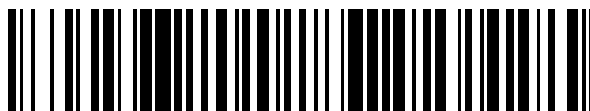


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 525 095**

51 Int. Cl.:

C07K 14/435 (2006.01)

D01F 4/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.07.2005 E 05763003 (0)**

97 Fecha y número de publicación de la concesión europea: **10.09.2014 EP 1773875**

54 Título: **Proteínas recombinantes de la seda de araña**

30 Prioridad:

22.07.2004 US 590196 P

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

17.12.2014

73 Titular/es:

**AMSILK GMBH (100.0%)
AM KLOPFERSPITZ 19
82152 PLANEGG/MARTINSRIED, DE**

72 Inventor/es:

**SCHEIBEL, THOMAS;
HUEMMERICH, DANIEL y
ACKERSCHOTT, CHRISTIAN**

74 Agente/Representante:

LAZCANO GAINZA, Jesús

ES 2 525 095 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 recombinantes de la seda de araña

5 La presente invención se dirige a proteínas recombinantes de la seda de araña, ácidos nucleicos, que codifican para estas proteínas recombinantes de la seda de araña, así como también huéspedes adecuados para la expresión de los ácidos nucleicos. Además, la presente invención se dirige a un método de agregación de proteínas de seda de araña y el uso de las proteínas en el campo de la biotecnología y/o medicina y otros campos industriales, en particular en la fabricación de piezas de automóviles, en la construcción de aviones, en el procesamiento de textiles y cuero, así como en la fabricación y procesamiento de papel, cosméticos, alimentos, dispositivos electrónicos, administración de fármacos y similares.

En esta solicitud se usan las siguientes abreviaturas:

15 NR, no repetitivo; Ap^r, gen de resistencia a ampicilina; IPTG, isopropilo-β-D-tiogalactosido; GdmCl, cloruro de guanidinio; GdmSCN, tiocianato de guanidinio; SDS, dodecilsulfato de sodio; PAGE, electroforesis en gel de poliacrilamina; Tris, tris(hidroximetil)aminometano; CD, dicroísmo circular; rep-proteínas, proteínas repetitivas; Da, Dalton; cps, cuentas por segundo; MRW, peso promedio por residuo; n.d., no determinado.

20 Las sedas de araña son polímeros de proteínas que muestran propiedades físicas extraordinarias (1). Entre los diferentes tipos de sedas de araña, las draglines son las más intensamente estudiadas. Las sedas Dragline son utilizadas por las arañas que tejen orbe para construir el marco y los radios de las redes, como líneas de vida que se arrastran detrás de forma permanente. Para estos fines se requiere una alta resistencia a la tracción y elasticidad. La combinación de tales propiedades da como resultado una dureza que es mayor que la de la mayoría de otros materiales conocidos (1;2). Las sedas Dragline se componen generalmente de dos proteínas principales cuyas estructuras primarias comparten una arquitectura común repetitiva (3;4).

30 Las variaciones de una sola unidad de repetición, que puede comprender hasta 60 aminoácidos, son iterados varias veces para representar la parte más grande de una secuencia de seda dragline. Estas unidades de repetición comprenden un conjunto limitado de motivos de aminoácidos diferentes. Un motivo que se encuentra en las unidades de repetición de la seda dragline es un bloque típicamente de 6 - 9 residuos de alanina. En los hilos de seda muchos motivos de poli-alanina forman pilas de láminas β cristalinas que conducen a resistencia a la tracción (5;6).

35 Los motivos ricos en glicina tales como GGX o GPGXX adoptan estructuras helicoidales flexibles que conectan las regiones cristalinas y proporcionan elasticidad al hilo (7).

Además, todas las proteínas de la seda dragline investigadas comprenden regiones en sus extremos carboxilo que no muestran ningún patrón de repetición obvio (regiones no repetitivas o NR). Hasta el momento ninguna función podría asignarse a estas regiones en el hilo final.

40 El ensamblaje de la seda *in vivo* es un proceso notable. Las proteínas de la seda de araña dragline se almacenan en concentraciones de hasta 50 % (p/v) (8) en la denominada glándula ampulácea mayor. Aunque se ha propuesto una "estructura helicoidal suelta dinámica" para las proteínas dentro de la glándula ampulácea mayor(8), datos más recientes sugieren una conformación aleatoria para las proteínas de la denominada Zona A, que representa la mayor parte de la glándula (9;10). La solución proteica altamente concentrada forma la base de seda (solución de hilatura), que muestra propiedades de un cristal líquido (11-13).

50 El ensamblaje del hilo se inicia durante un paso de la base a través del conducto de hilatura acompañado de extracción de agua, sodio y cloruro (14;15). Al mismo tiempo las concentraciones de los iones más liotrópicos potasio y fosfato aumentan y el pH desciende de 6.9 a 6.3 (14-16). Finalmente el ensamblaje se desencadena por tensión mecánica, que es causada al tirar el hilo hacia fuera del abdomen de la araña (17).

Para varios propósitos los hilos de seda natural no se pueden utilizar directamente, sino que tienen que ser disueltos y reagrupados en otras morfologías tales como películas, espumas, esferas, nanofibras, hidrogeles y similares.

55 La mayoría de las investigaciones relativas a películas fabricadas a partir de proteínas de la seda se han realizado con fibroína de seda, el componente proteico principal de la seda del gusano de la seda *Bombyx mori*. Las películas de fibroína de seda pueden moldearse a partir de soluciones acuosas o a partir de soluciones que contienen hexafluoroisopropanol (HFIP), ácido fórmico, y ácido trifluoroacético. En solución, las fibroínas de seda tienden a adoptar conformaciones helicoidales o aleatorias, dependiendo del solvente usado. Cuando se moldea en películas, las proteínas mantienen la conformación del estado soluble o adoptan una conformación más rica en lámina β. En la mayoría de los casos el

procesamiento de las películas con metanol conduce a un aumento adicional del contenido de láminas β y de la cristalinidad. Además de fibroína de seda, se han empleado además otras proteínas de la seda para moldear películas. Vollrath y colaboradores investigaron películas fabricadas de proteínas extraídas de la glándula ampulácea mayor de la araña *Nephila senegalensis*. Las películas moldeadas contenían principalmente proteínas en una conformación aleatoria cuando se prepararon a partir de una solución acuosa. Sus estructuras cambiaron a lámina β tras la adición de cloruro de potasio. Adicionalmente, se han fabricado películas a partir de una proteína de seda sintética derivada de la proteína de seda dragline MaSpl de la araña *Nephila clavipes* mediante el uso de HFIP como solvente. En solución la proteína adoptó una estructura de α -hélice que cambia a una conformación más rica en lámina β cuando se moldea en una película.

Desafortunadamente, la generación de materiales de película funcionales de fibroína de seda natural está restringida por su secuencia de aminoácidos. La modificación química selectiva de la fibroína de seda sólo es posible en una extensión muy limitada debido a la baja abundancia (<1.5%) de cadenas laterales de aminoácidos químicamente reactivos que contienen grupos tiol, amino o carboxilo. Adicionalmente, la modificación genética dentro del huésped natural para alterar las proteínas de la seda y por lo tanto las propiedades de las películas es tedioso.

Aunque se han desentrañado algunos aspectos estructurales de las proteínas de la seda de araña, todavía se sabe poco sobre la contribución de las proteínas de seda individuales y sus elementos de estructura primaria para el proceso de ensamble. Los estudios comparativos de las dos principales proteínas de seda dragline de la araña de jardín *Araneus diadematus*, ADF-3 y ADF-4, revelaron que, aunque sus secuencia de aminoácidos son bastante similares (4), ellas muestran características de solubilidad y de ensamblaje notablemente diferentes: Mientras ADF-3 es soluble incluso a altas concentraciones (18), ADF-4 es virtualmente insoluble y se auto ensambla en estructuras filamentosas bajo condiciones específicas (resultados no publicados).

El interés científico y comercial inició la investigación de la fabricación a escala industrial de la seda de araña. La producción de seda de araña nativa es poco práctica debido al canibalismo de las arañas, y la producción artificial ha encontrado problemas para lograr tanto un rendimiento suficiente de proteínas como un ensamblaje del hilo con calidad. La expresión bacteriana arrojó niveles de proteína bajas, probablemente causados por un uso de codones diferentes en bacterias y en arañas. Los genes sintéticos con un uso de codones adaptado para el huésped de expresión llevaron a rendimientos más altos, pero las proteínas sintetizadas de los mismos mostraron características diferentes en comparación con las sedas de araña nativas. La expresión de los ADNc de seda dragline parciales en líneas celulares de mamífero sí rindió proteínas de la seda (por ejemplo ADF-3) que podría hilarse artificialmente en hilos 'de seda', aunque todavía de calidad inferior.

WO03060099 se refiere a métodos y dispositivos para la hilatura de proteínas de biofilamentos en fibras. Esta invención es particularmente útil para la hilatura proteínas de seda recombinantes a partir de soluciones acuosas y mejorar la resistencia de las fibras y la viabilidad de fabricación para hacer que la producción comercial y el uso de tales fibras sea practicable. En la misma, se describe la expresión de proteínas de seda de araña en células de mamífero, por ejemplo, células de la glándula mamaria de cabra transgénica.

La expresión de los genes de seda de araña auténticos en huéspedes bacterianos es - como se mencionó anteriormente - ineficiente (24) ya que algunas secciones de genes contienen codones no eficazmente traducidos en las bacterias. Además, la manipulación de genes y la amplificación por PCR es difícil debido a la naturaleza repetitiva de las sedas. Para investigar las propiedades de las proteínas de seda de araña, las estrategias de clonación se emplean usando módulos de ADN sintético con un uso de codones adaptado al huésped de expresión correspondiente. Se obtuvieron genes sintéticos que codificaban para las proteínas que se asemejan a las regiones repetitivas de las sedas de araña (25-28). Sin embargo, ninguno de estos diseños de proteínas incluyó las regiones NR del carboxilo terminal que se encuentran en todas las sedas dragline.

Por lo tanto, es un objeto subyacente de la presente invención proporcionar proteínas de seda de araña recombinante que tengan características mejoradas, en particular, la capacidad de mejorada de expresarse con alto rendimiento y mejorar la resistencia y la flexibilidad, es decir, una mejor calidad. Además, es un objetivo de la presente invención proporcionar proteínas recombinantes de la seda de araña, que puedan expresarse convenientemente en sistemas de expresión ya conocidos. Es además un objetivo de la presente invención proporcionar un método mejorado para la agregación de las proteínas de seda de araña y un método para formar hilos fabricados de esas proteínas. Adicionalmente, es un objetivo de la presente invención proporcionar productos de papel, textil y de piel mejorados. Objetos adicionales son proporcionar nuevas proteínas y otros materiales basados en proteínas de seda de araña tales como esferas, nanofibrillas, hidrogeles, hilos, espumas, películas para usar en aplicaciones en biotecnología, medicina, farmacéutica y alimentarias, cosméticos, en dispositivos electrónicos y para otros fines comerciales.

Estos objetos se resuelven mediante la materia objeto de las reivindicaciones independientes. Las modalidades preferidas se exponen en las reivindicaciones dependientes.

El presente enfoque de ingeniería de proteínas, que proporciona proteínas recombinantes de seda de araña, que comprende o que consiste de secuencias de proteínas de seda de araña repetitivas sintéticas y/o regiones NR- auténticas (no repetitivas), revela que las proteínas muy parecidas a las proteínas de seda auténticas se pueden producir con altos rendimientos. En particular, el sistema de expresión bacteriana, así como el proceso de purificación simple y barato proporcionado en este documento, que se puede escalar fácilmente, proporciona la base para la producción a escala industrial rentable de proteínas del tipo de seda de araña.

Las proteínas de seda de araña principalmente han sido investigadas en cuanto a su contribución a las propiedades mecánicas del hilo de seda. Sin embargo, poco se sabe acerca de los mecanismos moleculares de ensamble de la seda. Como una primera etapa hacia la caracterización de este proceso, los inventores identificaron elementos de la estructura primaria de las proteínas principales de la seda dragline ADF-3 y ADF-4 de la araña de jardín (*Araneus diadematus*) que determinan la solubilidad de la proteína. Además, se investigó la influencia de las condiciones implicadas en la mediación de ensamblaje del hilo natural, en la agregación de proteínas. Se generaron genes que codifican para proteínas similares a la seda de araña mediante el uso de una estrategia de clonación de nuevo desarrollo, que se basa en una combinación de módulos de ADN sintéticos y secuencias génicas auténticas amplificadas por PCR. La comparación de las propiedades de la estructura secundaria, solubilidad y agregación de las proteínas sintetizadas reveló que los elementos estructurales primarios individuales tienen diversas influencias sobre las características de las proteínas. Las regiones repetitivas que representan la mayor parte de las proteínas de la seda dragline determinaron la solubilidad de las proteínas sintéticas, que difirieron mucho entre las construcciones derivadas de ADF-3 y ADF-4. Los factores, tales como la acidificación y el aumento de la concentración de fosfato, que promueven el ensamblaje de la seda *in vivo*, generalmente disminuyeron la solubilidad de las proteínas de la seda *in vitro*. De manera sorprendente este efecto fue pronunciado en las proteínas modificadas genéticamente que comprendían las regiones no repetitivas del carboxilo terminal de ADF-3 o ADF-4, lo que indica que estas regiones juegan un papel importante en la iniciación del ensamblaje de las proteínas de la seda de araña.

De acuerdo con un primer aspecto, la presente invención se dirige a una proteína recombinante de seda de araña que comprende una o más secuencias de proteínas de seda de araña repetitivas sintéticas.

(a) en donde la secuencia repetitiva sintética comprende entre 5 a 50 unidades de repetición, y en donde una unidad de repetición consiste de la secuencia de aminoácido de la sec. con núm. de ident.: 5 o variantes de esta, o

(b) en donde la secuencia repetitiva sintética comprende entre 10 a 50 unidades de repetición, y en donde una unidad de repetición consiste de la secuencia de aminoácidos de la sec. con núm. de ident.: 3 o variantes de esta y se combina con una unidad de repetición que consiste en la secuencia de aminoácido de la sec. con núm. de ident.: 4 o variantes de esta,

en donde las variantes en cada caso tienen un sustitución de aminoácido, deleciones de aminoácido y/o inserciones de aminoácido, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácido consisten en entre 1 a 2 deleciones de aminoácido y/o las inserciones de aminoácido consisten en entre 1 a 2 inserciones de aminoácido.

El término "secuencia repetitiva sintética" como se usa en la presente se entiende como secuencia de proteína recombinante, que no puede encontrarse en la naturaleza, que sin embargo, se deriva de unidades de repetición, que son de origen natural en proteínas de seda de araña. Como se indicó anteriormente, las secuencias repetitivas comprenden una o más unidades de repetición individuales, que comprenden hasta 60 aminoácidos. Las unidades de repetición de origen natural comprenden un conjunto limitado de motivos de aminoácido distintos. Las unidades de repetición confieren entre otros resistencia a la tracción y elasticidad al hilo, que pueden formarse más tarde a partir de la proteína de seda de araña.

Los diferentes tipos de unidades de repetición, que pueden formar la base para la secuencia repetitiva sintética de la invención, se explicarán en detalle más abajo.

El segundo componente de la proteína recombinante de seda de araña de la invención, que puede estar presente en adición a la secuencia repetitiva sintéticas o solo, comprende una o más secuencias de proteína no repetitivas auténticas. Estas secuencias no repetitivas desempeñan un papel funcional importante en el ensamble de hilos.

Se observa que en la presente invención, también se contemplan las proteínas recombinantes de seda de araña, que sólo comprenden secuencias repetitivas sintéticas. Aunque las proteínas recombinantes de la invención que muestran los dos componentes, es decir secuencias repetitivas sintéticas así como secuencias no repetitivas auténticas, tienen una variedad más amplia de utilidad y pueden producirse en mayores cantidades (ver el capítulo de Ejemplos más abajo), las proteínas recombinantes de la seda de araña que sólo tienen secuencias repetitivas sintéticas incluidas puede utilizarse para algunas aplicaciones específicas.

Estas aplicaciones son - entre otras - piezas de automóviles y aviones, revestimientos de superficies, así como sistemas de cierre de heridas y vendajes de heridas. O en otras palabras, las aplicaciones, en las que no se requieren estructuras de hilos de proteínas de seda de araña.

5

El término "auténtico" como se usa en la presente significa que las secuencias de ácidos nucleicos subyacentes se aíslan de su medio ambiente natural sin realizar enmiendas sustanciales en la secuencia en sí. La única modificación, que se acepta que se produzca, es donde la secuencia de ácido nucleico no repetitiva auténtica es modificada con el fin de adaptar dicha secuencia a la expresión en un huésped sin cambiar la secuencia de aminoácidos codificada. Las secuencias preferidas son NR3 (sec. con núm. de ident.: 10; derivada de ADF-3) y NR4 (sec. con núm. de ident.: 11; derivada de ADF-4). En las dos secuencias, para una traducción eficiente, el codón AGA (Arg), que se traduce rara vez en *E. coli*, se mutó a CGT (Arg) usando mutagénesis por PCR.

10

Secuencias no repetitivas auténticas preferidas de proteínas flageliformes son la secuencia de aminoácidos y la secuencia de ácido nucleico de FlagN-NR (sec. con núms. de ident.: 31 y 32) y FlagC-NR (sec. con núms. de ident.: 33 y 34).

15

De acuerdo con una modalidad preferida, las proteínas recombinantes de la seda de araña de la invención generalmente se derivan de proteínas dragline de araña a partir de glándula ampulácea mayor de la araña y/o a partir de proteínas derivadas de la glándula flageliforme.

20

De acuerdo con otra modalidad preferida, las secuencias no repetitivas auténticas se derivan de la región no repetitiva amino terminal (proteínas flageliformes) y/o la región no repetitiva carboxilo terminal (proteínas flageliformes y dragline) de una proteína de seda de araña de origen natural. Los ejemplos preferidos de estas proteínas se indicarán más abajo.

25

Generalmente se prefiere seleccionar las secuencias dragline y/o flageliformes a partir de proteínas dragline y/o flageliformes de arañas con telas en espiral (Araneidae y Araneoides).

Con mayor preferencia las proteínas dragline y/o proteínas flageliformes se derivan de una o más de las siguientes arañas: Arachnura higginsi, Araneus circulissparsus, Araneus diadematus, Argiope picta, araña de jardín de bandas (Argiope trifasciata), Araña de tela en oro de batik (Nephila antipodiana), araña telarañosa de Beccari (Cyrtophora beccarii), araña "caca de pájaro" (Celaenia excavata), araña tejedoras espinosas blanca y negra (Gasteracantha kuhlii), araña de jardín negra y amarilla (Argiope aurantia), araña de bolas (Ordgarius furcatus), araña de bolas - araña magnificente (Ordgarius magnificus), araña marrón (Neoscona nautica), araña de patas marrón (Neoscona rufofemorata), araña de cabeza negra (Zygiella calyptata), araña común de jardín (Parawixia dehaani), tejedora de orbe común (Neoscona oxancensis), tejedora de orbe cangrejo espinoso (Gasteracantha cancriformis (elipsoides)), araña espinosa curvada (Gasteracantha arcuata), Cyrtophora moluccensis, Cyrtophora parnasia, Dolophones conifera, Dolophones turrigera, araña espinosa de Doria (Gasteracantha doriae), Araña tejedora espinosa asiática (Gasteracantha mammosa), Araña telarañosa de dos colas (Cyrtophora exanthematica), Aculeperia ceropegia, Eriophora pustulosa, Flat Anepsion (Anepsion depressum), Araña tejedora cuadriespinosa (Gasteracantha quadrispinosa), Araña del tejedor del orbe del jardín (Eriophora transmarina), Araña tejedora de orbe gigante (Araneus bicentenarius), Arañas de seda dorada (Nephila maculata), araña espinosa de Hasselt (Gasteracantha hasseltii), Tegenaria atrica, Heurodes turriga, Araña cyclosa de las islas (Cyclosa insulana), araña joya (Astracantha minax), tejedor del orbe pálido (Araneus mitificus), araña de jardín de Laglaise (Eriovixia laglaisei), Araña cyclosa bifida (Cyclosa bifida), araña Malabar (Nephilengys malabarensis), araña cruz de San Andrés multi-coloreada (Argiope versicolor), Araña de tronco de árbol ornamental (Herennia ornatissima), araña cruz de San Andrés ovalada (Argiope aemula), araña tejedora roja (Cyrtophora unicolor), araña tejedora rusa (Cyrtophora hirta), araña cruz de San Andrés (Argiope keyserlingi), Scarlet Acusilas (Acusilas coccineus), Argiope plateada (Argiope argentata), araña tejedora con espalda espinosa (Gasteracantha cancriformis), tejedora de orbe manchada (Neoscona domiciliorum), cruz de San Andrés (Argiope aetheria), Araña Cruz de San Andrés (Argiope Keyserlingi), araña tocón de árbol (Poltys illepidus), araña triangular (Arkys clavatus), araña triangular (Arkys lancearius), araña doble espinada (Poecilopachys australasia), especies Nephila, por ejemplo Nephila clavipes, Nephila senegalensis, Nephila madagascariensis y muchas más (para otras especies de araña, ver más abajo). Las más preferidas, las proteínas de dragline se derivan de Araneus diadematus y las proteínas flageliformes se derivan de Nephila clavipes.

30

35

40

45

50

En el contexto de esta invención, debería ser evidente que una proteína recombinante de seda de araña no sólo puede comprender secuencias de proteína de una especie, sino que también puede contener secuencias derivadas de diferentes especies de arañas. Como un ejemplo, la una o más secuencias de proteínas de seda de araña repetitivas sintéticas pudieran derivarse de una especie, la una o más secuencias de proteína de seda de araña repetitivas no auténticas de otra. Como un ejemplo adicional, también es posible diseñar una proteína de seda de araña recombinante, que contiene más de un tipo de una secuencia repetitiva, en la que los diferentes tipos se derivan de diferentes especies.

55

De acuerdo con una modalidad preferida, la proteína dragline es ADF-3, ADF-4, MaSp I, MaSp II de tipo silvestre y la proteína flageliforme es FLAG. El término ADF-3/-4 se usa en el contexto de proteínas MaSp producidas por *Araneus diadematus* (fibroína de *Araneus diadematus* -3/-4). Ambas proteínas, ADF-3 y -4 pertenecen a la clase de proteínas MaSp II (espidroína II ampulácea mayor).

La fibra de seda tiene regiones cristalinas de lámina β intercaladas con segmentos amorfos elásticos similares a los polímeros cristalinos líquidos. Estos dos segmentos están representados por dos clases diferentes de proteínas, MaSp I (espidroína I ampulácea mayor) y MaSp II (espidroína II ampulácea mayor) codificada por genes diferentes.

Otras proteínas de seda de araña, que se pueden utilizar en esta invención (es decir, solo o en combinación con otras proteínas) y sus números de acceso de base de datos son:

espidroína 2 [*Araneus bicentarius*][gi|2911272
 proteína-1 de seda dragline de la glándula ampulácea mayor[*Araneus ventricosus*] gi|27228957
 proteína-2 de seda dragline de la glándula ampulácea mayor[*Araneus ventricosus*][gi|27228959 ampullate
 espidroína I [*Nephila madagascariensis*][gi|13562006
 espidroína 1 de ampulácea mayor[*Nephila senegalensis*][gi|13562010
 espidroína 1 de ampulácea mayor[*Latrodectus geometricus*][gi|13561998
 espidroína 1 de ampulácea mayor[*Argiope trifasciata*][gi|13561984
 espidroína 1 de ampulácea mayor[*Argiope aurantia*][gi|13561976
 espidroína 2 de la proteína de la seda dragline [*Nephila clavata*][gi|16974791
 espidroína 2 de ampulácea mayor[*Nephila senegalensis*][gi|13562012
 espidroína 2 de ampulácea mayor[*Nephila madagascariensis*][gi|13562008
 espidroína 2 de ampulácea mayor[*Latrodectus geometricus*][gi|13562002

Sin embargo, también las secuencias flageliformes y ya conocidas publicadas pueden usarse en la presente, en particular, la siguiente:

Cds parciales de proteínas de seda flageliformes [*Nephila clavipes*][gi|2833646
 Cds parciales de proteínas de seda flageliformes [*Nephila clavipes*][gi|2833648

La secuencia consenso que contiene polialanina se deriva de ADF-3 y tiene la secuencia de aminoácidos de sec. con núm. de ident.: 3 (módulo A) o una variante de esta como se definió anteriormente. El módulo A contiene una polialanina que tiene 6 residuos de alanina. Una secuencia consenso que contiene polialanina adicionalmente preferida, derivada de ADF-4, es el módulo C (sec. con núm. de ident.: 5), que contiene 8 residuos de alanina.

En la proteína recombinante de seda de araña de la invención, la secuencia repetitiva sintética se deriva de ADF-3 y comprende una o más repeticiones de secuencia de aminoácido de la sec. con núm. de ident.: 4 (módulo Q) o una variante de esta como se definió anteriormente.

Cabe destacar que los módulos específicos para la secuencia repetitiva sintética de la invención pueden combinarse además entre sí, por ejemplo, módulos (unidades de repetición) que combinan A y Q, Q y C etc. también están abarcados por la presente invención. El número de los módulos que se introducirá en la proteína de seda de araña está en el intervalo de 5-50 módulos, con mayor preferencia 10-40 y con la máxima preferencia entre 15-35 módulos.

La secuencia repetitiva sintética preferentemente comprende una o más de (AQ) como unidades de repetición. Con la máxima preferencia, la secuencia repetitiva sintética es (AQ)₁₂, o (AQ)₂₄.

Siempre que la secuencia repetitiva sintética se deriva de ADF-4, esta puede comprender preferentemente una o más repeticiones de la secuencia de aminoácido de sec. con núm. de ident.: 5 (módulo C) o una variante de esta, como se mencionó anteriormente, en donde la secuencia repetitiva sintética total es C₁₆ o C₃₂.

Las modalidades preferidas para las proteínas recombinantes de la seda de araña completas de la invención son (AQ)₁₂NR₃, (AQ)₂₄NR₃, C₁₆NR₄ y C₃₂NR₄ es decir proteínas que comprenden o consisten en dichas secuencias.

Se señala que la configuración anterior de la secuencia repetitiva sintética (con el uso del sistema A, Q y C) se aplica además a todas las demás unidades de repetición descritas anteriormente, por ejemplo todas las secuencias que contienen polialanina puede tomarse por A y/o C y todas las secuencias ricas en glicina pueden usarse como Q.

Además, también es posible combinar aquellas secuencias derivadas de ADF-3 y ADF-4 y Flag en una secuencia recombinante.

5 Como se explicó anteriormente, las secuencias de aminoácidos descritas en la presente no se restringen a las secuencias exactas proporcionadas en las sec. con núms. de ident. Las secuencias de aminoácidos indicadas en la presente también comprenden variantes. Por lo tanto, las secuencias de aminoácidos de las proteínas de la presente invención abarcan además todas las proteínas de seda de araña repetitivas sintéticas que comprenden las variantes de las unidades de repetición que consisten en la secuencia de aminoácido de sec. con núm. de ident.: 3, 4, o 5 en donde las variantes en cada caso tienen una sustitución de aminoácido, deleciones de aminoácido y/o inserciones de aminoácido, en donde la sustitución de aminoácidos consiste de 1 sustitución de aminoácido, las deleciones de aminoácido consisten en entre 1 a 2 deleciones de aminoácido y/o las inserciones de aminoácido consisten en entre 1 a 2 inserciones de aminoácido.

15 Las "sustituciones" de aminoácidos son el resultado de la sustitución de un aminoácido por otro aminoácido que tiene propiedades estructurales y/o químicas similares, es decir, sustituciones de aminoácidos conservadoras. Las sustituciones de aminoácidos pueden realizarse sobre la base de similitud en polaridad, carga, solubilidad, hidrofobicidad, hidrofiliidad, y/o la naturaleza anfipática de los residuos en cuestión. Por ejemplo, los aminoácidos no polares (hidrófobo) incluyen alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptófano, y metionina; los aminoácidos polares neutros incluyen glicina, serina, treonina, cisteína, tirosina, asparagina, y glutamina; los aminoácidos positivamente cargados (básicos) incluyen arginina, lisina, e histidina; y los aminoácidos negativamente cargados (ácidos) incluyen ácido aspártico y ácido glutámico.

