

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 710 708**

51 Int. Cl.:

**A61K 38/00** (2006.01)

**C12N 9/12** (2006.01)

**C12N 9/22** (2006.01)

**C12N 15/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.09.2015 PCT/EP2015/069890**

87 Fecha y número de publicación internacional: **10.03.2016 WO16034553**

96 Fecha de presentación y número de la solicitud europea: **01.09.2015 E 15756660 (5)**

97 Fecha y número de publicación de la concesión europea: **07.11.2018 EP 3143136**

54 Título: **Recombinasa a medida bien tolerada y muy específica para recombinar los sitios diana asimétricos en una pluralidad de cepas de retrovirus**

30 Prioridad:

**02.09.2014 EP 14183277**

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

**26.04.2019**

73 Titular/es:

**HEINRICH-PETTE-INSTITUT LEIBNIZ-INSTITUT FÜR EXPERIMENTELLE VIROLOGIE-STIFTUNG BÜRGERLICHEN RECHTS - (50.0%)  
Martinistrasse 52  
20251 Hamburg, DE y  
TECHNISCHE UNIVERSITÄT DRESDEN (50.0%)**

72 Inventor/es:

**HAUBER, JOACHIM;  
CHEMNITZ, JAN;  
BUCHHOLZ, FRANK y  
KARPINSKI, JANET**

74 Agente/Representante:

**SALVÀ FERRER, Joan**

ES 2 710 708 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

Recombinasa a medida bien tolerada y muy específica para recombinar los sitios diana asimétricos en una pluralidad de cepas de retrovirus

5 La presente invención se refiere a un procedimiento para preparar un vector de expresión que codifica una recombinasa a medida ("tailored") bien tolerada y muy específica, cuya recombinasa a medidas capaz de recombinar secuencias diana asimétricas dentro de la repetición terminal larga (LTR) del ADN proviral de una pluralidad de cepas de retrovirus que puede insertarse en el genoma de una célula huésped, en el que el retrovirus es VIH-1 y la secuencia diana asimétrica tiene la secuencia expuesta como la SEQ ID NO: 1, así como al vector de expresión obtenido, las células transfectadas con ésta, la recombinasa expresada y composiciones farmacéuticas que comprenden el vector de expresión, células y/o recombinasa. Las composiciones farmacéuticas son útiles, por ejemplo, en el tratamiento y/o prevención de la infección por retrovirus, en particular, la infección por VIH. En particular, la invención se refiere a recombinasas a medida bien toleradas y muy específicas capaces de combinar secuencias diana asimétricos en más del 90% de cepas VIH-1, escindiendo de este modo las secuencias de VIH-1, y vectores de expresión que las codifican.

Las infecciones retrovirales, tales como, por ejemplo, infecciones por el virus de la inmunodeficiencia humana (VIH) son todavía una de las enfermedades humanas más importantes y más generalizadas.

Un enfoque para el tratamiento de retrovirus, por ejemplo, VIH, es reconocer el provirus insertado en el genoma de la célula huésped. La escisión del ADN proviral del genoma del huésped, por ejemplo, evitaría la replicación adicional del VIH y se diferencia de las metodologías actuales en que tiene el potencial para erradicar incluso el virus latente presente en el genoma del huésped.

Una clase de proteínas que se consideraron para su uso en este enfoque alternativo son recombinasas específicas de sitio (FLOWERS et al, 1997). Las recombinasas específicas de sitio median una multitud de funciones en la naturaleza desde el reordenamiento de genes a la segregación del genoma, tales como por ejemplo la escisión, inversión o integración de las unidades de ADN definidas (revisado en STARK et al, 1992).

Una de las recombinasas más simples y mejor entendidas es la recombinasa Cre del bacteriófago P1 que resuelve dímeros del genoma en monómeros por recombinación entre dos sitios de ADN bicatenario idénticos, es decir, simétricos, de una secuencia particular (HOESS Y ABREMSKI, 1985). La recombinasa Cre ha encontrado un amplio uso en la genética del ratón (NAGY, 2000). Cre es una proteína de 38 kDa que lleva el nombre de su función, ya que causa la recombinación (STERNBERG & HAMILTON, 1981). El prerrequisito para esta recombinación es la alineación de dos sitios de recombinación reconocidos por Cre en orientación antiparalela que a continuación son unidos por cuatro subunidades Cre idénticas que se unen para formar un anillo en el que cada subunidad pone en contacto dos subunidades adyacentes y medio sitio de un sitio de recombinación (HOESS Y ABREMSKI, 1985). El sitio de recombinación reconocido por Cre es una secuencia de DNA bicatenario de 34 pb conocida como loxP (de locus of crossing over ( $\chi$ ), P1; STERNBERG & HAMILTON, 1981), que es palindrómica con la excepción de sus ocho pares de bases más interiores (referidas como el espaciador), que imparten direccionalidad al sitio.

Algunos sistemas de recombinación específicos de sitio, incluyendo el sistema Cre/loxP, funcionan sin proteínas accesorias o cofactores y funcionan bajo una amplia variedad de condiciones celulares. Sin embargo, dado que las recombinasas específicas de sitio funcionan a través de interacciones específicas de las subunidades de la enzima recombinasa con sus secuencias diana de ADN afines, el uso de estas enzimas está restringido por el requisito de que las regiones de ADN específicas deben contener adecuadamente sitios diana posicionados (LEWANDOSKI, 2001). Hasta la fecha, no se ha identificado recombinasa de tipo salvaje que reconozca secuencias retrovirales nativos como sus secuencias diana de ADN.

En los últimos años, se han llevado a cabo extensos análisis de mutacionales y estructurales de recombinasas específicas de sitio para alterar sus propiedades y para lograr un mejor entendimiento de los mecanismos intrínsecos de estas enzimas (para una revisión véase VAN DUYN, 2001; y COATES et al, 2005). Una gran cantidad de estudios se centraron en la recombinasa Cre para explorar su capacidad de evolución. Varios estudios demostraron que la especificidad de diana Cre podía alterarse cuando se cambiaron algunos nucleótidos en su sitio de reconocimiento loxP (BUCHHOLZ & STEWART, 2001; SANTORO Y SCHULTZ, 2002; RUFER Y SAUER, 2002). Otros estudios abordan la ingeniería de sitios diana loxP mutados que contienen secuencias de la LTR del VIH-1 para desarrollar posibles sitios diana para el uso de Cre como estrategia antiviral (LEE & PARK, 1998; Lee et al, 2000).

El procedimiento de evolución dirigida es un poderoso procedimiento para seleccionar enzimas con especificidades alteradas (revisado en Yuan et al, 2005; y JOHANNES Y ZHAO, 2006). En el principio se utilizó este procedimiento para aislar enzimas mejoradas sobre la base de ARN mediante la selección de moléculas de ARN con los sitios de sustrato alterados. El uso de procedimientos basados en la PCR permite el cribado de bibliotecas muy grandes y la recuperación de las regiones codificantes satisfactorias de un grupo de candidatos. En la evolución dirigida de proteínas, por el contrario, el cribado y la recuperación de los mutantes mejorados, que se identifican por las

alteraciones en las propiedades de la proteína, requiere un procedimiento para recuperar la secuencia de ácido nucleico que codifica la proteína. El enlace entre la proteína y su secuencia de codificación a menudo se ha mantenido por compartimentación. En consecuencia, el cribado de la biblioteca en la evolución dirigida de proteínas se ha limitado a enfoques "uno por uno" que mantienen los compartimentos y hay disponibilidad de ventajas asociadas con el cribado de grupos de candidatos.

Esta limitación ha sido superada por el desarrollo de procedimientos que permiten la reticulación de proteínas a sus respectivas moléculas de ARN mensajero (ARNm) utilizando fusiones de ARNm-proteína y expresión en ribosomas. Los cribados funcionales para la mejora de propiedades de la proteína se acoplaron así a la recuperación directa de moléculas de codificación correspondientes y se han cribado grandes grupos *in vitro* (véase, por ejemplo BUCHHOLZ et al, 1998). Se logró una mejora adicional de la evolución dirigida de proteínas mediante la llamada evolución de proteínas ligada a sustrato (SLIPE; BUCHHOLZ & STEWART, 2001), en la que el sustrato de la recombinasa se colocó en la misma molécula de ADN que la región de codificación de la proteína. De esta manera, cuando la recombinasa se expresó en un compartimiento, su acción alteró el sustrato de ADN junto a su propia región codificante. En consecuencia, se pudo cribar una biblioteca como un grupo mediante PCR para amplificar solo las regiones codificantes candidatas que estaban al lado de un sustrato alterado. Esto permite el cribado de grandes bibliotecas convenientemente para una rápida recuperación de regiones codificantes satisfactorias. Este procedimiento se aplicó para alterar la especificidad de ADN de la recombinasa Cre y adaptarla a un nuevo sitio diana de reconocimiento (BUCHHOLZ & Stewart, 2001).

En vista del potencial de recombinasas específicas de sitio y la necesidad de encontrar una terapia contra el SIDA erradicar el VIH-1 provirus del genoma de una célula huésped, el documento WO 2008/083931 da a conocer la generación de una recombinasa a medida (TRE) que es capaz de recombinar asimétrica sitios diana dentro de la LTR de ADN proviral de un retrovirus insertan en el genoma de una célula huésped, escindiendo de este modo el provirus del genoma de la célula huésped. La recombinasa ingeniería se describe en los ejemplos, Tre, reconoce un sitio asimétrica específica presente en un particular el VIH-1 cepa. El sitio diana asimétrica tiene una cierta homología con el sitio loxP simétrica reconocido por Cre. WO 2008/083931 apreciará que, debido a la alta variabilidad de secuencia de los retrovirus, en particular, el VIH, para el tratamiento de un paciente con una cepa diferente del VIH, una recombinasa a medida diferente podría tener que ser adaptado, o una colección de recombinasas preparada con un contenido adaptado recombinasas específica para una variedad de secuencias diana.

Por el contrario, el documento WO 2011/147590 A2 proporciona una recombinasa a medida capaz de escindir una pluralidad de retrovirus, por ejemplo, cepas de VIH. Por lo tanto, la recombinasa generada se puede emplear para una pluralidad de infecciones por VIH, sin generación de una nueva recombinasa para cada cepa. Los inventores encontraron que a pesar de la alta variabilidad de secuencia de los retrovirus, usando un enfoque innovador, era posible identificar secuencias diana asimétricas presentes en una alta proporción de los virus de un subtipo particular. Sorprendentemente, fue posible identificar una secuencia diana presente en el 96% de las cepas de subtipo B de VIH-1, es decir, las cepas prevalentes en Europa y América (SEQ ID NO: 1). También se identificó una secuencia diana adicional presente en un porcentaje menor de cepas de VIH-1 (SEQ ID NO: 2). Usando Cre (SEQ ID NO: 6) como base para la evolución molecular dirigida, también identificaron varias recombinasas a medida capaces de recombinar dichas secuencias diana asimétricos y proporcionaron secuencias de consenso de estas recombinasas a medida, por ejemplo, SEQ ID NO: 7 o Tre 3.0 (SEQ ID NO: 8, capaz de recombinar la SEQ ID NO: 1).

A la luz de esto, los inventores abordaron el problema de proporcionar una recombinasa a medida mejorada capaz de recombinar secuencias diana asimétricas presentes en una pluralidad de cepas de VIH-1. Los inventores sorprendentemente han encontrado que recombinasas a medida que tienen una secuencia que difiere de las secuencias de consenso SEQ ID NO: 8, como se enseña en el documento WO 2011/147590 A2, también son muy activas en la secuencia diana asimétrica SEQ ID NO: 1 presente en el 96% de todas las cepas de subtipo B de VIH-1, es decir, las cepas prevalentes en Europa y América, y han mejorado las características. Las recombinasas a medida de acuerdo con la invención comprenden preferiblemente la secuencia de aminoácidos de consenso de SEQ ID NO: 10 o cualquiera de las SEQ ID NO: 11-13. Las recombinasas a medida de la presente invención han mejorado la especificidad y por lo tanto son mejor toleradas por los seres humanos, en particular, en células T humanas, que las recombinasas a medida de acuerdo con el estado de la técnica. Las recombinasas son preferiblemente muy específicas, ya que no tienen ninguna actividad residual detectable en secuencias diana conocidas de la recombinasa de la que fueron desarrolladas, por ejemplo, en loxP (SEQ ID NO: 4), loxH (SEQ ID NO: 5) o, también en Tre 1.0 de loxLTR (SEQ ID NO: 3).

La presente invención proporciona por primera vez un procedimiento para generar un vector de expresión que codifica una recombinasa a medida bien tolerada y muy específica capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus de una especie insertada en el genoma de una célula huésped, en el que el retrovirus es VIH-1. Las recombinasas se han adaptado para reconocer sitios diana asimétricos diferentes de sus sitios diana simétricos nativos, que pueden estar presentes en una pluralidad de cepas de retrovirus, mediante la división del sustrato en una serie de nuevos subconjuntos con diferencias más pequeñas de la diana original y la adaptación de recombinasas por etapas para reconocer estos

subconjuntos (WO 2008/083931 y WO 2011/147590). Un enfoque combinatorio permite la selección de moléculas funcionales que reconocen el sitio diana asimétrico dentro de una secuencia determinada. Por lo tanto, cruzando a través de intermedios de sustrato durante la evolución molecular dirigida, ha sido posible producir enzimas con nuevas especificidades remotas de diana asimétrica. Este enfoque también se emplea por la presente invención. La presente invención suma a los procedimientos enseñados por los documentos WO 2008/083931 y WO 2011/147590, ya que introduce etapas de seleccionar recombinasas a medida bien toleradas por las células humanas, en particular, células T humanas.

Específicamente, la invención proporciona un procedimiento para preparar un vector de expresión que codifica una recombinasa a medida bien tolerado, cuya recombinasa a medida es capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus que puede insertarse en el genoma de una célula huésped, en el que el retrovirus es VIH-1, que comprende las etapas de identificar en la secuencia de la LTR de ADN proviral de una pluralidad de secuencias de cepas de retrovirus con una homología de al menos el 30% con la secuencia de hemisitio izquierdo y la secuencia de hemisitio derecho de al menos un sitio diana para recombinasa conocido, en el que las secuencias homólogas están separadas por un espaciador de 5-12 nucleótidos, y en el que la secuencia diana asimétrica tiene la secuencia expuesta como SEQ ID NO: 1, y se encuentra en una pluralidad de cepas de retrovirus; y generar, a través de etapas repetidas de

i) evolución molecular dirigida en al menos una recombinasa que reconoce el sitio diana homólogo conocido utilizando como sustrato secuencias diana modificadas basadas en la secuencia de la secuencia diana asimétrica, pero modificada para contener sólo un número limitado de variaciones de la secuencia diana conocida; en que, en cada ronda, la secuencia diana puede variar de la secuencia diana en que se conoce que la recombinasa actúa en uno, dos o tres nucleótidos; y

ii) barajar las bibliotecas de recombinasas para obtener bibliotecas de recombinasas capaces de recombinar secuencias diana más homólogas a la secuencia diana asimétrica;

hasta que se obtiene al menos una recombinasa que es activa en la secuencia diana asimétrica dentro de la LTR del ADN de retrovirus;

etapas repetidas de selección negativa contra la recombinación del sitio diana conocido mediante evolución molecular dirigida y barajado de las bibliotecas, es decir, se lleva a cabo selección negativa para la recombinación de loxP y loxH;

seleccionar dicha recombinasa o recombinasas a medida mediante la expresión de la biblioteca en células humanas, en particular, células T humanas, y cultivar dichas células humanas que expresan la recombinasa o recombinasas a medida durante al menos 1 semana, preferiblemente, al menos 2 semanas, y aislar el ácido o ácidos nucleicos de la recombinasa o recombinasas de las células cultivadas que expresan el marcador seleccionable;

y, opcionalmente, clonar el ácido nucleico que codifica la recombinasa o recombinasas en un vector de expresión adecuado.

La presente invención en particular proporciona un procedimiento para preparar un ácido nucleico o un vector de expresión que codifica una recombinasa a medida bien tolerada y muy específica, cuya recombinasa a medida es capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus, en el que el retrovirus es VIH-1, que comprende las etapas de

(a) identificar secuencias con una homología de al menos el 30% con la secuencia de hemisitio izquierdo y la secuencia de hemisitio derecho de al menos un sitio diana de recombinasa conocido en la secuencia de la LTR de ADN proviral de una pluralidad de cepas de retrovirus, en el que las secuencias homólogas están separadas por un espaciador de 5-12 nucleótidos, y en el que la secuencia diana asimétrica tiene la secuencia expuesta como la SEQ ID NO: 1, y se encuentra en una pluralidad de cepas de retrovirus;

(b) identificar dos secuencias, en el que la primera secuencia corresponde a la secuencia de la secuencia diana asimétrica de la etapa (a) homóloga al hemisitio izquierdo de dicho sitio diana conocido y se denomina "secuencia de hemisitio 1", y en el que la segunda secuencia corresponde a la secuencia de la secuencia diana asimétrica de la etapa (a) homóloga al hemisitio derecho y se denomina como "secuencia de hemisitio 2";

(c) determinar los nucleótidos dentro de las secuencias de la etapa (b) que se desvían de las secuencias de hemisitio izquierdo y de hemisitio derecho homólogas correspondientes de dicho al menos un sitio diana homólogo conocido de la etapa (a);

(d) generar un primer subconjunto de dos ácidos nucleicos diana que comprenden secuencias diana, en el que la primera secuencia diana se designa subsitio 1 y comprende, adyacentes entre sí y en orden 5' a 3', la secuencia de hemisitio 1 de la etapa (b), la secuencia espaciadora de la secuencia diana asimétrica y una repetición invertida de la secuencia de hemisitio 1, y en el que la segunda secuencia diana se designa subsitio 2 y comprende, adyacentes entre sí y en orden 5' a 3', una repetición invertida de la secuencia de hemisitio 2, la secuencia espaciadora de la secuencia diana asimétrica y la secuencia de hemisitio 2 de la etapa (b);

(e) generar un segundo subconjunto de los ácidos nucleicos diana que comprende secuencias diana modificadas en base a las secuencias diana en el primer subconjunto de la etapa (d), en el que, en las secuencias basadas en el subsitio 1, en la secuencia de hemisitio izquierdo, una parte de los nucleótidos que se desvían de la correspondiente secuencia de hemisitio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontrados en dicho sitio diana conocido, hasta que dicha secuencia de hemisitio contiene uno, dos o tres nucleótidos que se desvían de dicho sitio diana conocido, en el que el hemisitio derecho de dicha secuencia diana modificada está formado por una repetición invertida de dicha secuencia de hemisitio izquierdo modificada, que está separada de dicha secuencia de hemisitio izquierdo modificada por la secuencia espaciadora

de la secuencia diana asimétrica, y

en el que, en secuencias basadas en el subsitio 2, en la secuencia de hemisitio derecho, una parte de los nucleótidos que se desvían de la correspondiente secuencia de hemisitio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontrados en dicho sitio diana conocido, hasta

5 que dicha secuencia de hemisitio contiene uno, dos o tres nucleótidos que se desvían de dicho sitio diana conocido, en el que el hemisitio izquierdo de dicha secuencia diana modificada está formada por una repetición invertida de dicha secuencia de hemisitio derecho modificada, que está separada de dicha secuencia de hemisitio derecho modificada por la secuencia espaciadora de la secuencia diana asimétrica,

10 de manera que en todas las secuencias de hemisitio modificadas procedentes de una secuencia diana del primer subconjunto de la etapa (d) tomadas en conjunto, se pueden encontrar todos los nucleótidos que se desvían, mientras que ninguna de dichas secuencias de hemisitio modificadas solas comprende todos los nucleótidos que se desvían,

(f) aplicar por separado la evolución molecular dirigida en al menos una recombinasa que reconoce un sitio diana homólogo conocido de acuerdo con la etapa (a) usando cada ácido nucleico del segundo subconjunto obtenido en la

15 etapa (e) como sustrato;

(g) barajar las bibliotecas de recombinasas desarrolladas en la etapa (f), en el que todas las bibliotecas de recombinasas desarrolladas en secuencias basadas en el subsitio 1 se combinan y se barajan, y en el que todas las bibliotecas de recombinasas desarrolladas en secuencias basadas en el subsitio 2 se combinan y se barajan;

(h) aplicar la evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en las bibliotecas barajadas obtenidas en la etapa (g) usando cada ácido nucleico del subconjunto según la etapa (d) como sustrato;

(i) barajar las bibliotecas de recombinasas desarrolladas en la etapa (h);

(j) aplicar evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en la biblioteca barajada obtenida en la etapa (g) usando un ácido nucleico que comprende la secuencia diana asimétrica de la

25 etapa (a) como sustrato, hasta que se obtiene al menos una recombinasa que es activa en la secuencia diana asimétrica dentro de la LTR del ADN del retrovirus de la etapa (a);

(k) aislar el ácido nucleico que codifica dicha al menos una recombinasa obtenida en la etapa (j) de la biblioteca y clonarlo en un vector de evolución que permite la selección negativa de recombinasas a medida que recombinan el sitio diana conocido de acuerdo con la etapa (a), en el que se lleva a cabo la selección negativa para la

30 recombinación de loxP y loxH, obteniendo de este modo una biblioteca;

(l) aplicar la evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en la biblioteca obtenida en la etapa (k);

(m) barajar las bibliotecas obtenidas en la etapa (l);

(n) aislar el ácido nucleico que codifica dicha al menos una recombinasa a medida obtenida en la etapa (m) y clonarlo en un vector para la expresión de la recombinasa codificada y un marcador seleccionable en una célula humana, obteniendo de este modo una biblioteca de vectores,

(o) transformar células humanas, preferiblemente células T humanas, con dicha biblioteca de vectores obtenida en la etapa (n);

(p) cultivar las células que expresan dicho marcador seleccionable durante al menos 1 semana y seleccionar el marcador seleccionable con una alta expresión;

(q) aislar el ácido o ácidos nucleicos que codifican la recombinasa de las células que expresan dicho marcador seleccionable obtenido en la etapa (p);

(r) seleccionar un ácido nucleico que codifica una recombinasa capaz de la recombinación de la secuencia diana asimétrica de la etapa (a);

(s) aislar el ácido nucleico que codifica dicha al menos una recombinasa obtenida en la etapa (r) de la biblioteca; y,

(t) opcionalmente, clonar el ácido nucleico obtenido en la etapa (s) en un vector de expresión adecuado.

En la etapa (a) del procedimiento de la presente invención, se puede determinar la secuencia de la LTR del ADN proviral, tal como por ejemplo mediante secuenciación de ADN utilizando inhibidores de terminación de cadena (Sanger et al, 1977). Sin embargo, si ya se ha determinado la secuencia de la LTR del ADN retroviral insertado en el genoma del huésped, la secuencia puede determinarse por referencia a una base de datos. En base a la información de la secuencia, se realiza un análisis por ordenador de la información de la secuencia para identificar en la misma secuencias con una homología de al menos 30% con las secuencias de hemisitio izquierdo y hemisitio derecho de sitios diana conocidos, respectivamente, de recombinasas conocidas que están separadas por un espaciador adecuado de 5-12 nucleótidos, en que la secuencia diana asimétrica se encuentra en una pluralidad de cepas del retrovirus. Preferiblemente, la homología con las secuencias del hemisitio izquierdo y el hemisitio derecho de sitios diana conocidos es de al menos 40% o al menos 50%. Preferiblemente, estas cepas de retrovirus son de una especie o un subtipo de la misma. Preferiblemente, una pluralidad de cepas comprende más de 10 cepas, más preferiblemente, más de 100 cepas, más de 130 cepas, más de 200 cepas o más de 300 cepas, es decir, cepas de VIH-1. Las cepas pueden ser de un subtipo del virus, es decir, VIH-1, VIH-1 subtipo A, B y C, preferiblemente, el VIH-1 subtipo B. Por lo tanto, la recombinasa obtenida o el vector de expresión que codifica la misma pueden utilizarse para tratamiento de la infección con una pluralidad de cepas, por ejemplo, más de 50%, más del 70%, más del 80%, más del 90% o todas las cepas conocidas de un retrovirus o subtipo de las mismas.

