

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 784 211**

51 Int. Cl.:

A61K 38/16 (2006.01)

C07K 14/415 (2006.01)

G01N 33/53 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.02.2014** **PCT/GB2014/050435**

87 Fecha y número de publicación internacional: **21.08.2014** **WO14125290**

96 Fecha de presentación y número de la solicitud europea: **14.02.2014** **E 14705565 (1)**

97 Fecha y número de publicación de la concesión europea: **18.03.2020** **EP 2956156**

54 Título: **Proteínas de armazón derivadas de cistatinas vegetales**

30 Prioridad:

14.02.2013 GB 201302597

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
23.09.2020

73 Titular/es:

UNIVERSITY OF LEEDS (100.0%)
Leeds LS2 9JT, GB

72 Inventor/es:

MCPHERSON, MICHAEL y
TOMLINSON, DARREN

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 784 211 T3

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

DESCRIPCIÓN

Proteínas de almacén derivadas de cistatinas vegetales

5 La presente invención se refiere a nuevas proteínas, ácidos nucleicos que las codifican, y a métodos para su uso. En particular, la presente invención se refiere a proteínas con utilidad como armazones para mostrar secuencias peptídicas, y al uso de estos armazones en áreas tales como terapia, diagnóstico, monitoreo ambiental y de seguridad, biología de síntesis e investigación.

10 Antecedentes de la invención

El uso de los 'denominados' armazones de proteínas está llamando la atención en bioquímica como una ruta potencial para generar nuevas proteínas de unión a ligandos para su uso en investigación y medicina. Dichas proteínas de almacén, utilizadas para mostrar una o más secuencias de péptidos, pueden proporcionar potencialmente una alternativa a los anticuerpos o fragmentos de anticuerpos.

15 El término 'almacén de proteínas' se usa para describir un tipo de estructura de polipéptidos que se observa en diferentes contextos y con distintas funciones bioquímicas. Debido a su estabilidad conformacional intrínseca, se ha razonado que tales armazones podrían ser susceptibles a manipulación de proteínas.

20 Se piensa convencionalmente que las inmunoglobulinas (anticuerpos) deben su función a la composición de una región marco conservada y un sitio de unión a antígeno espacialmente bien definido hecho de segmentos de péptidos hipervariables, estos segmentos son variables tanto en secuencia como en conformación. Después de que los métodos de manipulación de anticuerpos junto con las técnicas de biblioteca dieron como resultado éxitos en la selección de fragmentos de anticuerpos funcionales, comenzó a crecer el interés en usar otras arquitecturas de proteínas para sintetizar proteínas de unión útiles.

25 Las descripciones tomadas de las propiedades deseables de un almacén de proteína adecuado incluyen las siguientes:

30 "Los candidatos para armazones de proteínas adecuados deberían exhibir un núcleo compacto y estructuralmente rígido que sea capaz de presentar bucles de superficie de secuencia y longitud variables o de lo contrario tolerar reemplazos de cadena lateral en una región de superficie contigua, incluyendo residuos hidrófobos expuestos, sin cambios significativos en sus propiedades de plegado". Skerra A: Engineered protein scaffolds for molecular recognition. J Mol Recognit 2000, 13:167-187 y 'El término 'almacén', como se usa en manipulación de proteínas, describe un marco polipeptídico de cadena sencilla típicamente de tamaño reducido (<200 AA) y que contiene un núcleo altamente estructurado asociado con porciones variables de alta tolerancia conformacional que permite inserciones, eliminaciones u otras sustituciones'. Wurch et al., Trends in Biotechnology, noviembre de 2012, 30, 575-582.

40 Sin embargo, no todas las clases de pliegues de polipéptidos que pueden parecer atractivas para la manipulación de regiones bucle a primera vista permitirán de hecho la construcción de sitios de unión a ligandos independientes con altas afinidades y especificidades.

45 Los armazones del estado de la técnica incluyen nucleasa estafilocócica inactivada, proteína fluorescente verde (GFP) y tioredoxina A (TrxA), el dominio de fibronectina tipo III ('Fn3'), proteínas de la familia de la lipocalina, proteína de unión a bilina (BBP), así como pliegues de proteína aislados tales como el dominio Z de la proteína A estafilocócica, "aficuerpos", anticalinas y repeticiones de anquirina, y otras.

50 El documento WO 2006/131749 describe varias mutaciones racionales hechas en Estefina A para mejorarla como un almacén. El almacén de Estefina A modificada comprende mutaciones en los siguientes tres sitios Lys71-Leu73, V48D y G4W y se conoce como STM (mutante triple de Estefina A). El documento WO2009/136182 describe mejoras adicionales de los armazones de STM.

55 Subsiste la necesidad de proteínas de almacén mejoradas. En particular, se ha encontrado que los armazones del estado de la técnica a menudo son incapaces de estabilizar suficientemente los aptámeros peptídicos, y la presencia de los aptámeros en el almacén en realidad hace que el almacén se deforme significativamente. Véase, por ejemplo, Woodman et al., "Design and Validation of a Neutral Protein Scaffold for the Presentation of Peptide Aptamers", J. Mol. Biol. (2005) 352, 1118-1133.

60 Idealmente, tales proteínas de almacén mejoradas proporcionarían uno o más de los siguientes beneficios:

- Estabilidad mejorada para proporcionar un marco rígido que no se deforma;
- Tamaño más pequeño;
- 65 - Capacidad mejorada para soportar bibliotecas de alta calidad y complejidad;
- Afinidad mejorada de proteínas de unión seleccionadas por su objetivo; y

- Simplicidad de manipulación adicional debido a los extremos terminales N y C accesibles.

Declaraciones de la invención

5 La invención se define en las reivindicaciones adjuntas.

De acuerdo con la presente invención, se proporciona una proteína sintética que tiene una secuencia derivada o relacionada con una cistatina vegetal.

10 De manera adecuada, la proteína sintética comprende una secuencia de consenso de varias (por ejemplo, 10 o más, 20 o más, o 50 o más) proteínas cistatina vegetales.

De manera adecuada, la proteína sintética es una proteína de almacén que comprende sitios adaptados o adecuados para la inserción de secuencias peptídicas heterólogas.

15 La proteína de almacén de la presente invención, por lo tanto, preferiblemente no se basa en una única secuencia de proteína natural, sino que es una proteína nueva que no existe en la naturaleza, que se deriva a través del diseño cuidadoso de la secuencia de consenso de representantes de la clase de cistatinas vegetales

20 En una realización, la proteína sintética comprende la secuencia de aminoácidos
NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYEAK
VWVKPWENFKELQEFKPVG DA (SEQ ID NO 1), o una variante de la misma. Preferiblemente, la variante tiene una secuencia al menos 50%, más preferiblemente 70%, idéntica a la misma.

25 Se ha encontrado que una proteína sintética que comprende dicha secuencia proporciona una proteína de almacén altamente estable y útil.

En otra realización, la proteína sintética comprende la secuencia de aminoácidos
GNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKDGGKKKLYE
30 AKVWVKPWENFKELQEF KPVGDA (SEQ ID NO 2), o una variante de la misma. Preferiblemente, la variante tiene una secuencia al menos 50%, más preferiblemente 70%, idéntica a la misma.

En otra realización, la proteína sintética comprende la secuencia de aminoácidos
ATGVRVPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLEAKD
35 GGKKKLYEAKVWVKPWE NFKELQEFKPVGDA (SEQ ID NO 3), o una variante de la misma. Preferiblemente, la variante tiene una secuencia al menos 50%, más preferiblemente 70%, idéntica a la misma.

Más preferiblemente, la proteína sintética de la presente descripción comprende una secuencia de aminoácidos que es al menos 75%, 80%, 85%, 90%, 95% o 99% idéntica a las SEQ ID NOS 1, 2 o 3 anteriores. Generalmente se
40 prefieren niveles más altos de identidad.

En algunos casos, la proteína sintética comprende o consiste en una secuencia de aminoácidos que es idéntica, o sustancialmente idéntica, a la SEQ ID NO 1, 2 o 3.

45 En las realizaciones preferidas de la presente invención, la proteína sintética expuesta anteriormente comprende al menos una secuencia peptídica heteróloga insertada en la misma. Las proteínas sintéticas de la presente invención tienen una utilidad particular como proteínas de almacén usadas para restringir y desplegar secuencias de péptidos. Por lo tanto, la presente invención se extiende tanto a la proteína de almacén 'vacía' (es decir, sin ninguna secuencia peptídica heteróloga presente) como a la proteína de almacén con una o más secuencias heterólogas insertadas en
50 ella.

Por consiguiente, una realización preferida de la presente invención es una proteína de almacén sintética que tiene una secuencia como se describe en el presente documento, la proteína de almacén despliega uno o más péptidos heterólogos insertados en puntos apropiados en el almacén. Por 'despliegue' se entiende que la secuencia peptídica
55 se inserta en la proteína de almacén en una ubicación que permite que el péptido se exponga en la superficie del almacén bajo condiciones adecuadas, por ejemplo, condiciones no desnaturizantes, tales como un ensayo *in vivo* o *in vitro*. Tales proteínas de almacén que despliegan uno o más péptidos heterólogos a menudo se denominan aptámeros peptídicos o aptámeros peptídicos múltiples, aunque debe tenerse en cuenta que el término aptámeros se usa de manera algo inconsistente en la técnica para referirse a los péptidos mismos o a la proteína completa.

60 Preferiblemente, la secuencia peptídica heteróloga tiene una longitud de 3 a 30 aminoácidos, más preferentemente de 4 a 20 aminoácidos, más preferiblemente de 4 a 16 aminoácidos. Se ha encontrado que las proteínas sintéticas de la presente invención son particularmente adecuadas para desplegar secuencias heterólogas de 5 a 13 aminoácidos de longitud.

65

Preferiblemente, la secuencia peptídica heteróloga se inserta en una región bucle de la proteína sintética o en el extremo terminal N de la proteína. Las regiones bucle se pueden definir como regiones que no están involucradas con estructuras secundarias o terciarias ordenadas de la proteína cuando se encuentran en condiciones convencionales (es decir, la proteína está plegada correctamente y no en condiciones desnaturalizantes), pero pueden contribuir a la función y/o la organización espacial correcta de elementos de estructura secundaria en la proteína. La posición de las regiones bucle dentro de la proteína de armazón puede, por supuesto, variar entre diferentes variantes de proteína. Las regiones bucle de la proteína sintética se pueden determinar examinando la estructura secundaria y terciaria de la proteína, y los métodos para lograr esto son bien conocidos en la técnica, incluida la cristalografía de rayos X, RMN y también en métodos *in silico*.

En la presente invención, se prefiere que los péptidos heterólogos se inserten en al menos una de las siguientes posiciones en la proteína:

- el bucle entre una primera y una segunda región de la lámina β (conocido como BUCLE1); y
- el bucle entre una tercera y una cuarta región de la lámina β (conocido como BUCLE2).

Primero, segundo, tercero y cuarto, por supuesto, deben interpretarse como relativos a la secuencia de la proteína, es decir, del extremo terminal N al C de la proteína.

Preferiblemente, se insertan péptidos heterólogos en ambas posiciones, es decir, en BUCLE1 y BUCLE2.

Se prefiere que la longitud del bucle entre regiones adyacentes de la lámina beta sea de 3 a 20 aminoácidos de longitud, más preferiblemente de 5 a 13 aminoácidos de longitud, y se cree que una longitud de bucle de aproximadamente 9 aminoácidos de longitud es óptima.

Los ejemplos de pares de péptidos heterólogos para inserción en BUCLE1 y BUCLE2 se exponen en la Tabla 2, es decir, las SEQ ID NOS 25 a 72. Estos ejemplos forman realizaciones preferidas de la presente invención.

Además, un punto de inserción adicional para un péptido heterólogo está en o cerca (por ejemplo, dentro de 4 aminoácidos) del extremo terminal N de la proteína. Se cree que la secuencia peptídica en este punto es menos crítica para unir un objetivo que las inserciones en las otras dos posiciones mencionadas anteriormente, pero no obstante parecería tener un papel en la unión.

Por lo tanto, en una realización de la presente invención, la proteína de armazón sintética comprende tres péptidos heterólogos, uno en cada una de las ubicaciones discutidas anteriormente.

En el caso de las secuencias específicas expuestas anteriormente, las regiones bucle preferidas dentro de la secuencia de la proteína se encuentran subrayadas:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLLEAKDGGKKKLYEAKV

VWKPWENFKELQEFKPVGDA (SEQ ID NO 1)

GNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLLEAKDGGKKKLYE

AKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 2)

MATGVRVAPGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQVVAGTMYYLTLLEAK

DGGKKKLYEAKVWVKPWENFKELQEFKPVGDA (SEQ ID NO 3)

Las regiones subrayadas podrían reemplazarse total o parcialmente por el péptido heterólogo, o el péptido heterólogo podría insertarse dentro de las regiones subrayadas sin eliminar las regiones bucle. Además, se puede agregar un péptido heterólogo al extremo terminal N de la SEQ ID NO 1 o 2.

Sin embargo, debe observarse que el péptido heterólogo podría insertarse en otras regiones bucle, por ejemplo, las regiones bucle entre la hélice α y la primera lámina β y/o la segunda y tercera región de la lámina β (que pueden denominarse como LOOP 3 y LOOP 4 respectivamente por conveniencia). Cuando la proteína del armazón se pliega correctamente, estas regiones bucle se colocan en el lado opuesto del armazón de aquellas mencionadas anteriormente (LOOP 1 y LOOP 2). Al insertar regiones bucle de péptidos heterólogos en ambos lados del armazón, es posible producir una fracción divalente, capaz de unir un primer objetivo en un lado del armazón y un segundo objetivo en el otro lado (generalmente opuesto), y esto puede ser una realización deseable en ciertas situaciones.

Cuando la proteína sintética comprende una secuencia de aminoácidos como se estableció anteriormente, la secuencia puede ser contigua o no contigua. Por ejemplo, la secuencia no será contigua cuando se haya insertado un péptido heterólogo dentro de la proteína sintética, por ejemplo, en una de las regiones bucle mencionadas anteriormente, el péptido heterólogo separa así porciones de las secuencias mencionadas anteriormente.

Se observará que los cálculos de identidad de secuencia para la presente invención, en general, deben adaptarse para tener en cuenta la situación en la que se inserta un péptido heterólogo en la proteína sintética. Cuando se ha producido una inserción de este tipo, la secuencia peptídica insertada normalmente no se tiene en cuenta al calcular la identidad de la secuencia. Esto se debe a que las proteínas sintéticas de la presente invención funcionan como armazones, en cuyo caso los péptidos insertados se insertan en puntos de la proteína sintética en los que se desplegaran en la superficie de la proteína. Las secuencias peptídicas insertadas son, por su propia naturaleza y propósito, muy variables. En tal caso, lo que preocupa es la secuencia de la porción de armazón de la proteína, ya que proporciona el marco estable para el despliegue de los péptidos, en lugar de las secuencias peptídicas insertadas intencionalmente altamente variables.

El péptido heterólogo puede insertarse en la proteína sintética con o sin la eliminación de los aminoácidos que normalmente se encuentran en la proteína sintética. Es decir, el péptido heterólogo puede insertarse en un extremo de la proteína sintética o entre dos aminoácidos dentro de la proteína sintética, sin que se eliminen los aminoácidos normales de la proteína sintética. Alternativamente, cuando se inserta un péptido en la proteína sintética, se pueden eliminar/reemplazar uno o más aminoácidos normalmente presentes en la proteína de armazón sintética, por ejemplo, los aminoácidos del 'bucle' VVAG, PWE y ATG (cuando están presentes) se pueden eliminar en las secuencias 1, 2 y 3 anteriores, o las secuencias de bucle correspondientes en una variante de secuencia.

Las proteínas de armazón particularmente preferidas de acuerdo con la invención se exponen en las SEQ ID NOS 74 a 79 a continuación, y la descripción asociada de modificaciones potenciales a los armazones.

El término "identidad" con respecto a la secuencia de proteínas se refiere a un grado de similitud entre proteínas en vista de las diferencias en los aminoácidos, pero que tiene en cuenta diferentes aminoácidos que son funcionalmente similares en vista del tamaño, la lipofilia, la acidez, etc. Se puede calcular un porcentaje de identidad mediante la alineación óptima de las secuencias usando una matriz de puntuación de similitud tal como la matriz Blosum62 descrita en Henikoff S. y Henikoff JG, PNAS Estados Unidos 1992, 89: 10915-10919. El cálculo del porcentaje de identidad y la alineación óptima de dos secuencias utilizando la matriz de similitud Blosum62 y el algoritmo de Needleman y Wunsch (J. Mol. Biol. 1970, 48: 443-453) se pueden realizar utilizando el programa GAP del Genetics Computer Group (GCG, Madison, WI, EE. UU.) utilizando los parámetros predeterminados del programa. Los parámetros específicos para calcular el porcentaje de identidad para secuencias de proteínas y secuencias de ácido nucleico con respecto a la presente invención se describen a continuación.

Las variantes de las proteínas que también forman parte de la presente invención pueden contener variaciones en la secuencia de aminoácidos debido a eliminaciones, sustituciones, inserciones, inversiones o adiciones de uno o más aminoácidos en dicha secuencia o debido a una alteración en una fracción químicamente ligada a la proteína. Por ejemplo, una variante de proteína puede comprender una estructura de carbohidrato o PEG unida a una proteína. Las proteínas de la invención pueden incluir una o más de tales modificaciones de proteínas. Las proteínas de la invención pueden incluir una eliminación de 1 o más aminoácidos, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, aminoácidos del extremo terminal N o C, siempre que la variante conserve la función deseada.

Las variantes de sustitución de proteínas son aquellas en las que se ha eliminado al menos un residuo en la secuencia de aminoácidos y se ha insertado un residuo diferente en su lugar. Las proteínas de la presente invención pueden contener sustituciones conservadoras o no conservadoras.

El término "sustitución conservadora" se refiere a la sustitución de una o más sustituciones de aminoácidos por residuos de aminoácidos que tienen propiedades bioquímicas similares. Típicamente, las sustituciones conservadoras tienen poco o ningún impacto en la actividad de una proteína resultante. Por ejemplo, una sustitución conservadora en una proteína puede ser una sustitución de aminoácido que no afecta sustancialmente la capacidad de la proteína de plegarse correctamente y de lo contrario realizar su función biológica habitual. El cribado de variantes de las proteínas de la presente invención puede usarse para identificar qué residuos de aminoácidos pueden tolerar una sustitución de aminoácidos. En un ejemplo, la temperatura de fusión relevante o la cantidad de hélice α y lámina β de una proteína modificada no se reduce en más del 25%, preferiblemente no más del 20%, especialmente no más del 10%, cuando se realizan uno o más sustituciones conservadoras de aminoácidos.

Se pueden incluir una o más sustituciones conservadoras en una proteína de la presente invención. En un ejemplo preferido, se incluyen 10 o menos sustituciones conservadoras en la proteína. Por lo tanto, una proteína de la invención puede incluir 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más sustituciones conservadoras. Se puede producir un polipéptido que contenga una o más sustituciones conservadoras manipulando la secuencia de nucleótidos que codifica ese polipéptido usando, por ejemplo, procedimientos estándar tales como mutagénesis dirigida al sitio o PCR. Alternativamente, se puede producir un polipéptido que contenga una o más sustituciones conservadoras usando métodos de síntesis de péptidos como se conocen en la técnica.

Los ejemplos de aminoácidos que pueden sustituirse por un aminoácido original en una proteína y que se consideran sustituciones conservadoras incluyen: Ser por Ala; Lys por Arg; Gln o His por Asn; Glu por Asp; Asn por Gln; Asp por Glu; Pro por Gly; Asn o Gln por His; Leu o Val por Ile; Ile o Val por Leu; Arg o Gln por Lys; Leu o Ile por Met; Met, Leu

o Tyr por Phe; Thr por Ser; Ser por Thr; Tyr por Trp; Trp o Phe por Tyr; e Ile o Leu por Val. En una realización, las sustituciones están entre Ala, Val, Leu e Ile; entre Ser y Thr; entre Asp y Glu; entre Asn y Gln; entre Lys y Arg; y/o entre Phe y Tyr. Se puede encontrar más información sobre sustituciones conservadoras en, entre otros sitios, Ben-Bassat et al., (J. Bacteriol. 169: 751-7, 1987), O'Regan et al., (Gene 77: 237-51, 1989), Sahin-Toth et al., (Protein Sci. 3: 240-7, 1994), Hochuli et al., (Bio/Technology 6: 1321-5, 1988), WO 00/67796 (Curd et al.,) y en los libros de texto estándar de genética y biología molecular.

Otras variantes pueden ser, por ejemplo, variantes funcionales tales como sales, amidas, ésteres y específicamente ésteres del extremo terminal C y derivados de N-acilo. También se incluyen péptidos que se modifican *in vivo* o *in vitro*, por ejemplo por glicosilación, amidación, carboxilación o fosforilación.

Las proteínas de acuerdo con la presente invención pueden modificarse mediante una variedad de técnicas químicas para producir derivados que tienen esencialmente la misma actividad que los péptidos no modificados, y opcionalmente que tienen otras propiedades deseables. Por ejemplo, los grupos de ácido carboxílico de la proteína, ya sea del extremo terminal carboxilo o de cadena lateral, pueden proporcionarse en forma de una sal de una presentación farmacéuticamente aceptable o esterificarse, por ejemplo para formar un éster de alquilo C1-C6, o convertirse en una amida, por ejemplo de fórmula CONR^1R^2 en donde R1 y R2 son cada uno independientemente H o alquilo C1-C6, o combinarse para formar un anillo heterocíclico, tal como un anillo de 5 o 6 miembros. Los grupos amino del péptido, ya sea del extremo terminal amino o cadena lateral, pueden estar en forma de una sal de adición de ácido farmacéuticamente aceptable, tal como el HCl, HBr, acético, benzoico, toluenosulfónico, maleico, tartárico y otras sales orgánicas, o puede modificarse a alquilo C1-C6 o dialquil amino o convertirse adicionalmente en una amida. Los grupos hidroxilo de las cadenas laterales peptídicas pueden convertirse en grupos alcoxi o éster, por ejemplo alcoxi C1-C6 o éster de alquilo C1-C6, usando técnicas bien reconocidas. Los anillos fenilo y fenólico de las cadenas laterales peptídicas pueden estar sustituidos con uno o más átomos de halógeno, tales como F, Cl, Br o I, o con alquilo C1-C6, alcoxi C1-C6, ácidos carboxílicos y sus ésteres, o amidas de tales ácidos carboxílicos. Los grupos metileno de las cadenas laterales peptídicas pueden extenderse a alquilenos C2-C4 homólogos. Los tioles pueden protegerse con uno cualquiera de varios grupos protectores bien reconocidos, tales como los grupos acetamida. Los expertos en la materia también reconocerán métodos para introducir estructuras cíclicas en los péptidos de esta descripción para seleccionar y proporcionar restricciones conformacionales a la estructura que den como resultado una estabilidad mejorada.

En las realizaciones preferidas de la presente invención, la proteína sintética tiene una temperatura de fusión (T_m) de al menos 90 °C, más preferiblemente al menos 95 °C, y lo más preferiblemente al menos 100 °C. La T_m en el caso de las proteínas también se conoce como el punto medio de desnaturalización, y se define como la temperatura a la que tanto los estados plegados como los desplegados están igualmente poblados. La T_m generalmente se determina para la proteína sintética en la que no se han insertado secuencias peptídicas heterólogas. La inserción de secuencias heterólogas puede, y a menudo lo hace, reducir la T_m y, por lo tanto, el valor más significativo se obtiene cuando se comparan las proteínas de armazón vacías, y esta es la base de los valores de T_m mencionados anteriormente. Es una ventaja sorprendente de la presente invención que se hayan obtenido tales proteínas altamente estables. La temperatura de fusión es un indicador particularmente útil de la estabilidad de la proteína. Las proporciones relativas de proteínas plegadas y desplegadas se pueden determinar mediante muchas técnicas conocidas por la persona experta, que incluyen calorimetría diferencial de barrido, espectroscopía de diferencia UV, fluorescencia, dicroísmo circular (CD) y RMN (véase Pace, C. Nick y J. Martin Scholtz. "Measuring the conformational stability of a protein" 299-321).

La estabilidad a temperatura extremadamente alta de las proteínas sintéticas de la presente invención también es un indicador de la muy alta estabilidad estructural de las proteínas a temperaturas normales (por ejemplo, alrededor de 25 °C *in vitro* y 37 °C *in vivo*). Esto significa que las proteínas sintéticas de la presente invención son muy adecuadas para actuar como armazones para desplegar péptidos heterólogos. Su estabilidad estructural significa que dichos péptidos heterólogos se desplegarán de manera consistente y precisa, sin alterar la estructura proteica del armazón. Debido a que las proteínas sintéticas de la presente invención son tan estables, es una indicación de que funcionarán mejor que los armazones conocidos que no tienen el mismo nivel de estabilidad.

Es sorprendente que una proteína de armazón tan pequeña tenga una temperatura de transición térmica de, por ejemplo, 101 °C, basada en la fuente de las secuencias de proteínas parentales de las plantas terrestres que se usaron como base para su diseño. Los progenitores de las plantas terrestres comúnmente crecen a temperaturas externas ambientales, por lo que solo se espera que algunos cultivos como el arroz crezcan en temperaturas tropicales, mientras que la mayoría crece en climas del norte moderados a fríos.

Como se mencionó anteriormente, la adición de péptidos heterólogos puede, y típicamente lo hace, reducir la estabilidad de las proteínas de armazón sintéticas de la presente invención. Sin embargo, se ha encontrado que la estabilidad de los armazones de la presente invención sigue siendo mayor que los armazones del estado de la técnica. Por consiguiente, en una realización preferida, las proteínas de armazón sintéticas que tienen al menos un péptido heterólogo insertado en el mismo tienen una T_m de 60 °C o más, más preferiblemente 70 °C o más, y especialmente 80 °C o más. Esto indica que las proteínas de armazón que tienen al menos un péptido heterólogo insertado en ellas son altamente estables.

Preferiblemente, la proteína sintética de la presente invención comprende un enlazador o etiqueta. El enlazador o etiqueta puede ser un enlazador de aminoácidos o etiqueta, u otro tipo de enlazador o etiqueta. La etiqueta puede ser cualquier etiqueta que proporcione una funcionalidad deseada a la proteína sintética, por ejemplo, una que permite un fácil aislamiento o purificación de la proteína. Un ejemplo de etiqueta es una etiqueta de polihistidina (también conocida como etiqueta His), y otras etiquetas bien conocidas incluyen etiqueta Myc, etiqueta SBP, etiqueta S, etiqueta calmodulina, etc.

En una realización adicional de la presente invención, se proporciona una proteína sintética como se expuso anteriormente conectada a un sustrato o fracción. La fracción puede ser un marcador, vehículo, proteína o similar. La proteína sintética se puede conectar directamente al sustrato o fracción, o se puede conectar a través de un enlazador. La proteína sintética se puede conectar covalentemente o no covalentemente al sustrato o fracción. Los sistemas no covalentes para unir proteínas a un sustrato o fracción incluyen, pero no se restringen a, el sistema biotina-avidina.

[0055] Cuando la fracción es un marcador, puede ser un marcador fluorescente, marcador radiactivo, marcador enzimático o cualquier otro marcador conocido por el experto en la materia.

Los ejemplos de marcadores fluorescentes incluyen, pero no se limitan a, colorantes orgánicos (por ejemplo, cianina, fluoresceína, rodamina, los Alexa Fluor, los DyLight Fluor, colorantes ATTO, colorantes BODIPY, etc.), fluoróforos biológicos (por ejemplo, proteína fluorescente verde (GFP), R-Ficoeritrina, etc.) y puntos cuánticos.

Los ejemplos de marcadores enzimáticos incluyen, pero no se limitan a, peroxidasa de rábano picante (HRP), fosfatasa alcalina (AP), glucosa oxidasa y β -galactosidasa.

Otro marcador bien conocido es la biotina. Los marcadores de biotina se componen típicamente del grupo biotinilo, un brazo espaciador y un grupo reactivo que es responsable de la unión a los grupos funcionales objetivo en las proteínas. La biotina puede ser útil para unir la proteína marcada a otras fracciones que comprenden una fracción de avidina.

Las personas expertas conocen bien diversas estrategias para marcar proteínas, y podrían aplicarse fácilmente a las proteínas sintéticas de la presente invención.

El sustrato al que se une la proteína puede ser cualquier superficie adecuada, por ejemplo, la superficie de un recipiente tal como una placa de 96 pozos.

Cuando la fracción es un vehículo, puede ser, por ejemplo, una perla (por ejemplo, una perla magnética) u otra partícula.

Normalmente, se añaden marcadores, enlazadores o etiquetas en los extremos terminales C y/o N de la proteína. Sin embargo, también se puede agregar a través de cualquier región de la proteína que esté suficientemente separada de los bucles en los que se insertan péptidos heterólogos para no interferir con la unión al objetivo, lo más preferiblemente en el lado opuesto de los bucles con inserciones.

En una realización de la presente invención, se proporciona una proteína de fusión que comprende una proteína sintética como se expuso anteriormente conectada a una segunda proteína. La segunda proteína puede ser una proteína que tenga una actividad deseada. En ciertas realizaciones de la invención, la segunda proteína puede ser una proteína de recubrimiento de fago u otra proteína útil en un sistema de despliegue en superficie, y los usos para tales proteínas de fusión en los métodos de barrido se describen con más detalle a continuación. La segunda proteína podría ser, por supuesto, otra proteína sintética, proporcionando así una proteína de armazón homo o hetero multimérica.

En un aspecto adicional de la presente invención, se proporciona una biblioteca que comprende una población de proteínas sintéticas, tal como se describió anteriormente, en la que varios miembros de la población comprenden una variedad de péptidos heterólogos que tienen diferentes secuencias. Se pueden crear bibliotecas adecuadas de péptidos heterólogos utilizando técnicas combinatorias conocidas por el experto en la materia, y las secuencias de ácido nucleico que codifican un conjunto adecuado de secuencias de péptidos heterólogos se pueden obtener a través de numerosas fuentes comerciales. Preferiblemente, dicha biblioteca tendrá una complejidad de 10^8 o superior, más preferiblemente de 10^9 o superior, y lo más preferiblemente de 10^{10} o superior.

Una biblioteca de este tipo tiene utilidad en la selección de proteínas de armazón sintéticas particulares que se unen a una entidad objetivo. La entidad objetivo puede ser cualquier entidad a la que sea deseable que una proteína se una específicamente. Los ejemplos de objetivos incluyen, pero no se limitan a, proteínas/péptidos (por ejemplo, receptores o ligandos), moléculas pequeñas (por ejemplo, moléculas farmacéuticas), ácidos nucleicos (por ejemplo, ADN o ARN) y compuestos inorgánicos.

La biblioteca puede comprender una población de las proteínas sintéticas de la presente invención adaptadas para su despliegue en un sistema de despliegue adecuado, por ejemplo, un sistema de despliegue en superficie. El sistema

de despliegue en superficie puede ser convenientemente un despliegue en fago. Alternativamente, el sistema de despliegue en superficie puede ser un despliegue en ARNm, despliegue en ribosomas, despliegue en CIS u otro sistema de despliegue en proteínas covalente o no covalente, despliegue en bacterias o despliegue en levadura. El sistema o expresión de dos híbridos de levadura en sistemas de células bacterianas o de mamíferos se usan comúnmente en sistemas *in vivo* para el cribado de aptámeros que se unen al objetivo. Colectivamente, estos sistemas de despliegue a menudo se denominan sistemas de biopaneos.

Por ejemplo, un despliegue en fagos puede ser un sistema de despliegue en fagos filamentosos (por ejemplo, M13). Por lo tanto, la proteína sintética puede fusionarse con proteínas pIII, pVIII, pVI, pVII o pIX. Otros sistemas de despliegue de fagos que pueden usarse en la presente invención incluyen despliegue en fagos T4, T7 y λ . Debe observarse que las proteínas sintéticas de la presente invención son versátiles y pueden usarse con cualquier sistema de despliegue adecuado, y los expertos conocen bien varios sistemas adecuados. El despliegue en fagos se usa comúnmente en la selección de anticuerpos, pero las técnicas son directamente aplicables a las proteínas de la presente invención. Los resúmenes de las tecnologías de despliegue en fagos se pueden encontrar, por ejemplo, en Lowman H.B., Clackson T. (2004) Phage display: a practical approach. Oxford University Press. páginas 10-11.

En una realización particularmente preferida de la presente invención, la proteína sintética de la presente invención, especialmente cuando se usa en la construcción de una biblioteca, comprende la secuencia:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X_n)TMYLTLEAKDG
GKKKLYEAKVWVK(X_n)NFKELQEFKPVGDA (SEQ ID NO 4)

en la que X es cualquier aminoácido y n es el número de aminoácidos en la secuencia, y en la que n es preferiblemente de 3 a 30, más preferiblemente de 4 a 20, y aún más preferiblemente de 4 a 15.

Preferiblemente, la proteína sintética comprende la secuencia:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X₅₋₁₃)TMYLTLEA
KDGGKKKLYEAKVWVK(X₅₋₁₃)NFKELQEFKPVGDA (SEQ ID NO 5)

Más preferiblemente, la proteína sintética comprende la secuencia:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X₉)TMYLTLEA
KDGGKKKLYEAKVWVK(X₉)NFKELQEFKPVGDA (SEQ ID NO 6)

Como se discutió anteriormente, las proteínas que comprenden una secuencia que es el 50%, más preferiblemente el 70%, idéntica a la SEQ ID NO 4, 5 o 6 están dentro del alcance de la presente descripción. Más preferiblemente, la proteína sintética de la presente descripción comprende una secuencia de aminoácidos que es al menos 75%, 80%, 85%, 90%, 95% o 99% idéntica a la SEQ ID NO 4, 5 o 6. Generalmente se prefieren mayores niveles de identidad.

En cualquiera de los ejemplos anteriores, la proteína puede contener una secuencia de aminoácidos adicional en el extremo terminal N. Esto puede comprender parte o la totalidad de la secuencia MATGVRAVPGNE (SEQ ID NO 80). Alternativa o adicionalmente puede comprender una secuencia peptídica heteróloga adicional, por ejemplo, de 3 a 20 aminoácidos de longitud.

En las realizaciones de la presente invención, puede preferirse que el aminoácido del extremo terminal N sea una metionina. Esto se puede lograr, por ejemplo, agregando una metionina adicional o reemplazando el aminoácido del extremo terminal N con una metionina.

