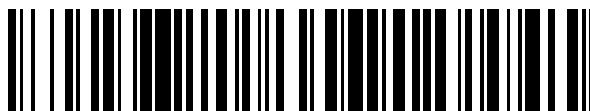


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 563 491**

51 Int. Cl.:

C07K 16/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.07.2005** **E 05770048 (6)**

97 Fecha y número de publicación de la concesión europea: **20.01.2016** **EP 1781680**

54 Título: **Bibliotecas universales de anticuerpos**

30 Prioridad:

06.07.2004 US 585931 P

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

15.03.2016

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

ES 2 563 491 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Bibliotecas universales de anticuerpos

Información relacionada

La solicitud reivindica prioridad sobre la solicitud de patente de EE.UU. provisional con número de solicitud 60/585.931, presentada el 6 de julio de 2004.

Antecedentes de la invención

Los anticuerpos tienen una profunda importancia como herramientas de investigación y en aplicaciones diagnósticas y terapéuticas. Sin embargo, la identificación de dichos anticuerpos útiles es difícil y, con frecuencia, especialmente si se conciben aplicaciones terapéuticas, requiere un considerable rediseño o "humanización" antes de que el anticuerpo sea adecuado para su administración.

Los procedimientos previos para la identificación de anticuerpos deseables han implicado típicamente la presentación en fagos de anticuerpos representativos, por ejemplo, bibliotecas humanas o bibliotecas sintéticas; sin embargo, estos planteamientos tienen limitaciones. Por ejemplo, la mayoría de las bibliotecas humanas contienen únicamente la diversidad de secuencias de anticuerpo que pueden capturarse o clonarse de forma experimental a partir del tejido de origen. Por consiguiente, la biblioteca humana puede carecer o infrarrepresentar otras secuencias de anticuerpo valiosas. Las bibliotecas sintéticas o de consenso tienen otras limitaciones tales como el potencial de codificar una secuencia de aparición no natural que tiene el potencial de ser inmunogénica. Además, las bibliotecas sintéticas, en un esfuerzo por ser completas, contienen frecuentemente demasiada diversidad y son difíciles de explorar. Es más, cuando se usan para identificar un anticuerpo candidato que se une a una diana concreta, estas bibliotecas no están abiertas a técnicas razonables de seguimiento y maduración de la afinidad para mejorar la unión de la molécula candidata. Por ejemplo, los procedimientos para la mejora posterior del anticuerpo suelen implicar mutagénesis *in vitro*, tal como mutagénesis aleatoria, mutagénesis por saturación, PCR propensa a errores, transposición de genes, y transposición de cadena de anticuerpo. Estas estrategias son fortuitas por naturaleza y suelen requerir la construcción de bibliotecas sumamente grandes para estudiar cualquier diversidad de secuencia significativa. A medida que aumenta el número de posiciones que hay que mutar en un anticuerpo, el tamaño de la biblioteca resultante aumenta más de lo que puede explorarse de forma factible.

Por consiguiente, existe una necesidad de una biblioteca universal de anticuerpos que represente de forma sistemática anticuerpos candidatos que no sean inmunogénicos y tengan propiedades deseadas, por ejemplo, una diversidad representativa que pueda explorarse fácilmente.

Sumario de la invención

La invención resuelve los problemas anteriores mediante la provisión de una biblioteca universal de anticuerpos (BUA) que representa a todos los anticuerpos candidatos deseables contra una clase antigénica dada, así como de los procedimientos de fabricación y exploración de dichas bibliotecas de anticuerpos. Es más, los anticuerpos de la biblioteca universal de anticuerpos proceden de secuencia humana, y son, por tanto, no inmunogénicos, y, por tanto, adecuados para aplicaciones terapéuticas, por ejemplo, para la administración a pacientes humanos para la prevención o el tratamiento de trastornos o enfermedades humanas.

Las bibliotecas de la invención, tienen una diversidad que se introduce de forma eficaz usando, por ejemplo, técnicas de mutagénesis tales como mutagénesis guiada (*walk-through mutagenesis*, WTM) o mutagénesis de recorrido (*look-through mutagenesis*, LTM) (véanse respectivamente, p. ej., las patentes de EE. UU. nº 6.649.340; 5.830.650; 5.798.208; y el documento U.S.S.N. 60/483.282) dependiendo de si hay que introducir diversidad de restos múltiples o diversidad de un residuo individual en un sitio dado, p. ej., dentro de una o más regiones determinantes de complementariedad (CDR, de *complementarity determining regions*). De manera importante, estas técnicas permiten aumentar al máximo la cantidad de diversidad productiva y reducir al mínimo la cantidad de diversidad no productiva, es decir, el simple ruido de fondo o azar. Por consiguiente, las bibliotecas universales de anticuerpos de la invención pueden ser más pequeñas que los anticuerpos existentes, pero comprender más diversidad razonable a fin de identificar de forma más eficaz moléculas candidatas de unión a anticuerpo.

En una realización, la biblioteca universal de anticuerpos desvelada es la aplicación de la tecnología WTM o LTM para crear una biblioteca completamente sintética que muestre una diversidad deseada en una o más CDR de las cadenas ligeras y/o pesadas. Las secuencias de anticuerpo, por ejemplo, las regiones marco conservadas y las CDR se seleccionan de acuerdo con ciertos criterios. Por ejemplo, un criterio es que la secuencia de anticuerpo ha de tener una frecuencia umbral mínima (p. ej., alrededor del 10 por ciento o más) de aparición dentro de las secuencias expresadas (reordenadas) de anticuerpo, p. ej., secuencias de anticuerpo humanas, y preferentemente, en respuesta a una clase concreta de antígenos. Opcionalmente, otro criterio más es que la secuencia expresada (reordenada) de anticuerpo se origine (o proceda), con una frecuencia umbral mínima (p. ej., alrededor de un 10 por ciento o más), de una secuencia de línea germinal. Es aún otro criterio/criterios hacer una comparación entre CDR (por ejemplo, CDR expresadas), que sean de una longitud, estructura canónica, y/o interdependencia de CDR dadas (p. ej., CDR1 contra CDR2, y/o 3. La diversidad se identifica y después se modifica por ingeniería genética en un

formato genético convencional, p. ej., un formato de anticuerpo monocatenario (scFv), usando oligonucleótidos que permitan el ensamblaje completo de las secuencias de región marco conservada y CDR mediante ingeniería genética (p. ej., reacción en cadena de la polimerasa (PCR), extensión por solapamiento único (SOE, de *single overlap extension*) y/o mutagénesis mediada de Kunkel), de una forma sistemática.

- 5 De manera importante, la invención reduce al mínimo cualquier mutación que pueda conducir a proteínas no funcionales mediante la evitación de mutaciones injustificadas que aparecen típicamente cuando se usan sondas mixtas. Además, el nivel de precisión capaz cuando se usa WTM contrasta con las tecnologías de mutagénesis aleatoria y/o transposición de genes. Es más, mediante el control de la selección de la secuencia marco conservada y el nivel de diversidad de secuencia respecto a la posición y tipo de aminoácido, se aumenta al máximo el reconocimiento de clases "de antígeno" por la biblioteca. Además, esta metodología *in vitro* evita la selección inmunológica negativa de autoantígenos y cualquier sesgo genético debido a la exposición medioambiental del organismo.

- 10 Por consiguiente, la invención proporciona la ventaja de ser capaz de empezar con una biblioteca de exploración dimensionada para ser informativa sin ser innecesariamente grande. Tras la identificación del primer conjunto de clones, las bibliotecas posteriores de maduración de la afinidad pueden compartir conjuntos comunes de nucleótidos de LTM y/o WTM, ahorrando tiempo y costes de reactivos. Más aún, las bibliotecas universales de anticuerpos son capaces de producir de forma rápida y eficaz anticuerpos muy específicos contra una variedad de antígenos, especialmente, p. ej., autoantígenos, que son difíciles de obtener mediante cualquier otro método.

- 15 La biblioteca universal de anticuerpos se genera y se explora mediante la síntesis inicial de polinucleótidos individuales que codifican una región o regiones definidas de un anticuerpo en el que, de forma colectiva, los polinucleótidos representan todos los posibles anticuerpos variantes de acuerdo con los criterios descritos en el presente documento. Los anticuerpos 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 en bacterias, o presentación en levaduras.

- 20 Después, los anticuerpos expresados se exploran y se seleccionan usando ensayos funcionales, tales como ensayos de ligamiento. En una realización, los polipéptidos se expresan en asociación con el polinucleótido que codifica la molécula de unión al anticuerpo, p. ej., un anticuerpo monocatenario (scFv), permitiendo así la identificación de la secuencia de polinucleótido que codifica la molécula de unión al anticuerpo (p. ej., scFv). En una realización relacionada, las anticuerpos se segregan y se presentan sobre la membrana de un procariota, tal como *E. coli*, usando, p. ej., la tecnología descrita en, p. ej., los documentos US20040072740A1; US20030036092A1; y US20030100023A1.

- 25 El método puede usarse para identificar secuencias de anticuerpos humanos para desarrollar anticuerpos nuevos o mejorados o fragmentos de los mismos, p. ej., anticuerpos monocatenarios (scFv). Además, el método puede realizarse con el beneficio de una información *a priori*, p. ej., por medio de elaboración de modelos por ordenador y bioexploración de bases de datos electrónicas, que pueden usarse para seleccionar un subgrupo inicial de secuencias para diversificar, p. ej., de acuerdo con el criterio descrito en el presente documento, usando, p. ej., mutagénesis WTM o LTM.

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

40 **Breve descripción de las figuras**

La **figura 1** muestra un esquema para llevar a cabo la construcción de una biblioteca universal de anticuerpos de la invención usando bioexploración de bases de datos asistida por ordenador.

- 45 La **figura 2** destaca etapas de ejemplo (y diversas estadísticas de bases de datos) para identificar y seleccionar CDR y componentes de región marco conservadas para su uso en la síntesis de bibliotecas universales de anticuerpos.

Las **figuras 3-4** muestran frecuencias umbrales de ejemplo de análisis de aparición, es decir, una identificación de las regiones marco conservadas de línea germinal de uso más frecuente usadas en una respuesta inmunitaria de anticuerpo humano frente a una clase antigénica dada. En la **figura 4**, la frecuencia relativa de la contribución de la línea germinal a los anticuerpos reordenados está tabulada dentro de cada familia de VH de línea germinal.

- 50 La **figura 5** muestra la secuencia de siete regiones marco conservadas de cadena pesada de alta frecuencia, usadas en respuesta frente a una clase antigénica dada (p. ej., un antígeno a base de proteína) y su disposición para funcionar como aceptores para regiones CDR sintéticas. Las CDR están de acuerdo con la definición del contacto (Mac Callum y cols.). Se ilustran los segmentos V de línea germinal enumerados de VH1, VH3 y VH4.

- 55 La **figura 6** muestra la diversidad de secuencia de una CDR1 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR (distribución de frecuencia). Longitud de 5 de la CDR1 de acuerdo con la definición de CDR de Kabat.

La **figura 7** muestra la diversidad de secuencia generada de una CDR1 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR y una matriz que muestra las posiciones de los restos y la diversidad potencial. Longitud de 6 de la CDR1 de acuerdo con la definición de contacto de CDR.

5 La **figura 8** muestra la diversidad de secuencia de una CDR2 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR (distribución de frecuencia). Longitud de 17 de la CDR2 de acuerdo con la definición de CDR de Kabat.

La **figura 9** muestra la diversidad de secuencia generada de una CDR2 de VH1 y VH3 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR y una matriz que muestra las posiciones de los restos y la diversidad potencial. Longitud de 13 de la CDR2 de acuerdo con la definición de contacto de CDR.

10 La **figura 10** muestra la distribución de longitudes de CDR3 de VH de tamaños de 9 a 18 aminoácidos que cubren alrededor del 75% el espacio de CDR disponible. Se realizó un análisis por separado para cada longitud (véase la fig. 12). Los tamaños de CDR3 de VH están de acuerdo con la definición de contacto de CDR.

15 La **figura 11** muestra la diversidad de secuencia de una CDR3 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR (distribución de frecuencia). Longitud de 13 de la CDR3 de acuerdo con la definición de CDR de Kabat.

La **figura 12** muestra la diversidad de secuencia generada de una CDR3 sintética de cadena pesada de ejemplo en forma de un perfil de variabilidad de CDR y una matriz que muestra las posiciones de los restos y la diversidad potencial. Tamaños de longitud de 9-18 de la CDR3 de acuerdo con la definición de contacto de CDR.

20 La **figura 13** muestra la diversidad de secuencia de cada CDR de cadena pesada, así como la diversidad de la biblioteca combinada de cadenas pesadas. El número de posiciones variables y tamaños de CDR están de acuerdo con la definición de Kabat.

La **figura 14** muestra la construcción de la biblioteca de cadena pesada usando una combinación de oligonucleótidos solapados no degenerados y degenerados, los cuales pueden convertirse a ácidos nucleicos bicatenarios usando la reacción en cadena de polimerasa de extensión por solapamiento único (SOE-PCR).

25 La **figura 15** muestra la combinación de la biblioteca de cadena pesada de la biblioteca universal de anticuerpos con una biblioteca kappa y/o lambda de cadena ligera para más diversidad.

30 La **figura 16** muestra la secuencia de siete regiones marco conservadas de cadena ligera de alta frecuencia (es decir, tres regiones marco conservadas de cadena ligera kappa y cuatro lambda) usadas en respuesta frente a una clase antigénica dada (p. ej., proteína) y su disposición para funcionar como aceptores para regiones CDR sintéticas. Las regiones CDR se identifican de acuerdo a la definición de contacto. Se ilustran los segmentos V de línea germinal enumerados de Vkappa I-L1, Vkappa III-A27, Vkappa III-L6, Vlambda 1-1b, Vlambda 2-2a2, Vlambda 3-31, y Vlambda 3-3r.

35 La **figura 17** muestra la diversidad de secuencia generada de CDR1 sintéticas de cadena ligera de Vkappa I y Vkappa III de ejemplo en forma de perfiles de variabilidad (distribuciones de frecuencia) y matrices de permutación. Tamaños de longitud de 7 y 8 de la CDR1 de acuerdo con la definición de contacto de CDR.

La **figura 18** muestra la secuencia de ácido nucleico y la secuencia de aminoácidos de un anticuerpo monocatenario de ejemplo (scFv) de la invención.

40 La **figura 19** muestra la diversidad de secuencia generada de CDR2 sintéticas de cadena ligera Vkappa I y Vkappa III de ejemplo en forma de perfiles de variabilidad (distribuciones de frecuencia) y matriz que muestra las posiciones de los restos y la diversidad potencial. Tamaño de longitud de 10 de la CDR2 de acuerdo con la definición de contacto de CDR.

La **figura 20** muestra la distribución de longitud de CDR3 de Vkappa de tamaños de 8 y 9 aminoácidos que cubren alrededor del 80 % el espacio de CDR disponible. Se realizó un análisis por separado para cada longitud (véase la fig. 21). Los tamaños de la CDR3 de VH están de acuerdo con la definición de contacto de CDR.

45 La **figura 21** muestra la diversidad de secuencia generada de CDR3 sintéticas de cadena ligera (Vkappa) de ejemplo en forma de perfiles de variabilidad (distribuciones de frecuencia) y matriz que muestra las posiciones de los restos y la diversidad potencial. Tamaños de longitud de 8-9 de la CDR3 de Vkappa de acuerdo con la definición de contacto de CDR.

50 La **figura 22** muestra la diversidad de secuencia generada de CDR1 sintéticas de cadena ligera (Vlambda) de ejemplo en forma de perfiles de variabilidad (distribución de frecuencias) y matriz que muestra las posiciones de los restos y la diversidad potencial. Tamaños de longitud de 9, 10 y 7 de la CDR1 de V₁ de acuerdo con la definición de contacto de CDR.

La **figura 23** muestra la diversidad de secuencia generada de CDR2 sintéticas de cadena ligera Vlambda 1,

Vlambda 2 y Vlambda 3 de ejemplo en forma de perfiles de variabilidad (distribuciones de frecuencia) y matriz que muestra las posiciones de los restos y la diversidad potencial. Tamaños de longitud 10 de la CDR2 de Vlambda 1, Vlambda 2 y Vlambda 3, de acuerdo con la definición de contacto.

La **figura 24** muestra la distribución de longitud de CDR3 de Vlambda de tamaños de 8 a 11 aminoácidos que cubren alrededor del 90 % el espacio de CDR disponible. Se realizó un análisis por separado para cada longitud (véase la fig. 25). Los tamaños de la CDR3 de VH están de acuerdo con la definición de contacto de CDR.

La **figura 25** muestra la diversidad de secuencia generada de CDR3 sintéticas de cadena ligera Vlambda de ejemplo en forma de perfiles de variabilidad (distribuciones de frecuencia) y matriz que muestran las posiciones de los restos y la diversidad potencial. Tamaños de longitud de 8, 9, 10 y 11 de la CDR3 de Vlambda de acuerdo con la definición de contacto de CDR.

La **figura 26** muestra la incorporación de diversidad de CDR dentro de un molde de expresión usando mutagénesis de Kunkel.

La **figura 27** muestra la diversidad de secuencia de cada CDR de cadena ligera, así como la diversidad combinada de la biblioteca de cadenas ligeras. El número de posiciones variables y tamaños de CDR están de acuerdo con la definición de Kabat.

La **figura 28** muestra la construcción de la biblioteca de cadena ligera usando una combinación de oligonucleótidos solapados no degenerados y degenerados, los cuales pueden convertirse a ácidos nucleicos bicatenarios usando reacción en cadena de polimerasa de extensión por solapamiento único (SOE-PCR).

La **figura 29** muestra la maduración de la afinidad de un anticuerpo de ensayo (panel izquierdo) y la diversidad de frecuencia resultante obtenida (panel inferior) y la unión mejorada de varios clones representativos (panel derecho).

La **figura 30** muestra una comparación de tres definiciones de CDR: Kabat (Kabat y cols.), Chothia (Chothia y cols.), consideraciones del contacto (MacCallum y cols.) para las cadenas VH (A) y VL (B). Los triángulos pequeños sobre los segmentos de CDR señalan las localizaciones en las que se producen las inserciones. El número de aminoácidos de cada CDR (es decir, las longitudes de las CDR) se muestra debajo de las dos gráficas.

La figura 31 muestra una muestra de segmentos VBASE analizados sintácticamente y almacenados en formato FR1-CDR1-FR2-CDR2-FR3. Las localizaciones de las CDR se identifican de acuerdo con la definición de contacto (MacCallum y cols.) y el esquema de numeración está de acuerdo con Chothia (Chothia y cols.). A partir de estos conjuntos de datos pueden extraerse datos individuales para cada FR o CDR.

La **figura 32** muestra las secuencias VH de línea germinal de VBASE almacenadas como FR123 en formato FASTA. La longitud de las regiones marco conservadas se refiere a la definición de contacto de CDR.

La **figura 33** muestra un árbol jerárquico obtenido a partir de segmentos V_H de VBASE en un formato FR123. El algoritmo de agrupamiento UPGMA se ha usado con la matriz de distancia computerizada usando la distancia p (fracción de desapareamientos).

La **figura 34** muestra un árbol jerárquico obtenido a partir de segmentos V_{κ} de VBASE en un formato FR123. El algoritmo de agrupamiento UPGMA se ha usado con la matriz de distancia computerizada usando la distancia p (fracción de desapareamientos).

La **figura 35** muestra un árbol jerárquico obtenido a partir de segmentos Vlambda de VBASE en un formato FR123. El algoritmo de agrupamiento UPGMA se ha usado con la matriz de distancia computerizada usando la distancia p (fracción de desapareamientos).

La **figura 36** muestra un conjunto de datos de entrada Kabat de VH (formato FR123) de anticuerpos antiproteína conocidos, visualizado como un árbol jerárquico (UPGMA).

La **figura 37** muestra un conjunto de datos de entrada Kabat de V_{κ} (formato FR123) de anticuerpos antiproteína conocidos, visualizado como un árbol jerárquico (UPGMA).

La **figura 38** muestra un conjunto de datos de entrada Kabat de Vlambda (formato FR123) de anticuerpos antiproteína conocidos, visualizado como un árbol jerárquico (UPGMA).

La **figura 39** muestra las secuencias de aminoácidos de 12 segmentos de línea germinal seleccionados. En esta figura, también se incluyen las secuencias de CDR de línea germinal. Éstas se sustituyen por secuencias mutadas como se describe en la siguiente sección.

La **figura 40** muestra un árbol de regiones marco conservadas, apropiado para la clase antigénica de los polisacáridos.

La **figura 41** muestra un diagrama de flujo que ilustra la selección del perfil de variabilidad de CDR1 y CDR2. [Véase el ejemplo 3].

La **figura 42** muestra un diagrama de flujo que ilustra la selección del perfil de variabilidad de CDR3. [Véase el ejemplo 3].

5 La **figura 43** muestra el proceso de selección para conjuntos de subgrupos de secuencias de cadena pesada, la división en subclases de secuencias de subgrupos, y la división posterior de las poblaciones de subclases basada en la estructura canónica.

La **figura 44** muestra un histograma de la prevalencia de restos de aminoácido en cada posición dentro de una población VH1_CDR1_6.

10 La **figura 45** muestra un histograma de la prevalencia de restos de aminoácido en cada posición dentro de una población VH1_CDR1_6_CS1 (estructura canónica 2).

La **figura 46** muestra un histograma de la prevalencia de restos de aminoácido en cada posición dentro de una población VH1_CDR1_6_CS1-2.

15 La **figura 47** muestra un histograma de la prevalencia de restos de aminoácido en cada posición dentro de una población VH1_CDR1_6_CS1-3.

La **figura 48** muestra un histograma de la prevalencia de restos de aminoácido en cada posición dentro de una población VH1_CDR1_6_CS2_13.

Las **Figuras 49A y 49B** muestran histogramas de prevalencia de restos de aminoácidos en cada posición dentro de una población VH1_CDR2_13_CS2 y una población VH1_CDR2_13_CS3 respectivamente,

20 Las **Figuras 50A y 50B** muestran histogramas de prevalencia de restos de aminoácidos en cada posición dentro de una población de VH1_CDR2_13_CS2-1 y una población VH1_CDR2_13_CS3-1, respectivamente.

Las **Figuras 51A, 51B y 51C** muestran histogramas de prevalencia de restos de aminoácidos en cada posición dentro de una población VH_CDR3-9, una población VH_CDR3-15, y una población VH_CDR3-18, respectivamente.

25 **Descripción detallada de la invención**

A fin de proporcionar una comprensión clara de la especificación y de las reivindicaciones, a continuación se proporcionan las siguientes definiciones.

Definiciones

30 Tal como se usa en el presente documento, la expresión "regiones de unión a anticuerpo" se refiere a una o más porciones de una región variable de una inmunoglobulina o un anticuerpo, capaz de unir un antígeno(s). Típicamente, la región de unión a anticuerpo es, por ejemplo, una cadena ligera de anticuerpo (VL) (o una región variable de la misma), una cadena pesada de anticuerpo (VH) (o una región variable de la misma), una región Fd de cadena pesada, una cadena de anticuerpo combinada ligera y pesada (o una región variable de la misma) tal como un Fab, F(ab')₂, dominio único, o anticuerpo monocatenario (scFv), o un anticuerpo de longitud completa, por
35 ejemplo, una IgG (p. ej., un subtipo IgG1, IgG2, IgG3, o IgG4), o un anticuerpo IgA1, IgA2, IgD, IgE, o IgM.

La expresión "región marco conservada" se refiere a las porciones reconocidas en la técnica de la región variable de un anticuerpo que existen entre las regiones CDR más divergentes. Dichas regiones marco conservadas se denominan típicamente regiones marco conservadas 1 hasta 4 (FR1, FR2, FR3, y FR4) y proporcionan un armazón para sostener en el espacio tridimensional a las tres CDR que se encuentran en una región variable de cadena
40 pesada o ligera de anticuerpo, de modo que las CDR pueden formar una superficie de unión a antígeno.

La expresión "frecuencia umbral de aparición" se refiere a un criterio de la invención que requiere que una secuencia seleccionada para su uso en la biblioteca universal de anticuerpos proceda de una secuencia que se haya determinado que es una secuencia privilegiada para que la expresen las células del sistema inmunitario, por ejemplo, al responder a una clase concreta de antígenos. Típicamente, dichas secuencias expresadas (reordenadas) que se ha determinado que cumplen la frecuencia umbral de aparición son secuencias que se expresan con un
45 porcentaje de aparición de alrededor del 10 % o más.

La expresión "frecuencia umbral de origen de línea germinal" se refiere a un criterio de la invención que requiere que una secuencia seleccionada (es decir, una secuencia expresada o reordenada) para su uso en la biblioteca universal de anticuerpos proceda de una secuencia que se haya determinado que es una secuencia de línea germinal privilegiada para que la expresen células del sistema inmunitario, por ejemplo, al responder a una clase concreta de antígenos. Típicamente, las secuencias que se ha determinado que cumplen la frecuencia umbral de origen de línea
50 germinal son secuencias que proceden o se originan a partir de una secuencia de línea germinal con un porcentaje

de aparición de alrededor del 10 % o más.

La expresión "clase antigénica predeterminada", o "clase de antígenos" o "clase antigénica" se refiere a antígenos que son estructuralmente/ químicamente similares en cuanto a su composición básica. Son clases antigénicas típicas proteínas (polipéptidos), péptidos, polisacáridos, polinucleótidos, y moléculas pequeñas.

- 5 La expresión "estructura canónica" incluye consideraciones como la secuencia lineal del anticuerpo, por ejemplo, como se cataloga en la base de datos Kabat. El esquema de numeración Kabat es un patrón ampliamente adoptado para la numeración de los restos de aminoácidos del dominio variable de un anticuerpo de una forma coherente. También pueden usarse consideraciones estructurales adicionales, por ejemplo, aquellas diferencias no reflejadas totalmente mediante la numeración Kabat, por ejemplo, como describen Chothia y cols. y reveladas, por ejemplo, mediante cristalografía y elaboración de modelos tridimensionales, para determinar la estructura canónica de un anticuerpo. Por consiguiente, una secuencia de anticuerpo dada puede colocarse dentro de una clase canónica, lo cual permite, entre otras cosas, la identificación de secuenciasceptoras apropiadas. La numeración Kabat de la secuencia de aminoácidos de un anticuerpo y las consideraciones estructurales, por ejemplo, como describen Chothia y cols., y su implicación para construir aspectos canónicos de un anticuerpo dado, se describen en la bibliografía (véase también, p. ej., Materiales y procedimientos, más adelante). La expresión "estructura canónica" también se refiere a la conformación de la cadena principal que adopta uno de los bucles de unión a antígeno. A partir de comparaciones estructurales comparativas, se ha encontrado que cinco de los seis bucles de unión a antígeno tienen solo un repertorio limitado de conformaciones disponibles. Cada estructura canónica puede caracterizarse mediante los ángulos de torsión del esqueleto de polipéptido. Los bucles correspondientes entre anticuerpos pueden tener, por tanto, estructuras tridimensionales muy similares, a pesar de la alta variabilidad de secuencia en la mayoría de las partes de los bucles (Chothia y Lesk, 1987 J. Mol. Biol. 196, 901-917 Chothia y cols., 1989 Nature 342, 877-883 Martin y Thornton, 1996 J. Mol. Biol. 263, 800-815). Además, existe una relación entre la estructura de bucle adoptada y las secuencias de aminoácidos que la rodean. La conformación de una clase canónica concreta está determinada por la longitud del bucle y los restos de aminoácido en posiciones clave, que interactúan dentro del bucle y fuera de la región marco conservada. Estos aminoácidos clave suelen interactuar a través de enlaces de hidrógeno. La asignación a una clase canónica concreta puede hacerse, por tanto, sobre la base de la presencia de estos restos de aminoácidos clave.

- 30 La expresión "región CDR definida" se refiere a una región determinante de complementariedad (*complementarity determining region*, CDR), de las cuales tres constituyen el carácter de unión de una región variable de cadena ligera y/o una región variable de cadena pesada de una molécula de unión. Hay tres CDR en cada una de las secuencias variables pesadas y variables ligeras, designadas como CDR1, CDR2 y CDR3, para cada una de las regiones variables. Las regiones CDR definidas contribuyen a la actividad funcional de una molécula de anticuerpo y pueden estar separadas por secuencias de aminoácidos que son simples regiones de armazón o marco conservadas. La definición exacta de los límites y las longitudes de las CDR está sometida a diferentes sistemas de clasificación. Por tanto, puede hacerse referencia a las CDR mediante Kabat, Chothia, contacto, o cualquier otra de las definiciones de los límites. A pesar de los límites diferentes, todas ellas tienen algunos restos que se solapan, en lo que constituyen las llamadas "regiones hipervariables", dentro de las secuencias variables. Estas definiciones de CDR diferirán, por tanto, en la longitud y los límites de sus áreas con respecto a la región marco conservada adyacente. Véase, por ejemplo, Kabat, Chothia, y/o MacCallum y cols., (véase, p. ej., Kabat y cols., en "Sequences of Proteins of Immunological Interest," U.S. Department of Health and Human Services, 1983; Chothia y cols., J. Mol Biol. 196:901-917, 1987; y MacCallum y cols., J. Mol Biol. 262:732-745 (1996).

- 45 La expresión "resto de aminoácido conservado" se refiere un resto de aminoácido determinado que se ha determinado que aparece con una frecuencia entre la de secuencia de línea germinal y la de secuencia CDR, o entre CDR de una clase canónica y/o una longitud dadas, que es elevada, típicamente, al menos 50 por ciento o más (p. ej., de alrededor del 60 por ciento, 70 por ciento, 80 por ciento, 90 por ciento, 95 por ciento, o más), para una posición de un resto dada. Cuando se determina que un resto dado aparece en una frecuencia tan elevada, se determina que está conservado y, por tanto, se representa en las bibliotecas de la invención como un resto "fijo" o "constante", al menos para esa posición del resto de aminoácido en la región CDR que se está analizando. Típicamente, no se introduce ninguna mutagénesis/ variabilidad de ácido nucleico para una posición conservada de aminoácido (codón), sino que, más bien, el resto es fijo y predeterminado

- 55 La expresión "resto de aminoácido semiconservado" se refiere restos de aminoácido que se ha determinado que aparecen con una frecuencia entre la de secuencia de línea germinal y la de secuencia CDR, o entre CDR de una clase canónica y/o una longitud dadas, que es elevada, para 2 a 3 restos para una posición de resto dada. Cuando 2-3 restos, preferentemente 2 restos que, juntos, están representados con una frecuencia de alrededor del 40 % del tiempo o más (p. ej., 50 %, 60 %, 70 %, 80 %, 90 % o más), se determina que los restos están semiconservados y, por tanto, se representan en las bibliotecas de la invención como "semifijos", al menos para esa posición del resto de aminoácido en la región CDR que se está analizando. Típicamente, se introduce un nivel apropiado de mutagénesis/ variabilidad de ácido nucleico para una posición semiconservada de aminoácido (codón) tal que están representados apropiadamente los 2 a 3 restos. Por tanto, puede decirse que cada uno de los 2 a 3 restos son "semifijos" para esta posición.

La expresión "resto de aminoácido variable" se refiere a restos de aminoácido que se ha determinado que aparecen

con una frecuencia entre la de secuencia germinal y la de secuencia CDR, o entre CDR de una clase canónica y/o una longitud dadas, que es variable para una posición de resto dada. Cuando aparecen muchos restos en una posición dada, se determina que la posición de resto es variable y, por tanto, se representa en las bibliotecas de la invención como variable, al menos para esa posición de resto de aminoácido en la región CDR que se está analizando. Típicamente, se introduce un nivel apropiado de mutagénesis/ variabilidad de ácido nucleico para una posición variable de aminoácido (codón) tal que está representado apropiadamente un espectro preciso de restos. Por supuesto, se entiende que, si se desea, las consecuencias de la variabilidad de cualquier posición de resto de aminoácido, es decir, conservada, semiconservada, o variable, pueden representarse, explorarse o alterarse usando, de forma apropiada, cualquiera de los procedimientos de mutagénesis desvelados en el presente documento, p. ej., LTM, WTM, WTM con dopaje, y/o WTM extendida.

La expresión "perfil de variabilidad" se refiere al catálogo de aminoácidos y de sus respectivos índices de frecuencia de aparición presentes en una posición concreta de CDR. Las posiciones de CDR proceden de un conjunto de datos alineados de CDR agrupados de acuerdo con características deseadas. En cada posición de CDR, se añaden frecuencias de aminoácidos ordenadas al perfil de variabilidad de esa posición hasta que las frecuencias combinadas de los aminoácidos alcanzan un valor umbral "elevado" predeterminado.

La expresión "aminoácido" o "resto de aminoácido" se refiere típicamente a un aminoácido cuya definición reconocida en la técnica es la de un aminoácido seleccionado del grupo que consiste en: alanina (Ala); arginina (Arg); asparagina (Asn); ácido aspártico (Asp); cisteína (Cys); glutamina (Gln); ácido glutámico (Glu); glicina (Gly); histidina (His); isoleucina (Ile); leucina (Leu); lisina (Lys); metionina (Met); fenilalanina (Phe); prolina (Pro); serina (Ser); treonina (Thr); triptófano (Trp); tirosina (Tyr); y valina (Val), aunque pueden usarse aminoácidos modificados, sintéticos o raros de la forma deseada. Por lo general, los aminoácidos pueden agruparse como de cadena lateral no polar (p. ej., Ala, Cys, Ile, Leu, Met, Phe, Pro, Val); de cadena lateral con carga negativa (p. ej., Asp, Glu); de cadena lateral con carga positiva (p. ej., Arg, His, Lys); o de cadena lateral polar sin carga (p. ej., Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, y Tyr).

El término "biblioteca" se refiere a dos o más moléculas de anticuerpo (o fragmentos de las mismas) que tienen una diversidad como se describe en el presente documento, mutadas de acuerdo con el procedimiento de la invención. Los anticuerpos de la biblioteca pueden estar en forma de polinucleótidos y polipéptidos, polinucleótidos y polipéptidos en un extracto acelular, o como polinucleótidos y/o polipéptidos en el contexto de un fago, células procariontas, o en células eucariotas.

El término "polinucleótido(s)" se refiere a ácidos nucleicos tales como moléculas de ADN y moléculas de ARN, y análogos de las mismas (p. ej., ADN o ARN generado usando análogos de nucleótidos o usando química de ácidos nucleicos). De forma deseada, los polinucleótidos pueden fabricarse de forma sintética, p. ej., usando química de ácidos nucleicos reconocida en la técnica, o de forma enzimática usando, p. ej., una polimerasa y, si se desea, modificarse. Las modificaciones típicas incluyen metilación, biotilación, y otras modificaciones conocidas en la técnica. Además, la molécula de ácido nucleico puede ser monocatenaria o bicatenaria y, cuando se desee, ligarse a un resto detectable.

El término "mutagénesis" se refiere, a menos que se especifique otra cosa, a cualquier método reconocido en la técnica para alterar la secuencia de un polinucleótido o un polipéptido. Los tipos preferidos de mutagénesis incluyen mutagénesis guiada (WTM), mutagénesis guiada beneficiosa, mutagénesis de recorrido (LTM), mutagénesis de recorrido mejorada (LTM2), WTM usando nucleótidos dopados para conseguir sesgo del codón, WTM extendida para mantener regiones cortas de la secuencia como constantes o fijas dentro de una región de mayor diversidad, o combinaciones de las mismas.

La expresión "mutagénesis beneficiosa combinatoria" se refiere a una biblioteca de combinación de secuencias codificadoras que codifican mezclas degeneradas de variaciones de secuencias de aminoácidos de CDR de V_L y/o V_H identificadas inicialmente a partir de la exploración predeterminada por mutagénesis LTM de aminoácidos, como poseedoras de una alteración en una propiedad mensurable,. En el procedimiento de mutación beneficiosa combinatoria, se generan oligonucleótidos codificadores de secuencias, los cuales representan combinaciones de estas mutaciones beneficiosas identificadas mediante LTM. Estas combinaciones pueden ser combinaciones de diferentes mutaciones beneficiosas dentro de una única CDR, mutaciones dentro de una o más CDR dentro de una única cadena de anticuerpo, o mutaciones dentro de CDR de diferentes cadenas de anticuerpo.

Descripción detallada

Visión de conjunto

Los anticuerpos son poderosas herramientas diagnósticas y terapéuticas. Son deseables las bibliotecas de anticuerpos que comprenden moléculas de unión candidatas que pueden seleccionarse fácilmente contra dianas. La promesa firme de una biblioteca universal de anticuerpos extensa ha sido incierta. Las bibliotecas sintéticas están afectadas por ruido de fondo y demasiada diversidad, que no aparece de forma natural. Las bibliotecas totalmente humanas están sesgadas hacia ciertas clases antigénicas y solo son tan diversas como lo permitan los métodos de captura. La presente invención proporciona una biblioteca universal de anticuerpos que es extensa y puede

explorarse fácilmente usando, por ejemplo, procedimientos de alto rendimiento para obtener nuevos agentes terapéuticos.

En particular, la biblioteca universal de anticuerpos (BUA) tiene el potencial de reconocer cualquier antígeno. Otras ventajas significativas de la biblioteca incluyen mayor diversidad, por ejemplo, frente a autoantígenos que suelen perderse en una biblioteca humana expresada porque el sistema inmunitario del donante elimina los anticuerpos autorreactivos mediante selección negativa. Otro aspecto es que la exploración de la biblioteca universal de anticuerpos (BUA) usando selección positiva de clones mediante FACS (*fluorescence activated cell sorter*, clasificador de células activadas por fluorescencia) evita la metodología convencional y tediosa de generar una biblioteca de hibridomas y la exploración del sobrenadante. Más aún, la biblioteca BUA puede reexplorarse para descubrir anticuerpos adicionales contra otras dianas deseadas.

1.1 Identificación y selección de componentes universales de anticuerpo mediante el uso de bioinformática

La primera etapa de la construcción de una biblioteca universal de anticuerpos (BUA) desvelada en el presente documento es la selección de secuencias que cumplan ciertos criterios predeterminados. Por ejemplo, la base de datos Kabat, una base de datos electrónica que contiene secuencias reordenadas no redundantes de anticuerpo, puede consultarse en relación a esas secuencias, que son las representadas con mayor frecuencia, en particular, contra una clase antigénica concreta. La clase antigénica puede incluir, por ejemplo, antígenos de proteína y péptido, pero también moléculas pequeñas, polisacáridos, y polinucleótidos. Se realiza un análisis de grupos de las secuencias de regiones marco conservadas de estos anticuerpos seguido de una comparación (usando el algoritmo de investigación de BLAST) con las secuencias de línea germinal (base de datos V BASE) para determinar las familias de línea germinal usadas con mayor frecuencia, que posteriormente se reordenan para generar anticuerpos funcionales que reconocen una clase antigénica dada, por ejemplo, antígenos o dianas proteicos.

Después se eligen las secuencias de regiones marco conservadas candidatas que representan a los grupos mayores y estructuralmente más diversos de anticuerpos funcionales, y después se determinan las estructuras canónicas de CDR1 y CDR2, para determinar la longitud de las CDR y, por tanto, la diversidad que puede tener cabida dentro de las regiones marco conservadas. Para CDR3, se realiza una distribución por tamaño de las longitudes para identificar un análisis de frecuencia de secuencias reordenadas de anticuerpo.

El procedimiento para obtener secuencias de aminoácidos a partir de las CDR incluye un análisis de frecuencia y la generación de los correspondientes perfiles de variabilidad (PV) de las secuencias reordenadas de anticuerpo existentes. Las posiciones invariables son fijas mientras que los aminoácidos de mayor frecuencia se eligen como naturales en otras posiciones. Estos aminoácidos naturales se alteran después de forma sistemática usando mutagénesis, p. ej., mutagénesis guiada (WTM), para generar la biblioteca universal de anticuerpos.

La estrategia de construcción de la biblioteca universal implica la selección de secuencias de regiones marco conservadas seguida del diseño de los bucles hipervariables de CDR. Para la selección de secuencias de regiones marco conservadas, se dispone un subgrupo de todos los armazones disponibles de regiones marco conservadas que se ha determinado que se han expresado en respuesta frente a un antígeno concreto. Mediante la determinación de las regiones marco conservadas que se expresan con mayor frecuencia en la naturaleza en respuesta frente a una clase antigénica dada se selecciona una región marco conservada aceptora apropiada. Por ejemplo, para determinar las regiones marco conservadas aceptoras preferidas que se expresan en respuesta frente a antígenos a base de proteínas, se buscan regiones marco conservadas "dirigidas a proteínas" en la base de datos Kabat (accesible en <http://www.kabatdatabase.com>). Si se necesitan secuencias aceptoras preferidas para presentar CDR contra una clase antigénica diferente, y/o secuencias aceptoras de una especie concreta, se configura el filtro Kabat de secuencia de proteína como corresponda. Por ejemplo, para determinar secuencias para su uso como agentes terapéuticos humanos contra dianas a base de proteínas, el filtro se configura para centrarse únicamente en secuencias de anticuerpos humanos (no en secuencias de ratón, rata, o pollo, etc.) que reconocen antígenos de proteína/péptido. Esto reduce enormemente la redundancia en el conjunto de datos y la información de secuencia que sesgaría los resultados.