El término "recombinasa", tal como se utiliza en el presente documento, se refiere a una proteína implicada en la recombinación. Como tales, las recombinasas reconocen y se unen a dos secuencias específicas de ADN

- denominadas "sitios de recombinación" o "sitios diana" y median en la recombinación entre estos dos sitios diana. Por consiguiente, el término "recombinasa" pretende referirse a cualquier componente de proteína de cualquier sistema recombinante que media en reordenamientos de ADN en un locus de ADN específico. Las recombinasas naturales reconocen sitios diana simétricos que consisten en dos secuencias idénticas denominadas "hemisito" de aproximadamente 9-20 pb que forman una repetición invertida, en la que las secuencias de hemisito están separadas por una secuencia espaciadora de 5-12 pb. Las recombinasas de la familia de tirosina integrasa se caracterizan por tener una tirosina como el nucleófilo del sitio activo que se utiliza para la escisión de ADN, mientras que recombinasas de la familia de serina integrasa utilizan una serina en lugar de una tirosina.
- En una realización de la presente invención, dicha al menos una recombinasa conocida, cuya secuencia diana se utiliza en la etapa (a) y sobre la cual se aplica la evolución molecular dirigida en las etapas (h) y (j), pertenece a la familia de serina integrasas. Las recombinasas preferidas que pertenecen a la familia de serina integrasas se seleccionan del grupo que consiste en intergrasa phiC31 (COMBES et al., 2002), cualquier componente de los sistemas de recombinación Gin o Hin, resolvasa Tn3 (KRASNOW Y COZZARELLI, 1983) o cualquier otro miembro de las grandes recombinasas de serina, Rag1, Rag2 o cualquier otro componente del sistema de recombinación VDJ o variantes del mismo.
- En otra realización, dicha recombinasa pertenece a la familia de las tirosina integrasas. Las recombinasas preferidas que pertenecen a la familia de tirosina integrasas se seleccionan del grupo que consiste en Cre del fago PI (ABREMSKI et al, 1983, 1984), la recombinasa FLP de levadura (VOLERT y BROACH, 1986), Dre del fago D6 (SAUER & MCDERMOTT, 2004), recombinasa R de plásmido pSR1 de *Zygosaccharomyces rouxii*, una recombinasa del plásmido PKDI de *Kluyveromyces drosophilarius*, una recombinasa del plásmido PKW1 de *Kluyveromyces waltii*, Tnpl del transposón Tn4430 de Bacillus, cualquier componente del sistema de recombinación  $\lambda$  Int o variantes de los mismos. Preferiblemente, dicha recombinasa es recombinasa Cre o una variante de la misma.
- El término variante en este contexto se refiere a las proteínas que derivan de las proteínas anteriores por delección, sustitución y/o adición de aminoácidos y que retienen algunas o todas de la función inherente a la proteína de la que deriva.
- En una realización preferida, la recombinasa conocida es una recombinasa quimérica obtenida mediante, por ejemplo, "barajado de familia" como se describe por CRAMERI et al. (1998). El requisito para el empleo del barajado de familia es una homología significativa entre las recombinasas utilizadas para la generación de las recombinasas quiméricas. Un ejemplo para una recombinasa quimérica que se puede utilizar en la presente invención es una recombinasa quimérica que consiste en secuencias de recombinasa Cre y de Dre recombinasa, respectivamente.
- En una realización más preferida, la recombinasa es la recombinasa Cre que reconoce un sitio diana simétrico de 34 pb conocido como loxP. El sitio loxP (y también otros sitios de recombinación de recombinasas de tipo salvaje) es palindrómico con dos repeticiones de 13 pb separadas por los ocho pares de bases más interna, que representan el llamado espaciador, que imparte direccionalidad al sitio. La recombinación tiene lugar mediante la escisión dentro de la secuencia espaciadora. Dependiendo de la ubicación relativa y la orientación de los dos sitios loxP participantes, Cre cataliza la integración, la escisión o el reordenamiento del ADN (HOESS Y ABREMSKI, 1985).
- Una recombinasa útil es ZRE aislada de *Salmonella enterica*, o variantes, fragmentos y homólogos de la misma, por ejemplo, que tienen una homología de al menos aproximadamente 70%, al menos aproximadamente 80%, al menos aproximadamente 90% o al menos aproximadamente 95% con la secuencia de tipo salvaje, y que tiene una función recombinasa. Las Zre recombinasas recombinan el ADN en sitios zox. Pueden ser utilizadas para iniciar el procedimiento de la invención ya sean solas o en el contexto de una biblioteca.
- En la realización más preferida, se utiliza una biblioteca de recombinasa como punto de partida para la evolución molecular, por ejemplo, una biblioteca de recombinasas que comprende recombinasas diferentes de tipo salvaje y/o adaptadas/barajadas, por ejemplo, tal como se describe, por ejemplo, en el Ejemplo 2 del documento WO 2011/147590 A2. Dicha biblioteca se usa preferiblemente como punto de partida utilizado para la generación de las recombinasas a medida capaces de reconocer SEQ ID NO: 1.
- La recombinasa a medida obtenida por el procedimiento de la presente invención es capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus, en el que el retrovirus es VIH-1. El ADN proviral reconocido por la recombinasa puede ser insertado en el genoma de una célula huésped. Alternativamente, la recombinasa a medida de la invención puede recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus que no está (todavía) integrado en el genoma de una célula huésped, es decir, que está presente como un complejo de preintegración no integrado (PIC), en el que el retrovirus es VIH-1. Por lo tanto, el VIH que aún no se ha integrado en el genoma de la célula huésped, así como el VIH que ya se ha integrado, pueden ser inactivados por la recombinasa a medida de la invención.
- Cabe indicar que en la presente invención y también en la técnica los términos "secuencia diana", "sitio diana" y

"sitio de recombinación" se usan indistintamente.

Contrariamente a las recombinasas naturales que reconocen sitios diana simétricos, el procedimiento de la presente invención proporciona recombinasas a medida que reconocen sitios diana que no consisten en secuencias palindrómicas separadas por un espaciador. En cambio, en los sitios diana asimétricos las secuencias no forman una repetición invertida simétrica. Por consiguiente, una recombinasa a medida capaz de reconocer un sitio diana asimétrico debe reconocer y recombinar sitios diana que consisten en hemisitios de secuencia variable.

Dentro de un sitio diana asimétrico, las secuencias que se refieren como "hemisito izquierdo" y "hemisito derecho", respectivamente, están definidos por su homología al hemisito izquierdo y derecho de un sitio diana conocido. La secuencia situada entre las secuencias homólogas al hemisito izquierdo y derecho de un sitio diana conocido se denomina espaciadora.

Sin embargo, si las secuencias se encuentran en la LTR que sólo tienen homología con la secuencia del hemisito izquierdo o derecho de un sitio diana conocido, estas secuencias podrían utilizarse, sin embargo, en la práctica de la presente invención. El tamaño del sitio diana que pertenece a la recombinasa, cuya secuencia diana nativa muestra homología con secuencias dentro de la LTR, es conocido por la persona experta en la materia. Por ejemplo, si la homología se encuentra dentro de la secuencia de LTR con respecto a una secuencia diana reconocida por la recombinasa Cre, un sitio diana asimétrico a reconocer por la recombinasa Cre debe consistir en 34 nucleótidos con dos secuencias de hemisito de 13 nucleótidos cada una separada por un espaciador de 8 nucleótidos. Por consiguiente, la secuencia homóloga dentro de la LTR se define como el hemisito izquierdo o derecho o el espaciador del sitio diana asimétrico que depende de la homología con la secuencia del sitio diana conocido. Por lo tanto, las secuencias con homología con el hemisito izquierdo de una secuencia diana conocida se definen como hemisito izquierdo, las secuencias con homología al hemisito derecho de una secuencia diana conocida se definen como hemisito derecho. A partir de esta definición, las otras partes de los sitios diana asimétricos se definen bajo la consideración de la estructura del sitio diana conocido. Por lo tanto, habiendo definido, por ejemplo, una secuencia de hemisito derecho dentro de la LTR sobre la homología con un sitio loxP (reconocido por recombinasa Cre), las otras secuencias correspondientes al espaciador y el hemisito izquierdo de la secuencia diana asimétrica se pueden definir fácilmente. La secuencia de espaciador se define, por ejemplo, contando 8 nucleótidos en dirección 5' del extremo 5' de la secuencia definida como secuencia de hemisito derecho, mientras que la secuencia de hemisito izquierdo se define de forma similar contando 13 nucleótidos en dirección 5' del extremo 5' de la secuencia espaciadora definida previamente.

Homología, en este contexto, así como en toda la solitud, significa identidad de secuencia. Una comparación preferida para los propósitos de homología es comparar al menos dos secuencias utilizando técnicas estándar conocidas en el sector, incluyendo, pero no limitado a, el algoritmo de homología local de SMITH Y WATERMAN (1981), el algoritmo de alineación por homología de NEEDLEMAN Y WUNSCH (1970), o el procedimiento de búsqueda de similitud de PEARSON Y LIPMAN (1988). Para los fines de la presente solicitud, la homología de secuencia se determina preferiblemente utilizando el programa informático ClustalW disponible en el Instituto Europeo de Bioinformática (EBI), a menos que se indique lo contrario.

En vista de la exigencia de dos sitios diana idénticos que deben estar presentes en el genoma de los provirus para permitir a la recombinasa escindir la secuencia entre estos dos sitios diana, las secuencias del ADN proviral se rastrean en la etapa (a) del procedimiento de la presente invención que están presentes al menos dos veces en el genoma. Tales secuencias son, por ejemplo, las secuencias LTR del ADN proviral. Por consiguiente, la secuencia de la LTR se rastrea preferiblemente, ya que los 5'-LTR y 3'-LTR del ADN proviral son idénticos. Un sitio diana asimétrico presente en 5'-LTR también está presente en 3'-LTR y por lo tanto permite la escisión del ADN proviral situado entre las LTR.

De las secuencias identificadas dentro de la secuencia LTR con suficiente homología con los sitios diana conocidos, las secuencias se eligen preferiblemente de manera que tengan la más alta homología con la secuencia del sitio diana de recombinasas conocidas. Sin embargo, también es posible seleccionar secuencias distintas de las de la homología más alta, por ejemplo, aquellas que están presentes en el mayor número de cepas de retrovirus, o en las cepas de retrovirus de interés, por ejemplo, si un paciente está infectado con una cepa particular.

Cabe indicar que el potencial del procedimiento de la presente invención permite incluso la adaptación de recombinasas que reconocen sitios diana asimétricos con menos del 30% de homología con sitios diana conocidos, por ejemplo, al menos 11% o al menos 20% de homología. Sin embargo, para asegurar la presencia de actividad de recombinación residual para el respectivo sitio diana asimétrico o subsitios para el mismo, se rastrea preferiblemente para secuencias que tienen una homología de al menos el 30% con las secuencias de hemisito izquierdo y hemisito derecho de sitios diana conocidos de recombinasas conocidas. En realizaciones preferidas adicionales, se seleccionan las secuencias diana asimétricas que tiene una homología de al menos 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80%, más preferiblemente 85%), particularmente preferiblemente 90% y lo más preferiblemente 95% con las secuencias de hemisito izquierdo y hemisito derecho de sitios diana conocidos de recombinasas conocidas.

En una realización de la presente invención, la secuencia seleccionada dentro de la LTR tiene homología con los sitios diana loxP simétricos reconocidos por la recombinasa Cre específica de sitio.

5 En una realización preferida, se utiliza una biblioteca de recombinasas como punto de partida para la evolución molecular, por ejemplo, una biblioteca de recombinasas que comprende diferentes recombinasas de tipo salvaje y/o adaptadas/barajadas, tal como la biblioteca descrita en el Ejemplo 2 del documento WO 2011/147590 A2. Una biblioteca de ejemplo comprende Cre y recombinasas derivadas de la misma. También puede comprender Tre, Dre, recombinasas de Salmonella y Shewanella y/o recombinasas derivadas de las mismas. La biblioteca puede comprender, por ejemplo, Cre, Dre, Dre "Cre-ed", Shewanella recombinasa (Shew), Shew "Cre-ed", y/o Zre, como se describe en el documento WO 2011/147590 A2. Tre es una recombinasa a medida como se describe en el documento WO 2008/083931, que también se denomina en adelante como Tre 1.0.

15 En una realización, todas las recombinasas en la biblioteca reconocen una secuencia diana con la misma longitud de espaciador. La longitud total de las secuencias de hemisio 1 y 2, incluyendo espaciador, es preferiblemente de 34 nucleótidos.

Si dicha al menos una recombinasa es una biblioteca de recombinasas, la homología es homología con respecto a un grupo de sitios diana de recombinasa conocidos (es decir, la homología en una posición determinada a al menos una de una secuencia diana se define como homología). En consecuencia, en la etapa (c), sólo los nucleótidos que no se corresponden a un nucleótido en al menos una de las secuencias diana conocidas se definen como nucleótidos que se desvían. En el caso de una biblioteca de recombinasas, un "nucleótido nativo" en la etapa (e) puede ser un nucleótido presente en esa posición en cualquiera de las secuencias diana conocidas, preferiblemente, es un nucleótido presente en esa posición en varias o la mayoría de las secuencias diana conocidas.

25 Para identificar las secuencias diana presentes en una pluralidad de cepas de retrovirus, los sitios de reconocimiento conocidos de recombinasas, que han sido descritos en la literatura, se pueden utilizar como un objetivo de búsqueda secuencias diana asimétricas conservadas contra un tramo genómico. Dada la naturaleza repetitiva de las regiones, sin embargo, el uso de herramientas de búsqueda estándar de similitud de secuencia está impedido. Sarkar et al., 2007, utilizaron BLAST (Altschul et al., 1997) para encontrar un sitio de unión de tipo lox a través de cepas de VIH. La búsqueda BLAST para el sitio de tipo lox cuando se realizó a través de secuencias LTR de VIH-1 dio lugar al descubrimiento de sólo un sitio presente en una sola cepa. BLAST no funciona bien con tales secuencias redundantes cortas, y programas alternativos, tales como HMMER (EDDY et al, 1998), RepeatMasker o el programa de palíndromo del paquete Emboss tampoco resultaron ser adecuados. Con un programa específico usando un algoritmo basado en una matriz de peso de la posición para las regiones flanqueantes basadas en un sitio de reconocimiento conocido de una recombinasa, y usando operaciones binarias en las secuencias después de transformarse en cadenas de bits, se identificaron secuencias diana asimétricas halladas en una pluralidad de cepas de retrovirus se identificaron WO 2011/147590 A2.

40 Para el VIH-1, se determinaron secuencias diana asimétricas adecuadas, que tienen una secuencia expuesta como SEC ID NO: 1 o SEQ ID NO: 2 a continuación. Esto hace posible generar recombinasas, que son a la práctica útiles como agentes terapéuticos contra genomas retrovirales en un número significativo de pacientes, ya que estas recombinasas se dirigen a sitios de reconocimiento presentes a través de tantas cepas del retrovirus como sea posible. Para las realizaciones de la presente invención, la secuencia del sitio diana asimétrico tiene la secuencia expuesta como la SEQ ID NO: 1.

Las secuencias del hemisio izquierdo y el hemisio derecho de la SEQ ID NO: 1 y 2 están subrayadas y el espaciador se muestra en negrita:

50 SEQ ID NO: 1 AACCCACTGCTTA**AGCCTCAATAAAGCTTGCCTT**  
SEQ ID NO: 2 CTGGGCGGGACTG**GGGAGTGGCGAGCCCTCAGAT**

SEQ ID NO: 1 está presente en el 96% de las cepas de VIH-1 de subtipo B buscadas (1024/1067), en el 92% de las cepas de VIH-1 de subtipo C buscadas (624/679) y en el 82% de las cepas de VIH-1 de subtipo A buscadas (71/87). SEQ ID NO: 2 es idéntica en un menor porcentaje de cepas de subtipo B y C.

55 SEQ ID NO: 1 tiene un 54% de homología con un grupo de sitios diana de recombinasa conocidos, y SEQ ID NO: 2 tiene un 42% de homología con el grupo de estas secuencias (con respecto a los hemisios izquierdo y derecho, respectivamente). La homología a los sitios diana conocidos individuales es inferior, por ejemplo, al menos 30% para la SEQ ID NO: 1 y al menos 11% para la SEQ ID NO: 2. En particular en el caso de baja homología individual a sitios diana conocidos, puede ser ventajoso utilizar una biblioteca de recombinasas como material de partida, por ejemplo, para la generación de una recombinasa a medida capaces de recombinar SEQ ID NO: 1 o SEQ ID NO: 2, comprendiendo una biblioteca Cre, Fre, Dre, Zre y Tre.

65 En la etapa (b) del procedimiento de la presente invención, la secuencia del sitio diana asimétrico dentro de la LTR de los provirus que es homóloga al hemisio izquierdo del sitio diana conocido se define como "secuencia de hemisio 1". La secuencia del sitio diana asimétrico dentro de la LTR de los provirus que es homóloga al hemisio

derecho del sitio diana conocido se define como secuencia de hemisio 2. La secuencia entre las secuencias que representan el hemisio izquierdo y derecho se conoce como el espaciador.

En la etapa (c), los nucleótidos dentro de "secuencia de hemisio 1" y la "secuencia de hemisio 2", respectivamente, de las secuencias de la etapa (b) que se desvían de las correspondientes secuencias homólogas de hemisio izquierdo y hemisio derecho de la diana conocida se determinan mediante la alineación de secuencias y la comparación de secuencias. En este contexto, la secuencia de la "secuencia de hemisio 1" se compara con el hemisio nativo correspondiente, que es preferiblemente la secuencia del hemisio izquierdo, mientras que la secuencia de "secuencia de hemisio 2" se compara con el otro hemisio que forma el sitio diana nativo palindrómico, que es preferiblemente la secuencia de hemisio derecho.

La Fig. 1 del documento WO 2011/147590 A2 muestra el resultado de esta comparación para las SEQ ID NO: 1 y 2, en comparación con una biblioteca de recombinasas. Los nucleótidos que se desvían se muestran ante un fondo oscuro.

Esta comparación no debe necesariamente llevarse a cabo después de la etapa (b) y antes de la etapa (d) del procedimiento de la invención, pero también se puede realizar en una fase diferente del procedimiento después de la etapa (a) y antes de la etapa (e).

En la etapa (d), se genera un primer subconjunto de dos ácidos nucleicos diana que comprende secuencias diana, en el que la primera secuencia diana se designa subsitio 1 y comprende, adyacentes entre sí y en orden 5' a 3', secuencia de hemisio 1 de la etapa (b), la secuencia espaciadora de la secuencia diana asimétrica y una repetición invertida de la secuencia de hemisio 1, y en donde la segunda secuencia diana se designa subsitio 2 y comprende, adyacentes entre sí y en orden 5' a 3', una repetición invertida de la secuencia de hemisio 2, la secuencia espaciadora de la secuencia diana asimétrica y la secuencia de hemisio 2 de la etapa (b). Las secuencias diana del primer subconjunto son secuencias de oligonucleótidos palindrómicas que tienen la estructura de un sitio diana simétrico. Estos sitios diana simétricos artificiales son sintetizados sobre la base de las secuencias de hemisio de la etapa (b) mediante el complemento de la secuencia de hemisio carente en cada secuencia de oligonucleótidos como repetición invertida, en la que la secuencia de la "secuencia de hemisio 1" y "secuencia de hemisio 2", respectivamente, se utiliza para complementar la segunda secuencia de hemisio en el extremo opuesto de la secuencia espaciadora. Por consiguiente, la primera secuencia diana en el primer subconjunto (referido como "subsitio 1") comprende una repetición invertida que consiste en la "secuencia de hemisio 1" y la "secuencia de hemisio" inversamente repetida separadas por la secuencia espaciadora, mientras que la segunda secuencia diana en el primer subconjunto (referido como "subsitio 2") comprende una repetición invertida que consiste en la "secuencia de hemisio 2" inversamente repetida y la "secuencia de hemisio 2" separada por la secuencia espaciadora. En el "subsitio 1" las secuencias están dispuestas de la siguiente manera: 5' - "secuencia de hemisio 1" - espaciador- "repetición invertida de la secuencia de hemisio 1" -3', en el "subsitio 2" la secuencia está dispuesta de la siguiente manera: 5'- "repetición invertida de la secuencia de hemisio 2" - espaciador- "secuencia de hemisio 2" - 3'.

Las secuencias espaciadoras dentro de cada dos secuencias diana sintéticas del primer subconjunto son preferiblemente idénticas y corresponden a la secuencia de la LTR que representa o que se define como la secuencia espaciadora del sitio diana asimétrico. Sin embargo, en una realización adicional, las secuencias espaciadoras pueden comprender una o dos desviaciones de la secuencia procedentes de sustituciones de nucleótidos.

Generalmente, esta etapa representa un primera división de las secuencias del sitio diana asimétrico seleccionado para la adaptación de una recombinasa específica (ver Figura 1 del documento WO 2008/083931, que se incorpora completamente aquí por referencia, y la Fig. 2 del documento WO 2011/147590 A2, que también se incorpora completamente aquí por referencia). Se generan secuencias en esta etapa que albergan sitios diana simétricos derivados de los hemisios del sitio diana asimétrico seleccionado para la adaptación de una recombinasa específica. Como consecuencia, cada mutación (es decir, la diferencia con el sitio o sitios diana reconocidos por la recombinada o recombinasas (de tipo salvaje)) presente en un hemisio de dicho sitio diana asimétrico ahora se ha extendido entre las secuencias diana simétricas en el primer subconjunto.

En la etapa (e) del procedimiento de la invención, se genera un segundo subconjunto de los ácidos nucleicos diana que comprenden secuencias diana modificadas en base a las secuencias diana en el primer subconjunto de la etapa (d). En secuencias basadas en el subsitio 1, en la secuencia de hemisio izquierdo, una parte de los nucleótidos que se desvían de la correspondiente secuencia de hemisio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontradas en dicho sitio diana conocido, hasta que dicha secuencia de hemisio contiene uno, dos o tres (preferentemente, dos) nucleótidos que se desvían de dicho sitio diana conocido, en la que el hemisio derecho de dicha secuencia diana modificada está formado por una repetición invertida de dicha secuencia de hemisio izquierdo modificada, que está separada de dicha secuencia de hemisio izquierdo modificada por la secuencia espaciadora de la secuencia diana asimétrica.

En secuencias basadas en el subsitio 2, en la secuencia de hemisio derecho, una parte de los nucleótidos que se

desvían de la correspondiente secuencia de hemisio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontrados en dicho sitio diana conocido, hasta que dicha secuencia de hemisio contiene uno, dos o tres (preferentemente, dos) nucleótidos que se desvían de dicho sitio diana conocido, en el que dicho hemisio izquierdo de dicha secuencia diana modificada está formado por una repetición invertida de dicha secuencia de hemisio derecho conocido, que se separa de dicho modificada secuencia media de sitio junto a la secuencia espaciadora de la secuencia diana asimétrica.

Por ejemplo, si un subsitio comprende seis nucleótidos que se desvían, tales como los dos subsitios basados en SEQ ID NO: 1 o subsitio 2 de la SEQ ID NO: 2 con respecto a la biblioteca de recombinasas mostrada en la Fig 1 del documento WO 2011/147590 A2, se pueden generar tres secuencias diana modificadas basándose en el subsitio, que contienen cada uno dos nucleótidos que se desvían (diferentes) en el hemisio izquierdo (si se basa en el subsitio 1) o hemisio derecho (si se basa en el subsitio 2). En consecuencia, en cada secuencia diana modificada, la secuencia del respectivo subsitio se modifica para corresponder a la secuencia de la secuencia diana conocida (o al menos una secuencia diana conocida) en cuatro nucleótidos (Fig. 2 del documento WO 2011/147590 A2). Por supuesto, también es posible generar seis secuencias diana modificadas que contienen cada una uno de los nucleótidos que se desvían, o dos secuencias diana que contienen cada una tres de los nucleótidos que se desvían.

En otro ejemplo, si un subsitio comprende nueve nucleótidos que se desvían, tal como subsitio 1 de la SEQ ID NO: 2 con respecto a la biblioteca de recombinasas que se muestra en la Fig 1 del documento WO 2011/147590 A2, pueden generarse tres secuencias diana modificadas en base al subsitio, que contienen cada una tres nucleótidos que se desvían (diferentes) en el hemisio.

Como consecuencia, en todas las secuencias de hemisio modificadas originarias de una secuencia diana del primer subconjunto de la etapa (d) tomadas juntos, se pueden encontrar todos los nucleótidos que se desvían, mientras que ninguna de dichas secuencias de hemisio modificadas solas comprende todos los nucleótidos que se desvían.

Una vez más, se genera una repetición invertida en base a la secuencia de hemisio modificada, de tal manera que la secuencia espaciadora que separa las dos secuencias que forman la repetición invertida (véase la Fig. 2 del documento WO 2011/147590 A2). Las secuencias espaciadoras dentro de cada secuencia diana modificada de un nuevo subconjunto que derivan de una secuencia diana de un subconjunto superior son preferiblemente iguales y corresponden a la secuencia de la LTR que representa o se define como la secuencia espaciadora del sitio diana asimétrico. Sin embargo, en una realización adicional, las secuencias espaciadoras pueden comprender una o dos desviaciones de secuencia procedentes de sustituciones de nucleótidos. Usando este enfoque, el número de mutaciones (es decir, las diferencias con respecto al sitio diana reconocido por la recombinasa de tipo salvaje) en las secuencias diana que representan a cada subconjunto es menor que en la secuencia diana asimétrica de partida, pero todas las mutaciones todavía están representadas en una de las secuencias diana (ver Figura 1 del documento WO 2008/083931, Fig. 2 del documento WO 2011/147590 A2).

El término "nucleótidos que se desvían" tal como se utiliza en el presente documento, se refiere a un nucleótido dentro de la secuencia diana asimétrica identificada o definida dentro de la LTR o dentro de una secuencia diana de un subconjunto generado de acuerdo con la presente invención que se desvía (es decir, es diferente) del nucleótido presente en la misma posición en la secuencia homóloga correspondiente de la secuencia diana simétrica homóloga conocida de una recombinasa conocida elegida en la etapa (a) del procedimiento de la presente invención. En este contexto, los términos "nucleótidos que se desvían" y "mutaciones" se utilizan indistintamente.

El documento WO 2008/083931 enseña que las recombinasas pueden adaptarse utilizando la evolución molecular dirigida usando secuencias diana como sustrato, si la secuencia diana usada como sustrato difiere en no más de 3 nucleótidos de la secuencia diana nativa. Por lo tanto, la generación de subconjuntos de diferentes órdenes descritos anteriormente sirve para reducir el número de nucleótidos que se desvían por secuencia diana a 3 o menos (véase la figura 1 del documento WO 2008/083931). La reducción gradual del número de nucleótidos que se desvían finalmente produce un número de subconjuntos de secuencias diana de diferentes órdenes con un número decreciente de nucleótidos que se desvían hasta que se crea un subconjunto final que se puede utilizar como sustrato para la evolución molecular dirigida. Durante la creación de los diferentes subconjuntos y la reducción de este modo del número de nucleótidos que se desvían, las diferencias con respecto al sitio diana reconocido por la recombinasa de tipo salvaje se extienden entre varias secuencias diana que no comprenden más de 3 de estos nucleótidos que se desvían cada una, mientras que las secuencias diana del orden final como un todo aún representan todos los nucleótidos que se desvían.

Opcionalmente, en el procedimiento de la invención, se pueden generar más subconjuntos de secuencias diana a partir de las secuencias diana del segundo subconjunto mediante la repetición por etapas del proceso de la etapa (e), es decir, la división de las secuencias diana en las respectivas secuencias de hemisio y la generación de nuevas estructuras palindrómicas sobre la base de estas secuencias de hemisio después de la alteración de la secuencia de hemisio derivada de una secuencia diana del segundo subconjunto, generando cada vez un nuevo subconjunto de secuencias diana, en el que las secuencias de hemisio utilizadas para generar las repeticiones invertidas contienen menos nucleótidos que se desvían de la correspondiente secuencia de hemisio homóloga de dicho al menos un sitio diana conocido. Estas secuencias diana adicionales se pueden utilizar para etapas

adicionales de la evolución molecular dirigida y el barajado de las bibliotecas de recombinasas. Por supuesto, dicha etapa adicional puede también ser realizada solamente para algunas de las secuencias, por ejemplo, para las secuencias en las que se obtienen recombinasas con una baja eficiencia de recombinación. Si se generan subconjuntos adicionales y se desarrollan recombinasas sobre estos, la biblioteca desarrollada de recombinasas se utiliza en la etapa (f) del procedimiento de la invención.

