

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 763 563**

51 Int. Cl.:

C12N 15/10 (2006.01)

G01N 33/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.11.2015 PCT/GB2015/053573**

87 Fecha y número de publicación internacional: **02.06.2016 WO16083793**

96 Fecha de presentación y número de la solicitud europea: **24.11.2015 E 15816204 (0)**

97 Fecha y número de publicación de la concesión europea: **09.10.2019 EP 3224360**

54 Título: **Método**

30 Prioridad:
24.11.2014 GB 201420852

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.05.2020

73 Titular/es:
**ACTOME GMBH (100.0%)
Georges-Köhler-Allee 103
79110 Freiburg, DE**

72 Inventor/es:
JENEY, CSABA

74 Agente/Representante:
**INGENIAS CREACIONES, SIGNOS E
INVENCIONES, SLP**

ES 2 763 563 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

Método

5 **Campo de la invención**

La presente invención se refiere, en general, a métodos y usos de kits para detectar interacciones de unión, en particular interacciones proteína-proteína, y particularmente a métodos de alto rendimiento de marcado, análisis, detección y medición de interacciones proteína-proteína.

10

Antecedentes

La arquitectura celular se define por sus complejos, las máquinas moleculares que en realidad constituyen una célula. La biología celular identifica tradicionalmente proteínas basadas en sus acciones individuales como catalizadores, moléculas de señalización, o elementos estructurales de células y microorganismos. Actualmente, los presentes inventores presencian la aparición de una perspectiva post-genómica que expande la función de la proteína, considerándola también un elemento en una red de interacciones proteína-proteína, con una función 'contextual' o 'celular' dentro de los módulos funcionales. La caracterización cualitativa y cuantitativa de redes complejas de proteína-proteína y la importante identificación de proteínas de interacción específica del tipo de célula son fundamentales para entender los procesos fisiológicos y las alteraciones de interacciones proteína-proteína en una multitud de enfermedades humanas tales como cáncer, enfermedades autoinmunitarias y otros trastornos. Los conocimientos detallados en las redes proteína-proteína y la identificación de diferencias asociadas a enfermedad pueden conducir a nuevas formas para el diseño racional y el desarrollo de fármacos específicos. El patrón de interacciones proteína-proteína en una célula o tejido también se puede usar como herramienta para diagnósticos moleculares.

25

Las proteínas participan en complejas interacciones que representan la base mecánica para gran parte de la fisiología y función de la célula. Estas interacciones proteína-proteína están organizadas en redes exquisitamente complejas. Se propuso que la arquitectura de las redes de interacción proteína-proteína estaba libre de escala, teniendo la mayoría de las proteínas solo una o dos conexiones, pero con relativamente menos 'ejes' que poseían decenas, centenas o más conexiones. Las redes de interacción son altamente dinámicas, permitiendo los rápidos cambios en el interactoma, por ejemplo, a estímulos externos o incluso procesos de desarrollo. Es probable que las interacciones entre proteínas del núcleo y entre dos o más proteínas de módulo estén mediadas por interacciones dominio-dominio. Es menos probable que las interacciones dentro de y entre proteínas de unión ocurran de este modo. A pesar de la contribución de los complejos de proteína y las interacciones con la regulación y ejecución de los procesos biológicos, relativamente pocos complejos son bien entendidos en términos de estructura y función.

30

Los intentos por obtener experimentalmente constantes cinéticas para interacciones celulares son escasos. Estos parámetros cuantitativos permitirán el desarrollo de modelos cinéticos basados en ecuaciones diferenciales de procesos celulares. Dichos modelos son necesarios para el entendimiento de la acción de fármacos y promoverán el descubrimiento de nuevos fármacos para muchas enfermedades complejas. El desarrollo de modelos cuantitativos a múltiples escalas puede proporcionar un entendimiento teórico de la acción terapéutica y efectos adversos de fármacos a un nivel celular.

40

El término 'muestreo' se usa para diseños experimentales donde solo se interroga un subconjunto de la población. El muestreo representativo es común en la generación de conjuntos de datos de interacciones de proteínas, donde el muestreo ha sido frecuentemente guiado por prioridades biológicas. La 'cobertura' resume qué parte del conjunto total de interacciones posibles se ha probado en realidad. En vista de las actuales tecnologías, no es válido hacer inferencias sobre el 'interactoma', por ejemplo, el conjunto de todas las interacciones físicas que tienen lugar en una célula en las condiciones que se estudian.

50

Se han ideado varios métodos para estudiar la interacción proteína-proteína que incluyen métodos físicos para seleccionar y detectar proteínas que se unen a otra proteína, tales como cromatografía de afinidad por proteína, transferencia de afinidad, inmunoprecipitación (incluyendo electroforesis en gel 2D y espectrometría de masas), reticulación; métodos basados en bibliotecas: sondeo de proteínas, presentación en fagos, sistema de dos híbridos, otros métodos basados en bibliotecas y métodos genéticos: supresores extragenéticos, efectos letales sintéticos, fenotipos de producción en exceso, producción en exceso de proteínas naturales y producción en exceso de proteínas mutantes; y no complementación no unida.

55

Muchos de estos métodos no son aptos para el análisis de interacciones de proteínas de alto rendimiento. Las tecnologías de alto rendimiento más prometedoras están disponibles por el desarrollo de técnicas de cribado de bibliotecas de péptidos y proteínas, tales como la estrategia de dos híbridos de levadura, que es un método para identificar y clonar genes para proteínas que interactúan con una proteína de interés; matrices de dos híbridos, donde se llevan a cabo experimentos a gran escala en un formato de matrices de colonias, en el que cada colonia de levadura expresa un par definido de proteínas de 'cebo' y 'presa' que se pueden puntuar para actividad de indicador génico - que indica interacción - en un modo automatizado; presentación en fagos donde una biblioteca de

60

65

proteínas se inmunopurifica contra una proteína de "cebo" y purificación por afinidad / espectrometría de masas (AP-MS), especialmente para definir todos los complejos en la célula (el 'complejoma') y sus proteínas constituyentes; y purificación por afinidad en tándem (TAP). La TAP revela proteínas de interacción como proteínas de núcleo, módulo, o de unión, según la frecuencia de su aparición en las diversas formas de ese complejo.

5 Todos estos métodos tienen ventajas y desventajas relacionadas con la fiabilidad, completitud y facilidad de información obtenida usando de estas técnicas. El método ideal captura la información del interactoma de un modo rápido y rentable, que permite muestreo aleatorio y alta redundancia del muestreo. Proporciona cobertura dinámica, basada en el contexto celular original, basada en la interacción proteína nativa-proteína, y completa, suficientemente grande, de datos cuantitativos de interacción de complejos de proteína de múltiples unidades incluso grandes. Suprime los efectos de variables aleatorias, tales como la detección de proteínas no específicas, que interaccionan accidentalmente. Por tanto, disminuye el efecto de las variables, que son variables relacionadas con cualquier evento de unión implicado en el principio de detección distinto de la interacción proteína-proteína original.

15 Los cribados de dos híbridos, especialmente las técnicas basadas en matriz, permiten la generación de información de interactomas a gran escala. Sin embargo, existen desventajas importantes debido a su detección binaria por parejas, ausencia de información dinámica basada en contexto original, agente de unión artificial (proteínas híbridas) y el principio restringido al contexto celular de levadura (por ejemplo, modificación postraduccional sesgada en comparación con el hospedador original). Casi todos estos han sido resueltos parcialmente por diversas formas. Sin embargo, no se ha ideado un método que combine todas estas características requeridas.

Los métodos basados en afinidad, especialmente aquellos que usan espectrometría de masas como principio de detección, generan una alta cantidad de datos semi-cuantitativos del interactoma, parcialmente en el contexto celular correcto. Sin embargo, están influidos por variables aleatorias y relacionadas con la unión (afinidad). Detectan eventos de unión no específicos accidentales. Para generar un conjunto de datos muestreados aleatoriamente, de alta cobertura, complejo, se requeriría una cantidad significativa de tiempo y coste, que compromete el beneficio de su potencial para detectar la naturaleza dinámica del interactoma. Algunas de estas cuestiones han sido resueltas, especialmente usando purificación por afinidad en tándem (TAP), donde se reducen a un mínimo los eventos unión no específica accidentales; sin embargo, a costa de recuperación menos fiable de complejos proteína-proteína.

30 Estas técnicas han acelerado la generación de datos de interacciones proteína-proteína (PPI) a gran escala. Después del estudio pionero sobre el interactoma, se han llevado a cabo varios estudios a gran escala dando como resultado algunos conjuntos de datos de alta calidad de interacciones proteína-proteína por parejas. Por ejemplo, el interactoma de levadura filtrada (FYI) es una intersección de diferentes conjuntos de datos, que incluyen datos de Y2H, datos de AP-MS, predicciones *in silico*, interacciones físicas del Centro de Información de Secuencias de Proteínas de Múnich, y complejos de proteínas informados en la bibliografía.

35 Niro et al. (Nuc. Acid Res. (2010) 38:9; e100) describen un método de determinación de una interacción de unión entre una biblioteca de presentación en fagos de fragmentos del marco de lectura abierto y transglutaminasa (TG2). Después de dos rondas de inmunopurificación con TG2, los componentes de unión se secuencian con secuenciación de 454.

45 D'Angelo et al. (Protein Engineering, design and selection (2014) 27:10; 301-307) desvelan un método de determinación de una interacción de unión entre una biblioteca de fagos de scFv y dos proteínas CDK2. Se usa un conjunto de cebadores directos en combinación con códigos de barras para la unión y secuenciación.

Ravn et al. (Nuc. Acid Res. (2010) 38:21; e193.1-e193.11) desvelan un método de determinación de una interacción de unión entre una biblioteca de fagos de scFv y un anticuerpo monoclonal anti-ratón y hIFN γ biotinilado soluble. Se usa un conjunto de cebadores directos en combinación con códigos de barras para la unión y secuenciación.

50 Ravn et al. (Methods (2013) 60:1; 99-110) y el documento de patente WO2010/135558 desvelan un método de determinación de una interacción de unión entre cinco bibliotecas de fagos de scFv y dos antígenos. Un método de determinación de una interacción de unión entre una biblioteca de fagos de scFv y Ag85 se describe por D'Angelo et al. (MABS (2013) 6:1; 160-172)

55 El documento WO2010/09589 desvela un método de determinación de interacciones de unión entre anticuerpos y antígenos cribando la unión, por ejemplo, bibliotecas de anticuerpos de fagos contra bibliotecas de antígenos de fagos.

60 Bowley et al. (PNAS (2009) 105:5; 1380-1385) describen métodos de determinación de interacciones de unión entre anticuerpos y antígenos cribando la unión entre bibliotecas combinatorias de anticuerpos contra bibliotecas de antígenos eucariotas.

65 Como los enfoques metodológicos existentes no cumplen completamente las necesidades de interacción proteína-proteína y estudios de interactoma, se necesitan nuevos métodos para el análisis y las caracterizaciones de redes de complejos proteína-proteína.

La presente invención proporciona métodos y usos de kits para detectar interacciones de unión, en particular interacciones proteína-proteína al nivel celular. Los métodos y usos se pueden usar para detectar simultáneamente todas, o un subconjunto de, las proteínas de interacción en redes de complejos de proteínas, preferentemente en el contexto original de células. Los métodos y usos proporcionan cobertura dinámica, basada en el contexto celular original, basada en interacciones proteína nativa-proteína, y completa, suficientemente grande, de datos de interacción cinética cuantitativa y potencialmente cinética de complejos de proteínas de múltiples unidades incluso grandes.

Los métodos descritos en el presente documento se pueden usar para detectar interacciones proteína-proteína usando tecnología de presentación de anticuerpos, usando una pluralidad de fagos de anticuerpos como agentes de unión. Los métodos también se pueden usar para detectar interacciones proteína-proteína usando tecnología de aptámeros, usando una pluralidad de aptámeros como agentes de unión. La complejidad de la pluralidad de agentes de unión se puede variar en amplios intervalos entre algunos agentes de unión a decenas de miles o cientos de miles o millones o decenas de millones o cientos de millones de agentes de unión. Para obtener agentes de unión de baja complejidad a partir de agentes de unión de alta complejidad adecuados para el método inventado, se idea un método de reducción de la complejidad (enriquecimiento).

Se pueden identificar y monitorizar interacciones más detalladas entre las moléculas diana. Por ejemplo, se pueden detectar interacciones proteína-proteína. La presencia de dos o más agentes de unión dentro de un complejo agente de unión/diana puede indicar que dos o más dianas pueden estar presentes dentro del complejo. Esto indica que las dos o más dianas pueden estar interaccionando, o unidas, entre sí. Si se conoce una parte identificable del agente de unión específica, por ejemplo la secuencia de ácidos nucleicos, entonces se pueden identificar las dianas. Este método se puede llevar a cabo usando amplificación por PCR altamente paralela conectando las secuencias de ácidos nucleicos identificables de fagos de anticuerpo presentados unidos, es decir, aquellos con características de unión predeterminadas, por ejemplo con secuencias de epítopes conocidas, o que se conoce que se unen a una molécula específica. Esto se puede hacer preferentemente por PCR en emulsión. Esto se puede llevar a cabo a bajas concentraciones de complejo de proteína, preferentemente en compartimentos. Las interacciones entre dianas, por ejemplo interacciones proteína-proteína, se pueden detectar por amplificación por PCR altamente paralela, preferentemente usando agentes de detección de la unión de complejidad reducida.

La información de la interacción diana-diana, por ejemplo proteína-proteína, se obtiene por secuenciación de las secuencias identificables conectadas, preferentemente por secuenciación de ADN altamente paralela o por otros medios de detección de secuencias. Se puede usar variar la cantidad de material de entrada, por ejemplo, la diana, para recoger datos de cinética de unión de ligando. Además, el método se puede llevar a cabo en presencia y ausencia de compuestos para determinar si los compuestos tienen algún efecto sobre la interacción de dianas, y si este efecto es agonista o antagonista.

El método también puede usar tecnología de presentación de proteínas, que presenta fragmentos de proteínas de un organismo y que determina las características de unión de una multitud de anticuerpos presentados, teniendo cada anticuerpo información de secuencias identificables únicas y teniendo cada fragmento de proteínas presentadas información de secuencias identificables. Preferentemente, la información de secuencias identificables para los fragmentos de proteínas presentadas es la secuencia que codifica la secuencia de aminoácidos presentada. La identidad de los anticuerpos unidos se puede determinar a partir de la información de secuencias identificables para cada complejo anticuerpo-proteína. Opcionalmente, se puede identificar la identidad del fragmento de proteína unido, dentro de cada complejo anticuerpo-proteína. Opcionalmente, la identidad de los anticuerpos unidos y la identidad del fragmento de proteína unido se pueden determinar a partir de la información de secuencias identificables conectadas para cada complejo anticuerpo-proteína. También se pueden determinar las características cinéticas de unión usando diferentes cantidades de la diana, por ejemplo proteínas y agentes de unión, tales como, proteínas presentadas o anticuerpos presentados.

Los métodos y las composiciones descritas en el presente documento también se pueden usar para identificar compuestos que pueden agonizar o antagonizar dichas interacciones proteína-proteína. La presente divulgación proporciona métodos y kits para detectar interacciones de unión con compuestos antagonistas (interrupción) o agonistas (promoción). La divulgación proporciona métodos y kits para detectar simultáneamente las interacciones de unión de compuestos antagonistas y/o agonistas en redes de proteínas complejas, preferentemente en el contexto original de las células. Los métodos y kits proporcionan datos basados en el contexto celular original, basados en la interacción proteína nativa-proteína, que son completos, y tienen cobertura suficientemente grande de tanto los datos de interacción cuantitativa como, posiblemente, cinética, incluso para complejos grandes de proteínas de múltiples unidades.

Sumario de la invención

La invención proporciona un método de determinación de una interacción de unión entre un agente de unión y una diana que comprende

- a) poner en contacto una biblioteca de agentes de unión con una diana para permitir la formación de complejos

agente de unión/diana en donde cada miembro de dicha biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única;

5 b) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;

10 c) unir las secuencias de nucleótidos únicas asociadas al (a los) agente(s) de unión en el complejo agente de unión/diana para formar una secuencia de nucleótidos unida en donde el aislamiento de los complejos aislados se mantiene durante la etapa de unión;

d) identificar el (los) agente(s) de unión presente(s) en el complejo de la secuencia de nucleótidos unida.

