



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 265 917**

(51) Int. Cl.:

G01N 33/532 (2006.01)

C12Q 1/37 (2006.01)

C12N 15/10 (2006.01)

C07K 16/00 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **00911077 .6**

(86) Fecha de presentación : **17.03.2000**

(87) Número de publicación de la solicitud: **1163519**

(87) Fecha de publicación de la solicitud: **19.12.2001**

(54) Título: **Aislamiento y análisis de proteínas.**

(30) Prioridad: **23.03.1999 GB 9906551**
29.03.1999 GB 9907057
06.04.1999 GB 9907641
28.06.1999 GB 9914874
02.07.1999 GB 9915363
06.07.1999 GB 9915677
14.07.1999 GB 9916511
31.08.1999 GB 9920503
21.09.1999 GB 9922285

(45) Fecha de publicación de la mención BOPI:
01.03.2007

(45) Fecha de la publicación del folleto de la patente:
01.03.2007

(73) Titular/es: **Biovation Limited**
Investment House, 6 Union Row
Aberdeen AB10 1DQ, GB

(72) Inventor/es: **Carr, Francis, J.**

(74) Agente: **Carvajal y Urquijo, Isabel**

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

Aislamiento y análisis de proteínas.

5 La presente invención tiene que ver con el aislamiento y análisis de proteínas, especialmente por análisis de espectrometría de masas. La invención tiene aplicación particular en el aislamiento de proteínas enlazantes tales como los anticuerpos.

10 En cuanto al aislamiento de proteínas, la invención suministra nuevos métodos para aislar proteínas específicas a partir de una mezcla compleja de tales proteínas mediante el enlazamiento a un compuesto blanco específico. La invención provee, en particular, métodos para aislar dominios específicos de anticuerpos a partir de una mezcla de tales dominios derivada de una genoteca, enlazándose a un antígeno blanco específico. Para el análisis de las proteínas la invención provee nuevos métodos para analizar mezclas complejas de proteínas, especialmente para cotejar proteínas entre dos o más muestras diferentes.

15 El invento tiene que ver, en particular, con genotecas específicas de proteínas individuales que incluyen tractsos de secuencias identificadoras mediante las cuales se puede identificar proteínas individuales. La invención tiene que ver, más aún, con métodos de identificación y tamizado de proteínas que pertenecen a dichas genotecas. La invención tiene también que ver con métodos de amplificación y expresión de los genes codificados en una proteína individual perteneciente a dicha genoteca.

20 Para el aislamiento de las proteínas de las mezclas complejas mediante el enlace de éstas a un blanco específico, y donde la identidad o secuencia de amino ácidos de la proteína se conoce de antemano, usualmente ha sido muy difícil el aislar suficiente cantidad de proteína para que se adhiera al blanco con el fin de lograr la caracterización directa de la misma. Para seleccionar una proteína de interés a partir de una genoteca grande de proteínas naturales, sintéticas o semisintéticas, se han desarrollado métodos de “visualización de proteínas” donde proteínas recombinantes se producen físicamente adheridas a sus genes de tal forma que la recuperación de las proteínas permita la subsecuente recuperación rápida de los genes. Tales métodos incluyen métodos de visualización “*in-vivo*” tales como visualización sobre bacteriófago (“fagovisualización”), bacteria y levadura, e incluyen métodos de visualización “*in-vivo*” tales como visualización sobre ribosomas (“visualización ribosómica”). Los genes recuperados pueden ser secuenciados para determinar la identidad de la proteína recuperada o pueden ser utilizados para regenerar la proteína recuperada. El documento WO 9215679 revela librerías de fagovisualización mejoradas que presentan varios dominios proteicos con potencial de unirse a un material blanco de interés donde se puede incorporar enlazantes rompibles de proteasa de sitio específico. Si una genoteca es sujeta a métodos de visualización de proteínas mediante los cuales las proteínas sean seleccionadas de acuerdo con características particulares tales como su capacidad de enlazarse a un antígeno (para regiones anticuerpo variables), entonces, con cada ciclo de selección, los genes recuperados se enriquecerán en aquellas proteínas que codifican tales características en particular. Las desventajas de los actuales métodos de visualización “*in-vivo*” incluyen un límite a la cantidad de proteínas funcionales visualizadas (la fagovisualización usualmente está limitada a polipéptidos menores a 40 kDa), la necesidad usual de fusionar la proteína recombinante a una proteína huésped (la cual puede interferir con la función o capacidad enlazante de la proteína recombinante), y una inhabilidad para variar el número de proteínas visualizadas por partícula visualizada; Este último es también un problema con métodos de visualización “*in-vivo*” tales como el de la visualización ribosómica. Adicionalmente, los métodos para la selección de proteínas con características particulares, tales como el poderse enlazar a un antígeno, están limitados debido al pequeño tamaño de las partículas de visualización, de tal suerte que métodos tales como Clasificación Celular Activada por Fluorescencia (FACS) no pueden ser empleados fácilmente. De ésta manera, queda la necesidad de nuevos métodos para mejorar el aislamiento de proteínas a partir de mezclas complejas, en particular para mejorar el aislamiento de regiones variables de anticuerpos (Fv's) a partir de mezclas complejas de Fv's. Esta, la presente invención, suministra métodos mejorados de aislamiento de proteínas a partir de mezclas complejas. En particular, la presente invención combina el uso de librerías de proteínas generadas a partir de genotecas con mejoramientos en espectrometría de masas y especialmente, mejoramientos en espectrometría de tiempo de vuelo con desorción-ionización con láser asistido por matriz (Maldi-ToF) y la habilidad de secuenciar directamente péptidos separados mediante tiempo de vuelo por espectrometría de masas en tándem (MS-MS), y más recientemente, la habilidad de combinar ToF y MS-MS en un sólo equipo (Q-ToF) y la de combinar HPLC con espectrometría de masas con cañón de electrones (ES) en tándem.

55 Así, la presente invención provee un método de identificación, tamizado y/o inclusión de secuenciación de proteínas suministrando una librería de proteínas individuales, una o más de las cuales pueden enlazarse a un blanco de interés, donde cada proteína individual incluye en su secuencia un “código de barra” que puede ser utilizado para identificar cada proteína individual en la librería.

