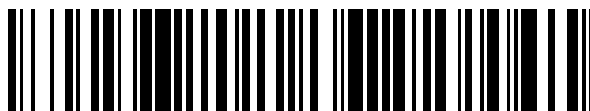


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 532 725**

51 Int. Cl.:

C12N 15/09 (2006.01)

C12N 15/13 (2006.01)

G01N 33/50 (2006.01)

A61K 39/395 (2006.01)

C07K 16/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.06.2008 E 08757277 (2)**

97 Fecha y número de publicación de la concesión europea: **07.01.2015 EP 2164961**

54 Título: **Modificación por ingeniería genética basada en secuencia y optimización de anticuerpos de cadena sencilla**

30 Prioridad:

25.06.2007 US 937112 P

12.03.2008 US 69057 P

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

31.03.2015

73 Titular/es:

**ESBATECH, AN ALCON BIOMEDICAL
RESEARCH UNIT LLC (100.0%)**

**Wagistrasse 21
8952 Schlieren, CH**

72 Inventor/es:

**BORRAS, LEONARDO y
URECH, DAVID**

74 Agente/Representante:

MILTENYI, Peter

ES 2 532 725 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

Modificación por ingeniería genética basada en secuencia y optimización de anticuerpos de cadena sencilla.

Antecedentes de la invención

5 Se ha demostrado que los anticuerpos son agentes terapéuticos muy eficaces y satisfactorios en el tratamiento de cáncer, enfermedades autoinmunitarias y otros trastornos. Aunque normalmente se han usado clínicamente anticuerpos de longitud completa, existen varias ventajas que el uso de un fragmento de anticuerpo puede proporcionar, tales como aumento de la penetración tisular, ausencia de función efectora de Fc combinada con la capacidad de añadir otras funciones efectoras y la probabilidad de menos efectos secundarios sistémicos que resultan de una semivida más corta *in vivo* de manera sistémica. Las propiedades farmacocinéticas de fragmentos de anticuerpo indican que pueden ser particularmente muy adecuados para enfoques terapéuticos locales. Además, los fragmentos de anticuerpo pueden ser más fáciles de producir que los anticuerpos de longitud completa en determinados sistemas de expresión.

15 Un tipo de fragmento de anticuerpo es un anticuerpo de cadena sencilla (scFv), que está compuesto por un dominio variable de cadena pesada (V_H) conjugado con un dominio variable de cadena ligera (V_L) por medio de una secuencia de ligador. Por tanto, los scFv carecen de todos los dominios de región constante de anticuerpo y los residuos de aminoácido de la primera superficie de contacto de dominio variable/constante (residuos de superficie de contacto) quedan expuestos al disolvente. Un scFv puede prepararse a partir de un anticuerpo de longitud completa (por ejemplo, molécula de IgG) a través de técnicas de modificación por ingeniería genética recombinantes. Sin embargo, la transformación de un anticuerpo de longitud completa en un scFv da como resultado a menudo escasa estabilidad y solubilidad de la proteína, bajos rendimientos de producción y una alta tendencia a agregarse, lo que eleva el riesgo de inmunogenicidad.

25 Por consiguiente, se han hecho intentos por mejorar propiedades tales como solubilidad y estabilidad de los scFv. Por ejemplo, Nieba, L. *et al.* (Prot. Eng. (1997) 10:435-444) seleccionaron tres residuos de aminoácido que se sabe que son residuos de superficie de contacto y los mutaron. Observaron un aumento de la expresión periplasmática del scFv mutado en bacterias, así como una disminución de la tasa de agregación inducida térmicamente, aunque la estabilidad y solubilidad termodinámicas no se alteraron significativamente. También se han notificado otros estudios en los que se llevó a cabo mutagénesis dirigida al sitio en residuos de aminoácido particulares dentro de scFv (véase por ejemplo, Tan, P.H. *et al.* (1988) Biophys. J. 75:1473-1482; Worn, A. y Plückthun, A. (1998) Biochem. 37:13120-13127; Worn, A. y Plückthun, A. (1999) Biochem. 38:8739-8750). En estos diversos estudios, los residuos de aminoácido fueron seleccionados para la mutagénesis basándose en sus posiciones conocidas dentro de la estructura de scFv (por ejemplo, a partir de estudios de modelado molecular).

35 En otro enfoque, se injertaron las regiones determinantes de complementariedad (CDR) de un scFv escasamente expresado en las regiones de marco de un scFv que se había demostrado que tiene propiedades favorables (Jung, S. y Plückthun, A. (1997) Prot. Eng. 10:959-966). El scFv resultante mostró estabilidad termodinámica y expresión soluble mejoradas.

40 El progreso de la modificación por ingeniería genética de scFv para mejorar propiedades funcionales se revisa en, por ejemplo, Worn, A. y Plückthun, A. (2001) J. Mol. Biol. 305:989-1010. Sin embargo, todavía se necesitan nuevos enfoques que permitan el diseño racional de scFv con propiedades funcionales superiores, en particular enfoques que ayuden al experto en la técnica en la selección de residuos de aminoácido posiblemente problemáticos para la modificación por ingeniería genética.

Monsellier E. *et al.* (Journal of Molecular Biology, 2006, vol. 362, n.º 3, páginas 580-593) dan a conocer un método en el que se comparó una secuencia de interés con la secuencia consenso global para VH/VL y se recuperaron las frecuencias de cada aminoácido en cada posición de un conjunto de bases de datos.

45 Pommié Christelle *et al.* (Journal of Molecular Recognition, 1995, vol. 17, n.º 1, páginas 17-232) describen los criterios normalizados de IMGT para el análisis estadístico de propiedades de aminoácidos de inmunoglobulinas.

Déret S. *et al.* (Computer Applications in the Biosciences: Cabios, 1995, vol. 11, n.º 4, páginas 435-439) describen el programa SUBIM que permite buscar en la base de datos Kabat y determinar el subgrupo de una región variable.

Johnson George *et al.* (Methods in Molecular Biology, 2004, vol. 248, páginas 11-25) es un artículo de revisión sobre la base de datos Kabat que comprende inmunoglobulinas humanas y murinas, entre otras.

50 Ewert S. *et al.* (Biochemistry, 2003, vol. 42, n.º 6, páginas 1517-1528) dan a conocer la alineación de secuencias de dominios de VH consenso humanos agrupados por dominios de VH con propiedades biofísicas favorables y dominios de VH con propiedades menos favorables.

55 Knappik A. *et al.* (Journal of Molecular Biology, 2000, vol. 296, n.º 1, páginas 57-86) describen la generación de secuencias HuCAL; se calcularon para cada subgrupo secuencias de inmunoglobulina reordenadas compiladas, alineadas, designadas a su subfamilia y regiones de marco consenso.

Honegger A. *et al.* (Journal of Molecular Biology. 2001, vol. 309, n.º 3, páginas 657-670) describen el esquema de numeración AHo para anticuerpos y proporcionan algunas macros de EXCEL que permiten la reenumeración semiautomática de archivos PDB.

5 Worn *et al.*, (Journal of Molecular Biology, 2001, vol. 305, n.º 5, páginas 989-1010) es un artículo de revisión sobre la estabilidad de la modificación por ingeniería genética de fragmentos scFv.

El documento WO 01/48017 (ESBATEch AG, 2001) describe el sistema de control de calidad (QC, *Quality Control*), es decir, un sistema de selección que permite el aislamiento de intracuerpos que son estables y solubles en un entorno reductor.

10 El documento WO 03/008451 (Univ. de Zúrich, 2003) se refiere a un método para la optimización de constructos de cadena pesada (VH) y ligera (VL) variables de inmunoglobulinas humanas aisladas.

Sumario de la invención

15 Esta invención proporciona métodos que permiten la identificación de residuos de aminoácido dentro de una secuencia de scFv que son posiblemente problemáticos para la estabilidad y/o solubilidad usando análisis basados en secuencia. Además, pueden seleccionarse residuos de aminoácido identificados según los métodos de la invención para la mutación y pueden prepararse scFv modificados por ingeniería genética que se han mutado y examinarse para detectar propiedades funcionales mejoradas tales como estabilidad y/o solubilidad. En una realización particularmente preferida, la invención proporciona métodos en los que se usa una base de datos de scFv seleccionados funcionalmente para identificar posiciones de residuos de aminoácido que son o bien más o bien menos tolerantes a la variabilidad que las posiciones correspondientes en secuencias de inmunoglobulina de anticuerpos maduros y/o de línea germinal, indicando de ese modo que tales posiciones de residuos identificadas pueden ser adecuadas para la modificación por ingeniería genética para mejorar la funcionalidad de scFv tal como estabilidad y/o solubilidad. Por tanto, la invención proporciona, y demuestra el beneficio de, un enfoque de "consenso funcional" basado en el uso de una base de datos de secuencias de scFv seleccionadas funcionalmente.

25 En aún otras realizaciones preferidas, la invención proporciona métodos para identificar residuos de aminoácido preferidos que van a sustituirse (o alternativamente, residuos de aminoácido que van a excluirse) posiciones de aminoácido de interés (por ejemplo, posiciones de aminoácido identificadas comparando una base de datos de secuencias de scFv que tienen al menos una propiedad deseable, por ejemplo, tal como se seleccionan con un ensayo de QC, frente a una base de datos de secuencias de anticuerpos maduros, por ejemplo, la base de datos Kabat) en un agente de unión inmunológica. Por tanto, la invención proporciona además métodos de "enriquecimiento/exclusión" para seleccionar un residuo de aminoácido particular. Todavía adicionalmente, la invención proporciona métodos de modificación por ingeniería genética de agentes de unión inmunológica (por ejemplo, scFv, inmunoglobulinas de longitud completa, fragmentos Fab, anticuerpos con un solo dominio (por ejemplo, Dab) y Nanobodies) mutando posiciones de aminoácido de región de marco particulares identificadas usando el enfoque de "consenso funcional" descrito en el presente documento. En realizaciones preferidas, las posiciones de aminoácido de región de marco se mutan sustituyendo el residuo de aminoácido existente por un residuo que se encuentra que es un residuo "enriquecido" usando los métodos de análisis de "enriquecimiento/exclusión" descritos en el presente documento.

40 En un aspecto, la invención se refiere a un método de identificación de una posición de aminoácido para mutación en un anticuerpo de cadena sencilla (scFv), teniendo el scFv secuencias de aminoácidos de V_H y V_L, comprendiendo el método:

a) introducir las secuencias de aminoácidos de V_H, V_L o V_H y V_L de scFv en una base de datos que comprende una multiplicidad de secuencias de aminoácidos de V_H, V_L o V_H y V_L de anticuerpos de manera que las secuencias de aminoácidos de V_H, V_L o V_H y V_L de scFv se alinean con las secuencias de aminoácidos de V_H, V_L o V_H y V_L de anticuerpos de la base de datos;

45 b) comparar una posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv con una posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos;

c) determinar si la posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv está ocupada por un residuo de aminoácido que está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

50 d) identificar la posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando la posición de aminoácido está ocupada por un residuo de aminoácido que no está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

55 El método puede comprender además mutar la posición de aminoácido identificada para mutación dentro de la secuencia de aminoácidos de V_H o V_L de scFv. Por ejemplo, la posición de aminoácido identificada para mutación puede sustituirse por un residuo de aminoácido que está conservado en la posición correspondiente dentro de las

5 secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Adicional o alternativamente, la posición de aminoácido identificada para mutación puede mutarse mediante mutagénesis al azar o mediante mutagénesis sesgada para generar una biblioteca de scFv mutados, seguido por el examen de la biblioteca de scFv mutados y la selección de scFv que tienen al menos una propiedad funcional mejorada (por ejemplo, mediante el examen de la biblioteca usando un sistema de control de calidad (sistema de QC) de levaduras).

10 Pueden usarse diversos tipos de bases de datos en los métodos de la invención. Por ejemplo, en una realización, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de línea germinal. En otra realización, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos reordenados, madurados por afinidad. En aún otra realización, particularmente preferida, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv seleccionadas por tener al menos una propiedad funcional deseable (tal como estabilidad de scFv o solubilidad de scFv). En aún otra realización, puede usarse más de una base de datos para fines de comparación. Por ejemplo, en una realización particularmente preferida, se usa una base de datos de scFv seleccionada por tener al menos una propiedad funcional deseable, así como una o más bases de datos de línea germinal o bases de datos de anticuerpos reordenados, madurados por afinidad, en las que los resultados de comparación de secuencias de la base de datos de scFv se comparan con los resultados de la(s) otra(s) base(s) de datos.

20 Los métodos de la invención pueden usarse para analizar, por ejemplo, la región V_H de un scFv, la región V_L de un scFv o ambas.

25 Aunque los métodos de la invención pueden usarse para analizar una única posición de aminoácido de interés dentro de un scFv de interés, más preferiblemente los métodos se usan para analizar múltiples posiciones de aminoácido a lo largo de la secuencia de scFv. Por tanto, en una realización preferida, en la etapa b) del método, se comparan múltiples posiciones de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv con posiciones correspondientes dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Por ejemplo, en una realización preferida, cada posición de región de marco dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv se compara con cada posición de región de marco correspondiente dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos. Adicional o alternativamente, pueden analizarse una o más posiciones dentro de una o más CDR del scFv. En aún otra realización, cada posición de aminoácido dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv se compara con cada posición de aminoácido correspondiente dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos.

35 Una posición de aminoácido que está "conservada" entre las secuencias de la base de datos puede estar ocupada por uno o más tipos particulares de residuos de aminoácido. Por ejemplo, en una realización, la posición "conservada" está ocupada por un residuo de aminoácido particular que aparece a una frecuencia muy alta en esa posición. Es decir, en la etapa c) del método, el residuo de aminoácido que está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos es el residuo de aminoácido que está lo más frecuentemente en esa posición o lo más significativamente enriquecido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En esta situación, para crear un scFv modificado por ingeniería genética, la posición de aminoácido identificada para mutación puede sustituirse por el residuo de aminoácido que se observa lo más frecuentemente o está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En otra realización, una posición de aminoácido que está "conservada" entre las secuencias de la base de datos puede estar ocupada, por ejemplo, por (i) residuos de aminoácido hidrófobos, (ii) residuos de aminoácido hidrófilos, (iii) residuos de aminoácido que pueden formar un enlace de hidrógeno o (iv) residuos de aminoácido que tienen una propensión a formar una lámina β . Es decir, en la etapa c) del método, la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos está conservada con: (i) residuos de aminoácido hidrófobos, (ii) residuos de aminoácido hidrófilos, (iii) residuos de aminoácido que pueden formar un enlace de hidrógeno o (iv) residuos de aminoácido que tienen una propensión a formar una lámina β .

40 Por consiguiente, para crear un scFv modificado por ingeniería genética, cuando la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos está conservada con residuos de aminoácido hidrófobos, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido hidrófobo que está lo más frecuentemente en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido hidrófobo que se selecciona como el que mejor se ajusta a esa posición dentro del scFv (por ejemplo, el residuo hidrófobo que lo más probablemente mantenga la estructura y función del scFv basándose en estudios de modelado molecular). Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un panel de residuos de aminoácido hidrófobos por medio de mutagénesis dirigida al sitio para crear una biblioteca de scFv modificados por ingeniería genética y la(s) sustitución/sustituciones más preferida(s) puede(n) seleccionarse mediante el examen de la biblioteca para detectar propiedades funcionales deseables (por ejemplo,

en un sistema de QC de levaduras).

Además, para crear un scFv modificado por ingeniería genética, cuando la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos está conservada con residuos de aminoácido hidrófilos, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido hidrófilo que está lo más frecuentemente en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido hidrófilo que se selecciona como el que mejor se ajusta a esa posición dentro del scFv (por ejemplo, el residuo hidrófilo que lo más probablemente mantenga la estructura y función del scFv basándose en estudios de modelado molecular). Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un panel de residuos de aminoácido hidrófilos por medio de mutagénesis dirigida al sitio para crear una biblioteca de scFv modificados por ingeniería genética y la(s) sustitución/sustituciones más preferida(s) puede(n) seleccionarse mediante el examen de la biblioteca para detectar propiedades funcionales deseables (por ejemplo, en un sistema de QC de levaduras).

Todavía adicionalmente, para crear un scFv modificado por ingeniería genética, cuando la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos está conservada con residuos de aminoácido que pueden formar un enlace de hidrógeno, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por el residuo de aminoácido que puede formar un enlace de hidrógeno que está lo más frecuentemente en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido que puede formar un enlace de hidrógeno que se selecciona como el que mejor se ajusta a esa posición dentro del scFv (por ejemplo, el residuo que lo más probablemente mantenga la estructura y función del scFv basándose en estudios de modelado molecular). Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un panel de residuos de aminoácido que pueden formar un enlace de hidrógeno, por medio de mutagénesis dirigida al sitio, para crear una biblioteca de scFv modificados por ingeniería genética y la(s) sustitución/sustituciones más preferida(s) puede(n) seleccionarse mediante el examen de la biblioteca para detectar propiedades funcionales deseables (por ejemplo, en un sistema de QC de levaduras).

Todavía adicionalmente, para crear un scFv modificado por ingeniería genética, cuando la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos está conservada con residuos de aminoácido que tienen una propensión a formar una lámina β , la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por el residuo de aminoácido que tiene una propensión a formar una lámina β que está lo más frecuentemente en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un residuo de aminoácido que tiene una propensión a formar una lámina β que se selecciona como el que mejor se ajusta a esa posición dentro del scFv (por ejemplo, el residuo que lo más probablemente mantenga la estructura y función del scFv basándose en estudios de modelado molecular). Adicional o alternativamente, la posición de aminoácido identificada para mutación dentro del scFv puede sustituirse por un panel de residuos de aminoácido que tienen una propensión a formar una lámina β , por medio de mutagénesis dirigida al sitio, para crear una biblioteca de scFv modificados por ingeniería genética y la(s) sustitución/sustituciones más preferida(s) puede(n) seleccionarse mediante el examen de la biblioteca para detectar propiedades funcionales deseables (por ejemplo, en un sistema de QC de levaduras).

En otra realización, el método para identificar una posición de aminoácido para mutación en un scFv puede realizarse usando una base de datos que es una base de datos restringida en la que sólo las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos que tienen alta similitud con las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv se incluyen en la base de datos.

En una realización preferida, para cuantificar la conservación de la posición de aminoácido que está analizándose (es decir, la "posición correspondiente" dentro de la secuencia de aminoácidos de V_H o V_L de anticuerpos de la base de datos), a la posición de aminoácido se le asigna un grado de conservación usando el índice de Simpson.

Los métodos de la invención también pueden usarse para examinar pares de posiciones de aminoácido dentro de la secuencia de scFv, para identificar posiciones de aminoácido que varían conjuntamente entre sí de manera que una o ambas de estas posiciones de pares covariantes pueden identificarse para mutación. Por tanto, en otra realización, la invención se refiere a un método que comprende:

- a) llevar a cabo un análisis de covarianza en secuencias de aminoácidos de V_H o V_L de anticuerpos de una base de datos para identificar un par covariante de posiciones de aminoácido;
- b) comparar el par covariante de posiciones de aminoácido con posiciones correspondientes dentro de una secuencia de aminoácidos de V_H o V_L de scFv;
- c) determinar si las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv están

ocupadas por residuos de aminoácido que están conservados en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

- 5 d) identificar una o ambas de las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando una o ambas de las posiciones correspondientes dentro del scFv está ocupada por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

Este análisis de covarianza también puede combinarse con el análisis de posiciones de aminoácido individuales de manera que el método de identificación de una posición de aminoácido para mutación en un scFv descrito anteriormente con las etapas a) - d) puede comprender además las etapas de:

- 10 e) llevar a cabo un análisis de covarianza en las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos para identificar un par covariante de posiciones de aminoácido;

f) comparar el par covariante de posiciones de aminoácido con posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv;

- 15 g) determinar si las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv están ocupadas por residuos de aminoácido que están conservados en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

- 20 h) identificar una o ambas de las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando una o ambas de las posiciones correspondientes dentro del scFv está ocupada por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

Los métodos de análisis de covarianza pueden aplicarse a un único par covariante o, alternativamente, pueden identificarse múltiples pares covariantes de posiciones de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de anticuerpos de la base de datos y compararse con las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv.

- 25 El método puede comprender además mutar una o ambas de las posiciones correspondientes dentro del scFv que están ocupadas por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Por ejemplo, en una realización, una de las posiciones correspondientes dentro del scFv que está ocupada por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido se sustituye por un residuo de aminoácido que está lo más frecuentemente en la posición de aminoácido de par covariante. En otra realización, ambas de las posiciones correspondientes dentro del scFv que están ocupadas por residuos de aminoácido que no están conservados en el par covariante de posiciones de aminoácido se sustituyen por residuos de aminoácido que están lo más frecuentemente en las posiciones de aminoácido de par covariante.
- 30

- 35 Los métodos basados en secuencia para identificar posiciones de aminoácido para mutación dentro una secuencia de scFv pueden combinarse con otros métodos que permiten el análisis estructural de scFv. Por ejemplo, en un aspecto, los métodos basados en secuencia pueden combinarse con métodos de modelado molecular para analizar adicionalmente la estructura del scFv. Por tanto, en un aspecto, los métodos descritos anteriormente con las etapas a) - d) pueden comprender además, por ejemplo, las etapas de:

e) someter las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv a modelado molecular; y

- 40 f) identificar al menos una posición de aminoácido adicional dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv para mutación.

Este método puede comprender además mutar la al menos una posición de aminoácido adicional dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv identificadas para mutación mediante modelado molecular.

- 45 En otro aspecto, la invención se refiere a composiciones de scFv preparadas según los métodos de la invención en las que se realizan una o más mutaciones en una o más posiciones de aminoácido identificadas para mutación. También se proporcionan formulaciones farmacéuticas, formulaciones que comprenden normalmente la composición de scFv y un portador farmacéuticamente aceptable.

- 50 En aún otro aspecto, la invención se refiere a un método de identificación de una o más posiciones de aminoácido de región de marco para mutación en un anticuerpo de cadena sencilla (scFv), teniendo el scFv secuencias de aminoácidos de V_H y V_L , comprendiendo el método:

a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H , V_L o V_H y V_L ;

b) proporcionar una segunda base de datos de secuencia de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de scFv seleccionadas por tener al menos una propiedad funcional deseable;

c) determinar la variabilidad de aminoácidos en cada posición de región de marco de la primera base de datos y en cada posición de región de marco de la segunda base de datos;

5 d) identificar una o más posiciones de región de marco en las que el grado de variabilidad de aminoácidos difiere entre la primera base de datos y la segunda base de datos para identificar de ese modo una o más posiciones de aminoácido de región de marco para mutación en un anticuerpo de cadena sencilla (scFv).

10 Preferiblemente, la variabilidad de aminoácidos en cada posición de región de marco se determina asignando un grado de conservación usando el índice de Simpson. En una realización, la una o más posiciones de aminoácido de región de marco se identifican para mutación basándose en la una o más posiciones de aminoácido de región de marco que tienen un valor del índice de Simpson (SI) inferior en la segunda base de datos en comparación con la primera base de datos. En otra realización, la una o más posiciones de aminoácido de región de marco se identifican para mutación basándose en la una o más posiciones de aminoácido de región de marco que tienen un valor del índice de Simpson superior en la segunda base de datos en comparación con la primera base de datos (por ejemplo, base de datos de línea germinal). En una realización preferida, la posición de aminoácido de la segunda base de datos (por ejemplo, una base de datos de QC) tiene un valor de IS que es al menos 0,01 menor, y más preferiblemente 0,05 menor (por ejemplo, 0,06, 0,07, 0,08, 0,09 ó 0,1), que el valor de IS de la posición de aminoácido correspondiente en la primera base de datos (por ejemplo, una base de datos Kabat). En más realizaciones preferidas, la posición de aminoácido de la segunda base de datos tiene un valor de IS que es al menos 0,1 menor (por ejemplo, 0,1, 0,15, 0,2, 0,25, 0,3, 0,35, 0,4, 0,45 ó 0,5 menor) que el valor de IS de la posición de aminoácido correspondiente en la primera base de datos.

20 En otro aspecto, la invención se refiere a un método de identificación de un residuo de aminoácido preferido para sustitución en un agente de unión inmunológica, comprendiendo el método:

a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H o V_L agrupadas (por ejemplo, secuencias de anticuerpos maduros y/o de línea germinal agrupadas según el subtipo de familia Kabat);

25 b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H o V_L de anticuerpos scFv agrupadas seleccionadas por tener al menos una propiedad funcional deseable (por ejemplo, según un ensayo de QC);

c) determinar la frecuencia de aminoácidos para un residuo de aminoácido en una posición de región de marco de la primera base de datos y en una posición de región de marco correspondiente de la segunda base de datos;

30 d) identificar el residuo de aminoácido como un residuo de aminoácido preferido para sustitución en una posición de aminoácido correspondiente del agente de unión inmunológica cuando el residuo de aminoácido aparece a una frecuencia superior en la segunda base de datos en relación con la primera base de datos (es decir, un residuo enriquecido).

En aún otro aspecto, la invención se refiere a un método de identificación de un residuo de aminoácido que va a excluirse de un agente de unión inmunológica, comprendiendo el método:

35 a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H o V_L agrupadas (por ejemplo, secuencias de anticuerpos maduros y/o de línea germinal agrupadas según el subtipo de familia Kabat);

b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H o V_L de anticuerpos de scFv agrupadas seleccionadas por tener al menos una propiedad funcional deseable (por ejemplo, según un ensayo de QC);

40 c) determinar la frecuencia de aminoácidos para un residuo de aminoácido en una posición de región de marco de la primera base de datos y en una posición de región de marco correspondiente de la segunda base de datos;

45 d) identificar el residuo de aminoácido como un residuo de aminoácido desfavorecido para sustitución en una posición de aminoácido correspondiente del agente de unión inmunológica cuando el residuo de aminoácido aparece a una frecuencia inferior en la segunda base de datos en relación con la primera base de datos, en el que dicho tipo de residuo de aminoácido es un residuo de aminoácido desfavorecido (es decir, un residuo excluido). En determinadas realizaciones preferidas, el residuo de aminoácido desfavorecido se identifica si el factor de enriquecimiento (FE) es inferior a 1.

50 En determinadas realizaciones, la primera base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y V_L de línea germinal. En otras realizaciones, la primera base de datos consiste en secuencias de aminoácidos de V_H , V_L o V_H y V_L de línea germinal. En aún otra realización, la primera base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras. En otra realización, la primera base de datos consiste en secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras. En una realización a modo de ejemplo, las secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras son de una base de datos Kabat (KDB).

En determinadas realizaciones, la segunda base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y

V_L de anticuerpos scFv seleccionadas a partir de un ensayo de QC. En otra realización, la segunda base de datos consiste en secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv seleccionadas a partir de un ensayo de QC.

5 En una realización, la propiedad funcional deseable es estabilidad mejorada. En otra realización, la propiedad funcional deseable es solubilidad mejorada. En aún otra realización, la propiedad funcional deseable es no agregación. En todavía otra realización, la propiedad funcional deseable es una mejora en la expresión (por ejemplo, en una célula procarionota). En aún otra realización, la propiedad funcional es una mejora en el rendimiento de repliegamiento tras un procedimiento de purificación de cuerpos de inclusión. En determinadas realizaciones, la
10 propiedad funcional deseable no es una mejora en la afinidad de unión a antígeno. En otra realización, se logra una combinación de una o más funciones deseables.

En aún otro aspecto, la invención se refiere a un método de modificación por ingeniería genética de un agente de unión inmunológica, comprendiendo el método:

a) seleccionar una o más posiciones de aminoácido dentro del agente de unión inmunológica para mutación; y

b) mutar la una o más posiciones de aminoácido seleccionadas para mutación,

15 en el que si la una o más posiciones de aminoácido seleccionadas para mutación son posiciones de aminoácido de cadena pesada, dicha posición de aminoácido se selecciona del grupo que consiste en las posiciones 1, 6, 7, 10, 12, 13, 14, 19, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87, 89, 90, 92, 95, 98, 103 y 107 usando la numeración AHO (posiciones de aminoácido 1, 6, 7, 9, 11, 12, 13, 18, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76, 78, 79, 81, 82b, 84, 89 y 93 usando la numeración Kabat), y

20 en el que si la una o más posiciones de aminoácido seleccionadas para mutación son posiciones de aminoácido de cadena ligera, dichas posiciones de aminoácido se seleccionan del grupo que consiste en las posiciones 1, 2, 3, 4, 7, 10, 11, 12, 14, 18, 20, 24, 46, 47, 50, 53, 56, 57, 74, 82, 91, 92, 94, 101 y 103 usando la numeración AHO (posiciones de aminoácido 1, 2, 3, 4, 7, 10, 11, 12, 14, 18, 20, 24, 38, 39, 42, 45, 48, 49, 58, 66, 73, 74, 76, 83 y 85 usando la numeración Kabat).