La etapa anterior reduce al mínimo la necesidad de generar numerosos armazones sintéticos diferentes de regiones marco conservadas y da como resultado típicamente un conjunto de datos de aceptores potenciales del alrededor de 600 secuencias o menos. Por consiguiente, el número resultante de secuencias es fácilmente manejable en análisis posteriores para determinar las secuencias precursoras de línea germinal que dan lugar a las secuencias gónicas reordenadas que se seleccionan por clase antigénica. Esta segunda determinación del origen de la línea germinal perfecciona la selección de las secuencias de anticuerpo que han sido seleccionados por una clase antigénica, porque identifica si hay secuencias de regiones marco conservadas de línea germinal óptimas (o de alta frecuencia) que estén sobrerrepresentadas. De hecho, se ha observado que en algunas respuestas policlonales contra ciertos antígenos, en las que se produce un gran número de secuencias reordenadas de anticuerpo, solo se usan unas cuantas secuencias de región marco conservada aceptora. En tal caso, la secuencia de anticuerpo y la diversidad de unión a los antígenos se localiza sobre todo en las CDR, no en las regiones marco conservadas. El análisis bioinformático anterior se centra en los genes V_H a efectos descriptivos, pero se entenderá que tanto los genes de V_λ como los de V_κ se evalúan de forma similar.

1.2 Estrategias de diseño para aumentar al máximo la diversidad de CDR

La elección de regiones marco conservadas candidatas basada en los criterios desvelados en el presente documento impone tanto los tamaños de CDR a introducir como la diversidad inicial de secuencia de aminoácidos. Cuando en las secuencias de anticuerpo se identifican 1) frecuencia de aparición contra una clase antigénica y 2) frecuencia de línea germinal, las secuencias pueden disponerse después de acuerdo con su clase canónica. La clase canónica se determina usando las convenciones, como describe Chothia (véase Materiales y procedimientos, más adelante). La mayoría de las secuencias de anticuerpo identificadas de un conjunto de secuencias de anticuerpo dado puede estar incluida dentro de una cierta clase canónica. La clase canónica impone después el número de restos de aminoácido que pueden tener cabida en las CDR. Por ejemplo, si la clase canónica es 1-3, entonces la CDR1 tendría un bucle de 6 aminoácidos y la CDR2 tendría un bucle de 13 aminoácidos. Para la secuencia variable de cadena pesada, la contribución de la secuencia del segmento J está relativamente bien conservada de modo que, típicamente, solo es necesario considerar la secuencia que se ajuste mejor de entre un subgrupo de solo seis secuencias. Un análisis de frecuencia de aminoácidos de CDR de las bases de datos Kabat y V BASE permite la identificación de las posiciones de restos de aminoácidos de CDR que están incluidas en dos categorías, p. ej., 1) posiciones que deberían estar conservadas, y 2) posiciones que son adecuadas para generación de diversidad.

Al diseñar la VH-CDR2, el análisis de diversidad en las bases de datos V BASE y Kabat se plantea de una forma similar al que se realizó anteriormente para VH-CDR.

Al diseñar la diversidad de las V_H CDR3, las secuencias CDR3 de los anticuerpos de la base de datos Kabat se alinean de acuerdo con su tamaño y clase antigénica. Las longitudes de las CDR3 de anticuerpos que reconocen antígenos no proteicos y de proteína/péptido se comparan y se realiza un análisis de frecuencia y se usa una frecuencia umbral del 10 % para identificar las secuencias más favorables para usar en la designación de la diversidad de las CDR3. Como el análisis del tamaño y de frecuencia de restos de aminoácidos de CDR3 se realiza usando, p. ej., las secuencias génicas reordenadas (D) y J de inmunoglobulina, no existen equivalentes de "CDR3" de línea germinal para las comparaciones directas de Kabat filtrada y V BASE. Sin embargo, puede generarse un análisis de frecuencia de Kabat filtrado o un perfil de variabilidad (PV) (figuras 11 y 12) para cada tamaño de CDR3 reordenada, el cual revela, para cada clasificación por tamaño, el aminoácido más frecuente a lo largo de las posiciones de la CDR3 y da como resultado una secuencia "natural" de consenso. Sorprendentemente, este planteamiento de "consenso" o de "frecuencia" identifica aquellos aminoácidos concretos bajo presión selectiva elevada. Por consiguiente, estas posiciones de restos son típicamente fijas, introduciéndose la diversidad dentro de las posiciones de aminoácido restantes (teniendo en cuenta la preferencia identificada en ciertos aminoácidos de estar presentes en estas posiciones).

Cuando se diseña la diversidad para cualquiera de las CDR mencionadas anteriormente, pueden incorporarse en las CDR de la forma deseada restos de aminoácidos modificados, por ejemplo, restos fuera de los 20 aminoácidos tradicionales usados en la mayoría de los polipéptidos, p. ej., homocisteína. Esto se lleva a cabo usando métodos reconocidos en la técnica, los cuales introducen típicamente codones de terminación dentro del polinucleótido en el que se desea el resto de aminoácido modificado. El método proporciona entonces un ARNt modificado ligado al aminoácido modificado que se va a incorporar (un ARNt denominado ARNt supresor de, p. ej., el codón de terminación ámbar, ópalo, u ocre) dentro del polipéptido (véase, p. ej., Kohrer y cols., Import of amber and ochre suppressors tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins, PNAS, 98, 14310-14315 (2001)).

2. Construcción de biblioteca universal de anticuerpos (BUA) asistida por ordenador

Las bibliotecas universales de anticuerpos de la invención y su construcción se han efectuado con el beneficio de la información de secuencia y estructural relativa a la diversidad del anticuerpo que se va a generar, de modo que el potencial para la generación de anticuerpos mejorados aumenta. La información de la elaboración de modelos también puede usarse para guiar la selección de la diversidad de aminoácidos que se va a introducir en las regiones definidas, p. ej., las CDR. Más aún, los resultados reales obtenidos con los anticuerpos de la invención pueden guiar la selección (o exclusión), p. ej., maduración de la afinidad, de anticuerpos posteriores que se van a fabricar y explorar de una manera iterativa.

En una divulgación particular, la elaboración de modelos por ordenador se usa para eliminar la producción de cualquiera de los anticuerpos para los que se haya predicho una estructura o función deficiente o no deseada. De esta forma, el número de anticuerpos que se va a producir puede reducirse de forma precisa, aumentando así la relación señal-ruido de fondo en los ensayos de exploración posteriores. En otra divulgación concreta, la elaboración de modelos por ordenador se actualiza continuamente con información de elaboración de modelos adicional de cualquier fuente pertinente, p. ej., de bases de datos de secuencias génicas y de proteínas y tridimensionales y/o resultados de anticuerpos analizados con anterioridad, de modo que la base de datos informatizada se hace más precisa en cuanto a su capacidad predictiva (fig. 1).

En otra divulgación más, se proporciona la base de datos informatizada con los resultados del ensayo, p. ej., afinidad de unión/ avidez de los anticuerpos analizados previamente, y categoriza los basándose en el criterio o criterios del

ensayo, como de respuesta positiva o de respuesta negativa, p. ej., como anticuerpos que se unen bien o no tan bien. De esta forma, la maduración de la afinidad de la invención puede equiparar una gama de respuestas funcionales con una información de secuencia y estructural concreta y usar dicha información para dirigir la producción de futuros anticuerpos para analizar. El procedimiento es especialmente adecuado para explorar una

5 afinidad de unión concreta a un antígeno diana, de anticuerpos o fragmentos de anticuerpo usando, p. ej., un ensayo Biacore.

Por consiguiente, la mutagénesis de restos no contiguos dentro de una región puede ser deseable si se conoce, p. ej., a través de la elaboración de modelos por ordenador, que ciertos restos de la región no participarán en la función deseada. La estructura coordinada y la interrelación espacial entre las regiones definidas, p. ej., los restos

10 funcionales de aminoácidos en las regiones definidas del anticuerpo, p. ej., la diversidad que se ha introducido, puede considerarse y elaborarse un modelo. Dichos criterios de elaboración del modelo incluyen, p. ej., química de los grupos laterales de los restos de aminoácidos, distancias atómicas, datos cristalográficos, etc. Por consiguiente, el número de anticuerpos que se va a producir puede reducirse al mínimo de forma inteligente.

En una divulgación preferida, una o más de las etapas anteriores son asistidas por ordenador. En una divulgación concreta, la etapa asistida por ordenador comprende, p. ej., minería de la base Kabat y, opcionalmente, el cruce de los resultados con V BASE, por medio de lo cual, se determinan ciertos criterios desvelados y se usan para diseñar la diversidad deseada de CDR (Fig. 1-2). El procedimiento también está abierto a su realización total o parcial mediante un dispositivo, p. ej., un dispositivo controlado por ordenador. Por ejemplo, la selección de secuencia de anticuerpo por minería de base de datos, el diseño de diversidad, la síntesis de oligonucleótidos, el ensamblaje de los anteriores mediado por PCR, y la expresión y selección de los anticuerpos candidatos que se unen a una diana

20 dada, pueden llevarse a cabo total o parcialmente por dispositivos entrelazados. Además, las instrucciones para llevar a cabo el procedimiento total o parcialmente pueden conferirse a un medio adecuado para su uso en un dispositivo electrónico para llevar a cabo las instrucciones. En resumen, los procedimientos desvelados de la invención pueden adaptarse a un planteamiento de alto rendimiento que comprenda un programa informático (p. ej., instrucciones legibles por ordenador) y un disco duro (p. ej., ordenadores, robots, y chips).

3. Síntesis de bibliotecas universales de anticuerpos

En una divulgación, las bibliotecas universales de anticuerpos (BUA) se generan para exploración mediante la síntesis de oligonucleótidos individuales que codifican la región definida del polipéptido y no tienen más de un codón para el aminoácido predeterminado. Esto se consigue mediante la incorporación en cada posición del codón dentro del oligonucleótido, bien del codón requerido para la síntesis del polipéptido natural, o bien de un codón para el aminoácido predeterminado, y se denomina mutagénesis de recorrido (LTM) (véase, p. ej., el documento U.S.S.N. 60/483282).

30

En otra divulgación, cuando se requiere diversidad en múltiples posiciones de aminoácidos, puede usarse mutagénesis guiada (WTM) (véanse, p. ej., las patentes de EE. UU. nº 6.649.340; 5.830.650; y 5.798.208; y el documento U.S.S.N. 60/483,282. La WTM permite realizar múltiples mutaciones con número mínimo de oligonucleótidos. Los oligonucleótidos pueden producirse de forma individual, en lotes, usando, p. ej., técnicas de dopaje, y después mezclarse o agruparse de la forma deseada.

35

La mezcla de oligonucleótidos para la generación de la biblioteca puede sintetizarse fácilmente mediante procedimientos conocidos para la síntesis de ADN. El procedimiento preferido implica el uso de química de beta cianoetil fosforamidito en fase sólida (p. ej., véase la patente de EE.UU. nº No. 4.725.677). Por comodidad, puede usarse un instrumento para síntesis automática de ADN que contenga recipientes de reactivos especificados de nucleótidos. Los polinucleótidos también pueden sintetizarse para contener sitios de restricción o sitios de hibridación de cebadores para facilitar la introducción o ensamblaje de los polinucleótidos que representan, p. ej., una región definida, dentro de un contexto genético mayor.

40

Los polinucleótidos sintetizados pueden insertarse dentro de un contexto genético mayor, p. ej., un anticuerpo monocatenario (scFv), usando técnicas convencionales de ingeniería genética. Por ejemplo, puede hacerse que los polinucleótidos contengan sitios adyacentes de reconocimiento para enzimas de restricción (p. ej., véase la patente de EE.UU. nº No. 4.888.286). Los sitios de reconocimiento pueden diseñarse para corresponder a sitios de reconocimiento, bien que existen de forma natural, o bien que se introducen en el gen próximo a la región codificadora de ADN. Tras la conversión a la forma bicatenaria, los polinucleótidos se ligan dentro del gen o del vector génico mediante técnicas convencionales. Por medio de un vector apropiado (incluyendo, p. ej., vectores de fago, plásmidos) los genes pueden introducirse dentro de un extracto acelular, fago, célula procariota, o célula eucariota adecuada para la expresión de los anticuerpos.

45

Como alternativa, se diseñan polinucleótidos que solapan parcialmente, típicamente de alrededor de 20-60 nucleótidos de longitud. Los polinucleótidos internos se aparean después con su pareja complementaria para dar una molécula de ADN bicatenario con elongaciones monocatenarias útiles para el apareamiento posterior. Las parejas apareadas pueden mezclarse después juntas, elongarse, y ligarse para formar moléculas bicatenarias de longitud completa usando PCR (véase, p. ej., el ejemplo 3). Pueden diseñarse sitios de restricción convenientes cerca de los extremos del gen sintético para la clonación en un vector adecuado. Las moléculas de longitud

55

completa pueden ligarse después dentro de un vector adecuado.

Cuando en el ensamblaje genético se usan polinucleótidos que solapan parcialmente, también puede incorporarse directamente un conjunto de nucleótidos degenerados en el lugar de uno de los polinucleótidos. La cadena complementaria apropiada se sintetiza durante la reacción de elongación a partir de un polinucleótido de la otra cadena parcialmente complementario mediante elongación enzimática con una polimerasa. La incorporación de los polinucleótidos degenerados en la etapa de síntesis también simplifica la clonación, en la que más de un dominio o región definida de un gen se muta o se modifica por ingeniería genética para tener diversidad.

En otro planteamiento, el anticuerpo está presente en un plásmido monocatenario. Por ejemplo, el gen puede clonarse dentro de un vector de fago o un vector con un origen de replicación de fago filamentosos que permite la propagación de moléculas monocatenarias con el uso de un fago auxiliar. El molde monocatenario puede aparearse con un conjunto de polinucleótidos degenerados que representan las mutaciones deseadas, y elongarse y ligarse, incorporando, por tanto, cada cadena análoga dentro de una población de moléculas que pueden introducirse en un anfitrión adecuado (véase, p. ej., Sayers, J. R. y cols., *Nucleic Acids Res.* 16: 791-802 (1988)). Este planteamiento puede evitar múltiples etapas de clonación en las que se seleccionan múltiples dominios para mutagénesis.

También puede usarse metodología de reacción en cadena de polimerasa (PCR) para incorporar polinucleótidos dentro de un gen, por ejemplo, diversidad de CDR dentro de regiones marco conservadas. Por ejemplo, los propios polinucleótidos pueden usarse como cebadores para la elongación. En este planteamiento, los polinucleótidos que codifican los casetes mutagénicos correspondientes a la región definida (o porción de la misma) son complementarios entre sí, al menos parcialmente, y pueden elongarse para formar un casete génico grande (p. ej., un scFv) usando una polimerasa, p. ej., usando amplificación por PCR.

El tamaño de la biblioteca variará dependiendo de la longitud de la CDR y de la cantidad de diversidad de CDR que se necesite representar usando, p. ej., WTM o LTM. Preferentemente, la biblioteca se diseñará para contener menos de 10^{15} , 10^{14} , 10^{13} , 10^{12} , 10^{11} , 10^{10} , 10^9 , 10^8 , 10^7 , y más preferentemente, 10^6 anticuerpos o menos.

La descripción anterior se ha centrado en la representación de la diversidad de anticuerpo mediante la alteración del polinucleótido que codifica el polipéptido correspondiente. Se entiende, sin embargo, que el ámbito de la divulgación también abarca procedimientos de representación de la diversidad de anticuerpo desvelados en el presente documento mediante la síntesis directa de las regiones de polipéptido deseadas usando química de proteínas. Al llevar a cabo este planteamiento, los polipéptidos resultantes siguen incorporando los aspectos desvelados, exceptuando que puede eliminarse el uso de un polinucleótido intermedio.

Para las bibliotecas descritas anteriormente, sin importar si es en forma de polinucleótidos y/o de los polipéptidos correspondientes, se entiende que las bibliotecas también pueden fijarse a un soporte sólido, tal como una micromatriz, y, preferentemente, disponerse, usando métodos reconocidos en la técnica.

El procedimiento de esta divulgación es especialmente útil para modificar moléculas de anticuerpo candidatas por medio de maduración de la afinidad. Pueden introducirse alteraciones dentro de la región variable y/o dentro de la región marco conservada (constante) de un anticuerpo. La modificación de la región variable puede producir anticuerpos con mejores propiedades de unión a antígeno, y, si se desea, propiedades catalíticas. La modificación de la región marco conservada también puede conducir a la mejora de propiedades quimiofísicas, tales como la solubilidad o la estabilidad, que son especialmente útiles, por ejemplo, en la producción comercial, la biodisponibilidad, y la afinidad por 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 está constituida de regiones variables de dos cadenas, una de la cadena pesada (VH) y una de la cadena ligera (VL). Una vez que se han identificado las características de unión a antígeno deseadas, la(s) región(es) variables pueden modificarse por ingeniería genética dentro de una clase de anticuerpo apropiada, tal como IgG, IgM, IgA, IgD, o IgE. En una divulgación preferida, una molécula de unión candidata identificada se somete a maduración de la afinidad para aumentar la afinidad/avidez de la molécula de unión por una diana/antígeno.

4. Sistema de expresión y exploración

Las bibliotecas de polinucleótidos generadas mediante cualquiera de las técnicas anteriores u otras técnicas adecuadas pueden expresarse y explorarse para identificar anticuerpos que tengan la estructura y/o la actividad deseadas. La expresión de los anticuerpos puede llevarse a cabo usando extractos acelulares (y, p. ej., presentación en ribosoma, presentación en fagos, células procariotas, o células eucariotas (p. ej., presentación en levaduras).

En una divulgación, los polinucleótidos se modifican por ingeniería genética para servir como moldes que pueden expresarse en un extracto acelular. Pueden usarse los vectores y extractos como se describe, por ejemplo, en las patentes de EE.UU. nº 5.324.637; 5.492.817; 5.665.563, y muchos están disponibles en el mercado. Pueden usarse la presentación en ribosoma y otras técnicas acelulares para ligar un polinucleótido (es decir, un genotipo) a un polipéptido (es decir, un fenotipo), p. ej., Profusion™ (véanse, p. ej., las patentes de EE.UU. nº 6.348.315; 6.261.804; 6.258.558; y 6.214.553).

Como alternativa, los polinucleótidos de la invención pueden expresarse en un sistema de expresión de *E. coli*

conveniente, tal como el que describen Pluckthun y Skerra. (Pluckthun, A. y Skerra, A., Meth. Enzymol. 178: 476-515 (1989); Skerra, A. y cols., Biotechnology 9: 273-278 (1991)). Las proteínas mutantes pueden expresarse para su secreción en el medio y/o en el citoplasma de las bacterias como describen M. Better y A. Horwitz, Meth. Enzymol. 178: 476 (1989). En una divulgación, los dominios individuales que codifican las VH y las VL se fijan cada uno al extremo 3' de una secuencia que codifica una secuencia de señal, tal como la secuencia de señal ompA, phoA o pelB (Lei, S. P. y cols., J. Bacteriol. 169: 4379 (1987)). Estos genes de fusión se ensamblan en una construcción dicistrónica, de modo que pueden expresarse a partir de un solo vector, y segregarse dentro del espacio periplásmico de *E. coli*, donde se renaturalizarán y podrán recuperarse en forma activa (Skerra, A. y cols., Biotechnology 9: 273-278 (1991)). Por ejemplo, los genes de cadena pesada de anticuerpo pueden expresarse de forma simultánea con los genes de cadena ligera del anticuerpo para producir un anticuerpo o fragmentos de anticuerpo.

En otra divulgación, las secuencias de anticuerpo se expresan sobre la superficie de la membrana de un procarionte, p. ej., *E. coli*, usando una señal de secreción y un resto de lipidación como se describe, p. ej., en los documentos US20040072740A1; US20030100023 A1; y US20030036092A1.

En otra divulgación más, los polinucleótidos pueden expresarse en células eucariotas, tales como levaduras usando, por ejemplo, presentación en levaduras como se describe, p. ej., en las patentes de EE.UU. n° 6.423.538; 6.331.391; y 6.300.065. En este planteamiento, los anticuerpos de la biblioteca (p. ej., scFv) se fusionan a un polipéptido que se expresa y se presenta sobre la superficie de la levadura.

También pueden usarse células eucariotas superiores para la expresión de los anticuerpos de la invención, tales como células de mamífero, por ejemplo, células de mieloma (p. ej., células NS/O), células de hibridoma, o células de ovario de hámster chino (CHO, de *Chinese hamster ovary*). Típicamente, cuando los anticuerpos se expresan en células de mamífero, se diseñan para expresarse dentro del medio de cultivo, o expresarse sobre la superficie de dicha célula. El anticuerpo o los fragmentos de anticuerpo pueden producirse, por ejemplo, como moléculas de anticuerpo enteras o como fragmentos VH y VL individuales, fragmentos Fab, dominios individuales, o como cadenas individuales (sFv) (véase, p. ej., Huston, J. S. y cols., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988)).

La exploración de los anticuerpos expresados (o de los anticuerpos producidos por síntesis directa) puede hacerse mediante cualquier medio apropiado. Por ejemplo, la actividad de unión puede evaluarse mediante inmunoensayo y/o cromatografía de afinidad convencionales. La exploración de la función catalítica de los anticuerpos de la invención, p. ej., la función proteolítica, puede conseguirse usando un ensayo convencional en placa de hemoglobina, como se describe, por ejemplo, en la patente de EE.UU. n° 5.798.208. La determinación de la capacidad de los anticuerpos candidatos para unirse a dianas terapéuticas puede analizarse *in vitro* usando, p. ej., un equipo Biacore, que mide los índices de unión de un anticuerpo a una diana o antígeno dados. Pueden efectuarse ensayos *in vivo* usando uno cualquiera de una serie de modelos animales y analizarse posteriormente, de forma apropiada, en seres humanos.

Ejemplificación

A lo largo de los ejemplos, se usaron los siguientes materiales y procedimientos, a menos que se manifieste otra cosa.

Materiales y procedimientos

En general, la práctica de la presente invención emplea, a menos que se indique otra cosa, técnicas convencionales de química, biología molecular, tecnología de ADN recombinante, tecnología de PCR, inmunología (especialmente, p. ej., tecnología de anticuerpos), sistemas de expresión (p. ej., expresión acelular, presentación en fagos, presentación en ribosoma, y Profusion™), y cualquier cultivo celular necesario que esté dentro de la habilidad de la técnica y se explique en la bibliografía. Véase, p. ej., Sambrook, Fritsch y Maniatis, Molecular Cloning: Cold Spring Harbor Laboratory Press (1989); DNA Cloning, vol. 1 y 2, (D.N. Glover, ed. 1985); Oligonucleotide Synthesis (M.J. Gait, ed. 1984); PCR Handbook Current Protocols in Nucleic Acid Chemistry, Beaucage, ed. John Wiley & 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 cols., Academic Press (1990); PCR Essential Techniques: Essential Techniques, Burke, ed., John Wiley & Son 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 cols., C.S.H.L. Press, Pub. (1999); Current Protocols in Molecular Biology, ed. Ausubel y cols., John Wiley & Sons (1992); Large-Scale Mammalian Cell Culture Technology, Lubiniecki, 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 cols., Yeast surface display for screening combinatorial polypeptide libraries, Nature Biotechnology, 15(6):553-7 (1997); Border y cols., Yeast surface display for directed evolution of protein expression, affinity, and stability, Methods Enzymol., 328:430-44 (2000); presentación en ribosoma como describen Pluckthun y cols. en la patente de EE.UU. n° 6.348.315, y Profusion™ como describen Szostak y cols., en las patentes de EE.UU. n° 6.258.558; 6.261.804; y 6.214.553; y expresión periplásmica bacteriana como se describe en el documento US20040058403A1.

Pueden encontrarse más detalles en relación al análisis de secuencia de anticuerpo usando las convenciones Kabat, p. ej., en Johnson y cols., The Kabat database and a bioinformatics example, *Methods Mol Biol.* 2004; 248:11-25; Johnson y cols., Preferred CDRH3 lengths for antibodies with defined specificities, *Int Immunol.* 1998, Dec;10(12):1801-5; Johnson y cols., SEQHUNT. A program to screen aligned nucleotide and amino acid sequences, *Methods Mol Biol.* 1995;51:1-15; y Wu y cols., Length distribution of CDRH3 in antibodies; y Johnson y cols., *Proteins.* 1993 May;16(1):1-7. Revisión).

Pueden encontrarse más detalles en relación al análisis de secuencia de anticuerpo usando las convenciones Chothia, p. ej., in Chothia y cols., Structural determinants in the sequences of immunoglobulin variable domain, *J Mol Biol.* 1998 May 1;278(2):457-79; Morea y cols., Antibody structure, prediction and redesign, *Biophys Chem.* 1997 Oct;68(1-3):9-16; Morea y cols., Conformations of the third hypervariable region in the VH domain of immunoglobulins; *J Mol Biol.* 1998 Jan 16;275(2):269-94; Al-Lazikani y cols., Standard conformations for the canonical structures of immunoglobulins, *J Mol Biol.* 1997 Nov 7;273(4):927-48. Barre y cols., Structural conservation of hypervariable regions in immunoglobulins evolution, *Nat Struct Biol.* 1994 Dec;1(12):915-20; Chothia y cols., Structural repertoire of the human VH segments, *J Mol Biol.* 1992 Oct 5;227(3):799-817 Conformations of immunoglobulin hypervariable regions, *Nature.* 1989 Dec 21-28;342(6252):877-83; y Chothia y cols., Revisión. Canonical structures for the hypervariable regions of immunoglobulins, *J Mol Biol.* 1987 Aug 20;196(4):901-17).

Se describen más detalles en relación con el análisis de Chothia, por ejemplo, en Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J Mol Biol.* 1998 Jan 16;275(2):269-94; Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. Structural repertoire of the human VH segments. *J Mol Biol.* 1992 Oct 5;227(3):799-817; Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, y cols. Conformations of immunoglobulin hypervariable regions. *Nature.* 1989 Dec 21-28;342(6252):877-83; Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol.* 1987 Aug 20;196(4):901-17; y Chothia C, Lesk AM. The evolution of protein structures. *Cold Spring Harb Symp Quant Biol.* 1987;52:399-405.

Se describen más detalles en relación con las consideraciones sobre el contacto de las CDR, por ejemplo, en MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site Topography. *J Mol Biol.* 1996 Oct 11;262(5):732-45.

Se describen más detalles en relación con las secuencias de anticuerpo y las bases de datos a las que se hace referencia en el presente documento, p. ej., en Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. *J Mol Biol.* 1992 Oct 5;227(3):776-98; Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics.* 2001 Mar;17(3):282-3; [VBDB] www.mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901; [KBTDB] www.lcabatdatabase.com; [BLST] www.ncbi.nlm.nih.gov/BLAST/ [CDHIT] bioinformatics.ljcrf.edu/cd-hi/; [EMBOSS] www.hgmp.mrc.ac.uk/Soft-ware/EMBOSS/; [PHYLIP] evolution.genetics.washington.edu/phylip.html; and [FASTA] fasta.bioch.virginia.edu.

Las bibliotecas bacterianas de expresión se construyen típicamente de la forma siguiente. La secuencia molde (fig. 18) se clona dentro de un vector de expresión-presentación apropiado, tal como el sistema de presentación y expresión APEX descrito en Harvey y cols. *PNAS* 101 (25): 9193. (2004). Las bibliotecas de mutagénesis guiada y de mutagénesis guiada extendida se preparan mediante mutagénesis de Kunkel de la construcción preparada con la secuencia incorporada dentro del vector APEX. Se preparó un molde monocatenario para mutagénesis de Kunkel usando protocolos convencionales (Sidhu, S. S. (2000) *Methods Enzymol.* 328:333-63). La mutagénesis de Kunkel del molde se llevó a cabo de acuerdo con procedimientos convencionales, como se detalla, por ejemplo, en Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* 82:488-92; Kunkel, T. A. y cols. (1987) *Meth. Enzymol.* 154: 367-82; Zoller, M. J. y Smith, M. (1983) *Meth. Enzymol.* 100:468-500; Hanahan, D. (1983) *J. Mol. Biol.* 166:557-80; y Maniatis, T., Fritsch, E. F. y Sambrook, J. (1989) en *Molecular Cloning, A Laboratory Manual*.

La fig. 26 muestra las etapas generales en el procedimiento de mutagénesis de Kunkel para la introducción de una colección de oligonucleótidos de biblioteca de CDR dentro de una secuencia molde de codificadora de anticuerpo. Inicialmente, el molde monocatenario con uracilo se hace reaccionar con una colección de oligonucleótidos L3 (fragmento verde) que portan las sustituciones del codón seleccionado para CDRL3. Para el molde que utiliza la SEC ID N° 1, hay dos tamaños de CDRL3: Tamaño 8 y tamaño 9. Para cada uno de estos tamaños, cada uno de los aminoácidos guiados generará una mezcla única de oligonucleótidos. Se muestran nueve mezclas de oligonucleótidos guiados para el tamaño 8 (VKIII_3_8*), y nueve mezclas guiadas para el tamaño 9 (VKIII_3_9*). Las 18 mezclas de oligonucleótidos se combinan de forma equimolar y se realiza la mutagénesis de Kunkel (ETAPA 1) como se ha descrito previamente (Sidhu, S.S. (2000) *Methods Enzymol.* 328:333-63). Típicamente, una reacción con 10 microgramos de molde monocatenario dará un tamaño de biblioteca de 10^8 - 10^9 transformantes. Esta reacción de mutagénesis se transforma dentro de células DH5 alfa y se realiza una maxipreparación sobre la colección de biblioteca de CDRL3.

Esta colección de biblioteca L3 de ADN se transforma dentro de células CJ236 para la preparación de un molde de biblioteca L3 monocatenaria (ETAPA 2). Esto crea el molde monocatenario para la incorporación de mutagénesis adicionales de CDR. Los oligonucleótidos mutados de la biblioteca H3 (ETAPA 3) se hibridan con el molde de

CDRL3 de la biblioteca. Para el molde que utiliza la secuencia mostrada en la fig. 18, hay diez longitudes de CDRH3 usadas en el diseño inicial: Los tamaños 9-18. Para cada tamaño de CDRH3, se realizan reacciones separadas. Por tanto, en la etapa 3, se realizan diez reacciones separadas utilizando 20 microgramos de molde monocatenario para cada reacción. Dentro de cada tamaño de CDRH3, se utilizan nueve aminoácidos guiados. Por tanto, se mezclan juntos nueve oligonucleótidos degenerados para cada reacción, igual que para el tamaño 9 (VH_3_9*) y un número similar para el tamaño 10 (VH_3_10*), etc. Cada reacción con 20 microgramos de molde monocatenario da más de 10^9 transformantes. Por tanto, para esta biblioteca, que contiene diez tamaños de CDR, la biblioteca total de CDRL3/CDRH3 es mayor de 10^{10} transformantes totales. Cada reacción de mutagénesis se transforma dentro de células DH5 alfa, y tras la maxipreparación del ADN de plásmido, el ADN puede transformarse dentro de una línea celular de expresión-presentación apropiada, tal como las células DH12S para la presentación y exploración APEX (como se describe en Harvey y cols. PNAS 101 (25): 9193. 2004)) o puede volver a transformarse el plásmido dentro de células CJ236 para la incorporación posterior de mutaciones en otras CDR, tales como CDRH1, CDRH2, CDRL1 y CDRL2 (ETAPA 4).

Ejemplo 1

PROCEDIMIENTOS PARA LA IDENTIFICACIÓN GUIADA POR BIOINFORMÁTICA DE SECUENCIAS DE BIBLIOTECAS UNIVERSALES DE ANTICUERPOS

En este ejemplo, las secuencias de bibliotecas universales de anticuerpos se identifican y se seleccionan usando bioinformática y los criterios desvelados en el presente documento.

Brevemente, se buscó en la base electrónica Kabat, que contiene secuencias expresadas, es decir, reordenadas, de inmunoglobulina, usando ciertos algoritmos de filtro. En particular, los algoritmos de filtro se diseñaron para identificar únicamente secuencias humanas que se hubieran expresado en respuesta a una clase antigénica concreta. La clase antigénica seleccionada eran antígenos/dianas a base de proteína porque éste es un conjunto de dianas manejable para el desarrollo de agentes terapéuticos humanos. Se entiende, sin embargo, que la base de datos se consulta con igual facilidad para otras clases antigénicas, p. ej., péptidos, polisacáridos, polinucleótidos, y moléculas pequeñas, así como para secuencias de anticuerpo procedentes de otras especies, como secuencias de primates, ratón, rata, o pollo, etc., para el desarrollo de, p. ej., agentes terapéuticos para aplicación veterinaria. Los criterios anteriores se aplicaron a un conjunto inicial de 5971 secuencias de V_H (se destaca, sin embargo, que este conjunto de secuencias puede aumentar en número a medida que se clonan y se introducen nuevas secuencias en la base de datos).

La búsqueda y análisis de filtro anteriores devolvieron un conjunto de datos de ~380 secuencias génicas de V_H que representan clones reordenados no redundantes de anticuerpos humanos que reconocen antígenos de proteína. La siguiente etapa implicó la designación del precursor de línea germinal que generó estas secuencias génicas reordenadas, seguida de un análisis de frecuencia de estas secuencias candidatas de línea germinal. En otras palabras, una determinación de si hay secuencias marco conservadas de línea germinal óptimas o de alta frecuencia para los antígenos de proteína. A fin de determinar las secuencias de línea germinal empleadas por los genes reordenados en las secuencias de V_H filtradas (de Kabat), se usó V BASE. V BASE es un extenso directorio de todas las secuencias humanas de región variable de línea germinal recopiladas a partir de más de un millar de secuencias publicadas, incluyendo las de las publicaciones actuales de las bibliotecas de datos Genbank y EMBL (véanse respectivamente, p. ej., Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410 y bases de datos de acceso público organizadas por el Centro de ingeniería de Proteínas. MRC Centre, Hills Rd, Cambridge, RU, CB2 2QH). Actualmente, hay 51 segmentos V_H funcionales agrupados en 7 familias: (es decir., V_H 1-7), 40 segmentos V_K funcionales agrupados en 7 familias: (es decir., V_K I-VII), y 31 segmentos V_L funcionales agrupados en 10 familias: (es decir., V_L 1-10). Se realizó un conjunto BLAST de secuencias de línea germinal de V BASE frente a las secuencias KABAT filtradas (~380 V_H) para identificar los genes (y familias) de V_H de línea germinal que aportan con mayor frecuencia secuencias expresadas (reordenadas). El análisis identificó, por ejemplo, que seis de las ocho regiones marco conservadas más representadas (regiones marco conservadas 1, 2, y 3) pertenecen a la familia de línea germinal V_H3 , y los miembros de las familias V_H1 y V_H4 formaban un grupo de representación intermedia. Se realizó un análisis de frecuencia sobre las secuencias marco conservadas de línea germinal de V_H (1, 2, y 3) que están representadas en la base de datos Kabat filtrada (figuras 3 y 4). Para este análisis se permitieron hasta 4 mutaciones somáticas (barras punteadas) durante la clasificación de las secuencias reordenadas de un gen de línea germinal. Idénticas coincidencias entre las secuencias Kabat filtradas y los genes de línea germinal se identifican con las barras en blanco. La línea umbral se ajusta para identificar aquellos genes de línea germinal que están representados con mayor frecuencia en secuencias reordenadas de anticuerpos que reconocen antígenos de proteína/péptido (figuras 3 y 4).

La identificación de las secuencias marco conservadas V_H3 de mayor frecuencia tiene una importante consecuencia. La selección de la secuencia marco conservada V_H3 inicial impone las correspondientes limitaciones de diversidad de secuencia y tamaño de CDR como definen las estructuras canónicas convencionales (véase, p. ej., Chothia, C., y cols., Structural repertoire of the human VH segments. J Mol Biol, 1992. 227(3): pág. 799-817 y Tomlinson, I.M., y cols., The Structural repertoire of the human V kappa domain. Embo J, 1995. 14(18): pág. 4628-38). La búsqueda de BLAST en Kabat filtrada por V BASE también identifica posiciones que son los llamados "puntos calientes" para hipermutaciones somáticas que pueden mutarse durante la maduración de la afinidad de moléculas candidatas. Los

resultados preliminares Kabat-V BASE identificaron cuatro regiones marco conservadas V_H3 muy utilizadas; las regiones marco conservadas 3-07, 3-11, 3-23, y 3-33 y regiones marco conservadas V_H1 ; 1-e, y regiones marco conservadas V_H4 ; 4-34 y 4-30.4 (fig. 5). Al elegir múltiples regiones marco conservadas de partida se creó una diversidad estructural añadida fuera de las CDR para la unión potencial a antígeno. El análisis comparativo de las seis secuencias J_H (que codifican la región marco conservada 4) indica que cuatro de estas secuencias son idénticas y que existen únicamente dos diferencias de aminoácidos en las otras dos secuencias. Esta conservación de la secuencia permite el uso de una región marco conservada 4 común para las siete familias de regiones marco conservadas, reduciendo al mínimo la generación de diversidad no funcional (fig. 5).

Por tanto, se demostró que puede identificarse de forma racional un conjunto manejable de secuencias de anticuerpo aceptoras usando los criterios de la invención y usando un planteamiento bioinformático y bases de datos existentes. Es más, la identificación de estas secuencias proporciona el fundamento para aumentar al máximo la diversidad inteligente de CDR dentro de la biblioteca universal de anticuerpos, como se discute más adelante

Ejemplo 2

PROCEDIMIENTOS PARA EL DISEÑO DE LA DIVERSIDAD DE CDR PARA BIBLIOTECAS UNIVERSALES DE ANTICUERPOS

En este ejemplo, se presentan procedimientos para optimizar la diversidad de CDR de una biblioteca universal de anticuerpos.

La elección de regiones marco conservadas candidatas, como se ha destacado previamente, impone tanto los tamaños de CDR para introducir como la selección de la secuencia de aminoácidos inicial. Las seis familias génicas de V_H3 elegidas tienen las mismas estructuras canónicas de 1-3. La estructura canónica 1-3 requiere que CDR1 y CDR3 tengan, respectivamente, bucles de 5 y 17 aminoácidos. Un análisis de frecuencia de aminoácidos de CDR de las bases de datos Kabat y V BASE permite la identificación de los aminoácidos de la CDR para 1) la conservación absoluta de la secuencia, 2) la primera sesión de generación de diversidad, y 3) la posterior maduración de la afinidad mediante la imitación de la hipermutación somática. El diseño de cada CDR en las regiones variables de cadena pesada y ligera se discute de forma secuencial más adelante.

Para diseñar la primera CDR de la cadena pesada, de aquí en adelante "VH-CDR1", los criterios anteriores se consideran de la forma siguiente: Las posiciones de las CDR que están conservadas tanto en la línea germinal como en los genes reordenados son fijas; las posiciones de las CDR conservadas en la línea germinal, pero variables en los genes reordenados, son fijas en la construcción de la biblioteca inicial, pero se permite que muten durante la maduración de la afinidad; y las posiciones de las CDR que presentan diversidad en la línea germinal y en las secuencias reordenadas son posiciones para la incorporación de diversidad usando mutagénesis, por ejemplo, mutagénesis guiada (WTM™). Al empezar con las seis familias génicas de V_H3 identificadas (es decir, 3-07, 3-21, 3-23, 3-30.5, 3-48, y 3-74), el análisis comparativo de V BASE de la secuencia de 5 aminoácidos de CDR1 de línea germinal revela que S31, Y32 y M34 están conservadas entre los seis genes (fig. 6, panel izquierdo). Un análisis de frecuencia de todas las secuencias reordenadas de 5 aminoácidos de CDR1 en el conjunto de datos Kabat filtrado (fig. 6, panel derecho) ilustra tres hallazgos importantes: Primero, Y32 está muy conservada, segundo, las posiciones conservadas de línea germinal S31 y M34 están sometidas a mutaciones somáticas posteriores, y tercero, las posiciones 33 y 35 de CDR1 no están conservadas ni en la línea germinal, ni en las secuencias reordenadas de anticuerpo.

Por consiguiente, en VH-CDR1, Y32 es fija y nunca se somete a ninguna alteración, ya que la estricta conservación de Y32 indica fuertes presiones selectivas para su preservación. Las posiciones 33 y 35 de CDR1 son sitios para la creación de la diversidad de secuencia inicial de CDR1 mediante mutagénesis p. ej., WTM™. Las posiciones S31 y M34 son inicialmente "fijas" pero se identifican como sitios para mutagénesis durante la maduración de la afinidad en cualquiera de los clones candidatos de scFv. La razón para no crear diversidad en todos los sitios es restringir la diversidad inicial de la biblioteca para facilitar la expresión y la presentación.