Las proteínas de armazón particularmente preferidas de la presente invención se exponen en la SEQ ID NOS 74 a 79. Las variantes de estas proteínas de acuerdo con los criterios establecidos anteriormente también son realizaciones de la presente invención.

En otro aspecto, la presente invención proporciona un polinucleótido que codifica una proteína sintética de acuerdo con la presente invención. El polinucleótido puede ser ADN o ARN. Si el polinucleótido es ADN, puede estar en forma monocatenaria o bicatenaria. La cadena sencilla podría ser la cadena codificante o la cadena no codificante (antisentido).

Un polinucleótido de acuerdo con la presente descripción puede comprender una secuencia que codifica una proteína que tiene una secuencia de acuerdo con las SEQ ID NO 1, 2, 3, 4, 5 o 6, o cualquiera de las SEQ ID NO 74 a 79, o una variante de la misma, como se discutió anteriormente.

En una realización el polinucleótido comprende la secuencia:

aacgctctgctggaattcgttctgtgtttaagctaaagaacaggttgctgtgacctgtactacctgaccctggaagctaaagacggttgtaaaaagaaactgt acgaagctaaagttgggttaaaccgtgggaaaactcaagaactgcaggagttcaaacggttggtgacgct (SEQ ID NO 7), que codifica el aminoácido establecido en la SEQ ID NO 1.

5 [0078] En otra realización el polinucleótido comprende la secuencia:
ggtaacgaaaactccctggaatcgaagaactggctcgttgcgtgttgacgaacacacaaaaaagaaaacgctctgctggaattcgttctgtgtttaagctaaa
gaacaggttgctgtgacctgtactacctgaccctggaagctaaagacggttgtaaaaagaaactgtacgaagctaaagttgggttaaaccgtgggaaaactt
caaagaactgcaggagttcaaacggttggtgacgct (SEQ ID NO 8), que codifica el aminoácido establecido en la SEQ ID NO 2.

10 En otra realización el polinucleótido comprende la secuencia:
gctaccggtgttctgtcagttccgggtaacgaaaactccctggaatcgaagaactggctcgttgcgtgttgacgaacacacaaaaaagaaaacgctctgctgga
attcgttctgtgtttaagctaaagaacaggttgctgtgacctgtactacctggaagctaaagacggttgtaaaaagaaactgtacgaagctaaagttt
gggttaaaccgtgggaaaactcaagaactgcaggagttcaaacggttggtgacgct (SEQ ID NO 9), que codifica el aminoácido
establecido en la SEQ ID NO 3.

15 Los polinucleótidos que tienen una secuencia de ácido nucleico que es una variante de las secuencias de ácido
nucleico identificadas pueden aislarse mediante un método que comprende las etapas de: a) hibridar un ADN que
comprende toda o parte de la secuencia identificada como se refleja en cualquiera de las SEQ ID NOS 7, 8 o 9, bajo
condiciones estrictas contra ácidos nucleicos de interés; y b) aislar dichos ácidos nucleicos por métodos conocidos
20 por un experto en la materia. Las condiciones de hibridación son preferiblemente altamente rigurosas.

De acuerdo con la presente invención, el término "riguroso" significa condiciones de lavado de 1 x SSC, SDS al 0,1%
a una temperatura de 65 °C; las condiciones 'altamente rigurosas' se refieren a una reducción en SSC hasta 0,3 x
25 SSC, más preferiblemente hasta 0,1 x SSC. Preferiblemente, los dos primeros lavados se llevan a cabo posteriormente
dos veces cada uno durante 15-30 minutos. Si es necesario lavar en condiciones muy rigurosas, se realiza un lavado
adicional con 0,1 x SSC una vez durante 15 minutos. La hibridación se puede realizar durante la noche en tampón
fosfato 0,5 M pH 7,5 con SDS al 7% a 65 °C. Dichos métodos de hibridación se describen en cualquier libro de texto
estándar sobre clonación molecular, por ejemplo: Molecular Cloning: A Laboratory Manual, 3ª ed.; editores: Sambrook
et al., CSHL Press, 2001.

30 Las variantes de las secuencias representadas en las SEQ ID NOS 7, 8 o 9 también pueden identificarse comparando
la secuencia *in silico* con otras secuencias que pueden estar comprendidas en una base de datos informática. Las
secuencias pueden compararse con secuencias en bases de datos utilizando un programa BLAST (BLASTF 2.1.2 [19
de octubre de 2000]) (Altschul, SF, TL Madden, AA Schaffer, J Zhang, Z Zhang, W Miller y DJ Lipman, "Gapped BLAST
y PSI-BLAST: "Gapped BLAST and PSI-BLAST: a new generation of protein database Search programs", Nucleic
35 Acids Res. 1997,25: 3389-3402).

Las realizaciones preferidas de la divulgación son polinucleótidos que codifican proteínas que tienen al menos 50%,
más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 95% de identidad con las SEQ ID NOS 1-
40 6, y aún más preferidos son aquellos polinucleótidos que codifican proteínas que tienen al menos un 97% de identidad
con las SEQ ID NOS 1, 2, 3, 4, 5 o 6, o cualquiera de las SEQ ID NO 74 a 79, siendo más preferidas aquellas proteínas
codificantes que tienen al menos un 98 o 99% de identidad. Los más preferidos son los polinucleótidos que codifican
las proteínas de las SEQ ID NOS 1, 2, 3, 4, 5 o 6, o cualquiera de las SEQ ID NOS 74 a 79.

45 Debido a la degeneración del código genético, los polinucleótidos que codifican una secuencia de aminoácidos idéntica
o sustancialmente idéntica pueden utilizar diferentes codones específicos. Todos los polinucleótidos que codifican las
proteínas como se definió anteriormente se consideran parte de la invención.

A este respecto, debe tenerse en cuenta que las secuencias discutidas anteriormente se han optimizado para la
50 expresión en el huésped bacteriano *Escherichia coli*, ya que este es un sistema conveniente y se usa típicamente para
las técnicas tales como el despliegue en fagos y para la producción posterior de proteínas de unión seleccionadas.
Sin embargo, cualquier persona experta en la técnica podría optar por expresar las proteínas en un sistema de
hospedadores alternativo, y en ese caso un cambio en la secuencia de nucleótidos que codifica la proteína
seleccionada (es decir, la optimización del codón) puede ser beneficioso y, por lo tanto, son parte de la presente
55 invención. La optimización del uso de codones para cualquier huésped particular es una rutina para la persona experta.

En realizaciones preferidas particulares, los polinucleótidos de acuerdo con la descripción son polinucleótidos aislados
que comprenden una secuencia que tiene al menos un 80% de identidad con la secuencia de ácido nucleico de las
60 SEQ ID NOS 7, 8 o 9 (excluyendo secuencias que codifican péptidos heterólogos, cuando sea relevante).

Los polinucleótidos de acuerdo con la presente invención pueden, por supuesto, contener secuencias adicionales,
tales como secuencias que codifican péptidos heterólogos, secuencias de control de expresión o secuencias que
codifican otras proteínas o subunidades de proteínas.

65 Más preferidos son aquellos polinucleótidos que comprenden una secuencia que tiene al menos un 90% de identidad,
y aún más preferidos al menos un 95% de identidad, preferiblemente un 99% de identidad, lo más preferido un 100%

de identidad con respecto a la secuencia completa de las SEQ ID NOS 7, 8 o 9 (excluyendo secuencias que codifican péptidos heterólogos, cuando sea relevante).

5 También se incluyen dentro de la definición de polinucleótidos los ARN o ADN modificados. Se pueden hacer modificaciones en las bases del ácido nucleico, y se pueden incorporar bases como la inosina. Otras modificaciones pueden implicar, por ejemplo, modificaciones de la cadena principal.

10 El "% de identidad" define la relación entre dos o más polinucleótidos o polipéptidos sobre la base de una comparación entre sus secuencias alineadas.

La identidad se puede calcular por métodos conocidos. Los porcentajes de identidad u homología mencionados en este documento con respecto a la presente invención son aquellos que pueden calcularse con el programa GAP, que se ejecuta bajo GCG (Genetics Computer Group Inc., Madison, WI, EE. UU.).

15 Para la comparación de secuencias de polipéptidos:

- Algoritmo de alineación: Needleman y Wunsch, J. Mol. Biol. 1970, 48: 443-453.
- Como matriz de comparación para la similitud de aminoácidos, se usa la matriz Blosum62 (Henikoff y Henikoff, citado anteriormente).
- 20 - Se utilizan los siguientes parámetros de puntuación por hueco:
 - Penalización por hueco: 12
 - Penalización de longitud de hueco: 2
 - No hay penalización por huecos en los extremos.

25 Para la comparación de la secuencia de nucleótidos:

- Algoritmo de alineación: Needleman y Wunsch (citado anteriormente).
- Matriz de comparación:
 - coincidencia = +10, falta de coincidencia = 0.
 - Penalización por hueco: 50.
 - Penalización por longitud de hueco: 3.

35 Los enfoques para desarrollar proteínas estabilizadas de consenso que incluyen la incorporación de otros parámetros, tales como métodos filogenéticamente no sesgados como los descritos por Jaeckel, Bloom et al., 2010, también se pueden utilizar.

40 Sin embargo, como se discutió anteriormente, se observará que los diversos cálculos de identidad de secuencia deben modificarse para tener en cuenta la situación en la que uno o más péptidos heterólogos se insertan opcionalmente en la proteína sintética. Cuando se produce una inserción de este tipo, la secuencia insertada que codifica el péptido o péptidos generalmente no se tiene en cuenta al calcular la identidad de la secuencia por los motivos expuestos anteriormente.

45 Los ácidos nucleicos, especialmente el ADN, de acuerdo con la invención serán útiles para la expresión *in vivo* o *in vitro* de la proteína codificada. Cuando los polinucleótidos de acuerdo con la invención se usan para la expresión de las proteínas codificadas, los polinucleótidos pueden incluir, además de la secuencia codificante para la proteína, otras secuencias codificantes o no codificantes, por ejemplo, secuencias líder o porciones de fusión, secuencias enlazadoras, secuencias marcadoras, promotores, potenciadores y similares.

50 Los polinucleótidos de acuerdo con la invención pueden usarse en la producción de proteínas recombinantes de acuerdo con la invención. Esto se puede lograr mediante la expresión de las proteínas en una célula huésped adecuada u organismo multicelular. Cuando se usa para la expresión de un polipéptido codificado, los polinucleótidos pueden incluir ventajosamente, además de la secuencia de codificación para el polipéptido, otras secuencias de codificación, por ejemplo, secuencias señal, secuencias líder, secuencias de direccionamiento, porciones de fusión, secuencias marcadoras, secuencias para ayudar en purificación y similares.

55 Se puede usar una amplia variedad de combinaciones de célula huésped y vehículo de clonación para la clonación y la expresión. Los polinucleótidos de la presente invención pueden clonarse en cualquier sistema de expresión apropiado. Los sistemas de expresión adecuados incluyen un sistema de expresión bacteriano (por ejemplo, *Escherichia coli* DH5α y BL21 (DE3)), un sistema de expresión viral (por ejemplo, Baculovirus), un sistema de levadura (por ejemplo, *Saccharomyces cerevisiae*) o células eucariotas (por ejemplo, células COS-7, CHO, BHK, HeLa, HD11, DT40, CEF o HEK-293T). Un amplio rango de sistemas de expresión adecuados están disponibles comercialmente. Típicamente, el polinucleótido se clona en un vector apropiado bajo control de un promotor constitutivo o inducible

60 adecuado y luego se introduce en la célula huésped para la expresión.

65

Por lo tanto, en otro aspecto, la presente invención proporciona un vector recombinante que comprende un polinucleótido de acuerdo con la invención. Los vectores adecuados incluyen plásmidos bacterianos o de levadura, cósmidos, fagémidos, fósidos, plásmidos de amplio rango de huéspedes y vectores derivados de combinaciones de ADN de plásmidos y fagos o virus. Un origen de replicación y/o un marcador de selección dominante puede estar presente convenientemente en el vector.

Los vectores de acuerdo con la invención son adecuados para transformar una célula huésped. Ejemplos de vectores de clonación adecuados son vectores plasmídicos tales como pBR322, los diversos plásmidos pUC, pET, pEMBL y Bluescript, o vectores virales tales como retrovirus, lentivirus, adenovirus, virus adenoasociados. pcDNA3.1 es un vector particularmente preferido para la expresión en células animales.

Particularmente útiles para la clonación y expresión de las bibliotecas de proteínas de armazón son vectores de fagos o fagémidos tales como los derivados de fagos filamentosos tales como M13 o fd o fagos de cápside tal como el bacteriófago T7.

Cuando se usa en la expresión de las proteínas de la presente invención, un vector de acuerdo con la presente invención típicamente comprende una secuencia de control de expresión operativamente unida a la secuencia de ácido nucleico que codifica la proteína para controlar la expresión del polinucleótido relevante. Dichas secuencias de control de la expresión generalmente comprenden una secuencia promotora y secuencias adicionales que regulan la transcripción y la traducción y/o mejoran los niveles de expresión. Las secuencias de control de expresión adecuadas son bien conocidas en la técnica e incluyen el promotor eucariota, procariota o viral o la señal de poli-A. El control de la expresión y otras secuencias variarán, por supuesto, dependiendo de la célula huésped seleccionada o pueden hacerse inducibles. Ejemplos de promotores útiles son el promotor tac (Deboer, Comstock et al., 1983), el promotor T7 (Studier y Moffatt 1986), el promotor de SV-40 (Science 1983, 222: 524-527), el promotor de metalotioneína (Nature 1982, 296: 39-42), el promotor de choque térmico (Voellmy et al., PNAS USA 1985, 82: 4949-4953), el promotor gX de PRV (Mettenleiter y Rauh, J. Virol. Methods 1990, 30: 55-66), el promotor IE de CMV humano (US 5.168.062), el promotor LTR del virus del sarcoma de Rous (Gorman et al., PNAS USA 1982, 79: 6777-6781), o el promotor del factor 1 alfa de alargamiento humano o de ubiquitina. Las secuencias de control procariotas incluyen, por ejemplo, lac, tac T7, ara y otros promotores conocidos. Los promotores procariotas en general que pueden usarse en la presente invención se discuten en Goldstein MA, Doi RH, 'Prokaryotic promoters in biotechnology'. Biotechnol Annu Rev. 1995; 1: 105-28. Se conocen muchas otras secuencias de control adecuadas en la técnica, y sería una rutina para la persona experta seleccionar secuencias adecuadas para el sistema de expresión que se usará.

Después de que el polinucleótido se ha clonado en un vector apropiado, el constructo se puede transferir a la célula (por ejemplo, bacteria o levadura) por medio de un método apropiado, tal como electroporación, transfección con CaCl_2 o lipofectinas. Cuando se usa un sistema de expresión de baculovirus, el vector de transferencia que contiene el polinucleótido puede transfectarse junto con un genoma completo de baculovirus.

Estas técnicas son bien conocidas en el arte y los fabricantes de materiales de biología molecular (como Clontech, Agilent, Promega y/o Invitrogen Life Technologies) proporcionan reactivos e instrucciones adecuados sobre cómo usarlos. Además, hay una serie de libros de texto de referencia estándar que proporcionan más información al respecto, por ejemplo, Rodríguez, R. L. y D. T. Denhardt, ed., "Vectors: A survey of molecular cloning vectors and their uses", Butterworths, 1988; Current protocols in Molecular Biology, eds.: F. M. Ausubel et al., Wiley N. Y., 1995; Molecular Cloning: a laboratory manual, citado anteriormente; y DNA Cloning, Vol. 1-4, 2a edición 1995, eds.: Glover y Hames, Oxford University Press).

[0105] En otro aspecto, la presente invención también proporciona una célula capaz de expresar una proteína recombinante, caracterizada porque la célula comprende un polinucleótido de acuerdo con la invención que codifica la proteína recombinante a expresar. Adecuadamente, la célula es una célula huésped transformada con un polinucleótido o vector como se describió anteriormente. El polinucleótido o vector de acuerdo con la invención puede integrarse de manera estable en el material genómico de la célula o puede ser parte de un vector de replicación autónoma. "Recombinante" en este contexto se refiere a una proteína que no se expresa en la célula en la naturaleza.

Las células huésped pueden cultivarse en medios nutritivos convencionales que pueden modificarse, por ejemplo, para la selección, amplificación o inducción apropiadas de la transcripción y, por lo tanto, la expresión de la proteína recombinante. Las células huésped pueden ser procariotas o eucariotas. Las líneas celulares de mamífero, por ejemplo, de humano, pueden preferirse, especialmente cuando las proteínas expresadas están destinadas para uso *in vivo* en sujetos humanos o mamíferos. Los ejemplos de líneas celulares adecuadas se mencionaron anteriormente, y las condiciones de cultivo apropiadas para los diversos tipos celulares adecuados son bien conocidas por el experto en la materia.

En otro aspecto, la presente invención proporciona un cultivo celular que comprende células de acuerdo con la invención.

En otro aspecto, la presente invención proporciona un método de cribado para una proteína sintética que se une a un objetivo, comprendiendo el método:

- a) Proporcionar una biblioteca que comprende una población de proteínas de armazón sintéticas que comprenden una o más secuencias de péptidos heterólogos como se describió anteriormente;
- b) Exponer dicha biblioteca a un objetivo; y
- c) Seleccionar proteínas de armazón sintéticas que se unen a dicho objetivo.

El objetivo puede ser convenientemente una proteína/péptido, un ácido nucleico o una molécula pequeña o cualquier otro objetivo de interés. El objetivo puede inmovilizarse convenientemente sobre un sustrato, por ejemplo, a los pozos de una placa de microtitulación o en perlas magnéticas u otras partículas apropiadas.

El método de cribado puede ser convenientemente un sistema de despliegue, por ejemplo, un sistema de despliegue en superficie. Por ejemplo, un sistema de despliegue en superficie adecuado podría ser un despliegue en fagos, un despliegue en ARNm, un despliegue en ribosomas, despliegue en CIS (Odegrip, Coomber et al., 2004) u otro despliegue en proteínas no covalentes o covalentes (FitzGerald 2000), despliegue en bacterias (Lofblom 2011) y despliegue en levadura (Traxlmayr y Obinger 2012) o despliegue en otra célula eucariota (Makela y Oker-Blom 2008) (Ho y Pastan 2009).

En ciertas realizaciones de la presente invención, las proteínas utilizadas en el método de cribado son, por lo tanto, fusiones de la proteína de armazón con una proteína de cubierta de bacteriófago, de modo que las proteínas de armazón se despliegan en la superficie de la partícula viral. La proteína que se despliega en la partícula viral corresponde, por lo tanto, a la secuencia genética dentro de la partícula del fago. Por supuesto, cuando el sistema de despliegue en superficie es distinto de un sistema de despliegue en fagos, se pueden proporcionar otras proteínas de fusión adecuadas.

Cuando el objetivo se inmoviliza en un sustrato, la biblioteca desplegada en la superficie se expone al sustrato con el objetivo inmovilizado sobre el mismo y, después de un período adecuado para permitir que el tiempo del fago se una al objetivo, el sustrato se enjuaga. Las proteínas que se unen al objetivo permanecen unidas al sustrato, mientras que otras se eliminan por lavado.

A continuación, se eluyen convenientemente las proteínas que se unen al objetivo o el fago asociado se separa del objetivo y, en el caso de la presentación en fagos, se usa para crear más fagos por infección de huéspedes bacterianos adecuados. Esto da como resultado una nueva biblioteca que comprende una mezcla enriquecida, que contiene fagos considerablemente menos irrelevantes (es decir, fagos que contienen proteínas de armazón que no se unen) que estaban presentes en la biblioteca inicial.

Las etapas de exponer la biblioteca al objetivo, el enjuague, la elución y la infección se repiten opcionalmente una o más veces, enriqueciendo aún más la biblioteca de fagos en las proteínas de unión. Esto generalmente se conoce como paneo. Por lo tanto, el método puede implicar realizar múltiples etapas de paneo.

Una vez que se ha realizado un número deseado de etapas de paneo, el ácido nucleico unido a la proteína desplegada en la superficie puede secuenciarse para identificar las proteínas de armazón que se unen al objetivo.

Se ha encontrado que se pueden obtener proteínas de armazón de acuerdo con la presente invención que se unen con alta afinidad y alta especificidad a los objetivos. Por ejemplo, se han obtenido proteínas de armazón que se unen con una afinidad de 1×10^{-9} M. Inesperadamente, las bibliotecas formadas a partir de proteínas de armazón de acuerdo con la presente invención han revelado proteínas de armazón que se unen a proteínas contra las cuales ha resultado imposible generar anticuerpos o anticuerpos con la especificidad deseada. Por ejemplo, las proteínas de unión artificial basadas en el armazón han sido proteínas seleccionadas que se unen específicamente al virus del papiloma humano (HPV) 16 E5, una proteína poco estudiada debido a la falla repetida de la capacidad de seleccionar la falta de anticuerpos adecuados contra este objetivo. El HPV es un agente causante del carcinoma de cuello uterino. También se han seleccionado proteínas de unión artificial que discriminan entre las dos formas del modificador tipo ubiquitina pequeña humana (SUMO), hSUMO1 y hSUMO2, mientras que los anticuerpos previamente seleccionados demostraron ser incapaces de tal discriminación. Es razonable esperar que otras proteínas de unión útiles puedan surgir contra otros objetivos cuando las técnicas basadas en anticuerpos no han podido proporcionar una solución adecuada.

En otro aspecto, la presente invención proporciona una proteína de armazón sintética obtenida por un método de cribado descrito anteriormente.

En otro aspecto, la presente invención proporciona el uso de proteína sintética como se estableció anteriormente, o un ácido nucleico que codifica dicha proteína sintética, en investigación. Por ejemplo, tales proteínas se pueden usar en cualquier área de investigación en la que se usan típicamente anticuerpos o fragmentos de anticuerpos. Dichos métodos pueden incluir interferir con las interacciones proteína/proteína, marcar objetivos (por ejemplo, de manera fluorescente), etc.

En otro aspecto, la presente invención proporciona el uso de proteína sintética como se estableció anteriormente, o un ácido nucleico que codifica dicha proteína sintética, en el control ambiental y de seguridad, o en biología de síntesis.

5 En otro aspecto, la presente invención proporciona el uso de proteína sintética como se estableció anteriormente, o un ácido nucleico que codifica dicha proteína sintética, en la detección de objetivos.

10 En otro aspecto, la presente invención proporciona el uso de una proteína de armazón sintética como se estableció anteriormente, o un ácido nucleico que codifica dicha proteína sintética, en terapia o diagnóstico. Por ejemplo, tales proteínas se pueden usar en cualquier área de diagnóstico o terapia en la que se usan típicamente anticuerpos o fragmentos de anticuerpos.

En otro aspecto, la presente descripción proporciona una preparación farmacéutica que comprende una proteína sintética como se estableció anteriormente y un vehículo o excipiente farmacéuticamente aceptable.

15 En otro aspecto, la presente descripción proporciona un método para validar objetivos farmacológicos e identificar dominios farmacológicos en proteínas objetivo usando proteína sintética como se estableció anteriormente.

20 En otro aspecto, la presente descripción proporciona un método adecuado para seleccionar proteínas sintéticas que se unen a moléculas pequeñas que incluyen compuestos orgánicos y compuestos farmacológicos.

Adecuadamente, las proteínas sintéticas de la presente invención pueden unirse a un agente terapéutico adecuado. En tal situación, la proteína sintética puede proporcionar una función de direccionamiento para suministrar la 'carga útil' terapéutica a un objetivo deseado.

25 La preparación de preparaciones farmacéuticas de acuerdo con la presente descripción se lleva a cabo por medios convencionales para la persona experta. Los métodos para preparar composiciones administrables, ya sea para administración intravenosa o subcutánea o de otro modo, serán conocidos o evidentes para los expertos en la materia y se describen con más detalle en publicaciones tales como Remington: The Science and Practice of Pharmacy, Lippincott Williams y Wilkins; 21ª edición revisada (1 de mayo de 2005).

30 'Proteína sintética' se refiere a una proteína que no se encuentra en la naturaleza y que se ha formado mediante técnicas recombinantes.

35 El término 'proteína' generalmente se usa indistintamente con 'péptido' o 'polipéptido', y significa al menos dos aminoácidos unidos covalentemente conectados por un enlace peptídico. El término proteína abarca productos naturales purificados, o productos que pueden producirse parcial o totalmente utilizando técnicas recombinantes o sintéticas. El término proteína puede referirse a un agregado de una proteína, tal como un dímero u otro multímero, una proteína de fusión, una variante de proteína o un derivado de la misma. El término también incluye proteínas modificadas, por ejemplo, una proteína modificada por glicosilación, acetilación, fosforilación, pegilación, ubiquitinación, etc.

40 Una proteína puede comprender aminoácidos no codificados por un codón de ácido nucleico. Los aminoácidos no naturales pueden introducirse mediante modificación postraduccional, por ejemplo, mediante la introducción de una cisteína u otro residuo apropiado en una posición específica y luego la conversión de esta seguida por la purificación de proteína en una deshidroalanina y luego la reacción química con un aminoácido no natural convenientemente modificado. Alternativamente, se podría incorporar un aminoácido no natural en la secuencia durante la traducción, ya sea *in vitro* o *in vivo*, mediante la supresión de un codón de parada, normalmente UAG, a través de una pareja de ARNt/ARNt sintetasa heteróloga apropiada que carga un ARNt supresor con un aminoácido no natural que luego se puede incorporar a la proteína en la posición del codón UAG.

50 La proteína de la presente invención puede ser una proteína de subunidad única o una proteína de subunidad múltiple en la que al menos una subunidad comprende dicha secuencia.

55 En ciertas realizaciones, la proteína puede comprender dos o más proteínas sintéticas como se estableció anteriormente, por ejemplo, en forma de una proteína de fusión. En tal caso, cada una de las proteínas sintéticas puede unirse al mismo objetivo o pueden unirse a diferentes objetivos. La unión de los mismos puede aumentar la avidéz y la aparente afinidad de unión, mientras que esos objetivos distintos de unión pueden ser útiles para el desarrollo de módulos de reconocimiento doble o de entrecruzamiento.

60 'Péptido heterólogo' o 'Secuencia de péptido heterólogo' en el contexto de la presente invención se refiere a una secuencia peptídica que normalmente no se encuentra en la ubicación relevante, por ejemplo, en la región bucle de la proteína del armazón. Típicamente, tales péptidos son relativamente cortos, es decir, contienen menos de 30 aminoácidos, típicamente 20 o menos. Tal péptido puede tener esencialmente cualquier secuencia. De hecho, es deseable que las bibliotecas que contienen una gran cantidad de secuencias de péptidos heterólogos se desplieguen en proteínas de armazón.

65

Debe entenderse que las rasgos, números enteros, características, compuestos, fracciones químicas o grupos descritos junto con un aspecto, realización o ejemplo particular de la invención son aplicables a cualquier otro aspecto, realización o ejemplo descrito en el presente documento, a menos que sea incompatible con ellos. Todas las características divulgadas en esta memoria descriptiva (incluidas las reivindicaciones, el resumen y los dibujos adjuntos), y/o todas las etapas de cualquier método o proceso así divulgado, se pueden combinar en cualquier combinación, excepto combinaciones en las que al menos algunas de dichas características y/o etapas son mutuamente excluyentes. La invención no está restringida a los detalles de ninguna de las realizaciones anteriores. La invención se extiende a cualquier característica nueva, o cualquier combinación novedosa, de las características descritas en esta memoria descriptiva (incluidas las reivindicaciones, resumen y dibujos adjuntos), o cualquier característica nueva, o cualquier combinación novedosa, de las etapas de cualquier método o proceso así descrito.

La atención del lector se dirige a todos los artículos y documentos que se mencionan en esta solicitud, y que están abiertos a la inspección pública.

Realizaciones específicas de la invención

Las realizaciones de la invención se describirán ahora, solo a modo de ejemplo, con referencia a los dibujos adjuntos en los que:

- La Figura 1 muestra una fitocistatina de consenso PHYTC57 derivada de 57 secuencias de aminoácidos de fitocistatina (A) El gen sintético PHYTC57 se muestra como una secuencia de doble cadena junto con los oligonucleótidos superpuestos (P1 a P6) utilizados para generar el gen por PCR recursiva. La secuencia de codificación también se muestra como un código de aminoácidos de una letra. Se muestran las posiciones de los dos sitios de restricción *SfiI* y *NotI*. (B) Representación esquemática de la región de clonación de pDHisII que se basa en pHEN1. Se muestran las posiciones de las regiones relevantes que codifican la secuencia señal pelB, la etiqueta de His6 y la sección del extremo terminal N de la proteína PIII del fago M13. Las posiciones de los sitios de unión del cebador M13 estándar para M13R y P10 se indican como los sitios de restricción *SfiI* y *NotI* únicos.
- La Figura 2 muestra el gráfico de los resultados de los ensayos de inhibición de papaína para una orizacistatina modificada que carece del residuo Asp86 (OSA-IAD86) y PHYTC57 a concentraciones variables de fitocistatinas, que muestra la eficacia mejorada de PHYTC57.
- La Figura 3 muestra la medición de resonancia de plasmón superficial de la interacción entre cistatinas inmovilizadas en un chip sensor NTA y papaína. Los experimentos se realizaron a varias concentraciones de papaína y los datos se analizaron utilizando el paquete de software Biaevaluation3 (BIAcore) con ajuste global de los datos al modelo de unión de Langmuir 1:1. OSA-IAD86, una cistatina I de arroz modificado; CPA, cistatina de papaya; CUN, cistatina de naranja; CEWC, cistatina de clara de huevo de gallina; PHYTC57, cistatina de consenso.
- La Figura 4 muestra (A) La estabilidad térmica de OSA-IAD86 (cuadrado gris) y PHYTC57 (triángulo negro) mostrada como la actividad enzimática residual ensayada a 25 °C después de incubación a 100 °C para los tiempos mostrados. PHYTC57 muestra una mayor estabilidad térmica que OSA-IAD86, (B) Efecto del tratamiento con fluido gástrico simulado sobre las propiedades inhibitorias de OSA-IAD86 (cuadrado gris) y PHYTC57 (triángulo negro) que muestra que PHYTC57 retiene la actividad durante un período de tiempo más largo de lo que lo hace OSA-IAD86. (C) Análisis por SDS-PAGE de la estabilidad de PHYTC57 y OSA-IAD86 para la incubación con fluido gástrico simulado. El tiempo de incubación se muestra en segundos y las posiciones de las proteínas marcadoras (M) se indican a la izquierda en kDa con la posición de la pepsina del ensayo y la cistatina indicada. (D) Análisis de transferencia Western de PHYTC57 después del tratamiento con fluido gástrico simulado durante los tiempos indicados (segundos) usando un anticuerpo anti-etiqueta His. Las posiciones de las proteínas marcadoras (M) se indican en kDa.
- La Figura 5. Comparación de PHYTC57 y OSA-IAD86 que muestra las diferencias de aminoácidos entre las proteínas PHYTC57 y OSA-IAD86. (A) Alineación de secuencias de proteínas. (B) Representación de la posición de los cambios de aminoácidos de (A) en la estructura tridimensional de OSAI con los cambios de residuos etiquetados. La región de unión también se muestra con los bucles del extremo terminal N, QVVAG y PW etiquetados.
- La Figura 6 muestra una estructura cristalina de una Adhirona aislado de la biblioteca y la secuencia de un armazón de Adhirona derivado de PHYTC57. (A) Se muestra el armazón de Adhirona y se indican los bucles insertados. (B) Secuencia de ácido nucleico de codón optimizado y la secuencia de aminoácidos para el armazón de aminoácidos de Adhirona 81. La hélice alfa, las láminas beta y las regiones de inserción para el bucle y el bucle2 están resaltadas.
- La Figura 7 muestra la caracterización bioquímica del armazón de Adhirona y la secuenciación de la biblioteca. (A) Se realizó calorimetría diferencial de barrido para determinar la temperatura de fusión del armazón de Adhirona (T_m 101 °C). (B) Se utilizó dicroísmo circular para examinar la estructura del armazón de Adhirona y de tres proteínas Adhirona seleccionadas que contienen inserciones de bucle y todas muestran una estructura β muy alta. (C) La biblioteca de fagos de Adhirona se usó para infectar células *E. coli* ER2738. Se aislaron y secuenciaron 96 clones aleatorios. El gráfico representa el porcentaje de cada aminoácido dentro de las regiones del bucle. Una biblioteca ideal contendría 5,26% de cada aminoácido; la cisteína no se incluyó en la biblioteca.
- La Figura 8 muestra una comparación de la estabilidad de (A) el armazón de Adhirona comparado con (B) una proteína representativa pequeña soluble bien caracterizada, la lisozima, por calorimetría diferencial de barrido. (C) muestra que para una Adhirona seleccionado para unirse a un anticuerpo myc, la adición de los bucles en el

armazón reduce la T_m a 85 °C, pero esto todavía representa una temperatura de fusión mayor que la mayoría de las proteínas de armazón. Esta proteína Adhirona puede experimentar ciclos repetidos de desnaturalización y renaturalización como se muestra en la serie de escaneos.

