получением следующих конструкций: *NcoI-3Gly/Ala-NheI* и *NcoI-4Gly/Ala-NheI*, которые соединялись с фрагментом **СТ** (см. ниже); и клон *NcoI-4Gly/Ala-XhoI*, который экспрессировался индивидуально, где стоп-кодон вставляли непосредственно перед участком *XhoI*.

(В) фрагменты гена СТ

Последовательность ДНК, кодирующая неповторяющийся С-концевой домен А вида *Euprosthénops* sp (но с высокой степенью сходства с MaSp1 из *Nephila clavipes* и *Nephila senegalensis*) амплифицировали методом ПЦР с применением клона геномной ДНК, кодирующего С-концевой домен MaSp1 (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)). Участки узнавания ферментов рестрикции вводили в 5'- и 3'-концы с получением *NheI-2Gly/Ala-CT_{hyb}-XhoI*, которые соединяли с клонами неполных REP 3Gly/Ala и 4Gly/Ala (см. выше), и *NhoI-2Gly/Ala-CT_{hyb}-XhoI*, который экспрессировали индивидуально.

(С) Построение генов REP-СТ гибрида MaSp1

Клоны REP 3Gly/Ala и 4Gly/Ala соединяли клонами СТ с помощью вектора pCR[®] 2.1-TOPO[®] (Invitrogen). Затем полученные гибридные клоны 5Gly/Ala-CT_{hyb} и 6Gly/Ala-CT_{hyb} вырезали *NcoI* и *XhoI* и субклонировали в модифицированный вектор pET32 (Novagen), где участок расщепления тромбина был удален и новый тромбиновый участок был введен после участка расщепления энтерокиназы.

Пример 7. Экспрессия слитых белков MaSp1

Белки MaSp1, кодированные генами, построенными в примерах 5-6, экспрессировали в качестве слитых белков (типа **X-REP-СТ**), как изложено ниже, с применением модифицированного вектора pET32: тиоредоксин-таг/His-таг/S-таг/участок расщепления тромбина/ген MaSp1, кодирующий тиоредоксин/His₆/S-таг/участок расщепления тромбина/белок MaSp1, и ген устойчивости к ампициллину под контролем промотора T7.

Различные конструкции MaSp1 в векторах экспрессии pET32 трансформировали в клетки *Escherichia coli* BL21(DE3) (Merck Biosciences). Клетки выращивали при 30°C в среде Лурия-Бертани, содержащей ампициллин, до OD₆₀₀ 1,0-1,5, вызванной IPTG, и затем инкубировали 4 часа при комнатной температуре. Клетки собирали центрифугированием и разрушали ДНКазой I и лизозимом в 20 mM Tris-HCl, pH 8,0, 20 mM имидазоле с 0,5 M NaCl и затем очищали аффинной хроматографией с His-таг на агарозе Ni-NTA (Qiagen, West Sussex, UK). Связанные слитые белки элюировали из колонки с Ni-NTA 200 mM имидазолом в 20 mM Tris-HCl, pH 8,0, с 0,5 M NaCl, и проводили диализ против 20 mM Tris-HCl, pH 8,0. Полученные слитые белки имели чистоту >90%, согласно результатам окраски по Кумасси SDS-полиакриламидных гелей, и их растворяли в 20 mM Tris-HCl, pH 8,0. Данный процесс давал приблизительно 40 мг/л культуры слитого белка, который оставался стабильным в течение недель без заметного образования осадка.

В другом эксперименте слитые белки экспрессировали в виде белков His₆/тиоредоксин/His₆/участок расщепления тромбина/MaSp1 из плазмиды, содержащей соответствующий ген, и ген устойчивости к канамицину под контролем промотора T7.

Пример 8. Образование волокон из белков MaSp1

Расщепление тагов слитых белков, полученных в примере 7, проводили в 20

мМ Tris-HCl, pH 8, с соотношением тромбин:слитый белок 1:1000 (об./об.), при очень осторожном покачивании при комнатной температуре. Расщепление тромбина происходило в течение 30-60 мин, согласно результатам электрофореза в SDS-полиакриламидном геле. Полученные белки MaSp1 (фигура 2В, SEQ ID NOS: 9-13) спонтанно полимеризовались в макроскопические волокна различных размеров, показанных в таблице 1. Волокна исходно формировались на границе раздела вода/воздух. Образование можно было наблюдать невооруженным глазом с 1 часа инкубации (см. фигуры 4А, 4В) и после 5 часов не происходило дальнейшего роста волокон. Волокна 6Gly/Ala-CT_{hyb} имели длину приблизительно 2 см и волокна 5Gly/Ala-CT_{nat} имели длину ≥ 10 см. В повторных экспериментах получили волокна 5Gly/Ala-CT_{nat}, которые имели длину ≥ 20 см (см. фигуру 4С), и даже ≥ 2 м. Образование волокон можно было наблюдать невооруженным глазом после 10 мин инкубации.

Волокна выделяли и отмывали буфером и затем проводили анализ N-концевой аминокислотной последовательности, который показывал только последовательность белка MaSp1. Анализ показал отсутствие расщепленных тагов в волокнах.

Пример 9. Анализ белковых волокон MaSp1

А. Измерения предела прочности

Предел прочности волокон 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13) и 5Gly/Ala-CT_{nat} (SEQ ID NO: 9) примера 8 устанавливали следующим образом. Для манипуляций с короткими (1-2 см) волокнами 6Gly/Ala-CT_{hyb} в ходе измерения пределов прочности их инкубировали в течение короткого времени в 15% растворе глицерина в воде перед высушиванием на воздухе. Длинные волокна (10 см) 5Gly/Ala-CT_{nat} либо не обрабатывали, либо инкубировали в течение короткого времени в 15% глицерине, либо вытягивали рукой в 75% метаноле перед высушиванием на воздухе. Предел прочности высушенных на воздухе волокон измеряли путем растягивания волокон с помощью прибора Zwick Material Tester со скоростью 10 мм/мин. См. таблицу 1.

Предел прочности обработанных глицерином, высушенных на воздухе, длиной 1-2 см волокон из 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13) составлял 2 МПа, и прочность волокон длиной 10 см из 5Gly/Ala-CT_{nat} (SEQ ID NO: 9) составляла 4-5 МПа.

Десятисантиметровые волокна 5Gly/Ala-CT_{nat}, вытянутые в обезвоженном растворителе метаноле перед высушиванием на воздухе имели предел прочности 2-3 МПа, что немного меньше, чем предел прочности у обработанных глицерином волокон такой же длины. Самый высокий измеренный в эксперименте предел прочности составлял 10 МПа, и он был получен для волокна 5Gly/Ala-CT_{nat} длиной 10 см, высушенного на воздухе без другой обработки.

Диапазон предела прочности (2-10 МПа) сравним с более низкими значениями, опубликованными для регенерированных волокон шелка паука (2-320 МПа). Самые длинные спонтанно образованные волокна получены из конструкции 5Gly/Ala-CT_{nat}, и такие высушенные на воздухе волокна также имеют наибольший предел прочности. Потенциально, указанный предел прочности может быть связан с наличием в данном белке поли-Ala сегментов из 12-15 остатков, по сравнению с поли-Ala сегментами из 8-14 остатков в 6Gly/Ala-CT_{hyb}, что могло бы давать большую долю кристаллической конформации β -листа в первом белке.