25 En un aspecto, la invención se refiere a un método de modificación por ingeniería genética de un agente de unión inmunológica, comprendiendo el método:

a) seleccionar una o más posiciones de aminoácido dentro del agente de unión inmunológica para mutación; y

b) mutar la una o más posiciones de aminoácido seleccionadas para mutación,

30 en el que la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en:

(i) las posiciones de aminoácido 1, 6, 7, 89 y 103 de $VH3$ usando la numeración AHO (posiciones de aminoácido 1, 6, 7, 78 y 89 usando la numeración Kabat);

(ii) las posiciones de aminoácido 1, 6, 12, 13, 14, 19, 21, 90, 92, 95 y 98 de $VH1a$ usando la numeración AHO (posiciones de aminoácido 1, 6, 11, 12, 13, 18, 20, 79, 81, 82b y 84 usando la numeración Kabat);

35 (iii) las posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107 de $VH1b$ usando la numeración AHO (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76 y 93 usando la numeración Kabat);

(iv) las posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57, 91 y 103 de $V\kappa1$ usando la numeración AHO (posiciones de aminoácido 1, 3, 4, 24, 39, 42, 49, 73 y 85 usando la numeración Kabat);

40 (v) las posiciones de aminoácido 2, 3, 10, 12, 18, 20, 56, 74, 94, 101 y 103 de $V\kappa3$ usando la numeración AHO (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 48, 58, 76, 83 y 85 usando la numeración Kabat); y

(vi) las posiciones de aminoácido 1, 2, 4, 7, 11, 14, 46, 53, 82, 92 y 103 de $V\lambda1$ usando la numeración AHO (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 38, 45, 66, 74 y 85 usando la numeración Kabat).

45 En otra realización, la mutación comprende además una o más (preferiblemente todas) las sustituciones de cadena de una posición de aminoácido de cadena pesada seleccionada del grupo que consiste en las posiciones de aminoácido 12, 13 y 144 usando la numeración AHO (posiciones de aminoácido 12, 85 y 103 usando la numeración Kabat).

50 En determinadas realizaciones preferidas, la una o más posiciones de aminoácido seleccionadas para mutación se mutan a un residuo de aminoácido encontrado en una posición de aminoácido correspondiente en una secuencia de anticuerpo seleccionada por tener al menos una propiedad funcional deseable (por ejemplo, en un sistema de QC de levaduras). En aún otras realizaciones, la una o más posiciones de aminoácido seleccionadas para mutación se

mutan a un residuo de aminoácido (por ejemplo, un “residuo de aminoácido enriquecido”) identificado según la metodología de análisis de enriquecimiento/exclusión de la invención.

5 Preferiblemente, el agente de unión inmunológica es un scFv, pero otros agentes de unión inmunológica, tales como inmunoglobulinas de longitud completa y otros fragmentos de anticuerpo (por ejemplo, Fab o Dab), también pueden modificarse por ingeniería genética según el método. La divulgación también abarca agentes de unión inmunológica preparados según el método de modificación por ingeniería genética, así como composiciones que comprenden los agentes de unión inmunológica y un portador farmacéuticamente aceptable.

Breve descripción de las figuras

10 La invención se entenderá mejor y objetos distintos de los expuestos anteriormente resultarán evidentes cuando se tenga en consideración la siguiente descripción detallada de la misma. Tal descripción hace referencia a los dibujos adjuntos, en los que:

La figura 1 es un diagrama de flujo que resume análisis de scFv basados en secuencia generales según los métodos de la invención.

15 En una primera etapa, se proporciona una secuencia de un scFv que va a mejorarse en solubilidad y estabilidad (recuadro 1), que posteriormente se compara con bases de datos de secuencias de anticuerpos (recuadro 2), tales como bases de datos de secuencias de línea germinal de código abierto (por ejemplo, Vbase, IMGT; recuadro 3), bases de datos de secuencias de anticuerpos maduros de código abierto (por ejemplo, KDB; recuadro 4) o bases de datos de fragmentos scFv estables y solubles completamente humanos (por ejemplo, QC; recuadro 5).

20 La aplicación de bases de datos de secuencias de línea germinal de código abierto tal como se describe en el recuadro 3 permite la identificación de posiciones altamente conservadas que se seleccionaron durante la evolución y por tanto se cree que contribuyen a la estabilidad de dominios variables en el contexto de anticuerpos de longitud completa (recuadro 3'). Una comparación frente a bases de datos de secuencias de anticuerpos maduros de código abierto 4 permite la identificación de patrones que representan mejoras de estabilidad, solubilidad y/o unión que son independientes de las respectivas CDR (recuadro 4'). Además, la comparación frente a bases de datos de fragmentos scFv estables y solubles completamente humanos (recuadro 5) conduce a la identificación de residuos críticos para la estabilidad y/o solubilidad específicamente en el formato de scFv, así como la identificación de patrones que representan mejoras de estabilidad, solubilidad y/o unión independientes de las respectivas CDR, específicamente en el formato de scFv, por ejemplo combinaciones de V_L y V_H (recuadro 5').

30 En una siguiente etapa (recuadro 6), se hace una sustitución de residuos críticos por el aminoácido adecuado más frecuente tal como se identifica en la base de datos respectiva.

Finalmente (recuadro 7), puede realizarse mutagénesis al azar o sesgada de residuos críticos y examen posterior para detectar estabilidad y/o solubilidad mejoradas en el sistema de QC de levaduras. Los mutantes pueden someterse de nuevo al procedimiento mencionado anteriormente (flecha hasta el recuadro 2).

35 La figura 2 es un diagrama de flujo de un método de múltiples etapas a modo de ejemplo para el análisis basado en secuencia de scFv.

40 En una primera etapa (recuadro 1), se determina la frecuencia de cada residuo en la región de marco comparando la aparición de diferentes aminoácidos en cada posición, basándose en los resultados proporcionados por herramientas bioinformáticas. En una segunda etapa, se define un grado de conservación en cada posición, por ejemplo usando el índice de Simpson con la fórmula $D = \sum n_i (n_i - 1) / N(N - 1)$. En una tercera etapa, se determina la mejor sustitución que minimiza la energía libre global (por ejemplo, aplicando la ley de Boltzmann: $\Delta\Delta G_{th} = -RT \ln(f_{original}/f_{consenso})$). Finalmente (etapa 4), se determina el papel de posibles mutaciones de estabilización. Para este fin, pueden tenerse en cuenta factores tales como interacciones locales y no locales, residuos canónicos, superficies de contacto, grado de exposición y propensión a giros β .

45 La figura 3 es un diagrama esquemático de un sistema de control de calidad (QC) a modo de ejemplo para la selección de scFv estables y solubles en levadura. Con este sistema, se seleccionan células huésped que pueden expresar scFv estables y solubles en un entorno reductor debido a la presencia de un constructo indicador inducible cuya expresión depende de la presencia de una proteína de fusión scFv-AD-Gal 11p estable y soluble. La interacción de la proteína de fusión con Gal4 (1-100) forma un factor de transcripción funcional que activa la expresión de un marcador seleccionable (véase la figura 3A). scFv inestables y/o insolubles no pueden formar un factor de transcripción funcional ni inducir la expresión del marcador seleccionable y por tanto se excluyen de la selección (figura 3B). Los scFv seleccionados pueden obtener un plegamiento de proteína estable y soluble, incluso en condiciones reductoras, en las que los enlaces disulfuro no se pliegan, mientras que scFv inestables y/o insolubles tienden a desplegarse, agregarse y/o degradarse. En condiciones oxidantes, los scFv seleccionados revelan todavía características de solubilidad y estabilidad superiores.

55 La figura 4 es un diagrama esquemático de otro sistema de control de calidad (QC) a modo de ejemplo. El concepto global para seleccionar scFv estables y solubles es el mismo que se describió para la figura 3, sin embargo, en esta

versión, el scFv se fusiona directamente con un factor de transcripción funcional que comprende un dominio de activación (AD) y un dominio de unión a ADN (DBD). La figura 4A representa un scFv soluble y estable a modo de ejemplo que, cuando se fusiona con un factor de transcripción funcional, no dificulta la transcripción de un marcador seleccionable. En cambio, la figura 4B representa el escenario mediante el cual un scFv inestable se fusiona con el factor de transcripción dando lugar a un constructo de fusión no funcional que no puede activar la transcripción del marcador seleccionable.

La figura 5 es un diagrama esquemático del análisis de variabilidad en posiciones de región de marco (FW) particulares dentro de secuencias de línea germinal nativas antes de la mutación somática (figura 5A) y en las posiciones de FW correspondientes dentro de secuencias de anticuerpos maduros tras la mutación somática seleccionadas en el sistema de QC (figura 5B). Pueden asignarse valores de variabilidad diferentes a las respectivas posiciones de FW (por ejemplo, residuos de región de marco altamente variables ("hvFR")) dentro de las secuencias de línea germinal y de QC (es decir, valores de "G" y "C", respectivamente). Si $G > C$ para una posición particular, hay un número restringido de residuos de FW estables adecuados en esa posición. Si $G < C$ para una posición particular, esto puede indicar que el residuo se ha seleccionado de manera natural para solubilidad y estabilidad óptimas.

La figura 6 representa el perfil de desnaturalización observado para variantes de ESBA105 tras estrés inducido térmicamente a un intervalo de temperaturas de desde 25 hasta 95°C. Las variantes de ESBA105 que tienen retromutaciones a residuos consenso de línea germinal (V3Q, R47K o V103T) se indican mediante líneas discontinuas. Las variantes que comprenden sustituciones preferidas identificadas mediante los métodos de la invención (QC11.2, QC15.2 y QC23.2) se indican mediante líneas continuas.

La figura 7 representa una comparación de la estabilidad térmica para un conjunto de variantes de ESBA105 que comprenden o bien retromutaciones consenso (S-2, D-2, D-3), una mutación a alanina (D-1) o un residuo de QC (QC7.1, QC11.2, QC15.2, QC23.2). Se proporciona la estabilidad térmica de cada variante (en unidades de desplegamiento arbitrarias).

La figura 8 representa el perfil de desnaturalización observado para variantes de ESBA212 tras estrés inducido térmicamente a un intervalo de temperaturas de desde 25 hasta 95°C. Las variantes de ESBA212 que tienen retromutaciones a residuos consenso de línea germinal (V3Q o R47K) se indican mediante líneas discontinuas. La molécula de ESBA212 original se indica mediante una línea continua.

La figura 9 representa las curvas de solubilidad por precipitación en PEG de ESBA105 silvestre y variantes de solubilidad de la misma.

La figura 10 representa los perfiles de desnaturalización térmica para ESBA105 silvestre y variantes de solubilidad de la misma tal como se mide tras la exposición térmica a un amplio intervalo de temperaturas (25-96°C).

La figura 11 representa un gel de SDS-PAGE que muestra el comportamiento de degradación de diversos mutantes de solubilidad de ESBA105 tras dos semanas de incubación en condiciones de estrés térmico.

Descripción detallada de la invención

La invención se refiere a métodos para la modificación por ingeniería genética basada en secuencia y la optimización de las propiedades de agentes de unión inmunológica, y en particular las propiedades de scFv, incluyendo pero sin limitarse a, estabilidad, solubilidad y afinidad. Más específicamente, la presente invención da a conocer métodos para optimizar anticuerpos scFv usando análisis de secuencias de anticuerpos para identificar posiciones de aminoácido dentro de un scFv que van a mutarse para mejorar de ese modo una o más propiedades físicas del scFv. La invención también se refiere a agentes de unión inmunológica modificados por ingeniería genética, por ejemplo, scFv, producidos según los métodos de la invención.

La invención se basa, al menos en parte, en el análisis de la frecuencia de aminoácidos en cada posición de región de marco de cadena pesada y ligera en múltiples bases de datos de secuencias de anticuerpos. En particular, el análisis de frecuencia de bases de datos de anticuerpos maduros y/o de línea germinal se ha comparado con el análisis de frecuencia de una base de datos de secuencias de scFv que se han seleccionado por tener propiedades funcionales deseadas. Asignando un grado de variabilidad a cada posición de región de marco (por ejemplo, usando el índice de Simpson) y comparando el grado de variabilidad en cada posición de región de marco dentro de los diferentes tipos de bases de datos de secuencias de anticuerpos, ha sido posible ahora identificar posiciones de región de marco de importancia para las propiedades funcionales (por ejemplo, estabilidad, solubilidad) de un scFv. Esto permite ahora la definición de un "consenso funcional" para las posiciones de aminoácido de región de marco, en el que se han identificado posiciones de región de marco que son o bien más o bien menos tolerantes a la variabilidad que las posiciones correspondientes en secuencias de inmunoglobulina de anticuerpos maduros y/o de línea germinal. Por tanto, la invención proporciona, y demuestra el beneficio de, un enfoque de "consenso funcional" basándose en el uso de una base de datos de secuencias de scFv seleccionadas funcionalmente. Todavía adicionalmente, la invención proporciona métodos de modificación por ingeniería genética de agentes de unión inmunológica (por ejemplo, scFv) mutando posiciones de aminoácido de región de marco particulares identificadas usando el enfoque de "consenso funcional" descrito en el presente documento.

Para que la invención pueda entenderse más fácilmente, se definen en primer lugar determinados términos. A menos que se defina de otro modo, todos los términos científicos y técnicos usados en el presente documento tienen el mismo significado que entiende comúnmente un experto habitual en la técnica a la que pertenece esta invención. Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o las pruebas de la invención, se describen a continuación métodos y materiales adecuados.

El término “anticuerpo” tal como se usa en el presente documento es un sinónimo para “inmunoglobulina”. Los anticuerpos según la presente invención pueden ser inmunoglobulinas completas o fragmentos de las mismas, que comprenden al menos un dominio variable de una inmunoglobulina, tal como dominios variables únicos, Fv (Skerra A. y Plücker, A. (1988) *Science* 240:1038-41), scFv (Bird, R.E. *et al.* (1988) *Science* 242:423-26; Huston, J.S. *et al.* (1988; *Proc. Natl. Acad. Sci. USA* 85:5879-83), Fab, (Fab')₂ u otros fragmentos bien conocidos por un experto en la técnica.

El término “región de marco de anticuerpo” tal como se usa en el presente documento se refiere a la parte del dominio variable, o bien V_L o bien V_H, que sirve como armazón para los bucles de unión a antígeno de este dominio variable (Kabat, E. A. *et al.*, (1991) *Sequences of proteins of immunological interest*. Publicación del NIH 91-3242).

El término “CDR de anticuerpo” tal como se usa en el presente documento se refiere a las regiones determinantes de complementariedad del anticuerpo que consisten en los bucles de unión a antígeno tal como se definen por Kabat E.A. *et al.*, (1991) *Sequences of proteins of immunological interest*. Publicación del NIH 91-3242). Cada uno de los dos dominios variables de un fragmento Fv de anticuerpo contiene, por ejemplo, tres CDR.

El término “anticuerpo de cadena sencilla” o “scFv” pretende referirse a una molécula que comprende una región variable de cadena pesada (V_H) de anticuerpo y una región variable de cadena ligera (V_L) de anticuerpo conectadas mediante un ligador. Tales moléculas de scFv pueden tener las estructuras generales: NH₂-V_L-ligador-V_H-COOH o NH₂-V_H-ligador-V_L-COOH.

Tal como se usa en el presente documento, “identidad” se refiere a la coincidencia de secuencia entre dos polipéptidos, moléculas o entre dos ácidos nucleicos. Cuando una posición en ambas de las secuencias comparadas está ocupada por la misma subunidad de monómero de aminoácido o base (por ejemplo, si una posición en cada uno de dos polipéptidos está ocupada por una lisina), entonces las respectivas moléculas son idénticas en esa posición. El “porcentaje de identidad” entre dos secuencias es una función del número de posiciones coincidentes compartidas por las dos secuencias dividido entre el número de posiciones comparadas x 100. Por ejemplo, si 6 de 10 de las posiciones en dos secuencias coinciden, entonces las dos secuencias tienen una identidad del 60%. Generalmente, se hace una comparación cuando dos secuencias se alinean para dar una identidad máxima. Tal alineación puede proporcionarse usando, por ejemplo, el método de Needleman *et al.* (1970) *J. Mol. Biol.* 48: 443-453, implementado convenientemente mediante programas informáticos tales como el programa Align (DNASTar, Inc.).

Secuencias “similares” son las que, cuando se alinean, comparten residuos de aminoácido idénticos y similares, donde residuos similares son sustituciones conservativas para residuos de aminoácido correspondientes en una secuencia de referencia alineada. En este sentido, una “sustitución conservativa” de un residuo en una secuencia de referencia es una sustitución por un residuo que es física o funcionalmente similar al residuo de referencia correspondiente, por ejemplo, que tiene un tamaño, forma, carga eléctrica, propiedades químicas similares, incluyendo la capacidad para formar enlaces de hidrógeno o covalentes, o similares. Por tanto, una secuencia “modificada mediante sustitución conservativa” es una que difiere de una secuencia de referencia o una secuencia silvestre en que están presentes una o más sustituciones conservativas. El “porcentaje de similitud” entre dos secuencias es una función del número de posiciones que contienen residuos coincidentes o sustituciones conservativas compartidos por las dos secuencias dividido entre el número de posiciones comparadas x 100. Por ejemplo, si 6 de 10 de las posiciones en dos secuencias coinciden y 2 de 10 posiciones contienen sustituciones conservativas, entonces las dos secuencias tienen una similitud positiva del 80%.

“Secuencia de aminoácidos consenso” tal como se usa en el presente documento se refiere a una secuencia de aminoácidos que puede generarse usando una de al menos dos, y preferiblemente más, secuencias de aminoácidos alineadas, y permitiendo huecos en la alineación, de manera que es posible determinar el residuo de aminoácido más frecuente en cada posición. La secuencia consenso es la secuencia que comprende los aminoácidos que se representan lo más frecuentemente en cada posición. En el caso de que dos o más aminoácidos estén representados por igual en una única posición, la secuencia consenso incluye ambos o todos de esos aminoácidos.

La secuencia de aminoácidos de una proteína puede analizarse a diversos niveles. Por ejemplo, puede presentarse conservación o variabilidad a nivel de residuo individual, a nivel de múltiples residuos, múltiples residuos con huecos, etc. Los ejemplos de clases de aminoácidos incluyen grupos R polares pero no cargados (serina, treonina, asparagina y glutamina); grupos R cargados positivamente (lisina, arginina e histidina); grupos R cargados negativamente (ácido glutámico y ácido aspártico); grupos R hidrófobos (alanina, isoleucina, leucina, metionina, fenilalanina, triptófano, valina y tirosina); y aminoácidos especiales (cisteína, glicina y prolina). Un experto en la técnica conoce otras clases y pueden definirse usando determinaciones estructurales u otros datos para evaluar la

capacidad de sustitución. En ese sentido, un aminoácido sustituible puede referirse a cualquier aminoácido que puede sustituirse y mantiene la conservación funcional en esa posición.

Tal como se usa en el presente documento, cuando una secuencia de aminoácidos (por ejemplo, una primera secuencia de V_H o V_L) se alinea con una o más secuencias de aminoácidos adicionales (por ejemplo, una o más secuencias de V_H o V_L en una base de datos), puede compararse una posición de aminoácido en una secuencia (por ejemplo, la primera secuencia de V_H o V_L) con una "posición correspondiente" en la una o más secuencias de aminoácidos adicionales. Tal como se usa en el presente documento, la "posición correspondiente" representa la posición equivalente en la(s) secuencia(s) que está(n) comparándose cuando las secuencias se alinean de manera óptima, es decir, cuando las secuencias se alinean para lograr la mayor identidad en tanto por ciento o similitud en tanto por ciento.

Tal como se usa en el presente documento, el término "base de datos de anticuerpos" se refiere a una colección de dos o más secuencias de aminoácidos de anticuerpos (una "multiplicidad" de secuencias), y normalmente se refiere a una colección de decenas, centenares o incluso miles de secuencias de aminoácidos de anticuerpos. Una base de datos de anticuerpos puede almacenar secuencias de aminoácidos de, por ejemplo, una colección de regiones V_H de anticuerpos, regiones V_L de anticuerpos o ambas, o puede almacenar una colección de secuencias de scFv compuestas por regiones V_H y V_L . Preferiblemente, la base de datos se almacena en un medio fijo, en el que pueden realizarse búsquedas, tal como en un ordenador dentro de un programa informático en el que pueden realizarse búsquedas. En una realización, la base de datos de anticuerpos es una base de datos que comprende o consiste en secuencias de anticuerpos de línea germinal. En otra realización, la base de datos de anticuerpos es una base de datos que comprende o consiste en secuencias de anticuerpos maduros (es decir, expresados) (por ejemplo, una base de datos Kabat de secuencias de anticuerpos maduros, por ejemplo, una base de datos KBD). En aún otra realización, la base de datos de anticuerpos comprende o consiste en secuencias seleccionadas funcionalmente (por ejemplo, secuencias seleccionadas a partir de un ensayo de QC).

El término "agente de unión inmunológica" se refiere a una molécula que contiene la totalidad o una parte del sitio de unión a antígeno de un anticuerpo, por ejemplo, la totalidad o parte del dominio variable de cadena pesada y/o ligera, de manera que el agente de unión inmunológica reconoce específicamente un antígeno diana. Los ejemplos no limitativos de agentes de unión inmunológica incluyen scFv y moléculas de inmunoglobulina de longitud completa, así como fragmentos de anticuerpo, incluyendo pero sin limitarse a (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios V_L , V_H , C_L y C_H1 ; (ii) un fragmento $F(ab')_2$, un fragmento bivalente que comprende dos fragmentos Fab unidos por un puente disulfuro en la región bisagra; (iii) un fragmento Fab', que es esencialmente un Fab con parte de la región bisagra (véase, FUNDAMENTAL IMMUNOLOGY (Paul ed., 3ª ed. 1993); (iv) un fragmento Fd que consiste en los dominios V_H y C_H1 ; (v) un anticuerpo con un solo dominio tal como un fragmento Dab (Ward *et al.*, (1989) Nature 341:544-546), que consiste en un dominio V_H o V_L , un anticuerpo de camélido (véase Hamers-Casterman, *et al.*, Nature 363:446-448 (1993), y Dumoulin, *et al.*, Protein Science 11:500-515 (2002)) o de tiburón (por ejemplo, Nanobodies® de Ig-NAR de tiburón; y (vii) un Nanobody, una región variable de cadena pesada que contiene un único dominio variable y dos dominios constantes.

Tal como se usa en el presente documento, el término "propiedad funcional" es una propiedad de un polipéptido (por ejemplo, un agente de unión inmunológica) para el que es deseable y/o ventajosa una mejora (por ejemplo, en relación con un convencional) para un experto en la técnica, por ejemplo, con el fin de mejorar las propiedades de fabricación o la eficacia terapéutica del polipéptido. En una realización, la propiedad funcional es estabilidad mejorada (por ejemplo, estabilidad térmica). En otra realización, la propiedad funcional es solubilidad mejorada (por ejemplo, en condiciones celulares). En aún otra realización, la propiedad funcional es no agregación. En todavía otra realización, la propiedad funcional es una mejora en la expresión (por ejemplo, en una célula procarionota). En aún otra realización, la propiedad funcional es una mejora en el rendimiento de replegamiento tras un procedimiento de purificación de cuerpos de inclusión. En determinadas realizaciones, la propiedad funcional no es una mejora en la afinidad de unión a antígeno.

Análisis basado en secuencia de scFv

La invención proporciona métodos para analizar una secuencia de scFv que permiten la identificación de posiciones de aminoácido dentro de la secuencia de scFv que van a seleccionarse para mutación. Las posiciones de aminoácido seleccionadas para mutación son las que se predice que influyen en las propiedades funcionales del scFv, tales como solubilidad, estabilidad y/o unión a antígeno, en los que la mutación en tales posiciones se predice que mejora las prestaciones del scFv. Por tanto, la invención permite una modificación por ingeniería genética de scFv para optimizar las prestaciones más focalmente que simplemente mutar al azar posiciones de aminoácido dentro de la secuencia de scFv.

Determinados aspectos del análisis basado en secuencia de secuencias de scFv se representan de manera esquemática en el diagrama de flujo de la figura 1. Tal como se muestra en esta figura, se compara la secuencia de un scFv que va a optimizarse con las secuencias en una o más bases de datos de anticuerpos, incluyendo una base de datos de anticuerpos compuesta por secuencias de scFv seleccionadas como estables y solubles. Esto puede permitir la identificación de residuos críticos para la estabilidad y/o solubilidad específicamente en el formato de scFv, así como la identificación de patrones que representan mejoras en la estabilidad, solubilidad y/o unión

independientes de las respectivas CDR, específicamente en el formato de scFv (por ejemplo, combinaciones de V_L y V_H). Una vez que se han identificado residuos críticos, pueden sustituirse, por ejemplo, por el aminoácido más frecuente tal como se identifica en la respectiva base de datos y/o mediante mutagénesis al azar o sesgada.

5 Por tanto, en un aspecto, la invención se refiere a un método de identificación de una posición de aminoácido para mutación en un anticuerpo de cadena sencilla (scFv), teniendo el scFv secuencias de aminoácidos de V_H y V_L , comprendiendo el método:

10 a) introducir las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv en una base de datos que comprende una multiplicidad de secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de manera que las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv se alinean con las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos;

b) comparar una posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv con una posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos;

15 c) determinar si la posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv está ocupada por un residuo de aminoácido que está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

d) identificar la posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando la posición de aminoácido está ocupada por un residuo de aminoácido que no está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

20 Por tanto, en el método de la invención, la secuencia de un scFv de interés (es decir, la secuencia de la V_H , V_L o ambas) se compara con las secuencias de una base de datos de anticuerpos y se determina si una posición de aminoácido en el scFv de interés está ocupada por un residuo de aminoácido que está "conservado" en la posición correspondiente de las secuencias en la base de datos. Si la posición de aminoácido de la secuencia de scFv está ocupada por un residuo de aminoácido que no está "conservado" en la posición correspondiente dentro de las secuencias de la base de datos, entonces esa posición de aminoácido del scFv se elige para mutación. Preferiblemente, la posición de aminoácido que se analiza es una posición de aminoácido de región de marco dentro del scFv de interés. Incluso más preferiblemente, puede analizarse cada posición de aminoácido de región de marco dentro del scFv de interés. En una realización alternativa, pueden analizarse una o más posiciones de aminoácido dentro de una o más CDR del scFv de interés. En aún otra realización, puede analizarse cada posición de aminoácido con el scFv de interés.

35 Para determinar si un residuo de aminoácido está "conservado" en una posición de aminoácido particular dentro de las secuencias de la base de datos de anticuerpos (por ejemplo, una posición de región de marco), puede calcularse el grado de conservación en la posición particular. Existe una variedad de diferentes modos conocidos en la técnica en los que puede cuantificarse la diversidad de aminoácidos en una posición dada, todos los cuales pueden aplicarse a los métodos de la presente invención. Preferiblemente, el grado de conservación se calcula usando el índice de diversidad de Simpson, que es una medida de diversidad. Tiene en cuenta el número de aminoácidos presentes en cada posición, así como la abundancia relativa de cada aminoácido. El índice de Simpson (I.S.) representa la probabilidad de que dos secuencias de anticuerpo seleccionadas al azar contengan el mismo aminoácido en determinadas posiciones. El índice de Simpson tiene en cuenta dos factores principales cuando se mide la conservación, riqueza y uniformidad. Tal como se usa en el presente documento, la "riqueza" es una medida del número de clases diferentes de aminoácidos presentes en una posición particular (es decir, el número de residuos de aminoácido diferentes representados en la base de datos en esa posición es una medida de la riqueza). Tal como se usa en el presente documento, la "uniformidad" es una medida de la abundancia de cada uno de los aminoácidos presentes en la posición particular (es decir, la frecuencia con la que aparecen los residuos de aminoácido en esa posición dentro de las secuencias de la base de datos es una medida de la uniformidad).

45 Aunque puede usarse la riqueza de residuos como una medida en sí misma para examinar el grado de conservación en una posición particular, no tiene en cuenta la frecuencia relativa de cada residuo de aminoácido presente en una determinada posición. Da tanto peso a los residuos de aminoácido que aparecen muy infrecuentemente en una posición particular dentro de las secuencias de una base de datos como a los residuos que aparecen muy frecuentemente en la misma posición. La uniformidad es una medida de la abundancia relativa de los diferentes aminoácidos que constituyen la riqueza de una posición. El índice de Simpson tiene en cuenta ambas, la riqueza y la uniformidad, y por tanto es un modo preferido para cuantificar el grado de conservación según la presente invención. En particular, residuos poco frecuentes en cada posición conservada se consideran posiblemente problemáticos y por tanto pueden elegirse para mutación.

