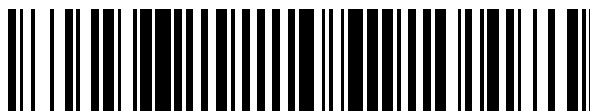


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 402 878**

51 Int. Cl.:

C12N 15/10	(2006.01) C07K 7/06	(2006.01)
G01N 33/68	(2006.01) G01N 33/569	(2006.01)
C40B 30/04	(2006.01) G01N 33/574	(2006.01)
C12N 15/11	(2006.01) C40B 40/10	(2006.01)
C12Q 1/04	(2006.01)	
C40B 40/08	(2006.01)	
C40B 50/06	(2006.01)	
C40B 40/06	(2006.01)	
C07K 7/08	(2006.01)	
C07K 14/00	(2006.01)	

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.05.2000 E 10005416 (2)**

97 Fecha y número de publicación de la concesión europea: **16.01.2013 EP 2230303**

54 Título: **Aislamiento de moduladores biológicos a partir de colecciones de fragmentos génicos biodiversos**

30 Prioridad:

05.05.1999 US 132711 P

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

10.05.2013

73 Titular/es:

**PHYLOGICA LIMITED (100.0%)
105 HAY STREET
SUBIACO WA 6008, AU**

72 Inventor/es:

**WATT, PAUL MICHAEL y
THOMAS, WAYNE ROBERT**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 402 878 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

Aislamiento de moduladores biológicos a partir de colecciones de fragmentos génicos biodiversos

Campo de la invención

5 La presente invención se refiere al campo del cribado de colecciones de genes y, más particularmente, a la generación y cribado de colecciones de dominios naturales derivadas de organismos de secuencias genómicas conocidas. Se describen métodos para aumentar la diversidad de dichas colecciones de fragmentos génicos biodiversos adicionalmente mediante procedimientos de mutagénesis. La presente invención proporciona también medios mediante los que pueden desarrollarse un amplio intervalo de productos terapéuticos, profilácticos y reactivos de diagnóstico basados en péptidos.

10 Generalidades

Los expertos en la materia apreciarán que la invención descrita en la presente memoria es susceptible de variaciones y modificaciones distintas de las descritas específicamente. Ha de entenderse que la invención, como se define por las reivindicaciones adjuntas, incluye todas dichas variaciones y modificaciones. La invención incluye también todas las etapas, rasgos, composiciones y compuestos a que se hace referencia o que se indican en la memoria descriptiva, individual o colectivamente, y todas y cada una de las combinaciones de dos cualesquiera o más de las etapas o rasgos que entren dentro del alcance de las reivindicaciones adjuntas.

15 La presente invención no ha de estar limitada en su alcance por las realizaciones específicas descritas en la presente memoria, que se pretenden solo con fines de ejemplificación. Productos, composiciones y métodos funcionalmente equivalentes que entren dentro del alcance de las reivindicaciones adjuntas están claramente dentro del alcance de la invención.

Los detalles bibliográficos de las publicaciones a las que se hace referencia numérica en esta memoria descriptiva se recogen al final de la descripción.

No se admite que ninguna de las referencias constituya técnica anterior.

25 Como se usa en la presente memoria, "derivado de" se entenderá que indica que puede obtenerse una entidad específica a partir de una fuente particular aunque no necesariamente directamente a partir de esa fuente.

A lo largo de esta memoria descriptiva y las reivindicaciones siguientes, a menos que el contexto requiera otra cosa, se entenderá que la palabra "comprender", o variaciones tales como "comprende" o "comprendiendo", implica la inclusión de una entidad o grupo de entidades citadas, pero no la exclusión de ningún otra entidad o grupo de entidades.

30 Antecedentes de la invención

Las interacciones/actividades biológicas tales como interacciones proteína:proteína, interacciones antígeno:anticuerpo, interacciones proteína:ácido nucleico, interacciones proteína:ligando e interacciones ácido nucleico:ácido nucleico están implicadas en una amplia variedad de procesos que ocurren en las células vivas. Por ejemplo, el agonismo y antagonismo de receptores por ligandos específicos, interacciones anticuerpo-antígeno, 35 incluyendo fármacos, hormonas, moléculas de segundo mensajero, etc. pueden afectar a una variedad de procesos biológicos tales como expresión génica, diferenciación y crecimiento celular, actividad enzimática, flujo metabólico y reparto metabólico entre los compartimentos celulares, entre otros. Las interacciones ADN:proteína y ARN:proteína son bien conocidas por sus efectos en la regulación de la expresión génica en células tanto procarióticas como eucarióticas, además de ser críticas para la replicación de ADN y, en el caso de ciertos virus, la replicación de ARN. 40 En casos en que la propagación de células es perjudicial tales como la replicación de un patógeno o una célula cancerosa, los agentes que se orientan a interacciones/actividades biológicas o estructuras funcionales son candidatos adecuados para terapia. Por ejemplo, los agentes que bloquean la función de canales de membrana o destruyen membrana citoplasmática por otros medios son dianas atractivas para terapias antimicrobianas contra patógenos. Adicionalmente, los agentes que interaccionan con funciones específicas de antígeno o no específicas de antígeno del sistema inmunitario pueden proporcionar moduladores inmunológicos o vacunas para alergia, 45 autoinmunidad, enfermedades infecciosas, fertilidad y envenenamiento. Por ejemplo, los agentes que tienen antigenicidad de antígenos microbianos, antígenos tumorales, alérgenos o autoantígenos pueden usarse para vacunas o inmunoterapia.

Una expresión génica y/o diferenciación celular, crecimiento celular y metabolismo indeseables o inapropiados pueden ser atribuibles también, al menos en muchos casos, a interacciones/actividades biológicas que implican la unión y/o actividad de moléculas proteicas, tales como factores de transcripción, hormonas peptídicas, moléculas receptoras y enzimas, entre otras. En estos casos, pueden preverse terapias que bloqueen dichas interacciones inapropiadas y/o que bloqueen la formación de estructuras celulares inapropiadas.

Producción de péptidos mediante técnicas de ADN recombinante

Los péptidos que pueden mediar en o interferir con un intervalo diverso de funciones biológicas incluyen péptidos naturales y péptidos sintetizados para representar una porción o porción modificada de una molécula conocida por mediar una función diana. Son una fuente de dichos péptidos las colecciones de péptidos aleatorios construidas con oligonucleótidos aleatorios (o semialeatorios) ligados en sitios de clonación de vectores de plásmido o fago.

Se transfectan o transforman vectores que contienen ADN que codifica diferentes péptidos en bacterias u otros hospedadores y se clonan mediante procedimientos de purificación en placa o colonia estándares. Los clones productores de péptidos con una actividad deseada pueden aislarse mediante una variedad de procedimientos de cribado o selección que son fundamentalmente iguales que los procedimientos de cribado usados para detectar polipéptidos codificados por ADNc o fragmentos de ADNc. Estos incluyen la producción de péptidos como fusiones con proteínas de cubierta de bacteriófago o fusiones con proteínas de superficie bacteriana, de modo que los péptidos puedan usarse como marcajes para procedimientos de purificación por afinidad; la producción de péptidos a partir de hospedadores infectados con fago o transformados con plásmidos para producir series de colonias o placas en que puede cribarse la actividad de unión a ligando o actividad biológica tal como inhibir el crecimiento de bacterias diana o inducir la activación de genes en bacterias diana; y en estrategias de selección positivas tales como sistemas de clonación de doble híbrido, en que el péptido producido en el microorganismo hospedador se une a proteínas diana formando complejos que activan la expresión de los genes indicadores clonados en el mismo hospedador. Una de las ventajas significativas de la tecnología de presentación en fago es que posibilita la construcción de colecciones de muy altas complejidades, concretamente de 10^{10} a 10^{11} clones individuales.

Igualmente, en sistemas de “doble híbrido inverso” o “doble híbrido dividido”, pueden cribarse en colecciones de péptidos expresados apropiadamente los bloqueantes de interacciones proteína/proteína particulares, lo que a su vez reduce la expresión de genes indicadores contraseleccionables que codifican productos tóxicos.

Modificación de péptidos para utilidad y optimización

Una vez se ha identificado el péptido activo o péptido de unión a ligando, pueden modificarse mediante una variedad de procedimientos para optimizar su utilidad. La modificación puede incluir: alteraciones de los residuos aminoácidos que interactúan con la diana para mejorar su especificidad de unión y afinidad; modificaciones que afectan a la presentación del péptido, incluyendo la valencia de unión y la limitación de conformaciones particulares y modificaciones para enlazar restos funcionales adicionales tales como marcadores, toxinas y coactivadores.

Los péptidos sintéticos pueden incluir residuos distintos de los 20 aminoácidos encontrados en la naturaleza y/o pueden ciclarse por medios tales como la oxidación de residuos de cisteína flanqueantes. En el caso de péptidos que imitan epítopos de anticuerpo, pueden añadirse mediante técnicas genéticas portadores que contienen los epítopos de linfocitos T necesarios para inducir respuestas inmunitarias de alta afinidad.

Ejemplos de péptidos que modulan sistemas biológicos

Los péptidos pueden aplicarse como productos terapéuticos o moléculas de partida para diseñar productos terapéuticos para enfermedades que incluyen infección, cáncer y trastornos metabólicos, así como agentes para vacunas e inmunoterapia, trasplante y diagnóstico. La utilidad potencial de dichos péptidos se ha demostrado por los siguientes ejemplos

Agentes peptídicos antimicrobianos

Se ha demostrado un efecto antimicrobiano por péptidos naturales producidos por ranas e insectos y péptidos catiónicos sintetizados artificialmente. Una gran variedad de antibióticos son péptidos o polipéptidos. Los gránulos de neutrófilos de mamífero producen familias de polipéptidos antimicrobianos que incluyen azurocidina, cathepsina G y péptidos catiónicos antimicrobianos (CAP57 y CAP37). Además, los neutrófilos producen al menos dos familias de péptidos antimicrobianos, las defensinas y las batenecinas. Además, muchos antibióticos y fármacos antifúngicos están compuestos por péptidos. Por ejemplo, la familia de péptidos α -helicoidales antimicrobianos de la magainina, aislada de la piel de la rana africana de uñas, *Xenopus laevis*, forma poros letales en las membranas celulares de ciertos microorganismos. De forma similar, ciertos péptidos α -helicoidales derivados de una variedad de géneros de insectos tienen actividad antimicrobiana. Recientemente, se han usado varios enfoques de diseño racional para aislar antibióticos peptídicos novedosos. Por ejemplo, Tiozzo *et al.* usaron un enfoque de “molde de secuencia” en que se diseñaban secuencias peptídicas candidatas a partir de alineamientos de péptidos antimicrobianos naturales [1]. La identificación de determinantes de virulencia en varios patógenos presenta otras dianas atractivas para la terapia antimicrobiana. Por ejemplo, Balaban y colaboradores (2) han identificado recientemente un autoinductor de virulencia en *Staphylococcus aureus* que controla la producción de toxinas bacterianas implicadas en la patogénesis. Los genes de toxina se inducen por una molécula de ARN reguladora, RNAIII, que está inducida por una concentración umbral de una proteína activadora de RNAIII endógena (RAP) [2]. Podría esperarse que los inhibidores peptídicos de RAP actuaran como determinantes de virulencia. Es más, se produce un inhibidor peptídico natural de RAP llamado RIP (péptido inhibidor de RNAIII) por una cepa no patógena de *Staphylococcus aureus* y parece inhibir el gen de RNAIII y causar una virulencia reducida [2].

Moduladores peptídicos de la regulación del crecimiento

Se ha demostrado la capacidad de péptidos para afectar a moduladores clave de la regulación del crecimiento por Brent y colaboradores, que usaron el cribado de doble híbrido para identificar “aptámeros” peptídicos restringidos de colecciones combinatorias que se unen estrechamente e inhiben la función de la cinasa dependiente de ciclina 2. Esto demuestra el potencial para el tratamiento de neoplasmas (3).

Los péptidos pueden exhibir una especificidad exquisita. Por ejemplo, se han identificado aptámeros peptídicos que pueden discriminar entre dos variantes alélicas estrechamente relacionadas de la proteína Ras (4). Además, un aptámero peptídico contra cinasa dependiente de ciclina 2 humana inhibe la actividad cinasa exclusivamente en ciertos sustratos particulares.

La especificidad peptídica se ha demostrado también *in vivo*. En un informe reciente, se ha mostrado que la expresión de aptámeros que reconocían cinasas dependientes de ciclina en moscas transgénicas causaba anomalías de desarrollo de forma negativa dominante (5). De forma importante, la especificidad de los dos aptámeros por Cdk particulares (como se determina mediante ensayos de doble híbrido de levadura) se retenía en el ensayo de *Drosophila in vivo*. Además, la coexpresión del aptámero específico de Cdk diana suprimía el fenotipo de desarrollo observado (5). Este informe de inhibición orientada exitosa de una enzima *in vivo* con aptámeros establece firmemente como practicable el principio de desarrollo de nuevas estrategias terapéuticas basadas la interferencia de péptidos.

Inhibición basada en péptidos: una estrategia terapéutica emergente

Se ha centrado recientemente mucha atención sobre los péptidos como agentes terapéuticos potenciales porque pueden ser altamente específicos y se sintetizan fácilmente. Las tecnologías de presentación en fago están empezando a probarse útiles para proporcionar moléculas de partida peptídicas en programas de descubrimiento de fármacos. El suministro eficaz de péptidos desde fuera de la célula al núcleo de células eucarióticas puede conseguirse ahora enlazando secuencias tales como el motivo orientador “penetratina”, que deriva de la proteína Antennapedia de *Drosophila*. Más recientemente, se ha identificado una familia de dichos péptidos orientadores (6). Por ejemplo, se ha mostrado que la conjugación de secuencias peptídicas con la proteína VP22 permite una exportación eficaz de la proteína de fusión a los núcleos de células adyacentes a los transfectantes primarios (7). Varios desarrollos recientes hacen factible seleccionar físicamente dominios peptídicos conformacionalmente restringidos para identificar péptidos que se unan con muy alta afinidad *in vivo*, favoreciendo una alta potencia. Se han reseñado péptidos miméticos que inhiben interacciones proteicas y/o funciones enzimáticas. Los ejemplos incluyen un nonapéptido derivado de ribonucleótido reductasa de herpesvirus simple que se ligó con una subunidad de enterotoxina para el suministro a células a través de su receptor. Se encontró que el conjugado peptídico inhibe la replicación de herpesvirus simple de tipo 1 en células Vero en reposo [8]. Usando el conocimiento detallado del dominio de interacción con PCNA de p21^{WAF1} derivado de cribados de doble híbrido, se ha diseñado un péptido que bloquea eficazmente la interacción. Este péptido de 20 unidades se unía con suficiente afinidad para bloquear la replicación de SV40. Se ha encontrado que una secuencia peptídica de 20 unidades derivada de p16 interacciona con Cdk4 y Cdk6 e inhibe la fosforilación de pRB y la progresión del ciclo celular [9]. Los autores acoplaron el péptido inhibidor específico con el péptido penetratina de 16 residuos para un suministro nuclear eficaz. Se ha mostrado que los péptidos funcionan como inhibidores en modelos animales. Por ejemplo, se ha mostrado también que un tetrapéptido que imita el sustrato de farnesil proteína transferasa bloquea el crecimiento de tumores dependientes de Ras en ratones atímicos.

Mimótopos peptídicos

Se han aislado péptidos que se parecen funcionalmente a epítopos (mimótopos) unidos por anticuerpos y se han usado como vacuna experimental para inducir anticuerpos que protejan frente a infecciones como los mostrados para hepatitis B, virus respiratorio sincitial, encefalitis japonesa y neumonía por *Streptococcus*. Los anticuerpos de alta afinidad se unen típicamente a estructuras complejas formadas por la conformación terciaria de un antígeno. Los mimótopos peptídicos convierten esencialmente un epítipo conformacional compuesto por una proteína completa en un péptido pequeño. Tiene ventajas cuando se desean solo ciertos epítopos, por ejemplo, para prevenir la inmunopatología en infección por RSV, o en la producción de epítopos recombinantes en que el polipéptido completo puede ser difícil de plegar, o cuando el antígeno entero tiene propiedades biológicas indeseables (toxinas estafilocócicas en el síndrome de choque tóxico). En el caso de antígenos de carbohidrato, pueden construirse polipéptidos que contienen el mimótopo convirtiendo un antígeno independiente de linfocitos T en un antígeno dependiente de linfocitos T para la producción de anticuerpos de alta afinidad e inmunogenicidad en animales jóvenes, incluyendo seres humanos. Al contrario que los carbohidratos, los mimótopos peptídicos pueden producirse como vacunas de ADN.

La posibilidad de usar mimótopos como antígenos para inmunoterapia del cáncer se ha demostrado para un antígeno de adenocarcinoma

Los mimótopos pueden usarse como antígenos para diagnosticar una enfermedad infecciosa detectando el anticuerpo. Se ha demostrado la posibilidad con infección por hepatitis C.

Se han demostrado mimótopos que representan antígenos reconocidos por autoanticuerpos contra tejido de islotes β en diabetes, y se ha propuesto que estos podrían usarse para monitorizar el desarrollo de la enfermedad (10). De forma similar, se han encontrado mimótopos para alérgenos de polen que podrían usarse en el diagnóstico de enfermedad alérgica. En ambos de estos casos, es también posible que los mimótopos pudieran usarse para terapia modulando la respuesta inmunitaria o en profilaxis.

Se han demostrado mimótopos que representan antígenos de transplante y por tanto pueden usarse como tolerógenos o bloqueantes para evitar el rechazo de transplante.

Interacciones de ligando o interacciones de receptor hormonal

Se han usado miméticos peptídicos como ligandos para purificar por afinidad moléculas biológicamente útiles como se muestra para la purificación de la proteína de coagulación sanguínea factor de von Willebrand.

Se ha demostrado también la modificación de la actividad enzimática con péptidos que imitan sustratos. Los miméticos peptídicos pueden usarse como hormonas como se muestra para eritropoyetina y pueden modificarse para aumentar la actividad biológica.

Métodos recombinantes para la producción de péptidos biológicamente activos

El uso de fragmentos de genes específicos o ADNc para producir péptidos que contienen una actividad biológica del polipéptido codificado por el gen o un inhibidor de la actividad puede ser a veces exitoso. En otros casos, la actividad puede depender de la conformación del polipéptido completo y no puede obtenerse mediante estas técnicas. En muchos casos, ha sido exitoso el uso de colecciones de péptidos aleatorios en fago o plásmidos para producir un péptido que imite la actividad biológica. Esto implica el cribado en grandes números de clones productores de una serie esencialmente aleatoria de péptidos para un péptido de la actividad deseada. La actividad está mediada a veces por un péptido que muestra una homología de secuencia aminoacídica que podría explicar su actividad biológica, mientras que en muchos casos el péptido actúa como mimético de la conformación del polipéptido o su ligando y no tiene homología de secuencia. Es más, el péptido puede ser un mimético de una molécula químicamente diferente tal como un carbohidrato. También es posible usar el enfoque de colección combinatoria para cribar inhibidores o mediadores de funciones complejas cuando no hay información sobre las interacciones moleculares requeridas.

La capacidad de aislar péptidos activos a partir de colecciones de fragmentos aleatorios puede ser sin embargo altamente variable y se han reseñado problemas con interacciones de baja afinidad, particularmente para péptidos que se requiere que presenten conformaciones complejas tales como epítopos discontinuos unidos por muchos anticuerpos. Existe impredecibilidad en que las colecciones que son una fuente rica en péptidos para un ligando pueden no contener péptidos para otros. Aunque la capacidad de obtener péptidos deseados debería aumentar con colecciones que contienen péptidos aleatorios mayores y más péptidos aleatorios, hay dificultades prácticas para realizar un cribado o purificación por afinidad de alto rendimiento, puesto que se ha mostrado que la purificación por afinidad de alta densidad es ineficaz. Existen también incertezas sobre el grado en que los péptidos aislados a partir de las colecciones de péptidos aleatorios retendrán su actividad de unión o biológica cuando se produzcan como parte de estrategias de suministro diferentes, tales como fusiones con diferentes polipéptidos. Existe por tanto la oportunidad de suplementar o mejorar la tecnología existente con nuevas estrategias.

