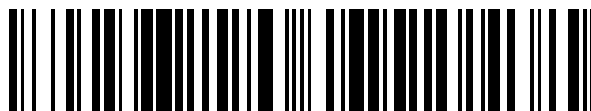


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 762 640**

51 Int. Cl.:

**C07K 16/44** (2006.01)

**A01K 67/00** (2006.01)

**C12N 15/13** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.03.2015 PCT/US2015/021884**

87 Fecha y número de publicación internacional: **24.09.2015 WO15143406**

96 Fecha de presentación y número de la solicitud europea: **20.03.2015 E 15715047 (5)**

97 Fecha y número de publicación de la concesión europea: **30.10.2019 EP 3119811**

54 Título: **Proteínas VL de unión a antígeno que exhiben diferentes características de unión**

30 Prioridad:

**21.03.2014 US 201461968896 P**

**13.11.2014 US 201462079078 P**

**05.12.2014 US 201462088117 P**

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

**25.05.2020**

73 Titular/es:

**REGENERON PHARMACEUTICALS, INC.**  
**(100.0%)**

**777 Old Saw Mill River Road**  
**Tarrytown, NY 10591-6707, US**

72 Inventor/es:

**BABB, ROBERT;**  
**RAFIQUE, ASHIQUE;**  
**HUANG, TAMMY T.;**  
**MACDONALD, LYNN y**  
**MURPHY, ANDREW J.**

74 Agente/Representante:

**PONS ARIÑO, Ángel**

**ES 2 762 640 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

Proteínas  $V_L$  de unión a antígeno que exhiben diferentes características de unión

5 **Campo técnico**

Esta divulgación generalmente se refiere a proteínas  $V_L$  de unión a antígeno que se unen a moléculas pequeñas y/o caracteriza las interacciones de las proteínas  $V_L$  de unión a antígeno y utiliza la información procedente de la caracterización para clasificar las proteínas  $V_L$  de unión a antígeno en grupos que pueden usarse como guía para la selección de una proteína  $V_L$  de unión a antígeno con una característica de unión no exhibida por anticuerpos convencionales.

**Antecedentes**

15 Los anticuerpos han surgido como una modalidad prometedora para la terapia y/o el diagnóstico biológico. Por ejemplo, los anticuerpos neutralizantes pueden interceptar e inactivar un patógeno antes de que establezca una infección. Los anticuerpos antagonistas pueden interferir con la señalización desregulada prevalente en, por ejemplo, progresión tumoral o autoinmunidad, y los anticuerpos agonistas se pueden usar para mejorar las respuestas inmunitarias. Estas capacidades se basan, en parte, en el reconocimiento específico de los anticuerpos y la afinidad por los epítopos, los sitios antigénicos a los que se unen los anticuerpos. Se puede generar un gran número de anticuerpos contra un antígeno diana, y cada anticuerpo puede variar sustancialmente en términos de afinidad y reconocimiento de epítipo o de ambos. Además, el diseño tradicional basado en anticuerpos puede estar limitado porque los sitios de unión a antígeno en los anticuerpos convencionales no son adecuados para todos los antígenos. La presente divulgación abarca el reconocimiento de que sigue existiendo una necesidad de mejorar y diversificar el diseño terapéutico basado en inmunoglobulina.

**Resumen**

30 Varios aspectos y aspectos descritos en el presente documento se basan en parte en el sorprendente descubrimiento de que animales no humanos genéticamente modificados que expresan proteínas de unión que contienen dominios variables de cadena ligera de inmunoglobulina unidos operativamente a una región constante de cadena pesada y dominios variables de cadena ligera de inmunoglobulina unidos operativamente a una región constante de cadena ligera pueden resolver varios problemas reconocidos en el presente documento y/o pueden proporcionar resultados sorprendentes. Por ejemplo, animales no humanos cuyo genoma incluye ambos (i) un locus de cadena pesada de inmunoglobulina que contiene segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, segmentos génicos  $V_L$  y  $J_L$ ); y (ii) un locus de cadena ligera de inmunoglobulina que contiene segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, segmentos génicos  $V_L$  y  $J_L$ ), puede proporcionar un repertorio más diversificado de proteínas de unión a antígeno, por ejemplo, proteínas  $V_L$  de unión, que han sido difíciles de obtener a partir de animales no humanos humanizados convencionales. Las proteínas  $V_L$  de unión a antígeno generadas en los animales genéticamente modificados desvelados en el presente documento se unen a moléculas pequeñas con una afinidad más alta que la que pueden lograr los anticuerpos convencionales, y también pueden exhibir una o más características o rasgos de unión que son distintos de los exhibidos por anticuerpos convencionales.

45 La invención se define en el conjunto de las reivindicaciones adjuntas. Las realizaciones y/o ejemplos de la siguiente descripción que no están cubiertos por las reivindicaciones adjuntas no se consideran parte de la presente invención.

La divulgación proporciona un método para producir una proteína  $V_L$  de unión a antígeno que se une específicamente a un esteroide que comprende las etapas de:

50 (a) inmunizar un ratón genéticamente modificado con dicho esteroide o dicho esteroide unido a un vehículo, en donde el ratón genéticamente modificado comprende

(i) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada de inmunoglobulina de ratón y

55 (ii) segmentos génicos variables de cadena ligera ( $V_L$ ) de inmunoglobulina y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina de ratón; y

60 (b) aislar una proteína  $V_L$  de unión a antígeno, o célula que expresa la proteína  $V_L$  de unión a antígeno, a partir del ratón inmunizado,

65 en donde la proteína  $V_L$  de unión a antígeno comprende un primer y un segundo dominio variable de cadena ligera de inmunoglobulina, en donde el primer dominio variable de cadena ligera de inmunoglobulina está unido operativamente a un dominio constante de cadena pesada de inmunoglobulina, en donde el segundo dominio variable de cadena ligera de inmunoglobulina está unido operativamente a un dominio constante de cadena ligera de inmunoglobulina, y en

donde el primer y el segundo dominio variable de cadena ligera de inmunoglobulina tienen secuencias heterogéneas y están asociados para formar un punto de unión de la proteína  $V_L$  de unión a antígeno que se une específicamente a dicho esteroide con una  $K_D$  de 50 nM o menos.

- 5 Generalmente, una proteína  $V_L$  de unión a antígeno como se desvela en el presente documento comprende una cadena de inmunoglobulina híbrida que comprende un dominio variable de cadena ligera de inmunoglobulina que se une específicamente a una molécula pequeña y que está unida operativamente a una región constante de cadena pesada. La proteína  $V_L$  de unión a antígeno también puede comprender el primer y el segundo dominio variable de cadena ligera de inmunoglobulina, en donde el primer y el segundo dominio variable de cadena ligera de
- 10 inmunoglobulina están asociados para formar un punto de unión que se une específicamente a una molécula pequeña. En algunos aspectos, la presente divulgación se refiere a una proteína de unión a antígeno que consiste esencialmente en el primer y segundo dominio variable de cadena ligera de inmunoglobulina que están asociados para formar un punto de unión, en donde la proteína de unión a antígeno se une específicamente a una molécula pequeña.
- 15 En algunos aspectos, el primer dominio variable de cadena ligera de inmunoglobulina está unido operativamente a un dominio constante de cadena pesada. Esta cadena de inmunoglobulina híbrida  $V_L$ - $C_H$  procede de un segmento génico variable de cadena ligera ( $V_L$ ) y un segmento génico de unión de cadena ligera ( $J_L$ ) unido operativamente a un gen de región constante de cadena pesada. El segundo dominio variable de cadena ligera de inmunoglobulina puede estar unido operativamente a un dominio constante de cadena ligera ( $V_L$ - $C_L$ ).
- 20 En algunos aspectos, cada cadena de una proteína  $V_L$  de unión a antígeno carece de una secuencia de aminoácidos codificada por y/o procedente de un segmento génico de región variable de cadena pesada de inmunoglobulina.
- En algunos aspectos, el primer dominio variable de cadena ligera de inmunoglobulina está codificado por un gen de
- 25 región variable de cadena ligera reordenado procedente de un segmento génico  $V_K$  humano seleccionado del grupo que consiste en  $V_K$  4-1,  $V_K$  1-5,  $V_K$  3-15,  $V_K$  3-20 y  $V_K$  1-33. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $J_K$  seleccionado del grupo que consiste en  $J_K$  1,  $J_K$  3,  $J_K$  4 y  $J_K$  5. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  1-5. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un
- 30 segmento génico  $V_K$  1-5, y el segundo dominio de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-20. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  1-5, y un segmento génico  $J_K$  seleccionado del grupo que consiste en  $J_K$  3,  $J_K$  4 y  $J_K$  5. En un aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  4-1. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  4-1 y un segmento
- 35 génico  $J_K$  1. En un aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  4-1 y el segundo dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  4-1 o  $V_K$  3-20. En un aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-20. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-20 y un segmento génico  $J_K$  1 o  $J_K$  2. En un aspecto, el primer dominio variable de cadena ligera
- 40 de inmunoglobulina procede de un segmento génico  $V_K$  3-20 y el segundo dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  4-1 o  $V_K$  1-5. En un aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-15. En otro aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-15 y un segmento génico  $J_K$  5. En un aspecto, el primer dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  3-15 y el segundo
- 45 dominio variable de cadena ligera de inmunoglobulina procede de un segmento génico  $V_K$  1-39. En otros aspectos, el primer y el segundo dominio variable proceden de los respectivos segmentos génicos  $V_{K1}:J_{K1}$   $V_{K2}:J_{K2}$  como se establece en la Tabla A.

**TABLA A**

Primer dominio variable		Segundo dominio variable	
$V_{K1}$	$J_{K1}$	$V_{K2}$	$J_{K2}$
3-20	4	4-1	2
3-20	4	1-5	2
3-20	3	4-1	1
4-1	1	4-1	3
4-1	1	3-20	3
4-1	1	3-20	2
4-1	3	3-20	2
1-33	3	3-20	5

50

(continuación)

Primer dominio variable		Segundo dominio variable	
V <sub>K1</sub>	J <sub>K1</sub>	V <sub>K2</sub>	J <sub>K2</sub>
1-33	1	1-33	3
3-15	5	1-39	3
1-5	5	3-20	1
1-5	5	3-20	2
1-5	4	3-20	1
1-5	4	3-20	2
1-5	4	3-20	3
1-5	3	3-20	2
1-5	3	3-20	3

En algunos aspectos, la longitud de la CDR3 de la cadena de inmunoglobulina híbrida V<sub>L</sub>-C<sub>H</sub> es más corta que la longitud de CDR3 del segundo dominio variable de cadena ligera de inmunoglobulina unido al dominio constante de cadena ligera (V<sub>L</sub>-C<sub>L</sub>). En algunos aspectos, la CDR3 de la cadena ligera de inmunoglobulina híbrida es al menos un aminoácido más corta que la CDR3 de la cadena ligera. En otros aspectos, las longitudes de la CDR3 difieren en al menos dos aminoácidos. En otros aspectos, las longitudes de la CDR3 difieren en al menos 3 aminoácidos. En otros aspectos, las longitudes de la CDR3 difieren en al menos 4 aminoácidos. En algunos aspectos, la CDR3 de la cadena de inmunoglobulina híbrida tiene 6 aminoácidos de longitud, y la CDR3 de la cadena ligera tiene aproximadamente 9 aminoácidos de longitud.

En algunos aspectos determinados, la región constante de cadena pesada es de un animal no humano. En algunos aspectos, la región constante de cadena ligera es de un animal no humano. En algunos aspectos, la región constante de cadena pesada se selecciona de una CH1, una bisagra, una CH2, una CH3, una CH4 y una combinación de las mismas. En algunos aspectos, la región constante de cadena pesada comprende una CH1, una bisagra, una CH2 y una CH3.

En algunos aspectos, el primer y/o el segundo dominio variable de cadena ligera de inmunoglobulina es un dominio variable de cadena ligera de inmunoglobulina humano. En algunos aspectos, el primer y/o el segundo dominio variable de cadena ligera de inmunoglobulina es de un roedor seleccionado de un ratón y una rata.

En algunos aspectos, la proteína V<sub>L</sub> de unión a antígeno desvelada en el presente documento se une a la molécula pequeña con mayor afinidad que una proteína de unión a antígeno que comprende dominios variables de cadena ligera y pesada de inmunoglobulina. En algunos aspectos, la proteína V<sub>L</sub> de unión a antígeno se une específicamente a una molécula pequeña con una K<sub>D</sub> de menos de 50 nM. En otros aspectos, la K<sub>D</sub> de la proteína V<sub>L</sub> de unión a antígeno es inferior a 40 nM. En aspectos adicionales, la K<sub>D</sub> de la proteína V<sub>L</sub> de unión a antígeno es inferior a 30 nM. En otro aspecto, la K<sub>D</sub> de la proteína V<sub>L</sub> de unión a antígeno es inferior a 20 nM. En otro aspecto, la K<sub>D</sub> de la proteína V<sub>L</sub> de unión a antígeno es inferior a 10 nM.

En un aspecto, se describen en el presente documento células o ácidos nucleicos que comprenden un gen de región variable de cadena ligera reordenado que codifica un dominio variable de una cadena de inmunoglobulina híbrida o una cadena ligera de una proteína V<sub>L</sub> de unión a antígeno que se une específicamente a una molécula pequeña como se desvela en el presente documento, y métodos de obtención de dichas células o ácidos nucleicos.

En algunos aspectos, se describen métodos para obtener una proteína V<sub>L</sub> de unión a antígeno específica para una molécula pequeña, que puede incluir la obtención de células o secuencias de ácido nucleico que comprenden y/o codifican uno o más dominios variables de cadena ligera (V<sub>L</sub>) de inmunoglobulina proteína V<sub>L</sub> de unión a antígeno que se une a una molécula pequeña. Los métodos generalmente comprenden aislar de un animal no humano genéticamente modificado como se desvela en el presente documento una proteína V<sub>L</sub> de unión que se une a una molécula pequeña y/o una célula que comprende una secuencia de ácido nucleico que codifica una proteína V<sub>L</sub> de unión a antígeno, en donde la proteína V<sub>L</sub> de unión se une específicamente a una molécula pequeña.

Los animales no humanos modificados genéticamente desvelados en el presente documento incluyen, por ejemplo, mamíferos y, en aspectos particulares, roedores (por ejemplo, ratones, ratas o hámsteres). En algunos aspectos, los animales no humanos incluyen aves, por ejemplo, pollos. En diversos aspectos, el roedor se selecciona entre un ratón y una rata.

En algunos aspectos, un genoma de un animal no humano como se desvela en el presente documento incluye tanto (i) un locus de cadena pesada de inmunoglobulina que contiene segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, segmentos génicos V<sub>L</sub> y J<sub>L</sub>) como (ii) un locus de cadena ligera de inmunoglobulina que contiene segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, segmentos génicos V<sub>L</sub> y J<sub>L</sub>). En algunos aspectos, los segmentos génicos V<sub>L</sub> y J<sub>L</sub> de inmunoglobulina humanos no reordenados de (i) están presentes

5 en el locus endógeno de cadena pesada de inmunoglobulina en el genoma. En algunos aspectos, el animal no humano carece de todos los segmentos génicos endógenos funcionales  $V_H$ ,  $D_H$  y  $J_H$ . En algunos aspectos, el animal no humano carece de todos los segmentos génicos endógenos, y funcionales  $V_H$ ,  $D_H$  y  $J_H$ , y el animal no humano comprende un gen Adam6a, un gen Adam6b o ambos. En algunos aspectos determinados, el gen Adam6a, el gen Adam6b o ambos están posicionados ectópicamente en el genoma.

10 En algunos aspectos, los segmentos génicos  $V_L$  y  $J_L$  de inmunoglobulina humanos no reordenados de (ii) están presentes en un locus endógeno de cadena ligera de inmunoglobulina del animal no humano. En algunos aspectos determinados, el locus endógeno de cadena ligera de inmunoglobulina es un locus de cadena ligera k.

15 En algunos aspectos, los segmentos génicos  $V_L$  y  $J_L$  de inmunoglobulina humanos no reordenados de (i) son segmentos génicos  $V_K$  y  $J_K$  humanos. En algunos aspectos, los segmentos génicos  $V_L$  y  $J_L$  de inmunoglobulina humanos no reordenados de (ii) son segmentos génicos  $V_K$  y  $J_K$  humanos. En algunos aspectos, los segmentos génicos  $V_L$  y  $J_L$  de inmunoglobulina humanos no reordenados de (ii) son segmentos génicos  $V_K$  y  $J_K$  humanos, y la secuencia de ácido nucleico de región constante de cadena ligera es una secuencia de ácido nucleico de región  $C_K$  de ratón o una secuencia de ácido nucleico de región  $C_K$  de rata.

20 En algunos aspectos, el animal no humano comprende una célula que expresa una proteína  $V_L$  de unión a antígeno que se une específicamente a una molécula pequeña. En algunos aspectos, la célula es un linfocito, por ejemplo, una célula NK, un linfocito T o un linfocito B. En algunos aspectos, la célula expresa una proteína  $V_L$  de unión que comprende una cadena híbrida  $V_L$ - $C_H$ . En algunos aspectos, la proteína  $V_L$  de unión comprende dos dominios variables de cadena ligera de inmunoglobulina idénticos. En otros aspectos, la proteína  $V_L$  de unión comprende dos dominios variables de cadena ligera de inmunoglobulina con secuencias heterogéneas.

25 En algún aspecto, la célula aislada de un animal como se desvela en el presente documento es un linfocito B. En otros aspectos, la célula es un linfocito B de memoria.

30 Los ácidos nucleicos que comprenden un gen de región variable de cadena ligera reordenado que codifica un dominio variable de una cadena de inmunoglobulina híbrida o una cadena ligera de una proteína  $V_L$  de unión a antígeno que se une específicamente a una molécula pequeña también pueden aislarse mediante la identificación, por ejemplo, a partir de una célula aislada de un animal no humano desvelado en el presente documento, la primera y segunda secuencia de ácido nucleico que codifican el primer y segundo dominio variable de cadena ligera de inmunoglobulina de una proteína  $V_L$  de unión que se une específicamente a la molécula pequeña. En algunos aspectos, los métodos para obtener una célula y/o ácido nucleico como se desvela en el presente documento comprenden (a) inmunizar un animal no humano con una molécula pequeña o la molécula pequeña unida a un vehículo, en donde el animal no humano comprende en su genoma (i) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada no humana, y (ii) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera no humana, (b) aislar una célula del animal no humano inmunizado, en donde la célula comprende la primera y segunda secuencia de ácido nucleico que codifican el primer y segundo dominio variable de cadena ligera de inmunoglobulina; y (c) identificar a partir de la célula la primera y la segunda secuencia de ácido nucleico que codifican el primer y el segundo dominio variable de cadena ligera de inmunoglobulina de una proteína  $V_L$  de unión que se une específicamente a la molécula pequeña.

45 En algunos aspectos, la inmunización de un animal no humano comprende cebar al animal no humano con la molécula pequeña o la molécula pequeña unida a un vehículo, permitiendo que el animal no humano descanse durante un período de tiempo y volver a inmunizar al animal con la molécula pequeña o la molécula pequeña unida al vehículo. En algunos aspectos, el período de tiempo es de unos pocos días, al menos una semana, al menos dos semanas, al menos tres semanas, al menos cuatro semanas o al menos un mes. En algunos aspectos, la inmunización del animal no humano comprende permitir que el animal no humano desarrolle una respuesta inmunitaria.

50 En algunos aspectos, la célula se obtiene mediante clasificación celular activada por fluorescencia (FACS, de sus siglas en inglés) o citometría de flujo. En algunos aspectos, la célula se obtiene de un tejido del animal no humano inmunizado, y en donde el tejido se selecciona del grupo que consiste en bazo, ganglio linfático, sangre y médula ósea.

60 En algunos aspectos, los métodos de la presente divulgación comprenden además fusionar el linfocito con una célula cancerosa, por ejemplo, para hacer un hibridoma. En algunos aspectos determinados, la célula cancerosa es una célula de mieloma. Por consiguiente, también se describen en el presente documento hibridomas y ácidos nucleicos aislados de los mismos, en donde los hibridomas expresan una proteína  $V_L$  de unión específica para una molécula pequeña.

65 En algunos aspectos, los métodos para hacer una proteína  $V_L$  de unión a antígeno específica para una molécula pequeña también pueden comprender: expresar un primer y un segundo ácido nucleico que codifican un primer y un segundo dominio variable de cadena ligera de inmunoglobulina de una proteína  $V_L$  de unión a antígeno específica

para la molécula pequeña en un sistema de expresión adecuado para expresar el primer y segundo dominio variable de cadena ligera de inmunoglobulina como un dímero que se une específicamente a la molécula pequeña.

También se describe en el presente documento un animal no humano que comprende (a) en su genoma: (i) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada no humana, y (ii) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera no humana; y (b) una proteína  $V_L$  de unión a antígeno que se une específicamente a una molécula pequeña.

En algunos aspectos, el animal no humano exhibe 2 veces o más, por ejemplo, al menos 3 veces, al menos 4 veces, al menos 5 veces, al menos 6 veces, al menos 7 veces, al menos 8 veces, al menos 10 veces o 20 veces o más linfocitos B positivos para antígeno que un animal no humano de referencia. En algunos aspectos, el animal no humano de referencia expresa anticuerpos quiméricos tras la inmunización, en donde los anticuerpos quiméricos tienen cadenas pesadas que comprenden dominios  $V_H$  humanos y dominios  $C_H$  de ratón y cadenas ligeras que tienen dominios  $V_L$  humanos y dominios  $C_L$  de ratón. En algunos aspectos determinados, el animal no humano de referencia es un animal no humano de tipo silvestre. En algunos aspectos, la inmunización comprende cebar al animal no humano con la molécula pequeña o la molécula pequeña unida a un vehículo, permitiendo que el animal no humano descanse durante un período de tiempo y volviendo a inmunizar al animal con la molécula pequeña o la molécula pequeña unida al vehículo. En algunos aspectos, el período de tiempo es de unos pocos días, al menos una semana, al menos dos semanas, al menos tres semanas, al menos cuatro semanas o al menos un mes. En algunos aspectos, los linfocitos B positivos para antígeno son linfocitos B de memoria.

En algunos aspectos, el animal no humano exhibe al menos 2 veces, al menos 3 veces, al menos 4 veces o al menos 5 veces o más una cantidad de anticuerpos mayor que un animal no humano de referencia. En algunos aspectos determinados, el animal no humano de referencia es un ratón genéticamente modificado, que expresa proteínas de unión a antígeno quiméricas tras la inmunización, y las proteínas de unión a antígeno quiméricas comprenden cadenas pesadas que contienen dominios  $V_H$  humanos y dominios  $C_H$  de ratón, y cadenas ligeras que tienen dominios  $V_L$  humanos y dominios  $C_L$  de ratón. En algunos aspectos determinados, el animal no humano de referencia es un animal no humano de tipo silvestre.

En algunos aspectos, una molécula pequeña de la presente divulgación es un hapteno y está unida a un vehículo. En algunos aspectos determinados, el vehículo comprende hemocianina de lapa californiana (KLH, de sus siglas en inglés), hemocianina de *Concholepas concholepas* (CCH, de sus siglas en inglés), seroalbúmina bovina (BSA, de sus siglas en inglés), una seroalbúmina bovina cationizada (cBSA, de sus siglas en inglés) u ovoalbúmina.

En algunos aspectos, una molécula pequeña de la presente divulgación es un compuesto orgánico cuyo peso molecular es inferior a 6 kDa.

En un aspecto, se desvelan, en el presente documento, métodos para identificar y/o aislar proteínas  $V_L$  de unión a antígeno específicas de antígeno que exhiben una característica de unión no exhibida por anticuerpos convencionales, proteínas  $V_L$  de unión a antígeno específicas de antígeno así identificadas y/o aisladas, ácidos nucleicos que las codifican, y/o células hospedadoras que las expresan.

En un aspecto, un método para identificar una o más proteínas  $V_L$  de unión a antígeno que exhiben una característica de unión única cuando se une específicamente a un antígeno no exhibido por anticuerpos convencionales que también se unen específicamente al antígeno como se desvela en el presente documento comprende (a) perfilar una o más características de unión de cada una de una pluralidad de proteínas de inmunoglobulina que se unen específicamente a un antígeno, en donde la pluralidad de proteínas de inmunoglobulina comprende proteínas  $V_L$  de unión a antígeno y anticuerpos convencionales, en donde cada proteína  $V_L$  de unión a antígeno comprende una cadena de inmunoglobulina híbrida que comprende (i) un dominio variable procedente de uno o más segmentos génicos de región variable de cadena ligera y (ii) un dominio constante procedente de uno o más segmentos génicos de región constante de cadena pesada, en donde cada anticuerpo convencional comprende una región variable de cadena pesada de inmunoglobulina procedente de una o más regiones variables de cadena pesada y un segmento génico de región variable de cadena ligera de inmunoglobulina procedente de uno o más segmentos génicos de región variable de cadena ligera; (b) agrupar la pluralidad de proteínas de inmunoglobulina en uno o más grupos basándose en al menos una característica de unión de cada una de las proteínas de inmunoglobulina, en donde las proteínas  $V_L$  de unión a antígeno y los anticuerpos convencionales que exhiben una característica de unión similar se agrupan en el mismo grupo; y (c) identificar un grupo que comprende todas o sustancialmente todas las proteínas  $V_L$  de unión a antígeno.

En algunos aspectos, una o más características de unión de cada una de la pluralidad de proteínas de inmunoglobulina se perfila mediante interrupción antigénica diferencial. En algunos aspectos, los métodos desvelados en el presente documento comprenden además mapear uno o más epítomos del antígeno unido por cada una de la pluralidad de proteínas de inmunoglobulina; en donde las proteínas de inmunoglobulina que se unen al mismo epítomo del antígeno se agrupan en el mismo grupo funcional. En algunos aspectos, el mapeo de uno o más epítomos del antígeno unido por cada una de la pluralidad de proteínas de inmunoglobulina comprende un ensayo de mapeo de epítomos

seleccionado del grupo que consiste en un ensayo de bloqueo cruzado, escaneo de alanina de mutantes de antígeno, transferencias de péptidos, análisis de escisión de péptidos, escisión de epítomos, extracción de epítomos, modificación química del antígeno y una combinación de los mismos.

- 5 En los métodos que se divulgan en el presente documento, una o más características de unión de una pluralidad de proteínas de unión a antígeno se determina usando antígeno inmovilizado en una superficie sólida. En algunos aspectos, la superficie sólida comprende chips de biosensor o perlas de poliestireno. En algunos aspectos, el antígeno se modifica después de la inmovilización y antes de la elaboración de perfiles. La modificación puede efectuarse con un producto químico (por ejemplo, clorhidrato de tris(2-carboxietil)fosfina (TCEP•HCl)/yodoacetamida, N-etil-N'-(dimetilaminopropil)carbodiimida (EDC)/etanolamina, yodoacetamida e hidracina, p-hidroxifenilgloxal (HPG, de sus siglas en inglés), peróxido de hidrógeno, N-bromosuccinimida, N-acetilimidazol, tetranitrometano, ácido arsanílico, cloruro de dansilo, glutaraldehído, ninhidrina, dietilpirocarbonato (DEPC), acetato de sulfosuccinimidilo (acetato de sulfo-NHS), polietilenglicol 5000 (PEG-5000), ácido 7-hidroxycumarin-3-carboxílico, succinimidil éster, y una combinación de los mismos) y/o una enzima (por ejemplo, tripsina porcina, endoproteinasa Glu-C, endoproteinasa Asp-N, quimotripsina, endoproteinasa Lys-C, y endoproteinasa Arg-C, pepsina, papaína, termolisina, subtilisina, proteasa K, proteasa específica de bromelina sulfhidrilo (ficina) y una combinación de los mismos).

La agrupación de acuerdo con los métodos desvelados en el presente documento puede comprender análisis de componentes principales (PCA, de sus siglas en inglés) y/o agrupamiento jerárquico. En un aspecto, se seleccionan dos componentes principales para presentar los datos. En un aspecto, la agrupación comprende el análisis de componentes principales. En otro aspecto, la agrupación comprende el agrupamiento jerárquico. En otro aspecto, la agrupación comprende tanto el análisis de componentes principales como el agrupamiento jerárquico. La agrupación puede basarse en uno o más perfiles de unión que comprenden una intensidad de señal de unión de cada proteína de inmunoglobulina a un panel de superficies de antígeno químicamente y/o enzimáticamente interrumpidas/modificadas como se describió anteriormente. Dichos resultados de agrupación pueden alinearse con otros datos de análisis típicos para un grupo de proteínas de inmunoglobulina tales como constantes de asociación, constantes de disociación, constantes de equilibrio, especificidades de unión hacia homólogos de antígeno de varias especies o miembros de la familia relacionados de la misma especie, datos de actividad funcional (por ejemplo, capacidad para bloquear el bloqueo de ligando, fosforilación de antígeno y/o internalización de antígeno en células) o cualquier combinación de los mismos. Los resultados de alineación, que pueden mostrarse como una "tabla de árbol", por ejemplo, un dendrograma de agrupamiento jerárquico procedente de datos de unión de interrupción de antígeno diferencial se alinean con otros diversos datos de ensayo para cada proteína de inmunoglobulina, pueden usarse para revelar patrones de comportamiento entre proteínas de inmunoglobulina que comparten un bin.

- 35 Algunos métodos de elaboración de perfiles como se desvelan en el presente documento comprenden además (d) aislar una o más proteínas  $V_L$  de unión a antígeno agrupadas en un grupo funcional identificado como que comprende todas o sustancialmente todas las proteínas  $V_L$  de unión a antígeno y/o (e) confirmar que la una o más proteínas  $V_L$  de unión a antígeno aisladas se unen a uno o más epítomos del antígeno que los anticuerpos convencionales no reconocen. La confirmación de que una o más proteínas  $V_L$  de unión a antígeno aisladas se une a uno o más epítomos del antígeno que los anticuerpos convencionales no reconocen puede comprender un ensayo de proteína de unión competitivo de alto rendimiento.

La secuencia de aminoácidos y/o la secuencia de ácido nucleico que codifica la misma puede determinarse para cualquiera de la una o más proteínas  $V_L$  de unión a antígeno aisladas de acuerdo con un método de elaboración de perfiles desvelado en el presente documento. Por consiguiente, también se describen en el presente documento proteínas  $V_L$  de unión a antígeno aisladas de acuerdo con un método de elaboración de perfiles desvelado en el presente documento, ácidos nucleicos aislados que comprenden una secuencia de nucleótidos que codifica una CDR de la región variable de una cadena de inmunoglobulina híbrida de una proteína  $V_L$  de unión a antígeno identificada y/o aislada, y células hospedadoras que expresan dichos ácidos nucleicos.

50 También se describe en el presente documento un método para identificar uno o más epítomos de un antígeno que están enmascarados con anticuerpos convencionales y se reconocen por una o más proteínas  $V_L$  de unión a antígeno específicas de antígeno que comprenden identificar una o más proteínas  $V_L$  de unión a antígeno que se unen a epítomos del antígeno no reconocido mediante anticuerpos convencionales usando métodos como los desvelados en el presente documento y (b) mapear el uno o más epítomos reconocidos por la una o más proteínas de unión a antígeno específicas de antígeno identificadas.

60 Otras características, objetos y ventajas de la presente invención son evidentes en la descripción detallada que sigue. Debería entenderse, sin embargo, que la descripción detallada, aunque indica realizaciones de la presente invención, se proporciona solo a modo de ilustración, no de limitación. Diversos cambios y modificaciones serán evidentes para los expertos en la técnica a partir de la descripción detallada.

### Breve descripción de las figuras

- 65 El dibujo incluido en el presente documento, que se compone de las siguientes Figuras, es solo para fines ilustrativos, no para limitación.

La **Figura 1** ilustra un esquema (no a escala) del locus de cadena pesada de ratón, en la parte superior y un esquema (no a escala) del locus de cadena ligera k humano, en la parte inferior. El locus de cadena pesada de ratón tiene aproximadamente 3 Mb de longitud y contiene aproximadamente 200 segmentos génicos variables de cadena pesada ( $V_H$ ), 13 segmentos génicos de diversidad de cadena pesada ( $D_H$ ) y 4 segmentos génicos de unión de cadena pesada ( $J_H$ ), así como potenciadores (Enh, de sus siglas en inglés) y regiones constantes de cadena pesada ( $C_H$ ). El locus de cadena ligera k humano se duplica en contigs distales y proximales de polaridad opuesta que abarca aproximadamente 440 kb y 600 kb, respectivamente. Entre los dos contigs hay aproximadamente 800 kb de ADN que se cree que está libre de segmentos génicos  $V_k$ . El locus de cadena ligera k humano contiene aproximadamente 76 segmentos génicos  $V_k$ , 5 segmentos génicos  $J_k$ , un potenciador (Enh) intrónico y una única región constante (Ck).