55 La fórmula para el índice de Simpson es $D = \sum n_i (n_i - 1) / (N(N - 1))$, en la que N es el número total de secuencias en la investigación (por ejemplo, en la base de datos) y n_i es la frecuencia de cada residuo de aminoácido en la posición que está analizándose. La frecuencia de un acontecimiento de aminoácido (i) en la base de datos es el número (n_i) de veces que aparece el aminoácido en la base de datos. Los propios recuentos n_i se facilitan en frecuencias

relativas, lo que significa que se normalizan mediante el número total de acontecimientos. Cuando se produce diversidad máxima, el valor de I.S. es cero y cuando se produce diversidad mínima, el valor de I.S. es 1. Por tanto, el intervalo de I.S. es de 0-1, con una relación inversa entre la diversidad y el valor del índice.

5 En la figura 2 se describe en detalle adicional un diagrama de flujo que resume las múltiples etapas para el análisis de posiciones de aminoácido de región de marco dentro de las secuencias de la base de datos.

Por consiguiente, en una realización preferida del método descrito anteriormente, a la posición correspondiente dentro de la secuencia de aminoácidos de V_H o V_L de anticuerpos de la base de datos se le asigna un grado de conservación usando el índice de Simpson. El valor de I.S. de esa posición correspondiente puede usarse como indicador de la conservación de esa posición.

10 En otras realizaciones, se usan en la presente invención alineaciones de confianza (es decir, alineaciones de secuencias para las que se considera similitud de la estructura de la proteína) de secuencias de anticuerpos estrechamente relacionadas para generar matrices de abundancia relativa de aminoácidos y grado de conservación de determinadas posiciones. Estas matrices se diseñan para su uso en comparaciones de bases de datos de anticuerpo-anticuerpo. Se calcula la frecuencia observada de cada residuo y se compara con las frecuencias esperadas (que son esencialmente las frecuencias de cada residuo en el conjunto de datos para cada posición).

15 El análisis de un anticuerpo scFv dado con el método descrito proporciona información sobre mutaciones admisibles biológicamente y residuos no habituales en determinadas posiciones en el anticuerpo scFv dado y permite la predicción de una posible debilidad dentro de su región de marco. Puede usarse la rutina para diseñar por ingeniería genética sustituciones de aminoácidos que se ajustan "de la mejor manera" a un conjunto de datos de frecuencia de aminoácidos, usando el valor de I.S. y la frecuencia relativa como criterio.

20 El análisis basado en secuencia descrito anteriormente puede aplicarse a la región V_H del scFv, a la región V_L del scFv o a ambas. Por tanto, en una realización, se introduce la secuencia de aminoácidos de V_H de scFv en la base de datos y se alinea con secuencias de aminoácidos de V_H de anticuerpos de la base de datos. En otra realización, se introduce la secuencia de aminoácidos de V_L de scFv en la base de datos y se alinea con secuencias de aminoácidos de V_L de anticuerpos de la base de datos. En aún otra realización, se introducen las secuencias de aminoácidos de V_H y V_L de scFv en la base de datos y se alinean con secuencias de aminoácidos de V_H y V_L de anticuerpos de la base de datos. Los algoritmos para alinear una secuencia con una colección de otras secuencias en una base de datos están bien establecidos en la técnica. Las secuencias se alinean de manera que se logre la mayor similitud o identidad en tanto por ciento entre las secuencias.

30 Los métodos de la invención pueden usarse para analizar una posición de aminoácido de interés dentro de una secuencia de scFv o, más preferiblemente, pueden usarse para analizar múltiples posiciones de aminoácido de interés. Por tanto, en la etapa b) del método descrito anteriormente, pueden compararse múltiples posiciones de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv con posiciones correspondientes dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Posiciones preferidas que van a analizarse son posiciones de región de marco dentro de las secuencias de V_H y/o V_L del scFv (por ejemplo, puede analizarse cada posición de región de marco de V_H y V_L). Adicional o alternativamente, puede analizarse una o más posiciones dentro de una o más CDR del scFv (aunque puede no preferirse mutar posiciones de aminoácido con las CDR, puesto que es más probable que mutaciones dentro de las CDR afecten a la actividad de unión a antígeno que mutaciones dentro de las regiones de marco). Todavía adicionalmente, los métodos de la invención permiten el análisis de cada posición de aminoácido dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv.

45 En los métodos de la invención, puede compararse la secuencia de un scFv de interés con las secuencias dentro de uno o más de una variedad de diferentes tipos de bases de datos de secuencias de anticuerpos. Por ejemplo, en una realización, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de línea germinal. En otra realización, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos reordenados, madurados por afinidad. En aún otra realización preferida, las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos de la base de datos son secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv seleccionadas por tener al menos una propiedad funcional deseable, tal como estabilidad de scFv o solubilidad de scFv (comentado adicionalmente a continuación).

50 Puede obtenerse, compilarse y/o generarse información de secuencias de anticuerpos a partir de alineaciones de secuencias de secuencias de línea germinal o a partir de cualquier otra secuencia de anticuerpo que se produzca en la naturaleza. Las fuentes de secuencias pueden incluir, pero no se limitan a, una o más de las siguientes bases de datos

55 - La base de datos Kabat (immuno.bme.nwu.edu) (a fecha de octubre de 2007); Johnson y Wu (2001) *Nucleic Acids Res.* 29: 205-206; Johnson y Wu (2000) *Nucleic Acids Res.* 28: 214-218). Los datos sin procesar de 2000 están disponibles mediante FTP en los EE.UU. y un sitio espejo en el R.U.

- Kabatman contiene una base de datos que permite al usuario buscar, en las secuencias Kabat, características no habituales de secuencias y permite al usuario encontrar asignaciones canónicas para las CDR en una secuencia de

anticuerpo específica.

- El sitio web AAAAA (www.bioc.unizh.ch/antibody/ (a fecha de octubre de 2007)), una página sobre anticuerpos preparada por Annemarie Honegger que proporciona información de secuencia y datos estructurales sobre anticuerpos.

5 - ABG: Directorio de estructuras 3D de anticuerpos - el directorio, creado por el Antibody Group (ABG), permite al usuario acceder a las estructuras de anticuerpos compiladas en el Banco de Datos de Proteínas (PDB, *Protein Data Bank*). En el directorio, cada entrada de PDB tiene un hipervínculo a la fuente original para producir información completa que se recupera fácilmente

10 - ABG: Directorios de genes de línea germinal de los segmentos de línea germinal VH y VK, parte de la página web del Antibody Group en el Instituto de Biotecnología, UNAM (Universidad Nacional de México)

15 - IMGT®, the international ImMunoGeneTics information system® - creado en 1989 por Marie-Paule Lefranc (Université Montpellier II, CNRS), IMGT es un recurso de conocimiento integrado especializado en inmunoglobulinas, receptores de células T y proteínas relacionadas del sistema inmunitario para seres humanos y otras especies de vertebrados. IMGT consiste en bases de datos de secuencias (IMGT/LIGM-DB, una base de datos exhaustiva de IG y TR de seres humanos y otros vertebrados, con traducción para secuencias completamente anotadas, IMGT/MHC-DB, IMGT/PRIMER-DB), una base de datos de genoma (IMGT/GENE-DB), una base de datos de estructuras (IMGT/3Dstructure-DB), un recurso web (repertorio de IMGT) (IMGT, the international ImMunoGeneTics information system®; imgt.cines.fr (a fecha de octubre de 2007); Lefranc *et al.* (1999) *Nucleic Acids Res.* 27: 209-212; Ruiz *et al.* (2000) *Nucleic Acids Res.* 28: 219-221; Lefranc *et al.* (2001) *Nucleic Acids Res.* 29: 207-209; Lefranc *et al.* (2003) *Nucleic Acids Res.* 31: 307-310).

20 - V BASE - un directorio exhaustivo de todas las secuencias de región variable de línea germinal humanas compiladas a partir de más de mil secuencias publicadas, incluyendo aquéllas en las versiones actuales de las bibliotecas de datos Genbank y EMBL.

25 En una realización preferida, la información de secuencias de anticuerpos se obtiene a partir de una biblioteca de scFv que tiene regiones de marco definidas que se han seleccionado por su estabilidad y solubilidad potenciadas en un entorno reductor. Más específicamente, se ha descrito un sistema de control de calidad (QC) de levaduras (véase por ejemplo, la publicación PCT WO 2001/48017; las solicitudes estadounidenses n.ºs 2001/0024831 y US 2003/0096306; las patentes estadounidenses n.ºs 7.258.985 y 7.258.986) que permite la selección intracelular de regiones de marco de scFv con estabilidad y solubilidad potenciadas en un entorno reductor. En este sistema, se transforma una biblioteca de scFv en células huésped que pueden expresar un antígeno conocido específico y que sólo sobreviven en presencia de interacción antígeno-scFv. Las células huésped transformadas se cultivan en condiciones adecuadas para la expresión del antígeno y el scFv y que permiten la supervivencia celular sólo en presencia de interacción antígeno-scFv. Por tanto, pueden aislarse los scFv expresados en las células supervivientes y que tienen regiones de marco definidas que son estables y solubles en un entorno reductor. Por consiguiente, el sistema de QC puede usarse para examinar una biblioteca de scFv grande para aislar de ese modo los scFv preferidos que tienen regiones de marco que son estables y solubles en un entorno reductor y las secuencias de esos scFv seleccionados pueden compilarse en una base de datos de secuencias de scFv. Entonces puede usarse una base de datos de scFv de este tipo para fines de comparación con otras secuencias de scFv de interés usando los métodos de la presente invención. Se describen en detalle adicional secuencias de región de marco de scFv preferidas que se han seleccionado y definido previamente usando el sistema de QC en la publicación PCT WO 2003/097697 y la solicitud estadounidense n.º 20060035320.

35 Se conocen en la técnica variantes del sistema de QC original. En una realización a modo de ejemplo, que se ilustra esquemáticamente en la figura 3, se fusiona una biblioteca de scFv con el dominio de activación (AD) del factor de transcripción de levaduras Gal4, que a su vez se fusiona con una parte de la denominada proteína Gal11p (11p). Entonces se transforma el constructo de fusión scFv-AD-Gal11p en células huésped que expresan los primeros 100 aminoácidos de Gal 4 y por tanto contienen el dominio de unión a ADN de Gal4 (DBD; Gal4(1-100)). Gal11p es una mutación puntual que se sabe que se une directamente a Gal4(1-100) (véase Barberis *et al.*, *Cell*, 81: 359 (1995)). Las células huésped transformadas se cultivan en condiciones que son adecuadas para la expresión de la proteína de fusión de scFv y que permiten la supervivencia celular sólo en el caso de que la proteína de fusión de scFv sea lo suficientemente estable y soluble como para interaccionar con Gal4(1-100) y de ese modo formar un factor de transcripción funcional que contiene un AD unido a un DBD (figura 3A). Por tanto, pueden aislarse los scFv expresados en las células supervivientes y que tienen regiones de marco definidas que son estables y solubles en un entorno reductor. Se describe una descripción adicional de este sistema de QC a modo de ejemplo en Auf der Maur *et al.*, *Methods*, 34: 215-224 (2004).

55 En otra realización a modo de ejemplo, se representa en la figura 4 un sistema de QC empleado en los métodos de la invención. En esta versión del sistema de QC, se fusiona directamente el scFv o la biblioteca de scFv con un factor de transcripción funcional y se expresa en una cepa de levaduras que contiene un marcador seleccionable. El marcador seleccionable sólo se activará en presencia de una fusión de factor de transcripción-scFv funcional, lo que significa que el constructo en su totalidad necesita ser estable y soluble (figura 4A). En el caso de que el scFv sea

inestable, formará agregados y finalmente se degradará, provocando también de ese modo la degradación del factor de transcripción fusionado con el mismo, de modo que ya no puede activar la expresión del marcador seleccionable (véase la figura 4B).

5 En los métodos de la invención, la secuencia de un scFv de interés puede compararse con todas las secuencias dentro de una base de datos de anticuerpos o, alternativamente, sólo una parte seleccionada de las secuencias en la base de datos puede usarse para fines de comparación. Es decir, la base de datos puede estar limitada, o restringida, a sólo las secuencias que tienen un alto porcentaje de similitud o identidad con el scFv de interés. Por tanto, en una realización del método de la invención, la base de datos es una base de datos restringida en la que sólo se incluyen en la base de datos las secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos que tienen alta similitud con la secuencia de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv.

10 Una vez que la secuencia de scFv de interés se introduce en la base de datos y se compara con las secuencias de anticuerpos dentro de la base de datos, se analiza la información de secuencia para proporcionar información sobre la frecuencia y variabilidad de aminoácidos de una posición dada y para predecir posiciones de aminoácido posiblemente problemáticas, en particular posiciones de aminoácido posiblemente problemáticas dentro de la región de marco del scFv. Tal información puede usarse también para diseñar mutaciones que mejoran las propiedades del scFv. Por ejemplo, puede mejorarse la solubilidad del anticuerpo reemplazando residuos hidrófobos expuestos al disolvente por residuos hidrófilos que por lo demás aparecen frecuentemente en esta posición.

15 En el método de la invención, hay varios tipos posibles de residuos de aminoácido que pueden estar "conservados" en una posición particular dentro de las secuencias de anticuerpos de la base de datos. Por ejemplo, puede encontrarse un residuo de aminoácido particular en esa posición a una frecuencia muy alta, indicando que este residuo de aminoácido particular se prefiere en esa posición particular. Por consiguiente, en una realización del método, en la etapa c), el residuo de aminoácido que está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos es el residuo de aminoácido que está lo más frecuentemente en esa posición dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En otras realizaciones, la posición puede estar "conservada" con un tipo o clase particular de residuo de aminoácido (es decir, la posición no está ocupada preferentemente por sólo un único residuo de aminoácido particular, sino que más bien está ocupada preferentemente por varios residuos de aminoácido diferentes cada uno de los cuales es del mismo tipo o clase de residuo). Por ejemplo, en la etapa c), la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos puede estar conservada con: (i) 20 residuos de aminoácido hidrófobos, (ii) residuos de aminoácido hidrófilos, (iii) residuos de aminoácido que pueden formar un enlace de hidrógeno o (iv) residuos de aminoácido que tienen una propensión a formar una lámina β .

25 En la etapa d) del método, una posición de aminoácido dentro de la secuencia de aminoácidos de V_H o V_L de scFv se identifica como una posición de aminoácido para mutación cuando la posición de aminoácido está ocupada por un residuo de aminoácido que no está conservado en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. Hay varias situaciones posibles que identificarían que una posición de aminoácido está ocupada por un residuo de aminoácido que "no está conservado" y por tanto que es posiblemente problemático. Por ejemplo, si la posición de aminoácido correspondiente dentro de la base de datos está conservada con un residuo hidrófobo y la posición en el scFv está ocupada por un residuo hidrófilo, esta posición podría ser posiblemente problemática en el scFv y la posición puede seleccionarse para mutación. 35 Asimismo, si la posición de aminoácido correspondiente dentro de la base de datos está conservada con un residuo hidrófilo y la posición en el scFv está ocupada por un residuo hidrófobo, esta posición podría ser posiblemente problemática en el scFv y la posición puede seleccionarse para mutación. Por tanto, en una realización preferida de los métodos dados a conocer en el presente documento, se sustituye un residuo de aminoácido hidrófobo por un residuo hidrófilo o viceversa. En todavía otros casos, si la posición de aminoácido correspondiente dentro de la base de datos está conservada con residuos de aminoácido que pueden formar un enlace de hidrógeno o que tienen una propensión a formar una lámina β , y la posición en el scFv está ocupada por un residuo que no puede formar un enlace de hidrógeno o que no tiene una propensión a formar una lámina β , respectivamente, esta posición podría ser posiblemente problemática en el scFv y la posición puede seleccionarse para mutación. Por tanto, en otra 40 realización preferida de los métodos de la presente invención, se sustituye un residuo de aminoácido que puede formar un enlace de hidrógeno por un residuo de aminoácido que tiene una propensión a formar una lámina β , o viceversa.

45 En una realización preferida, los métodos descritos en la presente invención pueden usarse solos o en combinación para crear listas combinatorias de sustituciones de aminoácidos para mejorar la estabilidad y/o solubilidad de fragmentos de cadena sencilla de anticuerpos.

55 Análisis de covarianza

La invención también se refiere a métodos para analizar la covarianza dentro de la secuencia de un scFv en comparación con secuencias de anticuerpos dentro de una base de datos. Residuos que pueden variar conjuntamente pueden ser, por ejemplo, (i) un residuo en una región de marco (FR) y un residuo en una CDR; (ii) un residuo en una CDR y un residuo en otra CDR; (iii) un residuo en una FR y un residuo en otra FR; o (iv) un residuo 60 en la V_H y un residuo en la V_L . Los residuos que interactúan entre sí en la estructura terciaria del anticuerpo

pueden variar conjuntamente de manera que residuos de aminoácido preferidos pueden estar conservados en ambas posiciones del par covariante y si un residuo está alterado, el otro residuo debe estar alterado también para mantener la estructura del anticuerpo. Se conocen en la técnica métodos para realizar un análisis de covarianza en un conjunto de secuencias de aminoácidos. Por ejemplo, Choulier, L. *et al.* (2000) *Protein* 41:475-484 describen la aplicación de un análisis de covarianza a alineaciones de secuencias de V_K y V_H de línea germinal humanas y de ratón.

Un análisis de covarianza puede combinarse con el método descrito anteriormente para analizar posiciones de aminoácido conservadas (etapas a)-d) en el método anterior), de manera que el método comprende además las etapas de:

5 e) llevar a cabo un análisis de covarianza en la secuencia de aminoácidos de V_H o V_L de anticuerpos de la base de datos para identificar un par covariante de posiciones de aminoácido;

f) comparar el par covariante de posiciones de aminoácido con posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv;

15 g) determinar si las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv están ocupadas por residuos de aminoácido que están conservados en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

20 h) identificar una o ambas de las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando una o ambas de las posiciones correspondientes dentro del scFv está ocupada por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

Adicional o alternativamente, puede realizarse un análisis de covarianza por sí mismo, de manera que la invención se refiere a un método que comprende las etapas de:

a) llevar a cabo un análisis de covarianza en secuencias de aminoácidos de V_H o V_L de anticuerpos de una base de datos para identificar un par covariante de posiciones de aminoácido;

25 b) comparar el par covariante de posiciones de aminoácido con posiciones correspondientes dentro de una secuencia de aminoácidos de V_H o V_L de scFv;

c) determinar si las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv están ocupadas por residuos de aminoácido que están conservados en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos; y

30 d) identificar una o ambas de las posiciones correspondientes dentro de la secuencia de aminoácidos de V_H o V_L de scFv como una posición de aminoácido para mutación cuando una o ambas de las posiciones correspondientes dentro del scFv están ocupadas por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

35 Los métodos de análisis de covarianza de la divulgación pueden usarse para analizar un par covariante, o más de un par covariante. Por tanto, en un aspecto del método, se identifican múltiples pares covariantes de posiciones de aminoácido dentro de la secuencia de aminoácidos de V_H y/o V_L de anticuerpos de la base de datos y se comparan con las posiciones correspondientes dentro de la secuencia de aminoácidos V_H y/o V_L de scFv

40 El método puede comprender además mutar una o ambas de las posiciones correspondientes dentro del scFv que están ocupadas por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En un aspecto, una de las posiciones correspondientes dentro del scFv que está ocupada por un residuo de aminoácido que no está conservado en el par covariante de posiciones de aminoácido se sustituye por un residuo de aminoácido que está lo más frecuentemente en la posición de aminoácido de par covariante. En otro aspecto, ambas de las posiciones correspondientes dentro del scFv que están ocupadas por residuos de aminoácido que no están conservados en el par covariante de posiciones de aminoácido se sustituyen por residuos de aminoácido que están lo más frecuentemente en las posiciones de aminoácido de par covariante.

Modelado molecular

45 Los métodos basados en secuencia de la invención para analizar scFv para detectar residuos posiblemente problemáticos pueden combinarse con otros métodos conocidos en la técnica para analizar relaciones de estructura/función de anticuerpos. Por ejemplo, en una realización preferida, los métodos analíticos basados en secuencia de la invención se combinan con modelado molecular para identificar residuos posiblemente problemáticos adicionales. Métodos y software para el modelado informático de estructuras de anticuerpo, incluyendo estructuras de scFv, están establecidos en la técnica y pueden combinarse con los métodos basados en secuencia de la invención. Por tanto, en otra realización, los métodos basados en secuencia descritos anteriormente

tal como se exponen en las etapas a) - d) comprenden además las etapas de:

e) someter las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv a modelado molecular; y

f) identificar al menos una posición de aminoácido adicional dentro de las secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv para mutación.

- 5 El método puede comprender además mutar la al menos una posición de aminoácido adicional dentro de secuencias de aminoácidos de V_H , V_L o V_H y V_L de scFv identificadas para la mutación mediante modelado molecular.

Análisis de “consenso funcional” frente a “consenso convencional”

10 En una realización particularmente preferida, se compara el grado de variabilidad en una o más posiciones de región de marco entre una primera base de datos de secuencias de anticuerpos (por ejemplo, una(s) base(s) de datos de línea germinal (por ejemplo, Vbase y/o IMGT) o una base de datos de anticuerpos maduros (por ejemplo, KBD)) y una segunda base de datos de scFv seleccionados como que tienen una o más propiedades deseables, por ejemplo, una base de datos de scFv seleccionados mediante selección de QC en levaduras, es decir, una base de datos de QC. Tal como se ilustra en la figura 5, puede asignarse un valor de variabilidad (por ejemplo, valor de índice de Simpson) a posiciones de región de marco dentro de la primera base de datos (por ejemplo, de línea germinal), denominados valores “G” en la figura 5, y puede asignarse un valor de variabilidad (por ejemplo, valor de índice de Simpson) a las posiciones de región de marco correspondientes dentro de la segunda base de datos (por ejemplo, base de datos de QC), denominados valores “Q” en la figura 5. Cuando el valor G es mayor que el valor Q en una posición particular (es decir, más variabilidad en las secuencias de línea germinal en esa posición que en las secuencias seleccionadas de scFv), esto indica que hay un número restringido de residuos de aminoácido de región de marco de scFv estables en esa posición, residuos de aminoácido de región de marco de scFv estables que pueden ser adecuados para su uso con cualquier CDR. Alternativamente, cuando el valor G es menor que el valor Q en una posición particular (es decir, más variabilidad en las secuencias seleccionadas de scFv en esa posición que en las secuencias de línea germinal), esto indica que esa posición particular es más tolerante a la variabilidad en el scFv y por tanto puede representar una posición en la que sustituciones de aminoácido pueden optimizar la estabilidad y/o solubilidad del scFv. La tabla A presenta una tabla de resumen del número de posiciones de aminoácido, y residuos de región de marco altamente variables (hvFR), en los que o bien G es mayor que Q o bien G es menor que Q. Tal como se indica en la tabla A, la variabilidad en el número total de aminoácidos (n.º de aa) y en los residuos de región de marco altamente variables (hvFR) aumenta significativamente entre FW de línea germinal y de QC. Las secuencias que se analizaron para generar la tabla A fueron aproximadamente 90 secuencias de scFv que se seleccionaron usando el ensayo de QC (tal como se describe en el documento WO03097697; denominadas en el presente documento “Q”) y todas las secuencias de V_H y V_L de línea germinal recuperadas de <http://www.bioc.unizh.ch/antibody/Sequences/index.html> en octubre de 2007 (denominadas en el presente documento “G”). Para el análisis de la tabla A, los dominios de V_H y V_L no se agruparon según su subtipo.

Tabla A: Tabla de resumen

	N.º de aa	G < Q (n.º de casos)	G > Q (n.º de casos)	X/Y	N.º de hvFR (Simpson < 0,4)	G < Q (n.º de casos)	G > Q (n.º de casos)	X/Y
V_L	108	61	11	5,5	16	13	3	4,3
V_H	116	50	18	2,8	27	22	5	4,4

- 35 En vista de lo anterior, en aún otro aspecto, la invención se refiere a un método de identificación de una o más posiciones de aminoácido de región de marco para mutación en un anticuerpo de cadena sencilla (scFv), teniendo el scFv secuencias de aminoácidos de V_H y V_L , comprendiendo el método:

a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H , V_L o V_H y V_L (por ejemplo, secuencias de anticuerpos maduros y/o de línea germinal);

- 40 b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv seleccionadas como que tienen al menos una propiedad funcional deseable;

c) determinar variabilidad de aminoácidos en cada posición de región de marco de la primera base de datos y en cada posición de región de marco de la segunda base de datos;

- 45 d) identificar una o más posiciones de región de marco en las que el grado de variabilidad de aminoácidos difiere entre la primera base de datos y la segunda base de datos para así identificar una o más posiciones de aminoácido de región de marco para mutación en un anticuerpo de cadena sencilla (scFv).

Preferiblemente, la variabilidad de aminoácidos en cada posición de región de marco se determina evaluando un grado de conservación usando el índice de Simpson. En una realización, la una o más posiciones de aminoácido de región de marco se identifican para mutación basándose en la una o más posiciones de aminoácido de región de marco que tienen un valor de índice de Simpson inferior en la segunda base de datos (de scFv) en comparación con la primera base de datos. En otra realización, la una o más posiciones de aminoácido de región de marco se

- 50

identifican para mutación basándose en la una o más posiciones de aminoácido de región de marco que tienen un valor de índice de Simpson superior en la segunda base de datos en comparación con la primera base de datos (por ejemplo base de datos de línea germinal o una base de datos de anticuerpos maduros).

5 En los ejemplos 2 y 3 a continuación se describen con más detalle análisis de variabilidad, e identificación de residuos para mutación, para tres familias de V_H humana y tres familias de V_L humana.

Análisis de enriquecimiento / exclusión

10 En otro aspecto, la invención proporciona métodos para seleccionar sustituciones de residuos de aminoácido preferidas (o, alternativamente, excluir sustituciones de aminoácidos particulares) en una posición de región de marco de interés dentro de un agente de unión inmunológica (por ejemplo, para mejorar una propiedad funcional tal como estabilidad y/o solubilidad). Los métodos de la invención comparan la frecuencia de un residuo de aminoácido en una posición de región de marco de interés en una primera base de datos de secuencias de anticuerpos (por ejemplo, base(s) de datos de línea germinal tales como Vbase y/o IMGT o, más preferiblemente, una base de datos de anticuerpos maduros tal como la base de datos Kabat (KBD)) con la frecuencia del residuo de aminoácido en una posición de aminoácido correspondiente en una segunda base de datos de scFv seleccionados como que tienen una o más propiedades deseables, por ejemplo, una base de datos de scFv seleccionados mediante selección de QC en levaduras, es decir, una base de datos de QC.

20 Al contrario que el método establecido para mutar un residuo de aminoácido en una posición dada hacia un residuo de aminoácido de una secuencia consenso, se ha encontrado sorprendentemente que el enfoque de "consenso funcional" condujo a anticuerpos con propiedades funcionales deseables mejoradas. Esto puede deberse al hecho de que las posiciones altamente conservadas en la naturaleza que muestran un determinado grado de variabilidad en una región de marco seleccionada deben tolerar la mutagénesis al azar y presentar una probabilidad aumentada de hallar aminoácidos alternativos superiores al residuo nativo en un formato de scFv. Además, una preferencia pronunciada por un aminoácido no común es una indicación de selección natural hacia un determinado residuo. Basándose en estas dos directrices estadísticas, pueden elegirse diferentes residuos dentro de las cadenas pesada y ligera como o bien posiciones flotantes (tolerantes a la variabilidad) o bien sustituciones preferidas (residuos no habituales).

30 En el caso de que la frecuencia relativa de un determinado residuo de aminoácido aumenta en la segunda base de datos que comprende secuencias con la característica funcional deseable con respecto a la primera base de datos, es decir una(s) base(s) de datos de línea germinal y/o una base de datos de anticuerpos maduros, el residuo respectivo se considera un residuo preferido en la posición dada para mejorar la estabilidad y/o solubilidad de un scFv. A la inversa, en el caso de que la frecuencia relativa de un determinado residuo de aminoácido disminuye en la segunda base de datos en comparación con la primera base de datos, el residuo respectivo se considera desfavorable en esa posición en el contexto de un formato de scFv.