Colecciones de dominios peptídicos biodiversos de fuentes genómicas definidas

Los péptidos presentan potencial terapéutico y son agentes profilácticos para muchas enfermedades humanas y animales, trastornos bioquímicos y efectos adversos de fármacos porque pueden interaccionar con otras moléculas con alta especificidad y afinidad. Sin embargo, un problema importante para superar en el campo de la terapia y profilaxis peptídicas es la identificación de las secuencias aminoacídicas específicas que tienen la actividad antagonista o agonista deseada frente a una actividad biológica particular en un entorno celular particular. Dichos candidatos a fármacos peptídicos pueden ser particularmente difíciles de identificar a partir de colecciones de péptidos verdaderamente aleatorias que carezcan de cualquier enriquecimiento en secuencias que codifican formas moleculares adecuadas para la unión a estructuras biológicas. En contraposición, la naturaleza ha ensamblado ya una fuente rica en dichos dominios con la pluralidad de péptidos, polipéptidos y proteínas codificados por el diverso intervalo de genomas que constituyen la biosfera.

Se han propuesto un amplio intervalo de métodos diferentes para facilitar el cribado de colecciones biológicas (tales como colecciones de ADNc) de manera conveniente para identificar las moléculas de proteína o polipéptido adecuadas. Se han preparado colecciones de miles, y en algunos casos incluso millones, de polipéptidos o péptidos mediante sistemas de expresión génica y presentado sobre soportes químicos o en sistemas biológicos adecuados para ensayar la actividad biológica. Generalmente, dichas colecciones están compuestas por genomas u organismos individuales que se cree que son fuentes ricas en nuevos fármacos (tales como especies bacterianas "extremófilas") o por una mezcla de genomas no caracterizados aislados directamente del entorno.

Aunque el cribado de colecciones biodiversas se ha probado valioso, dichas colecciones tienden a estar sesgadas hacia la frecuencia con que se encuentra un organismo particular en el entorno nativo y pueden no representar

necesariamente la población verdadera de la biodiversidad encontrada en una muestra biológica particular. Además, dichos cribados se pretenden normalmente para aislar genes que codifican enzimas, por ello a menudo se hacen intentos de sesgar dichas colecciones para que contengan insertos mayores que podría esperarse que codificaran enzimas biológicamente activas.

En la patente de EE.UU. 5.763.239 a nombre de Short *et al.*, se describe un procedimiento para normalizar colecciones de ADN genómico a partir de una muestra ambiental, en un intento de enfrentarse a este problema del sesgo. Debido a que las colecciones mencionadas en esa patente se generan a partir de muestras ambientales, por lo que se sabe poco sobre la constitución genómica de la colección, el procedimiento emplea métodos de normalización complicados para normalizar la constitución genómica de las colecciones. Aunque ese procedimiento permite cierta normalización de los genomas en una muestra ambiental, los métodos que describe son complicados, existe el riesgo de perder ADN genómico raro cuando se aplican los métodos y/o de que se introduzcan nuevos sesgos por el procedimiento.

Además de lo anterior, los métodos de cribado actuales a menudo se basan en el aislamiento de secuencias de ácido nucleico genómico usando procedimientos de amplificación por PCR, por lo que se sabe poco sobre las secuencias genómicas. En dichos casos, pueden introducirse sesgos mediante factores tales como la presencia de una representación desproporcionada de secuencias repetidas en ciertos genomas. Además, debido a que no se conoce información sobre la constitución genómica de la muestra ambiental, solo pueden derivarse datos bioinformáticos limitados a partir del cribado de la colección. Este problema se enfrenta en cierta medida en la patente de EE.UU. 5.763.239, que intenta aumentar la probabilidad de que una secuencia genómica de bajo número de copias en una muestra ambiental tenga la oportunidad de estar representada en una colección.

Sin embargo, no existen actualmente métodos disponibles para cribar colecciones de dominios peptídicos biodiversos normalizados *in vivo* en los que pueda estimarse exactamente la composición y complejidad completa de la colección y en los que el proceso de cribado proporcione datos bioinformáticos integrales tales que sean útiles para un diseño racional de fármacos. Además, no se han descrito métodos que estén diseñados específicamente para la construcción de colecciones de secuencias genómicas naturales que se han optimizado para la expresión de dominios *per se*, en lugar de polipéptidos enteros. Por consiguiente, existe la necesidad de desarrollar tecnologías que proporcionen un cribado a gran escala de colecciones de péptidos que estén enriquecidas en secuencias que codifican dominios bioactivos útiles en la determinación de productos terapéuticos peptídicos útiles, cuya base no esté necesariamente relacionada con el papel natural de los dominios peptídicos particulares.

Sumario de la invención

Las proteínas de diferente función muestran evidencias de evolución mediante reordenamiento de dominios (por ejemplo, factor de crecimiento nervioso y receptores de lipoproteína de baja densidad) o mediante pequeñas modificaciones de diferentes residuos en dominios conservados (serinproteasas). La presente invención intenta imitar esta evolución usando colecciones de péptidos codificados por fragmentos de secuencia nucleotídica conocidos y definidos que son una fuente rica en péptidos que contienen secuencias aminoacídicas evolucionadas para diversas interacciones moleculares no necesariamente relacionadas estrechamente con la función efectuada en el organismo donante. Se describen también medios para aumentar adicionalmente la diversidad de las colecciones de fragmentos génicos biodiversos mediante mutagénesis, tanto *in vitro* usando amplificación por PCR en condiciones mutagénicas como *in vivo* mediante replicación de la colección en cepas bacterianas "mutadoras" que contienen mutaciones en los genes implicados en la reparación de errores de emparejamiento del ADN.

La presente invención proporciona una colección de presentación en fago que comprende una pluralidad de diferentes secuencias nucleotídicas a partir de una pluralidad de organismos biodiversos que tienen cada uno un genoma secuenciado, y en la que los organismos son microorganismos y/o eucariotas que tienen genomas compactos.

La presente invención proporciona también un método de producción de una colección de presentación en fago que comprende una pluralidad de diferentes secuencias nucleotídicas a partir de organismos biodiversos, comprendiendo dicho método producir fragmentos de secuencia nucleotídica a partir de una secuencia nucleotídica de composición nucleotídica conocida, en el que dicha secuencia nucleotídica sea de un genoma secuenciado de un microorganismo y/o de un genoma compacto secuenciado de una especie eucariótica, y ligar los fragmentos en un vector de fago.

La presente invención proporciona también un método para identificar un modulador o mediador de una actividad biológica, incluyendo dicha actividad antigenicidad y/o inmunogenicidad, comprendiendo dicho método:

- (i) producir una colección de presentación en fago derivada de fragmentos de secuencia nucleotídica definida según el método descrito anteriormente y
- (ii) ensayar en al menos una secuencia aminoacídica derivada de la etapa (i) de la colección de presentación en fago una actividad biológica en la que la actividad sea diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo.

Las colecciones de la invención pueden usarse en diversos métodos de cribado y se describen en la presente memoria algunos de dichos métodos.

Se apreciará que la presente invención tiene una aplicación de amplio alcance para identificar secuencias aminoacídicas que tengan una actividad novedosa en comparación con la que puede reconocerse que tienen en su entorno natural ordinario. Por ejemplo, la presente invención es particularmente útil para cribar en colecciones de expresión de fragmentos genómicos secuencias aminoacídicas reactivas con anticuerpos particulares, por ejemplo, mediante cromatografía de afinidad de una colección de presentación en fago. Además, la presente invención proporciona un medio para definir los aminoácidos esenciales para modular una actividad biológica tal como, por ejemplo, unión a anticuerpo. Proporciona también un medio para aislar moduladores o mediadores de secuencia aminoacídica de una actividad biológica, que pueden funcionar independientemente de las limitaciones artificiales del sistema de cribado mediante el que se identificaron (por ejemplo, fusiones génicas, etc.).

En particular, la presente invención es particularmente útil para identificar productos terapéuticos novedosos tales como vacunas o antígenos inmunoterapéuticos, antibióticos o agentes inhibidores que puedan servir como candidatos a agonistas o antagonistas de cualquier actividad biológica. Por ejemplo, pueden usarse colecciones de fragmentos génicos biodiversos para producir antígenos que pueden usarse para vacunas o para inmunoterapia de enfermedades alérgicas o enfermedades autoinmunitarias. En el caso de inmunoterapia de alérgenos, es especialmente deseable obtener un péptido de alta afinidad (que es raro en colecciones de péptidos aleatorios) porque puede usarse como antígeno monovalente para evitar la reticulación de IgE en mastocitos.

Este sistema puede usarse también en el cribado de alto rendimiento de agentes orientados a interacciones proteína:ADN, péptido:ADN o péptido:proteína, proteína:proteína específicas o a una estructura tal como la pared celular o un componente de transporte de membrana.

Es una ventaja notable de la tecnología descrita en la presente memoria que, al tener un mayor control sobre la composición de una colección de expresión de secuencias aminoacídicas por saber su constitución definida, puede maximizarse intencionadamente la distancia filogenética entre los genomas constituyentes de la colección para asegurar el máximo grado de diversidad que, en principio, podría rivalizar con la diversidad de secuencia de muestras genómicas derivadas del entorno, a pesar del hecho de que dichas muestras pueden contener más diversidad de especies *per se*. Este enfoque se volverá más potente a medida que el intervalo de secuencias nucleotídicas disponibles aumente adicionalmente.

En una realización, se proporciona un método para identificar un modulador o mediador de una actividad biológica, incluyendo dicha actividad antigenicidad y/o inmunogenicidad, comprendiendo dicho método las etapas de:

- (i) producir una colección de expresión de fragmentos génicos derivada de fragmentos de secuencia nucleotídica definida, codificando dicha secuencia nucleotídica al menos una secuencia de aminoácidos;
- (ii) ensayar en al menos una secuencia aminoacídica de la colección de expresión derivada de la etapa (i) una actividad biológica, en la que la colección esté adaptada para presentar un intervalo de secuencias aminoacídicas cada una de las cuales puede variar en al menos un aminoácido y
- (iii) identificar aquellos aminoácidos esenciales para modular la actividad biológica, siendo dicha actividad diferente de la actividad con la que la secuencia no está normalmente asociada en su entorno nativo.

Puede seleccionarse una secuencia de aminoácidos que es particularmente eficaz para modular o mediar una actividad biológica (por ejemplo, antigenicidad o inmunogenicidad) comparando la actividad observada de una serie de diferentes secuencias aminoacídicas de constitución similar. Usando las diferencias en la actividad observada, es posible identificar aquellos aminoácidos esenciales para la actividad y aquellos que son deseables para la actividad o, como alternativa, aquellos que son un impedimento para conseguir una actividad eficaz.

En una segunda realización, el método puede emplearse para identificar péptidos antibacterianos novedosos que se liberan condicionalmente a partir de una proteína de fusión. Según esta realización, se proporciona un método de identificación de un péptido antibacteriano, que comprende:

- (i) transformar o transfectar una primera población bacteriana de células con una colección de expresión peptídica derivada de fragmentos de secuencia nucleotídica definida;
- (ii) cultivar dicha primera población bacteriana durante un tiempo y en condiciones suficientes para que ocurra la expresión de las secuencias aminoacídicas codificadas en dicha colección y para la liberación de las secuencias aminoacídicas a partir de sus fusiones asociadas;
- (iii) poner en contacto las secuencias aminoacídicas expresadas con bacterias patógenas;
- (iv) identificar aquella secuencia o secuencias que sean capaces de inhibir el crecimiento de bacterias patógenas o de destruir las bacterias patógenas y

- (v) seleccionar aquellas secuencias de la etapa de identificación de la etapa (iv) que no estén asociadas a la inhibición del crecimiento de bacterias patógenas o la destrucción de bacterias patógenas en su entorno nativo.

5 En una tercera realización, se proporciona un método para identificar un modificador de una actividad biológica asociada a una célula hospedadora, comprendiendo dicho método las etapas de:

- (i) expresar una molécula indicadora bajo el control operativo de la actividad biológica en la célula, en la que al menos una molécula asociada a la actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor;
- 10 (ii) incubar al menos una célula de la etapa (i) en presencia de una secuencia o secuencias aminoacídicas de una colección de expresión de fragmentos génicos derivada de una secuencia genómica definida, en condiciones que promuevan la interacción entre la secuencia o secuencias aminoacídicas y una secuencia nucleotídica o aminoacídica implicada en la actividad biológica; y
- (iii) identificar al menos una secuencia aminoacídica que, en presencia de las células, sea capaz de modificar la expresión de dicha molécula indicadora, o la actividad biológica; y
- 15 (iv) seleccionar aquellas secuencias de la etapa (iii) que no se reconoce generalmente que son capaces de modificar la expresión de dicha molécula indicadora, o la actividad biológica, en su entorno nativo.

Preferiblemente, el método descrito en esta realización se repite las veces que sea necesario para asegurar que sustancialmente todos los aminoácidos codificados por la secuencia nucleotídica definida se presentan ante la actividad biológica.

20 En una cuarta realización, se proporciona un método de identificación de un antagonista de una actividad biológica, comprendiendo dicho método las etapas de:

- (i) disponer la expresión de una molécula indicadora bajo el control operativo de una actividad biológica en una célula, en la que al menos un copartícipe de dicha actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor expresable bacteriano en un vector adecuado, en el que (a) la secuencia nucleotídica deriva de una secuencia nucleotídica de origen conocido y secuenciada y (b) la actividad biológica es diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo;
- 25 (ii) incubar la célula en presencia de un compuesto candidato para ensayar la capacidad de antagonizar la actividad biológica y
- 30 (iii) (iii) seleccionar células en las que se modifica la expresión de dicha molécula indicadora, o la actividad biológica.

35 Puede usarse en la presente invención cualquier secuencia nucleotídica de composición nucleotídica conocida. Preferiblemente, la secuencia nucleotídica deriva de un genoma sustancialmente secuenciado de un microorganismo y/o especie eucariótica compacta (concretamente, una especie con una alta proporción de secuencia que codifica polipéptido). Lo más preferiblemente, la secuencia nucleotídica deriva de un genoma secuenciado totalmente de un microorganismo y/o genoma compacto de una especie eucariótica que es un genoma que contiene un alto porcentaje de ADN que codifica polipéptidos.

40 Deseablemente, la presente invención emplea una colección de expresión peptídica preparada a partir de una secuencia genómica definida presente en el aislamiento o en combinación con otras secuencias genómicas definidas para identificar la secuencia o secuencias aminoacídicas que pueden ser candidatos adecuados para el diseño racional de fármacos, proporcionando sustancialmente al mismo tiempo datos bioinformáticos integrales sobre estos candidatos. Los datos bioinformáticos derivados del método pueden usarse para identificar aquellos aminoácidos importantes en la modulación de la actividad biológica.

45 En una quinta realización, se proporciona un método para identificar un modulador de una actividad biológica, comprendiendo dicho método las etapas de:

- (i) producir una colección de expresión de aminoácidos derivada de una secuencia genómica definida,
- (ii) poner en contacto una secuencia aminoacídica derivada de la colección de expresión con una molécula indicadora que está bajo el control operativo de una actividad biológica asociada a un hospedador y
- 50 (iii) identificar una secuencia aminoacídica capaz de modular la actividad biológica en la que esa actividad sea diferente de cualquier actividad que pueda tener la secuencia aminoacídica en su entorno nativo.

En una sexta realización, se proporciona un método para identificar una secuencia aminoacídica que es capaz de modular una actividad biológica en una célula hospedadora, comprendiendo dicho método las etapas de:

- (i) producir una colección en un hospedador en la que (a) las células transformadas de dicha colección contiene al menos una primera secuencia nucleotídica que comprende o codifica una molécula indicadora cuya expresión esté bajo el control operativo de dicha actividad biológica y una segunda secuencia nucleotídica derivada de una secuencia genómica conocida que sea capaz de codificar la secuencia aminoacídica cuando se dispone bajo el control operativo de una secuencia promotora adecuada, y en la que (b) sustancialmente toda la secuencia genómica conocida esté presente en la población de células transformadas que constituye dicha colección y la actividad biológica sea diferente de cualquier actividad que pueda tener la secuencia aminoacídica en su entorno nativo,
- (ii) cultivar dicho hospedador celular durante un tiempo y/o en condiciones suficientes para que ocurra la expresión de dicha segunda secuencia nucleotídica y
- (iii) seleccionar o cribar células en las que la expresión de dicha molécula indicadora esté modificada.

Preferiblemente, el método definido por la sexta realización incluye también las etapas de adicionales de:

- (iv) comparar el intervalo de secuencias aminoacídicas que pueden derivar de la secuencia genómica conocida frente a aquellas secuencias que exhibían actividad biológica y
- (v) determinar aquellos aminoácidos que son esenciales para modificar la actividad de la molécula indicadora.

En una forma particularmente preferida de la invención, pueden expresarse una pluralidad de secuencias genómicas definidas derivadas de diferentes organismos en la colección de fragmentos génicos. Cuando se usan en el método las secuencias genómicas de más de un organismo, cada una de las secuencias se proporciona preferiblemente en iguales cantidades molares para asegurar que se incluye una proporción igual de secuencias en el método.

La complejidad de la colección de expresión de fragmentos génicos puede aumentarse también sometiendo la secuencia o secuencias genómicas definidas derivadas de esas secuencias a métodos que malinterpreten o muten la secuencia o secuencias. Como alternativa, o además, la complejidad de la colección puede aumentarse también expresando la secuencia genómica definida en cada uno de sus diferentes marcos de lectura. Puede expresarse también en sus marcos de lectura inversos. Por tanto, al permitir la expresión de una secuencia génica en cada posible marco de lectura, para cualquier secuencia particular habría seis combinaciones posibles diferentes.

La presente invención contempla también secuencias aminoacídicas identificadas mediante el método de la presente invención, así como el uso de esas moléculas en una composición farmacéutica. La composición farmacéutica comprende una secuencia aminoacídica capaz de modular o mediar una actividad biológica o la función de una molécula biológica y un portador y/o diluyente farmacéuticamente aceptable.

La presente invención proporciona también un vector (o agrupamiento de hasta 3 vectores) capaz de expresar una secuencia nucleotídica en cada uno de sus posibles marcos de lectura, y en el que cada una de las secuencias aminoacídicas así producidas se expresa como una fusión con una segunda secuencia aminoacídica en que están restringidas conformacionalmente, en el que dicho vector comprende al menos:

- (i) un primer módulo de expresión que comprende:
 - (a) un sitio de clonación múltiple para la inserción de una secuencia nucleotídica que codifica dicha secuencia aminoacídica, en el que dicho sitio de clonación múltiple puede ser adyacente a una o más segundas secuencias nucleotídicas que codifican un bucle polipeptídico, de tal modo que puede producirse un polipéptido de fusión entre dichas primera y segunda secuencias aminoacídicas;
 - (b) una secuencia terminadora adyacente al sitio de clonación múltiple y distal de dicha secuencia promotora y segundas secuencias nucleotídicas;
- (ii) un medio para expresar la primera secuencia nucleotídica en cada uno de sus marcos de lectura;
- (iii) un origen de replicación bacteriano y/o un origen de replicación de bacteriófago y
- (iv) un segundo módulo de expresión que codifica un gen marcador de selección bacteriano.

Otro aspecto de la presente invención proporciona la modificación del microorganismo diana cuyo crecimiento o función alternativa puede inhibirse. Este microorganismo puede modificarse con fines de cribado de manera que facilite el cribado, tal como mediante:

- (i) la introducción de marcadores de resistencia a antibióticos novedosos mediante recombinación homóloga, mediante transformación de plásmidos o mediante mutagénesis aleatoria y selección,

- (ii) la introducción (mediante recombinación homóloga o transformación de plásmido) de uno o más genes indicadores (por ejemplo, luciferasa o β -galactosidasa) bajo el control de un promotor endógeno asociado a patología o virulencia. Por ejemplo, los promotores de los genes RNAIII o RAP de *Staphylococcus aureus* podrían usarse para controlar la expresión de un gen indicador que podría detectarse fácilmente. Dichos métodos son bien conocidos por los expertos en la materia, véase la patente internacional (PCT) WO 90/40979, por ejemplo.