60 Este aspecto de la presente invención tiene en cuenta librerías de proteínas, especialmente dominios de anticuerpos recombinantes tales como Fv's, donde miembros proteicos individuales de la librería incluyen, dentro de su secuencia de amino ácidos, un tracto de secuencia (un “código de barras”) que a su vez puede ser secuenciado con el fin de identificar qué proteína(s) ha sido enlazada al blanco específico (o, en el caso de Fv's, “antígeno”). Esta modalidad aplicará especialmente en los casos donde Fv's se derivan de genes humanos donde, a su vez, el Fv seleccionado puede ser adecuado para uso terapéutico en humanos o para diagnóstico. En esta aplicación en particular, se crea una genoteca de Fv's a partir de un banco de inmunoglobulinas tipo cADN, tales como las derivadas a partir de células B de sangre periférica en humanos, o tales como bancos sintéticamente creados utilizando regiones variables humanas

con CDR's (regiones determinantes de complementariedad) aleatorizadas ("combinatorial") en una o más posiciones. Si esta genoteca es creada de tal manera que un gen aleatorio (o semialeatorio) se incluye dentro de la región codificante Fv o en el extremo de ésta, entonces dicha secuencia aleatoria/semialeatoria de genes generará una secuencia, aleatoria/semialeatoria, de péptidos asociados con Fv's individuales. Tal secuencia aleatoria/semialeatoria de genes se crea utilizando métodos estándares tales como el de iniciación con oligonucleótidos/ADN polimerasa: Extensión o PCR donde una secuencia aleatoria/semialeatoria sintética de oligonucleótidos se utiliza como uno de un par de cebadores utilizados para amplificar fragmentos de genes de la inmunoglobulina durante la creación de la genoteca Fv. Si los miembros de la genoteca Fv constituyen dos cadenas (p.ej, cadenas laterales pesada y liviana derivadas de otras cadenas (VH y VL)) en contraposición a una sola (VH y VL unidas mediante un enlazante peptídico), entonces los "códigos de barra" individuales pueden ser asociados a cada una de las cadenas (o pueden ser asociados a una sola de ellas). Al crearse la genoteca, cada uno de los Fv's resultantes incluye uno o más "códigos de barra" únicos para ese Fv en particular o para un pequeño subconjunto de Fv's de la genoteca compleja. Preferiblemente el código de barras peptídico es C terminal a la región Fv de cadena sencilla o C terminal al VH o VL o a ambos e incluye; flanquearse entre sí mismo y la región Fv, uno o más sitios con actividad proteásica tales como sitios para la enterocinasa (se rompe después de Asp-Asp-Asp-Asp-Lys), factor Xa (se rompe luego de Ile-Glu/Asp-Gly-Arg) u otras endopeptinasas. Si una mezcla de tales Fv's es producida a partir de una genoteca adecuada, entonces ésta mezcla es a su vez mezclada con un antígeno blanco (o antígenos tales como los que se encuentran en la superficie celular), usualmente donde el antígeno es inmovilizado. Esto da como resultado que los Fv's se enlazan específicamente al antígeno blanco mientras que los no enlazantes (o enlazantes débiles, dependiendo de la severidad del lavado) son retirados. Luego de lavado el exceso de anticuerpos, el restante complejo antígeno/Fv es entonces usualmente liberado del Fv por digestión con la endoproteasa utilizada para la lisis del sitio sensible a la proteasa introducida. Estos péptidos "código de barras" liberados son entonces sujetos a una secuenciación vía análisis de masas/espectrometría de masas, ya sea en forma directa o, si se desea, siguiendo su captura en virtud a sus amino ácidos o secuencias específicas de amino ácidos lo cual permite capturar al péptido en una fase sólida tal como residuos de cistina que pueden ser biotinilados para su subsecuente captura sobre avidina o estreptovidina. En forma alternativa, se puede usar cualquier otro método para determinar la secuencia del péptido código de barras ya sea dentro del Fv o después de la liberación incluyendo el utilizar ligantes específicos que se unan al código de barras en una manera de secuencia específica. Luego de determinar las secuencias (o parte de las secuencias) de los códigos de barra derivados de los Fv's enlazados, los oligonucleótidos sintéticos correspondientes son producidos y utilizados para ampliar específicamente o enriquecer en genes específicos Fv de la genoteca. Estos genes Fv específicos o enriquecidos (o HV y VL) son entonces utilizados aún más para generar Fv's correspondientes que puedan entonces ser puestos a prueba nuevamente en su capacidad para enlazar antígenos, ya sea individualmente o como parte de un pequeño banco de Fv's aislados. Finalmente, mediante éste método se puede generar Fv's específicos con propiedades deseables de enlace a los antígenos y, si son de origen humano, con potencial terapéutico útil. Este aspecto incluye el uso de códigos de barras múltiples asociados con proteínas individuales o Fv's, por ejemplo dos códigos de barras adyacentes en el terminal C de los Fv's donde dos péptidos son liberados de cada Fv mediante digestión con proteasa, ya sea en forma simultánea para realzar la identidad de los Fv's que se unen al blanco o secuencialmente donde diferentes proteasas son utilizadas en rondas sucesivas de digestión para proveer un medio diferente de amplificar subsecuentemente genes Fv correspondientes a Fv's que se ligan al blanco. Este aspecto también incluye el uso de códigos de barra múltiples que se analizan simultáneamente con el fin de aumentar la diversidad de secuencias de códigos de barra globales para proveer una codificación específica de las proteínas individuales. Este aspecto también incluye el uso de códigos de barras dentro de proteínas individuales, por ejemplo dentro de una o más posiciones CDR de un Fv. Este aspecto también incluye el uso de proteasas que también pueden digerir los componentes proteicos de la mezcla proteína:blanco o, adicionalmente, cualquier agente proteico usado para inmovilizar al blanco, con la condición de que los péptidos códigos de barras liberados de la(s) proteína(s) de prueba enlazada(s) puedan ser aún detectadas y secuenciadas dentro del conjunto de los otros péptidos. En el formato preferido de este aspecto, se suministra una sola región de códigos de barras en el terminal C de las cadenas livianas que forman un fragmento FAB soluble donde VHs y VLs están codificados con el misma expresión casete o cistrón de tal forma que la secuencia código de barras se puede utilizar para acceder a ambos genes VH y VL. Tal fragmento FAB puede ser producido convenientemente utilizando un rango de sistemas de expresión, por ejemplo el sistema vector bacteriófago M13 donde, por introducción de secuencias líderes secretoras, las cadenas pesada y liviana de los FABs son secretadas en el espacio periplásmico de la bacteria huésped y en ese espacio son cultivadas. El sistema vector se prepara primero con códigos de barras enmarcados por clonación en mezclas de oligonucleótidos sintéticos. Para la formación de los códigos de barras adyacentes esto se efectúa convenientemente mediante clonación secuencial o mutagénesis de oligonucleótidos donde bancos de recombinantes M 13, que contienen el primer código de barras mezclado, son preparados como una plantilla para la clonación posterior del segundo código de barras enmarcado. Preferiblemente, la codificación de barras se designa de tal manera que la proteína codificada contiene sitios endonucleasa, tanto flanqueando los dos códigos de barras, como entre ellos e incluso donde una región "espaciadora" adyacente a uno de los códigos de barra crea un péptido que incluye ese código de barras, el cual tiene un peso molecular mayor que el del otro código de barras. Mediante el diseño juicioso de códigos de barras y el uso de códigos de barras múltiples en esta forma se provee una opción para simplificar el análisis de masas de los péptidos liberados de las endoproteasas mediante, por ejemplo, MALDI-ToF donde las secuencias de los péptidos pueden ser deducidas (o casi deducidas) de tal forma que los oligonucleótidos puedan ser diseñados para aislar (o enriquecer) proteínas específicas con el (los) código(s) de barras detectado(s) mediante el análisis MALDI-ToF. Tal deducción de éstas secuencias se consigue por diseño de secuencias donde amino ácidos específicos pueden presentarse en una o dos posiciones a lo largo del péptido. Por ejemplo, donde se diseña el péptido utilizando 17 de los 20 amino ácidos naturales (designados aquí como A-Q), entonces las secuencias pueden ser diseñadas con opciones para cualquiera de tres amino ácidos en cada posición a lo largo de la secuencia del péptido como sigue:

ES 2 265 917 T3

posición aa:	1	2	3	4	5	6	7	8
opciones de aminoácidos:	A	C	E	G	I	K	M	O
	B	D	F	H	J	L	N	P
	C	E	G	I	K	M	O	Q

5

10

15

Este diseño teóricamente daría 6561 diferentes secuencias códigos de barras peptídicas. Si un código de barras adyacente con una región espaciadora es diseñado también con esta misma base, daría entonces 6561 códigos de barras adicionales diferentes. En combinación, esto crearía 4×10^7 secuencias de códigos de barras que serían adecuadas para etiquetar de manera única a la mayoría de los miembros de una librería de proteínas de dicho tamaño. El empleo de secuencias de códigos de barras adyacentes adicionales o códigos de barras de mayor tamaño basados, por ejemplo, en el uso de dos amino ácidos específicos en cualquier posición de la secuencia (creando de esta manera 262.144 diferentes secuencias de códigos de barras en los que se emplean 19 amino ácidos) incrementaría la diversidad de códigos de barras provistos. En la práctica, la redundancia de codón se reduce a través de la elección juiciosa de los codones en cada posición de la secuencia durante el diseño de oligonucleótidos sintéticos mixtos. Un diseño de oligonucleótidos para un código de barras peptídico de 8 amino ácidos para secuenciación MS/MS es el siguiente:

20

Codones	NAC	NCC	NGG	NTG	TKC	VAG	GNV	CNT
Aminoácidos	N	T	R	L	F	Q	D	H
	D	P	G	M	C	E	V	L
	H	A	W	V		K	A	P
	Y	S					G	R

25

donde los codones

N = A, C, G o T

K = G o T

V = A, C o G

$4 \times 4 \times 3 \times 3 \times 2 \times 3 \times 4 \times 4 = 13824$ secuencias de códigos de barras

30

35

40

45

Codones específicos también pueden ser incorporados mediante síntesis discontinua de oligonucleótidos donde se adicionan secuencialmente codones específicos a mezclas separadas de oligonucleótidos previamente sintetizados ("mutagénesis de codones"). Una vez que se ha deducido un candidato a secuencia de código de barras por MALDI-ToF o MS/MS y donde la diversidad de códigos de barras individuales es menor que la de la librería, el correspondiente oligonucleótido específico (o mezcla de oligonucleótidos, si hay redundancia en el uso de los codones) se puede utilizar como un cebador PCR en conjunto con un cebador opuesto diseñado a partir de la proteína o sistema vector para enriquecer en codificación de genes a la proteína a partir de la cual fue detectado el código de barras. Donde se emplea códigos de barras adyacentes se puede usar entonces un segundo cebador anidado dentro del fragmento de un gen creado a partir del primer cebador, con el fin de enriquecer en su codificación genética a la proteína detectada. Si se requiere, el método anterior puede incorporar tres o más códigos de barras con el fin de aumentar la especificidad del enriquecimiento direccionado hacia el oligonucleótido de genes específicos que codifican a la proteína deseada. Será entendido por aquellos expertos en el arte, que éste primer aspecto de la invención puede cubrir un número de métodos variantes con el principio esencial de que una proteína específica es recuperada a partir de una librería de tales proteínas vía el análisis de masas o secuencial de uno o más péptidos asociados con o codificados por esa proteína específica y que, como tal, que éste aspecto tiene una amplia utilidad en el aislamiento de genes que codifican proteínas deseadas donde solo se determina o deduce la secuencia peptídica.

