(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 609 102**

(51) Int. CI.:

C12N 15/10 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **25.06.2004 PCT/US2004/020306**

(87) Fecha y número de publicación internacional: **13.01.2005 WO05003345**

(96) Fecha de presentación y número de la solicitud europea: **25.06.2004 E 04756045 (3)**

(97) Fecha y número de publicación de la concesión europea: **02.11.2016 EP 1660655**

(54) Título: **Mutagénesis por revisión**

(30) Prioridad:

27.06.2003 US 483282 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
18.04.2017

(73) Titular/es:

BIOREN, LLC (100.0%)
235 East 42nd Street
New York, NY 10017-5755 , US

(72) Inventor/es:

CREA, ROBERTO

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

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

Mutagénesis por revisión

Solicitudes relacionadas e información

La presente solicitud reivindica prioridad a la solicitud provisional de Estados Unidos N.º: 60/483282, presentada el 27 de junio de 2003.

Antecedentes de la invención

La mutagénesis es una herramienta potente en el estudio de la estructura y función de proteínas. Las mutaciones se pueden realizar en la secuencia de nucleótidos de un gen clonado que codifica una proteína de interés y el gen modificado se puede expresar para producir mutantes de la proteína. Comparando las propiedades de una proteína de tipo silvestre y los mutantes generados, a menudo es posible identificar aminoácidos individuales o dominios de aminoácidos que son esenciales para la integridad estructural y/o función bioquímica de la proteína, tal como su unión y/o actividad catalítica. Sin embargo, el número de mutantes que se puede generar a partir de la proteína individual hace difícil seleccionar los mutantes que serán informativos o que tengan la propiedad deseada, incluso si los mutantes seleccionados que abarcan las mutaciones están únicamente en las regiones supuestamente importantes de una proteína (por ejemplo, regiones que conforman un sitio activo de una proteína). Por ejemplo, la sustitución, supresión, o inserción de un aminoácido particular puede tener un efecto local o global sobre la proteína.

Los procedimientos anteriores para mutagenizar polipéptidos han sido bien demasiado restrictivos, bien demasiado inclusivos, o bien limitados para inactivar la función de las proteínas en lugar de ganar o mejorar la función. Por ejemplo, un enfoque altamente restrictivo es mutagénesis selectiva o dirigida al sitio que se usa para identificar la presencia de un sitio funcional particular o para entender las consecuencias de realización de una alteración muy especificada dentro del sitio funcional. Una aplicación común de mutagénesis dirigida al sitio es en el estudio de las fosfoproteínas donde un resto aminoácido, que ordinariamente se fosforilaría y permitiría que el polipéptido llevara a cabo su función, se altera para confirmar la unión entre la fosforilación y la actividad funcional. Este enfoque es muy específico para el polipéptido y el residuo que se está estudiando. Recíprocamente, un enfoque altamente inclusivo es la saturación o mutagénesis al azar que se diseña para producir un gran número de mutaciones que abarcan todas las alteraciones posibles dentro de una región definida de un gen o proteína. Esto se basa en el principio de que, generando esencialmente todas las posibles variantes de un dominio proteico relevante, la disposición apropiada de aminoácidos es probable que se produzca como uno de los mutantes generados al azar. Sin embargo, en la práctica, el vasto número de combinaciones al azar de mutaciones generadas pueden impedir la capacidad de seleccionar de manera significativa un candidato deseado debido a la presencia del así llamado “ruido” de tantos candidatos indeseados.

Otro enfoque, denominado mutagénesis “de recorrido” (véanse, v.g., las patentes números 5.830.650; 5.798.208) se ha usado para mutagenizar una región definida de un polipéptido sintetizando una mezcla de oligonucleótidos degenerados que, estadísticamente, contienen un conjunto deseado de mutaciones. Sin embargo, debido a que se emplea la síntesis de polinucleótidos degenerados, la mutagénesis de recorrido proporciona un número de alteraciones no deseadas además del conjunto deseado de mutaciones. Por ejemplo, para introducir secuencialmente una mutación a través de una región definida de cinco posiciones de aminoácidos solamente, se puede preparar (y seleccionar) un conjunto de alrededor de 100 polinucleótidos (véase, por ejemplo, la figura 6). De acuerdo con ello, para preparar y seleccionar, por ejemplo, dos o tres regiones llega a ser crecientemente compleja, es decir, que requiere la preparación y selección de 200 a 300 polinucleótidos, respectivamente, para la presencia de 10 a 15 mutaciones solamente.

Todavía otro enfoque que se ha usado para mutagenizar proteínas es la mutagénesis de exploración de alanina, donde un residuo de alanina se “explora” a través de una porción de una proteína para identificar posiciones donde la función de proteínas está interrumpida. Sin embargo, este enfoque solamente mira la pérdida de la función de proteínas por medio de la sustitución de un resto alanina neutro en una posición dada, más que la ganancia o mejora de una función. De este modo, no es un enfoque útil para generar proteínas que tengan estructura y función mejoradas.

De acuerdo con ello, permanece una necesidad de una forma sistemática de mutagenizar una proteína para una función nueva o mejorada.

Sumario de la invención

La presente divulgación pertenece a un procedimiento de mutagénesis para la generación de proteínas (o polipéptidos) novedosas o mejoradas y a bibliotecas de análogos de polipéptidos y de polipéptidos específicos generados por los procedimientos. El polipéptido dirigido para mutagénesis puede ser un polipéptido natural, sintético o modificado por ingeniería genética, que incluye fragmentos, análogos y las formas mutantes de los mismos.

La invención también se refiere a los siguientes aspectos según se definen en el siguiente párrafo numerado:

1) Una biblioteca de polipéptidos de expresión de presentación de levaduras que comprende polinucleótidos que codifican análogos de fragmentos variables monocatenarios (scFv) de anticuerpos de serotipo B (BoNT/B) anti toxina nerviosa botulínica, en la que cada scFv comprende

(a) un engarce poli-Gly-Ser de SEQ ID NO: 7,

(b) una etiqueta poli-His de SEQ ID N.º: 8,

(c) una etiqueta myc de SEQ ID N.º: 9

(d) una VH de SEQ ID N.º: 10 y VL de SEQ ID N.º: 11 en las que los residuos de regiones determinantes de complementariedad (CDR) están sustituidas en cada posición de aminoácido dentro de las seis CDR, presentando los polinucleótidos colectivamente todas las posibles combinaciones de variantes de acuerdo con los siguientes criterios:

i) cada polinucleótido contiene en cada posición de codón en cada una de las CDR, bien un codón requerido para el residuo de aminoácido de tipo silvestre del polipéptido o bien un codón para el residuo de aminoácido predeterminado y

ii) cada polinucleótido contiene no más de un codón para el residuo de aminoácido predeterminado en cada CDR; en las que

el residuo de aminoácido predeterminado está seleccionado del grupo que consiste en alanina, leucina, serina, ácido aspártico, glutamina, lisina, histidina, tirosina y prolina y en el que el scFv anti-BoNT/B se expresa como una pareja de fusión con la proteína Aga2.

En una divulgación, el procedimiento comprende la introducción de un aminoácido predeterminado esencialmente en cada posición dentro de una región definida (o varias regiones diferentes) de la secuencia de aminoácidos de un polipéptido. Se genera una biblioteca de polipéptidos de manera que contenga análogos de polipéptidos que no tienen individualmente más de un aminoácido predeterminado, pero que colectivamente tienen el aminoácido predeterminado en cada posición dentro de la(s) región/regiones definida(s). El procedimiento se puede denominar mutagénesis “por revisión” debido a que, en efecto, un aminoácido individual, predeterminado (y solamente el aminoácido predeterminado) se sustituye posición por posición a lo largo de una o más región/regiones definida(s) de un polipéptido. De este modo, la divulgación permite a alguien “revisar” las consecuencias estructurales y funcionales de sustituir de manera separada un aminoácido predeterminado en cada posición de aminoácido dentro de una región definida del polipéptido, por lo tanto segregando una química de proteínas específica a la región definida sin ninguna interferencia o “ruido” de la generación de análogos de polipéptidos no deseados (es decir, análogos que contienen sustituciones de aminoácidos distintas de las que siguen al esquema “por revisión”) (véase por ejemplo la figura 6).

De acuerdo con ello, la presente divulgación permite la evaluación altamente eficaz y sistemática precisa del papel de un cambio de aminoácidos específico en una o más regiones definidas de un polipéptido. Esto llega a ser particularmente importante cuando se evalúan (mediante mutación) dos o más regiones definidas, de manera que el número de análogos de polipéptidos requeridos se incrementa en gran medida y de este modo, también se incrementa la presencia de análogos no deseados. La presente divulgación obvia este problema mediante la eliminación completa de análogos no deseados y de este modo, obvia el potencial de que cualesquiera cambios en la estructura o función de proteína observada sean el resultado de cualquier cosa excepto la sustitución del aminoácido predeterminado. De este modo, el efecto de la segregación de una química de proteínas específica a incluso regiones múltiples con una proteína se puede estudiar con alta precisión y eficacia. De manera importante, esto incluye el estudio de cómo la mutagénesis puede efectuar la interacción de tales regiones, por lo tanto mejorando la estructura y función globales de la proteína.

En una divulgación particular de la invención, la biblioteca de los análogos de polipéptidos se genera y selecciona sintetizando primero los polinucleótidos individuales que codifican una región o regiones de un polipéptido en las que, de manera colectiva, los polinucleótidos presentan todos los polinucleótidos variantes posibles de acuerdo con el criterio de revisión descrito en la presente memoria. Los polinucleótidos variantes se expresan, por ejemplo, usando transcripción y traducción *in vitro* y/o usando una tecnología de presentación, tal como presentación en ribosomas, presentación en fagos, presentación bacteriana, presentación en levaduras, presentación en serie o cualquier otro sistema de presentación adecuada conocido en la técnica.

Los polipéptidos expresados se rastrean y seleccionan después usando ensayos funcionales, tales como ensayos de unión o ensayos enzimáticos/catalíticos. En una divulgación, los polipéptidos se expresan en asociación con el polinucleótido que codifica el polipéptido, por lo tanto permitiendo la identificación de la secuencia de polinucleótidos que codifica el polipéptido. Todavía en otra divulgación, los polipéptidos se sintetizan directamente usando química de proteínas.

De este modo, la presente divulgación proporciona un procedimiento de mutagénesis que se puede usar para generar bibliotecas de análogos de polipéptidos que son de un tamaño práctico para la selección, en parte, debido a que las bibliotecas están desprovistas de cualesquiera polipéptidos análogos no deseados o también llamados ruido. El procedimiento se puede usar para estudiar el papel de aminoácidos específicos en la estructura y función de polipéptidos y para desarrollar polipéptidos nuevos o mejorados tales como anticuerpos, fragmentos de unión o análogos de los mismos, anticuerpos de cadenas individuales, anticuerpos catalíticos, enzimas y ligandos. Además, el procedimiento se puede realizar con el beneficio de una información *a priori*., v.g., mediante modelado por ordenador, que se puede usar para seleccionar un subconjunto inicial de análogos de polipéptidos a producir y estudiar usando mutagénesis “por revisión”.

Otras ventajas y aspectos de la presente invención serán fácilmente evidentes a partir de la descripción y ejemplos siguientes.

Breve descripción de las figuras

La figura 1 ilustra las regiones definidas de manera ejemplar (o regiones D) que se pueden examinar usando mutagénesis por revisión (LTM) y los ensayos funcionales para identificar análogos de polipéptidos deseados a partir de tales regiones D.

La figura 2 ilustra el uso de LTM en tres regiones definidas en una región variable de anticuerpo (es decir, CDR1, CDR2 y CDR3 de una región variable de cadena pesada de anticuerpo). La región variable de la cadena ligera se puede explorar de manera similar, bien sola o bien en combinación con la región variable de la cadena pesada. Por conveniencia en los ensayos de selección posteriores, la región variable de la cadena pesada se puede explorar en el contexto de un anticuerpo monocatenario (sFv) como se muestra.

La figura 3 ilustra el uso de LTM en una región definida (es decir, posiciones 31-25 de CDR1) de una región variable de la cadena pesada.

La figura 4 ilustra el uso de LTM en una región definida (es decir, posiciones 55-68 de CDR2) de una región variable de la cadena pesada.

La figura 5 ilustra el uso de LTM en una región definida (es decir, posiciones 101-111 de CDR3) de una región variable de la cadena pesada.

La figura 6 ilustra las ventajas de LTM comparada con la mutagénesis de recorrido. La LTM de una región representativa definida, es decir la CDR1 de una región variable de cadena pesada de anticuerpo, da como resultado la alteración secuencial de cada posición de aminoácido mediante la región definida, sin la introducción de ningún residuo de aminoácido no deseado o así llamado ruido.

La figura 7 ilustra la integración de tres regiones definidas de una proteína de nuevo en un contexto de proteínas general siguiendo la mutagénesis por revisión, en particular, la integración de las tres CDR de una región variable de la cadena pesada de anticuerpo en un formato de anticuerpo monocatenario siguiendo la mutagénesis por revisión.

La figura 8 ilustra el uso de la reacción en cadena de la polimerasa (PCR) para construir regiones definidas de una cadena pesada y ligera de anticuerpo sometida a LTM en un contexto de gen mayor.

La figura 9 ilustra de manera ejemplar la diversidad de formatos de cada una de las CDR en una región variable de anticuerpo para obtener un sitio catalítico que comprende, por ejemplo, una serina, histidina, y/o ácido aspártico y cómo se pueden disponer.

La figura 10 ilustra la integración de todas las seis CDR de una región de unión de un anticuerpo que se ha sometido a LTM y la diversidad resultante si se usa un residuo de aminoácido predeterminado o veinte residuos de aminoácidos predeterminados diferentes.

La figura 11 ilustra la integración de todas las seis CDR de un anticuerpo monocatenario anti-TNF (sFv) que se ha sometido a LTM y la diversidad resultante si se usa un residuo de aminoácido predeterminado o tres residuos de aminoácidos predeterminados diferentes.

La figura 12 ilustra una biblioteca ordenada que presenta alguno de los posibles análogos de polipéptidos de las seis CDR de una región de unión de anticuerpos que se puede lograr usando LTM.

La figura 13 ilustra la selección de una biblioteca de expresión ordenada usando presentación de ribosomas libres.

La figura 14 ilustra la química combinatoria explorada cuando la región de unión de una región variable de anticuerpo (es decir, todas las seis CDR) se somete a LTM.

La figura 15 muestra la secuencia de la región variable (en formato monocatenario) de varias moléculas de unión anti-TNF representativas que están sometidas a LTM.

La figura 16 muestra un esquema para llevar a cabo el ensayo de selección de proteasa para candidatos catalíticos de selección de LTM.

La figura 17 muestra un esquema que detalla el mecanismo (y ventajas) del ensayo de selección de proteasa cuando se lleva a cabo en células bacterianas.

La figura 18 muestra un diagrama de flujo que detalla la mecánica de bibliotecas génicas de selección para actividad catalítica, por ejemplo, actividad de anticuerpo catalítica, que usa presentación bien en ribosomas o bien en levaduras.

La figura 19 muestra un esquema para llevar a cabo LTM donde se pueden coordinar una información a priori (por ejemplo, información de modelado por ordenador) y empírica (resultados de ensayos) para diseño y desarrollo de moléculas más eficaz. Este enfoque guiado de LTM se denomina mutagénesis “a través de guía”.

La figura 20 muestra una secuencia de tipo silvestre de V_H CDR3 hipotética (óvalos sombreados en la parte superior) y las secuencias resultantes en los miembros de las bibliotecas de la sustitución His por LTM (en óvalos en blanco). Las sustituciones His por LTM individuales están codificadas por oligonucleótidos individuales, por ejemplo, oligonucleótidos sintetizados de una manera de alto rendimiento. Las bibliotecas de

subconjuntos de LTM para las otras sustituciones de aminoácidos (LTM) en este dominio de CDR se construyen de una manera similar.

La figura 21 ilustra la generación de bibliotecas scFv. En la línea de la parte superior del eje x y la columna alejada de la parte izquierda del eje y los tres dígitos representan las 3 CDR sobre cada una de las cadenas ligeras y pesadas. Un “0” indica una secuencia de CDR de tipo silvestre, mientras que un “1” indica una CDR mutada por LTM. El número sobre la cuadrícula indica la complejidad de la biblioteca subconjunto. Por ejemplo en el extremo superior izquierdo de la matriz está un “0” cuando los ejes x e y correspondientes son “000” y ”000” indicando que ninguna de las CDR tanto en V_H como en V_L están respectivamente mutadas por LTM. Al mover una fila desde el extremo “0” a la posición de la cuadrícula vecina “1” estaría designada por “100” en el eje x y “000” en el eje indicando que V_H CDR1 está mutada mientras que V_L CDR permanece de tipo silvestre. De este modo, de manera similar, una numeración de cuadrícula de “4” significaría que existen cuatro CDR mutadas simultáneamente. Inicialmente las siete cadenas V_H y las siete cadenas V_L (indicadas por las flechas) se preparan usando SOE-PCR. Las cadenas V_H y V_L se amplifican después y se mezclan y se emparejan mediante cebador mega para generar todas las combinaciones V_H-V_L restantes en una etapa.

Descripción detallada de la invención

Con el fin de proporcionar un entendimiento claro de la memoria descriptiva y las reivindicaciones, se proporcionan a continuación las siguientes definiciones.

Definiciones

Como se usa en la presente memoria el término “análogo” se refiere a un polipéptido variante o mutante (o a un ácido nucleico que codifica tal polipéptido) que tiene una o más sustituciones de aminoácidos.

El término “molécula de unión” se refiere a cualquier molécula de unión, incluyendo proteínas, polipéptidos, péptidos y pequeñas moléculas, que se unen a un sustrato o diana. En una realización de la invención, la molécula de unión es un anticuerpo o fragmento de unión del mismo (por ejemplo, un fragmento Fab), anticuerpo de dominio individual, anticuerpo monocatenario (por ejemplo, sFv), o péptido capaz de unirse a un ligando.

El término “región definida” se refiere a una región seleccionada de un polipéptido. Típicamente, la región definida incluye todo o una porción de un sitio funcional, por ejemplo, el sitio de unión de un ligando, el sitio de unión de una molécula de unión o receptor, o un sitio catalítico. La región definida también puede incluir porciones múltiples de un sitio funcional. Por ejemplo, la región definida puede incluir todas las porciones, una porción, o porciones múltiples de una región determinante de complementariedad (CDR) o una región variable (VR) de cadena pesada o ligera completa de un anticuerpo. De este modo, el sitio funcional puede incluir una región definida individual o regiones definidas múltiples que contribuye(n) a la actividad funcional de la molécula.

El término “biblioteca” se refiere a dos o más moléculas mutagenizadas de acuerdo con el procedimiento de la divulgación. Las moléculas de la biblioteca pueden estar en forma de polinucleótidos, polipéptidos, polinucleótidos y polipéptidos, polinucleótidos y polipéptidos en un extracto libre de células, o como polinucleótidos y/o polipéptidos en el contexto de un fago, células procariotas, o en células eucariotas.

El término “mutagenización” se refiere a la alteración de una secuencia de aminoácidos. Esto se puede lograr alterando o produciendo un ácido nucleico (polinucleótido) capaz de codificar la secuencia de aminoácidos alterados, o mediante la síntesis directa de un polipéptido alterado usando química de proteínas.