"Inserciones" o "eliminaciones" están típicamente en el intervalo de aproximadamente 1 a 2 aminoácidos.

25 La variación permitida puede determinarse experimentalmente al producir inserciones, deleciones o sustituciones de aminoácidos de manera sistemática en una proteína mediante el uso de técnicas de ADN recombinante y pruebas de actividad para las variantes recombinantes resultantes. Esto no requiere más que experimentos de rutina para el experto en la técnica.

30 La presente invención, de acuerdo con un segundo aspecto, se dirige a una secuencia de ácido nucleico, que codifica una proteína recombinante de seda de araña como se describió anteriormente.

35 El término "secuencia de ácido nucleico" se refiere a un heteropolímero de nucleótidos o la secuencia de esos nucleótidos. Los términos "ácido nucleico" y "polinucleótido" se usan indistintamente en la presente descripción para referirse a un heteropolímero de nucleótidos.

40 La rigurosidad de la hibridación, como se usa en la presente, se refiere a las condiciones bajo las cuales los dúplex de polinucleótidos son estables. Como es conocido por los expertos en la técnica, la estabilidad del dúplex es una función de la concentración de iones de sodio y la temperatura (véase, por ejemplo, Sambrook y otros, Molecular Cloning: A Laboratory Manual 2o. Ed. (Cold Spring Harbor Laboratory, (1989)). Los niveles de rigurosidad usados para hibridar se pueden variar fácilmente por los expertos en la técnica.

45 Como se usa en la presente, la frase "condiciones moderadamente rigurosas" se refiere a condiciones que permiten que el ADN se una a un ácido nucleico complementario que tiene aproximadamente 60% de identidad, preferentemente aproximadamente 75% de identidad, con mayor preferencia aproximadamente 85% de identidad al ADN; y se prefiere especialmente más de aproximadamente 90% de identidad a dicho ADN. Preferentemente, las condiciones moderadamente rigurosas son condiciones equivalentes a la hibridación en 50% de formamida, 5 x solución de Denhart, 5 x SSPE, 0.2% SDS a 42°C, seguido por lavado en 0.2 x SSPE, 0.2% SDS, a 65°C.

50 De acuerdo con un tercer aspecto, se proporciona un vector que comprende los ácidos nucleicos anteriormente mencionados. Preferentemente, se proporciona un vector de expresión, que comprende dichos ácidos nucleicos. Este vector de expresión comprende preferentemente una o más secuencias reguladoras. El término "vector de expresión" se refiere generalmente a un plásmido o fago o virus o vector, para expresar un polipéptido/proteína de una secuencia de ADN (ARN). Un vector de expresión puede comprender una unidad transcripcional que comprende un ensamblaje de (1) un elemento genético o elementos que tienen un papel regulador en la expresión génica, por ejemplo, los promotores o potenciadores, (2) una secuencia estructural o de codificación, que se transcribe en el ARNm y se traduce en proteína, y (3) las secuencias de iniciación y terminación de la traducción y transcripción adecuadas. Las unidades estructurales destinadas para su uso en sistemas de expresión de levadura o eucariotas preferentemente incluyen una secuencia líder que permite la secreción extracelular de la proteína traducida por una célula huésped. Alternativamente, donde la proteína recombinante se expresa

sin una secuencia líder o de transporte, esta puede incluir un residuo de metionina amino-terminal. Este residuo puede ser escindido posteriormente de la proteína recombinante expresada para proporcionar un producto final.

5 De acuerdo con una modalidad preferida, el vector es un plásmido o un vector viral, que preferentemente es un sistema de baculovirus o un sistema de vector de virus vaccinia. En esta invención pueden usarse además otros sistemas de vectores virales. De un caso a otro, puede necesitarse una modificación del vector. Ejemplos de otros vectores virales son los adenovirus y todos los virus de ARN de cadena negativa, por ejemplo, la rabia, el sarampión, el RSV, etc.

10 De acuerdo con una modalidad preferida, el vector es el vector de clonación pAZL como se define en la Figura 6 o en la sec. con núm. de ident.: 55, o una variante de este como se definió anteriormente. Este vector muestra las siguientes propiedades y ventajas:

1. alta amplificación (más alta que otros vectores de clonación)
- 15 2. permite una construcción controlada y sin problemas de genes sintéticos (no se conoce otro vector que proporcione esta capacidad).

Un cuarto aspecto de la invención comprende un huésped, que se ha transformado con el vector como se definió anteriormente.

20 El huésped puede ser una célula procariota. En este caso se prefieren, *E. coli* o *Bacillus subtilis*.

Además, el huésped puede ser una célula eucariota, preferentemente una célula de mamífero, célula vegetal, célula de levadura o una célula de insecto.

25 La célula de mamífero preferentemente es una célula CHO, COS, HeLa, 293T, HEH o BHK.

Se prefiere usar además una célula de levadura como una célula hospedera, que preferentemente es *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Candida albicans* o *Hansenula polymorpha*.

30 Como células de insectos, las células de insectos Lepidópteros pueden usarse preferentemente, con mayor preferencia células de *Spodoptera frugiperda* y de *Trichoplusia ni*. Con la máxima preferencia, la célula de insecto es una célula Sf9, Sf21 o high-5.

35 Una ventaja del sistema de expresión en células de insecto, por ejemplo con respecto a los sistemas bacterianos, reside en el hecho de que las proteínas producidas están glicosiladas, por lo tanto son un blanco para la degradación por microorganismos. Esta característica puede ser de importancia, por ejemplo, en el campo de la medicina, siempre que las proteínas de la seda estén destinadas a un uso *in vivo*, en que se desea la degradación biológica. Esta característica puede encontrar aplicación particularmente en materiales de sutura y cierre de heridas y en sistemas de cobertura.

40 Cuando el huésped es una célula vegetal, la célula vegetal se deriva preferentemente del tabaco, papa, maíz y tomate.

De acuerdo con un quinto aspecto, se proporciona un método de agregación de proteínas de la seda de araña, que comprende las siguientes etapas:

- 45 a) preparar una solución de proteínas que contiene proteínas de seda de araña no orientadas como se define en la presente descripción;
- b) exponer la solución preparada en a) a un inductor de de agregación; y
- c) recuperar las proteínas de seda de araña precipitadas.

50 Preferentemente, las proteínas de seda de araña usadas en la etapa a) se producen por transformación de un huésped adecuado como se definió anteriormente con un vector o un ácido nucleico descrito en la presente, y que expresa el gen de la seda de araña bajo condiciones adecuadas.

55 El inductor de agregación se selecciona preferentemente de acidificación, preferentemente a un pH de aproximadamente 1, fosfato de potasio y tensión mecánica, preferentemente al rotar la solución de proteína y aplicar fuerzas de cizalla. La etapa desencadenante resultó ser esencial para realizar el método de esta invención.

60 Sorprendentemente los inventores mostraron que en particular los factores desencadenantes mencionados anteriormente mejoraron la agregación de proteínas de la seda de araña, que es un resultado muy deseado particularmente desde el punto de vista industrial. En este sentido se hace referencia al capítulo de "Resultados" más abajo, donde se explica la influencia

de estos factores desencadenantes sobre las proteínas recombinantes de la seda de araña de la invención: la influencia de cada factor desencadenante puede variar entre las diferentes proteínas recombinantes de la seda de araña de esta invención, sin embargo, puede verse como un concepto general que esos factores desencadenantes *in vitro* muestran una influencia inesperadamente alta sobre todas las proteínas recombinantes, que comprenden los componentes de la presente invención, es decir las regiones repetitivas y/o no repetitivas. Adicionalmente, de los resultados proporcionados en la presente puede derivarse que no solo un factor desencadenante, sino además combinaciones de ellos pueden conducir a la mejor manera de agregar las proteínas de la seda de araña de la invención.

Sin embargo, cabe señalar que este método no se limita a las proteínas de la seda de araña de la presente invención, sino que además puede aplicarse a todas las demás proteínas de la seda de araña disponibles, tanto si se producen de manera natural como sintética.

El método comprende además preferentemente la etapa de hilatura de dichas proteínas preparadas en la etapa a) o recuperadas en c) en filamentos, nanofibras e hilos mediante un método adecuado.

Para este propósito, pueden usarse métodos de hilatura, que se conocen per se en la técnica. Por ejemplo, una solución base de proteína de la seda de araña se extrude a través de una tobera de hilatura para formar un biofilamento. El biofilamento resultante puede alargarse o estirarse. Siempre que existan arreglos de moléculas tanto cristalinos como amorfos en los biofilamentos, el alargamiento o estiramiento aplicarán un esfuerzo de cizalla suficiente para orientar las moléculas y hacerlas más paralelas a las paredes del filamento y aumentar la resistencia a la tracción y la dureza del biofilamento.

La solución base puede contener las proteínas de la seda recombinantes de la invención y/o proteínas de la seda auténticas de una o más especies de arañas, o proteínas de la seda de diferentes géneros que producen seda, por ejemplo, una mezcla de proteínas de la seda de arañas y *B. mori*. En las modalidades más preferidas, las proteínas de la seda son sedas dragline y/o flageliformes de *N. clavipes* o *A. diadematus*, particularmente las proteínas MaSpl, MaSpII, ADF-3, ADF-4 y Flag. En modalidades alternativas, la solución base contiene una mezcla de proteínas de la seda y uno o más polímeros sintéticos o proteínas de biofilamentos naturales o sintéticos.

Preferentemente, la solución base es al menos 1%, 5%, 10%, 15% en peso/volumen (p/v) de proteína de la seda. Con mayor preferencia, la solución base es tanto como 20%, 25%, 30%, 35%, 40%, 45%, o 50% p/v de proteína de la seda. En modalidades preferidas, la solución base contiene proteína de la seda de araña sustancialmente pura. En modalidades preferidas, la base tiene un pH de aproximadamente 6.9.

Por "solución base" se entiende cualquier mezcla líquida que contiene proteína de la seda y es susceptible de extrusión para la formación de un biofilamento o moldeado de una película. Las soluciones base también pueden contener, además de los monómeros de proteínas, agregados de orden superior que incluyen, por ejemplo, dímeros, trímeros, y tetrámeros. Normalmente, las soluciones base son soluciones acuosas de pH 4.0-12.0 y que tienen menos de 40% de compuestos orgánicos o agentes caotrópicos (p/v). Preferentemente, las soluciones base no contienen solventes orgánicos o agentes caotrópicos, pero pueden incluir aditivos para mejorar la conservación, la estabilidad, o la facilidad de trabajo con la solución.

Por "filamento" se entiende una fibra de longitud indefinida, que está en el intervalo de longitud microscópica y de nanoescala a longitudes de una milla o mayores. La seda es un filamento natural, mientras que el nailon y el poliéster como ejemplo son filamentos sintéticos.

Más información respecto a cómo hilar las fibras de proteína de seda de araña puede encontrarse en WO03060099 (Karatzas y otros), publicado el 24 de julio de 2003, que se incorpora en la presente como referencia.

Además, las proteínas de seda de araña de la presente invención puede proporcionarse como películas o similares, por ejemplo, como un producto de proteína de seda de araña, para el que no se requiere una etapa de hilatura.

Para una descripción más detallada del proceso de fabricar películas se hace referencia al capítulo de Ejemplos.

Adicionalmente, el método de la presente invención puede incluir preferentemente en la etapa a) y/o c) un método de purificación, que comprende exponer las proteínas de la seda de araña expresadas a una desnaturalización por calor a 60-90, preferentemente 70- 80 °C seguido de la adición de sulfato de amonio de 600-1400 mM, preferentemente 800-1200 mM.

Como ya se ha explicado anteriormente, las proteínas/hilos tal como se define en la presente pueden usarse en el campo de

la biotecnología y/o la medicina, preferentemente para la fabricación de sistemas de cobertura o cierre de heridas, materiales de sutura para su uso en neurocirugía o cirugía oftálmica.

5 Además, las proteínas/hilos pueden usarse preferentemente para la fabricación de materiales de reemplazo, preferentemente materiales artificiales de cartílago o tendón.

10 Adicionalmente, los hilos/fibras de la invención pueden usarse en la fabricación de dispositivos médicos tales como tiras adhesivas médicas, injertos de piel, ligamentos de reemplazo, y malla quirúrgica; y en una amplia variedad de productos industriales y comerciales, tales como tejido para ropa, revestimiento de chaleco antibalas, tejido de recipientes, correas de bolsas o carteras, cable, cuerda, material de unión adhesiva, material de unión no adhesiva, material de precinto, cubiertas y partes de automóviles, material de construcción de aviones, material de impermeabilización, material de partición flexible, equipos deportivos; y, de hecho, en casi cualquier uso de fibra o tejido para el que se desean las características de alta resistencia a la tracción y elasticidad. La presente invención contempla además la adaptabilidad y el uso del producto de fibra estable en otras formas, tales como un revestimiento por rociado en seco, partículas similares a perlas, o su uso en una mezcla con otras composiciones.

15 Se señala explícitamente que las aplicaciones más preferidas de las proteínas de la seda de araña de la presente invención están en la fabricación y procesamiento de tejido para ropa (textiles) y cuero, cubiertas y partes de automóviles, materiales de construcción de aviones así como en la fabricación y procesamiento de papel.

20 Las proteínas recombinantes de la seda de araña de la presente invención pueden añadirse a productos con celulosa y queratina y colágeno y por lo tanto, la presente invención está dirigida además a un papel o producto de cuidado de la piel y cuidado del cabello, que comprende celulosa y/o queratina y/o colágeno y las proteínas de la seda de araña de la presente invención. Los papeles y productos de cuidado de la piel y cuidado del cabello, en los que se incorporan las proteínas de la presente invención muestran características mejoradas, particularmente resistencia a la tracción o resistencia al desgarramiento.

25 Además, las proteínas recombinantes de la seda de araña de la invención pueden usarse como un revestimiento para productos textiles y de cuero, lo que confiere de esta manera estabilidad y durabilidad al producto recubierto. Las proteínas de la seda en particular muestran aplicabilidad para el revestimiento de productos de cuero, ya que en este caso, el curtido y sus efectos negativos para el medio ambiente pueden evitarse o al menos reducirse.

30 A menos que se defina de cualquier otra forma, todos los términos técnicos y científicos que se usan en la presente descripción tienen el mismo significado que el que entiende comúnmente el experto en la técnica a la cual pertenece la invención. Todas las publicaciones, solicitudes de patentes, patentes y otras referencias mencionadas en la presente descripción se incorporan como referencia en su totalidad. En caso de conflicto, la presente descripción, incluyendo las definiciones, controlará. Además, los materiales, métodos, y ejemplos son ilustrativos solamente y no pretenden ser limitantes.

35 La invención se ilustra ahora adicionalmente mediante los ejemplos y los dibujos adjuntos, que muestran lo siguiente:

40 **Figura 1** Estrategia de clonación para construir genes sintéticos de la seda de araña. (A) El casete de clonación comprende los sitios de restricción requeridos para la multimerización de módulos (*BsgI* y *BseRI*) y para la escisión de los genes ensamblados (*NcoI*, *BamHI*, *HindIII*). Durante la construcción génica la región espaciadora se reemplazó por módulos y multímeros de módulos. (B) La conexión, de manera dirigida al sitio, de dos módulos se llevó a cabo al ligar dos fragmentos plasmídicos apropiados. El gen de resistencia a la ampicilina del vector (*Ap^r*) se reconstituyó. (C) Los nucleótidos necesarios para unir dos módulos se confinaron dentro del primer codón de cada módulo. (D) Los multímeros de módulos se conectaron como módulos individuales lo que resultó en un ensamblaje controlado de los genes sintéticos. (E) Las secuencias de aminoácidos de los módulos de seda diseñados se derivaron a partir de las proteínas de la seda dragline ADF-3 y ADF-4.

45 **Figura 2** Análisis de proteínas de la seda de araña. (A) Las etiquetas T7 de las proteínas recombinantes de la seda se detectaron después de una transferencia de tipo Western con un anticuerpo anti-etiqueta T7. (B) Las proteínas se sometieron a SDS-PAGE seguido de tinción con plata. Debido a una tinción débil de (AQ)₁₂ y (QAQ)₈ el contraste de la imagen se aumentó electrónicamente. (C) Se muestran espectros de emisión de fluorescencia de C₁₆NR4 purificada con longitudes de onda de excitación de 280 nm (línea recta) o 295 nm (línea de puntos), respectivamente.

50 **Figura 3** Estructura secundaria y transiciones de temperatura de proteínas de la seda de araña. (A) Los espectros de CD de proteínas rep (líneas rectas), proteínas repNR (líneas de puntos) y proteínas NR (guiones largos) se registraron a 20°C. (B) Las elipticidades del peso promedio por residuo (MRW) de las proteínas solubles de la seda de araña se

midieron a 220 nm al calentar las proteínas sintéticas de la seda a 90°C (línea recta), seguido de enfriamiento a 20°C (línea de puntos).

Figura 4 Agregación de proteínas sintéticas de la seda de araña. La agregación de las proteínas se determinó después de una incubación durante 1 hora en tampón (control), en presencia de 300 mM de NaCl, o 300 mM de KCl, a pH 1 o en presencia de 300 mM de fosfato de potasio. Las barras para las proteínas derivadas de ADF-3: gris claro; de ADF-4: gris oscuro.

Figura 5 Estrategia de clonación para la construcción de genes sintéticos de la seda de araña flageliforme (ver Figura 1). Los módulos individuales se conectaron como homo-multímeros (a), así como hetero-multímeros (b). (c) muestra las secuencias de aminoácidos de los módulos diseñados de la seda flageliforme derivados de una proteína de seda flageliforme (Flag) a partir de *Nephila clavipes*.

La **Figura 6** muestra un mapa de restricción del vector pAZL.

Figura 7: Formas de ensamblaje de proteínas de la seda de araña. **(A)** Esferas formadas por C₁₆ visualizadas mediante microscopía electrónica de barrido (SEM). **(B)** Nanofibrillas formadas por C₁₆NR4 visualizadas mediante microscopía de fuerza atómica (información de altura). **(C, D)** Microfibrilla formada por (AQ)₂₄NR3 investigada mediante SEM **(C)**. Para el corte de la fibrilla y la posterior visualización de la sección transversal se usó un haz de iones Ga⁺ enfocado **(D)**. **(E)** Espuma generada a partir de una solución de (AQ)₂₄NR3. **(F)** Espuma generada a partir de una solución de C₁₆NR4. **(G)** Gel reticulado formado por nanofibrillas de C₁₆NR4.

Figura 8: Espectros de CD de las proteínas sintéticas de la seda (AQ)₂₄NR3 y C₁₆ disueltas en 6 M de tiocianato de guanidina seguido por diálisis contra 5 mM de fosfato de potasio pH 8.0 (línea recta) o disueltas en HFIP (línea de puntos).

Figura 9: Moldeado de película de C₁₆ a partir de una solución de C₁₆ al 2% p/v en HFIP.

Figura 10: Espectros de CD de películas de proteínas producidas a partir de (AQ)₂₄NR3 y C₁₆. Las películas se moldearon a partir de una solución de proteínas en HFIP directamente sobre un vidrio de cuarzo plano y se analizaron por espectroscopía de CD (línea de puntos). La película se procesó posteriormente con 1 M de fosfato de potasio y se volvió a analizar. Debido a imprecisiones en la definición del espesor de las películas, la $\square\Theta_{MRW}$ no pudo determinarse.

Figura 11: Modificación de películas de C₁₆ moldeadas a partir de una solución de HFIP y procesadas con fosfato de potasio. **(A)** El acoplamiento eficiente de la fluoresceína (color amarillo) sólo ocurrió cuando los grupos carboxilo de C₁₆ se activaron (+) con el uso de EDC. En contraste sólo se unió poca fluoresceína a las películas sin activación con EDC (-). **(B)** La actividad de la β -galactosidasa acoplada se monitoreó mediante el uso de X-Gal como sustrato. La aparición de un precipitado azul indicó la actividad de la enzima sólo en películas que se habían activado con EDC (+), mientras que las películas no activadas sólo mostraron actividad enzimática residual (-).

Figura 12: Imagen de AFM de nanofibras de C₁₆.

Figura 13: Hidrogeles preparados de nanofibras de C₁₆

Figura 14: El comportamiento de tensión/deformación de los hidrogeles reticulados y no reticulados a una concentración de 10 mg/ml.

Figura 15: Dependencia de la frecuencia del módulo de almacenamiento (G') y el módulo de pérdida (G'') tanto para las redes de fibras reticuladas como no reticuladas a una concentración de 20 mg/ml.

Figura 16: Dependencia de la concentración del módulo de almacenamiento a una frecuencia de 0.5 Hz tanto para los hidrogeles reticulados como no reticulados. Ambas redes tienen módulos de almacenamiento que son proporcionales al cuadrado de la concentración [c]².

EJEMPLOS

Procedimientos experimentales

Materiales. Productos químicos se obtuvieron de Merck KGaA (Darmstadt, Alemania) si no se indica lo contrario. La manipulación y modificación del ADN se realizó como se describió previamente (19). Las enzimas de restricción se

obtuvieron de New England Biolabs (Beverly, MA, USA) y la ligasa de Promega Biosciences Inc. (San Luis Obispo, CA, USA). La purificación del ADN se realizó usando kits de Qiagen (Hilden, Alemania). Los oligonucleótidos sintéticos se obtuvieron de MWG Biotech AG (Ebersberg, Alemania). Todas las etapas de clonación se realizaron en la cepa de *E. coli* DH10B de Novagen (Madison, WI, USA).

5

Construcción del vector de clonación pAZL. Un casete de clonación con extremos coherentes complementarios a los generados por *BglIII* y *HindIII* se creó mediante hibridación de dos oligonucleótidos sintéticos CCI (GATCGAGGAGGATCCATGGGACGAATTCACGGCTAATGAAAGCTTACTGCAC) (sec. con núm. de ident.: 18) y CC2 (AGCTGTGCAGTAAGCTTTTCATTAGCCGTGAATTTCGTC CCATGGATCCTCCTC) (sec. con núm. de ident.: 19). La hibridación se llevó a cabo disminuyendo la temperatura de una solución de oligonucleótidos de 50 pmol/μl (cada una) de 95°C a 20°C con un incremento de 0.1°C/s. Las cadenas dobles no apareadas se desnaturalizaron a 70°C seguido de otro descenso de la temperatura a 20°C. Después de repetir el ciclo a 20°C-70°C-20°C diez veces, se realizaron diez ciclos adicionales con una temperatura de desnaturalización de 65°C. El casete de clonación resultante se ligó con un vector pFastBacI (Invitrogen, Carlsbad, California, Estados Unidos) digerido con *BglIII* y *HindIII*. Ambas secuencias de reconocimiento de enzimas de restricción se destruyeron tras esta etapa de clonación. El nuevo vector de clonación resultante se denominó pAZL.

20

Clonación de módulos de seda y regiones NR en el vector pAZL. Tres módulos de aminoácidos derivados de las proteínas de la seda dragline ADF-3 y ADF-4 (Fig.1E) se volvieron a traducir a una secuencia de ADN considerando el uso de codones bacterianos. Los oligonucleótidos de ADN complementarios correspondientes A1 (TCCGTACGGCCCAGGTGCTAGCGCCGACGGCAGCGGCTGGT GGCTACGGTCCGGGCTCTGGCCAGCAGGG) (sec. con núm. de ident.: 20) y A2 (CTGCTGGCCAGACCCGGACCGTAGCCACCAGCCGCTGCCGCTGCCGCGTAGCAC CTGGGCCGTACGGACC) (sec. con núm. de ident.: 21), Q1 (TCCGGGCCAGCAGGGCCCGGGTCAAC AGGGTCCTGGCCAGCAAGGTCCGGGCCAGCAGGG) (sec. con núm. de ident.: 22) y Q2 (CTGCT GGCCCGGACCTTGCTGGCCAGGACCCTGTTGACCCGGGCCCTGCTGGCCCGGACC) (sec. con núm. de ident.: 23), C1 (TTCTAGCGCGGCTGCAGCCGCGGCAGCTGCGTCCGGCCCGGG TGGCTACGGTCCGGAAAACCAGGGTCCATCTGGCCCGGGTGGCTACGGTCCTGGCG GTCCGGG) (sec. con núm. de ident.: 24) y C2 (CGGACCGCCAGGACCGTAGCCACCCGGGCCAG ATGGACCCTGGTTTTCCGGACCGTAGCCACCCGGGCCGGACGCAGCTGCCGCGGCTG CAGCCGCGCTAGAACC) (sec. con núm. de ident.: 25) se sintetizaron e hibridaron como se describió anteriormente y se ligaron al vector pAZL digerido con *BsgI* y *BseRI*. Las regiones NR de los genes de seda de araña *adf-3* (gi|1263286) y *adf-4* (gi|1263288) (obtenidos de Prof. Gosline, Vancouver, Canadá) se amplificaron por PCR usando los siguientes iniciadores: NR3f (GAAAACCATGGGTGCGGCTTCTGCAGCTGTATCTG) (sec. con núm. de ident.: 26), NR3r (GAAAAGAAGCTTTTCATTAGCCAGCAAGGGCTTGAGCTACAGATTG) (sec. con núm. de ident.: 27), NR4f (GAAAACCATGGGAGCATATGGCCCATCTCCTTC) (sec. con núm. de ident.: 28) y NR4r (GAAAAGAAGCTTTTCATTAGCCTGAAAGAGCTTGGCTAATCATTG) (sec. con núm. de ident.: 29).

40

Para las secuencias Flag, pueden usarse los siguientes iniciadores y casetes:

Iniciador de PCR:

45

FLAG-N-*chr*-sentido: (sec. con núm. de ident.: 43)

5'- GAAAACCATGGGCGAAAGCAGCGGAGGCGAT - 3'

45

FLAG-N-*chr*-anti: (sec. con núm. de ident.: 44)

5'- GAAAAGAAGCTTTTCATTAGCCTGGGCTGTATGGTCC - 3'

45

FLAG-C-*chr*-sentido: (sec. con núm. de ident.: 45)

5'- GAAAACCATGGGTGCTTATTATCCTAGCTCGC - 3'

50

FLAG-C-*chr*-anti: (sec. con núm. de ident.: 46)

5'- GAAAAGAAGCTTTTCATTAGCCATAAGCGAACATTCTCCTAC - 3'

Oligos para secuencias repetitivas a partir de las que se generaron casetes:

55

60

Módulo Y-(GPGGX)-ds: (**sec. con núm. de ident.:** 47)
 5' -
 5 TCCGGGCGGTGCGGGCCCAGGTGGCTATGGTCCGGGCGGTTCTGGGCCGGGTGGCT
 ACGGTCCTGGCGGTTCCGGCCCCGGGTGGCTACGG - 3'
 Módulo Y-(GPGGX)-cs: (**sec. con núm. de ident.:** 48)
 5' -
 10 GTAGCCACCCGGGCGGGAACCGCCAGGACCGTAGCCACCCGGCCCAGAACCGCCCCG
 GACCATAGCCACCTGGGCCCCGCACCGCCCCGGACC - 3'
 Módulo sp-(espaciador)-ds: (**sec. con núm. de ident.:** 49)
 5' -
 15 TGGCACCACCATCATTGAAGATCTGGACATCACTATTGATGGTGCGGACGGCCCCGAT
 CACGATCTCTGAAGAGCTGACCATCGG - 3'
 Módulo sp-(espaciador)-cs: (**sec. con núm. de ident.:** 50)
 5' -
 20 GATGGTCAGCTCTTCAGAGATCGTGATCGGGCCGTCCGCACCATCAATAGTGATGTC
 CAGATCTTCAATGATGGTGGTGCCACC - 3'
 Módulo K (GPGGAGGPY)-ds: (**sec. con núm. de ident.:** 51)
 5' -
 25 TCCGGGCGGTGCTGGCGGTCCGTACGGCCCTGGTGGCGCAGGTGGGCCATATGGTCC
 GGGCGGTGCGGGCGGTCCGTACGG - 3'

Módulo X (GGX)-ds: (**sec. con núm. de ident.:** 53)
 5'-TGGCGCTGGTGGCGCCGGTGGCGCAGGTGGCTCTGGCGGTGCGGGCGGTTCCGG -
 3'
 Módulo X (GGX)-cs: (**sec. con núm. de ident.:** 54)
 5'-GGAACCGCCCCGCACCGCCAGAGCCACCTGCGCCACCGGCGCCACCAGCGCCACC
 30 - 3'

Los productos de PCR y el vector pAZL se ligaron después de la digestión con *NcoI* y *HindIII*. La clonación de los módulos sintéticos así como los productos de PCR resultó en la sustitución del espaciador del casete de clonación, lo que preservó la disposición de sus elementos. Para una traducción más eficaz, el codón AGA (Arg), que rara vez se traduce en *E. coli*, se mutó a CGT (Arg) en NR3 y NR4 con el uso de mutagénesis por PCR (19).