- La Figura 9 muestra los resultados de ELISA en fagos para SUMO de levadura (ySUMO). (A) ELISA de fagos utilizando 24 clones aislados de la tercera ronda panorámica. Los fagos producidos por cada clon se incubaron en pozos que contenían ySUMO o control. La imagen se grabó tres minutos después de la adición del sustrato 3,3',5,5'-tetrametilbencidina (TMB). (B) Gráfico que muestra la absorbancia a 560 nm de ELISA de fagos para pozos con ySUMO (sombreados) y de control (blancos).
- La Figura 10 muestra la purificación y caracterización de una Adhirona específico para SUMO de levadura (Ad-ySUMO). (A) Ad-ySUMO se expresó en células BL21 y los lisados celulares se calentaron a 50 °C, 60 °C, 70 °C, 80 °C, 90 °C y 100 °C durante 20 minutos y el precipitado se retiró por centrifugación a 15.000 xg. Se separaron alícuotas de 5 µL de lisados aclarados para cada temperatura en un gel SDS-PAGE al 15% y se tiñeron con Coomassie para visualizar las proteínas. (B) Los lisados se incubaron con perlas de Ni-NTA durante 1 hora. El lisado posterior a la incubación (5 µL) y el Ad-ySUMO purificado (10 µL) se corrieron en un gel SDS-PAGE al 15% y se tiñeron con Coomassie para su visualización. (C) Se usaron Ad-ySUMO biotinilados para detectar ySUMO (barras sombreadas) y no detectaron SUMO humana (barras blancas) mediante ELISA. (D) Transferencias Western usando clones de Ad-ySUMO biotinilados 10, 15, 20 y 22 contra 0,5 µg de SUMO de levadura (panel superior) y mezclados con 20 µg de lisado de células HEK293 (panel inferior).
- La Figura 11 muestra los resultados de ELISA de fagos para Adhironas identificados en cribados contra una serie de objetivos, es decir, la proteína del factor de crecimiento FGF1 (A), un receptor de superficie celular CD31 (B), y un péptido (C). Los gráficos representan las lecturas de absorbancia de cada pozo después de la adición de TMB. Los pozos que contienen el objetivo se muestran como barras sombreadas, mientras que los pozos de control se muestran como barras blancas.
- La Figura 12 muestra imágenes de inmunofluorescencia de GFP E5 del HPV16 (objetivo) y GFP E5 del HPV16 sin epítipo para Adhirona (control). Adhirona E5 se conjugó con puntos cuánticos y se usó para detectar la proteína E5 en las células de mamíferos. Las células se tiñeron con DAPI (tinción de ADN), GFP y E5 (con los puntos cuánticos de Adhirona).
- La Figura 13 muestra una Adhirona dirigido a hSUMO2. (A) Se probó un ABP generado contra hSUMO2 para determinar la especificidad para hSUMO1 y hSUMO2 mediante ELISA. El hSUMO2 ABP se unió específicamente a hSUMO2, que se reflejó en la afinidad de unión. (B) Transferencia que muestra la capacidad de hSUMO2 ABP para inhibir la actividad ubiquitina ligasa dirigida a SUMO de RNF4. (C) Un vector de control o aglutinante de hSUMO2 se expresó en células y se analizó usando un anticuerpo anti-FLAG y se cultivó con arsénico para inducir cuerpos nucleares, PML (verde). RNF4 promueve la degradación de PML. El bloqueo de la interacción entre hSUMO2 y RNF4 altera la degradación de PML. Esta es la primera descripción de una proteína de unión a hSUMO2 que se une e inhibe específicamente a hSUMO2 sin interactuar con hSUMO1. La Adhirona hSUMO2 también bloquea específicamente el dominio con el que está interactuando y no afecta a otras funciones del hSUMO2, como se demuestra en (C).
- La Figura 14 muestra un gráfico que representa un ensayo de formación de coágulos sanguíneos y turbidez de lisis. La línea continua representa la formación normal y la lisis. Las líneas discontinuas representan el efecto de cinco Adhironas diferentes en este proceso. Los cinco Adhironas diferentes contienen epítipos diferentes y están teniendo diferentes efectos sobre la formación de coágulos y la lisis. Algunos prolongan el tiempo de coagulación, algunos prolongan el tiempo de lisis y otros prolongan el tiempo de coagulación y lisis. Uno de las Adhironas está completando la inhibición de la formación de coágulos. Esto demuestra que los diferentes Adhironas se unen e inhiben diferentes regiones de fibrinógeno y, por lo tanto, representan una forma realmente novedosa de estudiar la función de las proteínas.
- La Figura 15 muestra una imagen confocal de fibrinógeno marcado con FITC después de la formación del coágulo con una Adhirona de unión a fibrinógeno y una Adhirona de control que no se une al fibrinógeno. Esto demuestra la capacidad de la Adhirona para modificar la respuesta normal del coágulo.
- La Figura 16 muestra imágenes de fluoroscopia que demuestran la expresión de Adhironas funcionales en células de mamíferos. Se puede observar que la Adhirona de unión a SUMO2 humano altera la degradación de la fosfoproteína nuclear PML que conduce a un aumento en estos cuerpos nucleares de PML.
- La Figura 17 muestra la estructura de cocristales de FcγRIIIa y Adhirona unido. (A) La Adhirona se une a un sitio alostérico que afecta la unión del receptor a IgG. Este sitio también se ha identificado como un objetivo para el diseño de fármacos de molécula pequeña que proporciona evidencia de que las Adhironas pueden proporcionar información importante sobre sitios farmacológicos. (B) También se ha encontrado que las Adhironas se dirigen al sitio de unión directa de IgG en FcγRIIIa. La Adhirona contenía dos bucles y una secuencia del extremo terminal N no estructurada de 4 aminoácidos. Los péptidos del extremo terminal N también contribuyen a la interacción entre Adhirona y FcγRIIIa. La información obtenida al comprender la interacción de unión entre Adhirona y FcγRIIIa ayudará a identificar moléculas pequeñas a través de la detección *in silico*. Esto proporciona un enfoque novedoso e intrigante para futuras metodologías de descubrimiento de fármacos. Además, las estructuras cristalinas demuestran la capacidad del armazón para presentar bucles que pueden extender las láminas beta del armazón central o bucles más flexibles que crean superficies de interacción conformacional alternativas. Esto se ve facilitado por la estabilidad del núcleo inherente del armazón, que potencialmente hace que este armazón sea único.
- Figura 18 Espectros de RMN. (A) Superposición de espectros de huellas dactilares por HSQC 1H-15N para el armazón de ABP (gris claro) y la Adhirona 15 de SUMO de levadura (gris oscuro). (B) Superposición del espectro

de huellas dactilares de HSQC 1H-15N para Adhirona 15 de SUMO de levadura (gris oscuro) y el espectro HSQC TROSY 1H-15N para la proteína SUMO de levadura y el complejo SUMO de levadura-Adhirona 15 (gris claro).

- La Figura 19 muestra un gráfico del cambio de impedancia frente a la concentración de Adhirona de unión a SUMO. Esto proporciona un ejemplo del uso potencial de Adhironas en dispositivos biosensores basados en impedancia y muestra que las Adhironas unidos a una superficie tienen un rango dinámico mayor en comparación con los anticuerpos.
- Las Figuras 20 a 24 muestran alineamientos de secuencia de las regiones BUCLE1 y BUCLE2 de Adhironas seleccionados contra un rango de objetivos. Este análisis permite la identificación del rango de aglutinantes contra un objetivo dado y facilita el desarrollo de regiones de unión de consenso potenciales en algunos casos.
- La Figura 20 muestra un alineamiento de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen al receptor 1 de LDL oxidado tipo Lectina (LOX1).
- La Figura 21 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen a la Hormona de Crecimiento Humano (HGH).
- La Figura 22 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen al modificador tipo ubiquitina pequeña de levadura (SUMO).
- La Figura 23 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen a la proteína 2a de unión a penicilina (PBP2a).
- La Figura 24 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen a un objetivo peptídico.
- La Figura 25 muestra los resultados de ELISA de fagos para Adhironas identificados en cribados contra un compuesto orgánico, posaconazol. Los gráficos representan las lecturas de absorbancia de cada pozo después de la adición de TMB. Los pozos que contienen el objetivo se muestran como barras sombreadas, mientras que los pozos de control se muestran como barras blancas.
- Figura 26. Un gráfico del cambio de detección basado en la impedancia de varias concentraciones de fibrinógeno desde micromolar a attomolar por una Adhirona de unión a fibrinógeno. Esto proporciona un ejemplo del uso potencial de Adhironas en dispositivos biosensores basados en impedancia y muestra que las Adhironas unidos a una superficie pueden detectar bajas concentraciones de proteína objetivo dentro de una incubación de 15 minutos y mostrar un rango dinámico lineal grande. Los dos puntos de datos a 3×10^{-12} micromolar se midieron inmediatamente después de agregar el analito (punto de datos inferior) y después de una incubación de 15 minutos (punto de datos superior).
- La Figura 27 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen al dominio de homología 2 Src de la proteína 2 de unión al receptor del factor de crecimiento (Grb2).
- La Figura 28 muestra una alineación de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen al dominio de homología 2 Src del transductor de señal y del activador de transcripción 3 (STAT3).

Se ha establecido que "para generar un interés más amplio en un almacén de proteína particular, es necesario demostrar que se pueden generar especificidades para diferentes clases de ligandos relevantes, que las proteínas de unión derivadas son prácticamente útiles y que ofrecen al menos algunos beneficios sobre los fragmentos de anticuerpos convencionales. Estos criterios no se cumplen en muchos de los almacenes de proteínas propuestos hasta ahora y para la mayoría de ellos se han descrito simplemente esfuerzos iniciales de manipulación" (Skerra 2007). Por lo tanto, la presente invención tiene un claro valor al proporcionar un almacén de proteína útil, versátil y atractivo, como se demuestra a continuación.

El alto nivel de éxito en términos de identificación de proteínas de unión bioactivas importantes atribuibles al nuevo almacén de Adhirona y a la biblioteca derivada se debe en gran parte a la muy alta estabilidad del almacén central que proporciona un marco altamente rígido sobre el cual las regiones de bucle variable pueden desplegarse. Estas regiones bucle tienen suficiente flexibilidad para adoptar una gama de conformaciones y, por lo tanto, pueden interactuar con una amplia gama de características conformacionales en las moléculas objetivo, lo que permite la selección de reactivos de unión frente a una amplia gama de objetivos moleculares, incluidas las moléculas orgánicas pequeñas notables. Estos aspectos estructurales de este almacén único conducen a resultados funcionales que no se han logrado con otras proteínas de almacén.

También puede haber beneficios en el uso de una proteína vegetal para muchas aplicaciones, particularmente aquellas que involucran humanos ya que la proteína no se deriva de una proteína humana y por lo tanto no estará involucrada en interacciones naturales con proteínas humanas. Las únicas interacciones potenciales serían contra las cisteína proteasas, pero las regiones de unión activa que podrían interactuar con tales proteínas generalmente se reemplazan o eliminan en los almacenes de Adhirona. Sin embargo, los humanos entran en contacto y toleran constantemente las proteínas derivadas de plantas a través de, por ejemplo, alimentos, productos cosméticos y composiciones medicinales.

Creación de un ejemplo de proteína de almacén (denominada PHYTC57)

Introducción

Un enfoque de consenso para el diseño de proteínas comienza con una alineación de secuencia múltiple de miembros de una familia de proteínas para derivar una secuencia de consenso única en la que cada posición está normalmente

ocupada por el residuo que ocurre con mayor frecuencia. Los residuos que definen la estructura, la vía de plegamiento y la estabilidad de la proteína plegada tenderán a conservarse, mientras que los que se requieren para una función biológica común, tal como los residuos catalíticos en una enzima, o los residuos que interactúan con una proteína objetivo conservada, también es probable que se conserven. Una proteína natural generalmente no contendrá todos los residuos de consenso conservados porque las proteínas solo evolucionan para ser lo suficientemente estables como para realizar su función biológica *in vivo*, y en muchos casos esto puede incluir un grado de inestabilidad para facilitar la renovación y la regulación de los procesos biológicos (Steipe, Schiller et al., 1994).

Se estaba interesado en explorar si se podría aplicar un enfoque de consenso para mejorar las propiedades inhibitorias y la estabilidad de las fitocistatinas. Las fitocistatinas son inhibidores de proteínas pequeñas (~100 aa) de cisteína proteasas (Kondo et al., 1991). Un análisis filogenómico detallado de la superfamilia de cistatina revela las relaciones entre las diferentes clases de cistatinas (Kordis y Turk 2009). Las fitocistatinas comprenden tres regiones que participan en la unión a las cisteína proteasas, la región del extremo terminal N, un bucle QVVAG y un bucle PW (Margis et al., 1998). Anteriormente, con base en un modelo estructural y alineamientos de secuencia, se generó una variante de cistatina de arroz, orizacistatina I (OSA-I u OC-I), que carece del residuo Asp 86, cerca del bucle PW conservado y denominado OSA-IAD86. Esta variante mostró una mejora de 13 veces en K_i contra la papaína y la cisteína proteasa GCP-1 específica del intestino de *Caenorhabditis elegans* (Urwin et al., 1995; McPherson et al., 1997; Urwin et al., 1997; Urwin et al., 1998; Urwin et al., 2000; Urwin et al., 2001; Urwin et al., 2003; Lilley et al., 2004). Luego se demostraron buenos niveles de resistencia transgénica, contra nematodos de plantas, conferidos por OSA-IAD86 y otras cistatinas cuando se expresan en las células de alimentación especializadas que se desarrollan dentro de una planta parasitada por nematodos (Urwin et al., 1995; McPherson et al., 1997; Urwin et al., al., 1997; Urwin et al., 1998; Urwin et al., 2000; Urwin et al., 2001; Urwin et al., 2003; Lilley et al., 2004).

En el presente documento se describe el diseño, construcción y caracterización de una fitocistatina de consenso basada en la alineación de secuencias múltiples de las secuencias de aminoácidos de 57 fitocistatinas. Se mostró que esta fitocistatina de consenso muestra una inhibición eficaz de la cisteína proteasa de papaína y una mayor estabilidad en los ensayos de termoestabilidad y digestibilidad.

Materiales y métodos

Diseño de la fitocistatina de consenso denominada a continuación PHYTC57

Después de una búsqueda en tBLAST de la base de datos GenBank, se realizó una alineación múltiple de secuencias de fitocistatina usando CLUSTALW (<http://clustalw.genome.ad.jp/>; BLOSUM62; penalización por apertura de huecos = 12; penalización por extensión de huecos = 2) con una mayor alineación manual utilizando el programa CINEMA (<http://www.bioinf.manchester.ac.uk/dbbrowser/CINEMA2.1/>) (Parry-Smith et al., 1998; Lord et al., 2002). Se determinó el aminoácido más comúnmente utilizado en cada posición y se truncaron los extremos terminales N y C variables para obtener una proteína consenso de 95 aminoácidos de longitud. Se diseñó una región de codificación sintética para codificar la fitocistatina de consenso (phytc57) con el uso de codones optimizados para la expresión en *E. coli* utilizando la tabla de frecuencias de codones apropiada (<http://www.kazusa.or.jp/codon/>). La región de codificación se construyó a partir de seis oligonucleótidos (P1 - P6; véase la Figura 1) cada uno de aproximadamente 70 nt de longitud que incluía regiones de al menos 10 nt solapados entre oligonucleótidos adyacentes. Los oligonucleótidos flanqueantes P1 y P6 también contenían un sitio *Sfi*I y un sitio *Not*I, respectivamente, para la clonación en el vector pDHISII en el que la secuencia de codificación se fusionó con la región de codificación de la secuencia señal *pelB* derivada del vector. La retención de un sitio de escisión de la peptidasa señal se verificó sometiendo la secuencia del extremo terminal N de la proteína PELB/PHYTC57 al programa automático de predicción de secuencia señal Signal P (Bendtsen et al., 2004); (<http://www.cbs.dtu.dk/services/SignalP/>). Los pares de oligonucleótidos (P1 + P2), (P3 + P4) y (P5 + P6) se hibridaron y convirtieron en fragmentos bicatenarios por autocebado. Las reacciones (10 μ L) contenían 25 pmoles de cada cebador, 0,2 mM de cada dNTP, tampón 1x Pwo (Tris-HCl 10 mM pH 8,85, KCl 25 mM, $(\text{NH}_4)_2\text{SO}_4$ 5 mM, MgSO_4 2 mM) y 5 unidades de ADN polimerasa Pwo (Boehringer) y se sometieron a un ciclo de 95 °C durante 2 min, 25 °C durante 4 min y luego 72 °C durante 5 min. La posterior unión de los productos (P1/P2) a (P3/P4) y luego (P1/P2/P3/P4) a (P5/P6) se realizó de manera idéntica, para generar un producto de longitud completa. Se usó una parte alícuota de 5 μ L de la reacción final como molde en una PCR de 50 μ L junto con 50 pmoles de cada uno de los cebadores flanqueantes P1 y P6, 0,2 mM de cada dNTP, tampón 1x Pwo y 5 unidades de Pwo. La reacción se sometió a 95 °C durante 1 minuto, luego 30 ciclos de 95 °C, 30 segundos, 55 °C, 30 segundos y 72 °C, 30 segundos. El producto se digirió con *Sfi*I y *Not*I, se recuperó de un gel de agarosa y se clonó en pDHISII restringido con *Sfi*I y *Not*I. La región génica recién construida fue secuenciada para confirmar el correcto ensamblaje del cebador y la secuencia de ADN esperada. La secuencia de la región de codificación de consenso se muestra en la Figura 1A.

Síntesis de otros genes de fitocistatina

Las secuencias de ADN sintético optimizadas para la expresión de *E. coli* se diseñaron, construyeron y clonaron de manera similar para fitocistatinas de arroz (*Oryza sativa*; *osa-IAD86*, una forma modificada OC-I GenBank número de acceso U54702 que carece del codón para Asp 86 (Urwin et al., al., 1995), naranja satsuma (*Citrus unshiu*; *cun*, GenBank número de acceso C95263) y papaya (*Carcia papaya*; *cpa*, GenBank número de acceso X71124 (Song et

al., 1995). Se amplificó la secuencia del gen de cistatina de clara de huevo de pollo (*cewc*) a partir de un plásmido recombinante derivado de pQE30 usando los cebadores específicos de genes:

cewcF 5'ATTA GCGGCCAGCCGGCC ATGGCCAGCGAGGACCGCTCCCGGC3' (SEQ ID NO 10)

cewcR 5'CGCTGTACTTGCGGCCGCCCTGGCACTTGCTTTCCAGC3' (SEQ ID NO 11)

que introdujo un sitio *SfiI* y un sitio *NotI* (subrayado) respectivamente.

Las reacciones de PCR se llevaron a cabo con 50 pmoles de cada cebador, una concentración final de 0,2 mM de cada dNTP (Promega) y aproximadamente 10 ng de plantilla en un volumen de 50 µL que contiene tampón 1x SuperTaq (Tris-HCl 10 mM, pH 9,0, MgCl₂ 1,5 mM, KCl 50 mM, Triton X-100 al 0,1% (v/v)). Para asegurar una amplificación de alta fidelidad durante la PCR, se usó una mezcla 10:1 de SuperTaq (HT Biotechnology): ADN polimerasas Pfu (Boehringer Mannheim) con 1 unidad de mezcla de polimerasa por reacción. Las reacciones se sometieron a 1 ciclo de 95 °C durante 1 minuto y luego a 20 ciclos de 95 °C durante 30 segundos, 55 °C durante 30 segundos y 72 °C durante 30 segundos. Se usó una etapa final de 72 °C durante 30 segundos para asegurar que todos los productos fueran de longitud completa.

Construcción de pDHISII

Las regiones codificantes de fitocistatina se clonaron inicialmente en una versión modificada del vector fagémido pHEN1 (Hoogenboom et al., 1991) como fusiones del gen III. El vector pHEN se modificó mediante la adición de una región de hexahistidina codificada por oligonucleótidos complementarios que se fosforilaron e hibridaron para crear un enlazador con extremos de cadena sencilla apropiados para la unión con el vector escindido *NotI*. Las secuencias de oligonucleótidos fueron;

HisF: 5'-GGCCGC AGAGGATCGCATCACCATCACCATCACGG-3' (SEQ ID NO 12)

HisR: 5'-GGCCCC GTGATGGTGATGGTGATGCGATCCTCTGC-3' (SEQ ID NO 13) y los extremos complementarios *NotI* están subrayados.

El vector pHEN1 se digirió con *NotI* y se desfosforiló utilizando 5 unidades de fosfatasa alcalina de camarones (NEB) en una reacción de 50 µL que contenía NaCl 100 mM, Tris-HCl 50 mM (pH 7,9), MgCl₂ 10 mM y DTT 1 mM durante 30 minutos a 37 °C antes de calentar a 65 °C durante 15 min. El sitio distal *NotI* fue destruido por mutagénesis por PCR utilizando los cebadores

XhoF: 5'ATCACG CTCG AGCAGAACAAAACTCATCTCAG3' (SEQ ID NO 14)

XhoR: 5'TGTTCTG CTCGAG CGTGATGGTGATGGTGATGGCG3' (SEQ ID NO 15)

BamHIR: 5'TGGCCTTGATATTACAAACG3' (SEQ ID NO 16)

M13R: 5'AGCGGATAACAATTTACACAGGA3' (SEQ ID NO 17).

El cebador XhoF se usó junto con el cebador BamHIR y el XhoR junto con el cebador M13R para generar dos fragmentos que se hibridaron y amplificaron en una segunda PCR que incluía los cebadores M13R y BamHIR. El sitio de restricción *XhoI* introducido está subrayado. El producto resultante se clonó como un fragmento *SfiI/BamHI* en un vector pHEN1 digerido de manera similar. La presencia del sitio *XhoI* se cribó mediante análisis de restricción de ADN de fagémido aislado y el inserto en un clon positivo, pDHISII (Figura 1B), se confirmó mediante análisis de secuencia de ADN.

Expresión de cistatinas en pDHISII

Inicialmente, la expresión de cistatinas se realizó usando constructos en el vector pDHISII que agrega RGS(H)₆ al extremo terminal C y etiquetas myc. Los estudios de expresión en la cepa huésped HB2151 de *E. coli* permiten la supresión del codón ámbar dentro de la fusión cistatina-gen III, lo que resulta en la acumulación de cistatina en el periplasma. Los cultivos (1 L) se desarrollaron en medio 2xTY con glucosa al 0,1% (v/v) y 100 µg/mL de ampicilina, luego se indujeron mediante IPTG a 1 mM y se cultivaron a 30 °C, 16 horas. Los sedimentos celulares se resuspendieron en 20 mL de Tris 50 mM pH 8, sacarosa al 20% (p/v) a 4 °C y se realizaron preparaciones periplásmicas. A la fracción periplásmica se añadió 1 mL de resina Ni-NTA (GE Healthcare) y se incubó con mezcla durante 16 horas a 4 °C. La resina se lavó 8 veces con alícuotas de 1 mL de tampón de lavado (NaHPO₄ 50 mM, NaCl 500 mM, pH 6), luego 3 veces con tampón de lavado que contenía imidazol 40 mM. La cistatina se eluyó con 200 µL de tampón de lavado que contenía imidazol 250 mM a 4 °C durante 1 hora, antes de la diálisis contra PBS. La proteína se dividió en partes alícuotas y se congeló en nitrógeno líquido. Las muestras se analizaron mediante SDS-PAGE y espectrometría de masas con electroaspiración.

Expresión de cistatinas en pET101

Para una expresión más eficiente de la proteína soluble, los genes de fitocistatina se subclonaron junto con la etiqueta 6-His en el extremo terminal C de pDHISII en pET101 por clonación TOPO direccional (Invitrogen). Los cebadores

cystaSDFWD 5'CACCATGAAATCACTATTGCTTACG3' (SEQ ID NO 18)

cystaSDREV 5'CTACTAGTGATGGTGATGGTGATGCG3' (SEQ ID NO 19)

se usaron para amplificar por PCR los genes de fitocistatinas del vector de clonación original pDHisII usando las condiciones descritas anteriormente. Los clones positivos se confirmaron mediante análisis de secuencia de ADN y se introdujeron en células Star BL21 (DE3) (Invitrogen). Los cultivos se desarrollaron a 30 °C y se indujo la expresión de fitocistatinas mediante la adición de IPTG (a 1 mM) durante 16 h. Las células se recogieron por centrifugación (4.000 x g, 10 min), se sonicaron en hielo (3 x 1 min), los residuos celulares se sedimentaron por centrifugación (10.000 x g, 10 min). El sobrenadante se cargó en una columna quelante de metal (Pharmacia) cargada con NiCl₂ 0,1 mM. Después de un lavado exhaustivo, las fitocistatinas se eluyeron con imidazol 100 mM y se dializaron extensamente en tampón HBS (HEPES 10 mM, pH 7,4, NaCl 150 mM, P20 al 0,005%) y se almacenaron a -70 °C.

Ensayo de papaína

La inhibición de la papaína (Sigma) se ensayó utilizando pGlu-Phe-Leu-p-NA (Sigma), un sustrato sintético para cisteína proteinasas (Filippova et al., 1984). Se añadieron 100 ng de papaína, en 50 µL de tampón de incubación (MES/OH 0,15 M, pH 5,8, NaEDTA 4 mM, DTT 4 mM), a 50 µL de muestra, sin cistatina o con una concentración conocida de cistatina y previamente incubada por 30 minutos. Se añadieron 50 µL de pGlu-Phe-Leu-p-NA 1 mM (en tampón de incubación y DMSO al 30%) y se monitorizó la actividad de la papaína mediante un aumento lineal de OD a 415 nm, durante 15 minutos a 25 °C.

Experimentos de resonancia de plasmón superficial (SPR)

La papaína se obtuvo de Sigma como un polvo liofilizado. Para los experimentos de SPR, se preparó una solución patrón de 1 mg/mL disolviéndola en tampón de HBS que contenía DTT 1 mM. Este patrón se diluyó a las concentraciones requeridas con el mismo tampón. Todos los experimentos de SPR se realizaron en un instrumento BIACORE 3000 utilizando un chip sensor NTA. El tampón de proceso fue HBS-EP (HEPES 10 mM, NaCl 0,15 M, P20 al 0,005%, pH 7,4) y todos los experimentos se llevaron a cabo a 25 °C. Las fitocistatinas se inmovilizaron (aproximadamente 400 unidades de respuesta (RU) en celdas de flujo 2 a 4 con la celda de flujo 1 con blanco. Para controlar la unión de la papaína a las fitocistatinas, se inyectaron 240 µL de tampón de HBS sobre todas las celdas de flujo a un caudal de 80 µL/min (fase de asociación) seguido de flujo de tampón normal durante 3 min (fase de disociación) para proporcionar un blanco de sustracción. Esta inyección se repitió con papaína. Las proteínas se extrajeron del chip sensor usando EDTA y la superficie se volvió a cargar con níquel antes de repetir el experimento con una concentración diferente de papaína. Se realizaron al menos 5 ciclos para cada concentración para permitir el análisis cinético. Los datos de estos experimentos de unión se analizaron utilizando el paquete de software Biaevaluation3 (BIACORE) y ajustando globalmente los datos al modelo de unión de Langmuir 1:1.

Estabilidad térmica de las fitocistatinas.

Se prepararon alícuotas de PHYTC57 y OSA-IAD86 a razón de 0,5 µg de proteína/mL en tampón de fosfato 50 mM, pH 7,4. Las muestras se colocaron inmediatamente en un baño de agua hirviendo. Las muestras se retiraron en diversos momentos y se sumergieron en nitrógeno líquido, antes de almacenarlas a -70 °C. La actividad inhibitoria residual, para determinar el nivel restante de cistatina funcional, se determinó mediante el ensayo de inhibición de papaína.

Ensayo de digestibilidad

El fluido gástrico simulado se preparó como se describe (Astwood et al., 1996) inmediatamente antes de su uso y contenía pepsina al 0,32% (p/v) (Sigma) en NaCl 0,03 M ajustado a pH 1,2 con HCl concentrado. Se incubaron cuatro réplicas de PHYTC57 y de OSA-IAD86 a una concentración final de 0,2 mg/mL en SGF a 37 °C. A intervalos, se retiraron alícuotas de 200 µL y se mezclaron inmediatamente con 75 µL de Na₂CO₃ 0,2 M para terminar la digestión. Cada muestra se dividió en tres submuestras y, como se describió anteriormente (Atkinson et al., 2004), se separaron en geles SDS-PAGE al 15%. Se tiñeron dos geles para la proteína total con azul de Coomassie, uno para mostrar el efecto del fluido gástrico simulado en PHYTC57 y el segundo para comparar la digestibilidad de PHYTC57 con OSA-IAD86. El tercer gel se sometió a análisis de transferencia Western usando un anticuerpo de ratón para detectar la etiqueta 6His (Qiagen) de PHYTC57 con visualización usando actividad de fosfatasa alcalina. Se analizó una muestra final de cada punto temporal para determinar la actividad inhibitoria residual contra la papaína mediante el ensayo enzimático como se describió anteriormente.

Resultados

Síntesis de genes de fitocistatina

Para el diseño de la región de codificación de fitocistatina de consenso se realizó una búsqueda en tBLASTN de la base de datos GenBank usando OSA-I (*Oryza sativa*; U54702), ZMA2 (*Zea mays*; D38130) y secuencias de proteína HAN1 (*Helianthus annuus*; Q10993) como sondas de búsqueda. La lista de secuencias utilizadas para derivar la secuencia de consenso se muestra en la Tabla 1. Las secuencias se identificaron a partir de bases de datos mediante búsqueda de homología. La tabla muestra un nombre sistemático para cada cistatina junto con el nombre del

organismo y el nombre común de la planta y el número de acceso del GenBank. Las secuencias de codificación se tradujeron y alinearon usando el programa CLUSTALW y la alineación se mostró usando el programa CINEMA (Parry-Smith et al., 1998; Lord et al., 2002) para permitir la mejora mediante la alineación manual. Luego se obtuvo una secuencia de consenso identificando el aminoácido más común en cada posición (Figura 1A). La longitud de la proteína de consenso se fijó en 95 aminoácidos con el extremo terminal N colocado cuatro residuos antes del residuo de glicina del extremo terminal N conservado, y por lo tanto antes de la primera cadena β ($\beta 1$). El extremo terminal C se fijó 15 residuos después del motivo PW conservado y, por lo tanto, después de la última cadena β ($\beta 5$). Estos criterios se basaron en las estructuras de rayos X de CEWC (Bode et al., 1988) y estefina B humana (Stubbs et al., 1990) y la estructura de RMN de OSA-I (Nagata et al., 2000).

Tabla 1. Secuencias de fitocistatina utilizadas para derivar la secuencia de consenso (PHYTC57)

Fitocistatina	Nombre del organismo	Nombre común	Número de acceso
AAR	<i>Ambrosia artemisiifolia</i>	Ambrosía corta	L16624
ACE	<i>Allium cepa</i>	Cebolla	AA508918
ATH1	<i>Arabidopsis thaliana</i>	Arabidopsis	Z17618
ATH2	<i>Arabidopsis thaliana</i>	Arabidopsis	Z97341
ATH3	<i>Arabidopsis thaliana</i>	Arabidopsis	Z17675
ATH4	<i>Arabidopsis thaliana</i>	Arabidopsis	ATAJ110
ATH6	<i>Arabidopsis thaliana</i>	Arabidopsis	AC002409
ATH8	<i>Arabidopsis thaliana</i>	Arabidopsis	Z37263
AVU	<i>Artemisia vulgaris</i>	Artemisa	AF143677
BCA1	<i>Brassica campestris</i>	Col china	L41355
BCA2	<i>Brassica campestris</i>	Col china	L48182
BCA3	<i>Brassica campestris</i>	Col china	U51119
CPA	<i>Carcia papaya</i>	Papaya	X71124
CSA	<i>Cucumis sativus</i>	Pepino	AB014760
CSAT	<i>Castanea sativa</i>	Castaña	AJ224331
CUN	<i>Citrus unshiu</i>	Naranja Satsuma	C95263
DCA	<i>Daucus carota</i>	Zanahoria	D85623
DCAR	<i>Dianthus caryophyllus</i>	Clavel	AF064734
GHI2	<i>Gossypium hirsutum</i>	Algodón	A1728662
GHI3	<i>Gossypium hirsutum</i>	Algodón	A1726250
GMA1	<i>Glycine max</i>	Soja	D64115
GMA2	<i>Glycine max</i>	Soja	U51583
GMA3	<i>Glycine max</i>	Soja	U51855
GMA4	<i>Glycine max</i>	Soja	U51854
GMA5	<i>Glycine max</i>	Soja	A1495568
GMA7	<i>Glycine max</i>	Soja	A1938438
HAN1	<i>Helianthus annuus</i>	Girasol	Q10993
IBA	<i>Ipomoea batatas</i>	Batata	AF117334
MCR1	<i>Mesembryanthemum crystallinum</i>	Escarchada	AA856241
MCR2	<i>Mesembryanthemum crystallinum</i>	Escarchada	AA887617
MDO	<i>Malus domestica</i>	Manzano	AT000283
OSA1	<i>Oryza sativa</i>	Arroz	U54702
OSA2	<i>Oryza sativa</i>	Arroz	X57658
OSA5	<i>Oryza sativa</i>	Arroz	C25431
PAM	<i>Persea americana</i>	Aguacate	JH0269
PBA	<i>Populus balsamifera</i>	Álamo	A1167046
PCO	<i>Pyrus communis</i>	Pera	U82220
PTA	<i>Pinus taeda</i>	Pino	A1812403
PTR	<i>Populus tremula</i>	Álamo	A1162398
RCO1	<i>Ricinus communis</i>	Semillas de ricino	Z49697
RCO2	<i>Ricinus communis</i>	Sorgo	T23262
SBI	<i>Sorghum bicolor</i>	Sorgo	X87168
SLA	<i>Silene latifolia</i>	Campion blanca	Z93053
SLY1	<i>Lycopersicon esculentum</i>	Tomate	AF083253

SLY2	<i>Lycopersicon esculentum</i>	Tomate	X73986
SLY5	<i>Lycopersicon esculentum</i>	Tomate	AI1781497
STU1	<i>Solanum tuberosum</i>	Patata	L16450
STU2	<i>Solanum tuberosum</i>	Patata	L16450
STU3	<i>Solanum tuberosum</i>	Patata	L16450
STU4	<i>Solanum tuberosum</i>	Patata	L16450
STU5	<i>Solanum tuberosum</i>	Patata	L16450
STU6	<i>Solanum tuberosum</i>	Patata	L16450
STU7	<i>Solanum tuberosum</i>	Patata	L16450
STU8	<i>Solanum tuberosum</i>	Patata	L16450
STU10	<i>Solanum tuberosum</i>	Patata	X74985
VUN	<i>Vigna unguiculata</i>	Caupí	Z21954
ZMA1	<i>Zea mays</i>	Maíz	D10622
ZMA2	<i>Zea mays</i>	Maíz	D38130
ZMA4	<i>Zea mays</i>	Maíz	AI001246
ZMA5	<i>Zea mays</i>	Maíz	AI740162

El gen sintético que codifica PHYTC57 se generó a partir de seis oligonucleótidos (P1-P6) cada uno de aproximadamente 70 nt, diseñados para incluir regiones de al menos 10 nt solapados entre oligonucleótidos adyacentes como se describe en Materiales y métodos. Para el estudio comparativo de las fitocistatinas naturales, se adoptó un enfoque genético sintético similar utilizando oligonucleótidos para las regiones codificadoras de fitocistatina osa-1ΔD86, *cun* (Citrus unshiu; Satsuma) y *cpa* (Carcia papaya; Papaya) diseñadas a partir de la secuencia única de la proteína apropiada con cambios de codón para reflejar el uso del codón de *E. coli*. La región codificante de cistatina de clara de huevo de pollo (cewc) se amplificó por PCR a partir de un plásmido derivado de pQE. Los genes fueron clonados inicialmente en el vector de despliegue en fagos pDHisII explotando los sitios *SfiI* y *NotI* (Figura 1B).