La presente invención proporciona también un medio de explotación de los datos bioinformáticos referentes a secuencias homólogas que codifican dominios estructurales en genomas secuenciados, para diseñar colecciones definidas mediante técnicas tales como técnicas de PCR degenerada o síntesis química de ADN que se centran en un dominio de afinidad particular. La diversidad de dicha colección puede aumentarse adicionalmente mediante técnicas de mutagénesis conocidas por los expertos en la materia.

La presente invención proporciona también una técnica de cribado de alto rendimiento para la identificación de clones (de la colección) que produzcan secuencias aminoacídicas capaces de inhibir el crecimiento o reprimir los genes de virulencia del organismo diana patógeno.

Descripción detallada de las realizaciones preferidas

Los métodos de cribado descritos en la presente memoria difieren de los enfoques de diseño racional existentes que intentan modelar candidatos a péptidos terapéuticos basándose en homologías en las bases de datos con péptidos inhibidores naturales. Los enfoques existentes se centran en secuencias aminoacídicas que se han identificado anteriormente a partir de su fuente natural debido a sus propiedades inhibidoras. En contraposición, los métodos descritos en la presente memoria determinan empíricamente las secuencias aminoacídicas que pueden modular una actividad biológica, a partir de una amplia serie de candidatos codificados en una colección de expresión genómica derivada de secuencias nucleotídicas que se han determinado completamente, sin considerar la función original de esas secuencias en la naturaleza.

Los dominios peptídicos y polipeptídicos biológicamente interactivos naturales se cree que han evolucionado mediante selección desde un banco de dominios disponibles en cada organismo en los que surgen. En cualquier organismo, hay una enorme cantidad de diversa información de codificación. Para encauzar esta diversidad, se ha ideado un cribado genético que maximiza la diversidad de un agrupamiento de dominios biológicamente interactivos potenciales. Además, puesto que la información usada en el cribado deriva de información genética secuenciada, puede explotarse la información estructural que ya ha evolucionado en la naturaleza comparando moléculas biológicamente interactivas frente a secuencias similares de una secuencia nucleotídica secuenciada y ensayada. Esta información permite deseablemente la identificación de aminoácidos particulares que son esenciales para la acción de unión de la actividad biológica y/o posiblemente motivos particulares que son esenciales para, o al menos están implicados en, la reacción de unión. Por tanto, la presente invención proporciona métodos de cribado para identificar una secuencia o secuencias aminoacídicas potenciales que son capaces de modular o mediar actividades biológicas que implican péptidos, oligopéptidos, proteínas o secuencias de ácido nucleico.

Por lo tanto, la presente invención reside en un método para identificar un modulador o mediador de una actividad biológica, incluyendo dicha actividad antigenicidad e inmunogenicidad, comprendiendo dicho método las etapas de:

- (i) producir una colección de fragmentos génicos derivada de fragmentos de secuencia nucleotídica definida y
- (ii) ensayar en al menos una secuencia aminoacídica derivada de la etapa (i) de la colección de expresión una actividad biológica en la que la actividad sea diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo.

Como se usa en la presente memoria, el término "actividad biológica" se entenderá que incluye interacciones biológicas que conducen a una asociación física entre dos o más moléculas o "copartícipes". Dicha actividad debería interpretarse en su contexto más amplio e incluye, por ejemplo, interacciones tales como péptido:péptido, péptido:proteína, proteína:proteína, antígeno:anticuerpo, péptido:ácido nucleico, proteína:secuencia de ácido nucleico, péptido:ligando y proteína:ligando. Por ejemplo, la actividad incluye, pero sin limitación, cualquier interacción que module o medie la unión de anticuerpo o unión de antígeno o cualquier otra interacción basada en la secuencia aminoacídica descrita en la sección de antecedentes de esta memoria descriptiva. Preferiblemente, la asociación física implica un proceso celular o, como alternativa, se requiere que ocurra un evento celular, y en la que la actividad es diferente de cualquier actividad que pueda tener la secuencia aminoacídica en su entorno nativo. Además, incluirá actividad que conduce a la destrucción de una estructura y/o actividad biológica. La "asociación física" puede implicar la formación de un campo magnético o campo paramagnético inducido, la formación de enlace covalente tal como la formación de un puente disulfuro entre moléculas polipeptídicas, una interacción iónica tal como la que ocurre en un retícula iónica, un enlace de hidrógeno o, como alternativa, una interacción de van der Waals tal como una interacción dipolo-dipolo, interacción dipolo-dipolo inducido, interacción dipolo inducido-dipolo inducido o interacción repulsiva o cualquier combinación de las fuerzas de atracción anteriores.

Pueden usarse en la presente invención fragmentos de cualquier secuencia nucleotídica de composición nucleotídica conocida. Los expertos en la materia estarán al tanto de la variedad de métodos para producir fragmentos de secuencia nucleotídica incluyendo: cizallamiento mecánico (por ejemplo, mediante sonicación), digestión con una nucleasa (por ejemplo ADNasa I), digestión con enzima o enzimas de restricción, reacción en cadena de la polimerasa usando cebadores degenerados. Preferiblemente, la secuencia nucleotídica deriva de un genoma sustancialmente secuenciado de un microorganismo y/o especie eucariótica compacta. Más preferiblemente, la secuencia nucleotídica deriva de un genoma totalmente secuenciado de un microorganismo y/o especie eucariótica compacta. Lo más preferiblemente, se expresan una pluralidad de diferentes secuencias nucleotídicas en las colecciones de expresión de fragmentos génicos, derivando dichas secuencias de organismos biodiversos. Por tanto, se emplean deseablemente secuencias nucleotídicas biodiversas en el método de la invención para preparar las colecciones de expresión. Cuando se usan genomas secuenciados o fragmentos de los mismos de diferentes organismos en el método, cada uno de los genomas o fragmentos de los mismos debería proporcionarse en igual cantidad molar para asegurar que se incluye en el método una proporción igual de genomas secuenciados o fragmentos de los mismos.

Los expertos en el campo apreciarán que la colección de expresión de fragmentos génicos puede prepararse usando cualquier vector de expresión conocido en la materia. Preferiblemente, los vectores seleccionados para uso en la colección poseen promotores fuertes que potencian así la expresión de la secuencia aminoacídica. Por ejemplo, en un sistema bacteriano, los promotores expresables en bacteria que pueden usarse en el vector pueden incluir, pero sin limitación, pT7-Select, pET, pZero, pHook, pTYB o un derivado de los mismos. Se discuten con más detalle a continuación otros vectores que pueden usarse en el vector.

La secuencia o secuencias aminoacídicas derivadas de la colección de expresión de fragmentos génicos pueden expresarse en forma conformacionalmente restringida o conformacionalmente no restringida. Las secuencias aminoacídicas que se expresan en forma conformacionalmente restringida pueden expresarse en un segundo polipéptido como una proteína de fusión de tal modo que pueden "anidarse" eficazmente en la estructura secundaria del segundo polipéptido. Como alternativa, la secuencia o secuencias aminoacídicas pueden circularizarse mediante la oxidación de residuos de cisteína flanqueantes para limitar la diversidad conformacional. Esto puede ser particularmente beneficioso cuando la secuencia o secuencias aminoacídicas se anidan en un sitio expuesto a la superficie o funcional de una proteína, de tal modo que sean accesibles a la actividad biológica de interés. Por ejemplo, la secuencia o secuencias aminoacídicas pueden expresarse en un bucle polipeptídico de tiorredoxina (Trx). Sin limitarse a teoría o modo de acción alguno, la expresión de la secuencia o secuencias aminoacídicas en una forma conformacionalmente restringida limita los grados de libertad y el coste entrópico asociados a su unión, confiriendo un alto grado de afinidad y especificidad a la interacción.

Los expertos en el campo apreciarán que la presente invención tiene una aplicación de amplio alcance. A modo de ejemplificación, la presente invención es particularmente útil para cribar en colecciones de expresión de fragmentos génicos la secuencia o secuencias aminoacídicas reactivas con anticuerpos particulares, por ejemplo, mediante cromatografía de afinidad de una colección de presentación en fago. Como alternativa, pueden usarse colecciones de fragmentos génicos biodiversos para identificar secuencias antigénicas o inmunogénicas que pueden usarse para vacunas o para inmunoterapia de enfermedades alérgicas o enfermedades autoinmunitarias.

En una realización, se proporciona un método para identificar un modulador o mediador de una actividad biológica, incluyendo dicha actividad antigenicidad o inmunogenicidad, comprendiendo dicho método las etapas de:

- (i) producir una colección de expresión de fragmentos génicos derivada de fragmentos de secuencia nucleotídica definidos, codificando dicha secuencia nucleotídica al menos una secuencia aminoacídica;
- (ii) ensayar en al menos una secuencia aminoacídica derivada de la etapa (i) de la colección de expresión una actividad biológica, en la que la colección esté adaptada para presentar un intervalo de secuencias aminoacídicas cada una de las cuales puede variar en al menos un aminoácido y
- (iii) identificar aquellos aminoácidos esenciales para modular la actividad biológica, siendo dicha actividad diferente de la actividad con que la secuencia no está normalmente asociada en su entorno nativo.

Puede seleccionarse una secuencia aminoacídica que es particularmente eficaz para modular la actividad biológica comparando la actividad biológica observada de una serie de secuencias aminoacídicas diferentes de constitución similar. Usando las diferencias en la actividad biológica observada, es posible identificar aquellos aminoácidos esenciales para la actividad biológica y aquellos que son deseables para la actividad o, en caso alternativo, aquellos que son un impedimento para conseguir una actividad eficaz.

La presente invención tiene una aplicación de amplio alcance para identificar secuencias aminoacídicas que tienen una actividad novedosa en comparación con la que puede reconocerse que tienen en su entorno natural ordinario.

En una forma particularmente preferida de esta realización, se proporciona un método para identificar una secuencia aminoacídica que tiene actividad antigénica o inmunogénica, comprendiendo dicho método las etapas de:

- (i) producir una colección de expresión de fragmentos génicos derivada de fragmentos de secuencia nucleotídica definida, codificando dicha secuencia nucleotídica al menos una secuencia aminoacídica;
- (ii) ensayar en al menos una secuencia aminoacídica derivada de la etapa (i) de la colección de expresión una actividad antigénica o inmunogénica, en la que la colección se adapta para presentar un intervalo de secuencias aminoacídicas cualquiera de las cuales puede variar en al menos un aminoácido;
- (iii) identificar aquellas secuencias aminoacídicas esenciales para modular o mediar la actividad antigénica o inmunogénica y
- (iv) seleccionar aquellas secuencias de la etapa de identificación de la etapa (iii) que no están asociadas a actividad antigénica o inmunogénica en su entorno nativo.

Preferiblemente, las colecciones de fragmentos génicos empleadas en esta realización de la invención se usan para identificar o producir antígenos que pueden usarse para vacunas o inmunoterapia de enfermedades alérgicas o enfermedades autoinmunitarias. En el caso de inmunoterapia de alérgenos, es especialmente deseable identificar péptidos de alta afinidad (que son raros en colecciones de péptidos aleatorios) porque pueden usarse como antígenos monovalentes para evitar reacciones inmunológicas reticulantes tales como la reticulación de IgE en mastocitos.

En una segunda realización, las colecciones de péptidos de la presente invención pueden emplearse para identificar secuencias aminoacídicas antibacterianas novedosas que se liberan condicionalmente a partir de una proteína de fusión. Según esta realización, se proporciona un método de identificación de una secuencia aminoacídica antibacteriana que comprende:

- (i) transformar o transfectar una primera población bacteriana de células con una colección de expresión de péptidos derivada de fragmentos de secuencia nucleotídica definida;
- (ii) cultivar dicha primera población bacteriana durante un tiempo y en condiciones suficientes para que ocurra la expresión de las secuencias aminoacídicas codificadas en dicha colección y para la liberación de las secuencias aminoacídicas a partir de sus fusiones asociadas;
- (iii) poner en contacto las secuencias aminoacídicas expresadas con bacterias patógenas;
- (iv) identificar aquella secuencia o secuencias que son capaces de inhibir el crecimiento de las bacterias patógenas, o destruir las bacterias patógenas y
- (v) seleccionar aquellas secuencias de la etapa de identificación de la etapa (iv) que no están asociadas a la inhibición del crecimiento de bacterias patógenas, o la destrucción de bacterias patógenas, en su entorno nativo.

Debería apreciarse que el método descrito en esta realización tiene una aplicación de amplio alcance para identificar secuencias aminoacídicas novedosas que son capaces de inhibir el crecimiento de bacterias patógenas, o la destrucción de bacterias patógenas.

En una forma altamente preferida de esta realización, se insertan secuencias nucleotídicas que codifican un péptido o péptidos o fusiones peptídicas en el sitio de clonación de un vector de fago T7-Select (Invitrogen) con o sin la introducción de un sitio de escisión proteica condicional (tal como el elemento de corte y empalme de proteína sensible a la temperatura "inteína" modificado del elemento encontrado en el gen VMA1 de *Saccharomyces cerevisiae* (por ejemplo, IMPACT T7 System, New England Biolabs)) clonado en la intersección de fusión del vector. Se cultiva entonces la primera población bacteriana durante un tiempo y en condiciones suficientes para que ocurra la expresión de los péptidos codificados por dicha colección. En los casos en que se desee la escisión condicional del péptido a partir de su contexto de fusión (por ejemplo, el sistema de inteína), la población bacteriana/de fago puede ponerse en condiciones en que pueda ocurrir la escisión (por ejemplo, baja temperatura en el caso de escisión del mutante de inteína)]. Se separan entonces los clones individuales o agrupaciones de clones en dicha colección en series duplicadas. Se lisa entonces al menos una de dichas series duplicadas produciendo una serie de lisado. Obsérvese que esto no es necesario en el caso de vectores de fago líticos tales como T7-Select. Se pone entonces en contacto físico la serie de lisado con bacterias patógenas. Aquellos lisados que son capaces de inhibir el crecimiento de las bacterias patógenas, o de destruir las bacterias patógenas, pueden identificarse entonces mediante técnicas estándar.

Por conveniencia, la bacteria patógena descrita en esta realización puede estar contenida en un césped bacteriano sobre medios sólidos, sin embargo, esto no es esencial para el funcionamiento de esta realización.

Preferiblemente, el método en cuestión comprende adicionalmente la etapa de trazar el lisado de vuelta a la serie duplicada para localizar la célula bacteriana que expresa el mismo péptido antibacteriano que el expresado en dicho lisado. Más preferiblemente, se aísla la secuencia genética que codifica el péptido con fines de producir el péptido antibacteriano codificado en la misma.

En una ejemplificación de esta realización, se ensaya en lisados de BL21 de *Escherichia coli* que contienen proteína expresada a partir de colecciones de péptidos pET su capacidad de inhibir el crecimiento de microorganismos patógenos o, como alternativa, su capacidad de destruir microorganismos patógenos, en la que los clones individuales derivados de una población de células transformadas o transfectadas con la colección de péptidos en cuestión se siembran en placa por duplicado sobre membranas semipermeables, tales como membranas de nitrocelulosa o nailon o, como alternativa, se recolectan por duplicado en cultivos madre y cultivos en los que se va a inducir la expresión de la secuencia peptídica clonada, antes de la lisis. La siembra por duplicado y/o la recolección por duplicado pueden efectuarse manualmente o con ayuda de robots. Las muestras que comprenden aquellas colonias en que se va a inducir la expresión se lisan, por ejemplo, mediante exposición a cloroformo o mediante infección con un bacteriófago tal como el bacteriófago T7, y se recubren sobre un césped recién sembrado de bacterias patógenas.

En el caso de colecciones de fago lítico (tales como aquellas realizadas en el sistema T7-Select), puede usarse una placa Petri de doble cara. En este caso, el recubrimiento de fago ocupa una cara de las placas que está separada de las demás caras por una membrana semipermeable soportada (hecha de un material tal como nitrocelulosa o nailon) sobre la que se encuentra un césped sembrado de bacterias patógenas. Por tanto, la membrana semipermeable separa el recubrimiento de fago de las bacterias patógenas, que pueden cultivarse en diferentes medios respectivamente (véase el ejemplo).

Se ensaya la capacidad de los clones que expresan péptidos individuales de inhibir el crecimiento o destruir la bacteria patógena en cuestión mediante la detección de la presencia de un "claro" o "agujero" de tipo placa en el césped de las bacterias patógenas directamente debajo de la posición en que aparece el lisado que contiene el péptido antibacteriano expresado.

Los expertos en la materia reconocerán que este método proporciona la oportunidad de aislar un clon de fago o plásmido que expresa la actividad que da lugar al correspondiente agujero en el césped opuesto.

En una tercera realización, se proporciona un método para identificar un modificador de una actividad biológica asociada a una célula hospedadora, comprendiendo dicho método las etapas de:

- (i) expresar una molécula indicadora bajo el control operativo de la actividad biológica en la célula, en la que al menos una molécula asociada a la actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor;
- (ii) incubar al menos una célula de la etapa (i) en presencia de una secuencia o secuencias aminoacídicas de una colección de expresión de fragmentos génicos derivados de una secuencia genómica definida, en condiciones que promuevan la interacción entre la secuencia o secuencias aminoacídicas y la secuencia nucleotídica o aminoacídica implicada en la actividad biológica;
- (iii) identificar al menos una secuencia aminoacídica que, en presencia de las células, sea capaz de modificar la expresión de dicha molécula indicadora, o la actividad biológica y
- (iv) seleccionar aquellas secuencias de la etapa (iii) que no sean reconocidas generalmente por ser capaces de modificar la expresión de dicha molécula indicadora, o la actividad biológica, en su entorno nativo.

Preferiblemente, el método se repite las veces que sea necesario para asegurar que sustancialmente todos los aminoácidos codificados por la secuencia nucleotídica definida se presentan ante la actividad biológica.

En una forma particularmente preferida de la tercera realización, la colección de expresión de fragmentos génicos se prepara en un vector pET tal como aquellos que están comercialmente disponibles en Novagen. Los vectores pET como se describen en la presente memoria son particularmente útiles en dichas aplicaciones en virtud de la secuencia promotora T7 fuerte contenida en los mismos, que facilita la expresión bacteriana en cepas que expresan polimerasa de T7. Los expertos en la materia apreciarán que serán igualmente aplicables otros vectores de expresión bacteriana.

En una forma altamente preferida de esta realización, se incorpora una secuencia o secuencias nucleotídicas derivadas de una secuencia genética definida a un vector pET de tal modo que la secuencia nucleotídica esté ligada operativamente con una secuencia de inicio de la traducción bacteriana apropiada como se describe anteriormente. Puede expresarse una segunda secuencia nucleotídica adicionalmente en asociación con la primera secuencia nucleotídica de tal modo que el péptido resultante esté restringido al sitio activo bucle de tiorredoxina o a residuos de cisteína flanqueantes oxidados. Como con otras realizaciones de la invención, la segunda secuencia nucleotídica puede ser sintética y/o derivada de fuentes genómicas.

La expresión del vector pET se consigue mediante la infección de bacterias que contienen el plásmido de colección con el bacteriófago T7 o, como alternativa, usando cepas públicamente disponibles tales como BL21 de *E. coli*, que contiene el gen de polimerasa de T7 bajo el control de *lac*, porque en dichas cepas puede añadirse IPTG a los medios de crecimiento para inducir la expresión del gen de polimerasa de T7. Los derivados de la cepa BL21 (tales

como la cepa BL21trxB (DE3), que contiene una mutación en el gen de tiorredoxina reductasa trxB), son particularmente útiles para asegurar que los puentes disulfuro permanecen oxidados en el citoplasma bacteriano.

Esta realización es particularmente útil para identificar antagonistas de una actividad biológica. En dichas situaciones, la actividad biológica indeseable es preferiblemente funcional en ausencia del fármaco que se está cribando y se ensaya la perturbación de esa interacción en presencia de un compuesto candidato a fármaco, en las que la expresión del gen indicador modificado se detecta de la manera descrita para otras realizaciones de la invención.

