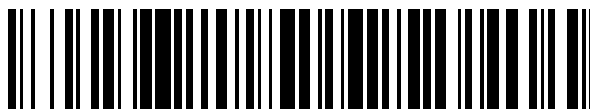


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 682 254**

51 Int. Cl.:

<b>C40B 40/10</b>	(2006.01)	<b>C07K 16/46</b>	(2006.01)
<b>C40B 40/08</b>	(2006.01)	<b>G01N 33/53</b>	(2006.01)
<b>C40B 40/02</b>	(2006.01)	<b>G01N 33/68</b>	(2006.01)
<b>C40B 30/04</b>	(2006.01)		
<b>C40B 50/06</b>	(2006.01)		
<b>C07K 16/44</b>	(2006.01)		
<b>C12N 15/13</b>	(2006.01)		
<b>A61K 39/395</b>	(2006.01)		
<b>A61K 31/7088</b>	(2006.01)		
<b>C07K 16/00</b>	(2006.01)		

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **17.08.2012 PCT/AU2012/000970**
- 87 Fecha y número de publicación internacional: **21.02.2013 WO13023251**
- 96 Fecha de presentación y número de la solicitud europea: **17.08.2012 E 12824386 (2)**
- 97 Fecha y número de publicación de la concesión europea: **02.05.2018 EP 2744931**

54 Título: **Polipéptidos solubles**

30 Prioridad:

**18.08.2011 AU 2011903298**

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

**19.09.2018**

73 Titular/es:

**AFFINITY BIOSCIENCES PTY LTD (100.0%)  
1 Dalmore Drive  
Scoresby, VIC 3179, AU**

72 Inventor/es:

**BEASLEY, MATTHEW DAVID;  
NIVEN, KEITH PHILIP y  
KIEFEL, BEN ROSS**

74 Agente/Representante:

**INGENIAS CREACIONES, SIGNOS E  
INVENCIONES, SLP**

ES 2 682 254 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

Polipéptidos solubles

## 5 Campo de la invención

La presente divulgación se refiere de manera general a polipéptidos, tales como moléculas de anticuerpos, que demuestran alta estabilidad y solubilidad. En particular, la presente divulgación se refiere a polipéptidos que comprenden dominios  $V_L$  y  $V_H$  emparejados que demuestran una expresión soluble y un plegamiento en un ambiente reductor o intracelular. La presente divulgación también se refiere a polinucleótidos que codifican tales polipéptidos, a bibliotecas de tales polipéptidos o polinucleótidos, y a métodos de uso de tales polipéptidos en investigación, en aplicaciones de diagnóstico y terapéuticas. Por ejemplo, se pueden usar los polipéptidos en métodos de selección sistemática para identificar un polipéptido que se une a una molécula diana particular.

## 15 Antecedentes de la invención

El repertorio de anticuerpos de vertebrados se formó mediante la duplicación y la diversificación de los genes ancestrales de un heterodímero de dos pliegues de inmunoglobulina (Ig). La diversidad generada por el sistema inmunológico no solo se basa en las familias de genes germinales de los genes de Ig, sino también de la recombinación *in vivo* de exones del subdominio durante el desarrollo de los cayados para formar numerosos linajes únicos con diversidad adicional en los límites del exón que se da en los bucles expuestos en la superficie de la proteína Ig. Este proceso de recombinación se denomina recombinación V(D)J, llamado así después de que los dos exones ligeros variables ( $V_L$ ) y los tres exones pesados variables ( $V_H$ ) se recombinen para formar los dominios de unión a antígeno de N-terminal de la cadena ligera y la cadena pesada del anticuerpo, respectivamente. Sin embargo, dado que los genes duplicados divergen de sus pares ancestrales, el efecto acumulado de las mutaciones ha dado como resultado un ajuste entre superficies de contacto menos perfecto entre las unidades heterodiméricas de los dominios variables. La presión de selección no se aplica a ningún gen, sino a la familia completa. Por lo tanto, la máxima diversidad, que es algo bueno para el sistema inmunológico, puede dar como resultado estabilidades en el plegamiento menos ideales para los miembros de la familia individual. Además, los propios dominios de unión pueden tener diferentes estabilidades de plegamiento. El requisito de formar un heterodímero funcional a partir de numerosas subunidades divergentes se compensa por la presencia de enlaces disulfuro conservados entre las láminas beta de los dominios. Sin embargo, la superficie de contacto aún puede no ser un ajuste estable, que requiere un punto de control de plegamiento en el RE.

Como resultado del enfoque de "consenso" para un ajuste de proteína aplicado por los dominios variables del anticuerpo, algunos emparejamientos tienen una baja estabilidad de plegamiento y son propensos tanto a una mala expresión en hospedadores bacterianos/mamíferos, como propensos a agregarse. Además, en casi todos los casos, hay un requisito total para que se formen enlaces disulfuro interlaminares en los dominios  $V_L$  y  $V_H$ . Se requiere que para la expresión de bibliotecas de anticuerpos en un hospedador bacteriano tal como *E. coli*, se exprese el anticuerpo en el periplasma de la célula, un espacio oxidante que tiene chaperonas disulfuro, y a menudo como una fusión entre los dominios  $V_L$  y  $V_H$  (anticuerpo monocatenario; scFv). Sin embargo, la exportación al periplasma requiere la excreción a través de la membrana interna, que se satura a los niveles deseados para la alta expresión del anticuerpo, dando como resultado rendimientos mucho más bajos que la expresión citoplásmica.

Además de la ventaja de una producción más barata de anticuerpos scFv en el citoplasma de *E. coli*, un almacén de anticuerpo que es competente para plegarse en un ambiente reductor también se podría usar como reactivo de afinidad en el citoplasma de los mamíferos. Esto permitiría la extensión de los usos de los anticuerpos como reactivos científicos en el citoplasma o en el núcleo para la obtención de imágenes o el bloqueo de funciones proteicas y, de manera similar, en agentes terapéuticos y de diagnóstico.

Como casi todos los anticuerpos de mamíferos son insolubles en el citoplasma, los grupos han investigado combinaciones de genes poco comunes que se pliegan para formar un heterodímero estable para su uso como un almacén para la construcción de diversidad adicional. La estrategia adoptada para encontrar anticuerpos solubles en citoplasma es la observación casual de que un clon de anticuerpo se expresa de manera estable en el citoplasma (Tavladoraki et al., 1999; Vaccaro et al., 2006) lo que puede ser la base de un almacén de anticuerpo intracelular ("intracuerpo") o, como alternativa, se puede tomar una estrategia evolutiva para llevar un gen de scFv hacia la estabilidad, bien *in vivo* (Martineau et al., 1998; Visintin et al., 1999; Auf der Maur et al., 2002; Fisher and DeLis, 2009) o *in vitro* (Contreras-Martinez y Delisa, 2007; Jermtus L., et al. 2001). Además, los anticuerpos de dominio único, en donde solo un único dominio variable no emparejado se une al antígeno diana, han demostrado ser solubles y estables en el citoplasma. Se han descrito dos anticuerpos de dominio único de camélido que se pliegan y son solubles cuando se expresan en el citoplasma (Kirchhofer AL., et al, 2010; Saerens et al., 2005).

Otra estrategia para la producción de anticuerpos intracelulares en el citosol bacteriano es el uso de mutantes de *E. coli* que tienen mutaciones que cambian el estado redox de proteínas en el citoplasma de reductor a oxidante. Esto produce scFv que se pliegan y se oxidan parcial y/o totalmente en el citoplasma de *E. coli* (He et al, 1995; Jurado P., et al., 2002).

Dos grupos que usaron el sistema de dos híbridos en levadura (Y2H) como una selección sistemática *in vivo* para los scFv que se unen a antígenos de las bibliotecas de scFv compilaron secuencias para sus clones solubles. El primer grupo (Tse et al., 2002) descubrió que el clado VH3 se empareja con los clados VLk 1 y 4. Alineando múltiples scFv solubles compilaron un consenso para los genes de VL y VH solubles que casi coinciden exactamente con el consenso de la familia compilado para la biblioteca Morphosys HuCAL™ para familias VH3 y VLk1 (Knappik et al., 2000). El segundo grupo que usó el Y2H informó en el documento WO 03/097697 que sus scFv solubles eran secuencias lo más estrechamente relacionadas con los miembros de los clados VH3, VH1a o VH1b combinados con las secuencias más estrechamente relacionadas con los miembros de los clados VLk1 o VLλ1 o VLλ3. Sin embargo, su configuración óptima fue VLλ3 emparejado con VH1b. Cabe destacar, sin embargo, que ninguna de las secuencias documentadas fueron coincidencias exactas con la traducción de la secuencia de la línea germinal del gen de inmunoglobulina homólogo más cercano, con múltiples mutaciones a través de la secuencia. Esto fue presumiblemente debido al uso por ambos grupos de bibliotecas de fagos preseleccionados para enriquecer los clones de unión a antígeno antes de la etapa limitante de la transformación de levadura. Sin embargo, esto implica que una, o más, de las mutaciones en cada gen puede estar confiriendo un efecto estabilizante sobre el plegamiento de scFv en el citoplasma.

El documento WO 2010/136698 se refiere a métodos de identificación de pares de clases VH y VL en el repertorio inmunológico humano, determinando los pares de clases VH y VL que son más prevalentes y los que tienen propiedades biofísicas favorables.

El documento WO 2016/028791 se refiere a una biblioteca de presentación de anticuerpos que contiene moléculas de anticuerpo de la línea germinal humana con variación en la CDR3 de VH, y en CDR3 de VL, y en la posición 52 de la CDR2 del VH, para la detección y la selección de moléculas de anticuerpo específicas para antígenos de interés.

El documento WO 2016/028791 desvela anticuerpos y moléculas relacionadas que se unen de manera específica a Reg IV. Tales anticuerpos tienen usos, por ejemplo, en la prevención y el tratamiento de cánceres gastrointestinales, enfermedad inflamatoria del intestino, y diabetes.

El documento WO 2011/075761 se refiere a métodos para seleccionar de manera sistemática un polipéptido para una actividad deseada frente a una molécula diana expresando el polipéptido en una célula bacteriana y permeabilizando la célula.

El documento US 2011/0118149 desvela composiciones y métodos para generar bibliotecas de secuencias de ADN que codifican polipéptidos homólogos, y usos de las bibliotecas para identificar las variantes de polipéptido diversificadas de forma natural.

Hasta la fecha, no se han publicado informes de un anticuerpo intracelular que tiene una identidad exacta con la secuencia de aminoácidos de la línea germinal humana de los correspondientes genes de VL y VH. Tal anticuerpo sería un almacén ventajoso para generar diversificación porque permitiría una alta producción de expresión citoplasmática, proporcionaría una mayor estabilidad en forma oxidada, proporcionaría mayor estabilidad estructural, asegurando una mayor tolerancia de la diversificación de bucles, y comprendería una secuencia completamente natural que da como resultado un mejor rechazo del paciente a partir de la producción de un anticuerpo completo.

Los presentes inventores informan en el presente documento de la aplicación de un método de presentación de proteínas descrito anteriormente en el documento WO 2011/075761 para la selección sistemática de una biblioteca de scFv humano y el aislamiento de genes de scFv solubles que tienen regiones marco idénticas a la secuencia de la línea germinal humana. Además, los presentes inventores demuestran la notable termoestabilidad y tolerancia de injerto de CDR3 en el almacén de scFv.

## Sumario de la invención

En un aspecto, la divulgación proporciona una biblioteca de polipéptidos que comprende una pluralidad de diferentes polipéptidos, que comprenden:

- i) una región variable de cadena pesada de anticuerpo (V<sub>H</sub>) que comprende una región de almacén que es al menos el 95 % idéntica a la región de almacén de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y
- i) una región variable de cadena ligera de anticuerpo (V<sub>L</sub>) que comprende una región de almacén que es al menos el 95 % idéntica a la región de almacén de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12);

en donde la V<sub>H</sub> y la V<sub>L</sub> son capaces de formar un sitio de unión a antígeno; y

en donde al menos dos de los polipéptidos difieren entre sí en la secuencia de aminoácidos presente en una o más regiones determinantes de complementariedad (CDR, del inglés *complementary determining regions*) en las

regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro. Preferentemente, la secuencia de aminoácidos en una o más de las CDR de los dominios variables  $V_H$  y/o  $V_L$  es aleatoria o semialeatoria o deriva de un anticuerpo humano.

En otro aspecto, la divulgación proporciona un procedimiento de construir una biblioteca de polipéptidos que es soluble en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro, comprendiendo el método la preparación de una pluralidad de diferentes polipéptidos, que comprenden:

- i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y
- i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

en donde al menos dos de los polipéptidos difieren entre sí en la secuencia de aminoácidos presente en una o más CDR en las regiones variables  $V_H$  y/o  $V_L$ .

En otro aspecto, la divulgación proporciona una biblioteca de polinucleótidos que comprende una pluralidad de diferentes polinucleótidos,

en donde cada polinucleótido codifica un polipéptido que comprende:

- i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y
- i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

en donde al menos dos de los polinucleótidos difieren entre sí codificando polipéptidos que comprenden una o más CDR diferentes en las regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

Preferentemente, los polinucleótidos codifican una secuencia de aminoácidos en una o más de las CDR de los dominios variables  $V_H$  y/o  $V_L$  que es aleatoria o semialeatoria o deriva de un anticuerpo humano.

En otro aspecto, la divulgación proporciona un método de construcción de una biblioteca de polinucleótidos, comprendiendo el método la preparación de una pluralidad de diferentes polinucleótidos que codifican un polipéptido, que comprende:

- i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y
- i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

en donde al menos dos de los polinucleótidos difieren entre sí codificando polipéptidos que comprenden una o más CDR diferentes en las regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

En otro aspecto, la divulgación proporciona un polipéptido aislado y/o recombinante que comprende:

- i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y
- i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al

menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12);

5 en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno y en donde el polipéptido es soluble en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

10 La  $V_L$  preferentemente comprende una región de armazón que es al menos el 90 % idéntica a la región de armazón de IGLV3-1 tal como se establece en la SEQ ID NO: 6.

Preferentemente, el polipéptido es un fragmento variable (Fv), tal como un fragmento Fab, un fragmento Fab', un fragmento F(ab'), un scFv, un diacuerpo, un triacuerpo, un tetracuerpo o un complejo polipeptídico de mayor orden.  
15 Más preferentemente, el polipéptido es un scFv y la  $V_H$  y la  $V_L$  están enlazadas entre sí a través de un enlazador peptídico.

Preferentemente, la región de armazón de las regiones variables  $V_H$  y  $V_L$  en el polipéptido desvelado en el presente documento es al menos el 95 %, el 96 %, el 97 %, el 98 % o el 99 % idéntica a la región del armazón de cualquiera de las secuencias dadas.  
20

El polipéptido desvelado en el presente documento es preferentemente en condiciones reductoras. Además, el polipéptido desvelado en el presente documento es preferentemente soluble y capaz de formar de manera estable un sitio de unión a antígeno cuando se produce en condiciones reductoras que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.  
25

En otra realización preferente, el polipéptido desvelado en el presente documento se conjuga con un compuesto. El compuesto se puede seleccionar del grupo que consiste en un radioisótopo, un marcador detectable, un compuesto terapéutico, un coloide, una toxina, un ácido nucleico, un péptido, una proteína, un compuesto que aumenta la semivida del polipéptido en un sujeto, y mezclas de los mismos.  
30

En otro aspecto, la divulgación proporciona un polinucleótido aislado y/o exógeno que codifica el polipéptido desvelado en el presente documento, o una región variable de cadena pesada o ligera del mismo.

35 En otro aspecto, la presente divulgación proporciona un método de selección sistemática que se une a una molécula diana tal como se describe anteriormente en donde el polinucleótido se expresa en el citoplasma y/o en el periplasma de una célula hospedadora, y en donde la célula hospedadora es una célula bacteriana, una célula de levadura o una célula de mamífero. Preferentemente, la célula hospedadora es una célula bacteriana y el método comprende:

- 40
- a) cultivar una célula bacteriana que comprende un polinucleótido que codifica un polipéptido de una cualquiera de las reivindicaciones 5-8, de manera que se produce el polipéptido,
  - b) permeabilizar la célula bacteriana, en donde el polinucleótido y el polipéptido se mantiene dentro de la célula bacteriana permeabilizada,
  - 45 c) poner en contacto la célula bacteriana permeabilizada con la molécula diana de manera que difunda a la célula bacteriana permeabilizada, y
  - d) determinar si el polipéptido de una cualquiera de las reivindicaciones 5-8 se une a la molécula diana.

En un aspecto adicional, la divulgación proporciona una biblioteca de células hospedadoras que comprende una pluralidad de células hospedadoras que comprenden un polipéptido desvelado en el presente documento, en donde al menos una célula hospedadora comprende un polipéptido que difiere de un polipéptido presente en otra célula hospedadora en la biblioteca en la secuencia de aminoácidos presente en una o más CDR en los dominios variables  $V_H$  y/o  $V_L$ . Una o más células hospedadoras en la biblioteca de células hospedadoras pueden comprender uno o más polinucleótidos que codifican el polipéptido desvelado en el presente documento. Por ejemplo, una célula hospedadora en la biblioteca de células hospedadoras puede contener un polinucleótido que codifica la  $V_H$  y otro polinucleótido que codifica la  $V_L$ .  
50  
55

En otro aspecto, la divulgación proporciona una composición que comprende el polinucleótido desvelado en el presente documento, y un vehículo farmacéuticamente aceptable.  
60

Como será evidente, las cualidades y características preferidas de un aspecto de la divulgación son aplicables a cualquier otros aspectos de la divulgación, haciendo los cambios necesarios.

A lo largo de la presente memoria descriptiva, la palabra "comprender", o variaciones tales como "comprende", o "comprendiendo", se entenderá que implican la inclusión de un elemento, número entero o etapa indicados, o grupo de elementos, números enteros o etapas, pero no la exclusión de cualquier otro elemento, número entero o etapa, o  
65

grupo de elementos, números enteros o etapas.

La divulgación de describe en lo sucesivo en el presente documento mediante los siguientes Ejemplos no limitantes y con referencia a las figuras adyacentes.

5

### Breve descripción de los dibujos adjuntos

La **FIGURA 1** demuestra la apariencia típica de un clon de scFv soluble, bien expresado (1A y recuadro), junto con un clon de scFv bien expresado pero insoluble (1B y recuadro).

10 La **FIGURA 2** muestra un alineamiento múltiple de los clones solubles seleccionados que tienen alta similitud o identidad total, con los genes de VL IGLV3-1, IGLV3-21 y IGLV6-57.

La **FIGURA 3** demuestra el comportamiento de dos clones, un IGLV3-1 y un IGLV3-21, con la expresión a temperaturas en aumento.

15 La **FIGURA 4** demuestra la solubilidad de un clon de IGLV3-1 cuando se expresa en el citosol de *E. coli* a 25 °C. La proteína de fusión scFv::I27::FLAG está por completo en fracción soluble (S).

La **FIGURA 5** demuestra el comportamiento de termoestabilidad del clon original (n.º 8.93) con sustitución de la región  $\lambda$  J por J1 o J2.

La **FIGURA 6A** demuestra la solubilidad y la elevada expresión de 4 clones independientes con la CDR3 de IGLV3-1 diversificada.

20 La **FIGURA 6B** demuestra una muestra de toda la población de clones con la CDR3 de IGHV3-23 diversificada.

La **FIGURA 7** ilustra las CDR ejemplares (en negrita y/o subrayado) en regiones variables preferidas descritas en el presente documento.

La **FIGURA 8** ilustra un ejemplo de una secuencia de polinucleótidos que codifica un armazón IGLV3-1::IGHV3-23 con regiones variables de CDR3, y la correspondiente, secuencia de aminoácidos traducida. Las CDR están subrayadas y en negrita. En cursiva está una secuencia de enlazador peptídico.

25 La **FIGURA 9** muestra la biblioteca IGLV3-1::IGHV3-23 scFv marcada con el ligando SNAP, que demuestra la alta frecuencia de miembros solubles de la biblioteca.

La **FIGURA 10** muestra el aislamiento de los scFv de unión a mAG de una selección de RED. El clon 34 fue positivo para la unión a mAG. El clon 25 fue negativo.

30 La **FIGURA 11** ilustra la naturaleza soluble del armazón IGLV3-1::IGHV3-23 scFv con la totalidad del scFv  $\alpha$ -mAG aislado de la selección de RED clonado como una proteína de fusión de C-terminal His6 FLAG presente en la fracción soluble (S), sin proteína en la fracción insoluble (P). La detección estaba usando un anticuerpo monoclonal  $\alpha$ -FLAG.

35 La **FIGURA 12** muestra la unión de mAG mediante la fusión de  $\alpha$ -mAG scFv His6FLAG unido a IMAC Ni-sefarosa.

La **FIGURA 13** demuestra la especificidad de la interacción  $\alpha$ -mAG scFv para mAG mediante un "despliegue" de mAG no purificado a partir del lisado de *E. coli*.  $\alpha$ -mAG scFv His6 FLAG se unió a resina IMAC Ni-Sepharose con la adición de mAG en lisado celular total de *E. coli* (carriles 6 y 7) dando como resultado la unión de una proteína del tamaño esperado de mAG (~ 26 kD).

40 La **FIGURA 14** muestra una captura de selección (superior) de la etapa FACS de la selección de biblioteca de MAG "dopada" que usa el bacteriófago defectuoso de lisis encapsulado que muestra la proteína de fusión gpD:: $\alpha$ -mAG scFv. Las células positivas para mAG que contienen fagos encapsulados están en la zona derecha. El bacteriófago recuperado de la selección de FACS se indujo para la replicación del bacteriófago y la expresión de gpD:: $\alpha$ -mAG y se marcó con mAG usando el método RED (abajo).

45

### LEYENDA DEL LISTADO DE SECUENCIAS

SEQ ID NO: 1 -	secuencia de polinucleótidos que codifica IGHV3-23 (Ref. del NCBI NT_026437.12).
SEQ ID NO: 2 -	secuencia de polinucleótidos que codifica IGHV3-23, excluyendo los intrones.
SEQ ID NO: 3 -	secuencia de aminoácidos de IGHV3-23
SEQ ID NO: 4 -	secuencia de polinucleótidos que codifica IGLV3-1 (Ref. del NCBI NT_011520.12).
SEQ ID NO: 5 -	secuencia de polinucleótidos que codifica IGLV3-1, excluyendo los intrones.
SEQ ID NO: 6 -	secuencia de aminoácidos de IGLV3-1
SEQ ID NO: 7 -	secuencia de polinucleótidos que codifica IGLV3-21 (Ref. del NCBI NT_011520.12)
SEQ ID NO: 8 -	secuencia de polinucleótidos que codifica IGLV3-21, excluyendo los intrones.
SEQ ID NO: 9 -	secuencia de aminoácidos de IGLV3-21
SEQ ID NO: 10 -	secuencia de polinucleótidos que codifica IGLV6-57 (Ref. del NCBI: NW_00183 8745.1)
SEQ ID NO: 11 -	secuencia de polinucleótidos que codifica IGLV6-57, excluyendo los intrones.
SEQ ID NO: 12 -	secuencia de aminoácidos de IGLV6-57
SEQ ID NO: 13 -	secuencia de polinucleótidos que codifica IGLV1-51 (Secuencia de referencia del NCBI: NT_011520.12)
SEQ ID NO: 14 -	secuencia de polinucleótidos que codifica IGLV1-51, excluyendo los intrones.
SEQ ID NO: 15 -	secuencia de aminoácidos de IGLV1-51
SEQ ID NO: 16 -	secuencia de polinucleótidos que codifica IGLV1-40 (Secuencia de referencia del NCBI: NT_011520.12)
SEQ ID NO: 17 -	secuencia de polinucleótidos que codifica IGLV1-40, excluyendo los intrones.
SEQ ID NO: 18 -	secuencia de aminoácidos de IGLV1-40

SEQ ID NO: 19 -	secuencia de polinucleótidos que codifica IGLV1-44 (Secuencia de referencia del NCBI: NT_011520.12)
SEQ ID NO: 20 -	secuencia de polinucleótidos que codifica IGLV1-44, excluyendo los intrones.
SEQ ID NO: 21 -	secuencia de aminoácidos de IGLV1-44
SEQ ID NO: 22 -	secuencia de polinucleótidos que codifica IGLV1-47 (Secuencia de referencia del NCBI: NT_011520.12)
SEQ ID NO: 23 -	secuencia de polinucleótidos que codifica IGLV1-47, excluyendo los intrones.
SEQ ID NO: 24 -	secuencia de aminoácidos de IGLV1-47
SEQ ID NO: 25 -	secuencia de polinucleótidos que codifica IGLV3-19 (Secuencia de referencia del NCBI: NT_011520.12)
SEQ ID NO: 26 -	secuencia de polinucleótidos que codifica IGLV3-19, excluyendo los intrones.
SEQ ID NO: 27 -	secuencia de aminoácidos de IGLV3-19
SEQ ID NO: 28 -	Enlazador peptídico preferido
SEQ ID NO: 29 -	Secuencia de variante de CDR
SEQ ID NO: 30 -	Secuencia de variante de CDR alternativa
SEQ ID NO: 31 -	Cebador HVK1 F1
SEQ ID NO: 32 -	Cebador HVK1 F2
SEQ ID NO: 33 -	Cebador HVK2 F
SEQ ID NO: 34 -	Cebador HVK3 F
SEQ ID NO: 35 -	Cebador HVK4 F
SEQ ID NO: 36 -	Cebador HVK5 F
SEQ ID NO: 37 -	Cebador HVK6 F
SEQ ID NO: 38 -	Cebador HVKCL R
SEQ ID NO: 39 -	Cebador HVL1 F1
SEQ ID NO: 40 -	Cebador HVL1 F2
SEQ ID NO: 41 -	Cebador HVL2 F
SEQ ID NO: 42 -	Cebador HVL3 F1
SEQ ID NO: 43 -	Cebador HVL3 F2
SEQ ID NO: 44 -	Cebador HVL4 F1
SEQ ID NO: 45 -	Cebador HVL4 F2
SEQ ID NO: 46 -	Cebador HVL5 F
SEQ ID NO: 47 -	Cebador HVL6 F
SEQ ID NO: 48 -	Cebador HVL7/8 F
SEQ ID NO: 49 -	Cebador HVL9/10 F
SEQ ID NO: 50 -	Cebador 01115 HVLCL R
SEQ ID NO: 51 -	Cebador 01116 HVLCL R2
SEQ ID NO: 52 -	Cebador HVK1 2F1
SEQ ID NO: 53 -	Cebador HVK1 2F2
SEQ ID NO: 54 -	Cebador HVK2 2F
SEQ ID NO: 55 -	Cebador HVK3 2F
SEQ ID NO: 56 -	Cebador HVK4 2F
SEQ ID NO: 57 -	Cebador HVK5 2F
SEQ ID NO: 58 -	Cebador HVK6 2F
SEQ ID NO: 59 -	Cebador HVKCL 2R
SEQ ID NO: 60 -	Cebador HVL1 2F1
SEQ ID NO: 61 -	Cebador HVL1 2F2
SEQ ID NO: 62 -	Cebador HVL2 2F
SEQ ID NO: 63 -	Cebador HVL3 2F1
SEQ ID NO: 64 -	Cebador HVL3 2F2
SEQ ID NO: 65 -	Cebador HVL4 2F1
SEQ ID NO: 66 -	Cebador HVL4 2F2
SEQ ID NO: 67 -	Cebador HVL5 2F
SEQ ID NO: 68 -	Cebador HVL6 2F
SEQ ID NO: 69 -	Cebador HVL7/8 2F
SEQ ID NO: 70 -	Cebador HVL9/10 2F
SEQ ID NO: 71 -	Cebador HVLCL 2R
SEQ ID NO: 72 -	Región J1 de Lamda J
SEQ ID NO: 73 -	Región J2 de Lamda J
SEQ ID NO: 74 -	Región J3 de Lamda J
SEQ ID NO: 75 -	Región J4 de Lamda J
SEQ ID NO: 76 -	Región J5 de Lamda J
SEQ ID NO: 77 -	Región J6 de Lamda J
SEQ ID NO: 78 -	Región J7 de Lamda J
SEQ ID NO: 79 -	Secuencia de la región J híbrida
SEQ ID NO: 80 -	Cebador de PCR
SEQ ID NO: 81 -	Secuencia traducida

SEQ ID NO: 82 - Cebador de PCR  
 SEQ ID NO: 83 - Secuencia traducida  
 SEQ ID NO: 84 - Secuencia de polinucleótidos que codifica un armazón IGLV3-1::IGHV3-23 con regiones variables de CDR3  
 SEQ ID NO: 85 - Secuencia de aminoácidos codificada por el polinucleótido de la SEQ ID NO: 84  
 SEQ ID NO: 86 - Secuencia de molde de CDR3  
 SEQ ID NO: 87 - Secuencia alternativa de molde de CDR3  
 SEQ ID NO: 88 - Secuencia marco de IGLV3-1 y la región J de IGHV3-23  
 SEQ ID NO: 89 - secuencia de intervención  
 SEQ ID NO: 90 - Cebador degenerado 1  
 SEQ ID NO: 91 - Cebador degenerado 2  
 SEQ ID NO: 92 - Bucle L1 de CDR3  
 SEQ ID NO: 93 - Bucle H1 de CDR3  
 SEQ ID NO: 94 - Bucle L2 de CDR3  
 SEQ ID NO: 95 - Bucle H2 de CDR3  
 SEQ ID NO: 96 - Bucle L3 de CDR3  
 SEQ ID NO: 97 - Bucle H3 de CDR3  
 SEQ ID NO: 98 - Bucle L4 de CDR3  
 SEQ ID NO: 99 - Bucle H4 de CDR3  
 SEQ ID NO: 100 - Bucle L5 de CDR3  
 SEQ ID NO: 101 - Bucle H5 de CDR3  
 SEQ ID NO: 102 - Bucle L6 de CDR3  
 SEQ ID NO: 103 - Bucle H6 de CDR3  
 SEQ ID NO: 104 - Bucle L8 de CDR3  
 SEQ ID NO: 105 - Bucle H8 de CDR3  
 SEQ ID NO: 106 - Bucle L9 de CDR3  
 SEQ ID NO: 107 - Bucle H9 de CDR3  
 SEQ ID NO: 108 - Bucle L10 de CDR3  
 SEQ ID NO: 109 - Bucle H10 de CDR3  
 SEQ ID NO: 110 - proteína mAG-BioHis6  
 SEQ ID NO: 111 - Secuencia de scFv anti-mAG-BioHis6  
 SEQ ID NO: 112 - Secuencia de polinucleótidos de construcción de fusión gpD::α-mAG scFv  
 SEQ ID NO: 113 - Secuencia de proteína de fusión gpD::α-mAG scFv  
 SEQ ID NO: 114 - IGLV 3-1 humana de tipo silvestre  
 SEQ ID NO: 115 - Clon soluble 8.93  
 SEQ ID NO: 116 - Clon soluble 8.184  
 SEQ ID NO: 117 - Clon soluble 8.174  
 SEQ ID NO: 118 - Clon 8.186 de IGLV 3-21 humano soluble  
 SEQ ID NO: 119 - Clon 8.39 de IGLV 3-21 humano soluble  
 SEQ ID NO: 120 - IGLV 3-21 humana de tipo silvestre  
 SEQ ID NO: 121 - Clon 9.19 de IGLV 3-21 humano soluble  
 SEQ ID NO: 122 - IGLV 6-57 humana de tipo silvestre  
 SEQ ID NO: 123 - Clon soluble 16.26  
 SEQ ID NO: 124 - Clon soluble 16.1  
 SEQ ID NO: 125 - Clon soluble 16.121

## Descripción detallada

### Técnicas y definiciones generales

- 5 Salvo que se defina específicamente de otra manera, todos los términos técnicos y científicos que se utilizan en el presente documento se deben interpretar como que tienen el mismo significado que el normalmente entendido por un experto en la materia (por ejemplo, en química de proteínas, bioquímica, cultivo celular, genética molecular, microbiología e inmunología).
- 10 Salvo que se indique lo contrario, la proteína recombinante, el cultivo celular y las técnicas inmunológicas utilizadas en la presente divulgación son procedimientos convencionales, bien conocidos por el experto en la materia. Tales técnicas se describen y se explican a través de las fuentes bibliográficas tales como, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley y Sons (1984), J. Sambrook et al., Molecular Cloning: A Laboratory Manual, 3ª edición, Cold Spring Harbour Laboratory Press (2001), R. Scopes, Protein Purification - Principals and Practice, 3ª edición, Springer (1994), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volúmenes 1 y 2, IRL Press (1991), D.M. Glover y B.D. Hames (editores), DNA Cloning: A Practical Approach, Volúmenes 1-4, IRL Press (1995 y 1996), y F.M. Ausubel et al. (editores), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, incluyendo todas las actualizaciones hasta la actualidad), Ed Harlow y David Lane (editores) Antibodies: A Laboratory Manual, Cold Spring Harbour Laboratory, (1988), y J.E. Coligan et al. (editores) Current Protocols in Immunology, John Wiley y Sons (incluyendo todas las actualizaciones hasta la
- 15
- 20



actualidad).

Los términos "polipéptido", "proteína" y "péptido" generalmente se usan de manera intercambiable en el presente documento. Tal como se usa en el presente documento, la expresión "polipéptido exógeno" se refiere a un polipéptido codificado por un polinucleótido exógeno. La expresión "polinucleótido exógeno" tal como se usa en el presente documento se refiere a un polinucleótido que es extraño para la célula en la que se ha introducido, o que la secuencia es homóloga para una secuencia en la célula en la que se ha introducido, pero en una posición en el ácido nucleico de la célula hospedadora en la que el polinucleótido no se encuentra normalmente.

Los términos "anticuerpo", "anticuerpos", "molécula de anticuerpo" y "moléculas de anticuerpo" tal como se usan en el presente documento incluyen anticuerpos policlonales, anticuerpos monoclonales, anticuerpos biespecíficos, diacuerpos, triacuerpos, multicuerpos, anticuerpos heteroconjugados, anticuerpos quiméricos que incluyen moléculas intactas así como fragmentos de los mismos, tales como Fab, F(ab')<sub>2</sub>, Fv y scFv y otras moléculas de tipo anticuerpo. El experto en la materia sabrá que generalmente se considera que un anticuerpo es una proteína que comprende una región variable hecha de una pluralidad de cadenas de polipéptidos, por ejemplo, una región variable de cadena ligera (V<sub>L</sub>) y una región variable de cadena pesada (V<sub>H</sub>). Un anticuerpo también puede comprender dominios constantes, que pueden estar dispuestos en una región constante o en un fragmento constante o fragmento cristizable (Fc). Los anticuerpos se pueden unir de manera específica a un antígeno o a unos pocos antígenos estrechamente relacionados. Los anticuerpos de longitud completa generalmente comprenden dos cadenas pesadas (de aproximadamente 50-70 kD) unidos covalentemente y dos cadenas ligeras (de aproximadamente 23 kD cada una). Una cadena ligera generalmente comprende una región variable y un dominio constante y en mamíferos es o bien una cadena ligera κ o una cadena ligera λ. Una cadena pesada generalmente comprende una región variable y uno o dos dominios constantes enlazados mediante una región de bisagra a un(os) dominio(s) constante(s) adicional(es). Las cadenas pesadas de mamíferos son de uno de los siguientes tipos α, δ, ε, γ o μ. Cada cadena ligera está unida covalentemente a una de las cadenas pesadas. Por ejemplo, las dos cadenas pesadas y las cadenas pesada y ligera se pueden mantener juntas mediante enlaces disulfuro intercatenarios y/o mediante interacciones no covalentes. El número de enlaces disulfuro intercatenarios (si están presentes) puede variar entre los diferentes tipos de anticuerpos. Cada cadena tiene una región variable en N-terminal (V<sub>H</sub> o V<sub>L</sub>, en donde cada una son de aproximadamente 110 aminoácidos de longitud) y uno o más dominios constantes en el extremo C-terminal. El dominio constante de la cadena ligera (C<sub>L</sub>, que es de aproximadamente 110 aminoácidos de longitud) se alinea a menudo con y se une mediante enlace disulfuro al primer dominio constante de la cadena pesada (C<sub>H</sub>, que es de aproximadamente 330-440 aminoácidos de longitud). La región variable de la cadena ligera a menudo se alinea con la región variable de la cadena pesada. La cadena pesada del anticuerpo puede comprender 2 o más dominios C<sub>H</sub> (tales como, C<sub>H</sub>2, C<sub>H</sub>3 y similares) y puede comprender una región de bisagra que se puede identificar entre los dominios constantes C<sub>H</sub>1 y C<sub>m</sub>. Los anticuerpos pueden ser de cualquier tipo (por ejemplo, IgG, IgE, IgM, IgD, IgA e IgY), clase (por ejemplo, IgG<sub>1</sub>, IgG<sub>2</sub>, IgG<sub>3</sub>, IgG<sub>4</sub>, IgA<sub>1</sub> e IgA<sub>2</sub>) o subclase. Preferentemente, el anticuerpo es un anticuerpo de murino (ratón o rata) o un anticuerpo de primate (preferentemente humano). El término "anticuerpo" también abarca anticuerpos humanizados, anticuerpos primatizados, anticuerpos humanos y anticuerpos quiméricos.

Tal como se usa en el presente documento, la expresión "región variable" se refiere a las partes de las cadenas ligera y pesada de un anticuerpo, tal como se define en el presente documento, que incluyen secuencias de aminoácidos de CDR; es decir, CDR1, CDR2 y CDR3, y regiones marco (FR, del inglés *Framework region*). V<sub>H</sub> se refiere a la región variable de la cadena pesada. V<sub>L</sub> se refiere a la región variable de la cadena ligera.

Tal como se usa en el presente documento, la expresión "región del armazón" se refiere a todos los restos de la región variable que no sean restos de CDR.

Tal como se usa en el presente documento, la expresión "región marco" (FR) se entenderá que significa una secuencia contigua de restos de región variable que no son restos de CDR. Por lo tanto, todas las FR juntas constituyen la "región de armazón". Cada región variable de un anticuerpo de origen natural típicamente tiene cuatro FR, identificadas como FR1, FR2, FR3 y FR4. Si se definen las CDR de acuerdo con Kabat, los restos ejemplares de la FR de cadena ligera (LCFR, de *light chain FR*) se colocan en aproximadamente los restos 1-23 (LCFR1), 35-49 (LCFR2), 57-88 (LCFR3) y 98-107 (LCFR4). Nótese que la λLCFR1 no comprende el resto 10, que se incluye en κLCFR1. Los restos ejemplares de la FR de cadena pesada (HCFR, de *heavy chain FR*) se colocan en aproximadamente los restos 1-30 (HCFR1), 36-49 (HCFR2), 66-94 (HCFR3) y 103-113 (HCFR4).

Tal como se usa en el presente documento, la expresión "regiones determinantes de complementariedad" (CDR; es decir, CDR1, CDR2 y CDR3 o región hipervariable) se refiere a los restos de aminoácidos de una región variable de inmunoglobulina cuya presencia es necesaria para la unión a antígeno. Cada región variable típicamente tiene tres regiones CDR identificadas como CDR1, CDR2 y CDR3. Cada CDR puede comprender restos de aminoácidos de una "región determinante de complementariedad" tal como se define por Kabat (1987 y/o 1991). Por ejemplo, en una cadena pesada, la región variable CDRH1 está entre los restos 31-35, la CDRH2 está entre los restos 50-65 y la CDRH3 está entre los restos 95-102. En una cadena ligera, la CDRL1 está entre los restos 24-34, la CDRL2 está entre los restos 50-56 y la CDRL3 está entre los restos 89-97. Estas CDR también pueden comprender numerosas inserciones, por ejemplo, tal como se describe en Kabat (1987 y/o 1991).