A partir del segundo subconjunto de secuencias diana obtenido en la etapa (e), se puede generar un tercer subconjunto, seguido por un cuarto, quinto, sexto, etc. subconjunto, si es necesario. Sin embargo, la generación del tercer subconjunto generalmente sólo es necesario, si las secuencias diana del segundo subconjunto todavía contienen más de tres nucleótidos que se desvían. Lo mismo se aplica a la generación de los próximos subconjuntos, que sólo son necesarios, si las secuencias diana del subconjunto anterior todavía contienen más de tres nucleótidos que se desvían. Cabe señalar que en una realización, se generarán subconjuntos de secuencias diana hasta que las secuencias diana del subconjunto final sólo comprenden un nucleótido que se desvía. Por consiguiente, en función del número de nucleótidos que se desvían en cada secuencia de hemisitio, puede diferir el número de subconjuntos generados para cada secuencia de hemisitio del sitio diana asimétrico. Por ejemplo, puede ser necesario generar sólo dos subconjuntos para la secuencia de hemisitio izquierdo, mientras que deben generarse tres o cuatro subconjuntos para el hemisitio derecho a efectos de propagar los nucleótidos que se desvían entre varias secuencias diana de manera que una única secuencia diana no comprende más de 3 de estos nucleótidos que se desvían.

El principio de la generación de más subconjuntos de las secuencias diana para reducir el número de nucleótidos que se desvían a los números por debajo de tres se ilustra en la Figura 1 del documento WO 2008/083931, y la Fig. 2 del documento WO 2011/147590 A2 proporciona ejemplos específicos de secuencias diana modificadas.

En la etapa (f), se aplica un procedimiento de evolución molecular dirigida en dicha al menos una recombinasa que reconoce un sitio diana homólogo conocido de la etapa (a), usando una secuencia diana del subconjunto final o segunda obtenido en la etapa (e) que contiene uno, dos o tres nucleótidos que se desvían de la secuencia de demisitio homóloga correspondiente de dicho sitio diana homólogo conocido como sustrato.

El término "subconjunto final", tal como se usa en el presente documento, se refiere al último subconjunto generado en la etapa (e), es decir, si no se generan subconjuntos adicionales, en el segundo subconjunto. Dependiendo del número de nucleótidos que se desvían en el sitio diana asimétrico y el número de subconjuntos que tenían que ser generados para reducir el número de nucleótidos que se desvían por secuencia diana por debajo de 3, el "subconjunto final" puede corresponder a cualquier subconjunto, por ejemplo, el segundo, tercero, cuarto o un subconjunto posterior, y puede ser diferente para las secuencias de hemisitio de la secuencia diana asimétrica dentro de la LTR. Si previamente se han desarrollado recombinasas en subconjuntos adicionales de secuencias diana modificadas que tienen menos nucleótidos que se desvían de la correspondiente secuencia de hemisitio homóloga de dicho sitio diana homólogo conocido, se usa la recombinasa obtenida en esa etapa.

Por supuesto, es posible iniciar el proceso de la invención con una recombinasa específica para una secuencia diana modificada específica y con otra recombinasa (o una biblioteca) para otra secuencia diana modificada específica. Los procedimientos de evolución molecular dirigida, también conocidos como la evolución en laboratorio o la evolución *in vitro*, son conocidos en la técnica (para una revisión ver YUAN et al, 2005 y referencias allí citadas; JOHANNES y ZHAO, 2006).

En una primera etapa de evolución molecular dirigida, las bibliotecas de secuencias de recombinasa mutadas al azar son generadas por procedimientos conocidos en la técnica, por ejemplo mediante el uso de PCR propensa a error y barajado de ADN (revisado en, por ejemplo YUAN et al., 2005), o los procedimientos descritos en la solicitud de patente internacional WO 2002/44409. Los plásmidos de cada biblioteca que comprenden la recombinasa mutada también contienen una de las secuencias diana del subconjunto final obtenido en la etapa (f). Después de la transfección de la biblioteca de plásmidos generada en las células apropiadas, se permite la expresión de la recombinasa está habilitado y se lleva a cabo la evolución molecular dirigida como es conocida por la persona experta en la técnica.

En una realización preferida, la evolución molecular dirigida empleada en la etapa (f) del procedimiento de la presente invención es la evolución de proteínas unida a sustrato (SLiPE; Buchholz y Stewart, 2001; Solicitud de Patente Internacional WO 02/44409). La evolución de proteínas unida a sustrato puede llevarse a cabo como se describe en detalle en los ejemplos del documento WO 2008/083931 o WO 2011/147590 A2. Brevemente, las secuencias diana obtenidas en la etapa (e) se clonan en un plásmido (el llamado vector de evolución) junto con una secuencia de codificación mutada aleatoriamente para la recombinasa. La mutación aleatoria se lleva a cabo mediante PCR propensa a error (véase BUCHHOLZ & STEWART, 2001). La biblioteca de plásmidos generada se transfecta a continuación en células de *E. coli* para permitir la expresión de la recombinasa. Mediante el uso de un promotor inducible para conducir la expresión de la recombinasa, es posible ajustar los niveles de expresión. Después de la incubación durante la noche, el ADN plásmido se aísla de las células y se digiere con NdeI para cortar los plásmidos que no se recombinaron y los plásmidos recombinados solamente son posteriormente amplificados con cebadores. El producto de PCR de la forma recombinada del plásmido produce una banda de 1,7

Kb. El producto de PCR se digiere con BsrGI y XbaI y se retrosubclona en el vector de evolución digerido de forma similar para el siguiente ciclo de evolución.

5 En la etapa (g), las bibliotecas de recombinasas desarrolladas en la etapa (f) se combinan y se barajan. La tecnología de barajado de ADN es conocida en el sector (para una revisión ver MINSHULL Y STEMMER, 1999; STEMMER, 1994). Las bibliotecas de recombinasas desarrolladas en secuencias diana modificadas basadas en el subsitio 1 se combinan y barajan, y, por separado, las bibliotecas de recombinasas desarrolladas en secuencias diana modificadas basadas en el subsitio 2 se combinan y se barajan. Las bibliotecas combinadas y barajadas a continuación se clonan en una nueva generación de vectores que comprenden las secuencias diana del siguiente subconjunto superior, es decir, si se generan dos subconjuntos, el subconjunto generado en la etapa (d). Por ejemplo, el vector para la biblioteca desarrollado en las secuencias basadas en el subsitio 1 comprende la secuencia del subsitio 1 como una secuencia diana, y el vector de la biblioteca desarrollado en las secuencias basadas en el subsitio 2 comprende la secuencia del subsitio 2 como una secuencia diana.

15 En la etapa (h), se aplica el procedimiento de evolución molecular dirigida en las bibliotecas barajadas obtenidas en la etapa (g) usando la secuencia diana del siguiente subconjunto superior, que, como se describe, puede ser el subconjunto según la etapa (d). En esta etapa, se puede utilizar el mismo procedimiento de evolución molecular dirigida que el aplicado antes en la etapa (f), pero también es posible utilizar un procedimiento diferente de evolución molecular dirigida en esta etapa del procedimiento de la presente invención. Ejemplos de diferentes procedimientos de evolución molecular dirigida se describen, por ejemplo por YUAN et al. (2005). Preferiblemente, también se aplica el procedimiento de evolución de proteínas unidas a sustrato en las bibliotecas combinadas y barajadas.

25 Este etapa produce recombinasas que reconocen y recombinan secuencias diana que albergan la combinación (y de este modo cantidades crecientes) de mutaciones de las diferentes secuencias diana del subconjunto inferior. La combinación de mutaciones de las diferentes bibliotecas de un subconjunto inferior de secuencias diana da lugar a efectos sinérgicos y conduce a la generación de recombinasas, que ahora recombinan secuencias diana de un subconjunto superior, lo que demuestra que se puede utilizar una estrategia de evolución a través de productos intermedios para lograr una actividad deseada.

30 En la etapa (i), las etapas (g), es decir, la combinación y el barajado de bibliotecas de recombinasas, y (j), es decir, la aplicación de la evolución molecular dirigida en las bibliotecas combinadas y barajadas, se repiten hasta que se consigue al menos una recombinasa que es activa en la secuencia diana asimétrica presente en la LTR del ADN proviral.

35 En un procedimiento en el que la generación de dos subconjuntos de secuencias diana era necesaria para generar secuencias diana con sólo una, dos o tres desviaciones de nucleótidos, las bibliotecas de recombinasas desarrolladas, por ejemplo, para el segundo subconjunto de secuencias diana se combinan y se barajan y se aplica una evolución molecular dirigida sobre esta biblioteca barajada utilizando las secuencias diana del primer subconjunto. En la siguiente etapa, la secuencia diana asimétrica de la etapa (a) dentro de la LTR del ADN proviral se utiliza para desarrollar la biblioteca de recombinasas que comprende recombinasas que reconocen las secuencias diana del primer subconjunto mediante evolución molecular dirigida para obtener al menos una recombinasa que es activa en la secuencia diana asimétrica dentro de la LTR del ADN retroviral. En esta etapa, el procedimiento de evolución molecular dirigida es, preferiblemente, el procedimiento de evolución de proteínas unidas a sustrato.

45 "Al menos una recombinasa" se refiere al hecho de que el procedimiento de la invención podría conducir a una o más recombinasas (individuales) que son, cada una por sí misma, activas en la recombinación de la secuencia diana asimétrica. No se pretende abarcar varias recombinasas diferentes que sólo juntas son capaces de recombinar la secuencia diana asimétrica. De hecho, el procedimiento de la invención no conduciría a la selección de recombinasas que necesitan combinarse con otras recombinasas diferentes para recombinar una secuencia diana asimétrica, ya que sólo se expresa una recombinasa por célula individual.

Las etapas del procedimiento (a) - (j) son, en esencia, conocidas en la técnica, en particular, de WO2011/147590.

55 Después de la etapa (j), la biblioteca de recombinasas a medida se selecciona negativamente para la recombinación del sitio diana simétrico conocido de acuerdo con la etapa (a), es decir, para la recombinación de loxP y loxH.

Esta selección puede realizarse mediante al menos un ciclo de una o más etapas que comprenden la evolución molecular dirigida y el barajado de la biblioteca de vectores.

60 Para este fin, el ácido nucleico que codifica dicha al menos una recombinasa a medida desarrollada en las etapas anteriores puede aislarse de los vectores utilizados en las mismas y se clonó en un vector de evolución adecuado. Dicho vector permite la selección negativa de recombinasas a medida que recombinan el sitio diana conocido de acuerdo con la etapa (a), es decir, para la recombinación de loxP (SEQ ID NO: 4) y loxH (SEQ ID NO: 5). Se obtiene de este modo una biblioteca de vectores. A continuación, se emplea la evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato (SLiPE), como es conocido por la persona experta y de

acuerdo con los principios descritos anteriormente.

Por ejemplo, el vector de la evolución puede construirse de manera que comprende tanto la secuencia diana asimétrica final (SEQ ID NO: 1) como el sitio diana conocido (loxP y/o loxH), cada uno dos veces para permitir la recombinación. La recombinación en el sitio diana conocido y después de digestión por restricción condujo a un producto lineal que no comprende dos sitios de cebadores específicos en un orden que permite la amplificación por PCR de un producto. Si no tiene lugar la recombinación en absoluto, el vector se linealiza por digestión por restricción, y no tiene lugar la amplificación por PCR. En cambio, la recombinación en el sitio diana asimétrico final (por ejemplo, SEQ ID NO: 1) corta el sitio de restricción, es decir, el vector no se linealiza por digestión por restricción, y la recombinasa a medida puede ser amplificada por PCR. La PCR, que inherentemente produce errores, se utiliza para generar variabilidad. La evolución puede ser llevada a cabo en *E. coli*.

La biblioteca obtenida después de uno o más, preferiblemente, después de aproximadamente diez ciclos de evolución molecular dirigida, puede barajarse. Pueden llevarse a cabo uno o más ciclos de evolución molecular dirigida y/o barajado.

Opcionalmente, la selección negativa para la recombinación de varios sitios diana conocidos, por ejemplo, para la recombinación de loxP y loxH puede llevarse a cabo alternativamente, por ejemplo, un ciclo de evolución con la selección negativa en loxP se pueden alternar con un ciclo de evolución con selección negativa en loxH. Por ejemplo, se pueden llevar a cabo de aproximadamente 10 a aproximadamente 30 o aproximadamente 15-20 ciclos de selección negativa, en combinación con aproximadamente dos rondas de barajado de ADN. Entre estos ciclos de evolución, se puede variar la cantidad del activador de transcripción L-arabinosa, por ejemplo, de 100 µg/ml a 1 µg/ml. Los vectores y procedimientos preferidos se muestran en la Fig. 2 y los ejemplos.

La especificidad de las recombinasas a medida sobre la secuencia diana asimétrica final y la potencial actividad residual en las secuencias diana conocidas se puede comprobar para uno o más clones después de llevar a cabo un número de ciclos de evolución.

Si la especificidad aún no es satisfactoria, deben llevarse a cabo más ciclos de evolución.

Como se muestra en la Fig. 3, la selección negativa elimina la actividad residual de recombinasas a medida, tales como las que se enseñan en el documento WO 2011/147590 A2 (TRE3) en las secuencias diana conocidas de loxP y loxH. No es detectable actividad residual de las recombinasas generadas de la invención en los sitios diana conocidos, es decir, loxP y loxH, incluso en presencia de cantidades altas (50 o 100 µg/ml) del activador de la transcripción L-arabinosa, es decir, en presencia de altas cantidades de recombinasa. Esto muestra que las recombinasas a medida obtenidas son muy específicas para su secuencia diana, en este caso, para la SEQ ID NO: 1, ya que la secuencia diana asimétrica específica se recombina, pero la secuencia diana simétrica conocida no.

Esto tiene la ventaja de que, cuando se aplica la recombinasa a medida de la invención para la terapia de los seres humanos, el riesgo de reacción cruzada con y recombinación de secuencias humanas en la célula huésped es mínima. Este es un factor que contribuye a la tolerabilidad de la recombinasa a medida en células humanas. Sin embargo, no sólo las consecuencias a corto plazo de la expresión de la recombinasa están en juego aquí, sino también los aspectos de seguridad, tales como los posibles efectos oncogénicos de recombinación inespecífica, incluso a baja eficacia. La eliminación de la actividad incluso residual de la recombinasa a medida contribuye así a la seguridad y la fiabilidad de la recombinasa a medida resultante en tratamientos terapéuticos.

La selección en contra de la recombinación de las secuencias diana simétricas conocidas va seguida por la selección para una recombinasa a medida bien tolerada en células humanas.

En la etapa (n) del procedimiento de la invención, el ácido nucleico que codifica dicha al menos una recombinasa a medida obtenida en la etapa (m) se aísla y se clona en un vector para la expresión de la recombinasa codificada y un marcador seleccionable en eucariotas, preferiblemente, células humanas, obteniendo de este modo una biblioteca de vectores. El ácido nucleico puede aislarse a partir del plásmido respectivo dentro de la biblioteca usando enzimas de restricción apropiadas. Los procedimientos de digestión con endonucleasas de restricción son conocidos por el experto. El ácido nucleico que codifica la recombinasa puede entonces recuperarse mediante procedimientos conocidos, tales como, por ejemplo electroforesis en gel. Se puede clonar en un vector de expresión apropiado para la expresión en células eucariotas, por ejemplo, humanas, tal como se conoce en el estado de la técnica o se describe a continuación. Por ejemplo, se puede usar un vector retroviral, por ejemplo, lentiviral, por ejemplo, tal como se muestra en la Fig. 4A. La expresión de la recombinasa a medida codificada y el marcador seleccionable preferiblemente es constitutiva, o puede ser inducida por agentes adecuados.

El marcador seleccionable puede conferir resistencia a un antibiótico, o puede ser una proteína fluorescente, tal como la proteína verde fluorescente (GFP) o un derivado de la misma (por ejemplo, EBFB, ECFP, YFP). Las proteínas fluorescentes, tales como GFP, permiten una fácil clasificación de células dependiente de la fuerza de expresión.

En la etapa (1), las células eucariotas, preferiblemente, humanas, preferiblemente, células T humanas, se transforman con dicha biblioteca de vectores obtenida en la etapa (k). Se pueden utilizar procedimientos conocidos en el estado de la técnica. Las células transformadas son generalmente células humanas, sin embargo, si se pretende el tratamiento para un paciente no humano, es aconsejable probar la tolerabilidad en células de esa especie de paciente. Las células humanas son preferiblemente células hematopoyéticas, por ejemplo, preferiblemente, células T, en particular, células T CD4+, pero también se pueden utilizar células madre, tales como células madre CD34+. Se pueden utilizar células primarias, por ejemplo, células T primarias, preferiblemente, células T CD4+ primarias, pero también se puede utilizar una línea de células, tales como las células T Jurkat.

En la etapa (p), las células que expresan dicho marcador seleccionable se cultivan durante un periodo de tiempo suficiente para seleccionar para recombinasas TRE bien toleradas por la célula humana. La selección se basa en la suposición de que la expresión del marcador y la recombinasa a medida están correlacionadas. Las células se seleccionan para la expresión del marcador, por ejemplo, se seleccionan células GFP-positivas, preferiblemente, células con fuerte expresión de GFP. Como la expresión del marcador seleccionable y la recombinasa a medida están correlacionadas, estas células también expresarán la recombinasa a medida. Por lo tanto, las células que expresan una recombinasa a medida que es perjudicial para su supervivencia o su capacidad para la propagación, se eliminan o reducen en cantidad. Preferiblemente, las células que expresan el marcador se cultivan durante al menos 1 semana, al menos dos semanas, al menos 3 semanas o al menos 4 semanas. En consecuencia, las recombinasas a medida expresadas en las células T serán bien toleradas por las células humanas, por ejemplo, células T humanas, es decir, que no serán tóxicas para dichas células, o, preferiblemente, de otro modo tampoco serán perjudiciales para la supervivencia y la propagación de dicha células. Preferiblemente, durante ese periodo de cultivo, se seleccionan las células al menos una vez, preferiblemente, 2, 3, o 4 veces, para una expresión elevada del marcador seleccionable. Por ejemplo, con una proteína fluorescente, la selección puede realizarse mediante la clasificación de células activadas por fluorescencia. Con un gen de resistencia a antibióticos, se pueden añadir cantidades crecientes de antibióticos al medio de cultivo.

Aunque la expresión de la recombinasa Cre de tipos salvaje en células humanas se ha establecido desde hace mucho tiempo, y, en niveles de expresión razonables, ha demostrado ser problemática, la sobreexpresión de Cre puede ser tóxica (LOONSTRA et al., 2001). Los inventores encontraron que un número significativo de recombinasas a medida mutadas eran perjudiciales para la supervivencia y/o la propagación de células T humanas tras una fuerte sobreexpresión. Curiosamente, a pesar de que podría esperarse que es la especificidad relativamente baja de las recombinasas a medida y la actividad residual en los sitios diana, tales como loxP y loxH (una secuencia presente en el genoma humano) lo que conduce a la baja tolerabilidad en células humanas, la selección para una tolerabilidad en células T humanas solas, sin selección previa para una alta especificidad, no era suficiente en la eliminación de reactividad cruzada residual en loxP o loxH. De este modo, sólo la combinación de ambas etapas de selección con el procedimiento conocido anteriormente de la invención conduce a una recombinasa a medida que es bien tolerada y muy específica.

En la etapa (q), las etapas de cultivo y selección van seguidas por el aislamiento del ácido nucleico que codifica dicha al menos una recombinasa de las células que expresan dicho marcador seleccionable obtenido en la etapa (p).

La etapa (r) añade otra selección para un ácido nucleico que codifica una recombinasa capaz de recombinar la secuencia diana asimétrica de la etapa (a), preferiblemente, para la recombinación con una alta actividad. La actividad de recombinación se prueba preferiblemente en células humanas, en particular, células T CD4+ humanas, pero también se puede probar en E. coli.

En la etapa (s), el ácido nucleico de una recombinasa que tiene actividad en la secuencia diana asimétrica de la etapa (a) dentro de la LTR del ADN retroviral se aísla de la biblioteca. El ácido nucleico puede aislarse a partir del plásmido respectivo dentro de la biblioteca usando enzimas de restricción apropiadas. Los procedimientos de digestión con endonucleasas de restricción son conocidos por el experto. El ácido nucleico que codifica la recombinasa puede entonces recuperarse mediante procedimientos conocidos, tales como, por ejemplo electroforesis en gel.

El ácido nucleico puede almacenarse (preferiblemente a temperaturas inferiores a -80°C) u opcionalmente se puede clonar en la etapa (t) en un vector de expresión para su uso en un análisis posterior, en procedimientos de expresión de proteínas, o para la administración a un sujeto para tratar y/o prevenir la infección por retrovirus, en particular, infección por VIH y/o SIDA. Los vectores de expresión adecuados son conocidos en el estado de la técnica o se describen a continuación.

El desarrollo de recombinasas a medida que reconocen específicamente secuencias asimétricas, tales como SEQ ID NO: 1 dentro de una pluralidad de LTR de VIH-1, permite la escisión del provirus respectivo de su integración cromosómica para la mayoría de los sujetos infectados con VIH-1. Un vector de expresión que codifica dicha recombinasa, las células transfectadas con el mismo y/o la proteína recombinasa derivada del mismo tiene usos médicos, por ejemplo, en el tratamiento y/o prevención de una infección por VIH-1. Los procedimientos preferidos de preparación de dicha recombinasa a medida o vector de expresión que la codifica se enseñan en el documento WO 2011/147590. Los presentes inventores añadieron al procedimiento descrito en el documento WO 2011/147590, una

etapa de selección activa para alta especificidad, es decir, sin reactividad cruzada detectable en la secuencia diana conocida de la etapa (a) (o en, por ejemplo, loxP y loxH), y para la tolerabilidad de la recombinasa a medida en células humanas, tales como células T humanas. Como se ha descrito, esto mejora significativamente el uso médico de la recombinasa a medida para escindir los genomas del provirus de VIH de células T humanas.

5 El ADN proviral que puede insertarse en el genoma de una célula huésped, o que aún no puede insertarse, preferiblemente es el ADN de un retrovirus. Los retrovirus comprenden una gran y diversa familia de virus de ARN con envoltura. La característica distintiva de la familia es su estrategia de replicación que incluye como etapas esenciales la transcripción inversa del ARN viral en ADN lineal de doble cadena y la posterior integración de este ADN (ADN proviral) en el genoma de la célula huésped. Los retrovirus se subdividen en siete grupos, definidos por el parentesco evolutivo. Cinco de estos grupos (retrovirus alfa, beta, delta, épsilon y gamma) representan retrovirus con potencial oncogénico y los otros dos grupos son los lentivirus y los spumavirus. Los virus de la leucemia humana de células T patógenas humanas de tipo I y tipo II (HTLV-I y HTLV-II) pertenecen al grupo de retrovirus delta, mientras que los virus del SIDA, virus de inmunodeficiencia humana de tipo 1 y tipo 2 (VIH-1 y VIH-2) pertenecen al grupo de los lentivirus (para una revisión véase el libro de texto estándar "Retroviruses" de COFFIN JM, HUGHES SH, VAR US HE (Eds.) 1997, Cold Spring Harbor Laboratory Press, Nueva York).

En las realizaciones de la presente invención, el retrovirus es un lentivirus, es decir un virus de inmunodeficiencia humana de tipo 1 (VIH-1).

La secuencia diana asimétrica identificada en la etapa (a) del procedimiento de la presente invención se localiza tanto en 5'-LTR como 3'-LTR del provirus del VIH. Preferiblemente, dicha secuencia diana asimétrica localizada en 5'-LTR y 3'-LTR de un provirus de VIH tiene la secuencia expuesta como la SEQ ID NO: 1.

En una realización preferida, el procedimiento de evolución molecular dirigida aplicada en el procedimiento de la presente invención es el procedimiento de evolución de proteínas unidas a sustrato (SLiPE; BUCHHOLZ & STEWART, 2001; véase también WO 02/44409).

Al llevar a cabo el procedimiento de la invención, tal como se describe en el presente documento, los inventores generaron varios ácidos nucleicos que codifican una recombinasa a medida bien tolerada, y las propias recombinasas a medida. La invención proporciona así una recombinasa a medida bien tolerada que comprende una secuencia de acuerdo con cualquiera de las SEQ ID NO: 10-13 o que consiste en las mismas, o los ácidos nucleicos que las codifican.

Se encontró sorprendentemente que estas recombinasas a medida difieren de las secuencias consenso SEQ ID NO: 7 y 8 de recombinasas a medida capaces de recombinar secuencias diana asimétricas, tal como se enseña en el documento WO 2011/147590, y de todas las otras recombinasas previamente conocidas.

En particular, las recombinasas a medida bien toleradas analizadas nuevas capaces de recombinar la secuencia diana asimétrica de SEQ ID NO: 1 con alta especificidad, sorprendentemente comprenden una mutación Q89L.

Una recombinasa a medida descrita en este documento comprende una secuencia de acuerdo con la SEQ ID NO: 9 (secuencia consenso Tre 3.1 85%). La SEQ ID NO: 9 representa una secuencia de consenso, estando cada mutación presente (en comparación con la secuencia de aminoácidos de Cre) con una probabilidad del 85%. Para esta determinación, 100 clones individuales, generados mediante el procedimiento descrito en el presente documento, se analizaron por secuenciación de Sanger, así como la biblioteca completa de Tre 3.1 generada mediante secuenciación de próxima generación (33.000 lecturas de secuencias únicas de 200 pb). Los aminoácidos variables se representan por una X, que puede representar cualquier aminoácido de origen natural (cf. Fig. 1). Aproximadamente un tercio de los aminoácidos de la SEQ ID NO: 9 son altamente variables, es decir, estas posiciones no tienen que conservarse para que la recombinasa sea capaz de recombinar la secuencia diana asimétrica de SEQ ID NO: 1 y sea bien tolerada por los seres humanos. Aproximadamente dos tercios de las posiciones de aminoácidos, por otro lado, parecen importantes para recombinar la secuencia diana asimétrica de SEQ ID NO: 1 con una alta especificidad y/o para ser bien tolerada por los seres humanos.

En una realización preferida, la recombinasa a medida de la presente invención comprende una secuencia de acuerdo con cualquiera de las SEQ ID NO: 11-13, más preferiblemente, SEQ ID NO: 11. Se seleccionaron estas recombinasas a medida para su especificidad, es decir, mientras que otras recombinasas generados por el procedimiento descrito en el presente documento todavía tenían una baja actividad de recombinación, pero detectable, en loxP, loxH o lox LTR 1.0, dicha actividad de recombinación no fue detectable con las recombinasas de la SEQ ID NO: 11-13, tal como se muestra en los ejemplos. La invención proporciona así una recombinasa a medida bien tolerada y muy específica capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de VIH-1 que puede insertarse en el genoma de una célula huésped (es decir, una recombinasa a medida funcional), que comprende preferiblemente la SEQ ID NO: 11-13, preferiblemente, SEQ ID NO: 11. Dicha recombinasa a medida se puede obtener, por ejemplo, de acuerdo con el procedimiento de la invención.