Preferiblemente, cuando la molécula indicadora es letal para la célula bacteriana, la expresión de la misma no debería ocurrir hasta que se proporciona el compuesto candidato de secuencia o secuencias aminoacídicas a la célula durante un tiempo y en condiciones suficientes para antagonizar la actividad biológica que conduce a la expresión del indicador. Por consiguiente, en una forma preferida, esta realización proporciona un método de identificación de un antagonista de una actividad biológica en una célula bacteriana que comprende las etapas de:

- (i) disponer la expresión de una molécula indicadora citostática o citotóxica bajo el control operativo de una actividad biológica en dicha célula, en la que al menos un copartícipe de unión en dicha actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor expresable en bacteria;
- (ii) incubar la célula en presencia de al menos un compuesto candidato de secuencia aminoacídica para ensayar su capacidad de antagonizar la actividad biológica durante un tiempo y en condiciones suficientes para que ocurra el antagonismo, en la que el compuesto candidato de secuencia aminoacídica deriva de una colección de expresión de fragmentos génicos derivada de una secuencia genómica definida;
- (iii) expresar el copartícipe de unión bajo el control del promotor expresable en bacteria durante un tiempo y en condiciones suficientes para dar como resultado la expresión de la molécula indicadora en ausencia de antagonismo y
- (iv) seleccionar las células supervivientes o en crecimiento.

Preferiblemente, el promotor expresable en bacteria inducible es el promotor de T7. En dichas circunstancias, la expresión de la molécula indicadora puede inducirse infectando células con el bacteriófago T7, que suministra la función polimerasa de T7. Como alternativa, la célula bacteriana puede ser una célula que contiene la polimerasa de T7 bajo el control de *lac* (por ejemplo, células BL21 de *E. coli*), en cuyo caso el promotor puede inducirse mediante la adición de IPTG al medio de crecimiento. El compuesto candidato puede ser cualquier molécula pequeña, fármaco, antibiótico u otro compuesto, siendo el único requisito que sea capaz de penetrar en o incorporarse activamente a la célula bacteriana o, como alternativa, que se modifique mediante la adición de una molécula portadora para facilitar dicha captación.

En una cuarta realización, se proporciona un método de identificación de un antagonista de una actividad biológica, comprendiendo dicho método las etapas de:

- (i) disponer la expresión de una molécula indicadora bajo el control operativo de una actividad biológica en una célula, en la que al menos un copartícipe de dicha actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor expresable en bacteria en un vector adecuado, en la que (a) la secuencia nucleotídica deriva de una secuencia nucleotídica de origen conocido y secuenciada y (b) la actividad biológica es diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo;
- (ii) incubar la célula en presencia de un compuesto candidato para ensayar su capacidad de antagonizar la actividad biológica y
- (iii) seleccionar células en las que la expresión de dicha molécula indicadora, o actividad biológica, esté modificada.

Este método es particularmente útil para identificar fármacos novedosos tales como antibióticos o agentes inhibidores que pueden servir como candidatos a agonistas y antagonistas de cualquier actividad biológica. Además, este sistema puede usarse en el cribado de alto rendimiento de antibióticos u otros agentes inhibidores novedosos que se orientan a interacciones de secuencia aminoacídica:secuencia de ácido nucleico o interacciones de secuencia aminoacídica:secuencia aminoacídica específicas.

Preferiblemente, cuando la molécula indicadora es letal para la célula bacteriana, la expresión de la misma no debería permitirse hasta que el compuesto candidato se proporcione a la célula durante un tiempo y en condiciones suficientes para antagonizar la actividad biológica que conduce a la expresión del indicador. Por consiguiente, un

aspecto preferido de esta realización proporciona un método de identificación de un antagonista de una actividad biológica en una célula bacteriana, que comprende:

- 5 (i) disponer la expresión de una molécula indicadora citostática o citotóxica bajo el control operativo de una actividad biológica en dicha célula, en la que al menos un copartícipe de unión en dicha actividad biológica comprende una secuencia aminoacídica codificada por una secuencia nucleotídica que está dispuesta en conexión operativa con un promotor expresable en bacteria, en la que (a) la secuencia nucleotídica se define por y deriva de una secuencia nucleotídica de origen conocido y (b) la actividad biológica es diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo;
- 10 (ii) incubar la célula en presencia de un compuesto candidato para ensayar su capacidad de antagonizar la actividad biológica durante un tiempo y en condiciones suficientes para que ocurra el antagonismo;
- (iii) expresar el copartícipe de unión bajo el control del promotor expresable en bacteria durante un tiempo y en condiciones suficiente para dar como resultado la expresión de la molécula indicadora en ausencia de antagonismo y
- 15 (iv) seleccionar las células supervivientes o en crecimiento.

En un ejemplo altamente preferido de esta realización, el promotor expresable en bacteria inducible es el promotor de T7. Un experto en el campo observará que puede usarse en la invención cualquier otro promotor inducible bacteriano. Esta realización se ejemplifica con respecto al promotor solo por conveniencia. En dichas circunstancias, la expresión de la molécula indicadora pueda inducirse infectando células con el bacteriófago T7, que suministra la función de polimerasa de T7. Como alternativa, la célula bacteriana puede ser una célula que contiene la polimerasa de T7 bajo el control de *lac* (por ejemplo, células BL21 de *E. coli*), en cuyo caso el promotor puede inducirse mediante la adición de IPTG al medio de crecimiento. El compuesto candidato puede ser cualquier molécula pequeña, fármaco, antibiótico u otro compuesto, siendo el único requisito que sea capaz de penetrar en o incorporarse activamente en la célula bacteriana o, como alternativa, que se modifique mediante la adición de una molécula portadora para facilitar dicha captación.

Deseablemente, la presente invención emplea una colección de fragmentos génicos preparada a partir de una secuencia genómica definida presente aislada o en combinación con otras secuencias genómicas definidas para identificar la secuencia o secuencias aminoacídicas que pueden ser candidatos adecuados para el diseño racional de fármacos, proporcionando sustancialmente al mismo tiempo datos bioinformáticos integrales sobre estos candidatos. Los datos bioinformáticos derivados del método pueden usarse para identificar aquellos aminoácidos importantes en la modulación de la actividad biológica.

Usando el conocimiento de la relación filogenética entre microorganismos, puede diseñarse una mezcla de genomas particulares para maximizar la diversidad de secuencia en la colección de expresión de péptidos. Este enfoque tiene varias ventajas notables frente a la clonación y expresión de ADN purificado directamente del entorno. En primer lugar, la colección puede acercarse más fácilmente a la diversidad y sesgo reales. Por tanto, pueden ejecutarse medidas para maximizar la diversidad de dominios y para minimizar el sesgo hacia los genomas de especies dominantes. En segundo lugar, el ADN agrupado artificialmente derivado de organismos conocidos distintos permite oportunidades únicas de examinar diversos genomas que pueden no aparecer conjuntamente en la naturaleza. Por ejemplo, los genomas de ciertas arqueobacterias podrían cribarse simultáneamente con aquellos de parásitos obligados tales como micoplasmas y/o diversos organismos grampositivos y/o gramnegativos. En tercer lugar, el alineamiento de secuencias derivadas de un cribado puede usarse para revelar motivos de consenso. Además, pueden excluirse otros motivos relacionados potenciales como candidatos a fármaco potenciales si no se identifican a partir de ninguno de los genomas en que aparecen teóricamente, a pesar de un cribado exhaustivo a una complejidad que predeciblemente cubriría todos los dominios potenciales codificados por el genoma o genomas, pero que no consiguen exhibir la actividad requerida. Esta información puede usarse para diseñar péptidos óptimos que imitan los motivos de consenso identificados en el cribado biológico mientras que carecen de los residuos alternativos de péptidos estructuralmente relacionados que se incluían presuntamente en el cribado exhaustivo. Finalmente, usar los genomas agrupados de organismos secuenciados facilita ciertos análisis bioinformáticos potentes que pueden ser útiles en el diseño de péptidos terapéuticos.

En una quinta realización, se proporciona un método para identificar un modulador de una actividad biológica, comprendiendo dicho método las etapas de:

- (i) producir una colección de expresión de fragmentos génicos derivada de una secuencia genómica definida;
- 55 (ii) poner en contacto una secuencia aminoacídica derivada de la colección de expresión con una molécula indicadora que está bajo el control operativo de una actividad biológica asociada a un hospedador y

- (iii) identificar una secuencia aminoacídica capaz de modular la actividad biológica, en el que la actividad es diferente de cualquier actividad que pueda tener la secuencia aminoacídica en su entorno nativo.

Preferiblemente, al menos uno de los copartícipes de la actividad biológica contemplados por esta realización es un péptido, polipéptido, proteína o molécula enzimática o un derivado de los mismos. Según esta realización, el copartípe o copartícipes restantes son una molécula seleccionada de la lista que comprende ácidos nucleicos tales como ARN o ADN monocatenario o bicatenario, un péptido, polipéptido, proteína, enzima, carbohidrato, aminoácido, nucleótido, nucleósido, lípido, lipoproteína, vitamina, coenzima, molécula receptora, hormona, compuesto químico, AMP cíclico, ión metálico o molécula de segundo mensajero, entre otros. Más preferiblemente, la actividad biológica es una interacción proteína:proteína o una interacción proteína:péptido o una interacción proteína:polipéptido.

- 10 En una forma particularmente preferida, la actividad biológica es entre un primer copartípe que comprende una secuencia aminoacídica y un segundo copartípe que comprende una molécula de ácido nucleico tal como ADN o ARN o, como alternativa, una secuencia aminoacídica o un derivado o análogo de la misma.

Según una sexta realización, se proporciona un método para identificar una secuencia aminoacídica que es capaz de modular la actividad biológica en una célula hospedadora, comprendiendo dicho método las etapas de:

- 15 (i) producir una colección en un hospedador en la que (a) las células transformadas de dicha colección contienen al menos una primera secuencia nucleotídica que comprende o codifica una molécula indicadora cuya expresión está bajo el control operativo de dicha actividad biológica, y una segunda secuencia nucleotídica derivada de una secuencia genómica conocida que puede codificar la secuencia aminoacídica cuando se dispone bajo el control operativo de una secuencia promotora adecuada, y en el que (b) sustancialmente toda la secuencia genómica conocida está presente en la población de células transformadas que constituye dicha colección y la actividad biológica es diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo;
- 20 (ii) cultivar dicho hospedador celular durante un tiempo y/o en condiciones suficientes para que ocurra la expresión de dicha segunda secuencia nucleotídica; y
- 25 (iii) seleccionar o cribar células en las que la expresión de dicha molécula indicadora esté modificada.

La segunda secuencia nucleotídica usada en el método puede derivar de cualquier secuencia genómica conocida. Al usar un número suficiente de segundas especies nucleotídicas para asegurar que se ensaya la secuencia completa de la secuencia genómica conocida, pueden reunirse datos bioinformáticos a partir no solo de secuencias que daban un resultado positivo en el sistema de ensayo, sino también de aquellas secuencias que no conseguían reaccionar.

- 30 Al comparar las secuencias aminoacídicas reactivas frente a secuencias similares en un genoma que causaban una reacción o, como alternativa, que no conseguían causar una reacción, pueden identificarse los motivos de secuencia así como aminoácidos individuales que pueden estar implicados en una actividad biológica. Además, si el cribado es suficientemente integral para asegurar una cobertura adecuada, puede mostrarse que ciertos residuos/motivos alternativos representados en la colección son subóptimos si se incorporan al diseño de inhibidores de la actividad.

- 35 Por tanto, en una forma preferida, esta realización proporciona un método de identificación de una secuencia o secuencias aminoacídicas que son capaces de modular una actividad biológica en un hospedador, comprendiendo dicho método las etapas de:

- 40 (i) producir una colección peptídica en un hospedadora en la que (a) las células transformadas de dicha colección contienen al menos una primera secuencia nucleotídica que comprende o codifica una molécula indicadora cuya expresión está bajo el control operativo de dicha actividad biológica, y una segunda secuencia nucleotídica derivada de una secuencia genómica conocida que puede codificar dicha secuencia o secuencias aminoacídicas cuando se dispone bajo el control operativo de una secuencia promotora adecuada, y en la que (b) sustancialmente toda la secuencia genómica conocida está presente en la población de células transformadas que constituye dicha colección;
- 45 (ii) cultivar dicho hospedador celular durante un tiempo y/o en condiciones suficientes para que ocurra la expresión de dicha segunda secuencia nucleotídica;
- (iii) seleccionar o cribar células en la que la expresión de dicha molécula indicadora está modificada;
- (iv) comparar el intervalo de secuencias aminoacídicas que pueden derivar de la secuencia genómica conocida frente a aquellas secuencias que modulaban la actividad biológica y
- 50 (v) determinar aquellos aminoácidos que son esenciales para modificar la actividad de la molécula indicadora.

En otra realización, la presente invención proporciona por lo tanto un vector capaz de expresar una secuencia nucleotídica en cada uno de sus posibles marcos de lectura y en el que cada una de dichas secuencias

aminoacídicas así producidas se expresa como una fusión con una segunda secuencia aminoacídica en la que pueden estar conformacionalmente restringidas, comprendiendo dicho vector al menos:

(i) un primer módulo de expresión que comprende:

5 (a) un sitio de clonación múltiple para la inserción de una primera secuencia nucleotídica que codifica dicha primera secuencia aminoacídica, en el que dicho sitio de clonación múltiple puede estar adyacente a una o más segundas secuencias nucleotídicas que codifican un bucle polipeptídico de tal modo que pueda producirse un polipéptido de fusión entre dichas primera y segunda secuencias aminoacídicas;

10 (b) una secuencia terminadora adyacente al sitio de clonación múltiple y distal de dicha secuencia promotora y segundas secuencias nucleotídicas;

(ii) un medio para expresar la primera secuencia nucleotídica en cada uno de sus marcos de lectura;

(iii) un origen de replicación bacteriano y/o un origen de replicación de bacteriófago y

(iv) un segundo módulo de expresión que codifica un gen marcador de selección bacteriano.

15 En una realización alternativa, el vector de expresión de la invención comprende adicionalmente un segundo módulo de expresión que comprende un gen marcador seleccionable ligado operativamente con dos o más secuencias promotoras y situado en dirección 5' de una secuencia terminadora, en el que una de dichas secuencias promotoras es un promotor expresable en bacteria y en el que una de dichas secuencias promotoras es un promotor expresable en levadura.

20 En otra realización alternativa, el vector en cuestión se modifica adicionalmente para proporcionar expresión extracelular inducible mediante fusiones de péptido señal y/o sistemas de lisis condicional. La lisis condicional puede conseguirse mediante la expresión de un gen lítico inducible en células bacterianas, introduciendo dichas secuencias en un módulo de expresión entre un promotor bacteriano inducible (tal como los promotores *lac*, *tac* o el más estrechamente regulado *araBAD*) y una secuencia de terminación transcripcional, en serie con las secuencias promotora y terminadora ya presentes en los módulos de expresión en cuestión.

25 En una realización adicional más, la lisis condicional de bacterias que expresan dicho péptido/polipéptido está causada por medios alternativos tales como mediante infección con un bacteriófago adecuado o mediante exposición a agentes químicos apropiados tales como cloroformo y/o SDS. En una forma particularmente preferida de la invención, el vector incluye también un tercer módulo de expresión que permite la expresión condicional de un gen lítico (tal como aquellos genes producidos por bacteriófagos).

30 La presente invención contempla también la secuencia o secuencias aminoacídicas identificadas mediante el método de la presente invención, así como el uso de esas moléculas en una composición farmacéutica. La composición farmacéutica comprende una secuencia o secuencias aminoacídicas capaces de modular una actividad biológica o la función de una molécula biológica y un portador y/o diluyente farmacéuticamente aceptable.

Fragmentos de secuencia nucleotídica biodiversos

35 Cuando se usan genomas secuenciados de diferentes organismos en las realizaciones anteriores, cada uno de los genomas debería proporcionarse en cantidades molares iguales para asegurar que se incluye una proporción igual de genomas secuenciados en el método. Debido a que los genomas son de un tamaño conocido, pueden aplicarse métodos de normalización estándares para asegurar que la concentración del genoma de un organismo no es proporcionalmente mayor que la del genoma de otro organismo. Dichos métodos para igualar las concentraciones genómicas son bien conocidos por los expertos en la materia e incluyen, a modo de ejemplo, la contribución de proporcionalmente más ADN al agrupamiento de los genomas que son mayores, para compensar la tendencia de que los fragmentos de dichos genomas estén subrepresentados si se combina una masa igual de ADN de cada genoma. Además, se contempla por la presente invención la normalización mediante otros medios conocidos por los expertos en la materia, tales como los dados a conocer en la patente de EE.UU. 5763239.

45 La presente invención intenta acelerar el proceso evolutivo combinando artificialmente dominios de diferentes genomas que habría sido improbable que coevolucionaran. Preferiblemente, las colecciones de expresión genómica se preparan a partir de organismos evolucionados diversamente. Por ejemplo, los organismos podrían derivar de: genomas eucarióticos compactos tales como de *Fugu rubripes*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*; y/o de microorganismos procarióticos que se han caracterizado genéticamente, tales como *E. coli*, *Aquifex aeliticus*,
50 *Methanococcus jannaschii*, *Bacillus subtilis*, *Haemophilus influenzae*, *Helicobacter pylori*, *Neisseria meningitidis*, *Synechocystis* sp, *Bordetella pertussis*, *Pasteurella multocida*, *Pseudomonas aeruginosa*, *Borrelia burgdorferi*, *Methanobacterium thermoautotrophicum*, *Mycoplasma pneumoniae*, *Archaeoglobus fulgidus* y *Vibrio harveyi*). Los expertos en la materia están al tanto de que el número de genomas secuenciados está aumentando rápidamente (pueden obtenerse fácilmente recopilaciones de genomas secuenciados por referencia al World Wide Web (por ejemplo, véanse las siguientes URL para detalles:
55

<http://www.tigr.org/tdb/mdb/html>

<http://www.sanger.ac.uk>

http://www.genome.ad.jp/kegg/java/org_proj.html

<http://www-fp.mcs.anl.gov/~gaasterland/genomes.html>

5 <http://www.ncgr.org/>

<http://www.cbs.dtu.dk/databases/DOGS/index.html>

<http://geta.life.uiuc.edu/~nikos/genomes.html>

y de que los métodos aquí descritos son aplicables a cualquier subconjunto del agrupamiento completo de genomas secuenciados.

10 La secuencia nucleotídica definida a partir de la que deriva la secuencia nucleotídica conocida no está limitada solo a aquellas secuencias que codifican aminoácidos en proteínas derivadas naturales, sino que incluye también secuencias nucleotídicas no codificantes. Por tanto, debería entenderse que la segunda secuencia nucleotídica puede derivar de una 5' UTR, un intrón (cuando sea aplicable), una 3' UTR o marcos de lectura/orientaciones alternativas del fragmento clonado.

15 La diversidad en un agrupamiento de secuencias nucleotídicas secuenciadas puede aumentarse también sometiendo las secuencias a métodos que malinterpreten o muten esos fragmentos. Por tanto, en una realización de la invención, los métodos descritos anteriormente pueden incluir también una etapa de mutación artificial de las colecciones de dominios. Dichos métodos son bien conocidos en la materia.

20 Un modo de conseguir este fin implicaría la mutación de la secuencia genómica conocida antes de la inserción en un vector de expresión. Por tanto, en una forma preferida, el método de la invención podría incluir la etapa de: someter la secuencia nucleotídica conocida a mutagénesis antes de la inserción en el vector de expresión. Esto puede conseguirse, por ejemplo, amplificando los genomas secuenciados usando procedimientos de PCR mutagénicos tales como aquellos que incluyen la etapa de efectuar la reacción de PCR en presencia de manganeso. Se ha calculado que, con una tasa de error de 0,5 bases por 100 pb/ciclo, ocho ciclos mutagénicos producirán cambios de base en un 90% de los productos de PCR y casi un 50% tendrán 2 o 3 sustituciones.

25 Otro modo en que las colecciones de dominios podrían mutarse sería mediante la expresión de fragmentos nucleotídicos en células que se modifican para mutar la información de secuencia. Dichas cepas son deficientes en ciertas enzimas, haciendo su tasa de mutación de aproximadamente 5.000 a 10.000 veces mayor que en la original de tipo silvestre. Por tanto, el método puede incluir la etapa de: expresar los fragmentos génicos biodiversos en una o más estirpes celulares que son deficientes en al menos una enzima de reparación de ADN. Por ejemplo, una vez construida, la colección de plásmido puede amplificarse en cepas bacterianas deficientes en reparación de errores de emparejamiento (por ejemplo, cepas que contienen la mutación *mutS*, *mutD* y/o *mutT*), dando como resultado la generación de mutaciones. En una ejemplificación de esta realización, se amplifican o propagan colecciones de péptidos derivadas de la expresión de ADN genómico en cepas bacterianas que son deficientes en la unidad epsilon (ε) de la ADN polimerasa III (concretamente, los alelos *dnaQ* y *mutD*) y/o son defectivas de reparación de errores de emparejamiento. Las cepas mutadoras de *Escherichia coli* que poseen los alelos *mutY* y/o *mutM* y/o *mutD* y/o *mutT* y/o *mutA* y/o *mutC* y/o *mutS* son particularmente útiles para dichas aplicaciones. Las cepas bacterianas que portan dichas mutaciones están fácilmente disponibles para los expertos en la materia.