A partir del análisis de frecuencia de Kabat anterior, CDR1 tiene una secuencia de consenso "natural" de SYAMH. Los restos A33 y H35 se eligen como secuencias naturales debido a su mayor frecuencia en la fig. 6. Al introducir la posterior diversidad de aminoácidos, la secuencia CDR1 sería entonces SYXMX, en la que X denota la posición en la que se efectúa la mutagénesis, p. ej., WTM. Por ejemplo, cuando la mutagénesis, p. ej., WTM, se efectúa sobre el resto de tirosina en las posiciones 33 y 35 de CDR1, las secuencias resultantes deseadas de CDR1 son SY $\overline{Y}$ MH, SYAM $\overline{Y}$ y SY $\overline{Y}$ M $\overline{Y}$. En este caso, se exploran los efectos de la introducción de una cadena lateral aromática. La secuencia de oligonucleótidos del codón para la posición natural A33 es GCX. Si se sustituye por Y33, la correspondiente secuencia de oligonucleótidos necesaria sería TAY. Por tanto, para una mezcla de oligonucleótidos A33 $\rightarrow$ Y33, las secuencias de codón resultantes son (G/T)(A/C)C. Los oligonucleótidos A33 $\rightarrow$ Y33 generados en este caso también pueden tener permutaciones de codón que codifican para glicina (GCC), aspartato (GAC) y serina (TCC). Estos "subproductos" adicionales contribuyen a una diversidad adicional en la posición 33. Para la posición próxima WTM™ 35, la secuencia del codón natural para H35 sería CAY y si se sustituye por Y35, la secuencia de oligonucleótidos requerida sería TAY. Por tanto, para una mezcla H35 $\rightarrow$ Y35, la secuencia de codón resultante es (C/T)AC. En este caso, no se formaría ningún "subproducto" adicional de aminoácidos.

En otro planteamiento, los subproductos se evitan mediante el empleo de mutagénesis de recorrido (LTM), que requiere típicamente la síntesis de un oligonucleótido para cada cambio deseado, pero elimina cualquier subproducto (ruido de fondo).

Para diseñar la segunda CDR de cadena pesada, de aquí en adelante "VH-CDR2", el análisis de secuencia anterior en las bases de datos VBASE y Kabat se planteó de una forma similar al de VH-CDR1 anterior. Se realizó un análisis de frecuencia para las secuencias VH-CDR2 y se construyó un alineamiento de las secuencias CDR2 de línea germinal de las seis regiones marco conservadas candidatas, y se seleccionó una frecuencia umbral del 10 % (fig. 8).

Al empezar con las mismas familias génicas de V_H3 (3-07, 3-21, 3-23, 3-30.5, 3-48, y 3-74), el análisis de frecuencia en V BASE (fig. 8, panel izquierdo) y Kabat filtrada (fig. 8, panel derecho) muestra que las posiciones I51 Y59, A60 y G65 de CDR2 están conservadas en toda la línea germinal y la mayoría de los genes reordenados y, por tanto, han de ser invariables en una CDR2 sintética. El análisis de frecuencia Kabat anterior indica que VH-CDR2 tendría una secuencia "natural" de consenso de GISGGTTYADSVKG. Como las posiciones 54, 55, 58, 61, 62, 63, y 64 de VH-CDR2 muestran conservación de secuencia en la línea germinal pero están sometidas a mutaciones somáticas posteriores, son, por tanto, "fijas" pero se permitió su mutación durante la maduración de la afinidad. Para la diversidad inicial de CDR2, los aminoácidos en investigación (subrayados) se incorporan en las posiciones 50, 52, 52a, 53, 56 y 57 (XIXXXGGXXYYADSVKG) y se introducen mediante mutagénesis, p. ej., WTM™.

A diferencia de la mutagénesis aleatoria, la WTM permite la colocación predeterminada de aminoácidos concretos. Por ejemplo, para realizar una WTM™ de CDR2 con un resto de tirosina (Y) en las posiciones 50, 52, 53, 56 y 57 (subrayadas), las secuencias WTM™ de CDR2 resultantes incluyen las siguientes (las alteraciones están subrayadas): Sustituciones únicas (YIXXXGGXXYYADSVKG, XIYXXGGXXYYADSVKG y etc.), sustituciones dobles (YIYXXGGXXYYADSVKG, YIXXYGGXXYYADSVKG y etc.), sustituciones triples (YIXXYGGYXXYYADSVKG y etc.), sustituciones cuádruples (YIYXXGGXXYYADSVKG o YIYXXGGYXXYYADSVKG) sustituciones quintuples (YIYXXGGYXXYYADSVKG), y sustituciones séxtuples (YIYXXGGYXXYYADSVKG). Típicamente, se prefieren 2-3 sustituciones por CDR y esto puede lograrse fácilmente mediante dopaje de la síntesis de oligonucleótidos (para los detalles técnicos véase, p. ej., el documento US20040033569A1).

La WTM™ para la CDR2 usando los nueve aminoácidos WTM™ preelegidos produce una diversidad de biblioteca de 9×2^6 o 576 miembros. A efectos comparativos, la mutagénesis por saturación de seis posiciones con los veinte aminoácidos sería de 20^6 o $6,4 \times 10^7$. Por consiguiente, al realizar mutagénesis por saturación en las 12 posiciones "no fijas" de la CDR2 sola, la diversidad de biblioteca sería de 20^{12} o 4×10^{15} , lo cual está más allá de las capacidades de la tecnología actual de presentación y exploración de bibliotecas. Esto ilustra una ventaja de la invención que, en cambio, permite que se construya una biblioteca de menor tamaño, pero más representativa. De hecho, los procedimientos de la invención proporcionan una biblioteca manejable en algunas posiciones de CDR, a fin de identificar la primera generación de moléculas de unión. La posterior mutagénesis por maduración de la afinidad en las otras posiciones de CDR optimiza después aquellas moléculas de unión identificadas.

Para diseñar la diversidad de las CDR3 de V_H, las secuencias CDR3 de los anticuerpos de la base de datos Kabat se alinearon de acuerdo con su tamaño y clase antigénica. Las longitudes de CDR3 de anticuerpos que reconocen antígenos no proteicos (en las barras sombreadas) y de proteína/péptido (barras en blanco), se muestran y se ajustan a las líneas de tendencia (continua para las primeras y punteada para las segundas) (fig. 11). También se realizó un análisis de frecuencia de las 13 secuencias de aminoácidos de CDR3 a partir del conjunto de datos Kabat filtrado y se seleccionó una frecuencia umbral del 10 % (fig. 11). Como el análisis del tamaño y de frecuencia de restos de aminoácidos de CDR3 se realiza usando, p. ej., las secuencias génicas reordenadas D y J de inmunoglobulina, no hay equivalentes de CDR3 de línea germinal para las comparaciones directas de Kabat filtrada y V BASE. No obstante, se examinó una base de datos Kabat filtrada y los resultados de la búsqueda indicaron que, en cuanto al tamaño de bucle de CDR3, hay una curva de distribución normal que oscila desde 6 a 24 aminoácidos, con un máximo en los 13-16 aminoácidos aproximadamente (figura 10).

Sin un análisis comparativo paralelo de V BASE frente a Kabat para posiciones de CDR3, se realizó un análisis de frecuencia de Kabat filtrada (figura 11) para cada tamaño de CDR3 reordenada. Dentro de cada clasificación por tamaño, la enumeración del aminoácido más frecuente en esa posición de la CDR3 da como resultado una secuencia "natural" de consenso. Sorprendentemente, este planteamiento de "consenso" identifica aminoácidos concretos bajo presiones selectivas elevadas. Por ejemplo, en una CDR3 de 13 aminoácidos de tamaño, la posición 101 estaba muy conservada como un aspartato (figura 11). Por tanto, como anteriormente, al diseñar la diversidad de VH-CDR1 y VH-CDR2, la D101 se mantiene como una posición de resto "fija" en la VH-CDR3 sintética de 13 aminoácidos. Las posiciones 96, 98, 100c, y 102 de VH-CDR3 muestran, sin embargo, una mayor preferencia por algunos aminoácidos y son, por tanto, "fijas" de forma preliminar, pero después se mutan durante la maduración de la afinidad. La distribución de frecuencias indica que las posiciones 95, 97, 99, 100, 100a, 100b, 100d, y 100e de CDR3 no mostraron ningún aminoácido preferente. Por tanto, en la secuencia de CDR3 de 13 aminoácidos, la fórmula XGXSXXXXXXXDY representa las posiciones (subrayadas) que son sitios de diversidad usando, p. ej., mutagénesis, tal como WTM. Puede efectuarse un análisis similar para todos los tamaños de secuencias de CDR3 entre 8 y 20 aminoácidos. La figura 10 ilustra que esta gama de tamaño abarca una mayoría de diversidad de longitud que se encuentra en la CDR3 de anticuerpos que reconocen antígenos/dianas proteicos.

Ejemplo 3**PROCEDIMIENTOS PARA GENERAR PERFILES DE VARIABILIDAD POSICIONAL (PV) PARA CDR DE ANTICUERPOS USANDO BIOINFORMÁTICA**

5 En otro planteamiento, las bibliotecas universales de anticuerpos (BUA) se diseñaron mediante la determinación de los perfiles de variabilidad posicional (PV) para CDR expresadas *in vivo*. Los perfiles de variabilidad posicional representan el catálogo de los diferentes aminoácidos, y de sus respectivos índices de aparición, presentes en una posición concreta en un conjunto de datos de secuencias alineadas de anticuerpos que se expresan de forma natural.

10 Por tanto, la determinación de los PV conlleva dos etapas, p. ej., etapa 1) recopilación y selección de (un conjunto de datos de) secuencias alineadas de aminoácidos que comparten una o más propiedades definidas de interés para crear un conjunto de datos. Las secuencias de CDR1, CDR2, y CDR3, bien de V_H , o bien de V_L , alineadas por separado, forman los conjuntos de datos iniciales para los fines típicos. Se dispone de varios planteamientos para extraer conjuntos de datos de CDR con los resultados de PV correspondientes. Típicamente, para efectuar la etapa (2), se enumera la variabilidad de aminoácidos de los conjuntos de datos (CDR) y sus frecuencias relativas para cada posición alineada (figura 43). El PV de cada conjunto de datos de CDR identifica después las características deseadas de una posición de CDR dada para la posterior introducción de la representación de diversidad.

15 Para efectuar la etapa 1, se ensambla una base de datos de secuencias alineadas. Se seleccionan secuencias que comparten uno o más aspectos definidos de interés. Por ejemplo, las proteínas pueden tener motivos, dominios, identificables y/o son miembros de familias relacionadas evolutivamente para permitir el alineamiento de secuencia total o por porciones entre ellas. El conjunto inicial de datos de entrada puede proceder de una recopilación previa de secuencias caracterizadas y agrupadas con anterioridad, tales como la base de datos Kabat de anticuerpos maduros expresados de forma endógena.

20 A partir de la base de datos Kabat, se recogieron de forma selectiva secuencias de inmunoglobulina humana y, en particular, de V_H para el conjunto de datos de la base de partida. Típicamente, el origen primordial de la línea germinal para cada secuencia reordenada humana de V_H se determina mediante análisis comparativo. La línea germinal de base correspondiente se denomina "subfamilia originaria" (ETAPA 1 en la figura 41). Dentro de este "conjunto de datos base" inicial pueden identificarse y definirse CDR adicionales en las secuencias de V_H usando los parámetros expuestos mediante la definición de contacto. La designación de las CDR y de los aminoácidos que comprenden también puede describirse mediante Kabat, Chothia o cualquier otra definición adecuada (ETAPA 2, en la figura 41).

25 Dentro del "conjunto de datos base" de secuencias V_H humanas de partida, aún es probable que las secuencias V_H recopiladas posean características enormemente diferentes. Las secuencias de las regiones marco conservadas de V_H variarán con respecto a: 1-agrupamientos de familias (V_H1 , V_H2 , V_H3 , V_H4 etc.). 2-"subfamilias originarias", 3-longitudes de CDR, 4-clases de estructura canónica de CDR, 5-especificidad de antígeno, entre otras. Debido a la disparidad de secuencia entre los miembros del "conjunto de datos base", el intento de obtener un análisis coherente puede requerir otra selección a partir del "conjunto de datos base" de partida, de modo que puedan identificarse los conjuntos de datos que comparten una o más de las propiedades de interés elegidas. Los miembros constituyentes que comparten esas propiedades respectivas pueden producir un conjunto de secuencias más "normalizado" para el análisis comparativo significativo dentro del subgrupo. Este procedimiento puede ser iterado, dando como resultado la generación de conjuntos de datos menores con mayor grado de relación, y esto se ejemplifica en la figura 43.

30 Las CDR pueden clasificarse de la forma siguiente. Empezando con el "conjunto de datos base" no redundante de todas las secuencias V_H humanas, solo se seleccionaron posteriormente las secuencias que generan secuencias V_H1 (figura 43). El filtro de no redundancia elimina las secuencias duplicadas de anticuerpo depositadas contra el mismo antígeno. Si hay diferentes anticuerpos producidos contra el mismo antígeno, estas secuencias se retienen en la base de datos. Dentro del subgrupo V_H1 , las secuencias CDR1 y CDR2 se identifican y después vuelven a dividirse como subgrupos de CDR1 o CDR2. Hay que destacar que en la división de las CDR dentro de las familias de V_H , los subgrupos de CDR1 o CDR2 aún pueden estar poblados con CDR de diferentes longitudes. La CDR2 de V_H1 aparece en longitudes tanto de 13 como también de 15 aminoácidos. Para CDR1 y CDR2 se seleccionan longitudes de 6 y 13 aminoácidos respectivamente, y los conjuntos de datos generados se denominan $VH-1_CDR1_6$ y $VH-1_CDR2_13$, respectivamente (figura 43).

35 Las estructuras canónicas se clasifican de la forma siguiente. Dentro de las secuencias CDR2 de V_H1 , se efectúa otra división de subgrupo para volver a filtrar entre las que eran, bien de la estructura canónica 2 (CS2), o bien de la estructura canónica (CS3). Las estructuras canónicas pueden definirse mediante la diferenciación de restos distintivos en restos clave. Los restos y posiciones de aminoácidos dependen de la definición de CDR que se utilice. En este ejemplo, las secuencias CDR2 de V_H1 que incorporan un aminoácido V, A, L o T en la posición de aminoácido 71 denotaron estructura canónica 2, mientras que un R en la misma posición de aminoácido significó estructura canónica 3. Estas operaciones generaron conjuntos de datos llamados, respectivamente, $VH-1_CDR2_13_CS2$ y $VH-1_CDR2_13_CS3$ (figura 43).

El subagrupamiento de VH-1_CDR2_13_CS2 y VH-1_CDR2_13_CS3 (figura 49) revela perfiles de variabilidad ligeramente diferentes entre ellos y la colección más general VH1_CDR2_13 (figura 48). Por ejemplo, en VH-1_CDR2_13_CS2, el "A" no es uno de los aminoácidos más preferidos en la posición 52a. Para VH-1_CDR2_13_CS3, se ha encontrado que en la posición 51 se introduce favorablemente un "M" que no está ni en VH-1_CDR2_13_CS2, ni en VH1_CDR2_13. Un ejemplo más espectacular entre VH-1_CDR2_13_CS2 y VH-1_CDR2_13_CS3 es que en la primera hay una preferencia casi idéntica entre A y T en la posición 57.

La CDR1 de VH1 tiene clase canónica 1 (CS1), con el requisito de que los restos de aminoácido 6 generen el subgrupo: VH-1_CDR1_6 (figura 44). En este caso, las CS1 de CDR1 que distinguen aminoácidos distintivos clave incluyen, por ejemplo, un T, A, V, G, o S en la posición 24, un G en la posición de aminoácido 26, y cualquiera de I, F, L, V, o S en la posición 29. Por tanto, es posible que el conjunto de datos de VH-1_CDR1_6 pueda contener secuencias que no pertenecen a la CS1, ya que algunas variantes de CDR de 6 aminoácidos no tienen las secuencias distintivas necesarias. El fin principal de la biblioteca universal de anticuerpos es emparejar de forma óptima secuencias marco conservadas con estructuras canónicas de CDR y sus secuencias variables dentro de ellas para obtener las configuraciones más estables y funcionales. Por tanto, estas secuencias no de CS1 pueden a "ruido de fondo" de secuencia en el conjunto de datos introduciendo aminoácidos no optimizados de forma natural para la estabilidad y funcionalidad de la CS1. Por tanto, un perfeccionamiento posterior puede dividir únicamente aquellas secuencias que tengan emparejamientos distintivos de CS1 para generar el conjunto de datos VH-1_CS1_CDR1_6 (figura 45).

El emparejamiento cruzado de CDR se realizó de la forma siguiente. Los subgrupos anteriores son ejemplos de colecciones de secuencias que se han filtrado y normalizado con respecto a subfamilia VH de línea germinal, tamaño de longitud de CDR y estructura canónica. Hay otros parámetros, fuera de las limitaciones estructurales inmediatas de CDR, que pueden influir de forma directa sobre las secuencias de CDR endógenas *in vivo*. De hecho, el fenómeno de que la estructura canónica de CDR puede influir sobre la secuencia de CDR interna puede demostrarse. Es posible que tener una estructura canónica de CDR pueda influir tanto sobre la estructura canónica como sobre la secuencia de otra CDR. Para demostrar esta interdependencia de CDR, se realizó un análisis de secuencia de CDR cuando los subgrupos se dividieron sobre la base de los apareamientos entre estructuras canónicas de CDR. Por ejemplo, se recogieron subgrupos de CDR a través de los cuales la CDR 1 (en este caso solo la CS1) se agrupó con cualquiera de CDR2, CS2 o CS3 en la secuencia original de anticuerpo compuesta de las mismas.

Siguiendo este razonamiento, VH-1_CDR1_6_CS1 se fragmentó así en cualquiera de VH-1_CDR1_6_CS1-2 y VH-1_CDR1_6_CS1-3 (figura 47) representando el "apareamiento" de CDR1_CS1 con CDR2_CS2 y CDR2_CS3 respectivamente. En general, el perfil de variabilidad es similar al de toda la colección VH-1_CDR1_6 pero hay preferencias individuales de CS en posiciones concretas de CDR. En la posición 31 de VH-1_CDR1_6_CS1-2 (figura 46), la "G" y la "D" no aparecerían en el perfil de variabilidad comparado con cualquiera de VH-1_CDR1_6_CS1 y VH-1_CDR1_6_CS1-3. Entre VH-1_CDR1_6_CS1-2 y VH-1_CDR1_6_CS1-3, la posición 33 también muestra alguna variabilidad fuera. En el apareamiento de VH-1_CDR1_6_CS1-2, el "A" es el aminoácido representado de forma dominante, junto con "G", "T", y "W", asociados de forma única, mientras que; en el apareamiento VH-1_CDR1_6_CS1-3, el aminoácido dominante es el "Y", con su aminoácido variable preferido "D". Aunque no se ha analizado, VH-1_CDR2 también aparece con una estructura canónica de "U" en una longitud de 15 aminoácidos, que también puede emparejarse de forma selectiva con sus secuencias CDR1 endógenas. Se predice que el perfil de variabilidad VH-1_CDR1_6_CS1-U resultante puede ser diferente de los perfiles de variabilidad VH-1_CDR1_6_CS1-3 y CS1-2 anteriores.

Por el contrario, la interdependencia entre secuencias CDR2 en relación con la estructura canónica 1 de CDR1 se analizó de una manera similar. La figura 50 ilustra el perfil de variabilidad de VH-1_CDR2_13_CS2-1 y VH-1_CDR2_13_CS3-1. En este caso, aunque los perfiles de variabilidad de CDR2 fueron casi idénticos, lo que demuestra que el diseño de CDR2 puede funcionar independientemente de CDR1 a este respecto. Sin embargo, para otras regiones marco conservadas, tales como la familia VH4, la CDR2 de VH4 solo tiene una estructura canónica (CS-1) con una longitud de CDR2 de 12 aminoácidos. Es la CDR1 de VH4 la que implica tres estructuras canónicas, CS-1, CS-2 y CS-3. En este caso, se anticipan preferencias posicionales de aminoácidos de CDR exclusivas a partir de los perfiles de variabilidad resultantes VH-4_CDR2_12_CS1-1, VH-4_CDR2_12_CS1-2, y VH-4_CDR2_12_CS1-1.

Estos resultados anteriores demuestran que, dependiendo de las estructuras canónicas tanto de CDR1 como de CDR2 elegidas para su utilización como aceptoras, los aminoácidos pueden "ajustarse" dependiendo de cuáles aminoácidos vayan a introducirse en las diversas posiciones de aminoácidos de la CDR para replicar la diversidad natural empleada, p. ej., mediante el emparejamiento de las secuencias que se van a encontrar con la mayor probabilidad con otras secuencias, si hay estabilización cruzada de la CDR.

La clasificación antigénica de CDR se realizó de la forma siguiente. Una vez agrupados sobre la base de clasificaciones estructurales, los miembros recogidos pueden clasificarse sobre la base de la especificidad antigénica (figura 43). Puede haber una correlación de aminoácidos preferidos dentro de las CDR para una clase antigénica dada y esto se observó para la preferencia de clase antigénica de ciertas regiones marco conservadas. Por tanto, es posible añadir un parámetro adicional, la especificidad de antígeno, en la división de las secuencias de

CDR.

La ampliación de las colecciones de secuencias de CDR se realizó de la forma siguiente. El análisis anterior ha demostrado la adición de combinaciones de exploración de múltiples parámetros para generar un subgrupo de interés. Esto tiene el efecto de "estrechar" los perfiles de variabilidad de las secuencias de CDR seleccionadas. Sin embargo, puede haber ocasión de realizar lo inverso, esto es, obtener conjuntos de datos mayores con menor grado de homogeneidad o propiedades compartidas. Esto se consigue, en efecto, mediante la combinación de diferentes conjuntos de datos (figura 43). En este ejemplo, se enumeró el perfil de variabilidad de todas las CDR2 de longitud de 12 con una estructura canónica 2 (CDR2_13_CS2), sin importar a qué familia de VH podían atribuirse esas CDR2. Esto da, efectivamente, un seguimiento más amplio de todos los aminoácidos que contribuyen a la diversidad CDR2_13_CS2.

El procedimiento selectivo para la elección de conjuntos de datos de los subgrupos de CDR1 y CDR2 se describe en la figura 43. La importante ventaja de nuestro procedimiento es que son posibles muchas vías de "selección" diferentes, y cada una de ellas genera un conjunto de datos diferente y, por tanto, un perfil de variabilidad (PV) diferente.

Esto se ejemplifica en las figuras de más adelante, en las que los perfiles de variabilidad para CDR1 se comparan entre VH1, VH3 y VH4. Aunque similares en cierto modo, aparece una importante diferencia en la posición 34. Para VH1, un "I" puede ser fijo, para VH3, un "M" puede ser fijo, y para VH4, un "W" puede ser fijo para el diseño de la CDR. Otra diferencia se relacionaría con el uso de consideraciones de frecuencia del 50 %-80 %. Para VH4, los dos aminoácidos encontrados con mayor frecuencia en la posición 35 son "S" y "H". De forma colectiva, el porcentaje agregado de esos dos aminoácidos sería superior al 80 %. Por definición, la posición 35 puede caracterizarse como "fija" y tanto "S" como "H" pueden introducirse en esa posición como coproductos forzados. Por el contrario, tanto para VH1 como para VH3, los dos aminoácidos encontrados con mayor frecuencia en la posición 35, su frecuencia de aparición no el agregado superior al 80 % y se caracterizarían como "variables" y se someterían a mutagénesis (p. ej., WTM) en esa posición.

Ejemplo 4

PROCEDIMIENTOS PARA LA MODIFICACIÓN POR INGENIERÍA GENÉTICA DE UNA BIBLIOTECA UNIVERSAL DE ANTICUERPOS

En este ejemplo, se describen las etapas de fabricación y ensamblaje de una biblioteca universal de anticuerpos usando técnicas de ingeniería genética.

Brevemente, los fragmentos V_L y V_H de los anticuerpos se clonan usando técnicas convencionales de biología molecular. Los oligonucleótidos que codifican las secuencias marco conservadas y las CDR de las regiones variables se ensamblan mediante la reacción en cadena de polimerasa (PCR). Estos fragmentos V_L y V_H se ligan posteriormente con un engarzador de poli-Gly-Ser (típicamente GGGGS-GGGSGGGGS) para generar anticuerpos monocatenarios (scFv). Las moléculas de longitud completa se amplifican después usando cebadores adyacentes en 5' y 3' que contienen sitios de restricción que facilitan la clonación dentro del(los) vector(es) de expresión-presentación. La diversidad total de las bibliotecas generada depende del número de secuencias marco conservadas y del número de posiciones en las CDR elegidas para mutagénesis. p. ej., usando WTM.

Típicamente, la diversidad media de la biblioteca de V_H , usando 9 aminoácidos para efectuar la WTM™, es de $3,5 \times 10^6$ (6 secuencias marco conservadas $\times$ 9 aminoácidos $\times$ (2^2 para CDR1 $\times$ 2^6 para CDR2 $\times$ 2^8 para la CDR3 de 13 aminoácidos)). La diversidad de la biblioteca de V_H es un límite superior y la diversidad de las bibliotecas de V_L y V_K es significativamente menor, lo que limita de esa forma la diversidad combinada de la biblioteca completa de scFV de 10^{10} a 10^{11} , lo cual está dentro del intervalo de las eficacias de transformación de los sistemas bacterianos.

Por consiguiente, se sintetizan 90 oligonucleótidos para abarcar las secuencias marco conservadas de las bibliotecas de V_H , V_L , y V_K , además de 2 oligonucleótidos que codifican para la región engarzadora y 2 oligonucleótidos que codifican para las inmunoetiquetas de His y Myc en los extremos N y C-terminales, respectivamente. Además, se sintetiza un subgrupo de 30-60 oligonucleótidos degenerados que muestran la diversidad en las CDR 1, 2 y 3 de cada una de las tres bibliotecas (total: 90-180). Estos oligonucleótidos se ensamblan mediante el procedimiento de PCR de extensión por solapamiento único (SOE) para generar las bibliotecas que incluyen las combinaciones V_H - V_L y V_H - V_K necesarias. Después se eligen clones aleatorios de cada biblioteca para la verificación de la secuencia y la evaluación de la calidad de la biblioteca.

Con respecto a la diversidad de CDR, se usa LTM para explorar pequeñas perturbaciones dentro de los bucles de CDR del anticuerpo (p. ej., un cambio por bucle). Para más mejoría, la WTM, que permite la incorporación de más de una sustitución dentro de una CDR, se usa a continuación para explorar de forma exhaustiva el panorama químico de la(s) CDR. Mediante el uso de WTM, el aminoácido natural y las variantes de aminoácidos deseadas se exploran en las posiciones diana de CDR, mediante la manipulación de la síntesis de oligonucleótidos. Se sintetiza un conjunto mixto de oligonucleótidos en el que un subgrupo de oligonucleótidos codifica para la secuencia natural y otro subgrupo codifica para la mutación diana en una posición específica. En el procedimiento de WTM, en cada etapa de la síntesis, una de las dos bases elonga la cadena de oligonucleótido en crecimiento. Una de las bases

codifica para el codón natural, mientras que la otra base pertenece a un codón para la mutación deseada.

Ejemplo 5

PROCEDIMIENTOS PARA LA EXPRESIÓN Y PRESENTACIÓN DE UNA BIBLIOTECA UNIVERSAL DE ANTICUERPOS

- 5 En este ejemplo, se describen procedimientos para la expresión y presentación de una biblioteca universal de anticuerpos para la exploración frente a dianas.

Brevemente, se usa un sistema bacteriano de expresión y presentación que tiene una fiabilidad demostrada para la expresión de moléculas scFv a partir de bibliotecas. El formato scFv consiste en las unidades funcionales de unión a antígeno (regiones V_H y V_L) unidas por un péptido engarzador (fig. 14). Dichas bibliotecas de la invención aumentan la diversidad del repertorio natural y una vez construidas pueden explorarse repetidamente para otros antígenos.

La biblioteca de scFv se transfecta dentro de los anfitriones bacterianos receptores usando técnicas convencionales. Las proteínas de fusión scFv expresadas se expresan en una localización de la superficie externa, lo cual permite la unión de antígenos marcados mediante fluorescencia. Las proteínas candidatas se marcan de forma individual mediante FITC (bien de forma directa, o bien indirecta, por medio de un enlace con biotina-estreptavidina). Aquellos miembros de la biblioteca que expresan clones de scFv adecuados que se unen de forma eficaz a los antígenos marcados se enriquecen después usando FACS. Esta población de células se vuelve a hacer crecer y se somete a posteriores sesiones de selección usando mayores niveles de rigurosidad para aislar un subgrupo menor de clones que reconocen la diana con mayor especificidad y afinidad. Las bibliotecas se abren con facilidad a formatos de alto rendimiento, usando, p. ej., etiquetas de anticuerpos anti Myc marcados con FITC y análisis de FACS para una identificación y confirmación rápidas.

Después se aíslan los clones candidatos y se realizan preparaciones de plásmido para obtener información de secuencia de scFv. El planteamiento permite una sustitución de los codones razonable motivada por hipótesis, necesaria para determinar y optimizar funcionalidad de aminoácidos en las regiones determinantes de complementariedad de las V_H y V_L del anticuerpo. El análisis comparativo de secuencias y los perfiles individuales de afinidad/especificidad de los clones determinan después qué clones se someten a maduración de la afinidad (véase el ejemplo 6).

Ejemplo 6

PROCEDIMIENTOS DE REALIZACIÓN DE MADURACIÓN DE LA AFINIDAD DE ALTO RENDIMIENTO DE CANDIDATOS DE UNA BIBLIOTECA UNIVERSAL DE ANTICUERPOS

- 30 En este ejemplo, se describen las etapas para la identificación y mejora de un anticuerpo candidato de una biblioteca universal de anticuerpos usando maduración de la afinidad.

Brevemente, a fin de validar el poder de la biblioteca universal y la capacidad de tomar una molécula de anticuerpo candidato y perfeccionar las propiedades de unión de la molécula, se designó un anticuerpo disponible en el mercado como anticuerpo de ensayo, y se mutó (usando, p. ej., tecnología de WTM™/LTM™), expresó, presentó, y mejoró de acuerdo con los procedimientos de la invención.

Brevemente, el anticuerpo problema se mutó en un formato scFv y después se expresó y presentó usando presentación en levaduras, aunque también puede usarse cualquiera de los sistemas de presentación en bacterias. Las selecciones cinéticas de las bibliotecas de scFv presentadas en levaduras implican el marcaje inicial de células con antígenos biotinilados seguido de captura dependiente del tiempo en presencia de un gran exceso de antígeno sin biotinilar. Los clones con cinéticas de disociación más lentas se identifican mediante marcaje con SAPE (de *streptavidin-phycoerythrin*, estreptavidina-ficoeritrina) tras el periodo de captura y se clasifican usando un clasificador FACS de alta velocidad. El panel izquierdo de la fig. 29 muestra la gráfica de puntos resultante del control natural y la ventana de clasificación, mostrando la gráfica de puntos la biblioteca y el número de clones de la ventana de clasificación, y la gráfica de puntos de los clones aislados a partir de la biblioteca después de la clasificación. En el panel derecho de la fig. 29, los datos de los ensayos de disociación para dos clones con afinidad madurada en comparación con la proteína natural se ajustaron a una curva exponencial individual para determinar las constantes del índice de disociación (K_{off}). Los clones 1 y 2 muestran K_{off} de los índices de disociación 5,2 y 4,3 veces más lentas que la molécula original.

La verificación de la secuencia de ADN de clones elegidos al azar indica que las bibliotecas son de gran calidad con respecto a la diversidad mutacional deseada, mutaciones puntuales, eliminaciones, e inserciones no intencionadas. Esta eficacia contrasta con las estrategias aleatorias/fortuitas de mutagénesis en las que la introducción incontrolada de diversas bases produce niveles mayores de efectos no deseados de cambios de base que conducen a baja expresión o funcionalidad del anticuerpo debidas a un uso de aminoácidos desfavorables y a codones de terminación involuntarios.

55 Es más, los datos tabulados de secuencias a partir del análisis de LTM de un anticuerpo de ensayo indican

diversidad productiva con muy poco ruido de fondo. El panel inferior de la fig. 29 muestra la secuencia natural y 29 mutaciones distintas que aumentan la afinidad de la de la molécula original en 1,5 veces o mejor, las cuales no estaban cubiertas en las seis CDR. Varios de estos cambios se aislaron múltiples veces, por ejemplo, en CDR3, se encontraron tres cambios de S a K y dos cambios de S a Q diferentes. En cambio, las columnas sombreadas indican las posiciones de CDR en las nunca se encontró que los cambios aumentarían la afinidad por el antígeno. Posteriormente, la combinación en una biblioteca de todas las mutaciones LTM individuales descubiertas facilita el aislamiento de clones que presentan avidez mejorada entre estas mutaciones de alta afinidad.

Ejemplo 7

PROCEDIMIENTOS DE EXPLORACIÓN DE UNA BIBLIOTECA UNIVERSAL DE ANTICUERPOS PARA IDENTIFICAR UN ANTICUERPO TERAPÉUTICO CANDIDATO PARA TRATAR DE LA ENFERMEDAD HUMANA

En este ejemplo, se describen procedimientos para la exploración de una biblioteca universal de anticuerpos de la invención para identificar un candidato terapéutico.

Brevemente, se eligió una enfermedad renal crónica de efectos devastadores que ha sido resistente a terapias previas como diana para la exploración frente a la biblioteca universal de anticuerpos de la invención. En particular, la enfermedad renal crónica (ERC) está reconocida como un importante problema de salud pública en los Estados Unidos, con más de 20 millones de individuos afectados. Un importante obstáculo en la comprensión de los procesos nefrogénicos es la falta de reactivos adecuados que reconozcan biomarcadores renales específicos que identifiquen 1) los diferentes tipos celulares implicados, y 2) las moléculas participantes que influyen sobre la diferenciación en estas células. Los anticuerpos que reconocen estos marcadores renales podrían aumentar significativamente el grupo actual de reactivos necesarios para investigar la organogénesis y el diagnóstico de la enfermedad renales.

Para comprender la biología renal, se identificaron seis anticuerpos candidatos humanos específicos renales, 1) un intercambiador de Na-H (isoformas NH3, NHE8) (14, 15), un intercambiador de aniones (isoformas SLC26A6, SLC 27A7), una molécula de adhesión caderina Ksp (16), y lipocalina, para su exploración frente a la biblioteca universal de anticuerpos.

Los hematólogos se han beneficiado durante mucho tiempo de reactivos de anticuerpos monoclonales (Mab, de *monoclonal antibodies*) que reconocen marcadores de superficie de los "grupos de diferenciación" (CD, de *cluster of differentiation*). La hematopoyesis, el proceso que genera los linajes linfóide y mieloide, ha mostrado a menudo numerosas ventajas como modelo de sistema de desarrollo. Muchas de las razones de su éxito residen en la facilidad con la que las células hematopoyéticas pueden identificarse, aislarse y manipularse mediante anticuerpos monoclonales (Mabs).

Para analizar los candidatos terapéuticos identificados en la exploración anterior, pueden usarse biomarcadores diagnósticos de enfermedad, p. ej., lipocalina asociada a la gelatinasa de neutrófilos (LAGN), un biomarcador en la detección de la insuficiencia renal aguda (IRA) precoz. Además, puede controlarse una enfermedad diana para el tratamiento terapéutico, por ejemplo, glomerulonefritis, con proteína de colágeno (IV) $\alpha 3$. Estas proteínas se marcan con biotina usando los protocolos existentes para facilitar la visualización por FACS usando estreptavidina ficoeritrina (SA-PE) y después se someten a 3-5 sesiones de selección para identificar una primera remesa de moléculas de unión a antígenos a partir de la biblioteca universal de anticuerpos. Los candidatos de anticuerpo iniciales se secuencian después y su afinidad se analiza con proteínas solubles purificadas usando un ensayo BIAcore. Después, si se desea, la afinidad de los candidatos de anticuerpo se madura como se describe en el ejemplo 6.

Después se realizan experimentos en el laboratorio para determinar su funcionalidad en el reconocimiento de la diana antigénica usando métodos reconocidos en la técnica tales como inmunohistoquímica, diagnóstico por inmunotransferencia de biomarcadores, y experimentos de bloqueo de anticuerpos *in vitro* e *in vivo*.

Ejemplo 8

PROCEDIMIENTOS DE IDENTIFICACIÓN GUIADA POR BIOINFORMÁTICA DE SECUENCIAS DE LA BIBLIOTECA UNIVERSAL DE ANTICUERPOS USANDO FILTRADO Y ANÁLISIS DE GRUPOS DE SECUENCIAS GÉNICAS

En este ejemplo, se describen procedimientos para la identificación de secuencias de la biblioteca universal de anticuerpos usando análisis de bases de datos.

Brevemente, V BASE y KABAT se seleccionaron como el principal origen de datos con el fin de diseñar bibliotecas universales de anticuerpos con óptima diversidad estructural y funcional. V BASE es una base de datos que contiene secuencias de ADN y polipéptidos de todos los segmentos de línea germinal humanos, alineados y anotados de acuerdo con las definiciones de CDR de Kabat (Kabat y cols.) con esquema de numeración basado en Chothia (Chothia y cols.). KABAT es la base de datos más extensa de secuencias reordenadas de anticuerpos de especies disparejas. Para mejorar la afinidad del anticuerpo mediante la introducción de diversidad en todas las CDR, tanto de las cadenas pesadas como ligeras, se seleccionó el esquema de definición de contacto para CDR (MacCallum y

cols.) como esquema alternativo al procedimiento guiado de Kabat (Kabat y cols.) y Chothia (Chothia y cols.). El procedimiento de definición de contacto permite la introducción de una diversidad estructural significativa y la mejora de la unión sin afectar a la estabilidad de la estructura del anticuerpo. Sin embargo, se usa la numeración de Chothia porque es el esquema óptimo para usar con el procedimiento de definición de contacto. En la fig. 30, se muestra una comparación de las tres definiciones de CDR más usadas. De acuerdo con nuestra elección del número de aminoácidos para región marco conservada

Tabla 1. Distribución de aminoácidos a lo largo de las cadenas pesadas y ligeras de acuerdo con la definición de contacto de MacCallum y cols.

	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4
VH	29	6-8	11	12-15	37	¿9-?	12
VK	29	7-13	9	10	33	¿8-?	11
VL	28	7-13	9	10	33/35	¿8-?	11

El análisis de V BASE se llevó a cabo para todos los segmentos V de línea germinal (51 VH, 40 VK, y 31 VL) que se han descargado y analizado sintácticamente de acuerdo con las definiciones anteriores y almacenado localmente en 3 archivos diferentes en un formato descrito mediante FR1-CDR1-FR2-CDR2-FR3, en el que FR se refiere a región marco conservada, p. ej., como se muestra en la fig. 31.

A partir de estos conjuntos de datos pueden extraerse datos individuales para cada FR o CDR. En particular, se construyeron secuencias como FR1xFR2xFR3 (llamada FR123) en la que "x" se usa como marcador de posición para las CDR 1 y 2. Los conjuntos de datos resultantes se almacenan en un formato compacto y conveniente, tal como FASTA (fig. 32).

Para cada una de las tres familias de línea germinal, las secuencias en FR123 se usaron para generar una matriz de distancias para analizar sus relaciones en el espacio de región marco conservada. Todas las secuencias idénticas se plegaron en una y todas las secuencias con alta semejanza se agruparon juntas. Cada grupo es una representación de características estructurales similares. La agrupación jerárquica y los correspondientes árboles de secuencias FR123 se analizaron por ordenador usando el procedimiento UPGMA dentro del paquete PHYLIP [PHYL]. Primero se analizó por ordenador una matriz de distancias usando PROTDIST [PHYL], y la fórmula simple de Kimura. Después, se efectuó el NEIGHBOR [PYHL] con el algoritmo UPGMA (véanse las figuras siguientes).

Se usó SeqhuntII [Johnson y cols.] para descargar las bases de datos completas (véase la tabla 2) de secuencias VH, VK y VL humanas a partir de la base de datos Kabat [KBTDB] en formato ASCII y se almacenaron localmente en tres archivos diferentes (datos sin procesar). Estos archivos se han analizado sintácticamente después y cada entrada de Kabat se ha almacenado en un paquete local de java DBMS A [KBTDB]. Se ha desarrollado un paquete de java (com.bioren-inc.kabatDB) que contiene clases para analizar sintácticamente y estudiar los conjuntos de datos anteriores. El paquete también proporciona una serie de procedimientos para convertir y escribir diferentes ensamblajes de las secuencias de entrada que permiten el aislamiento y análisis de regiones específicas. El diseño del paquete es flexible y permite cambiar fácilmente entre sistemas de numeración y/o definiciones de CDR. Se han ejecutado procedimientos para la identificación de clases canónicas de CDR1 y CDR2.

Tabla 2. Número de secuencias humanas descargadas de la base de datos Kabat

	Humana (H)
VH	5971
VK	2374
VL	2012

Los filtros del análisis de Kabat se configuraron de modo que los conjuntos de datos originales se filtraron en varias etapas secuenciales usando un paquete de java (com.biorenc.unilib) y algunas herramientas externas (PROTEIN SPECIFICITY (com.bioren-inc.unilib)). Para analizar un subgrupo de las secuencias reordenadas de inmunoglobulina almacenadas que reconocen únicamente antígenos de proteína (llamado filtro AP a partir de ahora) se seleccionó un filtro apropiado. Las secuencias almacenadas en la BDKabat que carecían de anotaciones antigénicas se excluyeron (Tabla 3).