Expresión de cistatina

Las cistatinas se expresaron inicialmente a partir de constructos de pDHisII y se purificaron por cromatografía de afinidad de Ni-NTA. El análisis de las cistatinas purificadas por espectrometría de masas de ionización por electroaspersión indicó que, inesperadamente, en cada caso hubo un truncamiento del extremo terminal C de la proteína expresada. La Tabla 2 muestra las masas esperadas de las cistatinas de longitud completa y aquellas con un truncamiento del extremo terminal C de 16 aminoácidos calculado utilizando el programa Protein Calculator (www.scripps.edu/~cdputnam/protcalc.html) junto con las masas moleculares determinadas. La secuencia del extremo terminal C de las proteínas se muestra con el sitio de truncamiento principal indicado por una flecha. Con la excepción de CUN, la masa molecular determinada está en excelente acuerdo con las formas de las proteínas de fusión que habían perdido los 16 aminoácidos del extremo terminal C, pero que retienen la etiqueta His (Tabla 2). En el caso de CUN hay truncamiento de menos de 16 residuos, pero esto no se caracterizó más. Para asegurar la expresión de secuencias proteicas definidas, las regiones codificantes de cistatina más la etiqueta His se subclonaron en pET101 mediante clonación facilitada por PCR y TOPO y se expresaron en células Star BL21 (DE3) (Stratagene). La expresión de proteínas se indujo mediante la adición de IPTG a 1 mM durante 16 horas y las cistatinas se purificaron en condiciones nativas por cromatografía de afinidad de Ni-NTA en virtud de su etiqueta de hexahistidina en el extremo terminal C. Las cistatinas se separaron mediante SDS-PAGE para examinar su pureza, que se estimó en > 98%, sus masas estaban en excelente acuerdo con los valores esperados por ESI-MS (datos no mostrados) y, por lo tanto, estas proteínas se usaron para análisis adicionales.

Tabla 2. Resultados de la espectrometría de masa con electroaspersión para cistatinas expresadas a partir de HB2151.

Fitocistatina	Masa predicha (Da)		Masa experimental (Da)
	Longitud completa	Truncada	
OSA1ΔD86	14329	12603	12603,3 ± 0,8
CPA	14321	12596	12593,5 ± 6,8
PHYTC57	13855	12129	12128,3 ± 0,8
CEWC	16399	14608	14607,4 ± 1,5
CUN	14222	12496	12728,0 ± 1,1
↓ RGSHHHHHHARAEQKLISEEDLNGAA (SEQ ID NO 20)			

Actividad de unión a cistatina

Los ensayos enzimáticos con papaína usando el sustrato artificial pGlu-Phe-Leu-p-nitroanilida confirmaron que PHYTC57 era un inhibidor activo de la cisteína proteasa. No se determinaron los valores de K_i , pero a partir de los valores de IC_{50} para PHYTC57 y OSA-IAD86 ($4,6 \times 10^{-8}$ M y $1,8 \times 10^{-7}$ M respectivamente) está claro que PHYTC57 es un inhibidor más potente que OSA-IAD86. Para medir directamente la interacción entre las cistatinas y la proteasa, se midieron la cinética de unión mediante el análisis de resonancia de plasmón superficial BIACore. Las cistatinas se inmovilizaron sobre chips sensores recubiertos de níquel mediante las etiquetas His del extremo terminal C. Luego se permitió que la papaína se uniera a la cistatina inmovilizada y se realizaron mediciones a varias concentraciones de papaína. Los sensorgramas para la unión de OSA-IAD86, CPA, CUN, CEWC y PHYTC57 a la papaína se muestran en la Figura 3. Los datos en cada concentración se ajustaron al modelo de unión de Langmuir 1:1 y se determinaron las constantes cinéticas. Los datos mostraron un buen ajuste al modelo consistente con la estequiometría 1:1 conocida de inhibición de cistatina de cisteína proteasas. En la Tabla 3 se muestra un resumen de las constantes cinéticas de estas cistatinas. PHYTC57 muestra una mayor asociación y una cinética de disociación más baja en comparación con las cistatinas de origen natural analizadas, con una constante de equilibrio K_D de $6,3 \times 10^{-12}$ M, lo que indica un complejo de unión fuerte con papaína. Este valor es dos órdenes de magnitud menor que el valor K_D medido para la cistatina de clara de huevo de gallina ($3,9 \times 10^{-10}$ M), tres órdenes de magnitud menor que la fitocistatina mejorada OSA-IAD86 ($4,7 \times 10^{-9}$ M) y cuatro órdenes de magnitud menor que las fitocistatinas CUN ($1,4 \times 10^{-8}$ M) y CPA (2×10^{-8} M).

Tabla 3. Parámetros cinéticos determinados por resonancia de plasmón superficial. Se muestran las constantes de velocidad de asociación y disociación (K_a y K_d) y las constantes de equilibrio (K_A y K_D).

	K_a (1/M.s)	K_d (1/s)	χ^2	K_A (1/M)	K_D (M)
OSAIA D86	$2,6 \times 10^5$	$1,2 \times 10^3$	0,98	$2,1 \times 10^8$	$4,7 \times 10^{-9}$
CUN	$2,6 \times 10^5$	$3,5 \times 10^3$	2,3	$7,3 \times 10^7$	$1,4 \times 10^{-8}$
CPA	$2,6 \times 10^5$	$5,2 \times 10^3$	3,2	$5,0 \times 10^7$	$2,0 \times 10^{-8}$
CEWC	$1,3 \times 10^6$	$4,9 \times 10^4$	5,3	$2,6 \times 10^9$	$3,9 \times 10^{-10}$
PHYTC57	$3,5 \times 10^5$	$2,2 \times 10^6$	13,9	$1,6 \times 10^{11}$	$6,3 \times 10^{-12}$

Estabilidad de la fitocistatina

Se estuvo interesado en explorar si PHYTC57 mostraba una mayor estabilidad en comparación con una fitocistatina parental bien caracterizada, OSA-IAD86. Las muestras de estas fitocistatinas se incubaron varias veces en un baño de agua hirviendo, se enfriaron y luego se analizaron para determinar la actividad inhibitoria residual en el ensayo de papaína (Figura 4A). La proteína de consenso PHYTC57 muestra una mayor termoestabilidad ($t_{1/2} = 17$ min) que OSA-IAD86 ($t_{1/2} = 6$ min), mientras que la actividad inhibitoria todavía se puede detectar a los 80 min con PHYTC57 en comparación con solo 58 min para OSA-IAD86.

También se probó la estabilidad de OSA-IAD86 y PHYTC57 usando un ensayo de digestibilidad de fluido gástrico simulado (SGF). Las proteínas se incubaron varias veces en SGF recién preparado antes de neutralizar la reacción. Luego, las muestras se analizaron de dos maneras, mediante ensayos enzimáticos para determinar la actividad inhibitoria residual (Figura 4B) y mediante SDS-PAGE (Figura 4C) para determinar si la proteína se digirió y se observó una mayor estabilidad de PHYTC57 en los estudios de digestión con PHYTC57 y OSA-IAD86 que muestran valores de $t_{1/2}$ de 260 segundos y 30 segundos respectivamente en ensayos enzimáticos. Estos datos de ensayo indican que aproximadamente el 99% de OSA-IAD86 se destruye en el fluido gástrico simulado entre 30 y 60 segundos después de la incubación. Para PHYTC57, este nivel de inactivación no ocurre hasta entre 2 y 5 minutos, lo que demuestra que PHYTC57 es más resistente a las condiciones de digestión. Los resultados de SDS-PAGE teñidos con Coomassie (Figura 4C), que identifican la proteína de longitud completa, respaldan estos datos con solo trazas de OSA-IAD86 presente a los 30 segundos, mientras que para PHYTC57 todavía hay algo de proteína a los 120 segundos. Para confirmar estos resultados para PHYTC57, se analizaron los productos de digestión de SGF mediante análisis de transferencia Western usando un anticuerpo anti-etiqueta6His. Como se muestra en la Figura 4D, esto revela que la mayoría de la proteína ha sido destruida por el tratamiento con SGF de 300 segundos, aunque quedan trazas de PHYTC57 intactas con un tratamiento de 300 e incluso 420 segundos. Si se usara PHYTC57 para la expresión de plantas transgénicas, el hecho de que la mayoría de la proteína PHYTC57 de longitud completa y toda la actividad inhibitoria se haya perdido después de una incubación de 10 minutos en SGF significaría que PHYTC57 debería destruirse fácilmente durante el proceso digestivo después de cualquier digestión involuntaria del huésped.

Discusión

El enfoque de consenso para el diseño de proteínas proporciona un método para generar una secuencia que no existe en la naturaleza. Dicha secuencia debería optimizar los parámetros de secuencia funcional conservados de una familia de proteínas homólogas. En particular, los residuos críticos que están involucrados en la estructura y el plegamiento de la familia probablemente estén altamente conservados (Lehmann y Wyss 2001; Main et al., 2003a; Main et al., 2003b; Main et al., 2005) (Forrer et al., 2004) (Steipe 2004). Dependiendo de los roles biológicos de los miembros individuales de la familia, los residuos funcionales pueden o no conservarse. En el caso de las fitocistatinas que muestran conservación funcional en forma de unión e inhibición de cisteína proteasa, existe un refuerzo de los motivos de secuencia altamente conservados que están involucrados en la interacción con el sitio activo de cisteína proteasa.

Por lo tanto, el enfoque de consenso debería dar lugar a una proteína optimizada en la que se mejoren la función biológica, la unión a la cisteína proteasa, así como la estabilidad general.

Las plantas contienen una clase de cistatinas que tienen características distintivas de las cistatinas y estefinas animales, y las plantas no contienen estefinas. En particular para las cistatinas y las estefinas animales, la secuencia Gly del extremo terminal N es importante para la inhibición de las cisteína proteasas, mientras que este no es el caso de las cistatinas vegetales [Abe, K. et al., (1988) J. Biol. Chem 263, 7655-7659]. Además, las cistatinas vegetales contienen una secuencia típica [LVI] [AGT] [RKE] [FY] [AS] [VI] X [EDQV] [HYFQ] N (SEQ ID NO 81) que no se encuentra en otras cistatinas o en estefinas.

El análisis de SPR de las cistatinas inmovilizadas con papaína reveló que PHYTC57 es más eficaz para formar y mantener un complejo de interacción proteína: proteína con papaína que las 3 fitocistatinas parentales probadas en el presente documento. En particular, la constante de velocidad de disociación se reduce, lo que indica que una vez unido, el complejo PHYTC57:papaína es más estable que los formados por las otras cistatinas probadas. Anteriormente se informó que las cistatinas derivadas de animales, que contienen enlaces disulfuro, son cistatinas más eficientes que las fitocistatinas caracterizadas, cuya eficacia varía de alrededor de 10 nM para OSA-I (Urwin et al., 1995) hasta pM para cistatinas de soja (Koiwa et al., 2001) contra la papaína. En este estudio se observó que PHYTC57 es más eficaz para unir papaína que la cistatina animal, cistatina de clara de huevo de gallina. Además, PHYTC57 muestra una estabilidad térmica mejorada, así como una mayor estabilidad en un ensayo de digestibilidad simulada.

El hecho de que PHYTC57 muestre propiedades mejoradas no significa que no pueda mejorarse más. El refuerzo de los motivos conservados que comprenden la región de unión puede representar la secuencia de unión biológicamente óptima, pero quizás no sea la más efectiva. Por ejemplo, Koiwa demostró que la alteración de la tercera región bucle en las cistatinas de soja mejoró la eficacia de las variantes resultantes (Koiwa et al., 2001). Ha habido un informe de que nuevas secuencias dentro de la región QVVAG confieren una actividad inhibidora mejorada (Melo et al., 2003). En términos de uso potencial en sistemas de plantas transgénicas, es poco probable que nuevas mejoras en la estabilidad sean beneficiosas debido a la necesidad de asegurar que los productos transgénicos no se acumulen en el medio ambiente.

Nos hemos enfocado en el presente documento en la eficacia de unión relativa de PHYTC57 en comparación con OSA- Δ D86. Esta variante de cistatina de arroz fue la primera fitocistatina mejorada generada y ha sido sometida a una amplia gama de estudios que conducen a la expresión de plantas transgénicas y a los ensayos de resistencia a los nematodos de las plantas (Urwin et al., 1995; McPherson et al., 1997; Urwin et al., 1997; Urwin et al., 1998; Urwin et al., 2000; Urwin et al., 2001; Urwin et al., 2003; Lilley et al., 2004). PHYTC57 muestra 34 diferencias de aminoácidos de OSA- Δ D86. Una comparación de las posiciones de estas diferencias de aminoácidos entre OSA- Δ D86 y PHYTC57 se muestra en la Figura 5 en la que las sustituciones están codificadas por colores y están representadas en el modelo estructural de OSA-I (código pdb 1EQK) (Nagata et al., 2000). Las diferencias se dispersan por todo el armazón estructural e incluyen tanto cambios conservadores como cambios no conservadores. Es interesante que 6 de los cambios involucren la introducción de aminoácidos cargados negativamente. Se podrían realizar estudios comparativos, bioquímicos, de análisis de secuencia y mutagénesis detallados de moléculas de cistatina parentales individuales con la proteína consenso para definir las sustituciones de "consenso" más críticas que conducen a propiedades mejoradas, mejorando así nuestra comprensión de las relaciones estructura-función de la proteína.

La estabilidad mejorada sorprendentemente significativa de PHYTC57 nos llevó a considerar el potencial de esta pequeña proteína de consenso como un nuevo armazón. Los armazones de proteínas para la selección de nuevas funciones de unión están demostrando ser útiles en una amplia gama de aplicaciones como reemplazos de anticuerpos (Skerra 2007) incluso en aplicaciones médicas (Wurch, Pierre et al., 2012). Se ha reportado un ejemplo del desarrollo de una cistatina como un armazón para la selección de aptámeros peptídicos basado en estefina humana A (Woodman, Yeh et al., 2005). Ha habido un informe reciente sobre el desarrollo de una proteína de consenso similar a Fn3 a partir de 15 secuencias de dominio similares a Fn3 de fibronectina o tenascina que se propone como un armazón potencial (Jacobs, Diem et al., 2012). El dominio 10 natural de Fn3 es un armazón bien estudiado desarrollado por Koide y colegas (Koide, Bailey et al., 1998; Karatan, Merguerian et al., 2004; Koide, Gilbreth et al., 2007). Debido a su termoestabilidad mejorada, su pequeño tamaño y la falta de cisteínas, se anticipó que la PHYTC57 de consenso demostrará ser útil como un armazón de proteínas de unión para desplegar bibliotecas de bucles de péptidos variables para el cribado contra un rango de moléculas objetivo para identificar nuevas proteínas de unión artificial.

PHYTC57, por lo tanto, ofrece beneficios potenciales para esquemas de defensa de plantas transgénicas como un inhibidor de cisteína proteasa mejorado dirigido a patógenos tales como nematodos de plantas, y para el desarrollo como una proteína de armazón para seleccionar nuevas funcionalidades de unión.

Creación de una biblioteca de proteínas de armazón, función de cribado y prueba de 'Adhironas'

Introducción

Los presentes inventores diseñaron una proteína de unión artificial (armazón) novedosa basada en la secuencia de consenso de 57 fitocistatinas derivadas de plantas descritas anteriormente (denominada PHYTC57 anteriormente, pero denominada a continuación como 'Adhirona'). Esta proteína artificial cumple con todos los requisitos (pequeña, monomérica, alta solubilidad y alta estabilidad y falta de enlaces disulfuro y sitios de glicosilación) para ser un buen armazón para la presentación de péptidos. Se eligieron las regiones VVAG y PWE del armazón Adhirona para la presentación de péptidos con nueve posiciones aleatorias en cada bucle.

Con base en los altos rendimientos de armazón de Adhirona expresados en *E. coli*, se planteó la hipótesis de que las alteraciones de proteínas dentro de los dos bucles tienen un efecto tolerable sobre el nivel de expresión de proteínas y la estabilidad del armazón. Por lo tanto, este armazón parece ser apto para su uso en la generación de bibliotecas combinatorias para el cribado con la tecnología de despliegue en fagos (Smith 1985). El éxito del sistema de despliegue en fagos depende de la calidad de la biblioteca de ADN inicial, que se deriva principalmente de su diversidad. Se puede lograr una diversidad de biblioteca mejorada mediante el uso de oligos sintetizados con trinucleótidos (trímeros) (Kayushin, Korosteleva et al., 1996) que proporcionan niveles teóricamente iguales de introducción de los diferentes aminoácidos, así como evitar codones de parada y cisteína (Virnekas, Ge et al., 1994; Krumpe, Schumacher et al., 2007). Además, las inserciones o eliminaciones de trímeros no conducirán a un cambio en la mutación del marco de lectura, produciendo proteínas potencialmente funcionales. Por lo tanto, se ha elegido una mezcla de trímeros que codifica los 19 aminoácidos naturales, excluyendo la cisteína, para los oligos aleatorizados del bucle.

Este trabajo demuestra que las Adhironas tienen un alto potencial para desempeñar un papel clave en la generación de reactivos de investigación, diagnósticos y terapéuticos (descubrimiento de fármacos).

El armazón de Adhirona muestra una estabilidad térmica notablemente alta (T_m aproximadamente 101 °C) por encima de lo informado para cualquier otra proteína de armazón no repetida, y puede expresarse a niveles altos en sistemas de expresión procariotas para producir proteína recombinante de forma reproducible. Se ha construido una biblioteca de despliegue en fagos basada en la inserción de secuencias de aminoácidos aleatorizadas para reemplazar los residuos en dos regiones bucle dentro de la Adhirona. La biblioteca tiene una complejidad de aproximadamente 3×10^{10} , con más del 86% de clones de longitud completa después de la producción del fago, lo que indica la muy alta calidad de la biblioteca. Como una demostración de la eficacia de la biblioteca, la proteína modificadora similar a la ubiquitina pequeña de levadura (SUMO) se cribó para identificar reactivos de proteína de unión artificial (Adhirona) capaces de unirse a esta proteína objetivo. Se identificaron más de 20 Adhironas individuales que se unen a SUMO de levadura (ySUMO) de acuerdo con lo evaluado por un ensayo inmunoabsorbente ligado a enzimas (ELISA) de fago. La secuenciación del ADN indicó que la mayoría muestra homología de secuencia parcial dentro de una de las dos regiones bucle con el motivo interactivo SUMO conocido (Val/Ile-X-Val/Ile-Val/Ile; en el que X es cualquier aminoácido). Se subclonaron cuatro regiones de codificación de Adhirona en el vector pET11 y la proteína recombinante se expresó, purificó y probó en análisis ELISA y Western blot. Los cuatro Adhironas tenían baja afinidad nanomolar por ySUMO y mostraron una alta especificidad por ySUMO con un bajo nivel de unión a la proteína SUMO1 humana. Además, se cribó la biblioteca de Adhirona contra otros objetivos, a saber, el factor de crecimiento de fibroblastos (FGF1), la molécula de adhesión de células endoteliales de plaquetas (PECAM-1), también conocida como grupo de diferenciación 31 (CD31), y un péptido de 10 aminoácidos con una cisteína en el extremo terminal N para la unión del tiol a la biotina (Cys-Thr-His-Asp-Leu-Tyr-Met-Ile-Met-Arg-Glu) y también se identificaron Adhironas contra estos objetivos de acuerdo con lo confirmado por ELISA de fagos. Por lo tanto, se ha desarrollado una proteína de armazón versátil, altamente estable y bien expresada, denominada Adhirona, que es capaz de mostrar bucles de péptidos aleatorios y se ha demostrado la capacidad para seleccionar reactivos de unión de alta afinidad y altamente específicos de una biblioteca de Adhirona contra un rango de objetivos para su uso en múltiples aplicaciones.

Materiales y métodos

Construcción de la biblioteca de Adhirona

Se identificó una secuencia de consenso derivada de la alineación de 57 secuencias de fitocistatina como se describió anteriormente y se sintetizó un gen modificado por codones diseñado para la expresión en *E. coli* (GenScript). La región de codificación del armazón de Adhirona y las regiones de codificación de la biblioteca de Adhirona se clonaron entre los sitios de restricción *NheI* y *NotI* para crear una región de codificación de fusión con la mitad 3' del gen III del bacteriófago M13 en un vector fagémido pBSTG1, un derivado de pDHisII que se deriva de pHEN1 (Hoogenboom, Griffiths et al., 1991) y que también contiene un péptido señal DsbA (pBSTG1-DsbA-Adhirona). La biblioteca se construyó mediante extensión de solapamiento de empalme (SOE) de dos productos de PCR (Horton, Cai et al., 1990) y todos los cebadores fueron sintetizados por Ella Biotech.

El primer producto de PCR se extendió desde la secuencia de codificación DsbA hasta el primer bucle insertado y fue generado por los cebadores:

Cebador directo 5' - TCTGGCGTTTTCTGCGTC - 3' (SEQ ID NO 21),

Cebador inverso 5' - CTGTTCTTTTCGCTTTAACAAC - 3' (SEQ ID NO 22).

El segundo producto de PCR introdujo dos regiones bucle de nueve aminoácidos en la proteína de armazón en el bucle 1 y el bucle 2 usando los siguientes cebadores. El sitio *Pst*I utilizado para la clonación está subrayado:

Bucle directo

5'GTTGTAAAGCGAAGAACAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACCATGTAC
CACTTGACCCTG – 3' (SEQ ID NO 23).

Bucle inverso

5'CTGCGGAAGTCTCTGCAGTTCTTTGAAGTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TTAAGCCAAACTTTTCGCTTCG - 3' (SEQ ID NO 24).

Las posiciones degeneradas (NNN) se introdujeron como trimeros que representan un codón único para cada uno de los 19 aminoácidos, excluyendo la cisteína y no hubo codones de terminación. Los cebadores también se diseñaron para introducir sitios de restricción *NheI* (directo) y *PstI* (inverso) para facilitar la clonación en el vector fagémido pBSTG1 que contiene un codón de parada ámbar en el marco para permitir la lectura traduccional para crear una proteína de fusión pIII truncada de Adhirona. La PCR se realizó usando polimerasa Phusion High Fidelity (NEB) a 98 °C durante 5 minutos seguido de 20 ciclos de 98 °C, 10 segundos; 56 °C, 15 segundos; 72 °C, 15 segundos seguido de 72 °C durante 5 minutos. Los productos de PCR se purificaron mediante extracción en gel (Qiagen) y se usaron para SOEing con 10 ciclos usando el protocolo anterior. El producto de PCR se digirió con *NheI* y *PstI* y se extrajo en gel y luego se clonó en el fagémido pBSTG1-Adhirona que también se había digerido con *NheI* y *PstI* para dejar la secuencia señal DsbA y la región de codificación del extremo terminal C de la Adhirona, para generar la biblioteca de Adhirona basada en ADN. Se usó la electroporación para introducir los productos de biblioteca ligados en células electrocompetentes *E. coli* ER2738 (Lucigen). En total, 20 mL de células ER2738 se electroporaron con 50 ng de ADN de biblioteca por 50 µL de células ER2738. Se permitió que las células se recuperaran durante 1 hora en medio 2TY y luego se cultivaron a 37 °C, 225 rpm a una OD₆₀₀ de 0,6 en 2 litros de medio 2TY. Se añadió 1 µL de fago auxiliar M13KO7 (NEB) (10¹⁴/mL) y se dejó infectar las células con agitación a 90 rpm durante 1 hora, y luego se permitió que el cultivo produjera partículas de fago durante la noche a 25 °C en presencia de kanamicina (50 µg/mL). Los fagos se precipitaron con polietilenglicol 8000 al 6% y NaCl 0,3 M, y se suspendieron en glicerol al 50% para almacenamiento. Se determinó que el tamaño de la biblioteca era aproximadamente 3 x 10¹⁰ con un fondo mínimo de vector solamente.

Preparación del objetivo y despliegue en fagos

Los siguientes protocolos se describen para SUMO de levadura pero se usó un protocolo idéntico para el cribado u otros objetivos. La proteína SUMO de levadura (ySUMO) se expresó en células BL21 (DE3) usando inducción con IPTG y se purificó por cromatografía de afinidad con resina Ni-NTA (Qiagen) de acuerdo con las instrucciones del fabricante. La pureza fue confirmada por SDS-PAGE. SUMO de levadura se biotiniló utilizando EZ-link NHS-SS-biotina (Pierce), de acuerdo con las instrucciones del fabricante. La biotinylación se confirmó usando estreptavidina conjugada con peroxidasa de rábano picante (HRP) para detectar la biotina en ySUMO absorbido en placas Immuno 96 Microwell^{MR} Nunc MaxiSorp^{MR} (Nunc). El biopaneó de la biblioteca de despliegue en fagos se realizó de la siguiente manera:

Se mezclaron 5 μ L de la biblioteca de fagémidos, que contenía 10^{12} partículas de fagémidos, con 95 μ L de solución salina tamponada con fosfato, Tween-20 al 0,1% (PBST) y se sometió a un paneo previo tres veces en pozos recubiertos con estreptavidina de alta capacidad de unión (Pierce) durante un total de 1 hora. Se agregaron 100 μ L de ySUMO biotinilada 100 nM al pozo de paneo durante 1 hora con agitación en una Heidolph VIBRAMAX 100 a una velocidad programada en 3 antes de agregar el fago previamente paneado durante 2,5 horas también en la plataforma vibratoria. Los pozos de paneo se lavaron 10 veces en 300 μ L de PBST usando un lavador de placas (Tecan Hydroflex), y se eluyeron con 100 μ L de glicina-HCl 50 mM (pH 2,2) durante 10 minutos, se neutralizaron con Tris-HCl 1 M (pH 9,1), y se eluyó adicionalmente con 100 μ L de trietilamina 100 mM durante 6 minutos y se neutralizó con 50 μ L de Tris-HCl 1 M (pH 7). Los fagos eluidos se incubaron con células ER2738 de crecimiento exponencial ($OD_{600} = 0,6$) durante 1 hora a 37 °C y 90 rpm. Las células se colocaron en placas de agar con caldo de Lisogenia suplementado con 100 μ g/mL de carbenicilina y se cultivaron a temperatura ambiente durante la noche. Al día siguiente, las colonias se rasparon en 5 mL de medio 2TY y se inocularon en 25 mL de medio 2TY suplementado con carbenicilina (100 μ g/mL) para alcanzar una OD_{600} de 0,2, se incubaron a 37 °C, 225 rpm durante 1 hora y se infectaron con aproximadamente 1×10^9 fagos auxiliares M13K07. Después de 1 hora de incubación a 90 rpm, se añadió kanamicina a razón de 25 μ g/mL, las células se incubaron durante la noche a 25 °C a 170 rpm, y el fago se precipitó con polietilenglicol 8000 al 6%, NaCl 0,3 M y se resuspendió en 1 mL de Tris 10 mM, pH 8,0, EDTA 1 mM (tampón de TE). Se usaron 2 μ L de esta suspensión de fagos para la segunda ronda de selección. Esta vez, el despliegue en fagos se realizó utilizando perlas magnéticas de estreptavidina (Invitrogen). Los fagos se panearon previamente con 10 μ L de perlas lavadas durante 1 hora en un rotor de velocidad fija Stuart SB2 (20 rpm), y se marcaron 10 μ L de perlas con 100 μ L de ySUMO biotinilada 100 nM durante cuatro horas en el mismo rotor. Las perlas marcadas con SUMO de levadura se lavaron tres veces en PBST antes de añadir el fago previamente paneado durante 1 hora. Las perlas se lavaron 5 veces en PBST usando un imán para separar las perlas de la solución después de cada lavado.

luego se eluyeron y se amplificaron como anteriormente. El paneo final se realizó usando placas de neutravidina de alta capacidad de unión (Pierce), como se describió previamente para la primera ronda de paneo, pero esta vez el fago se eluyó usando 100 µL de ditioneitol (DTT) 100 mM en una plataforma vibratoria durante 20 minutos antes de la infección de las células ER2738. Los fagos se recuperaron de los pozos que contenían proteína objetivo y los pozos de control para determinar el nivel de amplificación en los pozos objetivo.

ELISA de fagos

Las colonias individuales de ER2738 del paneo final se recogieron y se cultivaron en 100 µL de 2TY con 100 µg/mL de carbenicilina en una placa de 96 pozos profundos a 37 °C (900 rpm) durante 6 h. Se añadieron 25 µL del cultivo a 200 µL de 2TY que contenía carbenicilina y se cultivaron a 37 °C (900 rpm) durante 1 hora. Se añadieron fagos auxiliares M13K07 (10 µL de 10¹¹/mL), seguido de kanamicina a razón de 25 µg/mL y las bacterias se cultivaron durante la noche a 25 °C (450 rpm). Las placas recubiertas con estreptavidina (Pierce) se bloquearon con 2x de tampón de bloqueo de caseína (Sigma) durante la noche a 37 °C. Al día siguiente, las placas se marcaron con 0,4 nM de SUMO de levadura biotinilada durante 1 hora, las bacterias se recogieron por centrifugación a 3000 rpm durante 5 minutos y se añadieron 45 µL de medio de crecimiento que contenía el fago a los pozos que contenían SUMO de levadura biotinilada o un pozo que contiene el enlazador biotinilado y se incubó durante 1 hora. Los pozos se lavaron usando un lavador de placas Tecan Hydroflex 3 veces en 300 µL de PBST, y se añadió una dilución 1:1000 de anticuerpo anti-fago conjugado con HRP (Seramun) en 100 µL de PBST durante 1 hora. Los pozos se lavaron 10 veces en 300 µL de PBST y la unión se visualizó con 100 µL de sustrato líquido 3,3',5,5'-tetrametilbencidina (TMB) (Seramun) y se midió a 560 nm.

Producción de proteína Adhirona

Las secuencias codificantes de ADN de Adhironas que se unían a SUMO de levadura se amplificaron por PCR, el producto se digirió por restricción con *NheI* y *PstI* y se clonó en pET11a que contenía el armazón de Adhirona y se digirió con los mismos sitios de restricción. Las colonias se recogieron y se cultivaron durante la noche en 5 mL de medio LB a 37 °C, 225 rpm y el ADN plasmídico se purificó como minipreps (Qiagen) y se secuenciaron para confirmar la presencia del inserto correcto. Los plásmidos se transformaron en células BL21 (DE3) mediante choque térmico y las colonias se cultivaron durante la noche a 37 °C. Al día siguiente, el cultivo se añadió a 400 mL de medio LB, se dejó crecer hasta una OD₆₀₀ de 0,6 a 250 rpm a 37 °C y se añadió isopropil β-D-1-tiogalactopiranosido (IPTG) hasta una concentración final de 1 mM. Las células se cultivaron durante 6 horas más, se recogieron por centrifugación a 3000 g y se resuspendieron en 25 mL de 1x Bugbuster (Novagen). Se añadió Benzonasa de acuerdo con las instrucciones del fabricante y la suspensión se mezcló a temperatura ambiente durante 20 minutos, se calentó a 50 °C durante 20 minutos y se centrifugó durante 20 minutos a 9400 xg. El sobrenadante aclarado se mezcló con resina Ni-NTA 500 µL de suspensión durante 1 hora, se lavó 3 veces en 30 mL de tampón de lavado (PBS 50 mM, NaCl 500 mM, imidazol 20 mM, pH 7,4) y se eluyó en 1 mL de tampón de elución (PBS 50 mM, NaCl 500 mM, imidazol 300 mM, pH 7,4). Se biotinilaron 100 µg de las Adhironas de unión a SUMO (Ad-ySUMO) usando NHS SS-biotina (Pierce) de acuerdo con las instrucciones del fabricante para uso en ELISA y transferencia Western.

Análisis ELISA

Se absorbieron 5 ng (a menos que se indique lo contrario) de proteína objetivo en PBS sobre pozos de una placa Immuno 96 Microwell^{MR} Nunc MaxiSorp^{MR} durante la noche a 4 °C. Al día siguiente, se añadieron 200 µL de tampón de bloqueo 3x a los pozos y se incubaron a 37 °C durante 4 horas sin agitación. Los Adhironas de unión a SUMO de levadura biotinilados a razón de 100 µg/mL se diluyeron 1:1000 en PBST que contenía tampón de bloqueo 2x y se incubaron alícuotas de 50 µL en pozos objetivo durante 1 hora con agitación. Los pozos se lavaron 3 veces en 300 µL de PBST, y se añadió estreptavidina conjugada con peroxidasa de rábano picante (HRP) (Invitrogen) diluida 1:1000 en 50 µL de PBST a los pozos durante 1 hora. Los pozos se lavaron 6x en 300 µL de PBST y se visualizó la unión con 50 µL de sustrato líquido TMB y se midió la absorbancia a 560 nm.

Análisis de Transferencia Western

La proteína objetivo o la proteína objetivo mezclada con lisado de células HEK293 (20 µg) se mezcló con tampón de carga (Laemmli, Tris-HCl 60 mM pH 6,8, SDS al 2%, glicerol al 10%, β-mercaptoetanol al 5%, azul de bromofenol al 0,01%), se calentó durante 3 minutos, se centrifugó durante 1 minuto a 15.000 xg y luego se resolvió en un gel de SDS-poliacrilamida al 15%. Las proteínas se transfirieron a membranas de PVDF durante 45 minutos a 4 vatios (Amersham Biosciences) y se incubaron durante 1 hora en tampón de bloqueo (BSA al 5% en PBS Tween al 0,1%) seguido de incubación durante 1 hora con Ad-ySUMO (100 µg/mL diluidos 1:1000 con PBST). Los Ad-ySUMO unidos se detectaron usando HRP conjugada con estreptavidina y quimioluminiscencia (kit ECL Plus, Amersham).

Medida de la afinidad de interacción proteína-proteína

Los biosensores de estreptavidina de inmersión y lectura BLitz^{MR} (ForteBio) se usaron para estimar la afinidad de unión de los aglutinantes Ad-ySUMO biotinilados, de acuerdo con las instrucciones del fabricante. En resumen, se utilizaron al menos 4 lecturas a diferentes concentraciones de ySUMO (0,25 mM - 1 mM) para medir la afinidad de

cada Ad-ySUMO. Se utilizó un ajuste global para calcular la afinidad de cada Ad-ySUMO. Estas lecturas fueron comparables a las medidas de afinidad realizadas utilizando un instrumento de resonancia de plasmón superficial Biacore.