La **Figura 2** muestra una estrategia de direccionamiento ejemplar para la inserción progresiva de 40 segmentos génicos  $V_k$  humanos y 5  $J_k$  humanos en un locus de cadena pesada de ratón. Los casetes de selección de higromicina (hyg) y neomicina (neo) se muestran con sitios de reconocimiento de recombinasa (R1, R2, etc.). En la parte inferior se muestra un locus de cadena pesada de ratón modificado que comprende segmentos génicos  $V_k$  y  $J_k$  humanos unidos operativamente a regiones  $C_H$  de ratón.

La **Figura 3** muestra una estrategia de direccionamiento ejemplar para la inserción progresiva de  $V_\lambda$  humano y un segmento génico  $J_\lambda$  humano (o cuatro segmentos génicos  $J_\lambda$  humanos) en el locus de cadena pesada de ratón. Los casetes de selección de higromicina (hyg) y neomicina (neo) se muestran con sitios de reconocimiento de recombinasa (R1, R2, etc.). En la parte inferior se muestra un locus de cadena pesada de ratón modificado que comprende segmentos génicos  $V_\lambda$  y  $J_\lambda$  humanos (uno o cuatro) unidos operativamente a regiones  $C_H$  de ratón.

La **Figura 4** muestra el número total (izquierda) y el porcentaje (derecha) de anticuerpos antígeno positivos (o proteínas  $V_L$  de unión a antígeno) obtenidos de ratones KOH (MAID 1713/1242) y ratones humanizados VELOCIMMUNE® (VI3).

La **Figura 5** muestra la cinética de unión relativa de anticuerpos específicos para el antígeno B obtenida de ratones KOH (MAID 1713/1242) y ratones humanizados VELOCIMMUNE® (VI3).

La **Figura 6** proporciona un gráfico de Análisis de Componentes Principales (PCA) bidimensional de 739 proteínas de unión específicas para el antígeno C, una glucoproteína, que destaca un grupo de proteínas  $V_L$  de unión a antígeno específicas del antígeno C (●) que exhiben al menos una característica de unión distinta de anticuerpos específicos de antígeno A típicos (○) según lo determinado por la interrupción diferencial de antígeno (DAD, de sus siglas en inglés).

La **Figura 7** proporciona el número de proteínas  $V_L$  de unión (Número total; eje y) específicas para el antígeno A (■), antígeno B (▲) o antígeno C (□) y que tiene una determinada longitud de aminoácidos de CDR3 (eje x) en la cadena híbrida (A) o cadena ligera (B).

## Definiciones

Esta invención no se limita a los métodos ni a las condiciones experimentales descritas en particular, ya que dichos métodos y condiciones pueden variar. También debe entenderse que la terminología utilizada en el presente documento únicamente tiene el fin de describir realizaciones particulares, y no se pretende que sea limitante, ya que el alcance de la presente invención se define por las reivindicaciones.

A menos que se defina de otro modo, todos los términos y frases utilizadas en el presente documento incluyen que los términos y las frases que son habituales en la técnica, salvo que se indique claramente lo contrario o sea claramente evidente a partir del contexto en el que se usa un término o frase. Aunque puede usarse cualquier método y material similar o equivalente a los descritos en el presente documento en la práctica o ensayo de la presente invención, a continuación, se describen métodos y materiales particulares.

Una "proteína de unión a antígeno", "proteína de unión", "proteína de inmunoglobulina" o similar se refiere a una molécula de péptido mono o polimérico que comprende un sitio de unión a antígeno, que puede estar mutado somáticamente, capaz de reconocer y unir un antígeno (o porción de epítipo del mismo), por ejemplo, una sustancia capaz de inducir una respuesta inmunitaria y especialmente la producción de moléculas de inmunoglobulina maduras por afinidad. La proteína de unión a antígeno abarca proteínas  $V_L$  de unión a antígeno y anticuerpos convencionales. Un "sitio de unión a antígeno" de una proteína de unión a antígeno se refiere a la región de la proteína de unión a antígeno que se une al antígeno.

"Proteína  $V_L$  de unión a antígeno", "proteína  $V_L$  que se une al antígeno", "proteína  $V_L$  de unión" o similar, se refiere a una proteína de inmunoglobulina que comprende un dominio variable de cadena ligera de inmunoglobulina, que puede formar un sitio de unión a antígeno, unido operativamente a una región constante de cadena pesada. La "proteína  $V_L$  de unión a antígeno" incluye moléculas de inmunoglobulina que comprenden además una cadena ligera de manera



que la proteína  $V_L$  de unión comprende dos dominios variables de cadena ligera, que pueden formar un sitio de unión a antígeno. En un aspecto, al menos dos dominios variables de cadena ligera de las proteínas  $V_L$  de unión a antígeno son afines. En algunos aspectos, cada uno de los dos dominios variables de cadena ligera está codificado o procede de un segmento génico de región variable de cadena ligera ( $V_L$ ) y/o un segmento génico de región de unión de cadena ligera ( $J_L$ ). En aspectos preferidos, uno de los dos dominios variables de cadena ligera puede ser parte de una cadena de inmunoglobulina híbrida, y el otro de los dos dominios variables de cadena ligera puede ser parte de una cadena ligera (L) de inmunoglobulina. Dichos dominios de unión  $V_L$  se han descrito, véase, por ejemplo, la publicación de patente de EE.UU. n.º 20120096572, presentada el 2 de agosto de 2011.

El término "anticuerpo", "anticuerpo convencional", "anticuerpo típico", "anticuerpo de unión a antígeno", o similares, generalmente se refiere a una proteína de inmunoglobulina que comprende como mínimo un sitio de unión a antígeno que comprende (i) un dominio variable de cadena pesada procedente de un segmento génico variable de cadena pesada ( $V_H$ ), un segmento génico de diversidad de cadena pesada ( $D_H$ ) y/o un segmento génico de unión de cadena pesada ( $J_H$ ) y (ii) un dominio variable de cadena ligera procedente de un segmento génico variable de cadena ligera ( $V_L$ ) y/o un segmento génico de unión de cadena ligera ( $J_L$ ). En un aspecto preferido, los dominios  $V_H$  y  $V_L$  del anticuerpo son afines. Por consiguiente, el término anticuerpo, anticuerpo convencional, anticuerpo típico o similar abarca un fragmento variable de cadena sencilla (scFv, de sus siglas en inglés), una región de fragmento de unión a antígeno (Fab, de sus siglas en inglés), un fragmento  $F(ab')_2$ , etc. Dichos términos también abarcan moléculas tetraméricas, por ejemplo, moléculas que tienen dos cadenas pesadas (H) de inmunoglobulina y dos cadenas ligeras (L) de inmunoglobulina interconectadas mediante enlaces disulfuro.

Cada cadena pesada comprende un dominio variable de cadena pesada y una región constante de cadena pesada ( $C_H$ ). La región constante de cadena pesada comprende tres dominios,  $C_{H1}$ ,  $C_{H2}$  y  $C_{H3}$ . Cada cadena ligera comprende un dominio variable de cadena ligera y una región constante de cadena ligera ( $C_L$ ). Los dominios variables de cadena pesada y cadena ligera se pueden subdividir además en regiones de hipervariabilidad, denominadas regiones determinantes de la complementariedad (CDR), intercaladas con regiones que están más conservadas, denominadas regiones marco (FR, de sus siglas en inglés). Cada dominio variable de cadena pesada y ligera comprende tres CDR y cuatro FR, dispuestas desde el extremo amino hasta el extremo carboxilo en el siguiente orden: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4 (las CDR de cadena pesada pueden abreviarse como HCDR1, HCDR2 y HCDR3; las CDR de cadena ligera pueden abreviarse como LCDR1, LCDR2 y LCDR3). La expresión anticuerpo de "alta afinidad" incluye un anticuerpo que tiene una  $K_D$  con respecto a su epítipo diana de aproximadamente  $10^{-9}$  M o menos (por ejemplo, aproximadamente  $1 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M,  $1 \times 10^{-11}$  M, o aproximadamente  $1 \times 10^{-12}$  M). En un aspecto,  $K_D$  se mide mediante resonancia de plasmón superficial, por ejemplo, BIACORE™; en otro aspecto,  $K_D$  se mide mediante ELISA.

El término "aproximadamente" tal como se aplica en el presente documento a uno o más valores de interés, incluye un valor que es similar a un valor de referencia establecido. En determinados aspectos, el término "alrededor de" o "aproximadamente" incluye un intervalo de valores que se encuentran dentro de un 25 %, 20 %, 19 %, 18 %, 17 %, 16 %, 15 %, 14 %, 13 %, 12 %, 11 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, 1%, o menos en cualquier dirección (mayor o menor que) del valor de referencia indicado a menos que se indique lo contrario o sea evidente de otro modo en el contexto (excepto cuando dicho número exceda el 100 % de un posible valor).

La expresión "biológicamente activo" incluye una característica de cualquier agente que tenga actividad en un sistema biológico, *in vitro* o *in vivo* (por ejemplo, en un organismo). Por ejemplo, un agente que, cuando está presente en un organismo, tiene un efecto biológico dentro de ese organismo, se considera que es biológicamente activo. En aspectos particulares, cuando una proteína o polipéptido es biológicamente activo, una porción de esa proteína o polipéptido que comparte al menos una actividad biológica de la proteína o polipéptido se denomina normalmente una porción "biológicamente activa".

El término "vehículo", en el contexto de una molécula pequeña, por ejemplo, un vehículo unido a una molécula pequeña, se refiere a una macromolécula, generalmente una proteína, a la cual la molécula pequeña se puede acoplar para hacer que la molécula pequeña sea inmunogénica.

El término "afín", cuando se usa en el sentido de "afín a", *por ejemplo*, un primer dominio  $V_L$  que es "afín a" un segundo dominio  $V_L$ , pretende incluir una referencia a la relación entre dos dominios  $V_L$  de una misma proteína de unión producida por un ratón de acuerdo con la presente divulgación. Por ejemplo, un ratón que está genéticamente modificado de acuerdo con un aspecto de la presente divulgación, por ejemplo, un ratón que tiene un locus de cadena pesada en el que las regiones  $V_H$ ,  $D_H$  y  $J_H$  se reemplazan con regiones  $V_L$  y  $J_L$ , produce proteínas de unión similares a anticuerpos que tienen dos cadenas polipeptídicas idénticas hechas de la misma región  $C_H$  de ratón (por ejemplo, un isotipo IgG) fusionadas con un primer dominio  $V_L$  humano, y dos cadenas polipeptídicas idénticas hechas de la misma región  $C_L$  de ratón fusionadas con un segundo dominio  $V_L$  humano. Durante la selección clonal en el ratón, el primer y el segundo dominio  $V_L$  humano se seleccionaron mediante el proceso de selección clonal para aparecer juntos en el contexto de una única proteína de unión similar a un anticuerpo. Por lo tanto, El primer y segundo dominio  $V_L$  que aparecen juntos, como resultado del proceso de selección clonal, en una única molécula similar a un anticuerpo se denominan "afines". En contraste, un dominio  $V_L$  que aparece en una primera molécula similar a anticuerpo y un dominio  $V_L$  que aparece en una segunda molécula similar a anticuerpo no son afines, *a menos que* la primera y la

segunda molécula similar a anticuerpo tengan cadenas pesadas idénticas (es decir, a menos que el dominio VL fusionado con la primera región de cadena pesada humana y el dominio VL fusionado con la segunda región de cadena pesada humana sean idénticos).

5 La expresión "región determinante de la complementariedad", o el término "CDR", incluye una secuencia de aminoácidos codificada por una secuencia de ácido nucleico de los genes de inmunoglobulina de un organismo que, normalmente, (es decir, en un animal de tipo silvestre) aparece entre dos regiones marco en una región variable de una cadena ligera o pesada de una molécula de inmunoglobulina (por ejemplo, un anticuerpo o un receptor de linfocitos T). Una CDR puede ser codificada por, por ejemplo, una secuencia de línea germinal o una secuencia reordenada o  
10 no reordenada, y, por ejemplo, por un linfocito B o un linfocito T indiferenciado o maduro. Una CDR puede mutarse somáticamente (por ejemplo, variar de una secuencia codificada en la línea germinal de un animal), humanizarse y/o modificarse con sustituciones, adiciones o eliminaciones de aminoácidos. En algunas circunstancias (por ejemplo, para una CDR3), las CDR pueden codificarse por dos o más secuencias (por ejemplo, secuencias de la línea germinal) que no son contiguas (por ejemplo, en una secuencia de ácido nucleico no reordenada) pero que son contiguas en una secuencia de ácido nucleico de linfocitos B, por ejemplo, como resultado de empalmar o conectar las secuencias  
15 (por ejemplo, recombinación V-D-J para formar una CDR3 de cadena pesada).

El término "comparable" incluye dos o más agentes, entidades, situaciones, conjuntos de condiciones, etc. que pueden no ser idénticos entre sí, pero que son lo suficientemente similares como para permitir la comparación entre ellos, de modo que las conclusiones se puedan sacar razonablemente en base a diferencias o similitudes observadas. Los expertos en la técnica entenderán, en el contexto, qué grado de identidad se requiere en cualquier circunstancia dada para que dos o más de dichos agentes, entidades, situaciones, conjuntos de condiciones, se consideren comparables.

El término "conservadora" para describir una sustitución conservadora de aminoácidos incluye la sustitución de un resto de aminoácido por otro resto de aminoácido que tiene un grupo R de cadena lateral con propiedades químicas similares (por ejemplo, carga o hidrofobicidad). En general, una sustitución conservadora de aminoácidos no cambiará sustancialmente las propiedades funcionales de interés de una proteína, por ejemplo, la capacidad de un receptor para unirse a un ligando. Los ejemplos de grupos de aminoácidos que tienen cadenas laterales con propiedades químicas similares incluyen cadenas laterales alifáticas tales como glicina, alanina, valina, leucina e isoleucina; cadenas laterales hidroxil alifáticas tales como serina y treonina; cadenas laterales que contienen amida tales como asparagina y glutamina; cadenas laterales aromáticas tales como fenilalanina, tirosina y triptófano; cadenas laterales básicas tales como lisina, arginina e histidina; cadenas laterales ácidas tales como ácido aspártico y ácido glutámico; y, cadenas laterales que contienen azufre tales como cisteína y metionina. Los grupos de sustitución de aminoácidos conservadores incluyen, por ejemplo, valina/leucina/isoleucina, fenilalanina/arginina, lisina/arginina, alanina/valina, glutamato/aspartato, y asparagina/glutamina. En algunos aspectos, una sustitución de aminoácidos conservadora puede ser una sustitución de cualquier resto nativo en una proteína con alanina, como se usa en, por ejemplo, mutagénesis de barrido de alanina. En algunos aspectos, una sustitución conservadora es aquella que tiene un valor positivo en la matriz de probabilidad logarítmica PAM250 divulgada en Gonnet et al. ((1992) Exhaustive Matching of the Entire Protein Sequence Database, Science 256:1443-45. En algunos aspectos, una sustitución se considera  
40 "moderadamente conservadora" si tiene un valor no negativo en la matriz de probabilidad logarítmica PAM250.

En algunos aspectos, las posiciones de restos en una cadena ligera o pesada de inmunoglobulina difieren en una o más sustituciones conservadoras de aminoácidos. En algunos aspectos, las posiciones de restos en una cadena ligera de inmunoglobulina o fragmento funcional de la misma (por ejemplo, un fragmento que permite la expresión y secreción de, por ejemplo, un linfocito B) no son idénticas a una cadena ligera cuya secuencia de aminoácidos se enumera en el presente documento, pero difiere por una o más sustituciones conservadoras de aminoácidos.

El término "interrupción", cuando se usa fuera del contexto de "interrupción antigénica diferencial", incluye el resultado de un evento que interrumpe (por ejemplo, mediante recombinación homóloga) un ADN. En algunos aspectos, una interrupción puede lograr o representar una eliminación, inserción, inversión, modificación, reemplazo, sustitución o cualquier combinación de las mismas, de una secuencia o secuencias de ADN. En algunos aspectos, una interrupción puede lograr o representar la introducción de una mutación, como una mutación de aminoácido, interruptora o de cambio de marco, o cualquier combinación de las mismas, en una secuencia o secuencias codificantes en el ADN. En algunos aspectos, puede ocurrir una interrupción en un gen o locus génico endógeno a una célula. En algunos aspectos, las inserciones pueden incluir la inserción de genes completos o fragmentos de genes, por ejemplo, exones, en un sitio endógeno en una célula o genoma. En algunos aspectos, las inserciones pueden introducir secuencias que son de un origen diferente al de una secuencia endógena en la que se insertan. En algunos aspectos, una interrupción puede aumentar la expresión y/o actividad de un gen o producto génico (por ejemplo, de una proteína codificada por un gen). En algunos aspectos, una interrupción puede disminuir la expresión y/o actividad de un gen o producto génico.  
60 En algunos aspectos, una interrupción puede alterar la secuencia de un gen o producto génico (por ejemplo, una proteína codificada). En algunos aspectos, una interrupción puede truncar o fragmentar un gen o producto génico (por ejemplo, una proteína codificada). En algunos aspectos, una interrupción puede extender un gen o producto genético; en algunos de estos aspectos, una interrupción puede lograr el ensamblaje de una proteína de fusión. En algunos aspectos, una interrupción puede afectar el nivel pero no la actividad de un gen o producto genético. En algunos aspectos, una interrupción puede afectar la actividad pero no el nivel de un gen o producto genético. En algunos aspectos, una interrupción puede no tener un efecto significativo en el nivel de un gen o producto genético. En algunos

aspectos, una interrupción puede no tener un efecto significativo en la actividad de un gen o producto genético. En algunos aspectos, una interrupción puede no tener un efecto significativo tanto en el nivel como en la actividad de un gen o producto genético.

5 La expresión "locus endógeno" o "gen endógeno" incluye un locus genético encontrado en un progenitor u organismo de referencia antes de la introducción de una interrupción (por ejemplo, eliminación, inserción, inversión, modificación, reemplazo, sustitución o una combinación de las mismas como se describe en el presente documento). En algunos aspectos, un locus endógeno tiene una secuencia que se encuentra en la naturaleza. En algunos aspectos, un locus endógeno es de tipo silvestre. En algunos aspectos, un organismo de referencia que contiene un locus endógeno como se describe en el presente documento es un organismo de tipo silvestre. En algunos aspectos, un organismo de referencia que contiene un locus endógeno como se describe en el presente documento es un organismo genomanipulado. En algunos aspectos, un organismo de referencia que contiene un locus endógeno como se describe en el presente documento es un organismo criado en laboratorio (ya sea de tipo silvestre o genomanipulado).

15 La expresión "promotor endógeno" incluye un promotor que está naturalmente asociado, por ejemplo, en un organismo de tipo silvestre, con un gen endógeno.

20 La expresión "proteína de unión a epítipo" incluye una proteína que tiene al menos una CDR y que es capaz de reconocer selectivamente un epítipo, por ejemplo, es capaz de unir un epítipo con una  $K_D$  que es aproximadamente un micromolar o menos (por ejemplo, una  $K_D$  eso es aproximadamente  $1 \times 10^{-6}$  M,  $1 \times 10^{-7}$  M,  $1 \times 10^{-8}$  M,  $1 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M,  $1 \times 10^{-11}$  M, o aproximadamente  $1 \times 10^{-12}$  M). Las proteínas de unión a epítipo terapéuticas (por ejemplo, anticuerpos terapéuticos) requieren con frecuencia una  $K_D$  que está en el intervalo nanomolar o picomolar.

25 "Funcional", por ejemplo, en referencia a un polipéptido funcional, incluye un polipéptido que retiene al menos una actividad biológica normalmente asociada con la proteína nativa. En otro caso, un segmento génico de inmunoglobulina funcional puede incluir un segmento génico variable que es capaz de reordenación productiva para generar una secuencia de gen de inmunoglobulina reordenada.

30 La expresión "fragmento funcional" incluye fragmentos de proteínas de unión a epítipo que se puede expresar, secretar y se une específicamente a un epítipo con una  $K_D$  en el intervalo micromolar, nanomolar o picomolar. El reconocimiento específico incluye tener una  $K_D$  que esté al menos en el intervalo micromolar, el intervalo nanomolar o el intervalo picomolar.

35 La expresión "segmento génico", o "segmento" incluye una referencia a un segmento génico variable (V) (pesado o ligero), un segmento génico de diversidad (D) o un segmento génico de unión J (pesado o ligero), lo que incluye secuencias reordenadas en los loci de inmunoglobulina (en, por ejemplo, seres humanos y roedores) que pueden participar en una reordenación (mediada por, por ejemplo, recombinasas endógenas) para formar una secuencia génica V/J reordenada o V/D/J reordenada, cada una de las cuales puede estar unida operativamente a uno o más segmentos génicos constantes (C) (pesados o ligeros). A menos que se indique otra cosa, los segmentos V, D y J comprenden secuencias de señal de recombinación (RSS, de sus siglas en inglés) que permiten la recombinación V/J o la recombinación V/D/J de acuerdo con la regla 12/23. El segmento génico también incluye referencia a un segmento génico de región constante (pesado o ligero), que puede comprender en el extremo 5' del ADN repetitivo del segmento génico de región constante conocido como una región de cambio que permite la recombinación específica del sitio que da como resultado el cambio de isotipo. Una secuencia génica de región constante de cadena pesada puede comprender un segmento génico de región constante de cadena pesada o un grupo de segmentos génicos de región constante de cadena pesada, por ejemplo, en la organización de la línea germinal, cuyo grupo también puede comprender preferentemente en 5' de cada segmento génico de región constante de cadena pesada una región de cambio que permite el cambio de isotipo mediante recombinación específica del sitio. A menos que se indique otra cosa, los segmentos comprenden además secuencias con las que están asociados en la naturaleza o equivalentes funcionales de los mismos (por ejemplo, para el (los) promotor(es) y líder(es) de los segmentos V).

50 La expresión "línea germinal" en referencia a una secuencia de ácido nucleico de inmunoglobulina incluye una secuencia de ácido nucleico que puede pasarse a la prole.

55 La expresión "cadena pesada de inmunoglobulina", "cadena pesada" o similar generalmente se refiere a una proteína de inmunoglobulina de longitud completa que incluye, desde el extremo amino al terminal carboxilo, un dominio variable de cadena pesada ( $V_H$ ) y un dominio constante de cadena pesada ( $C_H$ ), e incluye cadenas pesadas que carecen de un dominio  $CH1$ , y, opcionalmente, que carecen adicionalmente de una región bisagra. Una secuencia de cadena pesada de inmunoglobulina puede ser de cualquier organismo.

60 Un "dominio variable de cadena pesada" se refiere a un dominio de inmunoglobulina que tiene una secuencia de aminoácidos que está codificada preferentemente por que procede de un gen de región variable de cadena pesada reordenado, que generalmente comprende secuencias de un segmento génico variable de cadena pesada ( $V_H$ ) (o una porción del mismo), un segmento génico de diversidad de cadena pesada ( $D_H$ ) (o una porción del mismo) y un segmento génico de unión de cadena pesada ( $J_H$ ) (o una porción del mismo). En aspectos preferidos, la secuencia génica de región variable de cadena pesada, por ejemplo, la secuencia génica  $V_H$ ,  $-D_H$ -  $J_H$  reordenada, procede de un

repertorio de segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  no reordenados, preferentemente segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  no reordenados, capaces de experimentar una reordenación productiva de genes, por ejemplo, capaces de unirse para formar una secuencia génica de región variable de cadena pesada en el marco. Los segmentos génicos  $V_H$ , los segmentos génicos  $D_H$  o los segmentos génicos  $J_H$  incluyen segmentos génicos  $V_H$ , segmentos génicos  $D_H$  o segmentos génicos  $J_L$  de cualquier organismo, incluidos, pero sin limitación, roedores (por ejemplo, ratones, ratas, etc.) y seres humanos. Un dominio variable de cadena pesada que comprende mutaciones somáticas (por ejemplo, aminoácidos no codificados por la secuencia de la línea germinal de un segmento génico  $V_H$ ,  $D_H$  y/o  $J_H$ ), y el gen de región variable de cadena pesada reordenado que codifica el mismo, puede considerarse independientemente procedente de los segmentos génicos  $V_H$ ,  $D_H$  y/o  $J_H$  de la línea germinal, o porciones de los mismos, que se reordenan productivamente para formar el gen que codifica el dominio variable de cadena pesada en primera instancia, por ejemplo, antes de la proliferación mediada por antígeno.

Un dominio variable de cadena pesada de inmunoglobulina normalmente incluye, desde el extremo amino al extremo carboxilo, tres regiones determinantes de la complementariedad (CDR) de cadena pesada y cuatro regiones marco (FR) de cadena pesada, por ejemplo, FRH1-CDRH1-FRH2-CDRH2-FRH3-CDRH3-FRH4, a menos que se especifique otra cosa. Un dominio  $V_H$  también puede referirse a la porción de una cadena pesada que se extiende (desde el extremo N hasta el extremo C) desde el extremo N de la cadena pesada hasta el extremo N de un dominio constante de cadena pesada.

Un dominio constante de cadena pesada ( $C_H$ ) se refiere a un dominio de inmunoglobulina que tiene una secuencia de aminoácidos que está codificada preferentemente por un segmento génico de región constante de cadena pesada, o una porción del mismo, de cualquier organismo. Los segmentos génicos de región constante de cadena pesada ejemplares incluyen, aunque no de forma limitativa, un segmento génico  $C_\mu$ , un segmento génico  $C_\delta$ , un segmento génico  $C_\gamma$  (por ejemplo,  $C_\gamma 1$ ,  $C_\gamma 2$ ,  $C_\gamma 3$ ,  $C_\gamma 4$ ), un segmento génico  $C_\alpha$  (por ejemplo,  $C_\alpha 1$ ,  $C_\alpha 2$ ), o un segmento génico  $C_\epsilon$ , que codifican un dominio constante de cadena pesada de IgM, IgD, IgG, IgA, o IgE, respectivamente. Un segmento génico de región constante de cadena pesada típico comprende normalmente exones que codifican cada uno un dominio  $C_{H1}$ , una bisagra, un dominio  $C_{H2}$ , un dominio  $C_{H3}$ , opcionalmente, un dominio  $C_{H4}$  (por ejemplo, en el caso de IgM o IgE) y, opcionalmente, un dominio transmembrana (M) (por ejemplo, en el caso de la inmunoglobulina unida a la membrana en linfocitos). Un dominio  $C_H$  también puede referirse a un dominio de inmunoglobulina que tiene una secuencia de aminoácidos que está codificada por un gen de región constante de cadena pesada que carece de una región  $C_{H1}$  funcional y, opcionalmente, carece adicionalmente de una región bisagra funcional. Generalmente, un dominio  $C_H$  también puede referirse a la porción de una cadena pesada que se extiende (desde el lado N-terminal al lado C-terminal) desde fuera de FR4 hasta el extremo C de la cadena pesada. Un dominio  $C_H$  también puede referirse a la porción de una cadena híbrida que se extiende (desde el lado N-terminal al lado C-terminal) desde fuera de FR4 hasta el extremo C de la cadena híbrida.

Los dominios constantes de cadena pesada con desviaciones menores, por ejemplo, truncamientos de uno, dos, tres o varios aminoácidos del extremo C, estarían abarcados por la expresión "dominio constante de cadena pesada", así como los dominios constantes de cadena pesada con modificaciones de secuencia, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 sustituciones de aminoácidos. Las sustituciones de aminoácidos se pueden realizar en una o más posiciones seleccionadas, por ejemplo, (con referencia a la numeración en la UE de un dominio constante de inmunoglobulina, por ejemplo, un dominio constante de IgG humano), 228, 233, 234, 235, 236, 237, 238, 239, 241, 248, 249, 250, 252, 254, 255, 256, 258, 265, 267, 268, 269, 270, 272, 276, 278, 280, 283, 285, 286, 289, 290, 292, 293, 294, 295, 296, 297, 298, 301, 303, 305, 307, 308, 309, 311, 312, 315, 318, 320, 322, 324, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 337, 338, 339, 340, 342, 344, 356, 358, 359, 360, 361, 362, 373, 375, 376, 378, 380, 382, 383, 384, 386, 388, 389, 398, 414, 416, 419, 428, 430, 433, 434, 435, 437, 438, and 439.

Por ejemplo, y no a modo de limitación, un dominio constante de cadena pesada puede modificarse para exhibir una semividua en suero mejorada (en comparación con el mismo dominio constante de cadena pesada sin las modificaciones mencionadas) y tener una modificación en la posición 250 (por ejemplo, E o Q); 250 y 428 (por ejemplo, L o F); 252 (por ejemplo, L/Y/F/W o T); 254 (por ejemplo, S o T) y 256 (por ejemplo, S/R/Q/E/D o T); o una modificación en 428 y/o 433 (por ejemplo, L/R/P/Q o K) y/o 434 (por ejemplo, H/F o Y); o una modificación en 250 y/o 428; o una modificación en 307 o 308 (por ejemplo, 308F, V308F) y 434. En otro ejemplo, la modificación puede comprender una modificación 428L (por ejemplo, M428L) y 434S (por ejemplo, N434S); una modificación 428L, 259I (por ejemplo, V259I) y 308F (por ejemplo, V308F); una modificación 433K (por ejemplo, H433K) y 434 (por ejemplo, 434Y); una modificación en 252, 254 y 256 (por ejemplo, 252Y, 254T y 256E); una modificación 250Q y 428L (por ejemplo, T250Q y M428L); una modificación en 307 y/o 308 (por ejemplo, 308F o 308P). Los restos se numeran según el sistema de numeración de la UE. En otro ejemplo no limitante, un dominio constante de cadena pesada puede modificarse para exhibir una afinidad cambiada a la proteína A, que puede ser útil en el aislamiento de anticuerpos bispecíficos, véase, por ejemplo, la Patente de EE.UU. N.º 8.586.713.

El término "heterólogo" incluye un agente o entidad de una fuente diferente. Por ejemplo, cuando se usa en referencia a un polipéptido, gen o producto génico o presente en una célula u organismo particular, el término aclara que el polipéptido, gen o producto génico oportuno 1) se genomanipuló por la mano del hombre; 2) se introdujo en la célula u organismo (o un precursor del mismo) a través de la mano del hombre (por ejemplo, mediante ingeniería genética); y/o 3) no se produce naturalmente ni está presente en la célula u organismo oportuno (por ejemplo, el tipo de célula o

tipo de organismo oportuno).

La expresión "célula hospedadora" incluye una célula en la que se ha introducido un ácido nucleico o proteína heterólogo (por ejemplo, exógeno). Las personas expertas al leer esta divulgación comprenderán que dichos términos incluyen no solo una célula sujeto particular, sino que también se usan para incluir la progenie de esa célula. Debido a que pueden producirse determinadas mutaciones en generaciones posteriores debido a una mutación o a influencias ambientales, dicha progenie puede, de hecho, no ser idéntica a la célula precursora, pero los expertos en la técnica todavía entienden que están incluidos dentro del alcance de la expresión "célula hospedadora". En algunos aspectos, una célula hospedadora es o comprende una célula procariota o eucariota. En general, una célula hospedadora es cualquier célula que sea adecuada para recibir y/o producir un ácido nucleico o proteína heterólogo, independientemente del Reino de la vida al que esté designada la célula. Las células ejemplares que pueden utilizarse como células hospedadoras de acuerdo con la presente divulgación incluyen las de procariotas y eucariotas (células individuales o células múltiples), células bacterianas (por ejemplo, cepas de *E. coli*, *Bacillus* spp., *Streptomyces* spp., etc.), células de micobacterias, células fúngicas, células de levadura (por ejemplo, *S. cerevisiae*, *S. pombe*, *P. pastoris*, *P. methanolica*, etc.), células vegetales, células de insecto (por ejemplo, SF-9, SF-21, células de insecto infectadas con baculovirus, *Trichoplusia ni*, etc.), células de animales no humanos, células humanas o células de fusión tales como, por ejemplo, hibridomas o cuadromas. En algunos aspectos, la célula es una célula humana, de mono, simios, hámster, rata o ratón. En algunos aspectos, la célula es eucariota y se selecciona de las siguientes células: CHO (por ejemplo, CHO K1, DXB-11 CHO, Veggie-CHO), COS (por ejemplo, COS-7), célula retiniana, Vero, CV1, de riñón (por ejemplo, HEK293, 293 EBNA, MSR 293, MDCK, HaK, BHK), HeLa, HepG2, WI38, MRC 5, Colo205, HB 8065, HL-60, (por ejemplo, BHK21), Jurkat, Daudi, A431 (epidérmica), CV-1, U937, 3T3, célula L, célula C127, SP2/0, NS-0, MMT 060562, célula de Sertoli, linfocito BRL 3A, célula HT1080, célula de mieloma, célula tumoral y una línea celular procedente de una célula mencionada anteriormente. En algunos aspectos, la célula comprende uno o más genes víricos, por ejemplo, una célula retiniana que expresa un gen vírico (por ejemplo, una célula PER.C6™). En algunos aspectos, una célula hospedadora es o comprende una célula aislada. En algunos aspectos, una célula hospedadora es parte de un tejido. En algunos aspectos, una célula hospedadora es parte de un organismo.

