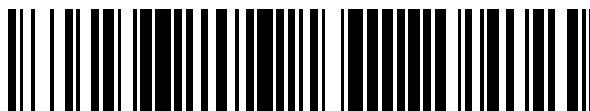


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 691 717**

51 Int. Cl.:

C12N 15/10 (2006.01)

C40B 30/04 (2006.01)

C40B 40/02 (2006.01)

C40B 50/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.10.2010** **PCT/EP2010/066250**

87 Fecha y número de publicación internacional: **05.05.2011** **WO11051333**

96 Fecha de presentación y número de la solicitud europea: **27.10.2010** **E 10771435 (4)**

97 Fecha y número de publicación de la concesión europea: **12.09.2018** **EP 2494046**

54 Título: **Bibliotecas universales del dominio de unión del lado inferior de tipo iii de la fibronectina de tipo III**

30 Prioridad:

30.10.2009 US 256409 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
28.11.2018

73 Titular/es:

NOVARTIS AG (100.0%)
Lichtstrasse 35
4056 Basel, CH

72 Inventor/es:

LOEW, ANDREAS y
VASH, BRIAN, EDWARD

74 Agente/Representante:

SALVÀ FERRER, Joan

ES 2 691 717 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

Bibliotecas universales del dominio de unión del lado inferior de tipo iii de la fibronectina de tipo III

5 REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

[0001] Esta solicitud reivindica el beneficio del número de serie de la solicitud de los EE. UU. 61/256 409 presentada el 30 de octubre de 2009.

10 ANTECEDENTES DE LA INVENCION

[0002] Las proteínas de unión basadas en andamiajes se están convirtiendo en alternativas legítimas a los anticuerpos en su capacidad para unirse a objetivos de ligandos específicos. Estas proteínas de unión de andamiaje comparten la calidad de tener un núcleo estructural estable que puede tolerar múltiples sustituciones en las regiones de unión del ligando. Algunos armazones de andamiajes tienen una arquitectura de dominio de proteína similar a la inmunoglobulina con bucles que se extienden desde un núcleo de sándwich beta. A continuación, se puede diseñar sintéticamente un núcleo de armazón de andamiaje a partir del cual se puede construir una biblioteca de diferentes variantes de secuencia. La diversidad de secuencias normalmente se concentra en las superficies exteriores de las proteínas tales como estructuras de bucle u otras superficies exteriores que pueden servir como regiones de unión de ligandos.

[0003] El dominio de la fibronectina de tipo III (Fn3) se identificó por primera vez como uno de los dominios repetitivos en la proteína de fibronectina. El dominio Fn3 constituye una proteína β -sándwich monomérica pequeña (~94 aminoácidos) formada por siete cadenas β con tres bucles de conexión. Los tres bucles cerca del extremo N de Fn3 son funcionalmente análogos a las regiones determinantes de complementariedad de los dominios de inmunoglobulina. Las bibliotecas de bucles Fn3 pueden modificarse para unirse a una variedad de dianas tales como citoquinas, factores de crecimiento y moléculas receptoras y otras proteínas.

[0004] Un posible problema al crear estas bibliotecas sintéticas es la alta frecuencia de variantes improductivas que dan lugar, por lo tanto, a selecciones de candidatos ineficientes. Por ejemplo, la creación de diversidad en las variantes a menudo implica técnicas *in vitro* tales como mutagénesis aleatoria, mutagénesis de saturación, PCR propensa a errores y mezcla de genes. Estas estrategias son intrínsecamente estocásticas y, a menudo, requieren la construcción de bibliotecas excesivamente grandes para explorar exhaustivamente la suficiente diversidad de secuencias. Además, no hay forma de enumerar el número, qué tipo y en qué parte de la proteína se han producido las mutaciones. Además, estas estrategias aleatorias crean sustituciones indiscriminadas que causan la desestabilización de la arquitectura proteica. Se ha demostrado que la mejora en una característica, como la optimización de afinidad, por lo general da lugar a una disminución de la estabilidad térmica en comparación con el armazón de andamiaje de la proteína original.

[0005] Por consiguiente, existe la necesidad de una biblioteca de dominio de unión a fibronectina que sea de construcción sistemática. Por diseño dirigido por bioinformática, los candidatos de bucle son flexibles para su inserción en múltiples andamiajes Fn3. Mediante sustituciones específicas de bucles dirigidos, se maximiza la estabilidad general del andamiaje mientras que al mismo tiempo se minimizan las sustituciones no inmunogénicas. Además, la biblioteca se puede adaptar a medida para que la diversidad general se pueda analizar fácilmente en diferentes sistemas. Además, la diversidad representativa de los bucles diseñados todavía es capaz de unirse a varias dianas de ligando predefinidos. Además, el diseño sistemático de bucles todavía permite la maduración por afinidad posterior de clones de unión recuperados.

[0006] El documento US 2002/019517 A1 describe un monocuerpo polipeptídico de fibronectina de tipo III (Fn3), una molécula de ácido nucleico que codifica el monocuerpo y una biblioteca de ácido nucleico variada que codifica el monocuerpo.

[0007] El documento WO2009/058379 A2 describe un armazón polipeptídico recombinante que comprende siete dominios de cadena beta designados A, B, C, D, E, F y G, unidos a seis regiones de bucle, en el que una región de bucle conecta cada cadena beta y se designa AB, BC, CD, DE, EF y FG. También se describe una biblioteca de presentación en polipéptidos que comprende el andamiaje.

[0008] Bloom y Calabro (2009) FN3: a new protein scaffold reaches the clinic, Drug Discovery Today 14:949-955 se refieren a "CT-322", una FN3 modificada que entró en ensayos de Fase II para glioblastoma multiforme.

[0009] Koide et al. (2002) Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor, Proc. Natl. Acad. Sci. USA 99:1253-1258 describe la selección de biblioteca combinatoria de monocuerpos para clones que se unen al receptor de estrógenos α .

5

[0010] El documento WO 2009/023184 A2 describe la mutagénesis directa y las bibliotecas de polipéptidos combinatorios de fibronectina de tipo III (FN3) de variante natural junto con su procedimiento de construcción y uso.

[0011] Koide et al. (1998) The fibronectin type III domain as a scaffold for novel binding proteins, J. Mol. Biol. 284:1141-1151 describe la preparación de una biblioteca de presentación en fagos de FN3 en la que se aleatorizaron los restos en dos bucles de superficie.

[0012] Karatan et al. (2004) Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain, Chem. Biol. 11:835-844 describe la construcción de una biblioteca presentada en fagos basada en el andamiaje de la décima fibronectina humana tipo III (FN3) aleatorizando restos en sus bucles FG y BC.

RESUMEN DE LA INVENCION

[0013] En un aspecto, esta descripción incluye una biblioteca combinatoria de variantes naturales de polipéptidos de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada. Los polipéptidos de la biblioteca incluyen (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre de un polipéptido o polipéptidos de fibronectina de tipo 3 nativos seleccionados, y (b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas (bucles inferiores). La Fn3 también puede tener regiones de bucle BC, DE y FG que tienen secuencias de aminoácidos de tipo silvestre, que tienen longitudes seleccionadas (bucles superiores) o secuencias de aminoácidos mutagenizadas.

[0014] De acuerdo con esto, en un aspecto, esta descripción se refiere a un procedimiento para formar una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada. El procedimiento comprende (i) alinear secuencias de bucle de aminoácidos AB, CD y EF en una colección de polipéptidos del dominio de fibronectina de tipo 3 nativo, (ii) segregar las secuencias de bucle alineadas según la longitud de bucle, (iii) para un bucle y longitud de bucle seleccionados de la etapa (ii), realizar análisis de frecuencia de aminoácidos posicionales para determinar las frecuencias de aminoácidos en cada posición del bucle, (iv) para cada bucle y longitud de bucle analizados en la etapa (iii), identificar en cada posición un aminoácido de consenso conservado o semiconservado seleccionado y otros aminoácidos de variantes naturales, (v) para al menos un bucle y longitud de bucle seleccionados, formar: (1) una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, el aminoácido de consenso, y si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido, o (2) una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, incluidos aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos, (vi) incorporar la biblioteca de secuencias codificantes dentro de las secuencias codificantes de Fn3 para formar una biblioteca de expresión de Fn3, y (vii) expresar los polipéptidos Fn3 de la biblioteca de expresión.

[0015] La biblioteca puede tener un umbral dado del 100 %, a menos que la posición de aminoácidos del bucle contenga solo un aminoácido dominante y uno variante, y los aminoácidos dominantes y variantes sean aminoácidos químicamente similares, en cuyo caso el umbral dado puede ser del 90 %. En este caso, la biblioteca contiene todas las variantes naturales o sus equivalentes químicos que tienen al menos cierta frecuencia razonable de aparición, por ejemplo, del 10 %, en el bucle y la posición del bucle seleccionados.

[0016] Las secuencias combinatorias de variantes naturales pueden estar en una combinación de bucles y longitudes de bucle seleccionadas de bucles AB y CD, AB y EF, y CD y EF, donde el bucle AB comprende 1-3 aminoácidos (AB/1-3), y preferentemente 3 aminoácidos (AB/3); el bucle CD comprende 4-10 aminoácidos (CD/4-10), y preferentemente 5, 6, o 7 aminoácidos (CD/5, CD/6, CD/7); y el bucle EF comprende 3-9 aminoácidos (EF/3-9), y preferentemente 6 aminoácidos (EF/6). Las secuencias combinatorias de variantes naturales pueden estar en

una combinación de bucles y longitudes de bucle seleccionadas del grupo que consiste en AB/1-3, AB/3, CD/4-10, CD/5, CD/6, CD/7 y EF/3-9, EF/6.

[0017] La biblioteca puede tener en dos de las combinaciones de bucle AB y CD, AB y EF, y CD y EF, mutaciones beneficiosas identificadas mediante la selección de una biblioteca combinatoria de variantes naturales que contiene variantes de aminoácidos en la combinación de dos bucles, y en el tercer bucle, identificados por AB, CD y EF, respectivamente.

[0018] En un caso, la biblioteca puede tener las secuencias de aminoácidos de tipo silvestre en las regiones A, AB, B, C, CD, D, E, EF, F y G del 10º módulo de fibronectina de tipo III de la fibronectina humana. En otro caso, la biblioteca puede tener las secuencias de aminoácidos de tipo silvestre en las regiones A, AB, B, C, CD, D, E, EF, F y G del 14º módulo de fibronectina de tipo III de la fibronectina humana.

[0019] En otro aspecto, esta descripción se refiere a una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad de unión o enzimática seleccionada, dichos polipéptidos que comprenden (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre de un polipéptido de fibronectina de tipo 3 nativo seleccionado, y (b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas, donde al menos una región de bucle seleccionada de una longitud seleccionada contiene una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido.

[0020] En otro aspecto, esta descripción se refiere a una biblioteca combinatoria de variantes naturales de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada. Los polipéptidos comprenden (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre de un polipéptido de fibronectina de tipo 3 nativo seleccionado, y (b) regiones de bucle AB, CD, y EF que tienen longitudes seleccionadas, donde al menos una región de bucle seleccionada de una longitud seleccionada contiene una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variante natural, incluidos los aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos.

[0021] La biblioteca tiene una biblioteca de secuencias combinatorias de variantes naturales en una combinación de bucles y longitudes de bucle seleccionadas de bucles AB y CD, AB y EF, y CD y EF, donde el bucle AB se selecciona de uno de AB/1-3, AB/3, y el bucle CD se selecciona de uno de CD/4-10, CD/5, CD/6, CD/7, o el bucle AB se selecciona de uno de AB/1-3, AB/3, y el bucle EF se selecciona de uno de EF/3-9, EF/6, o el bucle CD se selecciona de uno de CD/4-10, CD/5, CD/6, CD/7 y el bucle EF se selecciona de uno de EF/3-9, EF/6, o una combinación de los tres bucles seleccionados del grupo que consiste en AB/1-3, AB/3, CD/4-10, CD/5, CD/6, CD/7, EF/3-9 y EF/3. La biblioteca tiene un umbral dado del 100 %, a menos que la posición del aminoácido de bucle contenga solo un aminoácido dominante y uno variante, y los aminoácidos dominantes y variantes tengan cadenas laterales con propiedades fisicoquímicas similares, en cuyo caso el umbral dado es del 90 %. Los polipéptidos tienen las secuencias de aminoácidos de tipo silvestre en las regiones de cadena beta A, AB, B, C, CD, D, E, EF, F y G del 10º módulo de fibronectina de tipo III de la fibronectina humana, o 14º módulo de fibronectina de tipo III de la fibronectina humana.

[0022] En un caso, la biblioteca tiene una longitud de bucle AB de 3. En otro caso, la biblioteca tiene una longitud de bucle CD de 5, 6 o 7. En otro caso, la biblioteca tiene una longitud de bucle EF de 6. Los polipéptidos de la biblioteca están codificados por una biblioteca de expresión seleccionada del grupo que consiste en una biblioteca de presentación en ribosomas, una biblioteca de presentación en polisomas, una biblioteca de presentación en fagos, una biblioteca de expresión bacteriana y una biblioteca de presentación en levaduras.

[0023] En otro aspecto, esta descripción se refiere a una biblioteca de expresión de polinucleótidos que codifica la biblioteca de polipéptidos y se produce sintetizando polinucleótidos que codifican una o más regiones marco de cadena beta y una o más regiones de bucle en las que los polinucleótidos están predeterminados, donde

los polinucleótidos que codifican dichas regiones además comprenden suficiente secuencia de solapamiento por lo que las secuencias de polinucleótidos, en condiciones de reacción en cadena de la polimerasa (PCR), son capaces de ensamblarse en polinucleótidos que codifican dominios de unión a fibronectina completos.

- 5 **[0024]** En otro aspecto más, esta descripción se refiere a un procedimiento para identificar un polipéptido que tiene una afinidad de unión deseada con respecto a un antígeno seleccionado. El procedimiento comprende hacer reaccionar la biblioteca combinatoria de variantes naturales de polipéptidos de Fn3 con el antígeno seleccionado, y cribar los polipéptidos de Fn3 para seleccionar los que tienen una afinidad de unión deseada con respecto al antígeno seleccionado.

10

[0025] El procedimiento comprende además la etapa de identificar el polinucleótido que codifica el dominio de unión a fibronectina seleccionado.

- Basándose en la descripción que está contenida en el presente documento, la presente invención proporciona un procedimiento para formar una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles para
15 explorar la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada, en el que dichos polipéptidos tienen las secuencias de aminoácidos de tipo silvestre en las regiones de cadena beta A, AB, B, C, CD, D, E, EF, F y G del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, el procedimiento que comprende las etapas de:

- 20 (i) alinear secuencias de bucle de aminoácidos AB, CD y EF en una colección de polipéptidos del dominio de fibronectina de tipo 3 nativo,

(ii) segregar las secuencias de bucle alineadas según la longitud del bucle,

- (iii) para un bucle y una longitud bucle seleccionados de la etapa (ii), realizar un análisis de frecuencia posicional de aminoácidos para determinar las frecuencias de aminoácidos en cada posición del bucle, en el que la longitud del
25 bucle EF está fijada en 6 aminoácidos (EF/6),

(iv) para cada bucle y longitud de bucle analizados en la etapa (iii), identificar en cada posición un aminoácido de consenso conservado o semiconservado seleccionado y otros aminoácidos de variante natural,

(v) para al menos el bucle EF, formar:

- 30 (1) una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, el aminoácido de consenso, y si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido, o

- 35 (2) una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, incluidos aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos,

40

(vi) incorporar la biblioteca de secuencias codificantes en las secuencias codificantes de Fn3 marco para formar una biblioteca de expresión de Fn3, y

(vii) expresar los polipéptidos de Fn3 de la biblioteca de expresión.

- 45 **[0026]** La presente invención también proporciona una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad de unión o enzimática seleccionada, dichos polipéptidos que comprenden:

(a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, y

- 50 (b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas, en las que la longitud del bucle EF está fijada en 6 aminoácidos (EF/6), donde al menos la región de bucle EF contiene una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una
55 frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido.

[0027] Además, la presente invención proporciona una biblioteca combinatoria de variantes naturales de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más

polipéptidos que tienen una actividad de unión o enzimática seleccionada, dichos polipéptidos que comprenden:

- (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, y
- 5 (b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas, en las que la longitud del bucle EF está fijada en 6 aminoácidos (EF/6), donde al menos la región de bucle EF contiene una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un
- 10 50 %, otros aminoácidos de variantes naturales, que incluyen aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos.

[0028] Otros aspectos y realizaciones de la presente invención se exponen en las reivindicaciones adjuntas.

15

Breve descripción de las figuras

[0029]

- 20 La Fig. 1 es un diagrama esquemático que ilustra el procedimiento para construir una biblioteca de dominio de unión del lado inferior de la fibronectina usando biominería de bases de datos genéticas asistidas por ordenador y delineación de estructuras de bucles y andamiajes beta.

- La Fig. 2 representa un esquema que muestra las caras de bucle superior e inferior de Fn3 que ilustra el dominio de las dos láminas beta antiparalelas. Una mitad está compuesta de cadenas beta (ABE) y la otra mitad está
- 25 compuesta por (CDFG). Los 6 bucles similares a CDR también están indicados: AB, BC, CD, DE, EF y FG. Los bucles BC, DE y FG están presentes en el extremo N del dominio Fn3 y están dispuestos para formar superficies de unión del ligando. La secuencia RGD se encuentra en el bucle FG.

- La Fig. 3A es un gráfico de barras que muestra la diversidad de la longitud del bucle AB derivada del análisis bioinformático de todos los módulos de Fn3. La longitud del bucle AB 3 es el tamaño más predominante observado
- 30 en las secuencias de Fn3 expresadas.

La Fig. 3B es una tabla que muestra la distribución de la longitud del bucle en la longitud del bucle AB, mostrando que 3 es el tamaño más predominante observado en las secuencias expresadas de Fn3.

- La Fig. 4A es una tabla que muestra la distribución posicional de aminoácidos en las posiciones 1, 2 y 3 de un tamaño de longitud de bucle AB de 3 y los 7 porcentajes más altos de abundancia de aminoácidos de los
- 35 aminoácidos en cada posición.

La Fig. 4B es un gráfico de barras que muestra la diversidad de secuencia de una región de bucle ejemplar en forma de perfil de variabilidad de aminoácidos (distribución de frecuencia) para el tamaño de longitud de bucle AB de 3.

- La Fig. 5A es un gráfico de barras que muestra la diversidad de la longitud del bucle CD derivada del análisis bioinformático de todos los módulos de Fn3. Las longitudes del bucle CD de 5, 6 y 7 son los tamaños más
- 40 predominantes observados en las secuencias de Fn3 expresadas.

La Fig. 5B es una tabla que muestra la distribución de la longitud del bucle en la longitud del bucle CD. Las longitudes del bucle CD de 5, 6 y 7 son los tamaños más predominantes observados en las secuencias de Fn3 expresadas.

- La Fig. 6A es una tabla que muestra la distribución posicional de aminoácidos en las posiciones 1, 2, 3, 4 y 5 del tamaño de longitud del bucle CD de 5 y los 7 porcentajes más altos de abundancia de aminoácidos de los
- 45 aminoácidos en cada posición.

La Fig. 6B es un gráfico de barras que muestra la diversidad de secuencia de una región de bucle ejemplar en forma de perfil de variabilidad de aminoácidos (distribución de frecuencia) para el tamaño 5 de longitud de bucle CD.

- La Fig. 7A es una tabla que muestra la distribución posicional de aminoácidos en las posiciones 1, 2, 3, 4, 5 y 6 del tamaño de longitud del bucle CD de 6 y los 7 porcentajes más altos de abundancia de aminoácidos de los
- 50 aminoácidos en cada posición.

La Fig. 7B es un gráfico de barras que muestra la diversidad de secuencia de una región de bucle ejemplar en forma de perfil de variabilidad de aminoácidos (distribución de frecuencia) para el tamaño 6 de longitud de bucle CD.

- La Fig. 8A es una tabla que muestra la distribución posicional de aminoácidos en las posiciones 1, 2, 3, 4, 5, 6 y 7 del tamaño de longitud del bucle CD de 7 y los 7 porcentajes más altos de abundancia de aminoácidos de los
- 55 aminoácidos en cada posición.

La Fig. 8B es un gráfico de barras que muestra la diversidad de secuencias de una región de bucle ejemplar en forma de perfil de variabilidad de aminoácidos (distribución de frecuencia) para el tamaño 7 de longitud de bucle CD.

La Fig. 9A es una tabla que muestra la distribución de la longitud del bucle en la longitud del bucle EF, que muestra

que 6 es el tamaño más predominante observado en las secuencias de Fn3 expresadas.

La Fig. 9B es un gráfico de barras que muestra la diversidad de la longitud del bucle EF derivada del análisis bioinformático de todos los módulos de Fn3. La longitud del bucle EF de 6 es el tamaño más predominante observado en las secuencias de Fn3 expresadas.

- 5 La Fig. 10 A es una tabla que muestra la distribución posicional de aminoácidos en las posiciones 1, 2, 3, 4, 5 y 6 de un tamaño de longitud de bucle EF de 6 y los 7 porcentajes más altos de abundancia de aminoácidos de los aminoácidos en cada posición.

La Fig. 10B es un gráfico de barras que muestra la diversidad de secuencias de una región de bucle ejemplar en forma de perfil de variabilidad de aminoácidos (distribución de frecuencia) para el tamaño de longitud de bucle EF de 6.

Descripción detallada de la invención

Definiciones

15

[0030] Los términos a continuación tienen los siguientes significados a menos que se indique lo contrario en la especificación:

- [0031]** Como se usa en el presente documento, el término "dominio de fibronectina de tipo III" o "dominio Fn3" se refiere a un dominio Fn3 silvestre de cualquier organismo, así como a dominios Fn3 quiméricos contruidos a partir de cadenas beta de dos o más dominios Fn3 diferentes. Como es sabido en la materia, los dominios Fn3 naturales tienen una estructura beta-sándwich compuesta por siete cadenas beta, denominadas A, B, C, D, E, F y G, unidas por seis bucles, a las que se hace referencia como bucles AB, BC, CD, DE, EF y FG (Véase, por ejemplo, Bork y Doolittle, Proc. Natl. Acad. Sci. USA 89: 8990, 1992; Bork et al., Nature Biotech., 15: 553, 1997; Meinke et al., J. Bacteriol. 175: 1910, 1993; Watanabe et al., J. Biol. Chem. 265: 15659, 1990; Main et al., 1992; Leahy et al., 1992; Dickinson et al., 1994; Patente de Estados Unidos 6 673 901; Publicación del Tratado de cooperación en materia de patente WO/03104418; y, Solicitud de patente de Estados Unidos 2007/0082365).

- [0032]** Tres bucles están en la parte superior del dominio (los bucles BC, DE y FG) y tres bucles están en la parte inferior del dominio (los bucles AB, CD y EF) (Fig. 2). En un caso particular de esta descripción, por ejemplo en las realizaciones de la presente invención, el dominio Fn3 es del décimo dominio Fn3 de la fibronectina humana (¹⁰Fn3) (SEQ ID NO: 1). Los polipéptidos del dominio Fn3 individuales se denominan por el número de módulo y el nombre de la proteína, por ejemplo, el 10^o o 14^o módulo de la fibronectina humana (¹⁰Fn o ¹⁴Fn).

- [0033]** Como se usa en este documento, la expresión "molécula de unión basada en Fn3" o "molécula de unión basada en fibronectina de tipo III" se refiere a un dominio de Fn3 que se ha alterado para que contenga una o más secuencias de unión que no sean de Fn3. En un caso particular, uno o más de los bucles AB, CD y/o EF inferiores están alterados en comparación con el dominio Fn3 natural correspondiente para que contenga una o más secuencias que no se unen a Fn3. En otro caso, uno o más de los bucles AB, CD o EF inferiores y uno o más de los bucles superiores BC, DE y FG están alterados en comparación con el dominio Fn3 silvestre correspondiente para contener una o más secuencias de unión no Fn3. Dichas moléculas se denominan en el presente documento moléculas de unión basadas en Fn3 biespecíficas. En un ejemplo adicional, por ejemplo en una realización adicional de la invención, dos o más moléculas de unión basadas en Fn3 o moléculas de unión basadas en Fn3 biespecíficas están unidas entre sí. Dichas moléculas se denominan en la presente memoria "moléculas de unión basadas en Fn3 multiespecíficas" (véase, por ejemplo, el documento US n.º de serie 61/050142).

