

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 608 673**

51 Int. Cl.:

C07K 14/435 (2006.01)

C12N 15/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.10.2010 PCT/SE2010/051163**

87 Fecha y número de publicación internacional: **22.09.2011 WO11115538**

96 Fecha de presentación y número de la solicitud europea: **27.10.2010 E 10848078 (1)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 2547771**

54 Título: **Producción de proteínas y polipéptidos**

30 Prioridad:

21.04.2010 WO PCT/SE2010/050439
18.03.2010 EP 10156927

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.04.2017

73 Titular/es:

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

72 Inventor/es:

JOHANSSON, JAN;
RISING, ANNA;
HEDHAMMAR, MY y
NORDLING, KERSTIN

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 608 673 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

Producción de proteínas y polipéptidos

Campo técnico de la invención

5 La invención se refiere al campo de la producción de proteínas y polipéptidos, y más específicamente a la producción de proteínas de tela de araña (espidroínas), y de otras proteínas y polipéptidos que no son espidroínas. La presente invención proporciona un método para producir una proteína/polipéptido no espidroína deseada. También se proporcionan nuevos intermedios de proteína de fusión para producción de las proteínas y polipéptidos deseados, así como moléculas de ácido polinucleico que codifican dichos intermedios.

Antecedentes de la invención

10 La producción de proteínas y polipéptidos a partir de ADN se puede realizar en diversos hospedantes, pero un problema común es la formación de agregados de proteína/polipéptido insolubles. Esto puede dificultar seriamente o incluso impedir la producción de una proteína/polipéptido funcional. Una solución a este problema es expresar la proteína o el polipéptido deseado como una proteína de fusión con una proteína o polipéptido que proporcione la solubilidad requerida. La proteína de fusión puede romperse, y la proteína deseada se puede aislar.

15 El problema se agrava habitualmente con proteínas y polipéptidos de baja solubilidad, p.ej., proteínas y polipéptidos asociados a membrana. Por ejemplo, la proteína C tensioactiva pulmonar (SP-C; Tabla 6) es una proteína transmembrana que es producida por células alveolares de tipo II y es un constituyente de tensioactivo, necesario para evitar el colapso alveolar al final de la espiración. Los neonatos a menudo sufren insuficiencia respiratoria debido a cantidades insuficientes de tensioactivo. Hoy día esta afección es tratada con preparaciones tensioactivas extraídas de pulmones de animales. La SP-C-33 es una variante de la SP-C, en la que se han intercambiado los
20 residuos de la hélice transmembrana (de forma habitual principalmente valinas) por leucinas. La SP-C-33 retiene la función de la SP-C nativa, incluyendo una inserción apropiada en las membranas, pero es menos propensa a agregarse y por tanto es viable para producir en grandes cantidades el desarrollo de una preparación tensioactiva sintética. Puesto que hasta la fecha la SP-C-33 no ha podido ser producida a partir de ADN, hoy día se fabrica
25 mediante síntesis química.

Otros ejemplos de proteínas y polipéptidos que presentan dificultades cuando son expresados a partir de ADN recombinante son A β -péptido, IAPP, PrP, α -sinucleína, calcitonina, prolactina, cistatina, ATF y actina; SP-B, α -defensinas y β -defensinas; apolipoproteínas de clase A-H; LL-37, SP-C, SP-C33Leu, Brichos, GFP, neuroserpina; hormonas, que incluyen EPO y hormona del crecimiento (HC), y factores de crecimiento, que incluyen IGF-I e IGF-II;
30 avidina y estreptavidina; y proteasa 3C.

Hedhammar M. *et al.*, *Biochemistry* 47: 3407-3417 (2008), discuten las propiedades estructurales de las partes repetitivas y no repetitivas recombinantes de la espidroína ampulada mayor 1 procedente de *Euprosthrops australis* y las implicaciones en la formación de fibra de tela de araña.

Sumario de la invención

35 Un objetivo de la presente invención es proporcionar nuevos medios y métodos para la producción de proteínas y polipéptidos, y en particular de proteínas y polipéptidos no espidroína.

También es un objetivo de la presente invención proporcionar nuevos medios y métodos para la producción de proteínas y polipéptidos, y en particular proteínas y polipéptidos no espidroína, con baja solubilidad en agua, p.ej.,
40 proteínas y polipéptidos que sean propensos a agregarse cuando se producen a partir de ADN recombinante, proteínas y polipéptidos de membrana, y proteínas y polipéptidos formadores de amiloide.

Otro objetivo adicional de la presente invención es proporcionar medios y métodos alternativos para la producción de fármacos de proteína o polipéptido y de dianas farmacológicas.

Un objetivo de la presente invención es proporcionar nuevos medios y métodos para la producción de proteínas y polipéptidos que contienen disulfuros.

45 También es un objetivo de la presente invención proporcionar nuevos medios y métodos para la producción de apolipoproteínas.

Para éstos y otros objetivos que serán evidentes a partir de la siguiente especificación, la presente invención proporciona según un primer aspecto una proteína de fusión que es útil en un método para producir una proteína o polipéptido no espidroína deseada. La proteína de fusión puede ser útil como tal, o puede dividirse para obtener la
50 proteína o polipéptido en forma aislada. Las proteínas de fusión según la invención comprenden (i) al menos un resto potenciador de la solubilidad que deriva del fragmento N-terminal (NT) de una proteína de tela de araña; y (ii) al menos un resto que es una proteína o polipéptido deseado no espidroína; en donde cada resto potenciador de la solubilidad está unido directa o indirectamente al resto de proteína o polipéptido deseado no espidroína.

Cada resto potenciador de la solubilidad tiene al menos un 80% de identidad con respecto a la SEQ ID NO 6 o al menos un 50% de identidad con respecto a la SEQ ID NO 8. En realizaciones específicas de la proteína de fusión, cada resto potenciador de la solubilidad contiene entre 100 y 160 residuos de aminoácido.

5 En una realización, la proteína de fusión está sometida a la condición de que cuando la proteína de fusión comprenda un único resto potenciador de la solubilidad que derive del fragmento N-terminal (NT) de una proteína de tela de araña, entonces la proteína o polipéptido deseado es una proteína o polipéptido no espidroína.

La proteína o polipéptido deseado es una proteína o polipéptido no espidroína. En algunas realizaciones, la proteína o polipéptido deseado tiene menos de un 30% de identidad con respecto a cualquiera de las SEQ ID NO: 6-10.

10 En determinadas realizaciones, la proteína de fusión comprende al menos dos restos potenciadores de la solubilidad, cada uno de los cuales deriva del fragmento N-terminal (NT) de una proteína de tela de araña. En realizaciones específicas, la proteína de fusión comprende al menos dos restos potenciadores de la solubilidad consecutivos, cada uno de los cuales deriva del fragmento N-terminal (NT) de una proteína de tela de araña.

15 En algunas realizaciones de la proteína de fusión, al menos un resto potenciador de la solubilidad está ligado directa o indirectamente al extremo amino-terminal o al extremo carboxi-terminal de al menos un resto de la proteína o polipéptido deseado. En realizaciones específicas, al menos un resto potenciador de la solubilidad constituye el extremo amino-terminal y/o el extremo carboxi-terminal de la proteína de fusión.

En una realización, la proteína de fusión comprende además (iii) al menos un sitio de división dispuesto entre al menos un resto de proteína o polipéptido deseado y al menos un resto potenciador de la solubilidad.

20 En determinadas realizaciones de la proteína de fusión, la proteína o polipéptido deseado deriva de esponjas, ctenóforos ("portadores de peines"), medusas, corales, anémonas, gusanos planos, rotíferos, gusanos redondos, gusanos de cinta, almejas, caracoles, pulpos, gusanos segmentados, crustáceos, insectos, briozoos, braquiópodos, foronídeos, estrellas de mar, erizos de mar, tunicados, lanceolados, vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial. En realizaciones específicas, la proteína o polipéptido deseado deriva de moluscos, insectos, vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial. En realizaciones específicas adicionales, la proteína o polipéptido deseado deriva de vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial.

30 La proteína o polipéptido deseado se selecciona del grupo que consiste en proteínas y polipéptidos formadores de amiloide, SP-B y variantes de los mismos que contienen disulfuro, apolipoproteínas, proteínas y polipéptidos de membrana, fármacos de proteína y polipéptido y dianas farmacológicas, proteínas y polipéptidos propensos a agregación, y proteasas. En realizaciones específicas de la proteína de fusión, la proteína y polipéptido deseados se seleccionan del grupo que consiste en A β -péptido, IAPP, PrP, α -sinucleína, calcitonina, prolactina, cistatina, ATF y actina; SP-B, α -defensinas y β -defensinas; apolipoproteínas de clase A-H; LL-37, SP-C, SP-C33, SP-C33Leu, Bricos, GFP, neuroserpina; hormonas, que incluyen EPO y hormona de crecimiento (HC), y factores de crecimiento, que incluyen IGF-I e IGF-II; avidina y estreptavidina; y proteasa 3C.

35 Las realizaciones preferidas de la proteína de fusión se seleccionan del grupo que consiste en las SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 y 47, y las proteínas que tienen al menos un 80%, preferiblemente al menos un 90%, más preferiblemente al menos un 95% de identidad con respecto a dichas proteínas.

40 Según otro aspecto, la presente invención proporciona ácidos polinucleicos aislados que codifican las proteínas de fusión según la invención. Las realizaciones preferidas de los ácidos polinucleicos aislados se seleccionan del grupo que consiste en ácidos nucleicos que codifican una proteína de fusión seleccionada del grupo que consiste en las SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 y 47, y las proteínas que tienen al menos un 80%, preferiblemente al menos un 90%, más preferiblemente al menos un 95% de identidad con respecto a dichas proteínas; y el grupo de ácidos nucleicos que consiste en las SEQ ID NOS 27, 29, 31, 38, 40, 43 y 48.

45 Según un aspecto, la presente invención proporciona un nuevo uso de al menos un resto que se deriva del fragmento N-terminal (NT) de una proteína de tela de araña como un resto potenciador de la solubilidad en una proteína de fusión para la producción de una proteína o polipéptido no espidroína deseado. En una realización, la proteína de fusión está sujeta a la condición de que cuando la proteína de fusión comprende un único resto potenciador de la solubilidad que se deriva del fragmento N-terminal (NT) de una proteína de tela de araña, entonces la proteína o el polipéptido deseado es una proteína o polipéptido no espidroína. En una realización específica, la proteína o polipéptido deseado es una proteína o polipéptido espidroína. Cada resto potenciador de la solubilidad tiene al menos un 80% de identidad con respecto a la SEQ ID NO 6, o al menos un 50% de identidad con respecto a la SEQ ID NO: 8.

55 Según otro aspecto, la presente invención proporciona un método para producir una proteína de fusión, que comprende las etapas de: a) expresar en un hospedante adecuado una proteína de fusión según la invención; y b) obtener una mezcla que contenga la proteína de fusión, y opcionalmente aislar la proteína de fusión.

Según un aspecto relacionado, la presente invención proporciona un método para producir una proteína o polipéptido no espidroína deseado, que comprende las siguientes etapas: a) expresar en un hospedante adecuado una proteína de fusión según la invención; y b) obtener una mezcla que contenga la proteína o polipéptido de fusión, y opcionalmente aislar la proteína o polipéptido de fusión. En determinadas realizaciones, este método además comprende las siguientes etapas: c) dividir la proteína de fusión para proporcionar la proteína o polipéptido no espidroína deseado; y d) aislar la proteína o polipéptido no espidroína deseado; en donde dicha proteína de fusión comprende: (iii) al menos un sitio de división dispuesto entre al menos un resto de proteína o polipéptido no espidroína deseado y al menos un resto potenciador de la solubilidad.

En determinadas realizaciones de estos métodos, la etapa b) implica además la purificación de la proteína de fusión en un medio de afinidad con un resto NT inmovilizado y/o en un medio de intercambio aniónico. En realizaciones especificadas, la purificación de la proteína de fusión en un medio de afinidad se lleva a cabo en asociación a un medio de afinidad con un resto NT inmovilizado a un pH de 6,3 o inferior, seguido de la disociación del medio de afinidad con un medio de disociación deseado. En realizaciones específicas adicionales, el medio de disociación tiene un pH de 6,4 o superior, un pH de 4,1 o inferior y/o tiene una elevada fuerza iónica. En algunas realizaciones, la purificación de la proteína de fusión en un medio de intercambio aniónico se lleva a cabo en asociación al medio de intercambio aniónico en un pH de 6,4 o superior, seguido de la disociación del medio de intercambio aniónico con un medio de disociación que tiene una elevada fuerza iónica. En algunas realizaciones de estos métodos, la purificación de la proteína de fusión de la etapa b) se produce en una columna, sobre partículas magnéticas con superficies funcionalizadas, o en rellenos con superficies funcionalizadas.

20 Lista de secuencias anexas

	SEQ ID NO
	1 4Rep
	2 4RepCT
	3 NT4Rep
25	4 NT5Rep
	5 NT4RepCTHis
	6 NT
	7 CT
	8 secuencia NT de consenso
30	9 secuencia CT de consenso
	10 secuencia repetitiva de <i>Euprosthonops australis</i> MaSp1
	11 secuencia de segmento G de consenso 1
	12 secuencia de segmento G de consenso 2
	13 secuencia de segmento G de consenso 3
35	14 NT4Rep (ADN)
	15 NT4RepCT (ADN)
	16 NT5Rep (ADN)
	17 NT4RepCTHis 2
	18 NT4RepCTHis 2 (ADN)
40	19 ZbasicNT4RepCT
	20 NT4RepCT
	21 HisTrxHisThrNT4RepCT
	22 NT4RepCT 2
	23 HisNTNT4RepCT

	24	HisNTNT4RepCT (ADN)
	25	NT8RepCT
	26	HisNTMetSP-C33Leu
	27	HisNTMetSP-C33Leu (ADN)
5	28	HisNTNTMetSP-C33Leu
	29	HisNTNTMetSP-C33Leu (ADN)
	30	HisNTNTMetLL37
	31	HisNTNTMetLL37 (ADN)
	32	NTHis
10	33	HisNTNT8RepCT
	34	HisNTNTBrichos
	35	HisTrxHisSP-C33Leu
	36	HisTrxHisSP-C33Leu (ADN)
	37	HisTrxNtSP-C33Leu
15	38	HisTrxNtSP-C33Leu (ADN)
	39	2HisNtNtQGBrichos
	40	2HisNtNtQGBrichos (ADN)
	41	Brichos
	42	2HisNtNtQGGFP
20	43	2HisNtNtQGGFP (ADN)
	44	GFP (Proteína Fluorescente Verde)
	45	ZbGFP
	46	HisABPGFP
	47	2HisNtNtQGNS
25	48	2HisNtNtQGNS (ADN)
	49	NS (Neuroserpina)

Breve descripción de las figuras

Figura 1: muestra un alineamiento de secuencia de los dominios N-terminales de la espidroína.

Figura 2: muestra un alineamiento de secuencia de los dominios C-terminales de la espidroína.

30 Figura 3: muestra geles de electroforesis de proteínas de fusión.

Figura 4: muestra un gel de electroforesis de una proteína SP-C33Leu obtenida a partir de una proteína de fusión.

Figura 5: muestra la actividad superficial *in vitro* de suspensiones de tensioactivo que comprenden SP-C33Leu obtenida de una proteína de fusión.

Figura 6: muestra geles de electroforesis de proteínas de fusión SP-C33Leu.

35 Figura 7: muestra un gel de electroforesis de una proteína de fusión Brichos.

Figura 8: muestra un gel de electroforesis de una proteína de fusión de GFP y de la GFP obtenida a partir de la proteína de fusión.

Figura 9: muestra un gel de electroforesis de una proteína de fusión de neuroserpina y de la neuroserpina obtenida a partir de la proteína de fusión.

Descripción detallada de la invención

5 La presente invención se refiere a la producción de proteínas y polipéptidos, y en particular de proteínas y polipéptidos no espidroína. Dependiendo del propósito de dicha producción, el producto final puede variar. Por ejemplo, puede ser deseable obtener la proteína o polipéptido insertado en una membrana lipídica, en disolución o asociada a otras biomoléculas. También cabe destacar que también puede ser altamente deseable obtener la proteína o polipéptido deseado como parte de una proteína de fusión, lo que puede proporcionar un medio adecuado para la purificación y la detección y/o proporcionar propiedades deseables, p.ej., estabilidad y solubilidad.

10 La presente invención se basa de forma general en el descubrimiento de la utilidad del fragmento N-terminal (NT) de una proteína de tela de araña como resto potenciador de la solubilidad en una proteína de fusión que es producida a partir de ADN recombinante. De esta manera, la presente invención proporciona según un primer aspecto una proteína de fusión que comprende (i) al menos un resto potenciador de la solubilidad que se deriva del fragmento NT de una proteína de tela de araña; y (ii) al menos un resto que es una proteína o polipéptido no espidroína deseado.

15 En una realización preferida, la proteína de fusión consiste en (i) al menos un resto potenciador de la solubilidad que deriva del fragmento NT de una proteína de tela de araña; y (ii) al menos un resto que es una proteína o polipéptido no espidroína deseado, que incluye opcionalmente otras características descritas en la presente memoria, p.ej., un péptido ligando y/o un sitio de división entre el resto potenciador de la solubilidad y la proteína o polipéptido deseado. En determinados experimentos, de forma sorprendente, se han obtenido rendimientos elevados de diferentes proteínas de fusión en *E. coli*. La proteína de fusión puede ser útil como tal en forma aislada, p.ej., para estudios de proteínas, por lo demás agregadas y de baja solubilidad, en forma soluble, o en la cristalización asociada a cristalografía de rayos-X. La proteína de fusión también puede dividirse para liberar la proteína deseada.

20 El término "proteína de fusión" implica en la presente memoria una proteína que se prepara mediante expresión a partir de un ácido nucleico recombinante, es decir, ADN o ARN que es creado artificialmente combinando dos o más secuencias de ácido nucleico que normalmente no existirían juntas (ingeniería genética). Las proteínas de fusión según la invención son proteínas recombinantes, y por lo tanto no son idénticas a las proteínas naturales. En particular, las espidroínas naturales no son proteína de fusión según la invención, ya que no son expresadas a partir de un ácido nucleico recombinante, como se ha establecido anteriormente. Las secuencias de ácido nucleico combinadas codifican diferentes proteínas, proteínas parciales o polipéptidos con determinadas propiedades funcionales. La proteína de fusión resultante, o la proteína de fusión recombinante, es una proteína individual con propiedades funcionales derivadas de cada una de las proteínas, proteínas parciales o polipéptidos originales.

25 En determinadas realizaciones, la proteína de fusión según la invención y los correspondientes genes son quiméricos, es decir, los fragmentos de proteína/gen derivan de al menos dos especies diferentes. El resto potenciador de la solubilidad deriva del fragmento N-terminal de una proteína de tela de araña. Según este aspecto, se prefiere que la proteína o polipéptido deseado no sea una proteína espidroína. Esto implica que la proteína o el polipéptido deseado no deriva del fragmento C-terminal, repetitivo o N-terminal de una proteína de tela de araña.

30 La proteína de fusión según la invención también puede contener uno o más péptidos ligando. El(los) péptido(s) ligando puede(n) disponerse entre el resto potenciador de la solubilidad y el resto de proteína o polipéptido deseado, o puede(n) disponerse en cualquier extremo del resto potenciador de la solubilidad y la proteína o polipéptido deseado. Si la proteína de fusión contiene dos o más restos potenciadores de la solubilidad, el(los) péptido(s) ligando también puede(n) disponerse entre dos restos potenciadores de la solubilidad. El(los) ligando(s) puede(n) proporcionar un espaciador entre las unidades funcionales de la proteína de fusión, pero también pueden constituir un medio para la identificación y la purificación de la proteína de fusión, p.ej., una etiqueta His y/o Trx. Si la proteína de fusión contiene dos o más péptidos ligando para la identificación y la purificación de la proteína de fusión, es preferible que estén separados por una secuencia espaciadora, p.ej., His₆-espaciador-His₆. El ligando también puede constituir un péptido señal, tal como una partícula de reconocimiento de señal, que dirija la proteína de fusión a la membrana y/o que cause la secreción de la proteína de fusión desde la célula hospedante al medio circundante. La proteína de fusión también puede incluir un sitio de división en su secuencia de aminoácidos, lo que permite la división y la eliminación de el(los) ligando(s) y/o del resto o restos potenciadores de la solubilidad. Los especialistas en la técnica conocen varios sitios de división, p.ej., sitios de división para agentes químicos, tales como CNBr tras residuos de Met e hidroxilamina entre residuos Asn-Gly, sitios de división para proteasas, tales como trombina o proteasa 3C, y secuencias autodivisibles, tales como secuencias autodivisibles de inteína.

35 Cada resto potenciador de la solubilidad está unido directa o indirectamente al resto de proteína o polipéptido deseado. Un enlace directo implica una unión covalente directa entre los dos restos sin secuencias intervinientes, tales como ligandos. Un enlace indirecto también implica que los dos restos están unidos mediante enlaces covalentes, pero que hay secuencias intervinientes, tales como ligandos y/o uno o más restos potenciadores de la solubilidad adicionales.

40 El al menos un resto potenciador de la solubilidad puede disponerse en cualquier extremo de la proteína o polipéptido deseado, es decir, se puede disponer C-terminalmente o N-terminalmente. Se prefiere que el al menos

un resto potenciador de la solubilidad se disponga en el extremo N-terminal de la proteína o polipéptido deseado. Si la proteína de fusión contiene uno o más péptidos ligando para identificación y purificación de la proteína de fusión, p.ej., una etiqueta His o Trx, se prefiere que se disponga en el extremo N-terminal de la proteína de fusión. El al menos un resto potenciador de la solubilidad también puede integrarse dentro de la proteína o polipéptido deseado, por ejemplo entre los dominios o partes de una proteína deseada. En una realización preferida, al menos un resto potenciador de la solubilidad constituye el extremo N-terminal y/o el extremo C-terminal de la proteína de fusión, es decir, no hay presente ningún péptido ligando u otra secuencia terminal del resto potenciador de la solubilidad. Una proteína de fusión típica según la invención puede contener 1-6, tal como 1-4, tal como 1-2 restos potenciadores de la solubilidad.

En una realización preferida, la proteína de fusión comprende al menos dos restos potenciadores de la solubilidad, derivando cada uno de ellos del fragmento N-terminal (NT) de una proteína de tela de araña. Los restos potenciadores de la solubilidad, preferiblemente dos restos potenciadores de la solubilidad, pueden disponerse consecutivamente en cualquier extremo de la proteína o polipéptido deseado, es decir, disponerse C-terminalmente o N-terminalmente. Los restos potenciadores de la solubilidad dispuestos consecutivamente también pueden integrarse dentro de la proteína o polipéptido deseado, por ejemplo entre dominios o partes de una proteína deseada. Los restos potenciadores de la solubilidad también pueden disponerse no consecutivamente, tanto en cada extremo de la proteína o polipéptido deseado, es decir, disponerse C-terminalmente y N-terminalmente, como en un extremo de la proteína o polipéptido deseado e integrarse dentro de la proteína o polipéptido deseado. Una proteína de fusión preferida típica según la invención puede contener 2-6, tal como 2-4, restos potenciadores de la solubilidad.

