

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 422 424**

51 Int. Cl.:

C12N 9/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.02.2005 E 05722807 (4)**

97 Fecha y número de publicación de la concesión europea: **24.04.2013 EP 1844155**

54 Título: **Serina recombinasas específicas de sitio y métodos para su uso**

30 Prioridad:

02.02.2005 US 49552

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

11.09.2013

73 Titular/es:

**INTREXON CORPORATION (100.0%)
1750 Kraft Drive, Suite 1400
Blacksburg, VA 24060 , US**

72 Inventor/es:

PADIDAM, MALLA

74 Agente/Representante:

PÉREZ BARQUÍN, Eliana

ES 2 422 424 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

Serina recombinasas específicas de sitio y métodos para su uso

5 **Declaración sobre investigación con patrocinio federal**

Esta invención se realizó con ayuda del gobierno con la subvención nº 70NANB1H3062, otorgada por el National Institute of Standards and Technology, Advance Technology Program. El Gobierno puede tener ciertos derechos sobre la invención.

10

Campo de la invención

La presente invención se refiere al campo de la ingeniería genética. Específicamente, la invención se refiere a composiciones y métodos para integrar, deletar, invertir, intercambiar y translocar de un modo específico de sitio un polinucleótido en el genoma de una célula. La invención se refiere también a enzima, polinucleótidos, polipéptidos y construcciones de vectores.

15

Antecedentes de la invención

20

Muchos bacteriófagos y plásmidos integradores codifican sistemas de recombinación específicos de sitio que permiten la incorporación estable de su genoma en los de sus huéspedes y la escisión del genoma a partir del genoma del huésped. En estos sistemas, los requisitos mínimos para la reacción de recombinación son una enzima recombinasa que cataliza el evento de recombinación y dos sitios de recombinación (Sadowski (1986) J. Bacteriol. 165: 341-347; Sadowski (1993) FASEB J. 7: 760-767). Para los sistemas de integración en fagos, estos se denominan sitios de unión (*att*), con un elemento *attP* del ADN del fago y el elemento *attB* presente en el genoma bacteriano. Los dos sitios de unión pueden compartir una identidad de secuencia de tan solo unos pocos pares de bases. La proteína recombinasa se une a ambos sitios *att* y cataliza un intercambio conservador y recíproco de hebras de ADN que tiene como resultado la integración del fago circular o ADN plasmídico en el ADN del huésped. Para una reacción eficiente pueden ser necesarios fagos o factores del huésped adicionales, como la proteína de unión al ADN IHF, factor de integración del huésped (Friedman (1988) Cell 55: 545-554; Finkel y Johnson (1992) Mol. Microbiol. 6: 3257-3265). Las integrasas de los fagos, asociadas con otros factores del huésped y/o el fago, también escinden el genoma del fago a partir del genoma bacteriano durante la fase lítica del ciclo de crecimiento de los bacteriófagos. Con el fin de aclarar la relevancia y la función de determinados genes de interés se han desarrollado diversos métodos que permiten la manipulación de genomas de mamíferos. Entre ellos, el desarrollo de cepas de ratones transgénicos y de tecnologías dirigidas a genes ha resultado ser particularmente útil (Brandon, E. P., Idzerda, R. L. y McKnight, G. S. (1995) Curr Biol, 5, 625-34; Brandon, E. P., Idzerda, R. L. y McKnight, G. S. (1995) Curr Biol, 5, 758-65). Estas técnicas han experimentado un nuevo avance con la caracterización y aplicación de las recombinasas específicas de sitio (Kilby, N. J., Snaith, M. R. y Murray, J. A. (1993) Trends Genet, 9, 413-21).

25

30

35

40

Las recombinasas específicas de sitio se pueden separar en dos familias principales. La primera (la familia Int o familia de las tirosina recombinasas) comprende las enzimas que catalizan la recombinación entre sitios localizados en la misma molécula de ADN (recombinación intramolecular que conlleva resolución, escisión o inversión) o en moléculas de ADN distintas (recombinación intermolecular que da lugar a la integración) (Sauer, B. (1993) Methods Enzymol, 225, 890-900; Dymecki, S. M. (1996) Proc Natl Acad Sci EE.UU., 93, 6191-6; Abremski, K. y Hoess, R. (1984) J Biol Chem, 259, 1509-14; Nash, H. A. (1996) en Escherichia coli and Salmonella cellular and molecular biology, ed. F. C. Neidhart, R. I. Curtis, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Rezaikoff, M. Riley, M. Schaechter y H. E. Umbarger (A. S. M. Press, Washington D.C.), pág. 2363-7). La última propiedad se ha explotado para permitir la inserción dirigida de secuencias específicas en localizaciones precisas (Sauer, B. y Henderson, N. (1990) The New Biologist, 2, 441-9; Fukushige, S. y Sauer, B. (1992) Proc. Natl. Acad. Sci. EE.UU., 89, 7905-9). Las recombinasas que se han usado para manipular genomas de mamíferos han sido, principalmente las proteínas Cre y Flp, que pertenecen a la familia Int (Kilby, N. J., Snaith, M. R. y Murray, J. A. (1993) Trends Genet, 9, 413-21). Las secuencias diana de estas enzimas, denominados sitios loxP para la enzima Cre y FRT para la enzima Flp, consisten en una repetición invertida corta a la que se une la proteína. El proceso de recombinación es operativo a través de distancias largas (hasta 70 kb) en el genoma. Usando estas enzimas, diversos autores han comunicado recombinación de ADN específica de sitio y de tejido en modelos murinos (DiSanto, J. P., Muller, W., Guy, G. D., Fischer, A. y Rajewsky, K. (1995) Proc Natl Acad Sci EE.UU., 92, 377-81; Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. y Rajewsky, K. (1994) Science, 265, 103-6; Kuhn, R., Schwenk, F., Aguet, M. y Rajewsky, K. (1995) Science, 269, 1427-9; Orban, P. C., Chui, D. y Marth, J. D. (1992) Proc. Natl. Acad. Sci. EE.UU., 89, 6861-5), translocaciones cromosómicas en plantas y en animales (Deursen, J. v., Fornerod, M., Rees, B. v. y Grosveld, G. (1995) Proc. Natl. Acad. Sci. EE.UU., 92, 7376-80; Medberry, S. L., Dale, E., Qin, M. y Ow, D. W. (1995) Nucleic Acids Res, 23, 485-90; Osborne, B. I., Wirtz, U. y Baker, B. (1995) Plant J, 7, 687-701) e inducción dirigida de genes específicos (Pichel, J. G., Lakso, M. y Westphal, H. (1993) Oncogene, 8, 3333-42). El sistema Cre-loxP también se ha usado en combinación con promotores inducibles, tales como el promotor inducible del interferón gamma, que se usó para provocar ablación génica en el hígado con una eficiencia elevada y en menor medida en otros tejidos (Kuhn, R., Schwenk, F., Aguet, M. y Rajewsky, K. (1995) Science, 269, 1427-9). No obstante, este sistema de recombinación específica de sitio solamente permite la inducción de un número reducido de eventos de

45

50

55

60

65

recombinación en el mismo genoma. Dado que cada reacción de recombinación deja una secuencia diana para la recombinasa en el genoma en el sitio de cruzamiento y como las recombinasas (por ejemplo, Cre y Flp) pueden catalizar la recombinación intermolecular, la totalidad del proceso puede conducir a reorganizaciones cromosómicas indeseadas.

5 La segunda familia de recombinasas se denominan en conjunto la familia de resolvasas/invertasas o familia de serina (Grindley, N. D. F. (1994) en *Nucleic Acids and Molecular Biology*, ed. F. Eckstein y D. M. J. Lilley (Springer-Verlag, Berlín), pág. 236-67, (Smith, M. C. y Thorpe, H.M. (2000) *Mol. Microbiol.*, 44, 299-307)). Estas recombinasas específicas de sitio, que incluyen enzimas que catalizan reacciones intramoleculares e intermoleculares, podrían tener una ventaja sobre la familia Int de recombinasas. Las serina recombinasas que catalizan la integración en fagos (integrasas) están especialmente bien adaptadas para uso como herramientas de ingeniería genética. Hasta ahora se han estudiado tres serina recombinasas en células de mamífero ϕ C31, R4 y TP901-1, (Groth, A.C. y Calos, M.P. (2004) *J. Mol. Biol.* 335, 667-678). Se ha observado que estas recombinasas son autónomas, tienen secuencias *att* simples y tienen la capacidad de funcionar en células de mamífero. Como se ha observado ninguna o poca recombinación entre cualquier combinación de sitios distintos a *attP* o *attB*, las integraciones son unidireccionales y existe una frecuencia de integración elevada. Las serina recombinasas proporcionan una ventaja significativa sobre los sistemas de recombinación anteriores que emplean el uso de miembros de la familia Int de recombinasas. Estas enzimas tienen numerosas aplicaciones. Una forma es la introducción de sitios *att* en el genoma de un organismo y el uso como dianas para la recombinación.

20 El solicitante ha identificado nuevas serina recombinasas que demuestran actividad sólida en diversas células de mamífero y en células vegetales, así como la capacidad de integrar de forma estable un polinucleótido en el genoma de una célula huésped o de escindir un polinucleótido del genoma de una célula huésped.

25 **Sumario de la invención**

La presente invención proporciona composiciones y métodos para obtener una recombinación específica de sitio estable en una célula eucariota. Al contrario que en los métodos descritos anteriormente para la recombinación específica de sitio, las presentes recombinasas y métodos para su uso proporcionan una recombinación específica de sitio irreversible y estable.

30 La capacidad de las recombinasas de los fagos para una recombinación directa específica y eficiente entre secuencias de ADN en las células vivas las convierte en potencialmente útiles en diversas aplicaciones de ingeniería genética. Dichas aplicaciones incluyen integración, escisión, inversión, translocación e intercambio de casetes de secuencias de polinucleótidos.

35 **Breve descripción de los dibujos**

40 La figura 1 representa una representación esquemática del ensayo de recombinación intramolecular transitoria (TIRA) usado para analizar la capacidad de la recombinasa para detectar recombinación entre sitios *attP* y *attB* sobre una diana o plásmido de ensayo como se ha descrito en los ejemplos.

45 La figura 2 demuestra los resultados de TIRA para varias recombinasas realizado en células renales embrionarias humanas (HEK293).

La figura 3 demuestra los resultados de TIRA para varias recombinasas realizados en células NIH3T3 de ratón.

50 La figura 4 demuestra los resultados de TIRA para varias recombinasas realizados en células de ovario de hámster chino (CHO).

La figura 5 demuestra los resultados de TIRA para varias recombinasas realizados en células HeLa humanas.

55 La figura 6 demuestra los resultados de TIRA para varias recombinasas realizados en células estromales de médula ósea de rata.

La figura 7 demuestra los resultados de TIRA para varias recombinasas realizados en células madre neurales de ratón.

60 La figura 8 demuestra los resultados de TIRA para la recombinasa A118 realizados en células BY2 de tabaco.

La figura 9 representa una representación esquemática de integración estable de ADN plasmídico que contiene una secuencia *attP* o *attB* en cromosoma de HEK293 que contiene el sitio *attB* o el sitio *attP*.

65 La figura 10 demuestra los resultados de la amplificación por PCR de los sitios *attL* y *attR* tras la integración estable de ADN plasmídico que contiene una secuencia *attP* o *attB* en cromosoma de células HEK293 que contiene el sitio *attB* o el sitio *attP*.

La figura 11 representa una representación esquemática de la escisión de una secuencia STOP integrada de forma estable y la activación de la actividad luciferasa debido a la recombinasa.

5 La figura 12 demuestra los resultados de la escisión de una secuencia STOP integrada de forma estable y la activación de la actividad luciferasa debido a la recombinasa.

La figura 13 representa una representación esquemática de inserción o integración de un plásmido que contiene un sitio de recombinación *attP* o *attB* en el sitio nativo pseudo *attB* o pseudo *attP* presente en células HEK293.

10 La figura 14 demuestra las secuencias de nucleótidos de los sitios pseudo *attB* nativos para las recombinasas SF370.1 y SP β c2 identificadas en células HEK293.

15 Descripción de la forma de realización preferida

Definiciones

En la presente divulgación se usa una serie de términos y abreviaturas. Se proporcionan las definiciones siguientes y deberían ser útiles para la comprensión del ámbito y la práctica de la presente invención.

20 En una realización específica, el término "alrededor de" o "aproximadamente" significa dentro del 20 %, preferentemente dentro del 10 %, más preferentemente dentro del 5 % e incluso más preferentemente dentro del 1 %, de un valor o intervalo dado.

25 "Recombinasa", como se usa en el presente documento, hace referencia a un grupo de enzimas que puede facilitar la recombinación específica de sitio entre sitios definidos, en el que los sitios están separados físicamente en una molécula de ADN sencilla o en el que los sitios residen en moléculas de ADN distintas. Las secuencias de ADN de los sitios de recombinación definidos no son necesariamente idénticas. El inicio de la recombinación depende de la interacción de proteína-ADN, dentro del grupo en el que hay un gran número de proteínas que catalizan la integración y la escisión en fagos (por ejemplo, λ integrasa, ϕ C31), resolución de plásmidos circulares (por ejemplo, Tn3, gamma delta, Cre, Flp), inversión de ADN para la expresión de genes alternos (por ejemplo, Hin, Gin, Pin), ensamblaje de genes durante el desarrollo (por ejemplo, genes de fijación del nitrógeno de Anabaena) y transposición (por ejemplo, el transposón IS607). La mayoría de las recombinasas específicas de sitio entran dentro de una de las dos familias según la relación de la evolución y mecanística. Estas son tirosina recombinasas o de la familia de λ integrasas (por ejemplo, Cre, Flp, Xer D) y de la familia de resolvasas/integrasas o serina recombinasas (por ejemplo, ϕ C31, TP901-1, Tn3, gamma delta).

