

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 533 555**

51 Int. Cl.:

**C07K 14/435** (2006.01)

**C12N 15/62** (2006.01)

**D01F 4/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.12.2006 E 06835913 (2)**

97 Fecha y número de publicación de la concesión europea: **18.02.2015 EP 1976868**

54 Título: **Proteínas de seda de araña y métodos para producir proteínas de seda de araña**

30 Prioridad:

**30.12.2005 SE 0502932**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**13.04.2015**

73 Titular/es:

**SPIBER TECHNOLOGIES AB (100.0%)  
KTH Avd Proteinteknologi, Alba Nova Univers  
Centrum  
106 91 Stockholm , SE**

72 Inventor/es:

**JOHANSSON, JAN;  
HJÄLM, GÖRAN;  
STARK, MARGARETA;  
RISING, ANNA;  
GRIP, STEFAN;  
ENGSTRÖM, WILHELM y  
HEDHAMMAR, MY**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

Observaciones :

**Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes**

**ES 2 533 555 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Proteínas de seda de araña y métodos para producir proteínas de seda de araña

**Campo técnico de la invención**

5 La presente invención se refiere al campo de la producción recombinante de proteínas. Más específicamente, la presente invención tiene que ver con la producción recombinante de proteínas de seda de araña. La presente invención proporciona proteínas espidroínas de glándula ampulácea mayor aisladas y proteínas de fusión de espidroína de ampulácea mayor novedosas, así como métodos y moléculas de ácido polinucleico para producir tales proteínas. También se proporcionan polímeros de proteínas espidroínas de ampulácea mayor y métodos para producir tales polímeros.

**10 Antecedentes de la invención**

Las sedas de araña son polímeros de la naturaleza de alto rendimiento, que obtienen una dureza extraordinaria debido a una combinación de resistencia y elasticidad. Existen hasta siete glándulas especializadas en las arañas, que producen una variedad de tipos de fibras de seda con diferentes propiedades mecánicas y funciones. La seda Dragline, producida por la glándula ampulácea mayor, es la fibra más dura, y basándose en su peso supera a los materiales elaborados por el ser humano, tales como el acero de alta resistencia a la tensión y el Kevlar. Las propiedades de la seda dragline son atractivas en el desarrollo de nuevos materiales para fines médicos o técnicos.

15 La seda dragline consiste en dos polipéptidos principales, principalmente referidos como espidroínas de ampulácea mayor (MaSp) 1 y 2, pero ADF-3 y ADF-4 en *Araneus diadematus*. Estas proteínas tienen masas moleculares aparentes en el intervalo de 200-720 kDa, dependiendo de la edad de la muestra y de las condiciones del análisis, pero todavía no se ha informado sobre el gen completo de la seda de araña dragline. Las propiedades de los polipéptidos de la seda dragline se comentan en Huemmerich, D. et al. Novel assembly properties of recombinant spider dragline silk proteins. *Curr. Biol.* 14, 2070-2074 (2004). Las espidroínas de seda dragline conocidas están formadas por bloques altamente repetidos de segmentos ricos en alanina alternantes, formando láminas  $\beta$  cristalinas en la fibra, y segmentos ricos en glicina que son más flexibles y principalmente carecen de estructura ordenada. La región C-terminal no es repetitiva, está muy conservada entre especies, y adopta una conformación  $\alpha$ -helicoidal. La región N-terminal de las proteínas de la seda dragline no ha sido caracterizadas hasta hace muy poco, revelando un dominio N-terminal que está muy conservado entre las diferentes espidroínas, y también entre diferentes especies de arañas (Rising, A. et al. N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindrical spider silk proteins. *Biomacromolecules* 7, 3120-3124 (2006)).

20 Las propiedades mecánicas de la seda dragline varían entre especies; la seda dragline de *Euprosthenois sp* es más rígida, más fuerte (requiere más fuerza para romperse) y menos extensible que la seda dragline p. ej., de *Araneus diadematus* o *Nephila clavipes*. La seda dragline de *Euprosthenois sp* parece tener una mayor proporción de estructura en lámina  $\beta$  cristalina que la seda dragline de *Araneus diadematus*, muy probablemente debido a que la MaSp de *Euprosthenois sp* tiene el mayor contenido de polialanina entre todas las especies analizadas hasta ahora (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthenois sp.* (Pisauridae). *Comp Biochem Physiol B Biochem Mol Biol* 138, 371-376 (2004)).

25 Los intentos para producir sedas de araña artificiales han empleado fragmentos de genes naturales o sintéticos que codifican proteínas de seda dragline, puesto que no se ha informado todavía sobre el gen completo. Las proteínas de seda dragline recombinantes se han expresado en diversos sistemas incluyendo bacterias, levaduras, células de mamífero, plantas, células de insecto, gusanos de seda transgénicos y cabras transgénicas. Véanse p. ej. 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 producción 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); y Lazaris, A. et al. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-476 (2002).

El documento WO 2004/016651 (The University of York) describe secuencias de ácido nucleico que codifican partes repetitivas internas de proteínas MaSp1 de *Euprosthenois sp*. No se expresa la proteína.

30 Huemmerich, D. et al. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004) describen un gen sintético, "(AQ)<sub>12</sub>NR3", que codifica fragmentos ricos en Ala y ricos en Gly/Gln repetitivos y un fragmento no repetitivo, todos derivados de ADF3 de *Araneus*. El gen es expresado a una proteína soluble (59,8 kD, >528 aa), que se agrega pero no forma polímeros o fibras. El contenido de alanina de la proteína es de 10-15%.

35 El documento WO 03/057727 describe la expresión de polipéptidos de seda recombinantes solubles en líneas celulares de mamífero y animales. Un polipéptido de seda expresado (ADF-3; 60 kD, 652 aa) consiste en una unidad repetitiva y un dominio hidrófilo no repetitivo. Otro polipéptido de seda expresado (ADF-3 His; 63 kD, 677 aa)

consiste en una unidad repetitiva, un dominio hidrófilo no repetitivo, un epítipo c-myc y una etiqueta de seis Histidinas. La unidad repetitiva tiene un bajo contenido de Ala (10-20%). Los polipéptidos de seda obtenidos muestran una escasa solubilidad en medio acuoso y/o forman precipitados. Puesto que los polipéptidos de seda obtenidos no se polimerizan espontáneamente, se requiere centrifugación para obtener polímeros o fibras.

- 5 Varios factores complican la expresión de las proteínas de seda dragline. Debido a la naturaleza altamente repetitiva de los genes, y a la composición de aminoácidos restringida concomitante de las proteínas, se producen errores de transcripción y traducción. El agotamiento de las reservas de ARNt en los sistemas de expresión microbiana, con la subsiguiente traducción discontinua, que conduce a una terminación prematura de la síntesis de proteínas podría ser otra razón. Otras razones comentadas para el truncamiento de la síntesis de proteínas son la formación de la estructura secundaria del ARNm, y la recombinación de los genes. Se ha demostrado que los genes de MaSp nativos más grandes de 2,5 kb son inestables en anfitriones bacterianos. Adicionalmente, existen dificultades en el mantenimiento de las proteínas de seda recombinantes en forma soluble, puesto que los fragmentos de seda dragline obtenidos de forma natural como los copolímeros de bloques diseñados, especialmente las proteínas derivadas de MaSp1/ADF-4, se auto-ensamblan fácilmente en agregados amorfos, causando la precipitación y la pérdida de la proteína. Véase Huemmerich, D. et al. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43, 13604-13612 (2004) y Lazaris, A. et al. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295, 472-476 (2002).

El documento WO 03/020916 describe secuencias de ácido nucleico que codifican proteínas de seda de araña. Las proteínas forman espontáneamente agregados a una concentración elevada, y estos agregados se pueden reunir y centrifugar mecánicamente en fibras macroscópicas utilizando disolventes fuertemente polares.

Spohner, A. et al. The conserved C-termini contribute to the properties of spider silk proteins, *Biochem. Biophys. Res. Com.* 338, 897-902 (2005) proporcionan un análisis de los extremos C de las espidroínas derivadas de la glándula ampulácea mayor de arañas del género *Nephila* y comentan su posible influencia en la conformación y la solubilidad de las espidroínas nativas.

## 25 Compendio de la invención

Un objeto de la presente invención es proporcionar una proteína de seda de araña, que puede proporcionar fibras de seda de araña.

Otro objeto de la presente invención es proporcionar una proteína de seda de araña soluble en agua, que se pueda manipular fácilmente para autopolimerizarse en fibras a voluntad. Esto permite aplicaciones únicas, tales como cultivo de células eucarióticas sobre las fibras. Además, esta propiedad permite que todas las siguientes etapas se realicen en condiciones fisiológicas, lo que disminuye el riesgo de toxicidad y desnaturalización de las proteínas.

Otro objeto más de la presente invención es proporcionar fibras de una proteína de seda de araña novedosa.

Un objeto de la presente invención es proporcionar proteínas de seda de araña a gran escala, cuyas proteínas se pueden manipular fácilmente para autopolimerizarse en fibras a voluntad.

Es también un objeto de la invención proporcionar métodos para producir proteínas de seda y fibras de proteínas de seda de araña.

Para estos y otros objetos que resultarán evidentes a partir de la siguiente descripción, la presente invención proporciona de acuerdo con un aspecto una proteína espidroína de la ampulácea mayor aislada, en donde la proteína consiste en 150 a 420 residuos de aminoácido y se define mediante la fórmula **REP-CT**, en donde **REP** es un fragmento de proteína que tiene de 80 a 300 residuos de aminoácido, en donde dicho fragmento se selecciona del grupo de **L(AG)<sub>n</sub>L**, **L(AG)<sub>n</sub>AL**, **L(GA)<sub>n</sub>L**, **L(GA)<sub>n</sub>GL**, en donde n es un número entero de 4 a 8; cada segmento **A** individual es una secuencia de aminoácidos de 8 a 18 residuos de aminoácido, en donde de 0 a 3 de los residuos de aminoácido no son Ala, y los residuos de aminoácido restantes son Ala; cada segmento **G** individual es una secuencia de aminoácidos de 12 a 30 residuos de aminoácido, en donde al menos 40% de los residuos de aminoácido son Gly; y cada segmento **L** individual es una secuencia de aminoácidos conectora de 0 a 20 residuos de aminoácido; y **CT** es un fragmento de proteína que tiene de 70 a 120 residuos de aminoácido, cuyo fragmento es, o tiene al menos una identidad de 80% con, una secuencia de aminoácidos seleccionada del grupo que consiste en el SEQ ID NO: 4, los residuos de aminoácido 172-269 del SEQ ID NO: 9, y los residuos de aminoácido 172-269 del SEQ ID NO: 16.

La presente invención está basada en la identificación de un motivo de proteína, que es suficiente para formar fibras parecidas a la seda, y el uso de dicho motivo para la construcción de proteínas MaSp recombinantes, cuya producción es posible en anfitriones adecuados, tales como bacterias, preferiblemente *E. coli*.

En ciertas realizaciones de acuerdo con la invención, cada segmento **A** individual tiene al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1060-1073 del SEQ ID NO: 3; los residuos de aminoácido 31-42, 61-75, 90-104, 122-135 y 153-171 del SEQ ID NO: 9; los residuos de aminoácido 31-42 del SEQ ID NO: 14; y los residuos de aminoácido 122-135 del SEQ ID NO: 15. En realizaciones específicas, cada segmento A individual es una secuencia de aminoácidos seleccionada entre este grupo de secuencias de aminoácidos.

5 En algunas realizaciones de acuerdo con la invención, cada segmento G individual tiene al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1074-1092 del SEQ ID NO: 3; SEQ ID NO: 5; SEQ ID  
10 NO: 6; SEQ ID NO: 7; los residuos de aminoácido 11-30, 43-60, 76-89, 105-121 y 136-152 del SEQ ID NO: 9. En realizaciones específicas, cada segmento G individual es idéntico a una secuencia de aminoácidos seleccionada entre este grupo de secuencias de aminoácidos.

15 En ciertas realizaciones de acuerdo con la invención, dicho fragmento CT tiene al menos una identidad de 50% con el SEQ ID NO: 8 o al menos una identidad de 90% con una secuencia de aminoácidos seleccionada del grupo que consiste en el SEQ ID NO: 4, los residuos de aminoácido 172-269 del SEQ ID NO: 9 y los residuos de aminoácido 172-269 del SEQ ID NO: 16 así como cualquier secuencia de aminoácidos de la Fig. 3, en particular las secuencias de MaSp1 de la Fig. 3. En realizaciones específicas, dicho fragmento CT es una secuencia de aminoácidos seleccionada entre este grupo de secuencias de aminoácidos.

20 En ciertas realizaciones de acuerdo con la invención, el contenido de lipopolisacáridos (LPS) y otros pirógenos en la proteína espidroína de ampulácea mayor aislada es de 1 unidad de endotoxina (UE)/mg de proteína o inferior.

25 De acuerdo con otro aspecto, la presente invención proporciona una proteína de fusión aislada que consiste en un primer fragmento de proteína, que es una proteína espidroína de ampulácea mayor, y un segundo fragmento de proteína, en donde dicho segundo fragmento de proteína comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión, en donde dicho primer fragmento de proteína se acopla a través de dicho sitio de reconocimiento para el agente de escisión a dicho compañero de fusión.

30 La presente invención proporciona una proteína de fusión aislada seleccionada del grupo de **X-REP-CT**, y **REP-CT-X**, en donde **REP** y **CT** son fragmentos de proteína de acuerdo con la invención; y **X** es un fragmento de proteína que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión; en donde el fragmento de proteína **REP-CT** combinado se acopla a través de dicho sitio de reconocimiento para el agente de escisión a dicho compañero de fusión.

En ciertas realizaciones de acuerdo con la invención, el contenido de LPS y otros pirógenos en la proteína de fusión aislada es de 1 UE/mg de proteína o inferior.

35 De acuerdo con otro aspecto más, la presente invención proporciona un método para producir una proteína espidroína de ampulácea mayor de acuerdo con la invención, que comprende las etapas de: (i) proporcionar una disolución de una proteína de fusión de acuerdo con la invención en un medio líquido, (ii) añadir a dicho medio líquido un agente de escisión adecuado para lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y obtener de ese modo la proteína espidroína de ampulácea mayor; y opcionalmente (iii) aislar la proteína espidroína de ampulácea mayor obtenida en la etapa (ii) de dicho medio líquido.

40 La presente invención también proporciona un método para producir un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención, que comprende las etapas de: (i) proporcionar una disolución de una proteína de fusión de acuerdo con la invención en un medio líquido, (ii) añadir a dicho medio líquido un agente de escisión adecuado para lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y obtener de ese modo la proteína espidroína de ampulácea mayor; (iii) permitir que la proteína espidroína de ampulácea mayor obtenida en la etapa (ii) se polimerice en el medio líquido; y opcionalmente (iv) aislar el polímero obtenido en la etapa (iii) de dicho medio líquido. En un método preferido, dicha etapa (iii) comprende  
45 adicionalmente proporcionar una interfase entre dicho medio líquido y otra fase seleccionada del grupo que consiste en una fase gaseosa, una fase líquida y una fase sólida, en donde dicha polimerización se inicia en dicha interfase o en una región que rodea dicha interfase. En un método preferido, dicho medio líquido es un medio acuoso y dicha otra fase se selecciona del grupo que consiste en aire y disolventes orgánicos miscibles con agua.

50 De acuerdo con otro aspecto, la presente invención proporciona una molécula de poli(ácido nucleico) aislada que comprende una secuencia de ácido nucleico que codifica una proteína espidroína de ampulácea mayor de acuerdo con la invención, o su secuencia de ácido nucleico complementaria.

55 De acuerdo con otro aspecto más, la presente invención proporciona una molécula de poli(ácido nucleico) aislada que comprende una secuencia de ácido nucleico que codifica una proteína de fusión de acuerdo con la invención, o su secuencia de ácido nucleico complementaria.

Otro aspecto de la invención reside en un método para producir una proteína de fusión soluble de acuerdo con la invención, que comprende las etapas de: (i) expresar una molécula de poli(ácido nucleico) que codifica una proteína

de fusión soluble de acuerdo con la invención en un anfitrión adecuado; y (ii) aislar la proteína de fusión soluble obtenida en la etapa (i). Opcionalmente, dicha etapa (ii) para aislar la proteína de fusión soluble implica la eliminación de los LPS y otros pirógenos.

5 La presente invención también proporciona un método para producir una proteína espidroína de ampulácea mayor de acuerdo con la invención, que comprende las etapas de: (i) expresar una molécula de poli(ácido nucleico) que codifica una proteína de fusión soluble de acuerdo con la invención en un anfitrión adecuado; (ii) aislar la proteína de fusión soluble obtenida en la etapa (i); (iii) proporcionar una disolución de dicha proteína de fusión soluble obtenida en la etapa (ii) en un medio líquido, (iv) añadir a dicho medio líquido un agente de escisión adecuado para lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y obtener de ese modo la proteína espidroína de ampulácea mayor; y opcionalmente (v) aislar la proteína espidroína de ampulácea mayor obtenida en la etapa (iv) de dicho medio líquido. Adicionalmente de manera opcional, dicha etapa (ii) para aislar la proteína de fusión soluble, y opcionalmente la etapa (v) para aislar la proteína espidroína de ampulácea mayor, implica o implican la eliminación de los LPS y otros pirógenos.

15 La presente invención proporciona adicionalmente un método para producir un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención, que comprende las etapas de: (i) expresar una molécula de poli(ácido nucleico) que codifica una proteína de fusión soluble de acuerdo con la invención en un anfitrión adecuado; (ii) aislar la proteína de fusión soluble obtenida en la etapa (i); (iii) proporcionar una disolución de dicha proteína de fusión soluble obtenida en la etapa (ii) en un medio líquido, (iv) añadir a dicho medio líquido un agente de escisión adecuado para lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y obtener de ese modo la proteína espidroína de ampulácea mayor; (v) permitir que la proteína espidroína de ampulácea mayor obtenida en la etapa (iv) se polimerice en el medio líquido; y opcionalmente (vi) aislar el polímero obtenido en la etapa (v) de dicho medio líquido. En un método preferido, dicha etapa (v) comprende adicionalmente proporcionar una interfase entre dicho medio líquido y otra fase seleccionada del grupo que consiste en una fase gaseosa, una fase líquida y una fase sólida, en donde dicha polimerización se inicia en dicha interfase o en una región que rodea dicha interfase. En un método preferido, dicho medio líquido es un medio acuoso y dicha otra fase se selecciona del grupo que consiste en aire y disolventes orgánicos miscibles con agua.

25 De acuerdo con otro aspecto, la presente invención proporciona un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención. La presente invención también proporciona un polímero de una proteína espidroína de ampulácea mayor obtenible por medio de un método de acuerdo con la invención. En una realización preferida, dicho polímero es una fibra. En otras realizaciones preferidas, dicho polímero forma una estructura seleccionada del grupo que consiste en una espuma, un gel, una malla o una película.

30 De acuerdo con otro aspecto más, la presente invención proporciona un uso novedoso de un fragmento de proteína que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión para la fabricación de una proteína de fusión que comprende dicho fragmento de proteína acoplado a través de dicho sitio de reconocimiento para el agente de escisión a un fragmento de proteína de seda de araña. En realizaciones preferidas, dicho fragmento de proteína de seda de araña consiste en 150 a 420 residuos de aminoácido.