- [0034]** El término "secuencia de unión no Fn3" se refiere a una secuencia de aminoácidos que no está presente en el dominio Fn3 de origen natural (por ejemplo, de tipo silvestre) y que se une a una diana específica. Dichas secuencias de unión no Fn3 normalmente se introducen modificando (por ejemplo, mediante sustitución y/o adición) el dominio Fn3 de tipo silvestre (por ejemplo, dentro de los bucles inferiores y/o las regiones de bucle superior). Esto se puede lograr, por ejemplo, mediante mutación aleatoria o predeterminada de restos de aminoácidos dentro del dominio Fn3 de tipo silvestre.

- [0035]** El término "monoespecífico" como se usa en el presente documento se refiere a una molécula de unión basada en Fn3 que se une a una o más moléculas diana que comprenden dominios Fn3 en los que solo se usan para unir la región inferior del dominio Fn3 o la región superior del dominio Fn3, pero no ambas. Por ejemplo, una molécula de unión basada en Fn3 monoespecífica inferior es aquella que utiliza solo los bucles inferiores, como los bucles AB, CD o EF, o el C-terminal del dominio Fn3 para unirse a una diana, mientras que una molécula de unión basada en Fn3 monoespecífica superior solo utiliza los bucles superiores del dominio Fn3, como los bucles

BC, DE y FG, para unir la diana. Debe entenderse que no es necesario usar los tres bucles de la región superior o inferior para unir la molécula diana.

[0036] Los dominios monoespecíficos de Fn3 también se pueden unir entre sí (por ejemplo, en forma de perlas) para formar una molécula de unión basada en Fn3 multiespecífica que comprende, por ejemplo, al menos dos dominios Fn3 monoespecíficos que están unidos entre sí. Para moléculas de unión monoespecíficas inferiores, cada uno de los dominios Fn3 usa al menos un bucle inferior o región C-terminal para unirse a una o más moléculas diana. En un caso, por ejemplo, en una realización, esta molécula de unión multiespecífica basada en Fn3 se une a diferentes regiones objetivo de la misma molécula diana (por ejemplo, la Diana A). Por ejemplo, un dominio Fn3 puede unirse a una primera región objetivo de la Diana A y otro dominio Fn3 puede unirse a una segunda región objetivo de la Diana A. Esto puede usarse para aumentar la avidéz de la molécula de unión basada en Fn3 por la molécula diana. En otro caso, por ejemplo en otra realización, la molécula de unión multiespecífica basada en Fn3 se une a múltiples moléculas diana. Por ejemplo, un dominio Fn3 puede unirse a la Diana A y otro dominio Fn3 puede unirse a la Diana B (por ejemplo, extensor de la semivida). En otro caso más, por ejemplo en otra realización, la molécula de unión basada en Fn3 multiespecífica comprende al menos dos dominios Fn3 monoespecíficos que se unen a diferentes regiones objetivo de la Diana A y al menos dos dominios Fn3 monoespecíficos que se unen a diferentes regiones objetivo de la Diana B. El experto en la materia apreciará que cualquier número de dominios Fn3 se puede unir de esta manera para crear una molécula de unión multiespecífica basada en Fn3 que sea capaz de unirse a diferentes regiones objetivo de la misma molécula diana o de diferentes moléculas diana.

[0037] El término "biespecífico" como se usa en el presente documento se refiere a una molécula de unión basada en Fn3 que se une a una o más dianas usando tanto la región inferior del dominio Fn3 como la región superior del dominio Fn3. Por ejemplo, una molécula de unión biespecífica basada en Fn3 comprende dominios Fn3 que usan los bucles inferiores, como los bucles AB, CD o EF, o C-terminal de la molécula y los bucles superiores de la molécula, como los bucles BC, DE, y FG, para unir la diana. Las moléculas de unión biespecíficas basadas en Fn3 pueden usarse para unir la misma molécula diana, por ejemplo, la Diana A, que puede unirse a la parte superior e inferior de la molécula de unión biespecífica basada en Fn3 (véase Figura 3b). Como alternativa, la molécula de unión biespecífica basada en Fn3 se puede usar para unirse a dos moléculas diana diferentes, por ejemplo, la Diana A y la Diana B. En este caso, los bucles superiores se pueden usar para unirse a la Diana A y los bucles inferiores se pueden usar para unirse a la Diana B, o viceversa (véase Figura 3b). Las moléculas de unión biespecíficas basadas en Fn3 también se pueden unir entre sí (por ejemplo, de forma similar a perlas) para formar una molécula de unión multiespecífica basada en Fn3.

[0038] El término "multiespecífico" como se usa en el presente documento se refiere a una molécula de unión basada en Fn3 que comprende al menos dos moléculas de unión basadas en Fn3 monoespecíficas unidas entre sí o al menos dos moléculas de unión basadas en Fn3 biespecíficas unidas entre sí.

[0039] El término "región determinante de complementariedad (CDR)" se refiere a un bucle hipervariable de un dominio variable de anticuerpo o de un receptor de célula T. La posición de las CDR dentro de una región variable del anticuerpo se ha definido con precisión (véase, Kabat, EA, et al. Sequences of Proteins of Immunological Interest, quinta edición, Departamento de Salud y Servicios Humanos de los EE. UU., Publicación NIH n.º 91-3242, 1991, y Chothia, C. et al., J. Mol. Biol. 196: 901-917, 1987).

[0040] El término "anticuerpos de dominio único" se refiere a cualquier dominio variable único de origen natural y fragmentos de unión modificados correspondientes, que incluyen anticuerpos de dominio humano como los descritos por Domantis (Domantis/GSK (Cambridge, RU) o nanoanticuerpos de camélidos como se define a continuación.

[0041] El término "anticuerpo de cadena sencilla" se refiere a una porción de unión a antígeno de una región variable de cadena ligera y una porción de unión a antígeno de una región variable de cadena pesada, unidas, usando procedimientos recombinantes, por un enlazador sintético que permite que se preparen como una única cadena de proteína en la que las regiones VL y VH se aparean para formar moléculas monovalentes (conocidas como Fv de cadena sencilla (scFv); véase, por ejemplo, Bird et al. (1988) Science 242: 423-426 y Huston et al., (1988) Proc. Natl. Acad. Sci. USA 85: 5879-5883).

[0042] El término "SCALP" se refiere a una proteína de tipo cadena de anticuerpo de cadena sencilla. La agrupación SCALP incluye una cadena pesada aislada de un anticuerpo, o fragmento del mismo; o una región variable fusionada a otra. Por lo tanto, el grupo de SCALPS incluye V_{HH} (NANOBODY); y V_{HH} más una o más regiones constantes de la cadena pesada (por ejemplo, CH1, CH2 y/o CH3). Por lo tanto, una SCALP se ejemplifica

con un V_{HH}; a V_{HH} + CH1; un V_{HH} + CH2; un V_{HH} + CH3; un V_{HH} + CH1 + CH2; un V_{HH} + CH1 + CH2 + CH3; un V_{HH} + CH2 + CH3; un V_{HH} + CH1 + CH3; etc. Las SCALP incluyen anticuerpos de dominio único.

[0043] El término "NANOBODY®" se refiere a una región de un anticuerpo camélido que es el pequeño dominio variable único desprovisto de cadena ligera y que puede obtenerse mediante ingeniería genética para producir una proteína pequeña que tiene una alta afinidad por una diana, dando como resultado un bajo peso de proteína derivada de anticuerpos. Véase documento WO07042289 y Patente de EE.UU. número 5 759 808 expedida el 2 de junio de 1998; véase también Stijlemans, B. et al., 2004. Las bibliotecas manipuladas por ingeniería de anticuerpos camélidos y fragmentos de anticuerpos están disponibles en el mercado, por ejemplo, en Ablynx, Ghent, Bélgica. Como con otros anticuerpos de origen no humano, una secuencia de aminoácidos de un anticuerpo camélido se puede alterar de forma recombinante para obtener una secuencia que se asemeja más a una secuencia humana, es decir, el NANOBODY® puede estar "humanizado". Por lo tanto, la baja antigenicidad natural de los anticuerpos camélidos hacia los seres humanos se puede reducir aún más.

[0044] El término "SmALP" se refiere a cualquier proteína pequeña similar a anticuerpo. SmALP incluye fragmentos Fab, que comprenden las porciones VH, CH, VL y VH; Fv, que comprende las porciones VH y VL; ScFv, que comprende una porción VH y VL fusionadas entre sí; dAbs; di-ScFv, que comprende dos fragmentos ScFv; y Fcabs. Estas y otras SmALP se describen en Classical and Heavy Chain Antibodies, de Muyldermans, S. (2001), J Biotechnol. 74, 277-302, y son bien conocidos en la materia. Los dAbs (anticuerpos de dominio) corresponden a las regiones variables de las cadenas pesada (VH) o ligera (VL) de los anticuerpos humanos y tienen un peso molecular de aproximadamente 13 kDa, o menos de una décima parte del tamaño de un anticuerpo completo. Los dAbs están disponibles en Domantis Limited, una subsidiaria de GlaxoSmithKline. Un Fcab (Fc de unión a antígeno) es un anticuerpo comprimido que comprende un dominio CH2 y CH3, con dos sitios de unión a antígeno idénticos diseñados en los dominios CH3, como se puede obtener de f-star Biotechnologische Forschungs- und Entwicklungsges, M.b.H. Un cuerpo ASCP puede estar fusionado a una región CH1 y/o CH2 y/o CH3, y/o VL. El cuerpo ASCP puede estar dentro del contexto de SCALP o SmALP.

[0045] El término "diana" se refiere a un antígeno o epítipo reconocido por la molécula de unión basada en Fn3 de la invención. Las dianas incluyen, pero no se limitan a, epítopos presentes en proteínas, péptidos, carbohidratos y/o lípidos.

[0046] El término "conjugado" se refiere a una molécula de unión basada en Fn3 ligada química o genéticamente a uno o más restos no Fn3.

[0047] El término "resto no Fn3" se refiere a una entidad biológica o química que confiere funcionalidad adicional a una molécula a la que está unido. En una realización particular, el resto no Fn3 es un polipéptido, por ejemplo, albúmina de suero humano (HSA), o una entidad química, por ejemplo, polietilenglicol (PEG) que aumenta la semivida *in vivo* de la molécula de unión basada en Fn3.

[0048] El término "biblioteca" de polipéptidos Fn3 se refiere a una colección de polipéptidos Fn3 que tienen una variación o diversidad de secuencia seleccionada en al menos uno de los bucles AB, CD y EF de una longitud definida. El término "biblioteca" también se usa para referirse a la colección de secuencias de aminoácidos dentro de un bucle seleccionado AB, CD y EF de una longitud seleccionada, y a la colección de secuencias codificantes que codifican bibliotecas de aminoácidos de bucle o polipéptido.

[0049] El término "biblioteca universal de unión del lado inferior de Fn3" se refiere a una biblioteca de polipéptidos Fn3 en la que la diversidad de aminoácidos en una o más de las regiones de bucle AB, CD y EF está determinada por o refleja las variantes de aminoácidos presentes en una colección de secuencias conocidas de Fn3.

[0050] El término "biblioteca universal de unión a N+/-" o "bibliotecas N+/-" se refiere a una biblioteca más sofisticada o ajustada en la que los aminoácidos más frecuentes que rodean a un aminoácido fijo se determinan en el diseño de la biblioteca. Estas bibliotecas N+/- están construidas con variaciones en los bucles inferiores, AB, CD y EF, los bucles superiores, BC, DE, FG o cualquier combinación de bucles superiores e inferiores. Para "bibliotecas N+/-", N es el aminoácido predominante en una posición particular y los aminoácidos aguas arriba o aguas abajo se designan +N o -N, respectivamente. Por ejemplo, N+3 es un aminoácido 3 posiciones aguas arriba de N, mientras que N-3 es un aminoácido 3 posiciones aguas abajo de N en una estructura tridimensional de Fn3. Del mismo modo, N+2 y N+1 son aminoácidos en las posiciones 2 y 1 aguas arriba de N, respectivamente, mientras que N-2 y N-1 son aminoácidos en las posiciones 2 y 1 aguas abajo de N, respectivamente. Al alterar N del aminoácido predominantemente más abundante a un aminoácido menos abundante, se puede evaluar el efecto de esa

modificación en la abundancia de aminoácidos a 1, 2 o 3 posiciones de distancia de N. Al diseñar dicha biblioteca, se determinan la frecuencia y la abundancia de aminoácidos que rodean la posición N fija. Estas diferencias pueden usarse para generar bibliotecas universales de dominio de unión en el lado inferior de la fibronectina, bibliotecas de dominio de unión en el lado superior o una combinación de bibliotecas de dominio de unión tanto en el lado inferior como en el lado superior.

[0051] El término "resto de aminoácido conservado" o "aminoácido fijo" se refiere a un resto de aminoácido que se determina que se produce con una frecuencia que es alta, normalmente al menos un 50 % o más (por ejemplo, aproximadamente un 60 %, 70 %, 80 %, 90 %, 95 % o un 100 %), para una posición de resto dada. Cuando se determina que un resto dado se produce a una frecuencia tan alta, es decir, por encima de un umbral de aproximadamente un 50 %, se puede determinar que se conserva y se representa en las bibliotecas de esta descripción, por ejemplo, en las bibliotecas de la presente invención, como un resto "fijo" o "constante", al menos para esa posición de resto de aminoácido en la región de bucle que se analiza.

[0052] El término "resto de aminoácido semiconservado" se refiere a restos de aminoácidos que se determinan que se producen con una frecuencia que es alta, de 2 a 3 restos para una posición de resto dada. Cuando 2-3 restos, preferentemente 2 restos, que juntos, están representados a una frecuencia de aproximadamente un 40 % del tiempo o superior (por ejemplo, un 50 %, 60 %, 70 %, 80 %, 90 % o más), se determina que los restos están semiconservados y por lo tanto representados en las bibliotecas de esta descripción, por ejemplo en las bibliotecas de la presente invención, como "semifijos" al menos para esa posición de resto de aminoácido en la región de bucle que se analiza. Normalmente, se introduce un nivel apropiado de mutagénesis/variabilidad de ácido nucleico para una posición de aminoácido semiconservada (codón) de modo que los 2 a 3 restos estén representados apropiadamente. Así, cada uno de los 2 a 3 restos puede decirse que es "semifijo" para esta posición. Un "resto de aminoácido semiconservado seleccionado" es uno de los 2 o más restos de aminoácido semiconservados seleccionados, normalmente, aunque no necesariamente, el resto que tiene la mayor frecuencia de aparición en esa posición.

[0053] El término "resto de aminoácido variable" se refiere a restos de aminoácidos que se determinan que se producen con una frecuencia inferior (inferior al 20 %) para una posición de resto dada. Cuando muchos restos aparecen en una posición determinada, se determina que la posición del resto es variable y está representada en las bibliotecas de esta descripción, por ejemplo en las bibliotecas de la presente invención, como variable al menos para esa posición de restos de aminoácidos en la región del bucle analizada. Normalmente, se introduce un nivel apropiado de mutagénesis/variabilidad de ácido nucleico para una posición variable de aminoácido (codón) de manera que se represente adecuadamente un espectro preciso de restos. Por supuesto, se entiende que, si se desea, las consecuencias o la variabilidad de cualquier posición de resto de aminoácido, es decir, conservada, semiconservada o variable, se pueden representar, explorar o alterar utilizando, según sea apropiado, cualquiera de los procedimientos de mutagénesis descritos en este documento. Una frecuencia umbral más baja de aparición de aminoácidos variables puede ser, por ejemplo, del 5-10 % o inferior. Por debajo de este umbral, los aminoácidos variables pueden omitirse de los aminoácidos de variante natural en esa posición.

[0054] El término "consenso" se refiere a un aminoácido en un bucle AB, CD y EF de un polipéptido Fn3 que es un aminoácido conservado o uno seleccionado de aminoácidos semiconservados que aparece predominantemente en una posición dentro del bucle.

[0055] El término "aminoácidos de variante natural" incluye restos de aminoácidos conservados, semiconservados y variables observados, de acuerdo con sus frecuencias de aparición, en una posición dada en un bucle seleccionado de una longitud seleccionada. Los aminoácidos de variante natural pueden estar sustituidos por aminoácidos químicamente equivalentes y pueden excluir restos de aminoácidos variables por debajo de una frecuencia de aparición seleccionada, por ejemplo, el 5-10 %, o restos de aminoácidos que son químicamente equivalentes a otros aminoácidos de variante natural.

[0056] El término "biblioteca de secuencias de mutagénesis" se refiere a una biblioteca de secuencias dentro de un bucle Fn3 seleccionado y una longitud de bucle que se expresa mediante una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido. Así, para cada uno de los aminoácidos diana, la biblioteca de secuencias dentro de un bucle dado contendrá el aminoácido diana en todas las combinaciones de una a todas las posiciones dentro del bucle en el que el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a la frecuencia umbral dada. Si esta

frecuencia umbral se establece en el 100 %, cada posición en el bucle contendrá el aminoácido diana en al menos un miembro de la biblioteca. Las "secuencias de mutagénesis de la biblioteca" se pueden generar a partir de las tablas y figuras descritas en este documento usando proveedores comerciales tales como Geneart o DNA2.0.

- 5 **[0057]** El término "biblioteca de secuencias combinatorias de variantes naturales" se refiere a una biblioteca de secuencias dentro de un bucle Fn3 seleccionado y una longitud de bucle que se expresa mediante una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, incluidos
10 aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral de aparición mínimo seleccionado en esa posición, o sus equivalentes químicos. Por lo tanto, para cada posición de aminoácido en un bucle y longitud de bucle seleccionada, la biblioteca de secuencias combinatorias de variantes naturales contendrá el aminoácido de consenso en esa posición más otras variantes de aminoácidos identificadas como que tienen al menos cierta frecuencia mínima en esa posición, por ejemplo, al menos un 5-10 % de frecuencia,
15 o aminoácidos químicamente equivalentes. Además, las variantes naturales pueden sustituirse o descartarse si la secuencia codificante de ese aminoácido produce un número significativo de aminoácidos coproducidos, a través de la degeneración del codón.

- [0058]** El término "perfil de variabilidad" o "VP" se refiere a la catalogación de aminoácidos y sus respectivas
20 frecuencias de aparición presentes en una posición del bucle particular. Las posiciones del bucle se derivan de un conjunto de datos de fibronectina alineados.

- [0059]** El término "aminoácido" o "resto de aminoácido" normalmente se refiere a un aminoácido que tiene su definición reconocida en la materia, tal como un aminoácido seleccionado del grupo que consiste en: alanina (Ala, A); arginina (Arg, R); asparagina (Asn, N); ácido aspártico (Asp, D); cisteína (Cys, C); glutamina (Gln, Q); ácido glutámico (Glu, E); glicina (Gly, G); histidina (His, H); isoleucina (Ile, I); leucina (Leu, L); lisina (Lys, K); metionina (Met, M); fenilalanina (Phe, F); prolina (Pro, P); serina (Ser, S); treonina (Thr, T); triptófano (Trp, W); tirosina (Tyr, Y); y valina (Val, V), aunque pueden usarse aminoácidos modificados, sintéticos o raros según se desee.

- 30 **[0060]** El término "aminoácidos químicamente equivalentes" se refiere a aminoácidos que tienen propiedades estéricas, de carga y de solubilidad similares. Un esquema común agrupa aminoácidos de la siguiente manera: (1) glicina, que tiene una cadena lateral de hidrógeno; (2) alanina (Ala, A), valina (Val, V), leucina (Leu, L) e isoleucina (Ile, I), que tienen hidrógeno o una cadena lateral alifática no sustituida; (3) serina (Ser, S) y treonina (Thr, T) que tienen una cadena lateral alifática que lleva un grupo hidroxilo; (4) ácido aspártico (Asp, D) y glutámico (Glu, E), que
35 tienen una cadena lateral que contiene carboxilo; (5) asparagina (Asn, N) y glutamina (Glu, Q), que tienen una cadena lateral alifática que termina en un grupo amida; (6) arginina (Arg, R), lisina (Lys, L) e histidina (His, H), que tienen una cadena lateral alifática que termina en un grupo amino básico; (7) cisteína (Cys, C) y metionina (Met, M), que tienen una cadena lateral alifática que contiene azufre; (8) tirosina (Tyr, Y) y fenilalanina (Phe, F), que tienen una cadena lateral aromática; y (9) triptófano (Trp, W), prolina (Pro, P) e histidina (His, H), que tienen una cadena
40 lateral heterocíclica.

- [0061]** El término "polinucleótido(s)" se refiere a ácidos nucleicos tales como moléculas de ADN y moléculas de ARN y análogos de los mismos (por ejemplo, ADN o ARN generado usando análogos de nucleótidos o usando química de ácidos nucleicos). Según se desee, los polinucleótidos pueden prepararse sintéticamente, por ejemplo,
45 usando química de ácidos nucleicos reconocida en la materia o usando enzimáticamente, por ejemplo, una polimerasa, y, si se desea, modificarse. Las modificaciones típicas incluyen metilación, biotilación y otras modificaciones conocidas en la materia. Además, la molécula de ácido nucleico puede ser monocatenaria o bicatenaria y, cuando se desee, unirse a un resto detectable. Las bases polinucleotídicas y los pares de bases alternativos reciben sus abreviaturas habituales en el presente documento: Adenosina (A), Guanosina (G), Citidina (C), Timidina (T), Uridina (U), puRina (R = A/G), pirimidina (Y = C/T o C/U), aMino (M = A/C), cetó (K = G/T o G/U), fuerte (S = G/C), débil (W = A/T o A/U), V (A o C o G, pero no T), N o X, (cualquier base).
50

- [0062]** El término "mutagénesis" se refiere, a menos que se especifique lo contrario, a cualquier técnica reconocida en la materia para alterar una secuencia polinucleotídica o polipeptídica. Los ejemplos de mutagénesis
55 incluyen, pero no se limitan a, mutagénesis de paso (WTM), mutagénesis combinatoria de variantes naturales y mutagénesis combinatoria de variantes naturales beneficiosa, aunque se pueden emplear otras bibliotecas de mutagénesis, tales como mutagénesis directa (LTM).

II. Descripción general del procedimiento y las bibliotecas

[0063] Los andamiajes de anticuerpos artificiales que se unen a ligandos específicos se están convirtiendo en alternativas legítimas a los anticuerpos. Los anticuerpos han sido útiles como herramientas diagnósticas y terapéuticas. Sin embargo, la obtención de anticuerpos específicos que reconocen ciertas dianas ha sido difícil. Las bibliotecas de anticuerpos actuales están sesgadas contra ciertas clases de antígenos solo después de la exposición inmunológica. Por lo tanto, con frecuencia es necesario inmunizar un animal huésped con un antígeno particular antes de que pueda producirse la recuperación de anticuerpos específicos. Además, los anticuerpos son difíciles y caros de producir y requieren reactores de fermentación celular y procedimientos de purificación especiales.

10 **[0064]** Las limitaciones de los anticuerpos han estimulado el desarrollo de proteínas de unión alternativas basadas en pliegues de tipo inmunoglobulina u otras topologías de proteínas. Estos andamiajes que no son anticuerpos comparten la cualidad general de tener un núcleo estructural estable que es tolerante a las sustituciones múltiples en otras partes de la proteína.