En una realización, la recombinasa a medida puede comprender una secuencia de acuerdo con SEQ ID NO: 10. Esta secuencia, secuencia de consenso Tre 3,1 100% (3 clones) es una secuencia consenso de las tres recombinasas preferidas de SEQ ID NO: 11-13.

5 La recombinasa a medida también puede tener una identidad de aminoácidos de al menos 95%, preferiblemente, una identidad de aminoácidos de al menos 99% o una identidad de aminoácidos del 100% con la SEQ ID NO: 10, o puede variar de SEQ ID NO: 10 en sólo uno o dos aminoácidos, y comprende los siguientes intercambios de aminoácidos definidos en comparación con la secuencia de Cre (SEQ ID NO: 6): V7L, P12S, P15L, M30V, H40R, M44V, S51T, Y77H, K86N, Q89L, G93A, S108G, C155G, A175S, A249V, R259D, E262R, T268A, D278G, P307A, N317T, I320S. También puede comprender los intercambios: N160T, R241Q, K244I, N319E. Preferiblemente, comprende los siguientes intercambios de aminoácidos en comparación con Cre: N3I, V7L, N10S, P12S, P15L, V23A, M30V, F31L, H40R, M44V, S51T, Y77H, K86N, Q89L, G93A, S102F, S108G, N111D, K122R, A131T, S147A, D153E, C155G, N160T, F163L, I166V, I174V, A175S, V182I, G198S, D232S, R241Q, K244I, A249V, Q255R, R259D, A260V, E262R, G263K, T268A, D278G, P307A, N317T, N319E, I320S.

10 Estos intercambios específicos hacen que la enzima sea particularmente adecuada para la recombinación en una secuencia diana de SEQ ID NO: 1 o una secuencia diana que tiene una alta identidad de secuencia con la SEQ ID NO: 1 (por ejemplo, al menos 80%, al menos 90% o al menos 95% de identidad de secuencia con la SEQ ID NO: 1). Los inventores sorprendentemente pudieron demostrar que una única variación de aminoácido, a saber, Q89L, garantiza que la recombinasa a medida sea bien tolerada en células humanas, por ejemplo, células hematopoyéticas humanas o células T humanas, y tenga una alta especificidad, ya que no tiene ninguna actividad detectable en la recombinación de la secuencia diana original, loxP o loxH, y, preferiblemente, tampoco hay actividad detectable en Tre 1 de loxLTR. La actividad se puede detectar mediante electroforesis en gel de una muestra que comprende loxP (SEQ ID NO: 4), loxH (SEQ ID NO: 5), o, para la comparación, Tre 3 de loxLTR que comprende la SEQ ID NO: 1, cada uno en contacto con la recombinasa a medida, por ejemplo, por inducción de la expresión de la misma a partir de un vector adecuado, tal como se muestra en los ejemplos y en la figura 3. Como puede verse en dicha figura, la recombinasa Tre 3.0 (producida según el documento WO 2011/147590), aunque ya es bastante específica en comparación con otras recombinasas, tiene actividad residual en loxP y loxH bajo las condiciones mostradas, mientras que, con uTre, una recombinasa de la presente invención, el producto recombinado sólo puede observarse para loxLTR que comprende la SEQ ID NO: 1. Esta alta especificidad minimiza el riesgo de recombinación no deseada en el genoma humano.

15 La secuencia de la recombinasa a medida de la invención no se da a conocer en el documento WO 2008/083931 o el WO 2011/147590. En particular, la técnica anterior no enseña o sugiere que una recombinasa a medida capaz de recombinar la secuencia diana asimétrica de SEQ ID NO: 1 tenga un intercambio de aminoácidos Q89L. Por el contrario, el documento WO 2011/147590 enseña explícitamente que esta posición debe mantenerse como Q (véanse todas las secuencias específicas o secuencias de consenso de dicha publicación). La secuencia de la recombinasa a medida de la invención también varía de las recombinasas naturales, tales como Cre, Dre, Fre o Zre, lo cual es evidente a partir de la característica de que es capaz de la recombinación de secuencias diana asimétricas, preferiblemente, SEQ ID NO: 1, dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus que puede insertarse en el genoma de una célula huésped.

20 Si la recombinasa a medida capaz de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus insertado en el genoma de una célula huésped es para recombinar la secuencia diana de SEQ ID NO: 1, comprende preferiblemente la secuencia consenso Tre 3,1 de 100%, SEQ ID NO: 10, o una de las secuencias específicas SEQ ID NO: 11-13.

25 Las recombinasas a medida funcionales capaces de recombinación de secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de retrovirus que puede insertarse en el genoma de una célula huésped, que puede, por ejemplo, obtenerse mediante el procedimiento de la invención, pueden variar de dichas secuencias, pero las secuencias proporcionan una orientación valiosa para el experto para producir una recombinasa a medida capaces de recombinar caras diana asimétricas, tales como SEQ ID NO: 1, incluso sin llevar a cabo el procedimiento de la invención.

30 Preferiblemente, los intercambios de aminoácidos con respecto a la secuencia de referencia son sustituciones conservativas, que son bien conocidas por la persona experta (por ejemplo, Creighton (1984), Proteins. W.H. Freeman and Company (Ed.)). Por ejemplo, las sustituciones conservativas sustituyen un aminoácido del grupo de aminoácidos cargados negativamente por otro. Lo más preferiblemente, los intercambios conducen a uno de los aminoácidos presentes en cualquiera de las SEQ ID NO: 11-13 en la posición pertinente.

35 Las recombinasas a medida capaces de recombinar secuencias diana asimétricas dentro de la LTR de ADN proviral de una pluralidad de cepas de VIH-1 que puede insertarse en el genoma de una célula huésped también pueden comprender una combinación de 2 o más secuencias seleccionadas del grupo que consiste en SEQ ID NO: 11-13, en que una combinación de cualquiera de dichas secuencias comprende una parte C-terminal de cualquiera de estas secuencias, y la parte N-terminal de cualquiera de otras secuencias, en que la combinación comprende la SEQ ID NO: 10, por ejemplo, una parte C-terminal de cualquiera de estas secuencias, por ejemplo, SEQ ID NO: 11, y la

parte N-terminal de cualquier otra de estas secuencias, por ejemplo, SEQ ID NO: 12. La parte C-terminal puede tener una longitud de 1-342 aminoácidos. En una combinación de dos secuencias, la parte N-terminal puede tener una longitud de 1-342 aminoácidos. La recombinasa a medida también puede ser una combinación de tres o más partes derivadas de estas secuencias. La combinación comprende un motivo de consenso TRE 3.1, es decir, SEQ ID NO: 10.

La invención también proporciona un ácido nucleico que codifica una recombinasa a medida capaz de recombinar secuencias asimétricas diana tales como SEQ ID NO: 1 dentro de la LTR de ADN proviral de una pluralidad de cepas de VIH-1 que puede insertarse en el genoma de una célula huésped, comprendiendo la recombinasa a medida una secuencia de aminoácidos como se define anteriormente.

En el contexto de la invención, un ácido nucleico o una proteína que comprende una secuencia puede consistir en dicha secuencia.

Puede comprender alternativamente otras secuencias, por ejemplo, una secuencia de señal que proporciona la expresión/localización en un compartimiento celular específico, tal como una señal de localización nuclear, como en SEQ ID NO: 14 (la señal de localización nuclear está en las posiciones 2-9 de la SEQ ID NO: 14). Si una proteína se va a utilizar en una composición farmacéutica, se prefiere especialmente expresarla como una proteína de fusión con un dominio de transducción de proteína, tal como el dominio de transducción de proteína tat, que permite la transducción de la proteína de las células diana. Preferiblemente, una recombinasa a medida de la invención que se va a utilizar en una composición farmacéutica se prepara como una proteína de fusión con una secuencia de localización nuclear y con un dominio de transducción de proteína, por ejemplo, de tat, y un ácido nucleico que codifica una recombinasa a medida de la invención puede codificar dicha proteína de fusión. Por ejemplo, los siguientes dominios de transducción de proteína se pueden usar en una proteína de fusión con una recombinasa a medida de la invención, que incluye además, preferiblemente, una señal de localización nuclear:

- Dominio básico de transactivador tata de VIH-1 (Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J., Tat-mediated delivery of heterologous proteins into cells, Proc Natl Acad Sci USA. A. 18 enero 1994; 91 (2): 664-8)
- homeodominio de Drosophila Antennapedia (Antp) (Derossi D, Joliot AH, Chassaing G, Pro-chiantz A., The third hélix of the Antennapedia homeodomain translocates through biological membranes, J Biol. Chem 8 abril 1994; 269 (14): 10444-50).
- Factor de transcripción VP22 de HSV (Elliott G, O'Hare P., Intercellular trafficking and protein delivery by a herpesvirus structural protein, Cell 24 enero 1997; 88 (2): 223-33)
- Motivo de translocación permeable celular (TLM) del antígeno de superficie preS2 del virus de la hepatitis B (HBV) (Oess S, Hildt E., Novel cell permeable motif derived from the PreS2-domain of hepatitis-B virus Surface antigens, Gene Ther, 7 mayo 2000; 7 (9): 750-8).

En caso de que la proteína ha de ser purificada, también se puede añadir una etiqueta que facilita la purificación de una proteína tal como una etiqueta de His.

El uso del codón del ácido nucleico de la invención que codifica una recombinasa Tre, tal como se define anteriormente, puede ser elegido por la persona experta. Por ejemplo, un uso de codones adecuado para la expresión en una célula humana puede elegirse, en particular, si se pretende la expresión en una célula humana, por ejemplo, con fines terapéuticos. El uso de codones también se puede basar en el uso de codones de, por ejemplo, recombinasa Cre.

La recombinasa a medida o ácido nucleico que codifica dicha recombinasa a medida se puede obtener mediante el procedimiento de la invención, tal como se describe en el presente documento, o puede obtenerse mediante este procedimiento. También se puede obtener basado en las secuencias descritas en este documento, opcionalmente, mediante la combinación y/o variación adicional de estas secuencias, opcionalmente con el ensayo de la actividad en la recombinación de sitios diana asimétricos, tales como SEQ ID NO: 1.

La presente invención proporciona además una composición, por ejemplo, una biblioteca, que comprende dos o más de los ácidos nucleicos que codifican una recombinasa a medida, tal como se ha definido anteriormente, por ejemplo, que codifican dos o más recombinasas a medida que comprenden diferentes secuencias de la SEQ ID NO: 10 o cualquiera de las SEQ ID NO: 11-13. En una realización, la composición comprende ácidos nucleicos que codifican recombinasas a medida que comprenden dos o más, tres o más, cuatro o más, cinco o más, diez o más, 20 o más o 25 o más recombinasas que comprenden las secuencias de acuerdo con cualquiera de las SEQ ID NO: 10-13 o combinaciones de estas secuencias. Tales composiciones, en particular composiciones en las que el ácido nucleico es un vector de expresión, pueden ser particularmente adecuadas como composiciones farmacéuticas, tal como se describe a continuación.

En el procedimiento de la presente invención, el ácido nucleico que codifica una recombinasa a medida que está activa en la secuencia diana asimétrica dentro de la LTR del ADN retroviral se clona preferiblemente en un vector de expresión. Los vectores de expresión son construcciones genéticas para la expresión de las proteínas codificadas por los ácidos nucleicos dentro del vector. Tales vectores de expresión pueden ser vectores extracromosómicos

auto-replicantes o vectores que se integran en un genoma huésped. Generalmente, estos vectores de expresión incluyen ácido nucleico regulador de la transcripción y traducción unido operativamente al ácido nucleico que codifica la recombinasa a medida de la presente invención.

5 El término "secuencias de control" se refiere a secuencias de ADN necesarias para la expresión de una secuencia codificante unida operativamente en un organismo huésped particular. Las secuencias de control que son adecuadas para procariontes, por ejemplo, incluyen un promotor, opcionalmente una secuencia operadora, y un sitio de unión al ribosoma. Se sabe que las células eucariotas utilizan promotores, señales de poliadenilación y potenciadores.

10 Un ácido nucleico está "unido operativamente" cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, un promotor o potenciador está unido operativamente a una secuencia codificante si afecta a la transcripción de la secuencia; o un sitio de unión al ribosoma está unido operativamente a una secuencia codificante si está posicionado de manera que facilite la traducción. La unión se consigue mediante ligación en sitios de restricción convenientes. Si no existen tales sitios, se usan adaptadores o enlazadores de oligonucleótidos sintéticos de acuerdo con la práctica convencional. El ácido nucleico regulador transcripcional y traduccional será generalmente apropiado para la célula huésped usada para expresar la recombinasa a medida. Numerosos tipos de vectores de expresión apropiados, y secuencias reguladoras adecuadas son conocidos en la técnica para una variedad de células huésped.

20 El vector de expresión utilizado en la presente invención puede ser un vector retroviral, un vector lentiviral, un vector spumavirus, un vector adenoviral, o un vector de virus adeno-asociado. Sin embargo, en una realización preferida, el vector de expresión es un vector lentiviral seleccionado del grupo que consiste en vectores lentivirales derivados de VIH-1, SIV, FIV o EIAV. Los vectores lentivirales se describen por ejemplo por SCHAMBACH et al. (2006) o en la Solicitud de Patente Europea No. 1 1000751.5.

En realizaciones preferidas de la presente invención, el vector de expresión comprende un promotor celular, bacteriano, viral o híbrido.

30 En general, para el propósito de la presente invención, el promotor puede ser un promotor constitutivo o un promotor inducible. Además, los promotores pueden ser un promotor de origen natural, tal como un promotor bacteriano, celular o viral, o un promotor híbrido. Los promotores híbridos, que combinan elementos de más de un promotor, son conocidos en la técnica, y son útiles en la presente invención. Además, el promotor usado en la presente invención también puede ser un derivado de un promotor de origen natural. Un "derivado" de un promotor de origen natural como se usa en el presente documento puede ser una combinación de elementos activos en cis obtenidos a partir de promotores o secuencias de origen diferente o, alternativamente, puede obtenerse por delección o mutación de elementos activos en cis dentro de un promotor natural específico (EDELMAN et al, 2000; ALPER et al, 2006; HARTENBACH Y FUSSENEGGER, 2006).

40 En una realización de la presente invención, se selecciona o deriva el promotor constitutivo o derivado del mismo del grupo que consiste en promotores de citomegalovirus, virus del sarcoma de Rous, retrovirus relacionados con virus de leucemia murino, gen de fosfogliceroquinasa, virus que forma el foco de bazo murino o factor 1 alfa de elongación humana.

45 En una realización adicional de la presente invención, el promotor inducible o derivado del mismo se selecciona o deriva del grupo que consiste dentro de la LTR o derivados de la misma derivados de lentivirus, spumavirus y deltaretrovirus.

50 En este contexto, el término "LTR" se refiere a repeticiones terminales largas 5' y 3' de provirus que tienen la función del promotor (para una revisión ver el libro de texto "Retrovirus" (COFFIN JM, HUGHES SH, VARMUS HE (Eds.) 1997, Cold Spring Harbor Laboratory Press, Nueva York)).

Preferiblemente, el promotor inducible o derivado del mismo se selecciona o deriva de la LTR o derivados de la misma derivados de VIH-1, VIH-2, MVV, EIAV, CAEV, SIV, FIV, BIV, HTLV-I y HTLV-II.

55 La presente invención proporciona además un procedimiento para preparar una recombinasa a medida, en el que dicho procedimiento comprende el procedimiento mencionado anteriormente para la preparación de un vector de expresión que codifica una recombinasa a medida, y la etapa adicional de expresar dicha recombinasa a medida (o un polipéptido de fusión que comprende la secuencia de aminoácidos de dicha recombinasa a medida) a partir de dicho vector de expresión en una célula huésped adecuada.

60 Preferiblemente, las recombinasas finalmente obtenidas se ensayan en células de mamífero para asegurar que funcionen en un entorno de células de mamífero. Además, para obtener una buena expresión en células de mamífero las recombinasas pueden optimizarse para la expresión en estas células (por ejemplo, optimización de uso de codones usando procedimientos bien conocidos en la técnica. Véase, por ejemplo SHIMSHE et al, 2002) o se pueden añadir secuencias señal necesarias para dirigir la proteína en el núcleo de la célula de mamífero, tal como la

5 secuencia NLS (MACARA, 2001) al ácido nucleico de la recombinasa a medida. La expresión del ácido nucleico que codifica la recombinasa a medida clonado en un vector de expresión, por ejemplo, de acuerdo con la etapa (1) del procedimiento para preparar un vector de expresión que codifica una recombinasa a medida, se puede llevar a cabo usando, por ejemplo, sistemas de expresión bacterianos, de insectos o de mamíferos. Sin embargo, también se pueden emplear otros sistemas de expresión conocidos en la técnica. Los procedimientos de introducir ácido nucleico exógeno en huéspedes mamíferos, de insectos o bacterianos, así como otros huéspedes, son también bien conocidos en la técnica, y variarán con la célula huésped usada. Las técnicas incluyen transfección mediada por dextrano, precipitación con fosfato de calcio, transfección mediada por polibreno, fusión de protoplastos, electroporación, infección viral, encapsulación del polinucleótido o polinucleótidos en liposomas y microinyección directa del ADN en los núcleos.

15 Las proteínas de fusión se preparan mediante procedimientos bien conocidos en la técnica. Por ejemplo, el vector de expresión en el que el ácido nucleico que codifica la recombinasa a medida se clona ya comprende una secuencia nucleico que codifica un segundo polipéptido o proteína. Mediante la clonación del ácido nucleico que codifica la recombinasa a medida en el marco con la secuencia del segundo polipéptido o proteína, ambas secuencias se expresarán como proteína de fusión.

20 Las células huésped utilizadas para expresar la recombinasa a medida a partir del vector de expresión incluyen células procariotas, tales como, por ejemplo, células bacterianas o células de levadura, o, preferiblemente, células eucariotas, tales como por ejemplo células de insectos o mamífero, lo más preferiblemente, células humanas. La célula huésped puede ser una célula hematopoyética, por ejemplo, una célula madre hematopoyética adulta o una célula T, por ejemplo, una célula CD4+. La célula puede ser derivada de un sujeto infectado con el retrovirus y la célula se puede administrar de nuevo al tema después de la transformación, y, opcionalmente, cultivo y/o propagación.

25 La presente invención proporciona además un procedimiento para preparar una célula madre adulta transformada, en el que dicho procedimiento comprende el procedimiento mencionado anteriormente para la preparación de un vector de expresión que codifica una recombinasa a medida y la etapa adicional de introducir el vector de expresión obtenido en el procedimiento mencionado anteriormente para la preparación de un vector de expresión que codifica una recombinasa a medida *in vitro* en una célula madre adulta adecuada.

30 En un aspecto adicional, la presente invención está dirigida al ácido nucleico como se describe en el presente documento, y/o como se obtiene a partir del procedimiento de la presente invención mencionado anteriormente. Los ácidos nucleicos que codifican una recombinasa a medida definidos por una secuencia también se proporcionan en el presente documento.

35 Un "ácido nucleico", tal como se usa en el presente documento, es un compuesto polimérico comprendido de subunidades unidas covalentemente denominadas nucleótidos. El ácido nucleico incluye ácido polirribonucleico (ARN), por ejemplo, ARNm, y el ácido polidesoxirribonucleico (ADN), ambos pueden ser de cadena sencilla o de doble cadena. ADN incluye ADNc, ADN genómico, ADN sintético y ADN semi-sintético.

40 En un aspecto adicional, la presente invención también se dirige al vector de expresión como se puede obtener a partir del procedimiento de la presente invención anteriormente mencionado, y a un vector de expresión que comprende el ácido nucleico que codifica una recombinasa a medida, tal como se define en el presente documento.

45 El término "proteína" tal como se utiliza en este documento incluye proteínas, polipéptidos y péptidos. Como se enenderá por los expertos en la técnica, las secuencias de ácido nucleico de la invención se pueden utilizar para generar secuencias de proteínas. Un aspecto adicional de la invención es la proteína recombinasa a medida tal como se puede obtener, por ejemplo, a partir del procedimiento de la presente invención mencionado anteriormente, cuya recombinasa puede ser opcionalmente una proteína de fusión que comprende una recombinasa funcional. En una realización, la proteína recombinasa a medida se puede preparar como un polipéptido de fusión usando técnicas bien conocidas en el sector. En una realización preferida, la proteína recombinasa a medida está unida a un segundo polipéptido. Preferiblemente, el polipéptido de fusión se obtiene a partir del procedimiento de la presente invención mencionado anteriormente, en el que la recombinasa a medida se une a un segundo polipéptido.

50 En una realización, la proteína recombinasa a medida se prepara como un polipéptido de fusión para aumentar la expresión. En una realización adicional, la proteína recombinasa a medida se produce como un polipéptido de fusión para permitir la introducción del polipéptido en células vivas. Típicamente, las proteínas purificadas no pueden entrar en las células, debido a que no son capaces de pasar la membrana celular debido a su tamaño. Sin embargo, la fusión de secuencias de péptidos específicas a las proteínas puede dar lugar a la absorción de estas proteínas de fusión en las células. En la célula, la proteína puede entonces realizar su función. Las recombinasas específicas del sitio, incluyendo la recombinasa Cre, se han liberado con éxito en células con este enfoque (PEITZ et al., 2002). Las recombinasas permeables a las células se han descrito adicionalmente por NOLDEN et al. (2006) y LIN et al. (2004). Por lo tanto, esta estrategia puede usarse para suministrar las recombinasas a medida en células para eliminar el provirus de las células infectadas. Por lo tanto, el segundo polipéptido en el polipéptido de fusión puede comprender un péptido señal. El péptido señal puede ser un dominio de transducción de proteína, tal como el

péptido TAT o un péptido de la tercera hélice del homeodominio de Antennapedia (DEROSSI y otros, 1994, 1996; VIVES y otros, 1997; VIVES, 2003; RICHARD et al, 2005) o la NLS (secuencia de localización del núcleo) para suministrar el polipéptido de fusión en el núcleo de una célula eucariota (MACARA, 2001).

5 Un aspecto adicional de la presente invención se dirige a la célula madre adulta tal como se puede obtener a partir del procedimiento mencionado anteriormente para la preparación de una célula madre adulta transformada de la presente invención. Las células madre están infectadas o transfectadas preferiblemente con el vector de expresión de acuerdo con la presente invención. En una realización preferida, la célula madre adulta es una célula madre del linaje hematopoyético que expresa la recombinasa a medida, el polipéptido de fusión mencionado anteriormente o  
10 que comprende el vector de expresión antes mencionado. Las células madre hematopoyéticas (HSC) son células CD34+ derivadas de médula ósea, que, por ejemplo, pueden purificarse a partir de sangre periférica movilizada en G-CSF de donantes (por ejemplo, pacientes infectados por VIH) mediante leucoféresis de rutina (SCHERR y EDER, 2002). Las células modificadas genéticamente *in vitro* pueden entonces formularse para la reinfusión en los pacientes.

15 En el estado de la técnica, el término "células madre" designa células que (a) tienen la capacidad de auto-renovación y (b) la capacidad para formar al menos uno y a menudo un número de tipos de células especializadas debido a su capacidad de división asimétrica (DONOVAN Y GEARHART, 2001). Las células madre adultas pueden aislarse de diferentes tejidos de adultos, es decir, de individuos diferenciados. Tales células madre se conocen en el estado de la técnica como "células madre adultas multipotentes". La diferencia esencial entre las células madre pluripotentes embrionarias y células madre adultas multipotentes radica en el número de tejidos diferenciados que se pueden obtener a partir de las respectivas células.

20 En una realización adicional, el vector de expresión de la presente invención se utiliza para la transformación de células T, por ejemplo, células primarias CD4+ (células sanguíneas) de pacientes infectados por retrovirus (por ejemplo, VIH).

25 Alternativamente, la recombinasa a medida de la invención se puede formular para la administración por partículas similares a virus (VLP). Las VLP pueden usarse para empaquetar ARNm de Tre, proteína Tre, por ejemplo, proteína de fusión, o ADN, por ejemplo, plásmidos de ADN que expresan Tre, o una construcción que comprende Promotor-Tre ADNc-sitio polyA. En consecuencia, el ácido nucleico de la invención puede contener además una señal de empaquetamiento.

30 En un etapa adicional del procedimiento de la presente invención, el ácido nucleico de la invención, el vector de expresión que comprende la secuencia de ácido nucleico que codifica una recombinasa a medida de la invención, la proteína de recombinasa, la proteína de fusión o la célula madre adulta obtenida mediante los procedimientos de la presente invención se formulan como una composición farmacéutica para usar en la prevención y/o tratamiento de una infección por retrovirus y/o para la reducción de la carga viral en un sujeto infectado por un retrovirus, es decir, el VIH, en particular, VIH-1. Un objeto adicional de la presente invención es la composición farmacéutica obtenida  
35 mediante el procedimiento antes mencionado. La composición farmacéutica está preferiblemente presente en forma de una solución adecuada para la aplicación intravenosa (infusión).

40 La preparación farmacéutica puede comprender además uno o más portadores, excipientes y/o adyuvantes farmacéuticamente aceptables. Los portadores, excipientes y adyuvantes adecuados para su uso en una composición farmacéutica son conocidos en la técnica.

45 La composición farmacéutica de la presente invención reduce preferiblemente la carga de virus en un sujeto infectado por un retrovirus por debajo de 5.000 equivalentes de genoma/ml de plasma, preferiblemente por debajo de 500 equivalentes de genoma/ml de plasma y más preferiblemente por debajo de 50 equivalentes de genoma/ml de plasma cuando se administra al sujeto. De este modo, la composición farmacéutica de la presente invención que comprende un vector de expresión que codifica una recombinasa a medida (o la recombinasa a medida como un polipéptido o proteína de fusión o una célula madre que comprende el vector de expresión) es capaz de reducir la carga de virus en un sujeto infectado con un retrovirus mediante la erradicación de la reserva genética de los retrovirus dentro de las células huésped, evitando de este modo más ciclos de vida del virus.