35 Tal como se describe en detalle en el ejemplo 4 a continuación, las secuencias de anticuerpos (por ejemplo, secuencias de V_H o V_L) de la primera base de datos (por ejemplo, una base de datos de secuencias de anticuerpos maduros) pueden agruparse según su subtipo de familia de Kabat (por ejemplo, Vh1b, VH3, etc.). Dentro de cada subtipo de secuencia (es decir, subfamilia), la frecuencia de cada residuo de aminoácido (por ejemplo, A, V, etc.) en cada posición de aminoácido se determina como un porcentaje de todas las secuencias analizadas de ese subtipo. Lo mismo se hace para todas las secuencias de la segunda base de datos (es decir, una base de datos de scFv seleccionados como que tienen una o más propiedades deseables, por ejemplo, mediante selección de QC). Para cada subtipo, se comparan los porcentajes resultantes (frecuencias relativas) para cada tipo de residuo de aminoácido en una posición particular entre las bases de datos primera y segunda. Cuando la frecuencia relativa de un determinado residuo de aminoácido aumenta en la segunda base de datos (por ejemplo, una base de datos de QC) con respecto a la primera base de datos (por ejemplo, base de datos de Kabat), esto indica que el residuo respectivo se selecciona favorablemente (es decir, un "residuo enriquecido") y confiere propiedades favorables a la secuencia. A la inversa, cuando la frecuencia relativa del residuo de aminoácido disminuye en la segunda base de datos con respecto a la primera base de datos, esto indica que el residuo respectivo se desfavorece (es decir, un "residuo excluido"). Por consiguiente, los residuos enriquecidos son residuos preferidos para mejorar las propiedades funcionales (por ejemplo, estabilidad y/o solubilidad) de un agente de unión inmunológica, mientras que preferiblemente se evitan los residuos excluidos.

50 En vista de lo anterior, en una realización, la invención se refiere a un método de identificación de un residuo de aminoácido para sustitución en una posición particular de un agente de unión inmunológica, comprendiendo el método:

55 a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H y/o V_L agrupadas (por ejemplo, secuencias de anticuerpos maduros y/o de línea germinal agrupadas según subtipo de familia de Kabat);

b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H o V_L de anticuerpos scFv agrupadas seleccionadas como que tienen al menos una propiedad funcional deseable (por ejemplo, según ensayo de QC);

c) determinar la frecuencia de aminoácido para un residuo de aminoácido en una posición de región de marco de la primera base de datos y en una posición de región de marco correspondiente de la segunda base de datos;

5 d) identificar el residuo de aminoácido como un residuo de aminoácido preferido para sustitución en una posición de aminoácido correspondiente del agente de unión inmunológica cuando el residuo de aminoácido se produce con una frecuencia superior en la segunda base de datos con respecto a la primera base de datos (es decir, un residuo enriquecido).

10 El enriquecimiento de un residuo de aminoácido en la segunda base de datos (de scFv) (por ejemplo, una base de datos de QC) puede cuantificarse. Por ejemplo, puede determinarse la razón entre la frecuencia relativa de un residuo dentro de la segunda base de datos (RF2) y la frecuencia relativa de un residuo dentro de la primera base de datos (RF1). Esta razón (RF2:RF1) puede denominarse “factor de enriquecimiento” (EF). Por consiguiente, en determinadas realizaciones, el residuo de aminoácido en la etapa (d) se identifica si la razón de la frecuencia relativa del residuo de aminoácido entre las bases de datos primera y segunda (en el presente documento, el “factor de enriquecimiento”) es de al menos 1 (por ejemplo, de 1 a 10). En una realización preferida, el factor de enriquecimiento es mayor que 1,0 o de aproximadamente 1,0 (por ejemplo 1,0, 1,1, 1,2, 1,3, 1,4 ó 1,5). En aún otra realización preferida, el factor de enriquecimiento es de aproximadamente 4,0 a aproximadamente 6,0 (por ejemplo, 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9, 5,0, 5,1, 5,2, 5,3, 5,4, 5,5, 5,6, 5,7, 5,8, 5,9 ó 6,0). En otra realización, el factor de enriquecimiento es de aproximadamente 6,0 a aproximadamente 8,0 (por ejemplo, 6,0, 6,1, 6,2, 6,3, 6,4, 6,5, 6,6, 6,7, 6,8, 6,9, 7,0, 7,1, 7,3, 7,4, 7,5, 7,6, 7,7, 7,8, 7,9 ó 8,0). En otras realizaciones, el factor de enriquecimiento es mayor que 10 (por ejemplo, 10, 100, 1000, 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 o más). En determinadas realizaciones, pueden alcanzarse factores de enriquecimiento infinitos.

20 En otro aspecto, la invención se refiere a un método de identificación de un residuo de aminoácido que va a excluirse en una posición particular de un agente de unión inmunológica, comprendiendo el método:

a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H y/o V_L agrupadas (por ejemplo, secuencias de anticuerpos maduros y/o de línea germinal agrupadas según subtipo de familia de Kabat);

25 b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H y/o V_L de anticuerpos scFv agrupadas seleccionadas como que tienen al menos una propiedad funcional deseable (por ejemplo, según ensayo de QC);

c) determinar la frecuencia de aminoácido para un residuo de aminoácido en una posición de región de marco de la primera base de datos y en una posición de región de marco correspondiente de la segunda base de datos;

30 d) identificar el residuo de aminoácido como un residuo de aminoácido desfavorecido para sustitución en la posición de aminoácido correspondiente del agente de unión inmunológica cuando el residuo de aminoácido se produce a una frecuencia inferior en la segunda base de datos con respecto a la primera base de datos, en el que dicho tipo de residuo de aminoácido es un residuo de aminoácido desfavorecido (es decir, un residuo excluido).

35 En determinadas realizaciones preferidas, el residuo de aminoácido desfavorecido en la etapa (d) anterior se identifica si el factor de enriquecimiento (EF) es menor que 1.

Además, la invención se refiere a un método de mejora de un agente de unión inmunológica, teniendo el agente de unión inmunológica secuencias de aminoácidos de V_H y/o V_L , comprendiendo el método:

a) identificar una o más posiciones de aminoácido de región de marco para mutación;

40 b) identificar para cada posición particular de región de marco identificada en la etapa a) un residuo de aminoácido preferido para sustitución; y

c) mutar el residuo de aminoácido en cada posición particular de región de marco hacia el residuo de aminoácido preferido identificado en la etapa b).

45 Preferiblemente, la etapa a) se realiza determinando el grado de variabilidad en una posición de región de marco dada según los métodos dados a conocer en el presente documento, es decir evaluando un grado de conservación en cada posición de región de marco usando el índice de Simpson. Un aminoácido adecuado para sustitución se identifica preferiblemente según los métodos dados a conocer anteriormente, que tiene un factor de enriquecimiento de al menos 1, más preferiblemente de al menos 4 a 6. Entonces se muta el residuo de aminoácido correspondiente del agente de unión inmunológica hacia el residuo de aminoácido preferido usando métodos de biología molecular conocidos en la técnica. El agente de unión inmunológica es preferiblemente un anticuerpo scFv, una inmunoglobulina de longitud completa, un fragmento Fab, un Dab o un Nanobody.

50

Mutación de scFv

En los métodos de la invención, una vez que se han identificado una o más posiciones de aminoácido dentro de un scFv como que son posiblemente problemáticas con respecto a las propiedades funcionales del scFv, el método puede comprender además mutar estas una o más posiciones de aminoácido dentro de la secuencia de

aminoácidos de V_H o V_L de scFv. Por ejemplo, una posición de aminoácido identificada para mutación puede sustituirse por un residuo de aminoácido que está conservado o enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

5 Una posición de aminoácido identificada para mutación puede mutarse usando uno de varios métodos de mutagénesis posibles bien establecidos en la técnica. Por ejemplo, puede usarse mutagénesis dirigida al sitio para realizar una sustitución de aminoácido particular en la posición de aminoácido de interés. También puede usarse mutagénesis dirigida al sitio para crear un conjunto de scFv mutados en los que se ha introducido un repertorio limitado de sustituciones de aminoácido en la posición de aminoácido de interés.

10 Adicional o alternativamente, la(s) posición/posiciones de aminoácido identificada(s) para mutación puede(n) mutarse mediante mutagénesis al azar o mediante mutagénesis sesgada para generar una biblioteca de scFv mutados, seguido por el examen de la biblioteca de scFv mutados y la selección de scFv, preferiblemente selección de scFv que tienen al menos una propiedad funcional mejorada. En una realización preferida, la biblioteca se examina usando un sistema de control de calidad (sistema QC) de levaduras (descrito con más detalle anteriormente), lo que permite la selección de regiones de marco de scFv que tienen estabilidad y/o solubilidad potenciadas en un entorno reductor.

15 En la técnica se han descrito otras tecnologías de selección adecuadas para examinar bibliotecas de scFv, incluyendo, pero sin limitarse a tecnologías de presentación tales como presentación en fagos, presentación en ribosomas y presentación en levaduras (Jung *et al.* (1999) *J. Mol. Biol.* 294: 163- 180; Wu *et al.* (1999) *J. Mol. Biol.* 294: 151- 162; Schier *et al.* (1996) *J. Mol. Biol.* 255: 28-43).

20 En una realización, una posición de aminoácido identificada para mutación se sustituye por un residuo de aminoácido que está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En otra realización, la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos se conserva con residuos de aminoácido hidrófobos y la posición de aminoácido identificada para mutación dentro del scFv se sustituye por un residuo de aminoácido hidrófobo que está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En aún otra realización, la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos se conserva con residuos de aminoácido hidrófilos y la posición de aminoácido identificada para mutación dentro del scFv se sustituye por un residuo de aminoácido hidrófilo que está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En aún otra realización, la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos se conserva con residuos de aminoácido que pueden formar un enlace de hidrógeno y la posición de aminoácido identificada para mutación dentro del scFv se sustituye por un residuo de aminoácido que puede formar un enlace de hidrógeno que está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos. En todavía otra realización, la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos se conserva con residuos de aminoácido que tienen una propensión a formar una lámina β y la posición de aminoácido identificada para mutación dentro del scFv se sustituye por un residuo de aminoácido que tiene una propensión a formar una lámina β que está lo más significativamente enriquecido en la posición correspondiente dentro de las secuencias de aminoácidos de V_H o V_L de anticuerpos de la base de datos.

En una realización, la mejor sustitución que minimiza la energía libre global se selecciona como la mutación que va a realizarse en la(s) posición/posiciones de aminoácido de interés. La mejor sustitución que minimiza la energía libre global puede determinarse usando la ley de Boltzmann. La fórmula para la ley de Boltzmann es $\Delta\Delta G_{th} = RT\ln(f_{parental}/f_{consenso})$.

45 El papel de mutaciones posiblemente estabilizantes puede determinarse adicionalmente examinando, por ejemplo, interacciones locales y no locales, residuos canónicos, superficies de contacto, grado de exposición y propensión a giros β . Pueden aplicarse métodos de modelado molecular conocidos en la técnica, por ejemplo, al examinar adicionalmente el papel de mutaciones posiblemente estabilizantes. También pueden usarse métodos de modelado molecular para seleccionar sustituciones de aminoácido de "mejor ajuste" si está considerándose un panel de posibles sustituciones.

50 Dependiendo de la posición de aminoácido particular, pueden justificarse análisis adicionales. Por ejemplo, residuos pueden estar implicados en la interacción entre la cadena pesada y la ligera o pueden interactuar con otros residuos a través de puentes de sal o enlaces de H. En estos casos puede requerirse un análisis especial. En otra realización de la presente invención, puede cambiarse un residuo posiblemente problemático para la estabilidad por uno que es compatible con su homólogo en un par covariante. Alternativamente, el residuo homólogo puede mutarse con el fin de ser compatible con el aminoácido inicialmente identificado como problemático.

Optimización de la solubilidad

Los residuos posiblemente problemáticos para la solubilidad en un anticuerpo scFv incluyen aminoácidos hidrófobos

5 que están expuestos a disolvente en un scFv, pero que en el contexto de un anticuerpo de longitud completa quedarán enterrados en la superficie de contacto entre dominios variable y constante. En un scFv modificado por ingeniería genética, que carece de dominios constantes, los residuos hidrófobos que participaban en las interacciones entre los dominios variable y constante quedan expuestos a disolvente (véase, por ejemplo, Nieba *et al.* (1997) *Protein Eng.* 10: 435-44). Estos residuos sobre la superficie del scFv tienden a provocar agregación y por tanto problemas de solubilidad.

10 Se han descrito varias estrategias para sustituir aminoácidos hidrófobos que están expuestos a disolvente en anticuerpos scFv. Tal como conocen bien los expertos en la técnica, modificar residuos en determinadas posiciones afecta a las propiedades biofísicas de anticuerpos tales como estabilidad, solubilidad y afinidad. En muchos casos estas propiedades están interrelacionadas, lo que significa que el cambio de un único aminoácido puede afectar a varias de las propiedades mencionadas anteriormente. Por tanto, mutar residuos hidrófobos expuestos al disolvente de una manera no conservativa puede provocar reducción de la estabilidad y/o pérdida de afinidad por su antígeno.

15 Otros enfoques pretenden resolver problemas de solubilidad mediante uso exhaustivo de tecnologías de presentación de proteínas y/o esfuerzos de selección. Sin embargo, tales métodos requieren mucho tiempo, con frecuencia no logran proporcionar proteína soluble o dan como resultado una estabilidad inferior o reducción de la afinidad del anticuerpo. En la presente invención, se dan a conocer métodos para diseñar mutaciones de residuos hidrófobos expuestos a disolvente por residuos con una mayor hidrofilia usando un análisis basado en secuencia. Los residuos posiblemente problemáticos pueden sustituirse eligiendo el aminoácido hidrófilo representado con mayor frecuencia en posiciones definidas. Si se encuentra que un residuo interacciona con cualquier otro residuo en el anticuerpo, el residuo posiblemente problemático puede mutarse, no por el residuo más frecuente, sino por uno que es compatible con el segundo aminoácido del par covariante. Alternativamente, también puede mutarse un segundo aminoácido del par covariante con el fin de restaurar la combinación de aminoácidos. Además, puede tenerse en cuenta el porcentaje de similitud entre secuencias para ayudar a encontrar una combinación óptima de dos aminoácidos interrelacionados.

20 Se identifican aminoácidos hidrófobos sobre la superficie del scFv usando varios enfoques, incluyendo, pero sin limitarse a, enfoques basados en exposición a disolvente, información experimental e información de secuencia, así como modelado molecular.

25 En una realización de esta invención, la solubilidad se mejora sustituyendo residuos hidrófobos expuestos sobre la superficie del anticuerpo scFv por los residuos hidrófilos más frecuentes presentes en esas posiciones en bases de datos. Este fundamento se basa en el hecho de que es probable que los residuos que se producen con frecuencia no sean problemáticos. Tal como apreciarán los expertos en la técnica, las sustituciones conservativas tienen habitualmente un pequeño efecto de desestabilización de la molécula, mientras que las sustituciones no conservativas pueden ser perjudiciales para las propiedades funcionales del scFv.

30 Algunas veces los residuos hidrófobos sobre la superficie del anticuerpo pueden estar implicados en la interacción entre la cadena pesada y la ligera o pueden interactuar con otros residuos a través de puentes de sal o enlaces de H. En estos casos pueden requerirse análisis especiales. En otra realización de la presente invención, los residuos posiblemente problemáticos para la solubilidad pueden mutarse, no por el residuo más frecuente, sino por uno compatible con el par covariante o puede realizarse una segunda mutación para restaurar la combinación de aminoácidos covariantes.

35 Pueden usarse métodos adicionales para diseñar mutaciones en posiciones hidrófobas expuestas a disolvente. En otra realización de esta invención, se dan a conocer métodos que emplean restricción de la base de datos a aquellas secuencias que revelan la mayor similitud con el scFv que va a modificarse (comentado anteriormente). Aplicando tal base de datos de referencia restringida, se diseña la mutación de tal manera que es el mejor ajuste en el contexto de la secuencia específica del anticuerpo que va a optimizarse. En esta situación, el residuo hidrófilo elegido puede de hecho estar escasamente representado en su posición respectiva en comparación con un mayor número de secuencias (es decir, la base de datos sin restringir).

Optimización de la estabilidad

40 Los fragmentos de anticuerpo de cadena sencilla contienen un ligador peptídico que une covalentemente los dominios variables ligero y pesado. Aunque tal ligador es eficaz para evitar que se separen los dominios variables, y por tanto hace que el scFv sea superior con respecto al fragmento Fv, el fragmento scFv todavía es más propenso a desplegamiento y agregación en comparación con un fragmento Fab o con un anticuerpo de longitud completa, en ambos de los cuales V_H y V_L sólo están unidos de manera indirecta a través de los dominios constantes.

45 Otro problema común en scFv es la exposición de residuos hidrófobos sobre la superficie del scFv que conducen a agregación intermolecular. Además, algunas veces mutaciones somáticas adquiridas durante el proceso de maduración por afinidad colocan residuos hidrófilos en el núcleo de la lámina β . Tales mutaciones pueden tolerarse bien en el formato de IgG o incluso en un fragmento Fab, pero en un scFv esto contribuye claramente a la desestabilización y consiguiente desplegamiento.

Los factores conocidos que contribuyen a la desestabilización de scFv incluyen: residuos hidrófobos expuestos a

disolvente sobre la superficie del anticuerpo scFv; residuos hidrófilos no habituales enterrados en el núcleo de la proteína, así como residuos hidrófilos presentes en la superficie de contacto hidrófoba entre las cadenas pesada y ligera. Además, se sabe que interacciones de empaquetamiento de van der Waals entre residuos no polares en el núcleo desempeñan un papel importante en la estabilidad de la proteína (Monsellier E. y Bedouelle H. (2006) *J. Mol. Biol.* 362:580-93, Tan *et al.* (1998) *Biophys. J.* 75:1473-82; Worn A. y Plückthun A. (1998) *Biochemistry* 37:13120-7).

Por tanto, en una realización, con el fin de aumentar la estabilidad de anticuerpos scFv, se identifican aminoácidos no habituales y/o desfavorables en posiciones muy conservadas y se mutan por aminoácidos que son más comunes en esas posiciones conservadas. Tales aminoácidos no habituales y/o desfavorables incluyen: (i) residuos hidrófobos expuestos a disolvente sobre la superficie del anticuerpo scFv; (ii) residuos hidrófilos no habituales enterrados en el núcleo de la proteína; (iii) residuos hidrófilos presentes en la superficie de contacto hidrófoba entre las cadenas pesada y ligera; y (iv) residuos que alteran la superficie de contacto de V_H/V_L mediante impedimento estérico.

Por tanto, en una realización de la invención, puede lograrse un aumento de la estabilidad sustituyendo aminoácidos que están escasamente representados en sus posiciones por aminoácidos que se producen con mayor frecuencia en esas posiciones. La frecuencia de aparición proporciona generalmente una indicación de aceptación biológica.

Los residuos pueden estar implicados en la interacción entre la cadena pesada y la ligera o pueden interactuar con otros residuos a través de puentes de sal, enlaces de H o enlaces disulfuro. En estos casos pueden requerirse análisis especiales. En otra realización de la presente invención, puede cambiarse un residuo posiblemente problemático para la estabilidad por uno que es compatible con su homólogo en un par covariante. Alternativamente, puede mutarse el residuo homólogo con el fin de ser compatible con el aminoácido inicialmente identificado como que es problemático.

Pueden usarse métodos adicionales para diseñar mutaciones para mejorar la estabilidad. En otra realización de esta invención, se dan a conocer métodos que emplean restricción de la base de datos a las secuencias que revelan la mayor similitud con el scFv que va a modificarse (comentado anteriormente). Aplicando tal base de datos de referencia restringida, se diseña la mutación de tal manera que es el mejor ajuste en el contexto de la secuencia específica del anticuerpo que va a optimizarse. La mutación usa el aminoácido más frecuente que está presente en el subconjunto seleccionado de secuencias de la base de datos. En esta situación, el residuo elegido puede de hecho estar escasamente representado en su posición respectiva en comparación con un mayor número de secuencias (es decir, la base de datos sin restringir).

Composiciones y formulaciones de ScFv

Otro aspecto de la invención se refiere a una composición de scFv preparada según los métodos de invención. Por tanto, la invención se refiere a composiciones de scFv modificado por ingeniería genética en las que se han introducido una o más mutaciones en la secuencia de aminoácidos, en comparación con un scFv original de interés, en la que la(s) mutación(s) se ha(n) introducido en una(s) posición/posiciones que se predice que influye(n) sobre una o más propiedades biológicas, tales como estabilidad o solubilidad, en particular una o más posiciones de región de marco. En una realización, el scFv se ha modificado por ingeniería genética para contener una posición de aminoácido mutada (por ejemplo, una posición de región de marco). En otras realizaciones, el scFv se ha modificado por ingeniería genética para contener dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más de diez posiciones de aminoácido mutadas (por ejemplo, posiciones de región de marco).

Otro aspecto de la invención se refiere a formulaciones farmacéuticas de las composiciones de scFv de la invención. Tales formulaciones comprenden normalmente la composición de scFv y un portador farmacéuticamente aceptable. Tal como se usa en el presente documento, "portador farmacéuticamente aceptable" incluye todos y cada uno de los disolventes, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes de retardo de la absorción e isotónicos, y similares que son fisiológicamente compatibles. Preferiblemente, el portador es adecuado para, por ejemplo, administración intravenosa, intramuscular, subcutánea, parenteral, espinal, epidérmica (por ejemplo, mediante inyección o infusión), o tópica (por ejemplo, al ojo o a la piel). Dependiendo de la vía de administración, el scFv puede recubrirse en un material para proteger el compuesto frente a la acción de ácidos y otras condiciones naturales que pueden inactivar el compuesto.

Los compuestos farmacéuticos de la invención pueden incluir una o más sales farmacéuticamente aceptables. Una "sal farmacéuticamente aceptable" se refiere a una sal que conserva la actividad biológica deseada del compuesto original y no confiere ningún efecto toxicológico no deseado (véase, por ejemplo, Berge, S. M., *et al.* (1977) *J. Pharm. Sci.* 66:1-19). Los ejemplos de tales sales incluyen sales de adición de ácido y sales de adición de base. Las sales de adición de ácido incluyen las derivadas de ácidos inorgánicos no tóxicos, tales como clorhídrico, nítrico, fosfórico, sulfúrico, bromhídrico, yodhídrico, fosforoso y similares, así como de ácidos orgánicos no tóxicos tales como ácidos alifáticos mono y dicarboxílicos, ácidos alcanicos sustituidos con fenilo, ácidos hidroxialcanoicos, ácidos aromáticos, ácidos sulfónicos alifáticos y aromáticos y similares. Las sales de adición de base incluyen las derivadas de metales alcalinotérreos, tales como sodio, potasio, magnesio, calcio y similares, así como de aminas orgánicas no tóxicas, tales como N,N'-dibenciletilendiamina, N-metilglucamina, cloroprocaina, colina, dietanolamina, etilendiamina, procaína y similares.

Una composición farmacéutica de la invención también puede incluir un antioxidante farmacéuticamente aceptable. Los ejemplos de antioxidantes farmacéuticamente aceptables incluyen: (1) antioxidantes solubles en agua, tales como ácido ascórbico, clorhidrato de cisteína, bisulfato de sodio, metabisulfito de sodio, sulfito de sodio y similares; (2) antioxidantes solubles en aceite, tales como palmitato de ascorbilo, hidroxianisol butilado (BHA), hidroxitolueno butilado (BHT), lecitina, galato de propilo, alfa-tocoferol, y similares; y (3) agentes quelantes de metales, tales como ácido cítrico, ácido etilendiamina-tetraacético (EDTA), sorbitol, ácido tartárico, ácido fosfórico, y similares.

Los ejemplos de portadores acuosos y no acuosos adecuados que pueden emplearse en las composiciones farmacéuticas de la invención incluyen agua, etanol, polioles (tales como glicerol, propilenglicol, polietilenglicol, y similares), y mezclas adecuadas de los mismos, aceites vegetales, tales como aceite de oliva, y ésteres orgánicos inyectables, tales como oleato de etilo. Puede mantenerse la fluidez apropiada, por ejemplo, mediante el uso de materiales de recubrimiento, tales como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersiones, y mediante el uso de tensioactivos.

Estas composiciones también pueden contener adyuvantes tales como conservantes, agentes humectantes, agentes emulsionantes y agentes dispersantes. La prevención de la presencia de microorganismos puede garantizarse tanto mediante procedimientos de esterilización, citados anteriormente, como mediante la inclusión de diversos agentes antibacterianos y antifúngicos, por ejemplo, parabeno, clorobutanol, ácido fenol-sórbico, y similares. También puede ser deseable incluir agentes isotónicos, tales como azúcares, cloruro de sodio, y similares, en las composiciones. Además, la absorción prolongada de la forma farmacéutica inyectable puede provocarse mediante la inclusión de agentes que retrasan la absorción, tales como monoestearato de aluminio y gelatina.

Los portadores farmacéuticamente aceptables incluyen disoluciones o dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de disoluciones o dispersiones inyectables estériles. El uso de tales medios y agentes para sustancias farmacéuticamente activas se conoce en la técnica. Excepto en cuanto que cualquier medio convencional o agente sea incompatible con el compuesto activo, se contempla el uso de los mismos en las composiciones farmacéuticas de la invención. También pueden incorporarse compuestos activos complementarios en las composiciones.

Las composiciones terapéuticas normalmente deben ser estériles y estables en las condiciones de fabricación y almacenamiento. La composición puede formularse como una disolución, microemulsión, liposoma, u otra estructura ordenada adecuada para una alta concentración de fármaco. El portador puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol, y polietilenglicol líquido, y similares), y mezclas adecuadas de los mismos. La fluidez apropiada puede mantenerse, por ejemplo, mediante el uso de un recubrimiento tal como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersión y mediante el uso de tensioactivos. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol, o cloruro de sodio en la composición. La absorción prolongada de las composiciones inyectables puede provocarse incluyendo en la composición un agente que retarda la absorción, por ejemplo, sales de monoestearato y gelatina.

Pueden prepararse disoluciones inyectables estériles incorporando el compuesto activo en la cantidad requerida en un disolvente apropiado con un componente o una combinación de componentes indicados anteriormente, según se requiera, seguido por esterilización por microfiltración. Generalmente, las dispersiones se preparan incorporando el compuesto activo en un vehículo estéril que contiene un medio de dispersión básico y los demás componentes requeridos de los indicados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los métodos de preparación preferidos son secado a vacío y secado por congelación (liofilización) que proporcionan un polvo del principio activo más cualquier componente deseado adicional de una disolución anteriormente esterilizada por filtración del mismo.

La cantidad de principio activo que puede combinarse con un material portador para producir una forma de dosificación individual variará dependiendo del sujeto que esté tratándose, y del modo de administración particular. La cantidad de principio activo que puede combinarse con un material portador para producir una forma de dosificación individual será generalmente la cantidad de la composición que produce un efecto terapéutico. Generalmente, del cien por ciento, esta cantidad oscilará entre aproximadamente el 0,01 por ciento y aproximadamente el noventa y nueve por ciento de principio activo, preferiblemente entre aproximadamente el 0,1 por ciento y aproximadamente el 70 por ciento, lo más preferiblemente entre aproximadamente el 1 por ciento y aproximadamente el 30 por ciento de principio activo en combinación con un portador farmacéuticamente aceptable.

Los regímenes de dosificación se ajustan para proporcionar la respuesta deseada óptima (por ejemplo, una respuesta terapéutica). Por ejemplo, puede administrarse un único bolo, pueden administrarse varias dosis divididas a lo largo del tiempo o puede reducirse o aumentarse proporcionalmente la dosis según se indique por las exigencias de la situación terapéutica. Resulta especialmente ventajoso formular composiciones parenterales en forma unitaria de dosificación por su facilidad de administración y uniformidad de dosificación. La forma unitaria de dosificación tal como se usa en el presente documento se refiere a unidades físicamente diferenciadas adecuadas como dosificaciones unitarias para los sujetos que van a tratarse; cada unidad contiene una cantidad predeterminada de compuesto activo calculada para producir el efecto terapéutico deseado en asociación con el portador farmacéutico requerido. La especificación de las formas unitarias de dosificación de la invención viene

dictada por, y depende directamente de, (a) las características únicas del compuesto activo y el efecto terapéutico particular que va a lograrse, y (b) las limitaciones inherentes en la técnica de composición de tal compuesto activo para el tratamiento de la sensibilidad en individuos.

Modificación por ingeniería genética de agente de unión inmunológica basada en el enfoque de “consenso funcional”

5 Tal como se describe en detalle en los ejemplos 2 y 3, el enfoque de “consenso funcional” descrito en el presente documento, en el que se usa una base de datos de secuencias de scFv seleccionadas por las propiedades mejoradas para analizar la variabilidad de la posición de región de marco, permite la identificación de posiciones de aminoácido que son o bien más o bien menos tolerantes a la variabilidad en comparación con la variabilidad en las mismas posiciones en bases de datos de anticuerpos maduros y/o de línea germinal. Tal como se describe en detalle en los ejemplos 5 y 6, la retromutación de determinadas posiciones de aminoácido dentro de un scFv de muestra para dar el residuo consenso de línea germinal tiene un efecto o bien neutro o bien perjudicial, mientras que las variantes de scFv que contienen residuos de “consenso funcional” muestran una estabilidad térmica aumentada en comparación con la molécula de scFv silvestre. Por consiguiente, las posiciones de región de marco identificadas en el presente documento mediante el enfoque de consenso funcional son posiciones preferidas para la modificación de scFv con el fin de alterar, y preferiblemente mejorar, las propiedades funcionales del scFv. Tal como se usa en el presente documento, el término “propiedades funcionales” se refiere, por ejemplo, a estabilidad mejorada, solubilidad mejorada, ausencia de agregación, una mejora en la expresión, una mejora en el rendimiento de repliegamiento tras un procedimiento de purificación de cuerpos de inclusión o una combinación de dos o más de dichas mejoras. Preferiblemente, mejorar la propiedad funcional del scFv no implica una mejora en la afinidad de unión a antígeno.