50

55

60

65

Será entendido por aquellos expertos en el arte que, dentro del objetivo de ésta invención, hay muchas variaciones del primer aspecto. Por ejemplo, se entenderá que los códigos de barras peptídicos podrían ser incorporados en pares o grupos de proteínas a las cuales se les permite entonces ligarse con el fin de determinar qué proteínas se unen entre sí, gracias a la detección de los códigos de barras de cada una de las proteínas involucradas en la unión. Como una alternativa al aislamiento de las proteínas de mezclas complejas, las proteínas dentro de éstas mezclas complejas que muestran ciertas propiedades enlazantes tales como la de ligarse a otras macromoléculas como el ADN pueden ser detectadas utilizando el presente método. El presente método incluye una variedad de formas de adicionar códigos de barras peptídicos a las proteínas incluyendo métodos donde el código de barras es codificado dentro del fragmento genético que a su vez codifica a la proteína. Sin embargo, se puede adicionar códigos de barras a tales proteínas o a cualquiera otra mezcla adecuada de moléculas mediante el enlace directo de péptidos. Por ejemplo, se puede adicionar péptidos específicos a anticuerpos específicos o a proteínas utilizando un rango de métodos químicos o fotoquímicos. Una aplicación de tal método es etiquetar una mezcla compleja de proteínas con un código de barras (o selección de códigos de barra, por ejemplo con diferentes especificidades proteicas) y el otro código de barras a una mezcla compleja alternativa de proteínas. Por ejemplo para codificar con barras, de forma diferencial, proteínas provenientes de dos muestras diferentes que luego son mezcladas. Será entendido por parte de aquellos expertos en el arte, que el principio de adicionar códigos de barras peptídicos a las proteínas o a otras moléculas podría aplicarse también a códigos de barras no peptídicos donde dichos códigos puedan ser identificados directamente (o casi identificados) utilizando métodos de masas o de secuenciación. Como tales, los códigos de barras podrían incluir códigos de barra de ácidos nucleicos ligados a proteínas u otras moléculas incluyendo ácidos nucleicos. Al igual que con los códigos de barras peptídicos, tales códigos de barras de los ácidos nucleicos se pueden analizar por espectrometría de masas

Será entendido por aquellos expertos en el arte que, dentro del objetivo de la presente invención, hay aplicaciones del primer aspecto diferentes a las del aislamiento de proteínas, por ejemplo, la distribución de las proteínas u otros ligandos dentro de un organismo vivo puede ser dilucidada analizando los códigos de barras, por masas o por secuenciación, que están asociados con órganos específicos dentro de ese organismo. En el análisis de la especificidad de enlace de los péptidos o proteínas a otras moléculas, los códigos de barra pueden ser construidos como parte de las regiones enlazantes del péptido o proteína con el fin de determinar especificidad por análisis de masas o de secuenciación de los códigos de barras. Por ejemplo, se puede construir secuencias mixtas de códigos de barras peptídicos alrededor de residuos de anclaje conocidos de moléculas MHC y luego determinar, mediante elusión y análisis de masas o de secuenciación de los códigos de barras, el espectro de péptidos que se enlazan a moléculas MHC específicas.

Ejemplo 1

Los experimentos descritos en el presente ejemplo fueron realizados utilizando un par de genes anticuerpos modificados de cadena sencilla (scAbs). Se prepararon dos scAbs modificados consistentes en etiquetas epítomos N-terminales, la región variable de cadena pesada (VH), un enlace de 14 amino ácidos (EGKSSGSGSESKVD), la región variable de cadena liviana (VL) cada una fusionada al dominio b-zip desde, ya sea, el gen c-jun o c-fos.

Estas construcciones fueron clonadas al vector pET 5c (Rosenberg AH *et al.*, **56**:125-135, 1987) que provee un promotor T7 seguido por un sitio de enlace ribosómico a partir de T7 gen 10. Las ramificaciones scAb fueron insertadas en el vector en un sitio NdeI tal que la secuencia que codifica la etiqueta epitopo siguiera al primer ATG de T7 gen 10. La primer ramificación consistió de un scAb contra *Pseudomonas aeruginosa* (Molloy P. *et al. Journal of Applied Bacteriology*, **78**: 359-365, 1995) con la bandera epítipo (MDYKDDDK) (Knappik A y Pluckthun A, *Bio Techniques*, **17**: 754-761, 1994) adicionado al terminal N, y el dominio b-zip de c-fos (Abate, C. *et al Proc. Natl. Acad. Sci. USA*, **87**: 1032-1036, 1990) en la región terminal C de la proteína. El segundo consistió de un scAb construido a partir del antígeno anticuerpo anti-fetal 340 (Durrant LG *et al. Prenatal Diagnosis*, **14**: 131-140, 1994) con una etiqueta poli-Histidina en el terminal N, y el dominio b-zip de c-jun (Abate C. *et al. ibid*) en la región terminal C de la proteína.

Las *anti-Pseudomonas aeruginosa* (α -Ps-fos) scAb y el 340-jun scAb se construyeron como se describe a continuación:

Se amplificó ADN para el α -Ps scAb en el vector pPMIHis (Molloy P *et al.*, *ibid*) con los cebadores RD 5' FLAG: 5' gcggatcccatatggactacaagacgatgacgacaacaggtgcagctgcag3' (Genosys Biotechnologies Europe Ltd, Cambridge, Reino Unido) y RD 3': 5' gcgaattcgtggtggtggtggtggtggtgactctcc3' (Genosys) los cuales introdujeron la secuencia epítopo 5' de FLAG y removieron el codón 3' de detención respectivamente. La mezcla de reacción incluyó 0,1 μ g de ADN molde, 2,6 unidades de mezcla enzimática PCR Expand™ High Fidelity PCR (Boehringer Mannheim, Lewes, Reino Unido), solución amortiguadora HF Expand (Boehringer Mannheim), MgCl₂ 1,5 mM, trifosfatos deoxinucleótidos (dNTPs) 200 μ M (Life Technologies, Paisley, Reino Unido) y 25 pmoles de cada cebador. Los ciclos fueron de 5 minutos a 96°C, seguidos por [95°C 1 minuto, 50°C 1 minuto, 72°C 1 minuto] 5 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 1 minuto 30 segundos] 8 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 2 minutos] 5 veces, terminando con 72°C 5 minutos. El producto bp 1123 obtenido fue cortado con BamHI y EcoRI y clonado en el vector pUC19 (Boehringer Mannheim). La secuencia de ADN fue confirmada utilizando el kit ciclosecuenciador radioetiquetado Thermo Sequenase con [³²P] di-deoxinucleótidos (Amersham Life Science, Amersham, Reino Unido). El producto obtenido fue clonado en el vector pET5c (Promega UK Ltd, Southamptom, Reino Unido) como un fragmento Ndel a EcoRI (ver *Molecular Cloning, A Laboratory Manual* eds. Sambrook J, Fritsch EF, Maniatis T. Cold Spring Harbor Laboratory Press 1989, New York, EEUU). Se prepararon plásmidos de ADN utilizando el sistema de purificación Wizard® Plus SV Miniprep ADN (Promega UK Ltd), o para una escala mayor, el Qiagen Plasmid Midi Kit (Qiagen Ltd, Crawley, Reino Unido). El nuevo plásmido generado fue denominado pET5c FLAG- α Ps scAb.

