

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 926 548**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/62 (2006.01)

G01N 33/68 (2006.01)

C07K 16/12 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.10.2017 PCT/EP2017/077816**

87 Fecha y número de publicación internacional: **03.05.2018 WO18078167**

96 Fecha de presentación y número de la solicitud europea: **30.10.2017 E 17797583 (6)**

97 Fecha y número de publicación de la concesión europea: **29.06.2022 EP 3532612**

54 Título: **Colección de etiquetas y métodos para la detección de proteínas, preferentemente por espectrometría de masas**

30 Prioridad:

31.10.2016 EP 16196571

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.10.2022

73 Titular/es:

**UNIVERSITÄT ZÜRICH (100.0%)
Prorektorat VNW, Rämistrasse 71
8006 Zürich, CH**

72 Inventor/es:

**SEGER, MARKUS;
EGLOFF, PASCAL y
ZIMMERMANN, IWAN**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 926 548 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

Colección de etiquetas y métodos para la detección de proteínas, preferentemente por espectrometría de masas

5 La presente invención se relaciona con un método de sujeción de etiquetas de detección a un banco de proteínas y para uso posterior de las etiquetas para identificar y cuantificar proteínas que cumplen criterios biofísicos o farmacológicos.

Antecedentes de la invención

10 Los métodos de detección de proteínas y de presentación de proteínas son métodos del estado de la técnica para identificar o enriquecer proteínas que exhiben ciertas características (por ejemplo, alta afinidad de enlace a una molécula diana).

15 En las detecciones, las proteínas son analizadas una por una. Esto es muy laborioso y limitado a un número comparativamente bajo de pruebas. En una detección para enlazar proteínas, por ejemplo, los candidatos enlazadores individual se identifican mediante ELISA, y las secuencias similares de ELISA se caracterizan adicionalmente, por ejemplo, se caracterizan biofísicamente mediante cromatografía de exclusión de tamaño, experimentos de desplegado y su potencial terapéutico se prueban en modelos animales *in vivo*.

20 En los métodos de presentación, los grupos completos de proteínas (procedentes de bancos) son enriquecidos durante muchas rondas de selección. El procesamiento de grupos permite un enorme rendimiento sin mucho trabajo. Sin embargo, los métodos de presentación, tal como presentación de fagos, de ribosomas o de levaduras, requieren de una unión física entre el fenotipo (la proteína) y el genotipo (sus ácidos nucleicos codificantes). Esta es una severa limitación para la mayoría de los análisis, debido a que las entidades físicas para realizar la presentación (es decir, el fago, el ribosoma y el ADN o el ARN codificante) son usualmente 100 veces más grandes que la molécula de enlace real (por ejemplo, un fragmento de anticuerpo). Esto da lugar inevitablemente al sesgo en la selección y restringe las presiones de selección posibles a un subconjunto pequeño de presiones de selección imaginables - solo las presiones de selección que no son afectadas críticamente por el enorme tamaño de la partícula de presentación se puede aplicar actualmente (es decir, enlazarse).

30 Los documentos WO-A-2009/036157, WO-A-2007/134327, Kimple *et al.* Current Protocols in Protein Science 2013, 73, Unit 9.9, WO-A-97/07132, WO-A-2004/011676 y WO-A-00/31115 divulgan etiquetas y métodos basados en espectroscopia de masas usando etiquetas.

35 Con base en el estado de la técnica anteriormente mencionado, el objetivo de la presente invención es proveer el medio y los métodos para la identificación de proteínas individuales que cumplen criterios biofísicos o farmacológicos definidos de bancos de proteínas completas en ausencia de una unión física genotipo-fenotipo. Este objetivo se logra por las reivindicaciones de la presente memoria descriptiva.

40 Términos y definiciones

45 La persona experta es consciente de que, dentro de la presente memoria descriptiva, un número señalando el tamaño de un banco se relaciona con la diversidad de miembros del banco. Un banco I que es más grande que un banco corresponde a un banco I que comprende un número mayor de miembros únicos del banco que del banco II. Un banco de ácidos nucleicos con 100.000 miembros puede comprender muchos millones de moléculas de ácidos nucleicos, pero solo 100.000 miembros distintos del banco caracterizado cada uno por una secuencia de ácidos nucleicos dentro de dicho banco. De modo parecido, un banco con 1.000 miembros puede comprender millones de moléculas de polipéptido, pero solo 1.000 miembros únicos de polipéptidos del banco. La expresión "un miembro de un banco" se relaciona con un miembro particular del banco que puede estar presente en una pluralidad de copias idénticas.

50 Dentro del contexto de la presente memoria descriptiva, la expresión "dos secuencias de ácidos nucleicos están en marco" significa que el número de pares de bases entre el último codón de la primera secuencia de ácidos nucleicos y el primer codón de la segunda secuencia de ácidos nucleicos es divisible por tres.

55 Dentro del contexto de la presente memoria descriptiva, las expresiones "el polipéptido está asociado con la etiqueta de detección", respectivamente "etiqueta de polipéptido/detección está asociada con la etiqueta de afinidad" significa que ambos de los miembros anteriormente mencionados están comprendidos dentro de una secuencia primaria de aminoácidos, es decir, una cadena de polipéptidos continua. En particular, dicha etiqueta de detección y dicho polipéptido pueden estar separados por uno o más aminoácidos). Dicha etiqueta de detección y dicha etiqueta de afinidad pueden estar separadas también por uno o más aminoácidos.

60 Dentro del contexto de la presente memoria descriptiva, el término "elemento separable" se relaciona con una secuencia de péptidos susceptible de ser separada mediante agentes químicos o medios enzimáticos, por ejemplo, por proteasas. Las proteasas pueden ser específicas de secuencia (por ejemplo, trombina) o tener especificidad limitada a secuencia (por ejemplo, tripsina). Los elementos separables I y II pueden estar comprendidos dentro de la secuencia de aminoácidos de la etiqueta de detección o el polipéptido, particularmente en casos donde el último aminoácido de la etiqueta de

detección o el polipéptido es un K o R.

Dentro del contexto de la presente memoria descriptiva, el término "etiqueta de afinidad" se relaciona con un fragmento de molécula sujetado a un polipéptido para permitir la purificación de dicho polipéptido de una mezcla bioquímica. La purificación (purificación por afinidad) está basada en una interacción altamente específica (con una constante de disociación de $\leq 10^{-5}$) entre la etiqueta de afinidad y un compañero de enlace de la etiqueta de afinidad. Las etiquetas de afinidad consisten en una secuencia de aminoácidos, o pueden comprender una secuencia de aminoácidos al que se sujeta un fragmento de molécula mediante modificación postraduccional. A manera de ejemplo no limitante, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP (CBP [por sus siglas en inglés]: proteína de enlace a calmodulina), una etiqueta-CYD (CYD [por sus siglas en inglés]: péptido NorpD covalente pero disociable), una etiqueta-Strep, una etiqueta-StrepII, una etiqueta-FLAG, una etiqueta-HPC (HPC [por sus siglas en inglés]: proteína C de cadena pesada), una etiqueta-GST (GST [por sus siglas en inglés]: glutatión S transferasa), una etiqueta-Avi, a etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP (MBP [por sus siglas en inglés]: proteína de enlace a maltosa). Ejemplos adicionales de etiquetas de afinidad se pueden encontrar en Kimple *et al.*, *Curr Protoc Protein Sci.* 2013 Sep 24;73: Unit 9.9.

Dentro del contexto de la presente memoria descriptiva, el término "secuenciación profunda" se relaciona con secuenciación en paralelo de muchos miles de moléculas de ácidos nucleicos con una cobertura de $\geq 5x$, en particular $\geq 40x$. El término "cobertura" se relaciona con el número de veces que un nucleótido dado es leído durante el proceso de secuenciación profunda en promedio.

En el contexto de la presente memoria descriptiva, el término anticuerpo es usado en su significado conocido en la técnica de la biología celular y de la inmunología. Un anticuerpo completo es una glucoproteína que comprende al menos dos cadenas pesadas (H) y dos cadenas ligeras (L) interconectadas por enlaces bisulfuro. Cada cadena pesada está comprendida de una región variable de cadena pesada (VH [por sus siglas en inglés]) y una región constante de cadena pesada (CH [por sus siglas en inglés]). Cada cadena ligera está comprendida de una región variable de cadena ligera (abreviado en este documento como VL [por sus siglas en inglés]) y una región constante de cadena ligera (CL [por sus siglas en inglés]). Las regiones variables de las cadenas pesadas y ligeras contienen un dominio de enlace que interactúa con un antígeno. Las regiones constantes de los anticuerpos pueden mediar el enlace de la inmunoglobulina a tejidos o factores hospederos, incluyendo diversas células del sistema inmune (por ejemplo, células efectoras) y el primer componente del sistema complementario clásico.

Dentro del contexto de la presente memoria descriptiva, el término "nanocuerpo" se relaciona con un "anticuerpo de dominio único", es decir, un fragmento de anticuerpo que consiste en un único anticuerpo de dominio variable. Un nanocuerpo es capaz de unirse selectivamente a un antígeno específico. Tiene un peso molecular de solo 12-15 kDa (Harmsen *et al.*, *Appl. Microbiol. Biotechnol.* 77 (1): 13-22). Usualmente, los nanocuerpos se obtienen mediante inmunización de dromedarios, camellos, llamas, alpacas o tiburones y el aislamiento posterior del ARNm codificante para anticuerpos de cadena pesada. Los nanocuerpos se pueden derivar también de IgG con cuatro cadenas de murino o de humano.

Dentro del contexto de la presente memoria descriptiva, el término "sicuerpo" se relaciona con un nanocuerpo sintético. Los sicuerpos no se obtienen a través de inmunización con un antígeno, sino que se seleccionan *in vitro* de un banco sintético.

Dentro del contexto de la presente memoria descriptiva, el término "enriquecimiento" se relaciona con un proceso de aumentar la cantidad relativa de un cierto compuesto dentro de una mezcla de compuestos.

Dentro del contexto de la presente memoria descriptiva, el término "banco Flycode" se relaciona con un banco de secuencias de aminoácidos de acuerdo con la invención, que comprende una pluralidad de variantes de secuencia.

Dentro del contexto de la presente memoria descriptiva, el término "NestLink" se relaciona con un método en que las etiquetas de detección se sujetan a un banco de proteínas. Posteriormente, las etiquetas son usadas para identificar y cuantificar proteínas individuales que cumplen criterios definidos biofísica o farmacológicamente dentro del banco. NestLink combina beneficios clave de procedimientos de detección y presentación.

Dentro del contexto de la presente memoria descriptiva, el término "valor de hidrofobicidad" se relaciona con un valor predicho que caracteriza a un péptido. El valor de hidrofobicidad está calculado por el método que se describe en Krokhin *et al.*, *Mol Cell Proteomics.* 2004 Sep;3(9):908-19, de acuerdo con las fórmulas:

$$H = K_L * (\sum R_C + 0,42R_{cNt}^1 + 0,22R_{cNt}^2 + 0,05R_{cNt}^3)$$

si $H < 38$ y

$$H = K_L * (\sum R_C + 0,42R_{cNt}^1 + 0,22R_{cNt}^2 + 0,05R_{cNt}^3) - 0,3(K_L * (\sum R_C + 0,42R_{cNt}^1 + 0,22R_{cNt}^2 + 0,05R_{cNt}^3) - 38)$$

si $H \geq 38$;

si $H < 38$, $H_{final} = H$;

si $H \geq 38$, $H_{final} = H - 0,3*(H - 38)$;

donde H_{final} es el valor de hidrofobicidad y R_C son los coeficientes de retención características para tipos de aminoácidos de acuerdo con la tabla siguiente:

	R_C
Trp	11
Phe	10,5
Leu	9,6
Ile	8,4
Met	5,8
Val	5
Tyr	4
Ala	0,8
Thr	0,4
Pro	0,2
Glu	0
Asp	-0,5
Cys	-0,8
Ser	-0,8
Gln	-0,9
Gly	-0,9
Asn	-1,2
Arg	-1,3
His	-1,3
Lys	-1,9

5

R_{CNT} de un aminoácido X se define como:

$$R_{CNT}^X = (\sum R_C / 20) - R_C^X$$

10

N corresponde al número residuo de la etiqueta de detección que inicia con 1 del N-terminal. K_L se define como:

si $N < 10$, $K_L = 1 - 0,027*(10 - N)$

si $N > 20$, $K_L = 1 - 0,014*(N - 20)$

de otra manera $K_L = 1$.

15

Las secuencias de aminoácidos están dadas del amino al carboxilo terminal. Las letras mayúsculas para posiciones de secuencia se refieren a aminoácidos-L en el código de una letra (Stryer, Biochemistry, 3rd ed. p. 21).

Descripción detallada de la invención

20

El ámbito de protección se define por la materia de las reivindicaciones que se anexan.

Método para seleccionar un polipéptido de un banco de polipéptidos

25

De acuerdo con un primer aspecto, se provee de un método para seleccionar un polipéptido de un banco de polipéptidos de acuerdo con la reivindicación 1. El método comprende las siguientes etapas:

a. Se provee un primer banco de ácidos nucleicos. Cada miembro del primer banco de ácidos nucleicos comprende una secuencia codificante de polipéptidos que codifica un miembro de un primer banco de polipéptidos. Cada miembro del primer banco de ácidos nucleicos es diferente de cualquier otro miembro del primer banco de ácidos nucleicos.

30

b. Se provee un segundo banco de ácidos nucleicos. El segundo banco comprende una pluralidad de miembros. Cada miembro comprende una secuencia codificante de etiqueta que codifica una etiqueta de detección. Cada etiqueta de detección tiene las siguientes características:

35

i. La etiqueta está caracterizada por una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por el segundo banco de ácidos nucleicos.

ii. La etiqueta está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 900 y 2200 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 903 y 2180 Da.

40

iii. La etiqueta comprende un primer elemento separable.

La especificación de masa dada en ii se relaciona con la masa de la etiqueta después de que ha sido aislada, es decir, después de separar el primer elemento separable.

5 c. La secuencia codificante de polipéptido comprendida en el miembro del primer banco de ácidos nucleicos se inserta en un miembro de dicho segundo banco de ácidos nucleicos. De ese modo, se crea un banco de ácidos nucleicos etiquetados que codifica un banco de polipéptidos etiquetados. Cada miembro del banco de polipéptidos etiquetados comprende un polipéptido y una etiqueta de detección. La etiqueta de detección se separa del polipéptido por el primer elemento separable.

10 El banco de polipéptidos etiquetados es un "banco anidado", debido a que las secuencias codificantes de polipéptido del primer banco de ácidos nucleicos están "anidadas" dentro de los miembros del segundo banco de ácidos nucleicos. El segundo banco de ácidos nucleicos es muchas veces más grande que el banco de ácidos nucleicos etiquetados. El banco de ácidos nucleicos etiquetados es muchas veces más grande que el primer banco de ácidos nucleicos.

15 Dentro del banco de ácidos nucleicos, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con una secuencia codificante de etiqueta del segundo banco de ácidos nucleicos. La asociación tiene lugar en el marco. La secuencia codificante de polipéptido se inserta en una posición donde se someterá a transcripción y a una traducción posterior en un hospedero adecuado después de que se introduce un miembro del banco de ácidos nucleicos en un hospedero adecuado. La introducción en células bacterianas se puede lograr por transformación. La introducción en células no bacterianas se puede lograr por transfección. La persona experta está consciente de que no se requiere necesariamente un hospedero para la traducción: se pueden emplear también las técnicas de traducción *in vitro*. Para revisiones sobre los sistemas de expresión sin células, véase Rosenblum, FEBS Lett. 2014 Jan21; 588(2):261-8 y Zemella, Chembiochem. 2015 Nov; 16(17):2420-31. La secuencia codificante de polipéptido y la secuencia codificante de etiqueta se transcribirá dentro de la misma secuencia expresada.

25 Dentro del banco de ácidos nucleicos, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con una secuencia codificante de etiqueta del segundo banco de ácidos nucleicos. Cada miembro del banco de ácidos nucleicos etiquetados comprende solo una secuencia codificante de polipéptido y una secuencia codificante de etiqueta. Cada secuencia codificante de etiqueta está comprendida en solo un miembro del banco de ácidos nucleicos etiquetados. En otras palabras, cada secuencia codificante de etiqueta es única dentro del banco de ácidos nucleicos etiquetados. Sin embargo, cada secuencia codificante de polipéptido puede estar comprendida en muchos miembros del banco de ácidos nucleicos etiquetados (etiquetado redundante). En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con al menos una secuencia codificante de etiqueta del segundo banco de ácidos nucleicos. En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con al menos dos secuencias codificantes de etiqueta del segundo banco de ácidos nucleicos. En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con al menos cinco secuencias codificantes de etiqueta diferentes del segundo banco de ácidos nucleicos. En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con al menos diez secuencias codificantes de etiqueta diferentes del segundo banco de ácidos nucleicos. En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia en promedio con 10-30 secuencias codificantes de etiqueta diferentes del segundo banco de ácidos nucleicos. En ciertas realizaciones, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia en promedio con aproximadamente veinte secuencias codificantes de etiqueta diferentes del segundo banco de ácidos nucleicos.

40 d. Una pluralidad de secuencias de ácidos nucleicos se obtiene del banco de ácidos nucleicos etiquetados. En particular, una secuencia de ácidos nucleicos se obtiene para cada miembro del banco de ácidos nucleicos etiquetados. Cada una de dicha pluralidad de secuencias de ácidos nucleicos comprende una secuencia codificante de polipéptido y una secuencia codificante de etiqueta.

50 Con base en la información de secuenciación obtenida en la etapa d, se crea una base de datos. La base de datos comprende las secuencias de todos los polipéptidos y todas las etiquetas de detección en el banco de polipéptidos etiquetados. La persona experta está consciente de que la base de datos no comprende cada miembro único del banco de ácidos nucleicos etiquetados, debido a razones técnicas. Las secuencias pueden estar en forma de secuencias de ácidos nucleicos y/o de secuencias de aminoácidos. La base de datos comprende la información de que está comprendido el subconjunto de las secuencias codificantes de etiqueta del banco de ácidos nucleicos en el banco de ácidos nucleicos etiquetados. La base de datos comprende también la información a que la secuencia codificante de etiqueta, o respectivamente, a que las secuencias codificantes de etiqueta están asociadas con una secuencia codificante de polipéptido.

60 e. Un patrón de fragmentación de espectrometría de masas se predice para cada etiqueta de detección codificante por una secuencia codificante de etiqueta obtenida en la etapa d. La persona experta está consciente de que el patrón de fragmentación se predice para la etiqueta de detección *aislada*, que es para una etiqueta de detección que ha sido liberada de su polipéptido asociado por separación del primer elemento separable. La persona experta está consciente de que predecir el patrón de fragmentación también comprende predecir la masa total de la etiqueta de detección aislada.