5 Resultados

Diseño de la Adhirona y despliegue en fagos

El gen de Adhirona fue diseñado originalmente para crear un inhibidor de proteasa más potente, sin embargo, debido a su potencial como proteína de armazón para presentar regiones peptídicas restringidas para el reconocimiento molecular (Figura 6A) se decidió investigar su uso como armazón de este tipo. La secuencia del gen fue de codón optimizado para mejorar la expresión en un sistema de expresión de *E. coli* (Figura 6B). Se introdujeron sitios de restricción para facilitar la clonación del gen en el vector fagémido pBSTG1 para permitir la lectura traduccional en el marco a través de un codón de parada ámbar para permitir que se produzca una proteína de fusión Adhirona-pIII troncada cuando se expresa en células no supresoras tales como células ER2738 pero para permitir la producción de la Adhirona solo en células supresoras tales como JM83. La proteína de fusión Adhirona-pIII se expresó a partir del vector fagémido pBSTG1, mientras que los otros componentes para permitir la replicación y el empaquetamiento del ADN fagémido en partículas de fago M13 se introdujeron usando un fago auxiliar M13KO7. La expresión de la proteína de fusión Adhirona-pIII se confirmó mediante análisis de transferencia Western usando un anticuerpo anti-pIII. La estabilidad térmica del armazón de Adhirona se probó mediante calorimetría diferencial de barrido, que mostró una temperatura de fusión de 101 °C (Figura 7A). La integridad estructural de la secuencia de consenso se examinó utilizando dicroísmo circular, que demostró una alta proporción de láminas beta con respecto a hélices alfa y enrollamiento aleatorio (Figura 7B).

Luego se comparó la estabilidad térmica mediante calorimetría diferencial de barrido del armazón de Adhirona (Figura 8A) con una pequeña proteína representativa soluble bien caracterizada, lisozima (Figura 8B) que muestra que la lisozima es significativamente menos estable (T_m aproximadamente 65 °C) que la Adhirona. Luego se probó una Adhirona seleccionado para unirse a un anticuerpo myc, la adición de los bucles en el armazón reduce la T_m a 85 °C, pero esto todavía representa una temperatura de fusión más alta que la mayoría de las proteínas de armazón. Esta proteína Adhirona puede experimentar ciclos repetidos de desnaturalización y renaturalización como se muestra por la serie de barridos (Figura 8C).

Diseño de la biblioteca

La introducción de secuencias codificantes de péptidos adecuadas para el reconocimiento molecular fue guiada por las posiciones de bucle predichas dentro de la estructura de la Adhirona (Figura 6). El bucle1 se posicionó entre la primera y la segunda cadenas beta y el bucle2 se posicionó entre la tercera y la cuarta cadenas beta. Se introdujeron secuencias que comprenden nueve aminoácidos aleatorios (excluyendo la cisteína) en ambas posiciones del bucle, reemplazando cuatro y tres aminoácidos en el bucle1 y el bucle2, respectivamente. Para determinar si la extensión de estas regiones bucle altera la estructura de la Adhirona, se aislaron, expresaron y examinaron tres Adhironas individuales con inserciones de bucle mediante dicroísmo circular (Figura 7B). Los tres clones mantuvieron una alta proporción de estructura beta, con un clon que muestra un aumento en el contenido de la estructura beta que probablemente indica la extensión de las cadenas beta en las nuevas regiones bucle. Esto demuestra que la inserción del bucle no afecta la estructura del armazón. Se generó una biblioteca de despliegue en fagos de complejidad de aproximadamente 3×10^{10} . Para verificar la composición de aminoácidos, se aislaron 96 clones de células ER2738 infectadas con la biblioteca. Se examinó la secuencia de clones de fagos para determinar cualquier sesgo en la composición de aminoácidos u otras consecuencias indeseables introducidas durante la producción de fagos (Figura 7C). No se observó sesgo en la distribución de aminoácidos. El 86,5% de los clones eran variantes de longitud completa, mientras que el 3,1% de los clones eran el armazón de Adhirona sin insertos, y el 10,4% de los clones mostraban cambios de marco y, por lo tanto, probablemente no tenían valor en la biblioteca. Esta proporción muy alta de secuencias de codificación de longitud completa a nivel del genoma del fago demuestra la alta calidad de la biblioteca generada.

Cribado de la biblioteca

El cribado de la biblioteca se realizó inicialmente usando SUMO de levadura como objetivo. La SUMO de levadura se biotiniló para permitir la inmovilización de la proteína a través de proteínas de unión a avidina y para asegurar que el objetivo no se adsorbiera directamente sobre superficies de plástico o partículas que a veces pueden conducir a la desnaturalización de la proteína objetivo. Esto asegura que la proteína objetivo mantenga su estructura tridimensional permitiendo la selección de proteínas de unión que reconocen epítopos lineales o conformacionales. Se observó una amplificación de más de 1000 veces en la recuperación de colonias en comparación con las muestras de control mediante el paneo de la tercera ronda. Se aislaron veinticuatro clones y se confirmó su capacidad para unirse al objetivo SUMO mediante ELISA de fagos (Figura 9). Todos los clones probados mostraron una fuerte unión a SUMO de levadura con poca o ninguna unión a los pozos de control que demuestran la especificidad de las Adhironas. Se secuenciaron los clones, que identificaron 22 Adhironas distintos y las secuencias de las regiones bucle en estos clones se muestran en la Tabla 4.

Tabla 4. Muestra las dos secuencias de bucle para los 24 aglutinantes Ad-ySUMO identificados a partir del cribado

Ad-ySUMO	Bucle1 (SEQ ID No)	Bucle2 (SEQ ID No)
1	WDLTGNVDT (25)	WDDWGERFW(49)
2	IDLNSFAS(26)	DINQYWHSM(50)
3	INLMMVSPM(27)	GIQQNPSHA(51)
4	IDLTHSLNY(28)	GLTNEIQKM(52)
5	IDLTHSLNY(29)	GLTNEIQKM(53)
6	IDLTEWQDR(30)	PEPIHSHHS(54)
7	WVDMDYWR(31)	MDEIWAIEYA(55)
8	IDLTEWQDR(32)	EPGIPIVH(56)
9	IDLTDVWID(33)	GLMTQTNSM(57)
10	IIHENDAD(34)	GIMDGLNKY(58)
11	WILNNTQFI(35)	VLEGPDRTWT(59)
12	WYERSENWD(36)	RDYGFTLVP(60)
13	WDLTPINI(37)	YEDYQTPMY(61)
14	WFDDEYDWI(38)	DYAATDLYW(62)
15	IDLTPHDS(39)	YEEDEYWRM(63)
16	IDLTPHDS(40)	PIDSNFTGT(64)
17	WYLLDVMDD(41)	HDRRYKQAE(65)
18	WIDRGQYWD(42)	IHNQYTIMD(66)
19	WSEADNDWH(43)	LDLETWQHF(67)
20	IDLTPHDS(44)	PLWQYDAQY(68)
21	IDLTPHDS(45)	PSHHNYQTM(69)
22	IDLTPHDS(46)	PIDSNFTGT(70)
23	IDLTPHDS(47)	PHDELNWNM(71)
24	WEDFQTHWE(48)	DVGQLLSGI(72)

Los clones 4 y 5 son idénticos y los clones 16 y 22 también son idénticos. Curiosamente, los clones 15 y 23, así como 21 y 22 contienen la misma secuencia de aminoácidos en el bucle1 pero diferentes secuencias en el bucle2. Esta variación de secuencia es compatible con la naturaleza compleja de la biblioteca. El análisis de las secuencias identificó una secuencia común de IDLT en las posiciones 1 a 4 del bucle1 en 12 de los clones, lo que indica que este puede ser un motivo importante en la unión a al menos un epítipo en la ySUMO. También se produce una posición P o G en la posición 1 del segundo bucle en 9 clones distintos y se produce una P o G en una posición dentro de los residuos 2 a 5 en otros 6 clones, lo que puede indicar que alguna característica estructural puede ser importante en la unión. Curiosamente, el motivo IDLT es similar al sitio de unión de SUMO1 humana de la ligasa MEF2 E3 PIASx (VDV IDLT - SEQ ID NO 73) (Song, Durrin et al., 2004; Song, Zhang et al., 2005). Se seleccionaron cuatro clones para caracterización adicional; los clones 15 y 22, ya que esta secuencia del bucle1 se produjo más de una vez, el clon 20, ya que también contenía el motivo IDLT y el clon 10, ya que contenía un motivo distinto en el bucle1.

Caracterización de las proteínas Adhirona-ySUMO (Ad-YSUMO)

Debido a la alta estabilidad térmica del armazón de Adhirona (Figura 7A) se predijo que para ayudar a la purificación debería ser posible desnaturalizar por calor y precipitar la mayoría de las proteínas de E. coli sin afectar la integridad de las Adhironas expresados. Para probar esto, calentaron los lisados durante 20 minutos a 50 °C, 60 °C, 70 °C, 80 °C, 90 °C y 100 °C y se centrifugaron para sedimentar la proteína desnaturalizada y se analizaron los sobrenadantes por SDS-PAGE (Figura 10A). El calentamiento del lisado disminuyó drásticamente la cantidad de proteína bacteriana en el sobrenadante, pero no redujo significativamente los niveles de Adhirona. Una temperatura de 50 °C fue adecuada para precipitar la mayoría de las proteínas bacterianas, por lo que se adoptó en futuros estudios. La Figura 9B demuestra que los Ad-ySUMO purificados muestran una alta pureza utilizando un método de purificación de afinidad a metal por lotes y que en algunas muestras tal como el clon 10, la mayoría de la proteína no se aisló, potencialmente debido a la cantidad limitante de la resina utilizada durante esta purificación. El nivel estimado de proteína expresada fue de aproximadamente 100 mg/L. Las afinidades de las Adhironas se estimaron utilizando interferometría BioLayer con biosensores de inmersión y lectura BLITZ^{MR} (ForteBio). Las afinidades fueron 11,5, 2,4, 14,2 y 9,0 nM para Ad-ySUMO 10, 15, 20 y 22, respectivamente. Estos valores están en línea con las afinidades que normalmente se observan para anticuerpos buenos.

[0188] Para evaluar más a fondo el uso de las Adhironas como reactivos de investigación, las Ad-ySUMO se biotinilaron y se usaron en ELISA (Figura 10C) y análisis de transferencia Western (Figura 10D). Las Ad-ySUMO se unen a SUMO de levadura pero no a SUMO1 humana (datos no mostrados para las transferencias Western de SUMO1) (n = 3). Para determinar la especificidad de los reactivos, se mezcló SUMO de levadura con lisados de células

HEK293. Curiosamente, Ad-ySUMO10 y 15 muestran unión específica a SUMO de levadura sin unión a otras proteínas, pero Ad-ySUMO20 y 22 se unen a muchas proteínas en los lisados por transferencia Western (n = 3).

Otros ejemplos de cribados

Para evaluar más a fondo la capacidad de la biblioteca de Adhirona para identificar reactivos específicos capaces de unirse a un rango de objetivos, se cribó contra un factor de crecimiento (FGF1), un receptor (CD31) y una secuencia peptídica. Todos los cribados se realizaron en tres rondas de paneo. El ELISA de fagos se usó para examinar la capacidad de las Adhironas para unirse al objetivo correspondiente (Figura 11). Curiosamente, la mayoría de los clones probados para FGF1 y CD31 mostraron unión específica, mientras que solo tres clones del cribado de péptidos mostraron unión específica. Las rondas de paneo adicionales contra el péptido aumentaron la proporción de aciertos con respecto al fondo, de manera que el 80% de los clones recogidos mostraron unión al objetivo. Este resultado no es inesperado debido al pequeño tamaño y la probabilidad limitada de presentación apropiada del epítipo del péptido en comparación con los sitios de epítopos más grandes y, por lo tanto, potencialmente múltiples de las proteínas.

Para confirmar que las Adhironas expresados se unen a sus objetivos, se ha utilizado BLitz^{MR} para analizar tres Adhironas recombinantes distintos tanto para CD31 como para el péptido objetivo. Los Adhironas se expresaron y purificaron como proteínas solubles. Los valores de K_D para las Adhironas CD31 variaron de $8,5 \times 10^{-8}$ a $6,8 \times 10^{-9}$ M, mientras que los del péptido oscilaron entre $3,3 \times 10^{-8}$ y $3,5 \times 10^{-8}$.

Además, como se muestra en la Figura 25, se han identificado Adhironas que se unen a una molécula orgánica, posaconazol.

Estructura cristalina de la Adhirona

La Figura 17 muestra la estructura cristalina de una Adhirona complejo con un dominio del receptor de FcgRIIIa. Esto revela la estructura tridimensional de la Adhirona que muestra los elementos estructurales clave de 4 cadenas beta y una hélice alfa. La estructura es inesperada en términos de una naturaleza más compacta y una estructura aparentemente menos retorcida de lo que se ve para las estructuras de rayos X de otras cistatinas, incluida, por ejemplo, la estefina A. Es interesante que la estructura beta se extienda en las regiones bucle en diversos grados.

La importancia de mostrar dos bucles aleatorios para la selección de moléculas de unión para al menos algunos objetivos se destaca en esta estructura por la interacción íntima de ambos bucles con la proteína receptora. Estos bucles corresponden al BUCLE1 y BUCLE2 descritos con más detalle a continuación.

Discusión

Se ha desarrollado una nueva proteína de armazón basada en un diseño de consenso de proteínas de cistatina vegetal, denominada Adhirona, y que muestra una estabilidad térmica extremadamente alta con una T_m de 101 °C. Este armazón se ha utilizado para producir bibliotecas mediante la introducción de dos regiones variables de 9 aminoácidos. Estas secuencias variables fueron codificadas por oligonucleótidos en los que las posiciones variables son un subconjunto de trímeros que comprenden un codón único para cada aminoácido con la excepción de la cisteína. Esto dio como resultado una distribución muy similar de cada aminoácido dentro de la biblioteca. La biblioteca fue de muy alta calidad con un 86,5% de clones que representaban clones variantes de longitud completa.

La biblioteca se configuró en un formato de despliegue en fagos filamentosos como una fusión truncada de gp3 y se ha cribado contra diversas proteínas objetivo. El análisis de Adhironas identificados por cribado contra SUMO de levadura reveló una serie de proteínas con secuencias distintas en sus regiones variables. En algunos casos, existen similitudes que implican unirse al mismo sitio en la SUMO, mientras que otros clones no muestran la conservación de la secuencia. Todos los clones se unen a ySUMO y no a un rango de proteínas de control, lo que indica especificidad. También se han identificado Adhironas que se unen específicamente a un factor de crecimiento, FGF1 humano, un dominio de proteína del receptor humano de CD31, un péptido y un compuesto orgánico. La capacidad para seleccionar ligantes contra compuestos orgánicos, así como una amplia gama de proteínas, es un hallazgo importante, ya que la mayoría de los armazones tienen características estructurales que favorecen clases particulares de moléculas objetivo.

Los Adhironas se pueden purificar convenientemente mediante la inclusión de una etapa de temperatura a 50 °C que desnaturaliza muchas de las proteínas huésped de *E. coli* endógenas, mejorando así la eficiencia de la purificación por afinidad de la Adhirona. La estructura de rayos X de un complejo de una Adhirona que se une al receptor FcgRIII humano proporciona información útil no solo sobre el complejo sino también sobre la proteína Adhirona y revela una estructura beta más compacta y en gran parte compatible con los datos de CD. Es probable que la naturaleza compacta del armazón, que es más pronunciada que la observada en otras estructuras de estefinas o cistatinas, contribuya a su alta estabilidad térmica. La interfaz de interacción revelada por la estructura de rayos X indica que ambas regiones bucle están involucradas en la interacción con el dominio del receptor.

La demostración de bibliotecas exitosas y de alta calidad basadas en este armazón altamente estable, pequeño y fácilmente purificado, junto con nuestra estrategia robusta y efectiva para cribado contra una gama de proteínas, hace que nuestra biblioteca de Adhirona sea un recurso valioso y novedoso para el desarrollo de reactivos para una amplia gama de aplicaciones científicas, médicas y comerciales.

Variantes de Adhironas

Se han producido variantes del armazón de Adhirona. La secuencia que se muestra en la SEQ ID NO 1 y la Figura 6 es para la versión más corta examinada de Adhirona, y comprende 81 residuos de longitud antes de la adición de secuencias funcionales adicionales tales como un enlazador y etiqueta His u otras secuencias. Sin embargo, también se han producido dos proteínas de armazón más largas (SEQ ID NO 2 y 3), y cada una puede ser preferible en ciertas circunstancias. Es factible que sea posible una mayor eliminación del armazón, pero se cree que la SEQ ID NO 1, o sus variantes, están cerca de la longitud mínima óptima sin que la estabilidad de la proteína del armazón se vea comprometida indebidamente.

La longitud completa de 'Adhirona 92' (92 aa) tiene la siguiente secuencia (que consiste en la secuencia de armazón SEQ ID NO 3 más un enlazador y una etiqueta His, que están subrayados):

ATGVRVPGN ENSLEIEELA RFAVDEHNKK ENALLEFVRV VKAKEQVVAG
TMYLTLEAK DGGKKKLYEA KVVWKPWENF KELQEFKPVG DA AAAHHHHHH (SEQ ID
NO 74)

El 'Adhirona 84' corto (84 aa) tiene la siguiente secuencia (que consiste en la secuencia de armazón SEQ ID NO 2 más un enlazador y una etiqueta His, que están subrayados):

GNENSLEIEE LARFAVDEHN KKENALLEFV RVVKAKEQVV AGTMYLTLE
AKDGGKKKLY EAKVVWKPWE NFKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 75)

El 'Adhirona 81' más corto (81 aa), que se muestra en la Figura 6, tiene la siguiente secuencia (que consiste en la secuencia de armazón SEQ ID NO 1 más un enlazador y una etiqueta His, que están subrayados):

NSLEIEELAR FAVDEHNKKE NALLEFVRVV KAKEQVVAGT MYLTLEAKD
GGKKKLYEAK VVWKPWENFK ELQEFKPVGD A AAAHHHHHH (SEQ ID NO 76)

La secuencia subrayada comprende un enlazador con 3 Ala adicional y una etiqueta de detección/purificación de 6 His. Esta etiqueta no es parte del armazón mismo, pero es una adición útil a la proteína por razones obvias.

Ejemplos específicos de Adhironas para bibliotecas

Los ejemplos de secuencias de Adhirona que son útiles para preparar bibliotecas de proteínas de armazón, es decir, bibliotecas en las que se han insertado una variedad de péptidos en el armazón, son las siguientes:

Adhirona 92 (la secuencia que se muestra incluye, un Met adicional enlazador y etiqueta)

MATGVRVPGNEN**NSLEIEELARFAVDEHNKK**ENALLEFVRVV**KAKEQVVAGTMYLTLEAK**
DGGKKKLYEAKVVWKP**PWENF**KELQEFKPVGDA AAAHHHHHH (SEQ ID NO 77)

Se pueden hacer/tener una o más de las siguientes modificaciones:

- Se ha agregado un residuo de metionina adicional (en negrita) en el extremo terminal N para facilitar la traducción.
- Una porción del extremo terminal N se encuentra en los residuos de aminoácidos 2-4 (en negrita y cursiva), y convenientemente estos 3 aminoácidos se reemplazan por un inserto desde típicamente 3 hasta aproximadamente 20 aminoácidos.
- El BUCLE1 está ubicado en los residuos de aminoácidos 47-50 (numerados para excluir met del extremo terminal N) (en negrita y cursiva), y convenientemente estos 4 aminoácidos pueden ser reemplazados por un inserto típicamente desde 4 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
- El BUCLE2 se encuentra en los residuos de aminoácidos 76-78 (numerados para excluir met del extremo terminal N) (en negrita y cursiva), y convenientemente estos 3 aminoácidos pueden reemplazarse por un inserto de un inserto de típicamente 3 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
- Están presentes un enlazador del extremo terminal C y una etiqueta His. La longitud y la composición del enlazador pueden variar, y la etiqueta podría, por supuesto, adaptarse a cualquier sistema de purificación adecuado.

Adhirona 84 (la secuencia que se muestra incluye: un Met adicional, enlazador y etiqueta)

MGNENSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ**VVAG**TMYYLTLEAKDGGGKKKLY
EAKVWVK**PWEN**FKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 77)

5 Se pueden hacer/tener una o más de las siguientes modificaciones:

- Se ha agregado un residuo de metionina adicional (en negrita) en el extremo terminal N para facilitar la traducción.
 - Se puede añadir una secuencia del péptido del extremo terminal N al extremo terminal N de la Adhirona (es decir entre el residuo de metionina y la primera glicina como se muestra más arriba), y esta adición puede ser típicamente 3 hasta aproximadamente 20 aminoácidos.
 - 10 - El BUCLE1 está ubicado en los residuos de aminoácidos 39-42 (numerados para excluir met del extremo terminal N) (mostrado en negrita y cursiva), y convenientemente estos 4 aminoácidos pueden ser reemplazados por un inserto típicamente de 4 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
 - 15 - El BUCLE2 se encuentra en los residuos de aminoácidos 68-70 (numerados para excluir met del extremo terminal N) (mostrado en negrita y cursiva), y convenientemente estos 3 aminoácidos pueden reemplazarse por un inserto típicamente de 3 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
 - 20 - Están presentes un enlazador del extremo terminal C y una etiqueta His. La longitud y la composición del enlazador pueden variar, y la etiqueta podría, por supuesto, adaptarse a cualquier sistema de purificación adecuado.
- Adhirona 81 (excluye Met)

MNSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ**VVAG**TMYYLTLEAKDGGGKKKLYEAK
VWVK**PWEN**FKELQEFKPVGDA AAAHHHHHH (SEQ ID NO 78)

25 Se pueden hacer/tener una o más de las siguientes modificaciones:

- Se ha agregado un residuo de metionina adicional (en negrita) en el extremo terminal N para facilitar la traducción.
- Se puede añadir un bucle del extremo terminal N al extremo terminal N de la Adhirona y esta adición puede ser típicamente de 3 hasta aproximadamente 20 aminoácidos.
- 30 - El BUCLE1 está ubicado en los residuos de aminoácidos 36-39 (numerados para excluir met del extremo terminal N) (mostrado en negrita y cursiva), y convenientemente estos 4 aminoácidos pueden ser reemplazados por un inserto típicamente de 4 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
- 35 - El BUCLE2 se encuentra en los residuos de aminoácidos 65-67 (numerados para excluir met del extremo terminal N) (mostrado en negrita y cursiva), y convenientemente estos 3 aminoácidos pueden reemplazarse por un inserto típicamente de 3 hasta aproximadamente 20 aminoácidos. Se prefiere una longitud de bucle de 5 a 13 aminoácidos, y se cree que una longitud de bucle de 9 aminoácidos es óptima.
- Están presentes un enlazador del extremo terminal C y una etiqueta His. La longitud y la composición del enlazador pueden variar, y la etiqueta podría, por supuesto, adaptarse a cualquier sistema de purificación adecuado.

Por lo tanto, tomando al Adhirona 92 como ejemplo, una proteína de armazón particularmente adecuada para usar en un sistema de despliegue puede tomar la forma:

BUCLE1 DEL PÉPTIDO DEL EXTREMO TERMINAL N
45 MXXXXXXXXVRAVPGNENSLEIEELARFAVDEHNKKENALLEFV RVVVKAKEQXXXXXXXXXXTM

Enlazador y etiqueta del BUCLE2 YYLTLEAKDGGGKKKLYEAKVWVKXXXXXXXXXXNFKELQEFKPVGDA
AAAHHHHHH en el que X es cualquier aminoácido. (SEQ ID NO 79)

50 En general, se cree que el BUCLE1 y el BUCLE2 son de importancia primordial en la unión a objetivos, como lo soportan las estructuras cristalinas (Figura 17). En ciertas realizaciones de la invención, solo uno del BUCLE1 y el BUCLE2 puede reemplazarse con una secuencia peptídica, pero en general se prefiere que ambos se reemplacen. Aunque no se considera que el EXTREMO TERMINAL N sea tan importante como los BUCLES 1 y 2, en muchas circunstancias la inserción de un péptido adecuado en el EXTREMO TERMINAL N puede dar como resultado una

55 afinidad y especificidad de unión mejoradas.

Las Figuras 20 a 24 muestran alineamientos de secuencia de las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen a LOX1, HGH, SUMO de levadura, PBP2 y un péptido respectivamente.

60 Además, las Figuras 27 y 28 muestran las regiones BUCLE1 y BUCLE2 de varios Adhironas que se unen a Grb2 y STAT3.

Cabe señalar que los péptidos pueden insertarse opcionalmente en las regiones bucle sin eliminar los aminoácidos existentes, pero esto es típicamente menos preferido.

Ejemplos adicionales de las amplias utilidades de las Adhironas

Inmunofluorescencia:

Ejemplo: Reactivos para detectar proteínas virales

A pesar de los repetidos intentos durante muchos años, no se ha demostrado que sea posible generar un anticuerpo que identifique la proteína E5 del virus del papiloma humano. (Medición cuantitativa de los niveles de la oncoproteína E5 del virus del papiloma humano tipo 16 en líneas celulares epiteliales por espectrometría de masas. Sahab et al., J Virol. Septiembre de 2012; 86 (17): 9465-73; Wetherill, LF, Ross, R y Macdonald, A (2012). HPV E5: An enigmatic oncoprotein. Small DNA Tumour Viruses. Ed. K. Gaston. Caister Academic Press. Páginas 5-70)

La utilidad del sistema de Adhirona se demuestra mediante el ejemplo de que las Adhironas se produjeron contra la proteína viral E5 de HPV16 usando como objetivo un péptido con identidad en una región de la proteína E5. La Adhirona E5 se biotiniló y usó en inmunofluorescencia para detectar la proteína E5 en células sobreexpresadas. La Adhirona no reacciona de forma cruzada con otros serotipos de HPV. Además, las Adhironas E5 se han conjugado con puntos cuánticos y se han utilizado para detectar la proteína E5 en muestras humanas. La conjugación con puntos cuánticos aumentó la sensibilidad de los reactivos. Véase la Figura 12. La GFP 5 de HPV16 (objetivo) y la GFP 5 de HPV16 sin epítipo para Adhirona (control) se expresaron en células de mamífero. Adhirona E5 se conjugó con puntos cuánticos y se usó para detectar la proteína E5 en las células de mamíferos. Las células se tiñeron con DAPI (tinción de ADN), GFP y E5 (con los puntos cuánticos de Adhirona). La Adhirona solo se une al objetivo mostrando especificidad por la proteína E5.

Inhibición y modificación de las interacciones proteína - proteína:

Ejemplo 1. Reactivos que inhiben la unión a SUMO

No se han producido anticuerpos que puedan unirse específica y diferencialmente a SUMO 2 humana (hSUMO2). Se generaron Adhironas contra hSUMO2 y se identificaron múltiples Adhironas que se unen específicamente a SUMO2 en lugar de SUMO1 humana. La Figura 13 demuestra que las Adhironas hSUMO2 tienen un efecto funcional sobre la interacción proteína-proteína al unirse a hSUMO2 y evitar que RNF4, una ubiquitina ligasa E3 específica de poliSUMO se una a hSUMO2. La Adhirona hSUMO2 tiene este efecto sin afectar la ubiquitinación de otras proteínas. En presencia de ATP, hSUMO2 normalmente se une a RNF4 causando ubiquitinación de las proteínas objetivo (frotis negro en la parte superior del gel en el carril 2). En presencia de concentraciones crecientes de las Adhironas (carriles 3 a 9), el nivel de ubiquitinación disminuye.

Ejemplo 2. Reactivos que alteran el coágulo de fibrina y la lisis

El fibrinógeno se cribo para identificar Adhironas que podrían alterar la formación de coágulos y la lisis. Se han identificado numerosos Adhironas que alteran este proceso en muestras de plasma. El gráfico que se muestra en la Figura 14 representa la formación de coágulos y el ensayo de turbidez de lisis. La línea negra representa el curso normal del tiempo de formación de coágulos y lisis. Las líneas grises representan los efectos de cinco Adhironas diferentes en este proceso. Las Adhironas de control que no se unen a fibrinógeno no tienen efecto en este ensayo. Los efectos de las Adhironas incluyen, formación de coágulos reducida, mayor tiempo de lisis y mayor tiempo de coagulación. Esto demuestra la capacidad de las Adhironas para modular la función de la proteína al inhibir las interacciones proteína-proteína. La Figura 15 muestra una imagen confocal de fibrinógeno marcado con fluorescencia FITC después de la formación del coágulo en presencia de una Adhirona de unión a fibrinógeno y una Adhirona de control.

Expresión de Adhironas en células de mamífero:

Se generaron Adhironas contra SUMO2 humana como se describe en el Ejemplo 1 y se expresaron en células HEK293 de mamífero usando el vector de expresión de mamífero pcDNA3.1. Las Adhironas se fusionaron con una etiqueta FLAG y una señal de localización nuclear. Las células de control (sin expresión de Adhirona) y las células que expresan una Adhirona específica de SUMO2 humana se trataron con arsénico (As) durante 2 horas, luego se lavaron y se dejaron recuperar durante 12 y 24 horas. El arsénico provoca un aumento en los cuerpos proteicos de la leucemia promielocítica (PML), pero SUMO regula la degradación de estos cuerpos. Las células se tiñeron usando un anticuerpo anti-FLAG para identificar la Adhirona y para PML (organizador de cuerpos nucleares). La Adhirona SUMO2 humana altera la degradación de PML llevando a un aumento en estos cuerpos. Esto demuestra la capacidad de expresar Adhironas funcionales en células de mamíferos. Los resultados se muestran en la Figura 16.

Cocrystalización y otros métodos de biología estructural para identificar sitios farmacológicos en proteínas:

La Figura 17 muestra la estructura de cocristales de FcγRIIIa (gris) y Adhirona unida (blanco). Se identificaron Adhironas que se unen a FcγRIIIa y luego se usaron en una variedad de ensayos para mostrar que inhiben la unión de IgG, incluidos los ensayos basados en células y SPR. Las Adhironas se cocrystalizaron y se resolvió la estructura (diagrama anterior). Esto identificó sitios farmacológicos, incluido un sitio alostérico, en FcγRIIIa. La Adhirona también es adecuada para estudios de RMN y, como ejemplo, la Figura 18 muestra los espectros HSQC 1H-15N de una Adhirona anti-SUMO de levadura y el complejo Adhirona-SUMO de levadura. La capacidad de recopilar rápidamente datos estructurales sobre Adhironas tendrá muchas aplicaciones, incluida la identificación de posibles sitios de unión a fármacos.

Incorporación a dispositivos electrónicos para desarrollar dispositivos en el punto de atención:

La mutagénesis dirigida al sitio de la secuencia codificante de la Adhirona SUMO2 humana (descrita en el Ejemplo 1) permitió la introducción de una cisteína en el extremo terminal C de la etiqueta de oligohistidina. Esto permitió la inmovilización direccional de la Adhirona a la superficie de un dispositivo electrónico de modo que los bucles de reconocimiento molecular sean accesibles para el analito. Tras la unión de la proteína objetivo a Adhirona en el dispositivo, se puede medir un cambio en la impedancia. El cambio depende de la concentración (como se muestra en la Figura 19), lo que demuestra la presentación efectiva de la Adhirona y la capacidad de las Adhironas para incorporarse productivamente en dispositivos electrónicos como una plataforma para aplicaciones de biosensores.

Este protocolo se repitió para otro Adhirona, en este caso una Adhirona de fibrinógeno humano. Los resultados de este trabajo se muestran en la Figura 26. Una vez más se observó un cambio dependiente de la concentración, y esto se demostró en un rango de concentraciones atómolas hasta micromolares.

Se han identificado Adhironas que se unen a una gama de objetivos:

La biblioteca de Adhironas se ha utilizado para el cribado de una amplia gama de moléculas objetivo utilizando la metodología de despliegue descrita en este documento. La Tabla 5 enumera ejemplos de objetivos contra los cuales se han generado Adhironas; estas incluyen proteínas y moléculas pequeñas. Las Adhironas surgidas contra estos objetivos muestran una gran afinidad y especificidad. Esto demuestra que las Adhironas proporcionan una molécula de almacén versátil y que las bibliotecas creadas con este almacén son efectivas para identificar proteínas de unión artificiales capaces de unirse a una amplia gama de moléculas objetivo. En las Figuras 20-24 se muestran ejemplos de algunas de las secuencias que se han aislado de los cribados contra algunos de los objetivos. Recientemente también se han cribado contra partículas magnéticas producidas por bacterias magnetotrópicas y se han identificado Adhironas que se unen a los epítomos en estos complejos bioinorgánicos multicomponentes.

Tabla 5 - Objetivos contra los cuales han surgido con éxito las Adhironas.