El término “polinucleótido(s)” se refiere a ácidos nucleicos tales como moléculas de ADN y moléculas de ARN y análogos de los mismos (v.g., ADN o ARN generado usando análogos de nucleótidos o usando química de ácidos nucleicos). Según se desee, los polinucleótidos se pueden preparar de manera sintética, v.g., usando química de ácidos nucleicos reconocidos en la técnica o enzimáticamente usando, por ejemplo, una polimerasa. Las modificaciones típicas incluyen mutilación, biotinilación y otras modificaciones conocidas en la técnica. Además, la molécula de ácido nucleico puede ser de cadena sencilla o de cadena doble y donde se desea, estar unida o asociada (por ejemplo, de manera covalente o no covalente) a un resto detectable.

El término “polinucleótido variante” se refiere a un polinucleótido que codifica un análogo de polipéptido correspondiente (o porción del mismo) de la divulgación. De este modo, los polinucleótidos variantes contienen uno o más codones que se han cambiado para dar como resultado la expresión de un aminoácido diferente. El término “polipéptido(s)” se refiere a dos o más aminoácidos unidos mediante un enlace peptídico, v.g., péptidos (por ejemplo, desde 2 hasta 50 residuos de aminoácidos), así como secuencias de péptidos más largos v.g., secuencias de proteínas que comprenden típicamente secuencias de aminoácidos desde tan pocos como 50 residuos de aminoácidos hasta más de 1000 residuos de aminoácidos.

El término “agrupación” se refiere a la combinación de variantes de polinucleótidos o análogos de polipéptidos para formar bibliotecas que presentan la mutagénesis por revisión de una región polipeptídica entera. Las moléculas pueden estar en la forma de un polinucleótido y/o polipéptido y puede coexistir en la forma de una sub-biblioteca, como moléculas en un soporte sólido, como moléculas en solución, y/o como moléculas en uno o más organismos (por ejemplo, fago, células procariotas, o células eucariotas). El término “aminoácido predeterminado” se refiere a un residuo de aminoácido seleccionado para sustitución en cada posición dentro de una región definida o polipéptido a mutagenizar. Esto no incluye posición/posiciones dentro de la región que ya (por ejemplo, naturalmente) contiene el aminoácido predeterminado y de este modo, que necesita no estar sustituida con el aminoácido predeterminado. De acuerdo con ello, cada análogo de polipéptido generado de acuerdo con la presente divulgación contiene no más de un residuo “de aminoácido predeterminado” en una región determinada dada. Sin embargo, de manera colectiva, la biblioteca de

análogos de polipéptidos generados contiene el aminoácido predeterminado en cada posición dentro de la región que se está mutagenizando. Típicamente, un aminoácido predeterminado se selecciona para un tamaño particular o química particular usualmente asociados con el grupo lateral del aminoácido. Los aminoácidos predeterminados adecuados incluyen, por ejemplo, glicina y alanina (estéricamente pequeños); serina, treonina y cisteína (nucleófilos); valina, leucina, isoleucina, metionina y prolina (hidrófobos); fenilalanina, tirosina y triptófano (aromáticos); aspartato y glutamato (ácidos); asparagina, glutamina, e histidina (amidas); y lisina y arginina (básicos).

Descripción detallada

El estudio de proteínas ha revelado que ciertos aminoácidos juegan un papel crucial en su estructura y función. Por ejemplo, parece que solamente un número de aminoácidos participan en la unión de un anticuerpo a un antígeno o están implicados en el evento catalítico de una enzima.

Aunque está claro que ciertos aminoácidos son críticos para la actividad o función de proteínas, es difícil identificar qué aminoácidos están implicados, cómo están implicados y qué sustituciones pueden mejorar la estructura o función de la proteína. En parte, esto se debe a la complejidad de la configuración espacial de cadenas laterales de aminoácidos en polipéptidos y la interrelación de porciones diferentes del polipéptido que contribuye a formar un sitio funcional. Por ejemplo, la interrelación entre las seis CDR de las regiones variables de las cadenas pesadas y ligeras de un anticuerpo contribuyen al bolsillo de unión a antígeno o ligando.

Los procedimientos de mutagénesis anteriores, tales como mutagénesis selectiva (dirigida al sitio) y mutagénesis por saturación, son de utilidad limitada para el estudio de estructura de proteínas y función en vista del número enorme de posibles variantes en polipéptidos complejos. Esto es especialmente cierto dado que combinaciones deseables están a menudo acompañadas por la presencia de vastas cantidades de combinaciones no deseables o así llamadas ruido.

El procedimiento de esta divulgación proporciona un enfoque sistemático, práctico y altamente exacto para evaluar el papel de los aminoácidos particulares y su posición, dentro de una región definida de un polipéptido, en la estructura o función del polipéptido y de este modo, para producir polipéptidos mejorados.

1. Seleccionar una región definida

De acuerdo con la presente invención, una región o regiones definidas dentro de una proteína se seleccionan por mutagénesis. Típicamente, las regiones se cree que son importantes para la estructura o función de las proteínas (véase, por ejemplo, la figura 1). Esto se puede deducir, por ejemplo, a partir de qué aspectos estructurales y funcionales se conocen o se pueden deducir de la comparación de la(s) región(es) definidas con lo que se conoce a partir del estudio de otras proteínas y se pueden ayudar mediante modelado de la información. Por ejemplo, la región definida puede ser una que tenga un papel en un sitio funcional, por ejemplo, en unión, catálisis u otra función. En una divulgación, la región definida es una región hipervariable o región determinante de complementariedad (CDR) de una molécula de unión a antígenos. En otra divulgación, la región definida es una porción de una región determinante de complementariedad (CDR). En otras divulgaciones, dos o más regiones definidas, por ejemplo, CDR o porciones de las mismas, se seleccionan para mutagénesis.

2. Seleccionar un residuo de aminoácido predeterminado

El residuo de aminoácido elegido para sustitución dentro de la(s) región/regiones definida(s) generalmente se selecciona entre los conocidos que están implicados en la estructura o función de interés. Los veinte aminoácidos de origen natural difieren con respecto a su cadena secundaria. Cada cadena secundaria es responsable de propiedades químicas que hacen cada aminoácido único. Para el propósito de alteración de la unión o de creación de nuevas afinidades de unión, se puede seleccionar en general cualquiera de los veinte aminoácidos de origen natural. De este modo, los procedimientos anteriores de mutagénesis, que crean vastos números de análogos para cada sustitución, no eran prácticos para evaluar el efecto sobre la unión de proteína de sustitución de cada uno de los veinte aminoácidos. Por el contrario, los procedimientos de la presente divulgación crean un número práctico de análogos para cada una de las sustituciones de los aminoácidos y de este modo, permiten la evaluación de una diversidad de químicas de proteínas dentro de una región o regiones de una proteína.

En contraste con la unión de proteínas, solamente un subconjunto de restos de aminoácidos participan típicamente en episodios enzimáticos no catalíticos. Por ejemplo, a partir de las propiedades químicas de las cadenas secundarias, solamente un número seleccionado de aminoácidos naturales participan preferentemente en eventos catalíticos. Estos aminoácidos pertenecen al grupo de aminoácidos polares y neutros tales como Ser, Thr, Gln, Tyr y Cys, el grupo de aminoácidos cargados, Asp y Glu, Lys y Arg y especialmente el aminoácido His. Otras cadenas secundarias polares y neutras son las de Cys, Ser, Thr, Asn, Gln y Tyr. Gly se considera también que es un miembro límite de este grupo. Ser y Thr juegan un papel importante en la formación de enlaces hidrógeno. Thr tiene una asimetría adicional en el carbono beta, por lo tanto solamente se usa uno de los estereoisómeros. El aminoácido Gln y Asn también pueden formar enlaces de hidrógeno, los grupos amido que funcionan como dadores de hidrógeno y los grupos carbonilo que funcionan como aceptores. Gln tiene un grupo CH_2 más que Asn lo que hace que el grupo polar sea más flexible y reduce su interacción con la cadena principal. Tyr tiene un grupo hidroxilo muy polar (OH fenólico) que se puede disociar a valores de pH altos. Tyr se comporta de alguna manera como una cadena secundaria cargada; sus enlaces hidrógeno son bastante fuertes.

Los ácidos polares neutros se encuentran en la superficie así como en el interior de las moléculas de las

proteínas. Como residuos internos, usualmente forman enlaces de hidrógeno entre sí o con la estructura principal del polipéptido. Cys puede formar puentes disulfuro. Histidina (His) tiene una cadena lateral aromática heterocíclica con un valor de pK de 6,0. En el intervalo de pH fisiológico, su anillo imidazol puede estar bien no cargado o bien cargado, después de recoger un ion hidrógeno de la solución. Dado que estos dos estados están fácilmente disponibles, His es bastante adecuado para catalizar reacciones químicas. Se encuentra en la mayoría de los centros activos de las enzimas, por ejemplo, serina proteasas.

Asp y Glu están cargados negativamente a pH fisiológico. Debido a su corta cadena secundaria, el grupo carboxilo de Asp en bastante rígido con respecto a la cadena principal. Esta puede ser la razón por la que el grupo carbonilo en muchos sitios catalíticos se proporciona mediante Asp y no mediante Glu. Los ácidos cargados se encuentran generalmente en la superficie de un polipéptido.

Además, Lys y Arg se encuentran en la superficie. Tienen cadenas secundarias largas y flexibles que presentan múltiples rotámeros de energías similares. En varios casos, Lys y Arg toman parte en la formación de puentes de sal interna o ayudan en catálisis. Debido a su exposición a la superficie del polipéptido, Lys es un residuo reconocido más frecuentemente por enzimas que bien modifican la cadena secundaria o bien escinden la cadena peptídica en el extremo carbonilo de los restos de Lys.

Mientras que la química de grupo secundario de un aminoácido puede guiar la selección de un residuo de aminoácido predeterminado, la carencia de una química de grupo secundario deseado puede ser un criterio para excluir un residuo de aminoácido para uso como el aminoácido predeterminado. Por ejemplo, los aminoácidos estéricamente pequeños y químicamente neutros, tales como alanina, se pueden excluir de la mutagénesis por revisión por carecer de una química deseada.

3. Sintetizar bibliotecas de análogos de polipéptidos

En una realización, una biblioteca de análogos de polipéptidos se genera para seleccionar mediante síntesis de oligonucleótidos individuales que codifican la región definida del polipéptido y no tienen más de un codón para el aminoácido predeterminado. Esto se lleva a cabo incorporando en cada posición de codón dentro del oligonucleótido bien el codón requerido para la síntesis de polipéptido de tipo silvestre o bien un codón para el aminoácido predeterminado. Esto se diferencia de los oligonucleótidos producidos en mutagénesis por saturación, mutagénesis al azar, o mutagénesis por revisión porque, para cada oligonucleótido, solamente se realiza una mutación, como opuesto a mutaciones múltiples.

Los oligonucleótidos se pueden producir de manera individual y después se pueden mezclar o combinar como se desee. Cuando el codón de la secuencia de tipo silvestre y el codón predeterminado para el aminoácido predeterminado son el mismo, no se realiza ninguna sustitución.

De acuerdo con ello, el número de posiciones de aminoácidos dentro de la región definida determinará el número máximo de oligonucleótidos preparados. Por ejemplo, si cinco posiciones de codón están alteradas con el aminoácido predeterminado, después se sintetizan cinco oligonucleótidos más un polinucleótido que presenta la secuencia de aminoácidos de tipo silvestre. Pueden estar alteradas simultáneamente dos o más regiones.

La mezcla de oligonucleótidos para la generación de la biblioteca se puede sintetizar fácilmente mediante procedimientos conocidos para la síntesis de ADN. El procedimiento preferido implica el uso de química de beta-cianoetil fosforamidita de fase sólida. Véase la patente de Estados Unidos N.º: 4.725.677. Por conveniencia, se puede usar un instrumento para la síntesis de ADN automatizada que contenga recipientes de reactivos específicos de nucleótidos. Los polinucleótidos también se pueden sintetizar para que contengan sitios de restricción o sitios de hibridación de cebadores para facilitar la introducción o ensamblaje de los polinucleótidos que presentan, v.g., una región definida en un contexto de genes mayor.

Los polinucleótidos sintetizados se pueden insertar en un contexto de genes más grande del polipéptido que se está mutagenizando usando técnicas de ingeniería genética convencional. Por ejemplo, los polinucleótidos se pueden preparar para que contengan sitios de reconocimiento flanqueantes para las enzimas de restricción. Véase, Crea, R., patente de estados Unidos N.º: 4.888.286. Los sitios de reconocimiento se diseñan para que correspondan a sitios de reconocimiento que bien existen de manera natural o bien se introducen en el gen próximo al ADN que codifica la región. Después de la conversión en la forma de cadena doble, los polinucleótidos se ligan en el gen mediante técnicas convencionales. Mediante un vector apropiado (que incluye, por ejemplo, vectores de fagos, plásmidos) los genes se pueden introducir en un extracto sin células, fagos, células procariotas, o células eucariotas adecuadas para la expresión de los polipéptidos mutantes.

En los casos donde se conoce la secuencia de aminoácidos del polipéptido a mutagenizar o donde se conoce la secuencia de ADN, la síntesis de genes es un posible enfoque. Por ejemplo, polinucleótidos solapados parcialmente, típicamente se pueden diseñar aproximadamente 20-60 nucleótidos de longitud. Los polinucleótidos internos se fosforilan después hibridados a su pareja complementaria para proporcionar una molécula de ADN bicatenaria con extensiones monocatenarias útiles para hibridación adicional. Los pares hibridados se pueden después mezclar juntos y ligarse para formar una molécula bicatenaria de longitud completa (véase, por ejemplo, la figura 8). Los sitios de restricción convenientes se pueden diseñar cerca de los extremos del gen sintético para la clonación en un vector adecuado. Las moléculas de longitud completa se pueden escindir con las enzimas de restricción y ligarse en un vector adecuado. Los sitios de restricción convenientes también se pueden incorporar en la secuencia del gen sintético para facilitar la introducción de módulos mutagénicos.

Como una alternativa para sintetizar polinucleótidos que presentan el gen bicatenario de longitud completa, los polinucleótidos que se superponen parcialmente en sus extremos 3' (es decir, con extremos 3' complementarios) se pueden ensamblar en una estructura abierta y después se pueden cargar con una polimerasa adecuada para preparar un gen bicatenario de longitud completa. Típicamente, los polinucleótidos de superposición son entre 40-90 nucleótidos de longitud. Los polinucleótidos extendidos se ligan después. Los sitios de restricción convenientes se pueden introducir en los extremos y/o internamente para propósitos de clonación. Después de la digestión con una enzima o enzimas de restricción apropiada(s), el fragmento de gen se liga en un vector adecuado. Como alternativa, el fragmento de gen puede estar ligado en el extremo romo en un vector apropiado.

En estos enfoques, si están disponibles sitios de restricción convenientes (de manera natural o modificados por ingeniería genética) después del ensamblaje de genes, los polinucleótidos degenerados se pueden introducir posteriormente mediante clonación del módulo en un vector apropiado. Como alternativa, los polinucleótidos degenerados se pueden incorporar en la fase de ensamblaje de genes. Por ejemplo, cuando ambas cadenas del gen se sintetizan completamente químicamente, se pueden producir polinucleótidos degenerados solapantes y complementarios. Los pares complementarios se hibridarán entre sí.

Cuando se usan polinucleótidos parcialmente solapantes en el ensamblaje de genes, un conjunto de nucleótidos degenerados también se puede incorporar directamente en lugar de uno de los polinucleótidos. La cadena complementaria apropiada se sintetiza durante la reacción de extensión a partir de un polinucleótido complementario parcialmente de la otra cadena mediante extensión enzimática con una polimerasa. La incorporación de los polinucleótidos degenerados en la fase de síntesis también simplifica la clonación cuando se mutageniza más de un dominio o región definida de un gen.

En otro enfoque, el gen de interés está presente en un plásmido de cadena sencilla. Por ejemplo, el gen se puede clonar en un vector de fago o en un vector con un origen de fago filamentoso de replicación que permite la propagación de moléculas de cadena sencilla con el uso de un fago auxiliar. El molde de cadena sencilla se puede hibridar con un conjunto de polinucleótidos degenerados que presentan las mutaciones deseadas y alargarse y ligarse, incorporando de esta manera cada cadena análoga en una población de moléculas que se pueden introducir en un huésped apropiado (Sayers, J. R. y col., Nucleic Acids Res. 16: 791-802 (1988). Este enfoque puede evitar muchas etapas de clonación donde los dominios múltiples se seleccionan para mutagénesis.

La metodología de reacción en cadena de la polimerasa (PCR) también se puede usar para incorporar polinucleótidos en un gen. Por ejemplo, los propios polinucleótidos se pueden usar como cebadores para extensión. En este enfoque, los polinucleótidos que codifican los módulos mutagénicos que corresponden a la región definida (o porción de la misma) son complementarios entre sí, al menos en parte y se pueden extender para formar un módulo de genes grande que usa una polimerasa, por ejemplo, usando amplificación de PCR.

El tamaño de la biblioteca variará dependiendo de la longitud y número de regiones y aminoácidos dentro de una región que se mutageniza. Preferiblemente, la biblioteca se diseña para que contenga menos de 10^{15}, 10^{14}, 10^{13}, 12^{12}, 10^{11}, 10^{10}, 10^{9}, 10^{8}, 10^{7} y más preferiblemente, 10^{6} análogos de polipéptidos o menos.

La descripción anterior se ha centrado en la mutagénesis de polipéptidos y bibliotecas de polipéptidos alterando el polinucleótido que codifica el polipéptido correspondiente. Sin embargo, se entiende que el alcance de la divulgación también abarca procedimientos de mutagenización de polipéptidos mediante la síntesis directa de los análogos de polipéptidos deseados usando química de proteínas. Cuando se lleva a cabo este enfoque, los polipéptidos resultantes todavía incorporan las características de la divulgación excepto que se elimina el uso de un polinucleótido.

Para las bibliotecas descritas anteriormente, bien en la forma de polinucleótidos y/o de polipéptidos correspondientes, se entiende que las bibliotecas también pueden estar unidas a un soporte sólido, tal como un microchip y preferiblemente ordenadas, usando técnicas reconocidas en la técnica.

4. Sistemas de expresión y selección

Las bibliotecas de polinucleótidos generadas mediante cualquiera de las técnicas anteriores u otras técnicas adecuadas se pueden expresar y seleccionar para identificar análogos de polipéptidos que tienen la estructura y/o la actividad deseada. La expresión de los análogos de polipéptidos se puede llevar a cabo usando cualquier sistema de presentación de expresión adecuado conocido en la técnica incluyendo, pero no limitado a, sistemas de presentación en extractos sin células (por ejemplo, presentación en ribosomas y sistemas de presentación ordenados (por ejemplo, microordenados o macroordenados), sistemas de presentación bacteriana, sistemas de presentación en fagos, células procariotas, y/o células eucariotas (por ejemplo, sistemas de presentación en levaduras).

En una divulgación, los polinucleótidos se modifican por ingeniería genética para que sirvan como moldes que se pueden expresar en un extracto sin células. Los vectores y extractos como se describen, por ejemplo, en las patentes de Estados Unidos N.^{os}: 5.324.637; 5.492.817; 5.665.563, se pueden usar y muchos están comercialmente disponibles. Se puede usar la presentación en ribosomas y otras técnicas sin células para la unión a un polinucleótido (es decir, un genotipo) a un polipéptido (es decir, un fenotipo) por ejemplo, Profusion™ (véanse, v.g., las patentes de Estados Unidos números 6.348.315; 621.804; 6.258.558; y 6.214.553).

Como alternativa, los polinucleótidos de la invención se pueden expresar en un sistema de expresión conveniente de E. coli, tal como el descrito por Pluckthum y Skerra. (Pluckthum, A. y Skerra, A., Meth.