La expresión "región constante" (CR (de *constant region*) o fragmento cristalizante o Fc) tal como se usa en el presente documento, se refiere a una parte de un anticuerpo que comprende al menos un dominio constante y que generalmente (aunque no necesariamente) se glucosila y que se une a uno o más receptores y/o componentes de la cascada del complemento (por ejemplo, confiere funciones efectoras). La región constante de la cadena pesada se puede seleccionar a partir de cualquiera de los cinco isotipos:  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$  o  $\mu$ . Además, las cadenas pesadas de las diversas subclases (tales como las subclases de IgG de cadenas pesadas) son responsables de las diferentes funciones efectoras y, por lo tanto, eligiendo la región constante de cadena pesada, se pueden producir proteínas con la función efectora deseada. Las regiones constantes de cadena pesada preferidas son gamma 1 (IgG 1), gamma 2 (IgG 2) y gamma 3 (IgG 3).

Un "dominio constante" es un dominio en un anticuerpo cuya secuencia es altamente similar en anticuerpos del mismo tipo, por ejemplo, IgG o IgM o IgE. Una región constante de un anticuerpo generalmente comprende una pluralidad de dominios constantes, por ejemplo, la región constante de las cadenas pesadas  $\gamma$ ,  $\alpha$  y  $\delta$  comprenden tres dominios constantes y el Fc de las cadenas pesadas comprenden dos dominios constantes. Una región constante de las cadenas pesadas  $\mu$  y  $\epsilon$  comprende cuatro dominios constantes y la región del Fc comprende dos dominios constantes.

Tal como se usa en el presente documento, el término "Fv" se debe interpretar como que se refiere a cualquier proteína, bien comprendida de múltiples polipéptidos o un único polipéptido, en el que se asocian  $V_L$  y  $V_H$  y forman un complejo que tiene un sitio de unión a antígeno, es decir, capaz de unirse de manera específica a un antígeno. El  $V_H$  y el  $V_L$  que forman el sitio de unión a antígeno pueden estar en una única cadena de polipéptido o en diferentes cadenas de polipéptidos. Además, un Fv desvelado en el presente documento (así como cualquier polipéptido desvelado en el presente documento) puede tener múltiples sitios de unión a antígeno que pueden o no unirse al mismo antígeno. Debe entenderse que el término "Fv" abarca fragmentos que derivan directamente de un anticuerpo, así como de proteínas que se corresponden con tal fragmento producido usando medios recombinantes. En realizaciones preferidas, el  $V_H$  no está unido al dominio constante 1 ( $C_H$ ) de la cadena pesada y/o el  $V_L$  no está unido al dominio constante de una cadena ligera ( $C_L$ ). Los Fv ejemplares que contienen polipéptidos o proteínas incluyen un fragmento Fab, un fragmento Fab', un fragmento F(ab'), un scFv, un diacuerpo, un triacuerpo, un tetracuerpo o un complejo de mayor orden, o cualquiera de los anteriores enlazados a una región constante o dominio del mismo, por ejemplo, los dominios  $C_{H2}$  o  $C_{H3}$ . Un "fragmento Fab" consiste en un fragmento de unión a antígeno monovalente de un anticuerpo, y se puede producir, por ejemplo, mediante la digestión de un anticuerpo completo con la enzima papaína, para producir un fragmento que consiste en una cadena ligera intacta y en una parte de una cadena pesada o se puede producir usando medios recombinantes. Se puede obtener un "fragmento Fab" de un anticuerpo, por ejemplo, tratando una inmunoglobulina completa con pepsina, seguido por reducción, para producir una molécula que consiste en una cadena ligera intacta y en una parte de una cadena pesada. Se obtienen dos fragmentos Fab' por anticuerpo tratado de este modo. También se puede producir un fragmento Fab' por medios recombinantes. Un "fragmentos F(ab')<sub>2</sub>" de un anticuerpo consiste en un dímero de dos fragmentos Fab' que se mantienen juntos mediante dos enlaces disulfuro, y se obtiene tratando un anticuerpo completo con la enzima pepsina, sin la posterior reducción. Un fragmento "Fab<sub>2</sub>" es un fragmento recombinante que comprende dos fragmentos Fab enlazados usando, por ejemplo, una cremallera de leucina o un dominio  $C_{H3}$ .

Un "Fv de cadena simple" o "scFv" es una molécula recombinante que contiene el fragmento de la región variable (Fv) de una inmunoglobulina en la que la región variable de la cadena ligera y la región variable de la cadena pesada se unen de manera covalente mediante un enlazador polipeptídico flexible y adecuado. Un tratamiento detallado del Fv que contiene polipéptidos que se encuentra dentro del alcance de este término se proporciona en el presente documento a continuación.

Tal como se usa en el presente documento, la expresión "sitio de unión a antígeno" se puede considerar que significa una estructura formada por un polipéptido que es capaz de unirse de manera específica a un antígeno. El sitio de unión a antígeno no necesita ser una serie de aminoácidos contiguos, o incluso aminoácidos en una sola cadena de polipéptido. Por ejemplo, en un Fv producido a partir de dos cadenas de polipéptidos diferentes, el sitio de unión a antígeno está constituido por una serie de regiones de un  $V_L$  y un  $V_H$  que interactúan con el antígeno y que generalmente, sin embargo, no siempre están en una o más de las CDR en cada región variable.

Cualquiera de las posiciones de aminoácidos asignadas a las CDR y a las FR en el presente documento se definen de acuerdo con Kabat (1987 y 1991). El experto en la materia será fácilmente capaz de usar otros sistemas de numeración en la realización de la presente divulgación, por ejemplo, el sistema de numeración de bucles hipervariables de Chothia y Lesk (1987 y/o 1989) y/o Al-Lazikani *et al* (1997).

El experto en la materia sabrá que un "enlace disulfuro" es un enlace covalente formado mediante el acoplamiento de grupos tiol. El enlace también se denomina enlace SS- o puente disulfuro. En polipéptidos, un enlace disulfuro generalmente tiene lugar entre los grupos tioles de dos restos de cisteína.

El experto en la materia también sabrá que la expresión "condiciones no reductoras" incluye condiciones suficientes para la oxidación de grupos sulfhidrilo (-SH) en una proteína, por ejemplo, permisivas para la formación de enlaces disulfuro. Por consiguiente, el término "condiciones reductoras" incluye condiciones que no son suficientes para la

oxidación de grupos sulfhidrilo (-SH) en una proteína, por ejemplo, no permisivas para la formación de enlaces disulfuro.

Tal como se usa en el presente documento, el término "antígeno" debería interpretarse como cualquier composición de materia frente a la cual se puede generar una respuesta de anticuerpos. Los antígenos ejemplares incluyen proteínas, péptidos, polipéptidos, hidratos de carbono, grupos fosfato, fosfopéptidos o polipéptidos, péptidos glucosilados o péptidos, etc.

La descripción y las definiciones de regiones variables y partes de las mismas, de inmunoglobulinas, de anticuerpos y de fragmentos de los mismos en el presente documento se pueden clarificar adicionalmente mediante la discusión en Kabat (1987 y/o 1991), Bork *et al* (1994) y/o Chothia y Lesk (1987 y 1989) o Al-Lazikani *et al* (1997).

Tal como se usa en el presente documento, los términos "conjugar", "conjugado" o variaciones de los mismos se usan ampliamente para referirse a cualquier forma de asociación covalente o no covalente entre un compuesto útil en los métodos desvelados en el presente documento y otro agente.

El término "y/o", por ejemplo, "X y/o Y" se debe interpretar bien como "X e Y" o "X o Y", y se debe considerar que proporcionan un soporte explícito para ambos significados o para cualquiera de los significados.

El término "aproximadamente" tal como se usa en el presente documento, se refiere a un intervalo de +/-5 % del valor especificado.

Tal como se entenderá a partir de la siguiente descripción, los presentes inventores han aplicado métodos de presentación de proteínas para identificar polipéptidos, (por ejemplo, anticuerpos) que se pueden expresar en forma soluble en el citoplasma celular y que demuestran niveles sorprendentes de solubilidad, termoestabilidad y tolerancia a la diversificación de CDR. Los inventores han demostrado adicionalmente que el repertorio de inmunoglobulinas humanas tiene la capacidad de solubilidad y estabilidad citoplasmática usando solo secuencias de la línea germinal en las regiones marco de los dominios variables de los anticuerpos.

#### Presentación encapsulada retenida (RED):

Los presentes inventores han identificado polipéptido que se pueden expresar en forma soluble en el citoplasma celular y que demuestran niveles sorprendentes de solubilidad, termoestabilidad y tolerancia a la diversificación de CDR usando el método de presentación encapsulada retenida (RED, del inglés *Retained Encapsulated Display*). RED es una plataforma de presentación de proteínas para bacterias gram negativas que se describe en el documento WO 2011/075761. En RED, la proteína que se va a presentar se expresa bien en el periplasma o en el citoplasma de la célula. Las membranas celulares se permeabilizan después con detergentes o disolventes orgánicos mientras que la pared celular se deja intacta. La proteína de presentación está retenida por la pared celular, bien mediante la fusión a proteínas que aumenta su tamaño molecular por encima del límite de porosidad para la pared celular (por ejemplo, la fusión a monómeros de tetrámeros), o mediante la fusión a dominios de proteína que se unen o bien a ADN, a la propia pared celular, o a ambos. La unión fenotipo-genotipo requerida para un sistema de presentación se proporciona a través de la corretencción del plásmido y del ADN genómico en la pared celular de la célula permeabilizada.

#### Polipéptidos:

El repertorio de anticuerpos humanos contiene regiones variables funcionales y pseudogénicas (resumidas por Lefranc, 2000). Éstas se pueden clonar como exones de ADN genómico en linajes no inmunológicos o de ARNm procedente de células inmunológicas que se han sometido a recombinación V(D)J, con el fin de preparar una construcción genética que se pueda usar para expresar el anticuerpo. Durante tal proceso, los dominios variables de las cadenas ligera y pesada se pueden clonar bien como un scFv monomérico o en disposiciones que forman valencias bivalentes o de mayor orden. También se pueden clonar las regiones constantes aguas abajo de los dominios variables para crear anticuerpos Fab o de longitud completa.

En todas las formas, para lograr el correcto plegamiento y mantener la estabilidad y la solubilidad durante la producción de un anticuerpo, las construcciones genéticas que codifican el anticuerpo casi siempre se deben expresar en condiciones tales que se puedan formar enlaces disulfuro intradominio entre las láminas  $\beta$  (es decir, en condiciones no reductoras). Por lo tanto, en células de mamífero, los anticuerpos se insertan en el retículo endoplasmático (RE) y en el Golgi para la secreción o la inserción en la membrana. Si se expresan en un hospedador bacteriano tal como *E. coli*, se deben dirigir al espacio periplásmico, en donde residen las chaperonas del enlace disulfuro DsbA, B y C. Si un anticuerpo se expresa en un ambiente no oxidante, (tal como el citoplasma celular), la ausencia de enlaces disulfuro estabilizantes da como resultado el mal plegamiento y la degradación o, si se expresa en un elevado nivel en el citoplasma de *E. coli*, da como resultado la agregación como un cuerpo de inclusión subcelular.

Los presentes inventores han identificado polipéptidos que comprende armazones de región variable de anticuerpos,

que son capaces de formar un sitio de unión a antígeno incluso cuando los polipéptidos se expresan en un ambiente no oxidante (reductor).

Por consiguiente, la divulgación proporciona un polipéptido aislado y/o recombinante que comprende una región variable de cadena pesada de anticuerpos ( $V_H$ ) de la familia de  $V_H3$  de dominios variables de inmunoglobulina enlazada mediante un enlazador peptídico a una región variable de cadena ligera de anticuerpo ( $V_L$ ) de las familias  $V_L\lambda1$ , 3 o 6 de dominios variables de inmunoglobulinas, en donde el  $V_H$  y el  $V_L$  son capaces de formar un sitio de unión a antígeno.

El polipéptido desvelado en el presente documento se puede proporcionar en la forma de una cualquiera de las formas conocidas de anticuerpos o de fragmentos de anticuerpos. Por lo tanto, el polipéptido desvelado en el presente documento puede ser: (i) un anticuerpo; (ii) un anticuerpo de dominio único; (iii) un Fv de cadena simple (scFv); (iv) un diacuerpo, un triacuerpo o un tetracuerpo; (v) una proteína de fusión que comprende uno cualquiera de (ii)-(iv) y un dominio Fc de un anticuerpo o un dominio del mismo; (vi) una proteína de fusión que comprende uno cualquiera de (ii)-(iv) y una proteína capaz de unirse a una célula efectora inmunológica, o cualquier otra forma conocida de anticuerpo.

Preferentemente, el polipéptido desvelado en el presente documento es un Fv. Por ejemplo, el polipéptido desvelado en el presente documento es un fragmento Fab, un fragmento Fab', un fragmento F(ab'), un scFv, un diacuerpo, un triacuerpo, un tetracuerpo o un complejo polipeptídico de mayor orden.

Lo más preferentemente, el polipéptido desvelado en el presente documento es un scFv. Los scFv comprenden regiones  $V_H$  y  $V_L$  en una única cadena polipeptídica. Preferentemente, la cadena polipeptídica comprende adicionalmente un enlazador de polipéptidos entre el  $V_H$  y el  $V_L$  que permite que el scFv forme la estructura deseada para la unión a antígeno (es decir, para que el  $V_H$  y el  $V_L$  de la única cadena polipeptídica se asocien entre sí para formar el sitio de unión a antígeno). Esto es distinto de un diacuerpo o de un multímero de orden superior desvelado en el presente documento, en los que las regiones variables de diferentes cadenas polipeptídicas se asocian o se unen entre sí. El enlazador peptídico puede comprender 12 o más restos de aminoácidos. Por ejemplo, el enlazador peptídico puede comprender 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 aminoácidos o más. Preferentemente, el enlazador peptídico comprende más de 12 restos de aminoácidos, siendo  $(Gly_4Ser)_3$  (es decir, GGGSGGGSGGGGS (SEQ ID NO: 28)) uno de los enlazadores más favorecidos para un scFv. Se conocen otros enlazadores de polipéptidos en la materia. Los polipéptidos desvelados en el presente documento preferentemente comprenden una región de armazón de un  $V_H$  de la familia  $V_H3$  de inmunoglobulina variable y/o una región de armazón de un  $V_L$  de las familias  $V_L\lambda1$ , 3 o 6 de los dominios variables de inmunoglobulina. Por lo tanto, los polipéptidos desvelados en el presente documento preferentemente comprenden todos los restos de aminoácidos de cualquiera de las regiones variables desveladas en el presente documento, excluyendo los restos de CDR. Los restos de la CDR se pueden identificar fácilmente por el experto en la materia, con referencia a la discusión en Kabat (1987 y/o 1991), Bork *et al* (1994) y/o Chothia y Lesk (1987 y 1989) o Al-Lazikani *et al* (1997). Por lo tanto, los polipéptidos desvelados en el presente documento pueden comprender todas las FR de cualquier región variable desvelada en el presente documento. Los polipéptidos pueden comprender adicionalmente una o más de las CDR de las regiones variables desveladas en el presente documento. Los polipéptidos también pueden comprender una o más CDR que no están presentes en las regiones variables desveladas en el presente documento. Por lo tanto, se pueden insertar una o más CDR de un origen diferente en la región del armazón de las regiones variables desveladas en el presente documento. En el presente documento se incluye un tratamiento adicional de estas posibilidades, a continuación.

En una realización preferida, los scFv desvelados en el presente documento comprenden una región de un  $V_H$  de la familia  $V_H3$  de dominios variables de inmunoglobulina enlazados mediante un enlazador peptídico a un  $V_L$  de las familias  $V_L\lambda1$ , 3 o 6 de dominios variables de inmunoglobulina. En realizaciones preferidas adicionales, los scFv desvelados en el presente documento comprenden una región de armazón de IGHV3-23 y una región de armazón de una cualquiera de IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV3-1, IGLV3-19, IGLV3-21 y IGLV6-57. Lo más preferentemente, los scFv desvelados en el presente documento comprenden una región de armazón de IGHV3-23 y una región de armazón de IGLV3-1.

Los polipéptidos desvelados en el presente documento se pueden definir en términos de su porcentaje de identidad con una secuencia de referencia. El porcentaje de identidad se puede calcular mediante cualquier método adecuado conocido en la materia. Se conocen varios algoritmos para comparar secuencias alineadas, y se pueden usar para determinar el porcentaje de identidad de un polipéptido desvelado en el presente documento con una secuencia de referencia. Por ejemplo, las secuencias de aminoácidos y de polinucleótidos se pueden comparar de forma manual o usando herramientas de comparación e identificación de secuencias basadas en informática que emplean algoritmos tales como BLAST (del inglés *Basic Local Alignment Search Tool* o herramienta básica de búsqueda de alineación local; Altschul *et al.*, 1993); véase también [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)), el método de alineamiento Clustal (Higgins y Sharp, 1989) y otros, en donde los parámetros apropiados para cada comparación de secuencia específica se pueden seleccionar tal como entendería un experto en la materia.

Preferentemente, el polipéptido desvelado en el presente documento es un polipéptido aislado y/o recombinante. El

término "aislado" o "purificado" tal como se usa en el presente documento pretende significar un polipéptido que se ha separado generalmente de los lípidos, de los ácidos nucleicos, de otros polipéptidos y péptidos, y de otras moléculas contaminantes con las que se asocia en su estado natural. Preferentemente, el polipéptido aislado está al menos al 60 %, más preferentemente al menos al 75 % y más preferentemente al menos al 90 % sin otros componentes con los que se asocia de manera natural.

El término "recombinante" en el ámbito de un polipéptido se refiere al polipéptido cuando se produce por una célula, o en un sistema de expresión sin células, en una cantidad alterada o en una tasa alterada en comparación con su estado natural. En una realización, la célula es una célula que no produce de forma natural el polipéptido. Sin embargo, la célula puede ser una célula que comprende un gen no endógeno que provoca una cantidad alterada, preferentemente elevada, del péptido que se va a producir. Un polipéptido recombinante tal como se describe en el presente documento incluye polipéptidos que no se han separado de otros componentes de la célula transgénica (recombinante) o del sistema de expresión sin células en el que se produce, y los polipéptidos producidos en tales células o sistemas sin células que se purifican posteriormente sin al menos algunos otros componentes.

Por lo tanto, el polipéptido desvelado en el presente documento preferentemente comprende secuencias de aminoácidos que derivan de un anticuerpo de murino (ratón o rata) o de un anticuerpo de primate (preferentemente humano). Por lo tanto, las regiones variables y/o las regiones de armazón incluidas en los polipéptidos desvelados en el presente documento pueden ser regiones variables y/o regiones de armazón de murino (ratón o rata) o de primate (preferentemente, humano).

Preferentemente, los polipéptidos desvelados en el presente documento son solubles. Los métodos para determinar la solubilidad de un polipéptido son bien conocidos en la materia, por ejemplo, tal como se describe por J. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3ª edición, Cold Spring Harbour Laboratory Press (2001). Se puede determinar que los polipéptidos son solubles si, por ejemplo, no se pueden separar de una fracción celular lisada y/o permeabilizada mediante separación física (por ejemplo, mediante centrifugación). Además, se puede determinar que los polipéptidos desvelados en el presente son solubles si no forman cuerpos de inclusión en el citoplasma celular. Por lo tanto, se puede considerar que los polipéptidos son solubles si, cuando se expresan en una célula hospedadora, se retienen en una fracción soluble producida tras la lisis de la célula hospedadora mediante cualquiera de los métodos mecánicos, de detergentes y/o enzimáticos adecuados. Los métodos mecánicos adecuados incluyen, por ejemplo, el uso de ultrasonido. Los métodos de detergentes adecuados incluyen, por ejemplo, el uso de n-Octil-β-D-Tioglucoído (8TGP). Los métodos enzimáticos adecuados incluyen, por ejemplo, el uso de lisozima. Preferentemente, los polipéptidos desvelados en el presente documento se pueden retener en una fracción soluble de un lisado celular a un nivel de al menos el 25 %, tal como al menos el 50 %, al menos el 75 %, al menos el 90 %, al menos el 95 %, o al menos el 95 %.

Los polipéptidos desvelados en el presente documento preferentemente son capaces de formar de manera estable un sitio de unión a antígeno. Por lo tanto, los polipéptidos preferentemente son capaces de unirse a un antígeno diana a un nivel que es suficiente como para permitir la detección del complejo polipéptido-antígeno. Tal detección puede tener lugar en cualquiera de las condiciones experimentales adecuadas, tales como una temperatura de al menos 5 °C, al menos 10 °C, al menos 15 °C, al menos 20 °C, al menos 25 °C, al menos 30 °C, al menos 35 °C, al menos 40 °C, al menos 45 °C o al menos 50 °C.

#### Conjugados

El polipéptido desvelado en el presente documento se puede conjugar con uno o más compuestos usando cualquier método adecuado conocido en la materia. Los ejemplos de compuestos con los que se puede conjugar un polipéptido se seleccionan del grupo que consiste en un radioisótopo, un marcador detectable, un compuesto terapéutico, un coloide, una toxina, un ácido nucleico, un péptido, una proteína, un compuesto que aumenta la semivida de la proteína en un sujeto y mezclas de los mismos. Los agentes terapéuticos ejemplares incluyen, pero no se limitan a un agente antiangiogénico, un agente de antineovascularización y/u otro agente de vascularización, un agente antiproliferativo, un agente proapoptótico, un agente quimioterapéutico o un ácido nucleico terapéutico.

Una toxina incluye cualquier agente que sea perjudicial para las células (por ejemplo, las destruye). Para una descripción de estas clases de fármacos que se conocen en la materia y sus mecanismos de acción, véase Goodman et al., *"The Pharmacological Basis of Therapeutics"*, de Goodman y Gilman, 8ª edición, Macmillan Publishing Co., 1990. Las técnicas adicionales relevantes para la preparación de conjugados inmunoglobulina-inmunotoxina se proporcionan en, por ejemplo, Vitetta (1993) y en el documento US 5.194.594. Las toxinas ejemplares incluyen la cadena A de difteria, los fragmentos activos de no unión de la toxina de difteria, la cadena A de la exotoxina (de *Pseudomonas aeruginosa*), la cadena A de ricina, la cadena A de abrina, la cadena A de modeccina, sarcina alfa, las proteínas de *Aleurites fordii*, las proteínas de diantina, las proteínas de *Phytolaca americana* (PAPI, PAPII y PAP-S), el inhibidor de momordica charantia, curcina, crotina, el inhibidor de saponaria officinalis, gelonina, mitogelina, restrictocina, fenomicina, enomicina y los tricotecenos. Véase, por ejemplo, documento WO 93/21232.

Los agentes quimioterapéuticos adecuados para formar los inmunconjugados que comprenden los polipéptidos de

la presente divulgación incluyen auristatinas y maitansinas, taxol, citocalasina B, gramicidina D, bromuro de etidio, emetina, mitomicina, etopósido, tenopósido, vincristina, vinblastina, colchicina, doxorubicina, daunorrubicina, dihidroxiانترacindiona, mitoxantrona, mitramicina, actinomicina D, 1-des-hidrotestosterona, glucocorticoides, procaína, tetracaína, lidocaína, propranolol y puomicina, antimetabolitos (tales como metotrexato, 6-mercaptopurina, 6-tioguanina, citarabina, fludarabina, 5-fluorouracilo, dacarbazina, hidroxiurea, asparaginasa, gemcitabina, cladribina), agentes alquilantes (tales como mecloretamina, tioepa, clorambucilo, melfalán, carmustina (BSNU), lomustina (CCNU), ciclofosfamida, busulfán, dibromomanitol, estreptozotocina, dacarbazina (DTIC), procarbazona, mitomicina C, cisplatino y otros derivados del platino, tales como carboplatino), antibióticos (tales como dactinomicina (formalmente actinomicina), bleomicina, daunorrubicina (formalmente daunomicina), doxorubicina, idarrubicina, mitramicina, mitomicina, mitoxantrona, plicamicina, antramicina (AMC)).

Los ejemplos de inhibidores de angiogénesis adecuados (agentes antiangiogénicos) incluyen, aunque sin limitación, inhibidores de urocinasa, inhibidores de metaloproteasa de matriz (tales como marimastat, neovastat, BAY 12-9566, AG 3340, BMS-275291 y agentes similares), inhibidores de migración y proliferación de células endoteliales (tales como TNP-470, escualamina, 2-metoxiestradiol, combretastatinas, endostatina, angiostatina, penicilamina, SCH66336 (Schering-Plough Corp, Madison, NJ), R115777 (Janssen Pharmaceutica, Inc, Titusville, NJ) y agentes similares), antagonistas de factores de crecimiento angiogénico (tales como ZD6474, SU6668, anticuerpos contra agentes angiogénicos y/o sus receptores (tales como VEGF, bFGF, y angiopoyetina-1), talidomida, análogos de talidomida (tales como CC-5013), Sugen 5416, SU5402, ribozima antiangiogénica (tal como angiozima), interferón  $\alpha$  (tal como interferón  $\alpha 2a$ ), suramina y agentes similares), inhibidores de VEGF-R cinasa y otros inhibidores de tirosina cinasa antiangiogénica (tales como SU011248), inhibidores de señalización de integrina/supervivencia específica de endotelio (tales como vitaxina y agentes similares), antagonistas/quelantes de cobre (tales como tetratimolibdato, captopril y agentes similares), carboxiamido-triazol (CAI), ABT-627, CM101, interleucina-12 (IL-12), IM862, PNU145156E así como moléculas de nucleótidos que inhiben la angiogénesis (tales como ADNc antisentido de VEGF, ADNc que codifica angiostatina, ADNc que codifica p53 y ADNc que codifica el receptor-2 deficiente de VEGF) y agentes similares. Otros ejemplos de inhibidores de angiogénesis, neovascularización y/u otra vascularización son derivados antiangiogénicos de heparina y moléculas relacionadas (por ejemplo, heparinasa III), temozolomida, NK4, factor inhibidor de migración de macrófagos (MIF), inhibidores de ciclooxigenasa-2, inhibidores de factor 1 inducible por hipoxia, isoflavonas antiangiogénicas de soja, oltipraz, fumagilina y análogos de los mismos, análogos de somatostatina, polisulfato de pentosán, tecogalan sodio, dalteparina, tumstatina, trombospondina, NM-3, combrestatina, canstatina, avastatina, anticuerpos contra otras dianas relevantes (tales como mAb anti- $\alpha$ -v/ $\beta$ -3 integrina y anti-kininostatina) y agentes similares.

En un ejemplo, un polipéptido tal como se describe en el presente documento de acuerdo con cualquier realización se conjuga o se enlaza a otro polipéptido, incluyendo otro polipéptido desvelado en el presente documento o un polipéptido que comprende una región variable de inmunoglobulina, tal como una inmunoglobulina o un polipéptido que deriva de la misma, por ejemplo, tal como se describe en el presente documento. Otras proteínas no se excluyen. Las proteínas adicionales serán evidentes para el experto en la materia e incluyen, por ejemplo, un inmunomodulador o una proteína de prolongación de semivida o un péptido u otra proteína que se une a la albúmina del suero, entre otros.

Los inmunomoduladores ejemplares incluyen citocinas y quimiocinas. El término "citocina" es un término genérico para proteínas o péptidos liberados por una población celular que actúan en otra célula como mediadores intercelulares. Los ejemplos de citocinas incluyen linfocinas, monocinas, factores de crecimiento y hormonas de polipéptido tradicionales. Entre las citocinas se incluyen hormonas del crecimiento, tal como la hormona de crecimiento humana, la hormona de crecimiento humana con N-metionilo y la hormona de crecimiento bovina; la hormona paratiroidea, tiroxina, insulina, proinsulina, relaxina, prorrelaxina, hormonas glucoproteicas, tales como la hormona folículo estimulante (FSH), la hormona estimulante de la tiroides (TSH) y la hormona luteinizante (LH), el factor de crecimiento hepático; prostaglandina, el factor de crecimiento de fibroblastos, prolactina, lactógeno placentario, proteína OB, factores  $\alpha$  y  $\beta$  de necrosis tumoral; sustancia inhibidora mulleriana, péptido asociado a gonadotropina, inhibina, activina, factor de crecimiento endotelial vascular, integrina, trombopoyetina (TPO), factores de crecimiento neural, tal como NGF- $\beta$ , factor de crecimiento de plaquetas, factores de crecimiento transformante (TGF), tales como TGF- $\alpha$  y TGF- $\beta$ , factor de crecimiento I o II de tipo insulina, eritropoyetina (EPO), factores osteoinductivos, interferones tales como interferón- $\alpha$ , interferón- $\beta$ , o interferón- $\gamma$ ; factores estimulantes de colonias (CSF), tales como CSF de macrófagos (M-CSF), CSF de granulocitos-macrófagos (GM-CSF); y CSF de granulocitos (G-CSF), interleucinas (IL), tales como IL-1, IL-1 $\alpha$ , IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12; IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-21 y LIF.

Las quimiocinas generalmente actúan como quimioatrayentes para reclutar células inmunológicas efectoras al sitio de expresión de quimiocina. Las quimiocinas incluyen, aunque sin limitación, RANTES, MCAF, MIPI- $\alpha$  o MIPI-Beta. El experto en la materia reconocerá que también se sabe que determinadas citocinas tienen efectos quimioatrayentes y también podrían clasificarse bajo el término quimiocinas.

Los péptidos o proteínas de unión a albúmina sérica ejemplares se describen en los documentos US20060228364 o US20080260757.

Hay una variedad de radionucleidos disponible para la producción de proteínas radioconjugadas. Los ejemplos incluyen, aunque sin limitación, núcleos radiactivos de baja energía (por ejemplo, adecuados para fines de diagnóstico), tales como  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$ ,  $^{125}\text{I}$ ,  $^{123}\text{I}$ ,  $^{99}\text{Tc}$ ,  $^{43}\text{K}$ ,  $^{52}\text{Fe}$ ,  $^{67}\text{Ga}$ ,  $^{68}\text{Ga}$ ,  $^{111}\text{In}$  y similares. Preferentemente, el radionucleido es un radionucleido emisor de gama, de fotones o de positrones con una semivida adecuada para permitir la actividad o la detección tras el tiempo transcurrido entre la administración y la localización en el sitio de la imagen. La presente divulgación también abarca núcleos radiactivos de alta energía (por ejemplo, con fines terapéuticos), tales como  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{123}\text{I}$ ,  $^{111}\text{In}$ ,  $^{105}\text{Rh}$ ,  $^{153}\text{Sm}$ ,  $^{67}\text{Cu}$ ,  $^{67}\text{Ga}$ ,  $^{166}\text{Ho}$ ,  $^{177}\text{Lu}$ ,  $^{186}\text{Re}$  y  $^{188}\text{Re}$ . Estos isótopos típicamente producen partículas  $\alpha$  o  $\beta$  de alta energía que tienen longitud de onda corta. Tales radionucleidos destruyen las células con las que están en estrecha proximidad, por ejemplo, células neoplásicas a las que se ha unido o en las que ha entrado el conjugado. Tienen poco o no tienen efecto sobre células no localizadas y son esencialmente no inmunogénicos. Como alternativa, se pueden generar isótopos de alta energía mediante radiación térmica de un isótopo estable de otra forma, por ejemplo, como en la terapia de captura de neutrones de boro (Guan *et al.*, 1998).

En otra realización, el polipéptido se conjuga con un "receptor" (tal como estreptavidina) para la utilización en el predireccionamiento celular en donde el conjugado se administra al paciente, seguido de la eliminación del conjugado no unido de la circulación usando un agente clarificador y después la administración de un "ligando" (por ejemplo avidina) que se conjuga con un agente terapéutico (por ejemplo un radionucleótido).

Las proteínas de la presente divulgación se pueden modificar para que contengan restos no proteínicos adicionales que son conocidos en la materia y están fácilmente disponibles. Preferentemente, los restos adecuados para la derivatización de la proteína son polímeros solubles en agua. Los ejemplos no limitantes de polímeros hidrosolubles incluyen, aunque sin limitación, polietilenglicol (PEG), alcohol polivinílico (PVA), copolímeros de etilenglicol/propilenglicol, carboximetilcelulosa, dextrano, alcohol polivinílico, polivinilpirrolidona, poli-1, 3-dioxolano, poli-1, 3, 6-trioxano, copolímero de etileno/anhidrido maleico, poliaminoácidos (homopolímeros o copolímeros aleatorios), y dextrano o poli (n-vinilpirrolidona) polietilenglicol, homopolímeros de propilenglicol (PPG), copolímeros de óxido de propileno/óxido de etileno, polioles polioxiethylados (por ejemplo, glicerol; POG), alcohol polivinílico y mezclas de los mismos. El propionaldehído de polietilenglicol puede tener ventajas en la fabricación debido a su estabilidad en agua.

Las moléculas de polímeros típicamente se caracterizan por tener, por ejemplo, de aproximadamente 2 a aproximadamente 1000, o de aproximadamente 2 a aproximadamente 300 unidades de repetición.

Por ejemplo, los polímeros solubles en agua, que incluyen, pero sin limitación, PEG, poli(óxido de etileno) (PEO), polioxiethyleno (POE), alcoholes de polivinilo, hidroxietil celulosas, o dextranos, comúnmente se conjugan con proteínas para aumentar la estabilidad o el tamaño, etc., de la proteína.

PEG, PEO o POE se refiere a un oligómero o polímero de óxido de etileno. En el caso de PEG, estos oligómeros o polímeros se producen mediante, por ejemplo, la polimerización por apertura del anillo aniónico del óxido de etileno iniciada mediante ataque nucleofílico de un ión hidróxido sobre el anillo de epóxido. Una de las formas de PEG más útiles para la modificación de proteínas es monometoxi PEG (mPEG).

Los compuestos particularmente preferidos para la conjugación del polipéptido de la presente divulgación se establecen en la Tabla 1.

**Tabla 1:** Compuestos preferidos para la conjugación

Grupo	Detalle
<b>Radioisótopos (bien directa o indirectamente)</b>	• $^{123}\text{I}$ , $^{125}\text{I}$ , $^{130}\text{I}$ , $^{133}\text{I}$ , $^{135}\text{I}$ , $^{47}\text{Sc}$ , $^{72}\text{As}$ , $^{72}\text{Sc}$ , $^{90}\text{Y}$ , $^{88}\text{Y}$ , $^{97}\text{Ru}$ , $^{100}\text{Pd}$ , $^{101\text{m}}\text{Rh}$ , $^{101\text{m}}\text{Rh}$ , $^{119}\text{Sb}$ , $^{128}\text{Ba}$ , $^{197}\text{Hg}$ , $^{211}\text{At}$ , $^{212}\text{Bi}$ , $^{153}\text{Sm}$ , $^{169}\text{Eu}$ , $^{212}\text{Pb}$ , $^{109}\text{Pd}$ , $^{111}\text{In}$ , $^{67}\text{Ga}$ , $^{68}\text{Ga}$ , $^{67}\text{Cu}$ , $^{75}\text{Br}$ , $^{76}\text{Br}$ , $^{77}\text{Br}$ , $^{99\text{m}}\text{Tc}$ , $^{11}\text{C}$ , $^{13}\text{N}$ , $^{15}\text{O}$ , $^{18}\text{I}$ , $^{188}\text{Re}$ , $^{203}\text{Pb}$ , $^{64}\text{Cu}$ , $^{105}\text{Rh}$ , $^{198}\text{Au}$ , $^{199}\text{Ag}$ o $^{177}\text{Lu}$
<b>Prolongadores de semivida</b>	• Polietilenglicol • Glicerol • Glucosa
<b>Sondas fluorescentes</b>	• Ficoeritrina (PE) • Aloficocianina (APC) • Alexa Fluor 488 • Cy5.5
<b>Agentes biológicos</b>	• Proteínas fluorescentes tales como luciferasa de Renilla, GFP • Inmunomodulador • Toxinas • Una inmunoglobulina • Prolongadores de semivida tales como albúmina

Grupo	Detalle
<b>Agentes quimioterapéuticos</b>	<ul style="list-style-type: none"> <li>• Taxol</li> <li>• 5-Fluorouracil</li> <li>• Doxorubicina</li> <li>• Idarrubicina</li> </ul>

En otro ejemplo de la divulgación, se incluye un resto de espaciador entre el compuesto y el polipéptido con el que se conjuga. Los restos de espaciador desvelados en el presente documento pueden ser escindibles o no escindibles. Por ejemplo, el resto de espaciador escindible puede ser un resto de espaciador escindible por redox, de manera que el resto de espaciador es escindible en ambientes con un potencial redox más bajo, tales como el citoplasma y otras regiones con altas concentraciones de moléculas con grupos sulfhidrilo libres. Los restos de espaciador que se pueden escindir debido a un cambio en el potencial redox incluyen aquellos que contienen disulfuros. El estímulo de escisión se puede proporcionar tras la absorción intracelular de la proteína conjugada donde el potencial redox más bajo del citoplasma facilita la escisión del resto de espaciador.

En otro ejemplo, un descenso en el pH provoca la escisión del espaciador para liberar de este modo el compuesto en una célula diana. Un descenso en el pH está implicado en muchos procesos fisiológicos y patológicos, tales como el tráfico de endosomas, el crecimiento tumoral, la inflamación y la isquemia miocárdica. El pH cae desde un pH fisiológico de 7,4 a 5-6 en endosomas o 4-5 en lisosomas. Los ejemplos de restos de espaciador sensibles al ácido que se pueden usar para el direccionamiento de lisosomas o endosomas de células cancerosas, incluyen aquellos con uniones escindibles con ácido tales como los que se encuentran en acetales, cetales, ortoésteres, hidrazonas, tritilos, cis-aconitilos o tiocarbamoilos (véase por ejemplo, las Pat. de EE.UU. N.º 4.569.789, 4.631.190, 5.306.809, y 5.665.358). Otros restos de espaciadores sensibles al ácido ejemplares comprenden secuencias de dipéptido Phe-Lys y Val-Lys.

Los restos de espaciadores escindibles pueden ser sensibles a agentes de escisión suministrados por vía biológica que se asocian con una célula diana particular, por ejemplo, enzimas lisosómicas o asociadas al tumor. Los ejemplos de restos de unión que se pueden escindir de forma enzimática incluyen, aunque sin limitación, péptidos y ésteres. Los restos de unión escindibles por enzimas incluyen los que son sensibles a proteasas asociadas al tumor tales como catepsina B o plasmina. Los sitios escindibles por catepsina B incluyen las secuencias de dipéptido valina-citrulina, fenilalanina-lisina y/o valina-alanina.