El casete fos fue ensamblado mediante PCR de oligonucleótidos sobrepuestos:

Fos 1 para 5'-atggaattcctcgagaccgacaccctacaggecggaaccgaccagctgga

Fos85rev 5'-tcgcgatttcggtttcagcgcggatttttcgtcttcagctggtcggtt

Fos 71 para 5'-aaaccgaaatcgcgaacctgctgaaagaaaaagaaaagctggagttcatc

Fos 155rev 5'ggaagcttgaattccgccggacggtgtgccgccaggatgaactccagctt

Los anteriores oligonucleótidos fueron incluidos en una mezcla de reacción a 1 pmol cada uno, y la reacción fue iniciada empleando 10 pmol de cebadores Fos1fS; 5'-atggaattcctcgagacc y Fos 155rS 5'-ggaagcttgaaattccgcc utilizando polimeraza de alta fidelidad y componentes de reacción en la forma expuesta previamente. El producto 155bp resultante fue digerido con EcoRI, purificado y clonado en EcoRI corte pUC19 para análisis de secuencia utilizando

ES 2 265 917 T3

procedimientos estándares (ver *Molecular Cloning, A Laboratory manual Ibid*). El casete Fos fue subclonado en el plásmido pET5c FLAG- α Ps scAb como un fragmento XhoI-EcoRI mediante substitución del fragmento 320bp XhoI-EcoRI previamente existente el cual contiene el dominio de la región humana constante.

- 5 El 340 scAb fue producido substituyendo el VH y VK del anticuerpo 340 en lugar del α -Ps VH y VK en ppMI-His. El 340 VH fue amplificado con el cebador 5'cagctgcaggagctctggggaggcttag3' (Genosys) y 5'tcagtagacggtagcagggttccttgaccaccagta3' (Genosys). La mezcla de reacción incluía 0,1 μ g de templado ADN, 2,6 unidades de mezcla enzimática ExpandTM High Fidelity PCR, solución amortiguadora Expand HF, MgCl₂ 1,5 mM, 200 μ M dNTPs y 25 pmoles de cada cebador. Los ciclos fueron 96°C 5 minutos, seguido de [95°C 1 minuto, 50°C 1 minuto, 72°C 1 minuto] 5 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 1 minuto 30 segundos] 8 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 2 minutos] 5 veces, terminando con 72°C 5 minutos. El producto 357 bp se cortó con PstI y BatEI y fue clonado en PstI y BstEI cortado pPM1His (ver *Molecular Cloning, A Laboratory manual Ibid*). En forma similar, el 340 VK fue amplificado con los cebadores 5'gtgacattgagctcacacagtctct3' y 5'cagccctttatctcgagcttggtccg3' (Genosys). El producto 339 bp se cortó con SstI y XhoI y se clonó en SstI y XhoI cortado modificado pPM1His (producido anteriormente). La secuencia ADN fue confirmada usando el estuche de ciclosecuenciación radioetiquetado Thermo Sequenase con di-deoxinucleótidos [³³P] como antes. El ADN para el 340 scAb en el vector pPM1His fue amplificado con el cebador RD 5' HIS: 5'gcggatcccatgcaccatcatcaccacaccaggtgcagtcag3 (Genosy) y RD 3' (dados anteriormente) los cuales introdujeron los residuos 6-histidina en el extremo 5' y removieron el codón 3' stop respectivamente. Los reactivos y condiciones para la amplificación fueron exactamente los mismos que para la cadena α -Ps. 20 La secuencia ADN fue confirmada como antes y el producto fue clonado en el vector pRT5c como un NdeI para el fragmento EcoRI para generar el plásmido pEt5c HIS 340 scAb.

El casete jun fue ensamblado mediante PCR de los oligonucleótidos superpuestos:

- 25 junifor 5'-atgagaattctcgagcgtatcgtctgctgaagaaaaagttaaaccct
jun 85rev 5'-tagcgggtggaagccagtcggagttctgagcttcagggtttaactttt
jun 71 para 5'-tggtctccaccgctaacatgctgcgtgaacaggttgctcagctgaaacag
30 jun 146rev 5'-catgcgaattcgtggttcataactttctgttcagctgagcaacc

- Los oligonucleótidos anteriores fueron incluidos en una mezcla de reacción a 1 pmol cada uno, y la reacción se inició utilizando 10 pmol de cebadores Jim 1 para S: 5'atgagaattctcgagcg y Jun 146rev-S: 5'-catgcgaattcgtggttc utilizando polimeraza de alta fidelidad y componentes de reacción, igual que anteriormente. El producto 146bp resultante fue digerido con EcoRI, purificado y clonado en EcoRI cortado PUC19 para análisis de secuencia utilizando los métodos estándares (ver *Molecular Cloning, A Laboratory Manual. Ibid*). El casete Jun fue subclonado en el plásmido pEt5c HIS 340 scAb que lleva el dominio de región constante humana.

- 40 Los plásmidos his-340-jun y FLAG- α Ps-fos, fueron usados como templados para PCR utilizando el cebador biotinylated BioT7; 5'-agatctcgatccgcaaaatta y el cebador petrev; 5'-aaatagcgtatcacpggcc. Los cebadores fueron suministrados por Genosys (Cambridge, Reino Unido) y utilizados en la reacción a una concentración de 1 pmol. Los componentes y condiciones de PCR son las mismas que anteriormente. El producto de la reacción his-340-jun fue 992bp, y el producto de la reacción FLAG- α Ps-fos fue 1002bp. Los productos fueron purificados utilizando un cartucho de purificación giratorio (Qiagen, Crawley, Reino Unido) y diluidos a una concentración de 100 ng/ μ l. La cuantificación se hizo por absorbancia UV a 260 nm. 500 ng de ADN marcado con biotina se hizo reaccionar con 10 μ l de partículas magnéticas recubiertas con estreptavidina (Bangs labs, Fisher, EEUU). La reacción fue realizada en un tubo para microcentrífuga siliconizado en un volumen 500 μ l de PBS 1% (p/v) BSA durante 10 minutos a temperatura ambiente. Después del enlace, las partículas fueron recogidas con un imán (DynaI, Bromborough, Reino Unido) y lavadas tres veces utilizando PBS 1% BSA.