65 f. El banco de polipéptidos etiquetados se expresa del banco de ácidos nucleicos etiquetados. Como consecuencia del enfoque de etiquetado redundante descrito en la etapa c, el banco de polipéptidos etiquetados puede comprender un miembro dado de dicho primer banco de polipéptidos etiquetados con muchas etiquetas de detección diferentes (pero

solo una etiqueta por molécula). El etiquetado redundante se prefiere, debido a que facilita la detección sin ambigüedad de un miembro del primer banco de polipéptidos a través de las etiquetas de detección múltiple y minimiza las influencias potenciales de las etiquetas de detección en las propiedades biofísicas de los miembros del banco de polipéptidos etiquetados. La redundancia es una adición requerida por razones técnicas: algunas etiquetas de detección pueden no ser detectadas debido a que reducen los niveles de expresión, se pierden durante la preparación de muestras o no se eluyen dentro de la ventana de hidrofobicidad de la columna de fase inversa, que se analiza por espectrometría de masas.

g. Un miembro del banco de polipéptidos etiquetados se selecciona en una etapa de selección, produciendo un polipéptido seleccionado. Esta etapa de selección comprende aislar aquellos miembros del banco de polipéptidos etiquetados que cumplen con los criterios bioquímicos definidos. En otras palabras, se aplica una presión de selección al banco de polipéptidos etiquetados. Esta presión de selección debe llevar a una separación física de proteínas, de forma que se generen y se recolectan subgrupos separados físicamente. Una ventaja clave del método de acuerdo con la invención es que el intervalo de criterios de selección posible es mucho más alto que en los métodos de presentación de proteínas. A manera de ejemplo no limitante, los criterios se pueden seleccionar del grupo de criterios que comprenden la capacidad de unirse a una molécula diana con una afinidad definida, la estabilidad de un polipéptido en condiciones definidas, un cierto comportamiento de agregación (por ejemplo, la aparición predominante como un monómero) en condiciones definidas, la resistencia a las proteasas, la capacidad de penetración en los tejidos, la eliminación rápida o lenta del flujo sanguíneo, la capacidad de penetrar la barrera hematoencefálica y la capacidad de acumularse en los tumores.

h. El primer elemento separable se separa. De ese modo, la etiqueta de detección se separa del polipéptido seleccionado y se produce una etiqueta de detección aislada.

i. La etiqueta de detección aislada se identifica y cuantifica de la siguiente manera:

i. El patrón de fragmentación de la etiqueta de detección aislada se registra por espectrometría de masas. El patrón de fragmentación provee de información acerca de la masa y de la hidrofobicidad de los fragmentos de la etiqueta de detección aislada. El patrón de fragmentación produce información acerca de la secuencia de aminoácidos de la etiqueta de detección aislada.

ii. El patrón de masa y de fragmentación obtenido en la etapa i se hace coincidir con los patrones de masa y de fragmentación predichos en la etapa e. De ese modo, se identifica la etiqueta de detección aislada. La combinación de la información obtenida por la espectrometría de masas con la información obtenida por secuenciación del banco de ácidos nucleicos etiquetados permite la identificación sin ambigüedad de una etiqueta de detección dada.

La precisión de coincidencia de los patrones de fragmentación predichos y registrados se puede puntuar y permite la clasificación de los miembros del banco de polipéptidos. La comparación de clasificaciones de polipéptidos entre diferentes condiciones de selección se puede usar como una medida relativa de diversas características de polipéptidos (por ejemplo, constante de disociación, distribución de tejido, enlace de conformación específica, etc.). La comparación es la más exacta para miembros del banco de polipéptidos etiquetados de forma redundante, donde se promedia las diferencias en las eficiencias del registro del patrón de fragmentación de etiquetas individuales.

La puntuación de la precisión de coincidencia de los patrones de fragmentación registrados se puede usar como una medición de cantidades relativas de los miembros del banco de polipéptidos después de la selección. Las cantidades relativas son las más exactas para miembros del banco de polipéptidos etiquetados de forma redundante, donde se promedia las diferencias en las eficiencias del registro del patrón de fragmentación de etiquetas individuales.

j. La secuencia de ácidos nucleicos comprende la secuencia codificante de etiqueta que codifica la etiqueta de detección identificada en la etapa i se selecciona de la pluralidad de secuencias de ácidos nucleicos obtenidas en la etapa d. De ese modo, se identifica el miembro del banco de polipéptido etiquetado asociado con la etiqueta de detección identificada en la etapa i.

La persona experta está consciente de que las etapas g a j se realizan por un número de miembros diferentes de dicho banco de polipéptidos etiquetados en paralelo. Un grupo de muchos polipéptidos que demuestran los criterios definidos se selecciona en la etapa g, y todos estos polipéptidos se identifican a través del análisis de espectrometría de masas de sus etiquetas de detección. La persona experta está consciente de que, debido a razones técnicas, no se puede identificar cada polipéptido en esta etapa.

El análisis de espectrometría de masas realizado en la etapa i es cuantitativo, de este modo, el método de acuerdo con la invención permite no solo identificar un polipéptido sino también cuantificar la cantidad de este polipéptido en una muestra.

Para asegurar el etiquetado redundante y único, es importante que

i) el primer banco tenga un tamaño limitado y definido. En ciertas realizaciones, el primer banco de ácidos nucleicos tiene un tamaño de 5 a 100.000. En ciertas realizaciones, el primer banco de ácidos nucleicos tiene un tamaño de 100 a 50.000. En ciertas realizaciones, el primer banco de ácidos nucleicos tiene un tamaño de 500 a 5.000.

ii) el segundo banco de ácidos nucleicos tiene un tamaño de 10^3 a 10^{11} , particularmente 10^5 a 10^{10} , más particularmente 10^6 a 10^9 , incluso más particularmente aproximadamente 10^8 antes de la etapa de inserción del primer banco
 iii) después de la etapa de inserción, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido/etiqueta es de al menos 3x, particularmente al menos 5x, más particularmente al menos 15x, incluso más particularmente al menos 253 el número de miembros de dicho primer banco de ácidos nucleicos.
 iv) el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido/etiqueta es menor que 50 %, particularmente menor que 5 %, más particularmente menor que 0,5 %, incluso más particularmente menor que 0,05 % del número de miembros de dicho segundo banco de ácidos nucleicos.

El tamaño del banco se puede controlar por una etapa de restricción de diversidad anterior a la etapa a, en que se selecciona el primer banco como un subconjunto de un banco anterior más grande.

El método de acuerdo con la invención permite el análisis de bancos de proteínas en ausencia de la unión genotipo-fenotipo requerido para métodos de presentación de proteínas. Esto elimina la desventaja de tener grandes entidades físicas (por ejemplo, un fago o un ribosoma y la codificación de ADN o de ARN) sujetado a los miembros del banco de proteínas. Los bancos de proteínas enteros se pueden detectar como un grupo para criterios de selección, en vez de probar proteínas individuales como usualmente es el caso para detecciones de proteínas. Sin embargo, incluso a pesar de que se procesen grupos de proteínas enteros, la lectura es similar a las detecciones una vez que cada proteína se caracteriza individualmente. Esto es de particular importancia en el campo del desarrollo de proteínas de enlace (fármacos, herramientas de diagnóstico, de investigación etc.). Se puede analizar una gama de características proteicas en miles de candidatos a la vez. Una pregunta ejemplar sería: ¿Cuáles candidatos enlazadores son estables, solubles y monoméricos?

El método de acuerdo con la invención permite abordar la pregunta pertinente justo al inicio de la cadena de las proteínas terapéuticas: "¿Cuál enlazador tiene el potencial terapéutico más grande *in vivo*?" Las preguntas en relación con el potencial terapéutico son: ¿Cuál enlazador sobrevive las condiciones agrestes en el intestino en la administración oral? ¿Cuál enlazador cruza la barrera hematoencefálica? ¿Cuál enlazador muestra las propiedades adecuadas de eliminación renal a partir de la sangre? ¿Cuál enlazador, de entre miles, muestra una buena penetración en los tejidos en el tejido pertinente?

La etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -27 y 128. En ciertas realizaciones, la etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -1 y 70. Los valores de hidrofobicidad se relacionan con la masa de la etiqueta de detección después de que ha sido aislada, es decir, después de separar el primer elemento separable. El valor de hidrofobicidad no incluye la etiqueta de afinidad asociada.

En ciertas realizaciones, el miembro del banco de polipéptidos etiquetados se asocia con una etiqueta de afinidad. Tal etiqueta de afinidad puede simplificar la purificación del miembro seleccionado del banco de polipéptidos etiquetados y/o de la etiqueta de detección misma antes de la espectrometría de masas. La etiqueta de afinidad y el miembro del banco de polipéptidos etiquetados están comprendidos dentro de una secuencia de aminoácidos primaria. Cada miembro del banco de polipéptidos etiquetados comprende un polipéptido y una etiqueta de detección. La etiqueta de afinidad puede estar asociada con el polipéptido o con la etiqueta de detección.

En ciertas realizaciones, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-StrepII, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP.

En ciertas realizaciones, la etiqueta de detección está asociada con una etiqueta de afinidad. En estos casos, la etiqueta de afinidad está situada en el C-terminal de la etiqueta de detección. Esta disposición tiene la ventaja adicional de que la etiqueta de detección está protegida de la degradación por peptidasas y esto asegura que solo los polipéptidos no degradados asociados con etiquetas de detección completas son aislados durante la purificación de la proteína. La persona experta está consciente de que la expresión "la etiqueta de afinidad en el C-terminal de la etiqueta de detección" no implica necesariamente que la etiqueta de afinidad está situada inmediatamente en el C-terminal de la etiqueta de detección, sino que puede ser un enlazador de muchos aminoácidos separando la etiqueta de afinidad y la etiqueta de detección.

En ciertas realizaciones, la etiqueta de afinidad se separa de dicha etiqueta de detección por un segundo elemento separable, y dicho segundo elemento separable se separa antes de la etapa i. De este modo, solo se analiza la etiqueta de detección sin la etiqueta de afinidad asociada por espectrometría de masas.

Las especificaciones de los patrones de masa y de fragmentación de la etiqueta de detección se relacionan con el patrón de masa y de fragmentación de la etiqueta después de que se ha separado del polipéptido asociado y la etiqueta de afinidad, es decir, después de la separación del primer y el segundo elemento separable. La persona experta está consciente de que en casos donde la etiqueta de detección no esté liberada de una etiqueta de afinidad asociada antes de la espectrometría de masas, esto influirá los resultados del análisis de espectrometría de masas. Si todas las etiquetas de detección están asociadas con la misma etiqueta de afinidad, se pueden explicar los cambios en el patrón de masa y de fragmentación, por lo tanto, aún será posible identificar la etiqueta de detección, aunque no tan eficiente y con corte

limpio como en los casos en que la etiqueta de detección ha sido separada de la etiqueta de afinidad mediante separación del segundo elemento separable.

En ciertas realizaciones, la etiqueta de afinidad es una etiqueta-His.

5

En ciertas realizaciones, la etapa h comprende analizar la etiqueta de detección aislada a través de cromatografía de líquidos acoplada a espectrometría de masas por ionización de electroaspersión (LC-MS [por sus siglas en inglés]). En ciertas realizaciones, esta etapa comprende cromatografía de líquidos de fase inversa. Las etiquetas de detección aisladas se separan de acuerdo con su hidrofobicidad por cromatografía de fase inversa para reducir la complejidad de la muestra. Posteriormente, su patrón de masa y de fragmentación se registra por espectrometría de masas.

10

En ciertas realizaciones, la etapa d comprende secuenciar el banco de expresión etiquetado completo con una cobertura de $\geq 5x$. En ciertas realizaciones, la etapa d comprende secuenciación profunda del banco de expresión etiquetado.

15

En ciertas realizaciones, la etapa d comprende insertar la secuencia codificante de polipéptido y la secuencia codificante de etiqueta comprendida en el banco de ácidos nucleicos etiquetados en conjunto en un vector secuenciador. La secuenciación profunda comprende usualmente una etapa de amplificación por PCR. Los inventores se dieron cuenta de que la amplificación por PCR lleva a un número significativo de eventos de recombinación entre los segmentos del gen de los miembros etiquetados del banco. De este modo, se construyó un conjunto de plásmidos de secuenciación profunda, que permite la sujeción de elementos de secuencia requeridos para secuenciación profunda por digestión y ligadura de restricción, de esa manera se elimina la necesidad de amplificación por PCR del banco anidado antes de la secuenciación profunda.

20

En ciertas realizaciones, la etiqueta de detección aislada consiste en 5 a 30 aminoácidos contiguos y comprende uno y solo un aminoácido que tiene una cadena lateral cargada positivamente. En ciertas realizaciones, la etiqueta de detección aislada consiste en 7 a 21 aminoácidos contiguos y comprende uno y solo un aminoácido que tiene una cadena lateral cargada positivamente. En ciertas realizaciones, la etiqueta de detección aislada consiste en 11 a 15 aminoácidos contiguos y comprende uno y solo un aminoácido que tiene una cadena lateral cargada positivamente.

25

En ciertas realizaciones, el aminoácido que tiene una cadena lateral cargada positivamente está situado en el C-terminal de la etiqueta de detección aislada. En ciertas realizaciones, el aminoácido que tiene una cadena lateral cargada positivamente se selecciona de arginina (R) y lisina (K). En ciertas realizaciones, el aminoácido que tiene una cadena lateral cargada positivamente es una arginina (R) situada en el C-terminal de la etiqueta de detección aislada.

30

La persona experta está consciente de que además de que el aminoácido tenga una cadena lateral cargada positivamente, la etiqueta de detección aislada porte otra carga positiva en el pH neutral, que es la amina primaria en el N terminal de la etiqueta de detección aislada.

35

En ciertas realizaciones, la etiqueta de detección aislada comprende un elemento de secuencia I seleccionado de una colección de elementos de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P.

40

En ciertas realizaciones, el uno y único aminoácido que tiene una cadena lateral cargada positivamente está situado en el C-terminal de la etiqueta de detección aislada y los aminoácidos restantes son independientemente seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P. En ciertas realizaciones, el uno y único aminoácido que tiene una cadena lateral cargada positivamente es un R situado en el C-terminal de la etiqueta de detección aislada.

45

Las etiquetas de detección aisladas son detectables de manera óptima por espectrometría de masas, en particular por LC-MS (cromatografía de líquidos de fase inversa acoplada a ESI-MS). Los aminoácidos C y M fueron omitidos en el diseño de la etiqueta de detección debido a que son propensos a la oxidación. Los aminoácidos K, R y H fueron omitidos en el elemento de secuencia I, debido a que podrían agregar un aminoácido adicional con una cadena lateral cargada positivamente a la etiqueta, que no era deseable debido a que la etiqueta portaría una carga adicional durante la detección por ESI-MS y caería fuera del intervalo óptimo de detección. K y R agregarían sitios adicionales de escisión de la tripsina en la secuencia de etiqueta, que no era deseable.

50

Agregar un K a la secuencia de aminoácidos de la etiqueta de detección agregaría otra amina primaria, que complicaría el marcaje de la etiqueta de detección por etiquetas isobáricas para cuantificación relativa y absoluta por espectrometría de masas usando la química de la NHS.

55

En ciertas realizaciones, la etiqueta de detección aislada comprende

60

- a. el elemento de secuencia I, en donde el elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- b. el elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

65

En ciertas realizaciones, la etiqueta de detección aislada consiste en

- a. el elemento de secuencia III: GS;
- b. el elemento de secuencia I, en donde el elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- c. el elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

El orden de los elementos de secuencia del N-terminal al C-terminal es: el elemento de secuencia III, el elemento de secuencia I, elemento de secuencia II. Estas etiquetas de detección caen dentro de un intervalo de masa entre 903 y 2180 Da, que es óptimo para detección sensible por ESI-MS. Las etiquetas aisladas portan dos cargas positivas a pH fisiológico y por debajo de este, es decir, un R en C-terminal y el amino primario N-terminal. La carga positiva en el C-terminal de la etiqueta de detección aislada facilita la ionización de la etiqueta para la detección por espectrometría de masas y actúa como sitio único de escisión de la tripsina. Los péptidos con argininas o lisinas con C-terminal son particularmente bien detectados por espectrometría de masas (propiedades de ionización favorables). En cada etiqueta de detección aislada el amino N-terminal es la única amina primaria, que se usa para acoplamiento del amino a través de la química de la NHS. Esto permite que los marcajes se sujeten para la espectrometría de masas para que realicen, por ejemplo, iTRAQ (etiquetas isobáricas para cuantificación relativa y absoluta). Las etiquetas de detección fueron diseñadas para presentar un intervalo de hidrofobicidades idealmente adecuados para la separación de péptidos por columnas de cromatografía de fase inversa.

En ciertas realizaciones, todos los elementos de secuencia I comprendidos en el primer banco de ácidos nucleicos constituyen en conjunto una colección de elementos de secuencia I. Dentro de la colección de elementos de secuencia I, cada aminoácido tiene lugar con una frecuencia especificada en la tabla 1.

25

Tabla 1

A	S	T	N	Q	D	E	V	L	F	Y	W	G	P	Total
18	6	12	1	1	11	11	12	2	1	4	1	8	12	100

En ciertas realizaciones, uno de dicho primer y/o dicho elemento separable es o comprende una secuencia de reconocimiento de proteasa. En ciertas realizaciones, ambos de dichos primer y de dicho segundo elemento separable es o comprende una secuencia de reconocimiento de proteasa.

30

En ciertas realizaciones, el primer elemento separable es o comprende una secuencia de reconocimiento de trombina y/o el segundo elemento separable es o comprende una secuencia de reconocimiento de tripsina.

35 *Colección de polipéptidos*

De acuerdo con un segundo aspecto, se provee una colección de polipéptidos, de acuerdo con la reivindicación 5. Cada miembro de la colección

de polipéptidos está asociada con una etiqueta de detección. En ciertas realizaciones, cada miembro de la colección de polipéptidos está asociado con al menos una etiqueta de detección. La expresión "asociado con al menos una etiqueta de detección" se refiere al hecho de que cada miembro de la colección de polipéptidos puede estar asociado con más de una etiqueta de detección, pero solo una etiqueta por molécula de polipéptido. En otras palabras, la colección de polipéptidos puede comprender un polipéptido A asociado con la etiqueta de detección X y un polipéptido A asociado con la etiqueta de detección Y, pero no un polipéptido A asociado con ambas etiquetas de detección X y Y. En ciertas realizaciones, cada miembro de la colección de polipéptidos se asocia con al menos dos etiquetas de detección. En ciertas realizaciones, cada miembro de la colección de polipéptidos está asociado con al menos cinco etiquetas de detección. En ciertas realizaciones, cada miembro de la colección de polipéptidos está asociado con al menos diez etiquetas de detección. En ciertas realizaciones, cada miembro de la colección de polipéptidos está asociado con aproximadamente veinte etiquetas de detección. Cada etiqueta de detección tiene las siguientes características:

50

- a. La etiqueta está caracterizada por una secuencia etiqueta de aminoácidos diferentes de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por la pluralidad de vectores de expresión.
- b. La etiqueta está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre aproximadamente 900 y aproximadamente 2200 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 903 y 2180 Da.
- c. La etiqueta está separada de dicho miembro de dicha colección de polipéptidos por un primer elemento separable.

55

La etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -27 y 128. En ciertas realizaciones, la etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -1 y 70. En ciertas realizaciones del segundo aspecto de la invención, el miembro de la colección de polipéptidos está asociado con una etiqueta de afinidad.

60

En ciertas realizaciones del segundo aspecto de la invención, la etiqueta de detección está asociada con una etiqueta de afinidad. La etiqueta de afinidad y la etiqueta de detección están comprendidas dentro de la misma secuencia de aminoácidos primaria. La etiqueta de afinidad se separa de la etiqueta de detección por un segundo elemento separable. La etiqueta de detección se puede liberar de la etiqueta de afinidad a través de separar el segundo elemento separable.

5 En ciertas realizaciones, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-Strepll, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP. En ciertas realizaciones, la etiqueta de afinidad es una etiqueta-His.

10 La etiqueta de detección aislada consiste en 7 a 21 aminoácidos contiguos y comprende uno y solo un aminoácido que tiene una cadena lateral cargada positivamente. En ciertas realizaciones, la etiqueta de detección aislada consiste en 11 a 15 aminoácidos contiguos y comprende uno y solo un aminoácido que tiene una cadena lateral cargada positivamente.

15 El aminoácido que tiene una cadena lateral cargada positivamente está situado en el C-terminal de la etiqueta de detección aislada. El aminoácido que tiene una cadena lateral cargada positivamente se selecciona de arginina (R) y lisina (K). En ciertas realizaciones, el aminoácido que tiene una cadena lateral cargada positivamente es una arginina (R) situada en el C-terminal de la etiqueta de detección aislada.