Tabla 3 Tamaño de los conjuntos de datos después del filtro AP

	H	Prot. Ant. (PA)
VH	5971	758
VK	2373	454
VL	2012	217

Tabla 4. Tamaño de los conjuntos de datos después del filtro CF123

	H	AP	Redundancia (0,95)
VH	5971	903	547
VK	2373	618	268
VL	2012	310	140

- 5 Para evitar el sesgo causado por la redundancia de la base de datos, algunas secuencias se filtraron usando CD-HIT y la herramienta java (para comprobar doblemente los resultados). El algoritmo se basa en la generación de grupos de secuencias que tienen una identidad por encima del umbral elegido (95%), seguida de la selección de una secuencia representativa de cada grupo. La búsqueda de semejanza se hizo sobre secuencias de longitud completa (es decir, FR1-CDR1- FR2-CDR2- FR3-CDR3-FR4).
- 10 Como la mayoría del diseño de la biblioteca reside en el análisis ajustado de las regiones marco conservadas 1, 2 y 3, era muy importante tener datos de secuencia completa para evitar errores de clasificación y/o suposiciones erróneas (Tabla 5).

Tabla 5. Tamaño de los conjuntos de datos después del filtro de redundancia del 95 %

	H	AP	Redundancia 0,95	CF123
VH	5971	903	547	378
VK	2373	618	268	169
VL	2012	310	140	78

- 15 El diseño de la biblioteca se dividió en dos subproyectos relacionados: Regiones marco conservadas y CDR. Las regiones marco conservadas seleccionadas a partir del repertorio humano fueron los segmentos 1, 2, y 3 de regiones marco conservadas de línea germinal representativos de su uso en secuencias reordenadas de inmunoglobulina. Para este fin, los conjuntos de datos filtrados se analizaron sintácticamente y se reescribieron primero en un formato FR123. Este formato concreto se usó durante todo el procedimiento de selección de regiones marco conservadas para determinar cómo estaba distribuido el uso de la región marco conservada de línea germinal en el conjunto de datos de entrada, de modo que se identificaran las familias más populares. Para este fin se ejecutó una clasificación para cada secuencia en el conjunto de datos reordenados y se efectuó un análisis de semejanza usando BLASTP [BLST] para cada conjunto de datos Kabat-FR123 frente al VBASE-FR123 asociado, y después se analizaron sintácticamente. Los resultados se seleccionaron para revelar los aciertos con mayor puntuación de semejanza. La cardinalidad de cada grupo de familias se comparó y las más populares se eligieron como dianas para la elección de secuencias marco conservadas.
- 20
- 25

Tabla 6. Familias de línea germinal seleccionadas

Cadena	Subfamilias seleccionadas
VH	VH-1 VH-3
VK	Vk-I Vk-III
VL	VL-1 VL-2 VL-3

Tabla 7. Uso de la familia de línea germinal obtenido a partir del análisis de grupos

	subfamilia	secuencias	cobertura
VH	VH-1	103	27 %
	VH-3	153	41 %
	VH-4	93	25 %
VK	VK-I	80	47 %
	VK-III	57	34 %
VL	VL-1	22	28 %
	VL-2	19	24 %
	VL-3	33	42 %

Para visualizar grupos de los conjuntos de datos Kabat-FR123, se usó el paquete PHYLIP [PHYL] (véanse las fig. 36, 37, y 38). Los árboles se obtuvieron usando métodos de distancia (UPMGA). Las matrices de distancia se analizaron por ordenador usando la fórmula de Kimura. Las disposiciones resultantes de los árboles se emparejaron con el análisis BLAST previo.

Tras seleccionar las familias de línea germinal de interés, el análisis se ajustó mediante la investigación dentro de los segmentos muy utilizados de región marco conservada y su estructura canónica relativa. Cada conjunto de datos VBASE se analizó por BLAST frente al conjunto seleccionado de datos Kabat relacionado y se analizó sintácticamente el rendimiento dando únicamente secuencias con alta similitud. Cada segmento de región marco conservada de V BASE se ordenó después de acuerdo con el número de aciertos de alta similitud computados. Finalmente, para los grupos VBASE más populares (dentro de las familias seleccionadas), se eligieron representantes de sus miembros con calificación más alta. Si se desea, puede excluirse la secuencia en la posición más alta para el segmento más representativo del grupo seleccionado (es decir, el segmento que está a la distancia mínima de todos los demás segmentos dentro de un grupo; véase la tabla 8).

Tabla 8. Segmentos seleccionados de región marco conservada de línea germinal

V _H	CS	V _K	CS	V _L	cs
1-e	1-2	I-L1	2-1	1b	13-7
3-30*	1-3	III-A27	6-1	2a2*	14-7
3-23	1-3	III-L20	2-1	31	11-7
3-07	1-3			3r	11-7
3-11	1-3				
4-30.4	3-1				
4-34	1-1				

Ejemplo 9**PROCEDIMIENTOS PARA EL DISEÑO DE DIVERSIDAD DE CDR PARA BIBLIOTECAS UNIVERSALES DE ANTICUERPOS USANDO ANÁLISIS EXTENDIDO DE CDR**

En este ejemplo, se describen procedimientos para el diseño de CDR para bibliotecas universales de anticuerpos.

Brevemente, las secuencias marco conservadas seleccionadas (anteriores) se usaron para guiar la selección y el diseño de CDR: Tanto las longitudes como las secuencias de las CDR se diseñaron específicamente para cada familia de regiones marco conservadas para proporcionar total compatibilidad y óptima diversidad. Las longitudes de las CDR 1 y 2 se seleccionaron de acuerdo con la estructura canónica de las regiones marco conservadas de línea germinal seleccionadas (VH-1 VH-3, VK-I, etc.). Partiendo de esta base, se realizó un análisis completo para el

posterior diseño de las CDR1 y 2. El conjunto de datos Kabat original se ha filtrado únicamente para integridad de las regiones marco conservadas 1, 2 y 3 y para la redundancia (umbral de semejanza del 95 %) (véase la tabla 9). Para la compatibilidad región marco conservada-CDR, no se usaron filtros de especificidad.

Tabla 9. Filtrado de los conjuntos originales de datos Kabat para el diseño de CDR.

	Conjunto de datos de partida	CF123	R95
VH	5971	2842	1865
VK	2373	859	471
VL	2012	1282	744

5

Para cada clase de cadena (VH, VK, VL) las secuencias de entrada se han clasificado de acuerdo con las familias de línea germinal seleccionadas y se han guardado en diferentes conjuntos de datos (véase la tabla 9). Estos resultados se han obtenido usando el programa informático BLAST [BLST] y analizando sintácticamente los resultados como se ha descrito anteriormente.

- 10 Dentro de cada subfamilia seleccionada, se analizó la distribución de longitudes tanto de CDR1 como de CDR2, siguiendo la clasificación de las estructuras canónicas. En la siguiente sección se discuten más detalles acerca de este análisis, junto con la estrategia de diseño adoptada. En la tabla 10 se representa un sumario de los resultados.

Tabla 10. Clasificación de secuencias para el diseño de CDR

	Subfamilia	Secuencias	Cobertura
VH	VH-1	375	20 %
	VH-3	761	41 %
VK	VK-1	234	50 %
	VK-III	136	29 %
VL	VL-1	178	29 %
	VL-2	185	25 %
	VL-3	247	34 %

- 15 La longitud de CDR1 y CDR2 se determinó de la forma siguiente. La familia VH de línea germinal tiene longitudes de CDR1 de 6 y 8, estando la última presente únicamente en VH-2, que no se usa. La longitud de CDR2 varía desde 12 hasta 15, siendo 13 la más común y la única requerida por las secuencias marco conservadas seleccionadas.

- 20 La VH1 de línea germinal siempre tiene CDR1 con 6 aminoácidos y CDR2 con 13 aminoácidos (estructuras canónicas 1-3, 1-2). En el conjunto de datos reordenado -97 % de las secuencias identificadas como VH-1 tenían CDR de estas longitudes. Los criterios típicos de región marco conservada son: Longitud 1-e de CDR1: 6; longitud de CDR 2: 13, con una cobertura de clase esperada de >97 %.

- 25 La VH3 de línea germinal siempre tiene CDR1 con 6 aminoácidos y CDR2 con 13 y 15 aminoácidos. Las secuencias marco conservadas seleccionadas tienen longitudes de CDR de 6 y 13 respectivamente, por lo que la longitud 15 no se usó para CDR2. Aquí los datos mostraron que en el 99 % de las secuencias reordenadas, CDR1 tiene longitud 6, como se esperaba; en el 81 %, CDR2 tiene longitud 13. El espacio de uso restante de CDR2 está cubierto, muy probablemente, por las estructuras canónicas 1-U y 1-4 que tienen longitud 15 (en particular 3-15 tiene alguna popularidad de uso). Los criterios típicos de región marco conservada son: 3-07, 3-11, 3-23, longitud 3-30* longitud de CDR1: 6; longitud de CDR 2: 13 con una cobertura de clase esperada de: -81 %.

- 30 La CDR1 de la familia VK de línea germinal tiene un número de aminoácidos que varía de 7 a 13, teniendo la más popular 7 y 8 aminoácidos. La CDR2 siempre tiene 10 aminoácidos.

La VK-I de línea germinal siempre tiene CDR1 con 7 y CDR2 con 10 aminoácidos. El uso en las secuencias reordenadas muestra un emparejamiento perfecto con la información de la línea germinal. Los criterios típicos de región marco conservada son: longitud I-L1 de CDR1: 7; longitud de CDR 2: 10 con una cobertura de clase esperada de: >97 %.

5 La VK-III de línea germinal tiene CDR1 con 7 y 8 aminoácidos y siempre CDR2 con 10 aminoácidos. Aquí los datos muestran que el 50 % de las secuencias tienen una longitud 7 de CDR1 y el 48% tienen longitud 8. Dichos resultados se obtuvieron a causa de la presencia de dos estructuras canónicas muy diferentes y muy comunes. La longitud de CDR2 es 10 en >98 % de las secuencias. Con las regiones marco conservadas seleccionadas, se proporcionan las longitudes de CDR1 de modo que la cobertura esperada fue el 98 % del espacio de uso para la subfamilia. Los criterios típicos de región marco conservada son: longitud III-A27, III-L6 de CDR1: 7 y 8; longitud de CDR2: 10; y con una cobertura de clase esperada de: >98 %.

La CDR1 de la familia VL de línea germinal tiene un número de aminoácidos que varía entre 7 y 10, en la que 8 no es común y, por ello está excluido de forma selectiva. Las longitudes 7, 9 y 10 son bastante frecuentes en la familia.

10 La VL-1 de línea germinal tiene CDR1 con 9 y 10 aminoácidos y CDR2 con 10 aminoácidos. Los datos para secuencias reordenadas muestran que -74% tienen CDR1 de longitud 9 y -99 % de ellas tiene CDR2 de longitud 10. La longitud 10 para CDR1 se excluyó para un mejor ajuste y las regiones marco conservadas seleccionadas típicas fueron: longitud 1b de CDR1: 9; longitud de CDR 2: 10 con una cobertura de clase esperada de -75 %.

15 La VBASE de línea germinal: VL-2 tiene tanto CDR1 como CDR2 con 10 aminoácidos. Los criterios típicos de región marco conservada son: Longitud 2a2 de CDR1: 10; longitud de CDR 2: 10 con una cobertura de clase esperada de -95 %.

20 La VL-3 de línea germinal siempre tiene la CDR1 con 9 aminoácidos y la CDR2 con 10 aminoácidos. Las dos regiones marco conservadas seleccionadas de esta subfamilia se seleccionaron para proporcionar una cobertura más estructural porque esta subfamilia es la más usada en VL. Los criterios típicos de la región marco conservada son: longitud 3r, 31 de CDR1: 9; longitud de CDR 2: 10; y con una cobertura de clase esperada de -99 %.

Tabla 11. Longitudes de CDR 1 y 2 distribuidas a lo largo de las regiones marco conservadas seleccionadas con las secuencias seleccionadas indicadas.

	CDR	longitud	secuencias/total	
VH-1	1	6	369/375	√
	2	13	366/375	√
VH-3	1	6	752/761	√
	2	13	618/761	√
VK-I	1	7	228/234	√
	1	8	0/234	x
	2	10	232/234	√
VK-III	1	7	68/136	√
	1	8	65/136	√
	2	10	134/136	√
VL-I	1	7	0/178	√
	1	9	131/178	√
	2	10	45/178	x
	2	10	177/178	√
VL-2	1	7	1/185	x
	1	9	7/185	x
	2	10	176/185	√
	2	10	185/185	√
VL-3	1	7	244/247	x
	1	9	0/247	x

(continuación)

	CDR	longitud	secuencias/total	
	1	10	0/247	√
	2	10	247/247	√

Para cada una de las 15 CDR seleccionadas (tabla 11) se ejecutó un análisis de frecuencia separado para determinar el uso del aminoácido posicional en el contexto de la región marco conservada seleccionada. El propósito principal fue proporcionar una clasificación para cada posición dentro de cada CDR1 y 2 en 2 categorías diferentes: Una posición fija que muestra 1 o 2 aminoácidos y posiciones dominantes para la diversidad estructural inicial, es decir, mutagénesis.

Se ejecutó un sencillo análisis de frecuencia usando EMBOS/prophecy [EMB] generando una matriz que representa el uso posicional del aminoácido. Después, la matriz resultante se ha analizado sintácticamente y se ha filtrado a fin de tener datos de frecuencia relativa para cada posición. El analizador sintáctico proporciona un filtro muy sencillo basado en dos umbrales (bajo y alto). Para cada posición, el analizador sintáctico procesa únicamente aminoácidos con frecuencia relativa por encima del umbral bajo hasta que la frecuencia acumulada alcanza el umbral alto. Si no se alcanza el umbral alto, entonces el analizador sintáctico evalúa también los aminoácidos con frecuencia relativa por debajo del umbral bajo. Una buena combinación de umbral bajo-alto fue 10-80, porque proporciona buena sensibilidad para la clasificación de la posición. La salida del analizador sintáctico se visualiza como gráficos de frecuencia y los resultados se muestran en las siguientes figuras.

Clasificación cuantitativa de CDR

Las posiciones se clasifican como fijas cuando uno o más aminoácidos son dominantes de forma evidente sobre los otros. Normalmente, en estas situaciones, el analizador sintáctico, con un buen ajuste de los parámetros, es capaz de filtrar los aminoácidos poco comunes. El (los) aminoácido(s) dominante(s) se usan como naturales en la secuencia de CDR; si dos aminoácidos son dominantes, se usa un natural "degenerado", lo cual significa que se sintetiza un codón mixto para proporcionar ambos aminoácidos. Los parámetros elegidos son muy sensibles para identificar posiciones de alta variabilidad (posiciones WTM). En estas posiciones no hay aminoácidos dominantes evidentes, sino muchos diferentes con frecuencia media-baja. Aquí, la diversidad puede representarse usando mutagénesis, p. ej., LTM o WTM con los aminoácidos más frecuentes como naturales.

En las siguientes figuras, se publican todos los gráficos de frecuencia y las secuencias de aminoácidos para CDR1 y CDR2 que se desarrollaron para la biblioteca universal. La nomenclatura de CDR1y CDR2 está construida de la forma siguiente: TIPO DE CADENA-FAMILIA DE LÍNEA GERMINAL_TIPO DE CDR-LONGITUD DE CDR. Por ejemplo, el nombre VH-1_CDR1-6 se refiere a cadena pesada, familia VH-1, CDR1 que tiene 6 aminoácidos. La nomenclatura de las CDR3 es similar: No contiene la clasificación de familia de línea germinal.

Diseño de CDR

La CDR3 tanto de las cadenas pesadas como de las ligeras es la región más variable tanto en longitud como en secuencia, lo que proporciona la mayor parte de la diversidad estructural del sitio de unión del anticuerpo. Por ello, se ejecutó un análisis de longitud para cada cadena, tanto del conjunto de datos completo como de las cadenas específicas de proteína. Se encontró una diferencia significativa en la distribución de longitud de los dos conjuntos de datos, lo que demostró que los antígenos de proteína parecen preferir una CDR3 ligeramente más larga. La CDR3 de VH tiene una distribución bastante amplia, por lo que se seleccionaron longitudes desde 9 a 18 (~75 % del uso) (véase la fig. 10). La CDR3 de Vkappa tiene una distribución muy restringida y las longitudes más usadas son 8 y 9 (véase la fig. 20). La CDR3 de Vlambda tiene una distribución ligeramente más amplia y, en este caso, se seleccionaron para la biblioteca las longitudes 8, 9, 10 y 11 (véase la fig. 24).

En cada una de las longitudes seleccionadas se ejecutó un análisis de frecuencia similar al descrito en el diseño de CDR1 y CDR2. Para todas las longitudes, este análisis mostró una alta diversidad en las localizaciones del medio de las CDR y unas cuantas posiciones conservadas cerca de los límites con las regiones marco conservadas.

Al igual que para CDR1 y CDR2, las regiones de diversidad en CDR3 se seleccionaron como posiciones de alta variabilidad como dianas para mutagénesis. p. ej., WTM. Los restos elegidos como naturales fueron los aminoácidos más frecuentes en cada posición. En algunas posiciones WTM, se eligió el aminoácido natural y, típicamente, se determinó que la presencia de Gly era deseable. La estrategia de WTM se diseñó de una forma modificada: En lugar de escoger la combinación de bases de "distancia mínima" para proporcionar los aminoácidos diana y los naturales, se diseñaron codones mixtos a fin de proporcionar aminoácidos diana, naturales y Gly (es decir, Gly es un subproducto requerido). Las siguientes figuras muestran todos los gráficos de frecuencia y las secuencias de las CDR elegidas para la biblioteca universal de anticuerpos. La nomenclatura de CDR1y CDR2 se construyó de la forma siguiente: TIPO DE CADENA-FAMILIA DE LÍNEA GERMINAL_TIPO DE CDR-LONGITUD DE CDR. Por ejemplo, el nombre VH-1_CDR1-6 se refiere a las 6 posiciones de los restos de CDR1 de la cadena pesada, familia VH-1. La

nomenclatura de las CDR3 es similar pero no contiene la clasificación de familia de línea germinal.

Las condiciones adicionales del diseño de CDR3 son las siguientes. Las glicinas son una necesidad en la CDR3 para las estructuras de bucle funcionales. Se encuentran en la CDR3 aproximadamente en un 10-20 % a lo largo del bucle de CDR3 de V_H. Por tanto, las regiones de CDR3 se diseñaron para acomodar múltiples glicinas a lo largo del bucle. Por tanto, además del aminoácido natural, las glicinas eran un coproducto requerido en múltiples posiciones de CDR3 de V_H. En la posición 95, era muy común un Asp en la tabla de frecuencias para los anticuerpos contra proteínas y péptidos, por tanto, se usó un Asp como el aminoácido natural y Glicina como un coproducto requerido (D/G) para la WTM. De forma semejante, para la posición 96, la Arg fue bastante frecuente y, por tanto, se usó Arg como el aminoácido natural y Gly como un coproducto requerido (R/G). Para las posiciones 97-99, se usó una Ser y Gly como la base (S/G), ya que la serina era un aminoácido bastante común en los bucles de CDR y es, por tanto, bien tolerada. En la posición 101 se usó Asp (que permaneció constante), y también la posición directamente N-terminal del Asp (Phe (D-1)).

Para longitudes de CDR3 de V_H de 10 y por encima, en la posición que está a dos restos en N terminal al Asp (D-2) (p. ej., posición 100 en V_H_CDR3-10), se usó Tyr como el aminoácido base. La Tyr también es bien tolerada en los bucles de CDR de los anticuerpos. La preponderancia de Tyr N terminal al Asp aumenta con la longitud del bucle de CDR3. Por tanto, se añadieron Tyr adicionales como el aminoácido base, como se muestra en la tabla 12. Las demás posiciones N terminales de Asp 101 hasta la posición 99 usan una Ser como natural y Gly como un coproducto requerido, como se muestra en la tabla 12. A medida que se realiza la mutagénesis guiada, puede estructurarse cada bucle de CDR3 para crear anticuerpos funcionales, ya que las glicinas están presentes para la estructura de bucle (generalmente 10-25 %), y están presentes en el bucle aminoácidos bien tolerados. Se ganan interacciones funcionales de unión adicionales a través de los aminoácidos guiados y los coproductos funcionales.

En la tabla 12, más adelante, se expone un sumario de las secuencias de CDR identificadas para su uso en la biblioteca de anticuerpos universales de la invención. Los nombres de las CDR están normalizados: El primer campo en el nombre es la familia de línea germinal, el segundo campo es el tipo de CDR y el tercer campo es la longitud de la CDR (ejemplo: V_H1_CDR1-6 es la CDR1 de la familia de línea germinal V_H1 que tiene longitud 6). Las posiciones con letra única son posiciones fijas; las posiciones con dos letras son posiciones de combinación en las que la síntesis se realiza con una mezcla a fin de tener únicamente dos aminoácidos diana (ejemplo: T-S); y dos letras en las posiciones en las que la primera es "X" son posiciones WTM. El aminoácido que sigue a la X es el natural (ejemplo: X-V). Las posiciones de tres letras en las que la primera es "X" son posiciones WTM "optimizadas por coproducto". La letra del aminoácido que sigue a la "X" es la del natural. El último aminoácido (el de detrás de "/") es un coproducto requerido.

Tabla 12. Sumario de secuencias CDR para la biblioteca universal de anticuerpos

CDR1

VH1_CDR1-6	30	31	32	33	34	35
	T-S	S	Y	X-A	I-M	X-S

VH3_CDR1-6	S	S	Y	X-A	M	X-S
------------	---	---	---	-----	---	-----

VK1_CDR1-7	30	31	32	33	34	35	36
	S	S-N	X-Y	L	A-N	W	Y

VK3_CDR1-7	S	S-N	N-Y	L	A	W	Y
------------	---	-----	-----	---	---	---	---

VK3_CDR1-8	30	30A	31	32	33	34	35	36
	S	S-N	X-S	Y	L	A	W	Y

VL1_CDR1-9	30	30A	30B	31	32	33	34	35	36
	I	G	X-S	N	X-T	V	X-N	W	Y

VL2_CDR1-10	30	30A	30B	30C	31	32	33	34	35	36
	V	G	X-G	Y	N	Y	V	S	W	Y

VL3_CDR1-7	30	31	32	33	34	35	36
	X-SK-Q	X-Y	A-V	X-H	W	Y	

CDR2

VH1_CDR2-13	47	48	49	50	51	52	52A	53	54	55	56	57	58
	W	M	G	X-G	I	X-N	P	X-I	X-S	G	X-T	T-A	N

VH3_CDR2-13	W	V	S-A	X-V	I	S	X-G	D-S	G	G-S	X-S	T-K	Y
-------------	---	---	-----	-----	---	---	-----	-----	---	-----	-----	-----	---

VK1_CDR2-10	46	47	48	49	50	51	52	53	54	55
	L	L	I	Y	X-A	A	S	X-S	L	Q-E

VK3_CDR2-10	L	L	I	Y	G-D	A	S	X-S	R	A
-------------	---	---	---	---	-----	---	---	-----	---	---

VL1_CDR2-10	L	L	I	Y	X-S	N	N-S	X-N	R	P
-------------	---	---	---	---	-----	---	-----	-----	---	---

VL2_CDR2-10	L	M-I	I	Y	E-D	V	S-T	X-N	R	P
-------------	---	-----	---	---	-----	---	-----	-----	---	---

VL3_CDR2-10	L	V	I	Y	X-G	D	N-S	X-D	R	P
-------------	---	---	---	---	-----	---	-----	-----	---	---

CDR3

5

VK_CDR3-8	89	90	91	92	93	94	95	96
	Q	Q	Y	X-N	X-S	X-T	P	X-L

VK_CDR3-9	89	90	91	92	93	94	95	95a	96
	Q	Q	Y	X-N	X-S	X-T	P	P	X-L

VL_CDR3-8	89	90	91	92	93	94	95	96
	Q	S-A	W	D	X-S	S	X-N	X-V

VL_CDR3-9	89	90	91	92	93	94	95	95a	96
	Q	S-A	Y	D-A	X-S	S	X-N	X-T	X-V

VL_CDR3-10	89	90	91	92	93	94	95	95a	95b	96
	Q	S-A	W	D	X-S	S	L-S	X-N	X-G	X-V

VL_CDR3-11	89	90	91	92	93	94	95	95a	95b	95c	96
	Q	S-A	W	D	X-S	S	L-S	X-N	X-G	X-P	X-V

VH_CDR3-9	93	94	95	96	97	98	99	100	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/G	F	D				

VH_CDR3-10	93	94	95	96	97	98	99	100	100a	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/G	X-Y	F	D				

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/G	X-Y	X-Y	F	D				

VH_CDR3-12	93	94	95	96	97	98	99	100	100a	100b	100c	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	F	D					

VH_CDR3-13	93	94	95	96	97	98	99	100	100a	100b	100c	100d	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	X-Y	F	D					

VH_CDR3-14	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	X-Y	F	D						

VH_CDR3-15	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	X-Y	F	D							

VH_CDR3-16	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	X-Y	F	D								

VH_CDR3-17	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	101
	A	R	X-D/GX-R/GX-S/GX-S/GX-S/GX-S/G	X-Y	X-Y	X-Y	X-Y	F	D								

VH_CDR3-18	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	101
A	R	X-D/G	X-R/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y	X-Y	X-Y	X-Y	X-Y	F	D

Los diseños adicionales de CDR3 incorporan mayor diversidad dentro de varias posiciones de CDR3, especialmente en la región C terminal. Esta mayor diversidad refleja de forma más estrecha la diversidad observada en la base de datos Kabat. Esto se refleja en los diseños 2 y 3 de CDR (véanse las tablas más adelante).

- 5 Un diseño alternativo de CDR3 incorpora un diseño rico en tirosina que incorpora tirosinas de forma más amplia a lo largo del bucle de CDR3. Aunque las tirosinas se encuentran de forma más amplia a lo largo del bucle de CDR3, la mezcla de un codón de glicina con un codón de tirosina da como resultado codones de cisteína, así como el codón ámbar de terminación. El codón ámbar de terminación y la amplia incorporación de cisteína conduciría a secuencias no productivas de anticuerpo. Por tanto, las glicinas se incluyen como coproductos en posiciones clave en las que se observan cisteínas en las tablas de frecuencia Kabat. Las cisteínas pueden formar puentes disulfuro para estabilizar bucles de CDR3 largos, por tanto, la inclusión de cisteínas en estas posiciones clave puede ser útil para la funcionalidad de CDR3. El aumento de los tamaños de CDR3 también incluye mayor complejidad en posiciones internas, y esto se incorpora en el principio del diseño.

- 10 Lo ideal es que las tirosinas y glicinas puedan incorporarse en todas las posiciones. A fin de introducir estos restos en cada posición sin producir coproductos no deseados tales como el codón ámbar de terminación, se utiliza un procedimiento alternativo de síntesis en el que los grupos de codones se sintetizan por separado y después se combinan y se dividen para la siguiente sesión de síntesis (EA Peters, P J Schatz, S Johnson, y W J Dower, J Bacteriol. 1994 julio; 176(14): 4296-4305.). En este procedimiento, se utilizan dos grupos: El primer grupo utiliza el codón TMC, que codifica Y y S, y el segundo grupo utiliza el codón VRC, que codifica H, S, R, N, y D. Estas posiciones permiten, por tanto, una contribución hidrofóbica mediante la tirosina, y múltiples contribuciones polares con el segundo grupo. Todas las posiciones de diversidad que están anotadas en verde, más adelante, se generan usando grupos divididos de estos codones.

Todos estos diseños múltiples de CDR dan múltiples sub bibliotecas de las bibliotecas universales. El estado físico y el funcionamiento globales de cada diseño se ensayan de forma empírica contra múltiples antígenos.

25 Tabla 13 Diseño 2

VH_CDR3-9	93	94	95	96	97	98	99	100	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	F	D

VH_CDR3-10	93	94	95	96	97	98	99	100	100a	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	Y/A/N	F	D

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	Y/A/N	Y/A/N	F	D

VH_CDR3-12	93	94	95	96	97	98	99	100	100a	100b	100c	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	Y/A/N	Y/A/N	F	D

VH_CDR3-13	93	94	95	96	97	98	99	100	100a	100b	100c	100d	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	F	D

VH_CDR3-14	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	F	D

VH_CDR3-15	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	X-S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	F	D

VH_CDR3-16	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	X-S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	Y/A/N	F	D

VH_CDR3-17	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	X-S/G	X-S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	Y/A/N	F	D

VH_CDR3-18	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	101
	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-	X-
A	R	S/G	S/G	S/G	S/G	S/G	S/G	X-S/G	X-S/G	X-S/G	X-S/G	Y/A/N	Y/A/N	Y/A/N	Y/A/N	Y/A/N	F	D

Tabla 14 Diseño 3

VH_CDR3-9	93	94	95	96	97	98	99	100	101		
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	F	D		

VH_CDR3-10	93	94	95	96	97	98	99	100	100a	101	
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-Y/A	F	D	

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	F	D

VH_CDR3-12	93	94	95	96	97	98	99	100	100a	100b	100c	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	F	D

VH_CDR3-13	93	94	95	96	97	98	99	100	100a	100b	100c	100d	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	F	D

VH_CDR3-14	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	F	D	

VH_CDR3-15	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	F	D	

VH_CDR3-16	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	X-Y/A	F	D	

VH_CDR3-17	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	X-Y/A	F	D		

VH_CDR3-18	93	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	X-Y/A	X-Y/A	X-Y/A	F	D		

Ejemplo 10

5 DISEÑO DE OLIGONUCLEÓTIDOS PARA LA INTRODUCCIÓN DE DIVERSIDAD DE CDR USANDO WTM Y WTM EXTENDIDA

La construcción del oligo puede llevarse a cabo usando las secuencias expuestas en la tabla 12.

10 La mutagénesis guiada y la mutagénesis guiada extendida (para CDRH3) se llevaron a cabo en las posiciones apropiadas sombreadas en la tabla 12, donde se indicaron en la secuencia señalados con una X. La X se refiere al aminoácido guiado, y el (los) aminoácido(s) siguientes(s) al (guión)- se refieren al aminoácido base y a cualquiera de los coproductos requeridos señalados tras una (barra) /. Las posiciones en blanco con múltiples aminoácidos enumerados señalan una mezcla equivalente de aquellos aminoácidos con el número mínimo de coproductos. Esta mezcla refleja la mezcla predominante de estos aminoácidos presente en el perfil de variabilidad.

Por ejemplo, VH1_CDR1-6 se describe como:

15

Tabla 15

VH1_CDR1-6	30	31	32	33	34	35
	T-S	S	Y	X-A	I-M	X-S

Si se elige que el aminoácido guiado sea alanina, entonces se usan los siguientes codones para el diseño anterior.

Table 16.

5'-																	
30			31			32			33			34			35		
W	C	C	T	C	C	T	A	C	G	C	C	A	T	S	K	C	C
-3'																	

20 Para la posición 30 (numeración Chothia), TCC codifica serina y ACC codifica treonina, por tanto, la mezcla más eficaz es WCC.

Para la posición 31, TCC codifica S.

Para la posición 32, TAC codifica Y.

Para la posición 33, el aminoácido guiado es idéntico al aminoácido base, por tanto, se usa el codón del aminoácido base GCC, que codifica A.

- 5 Para la posición 34, ATS codifica I y M, en la que ATC codifica I y ATG codifica M.

Para la posición 35, se usa el procedimiento convencional de mutagénesis guiada. TCC es serina, y GCC es el emparejamiento de alanina más próximo. Por tanto, tanto G como T se requieren en la primera posición, C se requiere en la segunda posición, y C se requiere en la tercera posición, Por tanto, se usa KCC, que codifica A y S.

- 10 En la práctica, los oligonucleótidos se sintetizan con regiones adyacentes complementarias a la región variable del anticuerpo. Por tanto, se usa la siguiente secuencia:

Tabla 17

			<table><tr><td>30</td><td>31</td><td>32</td><td>33</td><td>34</td><td>35</td></tr><tr><td>W</td><td>C</td><td>C</td><td>T</td><td>C</td><td>C</td><td>T</td><td>A</td><td>C</td><td>G</td><td>C</td><td>C</td><td>A</td><td>T</td><td>S</td><td>K</td><td>C</td><td>C</td></tr></table>																30	31	32	33	34	35	W	C	C	T	C	C	T	A	C	G	C	C	A	T	S	K	C	C			
30	31	32	33	34	35																																								
W	C	C	T	C	C	T	A	C	G	C	C	A	T	S	K	C	C																												
VH1_1_6_WA	5'-	GCTTCGGGTGGCACAATTC																	TGGGTTAGACAGGCACCT	-3'																									

Los 20 aminoácidos y los aminoácidos no naturales que utilizan el codón ámbar pueden guiarse potencialmente a las posiciones apropiadas sombreadas en azul/verde. Para ejemplificar, se muestran nueve aminoácidos a continuación.

15

Tabla 18

VH1_1_6_WA	5'-	GCTTCGGGTGGCACAATTC	30	31	32	33	34	35	TGGGTTAGACAGGCACCT	-3'										
			W	G	C	T	C	C	T	A	C	G	C	C	A	T	S	K	C	C
VH1_1_6_WD	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	G	M	C	A	T	S	K	M	C
VH1_1_6_WS	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	K	C	C	A	T	S	T	C	C
VH1_1_6_WI	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	R	Y	C	A	T	S	A	K	C
VH1_1_6_WP	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	S	C	C	A	T	S	Y	C	C
VH1_1_6_WR	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	S	S	C	A	T	S	M	G	C
VH1_1_6_WY	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	K	M	C	A	T	S	T	M	C
VH1_1_6_WH	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	S	M	C	A	T	S	Y	M	C
VH1_1_6_WN	5'-	GCTTCGGGTGGCACAATTC	W	G	C	T	C	C	T	A	C	R	M	C	A	T	S	A	R	C

- 20 Para comprender la nomenclatura, VH1 es la región marco conservada VH1_1, se refiere a CDR1 de VH1, VH1_6 se refiere a CDR de tamaño 6 y W se refiere a guía y la letra final es el aminoácido guiado. Las secuencias anteriores ejemplifican guía con A (alanina) D (aspartato), S (serina), I (isoleucina), P (prolina), R (arginina), Y (tirosina), H (histidina), y N (asparagina).

La construcción de oligo usando la tabla 12 se llevó a cabo usando mutagénesis guiada extendida y dopaje de la forma siguiente.

- 25 La mutagénesis guiada y la mutagénesis guiada extendida (para CDRH3) se llevaron a cabo en las posiciones apropiadas sombreadas en azul o verde en la tabla 12, donde se indicaron en la secuencia señalados con una X. La X se refiere al aminoácido guiado, y el (los) aminoácido(s) siguientes(s) al (guión)- se refieren al aminoácido base y a cualquiera de los coproductos requeridos señalados tras una (barra) /. Las posiciones en blanco con múltiples aminoácidos enumerados señalan una mezcla equivalente de aquellos aminoácidos con el número mínimo de coproductos. Esta mezcla refleja la mezcla predominante de estos aminoácidos presente en el perfil de variabilidad.
- 30

Tabla 19.

VH_CDR3-10	93	94	95	96	97	98	99	100	100a	101
	A	R	X-D/G	X-R/G	X-S/G	X-S/G	X-S/G	X-Y	F	D

5

Para la posición 95, el aminoácido base es el aspartato, GAC, Alanina es GCC, y glicina es el coproducto requerido GGC. Por tanto, G está en la primera posición, A, G, y C están en la segunda posición y C está en la tercera posición.

10

15

Las posiciones 98 y 99 son idénticas a la posición 97, ya que utilizan los mismos aminoácidos base y coproductos necesarios.

20

Estos resultados se resumen a continuación.

Tabla 20.

5'-	95			96			97			98			99			100			-3'
	G	A	C	C	G	C	G	C	C	G	C	C	G	C	C	T	A	C	
		G		G	C		A	G		A	G		A	G		G	C		
		C																	

25

30

Tabla 21[illegible]

35

aminoácidos guiados para el tamaño 10 de CDR3 de VH.

Tabla 22

VH_3_10_WA	5'-	ACCGCTGTGTATTACTGT	95 96 97 98 99 100																				TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	G	C	G	C	G	C	G	C	G	C	G	T	
VH_3_10_WD	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	T	C	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	T	C	G	C	A	G	C	A	G	C	A	C	
VH_3_10_WS	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	C	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	C	G	C	A	G	C	A	G	C	A	C	
VH_3_10_WI	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	A	G	C	G	G	C	G	C	T	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	A	G	C	G	G	C	G	C	T	
VH_3_10_WP	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	C	T	G	G	T	G	G	T	G	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	C	T	G	G	T	G	G	T	G	
VH_3_10_WR	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	
VH_3_10_WY	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	
VH_3_10_WH	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	
VH_3_10_WN	5'-	ACCGCTGTGTATTACTGT	G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	TACTGGGTCAGGGCACACTG
			G	C	C	A	G	A	G	A	C	G	G	C	A	G	C	A	G	C	A	C	

La construcción de oligo usando los diseños 2, 3 y 4 de CDRH3 que utilizan mutagénesis guiada extendida y dopaje se realizó de la forma siguiente.

La mutagénesis guiada y la mutagénesis guiada extendida (para CDRH3) se realizaron en las posiciones apropiadas sombreadas en azul o verde en la tabla 12, donde se indicaron en la secuencia señalados con una X. La X se refiere al aminoácido guiado, y el (los) aminoácido(s) siguientes(s) al (guión)- se refieren al aminoácido base y a cualquiera de los coproductos requeridos señalados tras una (barra) /.

Tabla 23. Longitud 11 del diseño 2

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	A	R	X-S/G	X-S/G	X-S/G	X-S/G	X-S/G	X-Y/A/N	X-Y/A/N	F	D

Se ofrece este tercer ejemplo para ejemplificar más el proceso de mutagénesis guiada, en este caso, con los diseños adicionales de CDRH3.

Los 20 aminoácidos y los aminoácidos no naturales que utilizan el codón ámbar pueden guiarse potencialmente a las posiciones apropiadas sombreadas en verde. En este ejemplo, la histidina se guía usando el diseño 2, la prolina se guía usando el diseño 3, y la serina se guía usando el diseño 4. El listado de secuencias tiene estos aminoácidos guiados a lo largo de los diseños desde la longitud 9 a la 18. En el presente documento se muestra el análisis para la longitud 11.

Para el diseño 2, la histidina es el aminoácido de ejemplo de mutagénesis guiada/ mutagénesis guiada extendida. En la posición 95, el aminoácido base es serina, y AGC se elige sobre TCC para evitar la formación de cisteínas cuando

se añade glicina como coproducto. La glicina es el coproducto requerido y es GGC. La histidina está codificada como CAC. Por tanto, A, G y C están en la primera posición, G y A están en la segunda posición, y C está en la tercera posición. Las concentraciones de las mezclas se doparon para producir aproximadamente 15-25 % de glicinas, y favorecer la serina, el aminoácido base.

- 5 Las posiciones 96, 97, 98, y 99 utilizan el mismo diseño que la posición 95.

La posición 100 utiliza una tirosina como aminoácido base TAC, y alanina es GCC, y la asparagina AAC se requiere como coproducto. La histidina (CAC) es el aminoácido guiado extendido. Por tanto, la primera posición contiene T, G, A y C, la segunda posición contiene A y C, y la posición final contiene C. El dopaje se realiza para favorecer el aminoácido base tirosina.

- 10 La posición 100a utiliza el mismo principio de diseño que la posición 100.

Tras la adición de regiones adyacentes al cebador, el diseño del oligonucleótido para la longitud 11 de CDRH3 queda con histidina del diseño 2 se muestra a continuación y en el listado de secuencias.

Tabla 24[illegible]

- 15

Tabla 25 Longitud 11 del diseño 3

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	A	R	X-D/G	X-D/G	X-S/G	X-S/G	X-S/G	X-Y/A	X-Y/A	F	D

Para el diseño 3, se usa prolina para este ejemplo como aminoácido de mutagénesis quiada/quiada extendida.

- 20 En la posición 95, Asp (GAC) es el aminoácido base, y Gly (GGC) es el coproducto requerido. La guía de prolina (CCC) da como resultado G y C en la primera posición, A, G, y C en la segunda posición, y C en la tercera posición. Las glicinas se dopan para lograr entre 15-25 % de frecuencia como en los ejemplos previos.

La posición 96 utiliza el mismo principio de diseño que la posición 95.

- La posición 97 utiliza serina TCG para la mutagénesis guiada de prolina, ya que el aminoácido guiado no requiere una C o T en la tercera posición. Para un aminoácido guiado que requiere C o T, puede utilizarse AGC para el codón de serina para evitar coproductos de cisteína. Se prefiere TCG sobre AGC a causa del coproducto beneficioso de triptófano TGG frente a la codificación de AGG (arginina). La arginina es deseable, pero CGC ya está codificado en la mezcla final, haciendo a la arginina redundante.

Por tanto, TCG se usa para Serina, GGG para el coproducto requerido de glicina, y CCG para prolina. Por tanto, T, G, y C se usan en la primera posición, G y C en la segunda posición, y G en la posición final.

- 30 La posición 98 y 99 utiliza el mismo principio de diseño que la posición 97.

La posición 100 y 100a utiliza el mismo diseño que usa tirosina como el aminoácido base (TAC), el coproducto requerido Ala (GCC), y el aminoácido guiado extendido Pro (CCC). Por tanto, la primera posición contiene T, C y G, la segunda posición contiene A y C, y la tercera posición contiene C. El dopaje se realiza para favorecer el aminoácido base tirosina.