#### Complejos proteicos

Los polipéptidos desvelados en el presente documento se conjugan preferentemente con uno o más compuestos que los hace particularmente adecuados para su uso en el ensayo RED al que se refiere el presente documento. Por ejemplo, el polipéptido se puede asociar con al menos un segundo polipéptido (citado en lo sucesivo en el presente documento como "el segundo polipéptido") para formar un complejo proteico que tiene un tamaño molecular de manera que el complejo proteico se mantiene dentro de una célula bacteriana permeabilizada. El polipéptido se puede asociar con el segundo polipéptido mediante, por ejemplo, enlaces covalentes tales como puentes disulfuro o mediante asociación no covalente. "Asociación no covalente" se refiere a interacciones moleculares que no implican enlaces interatómicos. Por ejemplo, las interacciones no covalentes implican enlaces iónicos, enlaces de hidrógeno, interacciones hidrófobas y fuerzas de van der Waals. Las fuerzas no covalentes se pueden usar para mantener las cadenas de polipéptidos separadas juntas en proteínas o en complejos proteicos. Por lo tanto, el polipéptido y el segundo polipéptido se pueden expresar como polipéptidos separados bien del mismo vector o bien de vectores diferentes, o uno o ambos polipéptidos se pueden expresar a partir de ADN que codifica los polipéptidos que se ha integrado en el genoma de la célula bacteriana.

Como alternativa, el polipéptido y el segundo polipéptido que se asocian en un complejo proteico pueden ser una proteína de fusión. Tal como se usa en el presente documento, "proteína de fusión" se refiere a una proteína híbrida, que consiste en dos o más polipéptidos, o fragmentos de los mismos, que es resultado de la expresión de un polinucleótido que codifica al menos una parte de cada uno de los dos polipéptidos.

#### *Complejos proteicos retenidos en la célula bacteriana permeabilizada por tamaño molecular*

El segundo polipéptido puede ser cualquier polipéptido que tenga un tamaño molecular suficiente, es decir, suficiente peso molecular o radio molecular, de manera que al menos algunos de los complejos formados con el polipéptido que se seleccionan sistemáticamente para una actividad deseada sean incapaces de difundir a partir de la célula bacteriana permeabilizada. Por lo tanto, el complejo proteico se mantiene dentro de la célula bacteriana después de la permeabilización de la célula. El experto en la materia apreciará que la naturaleza del segundo polipéptido, incluyendo su peso molecular y si es una proteína globular o en bastón (filamentosa), determinará su capacidad para evitar o inhibir la difusión del complejo proteico a través de la pared celular bacteriana. En una realización, el peso molecular del segundo polipéptido es al menos de aproximadamente 30 kDa, o al menos aproximadamente 40, 50, 60, 70, 80, 90, 100, 120, 130, 140, 150 o más kDa. En una realización, el segundo polipéptido es al menos aproximadamente 120 kDa.



En una realización, el segundo polipéptido forma multímeros que tienen un tamaño molecular mayor que el tamaño de exclusión de poro de la célula bacteriana permeabilizada. Tal como se usa en el presente documento, el término "multímero" y las variaciones gramaticales del mismo se refieren a la formación de un complejo multimérico entre dos o más moléculas distintas. El multímero puede comprender, por ejemplo, dos o más moléculas de la misma proteína (es decir, un homomultímero) o una mezcla de dos o más proteínas diferentes o no idénticas (es decir, un heteromultímero). Las proteínas que forman multímeros adecuadas para su uso en los métodos desvelados en el presente documento incluyen los que forman dímeros, trímeros, tetrámeros, pentámeros, hexámeros y multímeros de mayor orden que comprenden siete o más subunidades.

Las proteínas multiméricas incluyen homodímeros, por ejemplo, receptor  $\alpha$  de PDGF, e isoformas  $\beta$ , receptor de eritropoyetina, MPL y receptor de G-CSF, cuyas subunidades del heterodímero tienen cada una dominios de unión al ligando y dominios efectores, por ejemplo, la isoforma  $\alpha\beta$  del receptor de PDGF, y los multímeros que tienen subunidades de componentes con funciones dispares, por ejemplo, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7 y receptores de GM-CSF. Los ejemplos no limitantes de otras proteínas multiméricas que se pueden utilizar en los métodos de la presente divulgación incluyen factores implicados en la síntesis o replicación de ADN, tales como las proteínas ADN polimerasa implicadas en la producción de ARNm, tales como TFIID y TFIIH; proteínas celulares, nucleares y otras proteínas asociadas a la membrana, tales como hormonas y otros receptores de transducción de la señal, proteínas de transporte activo y canales iónicos, proteínas multiméricas en la sangre, incluyendo la hemoglobina, el fibrinógeno y el Factor de Willabrand; proteínas que forman estructuras dentro de la célula, tales como actina, miosina y tubulina, y otras proteínas del citoesqueleto; proteínas que forman estructuras en el ambiente extracelular, tales como colágeno, elastina y fibronectina; proteínas implicadas en el transporte intracelular y extracelular, tales como cinesina y dineína, la familia SNARE de proteínas (receptor de proteína de unión NSF soluble) y clatrina; proteínas que ayudan a regular la estructura de la cromatina, tales como histonas y protaminas, Swi3p, Rsc8p y moira; factores de transcripción multiméricos tales como Fos, Jun y CBTF (factor de transcripción de caja CCAAT); enzimas multiméricas tales como acetilcolinesterasa y alcohol deshidrogenasa; proteínas chaperonas tales como GroE, Gro EL (chaperonina 60) y Gro ES (chaperonina 10); antitoxinas, tales como veneno de serpiente, toxina botulínica, super antígenos de *Streptococcus*; lisinas (enzimas de bacteriófagos y de virus); así como la mayoría de las proteínas alostéricas. En una realización, la proteína multimérica es una proteína de *E. coli*. Los ejemplos no limitantes de proteínas de *E. coli* que forman multímeros incluyen L-ramnosa isomerasa (RhnA; por ejemplo, el registro del NCBI CAA43002),  $\beta$ -galactosidasa ( $\beta$ -gal; por ejemplo, el registro del NCBI YP 001461520), betaína aldehído deshidrogenasa (BetB; por ejemplo, el registro del NCBI AAA23506), glutamato-5-cinasa (G5K; por ejemplo, el registro del NCBI AAB08662), glutation sintasa (GshB; por ejemplo, el registro del NCBI AP\_003504), y un aldehído deshidrogenasa de cadena media (YdcW; por ejemplo, el registro del NCBI AP\_002067).

En una realización, el polipéptido tiene un tamaño molecular suficiente como para mantener el polipéptido dentro de la pared celular bacteriana. Por lo tanto, el experto en la materia apreciará que tal polipéptido no necesariamente necesita estar asociado con un segundo polipéptido para mantener el polipéptido dentro de la célula bacteriana permeabilizada.

#### Proteínas de unión a ADN

Los presentes inventores han descubierto que el ADN se mantiene dentro de una célula bacteriana tras la permeabilización. Por lo tanto, en una realización, el polipéptido se asocia con una proteína de unión a ADN para formar un complejo proteico que se une al ADN y que se mantiene dentro de la célula bacteriana. Tal como se usa en el presente documento, "proteína de unión a ADN" se refiere a cualquier proteína que comprende un dominio de unión a antígeno que comprende al menos un motivo que reconoce ADN de cadena doble o de cadena sencilla. Tal como sabría un experto en la materia, los dominios de unión a ADN incluyen la hélice-giro-hélice, los dedos de zinc, la cremallera de leucina, la hélice alada, la hélice alada-giro-hélice, la hélice-bucle-hélice, el pliegue de inmunoglobulina que reconoce ADN, o los dominios B3. Asociar el polipéptido con una proteína de unión a ADN ventajosamente proporciona una mejor recuperación de ADN, por ejemplo, un plásmido, que codifica el polipéptido en los métodos de selección sistemática desvelados en el presente documento.

Los ejemplos de proteínas de unión a ADN incluyen proteínas de competencia bacteriana tales como, aunque sin limitación, proteínas de unión a ADN de *E. coli*, proteínas de unión a ADN de *Neisseria gonorrhoeae*, por ejemplo ComE, proteínas E2 de adenovirus, factor de transcripción AraC, factores de transcripción básicos de hélice-bucle-hélice, factores de transcripción básicos de cremallera de leucina, factor de respuesta de butirato, proteína B del centrómero, factores de transcripción COUP, factores de transcripción de respuesta de crecimiento temprano, factores de unión a la G-box, factores de transcripción GATA, proteínas HMGA, proteínas de homodominio, proteínas B I-kappa, factores de integración del hospedador, factores reguladores de interferón, factor 3 del gen estimulado por interferón, factores de transcripción de tipo Kruppel, proteína reguladora sensible a la leucina, proteínas de unión a la región de unión a la matriz, proteína de unión a metil-CpG, proteína MutS homóloga 2, proteína de leucemia mieloide-linfoide, NF-Kappa B, factores de transcripción NF1, factores respiratorios nucleares, proteína oncogénica p55, complejo de reconocimiento de origen, factores de transcripción de cajas emparejadas, factores de dominio POU, factores protooncogénicos, recombinasa Rad51, proteína de recombinación y reparación de ADN Rad52, proteína A de replicación, proteína C de replicación, proteína de retinoblastoma, proteínas Smad, factores de transcripción SOX, proteínas de dominio T-box, factores de transcripción TCF, proteína de unión a

telómero, receptor 9 de tipo Toll, transactivadores y factores de transcripción de hélice alada. En una realización, la proteína de unión a ADN es una proteína de unión a ADN de *E. coli*. En otra realización, la proteína de unión a ADN es una proteína de *Neisseria gonorrhoeae*, por ejemplo, ComE.

## 5 Proteínas de unión a pared celular

El polipéptido se puede asociar con una proteína de unión a la pared celular bacteriana. El experto en la materia entenderá que la elección de una proteína de unión a la pared celular dependerá de la especie de célula hospedadora, ya que las diferentes bacterias tienen diferentes composiciones de pared celular. Aunque las bacterias tienen paredes celulares hechas de peptidoglucano (PG), las modificaciones químicas entre especies pueden afectar a la unión de especies cruzadas. El experto en la materia será fácilmente capaz de determinar las proteínas de unión a la pared celular adecuadas para su uso en una especie bacteriana particular.

Las proteínas de unión a la pared celular bacteriana incluyen proteínas conocidas por tener una estructura de dominio, por lo que parte de la cadena polipeptídica en la estructura natural es capaz de reconocer y unirse a moléculas específicas o conformaciones moleculares en la pared celular bacteriana. Por lo tanto, el término "proteínas de unión a pared celular" incluye un dominio de proteína que es parte de la proteína que se une específicamente a la pared celular bacteriana. Los ejemplos de proteínas de unión a la pared celular bacteriana incluyen las hidrolasas de pared celular codificadas por bacteriófagos, las hidrolasas de pared celular de bacterias y diferentes autolisinas. También se incluyen moléculas de receptor codificadas por el ADN de bacteriófagos y otros virus. Cuando la proteína de unión a la pared celular bacteriana es de enzimas hidrolíticas de origen bacteriófago, que son capaces de unirse específicamente a las bacterias, la proteína de unión a la pared celular mantiene su capacidad de unión, pero preferentemente no tiene actividad hidrolítica significativa.

En una realización, la proteína de unión a la pared celular se une de manera no covalente a la pared celular de *E. coli*. Por ejemplo, para una célula hospedadora *E. coli* hay proteínas de unión a PG endógeno con un dominio de unión a PG de aproximadamente 100 aminoácidos conservados que se da en PAL, OmpA, YiaD, YfiB y MotB (Parsons et al., 2006). Sin embargo, se ha demostrado que las proteínas de otros organismos se expresan bien en *E. coli* y se unen a la pared celular con alta afinidad, por ejemplo, el dominio de unión a PG de aproximadamente 70 aminoácidos de *Pseudomonas*  $\phi$ KZ fago (KzPG) (Briers et al., 2009). Por lo tanto, un dominio de unión a PG de una proteína que se une a PG se puede usar como una proteína de unión a la pared celular bacteriana en los métodos desvelados en el presente documento.

En una realización ejemplar, el dominio de unión a PG se puede fusionar con el polipéptido desvelado en el presente documento y expresar en el citosol de la célula bacteriana. Tras la permeabilización de la membrana, el dominio de unión a PG se une a la pared celular, dando como resultado la retención del polipéptido de interés en la célula permeabilizada. Además, para mejorar potencialmente la retención del polipéptido de interés en la célula, el experto en la materia entenderá que el polipéptido se puede asociar con una proteína de unión a ADN además de una proteína de unión a la pared celular.

Como alternativa, el polipéptido se puede asociar con una proteína que es capaz de unirse covalentemente a la pared celular bacteriana. Preferentemente, la proteína comprende una señal de direccionamiento periplásmico. Por lo tanto, el polipéptido se expresa en el citosol de la célula bacteriana, pero se dirige al periplasma, en donde se enlaza a la pared celular antes de la permeabilización de la membrana.

A modo de ejemplo no limitante, la proteína de unión a la pared celular bacteriana que se une de manera covalente a la pared celular puede ser una lipoproteína capaz de unirse a la pared celular y que carece de una secuencia señal en N-terminal funcional necesaria para la unión a la membrana externa. Por ejemplo, la lipoproteína puede ser LPP de *E. coli*. La LPP es una proteína de *E. coli* abundante que forma una superhélice trimérica. En su forma natural, un extremo está unido a la membrana externa mediante lipidación y el otro está unido de manera covalente a la pared celular mediante una lisina en C-terminal. La lipoproteína puede comprender adicionalmente una secuencia que dirige a la lipoproteína al periplasma, por ejemplo, una secuencia de direccionamiento periplásmico OmpF. En una realización, la lipoproteína es lipoproteína de *E. coli* que carece de una secuencia señal funcional en N-terminal necesaria para la unión a la membrana externa.

A la luz de las enseñanzas de la presente divulgación, el experto en la materia será capaz de identificar o de diseñar proteínas que se unen de manera covalente a la pared celular bacteriana y que son adecuadas para su uso en los métodos de la presente divulgación.

En una realización de la divulgación, el polipéptido desvelado en el presente documento es un polipéptido de fusión que comprende un dominio KzPG y uno o más de otros dominios seleccionados de un espaciador, SNAP y/o DBP. En una realización particular, el polipéptido de fusión comprende uno o más espaciadores y los dominios KxPG, SNAP y DBP.

Polinucleótidos

La presente divulgación también proporciona un polinucleótido que codifica un polipéptido desvelado en el presente documento. Preferentemente, el polinucleótido es un polinucleótido aislado o recombinante.

El término "polinucleótido aislado" pretende significar un polinucleótido que se ha separado de manera general de las secuencias de polinucleótidos con las que se asocia o enlaza en su forma natural. Preferentemente, el polinucleótido aislado está al menos al 60 %, más preferentemente al menos al 75 % y más preferentemente al menos al 90 % sin otros componentes con los que se asocia de manera natural. Además, el término "polinucleótido" se usa de manera intercambiable en el presente documento con los términos "molécula de ácido nucleico", "gen" y "ARNm".

El término "recombinante" en el contexto de un polinucleótido se refiere a cuando el polinucleótido está presente en una célula o en un sistema de expresión sin célula, en una cantidad alterada en comparación con su estado natural. En una realización, la célula es una célula que no comprende de forma natural al polinucleótido. Sin embargo, la célula puede ser una célula que comprende un polinucleótido no endógeno, dando como resultado una cantidad de producción alterada, preferentemente aumentada, del polipéptido codificado. Un polinucleótido recombinante desvelado en el presente documento incluye polinucleótidos que no se han separado de otros componentes de la célula transgénica (recombinante) o del sistema de expresión sin célula, en el que está presente, y polinucleótidos producidos en tales células o sistemas sin células que posteriormente se purifican sin al menos algunos otros componentes.

"Polinucleótido" se refiere a un oligonucleótido, un polinucleótido o cualquier fragmento del mismo. Puede ser ADN o ARN de origen genómico o sintético, de cadena doble o de cadena sencilla, y combinado con carbohidratos, lípidos, proteínas u otros materiales para realizar una actividad particular definida en el presente documento.

El ADN que codifica un polipéptido que comprende una región variable se puede aislar usando métodos convencionales en la materia. Por ejemplo, se pueden diseñar cebadores para alinearse a regiones conservadas en una región variable que flanquea a la región de interés, y estos cebadores se pueden usar después para amplificar el ácido nucleico intermedio, por ejemplo, mediante PCR. Los métodos adecuados y/o cebadores se conocen en la materia y/o se describen, por ejemplo, en Borrebaeck (ed), 1995 y/o Froyen *et al.*, 1995. Las fuentes adecuadas de ADN molde para tales métodos de amplificación pueden derivar de, por ejemplo, hibridomas, transfectomas y/o de células que expresan proteínas que comprenden una región variable, por ejemplo, tal como se describe en el presente documento.

El polinucleótido desvelado en el presente documento puede codificar todo el polipéptido de la divulgación. Como alternativa, el polinucleótido puede codificar una única cadena pesada o ligera del polipéptido desvelado en el presente documento. Por lo tanto, dos polipéptidos, que codifican cada uno una de las cadenas pesada o ligera, se pueden producir y expresar en una única célula para producir el polipéptido desvelado en el presente documento.

Preferentemente, los polinucleótidos codifican la región de armazón de las regiones variables, y también una o más CDR. Lo más preferentemente, los polinucleótidos desvelados en el presente documento codifican la región de armazón de las regiones variables y las tres CDR. Los polinucleótidos desvelados en el presente documento se pueden mutagenizar con el fin de producir una variedad en las secuencias de aminoácidos de las CDR y posiblemente también en las secuencias de aminoácidos de las regiones de armazón. El experto en la materia sabrá de métodos adecuados para este fin.

El polinucleótido desvelado en el presente documento también puede codificar un conjugado de proteínas que está conjugado o es capaz de conjugarse con un polipéptido desvelado en el presente documento, tal como se describe en el presente documento.

Producción de polipéptidos

Los polipéptidos desvelados en el presente documento se pueden sintetizar mediante cualquiera de los métodos conocidos en la materia, tales como la producción y la recuperación de polipéptidos recombinantes, y mediante la síntesis química de los polipéptidos. Por lo tanto, la presente divulgación también proporciona un método de producción de los polipéptidos desvelados en el presente documento.

Los polipéptidos desvelados en el presente documento se pueden producir en condiciones reductoras o no reductoras. Preferentemente, los polipéptidos desvelados en el presente documento se producen en condiciones reductoras, tales como en el citoplasma de una célula hospedadora.

En el caso de un polipéptido recombinante, el ácido nucleico que codifica al mismo se coloca preferentemente en uno o más vectores de expresión, que después se transfectan a células hospedadoras, por ejemplo células de *E. coli*, células de levadura, células de insecto o células de mamífero, tales como células COS de simio, células de ovario de hámster chino (CHO) o células de mieloma que de otro modo no producen proteína inmunoglobulina, para

obtener la síntesis de proteínas en las células hospedadoras recombinantes. Los artículos de revisión sobre expresión recombinante en bacterias de ADN que codifican la inmunoglobulina incluyen Skerra *et al*, (1993) y Plückthun, (1992). Las técnicas de clonación molecular para lograr estos fines se conocen en la materia y se describen, por ejemplo, en Ausubel o Sambrook. Una amplia variedad de métodos de clonación y de amplificación *in vitro* son adecuados para la construcción de ácidos nucleicos recombinantes. Los métodos de producción de inmunoglobulinas recombinantes también se conocen en la materia. Véase los documentos US4.816.567; US5225539, US6054297, US7566771 o US5585089.

Tras el aislamiento, el ácido nucleico que codifica un polipéptido desvelado en el presente documento se inserta preferentemente en una construcción de expresión o vector replicable para la clonación adicional (amplificación del ADN) o para la expresión en un sistema sin células o en células. Preferentemente, el ácido nucleico está unido de manera operativa a un promotor.

Tal como se usa en el presente documento, el término "promotor" debe tomarse en su contexto más amplio e incluye las secuencias reguladoras transcripcionales de un gen genómico, incluyendo la caja TATA o el elemento iniciador, que se requiere para una iniciación de la transcripción precisa, con o sin elementos reguladores adicionales (por ejemplo, secuencias de activación aguas arriba, sitios de unión de factor de transcripción, potenciadores y silenciadores) que alteran la expresión de un ácido nucleico, por ejemplo, en respuesta a un estímulo de desarrollo y/o externo, o de una forma específica de tejido. En el presente contexto, el término "promotor" también se usa para describir un ácido nucleico recombinante, sintético o de fusión, o derivado que confiere, activa o mejora la expresión de un ácido nucleico al que está unido de manera operativa. Los promotores preferidos pueden contener copias adicionales de una o más elementos reguladores específicos para mejorar adicionalmente la expresión y/o alterar la expresión espacial y/o la expresión temporal de dicho ácido nucleico.

Tal como se usa en el presente documento, el término "unido de manera operativa a" significa colocar un promotor en relación con un ácido nucleico de manera que la expresión del ácido nucleico esté controlada por el promotor.

Los sistemas de expresión sin células también se contemplan mediante la presente divulgación. Por ejemplo, un ácido nucleico que codifica un polipéptido desvelado en el presente documento puede estar unido de manera operativa a un promotor adecuado, por ejemplo, un promotor de T7, y la construcción de expresión resultante expuesta a condiciones suficientes para la transcripción y la traducción. Se han descrito vectores de expresión típicos para la expresión *in vitro* o la expresión sin células e incluyen, pero sin limitación, los sistemas TNT T7 y TNT T3 (Promega), los vectores pEXP1-DEST y pEXP2-DEST (Invitrogen).

Hay muchos vectores disponibles para la expresión en células. Los componentes del vector generalmente incluyen, aunque sin limitación, uno o más de los siguientes: una secuencia señal, una secuencia que codifica una proteína de la presente divulgación (por ejemplo, derivada de la información proporcionada en el presente documento), un elemento potenciador, un promotor y una secuencia de terminación de la transcripción. El experto en la materia sabrá de secuencias adecuadas para la expresión de una proteína. Por ejemplo, las secuencias señal ejemplares incluyen señales de secreción en procariotas (por ejemplo, pelB, fosfatasa alcalina, penicilinas, lpp o enterotoxina II termoestable), señales de secreción en levadura (por ejemplo, líder de invertasa, líder de factor  $\alpha$  o líder de fosfatasa ácida) o señales de secreción en mamíferos (por ejemplo, señal gD de herpes simple).

En una realización preferida, el polinucleótido que codifica el polipéptido desvelado en el presente documento se inserta en un vector que es particularmente adecuado para la expresión en el sistema RED descrito en el presente documento. Por lo tanto, el vector puede ser particularmente adecuado para la expresión en una célula bacteriana. Por ejemplo, el vector puede comprender un sitio para insertar en el vector un polinucleótido que codifica un polipéptido desvelado en el presente documento, y un marco de lectura abierta que codifica un segundo polipéptido desvelado en el presente documento para formar un complejo proteico que se puede mantener dentro o se puede unir a la pared celular de una célula bacteriana permeabilizada. Los vectores adecuados se describen en el documento WO2011/075761.

Preferentemente, el vector también es capaz de replicarse en la célula bacteriana de manera independiente del genoma del hospedador. Los vectores adecuados incluyen plásmidos, virus y cósmidos así como elementos de ADN lineal, tal como el fago lineal N15 de *E. coli*, y/o ADN extracromosómico que se replica de manera independiente del genoma de la célula bacteriana.

El experto en la materia será capaz de determinar fácilmente cepas bacterianas adecuadas para la expresión de los polipéptidos en los métodos desvelados en el presente documento. Los expertos en la materia entenderían que las bacterias Gram negativas son adecuadas para su uso en los métodos desvelados en el presente documento, incluyendo, por ejemplo, *Salmonella*, *E. coli*, *Shigella*, *Campylobacter*, *Fusobacterium*, *Bordetella*, *Pasteurella*, *Actinobacillus*, *Haemophilus* y *Histophilus*. En una realización preferida, la bacteria Gram negativa es *E. coli*.

Los promotores ejemplares que se pueden incluir en el vector desvelado en el presente documento incluyen los activos en procariotas (por ejemplo, el promotor phoA, los sistemas promotores de  $\beta$ -lactamasa y lactosa, fosfatasa alcalina, un sistema de promotor de triptófano (trp) y promotores híbridos, tales como el promotor tac). Estos

promotores son útiles para la expresión en procariotas que incluyen eubacterias, tales como organismos Gram negativos o Gram positivos, por ejemplo, Enterobacteriaceae tales como *Escherichia*, por ejemplo, *E. coli*, *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, por ejemplo, *Salmonella typhimurium*, *Serratia*, por ejemplo, *Serratia marcescans*, y *Shigella*, así como *Bacilli* tales como *B. subtilis* y *B. licheniformis*, *Pseudomonas* tales como *P. aeruginosa*, y *Streptomyces*. Preferentemente, el hospedador es *E. coli*. Un hospedador de clonación de *E. coli* preferido es *E. coli* 294 (ATCC 31.446), aunque otras cepas tales como *E. coli* B, *E. coli* X 1776 (ATCC 31.537), y *E. coli* W3110 (ATCC 27.325), DH5 $\alpha$  o DH10B son adecuados.

Los promotores ejemplares activos en células de mamífero incluyen el promotor temprano inmediato del citomegalovirus (CMV-IE), el promotor de factor de elongación 1- $\alpha$  humano (EF1), promotores de ARN nuclear pequeño (U1a y U1b), el promotor de cadena pesada de  $\alpha$ -miosina, el promotor del virus de simio 40 (SV40), el promotor del virus del sarcoma de Rous (RSV), el promotor tardío principal de adenovirus, el promotor de  $\beta$ -actina; el elemento regulador híbrido que comprende un potenciador de CMV/ promotor de  $\beta$ -actina o un promotor de inmunoglobulina o fragmento activo de la misma. Los ejemplos de células de mamífero útiles son la línea CV1 de riñón de mono trasformada mediante SV40 (COS-7, ATCC CRL 1651); la línea de riñón embrionario humano (293 o células 293 subclonadas para su crecimiento en cultivo en suspensión; células de riñón de cría de hámster (BHK, ATCC CCL 10); o células de ovario de hámster chino CHO.

Los promotores típicos adecuados para la expresión en células de levadura tales como por ejemplo una célula de levadura seleccionada del grupo que comprende *Pichia pastoris*, *Saccharomyces cerevisiae* y *S. pombe*, incluyen, aunque sin limitación, el promotor *ADH1*, el promotor *GAL1*, el promotor *GAL4*, el promotor *CUP1*, el promotor *PHO5*, el promotor *nmt*, el promotor *RPR1* o el promotor *TEF1*.

Los promotores típicos adecuados para la expresión en células de insecto incluyen, aunque sin limitación, el promotor OPEI2, el promotor de actina de insecto aislado de *Bombyx muri*, el promotor *dsh* de *Drosophila sp.* (Marsh et al., 2000) y el promotor de metalotioneína inducible. Las células de insecto preferidas para la expresión de proteínas recombinantes incluyen una célula de insecto seleccionada del grupo que comprende, células BT1-TN-5B1-4 y células de *Spodoptera frugiperda* (por ejemplo, células sf19, células sf21). Los insectos adecuados para la expresión de los fragmentos de ácido nucleico incluyen, pero sin limitación, *Drosophila sp.*. También se contempla el uso de *S. frugiperda*.

Los medios para introducir la molécula de ácido nucleico aislada o una construcción génica que comprenda a la misma en una célula para la expresión son conocidos por los expertos en la materia. La técnica usada para una célula dada depende de las técnicas de éxito conocidas. Los medios para la introducción de ADN recombinante en las células incluyen la microinyección, la transfección mediada por DEAE-dextrano, la transfección mediada por liposomas tales como usando lipofectamina (Gibco, MD, EE.UU.) y/o cellfectin (Gibco, MD, EE.UU.), la absorción de ADN mediada por PEG, la electroporación y el bombardeo de micropartículas tales como mediante el uso de partículas de tungsteno o de oro recubiertas de ADN (Agracetus Inc., WI, EE.UU.) entre otras.

Las células hospedadoras usadas para producir la proteína desvelada en el presente documento se pueden cultivar en una variedad de medios, en función del tipo celular usado. Los medios comercialmente disponibles tales como F10 de Ham (Sigma), medio mínimo esencial ((MEM), (Sigma), RPMI-1640 (Sigma), medio Eagle modificado de Dulbecco ((DMEM), Sigma) son adecuados para cultivar las células de mamífero. Los medios para cultivar otros tipos celulares tratados en el presente documento son conocidos en la materia.

#### Secuencias de armazón:

Hasta la fecha, las secuencias de las regiones estructurales de intracuerpos (moléculas de anticuerpo cuyas secuencias se han diseñado genéticamente o han evolucionado hacia una mayor estabilidad de manera que se puedan plegar de manera productiva en el citoplasma), es decir, las secuencias no de CDR, han diferido sustancialmente de la secuencia genómica de la línea germinal afín. Tal como se desvela en el presente documento, los inventores han seleccionado sistemáticamente y han determinado secuencias de regiones variables de anticuerpos que son idénticas o están estrechamente relacionadas con, la secuencia genómica de la línea germinal afín, y que permiten el correcto plegamiento y una elevada estabilidad cuando se expresan en ambientes no oxidantes. Las secuencias preferidas para su uso en la presente invención se describen a continuación. Para cualquiera de las secuencias de región variable descritas en el presente documento, se apreciará que el experto en la materia será capaz de identificar las CDR (por ejemplo, muchas de las cuales se identifican en la base de datos del NCBI) y la región de armazón restante. Los ejemplos particulares de CDR en cada una de las regiones variables descritas en el presente documento se muestran en la Figura 7. El polipéptido de la presente divulgación comprende una región variable de cadena pesada ( $V_H$ ) de la familia  $V_H3$  de dominios variables de inmunoglobulina. Preferentemente, el  $V_H$  es IGHV3-23 (SEQ ID NO: 3).

IGHV3-23, también conocido como DP-47, pertenece a la familia  $V_H3$  de dominios variables de Ig humana. La familia  $V_H3$  tiene el 43 % (22/51) de los miembros funcionales de los genes de  $V_H$  y se ha citado IGHV3-23 como el gen más altamente expresado en el repertorio de  $V_H$  (Stewart et al., 1993). También se halla a alta frecuencia en reordenamientos de Ig productivos en linfocitos B (Brezinschek et al., 1997). Debido a su alta frecuencia en

repertorios de Ig naturales, también se ha aislado con frecuencia de las bibliotecas de expresión en fagos de regiones V humanas (Griffiths et al., 1994). También se ha usado como un compañero de armazón en bibliotecas sintéticas (Jirholt et al., 1998; Pini et al., 1998; Soderlind et al., 2000; Ge et al., 2010). IGHV3-23 se seleccionó como el compañero de región variable de cadena pesada en el estudio de los presentes inventores para identificar un armazón de región variable de anticuerpo soluble, estable.

Preferentemente, el polipéptido desvelado en el presente documento comprende una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3. Por ejemplo, la región de armazón puede ser al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3. Preferentemente, el polipéptido desvelado en el presente documento comprende una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 96 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3. la región de armazón comprende todos los restos de la región variable excluyendo los restos de CDR. Por lo tanto, el polipéptido desvelado en el presente documento puede comprender una región de armazón que comprende los aminoácidos 1 - 25, 33 - 51, 60 - 98 de la SEQ ID NO: 3 (o una región de armazón cuya secuencia de aminoácidos es, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la secuencia de estos aminoácidos). Como alternativa, el polipéptido desvelado en el presente documento puede comprender una región de armazón que comprende los aminoácidos 1 - 25 y 33 - 98 de la SEQ ID NO: 3 (o una región de armazón cuya secuencia de aminoácidos es al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la secuencia de estos aminoácidos). En otra alternativa, el polipéptido desvelado en el presente documento puede comprender una región de armazón que comprende los aminoácidos 1 - 51 y 60 - 98 de la SEQ ID NO: 3 (o una región de armazón cuya secuencia de aminoácidos es al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la secuencia de estos aminoácidos).

Los polipéptidos desvelados en el presente documento pueden comprender cualquier secuencia o secuencias de CDR. Por lo tanto, los polipéptidos desvelados en el presente documento pueden comprender las secuencias de CDR de IGHV3-23 (es decir, los aminoácidos 26 - 32, 52 - 59 de la SEQ ID NO: 3). Como alternativa, los polipéptidos desvelados en el presente documento pueden comprender cualquier otra secuencia o secuencias de CDR. Por lo tanto, la región de armazón del dominio variable de  $V_H3$  puede servir como un molde en el que se puede insertar cualquiera de las secuencias de CDR dadas. Las secuencias de CDR se pueden generar de manera aleatoria. Como alternativa o además, las secuencias de CDR se pueden generar de manera semialeatoria (asignando de manera aleatoria a cada posición de aminoácido particular en la CDR un resto de aminoácido seleccionado de un subconjunto de todos los aminoácidos posibles, sabiendo que el subconjunto es necesario o particularmente favorecido en una posición de aminoácido dada en la CDR).

Como alternativa, las secuencias de CDR pueden derivar de otro anticuerpo. Por lo tanto, las CDR de, por ejemplo, un anticuerpo humano se pueden insertar en el armazón del dominio variable de  $V_H3$ . Se apreciará que el experto en la materia pueda usar diversos métodos para asegurar que una región de armazón tal como se define en el presente documento comprende una o más secuencias de CDR tomadas de un anticuerpo humano. Preferentemente, tales métodos incluyen clonar una o más secuencias codificantes de CDR en un polinucleótido que codifica un polipéptido desvelado en el presente documento, tal como se describe con más detalle en el presente documento, a continuación. Las secuencias codificantes de CDR pueden variar de manera adicional mediante mutagénesis dirigida o aleatorizada, con el fin de proporcionar una pluralidad de polipéptidos que comprenden una pluralidad de diferentes secuencias de CDR. Tales métodos se pueden aplicar a una cualquiera o a una combinación de CDR1, CDR2 y CDR3.

Las secuencias de una CDR cualquiera o más, CDR1, CDR2 y CDR3 se pueden introducir en los armazones del dominio variable descritos en el presente documento, en cualquier combinación. Preferentemente, al menos la secuencia de una CDR3 se introduce en el armazón del dominio variable de  $V_H3$  descrito en el presente documento.

Además, la longitud de las secuencias de CDR introducidas en el armazón del dominio variable de  $V_H3$  descrito en el presente documento puede variar. Por ejemplo, se puede insertar una secuencia de CDR3 de 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 aminoácidos en el armazón. Los presentes inventores han descubierto que una secuencia de CDR3 acortada de menos de 12, tal como 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1, y lo más preferentemente, de 7 aminoácidos de longitud presenta una estabilidad mejorada.

#### *Compañeros de cadena ligera de IGL e IGL para IGHV3-23*

Los inventores de la presente divulgación, aplicando solo el criterio de que la fusión de scFv es soluble en su plataforma RED (al contrario de las selecciones funcionales realizadas anteriormente que requieren la unión del anticuerpo a un antígeno diana *in vivo* y que, por lo tanto, seleccionaban anticuerpos que estaban sustancialmente mutados a partir de sus respectivas secuencias de la línea germinal) fueron capaces de seleccionar cadenas ligeras no expuestas anteriormente que no se habían mutado en la región V. Por lo tanto, fueron capaces de identificar secuencias de la línea germinal que conferían solubilidad tras las fusiones del scFv IGHV3-23. Esto tiene el beneficio

significativo de asegurar que una biblioteca de armazón artificial construida de los dominios  $V_L$  y  $V_H$  es idéntica en secuencia a las abundantes proteínas de anticuerpos humanos, minimizando así el reconocimiento inmunológico y el rechazo en tiempo prolongado a cualquier derivado.

- 5 Por consiguiente, el polipéptido desvelado en el presente documento preferentemente comprende una región variable de cadena ligera de anticuerpo ( $V_L$ ) de las familias  $V_{L\lambda 1}$ , 3 o 6 de dominios variables de inmunoglobulinas combinado con la secuencia de la línea germinal IGHV3-23. Los miembros preferidos de la familia  $V_{L\lambda 1}$ , 3 o 6 incluyen IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), y IGLV6-57 (tal como se establece en la SEQ ID NO: 12).  
 10 Por lo tanto, el polipéptido desvelado en el presente documento preferentemente comprende una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), y IGLV6-57 (tal como se establece en la SEQ ID NO: 12). Por ejemplo, la región de armazón puede ser al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la región de armazón de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), y IGLV6-57 (tal como se establece en la SEQ ID NO: 12).

- Lo más preferentemente, el compañero de  $V_L$  de IGHV3-23 es IGLV3-1 (SEQ ID NO: 6). Por lo tanto, el polipéptido desvelado en el presente documento preferentemente comprende una región variable de cadena ligera de anticuerpos ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGLV3-1 (SEQ ID NO: 6). Por ejemplo, la región de armazón puede ser al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la región de armazón de IGLV3-1 (SEQ ID NO: 6). La región de armazón de IGLV3-1 puede comprender los aminoácidos 1 - 23, 32 - 48, 56 - 89 de la SEQ ID NO: 6. Por consiguiente, el polipéptido desvelado en el presente documento puede comprender una región variable de cadena ligera del anticuerpo ( $V_L$ ) que comprende una región de armazón que comprende los aminoácidos 1 - 23, 32 - 48, 56 - 89 de la SEQ ID NO: 6 (o una región de armazón cuya secuencia de aminoácidos es al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 % o el 100 % idéntica a la secuencia de estos aminoácidos).

- Un ejemplo preferido de una secuencia de polinucleótidos que codifica un armazón IGLV3-1::IGHV3-23 con regiones variables de CDR3, y la correspondiente secuencia de aminoácidos traducida se representa en la Figura 8.

- Preferentemente, el polipéptido desvelado en el presente documento comprende la región de armazón del dominio variable  $V_L$  (por ejemplo, el polipéptido desvelado en el presente documento puede comprender la región de armazón de IGHV3-1, por ejemplo, los aminoácidos 1 - 23, 32 - 48, 56 - 89 de la SEQ ID NO: 6). De nuevo, los polipéptidos desvelados en el presente documento pueden comprender cualquier secuencia o secuencias de CDR. Por lo tanto, los polipéptidos desvelados en el presente documento pueden las secuencias de CDR de cualquiera de IGLV1-40, IGLV1-44, IGLV1-47, IGLV3-1, IGLV3-19 y/o IGLV6-57. Como alternativa, los polipéptidos desvelados en el presente documento pueden comprender cualquier otra secuencia o secuencias de CDR. Por lo tanto, la región de armazón del dominio variable  $V_L$  puede servir como un molde en el que se puede insertar cualquiera de las secuencias de CDR dadas, tal como se describe anteriormente con respecto al armazón del dominio variable de  $V_H3$ . Por lo tanto, las secuencias de CDR se pueden generar de manera aleatoria. Como alternativa o además, las secuencias de CDR se pueden generar de manera semialeatoria (asignando aleatoriamente un resto de aminoácido seleccionado de un subconjunto de todos los aminoácidos posibles para cada posición particular de aminoácidos en la CDR, sabiendo que el subconjunto es particularmente favorecido en una posición de aminoácido dada en la CDR).

- Como alternativa, las secuencias de CDR pueden derivar de otro anticuerpo. Por lo tanto, las CDR de, por ejemplo, un anticuerpo humano se pueden insertar en el armazón del dominio variable de  $V_L$ . Se apreciará que hay diversos métodos diferentes disponibles para el experto en la materia para asegurar que una región de armazón tal como la que se define en el presente documento comprende una o más secuencias de CDR tomadas de un anticuerpo humano. Además, las secuencias de CDR de un anticuerpo humano se pueden mutagenizar aleatoriamente antes de la inserción en un armazón de dominio variable  $V_L$  descrito en el presente documento.

- Una cualquiera o más de las secuencias de CDR1, CDR2 y CDR3 se pueden insertar en el armazón del dominio variable  $V_L$  descrito en el presente documento, en cualquier combinación. Preferentemente, al menos la secuencia de una CDR3 se inserta en el armazón del dominio variable  $V_L$  descrito en el presente documento.