50 El término "carga de virus", como se usa en el presente documento, se refiere, por ejemplo, a los equivalentes de ARN (es decir, genomas) del VIH que se asocian con 1 ml de plasma del paciente (DYBUL et al, 2002). Por lo tanto, la carga de virus se determina midiendo el contenido de ADN viral en una muestra obtenida del paciente. Actualmente, hay tres tipos principales de ensayos de carga viral disponibles:

- 55
- 60 1) reacción en cadena de la polimerasa con transcripción inversa del ARN del VIH (RT-PCR): prueba de monitorización de VIH-1 Amplicor (TM); Roche Diagnostics
  - 2) ADN de cadena ramificada (bDNA): ensayo de ARN de VIH Versant (TM); Bayer Diagnostics; y
  - 3) Amplificación basada en secuencia de ácido nucleico (NASBA): ensayo NucliSens (TM); bioMérieux.

65 En una realización preferida, la composición farmacéutica de la presente invención es capaz de reducir la carga del virus en un sujeto infectado por un retrovirus por debajo de 5.000 equivalentes de genoma/ml de plasma,

preferiblemente por debajo de 500 equivalentes de genoma/ml de plasma y más preferiblemente por debajo de 50 equivalentes de genoma/ml de plasma. El paciente con una carga de virus por debajo de 5000 equivalentes de genoma/ml de plasma se considera que está relativamente bien ajustado para el tratamiento medicinal. Sin embargo, el objetivo en la terapia actual contra el SIDA es una reducción de la carga viral por debajo del límite de detección de los ensayos de carga del virus, que actualmente está por debajo de aproximadamente 50 equivalentes de genoma/ml de plasma. El retrovirus a tratar con la composición farmacéutica de la presente invención es un lentivirus, es decir, VIH, preferiblemente virus de inmunodeficiencia humana de tipo 1 (VIH-1). Sin embargo, es obvio para el experto en la técnica que la presente invención también es aplicable a las infecciones retrovirales por otros retrovirus a parte de los mencionados anteriormente.

El sujeto infectado por un retrovirus, al que se va a administrar la composición farmacéutica, se selecciona del grupo que consiste en seres humanos, primates, monos, ganado, caballos, cabras, ovejas y gatos domésticos. Sin embargo, el sujeto es preferiblemente un ser humano.

En general, una cantidad eficaz del vector de expresión, la recombinasa a medida o la célula transformada de la invención se administra al sujeto. La administración puede ser, por ejemplo, administración intravenosa o intramuscular.

En una realización, la composición farmacéutica se formula para administración concomitante con otros agentes activos de la terapia antirretroviral de gran actividad (TARGA). La terapia antirretroviral de gran actividad TARGA es una terapia de combinación dirigida a la transcriptasa inversa viral, la proteasa y la fusión (GULIC et al, 1997; LALEZARI et al, 2003).

En otra realización, la composición farmacéutica se formula para la administración concomitante o subsiguiente a la terapia de la activación inmunitaria global o la activación específica de la expresión génica de provirus. La premisa de la terapia de activación inmunitaria se basa en la hipótesis de que la activación deliberada de células infectadas por el VIH de forma latente puede acelerar la erradicación de los reservorios virales persistentes. La erradicación tendría lugar a través de la eliminación inmunitaria mediante la muerte programada de las células que expresan activamente los productos de VIH-1 (pro-apoptótica) (KULKOSKY y BRAY, 2006). La activación inmunitaria global (activación de las células inmunitarias, incluyendo células en reposo) se logra generalmente mediante, por ejemplo, la administración de inmunotoxinas, citoquinas (por ejemplo, IL-2), o anticuerpos que activan células T (por ejemplo, OKT3).

En vista del hecho de que la activación inmunitaria realizada para activar deliberadamente reservorios latentes resistentes a TARGA no consiguió desafortunadamente eliminar de forma permanente el VIH-1 y el rebote viral (para comentarios véase KULOSKY y BRAY 2006; MARCELLO, 2006; SHEHU-XHILAGA et al, 2005) debido al hecho de que la activación global de células T aparentemente también induce la replicación viral y aumenta el número de potenciales células diana de VIH-1 más allá del nivel que puede estar contenido por TARGA (FRASER et al, 2000), son necesarios tratamientos más específicos para tratar el VIH. Un enfoque es la activación de la transcripción de los genomas virales, por lo demás, quiescentes. La activación específica de la expresión génica de provirus latente puede conseguirse mediante la administración de éster de forbol prostratina o la citocina humana IL-7, que parecen ambos reactivar el VIH-1 latente en ausencia de proliferación celular (MARCELLO, 2006). Además, la activación de la transcripción selectiva de VIH-1 también se puede lograr mediante inhibidores de la histona-desacetilasa (HDACI), tal como, por ejemplo, ácido valproico, que finalmente induce el sobrecrecimiento de VIH-1 a partir de células en reposo en ausencia de activación celular (MARCELLO 2006; Lehrman et al, 2005).

Sin embargo, la terapia de activación inmunitaria global o la activación específica de la expresión génica de provirus o estrategias de terapia similares se benefician en gran medida de la eliminación simultánea de ADN proviral, reduciendo de ese modo en el paciente el conjunto de células infectadas.

La presente invención también proporciona un procedimiento de tratamiento y/o prevención de una infección por retrovirus, en particular, una infección por VIH, en un sujeto. En una realización, la secuencia del retrovirus que infecta el sujeto se analiza en una muestra obtenida del sujeto, y al menos un vector de expresión que codifica una recombinasa a medida, al menos una recombinasa a medida o al menos una célula transformada con dicho vector de expresión, por ejemplo, una célula madre adulta, se administra al sujeto, si el ADN proviral del sujeto comprende la secuencia diana asimétrica identificada en la etapa (a) en la que se ha seleccionado la recombinasa. La muestra obtenida del sujeto puede ser una muestra de sangre, por ejemplo, que comprende células CD4+ infectadas.

Breve descripción de las figuras

**Fig. 1:** La figura 1 proporciona un alineamiento de las secuencias de la proteína de (a) la recombinasa Cre SEQ ID NO: 6; (b) secuencia de consenso común de Tre, SEQ ID NO: 7, secuencia de consenso de recombinasas Tre específica para sitios diana asimétricos dentro de la LTR de VIH-1 según el documento WO 2011/147590; (c) secuencia de consenso de Tre recombinasas 3.0, SEQ ID NO: 8, recombinasas Tre específicas para SEQ ID NO: 1 según el documento WO 2011/147590;

(d) secuencia consenso de recombinasas Tre 3.1 consenso 85%, SEQ ID NO: 9, las mutaciones individuales en la secuencia de consenso frente a la secuencia de Cre están presentes en el 85% de todos los clones generados por el procedimiento de la presente invención y analizados mediante secuenciación de alto rendimiento (33.000 lecturas de secuencias de 200 pb únicas);

5 (e) secuencia consenso de recombinasas Tre 3.1 consenso 100%, SEQ ID NO: 10, secuencia de consenso de tres recombinasas Tre 3.1 seleccionadas por una alta especificidad para la SEQ ID NO: 1 (sin actividad en loxP, loxH o loxLTR Tre 1.0 (SEQ ID NOs: 3-5) y la tolerabilidad por los seres humanos; y

10 (f) recombinasa Tre 3.1 de ejemplo Utre, SEQ ID NO: 11, muy específica para SEQ ID NO: 1 y bien tolerada por los seres humanos La recombinasa a medida de acuerdo con esta secuencia se designa uTre, respectivamente "Tre universal". Los parches de letras en negrita indican aminoácidos conservados, posiciones variables y no especificadas se indican con una X. Las posiciones de Cre mutada en las SEQ ID NO: 9, 10 y/o 11 están subrayadas en la secuencia de Cre y la posición de la mutación se proporciona por encima de las secuencias. El intercambio Q89L encontrado únicamente en Tre3.1 está marcado en cursiva.

15 **Fig. 2:** La figura muestra un vector de evolución de ejemplo para la selección evolutiva contra recombinación en loxP. Los vectores correspondientes para seleccionar contra la recombinación en loxH pueden construirse fácilmente. loxLTR comprende la SEQ ID NO: 1. Se realizaron los ciclos de evolución de Tre3 51-69 con pEVOloxLTR-loxP y pEVOloxLTR-loxH, alternativamente, 100 - 1 µg/ml del activador de transcripción L-ara, incluyendo dos rondas de barajado de ADN).

20 **Fig. 3:** La Figura 3 muestra la alta especificidad de Utre vs. Tre3. Tre3: Clon fue aislado del ciclo 43 de la biblioteca Tre3.0. Utre: Clon fue aislado del ciclo 71 de la biblioteca Tre3.1. El producto recombinado se marca con un triángulo, no recombinado con dos triángulos. En condiciones en las que la recombinasa a medida se expresa (inducción por L-arabinosa), uTre recombina loxLTR que comprende la SEQ ID NO: 1 en E. coli. En cambio, Tre3 recombina loxLTR y loxP y loxH, es decir, tiene una especificidad relativamente relajada.

25 **Fig. 4:** La figura 4A muestra un vector de expresión lentiviral de ejemplo para la expresión constitutiva de un marcador seleccionable, EGFP, y la biblioteca de *tre* en células humanas. La Fig. 4B muestra el esquema de flujo para la selección celular de Tre bien tolerable y muy específico. La selección final para Tre muy activa confirma la actividad en las células humanas. Se seleccionaron clones individuales de recombinasas y se sometieron a análisis adicionales.

30 **Fig. 5:** La figura 5 muestra actividad uTre antiviral en cultivo de tejidos en dos cultivos representativos. Las células PM1 T se transdujeron con vectores que codificaban GFP sola (control, círculos sin relleno) o que codificaban uTre y GFP (uTre, cuadrados rellenos). Posteriormente, los cultivos se infectaron con VIH-1 y la carga viral se monitorizó con el tiempo usando ELISA con antígeno p24. El experimento muestra que la recombinasa a medida es eficaz en la reducción de la carga viral. Después de varias semanas (8 o 9 semanas), la carga viral no es detectable mediante una ELISA con antígeno p24.

35 **Fig. 6** Se muestra actividad uTre antiviral pronunciada en células CD4+ humanas primarias derivadas de un paciente infectado por el VIH. Las células se transdujeron con un vector que expresa GFP sola (experimento de control; panel de la izquierda) o con un vector que expresa uTre y GFP (uTre; panel derecho). La replicación del virus se controló mediante la liberación de antígeno p24 de VIH-1 (círculos sin relleno) y el porcentaje de células CD4+ humanas (GFP+) transducidas (cuadrados rellenos) se controló mediante FACS. La expresión de uTre dio lugar a un efecto antiviral pronunciado y protección de las células CD4+. Cabe destacar que la disminución de la carga viral entre el día 15 y el día 20 en el experimento de control refleja la muerte celular debido a la replicación del virus no inhibida.

40 **Fig. 7** La figura 7 muestra la actividad uTre antiviral en ratones humanizados infectados por el VIH. Los ratones inmunodeficientes fueron injertados con células madre hematopoyéticas CD34+ humanas/HSC (control), o con CD34+ HSC que expresan uTre (Animal # 1 y # 2). Posteriormente, los animales fueron infectados por VIH-1 y la carga viral (detectada como copias de ARN de VIH-1/ml; círculos sin relleno) y el porcentaje de células CD45+CD4+ humanas de todos los linfocitos (cuadrados rellenos) se monitorizaron con el tiempo.

## 55 Ejemplos

### Ejemplo 1:

60 Se utilizan materiales y procedimientos como se describen en WO 2008/083931, WO 2011/147590 y BUCHHOLZ & STEWART, 2001, si no se especifica lo contrario. Las recombinasas a medida capaces de recombinar secuencias diana asimétricas en una pluralidad de diferentes cepas de VIH-1 se prepararon como se describe en el documento WO 2011/147590. Las bibliotecas de *tre* resultantes se emplean en experimentos adicionales.

### Ejemplo 2:

65 Para mejorar la especificidad de uTre, se realizaron ciclos de evolución adicionales que seleccionan contra la actividad de recombinación en loxP y loxH. Para este propósito, se clonó la biblioteca de Tre desarrollada obtenida a

partir del ciclo de evolución 50 en un vector de evolución que contiene los dos sitios loxLTR (SEQ ID NO: 1) entrelazados con dos sitios loxP o dos sitios loxH, respectivamente. Un vector de ejemplo se muestra en la Figura 2. Después de la inducción de la expresión de recombinasa, la recombinación en loxLTR dio lugar a la eliminación del único sitio Ndel presente, mientras que la recombinación en loxP o loxH no. El ADN del plásmido aislado después de cada ciclo de evolución se digirió con Ndel y se amplificaron por PCR las secuencias codificantes de recombinasa que habían recombinado con éxito loxLTR en lugar de loxP o loxH y se subclonaron de nuevo en el vector de evolución para el siguiente ciclo de evolución. Se realizaron un total de 19 ciclos de evolución de Tre3 adicionales, incluyendo dos rondas de barajado de ADN, se llevaron a cabo, seleccionando alternativamente contra la recombinación en loxP y loxH.

### Ejemplo 3:

Para seleccionar recombinasas uTre con toxicidad celular (es decir citopaticidad) disminuida significativamente, las bibliotecas *tre* se ligaron en un vector lentiviral que expresa constitutivamente EGFP a partir de un promotor interno de LTR SFFV y la biblioteca *tre* del promotor constitutivo EF1alfa (Fig. 4A). La transducción de células T Jurkat permitió la clasificación secuencial (por FACS) de células que expresan de forma elevada GFP y el posterior aislamiento de clones de uTre no tóxicos (figura 4B). Para esto, se realizaron clasificaciones de células en los cultivos de células T transducidas en el día 3, día 10 y día 24 después de transducción con el aumento de rigurosidad en la expresión de EGFP. Después de otra semana de cultivo, la biblioteca *tre* restante se aisló y se analizaron los clones seleccionados con respecto a la actividad de Tre mejorada.

### Ejemplo 4:

Para analizar la actividad de uTre en líneas celulares, se transdujeron cultivos de células PM-1 T con vectores retrovirales derivados de ASLV que expresan Utre y GFP, o GFP sola (vector de control negativo). Cabe indicar, la expresión de GFP permitió el seguimiento de las células transducidas. A los 10 días después de la transducción, las células se infectaron con VIH-1<sub>Bal</sub>. El efecto de la expresión de uTre sobre la replicación de VIH-1 se controló mediante mediciones semanales de ELISA de la cantidad de antígeno p24 viral en los sobrenadantes de cultivo. Como se muestra (Figura 5), la liberación de p24 disminuyó notablemente en los cultivos transducidos con uTre, mientras que se mantiene estable o incluso aumenta en los cultivos de control (que expresan GFP solo).

### Ejemplo 5:

Análisis de la actividad de uTre en células CD4+ primarias derivadas de un paciente infectado por VIH-1. Las células CD4+ se estimularon con perlas magnéticas CD3/CD28 durante 48 h. Posteriormente, las células fueron transducidas con vectores lentivirales que expresaban GFP sola (que sirve como control negativo) o expresaban uTre junto con GFP. Las células se cultivaron en presencia de 100 UI de IL-2 durante 20 días. Las cargas virales (medidas mediante ELISA con antígeno p24) y el recuento de células CD4+ transducidas humanas (analizado por FACS) se monitorizaron en los días indicados después de transducción. Como se muestra en la Figura 6, la expresión de uTre da lugar a un efecto antiviral pronunciado (indicado por círculos abiertos) y la protección de células CD4+ (indicado por cuadrados rellenos). En cambio, la disminución de la carga viral entre el día 15 y día 20 en el experimento de control refleja la muerte celular debido a la replicación del virus no inhibida.

### Ejemplo 6:

Análisis de la actividad de uTre *in vivo*. Se injertaron ratones inmunodeficientes NOG (NOD.Cg-Prkdc<sup>scid</sup> IL2R<sup>tm1Wjl</sup>/SzJ) con células madre hematopoyéticas CD34+ humanas/HSC (control), o con CD34+ HSC que expresan uTre. Posteriormente, los animales se infectaron con VIH-1<sub>Bal</sub> y la carga viral (detectado por el ensayo basado en la PCR ultrasensible) y el porcentaje de células CD45+ CD4+ humanas (analizado por FACS) se monitorizaron con el tiempo. Como se muestra (Figura 7), la expresión de uTre dio lugar a actividades antivirales significativas *in vivo*.

### Lista de referencias

- Abremski K, Hoess RH, Sternberg N (1983) "Studies on the properties of PI site-specific recombination: evidence for topologically unlinked products following recombination." *Cell* 32, 1301-1311.
- Abremski K, Hoess R (1983) "Bacteriophage PI site-specific recombination. Purification and properties of the Cre recombinase protein." *J Biol. Chem.* 259, 1509-1514.
- Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA (1986) "Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone." *J Virol.* 59, 284-291.
- Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2006) "Tuning genetic control through promoter engineering" *Proc. Natl. Acad. Sci. USA* 102, 12678-12683.

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25, 3389-3402.
- 5 Beyer WR, Westphal M, Ostertag W, von Laer D (2002)"Oncoretrovirus and lentivirus vectors pseudotyped with lymphocytic choriomeningitis virus glycoprotein: generation, concentration and broad host range." *J Virol.* 76, 1488-1495.
- Blackard JT, Renjifo BR, Mwakagile D, Montano MA, Fawzi WW, Essex M (1999) "Transmission of human immunodeficiency type 1 viruses with intersubtype recombinant long terminal repeat sequences." *Virology* 254, 220-225.
- 10 Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH (2005) "Evolving strategies for enzyme engineering." *Curr. Opin. Struct. Biol.* 15, 447-452.
- 15 Buchholz F, Ringrose L, Angrand PO, Rossi F, Stewart AF (1996) "Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination." *Nucl. Acids Res.* 24, 4256-4262.
- Buchholz F, Angrand PO, Stewart AF (1998) "Improved properties of FLP recombinase evolved by cycling mutagenesis." *Nat. Biotechnol.* 16, 657-662.
- 20 Buchholz F, Stewart AF (2001) "Alteration of Cre recombinase site specificity by substrate-linked protein evolution." *Nat. Biotechnol.* 19, 1047-1052.
- 25 Chiu YL, Soros VB, Kreisberg JF, Stopak K, Yonemoto W, Greene WC (2005) "Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells." *Nature* 435, 108-114
- Chun T-W, Engel D, Berrey MM, Shea T, Corey L, Fauci AS (1998) "Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection." *Proc. Natl. Acad. Sci. USA* 95, 8869-8873.
- 30 Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF (2005) "Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools." *Trends Biotechnol.* 23, 407-419.
- Collins CH, Yokobayashi Y, Umeno D, Arnold FH, (2003) "Engineering proteins that bind, move, make and break DNA." *Curr. Opin. Biotechnol.* 14, 665.
- 35 Combes P, Till R, Bee S, Smith MC (2002) "The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31 -encoded site-specific recombination system." *J Bacteriol.* 184, 5746-5752.
- 40 Cramer A, Raillard SA, Bermudez E, Stemmer WP (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution." *Nature* 391, 288-291. Derossi D, Joliot AH, Chassaing G, Prochiantz A (1994) "The third helix of the Antennapedia homeodomain translocates through biological membranes." *J Biol Chem.* 269, 10444-10450.
- 45 Derossi D, Calvet S, Trembleau A, Chassaing G, Prochiantz A (1996) "Cell internalization of the third helix of the Antennapedia homeodomain is receptor-independent." *J Biol Chem* 271, 18188-18193. Donovan, P.J., Gearhart, J. (2001) "The end of the beginning for pluripotent stem cells." *Nature* 414, 92-97.
- 50 Donzella GA, Schols D, Lin SW, Este JA, Nagashima KA, Maddon PJ, Allaway GP, Sakmar TP, Henson G, De Clercq E, Moore JP (1998) "AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor." *Nature Medicine* 4, 72-77.
- Dybul M, Fauci AS, Bartlett JG, Kaplan JE, Pau AK (2002) "Guidelines for using antiretroviral agents among HIV infected adults and adolescents." *Annals of Internal Medicine* 137, 381-433.
- 55 Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14, 755- 763.
- Edelman GM, Meech R, Owens GC, Jones FS (2000) "Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity." *Proc. Natl. Acad. Sci. USA* 97, 3038-3043. Elliott G, O'Hare P., Intercellular trafficking and protein delivery by a herpesvirus structural protein., *Cell.* 1997 Jan 24;88 (2):223-33
- 60 Emerman M, Malim MH (1998) "HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology." *Science* 280, 1880-1884.
- 65 Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J., Tat-mediated delivery of heterologous proteins into cells., *Proc Natl Acad Sci U S A.* 1994 Jan 18;91(2):664-8.

- 5 Finzi D, Hemankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF (1997) "Identification of a reservoir for HIV 1 in patients on highly active antiretroviral therapy." *Science* 278, 1295-1300.
- 10 Flowers CC, Woffendin C, Petryniak J, Yang S, Nabel GJ (1997) "Inhibition of recombinant human immunodeficiency virus type 1 replication by a site-specific recombinase." *J. Virol.* 71, 2685-2692. Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, Richman DD, Valentine FT, Jonas L, Meibohm A, Emini EA, Chodakewitz JA (1997) "Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy." *N Engl. J. Med.* 337, 734-739.
- 15 Guzman LM, Belin D, Carson MJ, Beckwith J (1995) "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." *J. Bacteriol.* 177, 4121-4130. Hartenbach S, Fussenegger M (2006) "A novel synthetic mammalian promoter derived from an internal ribosome entry site." *Biotechnology and Bioengineering* 95, 547-559.
- 20 Hauber I, Bevec D, Heukeshoven J, Kratzer F, Horn F, Choidas A, Harrer T, Hauber J (2005) "Identification of cellular deoxyhypusine synthase as a novel target for antiretroviral therapy." *J. Clin. Invest.* 115, 76-85.
- 25 Hazuda DJ, Young SD, Guare JP, Anthony NJ, Gomez RP, Wai JS, Vacca JP, Handt L, Motzel SL, Klein HJ, Dornadula G, Danovich RM, Witmer MV, Wilson KA, Tussey L, Schleif WA, Gabryelski LS, Jin L, Miller MD, Casimiro DR, Emini EA, Shiver JW (2004) "Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques." *Science* 305, 528-532.
- 30 Hoess RH, Abremski K (1985) "Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system." *J. Mol. Biol.* 181, 351-362. Johannes TW, Zhao H (2006) "Directed evolution of enzymes and biosynthetic pathways." *Curr. Opin. Microbiol.* 9, 261-267.
- 35 Krasnow MA, Cozzarelli NR (1983) "Site-specific relaxation and recombination by the Tn3 resolvase: Recognition of the DNA path between oriented res sites." *Cell* 32, 1313-1324.
- 40 Kulkosky J, Bray S (2006) "HAART-persistent HIV-1 latent reservoirs: their origin, mechanisms of stability and potential strategies for eradication." *Curr. HIV Res.* 4, 199-208.
- 45 Lalezari JP, Henry K, O'Hearn M, Montaner JS, Piliero PJ, Trottier B, Walmsley S, Cohen C, Kuritzkes DR, Eron Jr. JJ, Chung J, DeMasi R, Donatucci L, Drobnies C, Delehanty J, Salgo M (2003) "Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America." *N. Engl. J. Med.* 348, 2175-2185.
- 50 Lee YS, Park JS (1998) "A novel mutant loxP containing part of long terminal repeat of HIV - 1 in spacer region: presentation of possible target site for antiviral strategy using site-specific recombinase." *Biochem. Biophys. Res. Comm.* 253, 588-593.
- 55 Lee YS, Kim ST, Kim GW, Lee M, Park JS (2000) "An engineered loxP sequence containing part of a long terminal repeat of HIV-1 permits Cre recombinase-mediated DNA excision." *Biochem. Cell Biol.* 78, 653-658.
- 60 Lehrman G, Hogue IB, Palmer S, Jennings C, Spina CA, Wiegand A, Landay AL, Coombs RW, Richman DD, Mellors JW, Coffin JM, Bosch RJ, Margolis DM (2005) "Depletion of latent HIV-1 infection *in vivo*: a proof-of-concept study" *Lancet* 366, 549-555.
- 65 Lewandoski, M. (2001) "Conditional control of gene expression in the mouse." *Nat. Rev. Genet.* 2, 743-755. Lin Q, Jo D, Gebre-Amlak KD, Ruley HE (2004) "Enhanced cell-permeant Cre protein for site-specific recombination in cultured cells." *BMC Biotechnol.* 4, 25.
- Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, Koup RA, Mellors JW, Connick E, Conway B, Kilby M, Wang L, Whitcomb JM, Hellmann NS, Richman DD (2002) "Antiretroviral-drug resistance among patients recently infected with HIV." *N. Engl. J. Med.* 347, 385-394.
- Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, Kanaar R, Berns A, Jonkers J (2001) "Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells." *PNAS* 98, 9209-9214.
- Macara IG (2001) "Transport into and out of the nucleus." *Microbiology and molecular biology reviews* 65, 570-594. Malim MH, Hauber J, Fenrick R, Cullen BR (1988) "Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes." *Nature* 335, 181-183.
- Marcello A (2006) "Latency: the hidden HIV-1 challenge." *Retrovirology* 3, 7.