Tal como se expone en las tablas 3-8 en el ejemplo 3, las siguientes posiciones de región de marco se han identificado como posiciones preferidas para modificación en las secuencias de V_H o V_L indicadas (la numeración usada a continuación es el sistema de numeración de AHo; en las tablas 1 y 2 en el ejemplo 1 se exponen tablas de conversión para convertir la numeración de AHo en la numeración del sistema de Kabat):

25 VH3: posiciones de aminoácido 1, 6, 7, 89 y 103;

VH1a: posiciones de aminoácido 1, 6, 12, 13, 14, 19, 21, 90, 92, 95 y 98;

VH1b: posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107;

Vκ1: posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57, 91, y 103;

Vκ3: 2, 3, 10, 12, 18, 20, 56, 74, 94, 101 y 103; y

30 Vλ1: 1, 2, 4, 7, 11, 14, 46, 53, 82, 92 y 103.

Por consiguiente, pueden seleccionarse una o más de estas posiciones de aminoácido para la modificación por ingeniería genética en agentes de unión inmunológica, tales como moléculas de scFv, para así producir formas variantes (es decir, mutadas) de los agentes de unión inmunológica. Preferiblemente, la modificación por ingeniería genética da como resultado agentes de unión inmunológica que tienen una o más propiedades funcionales mejoradas, sin afectar a la afinidad de unión a antígeno. En una realización preferida, los agentes de unión inmunológica resultantes tienen características de estabilidad y/o solubilidad mejoradas.

Por tanto, en aún otro aspecto, la invención se refiere a un método de modificación por ingeniería genética de un agente de unión inmunológica, en particular un método para mejorar una o más propiedades funcionales de un agente de unión inmunológica, comprendiendo el agente de unión inmunológica (i) una región variable de cadena pesada, o fragmento de la misma, de un tipo de dominio VH3, VH1a o VH1b, comprendiendo la región variable de cadena pesada residuos de región de marco de V_H y/o (ii) una región variable de cadena ligera, o fragmento de la misma, de un tipo de dominio Vκ1, Vκ3 o Vλ1, comprendiendo la región variable de cadena ligera residuos de región de marco de V_L , comprendiendo el método:

a) seleccionar una o más posiciones de aminoácido dentro del agente de unión inmunológica para mutación; y

45 b) mutar la una o más posiciones de aminoácido seleccionadas para mutación, en el que la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en:

(i) posiciones de aminoácido 1, 6, 7, 89 y 103 de VH3 usando la numeración de AHo (posiciones de aminoácido 1, 6, 7, 78 y 89 usando la numeración de Kabat);

50 (ii) posiciones de aminoácido 1, 6, 12, 13, 14, 19, 21, 90, 92, 95 y 98 de VH1a usando la numeración de AHo (posiciones de aminoácido 1, 6, 11, 12, 13, 18, 20, 79, 81, 82b y 84 usando la numeración de Kabat);

(iii) posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107 de VH1b usando la numeración de AHo (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76 y 93 usando la numeración de Kabat);

(iv) posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57, 91 y 103 de V κ 1 usando la numeración de AHo (posiciones de aminoácido 1, 3, 4, 24, 39, 42, 49, 73, y 85 usando la numeración de Kabat);

(v) posiciones de aminoácido 2, 3, 10, 12, 18, 20, 56, 74, 94, 101 y 103 de V κ 3 usando la numeración de AHo (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 48, 58, 76, 83 y 85 usando la numeración de Kabat); y

5 (vi) posiciones de aminoácido 1, 2, 4, 7, 11, 14, 46, 53, 82, 92 y 103 de V λ 1 usando la numeración de AHo (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 38, 45, 66, 74 y 85 usando la numeración de Kabat).

En una realización preferida, el método mejora una o más propiedades funcionales de un agente de unión inmunológica con la condición de que la una o más propiedades funcionales del scFv no implican una mejora en la afinidad de unión a antígeno.

10 En otra realización preferida, la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 6, 7, 89 y 103 de VH3 usando la numeración de AHo (posiciones de aminoácido 1, 6, 7, 78 y 89 usando la numeración de Kabat).

15 En otra realización preferida, la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 6, 12, 13, 14, 19, 21, 90, 92, 95 y 98 de VH1a usando la numeración de AHo (posiciones de aminoácido 1, 6, 11, 12, 13, 18, 20, 79, 81, 82b y 84 usando la numeración de Kabat).

20 En otra realización preferida, la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107 de VH1b usando la numeración de AHo (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76 y 93 usando la numeración de Kabat).

En otra realización preferida, la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57, 91 y 103 de V κ 1 usando la numeración de AHo (posiciones de aminoácido 1, 3, 4, 24, 39, 42, 49, 73 y 85 usando la numeración de Kabat).

25 En otra realización preferida, la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 2, 3, 10, 12, 18, 20, 56, 74, 94, 101 y 103 de V κ 3 usando la numeración de AHo (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 48, 58, 76, 83 y 85 usando la numeración de Kabat).

30 En otra realización preferida, una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 2, 4, 7, 11, 14, 46, 53, 82, 92 y 103 de V λ 1 usando la numeración de AHo (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 38, 45, 66, 74 y 85 usando la numeración de Kabat).

En diversas realizaciones, se seleccionan para mutación una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte o más de veinte de las posiciones de aminoácido descritas anteriormente.

35 En aún otro aspecto, la invención se refiere a un método para mejorar las propiedades de estabilidad y/o solubilidad de un agente de unión inmunológica, comprendiendo el agente de unión inmunológica (i) una región variable de cadena pesada, o fragmento de la misma, de un tipo de dominio VH3, VH1a o VH1b, comprendiendo la región variable de cadena pesada residuos de región de marco de V_H y/o (ii) una región variable de cadena ligera, o fragmento de la misma, de un tipo de dominio V κ 1, V κ 3 o V λ 1, comprendiendo la región variable de cadena ligera residuos de región de marco de V_L, comprendiendo el método:

a) seleccionar una o más posiciones de aminoácido dentro del agente de unión inmunológica para mutación; y

b) mutar la una o más posiciones de aminoácido seleccionadas para mutación, en el que la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en:

45 (i) posiciones de aminoácido 1, 89 y 103 de VH3 usando la numeración de AHo (posiciones de aminoácido 1, 78 y 89 usando la numeración de Kabat);

(ii) posiciones de aminoácido 1, 12, 13, 14, 19, 21, 90, 92, 95 y 98 de VH1a usando la numeración de AHo (posiciones de aminoácido 1, 11, 12, 13, 18, 20, 79, 81, 82b y 84 usando la numeración de Kabat);

50 (iii) posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107 de VH1b usando la numeración de AHo (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76 y 93 usando la numeración de Kabat);

(iv) posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57 y 91 de V κ 1 usando la numeración de AHo (posiciones

de aminoácido 1, 3, 4, 24, 39, 42, 49 y 73 usando la numeración de Kabat);

(v) posiciones de aminoácido 2, 3, 10, 12, 18, 20, 56, 74, 94 y 101 de $V_{\kappa 3}$ usando la numeración de AHO (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 48, 58, 76 y 83 usando la numeración de Kabat); y

5 (vi) posiciones de aminoácido 1, 2, 4, 7, 11, 14, 46, 53, 82 y 92 de $V_{\lambda 1}$ usando la numeración de AHO (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 38, 45, 66 y 74 usando la numeración de Kabat).

En un aspecto adicional, la invención se refiere a un método las propiedades de estabilidad y/o solubilidad de un agente de unión inmunológica sin afectar a la afinidad de unión a antígeno de un agente de unión inmunológica, comprendiendo el agente de unión inmunológica (i) una región variable de cadena pesada, o fragmento de la misma, de un tipo de dominio VH3, VH1a o VH1b, comprendiendo la región variable de cadena pesada residuos de región de marco de V_H y/o (ii) una región variable de cadena ligera, o fragmento de la misma, de un tipo de dominio $V_{\kappa 1}$, $V_{\kappa 3}$ o $V_{\lambda 1}$, comprendiendo la región variable de cadena ligera residuos de región de marco de V_L , comprendiendo el método:

a) seleccionar una o más posiciones de aminoácido dentro del agente de unión inmunológica para mutación; y

15 b) mutar la una o más posiciones de aminoácido seleccionadas para mutación, en el que la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en:

(i) posiciones de aminoácido 1 y 103 de VH3 usando la numeración de AHO (posiciones de aminoácido 1 y 89 usando la numeración de Kabat);

(ii) posiciones de aminoácido 1, 12, 13, 14, 19, 21, 90, 95 y 98 de VH1a usando la numeración de AHO (posiciones de aminoácido 1, 11, 12, 13, 18, 20, 79, 82b y 84 usando la numeración de Kabat);

20 (iii) posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55 y 77 de VH1b usando la numeración de AHO (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48 y 66 usando la numeración de Kabat);

(iv) posiciones de aminoácido 1 y 3 de $V_{\kappa 1}$ usando la numeración de AHO (posiciones de aminoácido 1 y 3 usando la numeración de Kabat);

25 (v) posiciones de aminoácido 2, 3, 10, 12, 18, 20, 94 y 101 de $V_{\kappa 3}$ usando la numeración de AHO (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 76 y 83 usando la numeración de Kabat); y

(vi) posiciones de aminoácido 1, 2, 4, 7, 11, 14, 82 y 92 de $V_{\lambda 1}$ usando la numeración de AHO (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 66 y 74 usando la numeración de Kabat).

30 Preferiblemente, el agente de unión inmunológica es un scFv, pero otros agentes de unión inmunológica, tales como inmunoglobulinas de longitud completa, fragmentos Fab, Dab, Nanobodies o cualquier otro tipo de agente de unión inmunológica descrito en el presente documento, también pueden modificarse por ingeniería genética según el método. La invención también abarca agentes de unión inmunológica preparados según el método de modificación por ingeniería genética, así como composiciones que comprenden los agentes de unión inmunológica y un portador farmacéuticamente aceptable.

35 En determinadas realizaciones a modo de ejemplo, un agente de unión inmunológica modificado por ingeniería genética según el método de la invención es un agente de unión inmunológica reconocido en la técnica que se une a un antígeno diana de importancia terapéutica o un agente de unión inmunológica que comprende regiones variables (regiones V_L y/o V_L) o una o más CDR (por ejemplo, CDRL1, CDRL2, CDRL3, CDRH1, CDRH2, y/o CDRH3) derivadas del agente de unión inmunológica de importancia terapéutica. Por ejemplo, agentes de unión inmunológica actualmente aprobados por la FDA u otras autoridades reguladoras pueden modificarse por ingeniería genética según los métodos de la invención. Más específicamente, estos agentes de unión inmunológica a modo de ejemplo incluyen, pero no se limitan a, anticuerpos anti-CD3 tales como muromonab (Orthoclone® OKT3; Johnson&Johnson, Brunswick, NJ; véase Arakawa *et al.* J. Biochem, (1996) 120:657-662; Kung y Goldstein *et al.*, Science (1979), 206: 347-349), anticuerpos anti-CD11 tales como efalizumab (Raptiva®, Genentech, South San Francisco, CA), anticuerpos anti-CD20 tales como rituximab (Rituxan®/Mabthera®, Genentech, South San Francisco, CA), tositumomab (Bexxar®, GlaxoSmithKline, Londres) o ibritumomab (Zevalin®, Biogen Idec, Cambridge MA) (véanse las patentes estadounidenses n.ºs 5.736.137; 6.455.043; y 6.682.734), anticuerpos anti-CD25 (IL2R α) tales como daclizumab (Zenapax®, Roche, Basilea, Suiza) o basiliximab (Simulect®, Novartis, Basilea, Suiza), anticuerpos anti-CD33 tales como gemtuzumab (Mylotarg®, Wyeth, Madison, NJ, véanse las patentes estadounidenses n.ºs 5.714.350 y 6.350.861), anticuerpos anti-CD52 tales como alemtuzumab (Campath®, Millennium Pharmaceuticals, Cambridge, MA), anticuerpos anti-Gp11b/g11a tales como abciximab (ReoPro®, Centocor, Horsham, PA), anticuerpos anti-TNF α tales como infliximab (Remicade®, Centocor, Horsham, PA) o adalimumab (Humira®, Abbott, Abbott Park, IL, véase la patente estadounidense n.º 6.258.562), anticuerpos anti-IgE tales como omalizumab (Xolair®, Genentech, South San Francisco, CA), anticuerpos anti-VSR tales como palivizumab (Synagis®, Medimmune, Gaithersburg, MD, véase la patente estadounidense n.º 5.824.307), anticuerpos anti-EpCAM tales como

5 edrecolomab (Panorex®, Centocor), anticuerpos anti-EGFR tales como cetuximab (Erbix®, Imclone Systems, Nueva York, NY) o panitumumab (Vectibix®, Amgen, Thousand Oaks, CA), anticuerpos anti-HER2/neu tales como trastuzumab (Herceptin®, Genentech), anticuerpos anti-integrina $\alpha 4$ tales como natalizumab (Tysabri®, BiogenIdec), anticuerpos anti-C5 tales como eculizumab (Soliris®, Alexion Pharmaceuticals, Cheshire, CT) y anticuerpos anti-VEGF tales como bevacizumab (Avastin®, Genentech, véase la patente estadounidense n.º 6.884.879) o ranibizumab (Lucentis®, Genentech).

10 A pesar de lo anterior, en diversas realizaciones, se excluyen determinados agentes de unión inmunológica de usarse en los métodos de modificación por ingeniería genética de la invención y/o se excluyen de ser la composición de agente de unión inmunológica producida mediante los métodos de modificación por ingeniería genética. Por ejemplo, en diversas realizaciones, existe la condición de que el agente de unión inmunológica no es alguno de los anticuerpos scFv, o variantes de los mismos, dados a conocer en las publicaciones PCT WO 2006/131013 y WO 2008/006235, tales como ESBA105 o variantes del mismo que se dan a conocer en las publicaciones PCT WO 2006/131013 y WO 2008/006235.

15 En diversas otras realizaciones, si el agente de unión inmunológica que va a modificarse por ingeniería genética según los métodos descritos anteriormente es cualquiera de los anticuerpos scFv, o variantes de los mismos, dados a conocer en las publicaciones PCT WO 2006/131013 o WO 2008/006235, entonces puede existir la condición de que la lista de posiciones de aminoácido posibles que pueden seleccionarse para sustitución según el método de modificación por ingeniería genética no incluye alguna o ninguna de las siguientes posiciones de aminoácido: posición de AHo 4 (Kabat 4) de $V_{\kappa 1}$ o $V_{\lambda 1}$; posición de AHo 101 (Kabat 83) de $V_{\kappa 3}$; posición de AHo 12 (Kabat 11) de VH1a o VH1b; posición de AHo 50 (Kabat 43) de VH1b; posición de AHo 77 (Kabat 66) para VH1b; posición de AHo 78 (Kabat 67) para VH1b; posición de AHo 82 (Kabat 71) para VH1b; posición de AHo 86 (Kabat 75) para VH1b; posición de AHo 87 (Kabat 76) para VH1b; posición de AHo 89 (Kabat 78) para VH3; posición de AHo 90 (Kabat 79) para VH1a; y/o posición de AHo 107 (Kabat 93) para VH1b.

20 En todavía diversas otras realizaciones, para cualquier agente de unión inmunológica que va a modificarse por ingeniería genética según los métodos descritos anteriormente, y/o cualquier agente de unión inmunológica producido según los métodos descritos anteriormente, puede existir la condición de que la lista de posiciones de aminoácido posibles que pueden seleccionarse para sustitución según el método de modificación por ingeniería genética no incluye alguna o ninguna de las siguientes posiciones de aminoácido: posición de AHo 4 (Kabat 4) de $V_{\kappa 1}$ o $V_{\lambda 1}$; posición de AHo 101 (Kabat 83) de $V_{\kappa 3}$; posición de AHo 12 (Kabat 11) de VH1a o VH1b; posición de AHo 50 (Kabat 43) de VH1b; posición de AHo 77 (Kabat 66) para VH1b; posición de AHo 78 (Kabat 67) para VH1b; posición de AHo 82 (Kabat 71) para VH1b; posición de AHo 86 (Kabat 75) para VH1b; posición de AHo 87 (Kabat 76) para VH1b; posición de AHo 89 (Kabat 78) para VH3; posición de AHo 90 (Kabat 79) para VH1a; y/o posición de AHo 107 (Kabat 93) para VH1b.

35 Otras realizaciones

Se entiende que la invención también incluye cualquiera de las metodologías, referencias, y/o composiciones expuestas en los apéndices (A-C) de la solicitud de patente estadounidense provisional con n.º de serie 60/905.365 y los apéndices (A-I) de la solicitud de patente estadounidense provisional con n.º de serie 60/937.112, incluyendo, pero son limitarse a, bases de datos identificadas, bioinformática, manipulación de datos *in silico* y métodos de interpretación, ensayos funcionales, secuencias preferidas, posiciones / alteraciones de residuo(s) preferidas, identificación y selección de región de marco, alteraciones de región de marco, alineación e integración de CDR, y alteraciones/mutaciones preferidas.

40 Puede encontrarse información adicional referente a estas metodologías y composiciones en los documentos estadounidenses con n.ºs de serie 60/819.378; y 60/899.907, y la publicación PCT WO 2008/006235, titulada "scFv Antibodies Which Pass Epithelial And/Or Endothelial Layers" presentadas en julio de 2006 y el 6 de febrero de 2007, respectivamente; el documento WO06131013A2 titulado "Stable And Soluble Antibodies Inhibiting TNF α " presentado el 6 de junio de 2006; el documento EP1506236A2 titulado "Immunoglobulin Frameworks Which Demonstrate Enhanced Stability In The Intracellular Environment And Methods Of Identifying Same" presentado el 21 de mayo de 2003; el documento EP1479694A2 titulado "Intrabodies ScFv with defined framework that is stable in a reducing environment" presentado el 18 de diciembre de 2000; el documento EP1242457B1 titulado "Intrabodies With Defined Framework That Is Stable In A Reducing Environment And Applications Thereof" presentado el 18 de diciembre de 2000; el documento WO03097697A2 titulado "Immunoglobulin Frameworks Which Demonstrate Enhanced Stability In The Intracellular Environment And Methods Of Identifying Same" presentado el 21 de mayo de 2003; y el documento WO0148017A1 titulado "Intrabodies With Defined Framework That Is Stable In A Reducing Environment And Applications Thereof" presentado el 18 de diciembre de 2000; y Honegger *et al.*, J. Mol. Biol. 309:657-670 (2001).

Además, se entiende que la invención también incluye metodologías y composiciones adecuadas para el descubrimiento y/o la mejora de otros formatos de anticuerpo, por ejemplo, anticuerpos de longitud completa o fragmentos de los mismos, por ejemplo Fab, Dab, y similares. Por consiguiente, los principios y residuos identificados en el presente documento como adecuados para selección o alteración para lograr propiedades biofísicas y/o terapéuticas deseadas pueden aplicarse a una amplia gama de agentes de unión inmunológica. En

una realización, los anticuerpos terapéuticamente relevantes, por ejemplo, anticuerpos aprobados por la FDA, se mejoran modificando una o más posiciones de residuo tal como se da a conocer en el presente documento.

Sin embargo, la invención no se limita a la modificación por ingeniería genética de agentes de unión inmunológica. Por ejemplo, un experto en la técnica reconocerá que los métodos de la invención pueden aplicarse a la modificación por ingeniería genética de otras moléculas de unión, distintas de inmunoglobulina, incluyendo, pero sin limitarse a, moléculas de unión a fibronectina tales como Adnectins (véase el documento WO 01/64942 y las patentes estadounidenses n.ºs 6.673.901, 6.703.199, 7.078.490, y 7.119.171), Affibodies (véanse, por ejemplo, las patentes estadounidenses 6.740.734 y 6.602.977 y en el documento WO 00/63243), Anticalins (también conocidas como lipocalinas) (véanse los documentos WO99/16873 y WO 05/019254), proteínas de dominio A (véanse los documentos WO 02/088171 y WO 04/044011) y proteínas de repetición de anquirina tales como Darpins o proteínas de repetición de leucina (véanse los documentos WO 02/20565 y WO 06/083275).

La presente divulgación se ilustra adicionalmente por los siguientes ejemplos, que no deben interpretarse como limitaciones adicionales.

EJEMPLO 1: Sistemas de numeración de la posición de anticuerpos

En este ejemplo, se proporcionan tablas de conversión para dos sistemas de numeración diferentes usados para identificar posiciones de residuo de aminoácido en las regiones variables de cadena pesada y ligera del anticuerpo. El sistema de numeración de Kabat se describe adicionalmente en Kabat *et al.* (Kabat, E. A., *et al.* (1991) Sequences of Proteins of Immunological Interest, quinta edición, U.S. Department of Health and Human Services, publicación de NIH n.º 91-3242). El sistema de numeración de AHo se describe adicionalmente en Honegger, A. y Plüchthun, A. (2001) *J Mol Biol.* 309:657-670).

Numeración de la región variable de cadena pesada

Tabla 1: Tabla de conversión para las posiciones de residuo en el dominio variable de cadena pesada

Kabat	AHo	Kabat	AHo	Kabat	AHo
1	1	44	51	87	101
2	2	45	52	88	102
3	3	46	53	89	103
4	4	47	54	90	104
5	5	48	55	91	105
6	6	49	56	92	106
7	7	50	57	93	107
*	8	51	58	94	108
8	9	52	59	95	109
9	10	52a	60	96	110
10	11	52b	61	97	111
11	12	52c	62	98	112
12	13	*	63	99	113
13	14	53	64	100	114
14	15	54	65	100a	115
15	16	55	66	100b	116
16	17	56	67	100c	117
17	18	57	68	100d	118
18	19	58	69	100e	119
19	20	59	70	100f	120
20	21	60	71	100g	121
21	22	61	72	100h	122
22	23	62	73	100i	123
23	24	63	74	*	124
24	25	64	75	*	125
25	26	65	76	*	126
26	27	66	77	*	127
*	28	67	78	*	128
27	29	68	79	*	129
28	30	69	80	*	130
29	31	70	81	*	131
30	32	71	82	*	132
31	33	72	83	*	133
32	34	73	84	*	134
33	35	74	85	*	135
34	36	75	86	*	136
35	37	76	87	101	137

ES 2 532 725 T3

35a	38	77	88	102	138
35b	39	78	89	103	139
*	40	79	90	104	140
*	41	80	91	105	141
*	42	81	92	106	142
36	43	82	93	107	143
37	44	82a	94	108	144
38	45	82b	95	109	145
39	46	82b	96	110	146
40	47	83	97	111	147
41	48	84	98	112	148
42	49	85	99	113	149
43	50	86	100		

Columna 1, posición de residuo en el sistema de numeración de Kabat. Columna 2, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 1. Columna 3, posición de residuo en el sistema de numeración de Kabat. Columna 4, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 3. Columna 5, posición de residuo en el sistema de numeración de Kabat. Columna 6, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 5

Numeración de la región variable de cadena ligera

Tabla 2: Tabla de conversión para las posiciones de residuo en el dominio variable de cadena ligera

Kabat	AHo	Kabat	AHo	Kabat	AHo
1	1	43	51	83	101
2	2	44	52	84	102
3	3	45	53	85	103
4	4	46	54	86	104
5	5	47	55	87	105
6	6	48	56	88	106
7	7	49	57	89	107
8	8	50	58	90	108
9	9	*	59	91	109
10	10	*	60	92	110
11	11	*	61	93	111
12	12	*	62	94	112
13	13	*	63	95	113
14	14	*	64	95a	114
15	15	*	65	95b	115
16	16	*	66	95c	116
17	17	51	67	95d	117
18	18	52	68	95e	118
19	19	53	69	95f	119
20	20	54	70	*	120
21	21	55	71	*	121
22	22	56	72	*	122
23	23	57	73	*	123
24	24	58	74	*	124
25	25	59	75	*	125
26	26	60	76	*	126
27	27	61	77	*	127
*	28	62	78	*	128
27a	29	63	79	*	129
27b	30	64	80	*	130
27c	31	65	81	*	131
27d	32	66	82	*	132
27e	33	67	83	*	133
27f	34	68	84	*	134
*	35	*	85	*	135
28	36	*	86	*	136
29	37	69	87	96	137
30	38	70	88	97	138
31	39	71	89	98	139
32	40	72	90	99	140

33	41	73	91	100	141
34	42	74	92	101	142
35	43	75	93	102	143
36	44	76	94	103	144
37	45	77	95	104	145
38	46	78	96	105	146
39	47	79	97	106	147
40	48	80	98	107	148
41	49	81	99	108	149
42	50	82	100		

Columna 1, posición de residuo en el sistema de numeración de Kabat. Columna 2, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 1. Columna 3, posición de residuo en el sistema de numeración de Kabat. Columna 4, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 3. Columna 5, posición de residuo en el sistema de numeración de Kabat. Columna 6, número correspondiente en el sistema de numeración de AHo para la posición indicada en la columna 5

EJEMPLO 2: Análisis basado en secuencia de secuencias de scFv

En este ejemplo, se describe en detalle el análisis basado en secuencia de secuencias de scFv. En la figura 1 se muestra un diagrama de flujo que resume el procedimiento del análisis.

Recogida y alineación de secuencias de inmunoglobulina humana

- 5 Se recogieron secuencias de dominios variables de líneas germinales y anticuerpos maduros humanos de diferentes bases de datos y se introdujeron en una base de datos personalizada como secuencias de aminoácidos de código de una letra. Se alinearon las secuencias de anticuerpo usando una implementación de EXCEL del algoritmo de alineación de secuencias de Needleman-Wunsch (Needleman *et al.*, J Mol Biol, 48(3):443-53 (1970)). Entonces se subdividió la base de datos en cuatro series diferentes (según la fuente de datos original) para facilitar el análisis y la comparación posterior, tal como sigue:

VBase: Secuencias de línea germinal humana

IMGT: Secuencias de línea germinal humana

Base de datos KDB: Anticuerpos maduros

- 15 Base de datos de QC: Base de datos interna de ESBATech que comprende regiones de marco de scFv seleccionadas mediante el examen de control de calidad

20 El sistema de examen de QC, y las secuencias de la región de marco de scFv que tienen propiedades funcionales deseables seleccionadas del mismo, se describen adicionalmente, por ejemplo, en la publicación PCT WO 2001/48017; la solicitud estadounidense n.º 20010024831; el documento US 20030096306; las patentes estadounidenses n.ºs 7.258.985 y 7.258.986; la publicación PCT WO 2003/097697 y la solicitud estadounidense n.º 20060035320.

25 La introducción de huecos y la nomenclatura de posiciones de residuo se realizaron siguiendo el sistema de numeración de AHo para el dominio variable de inmunoglobulina (Honegger, A. y Plückthun, A. (2001) J. Mol. Biol. 309:657-670). Posteriormente, se identificaron las regiones de marco y las regiones CDR según Kabat *et al.* (Kabat, E. A., *et al.* (1991) Sequences of Proteins of Immunological Interest, quinta edición, U.S. Department of Health and Human Services, publicación de NIH n.º 91-3242). Se descartaron las secuencias en la base de datos KDB completas en menos del 70% o que contenían múltiples residuos indeterminados en las regiones de marco. También se excluyeron las secuencias con una identidad de más del 95% con cualquier otra secuencia dentro de la base de datos para evitar ruido aleatorio en el análisis.

Asignación de secuencias a subgrupos

- 30 Se clasificaron las secuencias de anticuerpo en familias distintas agrupando los anticuerpos según métodos de clasificación basados en la homología de secuencia (Tomlinson, I. M. *et al.* (1992) J. Mol. Biol. 227:776-798; Williams, S.C. y Winter, G. (1993) Eur. J. Immunol. 23:1456-1461; Cox, J. P. *et al.* (1994) Eur. J. Immunol. 24:827-836). El porcentaje de homología para el consenso de familia se limitó a una similitud del 70%. En los casos en que las secuencias mostraron conflictos entre dos o más familias de línea germinal diferentes, o el porcentaje de homología estaba por debajo del 70% (para cualquier familia), se determinó el homólogo de línea germinal más próximo, se analizaron en detalle la longitud de las CDR, las clases canónicas y los residuos de subtipo de definición para asignar correctamente la familia.