Construcción de genes sintéticos de la seda de araña. La conexión de dos fragmentos génicos por ejemplo módulos individuales, multímeros de módulos o regiones NR representó la etapa básica de la estrategia de clonación. Para este propósito el vector pAZL, que contiene el fragmento génico 5' terminal designado se digirió con *BsaI* y *BsgI*, mientras que el vector que comprende el fragmento génico 3' terminal se digirió con *BseRI* y *BsaI*, respectivamente (Fig. 1B). La ligación de los fragmentos plasmídicos apropiados produjo la conexión de los dos fragmentos génicos y condujo a la reconstitución del gen de resistencia a la ampicilina (*Ap^r*) del vector pAZL que facilitó la identificación de las construcciones correctas.

Para la construcción de los genes, los módulos individuales se conectaron primero para producir unidades de repetición (Fig.1D + Fig.5). Estos se multimerizaron gradualmente y opcionalmente se unieron con regiones NR. Finalmente, las construcciones de los genes sintéticos así como las regiones NR se escindieron del vector pAZL con *BamHI* y *HindIII* y se ligaron con el vector de expresión bacteriano pET21a (Novagen) igualmente digerido, lo que proporciona una secuencia codificante para una etiqueta T7 (MASMTGGQQMGR) (sec. con núm. de ident.: 30) (20). La fidelidad de todas las construcciones se confirmó por secuenciación de ADN.

Expresión génica. Todos los genes de la seda se expresaron en la cepa de *E. coli* BLR [DE3] (Novagen). Las células se cultivaron a 37°C en medio LB a una OD₆₀₀ = 0.5. Antes de la inducción con 1 mM de IPTG (Isopropil-β-D-tiogalactosido), las células se cambiaron a 30°C en el caso de (AQ)₁₂, (AQ)₁₂NR3, (QAQ)₈, y (QAQ)₈NR3 y a 25°C en el caso de C₁₆, C₁₆NR4, NR3 y NR4 respectivamente. Alternativamente las células se cultivaron en un fermentador a una OD₆₀₀ = 40-50 con el uso de medios complejos (21) y la técnica de alimentación por lotes (22). De nuevo, antes de la inducción con 1 mM de IPTG las células se cambiaron a 25°C o 30°C, respectivamente. Las células que expresan (AQ)₁₂, (AQ)₁₂NR3, (QAQ)₈, (QAQ)₈NR3, C₁₆ y C₁₆NR4 se cosecharon después de 3-4 horas de la inducción mientras que las células que expresan NR3 y NR4 se cosecharon después de 16 horas.

Purificación de proteínas. Las células se resuspendieron con 5 ml/g de amortiguador que contenía 20 mM de *N*-(2-hidroxietil)piperazina-*N*-(ácido 2-etanosulfónico) (HEPES) pH 7.5, 100 mM de NaCl, 0.2 mg/ml de lisozima (Sigma-Aldrich, St. Louis, MO, Estados Unidos) y se incubaron a 4°C durante 30 min. Las células se lisaron por sonicación mediante el uso de un ultrasonificador HD/UW2200/KE76 (Bandelin, Berlín, Alemania) y el ADN genómico se digirió al incubar los lisados celulares con 0.1 mg/ml de DNasa I (Roche, Mannheim, Alemania) y 3 mM de MgCl₂ a 4°C durante 60 minutos. Los fragmentos celulares insolubles se sedimentaron a 50,000×g y 4°C durante 30 min. Las proteínas solubles de *E. coli* de los lisados que contenían (AQ)₁₂, (AQ)₁₂NR3, (QAQ)₈, (QAQ)₈NR3, C₁₆ y C₁₆NR4 se precipitaron mediante desnaturalización por calor a 80°C durante 20 min, mientras que los lisados que contenían NR3 y NR4 se calentaron a 70°C durante el mismo período de tiempo. Las proteínas precipitadas se eliminaron mediante sedimentación a 50,000×g durante 30 min. Las proteínas de la seda, que permanecieron solubles durante la desnaturalización por calor, se precipitaron con sulfato de amonio al 20% (800 mM) ((AQ)₁₂, (AQ)₁₂NR3, (QAQ)₈, (QAQ)₈NR3, C₁₆ y C₁₆NR4) o sulfato de amonio al 30% (1200 mM) (NR3 y NR4) a temperatura ambiente y se cosecharon mediante centrifugación a 10,000×g durante 10 min. Los sedimentos de (AQ)₁₂, (AQ)₁₂NR3, (QAQ)₈, (QAQ)₈NR3, NR3 y NR4 se enjuagaron con una solución que contenía la misma concentración de sulfato de amonio tal como se usó para la precipitación y se disolvieron en 6 M de cloruro de guanidinio (GdmCl). En contraste C₁₆ y C₁₆NR4 se lavaron con 8 M de urea y se disolvieron en 6 M de tiocianato de guanidinio (GdmSCN). Todas las proteínas se dializaron contra 10 mM de NH₄HCO₃. Los precipitados formados durante la diálisis se eliminaron por sedimentación a 50,000×g durante 30 min y las restantes proteínas solubles de la seda se liofilizaron. Antes del análisis la proteína liofilizada se disolvió en 6 M de GdmSCN seguido por diálisis frente a tampones adecuados. Los agregados se eliminaron por sedimentación a 125,000 ×g durante 30 min. Las concentraciones de proteínas se determinaron por fotometría en una cubeta de 1 cm de longitud de la trayectoria a 276 nm con el uso de coeficientes de extinción calculados (Tabla 1) (23). La identidad de las proteínas se confirmó por electroforesis en gel de dodecilsulfato sódico - poliacrilamida (SDS-PAGE); geles de Tris-glicina 10% para proteínas > 20 kDa y geles de Tris-Tricina 10 - 20% (Invitrogen) para proteínas < 20 kDa), seguido de transferencia sobre membranas de fluoruro de polivinilideno (PVDF) (Millipore, Billerica, MA, Estados Unidos) y la detección mediante el uso de un anticuerpo monoclonal anti-T7 de ratón (Novagen, 1:10,000) como primario y un conjugado de peroxidasa y anti-1gG de ratón (Sigma-Aldrich, 1:5,000) como anticuerpo secundario. La actividad peroxidasa se visualizó con el uso del kit de detección de membrana de Western ECL^{plus} de Amersham Biosciences (Piscataway, NJ, Estados Unidos).

Fluorescencia. Los espectros de fluorescencia se registraron en un Espectrofluorómetro FluoroMax (Jobin Yvon Inc, Edison, NJ, Estados Unidos). Los espectros se tomaron mediante el uso de una concentración de proteínas de 100 µg/ml en 10 mM de Tris (hidroximetil) aminometano (Tris) / HCl (pH 8.0) a temperatura ambiente. El tiempo de integración fue de 1 s, el tamaño del paso fue de 0.5 nm y los anchos de banda fueron de 5 nm (excitación) y 5 nm (emisión), respectivamente.

Análisis de la estructura secundaria. Los espectros de dicroísmo circular (CD) de UV lejano se obtuvieron mediante el uso de un espectropolarímetro Jasco 715 equipado con una unidad de control de temperatura (Jasco International Co. Ltd., Tokio, Japón). Todos los espectros se tomaron a una concentración de proteínas de 150 µg/ml en 5 mM de Tris/HCl (pH 8.0) en una cubeta de cuarzo de 0.1 cm de longitud de la trayectoria a 20°C. La velocidad de barrido fue de 20 nm/min, el tamaño del paso fue de 0.2 nm, el tiempo de integración se estableció en 1 s y el ancho de banda fue de 1 nm. Se promediaron cuatro barridos y se corrigieron con amortiguador. Las transiciones térmicas se analizaron con un incremento de calentamiento / enfriamiento de 1°C/min a 220 nm.

Ensayo de solubilidad. Para determinar la concentración máxima de proteínas solubles, una solución de 1 mg/ml (= 0.1% (p/v)) en 10 mM de Tris/HCl pH 8.0 se concentró por ultrafiltración mediante el uso de una membrana de poliéter sulfona con un corte para un peso molecular de 10,000 Da (Vivascience AG, Hannover, Alemania). Se tomaron muestras de la solución en distintos intervalos hasta que la proteína empezó a precipitar. Las muestras se diluyeron en 10 mM de Tris pH 8.0 para determinar la concentración de proteína por fotometría.

Ensayo de agregación. Todas las muestras se ajustaron a 1 mg/ml en 10 mM de Tris/HCl pH 8.0. Para probar los efectos iónicos sobre la agregación de proteínas de la seda, se añadieron sales hasta concentraciones finales de 300 mM. El efecto de la acidificación se investigó al añadir HCl hasta una concentración final de 100 mM (pH = 1). Todas las muestras se incubaron a temperatura ambiente durante 1 hora. Los precipitados de proteínas se eliminaron de todas las muestras por sedimentación a 125,000×g durante 25 min y la cantidad de proteína soluble restante se determinó fotométricamente. Dado que la suma de la proteína soluble y agregada tuvo que ser igual a la cantidad inicial de proteína soluble, el porcentaje de proteína agregada podría calcularse al restar la cantidad de proteína soluble de la cantidad de proteína usada inicialmente.

RESULTADOS

Una estrategia de cierre para diseñar proteínas similares a las de seda. La expresión de genes auténticos de la seda de araña en huéspedes bacterianos es ineficiente (24) ya que algunas secciones génicas contienen codones que no se

traducen de manera eficiente en bacterias. Además, la manipulación génica y la amplificación por PCR es difícil debido a la naturaleza repetitiva de las sedas. Para investigar las propiedades de las proteínas de la seda de araña, se han empleado estrategias de clonación mediante el uso de módulos de ADN sintético con un uso de codones adaptado al huésped de expresión correspondiente. Se obtuvieron los genes sintéticos que codifican para proteínas semejantes a las regiones repetitivas de las sedas de araña (25-28). Es importante destacar que ninguno de estos diseños de proteínas incluyó las regiones NR del carboxilo terminal que se encuentran en todas las sedas dragline.

Los inventores desarrollaron una estrategia de clonación sin problemas (29) que permite la combinación controlada de diferentes módulos de ADN sintético, así como fragmentos de genes auténticos. Se diseñó el vector de clonación pAZL que comprende un casete de clonación con un espaciador que actúa como marcador de posición para los genes sintéticos, y sitios de reconocimiento para las enzimas de restricción *BseRI* y *BsgI* (Fig.1A). Dado que los sitios de reconocimiento y escisión de estas enzimas están a 8 (*BseRI*) o 12 (*BsgI*) nucleótidos aparte, el inicio de la traducción y los codones de parada, así como sitios de restricción adicionales requeridos para la escisión de los genes ensamblados podrían colocarse cerca del espaciador.

En una primera etapa de clonación la región espaciadora de pAZL se reemplazó por un módulo de ADN sintetizado (para el diseño del módulo ver más abajo). Posteriormente pudieron unirse dos módulos de manera dirigida al sitio (ver materiales y métodos y Fig.1B). Las extensiones complementarias 3' de una sola cadena GG (sentido) y CC (antisentido) generadas mediante escisión con *BsgI* y *BseRI* se usaron para conectar dos módulos (Fig.1C). Así la secuencia de ADN requerida para unir dos módulos se confinó a un codón de glicina (GGX). La glicina es abundante de manera natural en proteínas de la seda de araña (~ 30%), por lo tanto los módulos pudieron diseñarse sin la necesidad de buscar sitios de reconocimiento para endonucleasas de restricción que, después de la traducción, hagan coincidir las secuencias auténticas de aminoácidos. Dado que la disposición de los elementos del casete de clonación se mantuvo sin cambios tras la clonación y la multimerización, pudo construirse una variedad de combinaciones de módulos (Fig.1D).

Diseño, síntesis y purificación de las sedas sintéticas de araña. Los inventores eligieron las proteínas de seda dragline ADF-3 y ADF-4 (3) de la araña de jardín *Araneus diadematus* como moldes para las construcciones sintéticas. La estructura primaria parcialmente identificada de ADF-3 en gran parte consiste de unidades de repetición, todas las cuales comprenden una secuencia consenso que incluye un motivo de poli-alanina. La longitud de las unidades de repetición individuales se determina al variar los números del motivo GPGQQ. Para imitar la secuencia repetitiva de ADF-3 se diseñaron dos módulos. Un módulo, denominado A, se derivó de la secuencia consenso que contiene poli-alanina (Fig.1E). Un segundo módulo denominado Q contenía cuatro repeticiones del motivo GPGQQ. Para estudiar unidades de repetición de diferente longitud, uno o dos módulos Q se combinaron con un módulo A para obtener (AQ) o (QAQ). Estas unidades de repetición se multimerizaron para generar genes sintéticos que codifican para proteínas repetitivas (proteínas rep) (AQ)₁₂ y (QAQ)₈.

La parte repetitiva de ADF-4 se compone generalmente de una sola unidad de repetición conservada que solamente presenta pequeñas variaciones. Los inventores combinaron estas variaciones y diseñaron un módulo consenso denominado C (Fig.1E), que los inventores multimerizaron para obtener la proteína rep C₁₆. Se eligió que el número de repeticiones de los módulos en todos los genes sintéticos codificara para proteínas de masa molecular similar (~50 kDa).

Tanto ADF-3 como ADF-4 presentan regiones NR homólogas en sus extremos carboxilo, que comprenden 124 y 109 aminoácidos respectivamente. Las secuencias génicas que codifican para estas regiones se amplificaron mediante PCR, y los codones problemáticos para la expresión bacteriana se cambiaron a codones más adecuados mediante mutagénesis dirigida a un sitio (ver materiales y métodos). Por lo tanto, todos los genes sintéticos usados pudieron combinarse con las regiones NR auténticas apropiadas lo que produce genes que codifican para las proteínas repNR (AQ)₁₂NR3, (QAQ)₈NR3 y C₁₆NR4. Adicionalmente NR3 y NR4 podrían expresarse solas.

Después de la síntesis bacteriana las proteínas de la seda se purificaron mediante una etapa de calor seguida de una precipitación con sulfato de amonio. La identidad de las proteínas se confirmó por inmunotransferencia, mediante el uso de anticuerpos dirigidos contra las secuencias peptídicas de etiquetas T7, unidas al extremo amino terminal de todas las proteínas de la seda (Fig.2A). Aunque todas las proteínas rep y todas las proteínas repNR tenían pesos moleculares similares (Tabla 1) éstas mostraron diferentes velocidades de migración cuando se sometieron a SDS-PAGE. Este efecto podría ser causado por una unión aberrante del dodecilsulfato a las proteínas debido a la composición de aminoácidos diferentes, lo que conduce a una variación de las cargas netas de las proteínas. Además de proteínas de longitud completa, la inmunotransferencia reveló trazas de proteínas con menor peso molecular dentro de las preparaciones de proteínas repNR. La unión del anticuerpo anti-etiqueta T7 a estas proteínas las identificó como proteínas de la seda que carecen de parte de su extremo carboxilo-terminal. Al analizar cada proteína purificada mediante SDS-PAGE y tinción con plata, no se detectaron otras proteínas en todas las preparaciones de proteínas (Fig.2B). Adicionalmente la pureza de las proteínas se determinó al medir la emisión de fluorescencia. La luz incidente de longitud de onda de 280 nm conduce a la excitación y la emisión de fluorescencia de tirosinas y triptófanos mientras que la luz de 295 nm excita exclusivamente a este último. Dado

que ninguna de las proteínas de la seda de araña diseñadas comprendía triptófanos, la emisión de fluorescencia tras la excitación con 295 nm habría sido un indicativo de proteínas de *E. coli* contaminantes, que en promedio contienen 1.5% de triptófano (30). Las mediciones de fluorescencia de todas las preparaciones de proteínas de la seda revelaron espectros de emisión similares al espectro de la tirosina, que se produce de manera abundante en las proteínas de la seda. En contraste, no se detectó ninguna fluorescencia de triptófano, lo que indica una alta pureza de las preparaciones de proteínas (los datos se muestran de manera ilustrativa para C₁₆NR4 en la Fig. 2B).

La producción bacteriana de proteínas de seda sintéticas en matraces Erlenmeyer produjo cantidades de proteínas similares para todas las construcciones. Los rendimientos de las preparaciones individuales están en el intervalo de 10 a 30 mg de proteína purificada por litro de medio de cultivo. La fermentación de las células se empleó para investigar la posibilidad de escalar la síntesis de proteínas. Los rendimientos de ((QAQ)₈NR3 y C₁₆NR4 por lo tanto pudieran aumentar a 140 y 360 mg/l, respectivamente.

Las proteínas RepNR consisten de una región repetitiva poco estructurada y un dominio no repetitivo altamente estructurado. La estructura secundaria se investigó mediante espectroscopía de CD. Las proteínas rep revelaron espectros típicos de proteínas no estructuradas de manera intrínseca. En contraste las proteínas NR revelaron espectros indicativos de un alto contenido de estructura secundaria. Estas regiones parecen representar dominios proteicos plegados de manera independiente. Los espectros de las proteínas repNR corresponden a grandes rasgos a una combinación de los espectros rep y NR ponderados de acuerdo con su participación en las proteínas repNR. Aunque no puede excluirse un cambio estructural menor dentro de las regiones rep o los dominios NR tras la unión mutua es probable que las proteínas repNR estén compuestas de una región que muestra mayormente una estructura aleatoria y un dominio proteico plegado en el carboxilo terminal. Sorprendentemente los espectros de las proteínas repNR fueron similares a los espectros de CD obtenidos a partir de la base de seda de la ampulácea mayor extraída directamente de arañas (*Nephila clavipes*) (9).

Las proteínas de la seda *se vuelven a plegar después de la desnaturalización térmica y química.* Al investigar los cambios estructurales mediante espectroscopía de CD tras un calentamiento, no se observaron transiciones de temperatura de cooperación para las proteínas rep entre 20°C y 90°C, un efecto que también se ha observado para otras proteínas desplegadas intrínsecamente (31; 32) (Fig.3). Dado que las proteínas repNR estaban al menos parcialmente estructuradas, el desplegamiento térmico de la región estructurada debe ser detectable a temperaturas elevadas. En consecuencia, se observaron transiciones térmicas de cooperación. Los puntos medios de las transiciones de temperatura fueron 67°C ((QAQ)₈NR3), 66°C ((AQ)₁₂NR3) y 72°C (C₁₆NR4), respectivamente (Fig.3B y Tabla 1). Además, todas las transiciones *térmicas* fueron completamente reversibles. La reversibilidad de los cambios estructurales tras el calentamiento explicó la alta recuperación de proteínas de la seda solubles después de la etapa de calor empleada durante la purificación de proteínas. El Tris se usó para amortiguar todas las soluciones investigadas por espectroscopía de CD, debido a las buenas propiedades espectrales y la poca capacidad de promover la agregación de proteínas de la seda. Debido a la fuerte dependencia de la temperatura de las soluciones amortiguadas con Tris, se esperaba que el pH de las muestras cambiara de pH 8 a pH 6 tras el calentamiento de 20°C a 90°C (19). Sin embargo, las transiciones de temperatura de las proteínas de la seda en amortiguador fosfato a pH 8, que muestran un valor de pK independiente de la temperatura, revelaron iguales temperaturas de punto medio (datos no mostrados) a pesar de que no fueron totalmente reversibles probablemente debido a la agregación de proteínas (ver más abajo). Esto indicó que las transiciones *térmicas* de las proteínas de la seda no se influenciaron por cambios de pH inducidos térmicamente en soluciones amortiguadas con Tris.

El efecto de la desnaturalización y la renaturalización químicas sobre la estructura secundaria se investigó mediante la medición del dicroísmo circular de las proteínas repNR en amortiguador Tris, después de la diálisis contra 6 M de GuaHCl y la renaturalización mediante diálisis frente al amortiguador Tris. Los espectros idénticos de las proteínas iniciales y las que volvieron a plegarse indicaron que la desnaturalización química es reversible (datos no mostrados).

La solubilidad de las proteínas de la seda se determina por sus secuencias repetitivas. Con el objetivo de obtener altas concentraciones de proteína en la base, las proteínas de la seda tienen que ser altamente solubles. Probamos las concentraciones máximas en que las proteínas rep y repNR se mantuvieron solubles para identificar los elementos de la estructura primaria que determinan la solubilidad. Todas las proteínas que comprenden los módulos A y Q pudieron concentrarse por ultrafiltración a más de 30% p/v sin formar agregados visibles, independientemente de la presencia del dominio NR. En contraste, las proteínas que contienen el módulo C sólo pudieron concentrarse hasta 8% p/v (C₁₆) y 9% p/v (C₁₆NR4), respectivamente (Tabla 1). Ambas proteínas formaron un sólido similar a un gel tras un aumento de la concentración (datos no mostrados). Por lo tanto, la solubilidad de las proteínas de la seda se determinó únicamente por sus secuencias repetitivas y no fue influenciada por el dominio NR.

El potasio no promueve agregación de las proteínas de seda sintéticas, independientemente de su estructura primaria. El pH, los iones, tales como potasio y fosfato, y la tensión mecánica están involucrados en el ensamblaje de la seda natural. Aquí hemos querido investigar cómo estos factores promueven el ensamblaje de las proteínas de seda sintéticas. Ya que

éramos incapaces de imitar el proceso de ensamblaje auténtico, que requiere una pre-orientación de las proteínas involucradas como se encuentra en la base líquida cristalina (33), realizamos un ensayo de agregación a partir de soluciones de proteínas que no presentan un orden de orientación. Ninguna de las proteínas probadas rep, repNR y NR mostraron una agregación significativa (< 5%) cuando se incubaron en amortiguador, lo que indica que todas las proteínas eran intrínsecamente solubles bajo las condiciones de prueba (Fig.4). Para investigar si la adición de iones causaba agregación por el aumento de la fuerza iónica, las proteínas se incubaron con cloruro de sodio. Sin embargo no se observó agregación. En contraste con el sodio, previamente se ha reportado que el potasio promueve específicamente la agregación de la seda (34). Sin embargo, el cloruro de potasio tampoco mostró influencia sobre la solubilidad de las proteínas de seda sintéticas (Fig.4).

La acidificación y la adición de fosfato inician la agregación de las proteínas rep en dependencia de su estructura primaria. La función exacta de la acidificación durante el ensamblaje de la seda de araña aún no se ha determinado. Sin embargo parece probable que los grupos cargados negativamente (por ejemplo los grupos fosforilo) están protonados lo que reduce así la carga neta y la repulsión de las proteínas de la seda de araña. Dado que las proteínas de seda sintéticas no contenían grupos químicos que mostraran un valor de pK_A dentro del intervalo de cambio de pH observado durante el proceso de hilatura, los inventores se propusieron imitar este efecto al protonar todos los grupos carboxilos de las cadenas terminal y lateral. $(QAQ)_8$ y $(AQ)_{12}$, que presentan sólo el grupo carboxilo terminal, no mostraron agregación (<5%) y mostraron una débil (18%) agregación a pH 1. Curiosamente la protonación de los 16 residuos de glutamato de C_{16} también causó solamente una agregación débil (8%) (Fig.4). El fosfato que se ha descrito que se añade a la base durante el proceso de hilatura no causó agregación de $(QAQ)_8$ y precipitación débil de C_{16} (12%). En contraste, $(AQ)_{12}$ mostró un aumento de la tendencia a agregarse (47%) después del tratamiento con fosfato de potasio. Se obtuvieron resultados similares con el uso de fosfato de sodio, lo que indica que el efecto es causado específicamente por los iones fosfato (datos no mostrados).

Los dominios NR amplifican la respuesta a los factores que promueven la agregación. Para investigar la influencia de los dominios NR, se probó la agregación de las proteínas repNR, así como las proteínas NR a un pH bajo y tras el tratamiento con fosfato. La acidificación de $(QAQ)_8NR3$ y $(AQ)_{12}NR3$, así como NR3 causó una débil agregación (10%, 15% y 13%), que estuvo en el intervalo mostrado por las correspondientes proteínas rep. Curiosamente, aunque el dominio NR4 no precipitó a pH 1 (0%), $C_{16}NR4$ mostró una fuerte agregación a pH 1 (70%). Por lo tanto la combinación de C_{16} repetitiva y el dominio NR4, que no se agregó significativamente tras la acidificación, condujo a una proteína altamente sensible a este factor promotor de la agregación. Se obtuvieron resultados similares para la adición de fosfato. Aunque ni NR3 ni NR4 mostraron agregación en presencia de fosfato (1% y 0%), la adición de los dominios NR a las regiones repetitivas provocó un aumento de la agregación de las proteínas repNR en comparación con las proteínas rep ($(QAQ)_8NR3$: 57%, $(AQ)_{12}NR3$: 81%, $C_{16}NR4$: 80%).

Con el uso de una estrategia de clonación que permite el ensamblaje controlado y sin problemas de los módulos de ADN, se construyeron genes sintéticos que codifican para proteínas similares a la seda de araña. El diseño de proteínas produjo diferentes combinaciones de unidades de repetición y regiones NR de origen natural, para probar de manera sistemática las propiedades de tales elementos individuales de la estructura primaria. El análisis estructural por espectroscopía de CD reveló que las regiones repetitivas en su mayoría no tienen estructura en su estado soluble, lo que muestra propiedades comunes a otras proteínas desplegadas intrínsecamente (31; 32). El mismo estado conformacional se ha propuesto para la mayor parte del contenido de la ampulácea mayor (10) que está dominado por secuencias de proteínas repetitivas. En contraste se encontraron regiones NR para representar independientemente dominios de la proteína plegable que adoptan su conformación después de la desnaturalización por calor así como el tratamiento con agentes caotrópicos. Debido a su pequeño tamaño relativo en comparación con las regiones repetitivas la influencia sobre las propiedades estructurales globales fue pequeña en las proteínas repNR.

En sedas de araña naturales que muestran regiones repetitivas de varios cientos de kDa, puede esperarse que la contribución estructural de las regiones NR sea incluso más pequeña, lo que explica la falta de evidencia para su presencia en las investigaciones del contenido de ampulácea mayor. Debido a la reversibilidad de la desnaturalización térmica y química de las proteínas repNR y la similitud de datos de CD presentados en este trabajo y obtenidos a partir de la solución de seda natural, se puede suponer que, incluso después del tratamiento con calor y reactivos caotrópicos durante la purificación y preparación de la muestra todos los componentes de seda de araña investigados en soluciones acuosas estaban en un estado de conformación comparable al de las proteínas de seda naturales dentro de la solución.

De acuerdo con Uversky y otros el desplegamiento intrínseco de las proteínas puede predecirse en base a su carga neta y a la hidropaticidad promedio. La carga neta de una proteína se usa para calcular una hidropaticidad "límite". Si la hidropaticidad promedio de la proteína está por debajo del valor "límite", se prevé que la proteína esté intrínsecamente desplegada (35;36). De acuerdo con los resultados presentados se prevé que las secuencias repetitivas $(QAQ)_8$ y $(AQ)_{12}$ estén intrínsecamente desplegadas (Tabla 1). El desplegamiento intrínseco de una proteína significa que las interacciones de los residuos de aminoácidos con el solvente que los rodea son más favorables que con los aminoácidos de la misma o

de otras cadenas polipeptídicas. En consecuencia, (QAQ)₈ y (AQ)₁₂ son solubles incluso a altas concentraciones. En contraste, C₁₆ muestra una hidropaticidad ligeramente por encima del valor límite. Aunque se revelan propiedades de las proteínas desplegadas intrínsecamente, las interacciones entre las cadenas polipeptídicas son cada vez más favorables a altas concentraciones lo que conduce a la agregación de la proteína y resulta en una menor solubilidad en comparación con (QAQ)₈ y (AQ)₁₂ (Tabla 1).