Таблица 1 Способность к образованию волокон у белков MaSp1				
Белок	SEQ ID NO	Способность образовывать волокна	Длина волокна (см)	Предел прочности волокна (МПа)

5Gly/Ala-CT _{nat}	9	++++	≥10	2-10
5Gly/Ala-CT _{nat} (пример 11)	9	+++++	≥20	180-230
4Gly/Ala	10	агрегаты	n.a.	n.a.
2Gly/Ala-CT _{hyb}	11	агрегаты	n.a.	n.a.
5Gly/Ala-CT _{hyb}	12	+	≤1	n.d.
6Gly/Ala-CT _{hyb}	13	+++	1-2	≤2
n.a. = не существует n.d. = не определен				

В. Сканирующая электронная микроскопия

Микроскопическую архитектуру волокон 6Gly/Ala-CT_{hyb} и 5Gly/Ala-CT_{nat} анализировали с помощью сканирующего электронного микроскопа (SEM) (фигура 5). Кратко, образцы помещали на целики SEM и покрывали в вакууме 6 нм

слоем золота и палладия. Образцы просматривали и фотографировали с помощью микроскопа SEM LEO 1550 FEG с применением ускоряющего напряжения 10 кВ. Данным методом выявлены диаметры одиночных волокон 10-30 мкм, при этом индивидуальные волокна демонстрировали довольно однородные диаметры (показаны на фигуре 5а 6Gly/Ala-CT_{hyb}, SEQ ID NO: 13). В дополнение к макроскопическим волокнам были обнаружены гелеподобные частицы. После высушивания на воздухе данные частицы 6Gly/Ala-CT_{hyb} помещали непосредственно на целик SEM, и были видны волокна с диаметром, приблизительно, 10-15 мкм (фигуры 5b, c). Диаметр макроскопических волокон 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), вытянутых в 75% метаноле и высушенных на воздухе, составлял 60-120 мкм, и они, очевидно, содержали несколько выровненных волокон (фигуры 5d-f). Волокно, скрученное перед высушиванием на воздухе (фигура 5е), конец волокна (фигура 5f).

С. Спектроскопия кругового дихроизма

Волокна, состоящие из белка 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13) или 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), в примере 8, отмывали в 20 мМ фосфатном буфере, pH 7, и суспендировали в 2% SDS в том же самом буфере. Спектры кругового дихроизма от 250 до 190 нм регистрировали при 22°C в кварцевой кювете с длиной пути 0,1 см с помощью спектрополяриметра Jasco J-810. Скорость сканирования составляла 50 нм/мин, время ответа 2 с, интервал поглощения 0,1 нм, и ширина полосы 1 нм.

Спектр, представленный на фигуре 6, представляет собой совокупность трех сканирований волокон белка 6Gly/Ala-CT_{hyb} (SEQ ID NO: 13). Спектр имеет минимум при 220 нм и максимум при 195 нм, т.е. параметры, которые характерны для антипараллельных β-листовых структур. Очень похожие спектры были получены для волокон 5Gly/Ala-CT_{nat} (не показаны). Таким образом, спонтанно образованные волокна демонстрируют сходную морфологию и структуру как, например, у нативных и регенирированных волокон шелка паука.

Пример 10. Биосовместимость рекомбинантного шелка паука

Поскольку желательна применение волокон шелка паука в биомедицинских областях, биосовместимость волокон оценивали путем выяснения эффектов рекомбинантно полученного шелка на различных типах клеток.

Белок MaSpl 5Gly/Ala-CT_{nat} (SEQ ID NO: 9) экспрессировали в бактериях, как описано в примерах 7-8. Очищенный белок применяли для получения волокон искусственного шелка длиной >10 см, и даже >20-200 см, и диаметром примерно 100 мкм.

А. Эмбриональные мышечные тучные клетки

Эмбриональные (12,5-дневные) мышинные тучные клетки (пролиферация *in vitro* в течение восьми недель с применением IL-8 и фактора тучных стволовых клеток) в двух различных клеточных плотностях, при этом высокая клеточная плотность была в четыре раза больше низкой плотности. Указанные клетки не прилипали к пластиковой поверхности и росли в суспензии. Кусочки шелкового волокна, каждый длиной, примерно, 0,5 см, добавляли в лунки. Тучные клетки инкубировали в течение трех дней в присутствии или в отсутствие шелкового волокна, и затем подсчитывали живые и мертвые клетки после окраски трипановым синим (фигура 7). Полосками показаны средние значения со стандартной ошибкой среднего, $n=2$, каждую пробу подсчитывали в триплете.

Тучные клетки не подвергались изменениям в присутствии шелковых волокон. После трех дней роста не наблюдали значительных различий в клеточной смерти или пролиферации, по сравнению с отрицательными контролями, которые выращивали без шелковых волокон.

Тестируемые тучные клетки не подвергались изменениям в присутствии рекомбинантных шелковых волокон, даже при сравнительно высоких количествах шелка. Указанные результаты показывают, что тестируемые волокна искусственного шелка имеют сходство с диким типом канатного шелка *Euprostheno australis* и являются нетоксичными и биосовместимыми. Таким образом, волокна искусственного шелка представляются подходящими для биомедицинских применений.

Пример 11. Механические свойства и структура волокон из белка MaSpl

Механические свойства волокон из 5Gly/Ala-CT_{nat} (SEQ ID NO: 9) проверяли с помощью тестов на определение предела прочности, при осуществлении которых получали кривые зависимости деформации от напряжения (фигура 9). Свойства прочности при растяжении определяли с помощью прибора Zwick Roell Z2.5 material tester (Zwick, Ulm, Germany). Тесты осуществляли в воздухе при комнатных условиях (20°C и 52% относительной влажности), используя скорость нагрузки 10 мм/мин. Кусочки волокон брали непосредственно из буфера, закрепляли в приборе и подвергали двум циклам растяжения-расслабления. Для получения однородной шелковой нити, подходящей для теста на растяжение, волокна растягивали с помощью циклов растяжения-расслабления. Сначала волокна удлиняли путем вытягивания с силой 0,1 Н. После расслабления волокна продолжали вытягивать до приложения силы в 0,25 Н.

В результате такой обработки получали вытянутые однородные волокна с диаметром приблизительно 80 мкм, как определяли по измерениям высоты с помощью прибора Mitutoyo IDC-112B instrument (Mitutoyo Corp, Tokyo, Japan) и подтверждали сканирующей электронной микроскопией (SEM), как описано далее. До и после циклов вытягивания-расслабления кусочки волокон наносили на целики SEM и высушивали на воздухе в течение ночи. Образцы покрывали под вакуумом 6 нм слоем золота и палладия. Образцы просматривали и фотографировали с помощью микроскопа LEO 1550 FEG (Carl Zeiss, Oberkochen, Germany) с применением ускоряющего напряжения 10 kV.

Вытянутые волокна были нарезаны на кусочки, концы которых закрепляли клеем между плотной бумагой (Loctite 420, Loctite, Göteborg, Sweden). Затем образцы волокон закрепляли в губках тисков прибора для испытания материалов и растягивали до их разрыва. Кривые зависимости деформации от напряжения построены с применением начальной площади поперечного сечения невытянутого волокна, исходя из круглого поперечного сечения. Значения напряжения нормализованы относительно начальной

площади сечения волокна. Деформация соответствует dL/L_0 , где L_0 представляет собой начальную длину волокна и dL представляет собой изменение длины волокна. На фигуре 9 показаны кривые зависимости напряжение-деформация для трех различных образцов дважды вытянутых волокон 5Gly/Ala-CT_{nat} (SEQ ID NO: 9), и их пределы прочности, оцененные приблизительно в 0,2 ГПа.

Микроскопическую архитектуру волокон анализировали с помощью SEM (фигура 10). Спонтанно сформированные волокна имеют однородный сглаженный внешний вид и ширину до нескольких сотен микрометров, в то время как высота оценивается в несколько десятков микрометров (фигура 10а, б).

После того как волокна подвергали циклам вытягивания-расслабления, их поперечное сечение принимало более округлую форму с компактной субструктурой из сильно выровненных фибрилл (фигура 10с-ф). Внешний вид разрезанных или разрушенных поверхностей (фигура 10е, ф) дополнительно подтверждает компактность полученного волокна.