40 Los "sitios de unión de recombinación" son secuencias de polinucleótidos específicas que son reconocidas por las enzimas recombinasas descritas en el presente documento. Normalmente están implicados dos sitios diferentes (denominados "sitios complementarios"), uno presente en el ácido nucleico diana (por ejemplo, un cromosoma o episoma de un eucariota o procariota) y otro en el ácido nucleico que se va a integrar en el sitio de recombinación diana. En el presente documento se usan los términos "*attB*" y "*attP*," que hacen referencia a los sitios de unión (o recombinación) originalmente de una diana bacteriana y un fago donante, respectivamente, aunque los sitios de recombinación para enzimas concretas pueden tener nombres diferentes. Normalmente, los sitios de recombinación incluyen brazos izquierdo y derecho separados por una región central o espaciadora. Por tanto, un sitio de recombinación *attB* consiste en BOB', en el que B y B' son los brazos izquierdo y derecho, respectivamente y O es la región central. De un modo similar, *attP* es POP', en el que P y P' son los brazos y O es, de nuevo, la región central. Tras la recombinación entre los sitios *attB* y *attP* y la integración concomitante de un ácido nucleico en la diana, los sitios de recombinación que flanquean el ADN integrado se denominan "*attL*" y "*attR*." Por tanto, los sitios *attL* y *attR*, usando la terminología anterior, consisten en BOP' y POB', respectivamente. En algunas representaciones en el presente documento, la "O" se omite y *attB* y *attP*, se designan, por ejemplo, como BB' y PP', respectivamente,

55 La expresión "sustancialmente libre" significa que una composición que comprende "A" (en la que "A" es una proteína sencilla, una molécula de ADN, un vector, una célula huésped recombinante etc.) está sustancialmente libre de "B" (en la que "B" comprende una o más proteínas, moléculas de ADN, vectores contaminantes etc.), cuando al menos aproximadamente el 75 % en peso de las proteínas, ADN, vectores (en función de la categoría de la especie a la que pertenecen A y B) en la composición es "A". Preferentemente, "A" comprende al menos aproximadamente 90 % en peso de la especie A + B en la composición, más preferentemente al menos aproximadamente 99 % en peso. También se prefiere que una composición, que carece sustancialmente de contaminación, contenga solamente una especie de peso molecular único que tenga la actividad o característica de la especie de interés.

65 El término "aislado" para los fines de la presente invención designa un material biológico (ácido nucleico o proteína) que se ha retirado de su ambiente original (el ambiente en el que está presente de forma natural). Por ejemplo, un polinucleótido presente en el estado natural en una planta o un animal no está aislado, aunque el mismo polinucleótido separado de los ácidos nucleicos adyacentes en los que está presente de forma natural, se considera "aislado". El término "purificado" no requiere que el material esté presente en una forma que exhiba pureza absoluta,

exclusivo de la presencia de otros compuestos. Es una definición bastante relativa.

Un polinucleótido está en el estado "purificado" tras la purificación del material de partida o del material natural en al menos un orden de magnitud, preferentemente 2 o 3 y preferentemente, 4 o 5 órdenes de magnitud.

Un "ácido nucleico" es un compuesto polimérico compuesto por subunidades unidas de forma covalente denominadas nucleótidos. El ácido nucleico incluye ácido polirribonucleico (ARN) y ácido polidesoxirribonucleico (ADN); ambos pueden ser monocatenarios o bicatenarios. ADN incluye, entre otros, ADNc, ADN genómico, ADN plasmídico, ADN sintético y ADN semisintético. El ADN puede ser lineal, circular o superenrollado.

Una "molécula de ácido nucleico" se refiere a la forma polimérica éster fosfato de los ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN") o cualquier análogo fosfoéster de los mismos, tales como fosforioatos y tioésteres, en forma monocatenaria, o una hélice bicatenaria. Son posibles las hélices bicatenarias de ADN-ADN, ADN-ARN y ARN-ARN. El término molécula de ácido nucleico y en particular, molécula de ADN o ARN, solamente se refiere a las estructuras primaria y secundaria de la molécula y no se limita a una forma terciaria concreta. Por tanto, este término incluye ADN bicatenario que se encuentra, entre otros, en moléculas de ADN lineales o circulares (por ejemplo, fragmentos de restricción), plásmidos y cromosomas. Al hablar de la estructura de moléculas de ADN bicatenarias concretas, las secuencias se pueden describir en el presente documento de acuerdo con la convención normal de proporcionar únicamente la secuencia en la dirección 5' a 3' a lo largo de la hebra no transcrita de ADN (es decir, la hebra que tiene una secuencia homóloga al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha sufrido una manipulación biológica molecular.

Se entenderá que el término "fragmento" quiere decir una secuencia de nucleótidos de longitud reducida respecto al ácido nucleico de referencia y que comprende, sobre la porción común, una secuencia de nucleótidos idéntica al ácido nucleico de referencia. Dicho fragmento de ácido nucleico de acuerdo con la invención puede incluirse, cuando sea adecuado, en un polinucleótido más grande del que es un constituyente. Dichos fragmentos comprenden o, como alternativa, consisten en oligonucleótidos de longitud variable de al menos 6, 8, 9, 10, 12, 15, 18, 20, 21, 22, 23, 24, 25, 30, 39, 40, 42, 45, 48, 50, 51, 54, 57, 60, 63, 66, 70, 75, 78, 80, 90, 100, 105, 120, 135, 150, 200, 300, 500, 720, 900, 1000 o 1500 nucleótidos consecutivos de un ácido nucleico de acuerdo con la invención.

Como se usa en el presente documento, un "fragmento de ácido nucleico aislado" es un polímero de ARN o ADN que es monocatenario o bicatenario, que opcionalmente contiene bases nucleotídicas sintéticas, no naturales o alteradas. Un fragmento de ácido nucleico aislado en forma de un polímero de ADN puede estar compuesto por uno o más segmentos de ADNc, ADN genómico o ADN sintético.

Un "gen" se refiere a un ensamblaje de nucleótidos que codifican un polipéptido e incluye ADNc y ácidos nucleicos de ADN genómico. "Gen" también se refiere a un fragmento de ácido nucleico que expresa una proteína o polipéptido específico, incluidas secuencias reguladoras precedentes (secuencias no codificantes en 5') y posteriores (secuencias no codificantes en 3') a la secuencia de codificación. "Gen nativo" se refiere a un gen como se encuentra en la naturaleza con sus propias secuencias reguladoras. "Gen quimérico" se refiere a cualquier gen que no es un gen nativo, que comprende secuencias reguladoras y/o de codificación que no se encuentran juntas en la naturaleza. En consecuencia, un gen quimérico puede comprender secuencias reguladoras y secuencias de codificación derivadas de diferentes fuentes o secuencias reguladoras y secuencias de codificación de la misma fuente, pero dispuestas de un modo distinto a como se encuentran en la naturaleza. Un gen quimérico puede comprender secuencias de codificación derivadas de diferentes fuentes y/o secuencias reguladoras derivadas de diferentes fuentes. "Gen endógeno" se refiere a un gen nativo en su localización natural en el genoma de un organismo. Un gen "extraño" o gen "heterólogo" se refiere a un gen que normalmente no se encuentra en el organismo huésped, pero que se introduce en el organismo huésped mediante transferencia génica. Los genes extraños pueden comprender genes nativos insertados en un organismo no nativo o genes quiméricos. Un "transgén" es un gen que se ha introducido en el genoma mediante un procedimiento de transformación.

ADN "heterólogo" se refiere a ADN que no se encuentra de forma natural en la célula o en un sitio cromosómico de la célula. Preferentemente, el ADN heterólogo incluye un gen extraño para la célula.

El término "genoma" incluye ADN o ARN cromosómico, además de mitocondrial, de cloroplastos o viral.

Una molécula de ácido nucleico es "hibrizable" con otra molécula de ácido nucleico, tal como un ADNc, ADN genómico o ARN, cuando una forma monocatenaria de la molécula de ácido nucleico puede hibridar con la otra molécula de ácido nucleico en las condiciones adecuadas de temperatura y fuerza iónica de la solución (véase Sambrook y col., 1989, más adelante). Las condiciones de hibridación y lavado son bien conocidas y se pueden encontrar ejemplos en Sambrook, J., Fritsch, E. F. y Maniatis, T. Molecular Cloning: A Laboratory Manual, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularmente el Capítulo 11 y la tabla 11.1 en dicho documento (incorporado íntegramente en el presente documento por referencia). Las condiciones de temperatura y fuerza iónica determinan la "rigurosidad" de la hibridación.

Las condiciones de rigurosidad se pueden ajustar para detectar de forma selectiva fragmentos moderadamente similares, tales como secuencias homólogas de organismos relacionados de forma distante, a fragmentos altamente similares, tal como genes que duplican las enzimas funcionales de organismos estrechamente relacionados. Para la detección selectiva preliminar de ácidos nucleicos homólogos se pueden usar condiciones de hibridación de baja rigurosidad correspondientes a una T_m de 55 °C, por ejemplo 5x SSC, 0,1 % de SDS, 0,25 % de leche y sin formamida, o 30 % de formamida, 5x SSC, 0,5 % de SDS). Las condiciones de hibridación de rigurosidad moderada corresponden a una T_m mayor, por ejemplo 40 % de formamida, con 5x o 6x SCC. Las condiciones de hibridación de rigurosidad alta corresponden a la T_m más alta, por ejemplo 50 % de formamida, con 5x o 6x SCC.

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque en función de la rigurosidad de la hibridación es posible que se formen apareamientos erróneos entre bases. El término "complementario" se usa para describir la relación entre las bases nucleotídicas que son capaces de hibridar entre sí. Por ejemplo, con respecto al ADN, la adenosina es complementaria de la timina y la citosina es complementaria de la guanina. De acuerdo con esto, la presente invención también incluye fragmentos de ácido nucleico aislado que son complementarios a las secuencias completas como se divulgan o usan en el presente documento, así como las secuencias de ácido nucleico sustancialmente similares.

En una realización específica de la invención, los polinucleótidos se detectan usando condiciones de hibridación que comprenden una etapa de hibridación a T_m de 55 °C y usando las condiciones indicadas anteriormente. En una realización preferida, la T_m es 60 °C, en una realización más preferida, la T_m es 63 °C, en una realización todavía más preferida, la T_m es 65 °C.

Los lavados posthibridación también determinan las condiciones de rigurosidad. Un conjunto de condiciones preferidas usa una serie de lavados que comienzan con 6X SSC, 0,5 % de SDS a temperatura ambiente durante 15 minutos (min), después se repiten con 2X SSC, 0,5 % de SDS a 45 °C durante 30 minutos y después, se repiten dos veces con 0,2X SSC, 0,5 % de SDS a 50 °C durante 30 minutos. Un conjunto más preferido de condiciones de rigurosidad usa temperaturas más altas a las que los lavados son idénticos a los anteriores a excepción de que la temperatura de los dos lavados finales de 30 minutos en 0,2X SSC, 0,5 % de SDS se aumentaba a 60 °C. Otro conjunto preferido de condiciones altamente rigurosas usa dos lavados finales en 0,1X SSC, 0,1 % de SDS a 65 °C. La hibridación requiere que los dos ácidos nucleicos comprendan secuencias complementarias, aunque dependiendo de la rigurosidad de la hibridación, son posibles los apareamientos erróneos entre bases.

La rigurosidad adecuada para hibridar ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementariedad, variables bien conocidas en la técnica. Cuanto mayor es el grado de similitud u homología entre dos secuencias de nucleótidos, mayor es el valor de la T_m para los híbridos de los ácidos nucleicos que tienen dichas secuencias. La estabilidad relativa (correspondiente a una T_m más alta) de las hibridaciones de los ácidos nucleicos disminuye en el orden siguiente: ARN:ARN, ADN:ARN, ADN:ADN. Para híbridos de una longitud mayor de 100 nucleótidos se han derivado ecuaciones para calcular la T_m (véase, Sambrook y col., *supra*, 9.50-10.51). Para la hibridación con ácidos nucleicos más cortos, es decir oligonucleótidos, la posición de los apareamientos erróneos se convierte en más importante y la longitud del oligonucleótido determina su especificidad (véase, Sambrook y col., *supra*, 11.7-11.8).

En una realización específica de la invención, los polinucleótidos se detectan usando condiciones de hibridación que comprenden una etapa de hibridación a menos de 500 mM de sal y al menos a 37 grados centígrados y una etapa de lavado en 2 x SSPE a al menos a 63 °C. En una realización preferida, las condiciones de hibridación comprenden sal a menos de 200 mM y al menos a 37 grados centígrados para la etapa de hibridación. En una realización más preferida, las condiciones de hibridación comprenden 2 x SSPE y 63 grados centígrados para las etapas de hibridación y de lavado.

En una realización, la longitud para un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Preferentemente, una longitud mínima para un ácido nucleico hibridables es, de al menos, aproximadamente 15 nucleótidos; más preferentemente, de al menos aproximadamente 20 nucleótidos y lo más preferentemente la longitud es de al menos 30 nucleótidos. Además, el experto en la técnica reconocerá que la temperatura y la concentración de sales de la solución de lavado pueden ajustarse en caso necesario de acuerdo con factores tales como la longitud de la sonda.

El término "sonda" se refiere a una molécula de ácido nucleico monocatenario cuyas bases se pueden aparear con un ácido nucleico diana monocatenario complementario para formar una molécula bicatenaria.

Como se usa en el presente documento, el término "oligonucleótido" se refiere a un ácido nucleico, en general de al menos 18 nucleótidos, que puede hibridar con una molécula de ADN genómico, una molécula de ADNc, un ADN de plásmido o una molécula de ARNm. Los oligonucleótidos se pueden marcar con, por ejemplo, ³²P-nucleótidos o nucleótidos a los que un marcador, tal como la biotina, se han conjugado de forma covalente. Un oligonucleótido marcado se puede usar como sonda para detectar la presencia de un ácido nucleico. Los oligonucleótidos (de los que uno o ambos pueden marcarse) se pueden usar como cebadores para PCR, bien para clonar la longitud completa o un fragmento de un ácido nucleico, o bien para detectar la presencia de un ácido nucleico. También se

puede usar un oligonucleótido para formar una triple hélice con una molécula de ADN. En general, los oligonucleótidos se preparan de forma sintética, preferentemente en un sintetizador de ácido nucleico. De acuerdo con esto, los oligonucleótidos se pueden preparar con enlaces análogos de fosfoéster, no naturales, tales como enlaces tioéster, etc.