15 **[0065]** La presente descripción proporciona dominios de unión universales del lado inferior de la fibronectina y una biblioteca de dominios de unión del lado inferior de la fibronectina que usan las regiones del bucle del lado inferior (AB, CD o EF). Estos dominios de unión a Fn3 del lado inferior y las bibliotecas de los dominios de unión a Fn3 del lado inferior son más completos y están diseñados para tener diversidad artificial en los bucles de unión a la diana. Al crear una diversidad artificial, el tamaño de la biblioteca se puede controlar de modo que se pueda cribar fácilmente utilizando, por ejemplo, procedimientos de alto rendimiento para obtener nuevas terapias. La biblioteca de fibronectina universal con regiones de bucle del lado inferior se puede explorar usando selección de clones físicos positivos mediante FACS, selección de fagos o retención selectiva de ligandos. Estas evaluaciones *in vitro* evitan la metodología convencional y tediosa inherente a la generación de una biblioteca de hibridomas de anticuerpos y la detección de sobrenadantes.

25 **[0066]** Además, la biblioteca de fibronectina universal con las regiones de bucle del lado inferior (AB, CD o EF) tiene el potencial de reconocer cualquier diana ya que los aminoácidos constituyentes en el bucle de unión diana se crean mediante técnicas de diversidad *in vitro*. Esto produce las ventajas significativas de la biblioteca que controla el tamaño de la diversidad y la capacidad de reconocer autoantígenos. Además, la biblioteca de dominio de unión del lado inferior de la fibronectina con las regiones de bucle del lado inferior (AB, CD o EF) se puede propagar y volver a seleccionar para descubrir módulos de unión de fibronectina adicionales contra otras dianas deseadas.

Revisión de fibronectina: (Fn)

35 **[0067]** Las proteínas de Fibronectina de Tipo III (Fn3) se refieren a un grupo de proteínas compuestas por subunidades monoméricas que tienen una estructura o motivo de Fibronectina de Tipo III (Fn3) compuesto de siete cadenas β con tres bucles de conexión. Las cadenas β A, B y E forman una mitad β -sándwich y cadenas las β C, D, F y G forman la otra mitad (véase Fig. 1), y que tiene pesos moleculares de aproximadamente 94 aminoácidos y pesos moleculares de aproximadamente 10 KDa. El pliegue global del dominio Fn3 está estrechamente relacionado con el de los dominios de inmunoglobulina, y los tres bucles cerca del extremo N-terminal de Fn3, denominados BC, DE y FG, pueden considerarse estructuralmente análogos a regiones determinantes de la complementariedad del dominio variable de anticuerpo (VH), CDR1, CDR2 y CDR3, respectivamente. Se ha pensado generalmente que los bucles inferiores de Fn3 confieren estabilidad estructural en lugar de usarse para unirse a dianas. Sin embargo, los procedimientos de la descripción demuestran que los bucles inferiores pueden usarse para la unión a dianas y para generar bibliotecas de moléculas de unión de Fn3 que usan los bucles superiores, los bucles inferiores o cualquier combinación de bucles superior e inferior para la unión.

[0068] La secuencia de aminoácidos de la fibronectina revela tres tipos de repeticiones o módulos internamente homólogos separados por secuencias de conexión (generalmente) cortas. Hay 12 módulos de tipo I, 2 de tipo II y 16 de tipo III, y se conocen como Fn I, Fn II y Fn III, respectivamente. Cada módulo Fn constituye una unidad plegada independientemente, a menudo denominada dominio. Dentro de la fibronectina en sí, hay dieciséis dominios Fn3 y tienen estructuras terciarias notablemente similares. Si bien la conformación de Fn3 está altamente conservada, la similitud entre diferentes módulos del mismo tipo dentro de una proteína de fibronectina dada es bastante baja, normalmente menos del 20 %.

Identificación y selección de componentes de andamiaje y fibronectina mediante bioinformática

[0069] El primer paso para construir una biblioteca de fibronectina universal de esta descripción es seleccionar secuencias que cumplan con ciertos criterios predeterminados. Se hizo una búsqueda en PFAM, ProSite

y bases de datos similares para secuencias que contengan dominios Fn3 (Fig. 1). Estas bases de datos electrónicas contienen fibronectina expresada catalogada y secuencias de proteínas de tipo fibronectina y se pueden consultar para aquellos módulos de Fn3 y secuencias similares (por ejemplo, usando el algoritmo de búsqueda BLAST). Las secuencias del módulo Fn3 se pueden agrupar según criterios predefinidos, como subclases de módulo, similitud de secuencia u organismo(s) de origen. La selección de secuencia marco también se puede realizar para proteínas de andamiaje tales como Fn I, Fn II o anquirina y otras proteínas, tales como SCALP, SmALP, NANOBOODIES y similares.

[0070] A continuación, se delinean secuencias de marco de andamiaje de la cadena β de Fn3 candidatas con lo que a continuación se identifican las regiones del bucle intermedio y los aminoácidos constituyentes. Esto determina entonces la longitud de los bucles existentes, los perfiles de aminoácidos para cada longitud de bucle y, por lo tanto, el tamaño físico y la diversidad de aminoácidos que se pueden acomodar dentro de estos marcos. Una vez que se identifican los bucles, las secuencias dentro de cada bucle se alinean, y las secuencias alineadas se dividen en grupos según la longitud del bucle. Se identificó la distribución de las longitudes del bucle para los bucles AB, CD y EF. Con esta información, se seleccionan los tamaños de bucle más comunes. En un ejemplo general de esta descripción, las longitudes del bucle seleccionadas son AB/3, CD/5, CD/6, CD/7 y EF/6, DE/6. Para cada cadena β , uno puede determinar los sitios preferidos de aceptores de bucle en los marcos basados tanto en análisis comparativo estructural como de secuencia. Por ejemplo, se puede usar la comparación de superposición estructural de los andamiajes de bucle entero y bucle β entre la fibronectina $^{10}\text{Fn3}$, $^{14}\text{Fn3}$ o cualquiera de los otros dominios Fn3 conocidos. En la identificación de posiciones del bucle precisas, la etapa anterior minimiza en gran medida las mutaciones del bucle de diversidad necesarias que no darían como resultado una especificidad de unión al ligando funcional.

[0071] Una vez que se seleccionan las longitudes del bucle, se realiza un análisis de frecuencia posicional de aminoácidos en cada posición del bucle, para determinar la frecuencia de aparición, en un conjunto de módulos de Fn3 nativos. Este procedimiento incluye un análisis de frecuencia y la generación de los correspondientes perfiles de variabilidad (VP) de las secuencias de bucle existentes (Véase Ejemplos y Figs. 3-10) Las posiciones de alta frecuencia (por ejemplo, > 50 %) se consideran conservadas o fijas. Los aminoácidos de frecuencia moderadamente alta o "semiconservados" o (cuando se combinan 2 o 3 representan > 40 %) se eligen como "tipo silvestre" en otras posiciones. Estos aminoácidos de tipo silvestre se alteran sistemáticamente usando mutagénesis, por ejemplo, mutagénesis directa, para generar la biblioteca de bucle universal. Las posiciones "variables" son aquellas en las que, por lo general, ningún aminoácido representa más del 20 % del conjunto representado.

[0072] La elección de las estructuras candidatas basadas en los criterios de esta descripción dicta tanto los tamaños de bucle que se introducirán como la diversidad inicial de la secuencia de aminoácidos.

[0073] Un análisis de perfil de variabilidad de bucle de las bases de datos de Fn3 permite la identificación de posiciones de restos de aminoácidos de bucle que se dividen en tres categorías, por ejemplo, 1) posiciones que deben conservarse o "fijas" y 2) posiciones semiconservadas y/o 3) posiciones variables que son adecuadas para la generación de diversidad. Se realiza un análisis de perfil de variabilidad y se utiliza una frecuencia umbral para identificar las secuencias más favorables que se utilizarán para designar la diversidad global de bucles.

[0074] La secuencia conservada o una secuencia semiconservada seleccionada (normalmente el aminoácido más frecuente en los restos semiconservados) se considera el resto de "tipo silvestre" o "consenso" en la secuencia de bucle. Este enfoque de "consenso" o "frecuencia" identifica aquellos aminoácidos particulares bajo alta presión selectiva que ocurre con mayor frecuencia en una posición particular.

[0075] De acuerdo con esto, estas posiciones de restos normalmente son fijas, introduciéndose la diversidad en las posiciones de aminoácidos restantes (teniendo en cuenta la preferencia identificada para que ciertos aminoácidos estén presentes en estas posiciones). El umbral para la frecuencia de aparición a la cual se introducirá la variación de aminoácidos puede variar entre niveles seleccionados tan bajos como del 40 %, preferentemente del 50 % hasta tan alto como del 100 %. En la frecuencia umbral del 100 %, la mutagénesis de los aminoácidos se puede introducir en todas las posiciones del bucle, y las únicas limitaciones en los aminoácidos de variantes naturales serán el número total de variantes y si hay equivalentes químicos disponibles.

[0076] Cuando se diseña la diversidad para cualquiera de los bucles mencionados anteriormente, los restos de aminoácidos modificados, por ejemplo, restos fuera de los 20 aminoácidos tradicionales usados en la mayoría de los polipéptidos, por ejemplo, homocisteína, pueden incorporarse en los bucles según se desee. Esto se lleva a cabo usando técnicas reconocidas en la materia que normalmente introducen codones de parada en el polinucleótido

donde se desea el resto de aminoácido modificado. La técnica proporciona entonces un ARNt modificado unido al aminoácido modificado que se va a incorporar (un denominado ARNt supresor de, por ejemplo, el codón de parada ámbar, ópalo u ocre) en el polipéptido (véase, por ejemplo, Köhrer et al., Import of amber and ochre suppressors tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins, 5 PNAS, 98, 14310-14315 (2001)).

[0077] El análisis bioinformático se centra en los genes del módulo Fn3 con fines descriptivos, pero se entenderá que los genes para otros módulos Fn y otras proteínas de andamiaje se evalúan de forma similar. En un caso, el análisis bioinformático se puede aplicar para proteínas de tipo anticuerpo de cadena sencilla (SCALP), 10 nanoanticuerpos y similares.

Construcción de la biblioteca de fibronectina universal asistida por ordenador

[0078] Las bibliotecas universales del bucle inferior de la fibronectina de esta descripción, por ejemplo las de 15 la invención, y su construcción se llevan a cabo con el beneficio de información secuencial y estructural de manera que se incrementa el potencial para generar dominios de unión a fibronectina mejorados. La información de modelado de reemplazo molecular estructural también se puede usar para guiar la selección de la diversidad de aminoácidos a introducir en las regiones de bucle definidas. Además, los resultados reales obtenidos con los dominios de unión a fibronectina de esta descripción, por ejemplo con los de la invención, pueden guiar la selección 20 (o exclusión), por ejemplo, la maduración de afinidad, de dominios de unión de fibronectina posteriores para realizar y cribar de forma iterativa.

[0079] En un caso, por ejemplo, en una realización, el modelado *in silico* se usa para eliminar la producción de cualquier dominio de unión a fibronectina que se prediga que tiene una estructura y/o función pobre o no 25 deseada. De esta forma, el número de dominios de unión a fibronectina que se van a producir puede reducirse bruscamente, aumentando de este modo la señal a ruido en ensayos de cribado posteriores. En otro ejemplo particular, por ejemplo en otra realización particular, el modelado *in silico* se actualiza continuamente con información de modelado adicional, desde cualquier fuente relevante, por ejemplo, desde secuencias de genes y proteínas y bases de datos tridimensionales y/o resultados de dominios de unión a fibronectina probados 30 previamente, de modo que la base de datos *in silico* se vuelve más precisa en su capacidad predictiva (Fig. 1).

[0080] En otro ejemplo más, por ejemplo en otra realización, la base de datos *in silico* se proporciona con los resultados del ensayo, por ejemplo, afinidad/avidez de unión de dominios de unión a fibronectina probados 35 previamente y categoriza los dominios de unión a fibronectina, basándose en el criterio o criterios de ensayo, como respondedores o no respondedores, por ejemplo, como moléculas de dominio de unión a fibronectina que se unen bien o no tan bien. De esta forma, la maduración de afinidad de esta descripción, por ejemplo la de la invención, puede equiparar un rango de respuestas funcionales con información estructural y de secuencia particular y usar dicha información para guiar la producción de futuros dominios de unión a fibronectina a analizar. El procedimiento es especialmente adecuado para el cribado de dominios de unión a fibronectina para una afinidad de unión particular 40 a un ligando diana usando, por ejemplo, un ensayo Biacore.

[0081] Por consiguiente, puede ser deseable la mutagénesis de restos no contiguos dentro de una región de bucle si se conoce, por ejemplo, a través de la modelización *in silico*, que ciertos restos en la región no participarán en la función deseada. La estructura de coordenadas y la interrelación espacial entre las regiones definidas, por 45 ejemplo, se pueden considerar y modelar los restos de aminoácidos funcionales en las regiones definidas del dominio de unión a fibronectina, por ejemplo, la diversidad que se ha introducido. Dichos criterios de modelado incluyen, por ejemplo, química del grupo lateral de restos de aminoácidos, distancias entre átomos, datos de cristalografía, etc. En consecuencia, el número de dominios de unión a fibronectina que se van a producir se puede minimizar de forma inteligente. 50

[0082] En un caso, por ejemplo, en una realización, una o más de las etapas anteriores están asistidas por ordenador. En un caso particular, por ejemplo en una realización particular, la etapa asistida por ordenador comprende, por ejemplo, extraer de las bases de datos NCBI, Genbank, PFAM y ProSite y, opcionalmente, hacer referencias cruzadas de los resultados contra la base de datos estructural PDB, por lo que ciertos criterios de la 55 invención se determinan y se usan para diseñar la diversidad de bucles deseada (Fig. 1). El procedimiento también se puede llevar a cabo, en parte o en su totalidad, por un dispositivo, por ejemplo, un dispositivo accionado por ordenador. Por ejemplo, la selección de secuencia del módulo de fibronectina de minería de datos, el diseño de diversidad, la síntesis de oligonucleótidos, el ensamblaje mediado por PCR de los anteriores, y la expresión y selección de dominios de unión de fibronectina candidatos que se unen a una diana dada, puede llevarse a cabo en

parte o completamente, mediante dispositivos entrelazados. Además, las instrucciones para llevar a cabo el procedimiento, en parte o en su totalidad, se pueden conferir a un medio adecuado para su uso en un dispositivo electrónico para llevar a cabo las instrucciones. En resumen, los procedimientos de esta descripción, que incluyen los de la invención, son modificables para un enfoque de alto rendimiento que comprende software (por ejemplo, instrucciones legibles por ordenador) y hardware (por ejemplo, ordenadores, robótica y chips).

Bibliotecas universales de mutagénesis

[0083] La presente descripción se refiere a una biblioteca de mutagénesis de polipéptidos del dominio de fibronectina de Tipo 3 útiles en el cribado para la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada. Los polipéptidos de la biblioteca incluyen (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre de un polipéptido o polipéptidos de fibronectina de tipo 3 nativos seleccionados, y (b) regiones de bucle AB, CD y EF que tienen una o más longitudes seleccionadas. Al menos una región de bucle seleccionada de una longitud seleccionada contiene una biblioteca de secuencias codificadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una aparición frecuencia igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido (aminoácidos producidos por las secuencias codificantes en una posición dada como resultado de la degeneración del codón).

[0084] Al construir una biblioteca dentro de un bucle dado de una longitud de bucle dada, el perfil de variabilidad se usa para definir una secuencia de posiciones fijas y "variables", es decir, posiciones en las que puede introducirse un aminoácido diana. El número de posiciones fijas dependerá de la frecuencia umbral seleccionada para el aminoácido de consenso en cada posición.

[0085] Una vez que se seleccionan las secuencias de bucle, se construye una biblioteca de oligonucleótidos de secuencia codificante que codifica todas las secuencias identificadas, haciendo las sustituciones de codones como se muestra que son eficaces para conservar el aminoácido de consenso existente, pero también codifican el aminoácido diana seleccionado, y cualquier otro aminoácido coproducto codificado por codones degenerados.

[0086] La biblioteca de secuencias codificantes para los bucles se añade a las secuencias marco, para construir la biblioteca de secuencias codificantes para las bibliotecas de polipéptidos. La biblioteca de polipéptidos puede estar codificada por un formato de biblioteca de expresión que incluye una biblioteca de presentación en ribosomas, una biblioteca de presentación en polisomas, una biblioteca de presentación en fagos, una biblioteca de expresión bacteriana o una biblioteca de presentación en levaduras.

[0087] Las bibliotecas se pueden usar en un procedimiento para identificar un polipéptido que tiene una afinidad de unión deseada, en el que la biblioteca combinatoria de variantes naturales se criba para seleccionar un dominio de unión a fibronectina que tiene una afinidad de unión deseada.

[0088] En otro caso, pueden generarse bibliotecas universales de mutagénesis combinatoria de variantes naturales. La descripción incluye una biblioteca combinatoria de variantes naturales de polipéptidos del dominio de fibronectina de Tipo 3 útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión seleccionada. Los polipéptidos de la biblioteca incluyen (a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre de un polipéptido o polipéptidos de fibronectina de tipo 3 nativos seleccionados, y (b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas. Al menos una región de bucle seleccionada de una longitud seleccionada contiene una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, incluidos aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral de aparición mínimo seleccionado en esa posición, o sus equivalentes químicos.

[0089] Al construir una biblioteca combinatoria de variantes naturales para un bucle y una longitud de bucle dados, el perfil de variabilidad se usa para definir una secuencia de posiciones fijas y "variables", es decir, posiciones en las que pueden introducirse variaciones de aminoácidos. En las bibliotecas, el número de posiciones fijas dependerá de la frecuencia umbral seleccionada para el aminoácido de consenso en cada posición. Si, por ejemplo, el umbral de frecuencia seleccionado se estableció en aproximadamente un 60 %, los restos conservados o

semiconservados y las sustituciones de variantes naturales no se realizarían en estas posiciones. Por el contrario, si la frecuencia umbral se establece en el 100 %, todas las posiciones se considerarían abiertas a la variación, reconociendo que un solo aminoácido con una frecuencia del 100 % en una posición del bucle no se sustituiría, y una posición que tenía un aminoácido muy dominante, por ejemplo, con una frecuencia del 90 %, podría sustituirse solo si la variante o variantes de baja frecuencia fueran químicamente diferentes al aminoácido dominante.

[0090] A partir del perfil de aminoácidos para un bucle y una longitud de bucle dados, y sabiendo cuál de las posiciones se mantendrá fija y cuáles admitirán variaciones, se pueden seleccionar las sustituciones de aminoácidos en cada posición variable. En general, el número de variaciones que se seleccionan (incluidos los aminoácidos coproducidos) dependerá del número de posiciones de sustitución variable en el bucle y del número promedio de variaciones por posición del bucle sustituido. Por supuesto, si las sustituciones de variante natural se introducen en un único bucle único, se pueden acomodar muchas más variaciones por posición.

[0091] Los aminoácidos de variantes naturales particulares que se seleccionan para cada posición generalmente incluirán los aminoácidos que tienen las frecuencias más altas, al tiempo que limitan el número de aminoácidos coproducidos, y de forma secundaria, preservan la diversidad química en cada sitio. Una vez que se seleccionan las secuencias de bucle de variante natural, se construye una biblioteca de oligonucleótidos de secuencia codificante que codifican todas las secuencias variantes naturales identificadas, realizando sustituciones de codones que son eficaces para conservar el aminoácido de consenso existente y codifican los aminoácidos de variantes seleccionadas, incluidas variantes codificadas por codones degenerados.

[0092] La biblioteca de secuencias codificantes para los bucles de variantes naturales se añade a las secuencias marco, para construir la biblioteca de secuencias codificantes para las bibliotecas de polipéptidos variantes naturales. En un caso, la biblioteca codificante incluye secuencias codificantes para un par de bucles AB/CD, AB/EF o CD/EF, donde cada bucle en el par tiene una longitud seleccionada. Después de seleccionar polipéptidos (o enzimáticos) de alta afinidad de unión de esta biblioteca, se puede construir una segunda biblioteca "beneficiosa" que incluya las mutaciones beneficiosas contenidas en una o ambas de la biblioteca original de variación natural de dos bucles, y aminoácidos de variantes naturales en el tercer bucle, es decir, el bucle de secuencia previamente fija. En otro caso, la biblioteca codificante incluye secuencias codificantes para los tres bucles, AB, CD y EF.

[0093] La biblioteca de polipéptidos puede estar codificada por una biblioteca de expresión que tiene el formato de una biblioteca de presentación en ribosomas, una biblioteca de presentación en polisomas, una biblioteca de presentación en fagos, una biblioteca de expresión bacteriana o una biblioteca de presentación en levaduras.

Síntesis de bibliotecas universales de dominio de unión del lado inferior de la fibronectina

[0094] En un caso, por ejemplo en una realización, los dominios universales de unión del lado inferior de la fibronectina de esta descripción, por ejemplo los de la invención, se generan para el cribado sintetizando oligonucleótidos individuales que codifican la región definida del polipéptido y no tienen más de un codón para el aminoácido predeterminado. Esto se lleva a cabo incorporando, en cada posición del codón dentro del oligonucleótido, el codón requerido para la síntesis del polipéptido de tipo silvestre o un codón para el aminoácido predeterminado y se denomina mutagénesis directa (LTM) (véase, por ejemplo, Publicación de Patente de Estados Unidos n.º 20050136428).

[0095] En otro caso, por ejemplo, en otra realización, cuando se requiere diversidad en múltiples posiciones de aminoácidos, puede usarse la mutagénesis de paso (WTM) (véase, por ejemplo, Patente de Estados Unidos n.º 6 649 340; 5 830 650; y 5 798 208; y Publicación de Patente de Estados Unidos n.º 20050136428). En otro caso, por ejemplo en otra realización, se puede crear diversidad usando los procedimientos disponibles de proveedores comerciales tales como DNA2.0 y Geneart proporcionando información sobre la longitud del bucle de los bucles AB, CD y EF, la distribución posicional de aminoácidos en cada posición del bucle, y la abundancia de 7 aminoácidos superiores en cada posición del bucle.

[0096] Además, los procedimientos de la descripción también proporcionan otras bibliotecas denominadas "bibliotecas N+/-". Estas bibliotecas N+/- están construidas con variaciones en los bucles inferiores, AB, CD y EF, los bucles superiores, BC, DE, FG o cualquier combinación de bucles superior e inferior. Para "bibliotecas N+/-", N es el aminoácido predominante en una posición particular y los aminoácidos aguas arriba o aguas abajo se designan +N o -N, respectivamente. Por ejemplo, N+3 es un aminoácido 3 posiciones aguas arriba de N, mientras que N-3 es un aminoácido 3 posiciones aguas abajo de N en una estructura tridimensional de Fn3. Del mismo modo, N+2 y N+1

son aminoácidos en las posiciones 2 y 1 aguas arriba de N, respectivamente, mientras que N-2 y N-1 son aminoácidos en las posiciones 2 y 1 aguas abajo de N, respectivamente. Alterando N del aminoácido predominantemente más abundante a un aminoácido menos abundante, el efecto de esa modificación puede evaluarse en la abundancia de aminoácidos a 1, 2 o 3 posiciones de distancia de N. Al diseñar dicha biblioteca, se

- 5 determinan la frecuencia y la abundancia de aminoácidos que rodean la posición N fija. Estas diferencias pueden usarse para generar bibliotecas universales de dominio de unión en el lado inferior, bibliotecas de dominio de unión en el lado superior o una combinación de bibliotecas de dominio de unión tanto en el lado inferior como en el lado superior de la fibronectina.
- 10 **[0097]** Solo con fines ilustrativos, la secuencia de consenso en el bucle CD/5 como se muestra en las Figs. 6AB, es GDGQP en las posiciones del bucle 1, 2, 3, 4 y 5, siendo G el aminoácido predominante. Usando la teoría N+/-, si G en la posición 3 del bucle se fija ya que es el aminoácido predominante (N), entonces se determina el efecto estructural y microambiental de G en la posición 1 del bucle (N-2), la posición 2 del bucle (N-1), la posición 4 del bucle (N+1) y la posición 5 del bucle (N+2). Se calcula la frecuencia de aminoácidos de cada posición N-2, N-1,
- 15 N+1, N+2 en el contexto de un G fijo en la posición N. Entonces, si G en la posición 3 del bucle se cambia a S, se determina el efecto de S en las posiciones N-2, N-1, N+1, N+2 (es decir, las posiciones del bucle 1, 2 y 4, 5), y así sucesivamente. Después de calcular todas las combinaciones posibles, la información producida es una distribución de aminoácidos (N-2, N-1, N, N+1, N+2) de una posición N dada dentro de una región de bucle predeterminada en el contexto de un aminoácido específico en esta posición N. Esta información se puede usar para generar una
- 20 biblioteca.