Como las secuencias repetitivas constituyen la fracción más grande de proteínas de seda de araña, probablemente determinan muchas de las propiedades de las proteínas. En consecuencia, las solubilidades de las proteínas repNR no difieren significativamente de las proteínas rep. La solubilidad y hidropaticidad calculada de (QAQ)₈ y (AQ)₁₂ se correlacionan bien con los valores de la ADF-3 auténtica (Tabla 1). C₁₆ y ADF-4 muestran menor solubilidad, aunque C₁₆ no comparte la insolubilidad intrínseca alta de ADF-4. Esta diferencia puede explicarse por la mayor hidropaticidad y carga neta inferior de ADF-4 comparada con C₁₆.

En contraste con las regiones repetitivas, los dominios NR representan sólo una fracción de las proteínas de seda de araña. Ambos dominios NR revelaron una estructura rica en α -hélices. Debido a la gran similitud entre los dominios NR de ADF-3 y ADF-4 (81% de similitud y 67% de identidad) puede suponerse que ambos podrían cumplir funciones relacionadas. Se obtuvo más información sobre la función de los dominios NR cuando se investigó la agregación de las proteínas de la seda tras el tratamiento con factores conocidos para inducir el ensamblaje de las proteínas de la seda *in vivo*. Se esperaba que la reducción de las cargas negativas por protonación de los grupos carboxilo de las proteínas de la seda afectara principalmente a las proteínas que comprenden el módulo C. En consecuencia, las proteínas compuestas por los módulos A y Q, que no contienen aspartatos o glutamatos, no mostraron más que una agregación débil. C₁₆, incluso después de la neutralización de sus 16 cargas negativas se mantuvo mayormente soluble. De manera sorprendente la combinación del dominio NR4, que no mostró ninguna respuesta a la acidificación por sí mismo, y la débil agregación de C₁₆ resultó en una proteína altamente sensible a la protonación. Por lo tanto se requiere una reducción de la carga de la región repetitiva y la presencia del dominio NR para la agregación eficiente. Se obtuvieron resultados similares cuando se añadió fosfato a las soluciones de proteínas. El fosfato, al igual que otros iones liotrópicos se sabe que aumenta la tensión superficial del agua, y promueve las interacciones hidrofóbicas (37). En el caso de proteínas de seda de araña es probable que la adición de fosfato inicie las interacciones entre los motivos poli-alanina hidrófobos, causando la agregación de las proteínas. En consecuencia la agregación de (AQ)₁₂ fue mayor que (QAQ)₈ que contiene un tercio menos de motivos poli-alanina que (AQ)₁₂. C₁₆ que muestra el número más largo y alto de motivos poli-alanina sin embargo no mostró la agregación más fuerte tras el tratamiento con fosfato. Una posible explicación de este resultado inesperado puede ser la repulsión de las cadenas laterales de glutamato cargadas negativamente e iones fosfato que conducen a su exclusión del solvente circundante y un debilitamiento de su efecto liotrópico. A pesar de que los dominios NR no respondieron a la adición de fosfato, su adición a las proteínas rep aumentó fuertemente la sensibilidad de fosfato. Aunque los datos presentados no son suficientes para llegar a una conclusión definitiva, parece probable que los dominios NR funcionan como potenciadores inespecíficos de la sensibilidad a la agregación de los factores de promoción. Para la agregación eficiente su presencia es tan importante como la capacidad de las regiones repetitivas de responder a estos factores.

El mecanismo de esta mejora podría implicar cambios en el estado oligomérico de las proteínas de la seda. Se ha encontrado que los dominios NR forman dímeros disulfuro puentado (38). Además, la oligomerización podría conducir a un aumento de las concentraciones locales de las secuencias de polipéptidos requeridos para iniciar la agregación que es asistida por las condiciones del solvente que favorezcan la formación de interacciones intermoleculares.

El presente enfoque de ingeniería de proteína, que combina secuencias repetitivas sintéticas con regiones NR auténticas, revela que las proteínas muy parecidas a las proteínas de seda auténticas se pueden producir con altos rendimientos. El sistema de expresión bacteriana, así como el proceso de purificación simple y barato, que se puede escalar fácilmente, proporciona la base para la producción a escala industrial rentable de proteínas del tipo de seda de araña. Con base en los estudios actuales, los mecanismos moleculares del ensamble de la seda de araña pueden ser investigados aún más, lo que proporcionará los conocimientos necesarios para la hilatura artificial de hilos de seda a partir de las proteínas recombinantes y para la obtención de nuevos materiales para aplicaciones biotecnológicas y médicas.

Ensamblaje de proteínas derivadas de la seda de araña

Los siguientes experimentos se realizaron para demostrar que las proteínas derivadas de las secuencias de la seda de araña ADF-3 (sec. con núm. de ident.: 1) o ADF-4 (sec. con núm. de ident.: 2) pueden ensamblarse en distintas formas morfológicas. Las proteínas (AQ)₂₄NR3 y C₁₆NR4 se construyeron, produjeron y prepararon en soluciones acuosas como se describe en Biochemistry 2004 Vol.43 págs. 13604-11362. Si no se menciona lo contrario las soluciones de proteína contenían 10 mM Tris-(hidroximetilo)-aminometano (Tris) pH 8.0.

1. Esferas

Esferas de proteínas que presentan diámetros en el intervalo entre 0.5 y 2 μm (Fig.7a) se generaron al añadir 0.8 M de sulfato amonio a una solución de C_{16} al 0.2% (p/v).

5 2. Nanofibrillas

Las nanofibrillas que mostraron diámetros entre 0.7 y 4 nm (Fig. 7b) se formaron por incubación de una solución de C_{16} NR4 1% (p/v) a temperatura ambiente por 2 semanas.

10 3. Microfibrillas

Para la formación de microfibrillas 5 - 10 μl de una solución (AQ)₂₄NR3 25% (p/v) se inyectaron lentamente en 0.5 M fosfato de potasio pH 8.0, formando una gota estable de solución de proteína. Después de la incubación durante 1 min la gota de proteína se retiró de la solución usando pinzas. Después de un tiempo de incubación adicional de 1 min en el aire, una fibrilla de proteína podría extraerse de la gota de proteína a una velocidad de aproximadamente 2 cm/s usando un segundo conjunto de pinzas. Las fibrillas mostraron una sección transversal redonda con un diámetro de 4 μm (Fig 7c,d).

15 4. Espumas

Se generaron espumas de proteína (Fig. 7e,f) a partir de las soluciones que contienen 2.5 mM peroxodisulfato de amonio (APS), 100 μM tris(2,2'-bipiridil)diclororutenio(II) (Rubpy) y 10% $\mu\text{(p/v)}$ (AQ)₂₄NR3 o 2% (p/v) C_{16} NR4. Las soluciones de proteína se espumaron con aire. Para estabilizar, las proteínas de estructura de espuma resultante se reticularon por exposición a la luz visible de una lámpara de tungsteno por 1 min (Protocolo: PNAS 1999 Vol.96 págs. 6020-6024). Las espumas se secaron subsecuentemente a 95°C.

25

5. Geles

Las nanofibrillas de C_{16} NR4 a una concentración de 1% (p/v) mostraron un aspecto de gel que fácilmente podría ser interrumpido por agitación o cizallamiento. Para mejorar las propiedades mecánicas del gel se permitió que el APS y el Rubpy entraran en el gel por difusión para producir concentraciones finales de 10 mM de APS y 100 μM de Rubpy. Después de la reticulación inducida por la luz (ver la sección 4) pudieron obtenerse geles dimensionalmente estables (Fig. 7g).

30

6. Películas

35 6.1 Estado soluble de las proteínas de seda de araña

Con el fin de formar las películas los inventores usaron las dos proteínas de seda sintéticas, (AQ)₂₄NR3 y C_{16} , que se derivan de las proteínas de seda dragline ADF-3 y ADF-4 de la araña de jardín *Araneus diadematus*(ver además anteriormente para más explicaciones). Estas dos proteínas diferentes se eligieron en base a observaciones anteriores donde ADF-3 y ADF-4 así como sus derivados muestran un comportamiento marcadamente diferente con respecto a la solubilidad y al ensamblaje. Las soluciones acuosas de ambas proteínas pudieron prepararse mediante la disolución de proteínas liofilizadas en 6 M de tiocianato de guanidinio y la posterior eliminación de la sal por diálisis frente a un amortiguador con bajo contenido de sal tal como 5 mM de fosfato de potasio pH 8.0. Las proteínas liofilizadas también pudieron disolverse directamente en HFIP. La medición del dicroísmo circular (CD) de las soluciones de proteínas reveló una influencia diferente de los dos solventes sobre la estructura secundaria. En solución acuosa ambas proteínas mostraron un espectro de CD con un único mínimo a una longitud de onda por debajo de 200 nm que es indicativo de una proteína principalmente aleatoria (Fig.8). En contraste, los espectros de ambas proteínas en HFIP mostraron un mínimo a 201 - 202 nm y un mínimo adicional ((AQ)₂₄NR3) u hombro (C_{16}) a 220 nm que es indicativo de un mayor contenido α -helicoidal (Fig.8).

50

6.2 Formación de la película

Las películas se formaron sobre una superficie de poliestireno (o sobre el cristal de cuarzo para mediciones de CD) a partir de soluciones HFIP que contienen 2% p/v de proteínas. Después de la evaporación del solvente, (AQ)₂₄NR3 y C_{16} formaron películas transparentes que podrían ser fácilmente desprendidas de la superficie (Fig.9 y datos no mostrados). Suponiendo la evaporación completa del solvente y que la densidad de la película de proteína es idéntica al valor reportado de 1.3 g/cm³ para la seda de araña dragline, se calculó que el espesor de las películas estaba en el intervalo de 0.5 a 1.5 μm . Las películas moldeadas (recién preparadas) fabricadas de cualquier proteína se disolvieron tras el contacto con agua. Dado que la insolubilidad en agua es un requisito previo para la mayoría de las aplicaciones de películas de proteínas, los inventores buscaron un método de procesamiento con el objetivo de producir películas insolubles. Se conoce que el fosfato

55

60

de potasio induce agregación y formación de estructuras químicamente estables de las proteínas de seda empleadas. En consecuencia, el procesamiento (incubación) de las películas moldeadas con 1 M de fosfato de potasio resultó en la conversión de las películas en un estado insoluble en agua.

5 6.3 Estructura secundaria

Para investigar las propiedades estructurales de las películas de proteína, su estructura secundaria se investigó por espectroscopia CD. Las películas moldeadas revelaron un espectro con dos mínimos a 208 nm y 220 nm, indicativo de un alto contenido α -helicoidal (Fig. 10). Después de procesar con 1 M de fosfato de potasio, las películas revelaron un espectro con un solo mínimo a 218 nm que es típico para una estructura rica en lámina β . Por lo tanto, la transición de la solubilidad en agua a la insolubilidad en agua fue paralela a una conversión de la estructura secundaria de la proteína de la α -hélice a la lámina β .

15 6.4 Estabilidad química

Para probar la estabilidad química, las películas se expusieron a 8 M urea, 6 M de clorhidrato guanidinio y 6 M de tiocianato de guanidinio (Tabla 2). Las películas moldeadas de las dos proteínas así como también las películas procesadas de (AQ)₂₄NR3 fueron solubles en estos desnaturalizantes. En contraste, las películas procesadas de C₁₆ sólo podían ser disueltas en tiocianato de guanidinio. Esta estabilidad química notable de las películas de C₁₆ es idéntica a aquella de ADF-4 producida de forma recombinante y agrupada y de la seda dragline natural. Estudios anteriores correlacionaron las propiedades de ensamble y las estabildades de las estructuras ensambladas directamente con las secuencias de aminoácidos de las proteínas de seda. Por lo tanto, se puede concluir, que las propiedades de las películas de seda de araña pueden ser modificadas directamente por alteración de la estructura primaria de la proteína de la seda a través de la manipulación del gen de seda correspondiente.

25 6.5 Modificación de la película

Muchas aplicaciones de las películas de proteínas requieren la presencia de funcionalidades específicas en la superficie de la película. A fin de demostrar, que nuestras películas de proteínas de seda de araña pueden ser modificadas con moléculas orgánicas pequeñas, así como macromoléculas biológicas como las proteínas, la fluoresceína del cromóforo y la enzima β -galactosidasa se acoplaron químicamente a las películas de C₁₆ procesadas. El acoplamiento se alcanzó mediante la activación de los grupos carboxilo expuestos a la superficie de C₁₆ usando 1-etil-3-(3-dimetilaminopropilo)carbodiimida (EDC) (para más detalles de las reacciones ver material complementario que se indica a continuación). Las películas se incubaron con etilendiamina lo que conduce a la formación de una amida. El grupo amino libre restante de etilendiamina se acopló posteriormente al fluoresceinisotiocianato resultando en el enlace covalente eficaz de la fluoresceína (Fig.11A) a través de la formación de un derivado de tiourea estable. De manera similar, la incubación de β -galactosidasa con películas de C₁₆ activadas con EDC condujo a la formación de enlaces amida entre los grupos carboxilo de C₁₆ y aminas primarias (por ejemplo de residuos de lisina) de la β -galactosidasa que estaban accesibles en la superficie de la enzima. Después de lavados repetidos de tales películas modificadas, la actividad β -galactosidasa pudo detectarse mediante el uso de 5-bromo-4-cloro-3-indolil- β -D-galactopiranosido (X-Gal) como sustrato (Fig.11B).

40 6.6 Conclusión

Aquí, se pudo demostrar que las películas de proteína se pueden obtener de proteínas de seda de araña sintéticas. Las películas, que inicialmente eran solubles en agua, pueden ser procesadas con fosfato de potasio que conduce a insolubilidad en agua, que es un requisito importante para muchas aplicaciones. La comparación de las estabildades químicas de las películas realizadas a partir de dos proteínas de seda de araña sintéticas diferentes sugiere que las propiedades de las películas se basan en la estructura primaria de las proteínas. Por lo tanto, será posible generar proteínas de seda que forman películas que exhiben propiedades específicas. Dado que diferentes moléculas funcionales pueden unirse covalentemente a la superficie de la película, una gran variedad de aplicaciones técnicas o médicas puede abordarse en el futuro.

45 6.7 Materiales y resultados complementarios

55 **Preparación de las soluciones de proteína**

Se llevó a cabo la producción y purificación de proteínas como se describió previamente. Para obtener soluciones acuosas de (AQ)₂₄NR3 y C₁₆, proteína liofilizada se disolvió en 6 M de tiocianato de guanidinio a una concentración de 10 mg/ml y subsecuentemente se dializó contra 5 mM de fosfato de potasio pH 8.0. Los agregados se eliminaron por sedimentación a 15,000×g por 10 min. Las concentraciones de proteína se determinaron por fotometría en una cubeta de 1 cm de longitud de

la trayectoria a 276 nm mediante el uso de coeficientes de extinción calculados de $73950 \text{ M}^{-1}\text{cm}^{-1}$ para (AQ)₂₄NR3 y $46400 \text{ M}^{-1}\text{cm}^{-1}$ para C₁₆. Alternativamente, las proteínas de seda liofilizadas se disolvieron directamente en hexafluoroisopropanol (HFIP).

5 Análisis de la estructura secundaria

Los espectros de dicroísmo circular (CD) de UV lejano se obtuvieron mediante el uso de un espectropolarímetro Jasco 715 (Jasco International Co. Ltd., Tokio, Japón). Los espectros de las proteínas solubles se tomaron a una concentración de proteína de 200 µg/ml en 5 mM de fosfato de potasio (pH 8.0) o HFIP en una cubeta de cuarzo de 0.1 cm de longitud de la trayectoria a 20°C. Para medir las películas, 100 µl de una solución de 2 mg/ml de proteína en HFIP se extendieron sobre un cristal de cuarzo plano de 4 cm² y se secaron al aire antes de la medición de CD. La velocidad de barrido fue de 20 nm/min, el tamaño del paso fue de 0.2 nm, el tiempo de integración se estableció para 1 s y el ancho de banda fue de 1 nm. Se promediaron cuatro barridos.

15 Modificación de la película

1. Acoplamiento de fluoresceína a superficies de películas de C₁₆

Las películas se prepararon al extender 15 µl por pocillo de 20 mg/ml de C₁₆ en HFIP sobre la parte inferior de una placa de 24 pocillos. Después de la evaporación del HFIP, las películas se incubaron durante 5 minutos con 1 M de fosfato de potasio. Después de enjuagar con agua, los grupos carboxilo se activaron mediante incubación durante 15 min con 100 mM de ácido 2-(N-morfolino) etanosulfónico (MES) pH 5.0, 100 mM de 1-etil-3-(3-dimetilaminopropil)carbodiimida (EDC) y 20 mM de N-hidroxisulfo-succinimida (NHS). Posteriormente se añadió etilendiamina para producir una concentración final de 500 mM. Después de 2 h de incubación las películas se enjuagaron exhaustivamente con agua. Finalmente, las películas se incubaron durante 1 h con 1 mg/ml de fluoresceinisotiocianato en 100 mM de carbonato de sodio pH 9.0, seguido de un enjuague con agua y secado al aire.

2. Acoplamiento de β-galactosidasa a las superficies de películas de C₁₆

Las películas se prepararon y activaron como se describió anteriormente. Después de 15 min de incubación con EDC / NHS, las películas se enjuagaron con agua y posteriormente se incubaron durante 2 h con una solución que contenía 100 µg/ml de β-galactosidasa, 4 mM de KH₂PO₄, 16 mM de Na₂HPO₄, 115 mM de NaCl (PBS). Después de un enjuague exhaustivo con PBS, se probó la actividad enzimática en la superficie de la película.

35 Ensayo de β-galactosidasa

Las películas acopladas con β-galactosidasa se incubaron durante 16 h a temperatura ambiente con una solución que contenía 100 mM de fosfato de sodio pH 7.0, 10 mM de cloruro de potasio, 1 mM de sulfato de magnesio, 50 mM de β-mercaptoetanol y 2 mg/ml de 5-bromo-4-cloro-3-indolil-β-D-galactopiranosido (X-Gal).

7. Hidrogeles adicionales

La parte repetitiva de ADF-4 se compone generalmente de una sola unidad de repetición conservada que solamente presenta pequeñas variaciones. Los inventores combinaron estas variaciones y diseñaron un módulo consenso denominado C (GSSAAAAAASGPGGYGPENQGPSGPGGYGPGGP) (sec. con núm. de ident.: 5), que se multimerizó para obtener la proteína rep C₁₆, lo que resultará en una proteína de una masa molecular de 48 kDa.

El gen para C₁₆ de la seda se expresó en la cepa de *E. coli* BLR [DE3] (Novagen). Las células se cultivaron a 37 ° C en medio LB a una OD₆₀₀ = 0.5. Antes de la inducción con 1 mM de IPTG (isopropil-β-D-tiogalactósido), las células se cambiaron a 25°C. Las células se cosecharon después de 3-4 horas de inducción.

La proteína C₁₆ se purificó como se describe en Huemmerich y otros (40). Los sedimentos de C₁₆ se lavaron con 8 M de urea y se disolvieron en 6 M de tiocianato de guanidina (GdmSCN) antes de la diálisis frente a 10 mM de NH₄HCO₃. Los precipitados formados durante la diálisis se eliminaron por sedimentación a 50,000xg durante 30 min y las restantes proteínas solubles de la seda se liofilizaron. Antes del análisis la proteína liofilizada se disolvió en 6 M de GdmSCN seguido por diálisis frente a 10mM de Tris / HCl. Los agregados se eliminaron por sedimentación a 125,000 xg durante 30 min. Las concentraciones de proteínas se determinaron por fotometría en una cubeta de 1 cm de longitud de la trayectoria a 276 nm con el uso del coeficiente de extinción calculado (40).

C₁₆ se auto-ensambla en nanofibras a concentraciones entre 5 y 30 mg/ml después de la adición de 10% p/v de metanol

(Fig. 12). Sorprendentemente, a las concentraciones usadas las nanofibras conducen a la formación de una red de fibras que representa los hidrogeles. Los hidrogeles de C₁₆ pudieron interrumpirse fácilmente por agitación o cizallamiento. Para mejorar las propiedades mecánicas del gel se permitió que el peroxodisulfato de amonio (APS), y el tris (2,2'-bipiridil) diclororutenio (II) (Rubpy) entraran en el gel por difusión para producir concentraciones finales de 10 mM de APS y 100 μM de Rubpy. Para obtener geles dimensionalmente estables las proteínas se reticularon por exposición a luz visible de una lámpara de tungsteno durante 1 min (IV) (Fig. 13).

Se realizaron mediciones reológicas dinámicas de los hidrogeles reticulados y no reticulados mediante el uso de un Physica MCR 301 con una geometría de placa-placa de 25 mm. El espacio entre la placa superior y el plato de la muestra se estableció al mover primero la placa superior aproximadamente 2 mm por encima de la superficie de la muestra. La placa superior se bajó muy lentamente (5 μm/s), mientras se monitoreó la fuerza normal y se detuvo en una fuerza normal límite de 0.1 N.

Después de encontrar tamaños de espacio adecuados para las muestras, las muestras se desviaron a 0.5 Hz y una deformación de 1% hasta que la fuerza normal se equilibró a un valor constante. Las mediciones reológicas dinámicas se realizaron a temperatura ambiente al aplicar una tensión constante a la muestra. Las medidas reológicas se llevaron a cabo en muestras con concentraciones de proteínas en el intervalo de 5 a 30 mg/ml.

Las imágenes de AFM de los hidrogeles secos indican que las nanofibras son de aproximadamente 3 nm de diámetro y parecen ser semiflexibles, con una longitud de persistencia en el mismo orden de magnitud que su longitud (Figura 12). Muchas de las nanofibras parecen tener además una estructura ramificada. De las imágenes de AFM no pudo determinarse si las estructuras en forma ramificada son ramas físicas en cada fibra polimérica o si son el resultado de la agrupación de nanofibras.

Similar a las redes de polímeros más concentrados, el hidrogel de la proteína recombinante de la seda de araña C₁₆ demuestra un comportamiento viscoelástico. Cuando se aplica una tensión a las redes viscoelásticas de C₁₆ de la seda la deformación cambia lentamente con el tiempo y es proporcional a la tensión aplicada. La Figura 14 muestra el comportamiento tensión/deformación de los hidrogeles reticulados y no reticulados a una concentración de 10 mg/ml. El hidrogel no reticulado de C₁₆ de la seda tiene un módulo de cizalla inicial de 38 Pa. Sin embargo, a medida que aumenta la tensión el hidrogel no reticulado muestra una mayor respuesta de deformación a la tensión, y después de una deformación de 20% la respuesta es relativamente lineal. Con el aumento de la tensión la red continúa deformándose hasta que se alcanza una deformación de 90%, donde el hidrogel no reticulado se rompe y fluye. A diferencia de las redes de fibras no reticuladas, las redes reticuladas muestran una respuesta viscoelástica lineal en todas las deformaciones, tiene un módulo de cizalla mucho más alto de 820 Pa, y se rompe a una menor deformación de 30%.

Las medidas viscoelásticas dinámicas de las redes de fibras no reticuladas a una concentración polimérica de 20 mg/ml revelan que el módulo de almacenamiento (G') y el módulo de pérdida (G'') son muy dependientes de la frecuencia de oscilación (ω), tanto en el intervalo de ω alto como en el de ω bajo (Figura 15). La red demuestra un comportamiento viscoso a bajas frecuencias y un comportamiento elástico a frecuencias moderadas con un cruce a 0.49 Hz. El comportamiento observado del hidrogel es similar al esperado para una red de polímeros enredados y no es similar a lo que se esperaría de una solución cristalina líquida o fluido viscoso.

El hidrogel no reticulado de C₁₆ de la seda muestra además un comportamiento viscoelástico dinámico que es muy diferente al que se observa en los hidrogeles reticulados químicamente (Figura 15). A diferencia del comportamiento de la red de fibras no reticuladas, el módulo de almacenamiento de la red de fibras reticuladas es casi constante a todas las frecuencias, excepto a las frecuencias más altas probadas. El hidrogel reticulado de C₁₆ de la seda demuestra además un módulo de almacenamiento mayor y de pérdida menor que el que se observa en la red no-reticulada.

Como era de esperar, el módulo de almacenamiento del hidrogel reticulado es mayor que el de la red no reticulada para todas las concentraciones probadas (Figura 16). Sin embargo, inesperadamente los módulos de almacenamiento tanto de las redes reticuladas como no reticuladas aumenta con la concentración [c] y tienen una dependencia de [c]². En el caso de las redes reticuladas de biopolímeros semiflexibles lineales, donde la longitud de persistencia es mayor que el tamaño de malla, se espera que el módulo de almacenamiento de la red de polímeros tenga una dependencia de [c], que es cercana a la del hidrogel reticulado de C₁₆ de la seda. En el caso de las redes de biopolímeros semiflexibles lineales que se enredan pero no se reticularan, se espera que el módulo de almacenamiento tenga una dependencia mucho más baja de la concentración de [c]. Tal dependencia ha demostrado ser válida para otros biopolímeros tales como F-actina, pero no describe la dependencia del hidrogel no reticulado de la seda.

Esta discrepancia podría explicarse si las estructuras con formas ramificadas observadas en las imágenes de AFM son ramas físicas reales en la red polimérica. Se espera que el módulo de almacenamiento de una red de polímeros

semiflexibles ramificados muestre una dependencia de la concentración entre lo que se esperaría para la red de polímeros reticulados y no reticulados.

Las imágenes de AFM y los datos de reología son consistentes con el modelo de una red de polímeros semiflexibles ramificados. Sin embargo, el comportamiento a escala del módulo de almacenamiento de los hidrogeles no puede explicarse en el marco de los modelos más ampliamente aceptados para redes de polímeros semiflexibles lineales.

TABLA

Tabla 1

Propiedades seleccionadas de las construcciones sintéticas de la seda y las proteínas auténticas de la seda de araña ADF-3 y ADF-4.

	QAQ ₈	AQ ₁₂	C ₁₆	NR ₃	NR ₄	QAQ ₈ N R ₃	AQ ₁₂ N R ₃	C ₁₆ NR 4	ADF- 3	ADF- 4
Masa molecular [kDa] ^a	47.5	48.1	47.7	13.3	11.9	59.3	59.8	58.1	56.1	34.9
Coefficiente de extinción (276nm) [M ⁻¹ cm ⁻¹] ^b	23200	34800	46400	4423	1523	27550	39150	47850	n.d.	n.d.
Residuos de aminoácidos cargados ^c (positivo/negativo)	0/0	0/0	0/16	2/2	2/2	2/2	2/2	2/18	4/2	2/6
Gran promedio de hidropaticidad (GRAVY) ^d	-1.252	-0.987	-0.464	0.401	0.438	-0.918	-0.710	-0.294	-0.628	-0.075
Hidropaticidad normalizada e hidropaticidad promedio "límite"	0.361 0.413	0.390 0.413	0.448 0.440	0.545 0.413	0.548 0.413	n.d.	n.d.	n.d.	0.399 0.415	0.464 0.417
Temperatura del punto medio del desplegamiento térmico ^f	no	no	no	n.d.	n.d.	67°C	66°C	72°C	n.d.	n.d.
Solubilidad (p/v) ^g	> 30%	> 30%	8%	n.d.	n.d.	> 30%	> 30%	7%	> 28%	< 1%

^a La masa molecular de las proteínas modificadas genéticamente incluye la etiqueta T7.

^b Los coeficientes de extinción se calcularon de acuerdo con Gill y Hippel (23).

^c Los residuos de aminoácidos cargados solo se refieren a las secuencias génicas de la seda; Las etiquetas T7 comprenden una arginina adicional

^d La hidropaticidad se calculó como se describe previamente (39). La hidrofobicidad aumenta con los valores de hidropaticidad.

^e La hidropaticidad se normalizó a un intervalo entre 0 y 1. La hidropaticidad "límite" se calculó de acuerdo con Uversky y otros (35;36). Si los valores de la hidropaticidad normalizada están por debajo del valor "límite" se prevé que las proteínas están desplegadas intrínsecamente. Los valores de ADF-3 y ADF-4 se refieren a sus secuencias repetitivas solamente.

^f Las temperaturas del punto medio se determinaron mediante espectroscopía de CD.

5 ^g Los valores para ADF-3 y ADF-4 se tomaron de (18) y de resultados no publicados.