40 De acuerdo con un aspecto final, la presente invención proporciona una molécula de poli(ácido nucleico) aislada que comprende una secuencia de ácido nucleico seleccionada del grupo que consiste en el SEQ ID NO: 1 y secuencias de ácido nucleico que codifican los SEQ ID NOS: 2-16, o sus secuencias de ácido nucleico complementarias. La presente invención también proporciona el uso de la molécula de poli(ácido nucleico) aislada para la fabricación de un gen no natural que codifica una proteína de seda de araña.

### Breve descripción de los dibujos

La Fig. 1 es un alineamiento de los segmentos dentro de la parte repetitiva de la proteína MaSp1 de *Euprosthonops australis*, esto es SEQ ID NO: 3.

45 La Fig. 2A ilustra una organización estructural pronosticada esquemática de la parte repetitiva de la proteína MaSp1 de *Euprosthonops australis* (SEQ ID NO: 3).

La Fig. 2B ilustra las organizaciones estructurales pronosticadas, esquemáticas de las proteínas espidroínas construidas de acuerdo con los ejemplos 5-8 (SEQ ID NOS: 9-13).

La Fig. 3 es un alineamiento de regiones C-terminales de MaSp1 y MaSp2, que ilustra su naturaleza conservada.

50 La Fig. 4 ilustra las apariencias macroscópicas de las fibras formadas a partir de proteínas espidroínas construidas de acuerdo con los ejemplos 5-8. (A): fibras de proteína 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13), barra 0,5 cm; (B): fibras de proteína 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), barra 1 cm. (C): fibras de proteína 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), barra 1 cm.

55 La Fig. 5 muestra las microfotografías de la microscopía electrónica de barrido (SEM) de fibras formadas a partir de proteínas espidroínas construidas de acuerdo con los ejemplos 5-8. Fibras individuales (a) y en fase de gel (b, c) de 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13). Fibras de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), estiradas en metanol del 75%, secadas al

aire y aplicadas sobre portaobjetos para SEM (d, e, f). Fibra retorcida antes del secado al aire (e), extremo de fibra (f).

La Fig. 6 presenta un espectro de dicroísmo circular (CD) de fibra de 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13).

5 La Fig. 7 ilustra los resultados de un estudio de toxicidad de mastocitos de ratón, que muestra los números de células vivas y muertas después de tres días de cultivo en presencia o ausencia de fibras de seda producidas *in vitro*.

La Fig. 8 es un cuadro de células HEK293 después de la exposición a fibras de seda producidas *in vitro* en un estudio de biocompatibilidad.

10 La Fig. 9 es una curva de tensión-deformación que muestra la resistencia a la tracción de fibras doblemente estiradas de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9).

15 La Fig. 10 muestra microfotografías SEM de fibras recombinantes de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9). a, b, Fibras formadas espontáneamente. La imagen ampliada (b) muestra la subestructura fibrilar. La fibrilla pequeña que protruye (flecha) tiene una anchura de aproximadamente 300 nm. c-f, fibras después de dos ciclos de estiramiento-relajación. c y d muestra la misma fibra a diferentes aumentos. e muestra un extremo de fibra cortada, y f muestra un punto de rotura después de la prueba de resistencia a la tracción.

### Descripción detallada de la invención

20 La presente invención se basa generalmente en la identificación de un motivo de proteína espidroína, que es suficiente para la producción recombinante de fibras de seda de araña. El motivo se basa en la secuencia de aminoácidos deducida a partir de la clonación y secuenciación de un ADNc de espidroína 1 mayor parcial (MaSp1) de *Euprosthonops australis*. De ello se deduce que el ADNc de MaSp1 aislado es útil como punto de partida para la construcción de genes de espidroína novedosos, tales como los referidos en la presente memoria. Los polímeros que se forman a partir de las proteínas resultantes de los ADNc de espidroína novedosos son útiles por sus propiedades físicas, especialmente la combinación útil de alta resistencia, elasticidad y peso ligero. También son útiles por su capacidad para soportar la adherencia y el crecimiento celulares. Las propiedades de la seda dragline son atractivas en el desarrollo de nuevos materiales para fines médicos o técnicos. En particular, las sedas de araña de acuerdo con la invención son útiles en dispositivos médicos, tales como implantes y productos médicos, tales como sistemas de cierre de heridas, vendas, suturas, apósitos de heridas, y matrices para la creación de tejidos artificiales y la regeneración celular guiada. Las sedas de araña de acuerdo con la invención son también particularmente útiles para su uso como tejidos o telas, por ejemplo en paracaídas, prendas blindadas, cinturones de seguridad, etc.

30 El término "fibra" según se utiliza en la presente memoria hace referencia a polímeros que tienen un grosor de al menos 1  $\mu\text{m}$ , preferiblemente polímeros macroscópicos que son visibles para el ojo humano, esto es, que tienen un grosor de al menos 1  $\mu\text{m}$ , y tienen una longitud considerable en comparación con su grosor, preferiblemente por encima de 5 mm. El término "fibra" no abarca los agregados o precipitados no estructurados.

35 Los términos "proteínas espidroínas de ampulácea mayor", "proteínas espidroínas" se utilizan indistintamente en toda la descripción y abarcan todas las proteínas espidroínas de ampulácea mayor conocidas, abreviadas típicamente como "MaSp", o "ADF" en el caso de *Araneus diadematus*. Estas proteínas espidroínas de ampulácea mayor son generalmente de dos tipos, 1 y 2. Estos términos incluyen además las nuevas proteínas de acuerdo con la invención, definidas en las reivindicaciones adjuntas, y otras proteínas no naturales con un alto grado de identidad y/o similitud con las proteínas espidroínas de ampulácea mayor conocidas.

40 Los autores de la presente invención han utilizado el motivo de proteína espidroína identificado para la construcción de constructos génicos novedosos, que codifican proteínas espidroínas no naturales. Se ha encontrado que una proteína espidroína de ampulácea mayor que consiste en 150 a 420 residuos de aminoácido, es decir más de o igual a 150, preferiblemente más de o igual a 220, preferiblemente más de o igual a 250, y menos de o igual a 420, preferiblemente menos de o igual a 380 residuos de aminoácido, preferiblemente menos de o igual a 320 residuos de aminoácido, preferiblemente menos de o igual a 280 residuos de aminoácido, tal como 220-360 residuos de aminoácido, se pueden producir recombinantemente, p. ej. en bacterias u otros organismos de producción adecuados. Las proteínas espidroínas resultantes forman espontáneamente fibras de seda macroscópicas de acuerdo con la invención. Éste es un resultado sorprendente, puesto que las proteínas espidroínas de origen natural y las proteínas espidroínas formadoras de fibras, producidas de manera recombinante, conocidas previamente, son considerablemente más largas que las proteínas de acuerdo con la invención. Por otra parte, las proteínas espidroínas de origen natural y las proteínas espidroínas formadoras de fibras, producidas de manera recombinante, conocidas previamente tienden a contener un gran número de repeticiones internas y requieren el uso de centrifugación y/o disolventes corrosivos para su polimerización.

55 Se muestra aquí por primera vez que las proteínas espidroínas pueden formar fibras espontáneamente *in vitro*. Los datos presentados en la presente memoria también muestran que solo se necesita que esté presente una fracción de la secuencia de espidroína para dictar la formación de fibra. Además, un híbrido de especie que contiene un

dominio repetitivo de *Euprosthénops* y un dominio C-terminal no repetitivo de *Nephila* (véase el Ejemplo 6C) forma también fibras, indicando que el potencial de formación de fibra de este motivo es robusto.

5 En su aspecto general, la proteína espidroína de ampulácea mayor de acuerdo con la invención se define mediante la fórmula **REP-CT**. El fragmento de proteína **REP** y el fragmento de proteína **CT** están acoplados covalentemente, típicamente a través de un enlace peptídico.

10 El fragmento de proteína **REP** tiene un carácter repetitivo, alternando entre tramos ricos en alanina y tramos ricos en glicina. El fragmento **REP** contiene generalmente más de 80, tal como más de 140, y menos de 300, preferiblemente menos de 240, tal como menos de 200, residuos de aminoácido, y puede dividirse en varios segmentos **L** (conector), segmentos **A** (ricos en alanina) y segmentos **G** (ricos en glicina), como se explicará con más detalle más adelante. Típicamente, dichos segmentos conectores, que son opcionales, están localizados en los extremos del fragmento **REP**, mientras que los segmentos restantes son a su vez ricos en alanina y ricos en glicina. De este modo, el fragmento **REP** puede tener generalmente cualquiera de las siguientes estructuras, en donde n es un número entero:

**L (AG)<sub>n</sub>L**, tal como **LA<sub>1</sub>G<sub>1</sub>A<sub>2</sub>G<sub>2</sub>A<sub>3</sub>G<sub>3</sub>A<sub>4</sub>G<sub>4</sub>A<sub>5</sub>G<sub>5</sub>L**;

15 **L(AG)<sub>nA</sub>L**, tal como **LA<sub>1</sub>G<sub>1</sub>A<sub>2</sub>G<sub>2</sub>A<sub>3</sub>G<sub>3</sub>A<sub>4</sub>G<sub>4</sub>A<sub>5</sub>G<sub>5</sub>A<sub>6</sub>L**;

**L(GA)<sub>n</sub>L**, tal como **LG<sub>1</sub>A<sub>1</sub>G<sub>2</sub>A<sub>2</sub>G<sub>3</sub>A<sub>3</sub>G<sub>4</sub>A<sub>4</sub>G<sub>5</sub>A<sub>5</sub>L**; o

**L(GA)<sub>n</sub>GL**, tal como **LG<sub>1</sub>A<sub>1</sub>G<sub>2</sub>A<sub>2</sub>G<sub>3</sub>A<sub>3</sub>G<sub>4</sub>A<sub>4</sub>G<sub>5</sub>A<sub>5</sub>G<sub>6</sub>L**.

20 De lo que se deduce que no es crítico que un segmento rico en alanina o un segmento rico en glicina esté adyacente a los segmentos conectores N-terminal o C-terminal. Se prefiere que n sea un número entero de 4 a 8, más preferido de 4 a 6, es decir n=4, n=5 o n=6.

25 En realizaciones preferidas, el contenido de alanina del fragmento **REP** de acuerdo con la invención está por encima de 20%, preferiblemente por encima de 25%, más preferiblemente por encima de 30%, y por debajo de 50%, preferiblemente por debajo de 40%, más preferiblemente por debajo de 35%. Esto es ventajoso, puesto que se contempla que un contenido de alanina más alto proporciona una fibra más rígida y/o más fuerte y/o menos extensible. Es probable que la razón para esto sea que un contenido más alto de alanina esté asociado con un contenido más alto de estructuras en lámina β en la fibra. De este modo, en una realización preferida, el contenido en lámina β en un polímero, tal como una fibra, de la proteína espidroína de ampulácea mayor de acuerdo con la invención está por encima de 50%, es decir más de 50% de la estructura secundaria de la proteína se encuentra en forma de lámina β.

30 En ciertas realizaciones, el fragmento **REP** carece de residuos de prolina, es decir no hay residuos de Pro en el fragmento **REP**.

35 Centrándose ahora en los segmentos que constituyen el fragmento **REP** de acuerdo con la invención, cabe destacar que cada segmento es individual, es decir dos segmentos **A** cualesquiera, dos segmentos **G** cualesquiera o dos segmentos **L** cualesquiera de un fragmento **REP** específico pueden ser idénticos o pueden no ser idénticos. De este modo, no es una característica general de la invención que cada tipo de segmento sea idéntico dentro de un fragmento **REP** específico. Más bien, la siguiente descripción proporciona a experto en la técnica pautas sobre cómo diseñar segmentos individuales y recogerlos en un fragmento **REP**, que es una porción de una proteína espidroína funcional de acuerdo con la invención.

40 Se ha concluido a partir de los datos experimentales presentados en la presente memoria que cada segmento **A** individual es una secuencia de aminoácidos que tiene de 8 a 18 residuos de aminoácido. Se prefiere que cada segmento **A** individual contenga de 13 a 15 residuos de aminoácido. También es posible que una mayoría, o más de dos, de los segmentos **A** contengan de 13 a 15 residuos de aminoácido, y que una minoría, tal como uno o dos, de los segmentos **A** contengan de 8 a 18 residuos de aminoácido, tal como 8-12 o 16-18 residuos de aminoácido. Una inmensa mayoría de estos residuos de aminoácido son residuos de alanina. Más específicamente, de 0 a 3 de los residuos de aminoácido no son residuos de alanina, y los residuos de aminoácido restantes son residuos de alanina. De este modo, todos los residuos de aminoácido en cada segmento **A** individual son residuos de alanina, sin excepción o con la excepción de uno, dos o tres residuos de aminoácido, que pueden ser cualquier aminoácido. Se prefiere que el aminoácido o los aminoácidos que remplazan a la alanina sea o sean aminoácidos de origen natural, preferiblemente seleccionados individualmente del grupo de serina, ácido glutámico, cisteína y glicina, más preferiblemente serina. Por supuesto, es posible que uno o más de los segmentos **A** sean segmentos con todo alanina, mientras que los segmentos **A** restantes contienen 1-3 residuos que no son alanina, tales como serina, ácido glutámico, cisteína o glicina.

55 En una realización preferida, cada segmento **A** contiene 13-15 residuos de aminoácido, incluyendo 10-15 residuos de alanina y 0-3 residuos que no son de alanina como se ha descrito anteriormente. En una realización más preferida, cada segmento **A** contiene 13-15 residuos de aminoácido, incluyendo 12-15 residuos de alanina y 0-1 residuos que no son de alanina como se ha descrito anteriormente.

Se prefiere que cada segmento **A** individual tenga al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1060-1073 del SEQ ID NO: 3. Cada secuencia de este grupo corresponde a un segmento de la secuencia de origen natural de la proteína MaSp1 de *Euprosthénops australis*, que es deducida de la clonación del ADNc correspondiente, véanse los Ejemplos 1-2 y la Fig. 1-2A. Alternativamente, cada segmento **A** individual tiene al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 31-42, 61-75, 90-104, 122-135 y 153-171 del SEQ ID NO: 9, residuos de aminoácido 31-42 del SEQ ID NO: 14 y residuos de aminoácido 122-135 del SEQ ID NO: 15. Cada secuencia de este grupo corresponde a un segmento de proteínas espidroínas no naturales expresadas de acuerdo con la invención, cuyas proteínas tienen la capacidad de formar fibras de seda en las condiciones apropiadas. Véanse los Ejemplos 5-8, 12 y la Fig. 2B. Si desear vincularse a ninguna teoría concreta, se contempla que los segmentos **A** de acuerdo con la invención formen estructuras helicoidales o láminas beta.

El término "% de identidad", según se utiliza a lo largo de la memoria descriptiva y las reivindicaciones adjuntas, se calcula de la siguiente manera. La secuencia problema se alinea con la secuencia diana utilizando el algoritmo CLUSTAL W (Thompson, J. D., Higgins, D. G. y Gibson, T. J., *Nucleic Acids Research*, 22: 4673-4680 (1994)). Los residuos de aminoácido de cada posición se comparan, y el porcentaje de posiciones en la secuencia problema que tienen correspondencias idénticas en la secuencia diana es referido como % de identidad.

El término "% de similitud", según se utiliza a lo largo de la memoria descriptiva y las reivindicaciones adjuntas, se calcula como se ha descrito para el "% de identidad", con la excepción de que los residuos hidrófobos Ala, Val, Phe, Pro, Leu, Ile, Trp, Met y Cys son similares; los residuos alcalinos Lys, Arg e His son similares; los residuos ácidos Glu y Asp son similares; y los residuos no cargados, hidrófilos Gln, Asn, Ser, Thr y Tyr son similares. El aminoácido natural Gly restante no es similar a ningún otro aminoácido en este contexto.

A lo largo de esta descripción, las realizaciones alternativas de acuerdo con la invención satisfacen, en lugar del porcentaje de identidad especificado, el porcentaje de similitud correspondiente. Otras realizaciones alternativas satisfacen el porcentaje de identidad especificado así como otro porcentaje de similitud, más alto, seleccionado del grupo de porcentajes de identidad preferidos para cada secuencia. Por ejemplo, una secuencia puede ser similar en 70% a otra secuencia; o puede ser idéntica en 70% a otra secuencia; o puede ser idéntica en 70% y similar en 90% a otra secuencia.

En realizaciones preferidas de acuerdo con la invención, cada segmento **A** individual tiene una identidad de al menos 90%, más preferiblemente 95%, lo más preferiblemente 100% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1060-1073 del SEQ ID NO: 3; los residuos de aminoácido 31-42, 61-75, 90-104, 122-135 y 153-171 del SEQ ID NO: 9; los residuos de aminoácido 31-42 del SEQ ID NO: 14; y los residuos de aminoácido 122-135 del SEQ ID NO: 15. De este modo, en ciertas realizaciones de acuerdo con la invención, cada segmento **A** individual es idéntico a una secuencia de aminoácidos seleccionada entre los segmentos de aminoácidos anteriormente mencionados.

Además, se llega a la conclusión a partir de los datos experimentales presentados en la presente memoria de que cada segmento **G** individual es una secuencia de aminoácidos de 12 a 30 residuos de aminoácido. Se prefiere que cada segmento **G** individual consista en 14 a 23 residuos de aminoácido. Al menos 40% de los residuos de aminoácido de cada segmento **G** son residuos de glicina. Típicamente el contenido de glicina de cada segmento **G** individual se encuentra en el intervalo de 40-60%.

Se prefiere que cada segmento **G** individual tenga al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1074-1092 del SEQ ID NO: 3. Cada secuencia de este grupo corresponde a un segmento de la secuencia de la proteína MaSp1 de *Euprosthénops australis* de origen natural, que es deducida de la clonación del ADNc correspondiente, véanse los Ejemplos 1-2 y la Fig. 1-2A. Alternativamente, cada segmento **G** individual tiene al menos una identidad de 80% con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 11-30, 43-60, 76-89, 105-121 y 136-152 del SEQ ID NO: 9. Cada secuencia de este grupo corresponde a un segmento de proteínas espidroínas no naturales, expresadas de acuerdo con la invención, cuyas proteínas tienen la capacidad de formar fibras de seda en las condiciones apropiadas. Véanse los Ejemplos 5-8, 12 y la Fig. 2B.