- [0098]** En otra ilustración, la secuencia de consenso de EF/6 como se muestra en las Figs. 10AB, es GLKPGT en las posiciones del bucle 1, 2, 3, 4, 5 y 6, respectivamente, del bucle EF. Usando la teoría N+/-, si L en la posición 2 se mantiene fijo ya que es el aminoácido predominante (N), entonces se determina el efecto microambiental estructural y local sobre G en la posición 1 (N-1), K en la posición 3 (N+1), P en la posición 4 (N+2), G en la posición 5 (N+3), y T en la posición 6 (N+4). Además, si L en la posición 2 se cambia a V, entonces se determina el efecto de este cambio en las posiciones N-1, N+1, N+2, N+3 y N+4 (es decir, las posiciones del bucle 1, 3, 4, 5 y 6).

- 30 **[0099]** La mezcla de oligonucleótidos para la generación de la biblioteca se puede sintetizar fácilmente mediante procedimientos conocidos para la síntesis de ADN. El procedimiento preferido implica el uso de la fase sólida química de beta-cianoetil fosoramidita (por ejemplo, véase la Patente de EE.UU. n.º 4 725 677). Por conveniencia, se puede usar un instrumento para la síntesis automatizada de ADN que contiene recipientes reactivos especificados de nucleótidos. Los polinucleótidos también se pueden sintetizar para que contengan sitios
- 35 de restricción o sitios de hibridación de cebador para facilitar la introducción o el ensamblaje de los polinucleótidos que representan, por ejemplo, una región definida, en un contexto de un gen más grande.

- [0100]** Los polinucleótidos sintetizados se pueden insertar en un contexto de un gen más grande, por ejemplo, un dominio de andamiaje único usando técnicas de ingeniería genética convencionales. Por ejemplo, se
- 40 puede hacer que los polinucleótidos contengan sitios de reconocimiento flanqueantes para las enzimas de restricción (por ejemplo, véase la Patente de EE.UU. n.º 4 888 286). Los sitios de reconocimiento pueden diseñarse para que correspondan a sitios de reconocimiento que existen de forma natural o se introducen en el gen próximo al ADN que codifica la región. Después de la conversión en una forma bicatenaria, los polinucleótidos se ligan en el gen o vector genético mediante técnicas convencionales. Por medio de un vector apropiado (que incluye, por
- 45 ejemplo, vectores de fagos, plásmidos), los genes pueden introducirse en un extracto libre de células, fago, célula procariota o célula eucariótica adecuada para la expresión de las moléculas del dominio de unión a fibronectina.

- [0101]** Cuando se usan polinucleótidos parcialmente solapantes en el ensamblaje del gen, también se puede incorporar directamente un conjunto de nucleótidos degenerados en lugar de uno de los polinucleótidos. La cadena
- 50 complementaria apropiada se sintetiza durante la reacción de extensión a partir de un polinucleótido parcialmente complementario de la otra cadena mediante extensión enzimática con una polimerasa. La incorporación de los polinucleótidos degenerados en la etapa de síntesis también simplifica la clonación en la que más de un dominio o región definida de un gen se muta o diseña para tener diversidad.

- 55 **[0102]** En otro enfoque, el dominio de unión a fibronectina está presente en un plásmido de cadena sencilla. Por ejemplo, el gen puede clonarse en un vector de fago o un vector con un origen de replicación de fago filamentoso que permite la propagación de moléculas monocatenarias con el uso de un fago auxiliar. La plantilla monocatenaria puede reasociarse con un conjunto de polinucleótidos degenerados que representan las mutaciones deseadas y alargarse y ligarse, incorporando así cada cadena análoga en una población de moléculas que pueden

introducirse en un huésped apropiado (véase, por ejemplo, Sayers, J. R. et al. al., *Nucleic Acids Res.* 16: 791-802 (1988)). Este enfoque puede evitar múltiples pasos de clonación donde se seleccionan múltiples dominios para la mutagénesis.

5 **[0103]** La metodología de la reacción en cadena de la polimerasa (PCR) también se puede usar para incorporar polinucleótidos en un gen, por ejemplo, la diversidad de bucles en las regiones marco de la cadena β . Por ejemplo, los propios polinucleótidos pueden usarse como cebadores para la extensión. En este enfoque, los polinucleótidos que codifican los casetes mutagénicos correspondientes a la región definida (o parte de la misma) son complementarios entre sí, al menos en parte, y pueden extenderse para formar un gran casete génico (por
10 ejemplo, un dominio de unión a fibronectina) usando una polimerasa, por ejemplo, usando amplificación por PCR.

[0104] El tamaño de la biblioteca variará dependiendo de la longitud del bucle y la cantidad de diversidad de secuencia que necesita representarse utilizando procedimientos de mutagénesis. Por ejemplo, la biblioteca está diseñada para contener menos de 10^{15} , 10^{14} , 10^{13} , 10^{12} , 10^{11} , 10^{10} , 10^9 , 10^8 , 10^7 y 10^6 dominios de unión a
15 fibronectina.

[0105] La descripción anterior se ha centrado en representar la diversidad del dominio de unión a fibronectina alterando el polinucleótido que codifica el polipéptido correspondiente. Se entiende, sin embargo, que el alcance de esta descripción también abarca los procedimientos de representación de la diversidad del dominio de unión a fibronectina descrita en este documento por síntesis directa de las regiones polipeptídicas deseadas utilizando
20 química de proteínas. Al llevar a cabo este enfoque, los polipéptidos resultantes todavía incorporan las características de esta descripción, excepto por que puede eliminarse el uso de un intermedio de polinucleótido.

[0106] Para las bibliotecas descritas anteriormente, ya sea en forma de polinucleótidos y/o polipéptidos
25 correspondientes, se entiende que las bibliotecas también pueden estar unidas a un soporte sólido, tal como un microchip, y preferentemente dispuestas en matrices, usando técnicas reconocidas en la materia.

[0107] El procedimiento de esta descripción es especialmente útil para modificar las moléculas de dominio de unión a fibronectina candidatas por medio de la maduración de afinidad. Se pueden introducir alteraciones en los
30 bucles y/o en la región marco (constante) de la cadena β de un dominio de unión a fibronectina. La modificación de las regiones de bucle puede producir dominios de unión a fibronectina con mejores propiedades de unión a ligando y, si se desea, propiedades catalíticas. La modificación de la región marco de la cadena β también puede conducir a la mejora de las propiedades químico-físicas, tales como la solubilidad o la estabilidad, que son especialmente útiles, por ejemplo, en la producción comercial, la biodisponibilidad y la afinidad por el ligando. Normalmente, la
35 mutagénesis se dirigirá a la región o regiones de bucle del dominio de unión a fibronectina, es decir, la estructura responsable de la actividad de unión a ligando que puede estar constituida por las tres regiones de bucle. En un caso preferido, una molécula de unión candidata identificada se somete a maduración por afinidad para aumentar la afinidad/avidez de la molécula de unión por un ligando diana.

40 Sistemas de expresión y cribado

[0108] Las bibliotecas de polinucleótidos generadas mediante cualquiera de las técnicas anteriores u otras técnicas adecuadas se pueden expresar y cribar para identificar moléculas de dominio de unión a fibronectina que tienen la estructura y/o actividad deseadas. La expresión de las moléculas del dominio de unión a fibronectina se
45 puede llevar a cabo usando extractos libres de células (y, por ejemplo, presentación en ribosomas), presentación en fagos, células procariotas o células eucarióticas (por ejemplo, presentación en levaduras).

[0109] En un caso, por ejemplo, en una realización, los polinucleótidos se modifican por ingeniería genética para que sirvan como plantillas que se pueden expresar en un extracto libre de células. Se pueden usar vectores y
50 extractos como se describe, por ejemplo, en la Patente de Estados Unidos n.º 5 324 637; 5 492. 817; 5 665 563, y muchos están disponibles en el mercado. Se puede usar la presentación en ribosomas y otras técnicas libres de células para unir un polinucleótido (es decir, un genotipo) a un polipéptido (es decir, un fenotipo), por ejemplo, Profusion™ (véase, por ejemplo, Patente de Estados Unidos n.º 6 348 315; 6 261 804; 6 258 558; y 6 214 553).

55 **[0110]** Como alternativa, los polinucleótidos de esta descripción, por ejemplo los de la invención, pueden expresarse en un sistema de expresión de *E. coli* conveniente, tal como el descrito por Pluckthun y Skerra. (Pluckthun, A. y Skerra, A., *Meth. Enzymol.* 178: 476-515 (1989); Skerra, A. et al., *Biotechnology* 9: 273-278 (1991)). Las proteínas mutantes se pueden expresar para la secreción en el medio y/o en el citoplasma de la bacteria, como se describe por M. Better y A. Horwitz, *Meth. Enzymol.* 178: 476 (1989). En un caso, por ejemplo, en una realización,

el dominio de unión a fibronectina está unido al extremo 3' de una secuencia que codifica una secuencia señal, tal como la secuencia señal ompA, phoA o pelB (Lei, SP et al., J. Bacteriol. 169: 4379 (1987)). Estas fusiones de genes se ensamblan en una construcción dicistrónica, de modo que se pueden expresar a partir de un solo vector, y se secretan en el espacio periplásmico de *E. coli* donde se replegarán y se pueden recuperar en forma activa. (Skerra, A, et al., Biotechnology 9: 273-278 (1991)).

[0111] En otro caso, por ejemplo en otra realización, las secuencias del dominio de unión a fibronectina se expresan en la superficie de la membrana de un procarionte, por ejemplo, *E. coli*, usando una señal de secreción y resto de lipidación como se describe, por ejemplo, en los documentos US20040072740A1; US20030100023A1; y US20030036092A1.

[0112] En otro caso más, por ejemplo, en otra realización más, los polinucleótidos pueden expresarse en células eucariotas tales como levadura usando, por ejemplo, presentación en levaduras como se describe, por ejemplo, en la Patente de Estados Unidos n.º 6 423 538; 6 331 391; y 6 300 065. En este enfoque, las moléculas del dominio de unión a fibronectina de la biblioteca se fusionan a un polipéptido que se expresa y se muestra en la superficie de la levadura.

[0113] También se pueden usar células eucariotas superiores para la expresión de las moléculas del dominio de unión a fibronectina de esta descripción, por ejemplo las de la invención, tales como células de mamífero, por ejemplo células de mieloma (por ejemplo, células NS/O), células de hibridoma o células de ovario de hámster chino (CHO). Normalmente, las moléculas del dominio de unión a fibronectina, cuando se expresan en células de mamífero, se diseñan para expresarse en el medio de cultivo, o expresarse en la superficie de dicha célula. El dominio de unión a fibronectina puede producirse, por ejemplo, como un solo módulo individual o como cadenas multiméricas que comprenden dímeros, trímeros, que pueden estar compuestos del mismo módulo o de diferentes tipos de módulos. (10Fn3-10Fn3: homodímero, 10Fn3-8Fn3: heterodímero).

[0114] El cribado del dominio de unión a fibronectina expresado (o dominio de unión a fibronectina producido por síntesis directa) puede realizarse por cualquier medio apropiado. Por ejemplo, la actividad de unión se puede evaluar mediante inmunoensayo convencional y/o cromatografía de afinidad. El cribado del dominio de unión a fibronectina de la invención para la función catalítica, por ejemplo, la función proteolítica, puede realizarse usando un ensayo de placa de hemoglobina convencional como se describe, por ejemplo, en la Patente de EE.UU. n.º 5 798 208. La determinación de la capacidad de un dominio de unión a fibronectina candidato para unirse a dianas terapéuticas puede analizarse *in vitro* usando, por ejemplo, un instrumento Biacore, que mide las velocidades de unión de un dominio de unión a fibronectina a una diana o ligando dados. Los ensayos *in vivo* se pueden realizar usando cualquiera de una serie de modelos animales y posteriormente se prueban, según corresponda, en seres humanos.

Análisis y cribado de bibliotecas Fn3 para la función catalítica.

[0115] Las bibliotecas de Fn3 también se pueden usar para detectar proteínas Fn3 que poseen actividad catalítica. El estudio de las proteínas ha revelado que ciertos aminoácidos desempeñan un papel crucial en su estructura y función. Por ejemplo, parece que solo un número discreto de aminoácidos participa en el evento catalítico de una enzima. Las serina proteasas son una familia de enzimas presentes en prácticamente todos los organismos, que han desarrollado un sitio catalítico estructuralmente similar caracterizado por la presencia combinada de serina, histidina y ácido aspártico. Estos aminoácidos forman una tríada catalítica que, posiblemente junto con otros determinantes, estabiliza el estado de transición del sustrato. El papel funcional de esta tríada catalítica se ha confirmado por sustituciones individuales y múltiples de serina, histidina y ácido aspártico por mutagénesis dirigida de serina proteasas y la importancia de la interacción entre estos restos de aminoácidos en la catálisis ahora ha sido bien establecida.

[0116] De manera similar, un gran número de otros tipos de enzimas se caracterizan por la conformación peculiar de su sitio catalítico y la presencia de ciertos tipos de restos de aminoácidos en el sitio que son los principales responsables del evento catalítico. Para una revisión extensa, véase Enzyme Structure and Mechanism, 1985, por A. Fersht, Freeman Ed., Nueva York.

[0117] Aunque está claro que ciertos aminoácidos son críticos para el mecanismo de catálisis, es difícil, cuando no imposible, predecir qué posición (o posiciones) debe ocupar un aminoácido para producir un sitio funcional tal como un sitio catalítico. Desafortunadamente, la configuración espacial compleja de las cadenas laterales de aminoácidos en las proteínas y la interrelación de las diferentes cadenas laterales en el bolsillo catalítico

de las enzimas no se comprenden lo suficiente como para permitir dichas predicciones. La mutagénesis selectiva dirigida a sitio y la mutagénesis de saturación son de utilidad limitada para el estudio de la estructura y función de proteínas en vista del enorme número de variaciones posibles en proteínas complejas.

5 **[0118]** Las bibliotecas de proteínas generadas por cualquiera de las técnicas anteriores u otras técnicas adecuadas pueden cribarse para identificar variantes de estructura o actividad deseadas.

[0119] Al comparar las propiedades de una proteína de tipo silvestre y las variantes generadas, es posible identificar aminoácidos individuales o dominios de aminoácidos que confieren actividad de unión y/o catalítica. Por lo general, la región estudiada será un dominio funcional de la proteína, como un dominio de unión. Por ejemplo, la región puede ser las regiones de unión del bucle AB, CD y EF del dominio Fn3. El cribado puede realizarse por cualquier medio apropiado. Por ejemplo, la actividad catalítica puede determinarse mediante ensayos adecuados para la conversión del sustrato y la actividad de unión se puede evaluar mediante inmunoensayo convencional y/o cromatografía de afinidad.

15 **[0120]** A partir de las propiedades químicas de las cadenas laterales, parece que solo un número seleccionado de aminoácidos naturales participa preferencialmente en un evento catalítico. Estos aminoácidos pertenecen al grupo de aminoácidos polares y neutros tales como Ser, Thr, Asn, Gln, Tyr y Cys, el grupo de aminoácidos cargados, Asp y Glu, Lys y Arg, y especialmente el aminoácido His. Las cadenas laterales polares y neutras típicas son las de Cys, Ser, Thr, Asn, Gln y Tyr. La Gly también se considera un miembro limítrofe de este grupo. La Ser y la Thr juegan un papel importante en la formación de enlaces de hidrógeno. La Thr tiene una asimetría adicional en el carbono beta, por lo tanto, solo se usa uno de los estereoisómeros. La amida de ácido Gln y Asn también pueden formar enlaces de hidrógeno, con los grupos amido que funcionan como donadores de hidrógeno y los grupos carbonilo que funcionan como aceptores. La Gln tiene un grupo CH₂ más que la Asn, lo que hace que el grupo polar sea más flexible y reduce su interacción con la cadena principal. La Tyr tiene un grupo hidroxilo muy polar (OH fenólico) que puede disociarse a valores de pH altos. La Tyr se comporta como una cadena lateral cargada; sus enlaces de hidrógeno son bastante fuertes.

30 **[0121]** La histidina (His) tiene una cadena lateral aromática heterocíclica con un valor de pK de 6,0. En el rango de pH fisiológico, su anillo de imidazol puede estar descargado o cargado, después de tomar un ion de hidrógeno de la solución. Dado que estos dos estados están fácilmente disponibles, la His es muy adecuada para catalizar reacciones químicas. Se encuentra en la mayoría de los centros activos de enzimas.

35 **[0122]** La Asp y la Glu están cargadas negativamente a pH fisiológico. Debido a su cadena lateral corta, el grupo carboxilo de la Asp es bastante rígido con respecto a la cadena principal. Esta puede ser la razón por la cual el grupo carboxilo en muchos sitios catalíticos es proporcionado por la Asp y no por la Glu. Los ácidos cargados generalmente se encuentran en la superficie de una proteína.

40 **[0123]** Por lo tanto, varias regiones diferentes o bucles de un dominio de proteína Fn3 pueden mutagenizarse simultáneamente. Esto permite la evaluación de sustituciones de aminoácidos en regiones relacionadas conformacionalmente tales como las regiones que, al plegar la proteína, se asocian para formar un sitio funcional tal como el sitio catalítico de una enzima o el sitio de unión de un anticuerpo. Este procedimiento proporciona una forma de crear sitios catalíticos modificados o completamente nuevos. Las tres regiones de bucle de Fn3, que pueden diseñarse para conferir unión al ligando diana, pueden mutagenizarse simultáneamente, o por separado dentro de los bucles AB, CD y EF para analizar las funciones catalíticas contribuyentes en este sitio de unión. Por lo tanto, la introducción de aminoácidos "catalíticamente importantes" adicionales en una región de unión a ligando de una proteína puede dar como resultado una actividad catalítica *de novo* hacia el mismo ligando diana.

50 **[0124]** Por lo tanto, se pueden construir nuevas estructuras en el "andamiaje" natural de una proteína existente mutando solo regiones relevantes mediante el procedimiento de esta descripción, por ejemplo, mediante el procedimiento de la invención. El procedimiento de esta descripción, por ejemplo, el procedimiento de la invención, es adecuado para el diseño de proteínas de unión catalítica *de novo* en comparación con el aislamiento de anticuerpos catalíticos de origen natural. En la actualidad, los anticuerpos catalíticos se pueden preparar mediante una adaptación de técnicas convencionales de fusión de células somáticas. En este proceso, un animal se inmuniza con un antígeno que se asemeja al estado de transición del sustrato deseado para inducir la producción de un anticuerpo que se une al estado de transición y cataliza la reacción. Las células productoras de anticuerpos se recogen del animal y se fusionan con una célula de inmortalización para producir células híbridas. Estas células se criban para detectar la secreción de un anticuerpo que cataliza la reacción. Este proceso depende de la disponibilidad de análogos del estado de transición de un sustrato. El proceso puede ser limitado porque es probable

que dichos análogos sean difíciles de identificar o sintetizar en la mayoría de los casos.

[0125] El procedimiento de esta descripción, por ejemplo, el procedimiento de la invención, se puede usar para producir muchas enzimas o anticuerpos catalíticos diferentes, que incluyen oxidorreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas. Entre estas clases, será de particular importancia la producción de proteasas, carbohidrasas, lipasas, dioxigenasas y peroxidasas mejoradas. Estas y otras enzimas que pueden prepararse mediante el procedimiento de esta descripción, por ejemplo mediante el procedimiento de la invención, tienen aplicaciones comerciales importantes para conversiones enzimáticas en el cuidado de la salud, cosméticos, alimentos, elaboración de cerveza, detergentes, medioambiente (por ejemplo, tratamiento de aguas residuales), agricultura, curtido, textiles y otros procesos químicos. Estos incluyen, entre otros, aplicaciones de diagnóstico y terapéuticas, conversiones de grasas, carbohidratos y proteínas, degradación de contaminantes orgánicos y síntesis de productos químicos. Por ejemplo, podrían diseñarse por ingeniería genética proteasas terapéuticamente efectivas con actividad fibrinolítica, o actividad contra las estructuras virales necesarias para la infectividad, tales como las proteínas de la cubierta viral. Dichas proteasas podrían ser agentes antitrombóticos o agentes antivirales útiles contra virus tales como el del sida, rinovirus, la gripe o la hepatitis. En el caso de las oxigenasas (por ejemplo, dioxigenasas), una clase de enzimas que requieren un cofactor para la oxidación de anillos aromáticos y otros dobles enlaces, aplicaciones industriales en procesos de biopulpaje, conversión de biomasa en combustibles u otros productos químicos, la conversión de contaminantes en aguas residuales, el bioprocesamiento de carbón y la detoxificación de compuestos orgánicos peligrosos son posibles aplicaciones de nuevas proteínas.

Procedimientos para generar bibliotecas universales de dominio de unión en el lado inferior de la fibronectina

[0126] En general, la práctica de la presente descripción, que incluye la práctica de la presente invención, emplea, a menos que se indique lo contrario, técnicas convencionales de química, biología molecular, tecnología de ADN recombinante, tecnología de PCR, inmunología (especialmente, por ejemplo, tecnología de anticuerpos), expresión sistemas (por ejemplo, expresión libre de células, presentación en fagos, presentación en ribosomas y Profusion™), y cualquier cultivo celular necesario que esté dentro de la experiencia de la técnica y se explica en la bibliografía. Véase, por ejemplo, Sambrook, Fritsch y Maniatis, Molecular Cloning: Cold Spring Harbor Laboratory Press (1989); DNA Cloning, Vols. 1 and 2, (D.N. Glover, Ed. 1985); Oligonucleotide Synthesis (M.J. Gait, Ed. 1984); PCR Handbook Current Protocols in Nucleic Acid Chemistry, Beaucage, Ed. John Wiley & Sons (1999) (Editor); Oxford Handbook of Nucleic Acid Structure, Neidle, Ed., Oxford Univ Press (1999); PCR Protocols: A Guide to Methods and Applications, Innis et al., Academic Press (1990); PCR Essential Techniques: Essential Techniques, Burke, Ed., John Wiley & Son Ltd (1996); The PCR Technique: RT-PCR, Siebert, Ed., Eaton Pub. Co. (1998); Current Protocols in Molecular Biology, eds. Ausubel et al., John Wiley & Sons (1992); Large-Scale Mammalian Cell Culture Technology, Lubiniecki, A., Ed., Marcel Dekker, Pub., (1990). Phage Display: A Laboratory Manual, C. Barbas (Ed.), CSHL Press, (2001); Antibody Phage Display, P O'Brien (Ed.), Humana Press (2001); Border et al., Yeast surface display for screening combinatorial polypeptide libraries, Nature Biotechnology, 15(6):553-7 (1997); Border et al., Yeast surface display for directed evolution of protein expression, affinity, and stability, Methods Enzymol., 328:430-44 (2000); presentación en ribosomas según lo descrito por Pluckthun et al. en la Patente de EE.UU. n.º 6 348 315 y Profusion™ como se describe en Szostak et al. en la Patente de EE.UU. n.º 6 258 558; 6 261 804; y 6 214 553, y expresión periplásmica bacteriana como se describe en la patente US20040058403A1.