20 En ciertas realizaciones del segundo aspecto de la invención, la etiqueta de detección comprende

- a. el elemento de secuencia I, en donde el elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- b. el elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

25 *Etiqueta de detección*

De acuerdo con un tercer aspecto que no pertenece a la invención reclamada como tal, se provee una etiqueta de detección de péptido que está diseñada para detección óptima por espectrometría de masas. La etiqueta de detección

30 consiste en 4 a 20 aminoácidos y tiene los siguientes atributos:

- a. La etiqueta de detección comprende solo un aminoácido que tiene una cadena lateral cargada positivamente.
 - b. La etiqueta de detección está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta de detección está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la
- 35 etiqueta de detección está caracterizada por una masa molecular de entre 900 y 2200 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 903 y 2180 Da.

En ciertas realizaciones del tercer aspecto de la invención, la etiqueta de detección consiste en 7 a 18 aminoácidos. En ciertas realizaciones del tercer aspecto de la invención, la etiqueta de detección consiste en 11 a 15 aminoácidos.

40

En ciertas realizaciones del tercer aspecto de la invención, la etiqueta de detección consiste esencialmente en

- a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
 - b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).
- 45

El aminoácido que tiene una cadena lateral cargada positivamente está situado en el C-terminal de la etiqueta de detección aislada. El aminoácido que tiene una cadena lateral cargada positivamente se selecciona de arginina (R) y lisina (K).

50

Colección de etiquetas de detección

De acuerdo con otro aspecto, se provee una colección etiquetas de péptidos, de acuerdo con la reivindicación 7. La colección de etiquetas de péptidos

55

comprende las etiquetas de péptidos de acuerdo con el tercer aspecto de la invención. Cada etiqueta de detección comprendida en la colección de etiquetas de péptidos consiste en 7 a 18 aminoácidos y está caracterizada por una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección comprendida en dicha colección de etiquetas de detección. En ciertas realizaciones, cada etiqueta de detección consiste en 11 a 15 aminoácidos. En ciertas realizaciones, la colección de etiquetas de péptidos comprende al menos 96 etiquetas de

60

péptidos. En ciertas realizaciones, la colección de etiquetas de péptidos comprende al menos 500.000 etiquetas de péptidos. En ciertas realizaciones, la colección de etiquetas de péptidos comprende al menos 10^7 etiquetas de péptidos. En ciertas realizaciones, la colección de etiquetas de péptidos comprende aproximadamente 10^8 etiquetas de péptidos.

En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección comprende solo un aminoácido que tiene una cadena lateral cargada positivamente y los aminoácidos restantes se seleccionan de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P.

65

La etiqueta está caracterizada por un valor de hidrofobicidad de entre -27 y 128. En ciertas realizaciones, la etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -1 y 70.

5 En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección está asociada con una etiqueta de afinidad. En ciertas realizaciones, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-Strep II, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP. En ciertas realizaciones la etiqueta de afinidad es una etiqueta-His. La etiqueta de afinidad y la etiqueta de detección están comprendidas dentro de la misma secuencia de aminoácidos primaria. La etiqueta de afinidad se separa de la etiqueta de detección por un elemento separable.

En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección consiste esencialmente en

15 a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

20 *Colección de vectores de plásmidos*

De acuerdo con aún otro aspecto, se provee una colección etiquetas de vectores de plásmidos, de acuerdo con la reivindicación 9. Cada miembro de dicha

25 colección de vectores de plásmidos comprende una secuencia de ácidos nucleicos que codifican una etiqueta de detección. Cada etiqueta de detección consiste en 7 a 18 aminoácidos y está caracterizada por una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por dicha colección de vectores de plásmidos. En ciertas realizaciones, cada etiqueta de detección consiste en 11 a 15 aminoácidos. En ciertas realizaciones, la colección de vectores de plásmidos comprende al menos 96 vectores de plásmidos. En ciertas realizaciones, la colección de vectores de plásmidos comprende al menos 500.000 vectores de plásmidos. En ciertas realizaciones, la colección de vectores de plásmidos comprende al menos 10^7 vectores de plásmidos. En ciertas realizaciones, la colección de vectores de plásmidos comprende aproximadamente 10^8 vectores de plásmidos.

La etiqueta de detección comprende solo un aminoácido que tiene una cadena lateral cargada positivamente.

35 En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta de detección está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la etiqueta de detección está caracterizada por una masa molecular de entre 900 y 2200 Da. En ciertas realizaciones, la etiqueta de detección está caracterizada por una masa molecular de entre 903 y 2180 Da.

40 La etiqueta está caracterizada por un valor de hidrofobicidad de entre -27 y 128. En ciertas realizaciones, la etiqueta de detección está caracterizada por un valor de hidrofobicidad de entre -1 y 70.

45 En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección está asociada con una etiqueta de afinidad. En ciertas realizaciones, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-StrepII, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP. En ciertas realizaciones la etiqueta de afinidad es una etiqueta-His. La etiqueta de afinidad y la etiqueta de detección están comprendidas dentro de la misma secuencia de aminoácidos primaria. La etiqueta de afinidad se separa de la etiqueta de detección por un segundo elemento separable.

En ciertas realizaciones de este aspecto de la invención, la etiqueta de detección consiste esencialmente en

55 a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos contiguos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

60 En ciertas realizaciones de este aspecto de la invención, cada miembro de la colección de vectores de plásmidos comprende

a. un casete de selección negativa flanqueado por 5' con un primer sitio de restricción de endonucleasa y 3' con un segundo sitio de restricción de endonucleasa;
b. un promotor situado en 5' del primer sitio de restricción de endonucleasa;
65 c. la secuencia etiqueta de ácido nucleico codificante de etiqueta de detección, situada en 3' del segundo sitio de restricción de endonucleasa. En ciertas realizaciones, la secuencia de ácido nucleico que codifica la etiqueta de detección

5 y el segundo sitio de restricción de endonucleasa están separados por menos que 100 pares de bases. En ciertas realizaciones, la secuencia de ácido nucleico que codifica la etiqueta de detección y el segundo sitio de restricción de endonucleasa están separados por menos que 50 pares de bases. En ciertas realizaciones, la secuencia de ácido nucleico que codifica la etiqueta de detección y el segundo sitio de restricción de endonucleasa están separados por aproximadamente 20 pares de bases. En ciertas realizaciones, los pares de bases situados entre la secuencia de ácidos nucleicos que codifica la etiqueta de detección y el segundo sitio de restricción de endonucleasa codifican un primer elemento separable.

10 En ciertas realizaciones de este aspecto de la invención, cada miembro de la colección de vectores de plásmidos comprende

- a. la secuencia etiqueta de ácidos nucleicos que codifica la etiqueta de detección, asociada dentro del mismo marco de lectura con una secuencia de ácidos nucleicos que codifica un polipéptido;
- 15 b. un elemento de diversidad que comprende bases no idénticas para impedir una sobrecarga de señal durante la secuenciación;
- c. un sitio de enlace del cebador para el enlace de cebadores secuenciadores;
- d. un elemento índice que comprende una de muchas secuencias de ácidos nucleicos definidas para multiplexar;
- e. un elemento adaptador para inmovilizar la molécula de ADN durante la secuenciación y
- 20 f. dos elementos flanqueantes de sitios de restricción de endonucleasa a-e para liberar el fragmento de ADN del vector plásmido antes de la secuenciación.

Los vectores de plásmidos descritos en la realización anterior sirven como plásmidos de secuenciación profunda. Preferentemente, estos vectores no comprenden la etiqueta de afinidad con el fin de reducir la longitud del fragmento que se va a secuenciar.

25 *Método de detección de proteína*

De acuerdo con otro aspecto, se provee un método de detección de proteína, de acuerdo con la reivindicación 11. El método comprende las siguientes etapas:

- a. Se provee un banco de ácidos nucleicos que codifica un banco de polipéptidos. Cada polipéptido comprendido en el banco de polipéptidos se asocia con una etiqueta de detección. La etiqueta de polipéptido y de detección están comprendidas dentro de la misma secuencia de aminoácidos primaria. Cada etiqueta de detección tiene las siguientes características:
 - 35 i. La etiqueta está caracterizada por una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por el banco de ácidos nucleicos.
 - ii. La etiqueta está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre aproximadamente 900 y aproximadamente 2200 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 903 y 2180 Da.
 - 40 iii. La etiqueta está separada del polipéptido asociado por un primer elemento separable.

45 Cada etiqueta de detección codificada por el banco de ácidos nucleicos es única con respecto a otra etiqueta de detección codificada por el banco de ácidos nucleicos. Cada polipéptido comprendido en el banco de polipéptidos se asocia con al menos una etiqueta de detección. En ciertas realizaciones, cada polipéptido comprendido en el banco de polipéptidos está asociado con al menos dos etiquetas de detección. En ciertas realizaciones, cada polipéptido comprendido en el banco de polipéptidos está asociado con al menos cinco etiquetas de detección. En ciertas realizaciones, cada polipéptido comprendido en el banco de polipéptidos está asociado con al menos diez etiquetas de detección. En ciertas realizaciones, cada polipéptido comprendido en el banco de polipéptidos está asociado con aproximadamente veinte etiquetas de detección. Cada molécula de polipéptido comprende solo una etiqueta de detección.

b. Se provee una base de datos. La base de datos comprende la siguiente información:

- 55 i. Una pluralidad de ácidos nucleicos y/o secuencias de aminoácidos. La pluralidad de secuencias comprende las secuencias de todos los miembros del banco de ácidos nucleicos. Cada una de las secuencias comprende una secuencia especificando un polipéptido y una secuencia especificando una etiqueta de detección.
- ii. Un patrón de fragmentación de espectrometría de masas para cada etiqueta de detección codificada por el banco de ácidos nucleicos.
- 60 c. El banco de polipéptidos se expresa del banco de ácidos nucleicos.
- d. Un miembro del banco de polipéptidos se selecciona en una etapa de selección, produciendo un polipéptido seleccionado.
- e. El primer elemento separable se separa. De ese modo, la etiqueta de detección se separa del polipéptido seleccionado y se produce una etiqueta de detección aislada.
- 65 f. La etiqueta de detección aislada se identifica de la siguiente manera:

- 5 i. El patrón de fragmentación de la etiqueta de detección aislada se registra por espectrometría de masas.
 ii. El patrón de fragmentación obtenido en la etapa i se hace coincidir con el patrón de fragmentación predicho en la base de datos provista. De ese modo, se identifica la etiqueta de detección aislada. La combinación de la información obtenida por la espectrometría de masas con la información obtenida por secuenciación del banco de ácidos nucleicos etiquetados permite la identificación sin ambigüedad de una etiqueta de detección dada.
- 10 g. La secuencia especificando la etiqueta de detección identificada en la etapa f se selecciona de la pluralidad de secuencias comprendidas en la base de datos. De ese modo, se identifica el miembro del banco de polipéptidos asociado con la etiqueta de detección identificada en la etapa f.
- En ciertas realizaciones, cada miembro de dicho banco de polipéptidos se asocia con una etiqueta de afinidad.
- 15 En ciertas realizaciones, cada etiqueta de detección está asociada con una etiqueta de afinidad.
- En ciertas realizaciones, la etiqueta de afinidad se selecciona del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-StrepII, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP.
- 20 En ciertas realizaciones, la etiqueta de afinidad se separa de dicha etiqueta de detección por un segundo elemento separable, y dicho segundo elemento separable se separa antes de la etapa f. De este modo, solo se analiza la etiqueta de detección sin la etiqueta de afinidad asociada por espectrometría de masas.
- 25 Las especificaciones de los patrones de masa y de fragmentación de la etiqueta de detección se relacionan con el patrón de masa y de fragmentación de la etiqueta después de que se ha separado del polipéptido asociado y la etiqueta de afinidad, es decir, después de la separación del primer y el segundo elemento separable. La persona experta está consciente de que en casos donde la etiqueta de detección no esté liberada de una etiqueta de afinidad asociada antes de la espectrometría de masas, esto influirá los resultados del análisis de espectrometría de masas. Como todas las etiquetas de detección están asociadas con la misma etiqueta de afinidad, se pueden explicar los cambios en el patrón de masa y de fragmentación, por lo tanto, aún será posible identificar la etiqueta de detección, aunque no tan eficiente y con corte limpio como en los casos en que la etiqueta de detección ha sido separada de la etiqueta de afinidad mediante separación del segundo elemento separable.
- 30 En ciertas realizaciones, la etiqueta de afinidad es una etiqueta-His.
- 35 La persona experta está consciente que las etapas d a g se realizan para un número de miembros diferentes del banco de polipéptidos en paralelo. Un grupo de muchos polipéptidos se selecciona en la etapa g, y todos estos polipéptidos se identifican a través de sus etiquetas de detección en el análisis de espectrometría de masas. La persona experta está consciente de que, debido a razones técnicas, no se puede identificar cada polipéptido en esta etapa. El análisis de espectrometría de masas realizado en la etapa f es cuantitativo, de este modo, el método de acuerdo con la invención permite no solo identificar un polipéptido sino también cuantificar la cantidad de este polipéptido en una muestra.
- 40 *Método para asociar un polipéptido con una etiqueta de detección única*
- 45 De acuerdo con aún otro aspecto, se provee un método, de acuerdo con la reivindicación 13, para asociar un polipéptido con una etiqueta de detección única. El método comprende las siguientes etapas:
- 50 a. Se provee un primer banco de ácidos nucleicos. Cada miembro del primer banco de ácidos nucleicos comprende una secuencia codificante de polipéptidos que codifica un miembro de un primer banco de polipéptidos;
 b. Se provee un segundo banco de ácidos nucleicos. Cada miembro del segundo banco de ácidos nucleicos comprende una secuencia codificante de etiqueta que codifica una etiqueta de detección. Cada etiqueta de detección tiene las siguientes características:
- 55 i. La etiqueta está caracterizada por una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por el segundo banco de ácidos nucleicos;
 ii. La etiqueta está caracterizada por una masa molecular de entre 200 y 5000 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 500 y 2500 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre aproximadamente 900 y aproximadamente 2200 Da. En ciertas realizaciones, la etiqueta está caracterizada por una masa molecular de entre 903 y 2180 Da.
 60 iii. la etiqueta está caracterizada por un valor de hidrofobicidad entre -27 y 128; y
 iv. la etiqueta comprende un elemento separable.
- c. La secuencia codificante de polipéptido comprendida en el miembro del primer banco de ácidos nucleicos se inserta en un miembro del segundo banco de ácidos nucleicos. De ese modo, se genera una pluralidad de plásmidos de combinación de etiqueta-polipéptido.
- 65

El primer banco de ácidos nucleicos tiene un tamaño de 5 a 100.000. En ciertas realizaciones, el primer banco de ácidos nucleicos tiene un tamaño de 100 a 50.000. En ciertas realizaciones, el primer banco de ácidos nucleicos tiene un tamaño de 500 a 5.000.

5 El segundo banco de ácidos nucleicos tiene un tamaño de 10^3 a 10^{11} . En ciertas realizaciones, el segundo banco de ácidos nucleicos tiene un tamaño de 10^5 a 10^{10} . En ciertas realizaciones, el segundo banco de ácidos nucleicos tiene un tamaño de 10^6 a 10^9 . En ciertas realizaciones, el segundo banco de ácidos nucleicos tiene un tamaño de aproximadamente 10^8 .

10 Dentro de la pluralidad de plásmidos de combinación polipéptido/etiqueta, cada secuencia codificante de polipéptido del primer banco de ácidos nucleicos se asocia con una secuencia codificante de etiqueta del segundo banco de ácidos nucleicos. La asociación tiene lugar dentro del mismo marco de lectura.

d. Se selecciona un subconjunto de la pluralidad de plásmidos de combinación de polipéptido-etiqueta. Esta etapa de selección comprende seleccionar un número definido de clones, en donde cada clon comprende un miembro de la pluralidad de plásmidos de combinación polipéptido-etiqueta. De ese modo, se genera un banco de ácidos nucleicos etiquetados que codifica un banco de polipéptidos etiquetados. Cada miembro del banco de polipéptidos etiquetados comprende un polipéptido y una etiqueta de detección. Cada etiqueta está comprendida en solo un miembro del banco de polipéptidos etiquetados. En otras palabras, cada etiqueta de detección es única dentro del banco de polipéptidos etiquetados. Sin embargo, cada polipéptido puede estar comprendido en muchos miembros del banco de polipéptidos etiquetados (etiquetado redundante).

En ciertas realizaciones, cada polipéptido está asociado con al menos una etiqueta de detección. En ciertas realizaciones, cada polipéptido está asociado con al menos dos etiquetas de detección. En ciertas realizaciones, cada polipéptido está asociado con al menos cinco etiquetas de detección. En ciertas realizaciones, cada polipéptido está asociado con al menos diez etiquetas de detección. En ciertas realizaciones, cada polipéptido está asociado con aproximadamente veinte etiquetas de detección.

En ciertas realizaciones de este aspecto de la invención, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es al menos $10 \times$ el número de miembros del primer banco de ácidos nucleicos. En ciertas realizaciones, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es al menos $20 \times$ el número de miembros del primer banco de ácidos nucleicos.

En ciertas realizaciones de este aspecto de la invención, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es menor que el 50 % del número de miembros del segundo banco de ácidos nucleicos. En ciertas realizaciones de este aspecto de la invención, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es menor que el 5 % del número de miembros del segundo banco de ácidos nucleicos. En ciertas realizaciones de este aspecto de la invención, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es menor que el 0,05 % del número de miembros del segundo banco de ácidos nucleicos.

Al seleccionar el tamaño óptimo del subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta, se asegura que, en el banco de polipéptidos etiquetados, cada etiqueta de detección es única (presente solamente una vez), pero cada polipéptido se presenta muchas veces, cada vez asociado con una etiqueta de detección diferente.

Donde sea que se encuentren alternativas para atributos separables como "realizaciones" en este documento, debe entenderse que tales alternativas se pueden combinar libremente para formar realizaciones discretas de la invención que se divulga en el presente documento.

50 La invención se ilustra además por los siguientes ejemplos y figuras, de los que se pueden extraer realizaciones y ventajas adicionales. Estos ejemplos pretenden ilustrar la invención, pero no limitar su alcance.

Breve descripción de las figuras

55 La Figura 1 muestra una revisión de la tecnología NestLink. A) Un banco de nanocuerpos se anida dentro de un banco Flycode codificado en el vector de expresión pNLx. Posteriormente, las secuencias de nanocuerpos con flycode se extirpan a través de digestión de restricción y se insertan en pNLs, que resulta en la sujeción de las secuencias adaptadoras requeridas para secuenciación profunda. Los adaptadores unidos a los nanocuerpos con flycode son extirpados a través de digestión de restricción y se someten a secuenciación profunda en la forma lineal. B) El banco anidado codificado en pNLx se expresa y se purifica. Se aplica una presión de selección (en este caso particular, las proteínas con un peso molecular aparente de un monómero de anticuerpo se seleccionan a través de cromatografía de exclusión de tamaño) y los Flycodes de los nanocuerpos seleccionados se aíslan a través de escisión de proteasa. C) Los datos de secuenciación profunda permiten la generación de una base de datos que asigna todos los Flycodes a su nanocuerpo correspondiente. Los Flycodes de cada nanocuerpo se concatenan. Los Flycodes anteriormente aislados (véase B) se someten a LC-MS y se generan listas de picos de los datos registrados de MS/MS. Los datos de MS/MS se buscan contra la base de datos que contiene los Flycodes concatenados, que permite la identificación y cuantificación

relativa de los nanocuerpos seleccionados.