5 Un “cebador” es un oligonucleótido que hibrida con una secuencia de ácido nucleico diana para crear un ácido nucleico bicatenario que puede servir como punto de iniciación para la síntesis de ADN en condiciones adecuadas. Dichos cebadores se pueden usar en una reacción en cadena de la polimerasa.

10 La “reacción en cadena de la polimerasa” se abrevia a PCR y significa un procedimiento *in vitro* para amplificar enzimáticamente secuencias de ácido nucleico específicas. La PCR implica una serie repetitiva de ciclos de temperatura, en la que cada ciclo comprende tres etapas: desnaturalización del ácido nucleico molde para separar las hebras de la molécula diana, hibridación un cebador oligonucleotídico monocatenario para PCR con el ácido nucleico molde y extensión del (los) cebador(es) hibridados mediante la ADN polimerasa. La PCR proporciona un medio para detectar la presencia de la molécula diana y en condiciones cuantitativas o semicuantitativas, para determinar la cantidad relativa de dicha molécula diana dentro de la combinación de partida de ácidos nucleicos.

15 “Reacción en cadena de la polimerasa-transcripción inversa” se abrevia a RT-PCR y significa un procedimiento *in vitro* para producir enzimáticamente una molécula o moléculas del ADNc diana a partir de una molécula o moléculas de ARN, seguido de amplificación enzimática de una secuencia o secuencias de ácido nucleico específico dentro de la molécula o moléculas de ADNc diana como se ha descrito anteriormente. La RT-PCR también proporciona un medio para detectar la presencia de la molécula diana y en condiciones cuantitativas o semicuantitativas, para determinar la cantidad relativa de dicha molécula diana dentro de la combinación de partida de ácidos nucleicos.

20 Una “secuencia codificadora” de ADN es una secuencia de ADN bicatenario que se transcribe y se traduce en un polipéptido en una célula *in vitro* o *in vivo* cuando se introduce sometida al control de las secuencias reguladoras adecuadas. “Secuencias reguladoras adecuadas” se refieren a secuencias nucleotídicas localizadas cadena arriba (secuencias no codificantes en 5’), dentro, o cadena abajo (secuencias no codificantes en 3’) de una secuencia de codificación y que influyen sobre la transcripción, el procesamiento o la estabilidad del ARN o la traducción de la secuencia de codificación asociada. Las secuencias reguladoras pueden incluir promotores, secuencias líder de la traducción, intrones, secuencias de reconocimiento de la poliadenilación, sitio de procesamiento del ARN, sitio de unión al efector y estructura tallo-bucle. Los límites de la secuencia codificadora están determinados por un codón de iniciación en el extremo 5’ (amino) y un codón de terminación de la traducción en el extremo 3’ (carboxilo). Una secuencia codificadora puede incluir, entre otras, secuencias procarióticas, ADNc de ARNm, secuencias de ADN genómico e incluso secuencias de ADN sintéticas. Si la secuencia codificadora está destinada a expresarse en una célula eucariota, en 3’ de la secuencia codificadora habitualmente habrá una señal de poliadenilación y una secuencia de terminación de la transcripción.

25 “Marco de lectura abierto” se abrevia a ORF y quiere decir una longitud de secuencia de ácido nucleico, ADN, ADNc o ARN, que comprende una señal de inicio o codón de iniciación de la traducción, tal como un ATG o AUG y un codón de terminación y potencialmente se puede traducir a una secuencia polipeptídica.

30 La expresión “cabeza-cabeza” se usa en el presente documento para describir la orientación de las dos secuencias polinucleotídicas en relación una con otra. Dos polinucleótidos están en una orientación cabeza—cabeza cuando el extremo 5’ de la hebra de codificación de un polinucleótido está adyacente al extremo 5’ de la hebra de codificación del otro polinucleótido, de modo que la dirección de transcripción de cada polinucleótido procede alejándose del extremo 5’ del otro polinucleótido. La expresión “cabeza-cabeza” se puede abreviar a (5’)-a-(5’) y también se puede indicar con los símbolos ($\leftarrow \rightarrow$) o ($3' \leftarrow 5' 5' \rightarrow 3'$).

35 La expresión “cola-cola” se usa en el presente documento para describir la orientación de las dos secuencias polinucleotídicas en relación una con otra. Dos polinucleótidos están en una orientación cola-cola cuando el extremo 3’ de la hebra de codificación de un polinucleótido está adyacente al extremo 3’ de la hebra de codificación del otro polinucleótido, de modo que la dirección de transcripción de cada polinucleótido procede acercándose al otro polinucleótido. La expresión “cola-cola” se puede abreviar a (3’)-a-(3’) y también se puede indicar con los símbolos ($\rightarrow \leftarrow$) o ($5' \rightarrow 3' 3' \leftarrow 5'$).

40 La expresión “cabeza-cola” se usa en el presente documento para describir la orientación de dos secuencias polinucleotídicas en relación una con otra. Dos polinucleótidos están en una orientación cabeza—cola cuando el extremo 5’ de la hebra de codificación de un polinucleótido está adyacente al extremo 3’ de la hebra de codificación del otro polinucleótido, de modo que la dirección de transcripción de cada polinucleótido procede en el mismo sentido que la del otro polinucleótido. La expresión “cabeza-cola” se puede abreviar a (5’)-a-(3’) y también se puede indicar con los símbolos ($\rightarrow \rightarrow$) o ($5' \rightarrow 3' 5' \rightarrow 3'$).

45 La expresión “cadena abajo” se refiere a una secuencia de nucleótidos que se localiza 3’ de la secuencia de nucleótidos de referencia. En concreto, las secuencias nucleotídicas cadena abajo se refieren, en general, a secuencias que siguen al punto de partida de la transcripción. Por ejemplo, el codón de iniciación de la traducción de

un gen se localiza cadena abajo del punto de partida de la transcripción.

La expresión “cadena arriba” se refiere a una secuencia de nucleótidos que se localiza 5’ de la secuencia de nucleótidos de referencia. En concreto, las secuencias nucleotídicas cadena arriba se refieren, en general, a secuencias que se localizan en el lado 5’ de una secuencia de codificación o punto de partida de la transcripción. Por ejemplo, la mayoría de los promotores se localizan cadena arriba del punto de partida de la transcripción.

Las expresiones “endonucleasas de restricción” y “enzimas de restricción” se refieren a una enzima que se une y corta dentro de una secuencia nucleotídica específica dentro de un ADN bicatenario.

“Recombinación homóloga” se refiere a la inserción de una secuencia de ADN extraño en otra molécula de ADN, por ejemplo la inserción de un vector en un cromosoma. Preferentemente, el vector está dirigido a un sitio cromosómico específico para recombinación homóloga. Para la recombinación homóloga específica, el vector contendrá regiones suficientemente largas de homología con secuencias del cromosoma para permitir la unión complementaria y la incorporación del vector en el cromosoma. Las regiones de homología más largas y los grados mayores de similitud de secuencia pueden incrementar la eficiencia de la recombinación homóloga.

Se pueden usar diversos procedimientos conocidos en la técnica para propagar un polinucleótido de acuerdo con la invención. Una vez que se han establecido el sistema huésped adecuado y las condiciones de crecimiento, los vectores de expresión recombinantes se pueden propagar y preparar en cantidad. Como se ha descrito en el presente documento, los vectores de expresión que se pueden usar incluyen, entre otros, los siguientes vectores o sus derivados: virus de origen humano o animal, tal como virus variolovacunal o adenovirus; virus de insectos, tales como baculovirus, vectores de levaduras, vectores bacteriófagos (por ejemplo, lambda) y vectores de ADN de plásmido y cósmido, por citar algunos.

Un “vector” es cualquier medio para la clonación y/o transferencia de un ácido nucleico en una célula huésped. Un vector puede ser un replicón al que se puede unir otro segmento de ADN de forma que se realiza la replicación del segmento unido. Un “replicón” es cualquier elemento genético (por ejemplo, plásmido, fago, cósmico, cromosoma, virus) que funciona como una unidad autónoma de replicación de ADN *in vivo*, es decir, capaz de replicarse sometido a su propio control. El término “vector” incluye medios virales y no virales para introducir el ácido nucleico en una célula *in vitro*, *ex vivo* o *in vivo*. Se puede usar un gran número de vectores conocidos en la técnica para manipular ácidos nucleicos e incorporar en los genes elementos de respuesta y promotores, etc. Posibles vectores incluyen, por ejemplo, plásmidos o virus modificados incluidos, por ejemplo, bacteriófagos, tales como los derivados del fago lambda, o plásmidos tales como los derivados de los plásmidos pBR322 o pUC, o el vector Bluescript. Por ejemplo, la inserción de los fragmentos de ADN correspondientes a los elementos de respuesta y los promotores en un vector adecuado se puede conseguir ligando los fragmentos de ADN adecuados en un vector elegido que tenga extremos complementarios cohesivos. Como alternativa, los extremos de las moléculas de ADN pueden modificarse enzimáticamente o se puede producir cualquier sitio ligando las secuencias nucleotídicas (engarces) en los extremos de ADN. Dichos vectores se pueden someter a ingeniería genética de modo que contengan genes marcadores seleccionables que proporcionen la selección de células que tienen incorporado el marcador en el genoma celular. Dichos marcadores permiten la identificación y/o la selección de células huésped que incorporan y expresan las proteínas codificadas por el marcador.

Se han usado vectores virales y particularmente vectores retrovirales, en una amplia variedad de aplicaciones de liberación génica en células, así como sujetos animales vivos. Los vectores virales que se pueden usar incluyen, entre otros, vectores de retrovirus, virus adenoasociados, virus de viruela, baculovirus, virus variolovacunal, herpes simple, Epstein-Barr, adenovirus, geminivirus y caulimovirus. Vectores no virales incluyen plásmidos, liposomas, lípidos con carga eléctrica (citofectinas), complejos ADN-proteína y biopolímeros. Además de un ácido nucleico, un vector puede comprender también una o más regiones reguladoras y/o marcadores seleccionables útiles en la selección, medición y monitorización de los resultados de transferencia de ácido nucleico (transferencia a qué tejidos, duración de la expresión etc.).

El término “plásmido” se refiere a un elemento cromosómico adicional que a menudo porta un gen que no es parte del metabolismo central de la célula y normalmente está en forma de moléculas de ADN bicatenario circular. Dichos elementos pueden ser secuencias de replicación autónoma, secuencias de integración de genoma, secuencias de fagos o nucleótidos, lineales, circulares o superenrolladas, de un ADN o ARN monocatenario o bicatenario, derivadas de cualquier fuente, en las que una serie de secuencias nucleotídicas se han unido o recombinado en una única construcción que es capaz de introducir en una célula un fragmento promotor y la secuencia de ADN para un producto génico seleccionado junto con la secuencia no traducida en 3’ adecuada.

Un “vector de clonación” es un “replicón”, que es una longitud de unidad de un ácido nucleico, preferentemente ADN, que se replica secuencialmente y que comprende un origen de replicación, tal como un plásmido, fago o cósmido, al que se puede unir otro segmento de ácido nucleico de forma que se realiza la replicación del segmento unido. Los vectores de clonación pueden replicarse en un tipo celular y expresarse en otro (“vector lanzadera”).

Los vectores se pueden introducir en las células huésped deseadas mediante procedimientos conocidos en la

técnica, por ejemplo transfección, electroporación, microinyección, transducción, fusión celular, DEAE dextrano, precipitación en fosfato cálcico, lipofección (fusión de lisosomas), el uso de una pistola génica o el uso de un vehículo del vector de ADN (véanse, por ejemplo, Wu y col., J. Biol. Chem., 267: 963-967; Wu y Wu, 1988, J. Biol. Chem. 263: 14621-14624; y Hartmut y col., Solicitud de patente canadienses N^o. 2.012.311, presentada el 15 de marzo de 1990).

También se puede introducir *in vivo* mediante lipofección un polinucleótido de acuerdo con la invención. Durante la última década, se ha producido un uso creciente de liposomas para encapsulación y transfección de ácidos nucleicos *in vitro*. Los lípidos catiónicos sintéticos diseñados para limitar las dificultades y peligros encontrados con la transfección mediada por liposomas se pueden usar para preparar liposomas para la transfección *in vivo* de un gen que codifica un marcador (Felgner y col., 1987, Proc. Natl. Acad. Sci. EE.UU. 84: 7413; Mackey y col., 1988, Proc. Natl. Acad. Sci. U.S.A 85: 8027-8031; y Ulmer y col., 1993, Science 259: 1745-1748). El uso de lípidos catiónicos puede estimular la encapsulación de ácidos nucleicos con carga negativa y también promueve la fusión con membranas celulares cargadas negativamente (Felgner y Ringold, 1989, Science 337: 387-388). Compuestos y composiciones lipídicas particularmente útiles para transferencia de ácidos nucleicos se describen en las publicaciones de patente internacional WO 95/18863 y WO 96/17823 y en la patente de EE.UU. N^o 5.459.127. El uso de la lipofección para introducir genes exógenos en los órganos específicos *in vivo* tiene ciertas ventajas prácticas. Dirigir los liposomas a moléculas de células específicas representa un área de beneficio. Está claro que dirigir la transfección a tipos celulares concretos sería particularmente preferido en un tejido con heterogeneidad celular, tal como el páncreas, el hígado, el riñón y el cerebro. Los lípidos pueden acoplarse químicamente a otras moléculas con el objetivo de dirigir (Mackey y col., 1988, *supra*). Los péptidos dirigidos, por ejemplo hormonas o neurotransmisores y proteínas, como anticuerpos o moléculas no peptídicas, podrían acoplarse a los liposomas químicamente.