- Después del lavado final se inició la reacción de translación *in vitro* mediante la adición de 25 μ l de mezcla T7 Quick Coupled de transcripción translación (Promega, Southampton, Reino Unido) suplementada con biotinil lisina tARN (Promega). La reacción de translación se efectuó a 30°C durante 60 minutos y luego se refrigeró con hielo. Las partículas fueron recolectadas con imanes y lavadas utilizando PBS frío que contenía BSA al 1%.

- En algunos experimentos se emplearon en las reacciones IVTT partículas no magnéticas de esteptavidina (Bangs Labs, Fishers, EEUU). En estos casos las partículas se recuperaron durante los ciclos de lavado por centrifugación.

- 60 En algunos experimentos se usaron en las reacciones IVTT partículas coloreadas, magnéticas y no magnéticas, de estreptavidina.

- En algunos experimentos, los productos de translación unidos a las partículas se detectaron empleando anticuerpos para, ya sea la bandera o el epítopo his6, diseñados en cada uno de las cadenas genéticas modelo. Se adicionaron anticuerpos a las partículas lavadas diluidas en PBS. Las incubaciones fueron durante 60 minutos a 4°C con agitación suave. Un reactivo secundario (conjugado anti-HRP de ratón-) se adicionó en la dilución recomendada en PBS y se incubó por otros 30 minutos a 4°C. Las partículas fueron lavadas tres veces utilizando 200 μ l de PBS antes del desarrollo de color con el sustrato cromogénico. Las reacciones fueron leídas a 492 nm.

Las reacciones de unión proteína-proteína fueron llevadas a cabo utilizando proteínas IVTT unidas a la superficie de la partícula. En tales experimentos las partículas no magnéticas de estreptavidina fueron “capturadas” mediante enlaces facilitados por proteínas (fos:jun) unidas a la superficie de las partículas magnéticas. Las partículas magnéticas con producto fos IVTT fueron mezclados cuidadosamente con partículas no metálicas con enlaces jun sobre la superficie. La reacción se efectuó en 100 μ l de PBS, BSA y se dejó avanzar a temperatura ambiente durante 30 minutos. En una reacción de control negativo, partículas no metálicas con una proteína Sea (a-Ps scAb; Molloy P *et al.*, *ibid*) enlazada a la superficie se mezclaron con las partículas magnéticas recubiertas con el producto fos IVTT. Después de la incubación, las partículas fueron capturadas mediante imanes y lavadas seis veces usando PBS, 1% BSA.

La presencia del gen proteico blanco capturado fue confirmada usando PCR y secuenciación ADN. Para la detección del gen modelo jun se usaron cebadores jun específicos 1for-S y Jun146rev en un ensayo PCR. El ensayo fue iniciado mediante la adición de 10% (v/v) en partículas, directamente a la mezcla PCR. Los componentes y condiciones de la reacción fueron los mismos que antes. El producto específico 146bp jun fue detectado mediante electroforesis en gel. Para detectar el gen fos modelo, se emplearon *cebadores Fos1S* y Fos 155rS en un ensayo PCR. Las condiciones de la reacción y detección del producto 155bp fos específico fueron iguales a las anteriores. Para la detección de la proteína control negativo se usaron cebadores Seq1scab 5'-agatccctactatagga y Seq2scab; 5'-ggtagctc-gatgtatcc para detectar un producto 115b en el gen proteico α -Ps scAb.

En los anteriores experimentos se detectó productos jun PCR luego de la captura mediante partícula magnética fos bajo condiciones donde no se podían detectar productos α -Ps scAb PCR luego de la interacción con las partículas fos magnéticas.

Ejemplo 2

En este ejemplo se produjo una librería de anticuerpos de una sola cadena incluyendo “códigos de barras” peptídicos únicos. Se preparó ARN linfocítico de sangre humana periférica de acuerdo con los procedimientos estándares. Brevemente, se prepararon linfocitos a partir de 10 ml de sangre heparinizada tomada de 16 donantes sanos. Los linfocitos fueron recolectados luego de un procedimiento de centrifugación por gradiente de densidad utilizando medio Lymphoprep (Sigma, Poole, Reino Unido). El ARN se preparó utilizando el sistema Quick-Prep y las instrucciones suministradas por el proveedor (Pharmacia, St Albans, Reino Unido). La síntesis del cADN fue realizada usando un kit de síntesis cADN (Pharmacia, St Albans, Reino Unido) y cebantes hexameros aleatorios con las condiciones recomendadas por el proveedor. La región variable de cadena pesada de la inmunoglobulina (Vh) y las regiones variables de cadenas livianas (VI) fueron amplificadas a partir del cADN en mezclas separadas PCR utilizando conjuntos cebadores diseñados para maximizar los repertorios Vh y VI. Los conjuntos cebadores fueron los mismos ya descritos (Marks J.D *et al* 1991, Eur. J. Immunol. 21: 985). Las reacciones Vh y VI PCR se efectuaron usando 2,6 unidades de mezcla enzimática ExpandTM High Fidelity PCR (Boehringer Mannheim, Lewes, Reino Unido), solución amortiguadora Expand HF (Boehringer), MgCl₂ 1,5 mM, trifosfatos doxinucleótidos (dNTPs) 200 μ M (Life Technologies, Paisley, Reino Unido) y 25 pmoles de cada banco de cebadores. Los ciclos fueron 96°C 5 minutos, seguido de [95°C 1 minuto, 50°C 1 minuto, 72°C 1 minuto] 5 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 1 minuto 30 segundos] 8 veces, [95°C 45 segundos, 50°C 1 minuto, 72°C 2 minutos] 5 veces, terminando con 72°C 5 minutos.