La Figura 2 muestra los diseños de plásmidos pertinentes para la tecnología NestLink antes (tiras superiores) y después de la inserción del banco (tiras inferiores). A) El fagémido usado para las selecciones de presentación de fago de los nanocuerpos contra las moléculas diana. El fagémido porta dos sitios de restricción SapI que permiten la inserción de bancos de nanocuerpos y su transferencia eficiente al vector de expresión de NestLink pNLx después del enriquecimiento por presentación de fago. B) el vector de expresión de NestLink pNLx que alberga la diversidad Flycode de aproximadamente 10^8 variantes. Los sitios SapI están diseñados para desvanecerse ante la inserción del banco de nanocuerpos. Los nanocuerpos en flycode se pueden extirpar específicamente del vector de expresión a través de restricción de la SfiI. La colocación de los sitios SfiI asegura la secuenciación profunda de los nanocuerpos enteros sujetos a sus Flycodes correspondientes, pero minimiza la longitud de la lectura de secuenciación profunda por la exclusión de secuencias redundantes, tales como PelB y la etiqueta-His. C) Se generó un conjunto de vectores de secuenciación profunda (pLNs) con diversos albergando cada uno todas las secuencias necesarias para la secuenciación con Illumina MiSeq. Los nanocuerpos en flycode se insertan en este vector a través de restricción y ligadura SfiI. Posteriormente, son liberados como un fragmento lineal que contiene todas las regiones adaptadoras MiSeq por restricción BseRI. De esta manera, no se requiere de PCR para generar fragmentos de ADN para el análisis MiSeq, que podría resultar en eventos de recombinación en las secuencias nanocuerpo-Flycode y, de ese modo, destruir la unión entre el Flycode y la secuencia de anticuerpo. D) Los adaptadores de secuenciación profunda se pueden sujetar también a través de oligonucleótidos adaptadores de doble cadena sintéticos a través de salientes de una sola cadena adecuada complementaria al sitio de restricción de la SfiI codificado en pNLx.

La Figura 3 muestra la cuantificación absoluta de miembros PLOI a través de Flycodes usando LC-MS. Se agregaron siete cantidades conocidas (eje de la x) de sicuerpos con flycode fueron agregados en dos diferentes muestras que contienen lisado de *E. coli* o de *M. smegmatis*, respectivamente (fondo). Los sicuerpos con flycode fueron agregados a 0,2, 0,4, 1,3, 4,1, 8,5, 18,0 y 27,5 unidades de absorbancia (280nm) y que contenían 28, 56, 112, 56, 112, 84 y 112 Flycodes como se determinó por secuenciación profunda. Los Flycodes se analizaron por LC-MS. Las intensidades de MS1 de todos los Flycodes de cada sicuerpo se sumaron usando el software Progenesis.

La Figura 4 muestra la identificación de sicuerpos que muestran las mejores constantes de disociación de $1'080$ candidatos enlazadores a través de NestLink. A: Los sicuerpos monoméricos coeluyen con la proteína diana biotilada en solución (SEC) fueron inmovilizados en dos columnas equivalentes de estreptavidina sefarosa. Una columna fue lavada con solución amortiguadora, la otra columna fue lavada por un exceso de proteína diana no biotilada durante 3 min. Posteriormente, los Flycodes de los sicuerpos unidos restantes se aislaron y cuantificaron a través de intensidades de LC-MS1. B: las intensidades de LC-MS1 (la suma de todos los flycodes) se determinaron para cada miembro del grupo y la razón entre las dos columnas se trazó en el eje de la y para cada sicuerpo individual (eje de la x). Los sicuerpos, que no se expresaron, que no eran monoméricos o que no se enlazaron a la diana en la solución no fueron detectables en ninguna de las columnas, una vez que fueron quitados como resultado de las presiones de preselección descritas en la prueba del experimento de prueba de principio (sicuerpos 320 - $1'080$). Los sicuerpos de enlace débiles solo fueron detectables después del lavado con solución amortiguadora, pero no bajo competencia con dianas en exceso (sicuerpos 187 - 320). Los sicuerpos 1-186 fueron detectados en ambas columnas y fueron clasificados de acuerdo con su constante de disociación. Los sicuerpos más prometedores para aplicaciones cuesta abajo son aquellos con las tasas de constante de disociación resultantes en una razón cercana a 1. C: Correlación de la lectura NestLink y de experimentos SPR de sicuerpos agregados individualmente. Las secuencias de ADN de 11 sicuerpos analizados en B) fueron sintetizadas (síntesis genética) y se expresaron los enlazadores correspondientes, se purificaron y analizaron por resonancia de plasmones de superficie uno por uno. Los datos SPR se trazan como la señal de enlace residual después de 3 minutos de lavado (como una medición de la constante de disociación) en el eje de la x en comparación con la razón del sicuerpo determinada por NestLink como se muestra en B) en el eje de la y.

Figura 5: Análisis de 3'469 nanocuerpos de una alpaca inmunizada y la identificación de estos, que muestran el enlace a antígeno más fuerte en solución. Después de eliminar aquellos miembros del grupo con pobres niveles de expresión (etapa 1) y de solubilidad (etapa 2, selección de nanocuerpos monoméricos), la fracción monomérica del grupo se incubó con el antígeno de proteína de membrana en tres diferentes razones estequiométricas y se analizó a través de SEC. Las muestras de LC-MS se recolectaron después de la etapa 1 (reportado en el nivel de expresión de cada miembro individual del grupo), en la etapa 2 (reportado en la solubilidad de cada miembro individual del grupo) y de todos los picos diana/complejo en la etapa 3. Los gráficos circulares representan la cantidad relativa de cada nanocuerpo en el grupo (aquellos que no son enlazadores o que son enlazadores débiles coloreados de manera colectiva en gris claro, la cantidad total de miembros del grupo corresponden al 100 %) en diferentes etapas del procedimiento de selección, según se determinó por la suma de todas las intensidades de MS1 para cada nanocuerpo (100 % = suma de todas las intensidades de MS1 de todos los flycodes de todos los nanocuerpos). Como se esperaba para la etapa 3, un aumento en la razón del grupo al antígeno lleva a un aumento en la competencia interna de los muchos miembros del grupo de enlace para la cantidad limitada de antígeno. Por lo tanto, la fracción de miembros del grupo con la afinidad más fuerte a antígeno aumenta a mayor competencia para los epítomos limitados.

Figura 6: A: Análisis de 1'456 sicuerpos de un grupo generado por unas selecciones *in vitro* (etapa 1) contra una diana proteína de membrana externa purificada, para superficie celular que se enlaza a las bacterias Gramnegativas de interés (etapa 2). En la etapa 2 (NestLink), aquellos miembros del grupo con pobres niveles de expresión y de solubilidad fueron

eliminados primero del ensamble, seguido por 4 diferentes experimentos de inmunoprecipitación usando 4 diferentes cepas bacterianas de interés. Después de quitar los miembros del grupo mediante lavado y que no se unieron con alta afinidad a las células, todos los Flycodes del grupo fueron aislados y analizados por LC- MS. La suma de todas las intensidades de MS1 de todos los Flycodes por sicuerpo se pueden usar después como una medición para la concentración relativa de cada sicuerpo individual en el grupo en cada una de las células diana. Esto permitió una lectura sin ambigüedad de especificidad celular (B) reportando para cada sicuerpo (eje de las x) su concentración relativa (en comparación con el grupo entero) en cada uno de los 4 tipos celulares. Por razones de claridad, solo el 25 % de todos los sicuerpos analizados se muestran en B.

10 Ejemplos

Banco de secuencia flycode

15 Se diseñó un banco aleatorizado de péptidos cortos codificados de ADN para que fuese óptimamente detectable por espectrometría de masas (MS), en particular por LC-MS (cromatografía de líquidos inversa acoplada a ESI-MS). Los péptidos cayeron dentro de un intervalo de masa entre 903 y 2180 Da, que es óptimo para detección sensible por ESI-MS. Los Flycodes portan dos cargas positivas a pH fisiológico y por debajo de este, es decir, un R en el C-terminal y el amino primario N-terminal. La carga positiva en el C-terminal de los flycodes facilita la ionización del péptido para la detección por espectrometría de masas y actúa como sitio único de escisión de la tripsina. En cada flycode el amino N-terminal es la única amina primaria, que se usa para acoplamiento del amino a través de simple química de la NHS. Esto permite que los marcajes se sujeten para la espectrometría de masas para que realicen, por ejemplo, iTRAQ (etiquetas isobáricas para cuantificación relativa y absoluta). Los flycodes fueron diseñados para presentar un intervalo de hidrofobicidades idealmente adecuado para la separación de péptidos por columnas de cromatografía de fase inversa.

25 El banco flycode consiste en dos partes más aminoácidos flanqueantes que son constantes, es decir, GS en el N- terminal y R en el C-terminal. La secuencia "GS" N-terminal es parte del sitio de escisión de la proteasa trombina, que permanece en el flycode después de la escisión.

30 Parte 1: La región código de barra abarca 7 posiciones de aminoácidos aleatorizados consecutivos. La frecuencia promedio de aminoácidos está dada en la anteriormente mencionada tabla 1 (en %).

35 No están presentes los veinte aminoácidos naturales en el código de barras (faltan C, M, K, R, H y I). C y M fueron omitidos debido a que son propensos a la oxidación. K, R y H fueron omitidos debido a que podrían agregar una carga positiva adicional a la secuencia flycode, que no era deseable debido a que el péptido podría portar en ese caso una carga adicional durante la detección ESI-MS y caer fuera del intervalo de detección óptimo. K y R agregarían sitios adicionales de escisión de la tripsina en la secuencia flycode, que no era deseable. K agregaría otra amina primaria, que complicaría el marcaje del péptido por química de la NHS. Se omitió la Isoleucina debido a que no se puede distinguir de la Leucina por masa.

40 Parte 2: El C-terminal se construyó en 5 diferentes variantes, que eran igualmente frecuentes en el banco flycode y que terminaban todos en un R. También carecían de C, M, K, H y I. Por lo tanto, los flycodes consisten en mínimamente 11 aminoácidos y máximo en 15 aminoácidos (GS + 7 residuos aleatorizados + 2-6 residuos C-terminal). Los 5 diferentes extremos C-terminal se listan aquí:
SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR), SEQ ID NO 05 (WQEGGR).

El vector de expresión de NestLink pLNx que contiene el banco flycode.

45 El vector de expresión de NestLink pLNx alberga el banco flycode con una diversidad de 10^8 variantes de secuencia (figura 2) y permite introducir un banco de proteínas de interés (PLOI) en el marco con los flycodes. El resultado de esta etapa es un "banco anidado" dado que dos bancos (el banco PLOI y el flycode) están anidados entre sí. El vector de expresión permite también la extracción del banco anidado (PLOI fusionado a flycodes) mediada por enzima de restricción, de forma que se puede insertar en el plásmido de secuenciación profunda o que dirige la ligadura del adaptador Illumina MiSeq que se puede realizar usando oligonucleótidos de doble cadena (adaptadores). Es de hacerse notar que el PLOI puede ser cualquier banco codificado genéticamente.

55 El PLOI se introduce en el vector de expresión al digerir por restricción un ADN fuente que codifica el banco, seguido de ligadura en el vector de expresión. Los inventores usaron un tipo de enzima de restricción IIS (SapI) para este fin. El ADN fuente usualmente procede de un fagémido obtenido después de las selecciones de presentación del fago, que contienen los sitios SapI orientados de forma que el PLOI se puede subclonar en el vector de expresión NestLink sin amplificación de PCR (descripción de este vector, véase a continuación). Cuando se inserta el PLOI, reemplaza un casete de selección negativa (ccdB), que mejora enormemente la eficiencia de la etapa de inserción.

60 El flycode se escinde del PLOI por trombina y la etiqueta-His se elimina del flycode por tripsina. Estas escisiones aseguran que los péptidos con masa óptima, hidrofobicidad y carga óptimas sean aislados para la espectrometría de masas (véase la descripción flycode anteriormente mencionada). También es concebible que cualquier otra combinación de proteasas se pueda usar para el mismo fin.

Es de hacerse notar que la arginina del C-terminal (R) del flycode juega un papel importante: primero, es la única amina cargada positivamente del flycode, una vez que las lisinas y otras argininas se omiten en el banco flycode. Por este motivo, la tripsina - una proteasa que se escinde después de los residuos cargados positivamente y, que, por lo tanto, se considera que es bastante inespecífica - se puede usar para escindir específicamente el enlace peptídico entre la arginina y la etiqueta-His (el flycode sería también demasiado pesado con la etiqueta-His para el análisis de espectrometría de masas y la etiqueta-His reduciría la separación en cromatografía de fase inversa antes de la espectrometría de masas). Segundo, se sabe que los péptidos con argininas en el C-terminal son particularmente bien detectados por espectrometría de masas (propiedades de ionización favorables). Y tercero, debido a este aminoácido cargado positivamente presente en el flycode, la carga total es consistentemente 2+ (N-terminal + arginina, todos los demás residuos son neutrales al pH bajo de la detección), lo que facilita el análisis de datos.

Un aspecto importante de la tecnología es el hecho de que es posible (y necesario) sujetar muchos flycodes únicos al mismo miembro del banco de proteínas de interés. Por ejemplo, para analizar un grupo de 100 proteínas diferentes, se sujetan 2000 flycodes a estas 100 proteínas de forma que, en promedio, cada proteína del grupo está unida 20 veces a un flycode diferente (la razón entre los miembros del grupo y los flycodes puede variarse de hecho según se desee). El etiquetado redundante facilita la detección sin ambigüedad de miembros del grupo a través de múltiples secuencias flycode y promedia las influencias potenciales de la secuencia flycode a las propiedades biofísicas de las proteínas analizadas de interés. El etiquetado redundante permite también la determinación de cantidades relativas de diferentes miembros del banco de proteínas dentro de una muestra seleccionada o del mismo miembro del banco de proteínas dentro de muestras seleccionadas de forma diferente. La redundancia se requiere además por razones técnicas: a pesar de que los flycodes están diseñados para la detección óptima por espectrometría de masas, algunos flycodes no se detectarán debido a que se perderán durante la preparación de la muestra o no se eluirán dentro de la ventana de hidrofobicidad de la columna de fase inversa que se analiza por espectrometría de masas.

Además, el vector de expresión NestLink contiene dos sitios de restricción de la Sfil que permiten la extracción del banco anidado (PLOI fusionado a flycodes), de forma que se pueden insertar en el plásmido de secuenciación profunda o que los adaptadores Illumina MiSeq se pueden ligar directamente usando oligonucleótidos de doble cadena (adaptadores). La base para esta etapa crucial se provee a continuación.

Es de hacerse notar que los sitios de restricción de la Sfil y/u otros sitios de restricciones dentro del PLOI o entre el PLOI y el flycode se pueden usar para agregar secuencias adicionales al banco anidado. De ahí que estas secuencias adicionales se pueden expresar como una fusión al banco anidado (entre el flycode y el PLOI o adyacente al banco anidado). De manera importante, tales secuencias no aumentan la longitud de la lectura de secuenciación profunda (que está limitada debido a razones técnicas) una vez que la transferencia al plásmido de secuenciación profunda (o la ligadura del adaptador de secuenciación profunda directa a través de oligonucleótidos) se realiza *antes* de introducir estas secuencias adicionales. Además, agregar secuencias adicionales de esta manera mantiene la unión física entre el flycode y el PLOI, que es absolutamente crucial para una correcta asignación de flycodes a los miembros del PLOI.

40 *Plásmidos de secuenciación profunda*

Los plásmidos de secuenciación profunda son un conjunto de vectores que portan todas las secuencias necesarias para secuenciado profundo con Illumina MiSeq y que permiten la inserción de miembros de grupos de bancos anidados del vector de expresión NestLink.

La transferencia del banco anidado al plásmido de secuenciación profunda (figuras 1 y 2) se realiza a través de digestión y ligadura de restricción. Los inventores usaron la enzima de restricción Sfil para este fin, una vez que posee suficiente especificidad, que es crucial cuando se digieren bancos enteros que pueden codificar por casualidad sitios de restricción. Además, el sitio de reconocimiento de la Sfil seleccionado traduce en aminoácidos razonablemente flexibles e hidrofílicos que se pueden usar como aminoácidos enlazadores en la construcción de expresión.

Los inventores son capaces de mostrar, mediante experimentos, que es crucial para el NestLink que la etapa de transferencia del vector de expresión NestLink en el plásmido de secuenciación profunda no incluya la etapa de amplificación por PCR del banco anidado. La amplificación por PCR de las secuencias flycode de proteínas lleva inevitablemente a la recombinación de regiones no homólogas (por ejemplo, CDRs) entre miembros del banco y de flycodes (la sujeción no esperada del flycode de una proteína de interés en otra donde no se sujetó en el vector de expresión NestLink). De ese modo, la unión entre el flycode y la proteína se destruye.

Como se describió anteriormente, el banco anidado se corta del vector de expresión a través de la Sfil. Posteriormente, se liga al plásmido de secuenciación profunda. Este reemplaza un casete de selección negativa (ccdB), que es crucial para la eficiencia de la etapa de inserción. Después de la inserción, es flanqueado por las secuencias que son necesarias (y usadas frecuentemente) para secuenciación profunda con Illumina MiSeq. La secuenciación tiene lugar desde ambos lados hacia el centro. Por lo tanto, las regiones importantes están presentes en ambos lados del inserto en dirección opuesta (secuencias complementarias inversas, excepto el índice).

65 Aquí hay una descripción de la secuencia de la parte interna (el inserto) hacia las regiones exteriores:

Sitios de la Sfil: Son usados para reemplazar el ccdB por el banco anidado.

5 **Diversidad:** La tecnología Illumina MiSeq genera la primera señal de secuenciación con base en la siguiente secuencia al sitio de enlace del cebador. Las primeras bases deben ser diversas (no idénticas) para impedir la sobrecarga de señal de los canales de detección y que se aborte la corrida de la secuenciación.

Sitio de enlace del cebador: El cebador de secuenciación se une aquí.

10 **Índice** (marcado con los números 501 y 701): La tecnología Illumina MiSeq permite multiplexar, es decir, se pueden analizar muchas muestras en una corrida de secuenciación. Para determinar qué lectura pertenece a qué muestra se lee también un índice (extensión variable de 8 pb). Con el fin de permitir secuenciar muchos experimentos NestLink en una sola corrida de secuenciación profunda, los investigadores generaron un conjunto de 11 plásmidos de secuenciación profunda, cada uno llevando un par diferente de índices (es de hacerse notar que hay una secuencia índice en ambos lados del inserto).

Adaptador: Este se usa para inmovilizar la plantilla de ADN para la secuenciación profunda en la celda de flujo de Illumina MiSeq.

20 **Sitio de restricción BseRI:** Este se usa para crear un fragmento lineal de ADN que es necesario para la secuenciación profunda con Illumina MiSeq. El hecho de que BseRI es una enzima de restricción tipo IIS (escinde fuera de su secuencia de reconocimiento) es particularmente útil para minimizar el saliente en el adaptador.

25 En el método tradicional, todos estos elementos de secuencia de Illumina MiSeq se sujetan al ADN para ser secuenciados por PCR, por ligadura de los adaptadores Illumina, seguido por amplificación por PCR o por el Kit de Prep de Muestra TRuSeq DNA PCR-free (Illumina). En el protocolo de los inventores, el ADN que se va a secuenciar (aquí, las secuencias proteína-flycode) se subclona de un vector donador (aquí, el vector de expresión NestLink) en el vector de secuenciación profunda por restricción y ligadura, de ese modo se evita la PCR. En una etapa final, el vector de secuenciación profunda se escinde usando BseRI. Esto libera la plantilla de secuenciación completa de Illumina MiSeq que se separa del esqueleto del vector por gel de agarosa de ADN y se purifica por extracción de gel.

Oligonucleótidos adaptadores de doble cadena para secuenciación profunda

35 Una segunda estrategia que permite la sujeción independiente de la PCR de las secuencias adaptadoras necesarias para la secuenciación profunda con Illumina MiSeq a la PLOI se basa en oligonucleótidos de doble cadena que portan el mismo conjunto de secuencias adaptadoras que se describen para los plásmidos de secuenciación profunda, que se pueden generar a través de síntesis de genes de oligonucleótidos de cadena sencilla complementarios y una posterior reacción de hibridación. Las cadenas sencillas se sintetizan con una diferencia de longitud, resultando en una saliente pegajosa del adaptador hibridado. Esta saliente corresponde a las secuencias complementarias del corte de los sitios de restricción de Sfil, que se generan cuando se extirpa el PLOI con flycode del vector de expresión NestLink. Por lo tanto, los oligonucleótidos hibridados se pueden ligar con alta eficiencia al PLOI con flycode para sujetar las secuencias adaptadoras requeridas para la secuenciación profunda con Illumina MiSeq. El producto de ligadura se purifica a través de gel de agarosa antes de la secuenciación profunda.