Otras moléculas también son útiles para facilitar la transfección de un ácido nucleico *in vivo*, tal como un oligopéptido catiónico (por ejemplo, el documento WO 95/21931), péptidos derivados de proteínas de unión a ADN (por ejemplo, el documento WO 96/25508), o un polímero catiónico (por ejemplo, el documento WO 95/21931).

También es posible introducir un vector *in vivo* como plásmido de ADN desnudo (véanse las patentes de EE.UU. 5.693.622, 5.589.466 y 5.580.859). También se pueden usar enfoques de liberación de ADN mediada por receptores (Curiel y col., 1992. Hum. Gene Ther. 3: 147-154; y Wu y Wu, 1987. J. Biol. Chem 262: 4429-4432).

El término "transfección" significa la captación e ARN o ADN exógeno o heterólogo por una célula. Una célula se ha "transfectado" con ADN o ARN exógeno o heterólogo cuando dicho ADN o ARN se ha introducido en el interior de la célula. Una célula se ha "transformado" con ADN o ARN exógeno o heterólogo cuando dicho ADN o ARN efectúa un cambio fenotípico. El ARN o ADN transformante puede integrarse (unirse de forma covalente) en el ADN cromosómico que compone el genoma de la célula.

"Transformación" se refiere a la transferencia de un fragmento de ácido nucleico en el genoma de un organismo huésped, que tiene como resultado una herencia genéticamente estable. Los organismos huésped que contienen los fragmentos de ácido nucleico transformados se denominan organismos "transgénicos" o "recombinantes" o "transformados".

La expresión "región genética" hace referencia a una región de una molécula de ácido nucleico o una secuencia nucleotídica que comprende un gen que codifica un polipéptido.

Además, el vector recombinante que comprende un polinucleótido de acuerdo con la invención puede incluir uno o más orígenes de replicación en los huéspedes celulares en los que se busca su amplificación o su expresión, marcadores o marcadores seleccionables.

La expresión "marcador seleccionable" quiere decir un factor de identificación, normalmente un gen de resistencia a un antibiótico o un gen de resistencia química, que es capaz de seleccionarse en base al efecto del gen marcador, es decir resistencia a un antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes y similares, en los que el efecto se usa para rastrear la herencia de un ácido nucleico de interés y/o identificar una célula u organismo que ha heredado el ácido nucleico de interés. Ejemplos de genes marcadores seleccionables conocidos y usados en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomycin, gentamicina, kanamicina, higromicina, herbicida bialafós, sulfonamida y similares y genes que se usan como marcadores fenotípicos, es decir genes reguladores de la antocianina, gen de la isopentaniiltransferasa y similares.

La expresión "gen indicador" quiere decir un ácido nucleico que codifica un factor de identificación que es capaz de identificarse en base al efecto del gen indicador, en el que el efecto se usa para rastrear la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés y/o para medir la inducción de la expresión génica o la transcripción. Ejemplos de genes indicadores conocidos y usados en la técnica incluyen: luciferasa (Luc), proteína fluorescente verde (GFP), cloranfenicol acetiltransferasa (CAT), β -galactosidasa (LacZ), β -glucuronidasa (Gus) y similares. Los genes marcadores seleccionables también se pueden

considerar genes indicadores.

“Promotor” se refiere a una secuencia de ADN capaz de controlar la expresión de una secuencia de codificación o ARN funcional. En general, una secuencia de codificación se localiza en 3' de una secuencia promotora. Los promotores pueden derivar en su totalidad de un gen nativo o estar compuestos por diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintético. Los expertos en la técnica entienden que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos celulares, o en diferentes etapas de desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas. Los promotores que hacen que un gen se exprese en la mayoría de los tipos celulares se denominan, en la mayoría de los casos, “promotores constitutivos”. Los promotores que hacen que un gen se exprese en un tipo celular específico normalmente se denominan “promotores específicos de células” o “promotores específicos de tejido”. Los promotores que hacen que un gen se exprese en una etapa específica de desarrollo o diferenciación celular normalmente se denominan “promotores específicos del desarrollo” o “promotores específicos de diferenciación celular”. Los promotores que se inducen y hacen que un gen se exprese tras la exposición o tratamiento de la célula con un agente, molécula biológica, sustancia química, ligando, luz o similar, que induce el promotor normalmente se denominan “promotores inducibles” o “promotores regulables”. Además se reconoce que, ya que en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido completamente, los fragmentos de ADN de diferentes longitudes pueden tener una actividad promotora idéntica.

Una “secuencia promotora” es una región reguladora de ADN capaz de unir la ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificadora cadena abajo (dirección 3'). Para los fines de definir la presente invención, la secuencia promotora se une a su extremo 3' mediante el sitio de inicio de la transcripción y se extiende cadena arriba (en dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del normal. Dentro de la secuencia promotora se encontrará un sitio de inicio de la transcripción (convenientemente definido mediante, por ejemplo, mapeo con la nucleasa S1), así como los dominios de unión a proteína (secuencias consenso) responsables de la unión de la ARN polimerasa.

Una secuencia codificadora está “sometida al control” de las secuencias de control de la transcripción y de la traducción en una célula cuando la ARN polimerasa transcribe la secuencia codificadora en ARNm, que después se somete a corte y empalme de ARN en trans (si la secuencia de codificación contiene intrones) y se traduce en la proteína codificada por la secuencia codificadora.

Las “secuencias de control de la transcripción y de la traducción” son secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una secuencia codificadora en una célula huésped. En las células eucariotas, las señales de poliadenilación son secuencias control.

La expresión “elemento de respuesta” significa uno o más elementos de ADN de acción en cis que confieren capacidad de respuesta en un promotor mediada por la interacción con los dominios de unión a ADN del primer gen quimérico. Este elemento de ADN puede ser de secuencia palindrómica (perfecta o imperfecta) o compuesto por motivos de secuencia o semisitios separados por un número variable de nucleótidos. Los semisitios pueden ser similares o idénticos y estar dispuestos como repeticiones directas o invertidas o como un semisitio único o multímeros de semisitios adyacentes en tándem. El elemento de respuesta puede comprender un promotor mínimo aislado de diferentes organismos en función de la naturaleza de la célula u organismo en el que se incorporará el elemento de respuesta. El dominio de unión a ADN de la primera proteína híbrida se une, en presencia o ausencia de un ligando, a la secuencia de ADN de un elemento de respuesta para iniciar o suprimir la transcripción de gen(es) cadena abajo sometido(s) a la regulación de este elemento de respuesta. Ejemplos de secuencias de ADN para los elementos de respuesta del receptor de ecdisona natural incluyen: RRGG/TTCANTGAC/ACY (véase Cherbas L. y col., (1991), Genes Dev. 5, 120-131); AGGTCAN_(n)AGGTCA, en la que N_(n) puede ser uno o más nucleótidos espaciadores (véase D'Avino PP. y col., (1995), Mol. Cell. Endocrinol, 113, 1-9); y GGGTTGAATGAATTT (véase Antoniewski C. y col., (1994). Mol. Cell Biol. 14, 4465-4474).

La expresión “operativamente unido” se refiere a la asociación de secuencias de ácido nucleico en un único fragmento de ácido nucleico de modo que la función de una está afectada por la otra. Por ejemplo, un promotor está operativamente unido a una secuencia de codificación cuando es capaz de afectar a la expresión de dicha secuencia de codificación (es decir, que la secuencia de codificación está sometida al control transcripcional del promotor). Las secuencias de codificación pueden estar operativamente unidas a secuencias reguladoras en orientación sentido o antisentido.

El término “expresión”, como se usa en el presente documento, se refiere a la transcripción y acumulación estables de ARN en sentido habitual (ARNm) o antisentido derivado de un ácido nucleico o polinucleótido. La expresión también se puede denominar traducción de ARNm en una proteína o polipéptido.

Las expresiones “casete”, “casete de expresión” y “casete de expresión génica” hacen referencia a un segmento de ADN que se puede insertar en un ácido nucleico o polinucleótido en sitios de restricción específicos o mediante recombinación homóloga. El segmento de ADN comprende un polinucleótido que codifica un polipéptido de interés y el casete y los sitios de restricción se han diseñado para garantizar la inserción del casete en un marco de lectura

adecuado para la transcripción y la traducción. “Casete de transformación” se refiere a un vector específico que comprende un polinucleótido que codifica un polipéptido de interés y que tiene elementos además del polinucleótido que facilitan la transformación de una célula huésped concreta. Los casetes, casetes de expresión, casetes de expresión génica y casetes de transformación de la invención pueden también comprender elementos que permiten la expresión potenciada de un polinucleótido que codifica un polipéptido de interés en una célula huésped. Estos elementos pueden incluir, entre otros: un promotor, un promotor mínimo, un potenciador, un elemento de respuesta, una secuencia de terminación, una secuencia de poliadenilación y similares.

Los términos “modulan” y “modula” quieren decir inducir, reducir o inhibir la expresión de ácido nucleico o génica, lo que tiene como resultado la correspondiente inducción, reducción o inhibición de la producción de proteínas o polipéptidos.

Los plásmidos o vectores de acuerdo con la invención pueden comprender además al menos un promotor adecuado para dirigir la expresión de un gen en una célula huésped. La expresión “vector de expresión” significa un vector, plásmido o vehículo diseñado para permitir la expresión de una secuencia de ácido nucleico insertada tras la transformación en el huésped. El gen clonado, es decir, la secuencia de ácido nucleico insertada, normalmente se coloca sometido al control de los elementos de control, tales como un promotor, un promotor mínimo, un potenciador o similares. Las regiones control de la iniciación o promotores que son útiles para dirigir la expresión de un ácido nucleico en una célula huésped deseada son numerosas y familiares para los expertos en la técnica. Casi cualquier promotor capaz de dirigir estos genes es adecuado para la presente invención, incluidos, entre otros: promotores virales, promotores bacterianos, promotores animales, promotores de mamíferos, promotores sintéticos, promotores constitutivos, promotores específicos de tejido, promotores específicos del desarrollo, promotores inducibles, promotores regulados por luz; *CYCI*, *HIS3*, *GAL1*, *GAL4*, *GAL10*, *ADH1*, *PGK*, *PHO5*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO*, *TPI*, promotores de la fosfatasa alcalina (útiles para expresión en *Saccharomyces*); promotor *AOX1* (útil para expresión en *Pichia*); promotores de β -lactamasa, *lac*, *ara*, *tet*, *trp*, *IPL*, *IPR*, *T7*, *tac* y *trc* (útiles para expresión en *Escherichia coli*); promotores regulados por luz, específicos de semillas, específicos de polen, específicos de ovarios, relacionados con la patogenicidad o con una enfermedad, promotores de 35 S del virus del mosaico de la coliflor, mínimo 35S de CMV, del virus del mosaico de las nervaduras (CsVMV), de la proteína de la unión a la clorofila a/b, de la ribulosa 1, 5-bisfosfato carboxilasa, específicos de brotes, específicos de raíz, de la quitinasa, inducibles por estrés, del virus baciliforme del tungro del arroz, súper promotor de plantas, de la leucina aminopeptidasa de la patata, de la nitrato reductasa, de la manopina sintasa, de la nopalina sintasa, de la ubiquitina, de la proteína zeína y de la antocianina (útiles para expresión en células vegetales); los promotores animales y de mamífero conocidos en la técnica incluyen, entre otros, la región promotora temprana del SV40 (SV40e), el promotor contenido en la repetición terminal larga en 3' (LTR) del virus del sarcoma de Rous (RSV), los promotores de E1A o de los genes tardíos mayores (MLP) de adenovirus (Ad), el promotor temprano del citomegalovirus (CMV), el promotor de la timidina cinasa (PGK) del virus del herpes simple (HSV) (TK), un promotor IE1 del baculovirus, un promotor del factor de elongación 1 alfa (EF1), un promotor de la fosfoglicerato cinasa (PGK), un promotor de la ubiquitina (Ubc), un promotor de la albúmina, las secuencias reguladoras del promotor de la L-metalotioneína de ratón y regiones de control de la transcripción, los promotores ubicuos (HPRT, vimentina, α -actina, tubulina y similares), los promotores de los filamentos intermedios (desmina, neurofilamentos, queratina, GFAP y similares), los promotores de los genes terapéuticos (de MDR, CFTR o el factor VIII y similares), promotores relacionados con patogenicidad o enfermedad y promotores que exhiben especificidad de tejido y que se han usado en animales transgénicos, tales como la región control del gen I de la elastasa que es activo en células acinares pancreáticas; la región control del gen de la insulina activo en las células pancreáticas beta, la región control del gen de la inmunoglobulina activa en células linfoides, la región control del virus del tumor de mama de ratón activa en células testiculares, de mama, linfoides y mastocitos; regiones control del gen de la albúmina, Apo AI y Apo AII activas en el hígado, región control del gen de la alfa-fetoproteína activa en el hígado, región control del gen de la alfa 1-antitripsina activa en el hígado, región control del gen de la beta-globina activa en células mieloides, región control del gen de la proteína básica de la mielina activa en células oligodendrocíticas en el cerebro, región control del gen de la cadena-2 ligera de la miosina activa en el músculo esquelético y región control del gen de la hormona liberadora de la gonadotropina activa en el hipotálamo, el promotor de la piruvato cinasa, el promotor de la vilina, el promotor de la proteína intestinal de unión a ácidos grasos, el promotor de la α -actina de las células de músculo liso y similares. Además, estas secuencias de expresión se pueden modificar mediante la adición de secuencias potenciadoras o reguladoras y similares.

Los potenciadores que se pueden usar en las realizaciones de la invención incluyen, entre otros: un potenciador de SV40, un potenciador de citomegalovirus (CMV), un potenciador del factor 1 de elongación (EF1), potenciadores de levaduras, potenciadores de genes virales y similares.

Las regiones control de la terminación, es decir las secuencias terminadoras o de poliadenilación, también pueden derivarse de varios genes nativos de los huéspedes preferidos. Opcionalmente, un sitio de terminación puede ser innecesario, aunque se prefiere su inclusión. En una realización preferida de la invención, la región control de la terminación puede estar compuesta por o derivada de una secuencia sintética, una señal de poliadenilación sintética, una señal de poliadenilación tardía de SV40, una señal de poliadenilación de SV40, una señal de poliadenilación de la hormona de crecimiento bovina (BGH), secuencias de terminación virales, o similares.