- Además, la longitud de las secuencias de CDR insertadas en el armazón del dominio variable  $V_L$  descrito en el presente documento puede variar. Por ejemplo, se puede insertar una secuencia de CDR3 de 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 aminoácidos en el armazón. Los presentes inventores han descubierto que una secuencia de CDR3 acortada de menos de 12, tal como 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1, y lo más preferentemente, de 7 aminoácidos de longitud presenta una estabilidad mejorada.

*Bibliotecas de polipéptidos sintéticos*

Una biblioteca de secuencias de polipéptidos se puede clonar y expresar en una variedad de plataformas de presentación de proteínas para seleccionar proteínas de afinidad frente a una diana deseada. Por lo tanto, la divulgación proporciona una biblioteca que comprende una pluralidad de polipéptidos desvelados en el presente documento. En una realización preferida, las bibliotecas se pueden preparar identificando un polipéptido y/o secuencia de polinucleótidos "parental" y alterando esa secuencia para crear una pluralidad de secuencias variantes para formar la biblioteca. La alteración se puede realizar mediante cualquiera de los medios adecuados, por ejemplo, mutagénesis de sitio dirigido o mutagénesis aleatorizada. Los métodos adecuados de la construcción de la biblioteca se conocen en la materia.

Tal como se ha indicado anteriormente, los dominios variables se pueden clonar directamente a partir de una fuente biológica, de manera que tanto las secuencias estructurales como las CDR estén presentes tal como se forman mediante recombinación V(D)J. Como alternativa, la biblioteca de anticuerpos puede ser parcial o totalmente sintética, con la CDR y las regiones estructurales ensambladas *de novo*. Por ejemplo, un único armazón artificial que representa un emparejamiento de genes de anticuerpos comúnmente expresados o particularmente estables podría recodificarse para una expresión optimizada en un organismo hospedador. Incluso todo el armazón y las CDR se pueden ensamblar en una única reacción usando oligonucleótidos solapantes, tales como los descritos por Ge et al. (2010).

Los métodos para generar diversidad en las CDR de unión a antígeno se han descrito por completo en la técnica anterior. Éstos incluyen - generar las CDR de ARNm, de células inmunológicas no expuestas anteriormente o preinmunizadas; diseñar y sintetizar las CDR mediante el análisis de las secuencias de anticuerpo clasificadas; diseñar y sintetizar las CDR con una distribución ponderada de aminoácidos basada en secuencias de anticuerpos clasificadas; adoptar una región de CDR aleatorizada y no sesgada.

Cada dominio de Ig,  $V_L$  y  $V_H$ , tiene tres regiones CDR, CDR1, CDR2 y CDR3, que son de longitud variable y tienen diferentes frecuencias de contactos entre superficies con el antígeno. La CDR más variante *in vivo* tanto para  $V_L$  como para  $V_H$  es CDR3, cuyo bucle se forma mediante recombinación entre la unión del exón de los dominios V-J ( $V_L$ ) o los dominios V-D-J ( $V_H$ ). Esto es representativo del sistema inmunológico no expuesto anteriormente. Sin embargo, una vez que un linfocito B se ha estimulado para la expansión, entonces actúa a menudo la hipermutación somática para diversificar también las CDR 1 y 2.

Sin embargo, para una biblioteca clonada de scFv de dominios variables construidos en un solo, o unos pocos, armazones de  $V_L$  y  $V_H$ , es común que la diversidad de CDR se limite a CDR3, con la composición de aminoácidos y la variación de la longitud del bucle considerando la unión a la diana.

En el caso de los polipéptidos estables descritos en la divulgación, esto permite que la totalidad del scFv, distinto de la región de bucle de CDR3 que varía de forma natural, sea idéntica, o casi igual, a la secuencia de la línea germinal de los genes de anticuerpos afines. Esto permite la selección sistemática de proteínas de afinidad que se parecen mucho al repertorio de anticuerpos humanos no expuestos anteriormente, minimizando de este modo la divergencia de la secuencia de un armazón diseñado genéticamente que podría desencadenar el reconocimiento inmunológico del paciente. Por lo tanto, en la presente invención, una biblioteca de polipéptidos puede comprender polipéptidos que difieren entre sí solo en la secuencia de CDR3.

Se puede construir una biblioteca de anticuerpos artificial usando un único armazón, o se puede constituir mediante una pluralidad de armazones. Por lo tanto, la biblioteca de la presente divulgación puede comprender uno o más polipéptidos que tiene una combinación particular de armazones de región variable de cadena pesada y ligera desvelados en el presente documento y uno o más polipéptidos que tienen otra combinación diferente de armazones de región variable de cadena pesada y ligera desvelados en el presente documento.

Por ejemplo, la biblioteca puede comprender:

- uno o más polipéptidos que comprenden una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3, y una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGLV3-1 (tal como se establece en la SEQ ID NO: 6), en donde el  $V_H$  y el  $V_L$  son capaces de formar un sitio de unión a antígeno; y
- uno o más polipéptidos que comprenden una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3, y una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de una cualquiera o más de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), o IGLV6-57 (tal como se establece en la SEQ ID NO: 12), en donde el  $V_H$  y el  $V_L$  son capaces de formar un sitio de unión a antígeno.



Por lo tanto, la biblioteca de la presente divulgación puede comprender uno o más polipéptidos que tienen cualquier combinación de armazones de región variable de cadena pesada y ligera desvelados en el presente documento.

Una pluralidad de armazones puede representar las dos amplias clases de genes de Ig humanas, es decir, una cadena pesada que se empareja con las clases de cadena ligera  $\kappa$  y  $\lambda$ , o pueden ser una mezcla de un solo armazón de cadena ligera con múltiples cadenas pesadas, o una mezcla de armazón de cadena ligera y una sola cadena pesada. También se podría extraer una pluralidad de armazones de un único miembro que sea el representante más estable de las diferentes subclases, o podría ser una combinación de los armazones más estables, pertenecientes a cualquier clase.

En el caso desvelado en el presente documento, una biblioteca de polipéptidos preferentemente está compuesta de regiones de armazón que son idénticas o casi idénticas (por ejemplo, al menos el 95, 96, 97, 98 o 99 % idénticas), a la región de armazón del gen IGHV3-23 humano, unido de manera operativa a una secuencia que es idéntica o casi idéntica (por ejemplo, al menos el 95, 96, 97, 98 o 99 % idénticas), a la región de armazón de los genes humanos para IGLV1-40, IGLV1-44, IGLV1-47, IGLV3-1, IGLV3-19 o IGLV6-53. La biblioteca puede constituir un único emparejamiento de armazón de un gen de cadena pesada y ligera en un formato reconocido (por ejemplo scFv) o puede ser un emparejamiento de armazón del gen IGHV3-23 con uno o más de los genes de cadena ligera mencionados anteriormente. Los inventores han demostrado que una biblioteca construida a partir de estos armazones tiene unas propiedades de estabilidad y solubilidad superiores y deseables en el citoplasma de *E. coli*. En efecto, es una biblioteca de intracuerpo superior cuya variación de la homología de secuencia ideal a sus genes de línea germinal afines existe solo en las regiones de bucle de CDR.

El experto en la materia reconocería que el método para la selección sistemática de polipéptidos solubles en citoplasma que tienen compañeros de VL para el gen de VH, IGHV3-23, también podrían aplicarse para la selección sistemática de compañeros solubles de otras regiones variables, bien VL o VH. Además, el método de selección sistemática de polipéptidos solubles en citoplasma podría repetirse usando variantes de un único armazón para encontrar mutaciones que aumentan su estabilidad y su solubilidad. Por ejemplo, cualquiera de los pares de armazón que se han identificado (IGHV3-23 con; IGLV1-40, IGLV1-44, IGLV1-47, IGLV3-1, IGLV3-19 o IGLV6-57) se podría usar como el molde para una biblioteca adicional de variantes en un único armazón con la intención de llevar a cabo la selección sistemática citoplasmática a una temperatura en la que el armazón parental tendría mala solubilidad.

Los expertos en la materia también reconocerían que una biblioteca de polipéptidos también puede estar constituida por los polipéptidos mencionados anteriormente presentes a menos del 100 % de abundancia. Una biblioteca compuesta del 50 % de los polipéptidos descritos en la divulgación; o del 25 % de los polipéptidos descritos en la divulgación; o del 10 % de los polipéptidos descritos en la divulgación, aún serviría para producir proteínas de afinidad de propiedades de estabilidad deseadas. Por lo tanto, la biblioteca de polipéptidos desvelada en el presente documento puede comprender polipéptidos que no sean los de la presente divulgación.

Además, aunque los inventores han descubierto sorprendentemente que un armazón de scFv que es idéntico o casi idéntico en la secuencia a los genes de la línea germinal para los dominios VH y VL descritos tiene unas propiedades de estabilidad y solubilidad superiores y deseables en el citoplasma de *E. coli*, los expertos en la materia reconocerían que estas secuencias podrían obtenerse para ser más polimórficas de lo que se ha informado, y aún funcionar como proteínas de afinidad con propiedades de estabilidad deseadas. Por lo tanto, la presente invención proporciona una región de armazón con una secuencia que diverge de las secuencias de la región del armazón desveladas en el presente documento en hasta el 10 % o el 5 %, y que aún sirve para producir proteínas de afinidad con propiedades de solubilidad y/o de estabilidad deseadas. Por lo tanto, los polipéptidos desvelados en el presente documento pueden comprender secuencias de la región del armazón que son al menos el 95 %, el 96 %, el 97 %, el 98 % o el 99 % idénticas a cualquiera de las secuencias de la región del armazón desveladas en el presente documento.

La presente invención proporciona tanto una biblioteca de polipéptidos como una biblioteca de polinucleótidos (por ejemplo, una biblioteca de ADN). Las bibliotecas de ADN son una colección de vectores recombinantes que contienen inserciones de ADN (fragmentos de ADN) que codifican el polipéptido. El origen de las inserciones de ADN puede ser genómico, ADNc, sintético o semisintético.

La clonación y construcción de bibliotecas de ADN que codifican polipéptidos desvelados en el presente documento se puede realizar usando métodos conocidos en la materia. Por ejemplo, Lutz y Patrick (2004) han revisado métodos de generar variabilidad de la biblioteca y estrategias para la recombinación génica para su uso en el diseño genético de proteínas. Para la selección sistemática de las variantes de polipéptido presentadas, se podrían adoptar las estrategias usadas para las bibliotecas de presentación en superficie y adaptarlas para los métodos de la presente divulgación (Becker et al., 2004; Kenrick et al., 2007; Miller et al., 2006; Daugherty et al., 2000).

Se puede introducir una biblioteca de ácidos nucleicos en una pluralidad de células bacterianas dando como resultado la expresión de un miembro de la biblioteca en cada una de las células bacterianas. Además de expresarse, los polipéptidos se mantienen en la célula bacteriana permeabilizada, o se unen a la pared celular, con

el fin de evaluar su función o característica. Las bibliotecas de ácido nucleico de un polipéptido, por ejemplo, se pueden generar mediante una variedad de métodos, incluyendo mediante la introducción de mutaciones tales como mutaciones puntuales, deleciones e inserciones, o mediante fenómenos de recombinación. Los métodos para la generación de bibliotecas de variantes se conocen en la materia e incluyen la PCR propensa a errores, la síntesis de ADN en bacterias comprometidas en la reparación del ADN, y la modificación química del ADN. Los métodos para la generación de bibliotecas mediante recombinación se conocen en la materia e incluyen el reordenamiento de genes, el ensamblaje de ADN en bacterias altamente recombinogénicas, el ensamblaje de la biblioteca de ácido nucleico sintético, etc., o cualquier combinación de los mismos. De este modo se puede introducir una biblioteca de polinucleótidos que codifica polipéptidos en una pluralidad de células bacterianas dando como resultado la expresión de uno de los miembros de la biblioteca en cada una de las células bacterianas.

En algunas realizaciones, una biblioteca comprende dos o más variantes de un polipéptido en donde cada una de las variantes comprende un único polipéptido con un cambio menor en la secuencia de aminoácidos, por ejemplo, en una secuencia de CDR. Una biblioteca puede tener al menos 2, al menos 5, al menos 10, al menos 50, al menos 100, al menos 1000, al menos 10.000, al menos 100.000, al menos 1.000.000, al menos  $10^7$  o más miembros.

#### Método de selección sistemática

Las bibliotecas de polipéptidos (o de anticuerpos) desveladas en el presente documento se pueden usar para seleccionar sistemáticamente un polipéptido que se une a una molécula diana. Se apreciará que el polipéptido se puede seleccionar o explorar en el contexto de una biblioteca de células que expresan cada una un polipéptido o variante de polipéptido diferente, o en el contexto de un único tipo de célula que expresa un único polipéptido. El término "molécula diana" se refiere a una molécula que se une a y/o que se modifica mediante el polipéptido y puede ser, por ejemplo, un antígeno, una enzima, un anticuerpo, un receptor, etc. Por tanto, "molécula diana" se puede usar para referirse a un sustrato tal como un sustrato enzimático o una molécula que está siendo evaluada para la unión (por ejemplo, un ligando, epítopo, antígeno, compañero de multimerización tal como un compañero homodimérico o heterodimérico, etc., o cualquier combinación de los mismos).

Por lo tanto, la divulgación proporciona un método de selección sistemática de un polipéptido que se une a un polipéptido diana, comprendiendo el método poner en contacto un polipéptido desvelado en el presente documento con el polipéptido diana, y determinar si el polipéptido desvelado en el presente documento se une al polipéptido diana. Preferentemente, en tales métodos se usa una pluralidad de polipéptidos desvelados en el presente documento.

Una serie de métodos adecuados de selección sistemática son conocidos en la materia, que se pueden usar de acuerdo con la presente invención.

Por ejemplo, el método puede comprender un método de presentación de proteínas. El método más antiguo de presentación de proteínas es la presentación en fago (Smith, 1985), en el que la proteína de interés se fusiona a una de las proteínas de recubrimiento externo del fago en donde puede estar presente junto con copias de tipo silvestre de la proteína. Por ejemplo, se puede usar una plataforma de presentación basada en el fago filamentoso M13 que usa fusiones a la proteína pIII.

Otros métodos de presentación adecuados incluyen los métodos de presentación '*in vitro*' en los que el polipéptido se expresa usando un extracto de traducción celular, y el acoplamiento entre el polipéptido y el ácido nucleico codificante se logra mediante la unión física (por ejemplo, presentación en ribosoma, presentación en ARNm) o mediante la unión a un armazón común o la encapsulación dentro de una membrana, tal como la compartimentalización *in vitro* (IVC, del inglés *in vitro compartmentalization*) en donde el ARNm se traduce en una suspensión de micela que también puede incluir un sistema de captura de microperlas (magnéticas o de sefarosa) tanto para el ARNm como para la proteína.

Otro método adecuado de presentación del polipéptido es la presentación en superficie microbiana, que implica la localización dirigida de los polipéptidos al exterior de una pared microbiana, bien de eubacterias gram negativas, o gram positivas o de levaduras. Los polipéptidos se fusionan a dominios de anclaje que los unen a la superficie celular. Los dominios de anclaje pueden tener motivos que dictan lipidación o unión covalente a la pared celular, o pueden ser una fusión a una proteína integral de membrana en una región de bucle expuesta.

La presente solicitud demuestra que los polipéptidos y los polinucleótidos desvelados en el presente documento son particularmente eficaces cuando se usan en métodos de selección sistemática que comprenden sistemas de expresión sin células. El uso de los polipéptidos y de los polinucleótidos de la presente divulgación en tales sistemas de expresión acelera enormemente el proceso de selección sistemática de polipéptidos que demuestran alta expresión, alta solubilidad y alta afinidad por un polipéptido diana. Las ventajas resultan de la alta solubilidad, estabilidad y expresión demostradas por los polipéptidos de la presente divulgación, en partícula en condiciones reductoras.

Por lo tanto, los polipéptidos y los polinucleótidos de la presente divulgación son particularmente adecuados para su

uso en un método de selección sistemática que comprende cualquiera de los sistemas de expresión *in vitro* sin células descritos en el presente documento y otros conocidos en la materia. Por ejemplo, los polipéptidos y los polinucleótidos de la presente divulgación son particularmente adecuados para su uso en un método de selección sistemática que comprende la presentación en ribosoma, la presentación en ARNm, la presentación en cis (en donde un polipéptido expresado permanece conjugado con su secuencia de polinucleótido codificante) u otro de tales métodos conocidos en la materia.

Además, la presente solicitud demuestra que los polipéptidos y los polinucleótidos desvelados en el presente documento son particularmente eficaces cuando se usan en métodos de selección sistemática basados en métodos de presentación de proteínas. Por ejemplo, los polipéptidos y los polinucleótidos desvelados en el presente documento son particularmente eficaces cuando se usan en métodos de selección sistemática que comprenden la presentación en fago (por ejemplo, el fago lambda lítico, el fago M13 filamentoso, el fago de lisis defectuosa y otros conocidos en la materia). En un ejemplo, los polipéptidos y los polinucleótidos desvelados en el presente documento son sorprendentemente eficaces cuando se usan en un método de selección sistemática que comprende el fago lambda.

Además, los polipéptidos y los polinucleótidos de la presente divulgación se pueden usar en un método de selección sistemática basado en un fago de lisis defectuosa (por ejemplo, tal como se describe en la Solicitud de Patente Internacional publicada como WO 2013/000023) en combinación con el sistema RED descrito en el presente documento y en el documento WO 2011/075761.

#### Kits

Los componentes necesarios para realizar los métodos desvelados en el presente documento se pueden proporcionar de manera conveniente en forma de un kit. Tal como entenderá un experto en la materia, los diversos componentes en el kit se pueden suministrar en recipientes individuales o alícuotas, o los componentes de la solución se pueden combinar en diferentes combinaciones y en diferentes concentraciones para lograr un rendimiento óptimo de los métodos desvelados en el presente documento. Está dentro de los conocimientos del destinatario experto en la materia determinar qué componentes del kit se pueden combinar de manera que los componentes se mantengan de forma estable antes de su uso.

Los kits desvelados en el presente documento típicamente contendrán como mínimo un vector que comprende un sitio para insertar en el vector un polinucleótido que codifica un polipéptido desvelado en el presente documento, y un marco de lectura abierta que codifica un segundo polipéptido que se asocia con el primer polipéptido para formar un complejo proteico que se puede mantener dentro o se puede unir a la pared celular de una célula bacteriana permeabilizada. Preferentemente, el kit también contiene un agente para permeabilizar una célula bacteriana. En una realización, el kit comprende adicionalmente células bacterianas, preferentemente, células bacterianas Gram negativas. Otros componentes adicionales pueden estar incluidos con el kit, u otros componentes se pueden suministrar por el usuario final, si fuera necesario.

#### Usos

Los polipéptidos de la presente divulgación son útiles en una variedad de aplicaciones, que incluyen la investigación, aplicaciones de diagnóstico y terapéuticas. En función del antígeno al que se une, el polipéptido puede ser útil para la administración de un componente a una célula, por ejemplo, para destruir una célula o evitar su crecimiento y/o para obtener imágenes y/o para ensayos *in vitro*. En un ejemplo, el polipéptido es útil tanto para la obtención de imágenes como para administrar un agente citotóxico a una células, es decir, se conjuga con un marcador detectable y un agente citotóxico o una composición comprende una mezcla de proteínas, algunas de las cuales se conjugan con un agente citotóxico, y algunas de las cuales se conjugan con un marcador detectable.

Los polipéptidos descritos en el presente documento también pueden actuar como inhibidores para inhibir (que puede ser la reducción o la prevención) (a) la unión (por ejemplo, de un ligando, de un inhibidor) a un receptor, (b) una función de señalización del receptor, y/o (c) una función estimuladora. Los polipéptidos que actúan como inhibidores de la función del receptor pueden bloquear la unión del ligando de forma directa o de forma indirecta (por ejemplo, provocando un cambio conformacional). Los polipéptidos descritos en el presente documento pueden ser particularmente adecuados para aplicaciones que implican una interacción de unión que tiene lugar dentro de una célula hospedadora, dada la estabilidad y el tamaño de los polipéptidos preferidos descritos en el presente documento.

#### Composiciones farmacéuticas y métodos de tratamiento

Los polipéptidos desvelados en el presente documento (sin. principios activos) son útiles para la administración parenteral, tópica, oral o local, la administración de aerosoles, o la administración transdérmica para tratamiento profiláctico o terapéutico. Las composiciones farmacéuticas se pueden administrar en una variedad de formas de dosificación unitaria en función del método de administración. Por ejemplo, las formas de dosificación unitaria adecuadas para la administración oral incluyen polvos, comprimidos, píldoras, cápsulas y pastillas para chupar o

mediante administración parenteral. Se reconoce que las composiciones farmacéuticas de la presente divulgación, cuando se administran por vía oral, se deberían de proteger de la digestión. Esto se logra típicamente bien complejando las proteínas con una composición para hacerlas resistente a la hidrólisis ácida y enzimática o mediante empaquetamiento del compuesto en un vehículo resistente de manera apropiada tal como un liposoma.

5 Los medios de protección de proteínas frente a la digestión se conocen en la materia.

Por lo general, se formulará una cantidad terapéuticamente eficaz del polipéptido en una composición para la administración a un sujeto. La frase "una cantidad terapéuticamente eficaz" se refiere a una cantidad suficiente como para promover, inducir y/o mejorar el tratamiento u otro efecto terapéutico en un sujeto. Como será evidente, la concentración de las proteínas de la presente divulgación en estas formulaciones puede variar ampliamente, y se seleccionará principalmente basándose en los volúmenes de líquido, en las viscosidades, en el peso corporal y similares, de acuerdo con el modo particular de administración seleccionado y con las necesidades del paciente. Dependiendo del tipo y de la gravedad de la enfermedad, una cantidad terapéuticamente eficaz puede ser de aproximadamente 1 µg/kg a 15 mg/kg (por ejemplo, 0,1-20 mg/kg) de molécula, ya sea, por ejemplo, mediante una o más administraciones separadas o mediante infusión continua. Una dosificación diaria típica puede variar desde aproximadamente 1 µg/kg a 100 mg/kg o más. Una dosificación ejemplar de la proteína a administrar al paciente está en el intervalo de aproximadamente 0,1 a aproximadamente 10 mg/kg de peso del paciente. Para administraciones repetidas a lo largo de varios días o un periodo mayor, dependiendo de la afección, se continuará el tratamiento hasta que suceda una supresión deseada de los síntomas de enfermedad. Un régimen de dosificación ejemplar comprende administrar una dosis de carga inicial de aproximadamente 4 mg/kg, seguido por una dosis de mantenimiento semanal de aproximadamente 2 mg/kg de la proteína. Otros regímenes de dosificación pueden ser útiles. El progreso de esta terapia se controla fácilmente mediante técnicas y ensayos convencionales.

Como alternativa, el polipéptido desvelado en el presente documento se formula a una dosis concentrada que está diluida a una dosis terapéuticamente eficaz antes de la administración a un sujeto.

Las composiciones farmacéuticas son particularmente útiles para la administración parenteral, por ejemplo, formuladas para inyección por vía intravenosa, intramuscular, subcutánea, transdérmica u otras de estas rutas, incluyendo la administración peristáltica y la instilación directa en un sitio de tumor o enfermedad (administración intracavitaria). Las composiciones para la administración comúnmente comprenderán una solución de las proteínas desveladas en el presente documento disuelta en un vehículo farmacéuticamente aceptable, preferentemente un transportador acuoso. Se puede usar una variedad de transportadores acuosos, por ejemplo, solución salina tamponada y similares. Otros vehículos ejemplares incluyen agua, solución salina, solución de Ringer, solución de dextrosa, y albúmina sérica humana al 5 %. Los vehículos no acuosos tales como aceites mezclados y oleato de etilo también se pueden usar. También se pueden usar liposomas como vehículos. Los vehículos pueden contener pequeñas cantidades de aditivos que mejoran la isotonicidad y la estabilidad química, por ejemplo, tampones y conservantes. Las composiciones pueden contener sustancias auxiliares farmacéuticamente aceptables tal como se requiere para aproximar las condiciones fisiológicas tal como el ajuste del pH y los agentes tamponantes, agentes de ajuste de la toxicidad, y similares, por ejemplo, acetato de sodio, cloruro de sodio, cloruro de potasio, cloruro de calcio, lactato de sodio y similares.

Las técnicas para preparar composiciones farmacéuticas son generalmente conocidas en la materia tal como se ejemplifica en "Pharmaceutical Sciences", de Remington, 16ª edición. Mack Publishing Company, 1980.

El documento WO2002/080967 describe composiciones y métodos para administrar composiciones en forma de aerosol que comprenden proteínas para el tratamiento de, por ejemplo, asma, que también son adecuadas para la administración de la proteína desvelada en el presente documento.

Las dosificaciones adecuadas de compuestos desvelados en el presente documento variarán dependiendo de la proteína específica, la afección a diagnosticar/tratar/prevenir y/o del sujeto que se está tratando. Está dentro de la capacidad de un médico experto el determinar una dosificación adecuada, por ejemplo, comenzando con una dosificación subóptima y modificando la dosis incrementalmente para determinar una dosificación óptima o útil. Como alternativa, para determinar una dosificación apropiada para el tratamiento/profilaxis, se usan datos de ensayos de cultivos celulares o de estudios en animales, en donde una dosis adecuada está en un intervalo de concentraciones en circulación que incluyen la DE50 del principio activo con poca o ninguna toxicidad. La dosificación puede variar dentro de este intervalo dependiendo de la forma de dosificación empleada y de la ruta de administración utilizada. Se puede estimar inicialmente una dosis terapéuticamente/profilácticamente eficaz a partir de ensayos de cultivo celular. Se puede formular una dosis en modelos animales para lograr un intervalo de concentración en plasma en circulación que incluye la CI50 (es decir, la concentración del compuesto que logra una inhibición semimáxima de los síntomas) tal como se determina en un cultivo celular. Dicha información se puede usar para determinar de manera más precisa las dosis útiles en seres humanos. Los niveles en plasma se pueden medir, por ejemplo, mediante cromatografía líquida de alto rendimiento.

Una proteína desvelada en el presente documento se puede combinar en una formulación de combinación farmacéutica o régimen de dosificación como terapia de combinación, con un segundo compuesto. El segundo compuesto de la formulación de combinación farmacéutica o régimen de dosificación preferentemente tiene

actividades complementarias para la proteína de la combinación de manera que no se afectan adversamente entre sí.

El segundo compuesto puede ser un agente quimioterapéutico, un agente citotóxico, una citocina, un agente inhibidor del crecimiento, un agente antihormonal y/o un cardioprotector. Dichas moléculas están presentes de manera conveniente en combinación en cantidades que son eficaces para el fin que se pretende. Una composición farmacéutica que contiene una proteína desvelada en el presente documento también puede tener una cantidad terapéuticamente eficaz de un agente quimioterapéutico tal como un inhibidor de la formación de tubulina, un inhibidor de la topoisomerasa o un aglutinante de ADN.

También se pueden usar cápsulas o composiciones farmacéuticas de "liberación lenta". Las formulaciones de liberación lenta generalmente se diseñan para dar un nivel de fármaco constante sobre un período prolongado y se pueden usar para administrar compuestos desvelados en el presente documento.

La presente divulgación también proporciona un método de tratamiento o de prevención de una afección en un sujeto, comprendiendo el método la administración de una cantidad terapéuticamente eficaz de una proteína desvelada en el presente documento a un sujeto que lo necesite.

Tal como se usa en el presente documento, los términos "previniendo", "prevenir" o "prevención" en el contexto de prevenir una afección incluye la administración de una cantidad de una proteína descrita en el presente documento suficiente como para detener o impedir el desarrollo de al menos un síntoma de una enfermedad o afección específica.

Tal como se usa en el presente documento, los términos "tratando", "tratar" o "tratamiento" incluye administrar una cantidad terapéuticamente eficaz de un(os) inhibidor(es) y/o agente(s) descritos en el presente documento suficiente como para reducir o eliminar al menos un síntoma de una enfermedad o afección especificada.

Tal como se usa en el presente documento, el término "sujeto" de debería de interpretar como cualquier animal, incluyendo los seres humanos, preferentemente un mamífero. Los sujetos ejemplares incluyen pero no se limitan a seres humanos, primates, ganado (por ejemplo, ovejas, vacas, caballos, burros, cerdos), animales de compañía (por ejemplo, perros, gatos), animales de ensayos de laboratorio (por ejemplo, ratones, conejos, ratas, cobayas o hámsteres), animales salvajes cautivos (por ejemplo, zorro, ciervo). Preferentemente el mamífero es un ser humano o un primate. Más preferentemente el mamífero es un ser humano.

Tal como se usa en el presente documento, una "afección" es una alteración de o una interferencia con la función normal, y no se limita a ninguna afección específica, e incluirá enfermedades o trastornos. En un ejemplo, la afección es un cáncer o un trastorno inmunopatológico.

Los cánceres ejemplares incluyen, aunque sin limitación, carcinoma, linfoma, blastoma, sarcoma y leucemia o tumores malignos linfoides. Más ejemplos particulares de dichos cánceres incluyen cáncer de células escamosas (por ejemplo cáncer de células escamosas epitelial), cáncer de pulmón incluyendo cáncer de pulmón microcítico, cáncer de pulmón no microcítico, adenocarcinoma del pulmón y carcinoma escamoso de pulmón, cáncer de peritoneo, cáncer hepatocelular, cáncer gástrico o de estómago incluyendo cáncer gastrointestinal, cáncer pancreático, glioblastoma, cáncer de cuello uterino, cáncer de ovario, cáncer de hígado, cáncer de vejiga, hepatoma, cáncer de mama, cáncer de colon, cáncer rectal, cáncer colorrectal, carcinoma de endometrio o uterino, carcinoma de las glándulas salivales, cáncer de riñón o renal, cáncer de próstata, cáncer de vulva, cáncer de tiroides, carcinoma hepático, carcinoma anal, carcinoma del pene, así como cáncer de cabeza y cuello.

La inmunopatología es el estudio de enfermedades que tienen una causa inmunológica y la enfermedad inmunológica es cualquier afección causada por las reacciones de las inmunoglobulinas a los antígenos. Por lo tanto, un "trastorno inmunopatológico" se puede definir como un trastorno que surge de la reacción del sistema inmunológico de un sujeto a los antígenos. Los trastornos inmunopatológicos incluyen las enfermedades autoinmunitarias y las respuestas hipersensibles (por ejemplo, Tipo I: anafilaxis, urticaria, alergias alimenticias, asma; Tipo II: anemia hemolítica autoinmunitaria, reacciones de transfusión de sangre; Tipo III: enfermedad del suero, vasculitis necrosante, glomerulonefritis, artritis reumatoide, lupus; Tipo IV: dermatitis de contacto, rechazo al injerto). Las enfermedades autoinmunitarias incluyen trastornos reumatológicos (tales como, por ejemplo, artritis reumatoide, síndrome de Sjögren, esclerodermia, lupus, tales como LES y nefritis lúpica, polimiositis/dermatomiositis, crioglobulinemia, síndrome de anticuerpos antifosfolípidos y artropatía psoriásica), artrosis, trastornos gastrointestinales y del hígado autoinmunitarios (tales como, por ejemplo, enfermedades inflamatorias del intestino (por ejemplo, colitis ulcerosa y enfermedad de Crohn), gastritis autoinmunitaria y anemia perniciosa, hepatitis autoinmunitaria, cirrosis biliar primaria, colangitis esclerosante primaria y celiaquía), vasculitis (tales como, por ejemplo, vasculitis asociada a ANCA, incluyendo la vasculitis de Churg-Strauss, la granulomatosis de Wegener y la poliarteritis), trastornos neurológicos autoinmunitarios (tales como, por ejemplo, esclerosis múltiple, síndrome de opsoclonía-mioclónia, miastenia grave, neuromielitis óptica y polineuropatías autoinmunitarias), trastornos renales (tales como, por ejemplo, glomerulonefritis, síndrome de Goodpasture y la enfermedad de Berger), trastornos dermatológicos autoinmunitarios (tales como, por ejemplo, psoriasis, urticaria, urticaria, pénfigo

vulgar, pénfigo ampoloso y lupus eritematoso cutáneo), trastornos hematológicos (tales como, por ejemplo, púrpura trombocitopénica, púrpura trombocitopénica trombótica, púrpura postransfusional y anemia hemolítica autoinmunitaria), aterosclerosis, uveítis, enfermedades autoinmunitarias de la audición (tales como, por ejemplo, enfermedad del oído interno y pérdida de audición), enfermedad de Behcet, síndrome de Raynaud, trasplante de órganos y trastornos endocrinos autoinmunitarios (tales como, por ejemplo, trastornos autoinmunitarios relacionados con la diabetes tales como diabetes mellitus insulín dependiente (DMID), enfermedad de Addison y enfermedad tiroidea autoinmunitaria (por ejemplo, enfermedad de Grave y tiroiditis)). Las más preferidas de tales enfermedades incluyen, por ejemplo, artritis reumatoide, colitis ulcerosa, vasculitis asociada a ANCA, lupus, esclerosis múltiple, síndrome de Sjögren, enfermedad de Grave, DMID, anemia perniciosa, tiroiditis y glomerulonefritis. El trastorno puede ser una enfermedad inflamatoria. La inflamación es una respuesta protectora de los tejidos corporales a la irritación o al daño y puede ser aguda o crónica. Por lo tanto, los trastornos inflamatorios incluyen enfermedades que implican neutrófilos, monocitos, mastocitos, basófilos, eosinófilos, macrófagos en donde tiene lugar la liberación de citocinas, la liberación de histamina, el estallido oxidativo, la fagocitosis, la liberación de otras enzimas granulares y la quimiotaxis. Las respuestas hipersensibles (definidas anteriormente en trastornos inmunopatológicos) también se pueden considerar como enfermedades inflamatorias (agudas o crónicas) dado que a menudo implican la activación del complemento y el reclutamiento/infiltración de diversos leucocitos tales como neutrófilos, mastocitos, basófilos, etc.

Las composiciones desveladas en el presente documento se administrarán de una manera compatible con la formulación de la dosificación y en una cantidad tal como sea terapéuticamente/profilácticamente eficaz. Las formulaciones se administran fácilmente en una variedad de formas, por ejemplo, mediante ingesta o inyección o inhalación.

Se pueden combinar otros regímenes terapéuticos con la administración de un polipéptido desvelado en el presente documento. La terapia de combinación puede administrarse como un régimen simultáneo o secuencial. Cuando se administra secuencialmente, la combinación se puede administrar en dos o más administraciones. La administración combinada incluye la coadministración, usando formulaciones separadas o una única formulación farmacéutica, y la administración consecutiva en cualquier orden, en la que, preferentemente, hay un período de tiempo en el que ambos (o todos) los agentes activos ejercen sus actividades biológicas de forma simultánea.

A continuación se describirá en detalle la invención con referencia a los siguientes ejemplos no limitantes.

## Ejemplos

### Ejemplo 1. Clonación de sub-bibliotecas de VL humano en un vector de presentación de IGHV3-23

Para seleccionar sistemáticamente los compañeros de la cadena ligera humana para IGHV3-23 que estarían bien expresados y fuesen solubles en el citoplasma de *E. coli*, los inventores clonaron las familias de cadena ligera, 10  $\lambda$  y 5  $\kappa$ , como fusiones de scFv para IGHV3-23. La biblioteca de scFv se clonó en una construcción de expresión que era inducible con arabinosa y se fusionó adicionalmente a los dominios aguas abajo que conferían la unión a la pared celular (dominio de unión a peptidoglucano (PG)), un dominio reportero de la expresión (SNAP; New England Biolabs) y un dominio de unión a ADN (DBP), en ese orden. Estos dominios aguas abajo permiten la retención del resto de scFv cuando las membranas externa e interna de la célula hospedadora bacteriana se permeabilizan mediante detergente o disolventes orgánicos anclando la proteína de fusión tanto al ADN como a la pared celular.

Las familias de cadena ligera  $\lambda$  y  $\kappa$  se amplificaron a partir de ADNc preparado a partir de células mononucleares de sangre periférica humana (PBMC). Las subfamilias de la cadena ligera  $\kappa$  y  $\lambda$  se amplificaron en 7 y 11 reacciones de PCR, respectivamente. Cada sub-biblioteca se seleccionó sistemáticamente por separado para caracterizar de forma amplia el porcentaje de la biblioteca que contenía aparentemente los miembros solubles. Las sub-bibliotecas que contenían un porcentaje apreciable de miembros solubles (> 1 %) se seleccionaron de manera sistemática como clones individuales.

Los cebadores de oligonucleótidos se basaron en las secuencias descritas por Hust y Dubel (2010) con modificaciones en los extremos para la clonación mediante *BsmBI*. Se hicieron cambios adicionales a las secuencias de cebadores inversos que originariamente se diseñaron frente al dominio constante C1 de la cadena ligera. Se consideró que esto incluía una secuencia innecesaria, y se diseñaron cebadores degenerados frente a las regiones J para la cadena ligera. Las secuencias de oligonucleótidos para la amplificación de las regiones de la cadena ligera se enumeran en la Tabla 2.