15 La presente invención describe métodos de análisis y caracterización de interacciones de unión de complejos, en particular redes o interactomas proteína-proteína. El método se basa en la identificación relacionada con la co-localización de agentes de unión y opcionalmente sus dianas, tales como proteínas, donde la información sobre la co-localización de agentes de unión y opcionalmente sus dianas, en una pluralidad de compartimentos, se unen por parejas y se traducen en un nucleótido. La identidad del agente de unión y opcionalmente las dianas también se puede determinar a partir de la secuencia de nucleótidos. Esta información puede ser revelada por secuenciación.

20 La presente invención también describe métodos de análisis y caracterización del efecto de compuestos antagonistas (interrupción) o agonistas (promoción) sobre las interacciones de moléculas diana. El método se basa en la identificación de agentes de unión y sus dianas, tales como proteínas, en presencia y ausencia del compuesto. La detección de complejos formados entre los agentes de unión y sus dianas, y la identificación de los agentes de unión, se lleva a cabo, en una pluralidad de compartimentos, por unión por parejas de secuencias de identificación
25 únicas de los agentes específicos de unión a diana unidos que entonces se traducen en un nucleótido. La alteración de la cantidad de complejos y la identidad de los agentes de unión y opcionalmente la diana implicada se pueden revelar por secuenciación.

30 El agente de unión es preferentemente un anticuerpo, o un aptámero.

Preferentemente, el agente de unión es un miembro de una biblioteca de presentación de anticuerpos o una biblioteca de anticuerpos en donde cada anticuerpo está marcado con dicha secuencia de nucleótidos única.

35 Preferentemente, la diana también está asociada con una secuencia de nucleótidos única.

La secuencia de nucleótidos asociada a un agente de unión en el complejo se puede unir a una segunda secuencia de nucleótidos asociada a un segundo agente de unión en el complejo. Este método se puede usar para identificar una pluralidad de agentes de unión que se unen a una única diana. Por ejemplo, cuando la diana es una proteína, el método puede identificar anticuerpos que se unen a diferentes epítopes sobre la proteína. Alternativamente, la diana
40 puede ser un complejo de proteína, y el método puede identificar una pluralidad de agentes de unión que se unen a diferentes proteínas dentro del complejo. Por ejemplo, la secuencia de nucleótidos asociada a un agente de unión en el complejo agente de unión/diana se puede unir a una secuencia de nucleótidos asociada a un segundo agente de unión en el complejo agente de unión/diana. Una vez se conoce la identidad de los agentes de unión (de la secuencia unida), puede ser posible identificar los componentes de la diana, y así, por ejemplo, las proteínas en la
45 diana que interaccionan naturalmente. Por ejemplo, si el agente de unión es un anticuerpo con características de unión conocidas, se puede identificar la proteína unida por el anticuerpo. Así, se puede identificar la identidad de las proteínas dentro de la diana. Esto permite interacciones proteína-proteína dentro de la muestra que va a detectarse e identificarse. Además, una vez se ha identificado una interacción proteína-proteína, el método se puede usar para monitorizar el efecto de un compuesto sobre la interacción.

50 Alternativamente, la secuencia de nucleótidos asociada a un agente de unión en el complejo agente de unión/diana se puede unir a una secuencia de nucleótidos asociada a una diana dentro del complejo agente de unión/diana. Este método se puede usar para identificar qué agente de unión interacciona con qué diana. Por ejemplo, se puede usar para identificar qué miembros de una biblioteca de agentes de unión pueden formar un complejo con una diana conocida. Esta información se puede usar para caracterizar los miembros de una biblioteca de agentes de unión para obtener información de características de unión.

60 Preferentemente, la producción de dichos productos de ácido nucleico unidos por pares al azar comprende utilizar al menos dos pares de cebadores de PCR para amplificar amplicones idénticos o no idénticos; en donde los cebadores de PCR en el extremo 5' tienen marcas de secuencia en donde la amplificación con cebadores marcados da como resultado productos de ácido nucleico unidos por pares al azar. Más preferentemente, la amplificación es amplificación por PCR en emulsión y la producción de dichos amplicones y productos de ácido nucleico unidos por pares al azar son procesos paralelos.

65 Preferentemente, dicha secuenciación de dichos productos de amplificación unidos es un método de secuenciación altamente paralelo.

El método de la presente invención se puede usar para investigar el efecto de un compuesto sobre la interacción entre los agentes de unión y las dianas o la interacción entre dos o más moléculas diana. La etapa de poner en contacto el agente de unión con una diana se puede llevar a cabo en presencia y ausencia de un compuesto, y los resultados se comparan para determinar si el compuesto afecta la interacción de unión entre el agente de unión y la diana o entre las moléculas diana. Este método puede ser utilizado para identificar posibles agentes farmacéuticos que se pueden usar para tratar enfermedades y afecciones médicas.

La invención también proporciona el uso de un kit para llevar a cabo el método de la invención, comprendiendo el kit

(i) Una biblioteca de agentes de unión en donde cada miembro de dicha biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única; y

(ii) un conjunto de al menos dos pares de cebadores para unir las secuencias de nucleótidos asociados al agente de unión; y opcionalmente instrucciones para su uso.

El kit puede comprender además una biblioteca de presentación de proteínas en donde cada miembro de dicha biblioteca está asociado con una secuencia de nucleótidos única.

Breve descripción de los dibujos

La Figura 1 representa un principio general del ensayo para detectar interacciones proteína-proteína usando métodos y composiciones de la presente invención. Se usan fagos de una biblioteca de anticuerpos con información de características de unión predeterminadas para revelar la información de unión de proteínas que interaccionan. La información de unión se determina en una pluralidad de compartimentos por dimerización por PCR. Los fagos se lisan durante el proceso liberando su ADN único. Estas secuencias únicas se amplifican usando cebadores universales y se dimerizan. Los productos dimerizados que codifican la información de unión se secuencian por secuenciación de nueva generación (NGS) y se determinan las identidades de proteína unida basándose en la detección de fagos específicos, con características de unión conocidas que incluyen las identidades de sus dianas de unión reconocidas.

La Figura 2 representa un principio general del ensayo para la caracterización de un agente de unión - para asociar la identidad de marca del ácido nucleico del agente de unión con su información de características de unión que incluye la identidad de sus dianas de unión reconocidas usando métodos y composiciones de la presente invención. La presentación de ADNc y los fagos de presentación de anticuerpos se unen entre sí. La información de unión se determina en una pluralidad de compartimentos por dimerización por PCR. Los fagos se lisan durante el proceso liberando su ADN único. Estas secuencias únicas se amplifican usando cebadores universales y se dimerizan. Los productos dimerizados que codifican la información de unión se secuencian por secuenciación de nueva generación (NGS) y se determina la identidad de proteínas de ADNc a partir de la secuencia por búsqueda en base de datos y se asocia con la identidad de marca del ácido nucleico del agente de unión.

Descripción detallada

El método ideal captura información del interactoma de un modo rápido y rentable, que permite muestreo aleatorio y alta redundancia del muestreo. El método proporciona cobertura completa de datos de interacción cuantitativa, incluso para grandes complejos de proteína de múltiples unidades. Estos datos se obtienen en un contexto celular original, por lo que se pueden medir las interacciones proteína nativa-proteína, y se pueden usar para detectar interacciones dinámicas. El método suprime los efectos de variables aleatorias, tales como detección no específica, interacción accidental de proteínas. También disminuye el efecto de variables, que son cualquier efecto relacionado con el evento de unión implicado en el principio de detección distinto de la interacción proteína-proteína original, por ejemplo auto-unión o una unión específica.

Una realización de la presente invención se resume en la Figura 1 y los diferentes componentes del sistema de ensayo se describen con detalle más adelante. Una realización adicional de la presente invención se resume en la Figura 2 y los componentes adicionales del sistema se describen con detalle más adelante.

La invención proporciona un método de determinación de una interacción de unión entre un agente de unión y una diana que comprende

a) poner en contacto una biblioteca de agentes de unión con una diana para permitir la formación de complejos agente de unión/diana en donde cada miembro de dicha biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única;

b) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;

c) unir las secuencias de nucleótidos únicas asociadas al (a los) agente(s) de unión en el complejo agente de unión/diana para formar una secuencia de nucleótidos unida en donde el aislamiento de los complejos aislados se mantiene durante la etapa de unión;

5

d) identificar el (los) agente(s) de unión presente(s) en el complejo de la secuencia de nucleótidos unida.

El método se puede llevar a cabo *ex vivo*, *in vivo* o *in vitro*.

10 Agente de unión

Preferentemente, el agente de unión es un anticuerpo, aptámero, o basado en un armazón de proteína manipulada. Alternativamente, el agente de unión puede ser un compuesto. El agente de unión puede ser un miembro de una biblioteca de presentación de anticuerpos o una biblioteca de anticuerpos, en donde cada anticuerpo está marcado con una secuencia de nucleótidos única. El método puede usar agentes de anticuerpo presentados como agente de unión, donde se conocen las características de unión, por ejemplo, la diana a la que se une el agente de unión, y se determinan las secuencias de nucleótidos únicas asociadas a la pluralidad de agentes de anticuerpo presentados y se correlacionan entre sí las características de unión y secuencias de nucleótidos únicas. Así, la invención proporciona métodos de determinación de las características de unión y que relacionan estas con la secuencia de nucleótidos única identificable de la pluralidad de agentes de anticuerpo presentados. Esto proporciona información de características de unión.

15

20

El agente de unión usado en la invención puede ser un anticuerpo. El término "anticuerpo", como se usa en el presente documento, se refiere a moléculas de inmunoglobulina y porciones inmunológicamente activas de moléculas de inmunoglobulina, es decir, moléculas que contienen un sitio de unión al antígeno que se une específicamente a un antígeno, tanto natural como parcialmente o completamente sintéticamente producido. El término "anticuerpo" incluye fragmentos de anticuerpos, derivados, equivalentes funcionales y homólogos de anticuerpos, anticuerpos humanizados, que incluyen cualquier polipéptido que comprende un dominio de unión de inmunoglobulina, tanto natural como completa o parcialmente sintética, y cualquier polipéptido o proteína que tenga un dominio de unión que es, o es homólogo a, un dominio de unión de anticuerpo. Por tanto, se incluyen las moléculas quiméricas que comprenden un dominio de unión de inmunoglobulina, o equivalente, fusionado con otro polipéptido. La clonación y expresión de anticuerpos quiméricos se describen en los documentos de patente EP-A-0120694 y EP-A-0125023. Los ejemplos de anticuerpos son los isotipos de inmunoglobulina (por ejemplo, IgG, IgE, IgM, IgD e IgA) y sus subclases isotípicas; fragmentos que comprenden un dominio de unión al antígeno tales como Fab, scFv, Fv, dAb, Fd; y diacuerpos. Los anticuerpos pueden ser policlonales o monoclonales.

25

30

35

Las regiones determinantes de la complementariedad (CDRs) son parte de las cadenas variables en las inmunoglobulinas (anticuerpos), generadas por linfocitos B, donde estas moléculas se unen a su antígeno específico. Como las partes más variables de las moléculas, las CDRs son cruciales para la diversidad de las especificidades de antígeno generadas por inmunoglobulinas. Existen tres CDRs (CDR1, CDR2 y CDR3), dispuestas no consecutivamente, sobre la secuencia de aminoácidos de un dominio variable de una inmunoglobulina. Puesto que las inmunoglobulinas normalmente están compuestas de dos dominios variables (sobre dos cadenas de polipéptidos diferentes, cadena pesada y ligera), existen seis CDRs para cada receptor de antígeno que se pueden poner conjuntamente en contacto con el antígeno.

40

45

Se ha mostrado que los fragmentos de un anticuerpo completo pueden realizar la función de antígenos de unión. Los ejemplos de fragmentos de unión son (i) el fragmento Fab que consiste en los dominios VL, VH, CL y CH1; (ii) el fragmento Fd que consiste en los dominios VH y CH1; (iii) el fragmento Fv que consiste en los dominios VL y VH de un único anticuerpo; (iv) el fragmento dAb (Ward, E.S. et al., Nature 341:544-546 (1989)) que consiste en un dominio VH; (v) CDR de regiones aisladas; (vi) fragmentos F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab unidos; (vii) moléculas Fv monocatenarias (scFv), en donde un dominio VH y un dominio VL se unen por un conector peptídico que permite que los dos dominios se asocien para formar un sitio de unión al antígeno (Bird et al., Science 242:423-426 (1988); Huston et al., PNAS USA 85:5879-5883 (1988)); (viii) dímeros Fv monocatenarios biespecíficos (documento de patente PCT/US92/09965) y (ix) "diacuerpos", fragmentos multivalentes o multiespecíficos construidos por fusión génica (documento de patente WO94/13804; P. Hollinger et al., Proc. Natl. Acad. Sci. USA 90: 6444-6448 (1993)).

50

55

Un "dominio de unión al antígeno" es la parte de un anticuerpo que comprende el área que se une específicamente y es complementaria a una parte o todo un antígeno. Donde un antígeno es grande, un anticuerpo solo se puede unir a una parte particular del antígeno, parte que se llama un epítipo. Un dominio de unión al antígeno se puede proporcionar por uno o más dominios variables de anticuerpo. Un dominio de unión al antígeno puede comprender una región variable de la cadena ligera del anticuerpo (VL) y una región variable de la cadena pesada del anticuerpo (VH).

60

65

Alternativamente, los agentes de unión se pueden basar en armazones de proteínas manipuladas. Los armazones de proteína derivan de estructuras estables, solubles, de proteína natural que se han modificado para proporcionar

un sitio de unión para una molécula diana de interés. Los ejemplos de armazones de proteínas manipuladas incluyen, pero no se limitan a, anticuerpos, que se basan en el dominio Z de la proteína A estafilocócica que proporciona una interfase de unión sobre dos de sus hélices α (Nygren, P. A. (2008). FEBS J 275(11): 2668-76); anticalinas, derivadas de lipocalinas, que incorporan sitios de unión para ligandos pequeños en el extremo abierto de un pliegue de barril beta (Skerra, A. (2008) FEBS J 275(11): 2677-83), nanocuerpos y DARPinas. Los armazones de proteínas manipuladas normalmente son dirigidos para la unión a las mismas proteínas antigénicas que los anticuerpos. También se pueden usar péptidos cortos para unir a una proteína diana. Los filómeros son péptidos estructurados naturales derivados de genomas bacterianos. Dichos péptidos representan una diversa matriz de pliegues estructurales de proteína y se pueden usar para inhibir/alterar interacciones proteína-proteína *in vivo* (Watt, P. M. (2006). Nat Biotechnol 24(2): 177-83)].

Alternativamente, el agente de unión puede ser un aptámero. Los aptámeros son oligonucleótidos sintéticos (ADN o ARN) que reconocen moléculas diana con alta afinidad y especificidad mediante una combinación de complementariedad de forma y enlaces químicos no covalentes (Blank & Blind, Current Opin. Chem. Biol., 2005, 9:336-342). Estos ligandos artificiales son bastante fáciles de obtener *in vitro* y se pueden desarrollar para reconocer una gran variedad de diferentes clases de moléculas que varían desde simples iones (por ejemplo Pb^{2+} , Liu & Lu, 2003. J Am Chem Soc., 125, 6642-6643) hasta nucleótidos, moléculas pequeñas, proteínas, virus, y células hasta organismos completos (Menger et al., 2006. Handbook of Experimental Pharmacology, 359-373). Se han seleccionado aptámeros de alta afinidad de unión mediante el método bien conocido SELEX (Ellington & Szostak, 1990. Nature, 346, 818-822) para la detección de moléculas de bajo peso molecular como teofilina (Jenison et al., 1994. Science, 263, 1425-1429), L-arginina (Geiger et al., 1996. Nucl. Acids Res., 24, 1029-1036), menomicina (Schuerer et al., 2001. Bioorg. Med. Chem., 92, 2557-2563), 17 β -estradiol (Kim et al., 2007. Biosens. Bioelectron., 22, 2525-2531), pero también para moléculas mayores como trombina (aptámero de unión a trombina: 5'-GGTTGGTGTGGTTGG-3') (Baldrich et al., Anal Chem. 2004, 76, 23,7053-63), toxina del cólera o proteína tat del VIH-1, entre otros (para una revisión véase Tombelli et al., 2007, Biomolec Eng., 24, 191-200). Algunos de los aptámeros anteriormente mencionados se han usado en ensayos de tipo ELISA sobre microplacas o sobre la superficie de transductores de biosensores (QCM, SPR). También se ha desarrollado un sistema colorimétrico de AuNP modificado por aptámero para la determinación de la proteína PDGF en un ensayo basado en sándwich (Huang et al., 2005, 77, 5735-5741).