Análisis estadístico

Una vez que se definieron las agrupaciones de familia, se realizaron análisis estadísticos para las coincidencias identificadas en el “examen de control de calidad (“QC”)” (tal examen de QC se describe en detalle en la publicación PCT WO 2003/097697). Sólo fueron posibles los análisis para las familias más representadas (VH3, VH1a, VH1b, Vκ1, Vκ3 y Vλ1) puesto que es necesario un número mínimo de secuencias para el análisis. Se calcularon las frecuencias de residuo, $f_i(r)$, para cada posición, i , por el número de veces que se observaba ese tipo de residuo particular dentro del conjunto de datos dividido entre el número total de secuencias. Se calculó la entropía posicional, $N(i)$, como una medida la variabilidad de cada posición de residuo (Shenkin, P.S. *et al.* (1991) *Proteins* 11:297-313; Larson, S.M. y Davidson, A.R. (2000) *Protein Sci.* 9:2170-2180; Demarest, S.J. *et al.* (2004) *J. Mol. Biol.* 335:41- 48) usando el índice de Simpson que es una medida matemática de la diversidad en un sistema que proporciona más información sobre la composición de aminoácidos que simplemente la riqueza. Se calculó el grado de diversidad para cada posición, i , teniendo en cuenta el número de aminoácidos diferentes presentes, así como la abundancia relativa de cada residuo.

$$D = \frac{\sum_{i=1}^r n(n-1)}{N(N-1)}$$

En la que: D es el índice de Simpson, N es el número total de aminoácidos, r es el número de aminoácidos diferentes presentes en cada posición y n es el número de residuos de un tipo de aminoácido particular.

Se examinó la base de datos de QC de las regiones de marco de Fv seleccionadas (seleccionadas mediante el examen de QC) usando diferentes criterios para definir las características únicas. Se usaron las diferentes series en la base de datos de secuencias para definir el grado de variabilidad de las posiciones de residuo dentro de las regiones de marco de Fv y para identificar las posiciones tolerantes a variación no comunes en la naturaleza que están presentes en las regiones de marco de Fv seleccionadas. Se definió como un umbral una diferencia en las puntuaciones de entropía posicional igual al, o de más del, 10%. Se seleccionaron posiciones adicionales si el residuo en una posición dada estaba ocupado por un aminoácido observado infrecuentemente en las otras series de secuencias, es decir, observado infrecuentemente en las bases de datos de línea germinal (VBase e IMGT) y la base de datos KDB. Si se encontraba que el comportamiento de un residuo era realmente diferente (poco o nada representado en cualquiera de las otras series de secuencias), la posición de residuo se definió como única.

El fundamento tras la identificación de las características únicas de las secuencias de la región de marco de Fv seleccionadas son las propiedades superiores demostradas de las regiones de marco y el posible uso de estos hallazgos para la formación de armazones mejorada. Se supuso que las posiciones altamente conservadas en la naturaleza que muestran un determinado grado de variabilidad en las regiones de marco seleccionadas deben tolerar mutagénesis al azar y presentar una probabilidad aumentada de encontrar aminoácidos alternativos superiores al residuo nativo en un formato de scFv. Además, una preferencia pronunciada por un aminoácido no común es una indicación de selección natural hacia determinado residuo. Basándose en estas dos directrices estadísticas se eligieron diferentes residuos dentro de las cadenas pesada y ligera como o bien posiciones flotantes (tolerantes a la variabilidad) o bien como sustituciones preferidas (residuos no habituales).

EJEMPLO 3: Identificación de posiciones de residuo tolerantes a la variabilidad y no habituales

Usando el enfoque de análisis de scFv basado en la secuencia descrito anteriormente en el ejemplo 2, se analizaron tres familias de región variable de cadena pesada (VH3, VH1a y VH1b) y tres familias de región variable de cadena ligera (Vκ1, Vκ3 y Vλ1) para identificar posiciones de aminoácido tolerantes a la variabilidad. En particular, se determinó el grado de diversidad, tal como se calcula usando el índice de Simpson, para cada posición de aminoácido para secuencias dentro de cuatro bases de datos diferentes, Vbase, IMGT, KDB y QC (scFv seleccionados), tal como se describió anteriormente. Se identificaron posiciones de aminoácido de residuo tolerantes a la variabilidad y no habituales basándose en las diferencias en los valores de índice de Simpson en esas posiciones para las bases de datos de línea germinal Vbase e IMGT en comparación con la base de datos de scFv seleccionados, QC. Adicionalmente, para las posiciones de interés identificadas, se identificó el residuo consenso de línea germinal y se determinó la frecuencia de ese residuo consenso en las bases de datos de QC y KDB.

Los resultados del análisis de variabilidad para las familias de región variable de cadena pesada VH3, VH1a y VH1b se muestran a continuación en las tablas 3, 4 y 5, respectivamente. Para cada tabla, las columnas son tal como sigue: columna 1: posición de residuo de aminoácido usando el sistema de numeración de AHo (la conversión al sistema de numeración de Kabat puede llevarse a cabo usando la tabla de conversión expuesta como tabla 1 en el ejemplo 1); columnas 2 a 5: diversidad calculada para cada serie de anticuerpos en la base de datos para la posición de residuo indicada en la columna 1; columna 6: residuo consenso de la familia de línea germinal correspondiente y KDB; columna 7: frecuencia de residuo relativa en la base de datos KDB para el residuo consenso en la columna 6; y columna 8: frecuencia de residuo relativa en base de datos de scFv seleccionados, QC, para el residuo consenso en la columna 6.

Tabla 3: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia VH3.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	en Residuo consenso	f (cons. en KDB)	f (cons. en QC)
1	0,68	0,65	0,50	0,53	E	66,67	53,57
6	1,00	1,00	0,57	0,86	E	92,56	68,97
7	1,00	0,91	0,65	0,93	S	96,33	77,59
89	0,86	0,83	0,55	0,71	L	84,06	70,18
103	0,73	0,76	0,38	0,76	V	86,85	55,36

Tabla 4: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia VH1a.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	en Residuo consenso	f (cons. en KDB)	f (cons. en QC)
1	0,82	0,83	0,62	0,77	Q	86,60	75,00
6	1,00	1,00	0,51	0,74	Q	84,31	58,30
12	1,00	1,00	0,72	0,93	V	96,29	83,30
13	1,00	1,00	0,72	0,86	K	92,59	83,30
14	1,00	1,00	0,60	0,93	K	96,29	75,00
19	1,00	1,00	0,72	1,00	V	100,00	83,30
21	0,83	0,83	0,72	0,96	V	98,14	83,30
90	1,00	1,00	0,47	0,89	Y	94,44	66,60
92	0,83	1,00	0,60	0,93	E	96,29	75,00
95	0,83	0,83	0,49	0,70	S	83,33	66,60
98	1,00	1,00	0,39	0,83	S	90,74	38,30

5 Tabla 5: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia VH1b.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	en Residuo consenso	f (cons. en KDB)	f (cons. en QC)
1	0,82	0,83	0,58	0,92	Q	95,65	70,59
10	0,82	0,83	0,52	0,73	A	85,00	70,59
12	1,00	1,00	0,64	0,86	V	92,59	76,47
13	1,00	1,00	0,52	0,86	K	92,59	70,59
14	1,00	1,00	0,54	0,88	K	93,83	70,59
20	1,00	1,00	0,61	0,86	K	92,59	76,47
21	0,83	0,83	0,47	0,84	V	91,36	64,71
45	0,70	0,83	0,64	0,90	R	95,06	76,47
47	0,83	1,00	0,31	0,95	A	97,53	47,06
50	0,70	0,70	0,48	0,76	Q	86,42	64,71
55	0,83	0,83	0,64	0,82	M	90,12	76,47
77	1,00	1,00	0,64	1,00	R	100,00	76,47
78	0,83	1,00	0,32	0,76	A	86,42	47,06
82	0,45	0,39	0,25	0,36	R	55,56	29,41
86	0,45	0,45	0,37	0,27	I	24,69	17,65
87	0,57	0,70	0,30	0,53	S	70,37	25,00
107	1,00	1,00	0,60	0,90	A	95,00	75,00

Los resultados del análisis de variabilidad para las familias de la región variable de cadena ligera V κ 1, V κ 3 y V λ 1 se muestran a continuación en las tablas 6, 7 y 8, respectivamente. Para cada tabla, las columnas son tal como sigue: columna 1: posición de residuo de aminoácido usando el sistema de numeración de AHo (la conversión al sistema de numeración de Kabat puede llevarse a cabo usando la tabla de conversión expuesta como tabla 1 en el ejemplo o 1); columnas 2 a 5: diversidad calculada para cada serie de anticuerpos en la base de datos para la posición de residuo indicada en la columna 1; columna 6: residuo consenso de la familia de línea germinal correspondiente y KDB; columna 7: frecuencia de residuo relativa en la base de datos KDB para el residuo consenso en la columna 6; y columna 8: frecuencia de residuo relativa en base de datos de scFv seleccionados, QC, para el residuo consenso en la columna 6.

10

15

Tabla 6: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia V κ 1.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	Residuo consenso	f (cons. en KDB)	f (cons. en QC)
1	0,52	0,47	0,61	0,68	D	81,5	23,3
3	0,76	0,72	0,66	0,55	Q	72,0	18,6
4	0,65	0,73	0,57	0,62	M	76,0	23,3
24	0,69	0,72	0,64	0,74	R	85,3	76,7
47	1,00	1,00	0,69	0,88	K	94,0	81,4
50	1,00	1,00	0,60	0,79	R	89,0	76,7
57	1,00	1,00	0,58	0,79	Y	88,6	74,4
91	0,83	0,81	0,70	0,77	L	86,6	81,4
103	0,91	1,00	0,67	0,90	T	81,4	95,7

Tabla 7: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia V κ 3.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	Residuo consenso	f (cons. en KDB)	f (cons. en QC)
2	1,00	1,00	0,72	0,69	I	82,47	83,33
3	1,00	1,00	0,72	0,64	V	77,93	83,33
10	1,00	1,00	0,72	0,93	T	96,19	83,33
12	1,00	1,00	0,72	0,98	S	98,84	83,33
18	1,00	1,00	0,72	0,92	R	95,86	83,33
20	1,00	1,00	0,68	0,95	T	97,30	66,67
56	1,00	1,00	0,72	0,91	I	95,31	83,33
74	1,00	1,00	0,50	0,86	I	92,61	66,67
94	1,00	1,00	0,72	0,82	S	90,29	83,33
101	1,00	1,00	0,50	0,91	F	95,14	66,67
103	1,00	1,00	0,50	0,82	F	90,47	66,67

5 Tabla 8: Análisis de variabilidad de residuos y frecuencias correspondientes del aminoácido consenso identificado en la línea germinal para la familia V λ 1.

Posición de residuo	Línea germinal en IMGT	Línea germinal en VBase	scFV seleccionados en QC	Sec. en KDB	Residuo consenso	f (cons. en KDB)	f (cons. en QC)
1	1,00	1,00	0,45	0,70	Q	81,10	62,50
2	1,00	1,00	0,27	0,73	S	85,13	37,50
4	1,00	1,00	0,60	0,85	L	92,00	75,00
7	1,00	1,00	0,77	0,99	P	99,32	87,50
11	0,59	0,52	0,53	0,51	V	59,88	37,50
14	0,59	0,52	0,49	0,51	A	59,95	31,25
46	1,00	1,00	0,70	0,80	Q	89,00	81,25
53	1,00	1,00	0,49	0,90	K	94,63	68,75
82	1,00	1,00	0,60	0,90	K	94,88	75,00
92	0,59	0,68	0,51	0,54	A	69,82	68,75
103	1,00	1,00	0,50	0,86	D	92,84	68,75

10 Tal como se expone en las tablas 3-8 anteriores, se encontró que un subconjunto de posiciones de residuo en las regiones de marco de scFv seleccionadas del sistema de QC estaban fuertemente sesgadas hacia determinados residuos no presentes o insuficientemente representados en las líneas germinales (VBase e IMGT) y en anticuerpos maduros (KDB), lo que sugiere que la estabilidad de scFv puede mejorarse racionalmente basándose en las características únicas de las secuencias de región de marco seleccionadas en el sistema de examen en levaduras de control de calidad.

EJEMPLO 4: Selección de residuos preferidos

15 Con el fin de seleccionar sustituciones de residuo de aminoácido preferidas (o, alternativamente, excluir residuos de aminoácido) en una posición de aminoácido particular que se sabe que mejora las propiedades funcionales (por ejemplo, estabilidad y/o solubilidad) de un scFv, se agruparon las secuencias de V_H y V_L de la base de datos de Kabat de secuencias de anticuerpo maduro según su subtipo de familia (por ejemplo, VH1b, VH3, etc.). Dentro de cada subfamilia de secuencias, se determinó la frecuencia de cada residuo de aminoácido en cada posición de aminoácido como un porcentaje de todas las secuencias analizadas de un grupo de subtipos. Se hizo lo mismo para

5 todas las secuencias de la base de datos de QC que consistían en anticuerpos que se preseleccionaron por su estabilidad y/o solubilidad potenciada por el sistema denominado de QC. Para cada subtipo, se compararon los porcentajes resultantes (frecuencias relativas) para cada residuo de aminoácido obtenido para las secuencias de Kabat y para las secuencias de QC en cada posición correspondiente. En el caso de que la frecuencia relativa de un determinado residuo de aminoácido estuviera aumentada en la base de datos de QC en relación con la base de datos de Kabat, el residuo respectivo se consideró un residuo preferido en la posición dada para mejorar la estabilidad y/o solubilidad de un scFv. A la inversa, en el caso de que la frecuencia relativa de un determinado residuo de aminoácido estuviera disminuida en la base de datos de QC en comparación con la base de datos de Kabat, el residuo respectivo se consideró desfavorable en esa posición en el contexto de un formato de scFv.

10 La tabla 9 representa un análisis a modo de ejemplo de la frecuencia de residuo en la posición de aminoácido H78 (numeración de AHo; posición de Kabat H67) para el subtipo VH1b en las diferentes bases de datos. Las columnas en la tabla 9 son tal como sigue: columna 1: tipo de residuo; columna 2: frecuencia de residuo en la base de datos de línea germinal IMGT; columna 3: frecuencia de residuo en la base de datos de línea germinal Vbase; columna 4: frecuencia de residuo en una base de datos de QC; columna 5: frecuencia de residuo en una base de datos de Kabat.

15 Tabla 9: Frecuencia de residuo relativa en la posición 78 (numeración de AHo) para el subtipo VH1b en dos bases de datos de línea germinal, una base de datos de QC y una base de datos de Kabat de anticuerpos maduros.

Residuo	IMGT_germ.	Vbase_germ.	Base de datos de QC	KDB_VH1B
D				
E				
K				
R				
H				
T				
S				
N				
Q				
G				
A			24	2
C				
P				
V	91	100	47	86
I			18	1
L			12	
M				
F	9			10
Y				
W				
Consenso	V	V	V	V
% de coincidencia	91	100	47	86
N.º de sec.*	11	11	17	81

*Número de secuencias recogidas para el análisis de frecuencia de residuo

20 En la base de datos de QC, se observó un residuo de alanina (A) a una frecuencia del 24%, un factor de 12 por encima de la frecuencia del 2% observada para el mismo residuo en una base de datos de Kabat de anticuerpos maduros (KDB_VH1B). Por consiguiente, un residuo de alanina en la posición H78 (numeración de AHo) se considera un residuo preferido en esa posición para potenciar las propiedades funcionales (por ejemplo, estabilidad y/o solubilidad) de un scFv. En contraposición, se observó un residuo de valina (V) en la base de datos de QC a una frecuencia relativa del 47%, mucho menor que la frecuencia del 86% observada en la base de datos de Kabat de anticuerpos maduros y la frecuencia de más del 90% observada para el mismo residuo en las bases de datos de línea germinal (el 91% en IMGT-germ y el 100% en Vbase-germ). Por tanto, se consideró que un residuo de valina (V) era un residuo desfavorable en la posición H78 en el contexto de un formato de scFv.

EJEMPLO 5: Comparación de variantes de scFv de ESBA105 a partir de dos enfoques diferentes

30 En este ejemplo, se comparó la estabilidad de variantes de scFv preparadas mediante dos enfoques diferentes. El anticuerpo scFv original fue ESBA 105, que se ha descrito anteriormente (véanse por ejemplo, las publicaciones PCT WO 2006/131013 y WO 2008/006235). Se seleccionó un conjunto de variantes de ESBA105 usando el sistema de examen en levaduras de control de calidad ("variantes de QC"), variantes que también se han descrito anteriormente (véanse por ejemplo, las publicaciones PCT WO 2006/131013 y WO 2008/006235). El otro conjunto de variantes se preparó mediante la retromutación de determinadas posiciones de aminoácido a la secuencia consenso de línea germinal preferida identificada mediante el análisis de secuencia descrito en los ejemplos 2 y 3 anteriores. Las retromutaciones se seleccionaron buscando dentro de las secuencias de aminoácidos posiciones que estaban conservadas en la secuencia de línea germinal, pero que contenían un aminoácido de frecuencia baja o

no habitual en el scFv seleccionado (denominado enfoque de modificación por ingeniería genética de consenso de línea germinal).

5 Se sometieron a prueba todas las variantes para determinar la estabilidad sometiendo las moléculas a una tensión inducida térmica. Mediante la exposición a un amplio intervalo de temperaturas (25-95°C) fue posible determinar puntos medios aproximados de las transiciones de desplegamiento térmico (TM) para cada variante. Se realizaron mediciones de termoestabilidad para las moléculas silvestres y las variantes con la espectroscopía FT-IR ATR en la que se guió luz IR a través de un interferómetro. La señal medida es el interferograma, realizando una transformada de Fourier en esta señal, el espectro final es idéntico al de la espectroscopía infrarroja convencional (dispersiva).

10 Los resultados del desplegamiento térmico se resumen a continuación en la tabla 10 y se representan gráficamente en la figura 6. Las columnas en la tabla 10 son tal como sigue: columna 1: variantes de ESBA105; columna 2: dominio que contiene la mutación; columna 3: mutación/mutaciones en la numeración de AHo; columna 4: puntos medios de TM calculados a partir de las curvas de desplegamiento térmico en la figura 6; columna 5: actividad relativa en comparación con ESBA105 original; columna 5: estrategia de mutagénesis para la variante especificada en la columna 1.

15 Tabla 10: Comparación de variantes de ESBA105 a partir de dos enfoques diferentes y su contribución a la estabilidad global medida en FT-IR (puntos medios calculados para las transiciones de desplegamiento térmico).

Variante	Dominio	Mutación	TM °C	Actividad	Descripción
E105			61,53		Molécula original
ESBA105_QC11.2	VH	F78L	66,26	1	Variante de QC
ESBA105_QC15.2	VH	K50R, F78I	65,47	1	Variante de QC
ESBA105_QC23.2	VH	F78L	66,53	1	Variante de QC
ESBA105_VL R47K	VL	R47K	62,4	0,9	Retromutada a consenso
ESBA105_VL V103T	VL	V103T	60,7	1	Retromutada a consenso
ESBA105_VL V3Q	VL	V3Q	61,9	1,2	Retromutada a consenso

20 En comparación con las variantes de QC, las retromutaciones a consenso de línea germinal tenían un efecto negativo o ningún efecto sobre la termoestabilidad y la actividad de ESBA105. Por tanto, estos resultados contradicen el enfoque de modificación por ingeniería genética de consenso que han usado otros para mejorar la estabilidad en diferentes anticuerpos y formatos (véanse por ejemplo, Steipe, B *et al.* (1994) J. Mol. Biol. 240:188-192; Ohage, E. y Steipe, B. (1999) J. Mol. Biol. 291:1119-1128; Knappik, A. *et al.* (2000) J. Mol. Biol. 296:57-86, Ewert, S. *et al.* (2003) Biochemistry 42:1517-1528; y Monsellier, E. y Bedouelle, H. (2006) J Mol. Biol. 362:580-593).

25 En un experimento separado, se compararon las variantes de QC anteriores (QC11.2, QC15.2, y QC23.2) y una variante de QC adicional (QC7.1) con variantes de un segundo conjunto que tenían o bien retromutaciones consenso (S-2, D-2 y D-3) o bien una retromutación a alanina (D-1) (véase la tabla 11). La identidad del residuo en las posiciones de región de marco seleccionadas se indica en la tabla 11 y la estabilidad térmica medida (en unidades de desplegamiento arbitrarias) se representa en la figura 7. Aunque algunas variantes de consenso (S-2 y D-1) mostraron un marcado aumento en la estabilidad térmica, este aumento fue inferior al aumento en la estabilidad térmica logrado por cada una de las cuatro variantes de QC.

30 Tabla 11: Se proporciona la identidad de los residuos de la región de marco en posiciones de región de marco seleccionadas de variantes de ESBA105 que comprenden o bien retromutaciones consenso (S-2, D-2, D-3), una mutación a alanina (D-1) o bien un residuo de QC (QC7.1, QC11.2, QC15.2, QC23.2). Los residuos que difieren del anticuerpo ESBA105 original se representan en cursiva y negrita. Las posiciones de aminoácido se proporcionan en la numeración de Kabat.

	Superficie de contacto de VL-CL	Superficie de contacto de VH-CH	Bucle exterior	Bucle exterior	Superficie de contacto de VH-CH
	L83	H43	H67	H69	H78
Original	V	K	F	F	V
QC7.1	<i>E</i>	K	F	F	<i>A</i>
QC11.2	V	K	<i>L</i>	F	V
QC15.2	V	<i>R</i>	<i>I</i>	F	V
QC23.2	V	K	<i>L</i>	F	V
S-2	V	K	<i>V</i>	F	V
D-1	V	K	<i>A</i>	F	V
D-2	V	K	<i>V</i>	<i>L</i>	V
D-3	V	K	F	<i>L</i>	V

Por consiguiente, los resultados en el presente documento demuestran que la presión de selección aplicada en el “sistema de examen en levaduras de control de calidad” produce una subpoblación de armazones que sí contienen características comunes observadas rara vez en la naturaleza (aún menos en seres humanos) y que presumiblemente son responsables de las propiedades biofísicas superiores de estas regiones de marco. Al exponer a 60°C diferentes variantes de ESBA105, fue posible volver a confirmar las propiedades superiores de las sustituciones preferidas identificadas en la base de datos de regiones de marco de scFv seleccionadas. Por tanto, se ha demostrado que el enfoque de “consenso funcional” descrito en el presente documento basado en las secuencias de scFv seleccionadas obtenidas del sistema de examen en levaduras de QC produce variantes de scFv que tienen estabilidad térmica superior a las variantes preparadas usando el enfoque de consenso de línea germinal.

10 EJEMPLO 6: Variantes de scFv de ESBA212

En este ejemplo, se comparó la estabilidad de variantes de consenso de línea germinal de un anticuerpo scFv (ESBA212) con una especificidad de unión diferente a la de ESBA105. Se prepararon todas las variantes de ESBA212 mediante la retromutación de determinadas posiciones de aminoácido a la secuencia consenso de línea germinal preferida identificada mediante el análisis de secuencia descrito en los ejemplos 2 y 3 anteriores. Las retromutaciones se seleccionaron buscando dentro de las secuencias de aminoácidos posiciones que estaban conservadas en la secuencia de línea germinal, pero que contenían un aminoácido de frecuencia baja o no habitual en el scFv seleccionado (denominado enfoque de modificación por ingeniería genética de consenso de línea germinal). Como en el ejemplo 5, se sometieron a prueba todas las variantes para determinar la estabilidad sometiendo las moléculas a una tensión inducida térmica.

20 Los resultados del desplegamiento térmico para las variantes de ESBA212 se resumen a continuación en la tabla 12 y se representan gráficamente en la figura 8. Las columnas en la tabla 12 son tal como sigue: columna 1: variantes de ESBA212; columna 2: dominio que contiene la mutación; columna 3: mutación/mutaciones en la numeración de AHo; columna 4: puntos medios de TM calculados a partir de las curvas de desplegamiento térmico en la figura 7; columna 5: actividad relativa en comparación con ESBA212 original; columna 5: estrategia de mutagénesis para la variante especificada en la columna 1.

Tabla 12: Comparación de variantes de ESBA212 retromutadas a residuo consenso de línea germinal y su contribución a la estabilidad global medida en FT-IR (puntos medios calculados para las transiciones de desplegamiento térmico).

Variante	Dominio	Mutación	TM °C	Actividad	Descripción
ESBA212			63,66		Molécula original
ESBA212_VL R47K	VL	R47K	59,94	2,8	Retromutada a consenso
ESBA212_VL V3Q	VL	V3Q	63,6	1,1	Retromutada a consenso

30 Tal como se observa para el anticuerpo scFv de ESBA105 no relacionado, las retromutaciones a consenso de línea germinal tenían un efecto negativo o ningún efecto sobre la termoestabilidad y la actividad de ESBA212. Por tanto, estos resultados sirven para destacar adicionalmente lo inadecuado de los enfoques basados en consenso convencionales. Estas deficiencias pueden abordarse empleando la metodología de consenso funcional de la invención.

EJEMPLO 7: Generación de scFv con solubilidad mejorada

35 En este ejemplo, se usó un enfoque basado en análisis de secuencia y modelado estructural para identificar mutaciones en regiones de marco de scFv que dan como resultado solubilidad mejorada.

a) Análisis estructural

40 Se modeló la estructura 3D de scFv de ESBA105 usando el servidor de modelado de homología de estructura de proteínas automatizado, accesible a través del servidor web ExPASy. Se analizó la estructura según la superficie relativa accesible al disolvente (rSAS) y se clasificaron los residuos tal como sigue: (1) Expuestos para residuos que muestran una rSAS $\geq 50\%$; y (2) parcialmente expuestos para residuos con $50\% \leq rSAS \leq 25\%$. Los residuos hidrófobos con una rSAS $\geq 25\%$ se consideraron parches hidrófobos. Para validar el área accesible al disolvente de cada parche hidrófobo encontrado, se realizaron cálculos a partir de 27 archivos PDB con alta homología a ESBA105 y una resolución superior a 2,7 Å. Se calcularon la rSAS promedio y la desviación estándar para los parches hidrófobos y se examinaron en detalle para cada uno de ellos (véase la tabla 13).

Tabla 13: Evaluación de los parches hidrófobos.

Residuo	Dominio	Superficie expuesta al disolvente %	STDE %	rSAS	Variabilidad de secuencia	Superficie de contacto de VH/antígeno	Superficie de contacto de VH/VL	Superficie de contacto de VH/CH
2	VH	23,06	19,26	10-25%	10-25%	>0-20%	>0-20%	0
4	VH	0,66	1,26	0-10%	0-10%	0		0
5	VH	61,85	12,96	50-75%	10-25%	0	>0-20%	0
12	VH	70,27	9,17	50-75%	10-25%	0	0	60-80%
103	VH	35,85	5,85	25-50%	10-25%	0	>0-2%	>0-2%
144	VH	62,17	7,82	50-75%	10-25%	0	0	>0-2%
15	VL	49,59	9,77	25-50%	10-25%	0	0	0
147	VL	31,19	23,32	25-50%	10-25%	0	0	60-80%

Columna 1, posición de residuo en el sistema de numeración de AHo. Columna 2, dominio para la posición indicada en la columna 1. Columna 3, Cálculos de área accesible al disolvente promedio a partir de 27 archivos PDB. Columna 4, desviaciones estándar de la columna 3. Columnas 5 a 9, papel estructural de los parches hidrófobos recuperados a partir de AHo.

La mayoría de los parches hidrófobos identificados en ESBA105 correspondían a la superficie de contacto de dominio variable-constante (VH/CH). Esto se correlacionó con los hallazgos anteriores de residuos hidrófobos expuestos a disolvente en un formato de scFv (Nieba *et al.*, 1997). Dos de los parches hidrófobos (VH2 y VH5) también contribuyeron a la interacción VL-VH y por tanto se excluyeron del análisis posterior.

b) Diseño de mutaciones de solubilidad

Se recuperaron un total de 122 secuencias de VL y 137 secuencias de VH de la página web de anticuerpos de Annemarie Honegger (www.bioc.uzh.ch/antibody). Las secuencias correspondían originalmente a 393 estructuras de anticuerpo en formato Fv o Fab extraídas del Banco de datos de proteínas (PDB) (www.rcsb.org/pdb/home/home.do). Se usaron las secuencias para el análisis independientemente de la especie o el subgrupo con el fin de aumentar la probabilidad de encontrar aminoácidos alternativos con mayor hidrofilia que el residuo nativo. Se excluyeron las secuencias que tenían una identidad de más del 95% con cualquier otra secuencia dentro de la base de datos para reducir el sesgo. Se alinearon las secuencias y se analizaron para determinar la frecuencia de residuos. Se aplicaron herramientas y algoritmos de análisis de secuencias para identificar y seleccionar mutaciones hidrófilas para alterar los parches hidrófobos en ESBA 105. Se alinearon las secuencias siguiendo el sistema de numeración de AHo para el dominio variable de inmunoglobulina (Honegger y Plückthun 2001). Se restringió el análisis a las regiones de marco.