Enzymol. 178: 476-515 (1989); Skerra, A. y col., Biotechnology 9: 273-178 (1991)). Las proteínas mutantes se pueden expresar para la secreción en el medio y/o en el citoplasma de la bacteria, como describen M. Better y A. Horwitz, Meth. Enzymol. 178: 476 (1989). En una realización, los dominios individuales que codifican VH y VL están cada uno unidos al extremo 3' de una secuencia que codifica una secuencia señal, tal como la secuencia señal ompA, phoA o pelB (Lei, S. P. y col., J. Bacterial. 169: 4379 (1987)). Estas fusiones de genes se ensamblan en una construcción dicistrónica, de manera que se pueden expresar a partir de un vector individual y secretarse en el espacio periplásmico de E. coli donde se replegarán y se pueden recuperar en una forma activa. (Skerra, A. y col., Biotechnology 9: 273-278 (1991)). Por ejemplo, los genes de cadena pesada de anticuerpos se pueden expresar simultáneamente con los genes de las cadenas ligeras de anticuerpos que producen anticuerpos o fragmentos de anticuerpos.

Todavía en otra realización de la invención, los polinucleótidos se pueden expresar en células eucariotas tales como levaduras, usando, por ejemplo, presentación en levaduras como se describe, v.g., en las patentes de Estados Unidos números 6.423.538; 6.331.391; y 6.300.065: En este enfoque, los análogos de polipéptidos de la biblioteca se fusionan a un polipéptido que se expresa y se presenta sobre la superficie de la levadura. También se pueden usar otras células eucariotas para la expresión de los polipéptidos de la divulgación tales como células de mamíferos, por ejemplo células de mieloma, células de hibridoma, o células de ovarios de hámster chino (CHO). Típicamente, los análogos de polipéptidos cuando se expresan en células de mamíferos se diseñan para que se expresen en el medio de cultivo, o para que se expresen sobre la superficie de tal célula. El anticuerpo o fragmento de anticuerpos se puede producir, por ejemplo, como moléculas de anticuerpo enteras o como fragmentos VH y VL individuales, fragmentos Fab, dominios simples, o como cadenas simples (sFv) (véase, Huston, J. S. y col., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988)).

La selección de los análogos de polipéptidos expresados (o polipéptidos producidos mediante síntesis directa) se pueden hacer mediante cualquier medio apropiado. Por ejemplo, la actividad de unión se puede evaluar mediante inmunoensayo convencional y/o cromatografía de afinidad y la actividad catalítica se puede determinar mediante ensayos adecuados para la conversión de sustratos. La selección de los análogos de polipéptidos de la divulgación para la función proteolítica se puede llevar a cabo usando un ensayo en placa de hemoglobina convencional como se describe, por ejemplo, en la patente de Estados Unidos N.º: 5.798.208.

5. *Mutagénesis por revisión asistida por modelado por ordenación*

La mutagénesis por revisión de la divulgación también se puede llevar a cabo con el beneficio de información estructural o de modelado concerniente a los análogos de polipéptidos a generarse, de manera que el potencial para generar análogos que tienen la función deseada mejorada se incrementa. La información estructural o de modelado se pueda también usar para guiar la selección de aminoácido predeterminado para introducir en las regiones definidas. Todavía adicionalmente, los resultados reales obtenidos con los análogos polipeptídicos de la invención pueden guiar la selección (o exclusión) de polipéptidos posteriores a prepararse y seleccionarse de una manera iterativa. De acuerdo con ello, la información estructural o de modelado se puede usar para generar subconjuntos iniciales de análogos de polipéptidos para uso en la divulgación, por lo tanto incrementando adicionalmente la eficacia de generación de polipéptidos mejorados.

En una divulgación particular, el modelado por ordenador se usa eliminar la producción de cualquier análogo de polipéptido que se predice que tiene una estructura y/o función pobre o indeseada. De esta forma, el número de análogos de polipéptidos a producirse se puede reducir notablemente incrementando por lo tanto la relación señal-ruido en ensayos de selección posteriores. En otra divulgación particular, el modelado por ordenador se actualiza continuamente con información modelado adicional, a partir de cualquier fuente relevante, por ejemplo, a partir de una secuencia de genes y de proteínas y de bases de datos de tres dimensiones y/o de resultados de los análogos ensayados previamente, de manera que la base de datos por ordenador se hace más precisa en su capacidad predictiva (figura 19).

Todavía en otra divulgación la base de datos por ordenador se proporciona con los resultados de ensayo de los análogos de polipéptidos previamente ensayados y clasifica los análogos, basándose en el criterio o criterios de ensayo, como efectores o no efectores, por ejemplo, como análogos de polipéptidos que se unen bien o no tan bien o que son enzimáticos/catalíticos o no tan enzimáticos/catalíticos. De esta forma la mutagénesis mediante guía de la divulgación puede igualar un intervalo de respuesta funcional con información estructural particular y usar tal información para guiar la producción de análogos de polipéptidos futuros a ensayarse. De acuerdo con ello, el procedimiento es especialmente adecuado para seleccionar anticuerpos o fragmentos de anticuerpos para una función particular, tal como afinidad de unión (por ejemplo, especificidad), estabilidad (por ejemplo, vida media) y/o función efectora (por ejemplo, activación de complementos y ADCC). De acuerdo con ello, la mutagénesis de residuos no contiguos dentro de una región puede ser deseable si se conoce, por ejemplo, por un modelado por ordenador que ciertos residuos en la región no participarán en la función deseada. La estructura coordinada y la interrelación espacial entre las regiones deseadas, v.g., los residuos de aminoácidos funcionales en las regiones definidas del polipéptido, por ejemplo, el/los aminoácido(s) predeterminado(s) que se han introducido, se pueden considerar y modelar. Tales criterios de modelado incluyen, por ejemplo, química de grupos secundarios de aminoácidos, distancias atómicas, datos de cristalografía, etc. De acuerdo con ello, el número de análogos polipeptídicos a producir se puede minimizar de manera inteligente.

En una divulgación preferida una o más de las etapas anteriores están asistidas por ordenador. El procedimiento también es capaz de llevarse a cabo, en parte o en conjunto, mediante un dispositivo, v.g., un dispositivo dirigido por ordenador. De acuerdo con ello, las instrucciones para llevar a cabo el procedimiento, en parte o en conjunto, se pueden conferir a un medio adecuado para uso en un dispositivo electrónico para llevar a cabo las instrucciones. En total, los procedimientos de la divulgación son capaces de un enfoque de alto rendimiento que comprende software (por ejemplo, instrucciones legibles de ordenador) y hardware (por

ejemplo, los ordenadores, robots y chips).

6. Explorar la química combinatoria de regiones múltiples definidas

La presente divulgación proporciona la importante ventaja de permitir la evaluación por mutagénesis de varias regiones o dominios diferentes de un polipéptido simultáneamente. Esto se puede hacer usando el mismo aminoácido o un aminoácido diferente predeterminado dentro de cada región, permitiendo la evaluación de sustituciones de aminoácidos en regiones relacionadas de manera conformacional, de forma que las regiones tras el plegamiento del polipéptido, están asociadas para preparar un sitio funcional (por ejemplo, el sitio de unión de un anticuerpo o el sitio catalítico de una enzima). Esto, a su vez, proporciona una forma eficaz de crear sitios funcionales nuevos o mejorados.

Por ejemplo, como se muestra en la figura 14, las seis CDR de un anticuerpo que preparan los aspectos únicos del sitio de unión de antígeno (región Fv), se puede mutagenizar simultáneamente, o separadamente en las cadenas VH o VL, para estudiar las interrelaciones tridimensionales de aminoácidos seleccionados en este sitio. En una divulgación, la química combinatoria de las tres o más regiones definidas se explora de manera sistemática usando mutagénesis por revisión y preferiblemente seis regiones definidas, por ejemplo las seis CDR de una región variable de la cadena pesada y la cadena ligera de anticuerpo. Para realizar mutagénesis por revisión sobre una CDR, típicamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, o más posiciones de aminoácidos se alteran.

De acuerdo con ello, la presente divulgación abre nuevas posibilidades para el diseño de tipos muy diferentes de polipéptidos nuevos y mejorados. El procedimiento se puede usar para mejorar una estructura o función existente de una proteína. Por ejemplo, se puede introducir un sitio de unión para un anticuerpo o fragmento de anticuerpo o afinidad para un antígeno preexistente, función efectora y/o estabilidad mejorada. Como alternativa, la introducción de aminoácidos “catalíticamente importantes” adicionales en un dominio catalítico de una enzima se puede realizar dando como resultado una actividad modificada o potenciada hacia un sustrato. Como alternativa, las estructuras, las especificidades o actividades enteramente nuevas se pueden introducir en un polipéptido. La síntesis de novo de actividad enzimática se puede lograr también. Las nuevas estructuras se pueden edificar sobre el “armazón” natural o de consenso de una proteína existente mediante mutación de regiones relevantes solamente mediante el procedimiento de la invención.

7. Mutagénesis por revisión para preparar anticuerpos nuevos o mejorados

El procedimiento de esta divulgación se usa específicamente para modificar moléculas de anticuerpos. Como se usa en la presente memoria, las moléculas de anticuerpos o anticuerpos se refieren a anticuerpos o porciones de los mismos, tales como anticuerpos de longitud completa, moléculas Fv, u otros fragmentos de anticuerpos, cadenas individuales o fragmentos de los mismos (por ejemplo, una cadena individual de Fv), anticuerpos monocatenarios y anticuerpos quiméricos. Se pueden introducir alteraciones en la región variable y/o en la región marco conservada (constante) de un anticuerpo. La modificación de la región variable puede producir anticuerpos con mejores propiedades de unión de antígeno y si se desea, con propiedades catalíticas. La modificación de la región marco también puede conducir a la mejora de las propiedades físico-químicas, tales como solubilidad o estabilidad (por ejemplo, vida media), que son especialmente útiles, por ejemplo, en producción comercial, biodisponibilidad, función efectora (por ejemplo, activación de complementos y/o ADCC) y afinidad de unión (por ejemplo, especificidad) para el antígeno. Típicamente, la mutagénesis se dirigirá a la región Fv de la molécula de anticuerpo, es decir, la estructura responsable de la actividad de unión a antígeno que se compone de regiones variables de dos cadenas, una de la cadena pesada (VH) y una de la cadena ligera (VL). Una vez que las características de unión de antígeno se identifican, la(s) región/regiones variable(s) se pueden modificar por ingeniaría genética en una clase de anticuerpo apropiada tal como IgG, IgM, IgA, IgD, o IgE.

8. Mutagénesis por revisión para preparar/mejorar polipéptidos catalíticos/enzimáticos

El procedimiento de la invención también es particularmente adecuado para el diseño de proteínas catalíticas, particularmente anticuerpos catalíticos. Actualmente, los anticuerpos catalíticos se pueden preparar mediante una adaptación de técnicas de fusión de células somáticas convencionales. En este procedimiento, un animal se inmuniza con un antígeno que se parece al estado de transición del sustrato deseado para inducir la producción de un anticuerpo que une el estado de transición y cataliza la reacción. Las células que producen anticuerpos se recogen a partir del animal y se fusionan con una célula de inmortalización para producir células híbridas. Después las células se seleccionan para la secreción de un anticuerpo que cataliza la reacción. Este procedimiento depende de la disponibilidad de análogos del estado de transición de un sustrato. El procedimiento puede estar limitado debido a que tales análogos es probable que sean difíciles de identificar o de sintetizar en la mayoría de los casos.

El procedimiento de la invención proporciona un enfoque diferente que elimina la necesidad de un análogo de estado de transición. Mediante el procedimiento de la divulgación, un anticuerpo se puede hacer catalítico mediante la introducción de aminoácidos adecuados en el sitio de unión de una inmunoglobulina (región Fv). La región del sitio de unión de antígeno (Fv) se compone de seis bucles hipervariables (CDR), tres derivados de la cadena pesada de inmunoglobulina (H) y tres de la cadena ligera (L), que conecta cadenas beta con cada subunidad. Los residuos de aminoácidos de los bucles CDR contribuyen casi enteramente a las características de cada anticuerpo monoclonal específico. Por ejemplo, las triadas catalíticas (constituidas por residuos de aminoácidos serina, histidina y ácido aspártico) modeladas después de serina proteasas se pueden crear en los segmentos hipervariables de la región Fv de un anticuerpo con afinidad conocida para la molécula de sustrato y se pueden seleccionar para actividad proteolítica del sustrato.

En particular, el procedimiento de la divulgación se puede usar para producir muchas enzimas diferentes o

anticuerpos catalíticos, incluyendo oxidorreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas. Entre estas clases, de particular importancia será la producción de proteasas, carbohidrasas, lipasas, dioxigenasas y peroxidasas mejoradas. Estas y otras enzimas que se pueden preparar mediante el procedimiento de la divulgación tienen aplicaciones comerciales importantes para las conversiones enzimáticas en asistencia sanitaria, cosméticos, alimentos, elaboración de cerveza, detergentes, medio ambiente (por ejemplo, tratamiento de aguas residuales), agricultura, curtido, textiles y otros procedimientos químicos. Estos incluyen, pero no se limitan a, aplicaciones diagnósticas y terapéuticas, conversiones de grasas, carbohidratos y proteína, degradación de contaminantes orgánicos y síntesis de productos químicos. Por ejemplo, las proteasas terapéuticamente eficaces con actividad fibrinolítica, o actividad contra estructuras virales necesarias para infectividad, tales como proteínas de revestimiento viral, se pueden modificar por ingeniería genética. Tales proteasas podrían ser agentes anti-tromboticos útiles o agentes anti-virales útiles contra virus tales como, por ejemplo, VIH, rinovirus, influenza, o hepatitis. En el caso de oxigenasas (por ejemplo, dioxigenasas), una clase de enzimas que requieren un co-factor para oxidación de anillos aromáticos y otros dobles enlaces, aplicaciones industriales en procesos de formación de biopulpa, conversión de biomasas en combustibles u otros agentes químicos, conversión de contaminantes de aguas residuales, bioprocesamiento de carbón y detoxificación de compuestos orgánicos peligrosos son posibles aplicaciones de proteínas novedosas.

La presente invención se ilustra adicionalmente en los siguientes ejemplos, que no se deben interpretar como limitantes.

Ejemplificación

A lo largo de los ejemplos, se usaron los siguientes materiales y procedimientos salvo que se especifique otra cosa.

Materiales y procedimientos

En general, la práctica de la invención emplea, salvo que se indique otra cosa, técnicas convencionales de química, biología molecular, tecnología de ADN recombinante, tecnología de PCR, inmunología (especialmente, por ejemplo, tecnología de anticuerpos), sistemas de expresión (por ejemplo, expresión sin células, presentación en fagos, presentación en ribosomas y Profusion™) y cualquier cultivo de células necesario que están dentro de la habilidad en la técnica y se explican en la bibliografía. Véanse, por ejemplo, Sambrook, Fritsch y Maniatis, *Molecular Cloning: Cold Spring Harbour Laboratory Press* (1989): *DNA Cloning*, Volúmenes 1 y 2, (D. N. Glover, Ed. 1985); *Oligonucleotides Synthesis* (M. J. Gait, Ed. 1984); *PCR Handbook Current Protocols in Nucleic Acid Chemistry*, Beaucage, Ed. John Wiley y Sons (1999) (Editor); *Oxford Handbook of Nucleic Acid Structure,* Neidle, Ed., Oxford Univ. Press (1999); *PCR Protocols: A Guide to Methods and Applications*, Innis y col., Academic Press (1990); *PCR Essential Techniques: Essential Techniques,* Burke, Ed., John Wiley y Sons Ltd (1996); *The PCR Technique: RT-PCR*, Siebert, Ed, Eaton Pub. Co. (1998); *Antibody Engineering Protocols (Methods in Molecular Biology),* 510, Paul, S., Humana Pr. (1996); *Antibody Engineering: A Practical Approach (Practical Approach Series, 169),* McCafferty, Ed., Irl Pr (1996); *Antibodies; A Laboratory Manual,* Harlow y col., C. S. H. L. Press, Pub. (1999): *Current Protocols in Molecular Biology*, eds., Ausubel y col., John Wiley y Sons (1992); *Large-Scale Mammalian Cell Culture Technology*, Lubieniecki, A., Ed., Marcel Dekker, Pub., (1990). *Phage Display: A Laboratory Manual*, C. Barbas (Ed.), CSHL Press, (2001); *Antibody Phage Display,* P. O'Brien (Ed.), Humana Press (2001); Border y col., Yeast surface display for screening combinatorial polypeptide libraries, *Nature Biotechnology,* 15 (6): 553-7 (1997; Border y col.; Yeast surface display for directed evolution of protein expression, affinity, and stability, *Methods Enzymol.*, 328: 430-44 (2000); presentación en ribosomas como describen Pluchthum y col., en la Patente de Estados Unidos N.º: 6.348.315 y Profusion™ como describen Szostak y col., Patente de Estados Unidos Números 6.258.558; 6.261.804; y 6.214.553.

Ejemplo 1

MUTAGÉNESIS POR REVISIÓN DE LAS TRES REGIONES DEFINIDAS EN UNA MOLÉCULA DE UNIÓN A ANTÍGENO

En este ejemplo, se describe la mutagénesis por revisión de tres CDR de un anticuerpo para mejorar la unión y proteolisis de un sustrato.

En particular, se realiza la mutagénesis "por revisión" de tres regiones determinantes de complementariedad (CDR) de un anticuerpo monoclonal. CDR1, CDR2 y CDR3 de la región variable de la cadena pesada (VH) son regiones definidas seleccionadas para la mutagénesis por revisión, Para esta divulgación, los aminoácidos predeterminados seleccionados son los tres restos de la triada catalítica de las serina proteasas, Asp, His y Ser. Asp se selecciona para VH CDR1, His se selecciona para VL CDR2 y Ser se selecciona para VH CDR3. La selección de estos tres aminoácidos predeterminados permite el uso de un ensayo de proteasa conveniente con el fin de detectar cuando los tres residuos están posicionados correctamente para mostrar una actividad funcional, es decir, proteolisis de un sustrato de ensayo.

Un anticuerpo ejemplar, MCPC 603, se reconoce como un buen modelo para investigar la unión y catálisis debido a que la región de unión de anticuerpo se ha caracterizado bien. La secuencia de aminoácidos y la secuencia de ADN del MCPC 603 VH y las regiones VL están públicamente disponibles (véanse, por ejemplo, Rudikoff, S. y Potter, M., Biochemistry 13: 4033 (1974); Plukthun, A. y col., Cold Spring Harbor Symp. Quant. Biol., Vol. LII: 105-112 (1987)). Las CDR para el anticuerpo MCPC 603 se han identificado como se muestra en la figura 2. En la cadena pesada, CDR1 abarca las posiciones de los residuos de aminoácidos 31-35, CDR2 abarca las posiciones 50-69 y CDR3 abarca las posiciones 101-111. En la cadena ligera, los residuos de aminoácidos de CDR1 son 24-40, CDR2 abarca los aminoácidos 55-62 y CDR3 abarca los aminoácidos

95-103.

El diseño de los oligonucleótidos para la mutagénesis por revisión en las CDR de MCPC 603 es tal que los análogos de polipéptidos resultan tener la secuencia de aminoácidos para CDR1, CDR2 y CDR3 como se muestra en las figuras 3-5. Se entiende que los oligonucleótidos sintetizados pueden ser mayores que las regiones CDR para alterar para facilitar la inserción en la construcción diana como se muestra en la figura 7. Un formato de anticuerpo monocatenario se elige por conveniencia en la expresión posterior y etapas de selección. Los oligonucleótidos se pueden convertir en cadenas bicatenarias mediante técnicas enzimáticas (véase, por ejemplo, Oliphant, A. R. y col., 1986, al principio) y después se ligan dentro de un plásmido restringido como se muestra en la figura 8. Los sitios de restricción pueden ser sitios de origen natural o sitios de restricción modificados por ingeniería genética.