В заключение следует отметить, что спонтанно образованные волокна и без формирования демонстрируют морфологию и механические свойства, сходные с нативными или регенерированными волокнами шелка паука.

Пример 12. Варианты спидроинового белка

Предполагают, что сильные межмолекулярные взаимодействия вносят свой вклад в необычайно высокий предел прочности шелка паука. Поэтому были созданы варианты небольших спидроинов, которые позволяют внутримолекулярное ковалентное перекрестное сшивание в волокнах. Два различных мутантных спидроиновых белка были сконструированы путем сайт-направленного мутагенеза для введения двух остатков цистеина в первый (SEQ ID NO: 14, положения 36 и 37) и в четвертый (SEQ ID NO: 15, положения 128 и 129) аланиновые блоки, соответственно. Данные варианты были экспрессированы и выделены с помощью того же протокола, который описан в примерах 7-8 для генов, сконструированных в примерах 5-6. Данные варианты (SEQ ID NOS: 14-15) образуют волокна таким же образом, как 5Gly/Ala-CT_{nat} (SEQ ID NO: 9).

Чтобы выяснить значение димеризации С-концевого домена, был построен вариант, в котором остаток цистеина в С-концевом домене заменяли остатком серина (SEQ ID NO: 16, положение 222). Несмотря на это данный вариант (SEQ ID NO: 16) формировал волокна таким же образом, как 5Gly/Ala-CT_{nat} (SEQ ID NO: 9).

Пример 13. Удаление LPS и других пирогенов из экспрессированных спидроиновых белков

Клетки E.coli, экспрессирующие желаемый спидроиновый слитый белок, отмывают следующими буферами:

A: 100 mM Tris, pH 8,

B: 5 mM CaCl₂, 100 mM Tris, pH 8,

C: 10 mM EDTA, 100 mM Tris, pH 8,

D: 100 mM Tris, pH 8, и

E: 100 mM Tris, pH 8.

Затем клетки разрушали в 20 mM Tris, pH 8 дополненном лизозимом и ДНКазой I. Затем пробу белка помещали на Ni-сефарозный матрикс и отмывали 20 mM Tris, 10-100 mM имидазолом, pH 8 перед элюированием 20 mM Tris, 100-300 mM имидазолом, pH 8. Соответствующие фракции собирали и проводили диализ против 20 mM Tris, pH 8 в течение ночи. Пробу белка затем дополняют 100 мкМ CaCl₂ и в заключение пропускают через колонку EndoTrap Blue, предварительно уравновешенную 20

мМ Tris, 100 мкМ CaCl₂, pH 8. Таким образом, могут быть получены пробы белка с содержанием пирогена 1 ЭЕ/мг белка согласно оценке IPT и кинетического теста LAL.

Затем слитый белок протеолитически расщепляют тромбином в соотношении 1: 1000 (масс./масс.) тромбин:слитый белок, что вызывает образование волокон (как описано выше). Волокна отмывали 3 раза в 20 мМ Tris, pH 8 и в заключение 3 раза в воде. Получают волокна с содержанием пирогенов 0,25 ЭЕ/мг волокна.

Структурные характеристики волокон не изменяются под действием автоклавирования при 125°C и 1,5 бар в течение 10 мин, что обеспечивает эффективную стерилизацию материала. Волокна являются химически стабильными и не растворяются в 8 М мочеvine, 6 М GuaHCl или в чистом HAc. Однако волокна могут растворяться в чистом HFIP или в муравьиной кислоте.

Формула изобретения

1. Рекомбинантный спидроиновый белок главной ампулярной железы, способный к полимеризации, отличающийся тем, что указанный белок состоит из 150-420 аминокислотных остатков и описывается формулой REP-CT, где

REP представляет собой фрагмент белка, имеющий 80-300 аминокислотных остатков, где указанный фрагмент выбран из группы L(AG)_nL, L(AG)_nAL, L(GA)_nL, L(GA)_nGL, где n представляет собой целое число от 4 до 8; каждый индивидуальный сегмент A представляет собой аминокислотную последовательность из 8-18 аминокислотных остатков, где от 0 до 3 аминокислотных остатков не являются Ala, и остальные аминокислотные остатки представляют собой Ala; каждый отдельный сегмент G является аминокислотной последовательностью из 12-30 аминокислотных остатков, где по меньшей мере 40% аминокислотных остатков являются Gly; и каждый индивидуальный сегмент L является линкерной аминокислотной последовательностью из 0-20 аминокислотных остатков; и

CT представляет собой фрагмент белка, имеющий от 70 до 120 аминокислотных остатков, являющийся С-концевым фрагментом, полученным из спидроинового белка главной ампулярной железы.

2. Рекомбинантный белок по п.1, в котором каждый индивидуальный сегмент A идентичен или, по меньшей мере, на 80% идентичен аминокислотной последовательности, выбранной из группы аминокислотных остатков 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 и 1060-1073 SEQ ID NO: 3; аминокислотных остатков 31-42, 61-75, 90-104, 122-135 и 153-171 SEQ ID NO: 9; аминокислотных остатков 12-25, 46-60, 75-88, 112-119, 150-158 и 173-180 SEQ ID NO: 13; аминокислотных остатков 31-42 SEQ ID NO: 14; и аминокислотных остатков 122-135 SEQ ID NO: 15;

каждый индивидуальный сегмент G идентичен или, по меньшей мере, на 80% идентичен аминокислотной последовательности, выбранной из группы аминокислотных остатков 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 и 1074-1092 SEQ ID NO: 3; SEQ ID NO: 5-7; аминокислотных остатков 11-30, 43-60, 76-89, 105-121 и 136-152 SEQ ID NO: 9; и аминокислотных остатков 1-11, 26-45, 61-74, 89-111, 120-149 и 159-172 SEQ ID NO: 13; и

указанный фрагмент СТ идентичен или, по меньшей мере, на 50% идентичен SEQ ID NO: 8; или где указанный фрагмент СТ идентичен или, по меньшей мере, на 80% идентичен аминокислотной последовательности, выбранной из группы, состоящей из SEQ ID NO: 4, аминокислотных остатков 172-269 SEQ ID NO: 9, аминокислотных остатков 181-276 SEQ ID NO: 13, и аминокислотных остатков 172-269 SEQ ID NO: 16.

3. Рекомбинантный белок по любому из пп.1-2, выбранный из группы, состоящей из SEQ ID NO: 9 и SEQ ID NO: 12-16.

4. Рекомбинантный слитый белок, способный к управляемой полимеризации, выбранный из группы X-REP-СТ и REP-СТ-X, где REP и СТ являются фрагментами белка по любому из пп.1-3; и

X является фрагментом белка, содержащим партнера по слитию и участок узнавания расщепляющего агента;

где комбинированный фрагмент белка REP-СТ соединяется через указанный участок узнавания расщепляющего агента с указанным партнером по слитию.

5. Полимер спидроинового белка главной ампулярной железы по любому из пп.1-3 для применения в качестве материала для медицинских или технических целей.

6. Композиция для управляемой полимеризации, включающая в себя эффективное количество рекомбинантного белка по п.4, в которой содержание LPS и других пирогенов составляет 1 ЭЕ/мг выделенного белка или менее.

7. Рекомбинантная молекула нуклеиновой кислоты, содержащая последовательность нуклеиновой кислоты, которая кодирует рекомбинантный спидроиновый белок главной ампулярной железы по любому из пп.1-3, или последовательность нуклеиновой кислоты, комплементарную ей.

8. Рекомбинантная молекула нуклеиновой кислоты, содержащая последовательность нуклеиновой кислоты, которая кодирует слитый белок по п.4, или последовательность нуклеиновой кислоты, комплементарную ей.