El agente de unión puede ser parte de una biblioteca, tal como una biblioteca de agentes de unión presentados, por ejemplo bibliotecas de presentación bacteriana, presentación en ARNm, presentación en bacteriófagos, de aptámeros, presentación en ribosomas o de presentación en levaduras. Preferentemente, la biblioteca de agentes de unión presentados es una biblioteca de anticuerpos de presentación en bacteriófagos. La biblioteca debe ser lo suficientemente grande tal que la biblioteca consista en una pluralidad de miembros de unión que se espera que se unan a al menos 75 % de las dianas de interés dentro de una muestra de diana. Más preferentemente, la biblioteca se diseña para unirse a al menos 80 %, 85 %, 90 %, 95 %, 97,5 % o 99 % de las dianas de interés dentro de una muestra. Por ejemplo, la biblioteca de agentes de unión comprende una pluralidad de miembros de unión a secuencias de proteínas o de péptidos con 95 % o mayor cobertura de proteínas esperadas o deseadas dentro de una muestra. Dichas bibliotecas se publican en la bibliografía. Cada miembro de la biblioteca tiene una marca de identidad de ácido nucleico detectable, que preferentemente es única para un miembro de la biblioteca. Preferentemente, las marcas de identidad de ácido nucleico únicas están unidas. "Unido" significa que el proceso de unión tiene el potencial de formar productos de ácido nucleico multiméricos al azar basándose en la co-localización de estas marcas de identidad de ácido nucleico en condiciones de ensayo adecuadas. Preferentemente, el producto multimérico es un dímero. Las condiciones de ensayo adecuadas incluyen el desmantelamiento de las partículas de bacteriófagos, en compartimentos separados, por ejemplo por tratamiento térmico en emulsión de lípidos, y la amplificación consenso específica de las secuencias únicas produce amplicones unibles. La unión de los amplicones unibles, por ejemplo, dominios de ácido nucleico específicos de presentación de la unión, forman marcas de identidad unidas, que codifican la información de co-localización de las marcas de identidad. Preferentemente, la secuencia única es los dominios de ácido nucleico específicos de presentación de la unión, por ejemplo, la secuencia que codifica una o más regiones CDR. La reacción de unión puede estar basada en la amplificación o implicar otras técnicas.

La unión basada en amplificación puede utilizar dos o más pares de cebadores de amplificación con capacidades de unión idénticas, pero con marcas de 5' complementarias o secuencias conectoras de dímero que dan como resultado la formación de dúplex de ácido nucleico extensibles por polimerasa. Las marcas o secuencias conectoras de dímeros significa que la secuencia amplificada por un par de cebadores se hibridará con secuencias amplificadas por el segundo par de cebadores. Así se unen las marcas de identidad.

Cada miembro de la biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única, que se puede usar para identificar el agente de unión. "Asociado", como se usa en el presente documento, significa que la presencia del agente de unión en el complejo se puede detectar por la presencia de la secuencia de ácidos nucleicos dentro de la secuencia unida generada en el método. La secuencia de nucleótidos se puede unir como una marca al agente de unión, ser parte del propio agente de unión, por ejemplo, aptámero, o estar presente dentro del agente de unión, por ejemplo, ácido nucleico dentro de un fago. Por ejemplo, cada miembro de la biblioteca puede ser marcado con la secuencia de nucleótidos única. Como se usa en el presente documento, "marcado" se

refiere a una secuencia de nucleótidos que está unida al miembro de la biblioteca. Se conocen en la técnica métodos de unión de nucleótidos a agentes de unión tales como anticuerpo o compuestos. Alternativamente, si la biblioteca de agentes de unión es una biblioteca de presentación, como se ha descrito anteriormente, la secuencia de nucleótidos única puede ser la secuencia que codifica una o más regiones CDR o el dominio de unión
 5 presentado. Por ejemplo, se puede generar una biblioteca de presentación insertando secuencias que codifican la secuencia de aminoácidos a presentar en un fago en una localización conocida. Entonces se pueden usar cebadores universales que amplificarán las secuencias insertadas y así identifican la secuencia de unión. Alternativamente, si el agente de unión es un aptámero, el propio aptámero puede ser la secuencia de nucleótidos única.

10 La secuencia de nucleótidos es un oligonucleótido y puede comprender ARN o ADN, monocatenario o bicatenario. Los nucleótidos usados para marcar el agente de unión o gen diana tienen, en general, 5-150 bases de longitud, por ejemplo 10-40, o 20-30 bases de longitud. Los nucleótidos que forman el ácido nucleico pueden ser químicamente modificados para aumentar la estabilidad de la molécula, para mejorar su biodisponibilidad, o para conferir actividad
 15 adicional a él. Por ejemplo, las bases de pirimidina se pueden modificar en las posiciones 6 u 8, y las bases de purina en la posición 5 con CH₃ o halógenos tales como I, Br o Cl. Las modificaciones de bases de pirimidinas también incluyen 2 NH₃, O⁶-CH₃, N⁶-CH₃ y N²-CH₃. Las modificaciones en la posición 2' son modificaciones de azúcar e incluyen normalmente un grupo NH₂, F o OCH₃. Las modificaciones también pueden incluir modificaciones de 3' y 5' tales como encapuchado.

20 Alternativamente, se pueden usar nucleótidos modificados, tales como nucleótidos de morfolino, ácidos nucleicos bloqueados (LNA) y ácidos nucleicos peptídicos (PNA). Los nucleótidos de morfolino se ensamblan a partir de diferentes subunidades de morfolino, conteniendo cada una de ellas una de las cuatro bases genéticas (adenina, citosina, guanina y timina) unidas a un anillo de morfolina de 6 miembros. Las subunidades se unen por enlaces no
 25 iónicos de inter-subunidades de fosfordiamidato dando un oligonucleótido de morfolino. Los monómeros de LNA se caracterizan por que la conformación de anillo de furanosa está restringida por un conector de metileno que conecta la posición 2'-O con la posición 4'-C. PNA es un análogo de ADN en el que el esqueleto es un pseudopéptido en vez de un azúcar.

30 Preferentemente, los agentes de unión son capaces de detectar más de una diana, preferentemente con diferentes afinidades aparentes. Alternativamente, los agentes de unión son capaces de detectar una única diana usando diferentes epítopes o sitios de unión, preferentemente con diferentes afinidades aparentes.

35 Se pueden predeterminar las características de unión de los miembros de la biblioteca de anticuerpos en fagos. Por ejemplo, se puede determinar qué epítipo se une por las CDRs que codificaron y expresaron el agente de unión (anticuerpo) del fago. Esta información se puede asociar a la secuencia de nucleótidos única que codifica las CDRs. Así, el epítipo unido por el anticuerpo expresado por el fago se puede identificar a partir de la secuencia de la secuencia de nucleótidos única. Una vez se conoce la secuencia de epítipo presente en la diana unida, puede ser posible identificar la proteína o el grupo de proteínas unidas.

40 Puede ser posible determinar las características de unión de los miembros de la biblioteca de anticuerpos en fagos usando los epítopes marcados en la secuencia de nucleótidos única o bibliotecas de epítopes marcados en la secuencia de nucleótidos única.

45 Diana

"Diana", como se usa en el presente documento, es la molécula o grupo de moléculas que forma un complejo con el agente de unión. El complejo se forma normalmente en condiciones fisiológicas normales del organismo de interés.

50 Preferentemente, la diana comprende una proteína. Más preferentemente, la diana es parte de una muestra de proteína. La muestra de proteína puede comprender una biblioteca de presentación de proteínas, preferentemente en donde cada miembro de dicha biblioteca está asociado con una secuencia de nucleótidos única. Preferentemente, la biblioteca de presentación de proteínas es una biblioteca de presentación en fagos de ADNc. Opcionalmente, la diana puede estar reticulada con otras dianas dentro de una pluralidad de dianas, por ejemplo,
 55 una muestra de proteína. Por ejemplo, una proteína dentro de una muestra puede estar reticulada con una o varias de otras proteínas dentro de la muestra.

60 La diana puede ser una diana conocida. Se pueden identificar agentes de unión que forman un complejo con la diana, que incluyen compuestos que interactúan con la diana. Alternativamente, la diana puede ser desconocida, y el método de la invención se usa para identificar la diana, o una pluralidad de moléculas diana que interactúan entre sí.

65 La diana se puede asociar a una secuencia de nucleótidos única. "Asociado" significa que la presencia de la diana dentro del complejo agente de unión/diana se puede detectar por la presencia de la secuencia de ácidos nucleicos dentro de la secuencia unida generada por el método. La secuencia de nucleótidos se puede unir como una marca a la diana, o estar presente dentro de la diana, por ejemplo, ácido nucleico dentro de un fago. Alternativamente, la

secuencia de nucleótidos puede ser parte de un aptámero conocido por unirse a la diana. Los complejos agente de unión/diana se pueden poner en contacto con los aptámeros para permitir que la diana presente se identifique mediante enlace de las secuencias de nucleótidos únicas, que incluyen el aptámero.

5 El ensayo de la presente invención se puede aplicar a cualquier muestra de proteína. Las proteínas pueden derivar de cualquier espécimen biológico que incluye, pero no se limita a, tejidos, especímenes citológicos, líquidos corporales, cultivos celulares, o cualquier otro material que contenga complejo de proteína. Las muestras de líquido corporal incluyen sangre, saliva, orina, líquido cefalorraquídeo, o suero. Alternativamente, la muestra se puede generar por métodos de expresión recombinante. La preparación de proteínas de especímenes se puede realizar usando métodos convencionales conocidos en la técnica. El espécimen puede ser químicamente tratado antes de la extracción, por ejemplo, se pueden usar diferentes agentes químicos fijadores o agentes de reticulación (por ejemplo, BS3 - (bis(sulfosuccinimidil)suberato). La muestra de proteína puede estar reticulada o sin reticular. Alternativamente, se pueden producir proteínas, por ejemplo, por sistemas de transcripción-traducción *in vitro*, o por sistemas de expresión recombinante. Dependiendo del objetivo experimental y el tipo de interacción proteína-proteína en investigación, las proteínas se pueden analizar o en su forma desnaturalizada o no desnaturalizada, y/o forma reticulada o no reticulada. La muestra de proteína se puede analizar en una pluralidad de condiciones para recoger información sobre las características de unión cuantitativas de la pluralidad de interacciones proteína-proteína. Por ejemplo, se puede variar la concentración o cantidad del agente de unión para determinar constantes de disociación y otros parámetros cinéticos.

20 Se puede preseleccionar la mezcla de proteínas. Por ejemplo, la mezcla de proteínas puede ser un enriquecimiento de proteínas específicas, por ejemplo, proteínas de una localización celular específica, de un tipo específico de célula, de un tamaño o carga electrostática similar, proteínas con propiedades de unión similares, características de secuencia similares, o funciones similares, por ejemplo, enzimas (Current Protocol in Molecular Biology (2006) 20.0.1-20.0.6 CAPÍTULO 20 Analysis of Protein Interactions). Preferentemente, las proteínas específicas comprenden fosfoproteínas, proteínas de membrana o proteínas naturalmente, postraduccionalmente, artificialmente modificadas. Las proteínas en la mezcla de proteínas del método se pueden desnaturalizar o no desnaturalizar y/o reticular o no reticular.

30 La proteína puede estar en forma de una biblioteca de presentación de proteínas. Los ejemplos comprenden bibliotecas de presentación bacteriana, presentación en ARNm, presentación en bacteriófagos y presentación en ribosomas, y presentación en levadura. Preferentemente, la biblioteca de presentación de proteínas es una biblioteca de presentación en bacteriófagos de proteínas, más preferentemente una biblioteca de presentación en fagos de ADNc. La biblioteca debe ser lo suficientemente grande tal que consista en una pluralidad de miembros de péptido o de proteína con al menos 70 % de cobertura de las proteínas que se espera detectar por el método en una muestra. Más preferentemente, la biblioteca es lo suficientemente grande como para proporcionar 75 %, 80 %, 85 %, 90 %, 95 %, 97,5 %, 99 % o mayor cobertura de las proteínas o péptidos en una muestra. Preferentemente, la biblioteca de presentación proporciona cobertura de cualquier entidad biológica adecuada, por ejemplo, una muestra de tejido u organismo completo, por ejemplo 95 % o más de cobertura de proteína de cualquier entidad biológica adecuada. Dichas bibliotecas se publican en la bibliografía (Danner S, Belasco JG. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc Natl Acad Sci U S A. 2001 Nov 6;98(23):12954-9. Epub 2001 Oct 23. PubMed PMID: 11606722; PubMed Central PMCID: PMC60806.). Cada miembro de la biblioteca está asociado con una secuencia de nucleótidos única, es decir, cada miembro tiene marcas de identidad de ácido nucleico detectables únicas. Preferentemente, las marcas de identidad de ácido nucleico únicas están unidas. "Unido" significa que el proceso de unión tiene el potencial de formar productos de ácido nucleico multiméricos al azar basándose en la co-localización de estas marcas de identidad de ácido nucleico en condiciones de ensayo adecuadas. Preferentemente, el producto multimérico es un dímero. Las condiciones de ensayo adecuadas incluyen el desmantelamiento de las partículas de bacteriófago, preferentemente en compartimentos separados, por ejemplo, por tratamiento térmico en emulsión de lípidos, y la amplificación consenso específica de las secuencias únicas producen amplicones unibles. La unión de los amplicones unibles, por ejemplo, dominios de ácido nucleico específicos de presentación de unión, forman marcas de identidad unidas, que codifican la información de co-localización de las marcas de identidad. La reacción de unión puede estar basada en amplificación o implicar otras técnicas. La unión basada en amplificación puede utilizar dos o más pares de cebadores de amplificación con capacidades de unión idénticas, pero con marcas de 5' complementarias o secuencias conectoras de dímeros que dan como resultado la formación de dúplex de ácidos nucleicos extensibles por polimerasa. Las marcas o secuencias conectoras de dímeros significan que la secuencia amplificada por un par de cebadores se hibridará con secuencias amplificadas por el segundo par de cebadores. Así se unen las marcas de identidad.

60 La marcas de identidad, es decir, las secuencias de nucleótidos únicas asociadas usadas en la biblioteca de agentes de unión y la biblioteca de dianas, tales como la biblioteca de presentación de proteínas y la biblioteca de anticuerpos, pueden ser diferentes en su origen biológico, y así el proceso de amplificación y unión se basa en dos pares de cebadores diferentes, por ejemplo, un par de cebadores amplifica secuencias diana, tales como marcas de identidad basadas en ADNc, y el segundo par de cebadores amplifica secuencias de nucleótidos específicas de agente de unión usadas como marcas de identidad. La unión de las diferentes marcas hace posible unir la información específica de agente de unión con la información de diana, por ejemplo, proteínas codificadas por

ADNcs presentados. Un ejemplo de este proceso se muestra en la Figura 2.

Un agente de unión, preferentemente un fago de anticuerpo presentado, puede reconocer una diana específica, por ejemplo, una proteína diana de presentación y la proteína correspondiente. Esto se llama especificidad. Alternativamente, una pluralidad de agentes de unión puede reconocer una diana, tal como una diana específica, por ejemplo, proteína de presentación y la proteína correspondiente. Esto se llama redundancia. Similarmente, un agente de unión puede reconocer más de una diana, tal como una especie de proteína, basándose en la similitud de la conformación de diana debido a, por ejemplo, la conformación de proteínas o secuencias de proteínas. Este fenómeno se llama reactividad cruzada. Además, el reconocimiento del agente de unión de una proteína diana se basa en la conformación de la proteína o su secuencia de proteínas. Esto se conoce como su reactividad. Las afinidades de unión de proteínas de agentes de unión tales como los agentes de unión presentados se pueden calcular a partir de la información cuantitativa de los conjuntos de datos de secuenciación. Las características de unión predeterminadas de los miembros de agentes de unión presentados pueden incluir reactividad y reactividad cruzada con especificidad y redundancia con afinidades calculadas.

Se pueden usar estas medidas para calcular las interacciones diana-diana detectadas, por ejemplo, proteína-proteína. Por cálculo basado en entradas de reactividad, especificidad y redundancia y las identidades y abundancia de marcas de identidad unidas, se adquiere información de interacciones diana-diana específicas, por ejemplo, proteína-proteína. Similarmente, usando estas medidas se puede disminuir la incertidumbre de los cálculos inducidos por la reactividad cruzada, en donde se tienen en cuenta la redundancia y las afinidades de agentes de anticuerpo presentados. Se puede usar una concentración variada de agente de unión y/o diana para calcular parámetros cuantitativos para la pluralidad de interacciones, tales como interacciones proteína-proteína. Así, el método se lleva a cabo preferentemente usando diferentes concentraciones del agente de unión y/o diana.