El término "humanizado" entendido en la técnica incluye ácidos nucleicos o proteínas cuyas estructuras (es decir, secuencias de nucleótidos o aminoácidos) incluyen porciones que corresponden sustancial o idénticamente con versiones de los ácidos nucleicos o proteínas oportunos que se encuentran en la naturaleza en animales no humanos y que son distinguibles de las versiones correspondientes que se encuentran en la naturaleza en seres humanos, y también incluyen porciones cuyas estructuras difieren de las presentes en las versiones de animales no humanos y, en cambio, corresponden más estrechamente con estructuras comparables encontradas en las versiones humanas. En algunos aspectos, un gen "humanizado" es uno que codifica un polipéptido que tiene sustancialmente la secuencia de aminoácidos como la de un polipéptido humano (por ejemplo, una proteína humana o una porción de la misma, por ejemplo, una porción característica de la misma). Por dar un solo ejemplo, en el caso de un receptor de membrana, un gen "humanizado" puede codificar un polipéptido con una porción extracelular cuya secuencia de aminoácidos es idéntica o sustancialmente idéntica a la de una porción extracelular humana, y cuya secuencia restante es idéntica o sustancialmente idéntica a la de un polipéptido no humano (por ejemplo, de ratón). En algunos aspectos, un gen humanizado comprende al menos una porción de una secuencia de ADN de un gen humano. En algún aspecto, un gen humanizado comprende una secuencia completa de ADN que se encuentra en un gen humano. En algunos aspectos, una proteína humanizada tiene una secuencia de aminoácidos que comprende una porción que aparece en una proteína humana. En algunos aspectos, una proteína humanizada tiene una secuencia de aminoácidos cuya secuencia completa se encuentra en una proteína humana. En algunos aspectos (incluidos, por ejemplo, algunos en los que una proteína humanizada tiene una secuencia de aminoácidos cuya secuencia completa se encuentra en una proteína humana), una proteína humanizada se expresa a partir de un locus endógeno de un animal no humano, cuyo locus endógeno corresponde al homólogo u ortólogo del gen humano oportuno que codifica la proteína.

El término "identidad" en relación con una comparación de secuencias incluye la identidad determinada por cualquiera de varios algoritmos diferentes conocidos en la técnica que pueden usarse para medir la identidad de secuencia de nucleótidos y/o aminoácidos. En algunos aspectos, como se describe en el presente documento, se determinan las identidades usando un alineación ClustalW v. 1.83 (lento) que emplea una penalización por apertura de hueco de 10,0, una penalización por extensión de hueco de 0,1, y que utiliza una matriz de similitud de Gonnet (MACVECTOR™ 10.0.2, MacVector Inc., 2008). El término "identidad" incluye la relación general entre moléculas poliméricas, por ejemplo, entre moléculas de ácido nucleico (por ejemplo, moléculas de ADN y/o moléculas de ARN) y/o entre moléculas polipeptídicas. En algunos aspectos, las moléculas poliméricas se consideran "sustancialmente idénticas" entre sí si sus secuencias son al menos un 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 % idénticas. Como entenderán los expertos en la técnica, hay una variedad de algoritmos disponibles que permiten la comparación de secuencias para determinar su grado de homología, incluso permitiendo espacios de longitud designada en una secuencia con respecto a otra cuando se consideran qué restos "corresponden" el uno al otro en diferentes secuencias. El cálculo del porcentaje de identidad entre dos secuencias de ácido nucleico, por ejemplo, se puede realizar mediante la alineación de las dos secuencias para fines de comparación óptimos (por ejemplo, se pueden introducir huecos en una o ambas de la primera y segunda secuencia de ácido nucleico para una alineación óptima y las secuencias no correspondientes se pueden descartar para fines de comparación). En determinados aspectos, la longitud de una secuencia alineada con fines comparativos es al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos

un 95 %, o sustancialmente el 100 % de la longitud de la secuencia de referencia. Luego se comparan los nucleótidos en las posiciones de nucleótidos correspondientes. Cuando se ocupa una posición en la primera secuencia por el mismo nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos, y la longitud de cada hueco, que necesita introducirse para un alineación óptimo de las dos secuencias. Los algoritmos representativos y los programas informáticos útiles para determinar el porcentaje de identidad entre dos secuencias de nucleótidos incluyen, por ejemplo, el algoritmo de Meyers y Miller (CABIOS, 1989, 4: 11-17), que se ha incorporado al programa ALIGN (versión 2.0) utilizando una tabla de restos de peso PAM120, una penalización de longitud de hueco de 12 y una penalización de hueco de 4. El porcentaje de identidad entre dos secuencias de nucleótidos se puede determinar, alternativamente, por ejemplo, usando el programa GAP en el paquete de programa informático GCG usando una matriz NWSgapdna.CMP.

El término "aislado" incluye una sustancia y/o entidad que ha sido (1) separada de al menos algunos de los componentes con los que se asoció cuando se produjo inicialmente (ya sea en la naturaleza y/o en un entorno experimental), y/o (2) diseñada, producida, preparada y/o fabricada por la mano del hombre. Las sustancias y/o entidades aisladas se pueden separar de aproximadamente un 10 %, aproximadamente un 20 %, aproximadamente un 30 %, aproximadamente un 40 %, aproximadamente un 50 %, aproximadamente un 60 %, aproximadamente un 70 %, aproximadamente un 80 %, aproximadamente un 90 %, aproximadamente un 91 %, aproximadamente un 92 %, aproximadamente un 93 %, aproximadamente un 94 %, aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 %, aproximadamente un 99 %, o más de aproximadamente un 99 % de los otros componentes con los que se asociaron inicialmente. En algunos aspectos, los agentes aislados son aproximadamente un 80 %, aproximadamente un 85 %, aproximadamente un 90 %, aproximadamente un 91 %, aproximadamente un 92 %, aproximadamente un 93 %, aproximadamente un 94 %, aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 %, aproximadamente un 99 %, o más de aproximadamente un 99 % puros. Una sustancia es "pura" si está sustancialmente libre de otros componentes. En algunos aspectos, como entenderán los expertos en la técnica, una sustancia todavía puede considerarse "aislada" o incluso "pura", después de haber sido combinada con determinados otros componentes tales como, por ejemplo, uno o más vehículos o excipientes (por ejemplo, tampón, disolvente, agua, etc.); en dichos aspectos, el porcentaje de aislamiento o pureza de la sustancia se calcula sin incluir dichos vehículos o excipientes. Por dar un solo ejemplo, en algunos aspectos, un polímero biológico tal como un polipéptido o polinucleótido que se produce en la naturaleza se considera "aislado" cuando, a) en virtud de su origen o fuente de procedencia no está asociado con algunos o todos los componentes que lo acompañan en su estado nativo en la naturaleza; b) está sustancialmente libre de otros polipéptidos o ácidos nucleicos de la misma especie de la especie que lo produce en la naturaleza; c) se expresa o está asociado con componentes de una célula u otro sistema de expresión que no es de la especie que lo produce en la naturaleza. Por lo tanto, por ejemplo, en algunos aspectos, un polipéptido que se sintetiza químicamente o se sintetiza en un sistema celular diferente del que lo produce en la naturaleza se considera un polipéptido "aislado". Alternativa o adicionalmente, en algunos aspectos, un polipéptido que ha sido sometido a una o más técnicas de purificación puede considerarse un polipéptido "aislado" en la medida en que se ha separado de otros componentes a) con los que está asociado en la naturaleza; y/o b) con los que estaba asociado cuando se produjo inicialmente.

Un "dominio variable de cadena ligera" se refiere a un dominio de inmunoglobulina que tiene una secuencia de aminoácidos que está codificada preferentemente por que procede de un gen de región variable de cadena ligera reordenado, que generalmente comprende secuencias de un segmento génico variable de cadena ligera ( $V_L$ ) (o una porción del mismo) y un segmento génico de unión de cadena ligera ( $J_L$ ) (o una porción del mismo). En aspectos preferidos, la secuencia génica de región variable de cadena ligera, por ejemplo, la secuencia génica  $V_L$ - $J_L$  reordenada, procede de un repertorio de segmentos génicos  $V_L$  no reordenados y/o  $J_L$  no reordenados, preferentemente segmentos génicos  $V_L$  no reordenados de la línea germinal y/o segmentos génicos  $J_L$  no reordenados de la línea germinal, capaces de experimentar una reordenación productiva de genes, por ejemplo, capaces de reordenarse para formar una secuencia génica de región variable de cadena ligera en el marco. Los segmentos génicos  $V_L$  o segmentos génicos  $J_L$  incluyen segmentos génicos  $V_L$  o segmentos génicos  $J_L$  de cualquier organismo, incluidos, pero sin limitación, roedores (por ejemplo, ratones, ratas, etc.) y seres humanos. Un dominio variable de cadena ligera que comprende mutaciones somáticas (por ejemplo, aminoácidos no codificados por la secuencia de la línea germinal de un segmento génico  $V_L$  y/o  $J_L$ ), y el gen de región variable de cadena ligera reordenado que codifica el mismo, puede considerarse independientemente procedente de los segmentos génicos  $V_L$  y  $J_L$  de la línea germinal, o porciones de los mismos, que se reordenan productivamente para formar el gen que codifica el dominio variable de cadena ligera en primera instancia, por ejemplo, antes de la proliferación mediada por antígeno.

Un dominio variable de cadena ligera de inmunoglobulina normalmente incluye, desde el extremo amino al extremo carboxilo, tres regiones determinantes de la complementariedad (CDR) de cadena ligera y cuatro regiones marco (FR), por ejemplo, FRL1-CDRL1-FRL2-CDRL2-FRL3-CDRL3-FRL4, a menos que se especifique otra cosa. Un dominio  $V_L$  también puede referirse a la porción de una cadena ligera que se extiende (desde el extremo N hasta el extremo C) desde el extremo N de la cadena ligera hasta el extremo N de un dominio constante de cadena ligera de la cadena ligera. Un dominio  $V_L$  también puede referirse a la porción de una cadena híbrida que se extiende (desde el extremo N hasta el extremo C) desde el extremo N de la cadena híbrida hasta el extremo N de un dominio constante de cadena pesada de la cadena híbrida.

Un dominio constante de cadena ligera (C<sub>L</sub>) se refiere a un dominio de inmunoglobulina que tiene una secuencia de aminoácidos que está codificada preferentemente por un gen de región constante de cadena ligera de cualquier organismo, tal como, pero sin limitación, una secuencia de aminoácidos codificada por un segmento génico C<sub>k</sub> o C<sub>λ</sub>, por ejemplo, un segmento génico C<sub>k</sub> o C<sub>λ</sub> de roedor o humano. Dichos dominios C<sub>k</sub> o C<sub>λ</sub> son bien conocidos en la técnica. Generalmente, un dominio C<sub>L</sub> también puede referirse a la porción de una cadena ligera que se extiende (desde el extremo N al extremo C) fuera de un FRL4 al extremo C de la cadena ligera.

La expresión "intervalo micromolar" pretende significar 1-999 micromolar; la expresión "intervalo nanomolar" pretende significar 1-999 nanomolar; la expresión "intervalo picomolar" pretende significar 1-999 picomolar.

La expresión "cadena de inmunoglobulina híbrida", "cadena híbrida", "cadena híbrida de inmunoglobulina", o similar se refiere a una proteína de inmunoglobulina que incluye, desde el extremo amino al carboxilo, un dominio variable de cadena ligera (que puede o no estar mutado somáticamente) y un dominio constante de cadena pesada. Generalmente, una cadena híbrida está codificada por una secuencia génica de región variable de cadena ligera reordenada unida operativamente a una secuencia génica de región constante de cadena pesada. La secuencia génica de región variable de cadena ligera de una cadena de inmunoglobulina híbrida puede comprender generalmente secuencias del segmento génico variable de cadena ligera (V<sub>L</sub>) (o una porción del mismo) y un segmento génico de unión de cadena ligera (J<sub>L</sub>). En aspectos preferidos, la secuencia génica de región variable de cadena ligera, por ejemplo, la secuencia génica V<sub>L</sub>-J<sub>L</sub> reordenada, que codifica el dominio variable de cadena procede de un repertorio de segmentos génicos V<sub>L</sub> y J<sub>L</sub> no reordenados, preferentemente segmentos génicos V<sub>L</sub> y J<sub>L</sub> no reordenados de la línea germinal, que son (a) capaces de experimentar un reordenamiento productivo de genes, por ejemplo, capaces de reorganizarse para formar una secuencia génica de región variable de cadena ligera dentro del marco y (b) unirse operativamente a uno o más segmentos génicos de región constante de cadena pesada, por ejemplo, un segmento génico de región constante o un grupo de segmentos génicos de región constante no reordenado.

La expresión "animal no humano" incluye un organismo vertebrado que no es humano. En algunos aspectos, un animal no humano es un ciclostomo, un pez óseo, un pez cartilaginoso (por ejemplo, un tiburón o una raya), un anfibio, un reptil, un mamífero o un pájaro. En algunos aspectos, un mamífero no humano es un primate, una cabra, una oveja, un cerdo, un perro, una vaca o un roedor. En algunos aspectos, un animal no humano es un roedor tal como una rata o un ratón.

La expresión "ácido nucleico" en su sentido más amplio, incluye cualquier compuesto y/o sustancia que esté o pueda incorporarse en una cadena de oligonucleótidos. En algunos aspectos, un ácido nucleico es un compuesto y/o sustancia que se incorpora o puede incorporarse a una cadena de oligonucleótidos a través de un enlace fosfodiéster. Como quedará claro por el contexto, en algunos aspectos, "ácido nucleico" incluye uno o más restos de ácido nucleico individuales (por ejemplo, nucleótidos y/o nucleósidos); en algunos aspectos, "ácido nucleico" incluye una cadena de oligonucleótidos que comprende restos de ácido nucleico individuales.

"Unido operativamente" también se refiere a una relación en donde los componentes unidos operativamente funcionan de la manera prevista. En un caso, una secuencia de ácido nucleico que codifica una proteína puede unirse operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) con el fin de retener la regulación de la transcripción adecuada. En un caso, una secuencia de ácido nucleico de una región variable de inmunoglobulina (o segmentos V(D)J) puede estar unida operativamente a una secuencia de ácido nucleico de una región constante de inmunoglobulina para permitir la recombinación adecuada entre las secuencias en una secuencia génica de cadena pesada o ligera de inmunoglobulina reordenada.

El término "polipéptido" incluye cualquier cadena polimérica de aminoácidos. En algunos aspectos, un polipéptido tiene una secuencia de aminoácidos que se produce en la naturaleza. En algunos aspectos, un polipéptido tiene una secuencia de aminoácidos que no se produce en la naturaleza. En algunos aspectos, un polipéptido tiene una secuencia de aminoácidos que está genomanipulada para que esté diseñada y/o producida por la acción de la mano del hombre.

El término "recombinante" pretende incluir polipéptidos (por ejemplo, proteínas del factor de activación de linfocitos B como se describe en el presente documento) que están diseñados, genomanipulados, preparados, expresados, creados o aislados por medios recombinantes, tales como los polipéptidos expresados usando un vector de expresión recombinante transfectado en una célula hospedadora, polipéptidos aislados de una biblioteca recombinante combinatoria de polipéptidos humanos (Hoogenboom HR, (1997) TIB Tech. 15:62-70; Azzazy H., y High-smith W. E., (2002) Clin. Biochem. 35:425-445; Gaviolondo J. V., y Larrick J. W. (2002) BioTechniques 29:128-145; Hoogenboom H., y Chames P. (2000) Immunology Today 21:371-378), anticuerpos aislados de un animal (por ejemplo, un ratón) que es transgénico para genes de inmunoglobulina humana (véase, por ejemplo, Taylor, L. D., et al. (1992) Nucl. Acids Res. 20:6287-6295; Kellermann S-A., y Green L. L. (2002) Current Opinion in Biotechnology 13:593-597; Little M. et al (2000) Immunology Today 21:364-370) o polipéptidos preparados, expresados, creados o aislados por cualquier otro medio que implique empalmar elementos de secuencia seleccionados entre sí. En algunos aspectos, uno o más de dichos elementos de secuencia seleccionados se encuentran en la naturaleza. En algunos aspectos, uno o más de dichos elementos de secuencia seleccionados se diseñan *in silico*. En algunos aspectos, uno o más de dichos

elementos de secuencia seleccionados resultan de la mutagénesis (por ejemplo, *in vivo* o *in vitro*) de un elemento de secuencia conocido, por ejemplo, de una fuente natural o sintética. Por ejemplo, en algunos aspectos, un polipéptido recombinante se compone de secuencias encontradas en el genoma de un organismo fuente de interés (por ejemplo, ser humano, ratón, etc.). En algunos aspectos, un polipéptido recombinante tiene una secuencia de aminoácidos que resultó de la mutagénesis (por ejemplo, *in vitro* o *in vivo*, por ejemplo, en un animal no humano), de modo que las secuencias de aminoácidos de los polipéptidos recombinantes son secuencias que, mientras que se originan y se relacionan con secuencias polipeptídicas, pueden no existir naturalmente dentro del genoma de un animal no humano *in vivo*.

El término "referencia" se usa en el presente documento para describir un agente o valor estándar o de control con el que se compara un agente o valor de interés. En algunos aspectos, se prueba un agente de referencia y/o se determina un valor de referencia sustancialmente de manera simultánea con la prueba o determinación del agente o valor de interés. En algunos aspectos, un agente o valor de referencia es una referencia histórica, opcionalmente incorporada en un medio tangible. Normalmente, como entenderían los expertos en la técnica, un agente o valor de referencia se determina o caracteriza en condiciones comparables a las utilizadas para determinar o caracterizar el agente o valor de interés. En algunos aspectos, los animales no humanos de control o de referencia (por ejemplo, ratones) se proporcionan en el presente documento e incluyen animales no humanos genéticamente modificados cuyos genomas expresan moléculas de inmunoglobulina tradicionales (es decir, inmunoglobulinas que tienen dominios V<sub>H</sub> y V<sub>L</sub> afines). En algunos aspectos determinados, los animales no humanos genéticamente modificados de control incluyen ratones humanizados VELOCIMMUNE® (véanse, por ejemplo, las patentes de EE.UU. n.º 8.502.018 y 8.642.835) y/o "ratones ULC" (véanse los documentos US 2011-0195454A1, US 2012-0021409A1, US 2012-0192300A1, US 2013-0045492A1, US 2013-0185821A1 y US 2013-0302836A1).

El término "reemplazo" se usa en el presente documento para incluir un proceso a través del cual una secuencia de ácido nucleico "reemplazada" (por ejemplo, un gen) que se encuentra en un locus hospedador (por ejemplo, en un genoma) se elimina de ese locus y un ácido nucleico "de reemplazo" diferente se ubica en su lugar. En algunos aspectos, la secuencia de ácido nucleico reemplazada y las secuencias de ácido nucleico de reemplazo son comparables entre sí en el sentido de que, por ejemplo, son homólogas entre sí y/o contienen elementos correspondientes (por ejemplo, elementos codificadores de proteínas, elementos reguladores, etc.). En algunos aspectos, una secuencia de ácido nucleico reemplazada incluye uno o más de un promotor, un potenciador, un sitio donante de empalme, un sitio receptor de empalme, un intrón, un exón, una región no traducida (UTR, de sus siglas en inglés); en algunos aspectos, una secuencia de ácido nucleico de reemplazo incluye una o más secuencias codificantes. En algunos aspectos, una secuencia de ácido nucleico de reemplazo es un homólogo de la secuencia de ácido nucleico reemplazada. En algunos aspectos, una secuencia de ácido nucleico de reemplazo es un ortólogo de la secuencia reemplazada. En algunos aspectos, una secuencia de ácido nucleico de reemplazo es o comprende una secuencia de ácido nucleico humana. En algunos aspectos, incluyendo donde la secuencia de ácido nucleico de reemplazo es o comprende una secuencia de ácido nucleico humana, la secuencia de ácido nucleico reemplazada es o comprende una secuencia de roedor (por ejemplo, una secuencia de ratón). La secuencia de ácido nucleico así colocada puede incluir una o más secuencias reguladoras que son parte de la secuencia de ácido nucleico fuente utilizada para obtener la secuencia así colocada (por ejemplo, promotores, potenciadores, regiones no traducidas en 5' o 3', etc.). Por ejemplo, en diversos aspectos, el reemplazo es una sustitución de una secuencia endógena con una secuencia heteróloga que da como resultado la producción de un producto génico a partir de la secuencia de ácido nucleico así colocada (que comprende la secuencia heteróloga), pero no la expresión de la secuencia endógena; el reemplazo es de una secuencia genómica endógena con una secuencia de ácido nucleico que codifica una proteína que tiene una función similar a una proteína codificada por la secuencia endógena (por ejemplo, la secuencia genómica endógena codifica un dominio, y el fragmento de ADN codifica uno o más dominios variables humanos). En diversos aspectos, un gen endógeno o fragmento del mismo se sustituye con un gen humano correspondiente o fragmento del mismo. Un gen humano correspondiente o fragmento del mismo es un gen humano o fragmento que es un ortólogo de, o es sustancialmente similar o igual en estructura y/o función, que el gen endógeno o fragmento del mismo que se reemplaza.

La expresión "molécula pequeña" incluye un compuesto orgánico cuyo peso molecular, en ausencia de un vehículo, es menor que aproximadamente 6 kilodaltons (kD) de tamaño, y que puede extraerse de fuentes naturales o producirse sintéticamente (xenobiótico). Las "moléculas pequeñas" también pueden comprender compuestos orgánicos que además comprenden átomos inorgánicos, por ejemplo, metales en complejo. "Molécula pequeña" puede referirse a un hapteno, por ejemplo, una molécula que puede unirse a proteínas de unión a antígeno en formato de inmunoglobulina tradicional pero no puede provocar una respuesta inmunitaria adaptativa. En algunos aspectos, la molécula pequeña, en ausencia de un vehículo, es inferior a aproximadamente 5 kD, 4 kD, 3 kD, aproximadamente 2 kD, o aproximadamente 1 kD. En algunos aspectos, el peso molecular de la molécula pequeña, en ausencia de un vehículo, como se describe en el presente documento, varía de 1 kD a 6 kD. En algunos aspectos, el peso molecular de la molécula pequeña, en ausencia de un vehículo, es inferior a 1,5 kD. En algunos aspectos determinados, el peso molecular de la molécula pequeña, en ausencia de un vehículo, como se describe en el presente documento, es inferior a 1400 daltons (D), inferior a 1300 D, inferior a 1200 D, inferior a 1100 D, inferior a 1000 D, inferior a 900 D, inferior a 800 D, inferior a 700 D, inferior a 600 D, inferior a 500 D, inferior a 400 D, inferior a 300 D, inferior a 200 D, o inferior a 100 D. En algunos aspectos, la molécula pequeña, en ausencia de un vehículo, es inferior a aproximadamente 800 daltons (D), aproximadamente 600 D, aproximadamente 500 D, aproximadamente 400 D, aproximadamente 300 D,



aproximadamente 200 D, o aproximadamente 100 D. En algunos aspectos, una molécula pequeña, en ausencia de un vehículo, es inferior a aproximadamente 2000 g/mol, inferior a aproximadamente 1500 g/mol, inferior a aproximadamente 1000 g/mol, inferior a aproximadamente 800 g/mol o inferior a aproximadamente 500 g/mol. En algunos aspectos, una molécula pequeña no es un polímero. En algunos aspectos, una molécula pequeña no incluye una fracción polimérica. En algunos aspectos, una molécula pequeña no es una proteína o polipéptido (por ejemplo, no es un oligopéptido o péptido). En algunos aspectos, una molécula pequeña no es un polinucleótido (por ejemplo, no es un oligonucleótido). En algunos aspectos, una molécula pequeña no es un polisacárido. En algunos aspectos, una molécula pequeña no comprende un polisacárido (por ejemplo, no es una glucoproteína, proteoglicano, glucolípido, *etc.*). En algunos aspectos, una molécula pequeña no es un lípido. En algunos aspectos, una molécula pequeña es un agente modulador. En algunos aspectos, una molécula pequeña es biológicamente activa. En algunos aspectos, una molécula pequeña es detectable (por ejemplo, comprende al menos una fracción detectable). En algunos aspectos, una molécula pequeña es terapéutica).

La expresión "hipermutado somáticamente" incluye la referencia a una secuencia de ácido nucleico o secuencia de aminoácidos codificada por la secuencia de ácido nucleico somático, de un linfocito B que ha sufrido un cambio de clase, en donde la secuencia de ácido nucleico de una región variable de inmunoglobulina (por ejemplo, secuencia de nucleótidos que codifica un dominio variable de cadena ligera o que incluye una secuencia CDR o FR de cadena ligera) en el linfocito B con cambio de clase no es idéntica a la secuencia de ácido nucleico en el linfocito B antes del cambio de clase, tal como, por ejemplo, una diferencia en una secuencia de CDR o marco de ácido nucleico entre un linfocito B que no ha experimentado un cambio de clase y un linfocito B que ha experimentado un cambio de clase. "Mutada somáticamente" incluye referencias a secuencias de ácido nucleico o secuencias de aminoácidos codificadas por ellas de linfocitos B madurados por afinidad que no son idénticas a las secuencias de región variable de inmunoglobulina correspondientes en linfocitos B que no están madurados por afinidad (es decir, secuencias en el genoma de células de línea germinal). La expresión "mutada somáticamente" también incluye referencia a una secuencia de ácido nucleico de región variable de inmunoglobulina de un linfocito B después de la exposición del linfocito B a un epítipo de interés, en donde la secuencia de ácido nucleico difiere de la secuencia de ácido nucleico correspondiente antes de la exposición del linfocito B al epítipo de interés. La expresión "mutada somáticamente" incluye secuencias de inmunoglobulinas que se han generado en un animal, por ejemplo, un ratón que tiene secuencias de ácido nucleico de región variable de inmunoglobulina humanas, en respuesta a un desafío de antígeno, y que resultan de los procesos de selección inherentemente operativos en dicho animal.

El término "sustancialmente" incluye la condición cualitativa de exhibir la extensión o grado total o casi total de una característica o propiedad de interés. Una persona normalmente experta en las técnicas biológicas entenderá que los fenómenos biológicos y químicos rara vez, o nunca, llegan a completarse y/o proceden hasta completarse o logran o evitan un resultado absoluto. Por lo tanto, el término "sustancialmente" se usa en el presente documento para capturar la posible falta de integridad inherente a muchos fenómenos biológicos y químicos.

La expresión "homología sustancial" incluye una comparación entre las secuencias de aminoácidos o de ácido nucleico. Como apreciarán los expertos en la técnica, generalmente se considera que dos secuencias son "sustancialmente homólogas" si contienen restos homólogos en las posiciones correspondientes. Los restos homólogos pueden ser restos idénticos. Alternativamente, los restos homólogos pueden ser restos no idénticos con características estructurales y/o funcionales apropiadamente similares. Por ejemplo, como es bien sabido por los expertos en la técnica, determinados aminoácidos se clasifican normalmente como aminoácidos "hidrófobos" o "hidrófilos", y/o tienen cadenas laterales "polares" o "no polares". La sustitución de un aminoácido por otro del mismo tipo a menudo puede considerarse una sustitución "homóloga". Las clasificaciones de aminoácidos típicas se resumen en las Tablas 1 y 2.

Como es bien conocido en esta técnica, las secuencias de aminoácidos o de ácido nucleico pueden compararse usando cualquiera de una variedad de algoritmos, incluidos los disponibles en programas informáticos comerciales tales como BLASTN para secuencias de nucleótidos y BLASTP, BLACK interrumpido y PSI-BLAST para secuencias de aminoácidos. Dichos programas ejemplares se describen en Altschul, et al., Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, et al., *Methods in Enzymology*; Altschul, et al., "Gapped BLAST y PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis, et al., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley, 1998; y Misener, et al., (eds.), *Bioinformatics Methods and Protocols (Methods in Molecular Biology, Vol. 132)*, Humana Press, 1999. Además de identificar secuencias homólogas, los programas mencionados anteriormente normalmente proporcionan una indicación del grado de homología. En algunos aspectos, dos secuencias se consideran sustancialmente homólogas si al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de sus restos correspondientes son homólogos en un tramo oportuno de restos. En algunos aspectos, el tramo oportuno es una secuencia completa. En algunos aspectos, el tramo oportuno es al menos 9, 10, 11, 12, 13, 14, 15, 16, 17 o más restos. En algunos aspectos, el tramo oportuno incluye restos contiguos a lo largo de una secuencia completa. En algunos aspectos, el tramo oportuno incluye restos discontinuos a lo largo de una secuencia completa. En algunos aspectos, el tramo oportuno es al menos 10, 15, 20, 25, 30, 35, 40, 45, 50 o más restos.

#### TABLA 1

Alanina	Ala	A	no polar	neutro	1,8
Arginina	Arg	R	polar	positivo	-4,5
Asparagina	Asn	N	polar	neutro	-3,5
Acido aspártico	Asp	D	polar	negativo	-3,5
Cisteína	Cys	C	no polar	neutro	2,5
Acido glutámico	Glu	E	polar	negativo	-3,5
Glutamina	Gln	Q	polar	neutro	-3,5
Glicina	Gly	G	no polar	neutro	-0,4
Histidina	His	H	polar	positivo	-3,2
Isoleucina	Ile	I	no polar	neutro	4,5
Leucina	Leu	L	no polar	neutro	3,8
Lisina	Lys	K	polar	positivo	-3,9
Metionina	Met	M	no polar	neutro	1,9
Fenilalanina	Phe	F	no polar	neutro	2,8
Prolina	Pro	P	no polar	neutro	-1,6
Serina	Ser	S	polar	neutro	-0,8
Treonina	Thr	T	polar	neutro	-0,7
Triptófano	Trp	W	no polar	neutro	-0,9
Tirosina	Tyr	Y	polar	neutro	-1,3
Valina	Val	V	no polar	neutro	4,2

TABLA 2

Aminoácidos ambiguos	3 letras	1 letra
Asparagina o ácido aspártico	Asx	B
Glutamina o ácido glutámico	Glx	Z
Leucina o Isoleucina	Xle	J
Aminoácido no especificado o desconocido	Xaa	X

- La expresión "identidad sustancial" incluye una comparación entre las secuencias de aminoácidos o de ácido nucleico.
- 5 Como apreciarán los expertos en la técnica, generalmente se considera que dos secuencias son "sustancialmente idénticas" si contienen restos idénticos en las posiciones correspondientes. Como es bien conocido en esta técnica, las secuencias de aminoácidos o de ácido nucleico pueden compararse usando cualquiera de una variedad de algoritmos, incluidos los disponibles en programas informáticos comerciales tales como BLASTN para secuencias de nucleótidos y BLASTP, BLACK interrumpido y PSI-BLAST para secuencias de aminoácidos. Dichos programas
- 10 ejemplares se describen en Altschul, et al., Basic local alignment search tool, J. Mol. Biol., 215(3): 403-410, 1990; Altschul, et al., Methods in Enzymology; Altschul et al., Nucleic Acids Res. 25:3389-3402, 1997; Baxevanis et al., Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Wiley, 1998; y Misener, et al., (eds.), Bioinformatics Methods and Protocols (Methods in Molecular Biology, Vol. 132), Humana Press, 1999. Además de identificar secuencias idénticas, los programas mencionados anteriormente normalmente proporcionan una indicación
- 15 del grado de identidad. En algunos aspectos, dos secuencias se consideran sustancialmente idénticas si al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de sus restos correspondientes son idénticos en un tramo oportuno de restos. En algunos aspectos, el tramo oportuno es una secuencia completa. En algunos aspectos, el tramo oportuno es al menos 10, 15, 20, 25, 30, 35, 40, 45, 50 o más restos.
- 20
- La expresión "vector de direccionamiento" o "construcción de direccionamiento" incluye una molécula polinucleotídica que comprende una región de direccionamiento. Una región de direccionamiento comprende una secuencia que es idéntica o sustancialmente idéntica a una secuencia en una célula, tejido o animal diana y proporciona la integración
- 25 de la construcción de direccionamiento en una posición dentro del genoma de la célula, tejido o animal mediante recombinación homóloga. También se incluyen regiones de direccionamiento que se dirigen usando sitios de reconocimiento de recombinasa específicos del sitio (por ejemplo, sitios *loxP* o *Frt*). En algunos aspectos, una construcción de direccionamiento de la presente divulgación comprende además una secuencia de ácido nucleico o gen de particular interés, un marcador seleccionable, secuencias de control y/o reguladoras, y otras secuencias de
- 30 ácido nucleico que permiten la recombinación mediada por la adición exógena de proteínas que ayudan en o facilitan la recombinación que implica dichas secuencias. En algunos aspectos, una construcción de direccionamiento de la presente divulgación comprende además un gen de interés total o parcial, en donde el gen de interés es un gen heterólogo que codifica una proteína total o parcial que tiene una función similar a una proteína codificada por una secuencia endógena.
- 35
- La expresión "no reordenada", con referencia a una secuencia de ácido nucleico, incluye secuencias de ácido nucleico que existen en la línea germinal de una célula animal. Generalmente, durante el desarrollo de linfocitos B en animales

no humanos no modificados, el primer reordenamiento de segmentos génicos no reordenados es la unión de segmentos génicos  $D_H$  y  $J_H$  en un locus de cadena pesada, generando un prolinfocito B. Los reordenamientos posteriores incluyen la unión de  $V_H-D_HJ_H$  en un locus de cadena pesada y, si es productivo, el reordenamiento de segmentos génicos de región variable de cadena ligera, por ejemplo, la unión de un segmento génico  $V_L$  con un segmento génico  $J_L$  dentro de un locus de cadena ligera. Un reordenamiento se considera "productivo" si la unión está en marco ("productivo"). El reordenamiento productivo en un alelo puede dar como resultado la exclusión alélica, por ejemplo, el silenciamiento del otro alelo. "No reordenado" también se refiere a segmentos génicos  $V_L$  y  $J_L$  no reordenados capaces de sufrir un reordenamiento productivo para formar un gen de región variable de cadena ligera unido operativamente a un segmento génico de región constante de cadena pesada, dando como resultado dicho enlace operativo un gen que codifica una cadena de inmunoglobulina híbrida, que también puede dar como resultado la exclusión alélica de uno o más alelos endógenos de cadena pesada y/o el reordenamiento de segmentos génicos de región variable de cadena ligera en uno o más loci endógenos de cadena ligera.