[0127] Se pueden encontrar más detalles con respecto a la clasificación, identificación y análisis de fibronectina y secuencia de Fn3, por ejemplo, PFAM. A program to screen aligned nucleotide and amino acid sequences, Methods Mol. Biol. 1995; 51:1-15, y Wu et al. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001 March; 17(3):282-3; las bases de datos y los programas de búsqueda y análisis incluyen la base de datos PFAM en el Instituto Sanger (pfam.sanger.ac.uk); la base de datos ExPASy PROSITE (www.expasy.ch/prosite/); la web SBASE (hydra.icgeb.trieste.it/sbase/); BLAST (www.ncbi.nlm.nih.gov/BLAST/); CD-HIT (bioinformatics.ljcrf.edu/cd-hi/); EMBOSS (www.hqmp.mrc.ac.uk/Software/EMBOSS/); PHYLIP (evolution.genetics.washington.edu/phylip.html); y FASTA (fasta.bioch.virginia.edu).

Levadura:

[0128] La biblioteca de dominio de unión del lado inferior de la fibronectina se transfecta en los hospedadores bacterianos/de levadura receptores usando técnicas convencionales como se describe en los Ejemplos. La levadura puede acomodar fácilmente tamaños de biblioteca de hasta 10^7 , con 10^3 - 10^5 copias de cada proteína de fusión Fn II que se presentan en cada superficie celular. Las células de levadura se criban y separan fácilmente usando citometría de flujo y clasificación celular activada por fluorescencia (FACS) o perlas magnéticas. El sistema de

secreción eucariota de la levadura y las vías de glucosilación de la levadura también permiten que las moléculas de tipo Fn3 se muestren con azúcares unidos a N y O sobre la superficie celular. Los detalles de la presentación en levaduras se describen en la sección de Ejemplos.

- 5 **[0129]** En otro caso, por ejemplo, en otra realización, el sistema de presentación en levaduras utiliza el receptor de adhesión de levadura α -aglutinina para presentar las proteínas en la superficie celular. Las proteínas de interés, en este caso, las bibliotecas de Fn3, se expresan como socios de fusión con la proteína Aga2.

- [0130]** Estas proteínas de fusión se secretan desde la célula y se unen por puentes disulfuro a la proteína
10 Aga1, que está unida a la pared celular de la levadura (véase Invitrogen, literatura del producto pYD1 Yeast Display). El plásmido, por ejemplo, pYD1, preparado a partir de un huésped de *E. coli* mediante purificación de plásmido (Qiagen), se digiere con las enzimas de restricción, BamHI y NotI, desfosforiladas terminalmente con fosfatasa alcalina intestinal de ternera. La ligación de las bibliotecas de productos pYD1 y CR, la transformación y selección de *E. coli* (DH5 α) en placas de LB-ampicilina (50 mg/ml) se realizaron usando protocolos de biología molecular
15 convencional para amplificar las bibliotecas antes de la electroporación en huéspedes de células de levadura.

- [0131]** Los procedimientos para seleccionar variantes de bibliotecas de Fn3 expresadas que tienen afinidades sustancialmente más altas para ligandos diana (por ejemplo, TNF, VEGF, VEGF-R, etc.), en relación con el dominio
20 Fn3 de tipo silvestre de referencia, se pueden lograr de la siguiente manera.

- [0132]** Los ligandos de prueba candidatos (por ejemplo, TNF, VEGF, VEGF-R, etc.) están marcados fluorescentemente (directa o indirectamente mediante un enlace biotina-estreptavidina como se describe anteriormente). Aquellos clones de la biblioteca que se unen eficazmente a los antígenos marcados entonces se enriquecen mediante el uso de FACS. Esta población de células de levadura se vuelve a cultivar y se somete a
25 rondas de selección posteriores usando niveles aumentados de rigurosidad para aislar un subconjunto más pequeño de clones que reconocen la diana con mayor especificidad y afinidad. Las bibliotecas son fácilmente susceptibles a formatos de alto rendimiento, utilizando, por ejemplo, moléculas de dominio de unión a Fn3 marcadas con FITC anti-Myc-tag y análisis FACS para su identificación y confirmación rápidas. Además, hay etiquetas de carboxilo terminales que pueden utilizarse para controlar los niveles de expresión y/o normalizar las mediciones de afinidad de
30 unión.

- [0133]** Para verificar la visualización de la proteína de fusión Aga2-Fn3, una alícuota de células de levadura (8×10^5 células en 40 μ l) del medio de cultivo se centrifuga durante 5 minutos a 2300 rpm. El sobrenadante se aspira y el sedimento celular se lava con 200 μ l de tampón de PBS/BSA enfriado en hielo (PBS/BSA al 0,5 % en p/v). Las células se vuelven a sedimentar y se elimina el sobrenadante antes de volver a suspender en 100 μ l de tampón que contiene el TNF α biotinilado (200 nM). Se dejó que las células se unieran al TNF α a 20 °C durante 45 minutos, después de lo cual se lavaron dos veces con tampón PBS/BSA antes de la adición e incubación con estreptavidina-FITC (2 mg/l) durante 30 minutos en hielo. Se realizó otra ronda de lavado en tampón antes del volumen de resuspensión final de 400 μ l en PBS/BSA. Las células se analizaron entonces en FACSScan (Becton
40 Dickinson) usando el software CellQuest según las instrucciones del fabricante.

- [0134]** Para generar una biblioteca contra TNF α , las selecciones cinéticas de las bibliotecas del dominio de unión a fibronectina de TNF- α presentadas en levadura implican el marcaje inicial de células con ligando de TNF- α biotinilado seguido de persecución dependiente del tiempo en presencia de un gran exceso de ligando de TNF- α no biotinilado. Los clones con una cinética de disociación más lenta se pueden identificar mediante marcaje con estreptavidina-PE después del período de persecución y se pueden clasificar usando un clasificador de FACS de alta velocidad. Después de la inducción de Aga2-Fn3, las células se incuban con TNF α biotinilado a concentraciones de saturación (400 nM) durante 3 horas a 25 °C bajo agitación. Después de lavar las células, se realizó una persecución en frío de 40 horas usando TNF α no marcado (1 μ M) a 25 °C. Las células se lavan entonces dos veces
50 con tampón PBS/BSA, marcado con estreptavidina-PE (2 mg/ml) anti-HIS-FITC (25 nM) durante 30 minutos en hielo, se lavaron y se resuspendieron y a continuación se analizaron en un clasificador FACS ARIA.

Presentación en ribosomas:

- 55 **[0135]** La presentación en ribosomas utiliza una maquinaria de transcripción/traducción acoplada *in vitro* libre de células para producir bibliotecas de proteínas. Los genes de la biblioteca Fn3 se insertan aguas arriba en el gen de la inmunoglobulina kappa ligera que no tiene un codón de terminación que hace que el ribosoma se bloquee, pero no se libere, cuando alcanza el extremo del ARNm. Adicionalmente, el espaciador del dominio kappa sirve para distanciar físicamente la proteína Fn3 del complejo ribosómico de modo que el dominio de unión a Fn3 tenga un

mejor acceso para reconocer su ligando afín. La biblioteca de ARNm se introduce en preparaciones de extracto de ribosoma S30 de *E. coli* (Roche) o en lisado reticulado de conejo (Promega). En cualquier caso, el extremo 5' del ARNm naciente se puede unir a los ribosomas y someterse a la traducción. Durante la traducción, la proteína de unión al ligando permanece unida de forma no covalente al ribosoma junto con su progenitor de ARNm en un complejo macromolecular.

[0136] Las proteínas Fn3 funcionales pueden unirse a un ligando específico que está unido a perlas magnéticas o a la superficie del pocillo de microtitulación. Durante el proceso de enriquecimiento, las variantes no específicas se eliminan por lavado antes de que se eluyan los aglutinantes de Fn3 específicos. El ARNm unido se detecta mediante RT-PCR usando cebadores específicos para la porción 5' Fn3 y 3' del gen kappa, respectivamente. El ADNc de doble cadena amplificado a continuación se clona en un vector de expresión para el análisis de la secuencia y la producción de proteínas.

[0137] Para reacciones de traducción procariotas, la mezcla de reacción puede contener 0,2 M de glutamato de potasio, 6,9 mM de acetato de magnesio, 90 mg/ml de proteína disulfuro isomerasa (Fluka), 50 mM de Tris acetato (pH 7,5), 0,35 mM de aminoácidos, 2 mM de ATP, 0,5 mM de GTP, 1 mM de AMPc, 30 mM de acetil fosfato, 0, mg/ml de ARNt de *E. coli*, 20 mg/ml de ácido fólico, 1,5 % de PEG 8000, 40 ml de extracto S30 de *E. coli* y 10 mg de ARNm en un volumen total de 110 ml. La traducción se puede realizar a 37 °C durante 7 min, después de lo cual los complejos ribosómicos se pueden estabilizar mediante dilución 5 veces en tampón de selección enfriado con hielo (50 mM de Tris acetato (pH 7,5), 150 mM de NaCl, 50 mM de acetato de magnesio, 0,1 % de Tween 20, 2,5 mg/ml de heparina).

Selección de afinidad para ligandos diana.

[0138] Los complejos de ribosomas estabilizados pueden incubarse con hapteno biotinilado (50 nM de fluoresceína-biotina (Sigma)) o antígeno (100 nM de IL-13 (Peprotech) biotinilado) según corresponda a 4 °C durante 1-2 h, seguido de captura en perlas magnéticas recubiertas de estreptavidina M280 (Dyna). Las perlas se lavaron a continuación para eliminar complejos de ribosomas no específicamente unidos. Para las selecciones procariotas, se pueden realizar cinco lavados en tampón de selección enfriado en hielo. Para las selecciones eucariotas, se realizaron tres lavados en PBS que contenían BSA al 0,1 % y acetato de magnesio 5 mM, seguido de un único lavado en PBS solo. Entonces se pueden incubar los complejos de eucariotas con 10 U de ADNasa I en 40 mM de Tris-HCl, 6 mM de MgCl₂, 10 mM de NaCl, 10 mM de CaCl₂ durante 25 min a 37 °C, seguido por tres lavados con PBS, acetato de magnesio 5 mM, Tween 20 al 1 %.

Recuperación de ARNm de complejos ribosómicos seleccionados

[0139] Para el análisis de recuperación de ARNm sin una etapa de interrupción específica, los complejos ribosómicos unidos a perlas magnéticas pueden procesarse directamente en la reacción de transcripción inversa. Para la recuperación de ARNm de selecciones procariotas por disrupción del complejo ribosómico, se pueden incubar complejos seleccionados en EB20 [Tris acetato 50 mM (pH 7,5), NaCl 150 mM, EDTA 20 mM, 10 mg/ml de ARN de *Saccharomyces cerevisiae*] durante 10 min a 4 °C. Para evaluar la eficacia del EDTA 20 mM para la recuperación de ARNm de selecciones eucariotas, los complejos ribosómicos pueden incubarse en PBS20 (PBS, EDTA 20 mM, 10 mg/ml de ARN de *S. cerevisiae*) durante 10 minutos a 4 °C. El ARNm puede purificarse usando un kit comercial (High Pure RNA Isolation Kit, Roche). Para muestras procariotas, se realizó la opción de digestión con el kit de ADNasa I; sin embargo, este paso no es necesario para muestras eucariotas, ya que la digestión con ADNasa I se realizó durante los lavados posteriores a la selección. La transcripción inversa se puede realizar en 4 ml de ARN purificado o 4 ml de complejos ribosómicos seleccionados, inmovilizados (es decir, una suspensión de perlas).

[0140] Para muestras procariotas, las reacciones contenían Tris-HCl 50 mM (pH 8,3), KCl 75 mM, MgCl₂ 3 mM, DTT 10 mM, cebadores 1,25 mM, 0,5 mM de mezcla de nucleótidos PCR (Amersham Pharmacia), 1 URNasin (Promega) y 5 U SuperScript II (Invitrogen) y se realizaron mediante incubación a 50 °C durante 30 min. Para muestras eucariotas, las reacciones contenían Tris-HCl 50 mM (pH 8,3), KCl 50 mM, MgCl₂ 10 mM, espermina 0,5 mM, DTT 10 mM, cebadores de RT 1,25 mM, mezcla de nucleótidos de PCR 0,5 mM, 1 U de RNasin y 5 U de AMV transcriptasa inversa (Promega) y se puede realizar mediante incubación a 48 °C durante 45 min.

PCR de salidas de selección

[0141] La PCR de punto final se puede realizar para visualizar la amplificación de la construcción de longitud

completa. Una muestra de 5 ml de cada reacción de transcripción inversa se puede amplificar con 2,5 Utaq polimerasa (Roche) en Tris-HCl 20 mM (pH 8,4), KCl 50 mM, MgCl₂ 1 mM, 5 % de DMSO, que contiene mezcla de nucleótidos de PCR 0,25 mM, cebador directo 0,25 mM (T7B o T7KOZ para experimentos procarióticos y eucarióticos, respectivamente) y cebador RT 0,25 mM. El ciclo térmico comprendió 94 °C durante 3 min, a 5 continuación 94 °C durante 30 s, 50 °C durante 30 s y 72 °C durante 1,5 min durante 30 ciclos, con un paso final a 72 °C durante 5 min. Los productos de PCR se visualizaron mediante electroforesis en geles de agarosa teñidos con bromuro de etidio. Los productos de PCR aislados se pueden subclonar en un vector de expresión de pBAD bacteriano para la producción de proteína soluble.

10 Expresión y producción bacteriana:

[0142] Las células huésped de *E. coli* competentes se preparan según las instrucciones del fabricante (sistema de expresión de Invitrogen PBAD). Brevemente, 40 µl de células competentes LMG 194 y 0,5 µl de constructos de pBAD Fn3 (aproximadamente 1 µg de ADN) pueden incubarse juntas en hielo durante 15 minutos, 15 después de lo cual se aplica un choque térmico de 42 °C durante un minuto. Las células a continuación se dejan recuperar durante 10 minutos a 37 °C en medio SOC antes de colocarlas en placas LB-Amp y crecimiento a 37 °C durante la noche. Las colonias individuales se recogen al día siguiente para cultivos líquidos a pequeña escala para determinar inicialmente las concentraciones de inducción de L-arabinosa óptimas para la producción de Fn3. Las réplicas de cada clon después de alcanzar una DO₆₀₀ = 0,5 se pueden probar inducidas con valoraciones en serie 20 (1:10) de L-arabinosa (concentración final del 0,2 % al 0,00002 %) después del crecimiento durante la noche a temperatura ambiente. Se pueden recolectar cultivos de prueba (1 ml), se añadieron 100 µl de tampón 1xBBS (10 mM, NaCl 160 mM, ácido bórico 200 mM, pH = 8,0) para resuspender las células antes de la adición de 50 µl de solución de lisozima durante 1 hora (37 °C). Los sobrenadantes celulares de las digestiones de lisozima se pueden recoger después de la centrifugación y puede añadirse MgSO₄ a una concentración final de 40 mM. Esta solución 25 se puede aplicar a columnas de Ni-NTA preequilibradas con PBS. Las muestras de Fn3 unidas con His se lavan dos veces con tampón PBS sobre el que se puede lograr la elución con la adición de imidazol 250 mM. La pureza de la expresión de Fn3 soluble se puede examinar mediante SDS-PAGE.

Ejemplificación

30 Ejemplo 1: Procedimientos para la identificación guiada por bioinformática de secuencias universales de bibliotecas de dominios de unión al lado inferior de la fibronectina

[0143] En este ejemplo, las secuencias universales de bucle AB, CD y EF para secuencias de bibliotecas de dominio de unión a fibronectina se identifican y seleccionan usando bioinformática y los criterios de la invención. Un 35 esquema generalizado de este proceso se presenta en la Fig. 1.

[0144] Brevemente, se buscó en la base de datos PFAM una alineación múltiple de secuencias que contenía solo secuencias pertenecientes a la familia Fibronectina de Tipo III en todas las especies conocidas, incluyendo mamíferos y seres humanos (Fn3, PFAM ID PF00041) en el formato de alineación de Stockholm 1.0 40 (<http://pfam.janelia.org/family?entry=fn3#tabview=tab2>).

[0145] Esta búsqueda arrojó un conjunto de datos inicial de 15 520 secuencias de proteínas "conjunto de datos de base". Sin embargo, se observa que este conjunto de secuencias puede aumentar en número a medida que se clonan secuencias adicionales y se ingresan en la base de datos. Dentro de las 15 520 secuencias 45 compiladas iniciales, las secuencias variarán con respecto a las especies de origen, los módulos de origen, las longitudes del bucle y otras cualidades. De este modo, se llevó a cabo un refinamiento adicional a partir de las secuencias compiladas de partida de tal manera que los conjuntos de datos de subconjuntos comparten una o más propiedades elegidas de interés, es decir, la presencia de bucles. Por lo tanto, a partir de las 15 520 secuencias descargadas iniciales, se eliminaron las secuencias redundantes o duplicadas, o las secuencias incompletas y se 50 identificaron las que contenían las tres secuencias de bucle, reduciendo el número de secuencias a 12 452 "conjunto de datos convencional".

[0146] El siguiente paso implicaba la designación del marco de andamiaje β, longitudes del bucle y secuencias de aminoácidos de bucle, seguido de un análisis de frecuencia de estas secuencias de bucle candidato. 55 Por lo tanto, la determinación de los perfiles de variabilidad implica la recopilación y selección de secuencias de aminoácidos alineados que comparten una o más propiedades definidas de interés para crear un conjunto de datos. Se diseñó un sistema de clasificación posicional de cadenas β y bucles utilizando la arquitectura de numeración posicional de los dominios Fn3 como punto de referencia. Las cadenas β se identificaron primero en base a la estructura cristalina de todos los módulos de proteína de fibronectina Fn3 disponibles en RCSB Protein Data Bank

(www.rcsb.org, HM Berman, J. Westbrook, Z. Feng, G. Gilliland, TN Bhat, H. Weissig, IN Shindyalov, PE Bourne; "The Protein Data Bank", Nucleic Acid Research, 28 págs. 235-242 (2000)). Una alineación de todas las estructuras cristalinas identificó regiones de cadenas β y regiones potenciales de bucles.

- 5 **[0147]** Se escribió una secuencia de comandos en Python para extraer datos sobre los bucles por alineación y agrupamiento según la longitud del bucle. El experto en la materia apreciará que, para este propósito, puede usarse cualquier metodología que alinee las regiones de bucle e identifique los aminoácidos en cada posición. El análisis identificó las siguientes longitudes del bucle para los bucles del lado inferior:
- 10 El bucle AB contenía de 1 a 3 aminoácidos.
El bucle CD contenía de 4 a 10 aminoácidos.
El bucle EF contenía de 3 a 9 aminoácidos.
Las secuencias aceptables contenían los 3 bucles.
- 15 **[0148]** La secuencia de comandos en Python acepta 12 452 de las 15 520 secuencias descargadas (80 %)
- [0149]** El comando en Python analizó cada ciclo para las secuencias aceptadas en cada posición para cada longitud y catalogó la longitud del bucle, la distribución de aminoácidos y el porcentaje de abundancia de restos de aminoácidos en cada posición de los bucles. Un experto en la materia puede apreciar que otras secuencias génicas relacionadas con Fn1, Fn2 y otros dominios del marco de andamiaje de proteína, tales como SCALPs, o nanoanticuerpos también pueden buscarse en estas bases de datos e identificarse de manera similar.
- 20 **[0150]** Los resultados de estos análisis se muestran en las Figs. 3-10. Por ejemplo, los límites externos del bucle AB se determinó que estaban entre 1-3, con la mayoría de las secuencias que requieren una longitud de bucle de 3 (12, 265, Figs. 3A y B). Después de establecer los límites externos, los datos se segregaron en tamaños de bucle AB y se determinó su composición de frecuencia de aminoácidos (Figs. 4A y B). El mismo principio se aplicó a los bucles CD y EF como se muestra en las Figs. 5-10A-B.
- 25 **[0151]** En base a estas definiciones de límites alineados de la cadena β , no se encontraron bucles de un tamaño, sino que se producen en longitudes diferentes. Se analizó la distribución de frecuencias de los tamaños de los bucles AB (Figs. 3AB), CD (Figs. 5AB) y EF (Figs. 9AB). La clasificación y la descripción del tamaño de los bucles siguen la nomenclatura de: Fn3 BUCLE/LONGITUD. Por ejemplo, AB/1-3 se refiere a la longitud del bucle Fn3 AB de 1 a 3 aminoácidos, y AB/3 se refiere a la longitud del bucle Fn3 AB de 3 aminoácidos. En la definición del bucle AB, los límites del bucle del andamiaje se establecieron originalmente para 1-3 aminoácidos de longitud, AB/1-3. La distribución de las 12 452 secuencias identificó la longitud 3 del bucle como la más predominante en las secuencias, AB/3. Se realizó un examen adicional de la longitud del bucle AB de 3 (Figs. 3-4 AB).
- 30 **[0152]** En la definición del bucle CD, los límites del bucle del andamiaje se establecieron originalmente para una longitud de 4-10 aminoácidos, CD/4-10. La distribución de las 12 452 secuencias identificó las longitudes del bucle 5, 6 y 7 como las más predominantes en las secuencias, CD/5, CD/6 y CD/7. Se realizó un examen adicional de las longitudes del bucle CD de 5, 6 y 7, ya que captura aproximadamente el 80 % de las moléculas de Fn3 del conjunto de datos de 12 452 secuencias (Figs. 5-8AB).
- 35 **[0153]** En la definición del bucle EF, los límites del bucle del andamiaje se establecieron originalmente para una longitud de 3-9 aminoácidos, EF/3-9. La distribución de las 12 452 secuencias identificó que la longitud 6 del bucle fue la más predominante en las secuencias, EF/6. Se realizó un examen adicional de la longitud del bucle 6 (Figs. 9-10AB).

50 Ejemplo 2: Evaluación de perfiles de variabilidad de bucle mediante bioinformática por filtrado y análisis de clúster de secuencias génicas

- [0154]** Las bibliotecas universales de dominio de unión a fibronectina se diseñaron determinando los perfiles de variabilidad para los bucles expresados *in vivo*. Los perfiles de variabilidad representan la catalogación de los diferentes aminoácidos y sus respectivas tasas de aparición, presentes en una posición particular en un conjunto de datos de secuencias alineadas. Se pueden identificar y delinear las familias de secuencias de bucle relacionadas con el tamaño usando los parámetros establecidos anteriormente dentro de este "conjunto de datos base" de inicio. El análisis comparativo de estos bucles alineados múltiples proporciona información de perfil de variabilidad en cuanto a la diversidad existente y "tolerada" para introducir cambios de aminoácidos que pueden conducir a la unión potencial del ligando. La designación de bucles y sus aminoácidos que también comprenden se puede describir para

otras proteínas de tipo andamiaje usando definiciones similares.

[0155] La distribución de frecuencia de los seis tamaños de bucle se generó para determinar si existían tamaños de bucle AB, CD y EF preferidos para secuencias de Fn3. Para el bucle AB, el tamaño de bucle de 3 aminoácidos fue el más predominante y representó el 98,50 % de la población del bucle AB (Fig. 3AB). En este caso, se eligió el tamaño de bucle AB de 3 para un análisis de perfil de variabilidad adicional, AB/3. (Fig. 4AB)

[0156] Para el bucle CD, los tamaños de bucle CD 5, 6 y 7 fueron los tamaños comunes que ocurren en el 15,56 %, 40,59 % y 21,00 % de la población del bucle CD, respectivamente (Fig. 5AB). Los tamaños de bucle CD 5, 6 y 7 fueron elegidos para el análisis del perfil de variabilidad, CD/5 (Fig. 6AB), CD/6 (Fig. 7AB) y CD7 (Fig. 8AB), respectivamente.

[0157] El análisis de tamaño de frecuencia del bucle EF demostró que el tamaño del bucle 6 se producía en casi el 98,83 % de las secuencias de Fn3 analizadas (Fig. 9AB). En este caso, solo se eligió el tamaño de bucle EF 6 para un análisis de perfil de variabilidad adicional, EF/6 (Fig. 10AB).