## ES 2 710 708 T3

- Matsumura I, Ellington AD (2001) "*In vitro* evolution of beta-glucuronidase into a beta- galactosidase proceeds through non-specific intermediates." J. Mol. Biol. 305, 331-339.
- Minshull J, Stemmer WP. (1999) "Protein evolution by molecular breeding." Curr. Opin. Chem. Biol. 3, 284-290.
- 5 Nagy A (2000) "Cre recombinase: the universal reagent for genome tailoring." Genesis 26, 99-109. Needleman SB, Wunsch CD (1970) "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J. Mol. Biol. 48, 443-453.
- 10 Nolden L, Edenhofer F, Haupt S, Koch P, Wunderlich FT, Siemen H, Brustle O. (2006) "Site- specific recombination in human embryonic stem cells induced by cell-permeant Cre recombinase." Nat. Methods 3, 461-467.
- Oess S, Hildt E., Novel cell permeable motif derived from the PreS2-domain of hepatitis-B virus surface antigens., Gene Ther. 2000 May; 7(9):750-8
- 15 O'Doherty U, Swiggard WJ, Malim MH (2000) "Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding." J Virol. 74, 10074-10080. Pearson WR, Lipman DJ (1988) "Improved tools for biological sequence comparison." Proc Natl Acad Sci USA 85, 2444-2448.
- 20 Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F. (2002) "Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: A tool for efficient genetic engineering of mammalian genomes." Proc. Natl. Acad. Sci. USA 99, 4489-4494.
- 25 Ratner L, Starcich B, Josephs SF, Hahn BH, Reddy EP, Livak KJ, Petteway SR, Jr., Pearson ML, Haseltine WA, Arya SK, (1985) "Polymorphism of the 3' open reading frame of the virus associated with the acquired immune deficiency syndrome, human T-lymphotropic virus type III." Nucl. Acids Res. 13, 8219-8229.
- Richard JP, Melikov K, Brooks H, Prevot P, Lebleu B, Chemomordik LV (2005) "Cellular uptake of the unconjugated TAT peptide involves clathrin-dependent endocytosis and heparin sulfate receptors." J. Biol. Chem. 280, 15300-15306.
- 30 Rufer AW, Sauer B (2002) "Non-contact positions impose site selectivity on Cre recombinase." Nucl. Acids Res. 30, 2764-2771.
- 35 Ruhl M, Himmelspach M, Bahr GM, Hammerschmid F, Jaksche H, Wolff B, Aschauer H, Farrington GK, Probst H, Bevec D, Hauber J (1993) "Eukaryotic initiation factor 5 A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation" J Cell Biol. 123, 1309-1320.
- 40 Sanger F, Nickler S, Coulson AR (1977) "DNA sequencing with chain-terminating inhibitors." Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Santoro SW, Schultz PG (2002) "Directed evolution of the site specificity of Cre recombinase." Proc. Natl. Acad. Sci. USA 99, 4185-4190.
- 45 Saraf-Levy T, Santoro SW, Volpin H, Kushnirsky T, Eyal Y, Schultz PG, Gidoni D, Carmi N (2006) "Site-specific recombination of asymmetric lox sites mediated by a heterotetrameric Cre recombinase complex." Bioorg. Med. Chem. 14, 3081-3089.
- 50 Sauer B, McDermott J (2004) "DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-cl regions of PI -related phages." Nucl. Acids. Res. 32, 6086- 6095. Schambach A, Bohne J, Chandra S, Will E, Margison GP, Williams DA, Baum C (2006) "Equal potency of gammaretroviral and lenti viral SIN vectors for expression of O6- methylguanine-DNA methyltransferase in hematopoietic cells." Molecular Therapy 13, 391- 400.
- 55 Scherr M, Eder M (2002) "Gene Transfer into Hematopoietic Stem Cells Using Lentiviral Vectors." Current Gene Therapy 2, 45-55.
- 60 Shehu-Xhilaga M, Tachedjian G, Crowe SM, Kedzierska K. (2005) "Antiretroviral compounds: mechanisms underlying failure of HAART to eradicate HIV-1." Curr. Med. Chem. 12, 1705-1719. Shimshek DR, Kim J, Hubner MR, Spergel DJ, Buchholz F, Casanova E, Stewart AF, See- burg PH, Sprengel R (2002) "Codon-improved Cre recombinase (iCre) expression in the mouse." Genesis 32(1), 19-26.
- Smith Tf, Waterman MS (1981) "Overlapping genes and information theory." J Theor. Biol. 91, 379-380.
- 65 Stark WM, Boocock MR, Sherratt DJ (1992) "Catalysis by site-specific recombinases." Trends Genet. 8, 432-439.
- Stemmer WPC (1994) "Rapid evolution of a protein *in vitro* by DNA shuffling." Nature 370, 389-391.

Sternberg N, Hamilton D (1981) "Bacteriophage PI site-specific recombination. I. Recombination between loxP sites." J. Mol. Biol. 150, 467-486.

5 Van Duyne GD (2001) "A structural view of cre-loxp site-specific recombination." Annu. Rev. Biophys. Biomol. Struct. 30, 87-104.

Vives E, Brodin P, Lebleu B (1997) "A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus." J. Biol. Chem. 272, 16010-16017. Vives E (2003) "Cellular uptake of the TAT peptide: an endocytosis mechanism following<"> ionic interactions." J. Mol. Recognit. 16, 265-271.

Volkert FC, Broach JR (1986) "Site-specific recombination promotes plasmid amplification in yeast." Cell 46, 541-550.

15 Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M (2003) "Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site." J. Mol. Biol. 326, 65-76.

Yuan L, Kurek I, English J, Keenan R (2005) "Laboratory-directed protein evolution" Microbiol. Mol. Biol. Rev. 69, 373-92.

20 WO 2002/44409

WO 2008/083931

25 WO 2011/147590.

LISTADO DE SECUENCIAS

30 <110> Heinrich-Pette-Institut Leibniz-Institut für Exp. Virologie  
Technische Universität Dresden

<120> Recombinasa a medida bien tolerada y muy específica para recombinar sitios diana asimétricos en una pluralidad de cepas de retrovirus

35 <130> HEI15463PCT

<160> 14

40 <170> BISSAP 1.3

<210> 1

<211> 34

<212> ADN

<213> Virus de inmunodeficiencia humana 1

45 <400> 1  
aaccactgc ttaagcctca ataaagcttg cctt 34

50 <210> 2  
<211> 34  
<212> ADN  
<213> Virus de inmunodeficiencia humana 1

55 <400> 2  
ctgggcggga ctggggagtg gcgagccctc agat 34

60 <210> 3  
<211> 34  
<212> ADN  
<213> Virus de inmunodeficiencia humana 1

65 <400> 3  
acaacatcct attacaccct atatgccaac atgg 34

<210> 4  
 <211> 34  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <223> loxP  
 <400> 4  
 10 ataacttcgt ataatgtatg ctatacgaag ttat 34  
 <210> 5  
 <211> 34  
 <212> ADN  
 <213> Secuencia artificial  
 15  
 <220>  
 <223> loxH  
 20  
 <400> 5  
 atatatacgt atatagacat atatacgtat atat 34  
 <210> 6  
 <211> 343  
 <212> PRT  
 <213> Fago P1 de Enterobacteria  
 25  
 <400> 6  
 30 Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val  
 1 5 10 15  
 Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg  
 20 25 30  
 35 Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val  
 35 40 45  
 Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe  
 50 55 60  
 40 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala  
 65 70 75 80  
 Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn  
 85 90 95  
 Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala  
 100 105 110  
 45 Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln  
 130 135 140  
 50 Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn  
 145 150 155 160  
 Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu  
 165 170 175  
 Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg  
 180 185 190  
 55 Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp  
 210 215 220  
 60 Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys  
 225 230 235 240  
 Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu  
 245 250 255  
 Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile  
 260 265 270  
 65 Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly  
 275 280 285  
 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300

ES 2 710 708 T3

Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile  
 305 310 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val  
 325 330 335  
 5 Arg Leu Leu Glu Asp Gly Asp  
 340

- <210> 7
- <211> 343
- 10 <212> PRT
- <213> Secuencia artificial
- <220>
- 15 <223> Secuencia consenso común Tre 100% (wo 2011/147590)
- <220>
- <221> VARIANTE
- <222> 3..5
- 20 <220>
- <221> VARIANTE
- <222> 7..10
- 25 <220>
- <221> VARIANTE
- <222> 12..12
- 30 <220>
- <221> VARIANTE
- <222> 15..16
- 35 <220>
- <221> VARIANTE
- <222> 18..19
- 40 <220>
- <221> VARIANTE
- <222> 28..31
- 45 <220>
- <221> VARIANTE
- <222> 34..35
- 50 <220>
- <221> VARIANTE
- <222> 37..37
- 55 <220>
- <221> VARIANTE
- <222> 39..40
- 60 <220>
- <221> VARIANTE
- <222> 48..48
- 65 <220>
- <221> VARIANTE
- <222> 51..51
- <220>

<221> VARIANTE  
 <222> 53..53  
  
 5 <220>  
 <221> VARIANTE  
 <222> 57..60  
  
 10 <220>  
 <221> VARIANTE  
 <222> 62..63  
  
 15 <220>  
 <221> VARIANTE  
 <222> 66..67  
  
 20 <220>  
 <221> VARIANTE  
 <222> 69..70  
  
 25 <220>  
 <221> VARIANTE  
 <222> 73..73  
  
 30 <220>  
 <221> VARIANTE  
 <222> 77..77  
  
 35 <220>  
 <221> VARIANTE  
 <222> 80..80  
  
 40 <220>  
 <221> VARIANTE  
 <222> 84..84  
  
 45 <220>  
 <221> VARIANTE  
 <222> 86..86  
  
 50 <220>  
 <221> VARIANTE  
 <222> 88..88  
  
 55 <220>  
 <221> VARIANTE  
 <222> 93..94  
  
 60 <220>  
 <221> VARIANTE  
 <222> 98..98  
  
 65 <220>  
 <221> VARIANTE  
 <222> 102  
  
 <220>  
 <221> VARIANTE  
 <222> 105  
  
 <220>  
 <221> VARIANTE  
 <222> 107..108  
  
 <220>  
 <221> VARIANTE  
 <222> 110..112  
  
 <220>

<221> VARIANTE  
 <222> 116..117  
  
 5 <220>  
 <221> VARIANTE  
 <222> 122  
  
 10 <220>  
 <221> VARIANTE  
 <222> 131  
  
 15 <220>  
 <221> VARIANTE  
 <222> 138  
  
 20 <220>  
 <221> VARIANTE  
 <222> 140  
  
 25 <220>  
 <221> VARIANTE  
 <222> 142..147  
  
 30 <220>  
 <221> VARIANTE  
 <222> 149..151  
  
 35 <220>  
 <221> VARIANTE  
 <222> 153-156  
  
 40 <220>  
 <221> VARIANTE  
 <222> 158  
  
 45 <220>  
 <221> VARIANTE  
 <222> 160  
  
 50 <220>  
 <221> VARIANTE  
 <222> 163  
  
 55 <220>  
 <221> VARIANTE  
 <222> 166  
  
 60 <220>  
 <221> VARIANTE  
 <222> 174-175  
  
 65 <220>  
 <221> VARIANTE  
 <222> 177..180  
  
 <220>  
 <221> VARIANTE  
 <222> 182..183  
  
 <220>  
 <221> VARIANTE  
 <222> 185  
  
 <220>  
 <221> VARIANTE  
 <222> 187  
  
 <220>

<221> VARIANTE  
 <222> 189

5 <220>  
 <221> VARIANTE  
 <222> 193

10 <220>  
 <221> VARIANTE  
 <222> 197..199

15 <220>  
 <221> VARIANTE  
 <222> 207

20 <220>  
 <221> VARIANTE  
 <222> 216..217

25 <220>  
 <221> VARIANTE  
 <222> 219

30 <220>  
 <221> VARIANTE  
 <222> 221

35 <220>  
 <221> VARIANTE  
 <222> 225

40 <220>  
 <221> VARIANTE  
 <222> 227

45 <220>  
 <221> VARIANTE  
 <222> 232..236

50 <220>  
 <221> VARIANTE  
 <222> 241..245

55 <220>  
 <221> VARIANTE  
 <222> 247

60 <220>  
 <221> VARIANTE  
 <222> 249

65 <220>  
 <221> VARIANTE  
 <222> 251

<220>  
 <221> VARIANTE  
 <222> 253..255

<220>  
 <221> VARIANTE  
 <222> 258..260

<220>  
 <221> VARIANTE  
 <222> 262..263

<220>

ES 2 710 708 T3

<221> VARIANTE  
 <222> 266..268  
  
 <220>  
 5 <221> VARIANTE  
 <222> 270..273  
  
 <220>  
 10 <221> VARIANTE  
 <222> 277..278  
  
 <220>  
 15 <221> VARIANTE  
 <222> 281  
  
 <220>  
 20 <221> VARIANTE  
 <222> 284..285  
  
 <220>  
 25 <221> VARIANTE  
 <222> 305  
  
 <220>  
 30 <221> VARIANTE  
 <222> 307  
  
 <220>  
 35 <221> VARIANTE  
 <222> 316  
  
 <220>  
 40 <221> VARIANTE  
 <222> 319..320  
  
 <220>  
 45 <221> VARIANTE  
 <222> 332  
  
 <220>  
 50 <221> VARIANTE  
 <222> 333  
  
 <220>  
 55 <221> VARIANTE  
 <222> 341..343  
  
 <400> 7  
 45 Met Ser Xaa Xaa Xaa Thr Xaa Xaa Xaa Xaa Leu Xaa Ala Leu Xaa Xaa  
 1 5 10 15  
 Asp Xaa Xaa Ser Asp Xaa Xaa Xaa Xaa Leu Xaa Xaa Xaa Xaa Arg  
 20 25 30  
 50 Asp Xaa Xaa Ala Xaa Ser Xaa Xaa Thr Trp Xaa Xaa Leu Leu Ser Xaa  
 35 40 45  
 Cys Arg Xaa Trp Xaa Ala Trp Cys Xaa Xaa Xaa Xaa Arg Xaa Xaa Phe  
 50 55 60  
 Pro Xaa Xaa Pro Xaa Xaa Val Arg Xaa Tyr Leu Leu Xaa Leu Gln Xaa  
 65 70 75 80  
 55 Arg Gly Leu Xaa Val Xaa Thr Xaa Gln Gln His Leu Xaa Xaa Leu Asn  
 85 90 95  
 Met Xaa His Arg Arg Xaa Gly Leu Xaa Arg Xaa Xaa Asp Xaa Xaa Xaa  
 100 105 110  
 60 Val Ser Leu Xaa Xaa Arg Arg Ile Arg Xaa Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Xaa Lys Gln Ala Leu Ala Phe Xaa Arg Xaa Asp Xaa Xaa Xaa  
 130 135 140  
 Xaa Xaa Xaa Leu Xaa Xaa Xaa Ser Xaa Xaa Xaa Xaa Asp Xaa Arg Xaa  
 145 150 155 160  
 65 Leu Ala Xaa Leu Gly Xaa Ala Tyr Asn Thr Leu Leu Arg Xaa Xaa Glu  
 165 170 175  
 Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp Xaa Ser Xaa Thr Xaa Gly Gly Arg  
 180 185 190

ES 2 710 708 T3

Xaa Leu Ile His Xaa Xaa Xaa Thr Lys Thr Leu Val Ser Thr Xaa Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Xaa Xaa Thr Xaa Leu Xaa Glu Arg Trp  
 210 215 220  
 5 Xaa Ser Xaa Ser Gly Val Ala Xaa Xaa Xaa Xaa Xaa Tyr Leu Phe Cys  
 225 230 235  
 Xaa Xaa Xaa Xaa Xaa Gly Xaa Ala Xaa Pro Xaa Ala Xaa Xaa Xaa Leu  
 245 250 255  
 10 Ser Xaa Xaa Xaa Leu Xaa Xaa Ile Phe Xaa Xaa Xaa His Xaa Xaa Xaa  
 260 265  
 Xaa Gly Ala Lys Xaa Xaa Ser Gly Xaa Arg Tyr Xaa Xaa Trp Ser Gly  
 275 280 285  
 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 15 Xaa Ile Xaa Glu Ile Met Gln Ala Gly Gly Trp Xaa Thr Val Xaa Xaa  
 305 310 315  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Xaa Gly Ala Met Val  
 325 330 335  
 20 Arg Leu Leu Glu Xaa Xaa Xaa  
 340

- <210> 8
- <211> 343
- <212> PRT
- 25 <213> Secuencia artificial
- <220>
- <223> Secuencia consenso Tre 3.0 100% (WO 2011/147590)
- 30 <220>
- <221> VARIANTE
- <222> 3..5
- 35 <220>
- <221> VARIANTE
- <222> 7..10
- 40 <220>
- <221> VARIANTE
- <222> 16
- 45 <220>
- <221> VARIANTE
- <222> 18..19
- 50 <220>
- <221> VARIANTE
- <222> 22..26
- 55 <220>
- <221> VARIANTE
- <222> 28..31
- 60 <220>
- <221> VARIANTE
- <222> 34..35
- 65 <220>
- <221> VARIANTE
- <222> 37
- <220>
- <221> VARIANTE
- <222> 39..40
- <220>
- <221> VARIANTE
- <222> 43

	<220>	
	<221>	VARIANTE
	<222>	48
5		
	<220>	
	<221>	VARIANTE
	<222>	51
10		
	<220>	
	<221>	VARIANTE
	<222>	53
15		
	<220>	
	<221>	VARIANTE
	<222>	57..60
20		
	<220>	
	<221>	VARIANTE
	<222>	62..63
25		
	<220>	
	<221>	VARIANTE
	<222>	66..67
30		
	<220>	
	<221>	VARIANTE
	<222>	73
35		
	<220>	
	<221>	VARIANTE
	<222>	77
40		
	<220>	
	<221>	VARIANTE
	<222>	80
45		
	<220>	
	<221>	VARIANTE
	<222>	84
50		
	<220>	
	<221>	VARIANTE
	<222>	94
55		
	<220>	
	<221>	VARIANTE
	<222>	98
60		
	<220>	
	<221>	VARIANTE
	<222>	102
65		
	<220>	
	<221>	VARIANTE
	<222>	105
	<220>	
	<221>	VARIANTE
	<222>	107..108

	<220>	
	<221>	VARIANTE
	<222>	111..112
5	<220>	
	<221>	VARIANTE
	<222>	116..117
10	<220>	
	<221>	VARIANTE
	<222>	122
15	<220>	
	<221>	VARIANTE
	<222>	131
20	<220>	
	<221>	VARIANTE
	<222>	138
25	<220>	
	<221>	VARIANTE
	<222>	140
30	<220>	
	<221>	VARIANTE
	<222>	142..147
35	<220>	
	<221>	VARIANTE
	<222>	149..151
40	<220>	
	<221>	VARIANTE
	<222>	153..156
45	<220>	
	<221>	VARIANTE
	<222>	158
50	<220>	
	<221>	VARIANTE
	<222>	160
55	<220>	
	<221>	VARIANTE
	<222>	163
60	<220>	
	<221>	VARIANTE
	<222>	166
65	<220>	
	<221>	VARIANTE
	<222>	174
	<220>	
	<221>	VARIANTE
	<222>	177..180
	<220>	
	<221>	VARIANTE
	<222>	182..183
	<220>	
	<221>	VARIANTE
	<222>	185

	<220>	
	<221>	VARIANTE
	<222>	187
5	<220>	
	<221>	VARIANTE
	<222>	189
10	<220>	
	<221>	VARIANTE
	<222>	193
15	<220>	
	<221>	VARIANTE
	<222>	197..199
20	<220>	
	<221>	VARIANTE
	<222>	207
25	<220>	
	<221>	VARIANTE
	<222>	216..217
30	<220>	
	<221>	VARIANTE
	<222>	219
35	<220>	
	<221>	VARIANTE
	<222>	221
40	<220>	
	<221>	VARIANTE
	<222>	225
45	<220>	
	<221>	VARIANTE
	<222>	232..236
50	<220>	
	<221>	VARIANTE
	<222>	241..245
55	<220>	
	<221>	VARIANTE
	<222>	247
60	<220>	
	<221>	VARIANTE
	<222>	249
65	<220>	
	<221>	VARIANTE
	<222>	251
	<220>	
	<221>	VARIANTE
	<222>	253..255
	<220>	
	<221>	VARIANTE
	<222>	258..260

ES 2 710 708 T3

<220>  
 <221> VARIANTE  
 <222> 262..263  
 5  
 <220>  
 <221> VARIANTE  
 <222> 266..268  
 10  
 <220>  
 <221> VARIANTE  
 <222> 270..273  
 15  
 <220>  
 <221> VARIANTE  
 <222> 277..278  
 20  
 <220>  
 <221> VARIANTE  
 <222> 281  
 25  
 <220>  
 <221> VARIANTE  
 <222> 284..285  
 30  
 <220>  
 <221> VARIANTE  
 <222> 305  
 35  
 <220>  
 <221> VARIANTE  
 <222> 316  
 40  
 <220>  
 <221> VARIANTE  
 <222> 319..320  
 45  
 <220>  
 <221> VARIANTE  
 <222> 332  
 50  
 <220>  
 <221> VARIANTE  
 <222> 341..343  
 <400> 8  
 Met Ser Xaa Xaa Xaa Thr Xaa Xaa Xaa Xaa Leu Ser Ala Leu Leu Xaa  
 1 5 10 15  
 Asp Xaa Xaa Ser Asp Xaa Xaa Xaa Xaa Xaa Leu Xaa Xaa Xaa Xaa Arg  
 20 25 30  
 Asp Xaa Xaa Ala Xaa Ser Xaa Xaa Thr Trp Xaa Val Leu Leu Ser Xaa  
 35 40 45  
 Cys Arg Xaa Trp Xaa Ala Trp Cys Xaa Xaa Xaa Xaa Arg Xaa Xaa Phe  
 50 55 60  
 Pro Xaa Xaa Pro Xaa Xaa Val Arg Xaa Tyr Leu Leu Xaa Leu Gln Xaa  
 65 70 75 80  
 Arg Gly Leu Xaa Val Asn Thr Xaa Gln Gln His Leu Ala Xaa Leu Asn  
 85 90 95  
 Met Xaa His Arg Arg Xaa Gly Leu Xaa Arg Xaa Xaa Asp Ser Xaa Xaa  
 100 105 110  
 Val Ser Leu Xaa Xaa Arg Arg Ile Arg Xaa Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Xaa Lys Gln Ala Leu Ala Phe Xaa Arg Xaa Asp Xaa Xaa Xaa  
 130 135 140  
 Xaa Xaa Xaa Leu Xaa Xaa Xaa Ser Xaa Xaa Xaa Xaa Asp Xaa Arg Xaa  
 145 150 155 160  
 Leu Ala Xaa Leu Gly Xaa Ala Tyr Asn Thr Leu Leu Arg Xaa Ser Glu  
 165 170 175

ES 2 710 708 T3

Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp Xaa Ser Xaa Thr Xaa Gly Gly Arg  
 180 185 190  
 Xaa Leu Ile His Xaa Xaa Xaa Thr Lys Thr Leu Val Ser Thr Xaa Gly  
 195 200 205  
 5 Val Glu Lys Ala Leu Ser Leu Xaa Xaa Thr Xaa Leu Xaa Glu Arg Trp  
 210 215 220  
 Xaa Ser Xaa Ser Gly Val Ala Xaa Xaa Xaa Xaa Xaa Tyr Leu Phe Cys  
 225 230 235 240  
 10 Xaa Xaa Xaa Xaa Xaa Gly Xaa Ala Xaa Pro Xaa Ala Xaa Xaa Xaa Leu  
 245 250 255  
 Ser Xaa Xaa Xaa Leu Xaa Xaa Ile Phe Xaa Xaa Xaa His Xaa Xaa Xaa  
 260 265 270  
 Xaa Gly Ala Lys Xaa Xaa Ser Gly Xaa Arg Tyr Xaa Xaa Trp Ser Gly  
 275 280 285  
 15 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 Xaa Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Xaa Thr Val Xaa Xaa  
 305 310 315 320  
 20 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Xaa Gly Ala Met Val  
 325 330 335  
 Arg Leu Leu Glu Xaa Xaa Xaa  
 340

25 <210> 9  
 <211> 343  
 <212> PRT  
 <213> Secuencia artificial

30 <220>  
 <223> Secuencia consenso Tre 3.1 85% (todos los clones)

35 <220>  
 <221> VARIANTE  
 <222> 3..5

40 <220>  
 <221> VARIANTE  
 <222> 8..10

45 <220>  
 <221> VARIANTE  
 <222> 16

50 <220>  
 <221> VARIANTE  
 <222> 18..19

55 <220>  
 <221> VARIANTE  
 <222> 22..26

60 <220>  
 <221> VARIANTE  
 <222> 28..29

65 <220>  
 <221> VARIANTE  
 <222> 31

70 <220>  
 <221> VARIANTE  
 <222> 34..35

75 <220>  
 <221> VARIANTE  
 <222> 37

<220>

<221> VARIANTE  
 <222> 39  
  
 5 <220>  
 <221> VARIANTE  
 <222> 43  
  
 10 <220>  
 <221> VARIANTE  
 <222> 48  
  
 15 <220>  
 <221> VARIANTE  
 <222> 53  
  
 20 <220>  
 <221> VARIANTE  
 <222> 57..60  
  
 25 <220>  
 <221> VARIANTE  
 <222> 62..63  
  
 30 <220>  
 <221> VARIANTE  
 <222> 66..67  
  
 35 <220>  
 <221> VARIANTE  
 <222> 69..70  
  
 40 <220>  
 <221> VARIANTE  
 <222> 73  
  
 45 <220>  
 <221> VARIANTE  
 <222> 80  
  
 50 <220>  
 <221> VARIANTE  
 <222> 84  
  
 55 <220>  
 <221> VARIANTE  
 <222> 88  
  
 60 <220>  
 <221> VARIANTE  
 <222> 94  
  
 65 <220>  
 <221> VARIANTE  
 <222> 98  
  
 <220>  
 <221> VARIANTE  
 <222> 102  
  
 <220>  
 <221> VARIANTE  
 <222> 105  
  
 <220>  
 <221> VARIANTE  
 <222> 107  
  
 <220>

<221> VARIANTE  
 <222> 111..112  
  
 5 <220>  
 <221> VARIANTE  
 <222> 116..117  
  
 10 <220>  
 <221> VARIANTE  
 <222> 122  
  
 15 <220>  
 <221> VARIANTE  
 <222> 131  
  
 20 <220>  
 <221> VARIANTE  
 <222> 138  
  
 25 <220>  
 <221> VARIANTE  
 <222> 140  
  
 30 <220>  
 <221> VARIANTE  
 <222> 142..147  
  
 35 <220>  
 <221> VARIANTE  
 <222> 149..151  
  
 40 <220>  
 <221> VARIANTE  
 <222> 153..156  
  
 45 <220>  
 <221> VARIANTE  
 <222> 158  
  
 50 <220>  
 <221> VARIANTE  
 <222> 163  
  
 55 <220>  
 <221> VARIANTE  
 <222> 166  
  
 60 <220>  
 <221> VARIANTE  
 <222> 174  
  
 65 <220>  
 <221> VARIANTE  
 <222> 177..180  
  
 <220>  
 <221> VARIANTE  
 <222> 182..183  
  
 <220>  
 <221> VARIANTE  
 <222> 185  
  
 <220>  
 <221> VARIANTE  
 <222> 187  
  
 <220>

<221> VARIANTE  
 <222> 189

5 <220>  
 <221> VARIANTE  
 <222> 193

10 <220>  
 <221> VARIANTE  
 <222> 197..199

15 <220>  
 <221> VARIANTE  
 <222> 207

20 <220>  
 <221> VARIANTE  
 <222> 216..217

25 <220>  
 <221> VARIANTE  
 <222> 219

30 <220>  
 <221> VARIANTE  
 <222> 221

35 <220>  
 <221> VARIANTE  
 <222> 225

40 <220>  
 <221> VARIANTE  
 <222> 227

45 <220>  
 <221> VARIANTE  
 <222> 232..236

50 <220>  
 <221> VARIANTE  
 <222> 242..243

55 <220>  
 <221> VARIANTE  
 <222> 245

60 <220>  
 <221> VARIANTE  
 <222> 247

65 <220>  
 <221> VARIANTE  
 <222> 249

<220>  
 <221> VARIANTE  
 <222> 251

<220>  
 <221> VARIANTE  
 <222> 253..255

<220>  
 <221> VARIANTE  
 <222> 258

<220>

ES 2 710 708 T3

<221> VARIANTE  
<222> 260

5 <220>  
<221> VARIANTE  
<222> 263

10 <220>  
<221> VARIANTE  
<222> 266..267

15 <220>  
<221> VARIANTE  
<222> 270..273

20 <220>  
<221> VARIANTE  
<222> 277

25 <220>  
<221> VARIANTE  
<222> 281

30 <220>  
<221> VARIANTE  
<222> 284..285

35 <220>  
<221> VARIANTE  
<222> 305

40 <220>  
<221> VARIANTE  
<222> 316

45 <220>  
<221> VARIANTE  
<222> 332

50 <220>  
<221> VARIANTE  
<222> 332

55 <220>  
<221> VARIANTE  
<222> 341..343

60 <400> 9

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

ES 2 710 708 T3

Xaa Leu Ile His Xaa Xaa Xaa Thr Lys Thr Leu Val Ser Thr Xaa Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Xaa Xaa Thr Xaa Leu Xaa Glu Arg Trp  
 210 215 220  
 5 Xaa Ser Xaa Ser Gly Val Ala Xaa Xaa Xaa Xaa Xaa Tyr Leu Phe Cys  
 225 230 235 240  
 Gln Xaa Xaa Ile Xaa Gly Xaa Ala Val Pro Xaa Ala Xaa Xaa Xaa Leu  
 245 250 255  
 10 Ser Xaa Asp Xaa Leu Arg Xaa Ile Phe Xaa Xaa Ala His Xaa Xaa Xaa  
 260 265 270  
 Xaa Gly Ala Lys Xaa Gly Ser Gly Xaa Arg Tyr Xaa Xaa Trp Ser Gly  
 275 280 285  
 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 15 Xaa Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Xaa Thr Val Glu Ser  
 305 310 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Xaa Gly Ala Met Val  
 325 330 335  
 20 Arg Leu Leu Glu Xaa Xaa Xaa  
 340