La expresión "dominio variable" incluye una secuencia de aminoácidos de una cadena ligera o pesada de inmunoglobulina (modificada según se desee) que comprende las siguientes regiones de aminoácidos, en secuencia desde el extremo N hasta el extremo C (a menos que se indique lo contrario): FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4.

El término "variante" incluye una entidad que muestra una identidad estructural significativa con una entidad de referencia, pero difiere estructuralmente de la entidad de referencia en presencia o nivel de una o más fracciones químicas en comparación con la entidad de referencia. En muchos aspectos, una variante también difiere funcionalmente de su entidad de referencia. En general, si una entidad particular se considera adecuadamente como una "variante" de una entidad de referencia se basa en su grado de identidad estructural con la entidad de referencia. Como apreciarán los expertos en la técnica, cualquier entidad de referencia biológica o química tiene determinados elementos estructurales característicos. Una variante, por definición, es una entidad química distinta que comparte uno o más de dichos elementos estructurales característicos. Para dar solo algunos ejemplos, una molécula pequeña puede tener un elemento estructural central característico (por ejemplo, un núcleo macrociclo) y/o una o más fracciones colgantes características para que una variante de la molécula pequeña sea una que comparta el elemento estructural central y las fracciones colgantes características pero difieren en otras fracciones colgantes y/o en los tipos de enlaces presentes (simple frente a doble, E frente a Z, etc.) dentro del núcleo, un polipéptido puede tener un elemento de secuencia característico compuesto por una pluralidad de aminoácidos que tienen posiciones designadas con respecto a otro en un espacio lineal o tridimensional y/o que contribuyen a una función biológica particular, un ácido nucleico puede tener un elemento de secuencia característico compuesto por una pluralidad de restos de nucleótidos que tienen posiciones designadas con respecto a otro en un espacio lineal o tridimensional. Por ejemplo, un polipéptido variante puede diferir de un polipéptido de referencia como resultado de una o más diferencias en la secuencia de aminoácidos y/o una o más diferencias en fracciones químicas (por ejemplo, hidratos de carbono, lípidos, etc.) unidos covalentemente a la cadena principal del polipéptido. En algunos aspectos, un polipéptido variante muestra una identidad de secuencia global con un polipéptido de referencia que es al menos un 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 % o 99 %. Alternativa o adicionalmente, en algunos aspectos, un polipéptido variante no comparte al menos un elemento de secuencia característico con un polipéptido de referencia. En algunos aspectos, el polipéptido de referencia tiene una o más actividades biológicas. En algunos aspectos, un polipéptido variante comparte una o más de las actividades biológicas del polipéptido de referencia. En algunos aspectos, un polipéptido variante carece de una o más de las actividades biológicas del polipéptido de referencia. En algunos aspectos, un polipéptido variante muestra un nivel reducido de una o más actividades biológicas en comparación con el polipéptido de referencia. En muchos aspectos, un polipéptido de interés se considera una "variante" de un polipéptido precursor o de referencia si el polipéptido de interés tiene una secuencia de aminoácidos que es idéntica a la del precursor pero para un pequeño número de alteraciones de secuencia en posiciones particulares. Normalmente, menos de un 20 %, 15 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 % de los restos en la variante se sustituyen en comparación con el precursor. En algunos aspectos, una variante tiene 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 resto sustituido en comparación con un precursor. A menudo, una variante tiene un número muy pequeño (por ejemplo, menos de 5, 4, 3, 2 o 1) de restos funcionales sustituidos (es decir, restos que participan en una actividad biológica particular). Además, una variante normalmente no tiene más de 5, 4, 3, 2 o 1 adiciones o eliminaciones y, a menudo, no tiene adiciones o eliminaciones, en comparación con el precursor. Por otra parte, cualquier adición o eliminación es normalmente menor que aproximadamente 25, aproximadamente 20, aproximadamente 19, aproximadamente 18, aproximadamente 17, aproximadamente 16, aproximadamente 15, aproximadamente 14, aproximadamente 13, aproximadamente 10, aproximadamente 9, aproximadamente 8, aproximadamente 7, aproximadamente 6, y comúnmente es menor que aproximadamente 5, aproximadamente 4, aproximadamente 3 o aproximadamente 2 restos. En algunos aspectos, el polipéptido precursor o de referencia es uno que se encuentra en la naturaleza. Como apreciarán los expertos en la técnica, una variedad de variantes de un polipéptido particular de interés puede encontrarse comúnmente en la naturaleza, particularmente cuando el polipéptido de interés es un polipéptido agente infeccioso.

El término "vector" incluye una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que está asociado. En algunos aspectos, los vectores son capaces de replicación cromosómica extra y/o expresión de ácidos nucleicos a los que están unidos en una célula hospedadora tal como una célula eucariota y/o procariota. Los vectores que pueden dirigir la expresión de genes unidos de forma operativa se denominan en el presente documento "vectores

de expresión".

El término "tipo silvestre" tiene su significado entendido en la técnica que incluye una entidad que tiene una estructura y/o actividad tal como se encuentra en la naturaleza en un estado o contexto "normal" (en contraste con mutante, enfermo, alterado, etc.). Los expertos en la técnica apreciarán que los genes y polipéptidos de tipo silvestre a menudo existen en múltiples formas diferentes (por ejemplo, alelos).

#### Descripción detallada de determinados aspectos

La presente divulgación proporciona, entre otras cosas, métodos para usar animales no humanos genéticamente modificados que tienen material genético humano que codifica dominios variables de cadena ligera (por ejemplo, regiones  $V_L$ ). En determinados aspectos, dichos animales no humanos son útiles, por ejemplo, para la producción y aislamiento de dominios  $V_L$  humanos, y las regiones determinantes de complementariedad (CDR) comprendidas en dichos dominios  $V_L$  humanos, que se unen a determinantes antigénicos que evaden los formatos de inmunoglobulina tradicionales. Se contempla que dichos animales no humanos proporcionen un nuevo sistema *in vivo* para la generación y maduración por afinidad de dominios  $V_L$  humanos que exhiben características únicas de unión a antígeno. Dichas proteínas de unión a antígeno tienen la capacidad de reconocer antígenos extraños que pueden eludir las inmunoglobulinas naturales. En algunos aspectos, los animales no humanos de la presente divulgación son capaces de generar dominios  $V_L$  humanos afines que se unen al antígeno en comparación con los animales no humanos genéticamente modificados de control; en algunos aspectos, dichos mamíferos no humanos desarrollan y/o tienen una población de linfocitos B que expresan proteínas de unión que se parecen a las inmunoglobulinas en la estructura, pero carecen de cualquier secuencia variable de cadena pesada. En algunos aspectos, las proteínas de unión a antígeno expresadas por dichos animales no humanos se caracterizan por que la porción de unión a antígeno comprende exclusivamente dominios  $V_L$  humanos. En algunos aspectos, los animales no humanos de la presente divulgación comprenden un locus endógeno de cadena pesada de inmunoglobulina que contiene material genético del animal no humano y una especie heteróloga (por ejemplo, un ser humano) y comprenden un locus endógeno de cadena ligera de inmunoglobulina que contiene material genético del animal no humano y una especie heteróloga (por ejemplo, ser humano). En algunos aspectos, los animales no humanos de la presente divulgación comprenden un locus de cadena pesada de inmunoglobulina que incluye segmentos génicos  $V_L$  y  $J_L$  humanos no reordenados y un locus de cadena ligera de inmunoglobulina que incluye segmentos génicos  $V_L$  y  $J_L$  humanos no reordenados. En algunos aspectos, la expresión de las proteínas de unión a antígeno está bajo el control de material genético de inmunoglobulina no humano (por ejemplo, un promotor y/o potenciador de inmunoglobulina no humano).

Diversos aspectos de la divulgación se describen en detalle en las secciones siguientes. El uso de secciones no pretende limitar la divulgación. Cada sección puede aplicarse a cualquier aspecto de la divulgación.

#### ***Proteínas de unión de tipo inmunoglobulina específicas para moléculas pequeñas***

En un aspecto, se describe una proteína  $V_L$  de unión a antígeno que se une específicamente a una molécula pequeña. Los aspectos de la proteína  $V_L$  de unión a antígeno descritos en el presente documento incluyen proteínas  $V_L$  de unión a antígeno que comprenden una cadena híbrida codificada por un gen de inmunoglobulina híbrido que comprende o procede de un segmento génico  $V_L$  (o parte del mismo), preferentemente no reordenado y más preferentemente humano, reordenado con un segmento génico  $J_L$  (o parte del mismo), preferentemente no reordenado y más preferentemente humano, unido operativamente a secuencias de nucleótidos que codifican uno o más dominios constantes de cadena pesada. Tras la reordenación de los segmentos génicos de cadena ligera, se obtiene una secuencia de nucleótidos reordenada que comprende una secuencia que codifica una región variable de cadena ligera fusionada con una secuencia que codifica una región constante de cadena pesada. Esta secuencia codifica una cadena de inmunoglobulina híbrida que tiene un dominio variable de cadena ligera fusionado con un dominio constante de cadena pesada. Por lo tanto, en un aspecto, la inmunoglobulina híbrida consiste esencialmente en, desde el extremo N al extremo C, un dominio  $V_L$  y un dominio  $C_H$ . En un aspecto, el dominio  $C_H$  comprende una región  $C_{H1}$ , una bisagra, una región  $C_{H2}$ , una región  $C_{H3}$  y, opcionalmente, una región  $C_{H4}$ . En otro aspecto, el dominio  $C_{H1}$  carece de un dominio  $C_{H1}$  funcional, por ejemplo, carece de un dominio  $C_{H1}$  en su totalidad o en parte, y además puede carecer de una región bisagra.

En algunos aspectos, la proteína  $V_L$  de unión a antígeno comprende una cadena de inmunoglobulina híbrida que comprende un dominio variable de cadena ligera de inmunoglobulina que se une específicamente a una molécula pequeña, en donde el dominio variable de cadena ligera de inmunoglobulina está unido operativamente a una región constante de cadena pesada. En algunos aspectos, la proteína  $V_L$  de unión a antígeno comprende el primer y el segundo dominio variable de cadena ligera de inmunoglobulina, en donde el primer y el segundo dominio variable de cadena ligera de inmunoglobulina pueden asociarse para formar un punto de unión que se une específicamente a una molécula pequeña. En un aspecto, se proporciona una proteína de unión a antígeno que consiste esencialmente en el primer y segundo dominio variable de cadena ligera de inmunoglobulina que están asociados para formar un punto de unión, en donde la proteína de unión a antígeno se une específicamente a una molécula pequeña.

En un aspecto, el primer y/o el segundo dominio variable de cadena ligera de inmunoglobulina es un dominio variable de cadena ligera de inmunoglobulina humano. En un aspecto, el primer y/o el segundo dominio de cadena ligera de

inmunoglobulina es de un roedor. En un aspecto, el roedor se selecciona entre un ratón o una rata.

En diversos aspectos, las proteínas  $V_L$  de unión a antígeno como se desvela en el presente documento, por ejemplo, las producidas por los animales no humanos genéticamente modificados, por ejemplo, ratones, desvelados en el presente documento, pueden ser en promedio más pequeñas que los anticuerpos convencionales y poseen ventajas asociadas con un tamaño más pequeño. Un tamaño más pequeño se obtiene al menos en parte a través de la ausencia de una secuencia de aminoácidos codificada por una región  $D_H$ , normalmente presente en un dominio  $V_H$ . También se puede obtener un tamaño más pequeño en la formación de una CDR3 que procede, por ejemplo, de una región  $V_K$  and y una región  $J_K$ .

En un aspecto, los dominios variables de cadena ligera se unen a la molécula pequeña con mayor afinidad que un punto de unión de una proteína de unión a antígeno humana que se forma a partir de dominios variables de cadena ligera y pesada de inmunoglobulina humanos.

En un aspecto, el primer y/o el segundo dominio variable de cadena ligera de inmunoglobulina son dominios variables de cadena ligera humanos. En un aspecto, el punto de unión de los dominios variables de cadena ligera se unen a la molécula pequeña con mayor afinidad que un punto de unión de un anticuerpo humano que se forma a partir de dominios variables de cadena ligera y pesada de inmunoglobulina humanos.

En un aspecto, el primer dominio variable de cadena ligera está unido a una primera región constante de cadena pesada de inmunoglobulina. En un aspecto, la primera región constante de cadena pesada de inmunoglobulina es de un animal no humano. En un aspecto, el animal no humano es un roedor. En un aspecto, el roedor se selecciona entre un ratón o una rata. En un aspecto, el animal no humano es un pollo. En un aspecto, la primera región constante de cadena pesada de inmunoglobulina se selecciona de una CH1, una bisagra, una CH2, una CH3, una CH4 y una combinación de las mismas. En un aspecto, la primera región constante de cadena pesada de inmunoglobulina comprende una CH1, una bisagra, una CH2 y una CH3.

En un aspecto, el segundo dominio variable de cadena ligera de inmunoglobulina está unido a una segunda región constante de cadena ligera de inmunoglobulina. En un aspecto, la segunda región constante de cadena ligera de inmunoglobulina es de un animal no humano. En un aspecto, el animal no humano es un roedor. En un aspecto, el roedor se selecciona entre un ratón o una rata. En un aspecto, el animal no humano es un pollo.

En un aspecto, la proteína  $V_L$  de unión a antígeno comprende dos dominios variables de cadena ligera idénticos. En un aspecto, la proteína  $V_L$  de unión a antígeno comprende dos dominios variables de cadena ligera con secuencias heterogéneas.

Una proteína  $V_L$  de unión a antígeno que se une a una molécula pequeña se puede obtener de un animal no humano genéticamente modificado como se desvela en el presente documento o proceder de células y/o ácidos nucleicos aislados de dicho animal después de la inmunización con la molécula pequeña.

***Animales no humanos genéticamente modificados que expresan proteínas  $V_L$***

Se describen animales no humanos que expresan proteínas  $V_L$  de unión a antígeno que comprenden cadenas de inmunoglobulina híbridas que tienen un dominio constante de cadena pesada fusionado con un dominio variable de cadena ligera de inmunoglobulina. Además, se describen múltiples estrategias para modificar genéticamente un animal no humano, por ejemplo, un roedor, que incluye, pero no se limita a, ratas y ratones, para expresar una cadena híbrida como parte de una proteína  $V_L$  de unión a antígeno, en donde la cadena híbrida se codifica o procede de un ácido nucleico que codifica una región  $V_L$  unida operativamente a una secuencia de nucleótidos que codifica una región  $C_H$ . Dichos animales no humanos genéticamente modificados representan una fuente para generar poblaciones de proteínas  $V_L$  de unión a antígeno que tienen la estructura tetramérica de algunos anticuerpos convencionales, pero exhiben una característica de unión única en comparación con los anticuerpos tradicionales.

Los animales no humanos modificados descritos en el presente documento pueden generar proteínas  $V_L$  de unión a antígeno que también comprenden una cadena ligera afín emparejada con una cadena híbrida para producir una proteína  $V_L$  de unión a antígeno que es similar a un anticuerpo, por ejemplo, puede ser tetramérica, pero en donde en vez de una cadena pesada (o par de cadenas pesadas) la proteína  $V_L$  de unión a antígeno comprende una cadena híbrida (o un par de cadenas híbridas) que comprende un dominio  $V_L$ , no un dominio  $V_H$ , fusionado a un dominio  $C_H$ .

En diversos aspectos, los animales no humanos modificados producen proteínas  $V_L$  de unión a antígeno, en donde el dominio  $V_L$  de una cadena híbrida exhibe un grado mejorado de hipermutación somática sobre un dominio  $V_L$  de una cadena ligera. En algunos aspectos, una región  $V_L$  de una cadena híbrida exhibe alrededor de 1,5 veces, 2 veces, 2,5 veces, 3 veces, 3,5 veces, 4 veces, 4,5 veces o 5 veces o más hipermutaciones somáticas que una región  $V_L$  fusionada con una región  $C_L$ . En algunos aspectos, el animal no humano modificado, por ejemplo, ratón, en respuesta a un antígeno, exhibe una población de proteínas de unión a antígeno que comprenden un dominio  $V_L$  de una cadena híbrida, en donde la población de proteínas  $V_L$  de unión a antígeno exhibe un promedio de aproximadamente 1,5 veces, 2 veces, 2,5 veces, 3 veces, 3,5 veces, 4 veces, 4,5 veces, 5 veces o más hipermutaciones somáticas en el

dominio  $V_L$  de la cadena híbrida de lo que se observa en una población de proteínas de unión a antígeno, por ejemplo, un dominio  $V_L$  de una cadena ligera, exhibido por un ratón de tipo silvestre en respuesta al mismo antígeno.

5 En un aspecto, las hipermutaciones somáticas en el dominio  $V_L$  de la cadena híbrida comprenden una o más o dos o más adiciones de N en una CDR3. En diversos aspectos, las proteínas  $V_L$  de unión a antígeno comprenden cadenas híbridas que comprenden dominios variables codificados por secuencias de cadena ligera de inmunoglobulina que comprenden un mayor número de adiciones de N que las observadas en la naturaleza para cadenas ligeras reordenadas desde un locus endógeno de cadena ligera, por ejemplo, los segmentos génicos  $V_L$  y  $J_L$  humanos se reorganizan para formar un gen de región variable reordenado unido operativamente con un gen de región constante de cadena pesada, en donde la región variable de cadena ligera reordenada comprende 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 o más adiciones de N.

15 En un aspecto, se describe un animal no humano, por ejemplo, un ratón, que comprende un locus de cadena híbrida de inmunoglobulina. En un aspecto, el locus de cadena híbrida se crea dentro de un locus endógeno de cadena pesada, en donde uno o más segmentos génicos de región variable de cadena pesada ( $V_H$ ) de inmunoglobulina, segmentos génicos de diversidad de cadena pesada ( $D_H$ ) y segmentos génicos de unión de cadena pesada ( $J_H$ ) en un locus endógeno de cadena pesada de inmunoglobulina de ratón se reemplazan con uno o más segmentos génicos de región variable de cadena ligera ( $V_L$ ) y uno o más segmentos génicos de unión de cadena ligera ( $J_L$ ). En un aspecto, se describe un animal no humano, que comprende un locus de cadena híbrida que reemplaza el locus endógeno de cadena pesada de inmunoglobulina, por ejemplo, se reemplazan todos o sustancialmente todos los segmentos génicos endógenos  $V_H$ ,  $D_H$  y  $J_H$  de uno o ambos loci de cadena pesada con uno o más segmentos génicos  $V_L$  y uno o más segmentos génicos  $J_L$  que forman una secuencia génica  $V_L$  reordenada en un locus endógeno de cadena pesada capaz de recombinarse con un gen  $C_H$  endógeno de ratón para formar un gen reordenado que procede de un segmento génico  $V_L$ , un segmento génico  $J_L$  y un gen  $C_H$  endógeno del ratón.

25 Los animales no humanos también abarcan la humanización de los loci de inmunoglobulina que dan como resultado la expresión de proteínas de unión, por ejemplo, proteínas  $V_L$  de unión a antígeno, que se asemejan a la estructura tetramérica de algunos anticuerpos convencionales pero difieren en las características de unión, y dan como resultado la expresión de dichas proteínas  $V_L$  de unión a antígeno en la superficie de la membrana de las células del animal no humano. En algunos aspectos, los animales no humanos de la presente divulgación son capaces de generar dominios  $V_L$  humanos, en cualquiera o ambas cadenas híbridas y ligeras de la proteína  $V_L$  de unión a antígeno, que se unen al antígeno; en algunos aspectos, dichos mamíferos no humanos desarrollan y/o tienen una población de linfocitos B que expresan proteínas de unión que comprenden dominios variables que no están codificados ni proceden de ninguna secuencia de segmentos génicos  $V_H$ ,  $D_H$  y/o  $J_H$ . En algunos aspectos, las proteínas  $V_L$  de unión a antígeno expresadas por dichos animales no humanos se caracterizan por que la porción de unión a antígeno comprende exclusivamente dominios  $V_L$  humanos. En algunos aspectos, los animales no humanos de la presente divulgación comprenden material genético de un locus endógeno de cadena pesada de inmunoglobulina del animal no humano y una especie heteróloga (por ejemplo, un ser humano) y comprenden material genético de un locus endógeno de cadena ligera de inmunoglobulina del animal no humano y una especie heteróloga (por ejemplo, ser humano).

40 En algunos aspectos, los animales no humanos de la presente divulgación comprenden un locus de cadena híbrida de inmunoglobulina que incluye segmentos génicos  $V_L$  humanos no reordenados y/o segmentos génicos  $J_L$  humanos y, preferentemente, un locus de cadena ligera de inmunoglobulinas que incluye segmentos génicos  $V_L$  humanos no reordenados y/o segmentos génicos  $J_L$  humanos. En algunos aspectos, la expresión de las proteínas  $V_L$  de unión a antígeno está bajo el control de material genético de inmunoglobulina no humano (por ejemplo, un promotor y/o potenciador de inmunoglobulina no humano).

50 En un aspecto, los segmentos  $V_L$  son  $V_L$  humanos. En un aspecto, los segmentos  $J_L$  son  $J_L$  humanos. En un aspecto específico, los segmentos  $V_L$  y  $J_L$  son segmentos  $V_L$  humanos y  $J_L$  humanos.

55 En un aspecto, se reemplazan todos o sustancialmente todos los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  con al menos seis segmentos génicos  $V_K$  humanos y al menos un segmento génico  $J_K$ . En un aspecto, se reemplazan todos o sustancialmente todos los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  con al menos 16 segmentos génicos  $V_K$  humanos ( $V_K$  humanos) y al menos un segmento génico  $J_K$ . En un aspecto, se reemplazan todos o sustancialmente todos los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  con al menos 30 segmentos génicos  $V_K$  humanos y al menos un segmento génico  $J_K$ . En un aspecto, se reemplazan todos o sustancialmente todos los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  con al menos 40 segmentos génicos  $V_K$  humanos y al menos un segmento génico  $J_K$ . En un aspecto, el al menos un segmento génico  $J_K$  comprende dos, tres, cuatro o cinco segmentos génicos  $J_K$  humanos.

60 En un aspecto, los segmentos  $V_L$  son segmentos  $V_K$  humanos. En un aspecto, los segmentos  $V_K$  humanos comprenden 4-1, 5-2, 7-3, 2-4, 1-5 y 1-6. En un aspecto, los segmentos  $V_K$  comprenden 3-7, 1-8, 1-9, 2-10, 3-11, 1-12, 1-13, 2-14, 3-15, 1-16. En un aspecto, los segmentos  $V_K$  humanos comprenden 1-17, 2-18, 2-19, 3-20, 6-21, 1-22, 1-23, 2-24, 3-25, 2-26, 1-27, 2-28, 2-29 y 2-30. En un aspecto, los segmentos  $V_K$  humanos comprenden 3-31, 1-32, 1-33, 3-34, 1-35, 2-36, 1-37, 2-38, 1-39 y 2-40.

65 En un aspecto, los segmentos  $V_L$  son segmentos  $V_K$  humanos y comprenden 4-1, 5-2, 7-3, 2-4, 1-5, 1-6, 3-7, 1-8, 1-9,

2-10, 3-11, 1-12, 1-13, 2-14, 3-15 y 1-16. En un aspecto, los segmentos  $V_K$  comprenden además 1-17, 2-18, 2-19, 3-20, 6-21, 1-22, 1-23, 2-24, 3-25, 2-26, 1-27, 2-28, 2-29 y 2-30. En un aspecto, los segmentos  $V_K$  comprenden además 3-31, 1-32, 1-33, 3-34, 1-35, 2-36, 1-37, 2-38, 1-39 y 2-40.

5 En un aspecto, los segmentos  $V_L$  son segmentos  $V_\lambda$  humanos, y comprenden un fragmento del grupo A del locus de cadena ligera  $\lambda$  humano. En un aspecto específico, el fragmento del grupo A del locus de cadena ligera  $\lambda$  humano se extiende desde hV $\lambda$ 3-27 hasta hV $\lambda$ 3-1.

10 En un aspecto, los segmentos  $V_L$  comprenden un fragmento del grupo B del locus de cadena ligera  $\lambda$  humano. En un aspecto específico, el fragmento del grupo B del locus de cadena ligera  $\lambda$  humano se extiende desde hV $\lambda$ 5-52 hasta hV $\lambda$ 1-40.

15 En un aspecto, los segmentos  $V_L$  comprenden una secuencia de región variable de cadena ligera  $\lambda$  humana que comprende un fragmento genómico del grupo A y un fragmento genómico del grupo B. En un aspecto, la secuencia de región variable de cadena ligera  $\lambda$  humana comprende al menos un segmento génico del grupo A y al menos un segmento génico del grupo B.

20 En un aspecto, los segmentos  $V_L$  comprenden al menos un segmento génico del grupo B y al menos un segmento génico del grupo C.

25 En un aspecto, los segmentos  $V_L$  comprenden hV $\lambda$  3-1, 4-3, 2-8, 3-9, 3-10, 2-11 y 3-12. En un aspecto específico, los segmentos  $V_L$  comprenden una secuencia contigua del locus de cadena ligera  $\lambda$  humano que se extiende desde V $\lambda$ 3-12 a V $\lambda$ 3-1. En un aspecto, la secuencia contigua comprende al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 hV $\lambda$ s. En un aspecto específico, las hV $\lambda$  incluyen 3-1, 4-3, 2-8, 3-9, 3-10, 2-11 y 3-12. En un aspecto específico, la hV $\lambda$ s comprende una secuencia contigua del locus  $\lambda$  humano que se extiende desde V $\lambda$ 3-12 a V $\lambda$ 3-1.

30 En un aspecto, la hV $\lambda$ s comprende de 13 a 28 o más hV $\lambda$ s. En un aspecto específico, las hV $\lambda$  incluyen 2-14, 3-16, 2-18, 3-19, 3-21, 3-22, 2-23, 3-25, y 3-27. En un aspecto específico, las hV $\lambda$ s comprenden una secuencia contigua del locus  $\lambda$  humano que se extiende desde V $\lambda$ 3-27 a V $\lambda$ 3-1.

35 En un aspecto, los segmentos  $V_L$  comprenden de 29 a 40 hV $\lambda$ s. En un aspecto específico, los segmentos  $V_L$  comprenden una secuencia contigua del locus  $\lambda$  humano que se extiende desde V $\lambda$ 3-29 a V $\lambda$ 3-1, y una secuencia contigua del locus  $\lambda$  humano que se extiende desde V $\lambda$ 5-52 a V $\lambda$ 1-40. En un aspecto específico, toda o sustancialmente toda la secuencia entre hV $\lambda$ 1-40 y hV $\lambda$ 3-29 en el ratón genéticamente modificado consiste esencialmente en una secuencia  $\lambda$  humana de aproximadamente 959 pb que se encuentra en la naturaleza (por ejemplo, en la población humana) cadena abajo del segmento génico hV $\lambda$ 1-40 (cadena abajo de la porción no traducida en 3'), un sitio de enzimas de restricción (por ejemplo, PI-SceI), seguido de una secuencia  $\lambda$  humana de aproximadamente 3.431 pb cadena arriba del segmento génico hV $\lambda$ 3-29 que se encuentra en la naturaleza.

40 En un aspecto, el  $J_K$  es humano y se selecciona del grupo que consiste en  $J_{K1}$ ,  $J_{K2}$ ,  $J_{K3}$ ,  $J_{K4}$ ,  $J_{K5}$  y una combinación de los mismos. En un aspecto específico, el  $J_K$  comprende de  $J_{K1}$  a  $J_{K5}$ .

45 En un aspecto, los segmentos  $V_L$  son segmentos  $V_\lambda$  humanos, y el segmento génico  $J_K$  comprende una RSS que tiene un espaciador de 12 meros, en donde la RSS está yuxtapuesta en el extremo cadena arriba del segmento génico  $J_K$ . En un aspecto, los segmentos génicos  $V_L$  son  $V_\lambda$  humanos y el locus  $V_{LH}$  comprende dos o más segmentos génicos  $J_K$ , cada uno con una RSS que tiene un espaciador de 12 meros en donde la RSS se yuxtapone en el extremo cadena arriba de cada segmento génico  $J_K$ .

50 En un aspecto específico, los segmentos  $V_L$  comprenden segmentos génicos  $k$  humanos contiguos que abarcan el locus  $k$  humano de  $V_{K4-1}$  a  $V_{K2-40}$ , y los segmentos  $J_L$  comprenden segmentos génicos contiguos que abarcan el locus  $k$  humano de  $J_{K1}$  a  $J_{K5}$ .

55 En un aspecto, donde los segmentos  $V_L$  son segmentos  $V_\lambda$  y no hay ningún segmento  $D_H$  entre los segmentos  $V_L$  y los segmentos  $J$ , los segmentos  $V_L$  están flanqueados cadena abajo (es decir, yuxtapuestos en el lado cadena abajo) con RSS de 23 meros y los segmentos  $J_K$  si están presentes o  $J_L$  si están presentes están flanqueados cadena arriba (es decir, yuxtapuestos en el lado cadena arriba) con RSS de 12 meros.

60 En un aspecto, donde los segmentos génicos  $V$  son segmentos génicos  $V_K$  y no hay un segmento génico  $D_H$  presente entre los segmentos génicos  $V$  y los segmentos génicos  $J$ , los segmentos génicos  $V_K$  están yuxtapuestos en el lado cadena abajo con una RSS de 12 meros, y los segmentos  $J_K$  si están presentes o los segmentos  $J_L$  si están presentes están yuxtapuestos en el lado cadena arriba con una RSS de 23 meros.

65 En un aspecto, se proporciona un célula, que comprende un locus de inmunoglobulina modificado como se describe en el presente documento. En un aspecto, la célula se selecciona de una célula totipotente, una célula pluripotente, una célula madre pluripotente inducida (iPS, de sus siglas en inglés) y una célula ME. En un aspecto, la célula MS es una línea ME F1 (F1H4; Valenzuela et al. 2007, supra) procedente de embriones heterocigotos 129S6/SvEvTac y

C57BL/6NTac que además contenían un reemplazo in situ de los segmentos génicos de cadena ligera k de ratón con segmentos génicos de cadena ligera k humanos (por ejemplo, véanse las Patentes de EE.UU. N.º 6.596.541 y 8.642.835). En un aspecto, la modificación genética se lleva a cabo en una línea celular ME híbrida cuyo genoma comprende 50 % de BALB/c[Tac], 25 % de C57BL/6N[Tac] y 25 % de 129S4/SvJae(V17).

5 En un aspecto específico, la célula una célula de ratón, por ejemplo, una célula ME de ratón. En un aspecto, la célula es homocigota para el locus de inmunoglobulina modificado. En un aspecto, la célula una célula de rata, por ejemplo, una célula ME de rata (véase el documento US-2014-0310828-A1).

## 10 **Moléculas pequeñas**

En un aspecto, la molécula pequeña es un hapteno y la molécula pequeña está unida a un vehículo. En un aspecto, el vehículo comprende hemocianina de lapa californiana (KLH, de sus siglas en inglés), hemocianina de Concholepas concholepas (CCH, de sus siglas en inglés), seroalbúmina bovina (BSA, de sus siglas en inglés), una seroalbúmina bovina cationizada (cBSA, de sus siglas en inglés) u ovoalbúmina. En un aspecto, la molécula pequeña es un compuesto orgánico cuyo peso molecular es inferior a 6 kDa.