45 Aquí hay una descripción de la secuencia de la plantilla de secuenciación profunda final de la parte interna (el inserto) hacia las regiones exteriores:

PLOI con Flycode: El PLOI con flycode se extirpa del vector de expresión NestLink a través de digestión de restricción con Sfil.

50 **Resto de los sitios de restricción de Sfil:** Esta enzima permite la extracción del vector de expresión NestLink y los extremos pegajosos generados son usados para sujetar los adaptadores de secuenciación profunda específicamente en el sitio.

55 **Diversidad:** La tecnología Illumina MiSeq genera la primera señal de secuenciación con base en la siguiente secuencia al sitio de enlace del cebador. Las primeras bases deben ser diversas (no idénticas) para impedir la sobrecarga de señal de algunos canales de detección y que se aborte la corrida de la secuenciación.

Sitio de enlace del cebador: El cebador de secuenciación se une aquí.

60 **Índice** (marcado con los números 501 y 701): La tecnología Illumina MiSeq permite multiplexar, es decir, se pueden analizar muchas muestras en una corrida de secuenciación. Para determinar qué lectura pertenece a qué muestra se lee también un índice (extensión variable de 8 pb). Con el fin de permitir secuenciar muchos experimentos NestLink en una sola corrida de secuenciación, los inventores generaron 7 adaptadores de secuenciación profunda (3 para un extremo y 4 para el otro extremo), que permite la generación de 12 pares de índices diferentes.

Adaptador: Este se usa para inmovilizar la plantilla de ADN para la secuenciación profunda en la celda de flujo de Illumina MiSeq.

Cuantificación de miembros PLOI a través de Flycodes

Muchas aplicaciones NestLink requieren la absoluta cuantificación de los miembros PLOI con flycode. Aunque la LC-MS es inexacta en la cuantificación de péptidos individuales en proteómica, los beneficios de NestLink de múltiples Flycodes sujetos a cada miembro PLOI y de un banco Flycode homogéneo, que está diseñado para la detección óptima por espectrometría de masas. Con base en esta consideración, los inventores hicieron la hipótesis de que las intensidades de MS1 sumadas de todos los Flycodes de cualquier miembro PLOI dado debe ser proporcional a la cantidad de este miembro PLOI en una muestra. Los inventores probaron esta hipótesis al agregar cantidades conocidas de ocho sicuerpos unidos a números variables de Flycodes en dos muestras que contenían lisados de *E. coli* y de *M. smegmatis*, respectivamente (figura 3). La relación lineal observada entre las intensidades de MS1 sumadas de todos los Flycodes de cada sicuerpo con flycode y su cantidad agregada proveyó la corrección de la hipótesis y demostró que el procedimiento NestLink descrito en este documento, se puede usar para cuantificar miembros PLOI individuales dentro de un grupo. La cantidad absoluta de miembros PLOI individuales se puede determinar si se agregan cantidades de una o más proteínas con flycode (estándares) en una muestra antes del aislamiento flycode para LC-MS.

Fagémido para las selecciones de presentación de fago (antes de NestLink)

En la mayoría de las aplicaciones actuales de los inventores, el PLOI es un grupo de nanocuerpos sintéticos enriquecidos, llamados sicuerpos. Usualmente, un banco de sicuerpos grandes se enriquece usando presentación de fagos para unirse a una proteína diana. Con el fin de evitar la recombinación de regiones no homólogas (es decir, las CDRs), el PLOI no debe ser amplificado por PCR después de las selecciones de presentación de fago. Para este fin, el vector fagémido (figura 2A) se construyó de forma que el PLOI se puede subclonar a través de sitios de restricción de SapI en el vector de expresión NestLink. Es de hacerse notar que los sitios de SapI son parte del producto de traducción, que se presenta en la superficie del fago. Los inventores pueden mostrar experimentalmente que estos aminoácidos adicionales derivados de los sitios de SapI no interfieren con la eficiencia de presentación del fago.

Además de los sitios SapI, el vector de presentación de fago contiene todos los elementos usualmente presentes en los fagémidos usados para presentar proteínas en el fago M13 y es un derivado del vector pMESy4 (banco de genes KF415192).

Una nota general adicional importante para todos los vectores se describe aquí: Con el fin de transferir la eficiencia de un inserto de un vector en otro, es crucial que los vectores porten diferentes resistencias a antibióticos. Por lo tanto, el vector de expresión NestLink porta un marcador de resistencia a cloranfenicol y los vectores de secuenciación profunda un marcador a kanamicina. Además, el fagémido para las selecciones de presentación de fago contienen un marcador de resistencia a ampicilina.

Prueba del experimento concepto

En este experimento, los inventores demostraron que NestLink se puede usar para caracterizar proteínas individuales dentro de un grupo grande de candidatos proteínas de una forma sin precedentes y que se pueden identificar los miembros del grupo con las características más prometedoras para aplicaciones cuesta abajo de selección.

Más específicamente, la prueba del experimento concepto que se describe a continuación demuestra que i) se desarrolló un método eficiente para el anidado de bancos en diversidades de bancos bien controlados y ii) que los bancos anidados pueden servir como base para presiones de selección sin precedentes en grupos de enlazadores.

En este ejemplo, los inventores trabajaron con un PLOI que consiste en un grupo de sicuerpos que se enriquecieron previamente a través de ribosomas y presentación de fagos (no se describen) contra la proteína de enlace a maltosa (MBP [por sus siglas en inglés])

Los inventores usaron el método NestLink descrito en esta patente para imponer las siguientes presiones de selección en un grupo diverso de sicuerpos a la vez: i) selección de los sicuerpos de mayor expresión, ii) selección de los sicuerpos con la mayor solubilidad, y iii) selección de los sicuerpos que se unen a la diana en un ensayo de enlace en solución.

Usando el protocolo descrito en la sección de materiales y métodos, los inventores pretenden unir cerca de 1200 miembros del grupo de sicuerpos distintos a aproximadamente 17'000 Flycodes únicos, resultando en un llamado "banco anidado". Esto se llevó a cabo al cultivar primero un número adecuado de clones de células que contenían fagémidos que codifican sicuerpos en un recipiente, seguido de aislamiento de su ADN de plásmido. En vez de recoger individualmente los clones del sicuerpo, se estimó el número de unidades formadoras de colonias (cfu [por sus siglas en inglés]) por volumen de bacterias recuperadas después de la transformación a través de su depósito en placas de agar. De ahí, un volumen adecuado de bacterias recuperadas (aproximadamente 1'200 cfu) se usó para inocular un cultivo que se cosechó posteriormente para el aislamiento del ADN del plásmido. Los insertos de ADN de estos fagémidos de diversidad restringida se ligaron al vector de expresión pNLx que contenía el banco Flycode de aproximadamente 10⁸ variantes

distintas. Usando estimaciones de cfu, como se señaló anteriormente, el número de clones se restringió a aproximadamente 17'000. Dado que solo aproximadamente 17'000 vectores que contenían Flycode (según se determinó por estimación de cfu) se usaron 10^8 variantes, los inventores calcularon que 99,974 % de los Flycodes son únicos, y, por lo tanto, la vasta mayoría de los Flycodes están etiquetando un sicuerpo único. Además, dado que anidaron cerca de 1'200 genes de sicuerpo dentro de cerca de 17000 vectores que contienen Flycodes, se esperaba que el sicuerpo promedio estuviese etiquetado con 14 diferentes Flycodes.

El banco anidado en el vector pNLx se expresó en bacterias en un solo matraz y se purificó como un grupo enlazador con flycode para realizar los experimentos de selección (véase a continuación). Para secuenciar el banco anidado, se transfirieron los sicuerpos codificados con flycode al vector de secuenciación profunda pNL que alberga todas las secuencias pertinentes para la secuenciación profunda de Illumina usando un dispositivo MiSeq. La secuenciación profunda del banco anidado proporcionó una asignación sin ambigüedad de cada Flycode a su sicuerpo correspondiente. Los datos de secuenciación profunda estuvieron en concordancia con los sicuerpos esperados y el número de Flycode dentro del banco anidado, una vez que se obtuvieron 1080 secuencias de sicuerpos distintos, unidos a 13'620 Flycodes únicos después del filtrado de los datos. Por lo tanto, en promedio cada sicuerpo se unió 12,61 veces a uno diferente y al Flycode. Los inventores no observaron uniones Flycode ambiguas a los sicuerpos después del filtrado de los datos de secuenciación (es decir, el mismo Flycode sujetos a dos o más sicuerpos diferentes). Este intento exitoso de anidar bancos unos dentro de otros usando diversidades bien controladas es sin precedentes de acuerdo con el conocimiento de los inventores.

Usando los datos de secuenciación profunda, una base de datos que alberga la información de secuencia entera del banco anidado se construyó por concatenación de todos los Flycodes de cada sicuerpo en una secuencia de proteínas continuas con el correspondiente sicuerpo como un identificador. Esta base de datos se cargó después en un servidor Mascot, para su uso posterior en búsquedas de iones con MS/MS

Como ejemplo de una aplicación novedosa de esta tecnología, los inventores usaron la biblioteca anidada y seleccionaron e identificaron específicamente aquellos sicuerpos con un cierto radio hidrodinámico aparente y aquellos que exhibieron una interacción de alta afinidad con MBP en solución. Ambas características se determinaron mediante cromatografía de exclusión de tamaño (SEC [por sus siglas en inglés]) y no se pueden utilizar con los sistemas de visualización del estado de la técnica que requieren un enlace genotipo-fenotipo, porque el genotipo aumenta el tamaño de la proteína visualizada generalmente en más de 100 veces, lo que hace que la partícula de visualización sea insensible a las pequeñas diferencias de tamaño a nivel de proteína.

Para este fin, el banco anidado se expresó y los enlazadores con flycode se purificaron a través de resina Ni-NTA y se sometieron a SEC. Las fracciones eluidas de sicuerpos que corresponden a proteínas monoméricas (los candidatos a enlazador con la solubilidad más alta) se agruparon y se dividieron en dos alícuotas equivalentes. Una alícuota se incubó con MBP y la otra solo con solución amortiguadora. Las dos muestras se analizaron por separado en SEC (la corrida sin MBP se usó como un control) y se recolectaron las fracciones de elución que corresponden al tamaño del complejo sicuerpo-MBP. Los Flycodes de las fracciones recolectadas de la MBP y de la corrida de control fueron aisladas posteriormente y se sometieron a dos corridas LC-MS por separado o combinadas en una corrida LC-MS/MS después del marcaje de etiqueta isobárico de los Flycodes aislados. La base de datos de secuenciación profunda generada anteriormente (asignaciones Flycode a sicuerpo) se pueden usar después para identificar los Flycodes en una búsqueda en Mascot, de ese modo, se identifica sin ambigüedad a sicuerpos que eluyen al mismo tamaño del complejo sicuerpo-MBP. Este experimento permitió a los inventores identificar más de 300 sicuerpos únicos, que están todos bien expresados, son monoméricos y se unen a la proteína diana en solución.

Aplicación de NestLink para la determinación de la constante de disociación

Con el fin de puntuar sicuerpos específicos a MBP, que fueron identificados en la prueba del experimento principio descrito anteriormente, de acuerdo con sus constantes de disociación de enlace, los inventores inmovilizaron cantidades iguales de los complejos MBP-sicuerpos aislados a través de la MBP biotilada en dos columnas de estreptavidina-sefarosa (figura 4). Una selección de la constante de disociación con exceso de MBP no biotilada (lavado durante 3 min) se realizó en una columna, mientras que la otra columna se lavó solo con solución amortiguadora. Después de las lavadas, los sicuerpos restantes de ambas columnas fueron eluidos y sus Flycodes se sometieron a las dos corridas de LC-MS/MS. De modo parecido al experimento de enlace en solución descrito anteriormente (corridas de SEC), la base de datos de secuenciación profunda se usó en búsquedas en Mascot para la identificación de los sicuerpos a través de los Flycodes. Además, las intensidades de MS1 para todos los Flycodes identificados se sumaron para cada sicuerpo usando el software Progenesis. Debido a la naturaleza cuantitativa de las intensidades pico de MS1 según se determinó anteriormente, los inventores esperaban que la razón entre las sumas flycode-intensidad para cada sicuerpo entre las dos columnas corresponderían a sus concentraciones relativas antes y después de la selección de constante de disociación con exceso de diana biotilada. Asumiendo que cada reacción de disociación sigue a una sola decaída exponencial y usando el conocimiento acerca del tiempo de lavado con diana en exceso (3 min), por lo tanto, los autores pudieron determinar las constantes de disociación aproximadas para más de 300 enlazadores al mismo tiempo. Este análisis se confirmó al medir las constantes de disociación de 11 enlazadores individuales usando resonancia de plasmones de superficie. Determinar las constantes de disociación dentro de un grupo de candidatos de enlazador en un solo experimento es sin precedentes de acuerdo con el conocimiento del autor. Un proceso que requería anteriormente de

muchas semanas debido a la necesidad del procesamiento individual de proteínas se puede realizar ahora al mismo tiempo usando la tecnología descrita en este documento.

Aplicación de NestLink para la identificación de enlazador de camélidos inmunizados

NestLink se aplicó a un grupo de nanocuerpos naturales, que se obtuvo a través de aislamiento de ADNc de células B de una alpaca inmunizada (camélido). El antígeno usado para la inmunización fue TM287/288, un transportador ABC (proteína integral de membrana) de *Thermotoga maritima*. En oposición al protocolo tradicional de la generación de nanocuerpos a partir de camélidos, este grupo de nanocuerpos *no* se enriqueció contra la diana usando presentación de fago.

Los nanocuerpos se amplificaron por PCR, la diversidad se restringió y se entrelazaron con el banco Flycode para producir 3'469 secuencias de nanocuerpos únicos sujetos a 59'974 Flycodes únicos según se determinó por secuenciación profunda con Illumina MiSeq (véase la sección de materiales y métodos). El banco anidado se expresó y se purificó a través de Ni-NTA, seguido de aislamiento de los miembros monoméricos agrupados a través de SEC. De forma análoga a la prueba del experimento principio (descrito anteriormente), los candidatos enlazadores no favorables, que no expresaron o que no fueron solubles, se eliminaron en estas etapas de preselección. La muestra de LC-MS se recolectó después de la elución de la columna de Ni-NTA y de la fracción monomérica de la corrida de SEC. Posteriormente, se incubaron cantidades crecientes del grupo con TM287/288 a razones de aproximadamente 0,1:1, 2:1 y 100:1 y las mezclas de antígeno/grupo se sometieron de nuevo a tres corridas de SEC (figura 5). Las fracciones correspondientes al tamaño del complejo diana/nanocuerpo se recolectaron. Los Flycodes de todas las muestras recolectadas se aislaron por separado y se analizaron por LC-MS/MS, lo que permitió la comparación de los niveles de expresión, solubilidad (monomérica en SEC) y de fuerza de enlace al antígeno en solución para todos los enlazadores a la vez.

En este análisis de 3'469 nanocuerpos únicos de un camélido inmunizado, los inventores identificaron 27 familias enlazadoras de alta afinidad con estabilidad favorable, niveles de expresión y solubilidad. De manera notable, NestLink fue mucho más eficiente que las selecciones de presentación por fago y las detecciones excesivamente convencionales, usando ELISA y secuenciación Sanger, que identificaron solo 5 de estas familias en el mismo grupo dentro de un tiempo de procesamiento significativamente más largo. En resumen, se puede establecer, por lo tanto, que NestLink se puede usar para identificar las biomoléculas candidatas más prometedoras de camélidos inmunizados, con un rendimiento y una exactitud que no satisfacen los procedimientos del estado de la técnica actuales.

Aplicación de NestLink para identificar enlazadores que se dirigen a una proteína en la superficie celular

Los experimentos descritos anteriormente se realizaron con el objetivo de identificar proteínas de enlace contra dianas/antígenos purificados en solución, que produjeron herramientas de búsqueda favorable para aplicaciones *in vitro*, tales como cristalografía. Aquí, los inventores pretendieron resolver un cuello de botella nuclear del desarrollo de fármacos, que es la identificación de enlazadores de proteína de membrana que reconocen la proteína diana con alta especificidad y afinidad en la superficie celular. El desarrollo de fármacos biomoleculares contra una proteína de membrana diana requiere usualmente dos etapas consecutivas, que son fundamentalmente diferentes. Primero, un grupo diverso de candidatos enlazadores se genera a través de procedimientos de presentación o de inmunización. Segundo, el grupo diverso se detecta para enlace y función en ensayos celulares. Este último es inherentemente ineficiente y lento debido a que requiere de análisis uno por uno de candidatos individuales a enlazador (usualmente en formato miniatura). En este experimento, los inventores reemplazaron la segunda etapa (detección) por NestLink, con el fin de identificar enlazadores de superficie celular, específicos contra una proteína de membrana integral diana sin el análisis laborioso uno por uno de candidatos individuales a enlazador.

Los inventores realizaron primero la presentación *in vitro* de un banco sicuerpo contra un antígeno a proteína de membrana externa, puro y solubilizado en detergente de una bacteria Gramnegativa (etapa 1, generación de un grupo diverso de candidatos a enlazador). En lugar de probar cada candidato individual a enlazante de este grupo diverso de forma individual para el enlace a superficie celular (normalmente el paso 2), los inventores realizaron NestLink y probaron un grupo grande de candidatos a la vez (figura 6A). 1'456 sicuerpos se entrelazaron con el banco flycode, resultando en la unión de 31'500 Flycodes (en promedio 22 Flycodes/sicuerpo) Como se describió anteriormente, la asignación de flycode a enlazador se obtuvo a través de secuenciación profunda y el banco anidado se expresó, purificó y los miembros del grupo monomérico se aislaron (contra selección/eliminación de candidatos a enlazador no deseados). De ahí, los miembros del grupo con pobres niveles de expresión y con pobre solubilidad se eliminaron primero y se monitorearon los niveles de expresión y las características de solubilidad de cada miembro del grupo. De este modo, el proceso NestLink canalizó exclusivamente los candidatos a enlazador prometedores en la selección de superficie celular, que se realizó de la siguiente manera: los miembros del grupo monomérico se dividieron en 4 fracciones equivalentes y cada fracción se incubó con otra cepa bacteriana. Los candidatos de sicuerpo no enlazantes se eliminaron mediante formación de pellas y volver a suspender/lavado usando solución amortiguadora. Posteriormente, todos los Flycodes de los sicuerpos que se unieron a una de las cepas bacterianas se aislaron y se sometieron a análisis LC-MS. La suma de todas las intensidades de MS1 de todos los Flycodes por sicuerpo se usaron como una medición para la concentración relativa de cada sicuerpo individual en el grupo en cada una de las células diana. Esto permitió una lectura exacta de célula-especificidad (figura 6B).

De los 1'456 candidatos a enlazador en el grupo, se identificaron 6 sicuerpos bien expresados y solubles, que reconocieron específicamente la proteína diana en su forma natural incorporada en la membrana externa de las bacterias Gramnegativas de interés (cepa 4). Los inventores confirmaron este hallazgo al analizar individualmente estos 6 sicuerpos identificados por citometría de flujo contra las 4 cepas (después de marcarlas por fluorescencia). Todos los candidatos probados mostraron el mismo perfil de especificidad en este ensayo de un solo clon, como se observó a través de NestLink. Es de hacerse notar que cada uno de los enlazadores identificados estuvo presente solo al < 0,05 % en el grupo anidado, según se determinó por secuenciación profunda con Illumina MiSeq. Considerando que las detecciones del estado de la técnica solo toman en consideración una característica de un candidato a enlazador (por ejemplo, enlace a la diana), pero se quedan cortas para reportar los niveles de expresión o las propensiones de solubilidad/oligomerización, es probable que cualquiera de estos seis enlazadores pudiera haber sido identificado por un enfoque clásico de detección de un solo clon. De ahí, este experimento demuestra que NestLink permite a bancos de enlazadores detectados a una profundidad sin precedentes gracias a la ausencia de una unión genotipo-fenotipo y al intercalado de dos bancos.