La naturaleza cuantitativa de la detección permite la determinación y el cálculo de las lecturas de secuenciación no informativas de origen, producidas por la co-localización no específica de marcas de identidad o auto-unión de las mismas marcas de identidad. Sin embargo, la detección de marcas de auto-unión contiene información sobre la calidad de los conjuntos de datos.

Para lograr cobertura suficiente, se pueden usar bibliotecas enriquecidas. Se pueden enriquecer la biblioteca de agentes de unión y/o la biblioteca de dianas. Las dianas presentadas, tales como las proteínas, se pueden enriquecer para cubrir todos los posibles componentes de unión en el contexto experimental. Similarmente, se pueden seleccionar los agentes de unión presentados para tener especificidades de unión enriquecidas hacia las dianas detectables en el contexto experimental. Por ejemplo, la biblioteca se puede limitar a dianas de interés específicas. Esto se puede hacer por selección basada en inmunopurificación, por ejemplo, se inmoviliza sobre una superficie sólida una muestra de diana con complejidad controlada y se pone en contacto con la biblioteca de agentes de unión seleccionando los agentes unidos de la biblioteca de agentes de unión por lavado y posterior elución de los agentes unidos. Similarmente, se inmoviliza una muestra de agente de unión con complejidad controlada sobre la superficie sólida y se pone en contacto con la biblioteca de dianas seleccionando las dianas unidas de la biblioteca de dianas por lavado y posterior elución de dianas unidas.

Preferentemente, enriquecimiento significa la retirada de agentes de unión presentados de auto-unión, es decir, retirada de los agentes de unión que se unen a otros agentes de unión en condiciones de ensayo. Más preferentemente, enriquecimiento significa reducir la complejidad de las bibliotecas de agentes de unión presentados, pero asegurando de que todavía tenga alta cobertura para la detección de dianas en el contexto experimental. La reducción de la complejidad puede implicar reducir el número total de miembros dentro de la biblioteca, eliminar miembros que se unen a proteínas no diana, o seleccionar solo los miembros que se unen a una diana de interés.

Se puede enriquecer una biblioteca de agentes de unión obteniendo una mezcla de proteínas con componentes de interés, por ejemplo, estudiando interacciones proteína-proteína conocidas y validando dinámicamente sus interacciones, o probando el efecto de compuestos agonistas o antagonistas. Por ejemplo, la inmovilización de una diana de interés, por ejemplo mezcla de proteínas sobre una superficie sólida, permite la unión de los miembros de una biblioteca de agentes de unión a la diana deseada, por ejemplo, proteína. Los agentes de unión unidos a la diana de interés se pueden separar de los agentes de unión no unidos por dilución u otros medios. Los miembros unidos de la biblioteca de agentes de unión se pueden eluir y usar, como una biblioteca de complejidad reducida, como se ha descrito anteriormente, en el método de la invención.

El uso de bibliotecas de agentes de unión enriquecidas es, en general, ventajoso ya que los métodos de separación (por ejemplo, micromatrices o separaciones basadas en emulsión) tienen capacidad limitada para producir complejos individuales dentro de un compartimento, de manera que se puedan obtener secuencias identificables unidas. La reducida complejidad de las bibliotecas de agentes de presentación enriquecidas es directamente traducible en el número de separaciones que se necesitan llevar a cabo para obtener los productos de amplificación basados en separación, emparejados aleatoriamente, unidos.

Unión de agentes de unión

El método de la invención se lleva a cabo preferentemente en condiciones fisiológicas, de manera que las interacciones se pueden detectar en su contexto original, es decir, en las mismas condiciones que se presentan en la célula. Esto proporciona información sobre las interacciones de unión que ocurren naturalmente.

5 La etapa de poner en contacto la biblioteca de agentes de unión con la diana se lleva a cabo normalmente en sistemas tampón conocidos, por ejemplo, en sistemas tampón que ya se han usado para estudios de interacciones proteína-proteína (por ejemplo, TBST-tampón). Dependiendo de la afinidad, la reacción se puede llevar a cabo a temperatura ambiente o 4 °C. Para conseguir señales reproducibles, se determinan el tiempo óptimo, temperatura
10 óptima, y otras condiciones de ensayo que incluyen las etapas de unión, lavado y detección. El experto en la técnica puede determinar condiciones óptimas.

Separación de complejos de proteína con los agentes de anticuerpo presentados unidos

15 Los complejos agente de unión/diana necesitan ser separados, es decir, aislados de otros complejos antes de la unión de las secuencias de marca, es decir, las secuencias de nucleótidos únicas asociadas. La separación se lleva a cabo por métodos conocidos en la técnica. Por ejemplo, la separación se puede llevar a cabo por dilución, unión específica, o separación por propiedades físicas y/o químicas. Los complejos se separan en compartimentos, tales como gotitas de emulsión, micro-cavidades etc., preferentemente compartimentos separados o limitados a la
20 difusión.

Preferentemente, dicha separación comprende uno cualquiera o más de unión a superficie sólida, dilución o separación de fases, entre otros, o proporcionando compartimentos separados o limitados a la difusión. La separación limita el número de agentes de unión no unidos en un compartimento, a uno, de media. Por ejemplo, el número medio de agentes de unión no unidos a un compartimento es uno. Los compartimentos pueden ser gotitas individuales dentro de una emulsión, o cámaras físicas individuales, tales como micro-cavidades. Los complejos se pueden separar según características físicas o químicas. Preferentemente, dicha dilución es dilución limitada.

25 Esta compartimentalización (por ejemplo, separación eficaz o aislamiento de grandes números de reacciones) se basa en la separación basada en la distribución de Poisson de fagos no unidos individuales; por ejemplo, emulsiones y micromatriz son el mejor método conocidos del estado de la técnica.

La separación de los complejos agente de unión/diana suficientemente antes del análisis adicional proporcionará las circunstancias donde se generan pares de marcas de ácido nucleico, que se basan en la co-localización de agentes de unión. Si los complejos no se separan suficientemente, se unirán las marcas de ácido nucleico de miembros de diferentes complejos y así proporcionarán información falsa. La separación reduce la cantidad de co-localizaciones no específicas de marcas de identidad de ácido nucleico y permite la identificación de componentes de unión específica, especialmente cuando se investigan mezclas de proteínas complejas. Por ejemplo, la separación puede resultar, de media, en un único agente de unión no unido por compartimento, donde la unión proporcionará solo
35 secuencias de nucleótidos auto-unidas, por consiguiente reducirá la posibilidad de unión aleatoria entre los miembros de la biblioteca de agentes de unión. Como la distribución de cualquier agente se basa en la distribución de Poisson, se pueden calcular las medidas necesarias para lograr la separación adecuada de agentes. Esto da preferentemente como resultado un único complejo dentro de un compartimento. Preferentemente, se usa una emulsión, que puede ser utilizada en la amplificación y la unión de dominios de ácido nucleico específicos para formar marcas de identidad unidas. El experto en la técnica conoce bien los métodos de amplificación en emulsión, por ejemplo, Schütze et al., Anal. Biochem. 1 de marzo de 2011; 410 (1):155-7.

Superficie sólida para la inmovilización

50 Opcionalmente, el complejo agente de unión/diana se inmoviliza tras la separación. Por ejemplo, los complejos agente de unión/diana pueden ser capturados sobre una superficie, por ejemplo como parte de una matriz. Esto puede ayudar a mantener la separación entre los complejos. Preferentemente, la separación de los complejos se mantiene, por ejemplo, durante la etapa de unión.

55 Los complejos agente de unión/diana se inmovilizan opcionalmente sobre superficies de soporte sólido, que incluyen, pero no se limitan a, membranas, por ejemplo, poli(fluoruro de vinilideno) (PVDF) o nitrocelulosa, superficies de plástico (por ejemplo, poliestireno) o pueden ser covalentemente acopladas a perlas apropiadas (por ejemplo, perlas activadas con epoxi). La unión o acoplamiento a superficies sólidas se realiza por métodos convencionales para proteínas ("Antibodies, a Laboratory Manual." Harlow, E., y Lane, D., eds. Cold Spring Harbor Press, Cold Spring Harbor N.Y., 1988), o por unión específica que incluye anticuerpos u otra interacción de unión específica, por ejemplo, biotina-avidina.

65 La inmovilización puede comprender (a) obtener soporte sólido con capacidad de unión universal de complejos de proteína reconocidos, es decir, el soporte sólido es capaz de unir todos los miembros de la biblioteca de agentes de unión o moléculas diana. La cantidad de complejos de proteína reconocidos y el número de sitios de unión disponibles sobre el soporte sólido deben ser equilibrados para lograr separación suficiente entre complejos de

proteína reconocidos unidos. Preferentemente, dicho soporte sólido comprende una membrana, una superficie de plástico, o una perla. Más preferentemente, el soporte sólido es una perla y la separación se logra en donde de media se une un complejo agente de unión/diana a una perla. Más preferentemente, la capacidad de unión universal de complejos de proteína reconocidos se proporciona por un anticuerpo anti-bacteriófago. Preferentemente, la separación comprende separación física en cámaras de reacción o separación fisicoquímica en gotitas dentro de una emulsión.

Preferentemente, la producción de dichos productos de ácido nucleico unidos emparejados al azar comprende utilizar al menos dos pares de cebadores de PCR para amplificar amplicones idénticos o no idénticos; en donde los cebadores de PCR en el extremo 5' tienen marcas de secuencia en donde la amplificación con cebadores marcados da como resultado productos de ácido nucleico unidos emparejados al azar. Más preferentemente, la amplificación es amplificación por PCR en emulsión y la producción de dichos amplicones y productos de ácido nucleico unidos emparejados al azar son procesos paralelos. Preferentemente, dicha secuenciación de dichos productos de amplificación unidos es un método de secuenciación altamente paralelo. La separación permite la formación específica de marcas de identidad unidas, en donde las moléculas que están co-localizadas en compartimentos separados o limitados a la difusión (por ejemplo, sobre superficie sólida, en emulsión) tienen tendencia a unir específicamente las secuencias de marcas de identidad de solo los agentes de unión unidos. Sin embargo, también se pueden unir marcas de identidad de agentes de unión no unidos, pero accidentalmente co-localizados. Similarmente, se pueden unir las marcas de identidad de agentes de unión que no se unen dentro de un complejo. El proceso de unión puede proporcionar una marca de identidad unida con dos marcas de identidad idénticas, por ejemplo, donde solo está presente un agente de unión en el compartimento. Similarmente, en el caso de más de un agente de unión que reconoce una diana, se pueden producir diferentes marcas de identidad unidas con las mismas especificidades de unión. La etapa de unión une la marca de identidad de uno o más miembros de la biblioteca de agentes de unión presentes. Además, la etapa de unión puede unir la marca de identidad de un miembro de una biblioteca de agentes de unión a una secuencia de ácidos nucleicos asociada a la diana. El proceso de unión no se basa en los ácidos nucleicos que interactúan entre sí debido a su estrecha proximidad, por ejemplo, hibridándose entre sí. Las secuencias se unen juntas usando un proceso de unión, por ejemplo, por métodos de amplificación como se describe en el presente documento. El uso de estos métodos, en lugar de basarse en ensayos de unión de gran proximidad física, tales como aquellos en el documento de patente US7306904, permite detectar múltiples interacciones en paralelo. Las marcas no necesitan estar en gran proximidad para ser unidas y así detectadas. Simplemente necesitan estar dentro del mismo compartimento.

Unión de secuencias de nucleótidos

Las secuencias de nucleótidos únicas asociadas a miembros del complejo pueden ser unidas por un método que comprende:

(i) Amplificar la secuencia de nucleótidos asociada a los agentes de unión, y opcionalmente si está presente la secuencia asociada a la diana, usar al menos dos pares de cebadores de PCR para producir al menos dos conjuntos de amplicones, en donde los cebadores se diseñan de manera que los amplicones del primer conjunto comprendan una secuencia, que es complementaria a una secuencia en el segundo conjunto de amplicones;

(ii) Hibridar los al menos dos conjuntos de amplicones; y

(iii) Llevar a cabo una reacción de amplificación para producir una secuencia de nucleótidos unida

Las etapas i-iii se pueden llevar a cabo secuencialmente o simultáneamente.

Cada par de cebadores comprende un cebador directo y un cebador inverso. Las secuencias de estos cebadores se diseñan de manera que permitan la amplificación de las secuencias identificables. Preferentemente, estos son cebadores universales, es decir, se unen a todas las secuencias identificables dentro de una biblioteca, por ejemplo, todos los miembros de una biblioteca de agentes de unión, o biblioteca de dianas. La secuencia entre los cebadores que se amplifica es única para un miembro de la biblioteca, permitiendo la identificación. Preferentemente, los pares de cebadores de PCR se diseñan para producir una secuencia de ácidos nucleicos unida dimerizada, de otro modo se produce la secuencia de ácidos nucleicos unida de multímeros. Esto se logra por que tienen secuencias conectoras de dímeros en el extremo 5' de al menos dos cebadores que amplifican los amplicones miembros del dímero o el multímero. Así, estos amplicones pueden formar productos de hibridación que se solapan parcialmente y extensibles por polimerasa en sus extremos 3' (por ejemplo, como se muestra en la Figura 2).

Preferentemente, la amplificación es amplificación por PCR en emulsión con separación adecuada de las secuencias de nucleótidos asociadas a los agentes de unión, o agente de unión y diana.

Opcionalmente, las secuencias de nucleótidos unidas de más de un complejo se combinan antes de identificar los agentes de unión y/o dianas unidos presentes. La identidad del (de los) agente(s) de unión y/o diana(s) se puede determinar, por ejemplo, por secuenciación de las secuencias de nucleótidos unidas. Esto se puede llevar a cabo usando un sistema altamente paralelo. Las secuencias unidas se pueden combinar de manera que se pueda llevar a

cabo una única reacción para identificar todas las secuencias unidas. Por ejemplo, todos los nucleótidos unidos se pueden secuenciar en una única reacción. Las secuencias unidas se pueden determinar cuantitativamente, para medir las abundancias relativas de las secuencias unidas.

5 Determinación de la información de dichos productos de ácido nucleico unidos

Basándose en las características de unión predeterminadas de los agentes de unión, la información de co-localización se puede deducir de la marca de identidad unida. La pluralidad de marcas de identidad unidas, es decir, las secuencias de nucleótidos únicas asociadas al agente de unión y opcionalmente la diana, en forma de las cifras de lecturas de secuenciación, proporciona información de las identidades de los miembros de complejos agente de unión/diana y su abundancia relativa. Otra información también es inferible, en donde se tiene en cuenta la información de todas las marcas de identidad unidas. Los ejemplos incluyen, pero no se limitan a, comparar afinidades relativas con afinidades predeterminadas de los agentes de unión, comparar abundancias relativas calculadas para los diferentes agentes de unión, determinación de diana unida y no unida, o relaciones de proteína.

Basándose en la unión por PCR de múltiplos de regiones CDR (CDR1, CDR2 y CDR3) de agentes de unión de anticuerpo, la secuencia de todas las regiones CDR se puede determinar usando secuenciación de NGS. Como las características de unión predeterminada de los agentes de unión se basan, preferentemente, en una identidad de secuencia de región CDR única, en una realización preferida, toda la información de secuencia de agentes de unión de anticuerpo se puede asociar a sus características de unión predeterminadas.

Determinación de los efectos del compuesto sobre la pluralidad de interacciones proteína-proteína

El método de la invención puede comprender la etapa de poner en contacto la biblioteca de agentes de unión con una diana en presencia o ausencia de un compuesto; para determinar si dicho compuesto efectúa la interacción de unión. Este método se puede usar para determinar los efectos de un compuesto u otros restos químicos sobre una pluralidad de interacciones de unión, por ejemplo interacciones proteína-proteína. Preferentemente, el método se usa para detectar compuestos o restos químicos que pueden promover o interrumpir ciertas interacciones proteína-proteína, u otras detectables, en donde los compuestos o restos químicos actúan de fármacos o eliminan o suprimen las consecuencias patológicas de dichas situaciones. Preferentemente, dichos fármacos se pueden usar para tratar diferentes enfermedades que incluyen, pero no se limitan a, cáncer, enfermedades infecciosas, enfermedades autoinmunitarias y otras.

Un compuesto como se usa en el presente documento se refiere a dos o más átomos que están conectados por enlaces covalentes. Un resto químico es parte de un compuesto, que forma un grupo funcional. El compuesto puede ser un agente farmacéutico conocido.

El método de la invención puede comprender la etapa de poner en contacto la biblioteca de agentes de unión con una diana en presencia o ausencia de un compuesto; para determinar si dicho compuesto efectúa las interacciones de unión.