En una realización preferida, la proteína de fusión según la invención tiene al menos un sitio de división dispuesto entre al menos una proteína o polipéptido deseado y al menos un resto potenciador de la solubilidad. Esto permite la división de la proteína de fusión y la purificación de la proteína deseada. Sin embargo, cabe destacar que puede ser deseable obtener la proteína o polipéptido deseado como parte de una proteína de fusión, lo cual puede proporcionar un medio adecuado para la purificación y la detección y/o proporcionar propiedades deseables, p.ej., estabilidad y solubilidad. En este caso, el sitio de división puede omitirse, o el sitio de división puede incluirse pero la etapa de división omitirse.

Una proteína de fusión preferida tiene la forma de un resto potenciador de la solubilidad dispuesto N-terminalmente, acoplado mediante un péptido ligando de 1-30 residuos de aminoácido, tal como 1-10 residuos de aminoácido, a una proteína o polipéptido deseado dispuesto C-terminalmente. El péptido ligando puede contener un sitio de división. Opcionalmente, la proteína de fusión tiene un péptido ligando N-terminal o C-terminal, que puede contener una etiqueta de purificación, tal como una etiqueta His, y un sitio de división.

Otra proteína de fusión preferida tiene la forma de un resto potenciador de la solubilidad dispuesto N-terminalmente acoplado directamente a una proteína o polipéptido dispuesto C-terminalmente. Opcionalmente, la proteína de fusión tiene un péptido ligando N-terminal o C-terminal, que puede contener una etiqueta de purificación, tal como una etiqueta His, y un sitio de división.

Una proteína de fusión preferida tiene la forma de dos restos potenciadores de la solubilidad dispuestos N-terminalmente consecutivos, acoplados mediante un péptido ligando de 1-30 residuos de aminoácido, tal como 1-10 residuos de aminoácido, a una proteína o polipéptido deseado dispuesto C-terminalmente. El péptido ligando puede contener un sitio de división. Opcionalmente, la proteína de fusión tiene un péptido ligando N-terminal o C-terminal, que puede contener una etiqueta de purificación tal como una etiqueta His, y un sitio de división.

Otra proteína de fusión preferida tiene la forma de dos restos potenciadores de la solubilidad dispuestos N-terminalmente consecutivos acoplados directamente a una proteína o polipéptido deseado dispuesto C-terminalmente. Opcionalmente, la proteína de fusión tiene un péptido ligando N-terminal o C-terminal, que puede contener una etiqueta de purificación, tal como una etiqueta His, y un sitio de división.

El resto potenciador de la solubilidad es derivado del fragmento NT de una proteína de tela de araña, o espidroína. Aunque los ejemplos necesariamente se refieren a fragmentos NT específicos, en este caso a proteínas derivadas de la espidroína mayor 1 (MaSp1) de *Euprosthénops australis*, se considera que el método descrito en la presente memoria es aplicable a cualquier resto de proteína similar. Los términos "espidroínas" y "proteínas de tela de araña" se usan de forma intercambiable a lo largo de toda la descripción, y abarcan todas las proteínas de tela de araña conocidas, que incluyen las proteínas de tela de araña ampuladas mayores, normalmente abreviadas como "MaSp", o "ADF" en el caso de *Araneus diadematus*. Estas proteínas de tela de araña ampuladas mayores generalmente son de dos tipos, 1 y 2. Estos términos incluyen además los nuevos fragmentos NT de proteína según la invención, tal como se definen en las reivindicaciones anexas y en las realizaciones incluidas en la presente memoria, y otras proteínas no naturales con un grado elevado de identidad y/o similitud con respecto a los fragmentos de proteínas NT de tela de araña conocidos.

El resto potenciador de la solubilidad presenta un grado elevado de similitud con respecto a la secuencia de aminoácidos N-terminal (NT) de las proteínas de tela de araña. Tal como se muestran en la Figura 1, esta secuencia de aminoácidos está bien conservada entre diversas especies y con respecto a proteínas de tela de araña,

incluyendo la MaSp1 y la MaSp2. En la Figura 1, los siguientes fragmentos NT de espidroína están alineados, denotados con las entradas de acceso al GenBank, cuando es aplicable:

TABLA 1 – Fragmentos NT de espidroína

Código	Especie y proteína espidroína	Nº acceso GenBank
Ea MaSp1	<i>Euprosthrops australis</i> MaSp 1	AM259067
Lg MaSp1	<i>Latrodectus geometricus</i> MaSp 1	ABY67420
Lh MaSp1	<i>Latrodectus hesperus</i> MaSp 1	ABY67414
Nc MaSp1	<i>Nephila clavipes</i> MaSp 1	ACF19411
At MaSp2	<i>Argiope trifasciata</i> MaSp 2	AAZ15371
Lg MaSp2	<i>Latrodectus geometricus</i> MaSp 2	ABY67417
Lh MaSp2	<i>Latrodectus hesperus</i> MaSp 2	ABR68855
Nim MaSp2	<i>Nephila inaurata madagascariensis</i> MaSp 2	AAZ15322
Nc MaSp2	<i>Nephila clavipes</i> MaSp 2	ACF19413
Ab CySp1	<i>Argiope bruennichi</i> espidroína cilíndrica 1	BAE86855
Ncl CySp1	<i>Nephila clavata</i> espidroína cilíndrica 1	BAE54451
Lh TuSp1	<i>Latrodectus hesperus</i> espidroína tubuliforme	ABD24296
Nc Flag	<i>Nephila clavipes</i> proteína de tela flageliforme	AF027972
Nim Flag	<i>Nephila inaurata madagascariensis</i> proteína de tela flageliforme	AF218623 (traducida)

5 Para cada secuencia solo se muestra la parte correspondiente al dominio N-terminal, omitiendo el péptido señal. Nc flag y Nlm flag están traducidas y editadas de acuerdo a Rising A. *et al.* Biomacromolecules 7, 3120-3124 (2006)).

10 No es crítico qué resto potenciador de la solubilidad específico está presente en las proteínas de fusión según la invención, siempre que el resto potenciador de la solubilidad no esté completamente ausente. Por tanto, el resto potenciador de la solubilidad según la invención puede seleccionarse de cualquiera de las secuencias de aminoácidos mostradas en la Figura 1, o de secuencias con un alto grado de similitud. Se puede usar una amplia variedad de secuencias potenciadoras de la solubilidad en la proteína de fusión según la invención. En base a las secuencias homólogas de la Figura 1, la siguiente secuencia constituye una secuencia de aminoácidos potenciadora de la solubilidad de consenso:

15 QANTPWSSPNLADAFINSF(ML)SA(A/I)SSSGAFSADQLDDMSTIG(D/N/Q)T
LMSAMD(N/S/K)MGRSG(K/R)STKSKLQALNMAFASSMAEIAAAESGG(G/Q)
SVGVKTNAISDALSSAFYQTTGSVNPQFV(N/S)EIRSLI(G/N)M(F/L)(A/S)QAS
ANEV (SEQ ID NO 8).

20 La secuencia del resto potenciador de la solubilidad según la invención tiene al menos un 50% de identidad, preferiblemente al menos un 60% de identidad, con respecto a la secuencia de aminoácidos de consenso SEQ ID NO 8, que se basa en las secuencias de aminoácidos de la Figura 1. En una realización preferida, la secuencia del resto potenciador de la solubilidad según la invención tiene al menos un 65% de identidad, preferiblemente al menos un 70% de identidad, con respecto a la secuencia de aminoácidos de consenso SEQ ID NO 8. En realizaciones preferidas, el resto potenciador de la solubilidad según la invención tiene además un 70%, preferiblemente un 80%, de similitud con respecto a la secuencia de aminoácidos de consenso SEQ ID NO 8.

25 Un resto potenciador de la solubilidad representativo según la invención es la secuencia de *Euprosthrops australis* SEQ ID NO 6. Según una realización preferida de la invención, el resto potenciador de la solubilidad tiene al menos un 80% de identidad con respecto a la SEQ ID NO 6 o a cualquier secuencia de aminoácidos individual de la Figura 1. En realizaciones preferidas de la invención, el resto potenciador de la solubilidad tiene al menos un 90%, tal como al menos un 95% de identidad con respecto a la SEQ ID NO 6 o a cualquier secuencia de aminoácidos individual de

la Figura 1. En realizaciones preferidas de la invención, el resto potenciador de la solubilidad es idéntico a la SEQ ID NO 6 o a cualquier secuencia de aminoácidos individual de la Figura 1, en particular la Ea MaSp1.

La expresión “% de identidad”, tal como se usa a lo largo de la especificación y las reivindicaciones anexas, se calcula como se indica a continuación. La secuencia de consulta es alineada con la secuencia diana usando el algoritmo CLUSTAL W (Thompson, J.D., Higgins, D.G. y Gibson, T.J., *Nucleic Acids Research*, 22: 4673-4680 (1994)). Se realiza una comparación para la ventana correspondiente a la más corta de las secuencias alineadas. Se comparan los residuos de aminoácido de cada posición, y el porcentaje de posiciones de la secuencia de consulta que tienen correspondencias idénticas en la secuencia diana se presenta como % de identidad.

La expresión “% de similitud”, tal como se usa a lo largo de la especificación y de las reivindicaciones anexas, se calcula como se ha descrito para el “% de identidad”, con la excepción de que los residuos hidrofóbicos Ala, Val, Phe, Pro, Leu, Ile, Trp, Met y Cys son similares; los residuos básicos Lys, Arg y His son similares; los residuos ácidos Glu y Asp son similares; y los residuos hidrofílicos no cargados Gln, Asn, Ser, Thr y Tyr son similares. El aminoácido natural restante Gly no es similar a ningún otro aminoácido en este contexto.

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

El resto potenciador de la solubilidad contiene entre 100 y 160 residuos de aminoácido. Es preferible que el resto potenciador de la solubilidad contenga al menos 100, o más de 110, preferiblemente más de 120, residuos de aminoácido. También se prefiere que el resto potenciador de la solubilidad contenga como mucho 160, o menos de 140 residuos de aminoácido. Un resto potenciador de la solubilidad típico contiene aproximadamente 130-140 residuos de aminoácido.

Según otro aspecto, la proteína o polipéptido deseado según la invención es una proteína o polipéptido no espidroína cuando la proteína de fusión comprende un resto potenciador de la solubilidad que deriva del fragmento N-terminal (NT) de una proteína de tela de araña. En una realización preferida, la proteína o polipéptido deseado es una proteína o polipéptido no espidroína. La secuencia de una proteína o polipéptido deseado no espidroína según la invención preferiblemente tiene menos de un 30% de identidad, tal como menos de un 20% de identidad, preferiblemente menos de un 10% de identidad, con respecto a cualquiera de las secuencias de aminoácido de espidroína descritas en la presente memoria, y específicamente con respecto a cualquiera de las SEQ ID NO: 6-10.

En una realización preferida, la proteína o polipéptido deseado no espidroína deriva de esponjas, ctenóforos (“portadores de peines”), medusas, corales, anémonas, gusanos planos, rotíferos, gusanos redondos, gusanos de cinta, almejas, caracoles, pulpos, gusanos segmentados, crustáceos, insectos, briozoos, braquiópodos, foronídeos, estrellas de mar, erizos de mar, tunicados, lanceolados, vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial. La palabra “deriva” significa que la secuencia de una proteína o polipéptido deseado no espidroína según la invención tiene preferiblemente al menos un 50% de identidad, preferiblemente la menos un 60%, preferiblemente al menos un 70%, más preferiblemente al menos un 80% de identidad, o incluso al menos un 90% de identidad, tal como un 95-100% de identidad, con respecto a una proteína natural correspondiente, y que presente una función mantenida. En una realización preferida, la proteína o polipéptido deseado no espidroína deriva de moluscos, insectos, vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial. En una realización preferida, la proteína o polipéptido deseado no espidroína deriva de vertebrados, que incluyen humanos, plantas, hongos, levaduras, bacterias, arqueobacterias o virus, o es una proteína o polipéptido artificial.

En una realización preferida, la proteína o polipéptido deseado no espidroína se selecciona del grupo que consiste en proteínas y polipéptidos formadores de amiloide, proteínas y polipéptidos que contienen disulfuros, apolipoproteínas, proteínas y polipéptidos de membrana, fármacos y dianas farmacológicas de proteínas y polipéptidos, proteínas y polipéptidos propensos a la agregación, y proteasas.

Las proteínas y polipéptidos formadores de amiloide según la invención incluyen proteínas y polipéptidos que están asociados a amiloide de enfermedad y funcional. Los ejemplos de proteínas y polipéptidos formadores de amiloide incluyen amiloide beta péptido (A β -péptido), polipéptido amiloide de isleta (amilina o IAPP), proteína de prion (PrP), α -sinucleína, calcitonina, prolactina, cistatina, factor natriurético atrial (ATF) y actina. En la Tabla 3 se enumeran ejemplos de proteínas y polipéptidos formadores de amiloide según la invención.

TABLA 3 – Proteínas y polipéptidos formadores de amiloide

Proteína	ID Uniprot
Aβ1-42	P05067
Apolipoproteína SAA	P02735
Cistatina C	P01034
Transtiretina	P02766
Lisozima	P61626
α-sinucleína	P37840
Proteína de prion	P04156
ODAM	A1E959
Lactadherina	Q08431
Tau	P10636
Gelsolina	P06396
ABri, ADan	Q9Y287
Insulina	P01308
Apolipoproteína A-II	P02652
Apolipoproteína A-IV	P06727
Semenogelina I	P04279
Queratoepitelina	Q15582
Lactotransferrina	P02788
Fibrinógeno cadena α	P02671
ANF	P01160
IAPP	P10997
β2-microglobulina	P61769
Calcitonina	P01258
Prolactina	P01236
Apolipoproteína A-I	P02647
CsgA	P28307
Sup35	C7GN25
Pmel17	P40967
HET-s	A8HR89
Ure2p	Q8NIE6

Los ejemplos de proteínas y polipéptidos que contienen disulfuro incluyen proteína tensioactiva B (SP-B) y variantes de la misma, tales como Mini-B, Mini-B27, Mini-BLeu, α-defensinas y β-defensinas. Sin pretender establecer ninguna teoría específica, se contempla que el resto potenciador de la solubilidad promueve la formación deseada de enlaces disulfuro intracadena sobre los enlaces disulfuro intercadena en las defensinas y otras proteínas y

polipéptidos que contienen disulfuros. En la Tabla 4 se enumeran ejemplos de proteínas y polipéptidos que contienen disulfuros según la invención.

TABLA 4 – Proteínas y polipéptidos que contienen disulfuros

Proteína	Secuencia / ID Uniprot
SP-B humana	FPIPLPYCWLCRALIKRIQAMIPK GALAVAVAQVCRVWPL VAGGICQCLAERYSVILLDTLLGRMLPQLVCRLVLRCSM ^a
SP-B de ratón	LPIPLPFCWLCRRTLKRIKRVQAVIPKGV LAVAVSQVCHVWPL VGGICQCLAERYTVLLLDALLGRVVPQLVCGLVLRCS T ^a
SP-B de cerdo	FPIPLPFCWLCRRTLKRIQAVVPKGVLLKAVAQVCHVWPL PVGICQCLAERYIVICLNMLLDRTLPQLVCGLVLRCS S ^a
SP-B de conejo	FPIPLPLCWLCRRTLKRIQAMIPKGV LAMAVAQVCHVWPL VGGICQCLAERYTVILLEVLLGHVLPQLVCGLVLRCS S ^a
SP-B de rata	LPIPLPFCWLCRRTLKRIKRVQAVIPKGV LAVAVSQVCHVWPL VGGICQCLAERYTVLLLDALLGRVVPQLVCGLVLRCS T ^a
Mini-B	CWLCRALIKRIQAMIPKGG RMLPQLVCRLVLRCS ^b
Mini-BLeu	CWLCRALIKRIQALIPKGG RLLPQLVCRLVLRCS ^b
Mini-B27	CLLCRALIKRFNRYLTPQLVCRLVLRCS ^c
1a AA	CWLARALIKRIQALIPKGG RLLPQLVARLVLRCS ^d
1b AA	AWLCRALIKRIQALIPKGG RLLPQLVCRLVLRAS ^e
1a LL	CWLLRALIKRIQALIPKGG RLLPQLVLRLLVLRCS ^d
1b LL	LWLCRALIKRIQALIPKGG RLLPQLVCRLVLRLS ^e
Proinsulina	P01308
CAR D1 ^f	P78310
Brichos	SEQ ID NO: 41
^a Enlaces Cys8-Cys77, Cys11-Cys71, Cys35-Cys46 y enlace intermolecular Cys48-Cys48 ^b Enlaces Cys1-Cys33 y C4-C27 ^c Enlaces Cys1-Cys27 y C4-C21 ^d Enlace Cys1-Cys33 ^e Enlace Cys4-Cys27 ^f Receptor de virus de Coxsackie y de adenovirus	

5 Los ejemplos de apolipoproteínas incluyen apolipoproteínas de clase A-H. En la Tabla 5 se incluyen ejemplos de apolipoproteínas según la invención.

TABLA 5 – Apolipoproteínas

Proteína	Secuencia / ID Uniprot
Apolipoproteína B-100	P04114
Apolipoproteína C-1	P02654
Apolipoproteína D	P05090
Apolipoproteína E	P02649

Los ejemplos de proteínas y polipéptidos de membrana incluyen receptores asociados a membrana, que incluyen receptores de citocina, KL4, LL-37, proteína tensioactiva C (SP-C) y variantes de la misma, tales como SP-C(Leu), SP-C33, SP-C30 y SP-C33Leu. Otros ejemplos específicos incluyen las proteínas SP-C33Leu fusionada a Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL ó SP-B, opcionalmente a través de un ligando, p.ej., Gly_n, Leu_n, Gly-Ala_n o similar. La SP-C33Leu puede estar dispuesta N-terminalmente o, preferiblemente, C-terminalmente respecto a la proteína Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL ó SP-B. En la Tabla 6 se incluyen ejemplos de proteínas y polipéptidos de membrana según la invención.

5

TABLA 6 – Proteínas y polipéptidos de membrana

Proteína	Secuencia / ID Uniprot
SP-C	FGIPCCPVHLKRLLIIVVVVVVLIIVVVIVGALLMGL *
SP-C(Leu)	FGIPSSPVHLKRLLLLLLLLLLILLLLILGALLMGL
SP-C33	IPSSPVHLKRLKLLLLLLLLLILLLLILGALLMGL
SP-C30	IPSSPVHLKRLKLLLLLLLLLILLLLILGALL
SP-C33(Leu)	IPSSPVHLKRLKLLLLLLLLLILLLLILGALLLGL
LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLPRTES
KL4	KLLLLKLLLLKLLLLKLLLL
* Cys-5 y Cys-6 de la SP-C nativa están palmitoilados	
Proteína	
ID Uniprot	
Receptor de hormona de crecimiento	P10912
Receptor 35 acoplado a proteína G	Q9HC97
Receptor de insulina	P06213
Receptor de hormona de liberación de gonadotropina	P30968
Receptor de lipoproteína de densidad muy baja	P98155
Receptor beta de TFG, tipo 1	P36897
Receptor D2 de prostaglandina	Q13258
Receptor de tirosina-proteína quinasa erbB-2 (HER2)	P04626
Receptor de tirosina-proteína quinasa erbB-4 (HER4)	Q15303
Receptor de tirosina-proteína quinasa erbB-3 (HER3)	P21860
Acuaporina-1	P29972
Acuaporina-2	P41181

10

Proteína de canal de cloruro CIC-Ka	P51800
Proteína de canal de cloruro CIC-Kb	P51801
Proteína de membrana integral DGCR2/IDD	P98153
Receptor de interleucina 9	Q01113

5 Los ejemplos de proteínas y polipéptidos fármacos y dianas farmacológicas incluyen hormonas que se producen recombinantemente, que incluyen hormonas de péptido y de proteína, tales como eritropoyetina (EPO) y hormona del crecimiento (HC), citocinas, factores de crecimiento, tales como los factores de crecimiento de tipo insulina (IGF-I e IGF-II), KL4, LL-37, proteína tensoactiva C (SP-C) y sus variantes, tales como SP-C(Leu), SP-C33, SP-C30 y SP-C33Leu. Otros ejemplos específicos incluyen SP-C33Leu fusionada a proteínas Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL o SP-B, opcionalmente a través de un ligando, p.ej., Gly_n, Leu_n, Gly-Ala_n o similares. La SP-C33Leu se puede disponer N-terminal o, preferiblemente, C-terminal con respecto a la proteína Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL o SP-B. En la Tabla 7 se presentan ejemplos de proteínas y polipéptidos fármacos o dianas farmacológicas según la invención.

TABLA 7 – Proteínas y polipéptidos fármacos y dianas farmacológicas

Proteína	Secuencia / ID Uniprot
Factor de crecimiento de tipo insulina IA	P01243
Factor de crecimiento de tipo insulina IB	P05019
Hormona de crecimiento 1, variante 1	Q6IYF1
Hormona de crecimiento 1, variante 2	Q6IYF0
Hormona de crecimiento 2, variante 2	B1A4H7
Insulina	P01308
Eritropoyetina	P01588
Factor de coagulación VIII	P00451
Factor de coagulación IX	P00740
Protrombina	P00734
Albúmina de suero	P02768
Antitrombina III	P01008
Interferón alfa	P01563
Somatotropina	P01241
Alérgeno de polen mayor Bet v 1-A	P15494
OspA (<i>Piscirickettsia salmonis</i>)	Q5BMB7
Variante de antígeno de 17 kDa de OspA (<i>P. salmonis</i>)	Q9F9K8
Factor de crecimiento transformante beta-1	P01137
Factor de crecimiento transformante beta-2	P61812
Factor de crecimiento transformante beta-3	P10600
Interleucina 1 beta	P01584
Interleucina 1 alfa	P01583
Interleucina 2	P60568

ES 2 608 673 T3

Interleucina 3	P08700
Interleucina 4	P05112
Interleucina 5	P05113
Interleucina 6	P05231
Interleucina 7	P13232
Interleucina 8	P10145
Interleucina 9	P15248
Interleucina 10	P22301
Interleucina 12 subunidad alfa	P29459
Interleucina 12 subunidad beta	P29460
Interleucina 18	Q14116
Interleucina 21	Q9HBE4
Linfopoyetina estromal tímica	Q969D9
Brichos	SEQ ID NO: 41
Neuroserpina	SEQ ID NO: 49
Proteína	Secuencia
SP-C	FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL ^a
SP-C(Leu)	FGIPSSPVHLKRLKLLLLLLLLLILGALLMGL
SP-C33	IPSSPVHLKRLKLLLLLLLLLILGALLMGL
SP-C30	IPSSPVHLKRLKLLLLLLLLLILGALL
SP-C33(Leu)	IPSSPVHLKRLKLLLLLLLLLILGALLGL
LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES
KL4	KLLLLKLLLLKLLLLKLLLLK
1a AA	CWLRALIKRIQALIPKGGRLLPQLVARLVLRCS ^b
1b AA	AWLCRALIKRIQALIPKGGRLLPQLVCRLVLRAS ^c
0 AAAA	AWLRALIKRIQALIPKGGRLLPQLVARLVLRAS
1a LL	CWLLRALIKRIQALIPKGGRLLPQLVLRVLVLRCS ^b
1b LL	LWLCRALIKRIQALIPKGGRLLPQLVCRLVLRAS ^c
0 LLLL	LWLLRALIKRIQALIPKGGRLLPQLVLRVLVLRAS
^a Cys5 y Cys6 de la SP-C nativa están palmitoilados	
^b Enlace Cys1-Cys33	
^c Enlace Cys4-Cys27	

Los ejemplos de proteínas y polipéptidos propensos a la agregación incluyen avidina, estreptavidina y partes de unión a ligando extracelulares de receptores de citocina. En la Tabla 8 se presentan proteínas y polipéptidos propensos a la agregación según la invención.