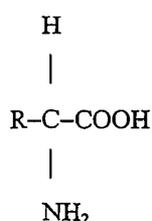
Las expresiones “secuencias no codificantes en 3'” o “región no traducida en 3' (UTR)” hacen referencia a secuencias de ADN localizadas cadena abajo (3') de una secuencia de codificación y pueden comprender secuencias de reconocimiento de poliadenilación [poli(A)] y otras secuencias que codifican señales reguladoras capaces de afectar al procesamiento del ARNm o a la expresión génica. La señal de poliadenilación normalmente se caracteriza porque afecta a la adición de tramos de ácido poliadenílico al extremo 3' del precursor de ARNm.

“Región reguladora” quiere decir una secuencia de ácido nucleico que regula la expresión de una segunda secuencia de ácido nucleico. Una región reguladora puede incluir secuencias responsables de forma natural de expresar un ácido nucleico concreto (una región homóloga) o puede incluir secuencias de un origen diferente que son responsables de la expresión de diferentes proteínas o incluso de proteínas sintéticas (o región heteróloga). En concreto, las secuencias pueden ser secuencias de genes procariotas, eucariotas o virales, o secuencias derivadas que estimulan o reprimen la transcripción de un gen de un modo específico o inespecífico y de un modo inducible o no inducible. Las regiones reguladoras incluyen orígenes de replicación, sitios de corte y empalme de ARN, promotores, potenciadores, secuencias de terminación de la transcripción y secuencias señal que dirigen al polipéptido a las vías secretoras de la célula diana.

Una región reguladora de una “fuente heteróloga” es una región reguladora que no está asociada de forma natural con el ácido nucleico expresado. Incluidas entre las regiones reguladoras heterólogas están las regiones reguladoras de una especie diferente, las regiones reguladoras de un gen diferente, regiones reguladoras híbridas y regiones reguladoras que no se producen en la naturaleza pero que están diseñadas por un experto en la técnica.

“Transcrito de ARN” hace referencia al producto resultante de la transcripción catalizada por la ARN polimerasa de una secuencia de ADN. Cuando el transcrito de ARN es una copia complementaria perfecta de la secuencia de ADN, se denomina transcrito primario o puede ser una secuencia de ARN derivada del procesamiento postranscripcional del transcrito primario y se denomina ARN maduro. “ARN mensajero (ARNm)” se refiere al ARN que no tiene intrones y que se puede traducir en proteínas en la célula. “ADNc” se refiere a un ADN bicatenario que es complementario y derivado del ARNm. ARN “sentido” se refiere al transcrito de ARN que incluye ARNm y de este modo, se puede traducir en proteínas en la célula. “ARN antisentido” se refiere a un transcrito de ARN que es complementario a todo o parte de un transcrito primario diana o ARNm y que bloquea la expresión de un gen diana. La complementariedad de un ARN antisentido puede ser con cualquier parte del transcrito génico específico, es decir en la secuencia no codificante en 5', la secuencia no codificante en 3' o la secuencia de codificación. “ARN funcional” se refiere a ARN antisentido, ARN de ribozima u otro ARN que todavía no está traducido, aunque tiene un efecto sobre los procesos celulares.

Un “polipéptido” es un compuesto polimérico compuesto por residuos de aminoácidos unidos de forma covalente. Los aminoácidos tienen la siguiente estructura general:



Los aminoácidos se clasifican en siete grupos según la cadena lateral R: (1) cadenas laterales alifáticas, (2) cadenas laterales que contienen un grupo hidroxilo (OH), (3) cadenas laterales que contienen átomos de azufre, (4) cadenas laterales que contienen un grupo ácido o amida, (5) cadenas laterales que contienen un grupo básico, (6) cadenas laterales que contienen un anillo aromático y (7) prolina, un aminoácido en el que la cadena lateral está condensada con el grupo amino. Un polipéptido de la invención comprende, preferentemente, al menos 14 aminoácidos.

Una “proteína” es un polipéptido que desempeña un papel estructural o funcional en una célula viva.

Un “polipéptido aislado” o “proteína aislada” es un polipéptido o proteína que carece sustancialmente de aquellos compuestos que normalmente están asociados con ellos en su estado natural (por ejemplo, otras proteínas o polipéptidos, ácidos nucleicos, hidratos de carbono, lípidos). Con “aislado” no se quiere decir que se excluyen las mezclas artificiales o sintéticas con otros compuestos o la presencia de impurezas que no interfieren en la actividad biológica y que pueden estar presentes, por ejemplo, debido a una purificación incompleta, a la adición de estabilizantes o a la formación de composición en un preparado farmacéuticamente aceptable.

Una “variante” de un polipéptido o proteína es cualquier análogo, fragmento, derivado o mutante que deriva de un polipéptido o proteína y que conserva al menos una propiedad biológica del polipéptido o proteína. En la naturaleza pueden existir diferentes variantes del polipéptido o proteína. Estas variantes pueden ser variaciones alélicas caracterizadas por diferencias en las secuencias de nucleótidos del gen estructural que codifica la proteína o pueden implicar corte y empalme diferencial o modificación postraduccional. El experto en la técnica puede producir

variantes que tienen una o varias sustituciones, deleciones, adiciones o reemplazos de aminoácidos. Estas variantes pueden incluir, entre otras: (a) variantes en las que uno o más residuos de aminoácidos están sustituidos por aminoácidos conservadores o no conservadores, (b) variantes en las que se añaden uno o más aminoácidos al polipéptido o proteína, (c) variantes en las que uno o más de los aminoácidos incluyen un grupo sustituyente y (d) variantes en las que el polipéptido o proteína se condensa con otro polipéptido tal como seroalbúmina. Los expertos en la materia conocen las técnicas para obtener estas variantes, incluidas técnicas genéticas (supresiones, deleciones, mutaciones etc.), químicas y enzimáticas.

Una "proteína heteróloga" hace referencia a una proteína que no se produce de forma natural en la célula.

Una "proteína madura" hace referencia a un polipéptido procesado postraduccionalmente, es decir uno del que se ha eliminado cualquier pre o propéptido presente en el producto primario de la traducción. Proteína "precursora" hace referencia al producto primario de la traducción del ARNm, es decir cuando todavía hay pre y propéptidos. Los prepéptidos y los propéptidos pueden ser, entre otros, señales de localización intracelular.

La expresión "péptido señal" se refiere a un polipéptido amino terminal precedente a una proteína madura secretada. El péptido señal se escinde y por tanto, no está presente en la proteína madura. Los péptidos señal tienen la función de dirigir y translocar las proteínas secretadas a través de las membranas celulares. El péptido señal también se denomina proteína señal.

Al principio de la secuencia de codificación de una proteína que se va a expresar sobre la superficie de una célula se incluye una "secuencia señal". Esta secuencia codifica un péptido señal en el extremo N del polipéptido maduro que dirige la célula huésped para translocar el polipéptido. La expresión "secuencia señal de translocación" se usa en el presente documento para hacer referencia a este tipo de secuencia señal. Las secuencias señal de translocación se pueden encontrar asociadas con una diversidad de proteínas nativas de procariontes y eucariotes y a menudo funcionales en ambos tipos de organismos.

El término "homología" se refiere al porcentaje de identidad entre dos restos polinucleotídicos o dos polipeptídicos. La correspondencia entre la secuencia de un resto y otra se puede determinar mediante técnicas conocidas en la material. Por ejemplo, la homología se puede determinar mediante una comparación directa de la información de la secuencia entre dos moléculas polipeptídicas alineando la información de la secuencia y usando programas informáticos disponibles con facilidad. Como alternativa, la homología se puede determinar mediante hibridación de polinucleotídicos en condiciones que forman dúplex estables entre regiones homólogas, seguida de digestión con nucleasa(s) específicas de una cadena y determinación del tamaño de los fragmentos digeridos.

Como se usa en el presente documento, el término "homólogo" y todas sus formas gramaticales y variaciones ortográficas hacen referencia a la relación entre proteínas que poseen un "origen evolutivo común", incluidas proteínas de superfamilias (por ejemplo, la superfamilia de las inmunoglobulinas) y proteínas homólogas de diferentes especies (por ejemplo, cadena ligera de la miosina, etc.) (Reeck y col., 1987, Cell 50: 667.). Dichas proteínas (y sus genes de codificación) tienen homología de secuencia, como refleja su grado elevado de similitud de secuencia. No obstante, en el uso habitual y en la presente solicitud, el término "homólogo", cuando se modifica con un adverbio tal como "considerablemente" puede hacer referencia a la similitud de secuencia y no a un origen evolutivo común.

De acuerdo con lo anterior, la expresión "similitud de secuencia" en todas sus formas gramaticales hace referencia al grado de identidad o a la correspondencia entre secuencias de ácido nucleico o de aminoácidos de proteínas que pueden o no compartir un origen evolutivo común (véase, Reeck y col., 1987, Cell 50: 667).

En una realización específica, dos secuencias de ADN son "sustancialmente homólogas" o "sustancialmente similares" cuando al menos aproximadamente el 50 % (preferentemente al menos aproximadamente el 75 % y más preferentemente al menos aproximadamente 90 o 95 %) de los nucleótidos coinciden con una longitud definida de las secuencias de ADN. Las secuencias que son sustancialmente homólogas se pueden identificar comparando las secuencias usando un software estándar disponible en los bancos de datos de secuencias o en un experimento de hibridación de tipo Southern en, por ejemplo, condiciones rigurosas tal y como se definen para ese sistema concreto. La definición de las condiciones de hibridación adecuadas está dentro de la experiencia en la técnica. Véase, por ejemplo, Sambrook y col., 1989, *supra*.

Como se usa en el presente documento, "sustancialmente similar" hace referencia a fragmentos de ácido nucleico en los que los cambios en una o más bases nucleotídicas tienen como resultado la sustitución de uno o más aminoácidos, pero no afectan a las propiedades funcionales de la proteína codificada por la secuencia de ADN. "Sustancialmente similar" también hace referencia a fragmentos de ácido nucleico en los que los cambios de una o más bases de nucleótidos no afectan a la capacidad del fragmento de ácido nucleico a participar en la alteración de la expresión génica mediante tecnología antisentido o de co-supresión. "Sustancialmente similar" también hace referencia a modificaciones de los fragmentos de ácido nucleico de la presente invención, tales como deleción o inserción de una o más bases nucleotídicas, que no afectan sustancialmente a las propiedades funcionales del transcrito resultante. Por tanto, se entiende que la invención abarca más que las secuencias de ejemplo específicas.

Cada una de las modificaciones propuestas está muy bien dentro de la experiencia en la técnica, como también lo está la determinación de la retención de la actividad biológica de los productos codificados.

Además, el experto en la técnica reconoce que las secuencias sustancialmente similares abarcadas por la presente invención también se definen por su capacidad para hibridar, en condiciones rigurosas (0,1 X SSC, 0,1 % de SDS, 65 °C y lavadas con 2 X SSC, 0,1 % de SDS seguido de 0,1 X SSC, 0,1 % de SDS), con las secuencias de ejemplo en el presente documento. Fragmentos de ácido nucleico sustancialmente similares de la presente invención son los fragmentos de ácido nucleico cuyas secuencias de ADN tienen una identidad de al menos un 70 % con la secuencia de ADN de los fragmentos de ácido nucleico comunicados en el presente documento. Fragmentos de ácido nucleico sustancialmente preferidos de la presente invención son los fragmentos de ácido nucleico cuyas secuencias de ADN tienen una identidad de al menos un 80 % con la secuencia de ADN de los fragmentos de ácido nucleico comunicados en el presente documento. Fragmentos de ácido nucleico más preferidos tienen una homología de al menos un 90 % con la secuencia de ADN de los fragmentos de ácido nucleico comunicados en el presente documento. Incluso más preferidos son los fragmentos de ácido nucleico que tienen una homología de al menos un 95 % con la secuencia de ADN de los fragmentos de ácido nucleico comunicados en el presente documento.

Dos secuencias de aminoácidos son “sustancialmente homólogas” o “sustancialmente similares” cuando más de aproximadamente el 40 % de los aminoácidos son idénticos o más del 60 % son similares (funcionalmente idénticos). Preferentemente, las secuencias similares u homólogas se identifican mediante alineación usando, por ejemplo, el programa en cadena GCG (Genetics Computer Group, Program Manual for the GCG Package, Versión 7, Madison, Wisconsin).

El término “correspondiente” se usa en el presente documento para hacer referencia a secuencias similares u homólogas, sea la posición exacta idéntica o diferente de la molécula en la que se mide la similitud u homología. Una alineación de la secuencia de ácido nucleico o de aminoácidos puede incluir espacios. Por tanto, la expresión “correspondiente a” hace referencia a la similitud de la secuencia y no a la numeración de los residuos de aminoácidos o bases nucleotídicas.

Una “porción sustancial” de una secuencia de aminoácidos o de nucleótidos comprende suficiente de la secuencia de aminoácidos de un polipéptido o de la secuencia de nucleótidos de un gen para identificar de forma supuesta dicho polipéptido o gen, mediante evaluación manual de la secuencia por parte de un experto en la técnica o mediante comparación e identificación automáticas de secuencias por ordenador usando algoritmos tales como BLAST (Basic Local Alignment Search Tool; Altschul, S. F. y col., (1993) J. Mol. Biol. 215: 403-410; véase también www.ncbi.nlm.nih.gov/BLAST/). En general se necesita una secuencia de diez o más aminoácidos contiguos o de treinta o más nucleótidos con el fin de identificar aparentemente un polipéptido o una secuencia de ácido nucleico como homóloga a una proteína o gen conocidos. Además, con respecto a las secuencias de nucleótidos, se pueden usar sondas oligonucleotídicas específicas del gen que comprende 20-30 nucleótidos contiguos en procedimientos dependientes de la secuencia de identificación (por ejemplo, hibridación de tipo Southern) y aislamiento (por ejemplo, hibridación *in situ* de colonias bacterianas o de placas de bacteriófagos) de genes. Además, se pueden usar oligonucleótidos cortos de 12-15 bases como cebadores en PCR con el fin de obtener un fragmento de ácido nucleico concreto que comprenda los cebadores. De acuerdo con lo anterior, una “porción sustancial” de una secuencia de nucleótidos comprende suficiente secuencia para identificar específicamente y/o aislar un fragmento de ácido nucleico que comprenda la secuencia.