FcγRIIIa - proteína	Interleucina 8 - proteína
P7 - péptido	Albúmina de suero humano - proteína
E5 - péptido	Proteína C reactiva - proteína
3D - proteína	Microglobulina beta 2 - proteína
gB de HSV-1 - proteína	Amiloide P de suero - proteína
gD de HSV-1 - proteína	Receptor 2 del factor de crecimiento endotelial vascular - proteína
gD de HSV-2 - proteína	Receptor 1 de lipoproteína oxidada de baja densidad - proteína
M2 - proteína	Factor 1 inflamatorio de aloinjerto - proteína
HE4 - proteína	CD30 - proteína
Proteína B de unión a calcio S100 - proteína	Proteína CD31
SUMO de levadura - proteína	Beta secretasa 1 - proteína
SUMO1 humana - proteína	Proteína convertasa subtilisina/kexina tipo 9 - proteína
SUMO2 humana - proteína	Miosina e1 - proteína
GST - proteína	Proteína fluorescente verde mejorada - proteína
Hormona de crecimiento - proteína	Fyn - proteína
Factor 1 de crecimiento de fibroblastos - proteína	Lck - proteína
Receptor 1 del factor de crecimiento de fibroblastos - proteína	ZAP70 - proteína
Receptor 3 del factor de crecimiento de fibroblastos - proteína	TUBA8 - péptido
Receptor 3 del factor de crecimiento de fosfo fibroblastos	MRP1 - péptido
Receptor 1 del factor de crecimiento epidérmico - proteína	Proteína 2a de unión a penicilina - proteína
HER2 - proteína	IgE de perro - proteína

Epirregulina - proteína	IgG de caballo - proteína
Anfirregulina - proteína	IgG de perro - proteína
Fibrinógeno - proteína	CRP de perro - proteína
Complemento C3 - proteína	3 moléculas pequeñas
Mioglobina - proteína	Posaconazol - molécula pequeña
CK19 - proteína	GP-73 - proteína
HE4 - proteína	Tiorredoxina - proteína
CD27 - proteína	Transductor y activador de señal de transcripción 1 - proteína
Transductor y activador de señal de transcripción 3 - proteína	Transductor y activador de señal de transcripción 4 - proteína
Transductor y activador de señal de transcripción 5 - proteína	Fosfoinositida 3-quinasa p85 alfa - proteína
Fosfoinositida 3-quinasa p85 beta - proteína	Fosfoinositida 3-quinasa p55 - proteína
Proteína de diferenciación de células de leucemia mieloide - proteína	Linfoma de células β extra grande - proteína
Proteína C transportadora de nucleósidos - proteína de membrana	Factor XIII - proteína
Proteína de la región de la agrupación del punto de interrupción - proteína	Caseína quinasa II A1 - proteína
Caseína quinasa II A2 - proteína	Proteína quinasa C zeta - proteína
Proteína quinasa cGMP dependiente tipo II - proteína	Quinasa 1 relacionada con vacuna - proteína
Hidrofobina 1 - proteína	FusB - proteína
Interleucina 17A - proteína	Interleucina 6 - proteína
Osteocalcina - proteína	Osteopontina - proteína
Hormona paratiroidea - proteína	Esporas bacterianas - proteína
Metaloproteinasas 3 de matriz - proteína	Aminotransferasa - proteína
Citoqueratina 8 - proteína	S100 - A3 - proteína
A6 de S100 - proteína	Transacetilasa - proteína
Inhibidores A1 de serina proteasa - proteína	Inhibidores de serina proteasa A3 - proteína
GAPDH - proteína	p53 - proteína
Resistina - proteína	Lipocalina 2 - proteína
Procalcitonina - proteína	

Referencias

- 5 Astwood, J. D., J. N. Leach, et al., (1996). "Stability of food allergens to digestion in vitro." *Nature Biotechnology* 14(10): 1269-1273.
- Atkinson, H. J., K. A. Johnston, et al., (2004). "Prima facie evidence that a phytocystatin for transgenic plant resistance to nematodes is not a toxic risk in the human diet." *Journal of Nutrition* 134(2): 431-434.
- Bendtsen, J. D., H. Nielsen, et al., (2004). "Improved prediction of signal peptides: SignalP 3.0." *Journal of Molecular Biology* 340(4): 783-795.
- 10 Binz, H. K., M. T. Stumpp, et al., (2003). "Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins." *Journal of Molecular Biology* 332(2): 489-503.
- Bode, W., R. Engh, et al., (1988). "The 2.0 Å X-Ray Crystal-Structure of Chicken Egg-White Cystatin and Its Possible Mode of Interaction With Cysteine Proteinases." *Embo Journal* 7(8): 2593-2599.
- 15 Carter, P. J. (2011). "Introduction to current and future protein therapeutics: A protein engineering perspective." *Experimental Cell Research* 317(9): 1261-1269.
- Dai, M. H., H. E. Fisher, et al., (2007). "The creation of a novel fluorescent protein by guided consensus engineering." *Protein Engineering Design & Selection* 20(2): 69-79.
- Deboer, H. A., L. J. Comstock, et al., (1983). "THE TAC PROMOTER - A FUNCTIONAL HYBRID DERIVED FROM THE TRP AND LAC PROMOTERS." *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 80(1): 21-25.
- 20 Filippova, I. Y., E. N. Lysogorskaya, et al., (1984). "L-Pyroglutamyl-L-Phenylalanyl-L-Leucine-Para-Nitroanilide - a Chromogenic Substrate for Thiol Proteinase Assay." *Analytical Biochemistry* 143(2): 293-297.
- FitzGerald, K. (2000). "In vitro display technologies - new tools for drug discovery." *Drug Discovery Today* 5(6): 253-258.
- 25 Forrer, P., H. K. Binz, et al., (2004). "Consensus design of repeat proteins." *ChemBioChem* 5: 183-189.
- Forrer, P., H. K. Binz, et al., (2004). "Consensus Design of Repeat Proteins." *ChemBioChem* 5(2): 183-189.
- Gebauer, M. and A. Skerra (2009). "Engineered protein scaffolds as next-generation antibody therapeutics." *Current Opinion in Chemical Biology* 13(3): 245-255.

- Grebien, F., O. Hantschel, et al., (2011). "Targeting the SH2-Kinase Interface in Bcr-Abl Inhibits Leukemogenesis." *Cell* 147(2): 306-319.
- Ho, M. and I. Pastan (2009). Mammalian Cell Display for Antibody Engineering. *Methods in Molecular Biology*. A. S. Dimitrov. 525: 337-352.
- 5 Hoffmann, T., L. K. Stadler, et al., (2010). "Structure-function studies of an engineered scaffold protein derived from stefin A. I: Development of the SQM variant." *Protein Eng Des Sel* 23(5): 403-413.
- Hoogenboom, H. R., A. D. Griffiths, et al., (1991). "Multisubunit Proteins on the Surface of Filamentous Phage - Methodologies for Displaying Antibody (Fab) Heavy and Light-Chains." *Nucleic Acids Research* 19(15): 4133-4137.
- Horton, R., Z. Cai, et al., (1990). "Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction." *Biotechniques* 8: 528-535.
- 10 Hutchison, C. A., S. Phillips, et al., (1978). "Mutagenesis at a specific position in a DNA sequence." *Journal of Biological Chemistry* 253(18): 6551-6560.
- Jacobs, S. A., M. D. Diem, et al., (2012). "Design of novel FN3 domains with high stability by a consensus sequence approach." *Protein Engineering Design & Selection* 25(3): 107-117.
- 15 Jacobs, S. A., M. D. Diem, et al., (2012). "Design of novel FN3 domains with high stability by a consensus sequence approach." *Protein Eng Des Sel* 25(3): 107-117.
- Jaekel, C., J. D. Bloom, et al., (2010). "Consensus Protein Design without Phylogenetic Bias." *Journal of Molecular Biology* 399(4): 541-546.
- Karatan, E., M. Merguerian, et al., (2004). "Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain." *Chemistry & Biology* 11(6): 835-844.
- 20 Kayushin, A., M. Korosteleva, et al., (1996). "A convenient approach to the synthesis of trinucleotide phosphoramidites-synthons for the generation of oligonucleotide/peptide libraries." *Nucleic Acids Res* 24: 3748 - 3755.
- Knappik, A., L. Ge, et al., (2000). "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides." *Journal of Molecular Biology* 296(1): 57-86.
- 25 Kohler, G. and C. Milstein (1975). "Continuous cultures of fused cells secreting antibody of predefined specificity." *Nature* 256(5517): 495-497.
- Koide, A., C. W. Bailey, et al., (1998). "The fibronectin type III domain as a scaffold for novel binding proteins." *Journal of Molecular Biology* 284(4): 1141-1151.
- Koide, A., C. W. Bailey, et al., (1998). "The fibronectin type III domain as a scaffold for novel binding proteins." *Journal of Molecular Biology* 284(4): 1141-1151.
- 30 Koide, A., R. N. Gilbreth, et al., (2007). "High-affinity single-domain binding proteins with a binary-code interface." *Proceedings of the National Academy of Sciences of the United States of America* 104(16): 6632-6637.
- Koiwa, H., M. P. D'Urzo, et al., (2001). "Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase." *Plant Journal* 27(5): 383-391.
- 35 Komor, R. S., P. A. Romero, et al., (2012). "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods." *Protein engineering, design & selection : PEDS* 25(12): 827-833.
- Komor, R. S., P. A. Romero, et al., (2012). "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods." *Protein Engineering Design and Selection* 25(12): 827-833.
- Kondo, H., K. Abe, et al., (1991). "Gene Organization of Oryzacystatin-II, a New Cystatin Superfamily Member of Plant-Origin, Is Closely Related to That of Oryzacystatin-I But Different from Those of Animal Cystatins." *FEBS Letters* 278(1): 87-90.
- 40 Kordis, D. and V. Turk (2009). "Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes." *Bmc Evolutionary Biology* 9.
- Krumpe, L., K. Schumacher, et al., (2007). "Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library." *BMC Biotechnology* 7(1): 65.
- 45 Lee, S.-C., K. Park, et al., (2012). "Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering." *Proceedings of the National Academy of Sciences* 109(9): 3299-3304.
- Lehmann, M., C. Loch, et al., (2002). "The consensus concept for thermostability engineering of proteins: further proof of concept." *Protein Engineering* 15(5): 403-411.
- 50 Lehmann, M., R. Lopez-Ulibarri, et al., (2000). "Exchanging the active site between phytases for altering the functional properties of the enzyme." *FEBS Letters* 478(2): 186-190.
- Lehmann, M. and M. Wyss (2001). "Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution." *Current Opinion in Biotechnology* 12(4): 371-375.
- Lilley, C. J., P. E. Urwin, et al., (2004). "Preferential expression of a plant cystatin at nematode feeding sites confers resistance to *Meloidogyne incognita* and *Globodera pallida*." *Plant Biotechnology Journal* 2(1): 3-12.
- 55 Lofblom, J. (2011). "Bacterial display in combinatorial protein engineering." *Biotechnology Journal* 6(9): 1115-1129.
- Lord, P. W., J. N. Selley, et al., (2002). "CINEMA-MX: a modular multiple alignment editor." *Bioinformatics* 18(10): 1402-1403.
- Main, E. R. G., S. E. Jackson, et al., (2003). "The folding and design of repeat proteins: reaching a consensus." *Current Opinion in Structural Biology* 13(4): 482-489.
- 60 Main, E. R. G., A. R. Lowe, et al., (2005). "A recurring theme in protein engineering: the design, stability and folding of repeat proteins." *Current Opinion in Structural Biology* 15(4): 464-471.
- Main, E. R. G., Y. Xiong, et al., (2003). "Design of stable alpha-helical arrays from an idealized TPR motif." *Structure* 11(5): 497-508.
- 65 Makela, A. R. and C. Oker-Blom (2008). "The baculovirus display technology - An evolving instrument for molecular screening and drug delivery." *Combinatorial Chemistry & High Throughput Screening* 11(2): 86-98.

- Margis, R., E. M. Reis, et al., (1998). "Structural and phylogenetic relationships among plant and animal cystatins." *Archives of Biochemistry and Biophysics* 359(1): 24-30.
- McPherson, M. J., P. E. Urwin, et al., (1997). Engineering plant nematode resistance by an anti-feedant approach. *Cellular and Molecular Basis for Plant-Nematode Interactions*. C. Fenoll, S. Ohl and F. Grundler. The Netherlands, Kluwer: 237-249.
- Melo, F. R., M. O. Mello, et al., (2003). "Use of phage display to select novel cystatins specific for *Acanthoscelides obtectus* cysteine proteinases." *Biochimica Et Biophysica Acta-Proteins and Proteomics* 1651(1-2): 146-152.
- Mosavi, L. K., T. J. Cammett, et al., (2004). "The ankyrin repeat as molecular architecture for protein recognition." *Protein Science* 13(6): 1435-1448.
- Mosavi, L. K., D. L. Minor, et al., (2002). "Consensus-derived structural determinants of the ankyrin repeat motif." *Proceedings of the National Academy of Sciences* 99(25): 16029-16034.
- Mullis, K., F. Faloona, et al., (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." *Cold Spring Harb Symp Quant Biol* 51 Pt 1: 263-273.
- Nagata, K., N. Kudo, et al., (2000). "Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. japonica." *Biochemistry* 39(48): 14753-14760.
- Nixon AE, W. C. (2006). "Engineered protein inhibitors of proteases." *Curr Opin Drug Discov Devel* 9(2): 261-268.
- Nord, K., J. Nilsson, et al., (1995). "A combinatorial library of an α -helical bacterial receptor domain." *Protein Engineering* 8(6): 601-608.
- Odegrip, R., D. Coomber, et al., (2004). "CIS display: In vitro selection of peptides from libraries of protein-DNA complexes." *Proceedings of the National Academy of Sciences of the United States of America* 101(9): 2806-2810.
- Parizek, P., L. Kummer, et al., (2012). "Designed Ankyrin Repeat Proteins (DARPs) as Novel Isoform-Specific Intracellular Inhibitors of c-Jun N-Terminal Kinases." *ACS Chemical Biology* 7(8): 1356-1366.
- Parmeggiani, F., R. Pellarin, et al., (2008). "Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core." *Journal of Molecular Biology* 376(5): 1282-1304.
- Parry-Smith, D. J., A. W. R. Payne, et al., (1998). "CINEMA - a novel colour interactive editor for multiple alignments (Reprinted from *Gene*, vol 221, pg GC57-GC63, 1998)." *Gene* 221(1): GC57-GC63.
- Polizzi, K. M., J. F. Chaparro-Riggers, et al., (2006). "Structure-guided consensus approach to create a more thermostable penicillin G acylase." *Biotechnology Journal* 1(5): 531-536.
- Reichert, J. M. (2010). "Antibodies to watch in 2010." *MAbs*. 2(1): 84-100.
- Saiki RK, S. S., Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." *Science* 230(4732): 1350-1354.
- Schlehuber, S. and A. Skerra (2005). "Anticalins as an alternative to antibody technology." *Expert Opinion on Biological Therapy* 5(11): 1453-1462.
- Skerra, A. (2007). "Alternative non-antibody scaffolds for molecular recognition." *Current Opinion in Biotechnology* 18: 295-304.
- Smith, G. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." *Science* 228(4705): 1315-1317.
- Song, I., M. Taylor, et al., (1995). "Inhibition of Cysteine Proteinases by Carica-Papaya Cystatin Produced in *Escherichia-Coli*." *Gene* 162(2): 221-224.
- Song, J., L. K. Durrin, et al., (2004). "Identification of a SUMO-binding motif that recognizes SUMO-modified proteins." *Proc Natl Acad Sci U S A* 101(40): 14373-14378.
- Song, J., Z. Zhang, et al., (2005). "Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation." *J Biol Chem* 280(48): 40122-40129.
- Stadler, L. K., T. Hoffmann, et al., (2011). "Structure-function studies of an engineered scaffold protein derived from Stefin A. II: Development and applications of the SQT variant." *Protein Eng Des Sel* 24(9): 751-763.
- Steipe, B. (2004). "Consensus-based engineering of protein stability: From intrabodies to thermostable enzymes." *Protein Engineering* 388: 176-186.
- Steipe, B., B. Schiller, et al., (1994). "Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain." *Journal of Molecular Biology* 240(3): 188-192.
- Steipe, B., B. Schiller, et al., (1994). "Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain." *Journal of Molecular Biology* 240(3): 188-192.
- Stubbs, M. T., B. Laber, et al., (1990). "The Refined 2.4a X-Ray Crystal-Structure of Recombinant Human Stefin- B in Complex with the Cysteine Proteinase Papain - a Novel Type of Proteinase-Inhibitor Interaction." *Embo Journal* 9(6): 1939-1947.
- Studier, F. W. y B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." *J Mol Biol* 189(1): 113-130.
- Theurillat, J.-P., B. Dreier, et al., (2010). "Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer." *Mod Pathol* 23(9): 1289-1297.
- Traxlmayr, M. W. and C. Obinger (2012). "Directed evolution of proteins for increased stability and expression using yeast display." *Archives of Biochemistry and Biophysics* 526(2): 174-180.
- Urwin, P. E., H. J. Atkinson, et al., (1995). "Engineered Oryzacystatin-I Expressed in Transgenic Hairy Roots Confers Resistance to *Globodera-Pallida*." *Plant Journal* 8(1): 121-131.
- Urwin, P. E., J. Green, et al., (2003). "Expression of a plant cystatin confers partial resistance to *Globodera*, full resistance is achieved by pyramiding a cystatin with natural resistance." *Molecular Breeding* 12(3): 263-269.

- Urwin, P. E., A. Levesley, et al., (2000). "Transgenic resistance to the nematode *Rotylenchulus reniformis* conferred by *Arabidopsis thaliana* plants expressing proteinase inhibitors." *Molecular Breeding* 6(3): 257-264.
- Urwin, P. E., C. J. Lilley, et al., (1997). "Resistance to both cyst and root-knot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin." *Plant Journal* 12(2): 455-461.
- 5 Urwin, P. E., M. J. McPherson, et al., (1998). "Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs." *Planta* 204(4): 472-479.
- Urwin, P. E., K. M. Troth, et al., (2001). "Effective transgenic resistance to *Globodera pallida* in potato field trials." *Molecular Breeding* 8(1): 95-101.
- 10 Virnekas, B., L. Ge, et al., (1994). "Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis." *Nucleic Acids Res* 22: 5600 - 5607.
- Von Behring, E., Kitasato, S., (1890). "Über das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren." *Deutsche Medizinische Wochenschrift* 16: 1113-1114.
- Wojcik, J., O. Hantschel, et al., (2010). "A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain." *Nat Struct Mol Biol* 17(4): 519-527.
- 15 Woodman, R., J. T. H. Yeh, et al., (2005). "Design and validation of a neutral protein scaffold for the presentation of peptide aptamers." *Journal of Molecular Biology* 352(5): 1118-1133.
- Wurch, T., A. Pierre, et al., (2012). "Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept." *Trends in Biotechnology* 30(11): 575-582.
- 20 Listado de secuencias
- <110> University of Leeds
- <120> Proteínas sintéticas novedosas
- 25 <130> P206256GB
- <160> 81
- 30 <170> PatentIn versión 3.5
- <210> 1
- <211> 81
- <212> PRT
- 35 <213> Secuencia Artificial
- <220>
- <223> Armazón Sintético
- 40 <400> 1
- Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His
1 5 10 15
- Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala
20 25 30
- Lys Glu Gln Val Val Ala Gly Thr Met Tyr Tyr Leu Thr Leu Glu Ala
35 40 45
- Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala Lys Val Trp Val Lys
50 55 60
- Pro Trp Glu Asn Phe Lys Glu Leu Gln Glu Phe Lys Pro Val Gly Asp
65 70 75 80
- Ala
- <210> 2
- 45 <211> 84
- <212> PRT

ES 2 784 211 T3

<213> Secuencia Artificial

<220>

<223> Armazón Sintético

5

<400> 2

Gly Asn Glu Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val
1 5 10 15

Asp Glu His Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val
20 25 30

Val Lys Ala Lys Glu Gln Val Val Ala Gly Thr Met Tyr Tyr Leu Thr
35 40 45

Leu Glu Ala Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala Lys Val
50 55 60

Trp Val Lys Pro Trp Glu Asn Phe Lys Glu Leu Gln Glu Phe Lys Pro
65 70 75 80

10

Val Gly Asp Ala

<210> 3

<211> 92

<212> PRT

15

<213> Secuencia Artificial

<220>

<223> Armazón Sintético

20

<400> 3

Ala Thr Gly Val Arg Ala Val Pro Gly Asn Glu Asn Ser Leu Glu Ile
1 5 10 15

Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His Asn Lys Lys Glu Asn
20 25 30

Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala Lys Glu Gln Val Val
35 40 45

Ala Gly Thr Met Tyr Tyr Leu Thr Leu Glu Ala Lys Asp Gly Gly Lys
50 55 60

Lys Lys Leu Tyr Glu Ala Lys Val Trp Val Lys Pro Trp Glu Asn Phe
65 70 75 80

Lys Glu Leu Gln Glu Phe Lys Pro Val Gly Asp Ala
85 90

25

<210> 4

<211> 74

<212> PRT

<213> Secuencia Artificial

<220>
 <223> Armazón Sintético

5 <220>
 <221> NON_CONS
 <222> (36)..(37)
 <223> Región variable entre estos residuos

10 <220>
 <221> NON_CONS
 <222> (60)..(61)
 <223> Región variable entre estos residuos
 <400> 4

15

Asn	Ser	Leu	Glu	Ile	Glu	Glu	Leu	Ala	Arg	Phe	Ala	Val	Asp	Glu	His
1				5					10					15	
Asn	Lys	Lys	Glu	Asn	Ala	Leu	Leu	Glu	Phe	Val	Arg	Val	Val	Lys	Ala
			20					25					30		
Lys	Glu	Gln	Thr	Met	Tyr	Tyr	Leu	Thr	Leu	Glu	Ala	Lys	Asp	Gly	Gly
		35					40					45			
Lys	Lys	Lys	Leu	Tyr	Glu	Ala	Lys	Val	Trp	Val	Lys	Asn	Phe	Lys	Glu
	50					55					60				
Leu	Gln	Glu	Phe	Lys	Pro	Val	Gly	Asp	Ala						
65					70										

<210> 5
 <211> 74
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Armazón sintética

25 <220>
 <221> NON_CONS
 <222> (35)..(36)
 <223> Región variable entre estos residuos

30 <220>
 <221> NON_CONS
 <222> (60)..(61)
 <223> Región variable entre estos residuos

35 <400> 5

ES 2 784 211 T3

Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His
1 5 10 15

Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala
20 25 30

Lys Glu Gln Thr Met Tyr Tyr Leu Thr Leu Glu Ala Lys Asp Gly Gly
35 40 45

Lys Lys Lys Leu Tyr Glu Ala Lys Val Trp Val Lys Asn Phe Lys Glu
50 55 60

Leu Gln Glu Phe Lys Pro Val Gly Asp Ala
65 70

- 5 <210> 6
<211> 92
<212> PRT
<213> Secuencia Artificial
<220>
10 <223> Armazón Sintético

<220>
<221> característica nueva
<222> (36)..(44)
15 <223> Xaa puede ser cualquier aminoácido natural

<220>
<221> característica nueva
<222> (70)..(78)
20 <223> Xaa puede ser cualquier aminoácido natural

<400> 6

Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His
1 5 10 15

Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala
20 25 30

Lys Glu Gln Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Met Tyr Tyr
35 40 45

Leu Thr Leu Glu Ala Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala
50 55 60

Lys Val Trp Val Lys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Phe
65 70 75 80

Lys Glu Leu Gln Glu Phe Lys Pro Val Gly Asp Ala
85 90

- 25 <210> 7
<211> 183
<212> ADN

ES 2 784 211 T3

	<213> Secuencia Artificial	
	<220>	
	<223> Secuencia sintética	
5	<400> 7	
	aacgctctgc tggaattcgt tcgtggtggt aaagctaaag aacagggtgt tgctggtacc	60
	atgtactacc tgaccctgga agctaaagac ggtggtaaaa agaaactgta cgaagctaaa	120
	gtttgggtta aaccgtggga aaacttcaaa gaactgcagg agttcaaacc ggttggtgac	180
10	gct	183
	<210> 8	
	<211> 252	
	<212> ADN	
15	<213> Secuencia Artificial	
	<220>	
	<223> Secuencia sintética	
20	<400> 8	
	ggtaacgaaa actccctgga aatcgaagaa ctggctcggt tcgctggtga cgaacacaac	60
	aaaaaagaaa acgctctgct ggaattcgtt cgtggtgtta aagctaaaga acagggtggt	120
	gctggtacca tgtactacct gaccctggaa gctaaagacg gtggtaaaaa gaaactgtac	180
	gaagctaaag tttgggttaa accgtgggaa aacttcaaag aactgcagga gttcaaaccg	240
	gttggtgacg ct	252
	<210> 9	
	<211> 276	
25	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> Secuencia sintética	
30	<400> 9	
	gctaccggtg ttcgtgcagt tccgggtaac gaaaactccc tggaaatcga agaactggct	60
	cgtttcgctg ttgacgaaca caacaaaaaa gaaaacgctc tgctggaatt cgttcgtggt	120
	gttaaagcta aagaacagggt tgttgctggt accatgtact acctgaccct ggaagctaaa	180
	gacggtggta aaaagaaact gtacgaagct aaagtttggg ttaaaccgtg ggaaaacttc	240
	aaagaactgc aggagttaa accggttggt gacgct	276
35	<210> 10	
	<211> 44	
	<212> ADN	
	<213> Secuencia Artificial	
40	<220>	
	<223> Cebador cewcF	
	<400> 10	
45	attagcgcc cagccggcca tggccagcga ggaccgctcc cggc	44

<210> 11
 <211> 38
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> cebador cewcR
 <400> 11
 10 cgctgtactt gcggcgcgcc tggcacttgc ttccagc 38
 <210> 12
 <211> 35
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido HisF
 20 <400> 12
 ggccgcagag gatcgcatca ccatcacat cacgg 35
 <210> 13
 <211> 35
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido HisR
 30 <400> 13
 ggccccgtga tggatgatgt gatgcgatcc tctgc 35
 <210> 14
 <211> 33
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 40 <223> Cebador XhoF
 <400> 14
 atcacgctcg agcagaacaa aaactcatct cag 33
 45 <210> 15
 <211> 35
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> Cebador XhoR
 <400> 15
 55 tgttctgctc gagcgtgatg gtgatggtga tggcg 35
 <210> 16
 <211> 21
 <212> ADN
 <213> Secuencia Artificial
 60 <220>
 <223> Cebador BamHIR
 <400> 16
 65 tggccttgat attcacaac g 21

5 <210> 17
 <211> 24
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador M13R
 10 <400> 17
 agcggataac aatttcacac agga 24
 <210> 18
 <211> 25
 15 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador CystaSDFW
 20 <400> 18
 caccatgaaa tcactattgc ttacg 25
 <210> 19
 25 <211> 26
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> Cebador CystaSDREV
 <400> 19
 ctactagtga tggatgatggt gatgcg 26
 35 <210> 20
 <211> 26
 <212> PRT
 <213> Secuencia Artificial
 40 <220>
 <223> Etiqueta HIS
 <400> 20
 Arg Gly Ser His His His His His His Ala Arg Ala Glu Gln Lys Leu
 1 5 10 15
 45 Ile Ser Glu Glu Asp Leu Asn Gly Ala Ala
 20 25
 <210> 21
 <211> 18
 <212> ADN
 50 <213> Secuencia Artificial
 <220>
 <223> Cebador directo
 55 <400> 21
 tctggcggtt tctgcgtc 18
 <210> 22
 <211> 21
 60 <212> ADN

<213> Secuencia Artificial

<220>

<223> Cebador inverso

5

<400> 22
ctgttcttc gcttaacaa c 21

<210> 23

10 <211> 69
<212> ADN
<213> Secuencia Artificial

<220>

15 <223> Cebador de bucle directo

<220>

<221> característica nueva

20 <222> (22)..(48)
<223> n es a, c, g, o t

<400> 23
gttggttaaag cgaaagaaca gnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnac catgtaccac 60

ttgaccctg 69

25 <210> 24
<211> 78
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Cebador de bucle inverso

<220>

35 <221> característica nueva
<222> (30)..(56)
<223> n es a, c, g, o t

<400> 24

40 ctgcggaact cctgcagttc tttgaagttn nnnnnnnnnnn nnnnnnnnnnn nnnnnnctta 60

acccaaactt tcgcttcg 78

<210> 25
<211> 9

45 <212> PRT
<213> Secuencia Artificial

<220>

<223> Péptido de bucle

50 <400> 25

Trp Asp Leu Thr Gly Asn Val Asp Thr
1 5

55 <210> 26

<211> 9
<212> PRT
<213> Secuencia Artificial

60 <220>

<400> 31

Trp Val Asp Met Asp Tyr Tyr Trp Arg
1 5

5 <210> 32
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Péptido de bucle

<400> 32

15 Ile Asp Leu Thr Gln Thr Glu Ile Val
1 5

<210> 33
<211> 9
<212> PRT
<213> Secuencia Artificial

20 <220>
<223> Péptido de bucle

25 <400> 33

Ile Asp Leu Thr Asp Val Trp Ile Asp
1 5

30 <210> 34
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Péptido de bucle

<400> 34

Ile Ile Ile His Glu Asn Asp Ala Asp
1 5

40 <210> 35
<211> 9
<212> PRT
<213> Secuencia Artificial

45 <220>
<223> Péptido de bucle

50 <400> 35

Trp Ile Leu Asn Asn Thr Gln Phe Ile
1 5

55 <210> 36
<211> 9
<212> PRT
<213> Secuencia Artificial

60 <220>
<223> Péptido de bucle

<400> 36

Trp Tyr Glu Arg Ser Glu Asn Trp Asp
1 5

5 <210> 37
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Péptido de bucle

<400> 37

15 Trp Asp Leu Thr Thr Pro Ile Asn Ile
1 5

<210> 38
<211> 9
<212> PRT
<213> Secuencia Artificial

20 <220>
<223> Péptido de bucle

25 <400> 38

Trp Phe Asp Asp Glu Tyr Asp Trp Ile
1 5

30 <210> 39
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Péptido de bucle

<400> 39

Ile Asp Leu Thr Gln Pro His Asp Ser
1 5

40 <210> 40
<211> 9
<212> PRT
<213> Secuencia Artificial

45 <220>
<223> Péptido de bucle

50 <400> 40

Ile Asp Leu Thr Gln Ser Phe Asp Met
1 5

<210> 41

55 <211> 9
<212> PRT
<213> Secuencia Artificial

60 <220>
<223> Péptido de bucle

<400> 41

Trp Tyr Leu Leu Asp Val Met Asp Asp
1 5

5 <210> 42
<211> 9
<212> PRT

10 <213> Secuencia Artificial
<220>
<223> Péptido de bucle

15 <400> 42

Trp Ile Asp Arg Gly Gln Tyr Trp Asp
1 5

20 <210> 43
<211> 9
<212> PRT
<213> Secuencia Artificial
<220>
<223> Péptido de bucle

25 <400> 43

Trp Ser Glu Ala Asp Asn Asp Trp His
1 5

30 <210> 44
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Péptido de bucle
<400> 44

40 Ile Asp Leu Thr Gly Gln Trp Leu Phe
1 5

<210> 45
<211> 9
<212> PRT
<213> Secuencia Artificial
<220>
<223> Péptido de bucle

45 <400> 45

Ile Asp Leu Thr Gln Ser Phe Asp Met
1 5

50 <210> 46
<211> 9
<212> PRT
<213> Secuencia Artificial

60 <220>

<223> Péptido de bucle

<400> 46

5 Ile Asp Leu Thr Gln Ser Phe Asp Met
 1 5

<210> 47
<211> 9
<212> PRT
10 <213> Secuencia Artificial

<220>
<223> Péptido de bucle

15 <400> 47

 Ile Asp Leu Thr Gln Pro His Asp Ser
 1 5

<210> 48
20 <211> 9
<212> PRT
<213> Secuencia Artificial

<220>
25 <223> Péptido de bucle

<400> 48

 Trp Glu Asp Phe Gln Thr His Trp Glu
 1 5

30 <210> 49
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Péptido de bucle

<400> 49

40 <210> 50
<211> 9
<212> PRT
45 <213> Secuencia Artificial

<220>
<223> Péptido de bucle

50 <400> 50

 Asp Ile Asn Gln Tyr Trp His Ser Met
 1 5

55 <210> 51
<211> 9
<212> PRT
<213> Secuencia Artificial

60 <220>
<223> Péptido de bucle

	<400> 51	
		Gly Ile Gln Gln Asn Pro Ser His Ala
5		1 5
	<210> 52	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
10		
	<220>	
	<223> Péptido de bucle	
	<400> 52	
15		Gly Leu Thr Asn Glu Ile Gln Lys Met
		1 5
	<210> 53	
	<211> 9	
20		
	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
	<223> Péptido de bucle	
25		
	<400> 53	
		Gly Leu Thr Asn Glu Ile Gln Lys Met
		1 5
30		
	<210> 54	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
35		
	<220>	
	<223> Péptido de bucle	
	<400> 54	
		Pro Glu Pro Ile His Ser His His Ser
40		1 5
	<210> 55	
	<211> 9	
	<212> PRT	
45		
	<213> Secuencia Artificial	
	<220>	
	<223> Péptido de bucle	
50		
	<400> 55	
		Met Asp Glu Ile Trp Ala Glu Tyr Ala
		1 5
	<210> 56	
55		
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
60		
	<223> Péptido de bucle	

ES 2 784 211 T3

<400> 56

Glu Pro Gly Ile Ile Pro Ile Val His
1 5

5 <210> 57
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Péptido de bucle

<400> 57

Gly Leu Met Thr Gln Thr Asn Ser Met
1 5

15 <210> 58
<211> 9
<212> PRT
<213> Secuencia Artificial

20 <220>
<223> Péptido de bucle

<400> 58

25 Gly Ile Met Asp Gly Leu Asn Lys Tyr
1 5

<210> 59
<211> 10
<212> PRT
<213> Secuencia Artificial

30 <220>
<223> Péptido de bucle

35 <400> 59

Val Leu Glu Gly Pro Asp Arg Trp Thr Val
1 5 10

40 <210> 60
<211> 9
<212> PRT
<213> Secuencia Artificial

45 <220>
<223> Péptido de bucle

<400> 60

Arg Asp Tyr Gly Phe Thr Leu Val Pro
1 5

50 <210> 61
<211> 9
<212> PRT
<213> Secuencia Artificial

55 <220>
<223> Péptido de bucle

<400> 61

60

ES 2 784 211 T3

		Tyr Glu Asp Tyr Gln Thr Pro Met Tyr
		1 5
	<210> 62	
	<211> 9	
5	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
	<223> Péptido de bucle	
10	<400> 62	
		Asp Tyr Ala Ala Thr Asp Leu Tyr Trp
		1 5
15	<210> 63	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
20	<220>	
	<223> Péptido de bucle	
	<400> 63	
		Tyr Glu Glu Asp Glu Tyr Trp Arg Met
		1 5
25	<210> 64	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
30	<220>	
	<223> Péptido de bucle	
	<400> 64	
35		Pro Ile Asp Ser Asn Phe Thr Gly Thr
		1 5
	<210> 65	
	<211> 9	
40	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
	<223> Péptido de bucle	
45	<400> 65	
		His Asp Arg Arg Tyr Lys Gln Ala Glu
		1 5
50	<210> 66	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
55	<220>	
	<223> Péptido de bucle	
	<400> 66	
		Ile His Asn Gly Tyr Thr Ile Met Asp
60		1 5