TABLA 8 – Proteínas y polipéptidos propensos a la agregación

Proteína	ID Uniprot / otra referencia
Estreptavidina, <i>Streptomyces avidinii</i>	P22629
Estreptavidina, <i>Streptomyces lavendulae</i>	B8YQ01
Estreptavidina V1, <i>Streptomyces venezuelae</i>	Q53532
Estreptavidina V2, <i>Streptomyces venezuelae</i>	Q53533
Estreptavidina putativa, <i>Burkholderia mallei</i> (cepa SAVP1)	A1V7Z0
Estreptavidina putativa, <i>Burkholderia thailandensis</i>	Q2T1V4
Estreptavidina putativa, <i>Burkholderia mallei</i>	Q62EP2
Estreptavidina de núcleo	GenBank: CAA77107.1
M4 (muteína cuádruple de estreptavidina)	J Biol Chem 280(24): 23225-23231 (2005)
Avidina, <i>Gallus gallus</i>	P02701 GenBank: CAC34569.1
Actina	P68133
Receptor de interleucina 6 subunidad alfa	P08887
Receptor de interleucina 6 subunidad beta	P40189
Receptor de interleucina 2 subunidad alfa	P01589
Receptor de interleucina 2 subunidad beta	P14784
Receptor de citocina común subunidad gamma	P31785
Proteína Fluorescente Verde (GFP)	SEQ ID NO: 44

Los ejemplos de proteasas incluyen la proteasa 3C de virus de coxsackie o de rinovirus humano. En la Tabla 9 se presentan otros ejemplos de proteasas según la invención.

5

TABLA 9 – Proteasas

Proteasa	Clase	Nº de acceso
Tripsina (bovina)	serina	P00760
Quimotripsina (bovina)	serina	P00766
Elastasa (porcina)	serina	P00772
Endoproteinasa Arg-C (glándula maxilar de ratón)	serina	
Endoproteinasa Glu-C (proteasa V8) (<i>Staphylococcus aureus</i>)	serina	P04188
Enzima liberadora de ácido acilamino (porcina)	serina	P19205
Carboxipeptidasa (<i>Penicillium janthinellum</i>)	serina	P43946
Proteinasa K (<i>Tritirachium album</i>)	serina	P06873
Subtilisina (<i>Bacillus subtilis</i>)	serina	P04189

		P029122
Carboxipeptidasa Y (levadura)	serina	P00729
Endoproteinasa Lys-C (<i>Lysobacter enzymogenes</i>)	serina	S77957
Enteropeptidasa (humana)	serina	P98073
Protrombina	serina	P00734
Factor X	serina	P00742
Pepsina	aspártico	P00791 P00790
Catepsina D (humana)	aspártico	P07339
Proteasa de VIH-1	aspártico	Q9YQ34
Catepsina C	cisteína	
Clostripaína (endoproteinasa-Arg-C) (<i>Clostridium histolyticum</i>)	cisteína	P09870
Papaína (<i>Carica papaya</i>)	cisteína	P00784
Proteasa 3C	cisteína	Q04107
Virus del grabado del tabaco (TEV)	cisteína	Q0GDU8
Termolisina (<i>Bacillus thermo-proteolyticus</i>)	metaló	P00800
Endoproteinasa Asp-N (<i>Pseudomonas fragi</i>)	metaló	Q9R4J4
Carboxipeptidasa A (bovina)	metaló	P00730
Carboxipeptidasa B (porcina)	metaló	P00732
Proteasa IgA	metaló	Q97QP7

5 En realizaciones preferidas de la invención, la proteína deseada no espidroína se selecciona entre proteína tensioactiva B (SP-B) y sus variantes, tales como Mini-B, Mini-B27, Mini-BLeu, KL4, LL-37, y proteína tensioactiva C (SP-C) y sus variantes, tales como SP-C(Leu), SP-C33, SP-C30 y SP-C33Leu. Otras proteínas no espidroína preferidas según la invención son neuroserpina, GFP y las proteínas 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL y 0 LLLL.

En determinadas realizaciones preferidas de la invención, la proteína de fusión se selecciona del grupo que consiste en las SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 y 47; y proteínas que tienen al menos un 80%, preferiblemente al menos un 90%, más preferiblemente al menos un 95% de identidad con respecto a dichas proteínas.

10 Según otro aspecto, la presente invención proporciona un ácido polinucleico aislado que codifica una proteína de fusión según la invención. En una realización preferida, el ácido polinucleico aislado se selecciona del grupo que consiste en las SEQ ID NOS 27, 29, 31, 38, 40 y 48. En otra realización preferida, el ácido polinucleico aislado se selecciona del grupo que consiste en las SEQ ID NOS 14-16, 18 y 24.

15 Según un aspecto, la presente invención proporciona un nuevo uso de al menos un resto que se deriva del fragmento N-terminal (NT) de una proteína de tela de araña como un resto potenciador de la solubilidad en una proteína de fusión para la producción de una proteína o polipéptido deseado no espidroína. Cuando la proteína de fusión comprende un único resto potenciador de la solubilidad que deriva del fragmento N-terminal (NT) de una proteína de tela de araña, entonces una alternativa preferida es que la proteína deseada sea una proteína o polipéptido no espidroína. La proteína o polipéptido deseado es una proteína o polipéptido no espidroína.

20 Según otro aspecto, la presente invención proporciona un método para producir una proteína de fusión. La primera etapa implica la expresión en un hospedante adecuado de una proteína de fusión según la invención. Los hospedantes adecuados son bien conocidos por los especialistas en la técnica e incluyen, p.ej., bacterias y células eucarióticas, tales como levaduras, líneas celulares de insecto y líneas celulares de mamífero. Típicamente, esta etapa implica la expresión de una molécula de ácido polinucleico que codifica la proteína de fusión en *E. coli*.

La segunda etapa del método implica la obtención de una mezcla que contenga la proteína de fusión. La mezcla puede obtenerse por ejemplo lisando o rompiendo mecánicamente las células hospedantes. La mezcla también puede obtenerse recolectando el medio de cultivo celular, si la proteína de fusión es secretada por la célula hospedante. La proteína obtenida de esta manera puede ser aislada usando procedimientos estándar. Si se desea, esta mezcla puede someterse a centrifugación, y se puede recolectar la fracción apropiada (precipitado o sobrenadante). La mezcla que contiene la proteína de fusión también puede someterse a filtración en gel, cromatografía, p.ej., cromatografía de intercambio aniónico, diálisis, separación de fases o filtración para producir la separación. Opcionalmente, se eliminan activamente los lipopolisacáridos y otros pirógenos en esta etapa. Si se desea, los péptidos ligando pueden eliminarse mediante división en esta etapa.

En una realización preferida, la mezcla obtenida comprende la proteína de fusión disuelta en un medio líquido, típicamente un tampón de sales o un medio de cultivo celular. En una realización preferida, la mezcla tiene un pH inferior a 6,3 y preferiblemente inferior a 6, lo que promueve la conformación de dominios NT solubles. En otra realización preferida, la mezcla tiene un pH superior a 6,4, y preferiblemente superior a 7, lo que evita o disminuye la formación de dominios NT solubles. Un pH superior a 6,4, tal como superior a 7, puede ser particularmente útil para mejorar la solubilidad de las proteínas de fusión según la invención, en donde la proteína/polipéptido deseado deriva de una proteína espidroína o en donde la proteína/polipéptido deseado es una proteína/polipéptido formador de amiloide o propenso a la agregación.

Según un aspecto relacionado, la presente invención proporciona un método para producir una proteína o polipéptido deseado. La primera etapa implica la expresión en un hospedante adecuado de una proteína de fusión según la invención. Los hospedantes adecuados son bien conocidos por los especialistas en la técnica e incluyen, p.ej., bacterias y células eucarióticas, tales como levaduras, líneas celulares de insecto y líneas celulares de mamífero. Típicamente, esta etapa implica la expresión de una molécula de ácido polinucleico que codifica la proteína de fusión en *E. coli*.

La segunda etapa del método implica la obtención de una mezcla que contiene la proteína de fusión. La mezcla puede obtenerse, por ejemplo, lisando o rompiendo mecánicamente, p.ej., sonicando, las células hospedantes. La mezcla también puede obtenerse recolectando el medio de cultivo celular, si la proteína de fusión es secretada por la célula hospedante. La proteína obtenida de esta manera puede ser aislada usando procedimientos estándar. Si se desea, esta mezcla puede someterse a centrifugación, y se puede recolectar la fracción apropiada (precipitado o sobrenadante). La mezcla que contiene la proteína de fusión también puede someterse a filtración en gel, cromatografía, p.ej., cromatografía de intercambio aniónico, diálisis, separación de fases o filtración para producir la separación. Opcionalmente, se eliminan activamente los lipopolisacáridos y otros pirógenos en esta etapa. Si se desea, los péptidos ligando pueden eliminarse mediante división en esta etapa. Como se ha establecido antes, ésta puede ser la forma más adecuada de la proteína o polipéptido deseado, es decir, como parte de una proteína de fusión. Puede proporcionar un medio adecuado para la purificación y la detección y/o proporcionar propiedades deseables, p.ej., estabilidad y en particular solubilidad.

En una realización preferida, el método también puede comprender la etapa de dividir la proteína de fusión para proporcionar la proteína o polipéptido deseado. En esta realización, la proteína de fusión comprende al menos un sitio de división dispuesto entre al menos un resto de proteína o polipéptido y al menos un resto potenciador de la solubilidad. En una proteína de fusión típica, esto implica la presencia de un único sitio de división entre el resto o restos potenciadores de la solubilidad y la proteína o polipéptido deseado. La división puede lograrse usando procedimientos estándar, por ejemplo mediante división con bromuro de cianógeno (CNBr) detrás de residuos Met, división mediante hidroxilamina entre residuos Asn y Gly, división mediante proteasa 3C entre residuos Gln y Gly en sitios -XLETLFQGX-, y en diversos otros sitios de proteasa que son bien conocidos por el especialista en la técnica.

La proteína o polipéptido deseado así obtenido puede aislarse usando procedimientos estándar. Si se desea, esta mezcla puede someterse a centrifugación, y se puede recolectar la fracción apropiada (precipitado o sobrenadante). La mezcla que contiene la proteína de fusión también puede someterse a filtración en gel, cromatografía, p.ej., cromatografía de intercambio aniónico, diálisis, separación de fases o filtración para producir la separación. Opcionalmente, se eliminan activamente los lipopolisacáridos y otros pirógenos en esta etapa. Si se desea, los péptidos ligando pueden eliminarse mediante división en esta etapa.

En una realización preferida, la mezcla obtenida comprende la proteína de fusión disuelta en un medio líquido, típicamente un tampón de sales o un medio de cultivo celular. En una realización preferida, la mezcla tiene un pH inferior a 6,3 y preferiblemente inferior a 6, tal como en el intervalo 4,2-6,3 ó 4,2-6, lo cual promueve la conformación de dominios NT solubles. En otra realización preferida, la mezcla tiene un pH superior a 6,4, y preferiblemente superior a 7, lo que evita o disminuye la formación de dominios NT solubles. Un pH por encima de 6,4, tal como superior a 7, puede ser particularmente útil para mejorar la solubilidad de proteínas de fusión según la invención en donde la proteína/polipéptido deriva de una proteína espidroína o en donde la proteína/polipéptido deseado es una proteína/polipéptido formador de amiloide o propenso a la agregación.

Por tanto, la proteína de fusión típicamente se obtiene como una disolución en un medio líquido. Con los términos "soluble" y "en disolución" se pretende indicar que la proteína de fusión no está agregada de forma visible y no precipita en el disolvente a 60 000xg. El medio líquido puede ser cualquier medio adecuado, tal como un medio

acuoso, preferiblemente un medio fisiológico, típicamente un medio acuoso tamponado, tal como tampón Tris-HCl 10-50 mM o tampón de fosfato. El medio líquido preferiblemente tiene un pH de 6,4 o superior, tal como 7 o superior, y/o una composición iónica que evita la polimerización del resto potenciador de la solubilidad. Es decir, el medio líquido típicamente presenta un pH de 6,4 o superior, tal como 7 o superior, o una composición iónica que evita la polimerización del resto potenciador de la solubilidad, o ambos.

Las composiciones iónicas que evitan la polimerización del resto potenciador de la solubilidad pueden ser preparadas fácilmente por un especialista en la técnica. Una composición iónica preferida que evita la polimerización del resto potenciador de la solubilidad tiene una fuerza iónica de más de 300 mM. Los ejemplos específicos de composiciones iónicas que evitan la polimerización del resto potenciador de la solubilidad incluyen NaCl por encima de 300 mM, fosfato 100 mM y combinaciones de dichos iones que presenten el efecto preventivo deseado sobre la polimerización del resto potenciador de la solubilidad, p.ej., una combinación de fosfato 10 mM y de NaCl 300 mM.

Se ha descubierto de manera sorprendente que la presencia de un resto potenciador de la solubilidad mejora la estabilidad de la disolución y evita la formación de polímeros en estas condiciones. Esto puede ser ventajoso cuando la polimerización inmediata puede ser indeseable, p.ej., durante la purificación de la proteína, en la preparación de lotes grandes, o cuando es necesario optimizar otras condiciones. Se prefiere que el pH del medio líquido se ajuste a 6,7 ó superior, tal como 7,0 ó superior, para alcanzar una elevada solubilidad de la proteína de fusión. También puede ser ventajoso que el pH del medio líquido se ajuste al rango de 6,4-6,8, lo que proporciona una solubilidad suficiente de la proteína de tela de araña y facilita el posterior ajuste de pH a 6,3 ó inferior.

Otro aspecto de la invención se basa en la observación de que el dominio NT forma grandes estructuras solubles cuando el pH se reduce desde aproximadamente 7 a 6, o más específicamente desde por encima de 6,4 hasta por debajo de 6,3. La formación de dichas estructuras se produce más eficientemente a un pH por encima de 4,2, es decir en el rango de 4,2-6,3, tal como 4,2-6. Esta propiedad puede ajustarse para purificación por afinidad, p.ej., si se inmoviliza NT en una columna. Esta estrategia permite la liberación de proteínas ligadas por un cambio de pH dentro de un intervalo fisiológicamente relevante, ya que la estructura se resolverá cuando el pH sea elevado, desde aproximadamente 6 a 7.

En una realización preferida de los métodos según la invención, la etapa de aislar la proteína de fusión implica la purificación de la proteína de fusión en un medio de afinidad, tal como una columna de afinidad, con un resto NT inmovilizado y/o en un medio de intercambio aniónico, tal como una columna de intercambio aniónico. La purificación de la proteína de fusión en un medio de afinidad se lleva a cabo preferiblemente en asociación a un medio de afinidad con un resto NT inmovilizado a un pH de 6,3 o inferior, preferiblemente en el rango de 4,2-6,3, seguido de la disociación desde el medio de afinidad con un medio de disociación deseado, p.ej., que tenga un pH de 6,4 o superior, un pH de 4,1 o inferior y/o que tenga una elevada fuerza iónica. La purificación de la proteína de fusión en un medio de intercambio aniónico se lleva a cabo de forma preferible en asociación al medio de intercambio aniónico a un pH de 6,4 o superior, seguida de la disociación desde el medio de intercambio aniónico con un medio de disociación que tenga una elevada fortaleza iónica. Si se desea, la purificación de la proteína de fusión en un medio de afinidad, tal como una columna de afinidad, con un resto NT inmovilizado, se puede combinar con la purificación en un medio de intercambio aniónico, tal como una columna de intercambio aniónico. Un medio de disociación que tiene una elevada fortaleza iónica típicamente presenta una fortaleza iónica de más de 300 mM, tal como NaCl por encima de 300 mM.

Estos dos procedimientos basados en afinidad utilizan las propiedades inherentes del resto potenciador de la solubilidad según la invención. De particular interés es la fuerte tendencia de los fragmentos de proteína NT espidroína a asociarse a un pH inferior a 6,3, en particular en el rango de 4,2-6,3. Esto puede utilizarse de forma ventajosa como una potente herramienta de purificación por afinidad, que permite la purificación en una etapa de proteínas de fusión según la invención a partir de mezclas complejas. Aunque la cromatografía es preferible, obviamente se pueden utilizar otros métodos de purificación basados en afinidad, tales como partículas magnéticas con superficies funcionalizadas o filtros con superficies funcionalizadas.

La presente invención se ilustrará adicionalmente a continuación a través de los siguientes ejemplos no limitativos.

Ejemplos

Ejemplo 1 – Producción de una proteína de fusión SP-C33Leu

Se construyó un vector de expresión que comprende un gen que codifica NT-MetSP-C33Leu como una fusión a His₆ (SEQ ID NOS: 26-27). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron durante otras 3 horas a 25°C. Las células fueron recolectadas mediante centrifugación y resuspendidas en Tris-HCl 20 mM, pH 8.

Se añadió lisozima, y las células se incubaron durante 30 minutos en hielo. Se añadió Tween a una concentración final de 0,7%. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. El sobrenadante se cargó en una columna de Ni-NTA sepharose, equilibrada con tampón Tris-HCl 20 mM, pH 8, que

contenía un 0,7% de Tween. La columna se lavó con tampón Tris-HCl 20 mM, pH 8, que contenía un 0,7% de Tween, y la proteína ligada se eluyó con tampón de Tris-HCl 20 mM, imidazol 300 mM, pH 8, que contenía un 0,7% de Tween.

- 5 El eluato se sometió a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras. En la Figura 3A se indica con una flecha una banda principal correspondiente a la proteína de fusión. El rendimiento se determinó por mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO_{600} de 1. El rendimiento fue de 64 mg/L. Se concluye que una proteína de fusión que contiene un único resto NT da como resultado un rendimiento sorprendentemente elevado en presencia de detergente en el lisato celular.

Ejemplo 2 – Producción de una proteína de fusión SP-C33Leu

- 10 Se construyó un vector de expresión que comprende un gen que codifica NT₂-MetSP-C33Leu (es decir, NTNT-MetSP-C33Leu) como una fusión a His₆ (SEQ ID NOS: 28-29). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO_{600} de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron durante otras 3 horas a 25°C. Las células fueron recolectadas mediante centrifugación y resuspendidas en Tris-HCl
15 20 mM, pH 8.

- Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. El Tween fue añadido, o no, hasta una concentración final de 0,7%. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. El sobrenadante se cargó en una columna de Ni-NTA sepharose, equilibrada con tampón Tris-HCl 20 mM,
20 pH 8, ± 0,7% de Tween. La columna se lavó con tampón Tris-HCl 20 mM, pH 8, ± 0,7% de Tween, y la proteína ligada se eluyó con tampón de Tris-HCl 20 mM, imidazol 300 mM, pH 8, ± 0,7% de Tween.

- El eluato se sometió a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras. En la Figura 3B se indica con una flecha una banda principal correspondiente a la proteína de fusión. El rendimiento se determinó por mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO_{600} de 1. El rendimiento fue de 40 mg/L en ausencia de Tween, y de 68 mg/L en presencia de Tween al 0,7%. Se concluye que una proteína de fusión que contiene dos restos NT consecutivos da como resultado un rendimiento sorprendentemente elevado en ausencia de detergente en el lisato celular, y un rendimiento incluso más elevado en presencia de detergente en el lisato celular.

Ejemplo 3 – Producción de proteínas de fusión SP-C33Leu

- 30 Se construyeron vectores de expresión que comprenden un gen que codifica NT-MetSP-C33Leu, NT₂-MetSP-C33Leu y NT-MetSP-C33Leu-NT, respectivamente. Los vectores se usaron para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO_{600} de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron durante otras 3 horas a 25°C. Las células fueron recolectadas mediante centrifugación y resuspendidas en Tris-HCl 20 mM, pH 8.

- 35 Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. El Tween fue añadido, o no, hasta una concentración final de 0,7%. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. Los lisatos celulares se centrifugaron a 20 000 x g durante 30 minutos.

Ejemplo 4 – Preparación de NT-Sepharose

- 40 Se usa una construcción CysHis₆T para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences). Las células son cultivadas a 30°C en medio Luria-Bertani que contiene kanamicina hasta una DO_{600} de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), e incubadas adicionalmente durante hasta 4 horas a temperatura ambiente. Después, las células son recolectadas y resuspendidas en Tris-HCl 20 mM, pH 8, suplementado con lisozima y DNasa I. Tras completar la lisis, los sobrenadantes de 15000g son cargados en una columna
45 empaquetada con Ni sepharose (GE Healthcare). La columna se lava intensivamente, y a continuación las proteínas ligadas son eluidas con imidazol 100-300 mM. Las fracciones que contienen las proteínas diana son agrupadas y dializadas contra Tris-HCl 20 mM, pH 8,0. La proteína Cys-His₆-NT purificada es acoplada a tior Sepharose activada usando un protocolo estándar (GE Healthcare).

Ejemplo 5 – Purificación de proteínas de fusión usando NT Sepharose

- 50 Los lisatos procedentes del Ejemplo 3 se cargan en una columna empaquetada con NT sepharose, se pre-equilibran con NaPi 20 mM, pH 6. La columna se lava intensamente con NaPi 20 mM, pH 6, y después se eluyen las proteínas ligadas con NaPi 20 mM, pH 7. Se agrupan las fracciones que contienen las proteínas diana. Las muestras de proteína se separan en geles de SDS-PAGE y a continuación se tifican con Azul Coomassie Brillante R-250. Se determina el contenido de proteína a partir de la absorbancia a 280 nm.

55

Ejemplo 6 – Purificación de proteínas de fusión en intercambiador aniónico

Los lisatos celulares procedentes del Ejemplo 3 son cargados en una columna HiTrap Q FF (GE Healthcare), se pre-equilibran con NaP 20 mM, pH 6,5. La columna se lava intensivamente y a continuación las proteínas ligadas son eluidas con un gradiente lineal de NaCl hasta 1 M. Las fracciones que contienen las proteínas diana son agrupadas. Las muestras de proteína se separan en geles de SDS-PAGE y a continuación se tiñen con Azul Coomassie Brillante R-250. Se determina el contenido de proteína a partir de la absorbancia a 280 nm.

Ejemplo 7 – División y aislamiento de la proteína deseada

Las proteínas de fusión de los Ejemplos 3, 5 y 6 se disuelven en ácido fórmico al 70% en agua, suplementado con 0,1 g/mL de CNBr y se dejan reposar a temperatura ambiente durante 24 horas. Después de eso se secan las mezclas y se separan en el sistema bifásico cloroformo/metanol/agua, 8:4:3, en volumen. La SP-C33Leu se obtiene en la fase orgánica y puede purificarse posteriormente de forma opcional mediante HPLC de fase inversa usando una columna C18. La actividad de la SP-C33Leu mezclada con fosfolípidos sintéticos se puede evaluar *in vitro* o *in vivo*, tal como se describe, p.ej., en J. Johansson *et al.*, J. Appl. Physiol. 95, 2055-2063 (2003).