- <210> 10
- <211> 343
- <212> PRT
- 25 <213> Secuencia artificial
- <220>
- <223> Consenso Tre 3.1 100% (3 clones)
- 30 <220>
- <221> VARIANTE
- <222> 3
- 35 <220>
- <221> VARIANTE
- <222> 5
- 40 <220>
- <221> VARIANTE
- <222> 10
- 45 <220>
- <221> VARIANTE
- <222> 18..19
- 50 <220>
- <221> VARIANTE
- <222> 23
- 55 <220>
- <221> VARIANTE
- <222> 63
- 60 <220>
- <221> VARIANTE
- <222> 80
- 65 <220>
- <221> VARIANTE
- <222> 84
- <220>
- <221> VARIANTE
- <222> 88

	<220>	
	<221>	VARIANTE
	<222>	111
5		
	<220>	
	<221>	VARIANTE
	<222>	117
10		
	<220>	
	<221>	VARIANTE
	<222>	151
15		
	<220>	
	<221>	VARIANTE
	<222>	153
20		
	<220>	
	<221>	VARIANTE
	<222>	160
25		
	<220>	
	<221>	VARIANTE
	<222>	163
30		
	<220>	
	<221>	VARIANTE
	<222>	182..183
35		
	<220>	
	<221>	VARIANTE
	<222>	198
40		
	<220>	
	<221>	VARIANTE
	<222>	219
45		
	<220>	
	<221>	VARIANTE
	<222>	232
50		
	<220>	
	<221>	VARIANTE
	<222>	244
55		
	<220>	
	<221>	VARIANTE
	<222>	247
60		
	<220>	
	<221>	VARIANTE
	<222>	255
65		
	<220>	
	<221>	VARIANTE
	<222>	263
	<220>	
	<221>	VARIANTE
	<222>	266

ES 2 710 708 T3

<220>  
 <221> VARIANTE  
 <222> 272  
 5  
 <220>  
 <221> VARIANTE  
 <222> 319  
 10 <400> 10  
 Met Ser Xaa Leu Xaa Thr Leu His Gln Xaa Leu Ser Ala Leu Leu Val  
 1 5 10 15  
 Asp Xaa Xaa Ser Asp Glu Xaa Arg Lys Asn Leu Met Asp Val Leu Arg  
 20 25 30  
 15 Asp Arg Gln Ala Phe Ser Glu Arg Thr Trp Lys Val Leu Leu Ser Val  
 35 40 45  
 Cys Arg Thr Trp Ala Ala Trp Cys Xaa Leu Asn Asn Arg Lys Xaa Phe  
 50 55 60  
 20 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu His Leu Gln Xaa  
 65 70 75 80  
 Arg Gly Leu Xaa Val Asn Thr Xaa Leu Gln His Leu Ala Gln Leu Asn  
 85 90 95  
 Met Leu His Arg Arg Phe Gly Leu Pro Arg Pro Gly Asp Ser Xaa Ala  
 100 105 110  
 25 Val Ser Leu Val Xaa Arg Arg Ile Arg Arg Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Thr Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln  
 130 135 140  
 30 Val Arg Ala Leu Met Glu Xaa Ser Xaa Arg Gly Gln Asp Ile Arg Xaa  
 145 150 155 160  
 Leu Ala Xaa Leu Gly Val Ala Tyr Asn Thr Leu Leu Arg Xaa Ser Glu  
 165 170 175  
 Ile Ala Arg Ile Arg Xaa Xaa Asp Ile Ser Arg Thr Asp Gly Gly Arg  
 180 185 190  
 35 Met Leu Ile His Ile Xaa Arg Thr Lys Thr Leu Val Ser Thr Ala Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Gly Val Thr Xaa Leu Val Glu Arg Trp  
 210 215 220  
 40 Ile Ser Val Ser Gly Val Ala Xaa Asp Pro Asn Asn Tyr Leu Phe Cys  
 225 230 235 240  
 Xaa Val Arg Xaa Asn Gly Xaa Ala Val Pro Ser Ala Thr Ser Xaa Leu  
 245 250 255  
 Ser Thr Asp Val Leu Arg Xaa Ile Phe Xaa Ala Ala His Arg Leu Xaa  
 260 265 270  
 45 Tyr Gly Ala Lys Asp Gly Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly  
 275 280 285  
 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 50 Ser Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Thr Thr Val Xaa Ser  
 305 310 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val  
 325 330 335  
 Arg Leu Leu Glu Asp Gly Asp  
 340  
 55  
 <210> 11  
 <211> 343  
 <212> PRT  
 <213> Secuencia artificial  
 60  
 <220>  
 <223> uTre  
 <400> 11  
 65 Met Ser Ile Leu Leu Thr Leu His Gln Ser Leu Ser Ala Leu Leu Val  
 1 5 10 15  
 Asp Ala Thr Ser Asp Glu Ala Arg Lys Asn Leu Met Asp Val Leu Arg  
 20 25 30

ES 2 710 708 T3

Asp Arg Gln Ala Phe Ser Glu Arg Thr Trp Lys Val Leu Leu Ser Val  
 35 40 45  
 Cys Arg Thr Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe  
 50 55 60  
 5 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu His Leu Gln Ala  
 65 70 75 80  
 Arg Gly Leu Ala Val Asn Thr Ile Leu Gln His Leu Ala Gln Leu Asn  
 85 90 95  
 10 Met Leu His Arg Arg Phe Gly Leu Pro Arg Pro Gly Asp Ser Asp Ala  
 100 105 110  
 Val Ser Leu Val Met Arg Arg Ile Arg Arg Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Thr Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln  
 130 135 140  
 15 Val Arg Ala Leu Met Glu Asn Ser Glu Arg Gly Gln Asp Ile Arg Thr  
 145 150 155 160  
 Leu Ala Leu Leu Gly Val Ala Tyr Asn Thr Leu Leu Arg Val Ser Glu  
 165 170 175  
 20 Ile Ala Arg Ile Arg Ile Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg  
 180 185 190  
 Met Leu Ile His Ile Ser Arg Thr Lys Thr Leu Val Ser Thr Ala Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp  
 210 215 220  
 25 Ile Ser Val Ser Gly Val Ala Ser Asp Pro Asn Asn Tyr Leu Phe Cys  
 225 230 235 240  
 Gln Val Arg Ile Asn Gly Val Ala Val Pro Ser Ala Thr Ser Arg Leu  
 245 250 255  
 30 Ser Thr Asp Val Leu Arg Lys Ile Phe Glu Ala Ala His Arg Leu Ile  
 260 265 270  
 Tyr Gly Ala Lys Asp Gly Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly  
 275 280 285  
 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 35 Ser Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Thr Thr Val Glu Ser  
 305 310 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val  
 325 330 335  
 40 Arg Leu Leu Glu Asp Gly Asp  
 340

<210> 12

<211> 343

<212> PRT

45 <213> Secuencia artificial

<220>

<223> recombinasa a medida

50 <400> 12

Met Ser Ser Leu Gln Thr Leu His Gln Asn Leu Ser Ala Leu Leu Val  
 1 5 10 15  
 Asp Val Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Val Leu Arg  
 20 25 30  
 55 Asp Arg Gln Ala Phe Ser Glu Arg Thr Trp Lys Val Leu Leu Ser Val  
 35 40 45  
 Cys Arg Thr Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Arg Phe  
 50 55 60  
 60 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu His Leu Gln Thr  
 65 70 75 80  
 Arg Gly Leu Thr Val Asn Thr Ile Leu Gln His Leu Ala Gln Leu Asn  
 85 90 95  
 Met Leu His Arg Arg Phe Gly Leu Pro Arg Pro Gly Asp Ser Asp Ala  
 100 105 110  
 65 Val Ser Leu Val Ile Arg Arg Ile Arg Arg Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Thr Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln  
 130 135 140

ES 2 710 708 T3

Val Arg Ala Leu Met Glu Asn Ser Asp Arg Gly Gln Asp Ile Arg Ser  
 145 150 155 160  
 Leu Ala Phe Leu Gly Val Ala Tyr Asn Thr Leu Leu Arg Ile Ser Glu  
 165 170 175  
 5 Ile Ala Arg Ile Arg Val Arg Asp Ile Ser Arg Thr Asp Gly Arg  
 180 185 190  
 Met Leu Ile His Ile Ser Arg Thr Lys Thr Leu Val Ser Thr Ala Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp  
 210 215 220  
 10 Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys  
 225 230 235 240  
 Pro Val Arg Val Asn Gly Ala Ala Val Pro Ser Ala Thr Ser Gln Leu  
 245 250 255  
 15 Ser Thr Asp Val Leu Arg Gly Ile Phe Glu Ala Ala His Arg Leu Val  
 260 265 270  
 Tyr Gly Ala Lys Asp Gly Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly  
 275 280 285  
 20 His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val  
 290 295 300  
 Ser Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Thr Thr Val Glu Ser  
 305 310 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val  
 325 330 335  
 25 Arg Leu Leu Glu Asp Gly Asp  
 340

<210> 13  
 <211> 343  
 30 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> recombinasa a medida

35 <400> 13  
 Met Ser Ser Leu Leu Thr Leu His Gln Ser Leu Ser Ala Leu Leu Val  
 1 5 10 15  
 40 Asp Val Ala Ser Asp Glu Ala Arg Lys Asn Leu Met Asp Val Leu Arg  
 20 25 30  
 Asp Arg Gln Ala Phe Ser Glu Arg Thr Trp Lys Val Leu Ser Val  
 35 40 45  
 Cys Arg Thr Trp Ala Ala Trp Cys Glu Leu Asn Asn Arg Lys Arg Phe  
 50 55  
 45 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu His Leu Gln Thr  
 65 70 75 80  
 Arg Gly Leu Thr Val Asn Thr Val Leu Gln His Leu Ala Gln Leu Asn  
 85 90 95  
 50 Met Leu His Arg Arg Phe Gly Leu Pro Arg Pro Gly Asp Ser Asn Ala  
 100 105 110  
 Val Ser Leu Val Ile Arg Arg Ile Arg Arg Glu Asn Val Asp Ala Gly  
 115 120 125  
 Glu Arg Thr Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln  
 130 135 140  
 55 Val Arg Ala Leu Met Glu Asp Ser Asp Arg Gly Gln Asp Ile Arg Ser  
 145 150 155 160  
 Leu Ala Phe Leu Gly Val Ala Tyr Asn Thr Leu Leu Arg Ile Ser Glu  
 165 170 175  
 60 Ile Ala Arg Ile Arg Val Arg Asp Ile Ser Arg Thr Asp Gly Arg  
 180 185 190  
 Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly  
 195 200 205  
 Val Glu Lys Ala Leu Ser Leu Gly Val Thr Glu Leu Val Glu Arg Trp  
 210 215 220  
 65 Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys  
 225 230 235 240  
 Gln Val Arg Ile Asn Gly Val Ala Val Pro Ser Ala Thr Ser Gln Leu  
 245 250 255

ES 2 710 708 T3

Ser Thr Asp Val Leu Arg Gly Ile Phe Ala Ala Ala His Arg Leu Ile  
 Tyr Gly Ala Lys Asp Gly Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly  
 5 His Ser Ala Arg Val Gly Ala Arg Asp Met Ala Arg Ala Gly Val  
 Ser Ile Ala Glu Ile Met Gln Ala Gly Gly Trp Thr Thr Val Asp Ser  
 305 290 275 260 310 295 280 265 300 315 320  
 Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val  
 10 Arg Leu Leu Glu Asp Ser Asp 330 335 340

<210> 14  
 <211> 351  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> uTre con señal de localización nuclear

<220>  
 <221> SEÑAL  
 <222> 2..9  
 <223> secuencia de localización nuclear (NLS)

<400> 14  
 Met Val Pro Lys Lys Lys Arg Lys Val Ser Ile Leu Leu Thr Leu His  
 1 5 10 15  
 30 Gln Ser Leu Ser Ala Leu Leu Val Asp Ala Thr Ser Asp Glu Ala Arg  
 Lys Asn Leu Met Asp Val Leu Arg Asp Arg Gln Ala Phe Ser Glu Arg  
 35 Thr Trp Lys Val Leu Leu Ser Val Cys Arg Thr Trp Ala Ala Trp Cys  
 50 55 60  
 Lys Leu Asn Asn Arg Lys Trp Phe Pro Ala Glu Pro Glu Asp Val Arg  
 65 70 75 80  
 Asp Tyr Leu Leu His Leu Gln Ala Arg Gly Leu Ala Val Asn Thr Ile  
 85 90 95  
 40 Leu Gln His Leu Ala Gln Leu Asn Met Leu His Arg Arg Phe Gly Leu  
 100 105 110  
 Pro Arg Pro Gly Asp Ser Asp Ala Val Ser Leu Val Met Arg Arg Ile  
 115 120 125  
 45 Arg Arg Glu Asn Val Asp Ala Gly Glu Arg Thr Lys Gln Ala Leu Ala  
 130 135 140  
 Phe Glu Arg Thr Asp Phe Asp Gln Val Arg Ala Leu Met Glu Asn Ser  
 145 150 155 160  
 Glu Arg Gly Gln Asp Ile Arg Thr Leu Ala Leu Leu Gly Val Ala Tyr  
 165 170 175  
 50 Asn Thr Leu Leu Arg Val Ser Glu Ile Ala Arg Ile Arg Ile Lys Asp  
 180 185 190  
 Ile Ser Arg Thr Asp Gly Gly Arg Met Leu Ile His Ile Ser Arg Thr  
 195 200 205  
 Lys Thr Leu Val Ser Thr Ala Gly Val Glu Lys Ala Leu Ser Leu Gly  
 210 215 220  
 Val Thr Lys Leu Val Glu Arg Trp Ile Ser Val Ser Gly Val Ala Ser  
 225 230 235 240  
 Asp Pro Asn Asn Tyr Leu Phe Cys Gln Val Arg Ile Asn Gly Val Ala  
 245 250 255  
 60 Val Pro Ser Ala Thr Ser Arg Leu Ser Thr Asp Val Leu Arg Lys Ile  
 260 265 270  
 Phe Glu Ala Ala His Arg Leu Ile Tyr Gly Ala Lys Asp Gly Ser Gly  
 275 280 285  
 Gln Arg Tyr Leu Ala Trp Ser Gly His Ser Ala Arg Val Gly Ala Ala  
 290 295 300  
 65 Arg Asp Met Ala Arg Ala Gly Val Ser Ile Ala Glu Ile Met Gln Ala  
 305 310 315 320  
 Gly Gly Trp Thr Thr Val Glu Ser Val Met Asn Tyr Ile Arg Asn Leu

ES 2 710 708 T3

Asp Ser Glu Thr <sup>325</sup>Gly Ala Met Val Arg <sup>330</sup>Leu Leu Glu Asp Gly <sup>335</sup>Asp  
                  340                  345                  350

5

10

## REIVINDICACIONES

1. Ácido nucleico que codifica una recombinasa a medida, cuya recombinasa a medida es capaz de recombinar secuencias diana asimétricas de SEQ ID NO: 1 dentro de la repetición terminal larga del ADN proviral de una pluralidad de cepas de VIH-1, en el que la secuencia de aminoácidos de la recombinasa a medida tiene al menos un 95% de identidad de secuencia, más preferiblemente, un 99% de identidad de secuencia con una secuencia según la SEQ ID NO: 10, en el que dicha recombinasa a medida comprende los intercambios de aminoácidos definidos en comparación con la SEQ ID NO: 6 de V7L, P12S, P15L, M30V, H40R, M44V, S51T, Y77H, K86N, Q89L, G93A, S108G, C155G, A175S, A249V, R259D, E262R, T268A, D278G, P307A, N317T, I320S.
2. Ácido nucleico, según la reivindicación 1, en el que la recombinasa a medida comprende una secuencia de aminoácidos según la SEQ ID NO: 10.
3. Ácido nucleico, según cualquiera de las reivindicaciones 1 ó 2, en el que la recombinasa a medida comprende la secuencia de aminoácidos según las SEQ ID NO: 11-13, o una combinación de cualquiera de dichas secuencias, en el que una combinación de cualquiera de dichas secuencias comprende una parte C-terminal de cualquiera de estas secuencias, y la parte N-terminal de cualquier otra de estas secuencias, en el que la combinación comprende la SEQ ID NO: 10.
4. Ácido nucleico, según cualquiera de las reivindicaciones 1 a 3, en el que la recombinasa a medida es muy específica para la recombinación de la SEQ ID NO: 1, ya que dicha recombinasa a medida no recombina secuencias de loxP (SEQ ID NO: 4) o loxH (SEQ ID NO: 5) con actividad detectable.
5. Recombinasa a medida codificada por el ácido nucleico, según cualquiera de las reivindicaciones 1 a 4, que se expresa opcionalmente como una proteína de fusión; o una célula transformada que comprende el ácido nucleico, según cualquiera de las reivindicaciones 1 a 4, en que dicha célula es preferiblemente una célula madre de linaje hematopoyético.
6. Composición farmacéutica que comprende un ácido nucleico, según cualquiera de las reivindicaciones 1 a 4, una recombinasa a medida, según la reivindicación 5, y/o una célula transformada, según la reivindicación 5.
7. Composición farmacéutica, según la reivindicación 6, en la que la composición farmacéutica es para usar en el tratamiento o la prevención de una infección por retrovirus en un sujeto, en la que el retrovirus es el VIH, en particular el VIH-1, en la que la composición farmacéutica se formula opcionalmente para la administración a un sujeto, si el ADN proviral encontrado en una muestra obtenida del sujeto comprende la secuencia diana asimétrica de SEQ ID NO: 1 sobre la que se ha seleccionado la recombinasa.
8. Procedimiento para preparar un ácido nucleico o un vector de expresión que codifica una recombinasa a medida bien tolerada, cuya recombinasa a medida es capaz de recombinar secuencias diana asimétricas dentro de la LTR del ADN proviral de una pluralidad de cepas de retrovirus, en el que el retrovirus es el VIH 1, que comprende las etapas de
- (a) identificar secuencias con una homología de al menos el 30% con la secuencia del hemisitio izquierdo y la secuencia del hemisitio derecho de al menos un sitio diana de recombinasa conocido en la secuencia de la LTR de ADN proviral de una pluralidad de cepas de retrovirus, en el que las secuencias homólogas están separadas por un espaciador de 5-12 nucleótidos, y en el que la secuencia diana asimétrica tiene la secuencia expuesta como la SEQ ID NO: 1, y se encuentra en una pluralidad de cepas de retrovirus;
- (b) identificar dos secuencias, en el que la primera secuencia corresponde a la secuencia de la secuencia diana asimétrica de la etapa (a) homóloga al hemisitio izquierdo de dicho sitio diana conocido y se denomina "secuencia de hemisitio 1", y en el que la segunda secuencia corresponde a la secuencia de la secuencia diana asimétrica de la etapa (a) homóloga al hemisitio derecho y se denomina como "secuencia de hemisitio 2";
- (c) determinar los nucleótidos dentro de las secuencias de la etapa (b) que se desvían de las secuencias de hemisitio izquierdo y de hemisitio derecho homólogas correspondientes de dicho al menos un sitio diana homólogo conocido de la etapa (a);
- (d) generar un primer subconjunto de dos ácidos nucleicos diana que comprenden secuencias diana, en el que la primera secuencia diana se designa subsitio 1 y comprende, adyacentes entre sí y en orden 5' a 3', la secuencia de hemisitio 1 de la etapa (b), la secuencia espaciadora de la secuencia diana asimétrica y una repetición invertida de la secuencia de hemisitio 1, y en el que la segunda secuencia diana se designa subsitio 2 y comprende, adyacentes entre sí y en orden 5' a 3', una repetición invertida de la secuencia de hemisitio 2, la secuencia espaciadora de la secuencia diana asimétrica y la secuencia de hemisitio 2 de la etapa (b);
- (e) generar un segundo subconjunto de los ácidos nucleicos diana que comprende secuencias diana modificadas en base a las secuencias diana en el primer subconjunto de la etapa (d), en el que, en las secuencias basadas en el subsitio 1, en la secuencia de hemisitio izquierdo, una parte de los nucleótidos que se desvían de la correspondiente secuencia de hemisitio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontrados en dicho sitio diana conocido, hasta que dicha secuencia de hemisitio contiene uno, dos o tres nucleótidos que se desvían de dicho sitio diana conocido, en el que el hemisitio derecho de dicha secuencia diana modificada está formado por una repetición invertida de

- dicha secuencia de hemisio izquierdo modificada, que está separada de dicha secuencia de hemisio izquierdo modificada por la secuencia espaciadora de la secuencia diana asimétrica, y en el que, en secuencias basadas en el subsitio 2, en la secuencia de hemisio derecho, una parte de los nucleótidos que se desvían de la correspondiente secuencia de hemisio homóloga de dicho al menos un sitio diana conocido de la etapa (a) se sustituye por los nucleótidos nativos encontrados en dicho sitio diana conocido, hasta que dicha secuencia de hemisio contiene uno, dos o tres nucleótidos que se desvían de dicho sitio diana conocido, en el que el hemisio izquierdo de dicha secuencia diana modificada está formada por una repetición invertida de dicha secuencia de hemisio derecho modificada, que está separada de dicha secuencia de hemisio derecho modificada por la secuencia espaciadora de la secuencia diana asimétrica,
- de manera que en todas las secuencias de hemisio modificadas procedentes de una secuencia diana del primer subconjunto de la etapa (d) tomadas en conjunto, se pueden encontrar todos los nucleótidos que se desvían, mientras que ninguna de dichas secuencias de hemisio modificadas solas comprende todos los nucleótidos que se desvían,
- (f) aplicar por separado la evolución molecular dirigida en al menos una recombinasa que reconoce un sitio diana homólogo conocido de acuerdo con la etapa (a) usando cada ácido nucleico del segundo subconjunto obtenido en la etapa (e) como sustrato;
- (g) barajar las bibliotecas de recombinasas desarrolladas en la etapa (f), en el que todas las bibliotecas de recombinasas desarrolladas en secuencias basadas en el subsitio 1 se combinan y se barajan, y en el que todas las bibliotecas de recombinasas desarrolladas en secuencias basadas en el subsitio 2 se combinan y se barajan;
- (h) aplicar la evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en las bibliotecas barajadas obtenidas en la etapa (g) usando cada ácido nucleico del subconjunto según la etapa (d) como sustrato;
- (i) barajar las bibliotecas de recombinasas desarrolladas en la etapa (h);
- (j) aplicar evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en la biblioteca barajada obtenida en la etapa (g) usando un ácido nucleico que comprende la secuencia diana asimétrica de la etapa (a) como sustrato, hasta que se obtiene al menos una recombinasa que es activa en la secuencia diana asimétrica dentro de la LTR del ADN del retrovirus de la etapa (a);
- (k) aislar el ácido nucleico que codifica dicha al menos una recombinasa obtenida en la etapa (j) de la biblioteca y clonarlo en un vector de evolución que permite la selección negativa de recombinasas a medida que recombinan el sitio diana conocido de acuerdo con la etapa (a), en el que se lleva a cabo la selección negativa para la recombinación de loxP y loxH, obteniendo de este modo una biblioteca;
- (l) aplicar la evolución molecular dirigida, preferiblemente, la evolución de proteínas unidas a sustrato, en la biblioteca obtenida en la etapa (k);
- (m) barajar las bibliotecas obtenidas en la etapa (l);
- (n) aislar el ácido nucleico que codifica dicha al menos una recombinasa a medida obtenida en la etapa (m) y clonarlo en un vector para la expresión de la recombinasa codificada y un marcador seleccionable en una célula humana, obteniendo de este modo una biblioteca de vectores,
- (o) transformar células humanas, preferiblemente células T humanas, con dicha biblioteca de vectores obtenida en la etapa (n);
- (p) cultivar las células que expresan dicho marcador seleccionable durante al menos 1 semana y seleccionar el marcador seleccionable con una alta expresión;
- (q) aislar el ácido o ácidos nucleicos que codifican la recombinasa de las células que expresan dicho marcador seleccionable obtenido en la etapa (p);
- (r) seleccionar un ácido nucleico que codifica una recombinasa capaz de la recombinación de la secuencia diana asimétrica de la etapa (a);
- (s) aislar el ácido nucleico que codifica dicha al menos una recombinasa obtenida en la etapa (r) de la biblioteca; y,
- (t) opcionalmente, clonar el ácido nucleico obtenido en la etapa (s) en un vector de expresión adecuado.
9. Procedimiento, según la reivindicación 8, en el que dicha al menos una recombinasa conocida cuya secuencia diana se utiliza en la etapa (a) y sobre la cual se aplica la evolución molecular dirigida en la etapa (f) es parte de una biblioteca de recombinasas.
10. Procedimiento, según cualquiera de las reivindicaciones 8 a 9, en el que el vector de expresión en la etapa (t) se selecciona del grupo que consiste en vectores retrovirales, vectores lentivirales, vectores de spumavirus, vectores adenovirales y vectores de virus adeno-asociados.
11. Procedimiento para preparar una recombinasa a medida, en el que dicho procedimiento comprende el procedimiento, según cualquiera de las reivindicaciones 8 a 10, y la etapa adicional de expresar la recombinasa a medida a partir del ácido nucleico que codifica la recombinasa insertado en el vector de expresión en una célula huésped adecuada, en el que la recombinasa se expresa opcionalmente como un polipéptido de fusión que comprende la secuencia de aminoácidos de la recombinasa a medida.
12. Procedimiento para preparar una célula transformada, en el que dicho procedimiento comprende el procedimiento, según cualquiera de las reivindicaciones 8 a 11, y la etapa adicional de introducir el vector de expresión en una célula *in vitro*, en el que la célula es preferiblemente una célula madre adulta.