15 *Aplicación de NestLink para monitorear la biodistribución y los parámetros de farmacocinética en organismos modelo*

En los ejemplos anteriores, los inventores mostraron que la selección NestLink permite presiones de selección sin precedentes, debido a la ausencia de la unión genotipo-fenotipo (por ejemplo, la selección de grupos monoméricos/miembros de banco en SEC). Aquí, se introduce otra presión de selección, que no se puede lograr en el caso de unión física genotipo-fenotipo: la selección de proteínas con biodistribución particular y las propiedades farmacocinéticas en organismos vivos. Un grupo anidado (etiquetado-flycode) de candidatos terapéuticos biomoleculares se puede inyectar en un modelo animal y la concentración relativa de cada miembro del grupo se puede medir después del transcurso de cierto tiempo en cada posición en el cuerpo (por ejemplo, en diferentes órganos, tejidos o tumores, etc.) por LC-MS. Un análisis de este tipo resultaría en un análisis de biodistribución exhaustivo/global para cada miembro del grupo individual dentro del cuerpo en un momento de evaluación específico. Si muchos individuos análogos de la misma especie fueran sometidos a este análisis después de diversos momentos de evaluación, el análisis de biodistribución de NestLink se podría extender en la dimensión temporal, de este modo, permitiendo la obtención de datos farmacocinéticos a baja o mediana resolución temporal para cada candidato.

Los inventores configuraron la base de este tipo de análisis al probar y optimizar los procedimientos de extracción Flycode de tejidos homogenizados de ratones a los que se agregaron anteriormente diferentes cantidades de sicuerpos con flycode. Con mayor detalle, muchos sicuerpos se unieron a una pequeña cantidad de Flycodes (20 - 30) y la asignación sicuerpo-a-flycode se determinó por secuenciación profunda con Illumina MiSeq. Los sicuerpos etiquetados-Flycode se expresaron después y se purificaron de forma individual y su concentración se determinó por mediciones de absorbancia. Los sicuerpos individuales se combinaron después a diferentes concentraciones (que abarcaron un orden de magnitud).

En paralelo, los órganos congelados (hígado, pulmón, riñón) y la sangre de los ratones se descongelaron y homogeneizaron usando condiciones de solución amortiguadora desnaturalizada y un crisol. La mezcla de valoración preparada anteriormente se agregó en los homogenados y se incubó durante 30 min a temperatura ambiente para permitir que actúen proteasas potenciales o enzimas modificadoras de flycode. Posteriormente los sicuerpos junto con los Flycodes restantes fueron extraídos, los Flycodes fueron aislados por escisión de proteasa y se analizaron a través de LC-MS. Con base en la detección de sicuerpos individuales de la mezcla de valoración, los inventores determinaron que la detección de sicuerpos a través de LC-MS de tales órganos y tejidos homogenizados son usualmente confiables a la cantidad de 30 - 100 ng (sicuerpos). Dado que se puede inyectar usualmente hasta 1 mg de agentes terapéuticos en un modelo de ratón, es claro que, en las posiciones más importantes en el cuerpo, estarán presentes docenas de microgramos después de la inyección de un grupo anidado. Así, suficientes Flycodes no degradados y no modificados están presentes para monitorear las biodistribuciones globales y para conducir análisis farmacocinéticos de un grupo enlazador.

50 *Materiales y métodos*

A continuación, se provee un protocolo general del método NestLink. Este abarca todas las etapas requeridas para llevar a cabo los experimentos como se señaló anteriormente y provee de detalles en relación con anidado de banco, secuenciación profunda, expresión y purificación de grupos enlazadores con flycode, extracción de Flycode, LC-MS y análisis de datos.

55 *Clonado de nanocuerpos con flycode por anidado de banco*

1. Restricción de diversidad del grupo sicuerpo/nanocuerpo
 Los experimentos NestLink se han llevado a cabo con grupos de sicuerpos o de nanocuerpos naturales, que se obtuvieron de selecciones de enlazador *in vitro* por presentación de fago o por inmunizaciones, respectivamente. En el caso de presentación de fago se usó para la selección de enlazador, 200 ng del grupo seleccionado *in vitro* de enlazadores potenciales codificados en un fagémido que se transformó en 50 μ l de células químicamente competentes MC1061 de *E. coli* (competencia lograda por protocolo de Promega Corporation, Subcloning Notebook 2004). Una serie de diluciones se depositaron en placas de agar que contenían 120 μ g/ml de ampicilina y se incubaron durante toda la noche a 30 °C. Las colonias de una placa contenían las unidades formadoras de colonia deseadas (en los ejemplos anteriores, el número oscilaba entre 1000 y 1500 cfu) se volvió a suspender por 2 ml de medio LB que contenía 100 μ g/ml de ampicilina y la

suspensión se transfirió a 200 ml de cultivo de medio LB que contenía 100 µg/ml de ampicilina. Este cultivo creció durante toda la noche a 37 °C y se usó para la preparación del ADN (Kit: #740412.10, MACHEREY-NAGEL). 15 µg del fagémido preparado se digirió por 100 unidades de BspQI (New England Biolabs, # R0712L) en solución amortiguadora NEB 3.1 (New England Biolabs, # B7203S) en un volumen de reacción de 140 µl a 50 °C durante 1 h, seguido de inactivación por calor de la enzima a 80 °C durante 20 min. Se realizó electroforesis en un gel de agarosa al 2 % (p/v) y la banda que correspondía al grupo enlazador se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL). En el caso de alpacas inmunizadas, las secuencias de nanocuerpos se amplificaron de ADNc de células B como se describió (Pardon *et al.*, Nat Protoc., 2014 Mar;9(3):674-93) y se amplificaron con los cebadores que contenían sitios de restricción BspQI. 5 µg del producto de la PCR se digirió por 100 unidades de BspQI (New England Biolabs, # R0712L) en solución amortiguadora NEB 3.1 (New England Biolabs, # B7203S) en un volumen de reacción de 140 µl a 50 °C durante 1 h, seguido de inactivación por calor de la enzima a 80 °C durante 20 min. Se realizó electroforesis en un gel de agarosa al 2 % (p/v) y la banda que correspondía al grupo enlazador se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL). El fragmento de la PCR digerido se clonó en el vector inicial de clonación FX con un marcador de resistencia Kanamicina (Geertsma *et al.*, Biochemistry, 2011 Apr 19;50(15):3272-8) y se volvieron a suspender 3.500 cfu por 2 ml de medio LB que contenía 50 µg/ml de kanamicina y la suspensión se transfirió a un cultivo de 200 ml de medio LB que contenía 50 µg/ml de kanamicina. Este cultivo creció durante toda la noche a 37 °C y se usó para la preparación del ADN (Kit: #740412.10, MACHEREY-NAGEL). 15 µg del fagémido preparado se digirió por 100 unidades de BspQI (New England Biolabs, # R0712L) en solución amortiguadora NEB 3.1 (New England Biolabs, # B7203S) en un volumen de reacción de 140 µl a 50 °C durante 1 h, seguido de inactivación por calor de la enzima a 80 °C durante 20 min. Se realizó electroforesis en un gel de agarosa al 2 % (p/v) y la banda que correspondía al grupo enlazador se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL).

2. Sujeción de los Flycodes a grupos de sicuerpos/nanocuerpos y restricción de la diversidad Flycode
El vector pNLx que contenía el banco Flycode se digirió por BspQ1, como se describió anteriormente para el fagémido y se realizó electroforesis en un gel de agarosa al 1 % (p/v). La banda que corresponde al vector abierto se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL). 200 ng del grupo enlazador fueron ligados a 400 ng del pNLx digerido usando 2,5 unidades de ligasa T4 (Fermentas #EL0011) en solución amortiguadora de ligasa T4 (Fermentas #B69) en un volumen de reacción de 28 µl a 37 °C durante 1h, seguido por inactivación por calor a 65 °C durante 10 min. 25 µl de la reacción de ligadura se usaron para la transformación en 150 µl de células electro-competentes MC1061 de *E. coli* (preparadas de acuerdo con Howard and Kaser 2007, Making and using antibodies, página 170). Las células se recuperaron durante 30 min a 37 °C en medio SOC y se inoculó un cultivo de 200 ml que contenía 25 µg/ml de cloranfenicol con un volumen de las bacterias recuperadas que correspondían al número deseado de unidades formadoras de colonia como se determinó por depósito en placa de una muestra diluida en una placa de agar que contenía 25 µg/ml de cloranfenicol (en los ejemplos anteriores el número de las cfu oscilaba entre 13'000 y 30'000). El cultivo creció durante toda la noche a 37 °C, seguido por la preparación del ADN (Kit: #740412.10, MACHEREY-NAGEL) y la generación de una solución concentrada de glicerol que contenía 1 ml de cultivo de fase estacionaria mezclado con 1 ml de 50 % (v/v) de glicerol.

Secuenciación profunda

1. Sujeción de secuencias adaptadoras Illumina

15 µg de pNLx que contenían los enlazadores con flycode se digirieron por 120 unidades de Sfil (Fermentas # ER1821) en Solución amortiguadora G (Fermentas # BG5) en un volumen de reacción de 140 µl a 50 °C durante 3 h, seguido por el agregado de 12 µl de 0,5 M de EDTA para inactivación enzimática. Se realizó electroforesis en un gel de agarosa al 2 % y la banda correspondiente al grupo enlazador unido a los Flycodes se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL). Para el primer ejemplo con los sicuerpos anti-MBP, el vector pNLs que contenía los adaptadores pertinentes para la secuenciación profunda de ADN a través de Illumina MiSeq con un índice adecuado (en este caso 502 y 703 se usó para indexado dual) se digirió por Sfil, como se describió anteriormente para pNLx, y se realizó electroforesis en un gel de agarosa al 1 %. La banda que corresponde al esqueleto del vector se extirpó y se extrajo (Kit: #740609.250, MACHEREY-NAGEL). 400 ng del grupo enlazador con flycode fueron ligados a 300 ng del pNLx digerido usando 2,5 unidades de ligasa T4 (Fermentas #EL0011) en solución amortiguadora de ligasa T4 (Fermentas #B69) en un volumen de reacción de 28 µl a 37 °C durante 1h, seguido por inactivación por calor a 65 °C durante 10 min. 25 µl de la reacción de ligadura se usaron para la transformación en 250 µl de células electro-competentes MC1061 de *E. coli* (preparadas de acuerdo con Howard and Kaser 2007, Making and using antibodies, página 170). Las células se recuperaron durante 45 min a 37 °C en medio SOC y se inoculó un cultivo de 200 ml que contenía 30 µg/ml de kanamicina con todas las células recuperadas. Se depositó una muestra de prueba en placas de agar selectivo a kanamicina con el fin de confirmar que eficiencia de la ligadura y de la transformación fue suficiente para transferir el banco anidado entero (> 200'000 cfu en total). El cultivo creció durante toda la noche a 37 °C, seguido por la preparación del ADN (Kit: #27106, QUIAGEN). Una digestión de restricción de 1 µg de los pNLs preparados que contenían el grupo enlazador con flycode se realizó usando 5 unidades de BseRI (New England Biolabs, #R0581S) en solución amortiguadora CutSmart (New England Biolabs, #B7204S) en un volumen de reacción total de 20 µl a 37 °C durante 2 h, seguido por inactivación por calor a 80 °C durante 20 min. Es de hacerse notar que en este punto muchos grupos con flycode contra diversas dianas se pueden agrupar (antes de la digestión con BseRI), cada uno colocado en pNLs indexados de forma diferente. El inserto que contenía el grupo enlazador con flycode sujetado a los adaptadores MiSeq se extrajo posteriormente de un gel de agarosa al 1 %.

Para los otros ejemplos provistos anteriormente, 300 - 400 ng de oligonucleótidos hibridados que contenían Sfil sobresalientes pegajosas se mezclaron con 600 ng del grupo enlazador con flycode extirpado de pNLx por Sfil usando 5 unidades de ligasa T4 (Fermentas #EL0011) en solución amortiguadora de ligasa T4 (Fermentas #B69) en un volumen de reacción de 20 μ l a 37 °C durante 1h, seguido por inactivación de calor a 65 °C durante 10 min. El grupo enlazador con flycode sujetado a los adaptadores MiSeq se extrajo posteriormente de un gel de agarosa al 2 % (Kit: #740609.250, MACHERY-NAGEL). Es de hacerse notar en este punto que se pueden agrupar muchos grupos con flycode contra diversas dianas, cada uno conteniendo un par diferente de adaptadores ligados.

2. Determinación de uniones nanocuerpo-Flycode

Se realizó secuenciación profunda en un dispositivo MiSeq de Illumina usando un protocolo de extremos emparejados (MiSeq Reagent Kit v2 (300 ciclos)). En una primera etapa del análisis, las lecturas de extremos emparejados fueron asociadas usando software estándar (Illumina). Para cualquier par de índices, un total de 800'000 - 8 Mio lecturas se obtuvieron, que corresponde a una redundancia de lectura promedio de 25 - 70 (este número iguala el número de lectura total dividido por el total del número flycode esperado para un banco anidado dado). Usando una secuencia de comandos personalizada, las lecturas sin procesar resultantes se filtraron aplicando los siguientes criterios positivos: i) patrón de flanqueo correcto de partes invariables del Flycode, ii) patrón de flanqueo correcto de partes invariables de nanocuerpos, iii) la secuencia no contiene N , iv) la secuencia está dentro del rango de tamaño esperado de posibles fusiones nanocuerpo-Flycode, v) la secuencia de la fusión nanocuerpo-Flycode está en el marco (es decir, se puede dividir por 3) vi) la secuencia no tiene codones de terminación. Después de filtrar, se generó una lista de Flycodes únicos. Los Flycodes que se leyeron al menos cinco veces se consideraron que eran correctos. Para cada Flycode correcto, se generó una secuencia consenso de todas las secuencias de nanocuerpos unidos. El enfoque de secuencia consenso se requirió para corregir errores de secuenciado en la secuencia de nanocuerpo. Una puntuación de consenso se introdujo para monitorear la variabilidad entre las secuencias de nanocuerpo sujetadas al mismo Flycode. La puntuación otorga grandes penalizaciones en caso de que uno o varios nanocuerpos sujetados al Flycode idéntico sean claramente diferentes de los demás, de esa manera se eliminan Flycodes unidos a dos o más nanocuerpos diferentes del análisis adicional. Solo los pares de nanocuerpo-Flycode con puntuación con alto consenso se consideraron en adelante. En una etapa final, las secuencias idénticas (consensos) y todos sus Flycodes unidos (en promedio 12 - 40 Flycodes por nanocuerpos en los ejemplos anteriores) se identificaron. Todos los flycodes conectados al mismo nanocuerpo se concatenaron en una secuencia de proteínas hipotéticas usando la secuencia de nanocuerpo como un identificador y esta base de datos se guardó en formato fasta-file.

Expresión y purificación de sicuerpos/nanocuerpos monoméricos con flycode

Las soluciones concentradas de glicerol con MC1061 *E. coli* que contienen pNLx que albergan el grupo enlazador con flycode se usaron para la inoculación de un precultivo de 50 ml de LB que contenía 1 % de glucosa, que se cultivó durante toda la noche a 37 °C. Se inocularon 600 ml de cultivo TB por el precultivo a un OD de 0,05 y se cultivó por 1,5 h a 37 °C seguido de cultivo durante toda la noche a 20°C. Se llevó a cabo la inducción a OD₆₀₀ de 0,8 por 0,05 % (p/v) de arabinosa. Las células se cosecharon mediante espinela a 5'000 g durante 20 min. El sobrenadante se decantó y las células se volvieron a suspender en 25 ml de 50 mM de Tris-HCl con pH 7,5 (20 °C), 150 mM de NaCl, 15 mM de imidazol con pH 8,0 (20 °C), complementado con una pizca de DNaseI (SIGMA #DN25). Las células se lisaron usando un microfluidificador (Microfluidics #110P) a 206842,5 kPa (30'000 psi) por dos rondas, al mismo tiempo que se enfría en hielo. Los restos de las células se hicieron en pellas a 5'000 g durante 30 min y el sobrenadante se aplicó a 1,5 ml de Ni-NTA en una columna de superflujo (QUIAGEN # 1018142) por flujo de gravedad. La columna se lavó con 30 ml de solución amortiguadora de lavado que contenía 20 mM de Tris-HCl con pH 7,5 (20 °C), 150 mM de NaCl y 30 mM de imidazol con pH 8 (20 °C). La columna se eluyó con 6 ml de 20 mM de Tris-HCl con pH 7,5 (20 °C), 150 mM de NaCl y 300 mM de imidazol con pH 8 (20 °C). Se inyectaron 5 ml de la elución en un HiLoad 16/600 Superdex 200 pg (GE Healthcare Life Sciences #28989335) y la región correspondiente a la fracción monomérica se recolectó y se concentró a un volumen de 1,2 ml a una absorbancia (280 nm) de 2,1 en un Nanodrop 2000c (Thermo Scientific) contra solución amortiguadora para experimentos de selección adicionales como se señaló en los ejemplos anteriores.

Aislamiento de Flycodes

Muestras que contenían PLOI con Flycode se diluyeron 10-20 veces por Solución Amortiguadora Ex (20 mM de Tris-HCl a pH 8,5, 150 mM de NaCl, 0,5 % (v/v) Tritón X-100, 0,125 % (p/v) de sodio desoxicolato, 10 mM de imidazol a pH 8,0, 4,5 M de GdmCl), se filtró (filtro de jeringa 0,2 μ m de corte) y se incubaron con 100 μ l de suspensión acuosa de superflujo Ni-NTA (QUIAGEN # 1018142) durante 2 h a temperatura ambiente. La resina se hizo pellas posteriormente a 500 g por 10 min y se transfirieron a una columna cromatográfica mini bio-spin, seguido por lavados 3 \times 500 μ l usando solución amortiguadora Ex, 3 \times 500 μ l usando solución amortiguadora TH (20 mM TEAB pH 8,0, 150 mM NaCl, 2,5 mM CaCl₂) que contenía 30 mM de imidazol pH 8,0 y 3 \times 500 μ l de solución amortiguadora TH. Después de cerrar el extremo inferior de la columna, la resina se volvió a suspender en 100 μ l de solución amortiguadora TH que contenía 2,4 U de Trombina (MILLIPORE #69671-3) seguido de incubación durante toda la noche a temperatura ambiente. La columna se drenó después y se lavó por 3 \times 500 μ l de solución amortiguadora TH que contenía 30 mM de imidazol con pH 8,0, seguido por 3 \times 500 μ l de solución amortiguadora TRY (20 mM TEAB pH 8,0, 50 mM de NaCl, 2,5 mM de CaCl₂) y con elución

por de solución amortiguadora TRY que contenía 300 mM de imidazol pH 8,0. El eluato se centrifugó (15'000 g) a través de un concentrador de corte Microcon de 10 kDa equilibrado antes (H₂O) (AMICON: YM-10) y 1 µg de tripsina (PROMEGA #V5113) se agregó al filtrado seguido por incubación durante toda la noche a 37 °C.