En una realización adicional, la invención proporciona métodos de determinación de datos de interactoma. En una pluralidad de compartimentos, los agentes de unión, tales como los agentes de anticuerpo presentados, se co-localizan según sus características de unión, es decir, dos o más agentes de unión están presentes dentro de un compartimento, ya que se unen a la misma diana, o se unen a dianas que ellas mismas interactúan o se unen entre sí. Sus secuencias identificables, es decir, secuencias de nucleótidos únicas, se unen llevando la información de co-localización en forma de identidades unidas. Basándose en las características de información de unión e información de co-localización, se puede determinar la información en tanto las interacciones proteína-proteína como identidades de proteínas.

La invención también describe un método de determinación de la interacción proteína-proteína en un espécimen obtenido que comprende:

- (a) obtener una mezcla de proteínas;
- (b) poner en contacto una biblioteca de agentes de unión con dicha mezcla de proteínas se manera que se formen complejos agente de unión-proteína, en donde cada miembro de la biblioteca de agentes de unión se asocia con una secuencia de nucleótidos única;
- (c) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;
- (d) opcionalmente, inmovilizar dichos complejos agente de unión-proteína sobre una superficie sólida;
- (e) detectar dicha secuencia de nucleótidos única de dichos agentes de unión dentro de dichos complejos agente de unión-proteína y unir dicha secuencia de nucleótidos única de dichos agentes de unión para proporcionar

productos de ácido nucleico unidos basados en la co-localización proporcionada por dicha separación

(f) opcionalmente, combinar dichos productos de ácido nucleico unidos; y

5 (g) secuenciar dichos productos de ácido nucleico unidos correspondientes a las características de unión de proteínas de dichos agentes de anticuerpo presentados. La secuencia de los productos de ácido nucleico unidos se usa para inferir la presencia de una interacción proteína-proteína en dicha mezcla de proteínas en dicho espécimen. La secuencia de nucleótidos única permite que se identifique el agente de anticuerpo presente, y las características de unión de proteínas correspondientes de dichos agentes de anticuerpo presentados.

10 Los datos de la interacción proteína-proteína pueden ser validados por medios estadísticos, que incluyen determinación de origen y resta del nivel de interacciones proteína-proteína, determinación de afinidades relativas aparentes y abundancias relativas de proteínas e interacciones proteína-proteína.

15 La invención también se refiere a un método de determinación de los efectos de un compuesto sobre la interacción proteína-proteína presente en un espécimen que comprende:

(a) obtener una mezcla de proteínas en presencia y ausencia de un compuesto;

20 (b) poner en contacto una biblioteca de agentes de unión con dicha mezcla de proteínas de manera que se formen complejos agente de unión-proteína en donde cada miembro de la biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única;

25 (c) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;

(d) opcionalmente, inmovilizar dichos complejos agente de unión-proteína sobre la superficie sólida;

30 (e) detectar dicha secuencia de nucleótidos única asociada a dichos agentes de unión dentro de dichos complejos agente de unión-proteína y unir dicha secuencia de nucleótidos única asociada a dichos agentes de unión para proporcionar productos de ácido nucleico unidos emparejados al azar basados en la co-localización proporcionada por dicha separación

(f) opcionalmente, combinar dichos productos de ácido nucleico unidos; y

35 (g) secuenciar dichos productos de ácido nucleico unidos

(h) comparar la información obtenida en presencia y ausencia de un compuesto;

40 La secuencia de nucleótidos única del agente de unión permite que el agente de unión presente se identifique, y así se usen las características de unión de proteínas predeterminadas de dichos agentes de unión para inferir la presencia de una interacción proteína-proteína en dicha mezcla de proteínas en dicho espécimen.

45 Preferentemente, la determinación del efecto del compuesto se determina usando una configuración experimental de alto rendimiento.

En una realización adicional de la invención, se proporciona un método de determinación de las características de unión de proteínas de los miembros de una biblioteca de agentes de unión que comprende:

50 (a) obtener una biblioteca de proteínas presentadas en donde cada miembro está asociado con una secuencia de nucleótidos única;

55 (b) poner en contacto una biblioteca de agentes de unión con dicha biblioteca de proteínas presentadas para permitir la formación de complejos agente de unión/proteína en donde cada miembro de dicha biblioteca de agentes de unión se asocia a una secuencia de nucleótidos única;

(c) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;

60 (d) opcionalmente, inmovilizar dichos complejos agente de unión/proteína separados sobre una superficie sólida;

(e) unir las secuencias de nucleótidos únicas asociadas a los agentes de unión y las proteínas para producir un producto de ácido nucleico unido; opcionalmente mientras se mantiene la separación de los complejos,

65 (f) opcionalmente, combinar los productos de ácido nucleico unidos;

(g) determinar la secuencia de los productos de ácido nucleico unidos.

Las características de unión de proteínas de los miembros de la biblioteca de agentes de unión se pueden determinar a partir de la información dentro de los productos de ácido nucleico unidos a secuencia. La detección de la secuencia de nucleótidos unida única asociada a los miembros de bibliotecas de agente presentados y la proteína indica el reconocimiento y la unión a ciertos miembros de la biblioteca de proteínas presentadas. La secuencia puede proporcionar información en cuanto a qué miembros de la biblioteca se unen a qué proteínas.

Se pueden calcular las características de unión de proteínas de una pluralidad de dichos agentes de unión. Se puede combinar toda la información de características de unión para todos los miembros de dicha biblioteca de agentes de unión, como información de unión.

La unión de dicha secuencia de nucleótidos única asociada a dichos agentes de unión y dichas proteínas que proporcionan productos de ácido nucleico unidos emparejados al azar basándose en la co-localización proporcionada por dicha separación.

Preferentemente, dichas bibliotecas de agentes de unión y/o dianas (por ejemplo, muestras de proteína) se usan en concentración variada en la pluralidad de mediciones para permitir el cálculo de la información de unión cuantitativa que incluye constantes de disociación de dicha pluralidad de interacciones proteína-proteína. Los métodos para calcular estas medidas se conocen por el experto en la técnica.

La invención también se refiere a usos de kits en los que llevar a cabo los métodos de la invención. El kit comprende

- a) una biblioteca de agentes de unión en donde cada miembro de dicha biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única; y
- b) un conjunto de al menos dos pares de cebadores para unir las secuencias de nucleótidos asociadas al agente de unión; y opcionalmente instrucciones para su uso.

El kit también puede comprender un medio para detectar interacciones proteína-proteína en donde se proporcionan reactivos, y opcionalmente materiales, para llevar a cabo una cualquiera o más de las siguientes etapas: separación, inmovilización, detección de secuencias de nucleótidos únicas, unión de secuencias de nucleótidos, y/o detección de secuencias de nucleótidos unidas para obtener dicha información de co-localización. El kit también puede comprender además instrucciones para llevar a cabo los métodos de la invención y utilizar el kit.

Los kits pueden comprender además una biblioteca de presentación de proteínas en donde cada miembro de dicha biblioteca está asociado con una secuencia de nucleótidos única.

COMPARACIÓN DE LOS MÉTODOS DE LA PRESENTE INVENCION CON LOS MÉTODOS EXISTENTES

El método de la presente invención es un enfoque novedoso para capturar la información del interactoma de un modo rápido y rentable, que permite muestreo aleatorio y alta redundancia del muestreo. Proporciona cobertura dinámica, basada en el contexto celular original, por ejemplo fisiológico, basada en la interacción proteína nativa-proteína, y completa, de datos de interacción cuantitativa de complejos de proteína de múltiples unidades incluso grandes. Suprime los efectos de variables aleatorias, tales como la detección de proteínas no específicas, que interaccionan accidentalmente. Por tanto, disminuye el efecto de las variables, tales como cualquier variable relacionada con eventos de unión implicados en el principio de detección distinto de la interacción proteína-proteína original. El método es adecuado para la detección *in vitro* de proteínas que interaccionan no solo con la proteína de cebo, sino también con ADN, ARN y compuestos químicos.

Se usa detección basada en PCR de un ensayo de ligación de proximidad (PLA) para determinar la expresión relativa de una diana. Como los perfiles de expresión de proteína y de ARNm no son idénticos, las diferencias observadas podrían ser significativas con respecto a los procesos biológicos. El ensayo es capaz de detectar varias dianas en un experimento dado. Sin embargo, como es necesario preparar un gran número de anticuerpos específicos que están artificialmente marcados, es inviable al nivel de interactoma. Sin anticuerpos altamente específicos, la reactividad cruzada reduce la capacidad del ensayo de ligación de proximidad para distinguir claramente interacciones específicas y no específicas. La detección paralela de dianas a números muy grandes llega a ser cara y engorrosa.

Se ha desarrollado una forma multiplexada de PLA. En este ensayo, un anticuerpo inmovilizado sobre un soporte sólido actúa de reactivo de captura para enriquecer localmente un antígeno de una mezcla compleja de proteínas. Después de lavar, se añade un par de sondas del ensayo de ligación de proximidad (PLA). Esto va seguido por lavados adicionales y ligación de oligonucleótidos acercados. Esto permite mayor especificidad basándose en la necesidad de tres eventos de unión. Esto, en combinación con el uso de amplificación por PCR, permite alta especificidad y sensibilidad, y un amplio intervalo dinámico para la cuantificación de proteínas. Se puede usar este método acoplado con secuenciación de nueva generación (NGS) para registrar digitalmente patrones de abundancia de proteína para demostrar la detección simultánea de 36 analitos de proteína.

Se describe una variación de PLA como un ensayo extremadamente sensible y específico (4PLA) para la detección de estructuras de diana complejas tales como microvesículas en las que la diana se captura primero mediante un anticuerpo inmovilizado y posteriormente se detecta usando cuatro otros anticuerpos con cadenas de ADN unidas.

5 El requisito para la unión coincidente por cinco anticuerpos para generar un indicador amplificable da como resultado tanto elevada especificidad como sensibilidad.

Todos los tipos de ensayo de ligación de proximidad que usan estrecha proximidad necesitan validación experimental debido a limitaciones estéricas muy frecuentes.

10 En el caso de los métodos de la invención, la baja especificidad no es un problema, y se puede incluso usar como información de validación. Los métodos de invención usan la co-localización y compartimentalización, en donde las marcas de identidad amplificadas difunden libremente en compartimentos que permiten condiciones estéricas más relajadas.

15 Los vectores de destino basados en recombinasa de expresión dual (DERB) codifican individualmente dos conjuntos de secuencias reconocibles de recombinasa para insertar el marco de lectura abierto (ORF) de proteínas de interés, dos conjuntos de promotores y marcas indicadoras en marco con las ORFs para detectar interacciones. La introducción de los vectores en células vivas (procariotas y eucariotas) permite la detección de interacciones de proteína por transferencia de energía por resonancia de fluorescencia (FRET) o complementación de fluorescencia bimolecular (BiFC). La plataforma de DERB muestra ventajas con respecto a los sistemas actualmente comercializados introduciendo vectores de clonación basados en recombinasa y aceptores compatibles validados mediante experimentos de prueba de principio y la identificación de una interacción desconocida. El sistema necesita grandes números de interacciones cribadas y, por consiguiente, grandes esfuerzos y costes, que solo se adaptan a sistemas robóticos a nivel internacional y usando condiciones de ensayo artificiales (proteínas de fusión y promotores artificiales).

20

25

30 El cribado de dos híbridos de levadura (Y2H) es implementación específica del ensayo de complementación de fragmentos de proteínas, o PCA, donde la identificación de interacciones proteína-proteína se basa en dos fragmentos de proteínas, cada uno ligado covalentemente a fragmentos incompletos de una tercera proteína (por ejemplo, DHFR, que actúa de indicador). La interacción entre las proteínas acerca los fragmentos de la proteína indicadora en proximidad suficiente para permitir que formen una proteína indicadora funcional cuya actividad se puede medir. Este principio se puede aplicar a muchas proteínas indicadoras diferentes, como el cribado de dos híbridos de levadura usando el factor de transcripción GAL4. El cribado de dos híbridos de levadura investiga la interacción entre proteínas de fusión artificiales dentro del núcleo de levadura. El método tiene una alta tasa de positivos falsos, que hace necesario verificar las interacciones identificadas por otros medios. El método está fuera del contexto celular y la ausencia de entorno natural limita su uso para interacciones gobernadas por modificaciones específicas del contexto celular de la proteína, o en el caso de interacciones de baja afinidad. Al nivel de interactoma, requiere optimización adicional y el uso de matrices, para hacer posible construir los conjuntos de datos muy grandes del interactoma. Esto implica altos costes, pero todavía no vencen todas las limitaciones de este tipo de ensayo.

35

40

Se puede usar un sistema de cebo dual que mejora la exactitud de cribados de bibliotecas con una selección inmediata para eliminar positivos falsos.

45 Se ha descrito un ensayo de trampa de interacción funcional de heterodimerización mediada por bobina en espiral, donde los dominios de heterodimerización de bobina en espiral se sustituyen por dominios de unión de proteína modular. Esto puede ser útil para validar interacciones proteína-proteína funcionalmente relevantes, que dirigen enzimas a sustratos específicos, y cribar bibliotecas de fusión para componentes de interacción funcionalmente importantes.

50

En respuesta a las limitaciones conocidas de cribados de Y2H, se desarrolló un sistema de dos híbridos basado en células de mamífero (M2H). Este sistema M2H es similar al de los dos híbridos de levadura, en el que las interacciones se investigan fusionando cada par de proteínas de interés con un dominio de unión de ADN y uno de trans-activación transcripcional, respectivamente. Las técnicas de dos híbridos basados en células de mamífero tienen varias ventajas, en comparación con los ensayos basados en levadura y la resolución de algunos de los problemas conocidos. Como las levaduras carecen de proteínas clave implicadas en las modificaciones postraduccionales, no se puede ensayar la interacción basada en estas proteínas. Además, se pueden usar varios contextos de células de mamífero diferentes para proporcionar datos de interactoma específicos del contexto celular. Sin embargo, son difíciles de producir grandes conjuntos de datos, por lo que no se puede lograr el nivel de interacciones del interactoma debido a la necesidad de manipular un número muy grande de especímenes de cultivo celular de mamífero.

55

60

Una variación de la detección de interacciones proteína-proteína en cribados de Y2H es usar un método de secuenciación basado en PCR específica, denominado secuenciación por fusión. Esto es una fusión por PCR, que sitúa un par de secuencias que codifican proteínas interactuantes sobre el mismo amplicón de PCR. La fusión por

65

PCR consiste en dos rondas de PCR. En la primera ronda, X e Y (presentes sobre los vectores Y2H DB-X y AD-Y) se amplifican con cebadores en la dirección 5' específicos del vector DB y AD, respectivamente. Se usaron los amplicones de la primera ronda como moldes para producir un producto de PCR concatenado compuesto de ORFs X e Y conectados por una secuencia conectora de 82 pb. Los productos de PCR se reúnen y se secuencian por secuenciación de ADN de nueva generación para producir ISTs fusionados (sISTs). La secuenciación por fusión ha eliminado el obstáculo de algunos protocolos de Y2H, pero todavía no resuelve los problemas asociados a etapas clave.

Una patente que describe las siguientes mejoras de PCA: 1) genes indicadores (y métodos de detección de su expresión) que permiten fácilmente el análisis de grandes bibliotecas ($>10^7$ en tamaño) y cuya selectividad puede ser fácilmente "ajustada", modificada y/o monitorizada, 2) métodos para la medición simultánea e independiente de múltiples interacciones (como se demuestra por la expresión de diferentes genes indicadores), y 3) construcción de bibliotecas usando un sistema basado en fagémidos que proporciona a) un método eficiente automatizable para realizar experimentos biblioteca frente a biblioteca y b) un método para simplificar el análisis de candidatos positivos de cualquier cribado/selección realizados en los PCA procariotas. El uso de esta tecnología basada en fagémidos para cribar biblioteca frente a biblioteca implica cruzar bibliotecas, por ejemplo, una infecta la biblioteca de cebo de células con la biblioteca de presa de fago (usando un exceso de células con respecto a fago para garantizar que cada célula esté de media solo infectada por un fago) y buscar expresión activada del gen indicador. Esto es una etapa significativa hacia barridos del interactoma a gran escala, sin embargo no está basada ni en dinámica ni en el contexto celular.

Para los experimentos de biblioteca frente a biblioteca realizados en levadura, se usa la formación de una célula diploide α que aloja el ADN de ambas células haploides de partida. Así, las células que alojan una biblioteca de híbridos de presa se pueden aparear con células α que alojan un híbrido de cebo de prueba. Aunque esto elimina las cuestiones de eficiencia de transformación, no trata los otros requisitos de barridos ideales del interactoma.