**Tabla 2. Secuencias de oligonucleótidos para la amplificación de las regiones de cadena ligera.**

Primera ronda $\kappa$	Secuencia de oligonucleótidos
HVK1 F1	GAC ATC CAG ATG ACC CAG TCT CC (SEQ ID NO: 31)
HVK1 F2	GMC ATC CRG WTG ACC CAG TCT CC (SEQ ID NO: 32)
HVK2 F	GAT RTT GTG ATG ACY CAG WCT CC (SEQ ID NO: 33)
HVK3 F	GAA ATW GTG WTG ACR CAG TCT CC (SEQ ID NO: 34)

**Primera ronda κ**

HVK4 F	GAC ATC GTG ATG ACC CAG TCT CC (SEQ ID NO: 35)
HVK5 F	GAA ACG ACA CTC ACG CAG TCT CC (SEQ ID NO: 36)
HVK6 F	GAW RTT GTG MTG ACW CAG TCT CC (SEQ ID NO: 37)
HVKCL R	ACA CTC TCC CCT GTT GAA GCT CTT (SEQ ID NO: 38)

**Secuencia de oligonucleótidos****Primera ronda λ**

HVL1 F1	CAG TCT GTG CTG ACT CAG CCA CC (SEQ ID NO: 39)
HVL1 F2	CAG TCT GTG YTG ACG CAG CCG CC (SEQ ID NO: 40)
HVL2 F	CAG TCT GCC CTG ACT CAG CCT (SEQ ID NO: 41)
HVL3 F1	TCC TAT GWG CTG ACW CAG CCA CC (SEQ ID NO: 42)
HVL3 F2	TCT TCT GAG CTG ACT CAG GAC CC (SEQ ID NO: 43)
HVL4 F1	CTG CCT GTG CTG ACT CAG CCC (SEQ ID NO: 44)
HVL4 F2	CAG CYT GTG CTG ACT CAA TCR YC (SEQ ID NO: 45)
HVL5 F	CAG SCT GTG CTG ACT CAG CC (SEQ ID NO: 46)
HVL6 F	AAT TTT ATG CTG ACT CAG CCC CA (SEQ ID NO: 47)
HVL7/8 F	CAG RCT GTG GTG ACY CAG GAG CC (SEQ ID NO: 48)
HVL9/10 F	CAG SCW GKG CTG ACT CAG CCA CC (SEQ ID NO: 49)
01115 HVLCL R	TGA ACA TTC TGT AGG GGC CAC TG (SEQ ID NO: 50)
01116 HVLCL R2	TGA ACA TTC CGT AGG GGC AAC TG (SEQ ID NO: 51)

**Segunda ronda κ**

HVK1 2F1	ATCTAGAATG GGA GAC GGT GAC ATC CAG ATG ACC CAG TCT CC (SEQ ID NO: 52)
HVK1 2F2	ATCTAGAATG GGA GAC GGT GMC ATC CRG WTG ACC CAG TCT CC (SEQ ID NO: 53)
HVK2 2F	ATCTAGAATG GGA GAC GGT GAT RTT GTG ATG ACY CAG WCT CC (SEQ ID NO: 54)
HVK3 2F	ATCTAGAATG GGA GAC GGT GAA ATW GTG WTG ACR CAG TCT CC (SEQ ID NO: 55)
HVK4 2F	ATCTAGAATG GGAGAC GGT GAC ATC GTG ATG ACC CAG TCT CC (SEQ ID NO: 56)
HVK5 2F	ATCTAGAATG GGA GAC GGT GAA ACG ACA CTC ACG CAG TCT CC (SEQ ID NO: 57)
HVK6 2F	ATCTAGAATG GGA GAC GGT GAW RTT GTG MTG ACW CAG TCT CC (SEQ ID NO: 58)
HVKCL 2R	GATCAG GGT CTG AGA CGA TTT RAT HTC CAS YYK KGT CCC HBS GCC RAA VGT (SEQ ID NO: 59)

**Segunda ronda λ**

HVL1 2F1	ATCTAGAATG GGAGAC GGT CAG TCT GTG CTG ACT CAG CCACC (SEQ ID NO: 60)
HVL1 2F2	ATCTAGAATG GGA GAC GGT CAG TCT GTG YTG ACG CAG CCG CC (SEQ ID NO: 61)
HVL2 2F	ATCTAGAATG GGA GAC GGT CAG TCT GCC CTG ACT CAG CCT (SEQ ID NO: 62)
HVL3 2F1	ATCTAGAATG GGA GAC GGT TCC TAT GWG CTG ACW CAG CCA CC (SEQ ID NO: 63)
HVL3 2F2	ATCTAGAATG GGA GAC GGT TCT TCT GAG CTG ACT CAG GAC CC (SEQ ID NO: 64)
HVL4 2F1	ATCTAGAATG GGA GAC GGT CTG CCT GTG CTG ACT CAG CCC (SEQ ID NO: 65)
HVL4 2F2	ATCTAGAATG GGAGAC GGT CAG CYT GTG CTG ACT CAA TCR YC (SEQ ID NO: 66)
HVL5 2F	ATCTAGAATG GGA GAC GGT CAG SCT GTG CTG ACT CAG CC (SEQ ID NO: 67)

**Segunda ronda  $\lambda$** 

HVL6 2F	ATCTAGAATG GGA GAC GGT AAT TTT ATG CTG ACT CAG CCC CA (SEQ ID NO: 68)
HVL7/8 2F	ATCTAGAATG GGA GAC GGT CAG RCT GTG GTG ACY CAG GAG CC (SEQ ID NO: 69)
HVL9/10 2F	ATCTAGAATG GGA GAC GGT CAG SCW GKG CTG ACT CAG CCA CC (SEQ ID NO: 70)
HVLCL 2R	GATCAG GGT CTG AGA CGA RRY GRT SAS CTB SGT BCC HBY DCC RAA BAC (SEQ ID NO: 71)

Los genes VL para las sub-bibliotecas de la cadena ligera  $\lambda$  y  $\kappa$  se amplificaron en dos rondas de PCR usando la ADN polimerasa Vent (New England Biolabs). Cada sub-biblioteca se clonó por separado en el vector de presentación RED usando *BsmBI* (New England Biolabs). Se estimó que cada biblioteca producía aproximadamente 20 - 40.000 colonias.

**Ejemplo 2. Selección sistemática de fusiones de scFv humanos usando la presentación encapsulada retenida (RED)**

Como una selección sistemática inicial de solubilidad, cada placa de la biblioteca se raspó y se suspendió en 10 ml de LB/glicerol (al 10 %). Se cultivó una parte de la suspensión (~50  $\mu$ l) en 1 ml de medio LB (10 g de triptona, 5 g de extracto de levadura, 10 g de NaCl por litro) a 37 °C durante 1 hora y después se indujo con arabinosa al 0,2 % y se cultivó durante 2 horas adicionales a 25 °C. En este punto, las células se permeabilizaron mediante resuspensión del sedimento bacteriano en n-Octil- $\beta$ -D-Tioglucoído (8TGP) al 0,5 % en medio LB durante 10 minutos a 25 °C. Las células permeabilizadas se lavaron una vez mediante sedimentación y resuspensión en medio LB antes de que la proteína de fusión de scFv se marcara mediante la adición de reactivo SNAP-surface 488 (S9124S; New England Biolabs) e incubación a 25 °C durante 20 minutos. Las células marcadas se lavaron después de nuevo mediante sedimentación y se resuspendieron en PBS antes de que se montase una muestra para su visionado mediante microscopía de fluorescencia.

El examen a microscopio mostró que aunque todas las bibliotecas tenían algunas células en cada campo de visión que parecían bien expresadas y solubles, se descubrió que solo las sub-bibliotecas que representaban los clados VL1, VL3 y VL6 tenían más del 1 % de las células que tenían una morfología soluble. Por lo tanto, se consideró que todas las sub-bibliotecas, salvo las VL1, VL3 y VL6, no tenían una frecuencia de clones solubles lo suficientemente alta y no se seleccionaron más.

Las sub-bibliotecas VL1, VL3 y VL6 se colocaron en diluciones que producían clones únicos y se seleccionaron sistemáticamente para la solubilidad de manera individual. Por lo tanto, las sub-bibliotecas VL1, VL3 y VL6 se colocaron en placas a diluciones que permitían obtener picos limpios de colonias individuales, a las que se indujo después su expresión y se prepararon para microscopía, tal como se ha descrito anteriormente.

La Figura 1 demuestra la apariencia típica de un clon de scFv soluble, bien expresado (1A y recuadro), junto con un clon de scFv bien expresado pero insoluble (1B y recuadro). Las células se marcan con fluoróforo SNAP tras la permeabilización tal como se describe anteriormente. La agrupación distintiva de un clon insoluble se contrasta con la localización más difusa y periférica del clon soluble.

Los clones de scFv que demostraron una expresión soluble se cultivaron después durante toda una noche a 37 °C bajo la selección de 100  $\mu$ g/ml de ampicilina, una preparación de plásmido realizada usando métodos convencionales (Sambrook et al., Molecular Cloning: A Laboratory Manual, 3ª edición, Cold Spring Harbour Laboratory Press (2001) y el ADN se secuenció con un cebador en la región de promotor aguas arriba de la construcción de expresión. Después se analizaron los archivos de secuencias frente a la base de datos del genoma humano GenBank usando el algoritmo BLAST del NCBI.

La Figura 2 representa un alineamiento múltiple de los clones solubles seleccionados que tienen alta similitud o identidad total, con los genes de VL IGLV3-1, IGLV3-21 y IGLV6-57. El alineamiento múltiple se preparó usando CLUSTAL X. Las imágenes de los clones solubles (enumeradas por número de aislado) por encima de la alineación se corresponden con las secuencias alineadas abajo.

Los clones se comprobaron mediante aislamiento y secuenciación.

La selección sistemática de 779 miembros de sub-biblioteca para individuos solubles produjo 11 clones de IGLV1-40 (SEQ ID NO: 18); 2 clones de IGLV1-44 (SEQ ID NO: 21); 3 clones de IGLV1-47 (SEQ ID NO: 24); 3 clones de



IGLV1-51 (SEQ ID NO: 15); 25 clones de IGLV3-1 (SEQ ID NO: 6); 2 clones de IGLV3-19 (SEQ ID NO: 27); 4 clones de IGLV3-21 (SEQ ID NO: 9); 18 clones de IGLV6-57 (SEQ ID NO: 12). En análisis de las secuencias de los clones de scFv solubles demostró que aparentemente eran secuencias que no se habían expuesto anteriormente para los miembros IGLV1-40, IGLV1-51, IGLV3-1 e IGLV3-19 que no se habían seleccionado por afinidad o madurado *in vivo* durante la presentación de linfocitos B. Además, determinados clones de IGLV6-57 tenían alta solubilidad con solo 1 (2 clones) o 2 (1 clon) cambios de aminoácidos a partir de la traducción de la secuencia de la línea germinal, para una identidad total del 99 % y del 98 %, respectivamente.

Por lo tanto, al contrario de los resultados de la selección sistemática previa de anticuerpos solubles y estables humanos en el citoplasma de la levadura por Tse et al., (2002), que descubrieron anticuerpos solubles que comprendían un dominio de V<sub>H</sub>3 estaban emparejados por completo con compañeros VL<sub>k</sub> 1 y 4, los inventores descubrieron que las subfamilias VL<sub>k</sub> tenían una mala solubilidad aparente como clase, con más del 99 % de las sub-bibliotecas de clones emparejadas de manera específica con el dominio IGHV3-23 bien expresado mal en *E. coli* o que presenta signos de mal plegamiento. Tissot et al. (documento WO 03/097697) también llevaron a cabo una selección sistemática de Y2H para scFv humanos solubles, y documentaron que sus scFv solubles eran secuencias lo más estrechamente relacionadas con miembros de los clados VH1a, VH1b o VH3, combinados con secuencias lo más estrechamente relacionadas con miembros de los clados VL<sub>k</sub>1 o VL<sub>k</sub>1 o VL<sub>k</sub>3. Sin embargo, su configuración óptima fue VL<sub>k</sub>3 emparejado con VH1b.

Sin embargo, dado que tanto Tse et al. (2002) como Tissot (documento WO 03/097697) aplicaron una selección sistemática funcional (es decir, la unión de un anticuerpo a un antígeno diana *in vivo*) como un requisito adicional para la solubilidad, sus anticuerpos resultantes que tenían una señal de Y2H positiva requerían tanto 1) solubilidad; como 2) unión a la diana, y por lo tanto, necesariamente, se mutaron sustancialmente a partir de su secuencia de la línea germinal.

La mayoría de los miembros de VL aislados en la selección sistemática para fusiones solubles al dominio IGHV3-23 fueron IGLV3-1, también conocido como DPL23. Aunque algunos clones tenían numerosas mutaciones, un número significativo fueron idénticas a la secuencia V de la línea germinal de IGLV3-1 (SEQ ID NO: 4), lo que indica que la secuencia de la línea germinal es inherentemente estable y soluble en el citoplasma cuando se asocia con IGHV3-23.

IGLV3-1 tiene una expresión moderadamente alta en el sistema inmunológico humano, lo que representa el 15 % de las cadenas ligeras λ (Knappik et al., 2000) pero no es el miembro λ expresado de manera más abundante (DPL11). En las referencias publicadas no está caracterizado, carece de cualquier cita específica y no se han documentado estructuras con alta identidad. Aunque las bibliotecas de armazón de scFv artificial que usan IGHV3-23 se han creado anteriormente, los compañeros de VL se eligieron principalmente basándose en los niveles de expresión relativa *in vivo*, es decir, PK22 altamente expresado (Pini et al., 1998; y Ge et al., 2010), DPL3 (también conocido como IGLV1-47) (Kobayashi et al, 1997; Soderlind et al, 2000) y DPL16 (también conocido como IGLV3-19) (Viti et al., 2000).

El único informe publicado de un análisis global de la termoestabilidad del repertorio del dominio variable humano se realizó sobre la biblioteca Morfosys HuCAL<sup>™</sup> por Ewert et al. (2003). En su artículo titulado "Biophysical Properties of Human Antibody Variable Domains", los autores examinaron tanto la estabilidad de los dominios individuales, como la estabilidad de las parejas de dominios (V<sub>L</sub>:V<sub>H</sub>), cuando se expresan en el periplasma de *E. coli*.

El consenso de VH3, con el que se relaciona IGHV3-23, se declaró el más estable en estabilidad y solubilidad termodinámica de las regiones variables de cadena pesada, mientras que el V<sub>k</sub>3 consenso fue la región variable de cadena ligera más estable.

Las combinaciones de V<sub>H</sub>:V<sub>L</sub> que producían los emparejamientos más estables fueron las formadas entre H3::κ3, H1b::κ3, H5::κ3 y H3::κ1. Es destacable que ninguno de los emparejamientos más estables incluyó la familia V<sub>L</sub>3. Además, la construcción de la biblioteca de scFv de los presentes inventores usando un compañero de V<sub>H</sub>3 (IGHV3-23) no fue, en sí misma, suficiente como para conferir estabilidad en la proteína de fusión cuando se expresó en el citoplasma de *E. coli*, ya que la mayoría de los clones en la mayoría de las sub-bibliotecas estaban mal plegados o mal expresados.

En resumen, basándose en la técnica anterior, la combinación de uso de los dominios IGHV3-23 e IGLV3-1 como una fusión de scFv no podría predecirse como que poseen una estabilidad y solubilidad mejoradas cuando se expresan en un ambiente reductor, tal como el citoplasma de *E. coli*.

### Ejemplo 3. Ensayo de termoestabilidad de clones de scFv

Tras la selección sistemática inicial de las sub-bibliotecas de V<sub>L</sub> con inducción y expresión a 25 °C, los clones de scFv se sometieron a una selección sistemática adicional para clasificar los clones y las familias en función de la termoestabilidad.

Cada clon se indujo a temperaturas de 26 °C, 28 °C, 30 °C, 32 °C, 34 °C, 36 °C y 38 °C durante 90 minutos antes de la permeabilización y el marcaje con SNAP, tal como se describe para el Ejemplo 2. Se puntuó la solubilidad de los clones usando la microscopía de fluorescencia, tal como se describe para el Ejemplo 2.

- 5 La Figura 3 demuestra el comportamiento de dos clones, un IGLV3-1 y un IGLV3-21, con la expresión a temperaturas en aumento. Las proteínas de fusión de scFv permanecen solubles hasta al menos los 36 °C para el clon IGLV3-1, aunque el clon IGLV3-21 presenta signos de mal plegamiento entre los 32 y los 34 °C.

- 10 Este gradiente de temperatura de expresión demostró que los clones de scFv relacionados o idénticos a, IGLV3-1 e IGLV6-57 se juzgaron como que tienen la mejor solubilidad como una clase, aunque los clones individuales de los otros genes  $\lambda$  también demostraron niveles variables de solubilidad. La Figura 4 demuestra la solubilidad de los clones representativos de cada especie del dominio VL aislado de la selección.

- 15 Que la aparente solubilidad de los scFv por microscopía no era artificial, se confirmó subclonando el fragmento scFv, junto con el dominio 127 aguas abajo de titina de humano, en una construcción de expresión con un epítipo FLAG en C-terminal. La proteína de fusión scFv::I27::FLAG se indujo con arabinosa a temperaturas que varían desde 26 °C a 38 °C y las células de *E. coli* se lisaron usando ultrasonidos. Las proteínas solubles se separaron de los sedimentos insolubles y los agregados proteicos mediante centrifugación (14K 1 min).

- 20 La Figura 4 demuestra la excelente solubilidad de un clon de IGLV3-1 cuando se expresa en el citosol de *E. coli* a 25 °C. La proteína de fusión scFv::I27::FLAG está por completo en fracción soluble. Esto demuestra algo de escisión en N-terminal de una fracción menor de la proteína total, aunque esto se eliminó cuando la proteína se extrajo en condiciones desnaturizantes, lo que sugiere que fue causada por la interacción de proteasas periplasmáticas que se liberaron con la permeabilización mediante el detergente 8TGP.

- 25 Por lo tanto, debido a la alta frecuencia de recuperación de IGLV3-1 de la selección sistemática de solubilidad en *E. coli*, se caracterizó adicionalmente por los rasgos necesarios para un armazón ejemplar de una biblioteca de scFv soluble - estabilidad en el citoplasma a temperaturas cercanas a 37 °C, y tolerancia para un bucle CDR3 diversificado. El IGLV3-1 se ensayó para termoestabilidad en el citoplasma de *E. coli* a una temperatura que varía de 28 °C a 38 °C. Se descubrió que era altamente soluble a 36 °C cuando se acoplaba con las regiones J1 y J2 de cadena ligera, así como con las regiones J que se formaron usando los oligonucleótidos degenerados como cebadores durante la PCR del ADNc de las PBMC. A 36 °C y por encima, se demuestra un grado de mal plegamiento. Esto se confirmó tanto mediante inmunofluorescencia como mediante análisis por transferencia de Western de scFv marcado con FLAG.

35

#### Ejemplo 4. Intercambio de dominio J de IGLV3-1

- Las secuencias de oligonucleótidos degenerados enumeradas en la Tabla 2 que se usaron para amplificar los dominios VL tuvieron que cebar las 7 regiones  $\lambda$  J en el genoma humano (Tabla 3). Por tanto, los clones aislados de la selección sistemática tenían regiones  $\lambda$  J híbridas que representaban una secuencia no canónica que podía haber reducido su estabilidad del plegamiento.

40

**Tabla 3. Regiones J de  $\lambda$  humanas**

Región lambda J	Secuencia de aminoácidos
J1	VFGTGKTVTs (SEQ ID NO: 72)
J2	VFGGGTKLTVs (SEQ ID NO: 73)
J3	VFGGGTKLTVs (SEQ ID NO: 74)
J4	VFGGGTQLIIs (SEQ ID NO: 75)
J5	VFGEGTELTVs (SEQ ID NO: 76)
J6	VFGSGTKTVTs (SEQ ID NO: 77)
J7	VFGGGTQLTAs (SEQ ID NO: 78)

- 45 La comparación de las regiones J de los clones IGLV3-1 más estables que tenían la secuencia de la línea germinal de las regiones marco demostró la similitud más alta con las regiones J 1 y 2/3. Por lo tanto, la región J híbrida ("VFGTGTKLIIs" (SEQ ID NO: 79)) se reemplazó con las secuencias  $\lambda$  J1 o J2/3 de la línea germinal (Tabla 3) para probar si la termoestabilidad del armazón IGLV3-1 se podría mejorar adicionalmente. Las variantes se probaron a temperaturas entre 30 °C, 32 °C, 34 °C, 36 °C y 38 °C. De manera subjetiva, se consideró que  $\lambda$ J1 dio un plegamiento y una solubilidad ligeramente mejores que J2/3 o la región J híbrida original del clon ensayado.

50

La Figura 5 demuestra el comportamiento de termoestabilidad del clon original (n.º 8.93) con sustitución de la región  $\lambda$  J por J1 o J2.

55

**Ejemplo 5. Tolerancia de IGLV3-1 y IGHV3-23 a la diversificación de CDR3**

Para que un scFv sea útil como un marco para una biblioteca de afinidad, necesita ser tolerante a sustituciones en la región CDR3. Esto es especialmente cierto para scFv que se expresan en un ambiente reductor, tal como el citosol de *E. coli*, donde los enlaces disulfuro intra-dominio estabilizadores están ausentes.

Para ensayar la estabilidad del armazón de scFv preferido, IGLV3-1::IGHV3-23, se diversificó por separado la región CDR3 de cada dominio. Por lo tanto, ambos genes IGLV3-1 y IGHV3-23 se ensayaron para su tolerancia a la diversificación de CDR3. La Figura 2 muestra la región alrededor de CDR3 para ambas secuencias, así como la diversificación propuesta. La CDR3 de IGLV3-1 de 2 aminoácidos se reemplazó con una región "-NNNGGNNN-" (SEQ ID NO: 29) (en donde 'N' es un aminoácido que no sea Trp, Gln, Lys, Glu, Met). Asimismo, la CDR3 de IGHV3-23 de 12 aminoácidos se reemplazó con una región "-NNNGNNN-" (SEQ ID NO: 29). Cada dominio se ensayó por separado para la permeabilidad y la expresión como una biblioteca agrupada de clones. Además, los clones individuales tomados de manera aleatoria también se ensayaron y se secuenciaron para confirmar la diversidad esperada.

La IGLV3-1 CDR3 se reemplazó desde el resto 91 en adelante modificando el dominio IGLV3-1 mediante PCR usando la secuencia de oligonucleótidos (inversa):

```
GATCAGGGTCTGAGACGAGACCGTCACTTTCGTACCGGTGCCGAACACCAC
AGTANNANNANTCCGCCANNANNANNGTCCACGCCTGACAGTAATAGT
CAGC (SEQ ID NO: 80)
```

La traducción (inv/comp) de esta secuencia da (cuando "N" es cualquier otro aminoácido que no sea Trp, Gln, Lys, Glu, Met):

```
...ADYYCQAWD(91) NNNGGNNN TVVFGTGKTVSS (SEQ ID NO: 81)
```

El reemplazo de la CDR3 del IGLV3-1 fue un éxito rotundo. La aparición de la población con inducción proteica a 30 °C fue muy soluble clones con buena expresión. Se analizaron de manera individual 40 clones y 36 se clasificaron como excelentes para la solubilidad y la expresión. Los 4 clones fallidos, y otros 16 con buena solubilidad y expresión, se secuenciaron en todo el dominio de VL. Se confirmó que los cuatro clones fallidos fallaron debido a desplazamientos del marco en el cebador oligonucleotídico largo usado para amplificar el gen. Los otros clones examinados que tuvieron el correcto marco de lectura tuvieron una mezcla aleatoria de aminoácidos, y demuestran que el marco de la línea germinal IGLV3-1 es muy tolerante de la diversificación de CDR3.

Por lo tanto, para IGLV3-1, la solubilidad de una biblioteca de CDR3 diversificada cuando se expresaba a 30 °C en el citoplasma de *E. coli* era sorprendentemente alta. Aproximadamente el 90 % de los clones fueron solubles con alta expresión. El 10 % de los clones con baja expresión o sin expresión, o se plegaron incorrectamente, se secuenciaron y se demostró que estaban desplazados al marco, predominantemente en la región del cebador inverso que era por necesidad de aproximadamente 100 bases de longitud. Las deleciones de bases son un error común cuando se construyen bibliotecas sintéticas usando oligonucleótidos largos y otros grupos han desarrollado estrategias de selección previa basadas en la selección de antibióticos para enriquecer los alelos en fase (por ejemplo, Ge et al., 2010).

La región de CDR3 del dominio de VH, IGHV3-23 se reemplazó usando un oligonucleótido inverso de manera similar al método descrito anteriormente desde el resto 98 en adelante modificando el dominio IGHV3-23 mediante PCR usando la secuencia de oligonucleótidos (inversa):

```
GATCAGGGTCTGAGACCCGCTGCTCACGGTAACCATGGTACCTTGACCCCA
AATATCAAACGCANNANNANNGCCANNANNANNTTTCGCACAGTAGTAAA
CAGC (SEQ ID NO: 82)
```

La traducción (inv/comp) de esta secuencia da (cuando "N" es cualquier otro aminoácido que no sea Trp, Gln, Lys, Glu, Met):

```
VYYCAK(98) NNNGNNN AFDIWGQGTMVT (SEQ ID NO: 83)
```

El IGHV3-23 demostró ser tan robusto para la diversificación de CDR3 como el dominio IGLV3-1, con un 80 % de clones probados que muestran alta expresión soluble y el 20 % que se expresaron mal fueron explicables tras la secuenciación debido a discrepancias conservativas en el marco, o más comúnmente, deleciones de un solo par de bases en la región del cebador oligonucleotídico largo, cambiando así el marco de traducción de la proteína.

Por lo tanto, para IGHV3-23, la solubilidad de una biblioteca de CDR3 diversificada cuando se expresaba a 30 °C en el citoplasma de *E. coli* era, de nuevo, sorprendentemente alta (aproximadamente el 80 %). De nuevo, el desplazamiento del marco de la proteína de fusión fue responsable de muchos aspectos negativos. El acortamiento del bucle de CDR3 de 12 a 7 aminoácidos también mejoró la solubilidad de esta biblioteca en comparación con el clon parental.

La Figura 6A demuestra la solubilidad y la elevada expresión de 4 clones independientes con la CDR3 de IGLV3-1 diversificada. La Figura 6B demuestra una muestra de toda la población de clones con la CDR3 de IGHV3-23 diversificada. Por lo tanto, en resumen, el marco IGLV3-1::IGHV3-23 es un armazón ejemplar para construir una biblioteca de afinidad, que es idéntico a la secuencia de la línea germinal humana y permaneciendo robustamente soluble con sustitución de los bucles de CDR3 con la secuencia diversificada. Además, la combinación del armazón con el método de presentación de proteínas RED en el citoplasma de *E. coli* permite la selección sistemática simultánea de la estabilidad de la proteína de afinidad, la expresión y la unión a la molécula diana. El armazón es altamente estable y soluble en el ambiente reductor del citoplasma de *E. coli* donde carece de los enlaces disulfuro intradominio de estabilización que son un requisito inicial para el plegamiento y la estabilidad de casi todas las otras proteínas de scFv. Este armazón permitirá la producción a bajo coste de reactivos de afinidad en el citoplasma de *E. coli* para investigación, usos terapéuticos o de diagnóstico, así como el uso de tales reactivos en el citoplasma de células de mamífero para dirigir proteínas endógenas.

#### **Ejemplo 6. Construcción de una biblioteca diversificada de IGLV3-1::IGHV3-23 scFv**

El armazón IGLV3-1::IGHV3-23 se diversificó usando la estrategia descrita para el Ejemplo 5 para introducir las secuencias de aminoácidos 'NNNGGNNN' (SEQ ID NO: 86) y 'NNNGNNN' (SEQ ID NO: 87) en las regiones CDR3 de VL y VH, respectivamente.

La diversidad se introdujo, en primer lugar, creando un armazón de base que consistió en la secuencia del marco de IGLV3-1 y la región J para IGHV3-23 tal como sigue:

Secuencia marco:

```

ATG  GGA  GAC  GGT  CAG  TCT  GTG  CTG  ACT  CAG  CCA  CCC
TCAGTGTCCGTGTCCCCAGGACAGACAGCCAGCATCACCTGCTCTGGAGAT
AAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCCAGTCC
CCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCTGAG
CGATTCTCTGGCTCCAACCTCTGGGAACACAGCCACTCTGACCATCAGCGGG
ACCCAGGCTATGGATGAGGCTGACTATTACTGTCAGGCG   TGG   GAC

tgagacctagacggtctct gcg TTT GAT ATT TGG GGT CAA GGT ACC ATG GTT ACC
GTG AGC AGC TCG TCT CaG ACC (SEQ ID NO: 88).

```

Este marco se clonó en el vector de expresión citoplásmica con el PG y elementos de dominio de unión a ADN a través de sitios *BsmBI* flanqueantes. La secuencia de intervención (la región VL J y marco IGHV3-23) (SEQ ID NO: 89) estaba codificada en un plásmido separado que sirvió como un molde para una PCR usando cebadores degenerados (SEQ ID NO: 90 y 91) que contenían la diversidad de CDR3 de ambas regiones VL y VH en los extremos 5' y 3', respectivamente. Estas secuencias de cebadores tenían sitios de restricción *BsaI* terminales que permitían la clonación sin interrupciones del producto de PCR en sitios *BsaI* orientados apropiadamente en el armazón.

Secuencia de intervención:

ACT GTG GTG TTC GGC acc ggt acg aaa gtg acT gtc TCA TCT CAG ACC  
 GGTGGTTCTGGTGGTGGTGGTTCTGGCGGCGGCGGCTCCGGTGGTGGTGA  
 TCCGAAGTCCAAGTCTGGAGTCCGGCGGTGGCCTGGTGCAGCCAGGTGGC  
 AGCCTGCGCCTGAGCTGCGCCGCATCCGGTTTTACTTTTCAGCAGCTACGCG  
 ATGTCGTGGGTGCGCCAGGCACCGGGCAAGGGCCTGGAGTGGGTGAGCGC  
 CATCAGCGGTAGCGGCGGTTCTACGTATTATGCGGACAGCGTCAAGGGCCG  
 TTTCACCATCAGCCGTGACAATTCCAAAAACACCCTGTACTTGCAGATGAA  
 CAGCTTGCGTGCGGAAGATACGGCTGTTTACTACTGTGCGAAA (SEQ ID NO:  
 89)

Cebador degenerado 1:

gatcag ggtctca ggac NNT NNT NNT ggc gga NNT NNT NNT ACT GTG GTG TTC  
 GGC acc ggt acg aaa gtg (SEQ ID NO: 90)

Cebador degenerado 2:

GATCAGGGTCTCAACGCANNANNANNGCCANNANNANNTTTCGCACAGTA  
 GTAAACAGCCGTATCTTC (SEQ ID NO: 91)

10 µg del vector de armazón de la base se cortaron con  $\beta$ sal. El vector cortado se precipitó usando Sureclean (Bioline). El inserto, que contiene las regiones de diversidad de CDR3, se generó por PCR a partir del marco núcleo como molde usando las SEQ ID NOS: 90 y 91. Se purificaron en gel 2 µg de PCR de inserto antes de la digestión con  $\beta$ sal. El digesto de PCR se precipitó usando Sureclean. Las cantidades equimolares del vector digerido y del inserto de PCR se ligaron usando una ADN ligasa T4. El ligamiento se eliminó por calor y se sometió a electroporación en serie en células Argentum de *E. coli* (Alchemy Bio). Las células electroporadas se extendieron sobre placas de agar de 15 cm de LB + carbenilicina (40 µg/ml) + glucosa (al 0,1 %). El tamaño total de la biblioteca fue mayor de  $1 \times 10^8$  clones independientes.

La calidad de la construcción de la biblioteca se evaluó mediante la expresión de los scFv diversificados. Tal como se observó anteriormente, la expresión de los compañeros de fusión solubles, parcialmente solubles o insolubles se puede evaluar directamente mediante la aparición del scFv en el sistema de presentación RED utilizando el dominio de unión a peptidoglucano (PG) y un reportero de expresión cromogénica tal como SNAP (New England Biolabs). Una proteína de fusión soluble está notablemente distribuida de manera uniforme alrededor del perímetro de la célula ya que es libre para difundirse y unirse a la pared celular una vez que las membranas se han permeabilizado (por ejemplo, Figura 1A). En comparación, una proteína de fusión insoluble forma un agregado densamente teñido que no migra a la pared celular (por ejemplo, la Figura 1B). Una fusión parcialmente soluble tiene algunas características de cada. Los presentes inventores han descubierto una excelente correlación entre la apariencia de una proteína de fusión, tal como se describe anteriormente, y la cantidad que aparece en las porciones solubles/insolubles en los análisis por transferencia de Western tal como el scFv en la Figura 4.

Según este estándar empírico, la biblioteca de los presentes inventores de scFv diversificado estaba compuesta de aproximadamente el 90 % de miembros solubles y bien expresados, lo que indicaba la tolerancia del armazón IGLV3-1::IGHV3-23 para la diversidad de CDR3 insertada. La Figura 9 es una imagen marcada con SNAP de una muestra de la biblioteca expresada.

Para confirmar adicionalmente que la biblioteca estaba compuesta de las regiones de CDR3 de  $V_L$  y  $V_H$  aleatorizadas, se secuenciaron 10 clones independientes. La secuenciación mostró que la composición de los bucles de CDR3 (mostrada en la Tabla 4) era la esperada para una diversidad de codón creada por el triplete de nucleótidos 'NTT' usado para degeneración, es decir, la ausencia de codones de parada y los aminoácidos W, Q, M, K y E.

**Tabla 4. Muestra de secuencias de bucle de CDR3 en la biblioteca aleatorizada**

Clon	CDR3 de VL	CDR3 de VH
1	PFGGGGYV (SEQ ID NO:92)	PPHGAPA (SEQ ID NO:93)
2	LCIGGVAS (SEQ ID NO:94)	HNSGNNF (SEQ ID NO:95)

Clon	CDR3 de VL	CDR3 de VH
3	FVSGGIST (SEQ ID NO:96)	FNFGNAY (SEQ ID NO:97)
4	INSGGASF (SEQ ID NO:98)	XXXGTNY (SEQ ID NO:99)
5	SRAGGCNG (SEQ ID NO:100)	FDYGHCI (SEQ ID NO:101)
6	TNRGGVCA (SEQ ID NO:102)	TAAGVPF (SEQ ID NO:103)
7	Clon mezclado	Clon mezclado
8	FSTGGCAF (SEQ ID NO:104)	AICGATA (SEQ ID NO:105)
9	FXGGGDGT (SEQ ID NO:106)	PYRGSFF (SEQ ID NO:107)
10	IIPGGLYA (SEQ ID NO:108)	PVIGSNT (SEQ ID NO:109)

#### Ejemplo 7. Selección sistemática de la biblioteca IGLV3-1::IGHV3-23 para la unión a la diana mAG1

La biblioteca diversificada se seleccionó sistemáticamente para clones que se unían a la proteína diana, mAG. El verde Azami (AG) es un ortólogo distante de la proteína verde fluorescente (GFP, del inglés *green fluorescent protein*) de *Aequorea victoria*. Aunque es de baja identidad de secuencia (5 %), es similarmente verde fluorescente con un pico de absorción a 492 nm y un pico de emisión a 510 nm. Se ha documentado una forma monomérica (mAG) por Karasawa et al. (2003) y se recodificó para la expresión óptima en *E. coli* mediante DNA2.0 (EE.UU.). Se incluyó un motivo de biotilación de *E. coli* BirA en C-terminal y un marcador 6 x His para ayudar en la purificación y la unión de la matriz de mAG. La secuencia de aminoácidos de la proteína mAG-BioHis6, se enumera como la SEQ ID NO:110.

Secuencia de la proteína mAG BioHis6:

MVSVIKPEMKIKLCMRGTVNGHNFVIEGEGKGNPYEGTQILDNLNVTGAPLPF  
 AYDILTTVFQYGNRAFTKYPADIQDYFKQTFPEGYHWERSMTYEDQGICTATS  
 NISMRGDCFFYDIRFDGTNFPNGPVMQKKTLKWEPSTEKMYVEDGVLKGDV  
 NMRLLEGGGGHYRCDFKTTYKAKKEVRLPDAHKIDHRIEILKHDKDYNKVKL  
 YENAVARYSMLPSQAKSGGLNDIFEAQKIEWHEDTGGSHHHHHH (SEQ ID  
 NO: 110)

10<sup>10</sup> células de la biblioteca diversificada, que representan una redundancia de aproximadamente 100 veces se indujeron para la presentación RED tal como se describe en el Ejemplo 2 y en el documento WO 2011/075761. Las células permeabilizadas se suspendieron en 50 ml de PBS y se marcaron con mAG purificado que se había preunido a microperlas conjugadas con estreptavidina de MACS (130-048-102, Miltenyi Biotec). Las células y las microperlas se agitaron suavemente durante toda una noche a 4 °C. Después se cargaron en columnas de 3 x LS (130-042-401, Miltenyi Biotec) que se fijaron a un soporte magnético. Cada columna se lavó con 50 ml de PBS. Las células se eluyeron en 10 ml de PBS, se agruparon y se sedimentaron. El ADN del plásmido que codifica la biblioteca en el vector de presentación de RED se aisló del sedimento celular mediante la lisis alcalina. Después se electroporó de nuevo el plásmido en células *Argentum* y la inducción, la unión y la purificación en columna se repitieron. Tras cuatro iteraciones de la selección por afinidad, se observó una baja abundancia de células permeabilizadas por RED mediante microscopía de fluorescencia que se unían a la proteína mAG. En la quinta iteración, las células permeabilizadas se clasificaron para la unión de mAG por FACS. Las células se marcaron para FACS usando el ligando SNAP, para normalizar la expresión de la proteína de fusión y mAG. El FACS de la población celular durante la colección de 4.428 eventos positivos en mAG de 2,46 x 10<sup>8</sup> eventos totales mostró una abundancia de aproximadamente 1 evento de unión en 10<sup>5</sup> células. El scFv de las células positivas para mAG que eran la salida de FACS se recuperó mediante PCR usando cebadores de oligonucleótidos que flanquean la secuencia de scFv y el producto se volvió a clonar en el vector de presentación RED. El análisis del resultado de la selección final para las células que fueron positivas para la unión de mAG mostró que aproximadamente el 40 % (23/60) de los clones eran positivos para mAG1. Por lo tanto, la etapa de FACS era capaz de un enriquecimiento de aproximadamente 10<sup>5</sup> veces de células positivas del fondo de la biblioteca.

La Figura 10 muestra la unión de mAG a células permeabilizadas de RED para clones que fue negativa (clon 25) y positiva (clon 34) para la unión a mAG.

Se secuenciaron 20 clones positivos en mAG y se descubrió que los 20 eran idénticos. La secuencia de la proteína del clon IGLV3-1::IGHV3-23 de unión a mAG se enumeran en la SEQ ID NO: 111 a continuación (con las secuencias de CDR3 en negrita y agrandado, y el enlazador peptídico subrayado). Se descubrió que la CDR3 de VL era 'FNLGGCGD' y la CDR3 de VH era 'HIDGPVA', lo que se ajusta a la diversidad diseñada.

scFv de unión a anti-mAG:

MGDGQSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRP  
SGIPERFSGSNSGNTATLTISGTQAMDEADYYCQAWDFN**LGCGD**TVVFGTGTKVTVSSQ  
TGGSGGGGSGGGGSGGGGSEVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAP  
GKGLEWVSAISGSGGSTYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYCAK**HID**  
**GPV**AAFDIWGGQTMVTVSSSSQTSILVA (SEQ ID NO: 111)

- 5 Para determinar las propiedades de scFv  $\alpha$ -mAG, se clonó el gen en un vector de expresión con 6x His en C-terminal y un marcador de epítipo FLAG. La expresión de scFv se indujo mediante arabinosa y las células se permeabilizaron con 8TGP al 0,5 % para liberar scFv soluble en el sobrenadante. El material celular insoluble se sedimentó y las muestras de ambos extractos se hirvieron con colorante de carga SDS-PAGE y se sometieron a electroforesis en un gel al 15 % de SDS-PAGE. Las proteínas redisueltas se transfirieron a una membrana de nitrocelulosa y se sondearon con un anticuerpo monoclonal de ratón  $\alpha$ -FLAG. La Figura 11 demuestra que el scFv  $\alpha$ -mAG estaba casi exclusivamente en la fracción soluble.

- 15 Para demostrar que el scFv  $\alpha$ -mAG era específico para proteína mAG, y no meramente un anticuerpo 'pegajoso', las células permeabilizadas de  $\alpha$ -mAG scFv se marcaron con proteína relacionada estructuralmente y funcionalmente con mAG, EGFP. Estas células, mientras que se unen a mAG, no se unen a EGFP (datos no mostrados). Para evaluar adicionalmente la especificidad de scFv  $\alpha$ -mAG, la proteína  $\alpha$ -mAG scFv His6-FLAG también se unió a la resina IMAC Ni-sepharose. Un lisado celular crudo de proteína mAG 'limpia' (con los marcadores His6 y FLAG eliminados) se mezcló con la resina. Las proteínas no unidas se eliminaron con el lavado. Las imágenes de microscopía de fluorescencia en la Figura 12 demuestran que las perlas de resina con scFv unido a  $\alpha$ -mAG se unieron a mAG, mientras que las perlas de control no se unieron. Las proteínas unidas se eluyeron con imidazol y se sometieron a electroforesis en gel SDS-PAGE. La tinción de Coomassie del gel (Figura 13) demostró una banda en la muestra de scFv  $\alpha$ -mAG que era del tamaño correcto para ser proteína mAG sin otras bandas específicas para el lisado celular de mAG evidentes.