En realizaciones preferidas de acuerdo con la invención, cada segmento **G** individual tiene una identidad de secuencia de al menos 90%, más preferiblemente 95%, lo más preferiblemente 100%, con una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1074-1092 del SEQ ID NO: 3; los residuos de aminoácido 11-30, 43-60, 76-89, 105-121 y 136-152 del SEQ ID NO: 9. De este modo, en ciertas realizaciones de acuerdo con la invención, cada segmento **G** individual es idéntico a una secuencia de aminoácidos seleccionada entre los segmentos de aminoácidos anteriormente mencionados.
- En ciertas realizaciones, los dos primeros residuos de aminoácido de cada segmento **G** de acuerdo con la invención no son -Gln-Gln-.
- En ciertas realizaciones, la posición correspondiente al residuo de Tyr conservado (es decir correspondiente a la posición 16 en el SEQ ID NO: 5, la posición 10 en SEQ ID NO: 6 y la posición 7 en el SEQ ID NO: 7) no es Phe en ningún segmento **G** de acuerdo con la invención.
- En ciertas realizaciones, la posición correspondiente al residuo de Tyr conservado (es decir correspondiente a la posición 16 en el SEQ ID NO: 5, la posición 10 en el SEQ ID NO: 6 y la posición 7 en el SEQ ID NO: 7) es Tyr en cada segmento **G** de acuerdo con la invención.
- De ello se desprende que ciertas realizaciones de las proteínas de acuerdo con la invención presentan una combinación de las limitaciones anteriormente mencionadas, es decir los dos primeros residuos de aminoácido de cada segmento **G** de acuerdo con la invención no son -Gln-Gln-, y el residuo de Tyr conservado (es decir correspondiente a la posición 16 en el SEQ ID NO: 5, la posición 10 en el SEQ ID NO: 6 y la posición 7 en el SEQ ID NO: 7) es Tyr en cada segmento **G** de acuerdo con la invención. En ciertas realizaciones, las limitaciones anteriormente mencionadas, tomadas por separado o en cualquier combinación posible, se pueden combinar adicionalmente con la limitación de que el fragmento **REP** carece de residuos de prolina, como se ha comentado aquí anteriormente.
- Con referencia a las Figs. 1-2 y a los Ejemplos 3-4, hay tres subtipos de segmento **G** de acuerdo con la invención. Esta clasificación está basada en el análisis cuidadoso de la secuencia de la proteína MaSp1 de *Euprosthenois australis* (Figs. 1-2A), y la información se ha empleado y verificado en la construcción de proteínas de seda de araña no naturales, novedosas (Fig. 2B).
- El primer subtipo de segmento **G** de acuerdo con la invención está representado por la secuencia consenso de una letra de aminoácidos GQG(G/S)QGG(Q/Y)GG (L/Q)GQGGYGGQA GSS, como se muestra en la Fig. 2A y el SEQ ID NO: 5. Este primer subtipo de segmento **G**, y generalmente el más largo, contiene típicamente 23 residuos de aminoácido, pero puede contener tan pocos como 17 residuos de aminoácido, y carece de residuos cargados o contiene un residuo cargado. Así, se prefiere que este primer subtipo de segmento **G** contenga 17-23 residuos de aminoácido, pero se contempla que pueda contener tan pocos como 12 o tantos como 30 residuos de aminoácido. Sin desear vincularse a ninguna teoría concreta, se contempla que este subtipo forme estructuras enrolladas o estructuras en hélice 3<sub>1</sub>. Los segmentos **G** representativos de este primer subtipo son los residuos de aminoácido 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 y 1074-1092 del SEQ ID NO: 3; los residuos de aminoácido 11-30, 105-121 y 136-152 del SEQ ID NO: 9; y los residuos de aminoácido 26-45 y 89-111 del SEQ ID NO: 13. Los segmentos **G** alternativos de este primer subtipo son los residuos de aminoácido 120-149 y 159-172 del SEQ ID NO: 13. En ciertas realizaciones, los dos primeros residuos de aminoácido de cada segmento **G** de este primer subtipo de acuerdo con la invención no son -Gln-Gln-.
- El segundo subtipo de segmento **G** de acuerdo con la invención está representado por la secuencia consenso de una letra de aminoácidos GQGGQGG(G/R)Y GQG(A/S)G(S/G)S, como se muestra en la Fig. 2A y el SEQ ID NO: 6. Este segundo subtipo de segmento **G**, generalmente de tamaño medio contiene típicamente 17 residuos de aminoácido y carece de residuos cargados o contiene un residuo cargado. Se prefiere que este segundo subtipo de segmento **G** contenga 14-20 residuos de aminoácido, pero se contempla que pueda contener tan pocos como 12 o tantos como 30 residuos de aminoácido. Sin desear vincularse a ninguna teoría concreta, se contempla que este subtipo forme estructuras enrolladas. Los segmentos **G** representativos de este segundo subtipo son los residuos de aminoácido 249-265, 471-488, 631-647 y 982-998 del SEQ ID NO: 3; y los residuos de aminoácido 43-60 del SEQ ID NO: 9.
- El tercer subtipo de segmento **G** de acuerdo con la invención está representado por la secuencia consenso de una letra de aminoácidos G(R/Q)GQG(G/R)YGGG (A/S/V)GGN, como se muestra en la Fig. 2A y el SEQ ID NO: 7. Este tercer subtipo de segmento **G** contiene típicamente 14 residuos de aminoácido, y es generalmente el más corto de los subtipos de segmento **G** de acuerdo con la invención. Se prefiere que este tercer subtipo de segmento **G** contenga 12-17 residuos de aminoácido, pero se contempla que pueda contener tantos como 23 residuos de aminoácido. Sin desear vincularse a ninguna teoría concreta, se contempla que este subtipo forme estructuras con giros. Los segmentos **G** representativos de este tercer subtipo son los residuos de aminoácido 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 del SEQ ID NO: 3; los residuos de aminoácido 76-89 del SEQ ID NO: 9; y los residuos de aminoácido 61-74 del SEQ ID NO: 13. Un segmento **G** alternativo de este tercer subtipo son los residuos de aminoácido 1-11 del SEQ ID NO: 13.

De este modo, en realizaciones preferidas, cada segmento G individual tiene una identidad de secuencia de al menos 80%, preferiblemente 90%, más preferiblemente 95% con una secuencia de aminoácidos seleccionada entre el SEQ ID NO: 5, SEQ ID NO: 6 y el SEQ ID NO: 7.

5 En una realización preferida de la secuencia alternante de los segmentos **A** y **G** del fragmento **REP**, cada segundo segmento **G** es del primer subtipo, mientras que los segmentos **G** restantes son del tercer subtipo, p. ej. ...**A**<sub>1</sub>**G**<sub>corto</sub>**A**<sub>2</sub>**G**<sub>largo</sub>**A**<sub>3</sub>**G**<sub>corto</sub>**A**<sub>4</sub>**G**<sub>largo</sub>**A**<sub>5</sub>**G**<sub>corto</sub>... en otra realización preferida del fragmento **REP**, un segmento **G** del segundo subtipo interrumpe la regularidad del segmento **G** a través de una inserción, p. ej. ...**A**<sub>1</sub>**G**<sub>corto</sub>**A**<sub>2</sub>**G**<sub>largo</sub>**A**<sub>3</sub>**G**<sub>medio</sub>**A**<sub>4</sub>**G**<sub>corto</sub>**A**<sub>5</sub>**G**<sub>largo</sub>...

10 Cada segmento **L** individual representa una secuencia de aminoácidos conectora opcional, que puede contener de 0 a 20 residuos de aminoácido, tal como de 0 a 10 residuos de aminoácido. Si bien este segmento es opcional y no es funcionalmente crítico para la proteína espidroína, su presencia todavía permite proteínas espidroínas completamente funcionales, que forman fibras de seda de araña de acuerdo con la invención. También hay secuencias de aminoácidos conectoras presentes en la porción repetitiva (SEQ ID NO: 3) de la secuencia de aminoácidos deducida de la proteína MaSp1 de *Euprosthrops australis*. En concreto, la secuencia de aminoácidos de un segmento conector puede asemejarse a cualquiera de los segmentos **A** o **G** descritos, pero usualmente no satisface suficientemente sus criterios como se define en la presente memoria.

20 Como se muestra en la Fig. 2A, un segmento conector dispuesto en la porción C-terminal del fragmento **REP** puede estar representado por las secuencias consenso de una letra de aminoácidos ASASAAASAA STVANSVS y ASAASAAA, que son ricas en alanina. De hecho, se puede considerar que la segunda secuencia es un segmento **A** de acuerdo con la invención, mientras que la primera secuencia tiene un alto grado de similitud con los segmentos **A** de acuerdo con la invención. Otro ejemplo de un segmento conector de acuerdo con la invención tiene una secuencia de aminoácidos de una letra GSAMGQGS, que es rica en glicina y tiene un alto grado de similitud con los segmentos **G** de acuerdo con la invención.

25 Los segmentos **L** representativos son los residuos de aminoácido 1-6 y 1093-1110 del SEQ ID NO: 3; los residuos de aminoácido 1-10 y 153-171 del SEQ ID NO: 9; y los residuos de aminoácido 173-180 del SEQ ID NO: 13, pero el experto en la técnica advertirá fácilmente que hay muchas secuencias de aminoácidos alternativas para estos segmentos. En una realización del fragmento **REP** de acuerdo con la invención, uno de los segmentos **L** contiene 0 aminoácidos, es decir uno de los segmentos **L** está vacío. En otra realización del fragmento **REP** de acuerdo con la invención, ambos segmentos **L** contienen 0 aminoácidos, es decir ambos segmentos **L** están vacíos. De este modo, estas realizaciones de los fragmentos **REP** de acuerdo con la invención se pueden representar esquemáticamente de la siguiente manera: **(AG)**<sub>n</sub>**L**, **(AG)**<sub>n</sub>**AL**, **(GA)**<sub>n</sub>**L**, **(GA)**<sub>n</sub>**GL**; **L(AG)**<sub>n</sub>, **L(AG)**<sub>n</sub>**A**, **L(GA)**<sub>n</sub>, **L(GA)**<sub>n</sub>**G**; y **(AG)**<sub>n</sub>, **(AG)**<sub>n</sub>**A**, **(GA)**<sub>n</sub>, **(GA)**<sub>n</sub>**G**. Cualquiera de estos fragmentos **REP** es adecuado para su uso con cualquier fragmento **CT** como se define a continuación.

35 El fragmento C-terminal (**CT**) de la proteína espidroína de acuerdo con la invención tiene un alto grado de similitud con la secuencia de aminoácidos C-terminal de las proteínas espidroínas. Como se muestra en la Fig. 3, esta secuencia de aminoácidos está bien conservada entre varias especies y proteínas espidroínas, incluyendo MaSp1 y MaSp2. Se demuestra en los siguientes ejemplos que no es crítico qué fragmento **CT** específico esté presente en las proteínas espidroínas de acuerdo con la invención, con tal que el fragmento **CT** no se pierda por completo. Así, el fragmento **CT** de acuerdo con la invención se puede seleccionar a partir de cualquiera de las secuencias de aminoácidos mostradas en la Fig. 3 o secuencias con un alto grado de similitud. Es notable que el fragmento **CT**<sub>hyb</sub> del SEQ ID NO: 13 tenga una identidad de 96% con la secuencia de aminoácidos consenso del SEQ ID NO: 8, si bien el fragmento **CT**<sub>nat</sub> del SEQ ID NO: 9 presenta solo una identidad de 59% con la secuencia de aminoácidos consenso del SEQ ID NO: 8. Esto ilustra que se puede utilizar una amplia variedad de secuencias C-terminales en la proteína espidroína de acuerdo con la invención.

45 La secuencia del fragmento **CT** de acuerdo con la invención tiene al menos una identidad de 50%, preferiblemente una identidad de al menos 60%, con la secuencia de aminoácidos consenso del SEQ ID NO: 8, que está basada en la secuencia de aminoácidos de la Fig. 3. En una realización preferida, la secuencia del fragmento **CT** de acuerdo con la invención tiene una identidad de al menos 65%, preferiblemente una identidad de al menos 70%, con los residuos de aminoácido 1-71 de la secuencia de aminoácidos consenso del SEQ ID NO: 8. En realizaciones preferidas, el fragmento **CT** de acuerdo con la invención tiene además una identidad de secuencia de 70%, preferiblemente 80%, con la secuencia de aminoácidos consenso del SEQ ID NO: 8, o los residuos de aminoácido 1-71 de la misma.

55 Los fragmentos **CT** representativos de acuerdo con la invención son la secuencia de *Euprosthrops australis* SEQ ID NO: 4, los residuos de aminoácido 172-268 derivados de *Euprosthrops australis* SEQ ID NO: 9, que se alega que derivan de *Euprosthrops sp* (Pouchkina-Stantcheva, N. N. & McQueen-Mason, S. J. Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthrops sp.* (Pisauridae). Comp Biochem Physiol B Biochem Mol Biol 138, 371-376 (2004)), pero con un alto grado de similitud con MaSp1 de *Nephila clavipes* y *Nephila senegalensis*. De este modo, de acuerdo con un aspecto preferido de la invención, el fragmento **CT** tiene al menos una identidad de 80% con el SEQ ID NO: 4, los residuos de aminoácido 172-269 del SEQ ID NO: 9, los residuos de aminoácido 172-269 del SEQ ID NO: 16 o cualquier secuencia de aminoácidos de MaSp1/ADF-4 individual de la Fig.

60

3 y el Ejemplo 4. En aspectos preferidos de la invención, el fragmento **CT** tiene una identidad de al menos 90%, por ejemplo al menos 95%, con el SEQ ID NO: 4, los residuos de aminoácido 172-269 del SEQ ID NO: 9, los residuos de aminoácido 172-269 del SEQ ID NO: 16 o cualquier secuencia de aminoácidos de MaSp1/ADF-4 individual de la Fig. 3 y el Ejemplo 4. En aspectos preferidos de la invención, el fragmento **CT** es idéntico al SEQ ID NO: 4, los residuos de aminoácido 172-269 del SEQ ID NO: 9, los residuos de aminoácido 172-269 del SEQ ID NO: 16 o cualquier secuencia de aminoácidos de MaSp1/ADF-4 individual de la Fig. 3 y el Ejemplo 4.

El fragmento **CT** consiste típicamente en 70 a 120 residuos de aminoácido. Se prefiere que el fragmento **CT** contenga al menos 70, o más de 80, preferiblemente más de 90, residuos de aminoácido. También se prefiere que el fragmento **CT** contenga a lo sumo 120, o menos de 110 residuos de aminoácido. Un fragmento **CT** típico contiene aproximadamente 100 residuos de aminoácido.

De acuerdo con otro aspecto, la presente invención proporciona una proteína de fusión aislada que consiste en un primer fragmento de proteína, que es una proteína espidroína de ampulácea mayor, que consiste preferiblemente en 150 a 420 residuos de aminoácido, y un segundo fragmento de proteína, que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión. El primer fragmento de proteína se acopla a través del sitio de reconocimiento para el agente de escisión al compañero de fusión, es decir el compañero de fusión puede ser separado por escisión tratando la proteína de fusión con un agente de escisión adecuado en las condiciones apropiadas, proporcionando una proteína espidroína de ampulácea mayor, que consiste preferiblemente en 150 a 420 residuos de aminoácido. Una ventaja con esta proteína de fusión es que se pueden producir grandes cantidades de la misma en disolución, preferiblemente en un medio fisiológico, típicamente un medio acuoso tamponado, tal como un tampón Tris-HCl 10-100 mM, pH 6-9, sin causar precipitación ni otros problemas de producción cuando se producen en anfitriones adecuados, tales como bacterias, preferiblemente *E. coli*. Las proteínas de fusión en disolución son solubles durante largos períodos de tiempo, típicamente días o semanas, lo que facilita la producción a gran escala y disminuye el riesgo de agregación de las proteínas. Por los términos "soluble" y "en disolución" se quiere significar que la proteína no está visiblemente agregada y no precipita del disolvente a 60000×g. A voluntad, las proteínas de fusión en disolución se pueden someter a escisión utilizando un agente de escisión adecuado, proporcionando una proteína espidroína de ampulácea mayor que forma espontáneamente fibras de seda.

En un aspecto preferido, la presente invención proporciona una proteína de fusión aislada seleccionada del grupo de **X-REP-CT** y **REP-CT-X**, preferiblemente **X-REP-CT**. **REP** y **CT** son fragmentos de proteína de acuerdo con la invención, lo que implica que la proteína MaSp resultante de la forma **REP-CT** es una proteína MaSp de acuerdo con la invención. **X** es un fragmento de proteína que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión como se ha definido anteriormente. El fragmento combinado de proteína **REP-CT** se acopla a través del sitio de reconocimiento para el agente de escisión al compañero de fusión.

Los compañeros de fusión de acuerdo con la invención incluyen cualquier fragmento de proteína que mejora la solubilidad y/o la estabilidad de su fragmento de proteína compañero, aquí la proteína MaSp de acuerdo con la invención. El compañero de fusión también proporciona una sujeción adecuada para la purificación por afinidad. Sin estar limitado a esto; los ejemplos de los compañeros de fusión de acuerdo con la invención incluyen tiorredoxina, proteína de unión a maltosa, glutatión S-transferasa (GST), MTB32-C, Gbl, ZZ y Nus A. El experto en la técnica es muy consciente de los compañeros de fusión alternativos adecuados. En una realización preferida de la invención, el compañero de fusión es un radical tiorredoxina (ThrX) combinado con una etiqueta de His y una etiqueta de S. En una realización preferida de la invención, el compañero de fusión es un radical ThrX combinado con dos etiquetas de His, es decir etiqueta de His/ThrX/etiqueta de His. En otra realización preferida de la invención, el compañero de fusión es un radical tiorredoxina (ThrX).

El sitio de reconocimiento para el agente de escisión está situado en ese fragmento **X** de proteína terminal que está acoplado al fragmento de proteína MaSp, de manera que la escisión en el sitio de reconocimiento da como resultado una proteína MaSp y un compañero de fusión. Sin estar limitados a esto, los ejemplos del sitio de reconocimiento para el agente de escisión de acuerdo con la invención incluyen un sitio de reconocimiento de trombina que tiene la secuencia de aminoácidos LVPRGS (escinde entre R y G); un sitio de reconocimiento de enteroquinasa que tiene la secuencia de aminoácidos DDDK (escinde después de K); un sitio de reconocimiento de hidroxilamina que tiene la secuencia de aminoácidos NG (escinde entre N y G); un sitio de reconocimiento de proteasa HRV 3C que tiene la secuencia de aminoácidos LGVLFQGP (escinde entre Q y G); un sitio de reconocimiento de Factor Xa que tiene la secuencia de aminoácidos I(E/D)GR (escinde después de R); un sitio de reconocimiento de TEV que tiene la secuencia de aminoácidos EXXYXQ(G/S), comúnmente ENLYFQG (escinde entre Q y G/S), un sitio de reconocimiento de Ac-TEV que tiene la secuencia de aminoácidos EDNLYFQG (escinde entre Q y G); y un sitio de reconocimiento de PreScission que tiene la secuencia de aminoácidos LEVLFQGP (escinde entre Q y G). Otros sitios de reconocimiento adecuados son los sitios de escisión para tripsina, endoproteínasa, proteasa V8, pepsina y CNBr. Otros ejemplos de sitios de reconocimiento de escisión adecuados se encuentran al alcance de los expertos en la técnica. En una realización preferida de la invención, el sitio de reconocimiento para el agente de escisión es un sitio de reconocimiento de trombina.

En una realización preferida, el fragmento **X** de acuerdo con la invención tiene la estructura ThrX/etiqueta de His/etiqueta de S/sitio de escisión de trombina, y el fragmento **X** se acopla al extremo N del fragmento de proteína **REP-CT** de acuerdo con la invención.

En una realización preferida, el fragmento **X** de acuerdo con la invención tiene la estructura etiqueta de His/ThrX/etiqueta de His/sitio de escisión de trombina, y el fragmento **X** se acopla al extremo N del fragmento de proteína **REP-CT** de acuerdo con la invención.

5 De acuerdo con otro aspecto, la presente invención proporciona un método para producir una proteína espidroína de ampulácea mayor de acuerdo con la invención. En la primera etapa, se proporciona una disolución de una proteína de fusión de acuerdo con la invención en un medio líquido. Preferiblemente, la proteína de fusión no se agrega, y por lo tanto, no se requieren procedimientos de resolubilización. La proteína de fusión se puede producir de manera recombinante y purificar utilizando una sujeción de afinidad en la proteína de fusión, tal como una etiqueta de His o cualquier epítipo adecuado en la proteína de fusión. El medio líquido puede ser cualquier medio adecuado, preferiblemente un medio fisiológico, típicamente un medio acuoso tamponado, tal como un tampón Tris-HCl 10-100 mM, pH 6-9. En la segunda etapa, se añade un agente de escisión de acuerdo con la invención al medio líquido con el fin de lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión. Como se ha descrito anteriormente, la proteína espidroína de ampulácea mayor de acuerdo con la invención se obtiene de esta manera. En una tercera etapa opcional, la proteína espidroína de ampulácea obtenida de este modo se aísla del medio líquido utilizando técnicas de aislamiento adecuadas, tales como cromatografía y/o filtración.