La expresión “porcentaje de identidad”, como se conoce en la técnica, es la relación entre dos o más secuencias polipeptídicas o dos o más secuencias polinucleotídicas, determinada comparando las secuencias. En la técnica, “identidad” también significa el grado de relación de la secuencia entre secuencias polipeptídicas o polinucleotídicas, como puede ser el caso según se determine mediante la correspondencia entre cadenas de dichas secuencias. “Identidad” y “similitud” se pueden calcular fácilmente mediante procedimientos conocidos, incluidos, entre otros, los descritos en: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, Nueva York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, Nueva York (1993); Computer Analysis of Sequence Data, parte I (Griffin, A. M. y Griffin, H. G., eds.) Humana Press, Nueva Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); y Sequence Analysis Primer (Gribskov, M. y Devereux, J., eds.) Stockton Press, Nueva York (1991). Se han diseñado procedimientos preferidos para determinar la identidad para dar la mayor coincidencia entre las secuencias analizadas. Los procedimientos para determinar la identidad y la similitud se codifican en programas informáticos disponibles para el público. Los cálculos de las alineaciones de secuencias y el porcentaje de identidad se pueden realizar usando el programa Megalign de LASERGENE bioinformatic computing suite (DNASTAR Inc., Madison, WI). Se pueden realizar múltiples alineaciones de las secuencias usando el procedimiento Clustal de alineación (Higgins y Sharp (1989) CABIOS. 5: 151-153) con los parámetros por defecto (PENALIZACIÓN DE ESPACIO= 10, PENALIZACIÓN DE LONGITUD DE ESPACIO=10). Se pueden seleccionar los parámetros por defecto para alineaciones por pares usando el procedimiento Clustal. KTUPLE 1, PENALIZACIÓN DE ESPACIO=3, VENTANA=5 y DIAGONALES GUARDADAS=5.

La expresión “software de análisis de secuencia” hace referencia a cualquier algoritmo informático o programa informático que es útil para el análisis de secuencias de nucleótidos o aminoácidos. El “software de análisis de

secuencia" puede estar disponible comercialmente o desarrollarse de forma independiente. El software de análisis de secuencia típico incluirá, entre otros, la suite GCG de programas (Paquete Wisconsin Versión 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul y col., J. Mol. Biol. 215: 403-410 (1990) y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE.UU.). Dentro del contexto de la

- 5 presente solicitud se entenderá que el software de análisis de la secuencia se usa para el análisis, que los resultados del análisis se basarán, a menos que se indique lo contrario, en los "valores por defecto" del programa al que se ha hecho referencia. Como se usa en el presente documento, "valores por defecto" significa cualquier conjunto de valores o parámetros que se cargaron originalmente en el software cuando se inicializó por primera vez.
- 10 Los "genes sintéticos" se pueden ensamblar a partir de bloques de construcción oligonucleotídicos que se sintetizan químicamente usando procedimientos conocidos en la técnica. Estos bloques de construcción se ligan e hibridan para formar segmentos génicos que, después, se ensamblan enzimáticamente para construir la totalidad del gen. "Químicamente sintetizado", en relación con una secuencia de ADN, significa que los nucleótidos componentes se ensamblaron *in vitro*. La síntesis química manual de ADN se puede realizar usando procedimientos bien
- 15 establecidos o la síntesis química automática se puede realizar usando una serie de máquinas disponibles comercialmente. De acuerdo con lo anterior, los genes se pueden ajustar para una expresión génica óptima basada en la optimización de la secuencia de nucleótidos de modo que reflejen el sesgo de los codones de la célula huésped. El experto en la técnica aprecia la probabilidad de éxito de la expresión génica si el uso de codones está sesgado hacia los codones favorecidos por el huésped. La determinación de los codones preferidos se puede basar
- 20 en una investigación de los genes derivados de la célula huésped de los que se dispone de información sobre la secuencia.

La invención

- 25 La presente invención proporciona composiciones y procedimientos para obtener una recombinación específica de sitio en células eucariotas. Más específicamente, la invención emplea recombinasas procariotas, tales como recombinasas de bacteriófagos, que son unidireccionales en cuanto a que pueden catalizar la recombinación entre dos sitios de recombinación complementarios, pero no pueden catalizar la recombinación entre los sitios híbridos que se forman mediante esta recombinación. El autor de la invención ha identificado nuevas recombinasas, cada una de las cuales dirige la recombinación únicamente entre un sitio de unión bacteriana (*attB*) y un sitio de unión del
- 30 fago (*attP*). La recombinasa no puede participar en la recombinación entre los sitios híbridos *attL* y *attR* que se forman tras la recombinación entre *attB* y *attP*. Dado que las recombinasas como estas no pueden catalizar solas la reacción inversa, la recombinación de *attB* y *attP* es estable. Esta propiedad es la que separa las composiciones y procedimientos de la presente invención de otros sistemas de recombinación usados actualmente para células eucariotas, tales como el sistema Cre-lox o FLP-FRT, en los que las reacciones de recombinación son reversibles.
- 35 El uso de los sistemas de recombinación de la presente invención proporciona nuevas oportunidades para dirigir el transgén estable y las reorganizaciones cromosómicas en células eucariotas.

- Los procedimientos de la presente invención implican poner en contacto un par de sitios de unión de recombinación, *attB* y *attP*, presentes en una célula eucariota, con una recombinasa correspondiente. A continuación, la recombinasa participa en la recombinación entre los sitios de unión para recombinación. Dependiendo de las ubicaciones relativas de los sitios de unión para recombinación se puede producir uno cualquiera de una serie de eventos como resultado de la recombinación. Por ejemplo, si los sitios de unión para recombinación están presentes en diferentes moléculas de ácido nucleico, la recombinación puede dar lugar a la integración de una molécula de
- 45 ácido nucleico en una segunda molécula. Por tanto, se puede obtener la integración de un plásmido que contiene un sitio de recombinación en un cromosoma de una célula eucariota que incluye el correspondiente sitio de recombinación. Dado que las recombinasas usadas en los procedimientos de la invención no pueden catalizar la reacción inversa, la integración es estable. Dichos procedimientos son útiles para, por ejemplo, obtener la integración estable en el cromosoma eucariota de un transgén que está presente en el plásmido.

- 50 Los sitios de unión para recombinación también pueden estar presentes en la misma molécula de ácido nucleico. En estos casos, el producto resultante normalmente depende de la orientación relativa de los sitios de unión. Por ejemplo, generalmente, la recombinación entre sitios que están paralelos o en orientación directa dará lugar a la escisión de todo el ADN que se encuentre entre los sitios de unión para recombinación. Por el contrario, la recombinación entre sitios de unión que están en orientación inversa puede dar lugar a la inversión del ADN intermedio. Asimismo, el ácido nucleico reorganizado resultante es estable en cuanto a que la recombinación es irreversible en ausencia de uno o más factores adicionales, generalmente codificados por el bacteriófago concreto o por la célula huésped del bacteriófago del que deriva la recombinasa, que normalmente no se encuentra en células eucariotas. Un ejemplo de una aplicación para la que este procedimiento es útil implica la colocación de un promotor
- 60 entre los sitios de unión para recombinación. Si el promotor está inicialmente en orientación opuesta con respecto a una secuencia de codificación que se va a expresar según el promotor y los sitios de recombinación que flanquean al promotor están en orientación invertida, el contacto de los sitios de unión para recombinación dará lugar a la inversión del promotor, lo que colocará al promotor en la orientación correcta para dirigir la expresión de la secuencia de codificación. De un modo similar, si el promotor está inicialmente en la orientación correcta para la expresión y los sitios de unión para recombinación están en la misma orientación, el contacto de los sitios de unión para recombinación con la recombinasa puede dar lugar a la escisión del fragmento promotor, de modo que se
- 65

detiene la secuencia de codificación.

Los procedimientos de la invención son también útiles para obtener translocaciones de cromosomas. Por ejemplo, en estas realizaciones, un sitio de unión para recombinación se coloca en un cromosoma y un segundo sitio de unión para recombinación que puede servir como sustrato para la recombinación con el primer sitio de unión para recombinación se coloca en un segundo cromosoma. Tras el contacto de los sitios de unión para recombinación con una recombinasa, se produce recombinación que da lugar al intercambio de los dos brazos del cromosoma. Por ejemplo, se pueden construir dos cepas de un organismo, una cepa que incluye el primer sitio de unión para recombinación y la segunda cepa que contiene el segundo sitio de unión para recombinación. Después, las dos cepas se cruzan para obtener una cepa progenie que incluye ambos sitios de unión para recombinación. Tras el contacto de los sitios de unión con la recombinasa se produce el intercambio de los brazos del cromosoma.

Recombinasas

Las recombinasas usadas en la práctica de la presente invención se pueden introducir en una célula diana antes, junto con o después de la introducción de un vector para dirigir. La recombinasa puede introducirse directamente en una célula como proteína usando, por ejemplo, liposomas, partículas recubiertas o microinyección. Como alternativa, se puede introducir en la célula un polinucleótido, ADN o ARN mensajero, que codifica la recombinasa usando un vector de expresión adecuado. Los componentes del vector a dirigir descrito con anterioridad son útiles en la construcción de casetes de expresión que contienen secuencias que codifican una recombinasa de interés. No obstante, la expresión de la recombinasa se puede regular de otras formas, por ejemplo colocando la expresión de la recombinasa sometida al control de un promotor regulable (es decir, un promotor cuya expresión puede inducirse o reprimirse de forma selectiva).

Las recombinasas para usar en la práctica de la presente invención se pueden producir de forma recombinante o purificar como se ha descrito anteriormente. Los polipéptidos que tienen la actividad recombinasa deseada se pueden purificar hasta obtener un grado de pureza deseado mediante procedimientos conocidos en la técnica como precipitación de proteínas con sulfato amónico, purificación, incluyendo, entre otros, fraccionamiento por tamaño, cromatografía de afinidad, HPLC, cromatografía de intercambio iónico, cromatografía de afinidad en heparina-agarosa (por ejemplo, Thorpe y Smith, Proc. Nat. Acad. Sci. 95: 5505-5510, 1998).

Los polipéptidos de recombinasa y los ácidos nucleicos que codifican los polipéptidos de recombinasa de la presente invención se describen en el ejemplo 1 y se pueden obtener usando procedimientos rutinarios conocidos por los expertos en la técnica. La recombinasa de la invención es una secuencia de polinucleótidos aislada que comprende un ácido nucleico que tiene una identidad de al menos un 90 % con la secuencia de ácido nucleico seleccionada del grupo que consiste en la SEC ID N°: 5 y la SEC ID N°: 9, en la que el polipéptido codificado por el ácido nucleico tiene actividad recombinasa. Más preferentemente, la recombinasa es una secuencia de polinucleótidos aislada que comprende la secuencia de ácido nucleico seleccionada del grupo que consiste en la SEC ID N°: 5 y la SEC ID N°: 9. Incluso más preferentemente, la recombinasa es una secuencia de polinucleótidos aislada que comprende la secuencia de ácido nucleico que codifica una recombinasa seleccionada del grupo que consiste en una recombinasa Bxb1 y una recombinasa ϕ Rv1.

Las recombinasas se pueden introducir en las células eucariotas que contienen los sitios de unión para recombinación en los que se desea la recombinación mediante cualquier procedimiento adecuado. En la técnica se conocen bien los procedimientos de introducción en las células de proteínas funcionales, por ejemplo mediante microinyección u otros procedimientos. La introducción de proteínas recombinasa purificadas garantiza una presencia transitoria de la proteína o su función, que a menudo es una realización preferida. Como alternativa, un gen que codifica la recombinasa se puede incluir en un vector de expresión usado para transformar la célula, en el que el polinucleótido que codifica la recombinasa está unido operativamente a un promotor que participa en la expresión del polinucleótido en la célula eucariota. El polipéptido de la recombinasa también se puede introducir en la célula eucariota mediante ARN mensajero que codifica el polipéptido de la recombinasa. Generalmente se prefiere que la recombinasa esté presente únicamente durante el tiempo necesario para la inserción de los fragmentos de ácido nucleico en el genoma que se está modificando. Por tanto, no cabe esperar que la ausencia de permanencia asociada con la mayoría de los vectores de expresión sea perjudicial. Se puede introducir el gen de la recombinasa en la célula antes, después o de forma simultánea a la introducción del polinucleótido exógeno de interés. En una realización, el gen de la recombinasa está presente dentro del vector que es portador del polinucleótido que tiene que insertarse; incluso el gen de la recombinasa puede incluirse dentro del polinucleótido. En otras realizaciones, el gen de la recombinasa se introduce en un organismo eucariota transgénico, por ejemplo una planta, animal, hongo transgénicos o similares, que después se cruza con un organismo que contiene los sitios de recombinación correspondientes. Se pueden crear células o animales transgénicos que expresan una recombinasa de forma constitutiva o sometida a promotores específicos de célula, específicos de tejido, específicos del desarrollo, específicos de los orgánulos o inducibles o reprimibles por moléculas pequeñas. Las recombinasas también se pueden expresar como proteína de fusión con otros péptidos, proteínas, péptidos señal de localización nuclear, péptidos señal o péptidos señal específicos de orgánulo (por ejemplo, péptidos de tránsito a mitocondrias o cloroplastos para facilitar la recombinación en mitocondrias o cloroplastos).

Vectores/construcciones

Las construcciones para dirigir contemplados por la invención pueden contener fragmentos de ácido nucleico adicionales, tales como secuencias control, secuencias marcadoras, secuencias de selección y similares como se trata más adelante.