5	<210> 67 <211> 9 <212> PRT <213> Secuencia Artificial <220> <223> Péptido de bucle	
10	<400> 67	Leu Asp Leu Glu Thr Trp Gln His Phe 1 5
15	<210> 68 <211> 9 <212> PRT <213> Secuencia Artificial <220> <223> Péptido de bucle	
20	<400> 68	Pro Leu Trp Gln Tyr Asp Ala Gln Tyr 1 5
25	<210> 69 <211> 9 <212> PRT <213> Secuencia Artificial <220> <223> Péptido de bucle	
30	<400> 69	Pro Leu Trp Gln Tyr Asp Ala Gln Tyr 1 5
35	<210> 70 <211> 9 <212> PRT <213> Secuencia Artificial <220> <223> Péptido de bucle	
40	<400> 70	Pro Ile Asp Ser Asn Phe Thr Gly Thr 1 5
45	<210> 71 <211> 9 <212> PRT <213> Secuencia Artificial <220> <223> Péptido de bucle	
50	<400> 71	Pro His Asp Glu Leu Asn Trp Asn Met 1 5
55	<210> 72	

<211> 9
 <212> PRT
 <213> Secuencia Artificial

5 <220>
 <223> Péptido de bucle

<400> 72

10 Asp Val Gly Gln Leu Leu Ser Gly Ile
 1 5

<210> 73
 <211> 7
 <212> PRT
 <213> Secuencia Artificial

15 <220>
 <223> Sitio de unión a SUMO1

20 <400> 73

Val Asp Val Ile Asp Leu Thr
 1 5

<210> 74
 <211> 101
 <212> PRT
 <213> Secuencia Artificial

25 <220>
 <223> Armazón Sintético

30 <400> 74

Ala Thr Gly Val Arg Ala Val Pro Gly Asn Glu Asn Ser Leu Glu Ile
 1 5 10 15

Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His Asn Lys Lys Glu Asn
 20 25 30

Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala Lys Glu Gln Val Val
 35 40 45

Ala Gly Thr Met Tyr Tyr Leu Thr Leu Glu Ala Lys Asp Gly Gly Lys
 50 55 60

Lys Lys Leu Tyr Glu Ala Lys Val Trp Val Lys Pro Trp Glu Asn Phe
 65 70 75 80

Lys Glu Leu Gln Glu Phe Lys Pro Val Gly Asp Ala Ala Ala Ala His
 85 90 95

His His His His His
 100

<210> 75
 <211> 93
 <212> PRT
 <213> Secuencia Artificial

35 <220>
 <223> Armazón sintética

40

ES 2 784 211 T3

<400> 75

Gly Asn Glu Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val
1 5 10 15

Asp Glu His Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val
20 25 30

Val Lys Ala Lys Glu Gln Val Val Ala Gly Thr Met Tyr Tyr Leu Thr
35 40 45

Leu Glu Ala Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala Lys Val
50 55 60

Trp Val Lys Pro Trp Glu Asn Phe Lys Glu Leu Gln Glu Phe Lys Pro
65 70 75 80

Val Gly Asp Ala Ala Ala Ala His His His His His His
85 90

- 5 <210> 76
- <211> 90
- <212> PRT
- <213> Secuencia Artificial
- <220>
- 10 <223> Armazón sintética

<400> 76

Asn Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His
1 5 10 15

Asn Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala
20 25 30

Lys Glu Gln Val Val Ala Gly Thr Met Tyr Tyr Leu Thr Leu Glu Ala
35 40 45

Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala Lys Val Trp Val Lys
50 55 60

Pro Trp Glu Asn Phe Lys Glu Leu Gln Glu Phe Lys Pro Val Gly Asp
65 70 75 80

Ala Ala Ala Ala His His His His His His
85 90

- 15 <210> 77
- <211> 102
- <212> PRT
- <213> Secuencia Artificial

- 20 <220>
- <223> Armazón sintética

<400> 77

25

ES 2 784 211 T3

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

5 <210> 78
 <211> 94
 <212> PRT
 <213> Secuencia Artificial

10 <220>
 <223> Armazón sintética
 <400> 78

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

15 <210> 79
 <211> 116
 <212> PRT
 20 <213> Secuencia Artificial
 <220>

<223> Armazón sintética
 <220>
 <221> característica nueva
 5 <222> (2)..(7)
 <223> Xaa puede ser cualquier aminoácido natural
 <220>
 <221> característica nueva
 10 <222> (51)..(59)
 <223> Xaa puede ser cualquier aminoácido natural
 <220>
 <221> característica nueva
 15 <222> (85).. (93)
 <223> Xaa puede ser cualquier aminoácido natural
 <400> 79
 Met Xaa Xaa Xaa Xaa Xaa Xaa Val Arg Ala Val Pro Gly Asn Glu Asn
 1 5 10 15
 Ser Leu Glu Ile Glu Glu Leu Ala Arg Phe Ala Val Asp Glu His Asn
 20 25 30
 Lys Lys Glu Asn Ala Leu Leu Glu Phe Val Arg Val Val Lys Ala Lys
 35 40 45
 Glu Gln Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Met Tyr Tyr Leu
 50 55 60
 Thr Leu Glu Ala Lys Asp Gly Gly Lys Lys Lys Leu Tyr Glu Ala Lys
 65 70 75 80
 Val Trp Val Lys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Phe Lys
 85 90 95
 Glu Leu Gln Glu Phe Lys Pro Val Gly Asp Ala Ala Ala Ala His His
 100 105 110
 His His His His
 115
 20 <210> 80
 <211> 12
 <212> PRT
 <213> Secuencia Artificial
 25 <220>
 <223> Proteína sintética
 <400> 80
 30 Met Ala Thr Gly Val Arg Ala Val Pro Gly Asn Glu
 1 5 10
 <210> 81
 <211> 10
 35 <212> PRT

<213> Secuencia Artificial

<220>

<223> Secuencia de consenso

5

<220>

<221> VARIANTE

<222> (1)..(1)

10 <223> Leu, Val o Ile

<220>

<221> VARIANTE

<222> (2)..(2)

<223> Ala, Gly o Tyr

15

<220>

<221> VARIANTE

<222> (3)..(3)

<223> Arg, Lys o Glu

20

<220>

<221> VARIANTE

<222> (4)..(4)

<223> Phe o Tyr

25

<220>

<221> VARIANTE

<222> (5)..(5)

<223> Ala o Ser

30

<220>

<221> VARIANTE

<222> (6)..(6)

<223> Val o Ile

35

<220>

<221> VARIANTE

<222> (7)..(7)

<223> Cualquier aminoácido

40

<220>

<221> VARIANTE

<222> (8)..(8)

<223> Glu, Asp, Gln o Val

45

<220>

<221> VARIANTE

<222> (9)..(9)

<223> His, Tyr, Phe o Gln

50

<400> 81

Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Asn
1				5						10

REIVINDICACIONES

1. Una proteína de armazón sintética que comprende la secuencia:

5 NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X_n)TMYLTLEAKDGGKKKLY-
EAKVWVK(X_n)NFKELQEFKPVGDA (SEQ ID NO: 4) o una secuencia de aminoácidos que es al menos 85% idéntica
a la secuencia de aminoácidos de la SEQ ID NO: 4,
en la que X_n es una secuencia peptídica heteróloga, en la que X es cualquier aminoácido y n es el número de
aminoácidos de la secuencia, y en la que n es de 3 a 30, en la que la proteína de armazón sintética tiene una
10 temperatura de fusión, T_m, de al menos 70 °C.

2. La proteína sintética de la reivindicación 1,
en la que la secuencia peptídica heteróloga tiene de 3 a 20 aminoácidos de longitud, más preferiblemente de 5 a 13
aminoácidos de longitud.

3. La proteína sintética de acuerdo con la reivindicación 1 o 2, en la que
n es de 4 a 20, más preferiblemente de 4 a 15, en la que convenientemente la proteína sintética comprende la
secuencia:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X₅₋₁₃)TMYLTLEA

KDGGKKKLYEAKVWVK(X₅₋₁₃)NFKELQEFKPVGDA (SEQ ID NO 5),

convenientemente que comprende la secuencia:

NSLEIEELARFAVDEHNKKENALLEFVRVVKAKEQ(X₉)TMYLTLEA

KDGGKKKLYEAKVWVK(X₉) NFKELQEFKPVGDA (SEQ ID NO 6).

4. La proteína sintética de cualquier reivindicación precedente, que comprende una secuencia de aminoácidos
adicional en o cerca del extremo terminal N,
convenientemente en la que la secuencia de aminoácidos adicional en el extremo terminal N comprende la secuencia
MATGVRVPGNE (SEQ ID NO 80), o
en la que la secuencia de aminoácidos adicional en el extremo terminal N comprende una secuencia peptídica
heteróloga adicional.

5. La proteína sintética de cualquier reivindicación precedente, que tiene una T_m de 80 ° C o superior.

6. La proteína sintética de cualquier reivindicación precedente que comprende un enlazador o etiqueta, que está
conectado a un sustrato o fracción,
convenientemente en la que la fracción es un marcador, vehículo o proteína,
y convenientemente en la que la fracción, sustrato, enlazador o etiqueta está unido a uno o ambos de los extremos
terminales C y/o N de la proteína,
convenientemente en la que la proteína sintética es una proteína de fusión.

7. Una biblioteca que comprende una población de proteínas sintéticas de acuerdo con una cualquiera de las
reivindicaciones 1 a 6, en la que miembros de la población comprenden una variedad de péptidos heterólogos que
tienen diferentes secuencias,
preferiblemente en la que la biblioteca tiene una complejidad de 10⁸ o superior, más preferiblemente 10⁹ o superior, y
lo más preferiblemente 10¹⁰ o superior,
preferiblemente en la que la biblioteca comprende una población de las proteínas sintéticas de la presente invención
adaptadas para su despliegue en un sistema de despliegue adecuado para biopaneos.

8. Un polinucleótido que codifica una proteína sintética de acuerdo con una cualquiera de las reivindicaciones 1 a 6.

9. Un vector recombinante que comprende un polinucleótido de acuerdo con la reivindicación 8 adaptado para la
expresión en una célula huésped,
convenientemente, en el que dicho vector comprende una secuencia de control de la expresión unida operativamente
a la secuencia de ácido nucleico que codifica la proteína sintética para controlar la expresión de la proteína sintética.

10. Una célula o cultivo de células adaptado para expresar una proteína de acuerdo con una cualquiera de las
reivindicaciones 1 a 6, preferiblemente en la que la célula o cultivo de células comprende un polinucleótido de acuerdo
con la reivindicación 8 o un vector de acuerdo con la reivindicación 9.

11. Un método de cribado para una proteína sintética que se une a un objetivo, comprendiendo el método:

a) proporcionar una biblioteca de acuerdo con la reivindicación 7;

b) exponer dicha biblioteca a un objetivo, convenientemente una proteína/péptido, un ácido nucleico o una molécula objetivo pequeña; y
c) seleccionar proteínas de armazón sintéticas que se unen a dicho objetivo,

5 preferiblemente en el que el método de cribado utiliza un sistema de despliegue, preferiblemente un método de despliegue en fagos.

10 12. Uso de una proteína sintética de acuerdo con una cualquiera de las reivindicaciones 1 a 6, o un ácido nucleico que codifica dicha proteína sintética, como armazón para la presentación del péptido en investigación, en control ambiental y de seguridad, en la biología de síntesis, en detección, o diagnóstico de objetivos.

13. Una proteína sintética de acuerdo con una cualquiera de las reivindicaciones 1 a 6, o un ácido nucleico que codifica dicha proteína sintética, para uso en terapia.

^{5'}ATTAGGGCCAGCGCGCATGGCGCTCTGCTGGGTGGTGTTCGTGCAGTTCGGGGTAACG3'
M A A L L G G V R A V P G N E N S L E I E E L A R F A V D E
ATGGCGCTCTGCTGGGTGGTGTTCGTGCAGTTCGGGTACGAAACTCCTCGAATCGAAGAACTGGCTGTTTCGCTGTTGACGAA
TACGGCGGAGACGCCACCCACACAGCAGGTCAAGGCCCATGCTTTGAGGGACCTTTAGCTTCTTGACCGAGCAAGCGACAACTGCTT
3'GCCCATTCGCTTTTGAGGGACCTTTAGCTTCTTGACCGAGCAAGCGGCAACTGCTT
P2

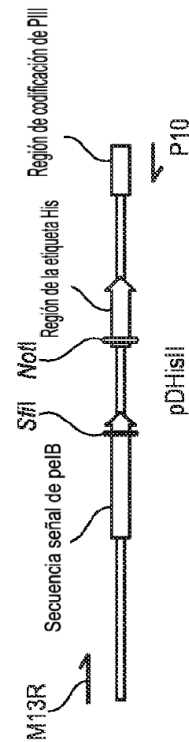
^{3'}CACAACAAAAAGAAACGCTCTGCTGGATTGCTGTTGTTAAAGCTAAAGCAAGGTTGTC3'
H N K K E N A L L E F V R V V K A K E Q V V A G T M Y V L T L E A K D G G
CACAACAAAAAGAAACGCTCTGCTGGATTGCTGTTGTTAAAGCTAAAGCAAGGTTGTTGCTGGTACCATGTACTACCTGACCGTGGAGCTAAAGACGGTGGT
GTGTTGTTTTCCTTTTCGAGACGACCTTAAGCAAGCACAAATTCGATTCTTGTCACACAGCACCATGGTACATGATGGACTGGGACCTTCGATTTCGCCACCA
GTTG5'
P2

^{3'}GTCCAACAACGACCATGGTACATGATGGACTGGGACCTTCGATTTCGCCACCA
P4

^{3'}AAAAGAACTGTACGAAGCTAAAGTTGGGTTAAACCGTGGGAAACCTCAAAGAAGTGC3'
K K K L V E A K V W V K P W E T F K E L Q E F K P V G D A
AAAAGAACTGTACGAAGCTAAAGTTGGGTTAAACCGTGGGAAACCTCAAAGAAGTGCAGGAATTCAAACCGTTGGTGACGCT
TTTTCTTTGACATGCTCGATTCAAACCCAAATTGGCACCCCTTTGAGAGTTCTTGACGTCCTTAAGTTTGGCCAAACCACTGCGA
TTTTCTTTGAC5'
P4

^{3'}GTTCTTGACGTCCTTAAGTTTGGCCAAACCACTGCGACCGCGCGCTTCATGTCGG
P6

^{3'}GTCTTGACGTCCTTAAGTTTGGCCAAACCACTGCGACCGCGCGCTTCATGTCGG
P6

1A
Fg²





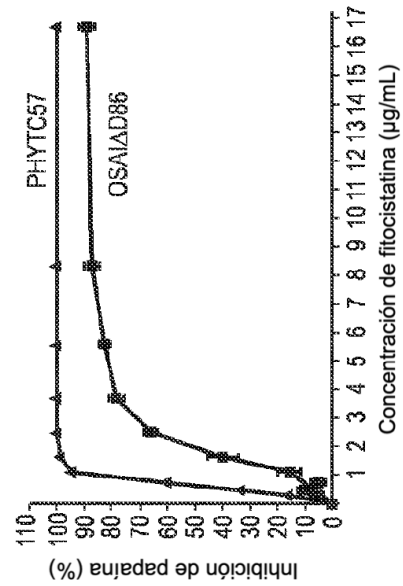


Fig. 2

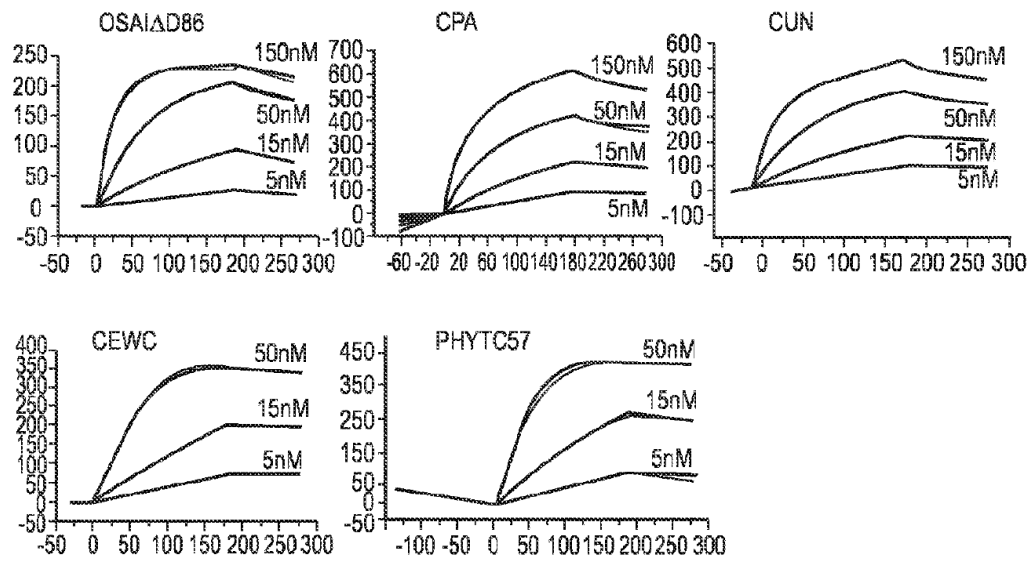


Fig. 3

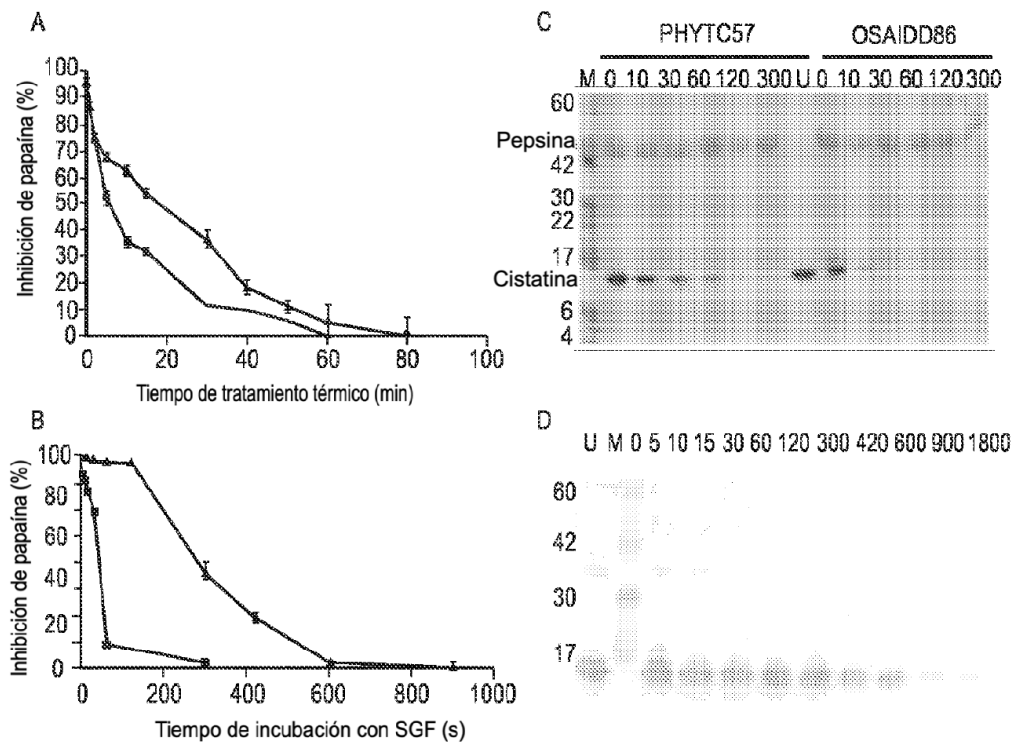


Fig. 4

A

OSAIDD86 MSSDGGPVLG GVEPV.GNEN DLHLVDLARF AVTEHNKKAN SLLEFEKLVS
 PHYTC57 . . . ,MAALLG GVRAPGNEN SLEIEELARF AVDEHNKKEN ALLEFVRVVK
 OSAIDD86 VKQQVVAGTL YYFTIEVKEG DAKKLYEAKV WEKPMM.FKE LQEFKPVASANA
 PHYTC57 AKEQVVAGTM YYLTLEAKDG GKKLYEAKV WKPWETFKE LQEFKPVGDA

B

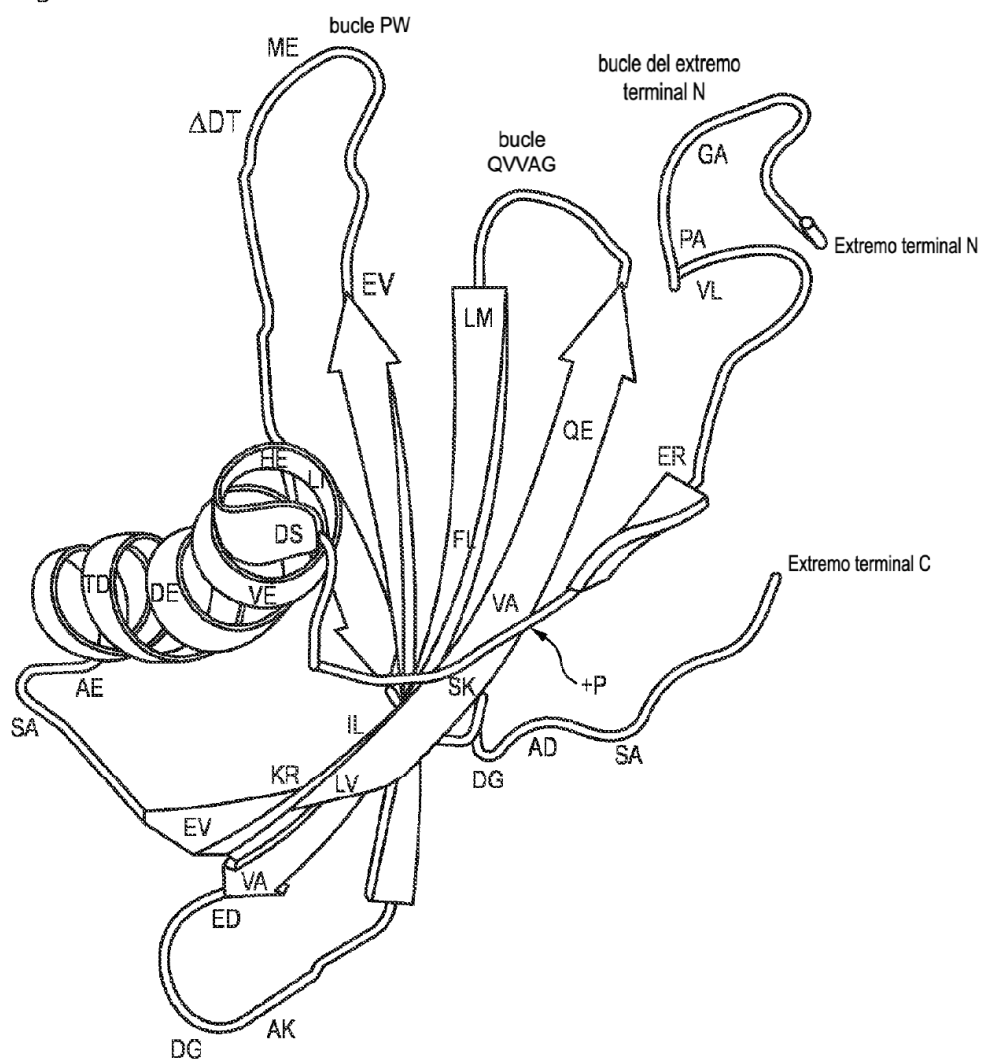
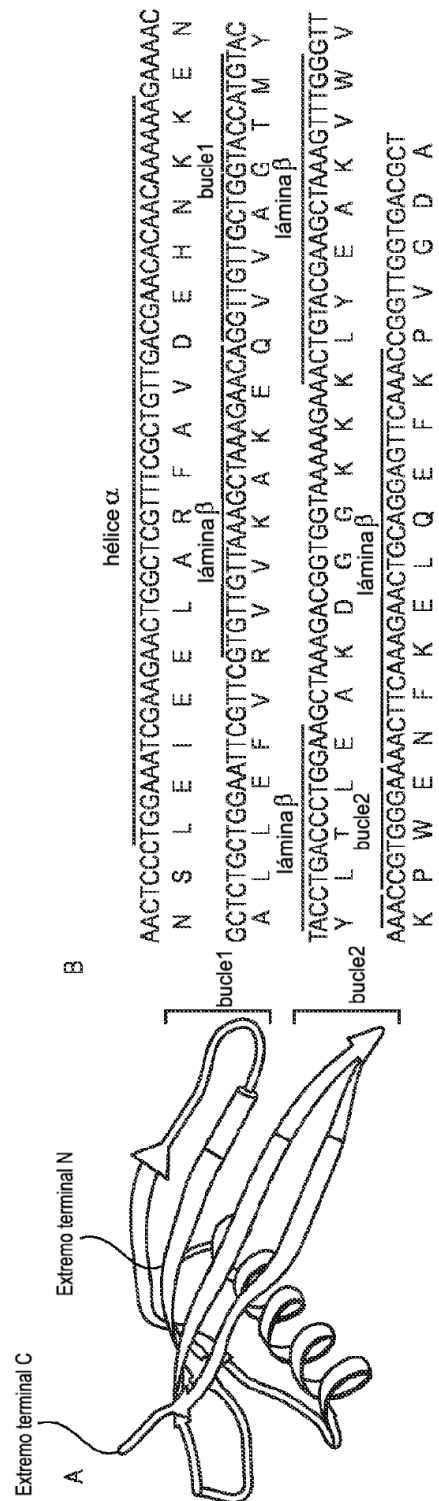