De acuerdo con otro aspecto más, la presente invención proporciona un método para producir un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención. En la primera etapa, se proporciona una disolución de una proteína de fusión de acuerdo con la invención en un medio líquido. Preferiblemente, la proteína de fusión no se agrega, y por lo tanto, no se requieren procedimientos de resolubilización. La proteína de fusión se puede producir de manera recombinante y purificar utilizando una sujeción por afinidad en la proteína de fusión, tal como una etiqueta de His o cualquier epítipo adecuado en la proteína de fusión. El medio líquido puede ser cualquier medio adecuado, preferiblemente un medio fisiológico, típicamente un medio acuoso tamponado, tal como un tampón Tris-HCl 10-100 mM, pH 6-9. En la segunda etapa, se añade un agente de escisión de acuerdo con la invención al medio líquido con el fin de lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión. Como se ha descrito anteriormente, la proteína espidroína de ampulácea mayor de acuerdo con la invención se obtiene de esta manera. En la tercera etapa, se permite que la proteína espidroína de ampulácea mayor obtenida de este modo se polimerice en el medio líquido. La polimerización se inicia típicamente en la interfase entre dos fases diferentes, tales como las interfases líquido/aire, líquido/sólido, y agua/aceite. De este modo, esta tercera etapa puede comprender también adicionalmente proporcionar una interfase entre el medio líquido y otra fase. La otra fase se selecciona del grupo que consiste en una fase gaseosa, una fase líquida y una fase sólida. Como se ha detallado anteriormente, el medio líquido es típicamente un medio acuoso, y las otras fases adecuadas son por ejemplo aire y disolventes orgánicos no miscibles con agua, tales como aceite, p. ej. aceite mineral adecuado para reacciones de PCR. La presencia de la interfase resultante estimula la polimerización en la interfase o en la región que rodea la interfase, cuya región se extiende en el medio líquido, de manera que dicha polimerización se inicia en dicha interfase o en dicha región de interfase. Las interfases preferidas incluyen interfases de agua/aire y agua/aceite. La polimerización se produce típicamente de manera espontánea en el plazo de minutos o unas pocas horas, tal como en el plazo de 1 min a 5 h, después de la incubación a temperatura ambiente. En una cuarta etapa opcional, el polímero de la proteína espidroína de ampulácea mayor obtenido de este modo se aísla del medio líquido utilizando técnicas de aislamiento adecuadas.

40 Como se ha comentado anteriormente, la formación de fibra se induce por medio de liberación proteolítica de la espidroína en miniatura a partir de la proteína de fusión. Si la reacción de escisión se realiza en un tubo que es movido suavemente de lado a lado, se forma una fibra en la interfase de aire-agua a lo largo del tubo. El tubo se puede elaborar de cualquier material adecuado, tal como plástico o vidrio. Si se deja reposar la mezcla de escisión, se forma una película en la interfase de aire-agua. Si se añade aceite a la parte superior de la mezcla de escisión acuosa, se forma una película en la interfase de aceite-agua, tanto si se deja reposar como si se mueve. Si se forma espuma en la mezcla de escisión, p. ej. haciendo burbujear aire o batiendo, la espuma es estable y se solidifica si se deja secar.

Utilizando el método o los métodos de la presente invención, es posible producir de manera recombinante grandes cantidades de proteínas de fusión de acuerdo con la invención, que se pueden escindir y dejar polimerizar a voluntad. Esto proporciona un mejor control del procedimiento de polimerización y permite la optimización de los parámetros para obtener fibras de seda con propiedades deseables.

La proteína espidroína de ampulácea mayor de acuerdo con la invención es producida típicamente de manera recombinante utilizando una variedad de anfitriones adecuados. De acuerdo con otro aspecto, la presente invención proporciona por lo tanto una molécula de poli(ácido nucleico) aislada que comprende una secuencia de ácido nucleico que codifica una proteína espidroína de ampulácea mayor de acuerdo con la invención, o su secuencia de ácido nucleico complementaria, tal como los SEQ ID NOS: 9-13, preferiblemente los SEQ ID NOS: 9, 12 y 13. Estas moléculas de poli(ácido nucleico) así como las moléculas de poli(ácido nucleico) que codifican las diversas proteínas descritas en la presente memoria (SEQ ID NOS: 2-16) pueden ser también útiles en desarrollos adicionales de proteínas espidroínas no naturales o sistemas de producción para las mismas.

60 La proteína de fusión de acuerdo con la invención es producida típicamente de manera recombinante utilizando una variedad de anfitriones adecuados, tales como bacterias, levaduras, células de mamífero, plantas, células de

insectos, y animales transgénicos. Se prefiere que la proteína de fusión de acuerdo con la invención sea producida en bacterias.

5 De acuerdo con otro aspecto, la presente invención proporciona por lo tanto una molécula de poli(ácido nucleico) aislada que comprende una secuencia de ácido nucleico que codifica una proteína de fusión de acuerdo con la invención, o su secuencia de ácido nucleico complementaria. La molécula de poli(ácido nucleico) puede ser también útil en desarrollos adicionales de proteínas espidroínas no naturales o sistemas de producción para las mismas.

10 Las moléculas de poli(ácido nucleico) de acuerdo con la invención pueden ser moléculas de ADN, incluyendo moléculas de ADNc, o moléculas de ARN. Como bien saben los expertos en la técnica, una secuencia de ácido nucleico también puede ser descrita por su secuencia de ácido nucleico complementaria. Por lo tanto, las secuencias de ácido nucleico que son complementarias a las secuencias de ácido nucleico de acuerdo con la invención también están incluidas en el alcance protector de la invención.

15 De acuerdo con un aspecto, la presente invención proporciona un método para producir una proteína de fusión soluble de acuerdo con la invención. En la primera etapa, una molécula de poli(ácido nucleico) que codifica una proteína de fusión de acuerdo con la invención es expresada en un anfitrión adecuado. En la segunda etapa, se aísla la proteína de fusión soluble obtenida de este modo en la etapa, p. ej. utilizando cromatografía y/o filtración. Opcionalmente, dicha segunda etapa de aislamiento de la proteína de fusión soluble implica la eliminación de los LPS y otros pirógenos.

20 La presente invención proporciona adicionalmente un método para producir una proteína espidroína de ampulácea mayor de acuerdo con la invención. En la primera etapa, una molécula de poli(ácido nucleico) que codifica una proteína de fusión de acuerdo con la invención es expresada en un anfitrión adecuado. En la segunda etapa, se aísla la proteína de fusión soluble obtenida de este modo, p. ej. utilizando cromatografía y/o filtración. En la tercera etapa, se proporciona una disolución de la proteína de fusión aislada, y en la cuarta etapa, se añade un agente de escisión adecuado al medio líquido. Esto logra la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y proporciona de este modo la proteína espidroína de ampulácea mayor. En una quinta etapa opcional, la proteína espidroína de ampulácea mayor obtenida de este modo es aislada del medio líquido. Adicionalmente de manera opcional, dicha segunda etapa de aislamiento de la proteína de fusión soluble, y opcionalmente la quinta etapa de aislamiento de la proteína espidroína de ampulácea mayor, implica o implican la eliminación de los LPS y otros pirógenos.

30 La presente invención también proporciona un método para producir un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención. En la primera etapa, una molécula de poli(ácido nucleico) que codifica una proteína de fusión de acuerdo con la invención es expresada en un anfitrión adecuado. En la segunda etapa, se aísla la proteína de fusión soluble obtenida de este modo, p. ej. utilizando cromatografía y/o filtración. En la tercera etapa, se proporciona una disolución de la proteína de fusión aislada, y en la cuarta etapa, se añade un agente de escisión adecuado al medio líquido. Esto logra la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y proporciona de este modo la proteína espidroína de ampulácea mayor. En la quinta etapa, la proteína espidroína de ampulácea mayor obtenida de este modo se polimeriza en el medio líquido. La polimerización típicamente se inicia en la interfase entre dos fases diferentes, tales como interfases líquido/aire, líquido/sólido, y agua/aceite. Por lo tanto, esta quinta etapa puede también comprender adicionalmente proporcionar una interfase entre el medio líquido y otra fase. La otra fase se selecciona del grupo que consiste en una fase gaseosa, una fase líquida y una fase sólida. Como se ha detallado anteriormente, el medio líquido es típicamente un medio acuoso, y las otras fases adecuadas son por ejemplo aire y disolventes orgánicos no miscibles con agua, tales como aceite, p. ej. aceite mineral adecuado para reacciones de PCR. La presencia de la interfase resultante estimula la polimerización en la interfase o en la región que rodea la interfase, cuya región se extiende al medio líquido, de manera que dicha polimerización se inicia en dicha interfase o en dicha región de la interfase. Las interfases preferidas incluyen interfases de agua/aire y agua/aceite. La polimerización se produce típicamente en minutos o en unas pocas horas, por ejemplo en 1 min a 5 h, tras la incubación a la temperatura ambiente. En una sexta etapa opcional, el polímero así obtenido se aísla del medio líquido.

50 Con el fin de obtener una proteína con bajo contenido de pirógeno, que es una obligación para el uso como biomaterial *in vivo*, se ha desarrollado un protocolo de purificación optimizado para la eliminación de lipopolisacáridos (LPS). Para evitar la contaminación por LPS liberados, se somete a las células bacterianas productoras a etapas de lavado con  $\text{CaCl}_2$  y EDTA alternante. Después de la lisis celular, todas etapas de purificación posteriores se realizan en tampones de baja conductividad con el fin de minimizar las interacciones hidrófobas entre la proteína diana y los LPS. El contenido de LPS se minimiza adicionalmente mediante el pase de la disolución de proteína a través de una columna Endotrap, que tiene un ligando que adsorbe específicamente los LPS. Para asegurar el bajo contenido constante de LPS y otros pirógenos, todos los lotes se analizan utilizando un ensayo de pirógenos *in vitro* (IPT) y/o un análisis cinético de producto lisado de amebocitos de *Limulus* (LAL). Aunque producidas en un anfitrión bacteriano gram-negativo, las proteínas espidroínas recombinantes se pueden purificar de manera que los niveles residuales de LPS y otros pirógenos se encuentren por debajo de los límites requeridos para los ensayos animales, es decir por debajo de 25 UE/implante. En ciertas realizaciones de acuerdo con la invención, el contenido de LPS y otros pirógenos en la proteína de fusión aislada es de 1 UE/mg de proteína o inferior. En ciertas realizaciones de acuerdo con la invención, el contenido de LPS y otros pirógenos en la proteína

espidroína de ampulácea mayor aislada es de 1 UE/mg de proteína o inferior, preferiblemente de 0,25 UE/mg de proteína o inferior.

De acuerdo con otro aspecto, la presente invención proporciona un polímero de una proteína espidroína de ampulácea mayor de acuerdo con la invención. En una realización preferida, el polímero de esta proteína es obtenible mediante uno cualquiera de los métodos para ello de acuerdo con la invención.

En realizaciones preferidas, el contenido de lámina  $\beta$  del polímero de la proteína espidroína de ampulácea mayor de acuerdo con la invención está por encima de 50%, es decir más de 50% de la estructura secundaria del polímero de esta proteína se encuentra en forma de lámina  $\beta$ . Esto es ventajoso, puesto que se contempla que un contenido más alto de estructuras en lámina  $\beta$  proporciona una fibra más rígida y/o fuerte y/o menos extensible.

Es preferible que el polímero de la proteína espidroína de acuerdo con la invención sea una fibra con un tamaño macroscópico, es decir con un diámetro por encima de 1  $\mu\text{m}$ , preferiblemente por encima de 10  $\mu\text{m}$  y una longitud por encima de 5 mm. Se prefiere que la fibra tenga un diámetro en el intervalo de 10-400  $\mu\text{m}$ , preferiblemente de 60-120  $\mu\text{m}$ , y una longitud en el intervalo de 0,5-300 cm, preferiblemente 1-100 cm. Otros intervalos preferidos son 0,5-30 cm y 1-20 cm. También se prefiere que el polímero de la proteína espidroína de acuerdo con la invención tenga una resistencia a la tracción por encima de 1 MPa, preferiblemente por encima de 2 MPa, más preferiblemente 10 MPa o superior. Se prefiere que el polímero de la proteína espidroína de acuerdo con la invención tenga una resistencia a la tracción por encima de 100 MPa, más preferiblemente 200 MPa o superior. La fibra tiene la capacidad de permanecer intacta durante la manipulación física, es decir se puede utilizar para procedimientos de hilado, tejido, torsión, crochet y similares.

En otras realizaciones preferidas, el polímero de la proteína espidroína de acuerdo con la invención forma una espuma, un gel, una malla o una película.

Un fragmento de proteína que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión es útil para la fabricación de una proteína de fusión. La proteína de fusión comprende dicho fragmento de proteína y un fragmento de proteína de seda de araña de acuerdo con la invención, y los dos fragmentos se acoplan a través de dicho sitio de reconocimiento para el agente de escisión. El fragmento de proteína de seda de araña preferiblemente consiste en 150 a 420 residuos de aminoácido.

La presente invención se ilustrará adicionalmente a continuación por medio de los siguientes Ejemplos no limitantes.

## Ejemplos

### Ejemplo 1 Clonación y secuenciación del ADNc de MaSp1 de *Euprosthrops australis*

Se utilizaron las glándulas ampuláceas mayores de aproximadamente 100 arañas *Euprosthrops australis* hembra adultas, recogidas en Sudáfrica, para construir una biblioteca de ADNc CloneMiner basada en pDONR222 (Invitrogen, Paisley, UK) hecha a medida. Los clones de ADNc que codificaban la proteína MaSp1 se obtuvieron escrutando la biblioteca con una sonda de ADNc que codificaba un fragmento rico en alanina y glicina procedente de arañas *Euprosthrops* de subespecies desconocidas. La transferencia de colonias y la detección se realizaron con un sistema de marcaje directo y detección ECL (Amersham Biosciences, Uppsala, Suecia) de acuerdo con las instrucciones del fabricante.

Se seleccionó un solo clon para su caracterización adicional. Para obtener la secuencia completa del inserto de ADNc a partir de este clon, se realizaron delecciones anidadas utilizando el Erase-a-Base System (Promega, Southampton, R.U.), y la secuenciación se llevó a cabo en un aparato MegaBase 1000 (Amersham Biosciences).

El ADNc resultante de 3,8 kb (SEQ ID NO: 1) codifica una proteína MaSp1 (SEQ ID NO: 2) de 1207 residuos de aminoácido, que contiene un fragmento repetitivo de 34 segmentos ricos en alanina y glicina (SEQ ID NO: 3), y un fragmento no repetitivo C-terminal de 97 residuos de aminoácido (SEQ ID NO: 4).

### Ejemplo 2 Análisis de secuencia del fragmento repetitivo de la proteína MaSp1 de *Euprosthrops australis*

El fragmento repetitivo de la secuencia de la proteína MaSp1 de *Euprosthrops australis* del Ejemplo 1 (SEQ ID NO: 3) se analizó adicionalmente mediante el alineamiento de los segmentos repetitivos del fragmento, véase la Fig. 1. El alineamiento se escrutó cuidadosamente y se llegó a la conclusión de la siguiente información estructural.

Los segmentos ricos en alanina de las proteínas MaSp1 de *Euprosthrops australis* son largos residuos de 13-15 aminoácidos y consisten solo en residuos de alanina o todos residuos de alanina menos un residuo, que es un residuo de serina, glutamato o glicina.

El fragmento repetitivo de la proteína MaSp1 de *Euprosthrops australis* contiene adicionalmente tres tipos de segmentos ricos en glicina relacionados, pero distintos, véase la Fig. 2A. Dos de los segmentos ricos en glicina difieren casi únicamente en la longitud y la aparición; el segmento rico en glicina más común contiene 23 residuos de aminoácido, mientras que una variante menos abundante contiene 17 residuos de aminoácido. Estos dos segmentos ricos en glicina carecen generalmente de residuos cargados o contienen un residuo cargado. En contraste, el

segmento rico en glicina más corto, que contiene 14 residuos de aminoácido, contiene únicamente la secuencia GRGQG o GQGQG en el extremo N-terminal, y GN en el extremo C-terminal.

5 El segmento rico en glicina más largo está representado por la secuencia consenso de una letra de aminoácidos QGG(G/S)QGG(Q/Y)GG (L/Q)GQGGYGQGA GSS (SEQ ID NO: 5), y carece de residuos cargados. Se pronostica que este segmento forma estructuras enrolladas o estructuras en hélice 3<sub>1</sub>. El segmento rico en glicina de tamaño medio está representado por la secuencia consenso de una letra de aminoácidos GQGGQGGQ(G/R)Y GQG(A/S)G(S/G)S (SEQ ID NO: 6), y carece de residuos cargados o contiene un residuo cargado. Se pronostica que este segmento forma estructuras enrolladas. El segmento rico en glicina más corto está representado por la secuencia consenso de una letra de aminoácidos G(R/Q)GQG(G/R)YGQG (A/S/V)GGN (SEQ ID NO: 7). Se pronostica que este segmento forma estructuras con giro.

10 El fragmento repetitivo de la proteína MaSp1 de *Euprosthénops australis* está formado de segmentos alternos ricos en alanina y glicina, p. ej. ....A<sub>1</sub>G<sub>1</sub>A<sub>2</sub>G<sub>2</sub>A<sub>3</sub>G<sub>3</sub>A<sub>4</sub>G<sub>4</sub>A<sub>5</sub>G<sub>5</sub>...

Se observa que cada uno de los segmentos ricos en glicina más cortos y más largos identificados anteriormente aparecen generalmente cada dos segmentos ricos en glicina, p. ej.

15 ...A<sub>1</sub>G<sub>corto</sub>A<sub>2</sub>G<sub>largo</sub>A<sub>3</sub>G<sub>corto</sub>A<sub>4</sub>G<sub>largo</sub>A<sub>5</sub>G<sub>corto</sub>...

En contraste, el fragmento rico en glicina de tamaño medio, menos abundante aparece generalmente entre un segmento rico en glicina del tipo más largo y un segmento rico en glicina del tipo más corto, p. ej.

...A<sub>1</sub>G<sub>corto</sub>A<sub>2</sub>G<sub>largo</sub>A<sub>3</sub>G<sub>medio</sub>A<sub>4</sub>G<sub>corto</sub>A<sub>5</sub>G<sub>largo</sub>...

### 20 **Ejemplo 3 Predicción de la estructura secundaria y terciaria del fragmento repetitivo de la proteína MaSp1 de *Euprosthénops australis***

Los polipéptidos de espidroína en disolución se pliegan típicamente mediante la formación de estructuras en horquilla, lo que prefigura la estructura en lámina β antiparalela de la fibra madura. Para discernir los posibles patrones de plegamiento para el fragmento repetitivo (SEQ ID NO: 3) de la proteína MaSp1 de *Euprosthénops australis* de los Ejemplos 1-2, se identificaron las regiones de la proteína que son compatibles con la formación de estructuras en horquilla o giro. Los segmentos ricos en alanina son candidatos poco adecuados para la formación de giro puesto que se pronostica que forman estructuras helicoidales, y lo más importante, estos segmentos son mantenidos generalmente para formar las láminas β en la fibra.