Ejemplo 8 – Producción de proteína de fusión LL-37

Se construyó un vector de expresión que comprende un gen que codifica NT₂-LL37 (es decir, NTNT-LL37) como una fusión a His₆ (SEQ ID NOS: 30-31). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron durante otras 3 horas a 25°C. Las células fueron recolectadas mediante centrifugación y resuspendidas en Tris-HCl 20 mM, pH 8.

Se añadió lisozima, y las células se incubaron durante 30 minutos en hielo. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna de Ni-NTA sepharose, equilibrada con tampón Tris-HCl 20 mM, pH 8, NaCl 250 mM. La columna se lavó con tampón Tris-HCl 20 mM, pH 8, NaCl 250 mM, y la proteína ligada se eluyó con tampón de Tris-HCl 20 mM, imidazol 300 mM, pH 8.

Ejemplo 9 – Producción de NT-REP₄-CT

Se construyó un vector de expresión para producir NT-REP₄-CT como fusión N-terminal a His₆ (SEQ ID NOS 17-18). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron hasta 4 horas más a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0) suplementado con lisozima y DNasa I.

Tras completar la lisis, los sobrenadantes de 15000g fueron cargados en una columna empaquetada con Ni sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensivamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas diana fueron agrupadas y dializadas contra Tris-HCl 20 mM (pH 8,0).

Las muestras de proteína se separaron mediante SDS-PAGE y a continuación se tiñeron con Azul Coomassie Brillante R-250. La proteína NT-REP₄-CT resultante se concentró mediante ultrafiltración usando un filtro de celulosa con un corte de masa molecular de 5 kDa (Millipore).

Ejemplo 10 – Producción de NT-REP₄-CT

Se construyó un vector de expresión para producir NT-REP₄-CT como fusión C-terminal a Zbasic (SEQ ID NO 19). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron hasta 2-4 horas más a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en fosfato Na 50 mM (pH 7,5) suplementado con lisozima y DNasa I.

Tras completar la lisis, los sobrenadantes de 15000g fueron cargados en un intercambiador catiónico (HiTrap S, GE Healthcare, Uppsala, Suecia). La columna se lavó intensivamente antes de que las proteínas ligadas fueran eluidas con un gradiente contra NaCl 500 mM. Las fracciones que contenían las proteínas diana fueron agrupadas y dializadas contra fosfato Na 50 mM (pH 7,5). La proteína NT-REP₄-CT (SEQ ID NO 20) fue liberada de las etiquetas Zbasic mediante división proteolítica usando una relación de proteasa 3C:proteína de fusión de 1:50 (p/p) a 4°C durante una noche. Para eliminar la etiqueta Zbasic liberada, la mezcla de división se cargó en un segundo intercambiador catiónico, y se recogió el flujo de salida.

Ejemplo 11 – Producción de NT-REP₄-CT

Se construyó un vector de expresión para producir NT-REP₄-CT como fusión C-terminal a HisTrxHis (SEQ ID NO 21). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-

D-tiogalactopiranosido (IPTG), y se incubaron hasta 2-4 horas más a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0) suplementado con lisozima y DNasa I.

5 Tras completar la lisis, los sobrenadantes de 15000g fueron cargados en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensivamente antes de que las proteínas ligadas fueran eluidas con un gradiente contra NaCl 500 mM. Las fracciones que contenían las proteínas diana fueron agrupadas y dializadas contra Tris-HCl 20 mM (pH 8,0). La proteína NT-REP₄-CT (SEQ ID NO 22) fue liberada de las etiquetas HisTrxHis mediante división proteolítica usando una relación de trombina:proteína de fusión de 1:1000 (p/p) a 4°C durante una noche. Para eliminar la HisTrxHis liberada, la mezcla de división se cargó en una segunda columna Ni-Sepharose, y se recogió el flujo de salida.

10 Ejemplo 12 – Producción de NT₂-REP₄-CT

Se construyó un vector de expresión que comprende un gen que codifica NT₂-REP₄-CT (es decir, NTNT-REP₄-CT) como fusión a His₆ (SEQ ID NOS: 23-24). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), e incubadas durante 3 horas a 25°C. Las células fueron recolectadas mediante centrifugación y resuspendidas en Tris-HCl 20 mM, pH 8,0.

15 Se añadió lisozima y DNasa, y las células fueron incubadas durante 30 minutos en hielo. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna de Ni-NTA sepharose, equilibrada con tampón Tris-HCl 20 mM, pH 8. La columna se lavó con tampón Tris-HCl 20 mM, pH 8, y la proteína ligada se eluyó con tampón Tris-HCl 20 mM, pH 8, imidazol 300 mM.

20 El eluato se sometió a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras. En la Figura 3C una flecha indica una banda principal correspondiente a la proteína de fusión. El rendimiento se determinó como mg de proteína purificada a partir de 1 litro de cultivo en matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 30 mg/L. Se concluyó que las proteínas espidroína en miniatura pueden ser expresadas de forma ventajosa como fusiones con dos restos NT.

25 Ejemplo 13 – Producción de NT-REP₄-CT, NT₂-REP₄-CT y NT-REP₈-CT

Se construyen vectores de expresión que comprende un gen que codifica NT-REP₄-CT (SEQ ID NOS: 20 y 22), NT₂-REP₄-CT (SEQ ID NO: 23) y NT-REP₈-CT (SEQ ID NO: 25), respectivamente. Los vectores se usan para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences) que son cultivadas a 30°C en medio Luria-Bertani que contiene kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incuban durante otras 3 horas a 25°C. Las células se recolectan mediante centrifugación y se resuspenden en Tris-HCl 20 mM, pH 8.

30 Se añade lisozima y las células se incuban durante 30 minutos en hielo. Se añade Tween, o no, hasta una concentración final de 0,7%. Los lisatos celulares son centrifugados a 20 000 x g durante 30 minutos. Se carga una porción de sobrenadante en una columna de intercambio aniónico de acuerdo al Ejemplo 6.

35 Se prepara un medio de afinidad NT como se ha descrito en el Ejemplo 4. Se carga otra porción del sobrenadante en una columna de afinidad NT de acuerdo al Ejemplo 5.

Los eluatos de la columna de intercambio aniónico y de la columna de afinidad NT son sometidos a electroforesis de gel.

Ejemplo 14 – Producción de NTHis, NT₂-REP₈-CT y NT₂-Brichos

40 A) NTHis

Se construyó un vector de expresión para producir NT como fusión N-terminal a His₆ (SEQ ID NO 32). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), y se incubaron adicionalmente durante hasta 4 horas a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0) suplementado con lisozima y DNasa I.

50 Tras completar la lisis, los sobrenadantes de 15000g fueron cargados en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas diana fueron agrupadas y dializadas contra Tris-HCl 20 mM (pH 8,0). Las muestras de proteína fueron separadas mediante SDS-PAGE y teñidas a continuación con Azul Coomassie Brillante R-250. La proteína NT resultante (SEQ ID NO 32) se concentró mediante ultrafiltración usando un filtro de celulosa con un corte de masa molecular de 5 kDa (Millipore). El rendimiento fue de 112 mg/litro de matraz agitado cultivado hasta una DO₆₀₀ de 1.

B) NT₂-REP₈-CT

Se construyó un vector de expresión para producir NT₂-REP₈-CT (NTNT8REPCT) como fusión N-terminal a His₆ (SEQ ID NO: 33). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), e incubadas adicionalmente durante hasta 4 horas a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0) suplementado con lisozima y DNasa I. Las muestras de proteína fueron separadas mediante SDS-PAGE y teñidas a continuación con Azul Coomassie Brillante R-250.

Tras completar la lisis, los sobrenadantes de 15000g son cargados en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lava intensamente antes de que las proteínas ligadas sean eluidas con imidazol 300 mM. Las fracciones que contienen las proteínas diana son agrupadas y dializadas contra Tris-HCl 20 mM (pH 8,0). Las muestras de proteína son separadas mediante SDS-PAGE y teñidas a continuación con Azul Coomassie Brillante R-250.

C) NT₂-Brichos

Se construyó un vector de expresión para producir NT₂-Brichos (NT-NT-Brichos) como fusión N-terminal a His₆ (SEQ ID NO: 34). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de ~1, inducidas con isopropil-β-D-tiogalactopiranosido (IPTG), e incubadas adicionalmente durante hasta 4 horas a temperatura ambiente. Después de eso, las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0) suplementado con lisozima y DNasa I. Además, las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado.

Tras completar la lisis, los sobrenadantes de 15000g fueron cargados en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas diana fueron agrupadas y dializadas contra Tris-HCl 20 mM (pH 8,0). Las muestras de proteína fueron separadas mediante SDS-PAGE y teñidas a continuación con Azul Coomassie Brillante R-250. La proteína NT₂-Brichos resultante (SEQ ID NO 34) se concentró mediante ultrafiltración usando un filtro de celulosa con un corte de masa molecular de 5 kDa (Millipore). El rendimiento fue de 20 mg/litro de matraz agitado cultivado hasta una DO₆₀₀ de 1.

Ejemplo 15 – NT para captura reversible dependiente del pH

Propósito: usar NT (y NTNT) inmovilizado covalentemente para capturar de forma reversible proteínas de fusión NT.

Estrategia: investigar la conformación dependiente del pH de proteínas de fusión NT (y NTNT) en fibras (y película) con NT (y NTNT) ligado covalentemente. Como control se usan fibras y películas sin NT.

A) Fibras

Se sumergieron fibras (~0,5 cm de longitud, ~50 μg) de NT-REP₄-CT (SEQ ID NO 20), NT₂-REP₄-CT (SEQ ID NO 23) y REP₄-CT (SEQ ID NO 2, control) en 100 μL de disolución de 5 mg/mL de NTHis soluble (SEQ ID NO 32) o NT₂-Brichos (SEQ ID NO 34) a pH 8 durante 10 minutos. El pH se redujo mediante la adición de 400 μL de tampón de fosfato sódico (NaP) a pH 6 y se incubó durante 10 minutos para permitir la incorporación de NT soluble a la fibra. Las fibras fueron transferidas a 500 μL de NaP a pH 6, y se lavaron dos veces. Finalmente, las fibras fueron transferidas a 500 μL de NaP a pH 7, y se incubaron 10 minutos para permitir la liberación de NT soluble. Se realizó lo mismo en presencia de NaCl 300 mM en todos los tampones de NaP pH 6. Las muestras de las diferentes disoluciones fueron analizadas en SDS-PAGE.

Usando las fibras NT₂-REP₄-CT y NT-REP₄-CT, se capturaron a pH 6 tanto NTHis como NT₂-Brichos. Tras aumentar el pH a 7, se volvió a liberar tanto NTHis como NT₂-Brichos y se pudieron detectar en SDS-PAGE. La adición de NaCl 300 mM disminuyó la captura a pH 6.

B) Película:

Se prepararon películas de NT-REP₄-CT (SEQ ID NO 20) y REP₄-CT (SEQ ID NO 2, control) moldeando 50 μL de una disolución de proteína de 3 mg/mL en un pocillo de plástico y dejando secar durante la noche. Al siguiente día, se añadieron 100 μL de disolución de 5 mg/mL de NTHis soluble (SEQ ID NO 32) a pH 8 a los pocillos con película, y se dejó reposar durante 10 minutos. A continuación se redujo el pH hasta 6 mediante la adición de 400 μL de NaP y se incubó durante 10 minutos para permitir la incorporación de NT soluble a la película. Después se lavaron las películas dos veces con 500 μL de NaP a pH 6. Para la liberación de NTHis soluble, se añadieron 500 μL de NaP a pH 7 y se incubó durante 10 minutos. Se realizó lo mismo en presencia de NaCl 300 mM en todos los tampones de NaP pH 6. Se analizaron muestras de las diferentes disoluciones en SDS-PAGE.

El análisis en SDS-PAGE mostró que una película de NT-REP₄-CT permitió la captura de NTHis a pH 6 y su liberación al aumentar el pH a 7.

Ejemplo 16 – NT para formación reversible dependiente de pH de proteínas de fusión

Propósito: usar NT como etiqueta reversible que permita el análisis de la interacción entre restos de proteína, p.ej., analizar la interacción de Brichos con dianas con estructuras de lámina beta, p.ej., proteína tensioactiva C (SP-C).

Se mezcla NT₂-Brichos (SEQ ID NO 34) con NT₂-MetSP-C33Leu (SEQ ID NO 28) o NTHis (SEQ ID NO 32) hasta un volumen total de 100 μ L a pH 8. Se añade tampón de NaP (400 μ L) para generar un pH final de 6, y la mezcla se incuba durante 20 minutos para permitir la formación de NT. Entonces se eleva nuevamente el pH hasta pH 7 para permitir la reversión de la formación de NT. Se analizan muestras de las diferentes disoluciones en gel nativo y en cromatografía de exclusión de tamaño (SEC).

Ejemplo 17 – División de Nt-MetSP-C33Leu y aislamiento de SP-C33Leu

Se disolvieron aproximadamente 58 mg de HisNT-MetSP-C33Leu liofilizada (SEQ ID NO: 26) obtenida en el Ejemplo 1 en 3 mL de ácido fórmico al 70% en agua sometiendo a vórtice y sonicación. A dicha disolución se añadieron 200 μ L de CNBr 5 M en acetonitrilo, y la mezcla se incubó a temperatura ambiente durante 24 h. Después de eso, los disolventes fueron evaporados en una corriente de nitrógeno, y el residuo se lavó tres veces mediante solubilización en ácido fórmico al 70% en agua y secado en atmósfera de nitrógeno.

A continuación se añadió al residuo seco 4,56 mL de cloroformo/metanol/agua (8:4:3, en volumen), tras lo cual la mezcla se llevó a vórtice y se centrifugó. Se retiró la fase superior y se añadió 1 mL de cloroformo/metanol/agua (8:4:3, en volumen) a la fase inferior, y se repitió el vórtice, la centrifugación y la eliminación de la fase superior. Las dos fases superiores se combinaron y se secaron a vacío. La fase inferior se secó en atmósfera de nitrógeno.

El contenido de la fase inferior (calle izquierda) y superior (calle derecha) se analizó mediante SDS-PAGE (Figura 4). Esto demostró que la fase inferior contiene una banda principal con una masa molecular estimada que coincide bien con la de SP-C33Leu. La identidad de SP-C33Leu se confirmó mediante espectrometría de masas ESI y secuenciación de aminoácidos, que mostró una masa monoisotópica de 3594,6 Da (calculada 3594,4) y la secuencia de aminoácidos esperada.

Ejemplo 18 – Análisis de actividad superficial de la mezcla SP-C33Leu/fosfolípido

Se disolvió 1,2-dipalmitoil-*sn*-glicero-3-fosfocolina (DPPC)/1-palmitoil-2-oleoil-*sn*-glicero-3-fosfoglicerol (POPG) (68:31, p/p) en cloroformo:metanol (1:1, v/v) y se mezcló con SP-C33Leu (obtenida en el Ejemplo 17) en el mismo disolvente. El contenido de péptido en las preparaciones fue del 2% referido al peso de fosfolípido. Los disolventes fueron evaporados con nitrógeno, y las preparaciones fueron resuspendidas en disolución salina hasta una concentración final de fosfolípido de 10 mg/mL mediante rotación lenta a 37°C.

Se midió la tensión superficial por triplicado en un alveolo mediante surfactómetro de burbuja cautiva (CBS) (Schurch S *et al.*, J. Appl. Physiol. 67: 2389-2396, 1989). En el CBS, están presentes el tensioactivo y una burbuja de aire que representa el alveolo pulmonar en una cámara cerrada hermética. Para evaluar la actividad superficial en circunstancias dinámicas, la cámara es comprimida y se puede calcular la tensión superficial estudiando la forma y la relación altura/anchura de la burbuja.

En el experimento, se insertaron 2 μ L de la preparación tensioactiva de SP-C33Leu (10 mg/mL) en la cámara de ensayo rellena de sacarosa. Tras la inserción, se creó una burbuja de aire y se midió la tensión superficial durante cinco minutos de adsorción. En los siguientes experimentos cíclicos cuasi-estáticos, la burbuja se comprimió por etapas desde el volumen inicial hasta alcanzar una tensión superficial inferior a 5 mN/m, alternativamente hasta una compresión de área máxima del 50%, y después se expandió durante cinco ciclos.

Los resultados se muestran en la Figura 5, en donde se presentan el primer y el quinto ciclo de un ejemplo representativo de las tres medidas realizadas. La actividad superficial de la mezcla SP-C33Leu/DPPC/POPG (Figura 5) fue muy similar a la de la SP-C33 sintética en la misma mezcla de fosfolípidos, véase p.ej., Johansson *et al.*, J. Appl. Physiol., 95: 2055-2063 (2003).

Ejemplo 19 – Producción de una proteína de fusión SP-C33Leu

A) sin NT (ejemplo comparativo)

Se construyó un vector de expresión que comprende un gen que codifica Tiorredoxina(TRX)-SP-C33Leu como fusión a 2xHis₆ (SEQ ID NOS: 35-36). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con isopropil- β -D-tiogalactopiranosido (IPTG), e incubadas durante 3 horas a 25°C. Las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0).

Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. Se añadió Tween hasta una concentración final de 0,7%. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna empaquetada de Ni-Sepharose (GE Healthcare, Uppsala, Suecia), equilibrada con Tris-HCl 20 mM (pH 8) + 0,7% de Tween. La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM + 0,7% de Tween.

La proteína diana se eluyó con imidazol 300 mM + 0,75 de Tween y se analizó mediante SDS-PAGE (Figura 6A). El eluato contenía una cantidad pequeña e impura de proteína diana.

B) con NT

Se construyó un vector de expresión que comprende un gen que codifica TRX-NT-SP-C33Leu como fusión a 2xHis₆ (SEQ ID NOS: 37-38). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con IPTG, e incubadas adicionalmente durante 3 horas a 25°C. Las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0).

Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. Se añadió Tween hasta una concentración final de 0,7%. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna empaquetada de Ni-Sepharose (GE Healthcare, Uppsala, Suecia), equilibrada con Tris-HCl 20 mM (pH 8) + 0,7% de Tween. La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM + 0,7% de Tween. Las fracciones que contenían las proteínas de fusión se agruparon y se dializaron contra agua desionizada.

El eluato se sometió a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras. En la Figura 6B una flecha indica una banda principal que corresponde a la proteína. El rendimiento se determinó como mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 30 mg/L.

Ejemplo 20 – Producción de Brichos

Se construyó un vector de expresión que comprende un gen que codifica NT₂-Brichos (es decir, NTNT-Brichos) como fusión a His₆LinkHis₆ (SEQ ID NOS: 39-40). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con IPTG, e incubadas adicionalmente durante 3 horas a 25°C. Las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0).

Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna empaquetada de Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas de fusión se agruparon y se dializaron contra Tris-HCl 20 mM (pH 8,0).

El eluato se sometió a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras. En la Figura 7 una flecha indica una banda principal que corresponde a la proteína de fusión. El rendimiento se determinó como mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 28 mg/L de la proteína de fusión.

La proteína Brichos (SEQ ID NO: 41) se libera de las etiquetas 2His₆NT₂ mediante ruptura proteolítica usando una relación de proteasa 3C:proteína de fusión de 1:100 (p/p) a 4°C. Para eliminar la etiqueta 2His₆NT₂, la mezcla de división se carga en una segunda Ni-Sepharose, y se recoge el flujo de salida.

Ejemplo 21 – Producción de Proteína Fluorescente Verde (GFP)

La GFP utilizada en este ejemplo es una variante S147P, ver Kimata, Y *et al.*, Biochem. Biophys. Res. Commun. 232: 69-73 (1997).

A) con NT

Se construyó un vector de expresión que comprende un gen que codifica NT₂-GFP (es decir, NTNT-GFP) como fusión a His₆LinkHis₆ (SEQ ID NOS: 42-43). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con IPTG, e incubadas adicionalmente durante 3 horas a 25°C. Las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0).

5 Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna empaquetada de Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensamente antes de que las proteínas ligadas fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas de fusión se agruparon y se dializaron contra Tris-HCl 20 mM (pH 8,0). La proteína GFP (SEQ ID NO: 44) se liberó de las etiquetas 2His₆NT₂ mediante ruptura proteolítica usando una relación de proteasa 3C:proteína de fusión de 1:100 (p/p) a 4°C. Para eliminar la etiqueta 2His₆NT₂, la mezcla de división se cargó en una segunda Ni-Sepharose, y se recogió el flujo de salida.

10 Los eluatos se sometieron a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras (Figura 8). En la Figura 8 las flechas indican bandas principales correspondientes a la proteína de fusión (primer eluato, calle izquierda) y la proteína diana (segundo eluato, calle derecha). El rendimiento se determinó como mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 44 mg/L de la proteína de fusión y de 16 mg/L de la proteína diana.

15 La GFP purificada era altamente fluorescente, lo que confirma el plegamiento correcto (barril beta con enlace a hélice alfa) que es obligado para la formación autocatalítica del cromóforo.

B) con otras etiquetas de purificación: Zb y His₆ABP (ejemplo comparativo)

20 Se cultivaron células BL21(DE3) que albergan los vectores (pT7ZbGFP, pT7His₆ABPGFP) durante una noche a 37°C en medio de caldo de soja triptico suplementado con kanamicina. A la mañana siguiente los cultivos fueron inoculados en 100 mL de medio fresco en matraces de agitación de 1 litro y se cultivaron hasta alcanzar una DO₆₀₀ de 1. A continuación se indujo la producción de proteína mediante la adición de IPTG hasta una concentración final de 1 mM, y la producción continuó durante 18 h. Las células fueron recolectadas y resuspendidas en tampón de fosfato sódico 50 mM (pH 7,5). Las células fueron perturbadas mediante sonicación en hielo durante 3 minutos, alternando 1 segundo encendido y 1 segundo apagado. El lisato celular se centrifugó a 10 000 x g durante 20 minutos. Los sobrenadantes se cargaron en columnas.

25 La proteína de fusión ZbGFP (SEQ ID NO: 45) se purificó en columnas HiTrap S HP de 1 mL en fosfato sódico 50 mM, pH 7,5, y se eluyó con el mismo tampón suplementado con NaCl 160 mM.

30 La proteína de fusión His₆ABPGFP (SEQ ID NO: 46) se purificó en columnas Talon de 1 mL en fosfato sódico 50 mM, pH 8, y se eluyó con el mismo tampón suplementado con ácido acético 30 mM y acetato sódico 70 mM, lo que proporciona un pH de 5,0.

Los eluatos fueron sometidos a SDS-PAGE en un gel de gradiente 10-20% en condiciones reductoras. El rendimiento se determinó como mg de proteína purificada/1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 10 mg/L para ZbGFP y de 7 mg/L para His₆ABPGFP.

Ejemplo 22 – Producción de neuroserpina

35 Se construyó un vector de expresión que comprende un gen que codifica NT₂-Neuroserpina (es decir, NTNT-Neuroserpina) como fusión a His₆LinkHis₆ (SEQ ID NOS: 47-48). El vector se usó para transformar células BL21(DE3) de *Escherichia coli* (Merck Biosciences) que fueron cultivadas a 30°C en medio Luria-Bertani que contenía kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con IPTG, e incubadas adicionalmente durante 3 horas a 25°C. Las células fueron recolectadas y resuspendidas en Tris-HCl 20 mM (pH 8,0).