Los polinucleótidos que codifican los análogos de polipéptidos se pueden expresar en cualesquiera de sistemas de expresión convenientes descritos en la presente memoria y se pueden seleccionar usando el ensayo de la serina proteasa descrito, por ejemplo, en la patente de Estados Unidos N.º: 5.798.208 (véanse también, las figuras 16-17). En resumen, los análogos de polipéptidos expresados se exponen a un sustrato de ensayo y se examinan para proteolisis del sustrato de ensayo. La cantidad de proteolisis, revelada por una zona de eliminación del sustrato, indica un análogo de polipéptido que tiene la actividad funcional deseada.

Ejemplo 2

MUTAGÉNESIS POR REVISIÓN DE SEIS REGIONES DEFINIDAS EN UNA MOLÉCULA DE UNIÓN A ANTÍGENOS

En este ejemplo, se describe la mutagénesis por revisión de los seis CDR de un anticuerpo para mejorar la unión y proteolisis de un sustrato. En particular, se realiza una mutagénesis “por revisión” de todas las seis regiones hipervariables o regiones determinantes de complementariedad (CDR) del anticuerpo de modelo mencionado anteriormente (MCPC 603), en este ejemplo, la mutagénesis por “revisión” se lleva a cabo a partir de dos a tres veces con un aminoácido diferente en una región o dominio dado. Por ejemplo, Asp, Ser e His se recorren secuencialmente en las cadenas pesadas y ligeras como se muestra en la figura 10.

La mutagénesis de los residuos continuos dentro de una región puede ser deseable si se conoce, o si se puede deducir, que ciertos residuos en la región no participarán en la función deseada. Además, el número de análogos se puede minimizar. Otras consideraciones en la selección del aminoácido predeterminado y las posiciones particulares a alterar son que los residuos pueden ser capaces de hidrogenar enlaces entre sí. Esta consideración puede imponer una restricción de proximidad sobre las variantes deseadas. De este modo, solamente ciertas posiciones dentro de las CDR pueden permitir que los aminoácidos de la triada catalítica interaccionen de manera apropiada. De este modo, el modelo molecular u otra información estructural se puede usar para enriquecer variantes funcionales.

En este caso, la información estructural conocida se usó para identificar restos en las regiones que pueden estar suficientemente cerca para permitir la formación de enlaces de hidrógeno entre Asp, His y Ser, así como el intervalo de residuos a mutagenizar. Roberts y col., han identificado regiones de estrecho contacto entre porciones de las CDR (Roberts, V. A. y col., Proc. Natl. Acad. Sci. USA 87: 6654-6658 (1990)). Esta información junto con los datos de la estructura de rayos x de MCPC 603 se usa para seleccionar áreas prometedoras de estrecho contacto entre las CDR dirigidas para mutagénesis. Este tipo de mutagénesis por revisión guiada por información estructural/de modelo se puede denominar como mutagénesis “a través de guía”.

La mutagénesis por revisión se lleva a cabo como se ilustra en la figura 10 en la que cada CDR se somete a un aminoácido predeterminado dando como resultado 8×10^5 y si se exploran todos los veinte aminoácidos, 5×10^{13} análogos de polipéptidos. Los polinucleótidos se pueden expresar en cualesquiera de los sistemas de expresión convenientes descritos en la presente memoria y se pueden seleccionar usando la serina proteasa descrita, por ejemplo, en la patente de Estados Unidos N.º: 5.798.208. En resumen, los análogos de polipéptidos expresados se exponen a un sustrato de ensayo y se examinan para evaluar proteolisis del sustrato de ensayo. La cantidad de proteolisis, revelada mediante una zona de eliminación del sustrato, indica un análogo de polipéptidos que tiene la actividad funcional deseada.

Ejemplo 3

MUTAGÉNESIS POR REVISIÓN DE MOLÉCULAS DE UNIÓN A ANTI-TNF PARA MEJORAR LA FUNCIÓN

En este ejemplo, se describe la mutagénesis por revisión de un anticuerpo anti-TNF para mejorar la unión.

En particular, se realiza la mutagénesis “por revisión” de todas las seis regiones hipervariables o regiones determinantes de complementariedad (CDR) de dos anticuerpos anti-TNF diferentes. Los anticuerpos anti-TNF tienen aplicación general en el tratamiento de enfermedad inmune en pacientes que tienen niveles inapropiados del ligando TNF (factor de necrosis tumoral). Existen dos anticuerpos anti-TNF disponibles comercialmente. Por conveniencia en la realización de mutagénesis por revisión y posterior selección, las regiones variables de cadena ligera y pesada (véanse las SEQ ID números 2-4) de estos anticuerpos se convirtieron a un formato monocatenario usando un engarce poli Gly-Ser (véase la figura 15). Las regiones definidas seleccionadas por mutagénesis por revisión con un aminoácido predeterminado se identifican mediante la presencia de una barra negra como se muestra en la figura 15. Estas regiones definidas corresponden a las CDR de los anticuerpos de las cadenas individuales.

Los polinucleótidos que presentan las seis regiones CDR y regiones flanqueantes suficientes para permitir el

ensamblaje de los polipéptidos en la secuencia monocatenaria completa mostrada en la figura 15 se sintetizan como se describe en la presente memoria. Los residuos de aminoácidos predeterminados se seleccionan para cada región CDR y separadamente y secuencialmente se introducen en cada posición de aminoácido a través de las seis regiones CDR. Los polinucleótidos se modifican por ingeniería genética adicionalmente para servir como moldes capaces de soportar la transcripción de un transcrito de ARN correspondiente que se puede después traducir en un polipéptido usando presentación en ribosomas.

Los polinucleótidos que codifican los análogos de polinucleótidos correspondientes se expresan usando una transcripción sin células y extractos de traducción. Las transcripciones de ARN están unidas covalentemente a un resto detectable tal como un resto fluorescente. Como alternativa, la secuencia de interés se puede fusionar en marco a un resto fluorescente tal como proteína fluorescente verde (GFP) para permitir la detección conveniente de expresión y normalización de aglutinantes frente a no aglutinantes. Preferiblemente, los polinucleótidos que codifican los análogos de polipéptidos están al menos parcialmente ordenados, v.g., expresados en un pocillo que tiene múltiples análogos de polipéptidos pero se pueden fácilmente des-replegar. De acuerdo con ello, cada pocillo contiene un subconjunto de análogos de polipéptidos unidos ahora al transcrito correspondiente unido a un resto fluorescente usando presentación en ribosomas. El pocillo se sonda con ligando diana, es decir TNF y se ensaya para análogos de polipéptidos que se unen y con los que la afinidad de unión se compara al polipéptido de tipo silvestre. Los análogos de polipéptidos que se unen mejor que los polipéptidos de tipo silvestre se vuelven a modificar después por ingeniería genética en un formato de anticuerpo IgG de longitud completa para realización de ensayo paralelo con los análogos de anticuerpos comerciales para unión mejorada usando técnicas convencionales.

Ejemplo 4

MUTAGÉNESIS POR REVISIÓN DE ANTICUERPOS CONTRA EL SEROTIPO B DE TOXINA NERVIOSA BOTULÍNICA (BoNT/B) Y EL SEROTIPO A DE TOXINA NERVIOSA BOTULÍNICA (BoNT/A) PARA MEJORAR LA FUNCIÓN

En este ejemplo, anticuerpos mejorados contra el serotipo B de toxina nerviosa botulínica (BoNT/B) y el serotipo A de toxina nerviosa botulínica (BoNT/A) se generan usando mutagénesis por revisión (LTM) para mejorar la función. El enfoque LTM se basa en la creación de mutaciones sencillas por CDR a través del bolsillo de unión, basándose en un subconjunto (el conjunto de LTM) de los 20 aminoácidos que exploran las características de tamaño, carga, hidrofobicidad y unión por hidrógeno. Los criterios para seleccionar los aminoácidos en el conjunto de LTM se describen más adelante.

1. Purificación de anticuerpos y secuenciación de genes y diseño de cadenas individuales (scFv)

Los fragmentos de anticuerpos murinos (Fab) con afinidad de unión a BoNT/B se pueden obtener como se describe en Emaneul y col., (1996) *Journal of Immunological Methods* 193: 189-197. Se pueden usar los anticuerpos *BotFab 5* (SEQ ID N.^{os}: 10 y 11, polipéptidos de cadena ligera y pesada, respectivamente), el *BotFab 20* (SEQ ID ID N.^{os}: 12 y 13, polipéptidos de cadena ligera y pesada, respectivamente) o el *BotFab 22* (SEQ ID ID N.^{os}: 12 y 13, polipéptidos de cadena ligera y pesada, respectivamente). Las secuencias foráneas se describen en la patente de Estados Unidos N.º: 5.932.449. El antígeno BoNT/B (toxina de longitud completa, cadena ligera, y/o cadena pesada) se puede obtener a partir de Metabiologics, Inc., WI.

Los anticuerpos anti-BoNT/A descritos en Pless y col., (2001) *Infect. Immun.* 69: 570 se pueden usar con el objetivo de mejorar sus afinidades mediante al menos un orden de magnitud. El antígeno de dominio de unión de cadena pesada BoNT/A (BoNT/A He) se puede obtener a partir de Metabiologics, Inc., WI.

Los fragmentos V_L y V_H del/de los anticuerpo(s) se clonan y se secuencian usando técnicas de biología molecular convencionales. Las regiones variables de las moléculas se amplifican mediante la reacción en cadena de la polimerasa (PCR) y se unen con un engarce poli-Gly-Ser (típicamente SGGGGSGGGGSGGGGS (SEQ ID N.º: 7)) para generar anticuerpos monocatenarios (csFv). Una etiqueta poli-His (HHHHHH (SEQ ID N.º: 8)) y una etiqueta myc (EQKLISEEDL (SEQ ID N.º: 9)) también se anexan al extremo C de los genes para facilitar la purificación y detección. Estas moléculas se muestran en cualquiera de las tecnologías bien conocidas (por ejemplo, tecnologías basadas en levaduras, bacterias o fagos) y se ensayan para evaluar su capacidad de unirse a BoNT/B o BoNT/A.

La versión monovalente de scFv de anticuerpos completos proporciona un buen formato para acometer estudios de mutagénesis y se conoce para reproducir generalmente el mecanismo de unión de la molécula completa, con la excepción de los efectos de avidez mostrados por moléculas multivalentes.

2. Mejora de anticuerpos usando mutagénesis por revisión (LTM) y diseño de genes

Para la mutagénesis por revisión (LTM), se eligen los siguientes nueve aminoácidos y sus características funcionales representativas: alanina y leucina (alifáticos), serina (grupo hidroxilo), ácido aspártico (ácido) y glutamina (amida), lisina e histidina (básicos), tirosina (aromático), prolina (hidrófobo). Estos aminoácidos muestran diversidad química adecuada en tamaño, carga, hidrofobicidad y unión por hidrógeno para proporcionar información inicial significativa sobre la funcionalidad química necesaria para mejorar las propiedades de anticuerpos. Las elecciones también se basan en la frecuencia de aparición de estos aminoácidos en las CDR de los anticuerpos. Por ejemplo, la dada entre tirosina y fenilalanina para representar aminoácidos con cadenas secundarias aromáticas, la primera se elige debido a su preponderancia significativamente más alta en las CDR de anticuerpos y a su capacidad para enlace de hidrógeno. LTM se emplea inicialmente para identificar aminoácidos específicos y propiedades químicas que son beneficiosas para unión, neutralización y/o cualesquiera propiedades adicionales deseadas en el

anticuerpo final. Sin embargo, LTM no se limita a estos nueve aminoácidos o a nueve aminoácidos totales. El análisis de LTM se puede realizar con cualquier combinación de aminoácidos y el subconjunto de LTM puede ser tan alto como 18 aminoácidos (1 corto de mutagénesis por saturación).

2.1 Síntesis de oligonucleótidos de LTM

En el análisis primario de LTM, el objetivo es explorar cada contribución de la cadena de aminoácidos a la afinidad de unión global dentro de cada CDR. Para generar eficazmente diversidad significativa, cada uno de los aminoácidos del subconjunto de LTM se dirige en cada posición individual dentro de la secuencia de CDR con una sustitución solamente por CDR. De este modo, cada oligonucleótido individual codifica solamente una mutación de CDR individual. Por ejemplo, con el fin de realizar la mutagénesis de histidina de LTM sobre un dominio V_H CDR3 de once aminoácidos hipotético, se sintetizan 11 oligonucleótidos que codifican para todas las 11 mutaciones de histidina individuales posibles (véase la figura 20). Un análisis tal ensaya los efectos de tener una amida voluminosa en cada posición en la CDR. Por lo tanto para generar una biblioteca de V_H CDR3 LTM, solamente se sintetizan 99 oligonucleótidos para el análisis de LTM (9 aminoácidos de LTM x 11 posiciones de V_H CDR3). Las secuencias de oligonucleótidos se ensayan para codones de parada accidentales, duplicación de tipo silvestre, uso de codón ineficaz, horquillas, bucles y otras estructuras secundarias que usan software disponible públicamente.

2.2 Síntesis de oligonucleótidos degenerados para las combinaciones de mutaciones de LTM beneficiosas

Al emplear el reemplazo de LTM sistemático de aminoácidos individuales dentro de un CDR, la preferencia de las funcionalidades químicas en cada posición está sin cubrir. Con el fin de combinar todas las mutaciones de las selecciones de LTM y explorar la posible aditividad y sinergia energética entre estas, se sintetizan los oligonucleótidos degenerados que codifican estas mutaciones y la secuencia de tipo silvestre. La combinación degenerada de oligonucleótidos con $1{,}6 \times 10^4$ variantes se usa posteriormente para generar una biblioteca de segunda generación que explora la naturaleza aditiva de estas sustituciones.

2.3 Diseño de oligonucleótidos asistido por ordenador, biblioteca y base de datos de los resultados.

El software acoplado con sintetizadores de ADN realizados a medida automatizados permite la síntesis rápida de oligonucleótidos. La primera etapa implica decidir qué aminoácidos diana se incorporarán en las CDR. El software determina la preferencia de codón (por ejemplo, preferencia de codón de levadura, bacteriano o de fago, dependiendo del sistema elegido) necesario para introducir los aminoácidos diana y también elimina cualquier duplicación de la secuencia de tipo silvestre que se puede generar mediante este procedimiento de diseño. Después analiza para evaluar las estructuras de los codones de parada potenciales, horquillas y otros bucles o para secuencias problemáticas que se corrigen posteriormente antes de la síntesis. El plan de diseño de LTM completado se envía después al sintetizador de ADN, que realiza una síntesis automática de los oligonucleótidos. De esta manera, los oligonucleótidos necesarios para crear las bibliotecas se pueden generar rápidamente.

Una base de datos electrónica puede almacenar información de todas las secuencias de oligonucleótidos de LTM, detalles de las sustituciones de las CDR de las bibliotecas de scFv y resultados de los ensayos de unión para un antígeno diana. El archivo de los datos de los oligonucleótidos permite las iteraciones racionales de las estrategias del diseño y la reutilización de los oligonucleótidos de reactivos.

3. Estrategia de LTM global

LTM se emplea inicialmente para identificar aminoácidos específicos y propiedades químicas que son beneficiosas para propiedades de unión, neutralización y/o cualesquiera propiedades adicionales deseadas en el anticuerpo final. También rápidamente identifica las regiones que no toleran ninguna mutagénesis sin pérdida significativa de afinidad (o cualquier propiedad física que se seleccione). Por lo tanto, no solamente es el análisis de LTM una buena metodología para explorar los requerimientos químicos en cada posición en todas las CDR de un anticuerpo, sino que también determina rápidamente los aminoácidos absolutamente requeridos para la unión a anticuerpo. Después de la identificación de las mutaciones beneficiosas mediante LTM, los esquemas de mutagénesis combinatoria se pueden emplear para que en primer lugar incorporen todas estas mutaciones de aminoácidos dispares para generar CDR múltiplemente mutadas. Adicionalmente o como alternativa, la mutagénesis “de recorrido” (WTM) se puede usar para sondar los efectos de mutaciones múltiples del mismo aminoácido en una CDR (como se describe, por ejemplo, en las patentes de Estados Unidos números 5.830.650; 5.798.208).

4. Bibliotecas LTM scFv

Las técnicas de LTM descritas anteriormente se usan para crear agrupaciones de oligonucleótidos con mutaciones en una CDR individual de la cadena ligera o pesada. Estos oligonucleótidos se sintetizan para que incluyan algunos de los marcos circundantes para facilitar el solapamiento y la hibridación durante PCR. Estas combinaciones de oligonucleótidos se utilizan para generar todas las posibles cadenas V_L y VH en las que existen mutaciones en CDR individuales, dobles y triples (combinaciones individuales, dobles y triples de CDR1, 2 y 3) usando PCR de extensión de solapamiento individual (SOE-PCR) (como se describe en Horton y col., (1989) *Gene* 77: 61-68). SOE-PCR es un procedimiento rápido y sencillo para combinar los fragmentos de ADN que no requieren sitios de restricción, endonucleasas de restricción, o ligasa de ADN. En dos regiones SOE-PCR en el gen se amplifican primero mediante PCR usando cebadores diseñados de manera que los productos de PCR compartan una secuencia complementaria en un extremo. En las condiciones de PCR, las secuencias complementarias se hibridan, formando un solapamiento. Las secuencias complementarias actúan después como cebadores, permitiendo la extensión mediante la ADN polimerasa para producir una molécula recombinante.

Por ejemplo, para crear la combinación de las cadenas V_H en las que tanto CDR-H1 como CDR-H2 están mutadas y la CDR-H3 es de tipo silvestre, que se llama “110” (1 significa una CDR mutante y 0 significa una CDR de tipo silvestre), los genes mutantes de CDR-H1 se usan como moldes y SOE-PCR se lleva a cabo para unir los oligonucleótidos de CDR-H2 para generar la combinación doblemente mutada (figura 21). Considerando que cada CDR puede ser bien de tipo silvestre o bien mutante, existen siete posibles combinaciones (mostradas mediante flechas en la figura 21) para cada una de las agrupaciones de las cadenas V_L y V_H (no incluyendo la molécula de tipo silvestre “000”). La combinación de las siete agrupaciones de V_L y de las ocho agrupaciones de V_H crea 63 combinaciones V_L-V_H de tipo no silvestre (scFv), (figura 21). Cada una de las 64 combinaciones V_L-V_H (incluyendo la secuencia de tipo silvestre) se denomina un “subconjunto” del agrupamiento de bibliotecas de LTM scFv entero. Un agrupamiento de bibliotecas de scFv se crea para cada aminoácido seleccionado para sustitución. El número de secuencias de aminoácidos presentadas dentro de cada biblioteca de subconjunto depende de la longitud de la CDR, la secuencia de aminoácidos dentro de la CDR y la estrategia de diseño de oligonucleótidos de LTM.