9. Способ получения растворимого слитого белка, описанного в п.4, включающий в себя стадии:

(i) экспрессии молекулы нуклеиновой кислоты, которая кодирует слитый белок по п.4, в подходящем хозяине; и

(ii) выделения растворимого слитого белка, полученного на стадии (i), если надо, включающая в себя удаление LPS и других пирогенов.

10. Способ получения полимера спидроинового белка главной ампулярной железы, описанного в любом из пп.1-3, включающий в себя стадии:

(i) предоставления раствора слитого белка по п.4 в жидкой среде,

(ii) добавления к указанной жидкой среде подходящего расщепляющего агента для расщепления слитого белка в участке узнавания расщепляющего агента, и получения благодаря этому спидроинового белка главной ампулярной железы;

(iii) вызывания полимеризации спидроинового белка главной ампулярной железы, полученного на стадии (ii) в жидкой среде; и, если надо, (iv) выделения полимера, полученного на стадии (iii) из указанной жидкой среды, если надо, включающее в себя удаление LPS и других пирогенов.

11. Способ по п.10 получения полимера спидроинового белка главной ампулярной железы, описанного в пп.1-3, в котором стадия (i) предоставления раствора слитого белка по п.4 состоит в:

(a) экспрессии молекулы нуклеиновой кислоты, которая кодирует слитый белок по п.4, в подходящем хозяине;

(b) выделении растворимого слитого белка, полученного на стадии (i), если надо,

включающего в себя удаление LPS и других пирогенов; и

(с) предоставлении раствора указанного растворимого слитого белка, полученного на стадии (и) в жидкой среде.

5 12. Способ по любому из пп.10 и 11, в котором указанная стадия (iii) вызывания полимеризации в жидкой среде спидроинового белка главной ампулярной железы, полученного на стадии (ii), далее включает в себя создание границы раздела между указанной жидкой средой и другой фазой, выбранной из группы, состоящей из газовой фазы, жидкой фазы и твердой фазы, где указанная полимеризация инициируется на 10 указанной границе раздела или в области, окружающей указанную границу раздела.

13. Способ по п.12, где указанная жидкая среда представляет собой водную среду, и указанная другая фаза выбрана из группы, состоящей из органических растворителей, несмешивающихся с воздухом и водой.

14. Способ культивирования эукариотических клеток, включающий стадию 15 выращивания указанных эукариотических клеток, за исключением эмбриональных клеток человека, на материале, который включает полимер по п.5.

15. Способ по п.14, где указанный материал включает волокна указанного полимера.

20

25

30

35

40

45

50

СПИСОК ПОСЛЕДОВАТЕЛЬНОСТЕЙ

- <110> Johansson, Jan
Grip, Stefan
Hjältn, Göran
Rising, Anna
Stark, Margareta
Engström, Wilhelm
Hedhammar, My
- <120> БЕЛКИ ШЕЛКА ПАУКА И СПОСОБЫ ПОЛУЧЕНИЯ БЕЛКОВ
ШЕЛКА ПАУКА
- <130> PC-21028983
- <150> SE 0502932-7
- <151> 2005-12-30
- <160> 16
- <170> PatentIn version 3.3
- <210> 1
- <211> 3766
- <212> ДНК
- <213> Euprosthenops australis
- <400> 1
- | | |
|---|------|
| gtcaaggtgc tggaggtaat gccgctgcag cagccgcagc agcagcagca gcagcagctg | 60 |
| gacagggcgg tcaaggtgga tatggtggac taggtcaagg aggatatgga caggggtgcag | 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 taatgccgct gcagcagccg | 420 |
| cagcagctgc agcagcagca gctggacaag gaggtcaagg aggacaaggt ggactaggtc | 480 |
| aaggaggata tggacaaggt gcaggaagtt ctgcagccgc cgcgcagca tcagcagccg | 540 |
| ccgcagcagc tggacgaggt caaggaggat atggtcaagg tgctggagggt aatgccgctg | 600 |
| cagcagccgc agcagctgcc gccgccgcag cagctggaca gggagggtcaa ggtggatatg | 660 |
| gtggactagg tcaaggagga tatggacaag gtgcaggaag ttctgcagcc gctgccgccg | 720 |
| cagcagcagc agccgccgcc gcaggtggac aagggtggaca aggtcaagga agatatggac | 780 |
| aagggtgcagg aagttctgca gccgctgccg ccgcagcagc agcagccgcc gcagcagctg | 840 |
| gacaaggtca agggaggatat ggtcaagggt ctggaggtaa tgccgctgca gcagccgcag | 900 |
| cagctgcagc agcagcagct ggacaaggag gtcaaggagg acaagggtgga ctaggtcaag | 960 |
| gaggatatgg acaagggtgca ggaagttctg ccgccgccgc cgcagcagca gcagccgccg | 1020 |
| cagcagctgg acgagggtcaa ggaggatatg gtcaagggtgc tggaggtaat gccgctgcag | 1080 |
| cagccgcagc agctgccgaa gccgcagcag ctggacaggg tggtaagggt ggatatggtg | 1140 |
| gactaggtca agggaggatat 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	acgaggtcaa	ggaggatatg	gtcaagggtc	tggaggtaat	gccgctgcag	1560
cagccgcagc	agctgcagca	gcagcatctg	gacaagggaag	tcaaggagga	caagggtggac	1620
aagggtcaagg	tggatatgga	caagggtgcag	gaagttctgc	agccgccgcc	gccgccgcag	1680
cagcagccgc	cgcagcatct	ggacgaggtc	aaggaggata	tgggtcaagg	gctggaggta	1740
atgccgctgc	tgcagccgca	gcagctgccg	ccgccgcagc	agctggacag	ggcgggtcaag	1800
gtggatatgg	tggactaggt	caaggaggat	atggacaagg	tgcaggaagt	tctgcagccg	1860
ctgccgccgc	cgcagcagcc	gccgcagcag	gtggacaagg	tggacaagg	caaggaggat	1920
atggacaagg	tgcaggaagt	tctgcagccg	ccgccgcagc	agcagcagca	gcagccgcag	1980
cagctggacg	aggtcaagga	ggatatggtc	aaggttctgg	aggtaatgcc	gctgcagcag	2040
ccgcagcagc	tgcagcagca	gcattctggac	aaggagggtca	aggaggacaa	ggtggacaag	2100
gtcaagggtg	rtatggacaa	ggtgcaggaa	gttctgcagc	cgccgccgcc	gcagcagcag	2160
cagccgccgc	agcagctgga	caagggtcaag	gaggatatgg	tcaagggtgct	ggaggtaatg	2220
ccgctgcagc	agccgcagca	gctgcagcag	cagcagctgg	acaaggagggt	caaggaggac	2280
aagggtggact	aggtcaagga	ggatatggac	aaggtgcagg	aagttctgca	gccgccgccg	2340
cagcmgcmgc	agcagccgcc	gcagcagctg	gacgaggtca	aggaggatat	ggtcaagggtg	2400
ttggaggtaa	tgccgtgca	gcagccgcag	cagctgcagc	agcagcagct	ggacaaggag	2460
gtcaaggagg	acaagggtgga	ctagggtcaag	gaggatatgg	acaagggtgca	ggtagttctg	2520
cagccgccgc	cgccgccgca	gcagcagccg	ccgcagcagc	tggacgaggt	caaggaggat	2580
atggtcaagg	ttctggaggt	aatgccgctg	cagcagccgc	agcagctgca	gcagcagcat	2640
ctggacaagg	aagtcaagga	ggacaagggtg	gacaagggtca	aggtggatat	ggacaagggtg	2700
caggaagttc	tgcagccgcc	gccgccgcag	cagcagcagc	cgccgcagca	tctggacgag	2760
gtcaaggagg	atatggtcaa	ggtgctggag	gtaatgccgc	tgctgcagcc	gcagcagctg	2820
ccgccgccgc	agcagctgga	cagggcggtc	aaggtggata	tggaggacta	ggtcaaggag	2880
gatatggaca	aggtgcagga	agttctgcag	ccgctgccgc	cgccgcagca	gccgccgcag	2940
caggtggaca	aggtggacaa	ggtcaaggag	gatatggaca	aggttcagga	ggttctgcag	3000
ccgccgccgc	cgccgcagca	gcagcagcag	ctgcagcagc	tggacgaggt	caaggaggat	3060
atggtcaagg	ttctggaggt	aatgctgctg	ccgcagccgc	tgccgccgcc	gccgccgctg	3120
cagcagccgg	acagggaggt	caagggtggat	atggttagaca	aagccaagggt	gctggttccg	3180
ctgctgctgc	tgctgctgct	gctgccgctg	ctgctgctgc	aggatctgga	caagggtggat	3240
acggtggaca	aggtcaagga	ggttatgggtc	agagtagtgct	ttctgcttca	gctgctgctg	3300