15 En algunos aspectos, la molécula pequeña es un hapteno en el sentido de que provoca una respuesta inmunitaria solo cuando está unida a un vehículo grande pero no produce una respuesta inmunitaria útil o significativa cuando está en condiciones comparables que carecen del vehículo u otro adyuvante, por ejemplo, empleada como un inmunógeno solo en ausencia de un adyuvante. Los ejemplos de haptenos incluyen, pero sin limitación, antibióticos, pesticidas, herbicidas, insecticidas, fármacos, vitaminas, esteroides, hormonas, toxinas, explosivos y colorantes (véase, por ejemplo, Gunther, S. et al., SuperHapten: a comprehensive database for small immunogenic compounds, Nucleic Acids Res., 2007, D906-910). También se puede encontrar una lista completa de haptenos y conjugados hapteno-vehículo correspondiente en la base de datos Hapten (Singh, M. et al., Bioinformatics, 2006, 22:253-255), a la que se puede acceder a través de Internet en la red mundial (www) en la URL "imtech.res.in/raghava/haptendb".

20 En algunos aspectos, el vehículo es una macromolécula que se une a un hapteno y le permite inducir una respuesta inmunitaria. En algunos aspectos, el vehículo es una proteína secretora o una proteína de la superficie celular. En algunos aspectos, el portador es un polímero. En algunos aspectos, el vehículo es hemocianina de lapa californiana (KLH, de sus siglas en inglés). En algunos aspectos, el vehículo es la preparación purificada de hemocianina de Concholepas concholepas (CCH, de sus siglas en inglés). En algunos aspectos, el vehículo es seroalbúmina bovina (BSA, de sus siglas en inglés). En algunos aspectos, el vehículo es una BSA catiónica (cBSA, de sus siglas en inglés) que se prepara mediante la modificación de la BSA nativa con un exceso de etilendiamina, que limita esencialmente los grupos carboxilo cargados negativamente con aminos primarias cargadas positivamente. En algunos aspectos, el vehículo es ovoalbúmina.

30 En algunos aspectos, la molécula pequeña es un esteroide natural. En algunos aspectos, la molécula pequeña es un esteroide caracterizado por una estructura molecular de 17 átomos de carbono dispuestos en cuatro anillos. Los ejemplos del esteroide como se describe en el presente documento incluyen, pero sin limitación, hormonas y alcaloides.

45 En algunos aspectos, el esteroide es un esteroide cardiotónico (CTS, de sus siglas en inglés). En algunos aspectos, el CTS es un inhibidor de Na<sup>+</sup>/K<sup>+</sup>-ATPasa. Los ejemplos de CTS incluyen, pero sin limitación, cardenólidos (ouabaína endógena), bufadienólidos, bufalina, marinobufagenina (MBG) y telocinobufagin. En algunos aspectos, el hapteno es marinobufagenina (MBG) y el vehículo es la seroalbúmina bovina. En algunos aspectos, el esteroide es cortisol.

50 En algunos aspectos, la molécula pequeña es un veneno o sustancia venenosa, incluyendo, pero sin limitación, paratión, malatión, tetraetilpirofosfato (TEPP, de sus siglas en inglés), 4,6-dinitro-o-cresol (DNOC), metacido, Demeton (systex), clordano, toxafeno, aldrina, hexacloruro de benceno, lindano, dieldrina, rotenona, pestex, diclorodifeniltricloroetano (DDT), un compuesto de selenio (silocida), fosfuro de zinc (Zn3P2), un compuesto de estricnina, warfarina y trióxido de arsénico.

55 En algunos aspectos, la molécula pequeña es un fármaco psicoactivo o sustancia psicotrópica que atraviesa la barrera hematoencefálica y actúa sobre el sistema nervioso central donde afecta la función cerebral, lo que da como resultado cambios en la percepción, el estado de ánimo, la conciencia, la cognición y el comportamiento. En algunos aspectos, la molécula pequeña es un estimulante, incluyendo, pero sin limitación, cafeína, nicotina, anfetaminas y cocaína. En algunos aspectos, la molécula pequeña es un alcaloide opioide, incluyendo, pero sin limitación, morfina, codeína, heroína, fentanilo, metadona y oxycodona. En algunos aspectos, la molécula pequeña es un fármaco psicodélico que distorsiona las percepciones sensoriales, incluidas la vista y el sonido. Los ejemplos del fármaco psicodélico incluyen, pero sin limitación, mesacalina, psilocibina, dimetiltriptamina (DMT), dietilamida del ácido lisérgico (LSD, de sus siglas en inglés), dimetoximetilanfetamina (DOM o "STP"), metilendioximetanfetamina (MDMA o "éxtasis").

65 En algunos aspectos, la molécula pequeña es un neurotransmisor, incluyendo, pero sin limitación, acetilcolina, noradrenalina, epinefrina, dopamina, serotonina, glutamato, glicina y ácido gamma-aminobutírico (GABA, de sus siglas en inglés).



En algunos aspectos, la molécula pequeña incluye, aunque no de forma limitativa, forskolina, solamarigina, crocina, compuestos de marihuana, alcaloides de opio, ginsenósidos, berberina, senósidos, peoniflorina, glicirricina, ácido ginkgólico, alcaloide de aconitina y baicalina.

5

**Construcciones de ácido nucleico, células y métodos para hacer las mismas**

En un aspecto, se describe un ácido nucleico que codifica un dominio variable de un dominio  $V_L$  de unión que se une específicamente a una molécula pequeña y una célula que expresa el ácido nucleico.

10

En un aspecto, se describe el uso de una secuencia de ácido nucleico de un ratón como se describe en el presente documento para producir una línea celular para la fabricación de un tratamiento humano. En un aspecto, el tratamiento humano es una proteína de unión que comprende una secuencia variable de cadena ligera humana (por ejemplo, procedente de un segmento  $V_\lambda$  humano o  $V_\kappa$  humano) fusionada con una secuencia constante de cadena pesada humana. En un aspecto, el tratamiento humano comprende un primer polipéptido que es una cadena ligera  $\lambda$  o  $\kappa$  humana de inmunoglobulina, y un segundo polipéptido que comprende una secuencia variable  $V_\lambda$  humana o  $V_\kappa$  humana fusionada con una secuencia constante de cadena pesada humana.

15

En un aspecto, se describe un sistema de expresión, que comprende una célula de mamífero que comprende un ácido nucleico que codifica un polipéptido que comprende un dominio  $V_L$  humano mutado somáticamente fusionado con un dominio  $C_H$  humano.

20

En un aspecto, el sistema de expresión comprende además una secuencia de nucleótidos que codifica un dominio  $V_L$  de inmunoglobulina fusionado con un dominio  $C_L$  humano, en donde el dominio  $V_L$  fusionado con el dominio  $C_L$  humano es una cadena ligera afin con el dominio  $V_L$  fusionado con el dominio  $C_H$  humano.

25

En un aspecto, la célula adecuada se selecciona de un linfocito B, un hibridoma, un cuadroma, una célula CHO, una célula COS, una célula 293, una célula HeLa y una célula retiniana humana que expresa una secuencia de ácido nucleico vírica (por ejemplo, una célula PERC.6™).

30

En un aspecto, se describe un método para fabricar una proteína de unión, aislar una célula o ácido nucleico de un animal no humano como se describe en el presente documento, en donde la célula o el ácido nucleico comprende o codifica una proteína  $V_L$  de unión que se une a una molécula pequeña. En algunos aspectos, el método comprende además clonar la secuencia de nucleótidos que codifica la secuencia de región  $V_L$  en marco con un gen que codifica una región  $C_H$  humana para formar una secuencia de proteína de unión humana, que expresa la secuencia de proteína de unión humana en una célula adecuada.

35

En un aspecto, el no humano se ha inmunizado con una molécula pequeña o una molécula pequeña unida a un vehículo, y la región  $V_L$  fusionada con la región  $C_H$  se une específicamente (por ejemplo, con una  $K_D$  en el intervalo micromolar, nanomolar o picomolar) a un epítipo de la molécula pequeña. En un aspecto, la secuencia de nucleótidos que codifica la región  $V_L$  fusionada con la región  $C_H$  está mutada somáticamente en el ratón.

40

En un aspecto, se describe un método para fabricar una proteína de unión a antígeno que se une a una molécula pequeña, comprendiendo el método (a) inmunizar un animal no humano con una molécula pequeña o la molécula pequeña unida a un vehículo, en donde el animal no humano comprende en su línea germinal (i) segmentos génicos variables ( $V_L$ ) de cadena ligera y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada no humana, y (ii) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera no humana; (b) permitir que el animal no humano desarrolle una respuesta inmunitaria a la molécula pequeña o la molécula pequeña unida al vehículo; (c) aislar una célula (por ejemplo, un linfocito) del animal no humano inmunizado, en donde la célula comprende la primera y segunda secuencia de ácido nucleico de región variable de inmunoglobulina que codifican el primer y segundo dominio variable de cadena ligera de inmunoglobulina; (d) identificar las primera y segunda secuencias de ácido nucleico de región variable de cadena ligera de inmunoglobulina que codifican los primer y segundo dominios variables de cadena ligera de inmunoglobulina que, cuando se combinan, se unen específicamente a la molécula pequeña o la molécula pequeña unida al vehículo; y, (e) expresar las secuencias de ácido nucleico de (d) en un sistema de expresión adecuado para expresar la proteína de unión a antígeno para formar una proteína de unión a antígeno que comprende un dímero del primer y segundo dominio variable de cadena ligera que se une a la molécula pequeña.

45

50

55

60

En algunos aspectos, las células (tales como los linfocitos B) se recuperan del animal (por ejemplo, del bazo o los ganglios linfáticos). Las células pueden fusionarse con una línea celular de mieloma para preparar líneas celulares de hibridoma inmortales, y dichas líneas celulares de hibridoma se criban y seleccionan para identificar líneas celulares de hibridoma que producen anticuerpos que contienen cadenas pesadas híbridas específicas para el antígeno utilizado para la inmunización.

65

En un aspecto, la inmunización comprende cebar al ratón con la molécula pequeña o la molécula pequeña unida a un vehículo, permitiendo que el animal no humano descanse durante un período de tiempo y volviendo a inmunizar al animal con la molécula pequeña o una molécula pequeña unida a un vehículo. En algunos aspectos, el período de tiempo es de unos pocos días, al menos una semana, al menos dos semanas, al menos tres semanas, al menos cuatro semanas o al menos un mes.

En un aspecto, se proporciona una región variable (VR) de inmunoglobulina (por ejemplo, que comprende una secuencia de  $V_L$  humana fusionada con una  $J_L$  humana) hecha en un ratón como se describe en el presente documento. En un aspecto específico, la VR de inmunoglobulina procede de un segmento génico humano de línea germinal seleccionado de un segmento  $V_K$  y un segmento  $V_\lambda$ , en donde la VR está codificada por una secuencia reordenada del ratón en donde la secuencia reordenada está hipermutada somáticamente. En un aspecto, la secuencia reordenada comprende de 1 a 5 hipermutaciones somáticas. En un aspecto, la secuencia reordenada comprende al menos 6, 7, 8, 9 o 10 hipermutaciones somáticas. En un aspecto, la secuencia reordenada comprende más de 10 hipermutaciones somáticas. En un aspecto, la secuencia reordenada se fusiona con una o más secuencias de región constante de cadena pesada humanas o de ratón (por ejemplo, seleccionada de una  $C_{H1}$ , bisagra,  $C_{H2}$ ,  $C_{H3}$  humanas o de ratón, y una combinación de las mismas).

En un aspecto, se describe una secuencia de aminoácidos de dominio variable de inmunoglobulina de una proteína de unión hecha en un ratón como se describe en el presente documento. En un aspecto, la VR se fusiona con una o más secuencias de región constante de cadena pesada humanas o de ratón (por ejemplo, seleccionada de una  $C_{H1}$ , bisagra,  $C_{H2}$ ,  $C_{H3}$  humanas o de ratón, y una combinación de las mismas).

En un aspecto, se describe un dominio variable de cadena ligera codificado por una secuencia de ácido nucleico procedente de un ratón como se describe en el presente documento.

En un aspecto, se describe una proteína de unión o un fragmento de unión a antígeno de la misma (por ejemplo, Fab,  $F(ab)_2$ , scFv) fabricada en un ratón como se describe en el presente documento, o procedente de una secuencia hecha en un ratón como se describe en el presente documento.

### 30 **Proteínas de unión biespecíficas**

Se describen proteínas de unión de tipo inmunoglobulina que comprenden una región constante de cadena pesada de inmunoglobulina fusionada con un dominio variable de cadena ligera de inmunoglobulina, así como proteínas de unión que tienen un dominio variable de cadena ligera de inmunoglobulina fusionado a un dominio constante de cadena ligera y un dominio variable de cadena ligera de inmunoglobulina fusionado a un dominio constante de cadena pesada. También se describen células que expresan dichas proteínas de unión, ratones que las producen y métodos y composiciones relacionados.

Las proteínas de unión descritas en el presente documento, y las secuencias de nucleótidos que las codifican, se pueden usar para hacer proteínas de unión multiespecíficas, por ejemplo, proteínas de unión biespecíficas. En este aspecto, un primer polipéptido que consiste esencialmente en un primer dominio  $V_L$  fusionado con una región  $C_H$  puede asociarse con un segundo polipéptido que consiste esencialmente en un segundo dominio  $V_L$  fusionado con una región  $C_H$ . Cuando el primer dominio  $V_L$  y el segundo dominio  $V_L$  se unen específicamente a un epítipo diferente, se puede hacer una molécula de unión biespecífica utilizando los dos dominios  $V_L$ . La región  $C_H$  puede ser igual o diferente. En un aspecto, por ejemplo, una de las regiones  $C_H$  se puede modificar para eliminar un determinante de unión a la proteína A, mientras que la otra región constante de cadena pesada no se modifica de esta manera (véase la Patente de EE.UU. N.º 8.586.713 B2). Esta disposición particular simplifica el aislamiento de la proteína de unión biespecífica de, por ejemplo, una mezcla de homodímeros (por ejemplo, homodímeros del primer o segundo polipéptido).

En un aspecto, se describen construcciones de ácidos nucleicos, células, embriones, ratones y métodos para fabricar proteínas que comprenden una o más secuencias de inmunoglobulina de región variable de cadena ligera  $\kappa$  y/o  $\lambda$  y una secuencia de región constante de cadena pesada de inmunoglobulina, que incluye proteínas que comprenden un dominio variable de cadena ligera  $\lambda$  o  $\kappa$  humano y una secuencia de región constante de cadena pesada humana o de ratón.

En un aspecto, se describen proteínas de unión que comprenden dominios variables de inmunoglobulina que proceden de dominios variables de inmunoglobulina de cadena ligera (*es decir*, kappa ( $\kappa$ ) y/o lambda ( $\lambda$ )), pero no de dominios variables de inmunoglobulina de cadena pesada de longitud completa. También se describen métodos y composiciones para fabricar proteínas de unión, incluidos ratones genéticamente modificados.

En un aspecto, los métodos y composiciones descritos en el presente documento se usan para fabricar proteínas de unión biespecíficas. En este aspecto, una primera  $V_L$  que se fusiona a una región  $C_H$  y una segunda  $V_L$  que se fusiona a una región  $C_H$  se clonan cada una independientemente en marco con una secuencia de IgG humana del mismo isotipo (por ejemplo, una IgG1, IgG2, IgG3, IgG4 humanas). La primera  $V_L$  se une específicamente a un primer epítipo, y la segunda  $V_L$  se une específicamente a un segundo epítipo. El primer y segundo epítopos pueden estar en

diferentes antígenos, o en el mismo antígeno.

5 En un aspecto, el isotipo IgG de la región C<sub>H</sub> fusionada a la primera V<sub>L</sub> y el isotipo IgG de la región C<sub>H</sub> fusionada a la segunda V<sub>L</sub> son el mismo isotipo, pero difieren en que un isotipo IgG comprende al menos una sustitución de aminoácidos. En un aspecto, la al menos una sustitución de aminoácidos hace que la cadena pesada que lleva la sustitución sea incapaz o sustancialmente incapaz de unirse a la proteína A en comparación con la cadena pesada que carece de la sustitución.

10 En un aspecto, la primera región C<sub>H</sub> comprende un primer dominio C<sub>H3</sub> de una IgG humana seleccionada de IgG1, IgG2 e IgG4; y la segunda región C<sub>H</sub> comprende un segundo dominio C<sub>H3</sub> de una IgG humana seleccionada de IgG1, IgG2 e IgG4, en donde el segundo dominio C<sub>H3</sub> comprende una modificación que reduce o elimina la unión del segundo dominio C<sub>H3</sub> a la proteína A (véase la Patente de EE.UU. 8.586.713 B2).

15 En un aspecto, el segundo dominio C<sub>H3</sub> comprende una modificación 435R, numerada de acuerdo con el sistema de numeración de la UE. En otro aspecto, el segundo dominio C<sub>H3</sub> comprende además una modificación 436F, numerada de acuerdo con el sistema de numeración de la UE.

20 En un aspecto, el segundo dominio C<sub>H3</sub> es el de una IgG1 humana que comprende una modificación seleccionada del grupo que consiste en D356E, L358M, N384S, K392N, V397M e V422I, numerada de acuerdo con el sistema de numeración de la UE.

En un aspecto, el segundo dominio C<sub>H3</sub> es el de una IgG2 humana que comprende una modificación seleccionada del grupo que consiste en N384S, K392N e V422I, numerada de acuerdo con el sistema de numeración de la UE.

25 En un aspecto, el segundo dominio C<sub>H3</sub> es el de una IgG4 humana que comprende una modificación seleccionada del grupo que consiste en Q355R, N384S, K392N, V397M, R409K, E419Q e V422I, numerada de acuerdo con el sistema de numeración de la UE.

30 En un aspecto, la proteína de unión comprende regiones C<sub>H</sub> que tienen una o más modificaciones como se menciona en el presente documento, en donde la región constante de la proteína de unión es no inmunogénica o sustancialmente no inmunogénica en un ser humano. En un aspecto específico, las regiones C<sub>H</sub> comprenden secuencias de aminoácidos que no presentan un epítipo inmunogénico en un ser humano. En otro aspecto específico, la proteína de unión comprende una región C<sub>H</sub> que no se encuentra en una cadena pesada humana de tipo silvestre, y la región C<sub>H</sub> no comprende una secuencia que genera un epítipo de linfocitos T.

35 En un aspecto, los dominios Fc pueden modificarse para que tengan una unión al receptor Fc alterada, lo que a su vez afecta la función efectora. Una región constante de cadena pesada (C<sub>H</sub>) genomanipulada, que incluye el dominio Fc, puede ser quimérica. Como tal, una región C<sub>H</sub> quimérica combina dominios C<sub>H</sub> procedentes de más de un isotipo de inmunoglobulina. Por ejemplo, una región C<sub>H</sub> quimérica comprende parte o la totalidad de un dominio C<sub>H2</sub> procedente de una molécula IgG1 humana, IgG2 humana o IgG4 humana, combinado con parte o la totalidad de un dominio C<sub>H3</sub> procedente de una molécula IgG1 humana, IgG2 humana o IgG4 humana. Una región C<sub>H</sub> quimérica también puede contener una región bisagra quimérica. Por ejemplo, una bisagra quimérica puede comprender una secuencia de aminoácidos de "bisagra superior" (restos de aminoácidos de las posiciones 216 a 227 de acuerdo con la numeración de la UE) procedente de una región de bisagra de IgG1 humana, IgG2 humana o IgG4 humana, combinada con una secuencia de "bisagra inferior" (restos de aminoácidos de las posiciones 228 a 236 de acuerdo con la numeración de la UE) procedente de una región bisagra de IgG1 humana, IgG2 humana o IgG4 humana. En un aspecto, la región bisagra quimérica comprende restos de aminoácidos procedentes de una bisagra superior de IgG1 humana o IgG4 humana y restos de aminoácidos procedentes de una bisagra inferior de IgG2 humana.

50 Para determinadas terapias, el dominio Fc puede genomanipularse para activar todas, algunas o ninguna de las funciones efectoras normales de Fc, sin afectar las propiedades farmacocinéticas deseadas de la proteína que contiene Fc (por ejemplo, de anticuerpo). Para ejemplos de proteínas que comprenden regiones C<sub>H</sub> quiméricas y que tienen funciones efectoras alteradas, véase la Solicitud de EE.UU. N.º 14/170.166, presentada el 31 de enero de 2014.

#### 55 ***Elaboración de perfiles de características de unión, agrupación y metodologías relacionadas***

60 En el presente documento se desvela el hallazgo inesperado de que una proteína V<sub>L</sub> de unión a antígeno, particularmente si se genera en animales no humanos que comprenden un gen de inmunoglobulina híbrido como se describe en el presente documento, puede exhibir una o más características de unión únicas o distintas cuando se une específicamente al antígeno, es decir, una característica de unión no exhibida por anticuerpos típicos o convencionales que se unen específicamente al mismo antígeno. La identificación y/o el aislamiento de dichas proteínas V<sub>L</sub> de unión a antígeno incluyen métodos para evaluar las características de unión de dichas proteínas V<sub>L</sub> de unión a antígeno específicas de antígeno a un antígeno, y también puede comprender comparar esas características de unión con las características de unión de anticuerpos típicos o convencionales que se unen específicamente al mismo antígeno. Algunos aspectos comprenden además aislar una secuencia de ácido nucleico que codifica una proteína V<sub>L</sub> de unión a antígeno que exhibe una o más características de unión distintas y,

opcionalmente, expresar la secuencia de ácido nucleico.

Como descripción general, los métodos para elaborar un perfil de las características de unión de una proteína de unión a antígeno comprenden (a) poner en contacto una proteína de unión específica de antígeno con el antígeno (incluidos fragmentos del mismo y/o fragmentos modificados del mismo) en condiciones que permitan la unión, preferentemente unión específica y (b) detectar el complejo proteína de unión-antígeno formado entre el antígeno (o fragmentos del mismo y/o fragmentos modificados del mismo) y la proteína de unión, si los hay. Una "característica de unión" como se usa en el presente documento se refiere a una cualquiera de las propiedades medibles bien conocidas, incluyendo, pero sin limitación, sensibilidad, especificidad, avidéz, afinidad, etc. Un experto en la técnica reconocerá que estas características de unión generales pueden ser como resultado una combinación de características de unión específicas, por ejemplo, especificidad de epítipo, constante de asociación, constante de disociación, constante de equilibrio, etc. Un perfil de unión comprende una cualquiera o más de dichas características de unión.

"Unión específica", "de unión específica", "que se une específicamente", "específica de antígeno" o similar se refiere a una proteína de unión a antígeno que forma un complejo con un antígeno que es relativamente estable en condiciones fisiológicas. La unión específica se caracteriza por una alta afinidad y una capacidad de baja a moderada, a diferencia de la unión no específica, que generalmente tiene una baja afinidad con una capacidad de moderada a alta. Normalmente, la unión se considera específica cuando la constante de asociación  $K_A$  es mayor que  $10^6 \text{ M}^{-1}$ . Si es necesario, la unión no específica se puede reducir sin afectar sustancialmente la unión específica mediante la variación de las condiciones de unión. Las condiciones de unión apropiadas, tal como la concentración de la proteína de unión a antígeno, la fuerza iónica de la solución, la temperatura, el tiempo permitido para la unión, la concentración de un agente bloqueante (por ejemplo, seroalbúmina, caseína de la leche), etc., pueden optimizarse por un experto en la técnica utilizando técnicas rutinarias.

Los métodos para elaborar perfiles de grandes cantidades de proteínas de unión a antígeno dirigidas contra un antígeno son bien conocidos en la técnica e incluyen, pero sin limitación, ensayos de bloqueo cruzado rutinarios, mapeo de epítomos, mutantes de barrido de alanina, transferencias de péptidos (Reineke (2004) *Methods Mol Biol* 248:443-63), análisis de escisión de péptidos, escisión de epítomos y extracción de epítomos y modificación química de antígenos (Tomer (2000) *Protein Science*: 9:487-496). Generalmente, estos métodos pueden incluir la inmovilización de un antígeno (o un fragmento, incluido un fragmento modificado, del mismo) en una superficie.

Generalmente, los soportes sólidos o semisólidos adecuados para la inmovilización, unión y/o enlace a un antígeno o fragmento del mismo (y modificaciones para hacer soportes sólidos adecuados para inmovilizar anticuerpos) son bien conocidos en la técnica. Los ejemplos no limitantes de un soporte sólido incluyen una matriz de chips de biosensor, una perla (por ejemplo, perlas de poliestireno, perlas magnetizadas), una placa de micropocillos, etc. Por lo tanto, por ejemplo, los nanocristales de núcleo-cubierta CdSe-CdS encerrados en una cubierta de sílice pueden proceder fácilmente para acoplarse a un antígeno o fragmento del mismo (Bruchez et al. (1998) *Science* 281: 2013-2016). Del mismo modo, los puntos cuánticos altamente fluorescentes (seleniuro de cadmio con cubierta de sulfuro de zinc) se han acoplado covalentemente a biomoléculas para su uso en la detección biológica ultrasensible (Warren y Nie (1998) *Science* 281: 2016-2018). Las perlas marcadas con fluorescencia están disponibles comercialmente en Luminex and Quantum Dot. Además, almohadillas, películas, nanopocillos o canales de microfluidos también pueden servir como un soporte sólido.

En algunos aspectos, el antígeno o fragmento del mismo (incluido un fragmento modificado del mismo) puede inmovilizarse, unirse o enlazarse sobre una superficie sólida o semisólida, tal como almohadillas de difluoruro de polivinilideno, nitrocelulosa, agarosa y/o de gel de poli(acrilamida). También se pueden usar portaobjetos de vidrio activados con aldehído, polilisina o un reticulador homofuncional. En algunos aspectos, el (los) antígeno(s) o fragmento(s) del mismo pueden estar dispuestos en una matriz tridimensional, por ejemplo, en la micromatriz de almohadilla de gel de poli(acrilamida) tridimensional descrito en Mirzabekov et al., *Nucleic Acids Res* 24(15): 2998-3004 (1996). En un aspecto preferido, el (los) antígeno(s) o fragmento(s) del mismo también pueden inmovilizarse en una superficie de chip biosensor, una perla de poliestireno o similar.

Los métodos y condiciones para la unión al antígeno son bien conocidos en la técnica y se describen adicionalmente en el presente documento. También son bien conocidos en la técnica los métodos y condiciones para detectar complejos de proteínas de unión a antígeno. La detección de complejos de proteínas de unión a antígeno puede ser cualitativa y/o cuantitativa. También se puede detectar la unión de una multiplicidad (generalmente, una gran multiplicidad) de proteínas de unión, por ejemplo, en un conjunto. Los métodos para detectar complejos de proteínas de unión a antígeno incluyen, por ejemplo, ELISA, ensayos inmunológicos fluorescentes, transferencias Western y de puntos, inmunoprecipitaciones, ensayos de competición con polipéptidos de la competencia e inmunoensayos focales, tecnología de resonancia de plasmón superficial (SPR, de sus siglas en inglés), ensayos de detección múltiple, etc.

### ***Interrupción diferencial de antígeno***

En un aspecto preferido, un método de elaboración de perfiles como se desvela en el presente documento se basa, en parte, en el principio de que el grado de similitudes entre los patrones de respuesta (por ejemplo, perfiles de unión) de dos proteínas de unión contra una macromolécula después de la introducción de una serie de cambios estables

independientes en la macromolécula reflejan el grado de similitud entre los epítomos de la macromolécula unida mediante las dos proteínas de unión. La evaluación de dichas interacciones macromoleculares después de realizar cambios en la macromolécula es un método conocido en la técnica como Perfiles Asistidos por Modificación (MAP, de sus siglas en inglés), Elaboración de Perfiles de Anticuerpos Basados en Estructura de antígeno (ASAP, de sus siglas en inglés) o Interrupción Diferencial de antígeno (DAD, de sus siglas en inglés). La DAD es un método que clasifica grandes cantidades de proteínas de unión a antígeno dirigidas contra el mismo antígeno de acuerdo con las similitudes del perfil de unión de cada proteína de unión a antígeno con antígenos, o fragmentos del mismo, modificados química o enzimáticamente (Publicación de Solicitud de Patente de EE.UU. N.º 2004/0101920; véase también Shi et al. (2006) J. Immunol. Methods 314:9-20)). Cada categoría puede reflejar una característica de unión (por ejemplo, un epítomo) bien diferente, o parcialmente solapada con, una característica de unión (por ejemplo, un epítomo) representada por otra categoría. Esta tecnología permite el filtrado rápido de proteínas de unión a antígeno genéticamente idénticas, de modo que la caracterización puede centrarse en proteínas de unión a antígeno genéticamente distintas. La DAD puede usarse para clasificar las proteínas  $V_L$  de unión a antígeno de la presente divulgación en grupos de proteínas de unión a antígeno que exhiben una característica de unión única en comparación con los anticuerpos convencionales, por ejemplo, proteínas  $V_L$  de unión a antígeno que se unen a epítomos enmascarados con anticuerpos típicos.

Preferentemente, la proteína antigénica puede inmovilizarse en superficies de chip biosensor o perlas de poliestireno. Los biosensores basados en afinidad emplean moléculas biológicas, tales como anticuerpos, receptores, ligandos, enzimas, hidratos de carbono o ácidos nucleicos, como transductores de señal en la interfaz entre la electrónica de estado sólido y la biología en fase de solución. Las propiedades de reconocimiento inherentes de estas interacciones biomoleculares pueden ser observadas y medidas por biosensores con un alto grado de sensibilidad y selectividad (para revisión, véase Baird y Myszk (2001) J. Molecular Recognition, 14:261-268).

Las ventajas del uso de biosensores incluyen la capacidad de recopilar datos en tiempo real, proporcionando rápidamente información detallada sobre una reacción de unión, y en segundo lugar, la reacción de unión entre las biomoléculas que interactúan no requiere el marcaje de las biomoléculas, por ejemplo, con marcadores fluorescentes o radioactivos para que se observe la reacción de unión. Biacore AB (Uppsala, Suecia) proporciona actualmente los instrumentos y la tecnología de biosensores más establecidos. Los instrumentos Biacore (modelos 1000, 2000 y 3000) son dispositivos SPR totalmente automatizados, basados en chips sensores que pueden aceptar muestras directamente de placas de 96 pocillos. Cuando se acopla a uno de estos instrumentos, una superficie del sensor, llamada chip, se divide en cuatro celdas de flujo independientes que pueden funcionar individualmente o en serie. Esta configuración de cubeta de lectura permite que el tampón pase continuamente sobre la superficie del sensor, aliviando así la necesidad de etapas de lavado que requieren mucho tiempo cuando se cambia la solución de analito por tampón. Además, los sistemas de flujo continuo aseguran que el ligando esté expuesto a una concentración constante de analito durante la duración del proceso de medición de unión. Además, la disponibilidad de cuatro celdas de flujo en cada chip sensor permite al usuario inmovilizar tres muestras diferentes y mantener una superficie de referencia dentro del mismo chip sensor. Los modelos Biacore 2000 y 3000 son capaces de monitorear interacciones de unión dentro de las cuatro celdas de flujo simultáneamente. La entrega de analito a cada superficie en serie permite la sustracción de referencia en línea y una calidad de datos mejorada (Myszka (1999) J. Mol. Recogn. 12:279-284; Rich et al. (2000) Curr. Opin. Biotechnol. 11:54-71). Otros biosensores tales como los instrumentos IASYS® de Affinity Sensors, SPR670 de Nippon Laser Electronics, Bio-Suplar II de Analytical  $\mu$ Systems y SPREETA™ de Texas Instruments también se pueden utilizar para practicar los métodos de la presente divulgación.

Las perlas de poliestireno pueden procesarse con, por ejemplo, un ensayo tal como un ensayo de detección LUMINEX™ múltiple (Luminex Corp., TX). Debido a la capacidad de LUMINEX™ para manejar el análisis múltiple con hasta 100 tipos diferentes de perlas, LUMINEX™ proporciona superficies de antígeno casi ilimitadas con varias modificaciones, lo que da como resultado una resolución mejorada en la elaboración de perfiles de epítomos de anticuerpos.

La modificación o alteración de la estructura del antígeno puede realizarse mediante un tratamiento químico que tiende a modificar específicamente las cadenas laterales de restos de aminoácidos particulares de la proteína antigénica, o mediante un tratamiento enzimático. Todas las modificaciones pueden realizarse preferentemente en el antígeno que está inmovilizado en una superficie, por ejemplo, una superficie de biosensor, una perla de poliestireno, etc. Se pueden realizar muchos tipos diferentes de modificaciones antigénicas, con cada superficie o perla que comprende antígeno modificado de una manera. Normalmente, se puede incluir en el análisis una superficie de control apropiada en la que se inmoviliza el antígeno no modificado.