30 Cuando se mutan los fragmentos antes de la generación de una colección de expresión, tanto los fragmentos mutados como no mutados deberían combinarse preferiblemente en la misma preparación y se expresan preferiblemente usando vectores descritos en la presente memoria. Las colecciones mutadas y no mutadas experimentarán los mismos procedimientos de selección. Debería compararse entonces y examinarse la especificidad y actividad biológica de los péptidos.

35 Para potenciar la diversidad en la colección de péptidos genómicos secuenciados, los genomas secuenciados fragmentados pueden expresarse también en cada uno de sus diferentes marcos de lectura. La expresión de dichas secuencias de esta manera puede conseguirse mediante cualquier método conocido en la materia incluyendo, por ejemplo, mediante ligamiento de los fragmentos con adaptadores y/o ligadores en los tres marcos de lectura diferentes o disponiendo los fragmentos bajo el control del sitio o sitios internos de entrada a ribosoma (IRES) y/o secuencias que confieren deslizamiento transcripcional/traducciona. Si se usan adaptadores, un único vector puede contener cada uno de los diferentes adaptadores o cada adaptador puede proporcionarse a un vector diferente.

40 Los fragmentos pueden expresarse también en marcos de lectura inversos. Por tanto, al permitir la expresión de una secuencia génica en cada marco de lectura posible, para cualquier secuencia peptídica particular habría seis posibles combinaciones diferentes.

- La presencia de clones en todos los marcos de lectura permite el cribado simultáneo de péptidos aleatorios expresados en marcos de lectura que no aparecen en la naturaleza, junto con una variedad de dominios peptídicos naturales clonados en el marco de lectura apropiado. Esto permite una comparación del éxito relativo de los inhibidores del aislamiento a partir de colecciones de péptidos naturales en contraposición con colecciones de péptidos aleatorios. Los métodos de cribado descritos en la presente memoria son también aplicables al cribado de colecciones de péptidos aleatorios restringidos o no restringidos derivadas de fuentes no biológicas artificiales.

Definiciones

- Como se usa aquí, la frase “no asociada normalmente en su entorno nativo” hará referencia a una actividad con la que la secuencia aminoacídica no está asociada típicamente. Adicionalmente, como se usa en la presente memoria, “entorno nativo” se entenderá que hace referencia al entorno biológico en que se encuentra típicamente la secuencia aminoacídica en la naturaleza.
- Como se usa en la presente memoria, el término “dominio” se entenderá que significa una unidad funcional de una secuencia o secuencias aminoacídicas que posee actividad en el aislamiento o en un contexto artificial y que no necesariamente implica rasgos estructurales.
- Como se usa en la presente memoria, “secuencia aminoacídica” incluirá péptidos, oligopéptidos y polipéptidos, incluyendo derivados y análogos de los mismos que comprenden un número de residuos en el intervalo de 1 a 500.
- Como se usa en la presente memoria, el término “aptámero” se entenderá que incluye los péptidos altamente específicos y conformacionalmente restringidos normalmente relacionados con la clase descrita por Brent y colaboradores (3).
- Como se usa en la presente memoria, el término “actividad” se entenderá que incluye cualquier actividad enzimática, cambio estructural o conformacional que ocurra fuera o dentro de la célula.
- Como se usa en la presente memoria, el término “colección de expresión de fragmentos génicos” se entenderá que incluye cualquier colección de expresión preparada usando insertos derivados de fragmentos genómicos o productos de PCR de una serie de distintos genomas procarióticos y/o genomas eucarióticos compactos.
- Como se usa en la presente memoria, el término “derivado” se entenderá que designa mutantes, partes o fragmentos de un polipéptido completo como se define en la presente memoria que son funcionalmente equivalentes. Los derivados incluyen péptidos modificados en los que los ligandos se enlazan con uno o más residuos aminoacídicos contenidos en los mismos, tales como grupos funcionales, carbohidratos, enzimas, proteínas, polipéptidos o moléculas indicadoras tales como radionucleidos o compuestos fluorescentes. Las formas glucosilada, fluorescente, acilada o alquilada de los péptidos en cuestión están también contempladas por la presente invención. Los procedimientos para derivatizar proteínas y péptidos son bien conocidos en la materia.
- “Análogos” de un péptido, proteína, polipéptido o enzima son moléculas funcionalmente equivalentes que comprenden uno o más análogos aminoacídicos de origen no natural conocidos por los expertos en la materia.
- Los términos “hospedador” y “hospedador celular” o término similar hacen referencia a células procarióticas y eucarióticas capaces de soportar la expresión de una molécula indicadora bajo el control de una actividad biológica, independientemente de si la actividad biológica o la molécula indicadora son endógenas de la célula.
- Los expertos en la materia estarán al tanto de que una “célula transformada” es una célula en la que se ha introducido un ácido nucleico exógeno, en la que el ácido nucleico exógeno se integra en el genoma de la célula hospedadora o, como alternativa, se mantiene en la misma como un elemento genético extracromosómico tal como un plásmido, episoma o cromosoma artificial, entre otros.
- La célula transformada puede ser cualquier célula capaz de soportar la expresión de ADN exógeno, tal como una célula bacteriana, célula de insecto, célula de levadura, célula de mamífero o célula de planta. Con particular preferencia, la célula es una célula bacteriana, una célula de mamífero o una célula de levadura. Con particular preferencia, la célula es una célula de levadura.
- El término “expresión” hace referencia a al menos la transcripción de una secuencia nucleotídica para producir una molécula de ARN. El término “expresión” puede hacer referencia también a la transcripción y traducción combinadas de una secuencia nucleotídica para producir un péptido, polipéptido, proteína o molécula enzimática o, como alternativa, al proceso de traducción del ARNm para producir un péptido, polipéptido, proteína o molécula enzimática.
- Se entiende por “bajo control operativo” que una primera entidad indicada está regulada o controlada por una segunda entidad indicada.
- En el presente contexto, cuando la expresión de la molécula indicadora está bajo el control operativo de una actividad biológica, dicha expresión se modifica (concretamente, se potencia, induce, activa, reduce o reprime) cuando se expresa un péptido, oligopéptido o polipéptido capaz de potenciar, inducir, activar, reducir o reprimir la

formación de dicha actividad biológica. Por consiguiente, habitualmente no es suficiente que solo esté presente un copartícipe en la actividad biológica para que ocurra dicha expresión modificada de la molécula indicadora, sin embargo, puede haber cierta expresión de la molécula indicadora en presencia de solo un copartícipe.

Como se usa en la presente memoria, el término “colección de péptidos” es un conjunto de diversas secuencias nucleotídicas que codifican un conjunto de secuencias aminoacídicas, en el que dichas secuencias nucleotídicas están preferiblemente contenidas en una molécula de vector de plásmido, cósmido, bacteriófago o virus adecuada que es adecuada para el mantenimiento y/o replicación en un hospedador celular. El término “colección de péptidos” engloba adicionalmente secuencias aminoacídicas aleatorias derivadas de una secuencia genómica conocida, en el que las secuencias aminoacídicas están codificadas por una segunda secuencia nucleotídica obtenida, por ejemplo, mediante cizallamiento o digestión parcial de ADN genómico usando endonucleasas de restricción o nucleasas tales como ADNasa I, entre otros enfoques.

Las colecciones de péptidos preferidas según esta realización de la invención son “colecciones representativas” que comprenden un conjunto de secuencias aminoacídicas o secuencias nucleotídicas que codifican las mismas, que incluye virtualmente todas las posibles combinaciones de secuencias aminoacídicas o nucleotídicas para una longitud definida y especificada anteriormente de molécula peptídica o de ácido nucleico, respectivamente.

Una colección de péptidos puede comprender partículas víricas o partículas bacteriófagas que comprenden un conjunto diverso de secuencias nucleotídicas que codifican un conjunto diverso de secuencias aminoacídicas, en la que los miembros de dicho conjunto diverso de secuencias nucleotídicas se disponen bajo el control operativo de una secuencia promotora que es capaz de dirigir la expresión de dicha secuencia nucleotídica en dicha célula, partícula vírica o partícula bacteriófaga.

Por consiguiente, la secuencia aminoacídica codificada por la segunda secuencia nucleotídica puede comprender cualquier secuencia de aminoácidos de al menos aproximadamente 1 a 100 aminoácidos de longitud, y preferiblemente de 1 a 60 aminoácidos de longitud, y puede derivar de la expresión de secuencias nucleotídicas conocidas que se preparan mediante uno cualquiera de una variedad de métodos tales como, por ejemplo, generación sintética aleatoria. Más preferiblemente, la unidad peptídica es un péptido de 6 a 20 aminoácidos. El uso de fragmentos nucleotídicos mayores, empleando particularmente ácido nucleico cizallado aleatoriamente derivado de genomas bacteriano, de levadura o animal, no está excluido.

Como alternativa o además, la secuencia aminoacídica puede expresarse como una proteína de fusión con un motivo orientador nuclear capaz de facilitar la orientación de dicho péptido al núcleo de dicha célula hospedadora donde ocurre la transcripción, en particular la señal de localización nuclear SV40 que es funcional en células de levadura y mamíferos.

Como alternativa o además, la secuencia aminoacídica puede expresarse como una proteína de fusión con una secuencia peptídica capaz de potenciar, aumentar o ayudar a la penetración o captación del péptido por una célula aislada, tal como cuando la secuencia aminoacídica en cuestión se sintetiza *ex vivo* y se añade a células aisladas en cultivo. En una realización particularmente preferida, la secuencia peptídica capaz de potenciar, aumentar o ayudar a la penetración o captación es funcional en células eucarióticas superiores; por ejemplo, la secuencia orientadora penetratina de *Drosophila*. Según esta realización, la proteína de fusión comprende al menos la secuencia aminoacídica:

CysArgGlnIleLysIleTrpPheGlnAsnArgArgMetLysTrpLysLys(Xaa)_nCys

o un homólogo, derivado o análogo de la misma, en la que Xaa es cualquier residuo aminoacídico y n tiene un valor mayor o igual a 1. Preferiblemente, el valor de n será de al menos 5, más preferiblemente de entre 5 y aproximadamente 20, todavía más preferiblemente de entre aproximadamente 15 y aproximadamente 35 y todavía aún más preferiblemente de entre aproximadamente 30 y aproximadamente 50 y aún más preferiblemente de entre aproximadamente 35 y aproximadamente 55. En una realización aún más preferida, el valor de n está entre al menos aproximadamente 40 y al menos aproximadamente 60.

La referencia en la presente memoria a un “promotor” ha de tomarse en su contexto más amplio e incluye las secuencias reguladoras transcripcionales de un gen genómico clásico, incluyendo la secuencia TATA que se requiere para un inicio de transcripción exacto en células eucarióticas, con o sin una secuencia CCAAT y elementos reguladores adicionales (concretamente, secuencias activadoras, potenciadoras y silenciadoras en dirección 5'). Los promotores pueden carecer también de motivo de secuencia TATA, sin embargo comprenden uno o más “elementos iniciadores” o, como en el caso de secuencias promotoras derivadas de levadura, comprenden una o más “secuencias activadoras en dirección 5” o elementos “UAS”. Para expresión en células procarióticas tales como bacterias, el promotor debería contener al menos las secuencias de -35 y -10.

Un promotor está habitualmente situado en dirección 5' de un gen estructural cuya expresión regula. Además, los elementos reguladores que comprenden un promotor están habitualmente situados a 2 kb del sitio de inicio de transcripción del gen.

En el presente contexto, el término “promotor” se usa también para describir una molécula sintética o de fusión, o un derivado, que confiere, activa o potencia la expresión de la molécula indicadora en cuestión en una célula. Los promotores preferidos pueden contener copias adicionales de uno o más elementos reguladores específicos para potenciar adicionalmente la expresión del gen y/o alterar la expresión espacial y/o expresión temporal. Por ejemplo, en levadura los elementos reguladores que confieren inducibilidad por galactosa, fosfato o cobre pueden disponerse adyacentes a una secuencia promotora heteróloga que activa la expresión del indicador, confiriendo así inducibilidad condicional a la expresión de dicho gen mediante la adición del inductor apropiado al medio de crecimiento.

Disponer un gen bajo el control operativo de una secuencia promotora significa situar dicho gen de tal modo que su expresión esté controlada por la secuencia promotora. Los promotores están generalmente situados en 5' a los genes que controlan. En la construcción de combinaciones de promotor heterólogo/gen estructural, se prefiere generalmente situar el promotor a una distancia del sitio de inicio de la transcripción génica que sea aproximadamente la misma que la distancia entre ese promotor y el gen que controla en su ambiente natural, concretamente, el gen del que deriva el promotor. Como es conocido en la materia, puede ajustarse cierta variación de esta distancia sin pérdida de la función promotora. De forma similar, la situación preferida de un elemento de secuencia reguladora con respecto a un gen heterólogo para situar bajo su control se define por la posición del elemento en su ambiente natural, concretamente, los genes de los que deriva. De nuevo, como es conocido en la materia, puede ocurrir también cierta variación de esta distancia.

Los ejemplos de promotores adecuados para uso en la regulación de la expresión de la molécula indicadora y/o la secuencia aminoacídica y/o el copartícipe de unión a polipéptido en una célula incluyen promotores derivados de virus, hongos, levaduras, insectos, animales y plantas. Los promotores preferidos son capaces de conferir la expresión a una célula eucariótica, especialmente una célula de levadura o mamífero. El promotor puede regular la expresión de un gen constitutiva o diferencialmente con respecto al tejido en que ocurre la expresión, o con respecto a la etapa de desarrollo en que ocurre la expresión, o en respuesta a estímulos externos tales como estrés ambiental u hormonas, entre otros.

Los promotores particularmente preferidos incluyen aquellos promotores de origen natural y sintéticos que contienen sitios de unión para factores de transcripción, más preferiblemente factores de transcripción de hélice-bucle-hélice (HLH), proteínas de dedo de cinc, proteínas de cremallera de leucina y similares. Los promotores preferidos pueden ser también secuencias sintéticas que comprenden una o más secuencias operadoras en dirección 5' tales como secuencias operadoras *LexA* o secuencias activadoras derivadas de cualquiera de los promotores a que se hace referencia en la presente memoria tales como sitios de unión a ADN *GAL4*.

Los expertos en la materia reconocerán que la elección del promotor dependerá de la naturaleza de la célula que se esté transformando y de la molécula para expresar. Dichas personas serán fácilmente capaces de determinar las combinaciones funcionales de secuencias promotoras y operadoras mínimas para los tipos celulares en que se efectúe el método de la invención.

Los promotores particularmente preferidos incluyen un promotor de levadura, un promotor de mamífero, una secuencia promotora bacteriana o bacteriófaga seleccionada de la lista que comprende las secuencias promotoras *GAL1*, *CUP1*, *PGK1*, *ADH2*, *PH05*, *PRB1*, *GUT1*, *SP013*, *ADH1*, *CMV*, *SV401 T7*, *SP6*, *lac* o *tac*.

Aunque la invención se efectúa preferiblemente en células de levadura, los inventores contemplan claramente modificaciones en las que la invención se efectúa enteramente en células de mamífero, utilizando promotores que son operativos en células de mamífero para activar la expresión de los diversos componentes de ensayo, en combinación con un gen indicador contraselectivo operativo en células de mamífero. Dichas realizaciones están dentro del entendimiento de los expertos en la materia.

Para expresión en células de mamífero, se prefiere que el promotor sea la secuencia promotora de CMV, más preferiblemente el promotor CMV-IE o, como alternativa, el promotor SV40 y, en particular, la secuencia promotora tardía de SV40. Estas y otras secuencias promotoras adecuadas para la expresión de genes en células de mamífero son bien conocidas en la materia.

Los ejemplos de células de mamífero contempladas en la presente memoria por ser adecuadas para expresión incluyen las estirpes celulares COS, VERO, HeLa, C 127 de ratón, ovario de hámster chino (CHO), WI-38, riñón de cría de hámster (BHK) o MDCK, entre otras. Una amplia variedad de estirpes celulares tales como estas están fácilmente disponibles para los expertos en la materia.

El prerrequisito para producir polipéptidos intactos en células bacterianas y, en particular, en células de *Escherichia coli*, es el uso de un promotor fuerte con un sitio de unión a ribosoma eficaz, tal como una secuencia de Shine-Dalgarno, que pueda incorporarse a vectores de expresión que portan la primera y segunda secuencias nucleotídicas, u otros constructos genéticos usados en la práctica de los diversos métodos alternativos descritos en la presente memoria. Los promotores típicos para expresión en células bacterianas tales como *E. coli* incluyen, pero sin limitación, el promotor *lacZ*, los promotores λ_L o λ_K sensibles a la temperatura, el promotor SP6, T3 o T7 o promotores compuestos tales como el promotor *tac* inducible por IPTG. Son bien conocidos en la materia una serie de otros sistemas de vectores para expresión moléculas de ácido nucleico en *E. coli*, y se describen, por ejemplo, en

Ausubel *et al.* (1987) y/o Sambrook *et al.* (1989). Están también públicamente disponibles numerosas fuentes de secuencias genéticas adecuadas para expresión en bacterias en diversos constructos de plásmido, tales como, por ejemplo: pKC30 (λ_L), pKK173-3 (tac), pET-3 (T7) o la serie pQE de vectores de expresión, entre otros.

5 Las células procarióticas adecuadas para expresión incluyen *Staphylococcus*, *Corynebacterium*, *Salmonella*, *Escherichia coli*, *Bacillus sp.* y *Pseudomonas sp.*, entre otras. Las cepas bacterianas que son adecuadas con los fines actuales son bien conocidas en la materia relevante.

Cuando se pretende que el promotor regule la expresión de la molécula indicadora, se prefiere particularmente que dicho promotor incluya una o más secuencias de reconocimiento para la unión de un dominio de unión a ADN derivado de un factor de transcripción, por ejemplo, un sitio de unión a GAL4 o secuencia operadora LexA.

10 Como se usa en la presente memoria, el término “molécula indicadora” se entenderá que hace referencia a cualquier molécula que sea capaz de producir un resultado identificable o detectable.

Una molécula indicadora puede ser una enzima, péptido, oligopéptido o polipéptido que comprende un producto visible o, al menos, cuando se incuba en presencia de una molécula de sustrato, puede convertir dicho sustrato en un producto visible, de tal modo que las células que expresan la molécula indicadora puedan detectarse fácilmente.

15 Por ejemplo, la expresión de genes indicadores que codifican polipéptidos que fluorescen por sí mismos o causan la fluorescencia de una segunda molécula puede conectarse operativamente con la actividad biológica que se esté ensayando, para facilitar la detección de células en las que esté presente o ausente la expresión de la molécula indicadora. Dichas aplicaciones son particularmente útiles en enfoques de cribado de fármacos de alto rendimiento, en los que es deseable cribar rápidamente en un gran número de candidatos a fármaco sus propiedades agonistas/antagonistas con respecto a la actividad biológica en cuestión. Las moléculas indicadoras preferidas según esta realización incluyen, pero sin limitación, la enzima β -galactosidasa de *Escherichia coli*, la proteína luciferasa de luciérnaga y la proteína fluorescente verde o mutantes de la misma que poseen espectros de emisión desplazados al rojo o desplazados al azul o un resultado potenciado. Los expertos en la materia estarán al tanto de cómo utilizar secuencias genéticas que codifican dichas moléculas indicadoras al efectuar la invención descrita en la presente memoria, sin experimentación indebida. Por ejemplo, la secuencia de codificación del gen que codifica dicha molécula indicadora puede modificarse para uso en la estirpe celular de interés (por ejemplo, células humanas o células de levadura) de acuerdo con las preferencias de uso de codón conocidas. Adicionalmente, la eficacia de traducción del ARNm derivado de fuentes no eucarióticas puede mejorarse mutando la correspondiente secuencia génica o introduciendo de otro modo en dicha secuencia génica un sitio de inicio de la traducción de consenso de Kozak.