30 Utilizando un algoritmo descrito recientemente para las predicciones de giro (Fuchs, PF & Alix, AJ, High accuracy prediction of beta-turns and their types using propensities and multiple alignments. *Proteins* 59, 828-839 (2005)), los segmentos ricos en glicina más cortos muestran una alta probabilidad de formación de giros β de tipo II, mientras que se pronostica que los dos segmentos ricos en glicina más largos forman estructuras enrolladas. El alto contenido de tripletes GGX en los segmentos ricos en Gly más largos sugiere que pueden formar estructuras en hélice 3<sub>1</sub>.

35 La naturaleza repetitiva de las secuencias de aminoácidos de la espidroína implica una naturaleza igualmente repetitiva del patrón de plegamiento. En conjunto, estas observaciones dan como resultado un plegamiento del fragmento repetitivo de la proteína MaSp1 de *Euprosthénops australis* como se muestra en la Fig. 2A. Es notable que los residuos cargados positivamente están localizados casi invariablemente en las estructuras de giro propuestas.

40 A partir del patrón de plegamiento del fragmento repetitivo de la proteína MaSp1 de *Euprosthénops australis*, se puede discernir un motivo que consiste en un segmento rico en alanina/segmento enrollado rico en glicina (más largo)/segmento rico en alanina (más corto) (segmento con giro rico en glicina (más corto)/segmento rico en alanina/segmento enrollado rico en glicina (más largo)/segmento rico en alanina (ilustrado esquemáticamente en la Fig. 2A).

### **Ejemplo 4 Análisis de secuencia del fragmento C-terminal no repetitivo de proteínas MaSp1**

45 Se alineó la estructura primaria del fragmento C-terminal no repetitivo (SEQ ID NO: 4) de la proteína MaSp1 de *Euprosthénops australis*, obtenida en el Ejemplo 1, con numerosos fragmentos C-terminales conocidos de proteínas MaSp1 y MaSp2, entre otros de *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)) y otros.

50 A partir del alineamiento mostrado en la Fig. 3, partiendo de la última Ser del fragmento repetitivo, resulta evidente que las regiones C-terminales de MaSp1 y MaSp2 están bien conservadas. *Euprosthénops* sp y *Nephila clavipes* tienen residuos idénticos en 95%; *Euprosthénops australis* y *Nephila clavipes* tienen residuos idénticos en 54%; y *Euprosthénops australis* y *Euprosthénops* sp tienen residuos idénticos en 55%. Se proporciona una secuencia

## ES 2 533 555 T3

consenso de las regiones C-terminales de MaSp1 y MaSp2 como SEQ ID NO: 8. En la Fig. 3, se alinean las siguientes proteínas MaSp, indicadas con entradas de acceso de GenBank cuando son aplicables:

Especie y proteína espidroína	Entrada
MaSp1 de <i>Euprosthonops</i> sp (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, <i>ídem</i> )	Cthyb_Esp
MaSp1 de <i>Euprosthonops australis</i> (SEQ ID NO: 4)	CTnat_Eau
MaSp1 de <i>Argiope trifasciata</i>	AF350266_At1
Sp1 de <i>Cyrtophora moluccensis</i>	AY666062_Cm1
MaSp1 de <i>Latrodectus geometricus</i>	AF350273_Lg1
MaSp1 de <i>Latrodectus hesperus</i>	AY953074_Lh1
Sp1 de <i>Macrothele holsti</i>	AY666068_Mh1
MaSp1 de <i>Nephila clavipes</i>	U20329_Nc1
MaSp1 de <i>Nephila pilipes</i>	AY666076_Np1
MaSp1 de <i>Nephila madagascariensis</i>	AF350277_Nm1
MaSp1 de <i>Nephila senegalensis</i>	AF350279_Ns1
Sp1 de <i>Octonoba varians</i>	AY666057_Ov1
Sp1 de <i>Psechrus sinensis</i>	AY666064_Ps1
MaSp1 de <i>Tetragnatha kauaiensis</i>	AF350285_Tk1
MaSp1 de <i>Tetragnatha versicolor</i>	AF350286_Tv1
Sp2 de <i>Araneus bicentenarius</i>	ABU20328_Ab2
MaSp2 de <i>Argiope amoena</i>	AY365016_Aam2
MaSp2 de <i>Argiope aurantia</i>	AF350263_Aau2
MaSp2 de <i>Argiope trifasciata</i>	AF350267_At2
MaSp2 de <i>Gasteracantha mammosa</i>	AF350272_Gm2
MaSp2 de <i>Latrodectus geometricus</i>	AF350275_Lg2
MaSp2 de <i>Latrodectus hesperus</i>	AY953075_Lh2
MaSp2 de <i>Nephila clavipes</i>	AY654293_Nc2



Especie y proteína espidroína	Entrada
MaSp2 de <i>Nephila madagascariensis</i>	AF350278_Nm2
MaSp2 de <i>Nephila senegalensis</i>	AF350280_Ns2
Fb1 de <i>Dolomedes tenebrosus</i>	AF350269_DtFb1
Fb2 de <i>Dolomedes tenebrosus</i>	AF350270_DtFb2
ADF-1 de <i>Araneus diadematus</i>	U47853_ADF1
ADF-2 de <i>Araneus diadematus</i>	U47854_ADF2
ADF-3 de <i>Araneus diadematus</i>	U47855_ADF3
ADF-4 de <i>Araneus diadematus</i>	U47856_ADF4

#### Ejemplo 5 Construcción de genes de MaSp1

Se amplificó la secuencia de ADN que codificaba la proteína derivada de *Euprosthénops australis* 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9) por medio de PCR con un kit Advantage GC2 (BD Biosciences, San Jose, CA, EEUU), utilizando un clon MaSp1 de la genoteca de ADNc del Ejemplo 1 como molde. Se introdujeron sitios de reconocimiento por las enzimas de restricción *Bam*HI y *Hind*III en los extremos 5' y 3', respectivamente, y se introdujo un codón de parada aguas arriba del sitio *Hind*III, mediante el uso de cebadores diseñados. El constructo *Bam*HI-5Gly/Ala-CT<sub>nat</sub>-*Hind*III se subclonó a continuación en un vector pET32 modificado (Merck Biosciences, Darmstadt, Alemania), preparado como se describe en el siguiente Ejemplo 6(C).

#### 10 Ejemplo 6 Construcción de genes quiméricos de MaSp1

##### (A) Fragmentos del gen de REP

Se amplificaron secuencias de ADN que codificaban los fragmentos repetitivos parciales (**REP**) indicados 3Gly/Ala y 4Gly/Ala por medio de PCR con *LA Taq* (TaKaRa Bio; Saint Germain-en-laye, Francia) en presencia de betaína (Henke W et al., Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* 25, 3957-3958 (1997)), utilizando un clon de ADNc parcial que codificaba una región repetitiva de proteína MaSp1 de *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)) (Entrada GenBank CQ974358 o CQ816656) como molde. Se introdujeron sitios de reconocimiento para enzimas de restricción en los extremos 5' y 3', dando los siguientes constructos: *Nco*I-3Gly/Ala-*Nhe*I y *Nco*I-4Gly/Ala-*Nhe*I que se van a unir con un fragmento CT (véase más abajo); y un clon *Nco*I-4Gly/Ala-*Xho*I que se va a expresar individualmente, cuando se inserte un codón de parada directamente aguas arriba del sitio *Xho*I.

##### (B) Fragmentos del gen de CT

Se amplificó una secuencia de ADN que codificaba el dominio C-terminal no repetitivo de *Euprosthénops sp* (pero con un alto grado de similitud con MaSp1 de *Nephila clavipes* y *Nephila senegalensis*) por medio de PCR utilizando un clon de ADN genómico que codificaba el dominio MaSp1 C-terminal (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)). Se introdujeron sitios de reconocimiento para enzimas de restricción en los extremos 5' y 3', proporcionando *Nhe*I-2Gly/Ala-CT<sub>hyb</sub>-*Xho*I, que se va a unir con los clones de REP parciales 3Gly/Ala y 4Gly/Ala (véase más arriba), y *Nco*I-2Gly/Ala-CT<sub>hyb</sub>-*Xho*I, que se va a expresar individualmente.

**(C) Construcción de los genes de MaSP1 híbridos de REP-CT**

Se unieron los clones de REP 3Gly/Ala y 4Gly/Ala con los clones de CT utilizando el vector pCR<sup>®</sup>2.1-TOPO<sup>®</sup> (Invitrogen). A continuación, se escindieron los clones 5Gly/Ala-CT<sub>hyb</sub> y 6Gly/Ala-CT<sub>hyb</sub> fusionados resultantes con *NcoI* y *XhoI*, y se subclonaron en un vector pET32 modificado (Novagen), donde el sitio de escisión de trombina original se retiró y se introdujo un nuevo sitio de trombina aguas abajo del sitio de escisión de enteroquinasa.

**Ejemplo 7 Expresión de proteínas de fusión de MaSp1**

Las proteínas MaSp1 codificadas por los genes construidos en los Ejemplos 5-6 se expresaron en forma de proteínas de fusión (del tipo **X-REP-CT**) como sigue, utilizando un vector pET32 modificado: etiqueta de tiorredoxina/etiqueta de His/etiqueta de S/sitio de escisión de trombina/gen de MaSp1, que codificaba tiorredoxina/His<sub>6</sub>/etiqueta de S/sitio de escisión de trombina/proteína MaSp1, y un gen de resistencia a ampicilina bajo el control del promotor T7.

Los diferentes constructos de MaSp1 de los vectores de expresión pET32 se transformaron en células de *Escherichia coli* BL21(DE3) (Merck Biosciences). Las células se hicieron crecer a 30°C en medio de Luria-Bertani que contenía ampicilina a una DO<sub>600</sub> de 1,0-1,5, inducidas con IPTG y adicionalmente incubadas durante 4 h a la temperatura ambiente. Las células se recogieron mediante centrifugación, y se lisaron por medio de ADNasa I y lisozima en Tris-HCl 20 mM, pH 8,0, imidazol 20 mM, con NaCl 0,5 M, y adicionalmente se purificaron mediante cromatografía de afinidad con etiqueta de His sobre agarosa Ni-NTA (Qiagen, West Sussex, Reino Unido). Las proteínas de fusión unidas se hicieron eluir de la columna de Ni-NTA con imidazol 200 mM en Tris-HCl 20 mM, pH 8,0, con NaCl 0,5 M, y se sometieron a diálisis frente a Tris-HCl 20 mM, pH 8,0. Las proteínas de fusión resultantes fueron >90% puras a juzgar por los geles de SDS-poliacrilamida teñidos con Coomassie y solubles en Tris-HCl 20 mM, pH 8,0. Este procedimiento produjo aproximadamente 40 mg/l de cultivo de proteína de fusión, que fue estable durante semanas sin una precipitación significativa.

En otro experimento, las proteínas de fusión se expresaron en forma de His<sub>6</sub>/tiorredoxina/His<sub>6</sub>/sitio de escisión de trombina/proteínas MaSp1 a partir de un plásmido que contenía el gen correspondiente y un gen de resistencia a kanamicina bajo el control del promotor T7.

**Ejemplo 8 Formación de fibras de proteínas MaSp1**

La escisión de las etiquetas de las proteínas de fusión resultantes del Ejemplo 7, se llevó a cabo en Tris-HCl 20 mM, pH 8, con una razón de trombina:proteína de fusión de 1:1000 (p/p), con un balanceo muy suave a la temperatura ambiente. La escisión de trombina se completó en 30-60 min, de acuerdo con SDS-PAGE. Las proteínas MaSp1 resultantes (Fig. 2B, SEQ ID NOS: 9-13) se polimerizaron espontáneamente en fibras macroscópicas en grados variables, véase la Tabla 1. Las fibras se formaron inicialmente en la interfase de agua/aire. La formación pudo ser observada a simple vista a partir de 1 hora de incubación (véase la Fig. 4A, 4B), y después de aproximadamente 5 horas no se produjo más crecimiento de las fibras. Las fibras de 6Gly/Ala-CT<sub>hyb</sub> tenían hasta aproximadamente 2 cm de largo, y las fibras de 5Gly/Ala-CT<sub>nat</sub> tenían ≥10 cm de largo. Los experimentos repetidos produjeron fibras de 5Gly/Ala-CT<sub>nat</sub> que tenían ≥20 cm (véase la Fig. 4C), e incluso ≥2 m de largo. La formación de fibras pudo ser observada a simple vista a partir de aproximadamente 10 min de incubación.

Las fibras se aislaron y se lavaron con tampón y después de eso se sometieron a análisis de la secuencia de aminoácidos N-terminal, que mostró solamente la secuencia de la proteína MaSp1. Esto demuestra que las etiquetas escindidas estaban ausentes de las fibras.

**Ejemplo 9 Análisis de fibras de proteína MaSp1****A. Mediciones de la resistencia a la tracción**

La resistencia a la tracción de las fibras de 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13) y 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9) del Ejemplo 8 se determinó como sigue. Con el fin de manipular las fibras de 6Gly/Ala-CT<sub>hyb</sub> más cortas (1-2 cm) para determinar las mediciones de la resistencia a la tracción, éstas se incubaron brevemente en glicerol en agua al 15% antes de ser secadas al aire. Las fibras de 5Gly/Ala-CT<sub>nat</sub> más largas (10 cm) o bien no se trataron, o bien se incubaron brevemente en glicerol al 15%, o bien se elaboraron a mano en metanol del 75% antes de secarlas al aire. La resistencia a la tracción de las fibras secadas al aire se midió tirando de las fibras en un Equipo de Ensayo de Materiales Zwick a una velocidad de 10 mm/min. Véase la Tabla 1.

La resistencia a la tracción de las fibras de 6Gly/Ala-CT<sub>hyb</sub> de 1-2 cm de largo secadas al aire, tratadas con glicerol (SEQ ID NO: 13) fue de aproximadamente 2 MPa, y la resistencia de las fibras de 10 cm de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9) fue de 4-5 MPa. Las fibras de 5Gly/Ala-CT<sub>nat</sub> de 10 cm de largo estiradas en metanol disolvente deshidratante antes del secado al aire mostraron una resistencia a la tracción de 2-3 MPa, que es ligeramente menos que para las fibras tratadas con glicerol del mismo tipo. La resistencia a la tracción más alta medida ahora fue de 10 MPa, que se encontró para una fibra de 5Gly/Ala-CT<sub>nat</sub> de 10 cm largo secada al aire sin tratamiento adicional.

5 El intervalo de resistencias a la tracción (2-10 MPa) es comparable a los valores inferiores referidos para fibras de seda de araña regeneradas (2-320 MPa). Las fibras formadas espontáneamente más largas derivan del constructo de 5Gly/Ala-CT<sub>nat</sub>, y tales fibras secadas al aire también muestran la mayor resistencia a la tracción. Potencialmente, esto pudo ser debido a sus segmentos de poli-Ala de 12-15 residuos de largo, con respecto a los segmentos de Ala de 8-14 residuos de 6Gly/Ala-CT<sub>hyb</sub>, que proporcionarían una mayor proporción de conformación en lámina β cristalina en la primera proteína.

**TABLA 1 Capacidad de formación de fibra de las proteínas MaSp1**

Proteína	SEQ ID NO	Capacidad de formación de fibra	Longitud de la fibra (cm)	Resistencia a la tracción de la fibra (MPa)
5Gly/Ala-CT <sub>nat</sub>	9	++++	≥10	2-10
5Gly/Ala-CT <sub>nat</sub> (Ejemplo 11)	9	+++++	≥20	180-230
4Gly/Ala	10	agregados	n.a.	n.a.
2Gly/Ala-CT <sub>hyb</sub>	11	agregados	n.a.	n.a.
5Gly/Ala-CT <sub>hyb</sub>	12	+	≤1	n.d.
6Gly/Ala-CT <sub>hyb</sub>	13	+++	1-2	<2
n.a. = no aplicable				
n.d. = no determinado				

### B. Microscopía Electrónica de Barrido

10 La arquitectura microscópica de las fibras de 6Gly/Ala-CT<sub>hyb</sub> y 5Gly/Ala-CT<sub>nat</sub> se analizó con microscopía electrónica de barrido (SEM) (Fig. 5). Brevemente, las muestras se aplicaron sobre portaobjetos de SEM y se cubrieron con capas de vacío con una capa de oro y paladio de 6 nm. Los especímenes se observaron y se fotografiaron en un LEO 1550 FEG SEM utilizando un voltaje de aceleración de 10 kV.

15 Esto reveló diámetros de 10-30 μm para las fibras individuales, con fibras individuales que mostraban diámetros bastante homogéneos (Fig. 5a que muestra 6Gly/Ala-CT<sub>hyb</sub>, SEQ ID NO: 13). Además de las fibras macroscópicas, se encontraron partículas de tipo gel. Después del secado al aire se observaron tales partículas de 6Gly/Ala-CT<sub>hyb</sub> directamente sobre un portaobjetos de SEM, fibras de aproximadamente 10-15 μm de diámetro (Fig. 5b, c). El diámetro de las fibras macroscópicas de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), estiradas en metanol del 75% y secadas al aire, fue de 60-120 μm y aparentemente contenían varias fibras alineadas (Fig. 5d-f). Fibra retorcida antes del secado al aire (Fig. 5e), extremo de la fibra (Fig. 5f).

### 20 C. Espectroscopía de Dicroísmo Circular

25 Las fibras que consistían en proteína 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13) o 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), preparadas en el Ejemplo 8, se lavaron en tampón fosfato 20 mM, pH 7, y se suspendieron en SDS al 2% en el mismo tampón. Los espectros de Dicroísmo Circular de 250 a 190 nm se registraron a 22°C en una cubeta de cuarzo con una longitud de camino de 0,1 cm, utilizando un espectropolarímetro Jasco J-810. La velocidad de barrido fue de 50 nm/min, el tiempo de respuesta de 2 seg., el intervalo de adquisición 0,1 nm, y la anchura de banda 1 nm.

30 El espectro mostrado en la Fig. 6 es una acumulación de tres barridos de fibras de proteína 6Gly/Ala-CT<sub>hyb</sub> (SEQ ID NO: 13). Presenta un mínimo a 220 nm y un máximo a 195 nm, rasgos que son característicos de estructuras en lámina β antiparalelas. Se obtuvieron espectros altamente similares para las fibras de 5Gly/Ala-CT<sub>nat</sub> (no mostrado). Las fibras formadas espontáneamente exhiben de este modo morfología y estructuras similares a las de las fibras de seda de araña nativas y regeneradas.

**Ejemplo 10 Biocompatibilidad de la seda de araña recombinante**

Puesto que es deseable utilizar fibras de seda de araña en aplicaciones biomédicas, la biocompatibilidad de las fibras ha sido evaluada mediante una investigación de los efectos de la seda producida de manera recombinante utilizando dos tipos celulares diferentes.

- 5 La proteína MaSp1 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9) se expresó en bacterias como se ha descrito en los Ejemplos 7-8. La proteína purificada se utilizó para producir fibras de seda artificiales con longitudes de >10 cm, e incluso >20-200 cm, y diámetros de alrededor de 100 µm.