Los ejemplos no limitantes de productos químicos que son adecuados para efectuar la alteración o modificación química incluyen ésteres de succinimidilo y sus derivados, compuestos que contienen aminas primarias, hidracinas y carbohidracinas, aminoácidos libres, homo y hetero-oligopéptidos que contienen de dos a veinte restos de longitud, clorhidrato de tris(2-carboxietil)fosfina (TCEP•HCl)/yodoacetamida, N-etil-N'-(dimetilaminopropil)carbodiimida (EDC)/etanolamina, yodoacetamida e hidracina, p-hidroxifenilglicoxal (HPG, de sus siglas en inglés), peróxido de hidrógeno, N-bromosuccinimida, N-acetilimidazol, tetranitrometano, ácido arsanílico, cloruro de dansilo, glutaraldehído, ninhidrina, dietilpirocarbonato (DEPC), sulfosuccina-acetato de imidilo (sulfo-NHS-acetato), polietilenglicol 5000 (PEG-5000) y ácido 7-hidroxycumarina-3-carboxílico, éster de succinimidilo. Los expertos en la técnica reconocerán que aún se pueden usar muchos otros productos químicos en la práctica de la DAD.

Los ejemplos no limitantes de enzimas, específicamente proteasas, que son adecuadas para efectuar la alteración o modificación enzimática del antígeno incluyen tripsina porcina modificada, endoproteinasa Glu-C, endoproteinasa Asp-N, quimotripsina, endoproteinasa Lys-C, y endoproteinasa Arg-C, pepsina, papaína, termolisina, subtilisina, proteasa K, bromelina y proteasa específica de sulfhidrilo (ficina). Una vez más, el experto en la técnica reconocerá fácilmente que podrían usarse otras proteasas en la práctica del método de la presente divulgación.

Usando la tecnología SPR, la unión puede medirse como unidades de resonancia (UR) usando configuraciones experimentales que permiten medir simultáneamente el complejo de proteína de unión-antígeno en todas las superficies, incluidas una superficie no modificada y tres modificadas de cada chip sensor. Las respuestas normalizadas pueden calcularse como porcentajes de respuestas de unión de cada una de las tres superficies modificadas a la superficie del sensor de control (sin modificar). Por lo tanto, se pueden recopilar nueve datos de respuesta (%) de cada muestra pasando cada muestra sobre tres chips sensores preparados por separado, cada uno con una superficie no modificada y tres superficies modificadas de manera diferente.

En un aspecto preferido, el antígeno puede inmovilizarse en una perla de poliestireno. Las perlas comprenden antígeno no modificado y no modificado generado de acuerdo con métodos bien conocidos en la técnica. Usando, por ejemplo, un ensayo de detección múltiple, por ejemplo, tal como el ensayo de detección LUMINEX™, los complejos de proteínas de unión-antígeno se pueden medir como la intensidad de fluorescencia media y se pueden calcular las respuestas normalizadas.

### **Agrupación**

En una aplicación particular y específica, la presente divulgación se refiere a un método para evaluar las interacciones entre proteínas de unión-antígeno, por ejemplo, proteínas  $V_L$  de unión a antígeno y anticuerpos típicos, y los antígenos a los que se dirigen, lo que permite un método rápido para clasificar el antígeno-proteínas de unión en grupos funcionales (también llamados agrupaciones o bins) cuyos miembros, llamados hermanos, exhiben una característica o perfil de unión único y similar a un antígeno, por ejemplo, a un antígeno modificado química o enzimáticamente. Se considera que las proteínas de unión que se agrupan en función de la similitud de sus características o perfiles de unión tienen una característica de unión similar, por ejemplo, se unen al mismo epítipo o epítipos similares. Estas agrupaciones se pueden mostrar opcionalmente en formato de matriz, o en formato de "árbol" como un dendrograma, o en un formato legible por computadora, o en cualquier formato compatible con el dispositivo de entrada de datos. La información sobre las agrupaciones puede capturarse desde una matriz, un dendrograma o mediante una computadora u otro dispositivo informático. La captura de datos puede ser visual, manual, automatizada o cualquier combinación de las mismas.

Como se usa en el presente documento, el término "bin" puede usarse como un sustantivo para referirse a agrupaciones de proteínas de unión identificadas que tienen perfiles de unión similares a un panel de superficies de antígeno modificadas/interrumpidas de acuerdo con los métodos de la presente divulgación. El término "bin" también puede usarse como un verbo para referirse a la práctica de los métodos de la presente divulgación, que incluye cualquier análisis de datos producidos por el ensayo.

Agrupación, como se describe en el presente documento, es el proceso de agrupar proteínas de unión en función de una o más características de unión, por ejemplo, los epítipos que reconocen. Más particularmente, la agrupación comprende métodos y sistemas para discriminar las propiedades de reconocimiento de epítipos de diferentes proteínas de unión, combinadas con procesos informáticos para agrupar proteínas de unión basadas en sus propiedades de reconocimiento de epítipos e identificar bins de proteínas de unión que tienen distintos perfiles de unión. Por consiguiente, los aspectos incluyen ensayos para determinar las propiedades de unión a epítipos de proteínas de unión como se discute en el presente documento, y procesos para analizar datos generados a partir de dichos ensayos.

La agrupación se puede lograr mediante cualquiera de los métodos de: 1) agrupar las características de unión, por ejemplo, mediante un examen visual, tratando cada unión a antígeno exhibida como una barra graduada (por ejemplo, como porcentaje de control de cada superficie de antígeno modificado); 2) calcular el valor determinante de cada matriz de proteína de unión y clasificar todos los determinantes calculados en grupos (véase "Calculus—One and Several Variables", 6ª edición, de Salas y Einar, págs. 715-717, 1990); o 3) aplicar algoritmos de reconocimiento de patrones y programas bioinformáticos relacionados a los datos de perfil de unión generados por los métodos y clasificar las proteínas de unión en grupos funcionales.

En un aspecto, los perfiles de respuesta normalizados para complejos de proteínas de unión-antígeno pueden organizarse en grupos usando el programas informáticos estadísticos apropiados. La agrupación también se puede lograr calculando el determinante de cada matriz de respuesta seguido de la clasificación de los determinantes en grupos y posiblemente inspeccionando visualmente la columna (perfil) de la barra de color graduada de cada grupo para verificar los resultados de la agrupación. Todo el "proceso de agrupación" se puede lograr mediante el reconocimiento de patrones bioinformáticos o el programa informático de extracción de datos. Los ejemplos no limitantes de dichos programas informáticos incluyen los programas disponibles comercialmente utilizados

habitualmente por análisis de micromatriz de ADN como J-express (DeNova, Inc. Vancouver, Columbia Británica), Stanford Gene Cluster Software (Universidad de Stanford, California), StatSoft of Statistica u otros programas no comerciales adecuados desarrollados por expertos.

5 Se pueden emplear varias técnicas para visualizar los perfiles generados como se describe anteriormente. Para que un observador humano haga comparaciones significativas, el espacio en el que se presentan los perfiles debe ser comprensible. Aunque puede ser difícil visualizar tendencias o agrupaciones significativas en espacios de alta dimensión, un aspecto comprende dos o tres dimensiones (características de unión) que se espera que sean más oportunas para un perfil en particular, aunque puede que no sea posible ver otras características de unión  
10 potencialmente significativas en el mismo espacio bidimensional o tridimensional.

Se pueden emplear varias técnicas para abordar este problema. Dichas técnicas crean un espacio dimensional inferior en el que las dimensiones individuales capturan dos o más características de los datos. Ejemplos de dichas técnicas incluyen análisis de componentes principales (PCA), análisis discriminante lineal y no lineal, graduación según una  
15 escala multidimensional y técnicas de búsqueda de proyección. Un enfoque particularmente preferido implica el uso de PCA. El PCA determina los vectores (dimensiones) a través de los cuales un conjunto de datos muestra la mayor variación en el espacio multidimensional. El primer componente principal muestra la dirección de mayor variación en los datos. El segundo componente principal muestra la dirección de la segunda mayor variación en los datos, etc. Se pueden seleccionar tantos componentes principales como sean adecuados para representar los datos. Normalmente,  
20 los primeros uno, dos o tres componentes principales se seleccionan para presentar datos a los observadores humanos. El análisis de componentes principales se describe más completamente en Jackson, J. E. (1991) A User Guide to Principal Components. Nueva York: John Wiley and Sons; y Jolliffe, I. T. (1986) Principal Component Analysis. Nueva York: Springer-Verlag.

25 Se encuentran disponibles varias herramientas disponibles comercialmente para realizar análisis de componentes principales. Los paquetes de cómputo estadístico ejemplares para realizar PCA pueden estar disponibles en Insightful Corporation (anteriormente MathSoft) de Seattle, Wash. o Partek Corporation de St. Louis, Mo, por ejemplo, Partek Genomic Suite Software. El análisis de componentes principales se puede aplicar a perfiles de unión cuantitativos de manera directa. Sin embargo, generalmente será necesario normalizar los conjuntos de datos de perfil antes de  
30 enviarlos al análisis de componentes principales. Esto se debe a que los diversos escalares que comprenden las características individuales de un perfil residen en escalas muy diferentes. Para llevar estas diversas características a una escala comparable para un análisis de PCA significativo, uno puede realizar transformaciones para normalizar los datos. En un aspecto preferido, cada una de las dimensiones se escala considerando todos los datos a lo largo de esa dimensión, restando la media de esos datos y dividiendo por la desviación estándar. Esto eficazmente escala los datos  
35 para la normalización.

En un aspecto preferido, los datos generados por la interrupción del antígeno diferencial pueden normalizarse de la siguiente manera. Los datos sin procesar pueden normalizarse dividiendo la señal de unión de una proteína de unión a antígeno a una superficie de antígeno modificada (o conjunto de perlas) por la señal de unión de la misma proteína  
40 de unión a antígeno a una superficie de antígeno no modificada (o conjunto de perlas). Posteriormente, todos los valores para una superficie dada (o conjunto de perlas) pueden dividirse por el valor medio de todas las proteínas de unión a esa superficie (o conjunto de perlas). Finalmente, todos los valores pueden transformarse usando log 2 como base.

45 En otro aspecto, los perfiles de unión se generan mediante un ensayo de proteína de unión competitiva de alto rendimiento, por ejemplo, el ensayo Multiplicado de agrupación de Anticuerpo Competitivo (MCAB, de sus siglas en inglés), y los datos de entrada se analizan usando el proceso de Reconocimiento de Patrón Competitivo (CPR, de sus siglas en inglés), los cuales se describen en la patente de EE.UU. N.º 8.568.992.

50 Tras la normalización de los perfiles de unión, por ejemplo, intensidades de señal, se pueden usar varios enfoques informáticos bien conocidos para identificar patrones subyacentes en datos complejos. Un enfoque que ha demostrado ser valioso para el análisis de grandes conjuntos de datos biológicos es la agrupación jerárquica. Aplicando este método, las proteínas de unión pueden ser forzadas a una estricta jerarquía de subconjuntos anidados en función de sus valores de disimilitud. En un aspecto ilustrativo, el par de proteínas de unión con el valor de disimilitud más bajo  
55 se agrupa primero. El par o grupo(s) de proteínas de unión con el siguiente valor más pequeño de disimilitud (o disimilitud promedio) se agrupa a continuación. Este proceso se repite reiteradamente hasta que quede una agrupación. De esta manera, las proteínas de unión se agrupan de acuerdo con cuán similares son sus perfiles de unión, en comparación con las otras proteínas de unión. En un aspecto, las proteínas de unión se agrupan en un dendrograma (a veces llamado "árbol filogenético") cuyas longitudes de rama representan el grado de similitud entre  
60 los patrones de unión de las dos proteínas de unión. Longitudes de ramas largas entre dos proteínas de unión indican que probablemente se unen a diferentes epítopos. Longitudes de ramas cortas indican que dos proteínas de unión probablemente compitan por el mismo epítopo.

Los grupos funcionales identificados de acuerdo con los métodos desvelados en el presente documento pueden verificarse utilizando métodos bien conocidos de acuerdo con el principio de que las proteínas de unión en el mismo  
65 grupo funcional deberían compartir una característica de unión única o distinta. En un aspecto, la característica de

unión única o distinta de las proteínas de unión en un único bin da como resultado que las proteínas de unión de ese bin se unan o compitan por el mismo epítipo(s) de un antígeno, en donde las proteínas de unión que representan diferentes grupos funcionales no deberían unirse o competir por el mismo epítipo(s) de un antígeno. En este aspecto, ELISA, ensayos de competición, ensayos de mapeo de epítopos, matrices de péptidos, etc., pueden usarse todos para verificar los bins determinados en el presente documento.

Un bin o grupo funcional comprende todas o sustancialmente todas las proteínas  $V_L$  de unión a antígeno cuando el bin comprende al menos un 90 %, preferentemente al menos un 95 %, más preferentemente al menos un 98 %, y lo más preferentemente al menos un 99 % de proteínas  $V_L$ . En un aspecto, un bin comprende el 100 % de proteínas  $V_L$  de unión a antígeno. En un aspecto, se perfilan cantidades suficientes de proteínas  $V_L$  de unión a antígeno específicas de antígeno y anticuerpos convencionales para una comparación y agrupación significativas. En un aspecto, las proteínas de unión en o aisladas del suero de animales no humanos que expresan proteínas  $V_L$  de unión a antígeno y que están inmunizados con un antígeno se perfilan y comparan con los perfiles de unión de proteínas de unión o se aíslan de suero de animales no humanos de control que están inmunizados con el mismo antígeno. En un aspecto, la inmunización comprende cebar, es decir, administrar el antígeno al animal no humano por primera vez, permitiendo que el animal no humano descanse durante un período de tiempo, por ejemplo, unos días, una semana, dos semanas, tres semanas, cuatro semanas, cinco semanas, etc., y volver a administrar el antígeno al animal no humano una o más veces.

## 20 Ejemplos

Los siguientes ejemplos no limitantes se presentan para proporcionar a los expertos en la técnica una divulgación y descripción completas de cómo preparar y usar animales no humanos descritos en el presente documento y ayudar a la comprensión de los mismos, y no pretenden representar que los experimentos a continuación son todos o los únicos experimentos realizados. Los ejemplos no incluyen descripciones detalladas de métodos convencionales que serían bien conocidos para los expertos en la técnica (técnicas de clonación molecular, etc.). Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (por ejemplo, cantidades, la temperatura, etc.) pero deben tenerse en cuenta algunos errores y desviaciones experimentales. A menos que se indique otra cosa, las partes son partes en peso, el peso molecular es el peso molecular promedio en peso, la temperatura está en grados centígrados y la presión es o está cerca de la atmosférica.

### Ejemplo 1. Generación de animales no humanos que tienen loci modificados de inmunoglobulina

Este ejemplo ilustra métodos ejemplares de genomanipulación de loci de inmunoglobulina de animales no humanos para contener (a) un locus de cadena pesada de inmunoglobulina que comprende segmentos génicos  $V_L$  y  $J_L$  de cadena ligera de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada de inmunoglobulina; y (b) un locus de cadena ligera de inmunoglobulina que comprende segmentos génicos  $V_L$  y  $J_L$  de cadena ligera de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina.

A continuación se describe una construcción de vectores de direccionamiento ejemplares para la inserción de segmentos génicos  $V$  y  $J$  de cadena ligera humanos (por ejemplo,  $V_K$  y  $J_K$ ) en un locus de cadena pesada de inmunoglobulina murino. La Figura 2 ilustra cuatro vectores de direccionamiento ejemplares que contienen una pluralidad de segmentos génicos de cadena ligera  $k$  humana para la inserción en un locus de cadena pesada de inmunoglobulina murino usando recombinación homóloga.

Se realizaron varias construcciones de direccionamiento utilizando la tecnología de genomanipulación VELOCIGENE® (véase, por ejemplo, la patente de EE.UU. 6.586.251 y Valenzuela, D.M., Murphy, A.J., Friendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yassenchak, J., Chernomorsky, R., Boucher, M., Elsasser, A.L., Esau, L., Zheng, J., Griffiths, J.A., Wang, X., Su, H., Xue, Y., Dominguez, M.G., Noguera, I., Torres, R., Macdonald, L.E., Stewart, A.F., DeChiara, T.M., Yancopoulos, G.D. (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* 21, 652-659) para modificar bibliotecas genómicas de cromosomas artificiales bacterianos (BAC, de sus siglas en inglés) de ratón. El ADN de BAC de ratón puede modificarse mediante recombinación homóloga para eliminar los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  endógenos para la inserción posterior de segmentos génicos  $V_L$  y  $J_L$  humanos no reordenados. Alternativamente, los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  endógenos pueden dejarse intactos e inactivados de modo que se inhiba la recombinación de segmentos génicos endógenos para formar una región variable funcional (por ejemplo, por inversión o interrupción de segmentos génicos).

Los ratones genéticamente modificados, y los métodos para fabricarlos, cuyo genoma contiene un locus de cadena pesada de inmunoglobulina que comprende segmentos génicos  $V_L$  y  $J_L$  de cadena ligera de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada de inmunoglobulina se describen en la publicación de solicitud de patente de EE.UU. N.º 2012-0096572 A1. Tal como se muestra en la Figura 2, se genomanipularon cuatro vectores de direccionamiento para insertar progresivamente 40 segmentos génicos  $V_K$  humanos y cinco segmentos génicos  $J_K$  humanos en un locus de cadena pesada de ratón inactivado (por ejemplo, segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$  endógenos eliminados) usando técnicas moleculares estándar



reconocidas en la técnica. La Tabla 3 expone el tamaño del ADN humano incluido en cada vector de direccionamiento, que contiene varios segmentos génicos de cadena ligera k humanos para su inserción en un locus de cadena pesada de inmunoglobulina de ratón. Se puede incluir cualquier número de segmentos génicos  $V_K$  y  $J_K$  humanos en los vectores de direccionamiento. Los vectores de direccionamiento ejemplares expuestos en la Figura 2 incluyen 5 segmentos génicos de cadena ligera k humanos que se encuentran de manera natural en el contig proximal del locus de cadena ligera k humano de la línea germinal (Figura 1). El locus de cadena pesada endógeno resultante después de la inserción sucesiva de los cuatro vectores de direccionamiento se muestra en la parte inferior de la Figura 2.

**TABLA 3**

Vector de direccionamiento	Tamaño de la secuencia K humana	Segmentos génicos K humanos añadidos	
		$V_K$	$J_K$
1	~ 110,5 kb	4-1, 5-2, 7-3, 2-4, 1-5, 1-6	1 - 5
2	~ 140 kb	3-7, 1-8, 1-9, 2-10, 3-11, 1-12, 1-13, 2-14, 3-15, 1-16	-
3	~ 161 kb	1-17, 2-18, 2-19, 3-20, 6-21, 1-22, 1-23, 2-24, 3-25, 2-26, 1-27, 2-28, 2-29, 2-30	-
4	~ 90 kb	3-31, 1-32, 1-33, 3-34, 1-35, 2-36, 1-37, 2-38, 1-39, 2-40	-

Usando un enfoque similar, se pueden construir otras combinaciones de dominios variables de cadena ligera humanos en el contexto de regiones constantes de cadena pesada murinas. Pueden proceder dominios variables adicionales de cadena ligera de segmentos génicos humanos  $V_\lambda$  y  $J_\lambda$ . Los vectores de direccionamiento ejemplares que incluyen ADN humano que incluyen varios números de segmentos génicos  $V_\lambda$  y  $J_\lambda$  humanos se exponen en la Figura 3.

El locus de cadena ligera  $\lambda$  humano se extiende más de 1.000 kb y contiene más de 80 genes que codifican segmentos variables (V) o de unión (J). De acuerdo con los informes publicados, entre los 70 segmentos génicos  $V_\lambda$  del locus de cadena ligera  $\lambda$  humano, entre 30 y 38 parecen ser segmentos génicos funcionales. Las 70 secuencias  $V_\lambda$  están organizadas en tres grupos, todos los cuales contienen diferentes miembros de distintos grupos de familias de genes V (grupos A, B y C). Dentro del locus de cadena ligera  $\lambda$  humano, más de la mitad de todos los dominios  $V_\lambda$  observados están codificados por los segmentos génicos 1-40, 1-44, 2-8, 2-14 y 3-21. Hay siete segmentos génicos  $J_\lambda$ , de los cuales solo cuatro se consideran generalmente segmentos génicos  $J_\lambda$  funcionales,  $J_\lambda 1$ ,  $J_\lambda 2$ ,  $J_\lambda 3$  y  $J_\lambda 7$ . En algunos alelos, un quinto par de segmentos génicos  $J_\lambda$ - $C_\lambda$ , es supuestamente un pseudogen ( $C_\lambda 6$ ). La incorporación de múltiples segmentos génicos  $J_\lambda$  humanos, en un locus de cadena pesada híbrido, como se describe en el presente documento, puede construirse mediante síntesis de novo. De esta forma, un fragmento genómico que contiene múltiples segmentos génicos  $J_\lambda$  humanos, en la configuración de la línea germinal, se genomanipula con múltiples segmentos génicos  $V_\lambda$  humanos y permite la recombinación V-J normal en el contexto de una región constante de cadena pesada. Un vector de direccionamiento ejemplar que incluye múltiples segmentos génicos  $J_\lambda$  se muestra en la Figura 3 (Vector de direccionamiento 1').

El acoplamiento de dominios variables de cadena ligera con regiones constantes de cadena pesada representa una fuente potencialmente rica de diversidad para generar proteínas  $V_L$  de unión a antígeno únicas con regiones  $V_L$  humanas en animales no humanos. Explotar esta diversidad del locus de cadena ligera  $\lambda$  humano (o locus k humano como se describió anteriormente) en ratones da como resultado la genomanipulación de cadenas pesadas híbridas únicas y da lugar a otra dimensión de proteínas de unión al repertorio inmunitario de animales genéticamente modificados y su posterior uso como plataforma de próxima generación para la generación de tratamientos.

Los vectores de direccionamiento descritos anteriormente se usan para electroporar células madre embrionarias (ME) de ratón para crear células ME modificadas para generar ratones quiméricos que expresan proteínas  $V_L$  de unión a antígeno (es decir, segmentos génicos de cadena ligera humanos unidos operativamente a regiones constantes de cadena pesada de ratón). Las células MS que contienen una inserción de segmentos génicos de cadena ligera humanos no reordenados se identifican mediante el ensayo cuantitativo de PCR, TAQMAN® (Lie y Petropoulos, 1998. Curr. Opin. Biotechnology 9:43-48). Los conjuntos y sondas de cebadores específicos están diseñados para la inserción de secuencias humanas y casetes de selección asociados, la pérdida de secuencias de cadena pesada de ratón y la retención de secuencias de ratón que flanquean el locus de cadena pesada endógeno.

Las células ME que llevan los segmentos génicos de cadena ligera humanos (por ejemplo,  $V_K$  y  $J_K$ ) pueden transfectarse con una construcción que expresa una recombinasa para eliminar cualquier casete de selección no deseado introducido por la inserción de los segmentos génicos de cadena ligera humanos. Opcionalmente, el casete de selección se puede eliminar criando a ratones que expresan la recombinasa (por ejemplo, documentos US 6.774.279). Opcionalmente, el casete de selección queda retenido en los ratones.

Las células ME dirigidas anteriormente descritas se utilizan como células ME donantes y se introducen en un embrión de ratón en la fase de 8 células mediante el método VELOCIMOUSE® (véase, por ejemplo, la patente de EE.UU. N.º 7.294.754 y Poueymirou, W.T., Auerbach, W., Frenthewey, D., Hickey, J.F., Escaravage, J.M., Esau, L., Dore, A.T., Stevens, S., Adams, N.C., Dominguez, M.G., Gale, N.W., Yancopoulos, G.D., DeChiara, T.M., Valenzuela, D. M. (2007). F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic

analyses. Nat Biotechnol 25, 91-99). VELOCIMICE® (ratones F0 totalmente procedentes de la célula ME donante) que portan independientemente segmentos génicos de cadena ligera humanos en un locus de cadena pesada de inmunoglobulina de ratón se identifican mediante genotipado usando una modificación del ensayo de alelos (Valenzuela et al., supra) que detecta la presencia de los segmentos génicos únicos de cadena ligera humanos en un locus endógeno de cadena pesada de inmunoglobulina. Las crías se genotipan y se selecciona una cría heterocigota u homocigota para el locus de cadena pesada de inmunoglobulina modificado genéticamente para caracterizar la expresión de cadenas pesadas que contienen V<sub>L</sub>.

La introducción de segmentos génicos de cadena ligera k humanos en un locus de cadena pesada de ratón se llevó a cabo en una línea ME F1 (F1H4; Valenzuela et al. 2007, supra) procedente de embriones heterocigotos 129S6/SvEvTac y C57BL/6NTac que además contenían un reemplazo *in situ* de los segmentos génicos de cadena ligera k de ratón con segmentos génicos de cadena ligera k humanos (por ejemplo, véanse las Patentes de EE.UU. N.º 6.596.541 y 8.642.835).

Se generaron ratones que comprenden loci de cadena pesada genéticamente modificados que contienen segmentos génicos V<sub>L</sub> y J<sub>L</sub> de cadena ligera de inmunoglobulina humanos no reordenados en un locus de cadena pesada (ratones KOH: MAID1713: 40 segmentos génicos V<sub>K</sub> humanos y cinco segmentos génicos J<sub>K</sub> humanos; MAID1994: 40 segmentos génicos V<sub>K</sub> humanos y cinco segmentos génicos J<sub>K</sub> humanos, y un gen *Adam6* integrado) como se describió anteriormente. En resumen, en ratones KOH, todos los segmentos génicos variables de cadena pesada funcionales endógenos se eliminaron y se reemplazaron con 40 segmentos génicos V<sub>K</sub> humanos no reordenados y cinco (5) segmentos génicos J<sub>K</sub> humanos no reordenados, que están unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada de inmunoglobulina.

Se criaron ratones humanizados homocigotos VELOCIMMUNE® (VI3; véase la Patente de EE.UU. N.º 8.642.835 y la Patente de EE.UU. N.º 8.502.018 B2) con ratones KOH homocigotos (MAID1713 o MAID 1994) para producir un ratón heterocigoto para el alelo de cadena ligera modificado y el alelo KOH. Los ratones heterocigotos F1 generados a partir de este cruce se criaron entre sí para obtener ratones homocigotos para cada alelo (MAID1713HO 1242HO, MAID1994HO 1242HO). Dichos ratones expresan proteínas V<sub>L</sub> de unión a antígeno que tienen una estructura que se asemeja a la de las inmunoglobulinas, pero que son distintas porque dichas proteínas de unión carecen de dominios variables de cadena pesada. La presencia de los alelos genéticamente modificados en los loci de cadena pesada y cadena ligera de inmunoglobulina se confirmó mediante cribado y cariotipo TAQMAN™ utilizando sondas y cebadores específicos descritos anteriormente. Los ratones KOH homocigotos comprenden una inserción de segmentos génicos de cadena ligera humanos no reordenados como se describe en el presente documento (por ejemplo, V<sub>K</sub> y J<sub>K</sub> humanos) en el locus de cadena pesada de ratón en el que se han eliminado todos los segmentos génicos V, D y J endógenos variables de cadena pesada y una inserción de segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, V<sub>K</sub> y J<sub>K</sub> humanos) en el locus de cadena ligera kappa (k) de ratón en el que se han eliminado todos los genes V<sub>K</sub> y J<sub>K</sub> de ratón. En algunos aspectos, los ratones KOH comprenden además un gen *Adam6* integrado.

Los ratones cuyo genoma comprende (i) un alelo de cadena pesada de inmunoglobulina que contiene una inserción de cuarenta (40) segmentos génicos V<sub>K</sub> y cinco (5) J<sub>K</sub> humanos no reordenados de modo que dichos segmentos génicos V<sub>K</sub> y J<sub>K</sub> humanos están unidos operativamente a regiones constantes de cadena pesada endógenas, y (ii) un alelo de cadena ligera de inmunoglobulina que contiene una inserción de cuarenta (40) segmentos génicos V<sub>K</sub> y cinco (5) J<sub>K</sub> humanos no reordenados, de modo que dichos segmentos génicos V<sub>K</sub> y J<sub>K</sub> humanos están unidos operativamente a una región constante de cadena ligera endógena, se denominan MAID1713/1242, "ratones KOH" (véase la publicación de solicitud de patente de EE.UU. n.º 2012-0096572 A1). Los ratones que tienen lo mismo y también un gen *Adam6* de ratón integrado se denominan MAID1994/1242 (véase la publicación de solicitud de patente de EE. UU. N.º 2013-0212719 A1).

## Ejemplo 2. Generación y caracterización de proteínas V<sub>L</sub> de unión a antígeno

El presente ejemplo describe la producción de proteínas de unión a antígeno a partir de ratones genomanipulados específicamente para expresar moléculas similares a inmunoglobulinas que comprenden dominios variables de cadena ligera de inmunoglobulina y carecen de dominios variables de cadena pesada (como se describió anteriormente). Este presente ejemplo ilustra específicamente la generación de proteínas de unión a antígeno ejemplares específicas para moléculas pequeñas (por ejemplo, un esteroide y un alcaloide del producto natural), que contienen (i) dos polipéptidos que comprenden cada uno un dominio variable de cadena ligera de inmunoglobulina unido a dominio constante de cadena ligera de inmunoglobulina, y (ii) dos polipéptidos que comprenden cada uno un dominio variable de cadena ligera de inmunoglobulina unido a un dominio constante de cadena pesada de inmunoglobulina.

Las proteínas V<sub>L</sub> de unión a antígeno se obtienen de ratones genéticamente modificados cuyo genoma incluye loci de cadena pesada y ligera de inmunoglobulina que contienen cada uno segmentos génicos de cadena ligera humanos no reordenados (por ejemplo, segmentos génicos V<sub>L</sub> y J<sub>L</sub>) unidos operativamente a regiones constantes de cadena pesada y ligera endógenas, respectivamente. Dichos ratones proporcionan un fuerte sistema *in vivo* para fabricar proteínas de unión a antígeno a dianas no proteináceas en comparación con ratones genéticamente modificados de tipo silvestre y/o de control.

**Inmunización**

5 Generalmente, un ratón como se describe en el presente documento se estimula con un antígeno, y las células (tales como los linfocitos B) se recuperan del animal (por ejemplo, del bazo o los ganglios linfáticos). Las células pueden fusionarse con una línea celular de mieloma para preparar líneas celulares de hibridoma inmortales, y dichas líneas celulares de hibridoma se criban y seleccionan para identificar líneas celulares de hibridoma que producen anticuerpos que contienen cadenas pesadas híbridas específicas para el antígeno utilizado para la inmunización. El ADN que codifica las regiones  $V_K$  humanas de las cadenas pesadas y cadenas ligeras híbridas puede aislarse y unirse a regiones constantes deseables, por ejemplo, cadena pesada y/o cadena ligera. Debido a la presencia de segmentos génicos  $V_K$  humanos fusionados a regiones constantes de cadena pesada de ratón, se produce un repertorio único similar a un anticuerpo y la diversidad del repertorio de inmunoglobulinas aumenta dramáticamente como resultado del formato único similar al anticuerpo creado. Esto confiere un nivel adicional de diversidad al repertorio específico de antígeno tras la inmunización. Las secuencias clonadas resultantes pueden producirse posteriormente en una célula, tal como una célula CHO. Alternativamente, el ADN que codifica las proteínas  $V_L$  de unión a antígeno específicas de antígeno o los dominios variables se puede aislar directamente de linfocitos específicos de antígeno (por ejemplo, linfocitos B; véase el documento US 7.582.298 B2).

20 Inicialmente, las proteínas  $V_L$  de unión a antígeno de alta afinidad se aíslan con regiones  $V_K$  humanas y regiones constantes de ratón. Como se ha descrito anteriormente, las proteínas  $V_L$  de unión a antígeno se caracterizan y seleccionan por características deseables, que incluyen afinidad, selectividad, epítopo, etc. Las regiones constantes de ratón pueden reemplazarse con una región constante humana deseada para generar proteínas  $V_L$  de unión a antígeno completamente humanas únicas que contienen dominios  $V_K$  humanos mutados somáticamente a partir de un locus de cadena pesada híbrido no reordenado de la presente divulgación. Las regiones constantes humanas adecuadas incluyen, por ejemplo, IgG1 o IgG4 de tipo silvestre o modificadas o, alternativamente, C $\kappa$  o C $\lambda$ .