Se calculó la frecuencia de residuos, $f(r)$, para cada posición, i , en la base de datos personalizada por el número de veces que se observa ese residuo particular dentro del conjunto de datos dividido entre el número total de secuencias. En una primera etapa, se calculó la frecuencia de aparición de los diferentes aminoácidos para cada parche hidrófobo. Se analizó la frecuencia de residuo para cada parche hidrófobo identificado en ESBA105 a partir de la base de datos personalizada descrita anteriormente. La tabla 14 notifica la frecuencia de residuo en los parches hidrófobos dividida entre la totalidad de los residuos presentes en la base de datos.

Tabla 14. Frecuencia de residuo de 259 secuencias de anticuerpos maduros en un formato scFv o Fab para los parches hidrófobos identificados en ESBA105

Residuo	VH 4	VH 12	VH 103	VH 144	VL 15	VL 147
A	0,23046215	0	0	0	3,8647343	0,176821923
C	0	0	0	0	0	0
D	0	0	0	0	0	0
E	0	0	0	0	0	0
F	0,483091787	0	0,483091787	0	0	0
G	0	0	0	0	0	0
H	0	0	0	0	0	0
I	0	2,415458937	9,661835749	0	5,314009662	70,38834951
K	0	0	0	0	0	0
L	96,61835749	89,85507246	7,246376812	27,0531401	45,89371981	15,53398058
M	0	0	10,62801932	1,93236715	0	0,970873786
N	0	0	0	0	0	0
P	0,966183575	0	0	0,966183575	21,73913043	0,485436893
Q	0	0	0	0,483091787	0	0
R	0	0	7,246376812	0	0	0
S	0	0,966183575	0	18,84057971	0	0

T	0	0	15,4589372	50,72463768	0,966183575	0
V	1,93236715	6,763285024	49,27536232	0	22,22222222	12,62135922
W	0	0	0	0	0	0
Y	0	0	0	0	0	0

Columna 1, tipo de residuo. Columnas 2 a 5, frecuencia relativa de residuos para los parches hidrófobos en la cadena pesada. Columna 6 y 7, frecuencia relativa de residuos para los parches hidrófobos en la cadena ligera.

5 En la segunda etapa, se usó la frecuencia de residuos hidrófilos en los parches hidrófobos para diseñar las mutaciones de solubilidad seleccionando el residuo hidrófilo más abundante en cada parche hidrófobo. La tabla 15 notifica los mutantes de solubilidad identificados usando este enfoque. Se calculó la hidrofobia de los residuos originales y mutantes como la hidrofobia promedio de los valores publicados en varios artículos y se expresó en función del nivel de exposición de la cadena lateral al disolvente.

Tabla 15. Diferentes mutaciones de solubilidad introducidas en ESBA105 para alterar los parches hidrófobos

Residuo	Dominio	Superficie expuesta al disolvente %	Residuo original	Hidrofobia del residuo original	Mutación de solubilidad	Hidrofobia de las mutaciones
4	VH	0,66	L	85,2	A	42,7
12	VH	70,27	V	73,2	S	28
103	VH	35,85	V	73,2	T	32,8
144*	VH	62,17	V	73,2	S	28
15	VL	49,59	V	73,2	T	32,8
147	VL	31,19	L	85,2	A	42,7

*Se cambió el parche hidrófobo en la posición 144, no por el residuo hidrófilo más abundante en la base de datos, sino por Ser, puesto que ésta estaba ya contenida en el donador de CDR de ESBA105.

Columna 1, posición de residuo en el sistema de numeración de AHo. Columna 2, dominio para la posición indicada en la columna 1. Columna 3, cálculos de área accesible al disolvente promedio a partir de 27 archivos PDB. Columna 4, residuos originales en ESBA105. Columna 5, hidrofobias promedio de la columna 4, recuperadas de AHo. Columnas 6, residuo hidrófilo más abundante en la posición indicada en la columna 1. Hidrofobia promedio de la columna 6 recuperada de AHo.

c) Pruebas de solubilidad de las variantes de ESBA105

10 Se introdujeron las mutaciones de solubilidad solas o en múltiples combinaciones y se sometieron a prueba para determinar el rendimiento de plegamiento, la expresión, la actividad y los patrones de estabilidad y agregación. La tabla 16 muestra las diversas combinaciones de mutaciones de solubilidad introducidas en cada variante optimizada de ESBA105 basándose en la posible contribución a la solubilidad y en el nivel de riesgo con que la mutación alterare la unión a antígeno.

15 Tabla 16: Diseño de variantes de solubilidad para ESBA105.

Residuo de superficie hidrófobo	Dominio	Residuo original	Mutantes**			
			Opt 1_0	Opt 0_2	Opt 1_2	Opt 2_4
15	VL	V	X		X	X
147*	VL	V				X
4*	VH	L				X
12	VH	V		X	X	X
103*	VH	V				X
144	VH	L		X	X	X

*Sometidos a prueba por separado en una segunda ronda

**El guión bajo separa el número de mutaciones contenidas en la cadena ligera y pesada, respectivamente.

Columna 1, posición de residuo en el sistema de numeración de AHo. Columna 2, dominio para la posición indicada en la columna 1. Columna 3, residuo original en ESBA105 en los parches hidrófobos diferentes. Columna 4, variantes diferentes que contienen mutaciones de solubilidad en las posiciones indicadas.

i. Mediciones de solubilidad

5 Se determinaron las solubilidades máximas de ESBA105 y variantes midiendo la concentración de proteínas en los sobrenadantes de mezclas de PEG-proteína centrifugadas. Se mezcló una concentración de partida de 20 mg/ml 1:1 con disoluciones de PEG que oscilaban entre el 30 y el 50% de saturación. Se eligieron estas condiciones basándose en el perfil de solubilidad observado para ESBA105 silvestre tras la determinación empírica de la dependencia lineal de Log S frente a la concentración de PEG (% p/v). En la figura 9 se representan curvas de solubilidad de varios ejemplos de ESBA105 variante que mostraron solubilidad superior. También se proporciona una lista completa de valores de solubilidad en la tabla 17.

Tabla 17. Actividad y solubilidad máxima estimada de los mutantes en comparación con ESBA105 original.

Molécula	E 105	E105 Opt 1_0	E105 Opt 0_2	E105 Opt 1_2	E105 VH V103T	E105 VL V147A
Punto de corte	1,956	2,228	2,179	2,163	2,223	2,047
Solubilidad máxima	90,36	169,04	151,01	145,55	167,11	111,43
Actividad en relación con ESBA105	1	1,4	1,5	1,5	1,2	2

ii. Mediciones de termoestabilidad

10 Se realizaron mediciones de termoestabilidad para ESBA105 original y seguimientos de solubilidad usando espectroscopía FT-IR ATR. Se expusieron térmicamente las moléculas a un amplio intervalo de temperaturas (de 25 a 95°C). Se obtuvo el perfil de desnaturalización aplicando una transformada de Fourier para las señales del interferograma (véase la figura 10). Se usaron los perfiles de desnaturalización para aproximar los puntos medios de las transiciones de desplegamiento térmico (TM) para cada variante de ESBA105 aplicando el modelo sigmoideo de Boltzmann (tabla 18).

Tabla 18: Puntos medios de las transiciones de desplegamiento térmico (TM) para cada variante de solubilidad.

	ESBA105	E105 Opt1.0	E105 Opt1.2	E105 Opt0.2	E105 VH V103T	E105 VL V147A
Sigmoideo de Boltzmann						
Valores de ajuste óptimo						
INFERIOR	0,3604	-0,405	0,7032	0,4516	0,4691	-0,6873
SUPERIOR	100,4	99,3	98,84	99,04	99,2	99,16
V50	61,53	59,91	59,39	60,86	62,08	55,89
PENDIENTE	2,935	2,886	3,117	2,667	2,682	3,551
Error estándar						
INFERIOR	0,5206	0,3471	0,6652	0,4953	0,3938	0,4754
SUPERIOR	0,5361	0,3266	0,6116	0,4891	0,4167	0,3714
V50	0,1047	0,06658	0,1328	0,0949	0,07811	0,0919
PENDIENTE	0,09039	0,05744	0,1146	0,08199	0,06751	0,08235
Intervalos de confianza del 95%						
INFERIOR	de -0,7432 a 1,464	de -1,141 a 0,3309	de -0,7071 a 2,114	de 0,5984 a 1,502	de -0,3658 a 1,304	de -1,695 a 0,3206
SUPERIOR	de 99,25 a 101,5	de 98,61 a 99,99	de 97,54 a 100,1	de 98,01 a 100,1	de 98,32 a 100,1	de 98,38 a 99,95
V50	de 61,31 a 61,75	de 59,77 a 60,06	de 59,11 a 59,67	de 60,66 a 61,06	de 61,91 a 62,24	de 55,70 a 56,09
PENDIENTE	de 2,743 a 3,127	de 2,764 a 3,007	de 2,874 a 3,360	de 2,494 a 2,841	de 2,539 a 2,825	de 3,376 a 3,725
Bondad del ajuste						
Grados de libertad	16	16	16	16	16	16
R ²	0,9993	0,9997	0,999	0,9994	0,9996	0,9996
Suma absoluta de cuadrados	26,18	10,8	37,2	24	16,14	15,11
Sy.x	1,279	0,8217	1,525	1,225	1,004	0,9719

iii. Mediciones de agregación

5 También se analizaron ESBA105 y sus variantes de solubilidad en una prueba dependiente del tiempo para evaluar el comportamiento de degradación y agregación. Para este fin, se incubaron proteínas solubles (20 mg/ml) a una temperatura elevada (40°C) en tampones fosfato a pH 6,5. Se mantuvieron muestras control a -80°C. Se analizaron las muestras tras un periodo de incubación de dos semanas para determinar la degradación (SDS-PAGE) y la agregación (SEC). Esto permitió descartar las variantes que eran propensas a la degradación (véase la figura 11) o mostraban tendencia a formar agregados solubles o insolubles (véase la tabla 19).

Tabla 19: Mediciones de agregación insoluble.

Proteína	Pérdida de proteína (agregados insolubles)
ESBA105	1,14%
ESBA105 Opt 1_0	8,17%
ESBA105 Opt 0_2	4,45%
ESBA105 Opt 1_2	46,60%
ESBA105 VH V103T	-1,95%

iv. Expresión y replegamiento de variantes de solubilidad

10 También se sometieron a prueba los mutantes de solubilidad para determinar la expresión y el rendimiento de replegamiento en relación con la molécula de ESBA105 original. Los resultados de estos estudios se muestran en la tabla 20.

Tabla 20. Expresión y replegamiento de variantes de solubilidad.

	Residuo de superficie hidrófobo							Expresión con relación a ESBA105	Rendimiento de replegamiento mg/l
	VH				VL				
	L4	V12	V103	L144	V15	F52	V147		
ESBA105								1,0	34
Opt 1_0					T			1,15	12,5
Opt 0_2		S		S				1,10	35
Opt 1_2		S		S	T			0,96	44
Opt 2_4	A	S	T	S	T		A	1,20	No puede producirse
VH L4A	A							1,0	No puede producirse
VH V103T			T					1,1	55
VL V147A							A	1,2	20

15 Aunque todos los mutantes de solubilidad hidrófilos mostraron solubilidad mejorada en comparación con la molécula de ESBA105 original, sólo algunas de estas moléculas mostraron ser adecuadas para otras propiedades biofísicas. Por ejemplo, muchas variantes tenían una termoestabilidad y/o rendimiento de replegamiento reducidos en relación con la molécula de ESBA105 original. En particular, la sustitución hidrófila en la posición VL147 disminuyó gravemente la estabilidad. Por tanto, se combinaron las mutaciones de solubilidad que no afectaban significativamente a la estabilidad térmica y se sometieron a tensión térmica adicional para confirmar sus propiedades.

20 Tres mutantes que contenían una combinación de cuatro mutaciones de solubilidad diferentes (Opt1.0, Opt0.2 y VH:V103T) mejoraron significativamente la solubilidad de ESBA105 sin afectar a la reproducibilidad, la actividad o la estabilidad térmica. Sin embargo, un mutante que tenía las mutaciones combinadas de Opt1.0 y Opt0.2 en ESBA105 (Opt 1_2) mostró una cantidad aumentada de agregados insolubles tras la incubación durante 2 semanas a 40°C (véase la tabla 19). Esto podría explicarse por el papel de la Val en la posición VL15 en un giro de lámina beta, puesto que la Val tiene la mayor propensión a lámina beta de todos los aminoácidos. Este resultado demostró que se tolera una única mutación de solubilidad en la posición VL 15, pero no en combinación con mutantes de solubilidad que alteran otros parches hidrófobos. Por tanto, se seleccionaron las mutaciones contenidas en Opt0_2 y VH:V103T como las mejores para mejorar las propiedades de solubilidad de las moléculas de scFv.

30 **EJEMPLO 8:** Generación de solubilidad y estabilidad potenciadas de scFv

35 Se optimizaron adicionalmente variantes de ESBA105 identificadas por diseño de solubilidad mediante sustitución con mutaciones de estabilización identificadas por el ensayo de control de calidad (QC). Se crearon un total de 4 constructos que contenían entre 1 y 3 de las mutaciones de solubilidad identificadas en el ejemplo 7 anterior, en combinación con todas las mutaciones de estabilización encontradas en QC 7.1 y 15.2 (es decir, D31N y V83E en el dominio VL y V78A, K43 y F67L en el dominio VH). Todos los constructos optimizados produjeron más proteína soluble que un scFv silvestre (véase la tabla 20). El mejor constructo mostró de manera constante un aumento de más de 2 veces en la solubilidad con respecto al tipo silvestre. Ni la actividad ni la estabilidad de las moléculas de scFv resultaron significativamente afectadas por la combinación de mutaciones que potencian la solubilidad y de

estabilización.

Tabla 21: ScFv con estabilidad y solubilidad optimizadas

Proteína	Mutaciones VL/VH	FTIR Tm (°C)	Solubilidad de PEG (mg/ml)	Actividad en relación con E105	kD
QC7.1D-N-15.2	VL: D31N; V83E VH: V78A; K43R; F67L	69,0	90	1,7	$9,06 \times 10^{-10}$
QC7.1D-N-15.2 VH V103T	VL: D31N; V83E VH: V78A; K43R; F67L; V103T	68,9	106	1,5	$8,79 \times 10^{-10}$
QC7.1D-N-15.2 Opt0_2	VL: D31N; V83E VH: V12S; V78A; K43R; F67L; L144S	66,5	121	1,2	$8,12 \times 10^{-10}$
QC7.1D-N-15.2 VH V103T Opt0_2	VL: D31N; V83E VH: V12S; V78A; K43R; F67L; V103T; L144S	67,3	186	1,5	$1,34 \times 10^{-9}$

Se usaron los valores de solubilidad para las 4 variantes para determinar la contribución de cada mutación a la solubilidad de scFv. Todas las mutaciones parecieron contribuir a la solubilidad del scFv de manera aditiva aunque varios de esos residuos estaban relativamente próximos entre sí tanto en la secuencia primaria como dentro de la estructura 3D. El análisis indicó que una combinación de tres mutaciones que potencian la solubilidad en el dominio VH (V12S, L144S, V103T (o V103S)) representa el ~60% de la solubilidad de scFv. Puesto que los parches hidrófobos están conservados en los dominios variables de todos los agentes de unión inmunológica, esta combinación de mutaciones óptima puede usarse para mejorar la solubilidad de prácticamente cualquier scFv u otra molécula de unión inmunológica.

5

10

REIVINDICACIONES

1. Método para identificar uno o más residuos de aminoácido para sustitución en una posición particular en un agente de unión inmunológica, comprendiendo el método:
 - 5 a) proporcionar una primera base de datos de secuencias de aminoácidos de V_H y/o V_L agrupadas según el subtipo;
 - b) proporcionar una segunda base de datos de secuencias de aminoácidos de V_H y/o V_L agrupadas según el subtipo y seleccionadas por tener al menos una propiedad funcional deseable, en el que la propiedad funcional deseable es estabilidad mejorada, solubilidad mejorada, no agregación, una mejora en la expresión y/o una mejora en el rendimiento de repliegamiento;
 - 10 c) determinar la frecuencia de aminoácidos para un residuo de aminoácido en una posición de región de marco de la primera base de datos y en una posición de región de marco correspondiente de la segunda base de datos;
 - 15 d) determinar el factor de enriquecimiento (RF2:RF1) del aminoácido estableciendo la razón entre la frecuencia relativa de un residuo dentro de la segunda base de datos (RF2) y la frecuencia relativa de un residuo dentro de la primera base de datos (RF1)
 - e) identificar el residuo de aminoácido
 - (i) como un residuo de aminoácido preferido para sustitución en la posición de aminoácido correspondiente del agente de unión inmunológica cuando el aminoácido tiene un factor de enriquecimiento de más de 1, o
 - 20 (ii) como un residuo de aminoácido que va a excluirse en la posición de aminoácido correspondiente del agente de unión inmunológica cuando el residuo de aminoácido tiene un factor de enriquecimiento de menos de 1.
2. Método según la reivindicación 1, en el que el residuo de aminoácido de la etapa d) tiene un factor de enriquecimiento de 4 a 6.
- 25 3. Método según una cualquiera de las reivindicaciones anteriores, en el que la primera base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y V_L de línea germinal.
4. Método según una cualquiera de las reivindicaciones 1-3, en el que la primera base de datos consiste en las secuencias de aminoácidos de V_H , V_L o V_H y V_L de línea germinal.
- 30 5. Método según una cualquiera de las reivindicaciones anteriores, en el que la primera base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras.
6. Método según una cualquiera de las reivindicaciones 1-5, en el que la primera base de datos consiste en secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras.
7. Método según la reivindicación 5 ó 6, en el que las secuencias de aminoácidos de V_H , V_L o V_H y V_L maduras son de la base de datos Kabat (KDB).
- 35 8. Método según una cualquiera de las reivindicaciones anteriores, en el que la segunda base de datos comprende secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpos scFv seleccionadas a partir de un ensayo de QC.
9. Método según una cualquiera de las reivindicaciones 1-7, en el que la segunda base de datos consiste en secuencias de aminoácidos de V_H , V_L o V_H y V_L de anticuerpo scFv seleccionadas a partir de un ensayo de QC.
- 40 10. Método de mejora de un agente de unión inmunológica, teniendo el agente de unión inmunológica secuencias de aminoácidos de V_H y/o V_L , comprendiendo el método:
 - a) identificar una o más posiciones de aminoácido de región de marco para mutación;
 - 45 b) identificar para cada posición de región de marco particular, tal como se identifica en la etapa a), un residuo de aminoácido preferido para sustitución según el método según una cualquiera de las reivindicaciones 1 a 9; y
 - c) mutar el residuo de aminoácido en cada posición de región de marco particular hacia el residuo de aminoácido preferido identificado en la etapa b)

en el que la etapa a) se realiza asignando un grado de conservación a cada posición de región de marco

usando el índice de Simpson, y el índice de Simpson para la posición de aminoácido en la segunda base de datos es al menos 0,01 menor que el valor índice de Simpson de la posición de aminoácido correspondiente en la primera base de datos.

- 5 11. Método según la reivindicación 10, en el que si el residuo de aminoácido seleccionado para mutación es un residuo de aminoácido de cadena pesada, está en una posición seleccionada del grupo que consiste en 1, 6, 7, 10, 12, 13, 14, 19, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87, 89, 90, 92, 95, 98, 103 y 107 usando la numeración AHo (posiciones de aminoácido 1, 6, 7, 9, 11, 12, 13, 18, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76, 78, 79, 81, 82b, 84, 89 y 93 usando la numeración Kabat), y
- 10 en el que si el residuo de aminoácido seleccionado para mutación es un residuo de aminoácido de cadena ligera, está en una posición seleccionada del grupo que consiste en 1, 2, 3, 4, 7, 10, 11, 12, 14, 18, 20, 24, 46, 47, 50, 53, 56, 57, 74, 82, 91, 92, 94, 101 y 103 usando la numeración AHo (posiciones de aminoácido 1, 2, 3, 4, 7, 10, 11, 12, 14, 18, 20, 24, 38, 39, 42, 45, 48, 49, 58, 66, 73, 74, 76, 83 y 85 usando la numeración Kabat).
- 15 12. Método según la reivindicación 11, en el que el agente de unión inmunológica comprende una región variable de cadena pesada VH3 y la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 6, 7, 89 y 103 de VH3 usando la numeración AHo (posiciones de aminoácido 1, 6, 7, 78 y 89 usando la numeración Kabat).
- 20 13. Método según la reivindicación 11, en el que el agente de unión inmunológica comprende una región variable de cadena pesada VH1a y la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 6, 12, 13, 14, 19, 21, 90, 92, 95 y 98 de VH1a usando la numeración AHo (posiciones de aminoácido 1, 6, 11, 12, 13, 18, 20, 79, 81, 82b y 84 usando la numeración Kabat).
- 25 14. Método según la reivindicación 11, en el que el agente de unión inmunológica comprende una región variable de cadena pesada VH1b y la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 10, 12, 13, 14, 20, 21, 45, 47, 50, 55, 77, 78, 82, 86, 87 y 107 de VH1b usando la numeración AHo (posiciones de aminoácido 1, 9, 11, 12, 13, 19, 20, 38, 40, 43, 48, 66, 67, 71, 75, 76 y 93 usando la numeración Kabat).
- 30 15. Método según la reivindicación 11, en el que el agente de unión inmunológica comprende una región variable de cadena ligera V κ 1 y la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 3, 4, 24, 47, 50, 57, 91 y 103 de V κ 1 usando la numeración AHo (posiciones de aminoácido 1, 3, 4, 24, 39, 42, 49, 73 y 85 usando la numeración Kabat).
- 35 16. Método según la reivindicación 11, en el que el agente de unión inmunológica comprende una región variable de cadena ligera V κ 3 y la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 2, 3, 10, 12, 18, 20, 56, 74, 94, 101 y 103 de V κ 3 usando la numeración AHo (posiciones de aminoácido 2, 3, 10, 12, 18, 20, 48, 58, 76, 83 y 85 usando la numeración Kabat).
- 40 17. Método según la reivindicación 11, en el que la una o más posiciones de aminoácido seleccionadas para mutación se seleccionan del grupo que consiste en las posiciones de aminoácido 1, 2, 4, 7, 11-, 14, 46, 53, 82, 92 y 103 de V λ 1 usando la numeración AHo (posiciones de aminoácido 1, 2, 4, 7, 11, 14, 38, 45, 66, 74 y 85 usando la numeración Kabat).
- 45 18. Método según una cualquiera de las reivindicaciones 11-17, en el que la mutación comprende además una o más (preferiblemente todas) las sustituciones de cadena de una posición de aminoácido de cadena pesada seleccionada del grupo que consiste en las posiciones de aminoácido 12, 13 y 144 usando la numeración AHo (posiciones de aminoácido 12, 85 y 103 usando la numeración Kabat).
- 50 19. Método según una cualquiera de las reivindicaciones 11-18, en el que la una o más posiciones de aminoácido seleccionadas para mutación se mutan a un residuo de aminoácido encontrado en una posición de aminoácido correspondiente en una secuencia de anticuerpo seleccionada a partir de un ensayo de QC.
20. Método según una cualquiera de las reivindicaciones 1-19, en el que el agente de unión inmunológica es un anticuerpo scFv, una inmunoglobulina de longitud completa, un fragmento Fab, un Dab o un Nanobody.
21. Método según una cualquiera de las reivindicaciones 1-9 u 11-20, en el que la propiedad funcional es estabilidad mejorada, solubilidad mejorada, no agregación, una mejora en la expresión, una mejora en el rendimiento de repliegamiento tras un procedimiento de purificación de cuerpos de inclusión o una combinación de dos o más de dichas mejoras.
- 55 22. Método según la reivindicación 21, en el que la mejora en la expresión se observa en una célula procariota.

23. Método según una cualquiera de las reivindicaciones 1-9 y 21-22, con la condición de que la propiedad funcional no sea una mejora en la afinidad de unión a antígeno.