Tabla 1: Longitudes de los bucles Fn3 AB, CD y EF analizados junto con el número de secuencias indican el número

Bucle	Longitud	Cantidad de secuencias	Porcentaje de secuencias
AB	3	12 265	98,50
CD	5	1937	15,56
CD	6	5054	40,59
CD	7	2615	21,00
EF	6	12 306	98,83

[0158] Para cada uno de los 5 bucles seleccionados (Tabla 1) se realizó un análisis de frecuencia separado para determinar el uso de aminoácidos posicionales en el contexto del bucle seleccionado. La salida se visualiza como gráficos de frecuencia y los resultados se muestran para el bucle AB/3 en las Figs. 4AB; para el bucle CD/5 en las Figs. 6AB, el bucle CD/6 en las Figs. 7AB, el bucle CD/7 en las Figs. 8AB; y para el bucle EF/6 en las Figs. 10AB.

Ejemplo 3: Identificación de posiciones del bucle fijas y no fijas con umbrales

[0159] En una realización, se diseña una biblioteca combinatoria de variantes naturales con un aminoácido de consenso conservado o semiconservado seleccionado de la siguiente manera.

[0160] Los conjuntos de datos de bucle Fn3 se enumeran como anteriormente para la variabilidad de aminoácidos y sus frecuencias relativas en cada posición alineada (Figs. 3-10). El análisis anterior identificó las preferencias posicionales en todos los bucles del módulo Fn3 y se denominan "perfiles de variabilidad". Por ejemplo, en el bucle AB de tamaño 3 (Fig. 4AB), se encuentra una treonina (T) en la posición 1 en aproximadamente el 35 %. No hubo posiciones del bucle donde un solo aminoácido representase más del 35 % de los aminoácidos posicionales representativos. Hubo una ligera preferencia por la serina (S) en la posición 2 con respecto a la alanina (A); y para la treonina (T) en la posición 3 sobre el ácido aspártico (D). Sin embargo, en este caso, no hubo ácidos predominantes por encima de un nivel umbral del 40 % para ser considerados aminoácidos "fijos" y todas las posiciones del bucle se consideran "variables". Para la construcción de la biblioteca, el aminoácido más común en cada posición del bucle sirve como aminoácido de partida para la mutagénesis de primera ronda. Por lo tanto, el bucle AB/tiene una secuencia de bucle de inicio totalmente "variable" de: X1 X2 X3, donde X1 es uno cualquiera de los aminoácidos T, S, N, D, K, Q; X2 es uno cualquiera de los aminoácidos S, A, K, P, D, E, T; y X3 es uno cualquiera de los aminoácidos T, D, S, N, E, R, G.

[0161] Para el bucle EF de 6 aminoácidos, se encuentra una leucina (L) en la posición 2 dentro del bucle en más del 90 % de todas las posiciones del bucle Fn3 que demuestran un alto grado de presión selectiva para su presencia. Por lo tanto, la posición 2 del bucle EF se consideraría "conservada" para la leucina (L), ya que se produjo por encima de un nivel umbral predeterminado del 40 % y fue más del doble de común que el siguiente aminoácido más frecuente en esa posición. Este resto "fijo" se ve como el aminoácido dominante con respecto a los otros aminoácidos que se producen en esa posición del bucle. Una posición "fija" puede no estar sujeta a la diversificación mutagénica en la construcción de la primera ronda de la biblioteca. Además, la prolina (P) en la posición 4 dentro del bucle ocurre a una frecuencia de > 46 % y también puede considerarse "fija". Al determinar primero estas posiciones "fijas" y posiciones no fijas (indicadas por "X"), EF/6 tendría una secuencia inicial "fija" de: G1 L2 X3 P4 G5 X6, donde X3 es uno cualquiera de los aminoácidos K, T, E, Q, L, S e I, y X6 es uno cualquiera de

los aminoácidos T, V, S, K, R, A y H. En bibliotecas posteriores, G1 también se puede alterar para que sea uno cualquiera de los aminoácidos G, N, D, K, E, S, R; L2 también se puede alterar para que sea uno cualquiera de los aminoácidos L, V, I, M, T, F, A; P4 también se puede alterar para que sea cualquiera de los aminoácidos P, E, A, K, S, T, N; y G5 también se puede alterar para que sea uno cualquiera de los aminoácidos G, N, D, A, S, Y, F. El experto en la materia puede apreciar que las bibliotecas se pueden generar usando cualquier combinación de aminoácidos en cualquier posición del bucle EF/6, como se describe en las Figs. 10AB.

[0162] Para el bucle CD/5, la glicina (G) en la posición 3 dentro del bucle se encuentra en > 49 % y se consideraría conservada y "fija". No parece haber ningún aminoácido predominante en las posiciones 1, 2, 4 y 5 del bucle CD/5, por lo tanto, se considerarían como posiciones "variables". Por lo tanto, una secuencia fija de inicio para el bucle CD/5 sería X1, X2, G3, X4, X5, donde X1 es uno cualquiera de los aminoácidos G, T, A, S, E, R, K; X2 es uno cualquiera de los aminoácidos D, G, N, S, E, Q, T; X4 es uno cualquiera de los aminoácidos Q, S, G, V, E, T, K; y X5 es uno cualquiera de los aminoácidos P, W, E, T, S, V, L. G3 también se puede alterar para que sea uno cualquiera de los aminoácidos G, S, K, D, A, T, N.

[0163] Para el bucle CD/6, el triptófano (W) en la posición 6 dentro del bucle se encuentra en > 59 % y se consideraría conservado y "fijo". No parece haber ningún aminoácido predominante en las posiciones 1, 2, 3, 4 y 5 del bucle CD/6, por lo tanto, se considerarían como posiciones "variables". Por lo tanto, una secuencia fija de inicio para el bucle CD/6 sería X1, X2, X3, X4, X5, W6, donde X1 es uno cualquiera de los aminoácidos K, T, A, V, S, L, E; X2 es uno cualquiera de los aminoácidos G, D, S, N, E, T, K; X3 es uno cualquiera de los aminoácidos S, T, G, R, E, D, A; X4 es uno cualquiera de los aminoácidos G, E, D, T, K, N, S; y X5 es uno cualquiera de los aminoácidos E, R, T, K, D, A, Q. W6 también se puede alterar para que sea uno cualquiera de los aminoácidos W, E, P, T, A, Y, S.

[0164] Para el bucle CD/7, no parece haber ningún aminoácido predominante en las posiciones 1, 2, 3, 4, 5, 6 o 7 dentro del bucle, por lo que se considerarían todas como posiciones "variables". Por lo tanto, se usaría la secuencia de bucle de inicio de X1 X2 X3 X4 X5 X6 X7, donde X1 es uno cualquiera de los aminoácidos K, A, E, T, S, V, G; X2 es uno cualquiera de los aminoácidos G, N, D, S, E, T, K; X3 es uno cualquiera de los aminoácidos G, S, E, T, A, K, D; X4 es uno cualquiera de los aminoácidos G, E, P, T, S, D, K; X5 es uno cualquiera de los aminoácidos G, E, S, D, T, R, A; X6 es uno cualquiera de los aminoácidos E, W, P, D, Y, T, A; y X7 es uno cualquiera de los aminoácidos T, W, V, E, A, S, P.

[0165] El perfil de variabilidad para cada conjunto de datos de bucle identifica las características deseadas de una posición del bucle dada para una mayor introducción de la representación de diversidad. Estos resultados anteriores demuestran que la diversidad de aminoácidos de bucle introducida en una biblioteca puede "afinarse" dependiendo del nivel de frecuencia de aparición. Estas posiciones del bucle "fijo" intentan reproducir parte de la diversidad natural para promover posibles efectos de estabilización estructural.

[0166] Esta "fijación" de posiciones también tiene el efecto de "estrechar" la diversidad de posiciones variables en secuencias de bucle de inicio. Sin embargo, puede haber la ocasión de hacer lo contrario, es decir, obtener bibliotecas más grandes y más diversas. En este caso, el efecto de "ensanchamiento" se logra elevando el umbral de frecuencia de aparición utilizado para designar aminoácidos "fijos". De esta forma, el perfil de variabilidad capturará menos de las posiciones del bucle más conservadas y las clasificará como posiciones "fijas". Las posiciones del bucle restantes serían parte de los aminoácidos "variables" más amplios que pueden diversificarse.

[0167] Cada una de las posiciones denotadas como X codificará para los aminoácidos presentes en el perfil de variabilidad relacionado o para un subconjunto de los químicamente equivalentes. De hecho, en algunos casos, dos o más aminoácidos presentes en una determinada posición son químicamente muy similares. En tales situaciones, es posible incluir en el diseño de mutagénesis solo un subconjunto de los aminoácidos y aún preservar las características de la química natural de esa posición. Esto reducirá el número total de mutantes y dará más flexibilidad para la optimización de la síntesis de oligonucleótidos.

Ejemplo 5: Identificación de posiciones del bucle fijas y no fijas sin umbrales

[0168] En otra realización, es posible diseñar diversidad combinatoria de variación natural en cada uno de los 5 bucles seleccionados (Tabla 1) sin definir un umbral de variabilidad, es decir, cuando el umbral seleccionado es del 100 %. En esta realización, cada bucle mutado está diseñado para contener aminoácidos que imitan su perfil de variabilidad en términos de características de variabilidad y química. En cada posición específica del bucle, la síntesis de oligonucleótidos se optimiza para contener un codón degenerado que coincidiría/imitaría la química y la variabilidad en esa posición. Las posiciones que tienen dos o más aminoácidos en sus perfiles de variabilidad se

mutarán independientemente del grado de variabilidad de esa posición.

Ejemplo 6: Procedimientos para diseñar diversidad de bucles para bibliotecas de dominio Fn3 con restricciones de umbral

[0169] En este ejemplo, se presentan procedimientos para optimizar la diversidad de bucles de una biblioteca de dominios de unión a Fn3. La elección de las estructuras candidatas, como se ha indicado anteriormente, dicta los tamaños de bucle a introducir y la selección inicial de la secuencia de aminoácidos. El procedimiento se ilustra particularmente para el bucle EF/6.

[0170] Para diseñar la biblioteca de bucles EF/6 para andamiajes basados en Fn3, las consideraciones de perfil de variabilidad son las siguientes: Como se ha indicado anteriormente, se determina que un resto de aminoácido "fijo" tiene lugar con una frecuencia umbral que normalmente es de al menos un 40 % (normalmente a al menos un 50 %) y es dos veces más frecuente que el siguiente aminoácido más frecuente para una posición dada de bucle. Tras la inspección del perfil de variabilidad de Fn3 EF/6 (Fig. 10AB), se puede ver que la leucina (L) en la posición 2 dentro del bucle EF/6 aparecía a > 92 %. El siguiente aminoácido que se produce con mayor frecuencia fue a una frecuencia inferior al 5 % y no se registró en el valor umbral mínimo. Cuando se determina que un resto dado, en este caso la leucina, se produce a una frecuencia tan alta, está altamente conservado y, por lo tanto, se representa en las bibliotecas de la invención como "fijo", lo que significa que no se mutará en la diversificación de bibliotecas de la primera ronda.

[0171] De manera similar, la glicina (G) en la posición 1, la prolina (P) en la posición 4 y la glicina (G) en la posición 5 dentro del bucle EF/6 están "altamente conservadas" ya que aparecen a una frecuencia del 40 %, > 46 % y > 45 % respectivamente y están "fijas" en la primera biblioteca de diversidad. En algunos casos, puede ser que haya dos restos de aminoácidos semiconservados en una posición del bucle dada. Una posición "semifija" es aquella que tiene una fuerte presión selectiva con un primer aminoácido (por ejemplo, > 60 %) y una menor presión selectiva para un segundo aminoácido en la misma posición (por ejemplo, > 30 %).

[0172] La razón para no crear diversidad en todos los sitios, excepto por el uso de una secuencia de inicio "fija", es restringir el tamaño de diversidad inicial de la biblioteca para facilitar la expresión y presentación eficiente de todas las variantes. Estas posiciones iniciales "fijas" indican fuertes presiones selectivas para su preservación. Sin embargo, todavía son sitios para un mayor refinamiento durante la maduración de afinidad. En otras palabras, las posiciones "fijas" iniciales pueden mutarse posteriormente. El objetivo general en las posiciones de "fijación" en la primera ronda de diversificación de la biblioteca es doble: 1) maximizar el número de clones funcionales incorporando la mayoría de los restos de bucle preferidos y 2) minimizar el tamaño total de la biblioteca.

[0173] El término resto de aminoácido "variable" se refiere a restos de aminoácidos que se determina que se producen con una frecuencia inferior (menor que el valor de umbral alto del 20 %) para una posición de resto dada. Tras la inspección del perfil de variabilidad EF/6 por ejemplo (Fig. 10AB), se puede ver que la posición variable 3 dentro del bucle no tiene un único aminoácido prevalente que se produce a una frecuencia superior al 40 %. Para la posición 3 de EF/6, cada uno de los bucles EF/6 tiene muchos aminoácidos diferentes que se producen a un nivel de frecuencia bastante bajo. En consecuencia, el sitio de la posición 3 se usa para la creación de la diversidad de secuencia de bucle inicial mediante mutagénesis. Todos los 20 aminoácidos y aminoácidos no naturales que utilizan el codón ámbar se pueden usar potencialmente para la mutagénesis. (Véase por ejemplo, el documento US2009/0176654).

Ejemplo 7: Producción de bibliotecas de dominios de unión en el lado inferior de la fibronectina

[0174] En este ejemplo, se describen las etapas para realizar y ensamblar una biblioteca de dominio de unión del lado inferior de la fibronectina universal usando técnicas de ingeniería genética.

[0175] Brevemente, los módulos de fibronectina se clonan usando técnicas convencionales de biología molecular. Los oligonucleótidos que codifican el armazón del andamiaje de la cadena beta y los bucles de diversidad de las regiones variables se pueden ensamblar usando reacciones de reacción en cadena de la polimerasa. Las moléculas de longitud completa a continuación se amplifican usando cebadores 5' y 3' flanqueantes que contienen sitios de restricción que facilitan la clonación en el vector o vectores de expresión-presentación. La diversidad total de las bibliotecas generadas depende del número de secuencias marco y número de posiciones en los bucles elegidos para la mutagénesis.

[0176] Luego se seleccionan los clones aleatorios de cada biblioteca para la verificación de la secuencia y la evaluación de la calidad de la biblioteca con respecto a la diversidad mutacional deseada, las mutaciones puntuales no deseadas, las delecciones y las inserciones. Esta eficacia contrasta con estrategias de mutagénesis aleatoria/estocástica donde la introducción descontrolada de varias bases produce niveles más altos de efectos de cambio de base no deseados que dan lugar a una baja expresión de la funcionalidad de dominios de unión a fibronectina debido al uso desfavorable de aminoácidos y codones de parada inadvertidos.

[0177] Más específicamente, ¹⁰F_n3 se usa para diseñar, aislar y manipular bibliotecas universales de dominio de unión del lado inferior de la fibronectina.

Construcción de la biblioteca

[0178] Se usó la secuencia de tipo silvestre ¹⁰F_n3 como se muestra en la SEQ ID NO: 1 como base para generar bibliotecas de aglutinantes que utilizan los bucles superior o inferior.

VSDVPRDLEVVAATPTSLISWDAPAVTVRYRITYGETGGNSPVQFTVPGSKSTA
TISGLKPGVDYTTITVYAVTGRGDSPASSKPISINYRTEI (SEQ ID NO: 1)

[0179] Usando los procedimientos descritos en los Ejemplos anteriores, la ¹⁰F_n3 de tipo silvestre se mutagenó en los bucles inferiores AB, CD, EF para generar una biblioteca de dominios de unión del lado inferior de la fibronectina (Biblioteca inferior). El mismo procedimiento puede usarse para generar una biblioteca de dominios de unión del lado superior de la fibronectina usando los bucles superiores BC, DE y FG (Biblioteca superior).

[0180] Las secuencias de ADN correspondientes a la biblioteca inferior se optimizan para la expresión en *E. coli* en Geneart AG, Alemania. La biblioteca inferior se ensambla a partir de oligonucleótidos sintéticos degenerados y genes correspondientes a fragmentos de longitud completa y se purifica en gel. La amplificación se realiza con los cebadores terminales y la ligación subsiguiente de la biblioteca amplificada en el vector de clonación pCR-Script que produce las bibliotecas de inicio. La Biblioteca inferior de inicio a continuación se criba para identificar los aglutinantes monoespecíficos como se describe en el Ejemplo 8 a continuación.

Ejemplo 8: Cribado de moléculas de unión monoespecíficas basadas en fibronectina

[0181] El presente ejemplo describe cómo seleccionar aglutinantes monoespecíficos de fibronectina generados a partir de la Biblioteca inferior descrita en el Ejemplo 7. La Biblioteca inferior se subclona en un vector de presentación en levadura tal como pYD1 (Invitrogen) usando procedimientos de recombinación homóloga y se transforma en una cepa adecuada tal como EBY100 utilizando técnicas convencionales de biología molecular.

[0182] La presentación y selección de aglutinantes basados en fibronectina contra una diana, por ejemplo, lisozima de huevo de gallina, se realizó siguiendo esencialmente el protocolo previamente publicado por Lipovsek, D. et al. (J Mol Biol. 11 de mayo de 2007; 368 (4): 1024-41) con algunas modificaciones menores

(i) Selección para la unión a la lisozima de clara de huevo de gallina utilizando clasificación de perlas magnéticas

[0183] Para todas las selecciones, los cultivos de levadura que presentan la Biblioteca inferior de moléculas basadas en ¹⁰F_n3 se inducen durante 18 h a 30 °C en medio que contiene galactosa (90 % de SG-CAA/10 % de SD-CAA, 50 µg/ml de kanamicina, 100 U/ml de penicilina G, 200 U/ml de estreptomycin). Las 10⁹ células de levadura inducidas de la Biblioteca inferior se lavan con 25 ml de solución salina tamponada con fosfato (PBS), pH 7,4, ácido etilendiaminotetraacético (EDTA) 2,2 mM, albúmina de suero bovino al 0,5 % (BSA) y a continuación se incuban en 5 ml del mismo tampón que contiene 1 µM de lisozima de clara de huevo de gallina biotinilada (HEL-b, Sigma, St. Louis, MO) durante 1 h a temperatura ambiente con rotación suave. Después de la incubación, la muestra se enfría en hielo, se lava con 25 ml de PBS helado, pH 7,4, EDTA 2 mM, BSA al 0,5 % y se resuspende en 2,5 ml del mismo tampón. Se añade una alícuota de 100 µl de perlas magnéticas Streptavidin MicroBeads (Milenyi Biotec, Auburn, CA) a la levadura y se incuba en hielo durante 10 min. Se añade PBS enfriado con hielo, pH 7,4, EDTA 2 mM, 0,5 % de BSA a la muestra hasta un volumen total de 25 ml inmediatamente antes de someterlo a separación en un separador de células AutoMACS (Milenyi Biotec), usando el programa preestablecido para la selección positiva de células raras (possas). Las células seleccionadas se recogen en 6 ml de SD-CAA, pH 4,5, 50 µg/ml de kanamicina, 100 U/ml de penicilina G, 200 U/ml de estreptomycin; se cuantifica por dilución en serie seguido de siembra en placas de agar SD-CAA; y se crece en 50 ml del mismo medio durante 2 días a 30 °C.

(ii) *Selección para la unión a lisozima de clara de huevo de gallina usando clasificación celular activada por fluorescencia*

- 5 **[0184]** Las siguientes rondas de selección se realizan por FACS, comenzando con 2×10^6 a 3×10^6 células de levadura inducidas. Las células se lavan con 1 ml de PBS, pH 7,4, BSA al 0,1 %, se resuspenden en 100 µl del mismo tampón que contiene lisozima de clara de huevo de gallina biotinilada, y se incuban a temperatura ambiente con rotación suave durante 1 h.
- 10 **[0185]** Después de lavar con 1 ml de PBS enfriado con hielo, pH 7,4, BSA al 0,1 %, las células se marcan con anticuerpos y estreptavidina. El anticuerpo monoclonal anti-c-myc conjugado con FITC (AbD Serotec) se usa para marcar la levadura para la presentación en la superficie de miméticos del anticuerpo marcado con c-myc, y la estreptavidina marcada con PE (Invitrogen) o anticuerpo anti-biotina (Miltényi) se utiliza para marcar HEL-b asociado con miméticos del anticuerpo de unión a la lisozima. Las clasificaciones FACS se realizan en células de levadura
- 15 marcadas con anticuerpo anti-c-myc de ratón conjugado con FITC y estreptavidina conjugada con PE (Invitrogen).
- [0186]** Las células de levadura doblemente marcadas se clasifican en un clasificador de células Dako MoFlo de alta velocidad con un láser de 488 nm, a 6000-10 000 células/s. Las puertas se ajustan para recoger las células de levadura con el 0,1-1 % más alto de la señal asociada a HEL-b (PE) y en la mitad superior de la señal asociada a la expresión (FITC). Las muestras duplicadas marcadas con el mismo anticuerpo y los mismos reactivos de estreptavidina, pero en ausencia de HEL-b se usan para evitar seleccionar las células que unen los reactivos de detección en lugar de la lisozima.
- 20 **[0187]** Para todas las bibliotecas, las primeras dos clasificaciones FACS se realizan en levadura marcada con HEL-b 1 µM. Una vez que se observa una población de células que está marcada con PE en presencia, pero no en ausencia de HEL-b, la concentración de HEL-b en la ronda siguiente disminuye en un orden de magnitud. Las células seleccionadas se recogen en 0,5 ml de SD-CAA, pH 4,5, 50 µg/ml de kanamicina, 100 U/ml de penicilina G y 200 U/ml de estreptomycin. Las células recolectadas se cultivan hasta la saturación en 5 ml del mismo medio, con agitación, durante 2 días a 30 °C, antes de inducirse y marcarse para la siguiente ronda de clasificación.
- 30 **[0188]** Después de varias rondas de clasificación de FACS, la población enriquecida final se pone en placas SDCAA y se incuba a 30 °C durante 2 días. Las colonias individuales se recogen usando un Genetix Clonepix y se vuelven a formar en placas de 96 pocillos que contienen medio SD-CAA. Después de la incubación durante 24 horas, las células se recogen por centrifugación y se resuspenden en medio SD-GAA para la inducción de moléculas
- 35 de Fn3 únicas expresadas en superficie. Los clones positivos se identifican mediante ELISA convencional. El ADN del plásmido correspondiente a los clones positivos de Fn3 únicos se purifica y se secuencia para identificar aglutinantes mono-específicos.
- [0189]** Una vez identificados y seleccionados los aglutinantes mono-específicos de la Biblioteca inferior,
- 40 pueden usarse para generar moléculas terapéuticas. En particular, los aglutinantes mono-específicos generados a partir de la Biblioteca inferior se pueden usar para generar nuevas moléculas de unión terapéutica contra una diana de interés. Diversos aglutinantes mono-específicos de la Biblioteca inferior se pueden combinar con enlazadores para producir una molécula de unión a Fn3 que es capaz de unirse a una o más regiones de una única diana (por ejemplo, TNF). Como alternativa, también puede diseñarse una molécula de unión a Fn3 que comprende
- 45 aglutinantes de la Biblioteca inferior para que se unan a una o más regiones de dianas múltiples (por ejemplo, una o más regiones de HSA y TNF).
- [0190]** Además, los aglutinantes mono-específicos generados a partir de la Biblioteca inferior se pueden combinar con la Biblioteca superior usando técnicas convencionales de biología molecular para generar aglutinantes
- 50 bis-específicos y multi-específicos.