40 Se añadió lisozima, y las células fueron incubadas durante 30 minutos en hielo. Las células fueron perturbadas mediante sonicación en hielo durante 5 minutos, alternando 2 segundos encendido y 2 segundos apagado. El lisato celular se centrifugó a 20 000 x g durante 30 minutos. Los sobrenadantes se cargaron en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia). La columna se lavó intensamente antes de que las proteínas de fusión fueran eluidas con imidazol 300 mM. Las fracciones que contenían las proteínas de fusión se agruparon y se dializaron contra Tris-HCl 20 mM (pH 8,0).

45 La proteína neuroserpina (SEQ ID NO: 49) se liberó de las etiquetas 2His₆NT₂ mediante ruptura proteolítica usando una relación de proteasa 3C:proteína de fusión de 1:100 (p/p) a 4°C. Para eliminar la etiqueta 2His₆NT₂, la mezcla de división se cargó en una segunda Ni-Sepharose, y se recogió el flujo de salida.

50 Los eluatos se sometieron a SDS-PAGE en un gel de Tris-Glicina al 12% en condiciones reductoras (Figura 9). En la Figura 9 las flechas indican bandas principales correspondientes a la proteína de fusión (primer eluato, calle izquierda) y la proteína diana (segundo eluato, calle derecha). El rendimiento se determinó como mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1. El rendimiento fue de 8 mg/L de la proteína de fusión y de 4 mg/L de la proteína diana. A modo de comparación, el rendimiento de expresión de neuroserpina con etiqueta His₆ fue de 1,7 mg/L (Belorgey *et al.* Eur. J. Biochem. 271(16): 3360-3367 (2004)).

Se determinó que la tasa de inhibición de tPa (activador de plasminógeno de tejido) por la neuroserpina expresada era igual que la publicada previamente (Belorgey *et al.*, J. Biol. Chem. 277, 17367-17373 (2002)).

Ejemplo 23 – Producción de proteínas de fusión de proteasa 3C

- 5 Se construyen vectores de expresión que comprenden un gen que codifica His₆NT-3C y His₆LinkHis₆NTNT3C, respectivamente (Gräslund T. *et al.*, Protein Expr. Purif. 9(1): 125-132 (1997); Cordingley M.G. *et al.*, J. Virol. 63(12): 5037-5045 (1989)). Los vectores se usan para transformar células BL21(DE3) de *Escherichia coli* (Merk Biosciences) que son cultivadas a 30°C en medio Luria-Bertani que contiene kanamicina hasta una DO₆₀₀ de 0,9-1, inducidas con IPTG, y se incuban durante otras 3 horas a 25°C. Las células se recolectan y se resuspenden en Tris-HCl 20 mM (pH 8).
- 10 Se añade lisozima y DNasa, y las células se incuban durante 30 minutos en hielo. Las células son perturbadas mediante sonicación en hielo durante 3 minutos, alternando 1 segundo encendido y un 1 segundo apagado. El lisato celular se centrifuga a 15 000 x g durante 30 minutos. Los sobrenadantes se cargan en una columna empaquetada con Ni-Sepharose (GE Healthcare, Uppsala, Suecia), equilibrada con Tris-HCl 20 mM (pH 8,0). La columna se lava intensivamente antes de que las proteínas ligadas sean eluidas con imidazol 300 mM. Las fracciones que contienen las proteínas de fusión se agrupan y se dializan contra agua desionizada. El eluato es sometido a SDS-PAGE en condiciones reductoras. El rendimiento se determina como mg de proteína purificada a partir de 1 litro de cultivo de matraz agitado desarrollado hasta una DO₆₀₀ de 1.
- 15

Listado de secuencias

<110> Spiber Technologies AB

5 <120> PRODUCCIÓN DE PROTEÍNAS Y POLIPÉPTIDOS

<130> PC-21046889

<150> EP 10156927

10 <151> 18-03-2010

<150> PCT/SE2010/050439

<151> 21-04-2010

15 <160> 49

<170> PatentIn versión 3.5

<210> 1

20 <211> 149

<212> PRT

<213> Euprostenops australis

<400> 1

25 Gly Ser Gly Asn Ser Gly Ile Gln Gly Gln Gly Gly Tyr Gly Gly Leu
 1 5 10 15

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

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

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

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

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

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

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

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

Gly Tyr Gly Gln Ser
 145

<210> 2

30 <211> 265

<212> PRT

<213> Euprostenops australis

ES 2 608 673 T3

<220>
 <221> DOMINIO
 <222> (1)..(167)
 <223> fragmento REP

5

<220>
 <221> DOMINIO
 <222> (168)..(265)
 <223> fragmento CT

10

<400> 2

Gly Ser Gly Asn Ser Gly Ile Gln Gly Gln Gly Gly Tyr Gly Gly Leu
 1 5 10 15

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

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

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

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

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

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

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

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

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

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

Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val
 180 185 190

ES 2 608 673 T3

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

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

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

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

Ala Asn Ala Met Ala Gln Val Met Gly
 260 265

<210> 3
 <211> 296
 5 <212> PRT
 <213> Euprostenops australis

<220>
 <221> DOMINIO
 10 <222> (1)..(137)
 <223> fragmento NT

<220>
 <221> DOMINIO
 15 <222> (138)..(296)
 <223> fragmento REP

<400> 3

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
 1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
 20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
 35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
 50 55 60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
 65 70 75 80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
 85 90 95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
 100 105 110

20

ES 2 608 673 T3

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
 115 120 125

Gln Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser
 130 135 140

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

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

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly
 180 185 190

Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala
 195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly
 210 215 220

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

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

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

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

Gly Gln Gly Gly Tyr Gly Gln Ser
 290 295

- <210> 4
- <211> 340
- 5 <212> PRT
- <213> Euprostenops australis

- <220>
- <221> DOMINIO
- 10 <222> (1)..(137)
- <223> fragmento NT

- <220>
- <221> DOMINIO
- 15 <222> (138)..(340)
- <223> fragmento REP

- <400> 4

ES 2 608 673 T3

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
 1 5 10 15
 Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
 20 25 30
 Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
 35 40 45
 Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
 50 55 60
 Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
 65 70 75 80
 Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
 85 90 95
 Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
 100 105 110
 Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
 115 120 125
 Gln Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser
 130 135 140
 Ala Ala Ala Ser Ala Gly Ala Pro Gly Tyr Ser Pro Ala Pro Ser Tyr
 145 150 155 160
 Ser Ser Gly Gly Tyr Ala Ser Ser Ala Ala Ser Ala Ala Ala Ala Ala
 165 170 175
 Gly Gln Gly Gly Pro Gly Gly Tyr Gly Pro Ala Pro Asn Gln Gly Ala
 180 185 190
 Ser Ser Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Pro Ser Gly
 195 200 205
 Pro Tyr Gly Thr Ser Tyr Gln Ile Ser Thr Gln Tyr Thr Gln Thr Thr
 210 215 220
 Thr Ser Gln Gly Gln Gly Tyr Gly Ser Ser Ser Ala Gly Ala Ala Ala
 225 230 235 240
 Ala Gly Ala Ala Gly Ala Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln
 245 250 255

ES 2 608 673 T3

Gly Gly Tyr Gly Gln Gly Ala Gly Gly Ala Ala Ala Ala Ala Ala Ala
 260 265 270

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

Gly Tyr Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly
 290 295 300

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 305 310 315 320

Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Pro
 325 330 335

Gly Ser Gly Gly
 340

5 <210> 5
 <211> 424
 <212> PRT
 <213> Euprostenops australis

10 <220>
 <221> DOMINIO
 <222> (1)..(136)
 <223> fragmento NT

15 <220>
 <221> DOMINIO
 <222> (137)..(313)
 <223> fragmento REP

20 <220>
 <221> DOMINIO
 <222> (314)..(411)
 <223> fragmento CT

25 <220>
 <221> DOMINIO
 <222> (412)..(424)
 <223> His tag

<400> 5

Met Lys Ala Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
 1 5 10 15

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
 20 25 30

30 Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
 35 40 45

ES 2 608 673 T3

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
50 55 60

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
65 70 75 80

Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
85 90 95

Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
100 105 110

Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
115 120 125

Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
130 135 140

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

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

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
180 185 190

Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
210 215 220

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

Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
245 250 255

Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
260 265 270

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

Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
290 295 300

ES 2 608 673 T3

Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
305 310 315 320

Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
325 330 335

Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
340 345 350

Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
355 360 365

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
370 375 380

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
385 390 395 400

Val Val Ala Asn Ala Met Ala Gln Val Met Gly Lys Leu Ala Ala Ala
405 410 415

Leu Glu His His His His His His
420

<210> 6

<211> 137

5 <212> PRT

<213> Euprostenops australis

<220>

<221> VARIANTE

10 <222> (6)..(6)

<223> deleción (deltaHis)

<400> 6

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50 55 60

15 Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65 70 75 80

ES 2 608 673 T3

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
85 90 95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
100 105 110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
115 120 125

Gln Ala Gly Met Asn Asp Val Ser Ala
130 135

<210> 7

<211> 98

<212> PRT

<213> Euprostenops australis

<400> 7

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

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

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

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

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

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

10 Met Gly

<210> 8

<211> 131

<212> PRT

15 <213> Secuencia Artificial

<220>

<223> Secuencia de consenso derivada de fragmentos de espidroína NT

20 <220>

<221> VARIANTE

<222> (20)..(20)

<223> Leu

25 <220>

<221> VARIANTE

<222> (42)..(42)

<223> Asn

30 <220>

<221> VARIANTE

<222> (42)..(42)

ES 2 608 673 T3

<223> Gln

<220>
 <221> VARIANTE
 5 <222> (50)..(50)
 <223> Ser

<220>
 <221> VARIANTE
 10 <222> (50)..(50)
 <223> Lys

<220>
 <221> VARIANTE
 15 <222> (56)..(56)
 <223> Arg

<220>
 <221> VARIANTE
 20 <222> (84)..(84)
 <223> Leu

<220>
 <221> VARIANTE
 25 <222> (114)..(114)
 <223> Ser

<220>
 <221> VARIANTE
 30 <222> (121)..(121)
 <223> Asn

<220>
 <221> VARIANTE
 35 <222> (123)..(123)
 <223> Leu

<220>
 <221> VARIANTE
 40 <222> (124)..(124)
 <223> Ser

<400> 8

Gln Ala Asn Thr Pro Trp Ser Ser Pro Asn Leu Ala Asp Ala Phe Ile
 1 5 10 15

Asn Ser Phe Met Ser Ala Ala Ser Ser Ser Gly Ala Phe Ser Ala Asp
 20 25 30

Gln Leu Asp Asp Met Ser Thr Ile Gly Asp Thr Leu Met Ser Ala Met
 35 40 45

45 Asp Asn Met Gly Arg Ser Gly Lys Ser Thr Lys Ser Lys Leu Gln Ala

ES 2 608 673 T3

50 55 60

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ala Glu
65 70 75 80

Ser Gly Gly Gly Ser Val Gly Val Lys Thr Asn Ala Ile Ser Asp Ala
85 90 95

Leu Ser Ser Ala Phe Tyr Gln Thr Thr Gly Ser Val Asn Pro Gln Phe
100 105 110

Val Asn Glu Ile Arg Ser Leu Ile Gly Met Phe Ala Gln Ala Ser Ala
115 120 125

Asn Glu Val
130

5 <210> 9
<211> 100
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Secuencia de consenso derivada de proteínas conocidas MaSp1 y MaSp2

15 <220>
<221> CARACTERÍSTICA_MISC
<222> (1)..(71)
<223> Longitud de secuencia presente en variantes de especies conocidas

20 <220>
<221> VARIANTE
<222> (7)..(7)
<223> Glu

<400> 9

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

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

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

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

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

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

25 Leu Gly Glu Phe
100

<210> 10

- <211> 1110
- <212> PRT
- <213> Euprostenops australis

- 5 <220>
- <221> REPETICIÓN
- <222> (7)..(19)

- 10 <220>
- <221> REPETICIÓN
- <222> (20)..(42)

- 15 <220>
- <221> REPETICIÓN
- <222> (43)..(56)

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

- 25 <220>
- <221> REPETICIÓN
- <222> (84)..(106)

- 30 <220>
- <221> REPETICIÓN
- <222> (107)..(120)

- 35 <220>
- <221> REPETICIÓN
- <222> (121)..(134)

- 40 <220>
- <221> REPETICIÓN
- <222> (135)..(147)

- 45 <220>
- <221> REPETICIÓN
- <222> (148)..(170)

- 50 <220>
- <221> REPETICIÓN
- <222> (171)..(183)

- 55 <220>
- <221> REPETICIÓN
- <222> (184)..(197)

- 60 <220>
- <221> REPETICIÓN
- <222> (198)..(211)

- 65 <220>
- <221> REPETICIÓN
- <222> (212)..(234)

- 65 <220>
- <221> REPETICIÓN
- <222> (235)..(248)

	<222> (249)..(265)
	<220>
5	<221> REPETICIÓN <222> (266)..(279)
	<220>
10	<221> REPETICIÓN <222> (280)..(293)
	<220>
	<221> REPETICIÓN <222> (294)..(306)
15	<220>
	<221> REPETICIÓN <222> (307)..(329)
20	<220>
	<221> REPETICIÓN <222> (330)..(342)
25	<220>
	<221> REPETICIÓN <222> (343)..(356)
30	<220>
	<221> REPETICIÓN <222> (357)..(370)
	<220>
	<221> REPETICIÓN <222> (371)..(393)
35	<220>
	<221> REPETICIÓN <222> (394)..(406)
40	<220>
	<221> REPETICIÓN <222> (407)..(420)
45	<220>
	<221> REPETICIÓN <222> (421)..(434)
50	<220>
	<221> REPETICIÓN <222> (435)..(457)
	<220>
	<221> REPETICIÓN <222> (458)..(470)
55	<220>
	<221> REPETICIÓN <222> (471)..(488)
60	<220>
	<221> REPETICIÓN <222> (489)..(502)
65	<220>
	<221> REPETICIÓN <222> (503)..(516)

- <220>
<221> REPETICIÓN
<222> (517)..(529)
- 5 <220>
<221> REPETICIÓN
<222> (530)..(552)
- 10 <220>
<221> REPETICIÓN
<222> (553)..(566)
- 15 <220>
<221> REPETICIÓN
<222> (567)..(580)
- 20 <220>
<221> REPETICIÓN
<222> (581)..(594)
- 25 <220>
<221> REPETICIÓN
<222> (595)..(617)
- 30 <220>
<221> REPETICIÓN
<222> (618)..(630)
- 35 <220>
<221> REPETICIÓN
<222> (631)..(647)
- 40 <220>
<221> REPETICIÓN
<222> (648)..(661)
- 45 <220>
<221> REPETICIÓN
<222> (662)..(675)
- 50 <220>
<221> REPETICIÓN
<222> (676)..(688)
- 55 <220>
<221> REPETICIÓN
<222> (689)..(711)
- 60 <220>
<221> REPETICIÓN
<222> (712)..(725)
- 65 <220>
<221> REPETICIÓN
<222> (726)..(739)
- 70 <220>
<221> REPETICIÓN
<222> (740)..(752)
- 75 <220>
<221> REPETICIÓN
<222> (753)..(775)

<222> (776)..(789)

<220>
<221> REPETICIÓN
5 <222> (790)..(803)

<220>
<221> REPETICIÓN
10 <222> (804)..(816)

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

15 <220>
<221> REPETICIÓN
<222> (840)..(853)

20 <220>
<221> REPETICIÓN
<222> (854)..(867)

25 <220>
<221> REPETICIÓN
<222> (868)..(880)

30 <220>
<221> REPETICIÓN
<222> (881)..(903)

35 <220>
<221> REPETICIÓN
<222> (904)..(917)

40 <220>
<221> REPETICIÓN
<222> (918)..(931)

45 <220>
<221> REPETICIÓN
<222> (932)..(945)

50 <220>
<221> REPETICIÓN
<222> (946)..(968)

55 <220>
<221> REPETICIÓN
<222> (969)..(981)

60 <220>
<221> REPETICIÓN
<222> (982)..(998)

65 <220>
<221> REPETICIÓN
<222> (999)..(1013)

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

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

ES 2 608 673 T3

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

5 <220>
 <221> REPETICIÓN
 <222> (1060)..(1073)

10 <220>
 <221> REPETICIÓN
 <222> (1074)..(1092)

<400> 10

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

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

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

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

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

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

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

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

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

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

ES 2 608 673 T3

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 165 170 175
 Ser Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln
 180 185 190
 Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 195 200 205
 Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
 210 215 220
 Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 225 230 235 240
 Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly
 245 250 255
 Arg Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 260 265 270
 Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln
 275 280 285
 Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 290 295 300
 Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly
 305 310 315 320
 Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 325 330 335
 Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 340 345 350
 Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Glu Ala Ala
 355 360 365
 Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
 370 375 380
 Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 385 390 395 400
 Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 405 410 415

ES 2 608 673 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ES 2 608 673 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

ES 2 608 673 T3

930

935

940

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

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

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

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

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

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

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

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

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

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

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

<210> 11

<211> 23

5 <212> PRT

<213> Secuencia Artificial

<220>

10 <223> Consensus sequence derived from internal REPETICIÓNs of Euprostenops australis MaSp1

<220>

<221> VARIANTE

<222> (4)..(4)

<223> Ser

15

<220>

<221> VARIANTE

<222> (8)..(8)

<223> Tyr

20

<220>

<221> VARIANTE

<222> (11)..(11)

<223> Gln

25

<400> 11

ES 2 608 673 T3

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

Gly Gln Gly Ala Gly Ser Ser
 20

5 <210> 12
 <211> 17
 <212> PRT
 <213> Secuencia Artificial

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

<220>
 <221> VARIANTE
 <222> (9)..(9)
 15 <223> Arg

<220>
 <221> VARIANTE
 <222> (14)..(14)
 20 <223> Ser

<220>
 <221> VARIANTE
 <222> (16)..(16)
 25 <223> Gly

<400> 12

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

30 **Ser**
 <210> 13
 <211> 14
 <212> PRT
 <213> Secuencia Artificial

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

40 <220>
 <221> VARIANTE
 <222> (2)..(2)
 <223> Gln

45 <220>
 <221> VARIANTE
 <222> (6)..(6)
 <223> Arg

50 <220>
 <221> VARIANTE
 <222> (11)..(11)
 <223> Ser

55 <220>
 <221> VARIANTE
 <222> (11)..(11)
 <223> Val

ES 2 608 673 T3

atgaaagcat cacacactac accatggaca aaccaggac tcgcagaaaa cttcatgaac 60
 agtttcatgc aaggcctgag ctcgatgcca ggtttcacgg caagccaatt ggatgatatg 120
 tcaaccatcg cacaatccat ggtacagtca atacaatcct tggcggcaca aggcaggaca 180
 tcaccgaata agctgcaggc ccttaacatg gcttttgcac cttcgatggc agaaatcgcg 240
 gcatccgaag aaggaggggg aagcctttcc accaaaacta gctctatagc cagtgcaatg 300
 tccaacgctg ttctgcaaac aactggagtg gtaaaccaac cgttcataaa tgaataact 360
 cagctcgta gcatgtttgc tcaagcaggt atgaatgatg tcagtgttc cgcacagca 420
 ggagcatccg cagcagcatc cgcaggagcg gctagcggtc aaggaggata tggaggacta 480
 ggtcaaggag gatatggaca aggtgcagga agttctgcag ccgctgccgc cgcgcagca 540
 gccgcgcag caggtggaca aggtggacaa ggtcaaggag gatatggaca aggttcagga 600
 ggttctgcag ccgcccgcgc cgcgcagca gcagcagcag ctgcagcagc tggacgaggt 660
 caaggaggat atggtcaagg ttctggaggt aatgctgctg ccgcagccgc tgcgcgcgc 720
 gccgcgcctg cagcagccgc acaggaggt caaggaggat atggtagaca aagccaaggt 780
 gctggttccg ctgctgctgc tgetgctgct gctgccgctg ctgctgctgc aggatctgga 840
 caaggaggat acggaggaca aggtcaagga ggttatggtc agagttagtc ttctgcttca 900
 gctgctgctg cagctgctag tactgtagct aattcgggtga gtcgcctctc atgccttcc 960
 gcagtatctc gagtttcttc agcagtttct agcttggttt caaatggtca agtgaatag 1020
 gcagcgttac ctaatatcat ttccaacatt tcttctctg tcagtgcac tgcctctggt 1080
 gcttctggat gtgaggtcat agtgaagct ctactcgaag tcatcactgc tcttgttcaa 1140
 atcgtagtt cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat 1200
 gttgttgcta atgcatggc tcaagtaatg ggcaagcttg cggccgcact cgagcaccac 1260
 caccaccac ac 1272

<210> 16
 <211> 1020
 <212> ADN
 <213> Euprostenops australis

5

<400> 16

ggttctggga attcacacac tacaccatgg acaaaccag gactcgcaga aaacttcatg 60
 aacagtttca tgcaaggcct gagctcgatg ccaggtttca cggcaagcca attggatgat 120
 atgtcaacca tcgcacaatc catggtacag tcaatacaat ccttggcggc acaaggcagg 180
 acatcaccga ataagctgca ggccctaac atggcttttg catcttcgat ggcagaaatc 240
 gcggcatccg aagaaggagg ggggaagcctt tccacaaaa ctagctctat agccagtgca 300
 atgtccaacg cgtttctgca aacaactgga gtggtaaacc aaccgttcat aatgaaata 360
 actcagctcg ttagcatggt tgctcaagca ggtatgaatg atgtcagtgc ttccgcatca 420

10

ES 2 608 673 T3

gcaggagcat ccgcagcagc atccgcagga gcgccagggt acagtccctgc accaagctac 480
 agttcgggag gttatgcttc aagtgctgcc tcagcagccg ctgcagcagg acaaggagga 540
 cctgggggat acggtccagc acctaacca ggagcttcat ctgccgctgc tgcagccgca 600
 ggatcaggac aaggaccatc aggaccgtac ggtacatctt accagataag tacacaatat 660
 actcaaacia cgacttcaca gggacaagga tatgggtcaa gtagcgctgg agccgcagct 720
 gcaggcgctg caggtgctgg acaagggggc tacggaggtc aaggtcaagg aggatatggt 780
 caaggagccg gaggtgctgc cgcagcagcc gccgctgccg cagccgctgc cgccgcagcc 840
 ggacaagggtg gacaagggtg aggaggatat ggacaaggag gacaaggagg acaaggagga 900
 caaggtcaag gaggatatgg acaaggtgca ggaagtcttg cagccgccgc cgccgcagca 960
 gcagcagccg ccgcagcagc aggacgaggt caaggaggat atggtccaggt ttctggaggt 1020

<210> 17

<211> 422

<212> PRT

<213> Euprostenops australis

<400> 17

Met Lys Ala Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
 1 5 10 15

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
 20 25 30

Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
 35 40 45

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
 50 55 60

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
 65 70 75 80

Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
 85 90 95

Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
 100 105 110

Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
 115 120 125

Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
 130 135 140

5

10

ES 2 608 673 T3

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

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

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
 180 185 190

Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
 195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
 210 215 220