Preferiblemente, la molécula indicadora permite la identificación colorimétrica de su expresión mediante fluorescencia directa (por ejemplo, proteína fluorescente verde) o mediante un cambio de color en presencia de un sustrato apropiado (por ejemplo, la producción de un color azul con β -galactosidasa en presencia del sustrato 5-bromo-4-cloro-3-indoil- β -D-galactósido (concretamente, X-GAL).

35 Las moléculas indicadoras particularmente preferidas son aquellas que producen un crecimiento o viabilidad celulares alterados, incluyendo la capacidad de inducir la muerte celular. En el presente contexto, la molécula indicadora comprende la primera molécula de ácido nucleico o está codificada por dicha primera molécula de ácido nucleico. Por consiguiente, los expertos en la materia estarán al tanto de que la molécula indicadora de dicha realización es preferiblemente un péptido, polipéptido, enzima, abzima u otra molécula proteica o, como alternativa, una molécula de ácido nucleico aislada.

Preferiblemente, la molécula indicadora puede inhibir, potenciar o modular de otro modo directa o indirectamente el crecimiento y/o viabilidad de la célula hospedadora. La modulación directa del crecimiento y/o viabilidad celulares es cuando la expresión de la molécula indicadora tiene una consecuencia directa sobre el crecimiento y/o viabilidad celulares. La modulación indirecta del crecimiento y/o viabilidad celulares es cuando la expresión de la molécula indicadora no tiene consecuencias directas sobre el crecimiento y/o viabilidad celulares, sin embargo, dicha expresión puede modular el crecimiento y/o viabilidad celulares cuando las células se cultivan en presencia de un cofactor o molécula de sustrato adecuado, entre otros.

50 Cuando la molécula indicadora es un péptido, polipéptido, enzima, abzima u otra molécula proteica que comprende un compuesto citostático, compuesto antimitótico, toxina, mitógeno o sustancia reguladora del crecimiento tal como una hormona o proteína que es esencial para el crecimiento o viabilidad celulares, puede tener un efecto directo sobre el crecimiento o viabilidad celulares cuando se expresa en las mismas. De forma similar, una molécula indicadora que comprende una molécula de ácido nucleico puede tener un efecto directo sobre el crecimiento y/o viabilidad celulares, por ejemplo, en el que la molécula indicadora es una ribozima, molécula anticodificante, minizima o molécula cosupresora que está orientada a la expresión de un gen que puede modificar el crecimiento y/o viabilidad celulares.

Cuando es deseable para la molécula indicadora tener un efecto indirecto sobre el crecimiento y/o viabilidad celulares, esto puede conseguirse, por ejemplo, acoplando la expresión de la molécula indicadora con la producción de un compuesto citostático, compuesto antimitótico, toxina o molécula reguladora negativa del crecimiento.

En una realización, la molécula indicadora es una enzima que, cuando se expresa en la célula hospedadora, cataliza la conversión de una molécula de sustrato que no puede alterar ni afectar al crecimiento y/o viabilidad celulares, produciendo un producto que comprende una toxina, compuesto citostático o compuesto antimetabólico. Según esta realización, la expresión de la molécula indicadora en presencia de dicho sustrato conduce a la producción de una concentración suficientemente alta de la toxina, compuesto citostático o compuesto antimetabólico para reducir el crecimiento celular o dar como resultado la muerte celular.

En una realización adicional, la molécula indicadora es una enzima que, cuando se expresa en la célula hospedadora, cataliza la conversión de una molécula de sustrato citostática o antimetabólica produciendo un producto es que incapaz de modificar el crecimiento y/o viabilidad celulares. Según esta realización, las células incubadas en presencia de la molécula de sustrato no crecen ni se dividen tan rápidamente como las células que no se incuban con la misma. Cuando las células incubadas en presencia de la molécula de sustrato citostática o antimetabólica expresan la molécula indicadora, se retoma la división celular y/o crecimiento celular cuando se reduce la concentración de dicho sustrato en dicha célula.

En una realización alternativa, la molécula indicadora potencia directa o indirectamente el crecimiento y/o viabilidad celulares, por ejemplo, acoplado la expresión de la molécula indicadora con la producción de un mitógeno o molécula reguladora positiva del crecimiento.

En una realización adicional, la molécula indicadora es una enzima que, cuando se expresa en la célula hospedadora, cataliza la conversión de un primer compuesto que es inactivo en la modulación del crecimiento y/o viabilidad celulares produciendo un mitógeno o producto molecular regulador positivo del crecimiento. Según esta realización, las células incubadas en presencia de la molécula de sustrato crecen y se dividen a una velocidad normal en comparación con otras células. La expresión de la molécula indicadora enzimática en presencia de la molécula de sustrato conduce a un crecimiento celular y/o división celular potenciadas a medida que aumenta en la célula la concentración de mitógeno o molécula reguladora positiva del crecimiento. Como consecuencia, las células en que se potencia la molécula indicadora como resultado de la actividad biológica crecen y/o se dividen más rápidamente que las células circundantes en la colección, facilitando su detección.

La secuencia aminoacídica que puede identificarse usando los métodos anteriormente descritos es capaz de modular la expresión de la molécula indicadora. Por consiguiente, la secuencia aminoacídica puede ser un agonista o antagonista de la actividad biológica bajo la que se ha dispuesto operativamente la expresión de la molécula indicadora. Cuando la secuencia aminoacídica es una molécula agonista, la expresión de la molécula indicadora aumentará o se potenciará o activará y, dependiendo de si la molécula indicadora aumenta o reduce directa o indirectamente o no el crecimiento y/o viabilidad celulares, el crecimiento celular aumentará o se reducirá, respectivamente.

Sin embargo, es claramente indeseable que la molécula indicadora dé como resultado la muerte celular, porque no sería posible recuperar las células que expresan el péptido deseado. Cuando la secuencia aminoacídica es un antagonista de la actividad biológica, la expresión de la molécula indicadora se reducirá o reprimirá o inactivará y, dependiendo de si la molécula indicadora aumenta o reduce directa o indirectamente o no el crecimiento y/o viabilidad celulares, el crecimiento celular se reducirá o aumentará, respectivamente. Cuando la molécula indicadora conduce directa o indirectamente a la muerte celular, el antagonismo de la actividad biológica por la secuencia aminoacídica antagonista facilita la supervivencia de la célula en comparación con células que no expresan el antagonista pero expresan la molécula indicadora.

Los ejemplos de genes indicadores seleccionables positivos en levadura adecuados (adecuados para el aislamiento de agonistas peptídicos) incluyen, pero sin limitación, *HIS3* y *LEU2*, cuyos productos proteicos permiten que las células que expresan estos genes indicadores sobrevivan en un medio de cultivo celular apropiado. A la inversa, existen varios genes indicadores contraseleccionables en levadura (adecuados para el aislamiento de antagonistas peptídicos), incluyendo el gen *URA3*, en el que la expresión de *URA3* es tóxica para una célula que expresa este gen en presencia del fármaco ácido 5-fluororótico (5FOA). Otros genes indicadores contraseleccionables incluyen *CYH1* y *LYS2*, que confieren letalidad en presencia de los fármacos cicloheximida y -aminoácido (αAA), respectivamente. Para contraselección en bacterias, están disponibles los correspondientes genes indicadores que codifican productos tóxicos, incluyendo: *SacB*, *CcdB* y el gen *GATA-1* de mamífero, cuya expresión es tóxica en *E. coli*.

Se usan métodos estándares para introducir la primera y segunda secuencias nucleotídicas en el hospedador celular. En el caso de células de levadura, esto puede conseguirse mediante apareamiento en masa o transformación.

La primera y segunda secuencias nucleotídicas pueden estar contenidas cada una dentro de un constructo genético separado que comprende adicionalmente un gen marcador seleccionable para facilitar la detección de las células transformadas, por ejemplo, un gen marcador seleccionable de resistencia a antibióticos. Preferiblemente, los genes marcadores seleccionables para cada constructo genético son diferentes, de tal forma que pueda facilitarse la presencia de uno o ambos constructos genéticos en una única célula. La primera y segunda secuencias

nucleotídicas pueden, por tanto, introducirse en el hospedador celular mediante cotransformación forzada y selección en un medio apropiado para seleccionar la presencia de ambos genes marcadores seleccionables.

Como alternativa, la primera y segunda secuencias de ácido nucleico pueden introducirse mediante transformación secuencial, acompañada de selección de los genes marcadores apropiados después de cada evento de transformación.

Como alternativa, la primera y segunda secuencias de nucleótidos pueden introducirse en poblaciones separadas de células hospedadoras que posteriormente se aparean, y aquellas poblaciones celulares que contienen ambas secuencias nucleotídicas se seleccionan en un medio que permite el crecimiento de las células hospedadoras transformadas con éxito con ambas primera y segunda moléculas de ácido nucleico.

Como alternativa, la primera y segunda secuencias nucleotídicas pueden estar contenidas en un único constructo genético e introducirse en la población de células hospedadoras en una sola etapa. Puede producirse una colección de péptidos aleatorios usando un vector que comprende al menos la primera secuencia nucleotídica dispuesta bajo el control operativo de un promotor adecuado con o sin una secuencia operadora y un gen marcador seleccionable, seleccionándose el sitio de inserción para la segunda secuencia nucleotídica de tal forma que la segunda secuencia nucleotídica pueda expresarse.

Estos rasgos son, además de las etapas que hay que efectuar en relación con la introducción de una o más moléculas de ácido nucleico adicionales que codifican uno o más coparticipes de unión polipeptídicos de la actividad biológica, variaciones de lo que se ha descrito anteriormente.

Las células hospedadoras seleccionadas pueden cribarse en medios que comprenden los componentes requeridos para utilizar la molécula indicadora contraseleccionable. Las células hospedadoras que expresan un péptido que inhibe la actividad biológica son incapaces de transcribir adecuadamente el gen indicador contraseleccionable, permitiendo así que la célula hospedadora viva en el medio de selección. Aquellas células hospedadoras que expresan las secuencias aminoacídicas que son incapaces de inhibir la actividad biológica transcriben el gen indicador, dando así como resultado la formación de un producto que es tóxico para la célula hospedadora en presencia del medio de selección.

El constructo genético puede estar en forma de un vector de replicación autónoma o puede comprender secuencias genéticas para facilitar la integración en un genoma de célula hospedadora.

Como alternativa, la primera secuencia nucleotídica que codifica la molécula indicadora puede integrarse en el cromosoma de la célula hospedadora mediante recombinación homóloga de los productos de la reacción en cadena de la polimerasa (PCR) o de secuencias en otra molécula de ADN que sea incapaz de replicarse de forma autónoma en células de levadura.

Según la naturaleza de la actividad biológica de interés, la primera secuencia nucleotídica puede disponerse en conexión operativa con cualquier secuencia promotora, siendo el único requisito que el promotor sea capaz de regular la expresión génica en la célula hospedadora seleccionada. Habitualmente, la célula hospedadora variará para adecuarse a la secuencia promotora. Pueden aislarse péptidos que son capaces de modular cualquier actividad biológica.

De hecho, la presente invención facilitará la identificación y aislamiento de una secuencia aminoacídica que modula o media la expresión de una molécula indicadora mediante el agonismo o antagonismo de cualquier etapa reguladora que se requiera para que ocurra la expresión, no simplemente las etapas posteriores en la ruta de transducción de señal, tales como las interacciones ADN-proteína o las interacciones entre factores de transcripción. Cuando se desee aislar una secuencia aminoacídica específica que sea capaz de modular una actividad biológica particular, solo es necesario conectar operativamente la expresión de la primera secuencia nucleotídica con la actividad biológica de interés. Esto se realiza disponiendo la primera secuencia nucleotídica en conexión operativa con una secuencia promotora que se regula por la actividad biológica o, como alternativa, manipulando genéticamente una secuencia promotora que está conectada operativamente con la primera molécula de ácido nucleico, disponiendo así el promotor bajo el control operativo de la actividad biológica.

En el caso de secuencias aminoacídicas que modulan o median una interacción proteína:ADN que se requiere para la expresión génica o modulación de la expresión génica, por ejemplo, para aislar una molécula peptídica que interacciona directamente con un elemento potenciador o silenciador de acción en cis o una proteína con la que se une dicho elemento, este objetivo puede conseguirse introduciendo el elemento de acción en cis en una secuencia promotora a la que está conectada operativamente la primera secuencia nucleotídica. Por este medio, la expresión de la molécula indicadora se dispone bajo el control operativo del elemento de acción en cis y ocurrirá la modulación de la expresión génica cuando la molécula de proteína apropiada se una al elemento de ADN de acción en cis o a la proteína que reconoce dicho elemento.

En el caso de una interacción proteína:proteína que controla la expresión génica, se selecciona el promotor que controla la expresión de la primera molécula de ácido nucleico de tal forma que contenga los elementos de acción en cis necesarios, a los que se une al menos una de las proteínas implicadas en la interacción. Cuando no se conocen

completamente las secuencias de acción en cis o los factores de acción en trans implicados en la regulación de la expresión génica, pero se conoce la secuencia promotora y el tipo celular en el que ocurre la expresión, la primera secuencia nucleotídica puede disponerse en conexión operativa con esa secuencia promotora e introducirse la molécula de ácido nucleico resultante en ese tipo celular. Dicha relación forma la base de los enfoques de cribado de “doble híbrido”. Cuando el péptido de interés antagoniza o agoniza cualquier etapa requerida para la expresión o la activación, represión o potenciación de la expresión génica, el efecto se identificará registrando la expresión alterada de la molécula indicadora.

Pueden detectarse secuencias aminoacídicas que modulan una actividad biológica en una célula de mamífero, en las que la expresión del gen indicador contraseleccionable se dispone bajo el control operativo de una secuencia promotora expresable en mamíferos, que es activa aberrantemente en la situación patogénica, por ejemplo un promotor oncogénico tal como MYC. La actividad de dicho promotor sería bloqueada directamente en células que expresan una secuencia aminoacídica capaz de inhibir el promotor oncogénico en una célula de mamífero.

Un método para identificar una secuencia aminoacídica que es capaz de antagonizar una interacción proteína:proteína en una célula hospedadora puede comprender las etapas de:

- (i) producir una colección de péptidos en una célula hospedadora, en la que las células transformadas de dicha colección contienen al menos una primera secuencia nucleotídica que comprende o codifica una molécula indicadora capaz de reducir el crecimiento y/o viabilidad de dicha célula hospedadora cuya expresión está bajo el control operativo de dicha interacción proteína:proteína, y una segunda secuencia nucleotídica derivada de una secuencia genómica definida que puede codificar dicha secuencia aminoacídica cuando se dispone bajo el control operativo de una secuencia promotora adecuada, y en la que (b) sustancialmente toda la secuencia genómica definida está presente en de la población de células transformadas que constituye dicha colección;
- (ii) cultivar dicho hospedador celular durante un tiempo y en condiciones suficientes para que ocurra la expresión de dicha segunda secuencia nucleotídica; y
- (iii) seleccionar las células en las que la expresión de dicha molécula indicadora se ha antagonizado, reprimido o reducido.

El método descrito anteriormente puede incluir la primera o posterior etapa adicional de introducir en el hospedador celular una o más moléculas de ácido nucleico adicionales que codifican uno o más copartícipes de unión polipeptídicos que están implicados en la actividad biológica bajo el control operativo de una o más secuencias promotoras. Dichos rasgos se describen en detalle anteriormente.

La molécula indicadora puede comprender un péptido, polipéptido, enzima u otra molécula proteica que sea capaz de convertir una molécula de sustrato inocua en un compuesto citostático, compuesto antimetabólico o toxina, de tal forma que la expresión antagonizada de la molécula indicadora por el péptido en cuestión evite la muerte celular o al menos evite una reducción en el crecimiento y/o viabilidad celulares en presencia del sustrato. Más preferiblemente, en el sistema de levadura, el gen indicador es *URA3* y/o *CYH2*, entre otros tales como *LYS2*. Con particular preferencia, la molécula indicadora es el producto del gen *URA3* que, cuando se expresa, convierte el ácido 5-fluororótico (5-FOA) en un producto tóxico.

Una ejemplificación de este método aprovecha el hecho de que los activadores de la transcripción eucarióticos más activos son modulares y comprenden un dominio de unión a ADN y un dominio de activación de ADN, en los que el dominio de unión a ADN y el dominio de activación de ADN pueden estar contenidos en la misma molécula de proteína o, como alternativa, en moléculas separadas que interaccionan para regular la expresión génica. La expresión de la molécula indicadora puede disponerse bajo el control operativo de una interacción proteína:proteína, por ejemplo entre las proteínas oncogénicas SCL y LMO2, que se unen formando un factor de transcripción artificial activo. La transcripción del gen indicador puede usarse por lo tanto como indicador de dos proteínas que interaccionan, en la que una de dichas proteínas de interés comprende al menos un dominio de unión a ADN y se une a un elemento promotor operador en dirección 5' del gen indicador, y dicha otra proteína de interés comprende al menos un dominio de activación de ADN. La unión de la proteína de unión a ADN al operador, en presencia de un dominio de activación de la función, inicia la transcripción del gen indicador. El indicador *URA3* actúa así como marcador contraseleccionable.

Este ejemplo puede adaptarse a la identificación de secuencias aminoacídicas que modulan otras interacciones proteína:proteína mediante la sustitución funcional del dominio de unión a ADN de un factor de transcripción por un dominio de unión a ADN diferente que es específico de un elemento de acción en cis diferente en la expresión regulada por promotor de la molécula indicadora. Los métodos para la producción de dichas proteínas de fusión son bien conocidos por los expertos en la materia. En dichos casos, la selección de un dominio de unión a ADN apropiado dependerá de la naturaleza del sitio de unión a ADN situado en dirección 5' del gen indicador.

Por ejemplo, pueden construirse proteínas de fusión entre una oncoproteína y un dominio de unión a ADN y/o un dominio de activación de ADN. Por ejemplo, una secuencia nucleotídica que codifica o es complementaria de una secuencia nucleotídica que codifica los restos 176 a 331 de SCL puede fusionarse con el dominio de unión a ADN

LexA, y una secuencia nucleotídica que codifica LMO2 puede fusionarse con un dominio de activación de ADN (o viceversa).

5 Los métodos descritos en la presente memoria son útiles para identificar secuencias aminoacídicas que inhiben las interacciones proteína:proteína que normalmente producen efectos perjudiciales (aparte del efecto perjudicial de ciertas moléculas indicadoras), por ejemplo las interacciones que implican productos oncogénicos. Los ejemplos específicos de oncogenes cuyos productos forman factores de transcripción que contribuyen a la tumorigénesis incluyen SCL y uno cualquiera o más de DRG, E47 y/o LMO2.

Un método para identificar una secuencia aminoacídica que es capaz de modificar una interacción proteína:proteína en una célula hospedadora puede comprender las etapas de:

- 10 (i) producir una colección de péptidos en un hospedador, en la que (a) las células transformadas de dicha colección contienen: (1) al menos una primera secuencia nucleotídica que comprende o codifica una molécula indicadora, en la que dicha secuencia nucleotídica está conectada operativamente con una secuencia operadora o un sitio de unión de factor de transcripción; (2) una segunda secuencia nucleotídica derivada de una secuencia genómica definida que codifica dicha secuencia aminoacídica
15 cuando se dispone bajo el control operativo de una secuencia promotora adecuada; y (3) una o más terceras secuencias nucleotídicas adicionales que codifican uno o más polipéptidos, proteínas o proteínas de fusión, en la que al menos uno de dichos polipéptidos, proteínas o proteínas de fusión incluye al menos un dominio de unión a ADN capaz de unirse a dicha secuencia operadora o sitio de unión de factor de transcripción, y al menos uno de dichos polipéptidos, proteínas o proteínas de fusión
20 incluye al menos un dominio de activación de ADN o derivado del mismo capaz de activar la expresión de dicha primera secuencia nucleotídica cuando se orienta al promotor/operador mediante interacción con otra proteína que porta el dominio de unión a ADN asociado; y (b) sustancialmente toda la secuencia genómica definida está presente en la población de células transformadas que constituye dicha colección;
- 25 (ii) cultivar dicho hospedador celular durante un tiempo y en condiciones suficientes para permitir que ocurra la expresión de dicha segunda y adicional secuencias nucleotídicas; y
- (iii) seleccionar las células en las que la expresión de dicha molécula indicadora se ha activado, inhibido o modificado de otra forma.