5. Revisión de bibliotecas de alto rendimiento y selección de anticuerpos mejorados

Está disponible una diversidad de procedimientos para la expresión y presentación de anticuerpos. Estos incluyen bacteriófago, *Escherichia coli* y levadura. Aunque cada uno de estos procedimientos se ha usado para mejora de anticuerpos, el sistema de presentación en levaduras permite varias ventajas (Poder y Wittrup (1997) *Nat. Biotechnol.* 15: 553-557). La levadura puede acomodar fácilmente tamaños de bibliotecas de hasta 10^7, con 10^3-10^5 copias de cada anticuerpo mostrándose en cada superficie celular. Las células de levaduras se pueden seleccionar fácilmente y separarse fácilmente usando citometría de flujo y separación de células activadas por fluorescencia (FACS) o perlas magnéticas. Las levaduras también permiten selección rápida y re-crecimiento rápido. El sistema de secreción eucariótico y las rutas de glucosilación de levadura permiten un subconjunto mucho mayor de moléculas scFv para plegarse y mostrarse correctamente sobre la superficie celular que los sistemas de presentación procarióticos. La presentación en levaduras acoplada con la evolución dirigida se ha usado para incrementar la K_D de un fragmento de anticuerpo scFv para fluoresceína a 48 fM, unión dos órdenes de magnitud más fuerte que cualquier ligando monovalente reseñado previamente (Boder y col., (2000) *PNAS* 97: 10701-10705). El sistema de presentación utiliza el receptor de adhesión de levaduras de aglutinina-a para presentar proteínas sobre la superficie celular. Las proteínas de interés, en el caso presente, bibliotecas de anti-BoNT/B scFv LTM o bibliotecas de anti-BoNT/A scFv LTM, se expresan como parejas de fusión con la proteína Aga2. Estas proteínas de fusión se seleccionan a partir de la célula y llegan a unirse por puentes disulfuro a la proteína Aga1, que se une a la pared de las células de levaduras (véase Invitrogen, bibliografía del producto de presentación de levaduras pYD1). Además, existen marcas carboxi terminales incluidas que se pueden utilizar para controlar los niveles de expresión y/o normalizar mediciones de afinidad de unión.

Las perlas magnéticas revestidas con estreptavidina (Spherotech) se usan para rastrear y seleccionar anticuerpos que se unen a BoNT/B o BoNT/A con alta afinidad. Esta metodología emplea unión de anticuerpos de alta afinidad a antígeno biotinilado que se une después a perlas revestidas con estreptavidina con el fin de seleccionar los clones de levaduras (Yeung y Wittrup (2002) *Biotechnol. Prog* 18: 212-220) y Feldhaus y col., (2003) *Nature Biotech.* 21: 163-170). El polipéptido BoNT/A o BoNT/A (Metabiologics) se biotinila usando protocolos convencionales (Pierce) y la selección se realiza usando procedimientos bien conocidos en la técnica. Usando selección basada en equilibrio y cinética, los anticuerpos con afinidades mejoradas se seleccionan a partir de estas bibliotecas. La eficacia de cada ronda de selección se controla mediante FACS analítica (FACScan). Además, las afinidades de unión relativa de las moléculas individuales presentadas sobre la superficie de levaduras se miden mediante valoración del antígeno. Esto permite una rápida identificación de moléculas con afinidad mejorada. Los clones de scFv se secuencian después para identificar mutaciones beneficiosas.

6. Generación de anticuerpos solubles para mediciones de afinidad BIAcore

Los anticuerpos de interés se subclonan en sistemas de expresión solubles (*Picchia pastoris* y/o E. coli) y se genera proteína soluble. Existen varios vectores comercialmente disponibles y líneas celulares para expresión de anticuerpos solubles, que incluyen los de Invitrogen (por ejemplo, pPIC9 para *P. pastoris*) y Novagen (pET20b para expresión periplásmica en *E. coli*). Estos sistemas se usan de manera rutinaria para generar anticuerpos monocatenarios o de longitud completa solubles. El sistema de expresión de *P. pastoris* (Invitrogen) produce de manera rutinaria 1-5 mg por litro de scFv purificado soluble. La purificación de proteínas está facilitada por la presencia de cola de polihistidina en el extremo C de la molécula, en el caso de cadenas individuales o mediante columnas de proteína A o proteína G para anticuerpos de longitud completa. Los anticuerpos monocatenarios y de longitud completa solubles se generan para obtener mediciones de velocidad cinética de afinidad BIAcore. Esta etapa es necesaria ya que moléculas de scFv de alta afinidad sobre la superficie celular de clones de levaduras se debe verificar como una molécula soluble libre de células. Los anticuerpos solubles monocatenarios y de longitud completa generados de la manera anterior se puede usar para obtener mediciones de afinidad BIAcore, así como en el ensayo de crecimiento hacia fuera de neuritas o en el ensayo de letalidad de ratón descrito más adelante.

7. Selección de anticuerpos seleccionados para neutralización *in vivo* o *in vitro*

Un ensayo basado en células que usa el crecimiento hacia fuera de neuritas de neuronas primaras de pollo como un indicador de intoxicación de BoNT y de esta manera, como un medio para cuantificar la neutralización de toxinas se puede usar para seleccionar los anticuerpos seleccionados. Los experimentos preliminares que usan cultivos de explantes de ganglios de la raíz dorsal de huevos de pollos fertilizados han indicado que existía menos crecimiento hacia fuera axonal de explantes tratados con BoNT/A que de los controles. Como alternativa, se puede usar un ensayo de unión neuromuscular de ganglio ciliar-músculo del iris de pollo (como se describe en Lomneth y col., (1990) *Neuroscience Letters* 113: 211-216).

El ensayo de letalidad de ratón (MLA, tal como describen, por ejemplo, Schantz y Kautter (1978) *J. Assoc. Off. Anal. Chem.* 61: 96-99) es otro procedimiento bien conocido y aceptado *in* vivo para ensayar la neutralización de BoNT. El ensayo implica la inyección interperitoneal de aproximadamente 0,5 ml de preparaciones de muestra de BoNT/B o BoNT/A con y sin el anticuerpo en 20 a 30 gramos, ratones blancos de la cepa ICR. La mortandad debida a la insuficiencia respiratoria se observa durante 1-4 días. La cuantificación de la neutralización requiere diluciones en serie de los MAb con niveles que varían de DL_{50} de BoNT/B o BoNT/A de ratón. La MLA se puede usar para determinar la capacidad de neutralización de los anticuerpos optimizados identificados mediante la selección *in vitro*.

La capacidad de neutralización de anticuerpos también se puede medir *in vitro* mediante el ensayo de protección de ratón (MPA, Goeschel y col., (1997) *Exp. Neurol.* 147: 96-102). En la MPA el nervio frénico izquierdo, junto con el hemidiafragma izquierdo, se escinde del ratón.

El nervio frénico se electroestimula continuamente en un baño de tejido. Los anticuerpos purificados se incuban con BoNT/B o BoNT/A y se añaden al baño de tejido. La parálisis inducida por toxina se define como un 50 % de reducción en la sacudida muscular inicial.

La invención descrita en la presente memoria está abarcada por las siguientes reivindicaciones.

LISTADO DE SECUENCIAS

<110> R. Crea & Co., et al.

<120> MUTAGÉNESIS POR REVISIÓN

<130> BRI-001PC

<150> 60/483,282
<151> 2003-06-27

<160> 13

<170> FastSEQ para Windows Versión 4.0

<210> 1
<211> 121
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 1
Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Val Lys Pro Gly Ser
1 5 10 15
Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp Tyr
20 25 30
Asn Val Asp Trp Val Lys Gln Ala Pro Gly Gln Gly Leu Gln Trp Ile
35 40 45
Gly Asn Ile Asn Pro Asn Asn Gly Gly Thr Ile Tyr Asn Gln Lys Phe
50 55 60
Lys Gly Lys Gly Thr Leu Thr Val Asp Lys Ser Thr Ser Thr Ala Tyr
65 70 75 80
Met Glu Leu Ser Ser Leu Thr Ser Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95
Ala Arg Ser Ala Phe Tyr Asn Asn Tyr Glu Tyr Phe Asp Val Trp Gly
100 105 110
Gln Gly Thr Thr Val Thr Val Ser Ser
115 120

<210> 2
<211> 119
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 2
Glu Val Lys Leu Glu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15
Ser Met Lys Leu Ser Cys Val Ala Ser Gly Phe Ile Phe Ser Asn His
20 25 30
Trp Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val
35 40 45
Ala Glu Ile Arg Ser Lys Ser Ile Asn Ser Ala Thr His Tyr Ala Glu
50 55 60

```
Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ala
65                  70                  75                  80

Val Tyr Leu Gln Met Thr Asp Leu Arg Thr Glu Asp Thr Gly Val Tyr
                85                  90                  95
Tyr Cys Ser Arg Asn Tyr Tyr Gly Ser Thr Tyr Asp Tyr Trp Gly Gln
            100                 105                 110
Gly Thr Thr Leu Thr Val Ser
        115
```

<210> 3
<211> 114
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

```
<400> 3
Asp Ile Met Met Thr Gln Ser Pro Ser Thr Leu Ser Ala Ser Val Gly
 1               5                  10                  15
Asp Arg Val Thr Ile Thr Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
            20                  25                  30
Asn Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
        35                  40                  45
Ala Pro Lys Leu Leu Ile Ser Trp Ala Ser Thr Arg Glu Ser Gly Val
    50                  55                  60
Pro Ser Arg Phe Ile Gly Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr
65                  70                  75                  80
Ile Ser Ser Leu Gln Pro Asp Asp Val Ala Thr Tyr Tyr Cys Gln Gln
                85                  90                  95
Tyr Tyr Asp Tyr Pro Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile
            100                 105                 110
Lys Arg
```

<210> 4
<211> 107
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

```
<400> 4
Asp Ile Leu Leu Thr Gln Ser Pro Ala Ile Leu Ser Val Ser Pro Gly
 1               5                  10                  15
Glu Arg Val Ser Phe Ser Cys Arg Ala Ser Gln Phe Val Gly Ser Ser
            20                  25                  30
Ile His Trp Tyr Gln Gln Arg Thr Asn Gly Ser Pro Arg Leu Leu Ile
        35                  40                  45
Lys Tyr Ala Ser Glu Ser Met Ser Gly Ile Pro Ser Arg Phe Ser Gly
    50                  55                  60
Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Asn Thr Val Glu Ser
65                  70                  75                  80
Glu Asp Ile Ala Asp Tyr Tyr Cys Gln Gln Ser His Ser Trp Pro Phe
                85                  90                  95
Thr Phe Gly Ser Gly Thr Asn Leu Glu Val Lys
            100                 105
```

<210> 5
<211> 16
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 5
Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 6
<211> 6
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 6
His His His His His His
1 5

<210> 7
<211> 10
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 7
Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 8
<211> 236
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 8
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
20 25 30
Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
35 40 45
Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
50 55 60
Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
65 70 75 80
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
85 90 95

```
Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
            100                 105                 110
Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
        115                 120                 125
Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
    130                 135                 140
Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
145                 150                 155                 160
Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
                165                 170                 175
Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
            180                 185                 190
Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
        195                 200                 205
Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
    210                 215                 220
Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
225                 230                 235
```

<210> 9
<211> 254
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

```
<400> 9
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1               5                  10                  15
Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
            20                  25                  30
Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
        35                  40                  45
Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
    50                  55                  60
Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
65                  70                  75                  80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
                85                  90                  95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
            100                 105                 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
        115                 120                 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
    130                 135                 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145                 150                 155                 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
                165                 170                 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
            180                 185                 190
Pro Ala Val Leu Gln Phe Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
        195                 200                 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
    210                 215                 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225                 230                 235                 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His His
                245                 250
```

<210> 10
<211> 236
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 10

Met	Lys	Tyr	Leu	Leu	Pro	Thr	Ala	Ala	Ala	Gly	Leu	Leu	Leu	Leu	Ala
1				5				10						15	
Ala	Gln	Pro	Ala	Met	Ala	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ala	Ser
			20					25					30		
Leu	Ser	Ala	Ser	Val	Gly	Glu	Thr	Val	Thr	Ile	Thr	Cys	Arg	Ala	Ser
		35					40					45			
Gly	Asn	Ile	His	Asn	Tyr	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Gln	Gly	Lys
	50					55					60				
Ser	Pro	Gln	Leu	Leu	Val	Tyr	Asn	Ala	Lys	Thr	Leu	Ala	Asp	Gly	Val
65					70					75					80
Pro	Ser	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Gln	Tyr	Ser	Leu	Lys
				85					90					95	
Ile	Asn	Ser	Leu	Gln	Pro	Glu	Asp	Phe	Gly	Ser	Tyr	Tyr	Cys	Gln	His
			100					105					110		
Phe	Trp	Ser	Thr	Pro	Trp	Thr	Phe	Gly	Gly	Gly	Thr	Lys	Leu	Glu	Ile
		115					120					125			
Lys	Arg	Ala	Asp	Ala	Ala	Pro	Thr	Val	Ser	Ile	Phe	Pro	Pro	Ser	Ser
	130					135					140				
Glu	Gln	Leu	Thr	Ser	Gly	Gly	Ala	Ser	Val	Val	Cys	Phe	Leu	Asn	Asn
145					150					155					160
Phe	Tyr	Pro	Lys	Asp	Ile	Asn	Val	Lys	Trp	Lys	Ile	Asp	Gly	Ser	Glu
				165					170					175	
Arg	Gln	Asn	Gly	Val	Leu	Asn	Ser	Trp	Thr	Asp	Gln	Asp	Ser	Lys	Asp
			180					185					190		
Ser	Thr	Tyr	Ser	Met	Ser	Ser	Thr	Leu	Thr	Leu	Thr	Lys	Asp	Glu	Tyr
		195					200					205			
Glu	Arg	His	Asn	Ser	Tyr	Thr	Cys	Glu	Ala	Thr	His	Lys	Thr	Ser	Thr
	210					215					220				
Ser	Pro	Ile	Val	Lys	Ser	Phe	Asn	Arg	Asn	Glu	Cys				
225					230					235					

<210> 11
<211> 254
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 11

Met	Lys	Tyr	Leu	Leu	Pro	Thr	Ala	Ala	Val	Gly	Leu	Leu	Leu	Leu	Ala
1				5				10						15	
Ala	Gln	Pro	Ala	Met	Ala	Glu	Val	Gln	Leu	Gln	Gln	Ser	Gly	Ala	Glu
			20					25					30		
Leu	Val	Lys	Pro	Gly	Ala	Ser	Val	Lys	Leu	Ser	Cys	Thr	Ala	Ser	Gly
		35					40					45			
Phe	Asn	Ile	Lys	Asp	Thr	Phe	Met	His	Trp	Val	Lys	Gln	Arg	Pro	Glu
	50					55					60				

```
Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
65                  70                  75                  80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
                85                  90                  95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
            100                 105                 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
        115                 120                 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
    130                 135                 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145                 150                 155                 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
                165                 170                 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
            180                 185                 190
Pro Ala Val Leu Gln Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
        195                 200                 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
    210                 215                 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225                 230                 235                 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His His
                245                 250
```

<210> 12
<211> 236
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

```
<400> 12
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1               5                  10                  15
Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
            20                  25                  30
Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
        35                  40                  45
Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
    50                  55                  60
Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
65                  70                  75                  80
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
                85                  90                  95
Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
            100                 105                 110
Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
        115                 120                 125
Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
    130                 135                 140
Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
145                 150                 155                 160
Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
                165                 170                 175
Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
            180                 185                 190
Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
        195                 200                 205
```

```
Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
    210                 215                 220
Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
225                 230                 235
```