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

Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
 245 250 255

Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
 260 265 270

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

Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 290 295 300

Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 305 310 315 320

Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 325 330 335

Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 340 345 350

Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 355 360 365

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
 370 375 380

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
 385 390 395 400

Val Val Ala Asn Ala Met Ala Gln Val Met Gly Ala Ala Ala Leu Glu
 405 410 415

His His His His His His
 420

- <210> 18
- 5 <211> 1266
- <212> ADN
- <213> Euprostenops australis

ES 2 608 673 T3

<400> 18

atgaaagcat cacacactac accatggaca aaccaggac tcgcagaaaa cttcatgaac	60
agtttcatgc aaggcctgag ctcgatgccca ggtttcacgg caagccaatt ggatgatatg	120
tcaaccatcg cacaatccat ggtacagtca atacaatcct tggcggcaca aggcaggaca	180
tcaccgaata agctgcaggc ccttaacatg gcttttgcac cttcgatggc agaaatcgcg	240
gcatccgaag aaggaggggg aagcctttcc accaaaacta gctctatagc cagtgcaatg	300
tccaacgcgt ttctgcaaac aactggagtg gtaaaccaac cgttcataaa tgaaataact	360
cagctcgta gcatgtttgc tcaagcaggt atgaatgatg tcagtgcctc cgcacagca	420
ggagcatccg cagcagcacc cgcaggagcg gctagcggtc aaggtagata tggtagacta	480
ggtcaaggag gatatggaca aggtgcagga agttctgcag ccgctgccgc cgcgcagca	540
gccgccgcag caggtggaca aggtggacaa ggtcaaggag gatatggaca aggttcagga	600
ggttctgcag ccgccgccgc cgcgcagca gcagcagcag ctgcagcagc tggacgaggt	660
caaggaggat atggtcaagg ttctggaggt aatgctgctg ccgcagccgc tgccgccgcc	720
gccgccgctg cagcagccgg acaggaggt caaggtagat atggttagaca aagccaaggt	780
gctggttccg ctgctgctgc tgctgctgct gctgccgctg ctgctgctgc aggatctgga	840
caaggtagat acggtggaca aggtcaagga ggttatggtc agagtagtgc ttctgcttca	900
gctgctgctg cagctgctag tactgtagct aattcggtag gtgcgctctc atcgcttcc	960
gcagtatctc gagtttcttc agcagtttct agcttggttt caaatggtea agtgaatatg	1020
gcagcgttac ctaatatcat ttccaacatt tcttcttctg tcagtgcacc tgctcctggt	1080
gcttctggat gtgaggtcat agtgcaagct ctactogaag tcatcactgc tcttgttcaa	1140
atcgtagtt cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat	1200
gttgttgcta atgccatggc tcaagtaatg ggcgcggccg cactcgagca ccaccaccac	1260
caccac	1266

- 5 <210> 19
- <211> 491
- <212> PRT
- <213> Euprostenops australis

10 <400> 19

ES 2 608 673 T3

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

Asn Lys Phe Asn Lys Glu Arg Arg Arg Ala Arg Arg Glu Ile Arg His
 20 25 30

Leu Pro Asn Leu Asn Arg Glu Gln Arg Arg Ala Phe Ile Arg Ser Leu
 35 40 45

Arg Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys
 50 55 60

Leu Asn Asp Ala Gln Ala Pro Lys Pro Asn Leu Glu Ala Leu Phe Gln
 65 70 75 80

Gly Pro Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
 85 90 95

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
 100 105 110

Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
 115 120 125

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
 130 135 140

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
 145 150 155 160

Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
 165 170 175

Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
 180 185 190

Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
 195 200 205

Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
 210 215 220

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

Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
 245 250 255

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln

ES 2 608 673 T3

260 265 270

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

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
 290 295 300

Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
 305 310 315 320

Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
 325 330 335

Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
 340 345 350

Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly
 355 360 365

Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 370 375 380

Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 385 390 395 400

Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 405 410 415

Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 420 425 430

Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 435 440 445

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
 450 455 460

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
 465 470 475 480

Val Val Ala Asn Ala Met Ala Gln Val Met Gly
 485 490

<210> 20

<211> 411

5 <212> PRT

<213> Euprostenops australis

<400> 20

ES 2 608 673 T3

Gly Pro Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
 1 5 10 15
 Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
 20 25 30
 Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
 35 40 45
 Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
 50 55 60
 Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
 65 70 75 80
 Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
 85 90 95
 Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
 100 105 110
 Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
 115 120 125
 Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
 130 135 140
 Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu
 145 150 155 160
 Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
 165 170 175
 Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
 180 185 190
 Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
 195 200 205
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
 210 215 220
 Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala
 225 230 235 240
 Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
 245 250 255

ES 2 608 673 T3

Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
 260 265 270

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

Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 290 295 300

Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 305 310 315 320

Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 325 330 335

Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 340 345 350

Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 355 360 365

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
 370 375 380

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
 385 390 395 400

Val Val Ala Asn Ala Met Ala Gln Val Met Gly
 405 410

<210> 21

<211> 551

5 <212> PRT

<213> Euprostenops australis

<400> 21

Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
 1 5 10 15

His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 20 25 30

Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
 35 40 45

Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
 50 55 60

10 Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
 65 70 75 80

ES 2 608 673 T3

Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
 85 90 95
 Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
 100 105 110
 Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly Ser Gly His Met His His
 115 120 125
 His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser Gly Asn Ser
 130 135 140
 His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn
 145 150 155 160
 Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln
 165 170 175
 Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln
 180 185 190
 Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu
 195 200 205
 Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu
 210 215 220
 Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met
 225 230 235 240
 Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile
 245 250 255
 Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn
 260 265 270
 Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala Ala Ala Ser Ala
 275 280 285
 Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly
 290 295 300
 Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 305 310 315 320
 Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Tyr Gly
 325 330 335

ES 2 608 673 T3

Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 340 345 350

Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
 355 360 365

Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 370 375 380

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

Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 405 410 415

Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr
 420 425 430

Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr
 435 440 445

Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg
 450 455 460

Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met
 465 470 475 480

Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala
 485 490 495

Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu
 500 505 510

Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly
 515 520 525

Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn
 530 535 540

Ala Met Ala Gln Val Met Gly
 545 550

<210> 22

<211> 412

5 <212> PRT

<213> Euprostenops australis

<400> 22

10 Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala

ES 2 608 673 T3

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

Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala
 290 295 300

Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro
 305 310 315 320

Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn
 325 330 335

Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser
 340 345 350

Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile
 355 360 365

Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser
 370 375 380

Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr
 385 390 395 400

Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
 405 410

<210> 23

<211> 534

5 <212> PRT

<213> Euprostenops australis

<400> 23

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
 1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
 20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
 35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
 50 55 60

10 Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 65 70 75 80

ES 2 608 673 T3

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180 185 190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
245 250 255

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
260 265 270

Gly Asn Ser Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr
275 280 285

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
290 295 300

Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Tyr Gly Gln
305 310 315 320

Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
325 330 335

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

ES 2 608 673 T3

agcatggcgg aaattgcggc gagcgaagaa ggcggcggca gcctgagcac caaaaccagc 300
 agcattgcga gcgcgatgag caacgcgttt ctgcagacca ccggcgtggt gaaccagccg 360
 tttattaacg aaattaccca gctggtgagc atgtttgcgc aggccggtat gaatgacggc 420
 ggcggcacc cgtggaccaa cccgggcctg gcggaact ttatgaacag ctttatgcag 480
 ggcctgagca gcatgccggg ctttaccgag agccagctgg atgatatgag caccattgag 540
 cagagcatgg tgcagagcat tcagagcctg gcggcgcagg gccgtaccag cccgaacaaa 600
 ctgcagggc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa 660
 ggcggcggca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt 720
 ctgcagacca ccggcgtggt gaaccagccg tttattaacg aaattaccca gctggtgagc 780
 atgtttgcgc aggcgggcat gaacgatgtg agcgcgggga attcagggtca aggtgatata 840
 ggtggactag gtcaaggagg atatggacaa ggtgcaggaa gttctgcagc cgctgccgcc 900
 gccgcagcag ccgccgcagc aggtggacaa ggtggacaag gtcaaggagg atatggacaa 960
 ggttcaggag gttctgcagc cgccgccgcc gccgcagcag cagcagcagc tgcagcagct 1020
 ggacgaggtc aaggaggata tggtaagggt tctggaggta atgctgctgc cgcagccgct 1080
 gccgccgccg ccgccgctgc agcagccgga caggagggtc aagggtgata tggtagacaa 1140
 agccaagggt ctggttccgc tgctgctgct gctgctgctg ctgccgctgc tgctgctgca 1200
 ggatctggac aagggtgata cgggtggacaa ggtcaaggag gttatggtca gagtagtct 1260
 tctgcttcag ctgctgcgct agctgctagt actgtagcta attcgggtgag tcgcctctca 1320
 tcgccttccg cagtatctcg agtttcttca gcagtttcta gcttggtttc aaatggtcaa 1380
 gtgaatatgg cagcgttacc taatatcatt tccaacattt cttcttctgt cagtgcattc 1440
 gctcctggtg cttctggatg tgaggtcata gtgcaagctc tactcgaagt catcactgct 1500
 cttgttcaaa tcgttagttc ttctagtgtt ggatatatta atccatctgc tgtgaaccaa 1560
 attactaatg ttgttgctaa tgccatggct caagtaatgg gc 1602

<210> 25

<211> 514

5 <212> PRT

<213> Euprostenops australis

<400> 25

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
 1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
 20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
 35 40 45

10

ES 2 608 673 T3

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50 55 60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65 70 75 80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
85 90 95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
100 105 110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
115 120 125

Gln Ala Gly Met Asn Asp Val Ser Ala Gly Tyr Gly Gln Gly Ala Gly
130 135 140

Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
145 150 155 160

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

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Gln Gly Ser Gln
180 185 190

Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly
195 200 205

Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
210 215 220

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn Ala Ala
225 230 235 240

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

Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala
260 265 270

Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly
275 280 285

Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly
290 295 300

ES 2 608 673 T3

Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
305 310 315 320

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

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly
340 345 350

Gly Gln Gly Gly Tyr Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala
355 360 365

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln
370 375 380

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

Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val
405 410 415

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
420 425 430

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
435 440 445

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
450 455 460

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
465 470 475 480

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
485 490 495

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
500 505 510

Met Gly

<210> 26

<211> 176

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Proteína de fusión

10

<400> 26

ES 2 608 673 T3

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
 1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
 20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
 35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
 50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
 85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
 100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
 115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Met Ile
 130 135 140

Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu Leu Leu
 145 150 155 160

Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Leu Gly Leu
 165 170 175

<210> 27
 <211> 528
 5 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

10 <400> 27

atgggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg 60

gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgcg 120

agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg 180

gcgggcgcagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc 240

agcatggcgg aaattgcgcc gagcgaagaa ggcggcggca gctgagcac caaaaccagc 300

agcattgcga gcgcgatgag caacgcgttt ctgcagacca ccggcgtggt gaaccagccg 360

tttattaacg aaattaccca gctggtgagc atgtttgcgc aggcgggcat gaacgatgtg 420

agcgcgatga ttccgagcag cccggtgcat ctgaaacgcc tgaaactgct gctgctgctg 480

15 ctgctgctga ttctgctgct gattctgggc gcgctgctgc tgggcctg 528

ES 2 608 673 T3

<210> 28
 <211> 309
 <212> PRT
 <213> Secuencia Artificial

5

<220>
 <223> Proteína de fusión

<400> 28

10

```

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1          5          10          15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20          25          30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35          40          45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50          55          60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65          70          75          80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85          90          95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100         105         110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115         120         125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130         135         140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145         150         155         160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165         170         175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180         185         190
    
```

ES 2 608 673 T3

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
 195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
 225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
 245 250 255

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 260 265 270

Gly Asn Ser Met Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys
 275 280 285

Leu Leu Leu Leu Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala
 290 295 300

Leu Leu Leu Gly Leu
 305

<210> 29
 <211> 927
 5 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

10 <400> 29

```

atgggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg      60
gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgcg      120
agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg      180
gcgggcgagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc      240
agcatggcgg aaattgcggc gagcgaagaa ggcggcggca gcctgagcac caaaaccagc      300
agcattgcga gcgcgatgag caacgcgttt ctgcagacca cggcgtggt gaaccagccg      360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggccggtat gaatgacggc      420
ggcggcacc cgtggaccaa cccgggcctg gcggaaaact ttatgaacag ctttatgcag      480
ggcctgagca gcatgccggg ctttaccgcg agccagctgg atgatatgag caccattgcg      540
cagagcatgg tgcagagcat tcagagcctg gcggcgcagg gccgtaccag cccgaacaaa      600
ctgcaggcgc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa      660
ggcggcggca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt      720
    
```

ES 2 608 673 T3

ctgcagacca ccggcgtggt gaaccagccg tttattaacg aaattaccca gctggtgagc 780
atgtttgcgc aggcggggcat gaacgatgtg agcgcgggga attctatgat tccgagcagc 840
ccggtgcatac tgaaacgcct gaaactgctg ctgctgctgc tgctgctgat tctgctgctg 900
attctggggc cgctgctgct gggcctg 927

<210> 30
<211> 313
5 <212> PRT
<213> Secuencia Artificial

<220>
10 <223> Proteína de fusión

<400> 30

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala

ES 2 608 673 T3

180	185	190																				
Gln	Gly	Arg	Thr	Ser	Pro	Asn	Lys	Leu	Gln	Ala	Leu	Asn	Met	Ala	Phe							
		195					200					205										
Ala	Ser	Ser	Met	Ala	Glu	Ile	Ala	Ala	Ser	Glu	Glu	Gly	Gly	Gly	Ser							
	210					215					220											
Leu	Ser	Thr	Lys	Thr	Ser	Ser	Ile	Ala	Ser	Ala	Met	Ser	Asn	Ala	Phe							
225					230					235					240							
Leu	Gln	Thr	Thr	Gly	Val	Val	Asn	Gln	Pro	Phe	Ile	Asn	Glu	Ile	Thr							
				245					250					255								
Gln	Leu	Val	Ser	Met	Phe	Ala	Gln	Ala	Gly	Met	Asn	Asp	Val	Ser	Ala							
			260					265					270									
Gly	Asn	Ser	Met	Leu	Leu	Gly	Asp	Phe	Phe	Arg	Lys	Ser	Lys	Glu	Lys							
		275					280					285										
Ile	Gly	Lys	Glu	Phe	Lys	Arg	Ile	Val	Gln	Arg	Ile	Lys	Asp	Phe	Leu							
	290					295					300											
Arg	Asn	Leu	Val	Pro	Arg	Thr	Glu	Ser														
305					310																	

<210> 31
 <211> 939
 5 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

10 <400> 31

atgggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg	60
gcggaaaact ttatgaacag ctttatgcag ggctgagca gcatgccggg ctttaccgcg	120
agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg	180
gcggcgcagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc	240
agcatggcgg aaattgcggc gagcgaagaa ggcgccggca gcctgagcac caaaaccagc	300
agcattgoga gcgcgatgag caacgcgttt ctgcagacca ccggcgtggt gaaccagccg	360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggccggtat gaatgacggc	420
ggcggcaccc cgtggaccaa cccgggcctg gcggaaaact ttatgaacag ctttatgcag	480
ggcctgagca gcatgccggg ctttaccgcg agccagctgg atgatatgag caccattgcg	540
cagagcatgg tgcagagcat tcagagcctg gcggcgcagg gccgtaccag cccgaacaaa	600
ctgcaggcgc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa	660

ES 2 608 673 T3

ggcggcggca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt 720
 ctgcagacca ccggcgtggt gaaccagccg tttattaacg aaattaccca gctggtgagc 780
 atgtttgcgc aggcgggcat gaacgatgtg agcgcgggga attctatgct gctgggtgat 840
 ttcttccgca aatctaaaga gaagattggc aaagagttta aaagaattgt ccagagaatc 900
 aaggatTTTT tgcggaatct tgtaccacagc acagagtcc 939

<210> 32
 <211> 158
 5 <212> PRT
 <213> Euprostenops australis

<400> 32

Met Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe
 1 5 10 15

Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala
 20 25 30

Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser
 35 40 45

Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln
 50 55 60

Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser
 65 70 75 80

Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser
 85 90 95

Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro
 100 105 110

Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly
 115 120 125

Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala Ala Ala
 130 135 140

10 Ser Ala Gly Ala Ala Ala Leu Glu His His His His His His
 145 150 155

<210> 33
 <211> 646
 <212> PRT
 15 <213> Euprostenops australis

<400> 33

ES 2 608 673 T3

His His His His His His Ser His Thr Thr Pro Trp Thr Asn Pro Gly
 1 5 10 15
 Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met
 20 25 30
 Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln
 35 40 45
 Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser
 50 55 60
 Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala
 65 70 75 80
 Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr
 85 90 95
 Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly
 100 105 110
 Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met
 115 120 125
 Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro Trp Thr Asn
 130 135 140
 Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser
 145 150 155 160
 Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile
 165 170 175
 Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg
 180 185 190
 Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser
 195 200 205
 Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr
 210 215 220
 Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr
 225 230 235 240
 Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val
 245 250 255

ES 2 608 673 T3

Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Gly Tyr Gly
 260 265 270

Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 275 280 285

Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly
 290 295 300

Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly
 305 310 315 320

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

Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 340 345 350

Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly
 355 360 365

Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 370 375 380

Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr
 385 390 395 400

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 405 410 415

Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln
 420 425 430

Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 435 440 445

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

Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 465 470 475 480

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

Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 500 505 510

Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr Gly

ES 2 608 673 T3

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180 185 190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
245 250 255

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
260 265 270

Gly Asn Ser Glu His Leu Val Thr Thr Ala Thr Phe Ser Ile Gly Ser
275 280 285

Thr Gly Leu Val Val Tyr Asp Tyr Gln Gln Leu Leu Ile Ala Tyr Lys
290 295 300

Pro Ala Pro Gly Thr Cys Cys Tyr Ile Met Lys Ile Ala Pro Glu Ser
305 310 315 320

Ile Pro Ser Leu Glu Ala Leu Thr Arg Lys Val His Asn Phe Gln Met
325 330 335

Glu Cys Ser Leu Gln Ala Lys Pro Ala Val Pro Thr Ser Lys Leu Gly

ES 2 608 673 T3

340

345

350

Gln Ala Glu Gly Arg Asp Ala Gly Ser Ala Pro Ser Gly Gly Asp Pro
 355 360 365

Ala Phe Leu Gly Met Ala Val Ser Thr Leu Cys Gly Glu Val Pro Leu
 370 375 380

Tyr Tyr Ile
 385

<210> 35

<211> 178

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Proteína de fusión

10

<400> 35

Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
 1 5 10 15

His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 20 25 30

Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
 35 40 45

Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
 50 55 60

Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
 65 70 75 80

Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
 85 90 95

Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
 100 105 110

Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly Ser Gly His Met His His
 115 120 125

His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser Gly Asn Ser
 130 135 140

Met Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu
 145 150 155 160

Leu Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Leu
 165 170 175

15 Gly Leu

<210> 36

<211> 537

ES 2 608 673 T3

<212> ADN
<213> Secuencia Artificial

5 <220>
<223> Proteína de fusión

<400> 36

```

atgggccatc atcatcatca tcatatggct agcagcgata aaattattca cctgactgac      60
gacagttttg acacggatgt actcaaagcg gacggggcga tcctcgtcga tttctgggca      120
gagtgggtgcg gtcctgtgcaa aatgatcgcc ccgattctgg atgaaatcgc tgacgaatat      180
cagggcaaac tgaccgttgc aaaactgaac atcgatcaaa accctggcac tgcgccgaaa      240
tatggcatcc gtggtatccc gactctgctg ctgttcaaaa acggtgaagt ggcggcaacc      300
aaagtgggtg cactgtctaa aggtcagttg aaagagttcc tcgacgctaa cctggccggt      360
tctggttctg gccatatgca ccatcatcat catcattctt ctggtctggt gccacgcggt      420
tctgggaatt ctatgattcc gagcagcccg gtgcatctga aacgcctgaa actgctgctg      480
ctgctgctgc tgctgattct gctgctgatt ctgggcgcgc tgctgctggg cctgtaa      537
    
```

10 <210> 37
<211> 313
<212> PRT
<213> Secuencia Artificial

15 <220>
<223> Proteína de fusión

<400> 37

20 Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
1 5 10 15

His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
20 25 30

Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
35 40 45

Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
50 55 60

Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
65 70 75 80

ES 2 608 673 T3

Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
85 90 95

Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
100 105 110

Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly Ser Gly His Met His His
115 120 125

His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser Gly Asn Ser
130 135 140

His Met Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn
145 150 155 160

Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr
165 170 175

Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln
180 185 190

Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu
195 200 205

Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala
210 215 220

Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala
225 230 235 240

Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln
245 250 255

Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala
260 265 270

Gly Met Asn Asp Val Ser Ala Met Ile Pro Ser Ser Pro Val His Leu
275 280 285

Lys Arg Leu Lys Leu Leu Leu Leu Leu Leu Leu Ile Leu Leu Leu
290 295 300

Ile Leu Gly Ala Leu Leu Leu Gly Leu
305 310

<210> 38

<211> 939

5 <212> ADN

<213> Secuencia Artificial

<220>

<223> Proteína de fusión

10

<400> 38

ES 2 608 673 T3

atgggccatc atcatcatca tcatatggct agcagcgata aaattattca cctgactgac 60
gacagttttg acacggatgt actcaaagcg gacggggcga tcctcgtcga tttctgggca 120
gagtgggtcg gtccgtgcaa aatgatcgcc ccgattctgg atgaaatcgc tgacgaatat 180
cagggcaaac tgaccgttgc aaaactgaac atcgatcaaa accctggcac tgcgccgaaa 240
tatggcatcc gtggtatccc gactctgctg ctgttcaaaa acggtgaagt ggcggcaacc 300
aaagtgggtg cactgtctaa aggtcagttg aaagagttcc tcgacgctaa cctggccggt 360
tctggttctg gccatatgca ccatcatcat catcattctt ctggtctggt gccacgcggt 420
tctgggaatt cccatatgag ccataccacc ccgtggacca acccgggcct ggcggaaaac 480
tttatgaaca gctttatgca gggcctgagc agcatgccgg gctttaccgc gagccagctg 540
gatgatatga gcaccattgc gcagagcatg gtgcagagca ttcagagcct ggcggcgcag 600
ggccgcacca gccccaacia actgcaggcg ctgaacatgg cgtttgcgag cagcatggcg 660
gaaattgctg cgagcgaaga aggcggcggc agcctgagca ccaaaaccag cagcattgctg 720
agcgcgatga gcaacgcggt tctgcagacc accggcgtgg tgaaccagcc gtttattaac 780
gaaattaacc agctgggtgag catgtttgctg caggcgggca tgaacgatgt gagcgcgatg 840
attccgagca gcccggtgca tctgaaacgc ctgaaattgc tgetgctgtt actgctgctg 900
attctgcttc tgattctggg cgcgctgctg ctgggcctg 939