- 35 Con las regiones adyacentes añadidas, el oligonucleótido para el diseño 3 con guía de prolina se muestra a continuación:

Tabla 26[illegible]

Tabla 27 Longitud 11 del diseño 4

VH_CDR3-11	93	94	95	96	97	98	99	100	100a	100b	101
	A	R	X-G/H	X-G/H	X-Y	X-Y	X-Y/H/D/N	X-Y	X-Y	F	D

5

Para el diseño 4, se usa serina como aminoácido de mutagénesis guiada/ guiada extendida.

Para las posiciones 95 y 96, la glicina (GGC) es el aminoácido base, con histidina (CAC) como el coproducto requerido, serina (AGC) es el codón guiado. Por tanto, A, G y C se usan en la primera posición, G en la segunda posición, y C en la tercera posición.

10 Para la posición 97, 98, 100 y 100a, la tirosina TAC es el aminoácido base, y TCC es el codón de guiado, dando T en la primera posición, A y C en la segunda posición, y C en la posición final.

Para la posición 99, la tirosina (TAC) es el aminoácido base, e histidina (CAC), aspartato (GAC), y asparagina (AAC) son los coproductos requeridos. TCC se utiliza para el codón de serina. Por tanto, A, C, G y T se usan en la primera posición, A y C se usan en la segunda posición, y C se usa en la tercera posición.

15 El dopaje se realiza para favorecer el aminoácido base, y las regiones adyacentes se añaden para dar la siguiente mezcla de oligonucleótidos:

Tabla 28

WH_3_11_WS Design 4 5- ACCGCTGTGTATTACTGT

	95	96	97	98	99	100	100a	
G	C	G	A	G	A	G	C	T
G	A	G	A	T	A	T	A	C
C	55	C	55	40	55	40	55	G
45	45	45	45	60	45	60	45	60
30		30		60		60		60
25		25		40		40		40
				20		20		20
				20		20		20

 TACTGGGTCAGGGCAGCTG

Ejemplo 11

20 DISEÑO DE GRUPO DIVIDIDO, MUTAGÉNESIS/SÍNTESIS DE OLIGONUCLEÓTIDO

Para la identificación de región de unión a antígeno/contactos de antígeno interesantes, pueden incorporarse tirosinas y glicinas en todas las posiciones de restos deseadas. A fin de introducir estos restos en cada posición sin producir coproductos no deseados tales como el codón ámbar de terminación, puede utilizarse un procedimiento alternativo de síntesis en el que los grupos de codones se sintetizan por separado y después se combinan y se dividen para la siguiente sesión de síntesis (EA Peters, P J Schatz, S S Johnson, y W J Dower, J Bacteriol. 1994 julio; 176(14): 4296-4305.). En este procedimiento, se utilizan dos grupos: El primer grupo utiliza el codón TMC, que codifica Y y S, y el segundo grupo utiliza el codón VRC, que codifica H, S, R, N, G y C. Estos grupos, permiten, por tanto, una contribución hidrofóbica mediante la tirosina, y múltiples contribuciones polares y glicina con el segundo grupo. En este diseño de grupo dividido, todas las posiciones de diversidad que se señalan con una X en las tablas de diversidad (figura 12) pueden contener grupos divididos de estos codones. Este ejemplo muestra los conjuntos de codones utilizados para la longitud 9 de CDR de VH3 como se muestra a continuación:

Tabla 29[illegible]

El primer grupo codifica Y y S en una proporción de 50-50. Sin embargo, el segundo grupo se dopa para incrementar la incorporación de glicina a un 15 % después del agrupamiento. El grupo de tirosina está codificado a 1/3 del tamaño del grupo de histidina para obtener una proporción de aminoácidos más equilibrada.

- 5 A fin de producir la mezcla de aminoácidos deseada, se utilizan cuatro columnas de oligonucleótidos. Primero, en las cuatro columnas, la porción 3' fija de los oligonucleótidos se sintetiza como se define mediante las regiones adyacentes y la porción fija de la CDRH3 mostrada anteriormente. Para la posición 99 de la secuencia de ejemplo anterior, la primera columna sintetiza el codón TMC (CMT en la síntesis de ADN 3'-5'). Las tres columnas restantes sintetizan el codón VRC (CRV en la síntesis de ADN 3'-5') utilizando las proporciones de nucleótidos subrayadas anteriormente. Tras acoplarse los tres nucleótidos, las cuatro columnas se abren, el soporte de la síntesis se elimina mediante lavado con acetonitrilo, y las resinas se agrupan. Después de mezclar, la resina se coloca en porciones equivalentes en las cuatro columnas. En este punto, se sintetiza la siguiente posición, la posición 98. Una columna sintetiza el codón TMC como se describe anteriormente, y tres columnas sintetizan la mezcla VRC. La resina se agrupa, se mezcla y se vuelve a fraccionar como se describe para la posición 99. Este procedimiento se repite para la posición 97, 96, y 95. En este punto, la región 5' fija y adyacente se añade a las cuatro columnas, y la mezcla de oligonucleótido resultante de las cuatro columnas puede agruparse e incorporarse dentro de un molde de anticuerpo utilizando un procedimiento de mutagénesis tal como mutagénesis de Kunkel.
- 10
- 15

Equivalentes

- Los expertos en la técnica reconocerán, o serán capaces de determinar usando no más que la experimentación rutinaria, muchos equivalentes a las realizaciones específicas de la invención descrita en el presente documento. Se pretende que las siguientes reivindicaciones engloben dichos equivalentes.
- 20

(i) una región marco conservada de cadena ligera de un clon de región marco conservada de cadena ligera seleccionada del grupo que consiste en V_{K1} -L1 y V_{K1} -L5;

(iii) una región V_{kl} CDR-L1 de longitud siete de acuerdo con la definición de contacto de CDR que comprende la secuencia de aminoácidos SX1X2I (A/N)WY, en la que X1 es A, D, S, I, P, R, Y, H, N, T o K y X2 es A, D, S, I, P

(iv) una región V_{KL} CDR-L2- de longitud diez de acuerdo con la definición de contacto de CDR que comprende la secuencia de aminoácidos LLIYX1ASX2L(Q/E), en la que X1 es A, D, S, I, P, R, Y, H, N, K o G y X2 es A, D, S, I, P, R, Y, H, N o T.

(vi) una región CDR-H1 de longitud seis de acuerdo con la definición de contacto de CDR que comprende la secuencia de aminoácidos (T/S)SYX1(I/M)X2, en la que X1 es A, Y, G, D, S, I, P, R, H, N o G, y X2 es A, D, S, I, P, R, Y, H, o N, o comprende la secuencia de aminoácidos SX1Y X2MX3, en la que X1 es A, D, S, I, P, R, Y, H, N o T y X2 es A, D, S, I, P, R, Y, H, N, W o G y X3 es A, D, S, I, P, R, Y, H, N;

y (viii) una región CDR-H3 seleccionada del grupo que consiste en:

(b) una región CDR-H3, que es una región de longitud diez de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, E, L, V o M y X2 es A, D, S, I, P, R, Y, H, N, G, L, V, K o Q y X3 es A, D, S, I, P, R, Y, H, N, G, T, E, L, V, Q o W y X4 es A, D, S, I, P, R, Y, H, N, G, L, V, W o Q y X5 es A, D, S, I, P, R, Y, H, N, G, T, E, W o L y X6 es A, D, S, I, P, R, Y, H, N, G, L o F:

(d) una región CDR-H3, que es una región de longitud doce de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6 X7 X8FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, E o V y X2 es A, D, S, I, P, R, Y, H, N, G, L, Q o T y X3 es A, D, S, I, P, R, Y, H, N, G, L, T, V o W y X4 es A, D, S, I, P, R, Y, H, N, G, W, L o V y X5 es A, D, S, I, P, R, Y, H, N, G, V, F, T o L y X6 es A, D, S, I, P, R, Y, H, N, G, L, T o E y X7 es A, D, S, I, P, R, Y, H, N, G, T, W o F y X8 es A, D, S, I, P, R, Y, H, N, G, F, T o W;

K y X2 es A, D, S, I, P, R, Y, H, N, G, L, Q o K y X3 es A, D, S, I, P, R, Y, H, N, G, L, V, K o M y X4 es A, D, S, I, P, R, Y, H, N, G, L, T o V y X5 es A, D, S, I, P, R, Y, H, N, G, T, W, L o Q y X6 es A, D, S, I, P, R, Y, H, N, G, V, L, E o T y X7 es A, D, S, I, P, R, Y, H, N, L, V, T, W o G y X8 es A, D, S, I, P, R, Y, H, N, G o F y X9 es A, D, S, I, P, R, Y, H, N, G, F, W o T;

39

R, Y, H, N, G, T, F, E o L y X9 es A, D, S, I, P, R, H, N, G, W o T y X10 es A, D, S, I, P, R, Y, H, N, G, F, W o L;

(g) una región CDR-H3, que es una región de longitud quince de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6 X7 X8 X9 X10 X11FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, E, V o T y X2 es A, D, S, I, P, R, Y, H, N, G, W, L, T o V y X3 es A, D, S, I, P, R, Y, H, N, G, E, T, F, L o W y X4 es A, D, S, I, P, R, Y, H, N, G, E, C, T o F y X5 es A, D, S, I, P, R, Y, H, N, G, T, W o L y X6 es A, D, S, I, P, R, Y, H, N, G, T o E y X7 es A, D, S, I, P, R, Y, H, N, G, T, V o W y X8 es A, D, S, I, P, R, Y, H, N, G, T, L, F, V, M o W y X9 es A, D, S, I, P, R, Y, H, N, G, V, C o K y X10 es A, D, S, I, P, R, Y, H, N, G, W o Q y X11 es A, D, S, I, P, R, Y, H, N, G, W, F o L;

(h) una región CDR-H3, que es una región de longitud dieciséis de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6 X7 X8 X9 X10 X11 X12FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, L, V o E y X2 es A, D, S, I, P, R, Y, H, N, G, L, V o E y X3 es A, D, S, I, P, R, Y, H, N, G o L y X4 es A, D, S, I, P, R, Y, H, N, G, L, T o E y X5 es A, D, S, I, P, R, Y, H, N, G, V, F, T o M y X6 es A, D, S, I, P, R, Y, H, N, G, T, F o W y X7 es A, D, S, I, P, R, Y, H, N, G, T, V L o E y X8 es A, D, S, I, P, R, Y, H, N, G, T, E o W y X9 es A, D, S, I, P, R, Y, H, N, G, L, F o W y X10 es A, D, S, I, P, R, Y, H, N, G, L, T o F y X11 es A, D, S, I, P, R, Y, H, N, G, W o T y X12 es A, D, S, I, P, R, Y, H, N, G, W o F;

(i) una región CDR-H3, que es una región de longitud diecisiete de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6 X7 X8 X9 X10 X11 X12 X13FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, V, E o L y X2 es A, D, S, I, P, R, Y, H, N, G, L o Q y X3 es A, D, S, I, P, R, Y, H, N, G, L, T, V o M y X4 es A, D, S, I, P, R, Y, H, N, G, V o C y X5 es A, D, S, I, P, R, Y, H, N, G, V o T y X6 es A, D, S, I, P, R, Y, H, N, G, F, V o T y X7 es A, D, S, I, P, R, Y, H, N, G, W, T, V o F y X8 es A, D, S, I, P, R, Y, H, N, G, L o V y X9 es A, D, S, I, P, R, Y, H, N, G, V, F, L o C y X10 es A, D, S, I, P, R, Y, H, N, G, F o L y X11 es A, D, S, I, P, R, Y, H, N, G, L, V o C y X12 es A, D, S, I, P, R, Y, H, N o G y X13 es A, D, S, I, P, R, Y, H, N, G o W; y

(j) una región CDR-H3, que es una región de longitud dieciocho de acuerdo con la definición de contacto de CDR, que comprende la secuencia de aminoácidos ARX1X2X3X4X5 X6 X7 X8 X9 X10 X11 X12 X13 X14FD, en la que X1 es A, D, S, I, P, R, Y, H, N, G, E, V o L y X2 es A, D, S, I, P, R, Y, H, N, G, L o K y X3 es A, D, S, I, P, R, Y, H, N, G, T, V, L o F y X4 es A, D, S, I, P, R, Y, H, N, G, V o T y X5 es A, D, S, I, P, R, Y, H, N, G, C, F, K, L o M y X6 es A, D, S, I, P, R, Y, H, N, G, V, F, T, W o C y X7 es A, D, S, I, P, R, Y, H, N, G, T, W o F y X8 es A, D, S, I, P, R, Y, H, N, G, W, V o F y X9 es A, D, S, I, P, R, Y, H, N, G, L o V.

Diagrama de flujo

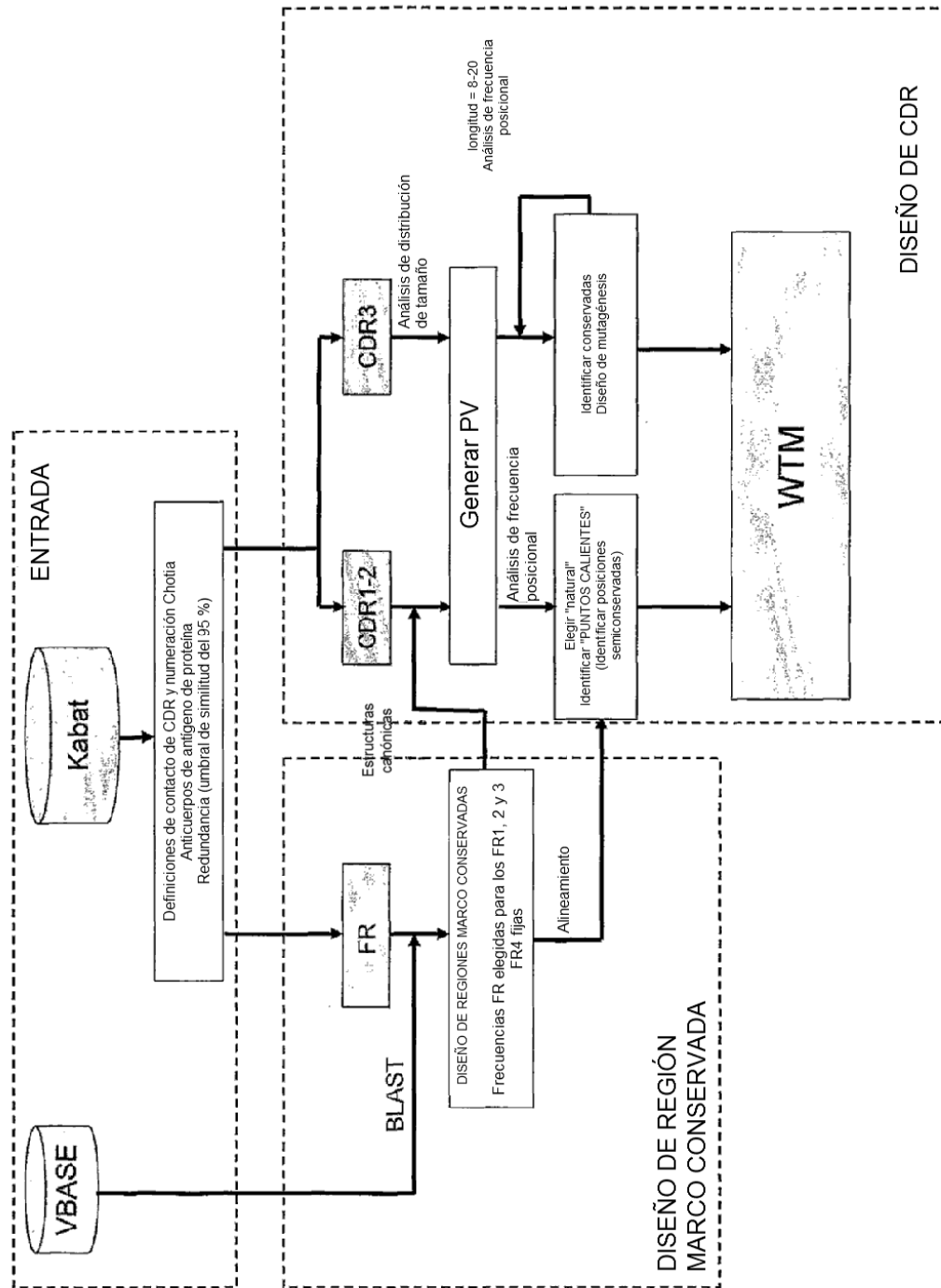
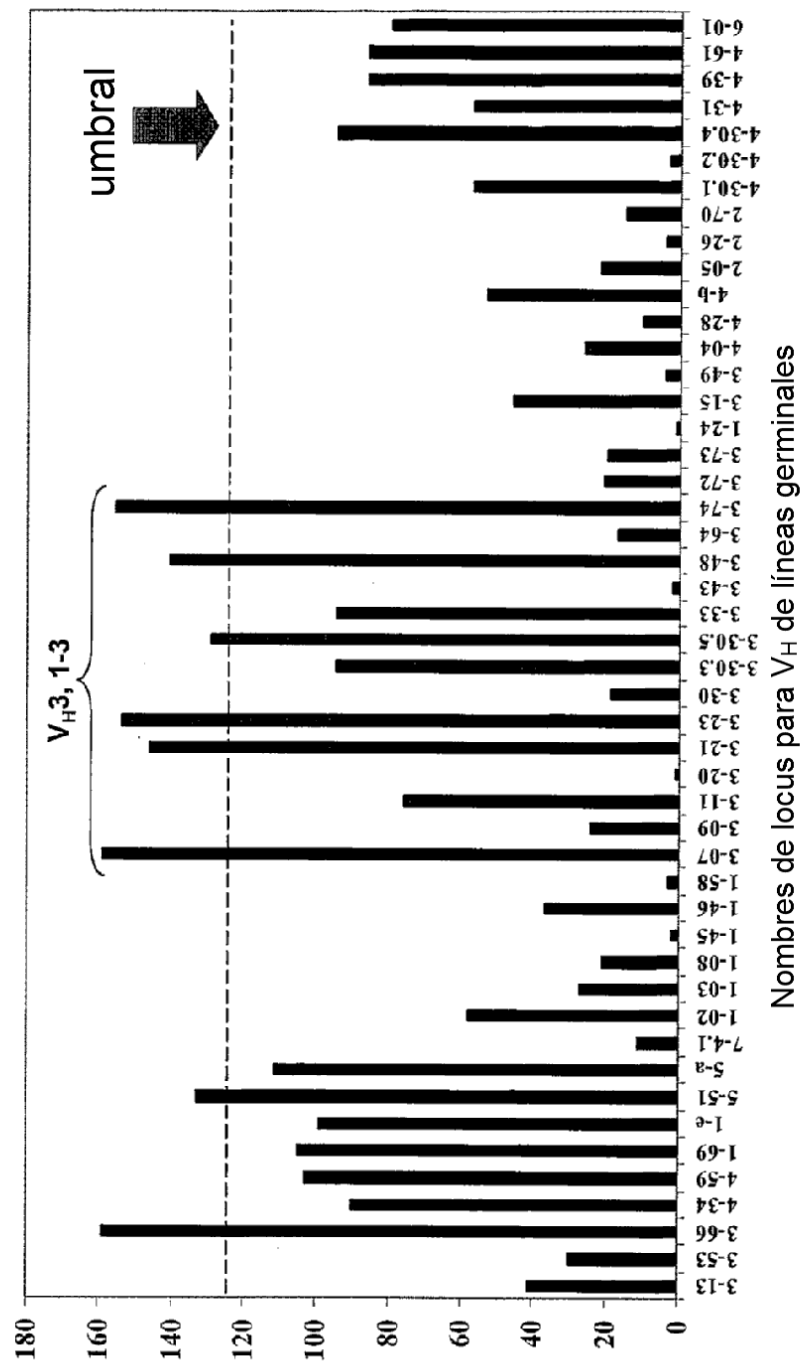


Fig. 1

Fig. 2



Análisis de frecuencia de regiones marco conservadas de V_H de línea germinal



Análisis de frecuencia de regiones marco conservadas de V_H de línea germinal

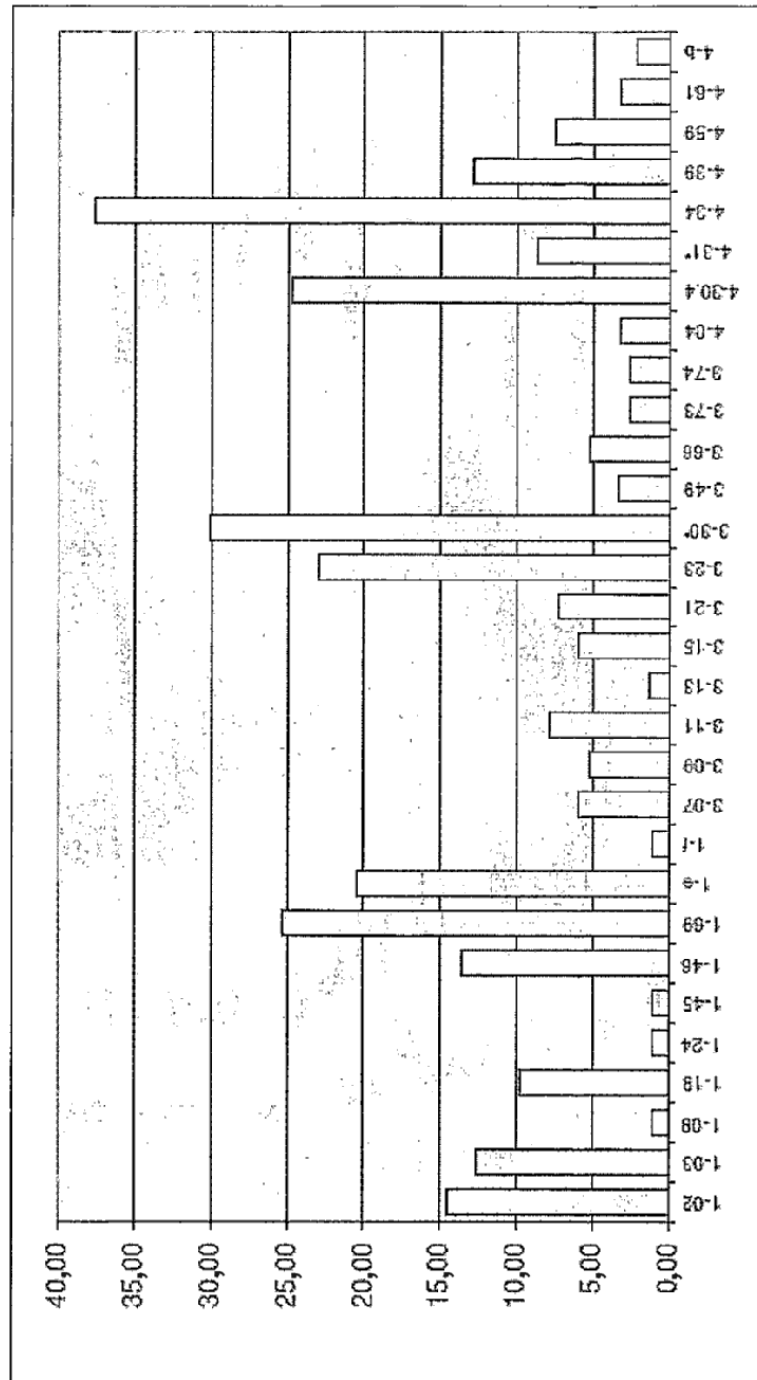


Fig. 4

Nombres de locus para V_H de líneas germinales

Regiones marco conservadas de V_H de alta frecuencia

VH

	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4	
	1	2	3	4	5	6	7	8
12345678901234567890123456789	01ab2345	67890123456	7890123456	789012345678	901234567890123456789012abc3456789012	34567890abc1	234567890123	
1-e	QVQLVQSGAEVKRPGSTVNVSCIASGETF	SS--YAIS	WVRQAPCKGLE	WVGGLPIFGTAN	IAQRFQGRVITIRADKSTSTAYNELSSLRSEDTAVYYC	APCGSGSGYAFD	YWGGGILVTVSS	
3-07	EVQLVESGGGLVQPGGSLRLSCAASGETF	SS--YQMS	WVRQAPCKGLE	WVANIKQDSSEKY	YVDSVKGSETTIERDNAKNELYLQNKSLPAEDTAVYYC	APCGSGSGYAFD	YWGGGILVTVSS	
3-11	QVQLVESGGSEWKPGGSLRSCIASGETF	SD--YVMS	WIRQAPCKGLE	WVSIVSSSGSTIY	YADSVKGEFTTIERDNAKNELYLQNKSLPAEDTAVYYC	APCGSGSGYAFD	YWGGGILVTVSS	
3-23	EVQLLESGGCHVQPGSSRLSCLASGETF	SS--YAMS	WVRQAPCKGLE	WVSAISGSGSTY	YADSVKGEFTTIERDNKNTLYLQNKSLPAEDTAVYYC	ANDGSGSGYAFD	YWGGGILVTVSS	
3-33	QVQLVESGGGVVQPGSLRLSCAASGETF	SS--YGMH	WVRQAPCKGLE	WVAIVWDGSKNY	YADSVKGEFTTIERDNKNTLYLQNKSLPAEDTAVYYC	ANDGSGSGYAFD	YWGGGILVTVSS	
4-34	QVQLQKQAGLHPSETISLTCAVYGGSEF	SG--YYMS	WIRQPPCKGLE	WIGEL-NHSGSTN	YNPSLAKSRVTISVDTSKNQFSLKLSSTVTRADTAVYYC	APDGSGSGYAFD	YWGGGILVTVSS	
4-36.4	QVQLQKQSGPLMKRPSQTLISLTCAVYGGESI	SGGDYYMS	WIRQPPCKGLE	WIGVI-YYSGSTY	YNPSLAKSRVTISVDTSKNQFSLKLSSTVTRADTAVYYC	ANDGSGSGYAFD	YWGGGILVTVSS	

V_H-CDR1

VBASE

	** *
1-3@3-07	SYWMS
1-3@3-21	SYSMN
1-3@3-23	SYAMS
1-3@3-30.5	SYGMH
1-3@3-48	SYSMN
1-3@3-74	SYWMH

Kabat

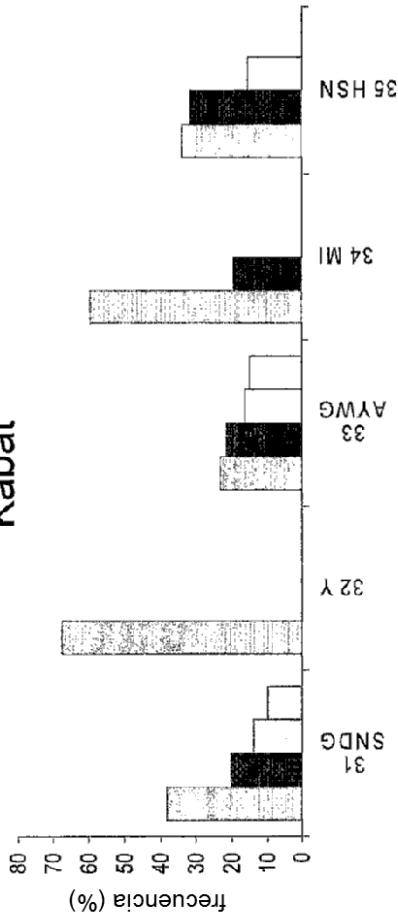


Fig. 6

• Todos los bucles CDR1-1 tienen 5 aminoácidos (definición de Kabat)

• La posición 32 está conservada

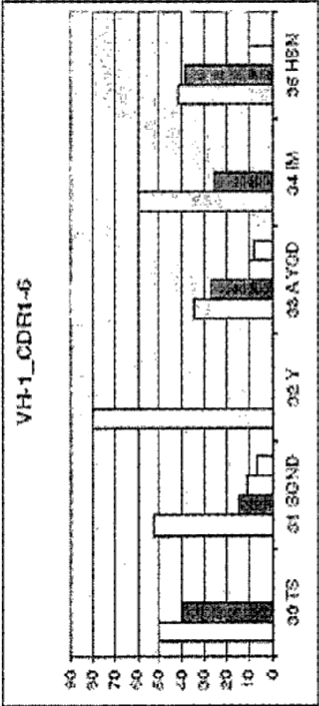
• Las posiciones 31 y 34 se mutan durante la maduración de la afinidad

• Las posiciones 33 y 35 se mutan

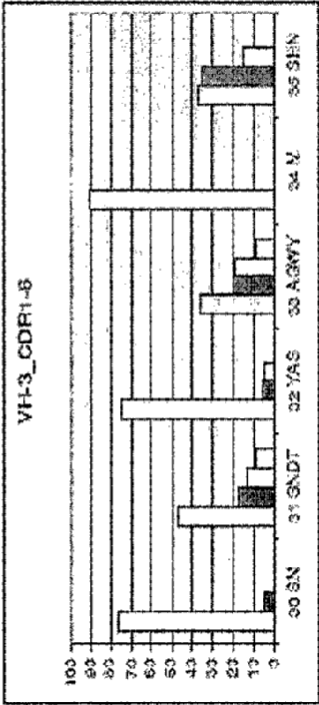
SYXMX

- POSICIONES INVARIABLES
- WTM EN BIBLIOTECAS UNIVERSALES
- WTM EN MADURACIÓN DE LA AFINIDAD

Matrices de secuencias de V_H-CDR1 (definición de contacto)



30	31	32	33	34	35
T-S	S	Y	X _A	I-M	X _S



30	31	32	33	34	35
S	X _S	Y	X _A	M	X _S

A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30																			
31																			
32																			
33																			
34																			
35																			

Fig. 7

A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30																			
31																			
32																			
33																			
34																			
35																			

V_H-CDR2

VBASE
* . . . *
1-3@3-07 NIKQDGSSEKYYVDSVKG
1-3@3-21 SSSSSSYIYYADSVKG
1-3@3-23 AISGSGSTYYADSVKG
1-3@3-30.5 VISYDGSNKYYADSVKG
1-3@3-48 YISSSSSTIYYADSVKG
1-3@3-74 RINS DGSSTSYADSVKG

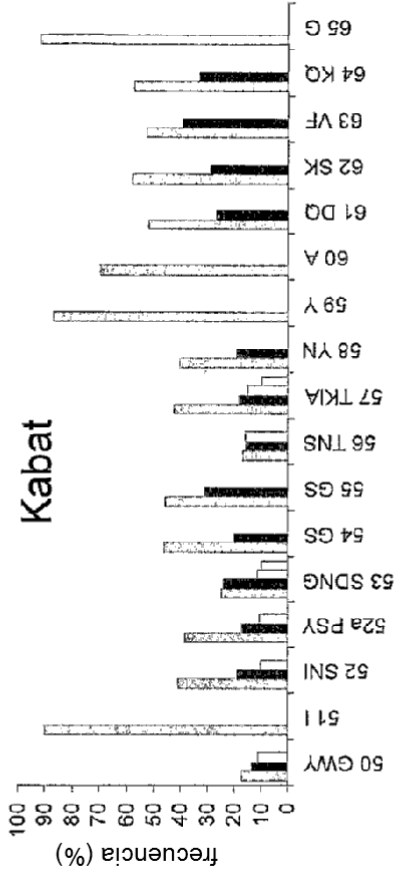


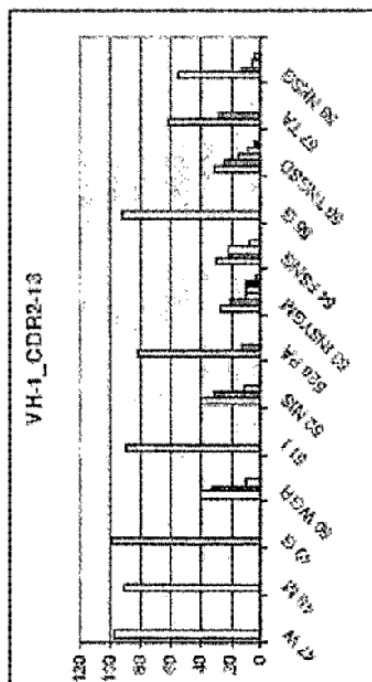
Fig. 8

- Todos los bucles CDR2-3,2 tienen 17 aminoácidos (definición de Kabat)
- Las posiciones 51, 59, 60 y 65 están conservadas como aminoácidos individuales
- Las posiciones 54, 55, 58, 61, 62, 63 y 64 se mutan durante la maduración de la afinidad
- Las posiciones 50, 52, 52a, 53, 56 y 57 se mutan

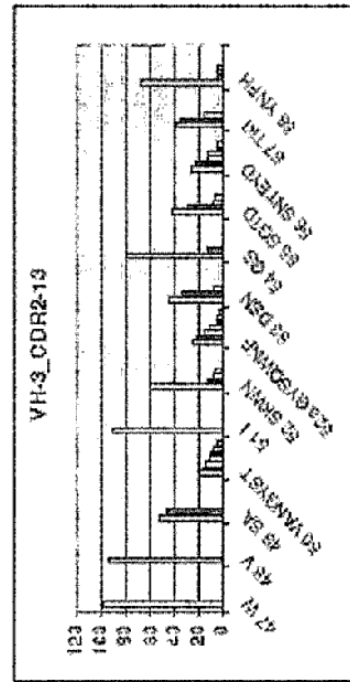
XIXXXGXXYYADSVKG

• POSICIONES INVARIABLES
• WTM EN BIBLIOTECAS UNIVERSALES
• WTM EN MADURACIÓN DE LA AFINIDAD

Matrices de secuencias de V_H-CDR2 (definición de contacto)



47	48	49	50	51	52	52A	53	54	55	56	57	58
W	M	G	X ₁	I	X ₂	P	XI	X ₃	G	X ₄	T-A	N



47	48	49	50	51	52	52A	53	54	55	56	57	58
W	V	S-A	X ₁	I	S	X ₂	X ₃	G	X ₄	X ₅	X ₆	Y

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
47																				X
48																				
49										X										
50	X	X	X	X	X	X	X	X	X						X					
51															X					
52	X	X	X	X	X	X	X	X	X											
52A																				
53	X	X	X	X	X	X	X	X	X											
54	X	X	X	X	X	X	X	X	X											
55	X	X	X	X	X	X	X	X	X											
56	X	X	X	X	X	X	X	X	X						X			X	X	
57	X	X	X	X	X	X	X	X	X											
58									X											

Fig. 9

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
47																				X
48																			X	
49	X	X	X	X	X	X	X	X	X											
50	X	X	X	X	X	X	X	X	X											
51																				
52	X	X	X	X	X	X	X	X	X											
52A	X	X	X	X	X	X	X	X	X						X					
53	X	X	X	X	X	X	X	X	X											
54															X					
55	X	X	X	X	X	X	X	X	X						X					
56	X	X	X	X	X	X	X	X	X											
57	X	X	X	X	X	X	X	X	X											
58									X											

Análisis de longitud de V_H -CDR3

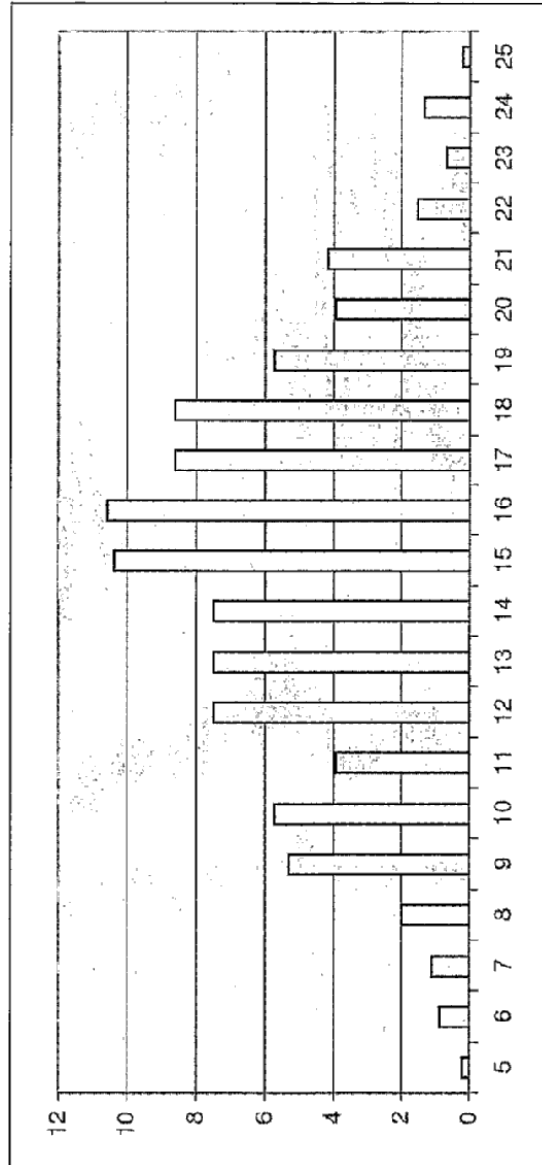


Fig. 10

- La variabilidad de longitud de la CDR3 de V_H oscila de 4 a 28 restos (definición de contacto)
- Sin embargo, la mayoría (~75 %) de las CDR3 de V_H de anticuerpos antiproteína oscila de 9 a 18 restos.
- Las 13 longitudes (9, 10,...18) se sintetizan por separado y se almacenan antes del ensamblaje genético en las proporciones deseadas.

V_H-CDR3 (definición de Kabat de longitud 13)

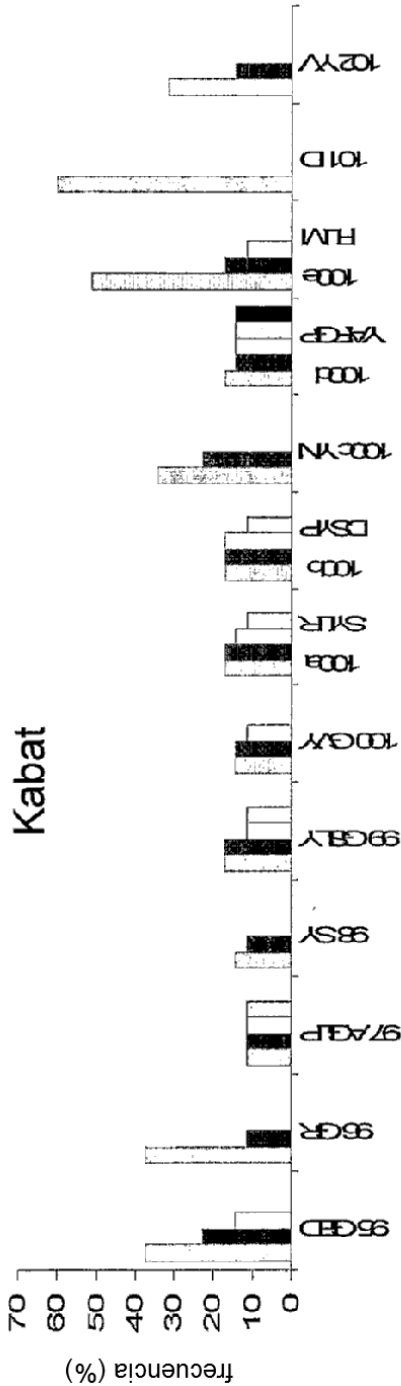


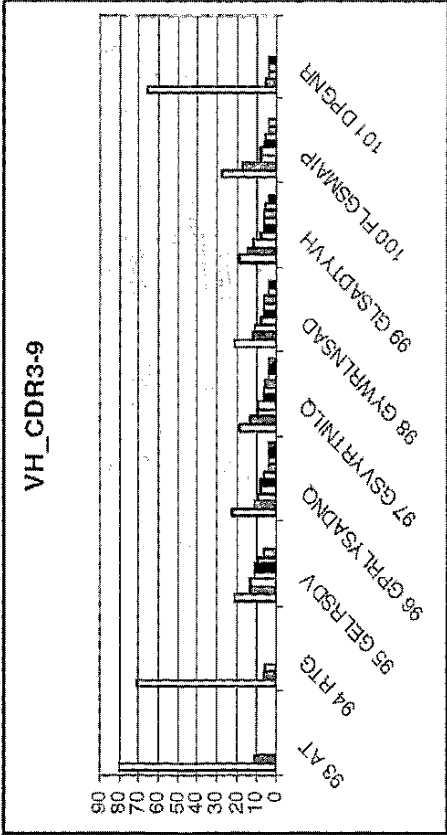
Fig. 11

- Las posiciones 100e y 101 están conservadas
- Las posiciones 95, 96, 97, 98, 99, 100, 100a, 100b, 100c, 100d y 102 se mutan.