<210> 13
<211> 254
<212> PRT
<213> Secuencia artificial

<400> 13

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1               5                  10                  15
Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
            20                  25                  30
Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
        35                  40                  45
Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
    50                  55                  60
Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
65                  70                  75                  80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
                85                  90                  95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
            100                 105                 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
        115                 120                 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
    130                 135                 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145                 150                 155                 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
                165                 170                 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
            180                 185                 190
Pro Ala Val Leu Gln Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
        195                 200                 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
    210                 215                 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225                 230                 235                 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His His
                245                 250
```

REIVINDICACIONES

1. Una biblioteca de polipéptidos de expresión de presentación de levaduras que comprende polinucleótidos que codifican análogos de fragmentos variables monocatenarios (scFv) de anticuerpos de serotipo B (BoNT/B) anti toxina nerviosa botulínica, en la que cada scFv comprende

(a) un engarce poli-Gly-Ser de SEQ ID NO: 7,

(b) una etiqueta poli-His de SEQ ID N.º: 8,

(c) una etiqueta myc de SEQ ID N.º: 9

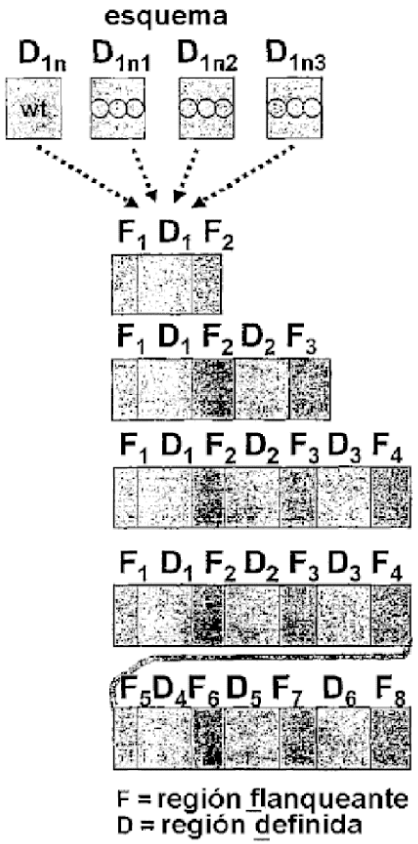
(d) una VH de SEQ ID N.º: 10 y VL de SEQ ID N.º: 11 en la que los residuos de regiones determinantes de complementariedad (CDR) están sustituidas en cada posición de aminoácido dentro de las seis CDR, presentando los polinucleótidos colectivamente todas las posibles combinaciones de variantes de acuerdo con los siguientes criterios:

i) cada polinucleótido contiene en cada posición de codón en cada una de las CDR, bien un codón requerido para el residuo de aminoácido de tipo silvestre del polipéptido o bien un codón para el residuo de aminoácido predeterminado y

ii) cada polinucleótido contiene no más de un codón para el residuo de aminoácido predeterminado en cada CDR; en la que

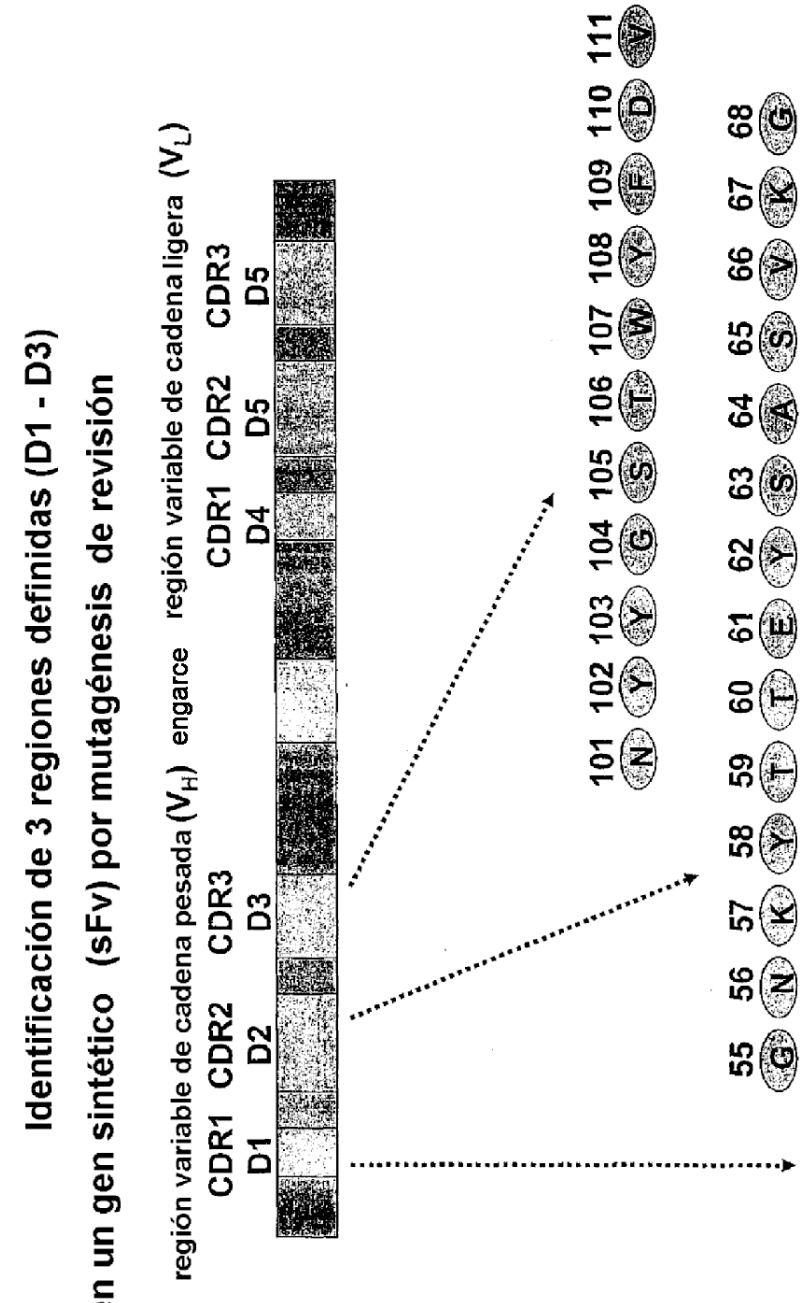
el residuo de aminoácido predeterminado está seleccionado del grupo que consiste en alanina, leucina, serina, ácido aspártico, glutamina, lisina, histidina, tirosina y prolina y en el que el scFv anti-BoNT/B se expresa como una pareja de fusión con la proteína Aga2.

Formatos de biblioteca de mutagénesis por revisión (LTM)



características estructurales (ejemplos)	actividad funcional (ensayos)