La presente invención también proporciona medios para la inserción dirigida de un polinucleótido (o secuencia(s) de ácido nucleico) de interés en un genoma, por ejemplo (i) proporcionando una recombinasa, en la que la recombinasa es capaz de facilitar la recombinación entre un primer sitio de recombinación y un segundo sitio de recombinación, (ii) proporcionando una construcción dirigida que tiene una primera secuencia de recombinación y un polinucleótido de interés, (iii) introduciendo la recombinasa y la construcción dirigida en una célula que contiene en su ácido nucleico el segundo sitio de recombinación, en el que dicha introducción se realiza en condiciones que permiten que la recombinasa facilite un evento de recombinación entre el primero y el segundo sitio de recombinación.

La presente invención también se refiere a un vector para la integración específica de sitio de una secuencia de polinucleótidos en el genoma de una célula eucariota aislada, comprendiendo dicho vector un polinucleótido de interés y un segundo sitio de recombinación *attB* o *attP*, en el que dicho segundo sitio de recombinación *attB* o *attP* comprende una secuencia de polinucleótidos que recombina con un primer sitio de recombinación *attP* o *attB* o seudositio *attP* o seudositio *attB* en el genoma de dicha célula eucariota aislada, en el que dicho vector comprende además un polinucleótido que codifica una recombinasa específica de sitio seleccionada del grupo que consiste en una recombinasa de fago de *Mycobacterium tuberculosis* y una recombinasa de fago de *Mycobacterium smegmatis*, con la condición de que cuando el primer sitio de recombinación es *attB* o seudo *attB*, el segundo sitio de recombinación es *attP* y cuando el primer sitio de recombinación es *attP* o seudo *attP*, el segundo sitio de recombinación es *attB*. La recombinasa se selecciona del grupo que consiste en una recombinasa ϕ Rv1 y una recombinasa Bxb1.

Los polinucleótidos de interés pueden incluir, entre otros, casetes de expresión que codifican productos polipeptídicos. Las construcciones dirigidas pueden ser circulares o lineales y también pueden contener marcadores seleccionables, un origen de replicación y otros elementos.

Diversos vectores de expresión son adecuados para usar en la práctica de la presente invención, tanto para expresión en procariontes como para expresión en eucariotas. En general, la construcción dirigida tendrá una o más de las características siguientes: un promotor, secuencias potenciadoras del promotor, una secuencia marcadora de selección, un origen de replicación, una secuencia de elemento inducible, una secuencia marcadora de epítipo y similares.

El promotor y las secuencias potenciadoras del promotor son secuencias de ADN a las que se une la ARN polimerasa y se inicia la transcripción. El promotor determina la polaridad del transcrito especificando qué hebra se transcribirá. Los promotores bacterianos consisten en secuencias consenso, -35 y -10 nucleótidos respecto al inicio de la transcripción, que están unidas por un factor sigma específico y ARN polimerasa. Los promotores eucariotas son más complejos. La mayoría de los promotores usados en los vectores de expresión se transcriben mediante la acción de la ARN polimerasa II. Los factores de transcripción generales (GTFs) se unen primero a secuencias específicas cerca del inicio y después reclutan la unión de la ARN polimerasa II. Además de estos elementos promotores mínimos, los elementos de secuencia pequeños son reconocidos específicamente por las proteínas modulares de unión a ADN/de transactivación (por ejemplo, AP-1, SP-1) que regulan la actividad de un promotor dado. Los promotores virales realizan la misma función que los promotores bacterianos o eucariotas y proporcionan una ARN polimerasa específica en trans (bacteriófago T7) o reclutan factores celulares y ARN polimerasa (SV40, RSV, CMV). Pueden ser preferibles los promotores virales, ya que son, generalmente, promotores particularmente fuertes.

Los promotores pueden ser, además, constitutivos o regulables (es decir, inducibles o reprimibles). Los elementos inducibles son elementos de secuencia de ADN que actúan junto con promotores y se unen a represores (por ejemplo, el sistema represor lacO/LAC Iq en *E. coli*) o inductores (por ejemplo, sistema de inductores gal1/GAL4 en levaduras). En cualquiera de los casos, la transcripción está prácticamente "apagada" hasta que el promotor se reprime o induce, punto en el que se "enciende" la transcripción.

Ejemplos de promotores constitutivos incluyen el promotor int del bacteriófago λ , el promotor bla de la secuencia del gen de la β -lactamasa de pBR322, el promotor CAT de la secuencia del gen de la cloranfenicol acetiltransferasa de pPR325 y similares. Ejemplos de promotores procariontes inducibles incluyen los promotores principales derecho e izquierdo del bacteriófago (P.sub.L y P.sub.R), los promotores trp, reca, lacZ, AraC y gal de *E. coli*, los promotores de la α -amilasa (Ulmanen Ett et al., J. Bacteriol. 162: 176-182, 1985) y específicos de 28 sigma de *B. subtilis* (Gilman y col., Gene sequence 32: 11-20 (1984)), los promotores de los bacteriófagos de *Bacillus* (Gryczan, en: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), promotores de *Streptomyces* (Ward y col., Mol. Gen. Genet. 203: 468-478, 1986) y similares. Ejemplos de promotores procariontes se revisan en Glick (J. Ind. Microbiol. 1: 277-282, 1987); Cenatiempo (Biochimie 68: 505-516, 1986); y Gottesman (Ann. Rev. Genet. 18: 415-442, 1984).

- Promotores eucariotas preferidos incluyen, entre otros, los siguientes: el promotor de la secuencia del gen de la metalotioneína I de ratón (Hamer y col., *J. Molec. Appl. Genet.* 1: 273-288, 1982), el promotor *TK* del virus Herpes (McKnight, *Cell* 31: 355-365, 1982); el promotor temprano de SV40 (Benoist y col., *Nature (London)* 290: 304-310, 1981); el promotor de la secuencia del gen de las agallas de levadura (Johnston y col., *Proc. Natl. Acad. Sci. (EE.UU.)* 79: 6971-6975, 1982); Silver y col., *Proc. Natl. Acad. Sci. (EE.UU.)* 81: 5951-5955, 1984), el promotor del CMV, el promotor de EF-1, el(los) promotor(es) respondedores a ecdisona, el promotor respondedor a tetraciclina y similares. Ejemplos de promotores para usar en la presente invención se seleccionan de un modo tal que sean funcionales en el tipo celular (y/o animales o plantas) en el que se introducen.
- 10 Los marcadores de selección son elementos valiosos en los vectores de expresión, ya que proporcionan un medio para seleccionar el crecimiento de únicamente las células que contienen un vector. Dichos marcadores son de dos tipos: de resistencia a fármacos y auxotróficos. Un marcador de resistencia a fármacos permite que las células destoxifiquen un fármaco añadido de forma exógena que, de otro modo, mataría a la célula. Los marcadores auxotróficos permiten que las células sinteticen un componente esencial (normalmente un aminoácido) cuando crecen en medio que carece del componente esencial.
- 15 Los genes de marcadores seleccionables habituales incluyen los de la resistencia a antibióticos, tales como ampicilina, tetraciclina, kanamicina, bleomicina, estreptomycin, higromicina, neomicina, Zeocina™ y similares. Los genes autotróficos seleccionables incluyen, por ejemplo, *hisD*, que permite el crecimiento en medio sin histidina en presencia de histidinol.
- 20 Un elemento adicional útil en un vector de expresión es un origen de replicación. Los orígenes de replicación son segmentos de ADN únicos que contienen múltiples secuencias repetidas cortas que son reconocidas por proteínas multiméricas de unión al origen y que desempeñan un papel principal en el ensamblaje de las enzimas de replicación de ADN en el sitio de origen. Orígenes de replicación adecuados para usar en los vectores de expresión empleados en el presente documento incluyen *E. coli* oriC, el origen plasmídico *colE1*, 2μ y ARS (ambos útiles en sistemas de levaduras), *sf1*, SV40, EBV oriP (útiles en sistemas de mamífero) y similares.
- 25 Las marcas de epítomos son secuencias peptídicas cortas que son reconocidas por anticuerpos específicos de epítomos. Una proteína de fusión que comprende una proteína recombinante y un marcador de epítomo puede purificarse simplemente y fácilmente usando un anticuerpo unido a una resina de cromatografía. La presencia de un marcador de epítomo permite además detectar la proteína recombinante en ensayos posteriores, tales como transferencias de tipo Western, sin tener que producir un anticuerpo específico para la propia proteína recombinante. Ejemplos de marcadores de epítomo de uso habitual incluyen V5, glutatión-S-transferasa (GST), hemaglutinina (HA), el péptido Phe-His-His-Thr-Thr, los dominios de unión a quitina y similares.
- 30 Un elemento adicional útil en un vector de expresión es un sitio de clonación múltiple o poliengarce. El ADN sintético que codifica una serie de sitios de reconocimiento de la endonucleasa de restricción se inserta en un vector plasmídico, por ejemplo cadena abajo del elemento promotor. Estos sitios se someten a ingeniería para una clonación conveniente del ADN en el vector en una posición específica.
- 40 Los elementos anteriores se pueden combinar para producir vectores de expresión adecuados para usar en los procedimientos de la invención. Los expertos en la técnica podrían seleccionar y combinar los elementos adecuados para uso en su sistema concreto en vista de las enseñanzas de la presente especificación. Vectores procariontes adecuados incluyen plásmidos tales como los capaces de replicarse en *E. coli* (por ejemplo, pBR322, *ColE1*, pSC101, PACYC 184, *itVX*, PRSET, pBAD (Invitrogen, Carlsbad, Calif.) y similares). Dichos plásmidos se divulgan en Sambrook (véase "Molecular Cloning: A Laboratory Manual," segunda edición, editado por Sambrook, Fritsch, y Maniatis, Cold Spring Harbor Laboratory, (1989)). Los plásmidos de *Bacillus* incluyen pC194, pC221, pT127 y similares y se divulgan en Gryczan (en: *The Molecular Biology of the Bacilli*, Academic Press, NY (1982), pág. 307-329). Plásmidos de *Streptomyces* adecuados incluyen *pli101* (Kendall y col., *J. Bacteriol.* 169: 4177-4183, 1987) y bacteriófagos de *Streptomyces* tales como ϕ C31 (Chater y col., en: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kiado, Budapest, Hungría (1986), pág. 45-54). Los plásmidos de *Pseudomonas* se revisan en John y col., (*Rev. Infect. Dis.* 8: 693-704, 1986), e Izaki (*Jpn. J. Bacteriol.* 33: 729-742, 1978).
- 50 Plásmidos eucariotas adecuados incluyen, por ejemplo, BPV, EBV, variolovacunal, SV40, círculo de 2 micrómetros, pCDNA3.1, pCDNA3.1/GS, pDual, pYES2/GS, pMT, pIND, pIND(Sp1), pVgRXR (Invitrogen) y similares, o sus derivados. Dichos plásmidos son bien conocidos en la técnica (Botstein y col., *Miami Wntr. SyTnp.* 19: 265-274, 1982; Broach, en: "The Molecular Biology of the Yeast *Saccharomyces*: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pág. 445-470, 1981; Broach, *Cell* 28: 203-204, 1982; Dilon y col., *J. Clin. Hematol. Oncol.* 10: 39-48, 1980; Maniatis, en: *Cell Biology: A Comprehensive Treatise*, Vol. 3, Gene Sequence Expression, Academic Press, NY, pág. 563-608, 1980. Los casetes dirigidos descritos en el presente documento se puede construir usando las metodologías conocidas en la técnica de la biología molecular (véase, por ejemplo, Ausubel o Maniatis), en vista de las enseñanzas de la memoria descriptiva. Como se ha descrito anteriormente, las construcciones dirigidas se ensamblan insertando en una estructura de vector adecuada, un sitio de unión para recombinación, secuencias que codifican polinucleótidos de interés unidas operativamente a un promotor de interés
- 65 y opcionalmente, una secuencia que codifica un marcador de selección positiva.

Un procedimiento preferido de obtención de polinucleótidos, incluidas las secuencias reguladoras adecuadas (por ejemplo, promotores) es la PCR. Los procedimientos generales para la PCR se enseñan en MacPherson y col. PCR: A PRACTICAL APPROACH, (IRL Press at Oxford University Press, (1991)). Las condiciones para la PCR para cada reacción de aplicación se pueden determinar empíricamente. Una serie de parámetros influyen sobre el éxito de una reacción. Entre estos parámetros se encuentran la temperatura y el tiempo de hibridación, el tiempo de extensión, la concentración de Mg^{2+} y ATP, el pH y la concentración relativa de los cebadores, moldes y desoxirribonucleótidos. Después de la amplificación, los fragmentos resultantes se pueden detectar mediante electroforesis en gel de agarosa, seguida de visualización con tinción con bromuro de etidio e iluminación ultravioleta.

Los casetes de expresión, construcciones dirigidas, vectores, recombinasas y secuencias de codificación de la recombinasa de la presente invención se pueden formular en kits. Los componentes de dichos kits pueden incluir, entre otros, recipientes, instrucciones, soluciones, tampones, desechables y hardware.

15 Métodos

La presente invención comprende un método para obtener una recombinación específica de sitio en una célula eucariota aislada, comprendiendo el método: proporcionar una célula eucariota aislada que comprende un primer sitio de recombinación y un segundo sitio de recombinación con una secuencia de polinucleótidos flanqueada por un tercer sitio de recombinación y un cuarto sitio de recombinación; poner en contacto los sitios de recombinación con un polipéptido de recombinasa procariota, que tiene como resultado la recombinación entre los sitios de recombinación, en el que el polipéptido de recombinasa puede participar en la recombinación entre el primero y el tercer sitios de recombinación y el segundo y el cuarto sitios de recombinación, el primero y el segundo sitios de recombinación son *attP* o *attB*, el tercero y el cuarto sitios de recombinación son *attB* o *attP* y la recombinasa se selecciona del grupo que consiste en una recombinasa del fago de *Mycobacterium tuberculosis* y una recombinasa del fago de *Mycobacterium smegmatis*, siempre que cuando el primero y el segundo sitios de recombinación son *attB*, el tercero y el cuarto sitios de recombinación son *attP* y cuando el primero y el segundo sitios de recombinación son *attP* el tercero y el cuarto sitios de recombinación son *attB*. La recombinasa se selecciona del grupo que consiste en una recombinasa ϕ Rv1 y una recombinasa Bxb1.