En un PCR aparte, un fragmento enlazante de la forma (Gly₄Ser)₃ (Huston J.S. *et al* 1988, PNAS, 85: 5879-5883) fue amplificado a partir de un templado clonado pSW1-ScFvD1.3 (McCafferty *et al*, 1990, *Nature* **348**: 522-554) utilizando los conjuntos de cebantes descritos anteriormente (Marks, J.D en *Antibody Engineering*, ed Borrebaek C.A.K New York O.U.P., 1995). El producto fragmento enlazante 93bp fue enlazado junto con una mezcla equimolar de los productos Vh y VI PCR. La mezcla fue amplificada aún más en una reacción “de tracción” utilizando para ello cebantes flaqueadores HuFORNot tal como lo detallan Vaughan *et al* (Vaughan T. J. *et al* 1996, *Nature Biotech.* **14**: 309-314). Todos los fragmentos utilizados en la reacción de tracción fueron purificados para librarlos de sus cebadores iniciales antes de incluirlos en la reacción. La purificación se realizó usando el sistema Wizard PCR Perps de Promega (Promega, Southampton, Reino Unido).

La secuencia contigua ensamblada de la forma Vh-enlazante-VI fue digerida con enzimas de restricción SfiI y NotI (Boehringer) utilizando condiciones estándares y purificada como se describió anteriormente. El fragmento purificado fue enlazado con una mezcla adaptadora de oligonucleótidos sintéticos de doble cadena con el fin de introducir un sitio de rompimiento de la proteasa V8 yuxtapuesto con un tracto de secuencia aleatorizada en consonancia con el terminal C del gen VI. Esta secuencia de código de barras V8/única fue producida templando un par de bancos de oligonucleótidos sintéticos de la forma

5'-ggccgcgaggaagaggaa[(atg)/(can)/(agn)/(aan)/(gan)/(ttn)]₂gc-3' y

5'-ggccgc[(naa)/(ntc)/(ngt)/(nct)(nag)/(cat)]₂ctccttctctcgc-3'

Este enlazante tiene extremos NotI compatibles (subrayados) y por lo tanto facilita la inserción del fragmento código de barras del anticuerpo-V8/único de cadena sencilla completo en el vector fagómico pCANTAB 5 SfiI-NotI preparado (Pharmacia).

La secuencia código de barras, única, fue diseñada para evitar la introducción de codones stop y sesgada aún más para excluir residuos codificantes con más de dos codones alternativos. Mediante ésta estrategia, el número de oligonu-

ES 2 265 917 T3

cleótidos requeridos para identificar una secuencia peptídica decodificada dada es minimizado. En suma, la secuencia código de barras única es capaz de codificar 11 de los 20 amino ácidos. Adicionalmente al sitio de rompimiento de la peptidasa V8 (una cuerda de 4 residuos de ácido glutámico), la secuencia código de barras tiene una longitud de 12 codones. De esta manera el repertorio de 11 amino ácidos (10 de los cuales son codificados por cualquiera de los dos codones) es capaz de codificar $11^{12/2} = \sim 1,5 \times 10^{12}$ péptidos diferentes.

El fragmento scfv ensamblado (Vh-enlazante-VI) con terminaciones SfiI y NotI fue, enlazado a la secuencia código de barras NotI adaptadora y re-purificado. Para experimentos que expresan la librería humana scfv mediante fago de despliegue, el fragmento completo fue enlazado al vector phagemid pCANTAB 5 (Pharmacia) SfiI-NotI preparado, y transformado en TG1 *E. coli* competente.

Para otros experimentos en los cuales se utilizó transcripción *in vitro* y translación (IVTT), la librería scfv ensamblada fue subclonada en el pCANTAB5-T7 SfiI NotI preparado. Este vector es el mismo que el pCANTAB5 disponible comercialmente excepto que fue modificado para incluir la secuencia promotora T7 (ttaatacgaactactata) insertada en el sitio HindIII en la posición 2235. La modificación fue lograda mediante el enlace de un enlazante ADN sintético de doble cadena de secuencia 5'-agctaatacgaactactata en un corte HindIII y pCANTAB5 defosforilado. Se seleccionaron clones recombinantes que contienen el promotor T7 utilizando un PCR de diagnóstico.

Luego del enlace y transformación en el TG1 *E. coli* competente, las células fueron cultivadas durante 1 hora en 1 ml de medio SOC y luego colocadas en placas sobre medio TYE con 100 µg/ml de ampicilina. Las colonias fueron retiradas de las placas en 25 ml de caldo de cultivo 2x TY que contiene ampicilina. La librería cultivada se utilizó para preparar ADN para reacciones IVTT.

La librería ADN pCANTAB5-T7 Sefv se empleó en una reacción de translación *in vitro*. El IVTT se llevó a cabo utilizando mezcla transcripción-traslación acoplada T7 Quick (Promega, Southampton, Reino Unido) y 10 µg de la librería ADN pCANTAB5-T7 Scfv en un volumen total de 50 µl. La reacción de translación fue efectuada a 30°C durante 90 minutos y luego refrigerada en hielo. En algunos experimentos se monitorearon las reacciones buscando presencia de productos de translación utilizando ensayos de modalidad de ³⁵S-metionina. Las reacciones fueron almacenadas a -70°C antes de su empleo en ensayos de enlace y tamizado.

La librería anticuerpos de cadena sencilla se utilizó en una reacción de enlace a la proteína humana recombinante p53 (Oncogene Research Products - Calbiochem, Nottingham, Reino Unido). La mezcla IVTT fue diluida 10 veces en PBS y utilizada en un ensayo de enlace a la proteína humana recombinante p53 inmovilizada en una microplaca de 96 pozos. La proteína p53 fue inmovilizada mediante incubación durante toda la noche a una concentración de 100 µg/ml en solución tampón fosfato a 4°C. La placa fue lavada usando PBS 0,5% (p/v) BSA y la mezcla IVTT diluida fue adicionada a los pozos de prueba y de control para su enlace. La reacción de enlace se efectuó a 37°C durante 90 minutos. La placa se lavó 3 veces usando PBS-T (PBS + 0,05% v/v tween-20) y luego fue sometida a digestión con proteasa V8 (Takara, Wokingham, Reino Unido). Los fragmentos proteicos fueron recolectados del sobrenadante y fraccionados según su tamaño para excluir a la proteasa V8 y otras especies grandes antes del análisis por MALDI-tof.