XXXXXXXXXXXXFDX

• POSICIONES INVARIABLES
• WTM EN BIBLIOTECAS UNIVERSALES
• WTM EN MADURACIÓN DE LA AFINIDAD

Matrices de secuencias de V_H-CDR3 (definición de contacto)



93	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V
94	X					X													
95	X	X	X	X	X	X	X	X	X			X			X				X
96	X	X	X	X	X	X	X	X	X			X			X				X
97	X	X	X	X	X	X	X	X	X			X			X				X
98	X	X	X	X	X	X	X	X	X			X			X				X
99	X	X	X	X	X	X	X	X	X			X			X				X
100														X					
101		X																	

Fig. 12

93	94	95	96	97	98	99	100	101
A	R	X _{D/G}	X _{R/G}	X _{S/G}	X _{S/G}	X _{S/G}	F	D

Fig. 12

Matrices de secuencias de V_H-CDR3 (continuación)

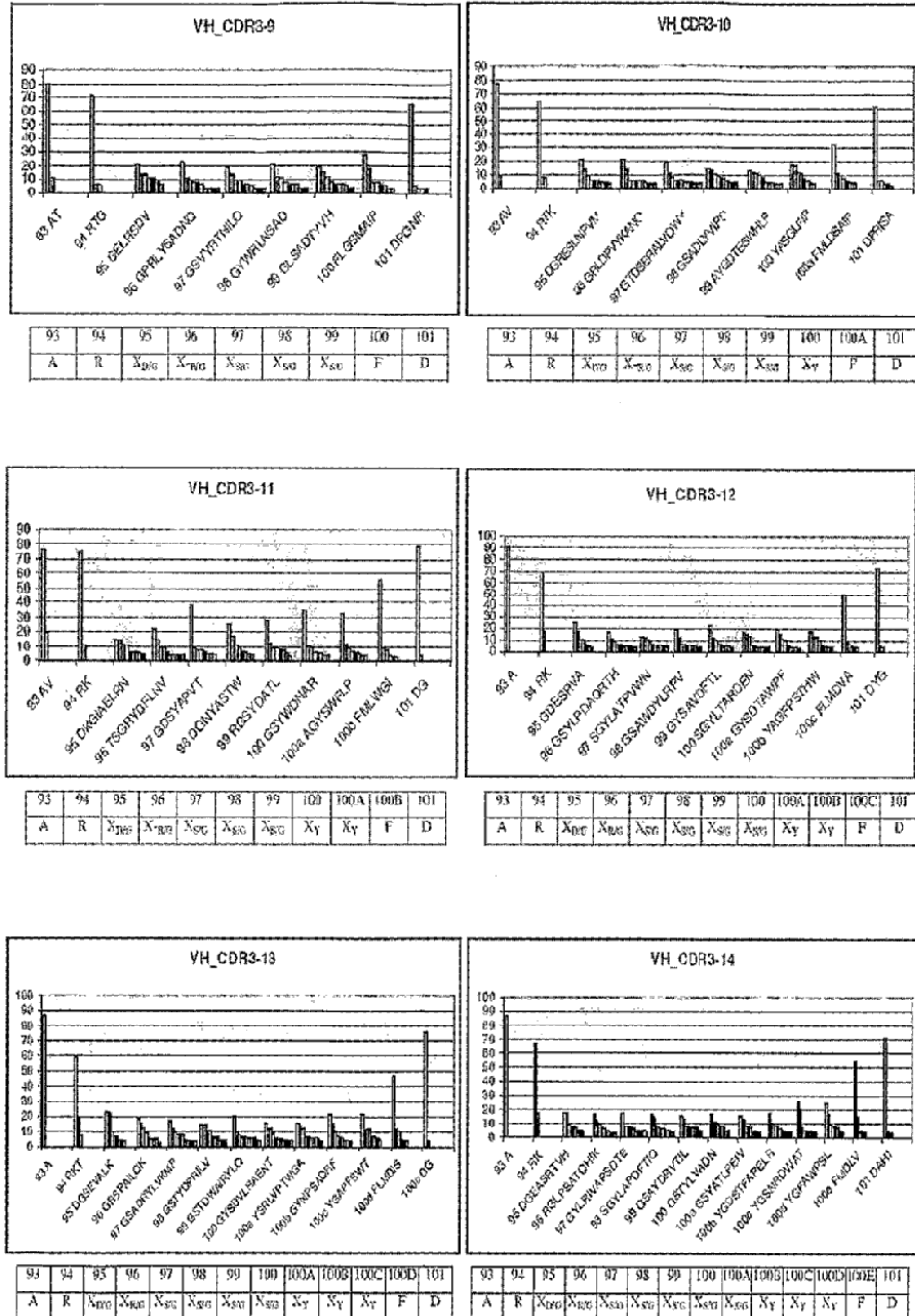
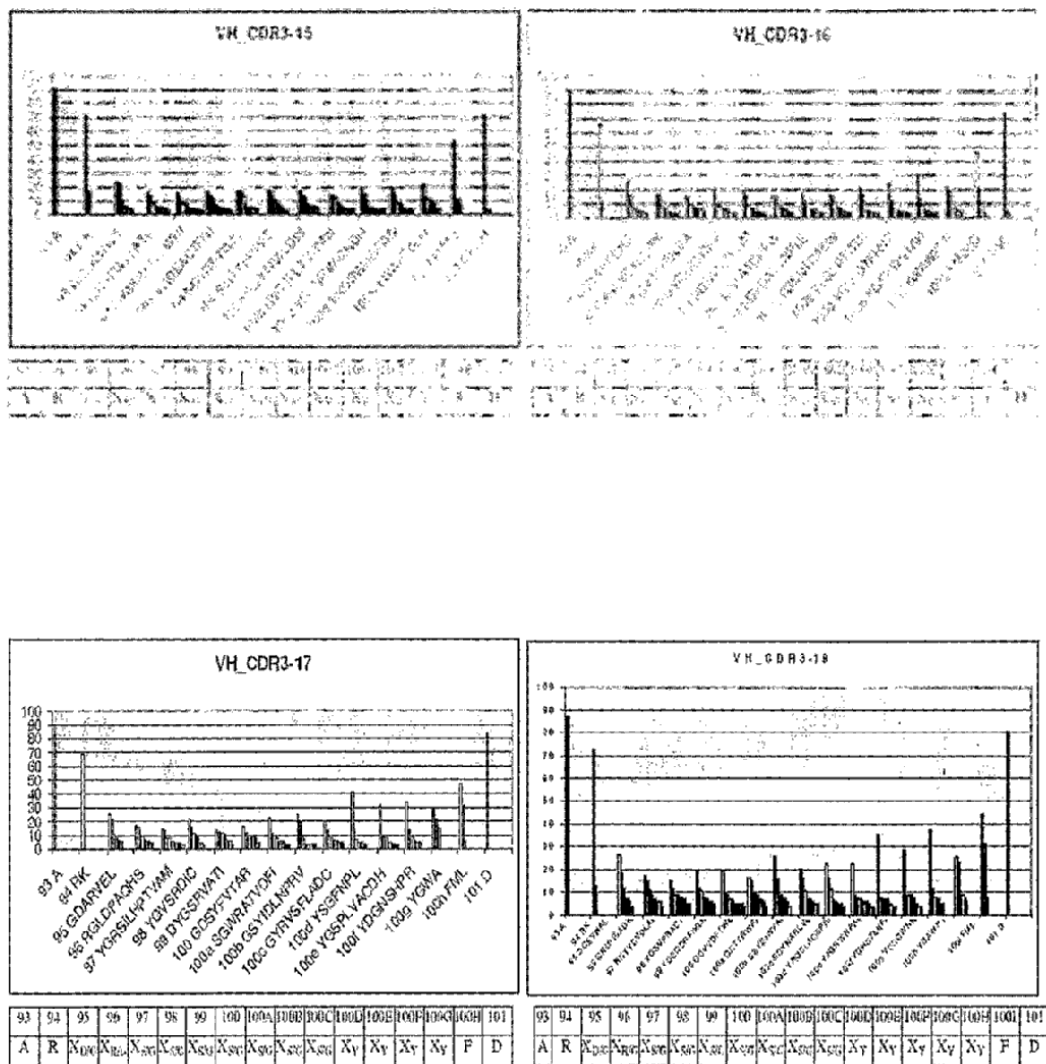


Fig. 12

Matrices de secuencias de V_H-CDR3 (continuación)



Matrices de secuencias de V_H-CDR3 (continuación)

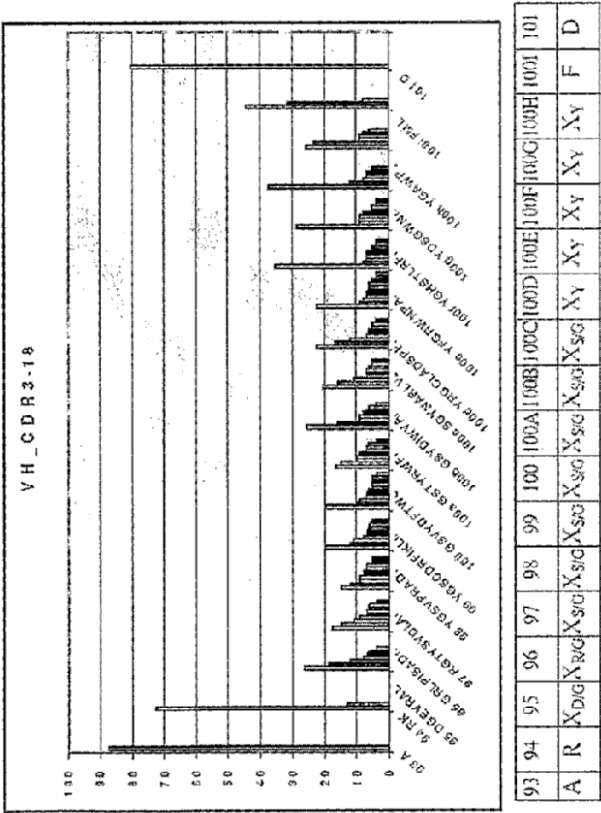


Fig. 12

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
93	X																			
94						X														
95	X	X	X	X	X	X	X	X	X						X				X	
96	X	X	X	X	X	X	X	X	X						X				X	
97	X	X	X	X	X	X	X	X	X						X				X	
98	X	X	X	X	X	X	X	X	X						X				X	
99	X	X	X	X	X	X	X	X	X						X				X	
100	X	X	X	X	X	X	X	X	X						X				X	
100a	X	X	X	X	X	X	X	X	X						X				X	
100b	X	X	X	X	X	X	X	X	X						X				X	
100c	X	X	X	X	X	X	X	X	X						X				X	
100d	X	X	X	X	X	X	X	X	X						X				X	
100e	X	X	X	X	X	X	X	X	X						X				X	
100f	X	X	X	X	X	X	X	X	X						X				X	
100g	X	X	X	X	X	X	X	X	X						X				X	
100h	X	X	X	X	X	X	X	X	X						X				X	
100i															X					
101	X																			

Diversidad de la biblioteca de V_H

Diversidad de CDR mediante WTM (bajo dopaje $\Rightarrow$ 1 a 2 sustituciones por molécula) con todos los aminoácidos excepto Cys.

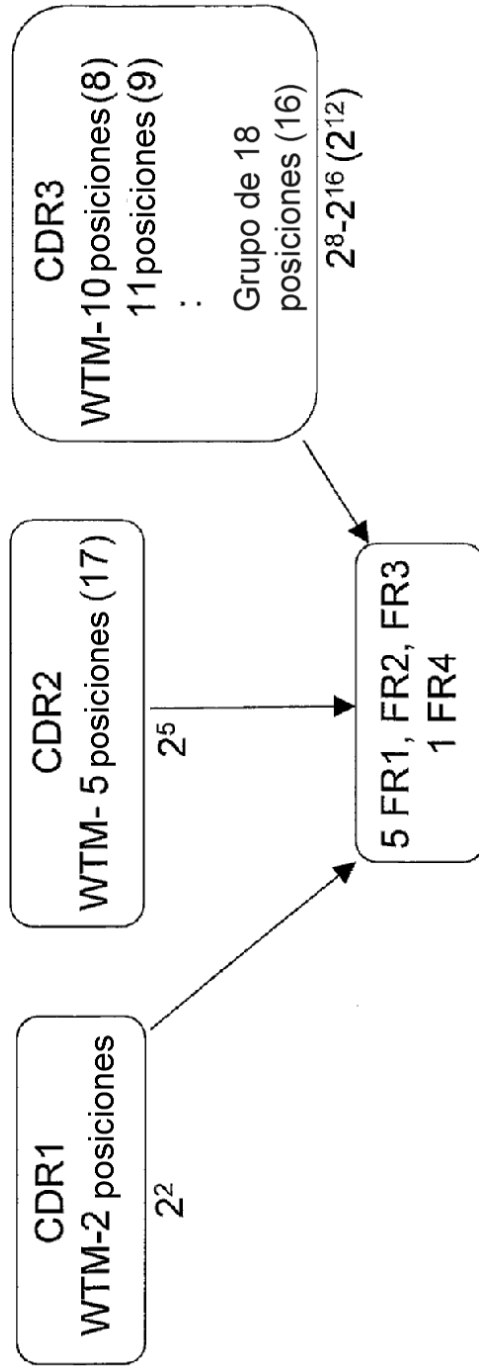


Fig. 13

- Combinar la biblioteca de V_H con la biblioteca de V_L ($V_k : V_\lambda$ a 2:1)
- Todos los oligos contienen 20-meros solapados no degenerados en los extremos 5' y 3'

CONSTRUCCIÓN DE BIBLIOTECA V_H

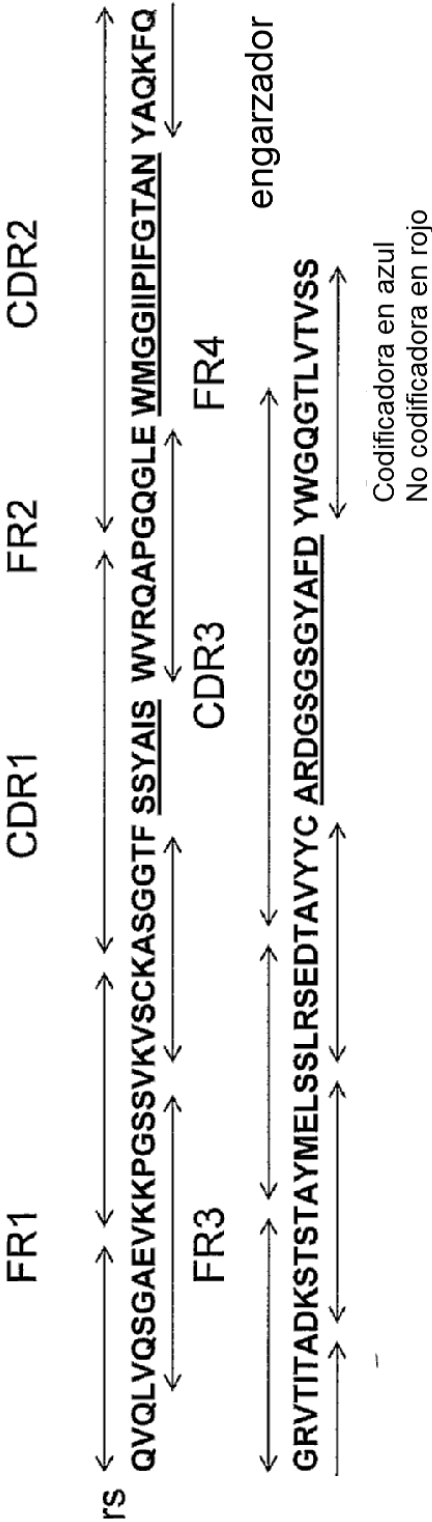


Fig. 14

Número necesario de oligos no degenerados

FR1:	4	} V_H 5 regiones marco conservadas => 55 oligos
FR2:	1	
FR3:	5	
FR4:	1	

Número necesario de oligos degenerados

CDR1:	2	} X 9 (N.º de AA para WTM)
CDR2:	2	
CDR3:	10	

**CONSTRUCCIÓN DE UNA BIBLIOTECA Y ESTRATEGIA DE SELECCIÓN
PARA BIBLIOTECAS MAYORES**

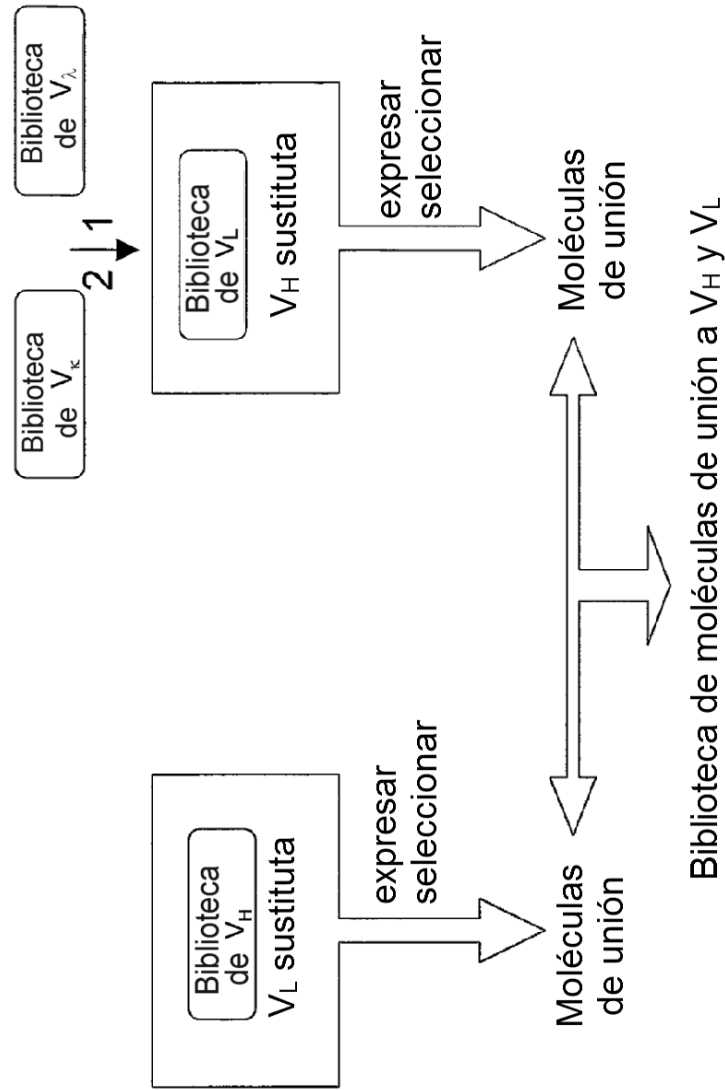


Fig. 15

REGIONES MARCO CONSERVADAS DE CADENA LIGERA KAPPA (κ) Y LAMBDA (λ)

Fig. 16

VK									
	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4		
1	1	2	3	4	5	6	7	8	9
12345678901234567890123456789	01a23456	789012345	6789012345	6789012345	678901234567890123456789012345678	90123456	7890123456a78		
I-21	DIQMIQSPSSLSASVGDRTIICRAQQGI	SN-VLWF	QQKPEHAK	SLIYAASSLQ	SGVPSHESGSGSIDFTLIISLQEDFAIYYC	QQYNSTPL	IFGQETKVELKRT		
III-27	EIVLTQSPGTLSEPERAFLSCRAQQSV	SSSVLAWY	QQKPEHAK	LLIYASGARA	TCIPDRFSGSGSIDFTLIISLQEDFAIYYC	QQYNSTPL	IFGQETKVELKRT		
III-26	EIVLTQSPAILSEPERAFLSCRAQQSV	SS-VLAWY	QQKPEHAK	LLIYASGARA	TCIPDRFSGSGSIDFTLIISLQEDFAIYYC	QQYNSTPL	IFGQETKVELKRT		
VL									
	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4		
2	3	4	5	6	7	8	9	0	
12345678901234567890123456789	01abc23456	789012345	6789012345	678901234567890123456789012345678	90123456a78	7890123456a78			
1b	QSVLTQPPSVSRAPGQVITISGSSSEN	IGNN-VVSWY	QQLEPETHAK	LLIYDNNRKP	SGIPDRFSGSGSGTEATLTITSLQEDFAIYYC	QQWDSGLNGV	VFGSETKLIVLIGQ		
2a2	QCALIQPASVSGSPESQITITGIGTISD	VGGNYVSWY	QQHPEHAK	LMIVYSNRKP	SGVSNRFSGSGSGTASLTITSLQEDFAIYYC	QQWDSGLNGV	VFGSETKLIVLIGQ		
31	SSELTQDRAVSVLIGQIVRIITQEDSIR	SY----VASWY	QQKPEHAK	LMIVYGNRKP	SGIPDRFSGSGSGTASLTITSLQEDFAIYYC	QQWDSGLNGV	VFGSETKLIVLIGQ		
3c	SVELTQPPSVSVSPSTASITCSSEKLG	DK----YACWY	QQHPEHAK	LMIVYQSKRP	SGIPDRFSGSGSGTASLTITSLQEDFAIYYC	QQWDSGLNGV	VFGSETKLIVLIGQ		

Matrices de secuencia de V_{L(kappa)}-CDR1 (definición de contacto)

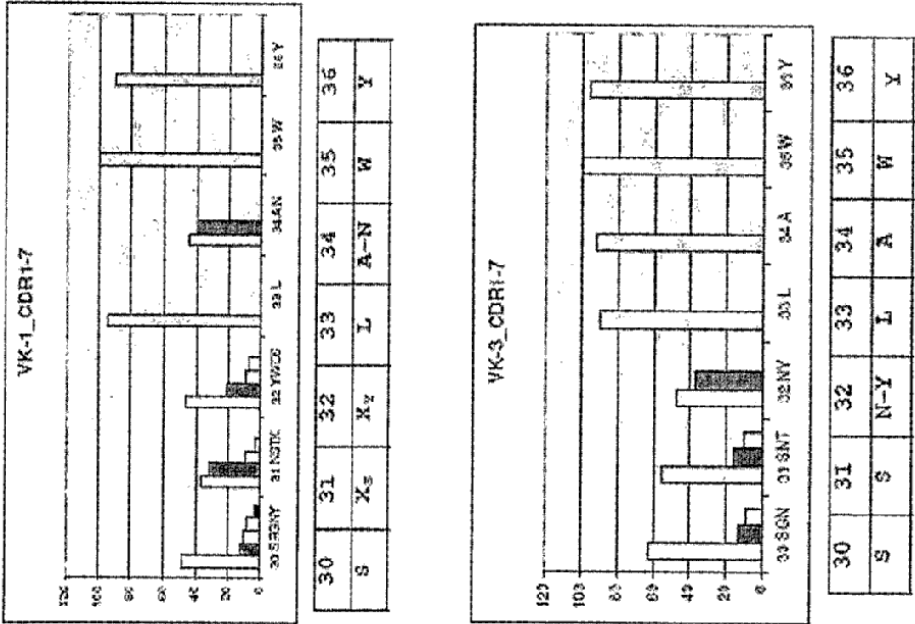


Fig. 17

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30			X																	
31	X	X	X	X	X	X	X	X	X											
32	X	X	X	X	X	X	X	X	X											
33																X				
34	X								X											
35																				
36								X												X

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30			X																	
31			X																	
32							X		X											
33																				
34			X													X				
35																				
36								X												X

Matrices de secuencia de V_{L(kappa)}-CDR1 (continuación)

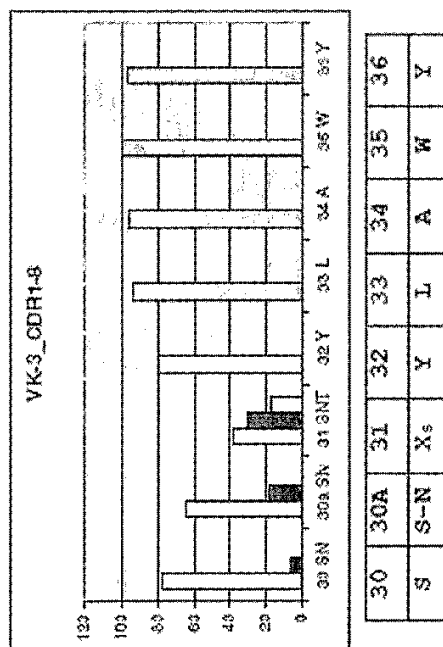
[illegible]

Fig. 17

Fig. 18

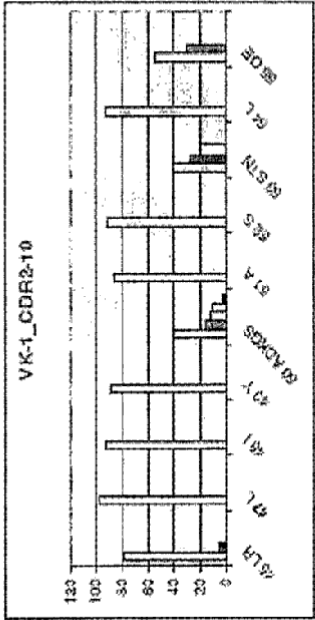
Secuencia VH3-23-VKIII-A27 SEC ID N°: 1

```

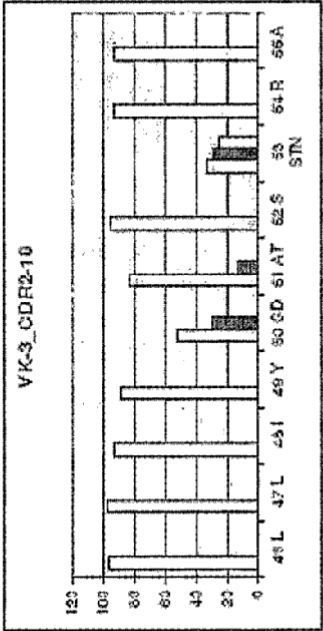
-----
CDR-H1
1  GlnValGlnLeu LeuGluSer GlyGlyGly LeuValGlnPro GlyGlySer LeuArgLeu SerCysAlaAla SerGlyPhe ThrPheSer SerTyzAlaMet.
   GAATGCAGC TGTTGGAGTC TGTGGAGGA TTGGTGCAGC CTGGCGGTAG CTTCGGCTG TCTTGTGCTG CCACTGGCTT TACCTTCTCT AGCTATGCTA
   CTTCAGTGG ACACCTTCAG ACCACTCTCT AACCACTGCG GACCGGCATC GAACGGCAGC AGACACGAGC GGTCAACGAA ATGGAAGAGA TCGATACCAT
   CDR-H1
-----
101  •NSerTyrVal ArgGlnAla ProGlyLysGly LeuGluTyr ValSerAla IleSerGlySer GlyGlySer ThrTyrTyr AlaAspSerVal LysGlyArg.
   TGAGTTGGGT TAGACAGGCT CTTCGCAGG GTTTGGATG GGTGCTGCT ATCTCTGCT CTGGCGGTAG CACCTACTAT GCAGTAGCG TCAAGGCCG
   ACTCAACCCA ATCTGTCCGA GGACCTTTC CAAACCTTAC CCACAGACA TAGAGACCGA GACCGCCATC GTGATGATA CGTCTATCGC AGTTTCGGC
   CDR-H2
-----
201  •PheThrIle SerArgAspAsn SerIysAsn ThrLeuTyr LeuGlnMetAsn SerLeuArg AlaGluAsp ThrAlaValTyr TyrCysAla LysAspGly
   CTTACCATC AGCGGGAFA ACAGTAAAAA CACCTGTATC TTGCAGATCA ACAGCTTCCG CGCCGAGAT ACCCTGTGT ATTACTGTGC TAAGATGGT
   GAAGTGGTAG TCGGCCCTAT TGTCAATTTT GTGGACATG AAGTCTACT TGTGGAGC GGGCTTCTA TGGGACACA TATGACACG ATTCTACCA
   CDR-H3
-----
301  SerGlySerGly TyrAlaPhe AspTyrTyr GlyGlnGlyThr LeuValThr ValSerSer GlyGlyGlyGly SerGlyGly GlyGlySer GlyGlyGlyGly.
   TCTGGTCCG GCTACGCTT CGATTACTGG GGTCAAGGCA CACTGGTATC CGTCTATAC GGTGGAGCG GTTCTGTGG AGCGGTTCG GGTGGCGAG
   AGACCAAGC CGATCGGAA GCTAATGACC CCACTCCGT GTGACCATG GCAGAGATG CCACCTCCG CAGACCAAC TCGGCCAAGC CCACCGCTC
   CDR-L1
-----
401  •GserGluIle ValLeuThr GlnSerProGly ThrLeuSer LeuSerPro GlyGluArgAla ThrLeuSer CysArgAla SerGlnSerVal SerSerSer.
   GTTCAGAAAT CGTCTGACA CAGTCTCCAG GCACCTTATC TCTGTCCCA GGCAGACCG CTACACTGTC CTGCAGAGCT TCTCAGTCCG TGTCTAGTTC
   CAAGTCTTTA GCAGACTGT GTCAGAGGTC CGTGGACAG AGACAGGCT CCGTTCGCG GATGTACAG GACCTCTCGA AGATCAGGC ACAGATCAAG
   CDR-L2
-----
501  •TyrLeuAla TrpTyrGlnGln LysProGly GlnAlaPro ArgLeuLeuIle TyrGlyAla SerSerArg AlaThrGlyIle ProAspArg PheSerGly
   CTATCTGCC TGTATTCAC AGAACCTTCG TCAGGCCCTT CGCTTCTCA TCTAGGTGC TTCTAGCAGA GCGACGCA TCCCTGTATAG ATTCTCTGGT
   GATGACCGG ACCATAGTTG TCTTTGGACC AGTCGGGGA GCGACGACT AGATGCCAGC AGATGGTCT CGGTGTCTGT AGGACTATC TAGAGACCA
   CDR-L3
-----
601  SerGlySerGly ThrAspPhe ThrLeuThr IleSerArgLeu GluProGly AspPheAla ValTyrTyrCys GlnGlnTyr AsnSerThr ProLeuThrPhe.
   AGCGGCTTC GCACAGATT CACACTACT ATCTCCGCTT TGGACACAGA AGATTCCGC GTTACTATT GCGACAGTA CACAGCACC CCATTGACAT
   TCGCCGACAC CGTCTTAAA GTGTACTCA TAGAGGCAA AACTTGTCT TCTAAGCGG CAATATGTA CCGTGTCTAT GTGTCTGCG GGTAACTCTA
   •PCLyGlnGly ThrLeuVal GluIleLysArg Thr
701  TCGGTACAG CACCAAGTG GAATCAAA CAAC
   ACCACTCC GTGGTTTAC CTTAGTTT CTGG

```

Matrices de secuencia de $V_{L(kappa)}$ -CDR2 (definición de contacto)



46	47	48	49	50	51	52	53	54	55
L	L	I	Y	X4	A	S	X2	L	Q-E



46	47	48	49	50	51	52	53	54	55
L	L	I	Y	G-D	A	S	X5	R	A

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
46																X				
47																X				
48				X																
49							X													
50			X	X	X	X	X	X	X											
51			X																	
52																				
53			X	X	X	X	X	X	X											
54																	X			
55													X							

Fig. 19

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
46																X				
47																X				
48				X																
49							X													
50			X												X					
51			X																	
52				X	X	X	X	X	X											
53			X	X	X	X	X	X	X											
54							X													
55			X																	

Distribución de longitud de CDR3 de V-kappa

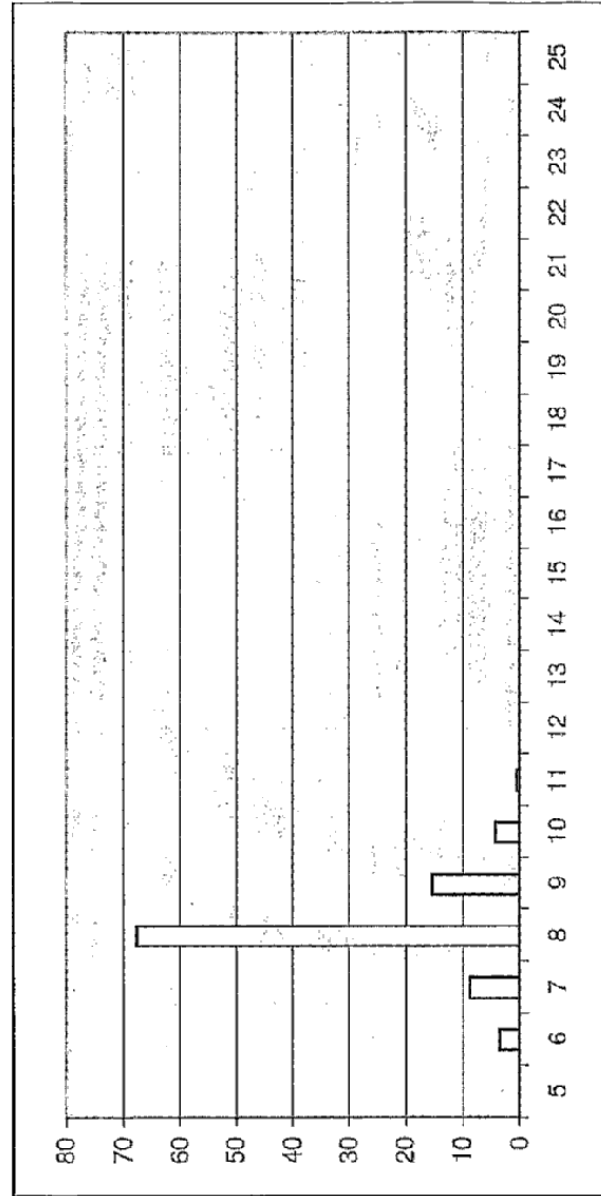
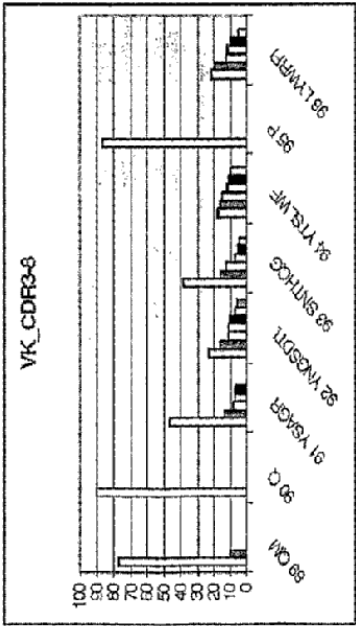


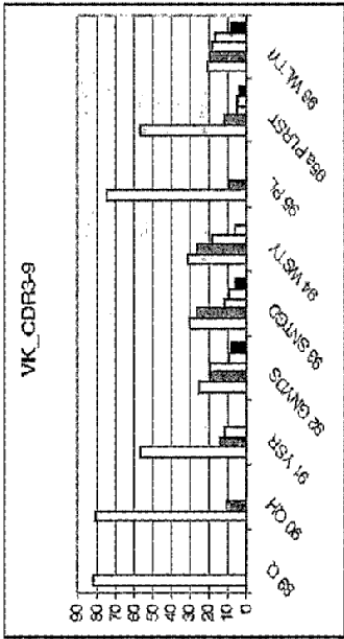
Fig. 20

- La variabilidad de longitud de CDR3 de V_k oscila de 6 a 11 restos (definición de contacto)
- Sin embargo, la mayoría (~80 %) de las CDR3 de V_k de anticuerpos antiproteína oscilan de 8 a 9 restos
- Las longitudes 8 y 9 se sintetizan por separado y se almacenan antes del ensamblaje genético en las proporciones deseadas

Matrices de secuencia de V_L(kappa)-CDR3 (definición de contacto)



89	Q	90	Q	91	Y	92	K	93	S	94	Y	95	P	96	L
----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---



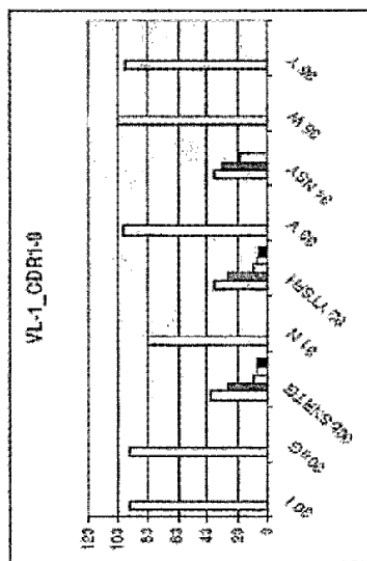
89	Q	90	Q	91	Y	92	K	93	S	94	Y	95	P	96	L
----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---

89	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
90																	X			
91							X										X			
92	X	X	X	X	X	X		X	X											
93	X	X	X	X	X	X	X	X	X									X		
94	X	X	X	X	X	X	X	X	X											
95																				
96	X	X	X	X	X	X	X	X	X							X				

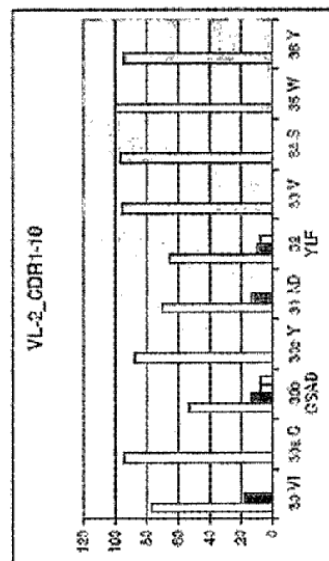
Fig. 21

89	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
90																	X			
91							X										X			
92	X	X	X	X	X	X	X	X	X											
93	X	X	X	X	X	X	X	X	X											
94	X	X	X	X	X	X	X	X	X											
95																				
96	X	X	X	X	X	X	X	X	X							X				

Matrices de secuencia de $V_{L(\text{Lambda})}\text{-CDR1}$ (definición de contacto)



30	30A	30B	31	32	33	34	35	36
I	G	S	N	Y	Y	Y	Y	Y



30	30A	30B	31	32	33	34	35	36
V	G	S	N	Y	Y	Y	Y	Y

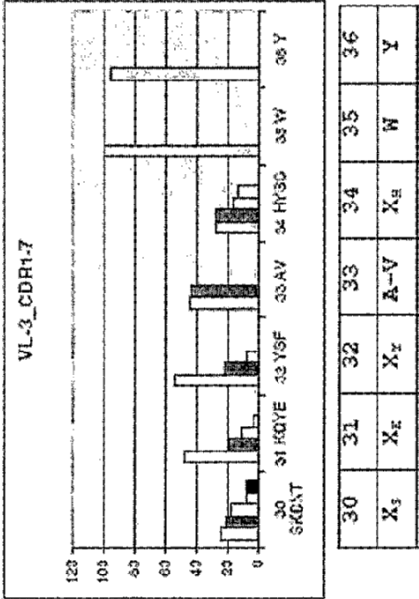
	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30							X													
30A															X					
30B	X	X	X	X	X	X	X	X	X											
31																				
32	X	X	X	X	X	X	X	X	X										X	
33																				
34	X	X	X	X	X	X	X	X	X											
35																				
36								X												X

Fig. 22

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30																				
30A															X				X	
30B	X	X	X	X	X	X	X	X	X						X	X				
30C																				
31																				
32								X										X		
33																				
34																				
35																				
36								X												X

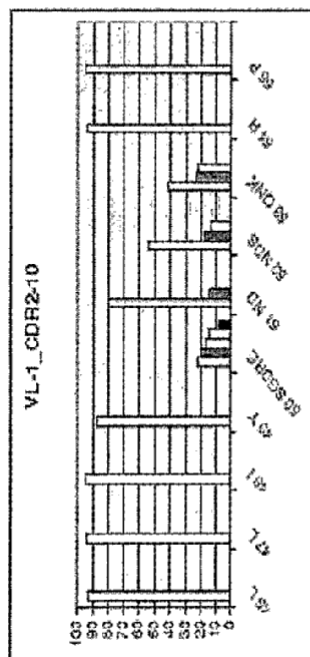
Matrices de secuencia de V_L(Lambda)-CDR1 (continuación)

Fig. 22

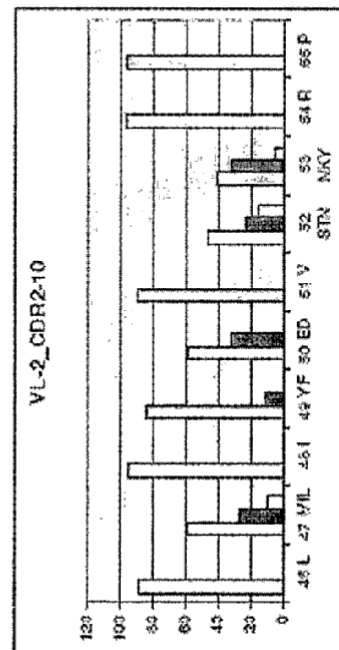


	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
30	X	X	X	X	X	X	X	X	X											
31	X	X	X	X	X	X	X	X	X											
32	X	X	X	X	X	X	X	X	X		X									
33	X	X	X	X	X	X	X	X	X									X		
34	X	X	X	X	X	X	X	X	X											
35							X													X
36																				

Matrices de secuencia de $V_{L(\text{Lambda})}$ -CDR2 (definición de contacto)



46	47	48	49	50	51	52	53	54	55
L	L	I	Y	X _S	N	N-D	X _H	R	P



46	47	48	49	50	51	52	53	54	55
L	M-I	I	Y	E-D	V	X _S	X _H	R	P

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
46																X				
47																X				
48																				
49																				
50	X	X	X	X	X	X	X	X	X											
51																				
52	X	X	X	X	X	X	X	X	X											
53	X	X	X	X	X	X	X	X	X											
54																				
55																				

Fig. 23

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
46																X				
47										X										
48																				
49																				
50																				
51																				
52	X	X	X	X	X	X	X	X	X											
53	X	X	X	X	X	X	X	X	X											
54																				
55																				

Matrices de secuencia de VL(Lambda)-CDR2 (continuación)