<210> 39

<211> 410

5 <212> PRT

<213> Secuencia Artificial

<220>

10 <223> Proteína de fusión

<400> 39

Met Gly His His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
1 5 10 15

Gly Ser His His His His His His Met Ser His Thr Thr Pro Trp Thr
20 25 30

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
35 40 45

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
50 55 60

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly

ES 2 608 673 T3

Met Lys Ile Ala Pro Glu Ser Ile Pro Ser Leu Glu Ala Leu Thr Arg
 340 345 350

Lys Val His Asn Phe Gln Met Glu Cys Ser Leu Gln Ala Lys Pro Ala
 355 360 365

Val Pro Thr Ser Lys Leu Gly Gln Ala Glu Gly Arg Asp Ala Gly Ser
 370 375 380

Ala Pro Ser Gly Gly Asp Pro Ala Phe Leu Gly Met Ala Val Ser Thr
 385 390 395 400

Leu Cys Gly Glu Val Pro Leu Tyr Tyr Ile
 405 410

<210> 40
 <211> 1233
 5 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

10 <400> 40

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcgggtg cagccatcat 60
 catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt 120
 atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat 180
 gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc ggcgcagggc 240
 cgtaccagcc cgaacaaact gcaggcgcctg aacatggcgt ttgcgagcag catggcggaa 300
 attgcggcga gcgaagaag cggcggcagc ctgagcacca aaaccagcag cattgcgagc 360
 gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa 420
 attaccagc tggtagcat gtttgcgcag gcggcatga acgatggcgg cggcaccocg 480
 tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc 540
 atgccgggct ttaccgcgag ccagctggat gatatgagca ccattgcgca gagcatggtg 600
 cagagcattc agagcctggc ggcgcagggc cgtaccagcc cgaacaaact gcaggcgcctg 660
 aacatggcgt ttgcgagcag catggcggaa attgcggcga gcgaagaag cggcggcagc 720
 ctgagcacca aaaccagcag cattgcgagc gcgatgagca acgcgtttct gcagaccacc 780
 ggcgtggtga accagccgtt tattaacgaa attaccagc tggtagcat gtttgcgcag 840
 gcgggatga acgatgtgag cgcgctggaa gcgctgttcc agggcccga ttcagagcac 900
 ctggttacca ctgccacctt ctccatcggc tccactggcc tcgtggtgta tgactaccag 960
 cagctgctga tcgcctacaa gccagcccct ggcacctgct gctacatcat gaagatagct 1020
 ccagagagca tcccagctct tgaggctctc actagaaaag tccacaactt ccagatggaa 1080
 tgctctctgc aggccaaagc cgcagtcct acgtctaagc tgggccagc agaggggcca 1140
 gatgcaggct cagcacctc cggaggggac ccggccttcc tgggcatggc cgtgagcacc 1200
 15 ctgtgtggcg aggtgccgct ctactacatc tag 1233

ES 2 608 673 T3

<210> 41
 <211> 116
 <212> PRT
 <213> Homo sapiens

5

<400> 41

Gly Pro Asn Ser Glu His Leu Val Thr Thr Ala Thr Phe Ser Ile Gly
 1 5 10 15
 Ser Thr Gly Leu Val Val Tyr Asp Tyr Gln Gln Leu Leu Ile Ala Tyr
 20 25 30
 Lys Pro Ala Pro Gly Thr Cys Cys Tyr Ile Met Lys Ile Ala Pro Glu
 35 40 45
 Ser Ile Pro Ser Leu Glu Ala Leu Thr Arg Lys Val His Asn Phe Gln
 50 55 60
 Met Glu Cys Ser Leu Gln Ala Lys Pro Ala Val Pro Thr Ser Lys Leu
 65 70 75 80
 Gly Gln Ala Glu Gly Arg Asp Ala Gly Ser Ala Pro Ser Gly Gly Asp
 85 90 95
 Pro Ala Phe Leu Gly Met Ala Val Ser Thr Leu Cys Gly Glu Val Pro
 100 105 110
 Leu Tyr Tyr Ile
 115

10

<210> 42
 <211> 537
 <212> PRT
 <213> Secuencia Artificial

15

<220>
 <223> Proteína de fusión

<400> 42

20

Met Gly His His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
 1 5 10 15
 Gly Ser His His His His His His Met Ser His Thr Thr Pro Trp Thr
 20 25 30

ES 2 608 673 T3

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
 35 40 45
 Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
 50 55 60
 Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
 65 70 75 80
 Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 85 90 95
 Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
 100 105 110
 Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
 115 120 125
 Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
 130 135 140
 Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
 145 150 155 160
 Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
 165 170 175
 Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
 180 185 190
 Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
 195 200 205
 Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
 210 215 220
 Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 225 230 235 240
 Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
 245 250 255
 Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
 260 265 270
 Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 275 280 285

ES 2 608 673 T3

Leu Glu Ala Leu Phe Gln Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe
290 295 300

Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly
305 310 315 320

His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly
325 330 335

Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro
340 345 350

Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser
355 360 365

Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met
370 375 380

Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly
385 390 395 400

Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val
405 410 415

Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile
420 425 430

Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile
435 440 445

Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
450 455 460

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
465 470 475 480

Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
485 490 495

Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
500 505 510

His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
515 520 525

Met Asp Glu Leu Tyr Lys Leu Ile Asn
530 535

<210> 43
<211> 1620
5 <212> ADN
<213> Secuencia Artificial

<220>
<223> Proteína de fusión

10 <400> 43

ES 2 608 673 T3

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcgggtg cagccatcat 60
catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt 120
atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat 180
gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc ggcgcagggc 240
cgtaccagcc cgaacaaact gcaggcgctg aacatggcgt ttgcgagcag catggcggaa 300
attgcggcga gcgaagaagg cggcggcagc ctgagcacca aaaccagcag cattgcgagc 360
gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa 420
attaccagc tggtgagcat gtttgcgagc gcgggcatga acgatggcgg cggcaccccg 480
tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc 540
atgccgggct ttaccgcgag ccagctggat gatatgagca ccattgcgca gagcatggtg 600
cagagcattc agagcctggc ggcgcagggc cgtaccagcc cgaacaaact gcaggcgctg 660
aacatggcgt ttgcgagcag catggcggaa attgcggcga gcgaagaagg cggcggcagc 720
ctgagcacca aaaccagcag cattgcgagc gcgatgagca acgcgtttct gcagaccacc 780
ggcgtggtga accagccgtt tattaacgaa attaccagc tggtgagcat gtttgcgagc 840
gcgggcatga acgatgtgag cgcgctggaa gcgctgttcc agggcccga ttcaaagggc 900
gaggagctgt tcaccggggt ggtgccatc ctggtcgagc tggacggcga cgtaaacggc 960
cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1020
aagttcatct gcaccaccgg caagctgcc gtgccctggc ccaccctcgt gaccaccctg 1080
acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1140
aagtcgccca tgcgggaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1200
aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1260
ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1320
tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaa 1380
ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 1440
aacaccccca tcggcgacgg ccccgctgctg ctgcccgaca accactacct gagcaccag 1500
tccgccctga gcaaagacc caacgagaag cgcgatcaca tggctcctgct ggagttcgtg 1560
accgccgccg ggatcactct cggcatggac gagctgtaca agttaattaa ctaatgataa 1620

<210> 44
<211> 243
<212> PRT
<213> Aequorea victoria

5

<400> 44

ES 2 608 673 T3

Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
 1 5 10 15

Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser
 20 25 30

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
 35 40 45

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr
 50 55 60

Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
 65 70 75 80

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln
 85 90 95

Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala
 100 105 110

Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys
 115 120 125

Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu
 130 135 140

Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys
 145 150 155 160

Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly
 165 170 175

Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp
 180 185 190

Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala
 195 200 205

Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
 210 215 220

Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235 240

Leu Ile Asn

- 5 <210> 45
- <211> 307
- <212> PRT
- <213> Secuencia Artificial
- 10 <220>
- <223> Proteína de fusión
- <400> 45

ES 2 608 673 T3

Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile
 1 5 10 15

Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln
 20 25 30

Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala
 35 40 45

Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys Leu Glu Ala Leu Phe Gln
 50 55 60

Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
 65 70 75 80

Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser
 85 90 95

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
 100 105 110

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr
 115 120 125

Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
 130 135 140

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln
 145 150 155 160

Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala
 165 170 175

Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys
 180 185 190

Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu
 195 200 205

ES 2 608 673 T3

Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys
 210 215 220

Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly
 225 230 235 240

Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp
 245 250 255

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

Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
 275 280 285

Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 290 295 300

Leu Ile Asn
 305

<210> 46
 <211> 368
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

<400> 46

His His His His His His Leu Ala Glu Ala Lys Val Leu Ala Asn Arg
 1 5 10 15

Glu Leu Asp Lys Tyr Gly Val Ser Asp Tyr His Lys Asn Leu Ile Asn
 20 25 30

Asn Ala Lys Thr Val Glu Gly Val Lys Asp Leu Gln Ala Gln Val Val
 35 40 45

Glu Ser Ala Lys Lys Ala Arg Ile Ser Glu Ala Thr Asp Gly Leu Ser
 50 55 60

Asp Phe Leu Lys Ser Gln Thr Pro Ala Glu Asp Thr Val Lys Ser Ile
 65 70 75 80

Glu Leu Ala Glu Ala Lys Val Leu Ala Asn Arg Glu Leu Asp Lys Tyr
 85 90 95

Gly Val Ser Asp Tyr Tyr Lys Asn Leu Ile Asn Asn Ala Lys Thr Val
 100 105 110

ES 2 608 673 T3

Glu Gly Val Lys Ala Leu Ile Asp Glu Ile Leu Ala Ala Leu Pro Asn
 115 120 125
 Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu
 130 135 140
 Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly
 145 150 155 160
 Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr
 165 170 175
 Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr
 180 185 190
 Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln His
 195 200 205
 Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr
 210 215 220
 Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys
 225 230 235 240
 Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp
 245 250 255
 Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr
 260 265 270
 Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile
 275 280 285
 Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln
 290 295 300
 Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val
 305 310 315 320
 Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys
 325 330 335
 Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr
 340 345 350
 Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Leu Ile Asn
 355 360 365

<210> 47
 <211> 692
 5 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Proteína de fusión

10 <400> 47

ES 2 608 673 T3

Met Gly His His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
 1 5 10 15

Gly Ser His His His His His His Met Ser His Thr Thr Pro Trp Thr
 20 25 30

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
 35 40 45

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
 50 55 60

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
 65 70 75 80

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 85 90 95

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
 100 105 110

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
 115 120 125

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
 130 135 140

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
 145 150 155 160

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
 165 170 175

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
 180 185 190

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
 195 200 205

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
 210 215 220

ES 2 608 673 T3

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 225 230 235 240

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
 245 250 255

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
 260 265 270

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 275 280 285

Leu Glu Ala Leu Phe Gln Gly Pro Asn Ser Thr Gly Ala Thr Ser Pro
 290 295 300

Glu Glu Ala Ile Ala Asp Leu Ser Val Asn Met Tyr Asn Arg Leu Arg
 305 310 315 320

Ala Thr Gly Glu Asp Glu Asn Ile Leu Phe Ser Pro Leu Ser Ile Ala
 325 330 335

Leu Ala Met Gly Met Met Glu Leu Gly Ala Gln Gly Ser Thr Gln Lys
 340 345 350

Glu Ile Arg His Ser Met Gly Tyr Asp Ser Leu Lys Asn Gly Glu Glu
 355 360 365

Phe Ser Phe Leu Lys Glu Phe Ser Asn Met Val Thr Ala Lys Glu Ser
 370 375 380

Gln Tyr Val Met Lys Ile Ala Asn Ser Leu Phe Val Gln Asn Gly Phe
 385 390 395 400

His Val Asn Glu Glu Phe Leu Gln Met Met Lys Lys Tyr Phe Asn Ala
 405 410 415

Ala Val Asn His Val Asp Phe Ser Gln Asn Val Ala Val Ala Asn Tyr
 420 425 430

Ile Asn Lys Trp Val Glu Asn Asn Thr Asn Asn Leu Val Lys Asp Leu
 435 440 445

Val Ser Pro Arg Asp Phe Asp Ala Ala Thr Tyr Leu Ala Leu Ile Asn
 450 455 460

Ala Val Tyr Phe Lys Gly Asn Trp Lys Ser Gln Phe Arg Pro Glu Asn
 465 470 475 480

ES 2 608 673 T3

Thr Arg Thr Phe Ser Phe Thr Lys Asp Asp Glu Ser Glu Val Gln Ile
 485 490 495

Pro Met Met Tyr Gln Gln Gly Glu Phe Tyr Tyr Gly Glu Phe Ser Asp
 500 505 510

Gly Ser Asn Glu Ala Gly Gly Ile Tyr Gln Val Leu Glu Ile Pro Tyr
 515 520 525

Glu Gly Asp Glu Ile Ser Met Met Leu Val Leu Ser Arg Gln Glu Val
 530 535 540

Pro Leu Ala Thr Leu Glu Pro Leu Val Lys Ala Gln Leu Val Glu Glu
 545 550 555 560

Trp Ala Asn Ser Val Lys Lys Gln Lys Val Glu Val Tyr Leu Pro Arg
 565 570 575

Phe Thr Val Glu Gln Glu Ile Asp Leu Lys Asp Val Leu Lys Ala Leu
 580 585 590

Gly Ile Thr Glu Ile Phe Ile Lys Asp Ala Asn Leu Thr Gly Leu Ser
 595 600 605

Asp Asn Lys Glu Ile Phe Leu Ser Lys Ala Ile His Lys Ser Phe Leu
 610 615 620

Glu Val Asn Glu Glu Gly Ser Glu Ala Ala Val Ser Gly Met Ile
 625 630 635 640

Ala Ile Ser Arg Met Ala Val Leu Tyr Pro Gln Val Ile Val Asp His
 645 650 655

Pro Phe Phe Phe Leu Ile Arg Asn Arg Arg Thr Gly Thr Ile Leu Phe
 660 665 670

Met Gly Arg Val Met His Pro Glu Thr Met Asn Thr Ser Gly His Asp
 675 680 685

Phe Glu Glu Leu
 690

<210> 48
 <211> 2085
 <212> ADN
 <213> Secuencia Artificial

5

<220>
 <223> Proteína de fusión

10

<400> 48

ES 2 608 673 T3

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcgggag cagccatcat 60
catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt 120
atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat 180
gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc ggcgcagggc 240
cgtaccagcc cgaacaaaact gcaggcgcctg aacatggcgt ttgcgagcag catggcggaa 300
attgcggcga gcgaagaagg cggcggcagc ctgagcacca aaaccagcag cattgcgagc 360
gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa 420
attaccagc tggtgagcat gtttgcgcag gcgggcatga acgatggcgg cggcaccacc 480
tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc 540
atgccgggct ttaccgcgag ccagctggat gatatgagca ccattgcgca gagcatggtg 600
cagagcattc agagcctggc ggcgcagggc cgtaccagcc cgaacaaaact gcaggcgcctg 660
aacatggcgt ttgcgagcag catggcggaa attgcggcga gcgaagaagg cggcggcagc 720
ctgagcacca aaaccagcag cattgcgagc gcgatgagca acgcgtttct gcagaccacc 780
ggcgtggtga accagccgtt tattaacgaa attaccagc tggtgagcat gtttgcgcag 840
gcgggcatga acgatgtgag cgcgtggaa gcgctgtcc agggcccga tcaacaggg 900
gccacttccc ctgaggaagc cattgctgac ttgtcagtga atatgtataa tcgtcttaga 960
gccactgggt aagatgaaaa tattctcttc totccattga gtattgctct tgcaatggga 1020
atgatggaac ttggggccca aggatctacc cagaaagaaa tccgccctc aatgggatat 1080
gacagcctaa aaaatggtga agaattttct ttcttgaagg agttttcaa catggtaact 1140
gctaaagaga gccaatatgt gatgaaaatt gccaatcct tgtttgtgca aaatggattt 1200
catgtcaatg aggagttttt gcaaatgatg aaaaaatatt ttaatgcagc agtaaatcat 1260
gtggacttca gtcaaatgt agccgtggcc aactacatca ataagtgggt ggagaataac 1320
acaaaacaatc tgggtgaaaga tttggtatcc ccaagggatt ttgatgctgc cacttatctg 1380
gccctcatta atgctgtcta tttcaagggg aactggaagt cgcagtttag gcctgaaaat 1440
actagaacct tttctttcac taaagatgat gaaagtgaag tccaaattcc aatgatgat 1500
cagcaaggag aatthtatta tggggaattt agtgatggct ccaatgaagc tggtggtatc 1560
taccagtcc tagaaatacc atatgaagga gatgaaataa gcatgatgct ggtgctgtcc 1620
agacaggaag ttctcttgc tactctggag ccattagtca aagcacagct ggttgaagaa 1680
tgggcaaaact ctgtgaagaa gcaaaaagta gaagtatacc tgcccagggt cacagtggaa 1740
caggaaattg atttaaaaga tgttttgaag gctcttggaa taactgaaat tttcatcaaa 1800
gatgcaaat tgacaggcct ctctgataat aaggagattt ttctttccaa agcaattcac 1860
aagtccttcc tagaggtaa tgaagaaggc tcagaagctg ctgctgtctc aggaatgatt 1920
gcaattagta ggtggctgt gctgtatcct caagttattg tcgaccatcc atttttcttt 1980
cttatcagaa acaggagaac tggtaacaatt ctattcatgg gacgagtcac gcatcctgaa 2040
acaatgaaca caagtggaca tgatttcgaa gaactttaat gataa 2085

- 5 <210> 49
- <211> 398
- <212> PRT
- <213> Homo sapiens

ES 2 608 673 T3

<400> 49

Gly Pro Asn Ser Thr Gly Ala Thr Ser Pro Glu Glu Ala Ile Ala Asp
 1 5 10 15

Leu Ser Val Asn Met Tyr Asn Arg Leu Arg Ala Thr Gly Glu Asp Glu
 20 25 30

Asn Ile Leu Phe Ser Pro Leu Ser Ile Ala Leu Ala Met Gly Met Met
 35 40 45

Glu Leu Gly Ala Gln Gly Ser Thr Gln Lys Glu Ile Arg His Ser Met
 50 55 60

Gly Tyr Asp Ser Leu Lys Asn Gly Glu Glu Phe Ser Phe Leu Lys Glu
 65 70 75 80

Phe Ser Asn Met Val Thr Ala Lys Glu Ser Gln Tyr Val Met Lys Ile
 85 90 95

Ala Asn Ser Leu Phe Val Gln Asn Gly Phe His Val Asn Glu Glu Phe
 100 105 110

Leu Gln Met Met Lys Lys Tyr Phe Asn Ala Ala Val Asn His Val Asp
 115 120 125

Phe Ser Gln Asn Val Ala Val Ala Asn Tyr Ile Asn Lys Trp Val Glu
 130 135 140

Asn Asn Thr Asn Asn Leu Val Lys Asp Leu Val Ser Pro Arg Asp Phe
 145 150 155 160

Asp Ala Ala Thr Tyr Leu Ala Leu Ile Asn Ala Val Tyr Phe Lys Gly
 165 170 175

Asn Trp Lys Ser Gln Phe Arg Pro Glu Asn Thr Arg Thr Phe Ser Phe
 180 185 190

Thr Lys Asp Asp Glu Ser Glu Val Gln Ile Pro Met Met Tyr Gln Gln
 195 200 205

5

ES 2 608 673 T3

Gly Glu Phe Tyr Tyr Gly Glu Phe Ser Asp Gly Ser Asn Glu Ala Gly
 210 215 220

Gly Ile Tyr Gln Val Leu Glu Ile Pro Tyr Glu Gly Asp Glu Ile Ser
 225 230 235 240

Met Met Leu Val Leu Ser Arg Gln Glu Val Pro Leu Ala Thr Leu Glu
 245 250 255

Pro Leu Val Lys Ala Gln Leu Val Glu Glu Trp Ala Asn Ser Val Lys
 260 265 270

Lys Gln Lys Val Glu Val Tyr Leu Pro Arg Phe Thr Val Glu Gln Glu
 275 280 285

Ile Asp Leu Lys Asp Val Leu Lys Ala Leu Gly Ile Thr Glu Ile Phe
 290 295 300

Ile Lys Asp Ala Asn Leu Thr Gly Leu Ser Asp Asn Lys Glu Ile Phe
 305 310 315 320

Leu Ser Lys Ala Ile His Lys Ser Phe Leu Glu Val Asn Glu Glu Gly
 325 330 335

Ser Glu Ala Ala Ala Val Ser Gly Met Ile Ala Ile Ser Arg Met Ala
 340 345 350

Val Leu Tyr Pro Gln Val Ile Val Asp His Pro Phe Phe Phe Leu Ile
 355 360 365

Arg Asn Arg Arg Thr Gly Thr Ile Leu Phe Met Gly Arg Val Met His
 370 375 380

Pro Glu Thr Met Asn Thr Ser Gly His Asp Phe Glu Glu Leu
 385 390 395

REIVINDICACIONES

1. Una proteína de fusión que comprende
- (i) al menos un resto potenciador de la solubilidad que deriva del fragmento N-terminal (NT) de una proteína de tela de araña; y
- 5 (ii) al menos un resto que es una proteína deseada no espidroína, o polipéptido, seleccionada del grupo que consiste en proteínas y polipéptidos formadores de amiloide, SP-B y variantes de la misma que contienen disulfuro, apolipoproteínas, proteínas y polipéptidos de membrana, fármacos de proteína y polipéptido, proteínas y polipéptidos propensos a la agregación, y proteasas,
- 10 en donde cada resto potenciador de la solubilidad está unido directa o indirectamente, con secuencias implicadas seleccionadas entre péptidos ligandos y/o restos potenciadores de la solubilidad adicionales, al resto de proteína o polipéptido deseado no espidroína, y en donde cada resto potenciador de la solubilidad tiene al menos un 80% de identidad con respecto a la SEQ ID NO 6, o al menos un 50% de identidad con respecto a la SEQ ID NO 8.
2. Una proteína de fusión según la reivindicación 1, en donde la proteína o polipéptido deseado no espidroína tiene menos de un 30% de identidad con respecto a cualquiera de las SEQ ID NO: 6-10.
- 15 3. Una proteína de fusión según cualquier reivindicación precedente, en donde cada resto potenciador de la solubilidad contiene entre 100 y 160 residuos de aminoácido.
4. Una proteína de fusión según cualquier reivindicación precedente, en donde la proteína de fusión comprende al menos dos restos potenciadores de la solubilidad, cada uno de los cuales deriva del fragmento N-terminal (NT) de una proteína de tela de araña.
- 20 5. Una proteína de fusión según cualquier reivindicación precedente, que además comprende
- (iii) al menos un sitio de división dispuesto entre al menos un resto de proteína o polipéptido deseado no espidroína y al menos un resto potenciador de la solubilidad.
- 25 6. Una proteína de fusión según cualquier reivindicación precedente, en donde la proteína o polipéptido deseado no espidroína se selecciona del grupo que consiste en A β -péptido, IAPP, PrP, α -sinucleína, calcitonina, prolactina, cistatina, ATF y actina; SP-B y variantes de la misma, mini-BLeu, α -defensinas y β -defensinas; apolipoproteínas de clase A-H; LL-37, SP-C y variantes de la misma, SP-C33, SP-C33Leu, Brichos, GFP, neuroserpina; hormonas, que incluyen EPO y hormona del crecimiento (HC), y factores de crecimiento, que incluyen IGF-I e IGF-II; avidina y estreptavidina; y proteasa 3C.
- 30 7. Una proteína de fusión según la reivindicación 6, en donde la proteína o polipéptido deseado no espidroína se selecciona de SP-B y variantes de la misma, y SP-C y variantes de la misma.
8. Una proteína de fusión según la reivindicación 7, en donde la proteína o polipéptido deseado no espidroína es mini-BLeu.
9. Una proteína de fusión según la reivindicación 7, en donde la proteína o polipéptido deseado no espidroína es SP-C33Leu.
- 35 10. Una proteína de fusión según cualquier reivindicación precedente, seleccionada del grupo que consiste en las SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 y 47; y de proteínas que tengan al menos un 80% de identidad con respecto a cualquiera de dichas proteínas.
11. Un ácido nucleico aislado seleccionado del grupo que consiste en ácidos nucleicos que codifican una proteína de fusión según cualquier reivindicación precedente y las SEQ ID NOS 27, 29, 31, 38, 40, 43 y 48.
- 40 12. El uso de al menos un resto que se deriva del fragmento N-terminal (NT) de una proteína de tela de araña como resto potenciador de la solubilidad en una proteína de fusión para la producción de una proteína o polipéptido deseado no espidroína; en donde cada resto potenciador de la solubilidad tiene una identidad de al menos el 80% con respecto a la SEQ ID NO 6, o una identidad de al menos el 50% con respecto a la SEQ ID NO 8.
- 45 13. Un método para la producción de una proteína de fusión o de una proteína o polipéptido deseado no espidroína, que comprende las siguientes etapas:
- a) expresar en un hospedante adecuado una proteína de fusión según una cualquiera de las reivindicaciones 1-10; y
- b) obtener una mezcla que contenga la proteína de fusión, y opcionalmente aislar la proteína de fusión.