Ejemplo 9: Generación de moléculas de unión basadas en fibronectina bifuncional

- [0191]** El modelado por ordenador de las regiones aleatorizadas en la estructura de rayos X de Fn3 humana
- 55 muestra que combinando aglutinantes mono-específicos de cada una de las bibliotecas A y B, se pueden crear moléculas de unión a fibronectina bis-específicas. Estas moléculas de unión pueden manipularse de manera que reconozcan regiones diferentes en la misma molécula diana, o que los diferentes sitios de unión de la molécula de fibronectina bis-específica o multi-específica puedan unirse a diferentes regiones en dos o más dianas diferentes.

[0192] Por ejemplo, una secuencia adecuada correspondiente a los Bucles inferiores (obtenida al seleccionar la Biblioteca Inferior) y una secuencia adecuada correspondiente a los Bucles superiores (obtenida al seleccionar la Biblioteca superior) se pueden combinar en una sola molécula, Aglutinante C que usa ambos Bucles inferiores y superiores para unir, generando así una molécula biespecífica.

5

[0193] El ADN correspondiente a la secuencia de aminoácidos combinada de Aglutinante C se sintetiza y optimiza para la expresión en *E. coli* en Geneart AG, Alemania. La clonación en *E. coli* y la posterior purificación sigue protocolos convencionales como se describe en el capítulo 4 (Procedimientos de fabricación).

10 **[0194]** Como alternativa, se puede generar una molécula de unión basada en Fn3 bifuncional uniendo dos o más moléculas de unión basadas en Fn3 monoespecíficas.

Equivalentes

15 **[0195]** Los expertos en la materia reconocerán o serán capaces de determinar, usando nada más que experimentación de rutina, muchos equivalentes de las realizaciones específicas de la invención descrita en el presente documento.

REIVINDICACIONES

1. Un procedimiento para formar una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles para explorar la presencia de uno o más polipéptidos que tienen una actividad enzimática o de unión
5 seleccionada, donde dichos polipéptidos tienen las secuencias de aminoácidos de tipo silvestre en las regiones de cadena beta A, AB, B, C, CD, D, E, EF, F y G del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, el procedimiento que comprende las etapas de:
 - (i) alinear secuencias de bucle de aminoácidos AB, CD y EF en una colección de polipéptidos del dominio de
10 fibronectina de tipo 3 nativo,
 - (ii) segregar las secuencias de bucle alineadas según la longitud del bucle,
 - (iii) para un bucle y una longitud bucle seleccionados de la etapa (ii), realizar un análisis de frecuencia posicional de aminoácidos para determinar las frecuencias de aminoácidos en cada posición del bucle, en el que la longitud del bucle EF está fijada en 6 aminoácidos (EF/6),
 - 15 (iv) para cada bucle y longitud de bucle analizados en la etapa (iii), identificar en cada posición un aminoácido de consenso conservado o semiconservado seleccionado y otros aminoácidos de variante natural,
 - (v) para al menos el bucle EF, formar:
 - (1) una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que
20 codifican, en cada posición del bucle, el aminoácido de consenso, y si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido, o
 - (2) una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de
25 secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, incluidos aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos,
- 30 (vi) incorporar la biblioteca de secuencias codificantes en las secuencias codificantes de Fn3 marco para formar una biblioteca de expresión de Fn3, y
- (vii) expresar los polipéptidos de Fn3 de la biblioteca de expresión.
2. El procedimiento de la reivindicación 1, en el que la frecuencia umbral dada es del 100 %.
3. El procedimiento de la reivindicación 1, en el que la frecuencia umbral dada es una frecuencia
35 seleccionada entre el 50-95 %.
4. El procedimiento de la reivindicación 1, en el que la etapa (ii) segrega los bucles y las longitudes de
40 bucle en el grupo que consiste en AB/1-3, y CD/4-10.
5. El procedimiento de la reivindicación 1, en el que la etapa (ii) segrega los bucles y las longitudes de bucle en el grupo que consiste en AB/3, CD/5, CD/6 y CD/7.
- 45 6. El procedimiento de la reivindicación 1, en el que la biblioteca formada tiene una biblioteca de secuencias combinatorias naturales en una combinación de bucles y longitudes de bucle seleccionadas de los bucles AB y EF, y CD y EF, donde el bucle AB se selecciona de uno de AB/1-3 y AB/3, y el bucle CD se selecciona de uno de CD/4-10, CD/5, CD/6 y CD/7, o una combinación de los tres bucles AB, CD y EF, donde los bucles AB y CD se seleccionan del grupo que consiste en AB/1-3, AB/3, CD/4-10, CD/5, CD/6 y CD/7.
- 50 7. El procedimiento de la reivindicación 6, en el que el umbral dado es del 100 %, a menos que la posición del aminoácido del bucle contenga solo un aminoácido dominante y uno variante, y los aminoácidos dominantes y variantes tengan cadenas laterales con propiedades fisicoquímicas similares, en cuyo caso el umbral dado es del 90 %.
- 55 8. Una biblioteca de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad de unión o enzimática seleccionada, dichos polipéptidos que comprenden:

(a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, y

(b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas, en las que la longitud del bucle EF está fijada en 6 aminoácidos (EF/6), donde al menos la región de bucle EF contiene una biblioteca de secuencias de mutagénesis expresadas por una biblioteca de secuencias codificantes que codifican, en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, un único aminoácido diana común y cualquier aminoácido coproducido.

9. La biblioteca de la reivindicación 8, en la que la frecuencia umbral dada es del 100 %.

10. La biblioteca de la reivindicación 8, en la que la frecuencia umbral dada es una frecuencia seleccionada entre el 50-95 %.

11. La biblioteca de la reivindicación 8, en la que los bucles y las longitudes de bucle se seleccionan del grupo que consiste en bucles AB y EF, y CD y EF, donde el bucle AB se selecciona de uno de AB/1-3 y AB/3, y el bucle CD se selecciona de uno de CD/4-10, CD/5, CD/6 y CD/7, o una combinación de los tres bucles AB, CD y EF, donde los bucles AB y CD se seleccionan del grupo compuesto por AB/1-3, AB/3, CD/4-10, CD/5, CD/6 y CD/7, y que tiene una biblioteca de secuencias de mutagénesis formadas en cada uno de los bucles y longitudes de bucle seleccionadas del grupo compuesto por AB/1-3, AB/3, CD/4-10, CD/5, CD/6 y CD/7.

12. Una biblioteca combinatoria de variantes naturales de polipéptidos del dominio de fibronectina de tipo 3 (Fn3) útiles en el cribado de la presencia de uno o más polipéptidos que tienen una actividad de unión o enzimática seleccionada, dichos polipéptidos que comprenden:

(a) regiones A, AB, B, C, CD, D, E, EF, F y G que tienen secuencias de aminoácidos de tipo silvestre del 10º módulo de fibronectina de tipo III de la fibronectina humana que tiene la secuencia de SEQ ID NO: 1, y

(b) regiones de bucle AB, CD y EF que tienen longitudes seleccionadas, en las que la longitud del bucle EF está fijada en 6 aminoácidos (EF/6), donde al menos la región de bucle EF contiene una biblioteca de secuencias combinatorias de variantes naturales expresadas por una biblioteca de secuencias codificantes que codifican en cada posición del bucle, un aminoácido de consenso conservado o semiconservado seleccionado y, si el aminoácido de consenso tiene una frecuencia de aparición igual o inferior a una frecuencia umbral seleccionada de al menos un 50 %, otros aminoácidos de variantes naturales, que incluyen aminoácidos semiconservados y aminoácidos variables cuya frecuencia de aparición está por encima de un umbral mínimo seleccionado en esa posición, o sus equivalentes químicos.

13. La biblioteca de la reivindicación 12, que tiene una biblioteca de secuencias combinatorias de variantes naturales en una combinación de bucles y longitudes de bucle seleccionadas de bucles AB y EF, y CD y EF, donde el bucle AB se selecciona de uno de AB/1-3 y AB/3, y el bucle CD se selecciona de uno de CD/4-10, CD/5, CD/6 y CD/7, o una combinación de los tres bucles AB, CD y EF, donde los bucles AB y CD se seleccionan del grupo que consiste en AB/1-3, AB/3, CD/4-10, CD/5, CD/6, CD/7.

14. La biblioteca de la reivindicación 13, en la que el umbral dado es del 100 %, a menos que la posición del aminoácido del bucle contenga solo un aminoácido dominante y uno variante, y los aminoácidos dominantes y variantes tengan cadenas laterales con propiedades fisicoquímicas similares, en cuyo caso el umbral dado es del 90 %.

15. La biblioteca de la reivindicación 13, en la que la longitud del bucle AB es 3.

16. La biblioteca de la reivindicación 13, en la que la longitud del bucle CD es 5.

17. La biblioteca de la reivindicación 13, en la que la longitud del bucle CD es 6.

18. La biblioteca de la reivindicación 13, en la que la longitud del bucle CD es 7.

19. La biblioteca de la reivindicación 13, en la que los polipéptidos están codificados por una biblioteca de expresión seleccionada del grupo que consiste en una biblioteca de presentación en ribosomas, una biblioteca de presentación en polisomas, una biblioteca de presentación en fagos, una biblioteca de expresión bacteriana y una biblioteca de presentación en levaduras.

20. Una biblioteca de expresión de polinucleótidos que codifica la biblioteca de polipéptidos de la reivindicación 13, y producida sintetizando polinucleótidos que codifican una o más regiones marco de la cadena beta y una o más regiones de bucle en las que los polinucleótidos están predeterminados, en la que los polinucleótidos que codifican dichas regiones comprenden adicionalmente suficiente secuencia de superposición mediante la cual las secuencias de polinucleótidos, en condiciones de reacción en cadena de la polimerasa (PCR), son capaces de ensamblarse en polinucleótidos que codifican dominios de unión a fibronectina completos.
21. Un procedimiento para identificar un polipéptido que tiene una afinidad de unión deseada con respecto a un antígeno seleccionado, que comprende hacer reaccionar la biblioteca combinatoria variante natural de polipéptidos Fn3 de la reivindicación 13 con el antígeno seleccionado, y cribar los polipéptidos Fn3 para seleccionar los que tienen una afinidad de unión deseada con respecto al antígeno seleccionado.
22. El procedimiento de la reivindicación 21, en el que el procedimiento comprende adicionalmente la etapa de identificar el polinucleótido que codifica el dominio de unión a fibronectina seleccionado.

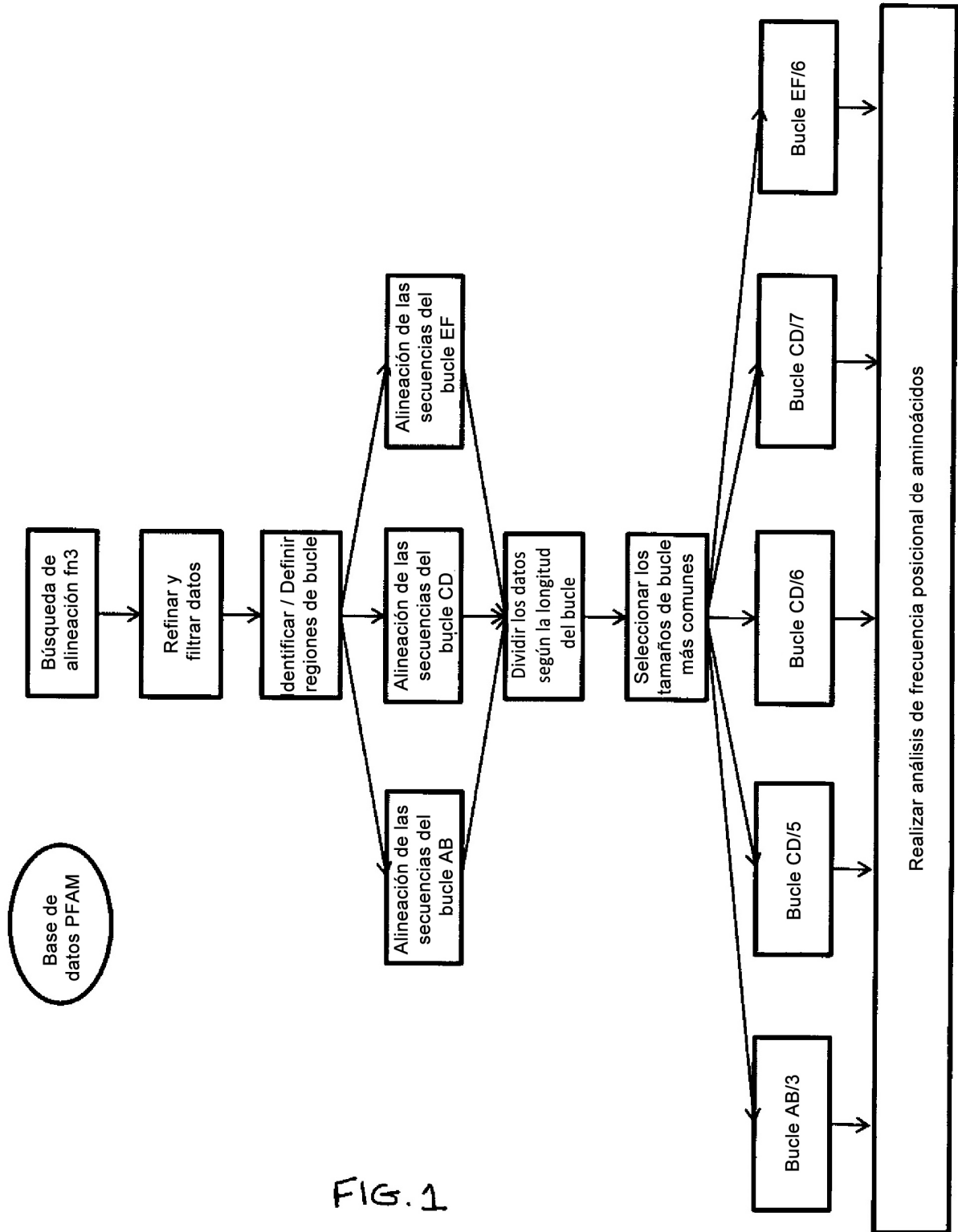


FIG. 1

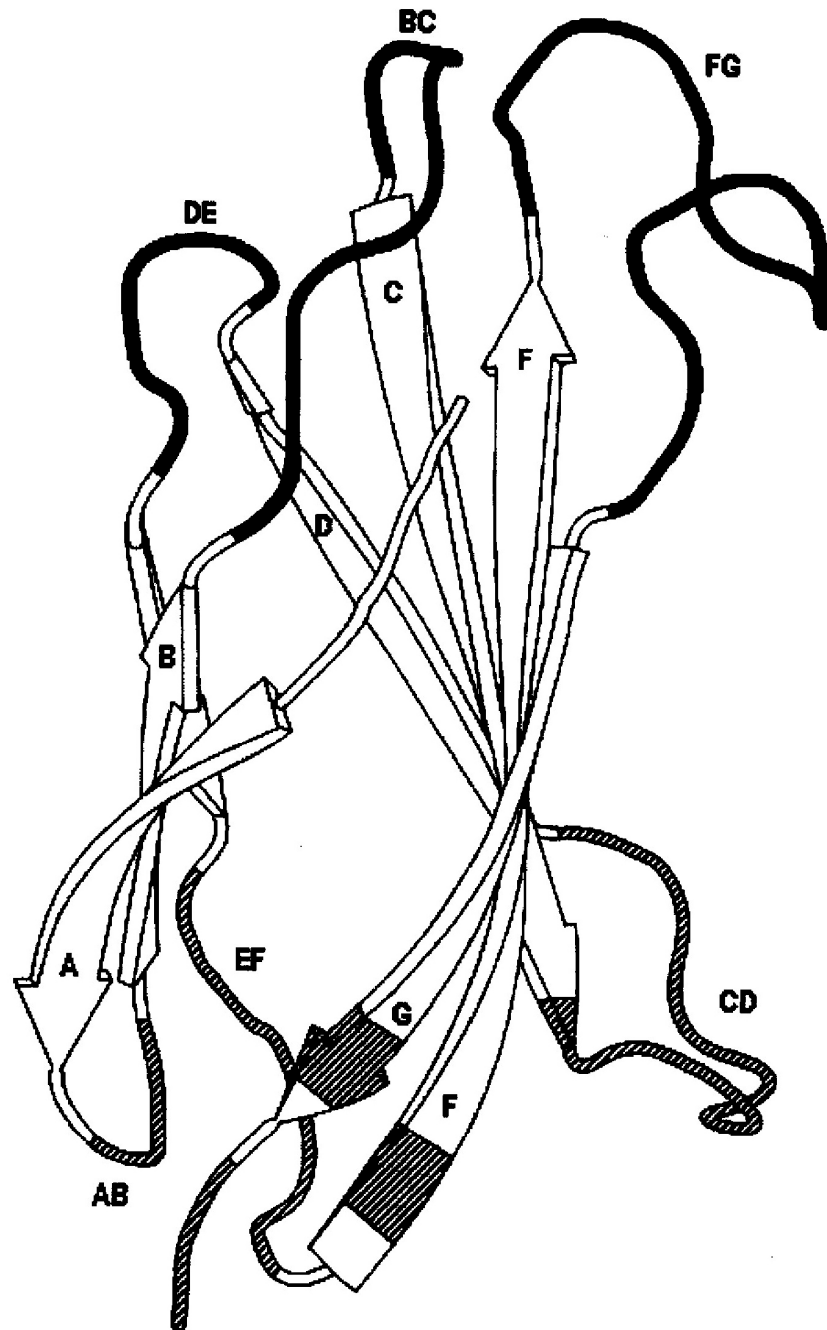


Fig. 2

Bucle AB

Distribución de la longitud del bucle

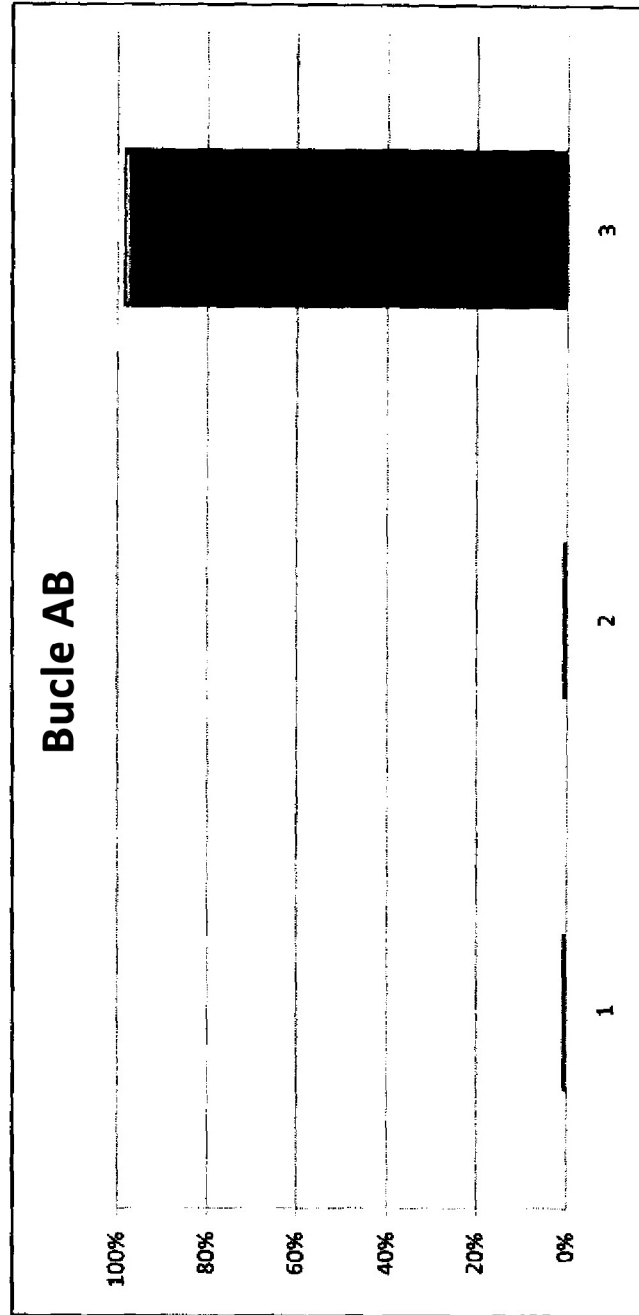


Fig. 3A

Bucle AB

Distribución de la longitud del bucle

Longitud del bucle	Número de secuencias	Porcentaje de secuencias
1	90	0,72
2	97	0,78
3	12.265	98,50

Fig. 3B

Bucle AB

*Distribución posicional de longitud de bucle 3
(12.265 secuencias)*

Longitud 3 del bucle AB		Abundancia de los 7 aminoácidos principales (%)						
Posición	Aminoácidos	1º	2º	3º	4º	5º	6º	7º
1	TSGNDKQ	35,4	21,6	10,9	6,1	5,8	2,9	2,9
2	SAKPDET	16,6	12,6	12,2	11,4	8,7	8,1	5,7
3	TDSNERG	23,5	18,1	14,4	14,2	6,5	5,9	3,4

Fig. 4A

Bucle AB

*Distribución posicional de longitud de bucle 3
(12.265 secuencias)*

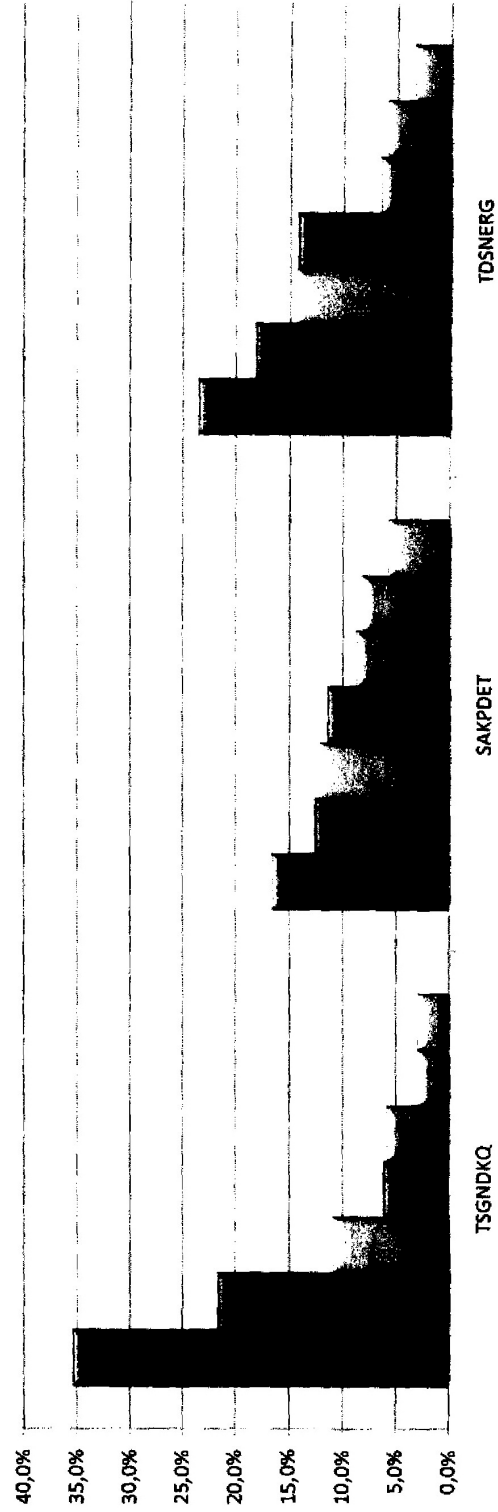


Fig. 4B

Bucle CD

Distribución de la longitud del bucle

Bucle CD

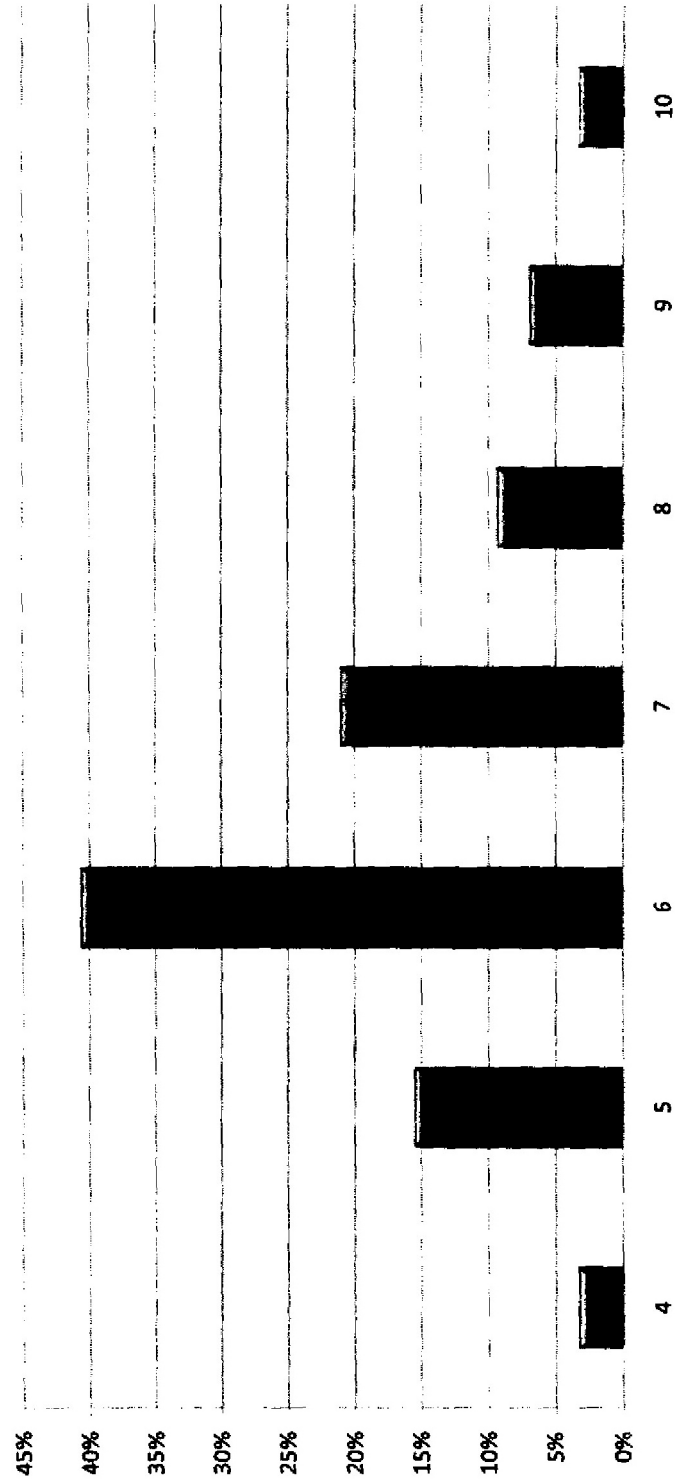


Fig. 5A

Bucle CD

Distribución de la longitud del bucle

Longitud del bucle	Número de secuencias	Porcentaje de secuencias
4	404	3,24
5	1.937	15,56
6	5.054	40,59
7	2.615	21,00
8	1.161	9,32
9	866	6,95
10	415	3,33