- 25 Por lo tanto, la presente invención se puede usar con éxito para generar una biblioteca de polipéptidos de scFv que contienen bucles de CDR3 aleatorizados y seleccionada sistemáticamente para identificar scFv que presenta actividad de unión específica.

#### **Ejemplo 8. Presentación en fago lambda usando el scFv $\alpha$ -mAG IGLV3-1::IGHV3-23**

- 30 Para demostrar la utilidad de un armazón que presentaba una estabilidad mejorada y un plegamiento productivo en el ambiente reductor del citoplasma, el scFv de  $\alpha$ -mAG IGLV3-1::IGHV3-23 se clonó como una fusión de C-terminal a la proteína de la cápside gpD del bacteriófago lambda. El bacteriófago Lambda se ha descrito durante mucho tiempo como un vehículo ejemplar para la presentación de proteínas, ya que tiene una serie de ventajas sobre el fago filamentoso. La proteína de la cápside de lambda, gpD, está presente en aproximadamente 400 copias por cabeza de fago, un es un compañero de presentación robusto y tolerante, que permite que más del 80 % de la gpD cargada por cápside sea proteínas de fusión recombinantes mientras se mantiene la viabilidad infecciosa (Vaccaro et al., 2006). Además, tolera fusiones bien a su extremo N-terminal o a C-terminal.

- 40 Por lo tanto, un bacteriófago lambda o vector empaquetado de manera equivalente, tiene una presentación multivalente de la proteína de la biblioteca, en comparación con la presentación nominal de molécula única de fago filamentoso. Esta presentación multivalente puede dar como resultado eficacias de captura fenomenales del fago a partir de una solución de unión de hasta casi el 100 % de captura (Mikawa et al., 1996). Adicionalmente, el ensamblaje de bibliotecas del bacteriófago lambda se facilita mediante la disponibilidad comercial de kits que permiten una elevada eficacia de empaquetamiento de lambda (de hasta  $2 \times 10^9$  ufp/ $\mu$ g).

- 50 Sin embargo, el bacteriófago lambda no ha disfrutado de la popularidad del uso del fago filamentoso para la presentación de anticuerpos debido al hecho singular de que lambda y fagos relacionados como P2/P4, P22, T7 y T4, tienen un estilo de vida lítico que resulta de su ensamblaje en el citoplasma. Como la gran mayoría de armazones no se pliegan de manera productiva sin puentes disulfuro interdominio oxidados, esto ha impedido en gran medida el uso del bacteriófago lambda para la presentación del anticuerpo.

- 55 La identificación de los presentes inventores de una serie de compañeros de IGLV para el dominio IGHV3-23 que forma un armazón de scFv citoplasmáticamente estable les ha permitido demostrar la aplicación ejemplar de presentación en lambda para la selección sistemática de anticuerpos. El scFv de  $\alpha$ -mAG se clonó como una fusión en C-terminal a la proteína de la cápside gpD de lambda con la expresión de la proteína de fusión bajo el control del promotor araBAD inducible por arabinosa y del regulador araC. Esta unidad se clonó en el genoma del bacteriófago lambda de manera similar a otras plataformas de presentación en lambda (Mikawa et al., 1996; Sternberg y Hoess, 1995; Minenkova et al., 2003), con la notable excepción de que el genoma de lambda usado era genéticamente

*c/857 gpD<sup>+</sup> RS*. La delección de los genes RS, que constituyen los genes de endolisina (R) y de porina (S) necesarios para la lisis celular, se describió en la Solicitud de Patente Internacional N.º PCT/AU2012/000761 para el uso de bacteriófagos de lisis defectuosa en la presentación lamdoide. Un vector de fago de lisis defectuosa usado para la presentación lamdoide permite el empaquetamiento de una partícula infectiva de bacteriófago en el citoplasma.

- 5 Estas partículas continúan acumulándose dentro de la célula, con su proteína de fusión de la cápside unida a su superficie a alta densidad, hasta que el crecimiento es detenido por el investigador que procesa las células bacterianas hospedadoras para la presentación citoplasmática de RED. De este modo, la preparación resultante se puede seleccionar de manera sistemática para la unión del antígeno de proteína de fusión por FACS. Para liberar las partículas del bacteriófago que están encapsuladas dentro de la célula permeabilizada que han sido clasificadas positivamente por FACS para la unión al antígeno simplemente se requiere la adición de una lisozima. Una preparación de lisozima altamente activa se puede comprar comercialmente para esta tarea (por ejemplo, Ready-Lyse de Epicentre). Para completar la recuperación de los clones seleccionados por afinidad, las partículas infecciosas de bacteriófago pueden infectarse en células hospedadoras de *E. coli* y crecer como lisógenos. Por lo tanto, los expertos en la técnica deben apreciar que el uso de un fago defectuoso de lisis, junto con un armazón de anticuerpo humano estable citoplásmicamente, permite altas frecuencias de captura de clones de biblioteca polivalente en el formato de bacteriófago libre, con la última selección llevada a cabo mediante FACS. De forma importante, este cambio en el formato de selección ocurre *sin ningún requisito para reformatear la construcción de expresión de la biblioteca*. Por lo tanto, este es un sistema de selección sistemática que tiene doble capacidad para la selección altamente paralela (selección de bacteriófagos libres) con baja selectividad clonal y una selección sistemática con alta selectividad clonal pero bajo rendimiento (FACS de bacteriófago encapsulado).

Para demostrar los beneficios de la presentación en fago lambda usando los polipéptidos de la presente divulgación, se clonó el modelo de fusión scFv  $\alpha$ -mAG (como uno de los muchos ejemplos adecuados de polipéptidos de la presente divulgación) como una fusión en C-terminal al gen de gpD de la cápside de lambda.

La secuencia de ADN de la construcción de fusión gpD::a-mAG scFv usada fue:

```

ATGACGAGCAAAGAAACCTTTACCCATTACCAGCCGCAGGGCAACAGTGA
CCCGGCTCATACCGCAACCGCGCCCGGCGGATTGAGTGCGAAAGCGCCTGC
AATGACCCCGCTGATGCTGGACACCTCCAGCCGTAAGCTGGTTGCGTGCGGA
TGGCACCACCGACGGTGCTGCCGTTGGCATTCTTGCGGTTGCTGCTGACCA
GACCAGCACCGCTGACGTTCTACAAGTCCGGCACGTTCCGTTATGAGGA
TGTGCTCTGGCCGGAGGCTGCCAGCGACGAGACGAAAAAACGGACCGCGT
TTGCCGGAACGGCAATCAGCATCGTTGGAGGTAGCGGCGGATCGGATGAC
GACGATAAGTCTAGAAATGGCGGAGACGGTCAGTCTGTGCTGACTCAGCCA
CCCTCAGTGTCCGTGTCCCCAGGACAGACAGCCAGCATCACCTGCTCTGGA
GATAAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCCAG
TCCCCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCT
GAGCGATTCTCTGGCTCCAACCTCTGGGAACACAGCCACTCTGACCATCAGC
GGGACCCAGGCTATGGATGAGGCTGACTATTACTGTCAGGCGTGGGACTTT
AATCTTGGCGGATGTGGTGATACTGTGGTGTTCGGCACCGGTACGAAAGTG
ACTGTCTCATCTCAGACCGGTGGTTCTGGTGGCGGTGGTTCTGGCGGCGGC
GGCTCCGGTGGTGGTGGATCCGAAGTCCAACCTGCTGGAGTCCGGTGGTGGC
CTGGTGCAGCCAGGTGGCAGCCTGCGCCTGAGCTGCGCCGCATCCGGTTTT
ACTTTCAGCAGCTACGCGATGTCGTGGGTGCGCCAGGCACCGGGCAAGGGC
CTGGAGTGGGTGAGCGCCATCAGCGGTAGCGGCGGTTCTACGTATTATGCG
GACAGCGTCAAGGGCCGTTTCACCATCAGCCGTGACAATTCCAAAAACACC
CTGTACTTGACAGATGAACAGCTTGCGTGCGGAAGATACGGCTGTTTACTAC
TGTGCGAAACATATTGATGGCCCTGTTGCTGCGTTTGATATTTGGGGTCAAG
GTACCATGGTTACCGTGAGCAACTCGAGCGATTACAAGGACGATGATGACA
AATAA (SEQ ID NO: 112)

```



La secuencia proteica de la proteína de fusión gpD::a-mAG scFv usada fue:

MTSKETFTTHYQPQGNSDPAHTATAPGGLSAKAPAMTPLMLDTSSRKLVAWDGTTDGAAVG  
 ILAVAADQTSTTLTFYKSGTFRYEDVLWPEAASDETKKRTAFAGTAISIVGGSGGSDDDD  
 KSRNGGDGQSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSK  
 RPSGIPERFSGSNSGNTATLTISGTQAMDEADYYCQAWDFNLGGCGDTVVFGTGTKVTVS  
 SQTGGSGGGGSGGGGSGGGGSEVQLLESGGGLVQPGSLRLSCAASGFTFSSYAMSWVRQ  
 APGKGLEWVSAISGSGGSTYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYCAKH  
 IDGPVAAFDIWGQGTMTVTVSNSSDYKDDDDK\* (SEQ ID NO: 113)

- 5 La fusión gpD:: a-mAG scFv se clonó entonces en el vector de presentación en lambda bajo el control del promotor araBAD. Las células huésped se indujeron para empaquetamiento lambda calentando el clon de lisógeno a 42 °C durante 15 minutos (el fondo genético lambda era c/857 gpD<sup>+</sup>RS- con el represor cl sensible a la temperatura). La proteína de fusión se indujo con arabinosa al 0,2 % inmediatamente después de la inducción térmica. Se dejó crecer el cultivo, con aireación, a 32 °C durante 75 minutos adicionales. Las células se sedimentaron y se resuspendieron en 1/3 de volumen de cultivo de medio LB + 8TGP al 0,5 % y se incubó a 25 °C durante 10 minutos para permeabilizar las células mediante el método RED para la selección sistemática. Para liberar el fago, se añadió 1/10.000 de volumen de lisozima Ready-Lyse (Epicentre) a la suspensión. Se añadió una gota de cloroformo para inactivar cualquier célula superviviente y las partículas del bacteriófago liberadas se titularon para unidades formadoras de lisógeno (ufc/ml). Se prepararon dos reservorios de bacteriófagos - uno con la construcción con la fusión clonada gpD::α-mAG scFv, el otro con una construcción vacía. La fusión gpD::α-mAG scFv se diluyó a 1 clon en 10<sup>9</sup> de construcción vacía, para simular una densidad de biblioteca de scFv inicial de solo unos pocos clones positivos. Esta biblioteca "dopada" se sometió entonces a cribado contra mAG biotinilado unido a un soporte de perlas de estreptavidina. El cribado se llevó a cabo de acuerdo con los métodos comúnmente usados para el cribado de fagos conocidos por los expertos en la materia. Se llevaron a cabo dos rondas de cribado, recuperándose la ronda final en la cepa del hospedador *E. coli* como lisógenos. La tercera ronda de selección sistemática se llevó a cabo mediante FACS. Las células de lisógeno se trataron tal como se describió anteriormente para la inducción de calor del bacteriófago junto con la inducción con arabinosa de la fusión gpD::α-mAG scFv. Sin embargo, en lugar de liberar partículas de bacteriófago con tratamiento de lisozima, las células permeabilizadas se incubaron con proteína mAG. Las células permeabilizadas se lavaron una vez, se resuspendieron en TBS + MgSO<sub>4</sub> 10 mM y después se clasificaron por la unión a mAG (es decir, las células positivas en mAG se marcarían de verde) mediante FACS. La Figura 14 (SUPERIOR) muestra una captura de pantalla de la clasificación de FACS en operación que demuestra la incidencia de células positivas en mAG. La incidencia final de células positivas en mAG tras el FACS, evaluada mediante microscopia de fluorescencia (Figura 14, INFERIOR), fue del 20 %.
- 30 Por lo tanto, se ha demostrado que la presentación en cápside de lambda, junto con los armazones de scFv estables y solubles de la presente divulgación, puede aislar de manera robusta los clones de unión a partir de una dilución de inicio relativamente alta (1 en 10<sup>9</sup>). Además, cuando se combina con un bacteriófago de lisis defectuosa y se trata mediante el método enseñado por RED, permite una magnificación adicional de las propiedades para incluir la capacidad de selección sistemática de FACS sin volver a clonar los miembros de la biblioteca.
- 35 La combinación de estos métodos acelera enormemente el proceso de selección para clones de anticuerpos con propiedades ideales (alta expresión, alta solubilidad y alta afinidad).
- 40 Los expertos en la materia apreciarán que se pueden hacer numerosas variaciones y/o modificaciones tal como se muestra en las realizaciones específicas. Las presentes realizaciones, por lo tanto, se deben considerar en todos los aspectos como ilustrativas y no como restrictivas.

Cualquier tratamiento de documentos, actas, materiales, dispositivos, artículos o similares que se han incluido en la presente memoria descriptiva es únicamente con el propósito de proporcionar un contexto para la presente invención. No debe tomarse como una admisión de que alguno o todos estos asuntos forman parte de la base de la técnica anterior o eran conocimientos generales comunes en el campo relevante para la presente invención tal como existía antes de la fecha de prioridad de cada reivindicación de la presente solicitud.

## REFERENCIAS

- 50 Al-Lazikani et al. (1997) J Mol Biol 273:927-948.  
 Altschul et al. (1993) J. Mol. Biol. 215: 403410.  
 Auf der Maur et al. (2002) JBC 277:45075-45085.  
 Becker et al. (2004) Curr. Opin. Biotech. 15:323-329.  
 55 Bork et al. (1994) J. Mol. Biol. 242: 309-320.

Borrebaeck (ed) (1995) Antibody Engineering. Oxford University Press.  
 Brezinschek et al. (1997) J. Clin. Invest. 99:2488-2501.  
 Briers et al. (2009) Biochem. Biophys. Res. Comm. 383:187-191.  
 Chen W et al. (2008) JMB 382:779-789.  
 5 Chothia and Lesk (1987) J. Mol Biol. 196:901-917.  
 Chothia et al. (1989) Nature 342:877-883.  
 Contreras-Martinez and DeLisa (2007) JMB 372:513-524.  
 Daugherty et al. (2000) J. Immunol. Methods 243:211-227.  
 Ewert et al. (2003) JMB 325: 531-553.  
 10 Fisher and Delisa (2009) JMB 385:299-311.  
 Froyen et al. (1995) Mol. Immunol. 37: 515-521.  
 Ge et al. (2010) Biotech Bioeng 106, 347-57.  
 Griffiths et al. (1994) EMBO J 13:3245-3260.  
 Guan et al. (1998) Proc. Natl. Acad. Sci. USA, 95: 13206-10.  
 15 He et al. (1995) NAR 23:4009-4010.  
 Higgins and Sharp (1989) CABIOS. 5: 151-153.  
 Hust and Dubel (2010) Antibody Eng. Chapter 5: Antibody Engineering, Vol 1; Springer.  
 Jermutus et al. (2001) PNAS 98:75-80.  
 Jirholt et al. (1998) Gene 215, 471-476.  
 20 Jurado et al. (2002) JMB 320:1-10.  
 Kabat (1987 and 1991) Sequences of Proteins of Immunological Interest. National Institutes of Health.  
 Karasawa et al. (2003) JBC 278:34167-34171.  
 Kenrick et al. (2007) Curr. Prot. Cyt. 4.6.1-4.6.27.  
 Kirchhofer et al. (2010) Nat. Struct. Mol. Bio. 17:133-139.  
 25 Knappik et al. (2000) JMB 296:57-86.  
 Kobayashi et al. (1997) Biotechniques 23:500-503.  
 Lefranc (2000) Curr. Prot. Imm. 1-37.  
 Lutz and Patrick (2004) Curr. Opin. Biot. 15:291-297.  
 Marsh et al. (2000) Hum. Mol. Genet. 9:13-25.  
 30 Martineau et al. (1998) JMB 280:117-127.  
 Mikawa et al. (1996) JMB 262:21-30.  
 Miller et al. (2006) Nat. Meth. 3:561-570.  
 Minenkova et al. (2003) Int J Can 106:534-544.  
 Parsons et. al. (2006) Biochem. 45:2122-2128.  
 35 Pini et al. (1998) JBC 273:21769-21776.  
 Plückerthun (1992) Immunol. Revs., 130:151-188.  
 Saeuens et al. (2005) JMB 352:597-607.  
 Skerra et al (1993) Curr. Opinion in Immunol. 5:256-262.  
 Smith (1985) Science 228:1315-1317.  
 40 Soderlind et al. (2000) Nat. Biotech. 18:852-856.  
 Sternberg and Hoess (1995) PNAS 92:1609-1613.  
 Stewart et al. (1993) J. Exp. Med. 177:409-418.  
 Tavladoraki et al. (1999) Eur. J. Biochem. 262:617-624.  
 Tse et al. (2002) JMB 317:85-94.  
 45 Vaccaro et al. (2006) J. Imm. Methods. 310:149-158.  
 Visintin et al. (1999) PNAS 96:11723-11728.  
 Vitetta et al. (1993) Immunol. Today 14: 252-259.  
 Viti et al. (2000) Meth. Enzy. 326:480-505

50 LISTADO DE SECUENCIAS

- <110> Affinity Biosciences Pty Ltd  
 Matthew, Beasley
- 55 <120> Anticuerpos solubles
- <130> 512188
- <160> 125
- 60 <170> PatentIn versión 3.5
- <210> 1
- <211> 456
- 65 <212> ADN
- <213> Homo sapiens

# ES 2 682 254 T3

<400> 1

tcttttcgcac agtaatatatac ggccgtgtcc tcggctctca ggctgttcat ttgcagatac	60
agcgtgttct tggaattgtc tctggagatg gtgaaccggc ccttcacgga gtctgcgtag	120
tatgtgctac caccactacc actaatagct gagaccact ccagcccctt ccctggagcc	180
tggcggaccc agctcatggc atagctgcta aaggtgaatc cagaggctgc acaggagagt	240
ctcagggacc cccagggctg taccaagcct ccccagact ccaacagctg cacctcacac	300
tggacacctg caaacaacaaa gaaaccctgg tcagaaactg ccacacgtat ccactgtttc	360
tctcactctt atccattcac actcaatttt tctatttctc catgaattac cttttaaaat	420
agccacaaga aaaagccagc tcagcccaaa ctccat	456

5 <210> 2  
 <211> 296  
 <212> ADN  
 <213> Homo sapiens

10 <400> 2

gaggtgcagc tggttgagtc tgggggagggc ttggtacagc ctgggggggtc cctgagactc	60
tcctgtgcag cctctggatt cacctttagc agctatgcca tgagctgggt ccgccaggct	120
ccaggggaagg ggctggagtg ggtctcagct attagtggta gtggtggtag cacatactac	180
gcagactccg tgaagggccg gttcaccatc tccagagaca attccaagaa cacgctgtat	240
ctgcaaataga acagcctgag agccgaggac acggccgtat attactgtgc gaaaga	296

15 <210> 3  
 <211> 98  
 <212> PRT  
 <213> Homo sapiens

20 <400> 3

Glu Val Gln Leu Leu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

# ES 2 682 254 T3

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr  
20 25 30

Ala Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val  
35 40 45

Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val  
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr  
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys  
85 90 95

Ala Lys

<210> 4  
<211> 481  
<212> ADN  
<213> Homo sapiens

<400> 4

atgaccctgc tgcaggtgga tgggctcggc ggggctgaaa tccccccaca cagtgtcat 60  
gtgtcacac tgccttaggg ctctttcatc cctggatctg tgtccaggcc aggcacgtgg 120  
gaagatttac ttggagttca gctcctcagt ttcaagcctt ttctctcccg ttttctctcc 180  
tgtaggatcc gtggcctcct atgagctgac tcagccaccc tcagtgtccg tgtccccagg 240  
acagacagcc agcatcacct gctctggaga taaattgggg gataaatatg cttgctggta 300  
tcagcagaag ccaggccagt cccctgtgct ggtcatctat caagatagca agcggccctc 360  
agggatccct gagcgattct ctggctccaa ctctgggaac acagccactc tgaccatcag 420  
cgggacccag gctatggatg aggctgacta ttactgtcag gcgtgggaca gcagcactgc 480

a 481

<210> 5  
<211> 285  
<212> ADN  
<213> Homo sapiens

<400> 5

# ES 2 682 254 T3

tcctatgagc tgactcagcc accctcagtg tccgtgtccc caggacagac agccagcatc 60  
 acctgctctg gagataaatt gggggataaa tatgcttgct ggtatcagca gaagccaggc 120  
 cagtcccctg tgctgggtcat ctatcaagat agcaagcggc cctcagggat ccctgagcga 180  
 ttctctgggt ccaactctgg gaacacagcc actctgacca tcagcgggac ccaggctatg 240  
 gatgaggctg actattactg tcaggcgtgg gacagcagca ctgca 285

<210> 6  
 <211> 95  
 <212> PRT  
 <213> Homo sapiens  
 <400> 6

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

<210> 7  
 <211> 781  
 <212> ADN  
 <213> Homo sapiens  
 <400> 7

# ES 2 682 254 T3

atggcctgga ccgttctcct cctcggcctc ctctctcact gcacaggtga tccccccagc	60
gtctcaccaa cctgcccagc ccaaggggttc tgggtccagc gtgtccttga ttctgagctc	120
aggagggccc ttctgtggt gggcaggatg ctcatgacct tgctgcaggg tgggaggctg	180
gtggggctga actccccca aactgtgctc aaaggcttgt gagagcctga gggactgcac	240
ctgccaggag agagtagtga gttttcagtt caaagtctcc atacaacagg aaagtcatgg	300
gccactgggg ctggggctga ttgcagggga taccctgagg gttcacagac tctctggagc	360
ttgtctggga cagcagggca agggatttca taagaagcat ctttcacctg caagccaacc	420
tctctcttat ttattttattt atttatttat ttattttattt atttattttt atctttgcag	480
gctctgtgac ctctatgtg ctgactcagc caccctcggt gtcagtggcc ccaggacaga	540
cggccaggat tacctgtggg ggaaacaaca ttggaagtaa aagtgtgcac tggtagcagc	600
agaagccagg ccaggcccct gtgctggtcg tctatgatga tagcgaccgg ccctcaggga	660
tccctgagcg attctctggc tccaactctg ggaacacggc caccctgacc atcagcaggg	720
tcgaagccgg ggatgaggcc gactattact gtcaggtgtg ggatagtagt agtgatcatc	780
c	781

<210> 8  
 <211> 290  
 <212> ADN  
 <213> Homo sapiens

<400> 8

tcctatgtgc tgactcagcc accctcggtg tcagtggccc caggacagac ggccaggatt	60
acctgtgggg gaaacaacat tgggaagtaa agtgtgcact ggtaccagca gaagccaggc	120
caggcccctg tgctggtcgt ctatgatgat agcgaccggc cctcagggat ccctgagcga	180
ttctctggct ccaactctgg gaacacggcc accctgacca tcagcagggt cgaagccggg	240
gatgaggccg actattactg tcaggtgtgg gatagtagta gtgatcatcc	290

<210> 9  
 <211> 97  
 <212> PRT  
 <213> Homo sapiens

<400> 9

# ES 2 682 254 T3

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ala Pro Gly Gln  
1 5 10 15

Thr Ala Arg Ile Thr Cys Gly Gly Asn Asn Ile Gly Ser Lys Ser Val  
20 25 30

His Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Val Leu Val Val Tyr  
35 40 45

Asp Asp Ser Asp Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser  
50 55 60

Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Arg Val Glu Ala Gly  
65 70 75 80

Asp Glu Ala Asp Tyr Tyr Cys Gln Val Trp Asp Ser Ser Ser Asp His  
85 90 95

Pro

<210> 10  
<211> 478  
<212> ADN  
<213> Homo sapiens

<400> 10

atggcctggg ctccactact tctcaccctc ctcgctcact gcacaggtgg ctgcctgcaa 60  
ggaattcagg gagcgttcct ggatgtcacc tgggctgatg atctgttcct cctgcctggg 120  
aaccagtctt catctctccc cactgatctc tgtgttgctc tcttcttgca ggttcttggg 180  
ccaattttat gctgactcag cccactctg tgtcggagtc tccggggaag acggtaacca 240  
tctcctgcac cggcagcagt ggcagcattg ccagcaacta tgtgcagtgg taccagcagc 300  
gcccgggcag tgccccacc actgtgatct atgaggataa ccaaagacc tctgggggtcc 360  
ctgatcggtt ctctggctcc atcgacagct cctccaactc tgcctccctc accatctctg 420  
gactgaagac tgaggacgag gctgactact actgtcagtc ttatgatagc agcaatca 478

<210> 11  
<211> 296  
<212> ADN  
<213> Homo sapiens

<400> 11

# ES 2 682 254 T3

```

aattttatgc tgactcagcc ccactctgtg tcggagtctc cggggaagac ggtaaccatc      60
tcctgcaccg gcagcagtgg cagcattgcc agcaactatg tgcagtggta ccagcagcgc      120
ccgggcagtg cccccaccac tgtgatctat gaggataacc aaagaccctc tggggtcctt      180
gatcggttct ctgggtccat cgacagctcc tccaactctg cctccctcac catctctgga      240
ctgaagactg aggacgaggc tgactactac tgtcagtctt atgatagcag caatca          296

```

<210> 12  
 <211> 98  
 <212> PRT  
 <213> Homo sapiens

<400> 12

```

Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu Ser Pro Gly Lys
1           5           10           15

Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser Ile Ala Ser Asn
          20           25           30

Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Ala Pro Thr Thr Val
          35           40           45

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

Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu Thr Ile Ser Gly
65           70           75           80

Leu Lys Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln Ser Tyr Asp Ser
          85           90           95

```

Ser Asn

<210> 13  
 <211> 462  
 <212> ADN  
 <213> Homo sapiens

<400> 13



# ES 2 682 254 T3

```

atgacctgct cccctctcct cctcaccctt ctcattcact gcacaggtgc ccagacacag      60
ggtcagggga ggggtccagg aagcccatga ggccctgctt tctccttctc tctctagacc      120
aagaatcacc gtgtctgtgt ctctcctgct tccaggggtcc tgggccccagt ctgtgttgac      180
gcagccgccc tcagtgtctg cggccccagg acagaagggtc accatctcct gctctggaag      240
cagctccaac attggaata attatgtatc ctggtaccag cagctcccag gaacagcccc      300
caaactcctc atttatgaca ataataagcg accctcaggg attcctgacc gattctctgg      360
ctccaagtct ggcacgtcag ccaccctggg catcaccgga ctccagactg gggacgaggc      420
cgattattac tgcggaacat gggatagcag cctgagtgtc gg                        462

```

<210> 14  
 <211> 296  
 <212> ADN  
 <213> Homo sapiens

<400> 14

```

cagtctgtgt tgacgcagcc gccctcagtg tctgcggccc caggacagaa ggtcaccatc      60
tcctgctctg gaagcagctc caacattggg aataattatg tatcctggta ccagcagctc      120
ccaggaacag cccccaaact cctcatttat gacaataata agcgaccctc agggattcct      180
gaccgattct ctggctccaa gtctggcacg tcagccaccc tgggcatcac cggactccag      240
actggggacg aggccgatta ttactgcgga acatgggata gcagcctgag tgcttg      296

```

<210> 15  
 <211> 99  
 <212> PRT  
 <213> Homo sapiens

<400> 15

```

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Ala Ala Pro Gly Gln
1          5          10          15

```

```

Lys Val Thr Ile Ser Cys Ser Gly Ser Ser Ser Asn Ile Gly Asn Asn
          20          25          30

```

# ES 2 682 254 T3

Tyr Val Ser Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu  
35 40 45

Ile Tyr Asp Asn Asn Lys Arg Pro Ser Gly Ile Pro Asp Arg Phe Ser  
50 55 60

Gly Ser Lys Ser Gly Thr Ser Ala Thr Leu Gly Ile Thr Gly Leu Gln  
65 70 75 80

Thr Gly Asp Glu Ala Asp Tyr Tyr Cys Gly Thr Trp Asp Ser Ser Leu  
85 90 95

Ser Ala Gly

<210> 16  
<211> 466  
<212> ADN  
<213> Homo sapiens

<400> 16

atggcctggt ctcctctcct cctcactctc ctgcctcact gcacaggtga ctggatacag 60  
gtccagggga ggggccctgg gaagcctatg gattcttgct ttctcctggt gtctctagaa 120  
gccgaataat gatgcctgtg tctctccac ttccagggtc ctgggccag tctgtgctga 180  
cgcagccgcc ctcaagtgtct ggggccccag ggcagagggt caccatctcc tgcactggga 240  
gcagctccaa catcggggca gggttatgatg tacactggta ccagcagctt ccaggaacag 300  
ccccaaact cctcatctat ggtaacagca atcgccctc aggggtccct gaccgattct 360  
ctggctccaa gtctggcacc tcagcctccc tggccatcac tgggctccag gctgaggatg 420  
aggctgatta ttactgccag tcctatgaca gcagcctgag tggttc 466

<210> 17  
<211> 299  
<212> ADN  
<213> Homo sapiens

<400> 17

cagtctgtgc tgacgcagcc gccctcagtg tctggggccc cagggcagag ggtcaccatc 60  
tcctgcactg ggagcagctc caacatcggg gcagggttatg atgtacactg gtaccagcag 120  
cttcaggaa cagcccccaa actcctcatc tatggtaaca gcaatcggcc ctccagggtc 180  
cctgaccgat tctctggctc caagtctggc acctcagcct ccctggccat cactgggctc 240  
caggctgagg atgaggctga ttattactgc cagtccatg acagcagcct gagggttc 299

<210> 18  
<211> 100  
<212> PRT  
<213> Homo sapiens

# ES 2 682 254 T3

<400> 18

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Gly Ala Pro Gly Gln  
1 5 10 15

Arg Val Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn Ile Gly Ala Gly  
20 25 30

Tyr Asp Val His Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu  
35 40 45

Leu Ile Tyr Gly Asn Ser Asn Arg Pro Ser Gly Val Pro Asp Arg Phe  
50 55 60

Ser Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Thr Gly Leu  
65 70 75 80

Gln Ala Glu Asp Glu Ala Asp Tyr Tyr Cys Gln Ser Tyr Asp Ser Ser  
85 90 95

Leu Ser Gly Ser  
100

5 <210> 19  
<211> 468  
<212> ADN  
<213> Homo sapiens

10 <400> 19

atggccagct tccctctcct cctcaccctc ctcactcact gtgcagggtga caggatgggg	60
accaagaaag gggccctggg aagcccatgg ggccctgctt tctcctcttg tctccttttg	120
tctcttgtca atcaccatgt ctgtgtctct ctcacttcca gggtcctggg ccagatctgt	180
gctgactcag ccaccctcag cgtctggggac ccccgggcag agggtcacca tctcttggtc	240
tggaagcagc tccaacatcg gaagtaatac tgtaaactgg taccagcagc tcccaggaac	300
ggccccaaa ctctcatct atagtaataa tcagcggccc tcaggggtcc ctgaccgatt	360
ctctgggtcc aagtctggca cctcagcctc cctggccatc agtgggctcc agtctgagga	420
tgaggctgat tattactgtg cagcatggga tgacagcctg aatggtcc	468

15 <210> 20  
<211> 296  
<212> ADN  
<213> Homo sapiens

20 <400> 20

# ES 2 682 254 T3

cagtctgtgc tgactcagcc accctcagcg tctgggaccc ccgggcagag ggtcaccatc 60  
 tcttggttctg gaagcagctc caacatcgga agtaatactg taaactggta ccagcagctc 120  
 ccaggaacgg cccccaaact cctcatctat agtaataatc agcggccctc aggggtccct 180  
 gaccgattct ctgggtccaa gtctggcacc tcagcctccc tggccatcag tgggctccag 240  
 tctgaggatg aggctgatta ttactgtgca gcatgggatg acagcctgaa tgggcc 296

<210> 21  
 <211> 99  
 <212> PRT  
 <213> Homo sapiens  
 <400> 21

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

Asn Gly Pro

<210> 22  
 <211> 468  
 <212> ADN  
 <213> Homo sapiens  
 <400> 22

# ES 2 682 254 T3

atggcgggct tccctctcct cctcaccctc ctcactcact gtgcagggtga caggatgggg	60
accaagagag gggccctggg aagcccatgg ggccctgctt tctcctcttg tctcctttcg	120
tctcttggtca atcaccatgt ctgtgtctct ctcacttcca gggtcctggg ccaggtctgt	180
gctgactcag ccaccctcag cgtctgggac ccccgggcag agggtcacca tctcttggtc	240
tggaagcagc tccaacatcg gaagtaatta tgtatactgg taccagcagc tcccaggaac	300
ggcccccaaa ctctcatct atagtaataa tcagcggccc tcaggggtcc ctgaccgatt	360
ctctgggtcc aagtctggca cctcagcctc cctggccatc agtgggctcc ggtccgagga	420

tgaggctgat tattactgtg cagcatggga tgacagcctg agtgggtcc	468
---	-----

<210> 23  
 <211> 296  
 <212> ADN  
 <213> Homo sapiens  
 <400> 23

cagtctgtgc tgactcagcc accctcagcg tctgggaccc ccgggcagag gggtcaccatc	60
tcttggttctg gaagcagctc caacatcgga agtaattatg tatactggta ccagcagctc	120
ccaggaacgg cccccaaact cctcatctat agtaataatc agcggccctc aggggtccct	180
gaccgattct ctgggtccaa gtctggcacc tcagcctccc tggccatcag tgggtcccg	240
tccgaggatg aggctgatta ttactgtgca gcatgggatg acagcctgag tgggtcc	296

<210> 24  
 <211> 99  
 <212> PRT  
 <213> Homo sapiens  
 <400> 24

# ES 2 682 254 T3

Gln Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln  
1 5 10 15

Arg Val Thr Ile Ser Cys Ser Gly Ser Ser Ser Asn Ile Gly Ser Asn  
20 25 30

Tyr Val Tyr Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu  
35 40 45

Ile Tyr Ser Asn Asn Gln Arg Pro Ser Gly Val Pro Asp Arg Phe Ser  
50 55 60

Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly Leu Arg  
65 70 75 80

Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu  
85 90 95

Ser Gly Pro

<210> 25  
<211> 493  
<212> ADN  
<213> Homo sapiens

<400> 25

atggcctgga cccctctctg gctcactctc ctcactcttt gcataggtgc tgcctcccag 60

ggctcaaccc catattatca tgctagctgt gccaacctgg ccctgagctt cggctcaaca 120

cagggagtag tgtaggggtgt gggactctag gcgtgaaacc cttatcctca cctcttctgt 180

cctcttttgc aggttctgtg gtttcttctg agctgactca ggaccctgct gtgtctgtgg 240

ccttgggaca gacagtcagg atcacatgcc aaggagacag cctcagaagc tattatgcaa 300

gctggtacca gcagaagcca ggacaggccc ctgtacttgt catctatggt aaaaacaacc 360

ggccctcagg gatcccagac cgattctctg gctccagctc aggaaacaca gcttccttga 420

ccatcactgg ggctcaggcg gaagatgagg ctgactatta ctgtaactcc cgggacagca 480

gtggtaacca tct 493

<210> 26  
<211> 290  
<212> ADN  
<213> Homo sapiens

<400> 26

tcttctgagc tgactcagga ccttgcctgtg tctgtggcct tgggacagac agtcaggatc 60  
 acatgccaaag gagacagcct cagaagctat tatgcaagct ggtaccagca gaagccagga 120  
 caggccccctg tactttgtcat ctatggtaaa aacaaccggc cctcagggat cccagaccga 180  
 ttctctggct ccagctcagg aaacacagct tccttgacca tcaactggggc tcaggcggaa 240  
 gatgaggctg actattactg taactcccgg gacagcagtg gtaaccatct 290

<210> 27  
 <211> 97  
 <212> PRT  
 <213> Homo sapiens  
 <400> 27

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

10 Leu  
 <210> 28  
 <211> 15  
 <212> PRT  
 15 <213> Secuencia artificial

<220>  
 <223> Enlazador peptídico

20 <400> 28

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

<210> 29  
 <211> 8  
 <212> PRT  
 25 <213> Secuencia artificial  
 <220>

<223> Secuencia de variante de CDR 1

<220>  
<221> VARIANTE  
5 <222> (1)..(3)  
<223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met

<220>  
<221> VARIANTE  
10 <222> (6)..(8)  
<223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met

<400> 29

15                                   Xaa Xaa Xaa Gly Gly Xaa Xaa Xaa  
                                  1                  5

<210> 30  
<211> 7  
<212> PRT  
20 <213> Secuencia artificial

<220>  
<223> Secuencia de variante de CDR

25 <220>  
<221> VARIANTE  
<222> (1)..(3)  
<223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met

30 <220>  
<221> VARIANTE  
<222> (5)..(7)  
<223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met

35 <400> 30

                                  Xaa Xaa Xaa Gly Xaa Xaa Xaa  
                                  1                  5

<210> 31  
40 <211> 23  
<212> ADN  
<213> Secuencia artificial

<220>  
45 <223> HVK1F1

<400> 31  
gacatccaga tgaccagtc tcc           23

50 <210> 32  
<211> 23  
<212> ADN  
<213> Secuencia artificial

55 <220>  
<223> HVK1 F2

<400> 32  
gmcatccrgw tgaccagtc tcc           23

60 <210> 33  
<211> 23  
<212> ADN



<213> Secuencia artificial  
 <220>  
 <223> HVK2 F  
 5  
 <400> 33  
 gatrttgta tgacycagwc tcc 23  
 <210> 34  
 10 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 15 <223> HVK 3F  
 <400> 34  
 gaaatwgtgw tgacrcagtc tcc 23  
 20 <210> 35  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 25 <220>  
 <223> GAC ATC GTG ATG ACC CAG TCT CC  
 <400> 35  
 30 gacatcgtga tgaccagtc tcc 23  
 <210> 36  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 35 <220>  
 <223> HVK5 F  
 <400> 36  
 40 gaaacgacac tcacgcagtc tcc 23  
 <210> 37  
 <211> 23  
 <212> ADN  
 45 <213> Secuencia artificial  
 <220>  
 <223> HVK6 F  
 50 <400> 37  
 gawrttgtn tgacwcagtc tcc 23  
 <210> 38  
 <211> 24  
 55 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> HVKCL R  
 60 <400> 38  
 acactctccc ctgtgaagc tctt 24  
 <210> 39  
 65 <211> 23  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> HVL1 F1  
 5  
 <400> 39  
 cagtctgtgc tgactcagcc acc 23  
 <210> 40  
 10 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 15 <223> HVL1 F2  
 <400> 40  
 cagtctgtgy tgacgcagcc gcc 23  
 <210> 41  
 20 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 25 <223> HVL2F  
 <400> 41  
 30 cagtctgccc tgactcagcc t 21  
 <210> 42  
 <211> 23  
 <212> ADN  
 35 <213> Secuencia artificial  
 <220>  
 <223> HVL3 F1  
 <400> 42  
 40 tcctatgwgcc tgacwcagcc acc 23  
 <210> 43  
 <211> 23  
 <212> ADN  
 45 <213> Secuencia artificial  
 <220>  
 <223> HVL3 F2  
 <400> 43  
 50 tcttctgagc tgactcagga ccc 23  
 <210> 44  
 <211> 21  
 55 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> HVL4 F1  
 60 <400> 44  
 ctgcctgtgc tgactcagcc c 21  
 <210> 45  
 65 <211> 23  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> HVL4 F2  
 5  
 <400> 45  
 cagcytgtgc tgactcaatc ryc 23  
 <210> 46  
 10 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 15 <223> HVL5F  
 <400> 46  
 cagcctgtgc tgactcagcc 20  
 20 <210> 47  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 25 <220>  
 <223> HVL6 F  
 <400> 47  
 aattttatgc tgactcagcc cca 23  
 30 <210> 48  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 35 <220>  
 <223> HVL7/8F  
 <400> 48  
 40 cagrcgtgtg tgacycagga gcc 23  
 <210> 49  
 <211> 23  
 <212> ADN  
 45 <213> Secuencia artificial  
 <220>  
 <223> HVL9/10F  
 50 <400> 49  
 cagscwgkgc tgactcagcc acc 23  
 <210> 50  
 <211> 23  
 55 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> 01115 HVLCL R  
 60 <400> 50  
 tgaacattct gtaggggcca ctg 23  
 <210> 51  
 65 <211> 23