14. Un método para la producción de una proteína o polipéptido deseado no espidroína según la reivindicación 13, en donde la proteína de fusión comprende además (iii) al menos un sitio de división dispuesto entre al menos un resto de proteína o polipéptido deseado no espidroína y al menos un resto potenciador de la solubilidad; y

en donde dicho método comprende además las siguientes etapas:

- 5 c) dividir la proteína de fusión para proporcionar la proteína o polipéptido deseado no espidroína; y
 d) aislar la proteína o polipéptido deseado no espidroína.

15. Método según una cualquiera de las reivindicaciones 13-14, en donde la etapa b) implica además la purificación de la proteína de fusión en un medio de afinidad con un resto NT inmovilizado y/o un medio de intercambio aniónico.

Ea	MaSp1	SHTTPWTPNPGLAENFMNSFMQGLSSMPGFTASQOLDMS TIAQSMVQSIQSLAAQGRITSPNKIQALNMFAFA
Lg	MaSp1	QANTPWSSKANADAFINSFISSAQNTGFSFSQDMEDMSLIGNTILMTAMDNG--GRITPSPKIQALDMFAFA
Lh	MaSp1	QANTPWSSKANADAFINSFISAAANTGFSFSQDMEDMSLIGNTILMAAMDNG--GRITPSPKIQALDMFAFA
Nc	MaSp1	-QANTPWSSSTELADAFINAFMNEAGRTGAFADQLDDMS TIGDTIKTAMDKMARSNKSSKKGKIQALNMFAFA
At	MaSp2	QGATPWENSQLAESFISRFELRFIQSCGAFSPNQLDMS SSGDITKTAIEKMAQSRKSSKSKLQALNMFAFA
Lg	MaSp2	--LRWSSKDNADRFINAFLOAASNSGAFSSDQVDDMS VIGNTILMTAMDNG--GRITPSPKIQALDMFAFA
Lh	MaSp2	QANTPWSSKENADAFIGAFMNAASQSGAFSSDQIDDM SVISNTILMAAMDNG--GRITQSKIQALDMFAFA
Nim	MaSp2	QANTPWSDTATADAFIQNFLGAVSGGAFTPDQLDDMS TVGDTIMSAMDKMARSNKSSKSKLQALNMFAFA
Nc	MaSp2	QARSPWSDTATADAFIQNFLAAVSGGAFPSDQLDDMS TIGDTIMSAMDKMARSNKSSQHKLQALNMFAFA
Ab	Cysp1	AVPSVFPSPNLASGFLQCLTFGIGNSPAFPTEQEQDLDLAKVILSAVTSNTDTSKSAR--AQALSTALA
Nc1	Cysp1	EVPSVFPSPNLASGFLGCLTFGIGLSPAFPTEQEQDLDLAKVILSAVTSNTDTSKSAR--AQALSTALA
Lh	TuSp1	ASVNI FNSPNAATSFNLCLRSNIESSPAFPTEQEQADLDSIAEVILSDVSS-VNTASSAT--SLALSTALA
Nc	flag	IANSFNSPNTAEAFARSEFVSNIVSSGEFGAQAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVALLA
Nim	flag	IVNSPFSNPNTAEAFARSEFVSNVSSGEFGAQAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVALLA
Ea	MaSp1	SSMAEIAASEEGGSLSTKTSIASAMSN AFLQTTCVVNQFFINEITQLVSMFAQAQMNDV
Lg	MaSp1	SSVAEIAASEG--GDLGVT TNAIADALTSAFYQ TTCVVNNRFISEIRSLISMFAQAASANDV
Lh	MaSp1	SSVAEIAASEG--GDLGVT TNAIADALTSAFYQ TTCVVNSRFISEIRSLIGMFAQAASANDV
Nc	MaSp1	SSMAEIAAVEQGLSVDAKTNAIADSLNSAFYQ TTCGAANPQFVNEIRSLINMFAQSSANEV
At	MaSp2	SSMAEIAVAEQGLSLEAKTNAIASALSAAFLETTGYVNOQFVNEIKTLIFMIAQAASNEI
Lg	MaSp2	SSVAEIAVADG--QNVGGATNAISNALRSAFYQ TTCVVNNQFISEISNLINMFAQVSAANEV
Lh	MaSp2	SSVAEIAVADG--QNVGAATNAISDALRSAFYQ TTCVVNNQFITGISSLIGMFAQVSGNEV
Nim	MaSp2	SSMAEIAAVEQGGQSMVDKTNAIANALDSAFYMTTCSTNQFVNEMRSLINMLSAAYNEV
Nc	MaSp2	SSMAEIAAVEQCGMSMAVKTNAIVDGLNSAFYMTTGAANPQFVNEMRSLISMI SAASANEV
Ab	Cysp1	SSLTDLLIAESAESNYSNQLSELGTILSDCFIQ TFGSDNPAFVSRIOQLISVLSQNDTNI
Nc1	Cysp1	SSLADLLISESSGSSYQTI SALTNILSDCFVTTTGSNNPAFVSRVQTLIGVLSQSSSNAI
Lh	TuSp1	SSLAE LLVTE SA EEDIDNQVVALS TILSQCFVE T TGSNPAFVA SVKSL LGVLSQSASNYE
Nc	flag	SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGS PNEEFVHEVQD LIQMLSQEQINEV
Nim	flag	SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGS PNEEFVHEVQD LIQMLSQEQINEV

Fig 1

CThyb_Esp	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSSTISNVVS	QIGASNPGLS
CTnat_Eau	SRLSSPSAVS	RVSSAVSSLV	SNG-QVNMAA	LPNIISNISS	SVSASAPGAS
AF350266_At1	SRLSSPGAAS	RVSSAVTSLV	SSGGPTNSAA	LSNTISNVVS	QISSSNPGLS
AY666062_Cm1	SHLSSPEASS	RVSSAVSNLV	SSG-STNSAA	LPNTISNVVS	QISSSNPGLS
AF350273_Lg1	SALAAPATSA	RISSHASTLL	SNG-PTNFAS	ISNVISNAVS	QISSSNPGAS
AY953074_Lh1	SALSAPATSA	RISSHASALL	SSG-PTNPAS	ISNVISNAVS	QISSSNPGAS
AY666068_Mh1	SHLSSPEASS	RVSSAVSNLV	SSG-STNSAA	LPNTISNVVS	QISSSNPGLS
U20329_Nc1	SRLSSPQASS	RVSSAVSNLV	ASG-PTNSAA	LSSTISNVVS	QIGASNPGLS
AY666076_Np1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISSSNPGLS
AF350277_Nm1	SRLSSPQASS	RVSSAVSNLV	ASG-PTNSAA	LSSTISNAVS	QIGASNPGLS
AF350279_Ns1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSSTISNVVS	QIGASNPGLS
AY666057_Ov1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISSSNPGLS
AY666064_Ps1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LPNTISNVVS	QISSSNPGLS
AF350285_Tk1	SLLSSPASNA	RISSAVSALA	SGA-ASGPGY	LSSVISNVVS	QVSSNSGGLV
AF350286_Tv1	SRLSSPASNA	RISSAVSALA	SGG-ASSPGY	LSSIIISNVVS	QVSSNNDGLS
ABU20328_Ab2	SRLSSSAASS	RVSSAVSSLV	SSG-PTTPAA	LSNTISSAVS	QISASNPGLS
AY365016_Aam2	-RLSSPQASS	RVSSAVSTLV	SSG-PTNPAS	LSNAIGSVVS	QVSASNPGLP
AF350263_Aau2	SRLSSPQASS	RVSSAVSTLV	SSG-PTNPAA	LSNAISSVVS	QVSASNPGLS
AF350267_Ai2	SRLSSPQASS	RVSSAVSTLV	SSG-PTNPAS	LSNAISSVVS	QVSSSNPGLS
AF350272_Gm2	SRLSSPQAGA	RVSSAVSALV	ASG-PTSPAA	VSSAISNVAS	QISASNPGLS
AF350275_Lg2	SALSSPTTHA	RISSHASTLL	SSG-PTNSAA	ISNVISNAVS	QVSASNPGLS
AY953075_Lh2	SALSSPTTHA	RISSHASTLL	SSG-PTNAAA	LSNVISNAVS	QVSASNPGLS
AY654293_Nc2	SRLASPDSCA	RVASAVSNLV	SSG-PTSSAA	LSSVISNAVS	QIGASNPGLS
AF350278_Nm2	SRLASPDSCA	RVASAVSNLV	SSG-PTSSAA	LSSVISNAVS	QIGASNPGLS
AF350280_Ns2	SRLASPDSCA	RVASAVSNLV	SSG-PTSSAA	LSSVIXNAVS	QIGASNPGLS
AF350269_DiFb1	SRLSSPEAAS	RVSSAVSSLV	SNG-QVNVDA	LPSIIISNLSS	SISASATTAS
AF350270_DiFb2	SRLSSPQAAS	RVSSAVSSLV	SNG-QVNVAA	LPSIIISNLSS	SISASSTAAS
U47853_ADF1	NRLSSAGAAS	RVSSNVAAIA	SAG----AAA	LPNVISNIYS	GVLSS--GVS
U47854_ADF2	SRLSSPSAAA	RVSSAVS-LV	SNGGPTSPAA	LSSSISNVVS	QISASNPGLS
U47855_ADF3	SRLSSPAASS	RVSSAVSSLV	SSG-PTKHA	LSNTISSVVS	QVSASNPGLS
U47856_ADF4	SVYLRLQFRL	EVSSAVSSLV	SSG-PTNGAA	VSGALNSLVS	QISASNPGLS
Consenso	SRLSSPQASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISASNPGLS

Fig 2

CThyb_Esp	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQLV	QGSVYQALGE	F
CTnat_Eau	GCEVIVQALL	EVITALVQIV	SSSSVGYINP	SAVNQITNVV	ANAMAQVMG-	-
AF350266_At1	GCDVLVQALL	EIVSALVHIL	GSANIGQVNS	SGVGRSASIV	QQSINQAFS-	-
AY666062_Cm1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350273_Lg1	SCDVLVQALL	ELVTALLTII	GSSNVGNVNY	DSSGQYAQVV	QGSVQNAFV-	-
AY953074_Lh1	ACDVLVQALL	ELVTALLTII	GSSNIGSVNY	DSSGQYAQVV	TQSVQNVFG-	-
AY666068_Mh1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVDY	GSAGQATQIV	QOSA-----	-
U20329_Nc1	GCDVLIQALL	EVVSALIQIL	GSSSIGQVNY	GSAGQATQIV	QGSVYQALG-	-
AY666076_Np1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350277_Nm1	GCDVLIQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQ--	-----	-
AF350279_Ns1	GCDVLIQALL	EVVSALVHIL	GSSSIGQVNY	GSAGQATQ--	-----	-
AY666057_Ov1	GCDVLVQALL	EVVSAPIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AY666064_Ps1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350285_Tk1	GCDTLVQALL	EAAAALVHVL	ASSSGGQVNL	NTAGYTSQL-	-----	-
AF350286_Tv1	GCDTVVQALL	EVAALVHVL	ASSNICQVNL	NTAGYTSQL-	-----	-
ABU20328_Ab2	GCDVLVQALL	EVVSALVHIL	GSSSVGQINY	GASAQYQMV	-----	-
AY365016_Aam2	SCDVLVQALL	EIVSALVHIL	GSSSIGQINY	SASSQYARLV	QGSIAQALG-	-
AF350263_Aan2	GCDVLVQALL	ELVSALVHIL	GSSSIGQINY	AAS-----	-----	-
AF350267_At2	GCDVLVQALL	EIVSALVHIL	GSSSIGQINY	AASSQYAQLV	QGSITQALG-	-
AF350272_Gm2	GCDVLVQALL	EIVSALVSIL	SSASIGQINY	GASGQYAMI	-----	-
AF350275_Lg2	SCDVLVQALL	ELITALISIV	DSSNIGQVNY	GSSGQYAQMV	G-----	-
AY953075_Lh2	SCDVLVQALL	EIITALISIL	DSSSVGQVNY	GSSGQYAQIV	QQSMQQAMG-	-
AY654293_Nc2	GCDVLIQALL	EIVSACVTIL	SSSSIGQVNY	GAASQFAQVV	QGSVLSAF--	-
AF350278_Nm2	GCDVLIQALL	EIVSACVTIL	SSSSIGQVNY	GAA-----	-----	-
AF350280_Ns2	GCDVLIXALL	EIVSACVTIL	SSSSIGQVNY	GAA-----	-----	-
AF350269_DtFb1	DCEVLVQVLL	EVVSALVQIV	CS-----	-----	-----	-
AF350270_DtFb2	DCEVLVQVLL	EIVSALVQIV	SSANVGYINP	EASGSLN-AV	GSALAAAMG-	-
U47853_ADF1	SSEALIQALL	EVISALIHVL	GSASIGNVSS	VGVNSALNAV	QNAVQAYAG-	-
U47854_ADF2	GCDILVQALL	EIISALVHIL	GSANIGPVNS	SSAGQASASIV	QGSVYRALS-	-
U47855_ADF3	GCDVLVQALL	EVVSALVSIL	GSSSIGQINY	GASAQYQMV	QGSVAQALA-	-
U47856_ADF4	GCDALVQALL	ELVSALVAIL	SSASIGQVNV	SSVSQSTQMI	SQALS-----	-
Consenso	GCDVLVQALL	EVVSALVHIL	GSSSIGQVNY	GSAGQATQIV	QGSVAQALGE	F

Fig 2 (continuación)

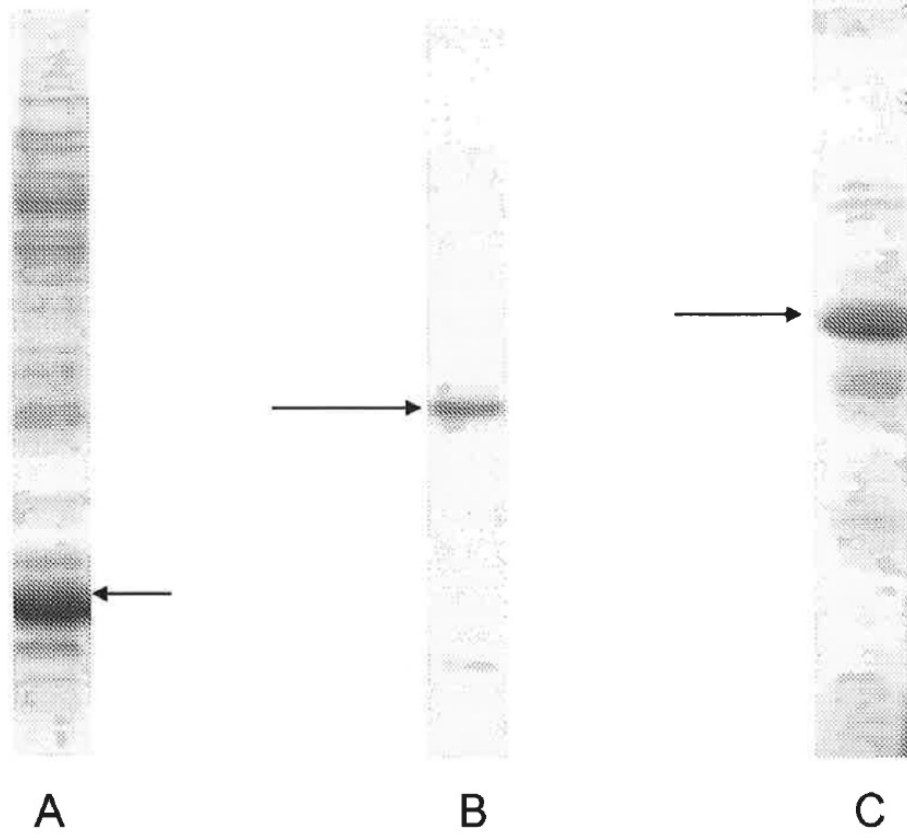


Fig. 3

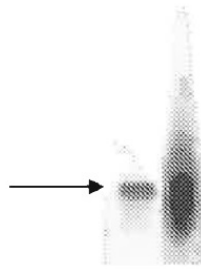


Fig. 4

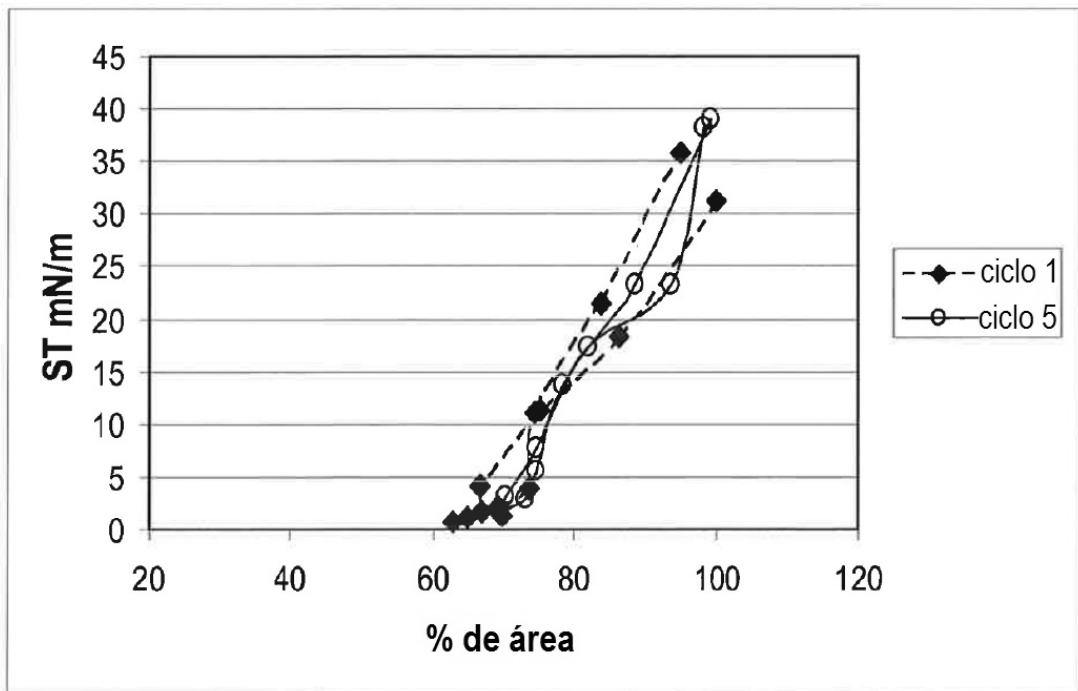


Fig. 5

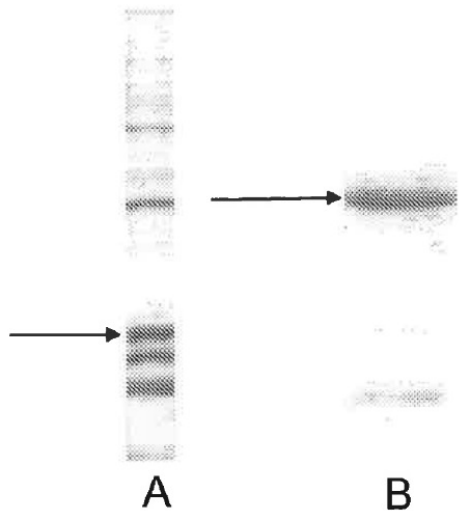


Fig. 6

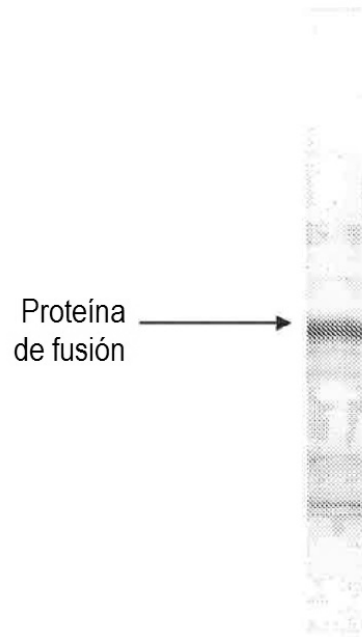


Fig. 7

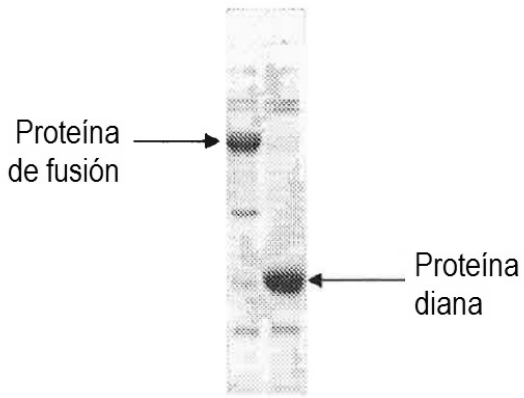


Fig. 8

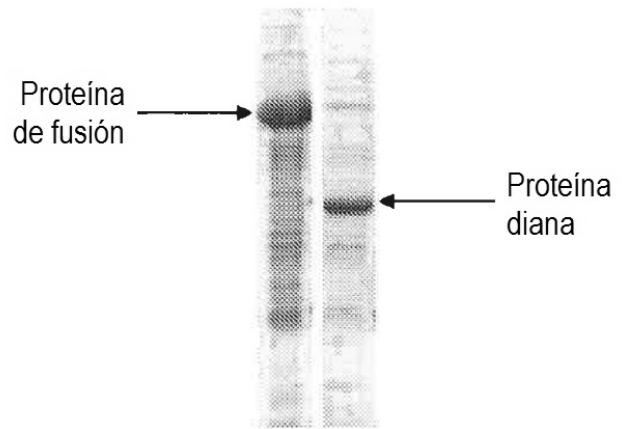


Fig. 9