Fig. 5B

Bucle CD

*Distribución posicional de longitud de bucle 5
(1937 secuencias)*

		Abundancia de los 7 aminoácidos principales (%)						
Posición	Aminoácidos	1º	2º	3º	4º	5º	6º	7º
1	GTASERK	13,8	12,8	10,3	8,5	7,6	7,6	7,2
2	DGNSEQT	25,7	18,9	12,7	9,3	7,3	5,5	4,5
3	GSKDATN	49,4	7,7	5,8	5,6	4,6	4,4	4,3
4	QSGVETK	15,5	11,0	10,5	7,7	7,1	6,7	6,1
5	PWETSVL	17,3	10,7	10,7	7,8	7,3	7,2	5,9

Fig. 6A

Bucle CD

Distribución posicional de longitud de bucle 5

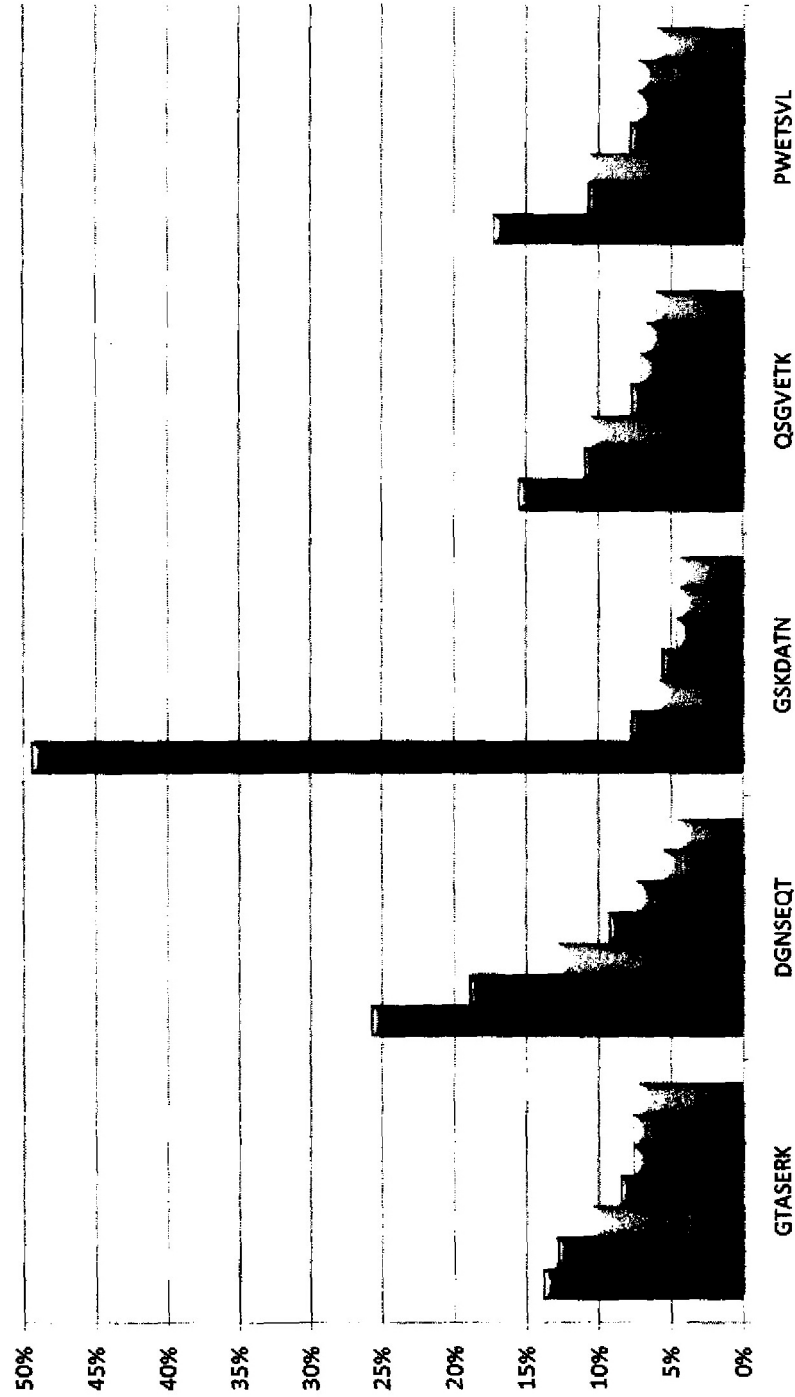


Fig. 6B

Bucle CD

*Distribución posicional de longitud de bucle 6
(5.054 secuencias)*

Bucle CD		Abundancia de los 7 aminoácidos principales (%)						
Posición	Aminoácidos	1º	2º	3º	4º	5º	6º	7º
1	KTAVSLE	17,2	12,0	10,2	9,9	7,6	6,7	6,3
2	GDSNETK	23,1	12,4	11,3	9,7	7,9	7,1	5,7
3	STGREDA	20,5	14,8	13,7	10,2	7,4	7,0	5,4
4	GEDTKNS	17,8	10,5	9,0	7,7	7,1	6,8	6,5
5	ERTKDAQ	15,5	10,5	10,3	8,0	7,8	6,3	5,8
6	WEPTAYS	59,1	5,6	4,2	3,7	3,0	2,9	2,5

Fig. 7A

Bucle CD

Distribución posicional de longitud de bucle 6

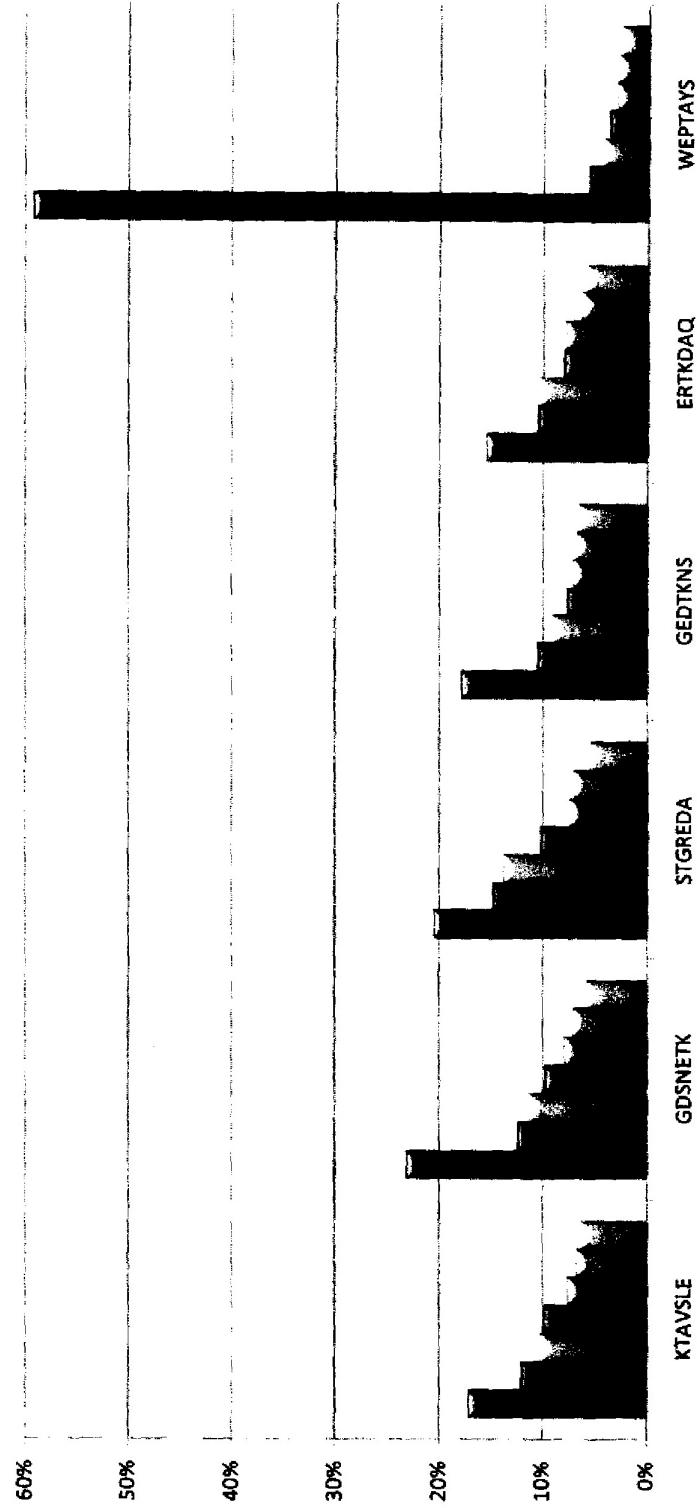


Fig. 7B

Bucle CD

*Distribución posicional de longitud de bucle 7
(2615 secuencias)*

COL		Abundancia de los 7 aminoácidos principales (%)						
Posición	Aminoácidos	1º	2º	3º	4º	5º	6º	7º
1	KAETSVG	18,5	13,6	11,9	9,8	7,9	7,9	4,7
2	GNDSETK	20,1	15,5	13,3	10,6	7,9	7,4	5,1
3	GSETAKD	19,1	14,6	9,7	9,0	7,5	6,7	6,5
4	GEPTSDK	13,7	13,3	11,2	9,7	7,1	6,7	5,9
5	GESDTRA	16,6	13,4	10,3	8,6	6,7	6,5	6,3
6	EWPDYTA	17,4	13,1	11,3	7,5	7,2	6,0	5,9
7	TWVEASP	13,6	13,3	10,5	8,6	7,1	7,0	5,7

Fig. 8A

Bucle CD

Distribución posicional de longitud de bucle 7

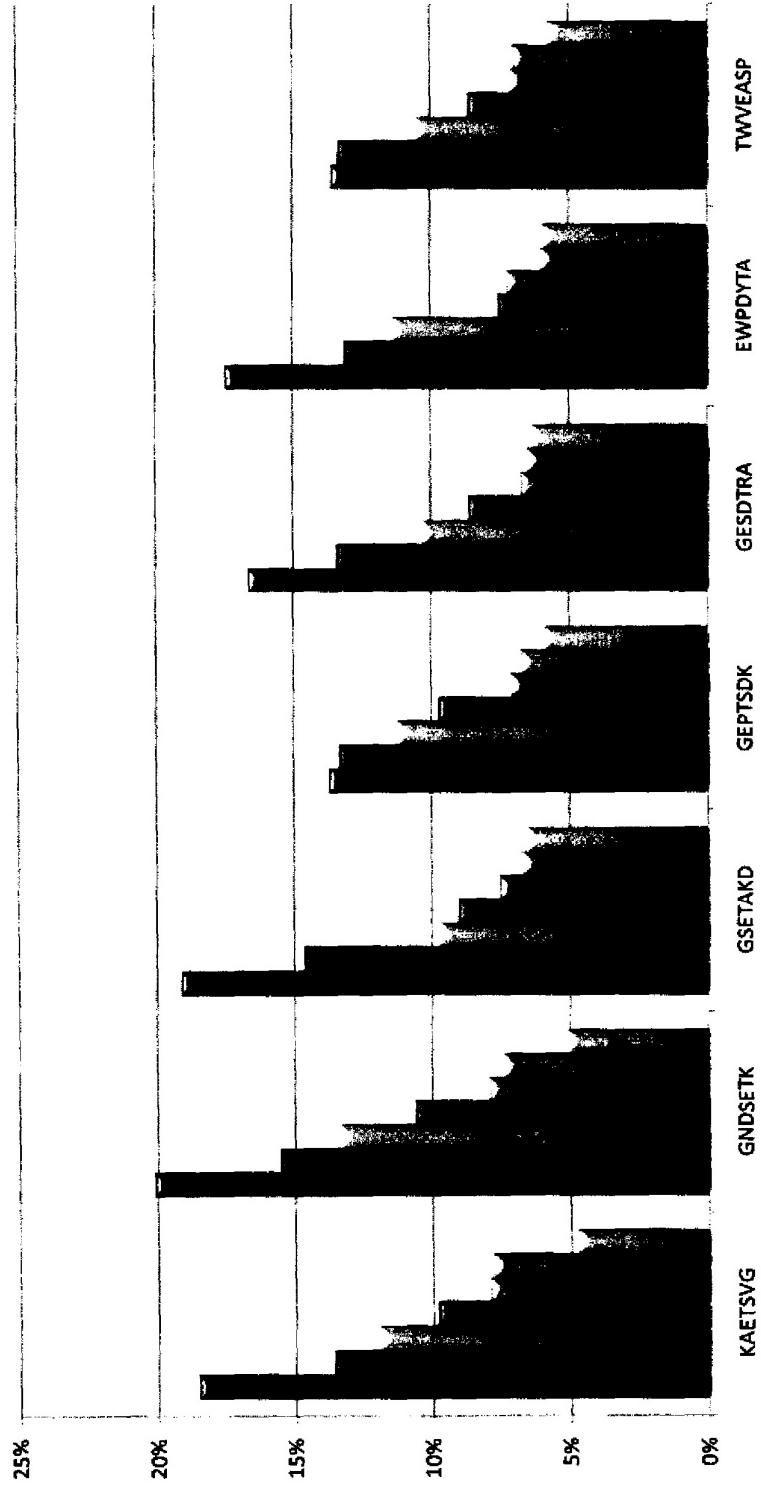


Fig. 8B

Bucle EF

Distribución de la longitud del bucle

Longitud del bucle	Número de secuencias	Porcentaje de secuencias
3	4	0,03
4	3	0,02
5	83	0,67
6	12.306	98,83
7	33	0,27
8	22	0,18
9	1	0,01

Fig. 9A

Bucle EF

Distribución de la longitud del bucle

Bucle EF

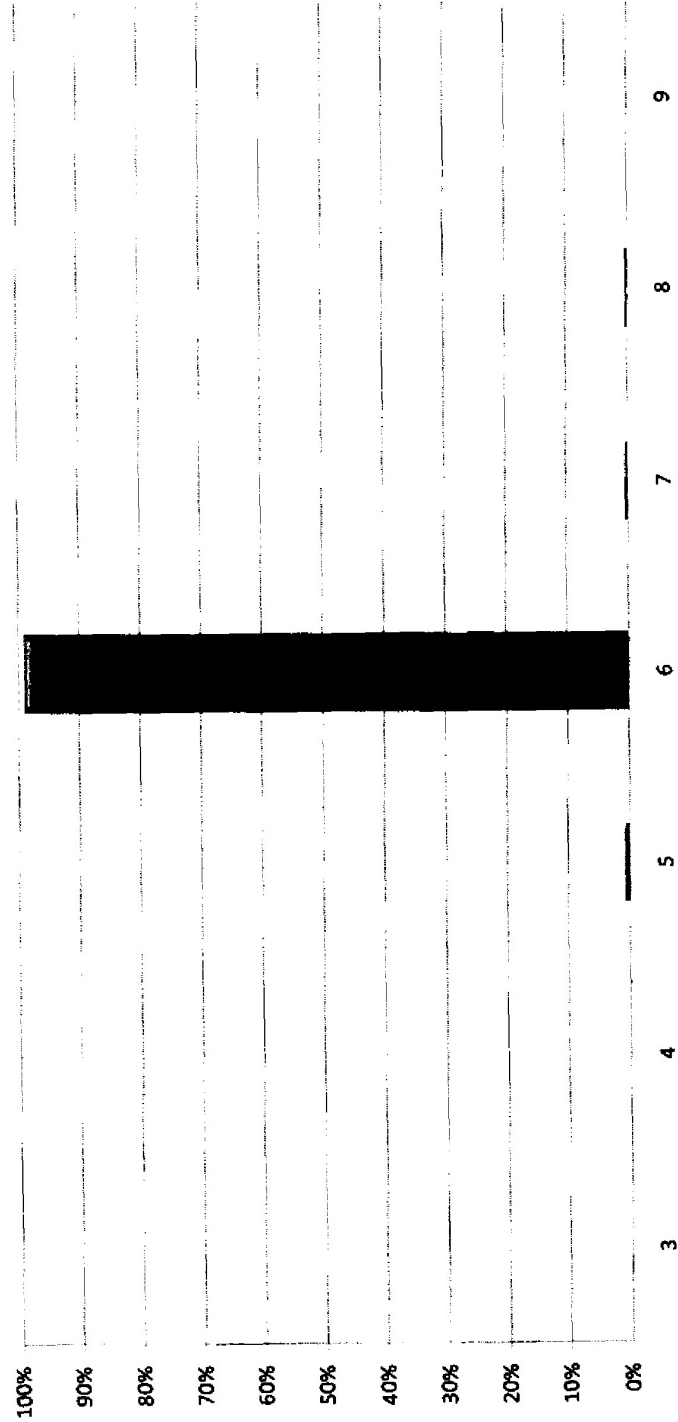


Fig. 9B

Bucle EF

*Distribución posicional de longitud de bucle 6
(12.306 secuencias)*

		Abundancia de los 7 aminoácidos principales (%)						
Posición	Aminoácidos	1º	2º	3º	4º	5º	6º	7º
1	GNDKESR	40,0	18,0	14,3	6,8	4,7	4,6	3,2
2	LVIMTFA	92,1	2,5	1,7	1,5	0,7	0,5	0,3
3	KTEQLSI	14,0	13,5	12,2	8,1	6,9	6,8	6,8
4	PEAKSTN	46,2	17,5	6,8	6,0	4,3	3,7	2,4
5	GNDASYF	45,5	13,1	8,6	5,8	5,5	4,9	3,2
6	TVSKRAH	37,9	8,6	7,7	7,4	6,2	5,8	4,5

Fig. 10A

Bucle EF

Distribución posicional de longitud de bucle 6

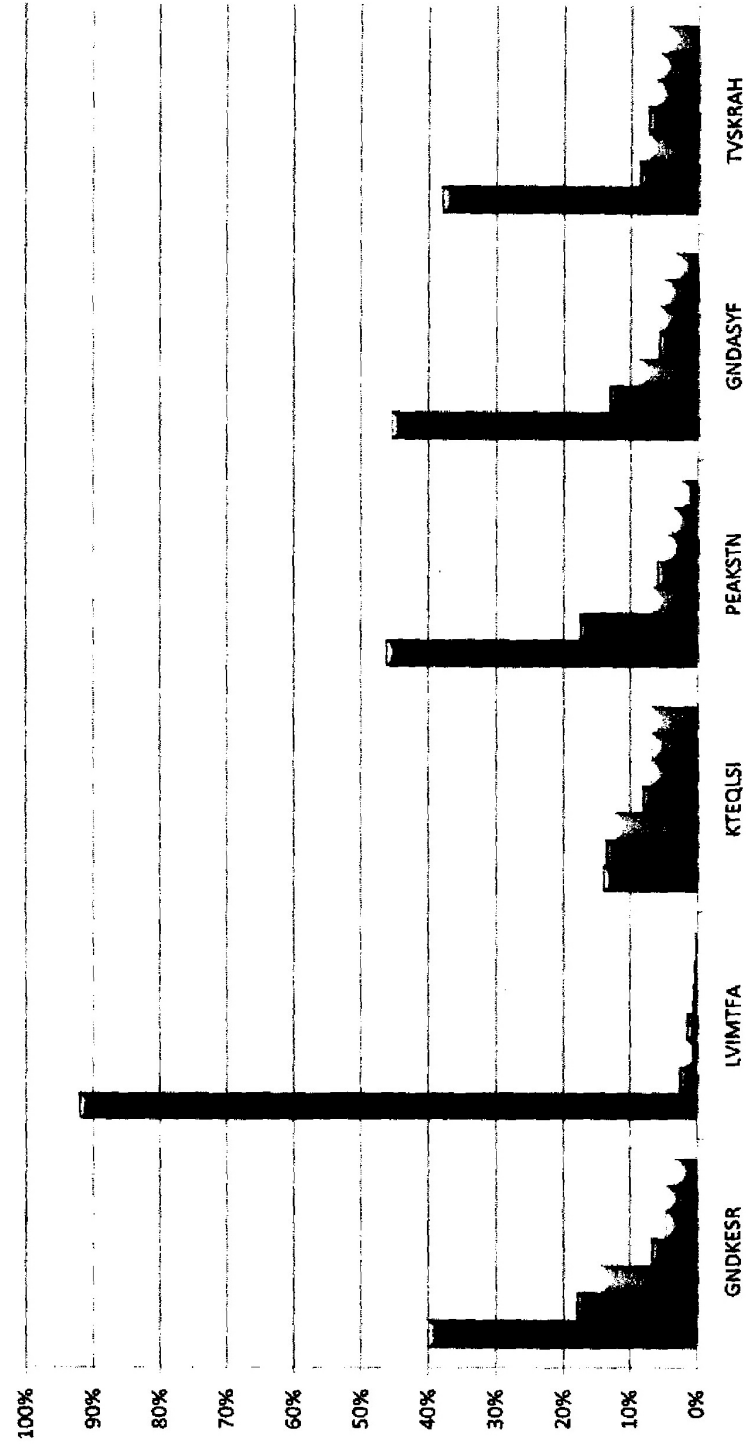


Fig. 10B