5 Las muestras eluidas de Flycode se sometieron posteriormente a un procedimiento de limpieza con ZipTip (MILLIPORE #ZTC18S960). Las ZipTips se lavaron antes con 200 µl de metanol, 200 µl de 60 % (v/v) de acetonitrilo (ACN) y 200 µl de 3 % (v/v) de acetonitrilo que contenía 0,1 % (v/v) de ácido trifluoroacético. Se cargaron 100 µl de la muestra Flycode, seguido de lavado con 200 µl de 3 % (v/v) de acetonitrilo que contenía 0,1 % (v/v) de ácido trifluoroacético y la elución por 10 2 × 40 µl de 60 % (v/v) de acetonitrilo que contenía 0,1 % (v/v) de ácido trifluoroacético. El solvente se evaporó posteriormente (speedvac) y los Flycodes se volvieron a suspender en 15 µl de 3 % (v/v) de acetonitrilo que contenía 0,1 % (v/v) de ácido fórmico.

LC-MS

15 Usando un sistema de HPLC Easy-nLC 1000, se inyectaron 2 µl de la solución Flycode vuelta a suspender en una columna capilar hecha para uso interno empacada con material de fase inversa (ReproSil-Pur 120 C18-AQ, 1,9 µm; dimensión de la columna 150 mm × 0,075 mm). La columna se equilibró con solvente A (0,1 % de ácido fórmico (FA [por sus siglas en inglés]) en agua). Los péptidos se eluyeron con una tasa de flujo de 0,3 µl/min usando el siguiente gradiente: 0 - 60 min; 5 - 20 % B (0,1 % FA en ACN), 60 - 70 min; 20 - 97 % B. Después de 10 min de lavado por 97 % B, la columna se volvió a equilibrar por solvente A durante 5 min. Los espectros de masas de alta precisión se adquirieron con un espectrómetro de masas Orbitrap Fusion (Thermo Scientific) usando el siguiente parámetro: rango de escaneo de 300-1500 m/z, objetivo AGC de 5e5, resolución de 120'000 (a m/z 190), y un tiempo máximo de inyección de 100 ms. Los espectros MS/MS dependientes de los datos se registraron en el modo de velocidad máxima en la trampa de iones lineal usando aislamiento de cuadrupolo (ventana de 1,6 m/z), objetivo AGC de 1e4, tiempo de inyección máximo de 35 ms, fragmentación HCD con energía de colisión del 30 %, un tiempo de ciclo máximo de 3 seg, y se habilitó todo el tiempo paralelizable disponible. Se seleccionaron señales precursoras monoisotópicas para MS/MS con estados de carga entre 2 y 6 y una intensidad de señal mínima de 5e4. La exclusión dinámica se estableció en 25 segundos y una ventana de exclusión de 10 ppm. Después de la recolección de datos se generaron listas usando Proteome Discoverer 1.4 (Thermo Scientific).

30 *Análisis y cuantificación de datos*

Las corridas de LC-MS (una corrida por extracción/muestra de flycode) fueron inspeccionadas anteriormente por el software Xcalibur y los archivos sin procesar de Xcalibur fueron importados y convertidos por Progenesis a archivos mznld. Se usó posteriormente Progenesis para alinear las corridas de LC-MS de interés (puntuación de alineación > 80 %) y se eliminaron los iones peptídicos de cargas +1 y +5 a +20 del análisis. Posteriormente, se exportó desde Progenesis un archivo mgf combinado de todas las corridas de LC-MS alineadas (umbral de rango < 5, recuento de fragmentos de iones > 1'000, eliminación de isótopos y desconvolución de carga) y se cargó en el servidor mascot junto con el flycode determinado con anterioridad a la asignación PLOI-miembro (base de datos de secuenciación profunda en formato fasta-file, véase lo anteriormente mencionado). Las identificaciones de Mascot se importaron directamente al software Scaffold, seguido de la conversión de datos y la exportación del informe de espectro, que luego se importó a Progenesis, permitiendo la asignación de atributos a sus flycodes correspondientes. Usando Progenesis, las intensidades de los atributos se normalizaron típicamente a los estándares agregados y todos los flycodes únicos de cada PLOI-miembro se usaron para la cuantificación. Las intensidades sin procesar y normalizadas se exportaron posteriormente (formato CSV) y se analizaron adicionalmente por Excel.

45 Listado de secuencias

<110> Universidad de Zúrich

50 <120> Detección de proteínas y método de detección

<130> uz282wo

<160> 5

55 <170> PatentIn version 3.5

<210> 1

<211>2

60 <212> PRT

<213> Secuencia artificial

<220>

<223> Secuencia etiqueta

ES 2 926 548 T3

<400> 1
Trp Arg
1
5 <210> 2
<211> 3
<212> PRT
<213> Secuencia artificial
<220>
10 <223> Secuencia etiqueta
<400> 2
Trp Leu Arg
1
15 <210> 3
<211> 4
<212> PRT
<213> Secuencia artificial
<220>
20 <223> Secuencia etiqueta
<400> 3
Trp Gln Ser Arg
1
25 <210> 4
<211> 5
<212> PRT
<213> Secuencia artificial
30 <220>
<223> Secuencia etiqueta
<400> 4
35 Trp Leu Thr Val Arg
1 5
<210> 5
<211> 6
<212> PRT
40 <213> Secuencia artificial
<220>
<223> Secuencia etiqueta
45 <400> 5
Trp Gln Glu Gly Gly Arg
1 5

REIVINDICACIONES

1. Un método para seleccionar un polipéptido de un banco de polipéptidos, que comprende las etapas de
- 5 a. proveer un primer banco de ácidos nucleicos, en donde cada miembro de dicho banco de ácidos nucleicos comprende una secuencia codificante de polipéptidos que codifica un miembro de un primer banco de polipéptidos;
- b. proveer un segundo banco de ácidos nucleicos, en donde dicho segundo banco de ácidos nucleicos comprende una pluralidad de miembros, en donde cada miembro comprende una secuencia codificante de etiqueta que codifica una etiqueta que codifica una etiqueta de detección, en donde dicha etiqueta de detección:
- 10 i. está **caracterizado por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por dicho segundo banco de ácidos nucleicos;
- ii. está **caracterizado por** una masa molecular de entre 200 y 5000 Da, particularmente entre 500 y 2500 Da, más particularmente entre aproximadamente 900 y 2200 Da; y
- 15 iii. comprende un primer elemento separable;
- iv. está **caracterizado por** un valor de hidrofobicidad de entre -27 y 128 según se calculó por el método descrito en este documento;
- c. insertar dicha secuencia codificante de polipéptido comprendida en dicho miembro de dicho primer banco de ácidos nucleicos en un miembro de dicho segundo banco de ácidos nucleicos, de ese modo creando un banco de ácidos nucleicos etiquetados que codifican un banco de polipéptidos etiquetados, en donde cada miembro de dicho banco de polipéptidos etiquetados comprende un polipéptido y una etiqueta de detección separada de dicho polipéptido por dicho primer elemento separable;
- 20 d. obtener una pluralidad de secuencias de ácidos nucleicos de dicho banco de ácidos nucleicos, en donde cada una de dicha pluralidad de secuencias de ácidos nucleicos comprende una secuencia codificante de polipéptido y una secuencia codificante de etiqueta;
- e. predecir un patrón de fragmentación de espectrometría de masas para cada etiqueta de detección codificada por una secuencia codificante de etiqueta obtenida en la etapa d;
- f. expresar dicho banco de polipéptido etiquetado de dicho banco de ácidos nucleicos etiquetados;
- 30 g. seleccionar un miembro de dicho banco de polipéptido etiquetado en una etapa de selección, produciendo un polipéptido seleccionado;
- h. separar dicho primer elemento separable, de ese modo separar dicha etiqueta de detección de dicho polipéptido seleccionado, produciendo una etiqueta de detección aislada;
- i. identificar dicha etiqueta de detección aislada por
- 35 i. registrar un patrón de fragmentación de dicha etiqueta de detección aislada por espectrometría de masas;
- ii. coincidir dicho patrón de fragmentación obtenido en la etapa i con dichos patrones de fragmentación predichos en la etapa e, de ese modo identificar dicha etiqueta de detección aislada;
- 40 j. seleccionar de dicha pluralidad de secuencias de ácidos nucleicos obtenida en la etapa d una secuencia de ácidos nucleicos que comprende una secuencia codificante de etiqueta que codifica dicha etiqueta de detección identificada en la etapa i, de ese modo identificar al miembro de dicho banco de polipéptidos etiquetados asociados con dicha etiqueta de detección identificada en la etapa i.
- 45 2. El método de acuerdo con la reivindicación 1, en donde dicha etiqueta de detección aislada está **caracterizada por** un valor de hidrofobicidad de entre -1 y 70.
3. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde dicha etiqueta de detección aislada comprende un elemento de secuencia I seleccionado de una colección de elementos de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P.
- 50 4. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde dicha etiqueta de detección aislada consiste en
- 55 a. un elemento de secuencia III, en donde dicho elemento de secuencia III es GS;
- b. dicho elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- 60 c. dicho elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR);
- en donde en particular el orden de dichos elementos de secuencia del N-terminal al C-terminal es el elemento de secuencia III, el elemento de secuencia I, elemento de secuencia II.
- 65 5. Una colección de polipéptidos, en donde cada miembro de dicha colección de polipéptidos se asocia con una etiqueta de detección, particularmente al menos uno, más particularmente al menos dos, incluso más particularmente al menos

cinco, incluso más particularmente al menos 10, incluso más particularmente aproximadamente veinte etiquetas de detección, y en donde dicha etiqueta de detección

- 5 a. está **caracterizada por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección de los polipéptidos;
- b. está **caracterizada por** una masa molecular de entre 200 y 5000 Da, particularmente entre 500 y 2500 Da, más particularmente entre aproximadamente 900 y 2200 Da;
- c. está separada de dicho miembro de dicha colección de polipéptidos por un primer elemento separable;
- 10 d. está **caracterizada por** un valor de hidrofobicidad de entre -27 y 128, particularmente entre -1 y 70 según se calculó por el método descrito en este documento;
- e. consiste en 7 a 21 aminoácidos contiguo y comprende solo un aminoácido que tiene una cadena lateral cargada positivamente que está situada en el C-terminal de la etiqueta de detección y se selecciona de arginina y lisina.

15 6. La colección de polipéptidos de acuerdo con cualquiera de las reivindicaciones 5, en donde dicha etiqueta de detección comprende

- a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- 20 b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

7. Una colección de etiquetas de detección que comprende al menos 96, más particularmente al menos 500.000 etiquetas de detección, incluso más particularmente al menos 10^7 etiquetas de detección, incluso más particularmente aproximadamente 10^8 etiquetas de detección, en donde cada etiqueta de detección:

- 25 a. consiste en 7 a 18, más particularmente 11 a 15 aminoácidos; y
- b. está **caracterizada por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección comprendida en dicha colección de etiquetas de detección;
- 30 c. comprende solo un aminoácido que tiene una cadena lateral cargada positivamente que está situada en el C-terminal de la etiqueta de detección y se selecciona de arginina y lisina;
- d. está **caracterizada por** una masa molecular de entre 200 y 5000 Da, particularmente entre 500 y 2500 Da, más particularmente entre 900 y 2200 Da;
- e. está **caracterizada por** un valor de hidrofobicidad de entre -27 y 128, según se calculó por el método descrito en este documento.

35 8. La colección de etiquetas de detección de la reivindicación 7, en donde cada etiqueta de detección consiste en

- a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- 40 b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

9. Una colección de vectores de plásmidos, particularmente al menos 96, más particularmente al menos 500.000, incluso más particularmente al menos 10^7 vectores de plásmidos, incluso más particularmente aproximadamente 10^8 vectores de plásmidos, en donde cada miembro de dicha colección de vectores de plásmidos comprende una secuencia de ácidos nucleicos codificante de etiqueta que codifica una etiqueta de detección, en donde cada etiqueta de detección consiste en 4 a 20, particularmente 7 a 18, más particularmente 11 a 15 aminoácidos y está **caracterizada por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por dicha colección de vectores de plásmidos; y en donde dicha etiqueta de detección codificada está **caracterizada por**

50 un valor de hidrofobicidad entre -27 y 128 según se calculó por el método descrito en este documento, y consiste en 7 a 18 aminoácidos y comprende solo un aminoácido que tenga una cadena lateral cargada positivamente que está situada en el C-terminal de la etiqueta de detección y se selecciona de arginina y lisina.

55 10. La colección de vectores de plásmidos de acuerdo con la reivindicación 9, en donde dicha etiqueta de detección consiste esencialmente en

- a. un elemento de secuencia I, en donde dicho elemento de secuencia I consiste en 5 a 10, particularmente 7 aminoácidos, independientes entre sí seleccionados de A, S, T, N, Q, D, E, V, L, I, F, Y, W, G y P; y
- 60 b. un elemento de secuencia II seleccionado de SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) y SEQ ID NO 05 (WQEGGR).

11. Un método de detección de proteínas que comprende

- a. proveer un banco de ácidos nucleicos que codifica un banco de polipéptidos
- 65 en donde dicho banco de polipéptidos comprende una pluralidad de miembros y cada miembro está asociado con una etiqueta de detección, y en donde dicha etiqueta de detección

- i. está **caracterizada por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por dicho banco de ácidos nucleicos;
- ii. está **caracterizada por** una masa molecular de entre 200 y 5000 Da, particularmente entre 500 y 2500 Da, más particularmente entre aproximadamente 900 y 2200 Da; y
- iii. está separada de dicho miembro de dicha colección de polipéptidos por un primer elemento separable;
- iv. está **caracterizada por** un valor de hidrofobicidad de entre -27 y 128, según se calculó por el método descrito en este documento;
- b. proveer una base de datos que comprende
- i. una pluralidad de ácidos nucleicos y/o secuencias de aminoácidos, en donde dicha pluralidad de secuencias comprende las secuencias de todos los miembros de dicho banco de aminoácidos, y en donde cada una de dichas secuencias comprende una secuencia que especifica un polipéptido y una secuencia que especifica una etiqueta de detección;
- ii. un patrón de fragmentación de espectrometría de masas predicho para cada etiqueta de detección codificada por dicho banco de ácidos nucleicos;
- c. expresar dicho banco de polipéptido de dicho banco de ácidos nucleicos;
- d. seleccionar un miembro de dicho banco de polipéptido en una etapa de selección, produciendo un polipéptido seleccionado;
- e. separar dicho primer elemento separable, de ese modo separar dicha etiqueta de detección de dicho polipéptido seleccionado, produciendo una etiqueta de detección aislada;
- f. identificar dicha etiqueta de detección aislada por
- i. registrar un patrón de fragmentación de dicha etiqueta de detección aislada por espectrometría de masas;
- ii. coincidir dicho patrón de fragmentación obtenido en la etapa i con dichos patrones de fragmentación predichos en dicha base de datos, de ese modo identificar dicha etiqueta de detección aislada;
- g. seleccionar de dicha pluralidad de secuencias comprendida en dicha base de datos una secuencia que especifica dicha etiqueta de detección identificada en la etapa f, de ese modo identificar al miembro de dicho banco de polipéptidos asociados con dicha etiqueta de detección identificada en la etapa f.
12. El método de acuerdo con la reivindicación 11, en donde cada miembro de dicho banco de polipéptidos o en donde cada etiqueta de detección se asocia con una etiqueta de afinidad, en particular una etiqueta de afinidad seleccionada del grupo que comprende una etiqueta-His, una etiqueta-CBP, una etiqueta-CYD, una etiqueta-Strep, una etiqueta-StrepII, una etiqueta-FLAG, una etiqueta-HPC, una etiqueta-GST, una etiqueta-Avi, una etiqueta de biotilación, una etiqueta-Myc, una etiqueta 3xFLAG y una etiqueta-MBP, y/o, en donde dicha etiqueta de afinidad está separada preferentemente de dicha etiqueta de detección por un segundo elemento separable, y dicho elemento separable se separa antes de la etapa f.
13. Un método para asociar un polipéptido con una etiqueta de detección única, que comprende las etapas de
- a. proveer un primer banco de ácidos nucleicos, en donde cada miembro de dicho banco de ácidos nucleicos comprende una secuencia codificante de polipéptidos que codifica un miembro de un primer banco de polipéptidos;
- b. proveer un segundo banco de ácidos nucleicos, en donde cada miembro de dicho segundo banco de ácidos nucleicos comprende una secuencia codificante de etiqueta que codifica una etiqueta de detección, en donde dicha etiqueta de detección:
- i. está **caracterizado por** una secuencia de aminoácidos diferente de la secuencia de aminoácidos de cualquier otra etiqueta de detección codificada por dicho segundo banco de ácidos nucleicos;
- ii. está **caracterizado por** una masa molecular de entre 200 y 5000 Da, particularmente entre 500 y 2500 Da, más particularmente entre aproximadamente 900 y 2200 Da;
- iii. está **caracterizado por** un valor de hidrofobicidad de entre -27 y 128, según se calculó por el método descrito en este documento;
- iv. comprende un primer elemento separable;
- c. insertar dicha secuencia codificante de polipéptido comprendida en dicho miembro de dicho primer banco de ácidos nucleicos en un miembro de dicho segundo banco de ácidos nucleicos, en donde
- i. dicho primer banco de ácidos nucleicos tiene un tamaño de 5 a 100.000, particularmente 100 a 50.000, más particularmente 500 a 5.000 y
- ii. dicho segundo banco de ácidos nucleicos tiene un tamaño de 10^3 a 10^{11} , particularmente 10^5 a 10^{10} , más particularmente 10^6 a 10^9 , incluso más particularmente aproximadamente 10^8 , de ese modo se genera una pluralidad de plásmidos de combinación polipéptido/etiqueta;
- d. seleccionar un subconjunto de dicha pluralidad de plásmidos de combinación polipéptido/etiqueta, generando de ese

modo un banco de ácidos nucleicos etiquetados que codifican un banco de polipéptidos etiquetados.