<212> ADN  
 <213> Secuencia artificial  
  
 5 <220>  
 <223> 01116 HVLCL R2  
  
 <400> 51  
 tgaacattcc gtaggggcaa ctg 23  
  
 10 <210> 52  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial  
  
 15 <220>  
 <223> HVK1 2F1  
  
 <400> 52  
 atctagaatg ggagacggtg acatccagat gacccagtct cc 42  
 20  
 <210> 53  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial  
 25  
 <220>  
 <223> HVK1 2F2  
  
 <400> 53  
 atctagaatg ggagacggtg mcatccrgwt gacccagtct cc 42  
 30  
 <210> 54  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial  
 35  
 <220>  
 <223> HVK2 2F  
  
 <400> 54  
 atctagaatg ggagacggtg mcatccrgwt gacccagtct cc 42  
 40  
 <210> 55  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial  
 45  
 <220>  
 <223> HVK3 2F  
  
 <400> 55  
 atctagaatg ggagacggtg aaatwgtgwt gacrcagtct cc 42  
 50  
 <210> 56  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <223> HVK4 2F  
  
 <400> 56  
 atctagaatg ggagacggtg acatcgtgat gacccagtct cc 42  
 60  
 <210> 57  
 <211> 42  
 65

	<212> ADN	
	<213> Secuencia artificial	
5	<220>	
	<223> HVK5 2F	
	<400> 57	
	atctagaatg ggagacggtg aaacgacact cacgcagtct cc	42
10	<210> 58	
	<211> 42	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> HVK6 2F	
	<400> 58	
	atctagaatg ggagacggtg awrtgtgmt gacwcagtct cc	42
20	<210> 59	
	<211> 51	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> HVKCL 2R	
	<400> 59	
30	gacgagggtc tgagacgatt trathccas yykkgccch bsgccraav t	51
	<210> 60	
	<211> 42	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> HVL1 2F1	
40	<400> 60	
	atctagaatg ggagacggtc agtctgtgct gactcagcca cc	42
	<210> 61	
	<211> 42	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> HVL1 2F2	
50	<400> 61	
	atctagaatg ggagacggtc agtctgtgyt gacgcagccg cc	42
	<210> 62	
55	<211> 40	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> HVL2 2F	
	<400> 62	
	atctagaatg ggagacggtc agtctgccct gactcagcct	40
65	<210> 63	

	<211> 42		
	<212> ADN		
	<213> Secuencia artificial		
5	<220>		
	<223> HVL3 2F1		
	<400> 63		
10	atctagaatg ggagacggtt cctatgwgt gacwcagcca cc	42	
	<210> 64		
	<211> 42		
	<212> ADN		
	<213> Secuencia artificial		
15	<220>		
	<223> HVL3 2F2		
	<400> 64		
20	atctagaatg ggagacggtt cttctgagct gactcaggac cc	42	
	<210> 65		
	<211> 40		
	<212> ADN		
25	<213> Secuencia artificial		
	<220>		
	<223> HVL4 2F1		
30	<400> 65		
	atctagaatg ggagacggtc tgctgtgct gactcagccc	40	
	<210> 66		
	<211> 42		
35	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> HVL4 2F2		
40	<400> 66		
	atctagaatg ggagacggtc agcytgtgt gactcaatc yc	42	
	<210> 67		
45	<211> 39		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
50	<223> HVL5 2F		
	<400> 67		
	atctagaatg ggagacggtc agsctgtgt gactcagcc	39	
55	<210> 68		
	<211> 42		
	<212> ADN		
	<213> Secuencia artificial		
60	<220>		
	<223> HVL6 2F		
	<400> 68		
65	atctagaatg ggagacggtc attttatgt gactcagccc ca	42	

# ES 2 682 254 T3

```

5      <210> 69
      <211> 42
      <212> ADN
      <213> Secuencia artificial

      <220>
      <223> HVL7/8 2F

10     <400> 69
      atctagaatg ggagacgggc agrctgtggt gacycaggag cc      42

      <210> 70
      <211> 42
      <212> ADN
15     <213> Secuencia artificial

      <220>
      <223> HVL9/10 2F

20     <400> 70
      atctagaatg ggagacgggc agscwggkgt gactcagcca cc      42

      <210> 71
      <211> 48
25     <212> ADN
      <213> Secuencia artificial

      <220>
      <223> HVLCL 2R

30     <400> 71
      gatcagggtc tgagacgarr ygrtsasctb sgtbcchbyd ccraabac      48

      <210> 72
      <211> 11
35     <212> PRT
      <213> Homo sapiens

      <400> 72
40
      Val Phe Gly Thr Gly Thr Lys Val Thr Val Ser
      1          5          10

      <210> 73
      <211> 11
45     <212> PRT
      <213> Homo sapiens

      <400> 73

      Val Phe Gly Gly Gly Thr Lys Leu Thr Val Ser
      1          5          10

50
      <210> 74
      <211> 11
      <212> PRT
55     <213> Homo sapiens

      <400> 74

      Val Phe Gly Gly Gly Thr Lys Leu Thr Val Ser
      1          5          10

60
      <210> 75
      <211> 11

```

# ES 2 682 254 T3

<212> PRT  
 <213> Homo sapiens  
  
 <400> 75  
 5                   Val Phe Gly Gly Gly Thr Gln Leu Ile Ile Ser  
                   1                   5                   10  
  
 <210> 76  
 <211> 11  
 10   <212> PRT  
      <213> Homo sapiens  
  
 <400> 76  
  
                   Val Phe Gly Glu Gly Thr Glu Leu Thr Val Ser  
 15                   1                   5                   10  
  
 <210> 77  
 <211> 11  
 <212> PRT  
 20   <213> Homo sapiens  
  
 <400> 77  
  
                   Val Phe Gly Ser Gly Thr Lys Val Thr Val Ser  
 25                   1                   5                   10  
  
 <210> 78  
 <211> 11  
 <212> PRT  
 30   <213> Homo sapiens  
  
 <400> 78  
  
                   Val Phe Gly Gly Gly Thr Gln Leu Thr Ala Ser  
                   1                   5                   10  
  
 35   <210> 79  
      <211> 11  
      <212> PRT  
      <213> Homo sapiens  
  
 40   <400> 79  
  
                   Val Phe Gly Thr Gly Thr Lys Leu Ile Ile Ser  
                   1                   5                   10  
  
 45   <210> 80  
      <211> 105  
      <212> ADN  
      <213> Secuencia artificial  
  
 50   <220>  
      <223> Cebador de PCR  
  
      <220>  
      <221> misc\_feature  
      <222> (56)..(57)  
 55   <223> n es a, c, g, o t  
  
      <220>  
      <221> misc\_feature  
      <222> (59)..(60)  
 60   <223> n es a, c, g, o t



<220>  
 <221> misc\_feature  
 <222> (62)..(63)  
 <223> n es a, c, g, o t  
 5  
 <220>  
 <221> misc\_feature  
 <222> (71)..(72)  
 <223> n es a, c, g, o t  
 10  
 <220>  
 <221> misc\_feature  
 <222> (74)..(75)  
 <223> n es a, c, g, o t  
 15  
 <220>  
 <221> misc\_feature  
 <222> (77)..(78)  
 <223> n es a, c, g, o t  
 20  
 <400> 80  
 gatcagggtc tgagacgaga ccgtcacttt cgtaccggtg ccgaacacca cagtannann 60  
 anntccgccca nnannanngt cccacgcctg acagtaatag tcagc 105  
 25 <210> 81  
 <211> 31  
 <212> PRT  
 <213> Homo sapiens  
 30 <220>  
 <221> VARIANTE  
 <222> (10)..(12)  
 <223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met  
 35 <220>  
 <221> VARIANTE  
 <222> (15)..(17)  
 <223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met  
 40 <400> 81  
 Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Xaa Xaa Xaa Gly Gly Xaa Xaa  
 1 5 10 15  
 Xaa Thr Val Val Phe Gly Thr Gly Thr Lys Val Thr Val Ser Ser  
 20 25 30  
 45 <210> 82  
 <211> 105  
 <212> ADN  
 <213> Homo sapiens  
 50 <220>  
 <221> misc\_feature  
 <222> (65)..(66)  
 <223> n es a, c, g, o t  
 55 <220>  
 <221> misc\_feature  
 <222> (68)..(69)  
 <223> n es a, c, g, o t

<220>  
 <221> misc\_feature  
 <222> (71)..(72)  
 <223> n es a, c, g, o t  
 5  
 <220>  
 <221> misc\_feature  
 <222> (77)..(78)  
 <223> n es a, c, g, o t  
 10  
 <220>  
 <221> misc\_feature  
 <222> (80)..(81)  
 <223> n es a, c, g, o t  
 15  
 <220>  
 <221> misc\_feature  
 <222> (83)..(84)  
 <223> n es a, c, g, o t  
 20  
 <400> 82  
 gatcaggggtc tgagacccgc tgctcacggt aaccatggta ccttgacccc aaatatcaaa 60  
 cgcannanna nngccannan nanntttcgc acagtagtaa acagc 105  
 25  
 <210> 83  
 <211> 25  
 <212> PRT  
 <213> Homo sapiens  
 30  
 <220>  
 <221> VARIANTE  
 <222> (7)..(9)  
 <223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met  
 35  
 <220>  
 <221> VARIANTE  
 <222> (11)..(13)  
 <223> X es cualquier aminoácido que no sea Trp, Gln, Lys, Glu, Met  
 40  
 <400> 83  
 Val Tyr Tyr Cys Ala Lys Xaa Xaa Xaa Gly Xaa Xaa Xaa Ala Phe Asp  
 1 5 10 15  
 Ile Trp Gly Gln Gly Thr Met Val Thr  
 20 25  
 45  
 <210> 84  
 <211> 771  
 <212> ADN  
 <213> Homo sapiens  
 50  
 <220>  
 <221> misc\_feature  
 <222> (286)..(287)  
 <223> n es a, c, g, o t  
 55  
 <220>  
 <221> misc\_feature  
 <222> (289)..(290)  
 <223> n es a, c, g, o t

5  
 <220>  
 <221> misc\_feature  
 <222> (292)..(293)  
 <223> n es a, c, g, o t  
  
 10  
 <220>  
 <221> misc\_feature  
 <222> (301)..(302)  
 <223> n es a, c, g, o t  
  
 15  
 <220>  
 <221> misc\_feature  
 <222> (304)..(305)  
 <223> n es a, c, g, o t  
  
 20  
 <220>  
 <221> misc\_feature  
 <222> (307)..(308)  
 <223> n es a, c, g, o t  
  
 25  
 <220>  
 <221> misc\_feature  
 <222> (706)..(707)  
 <223> n es a, c, g, o t  
  
 30  
 <220>  
 <221> misc\_feature  
 <222> (709)..(710)  
 <223> n es a, c, g, o t  
  
 35  
 <220>  
 <221> misc\_feature  
 <222> (712)..(713)  
 <223> n es a, c, g, o t  
  
 40  
 <220>  
 <221> misc\_feature  
 <222> (718)..(719)  
 <223> n es a, c, g, o t  
  
 45  
 <220>  
 <221> misc\_feature  
 <222> (721)..(722)  
 <223> n es a, c, g, o t  
  
 50  
 <220>  
 <221> misc\_feature  
 <222> (724)..(725)  
 <223> n es a, c, g, o t  
  
 <400> 84

# ES 2 682 254 T3

```

atgggagacg gtcagtctgt gctgactcag ccaccctcag tgtccgtgtc cccaggacag      60
acagccagca tcacctgctc tggagataaa ttgggggata aatatgcttg ctggtatcag      120
cagaagccag gccagtcccc tgtgctggtc atctatcaag atagcaagcg gccctcaggg      180
atccctgagc gattctctgg ctccaactct gggaacacag ccaactctgac catcagcggg      240
accagggcta tggatgaggc tgactattac tgtcaggcgt gggacnntnn tnntggaggt      300
nntnntnnta ctgtggtggt cggcacgggc accaagctca tcatttcgtc tcagaccggt      360
ggttctggtg gtggtggttc tggcggcggc ggctccggtg gtggtggatc cgaagtccaa      420
ctgctggagt ccggcggtag cctggtgcag ccagggtggca gcctgcgcct gagctgcgcc      480
gcatccggtt ttactttcag cagctacgcg atgtcgtggg tgcgccaggc accgggcaag      540
ggcctggagt gggtcagcgc catcagcggg agcggcgggt ctacgtatta tgcggacagc      600
gtcaagggcc gtttcaccat cagccgtgac aattccaaaa acaccctgta cttgcagatg      660
aacagcttgc gtgcggaaga tacggctggt tactactgtg cgaaanntnn tnntggannt      720
nntnntgcct ttgatatttg gggtaaggt accatggta ccgtgagcag c              771

```

<210> 85  
 <211> 257  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <222> (96)..(98)  
 <223> Xaa puede ser cualquier aminoácido de origen natural

<220>  
 <221> misc\_feature  
 <222> (101)..(103)  
 <223> Xaa puede ser cualquier aminoácido de origen natural

<220>  
 <221> misc\_feature  
 <222> (236)..(238)  
 <223> Xaa puede ser cualquier aminoácido de origen natural

<220>  
 <221> misc\_feature  
 <222> (240)..(242)  
 <223> Xaa puede ser cualquier aminoácido de origen natural

<400> 85

# ES 2 682 254 T3

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

# ES 2 682 254 T3

130

135

140

Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala  
145 150 155 160

Ala Ser Gly Phe Thr Phe Ser Ser Tyr Ala Met Ser Trp Val Arg Gln  
165 170 175

Ala Pro Gly Lys Gly Leu Glu Trp Val Ser Ala Ile Ser Gly Ser Gly  
180 185 190

Gly Ser Thr Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser  
195 200 205

Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg  
210 215 220

Ala Glu Asp Thr Ala Val Tyr Tyr Cys Ala Lys Xaa Xaa Xaa Gly Xaa  
225 230 235 240

Xaa Xaa Ala Phe Asp Ile Trp Gly Gln Gly Thr Met Val Thr Val Ser  
245 250 255

Ser

<210> 86  
<211> 8  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<222> (1)..(3)  
<223> Xaa puede ser cualquier aminoácido de origen natural

<220>  
<221> misc\_feature  
<222> (6)..(8)  
<223> Xaa puede ser cualquier aminoácido de origen natural

<400> 86

Xaa Xaa Xaa Gly Gly Xaa Xaa Xaa  
1 5

<210> 87  
<211> 7  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<222> (1)..(3)  
<223> Xaa puede ser cualquier aminoácido de origen natural

<220>

<221> misc\_feature  
 <222> (5)..(7)  
 <223> Xaa puede ser cualquier aminoácido de origen natural

5 <400> 87

Xaa Xaa Xaa Gly Xaa Xaa Xaa  
 1 5

10 <210> 88  
 <211> 361  
 <212> ADN  
 <213> Homo sapiens

15 <400> 88  
 atgggagacg gtcagtctgt gctgactcag ccaccctcag tgtccgtgtc cccaggacag 60  
 acagccagca tcacctgctc tggagataaa ttgggggata aatatgcttg ctggtatcag 120  
 cagaagccag gccagtcccc tgtgctgggc atctatcaag atagcaagcg gccctcaggg 180  
 atccctgagc gattctctgg ctccaactct gggaacacag ccaactctgac catcagcggg 240  
 acccaggcta tggatgaggg tgactattac tgtcaggcgt gggactgaga cctagacggt 300  
 ctctgcgttt gatatttggg gtcaagggtac catggttacc gtgagcagct cgtctcagac 360  
 c 361

20 <210> 89  
 <211> 396  
 <212> ADN  
 <213> Homo sapiens

<400> 89  
 actgtggtgt tcggcaccgg tacgaaagtg actgtctcat ctcagaccgg tggttctggt 60  
 ggtggtggtt ctggcggcgg cggctccggt ggtggtggat ccgaagtcca actgctggag 120  
 tccggcgggtg gcctggtgca gccagggtggc agcctgcgcc tgagctgcgc cgcattccggt 180  
 ttacttttca gcagctacgc gatgtcgtgg gtgcgccagg caccgggcaa gggcctggag 240  
 tgggtcagcg ccatcagcgg tagcggcgggt tctacgtatt atgcggacag cgtcaagggc 300  
 cgtttcacca tcagccgtga caattccaaa aacaccctgt acttgcagat gaacagcttg 360  
 25 cgtgcggaag atacggctgt ttactactgt gcgaaa 396

30 <210> 90  
 <211> 71  
 <212> ADN  
 <213> Homo sapiens

35 <220>  
 <221> misc\_feature  
 <222> (18)..(19)  
 <223> n es a, c, g, o t

<220>  
 <221> misc\_feature  
 <222> (21)..(22)

<223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 5 <222> (24)..(25)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 10 <222> (33)..(34)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 15 <222> (36)..(37)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 20 <222> (39)..(40)  
 <223> n es a, c, g, o t  
 <400> 90  
 gatcagggtc tcaggacnnt nntnntggcg ganntnntnn tactgtggtg ttcggcaccg 60  
 25 gtacgaaagt g 71  
 <210> 91  
 <211> 68  
 <212> ADN  
 30 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 35 <222> (19)..(20)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 40 <222> (22)..(23)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 45 <222> (25)..(26)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 50 <222> (31)..(32)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 55 <222> (34)..(35)  
 <223> n es a, c, g, o t  
 <220>  
 <221> misc\_feature  
 60 <222> (37)..(38)  
 <223> n es a, c, g, o t  
 <400> 91



# ES 2 682 254 T3

	gatcagggtc tcaacgcann annanngcca nnannanntt tcgcacagta gtaaacagcc	60
	gtatcttc	68
5	<210> 92 <211> 8 <212> PRT <213> Homo sapiens  <400> 92	
10	Pro Phe Gly Gly Gly Gly Tyr Val 1 5	
15	<210> 93 <211> 7 <212> PRT <213> Homo sapiens  <400> 93	
20	Pro Pro His Gly Ala Pro Ala 1 5	
25	<210> 94 <211> 8 <212> PRT <213> Homo sapiens  <400> 94	
30	Leu Cys Ile Gly Gly Val Ala Ser 1 5	
35	<210> 95 <211> 7 <212> PRT <213> Homo sapiens  <400> 95	
40	His Asn Ser Gly Asn Asn Phe 1 5	
45	<210> 96 <211> 8 <212> PRT <213> Homo sapiens  <400> 96	
50	Phe Val Ser Gly Gly Ile Ser Thr 1 5	
55	<210> 97 <211> 7 <212> PRT <213> Homo sapiens  <400> 97	
	Phe Asn Phe Gly Asn Ala Tyr 1 5	
	<210> 98	

# ES 2 682 254 T3

<211> 8  
 <212> PRT  
 <213> Homo sapiens  
 5 <400> 98  
 Ile Asn Ser Gly Gly Ala Ser Phe  
 1 5  
 <210> 99  
 10 <211> 7  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 15 <221> misc\_feature  
 <222> (1)..(3)  
 <223> Xaa puede ser cualquier aminoácido de origen natural  
 <400> 99  
 20 Xaa Xaa Xaa Gly Thr Asn Tyr  
 1 5  
 <210> 100  
 <211> 8  
 25 <212> PRT  
 <213> Homo sapiens  
 <400> 100  
 Ser Arg Ala Gly Gly Cys Asn Gly  
 30 1 5  
 <210> 101  
 <211> 7  
 <212> PRT  
 35 <213> Homo sapiens  
 <400> 101  
 Phe Asp Tyr Gly His Cys Ile  
 40 1 5  
 <210> 102  
 <211> 8  
 <212> PRT  
 <213> Homo sapiens  
 45 <400> 102  
 Thr Asn Arg Gly Gly Val Cys Ala  
 1 5  
 <210> 103  
 <211> 7  
 <212> PRT  
 <213> Homo sapiens  
 50 <400> 103  
 Thr Ala Ala Gly Val Pro Phe  
 1 5  
 <210> 104

# ES 2 682 254 T3

<211> 8  
 <212> PRT  
 <213> Homo sapiens  
 5 <400> 104  
 Phe Ser Thr Gly Gly Cys Ala Phe  
 1 5  
 <210> 105  
 10 <211> 7  
 <212> PRT  
 <213> Homo sapiens  
 <400> 105  
 15 Ala Ile Cys Gly Ala Thr Ala  
 1 5  
 <210> 106  
 <211> 8  
 20 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 25 <222> (2)..(2)  
 <223> Xaa puede ser cualquier aminoácido de origen natural  
 <400> 106  
 Phe Xaa Gly Gly Gly Asp Gly Thr  
 30 1 5  
 <210> 107  
 <211> 7  
 <212> PRT  
 35 <213> Homo sapiens  
 <400> 107  
 Pro Tyr Arg Gly Ser Phe Phe  
 40 1 5  
 <210> 108  
 <211> 8  
 <212> PRT  
 45 <213> Homo sapiens  
 <400> 108  
 Ile Ile Pro Gly Gly Leu Tyr Ala  
 50 1 5  
 <210> 109  
 <211> 7  
 <212> PRT  
 <213> Homo sapiens  
 55 <400> 109  
 Pro Val Ile Gly Ser Asn Thr  
 1 5  
 <210> 110  
 60 <211> 254

<212> PRT

<213> Aequorea victoria

<400> 110

5

# ES 2 682 254 T3

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

# ES 2 682 254 T3

<210> 111  
 <211> 266  
 <212> PRT  
 <213> Homo sapiens

5

<400> 111

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val  
 1 5 10 15

Ser Pro Gly Gln Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly  
 20 25 30

Asp Lys Tyr Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val  
 35 40 45

# ES 2 682 254 T3

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

<210> 112

<211> 1176

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Construcción de fusión

10

# ES 2 682 254 T3

<400> 112

atgacgagca aagaaacctt taccattac cagccgcagg gcaacagtga cccggctcat	60
accgcaaccg cgcccggcgg attgagtgcg aaagcgcctg caatgacccc gctgatgctg	120
gacacctcca gccgtaagct ggttgcgctg gatggcacca ccgacgggtg tgcggttggc	180
attcttgccg ttgctgctga ccagaccagc accacgctga cgttctacaa gtccggcacg	240
ttccgttatg aggatgtgct ctggccggag gctgccagcg acgagacgaa aaaacggacc	300
gcgtttgccg gaacggcaat cagcatcggt ggaggtagcg gcggatcgga tgacgacgat	360
aagtctagaa atggcggaga cggtcagtct gtgctgactc agccaccctc agtgtccgtg	420
tccccaggac agacagccag catcacctgc tctggagata aattggggga taaatatgct	480
tgctggtatc agcagaagcc aggccagtcc cctgtgctgg tcatctatca agatagcaag	540
cggccctcag ggatccctga gcgattctct ggctccaact ctgggaacac agccactctg	600
accatcagcg ggaccagggc tatggatgag gctgactatt actgtcaggc gtgggacttt	660
aatcttgccg gatgtggtga tactgtggtg ttccggcaccg gtacgaaagt gactgtctca	720
tctcagaccg gtggttctgg tggcggtggt tctggcggcg gcggctccgg tgggtggtgga	780
tccgaagtcc aactgctgga gtccggtggt ggctggtgc agccaggtgg cagcctgcgc	840
ctgagctgcg ccgcatccgg ttttactttc agcagctacg cgatgtcgtg ggtgcgccag	900
gcaccgggca agggcctgga gtgggtcagc gccatcagcg gtagcggcgg ttctacgtat	960
tatgcggaca gcgtcaaggg ccgtttcacc atcagccgtg acaattccaa aaacaccctg	1020
tacttgcala tgaacagctt gcgtgcggaa gatacggctg tttactactg tgcgaaacat	1080
attgatggcc ctggtgctgc gtttgatatt tggggtaag gtaccatggt taccgtgagc	1140
aactcgagcg attacaagga cgatgatgac aaataa	1176

5 <210> 113  
 <211> 391  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <223> Proteína de fusion  
 <400> 113

Met	Thr	Ser	Lys	Glu	Thr	Phe	Thr	His	Tyr	Gln	Pro	Gln	Gly	Asn	Ser
1				5					10					15	

Asp	Pro	Ala	His	Thr	Ala	Thr	Ala	Pro	Gly	Gly	Leu	Ser	Ala	Lys	Ala
			20					25					30		

15 Pro Ala Met Thr Pro Leu Met Leu Asp Thr Ser Ser Arg Lys Leu Val



# ES 2 682 254 T3

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

# ES 2 682 254 T3

Thr Phe Ser Ser Tyr Ala Met Ser Trp Val Arg Gln Ala Pro Gly Lys  
290 295 300

Gly Leu Glu Trp Val Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr  
305 310 315 320

Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser  
325 330 335

Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr  
340 345 350

Ala Val Tyr Tyr Cys Ala Lys His Ile Asp Gly Pro Val Ala Ala Phe  
355 360 365

Asp Ile Trp Gly Gln Gly Thr Met Val Thr Val Ser Asn Ser Ser Asp  
370 375 380

Tyr Lys Asp Asp Asp Asp Lys  
385 390

<210> 114  
<211> 89  
<212> PRT  
<213> Homo sapiens

<400> 114

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ser Pro Gly Gln  
1 5 10 15

Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly Asp Lys Tyr Ala  
20 25 30

Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val Leu Val Ile Tyr  
35 40 45

Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser  
50 55 60

Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly Thr Gln Ala Met  
65 70 75 80

Asp Glu Ala Asp Tyr Tyr Cys Gln Ala  
85

<210> 115  
<211> 93  
<212> PRT  
<213> Homo sapiens

<400> 115

# ES 2 682 254 T3

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val  
1 5 10 15

Ser Pro Gly Gln Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly  
20 25 30

Asp Lys Tyr Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val  
35 40 45

Leu Val Ile Tyr Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg  
50 55 60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly  
65 70 75 80

Thr Gln Ala Met Asp Glu Ala Asp Tyr Tyr Cys Gln Ala  
85 90

<210> 116  
<211> 93  
<212> PRT  
<213> Homo sapiens

<400> 116

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val  
1 5 10 15

Ser Pro Gly Gln Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly  
20 25 30

Asp Lys Tyr Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val  
35 40 45

Leu Val Ile Tyr Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg  
50 55 60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly  
65 70 75 80

Thr Gln Ala Met Asp Glu Ala His Tyr Tyr Cys Gln Thr  
85 90

<210> 117  
<211> 93  
<212> PRT  
<213> Homo sapiens

<400> 117

# ES 2 682 254 T3

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val  
1 5 10 15

Ser Pro Gly Gln Thr Ala Thr Ile Thr Cys Ser Gly Asp Lys Leu Gly  
20 25 30

Asp Lys Phe Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val  
35 40 45

Leu Val Ile Tyr Gln Asp Tyr Gln Arg Pro Ser Gly Ile Pro Glu Arg  
50 55 60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly  
65 70 75 80

Thr Gln Val Met Asp Glu Ala Asp Tyr Tyr Cys Gln Ala  
85 90

<210> 118

<211> 87

5 <212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

10 <222> (41)..(41)

<223> Xaa puede ser cualquier aminoácido de origen natural

<220>

<221> misc\_feature

15 <222> (61)..(61)

<223> Xaa puede ser cualquier aminoácido de origen natural

<400> 118

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

Ala Pro Glu Lys Thr Ala Thr Ile Ala Cys Gly Gly Asn Arg Ile Gly  
20 25 30

Ser Lys Ser Val His Trp Tyr Gln Xaa Lys Pro Gly Gln Ala Pro Val  
35 40 45

Leu Val Ile Tyr Asn Asp Asn Asp Arg Pro Ser Gly Xaa Pro Glu Arg  
50 55 60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Arg  
65 70 75 80

Val Glu Val Gly Asp Glu Ala  
85

20

<210> 119

# ES 2 682 254 T3

<211> 87  
<212> PRT  
<213> Homo sapiens

5 <400> 119

```

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val
1          5          10          15

Ala Pro Gly Lys Thr Ala Thr Ile Ala Cys Gly Gly Asn Arg Ile Gly
          20          25          30

Ser Lys Ser Val His Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Val
          35          40          45

Leu Val Ile Tyr Asn Asp Asn Asp Arg Pro Ser Gly Ile Pro Glu Arg
          50          55          60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Arg
65          70          75          80

Val Glu Val Gly Asp Glu Ala
          85

```

<210> 120  
<211> 83  
<212> PRT  
<213> Homo sapiens

10

<400> 120

15

```

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ala Pro Gly Lys
1          5          10          15

Thr Ala Arg Ile Thr Cys Gly Gly Asn Asn Ile Gly Ser Lys Ser Val
          20          25          30

His Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Val Leu Val Ile Tyr
          35          40          45

Tyr Asp Ser Asp Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser
50          55          60

Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Arg Val Glu Ala Gly
65          70          75          80

Asp Glu Ala

```

<210> 121  
<211> 87  
<212> PRT  
<213> Homo sapiens

20

<400> 121

# ES 2 682 254 T3

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val  
1 5 10 15

Ala Pro Gly Lys Thr Ala Thr Ile Thr Cys Gly Gly Asn Asn Ile Glu  
20 25 30

Asp Lys Ser Val His Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Val  
35 40 45

Leu Val Ile Tyr Tyr Asp Thr Asp Arg Pro Ser Gly Ile Pro Glu Arg  
50 55 60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Asn Arg  
65 70 75 80

Val Glu Ala Gly Asp Glu Ala  
85

<210> 122  
<211> 95  
<212> PRT  
<213> Homo sapiens

<400> 122

Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu Ser Pro Gly Lys  
1 5 10 15

Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser Ile Ala Ser Asn  
20 25 30

Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Ala Pro Thr Thr Val  
35 40 45

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

Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu Thr Ile Ser Gly  
65 70 75 80

Leu Lys Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln Ser Tyr Asp

85

90

95

<210> 123  
<211> 99  
<212> PRT  
<213> Homo sapiens

<400> 123

# ES 2 682 254 T3

Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu  
1 5 10 15

Ser Pro Gly Lys Thr Val Thr Ile Pro Cys Thr Gly Ser Ser Gly Ser  
20 25 30

Ile Ala Ser Asn Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Ala  
35 40 45

Pro Thr Thr Val Ile Tyr Glu Asp Lys Gln Lys Pro Ser Gly Val Pro  
50 55 60

Asp Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu  
65 70 75 80

Thr Ile Ser Gly Leu Lys Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln  
85 90 95

Ser Tyr Asp

<210> 124

<211> 99

<212> PRT

<213> Homo sapiens

<400> 124

Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu  
1 5 10 15

Ser Pro Gly Lys Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser  
20 25 30

Ile Ala Ser Asn Ser Val Gln Arg Tyr Gln Gln Arg Pro Gly Ser Ala  
35 40 45

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

Asp Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu  
65 70 75 80

Ile Ile Ser Gly Leu Gln Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln  
85 90 95

Ser Tyr Asp

<210> 125

<211> 99

<212> PRT

<213> Homo sapiens

# ES 2 682 254 T3

<400> 125

Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu  
1 5 10 15

Ser Pro Gly Lys Thr Val Thr Ile Ser Cys Thr Gly Arg Ser Gly Ser  
20 25 30

Ile Ala Asp Asn Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Val  
35 40 45

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

Glu Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu  
65 70 75 80

Thr Ile Ser Gly Leu Met Thr Glu Asp Glu Ala Asp Tyr His Cys Gln  
85 90 95

Ser Tyr Asp



## REIVINDICACIONES

1. Una biblioteca de polipéptidos que comprende una pluralidad de polipéptidos diferentes, que comprenden:

- 5 i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y  
 i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

15 en donde al menos dos de los polipéptidos difieren entre sí en la secuencia de aminoácidos presente en una o más regiones determinantes de complementariedad (CDR, del inglés *complementary determining regions*) en las regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

20 2. Un método de construir una biblioteca de polipéptidos que es soluble en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivas para la formación de enlaces disulfuro, comprendiendo el método la preparación de una pluralidad de diferentes polipéptidos, que comprenden:

- 25 i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y  
 i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

35 en donde al menos dos de los polipéptidos difieren entre sí en la secuencia de aminoácidos presente en una o más CDR en las regiones variables  $V_H$  y/o  $V_L$ .

3. Una biblioteca de polinucleótidos que comprende una pluralidad de polinucleótidos diferentes, en donde cada polinucleótido codifica un polipéptido que comprende:

- 40 i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y  
 i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

45 en donde al menos dos de los polinucleótidos difieren entre sí codificando polipéptidos que comprenden una o más CDR diferentes en las regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

50 4. Un método de construcción de una biblioteca de polinucleótidos, comprendiendo el método la preparación de una pluralidad de diferentes polinucleótidos que codifican un polipéptido, que comprende:

- 55 i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y  
 i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde la  $V_H$  y la  $V_L$  son capaces de formar un sitio de unión a antígeno; y

65 en donde al menos dos de los polinucleótidos difieren entre sí codificando polipéptidos que comprenden una o más

CDR diferentes en las regiones variables  $V_H$  y/o  $V_L$ , y en donde los polipéptidos son solubles en condiciones reductoras, que no son suficientes para la oxidación de grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

5. Un polipéptido aislado y/o recombinante que comprende:

i) una región variable de cadena pesada de anticuerpo ( $V_H$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de IGHV3-23 tal como se establece en la SEQ ID NO: 3; y

10 i) una región variable de cadena ligera de anticuerpo ( $V_L$ ) que comprende una región de armazón que es al menos el 95 % idéntica a la región de armazón de uno cualquiera de IGLV1-40 (tal como se establece en la SEQ ID NO: 18), IGLV1-44 (tal como se establece en la SEQ ID NO: 21), IGLV1-47 (tal como se establece en la SEQ ID NO: 24), IGLV3-1 (tal como se establece en la SEQ ID NO: 6), IGLV3-19 (tal como se establece en la SEQ ID NO: 27), IGLV6-57 (tal como se establece en la SEQ ID NO: 12); en donde el  $V_H$  y el  $V_L$  son capaces de formar un sitio de unión a antígeno, y

15 en donde el polipéptido es soluble en condiciones reductoras, que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

20 6. El polipéptido de la reivindicación 5, que es un fragmento variable (Fv) y/o en donde el polipéptido es un scFv y el  $V_H$  y el  $V_L$  están unidos entre sí mediante un enlazador peptídico.

7. El polipéptido de la reivindicación 5, que es un fragmento Fab, un fragmento Fab', un fragmento F(ab'), un scFv, un diacuerpo, un triacuerpo, un tetracuerpo o un complejo polipeptídico de mayor orden.

25 8. El polipéptido de una cualquiera de las reivindicaciones 5-7, que es soluble y capaz de formar de manera estable un sitio de unión a antígeno cuando se produce en condiciones reductoras que no son suficientes para la oxidación de los grupos sulfhidrilo (-SH) en una proteína y no son permisivos para la formación de enlaces disulfuro.

30 9. Un polinucleótido aislado y/o exógeno que codifica el polipéptido de una cualquiera de las reivindicaciones 5-8.

10. Un método de selección sistemática de un polipéptido que se une a una molécula diana, comprendiendo el método poner en contacto un polipéptido de una cualquiera de las reivindicaciones 5-8 o la biblioteca de la reivindicación 1 con la molécula diana, y determinar si el polipéptido de una cualquiera de las reivindicaciones 5-8 se une a la molécula diana.

35 11. El método de la reivindicación 10, en donde un polinucleótido que codifica el polipéptido de una cualquiera de las reivindicaciones 5-8 se expresa en una célula hospedadora o en un sistema de expresión sin célula para producir el polipéptido de una cualquiera de las reivindicaciones 5-8.

40 12. El método de la reivindicación 11, en donde el polinucleótido se expresa en el citoplasma y/o en el periplasma de una célula hospedadora, y en donde la célula hospedadora es una célula bacteriana, una célula de levadura o una célula de mamífero.

45 13. El método de la reivindicación 12, en donde la célula hospedadora es una célula bacteriana y el método comprende:

a) cultivar una célula bacteriana que comprende un polinucleótido que codifica un polipéptido de una cualquiera de las reivindicaciones 5-8, de manera que se produce el polipéptido,

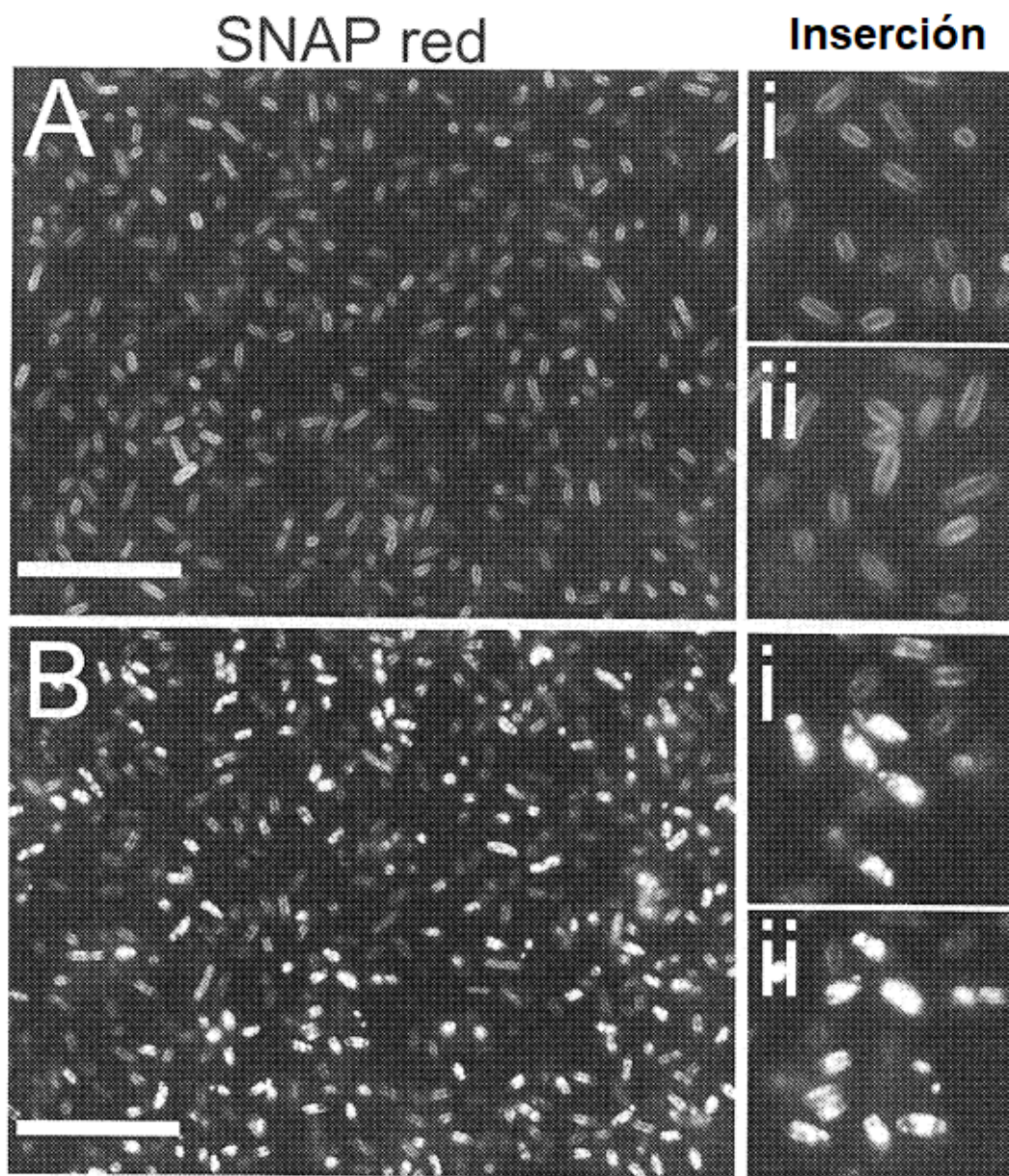
50 b) permeabilizar la célula bacteriana, en donde el polinucleótido y el polipéptido se mantiene dentro de la célula bacteriana permeabilizada,

c) poner en contacto la célula bacteriana permeabilizada con la molécula diana de manera que difunda a la célula bacteriana permeabilizada, y

d) determinar si el polipéptido de una cualquiera de las reivindicaciones 5-8 se une a la molécula diana.