El análisis de fragmentos MALDI-tof identificó un número de fragmentos peptídicos. Las secuencias peptídicas se emplearon para diseñar un grupo de oligonucleótidos sintéticos correspondientes. Los oligonucleótidos se usaron en un tamizado basado en PCR de la librería de cadena sencilla. Se utilizó ADN Pfu turbo (Stratagene Europe) para sintetizar las cadenas complementarias en miembros de la librería ADN anticuerpos humanos de cadena sencilla. Luego de 15 rondas de ciclos térmicos, el producto fue sometido a digestión DpnI. Este paso agotó la mezcla de moléculas plasmídicas parentales para asegurar que solamente los productos cebadores sintetizados se propagaran.

1 µl de la reacción fue transformado en células competentes TG1 y colocado sobre placas LB que contenían 100 µg/ml de ampicilina. Se escogieron los clones individuales, se expandieron y se prepararon con ADN de acuerdo con los procedimientos estándares. El ADN se utilizó directamente en una segunda vuelta de tamizado que involucró IVTT, enlace a antígeno, digestión con proteasa V8, análisis de fragmentos MALDI-tof. Luego de dos vueltas de selección, 6 scFv's fueron aislados los cuales enlazaban p53 recombinante.

REIVINDICACIONES

- 5 1. Una librería de proteínas individuales, siendo una o más de éstas capaz de enlazarse a un blanco de interés, cada una de dichas proteínas teniendo dentro de su secuencia de amino ácidos o en uno de sus extremos, uno o más trectos de secuencias identificadoras individuales sintéticas las cuales son "códigos de barras" únicas a la proteína en particular enlazada a un blanco de interés específico, y que además se encuentran flanqueadas por uno o más sitios sensibles a la proteasa, dicha librería, de ésta manera, provee proteínas que contienen una diversidad de los mencionados trectos de secuencias identificadoras individuales sintéticas.
- 10 2. Una librería de la reivindicación 1 donde dichos trectos de secuencias identificadoras han sido generadas aleatoria o semialeatoriamente.
- 15 3. Una librería de las reivindicaciones 1 ó 2, donde una proteína individual de dicha librería incluye múltiples trectos de secuencias identificadoras.
- 20 4. Una librería de acuerdo con la reivindicación 3, donde una proteína individual incluye dos trectos de secuencias identificadoras adyacentes.
- 25 5. Una librería de acuerdo con cualquiera de las reivindicaciones 1 a 4, donde dicho sitio sensible a las proteasas es un sitio de reconocimiento para la digestión de endoproteasas.
- 30 6. Una librería de la reivindicación 5, donde el sitio de reconocimiento es el sitio para la enteroquinasa o Factor Xa.
- 35 7. Una librería de acuerdo con cualquiera de las reivindicaciones 1 a 6, donde dicho blanco de interés es una proteína.
- 40 8. Una librería de cualquiera de las reivindicaciones 1 a 7, donde las proteínas individuales de la librería son anticuerpos o dominios de anticuerpos recombinantes, que incluyen regiones anticuerpo variables.
- 45 9. Una librería de la reivindicación 8, donde el dominio del anticuerpo es un dominio Fv consistente de una cadena VH y otra VL.
- 50 10. Una librería de acuerdo con la reivindicación 9, donde cada cadena tiene su propio tracto de secuencia identificadora.
- 55 11. Una librería de acuerdo con cualquiera de las reivindicaciones 8 a 10, donde dicho tracto de secuencia identificadora es C-terminal a la secuencia de dominio Fv.
- 60 12. Una librería de acuerdo con cualquiera de las reivindicaciones 8 a 11, donde el blanco de interés es un antígeno.
- 65 13. Un método de identificación de, o tamizado para una proteína que se enlaza a un blanco de interés donde la proteína es un miembro de la librería según se define en cualquiera de las reivindicaciones 1 a 12, el cual consta de los siguientes pasos:
 - (i) Colocación en contacto o asociación a cada una de las proteínas individuales de la librería que incluye a dicho, o más trectos de secuencias identificadoras individuales sintéticas y al (los) sitio(s) sensibles a la proteasa, con uno o más de los mencionados blancos de interés,
 - (ii) Aislamiento del complejo formado mediante una proteína individual y el blanco de interés,
 - (iii) Digestión de dicho complejo para romper los sitios sensibles a la proteasa introducidos, liberando dicho(s) tracto(s) de secuencias identificadoras individuales sintéticas, y
 - (iv) Determinación de dicho(s) tracto(s) de secuencia liberados mediante el (los) cual(es) dicha proteína individual es finalmente identificable.
14. Un método de reivindicación 13, donde la proteína individual ha sido codificada por una construcción de ADN que incluye codificaciones de secuencias nucleicas para dicho o más trectos de secuencia identificadora individuales y para dicho(s) sitio(s) sensible(s) a la proteasa.
15. Un método de reivindicación 14, donde dicha construcción de ADN es parte de un sistema vector.
16. Un método de acuerdo con cualquiera de las reivindicaciones 13 a 15, donde la proteína individual y blanco de interés se enlazan en solución.

ES 2 265 917 T3

17. Un método de acuerdo con cualquiera de las reivindicaciones 13 a 16, donde el complejo proteína/blanco es aislado por remoción de moléculas no enlazantes.

5 18. Un método de acuerdo con cualquiera de las reivindicaciones 13 a 17, donde la digestión de dicho complejo se logra mediante el uso de endoproteasa.

19. Un método de acuerdo con cualquiera de las reivindicaciones 13 a 18, donde la determinación del(los) tracto(s) de secuencias identificadoras se logra mediante espectrometría de masas.

10 20. Un método de reivindicación 19, donde la espectrometría es MALDI-ToF.

21. Un método de acuerdo con cualquiera de las reivindicaciones 13 a 20, donde las proteínas individuales de la librería son anticuerpos o dominios de anticuerpos que tienen una región variable y donde los blancos de interés son antígenos.

15 22. Un método para la amplificación y expresión de los genes que codifican a la proteína individual según la identificación realizada mediante un método de las reivindicaciones 13 a 21 utilizando la secuencia nucleica del tracto de secuencias identificadoras sintéticas como un cebador PCR en conjunción con el primer opuesto diseñado a partir de dicha proteína individual o del sistema vector.

20

25

30

35

40

45

50

55

60

65