**A. Mastocitos embrionarios de ratón**

- 10 Se sembraron mastocitos embrionarios (día 12,5) de ratón (que habían proliferado *in vitro* durante ocho semanas utilizando IL-8 y factor de células madre de mastocitos) a dos densidades celulares diferentes, siendo la densidad más alta de aproximadamente cuatro veces la densidad más baja. Estas células no se adhieren a la superficie de plástico, pero crecen en suspensión. Se añadieron a los pocillos piezas de fibra de seda, cada una de aproximadamente 0,5 cm largo. Los mastocitos se incubaron durante tres días, con o sin la presencia de fibra de seda, y después de eso se contaron las células vivas y muertas después de teñir con azul Tripán (Fig. 7). Las barras muestran los valores medios con el error típico de la media, n=2, cada muestra se cuenta por triplicado.
- 15

Los mastocitos no resultan afectados por la presencia de las fibras de seda. Después de tres días de crecimiento, no hay diferencias significativas en la muerte o proliferación celular en comparación con los controles negativos cultivados sin fibras de seda.

**B. Células de riñón embrionario humano (HEK) 293**

- 20 Se adsorbieron piezas de la fibra de seda, aproximadamente 0,5 cm de longitud, al fondo de placas de microtitulación de 6 pocillos dejándolas secar a partir de un pequeño volumen de tampón. Las fibras no se desprenden cuando se añade medio de crecimiento celular. Las células de riñón embrionario humano (HEK) 293 se cultivaron en placa a continuación a diferentes densidades celulares y se dejaron crecer durante un total de seis días. Las células HEK-293 se adhieren y crecen ancladas a la superficie de las células del plástico. Se estudió la capacidad de las células HEK-293 para crecer en la proximidad de las fibras, y el anclaje físico de las células a las fibras.
- 25

- Las células HEK-293 se anclaron y proliferaron normalmente en los pocillos que contenían fibras de seda (según se observó al microscopio óptico). Las células crecieron muy estrechamente a lo largo de los bordes de la fibra, y aparentemente incluso crecieron bajo una fibra parcialmente desprendida (Fig. 8). Después de siete días, las fibras se desprendieron cuidadosamente de la superficie de plástico, y se observó claramente que los grupos de células se anclaron físicamente a las fibras. La fibra cubre la mitad superior derecha de la figura. Se observan células HEK293 ancladas al borde de la fibra, y también crecen bajo la fibra.
- 30

- Los dos diferentes tipos celulares (mastocitos, HEK-293) estudiados no resultaron afectados por la presencia de fibras de seda recombinantes, incluso a cantidades de seda comparativamente altas. Esto indica que las fibras de seda artificiales sometidas a ensayo se asemejan a la seda dragline de tipo salvaje de *Euprosthenoops australis*, en que son no tóxicas y biocompatibles. Las fibras de seda artificiales parecen de este modo adecuadas para aplicaciones biomédicas.
- 35

**Ejemplo 11 Propiedades mecánicas y estructura de fibras de proteína MaSp1**

- 40 Las propiedades mecánicas de las fibras de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9) se examinaron utilizando ensayos de tracción realizados para proporcionar curvas de tensión-deformación (Fig. 9). Estas propiedades de tracción se caracterizaron utilizando un equipo de ensayo de materiales Zwick Roell Z2.5 (Zwick, Ulm, Alemania). Los ensayos se realizaron en aire a condiciones ambiente (20°C y 52% de humedad relativa) utilizando una velocidad de carga de 10 mm/min. Las piezas de fibra se transfirieron directamente desde el tampón, se montaron y se sometieron a dos ciclos de estiramiento-relajación. Con el fin de generar una hebra de seda homogénea adecuada para el ensayo de tracción, las fibras se alargaron utilizando ciclos de estiramiento-relajación. En primer lugar, las fibras se alargaron tirando con una fuerza de 0,1 N. Después de la relajación, se tiró de ellas hasta aplicar una fuerza de 0,25 N.
- 45

- Este tratamiento generó fibras homogéneas alargadas con un diámetro de aproximadamente 80 µm según se determinó mediante mediciones de la altura utilizando un aparato Mitutoyo IDC-112B (Mitutoyo Corp, Tokyo, Japón) y se confirmó mediante microscopía electrónica de barrido (SEM) como sigue. Antes y después de los ciclos de estiramiento-relajación, se aplicaron piezas de fibra sobre portaobjetos para SEM y se secaron al aire durante la noche. Las muestras se cubrieron con capas de vacío con una capa de oro y paladio de 6 nm. Los especímenes se observaron y se fotografiaron con un microscopio LEO 1550 FEG (Carl Zeiss, Oberkochen, Alemania) utilizando un voltaje de aceleración de 10 kV.
- 50

- Las fibras estiradas se cortaron en piezas, cuyos extremos se fijaron entre cartulina con pegamento (Loctite 420, Loctite, Göteborg, Suecia). A continuación se fijaron las muestras de fibra en las sujeciones del aparato de prueba
- 55

de materiales y se estiraron hasta que se rompieron. Se construyeron curvas de tensión-deformación utilizando el área transversal inicial de la fibra pre-estirada, suponiendo una sección transversal circular. Los valores de tensión se normalizan al área transversal inicial de la fibra. La deformación corresponde a  $dL/L_0$  donde  $L_0$  es la longitud inicial de la fibra y  $dL$  es el cambio en la longitud de la fibra. En la Fig. 9, se muestran las curvas de tensión-deformación para tres muestras diferentes de fibras doblemente estiradas de 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9), y su resistencia a la tracción media aproximadamente 0,2 GPa.

Se analizó la arquitectura microscópica de las fibras mediante SEM (Fig. 10). Las fibras formadas espontáneamente tienen una apariencia aplanada homogénea y una anchura de hasta varios cientos de micrometros, mientras la altura mide algunas decenas de micrometros (Fig. 10a, b).

Después de que las fibras hayan sido sometidas a ciclos de estiramiento-relajación, su sección transversal adoptó una forma más redondeada con una subestructura compacta de fibrillas estrechamente alineadas (Fig. 10c-f). La apariencia de superficies cortadas o fracturadas (Fig. 10e, f) dan fe, adicionalmente, de la compacidad de la fibra producida.

En conclusión, las fibras formadas espontáneamente muestran una morfología y unas propiedades mecánicas similares a las de las fibras de seda de araña nativas o regeneradas, incluso sin hilado.

### Ejemplo 12 Variantes de la proteína espidroína

Se piensa que las interacciones intermoleculares fuertes contribuyen a la impresionante resistencia a la tracción de la seda de araña. Por lo tanto, se han producido variantes de espidroínas en miniatura que permiten el entrecruzamiento covalente intermolecular en las fibras. Se han construido dos proteínas espidroínas mutantes diferentes mediante mutagénesis dirigida al sitio para introducir dos residuos de cisteína en el primer (SEQ ID NO: 14, posiciones 36 y 37) y el cuarto (SEQ ID NO: 15, posiciones 128 y 129) bloques de alanina, respectivamente. Estas variantes han sido expresadas y aisladas utilizando el mismo protocolo descrito en los Ejemplos 7-8 para los genes construidos en los Ejemplos 5-6. Estas variantes (SEQ ID NOS: 14-15) forman fibras de la misma manera que 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9).

Con el fin de averiguar la importancia de la dimerización del dominio C-terminal, se ha construido una variante donde el residuo de cisteína del dominio C-terminal se cambia por un residuo de serina (SEQ ID NO: 16, posición 222). No obstante, esta variante (SEQ ID NO: 16) forma fibras de la misma manera que 5Gly/Ala-CT<sub>nat</sub> (SEQ ID NO: 9).

### Ejemplo 13 Eliminación de LPS y otros pirógenos de proteínas espidroína expresadas

Se lavan células de *E. coli* que expresan la proteína de fusión de espidroína deseada con los siguientes tampones:

A: Tris 100 mM, pH 8,

B: CaCl<sub>2</sub> 5 mM, Tris 100 mM, pH 8,

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

D: Tris 100 mM, pH 8, y

E: Tris 100 mM, pH 8.

Después de eso, las células se lisan en Tris 20 mM, pH 8 con un suplemento de lisozima y ADNasa I. La muestra de proteína se carga a continuación sobre una matriz de Ni-sefarosa y se lava con Tris 20 mM, imidazol 10-100 mM, pH 8 antes de la elución con Tris 20 mM, imidazol 100-300 mM, pH 8. Las fracciones relevantes se reúnen y se someten a diálisis frente a Tris 20 mM, pH 8 durante la noche. La muestra de proteína se complementa a continuación con CaCl<sub>2</sub> 100 μM y finalmente se hace pasar a través de una columna EndoTrap Blue, previamente equilibrada con Tris 20 mM, CaCl<sub>2</sub> 100 μM, pH 8. De este modo, se pueden obtener muestras de proteína con un contenido de pirógeno de 1 UE/mg de proteína, como se evalúa mediante IPT y un análisis cinético LAL.

La proteína de fusión se escinde proteolíticamente a continuación con trombina utilizando una razón 1:1000 (p/p) de trombina:proteína de fusión, que induce la formación de fibra (como se ha descrito anteriormente). Las fibras se lavan 3 veces en Tris 20 mM, pH 8 y finalmente 3 veces en agua. Esto proporciona fibras con un contenido de pirógeno de 0,25 UE/mg de fibra.

Las características estructurales de las fibras no resultan afectadas después de pasar por el autoclave a 125°C y 1,5 bar durante 10 min, lo que permite una esterilización eficaz del material. Las fibras son químicamente estables y no pueden ser solubilizadas en urea 8 M, GuaHCl 6 M, o HAc puro. No obstante, las fibras pueden ser solubilizadas en HFIP o ácido fórmico puros.

**Lista de secuencias**

<110> Johansson, Jan Grip, Stefan Hjäl m, Göran Rising, Anna Stark, Margareta Engström, Wilhelm Hedhammar, My

5 <120> PROTEÍNAS DE SEDA DE ARAÑA Y MÉTODOS PARA PRODUCIR PROTEÍNAS DE SEDA DE ARAÑA

<130> PC-21028983

<150> SE 0502932-7

10 <151> 2005-12-30

<160> 16

<170> PatentIn versión 3.3

15

<210> 1

<211> 3766

<212> ADN

<213> Euprostenops australis

20

<400> 1

```

gtcaaggtgc tggaggtaat gccgctgcag cagccgcagc agcagcagca gcagcagctg      60
gacagggcgg tcaaggtgga tatggtggac taggtcaagg aggatatgga caggggtgcag      120
gaagttctgc agccgccgcc gccgcagcag cagcagcagc tgcagcagct ggacgaggtc      180
aaggaggata tggtaagggt tctggaggta atgccgctgc agcagccgca gcagctgcag      240
cagcagcatc tggacaagga ggtcaaggag gacaagggtg acaagggtcaa 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 cgccgcagca tcagcagccg      540
ccgcagcagc tggacgaggt caaggaggat atggtcaagg tgctggaggt aatgccgctg      600
cagcagccgc agcagctgcc gccccgcagc cagctggaca gggtggtcaa ggtggatatg      660
gtggactagg tcaaggagga tatggacaag gtgcaggaag ttctgcagcc gctgccgccg      720
cagcagcagc agccgccgcc gcaggtggac aaggtggaca aggtcaagga agatatggac      780
aaggtgcagg aagttctgca gccgctgccg ccgcagcagc agcagccgcc gcagcagctg      840
gacaaggtca aggaggatat ggtcaagggt ctggaggtaa tgccgctgca gcagccgcag      900
cagctgcagc agcagcagct ggacaaggag gtcaaggagg acaagggtgga ctaggtcaag      960
gaggatatgg acaagggtgca ggaagttctg ccgccgccgc cgcagcagca gcagccgccg     1020
cagcagctgg acgaggtcaa ggaggatat gtcaaggtgc tggaggtaat gccgctgcag     1080
cagccgcagc agctgccgaa gccgcagcag ctggacaggg tggtaagggt ggatatggtg     1140
gactaggtca aggaggatat ggacaagggt caggaagttc tgcagccgcc gccgcagcag     1200
cagcagccgc cgcagcagct ggacgaggtc aaggaggata tggtaagggt gctggaggta     1260

```

ES 2 533 555 T3

atgccgctgc agcagccgca gcagctgccg ccgccgcagc agctggacag ggtggtcaag 1320  
gtggatatgg tggactaggt caaggaggat atggacaagg tgcaggaagt tctgcagccg 1380  
ctgccgccgc agcagcagca gccgccgccg cagggtggaca aggtggacaa ggtcaaggaa 1440  
gatatggaca aggtgcagga agttctgcag ccgctgccgc cgcagcagca gcagcagccg 1500  
cagcagctgg acgaggtcaa ggaggatatg gtcaaggttc tggaggtaat gccgctgcag 1560  
cagccgcagc agctgcagca gcagcatctg gacaaggaag tcaaggagga caagggtggac 1620  
aaggtaagg tggatatgga caagggtcag gaagttctgc agccgccgcc gccgccgcag 1680  
cagcagccgc cgcagcatct ggacgaggtc aaggaggata tggtaagggt gctggaggta 1740  
atgccgctgc tgcagccgca gcagctgccg ccgccgcagc agctggacag ggcggtcaag 1800  
gtggatatgg tggactaggt caaggaggat atggacaagg tgcaggaagt tctgcagccg 1860  
ctgccgccgc cgcagcagcc gccgcagcag gtggacaagg tggacaagggt caaggaggat 1920  
atggacaagg tgcaggaagt tctgcagccg ccgccgcagc agcagcagca gcagccgcag 1980  
cagctggacg aggtcaagga ggatatggtc aaggttctgg aggtaatgcc gctgcagcag 2040  
ccgcagcagc tgcagcagca gcatctggac aaggagggtca aggaggacaa ggtggacaag 2100  
gtcaagggtg rtatggacaa ggtgcaggaa gttctgcagc cgccgccgcc gcagcagcag 2160  
cagccgccgc agcagctgga caaggtcaag gaggatatgg tcaagggtgct ggaggtaatg 2220  
ccgctgcagc agccgcagca gctgcagcag cagcagctgg acaaggagggt caaggaggac 2280  
aagggtgact aggtcaagga ggatatggac aagggtcagg aagttctgca gccgccgccg 2340  
cagcmgcmgc agcagccgcc gcagcagctg gacgaggtca aggaggatat ggtcaagggtg 2400  
ttggaggtaa tgccgctgca gcagccgcag cagctgcagc agcagcagct ggacaaggag 2460  
gtcaaggagg acaaggtgga ctaggtaag gaggatatgg acaagggtgca ggtagttctg 2520  
cagccgccgc cgccgccgca gcagcagccg ccgcagcagc tggacgaggt caaggaggat 2580  
atggtcaagg ttctggagggt aatgccgctg cagcagccgc agcagctgca gcagcagcat 2640  
ctggacaagg aagtcaagga ggacaagggt gacaagggtca aggtggatat ggacaagggtg 2700  
caggaagttc tgcagccgcc gccgccgcag cagcagcagc cgccgcagca tctggacgag 2760  
gtcaaggagg atatggtcaa ggtgctggag gtaatgccgc tgctgcagcc gcagcagctg 2820  
ccgccgccgc agcagctgga cagggcggtc aagggtggata tgggtggacta ggtcaaggag 2880  
gatatggaca aggtgcagga agttctgcag ccgctgccgc cgccgcagca gccgccgcag 2940  
cagggtggaca aggtggacaa ggtcaaggag gatatggaca aggttcagga ggttctgcag 3000  
ccgccgccgc cgccgcagca gcagcagcag ctgcagcagc tggacgaggt caaggaggat 3060  
atggtcaagg ttctggagggt aatgctgctg ccgcagccgc tgccgccgcc gccgccgctg 3120  
cagcagccgg acagggagggt caagggtgat atggtagaca aagccaagggt gctggttccg 3180  
ctgctgctgc tgctgctgct gctgccgctg ctgctgctgc aggatctgga caagggtgat 3240  
acggtggaca aggtcaagga ggttatggtc agagtagtgc ttctgcttca gctgctgcgt 3300

ES 2 533 555 T3

cagctgctag tactgtagct aattcgggta gtcgcctctc atcgccttcc gcagtatctc 3360  
 gagtttcttc agcagtttct agcttggttt caaatgggtca agtgaatatg gcagcgttac 3420  
 ctaatatcat ttccaacatt tcttcttctg tcagtgcacg tgctcctggg gcttctggat 3480  
 gtgaggtcat agtgcaagct ctactcgaag tcatcactgc tcttgttcaa atcgttagtt 3540  
 cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat gttggtgcta 3600  
 atgccatggc tcaagtaatg ggctgagggt tttaatagta aaaggtgtga tattcctcaa 3660  
 tgttttgaaa attattaatc gaatttttac cttgtgtgct atcagatata aattgaagta 3720  
 taataaataa atatttgcac tttcaaaaaa aaaaaaaaaa aaaaaa 3766

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

5

<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 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 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  
 10 ser Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln



ES 2 533 555 T3

180					185					190				
Gly	Ala	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
		195					200					205		
Ala	Ala	Ala	Gly	Gln	Gly	Gly	Gln	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
		225					230					235		240
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gln
				245					250					255
Arg	Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala
			260					265					270	
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gln	Gly	Gln	Gly	Gly	Tyr	Gly
			275					280					285	
Gly	Ala	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
		290					295					300		
Ala	Ala	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Leu	Gly	Gln
		305							310					315
Gly	Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala
				325					330					335
Ala	Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln
				340				345					350	
Ala	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Glu	Ala	Ala
			355					360					365	
Ala	Ala	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Gly	Leu	Gly	Gln
				370					375				380	
Gly	Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala
				385					390					395
Ala	Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln
				405					410					415
Ala	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			420						425					430
Ala	Ala	Gly	Gln	Gly	Gly	Gln	Gly	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
				450					455					460

ES 2 533 555 T3

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 Ala  
 610 615 620

Ala 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 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 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 Ala  
 705 710 715 720

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

ES 2 533 555 T3

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

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

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

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

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

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

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

ES 2 533 555 T3

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 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

5 <210> 3  
 <211> 1110  
 <212> PRT  
 <213> Euprostenops australis

10 <220>  
 <221> REPETICIÓN  
 <222> (7)..(19)

15 <220>  
 <221> REPETICIÓN  
 <222> (20)..(42)

<220>  
 <221> REPETICIÓN  
 <222> (43)..(56)

20 <220>  
 <221> REPETICIÓN  
 <222> (57)..(70)

	<220> <221> REPETICIÓN <222> (71)..(83)
5	<220> <221> REPETICIÓN <222> (84)..(106)
10	<220> <221> REPETICIÓN <222> (107)..(120)
15	<220> <221> REPETICIÓN <222> (121)..(134)
20	<220> <221> REPETICIÓN <222> (135)..(147)
25	<220> <221> REPETICIÓN <222> (148)..(170)
30	<220> <221> REPETICIÓN <222> (171)..(183)
35	<220> <221> REPETICIÓN <222> (184)..(197)
40	<220> <221> REPETICIÓN <222> (198)..(211)
45	<220> <221> REPETICIÓN <222> (212)..(234)
50	<220> <221> REPETICIÓN <222> (235)..(248)
55	<220> <221> REPETICIÓN <222> (249)..(265)
60	<220> <221> REPETICIÓN <222> (266)..(279)
65	<220> <221> REPETICIÓN <222> (280)..(293)
	<220> <221> REPETICIÓN <222> (294)..(306)
	<220> <221> REPETICIÓN <222> (307)..(329)
	<220> <221> REPETICIÓN