Tabla 2 Solubilidad de películas de proteína en desnaturalizantes. Las películas se consideraron como insolubles (-), en caso que una inmersión completa en el agente respectivo y agitación repetida durante un período de cinco minutos no resultó en un cambio de la apariencia óptica. En contraste, la solubilidad (+) se caracterizó por la desintegración completa de la película bajo las mismas condiciones.

	agua	8 M urea	6 M clorhidrato de guanidinio	6 M tiocianato de guanidinio
(AQ) ₂₄ NR3	+	+	+	+
as-cast				
(AQ) ₂₄ NR3	-	+	+	+
procesado				
C ₁₆	+	+	+	+
as-cast				
C ₁₆	-	-	-	+
procesado				

REFERENCIAS

1. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., y Savage, K. N. (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function, *J. Exp. Biol.* 202 Pt 23, 3295-3303.

2. Vollrath, F. y Knight, D. P. (2001) Liquid crystalline spinning of spider silk, *Nature* 410, 541-548.

3. Guerette, P. A., Ginzinger, D. G., Weber, B. H., y Gosline, J. M. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science* 272, 112-115.

4. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., y Lewis, R. (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences, *Science* 291, 2603-2605.

5. Simmons, A. H., Ray, E., y Jelinski, L. W. (1994) Solid-State ¹³C NMR of Nephila clavipes Dragline Silk Establishes Structure and Identity of Crystalline Regions, *Macromolecules* 27, 5235-5237.

6. Parkhe, A. D., Seeley, S. K., Gardner, K., Thompson, L., y Lewis, R. V. (1997) Structural studies of spider silk proteins in the fiber, *J. Mol. Recognit.* 10, 1-6.

7. van Beek, J. D., Hess, S., Vollrath, F., y Meier, B. H. (2002) The molecular structure of spider dragline silk: folding and orientation of the protein backbone, *Proc. Natl. Acad. Sci. U. S. A* 99, 10266-10271.

8. Hijirida, D. H., Do, K. G., Michal, C., Wong, S., Zax, D., y Jelinski, L. W. (1996) ¹³C NMR of Nephila clavipes major ampullate silk gland, *Biophys. J.* 71, 3442-3447.

9. Kenney, J. M., Knight, D., Wise, M. J., y Vollrath, F. (2002) Amyloidogenic nature of spider silk, *Eur. J. Biochem.* 269, 4159-4163.

10. Hronska, M., van Beek, J. D., Williamson, P. T., Vollrath, F., y Meier, B. H. (2004) NMR characterization of native liquid spider dragline silk from Nephila edulis, *Biomacromolecules.* 5, 834-839.

11. Kerkam, K., Viney, C., Kaplan, D., y Lombardi, S. (1991) Liquid crystallinity of natural silk secretions, *Nature* 349, 596-598.

12. Knight, D. P. y Vollrath, F. (1999) Liquid crystals and flow elongation in a spider's silk production line, *Proc. R. Soc. Lond.* 519-523.

13. Willcox, J., Gido, S., Muller, W., y Kaplan, D. (1996) Evidence of a Cholesteric Liquid Crystalline Phase in Natural Silk Spinning Processes, *Macromolecules* 29, 5106-5110.

14. Knight, D. P. y Vollrath, F. (2001) Changes in element composition along the spinning duct in a Nephila spider, *Naturwissenschaften* 88, 179-182.

15. Vollrath, F., Knight, D., y Hu, X. W. (1998) Silk production in a spider involves acid bath treatment, *Proc. R. Soc. Lond B Biol. Sci.* 265, 817-820.

16. Tillinghast, E. K., Chase, S. F., y Townley, M. A. (1984) Water extraction by the major ampullate duct during silk formation in the spider, *Argiope aurantia* Lucas, *J. Insect Physiol.* 30, 591-596.

17. Knight, D. P., Knight, M. M., y Vollrath, F. (2000) Beta transition and stress-induced phase separation in the spinning of spider dragline silk, *Int. J. Biol. Macromol.* 27, 205-210.
18. Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J. F., Duguay, F., Chretien, N., Welsh, E. A., Soares, J. W., y Karatzas, C. N. (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells, *Science* 295, 472-476.
- 5 19. Sambrook, J. y Russell, D. (2001) *Molecular Cloning*.
20. Kroll, D. J., Abdel-Malek Abdel-Hafiz, H., Marcell, T., Simpson, S., Chen, C. Y., Gutierrez-Hartmann, A., Lustbader, J. W., y Hoeffler, J. P. (1993) A multifunctional prokaryotic protein expression system: overproduction, affinity purification, and selective detection, *DNA Cell Biol.* 12, 441-453.
- 10 21. Reiling, H. E., Laurila, H., y Fiechter, A. (1985) Mass-Culture of *Escherichia-Coli* - Medium Development for Low and High-Density Cultivation of *Escherichia Coli-B/R* in Minimal and Complex Media, *Journal of Biotechnology* 2, 191-206.
22. Yee, L. y Blanch, H. W. (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*, *Biotechnology (N. Y.)* 10, 1550-1556.
23. Gill, S. C. y von Hippel, P. H. (1989) Calculation of Protein Extinction Coefficients from Amino-Acid Sequence Data, *Analytical Biochemistry* 182, 319-326.
- 15 24. Arcidiacono, S., Mello, C., Kaplan, D., Cheley, S., y Bayley, H. (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*, *Appl. Microbiol. Biotechnol.* 49, 31-38.
25. Prince, J. T., McGrath, K. P., DiGirolamo, C. M., y Kaplan, D. L. (1995) Construction, cloning, and expression of synthetic genes encoding spider dragline silk, *Biochemistry* 34, 10879-10885.
26. Fahnestock, S. R. e Irwin, S. L. (1997) Synthetic spider dragline silk proteins and their production in *Escherichia coli*, *Appl. Microbiol. Biotechnol.* 47, 23-32.
- 20 27. Lewis, R. V., Hinman, M., Kothakota, S., y Fournier, M. J. (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins, *Protein Expr. Purif.* 7, 400-406.
28. Scheller, J., Guhrs, K. H., Grosse, F., y Conrad, U. (2001) Production of spider silk proteins in tobacco and potato, *Nat. Biotechnol.* 19, 573-577.
- 25 29. Padgett, K. A. y Sorge, J. A. (1996) Creating seamless junctions independent of restriction sites in PCR cloning, *Gene* 168, 31-35.
30. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., y Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science* 277, 1453-1474.
- 30 31. Kim, T. D., Ryu, H. J., Cho, H. I., Yang, C. H., y Kim, J. (2000) Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes, *Biochemistry* 39, 14839-14846.
32. Uversky, V. N., Lee, H. J., Li, J., Fink, A. L., y Lee, S. J. (2001) Stabilization of partially folded conformation during alpha-synuclein oligomerization in both purified and cytosolic preparations, *J. Biol. Chem.* 276, 43495-43498.
- 35 33. Knight, D. P. y Vollrath, F. (2002) Biological liquid crystal elastomers, *Philos. Trans. R. Soc. Lond B Biol. Sci.* 357, 155-163.
34. Chen, X., Knight, D. P., y Vollrath, F. (2002) Rheological characterization of nephila spidroin solution, *Biomacromolecules.* 3, 644-648.
- 35 35. Uversky, V. N., Gillespie, J. R., y Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins* 41, 415-427.
- 40 36. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci.* 11, 739-756.
37. Arakawa, T. y Timasheff, S. N. (1985) Theory of protein solubility, *Methods Enzymol.* 114, 49-77.
38. Spohner, A., Unger, E., Grosse, F., y Weisshart, K. (2004) Conserved C-termini of Spidroins are secreted by the ampulácea mayorglands and retained in the silk thread, *Biomacromolecules.* 5, 840-845.
- 45 39. Kyte, J. y Doolittle, R. F. (1982) A simple method for displaying the hydrophatic character of a protein, *J. Mol. Biol.* 157, 105-132.
40. Huemmerich, D., Helsen, C.W., Oschmann, J., Rudolph, R. y Scheibel, T. (2004) Primary structure elements of dragline silks and their contribution to protein solubility and assembly, *Biochemistry* 43, 13604-13612.

LISTADO DE SECUENCIAS

- 50 <110> Technische Universität München
- <120> Proteínas recombinantes de la seda de araña
- <130> P19310
- <140>
- 55 <141>
- <150> US 60/590,196
- <151> 2004-07-22
- <160> 55
- <170> PatentIn versión 3.1

ES 2 525 095 T3

<210> 1
<211> 653
<212> PRT
<213> Araneus diadematus
<400> 1

5

10

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Asp Pro Asn Ser
1 5 10 15

15

Ala Arg Ala Gly Ser Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly
20 25 30

20

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

25

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

30

35

40

45

50

55

60

Pro Ser Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gly Gln Gly Pro

ES 2 525 095 T3

	65				70					75					80
5	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr	Gly
					85					90				95	
10	Pro	Gly	Ser	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly
				100				105						110	
15	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Asn	Gly	Pro	Gly
			115					120					125		
20	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro
				130				135					140		
25	Ala	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr	Gly	Pro	Gly
	145						150					155			160
30	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr
						165			170					175	
35	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr	Gly	Pro
				180					185					190	
40	Ser	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly
			195					200					205		
45	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr
		210					215						220		
50	Pro	Gly	Ser	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln
		225				230					235				240
55	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala
				245					250					255	
60	Ala	Ala	Ala	Gly	Gly	Tyr	Gly	Pro	Gly	Tyr	Gly	Gln	Gln	Gly	Pro
				260				265						270	
65	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser
			275					280					285		
70	Ala	Ser	Ala	Ala	Ser	Gly	Gly	Tyr	Gly	Pro	Gly	Ser	Gly	Gln	Gln
		290					295					300			
75	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly
						310					315				320

ES 2 525 095 T3

Ser Ala Ala Ala Ala Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln
 325 330 335
 5
 Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln
 340 345 350
 Gly Pro Gly Gly Gln Gly Pro Tyr Gly Pro Gly Ala Ser Ala Ala Ala
 355 360 365
 10
 Ala Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln Gln Gly Pro Gly
 370 375 380
 15
 Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln
 385 390 395 400
 Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln
 405 410 415
 20
 Gly Pro Gly Gln Gln Gly Pro Gly Gly Gln Gly Ala Tyr Gly Pro Gly
 420 425 430
 25
 Ala Ser Ala Ala Ala Gly Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly
 435 440 445
 30
 Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln
 450 455 460
 Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln
 465 470 475 480
 35
 Gly Pro Gly Gln Gln Gly Pro Tyr Gly Pro Gly Ala Ser Ala Ala Ala
 485 490 495
 40
 Ala Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln Gln Gly Pro Gly
 500 505 510
 Gln Gln Gly Pro Gly Gln Gln Gly Pro Val Gly Gln Gly Pro Tyr Gly
 515 520 525
 45
 Pro Gly Ala Ala Ser Ala Ala Val Ser Val Gly Gly Tyr Gly Pro Gln
 530 535 540
 50
 Ser Ser Ser Ala Pro Val Ala Ser Ala Ala Ala Ser Arg Leu Ser Ser
 545 550 555 560
 55
 Pro Ala Ala Ser Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser
 565 570 575
 60

ES 2 525 095 T3

Ser Gly Pro Thr Asn Gln Ala Ala Leu Ser Asn Thr Ile Ser Ser Val
 580 585 590
 Val Ser Gln Val Ser Ala Ser Asn Pro Gly Leu Ser Gly Cys Asp Val
 5 595 600 605
 Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala Leu Val Ser Ile Leu
 10 610 615 620
 Gly Ser Ser Ser Ile Gly Gln Ile Asn Tyr Gly Ala Ser Ala Gln Tyr
 15 625 630 635 640
 Thr Gln Met Val Gly Gln Ser Val Ala Gln Ala Leu Ala
 20 645 650
 <210> 2
 <211> 671
 <212> PRT
 25 <213> Araneus diadematus
 <400> 2
 Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ala Ala Arg Ala
 1 5 10 15
 Gly Ser Ser Ala Ala Ala Ala Ala Ala Ser Gly Ser Gly Gly Tyr
 20 25 30
 Gly Pro Glu Asn Gln Gly Pro Ser Gly Pro Val Ala Tyr Gly Pro Gly
 35 35 40 45
 Gly Pro Val Ser Ser Ala Ala Ala Ala Ala Ala Gly Ser Gly Pro
 40 50 55 60
 Gly Gly Tyr Gly Pro Glu Asn Gln Gly Pro Ser Gly Pro Gly Gly Tyr
 65 70 75 80
 Gly Pro Gly Gly Ser Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 45 85 90 95
 Ser Gly Pro Gly Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro
 50 100 105 110
 Gly Gly Ser Gly Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro
 55 115 120 125
 60

ES 2 525 095 T3

Gly Ala Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Pro Gly
 130 135 140
 Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Ala Tyr Gly
 145 150 155 160
 5
 Pro Gly Gly Pro Gly Ser Ser Ala Ala Ala Ser Gly Pro Gly Gly Tyr
 165 170 175
 10
 Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Gly Ser Gly Gly Tyr
 180 185 190
 15
 Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Gly Pro Gly Ala Ser
 195 200 205
 20
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Pro Gly Gly Tyr Gly
 210 215 220
 Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Ala Tyr Gly Pro Gly Gly
 225 230 235 240
 25
 Pro Gly Ser Ser Ala Ala Ala Ser Gly Pro Gly Gly Tyr Gly Pro Gly
 245 250 255
 30
 Ser Gln Gly Pro Ser Gly Pro Gly Ala Tyr Gly Pro Gly Gly Pro Gly
 260 265 270
 35
 Ser Ser Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Pro Gly Gly Tyr
 275 280 285
 40
 Gly Pro Gly Asn Gln Gly Pro Ser Gly Pro Gly Gly Tyr Gly Pro Gly
 290 295 300
 Gly Pro Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ser Gly Pro Gly
 305 310 315 320
 45
 Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Val Tyr Gly
 325 330 335
 50
 Pro Gly Gly Pro Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Gly Ser
 340 345 350
 Gly Pro Gly Gly Tyr Gly Pro Gly Asn Gln Gly Pro Ser Gly Pro Gly
 355 360 365
 55
 Gly Tyr Gly Pro Gly Gly Ser Gly Ser Ser Ala Ala Ala Ala Ala Ala

60

ES 2 525 095 T3

370 375 380

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

5 Gly Pro Gly Gly Ser Gly Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser
 405 410 415

10 Gly Pro Gly Ala Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly
 420 425 430

15 Pro Gly Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Ala
 435 440 445

20 Tyr Gly Pro Gly Gly Pro Gly Ser Ser Ala Ala Ala Ser Gly Pro Gly
 450 455 460

25 Gly Tyr Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Ala Tyr Gly
 465 470 475 480

30 Pro Gly Gly Tyr Gly Pro Gly Ser Ser Ala Ala Ala Ala Ala Ala Ser Gly
 485 490 495

35 Ser Arg Gly Tyr Gly Pro Gly Ser Gln Gly Pro Gly Gly Pro Gly Ala
 515 520 525

40 Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Pro Gly Gly Tyr
 530 535 540

45 Gly Pro Gly Ser Gln Gly Pro Ser Gly Pro Gly Tyr Gln Gly Pro Ser
 545 550 555 560

50 Gly Pro Gly Ala Tyr Gly Pro Ser Pro Ser Ala Ser Ala Ser Val Ala
 565 570 575

55 Ala Ser Arg Leu Ser Ser Pro Ala Ala Ser Ser Arg Val Ser Ser Ala
 580 585 590

60 Val Ser Ser Leu Val Ser Ser Gly Pro Thr Asn Gly Ala Ala Val Ser
 595 600 605

Gly Ala Leu Asn Ser Leu Val Ser Gln Ile Ser Ala Ser Asn Pro Gly
 610 615 620

ES 2 525 095 T3

Leu Ser Gly Cys Asp Ala Leu Val Gln Ala Leu Leu Glu Leu Val Ser
625 630 635 640

5 Ala Leu Val Ala Ile Leu Ser Ser Ala Ser Ile Gly Gln Val Asn Val
645 650 655

10 Ser Ser Val Ser Gln Ser Thr Gln Met Ile Ser Gln Ala Leu Ser
660 665 670

15

<210> 3
<211> 24
<212> PRT
20 <213> Secuencia artificial
<220>
<223> Módulo A (ADF-3)
<400> 3

25 Gly Pro Tyr Gly Pro Gly Ala Ser Ala Ala Ala Ala Ala Ala Gly Gly
1 5 10 15

Tyr Gly Pro Gly Ser Gly Gln Gln
20

30

<210> 4
35 <211> 20
<212> PRT
<213> Secuencia artificial
<220>
40 <223> Módulo Q (ADF-3)
<400> 4

Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly
1 5 10 15

45

Pro Gly Gln Gln
20

50

<210> 5
<211> 35
<212> PRT
55 <213> Secuencia artificial
<220>
<223> Módulo C (ADF-4)
<400> 5

60

ES 2 525 095 T3

1 Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Pro Gly Gly
 5 Tyr Gly Pro Glu Asn Gln Gly Pro Ser Gly Pro Gly Gly Tyr Gly Pro
 10 Gly Gly Pro
 15 <210> 6
 <211> 942
 <212> PRT
 <213> Nephila clavipes
 20 <400> 6
 25 Ala Cys Phe Thr Ser Ala Val Ile Phe Leu Phe Leu Ala Gln Cys Ala
 30 Ser Thr Tyr Gly Arg Gly Ile Ile Ala Asn Ser Pro Phe Ser Asn Pro
 35 Asn Thr Ala Glu Ala Phe Ala Arg Ser Phe Val Ser Asn Ile Val Ser
 40 Ser Gly Glu Phe Gly Ala Gln Gly Ala Glu Asp Phe Asp Asp Ile Ile
 45 Gln Ser Leu Ile Gln Ala Gln Ser Met Gly Lys Gly Arg His Asp Thr
 50 Lys Ala Lys Ala Lys Ala Met Gln Val Ala Leu Ala Ser Ser Ile Ala
 55 Glu Leu Val Ile Ala Glu Ser Ser Gly Gly Asp Val Gln Arg Lys Thr

ES 2 525 095 T3

				100					105					110			
	Asn	Val	Ile	Ser	Asn	Ala	Leu	Arg	Asn	Ala	Leu	Met	Ser	Thr	Thr	Gly	
			115					120					125				
5																	
	Ser	Pro	Asn	Glu	Glu	Phe	Val	His	Glu	Val	Gln	Asp	Leu	Ile	Gln	Met	
		130					135					140					
10																	
	Leu	Ser	Gln	Glu	Gln	Ile	Asn	Glu	Val	Asp	Thr	Ser	Gly	Pro	Gly	Gln	
	145					150					155					160	
15																	
	Tyr	Tyr	Arg	Ser	Ser	Ser	Ser	Gly	Gly	Gly	Gly	Gly	Gly	Gln	Gly	Gly	
					165					170					175		
20																	
	Pro	Val	Val	Thr	Glu	Thr	Leu	Thr	Val	Thr	Val	Gly	Gly	Ser	Gly	Gly	
				180					185					190			
25																	
	Gly	Gln	Pro	Ser	Gly	Ala	Gly	Pro	Ser	Gly	Thr	Gly	Gly	Tyr	Ala	Pro	
			195					200						205			
30																	
	Thr	Gly	Tyr	Ala	Pro	Ser	Gly	Ser	Gly	Ala	Gly	Gly	Val	Arg	Pro	Ser	
		210					215					220					
35																	
	Ala	Ser	Gly	Pro	Ser	Gly	Ser	Gly	Pro	Ser	Gly	Gly	Ser	Arg	Pro	Ser	
	225					230					235					240	
40																	
	Ser	Ser	Gly	Pro	Ser	Gly	Thr	Arg	Pro	Ser	Pro	Asn	Gly	Ala	Ser	Gly	
					245					250					255		
45																	
	Ser	Ser	Pro	Gly	Gly	Ile	Ala	Pro	Gly	Gly	Ser	Asn	Ser	Gly	Gly	Ala	
			260						265					270			
50																	
	Gly	Val	Ser	Gly	Ala	Thr	Gly	Gly	Pro	Ala	Ser	Ser	Gly	Ser	Tyr	Gly	
		275						280					285				
55																	
	Pro	Gly	Ser	Thr	Gly	Gly	Thr	Tyr	Gly	Pro	Ser	Gly	Gly	Ser	Glu	Pro	
		290					295					300					
60																	
	Phe	Gly	Pro	Gly	Val	Ala	Gly	Gly	Pro	Tyr	Ser	Pro	Gly	Gly	Ala	Gly	
	305					310					315					320	
65																	
	Pro	Gly	Gly	Ala	Gly	Gly	Ala	Tyr	Gly	Pro	Gly	Gly	Val	Gly	Thr	Gly	
				325						330					335		
70																	
	Gly	Ala	Gly	Pro	Gly	Gly	Tyr	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	
			340						345					350			

ES 2 525 095 T3

5 Tyr Gly Pro Gly Gly Ala Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ala
355 360 365

10 Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ala Gly Pro Gly Gly Tyr Gly
370 375 380

15 Pro Gly Gly Ala Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ala Gly Pro
385 390 395 400

20 Gly Gly Tyr Gly Pro Gly Gly Thr Gly Pro Gly Gly Tyr Gly Pro Gly
405 410 415

25 Gly Thr Gly Pro Gly Gly Val Gly Pro Gly Gly Ala Gly Pro Gly Gly
420 425 430

30 Tyr Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala
435 440 445

35 Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly
450 455 460

40 Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Ala Gly Pro
465 470 475 480

45 Ser Gly Ala Gly Leu Gly Gly Ala Gly Pro Gly Gly Ala Gly Leu Gly
485 490 495

50 Gly Ala Gly Pro Gly Gly Ala Gly Thr Ser Gly Ala Gly Pro Gly Gly
500 505 510

55 Ala Gly Pro Gly Gly Ala Gly Gln Gly Asp Ala Gly Pro Gly Gly Ala
515 520 525

60 Gly Arg Gly Gly Ala Gly Arg Gly Gly Val Gly Arg Gly Gly Ala Gly
530 535 540

65 Arg Gly Gly Ala Gly Arg Gly Gly Ala Arg Gly Ala Gly Gly Ala Gly
545 550 555 560

70 Gly Ala Gly Gly Ala Gly Gly Ser Gly Gly Thr Thr Ile Val Glu Asp
565 570 575

75 Leu Asp Ile Thr Ile Asp Gly Ala Asp Gly Pro Ile Thr Ile Ser Glu
580 585 590

80 Glu Leu Thr Ile Gly Gly Ala Gly Ala Gly Gly Ser Gly Pro Gly Gly
595 600 605

ES 2 525 095 T3

Ala Gly Pro Gly Asn Val Gly Pro Gly Arg Ser Gly Pro Gly Gly Val
 610 615 620

5 Gly Pro Gly Gly Ser Gly Pro Gly Gly Val Gly Pro Gly Ser Phe Gly
 625 630 635 640

10 Pro Gly Gly Val Gly Pro Gly Gly Ser Gly Pro Gly Gly Val Gly Ser
 645 650 655

15 Gly Gly Ser Gly Gln Gly Gly Val Arg Pro Ser Gly Ser Gly Pro Gly
 660 665 670

Gly Val Gly Thr Gly Gly Val Gly Pro Gly Gly Ala Gly Gly Pro Tyr
 675 680 685

20 Gly Pro Gly Gly Ser Gly Pro Gly Gly Ala Gly Ser Ala Gly Gly Thr
 690 695 700

25 Tyr Gly Pro Gly Gly Phe Gly Gly Pro Gly Gly Phe Gly Gly Pro Gly
 705 710 715 720

30 Gly Ala Gly Gly Pro Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly
 725 730 735

35 Pro Gly Gly Ala Gly Gly Pro Tyr Gly Pro Gly Gly Ala Gly Gly Pro
 740 745 750

40 Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly Pro Gly Gly Ala Gly
 755 760 765

45 Gly Ser Tyr Gly Leu Gly Gly Ala Gly Gly Ser Gly Gly Val Gly Pro
 770 775 780

Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ala Gly Pro Gly
 785 790 795 800

50 Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly
 805 810 815

55 Ser Gly Ser Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Ser
 820 825 830

Gly Pro Gly Gly Tyr Gly Pro Gly Gly Thr Gly Pro Gly Gly Ser Glu
 835 840 845

Ser Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Ser Gly Pro
 850 855 860

ES 2 525 095 T3

Gly Gly Ser Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly
 865 870 875 880

5 Gly Ser Gly Pro Ser Ser Phe Val Pro Gly Gly Ser Gly Pro Gly Gly
 885 890 895

10 Ser Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Val
 900 905 910

15 Gly Leu Gly Gly Ala Gly Arg Gly Gly Ala Gly Arg Gly Gly Ala Gly
 915 920 925

20 Ser Val Gly Ala Gly Arg Gly Gly Ala Gly Arg Gly Gly Thr
 930 935 940

25 <210> 7
 <211> 907
 <212> PRT
 <213> Nephila clavipes
 <400> 7

30 Gly Pro Gly Gly Val Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly
 1 5 10 15

35 Pro Gly Gly Ala Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro
 20 25 30

40 Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly
 35 40 45

45 Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly
 50 55 60

50 Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Tyr
 65 70 75 80

55 Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Thr Gly
 85 90 95

60 Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro
 100 105 110

Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly

ES 2 525 095 T3

		115				120					125					
5	Gly	Phe	Gly	Pro	Gly	Gly	Ser	Gly	Pro	Gly	Gly	Tyr	Gly	Pro	Gly	Gly
		130					135					140				
10	Ser	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Val	Gly	Pro	Gly	Gly	Phe
	145					150					155					160
15	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly
					165					170						175
20	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Pro
				180					185					190		
25	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Pro	Gly
			195					200					205			
30	Gly	Ala	Gly	Pro	Gly	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Ala
	210						215						220			
35	Gly	Gly	Ser	Gly	Gly	Ala	Gly	Gly	Ser	Gly	Gly	Thr	Thr	Ile	Ile	Glu
	225					230						235				240
40	Asp	Leu	Asp	Ile	Thr	Ile	Asp	Gly	Ala	Asp	Gly	Pro	Ile	Thr	Ile	Ser
					245					250						255
45	Glu	Glu	Leu	Thr	Ile	Ser	Gly	Ala	Gly	Gly	Ser	Gly	Pro	Gly	Gly	Ala
				260					265					270		
50	Gly	Pro	Gly	Gly	Val	Gly	Pro	Gly	Gly	Ser	Gly	Pro	Gly	Gly	Val	Gly
					275				280					285		
55	Pro	Gly	Gly	Ser	Gly	Pro	Gly	Gly	Val	Gly	Pro	Gly	Gly	Ser	Gly	Pro
					290			295					300			
60	Gly	Gly	Val	Gly	Pro	Gly	Gly	Ala	Gly	Gly	Pro	Tyr	Gly	Pro	Gly	Gly
	305					310						315				320
65	Ser	Gly	Pro	Gly	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Pro	Gly	Gly	Ala	Tyr
					325					330					335	
70	Gly	Pro	Gly	Gly	Ser	Tyr	Gly	Pro	Gly	Gly	Ser	Gly	Gly	Pro	Gly	Gly
					340				345					350		
75	Ala	Gly	Gly	Pro	Tyr	Gly	Pro	Gly	Gly	Glu	Gly	Pro	Gly	Gly	Ala	Gly
			355					360						365		

ES 2 525 095 T3

Gly Pro Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly Pro Gly Gly
 370 375 380
 5
 Ala Gly Gly Pro Tyr Gly Pro Gly Gly Glu Gly Gly Pro Tyr Gly Pro
 385 390 395 400
 10
 Gly Gly Ser Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly Pro Gly
 405 410 415
 15
 Gly Pro Tyr Gly Pro Gly Gly Glu Gly Pro Gly Gly Ala Gly Gly Pro
 420 425 430
 Tyr Gly Pro Gly Gly Val Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr
 435 440 445
 20
 Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ala Gly
 450 455 460
 25
 Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro
 465 470 475 480
 Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly
 485 490 495
 30
 Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Ser Gly Gly
 500 505 510
 35
 Ala Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr
 515 520 525
 Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Thr Gly
 530 535 540
 40
 Pro Gly Gly Thr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr Gly Pro
 545 550 555 560
 45
 Gly Gly Ser Gly Pro Gly Gly Ser Gly Pro Gly Gly Ser Gly Pro Gly
 565 570 575
 50
 Gly Tyr Gly Pro Ser Gly Ser Gly Pro Gly Gly Tyr Gly Pro Ser Gly
 580 585 590
 Ser Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr
 595 600 605
 55
 Gly Pro Gly Gly Ser Gly Ala Gly Gly Thr Gly Pro Gly Gly Ala Gly
 610 615 620
 60