librería de región D tipo silvestre (wt) y alteraciones secuenciales	
región 1 D (ligando)	unión a receptor (unión o ensayo de respuesta celular)
regiones 2 D (ligando o sitio catalítico)	unión a receptor o catálisis (unión, respuesta celular, o ensayo de catálisis)
regiones 3 D (VL, VL, dominio único o sitio catalítico)	unión a receptor o catálisis (unión, respuesta celular, o ensayo de catálisis)
regiones > 3 D (sFv, anticuerpo (v. g., IgG), o sitio catalítico)	unión a receptor o catálisis (unión, respuesta celular, o ensayo de catálisis)

ES 2 609 102 T3

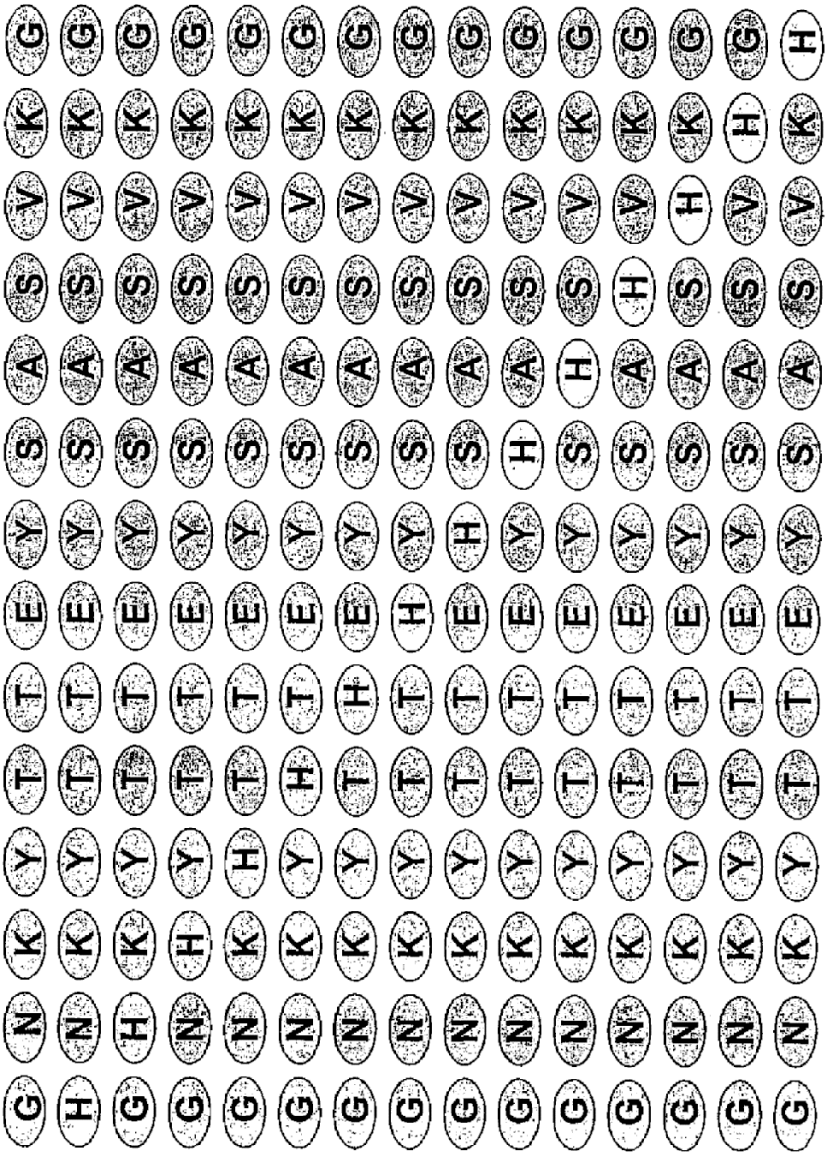

Identificación de 3 regiones definidas (D1 - D3)
en un gen sintético (sFv) por mutagénesis de revisión
región variable de cadena pesada (V_H) engarce región variable de cadena ligera (V_L)
CDR1 CDR2 CDR3 CDR1 CDR2 CDR3
D1 D2 D3 D4 D5 D5
31 32 33 34 35
D F Y M G
55 56 57 58 59 60 61 62 63 64 65 66 67 68
G N K Y T T E Y S A S V K G
101 102 103 104 105 106 107 108 109 110 111
N Y Y G S T W Y F D V

LTM secuencial de una primera región definida (D1)
posiciones de aminoácidos 31-35 de V_HCDR1

31	32	33	34	35
D	F	Y	M	G
D	D	Y	M	G
D	F	D	M	G
D	F	Y	D	G
D	F	Y	M	D

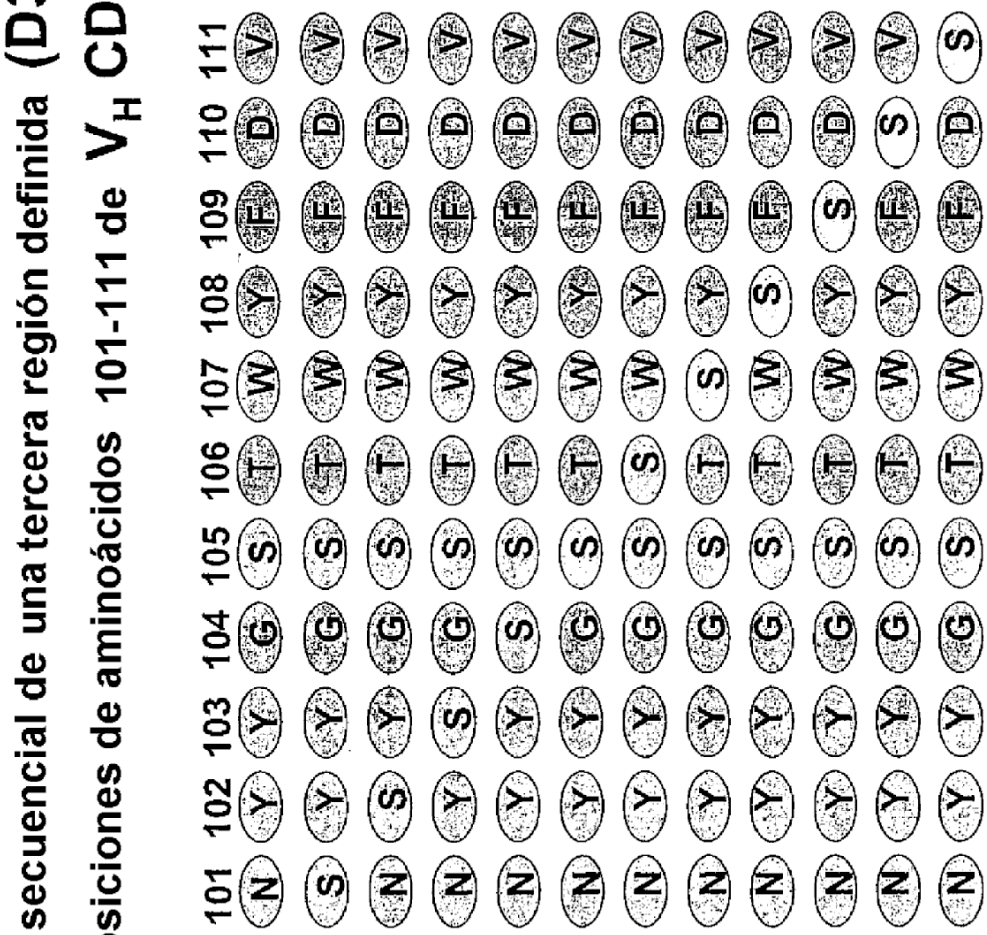
LTM secuencial de una segunda región definida (D2)

(posiciones de aminoácidos 55-68 de V_H CDR2)

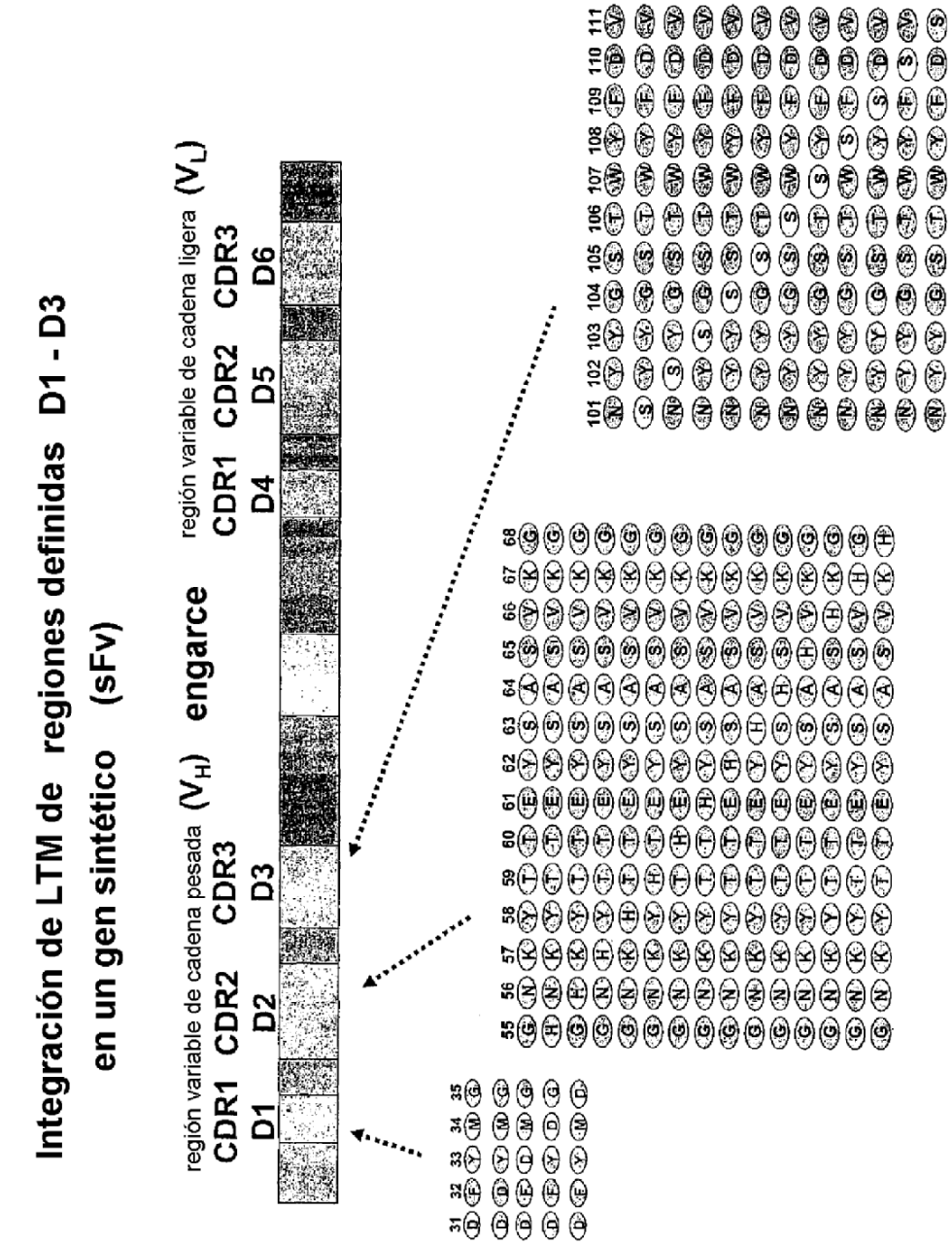


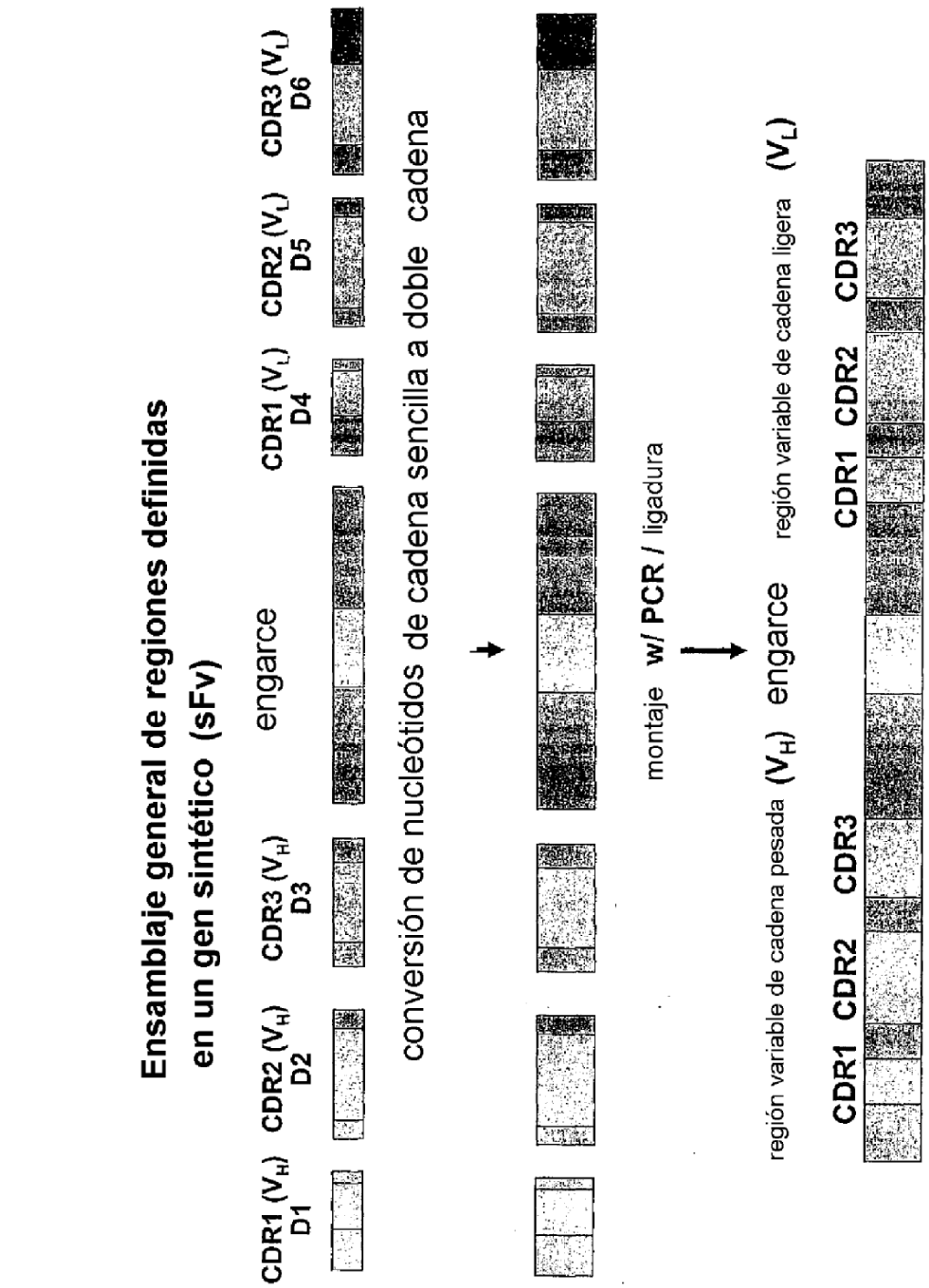
LTM secuencial de una tercera región definida (D3)

(posiciones de aminoácidos 101-111 de V_H CDR3)

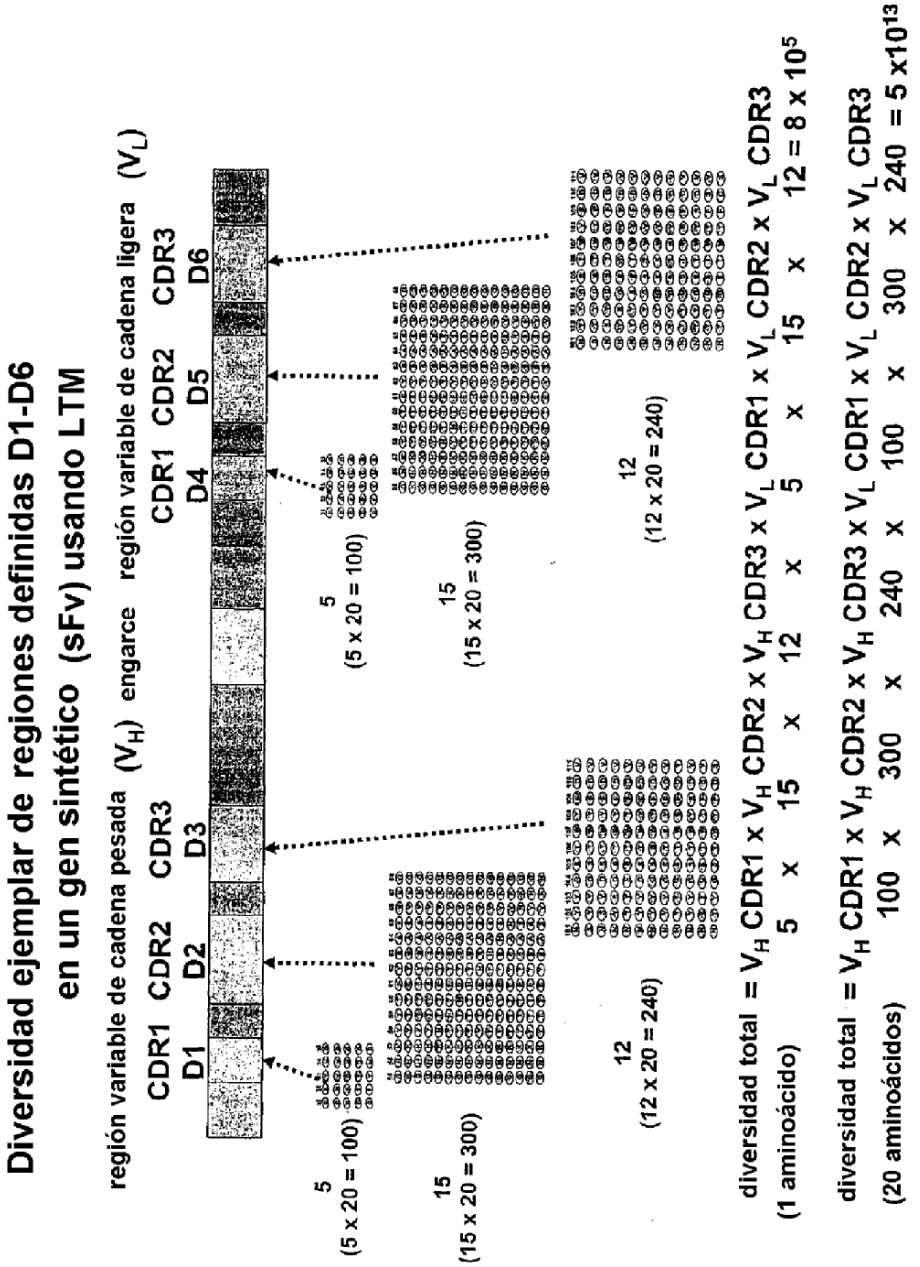


Mutagénesis de espera (WTM)
V_H CDR1 Asp D espera
31 32 33 34 35
D F Y M G
GAC TTC TAC ATG GAG
GA G GAT T
aminoácidos 1 4 2 7 2
D F Y M G
D D D D
1 oligo rendimiento de ensayo
2^7 = 128 oligo comb.
112 secuencias de aminoácidos
% de aminoácidos deseados (D) 100 25 50 12,5 50
% de aminoácidos no deseados 0 75 50 87,5 50
Mutagénesis de revisión (LTM)
V_H CDR1 Asp D revisión
31 32 33 34 35
D F Y M G
GAC TTC TAC ATG GAG
D D Y M G
GAC GAC TAC ATG GAG
D F D M G
GAC TTC GAC ATG GAG
D F Y D G
GAC TTC TAC GAC GAG
D F Y M D
GAC TTC TAC ATG GAC
5 oligo ensayos rendimiento 5 oligo combinaciones 5 secuencias de aminoácidos
% de aminoácidos deseados 100 100 100 100 100
% de aminoácidos no deseados 0 0 0 0 0


Integración de LTM de regiones definidas D1 - D3
en un gen sintético (sFv)
región variable de cadena pesada (V_H)
CDR1 CDR2 CDR3
D1 D2 D3
engarce
región variable de cadena ligera (V_L)
CDR1 CDR2 CDR3
D4 D5 D6


Ensamblaje general de regiones definidas
en un gen sintético (sFv)
CDR1 (V_H) D1
CDR2 (V_H) D2
CDR3 (V_H) D3
engarce
CDR1 (V_L) D4
CDR2 (V_L) D5
CDR3 (V_L) D6
conversión de nucleótidos de cadena sencilla a doble cadena
montaje w/ PCR / ligadura
región variable de cadena pesada (V_H)
engarce
región variable de cadena ligera (V_L)
CDR1 CDR2 CDR3
CDR1 CDR2 CDR3

Formatos de diversidad ejemplar de regiones definidas D1-D3 en un gen sintético


Diversidad ejemplar de regiones definidas D1-D6
en un gen sintético (sFv) usando LTM
región variable de cadena pesada (V_H) engarce región variable de cadena ligera (V_L)
CDR1 CDR2 CDR3 CDR1 CDR2 CDR3
D1 D2 D3 D4 D5 D6
5 (5 x 20 = 100)
15 (15 x 20 = 300)
12 (12 x 20 = 240)
5 (5 x 20 = 100)
15 (15 x 20 = 300)
12 (12 x 20 = 240)
diversidad total = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(1 aminoácido) 5 x 15 x 12 x 5 x 15 x 12 = 8 x 10^5
diversidad total = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(20 aminoácidos) 100 x 300 x 240 x 100 x 300 x 240 = 5 x 10^{13}