30 Se inmunizaron por separado cohortes separadas de ratones KOH con un producto alcaloide natural (antígeno A) y un esteroide (antígeno B). También se inmunizaron cohortes separadas de "VI3" (ratones humanizados VELOCIMMUNE®, Véase la Patente de EE.UU. N.º 8.642.835 y Patente de EE.UU. N.º 8.502.018 B2) y ratones "ULC" (véanse los documentos US 2011-0195454A1, US 2012-0021409A1, US 2012-0192300A1, US 2013-0045492A1, US 2013-0185821A1 y US 2013-0302836A1) para proporcionar perfiles de respuesta inmunitaria comparables.

35 En resumen, el antígeno A se conjugó con KLH y se usó como inmunógeno para inmunizar ratones KOH, VI3 y ULC. Para el antígeno B, se usó un conjugado de BSA como inmunógeno para inmunizar cepas de KOH y VI3. Se recogió suero inmunitario de los ratones antes del inicio de la inmunización. El inmunógeno se administró a 2,35  $\mu$ g de conjugado para la inmunización de cebado inicial mezclado con 10  $\mu$ g de oligonucleótido CpG (Invivogen) como adyuvante en un volumen de 25  $\mu$ l mediante inyección en la almohadilla plantar (f.p., de sus siglas en inglés). Posteriormente, los ratones se estimularon por la misma ruta con 2,35  $\mu$ g de inmunógenos respectivos junto con 10  $\mu$ g de CpG y 25  $\mu$ g de Adju-Phos (Brenntag) como adyuvantes en los días 3, 6, 11, 13, 17, 20 para un total de 6 estimulaciones. Los ratones fueron sangrados los días 15 y 22 después de la 4ª y 6ª estimulación, respectivamente. El antisuero se analizó para determinar los valores de los conjugados KLH del antígeno A. Para el antígeno B, los valores se analizaron en el antígeno conjugado con BSA y BSA. Para los ratones KOH, después de completar 6 estimulaciones, se permitió a los ratones una fase de reposo de 4 a 5 semanas, después de lo cual se administraron 4 estimulaciones adicionales con los inmunógenos. Se desangraron los ratones y se ensayaron los valores de antisuero.

50 Cuando se consigue una respuesta inmunitaria deseada, se recogen los esplenocitos y se fusionan con células de mieloma de ratón para conservar su viabilidad y formar líneas celulares de hibridoma. Las líneas celulares de hibridoma se criban y seleccionan para identificar líneas celulares que producen proteínas  $V_L$  de unión a antígeno específicas de antígeno. Usando esta técnica, se obtienen varias proteínas  $V_L$  de unión a antígeno específicas de antígeno (es decir, proteínas de unión que poseen dominios  $V_K$  humanos en el contexto de dominios constantes de cadena pesada y ligera de ratón).

55 Alternativamente, las proteínas  $V_L$  de unión a antígeno específicas de antígeno se aíslan directamente de linfocitos B positivos para antígeno sin fusión con las células de mieloma, como se describe en la Patente de EE.UU. N.º 7.582.298. Usando este método, se obtuvieron varias proteínas  $V_L$  de unión a antígeno específicas de antígeno completamente humanas (es decir, anticuerpos que poseen dominios  $V_K$  humanos y dominios constantes humanos).

**Determinación de valores de antisuero**

60 Los valores séricos contra un inmunógeno se determinaron mediante un ELISA estándar. A continuación se describe el ensayo en detalle. Noventa placas de microtitulación de seis pocillos (Thermo Scientific) se revistieron a 2  $\mu$ g/ml con conjugados BSA de antígeno A (un producto alcaloide natural aromático sustituido) o antígeno B (un esteroide) en solución salina tamponada con fosfato (PBS, Irvine Scientific) durante la noche a 4 °C. Al día siguiente, las placas se lavaron con solución salina tamponada con fosfato que contenía Tween 20 al 0,05 % (PBS-T, Sigma-Aldrich) cuatro veces usando un lavador de placas (Molecular Devices). Luego se bloquearon las placas con 250  $\mu$ l de seroalbúmina

bovina al 0,5 % (BSA, Sigma-Aldrich) en PBS y se incubaron durante 1 hora a temperatura ambiente. Las placas se lavaron cuatro veces con PBS-T. Los sueros de ratones inmunizados y presueros inmunitarios se diluyeron en serie tres veces en BSA-PBS al 0,5 % a partir de 1:300 o 1:1000, se añadieron a las placas bloqueadas por duplicado y luego se incubaron durante 1 hora a temperatura ambiente. Los dos últimos pocillos se dejaron en blanco para ser utilizados como control secundario de anticuerpos (control de fondo). Las placas se lavaron nuevamente cuatro veces con PBS-T en un lavador de placas. El anticuerpo secundario de cabra anti-IgG-Fc de ratón conjugado con peroxidasa de rábano picante (HRP, de sus siglas en inglés) (Jackson ImmunoResearch) se añadió a las placas a una dilución 1:5000/1:10.000 y se incubó durante 1 hora a temperatura ambiente. Luego se lavaron las placas ocho veces con PBS-T y se desarrollaron usando TMB/H<sub>2</sub>O<sub>2</sub> como sustrato. El sustrato se incubó durante 20 minutos y la reacción se detuvo con ácido sulfúrico 2 N (H<sub>2</sub>SO<sub>4</sub>, VWR, n.º de catálogo BDH3500-1) o ácido fosfórico 1 N (JT Baker, n.º de catálogo 7664-38-2). Las placas se leyeron en un espectrofotómetro (Victor, Perkin Elmer) a 450 nm. Los valores se calcularon utilizando el programa informático Graphpad PRISM.

La respuesta inmunitaria inducida en ratones al inmunógeno inyectado se midió como valores, que se define como el recíproco de la dilución sérica más alta en la que la absorbancia de unión al antígeno es dos veces mayor sobre el antecedente. Al final del curso de inmunización, tanto los ratones KOH como los VI3 obtuvieron valores altos comparables.

#### **Identificación de proteínas de unión mediante Luminex**

Para preparar perlas acopladas a antígeno para el cribado, se activaron 0,12 ml de suspensión de perlas Luminex (microesferas carboxiladas, Luminex Corp.) en tampón de fosfato de sodio 0,1 M (JTBaker, n.º de Cat. 4011-01) a pH 6,2 mediante la adición de 15 µl de 50 mg/ml de EDC (1-etil-3-(3-dimetilaminopropil)carbodiimida, Sigma, n.º de Cat. 03449) y 15 µl de 50 mg/ml de Sulfo-NHS (N-hidroxisuccinimida, Pierce, n.º de Cat. 24510) seguido de incubación a temperatura ambiente durante 10 minutos. Posteriormente, se añadieron 0,5 ml de 20 µg/ml de antígeno A conjugado con BSA (un producto alcaloide natural aromático sustituido) en tampón MES 50 mM a pH 5 (ACROS, n.º de Cat. 327765000) a las perlas activadas, y se dejó que la reacción de acoplamiento de amina primaria continuara durante dos horas, y los grupos reactivos restantes en las perlas se enfriaron mediante la adición de un volumen 1/10 de solución Tris 1 M a pH 8 (Teknova, n.º de Cat. T1080). Las perlas se lavaron con PBS (Life Technologies, n.º de Cat. 14190-144) que contenía Tween-20 al 0,05 % (Calbiochem, n.º de Cat. 655205), y se almacenaron en tampón PBS que contenía BSA al 2 % p/v (Sigma, n.º de Cat. A4503). De la misma manera, también se preparó un lote de perlas de control negativo con proteína BSA acoplada.

Para seleccionar las proteínas de unión, se distribuyeron alícuotas de 75 µl que contenían 3000 perlas de BSA-antígeno A a cada pocillo prehidratado de placas de filtro de 96 pocillos (Millipore, n.º de Cat. MSBVN1250). Cada muestra de proteína de unión (25 µl) se añadió a cada pocillo y las placas se incubaron durante la noche en un agitador de placas a 4 °C. En la mañana del segundo día, las perlas se lavaron con tampón PBS que contenía Tween-20 al 0,05 % (PBS-T) usando un colector de vacío, y se detectó proteína de unión unida a la perla mediante incubación de las perlas con 0,1 ml de 1,25 µg/ml de anticuerpo de cabra anti-IgG humana conjugado con R-ficoeritrina (Southern Biotech, n.º de Cat. 2063-09) en PBS-T durante 30 minutos a temperatura ambiente. Luego se lavaron las perlas y se suspendieron en 0,15 ml de PBS-T, y se midió la intensidad de fluorescencia media (MFI, de sus siglas en inglés) con un analizador basado en citometría de flujo Luminex. De una manera similar, se prepararon perlas de antígeno B (un esteroide) conjugado con BSA y se cribaron muestras que contenían proteínas de unión.

#### **Cinética de unión relativa**

Se incubó previamente 50 nM de Neutravidina con 200 nM de antígeno marcado con biotina durante al menos 24 horas antes del comienzo de la criba. El marcado de Neutravidina en la molécula pequeña mejoró la sensibilidad de la detección de afinidad de rendimiento de los sobrenadantes brutos de proteínas de unión al aumentar el peso de la masa de la molécula pequeña. La superficie del sensor Biacore, que se inmovilizó por primera vez con anticuerpo específico anti-Fc humano o anti-Fc de ratón, se usó para capturar anticuerpos de medios acondicionados en bruto. Luego se inyectaron las soluciones de molécula pequeña/Neutravidina sobre la superficie capturada de la proteína de unión durante dos minutos, seguido de la disociación del complejo unido durante 10 minutos. El experimento se realizó a 25 °C utilizando HBST como tampón de ejecución.

La Figura 4 establece el número total (izquierda) y el porcentaje (derecha) de anticuerpos antígeno positivos (es decir, proteínas V<sub>L</sub> de unión a antígeno) obtenidos de ratones KOH y ratones humanizados VELOCIMMUNE®. La Figura 5 establece una cinética de unión ejemplar de anticuerpos contra el antígeno B obtenido de ratones KOH y ratones humanizados VELOCIMMUNE®.

Los resultados mostraron que los ratones humanizados VELOCIMMUNE® (VI3) produjeron 10 de 528 muestras de proteínas de unión que tenían una MFI > 1000 en las perlas de antígeno A. Para las perlas de antígeno B, los ratones humanizados VELOCIMMUNE® (VI3) mostraron solo dos de 350 muestras de proteínas de unión que tienen una MFI superior a 1000. En contraste, los ratones KOH mostraron 453 de 528 muestras que tenían una MFI > 1000 en las perlas de antígeno A. En las perlas de antígeno B, los ratones KOH mostraron 74 de 339 muestras que tenían una MFI > 1000. Todas las muestras positivas para antígeno mostraron una unión mínima o insignificante en las perlas

BSA de control negativo (por ejemplo, MFI ~ 118).

**Uso del segmento génico k humano**

- 5 Para caracterizar aún más las proteínas  $V_L$  de unión a antígeno anti-antígeno A o anti-antígeno B producidas en los ratones de acuerdo con la presente divulgación, se clonaron y se secuenciaron ácidos nucleicos que codifican los dominios  $V_K$  humanos (tanto de cadenas pesadas como ligeras de la proteína  $V_L$  de unión a antígeno) usando métodos adaptados de los descritos en el documento US 2007/0280945A1. A partir de las secuencias de ácido nucleico y las secuencias de aminoácidos predichas de los anticuerpos, se identificó el uso de genes para la región variable de
- 10 cadena pesada híbrida de proteínas  $V_L$  de unión a antígeno seleccionadas y purificadas obtenidas de los ratones inmunizados con el antígeno A o B (descrito anteriormente). La Tabla 4 expone el uso de segmentos génicos  $V_K$  y  $J_K$  humanos de proteínas  $V_L$  de unión a antígeno anti-antígeno A seleccionadas. La Tabla 5 expone el uso de segmentos génicos  $V_K$  y  $J_K$  de proteínas  $V_L$  de unión a antígeno anti-antígeno B seleccionadas.
- 15 Los datos de uso de genes muestran que los ratones de acuerdo con la presente divulgación pueden generar regiones variables de cadena pesada híbridas únicas contra un antígeno pequeño, que proceden de una variedad de segmentos génicos  $V_K$  y  $J_K$  humanos en el locus de cadena pesada de inmunoglobulina. El uso de segmentos génicos  $V_K$  y  $J_K$  humanos demuestra aún más la reordenación diversa y variada dentro de su locus, así como en comparación con los segmentos génicos  $V_K$  y  $J_K$  de cadena ligera. Además, la diversidad es evidente en el uso del segmento génico entre
- 20 la cadena pesada híbrida y la cadena ligera.

TABLA 4

Proteína $V_L$	Cadena pesada híbrida		Cadena ligera	
	$V_K$	$J_K$	$V_K$	$J_K$
1	3-20	4	4-1	2
2	3-20	4	1-5	2
3	4-1	1	4-1	3
4	4-1	1	3-20	3
5	1-5	5	3-20	1
6	3-20	4	1-5	2

(continuación)

Proteína V <sub>L</sub>	Cadena pesada híbrida		Cadena ligera	
	V <sub>K</sub>	J <sub>K</sub>	V <sub>K</sub>	J <sub>K</sub>
7	4-1	1	3-20	2
8	3-20	4	1-5	2
9	4-1	1	3-20	3
10	4-1	1	3-20	3
11	1-33	1	1-33	3
12	4-1	1	3-20	3
13	4-1	1	3-20	3
14	4-1	1	3-20	2
15	3-20	3	4-1	1
16	1-33	1	3-20	3
17	3-20	3	4-1	1
18	4-1	1	3-20	1
19	4-1	1	3-20	3
20	4-1	1	3-20	3
21	4-1	1	3-20	3
22	4-1	1	3-20	3
23	1-33	3	3-20	5

TABLA 5

Proteína V <sub>L</sub>	Cadena pesada híbrida		Cadena ligera	
	V <sub>K</sub>	J <sub>K</sub>	V <sub>K</sub>	J <sub>K</sub>
24	1-5	3	3-20	3
25	3-15	5	1-39	3
26	1-5	4	3-20	2
27	1-5	4	3-20	3
28	1-5	5	3-20	2
29	1-5	3	3-20	2
30	4-1	3	3-20	2
31	4-1	1	3-20	2
32	1-5	4	3-20	1
33	1-5	5	3-20	1
34	4-1	1	3-20	2

**Determinación de afinidad**

5 Las constantes de disociación de equilibrio ( $K_D$ ) para los sobrenadantes de proteínas V<sub>L</sub> de unión a antígeno purificados específicos de antígeno B se determinaron mediante SPR (resonancia de plasma superficial) utilizando un instrumento BIACORE™ 2000 (GE Healthcare). Todos los datos se obtuvieron utilizando DPBS + DMSO al 0,1 % como muestra y tampón en funcionamiento a 25 °C.

10 En resumen, cada proteína V<sub>L</sub> de unión a antígeno purificada estaba en una superficie de chip sensor CM5 previamente derivatizada con una alta densidad de proteína A usando química de acoplamiento de amina estándar. Durante la etapa de captura, se inyectó proteína V<sub>L</sub> de unión a antígeno anti-antígeno B purificada a través de la superficie de la proteína A a una velocidad de flujo de 5 µl/min, durante un total de 3-4 minutos. La etapa de captura fue seguido por una inyección de tampón o analito en funcionamiento en un intervalo de concentración de dilución triple de solución madre 270 µM -13,7 nM durante 1,5 minutos a una velocidad de flujo de 100 µl/min. La disociación del antígeno de la proteína V<sub>L</sub> de unión a antígeno purificada capturada se controló durante al menos 5 minutos. La proteína V<sub>L</sub> de unión a antígeno purificada capturada se eliminó mediante una breve inyección de glicina 10 mM, pH 1,5. Todos los sensogramas se referenciaron dos veces restando los sensogramas de las inyecciones de tampón de los sensogramas de analito, eliminando así los artefactos causados por la disociación de la proteína V<sub>L</sub> de unión a antígeno purificada de la superficie de captura. Los datos de unión para cada proteína V<sub>L</sub> de unión a antígeno purificada se ajustaron a un modelo de unión 1:1 con transporte de masa usando el programa informático de evaluación Biacore T100 v2.1. La Tabla 6 proporciona los datos de unión para un anticuerpo disponible comercialmente específico para antígeno B, once proteínas V<sub>L</sub> de unión a antígeno específicas de antígeno B purificadas y 3 anticuerpos de control obtenidos de animales VI3 de control.

25

TABLA 6

Proteína VL/AcM	ka	kd	KD	t1/2 (segundos)
AcM comercial	1,03E+06	5,85E-02	56,9 nM	12
Proteína VL 1	4,82E+06	4,58E-02	9,49 nM	15
Proteína VL 2	6,40E+05	7,43E-03	11,6 nM	93
Proteína VL 3	NC	NC	NC	NC
Proteína VL 4	1,35E+06	5,98E-03	4,4 nM	116
Proteína VL 5	1,19E+06	7,11E-03	6,0 nM	97
Proteína VL 6	8,50E+05	7,41E-03	8,7 nM	94
Proteína VL 7	NU	NU	NU	NU
Proteína VL 8	1,01E+06	4,46E-03	4,4 nM	156
Proteína VL 9	1,04E+05	2,02E-01	1,93 µM	3
Proteína VL 10	2,42E+06	8,10E-02	33,4 nM	9
Proteína VL 11			≥ 270 µM	
AcM 1 de control	NU	NU	NU	NU
AcM 2 de control			≥ 270 µM	
AcM 3 de control	Estado estable		6,3 µM	

NC = análisis de ajuste de unión 1:1 no concluyente debido a baja señal  
NU = no unido

Las afinidades de unión de 11 proteínas VL de unión a antígeno anti-antígeno B purificadas variaron, todas exhibiendo una  $K_D$  en el intervalo de aproximadamente 4,4 nM a 1,93 µM. Notablemente, siete de las once proteínas VL de unión a antígeno exhibieron una  $K_D$  de aproximadamente 10 nM o menos. En contraste, el anticuerpo disponible comercialmente tenía una afinidad de unión al antígeno A de -57 nM, y ninguno de los tres anticuerpos aislados de los animales de control exhibió una  $K_D$  en el intervalo nanomolar. Las mediciones de  $T^{1/2}$  para las proteínas VL de unión a antígeno purificadas que exhiben una  $K_D$  en el intervalo nanomolar bajo variaron entre 15 y 156 segundos. Sin desear quedar ligado a ninguna teoría en particular, las fluctuaciones en los perfiles de unión de las proteínas VL de unión a antígeno purificadas mostradas en la Tabla 6 y, particularmente, las bajas afinidades o falta de unión por algunas de las proteínas VL de unión a antígeno purificadas, pueden ser un resultado de una o más proteínas VL de unión a antígeno que reconocen un epítipo del antígeno A que está presente solo cuando está unido al vehículo. Independientemente, los datos de afinidad que usan anticuerpos purificados son consistentes con las proteínas VL de unión a antígeno resultantes de la asociación combinatoria de dominios variables de cadena ligera humanos reordenados unidos a regiones constantes de cadena pesada y ligera (descritas en la Tabla 4) que son de alta afinidad, seleccionadas clonalmente, mutadas somáticamente, capaces de unir moléculas pequeñas con alta eficacia y, por lo tanto, terapéuticamente oportunas.

### Ejemplo 3: Elaboración de perfiles de características de unión

#### Inmunización

Las cohortes de ratones KOH se inmunizaron por separado con una glucoproteína secretada humana (Antígeno C) comprada en R&D systems. También se inmunizaron cohortes separadas de "Adam6/VI3" (ratones humanizados VELOCIMMUNE®, véase la Patente de EE.UU. N.º 8.642.835 y la Patente de EE.UU. N.º 8.502.018 B2 que tiene un gen Adam6 integrado), ratones "ULC" (véanse los documentos US 2011-0195454A1, US 2012-0021409A1, US 2012-0192300A1, US 2013-0045492A1, US 2013-0185821A1 y US 2013-0302836A1), y ratones Balb/c de tipo silvestre para proporcionar perfiles de respuesta inmunitaria comparables.

El antígeno C conjugado con hapteno se usó como inmunógeno para inmunizar ratones KOH, Adam6/VI3, ULC y Balb/c. Se recogió suero inmunitario de los ratones antes del inicio de la inmunización. El inmunógeno se administró a 2,35 µg de conjugado para la inmunización de cebado inicial mezclado con 10 µg de oligonucleótido CpG (Invivogen) como adyuvante en un volumen de 25 µl mediante inyección en la almohadilla plantar (f.p., de sus siglas en inglés). Posteriormente, los ratones se estimularon por la misma ruta con 2,35 µg de inmunógeno junto con 10 µg de CpG y 25 µg de Adju-Phos (Brenntag) como adyuvantes en los días 3, 6, 11, 13, 17, 20 para un total de 6 estimulaciones. Los ratones fueron sangrados los días 15 y 22 después de la 4ª y 6ª estimulación, respectivamente. El antisero se analizó para determinar los valores de anticuerpos para conjugar hapteno de antígeno C. Para ratones KOH, después de completar 6 estimulaciones, se permitió a los ratones una fase de reposo de 4 a 5 semanas, después de lo cual se administraron 4 estimulaciones adicionales con los inmunógenos. Se desangraron los ratones y se ensayaron los valores de antisero.

#### Preparación y modificación de antígeno C en perlas

Para preparar perlas acopladas a antígeno para el cribado, se activaron 0,12 ml de suspensión de perlas Luminex (microesferas carboxiladas, Luminex Corp.) en tampón de fosfato de sodio 0,1 M (JTBaker, n.º de Cat. 4011-01) a pH

6,2 mediante la adición de 15 µl de 50 mg/ml de EDC (1-etil-3-(3-dimetilaminopropil)carbodiimida, Sigma, n.º de Cat. 03449) y 15 µl de 50 mg/ml de Sulfo-NHS (N-hidroxisuccinimida, Pierce, n.º de Cat. 24510) seguido de incubación a temperatura ambiente durante 10 minutos. Posteriormente, se añadió 0,5 ml de 20 µg/ml de antígeno C en tampón MES 50 mM a pH 5 (ACROS, n.º de Cat. 327765000) a las perlas activadas, y se dejó que la reacción de acoplamiento de amina primaria continuara durante dos horas, y los grupos reactivos restantes en las perlas se enfriaron mediante la adición de un volumen 1/10 de solución Tris 1 M a pH 8 (Teknova, n.º de Cat. T1080). Las perlas se lavaron con PBS (Life Technologies, n.º de Cat. 14190-144) que contenía Tween-20 al 0,05 % (Calbiochem, n.º de Cat. 655205), y se almacenaron en tampón PBS que contenía BSA al 2 % p/v (Sigma, n.º de Cat. A4503). De la misma manera, también se preparó un lote de perlas de control negativo.

Diecinueve conjuntos de perlas junto con el antígeno C se trataron individualmente con uno de los siguientes reactivos de interrupción del antígeno diferencial: tripsina, Glu-C, Asp-N, quimotripsina, Lys-C, Arg C, Pepsina, acetato de sulfo-NHS, EDC/etanolamina, TCEP/Yodoacetamida, PEG-5000, papaína, termolisina, subtilisina, proteasa K, bromelaína, ficina y H1193 o ácido 7-hidroxycumarina-3-carboxílico, éster de succinimidilo. El tratamiento químico consistió en incubar el conjunto de perlas en productos químicos reactivos recién disueltos 10 mM en solución tamponada con fosfato (PBS) durante 90 minutos a temperatura ambiente. El tratamiento proteolítico consistió en incubar el conjunto de perlas en 10-100 mg de enzima recién disuelta en PBS u otro tampón recomendado durante 90 minutos a temperatura ambiente. Un conjunto de perlas adicional se incubó en PBS durante 90 minutos a temperatura ambiente y el antígeno C acoplado a este conjunto de perlas no se modificó. Después de las incubaciones anteriores, los conjuntos de perlas se lavaron en PBS que contenía Tween 20 al 0,05 % (PBS-T) y se almacenaron en PBS con BSA al 5 % y azida de sodio al 0,02 %.

Para seleccionar las proteínas de unión, se agruparon las 19 perlas de antígeno modificado y las perlas de antígeno de control no modificado, como se ha descrito anteriormente. Se distribuyeron alícuotas de setenta y cinco (75) µl que contenían 3000 perlas en cada pocillo prehidratado de placas de filtro de 96 pocillos (Millipore, n.º de Cat. MSBVN1250). Cada muestra de anticuerpo (25 µl) se añadió a cada pocillo y las placas se incubaron durante la noche en un agitador de placas a 4 °C. En la mañana del segundo día, las perlas se lavaron con PBS-T usando un colector de vacío, y el anticuerpo unido a la perla se detectó mediante incubación de las perlas con 0,2 ml de 1,25 µg/ml de anticuerpo de cabra anti-IgG de ratón o humana conjugado con R-ficoeritrina en PBS-T para 45 minutos a temperatura ambiente. Luego se lavaron las perlas y se suspendieron en 0,2 ml de PBS-T, y se midió la intensidad de fluorescencia media (MFI) con un espectrofotómetro de flúor Luminex. Los datos de unión se sometieron a análisis de datos bioinformáticos como se describió anteriormente.

La Figura 6 proporciona una visualización de PCA en 2D de 736 grupos de proteínas de unión a antígeno C en base a los datos de perfil de epítipo de interrupción de antígeno diferencial. Destacado por el rectángulo es un grupo de epítipos único que no comparte las características de unión del epítipo con los anticuerpos convencionales probados. Los miembros de este único epítipo bin son proteínas V<sub>L</sub> de unión a antígeno generadas en ratones que comprenden un locus de inmunoglobulina que codifica una cadena de inmunoglobulina híbrida que tiene una región variable codificada por uno o más segmentos génicos de región variable de cadena ligera fusionados operativamente a una región de cadena pesada codificada por uno o más genes de región constante de cadena pesada.

#### ***Preparación de antígeno C modificado en superficies de biosensor***

El antígeno C, una glucoproteína secretada, se acopla a una superficie de chip biosensor CM5 mediante un procedimiento de acoplamiento de amina mediado por NHS/EDC estándar. La cantidad de antígeno C acoplado a cada superficie de la cubeta de lectura está entre 3000 y 10000 UR. Para minimizar un efecto de hacinamiento, la densidad de acoplamiento preferida es de alrededor de 5000 UR. Se tiene cuidado de acoplar cantidades casi idénticas de antígeno C a las cuatro celdas de flujo para que se puedan hacer comparaciones justas entre la unión a las tres señales modificadas de la cubeta de lectura y la superficie de la cubeta de lectura de control no modificada.

Se utilizan seis enzimas proteolíticas de grado de secuenciación para modificar cada superficie de antígeno C acoplada: tripsina, endoproteinasa Glu-C y endoproteinasa Asp-N para modificar las celdas de flujo 2, 3 y 4 de un primer chip biosensor y quimotripsina, endoproteinasa Lys-C y endoproteinasa Arg-C para modificar las celdas de flujo 2, 3 y 4 de un segundo chip biosensor. El Biacore 2000 se establece en el modo de cubeta de lectura único a una velocidad de flujo de 2 µl/min y se inyecta 60 µl de tripsina 200 µg/ml en Tris-HCl 0,1 M, pH 8,0 en la cubeta de lectura 2. Se puede observar un sensograma de curvatura hacia abajo como un perfil de digestión proteolítica típico, lo que indica que la tripsina está eliminando específicamente la masa digerible con tripsina. La misma dosis de enzima se inyecta repetidamente en la cubeta de lectura hasta que se forme una superficie estable. Cuando se completa la digestión con tripsina en la cubeta de lectura 2, se inyectan 60 µl de 50 µg/ml de endoproteinasa Glu-C en el mismo tampón que la tripsina en la cubeta de lectura 3. De nuevo, la misma dosis de enzima se inyecta repetidamente en la misma cubeta de lectura hasta que se forme una superficie estable. De una manera similar, se inyectan 60 µl de 50 µg/ml de endoproteinasa Asp-N en el mismo tampón en la cubeta de lectura 4 para crear una superficie modificada endoAsp-N estable. Al final de los tratamientos con enzimas, el Biacore 2000 está configurado en modo de cubeta de lectura. Se ejecuta un tampón de regeneración en las cuatro superficies de antígeno C para generar superficies de trabajo finales estables.



Las proteínas de unión específicas para el antígeno C generadas en animales no humanos que tienen loci de inmunoglobulina modificados y animales de control como se describe en el Ejemplo 2, así como los anticuerpos comercialmente disponibles previamente caracterizados específicos para el antígeno C, se transfieren a una nueva placa de microtitulación de 96 pocillos y se mezclan con 75 µl de tampón de dilución 2X (Hepes 20 mM, pH 7,4, NaCl 300 mM, P-20 al 0,01 %, CMDX 40 mg/ml). El medio de control apropiado mezclado con el tampón de dilución 2X se usa como control negativo.

Cada muestra de proteína de unión se inyecta en las cuatro celdas de flujo y las señales de unión (UR) de cada cubeta de lectura se registran al final de la inyección y las superficies se regeneran. El ciclo de unión/regeneración para cada muestra de anticuerpos está controlado por el Programa de Asistente de Automatización provisto por el fabricante de Biacore.

Las celdas de flujo 2, 3 y 4 de un segundo chip que contiene una cantidad idéntica de antígeno C se digieren con quimotripsina, endoproteinasa Lys-C, y endoproteinasa Arg-C, respectivamente, de una manera similar a la descrita anteriormente en la preparación del primer chip. El mismo conjunto de muestras de proteínas de unión se inyecta en las cuatro celdas de flujo y sus señales de unión (UR) se recogen de la misma manera que el primer chip.

Cantidades idénticas de antígeno C se acoplan a las cuatro celdas de flujo de un tercer chip CM5 mediante un protocolo de acoplamiento de aldehído estándar (BIA Applications Handbook, 4.5). La cantidad de antígeno C acoplado a cada superficie de la cubeta de lectura está entre 3000 y 10000 UR, con la cantidad de acoplamiento preferida en alrededor de 5000 UR para minimizar cualquier efecto de hacinamiento. Para modificar cualquier amina ε de lisina en el antígeno C sin desnaturar su estructura, se inyecta acetato de sulfo-NHS 5 mM disuelto en solución salina tamponada con fosfato (PBS) a 5 µl/min en la cubeta de lectura 2 durante 20 minutos. Para modificar cualquier grupo de ácido carboxílico de cualquier resto de ácido glutámico y ácido aspártico en el antígeno C sin desnaturar su estructura, se inyecta EDC 200 mM disuelto en H<sub>2</sub>O en la cubeta de lectura 3 al mismo caudal durante 7 minutos seguido de una inyección de hidracina 50 mM disuelta en H<sub>2</sub>O durante 7 minutos. Para el tratamiento desnaturante del antígeno C, se inyecta TCEP 100 mM disuelto en Tris-HCl 0,1 M, pH 8,0, en la cubeta de lectura 4 al mismo caudal durante 20 minutos seguido de inyección de yodoacetamida 100 mM disuelta en Tris-HCl 0,1 M, pH 8,0. Al final de los tratamientos, el Biacore 2000 está configurado en modo de cubeta de lectura. Se inyecta un tampón de regeneración en las cuatro superficies de antígeno C tres veces para generar una superficie de trabajo final estable.

Cuando se recopilan los datos de unión de los tres chips separados que contienen las nueve superficies de antígeno C modificadas y tres superficies de control de antígeno C no modificadas, todos los nueve valores de UR de respuesta de cada proteína de unión a las nueve superficies de antígeno C modificadas se convierten en relaciones de respuesta a la de los controles no modificados. Los datos de respuesta de todas las preparaciones de proteínas de unión analizadas se someten a análisis de datos bioinformáticos como se describió anteriormente.

Los resultados de las distribuciones de grupos de epítopos se muestran mediante métodos de visualización de reconocimiento de patrones típicos (no supervisados). Uno de estos métodos de visualización son los árboles jerárquicos (Dendrogramas) que describen las relaciones de agrupación de las proteínas de unión en una disposición similar a un árbol. En el árbol jerárquico, las proteínas de unión que probablemente comparten epítopos se unirán mediante "brazos" relativamente más cortos, donde aquellos que probablemente compartan epítopos se unirán por "brazos" relativamente más largos.

#### **45 Verificación de grupos de proteínas de unión mediante mapeo de epítopos**

Las proteínas de unión de dos grupos funcionales diferentes (o agrupaciones o bins) según lo determinado por DAD se pueden verificar por otros métodos tales como ELISA, ensayo de competición, etc. Un ensayo de mapeo de epítopos es normalmente realizado por instrumentos Biacore u Octet. Los anticuerpos de dos grupos funcionales diferentes no deberían interactuar con el mismo epítipo. Por lo tanto, la unión de un primer anticuerpo de un grupo al antígeno inmovilizado no debe impedir la unión de un segundo anticuerpo de un grupo diferente en ningún grado significativo. En cambio, los anticuerpos del mismo grupo deben exhibir una competencia casi completa entre sí cuando se unen a su antígeno.