ES 2 525 095 T3

Gly Ala Gly Gly Ala Gly Gly Ser Gly Gly Ala Gly Gly Ser Gly Gly
 625 630 635 640

5

Ala Gly Gly Ser Gly Gly Ala Gly Gly Ser Gly Gly Val Gly Gly Ser
 645 650 655

10

Gly Gly Thr Thr Ile Thr Glu Asp Leu Asp Ile Thr Ile Asp Gly Ala
 660 665 670

15

Asp Gly Pro Ile Thr Ile Ser Glu Glu Leu Thr Ile Ser Gly Ala Gly
 675 680 685

20

Gly Ser Gly Pro Gly Gly Ala Gly Pro Gly Gly Val Gly Pro Gly Gly
 690 695 700

25

Ser Gly Pro Gly Gly Val Gly Pro Gly Val Ser Gly Pro Gly Gly Val
 705 710 715 720

30

Gly Pro Gly Gly Ser Gly Pro Gly Gly Val Gly Ser Gly Gly Ser Gly
 725 730 735

35

Pro Gly Gly Val Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Ser
 740 745 750

40

Gly Gly Val Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Gly Phe
 755 760 765

45

Tyr Gly Pro Gly Gly Ser Glu Gly Pro Tyr Gly Pro Ser Gly Thr Tyr
 770 775 780

50

Gly Ser Gly Gly Gly Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly
 785 790 795 800

55

Pro Gly Ser Pro Gly Gly Ala Tyr Gly Pro Gly Ser Pro Gly Gly Ala
 805 810 815

60

Tyr Tyr Pro Ser Ser Arg Val Pro Asp Met Val Asn Gly Ile Met Ser
 820 825 830

65

Ala Met Gln Gly Ser Gly Phe Asn Tyr Gln Met Phe Gly Asn Met Leu
 835 840 845

70

Ser Gln Tyr Ser Ser Gly Ser Gly Thr Cys Asn Pro Asn Asn Val Asn
 850 855 860

75

Val Leu Met Asp Ala Leu Leu Ala Ala Leu His Cys Leu Ser Asn His
 865 870 875 880

ES 2 525 095 T3

Gly Ser Ser Ser Phe Ala Pro Ser Pro Thr Pro Ala Ala Met Ser Ala
 885 890 895
 5
 Tyr Ser Asn Ser Val Gly Arg Met Phe Ala Tyr
 900 905
 10
 <210> 8
 <211> 636
 <212> PRT
 <213> Araneus diadematus
 15
 <400> 8
 Ala Arg Ala Gly Ser Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly
 1 5 10 15
 20
 Gln Gln Gly Pro Gly Gln Gln Gly Pro Tyr Gly Pro Gly Ala Ser Ala
 20 25 30
 25
 Ala Ala Ala Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln Gln Gly
 35 40 45
 30
 Pro Ser Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gly Gln Gly Pro
 50 55 60
 35
 Tyr Gly Pro Gly Ala Ser Ala Ala Ala Ala Ala Ala Gly Gly Tyr Gly
 65 70 75 80
 40
 Pro Gly Ser Gly Gln Gln Gly Pro Gly Gly Gln Gly Pro Tyr Gly Pro
 85 90 95
 45
 Gly Ser Ser Ala Ala Ala Ala Ala Ala Gly Gly Asn Gly Pro Gly Ser
 100 105 110
 50
 Gly Gln Gln Gly Ala Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly
 115 120 125
 55
 Ala Ser Ala Ala Ala Ala Ala Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly
 130 135 140
 60
 Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gly Gln Gly Pro Tyr Gly
 145 150 155 160
 Pro Gly Ala Ser Ala Ala Ala Ala Ala Ala Gly Gly Tyr Gly Pro Gly

ES 2 525 095 T3

				165					170						175	
5	Ser	Gly	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr
				180					185					190		
10	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr	Gly	Pro
			195					200					205			
15	Gly	Ser	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly
		210					215					220				
20	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala
	225					230					235					240
25	Ala	Ala	Gly	Gly	Tyr	Gly	Pro	Gly	Tyr	Gly	Gln	Gln	Gly	Pro	Gly	Gln
				245					250						255	
30	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala
			260					265					270			
35	Ser	Ala	Ala	Ser	Gly	Gly	Tyr	Gly	Pro	Gly	Ser	Gly	Gln	Gln	Gly	Pro
		275					280						285			
40	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser
		290					295					300				
45	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Tyr	Gly	Pro	Gly	Ser	Gly	Gln	Gln
	305					310					315					320
50	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly
				325					330						335	
55	Pro	Gly	Gly	Gln	Gly	Pro	Tyr	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala	Ala
			340					345						350		
60	Ala	Ala	Gly	Gly	Tyr	Gly	Pro	Gly	Ser	Gly	Gln	Gln	Gly	Pro	Gly	Gln
		355					360						365			
65	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln
		370					375					380				
70	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gln	Gln	Gly
	385					390					395					400
75	Pro	Gly	Gln	Gln	Gly	Pro	Gly	Gly	Gln	Gly	Ala	Tyr	Gly	Pro	Gly	Ala
				405					410						415	

ES 2 525 095 T3

Ser Ala Ala Ala Gly Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln
 420 425 430
 5
 Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln
 435 440 445
 10
 Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly Pro Gly Gln Gln Gly
 450 455 460
 15
 Pro Gly Gln Gln Gly Pro Tyr Gly Pro Gly Ala Ser Ala Ala Ala Ala
 465 470 475 480
 20
 Ala Ala Gly Gly Tyr Gly Pro Gly Ser Gly Gln Gln Gly Pro Gly Gln
 485 490 495
 25
 Gln Gly Pro Gly Gln Gln Gly Pro Gly Gly Gln Gly Pro Tyr Gly Pro
 500 505 510
 30
 Gly Ala Ala Ser Ala Ala Val Ser Val Gly Gly Tyr Gly Pro Gln Ser
 515 520 525
 35
 Ser Ser Val Pro Val Ala Ser Ala Val Ala Ser Arg Leu Ser Ser Pro
 530 535 540
 40
 Ala Ala Ser Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Ser
 545 550 555 560
 45
 Gly Pro Thr Lys His Ala Ala Leu Ser Asn Thr Ile Ser Ser Val Val
 565 570 575
 50
 Ser Gln Val Ser Ala Ser Asn Pro Gly Leu Ser Gly Cys Asp Val Leu
 580 585 590
 55
 Val Gln Ala Leu Leu Glu Val Val Ser Ala Leu Val Ser Ile Leu Gly
 595 600 605
 60
 Ser Ser Ser Ile Gly Gln Ile Asn Tyr Gly Ala Ser Ala Gln Tyr Thr
 610 615 620
 Gln Met Val Gly Gln Ser Val Ala Gln Ala Leu Ala
 625 630 635

ES 2 525 095 T3

<210> 9
 <211> 410
 <212> PRT
 <213> Araneus diadematus
 <400> 9

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

ES 2 525 095 T3

	225					230					235					240
5	Gly	Pro	Gly	Ser	Gln	Gly	Pro	Ser	Gly	Pro	Gly	Gly	Ser	Gly	Gly	Tyr
					245					250					255	
10	Gly	Pro	Gly	Ser	Gln	Gly	Gly	Ser	Gly	Pro	Gly	Ala	Ser	Ala	Ala	Ala
				260					265					270		
15	Ala	Ala	Ala	Ala	Ala	Ser	Gly	Pro	Gly	Gly	Tyr	Gly	Pro	Gly	Ser	Gln
			275					280					285			
20	Gly	Pro	Ser	Gly	Pro	Gly	Tyr	Gln	Gly	Pro	Ser	Gly	Pro	Gly	Ala	Tyr
		290					295					300				
25	Gly	Pro	Ser	Pro	Ser	Ala	Ser	Ala	Ser	Val	Ala	Ala	Ser	Val	Tyr	Leu
	305					310					315					320
30	Arg	Leu	Gln	Pro	Arg	Leu	Glu	Val	Ser	Ser	Ala	Val	Ser	Ser	Leu	Val
					325					330					335	
35	Ser	Ser	Gly	Pro	Thr	Asn	Gly	Ala	Ala	Val	Ser	Gly	Ala	Leu	Asn	Ser
				340					345					350		
40	Leu	Val	Ser	Gln	Ile	Ser	Ala	Ser	Asn	Pro	Gly	Leu	Ser	Gly	Cys	Asp
			355					360					365			
45	Ala	Leu	Val	Gln	Ala	Leu	Leu	Glu	Leu	Val	Ser	Ala	Leu	Val	Ala	Ile
		370					375					380				
50	Leu	Ser	Ser	Ala	Ser	Ile	Gly	Gln	Val	Asn	Val	Ser	Ser	Val	Ser	Gln
	385					390					395					400
55	Ser	Thr	Gln	Met	Ile	Ser	Gln	Ala	Leu	Ser						
					405					410						

ES 2 525 095 T3

<210> 10
 <211> 140
 <212> PRT
 <213> Secuencia artificial
 5 <220>
 <223> NR3 (ADF-3)
 <400> 10

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

15 Ala Ala Ser Ala Ala Val Ser Val Gly Gly Tyr Gly Pro Gln Ser Ser
 20 25 30

20 Ser Ala Pro Val Ala Ser Ala Ala Ala Ser Arg Leu Ser Ser Pro Ala
 35 40 45

25 Ala Ser Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Ser Gly
 50 55 60

30 Pro Thr Asn Gln Ala Ala Leu Ser Asn Thr Ile Ser Ser Val Val Ser
 65 70 75 80

35 Gln Val Ser Ala Ser Asn Pro Gly Leu Ser Gly Cys Asp Val Leu Val
 85 90 95

40 Gln Ala Leu Leu Glu Val Val Ser Ala Leu Val Ser Ile Leu Gly Ser
 100 105 110

45 Ser Ser Ile Gly Gln Ile Asn Tyr Gly Ala Ser Ala Gln Tyr Thr Gln
 115 120 125

50 Met Val Gly Gln Ser Val Ala Gln Ala Leu Ala Gly
 130 135 140

<210> 11
 <211> 125
 <212> PRT
 55 <213> Secuencia artificial
 <220>
 <223> NR4 (ADF-4)
 <400> 11

60

ES 2 525 095 T3

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

 Ala Tyr Gly Pro Ser Pro Ser Ala Ser Ala Ser Val Ala Ala Ser Arg
 5 20 25 30

 Leu Ser Ser Pro Ala Ala Ser Ser Arg Val Ser Ser Ala Val Ser Ser
 10 35 40 45

 Leu Val Ser Ser Gly Pro Thr Asn Gly Ala Ala Val Ser Gly Ala Leu
 50 55 60

 Asn Ser Leu Val Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu Ser Gly
 15 65 70 75 80

 Cys Asp Ala Leu Val Gln Ala Leu Leu Glu Leu Val Ser Ala Leu Val
 20 85 90 95

 Ala Ile Leu Ser Ser Ala Ser Ile Gly Gln Val Asn Val Ser Ser Val
 25 100 105 110

 Ser Gln Ser Thr Gln Met Ile Ser Gln Ala Leu Ser Gly
 115 120 125

30

<210> 12
 <211> 1959
 <212> ADN
 <213> Araneus diadematus

35

<400> 12

40

45

50

55

60

ES 2 525 095 T3

atggctagca tgactggtgg acagcaaatg ggtcgggatc cgaattcggc acgagccgga 60
 tctggacaac aaggacccgg acaacaagga cccggacaac aaggacccgg acaacaagga 120
 ccatatggac cgggtgcatc cgccgcagca gcagccgctg gaggttatgg acccggatct 180
 5 ggacaacaag gacccagcca acaaggacct ggccaacaag gacccggtgg tcaaggacca 240
 tatggacccg gtgcatccgc cgccgcagca gccgctggtg gatatggacc cggttccgga 300
 caacaaggac caggaggtca aggaccatat ggacctggtt catccgctgc cgcagcagcc 360
 10 gctggaggtta atggacccgg atctggacaa caagggcccg gtcaacaagg tcctggacaa 420
 caaggacccg gtgcatccgc cgccgcagca gccgctggag gatacggacc cggatctgga 480
 15 caacaaggac cgggacaaca aggaccagga ggtcaaggac catatggacc tgggtgcatcc 540
 gccgctgcag cagccgctgg aggatacggga cccggatctg gacaacaagg acccggacaa 600
 caaggaccag gaggtcaagg accatatgga cccggtgcat ccgctgcagc agcagccgct 660
 20 ggaggttatg gacccggatc tggacaacaa ggacccggac aacaaggacc tggacaacaa 720
 ggacccggtg gtcaaggacc atatggacc gccgcatccg ccgcccagc agccgctgga 780
 ggatacggac ccggttatgg acagcaagga ccaggacaac aaggaccagg aggtcaagga 840
 25 ccatatggac ctggtgcatc cgccgcctca gcagcctctg gaggatacgg acccggatct 900
 ggacaacaag gacccggaca acaaggacct ggaggtcaag gaccatatgg acctggtgca 960
 tccgccgcag cagcagccgc tggaggttat ggacccggat ctggacaaca aggaccaggc 1020
 30 caacaaggac ccggtcaaca aggacctgga caacaaggac ccggtggtca aggaccatat 1080
 ggacctggtg catccgccgc agcagcagcc gctggaggtt atggacccgg atctggacaa 1140
 35 caaggacccg gtcaacaagg acccggtaa caaggacccg gtcaacaagg acccggtaa 1200
 caaggacccg gccacaagg acccggtaa caaggacccg gccacaagg acctggtcaa 1260
 caaggtcccg gtggtcaagg ggcataatgga cctggtgcat ccgccgcagc aggagccgct 1320
 40 ggaggttatg gacccggatc tggacaacaa ggacccggac aacaaggacc cggacaacaa 1380
 ggacccggac aacaaggacc cggacaacaa ggacccggac aacaaggacc cggacaacaa 1440
 ggacccggac aacaaggacc atatggacct ggtgcatccg ccgcccagc agccgctgga 1500
 45 gggttatggac ccggatctgg acaacaagga cccggccaac aaggacctgg acaacaagga 1560
 cccggtggtc aaggaccata tggacctggt gcggcttctg cagctgtatc tgttgagga 1620
 50 tatggaccac aaagctcctc ggctcctggt gcatcagcag ccgcttctcg cctttcttct 1680
 ccagcggcca gttctagagt ttcacggct gtatcatctt tggatatctag tggacctact 1740
 aatcaagctg cactttctaa tactatcagt agcgttgtat cgcaagttag tgcaagtaat 1800
 55 cctggtcttt ctggttgcga tgtacttgtg caagcattgc tcgaagttgt atcggccctg 1860
 gtatctatcc ttggatcttc tagtatcggg caaatlaact atggtgcctc tgctcagtac 1920
 acccaaatgg taggtcaatc tgtagctcaa gcccttgct 1959
 60

ES 2 525 095 T3

<210> 13
<211> 2013
5 <212> ADN
<213> Araneus diadematus

<400> 13

10

15 atggctagca tgactggtgg acagcaaatg ggtcgcgcgg cacgagcagg atcttcagca 60
gcagcggcgg cggcagcaag tggatctgga ggatacggac ctgaaaacca aggaccatct 120
ggacctgtag catatggacc tggaggacce gtatcttcag ctgcagcagc agccgctgca 180
20 ggaagtggac ctggtggata cggacctgaa aaccaaggac catctggacc cggaggatat 240
ggacctggtg gttccggatc ttcagcagca gcagcagccg ctgcagcaag tggacctgga 300
ggatatggac ctggaagcca aggaccatct ggacctggtg gatccggagg atatggtccc 360
25 ggaagccaag ggccatctgg acctggtgca tcttcggcag cagcagcagc cgctgcagca 420
agtggacctg gaggatatgg acctggaagc caaggaccat ctggacctgg agcatatgga 480

30

35

40

45

50

55

60

ES 2 525 095 T3

cctggtggac cgggatcttc agctgcagca agtggacctg gaggatatgg acctggaagc 540
 caaggaccat ctggacctgg tggatccgga ggatatggtc ccggaagcca agggccatct 600
 5 ggacctggtg ggccctggtgc atctgcggca gcagcagcag ccgctgcagc aagtggacct 660
 ggaggatatg gacctggaag ccaaggacca tctggacctg gagcatatgg acctggtgga 720
 10 cccggatctt cagctgcagc aagtggacct ggaggatatg gacctggaag ccaaggacca 780
 tctggacctg gagcatatgg acctggtgga cccggatctt cagctgcagc agcagccgct 840
 15 gcaggaagtg gacctggtgg atacggacct ggaaaccaag gaccatctgg acccggagga 900
 tatggacctg gtggtcccgg atcttcagca gcagcagccg ctgcagcaag tggacctgga 960
 ggatatggac ctggaagcca aggaccatct ggacctggag tatatggacc tgggtgacct 1020
 20 ggatcttcag ctgcagcagc agccgctgca ggaagtggac ctggtggata cggacctgga 1080
 aaccaaggac catctggacc cggaggatat ggacctggtg gttccggatc ttcagcagca 1140
 25 gcagcagccg ctgcagcaag tggacctgga ggatatggac ctggaagcca aggaccatct 1200
 ggacctggtg gatccggagg atatggtccc ggaagccaag ggccatctgg acctggtgca 1260
 30 tcttcggcag cagcagcagc cgctgcagca agtggacctg gaggatatgg acctggaagc 1320
 caaggaccat ctggacctgg agcatatgga cctggtggac ccggatcttc agctgcagca 1380
 agtggacctg gaggatatgg acctggaagc caaggaccat ctggtcctgg agcatatgga 1440
 35 cctggtggac cgggatcttc agctgcagca gccgctgcag caagtggacc tggaggatat 1500
 ggacctggaa gccaaaggacc atctggacct ggtggatccc gaggatatgg tcccggaagc 1560
 40 caaggacctg gtgggcctgg agcatctgcg gcagcagcag cagccgctgc agcaagtgga 1620
 cctggaggat atggacctgg aagccaagga ccatctggac ctggatatca aggccctagt 1680
 45 ggtcctggag catatggccc atctccttct gcttccgcat ccgttgcagc ctctcgttta 1740
 tcttcgctg cagcctcgtc tagagtgtct tccgctgtat cgtctttagt gtctagcggga 1800
 cctacgaatg gtgctgctgt ttctggagct ttgaatagtt tagtatctca gattagtgca 1860
 50 agtaatccag gtttatcggg atgtgatgct cttgtgcagg cattattgga attagtgtct 1920
 gctcttgtgg caattctttc atctgcaagt attggccaag tcaacgtcag ctctgttagt 1980
 55 cagtcaactc aatgattag ccaagctctt tca 2013

60

ES 2 525 095 T3

<210> 14
 <211> 420
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> NR3 (ADF-3)
 <400> 14

10 atggctagca tgactggtgg acagcaaatg ggtcgcggat ccatgggtgc ggcttctgca 60
 gctgtatctg ttggaggata tggaccacaa agctcctcgg ctctctgttg atcagcagcc 120
 gcttctcggc tttcttctcc agcggccagt tctcgtgttt catcggctgt atcatctttg 180
 15 gtatctagtg gacctactaa tcaagctgca ctttctaata ctatcagtag cgttgtatcg 240
 caagttagtg caagtaatcc tggcttttct ggttgcgatg tacttgtgca agcattgctc 300
 gaagttgtat cggccctggt atctatcctt ggatcttcta gtatcgggca aattaactat 360
 20 ggtgcctctg ctcagtacac ccaaatggta ggtcaatctg tagctcaagc ccttgctggc 420

25
 <210> 15
 <211> 375
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> NR4 (ADF-4)
 <400> 15

35 atggctagca tgactggtgg acagcaaatg ggtcgcggat ccatgggagc atatggccca 60
 tctccttctg ctcccgcatc cgttgcagcc tctcgtttat ctccgcctgc agcctcgtct 120
 cgtgtgtctt ccgctgtatc gtcttttagt tctagcggac ctacgaatgg tgctgctggt 180
 40 tctggagctt tgaatagttt agtatctcag attagtgcaa gtaatccagg tttatcggga 240
 tgtgatgctc ttgtgcaggc attattggaa ttagtgtctg ctcttggtgg aattctttca 300
 tctgcaagta ttggccaagt caacgtcagc tctgttagtc agtcaactca aatgattagc 360
 45 caagctcttt caggc 375

50
 <210> 16
 <211> 2828
 <212> ADN
 55 <213> Nephila clavipes
 <400> 16
 gcttgcttta cctcggcagt gatatttctt ttcttagcgc agtgtgcgtc gacgtacgga 60

ES 2 525 095 T3

agggggatta tagccaactc ccctttctca aaccctaaca cagcgggaagc ttttgcacga 120
 tctttcgtga gcaatattgt ttctagtggg gaatttggag cccaaggagc cgaagacttc 180
 gatgacataa ttcagagtct catacaggcc cagagcatgg gcaaagggcg gcatgatacg 240
 5 aaggccaagg cgaaagcgat gcaggtagcc cttgcttctt ctatagccga attggttatt 300
 gcagaaagca gcggaggcga tgtgcaacgc aaaaccaacg ttatctccaa cgctttgaga 360
 10 aacgccttga tgtctacaac aggcagccca aacgaagagt tcgtccatga agttcaagac 420
 ctcatccaga tgttatctca agaacagatc aacgaggtag atacttcagg accagggcag 480
 tactacaggt cgtctctctc cggtggagga ggtggaggac aaggaggtcc tgtagttact 540
 15 gaaacactga cegttacagt tggcggatcc ggtggagggc aaccttcagg tgcaggtcct 600
 agtggtagac gtggatatgc accaactgga tacgccccaa gcggctcagg tgcaggtggc 660
 gttcgacctg gtgcctccgg tccaagtggg agtggacctg gtggtagatc tcgtcctagt 720
 20 agtagtggac ctagtggaac tcgtcccagc cctaattggtg caagtggatc tagccctggt 780
 ggtatcgcac ctggtggatc caattctggt ggtgctggag tatccggcgc aactggagga 840
 25 cctgcatcca gcggctccta cggaccagga agtacaggtg gaacatatgg acctagtgga 900
 ggaagtgaac ctttcggacc aggagtggct ggaggaccat acagcccagg tggagctgga 960
 cctggtggtg caggtggagc ctatggacca ggaggtgtag gaactggtgg agccggacca 1020
 30 ggaggttacg gacctggtgg agccggacca ggaggttatg gacctggtgg agccggacca 1080
 ggaggttacg gacctggtgg agctggacca ggaggttacg gacctggtgg agctggacct 1140
 ggaggttacg gacctggtgg agctggacct ggaggttacg gacctggtgg agctggacct 1200
 35 ggaggttacg gacctggtgg aactggacct ggtggatacg gacctggtgg aactggacct 1260
 ggaggagtgg gacctggagg agctggacca ggaggatatg gacctggtgg tgctggacct 1320
 40 ggtggtgctg gacctggtgg tgctggacct ggtggtgctg gacctggtgg tgctggacct 1380
 ggtggtgctg gacctggtgg atacggccct ggtggatctg gacctggtgg tgctggacct 1440
 agtggtgccg gacttggtgg tgctggacct ggaggtgccc gacttggtgg agcaggacct 1500
 45 ggaggagcag gaaccagtgg tgccggaccc ggtggagcag gaccgggtgg agcaggacaa 1560
 ggtgatgctg gaccgggtgg tgcaaggacgt ggaggagcag gtcgtggtgg tgtaggtcgt 1620
 ggtggtgacg gtcgtggagg tgcaaggacgt ggtggagcga gagggtgctgg tggagcagga 1680
 50 ggtgctggtg gacgaggagg atccggcggc acaacaatcg tagaggactt ggatattaca 1740
 attgatggtg cagatggccc gataacaata tcagaagaat taacaatcgg tggagcaggc 1800
 55 gctggaggtt ccggaccocg tggctctgga ccagaaacg ttggacctgg tcgctctgga 1860
 ccaggaggag taggacctgg tggctctgga ccaggaggcg taggacctgg tagctttgga 1920
 ccaggaggcg taggacctgg tggctccgga ccaggaggcg taggatctgg tggctccgga 1980
 60

ES 2 525 095 T3

caaggaggag taagacctag tggctccgga ccaggtggcg taggaactgg aggcgtagga 2040
cccggtagtg ctggaggacc ttacggctect ggtggttccg gaccggagg tgcaggaagc 2100
5 gctggaggaa cttatggacc tggtaggttc ggaggaccgc gtggtttcgg aggaccgggt 2160
ggtgctggtg gaccctacgg tccaggtggt gctggtggac cctacggacc aggtggtgct 2220
10 ggtggaccct acggaccagg tggtagctgt ggaccctacg ggccgggtgg tgctggtgga 2280
ccctacgggc cgggaggtgc tggtaggacc tacgggctgg gtggtgctgg tggatcagga 2340
ggtgtaggac ctggtggaag tggacctgga ggtagtaggac ccggtggagc gggacctgga 2400
15 ggttacggac ccggtggttc tggtagcagg ggtagcggac ctggcgggttc tggatctggt 2460
ggtagcggac ctggaggttc tggacctggt ggtagctggac ctggtggata cggacctggt 2520
20 ggtactggac ctggtggttc tgaatctggt ggtagcggac ctggtggata tggacctggc 2580
ggttctggac ctggtggata tggacctggc ggtagctggac ctggtggata cggacctggt 2640
ggttctggac ctggtggata tggacctggc ggtagctggac ctggtggata cggacctggt 2700
25 ggcgctggac ccggtggcg tggacctggt ggtgtaggac ttggaggtgc tggactggt 2760
ggagctggac gtggtggagc tggaggtgtt ggagctggac gtggtggagc tggactggt 2820
30 ggaactgg 2828

<210> 17
<211> 2724
<212> ADN
35 <213> Nephila clavipes
<400> 17

ggaccaggag gtgtaggacc tggtaggaagt ggacctggag gttatggacc cggtaggagct 60
40 ggacctggag gttacggacc tggtaggttct ggtccagggt gatacggacc cggtaggttcg 120
ggaccaggag gatacggacc tggcggttct ggacctggtg gatacggacc aggcggttct 180
45 ggacctggtg gatacggacc aggcggttct ggacctggtg gatacggacc tggtaggatat 240
ggacctggtg gttctggacc tggtaggatat ggacctggtg gtactggacc tggtaggttct 300
ggaccggcg gatacggacc tggtaggttct ggacctggcg gttctggacc tggtaggatac 360
50 ggacctggtg gttctggacc tggcggtttt ggacctggcg gttctggacc tggtaggatac 420
ggacctggtg gctctggacc cggtaggtgct ggtccgggtg gtgtaggacc cggtaggtttt 480
55 ggacctggtg gtgctggacc cggtaggagct ggacctggtg gtgctggacc tggtaggtgct 540
ggacctggtg gtgctggacc tggtaggagct ggacctggtg gtgctggacc tggtaggagct 600

60

ES 2 525 095 T3

	ggacctggtg gtgctggacc tggtaggact ggacctggtg gtgctggtgg cgctggagga	660
	gcaggcggag caggagggtc aggtggagca ggaggatccg gcggtacaac aatcatagaa	720
	gacttggata ttacaattga tggcgctgat ggcccgataa cgatttcaga agaattaaca	780
5	attagtgggtg ctggagggtc cggaccgggt ggtgctggac caggagggtg agggcctggt	840
	ggctccggac caggagggtg aggacctgga ggctctggac caggagggtg aggacctggt	900
10	ggttctggtc caggaggcgt aggacctggt ggtgctggtg gaccttacgg acctggcgggt	960
	tctggacctg gagggtgcagg cggagctgga ggacctggtg gagcatacgg acctgggtgga	1020
	tcatatggac ctggtggttc cggaggaccc ggtggtgctg gcggaccata cggacctgga	1080
15	ggtgaaggac ccggtggtgc tggcggaccc tacggacctg gtggtgcagg tggaccttac	1140
	ggcccagggtg gtgcagggtg acctacgga ccagggtggtg aagggtggacc ctacggacca	1200
	ggtggatcat acggaccggg tggtagtgggt ggaccatacg gaccagggtg acctacgga	1260
20	cctggagggtg aaggaccagg tggtagtggc ggacctatg gaccaggagg tgtaggacct	1320
	ggtggaagtg gacctggagg ttatggacct ggtggaagtg gacctggagg ttatggacct	1380
	ggtggagctg gacctggagg ttacggacct ggtggttctg gtccagggtg atacggaccc	1440
25	ggtggttctg gtccagggtg atacggaccc ggtggttccg gaccaggagg atacggacct	1500
	ggcggttctg gacctggtg atacggatct ggcggtagctg gacctggtg atacggacct	1560
30	ggcggttctg gacctggtg atacggatct ggaggttctg gacctggtg ttatggacct	1620
	ggtggtactg gacctggtg tactggacct ggtggttctg gacctggcgg atacggacct	1680
	ggtggttctg gacctggcgg ttctggacct ggcggttctg gacctggtg atacggacct	1740
35	agtggttcgg gacctggtg atacggacct agtggttctg gacctggcgg atacggacct	1800
	ggcggttctg gacctggtg atacggacct ggtggtctctg gagccggtg tactggacct	1860
	ggtggcgtg gaggagcagg cggagcagga ggttcagggtg gagcaggagg ttcagggtgt	1920
40	gcaggagggt cagggtggagc aggaggttca ggtggagtag gaggatccg cggtacaaca	1980
	atcacccaag acttggatat tacaattgat ggcgcagatg gcccgataac gatttcagaa	2040
45	gaattaacaa ttagtgggtc tggaggttct ggaccgggtg gtgctggacc aggtggtgta	2100
	ggcctggtg gctctggacc aggaggtgta ggacctggag tctctggacc aggaggcgtg	2160
	ggacctggtg gttctggacc aggaggcgtg ggttctggtg gttctggacc aggaggcgtg	2220
50	ggacctggtg gttacggacc tggaggttct ggatcaggag gcgtaggacc tgggtggttac	2280
	ggacctggag gttcaggagg attttacgga cctggagggt cagaaggacc ttatggacct	2340
	agtggaactt atggttctgg aggaggatat ggtcctggtg gtgctggagg accatatgga	2400
55	cctggaagtc ctggaggagc ttatggacct ggaagcctg gaggagctta ttatcctagc	2460
	tcgcgtgttc ccgatatggt gaatggtata atgagtgcta tgcaaggatc tggttttaac	2520
60		

ES 2 525 095 T3

taccaaatgt ttggtaatat gctatcacia tattcgtctg gttcaggaac atgcaatcca 2580
 aataatgta atgtttgat ggatgctttg ttagctgctt tgcactgtct aagtaaccac 2640
 5 ggatcatcat cttttgcacc ttctccaact cgggctgcta tgagtgcgta ttctaattct 2700

10 <210> 18
 <211> 52
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> oligonucleótido sintético.