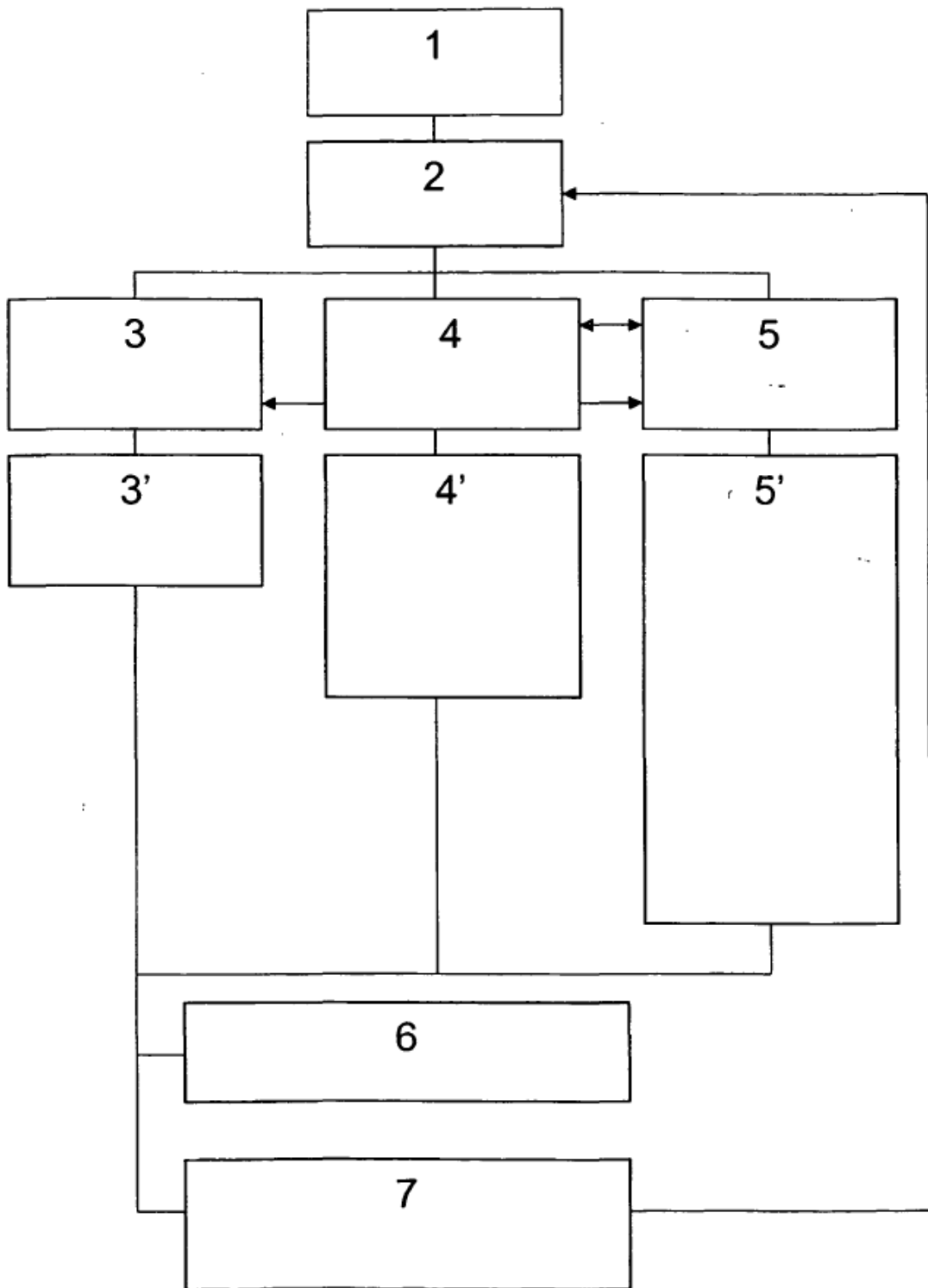


Fig. 1

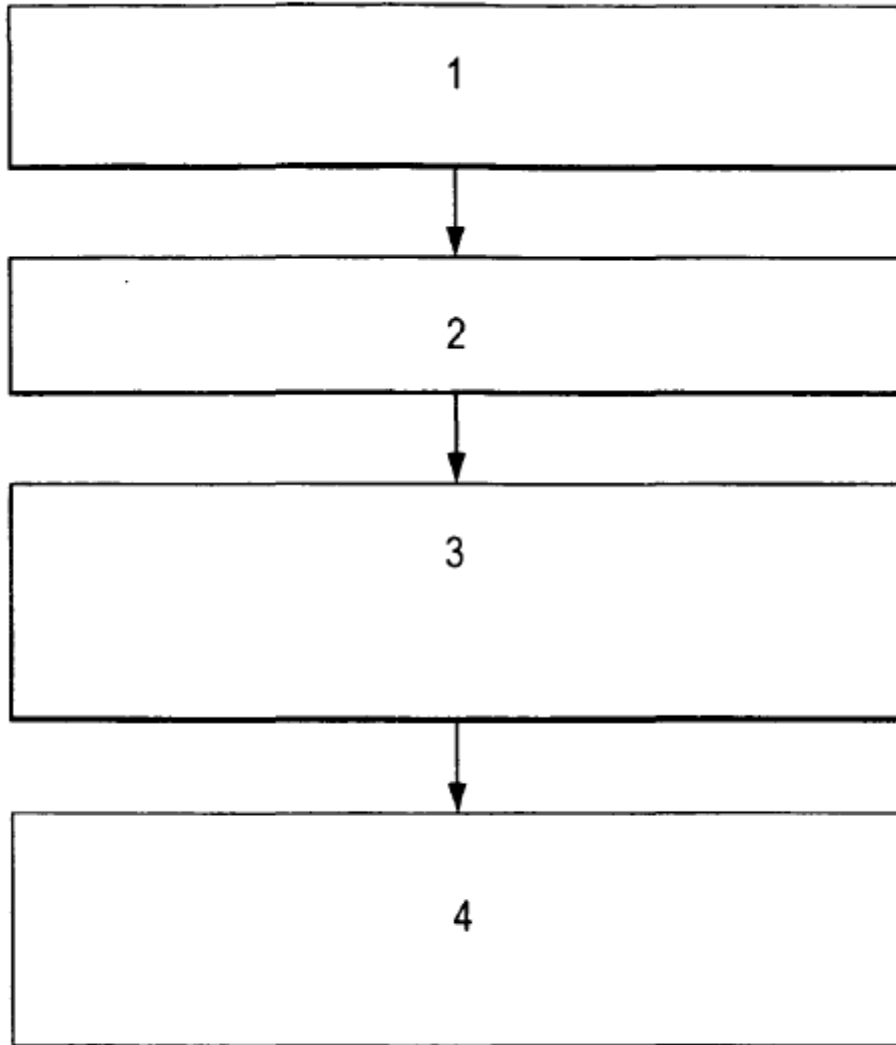


Fig. 2

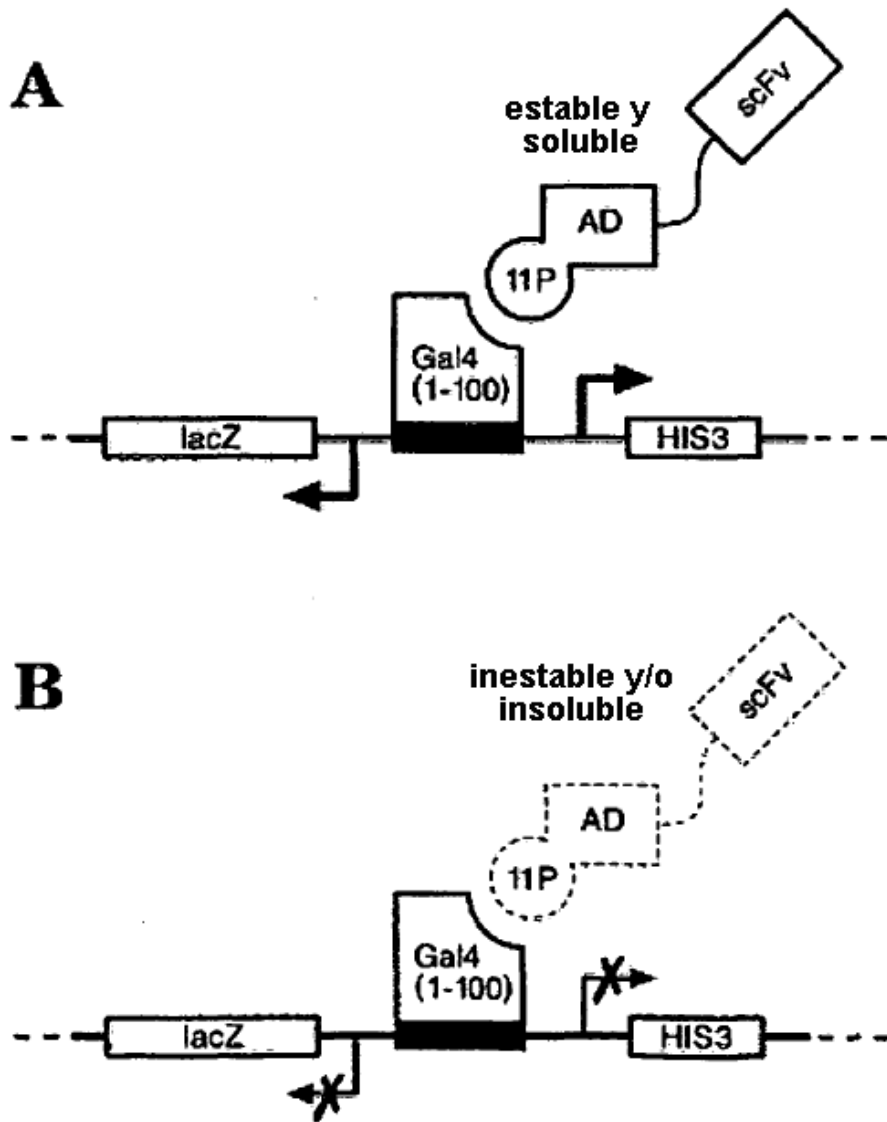
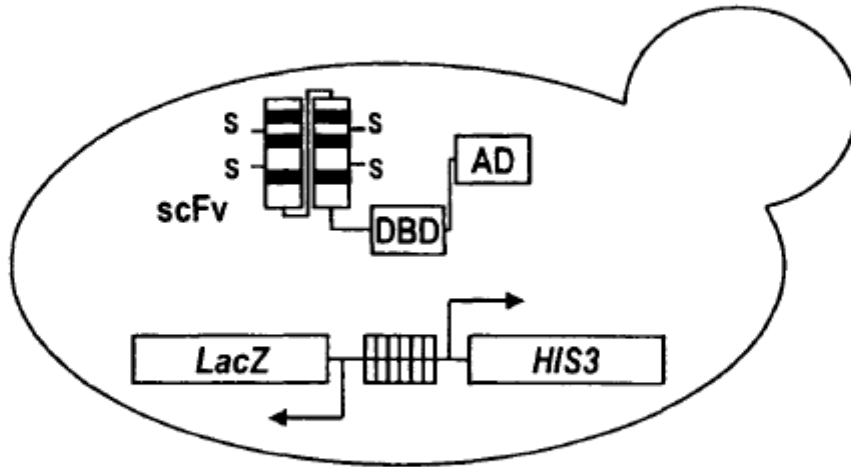


Fig. 3

A



B

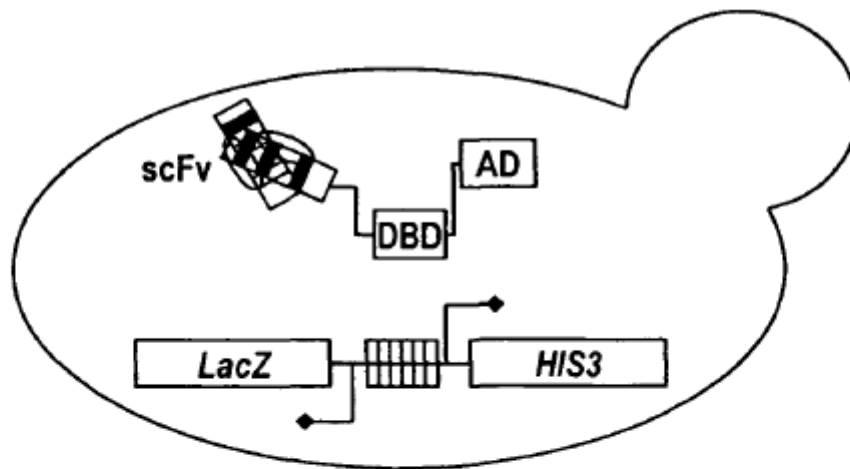
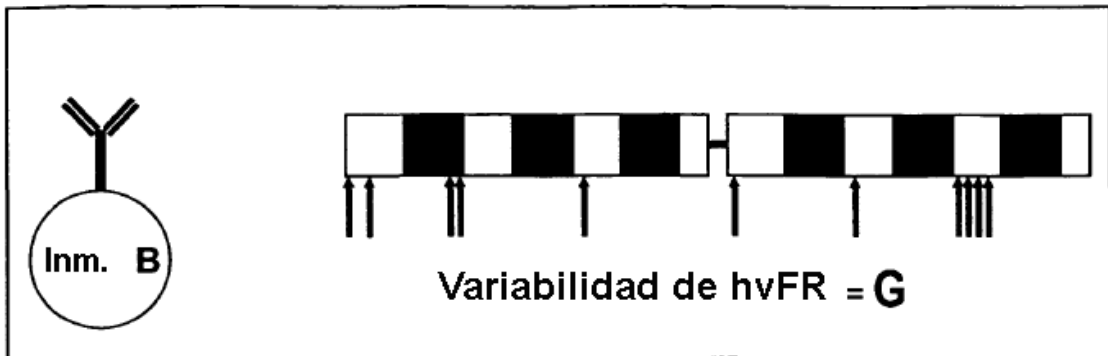


Fig. 4

A



B

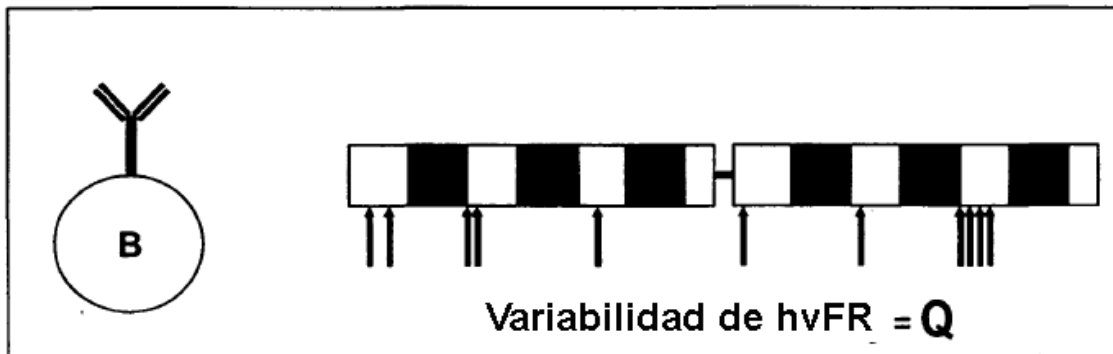


Fig. 5

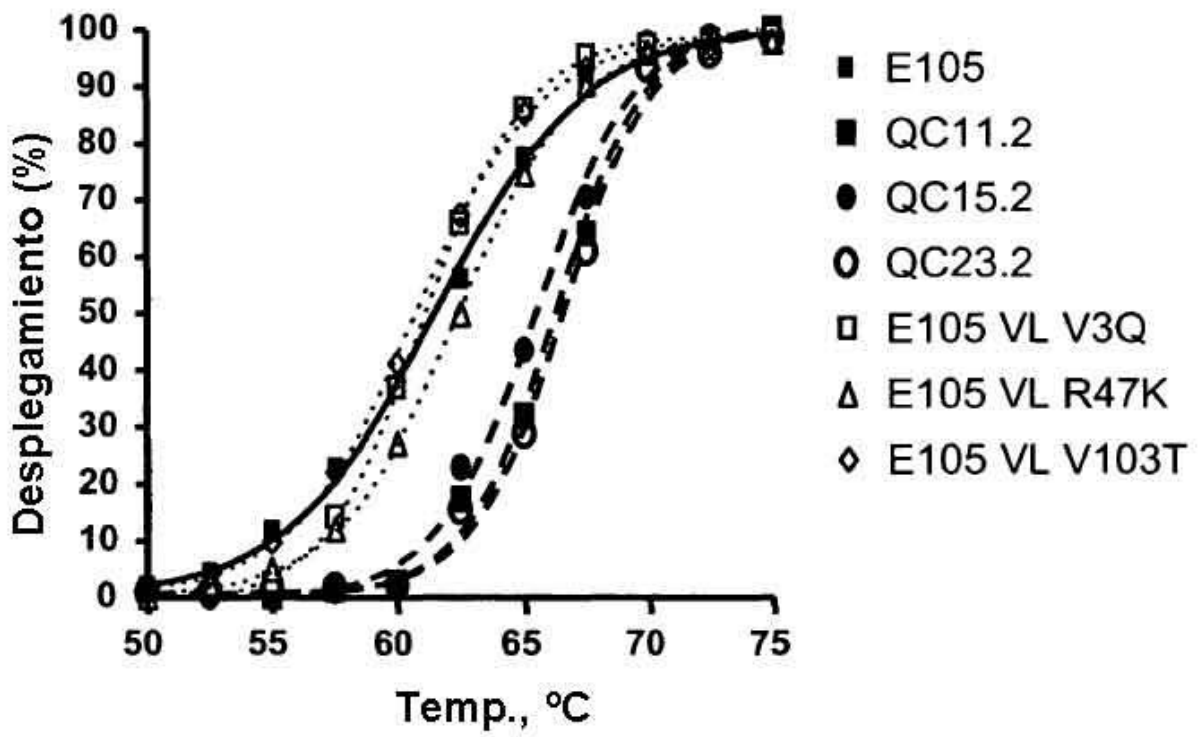


Fig. 6

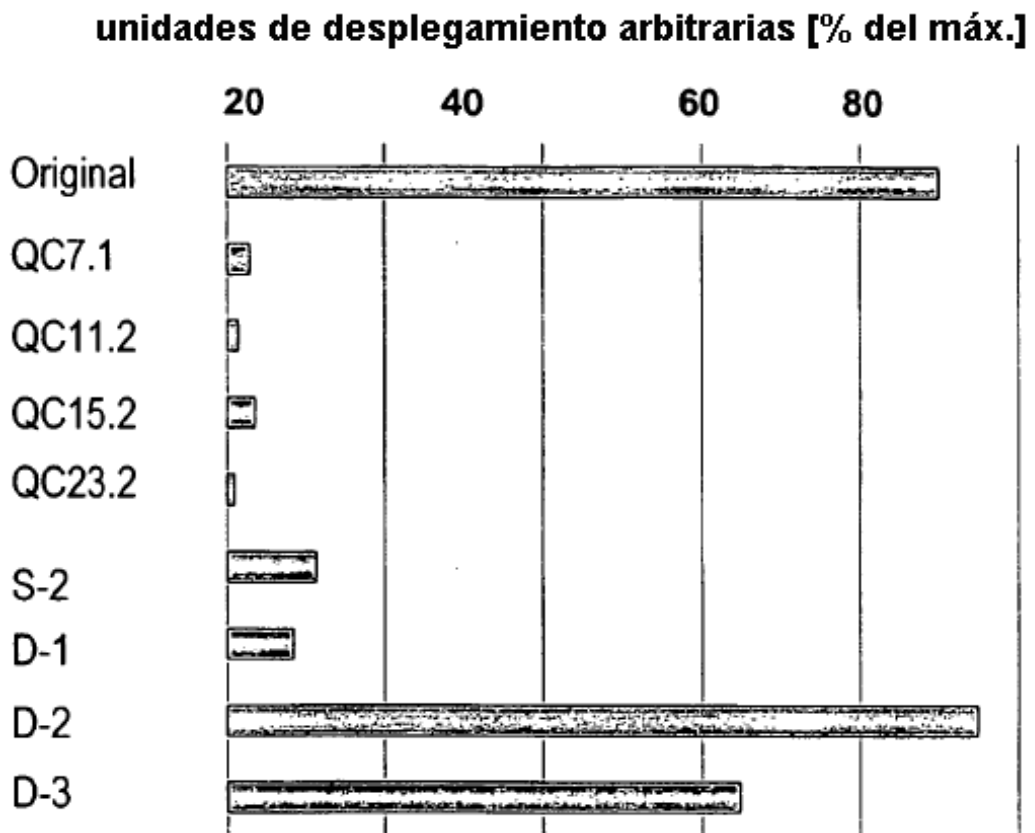


Fig. 7

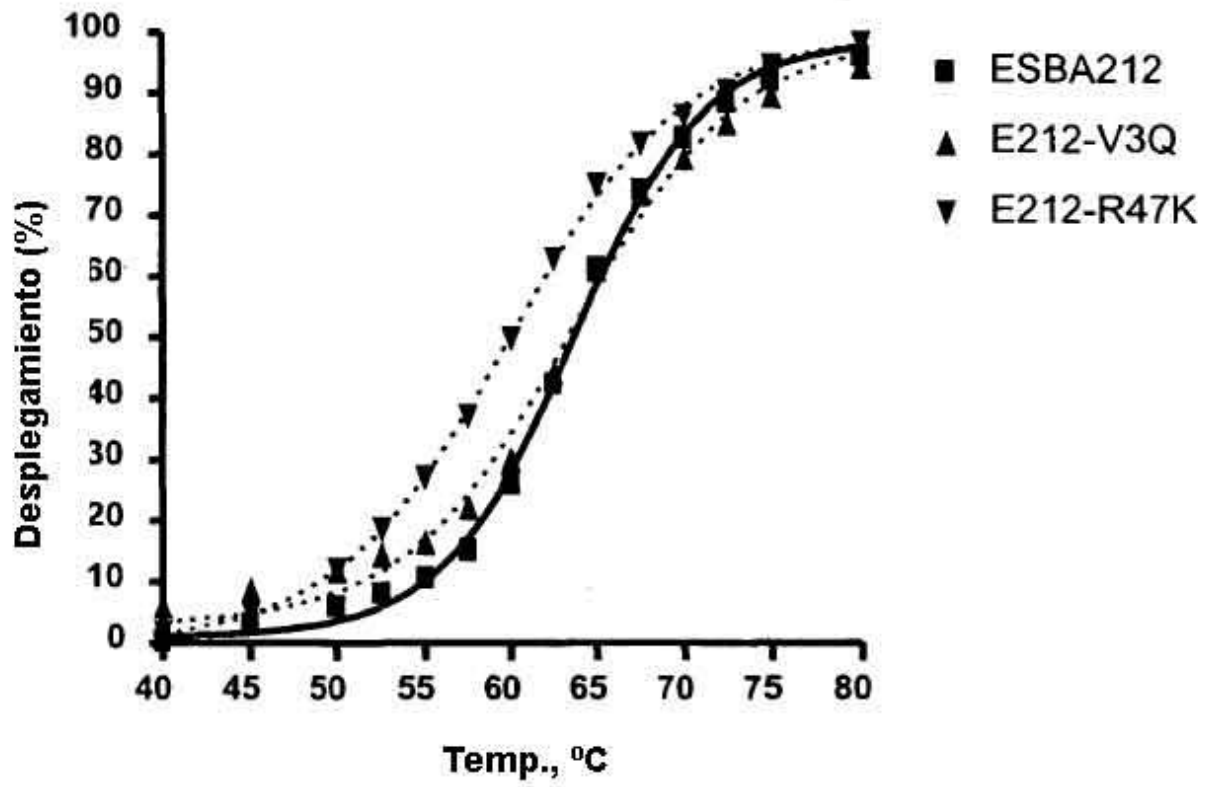


Fig. 8

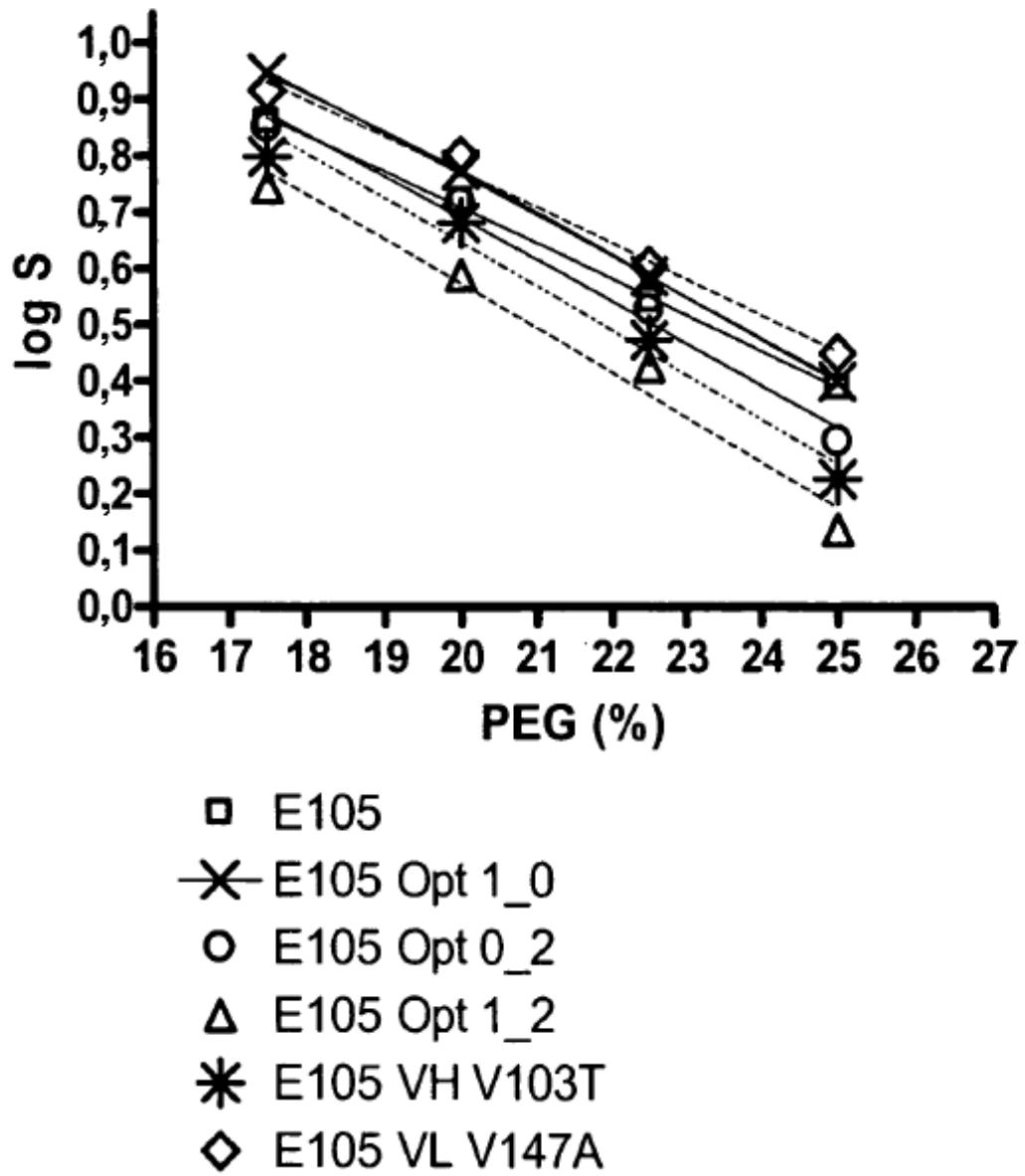
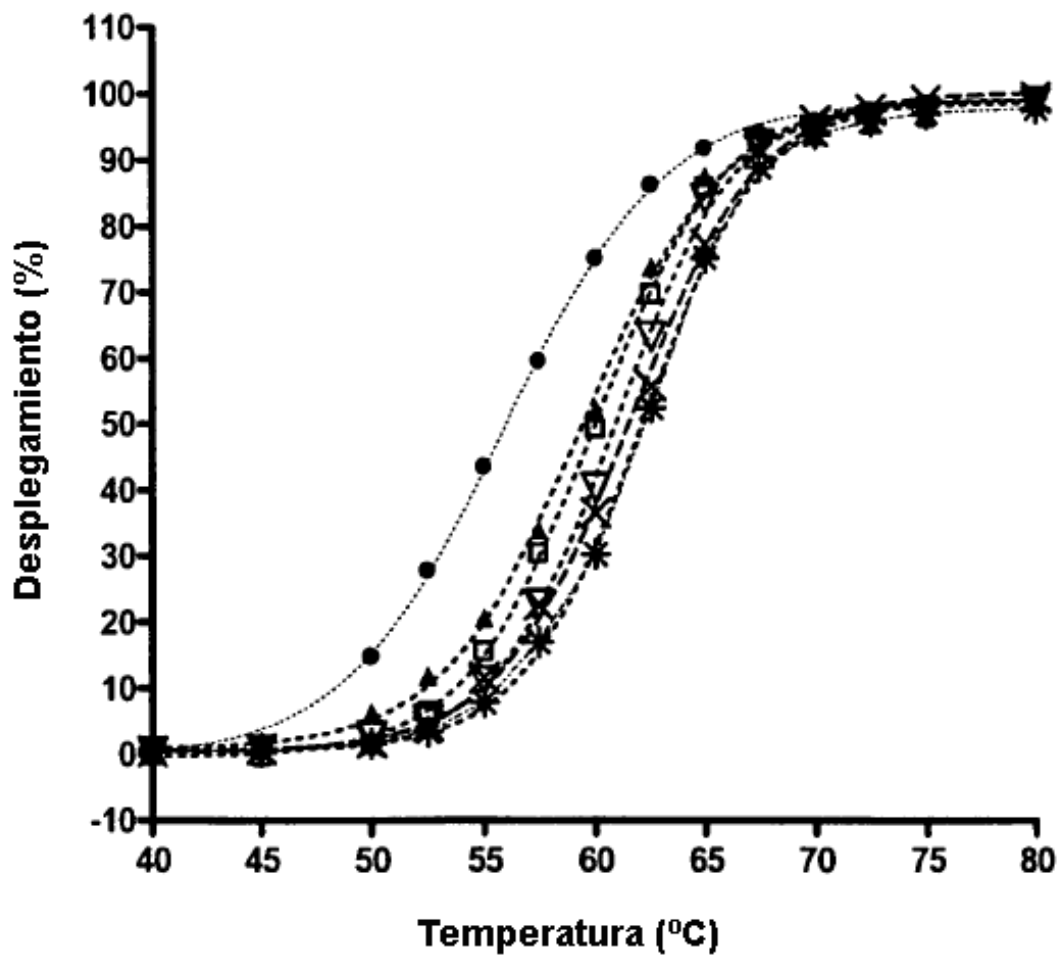


Fig. 9



- × ESBA105
- E105 Opt1.0
- ▲ E105 Opt1.2
- ▽ E105 Opt0.2
- * E105 VL V103T
- E105 VL V147A

Fig. 10

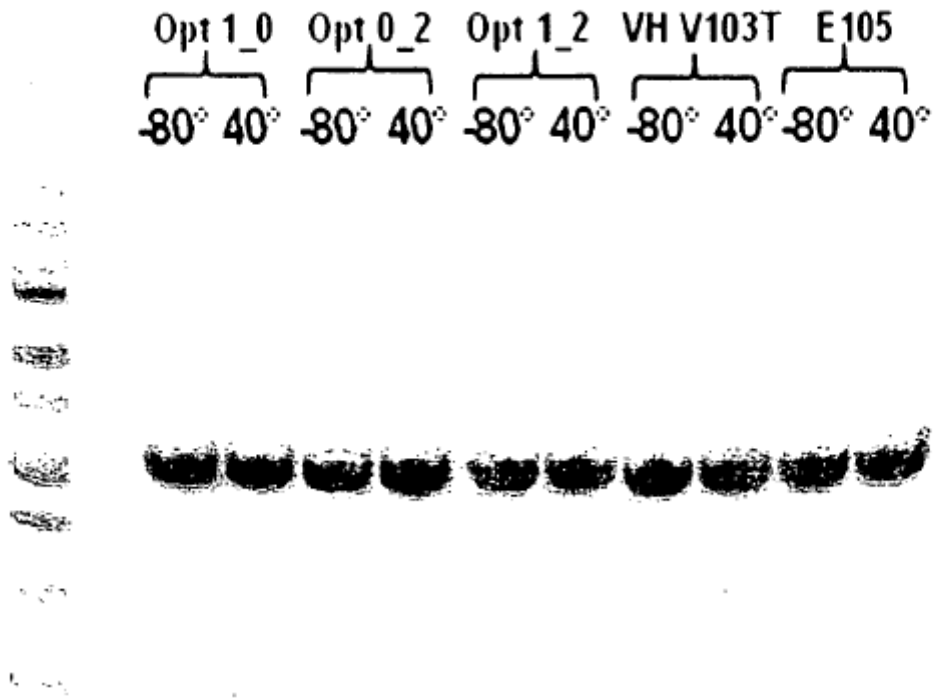


Fig. 11