cagctgctag tactgtagct aattcgggtga gtcgcctctc atcgccctcc gcagtatctc 3360
 gagtttcttc agcagtttct agcttggttt caaatgggtca agtgaatatg gcagcgttac 3420
 ctaatatcat ttccaacatt tcttctctcg tcagtgcacg tgctcctggg gcttctggat 3480
 gtgagggtcat agtgcaagct ctactcgaag tcatcactgc tcttggtcaa atcgttagtt 3540
 cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat gttgttgcta 3600
 atgccatggc tcaagtaatg ggctgagggt tttaatagta aaagggtgtga tattcctcaa 3660
 tgttttgaaa attattaatc gaatttttac cttgtgtgct atcagatata aattgaagta 3720
 taataaataa atatttgcac tttcaaaaaa aaaaaaaaaa aaaaaa 3766

<210> 2
 <211> 1207
 <212> PRT
 <213> Euprosthonops australis

<400> 2

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

180	185	190
Gly Ala Gly 195	Gly Asn Ala Ala Ala 200	Ala Ala Ala Ala Ala 205
Ala Ala Ala 210	Gly Gln Gly 215	Gln Gly Gly Tyr Gly 220
Gly Gly Tyr 225	Gly Gln Gly 230	Ala Gly Ser Ser Ala 235
Ala Ala Ala Ala 245	Ala Ala Gly Gly 250	Gln Gly Gly Gln Gly 255
Arg Tyr Gly 260	Gln Gly Ala Gly Ser 265	Ser Ala Ala Ala Ala 270
Ala Ala Ala 275	Ala Ala Ala Ala 280	Gly Gln Gly Gln Gly 285
Gly Ala Gly 290	Gly Asn Ala Ala 295	Ala Ala Ala Ala Ala 300
Ala Ala Gly 305	Gln Gly Gly Gln Gly 310	Gly Gly Leu Gly Gln Gly 315
Gly Tyr Gly 325	Gln Gly Ala Gly Ser 330	Ser Ala Ala Ala Ala 335
Ala Ala Ala 340	Ala Ala Ala Gly Arg 345	Gln Gly Gly Tyr Gly 350
Ala Gly Gly 355	Asn Ala Ala Ala 360	Ala Ala Ala Ala Ala 365
Ala Ala Gly 370	Gln Gly Gly Gln Gly 375	Gly Gly Tyr Gly 380
Gly Tyr Gly 385	Gln Gly Ala Gly Ser 390	Ser Ala Ala Ala Ala 395
Ala Ala Ala 405	Ala Ala Gly Arg Gly 410	Gln Gly Gly Tyr Gly 415
Ala Gly Gly 420	Asn Ala Ala Ala 425	Ala Ala Ala Ala Ala 430
Ala Ala Gly 435	Gln Gly Gly Gln Gly 440	Gly Tyr Gly Gly 445
Gly Tyr Gly 450	Gln Gly Ala Gly 455	Ser Ser Ala Ala Ala 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 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 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
 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
 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
 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
 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 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
 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
 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> Euprostenops australis

<220>
 <221> Повтор
 <222> (7)..(19)

<220>
 <221> Повтор
 <222> (20)..(42)

<220>
 <221> Повтор
 <222> (43)..(56)

 <220>
 <221> Повтор
 <222> (57)..(70)

 <220>
 <221> Повтор
 <222> (71)..(83)

 <220>
 <221> Повтор
 <222> (84)..(106)

 <220>
 <221> Повтор
 <222> (107)..(120)

 <220>
 <221> Повтор
 <222> (121)..(134)

 <220>
 <221> Повтор
 <222> (135)..(147)

 <220>
 <221> Повтор
 <222> (148)..(170)

 <220>
 <221> Повтор
 <222> (171)..(183)

 <220>
 <221> Повтор
 <222> (184)..(197)

 <220>
 <221> Повтор
 <222> (198)..(211)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

 <220>

<221> Повтор
<222> (581)..(594)

<220>
<221> Повтор
<222> (595)..(617)

<220>
<221> Повтор
<222> (618)..(630)

<220>
<221> Повтор
<222> (631)..(647)

<220>
<221> Повтор
<222> (648)..(661)

<220>
<221> Повтор
<222> (662)..(675)

<220>
<221> Повтор
<222> (676)..(688)

<220>
<221> Повтор
<222> (689)..(711)

<220>
<221> Повтор
<222> (712)..(725)

<220>
<221> Повтор
<222> (726)..(739)

<220>
<221> Повтор
<222> (740)..(752)

<220>
<221> Повтор
<222> (753)..(775)

<220>
<221> Повтор
<222> (776)..(789)

<220>
<221> Повтор
<222> (790)..(803)

<220>
<221> Повтор
<222> (804)..(816)

<220>
<221> Повтор
<222> (817)..(839)

<220>
<221> Повтор
<222> (840)..(853)

<220>
<221> Повтор

<222> (854)..(867)

<220>

<221> Повтор

<222> (868)..(880)

<220>

<221> Повтор

<222> (881)..(903)

<220>

<221> Повтор

<222> (904)..(917)

<220>

<221> Повтор

<222> (918)..(931)

<220>

<221> Повтор

<222> (932)..(945)

<220>

<221> Повтор

<222> (946)..(968)

<220>

<221> Повтор

<222> (969)..(981)

<220>

<221> Повтор

<222> (982)..(998)

<220>

<221> Повтор

<222> (999)..(1013)

<220>

<221> Повтор

<222> (1014)..(1027)

<220>

<221> Повтор

<222> (1028)..(1042)

<220>

<221> Повтор

<222> (1043)..(1059)

<220>

<221> Повтор

<222> (1060)..(1073)

<220>

<221> Повтор

<222> (1074)..(1092)

<400> 3

Gln Gly Ala Gly Gly Asn Ala 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 Ala

35	40	45
Ala 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 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		
Ala 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 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
 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 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 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 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 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 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 Ala
 580 585 590

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

Ala Ala Ala Ala Ala Gly Gln Gly Gln 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

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
 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 Ser
 865 870 875 880
 Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln 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 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
 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
 Val Gln Ile Val 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> Искусственный

<220>
 <223> Консенсусная последовательность, полученная из внутренних повторов
 Euprosthonops australis masp1

<220>
 <221> Вариант
 <222> (4)..(4)
 <223> Ser

<220>
 <221> Вариант
 <222> (8)..(8)
 <223> Tyr

<220>
 <221> Вариант
 <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
 20

<210> 6
 <211> 17
 <212> PRT
 <213> Искусственный

<220>
 <223> Консенсусная последовательность, полученная из внутренних повторов
Euprosthénops australis MaSp1

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

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

<220>
 <221> Вариант
 <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> Искусственный

<220>
 <223> Консенсусная последовательность, полученная из внутренних повторов
Euprosthénops australis MaSp1