Se fusionaron genéticamente dos bibliotecas de cremalleras de leucina, semi-aleatorizadas, en las posiciones adyacentes al núcleo hidrófobo, a uno cualquiera de dos fragmentos designados de la enzima dihidrofolato reductasa murina (mDHFR), y se co-transformaron en *E. coli*. Se requirió la interacción entre los polipéptidos de la biblioteca para la reconstitución de la actividad enzimática de mDHFR, permitiendo el crecimiento bacteriano. Esta estrategia, sin embargo, está limitada por la eficiencia de transformación que se puede lograr en células bacterianas.

También se concibe la adaptación de un ensayo de captura de GST a un formato de placa filtrante de 96 pocillos. El uso de una placa filtrante de múltiples pocillos hace posible ensayar más muestras en significativamente menos tiempo usando menos reactivos y procesamiento de muestras más eficiente que el ensayo tradicional de un solo tubo. Este tipo de ensayo resolvió algunos de los problemas que causan obstáculos técnicos; sin embargo, es inviable generar los conjuntos de datos muy grandes requeridos, usando este sistema.

El método de purificación por afinidad en tándem (TAP) se puede observar como una versión más específica de la co-inmunoprecipitación, permite la identificación de alto rendimiento de interacciones de proteínas. La exactitud del método se puede comparar con los experimentos de co-inmunoprecipitación a pequeña escala y las interacciones se detectan dentro del contexto celular correcto. Sin embargo, requiere dos etapas sucesivas de purificación de proteínas, por lo que no puede detectar fácilmente interacciones proteína-proteína transitorias. El método de TAP aplica la fusión de la marca de TAP al extremo C de la proteína en estudio. La marca de TAP consiste en péptido de unión a calmodulina (CBP) del extremo N, seguido por el sitio de escisión de la proteasa del virus del grabado del tabaco (proteasa TEV) y la proteína A. Es capaz de proporcionar la determinación real de componentes de proteína cuantitativamente en el contexto celular correcto, sin embargo, el método, al nivel del interactoma, requiere un gran esfuerzo e incurriría en grandes costes para cubrir todo el proteoma con las construcciones usadas.

En respuesta a este problema, se desarrolló la estrategia sin marca para realizar el fraccionamiento bioquímico sistemático altamente extenso del interactoma de proteína humana soluble usando múltiples técnicas de separación que incluyen cromatografía de intercambio iónico de múltiples lechos de alto rendimiento no desnaturizante, centrifugación en gradiente de sacarosa e isoelectroenfoque. El método necesita verificación y análisis estadístico para producir conjuntos de datos fiables, también necesita una cantidad significativa de un material de muestra y su coste limita su uso para el muestreo aleatorio y alta redundancia de muestreo.

Las micromatrices de proteína han introducido un nuevo enfoque para identificar y caracterizar interacciones de proteína, proporcionando la capacidad de identificar rápidamente nuevas interacciones entre miles de proteínas en un único experimento. Puesto que se conoce la localización e identidad de cada proteína sobre la matriz, se pueden desarrollar rápidamente mapas de interacción a partir del sondeo iterativo de matrices de proteína. Debido a que un experimento de micromatrices de proteínas se realiza en el plazo de un día, y las interacciones se evalúan en el contexto de miles de otras proteínas, el perfilado de la interacción en micromatrices puede acelerar enormemente la velocidad a la que se descubren novedosas interacciones de proteínas. Además, la naturaleza *in vitro* de los experimentos de micromatrices de proteínas permite el control con respecto a las condiciones de sondeo que afectan las interacciones de la proteína, tales como la concentración de proteína, modificaciones postraduccionales, y la presencia de cofactores, que pueden no ser posibles con otros métodos tales como el cribado de dos híbridos

de levadura. Sin embargo, la sonda clásica en un enfoque de tiempo no es adecuada para grandes experimentos a nivel de interactoma.

5 Se usa una versión de detección basada en matriz de proteínas, donde se aplica lisado de proteínas celulares o mezclas de péptidos sintéticos a la matriz de proteína con proteína/ péptido de cebo inmovilizado. Se lavan las proteínas/péptidos no específicos en diversas condiciones rigurosas y solo quedan sobre el chip las proteínas que interactúan específicamente con la proteína/péptido de cebo. Finalmente, los complejos de proteína/péptido de interacción capturados se analizan entonces por espectrometría de masas de SELDI-TOF y se confirman sus identidades por sus masas distintivas predichas. Es un enfoque altamente prometedor, sin embargo, la
10 secuenciación de proteínas por SELDI-TOF está limitada por varios factores (cantidad de proteína, separación, modificaciones postraduccionales) y las interacciones carecen del contexto natural.

15 La mezcla de proteínas también puede ser inmovilizada sobre un soporte sólido, y se pone en contacto con una pluralidad de dominios de interacción proteína-proteína no marcada en condiciones de unión apropiadas. En presencia de al menos un dominio de interacción proteína-proteína seleccionado marcado (siendo el dominio de interacción proteína-proteína marcado diferente de los dominios de interacción proteína-proteína no marcados), se mide la unión del dominio de interacción proteína-proteína marcado. Este método es específico del dominio de interacción, que limita su aplicación.

20 Una tecnología de presentación libre de células combinada con secuenciación de nueva generación (NGS) puede mejorar tanto la cobertura como la fiabilidad de conjuntos de datos del interactoma. El método completamente libre de células da un alto rendimiento y un espacio de detección, que prueba las interacciones sin usar clones. La información cuantitativa proporcionada por NGS reduce el número de positivos falsos. El método es adecuado para la detección *in vitro* de proteínas que interactúan no solo con la proteína de cebo, sino también con ADN, ARN y
25 compuestos químicos. El método emplea un tratamiento *in vitro* completo con bibliotecas de ADNc (extraídas de células y tejidos) y la selección en proteínas diana para obtener secuencias de ADNc seleccionadas para NGS. Las selecciones usando el método se realizan en condiciones libres de células, y la posterior secuenciación por NGS no se limita por etapas de clonación que usan cualquier tipo de células. Este método aplica un enfoque de uno en uno fuera del contexto celular, que limita su capacidad para generar grandes conjuntos de datos al nivel del interactoma.

30 Otro ensayo libre de células utiliza las denominadas inteínas, que son secuencias de péptidos capaces de dirigir el corte y empalme *in vitro* de proteínas. Una inteína es una secuencia de proteínas intermedia en un precursor de proteínas que se escinde del precursor de proteína durante el corte y empalme de proteínas. Se proporcionan construcciones de fusión de dos híbridos, donde una tiene un primer agente de prueba y un fragmento de inteína del extremo N o N-inteína, y el otro tiene un segundo agente de prueba y un fragmento de inteína del
35 extremo C o C-inteína. Además, una o ambas construcciones de fusión pueden tener un indicador que experimenta cambios detectables tras el corte y empalme *in vitro* de las construcciones de fusión. Tanto el alto rendimiento como las características libre de contexto celular son desventajas significativas de este método.

40 Otro método usado para estudiar interacciones proteína-ADN o proteína-proteína es el método de presentación en fagos. Las proteínas se presentan sobre la superficie de bacteriófago filamentoso (por ejemplo, M13) que codifica ADN de la proteína presentada. Las proteínas diana o secuencias de ADN de interés se inmovilizan sobre un soporte sólido y se usan para bibliotecas enriquecidas por afinidad de proteínas presentadas en fagos para candidatos que se unen a la diana. El método se ha usado para identificar y caracterizar tanto interacciones
45 proteína-ADN como proteína-proteína. La presentación en fagos es un proceso de enriquecimiento que requiere múltiples ciclos para deducir datos de interacción proteína-proteína. Los enriquecimientos se realizan *in vitro*, que influyen en las interacciones y favorece la interacción de alta afinidad en la detección. Ciertas proteínas (particularmente las grandes) no son muy aptas para el análisis por presentación en fagos. Una desventaja importante de la presentación en fagos de ORF basados en bacterias es que las proteínas presentadas sobre la
50 superficie de fagos carecen de modificaciones postraduccionales apropiadas, tales como glucosilación.

La presentación en fagos, como se usa en la materia, se limita a reproducir eventos de unión naturales, como se ha descrito anteriormente. Sin embargo, los repertorios de anticuerpos expresados como presentación en fagos son extremadamente satisfactorios en proporcionar moléculas farmacológicas de partida para anticuerpos terapéuticos o
55 anticuerpos de detección para ensayos de diagnóstico. La presente invención aplicó la tecnología de presentación en fagos en el último contexto, donde las bibliotecas de anticuerpos en fago extremadamente grandes contienen una pluralidad de especificidades que cubren casi cualquier diana identificable.

60 Una variación de la presentación en fagos se refiere a un método para la selección e identificación de componentes de interacción. Las moléculas diana (ligandos) se inmovilizan sobre la superficie de un vehículo de fase sólida de forma que son direccionables en la posición en un grado bidimensional y se ponen en contacto con virus de presentación de proteínas. El componente de interacción se identifica por detección y determinación de la posición de la unión entre el ligando inmovilizado y el componente de interacción. El método de detección preferido descrito es resonancia de plasmones superficiales (SPR).

65 La reticulación química se usa frecuentemente para "fijar" interacciones de proteína en su sitio antes de intentar

aislar/identificar proteínas de interacción. Los reticulantes comunes para la presente solicitud incluyen el reticulante no escindible [éster de NHS], [suberato de bis-sulfosuccinimidilo] (BS3); una versión escindible de BS3, [ditiobis(sulfosuccinimidilpropionato)](DTSSP); y el reticulante [imidoéster] [ditiobispropionimidato de dimetilo] (DTBP) que es popular por fijar interacciones en ensayos de ChIP.

5 Se ha desarrollado una tecnología para identificar proteínas que se unen específicamente a elementos reguladores de la transcripción predichos usando biblioteca de presentación en fagos de péptidos genómicamente codificados, que se unen a un ADN bicatenaria inmovilizado en la superficie, que contiene una secuencia de motivos de ADN de interés. Después del enriquecimiento para una interacción ADN-proteína específica, se amplifican los fagos unidos, y se secuencian los insertos del fago enriquecido para determinar las proteínas de interacción usando marcado e hibridación con la micromatriz de ADN.

15 Supuestamente se han medido uno o más analitos usando una matriz chemFET. La matriz puede incluir cualquiera de una variedad de sustancias químicas que proporcionan información relevante referente a un proceso químico o procesos químicos de interés que incluyen la unión de un anticuerpo a un antígeno. En algunos aspectos, la capacidad para medir los niveles o concentraciones de uno o más analitos, además de detectar simplemente la presencia de un analito, proporciona valiosa información a propósito del proceso o procesos químicos.

20 Como es evidente de la descripción anterior, la presente invención proporciona un poderoso sistema *in vitro* y versátil para detectar y caracterizar interacciones proteína-proteína, y para seleccionar compuestos capaces de modular interacciones proteína-proteína. El sistema se puede usar con gran conveniencia y puede ser fácilmente adaptado a procedimientos de cribado de alto rendimiento.

25 Aunque la invención anterior se ha descrito con algún detalle a modo de ilustración y ejemplo para los fines de claridad de entendimiento, será obvio que ciertos cambios y modificaciones pueden ser puestos en práctica dentro del alcance de las reivindicaciones adjuntas.

Ejemplo 1.

30 Lisis y detección de fagos en PCR digital en gotitas

Se han usado vector de fagémido pBluescript II SK(+) (Agilent, 212205) origen f1 en orientación (+), orientación de policonector Sac-->Kpn en cepa hospedadora: XL1-Blue MRF' para generar el fago pBluescript II SK(+) y se compró el fago auxiliar M13KO7 (NEB, N0315S). Los fagos se valoraron para determinar el número de fagos infectivos. Los fagos se diluyeron sucesivamente por factor de 10 para lograr menos que el fago individual por dilución de compartimento. Las diluciones se sometieron a PCR digital usando el sistema de PCR digital en gotitas QX200 (ddPCR™). Brevemente, se usaron 5'CTCAAGTCGGTGACGGTGAT3' (directo específico de M13KO7), 5'GACAAAAGGGCGACATTCAA3' (inverso específico de M13KO7) y/o 5'TCTTGATCCGGCAAACAAC3' (directo específico de pBluescript II SK(+)), 5'TTTTCTGCGCGTAATCTGCT3' (inverso específico de pBluescript II SK(+)) con las sondas 5'CTGGTAGCGGTGTTTTT3' (sonda específica de pBluescript II SK(+)/FAM-MGB marcada), 5'CCGTCAATATTTACCTTCCC3' (sonda específica de M13KO7 VIC-MGB marcada) para amplificar los fagos compartimentalizados, la amplificación se registró en dos canales diferentes y la generación de gotitas, PCR y detección fueron según el protocolo del fabricante.

45 Se detecta la eficaz distribución de Poisson de lisis de fagos y de detección de fagos individuales de recuentos indicando la sensibilidad de detección de fagos individuales.

Ejemplo 2.

50 PCR de dimerización

Se han usado vector de fagémido pBluescript II SK(+) (Agilent, 212205) origen f1 en orientación (+), orientación de policonector Sac-->Kpn en cepa hospedadora: XL1-Blue MRF' para generar el fago pBluescript II SK(+) y se compró el fago auxiliar M13KO7 (NEB, N0315S). Los fagos se marcaron con 20 de exceso molar de biotina según las instrucciones del fabricante usando EZ-Link Sulfo-NHSBiotin (Thermo, 21326) y se precipitaron usando precipitación en el punto isoelectrico como se describió (Dong D, Sutaria S, Hwangbo JY, Chen P. A simple and rapid method to isolate purer M13 phage by isoelectric precipitation. Appl Microbiol Biotechnol. 2013 Sep;97(18):8023-9.). Se mezclaron igual número de fagos purificados proporcionando una concentración de 10e+6/ml de fagos y se han combinado con avidina (A9275-1MG, Sigma-Aldrich) en equivalencia molar o usado como una mezcla y se sometieron a PCR digital (sistema de PCR digital en gotitas QX200 (ddPCR™)). Brevemente, se usaron 5'TAACGTGGGAATGGTGCTTCTCAAGTCGGTGACGGTGAT3' (directo específico de M13KO7), 5'GACAAAAGGGCGACATTCAA3' (inverso específico de M13KO7) y 5'GAAGCACCATTTCCACGTTATCTTGATCCGGCAAACAAC3' (directo específico de pBluescript II SK(+)), 5'TTTTCTGCGCGTAATCTGCT3' (inverso específico de pBluescript II SK(+)) con las sondas 5'CTGGTAGCGGTGTTTTT3' (sonda específica de pBluescript II SK(+)/FAM-MGB marcada), 5'CCGTCAATATTTACCTTCCC3' (sonda específica de M13KO7 VIC-MGB marcada) para amplificar los fagos

compartimentalizados, la amplificación se registró en dos canales diferentes y la generación de gotitas, PCR y detección fueron según el protocolo del fabricante. Se esperaron elevadas cifras de "unión" en presencia de avidina debido a la unión biotina-avidina. Para detectar productos de PCR dimerizados, se extrajo el ADN amplificado según la recomendación del fabricante y se sometió a PCR usando los cebadores 5'TTTTCTGCGCGTAATCTGCT3' (inverso específico de pBluescript II SK(+)), 5'GACAAAAGGGCGACATTCAA3' (inverso específico de M13KO7) y sondas 5'CTGGTAGCGGTGGTTTTT3' (sonda específica de pBluescript II SK(+) FAM-MGB marcada), 5'CCGTC AATATTTACCTTCCC3' (sonda específica de M13KO7 VIC-MGB marcada) en el instrumento de PCR en tiempo real. Solo los productos correctamente dimerizados son amplificables y la señal de amplificación esperada se genera por ambas sondas.

Ejemplo 3.

Identificación basada en compartimentos de proteínas/complejos de proteína

(a) La extracción de complejos de proteína se logra por medios estándar - que incluye cualquier método de separación, que proporciona incluso complejos de proteína parcialmente preservados y métodos para purificar parcialmente complejos de proteína (por ejemplo, según su modificación postraduccional u otra).

(b) Se combinan fagos de la biblioteca de anticuerpos con los complejos de proteína y se diluyen y compartimentalizan para lograr el nivel de separación de fagos individuales para fagos no unidos. La compartimentalización (por ejemplo, separación eficaz de reacciones de números grandes) se basa en la separación basada en la distribución de Poisson de fagos no unidos individuales; por ejemplo, emulsiones y micromatriz son el mejor método conocido del estado de la técnica. Los únicos requisitos son la co-localización no específica mínima con fagos y complejos de proteínas. Las secuencias de nucleótidos se unen usando la lisis térmica eficaz de fagos, seguido por proceso de PCR de unión.