13. Procedimiento, según cualquiera de las reivindicaciones 8 a 12, que comprende además la etapa de preparar el vector de expresión, la recombinasa o célula como una composición farmacéutica.

5 14. Ácido nucleico, tal como se puede obtener a partir del procedimiento, según cualquiera de las reivindicaciones 8 a 13, en el que el ácido nucleico es un ácido nucleico, según cualquiera de las reivindicaciones 1-4.

Fig. 1

- a) Cre SEQ ID NO:6
- b) consenso Tre común 100% (WO 2011/147590) SEQ ID NO:7
- c) consenso Tre 3.0 100% (WO 2011/147590) SEQ ID NO:8
- d) consenso Tre 3.1 85% (todos los clones) SEQ ID NO:9
- e) consenso Tre 3.1 100% (3 clones) SEQ ID NO:10
- f) secuencia de uTre SEQ ID NO:11

```

                31
1 3   7 10 12 15   23   30   40  44   51
MSNLLTVHQNLPALPVDATSEVVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLN a
MSXXXTXXXLXALXXDXXSDXXXXLXXXXRDXAXSXTWXXLLSXXCRXXAWCXXX b
MSXXXTXXXLSALLXDXXSDXXXXLXXXXRDXAXSXTWXXVLLSXXCRXXAWCXXX c
MSXXXTLXXXLSALLXDXXSDXXXXLXXVXRDXAXSXTWXXVLLSXXCRTWXXAWCXXX d
MSXLXTLHQXLSALLVDXXSDEXRKNLMDVLRDRQAFSERTWKVLLSVCRTWAAWCXLN e
MSILLTLHQXLSALLVDATSEARKNLMDVLRDRQAFSERTWKVLLSVCRTWAAWCKLN f

60                77       86 89  93       102  108111
NRKWFPAEPEDVRDYLLYLQARGLAVKTIQOHLGQLNMLHRRSGLPRPDSNAVSLVMRR a
XRXFXFPXXPXXVRXYLLXLQXRGLXVXTXQOHLXXLNMXHRRXGLXRXDXXXXVSLXXXR b
XRXFXFPXXPXXVRXYLLXLQXRGLXVNTXQOHLAXLNMXHRRXGLXRXDXSXXVSLXXXR c
XRXFXFPXXPXXVRXYLLHLQXRGLXVNTXLOHLAXLNMXHRRXGLXRXGDSXXVSLXXXR d
NRKWFPAEPEDVRDYLLHLQXRGLXVNTXLOHLAQLNMLHRRFGLPRPGDSXAVSLVXRR e
NRKWFPAEPEDVRDYLLHLQARGLAVNTIQLQHLAQLNMLHRRFGLPRPGSDAVSLVMRR f

                155       163       175
122       131       147  153  160  166  174
IRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAEIAR a
IRXENVDAGERXKQALAFXXDXXXXXXXXLXXXSXXXXXRXLAXLGXAYNTLLRXXEXXX b
IRXENVDAGERXKQALAFXXDXXXXXXXXLXXXSXXXXXRXLAXLGXAYNTLLRXXSEXXX c
IRXENVDAGERXKQALAFXXDXXXXXXXXLXXXSXXGXDXRTLAXLGXAYNTLLRXXSEXXX d
IRRENVDAGERTKQALAFERTDFDQVRALMEXSXRQDIXRLAXLGVAYNTLLRXXSEIAR e
IRRENVDAGERTKQALAFERTDFDQVRALMENSERQDIXRTLALLGVAYNTLLRVSEIAR f

182                198                232
IRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALS LGVTKLVERWISVSGVADDPNNYLF a
XRXDXSXTXGGRXLIHXXXTKTLVSTXGVEKALS LXXTXLXERWXSXSGVAXXXXXYLF b
XRXDXSXTXGGRXLIHXXXTKTLVSTXGVEKALS LXXTXLXERWXSXSGVAXXXXXYLF c
XRXDXSXTXGGRXLIHXXXTKTLVSTXGVEKALS LXXTXLXERWXSXSGVAXXXXXYLF d
IRXXDISRTDGGRMLIHIIXRTKTLVSTAGVEKALS LGVTKLVERWISVSGVAXDPNNYLF e
IRIKDISRTDGGRMLIHISRTTKTLVSTAGVEKALS LGVTKLVERWISVSGVASDPNNYLF f

244                260263
241       249       255 259262  268       278
CRVRKNGVAAPSATSQLS TRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDM a
CXXXXXGXAXPXAXXXLSXXXLXXIFXXXHXXXXGAKXXSGXRYXXWSGHSARVGAARDM b
CXXXXXGXAXPXAXXXLSXXXLXXIFXXXHXXXXGAKXXSGXRYXXWSGHSARVGAARDM c
CQXXIXGXAVPXAXXXLSXDLRXIFXXAHXXXXGAKXGSGXRYXXWSGHSARVGAARDM d
CXVRXNGXAVPSATSXLS TDVLRXIFXAAHRLXYGAKDGSGQRYLAWSGHSARVGAARDM e
CQVRINGVAVPSATSRLSTDVLRKIFEAAHRLIYGAKDGSGQRYLAWSGHSARVGAARDM f

319
300       307       317320
ARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSE TGAMVRLLEDGD* a
ARAGVXIXEIMQAGGWXTVXXVMNYIRNLDSE XGAMVRLLEXXX* b
ARAGVXIAEIMQAGGWXTVXXVMNYIRNLDSE XGAMVRLLEXXX* c
ARAGVXIAEIMQAGGWXTVESVMNYIRNLDSE XGAMVRLLEXXX* d
ARAGVSIAEIMQAGGWTTVXSVMNYIRNLDSE TGAMVRLLEDGD* e
ARAGVSIAEIMQAGGWTTVESVMNYIRNLDSE TGAMVRLLEDGD* f

```

Fig. 2

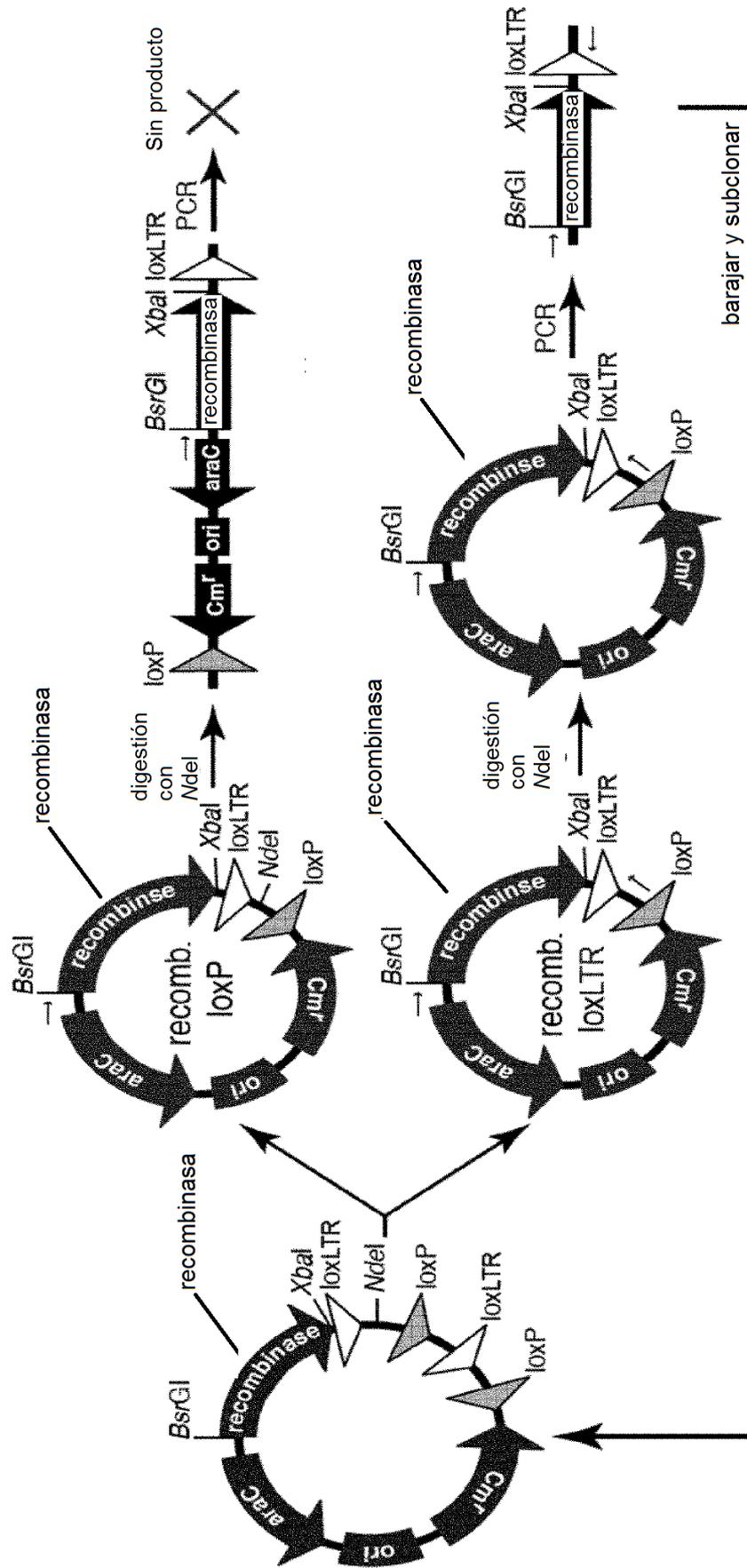


Fig. 3

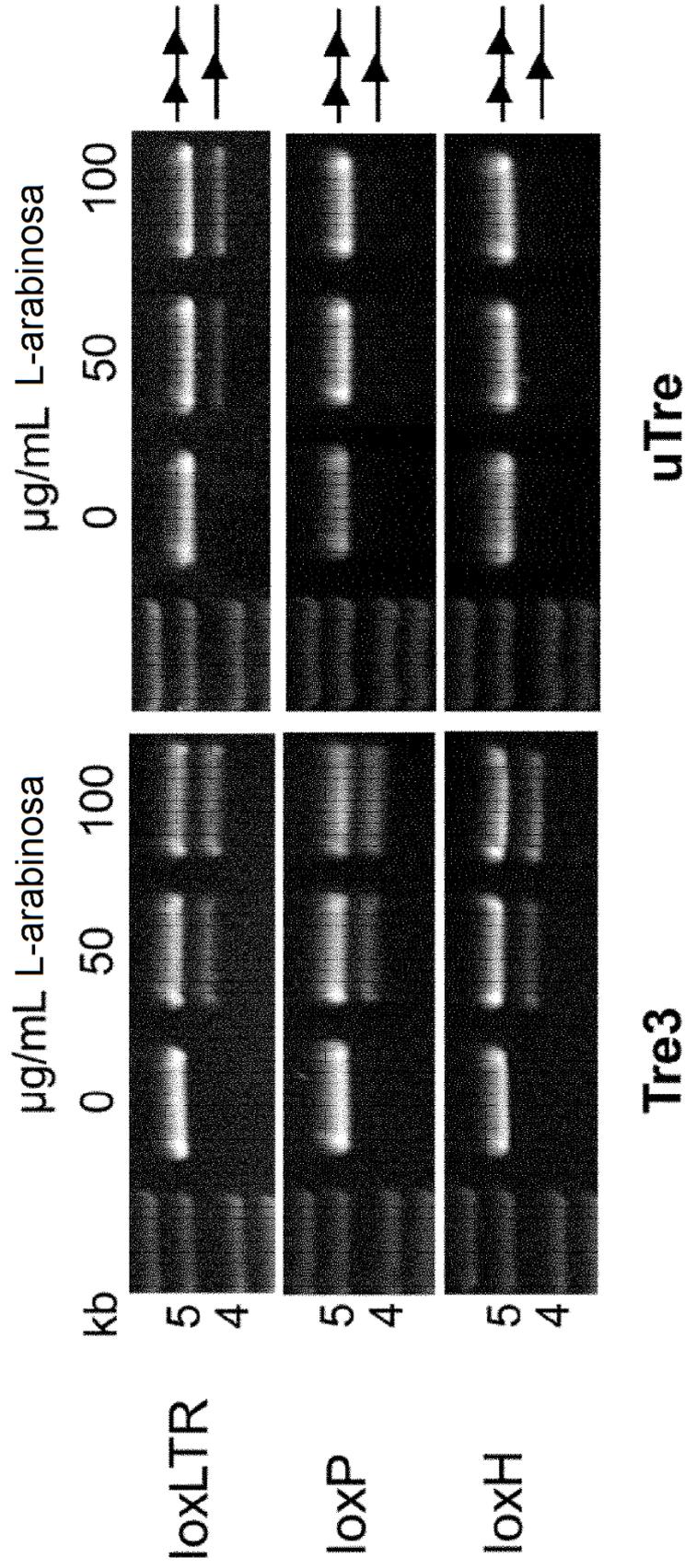


Fig. 4 A

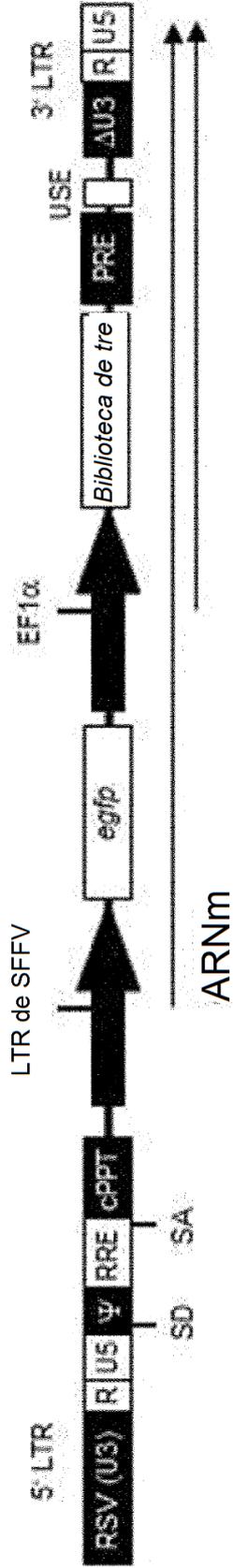


Fig. 4 B

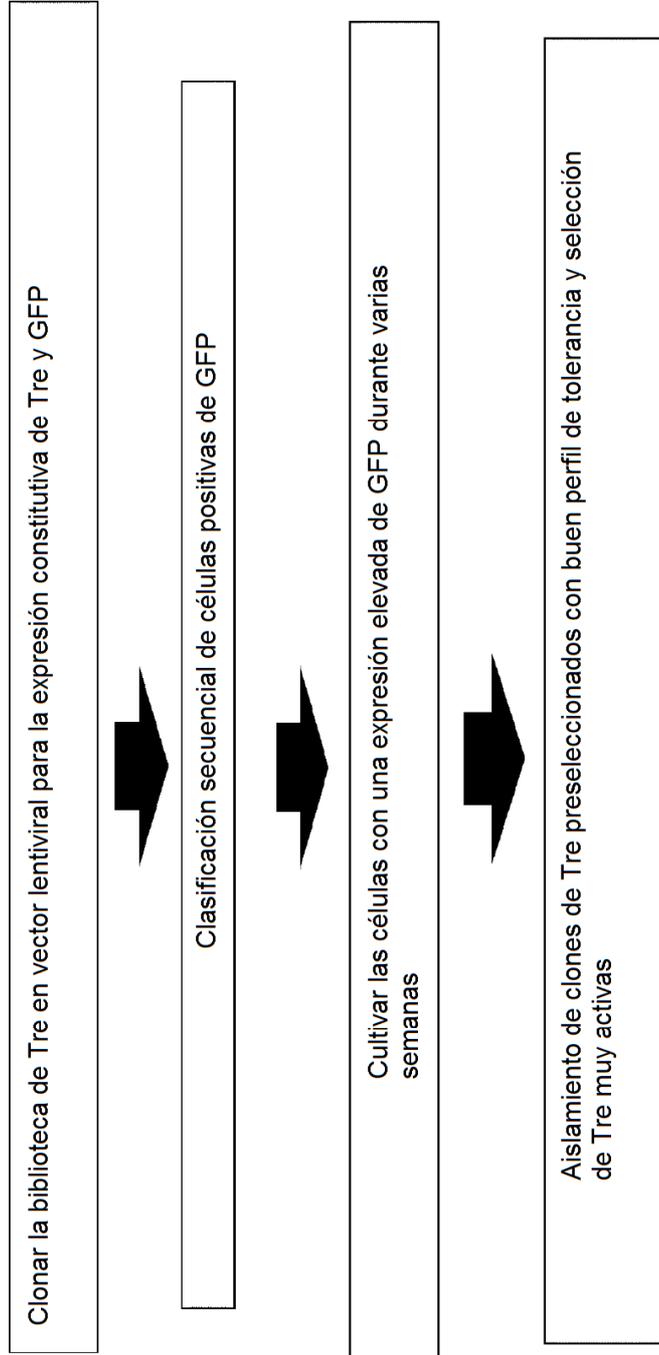
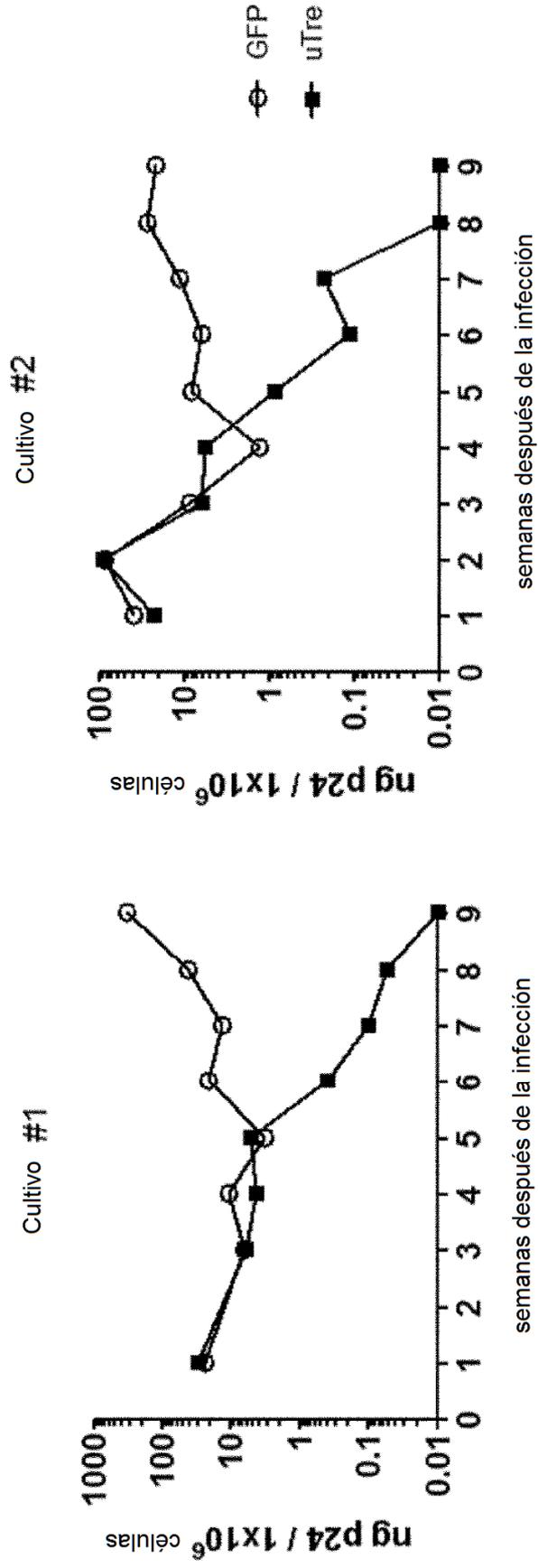


Fig. 5 A



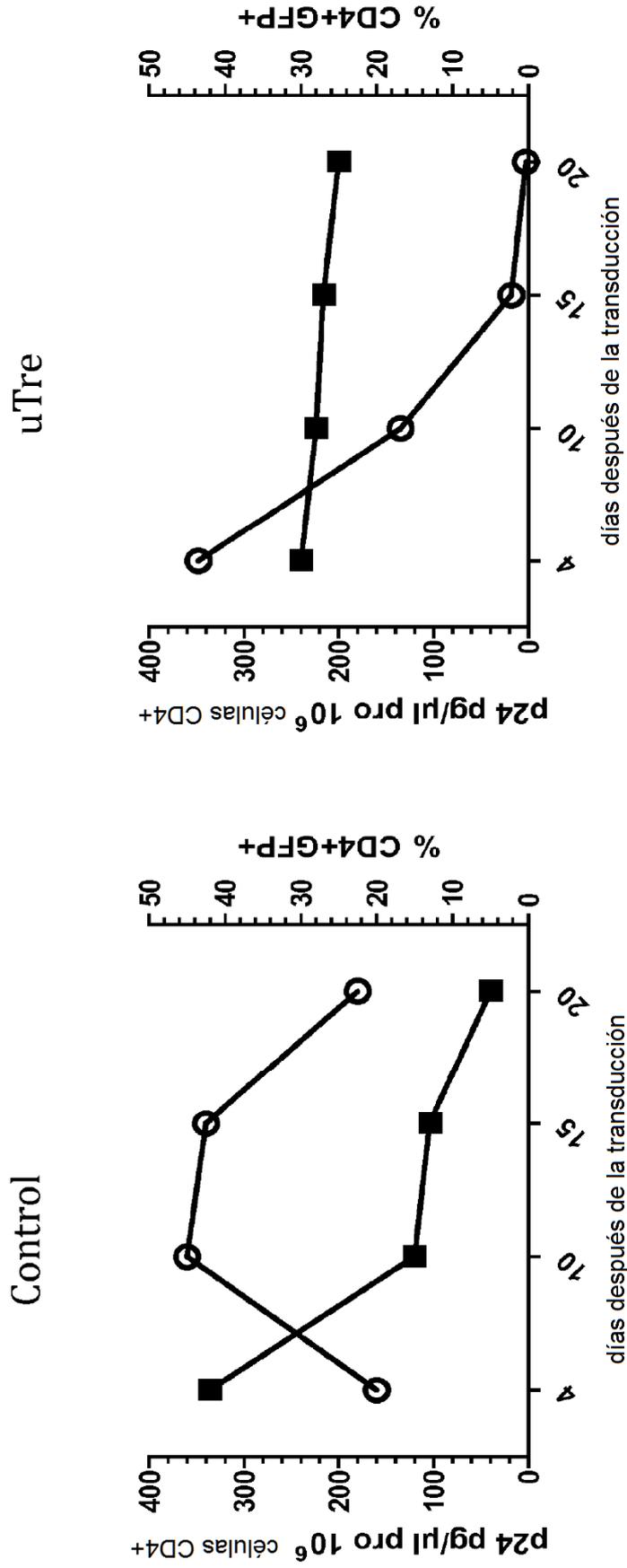
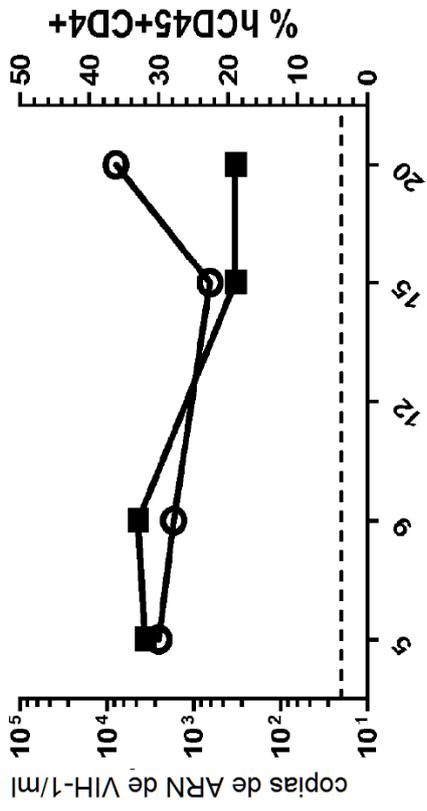
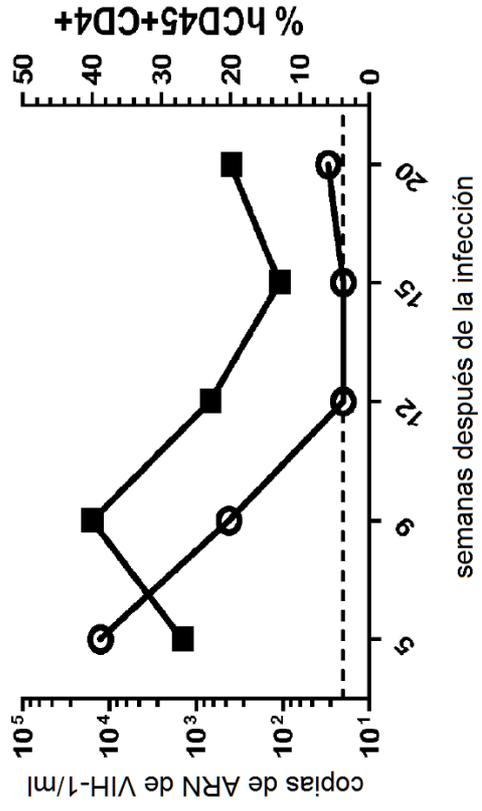


Fig. 6

Fig. 7 Control



Animal#1 / uTre



Animal#2 / uTre