14. El método de acuerdo con la reivindicación 13, en donde en la etapa d, el subconjunto seleccionado de la pluralidad de plásmidos de combinación polipéptido-etiqueta es de al menos 3x, particularmente al menos 5x, al menos 10x, al menos 15x, al menos 20x o al menos 25, el número de miembros del primer banco de ácidos nucleicos.
- 5

Fig. 1

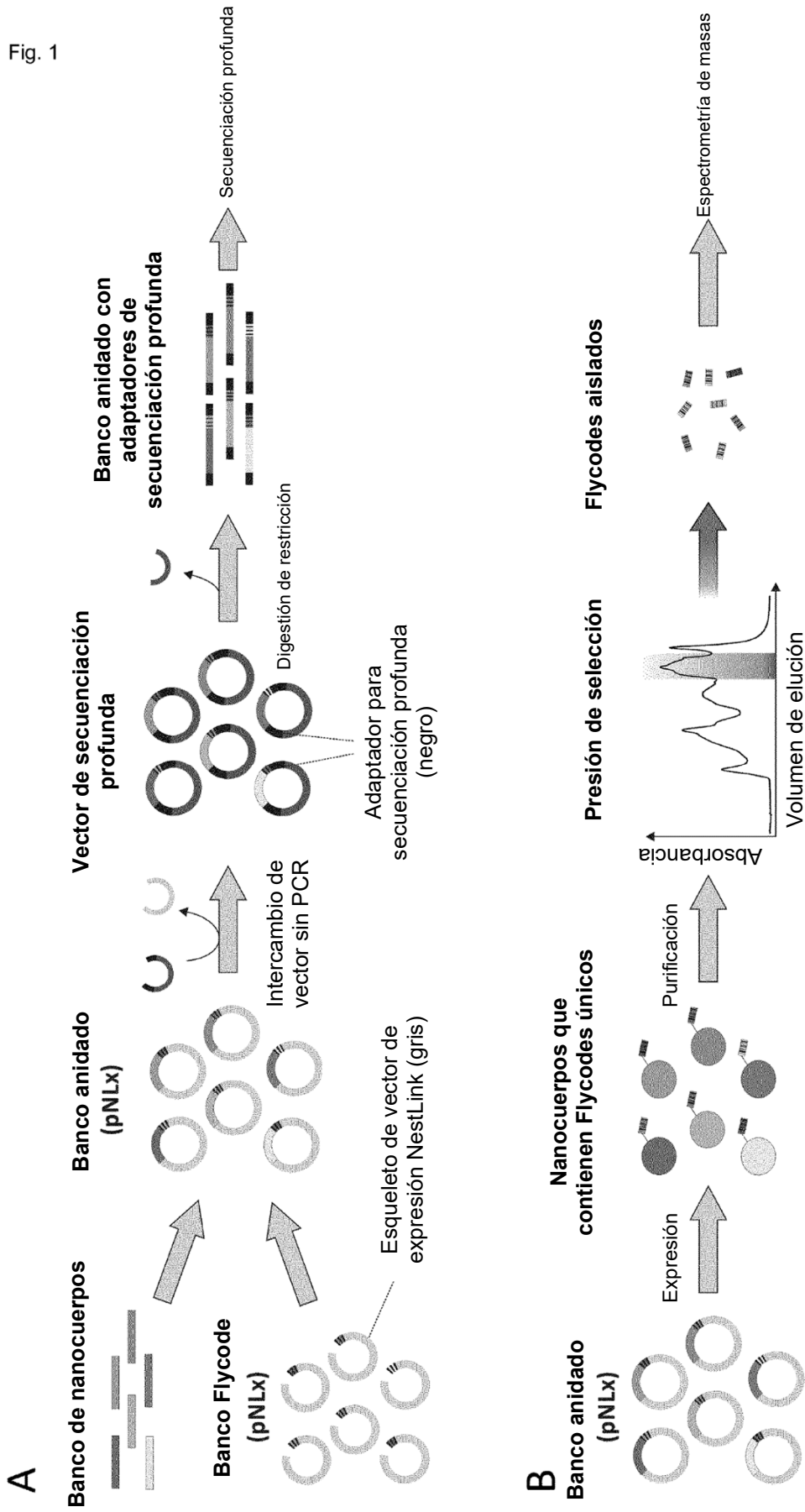


Fig. 1 (continuación)

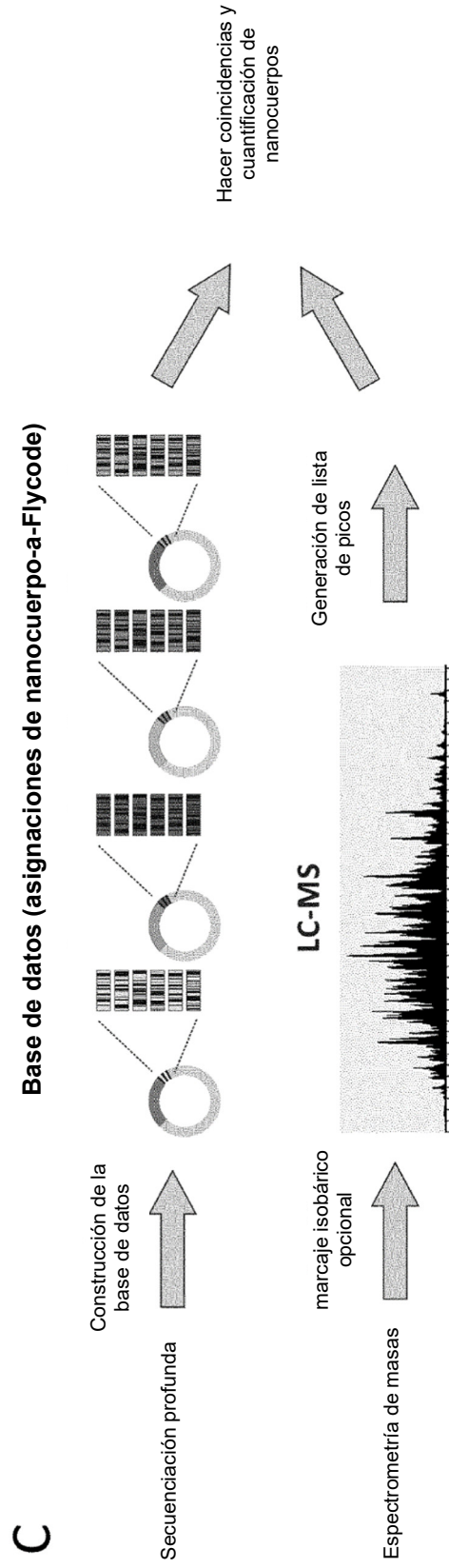


Fig. 2

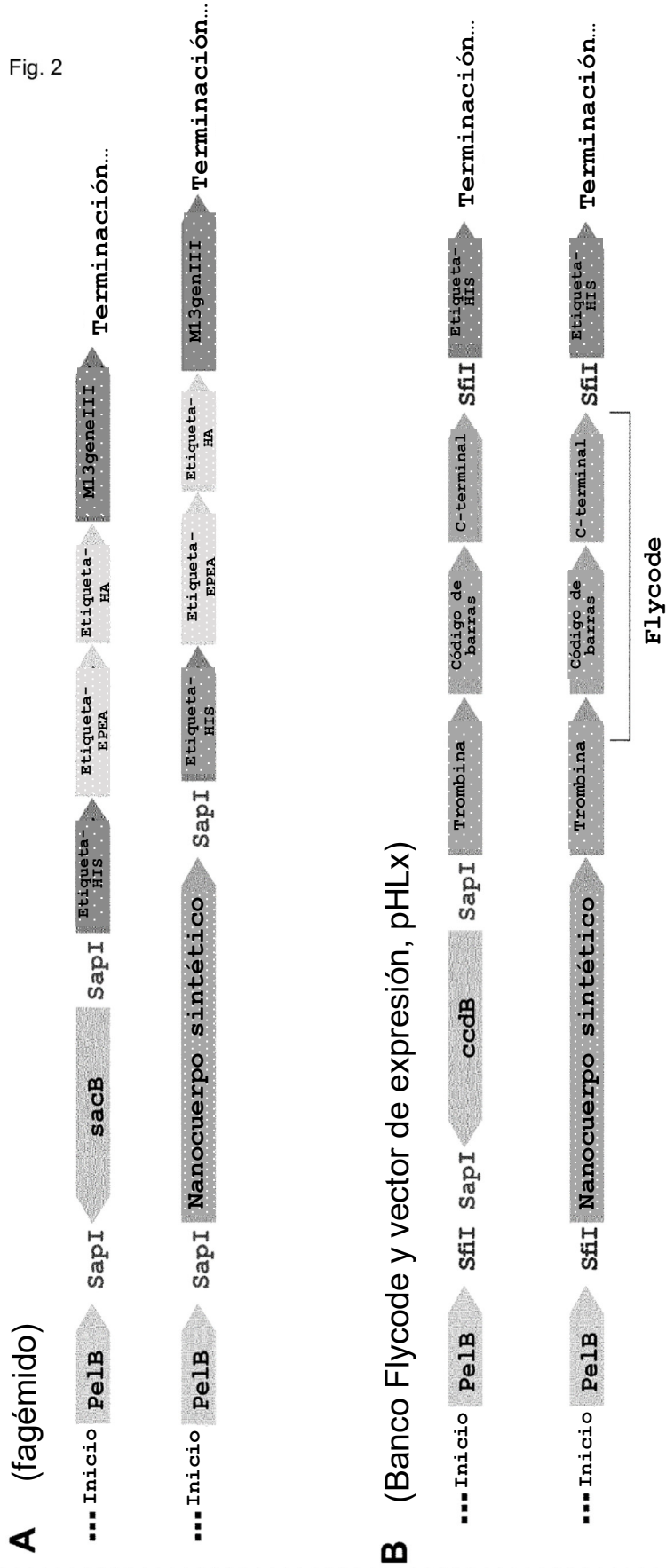


Fig. 3

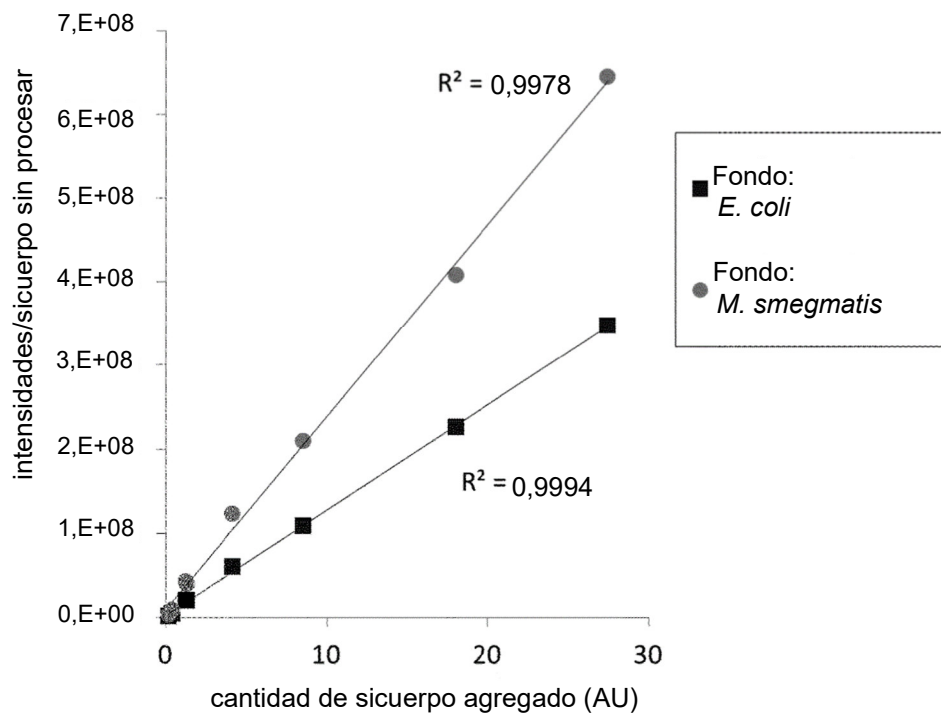
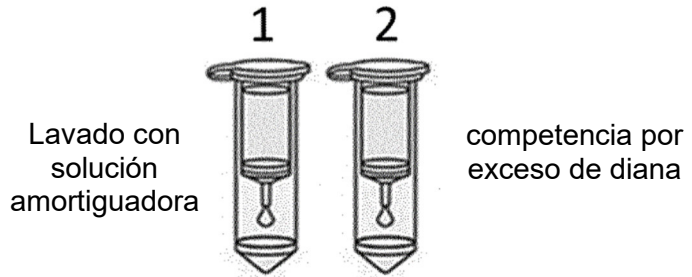


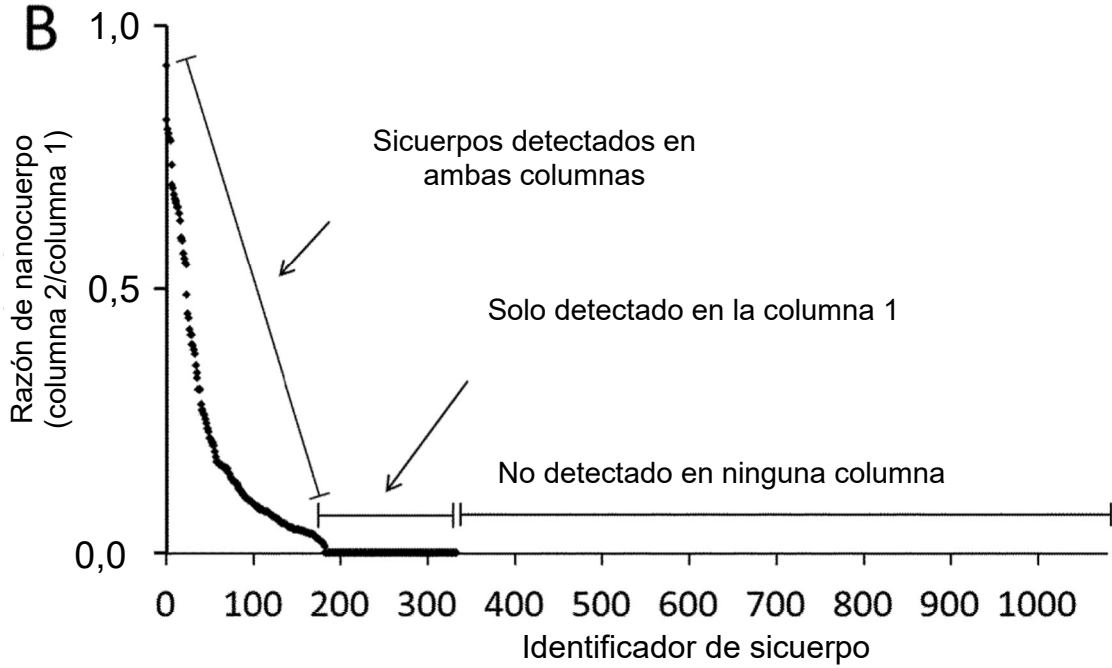
Fig. 4

A

Inmovilización de miembros del grupo en el complejo con diana biotinilada a dos columnas, seguido de selección por constante de disociación



B



C

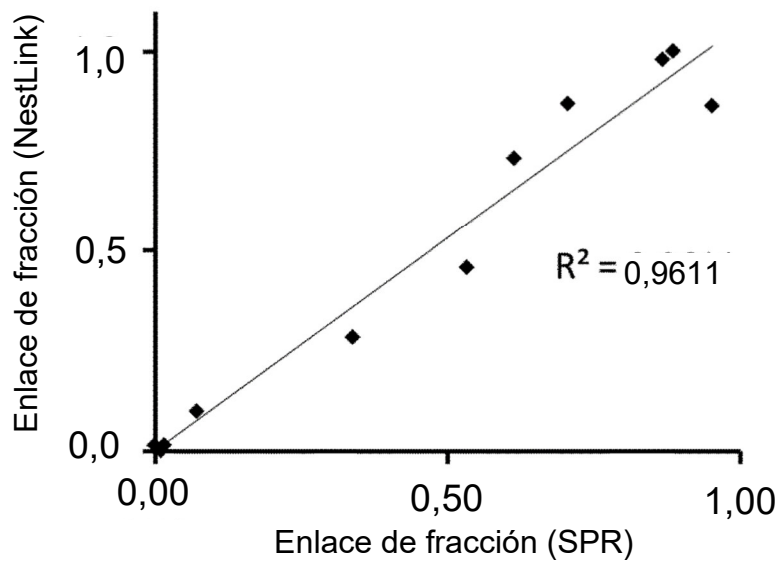
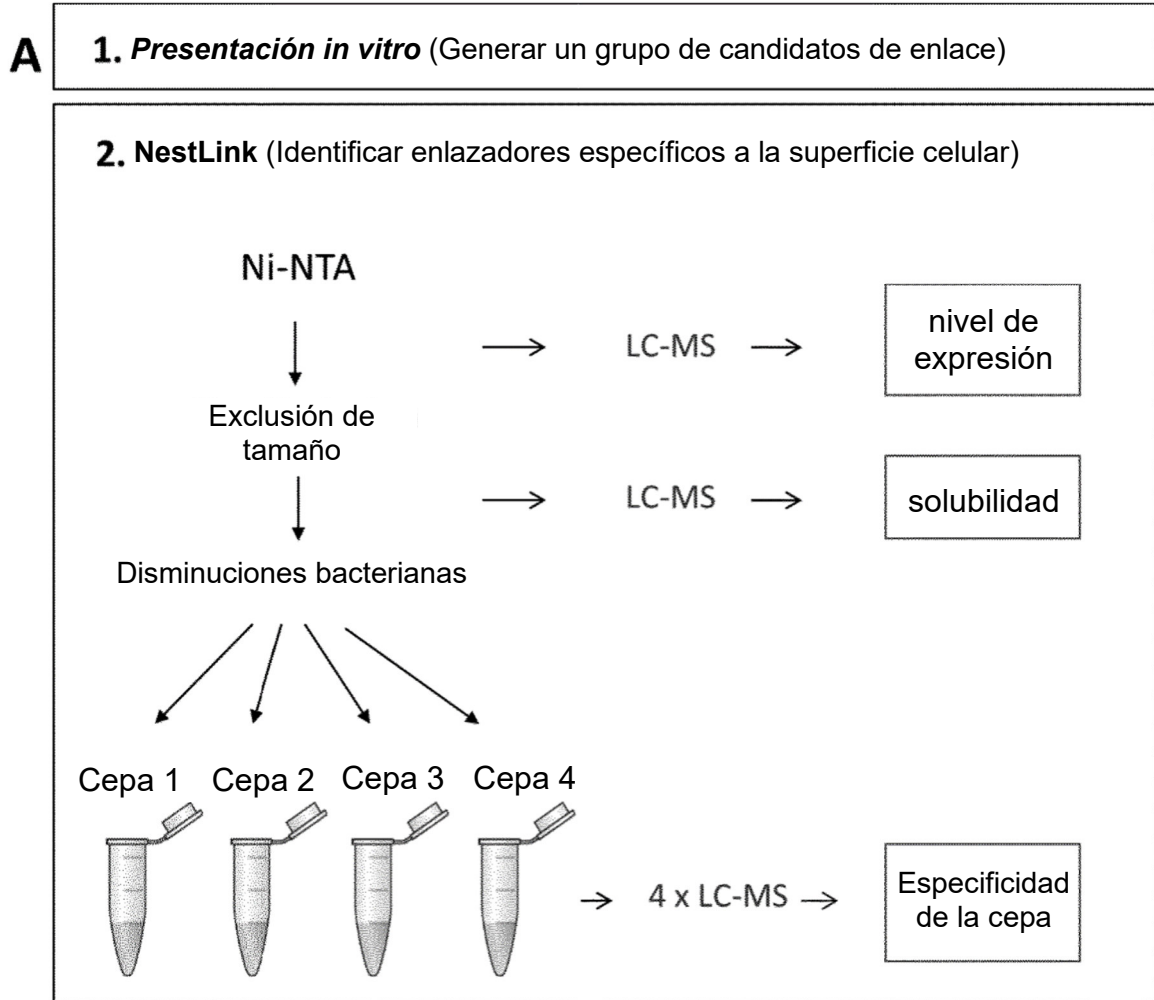


Fig. 5



B Lectura de especificidad a la cepa

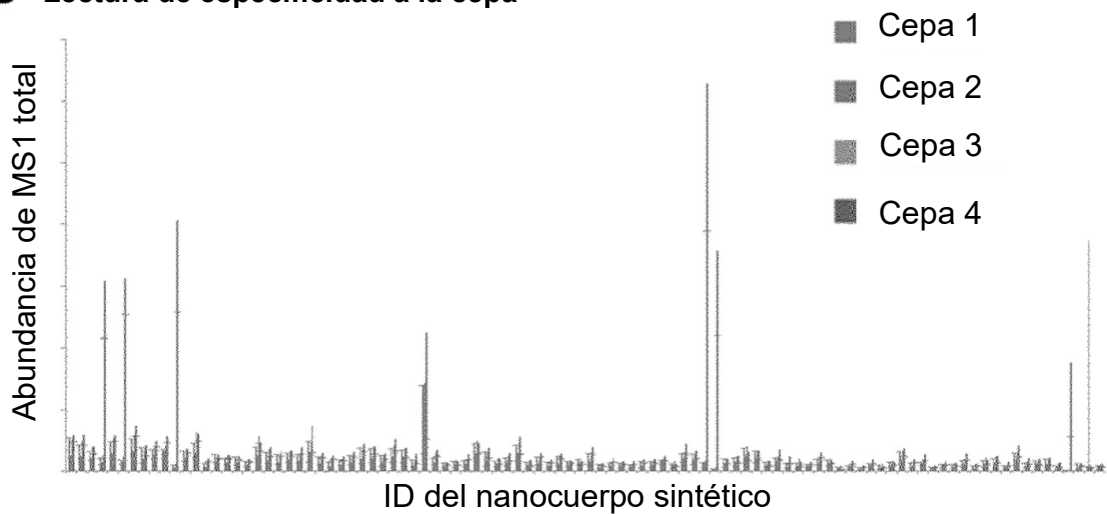


Fig. 6