<220>
 <221> Вариант
 <222> (2)..(2)
 <223> Gln

<220>
 <221> Вариант
 <222> (6)..(6)
 <223> Arg

<220>
 <221> Вариант
 <222> (11)..(11)
 <223> Ser

<220>
 <221> Вариант
 <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> Искусственный

 <220>
 <223> Консенсусная последовательность, полученная из известных белков MaSp1 и MaSp2

 <220>
 <221> MISC_FEATURE
 <222> (1)..(71)
 <223> Длина последовательности, присутствующая в известных видовых вариантах

 <220>
 <221> Вариант
 <222> (7)..(7)
 <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> Искусственный

<220>
 <223> Последовательность, полученная из *Euprosthonops australis* masp1

<220>
 <221> Домен
 <222> (1)..(171)
 <223> Повторяющийся фрагмент

<220>
 <221> Повтор
 <222> (11)..(30)

<220>
 <221> Повтор
 <222> (31)..(42)

<220>
 <221> Повтор
 <222> (43)..(60)

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

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

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

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

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

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

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

 <220>
 <221> Домен
 <222> (172)..(269)
 <223> С-концевой фрагмент

 <400> 9

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 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
 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
 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> Искусственный

<220>
<223> Гибридная последовательность, полученная из *Euprosthenops australis* and *Euprosthenops* sp MaSp1

<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

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> Искусственный

<220>
<223> Гибридная последовательность, полученная из *Euprosthenops australis* and *Euprosthenops* sp MaSp1

<400> 12

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

<220>
 <223> Гибридная последовательность, полученная из *Euprosthonops australis* and *Euprosthonops* sp MaSp1

<220>
 <221> Повтор
 <222> (1)..(11)

<220>
 <221> Домен
 <222> (1)..(180)
 <223> Повторяющийся фрагмент

<220>
 <221> Повтор
 <222> (12)..(25)

<220>
 <221> Повтор
 <222> (26)..(45)

<220>
 <221> Повтор
 <222> (46)..(60)

<220>
 <221> Повтор
 <222> (61)..(74)

<220>
 <221> Повтор
 <222> (75)..(88)

<220>
 <221> Повтор
 <222> (89)..(111)

<220>
 <221> Повтор
 <222> (112)..(119)

<220>
 <221> Повтор
 <222> (120)..(149)

<220>
 <221> Повтор
 <222> (150)..(158)

<220>
 <221> Повтор
 <222> (159)..(172)

<220>
 <221> Повтор
 <222> (173)..(180)

<220>
 <221> Домен
 <222> (181)..(276)
 <223> С-концевой фрагмент

<400> 13

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
 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 Glu Ala
275

<210> 14
<211> 269
<212> PRT
<213> Искусственный

<220>
<223> Последовательность, полученная из *Euprosthenops australis* masp1

<220>
<221> Домен
<222> (1) .. (171)
<223> Повторяющийся фрагмент

<220>
<221> Повтор
<222> (11) .. (30)

<220>
<221> Повтор
<222> (31) .. (42)

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

<220>
<221> Повтор
<222> (43) .. (60)

<220>
<221> Повтор
<222> (61) .. (75)

<220>
<221> Повтор
<222> (76) .. (89)

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

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

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

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

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

<220>
<221> Домен
<222> (172) .. (269)
<223> С-концевой фрагмент

<220>
<221> Домен

<222> (172)..(269)

<223> C-концевой фрагмент

<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 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

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> Искусственный
- <220>
 <223> Последовательность, полученная из *Euprosthenops australis* masp1
- <220>
 <221> Домен
 <222> (1)..(171)
 <223> Повторяющийся фрагмент
- <220>
 <221> Повтор
 <222> (11)..(30)
- <220>
 <221> Повтор
 <222> (31)..(42)
- <220>
 <221> Повтор
 <222> (43)..(60)
- <220>
 <221> Повтор
 <222> (61)..(75)
- <220>
 <221> Повтор
 <222> (76)..(89)
- <220>
 <221> Повтор
 <222> (90)..(104)
- <220>
 <221> Повтор
 <222> (105)..(121)
- <220>
 <221> Повтор
 <222> (122)..(135)
- <220>
 <221> Мутаген
 <222> (128)..(129)
- <220>
 <221> Повтор
 <222> (136)..(152)
- <220>
 <221> Повтор
 <222> (153)..(171)
- <220>
 <221> Домен
 <222> (172)..(269)
 <223> С-концевой фрагмент
- <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 Ala Cys
 115 120 125
 Cys 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> 16

<211> 269
 <212> PRT
 <213> Искусственный

<220>
 <223> Последовательность, полученная из *Euprosthonops australis* masp1

<220>
 <221> Домен
 <222> (1) .. (171)
 <223> Повторяющийся фрагмент

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

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

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

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

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

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

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

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

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

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

<220>
 <221> Домен
 <222> (172) .. (269)
 <223> С-концевой фрагмент

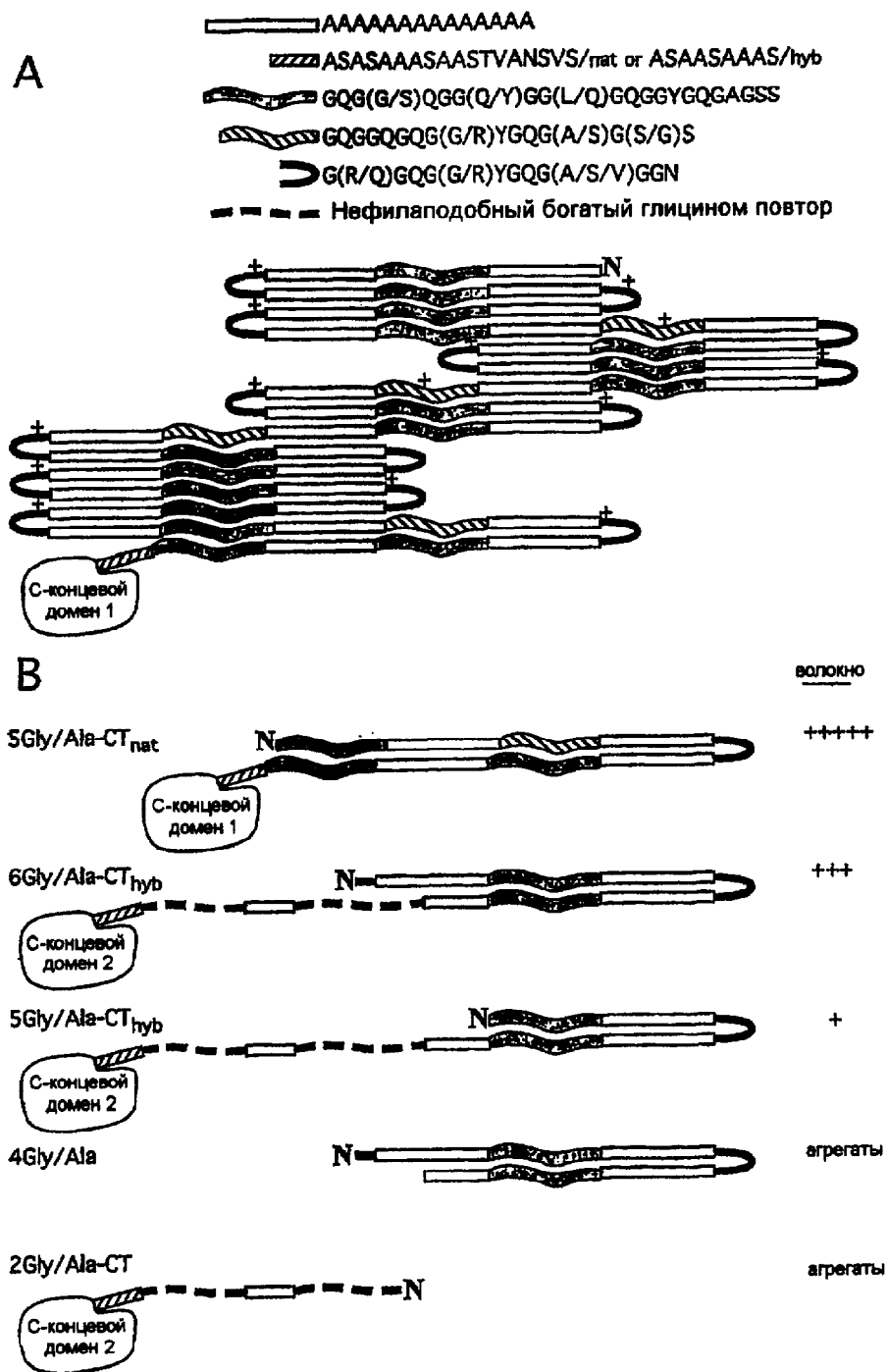
<220>
 <221> Мутаген
 <222> (222) .. (222)