30 Las proteínas implicadas en la actividad biológica de interés, que están codificadas por la segunda molécula de ácido nucleico, se sintetizan en la célula hospedadora, codificadas por una o más secuencias nucleotídicas extrañas transformadas en la célula hospedadora, o integradas en el genoma de dicha célula. Sin embargo, están también contempladas situaciones en las que estas secuencias están también codificadas por genes endógenos de la célula hospedadora.

35 El dominio de unión a ADN puede unirse a la secuencia operadora y, en presencia de la región activadora de ADN, ocurre la expresión de la molécula indicadora. Cuando la segunda secuencia nucleotídica codifica un péptido que antagoniza o inhibe la unión al ADN y/o la activación del ADN, la expresión de la molécula indicadora se reprime, reduce o inhibe de otra forma. Como alternativa, cuando la segunda secuencia nucleotídica codifica una secuencia aminoacídica que agoniza o potencia la unión al ADN y/o la activación del ADN, la expresión de la molécula indicadora se activa, potencia o aumenta de otra forma.

40 Los expertos en la materia reconocerán que el dominio de unión a ADN y el dominio de activación de ADN pueden estar contenidos en una única molécula aminoacídica o, como alternativa, pueden estar contenidos en moléculas aminoacídicas separadas que interaccionan entre sí para regular la expresión de genes indicadores.

45 De forma similar, la primera y/o segunda y/o adicionales secuencias nucleotídicas pueden estar contenidas en una única molécula de ácido nucleico, por ejemplo en un constructo genético o, como alternativa, una, dos, tres o más de dichas secuencias pueden estar contenidas en moléculas de ácido nucleico separadas. Cuando una o más de las secuencias nucleotídicas están contenidas en moléculas de ácido nucleico separadas, entonces cada secuencia nucleotídica está además preferiblemente conectada operativamente con su propia secuencia promotora. Como alternativa, cuando cualesquiera de dos o más de las secuencias nucleotídicas están contenidas en las mismas moléculas de ácido nucleico, las secuencias nucleotídicas pueden expresarse bajo el control de un único promotor o,
50 como alternativa, bajo el control de secuencias promotoras separadas.

Los expertos en la materia reconocerán que las alternativas descritas anteriormente son igualmente aplicables a los métodos descritos en la presente memoria.

55 El método descrito anteriormente puede comprender adicionalmente la etapa de aislar la segunda secuencia nucleotídica a partir de la célula hospedadora, de secuenciar la molécula de ácido nucleico y de derivar la secuencia aminoacídica codificada por la misma. Una vez se ha identificado la secuencia, puede compararse entonces con secuencias parecidas dentro de la secuencia nucleotídica conocida para identificar aquellos aminoácidos que son

esenciales para la modulación de la actividad biológica. Pueden producirse entonces péptidos sintéticos, basándose en la secuencia aminoacídica derivada así obtenida. Los expertos en la materia conocen bien dichas técnicas.

Las secuencias aminoacídicas pueden identificarse mediante los métodos descritos en la presente memoria.

- 5 Preferiblemente, las secuencias aminoacídicas son agonistas o antagonistas de interacciones proteína:proteína o proteína:ADN. Más preferiblemente, los péptidos, oligopéptidos y polipéptidos de la presente invención son antagonistas de interacciones proteína:proteína o interacciones proteína:ADN, y aún más preferiblemente, antagonistas de interacciones proteína:proteína.

- 10 Los péptidos pueden antagonizar o inhibir interacciones que producen efectos perjudiciales en células eucarióticas, en particular células humanas o animales. Más preferiblemente, las secuencias aminoacídicas de la invención antagonizan o inhiben interacciones que implican una o más oncoproteínas.

- 15 Dichas secuencias aminoacídicas o fragmentos o derivados de las mismas pueden usarse en el tratamiento profiláctico o terapéutico de un ser humano o animal. Los métodos de tratamiento incluyen su uso en regímenes de terapia de antibióticos peptídicos, tales como en protocolos de tratamiento para pacientes con infecciones bacterianas, fúngicas o víricas. Su uso en protocolos de tratamiento para dichos pacientes incluye su administración como medio para inhibir el crecimiento de los microorganismos infectantes y/o inhibir su virulencia. También se contempla el uso de dichos péptidos para potenciar los efectos de otros agentes antimicrobianos (por ejemplo, véase la solicitud PCT internacional WO 96/24684).

Una composición farmacéutica puede comprender un péptido, oligopéptido y polipéptido que sea capaz de modular una actividad biológica, y uno o más portadores y/o diluyentes farmacéuticamente aceptables.

- 20 En dicha composición farmacéutica, dicho péptido, oligopéptido y polipéptido pueden antagonizar el crecimiento y/o virulencia de un patógeno.

Se hace referencia a estos componentes como ingredientes activos.

- 25 Las formas farmacéuticas adecuadas para uso inyectable incluyen disoluciones (cuando son hidrosolubles) o dispersiones acuosas estériles, y polvos estériles para la preparación extemporánea de disoluciones o dispersiones inyectables estériles, o pueden estar en forma de una crema u otra forma adecuada para aplicación tópica. Como alternativa, las disoluciones inyectables pueden suministrarse encapsuladas en liposomas para ayudar a su transporte a través de la membrana celular. Como alternativa o además, dichas preparaciones pueden contener constituyentes de estructuras de poro autoensamblantes para facilitar el transporte a través de la membrana celular. Deben ser estables bajo las condiciones de fabricación y almacenamiento, y deben conservarse ante la acción contaminante/destructiva de microorganismos ambientales como bacterias y hongos. El portador puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido, y similares), mezclas adecuadas de los mismos y aceites vegetales. La fluidez apropiada puede mantenerse, por ejemplo, mediante el uso de un recubrimiento como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersión, y mediante el uso de tensioactivos. La prevención de la acción de microorganismos puede lograrse mediante diversos agentes antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido sórbico, timerosal y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares o cloruro de sodio. La absorción prolongada de composiciones inyectables puede lograrse mediante el uso en las composiciones de agentes retardantes de la absorción, por ejemplo, monoestearato de aluminio y gelatina.

- 40 Las disoluciones inyectables estériles se preparan incorporando los compuestos activos en la cantidad requerida al disolvente apropiado con varios de los otros ingredientes enumerados anteriormente, según se requiera, seguido de esterilización por filtración. En general, las dispersiones se preparan incorporando los diversos ingredientes activos esterilizados a un vehículo estéril que contiene el medio de dispersión básico y los demás ingredientes requeridos de los enumerados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los métodos preferidos de preparación son secado a vacío y la técnica de liofilización, que proporcionan un polvo del ingrediente activo más cualquier ingrediente adicional deseado a partir de una disolución del mismo previamente esterilizada mediante filtración.

- 50 Cuando los ingredientes activos están protegidos de forma adecuada, pueden administrarse por vía oral, por ejemplo, con un diluyente inerte o con un portador comestible asimilable, o pueden encerrarse en una cápsula de gelatina de cubierta dura o blanda, o pueden comprimirse en comprimidos, o pueden incorporarse directamente al alimento de la dieta. Para administración terapéutica oral, el compuesto activo puede incorporarse con excipientes y usarse en forma de comprimidos ingeribles, comprimidos bucales, trociscos, cápsulas, elixires, suspensiones, jarabes, obleas y similares. Dichas composiciones y preparaciones deberían contener al menos un 1% en peso del compuesto activo. El porcentaje de las composiciones y preparaciones puede variar, por supuesto, y puede estar convenientemente entre aproximadamente 5% y aproximadamente 80% del peso de la unidad. La cantidad de compuesto activo en dichas composiciones terapéuticamente útiles es tal que se obtenga una dosificación adecuada. Las composiciones o preparaciones preferidas se preparan de forma que una forma de dosificación unitaria contenga entre aproximadamente 0,1 µg y 20 g de compuesto activo.

Los comprimidos, trociscos, píldoras, cápsulas y similares pueden contener también los componentes que se enumeran a continuación: un ligante tal como goma, goma arábica, almidón de maíz o gelatina; excipientes tales como fosfato de dicalcio; un agente disgregante tal como almidón de maíz, almidón de patata, ácido alginico y similares; un lubricante tal como estearato de magnesio y pueden añadirse un agente edulcorante tal como sacarosa, lactosa o sacarina o un agente aromatizante tal como menta, aceite de gualteria o aroma de cereza. Cuando la forma de dosificación unitaria es una cápsula puede contener, además de los materiales del tipo anterior, un portador líquido. Diversos otros materiales pueden estar presentes como recubrimientos o pueden modificar de otro modo la forma física de la unidad de dosificación. Por ejemplo, los comprimidos, píldoras o cápsulas pueden recubrirse con goma laca, azúcar o ambos. Un jarabe o elixir puede contener el compuesto activo, sacarosa como agente edulcorante, metil- y propilparabenos como conservantes, un tinte y un aromatizante como aroma de cereza o naranja. Por supuesto, cualquier material usado en la preparación de cualquier unidad de dosificación debería ser farmacéuticamente puro y sustancialmente no tóxico en las cantidades empleadas. Además, el compuesto o compuestos activos pueden incorporarse a preparaciones y formulaciones de liberación sostenida.

Las formas adecuadas para aplicación tópica incluyen cremas, lociones y geles.

Los portadores y/o diluyentes farmacéuticamente aceptables incluyen todos y cada uno de los disolventes, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y retardantes de la absorción, y similares. El uso de dichos medios y agentes para sustancias farmacéuticamente activas es bien conocido en la materia. Excepto si cualquier medio o agente convencional es incompatible con el ingrediente activo, se contempla el uso de los mismos en composiciones terapéuticas. También pueden incorporarse ingredientes activos suplementarios a las composiciones.

Resulta especialmente ventajoso formular composiciones parenterales en una forma de dosificación unitaria por su facilidad de administración y uniformidad de dosificación. Forma de dosificación unitaria, como se usa en la presente memoria, hace referencia a unidades físicamente discretas adecuadas como dosificaciones unitarias para los sujetos mamíferos que se van a tratar; conteniendo cada unidad una cantidad predeterminada de material activo calculada para producir el efecto terapéutico deseado en asociación con el portador farmacéutico requerido. Las especificaciones para las formas de dosificación unitarias novedosas están dictadas y son directamente dependientes de (a) las características únicas del material activo y el efecto terapéutico particular que se va a conseguir, y (b) las limitaciones inherentes en la técnica de mezclado intensivo de dicho material activo para el tratamiento de una enfermedad en sujetos vivos que tienen una afección patológica en la que la salud corporal está deteriorada, como se describe en la presente memoria en detalle.

El ingrediente activo principal se mezcla intensivamente para una administración conveniente y eficaz en cantidades eficaces con un portador farmacéuticamente aceptable en una forma de dosificación unitaria. Una forma de dosificación unitaria puede contener, por ejemplo, el compuesto activo principal en cantidades en el intervalo de 0,5 µg a aproximadamente 2.000 mg. Expresado en proporciones, el compuesto activo está presente, en general, desde aproximadamente 0,5 µg a aproximadamente 2.000 mg/ml de portador. En el caso de composiciones que contienen ingredientes activos suplementarios, las dosificaciones se determinan por referencia a la dosis y vía de administración habituales de dichos ingredientes.

La composición farmacéutica puede comprender también moléculas genéticas tales como un vector capaz de transfectar células diana, en el que el vector porta una molécula de ácido nucleico capaz de inhibir dichas interacciones/actividades biológicas perjudiciales. El vector puede ser, por ejemplo, un vector vírico.

EJEMPLOS

Se describen con más detalle rasgos adicionales de la presente invención en los siguientes ejemplos no limitantes. Sin embargo, ha de entenderse que esta descripción detallada se incluye solo con fines de ejemplificar la presente invención. No debería entenderse en modo alguno como una restricción de la descripción amplia de la invención como se ha expuesto anteriormente y se define en las reivindicaciones adjuntas.

Los métodos de clonación molecular, inmunología y química de proteínas que no se describen explícitamente en los siguientes ejemplos se reseñan en la bibliografía y son conocidos por los expertos en la materia. Los textos generales que describen técnicas convencionales de biología molecular, microbiología y ADN recombinante dentro del conocimiento de un experto en la materia incluyen, por ejemplo: Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 2ª edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York (1989); Glover ed., "DNA Cloning: A Practical Approach", volúmenes I y II, MRL Press, Ltd., Oxford, R.U. (1985) y Ausubel, F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. "Current protocols in molecular biology". Greene Publishing Associates/Wiley Intersciences, Nueva York.

Ejemplo 1

La construcción de colecciones de fragmentos génicos biodiversos

Se redujo el ADN genómico de un panel diverso de microorganismos, elegidos para maximizar la diversidad genética en el panel, a fragmentos adecuados para expresar péptidos. Las técnicas adecuadas conseguir este

resultado incluyen: cizallamiento mecánico, digestión parcial con ADNasa 1 y el uso de combinaciones de endonucleasas de restricción.

Se añadió entonces cada genoma al agrupamiento en proporción directa a su tamaño y complejidad. Se añadió más ADN de genomas grandes que de genomas pequeños para asegurar una representación adecuada.

- 5 Se construyó entonces una colección de péptidos digiriendo alícuotas del ADN agrupado con todas las 6 combinaciones de 2 enzimas de restricción de un conjunto que contenía *Alu* I, *Bst* U I, *Hae* III y *Rsa* I. Estas enzimas son endonucleasas de restricción de corte de extremos romos que tienen secuencias de reconocimiento de 4 pares de bases distintas, y por tanto cada combinación producirá fragmentos con tamaños predominantes en el intervalo de 90-120 pb. Estos son adecuados para la clonación, y la longitud del ADN es suficiente para codificar péptidos de aproximadamente 30 restos aminoacídicos, que están en el intervalo de tamaños de secuencias reseñados en regiones estructuralmente conservadas de proteínas. En los casos en que se van a ligar ligadores en lugar de adaptadores a los fragmentos genómicos, el agrupamiento de digestiones genómicas puede protegerse de la posterior digestión mediante tratamiento con una metilasa apropiada (en este ejemplo, metilasa de *Eco*RI).

- 15 Se purificaron los fragmentos de digestión mediante electroforesis en gel de acrilamida nativa, seguida de una cromatografía de filtración en gel.

Se ligaron entonces los ligadores a los fragmentos de ADN mediante métodos estándares. Se usaron 3 marcos de lectura de ligadores. Cuando los fragmentos se van a clonar en un sitio *Eco*RI, pueden usarse cantidades equimolares de los siguientes 3 ligadores autoreasociativos:

d(pGGAATTCC), d(pCGGAATTCCG) y d(pCCGGAATTCCGG)

- 20 Cuando se pretendía que la clonación fuera direccional, se añadió una cantidad equimolar de otro ligador correspondiente al segundo sitio de restricción 3', por ejemplo, para la clonación en los sitios *Eco*RI y *Hind*III de un vector (por ejemplo, se añadió al ligamiento un número igual de moles (a los ligadores *Eco*RI combinados) del siguiente ligador: d(pCCAAGCTTGG)).

- 25 Se digirieron entonces los ligadores con la endonucleasa o endonucleasas correspondientes a sus secuencias de reconocimiento y se purificaron los fragmentos de digestión con el tamaño apropiado mediante técnicas estándares que incluyen electroforesis en gel de agarosa, gradientes de sacarosa o acetato de potasio, o cromatografía de exclusión por tamaño.

- 30 Se clonaron entonces los fragmentos genómicos que contienen ligadores o adaptadores flanqueantes (véase el ejemplo 4 a continuación) en un vector de expresión pT7-Select mediante la metodología estándar para la construcción de colecciones.

Ejemplo 2

Construcción de colecciones de BGF

- 35 Las colecciones de fragmentos génicos biodiversos pueden construirse utilizando fragmentos adaptados de ADN genómico agrupado procedente de un conjunto evolutivamente diverso de genomas compactos. Para maximizar la diversidad del agrupamiento, puede aumentarse la concentración relativa del ADN en el agrupamiento de los genomas más grandes en proporción al tamaño de genoma haploide total. Los insertos genómicos pueden fragmentarse utilizando cizallamiento mecánico (por ejemplo, sonicación), seguido de la reparación y ligamiento de los oligonucleótidos ligadores o adaptadores. Como alternativa, pueden prepararse mediante extensión con polimerasa de oligonucleótidos parcialmente degenerados reasociados con el ADN genómico desnaturalizado, seguido de amplificación usando la reacción en cadena de la polimerasa (PCR).

- 40 En este ejemplo, los oligonucleótidos utilizados en la extensión primaria con el fragmento de Klenow de la ADN polimerasa I (a 15-25°C) tenían las secuencias

(se usa * para representar una base universal, como 5-nitroindol):

Cebador de codificación: **GACTACAAGGACGACGACGACAAGNNNNNNNNN***

- 45 Cebador inverso: **ATTCCCGGGAAGCTTATCAATCAATCANNNNNNNNN***

N corresponde a nucleótidos degenerados (por ejemplo, dATP, dCTP, dGTP o dTTP). Además, cualquiera de las bases universales: desoxiinosina o 5-nitroindol (o un análogo funcionalmente equivalente) puede sustituirse en cualquiera o todas las posiciones "N" del cebador, en especial en la posición 3'-terminal. Por tanto, la longitud de la serie "N" puede variar de 6 a 8 nucleótidos.

Según este ejemplo, los cebadores para la amplificación por PCR anidada del producto de la reacción de extensión de Klenow eran:

Cebador de codificación: **GAGAGGAATTCAGACTACAAGGACGACGACGACAAG**

Cebador inverso: **GAGAGAATTCCCGGGAAGCTTATCAATCAATCA**

- 5 Se efectuó la amplificación por PCR usando un protocolo "Touchdown" con enzima "de inicio en caliente" para maximizar la especificidad.

La extensión inicial y amplificación por PCR se efectuaron enteramente con polimerasa de Klenow, añadiendo más polimerasa en cada ciclo, como en el informe inicial de la PCR. Esto permite efectuar la ciclación completa entre la temperatura de desnaturalización (90-100°C) y una temperatura de reasociación baja (15-25°C), minimizando el sesgo de reasociación potencial frente a la amplificación de secuencias ricas en A/T. Para este enfoque, los cebadores tenían la forma:

Cebador de codificación: **GAGAATTCANNNNNNNNN'**

Cebador inverso: **GAGAATTCNNNNNNNNN***

- 15 Pueden incluirse nucleótidos metilados en la reacción de PCR (pero no se incorporan a los cebadores) para proteger a los productos de la escisión interna con enzimas de restricción durante la clonación.

En una forma preferida del ejemplo, también puede emplearse una PCR mutagénica usando nucleótidos alternativos y/o el uso de un tampón de manganeso para aumentar la diversidad de las secuencias de los insertos.

Se digirieron los productos de PCR resultantes con EcoRI sola o con EcoRI y XmaI (cuando el cebador inverso contiene un sitio XmaI) antes de la clonación en vectores de la serie pBLOCK.

- 20 Se construyeron las colecciones según una metodología convencional usando las células competentes más eficaces comercialmente disponibles, o sea, XL10-Gold (Stratagene) para asegurar unas complejidades mayores de 10^7 clones independientes.

Ejemplo 3

Colecciones de mimótopos que usan fragmentos génicos biodiversos

- 25 Este ejemplo ilustra la detección de mimótopos del alérgeno principal del ácaro del polvo doméstico Der p 1.