<400> 18
 gatcgaggag gatccatggg acgaattcac ggctaataaa agcttactgc ac 52

20 <210> 19
 <211> 52
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> oligonucleótido sintético.

<400> 19
 agctgtgcag taagcttca ttagccgtga atctgtcca tggatcctcc tc 52

30 <210> 20
 <211> 72
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> oligonucleótido sintético.

40 <400> 20
 tccgtacggc ccaggtgcta gcgcccagc ggcagcggct ggtggctacg gtccgggctc 60

tggccagcag gg 72

45 <210> 21
 <211> 72
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> oligonucleótido sintético.

50 <400> 21
 ctgctggcca gagcccggac cgtagccacc agccgctgcc gctgcggcgc tagcacctgg 60

gccgtacgga cc 72

55 <210> 22
 <211> 60
 <212> ADN
 <213> Secuencia artificial

ES 2 525 095 T3

<220>
 <223> oligonucleótido sintético.

<400> 22
 5 tccgggccag caggcccgg gtaacaggg tctggccag caaggccgg gccagcagg 60

<210> 23
 <211> 60
 <212> ADN
 10 <213> Secuencia artificial

<220>
 <223> oligonucleótido sintético.

<400> 23
 15 ctgctggccc ggacctgct ggccaggacc ctgtgaccc gggccctgct ggcccggacc 60

<210> 24
 <211> 105
 <212> ADN
 20 <213> Secuencia artificial

<220>
 <223> oligonucleótido sintético.

<400> 24
 25 ttctagcgcg gctgcagccg cggcagctgc gtccggcccg ggtggctacg gtccggaaaa 60
 ccagggtcca tctggcccgg gtggctacgg tcctggcggt ccggg 105

<210> 25
 <211> 105
 <212> ADN
 30 <213> Secuencia artificial

<220>
 <223> oligonucleótido sintético.

<400> 25
 35 cggaccgcca ggaccgtagc caccggggcc agatggaccc tggttttccg gaccgtagcc 60
 acccggggcg gacgcagctg ccggcgctgc agccgcgcta gaacc 105

<210> 26
 <211> 36
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Iniciador

<400> 26
 50 gaaaaccat ggtgcggt tctgcagctg tatctg 36

<210> 27
 <211> 45
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Iniciador

<400> 27
 gaaaagaagc ttcattagc cagcaagggc ttgagctaca gattg 45

 5 <210> 28
 <211> 34
 <212> ADN
 <213> Secuencia artificial

 10 <220>
 <223> Iniciador

 <400> 28
 gaaaaacat gggagcatat ggcccatctc cttc 34

 15 <210> 29
 <211> 45
 <212> ADN
 <213> Secuencia artificial

 20 <220>
 <223> Iniciador

 <400> 29
 gaaaagaagc ttcattagc ctgaaagagc ttggctaadc atttg 45

 25 <210> 30
 <211> 12
 <212> PRT
 <213> Secuencia artificial

 30 <220>
 <223> etiqueta T7

 <400> 30
 35 Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg

 1 5 10

 <210> 31
 40 <211> 216
 <212> PRT
 <213> Secuencia artificial

 <220>
 45 <223> FlagN-NR
 <400> 31

 50

 55

 60

ES 2 525 095 T3

5 Gly Glu Ser Ser Gly Gly Asp Val Gln Arg Lys Thr Asn Val Ile Ser
1 5 10 15

10 Asn Ala Leu Arg Asn Ala Leu Met Ser Thr Thr Gly Ser Pro Asn Glu
20 25 30

15 Glu Phe Val His Glu Val Gln Asp Leu Ile Gln Met Leu Ser Gln Glu
35 40 45

20 Gln Ile Asn Glu Val Asp Thr Ser Gly Pro Gly Gln Tyr Tyr Arg Ser
50 55 60

25 Ser Ser Ser Gly Gly Gly Gly Gly Gln Gly Gly Pro Val Val Thr
65 70 75 80

30 Glu Thr Leu Thr Val Thr Val Gly Gly Ser Gly Gly Gly Gln Pro Ser
85 90 95

35 Gly Ala Gly Pro Ser Gly Thr Gly Gly Tyr Ala Pro Thr Gly Tyr Ala
100 105 110

40 Pro Ser Gly Ser Gly Ala Gly Gly Val Arg Pro Ser Ala Ser Gly Pro
115 120 125

45 Ser Gly Ser Gly Pro Ser Gly Gly Ser Arg Pro Ser Ser Ser Gly Pro
130 135 140

50 Ser Gly Thr Arg Pro Ser Pro Asn Gly Ala Ser Gly Ser Ser Pro Gly
145 150 155 160

55 Gly Ile Ala Pro Gly Gly Ser Asn Ser Gly Gly Ala Gly Val Ser Gly
165 170 175

60

ES 2 525 095 T3

	Ala Thr Gly Gly Pro Ala Ser Ser Gly Ser Tyr Gly Pro Gly Ser Thr	
	180	185 190
5	Gly Gly Thr Tyr Gly Pro Ser Gly Gly Ser Glu Pro Phe Gly Pro Gly	
	195	200 205
10	Val Ala Gly Gly Pro Tyr Ser Pro	
	210	215
15	<210> 32 <211> 648 <212> ADN <213> Secuencia artificial	
20	<220> <223> FlagN-NR	
25	<400> 32	
25	ggcgaaagca gcgaggcga tgtgcaacgc aaaaccaacg ttatctccaa cgctttgaga	60
	aacgccttga tgtctacaac aggcagccca aacgaagagt tcgtccatga agttcaagac	120
30	ctcatccaga tgttatctca agaacagatc aacgaggtag atacttcagg accagggcag	180
	tactacaggt cgtcttcttc cggtgaggga ggtggaggac aaggaggtcc tgtagttact	240
35	gaaacactga ccgttacagt tggcggatcc ggtggagggc aaccttcagg tgcaggtcct	300
	agtggtagac gtggatatgc accaactgga tacgccccaa gcggctcagg tgcaggtggc	360
40	gttcgacctg gtgcctccgg tccaagtggc agtggacctg gtggtaggac tcgtcctagt	420
	agtagtggac ctagtggaac tcgtcccagc cctaattggtg caagtggatc tagccctggc	480
45	ggtagtcgac ctggtaggac caattctggt ggtgctggag tatccggcgc aactggagga	540
	cctgcatcca gcggctccta cggaccagga agtacaggtg gaacatatgg acctagtgga	600
	ggaagtgaac ctttcggacc aggagtggct ggaggaccat acagccca	648
50	<210> 33 <211> 93 <212> PRT <213> Secuencia artificial	
55	<220> <223> FlagC-NR <400> 33	
60		

ES 2 525 095 T3

Gly Ala Tyr Tyr Pro Ser Ser Arg Val Pro Asp Met Val Asn Gly Ile
 1 5 10 15
 5 Met Ser Ala Met Gln Gly Ser Gly Phe Asn Tyr Gln Met Phe Gly Asn
 20 25 30
 10 Met Leu Ser Gln Tyr Ser Ser Gly Ser Gly Thr Cys Asn Pro Asn Asn
 35 40 45
 15 Val Asn Val Leu Met Asp Ala Leu Leu Ala Ala Leu His Cys Leu Ser
 50 55 60
 20 Asn His Gly Ser Ser Ser Phe Ala Pro Ser Pro Thr Pro Ala Ala Met
 65 70 75 80
 25 Ser Ala Tyr Ser Asn Ser Val Gly Arg Met Phe Ala Tyr
 85 90

30 <210> 34
 <211> 279
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> FlagC-NR
 <400> 34

40 ggtgcttatt atcctagetc gcgtgttccc gatatggtga atgggtataat gagtgctatg 60
 caaggatctg gttttaacta ccaaatgttt ggtaatatgc taccacaata ttgctctggt 120
 tcaggaacat gcaatccaaa taatgttaat gttttgatgg atgctttggt agctgctttg 180
 45 cactgtctaa gtaaccacgg atcatcatct tttgcacctt ctccaactcc ggctgctatg 240
 agtgcgtatt ctaattctgt aggaagaatg ttcgcttat 279

50
 <210> 35
 <211> 27
 <212> PRT
 55 <213> Secuencia artificial
 <220>
 <223> Módulo K
 <400> 35

60

ES 2 525 095 T3

Gly Pro Gly Gly Ala Gly Gly Pro Tyr Gly Pro Gly Gly Ala Gly Gly
 1 5 10 15

5 Pro Tyr Gly Pro Gly Gly Ala Gly Gly Pro Tyr
 20 25

10 <210> 36
 <211> 81
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Módulo K

<400> 36
 ggtccgggcg gtgctggcgg tccgtacggc cctggggcgg caggtggggc atatggtccg 60
 ggcgggtgcgg gcggtcgta c 81

20 <210> 37
 <211> 28
 <212> PRT
 <213> Secuencia artificial

25 <220>
 <223> Módulo sp

<400> 37

30 Gly Gly Thr Thr Ile Ile Glu Asp Leu Asp Ile Thr Ile Asp Gly Ala
 1 5 10 15

35 Asp Gly Pro Ile Thr Ile Ser Glu Glu Leu Thr Ile
 20 25

40 <210> 38
 <211> 84
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Módulo sp

<400> 38
 ggtggcacca ccatcattga agatctggac atcactattg atggtgcgga cggcccgatc 60

45 acgatctctg aagagctgac catc 84

50 <210> 39
 <211> 18
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Módulo X
 <400> 39

55

ES 2 525 095 T3

Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ser Gly Gly Ala Gly
1 5 10 15

5 Gly Ser

10 <210> 40
<211> 54
<212> ADN
<213> Secuencia artificial

15 <220>
<223> Módulo X

<400> 40
ggtaggcctg gtggcggcgg tggcgcaggt ggctctggcg gtgcgggagg ttcc 54

20 <210> 41
<211> 30
<212> PRT
<213> Secuencia artificial

25 <220>
<223> Módulo Y

<400> 41

30 Gly Pro Gly Gly Ala Gly Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly
1 5 10 15

Pro Gly Gly Tyr Gly Pro Gly Gly Ser Gly Pro Gly Gly Tyr
20 25 30

35

<210> 42
<211> 90
<212> ADN
<213> Secuencia artificial

45 <220>
<223> Módulo Y

<400> 42

ggtagggggc gtgcggggccc aggtggctat ggtagggggc gttctggggc gggtaggctac 60

ggtagcggcg gttcggggccc gggtaggctac 90

50 <210> 43
<211> 32
<212> ADN
<213> Secuencia artificial

55 <220>
<223> Iniciador de PCR

<400> 43
gaaaaacat gggcgaaagc agcggaggcg at 32

<210> 44
 <211> 36
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <223> Iniciador de PCR
 <400> 44
 10 gaaaagaagc ttcattagc ctgggctgta tggcc 36
 <210> 45
 <211> 33
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Iniciador de PCR
 20 <400> 45
 gaaaacatc ggggcttat tatcctagc cgc 33
 <210> 46
 <211> 42
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Iniciador de PCR
 30 <400> 46
 gaaaagaagc ttcattagc cataagcgaa cattctctc ac 42
 <210> 47
 <211> 90
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 40 <223> oligonucleótido sintético.
 <400> 47
 tccgggcggt gcgggccag gtggctatgg tccgggcggt tctgggccgg gtggctacgg 60
 tctggcggt tccggcccg gtggctacgg 90
 45 <210> 48
 <211> 90
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> oligonucleótido sintético.
 <400> 48
 gtagccaccc gggccggaac cgccaggacc gtagccaccc ggcccagaac cgcccggacc 60
 atagccacct gggcccgcac cgcccggacc 90
 55 <210> 49
 <211> 84

ES 2 525 095 T3

<212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> oligonucleótido sintético.
 <400> 49
 tggcaccacc atcattgaag atctggacat cactattgat ggtgcgagc gcccgatcac 60
 gatctctgaa gagctgacca tcgg 84
 10 <210> 50
 <211> 84
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> oligonucleótido sintético.
 <400> 50
 20 gatggtcagc tcttcagaga tcgtgatcgg gccgtccgca ccatcaatag tgatgtccag 60
 atcttcaatg atggtggtgc cacc 84
 <210> 51
 <211> 81
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> oligonucleótido sintético.
 30 <400> 51
 tccgggaggc gctggcggtc cgtaccggccc tgggtggcgca ggtgggcat atggtccggg 60
 cgggtgcgggc ggtccgtacg g 81
 <210> 52
 <211> 81
 35 <212> ADN
 <213> Secuencia artificial
 <220>
 40 <223> oligonucleótido sintético.
 <400> 52
 gtaccggaccg cccgcaccgc ccggaccata tggcccacct gcgccaccag gcccgtaccg 60
 accgccagca ccgccggac c 81
 45 <210> 53
 <211> 54
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> oligonucleótido sintético.
 <400> 53
 tggcgtggt ggcgccggtg ggcaggtg ctctggcgtt gcggcggtt ccgg 54

ES 2 525 095 T3

<210> 54
 <211> 54
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <223> oligonucleótido sintético.
 <400> 54
 10 ggaaccgccc gcaccgccag agccacctgc gccaccggcg ccaccagcgc cacc 54
 <210> 55
 <211> 3238
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Vector de clonación pAZL
 20 <400> 55
 tgtcgagaag tactagagga tcataatcag ccataccaca tttgtagagg ttttacttgc 60
 tttaaaaaac ctcccacacc tccccctgaa cctgaaacat aaaatgaatg caattgttgt 120
 25 tgттаacttg tttattgcag cttataatgg ttacaaaataa agcaatagca tcacaaattt 180
 caaaaataaa gcattttttt cactgcattc tagttgtggg ttgtccaaac tcatcaatgt 240
 30 atcttatcat gtctggatct gatcactgct tgagcctagg agatccgaac cagataagtg 300
 aaatctagtt ccaaaactatt ttgtcatttt taattttcgt attagcttac gacgctacac 360
 35 ccagttocca tctattttgt cactcttccc taaataatcc ttaaaaactc catttccacc 420
 cctcccagtt cccaactatt ttgtccgccc acagcggggc atttttcttc ctgttatggt 480
 tttaatcaaa catcctgcc aactccatgtg acaaaccgtc atcttcggct actttttctc 540
 40 tgtcacagaa tgaaaatfff tctgtcatct cttcgttatt aatgtttgta attgactgaa 600
 tatcaacgct tatttgcagc ctgaatggcg aatgggacgc gccctgtagc ggcgattaa 660
 45 gcgcggcggg tgtggtggtt acgcgcageg tgaccgctac acttgccagc gccctagcgc 720
 ccgctccttt cgctttcttc ccttcctttc tcgccacggt cgccggcttt ccccgtaag 780
 ctctaaatcg ggggctccct ttagggttcc gatttagtgc tttacggcac ctcgacccca 840
 50 aaaaacttga ttaggggtgat gggtcacgta gtgggccatc gccctgatag acggtttttc 900
 55
 60

ES 2 525 095 T3

gccctttgac gttggagtcc acgttcttta atagtggact ctgttccaa actggaacaa 960
 cactcaaccc tatctcggtc tattcttttg atttataagg gattttgccg atttcggcct 1020
 5 attggttaaa aaatgagctg atttaacaaa aatttaacgc gaattttaac aaaatattaa 1080
 egtttacaat ttcaggtggc acttttcggg gaaatgtgcg cggaaccctt atttgtttat 1140
 ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga taaatgcttc 1200
 10 aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgc cttattccct 1260
 tttttgcggc attttgctt cctgtttttg ctcaaccaga aacgctggtg aaagtaaaag 1320
 atgtgaaga tcagttgggt gcacgagtg gttacatcga actggatctc aacagcggta 1380
 15 agatccttga gagttttcgc ccgaagaac gttttccaat gatgagcact tttaaagttc 1440
 tgetatgtgg cgcggtatta tcccgattg acgcgggca agagcaactc ggtcgcgca 1500
 20 tacactattc tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg 1560
 atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat aacactgogg 1620
 ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca 1680
 25 tgggggatca tgtaactcgc ctgatcgtt gggaaaccgga gctgaatgaa gccataccaa 1740
 acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcgc aaactattaa 1800
 ctggcgaact acttactcta gcttccggc aacaattaat agactggatg gaggcggata 1860
 30 aagtgcgag accactctc cgtcggccc ttccggctgg ctggtttatt gctgataaat 1920
 ctggagccgg tgagcgtggg tctcgggta tcattgcagc actggggcca gatggaagc 1980
 35 cctcccgtat cgtagtatc tacacgacgg ggagtcaggc aactatggat gaacgaaata 2040
 gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca gaccaagttt 2100
 actcatacat actttagatt gatttaaaac ttcattttta atttaaaag atctaggtga 2160
 40 agatcctttt tgataatctc atgacaaaaa tccctaacg tgagttttcg ttccactgag 2220
 cgtcagacc cgtagaaaag atcaaaggat cttcttgaga tcttttttt ctgcgcgtaa 2280
 45 tctgctgctt gcaaaacaaa aaaccaccgc taccagcggg ggtttgtttg ccggatcaag 2340
 agctaccaac tctttttccg aaggttaactg gcttcagcag agcgcagata ccaaatactg 2400
 tcttctagt gtagecgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 2460
 50 acctcgctct gctaactctg ttaccagtggt ctgctgccag tggcgataag tcgtgtctta 2520
 ccgggttga ctcaagacga tagttaccgg ataaggcgcg gcggtcgggc tgaacggggg 2580
 gtctgtcac acagcccagc ttggagcga cgaacctacac cgaactgaga tacctacagc 2640
 55 gtgagcattg agaaagccc acgcttccc aagggagaaa ggcggacagg tatccggtaa 2700
 gcggcagggt cggaacagga gagcgcacga gggagcttc aggggaaac gcctggtatc 2760
 60 tttatagtcc tgtcgggtt cgcacctct gacttgagcg tcgatttttg tgatgctcgt 2820

ES 2 525 095 T3

5
10
15
20
25
30
35
40
45
50
55

cagggggg	cg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	ttcctggcct	2880
tttgctggcc	ttttgctcac	atgttctttc	ctgcggttatc	ccctgattct	gtggataacc	2940	
gtattaccgc	ctttgagtga	gctgataccg	ctcgccgcag	ccgaacgacc	gagcgcagcg	3000	
agtcagtgag	cgaggaagcg	gaagagcgcc	tgatgcggta	ttttctcctt	acgcatctgt	3060	
gcggtatttc	acaccgcaga	ccagccgcgt	aacctggcaa	aatcggttac	ggttgagtaa	3120	
taaatggatg	ccctgcgtaa	gcgggtgtgg	gcggacaata	aagtcttaaa	ctgaacaaaa	3180	
tagatcgagg	aggatccatg	ggacgaattc	acggctaata	aaagcttact	gcacagct	3238	

Reivindicaciones

- 5
1. Una proteína recombinante de seda de araña que comprende una o más secuencias de proteínas de seda de araña repetitivas sintéticas,
- 10
- (a) en donde la secuencia repetitiva sintética comprende entre 5 a 50 unidades de repetición, y en donde una unidad de repetición consiste de la secuencia de aminoácidos de la sec. con núm. de ident.: 5 o variantes de esta, o
- (b) en donde la secuencia repetitiva sintética comprende entre 10 a 50 unidades de repetición, y en donde una unidad de repetición consiste de la secuencia de aminoácidos de la sec. con núm. de ident.: 3 o variantes de esta y se combina con una unidad de repetición que consiste en la secuencia de aminoácidos de la sec. con núm. de ident.: 4 o variantes de esta,
- 15
- en donde las variantes en cada caso tienen una sustitución de aminoácido, deleciones de aminoácido y/o inserciones de aminoácido, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácido consisten en entre 1 a 2 deleciones de aminoácido y/o las inserciones de aminoácido consisten en entre 1 a 2 inserciones de aminoácido.
- 20
2. La proteína recombinante de seda de araña de la reivindicación 1, en donde la proteína recombinante de seda de araña comprende adicionalmente una o más secuencias de proteína de seda de araña repetitivas no auténticas, y en donde
- 25
- (i) la secuencia de ácido nucleico que codifica la secuencia no repetitiva auténtica es la sec. con núm. de ident.: 14 o una variante de esta, que codifica la secuencia de aminoácidos de sec. con núm. de ident.: 10 con una sustitución de aminoácido, deleciones de aminoácidos y/o inserciones de aminoácidos, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácidos consisten en entre 1 a 2 deleciones de aminoácidos y/o las inserciones de aminoácidos consisten en entre 1 a 2 inserciones de aminoácidos,
- 30
- (ii) la secuencia de ácido nucleico que codifica la secuencia no repetitiva auténtica es la sec. con núm. de ident.: 15 o una variante de esta, que codifica la secuencia de aminoácidos de sec. con núm. de ident.: 11 con una sustitución de aminoácido, deleciones de aminoácidos y/o inserciones de aminoácidos, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácidos consisten en entre 1 a 2 deleciones de aminoácidos y/o las inserciones de aminoácidos consisten en entre 1 a 2 inserciones de aminoácidos,
- 35
- (iii) la secuencia de ácido nucleico que codifica la secuencia no repetitiva auténtica es la sec. con núm. de ident.: 32 o una variante de esta, que codifica la secuencia de aminoácidos de sec. con núm. de ident.: 31 con una sustitución de aminoácido, deleciones de aminoácidos y/o inserciones de aminoácidos, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácidos consisten en entre 1 a 2 deleciones de aminoácidos y/o las inserciones de aminoácidos consisten en entre 1 a 2 inserciones de aminoácidos, o
- 40
- (iv) la secuencia de ácido nucleico que codifica la secuencia no repetitiva auténtica es la sec. con núm. de ident.: 34 o una variante de esta, que codifica la secuencia de aminoácidos de sec. con núm. de ident.: 33 con una sustitución de aminoácido, deleciones de aminoácidos y/o inserciones de aminoácidos, en donde la sustitución de aminoácidos consiste en 1 sustitución de aminoácido, las deleciones de aminoácidos consisten en entre 1 a 2 deleciones de aminoácidos y/o las inserciones de aminoácidos consisten en entre 1 a 2 inserciones de aminoácidos.
- 45
3. La proteína recombinante de seda de araña de una o más de las reivindicaciones 1 o 2, en donde la secuencia repetitiva sintética es (AQ)₁₂ o (AQ)₂₄, y en donde A representa la secuencia de aminoácidos de la sec. con núm. de ident.: 3 y Q representa la secuencia de aminoácidos de la sec. con núm. de ident.: 4.
- 50
4. La proteína recombinante de seda de araña de una o más de las reivindicaciones 1 o 2, en donde la secuencia repetitiva sintética es C₁₆ o C₃₂, y en donde C representa la secuencia de aminoácidos de la sec. con núm. de ident.: 5.
- 55
5. La proteína recombinante de seda de araña de una o más de las reivindicaciones 1 a 4, en donde las proteínas recombinantes de la seda de araña completas comprenden la fórmula (AQ)₁₂NR₃, (AQ)₂₄NR₃, C₁₆NR₄, C₃₂NR₄, (AQ)₁₂, (AQ)₂₄, C₁₆ o C₃₂, y en donde A representa la secuencia de aminoácidos de la sec. con núm. de ident.: 3, Q representa la secuencia de aminoácidos de la sec. con núm. de ident.: 4, C representa la secuencia de aminoácidos de la sec. con núm. de ident.: 5, NR₃ representa la secuencia de aminoácidos de la sec. con núm. de ident.: 10 y NR₄ representa la secuencia de aminoácidos de la sec. con núm. de ident.: 11.
- 60

- 5
6. Una secuencia de ácido nucleico, que codifica una proteína recombinante de seda de araña de una o más de las reivindicaciones 1 a 5.
7. Un vector, que comprende la secuencia de ácido nucleico de la reivindicación 6.
8. El vector de la reivindicación 7, que es un vector de clonación que tiene una secuencia de ácido nucleico de acuerdo con la sec. con núm. de ident.: 55.
- 10
9. El vector de la reivindicación 7, que comprende además una o más secuencias reguladoras, en donde el vector es un vector de expresión..
10. El vector de las reivindicaciones 7 a 9, que es un plásmido o un vector viral.
- 15
11. El vector de la reivindicación 10, en donde el vector es un sistema de baculovirus o un sistema de vector del virus vaccinia.
12. Un huésped no humano, que ha sido transformado con el vector de cualquiera de las reivindicaciones 7 a 11.
- 20
13. El huésped de la reivindicación 12, que es una célula procariota preferentemente *E. coli* o *Bacillus subtilis*.
14. El huésped de la reivindicación 12, que es una célula eucariota, preferentemente una célula de mamífero, célula vegetal, célula de levadura o una célula de insecto.
- 25
15. El huésped de la reivindicación 14, en donde la célula de mamífero es una célula CHO, COS, HeLa, 293T, HEH o BHK, o en donde la célula de levadura se selecciona de *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Candida albicans*, y *Hansenula polymorpha*.
- 30
16. El huésped de la reivindicación 14, en donde la célula de insecto es seleccionada de células de insecto Lepidoptera, preferentemente de *Spodoptera frugiperda* y de *Trichoplusia ni*, preferentemente es una célula Sf9, Sf21 o high-5.
- 35
17. El huésped de la reivindicación 14, en donde la célula vegetal se deriva de tabaco, papa, maíz, chícharo y tomate.
18. Un método de agregación de proteínas de seda de araña, que comprende las siguientes etapas:
- 40
- a) preparar una solución de proteína que contiene proteínas de seda de araña no orientadas como se define en una o más de las reivindicaciones 1 a 5;
- b) exponer la solución preparada en a) a un inductor de de agregación; y
- c) recuperar las proteínas de seda de araña precipitadas.
- 45
19. El método de la reivindicación 18, en donde las proteínas de seda de araña usadas en la etapa a) se producen por transformación de un huésped adecuado de una o más de las reivindicaciones 12 a 17 con un vector de la reivindicación 7 a 11 o un ácido nucleico de la reivindicación 6, y expresar la proteína de seda de araña bajo condiciones adecuadas.
- 50
20. El método de las reivindicaciones 18 o 19, en donde el inductor de de agregación es seleccionado de acidificación, preferentemente a un pH de aproximadamente 1, fosfato de potasio y tensión mecánica, preferentemente al rotar la solución de proteína y aplicar fuerzas de cizalla y/o que comprende además la etapa de hilar dichas proteínas proporcionadas en la etapa a) o recuperadas en la etapa c) en filamentos, nanofibras e hilos por un método adecuado o formar una película.
- 55
21. Uso de las proteínas/hilos tal como se define en una o más de las reivindicaciones precedentes en el campo de la biotecnología o la medicina para la fabricación de sistemas de cierre o de cobertura de la herida, preferentemente para la fabricación de materiales de sutura, que están destinados con mayor preferencia para usar en neurocirugía o cirugía oftálmica.
22. Uso de las proteínas/hilos tal como se define en una o más de las reivindicaciones precedentes para la fabricación de materiales de materiales de reemplazo, preferentemente materiales de cartílago o tendón artificiales.