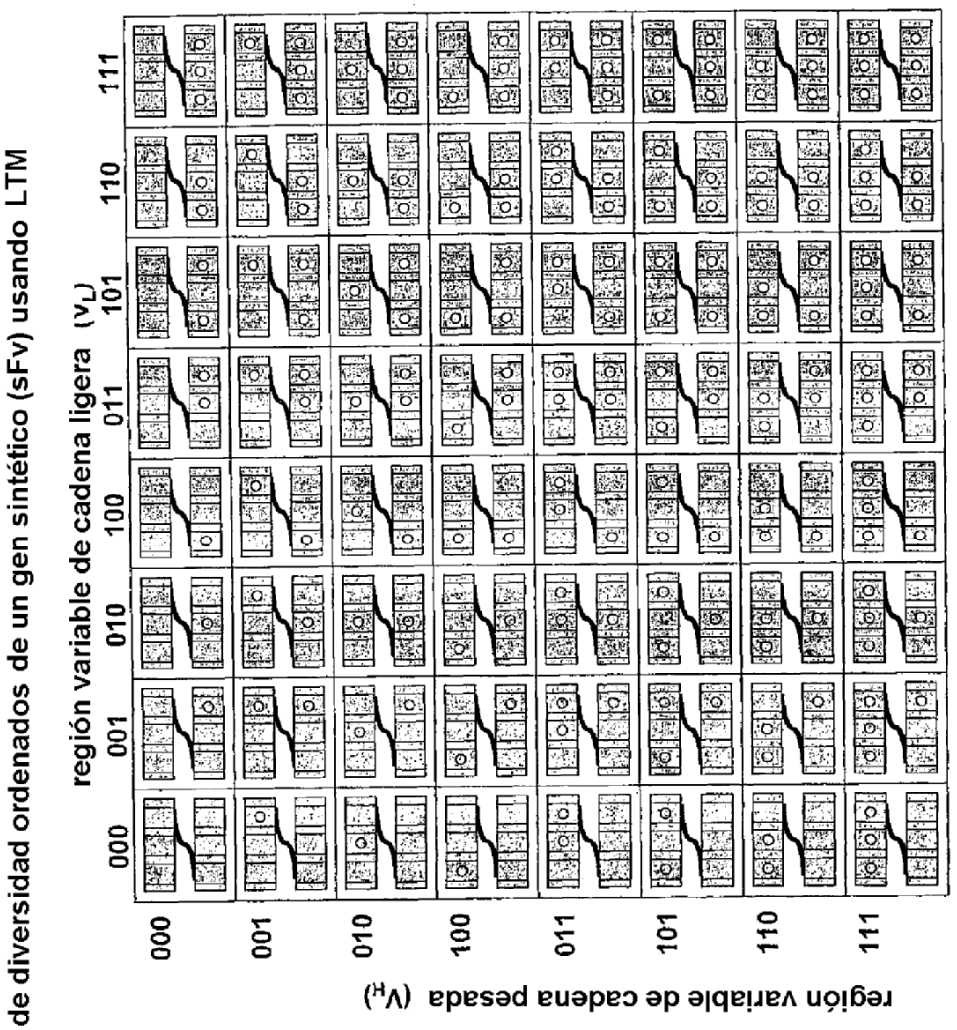
Diversidad ejemplar de regiones definidas D1-D6 en un gen sintético (a-TNF sFv) usando LTM

diversidad total = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(1 aminoácido) 5 x 15 x 12 x 5 x 15 x $12 = 8 \times 10^5$

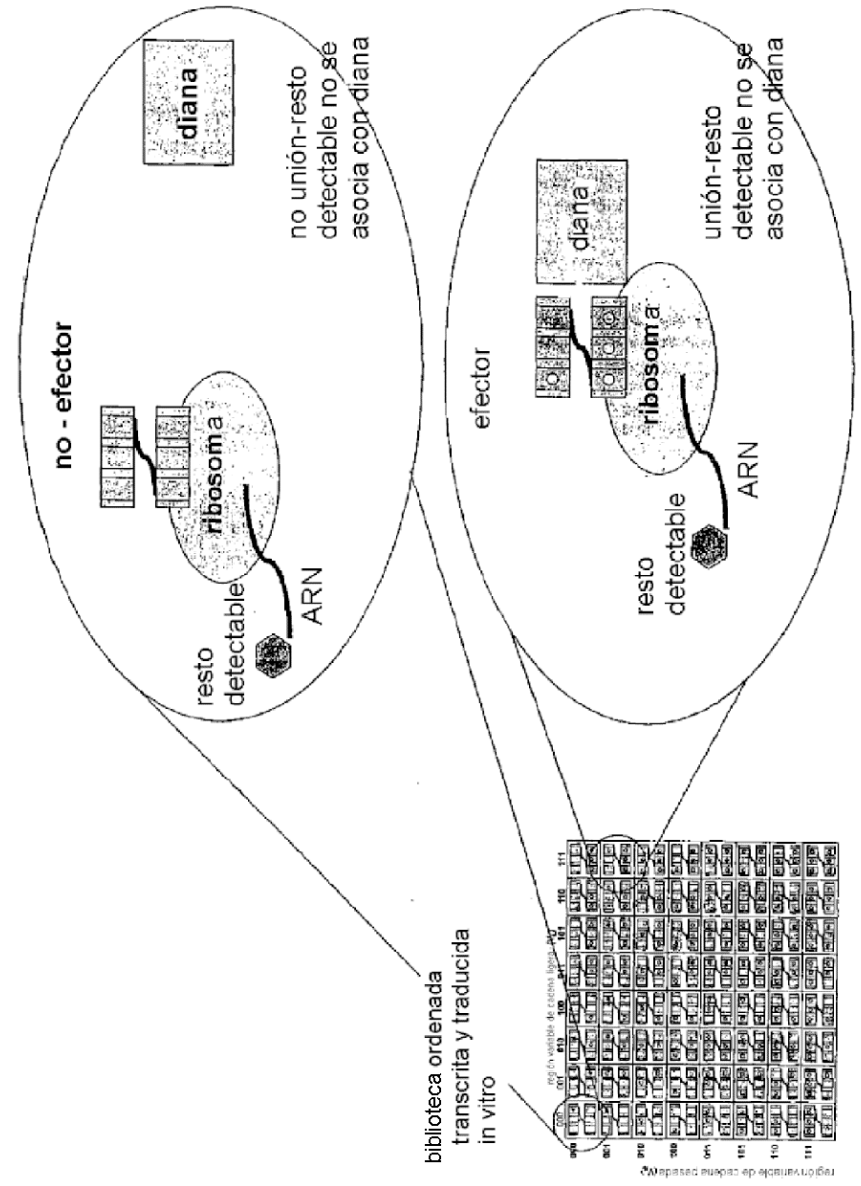
diversidad total = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(3 aminoácidos) 15 x 54 x 27 x 33 x 21 x $27 = 4.1 \times 10^8$

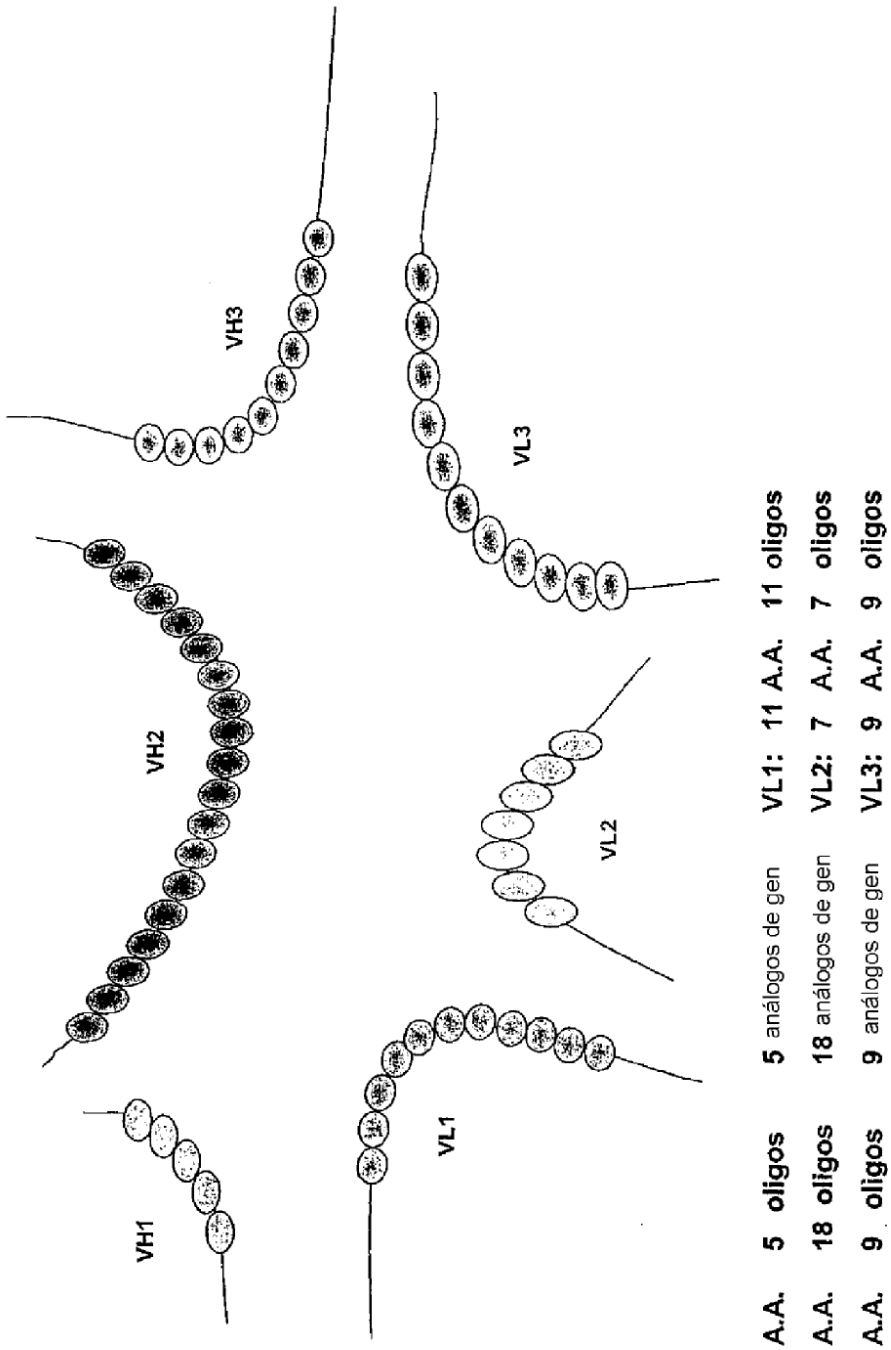
Asp, Ser, His

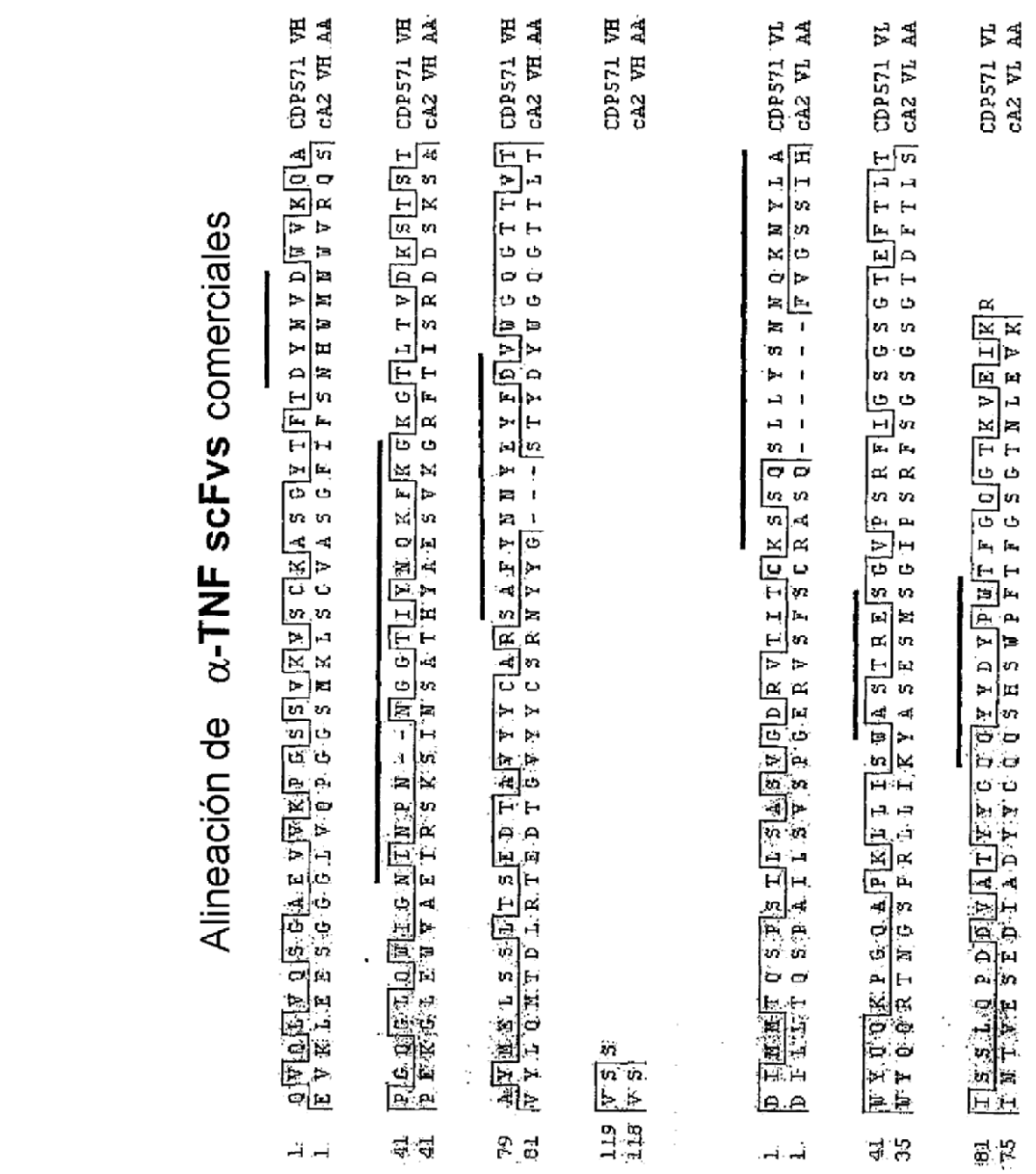
Formatos de diversidad ordenados de un gen sintético (sFv) usando LTM

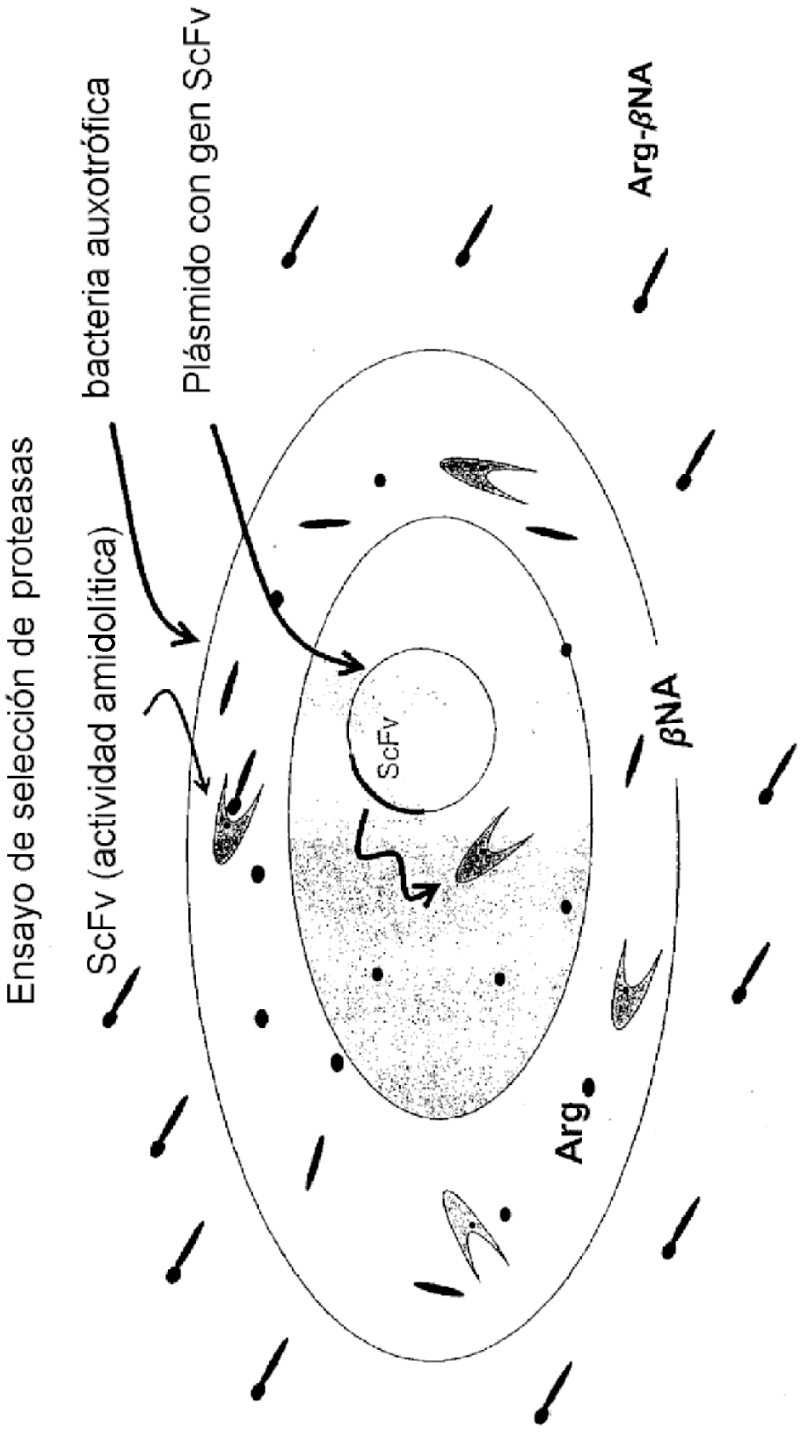


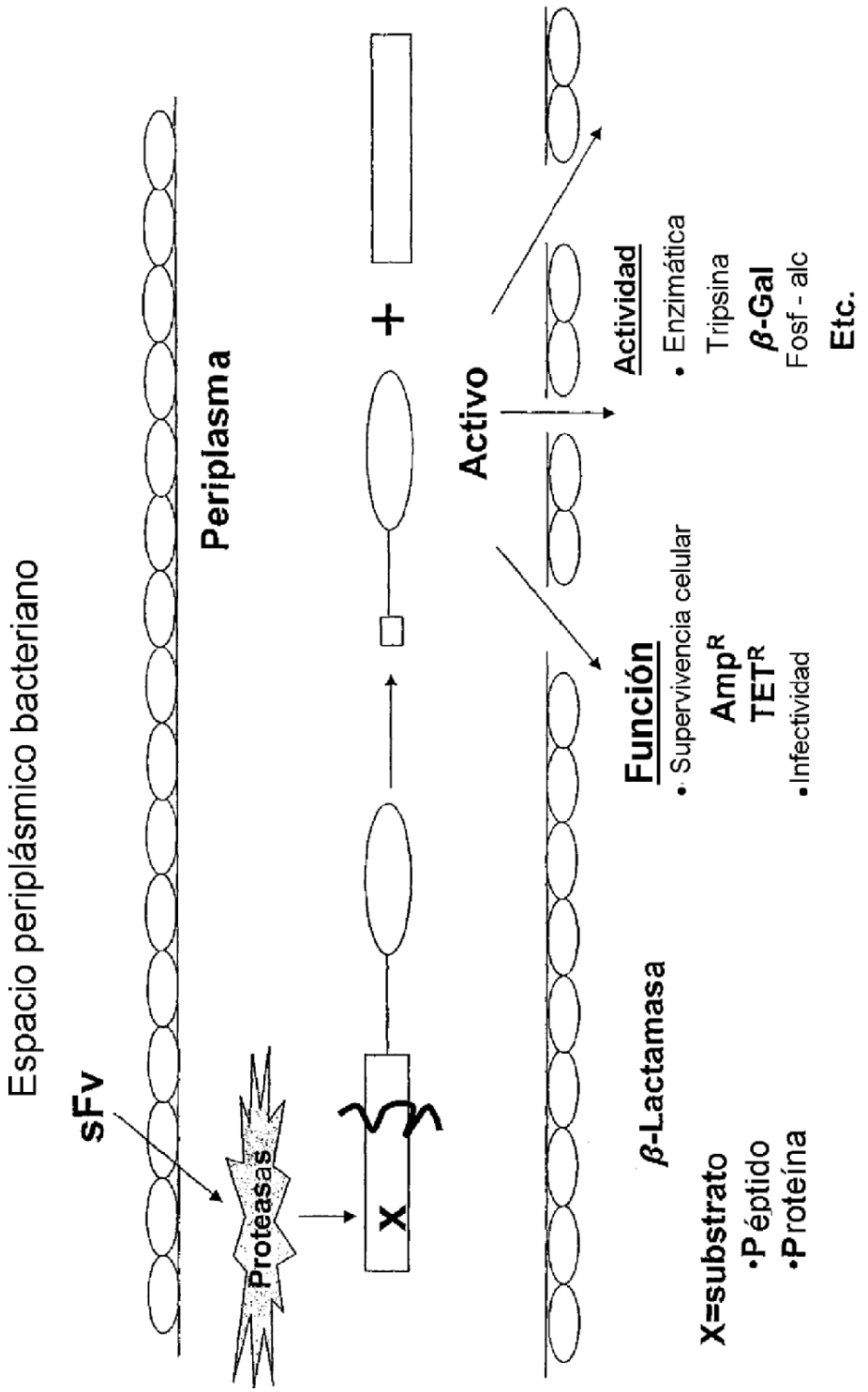
Sistema de detección universal para para seleccionar una biblioteca ordenada

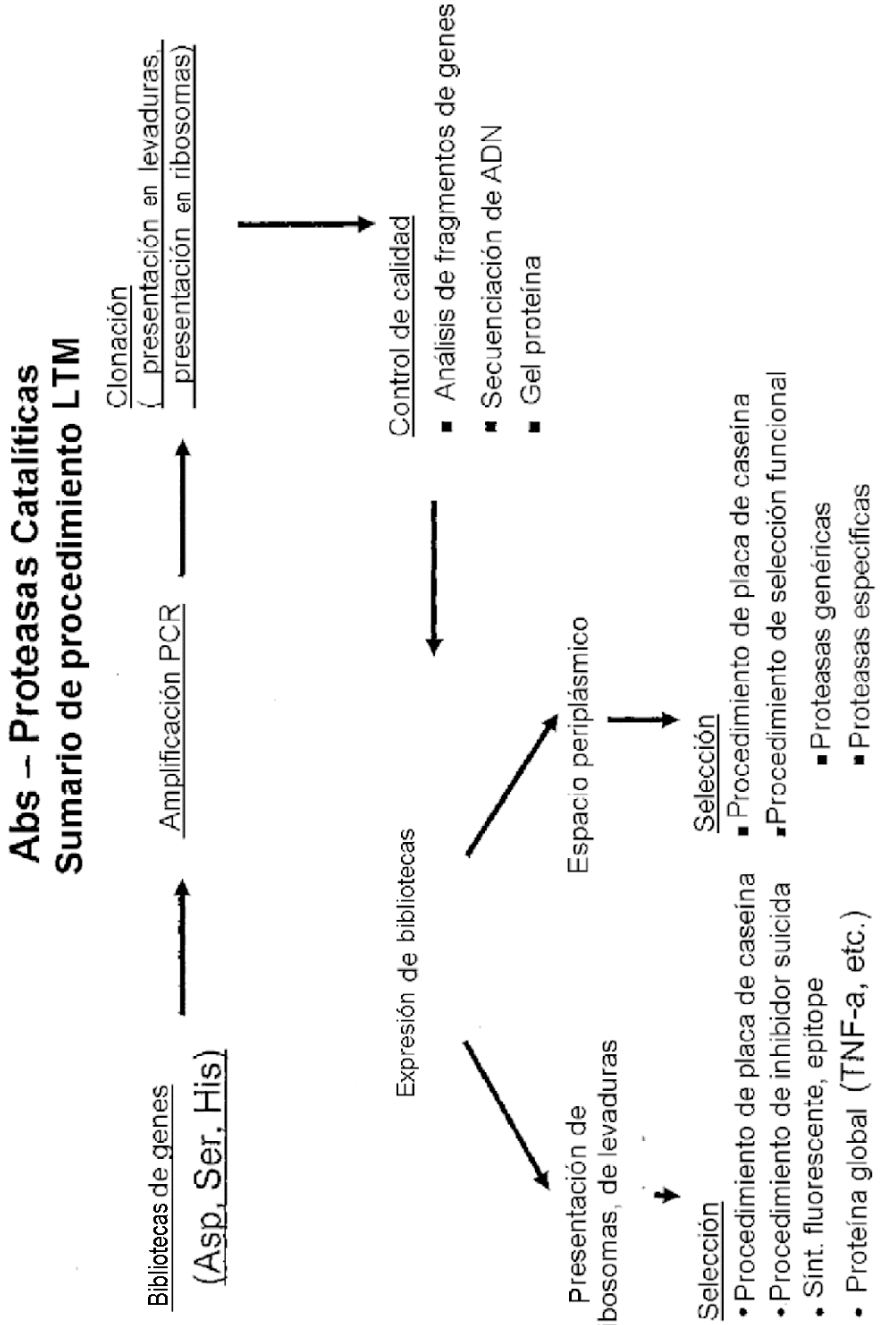


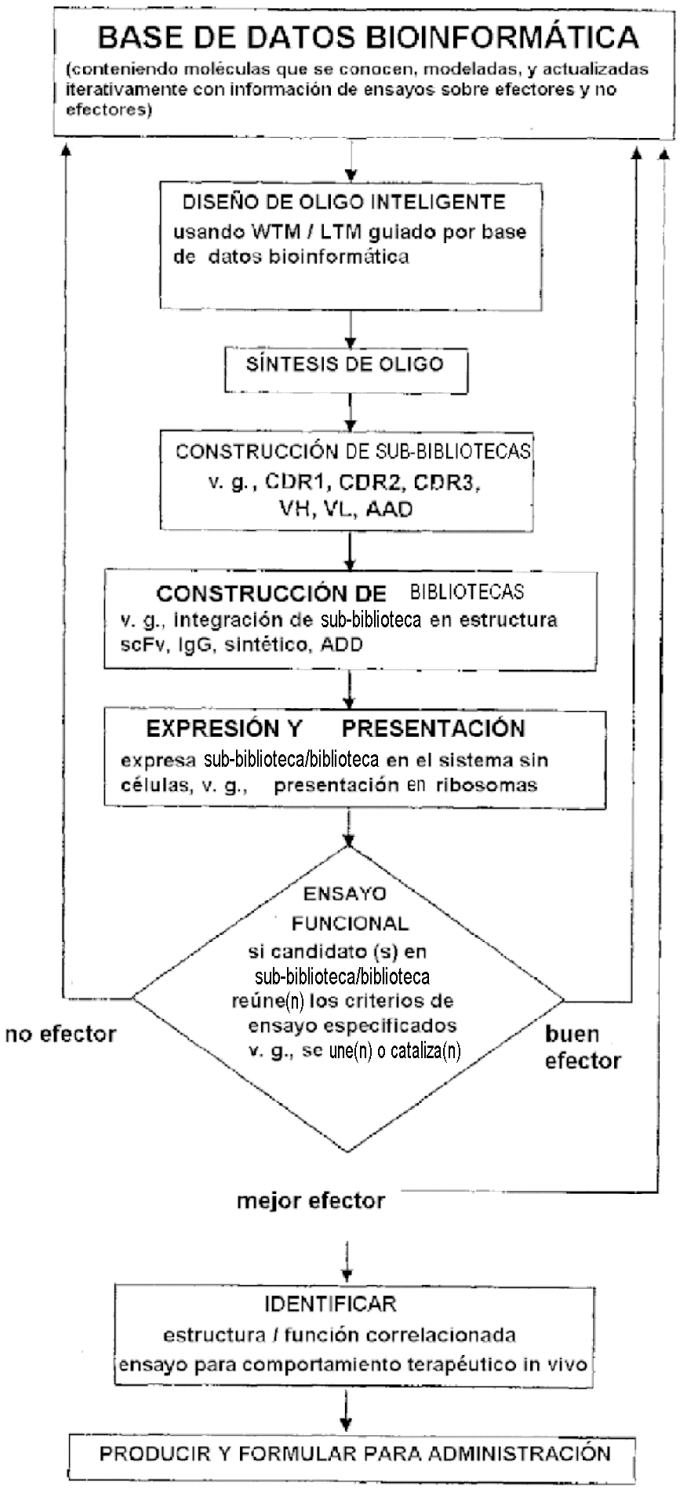

VH3
VH2
VH1
VL3
VL2
VL1
cA2
VH1: 5 A.A. 5 oligos 5 análogos de gen VL1: 11 A.A. 11 oligos
VH2: 18 A.A. 18 oligos 18 análogos de gen VL2: 7 A.A. 7 oligos
VH3: 9 A.A. 9 oligos 9 análogos de gen VL3: 9 A.A. 9 oligos

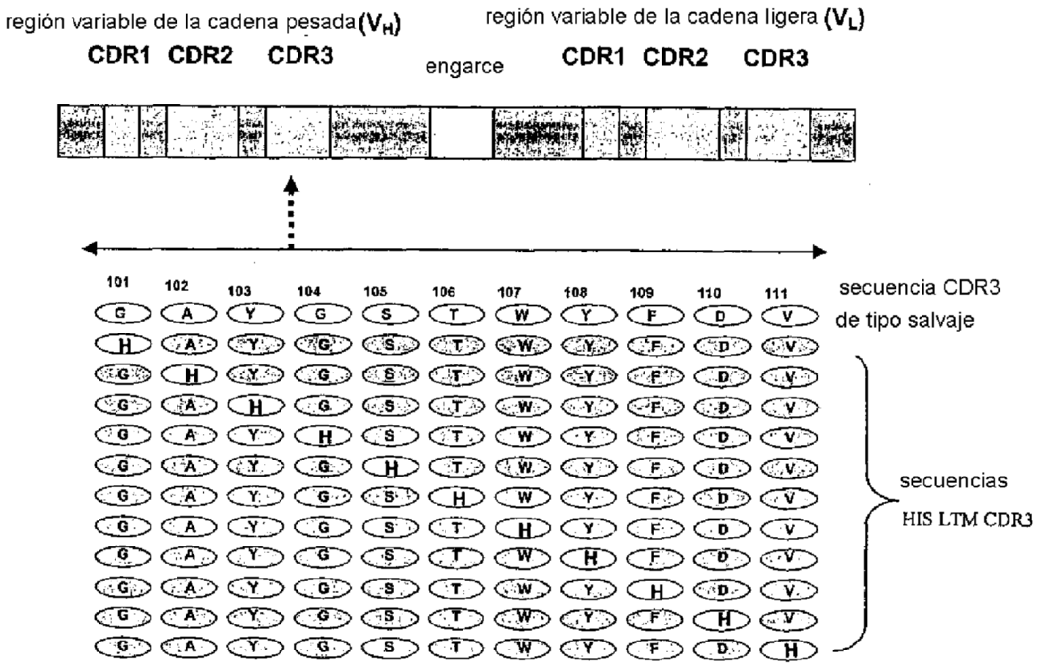

Alineación de α-TNF scFvs comerciales
CDP571 VH
cA2 VH AA
CDP571 VL
cA2 VL AA

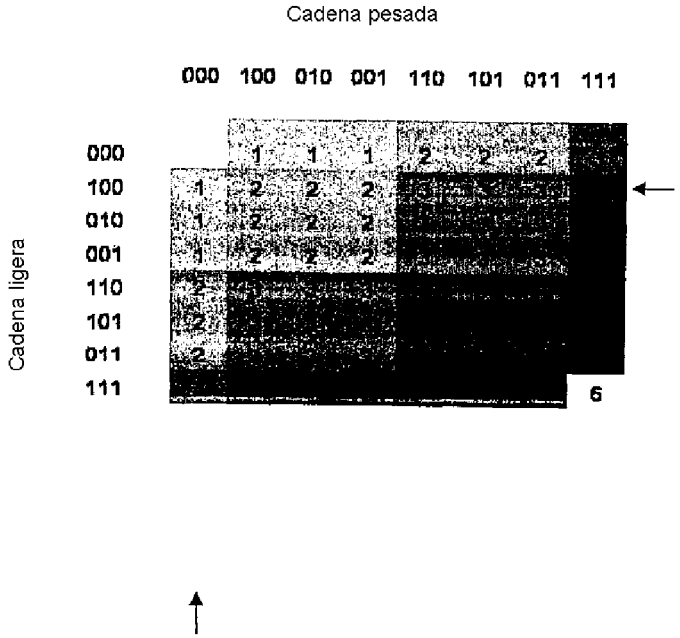

Ensayo de selección de proteasas
ScFv (actividad amidolítica)
bacteria auxotrófica
Plásmido con gen ScFv
Arg-βNA
ScFv
βNA
Arg.


Espacio periplásmico bacteriano
sFv
Periplasma
Proteasas
X
+
Activo
Actividad
• Enzimática
Tripsina
β-Gal
Fosf - alc
Etc.
Función
• Supervivencia celular
Amp^R
TET^R
•Infectividad
β-Lactamasa
X=substrato
•Péptido
•Proteína


Abs – Proteasas Catalíticas
Sumario de procedimiento LTM
Bibliotecas de genes
(Asp, Ser, His)
Amplificación PCR
Clonación
(presentación en levaduras,
presentación en ribosomas)
Control de calidad
Análisis de fragmentos de genes
Secuenciación de ADN
Gel proteína
Expresión de bibliotecas
Espacio periplásmico
Selección
Procedimiento de placa de caseína
Procedimiento de selección funcional
Proteasas genéricas
Proteasas específicas
Presentación de
ribosomas, de levaduras
Selección
Procedimiento de placa de caseína
Procedimiento de inhibidor suicida
Sínt. fluorescente, epitope
Proteína global (TNF-a, etc.)


BASE DE DATOS BIOINFORMÁTICA
(conteniendo moléculas que se conocen, modeladas, y actualizadas iterativamente con información de ensayos sobre efectores y no efectores)
DISEÑO DE OLIGO INTELIGENTE
usando WTM / LTM guiado por base de datos bioinformática
SÍNTESIS DE OLIGO
CONSTRUCCIÓN DE SUB-BIBLIOTECAS
v. g., CDR1, CDR2, CDR3, VH, VL, AAD
CONSTRUCCIÓN DE BIBLIOTECAS
v. g., integración de sub-biblioteca en estructura scFv, IgG, sintético, ADD
EXPRESIÓN Y PRESENTACIÓN
expresa sub-biblioteca/biblioteca en el sistema sin células, v. g., presentación en ribosomas
ENSAYO FUNCIONAL
si candidato (s) en sub-biblioteca/biblioteca reúne(n) los criterios de ensayo especificados v. g., se une(n) o cataliza(n)
no efector
buen efector
mejor efector
IDENTIFICAR
estructura / función correlacionada
ensayo para comportamiento terapéutico in vivo
PRODUCIR Y FORMULAR PARA ADMINISTRACIÓN


región variable de la cadena pesada (V_H)
CDR1 CDR2 CDR3
engarce
región variable de la cadena ligera (V_L)
CDR1 CDR2 CDR3
101 102 103 104 105 106 107 108 109 110 111
secuencia CDR3 de tipo salvaje
secuencias HIS LTM CDR3


Cadena pesada
000 100 010 001 110 101 011 111
Cadena ligera
000
100
010
001
110
101
011
111
6