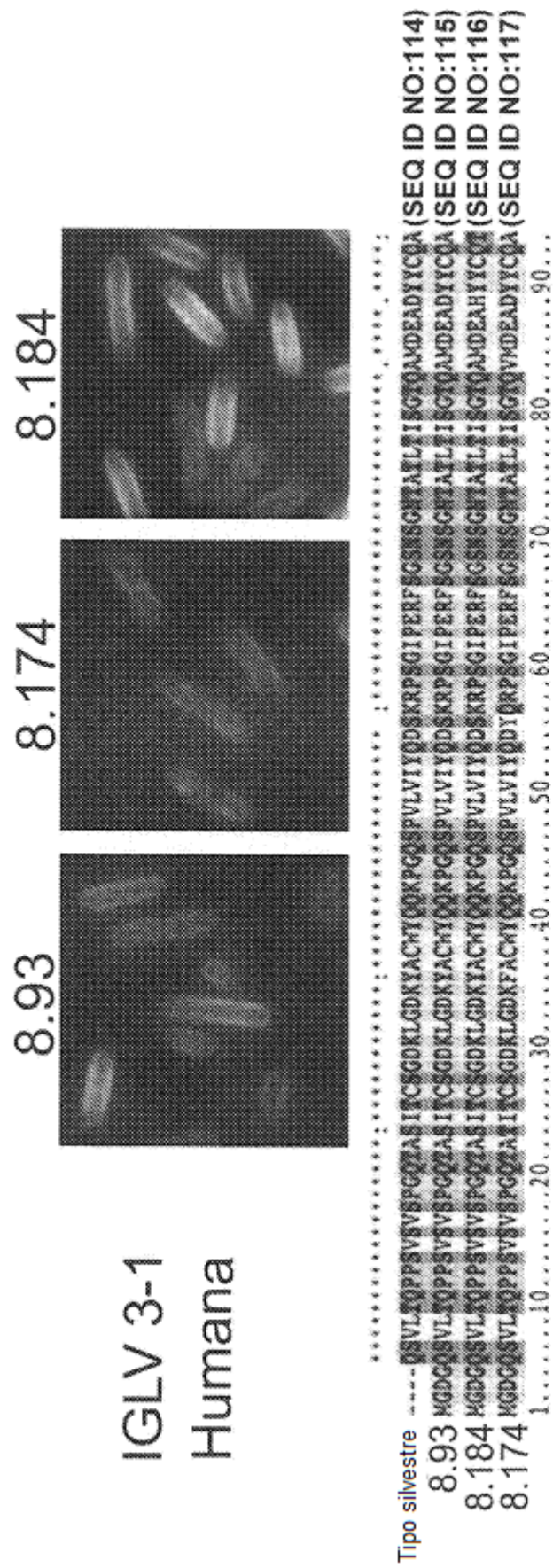
55 14. Una biblioteca de células hospedadoras que comprende una pluralidad de células hospedadoras que comprenden un polinucleótido de una cualquiera de las reivindicaciones 5-8, en donde al menos una célula hospedadora comprende un polipéptido que difiere de un polipéptido presente en otra célula hospedadora en la biblioteca en la secuencia de aminoácidos presente en una o más CDR en los dominios variables  $V_H$  y/o  $V_L$ .

60 15. Una composición que comprende el polipéptido de una cualquiera de las reivindicaciones 5-8, y/o el polinucleótido de la reivindicación 9, y un vehículo farmacéuticamente aceptable.



**Figura 1**





## Figura 2

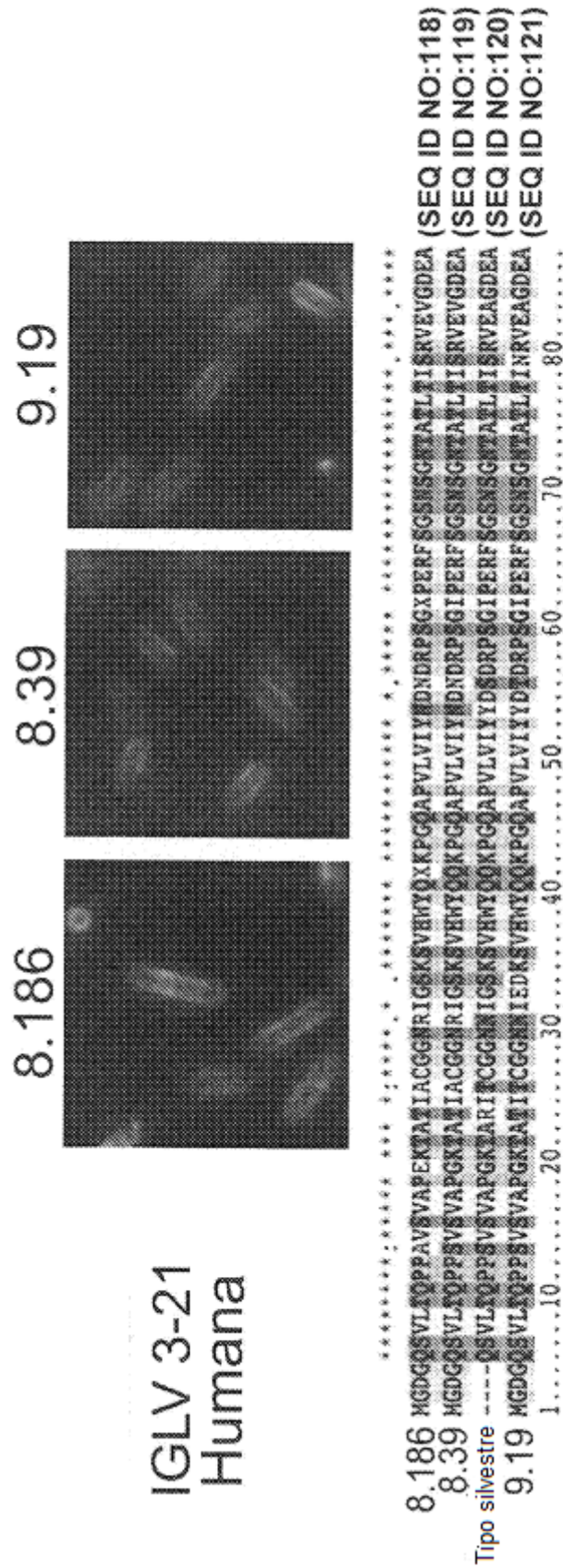


Figura 2 (continuación)

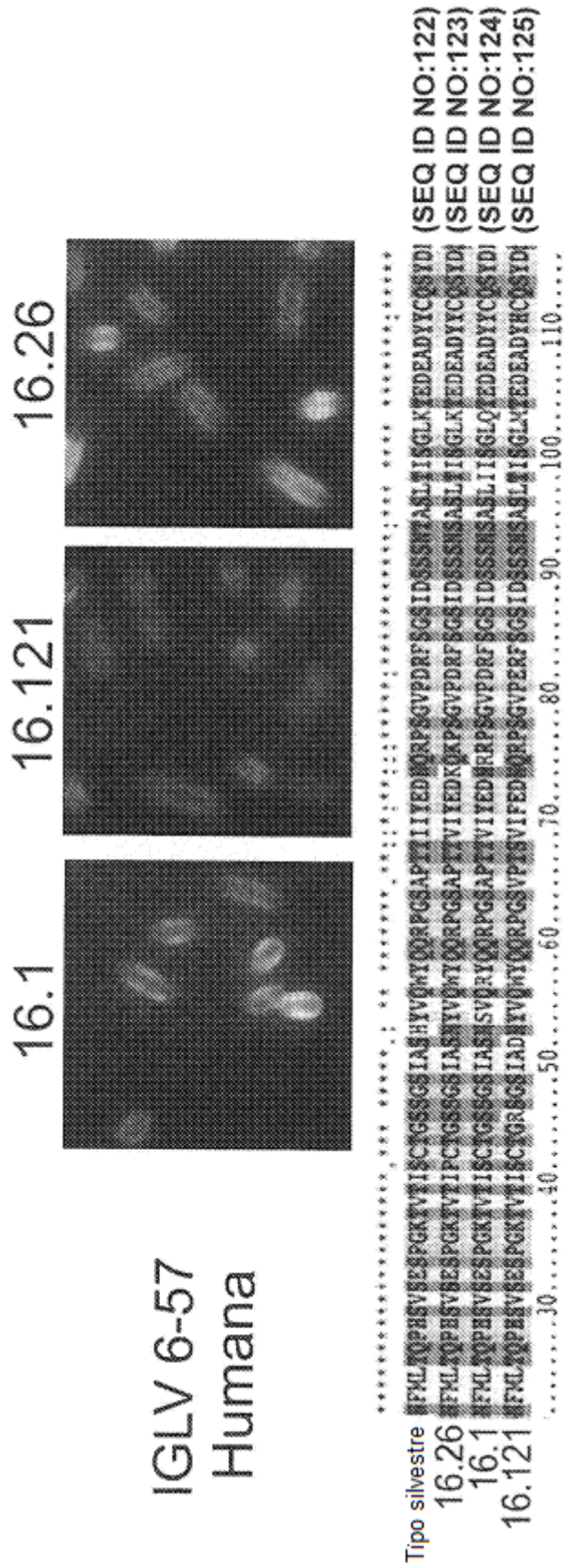
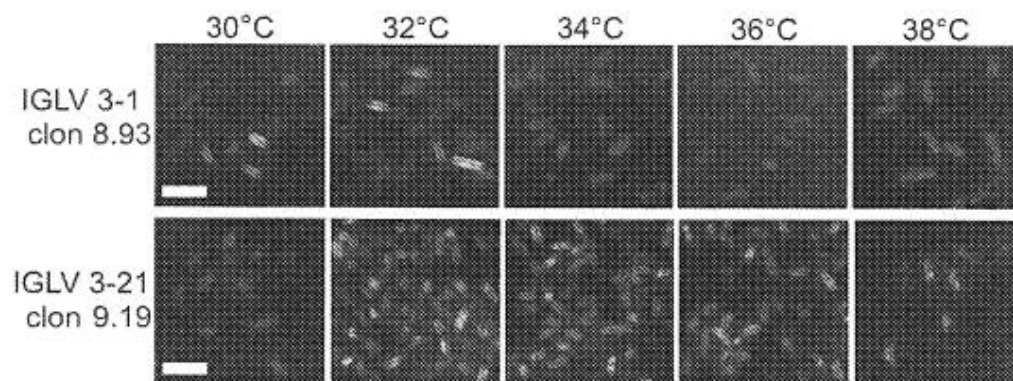
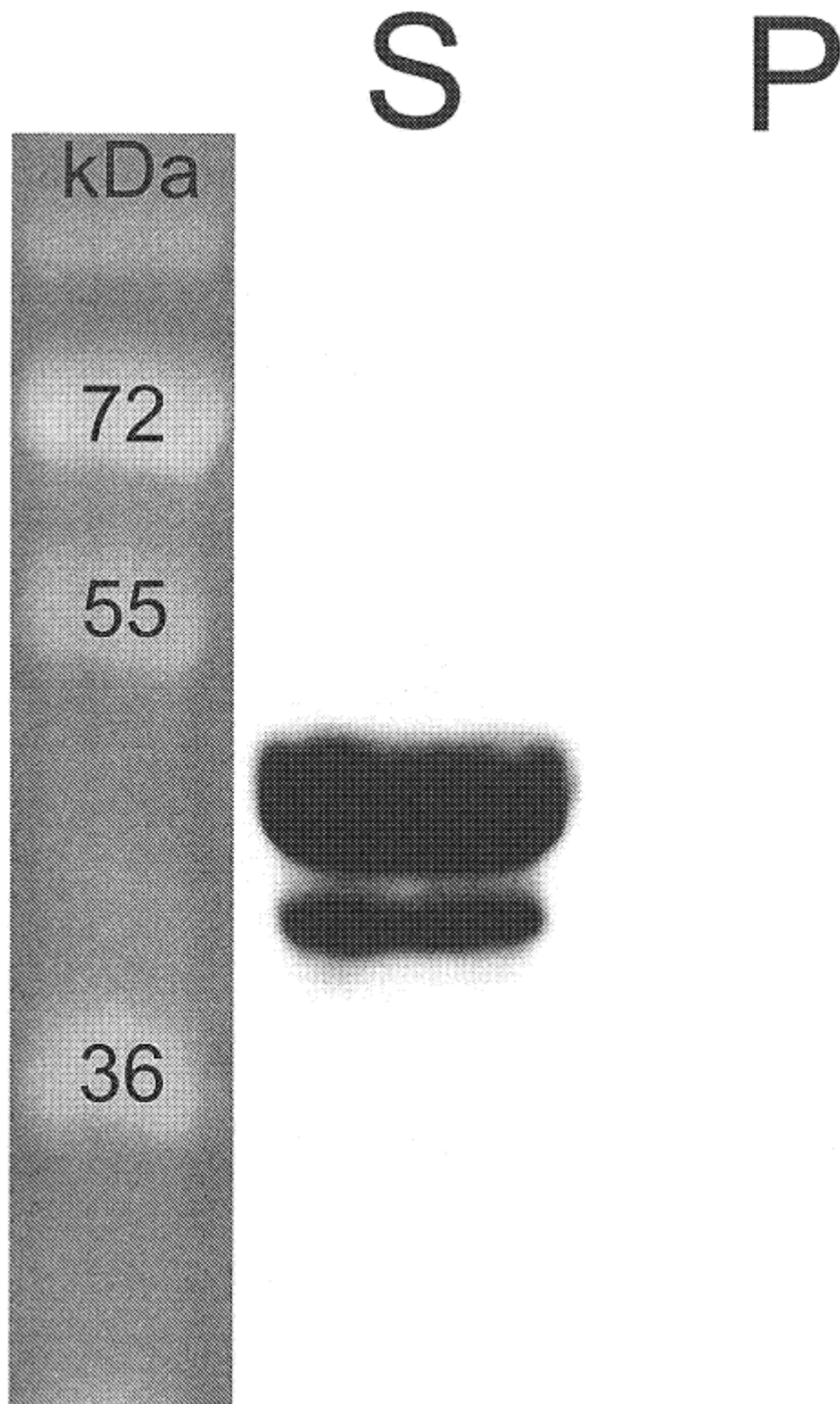


Figura 2 (continuación)



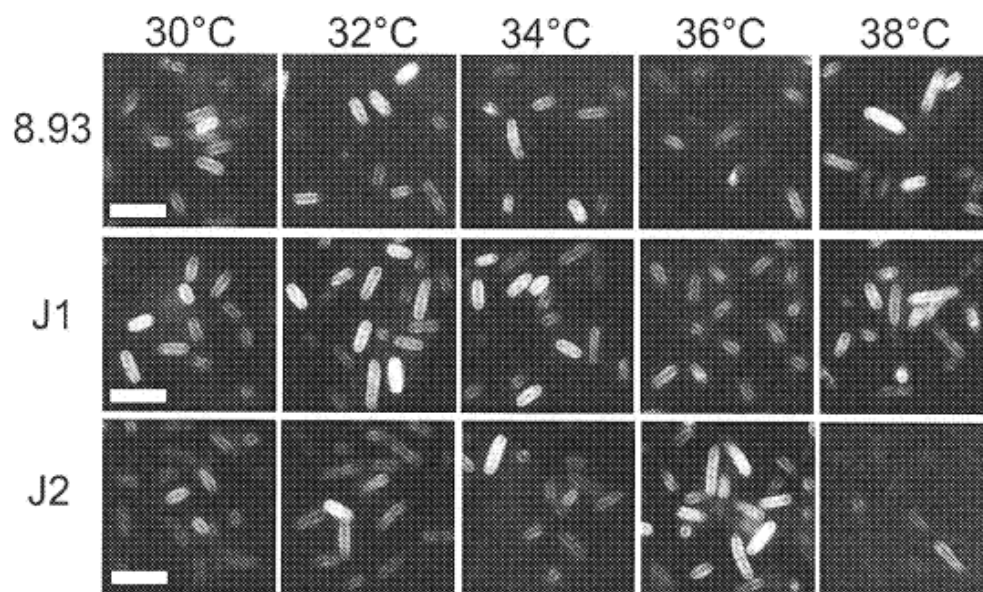


**Figura 3**

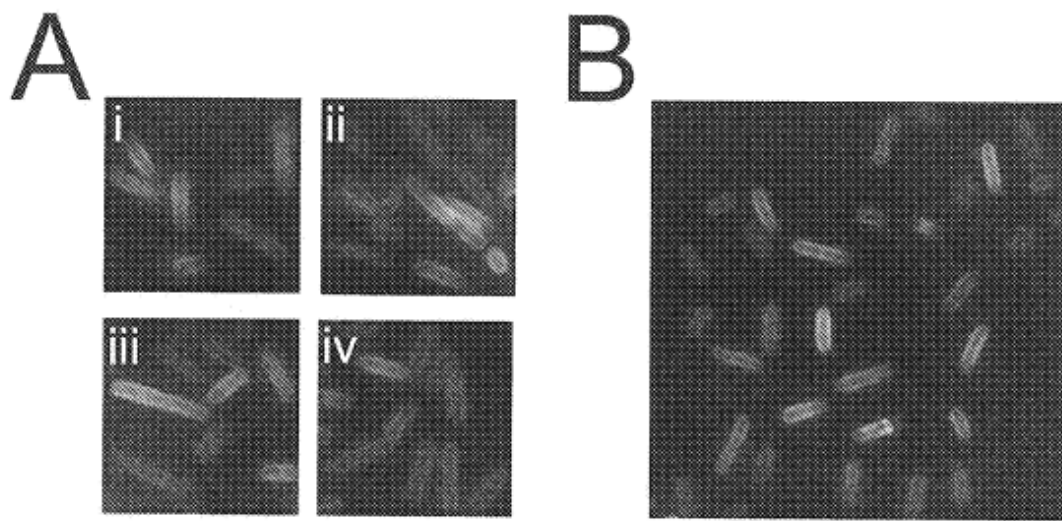


**Figura 4**





**Figura 5**



**Figura 6**

IGLV1-51(aka DPL5) (SEQ ID NO: 15)  
 QSVLTQPPSVSAAPGQKVITSCS**GSSSNIGNN**YVSWYQQLPGTAPKLLIY**DN**NKRPSGIPDRFSGSKSGTSATLGITGLQTGDEADYYCGT**WDSSLSAG**

IGLV1-40 (aka DPL6) (SEQ ID NO: 18)  
 QSVLTQPPSVSGAPGQRTVITSC**TGSSSNIGAGYD**VHWYQQLPGTAPKLLIY**GNSNRPS**GVDPDRFSGSKSGTSASLAITGLQAEDEADYYCQS**YDSSLSGS**

IGLV1-44 (aka DPL2) (SEQ ID NO: 21)  
 QSVLTQPPSASGTPGQRTVITSCS**GSSSNIGSNT**VNWYQQLPGTAPKLLIY**SNNQRPS**GVDPDRFSGSKSGTSASLAISGLQSEDEADYYCAA**WDDSLNGP**

IGV1-47 (aka DPL3) (SEQ ID NO: 24)  
 QSVLTQPPSASGTPGQRTVITSCS**GSSSNIGSN**YVWYQQLPGTAPKLLIY**SNNQRPS**GVDPDRFSGSKSGTSASLAISGLRSEDEADYYCAA**WDDSLSGP**

IGLV3-1 (aka DPL23) (SEQ ID NO: 6)  
 QSVLTQPPSVSVSPGQTASITCS**GDKLGDKYAC**WYQQKPGQSPVLVIY**QDSKRPS**GIPERFSGSNSGNTATLTISGTQAMDEADYYCQA**WDSSTA**

IGLV3-19 (SEQ ID NO: 27)  
 SSELTDPAVSVVALGQTVRITCQ**GDSLRSYY**ASWYQQKPGQAPVLVIY**GKNNRPS**GIPDRFSGSSSGNTASLTITGAQAEDEADYYCNS**RDSSGNHL**

IGLV3-21 (SEQ ID NO: 9)  
 QSVLTQPPSVSVAPGQTARITCG**GNNIGSKS**VHWYQQKPGQAPVLWY**DDSDRPS**GIPERFSGSNSGNTATLTISRVEAGDEADYYCQV**WDSSSDHP**

IGLV6-57 (SEQ ID NO: 12)  
 NFMLTQPHSVSESPGKTVITSC**TGSSGSIASNY**VQWYQQRPQSAPTIVY**EDNQRPS**GVDPDRFSGSIDSSNSASLTISGLKTEDEADYYCQS**YDSSN**

IGHV3-23 (aka DP47) (SEQ ID NO: 3)  
 EVQLLESGGGLVQPGGSLRLSCAAS**GFTFSSY**AMSWVRQAPGKGLEWWSA**ISGSGG**STY**YADSVKGRFTISRDN**SKNTLYLQMNSLRAEDTAVYYCAK

Figura 7

Armazón IGLV3-1::IGHV3-23 con regiones variables de CDR3 (SEQ ID NO: 84)

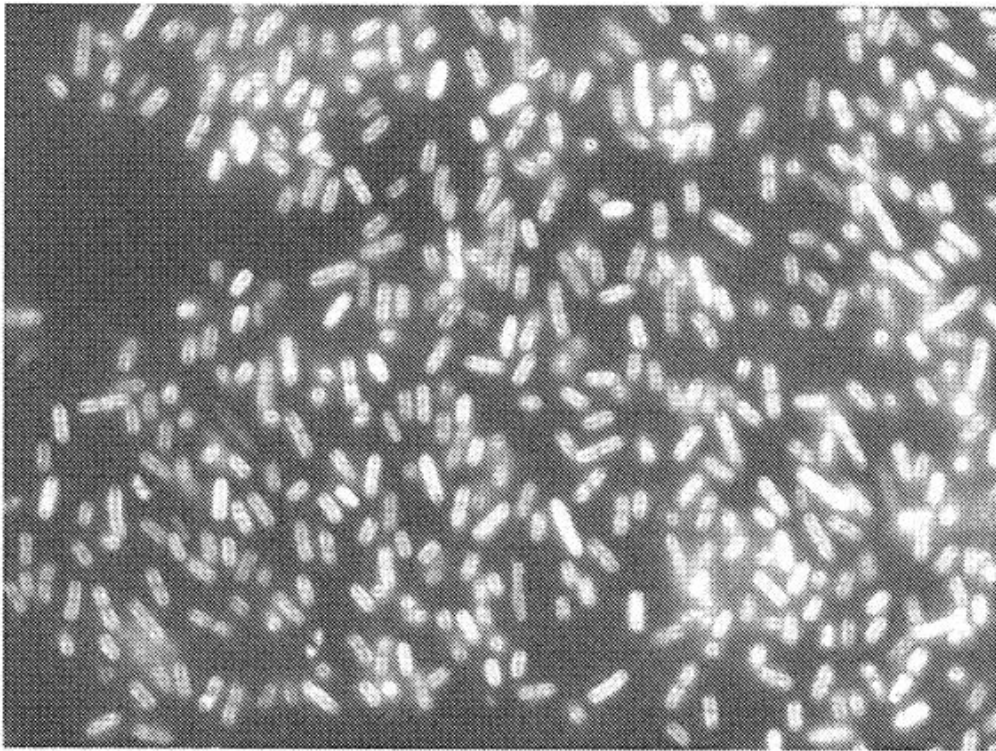
ATG GGA GAC GGT CAG TCT GTG CTG ACT CAG CCA CCC  
 TCAGTGTCCGTGTCCCCAGGACAGACAGCCAGCATCACCTGCTCTGGAGAT  
 AAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCCAGTCC  
 CCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCTGAG  
 CGATTCTCTGGCTCCAACCTCTGGGAACACAGCCACTCTGACCATCAGCGGG  
 ACCCAGGCTATGGATGAGGCTGACTATTACTGTCAGGCGTGGGAC  
NNNNNNNTGGAGGTNNNNNNNT  
 ACTGTGGTGTTCGGCACGGGCACCAAGCTCATCATTTTCGTCT  
**CAGACCGGTGGTTCTGGTGGTGGTGGTTCTGGCGGCGGCGGCTCCGGTGGT**  
**GGTGGATCC**  
 GAAGTCCAACCTGCTGGAGTCCGGCGGTGGCCTGGTGCAGCCAGGTGGCAGC  
 CTGCGCCTGAGCTGCGCCGCATCCGGTTTTACTTTCAGCAGCTACGCGATGT  
 CGTGGGTGCGCCAGGCACCGGGCAAGGGCCTGGAGTGGGTGAGCGCCATC  
 AGCGGTAGCGGCGGTTCTACGTATTATGCGGACAGCGTCAAGGGCCGTTTC  
 ACCATCAGCCGTGACAATTCCAAAAACACCCTGTACTTGCAGATGAACAGC  
 TTGCGTGCGGAAGATACGGCTGTTTACTACTGTGCGAAA  
NNNNNNNTGGANNTNNNNNT  
 GCCTTTGATATTTGGGGTCAAGGTACCATGGTTACCGTGAGC AGC

Traducción (SEQ ID NO: 85)

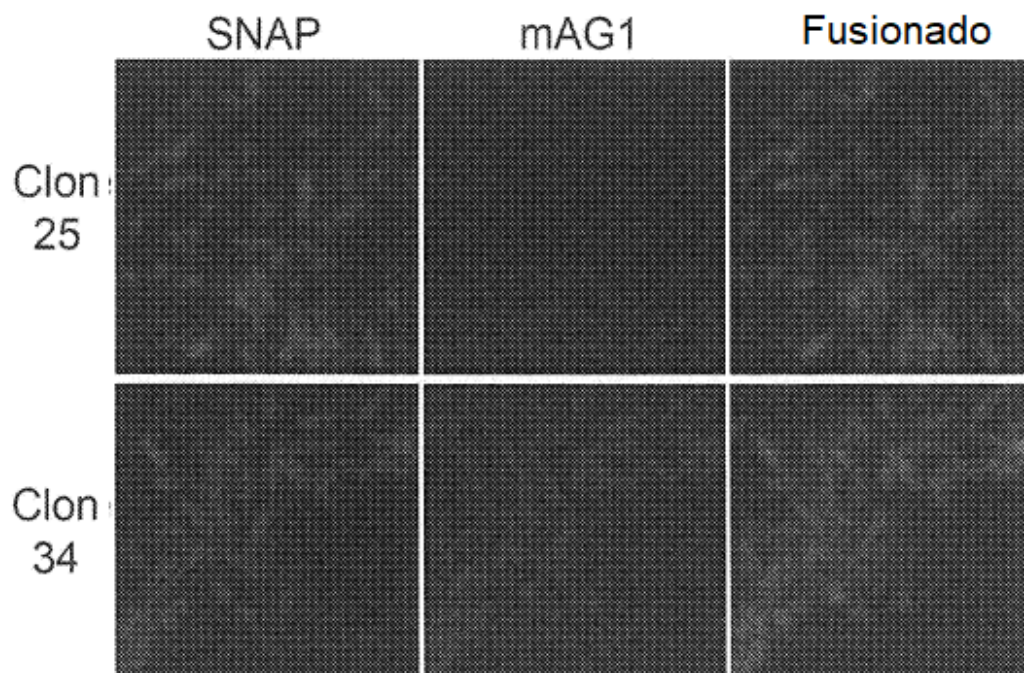
MGDGQSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSG  
 IPERFSGSNSGNTATLTISGTQAMDEADYYCQAWD NNNGGNNN TVVFGTGKVTVSS **QTGGSGGGG**  
**SGGGSGGGGS** EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKLEWVS  
 AISGSGGSTYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCAK NNNGNNN  
 AFDIWGQGTMTVTSS

**Figura 8**

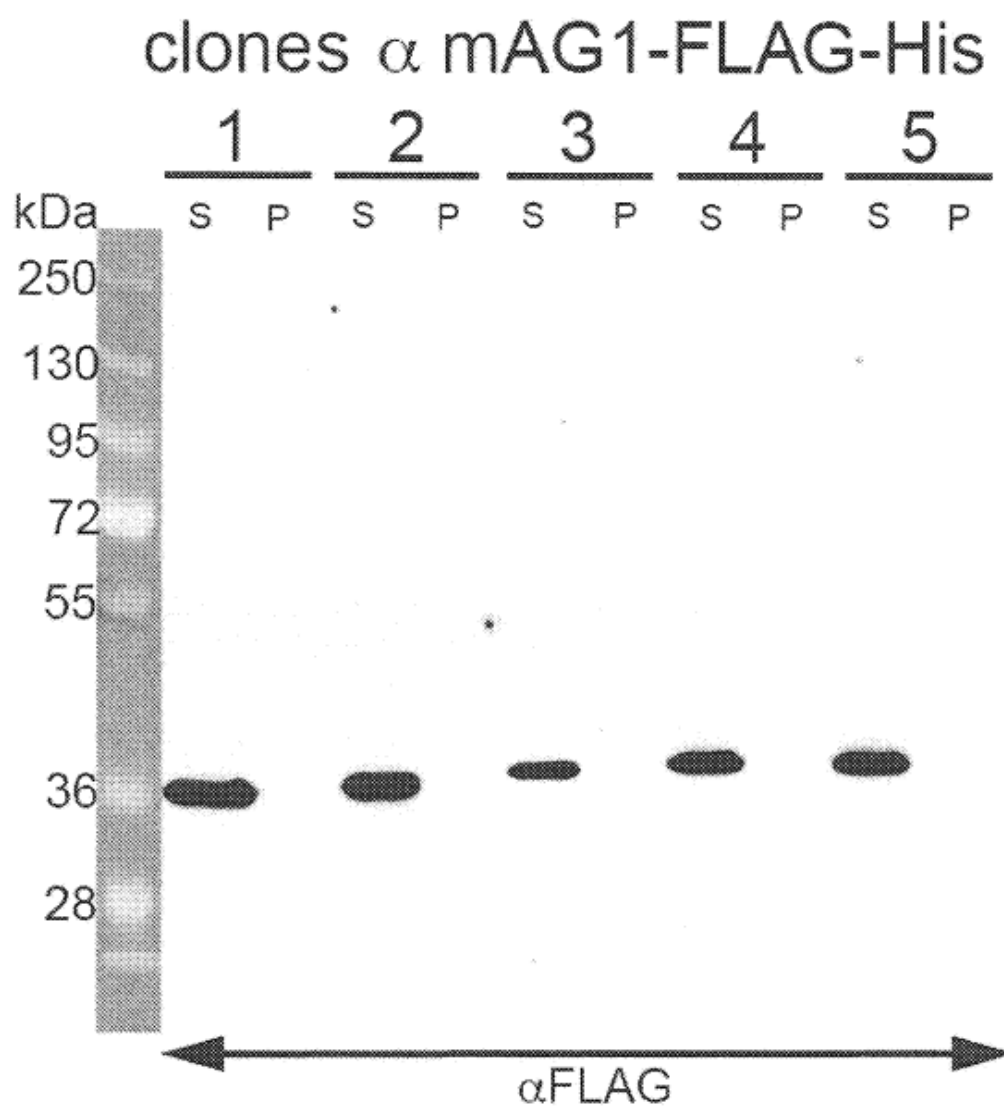




**Figura 9**

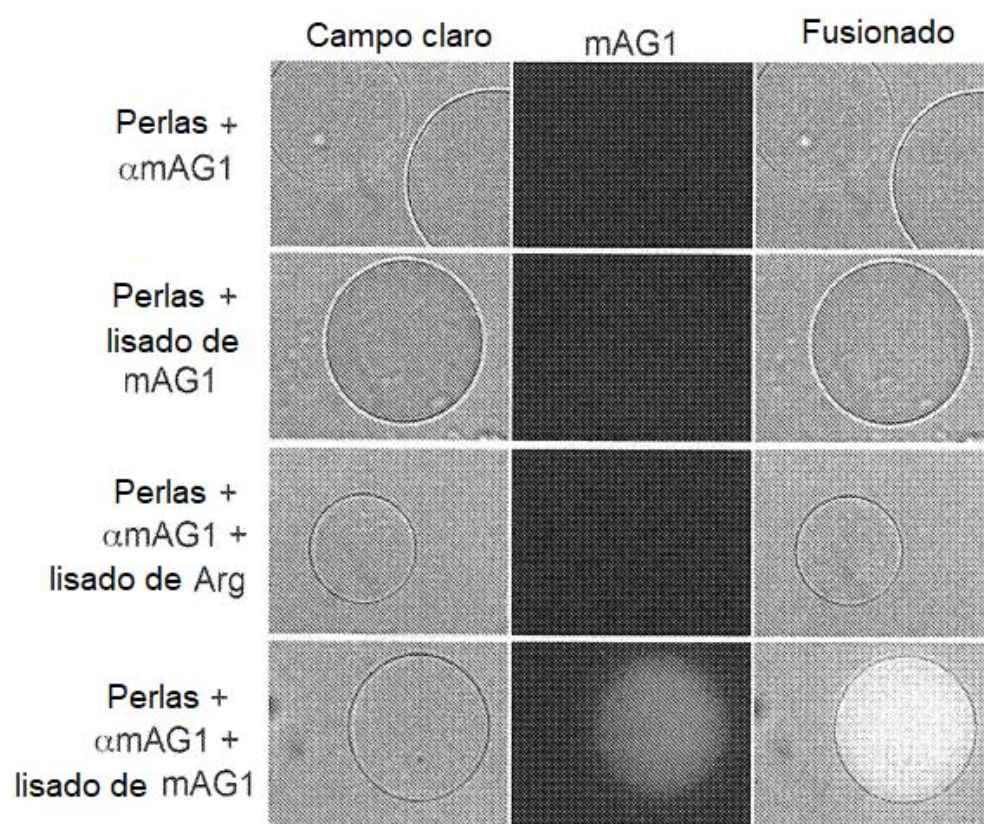


**Figura 10**



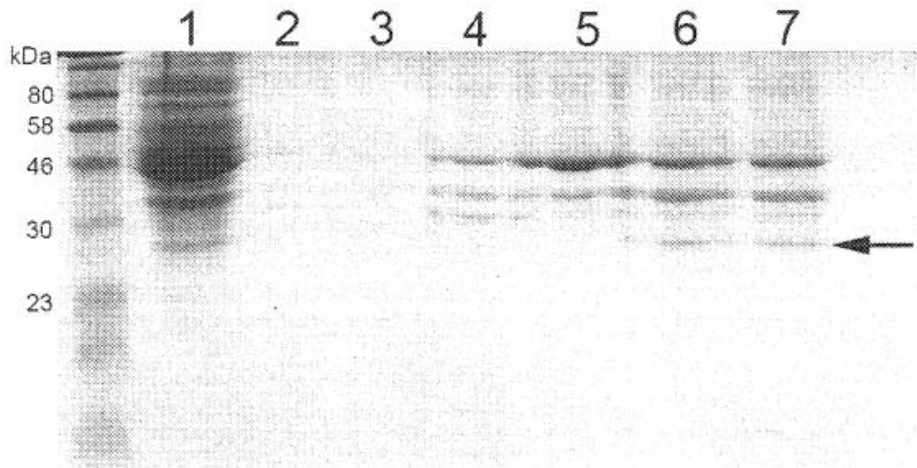
**Figura 11**



**Figura 12**



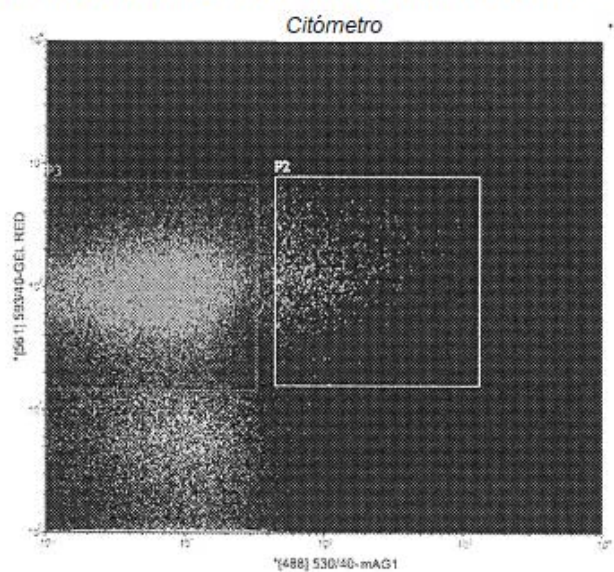
## Experimento de despegue de $\alpha$ mAG1



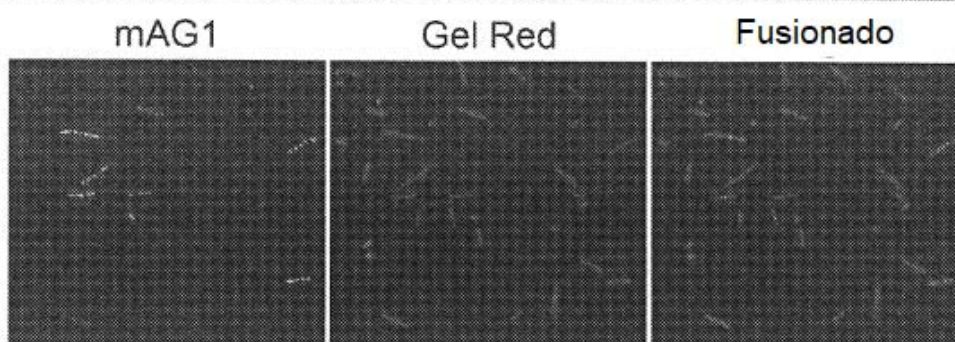
- 1 - Lisado de células completas con mAG1 limpia
- 2 - Perlas + mAG1 "sucia"
- 3 - Perlas + GFP
- 4 - Perlas +  $\alpha$ mAG1
- 5 - Perlas +  $\alpha$ mAG1 + Lisado completo de Arg
- 6 - Perlas +  $\alpha$ mAG1 (C1) + mAG1
- 7 - Perlas +  $\alpha$ mAG1 (C4) + mAG1

**Figura 13**

## Presentación lambdaoide de mAG1 - FACs



## Células clasificadas por FACs



**Figura 14**