23. Uso de las proteínas/hilos tal como se define en una o más de las reivindicaciones 1 a 5 en la fabricación de piezas de automóviles y aviones.
- 5 24. Sistemas de cierre o de cobertura de heridas, materiales de sutura, materiales de reemplazo, preferentemente cartilago artificial, materiales de tendón, partes o piezas de automóviles usados en la construcción de aviones, que comprenden una proteína de una o más de las reivindicaciones 1 a 5 o que se obtienen por un método de una o más de las reivindicaciones 18-20.
- 10 25. Un producto de papel, o un producto de textil o cuero, que comprende una proteína recombinante de seda de araña de una o más de las reivindicaciones 1 a 5.
26. El producto de textil o cuero de la reivindicación 25, en donde la proteína recombinante de seda de araña está presente como un recubrimiento.
- 15 27. Un gel o una espuma que comprende o que consiste en una proteína de una o más de las reivindicaciones 1 a 5, en donde el gel comprende o consiste preferentemente en proteínas basadas en (AQ)₁₂NR3, (AQ)₂₄NR3, C₁₆NR4, C₃₂NR4, (AQ)₁₂, (AQ)₂₄, C₁₆ o C₃₂.
- 20 28. Recubrimiento para los implantes y endoprótesis que comprenden o que consisten en una proteína de una o más de las reivindicaciones 1 a 5.
29. Una esfera, hilo o fibra, que comprende una proteína/hilo de una o más de las reivindicaciones 1 a 5 y una fibra adicional, la fibra no es de origen de araña y preferentemente es una fibra derivada de planta o una fibra sintética.
- 25 30. Una película que comprende o que consiste en una proteína de una o más de las reivindicaciones 1 a 5 o que se basa en (AQ)₁₂NR3, (AQ)₂₄NR3, C₁₆NR4, C₃₂NR4, (AQ)₁₂, (AQ)₂₄, C₁₆ o C₃₂ y que comprende o consiste en preferentemente proteínas basadas en (AQ)₂₄NR3 o C₁₆.
- 30 31. La película de la reivindicación 30, en donde la superficie de la película se modifica por las moléculas orgánicas pequeñas o macromoléculas biológicas, por ejemplo proteínas, fluoresceína o β-galactosidasa.

35

40

45

50

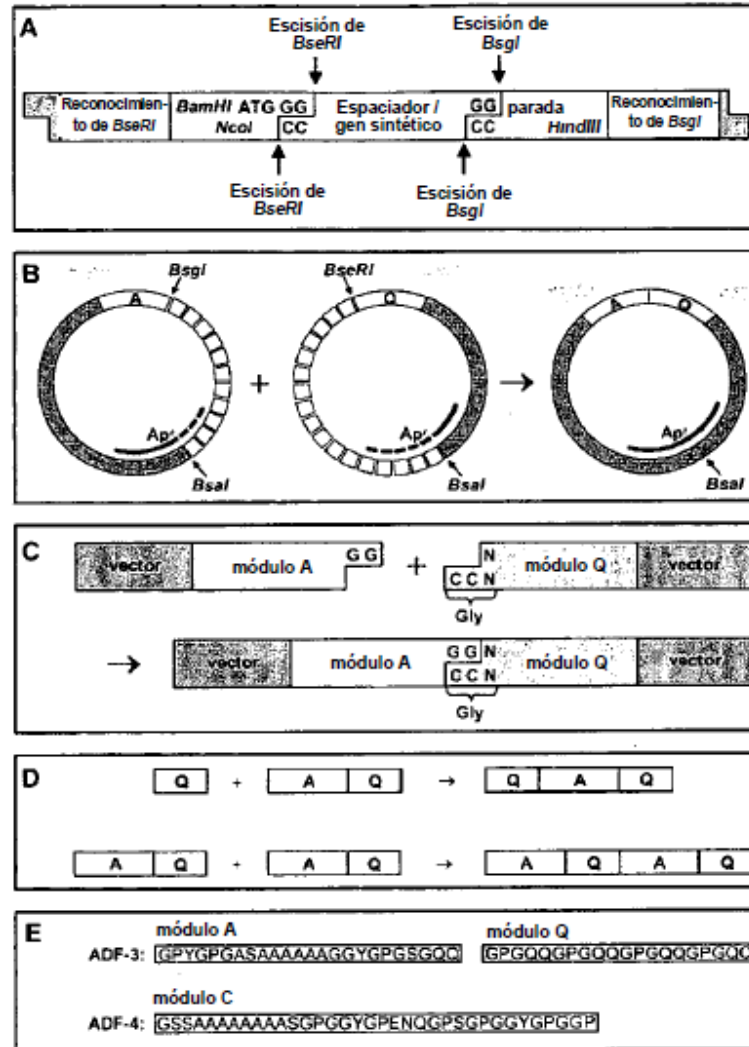


Figura 1

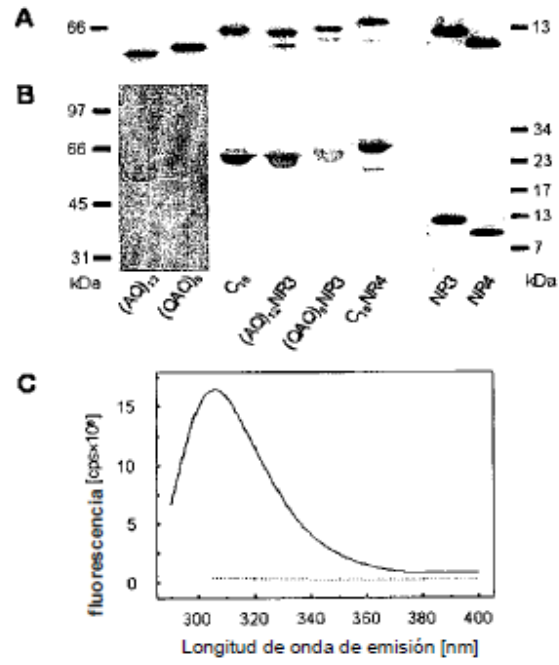


Figura 2

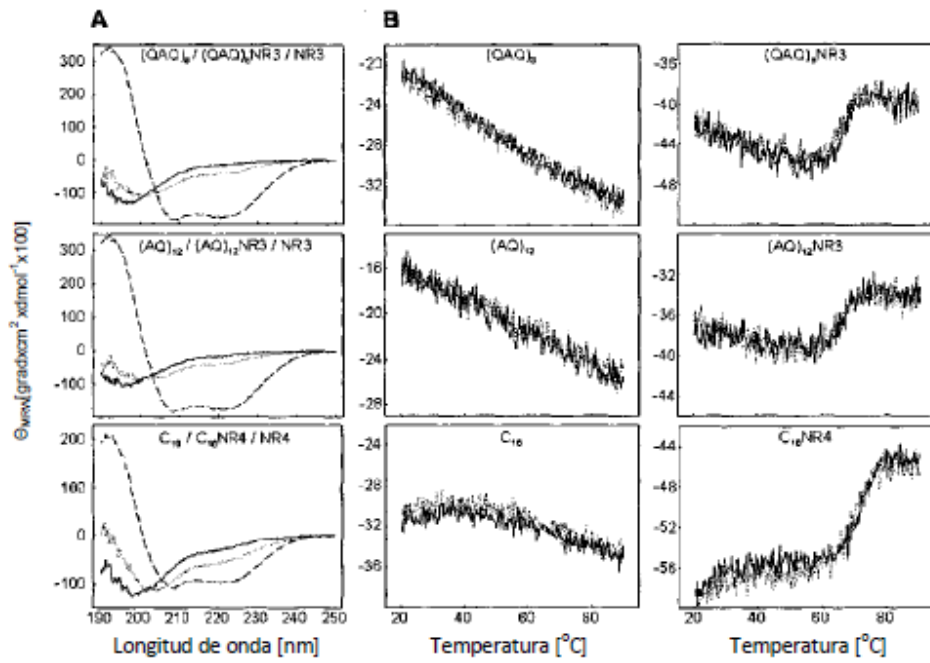


Figura 3

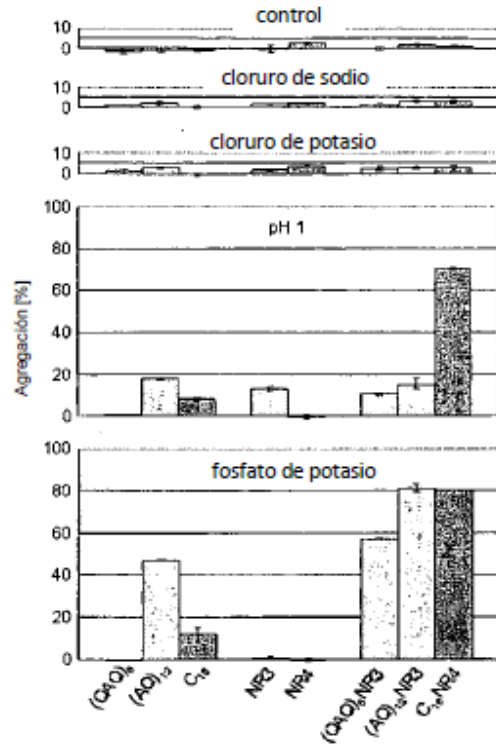


Figura 4

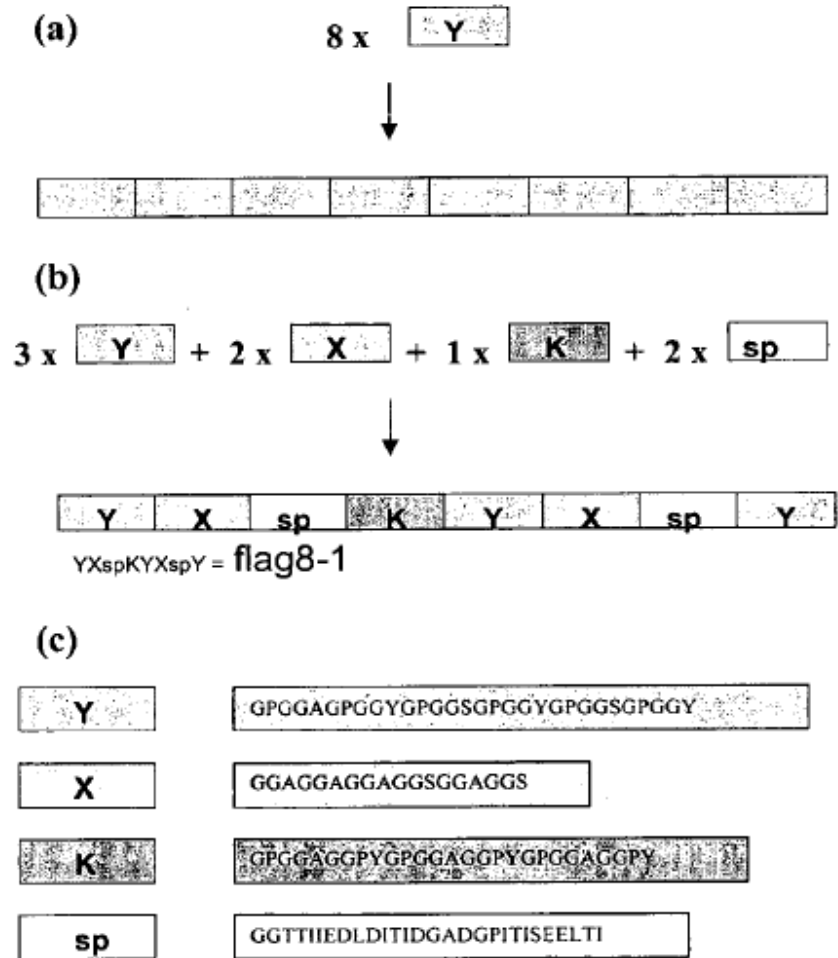


Figura 5

Vector pAZI

Fuente: pFastbacl (Invitrogen)
Elementos de pFastbacl: origen de replicación;
resistencia a la ampicilina
Elementos añadidos: casete de clonación

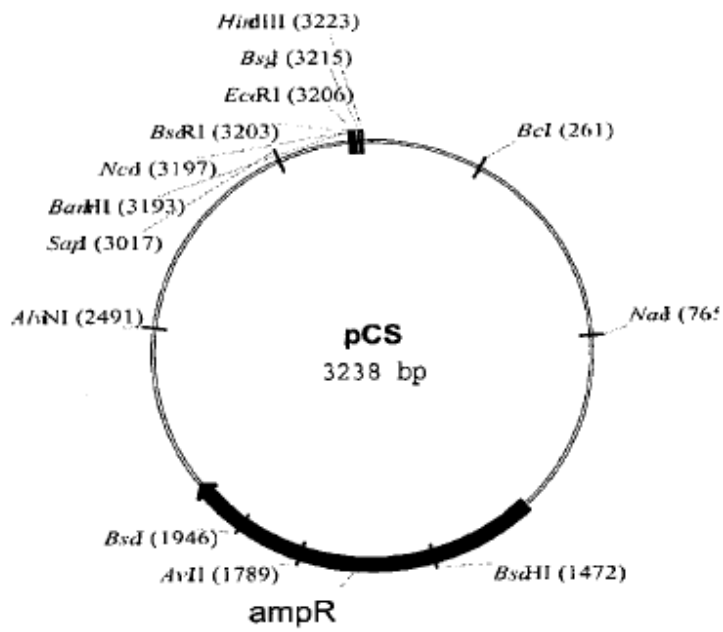


Figura 6

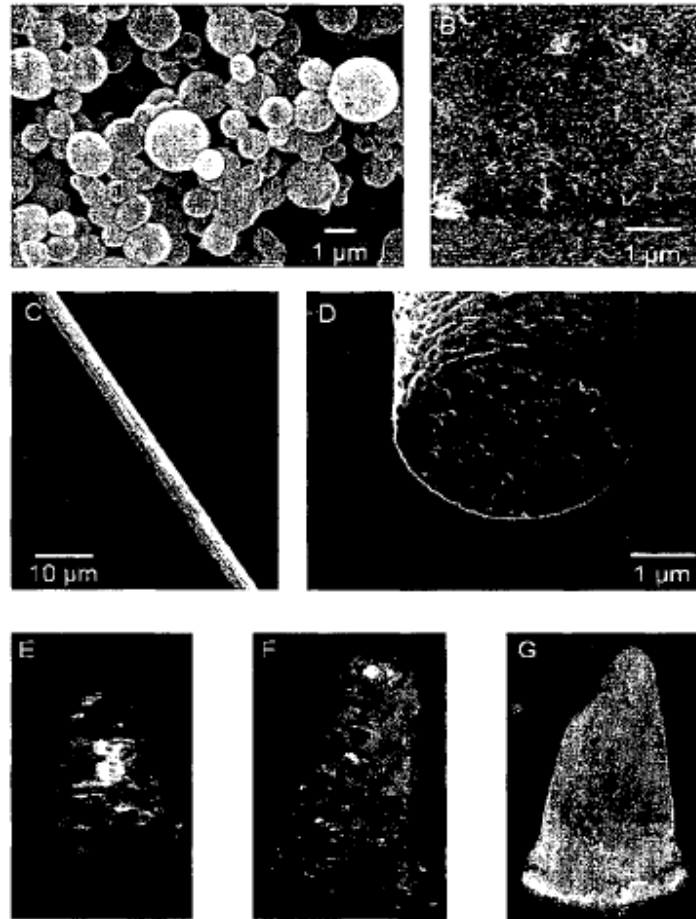


Figura 7: Formas de ensamblaje de proteínas de la seda de araña. (A) Esferas formadas por $C_{16}NR4$ visualizadas mediante microscopía electrónica de barrido (SEM). **(B)** Nanofibrillas formadas por $C_{16}NR4$ visualizadas mediante microscopía de fuerza atómica (información de altura). **(C, D)** Microfibrilla formada por $(AQ)_{24}NR3$ investigada mediante SEM **(C)**. Para el corte de la fibrilla y la posterior visualización de la sección transversal se usó un haz de iones Ga^+ enfocado **(D)**. **(E)** Espuma generada a partir de una solución de $(AQ)_{24}NR3$. **(F)** Espuma generada a partir de una solución de $C_{16}NR4$. **(G)** Gel reticulado formado por nanofibrillas de $C_{16}NR4$.

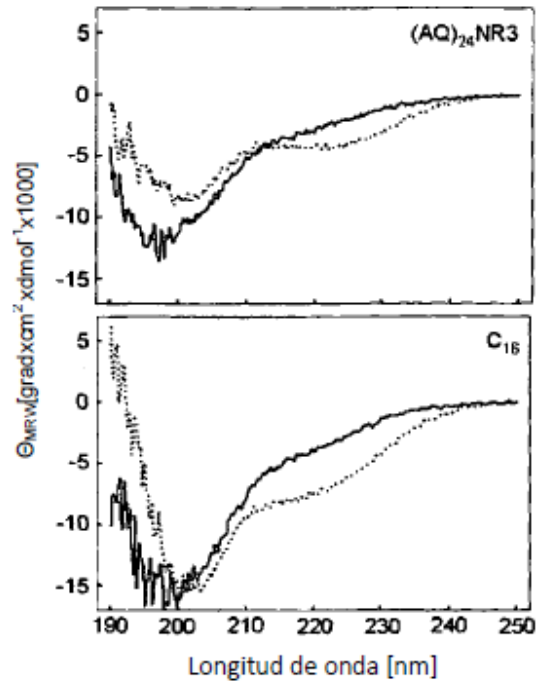


Figura 8

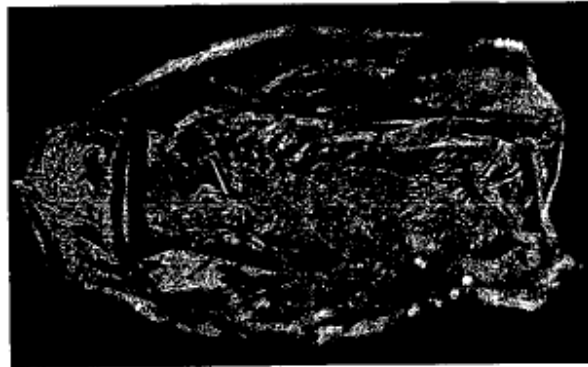


Figura 9

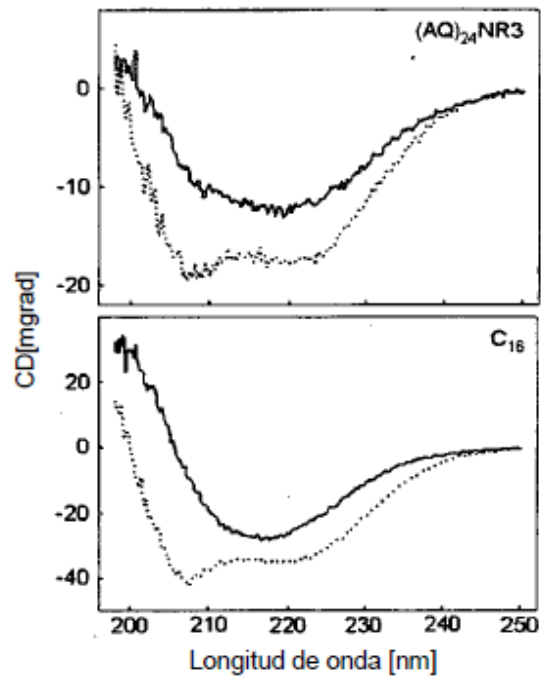


Figura 10

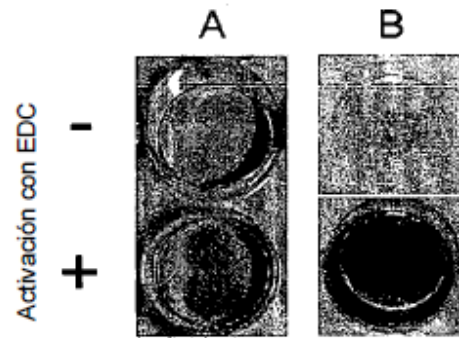


Figura 11

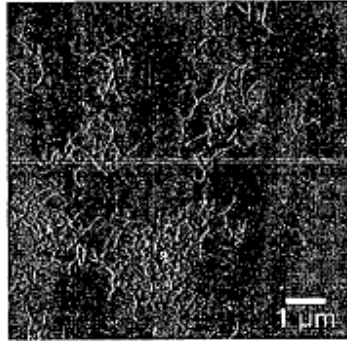


Figura 12



Figura 13

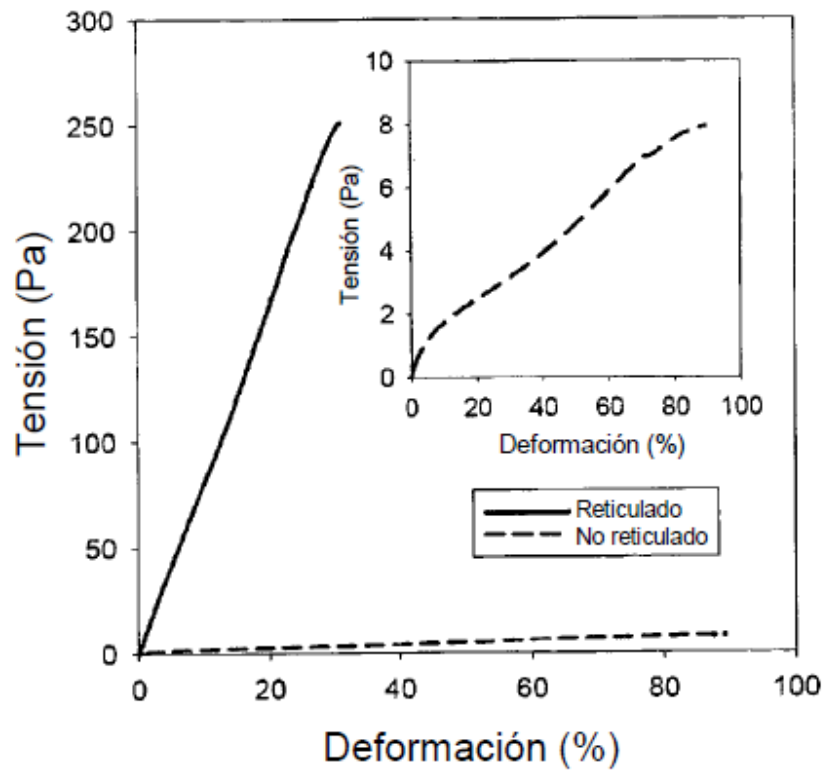


Figura 14

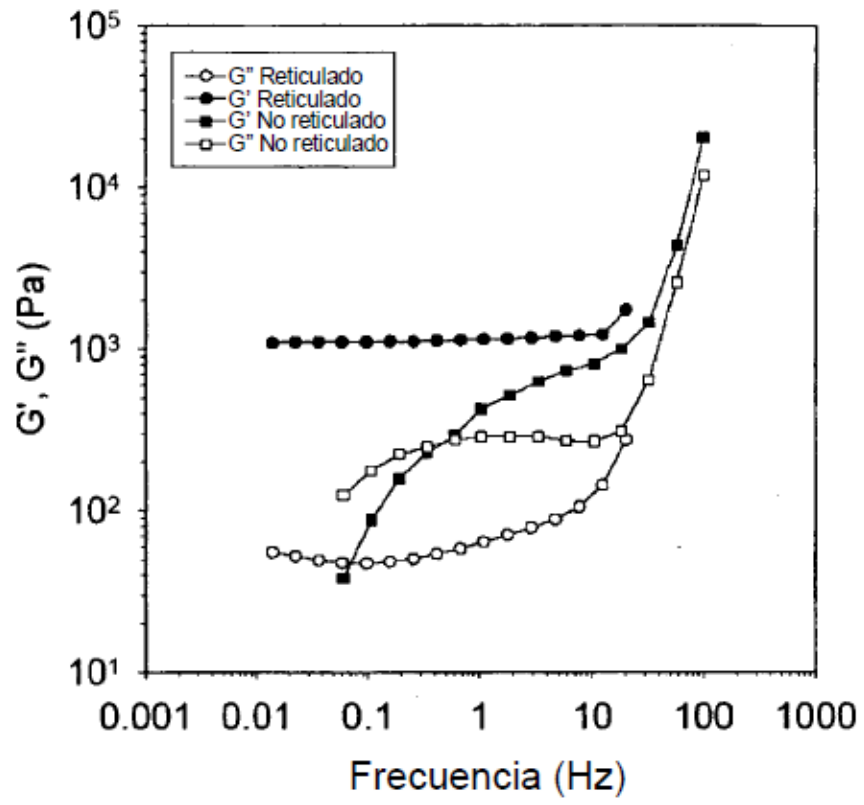


Figura 15

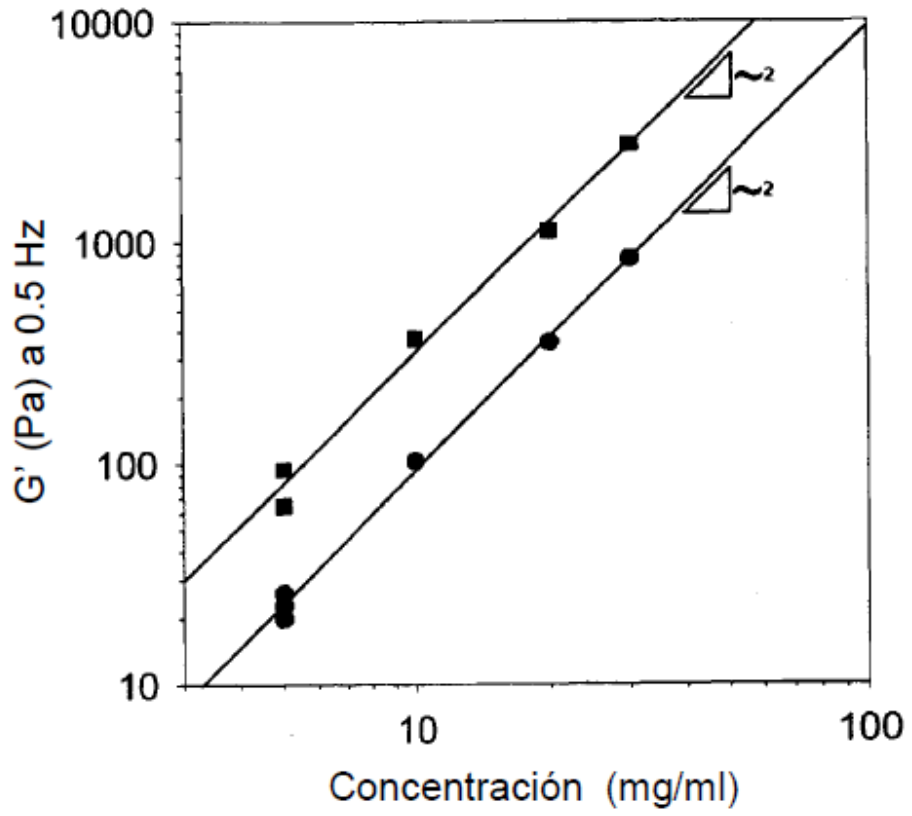


Figura 16