(c) Se amplifican el anticuerpo que codifica (basándose en cualquier región CDR de gen de anticuerpo en la biblioteca de fagos, preferentemente basándose en CDR3), fragmento específicos de ADN de fagos (unidos y sin unir) por compartimento (usando cebadores generales capaces de dimerización) y se unen (PCR de dimerización, para un ejemplo véase el Ejemplo 2 y 4.)

(d) Los amplicones de PCR dimerizados generados, incluso no purificados, se combinan preferentemente usando extracción de ADN de la emulsión o retirada física de las reacciones de las cavidades de la micromatriz u otros medios.

(e) La información de uniones de la multitud de amplicones de dímeros de PCR se reveló de un modo altamente paralelo y cuantitativo, preferentemente por secuenciación de nueva generación de ADN.

(f) La información predeterminada de características de unión de fagos individuales (véanse el Ejemplo 5 y 7) y la información de uniones de fagos unidos se usa para calcular el interactoma en base estadística que incluye la determinación de interacciones significativas basándose en la información de unión restada del origen (retirada de interacciones promiscuas o aleatorias o casuales); confirmación y filtración de interacción basándose en la información de uniones redundante de diferentes fagos con la misma información de características de unión posiblemente ponderadas por su información de constante de disociación conocida; determinación de interacciones de dímeros y multímeros basándose en información de uniones confirmada y filtrada, posiblemente ponderada; medición de la abundancia relativa de proteínas y complejos de proteína basándose en la información de unión cuantitativa; confirmación, cálculo de error estadístico de la determinación de interacciones de dímeros y multímeros, y la abundancia relativa de proteínas y complejos de proteína basándose en mediciones redundantes de la detección de varios fagos con la misma información de características de unión.

Ejemplo 4.

Identificación basada en compartimentos de proteínas de control/complejos de proteína

Se compraron fagos M13 que presentan anticuerpos de Source BioScience (6001_hDAb) que incluyen una biblioteca de anticuerpos de fagémido (-3x10⁹) (Dudgeon K, Famm K, Christ D. Sequence determinants of protein aggregation in human VH domains. Protein Eng Des Sel. 2008 Oct 28.) El fago auxiliar KM13, cepa bacteriana TG1Tr y anticuerpos anti-beta-galactosidasa y anti-ubiquitina bovina presentan fagos. Los fagos de control se secuenciaron por NGS (ThermoFisher Scientific PGM, Ion Xpress™ Plus Fragment Library Kit, 4471269) y sus afinidades de unión se verificaron por ELISA contra los antígenos respectivos. Los antígenos son la ubiquitina monobiotinilada (b-UBI) de LifeSensors (Nº S1280), biotina marcada con beta-galactosidasa (G5025) (b-BGAL). La avidina de clara de huevo (A9275) también es de Sigma-Aldrich.

Basándose en los resultados de secuenciación se diseñan cebadores generales: directo general - CCAAGAACACGCTGTATCTGCA; cebadores inversos generales capaces de dimerización - TGCGCATCCATTGTAGAGGTGAGACGGTGACCAGGGTTCC y

ACCTCTACAATGGATGCGCAGAGACGGTGACCAGGGTTCC. Para detectar productos dimerizados, se diseña una reacción de PCR en tiempo real específica de dímero: directo - AGTTGGAGTCTTGGGGTCAGG, inverso - AGGTGGTTCGATGTTTGACTACTG y sonda - FAM TCTCACCTCTACAATGGAT MGB.

- 5 También se diseñan sondas específicas de fago de control: anti-beta-galactosidasa - FAM GCTAGGGCTATGTATCC MGB; anti-ubiquitina bovina - VIC TGGGTCGATGTTTGACTAC MGB.

10 Según las instrucciones, se amplifican fagos de control (anti-beta-galactosidasa $3,8 \times 10^{12}/\text{ml} = 6,48 \text{ nM}$, anti-ubiquitina bovina $4,0 \times 10^{12}/\text{ml} = 6,7 \text{ nM}$). Se combinaron avidina (36 nM) (o se omitió), b-UBI (72nM) y b-BGAL (72 nM para el monómero) y se incubaron durante una hora a temperatura ambiente para formar complejos. Se combinaron los complejos diez veces diluidos y se incubaron durante la noche con fagos de control a 1,5 nM. Los complejos unidos a fago se diluyen 2×10^6 veces y según el protocolo de (sistema de PCR digital en gotitas QX200 (ddPCR™)) se generan gotitas de emulsión y se amplifican usando condiciones de PCR: Supermix de ddPCR for Probes (sin dUTP) (186-3023), la concentración de cebador directo general es 800 nM, concentración de cebadores inversos generales capaces de dimerización es 50 nM. En algunos casos también se añaden sondas específicas de fago de control a 250 nM. Las gotitas amplificadas son cloroformo extraído según el protocolo del fabricante para recuperar productos de dímero amplificados.

20 Los dímeros se detectan satisfactoriamente por la PCR en tiempo real específica de dímero demostrando su correcta estructura dimerizada. Se detectaron elevadas 'unión' – 'uniones' en presencia de avidina debido a los fagos de control y la unión del complejo avidina/antígeno si se incluyeron las sondas específicas de control en la reacción de PCR en emulsión, que indica detección de anti-beta-galactosidasa y se localizaron fagos anti-ubiquitina bovina en la misma gotita a una mayor tasa que solo por casualidad.

25 Ejemplo 5.

Predeterminación de la información de características de unión de un biblioteca de fagos de anticuerpos

30 (a) Se combinan anticuerpo (la biblioteca de fagos, cuya información de características de unión se va a determinar) y fagos de la biblioteca de ADNc (que constituyen los ADNcs, que serán determinados contra la información de características de unión de un biblioteca de fagos de anticuerpos) y se diluyen los complejos de fago de anticuerpo- ADNc y se compartimentalizan para lograr nivel de separación de fagos individuales para fagos no unidos – para más detalles véase el Ejemplo 3. parte (b),

35 (b) Se amplifican el anticuerpo que codifica (basado en cualquier región CDR de gen de anticuerpo en la biblioteca de fagos, preferentemente basada en CDR3) y el ADNc que codifica (el fragmento de ADNc) secuencias de ADN de fagos (unidos y no unidos) por compartimento (usando cebadores generales capaces de dimerización) y se unen (PCR de dimerización, como un ejemplo véase el Ejemplo 2.y 4.)

40 (c) Los amplicones de PCR dimerizados generado se combinan preferentemente usando extracción de ADN de la emulsión o retirada física de las reacciones de las cavidades de la micromatriz

45 (d) La información de unión de la multitud de amplicones de dímeros de PCR se revela de un modo altamente paralelo y cuantitativo, preferentemente por secuenciación de nueva generación de ADN.

50 (e) La información de unión de fagos unidos se usa para calcular la información de características de unión de la biblioteca de fagos de anticuerpos contra la biblioteca de fagos de ADNc que incluye determinación de interacciones significativas basándose en información de unión restada de la original (retirada de interacciones promiscuas o aleatorias o casuales); identificación de uniones de anticuerpo - ADNc significativas basándose en interacciones estadísticamente significativas; determinación de la información de características de unión para cada anticuerpo-fagos detectados que incluyen fragmentos de ADNc detectados, proteínas detectadas inferidas; confirmación, cálculo de error estadístico de la información de características de unión basándose en mediciones redundantes de la detección de varios fagos con la misma información de características de unión.

55 Ejemplo 6.

Enriquecimiento de una biblioteca de fagos de anticuerpos

60 (a) Se inmovilizan fagos de la biblioteca de ADNc (que constituyen los ADNcs, que necesitan ser detectados por la biblioteca de fagos de anticuerpos enriquecida) por medios separables (para la separación de fagos de anticuerpos sin unir), preferentemente sobre micropelras

65 (b) Se combinan fagos de la biblioteca de anticuerpos (la biblioteca de fagos, cuyo enriquecimiento está previsto) con fagos de ADNc inmovilizados para lograr la separación de fagos unidos y sin unir

(c) Se retiran fagos de anticuerpos sin unir, preferentemente lavando

(d) Se eluyen los fagos de anticuerpos unidos y se amplifican opcionalmente por medios adecuados para conseguir una preparación de alto título

5 (e) Opcionalmente, la preparación de alto título de fagos de anticuerpos enriquecidos se somete a la siguiente ronda de enriquecimiento

(f) Opcionalmente, los fagos unidos eluidos se verifican contra la biblioteca de fagos de ADNc usando el método descrito en el Ejemplo 5.

10

Ejemplo 7.

Predeterminación de la información de características de unión de una biblioteca de fagos de anticuerpos enriquecida

15

Se compraron fagos M13 que presentan anticuerpos de Source BioScience (6001_hDAb) que incluyen una biblioteca de anticuerpos de fagémido (-3×10^9) ((Dudgeon K, Famm K, Christ D. Sequence determinants of protein aggregation in human VH domains. *Protein Eng Des Sel.* 2008 Oct 28.), el fago auxiliar KM13 y la cepa bacteriana TG1Tr. Se amplificó la biblioteca, se infectó con fago auxiliar KM13 y se recogieron los fagos según el protocolo (Lee CM, Iorno N, Sierro F, Christ D. Selection of human antibody fragments by phage display. *Nat Protoc.* 2007;2(11):3001-8.). Se compró la biblioteca de péptidos de presentación en fagos PhD12 (E8110S) y la cepa hospedadora de ER2738 de *E. coli* de New England Biolabs. Se sembró la biblioteca PhD12 en placas LB/IPTG/Xgal y se seleccionaron 50 placas y se combinaron (biblioteca de cebo de antígeno). La biblioteca de cebo de antígeno se absorbió sobre placa de microtitulación y se llevó a cabo inmunopurificación usando la biblioteca entera de Source BioScience (6001_hDAb) según el protocolo de la biblioteca Source BioScience (6001_hDAb). Se sembraron en total 612 clones en placas de LB/ampicilina, que se amplificaron, se infectaron con fago auxiliar KM13 y se recogieron (biblioteca de anticuerpos enriquecida).

20

25

30

Basándose en las secuencias de los fagos se diseñan cebadores generales: específico de biblioteca general directo de Source BioScience (6001_hDAb) - CCAAGAACACGCTGTATCTGCA; cebador inverso general capaz de dimerización específico de biblioteca de Source BioScience (6001_hDAb) - TGCGCATCCATTGTAGAGGTGAGACGGTGACCAGGGTTCC- y específico de biblioteca de péptidos de presentación en fagos PhD12 general directo - CGCAATTCCTTTAGTGGTACCTTT; cebador inverso general capaz de dimerización específico de biblioteca de péptidos de presentación en fagos PhD12 - ACCTCTACAATGGATGCGCATCTGTATGGGATTTTGCTAAACAAC.

35

Para detectar productos dimerizados se diseña una reacción de PCR en tiempo real específica de dímero: directo - CGGACTGTTGAAAGTTGTTTAGCA, inverso - GGTCACCGTCTCACCTCTAC y sonda - VIC-CATACAGATGCGCATCC-MGB.

40

Se combinaron la biblioteca de cebo de 10^{12} antígenos y los fagos de la biblioteca de anticuerpos enriquecida y se incubaron durante la noche a temperatura ambiente. Se diluyeron los complejos de fago 2×10^6 veces y según el protocolo de (sistema de PCR digital en gotitas QX200 (ddPCR™)) se generan gotitas de emulsión y se amplifican usando condiciones de PCR: ddPCR Supermix for Probes (sin dUTP) (186-3023), la concentración de cebador directo general es 800 nM, la concentración de cebadores inversos generales capaces de dimerización es 50 nM. Las gotitas amplificadas se extraen con cloroformo según el protocolo del fabricante para recuperar los productos de dímero amplificados. Los productos de dímero amplificados se secuencian en NGS y se detectan la biblioteca de cebo específica y los productos dimerizados de la biblioteca de anticuerpos enriquecida que indica interacciones basadas en secuencia específicas entre los miembros de la biblioteca de cebo de antígenos y la biblioteca de anticuerpos de cobertura.

45

50

Ejemplo 8.

Determinación de la información de unión cuantitativa de los miembros de una biblioteca de fagos de anticuerpos

55

(a) Aplicar el método en el Ejemplo 3. con la modificación de que en la etapa b. se usan varias cantidades cuantificadas de proteína en condiciones de equilibrio dando como resultado varias determinaciones paralelas

(b) Basándose en la información cuantitativa obtenida en las determinaciones paralelas, se pueden construir curvas de unión cuantitativas para una multitud de interacciones proteína-fago y se pueden calcular la constante de disociación e información de capacidad de unión.

60

Ejemplo 9.

65 Estequiometría de la invención

Con respecto a la estequiometría de la unión, 1 nM de proteína en el volumen celular de *E. coli* es aproximadamente 1 molécula/célula y 2.000 moléculas/célula de mamífero (HeLa), y la concentración característica para las proteínas de señalización (como un ejemplo aquí) está en el intervalo 10 nM-1 microM. Además, como la constante de disociación (Kd) de anticuerpos de presentación en fagos está en el intervalo de 10 nM y hasta 0,1 nM y las velocidades de disociación de $10^{(-3)}$ a $10^{(-4)}$ s⁻¹ y como estos fagos pueden ser rutinariamente seleccionados, se espera que la estequiometría de unión a saturación para la mayoría de las proteínas/epítopes y velocidades de disociación proporcionen tiempo suficiente para compartimentalizar complejos sin disociación temprana.

Existen $2-4 \times 10^6$ proteínas por micrómetro cúbico (es decir, 1 fL) en levadura de bacterias, y células de mamífero (Bioessays. 2013 Dec;35(12):1050-5.) para el volumen de 5000 células eucariotas (10000 fL) es 10^{10} proteínas y la máxima concentración de fagos es alrededor 10^{16} /ml → en 10000 fL de volumen: 10^{11} fagos.

La complejidad del interactoma en el intervalo de $10^{(+4-5)}$, la multiplejidad interacción/fago deseada es aproximadamente 10, que corresponde a 10^5 fagos individuales. Además, es decir, 0,1 nM por fago individual (a 10^{16} /ml), y como cada proteína superior a 10^6 por 10000 fL (0,01 % de proteína total) tiene una concentración superior a 1 nM y el promedio 0,1-1 nM de Kd de un fago de anticuerpo (probabilidad subnanomolar HuCAL GOLD: 30 %) (J. Mol. Biol. (2008) 376, 1182-1200) puede proporcionar una saturación de 50-5 %. Esto es correspondiente a la saturación co-localizada de 25-0,25 % (saturación co-localizada significa que dos fagos unidos con distintas especificidades están localizados en el mismo compartimento).

En el caso del instrumento HiSEQ 2500 NGS a 0,25 % de tasa de co-localización, los 300 millones de lecturas (10-300 Gb, lecturas de 250 pb) corresponden al mínimo número de productos de PCR heterodiméricos de 7×10^5 , que significa una cobertura de secuenciación de la interacción binaria mínima: 375.

Una complejidad de biblioteca de fagos de anticuerpos completamente aleatorizada, primaria, de alta cobertura, tiene hasta $10^{(+13)}$ de clones de fagos individuales y el número de PCR en emulsión gestionable o compartimentos de micromatriz están en el intervalo de 10^5-10^8 (el número real de compartimentos de chip de NGS), estos números desproporcionados necesitan ser fusionados reduciendo la complejidad de bibliotecas en fagos de anticuerpos y aumentan la abundancia de fagos con capacidad de unión contra proteínas diana (que se dirigen a interactoma parcial o completo). Para reducir la complejidad se idea un proceso de selección específico - la selección de fagos de bibliotecas completamente aleatorizadas usando selección de bibliotecas (anticuerpo) contra el método de bibliotecas (ADNc) da como resultado biblioteca de fagos de baja complejidad, enriquecida por afinidad, intacta, de uso general, además el proceso se puede monitorizar detectando las uniones anticuerpo-proteína durante el proceso de selección o incluso se puede extraer la información de cinética de unión usando, por ejemplo, diferente cantidad de fagos de presentación de proteínas de entrada.

También es posible generar bibliotecas por construcción gradual de bibliotecas cada vez más complejas usando enfoques ascendentes (mezclando fagos con características de unión conocidas y añadiendo fagos de origen; estas son bibliotecas específicas adaptadas a tareas específicas), y descendentes (por ejemplo, reduciendo la complejidad seleccionando los fagos interactuantes)

Ejemplo 10.

Evaluación estadística de la información obtenida

(a) La extracción de complejos de proteína se logra por medios estándar, y se combinan los fagos de bibliotecas de anticuerpos y se diluyen los complejos formados y se compartimentalizan para lograr nivel de separación de fagos individuales para fagos no unidos.