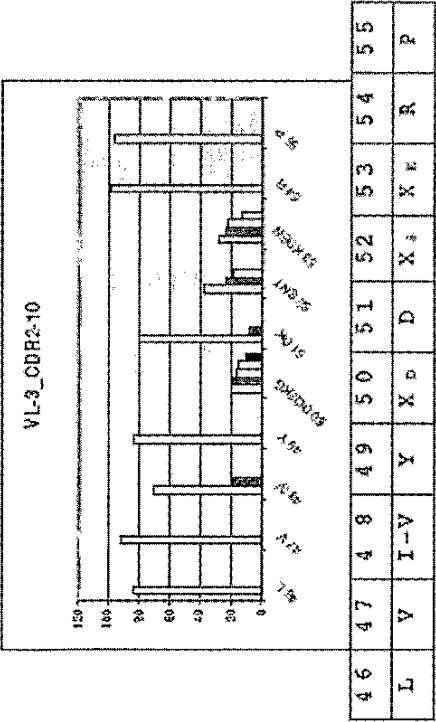


Fig. 23

	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
46																X				
47						X													X	
48																			X	
49							X													
50	X	X	X	X	X	X	X	X	X										V	
51		X																		
52	X	X	X	X	X	X	X	X	X											
53	X	X	X	X	X	X	X	X	X		X									
54																				
55																				

Distribución de longitud de CDR3 de V-lambda

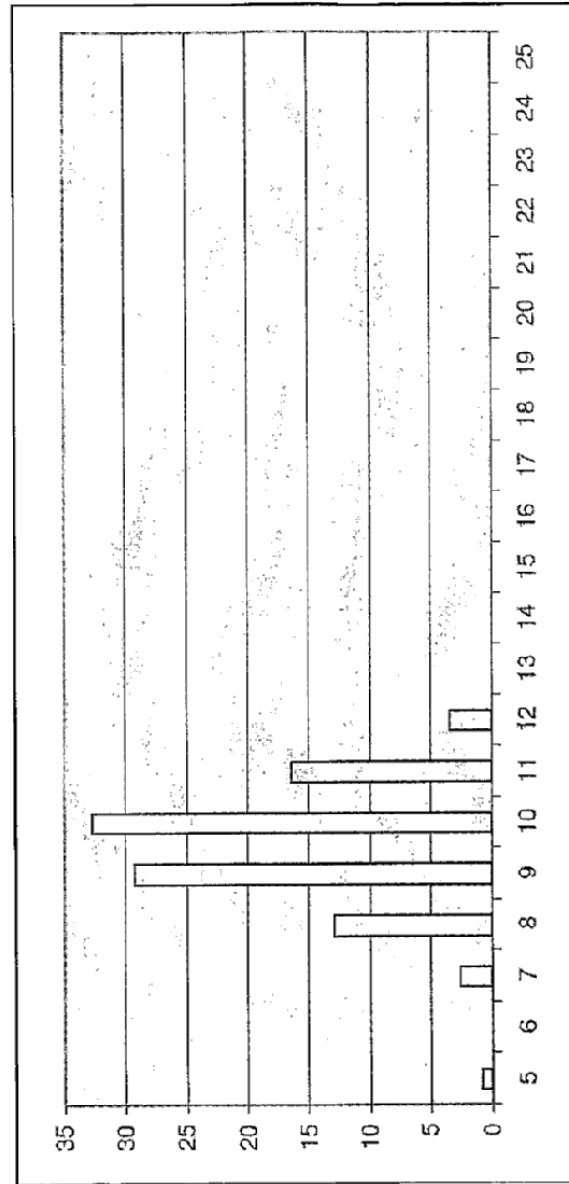
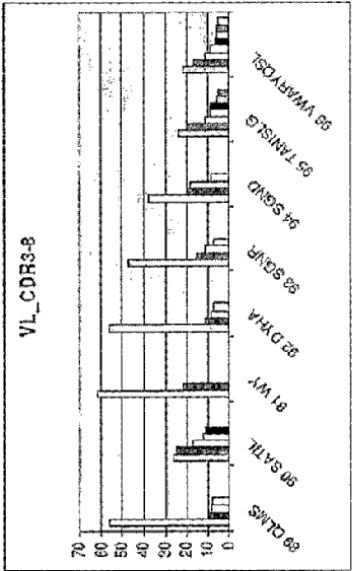


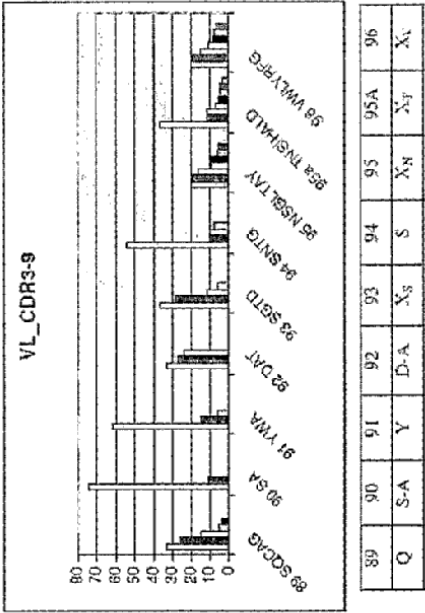
Fig. 24

- La variabilidad de longitud de CDR3 de $V\lambda$ oscila de 5 a 12 restos (definición de contacto)
- Sin embargo, la mayoría (~90 %) de las CDR3 de $V\lambda$ de los anticuerpos anti proteína oscilan de 8 a 11 restos
- Las longitudes de 8, 9, 10 y 11 se sintetizan por separado y se almacenan antes del ensamblaje genético en las proporciones deseadas

Matrices de secuencia de VL(Lambda)-CDR3 (definición de contacto)



89	90	91	92	93	94	95	96
Q	S-A	W	D	X _S	S	X _S	X _N



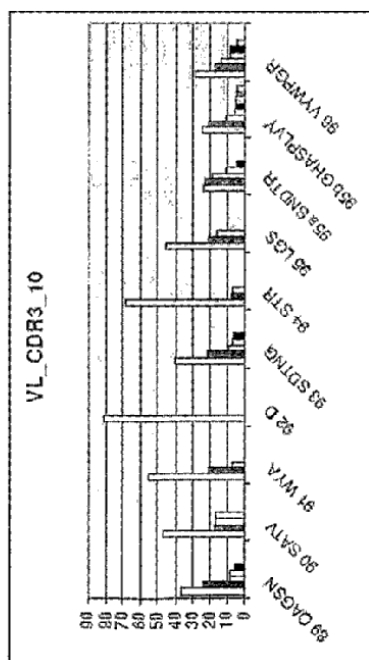
89	90	91	92	93	94	95	96
Q	S-A	Y	D-A	X _S	S	X _N	X _F

89	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
90	X																X			
91																				X
92	X																			
93	X	X	X	X	X	X	X	X	X											
94	X	X	X	X	X	X	X	X	X											
95	X	X	X	X	X	X	X	X	X											
96	X	X	X	X	X	X	X	X	X											X

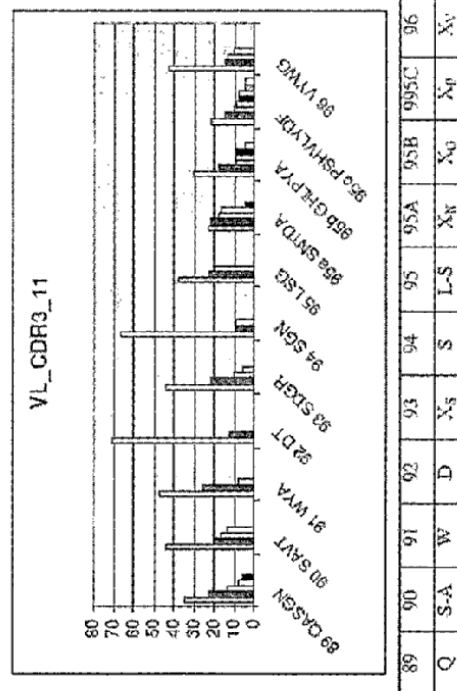
Fig. 25

89	A	D	S	I	P	R	Y	H	N	M	K	C	E	F	G	L	Q	T	V	W
90	X						X										X			
91																				
92	X	X	X	X	X	X	X	X	X											
93	X	X	X	X	X	X	X	X	X											
94	X	X	X	X	X	X	X	X	X											
95	X	X	X	X	X	X	X	X	X									X		
95a	X	X	X	X	X	X	X	X	X											
96	X	X	X	X	X	X	X	X	X										X	X

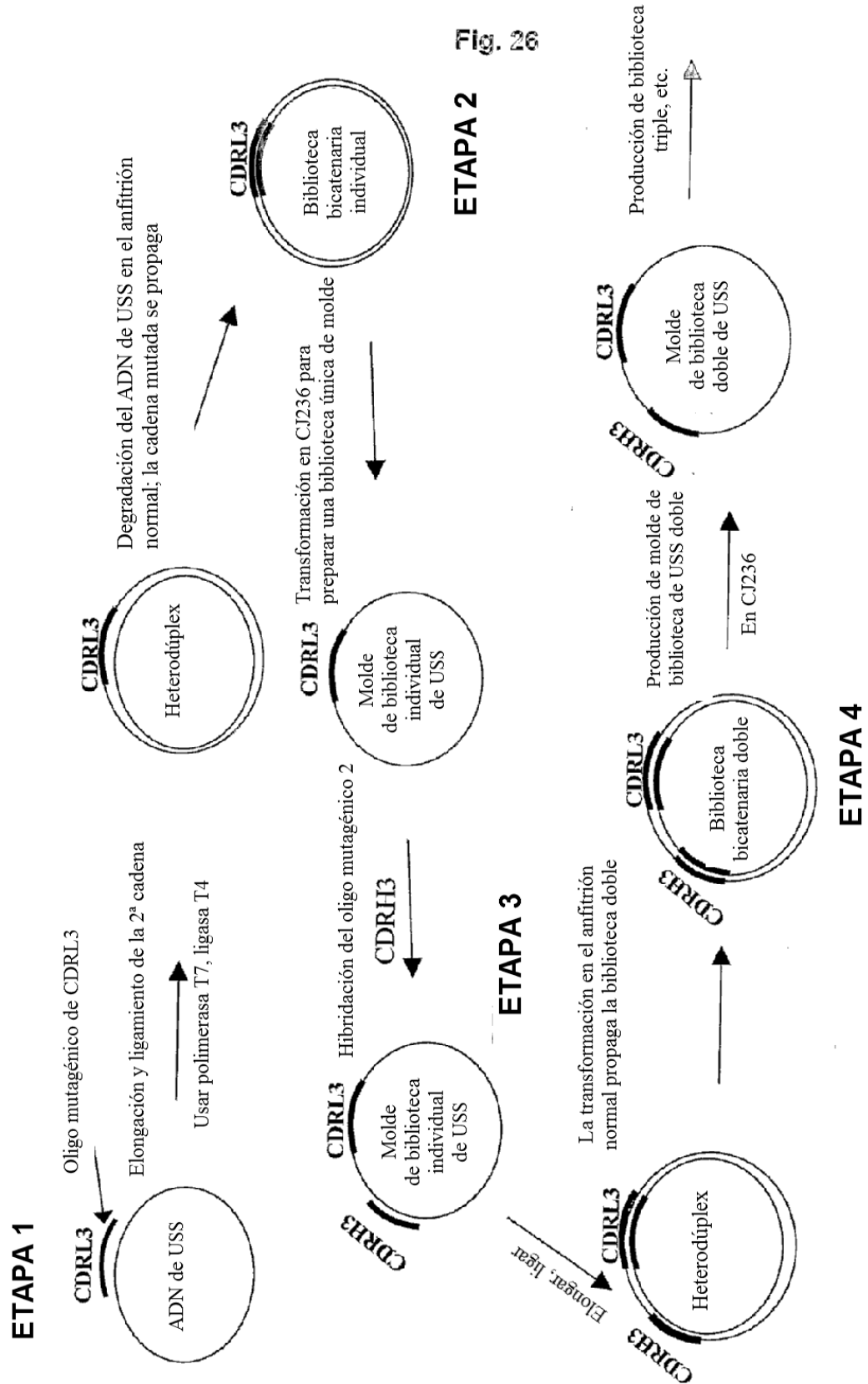
Matrices de secuencia de V_L(Lambda)-CDR3 (continuación)



Q	S-A	W	D	X _S	S	L-S	X _S	X _G	X _V
89									
90									
91									
92									
93									
94									
95									
96									

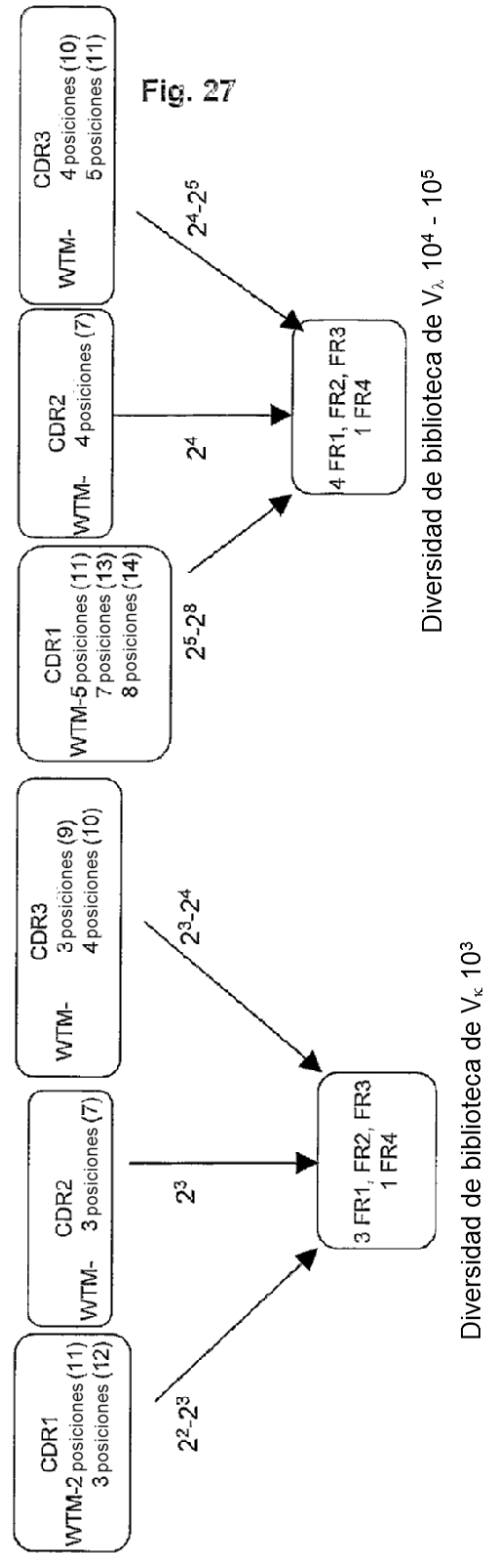


Q	S-A	W	D	X _S	S	L-S	X _S	X _G	X _V
89									
90									
91									
92									
93									
94									
95									
96									



Diversidad de la biblioteca de V_L

Diversidad de CDR mediante WTM (bajo dopaje => 1 o 2 sustituciones por molécula) con todos los aminoácidos excepto Cys.



- Combinar la biblioteca de V_L con la biblioteca de V_H (~10¹¹)
- Todos los oligos contienen 20-meros no degenerados en los extremos 5' y 3'

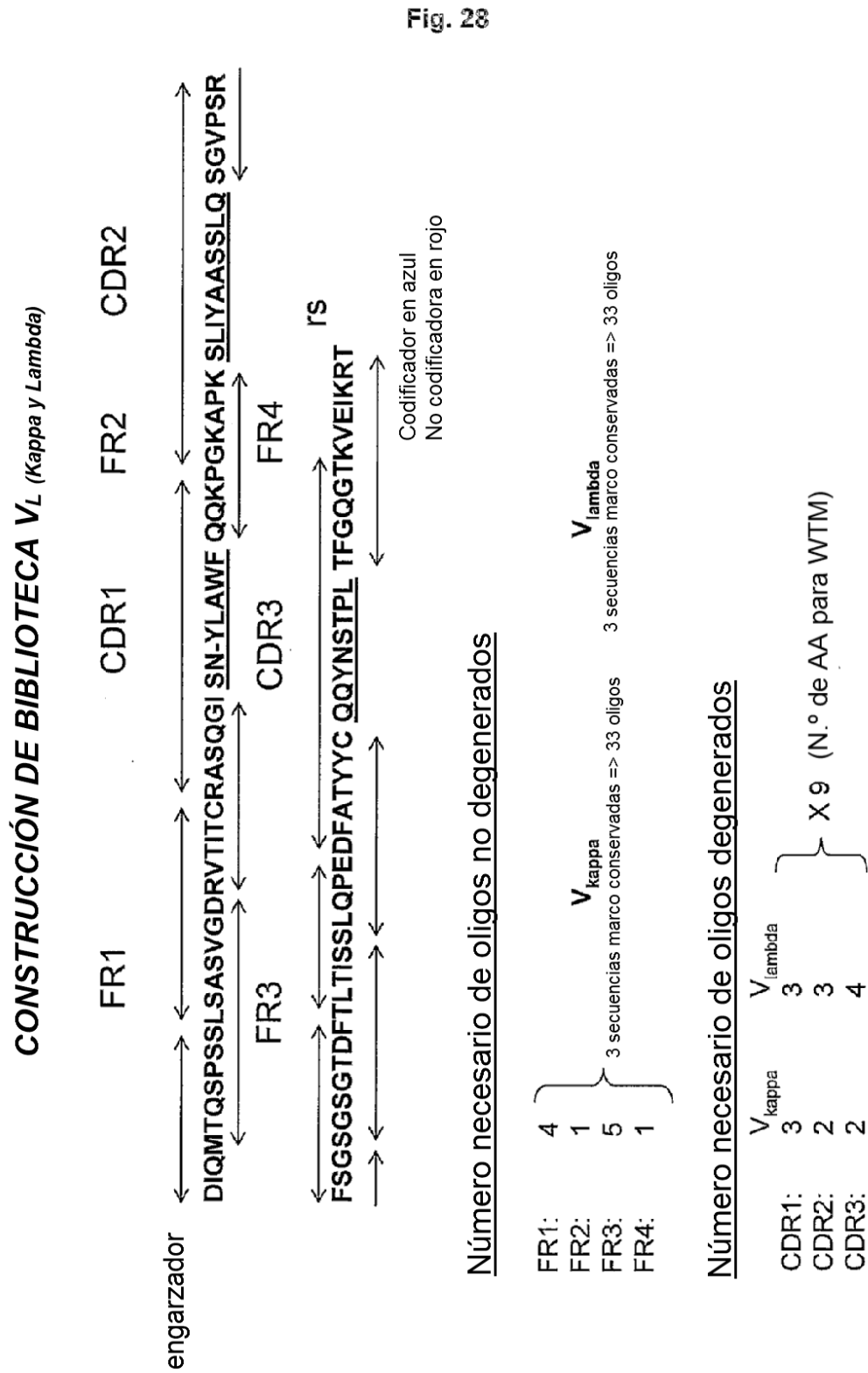


Fig. 28

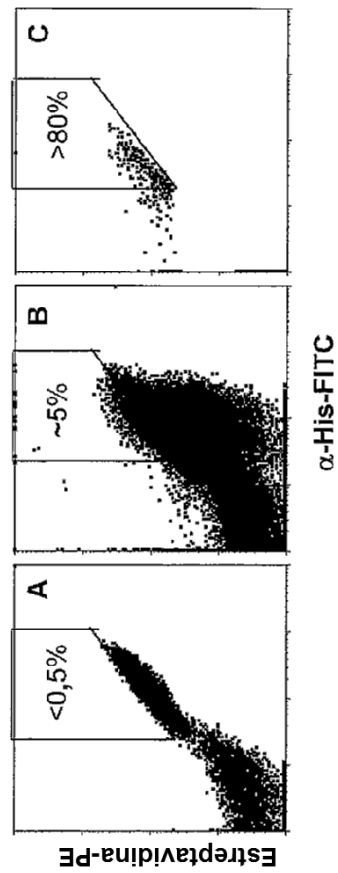
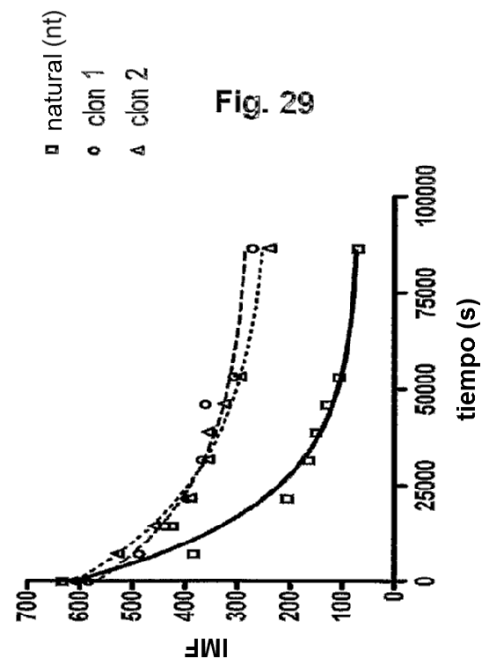
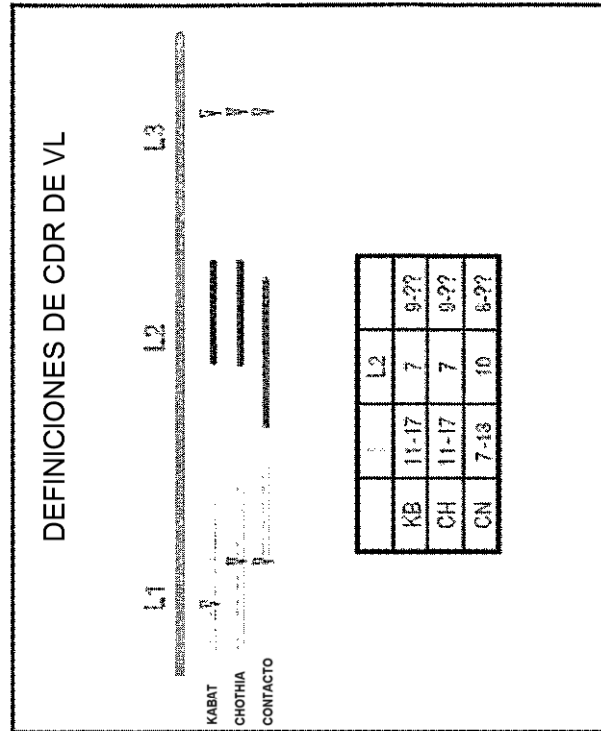
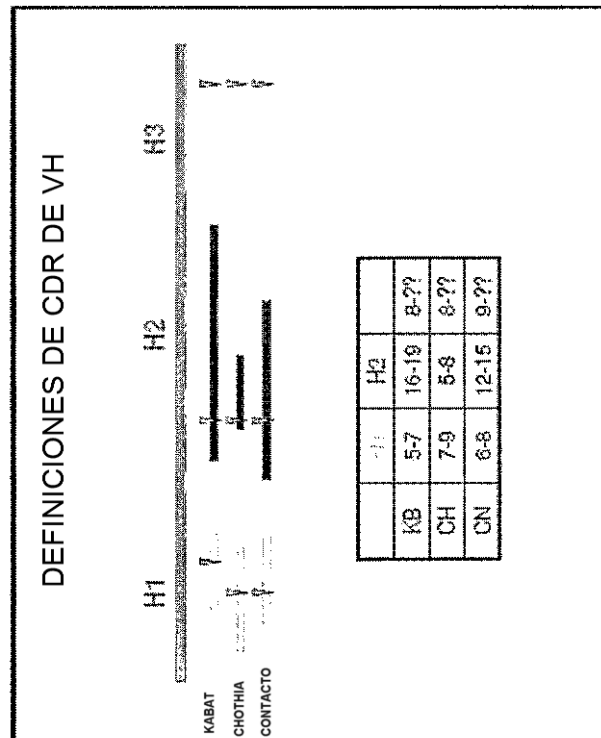
[illegible]

Fig. 30



(B)



(A)

Fig. 31

[illegible]

Fig. 32

```

VB_VH_FR123_CN.FASTA

>1-02
QVQLVQSGAEVKKPGASVKVSCKASGYTFxWVRQAPGQGLExYAQKFQGRVTMTTRDTSISTAYMELSLRSDDDTAVYYC
>1-03
QVQLVQSGAEVKKPGASVKVSCKASGYTFxWVRQAPGQGLExYSQKFQGRVTITRDTSASTAYMELSSLRSEDTAVYYC
>1-08
QVQLVQSGAEVKKPGASVKVSCKASGYTFxWVRQATGQGLExYAQKFQGRVTMTNRNTSISTAYMELSSLRSEDTAVYYC
>1-18
QVQLVQSGAEVKKPGASVKVSCKASGYTFxWVRQAPGQGLExYAQKLQGRVTMTTDTSTSTAYMELSLRSDDDTAVYYC
>1-24
QVQLVQSGAEVKKPGASVKVSCKVSGYTLxWVRQAPGKGLExYAQKFQGRVTMTEDTSTDYAMELSSLRSEDTAVYYC

```

Fig. 33

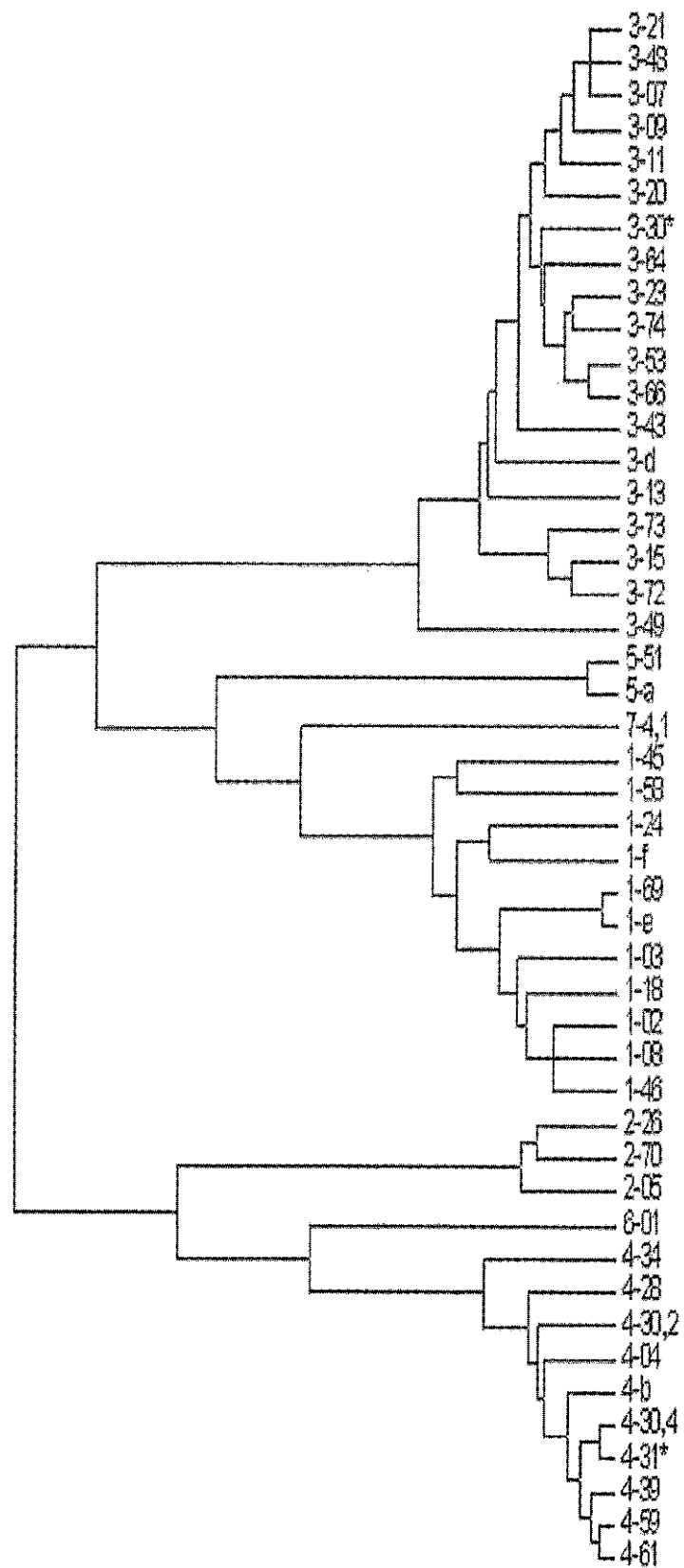


Fig. 34

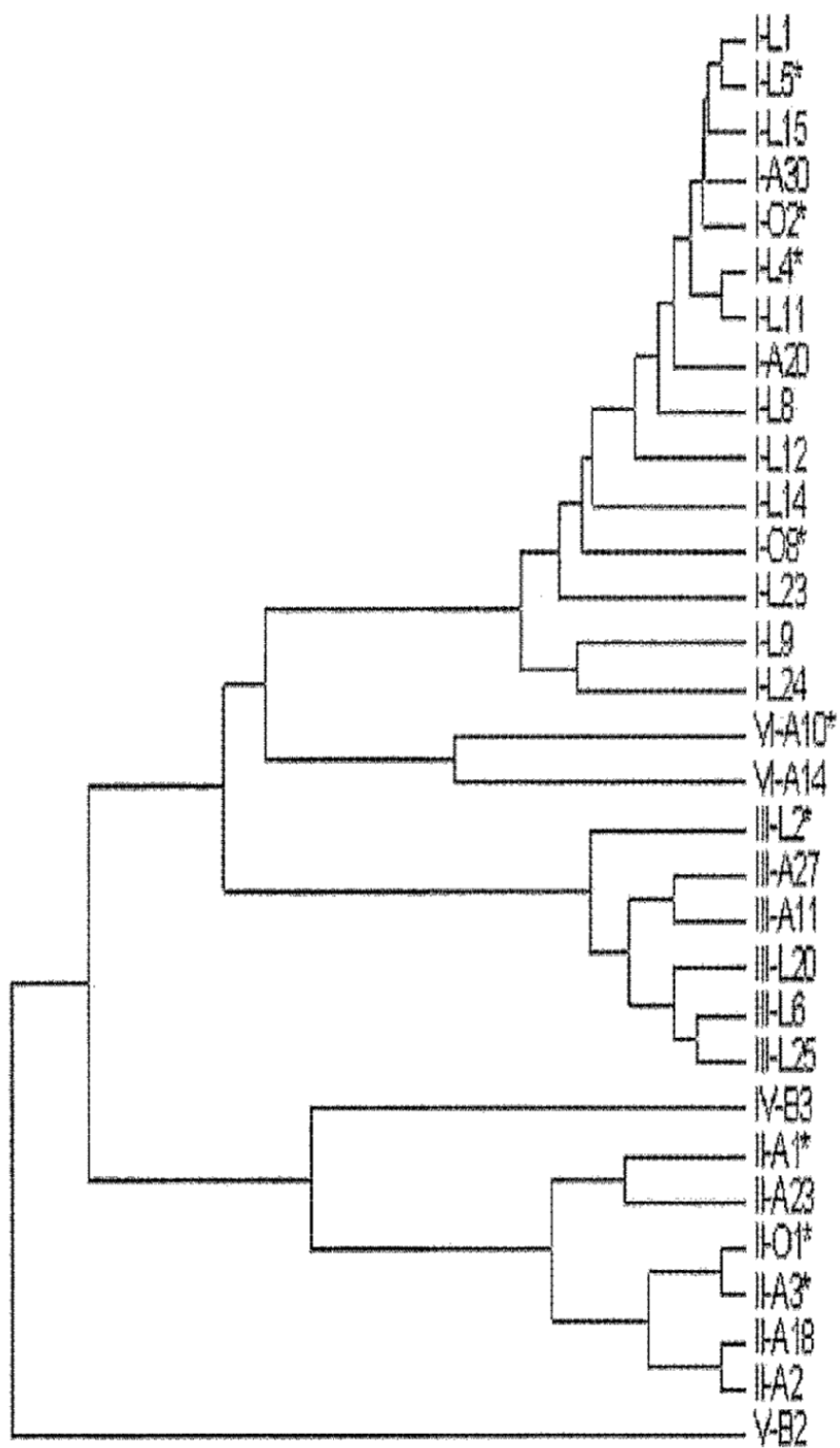


Fig. 35

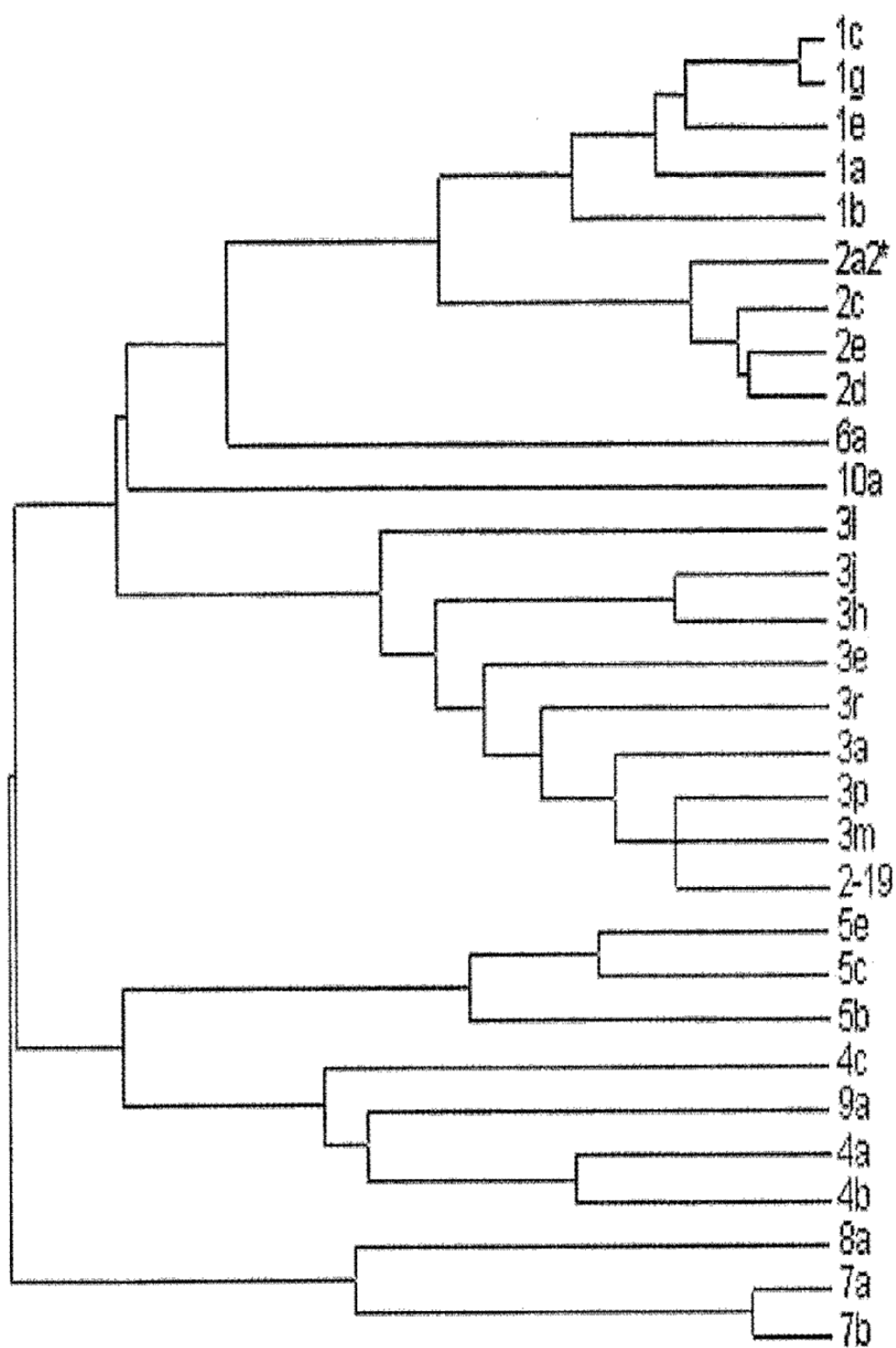


Fig. 36



Fig. 37

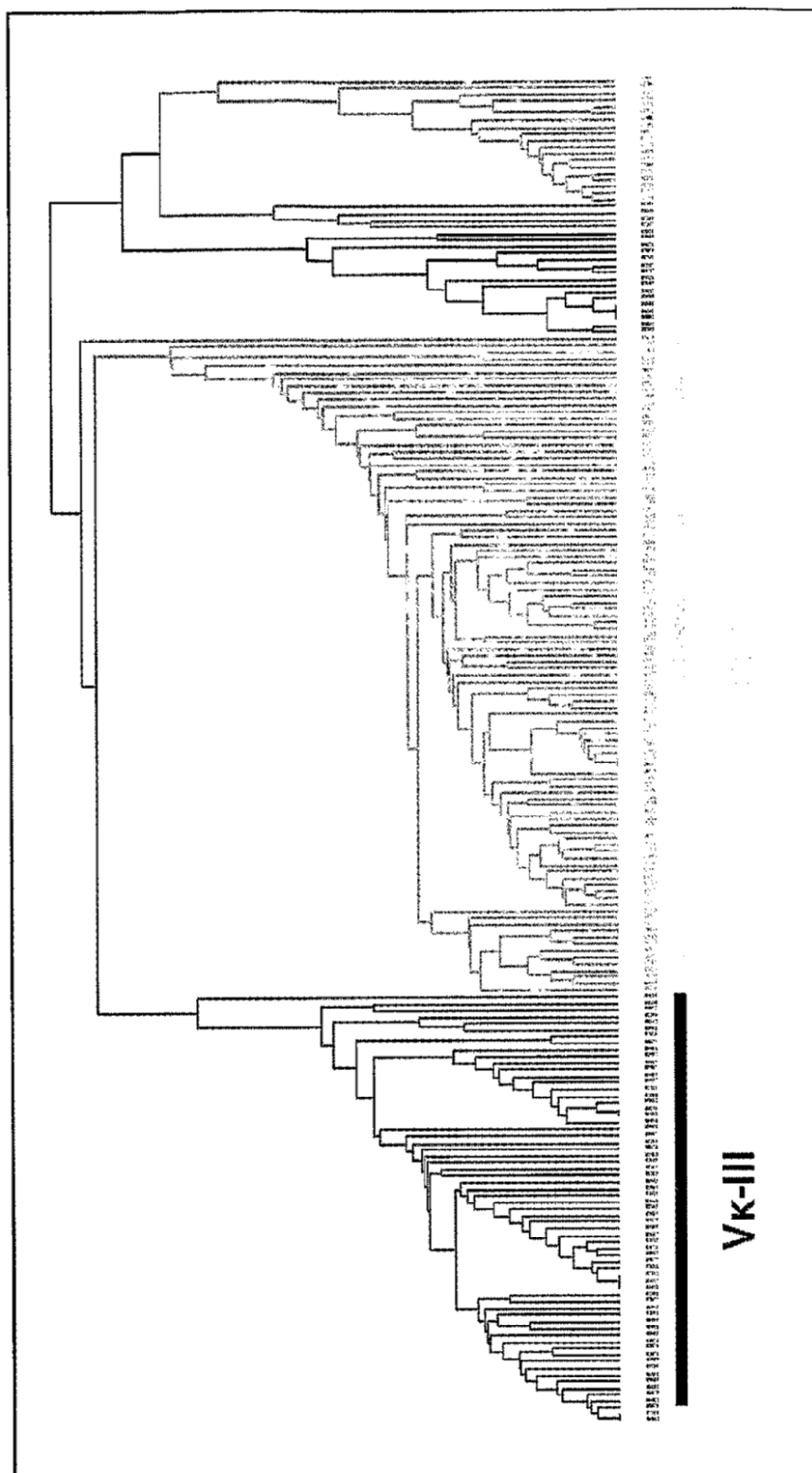


Fig. 38

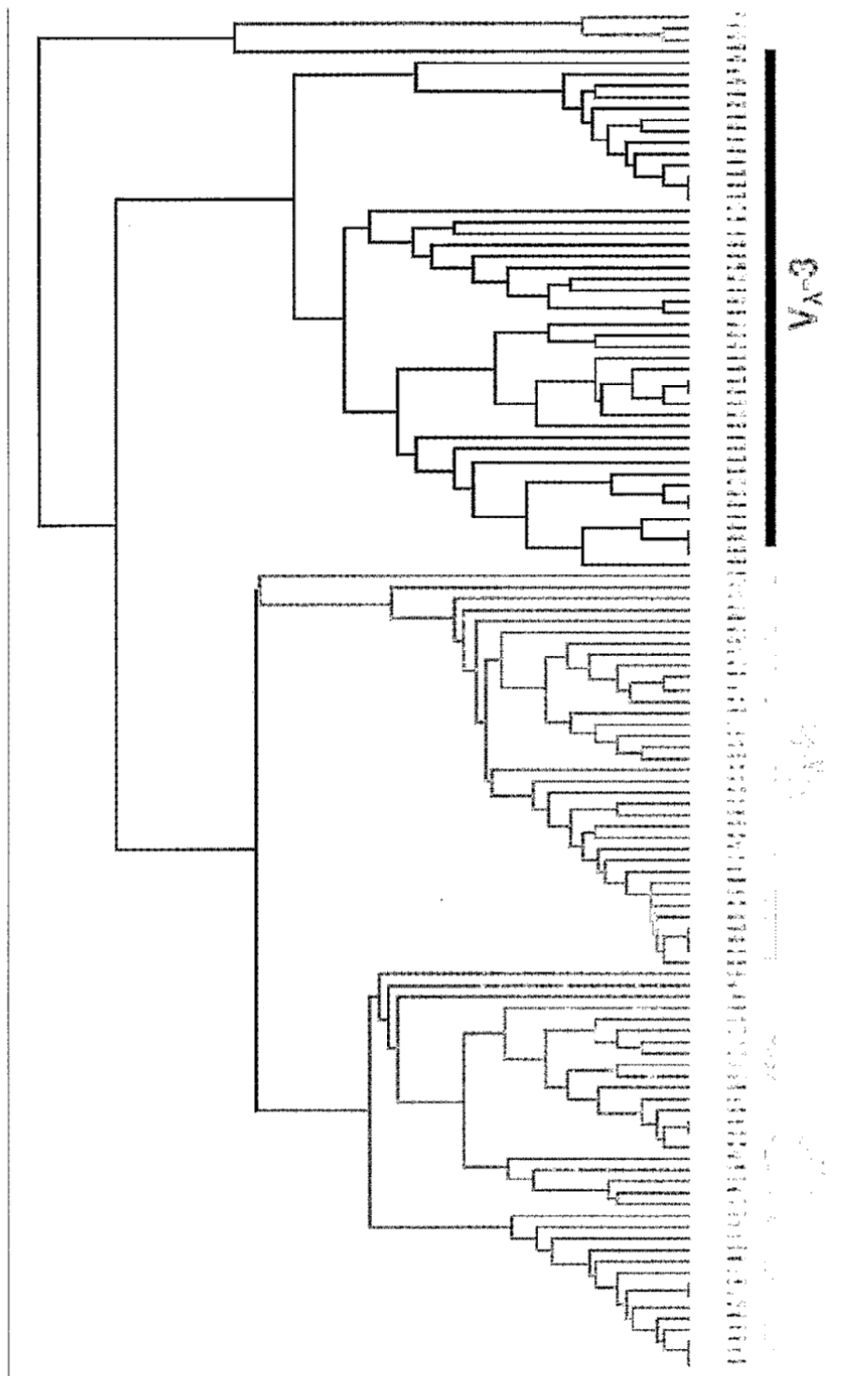
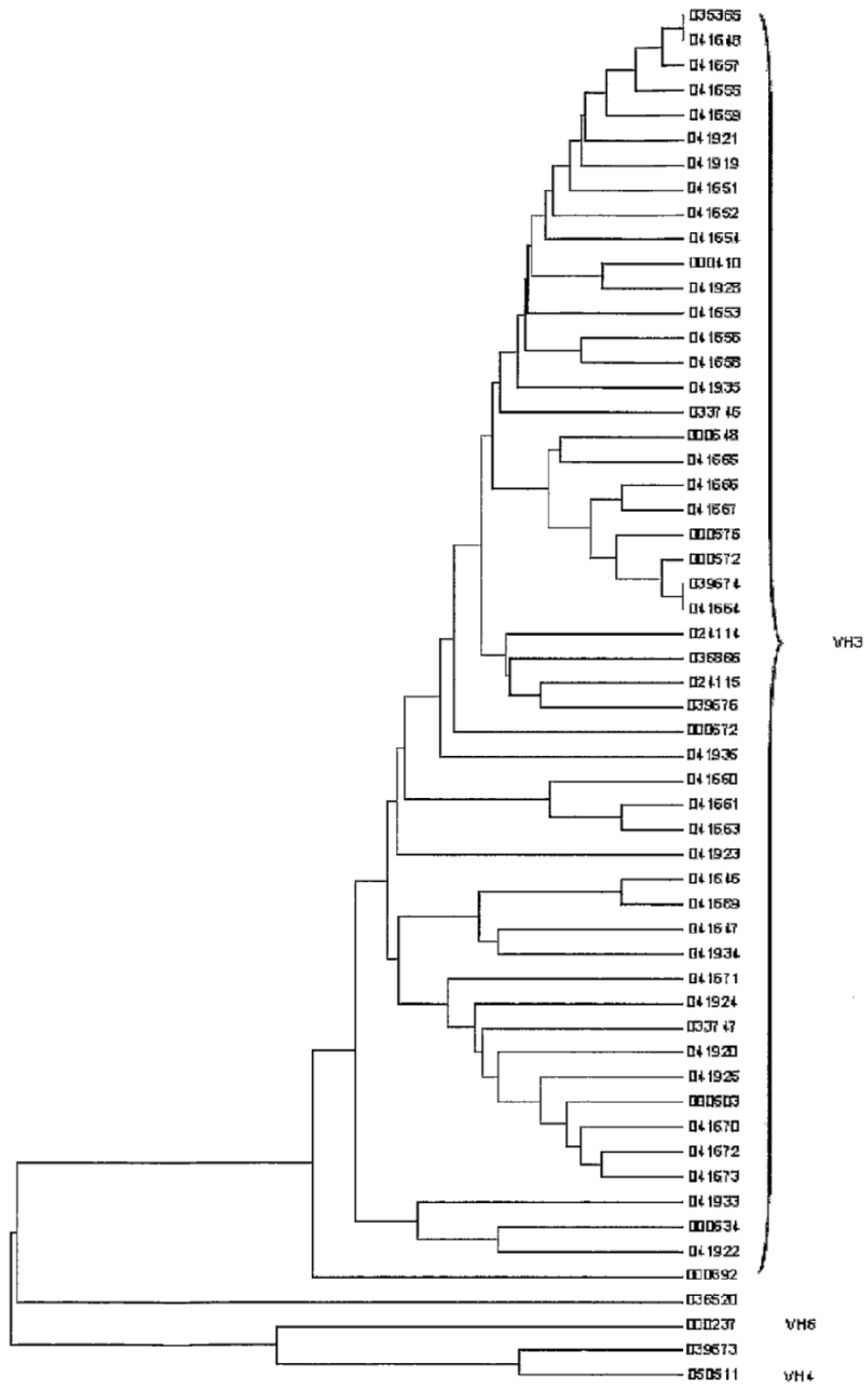