Los grupos funcionales identificados usando DAD también se verifican usando una matriz de péptidos procedentes de la secuencia primaria de antígeno C. Los péptidos procedentes del antígeno C o los péptidos superpuestos para cubrir la secuencia completa del antígeno C se preparan como matrices de puntos en una membrana de PVDF o se imprimen en portaobjetos de micromatrices de proteínas normales. Las proteínas de unión que representan diferentes grupos funcionales o proteínas de unión del mismo grupo funcional se incuban con las matrices de péptidos seguidas de una transferencia de puntos estándar o procedimientos de unión y tinción de matrices de proteínas. Las proteínas de unión del mismo grupo funcional, que reconocen el mismo epítipo, deben mostrar patrones de unión idénticos o casi idénticos en las láminas o diapositivas de la matriz de péptidos. En cambio, las proteínas de unión de diferentes grupos funcionales, que reconocen un epítipo diferente en el antígeno C, deben mostrar un patrón de unión diferente a la matriz de péptidos.

**Ejemplo 4. Evaluación de proteínas V<sub>L</sub> de unión específicas para moléculas pequeñas**

Se evaluaron las características estructurales de las proteínas V<sub>L</sub> de unión generadas contra el antígeno A, el antígeno B y el antígeno C como se desvela en los ejemplos 1-3. En particular, se determinó la longitud de la CDR3 de cadenas híbridas y ligeras de proteínas V<sub>L</sub> de unión específicas para el antígeno A (una molécula pequeña alcaloide; n = 132), el antígeno B (una molécula pequeña esteroidea; n = 87) o el antígeno C (una macromolécula de glucoproteína, n = 61). La Tabla 7 muestra el número de cadenas híbridas que tienen una longitud de aminoácidos de la CDR3 de 6, 7, 8, 9, 10, 11 o 12 de proteínas V<sub>L</sub> de unión específicas para el antígeno A, el antígeno B o el antígeno C. La Tabla 8 muestra el número de cadenas ligeras que tienen una longitud de aminoácidos de la CDR3 de 7, 8, 9 o 10 de estas mismas proteínas V<sub>L</sub> de unión. La Figura 7 proporciona estos datos en formato de gráfico de barras.

**TABLA 7**

Longitud de CDR3	Antígeno A	Antígeno B	Antígeno C	Total
6	3	38		41
7	2		6	8
8	1			1
9	31	17		48
10	94		45	139
11	1	30	10	41
12		2		2

**TABLA 8**

Longitud de CDR3	Antígeno A	Antígeno B	Antígeno C	Total
7	2		1	3
8	9	1	1	11
9	98	86	48	232
10	23		11	34

- 15 La longitud de la CDR3 en cadenas ligeras de proteínas V<sub>L</sub> de unión fue consistentemente de aproximadamente 9 aminoácidos independientemente de la especificidad del antígeno. En contraste, la longitud de la CDR3 en las cadenas híbridas de las proteínas V<sub>L</sub> de unión evaluadas fue más variable, particularmente para las proteínas V<sub>L</sub> de unión específicas de moléculas pequeñas. Las cadenas híbridas de proteínas V<sub>L</sub> de unión específicas para el antígeno C, una glucoproteína, tenían longitudes de la CDR3 de aproximadamente de 10 a 11 aminoácidos de longitud, y algunas
- 20 tenían menos de 10 aminoácidos. En contraste, es probable que la CDR3 de cadenas híbridas de proteínas V<sub>L</sub> de unión específicas de moléculas pequeñas, por ejemplo, el antígeno A o el antígeno B, tenga menos de 10 aminoácidos de longitud. Poco menos de la mitad (aproximadamente un 40 %) de las proteínas V<sub>L</sub> de unión específicas del antígeno B tenían una longitud de la CDR3 de 6 aminoácidos.
- 25 Tomados en conjunto, estos ejemplos demuestran que los animales no humanos, por ejemplo, roedores y ratones en particular, genéticamente modificados para producir las proteínas V<sub>L</sub> de unión a antígeno como se describe en el presente documento, proporcionan un fuerte sistema *in vivo* para la generación eficaz de proteínas V<sub>L</sub> de unión a antígeno específicas de antígeno que exhiben características de unión no exhibidas por anticuerpos típicos, por ejemplo, la capacidad de unir moléculas pequeñas con una alta afinidad, posiblemente mediante el uso de un paratopo
- 30 o superficie de unión nueva en la molécula pequeña que no es adecuada para la unión por anticuerpos convencionales.

## REIVINDICACIONES

1. Un método para producir una proteína  $V_L$  de unión a antígeno que se une específicamente a un esteroide que comprende las etapas de:

5 (a) inmunizar un ratón genéticamente modificado con dicho esteroide o dicho esteroide unido a un vehículo, en donde el ratón genéticamente modificado comprende

10 (i) segmentos génicos variables de cadena ligera ( $V_L$ ) y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena pesada de inmunoglobulina de ratón y

(ii) segmentos génicos variables de cadena ligera ( $V_L$ ) de inmunoglobulina y de unión de cadena ligera ( $J_L$ ) de inmunoglobulina humanos no reordenados unidos operativamente a una secuencia de ácido nucleico de región constante de cadena ligera de inmunoglobulina de ratón; y

15 (b) aislar una proteína  $V_L$  de unión a antígeno, o célula que expresa la proteína  $V_L$  de unión a antígeno, a partir del ratón inmunizado,

20 en donde la proteína  $V_L$  de unión a antígeno comprende un primer y un segundo dominio variable de cadena ligera de inmunoglobulina, en donde el primer dominio variable de cadena ligera de inmunoglobulina está unido operativamente a un dominio constante de cadena pesada de inmunoglobulina, en donde el segundo dominio variable de cadena ligera de inmunoglobulina está unido operativamente a un dominio constante de cadena ligera de inmunoglobulina, y en donde el primer y el segundo dominio variable de cadena ligera de inmunoglobulina tienen secuencias heterogéneas y están asociados para formar un punto de unión de la proteína  $V_L$  de unión a antígeno que se une específicamente a dicho esteroide con una  $K_D$  de 50 nM o menos.

2. El método de la reivindicación 1, en donde el método comprende además:

30 (c) recoger una proteína  $V_L$  de unión a antígeno del sobrenadante de un cultivo de hibridoma, en donde el hibridoma se produce a partir de la célula aislada en la etapa (b).

3. El método de la reivindicación 1, que además comprende

35 (c) identificar la primera y segunda secuencia de ácido nucleico de región variable de cadena ligera de inmunoglobulina que codifican el primer y segundo dominio variable de cadena ligera de inmunoglobulina; y

(d) expresar las secuencias de ácido nucleico de (c) en un sistema de expresión adecuado para expresar la proteína de unión a antígeno para formar una proteína de unión a antígeno que comprende un dímero del primer dominio variable de cadena ligera de inmunoglobulina fusionado con un dominio  $C_H$  humano y el segundo dominio variable de cadena ligera de inmunoglobulina fusionado con un dominio  $C_L$  humano, en donde el primer y segundo dominio variable de cadena ligera de inmunoglobulina se unen a dicho esteroide.

40 4. El método de la reivindicación 3, en donde el sistema de expresión comprende el primer y segundo ácido nucleico que codifica dichos primer y segundo dominios variables de cadena ligera de inmunoglobulina fusionados con un dominio  $C_H$  humano y un dominio  $C_L$  humano respectivamente.

45 5. El método de una cualquiera de las reivindicaciones 1-4, en donde dicho esteroide es una hormona, un alcaloide o un esteroide cardiotónico.

50 6. El método de una cualquiera de las reivindicaciones 1-4, en donde dicho esteroide unido a un vehículo es un hapteno.

7. El método de una cualquiera de las reivindicaciones 1-4, en donde el punto de unión de la proteína  $V_L$  de unión a antígeno se une específicamente a dicho esteroide con una  $K_D$  de 30 nM o menos.

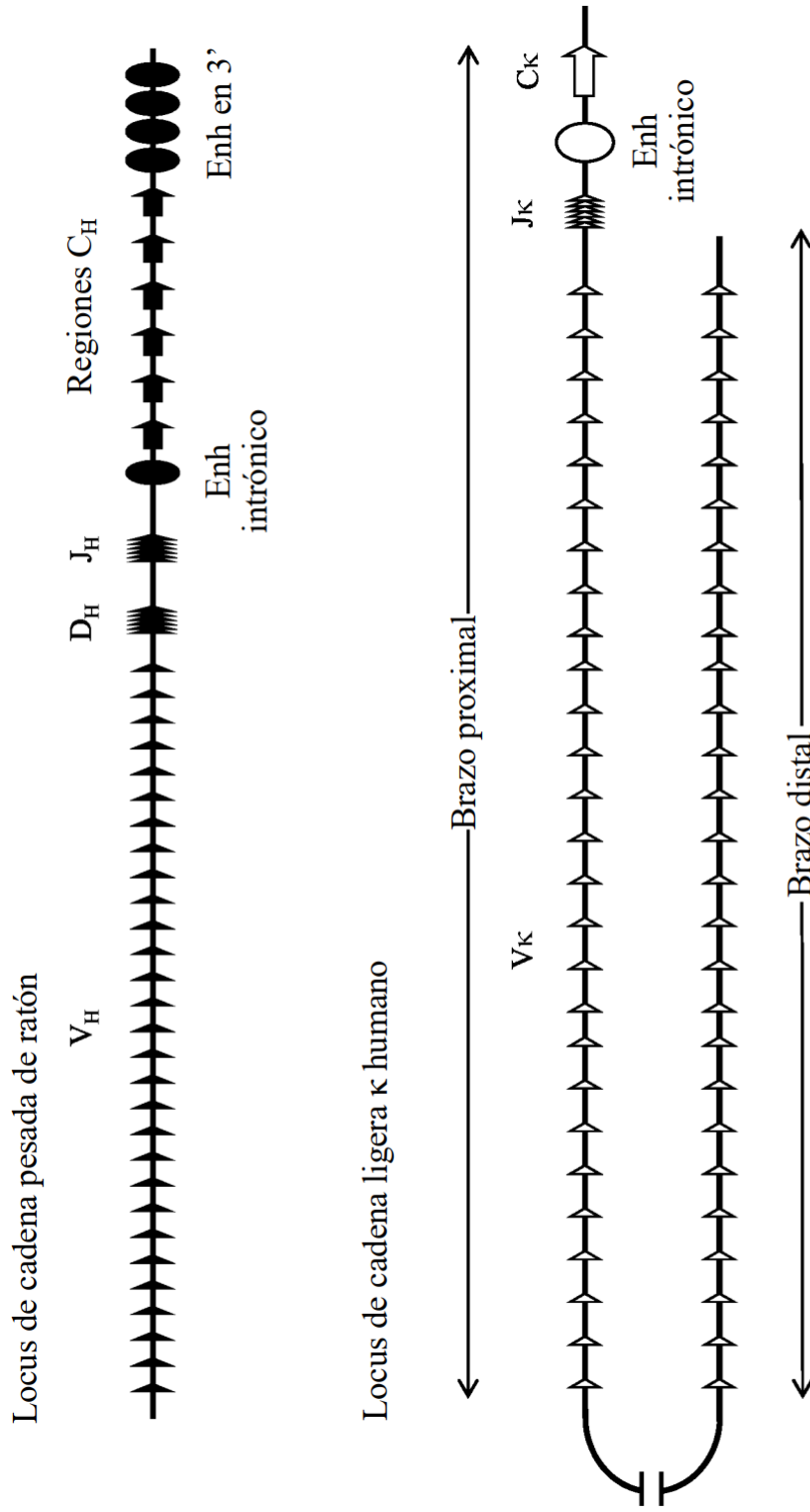


FIG. 1

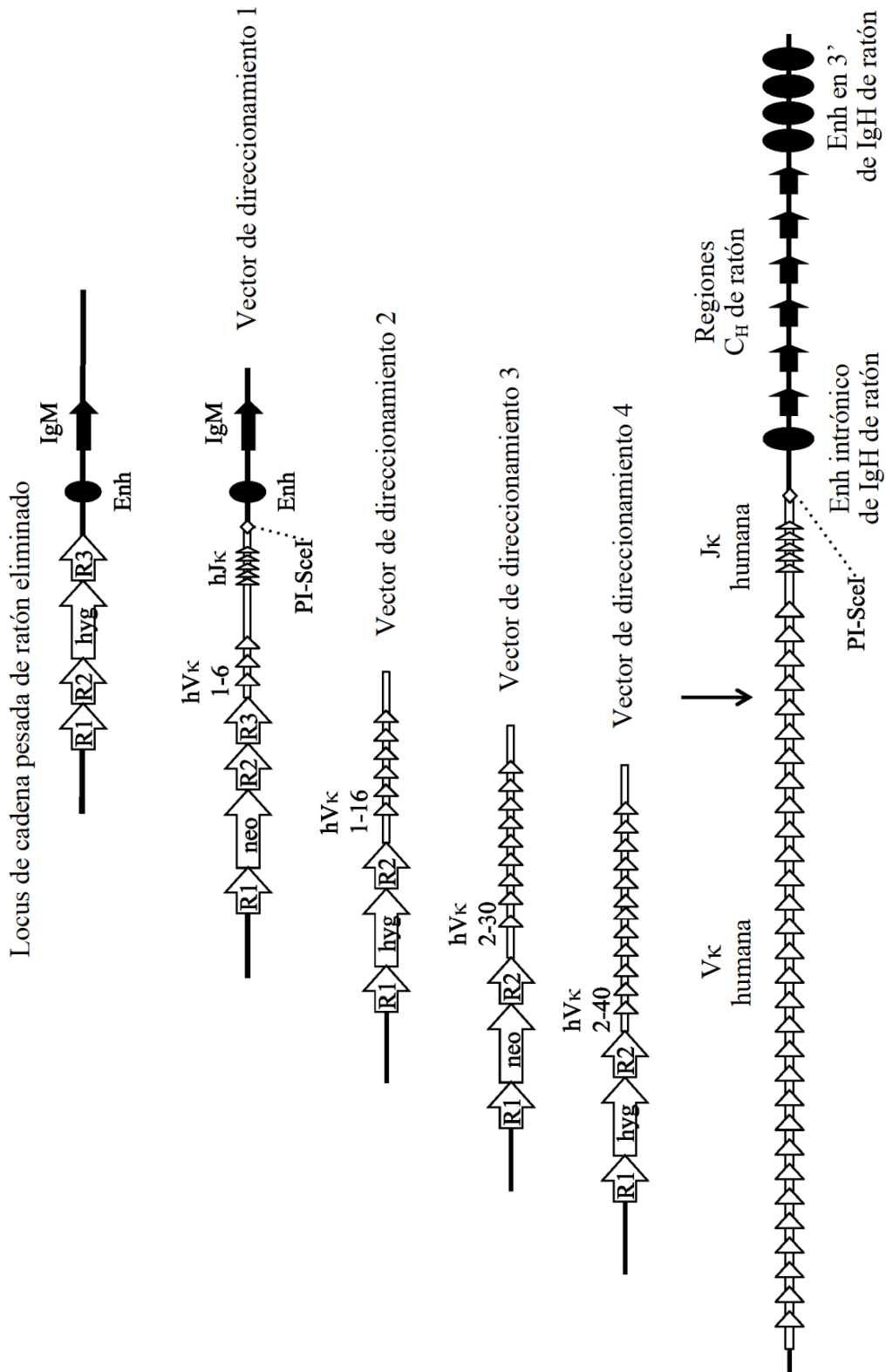


FIG. 2

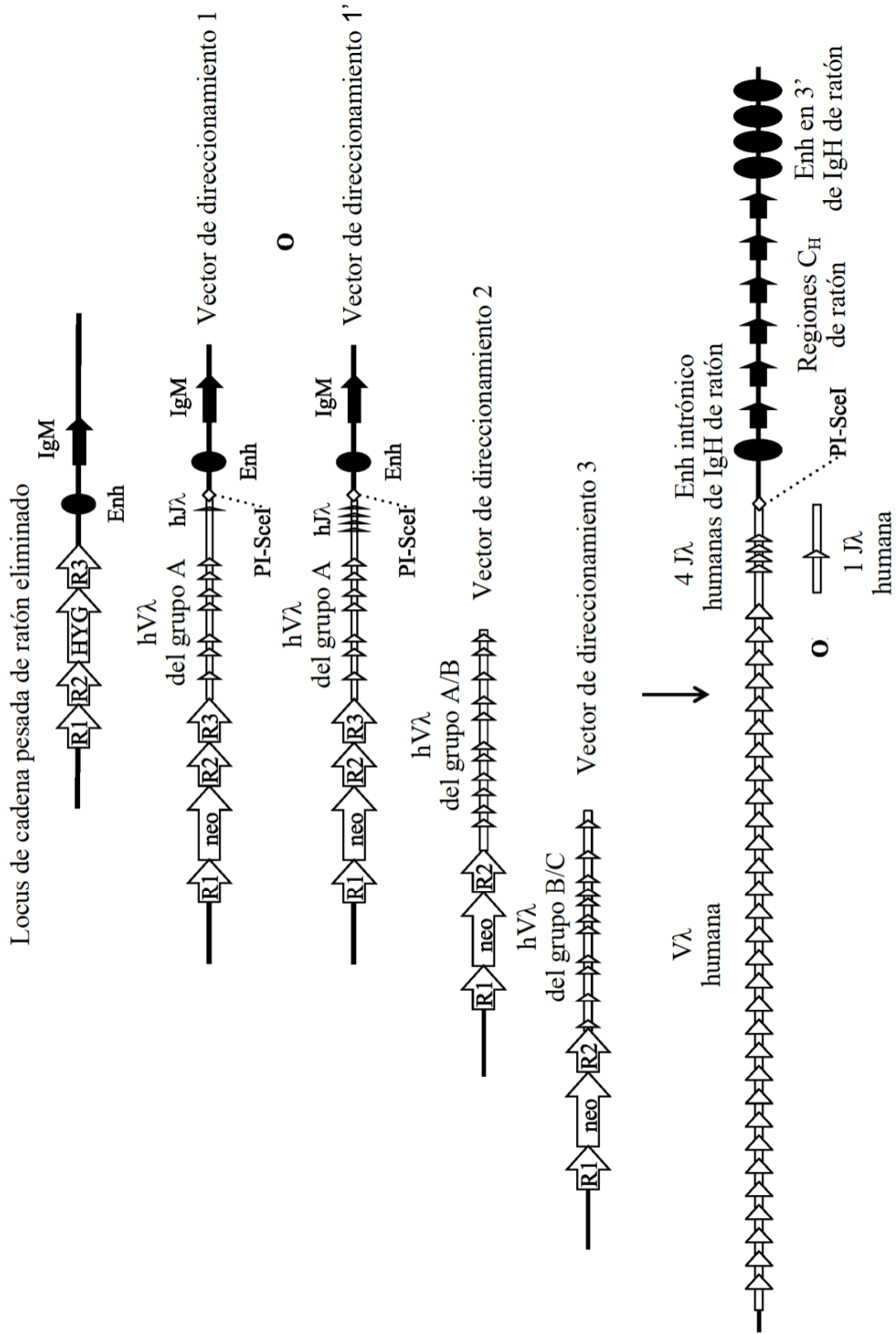


FIG. 3

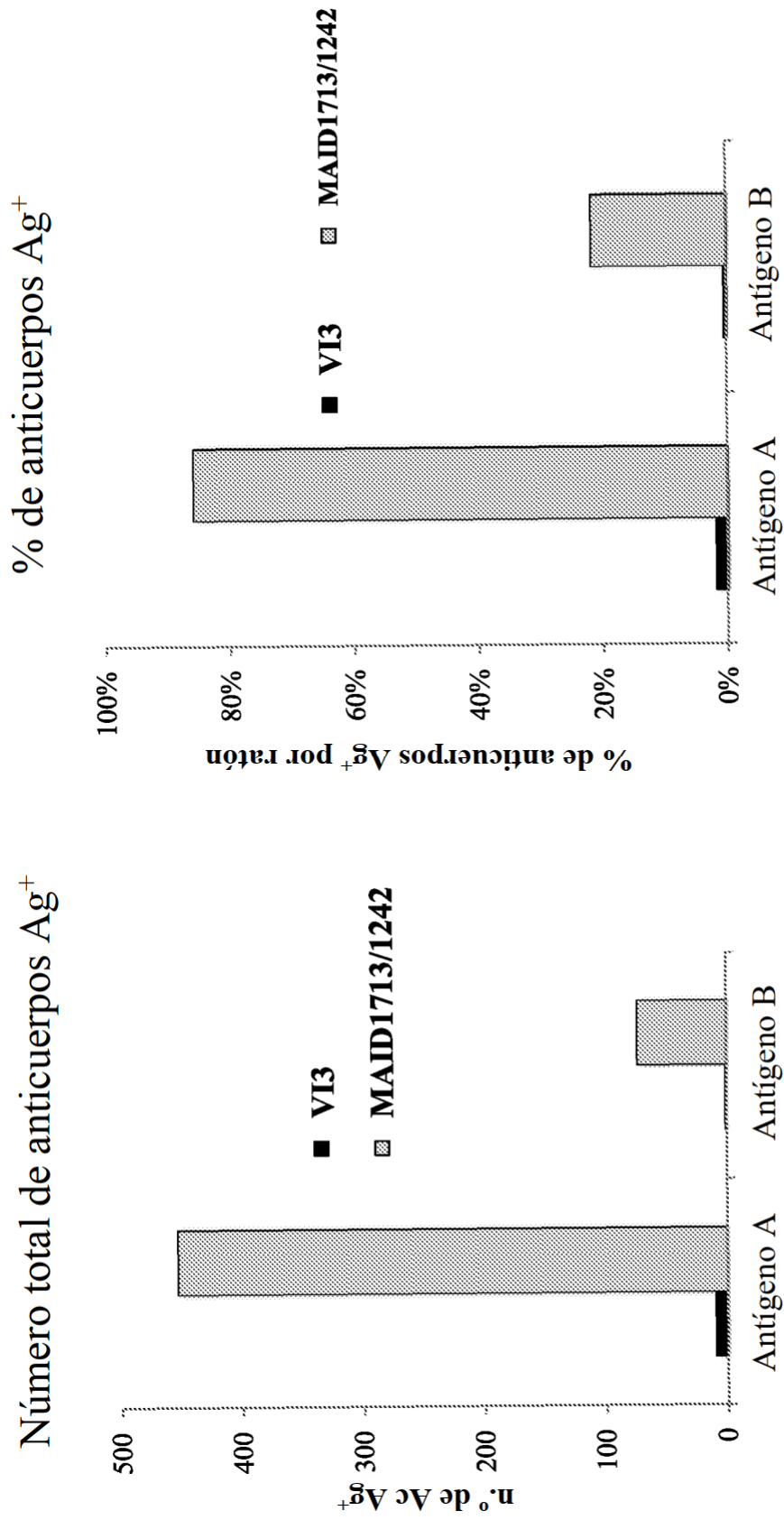


FIG. 4

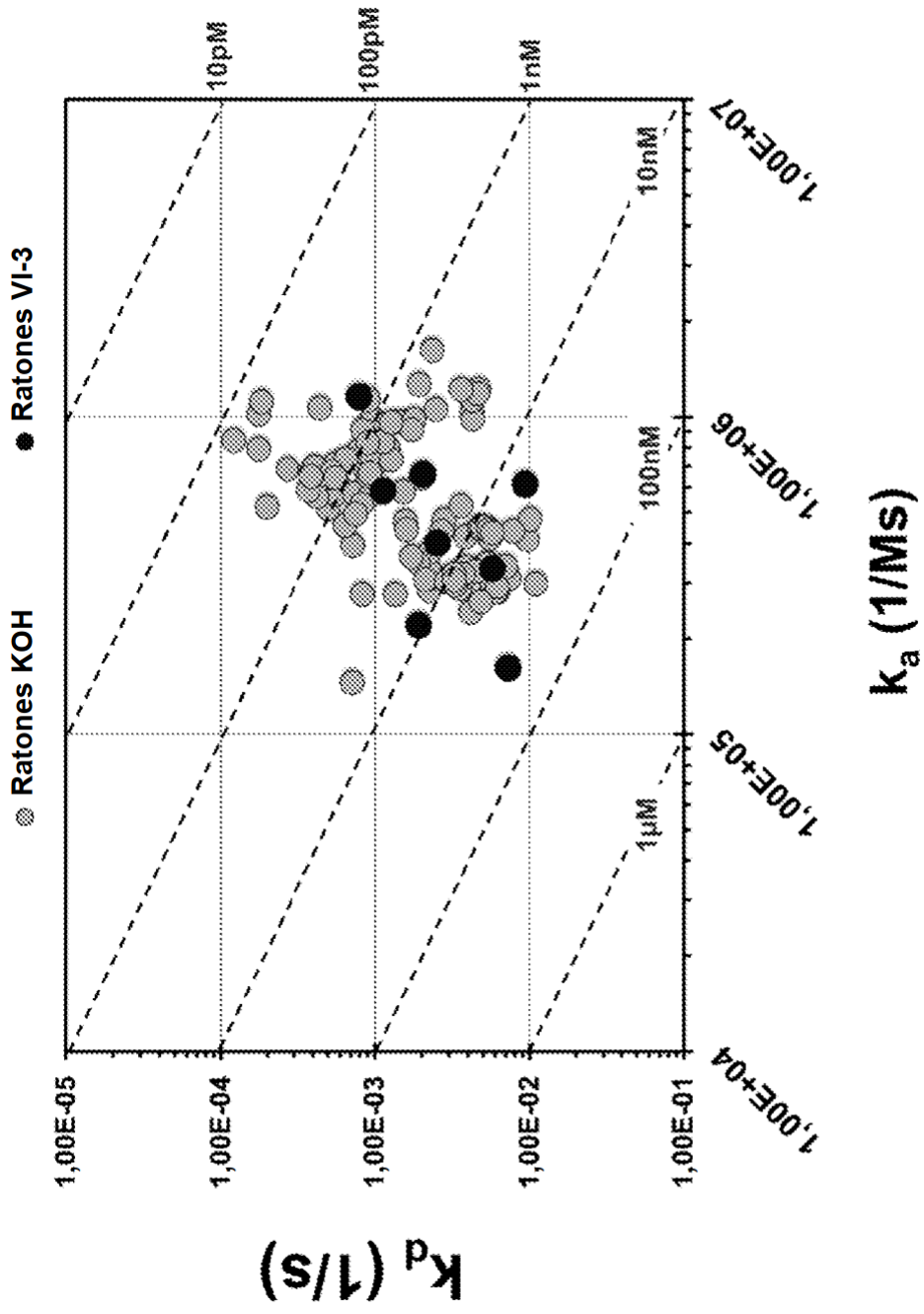


FIG. 5



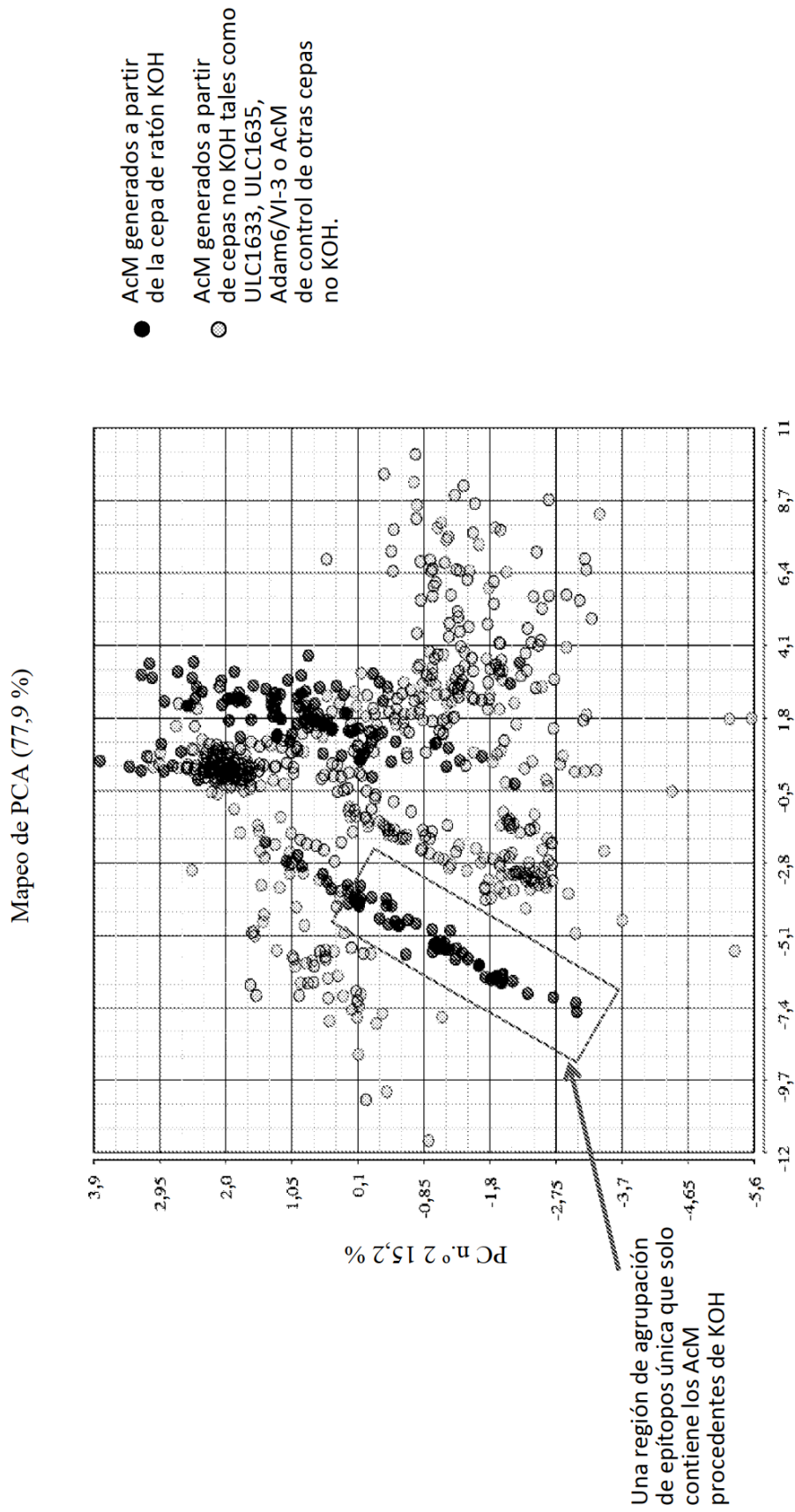


FIG. 6

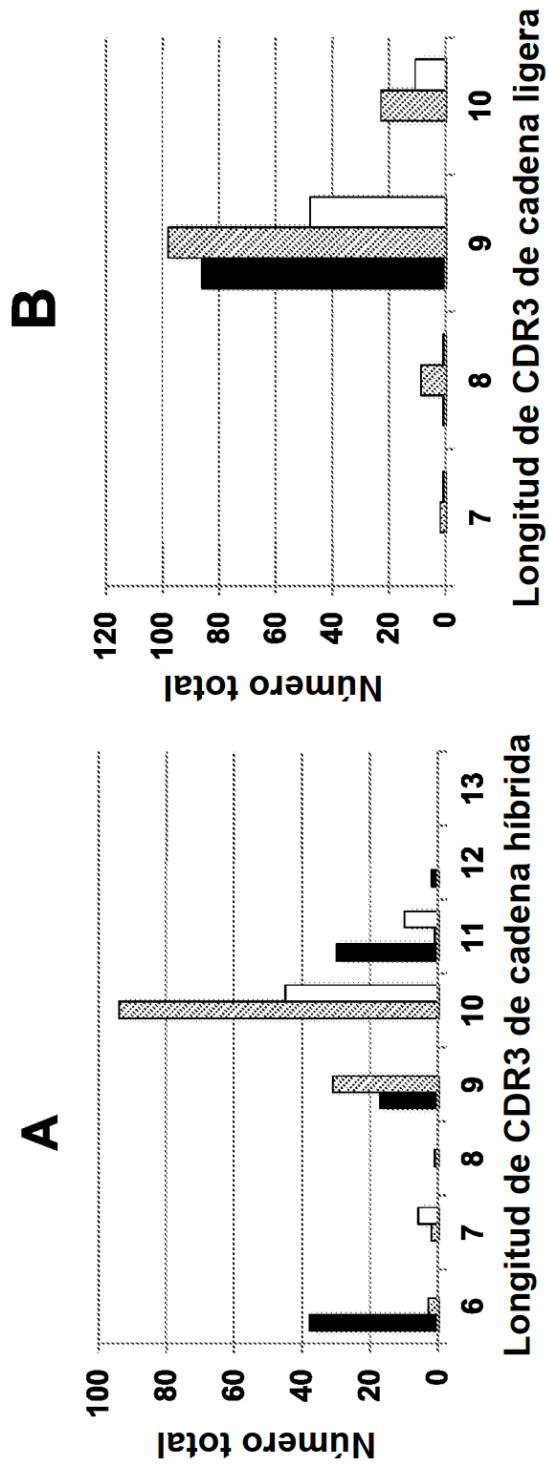


FIG. 7