(b) El anticuerpo que codifica fragmentos específicos de ADN de fagos (unidos y no unidos) por compartimento se amplifican y se unen juntos por PCR (preferentemente por número limitado de ciclos de amplificación).

(c) Esta información de unión se revela en un modo altamente paralelo y cuantitativo por secuenciación de nueva generación de ADN.

(d) La información de características de unión predeterminada de fagos individuales y la información de unión de fagos unidos se usa para calcular el interactoma.

El método se basa en la identificación basada en compartimentos de proteínas/complejos de proteína, en los que identidades proteína-anticuerpo de un complejo de proteína individual por compartimento se traducen en ADN. Se usa la identificación de todos los fragmentos de ADN unidos para determinar cuantitativamente las interacciones; sin embargo, los fagos no unidos accidentalmente atrapados en el mismo compartimento pueden contribuir a un origen. Este origen se puede manipular por medios estadísticos simples ya que es un acontecimiento al azar, que se puede diferenciar de acontecimientos específicos. La distribución de agentes de unión durante la compartimentalización está gobernada por la distribución de Poisson, que cuenta así las apariciones de cada agentes de unión (por la determinación de sus abundancias relativas de agentes de unión por NGS después de la amplificación por PCR

- basada en compartimentos limitados), si se conoce el número de compartimentos, así se puede calcular la detección del origen proteínas/complejos de proteína. Se usará una multitud de acontecimientos específicos de la unión para identificar las proteínas diana exactas de los anticuerpos, ya que la información de unión de complejos de múltiples proteínas es debida a la co-localización, e indica unión directa. Como para cada proteínas/complejo de proteína se calcula la detección del origen, cualquier variación de la detección de proteínas/complejos de proteína es debida a efectos reales de la unión, que se pueden calcular por simple resta retirando la detección del origen calculada de proteínas/complejos de proteína (o resta corregida de Poisson ya que las proteínas/complejos de proteína unidos a agentes de unión cambian el número total de agentes de unión). Si se usan combinaciones variadas, en condiciones de equilibrio, del anticuerpo de reacción y analitos de proteína, se puede construir un gráfico de Scatchard u otros cálculos de cinética de unión para calcular las Kds, u otros parámetros para las interacciones anticuerpo-proteína. También se pueden calcular datos cinéticos internos del interactoma para todas o varias interacciones usando diferente concentración de las proteínas interactuantes (cambiando las condiciones experimentales o usando análisis con adiciones).
- 5
- 10

REIVINDICACIONES

1. Un método de determinación de una interacción de unión entre un agente de unión y una diana que comprende
 - 5 a) poner en contacto una biblioteca de agentes de unión con una diana para permitir la formación de complejos agente de unión/diana en donde cada miembro de dicha biblioteca de agentes de unión está asociado con una secuencia de nucleótidos única;
 - b) aislar dichos complejos agente de unión/diana en compartimentos de manera que exista un único complejo agente de unión/diana en un compartimento;
 - 10 c) unir las secuencias de nucleótidos únicas asociadas al (a los) agente(s) de unión en el complejo agente de unión/diana para formar una secuencia de nucleótidos unida en donde el aislamiento de los complejos aislados se mantiene durante la etapa de unión;
 - d) identificar el (los) agente(s) de unión presente(s) en el complejo de la secuencia de nucleótidos unida.
- 15 2. El método según la reivindicación 1, en donde dicha biblioteca de agentes de unión comprende una biblioteca de anticuerpos.
3. El método de la reivindicación 2, en donde dicha biblioteca de anticuerpos comprende una biblioteca de presentación de anticuerpos o una biblioteca de anticuerpos en donde cada anticuerpo está marcado con dicha
 - 20 secuencia de nucleótidos única.
4. El método de cualquier reivindicación precedente, en donde la diana está asociada con una secuencia de nucleótidos única.
- 25 5. El método de cualquier reivindicación precedente, en donde dicha diana comprende una proteína.
6. El método de cualquier reivindicación precedente, en donde dicha diana comprende una biblioteca de presentación de proteínas, en donde cada miembro de dicha biblioteca está asociado con una secuencia de
 - 30 nucleótidos única.
7. El método de la reivindicación 6, en donde dicha biblioteca de presentación de proteínas es una biblioteca de presentación en fagos de ADNc.
8. El método de la reivindicación 5, en donde dicha proteína está dentro de una mezcla de proteínas.
- 35 9. El método de la reivindicación 8, en donde dicha mezcla de proteínas es una mezcla de proteínas enriquecida.
10. El método de la reivindicación 8 o la reivindicación 9, en donde dicha mezcla de proteínas está enriquecida en fosfoproteínas, proteínas de la membrana, y/o proteínas naturalmente o artificialmente modificadas.
- 40 11. El método de cualquier reivindicación precedente, en donde los agentes de unión que se unen a otros agentes de unión se retiran de la biblioteca de agentes de unión antes de uso.
12. El método de cualquier reivindicación precedente, en donde la biblioteca de agentes de unión se enriquece antes
 - 45 de uso.
13. El método de cualquier reivindicación precedente, en donde dicho compartimento es una localización fija sobre un soporte sólido.
- 50 14. El método de cualquier reivindicación precedente, en donde dicho compartimento es una gotita de emulsión, o compartimento separado o limitado a la difusión.
15. El método de cualquier reivindicación precedente, en donde la secuencia de nucleótidos asociada al agente de
 - 55 unión en el complejo se une a otra secuencia de nucleótidos asociada a un agente de unión en el complejo.
16. El método de una cualquiera de las reivindicaciones 4 a 14, en donde la secuencia de nucleótidos asociada a un
 - 60 agente de unión en el complejo se une a la secuencia de nucleótidos asociada a la diana dentro del complejo agente de unión/diana.
17. El método de una cualquiera de cualquier reivindicación precedente, en donde el agente de unión y/o diana
 - 65 presentes dentro del complejo se identifican a partir de la secuencia de nucleótidos unida.
18. El método de cualquier reivindicación precedente, en donde el (i) agente de unión o (ii) agente de unión y diana
 - presentes dentro del complejo se identifican por secuenciación de dicha secuencia de nucleótidos unida.
19. El método de cualquier reivindicación precedente, en donde dicho complejo agente de unión/diana se inmoviliza

tras el aislamiento.

- 5 20. El método de cualquier reivindicación precedente, en donde las secuencias de nucleótidos unidas de más de un complejo se combinan antes de identificar el agente de unión y/o diana presentes.
21. El método de cualquier reivindicación precedente, en donde la etapa (a) de poner en contacto la biblioteca de agentes de unión con una diana se lleva a cabo en presencia y/ o ausencia de un compuesto.
- 10 22. El método de las reivindicaciones 5-20, en donde dicha proteína se reticula con otras proteínas dentro de una mezcla de proteínas.
23. El método de cualquier reivindicación precedente, en donde el aislamiento se lleva a cabo por dilución, unión específica, o aislamiento por propiedades físicas y/o químicas de los complejos.
- 15 24. El método de las reivindicaciones 3-23, en donde la secuencia de nucleótidos única asociada a cada miembro de la biblioteca de fagos de anticuerpos es la secuencia de nucleótidos dentro del fago que codifica la CDR de dicho anticuerpo.
- 20 25. El método de las reivindicaciones 7 a 23, en donde la secuencia de nucleótidos única asociada a cada miembro de la biblioteca de presentación en fagos de ADNc es la secuencia dentro del fago que codifica la proteína presentada.
- 25 26. El método de cualquier reivindicación precedente, en donde la unión de las secuencias de nucleótidos comprende:
- 30 i. Amplificar la secuencia de nucleótidos asociada al agente de unión o agente de unión y diana usando al menos dos pares de cebadores de PCR para producir al menos dos conjuntos de amplicones, en donde los cebadores se diseñan de manera que los amplicones generados por un primer conjunto de cebadores comprendan una secuencia que es complementaria a una secuencia en los amplicones generados por el segundo conjunto de cebadores;
- 35 ii. Hibridar los conjuntos de amplicones;
- iii. Llevar a cabo una reacción de amplificación para producir una secuencia de nucleótidos unida.
27. El método de la reivindicación 26, en donde dicha amplificación se lleva a cabo usando amplificación por PCR en emulsión.
28. El método de la reivindicación 26 o la reivindicación 27, en el que las etapas i-iii se llevan a cabo simultáneamente.
- 40 29. El método de cualquier reivindicación precedente, en donde el método se repite y se varía la concentración de la diana o biblioteca de agentes de unión.
30. Uso de un kit para llevar a cabo el método de cualquier reivindicación precedente, comprendiendo dicho kit
- 45 a) Una biblioteca de agentes de unión en donde cada miembro de dicha biblioteca de agentes de unión tiene características de unión de proteínas predeterminadas y está asociada con una secuencia de nucleótidos única conocida; y
- b) un conjunto de al menos dos pares de cebadores para unir las secuencias de nucleótidos asociados al agente de unión; y opcionalmente instrucciones para su uso.
- 50 31. El uso de un kit de la reivindicación 30, comprendiendo dicho kit además una biblioteca de presentación de proteínas en donde cada miembro de dicha biblioteca está asociado con una secuencia de nucleótidos única.

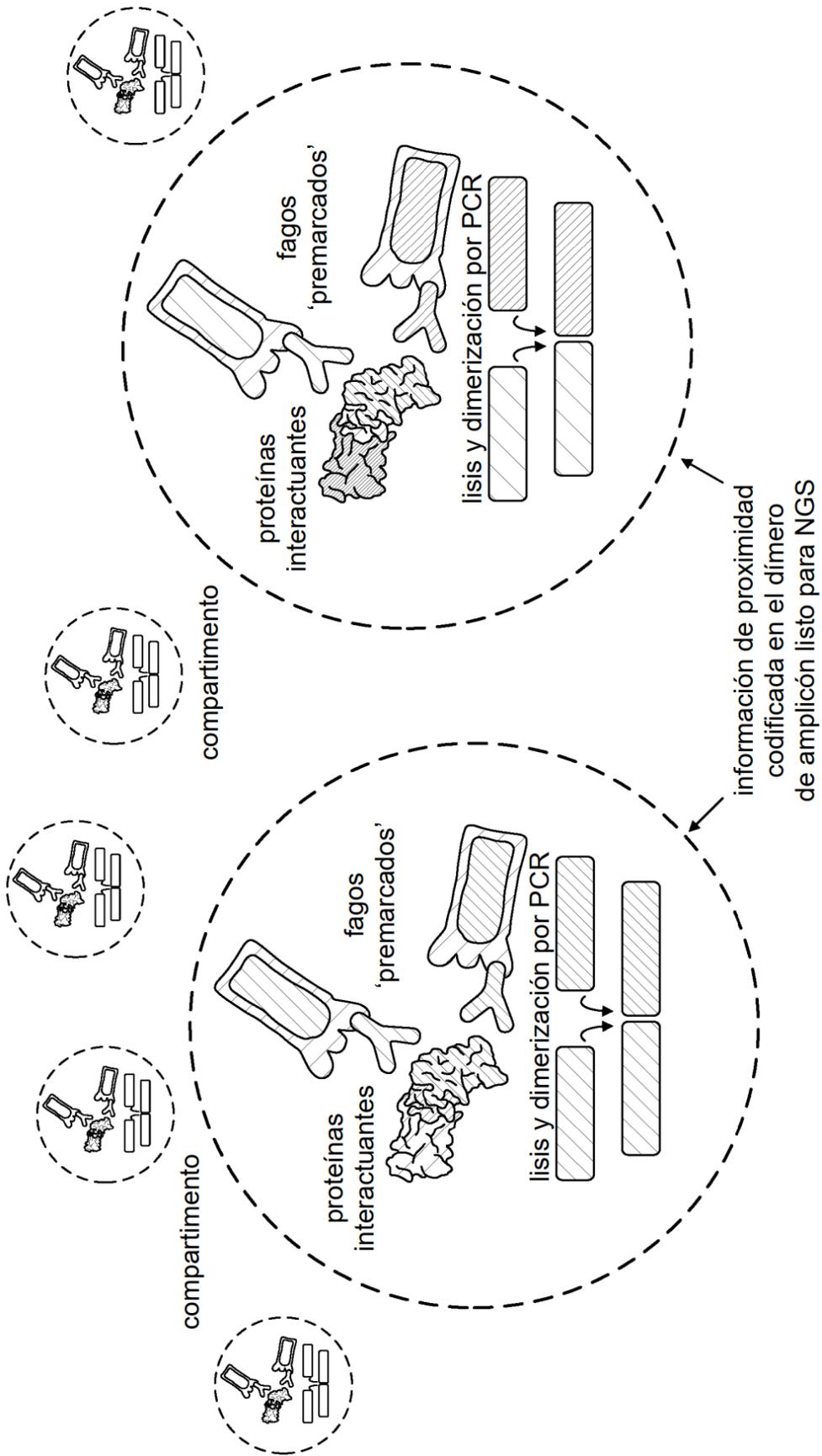


FIG. 1

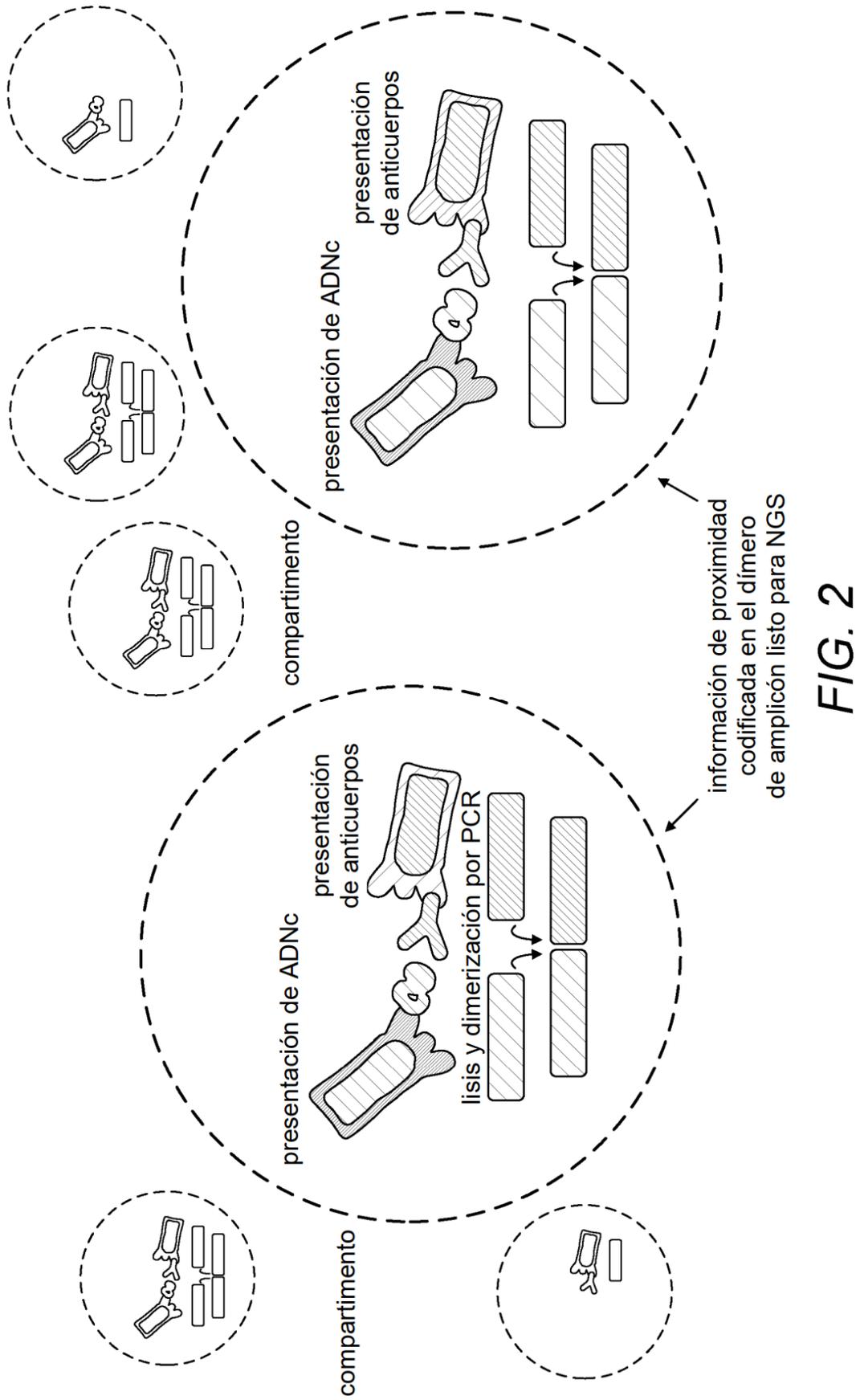


FIG. 2