<400> 16

Gly ser Ala met Gly tyr Leu trp ile Gln Gly Gln Gly Gly tyr Gly
 1 5 10 15

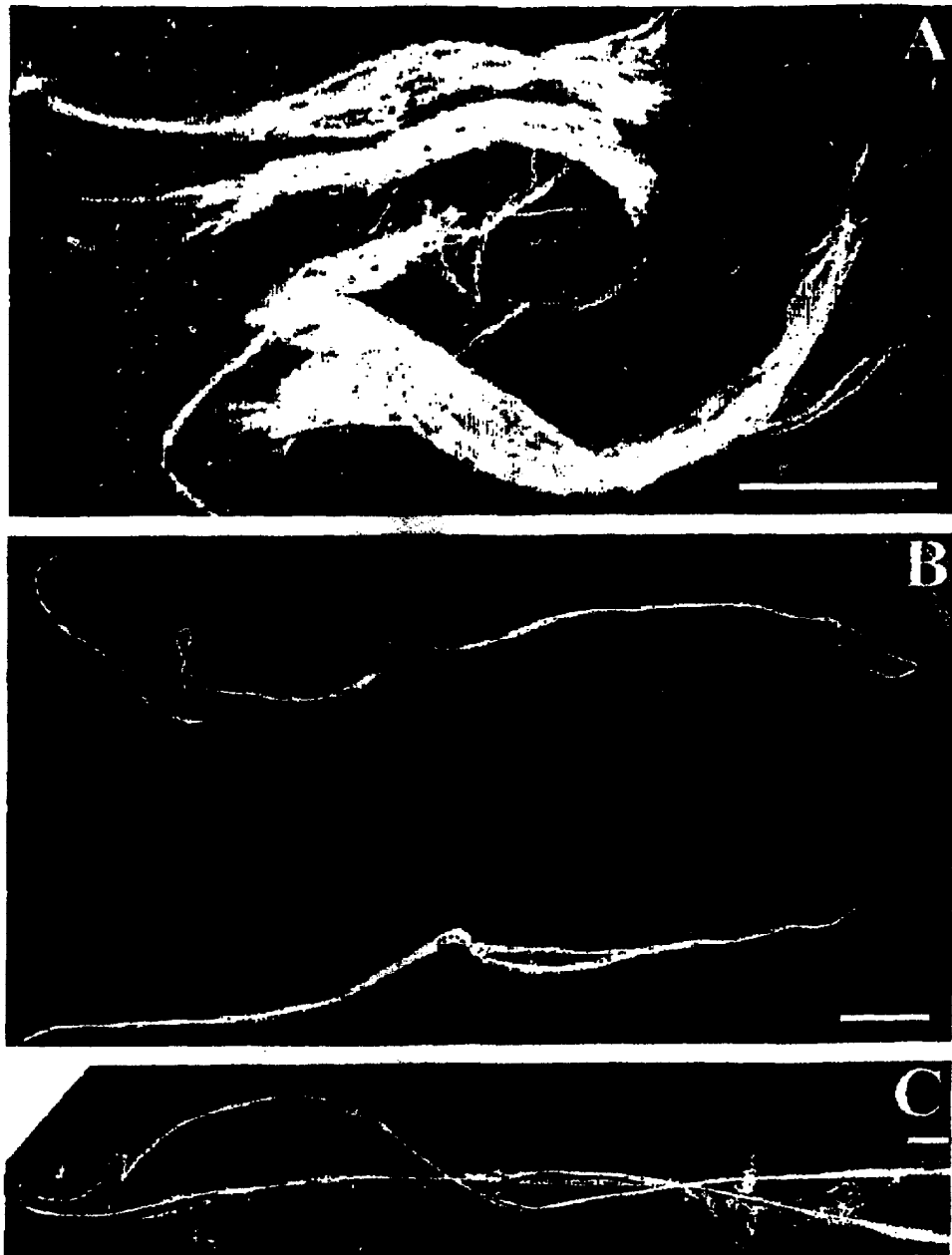
Gly Leu Gly Gln Gly Gly tyr Gly Gln Gly Ala Gly Ser ser Ala Ala

20	25	30
Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala 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
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 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
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
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	