Se aisló ADN en el intervalo de 90-120 pb de cada una de las dobles digestiones y se agrupó, se ligó con ligadores en todos los marcos de lectura y se clonó en el vector de presentación en fago T7 Select 1.1, o el vector T7 Select 415. Algunos fragmentos de ADN se encontraban fuera del intervalo de 90-120 pb y no se clonaron, pero la redundancia en los procedimientos de digestión debería permitir una representación de la mayoría de las secuencias. El uso de un agrupamiento de 3 marcos de lectura de ligadores y/o una señal de deslizamiento traduccional en la construcción de la colección aseguró que los 6 marcos de lectura de los insertos estuvieran representados. El tamaño total del genoma de un panel biodiverso de microorganismos era de aproximadamente 35 Mb. Este procedimiento generaba aproximadamente 12×10^6 fragmentos diferentes, permitiendo la clonación en todos los marcos de lectura y orientaciones. Teniendo en cuenta esto último, aproximadamente 1/6 de las secuencias codificaba péptidos naturales. El T7-Select es un sistema de clonación molecular con una alta eficacia de empaquetamiento y está diseñado para presentar los péptidos codificados por el ADN clonado como fusiones C-terminales sobre una proteína de cubierta de fago que es accesible para procedimientos de purificación por afinidad. Una minoría de los péptidos no naturales eran más pequeños que el intervalo de tamaño estimado, porque estarían truncados por codones de terminación. El T7-Select 1.1 o el vector T7-Select 415 presentan el péptido con bajo y alto número de copias, así que pueden usarse interacciones de alta y baja afinidad para purificación por afinidad.

Se generó aún más diversidad mediante mutagénesis por PCR, que realiza la amplificación en condiciones que favorecen altas tasas de error. Se ha calculado que, con una tasa de error de 0,5 bases por 100 pb/ciclo (que puede conseguirse), ocho ciclos mutagénicos producen cambios de base en un 90% de los productos de la PCR, y casi un 50% tendrán 2 o 3 sustituciones. Se añadieron ligadores para proporcionar las secuencias cebadoras para PCR, y se realizó una PCR final de alta fidelidad con ligadores extendidos para proporcionar sitios de clonación. Los fragmentos mutados tenían una diversificación 10x de las secuencias en una cantidad de ADN que se empaquetó con facilidad.

Se aislaron fagos procedentes de colecciones construidas anteriormente que presentan péptidos que se unen a IgG e IgE anti-Der p 1 humanas y de múdo usando métodos basados en los descritos para alérgenos del polen [11] y otros antígenos. Son esencialmente protocolos estándares de purificación por afinidad de fagos que presentan

antígenos. Dichos métodos se han descrito para sistemas de presentación en fagos filamentosos y el sistema de clonación T7-Select.

Se purificó el anticuerpo por afinidad en Sepharose™ acoplada con Der p 1 y se usó para recubrir pocillos de ELISA para inmunoseleccionar fagos mediante un procedimiento de inmunoabsorción. Se efectuaron varios ciclos de selección y amplificación de fagos, como se recomienda. Se han usado varios tipos de métodos de purificación por afinidad para seleccionar fagos, así que hay margen para usar una diversidad de procedimientos. Se aislaron anticuerpos de IgE humanos a partir del suero de sujetos alérgicos e IgG a partir del suero de sujetos alérgicos y no alérgicos. Se utilizaron anticuerpos de IgG monoclonales de ratón, que son conocidos por reconocer un epítipo diferente, para aislar péptidos que imitan diferente epítipo.

Después de la selección y amplificación de los fagos que presentan péptidos, puede obtenerse una purificación adicional usando inmunoensayos en placa efectuados con anticuerpos anti-Der p 1 [11, 12]. Dicho procedimiento permite el aislamiento de clones individuales que reaccionan con los anticuerpos. Se efectuaron inmunoensayos cruzados con diferentes preparaciones de anticuerpos humanos y de ratón para estimar la frecuencia y aislar los mimótopos peptídicos compartidos. Se seleccionaron entonces los fagos para un estudio adicional basado en la secuencia del péptido, la reactividad serológica y el uso previsto. Se ensayó la especificidad de la interacción de mimótomo-anticuerpo mediante ensayos de inhibición frente a otros alérgenos purificados de ácaros y mediante transferencia Western de fuentes proteicas complejas, alérgenos y extractos microbianos.

La actividad de unión a anticuerpos de los mimótopos a menudo puede mejorarse mediante ajustes finos de la secuencia aminoacídica. Los clones que codifican péptidos que reaccionan con anti-Der p 1 se optimizaron para la unión a anticuerpos mediante mutagénesis aleatoria utilizando PCR potenciada para el error de emparejamiento mediante Mn^{++} y altas concentraciones de nucleótidos. Se usarán como cebadores las secuencias que flanquean el sitio de clonación. Para la clonación, se efectuó una PCR final de alta fidelidad usando los cebadores extendidos para contener las secuencias de las enzimas de restricción para reclonación en el vector de presentación. Se usaron entonces los fagos que contenían los insertos mutados para transformar *E. coli* y producir placas para inmunocribado. Se recogieron los clones que muestran la mayor actividad de unión a anticuerpos.

Se ensayó en los péptidos identificados mediante el procedimiento de purificación descrito su capacidad no solo de imitar un epítipo del alérgeno Der p 1, sino también de ser un mimótomo que puede inmunizar a animales o seres humanos para inducir anticuerpos anti-Der p 1. Esto se efectuó de los siguientes modos: con un péptido sintético acoplado químicamente con un portador inmunogénico, con un péptido fusionado genéticamente con un portador mediante técnicas de clonación molecular y usando el fago que presenta el péptido como inmunógeno.

La capacidad de los péptidos de unirse a IgE frente al alérgeno Der p 1 puede usarse para técnicas de diagnóstico que no solo detectan la presencia de anticuerpo, sino que también pueden mostrar la diversidad de la respuesta inmunitaria y el patrón de reconocimiento de epítipo. La capacidad de actuar como mimótomo e inducir un anticuerpo de IgG antialérgico puede usarse en varias estrategias inmunoterapéuticas. De forma importante, pueden producirse constructos para posibilitar usar el péptido como inmunógeno monovalente y, por tanto, para evitar la reacción alérgica causada por moléculas de IgE de reticulación en pacientes alérgicos.

Ejemplo 4

Cribado de colecciones de péptidos que codifican fragmentos génicos biodiversos para agentes antimicrobianos

Para aislar péptidos novedosos con actividad antibiótica contra una cepa de *Staphylococcus aureus* multirresistente, se usó el siguiente enfoque.

Se preparó en primer lugar una colección de fragmentos génicos biodiversos mediante los procedimientos descritos en el ejemplo 1 en un vector de fago T7. Los ejemplos de vectores de fago T7 que pueden usarse en esta parte del método incluyen: T7-Select 415-1b, T7-Select 1-1b, T7-Select 1-2a, T7-Select 1-2b y/o T7-Select 1-2c (Novagen), que tienen una complejidad mayor de 1.000.000 clones individuales.

Se sembró en placa la colección a una multiplicidad de al menos uno sobre un césped de BL21 de *E. coli* (en el caso de T7-Select 415-1b) o en cualquiera de los hospedadores complementarios BLT5403 de *E. coli* o BLT5615 de *E. coli* (para los otros vectores), para permitir una densidad de placa por debajo de la semiconfluencia.

Las placas usadas tenían dos lados, preparadas de manera parecida a placas de cultivo duales unidas entre sí por el lado inferior. Por lo tanto, dichas placas tenían dos tapas en las caras opuestas. La cara compartida de los dos lados de las placas estaba hecha de una membrana de nitrocelulosa o nailon, soportada por una rejilla hecha de un material rígido tal como plástico. El lado opuesto de la placa al lado que contenía el recubrimiento de placa de T7 derivado de BL21 contenía un medio adecuado para el crecimiento de *Staphylococcus aureus*. Después del sembrado en placa de la colección, el *Staphylococcus aureus* estaba sobre la cara de la placa que contenía el medio apropiado a la densidad mínima requerida para obtener un césped.

Se incubaron entonces las placas a 37°C hasta que aparecieron las placas del fago T7 y aparece el césped de *Staphylococcus aureus*. Cualquier discontinuidad en el césped de *Staphylococcus aureus* puede reflejar la difusión de un fármaco inhibidor producido por una placa de fago en una posición correspondiente en el lado opuesto de la placa. Se purificaron entonces las placas hasta clonalidad y se ensayaron las propiedades inhibidoras en posteriores cribados secundarios, terciarios y/o cuaternarios.

Se amplificaron entonces los insertos de placas puras utilizando PCR y se secuenciaron usando cebadores de vector. Se subclonaron entonces los insertos de los clones y se purificaron mediante metodología de expresión bacteriana estándar, usando vectores tales como PET14b, pMAL-c2 o pTYB1, y se ensayó la concentración mínima inhibidora (MIC) mediante métodos conocidos por los expertos en la materia.

La secuencia de los péptidos inhibidores puede usarse entonces para diseñar candidatos a fármacos basados en péptidos sintéticos de los que se ensayaría la actividad antimicrobiana contra *Staphylococcus aureus*.

Ejemplo 5

Selección de bloqueantes de interacciones proteína/proteína a partir de colecciones de fragmentos génicos biodiversos en levaduras

Se construyeron colecciones de doble híbrido inversas y se cribaron usando el vector pBLOCK-1 como se describió en una memoria descriptiva anterior de los inventores (véase el documento PCT/AU99/00018), usando insertos genómicos como se describió anteriormente en el ejemplo 1, con la adición de ligadores EcoRI, y se clonaron en el sitio EcoRI del vector.

Los expertos en la materia conocerán variaciones obvias a este método tales como la posibilidad de usar adaptadores en lugar de ligadores, de usar sitios de clonación alternativos y de incluir secuencias de adición en los ligadores. Por ejemplo, podría usarse un agrupamiento de los siguientes adaptadores reasociados en lugar de los ligadores (cada hebra de la secuencia adaptadora se muestra para lectura de 5' a 3'):

Adaptador 1

AATTCAATCAATCACACACAGGAGGCCACCATGGATGCATGTGTGCAC
GTGCACACATGCATCCATGGTGGCCTCCTGTGTGTGATTGATTG

Adaptador 2

AATTCAATCAATCACACACAGGAGGCCACCATGGATGCATGTGTGCA
TGCACACATGCATCCATGGTGGCCTCCTGTGTGTGATTGATTG

Adaptador 3

AATTCAATCAATCACACACAGGAGGCCACCATGGATGCATGTGTGC
GCACACATGCATCCATGGTGGCCTCCTGTGTGTGATTGATTG

Los adaptadores tales como los mostrados aquí pueden codificar motivos útiles para la expresión o restricción conformacional (por ejemplo, en este caso, secuencias duales de Shine-Dalgarno y Kozak, restos de cisteína flanqueantes y codones de terminación).

Se transformó o apareó la colección en una cepa de levaduras que contiene las dos proteínas cuya interacción se pretende bloquear y que contiene los genes indicadores contraseleccionables cuya expresión depende de esa interacción. La metodología detallada para el cribado de doble híbrido inverso se describe en la memoria descriptiva de los inventores PCT/AU99/00018.

REFERENCIAS

1. Tiozzo, E., Rocco, G., Tossi, A. y Romeo, D. (1998). Biochemical and Biophysical Research Communications, 249, 202-206.
2. Balaban, N., Goldkorn, T., Nhan, R., Dang, L., Scott, RM, R., Rasooly, A., Wright, S., Larrick, J., Rasooly, R. y Carlson, J. (1998). Science, 280, 438-440.
3. Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J., Brent, R. (1996). Nature, 380, 548-550

4. Xu, C, Mendelsohn, A. y Brent, R. (1997). Proc.Natl. Acad. Sci. USA, 94, 12473- 12478.
5. Kolonin, M. y Finley, R. (1998). Proc. Natl. Acad. Sci. USA, 95, 14266- 14271.
6. Derossi, D., Joliot, A.H., Chassaing, G., Prochiantz, A. (1994). Journal of Biological Chemistry, 269, 10444-10450.
- 5 7. Phelan, A. (1998). Nature Biotechnology, 16, 440-443.
8. Marcello, A., Loregion, A., Cross, A., Marsden, H., Hirst, T., Palu, G. (1994). Proc. Natl. Acad. Sci. U.S.A, 91, 8994-8998.
9. Fåhræus, E., Paramio, J.M., Ball, K.L., Lain, S., Lane, D.P. (1996). Current Biology, 6, 84-91.
- 10 10. Mennuni, C, Santini, C, Lazzaro, D., Dotta, F., Farilla, L., Fierabracci, A., Bottazzo, G.F., Di Mario, U., Cortese, R. y Luzzago, A. (1997). Journal of Molecular Biology, 268, 599-606.
11. Leitner, A., Vogel, M., Radauer, C, Breiteneder, H., Stadler, B.M., Scheiner, O., Kraft, D. y Jensen-Jarolim, E. (1998). European Journal of Immunology, 28, 2921-7.
12. Pincus, S.H., Smith, M.J., Jennings, H.J., Burritt, J.B. y Glee, P.M. (1998). Journal of Immunology, 160, 293-8.

LISTA DE SECUENCIAS

<110> Phylogica Limited

5 <120> Aislamiento de moduladores biológicos a partir de colecciones de fragmentos génicos biodiversos

<130> 27.20.76784/21

<150> EP06003767.8

10 <151> 05-05-2000

<150> PCT/AU00/00414

<151> 2000-05-05

15 <150> US 60/132.711

<151> 05-05-1999

<160> 14

20 <170> PatentIn versión 3.3

<210> 1

<211> 19

<212> PRT

25 <213> secuencia artificial

<220>

<223> Proteína de fusión

30 <220>

<221> CARACTERÍSTICA MISCELÁNEA

<222> (18)..(18)

<223> X puede ser cualquier aminoácido y cualquier número de aminoácidos

35 <400> 1

Cys Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys

1 5 10 15

Lys Xaa Cys

<210> 2

<211> 8

<212> DNA

40 <213> secuencia artificial

<220>

<223> Enlazador

45 <400> 2

ggaattcc 8

<210> 3

<211> 10

50 <212> DNA

<213> secuencia artificial

<220>

<223> Enlazador

55 <400> 3

cggaattccg 10

<210> 4

60 <211> 12

<212> DNA

<213> secuencia artificial
 <220>
 <223> Enlazador
 5 <400> 4
 ccggaattcc gg 12
 <210> 5
 10 <211> 10
 <212> DNA
 <213> secuencia artificial
 <220>
 15 <223> Enlazador
 <400> 5
 ccaagcttgg 10
 20 <210> 6
 <211> 32
 <212> DNA
 <213> secuencia artificial
 25 <220>
 <223> Cebador directo
 <220>
 30 <221> característica miscelánea
 <222> (25)..(31)
 <223> N puede ser cualquier nucleótido
 <220>
 35 <221> característica miscelánea
 <222> (32)..(32)
 <223> N puede ser cualquier nucleótido que represente una base universal
 <400> 6
 40 gactacaagg acgacgacga caagnnnnnn nn 32
 <210> 7
 <211> 35
 <212> DNA
 <213> secuencia artificial
 45 <220>
 <223> Cebador inverso
 <220>
 50 <221> característica miscelánea
 <222> (28)..(34)
 <223> N puede ser cualquier nucleótido
 <220>
 55 <221> característica miscelánea
 <222> (35)..(35)
 <223> N puede ser cualquier nucleótido que represente una base universal
 <400> 7
 60 attcccgga agcttatcaa tcaatcannn nnnnn 35
 <210> 8
 <211> 36
 <212> DNA
 65 <213> secuencia artificial
 <220>

<223> Cebador directo
 <400> 8
 gagaggaatt cagactacaa ggacgacgac gacaag 36
 5
 <210> 9
 <211> 33
 <212> DNA
 <213> secuencia artificial
 10
 <220>
 <223> Cebador inverso
 <400> 9
 15 gagagaattc ccggaagct tatcaatcaa tca 33
 <210> 10
 <211> 17
 <212> DNA
 20 <213> secuencia artificial
 <220>
 <223> Cebador directo
 25 <220>
 <221> característica miscelánea
 <222> (10)..(16)
 <223> N puede ser cualquier nucleótido
 30 <220>
 <221> característica miscelánea
 <222> (17)..(17)
 <223> N puede ser cualquier nucleótido que represente una base universal
 35 <400> 10
 gagaattcan nnnnnnn 17
 <210> 11
 <211> 16
 40 <212> DNA
 <213> secuencia artificial
 <220>
 <223> Cebador inverso
 45 <220>
 <221> característica miscelánea
 <222> (9)..(15)
 <223> N puede ser cualquier nucleótido
 50 <220>
 <221> característica miscelánea
 <222> (16)..(16)
 <223> N puede ser cualquier nucleótido que represente una base universal
 55 <400> 11
 gagaattcnn nnnnnn 16
 <210> 12
 <211> 92
 60 <212> DNA
 <213> secuencia artificial
 <220>
 <223> Adaptador 1
 65 <400> 12

	aattcaatca atcacacaca ggaggccacc atggatgcat gtgtgcacgt gcacacatgc	60
	atccatgggtg gcctcctgtg tgtgattgat tg	92
5	<210> 13 <211> 90 <212> DNA <213> secuencia artificial	
10	<220> <223> Adaptador 2	
	<400> 13 aattcaatca atcacacaca ggaggccacc atggatgcat gtgtgcatgc acacatgcat	60
	ccatgggtggc ctcctgtgtg tgattgattg	90
15	<210> 14 <211> 88 <212> DNA <213> secuencia artificial	
20	<220> <223> Adaptador 3	
	<400> 14 aattcaatca atcacacaca ggaggccacc atggatgcat gtgtgcgcac acatgcatcc	60
25	atggtggcct cctgtgtgtg attgattg	88

REIVINDICACIONES

1. Una colección de presentación en fago que comprende una pluralidad de diferentes secuencias nucleotídicas de una pluralidad de organismos biodiversos que tienen cada uno un genoma secuenciado, y en la que los organismos son microorganismos y/o eucariotas que tienen genomas compactos.
- 5 2. La colección de presentación en fago según la reivindicación 1, en la que los fragmentos de secuencia de la colección son de organismos seleccionados del grupo consistente en *Fugu rubripes*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*; y/o de microorganismos procarióticos que se han caracterizado genéticamente, tales como *E. coli*, *Aquifex aelitus*, *Methanococcus jannaschii*, *Bacillus subtilis*, *Haemophilus influenzae*, *Helicobacter pylori*, *Neisseria meningitidis*, *Synechocystis* sp, *Bordetella pertussis*, *Pasteurella multocida*, *Pseudomonas aeruginosa*, *Borrelia burgdorferi*, *Methanobacterium thermoautotrophicum*, *Mycoplasma pneumoniae*, *Archaeoglobus fulgidus* y *Vibrio harveyi*.
- 10 3. La colección de presentación en fago según la reivindicación 1 o 2, construida usando fragmentos adaptados de ADN genómico agrupado de un panel evolutivamente diverso de genomas compactos.
- 15 4. La colección de presentación en fago según una cualquiera de las reivindicaciones 1 a 3, que comprende fragmentos de ADN de un panel biodiverso de microorganismos.
5. La colección de presentación en fago según una cualquiera de las reivindicaciones 1 a 3, que comprende fragmentos adaptados de ADN genómico agrupado de un conjunto evolutivamente diverso de genomas compactos.
6. La colección de presentación en fago según una cualquiera de las reivindicaciones 2 a 5, en la que los fragmentos de ácido nucleico de la colección comprenden de 90 pares de bases a 120 pares de bases de longitud.
- 20 7. La colección de presentación en fago según una cualquiera de las reivindicaciones 2 a 6, en la que los fragmentos de ácido nucleico de la colección son de suficiente longitud para codificar péptidos que comprenden aproximadamente 30 aminoácidos de longitud.
8. Un método de producción de una colección de presentación en fago que comprende una pluralidad de diferentes secuencias nucleotídicas de organismos biodiversos, comprendiendo dicho método producir fragmentos de secuencia nucleotídica a partir de una secuencia nucleotídica de composición nucleotídica conocida, en el que dicha secuencia nucleotídica es de un genoma secuenciado de un microorganismo y/o de un genoma compacto secuenciado de una especie eucariótica, y ligar los fragmentos en un vector de fago.
- 25 9. El método según la reivindicación 8, en el que el método comprende ligar los fragmentos en un sitio de clonación de un vector de fago T7-Select.
- 30 10. Un método para identificar un modulador o mediador de una actividad biológica, comprendiendo dicho método:
 - (i) producir una colección de presentación en fago derivada de fragmentos de secuencia nucleotídica definida según el método de la reivindicación 8 o 9, y
 - (ii) ensayar en al menos una secuencia aminoacídica derivada de la etapa (i) de la colección de presentación en fago una actividad biológica, en la que esa actividad es diferente de cualquier actividad que la secuencia aminoacídica pueda tener en su entorno nativo.
- 35