Fig. 39

[illegible]

Fig. 40



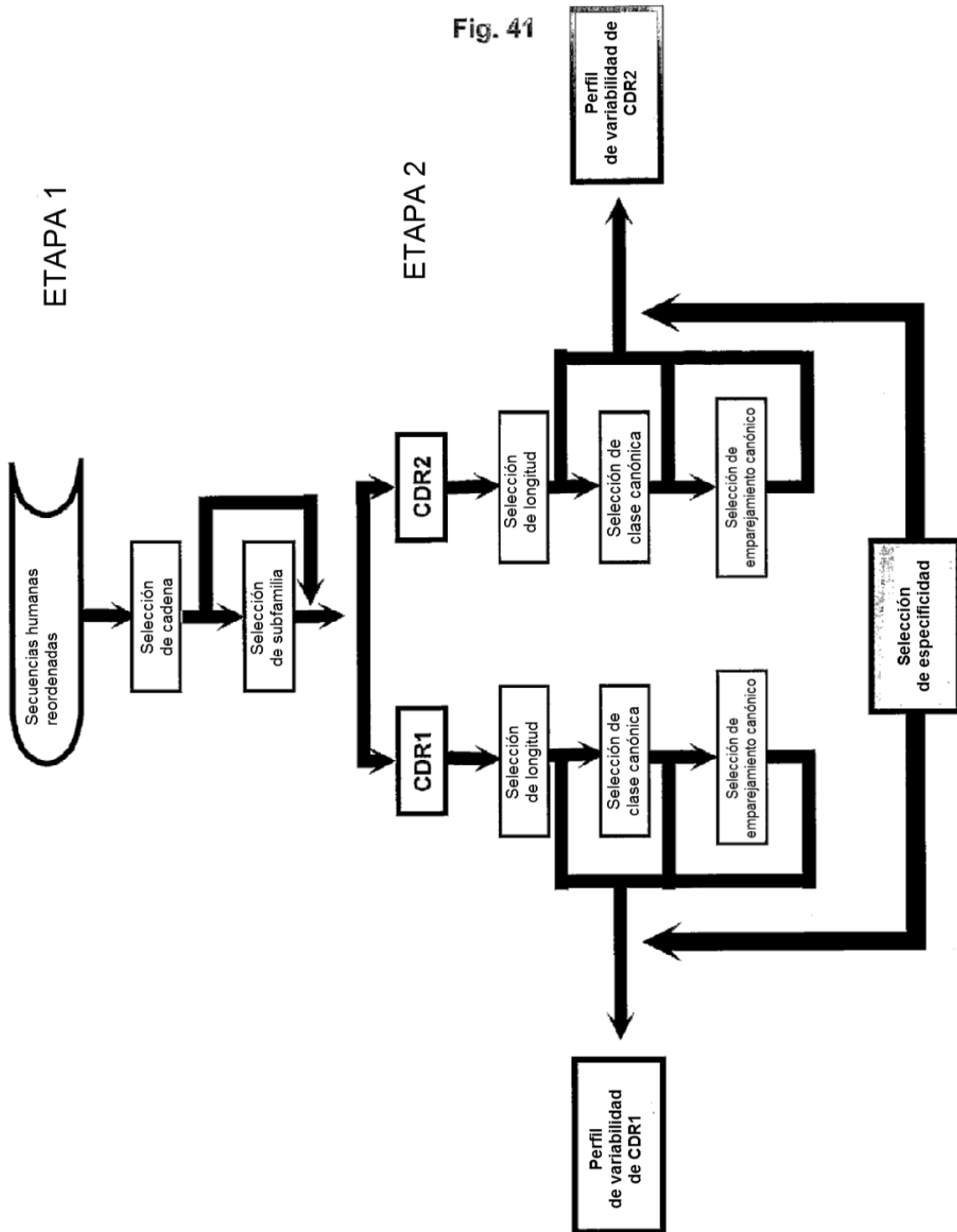


Fig. 42

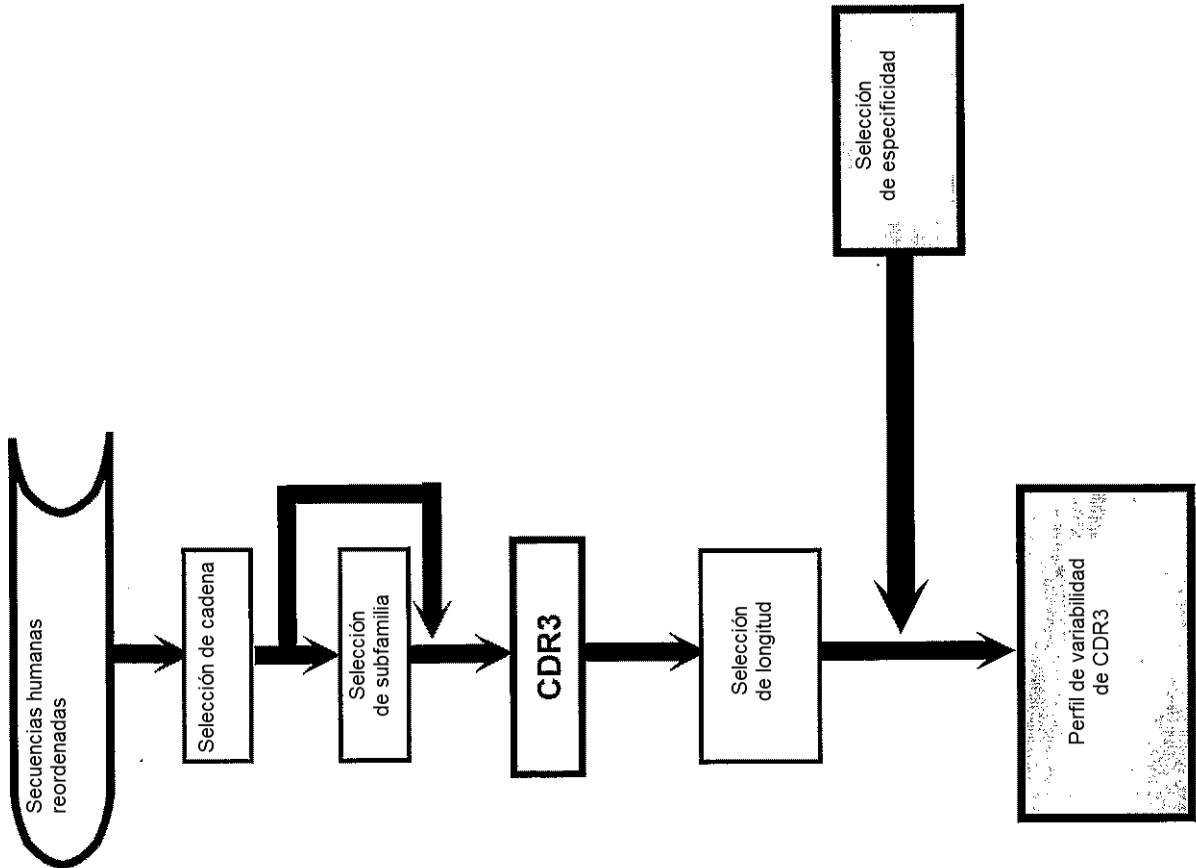


Fig. 43

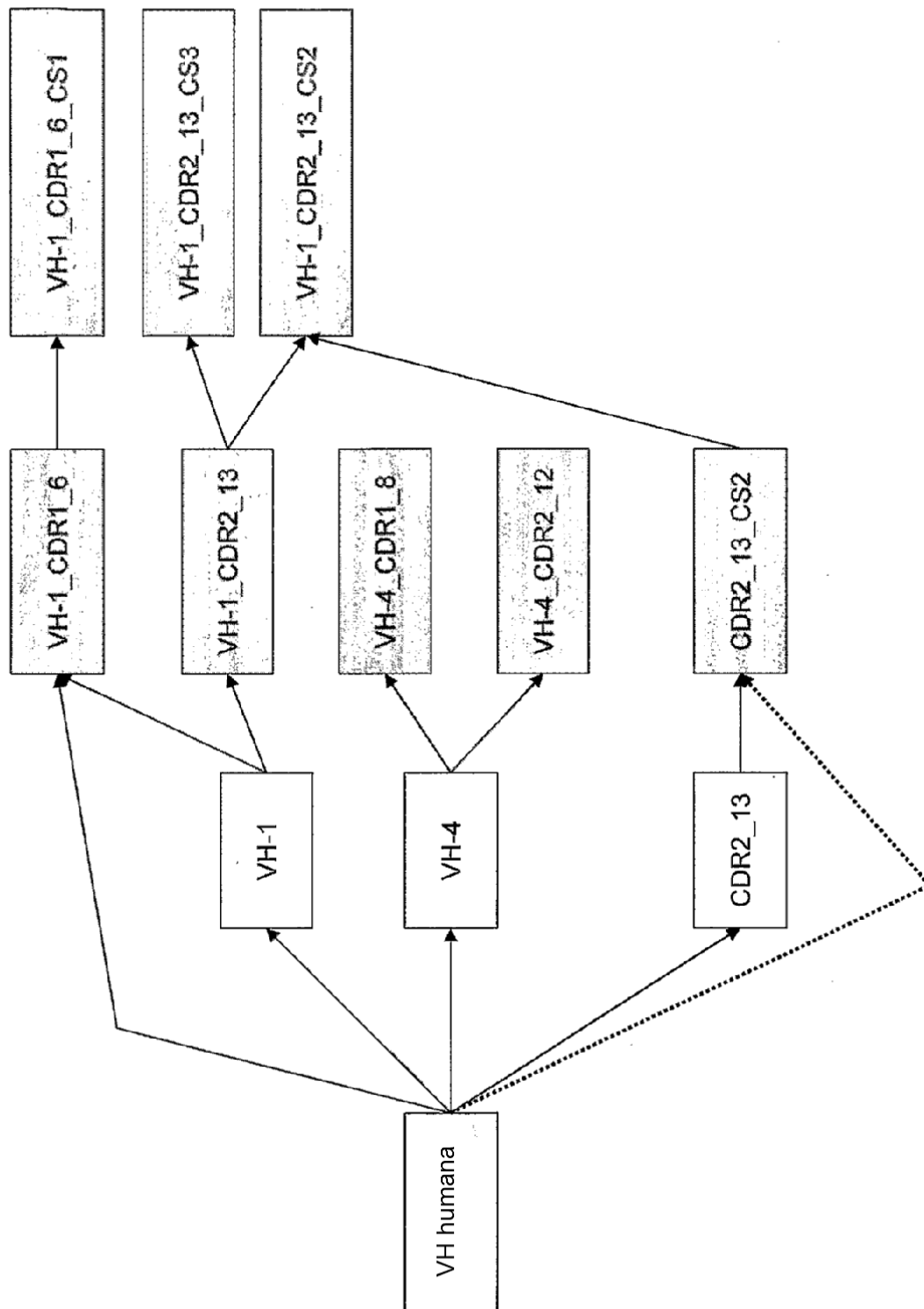


Fig. 44

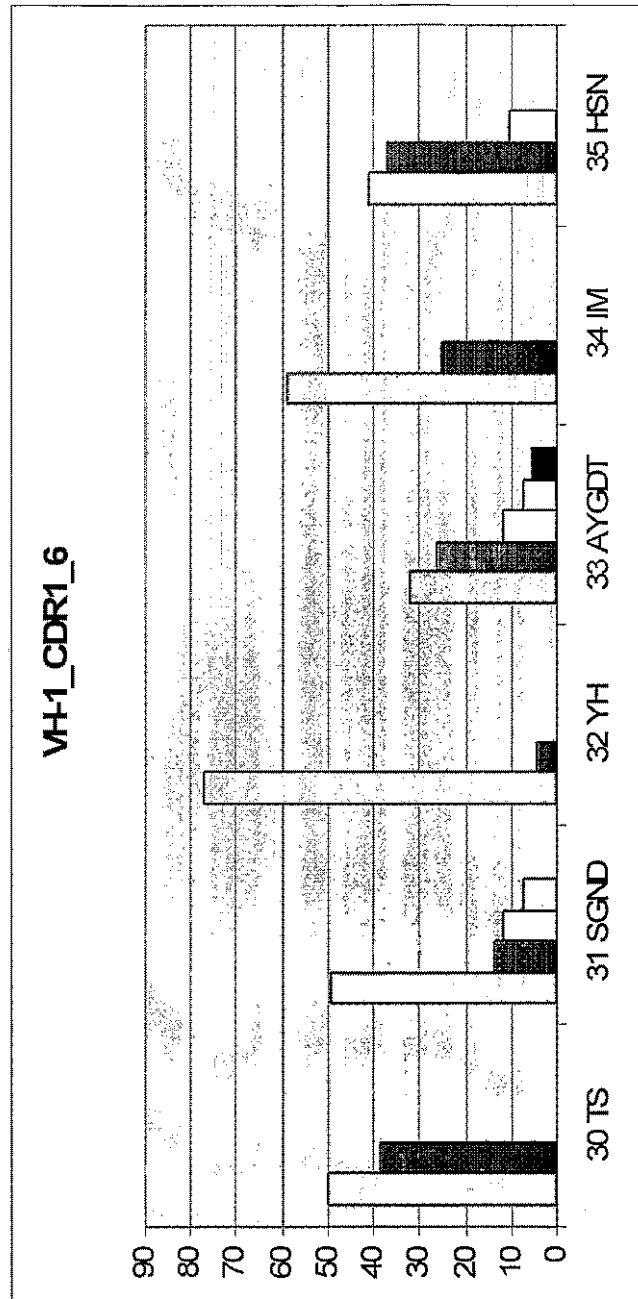


Fig. 45

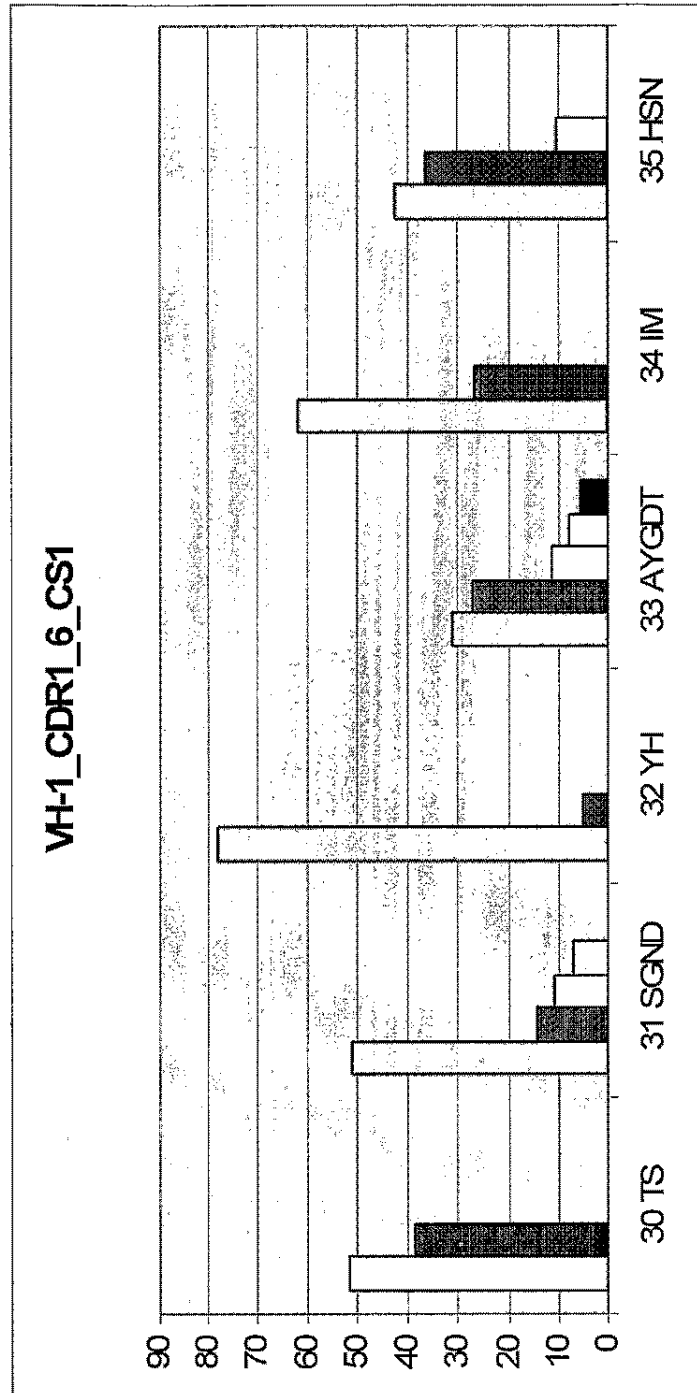


Fig. 46

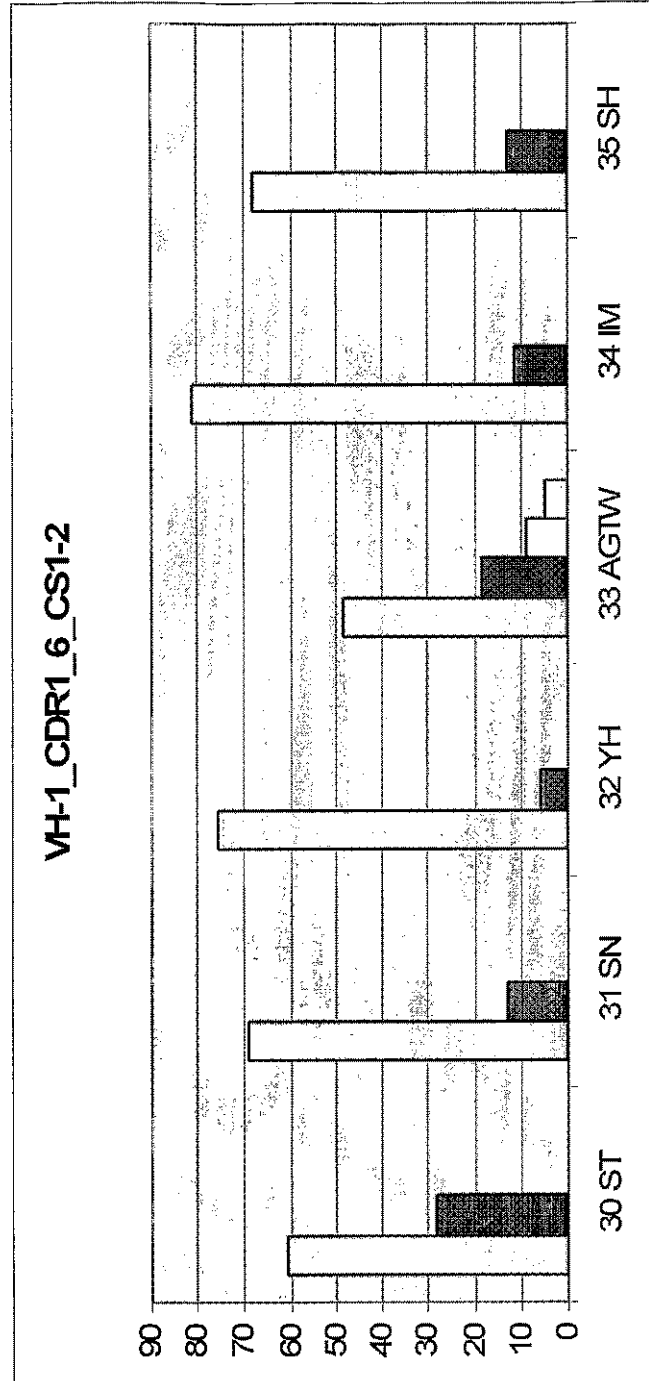


Fig. 47

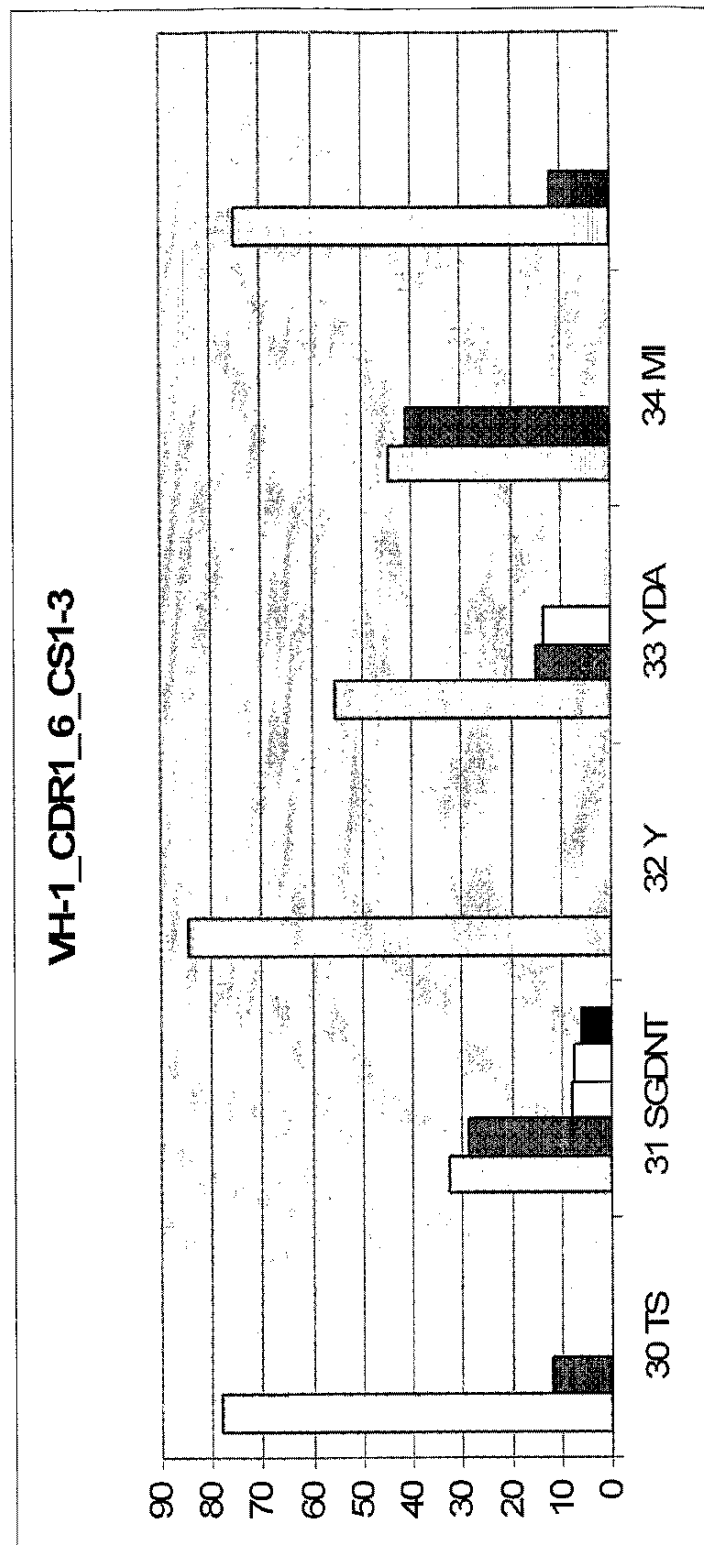


Fig. 48

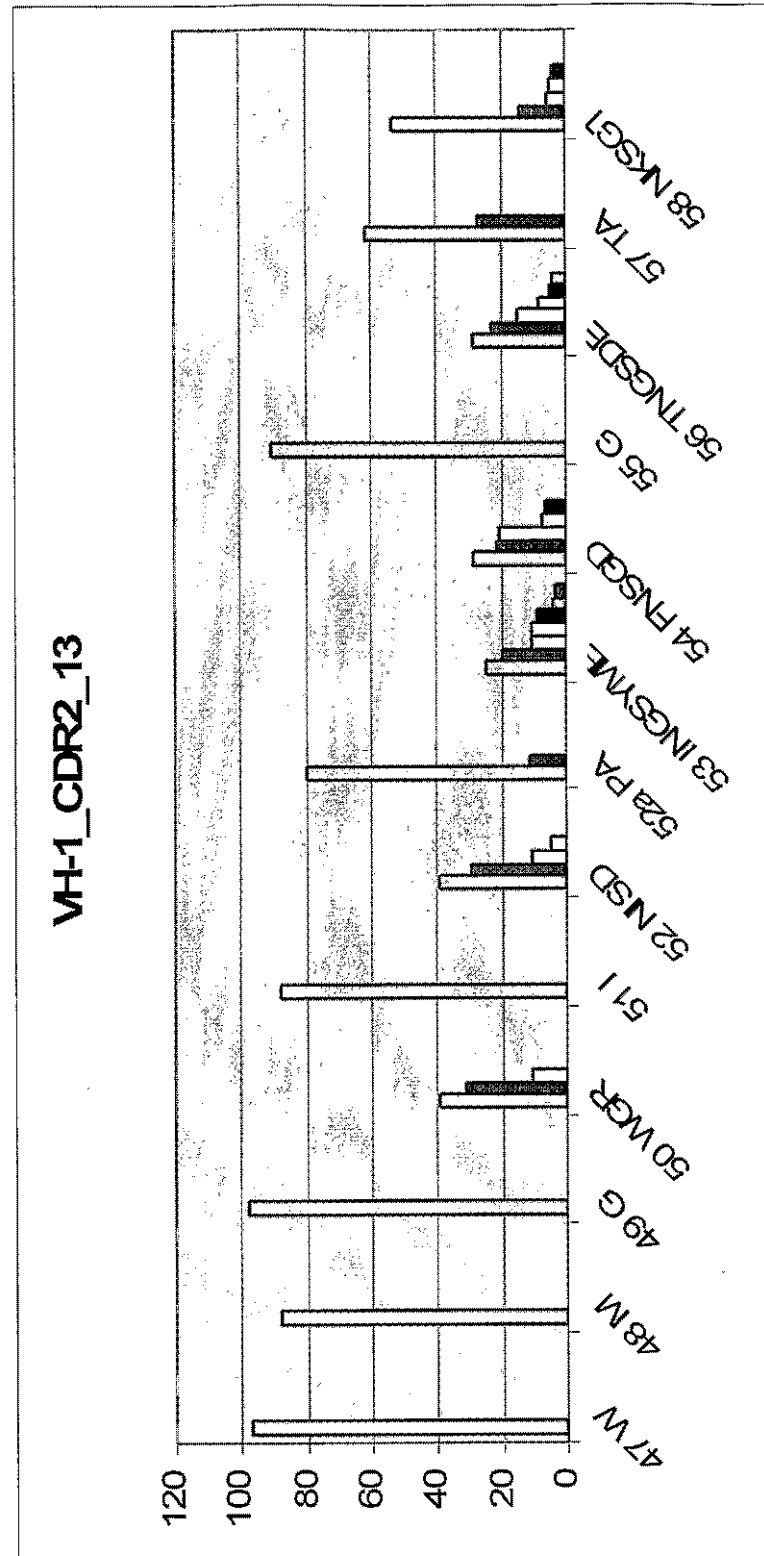


Fig. 49

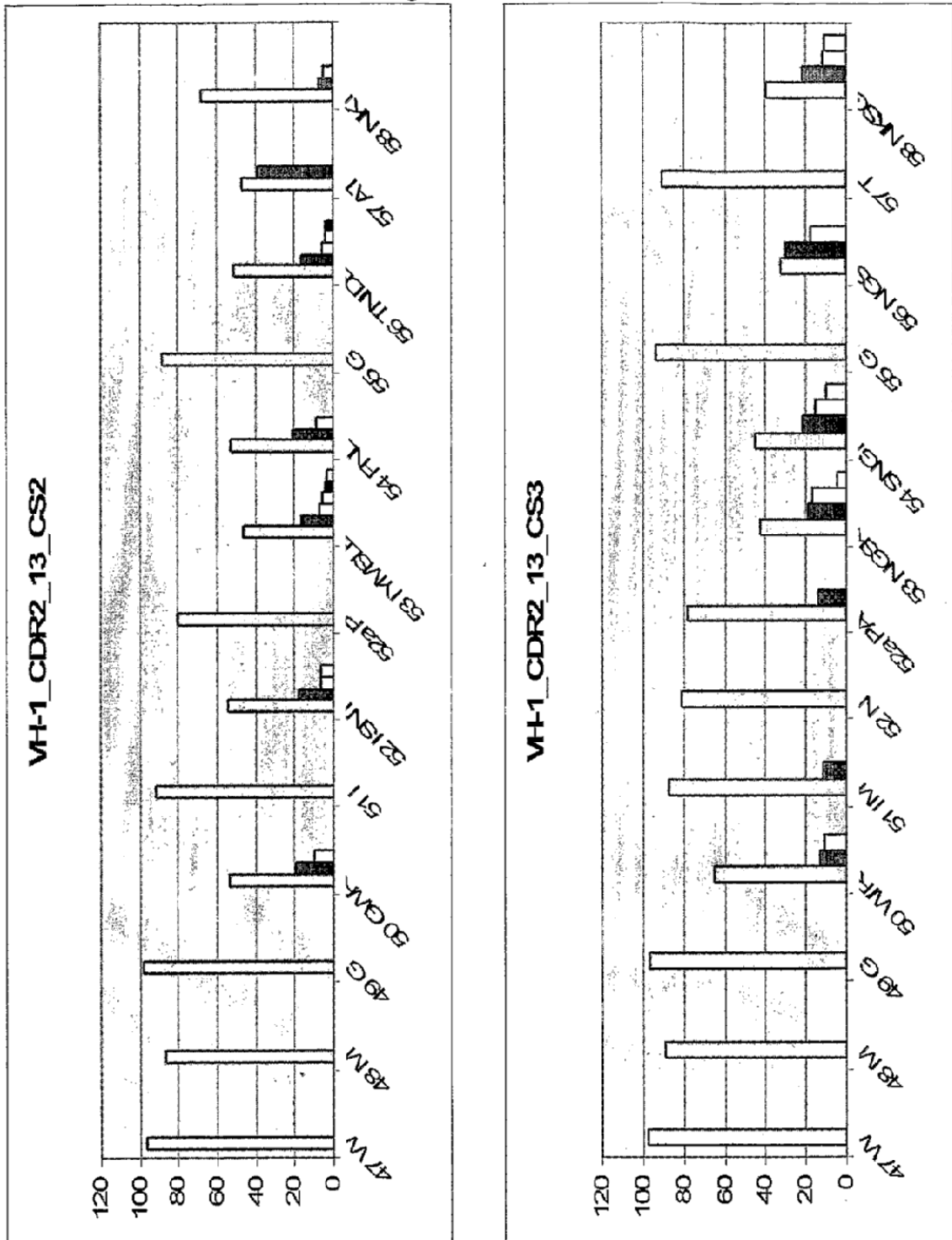
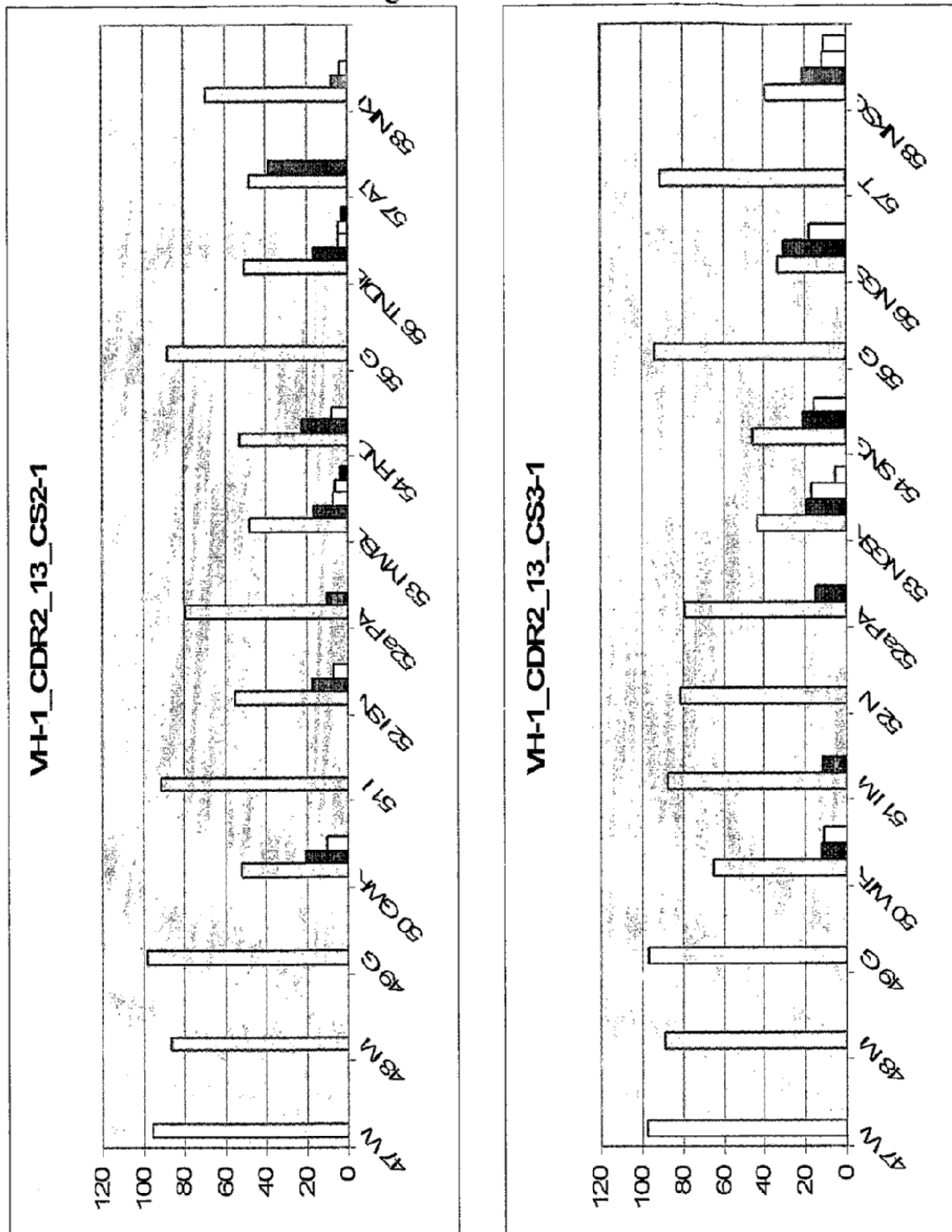


Fig. 50



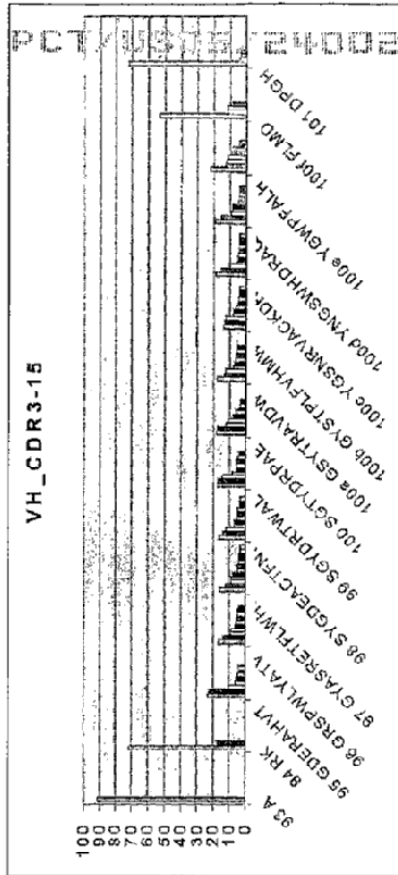


Fig. 51