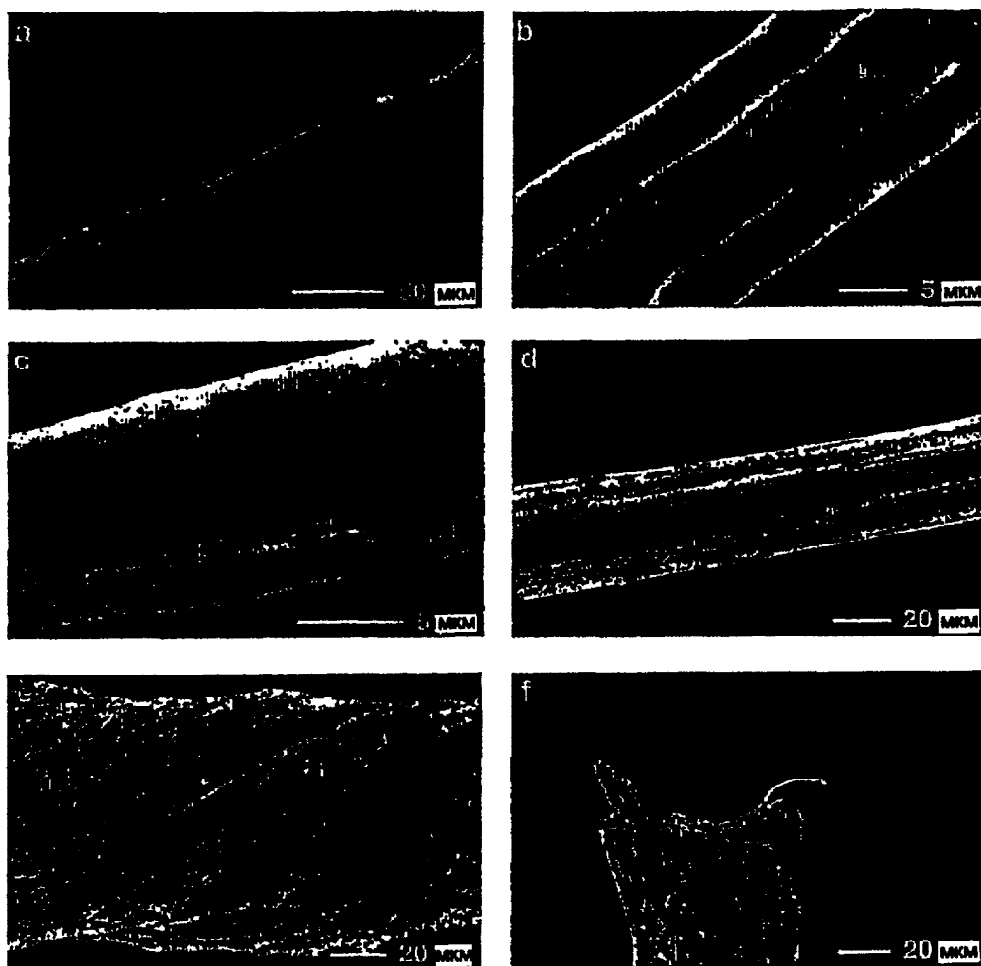


CTHb_Eop	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
CTHst_Eau	SRLSPHASS	KVSSAVSHLV	SSG-QVMDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359246_Acl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666062_Cml	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359273_Lgl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY959074_Lbl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666068_Mhl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
U28329_Mcl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666070_Mpl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359277_Nml	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359279_Nml	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666097_Orl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666094_Prl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359285_Tld	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359286_Tvl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
ABU28328_Akl	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY359016_Aml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359283_Aml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359287_Al2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359272_Qml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359278_Lgl2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY959075_Lgl2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AY666283_Nml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359278_Nml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359280_Nml2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359269_DlFl1	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
AF359270_DlFl2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
U47853_ADF1	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
U47854_ADF2	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
U47855_ADF3	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
U47856_ADF4	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F
Consensus	SRLSPHASS	KVSSAVSHLV	SSG-PTSDAA	LSSTISSVVS	QIGASHPGLS	CCDVLVQALL	EYVVALTHIL	GSSSIGQVNY	GSAGQATQIV	QGSVTOALCE	F

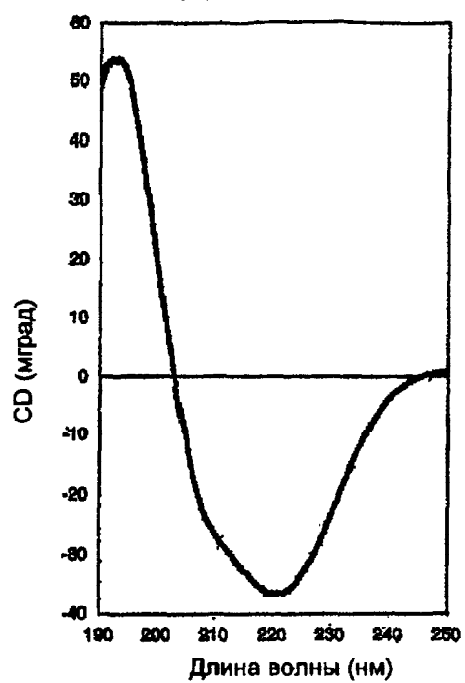
Фиг.3



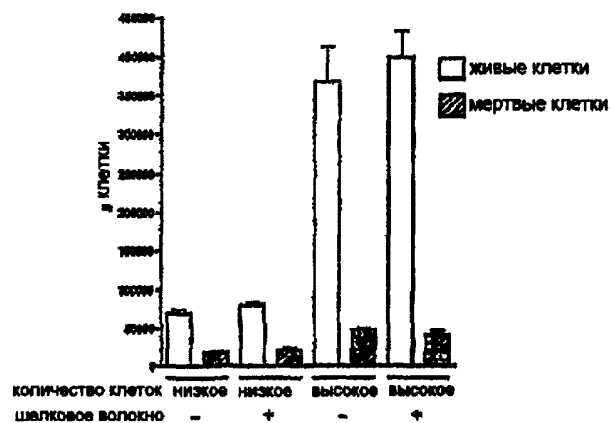
Фиг.4



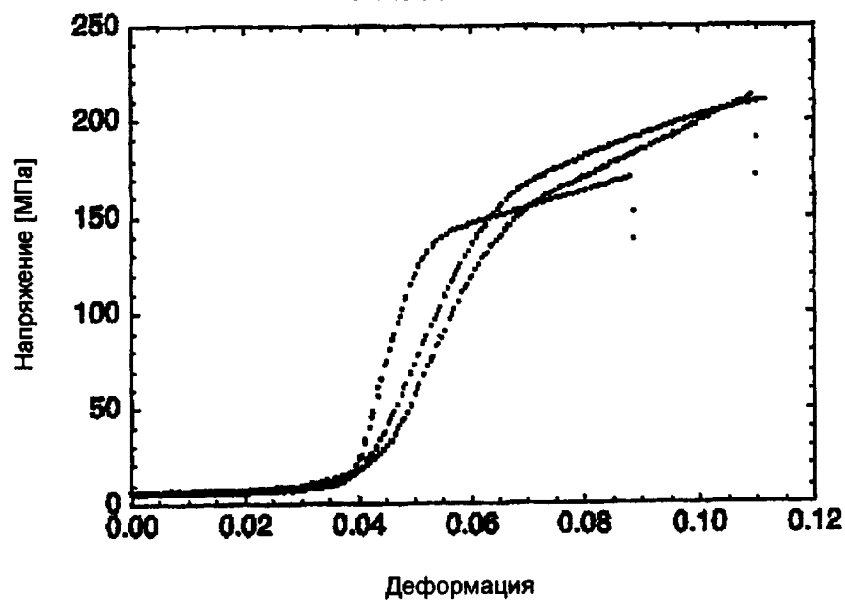
Фиг.5



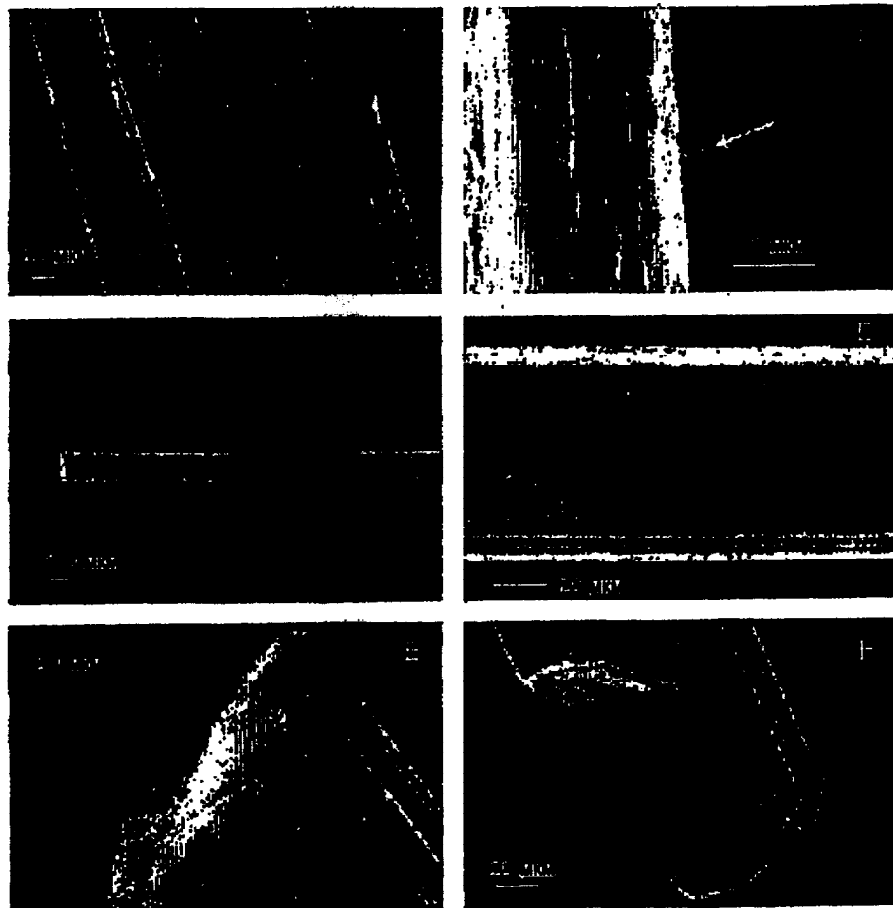
Фиг.6



Фиг.7



Фиг.9



Фиг.10