<222> (330)..(342)

<220>  
<221> REPETICIÓN  
5 <222> (343)..(356)

<220>  
<221> REPETICIÓN  
10 <222> (357)..(370)

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

15 <220>  
<221> REPETICIÓN  
<222> (394)..(406)

20 <220>  
<221> REPETICIÓN  
<222> (407)..(420)

<220>  
<221> REPETICIÓN  
25 <222> (421)..(434)

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

30 <220>  
<221> REPETICIÓN  
<222> (458)..(470)

35 <220>  
<221> REPETICIÓN  
<222> (471)..(488)

40 <220>  
<221> REPETICIÓN  
<222> (489)..(502)

<220>  
<221> REPETICIÓN  
45 <222> (503)..(516)

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

50 <220>  
<221> REPETICIÓN  
<222> (530)..(552)

55 <220>  
<221> REPETICIÓN  
<222> (553)..(566)

60 <220>  
<221> REPETICIÓN  
<222> (567)..(580)

<220>  
<221> REPETICIÓN  
65 <222> (581)..(594)

- <220>  
 <221> REPETICIÓN  
 <222> (595)..(617)
- 5 <220>  
 <221> REPETICIÓN  
 <222> (618)..(630)
- 10 <220>  
 <221> REPETICIÓN  
 <222> (631)..(647)
- 15 <220>  
 <221> REPETICIÓN  
 <222> (648)..(661)
- 20 <220>  
 <221> REPETICIÓN  
 <222> (662)..(675)
- 25 <220>  
 <221> REPETICIÓN  
 <222> (689)..(711)
- 30 <220>  
 <221> REPETICIÓN  
 <222> (712)..(725)
- 35 <220>  
 <221> REPETICIÓN  
 <222> (726)..(739)
- 40 <220>  
 <221> REPETICIÓN  
 <222> (740)..(752)
- 45 <220>  
 <221> REPETICIÓN  
 <222> (753)..(775)
- 50 <220>  
 <221> REPETICIÓN  
 <222> (776)..(789)
- 55 <220>  
 <221> REPETICIÓN  
 <222> (790)..(803)
- 60 <220>  
 <221> REPETICIÓN  
 <222> (804)..(816)
- 65 <220>  
 <221> REPETICIÓN  
 <222> (817)..(839)
- 70 <220>  
 <221> REPETICIÓN  
 <222> (840)..(853)
- 75 <220>  
 <221> REPETICIÓN

<222> (854)..(867)

<220>  
<221> REPETICIÓN  
5 <222> (868)..(880)

<220>  
<221> REPETICIÓN  
10 <222> (881)..(903)

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

15 <220>  
<221> REPETICIÓN  
<222> (918)..(931)

20 <220>  
<221> REPETICIÓN  
<222> (932)..(945)

25 <220>  
<221> REPETICIÓN  
<222> (946)..(968)

30 <220>  
<221> REPETICIÓN  
<222> (969)..(981)

35 <220>  
<221> REPETICIÓN  
<222> (982)..(998)

40 <220>  
<221> REPETICIÓN  
<222> (999)..(1013)

45 <220>  
<221> REPETICIÓN  
<222> (1014)..(1027)

50 <220>  
<221> REPETICIÓN  
<222> (1028)..(1042)

55 <220>  
<221> REPETICIÓN  
<222> (1043)..(1059)

60 <220>  
<221> REPETICIÓN  
<222> (1060)..(1073)

65 <220>  
<221> REPETICIÓN  
<222> (1074)..(1092)

60 <400> 3



ES 2 533 555 T3

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 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 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 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 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

ES 2 533 555 T3

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 Ala  
 485 490 495  
 Ala 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 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 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 Ala  
 580 585 590

ES 2 533 555 T3

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  
 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

ES 2 533 555 T3

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

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

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

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

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

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

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

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

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

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

Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala 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

5

ES 2 533 555 T3

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 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

5 <210> 5  
 <211> 23  
 <212> PRT  
 <213> Artificial

10 <220>  
 <223> Secuencia consenso derivada de repeticiones internas de MaSp1 de Euprosthénops australis

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

15 <220>  
 <221> VARIANTE  
 <222> (8)..(8)  
 <223> Tyr

20 <220>  
 <221> VARIANTE  
 <222> (11)..(11)  
 <223> Gln

25 <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

30 <210> 6  
 <211> 17  
 <212> PRT  
 <213> Artificial

35 <220>  
 <223> Secuencia consenso derivada de repeticiones internas de MaSp1 de Euprosthénops australis

40 <220>  
 <221> VARIANTE  
 <222> (9)..(9)  
 <223> Arg

<220>  
 <221> VARIANTE  
 <222> (14)..(14)  
 <223> Ser  
 5

<220>  
 <221> VARIANTE  
 <222> (16)..(16)  
 <223> Gly  
 10

<400> 6  
**Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser**  
**1 5 10 15**

**Ser**

15 <210> 7  
 <211> 14  
 <212> PRT  
 <213> Artificial

20 <220>  
 <223> Secuencia consenso derivada de repeticiones internas de Masp1 de Euprosthenops australis

25 <220>  
 <221> VARIANTE  
 <222> (2)..(2)  
 <223> Gln

30 <220>  
 <221> VARIANTE  
 <222> (6)..(6)  
 <223> Arg

35 <220>  
 <221> VARIANTE  
 <222> (11)..(11)  
 <223> Ser

40 <220>  
 <221> VARIANTE  
 <222> (11)..(11)  
 <223> val  
 <400> 7

45 **Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn**  
**1 5 10**

50 <210> 8  
 <211> 100  
 <212> PRT  
 <213> Artificial

55 <220>  
 <223> Secuencia consenso derivada de proteínas MaSp1 y Masp2 conocidas

60 <220>  
 <221> RASGO\_MISC  
 <222> (1)..(71)  
 <223> Longitud de la secuencia presente en variantes de especies conocidas

60 <220>  
 <221> VARIANTE  
 <222> (7)..(7)

ES 2 533 555 T3

<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

5 Leu Gly Glu Phe  
100

<210> 9

<211> 269

<212> PRT

10 <213> Artificial

<220>

<223> Secuencia derivada de MaSp1 de Euprostenops australis

15 <220>

<221> DOMINIO

<222> (1)..(171)

<223> Fragmento repetitivo

20 <220>

<221> REPETICIÓN

<222> (11)..(30)

25 <220>

<221> REPETICIÓN

<222> (31)..(42)

30 <220>

<221> REPETICIÓN

<222> (43)..(60)

35 <220>

<221> REPETICIÓN

<222> (61)..(75)

40 <220>

<221> REPETICIÓN

<222> (76)..(89)

45 <220>

<221> REPETICIÓN

<222> (90)..(104)

<220>

<221> REPETICIÓN

<222> (105)..(121)

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

5

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

10

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

15

<220>  
 <221> DOMINIO  
 <222> (172)..(269)  
 <223> Fragmento C-terminal

20

<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



ES 2 533 555 T3

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> Euprostenops australis

5

<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

10

<210> 11  
 <211> 152  
 <212> PRT  
 <213> Artificial

15

<220>

ES 2 533 555 T3

<223> Secuencia híbrida derivada de MaSp1 de Euprosthénops australis y Euprosthénops sp

<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

5

<210> 12

<211> 257

<212> PRT

10 <213> Artificial

<220>

<223> Secuencia híbrida derivada de MaSp1 de Euprosthénops australis y Euprosthénops sp

15 <400> 12

ES 2 533 555 T3

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 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

5 <212> PRT

<213> Artificial

<220>

<223> Secuencia híbrida derivada de MaSp1 de Euprostenops australis y Euprostenops sp

10

<220>

<221> REPETICIÓN

<222> (1)..(11)

15

<220>

<221> DOMINIO  
 <222> (1)..(180)

5 <223> Fragmento repetitivo  
 <220>  
 <221> REPETICIÓN  
 <222> (12)..(25)

10 <220>  
 <221> REPETICIÓN  
 <222> (26)..(45)

15 <220>  
 <221> REPETICIÓN  
 <222> (46)..(60)  
 <220>  
 <221> REPETICIÓN  
 <222> (61)..(74)

20 <220>  
 <221> REPETICIÓN  
 <222> (75)..(88)

25 <220>  
 <221> REPETICIÓN  
 <222> (89)..(111)

30 <220>  
 <221> REPETICIÓN  
 <222> (112)..(119)

35 <220>  
 <221> REPETICIÓN  
 <222> (120)..(149)

40 <220>  
 <221> REPETICIÓN  
 <222> (150)..(158)

45 <220>  
 <221> REPETICIÓN  
 <222> (159)..(172)

50 <220>  
 <221> DOMINIO  
 <222> (173)..(180)  
 <223> Fragmento C-terminal

<400> 13

ES 2 533 555 T3

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 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 Gln Ala Ser Ser Arg Val  
 180 185 190  
 Ser Ser Ala Val Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala  
 195 200 205  
 Ala Leu Ser Ser Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser  
 210 215 220  
 Asn Pro Gly Leu Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu  
 225 230 235 240  
 Val Val Ser Ala Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln  
 245 250 255  
 Val Asn Tyr Gly Ser Ala Gly Gln Ala Thr Gln Leu Val Gly Gln Ser  
 260 265 270  
 Val Tyr Gln Ala  
 275

5 <210> 14  
 <211> 269  
 <212> PRT  
 <213> Artificial

10 <220>  
 <223> Secuencia derivada de MaSp1 de Euprostenops australis

<220>

<221> DOMINIO  
 <222> (1)..(171)  
 <223> Fragmento repetitivo

5 <220>  
 <221> REPETICIÓN  
 <222> (11)..(30)

10 <220>  
 <221> REPETICIÓN  
 <222> (31)..(42)

15 <220>  
 <221> MUTÁGENO  
 <222> (36)..(37)

20 <220>  
 <221> REPETICIÓN  
 <222> (43)..(60)

25 <220>  
 <221> REPETICIÓN  
 <222> (76)..(89)

30 <220>  
 <221> REPETICIÓN  
 <222> (90)..(104)

35 <220>  
 <221> REPETICIÓN  
 <222> (105)..(121)

40 <220>  
 <221> REPETICIÓN  
 <222> (122)..(135)

45 <220>  
 <221> REPETICIÓN  
 <222> (153)..(171)

50 <220>  
 <221> DOMINIO  
 <222> (172)..(269)  
 <223> Fragmento C-terminal

55 <220>  
 <221> DOMINIO  
 <222> (172)..(269)  
 <223> Fragmento C-terminal

60 <400> 14

ES 2 533 555 T3

Gly Ser Ala Met Gly Tyr Leu Trp Ile Gln Gly Gln Gly Gly Tyr Gly  
1 5 10 15  
Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala  
20 25 30  
Ala Ala Ala Cys Cys Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln  
35 40 45  
Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala  
50 55 60  
Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly  
65 70 75 80  
Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala  
85 90 95  
Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr  
100 105 110  
Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala  
115 120 125  
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

5 <210> 15  
<211> 269  
<212> PRT  
<213> Artificial

10 <220>  
<223> Secuencia derivada de MaSp1 de Euprostenops australis

<220>  
<221> DOMINIO  
15 <222> (1)..(171)  
<223> Fragmento repetitivo

<220>  
 <221> REPETICIÓN  
 <222> (11)..(30)  
 5  
 <220>  
 <221> REPETICIÓN  
 <222> (31)..(42)  
 10  
 <220>  
 <221> REPETICIÓN  
 <222> (43)..(60)  
 15  
 <220>  
 <221> REPETICIÓN  
 <222> (61)..(75)  
 20  
 <220>  
 <221> REPETICIÓN  
 <222> (76)..(89)  
 25  
 <220>  
 <221> REPETICIÓN  
 <222> (90)..(104)  
 30  
 <220>  
 <221> REPETICIÓN  
 <222> (105)..(121)  
 35  
 <220>  
 <221> MUTÁGENO  
 <222> (128)..(129)  
 40  
 <220>  
 <221> REPETICIÓN  
 <222> (136)..(152)  
 45  
 <220>  
 <221> REPETICIÓN  
 <222> (153)..(171)  
 50  
 <220>  
 <221> DOMINIO  
 <222> (172)..(269)  
 <223> Fragmento C-terminal  
 <400> 15



ES 2 533 555 T3

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 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> Artificial

5

<220>  
 <223> Secuencia derivada de MaSp1 de Euprosthénops australis

10

<220>  
 <221> DOMINIO  
 <222> (1)..(171)  
 <223> Fragmento repetitivo

ES 2 533 555 T3

<220>  
 <221> REPETICIÓN  
 <222> (11)..(30)  
 5  
 <220>  
 <221> REPETICIÓN  
 <222> (31)..(42)  
 10  
 <220>  
 <221> REPETICIÓN  
 <222> (43)..(60)  
 15  
 <220>  
 <221> REPETICIÓN  
 <222> (61)..(75)  
 20  
 <220>  
 <221> REPETICIÓN  
 <222> (76)..(89)  
 25  
 <220>  
 <221> REPETICIÓN  
 <222> (90)..(104)  
 30  
 <220>  
 <221> REPETICIÓN  
 <222> (105)..(121)  
 35  
 <220>  
 <221> REPETICIÓN  
 <222> (122)..(135)  
 40  
 <220>  
 <221> REPETICIÓN  
 <222> (136)..(152)  
 45  
 <220>  
 <221> DOMINIO  
 <222> (172)..(269)  
 <223> Fragmento C-terminal  
 50  
 <220>  
 <221> MUTÁGENO  
 <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

ES 2 533 555 T3

20 25 30  
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln  
 35 40 45  
 Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala  
 50 55 60  
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly  
 65 70 75 80  
 Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala  
 85 90 95  
 Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr  
 100 105 110  
 Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala  
 115 120 125  
 Ala Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly  
 130 135 140  
 Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala  
 145 150 155 160  
 Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser  
 165 170 175  
 Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser  
 180 185 190  
 Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile  
 195 200 205  
 Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Ser Glu Val  
 210 215 220  
 Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val  
 225 230 235 240  
 Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile  
 245 250 255  
 Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly  
 260 265

## REIVINDICACIONES

1. Una proteína espidroína de la ampulácea mayor aislada,

caracterizada porque dicha proteína consiste en 150 a 420 residuos de aminoácido y se define mediante la fórmula **REP-CT**, en donde

5 **REP** es un fragmento de proteína que tiene de 80 a 300 residuos de aminoácido, en donde dicho fragmento se selecciona del grupo de **L(AG)<sub>n</sub>L**, **L(AG)<sub>n</sub>AL**, **L(GA)<sub>n</sub>L**, **L(GA)<sub>n</sub>GL**, en donde

n es un número entero de 4 a 8;

cada segmento A individual es una secuencia de aminoácidos de 8 a 18 residuos de aminoácido, en donde de 0 a 3 de los residuos de aminoácido no son Ala, y los residuos de aminoácido restantes son Ala;

10 cada segmento G individual es una secuencia de aminoácidos de 12 a 30 residuos de aminoácido, en donde al menos 40% de los residuos de aminoácido son Gly; y

cada segmento L individual es una secuencia de aminoácidos conectora de 0 a 20 residuos de aminoácido; y

15 **CT** es un fragmento de proteína que tiene de 70 a 120 residuos de aminoácido, cuyo fragmento es, o tiene al menos una identidad de 80% con, una secuencia de aminoácidos seleccionada del grupo que consiste en el SEQ ID NO: 4, residuos de aminoácido 172-269 del SEQ ID NO: 9, y residuos de aminoácido 172-269 del SEQ ID NO: 16.

2. Una proteína aislada de acuerdo con la reivindicación 1, en donde

20 cada segmento A individual es, o tiene al menos una identidad de 80% con, una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1060-1073 del SEQ ID NO: 3; residuos de aminoácido 31-42, 61-75, 90-104, 122-135 y 153-171 del SEQ ID NO: 9; residuos de aminoácido 31-42 del SEQ ID NO: 14; y los residuos de aminoácido 122-135 del SEQ ID NO: 15; y

25 cada segmento G individual es idéntico a, o tiene al menos una identidad de 80% con, una secuencia de aminoácidos seleccionada del grupo de residuos de aminoácido 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 y 1074-1092 del SEQ ID NO: 3; SEQ ID NOS: 5-7; y los residuos de aminoácido 11-30, 43-60, 76-89, 105-121 y 136-152 del SEQ ID NO: 9.

30 3. Una proteína aislada de acuerdo con una cualquiera de las reivindicaciones 1 o 2, en donde dicho fragmento **CT** es, o tiene al menos una identidad de 50% con, el SEQ ID NO: 8; o en donde dicho fragmento **CT** es, o tiene al menos una identidad de 90% con, una secuencia de aminoácidos seleccionada del grupo que consiste en el SEQ ID NO: 4, residuos de aminoácido 172-269 del SEQ ID NO: 9, y residuos de aminoácido 172-269 del SEQ ID NO: 16.

35 4. Una proteína aislada de acuerdo con la reivindicación 3, en donde dicho fragmento **CT** es, o tiene al menos una identidad de 95% con, una secuencia de aminoácidos seleccionada del grupo que consiste en el SEQ ID NO: 4, residuos de aminoácido 172-269 del SEQ ID NO: 9, y residuos de aminoácido 172-269 del SEQ ID NO: 16.

40 5. Una proteína aislada de acuerdo con una cualquiera de las reivindicaciones 1-4, en donde n es un número entero de 4 a 6.

6. Una proteína aislada de acuerdo con una cualquiera de las reivindicaciones 1-5, seleccionada del grupo que consiste en el SEQ ID NO: 9 y los SEQ ID NOS: 14-16.