Fig. 5



১৩

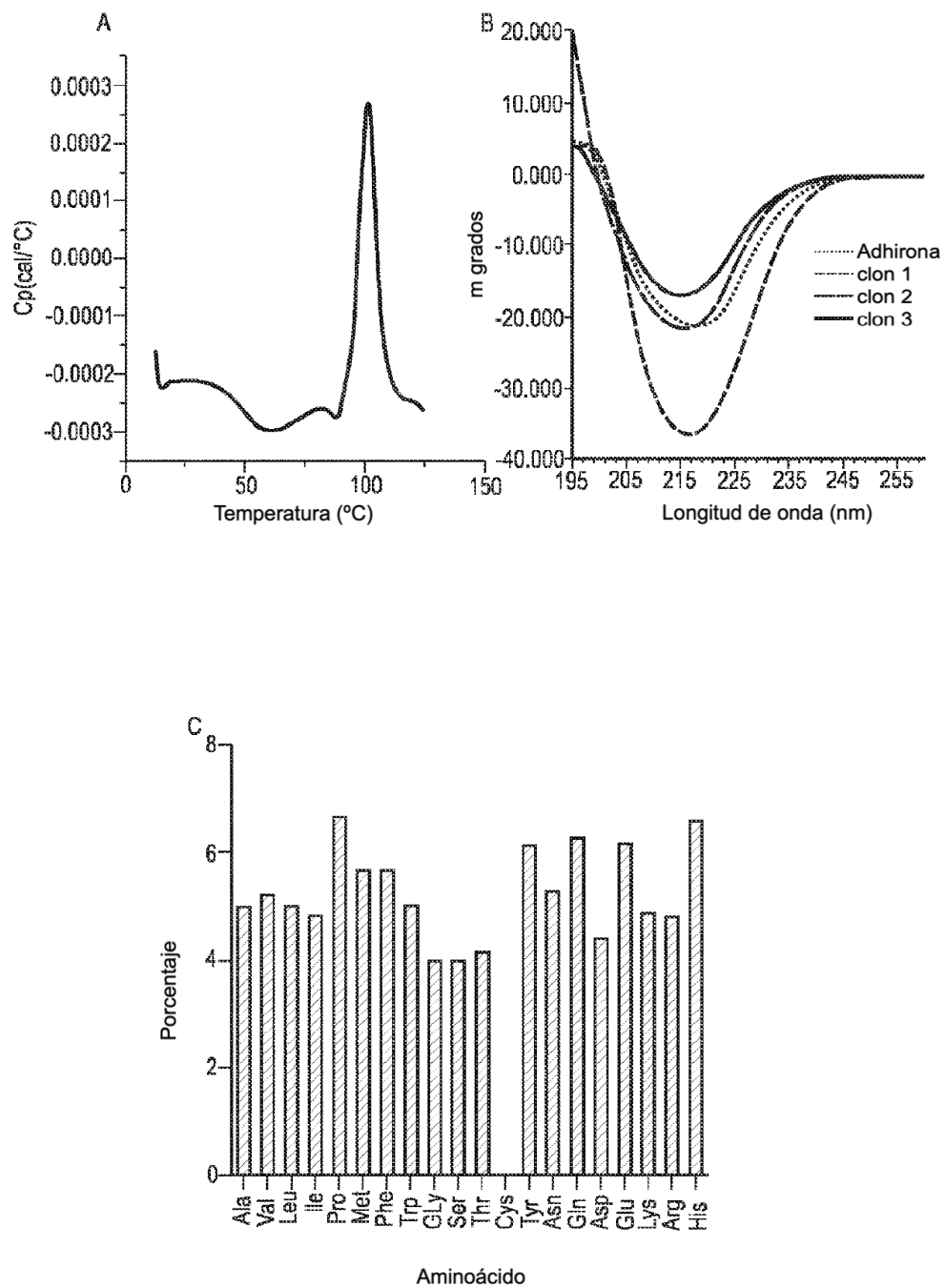


Fig. 7

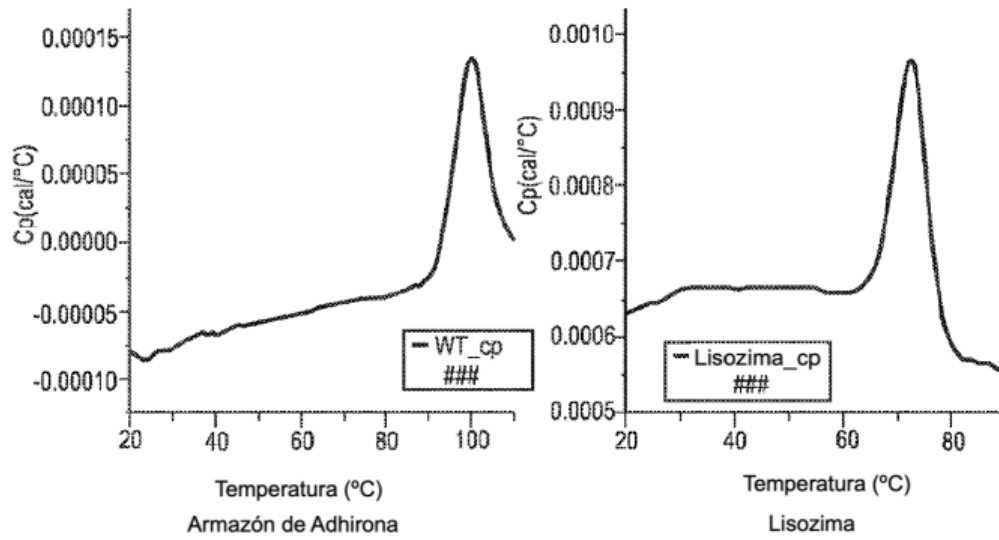


Fig. 8a

Fig. 8b

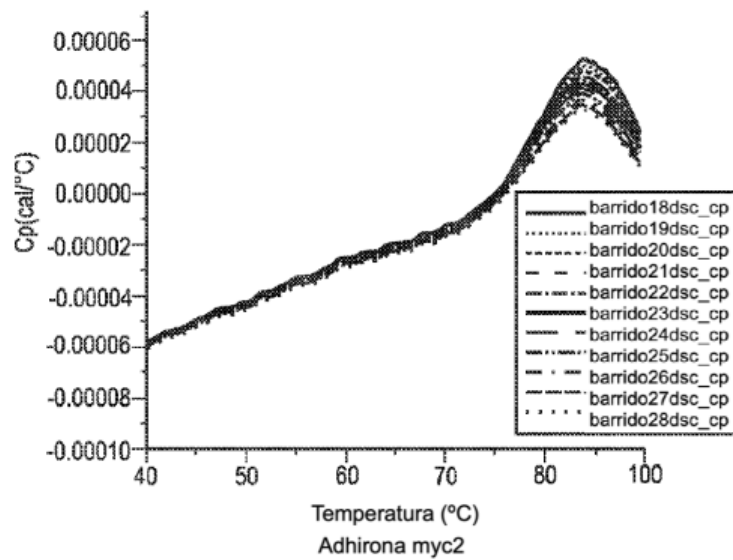


Fig. 8c

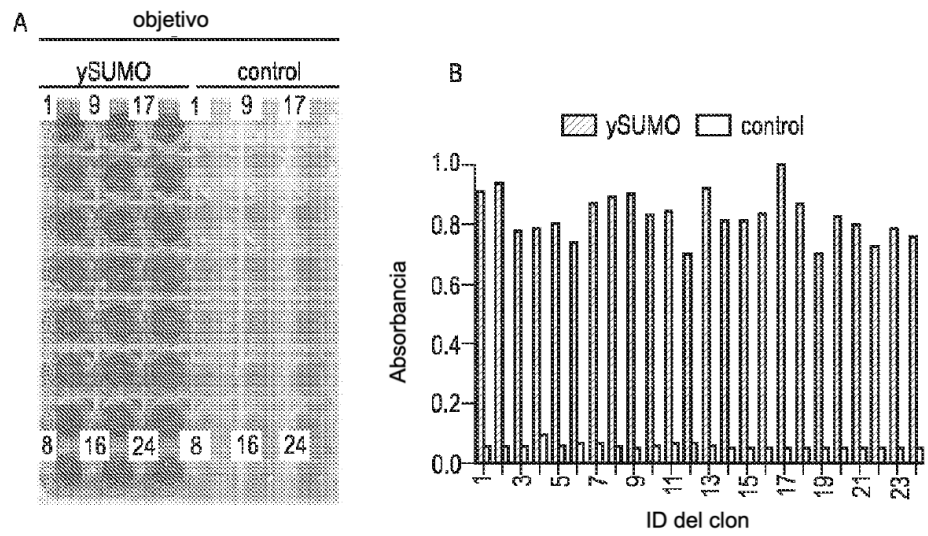


Fig. 9

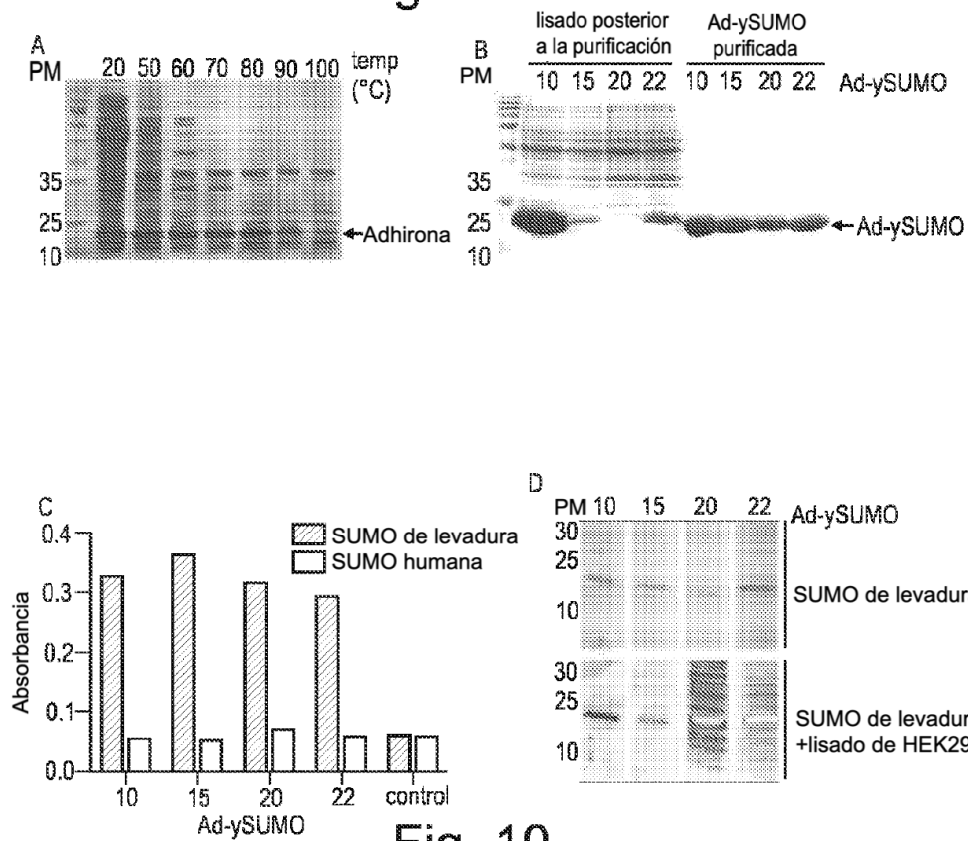


Fig. 10

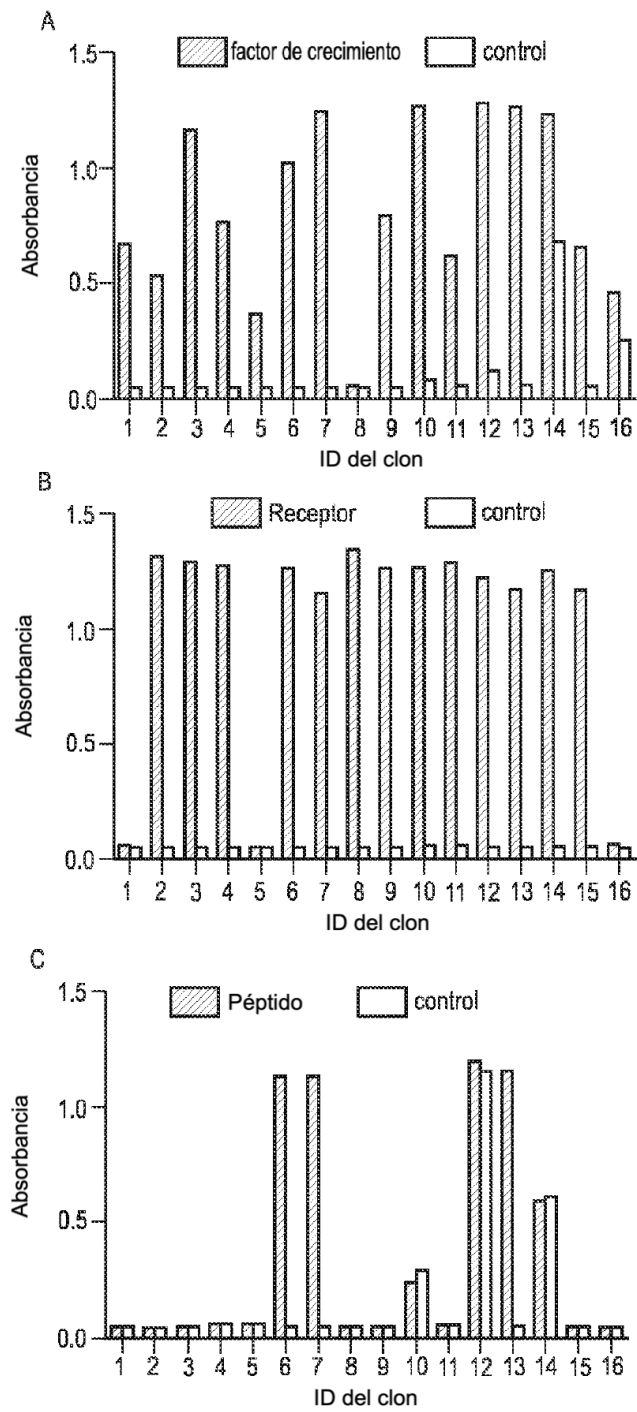


Fig. 11

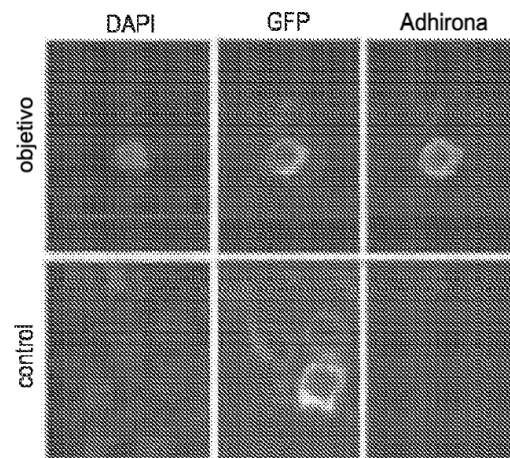


Fig. 12

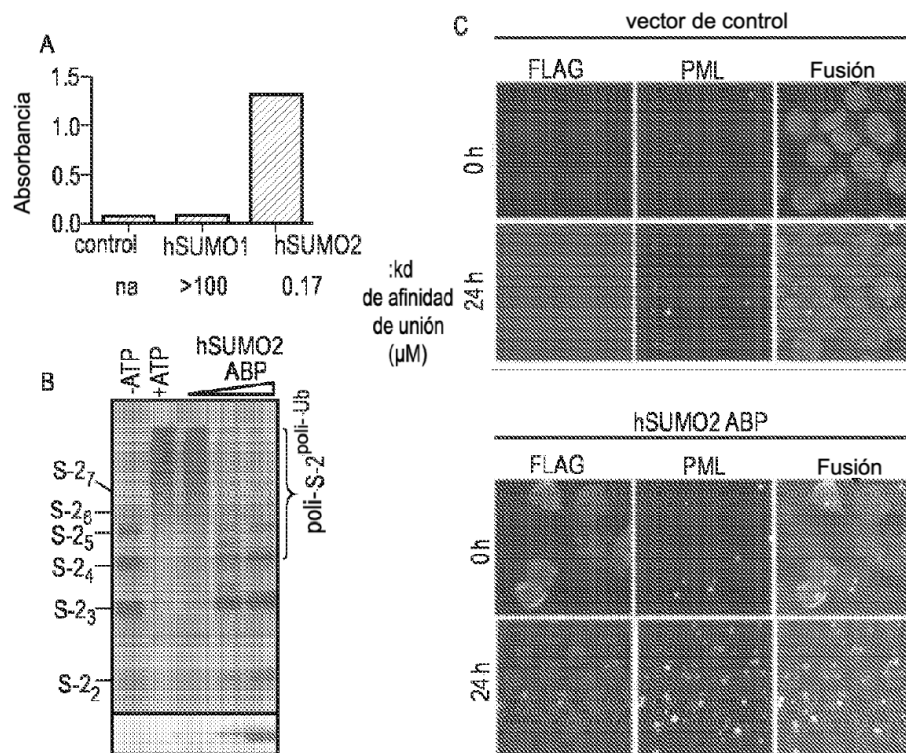


Fig. 13

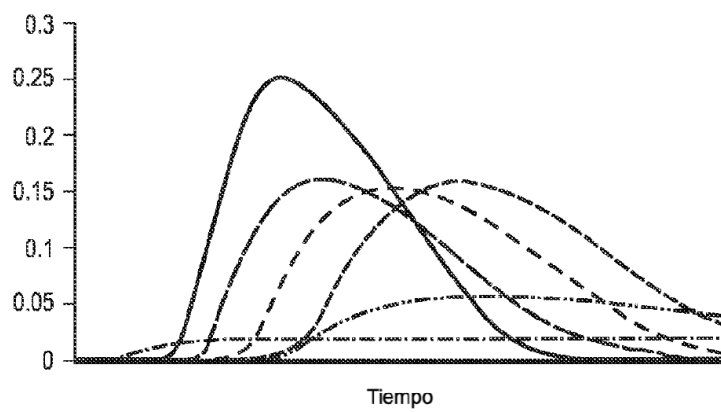


Fig. 14

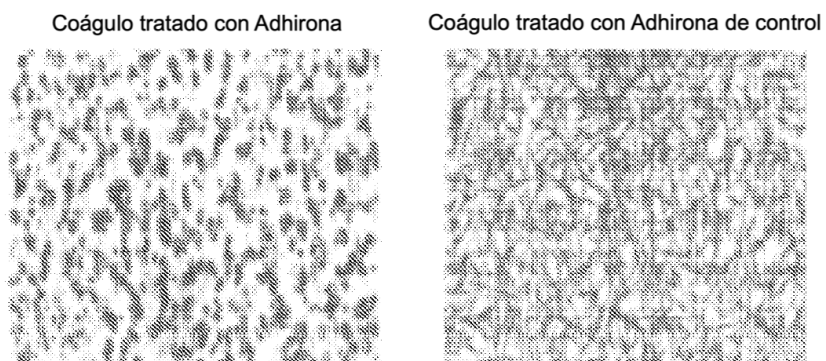


Fig. 15

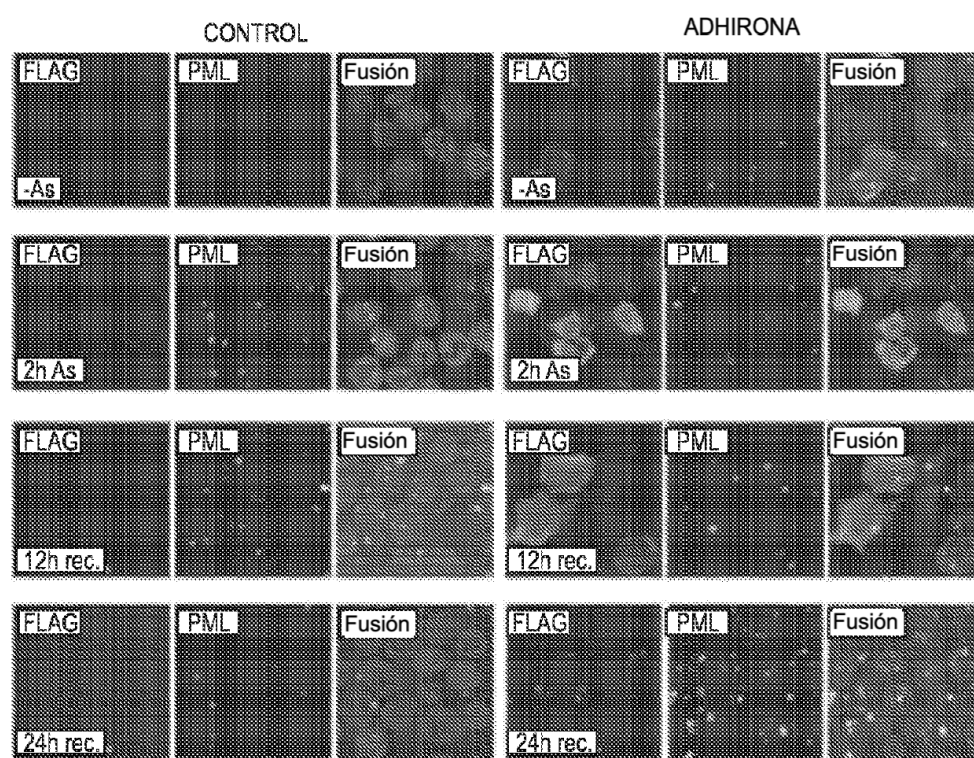


Fig. 16

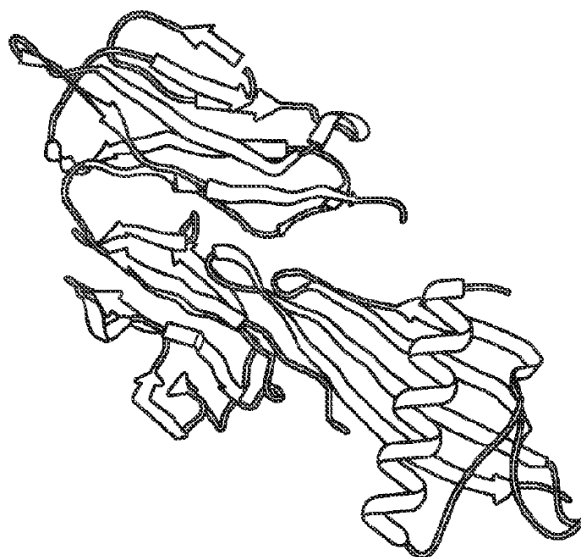


Fig. 17A



Fig. 17B

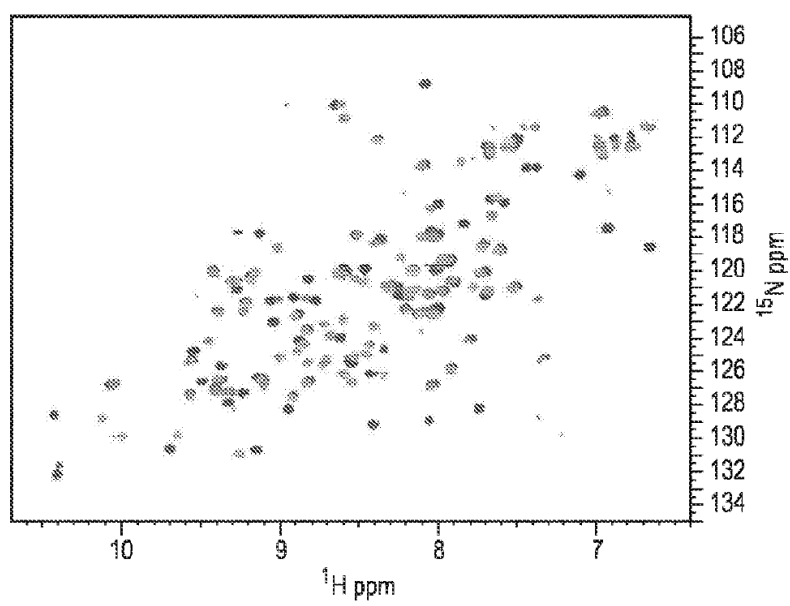


Fig. 18A

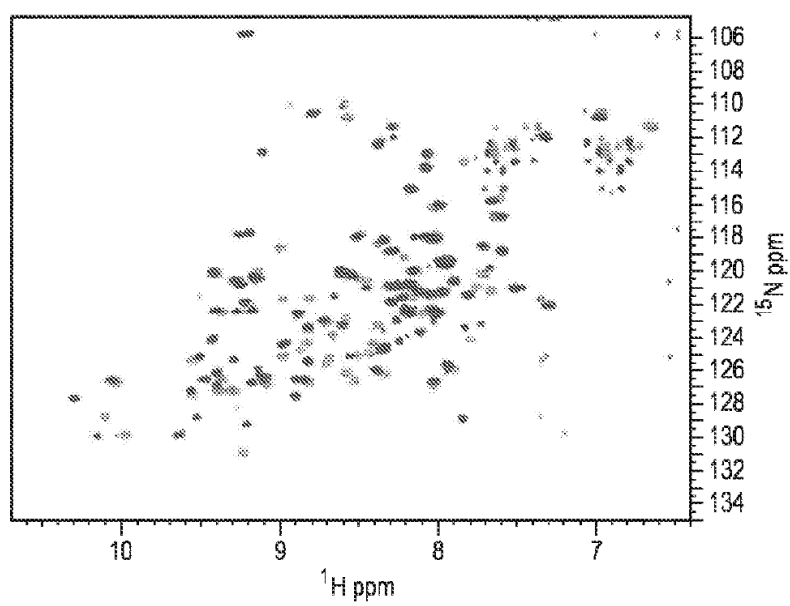


Fig. 18B

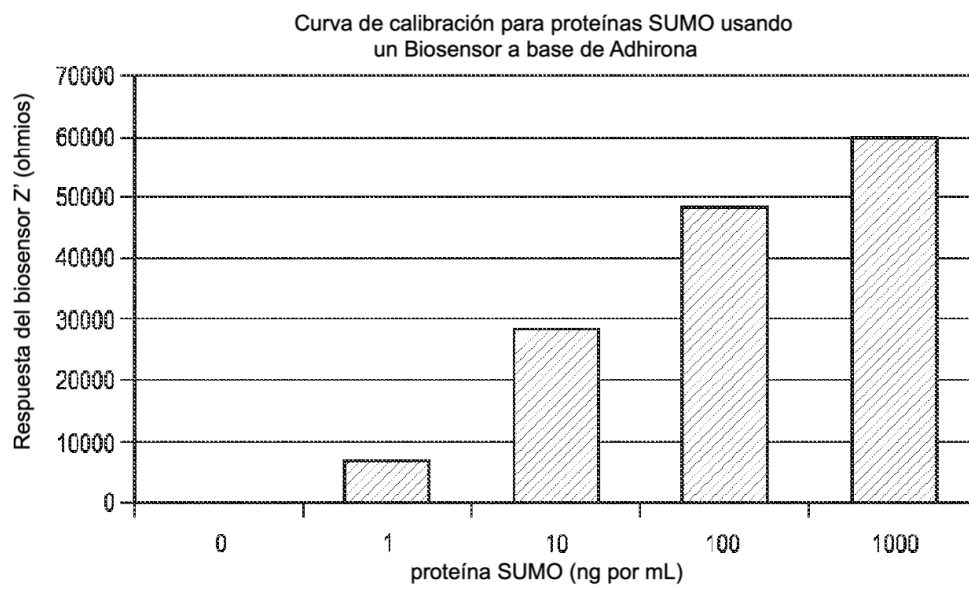


Fig. 19

Aglutinantes de Adhirona LOX1

Bucle 1										Bucle 2																			
70										100										110									
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TA	HL	DP	LM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TH	LG	ML	PP	IT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	TF	GR	EF	LG	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	PI	GE	HP	VN	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	PI	GE	HP	VN	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	PI	GE	HP	VN	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	WD	TQ	VS	PR	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	RW	GV	WE	TM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	SL	PN	AW	AQ	TM	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	EM	GP	ED	IM	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	WG	VR	PM	PV	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			
EQ	T	-	P	-	DT	MY	LY	LT	LE	AK	KG	KK	KK	LY	EA	KK	VW	VK	FW	MI	IS	DL	LI	FN	FK	EL			

Fig. 20

Agglutinantes de Adhirona de la hormona de crecimiento humano

Bucle 1										Bucle 2									
70	KEQAT	DS	EH	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQIT	DS	SP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQWA	RP	PN	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQGH	AR	AD	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDV	KP	NE	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQRS	DQ	PM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDN	D	WL	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQGF	DK	DT	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDD	GS	VI	MP	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQPD	VH	EW	PM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTD	H	SE	MM	PM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQMT	N	TE	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQPE	Q	DR	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQVR	TM	QY	T	DE	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDV	AG	L	P	PL	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQGY	D	GW	L	AR	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTE	P	PH	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTE	D	VR	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTE	D	VR	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTP	NS	AA	I	FW	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQTP	NS	AA	I	FW	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQPR	L	G	-	Y	I	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQAD	M	G	G	M	P	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDD	P	R	Y	G	M	I	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDD	P	R	Y	G	M	I	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQND	F	D	Y	M	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQGN	P	NE	Y	H	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDK	T	Q	T	Y	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQSV	T	G	Q	V	P	A	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDE	T	N	Y	T	I	G	M	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQDN	E	N	W	M	P	M	F	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQPE	H	A	P	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQAD	L	G	Y	L	E	N	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQEE	N	S	A	M	L	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQRY	S	Q	L	M	M	K	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
	KEQ																		

Fig. 21

Agutinantes de Achirona SUMO de levadura

Bucle 1

Bucle 2

70

80

90

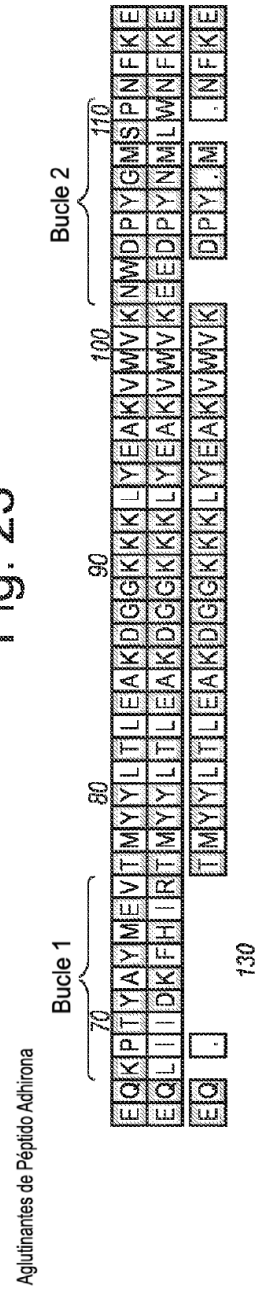
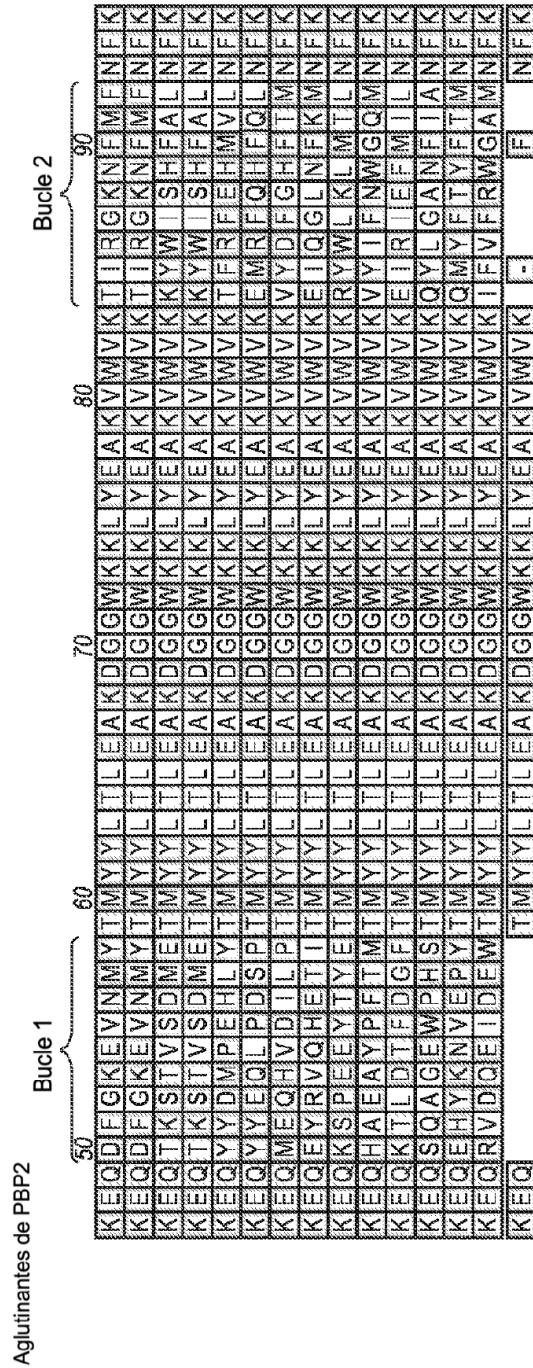
100

110

K	E	Q	W	D	L	T	G	N	V	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	W	D	W	G	E	K	F	W	N	F	K	E		
K	E	Q	I	D	L	T	N	S	F	A	S	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	D	I	N	Q	Y	W	M	S	M	N	F	K	E
K	E	Q	I	M	L	M	M	V	S	P	M	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	I	Q	Q	N	P	S	M	A	N	F	K	E
K	E	Q	I	D	L	T	K	S	L	N	Y	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	L	T	N	E	I	Q	K	M	N	F	K	E
K	E	Q	I	D	L	T	K	S	L	N	Y	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	L	T	N	E	I	Q	K	M	N	F	K	E
K	E	Q	I	D	L	T	F	W	Q	D	K	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	E	I	K	S	M	S	N	F	K	E		
K	E	Q	W	V	D	M	D	Y	Y	W	K	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	M	D	E	I	W	A	E	Y	A	N	F	K	E
K	E	Q	I	D	L	T	Q	T	E	I	V	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	E	P	G	I	I	P	I	V	M	N	F	K	E
K	E	Q	I	D	L	T	D	V	W	I	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	L	M	T	Q	T	N	S	M	N	F	K	E
K	E	Q	I	L	K	E	N	D	A	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	I	M	D	G	L	N	K	Y	N	F	K	E	
K	E	Q	W	I	L	N	T	Q	F	I	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	G	L	E	G	P	D	K	W	V	N	F	K	E	
K	E	Q	W	Y	E	K	S	E	N	W	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	K	D	Y	G	F	T	L	V	P	N	F	K	E
K	E	Q	W	D	L	T	T	P	I	N	I	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	Y	E	D	Y	Q	T	P	M	Y	N	F	K	E
K	E	Q	W	F	D	E	Y	D	W	I	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	D	Y	A	T	D	L	Y	W	N	F	K	E		
K	E	Q	I	D	L	T	Q	P	K	D	S	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	Y	E	D	E	Y	W	K	M	N	F	K	E	
K	E	Q	I	D	L	T	Q	S	F	D	M	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	I	D	S	N	F	T	G	T	N	F	K	E
K	E	Q	W	Y	L	D	V	M	D	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	H	D	R	Y	R	Q	A	E	N	F	K	E		
K	E	Q	W	I	D	K	G	Q	Y	W	D	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	I	H	N	G	Y	T	I	M	D	N	F	K	E
K	E	Q	W	S	E	A	D	N	D	W	M	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	L	D	L	E	T	W	Q	W	F	N	F	K	E
K	E	Q	I	D	L	T	G	Q	W	L	F	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	L	W	Q	Y	D	A	Q	Y	N	F	K	E
K	E	Q	I	D	L	T	Q	S	F	D	M	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	S	R	R	N	Y	Q	T	M	N	F	K	E
K	E	Q	I	D	L	T	Q	S	F	D	M	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	I	D	S	N	F	T	G	T	N	F	K	E
K	E	Q	I	D	L	T	Q	P	K	D	S	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	P	H	D	E	L	N	W	N	M	N	F	K	E
K	E	Q	W	E	D	F	Q	T	K	W	E	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K	D	V	G	Q	L	L	S	G	I	N	F	K	E

K	E	Q	I	D	L	T	T	T	M	Y	Y	L	T	L	E	A	K	D	G	G	W	K	K	L	Y	E	A	K	V	W	V	K																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Fig. 22



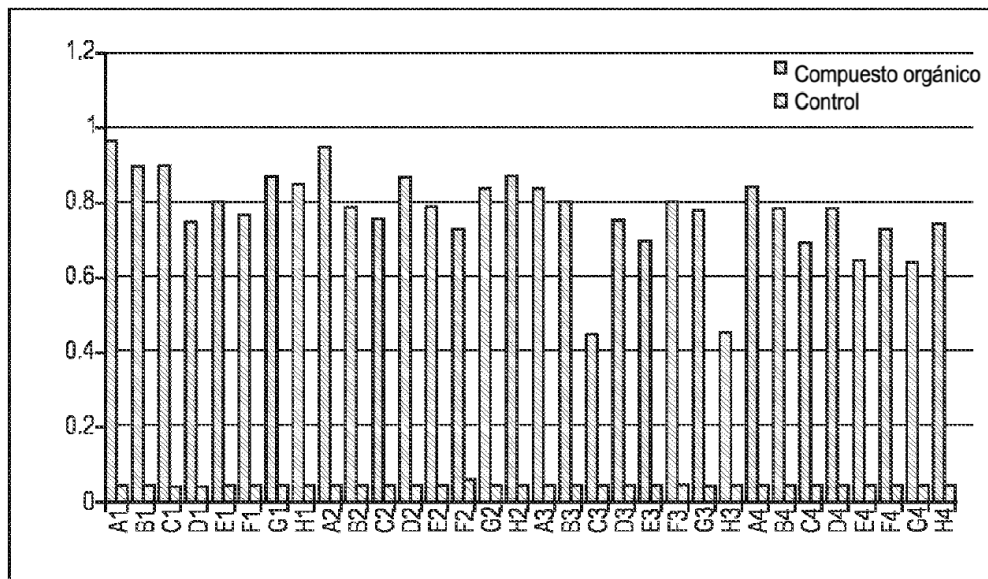


Fig. 25

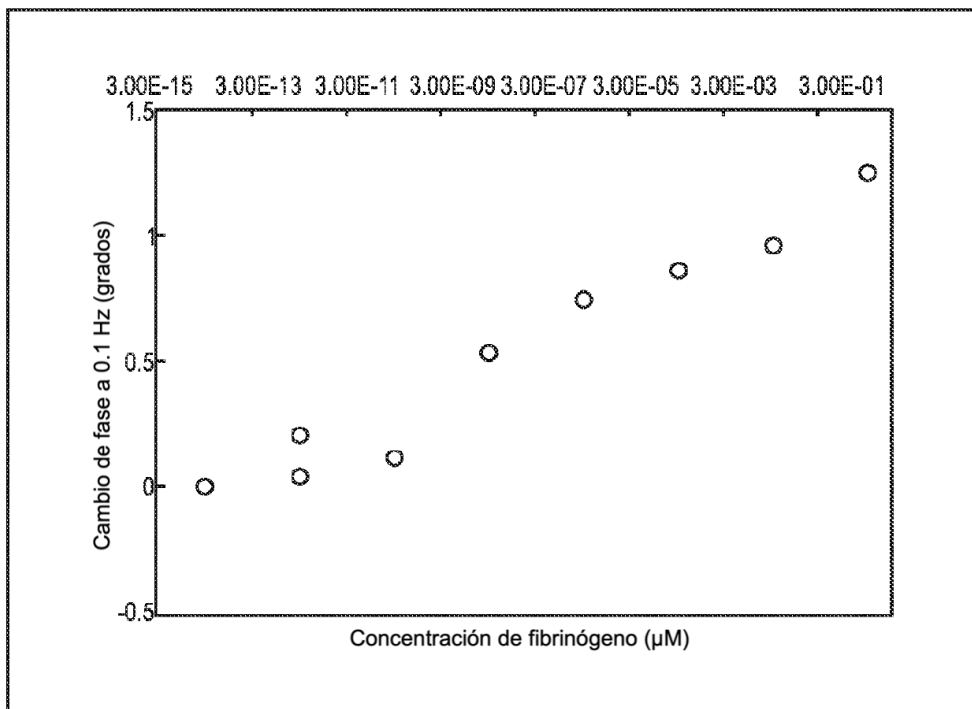


Fig. 26

[illegible]

Fig. 27

Aglutinantes STAT3 Adhirona									
50	60	70	80	90					
AKETDEFNQPNWNTMYLTL	EAKDGGKKKLYEAKVMVKTIIYINSFIMNFKE								
AKETDQQQKPNWNTMYLTL	EAKDGGKKKLYEAKVMVKKLFVNRNPNWNFKE								
AKEQVMEVPYFWQTMYYLT	EAKDGGKKKLYEAKVMVKKWILVANEQANFKE								
AKEQNDQDHYIHETMYLTL	EAKDGGKKKLYEAKVMVKKLILTNISFNFKE								
AKEQPVPIKAWETMYLTL	EAKDGGKKKLYEAKVMVKKIIYYNKNKINFKE								
AKEQSKQVNMKAHTMYLTL	EAKDGGKKKLYEAKVMVKSILMAHETFNFKE								
AKEQSKDPWTRLGTMYLTL	EAKDGGKKKLYEAKVMVKKEDPWIMFLANFKE								
AKEQWLRITYQDWFTMYLTL	EAKDGGKKKLYEAKVMVKKDTKSPDYGVNFKE								
AKEQMTKPKPGHSTMYLTL	EAKDGGKKKLYEAKVMVKKPEADIAFLWNFKE								
AKEQIIEVASPNWNTMYLTL	EAKDGGKKKLYEAKVMVKKIRWDGDWYNFKE								
AKEQSEGHTFIQETMYLTL	EAKDGGKKKLYEAKVMVKKSLMLTSGAFNFKE								
AKEQTLVITEDPQYTMYYLT	EAKDGGKKKLYEAKVMVKKWDMSIMFMWNFKE								
AKEQWLLMPMSPWSTMYLTL	EAKDGGKKKLYEAKVMVKSIIYRKQDVNFKE								
AKEQEPHMGIPWMTMYLTL	EAKDGGKKKLYEAKVMVKEIYMETHTIVNFKE								
AKEQWPFNTQTWNTMYLTL	EAKDGGKKKLYEAKVMVKKYFKFYGNMNFKE								
AKEQFPLDAMPWVTMYLTL	EAKDGGKKKLYEAKVMVKSIIKVHHDITINFKE								
AKEQWPWQDVVFSTMYLTL	EAKDGGKKKLYEAKVMVKKLAIDGVWMNFKE								
AKEQPESRTQAWQTMYYLT	EAKDGGKKKLYEAKVMVKEIRIKMHWYNFKE								
AKEQLVHDDVPWQTMYYLT	EAKDGGKKKLYEAKVMVKKALNYNGMWDNFKE								
AKEQWVLFIPYRMTMYLTL	EAKDGGKKKLYEAKVMVKKHQGIENNGNFKE								
AKEQLHESVTFEPTMYLTL	EAKDGGKKKLYEAKVMVKKQFMINHAWDNFKE								
AKEQSHASSRPWATMYLTL	EAKDGGKKKLYEAKVMVKKIIAHGFPAINFKE								
AKEQ	W	TMYYLTLEAKDGGKKKLYEAKVMVKK							NFKE

Fig. 28