7. Una proteína de fusión aislada seleccionada del grupo de **X-REP-CT**, y **REP-CT-X**, en donde

45 **REP** y **CT** son fragmentos de proteína de acuerdo con una cualquiera de las reivindicaciones 1-6; y

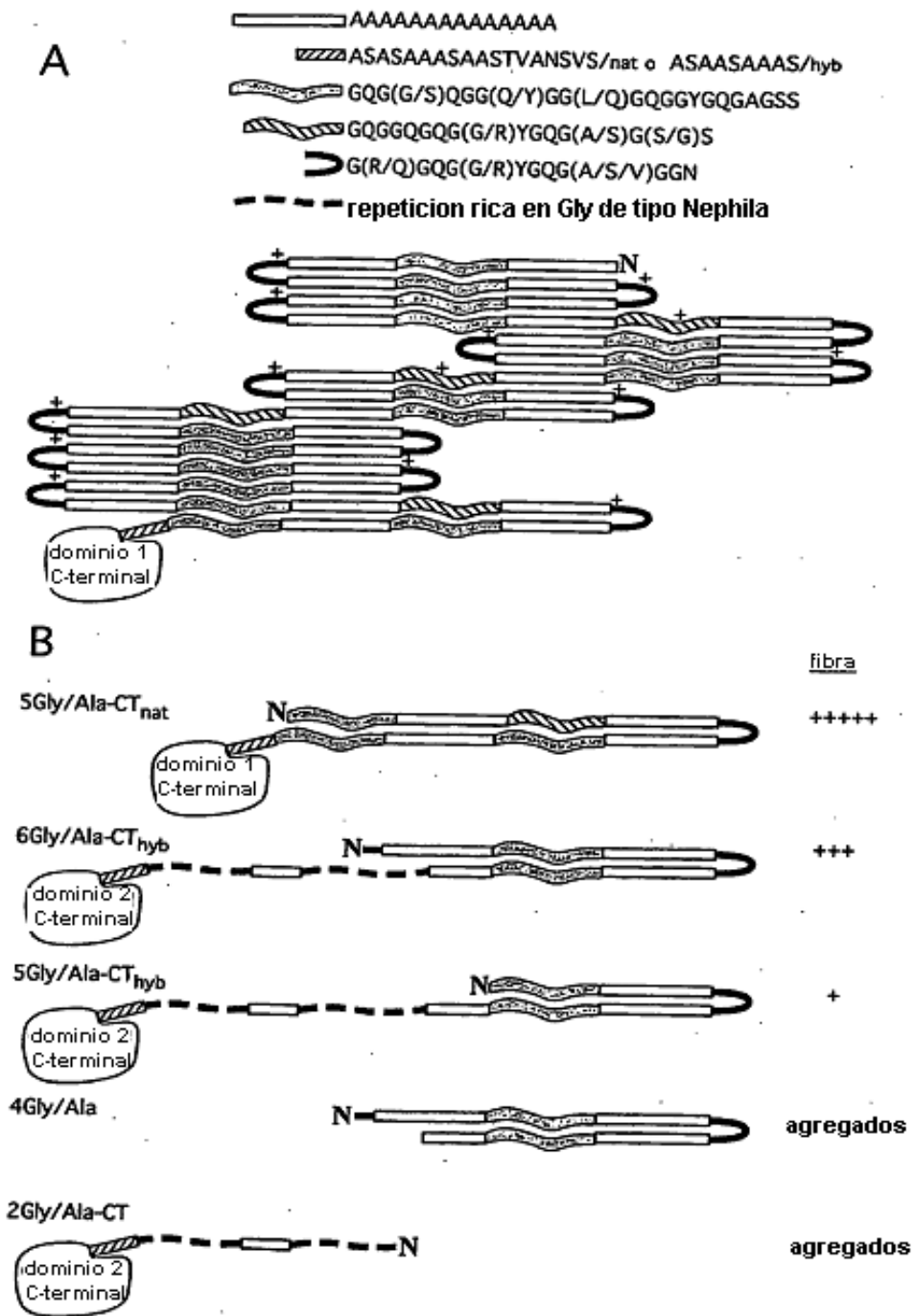
**X** es un fragmento de proteína que comprende un compañero de fusión y un sitio de reconocimiento para el agente de escisión;

en donde el fragmento combinado de proteína **REP-CT** se acopla a través de dicho sitio de reconocimiento para el agente de escisión a dicho compañero de fusión.

8. Un polímero de una proteína espidroína de ampulácea mayor de acuerdo con una cualquiera de las reivindicaciones 1-6.
9. Un polímero de acuerdo con la reivindicación 8, en donde dicho polímero tiene la forma de una fibra, una espuma, un gel, una malla o una película.
- 5 10. Un polímero de acuerdo con la reivindicación 9, en donde dicho polímero tiene la forma de una fibra.
11. Una composición que comprende una proteína aislada de acuerdo con una cualquiera de las reivindicaciones 1-7, en donde el contenido de LPS y otros pirógenos es 1 UE/mg de proteína aislada o inferior.
12. Una molécula de ácido nucleico aislada que comprende una secuencia de ácido nucleico seleccionada del grupo que consiste en el SEQ ID NO: 1; secuencias de ácido nucleico que codifican los SEQ ID NOS: 2-16, y sus secuencias de ácido nucleico complementarias; y secuencias de ácido nucleico que codifican una proteína de fusión de acuerdo con la reivindicación 7, y sus secuencias de ácido nucleico complementarias.
- 10 13. Un método para producir una proteína de fusión soluble como se define en la reivindicación 7, que comprende las etapas de:
- (i) expresar una molécula de ácido nucleico que codifica una proteína de fusión de acuerdo con la reivindicación 7 en un anfitrión adecuado; y
- 15 (ii) aislar la proteína de fusión soluble obtenida en la etapa (i), que implica opcionalmente la eliminación de los LPS y otros pirógenos.
14. Un método para producir un polímero de una proteína espidroína de ampulácea mayor como se define en una cualquiera de las reivindicaciones 1-6, que comprende las etapas de:
- (i) proporcionar una disolución de una proteína de fusión de acuerdo con la reivindicación 7 en un medio líquido,
- 20 (ii) añadir a dicho medio líquido un agente de escisión adecuado para lograr la escisión de la proteína de fusión en el sitio de reconocimiento para el agente de escisión, y obtener de ese modo la proteína espidroína de ampulácea mayor;
- (iii) permitir que la proteína espidroína de ampulácea obtenida en la etapa (ii) se polimerice en el medio líquido; y opcionalmente
- 25 (iv) aislar el polímero obtenido en la etapa (iii) de dicho medio líquido, que implica opcionalmente la eliminación de los LPS y otros pirógenos.
15. Un método de acuerdo con la reivindicación 14 para producir un polímero de una proteína espidroína de ampulácea mayor como se define en las reivindicaciones 1-6, en donde la etapa (i) de proporcionar una disolución de una proteína de fusión de acuerdo con la reivindicación 7 consiste en:
- (a) expresar una molécula de ácido nucleico que codifica una proteína de fusión de acuerdo con la reivindicación 7 en un anfitrión adecuado;
- (b) aislar la proteína de fusión soluble obtenida en la etapa (i), que implica opcionalmente la eliminación de los LPS y otros pirógenos; y
- 35 (c) proporcionar una disolución de dicha proteína de fusión soluble obtenida en la etapa (ii) en un medio líquido.
16. Un método de acuerdo con una cualquiera de las reivindicaciones 14-15, en donde dicha etapa (iii) de permitir que la proteína espidroína de ampulácea mayor obtenida en la etapa (ii) se polimerice en el medio líquido, comprende adicionalmente proporcionar una interfase entre dicho medio líquido y otra fase seleccionada del grupo que consiste en una fase gaseosa, una fase líquida y una fase sólida, en donde dicha polimerización se inicia en dicha interfase o en una región que rodea dicha interfase.
- 40 17. Un método de acuerdo con la reivindicación 16, en donde dicho medio líquido es un medio acuoso y dicha otra fase se selecciona del grupo que consiste en aire y disolventes orgánicos miscibles con agua.
18. El uso de una molécula de ácido nucleico aislada de acuerdo con la reivindicación 12 para la fabricación de un gen no natural que codifica una proteína de seda de araña.
- 45 19. El uso de una proteína espidroína de ampulácea mayor, proteína de fusión, polímero o composición de acuerdo con una cualquiera de las reivindicaciones 1-11 para la fabricación de un producto para diseñar tejidos o para la regeneración celular guiada.

20. El uso de un polímero de acuerdo con una cualquiera de las reivindicaciones 8-10 como soporte para la adherencia y el crecimiento celulares.





**FIG 2**





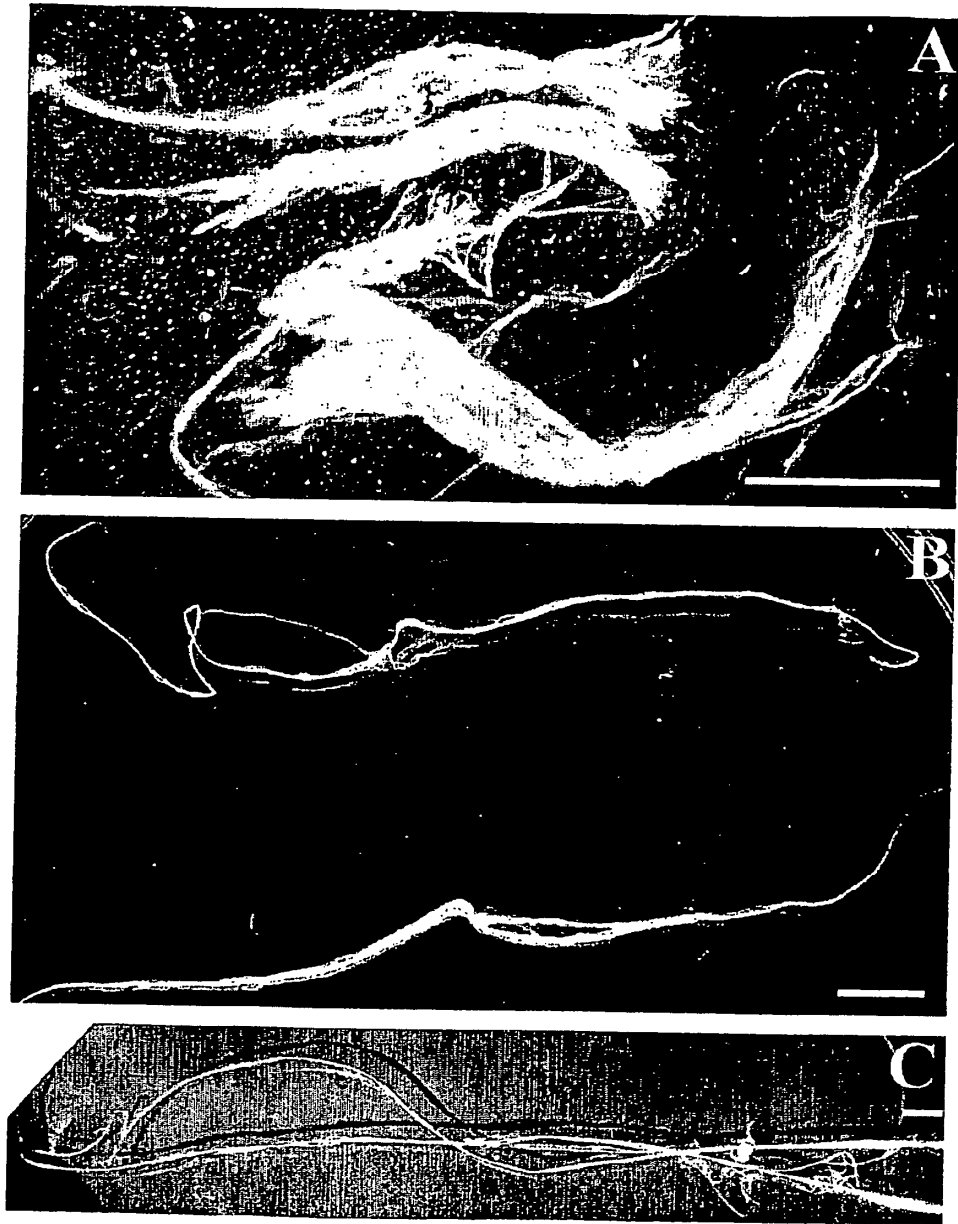


FIG 4

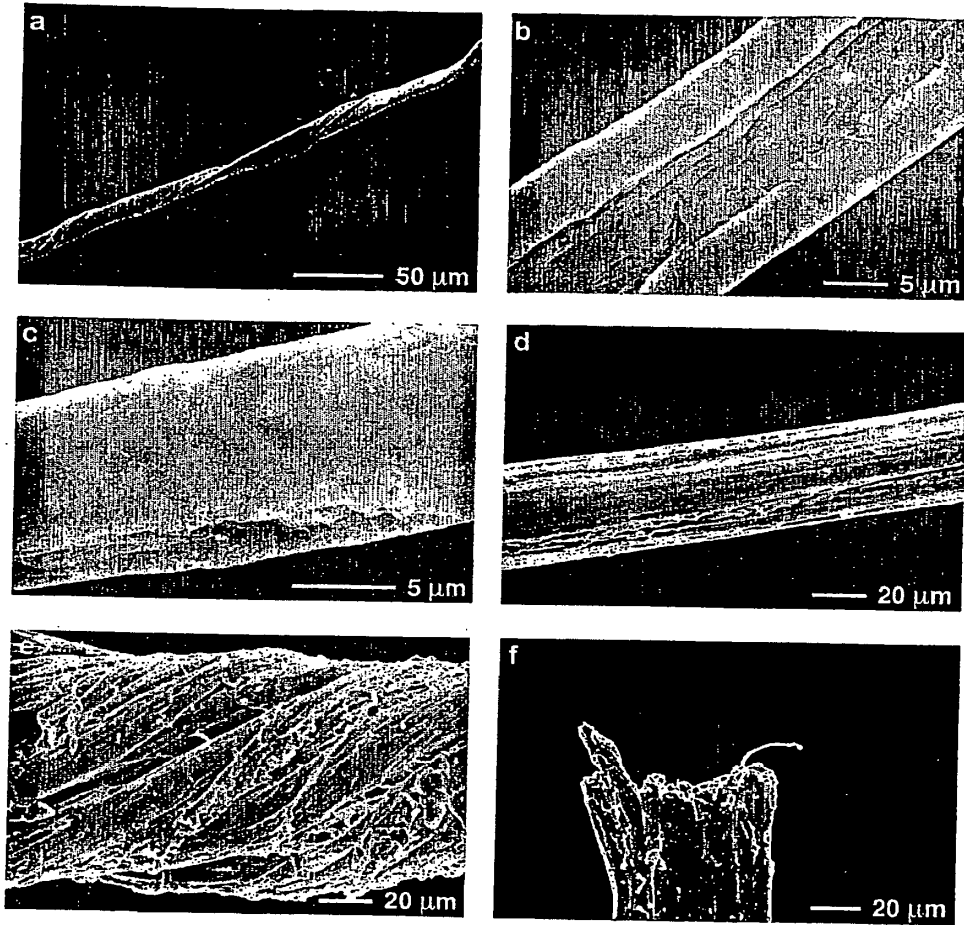


FIG 5

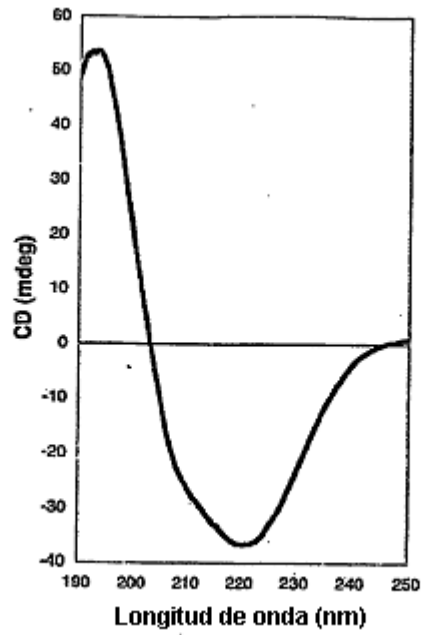


FIG 6

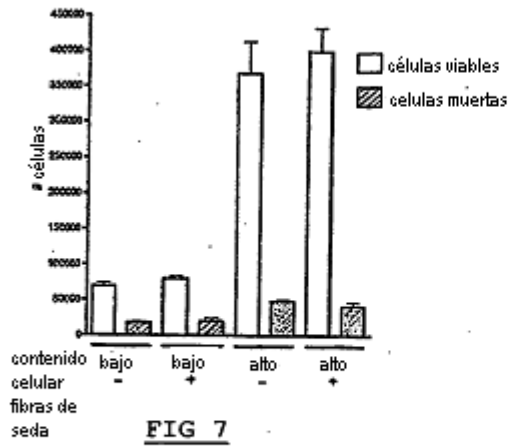


FIG 7

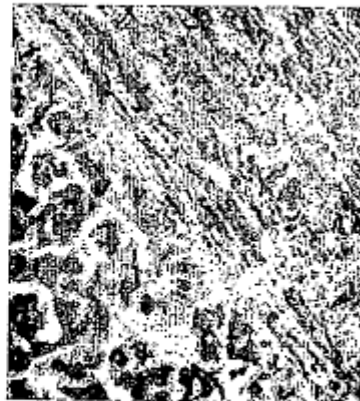


FIG 8

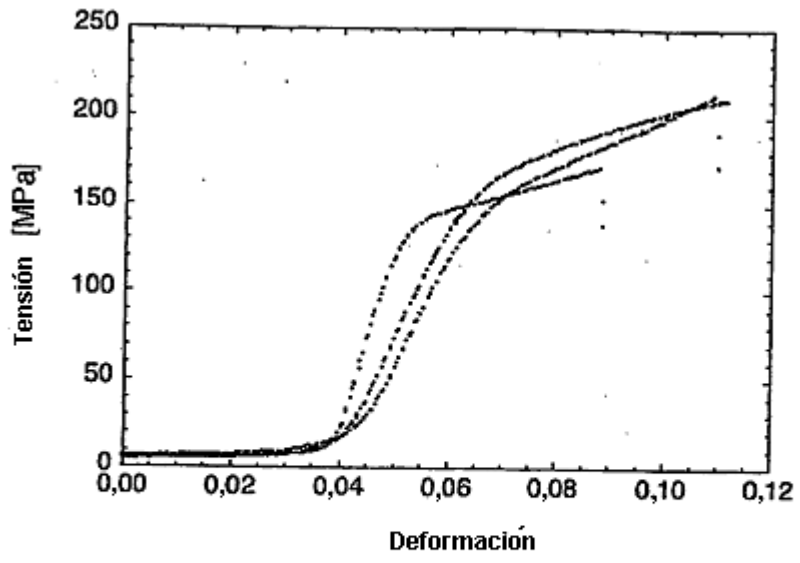


FIG 9

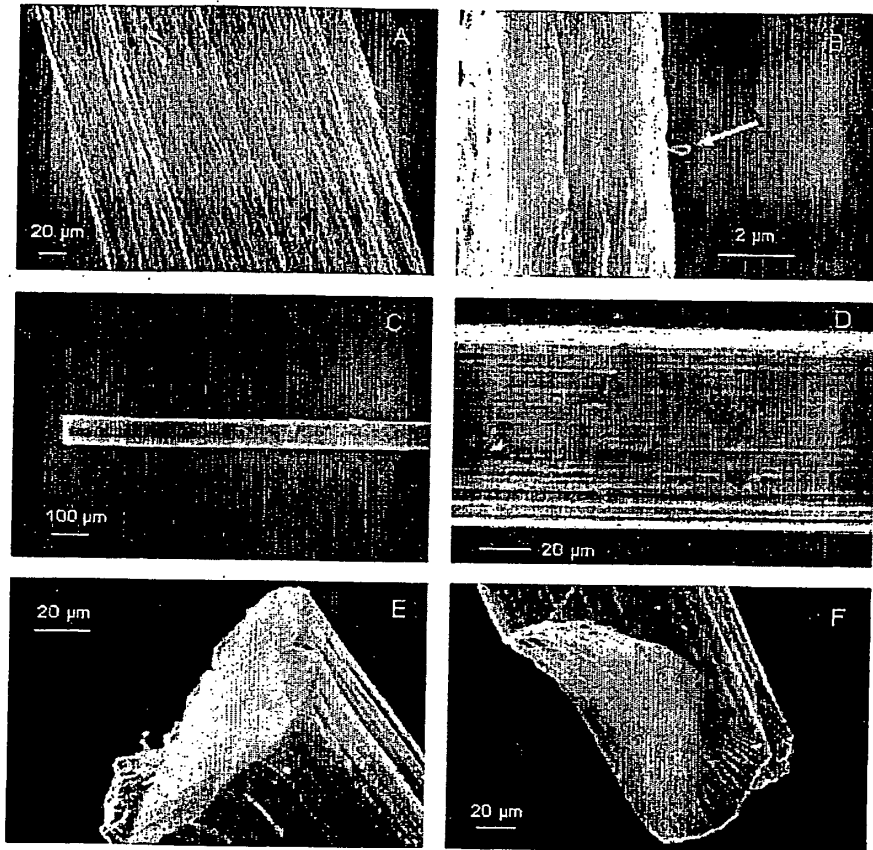


FIG 10