Otro método de la presente invención proporciona obtener múltiples recombinaciones específicas de sitio en una célula eucariota aislada, comprendiendo el método: proporcionar una célula eucariota aislada que comprende un primer sitio de recombinación y un segundo sitio de recombinación con una secuencia de polinucleótidos flanqueada por un tercer sitio de recombinación y un cuarto sitio de recombinación; poniendo en contacto el primero y el segundo sitios de recombinación con un primer polipéptido de recombinasa procariota, poniendo en contacto el tercero y el cuarto sitios de recombinación con un segundo polipéptido de recombinasa procariota que tiene como resultado la recombinación entre el primero y el segundo sitios de recombinación y la recombinación entre el tercero y el cuarto sitios de recombinación, en el que el primer polipéptido de recombinasa puede participar en la recombinación entre el primero y el segundo sitio de recombinación y el segundo polipéptido de recombinasa puede participar en la recombinación entre el tercero y el cuarto sitios de recombinación, la primera y la segunda recombinasa se seleccionan del grupo que consiste en una recombinasa de fago de *Mycobacterium tuberculosis* y una recombinasa de fago de *Mycobacterium smegmatis*, siempre que el primer polipéptido de recombinasa y el segundo polipéptido de recombinasa sean diferentes. El método puede comprender además un quinto sitio de recombinación y un sexto sitio de recombinación y un tercer polipéptido de recombinasa, en el que el tercer polipéptido de recombinasa puede participar en la recombinación entre el quinto y el sexto sitios de recombinación, siempre que el tercer polipéptido de recombinasa sea diferente al primero y al segundo polipéptidos de recombinasa.

Células

Las células adecuadas para modificación que usan los métodos de la invención incluyen células procariotas y células eucariotas. Las células procariotas son células que carecen de un núcleo definido. Ejemplos de células procariotas incluyen células bacterianas, células de micoplasma y células de arqueobacterias. Las células procariotas particularmente preferidas incluyen las que son útiles en varios tipos de sistemas de ensayo (tratados con mayor detalle más adelante) o las que tienen alguna utilidad industrial tales como *Klebsiella oxytoca* (producción de etanol), *Clostridium acetobutylicum* (producción de butanol) y similares (véase Green y Bennet, Biotech & Bioengineering 58: 215-221, 1998; Ingram y col., Biotech & Bioengineering 58: 204-206, 1998). Las células eucariotas adecuadas incluyen tanto células animales (tales como de insectos, roedores, vacas, cabras, conejos, ovejas, primates no humanos, humanos y similares) como células vegetales (tales como arroz, maíz, algodón, tabaco, tomate, patata y similares). Tipos de células aplicables a fines concretos se tratan con más detalla más adelante.

Otra realización más comprende células modificadas genéticamente aisladas. Células adecuadas pueden ser procariotas o eucariotas, como se ha tratado anteriormente. Las células modificadas genéticamente pueden ser organismos unicelulares o pueden derivar de organismos multicelulares. Por "aislado" en referencia a células modificadas genéticamente derivadas de organismos multicelulares se quiere decir las células que están fuera del cuerpo vivo, ya sea en un ambiente vegetal, animal o artificial. El uso del término aislado no implica que las células

modificadas genéticamente sean las únicas células presentes.

En una realización, las células modificadas genéticamente contienen una cualquiera de las construcciones de ácido nucleico de la invención. En una segunda realización, una recombinasa que reconoce específicamente secuencias de recombinación se introduce en células modificadas genéticamente que contienen una de las construcciones de ácido nucleico de la invención en condiciones tales que la o las secuencias de ácido nucleico de interés se insertarán en el genoma. Por tanto, las células modificadas genéticamente poseen un genoma modificado. En la técnica se conocen bien los procedimientos de introducción de dicha recombinasa y se tratan más adelante.

Las células modificadas genéticamente se pueden producir de diversos modos. Los organismos unicelulares se pueden modificar para producir sustancias comercialmente valiosas, tales como proteínas recombinantes, disolventes industriales, enzimas industrialmente útiles y similares. Organismos unicelulares preferidos incluyen hongos, tales como levaduras (por ejemplo, *S. pombe*, *Pichia pastoris*, *S. cerevisiae* (tales como INVSc1) y similares) *Aspergillus* y similares y bacterias tales como *Klebsiella*, *Streptomyces* y similares.

Las células aisladas de organismos multicelulares pueden ser útiles de forma similar, incluidas células de insectos, células de mamífero y células vegetales. Las células de mamífero que pueden ser útiles incluyen las derivadas de roedores, primates y similares. Incluyen células de ovario de hámster chino (CHO), células HeLa, células madre neurales de ratón, células estromales de médula ósea de rata, células de origen fibroblástico tales como VERO, 3T3 o CHOK1, células HEK 293 o células de origen linfóide (tales como células 32D) y sus derivados.

Además, también están disponibles como huésped células vegetales, tales como las células BY2 de tabaco y se dispone de secuencias control compatibles con las células vegetales, tales como las secuencias 35S y 19S del virus del mosaico de la coliflor, del promotor de la nopalina sintasa y de la señal de poliadenilación y similares. Se pueden usar células vegetales transgénicas adecuadas para producir plantas transgénicas.

Otro huésped preferido es una célula de insecto, por ejemplo de *Drosophila larvae*. Usando células de insecto como huéspedes se puede usar el promotor de la alcohol deshidrogenasa de *Drosophila* (Rubin, Science 240: 1453-1459, 1988). Como alternativa, se pueden modificar vectores de baculovirus para expresar cantidades grandes de péptido codificado por una secuencia de ácido nucleico deseada en células de insecto (Jasny, Science 238: 1653, 1987); Miller y col., en: Genetic Engineering (1986), Setlow, J. K. y col., eds., Plenum, vol. 8, pág. 277-297).

Las células modificadas genéticamente son útiles adicionalmente como herramientas para la detección selectiva de sustancias capaces de modular la actividad de una proteína codificada por un fragmento de ácido nucleico de interés. Por tanto, una realización adicional de la invención comprende procedimientos de detección selectiva que comprende poner en contacto células modificadas genéticamente de la invención con una sustancia de ensayo y monitorizar las células para detectar un cambio en el fenotipo celular, la proliferación celular, la diferenciación celular, la actividad enzimática de la proteína o la interacción entre la proteína y una pareja de unión natural de la proteína en comparación con las células de ensayo que entran en contacto con la sustancia de ensayo.

Se puede evaluar diversas sustancias de ensayo usando las células modificadas genéticamente que incluyen péptidos, proteínas, anticuerpos, compuestos orgánicos de bajo peso molecular, productos naturales derivados de, por ejemplo, células fúngicas o vegetales y similares. Por "compuesto orgánicos de bajo peso molecular" se quiere decir una especie química con un peso molecular de, generalmente, menos de 50-1.000. Las fuentes de sustancias de ensayo son bien conocidas para los expertos en la técnica.

Los expertos en la técnica también conocen varios procedimientos de ensayo que usan células. Incluyen, por ejemplo, ensayos de la actividad enzimática (Hirth y col., Patente de EE.UU. Nº 5.763.198, presentada el 9 de junio de 1998), ensayos de la unión de una sustancia de ensayo a una proteína expresada por las células modificadas genéticamente, ensayos de activación transcripcional de un gen indicador y similares.

Las células modificadas mediante los procedimientos de la presente invención se pueden mantener en condiciones que, por ejemplo, (i) las mantengan vivas pero que no estimulen su crecimiento, (ii) estimulen el crecimiento de las células y/o (iii) hagan que las células se diferencien o desdiferencien. Normalmente, las condiciones de cultivo celular permiten la acción de la recombinasa en las células, aunque la regulación de la actividad de la recombinasa también se puede modular mediante condiciones de cultivo (por ejemplo, elevar o disminuir la temperatura a la que se cultivan las células). Para una célula dada, en la técnica se conocen las condiciones de cultivo para el tipo de célula, tejido u organismo.

60 Plantas transgénicas y animales transgénicos no humanos

Los animales transgénicos se pueden producir usando los procedimientos de la presente invención para que sirvan como sistema modelo para el estudio de varios trastornos y para la detección selectiva de fármacos que modulan dichos trastornos. En el texto siguiente, las expresiones "células madre totipotenciales o pluripotenciales" excluirán las derivadas usando embriones humanos.

Una planta o animal “transgénico” hace referencia a una planta o animal modificado genéticamente o a la descendencia de plantas o animales modificados genéticamente. Una planta o animal transgénico normalmente contiene material de al menos un organismo no relacionado, tal como de un virus. El término “animal”, como se usa en el contexto de los organismos transgénicos significa todas las especies salvo los seres humanos. También incluye un animal individual en todas las etapas del desarrollo, incluidas las etapas embrionarias y fetal. Se incluyen los animales de granja (por ejemplo, pollos, cerdos, cabras, ovejas, vacas, caballos, conejos y similares), roedores (tales como ratones) y animales domésticos (por ejemplo, perros y gatos). En una realización preferida, el animal es un ratón o una rata.

El término planta o animal “quimérico” se usa para hacer referencia a plantas o animales en los que se encuentra el gen heterólogo o en los que el gen heterólogo se expresa en algunas pero no en todas las células de la planta o animal.

El término animal transgénico también incluye un animal transgénico de una línea de células germinales. Un “animal transgénico de una línea de células germinales” es un animal transgénico en el que la información genética proporcionada por el procedimiento de la invención ha sido captada e incorporada en una célula de la línea germinal, lo que le confiere la capacidad de transferir la información a la descendencia. Si, de hecho, dicha descendencia posee alguna o toda esa información, entonces también son animales transgénicos.

En la técnica se conocen procedimientos de generación de plantas y animales transgénicos y se pueden usar en combinación con las enseñanzas de la presente solicitud.

En una realización, un animal transgénico se produce introduciendo en un embrión de una célula una construcción de ácido nucleico que comprende un primer sitio de recombinación capaz de recombinarse con un segundo sitio de recombinación dentro del genoma del organismo del que derivó la célula y un fragmento de ácido nucleico de interés, de un modo tal que el fragmento de ácido nucleico de interés está integrado de forma estable en el ADN de las células de la línea germinal del animal maduro y se hereda siguiendo un patrón mendeliano. En esta realización, el fragmento de ácido nucleico de interés puede ser uno cualquiera de los fragmentos descrito anteriormente. Como alternativa, la secuencia de ácido nucleico de interés puede codificar un producto exógeno que rompa o interfiera en la expresión de una proteína producida endógenamente de interés, lo que da un animal transgénico con menor expresión de la proteína de interés.

Se dispone de varios procedimientos para producir animales transgénicos. Una construcción de ácido nucleico de la invención se puede inyectar en el pronúcleo, o el citoplasma, de un huevo fertilizado antes de la fusión de los pronúcleos de sexo masculino y femenino, o se puede inyectar en el núcleo de una célula embrionaria (por ejemplo, el núcleo de un embrión de dos células) tras el inicio de la división celular (Brinster y col., Proc. Nat. Acad. Sci. EE.UU. 82: 4438, 1985). Los embriones se pueden infectar con virus, especialmente retrovirus, modificados con un sitio de recombinación *attD* y una secuencia de ácido nucleico de interés. La célula puede además tratarse con una recombinasa específica de sitio como se ha descrito anteriormente para estimular la integración de la secuencia de ácido nucleico de interés en el genoma.

Solo a modo de ejemplo, para preparar un ratón transgénico se induce la hiperovulación de ratones hembra. Después de permitir el apareamiento, se sacrifica a las hembras mediante asfixia con CO₂ o dislocación cervical y se recuperan los embriones de los oviductos escindidos. Se extraen los cúmulos de células circundantes. Después, los embriones pronucleares se lavan y almacenan hasta el momento de la inyección. Ratones hembras adultos en ciclo aleatorio se aparean con machos vasectomizados. Las hembras receptoras se aparean al mismo tiempo como hembras donantes. Después, los embriones se transfieren quirúrgicamente. El procedimiento para generar ratas transgénicas es similar al de los ratones. Véase, Hammer y col., Cell 63: 1099-1112, 1990). Se pueden obtener roedores adecuados para experimentos transgénicos a partir de fuentes comerciales estándar tales como Charles River (Wilmington, Mass.), Taconic (Germantown, N.Y.), Harlan Sprague Dawley (Indianapolis, Ind.), etc.

Los expertos en la técnica conocen bien los procedimientos para manipular el embrión de roedores y para microinyección de ADN en el pronúcleo del cigoto (Hogan y col., *supra*). Los procedimientos de microinyección para peces, huevos de anfibio y aves se detallan en Houdebine y Chourrout, *Experientia* 47: 897-905, 1991). Otros procedimientos para la introducción de ADN en tejidos de animales se describen en la patente de EE.UU. N° 4,945,050 (Sandford y col., jul. 30, 1990).

Las células madre totipotenciales o pluripotenciales derivadas de la masa de células internas del embrión y estabilizadas en cultivo se pueden manipular en cultivo para incorporar secuencias de ácido nucleico usando los procedimientos de la invención. Se puede producir un animal transgénico a partir de estas células mediante inyección en un blastocisto que después se implanta en una madre sustituta y se le permite llegar a término.

Los expertos en la técnica también conocen los procedimientos para cultivar las células madre y la posterior producción de animales transgénicos mediante la introducción de ADN en las células madre usando procedimientos tales como electroporación, precipitación en fosfato cálcico/ADN, microinyección, fusión de liposomas, infección retroviral y similares. Véase, por ejemplo, *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, E. J.

Robertson, ed., IRL Press, 1987). Revisiones de procedimientos de laboratorio estándar para la microinyección de ADN heterólogo en huevos fertilizados de mamífero (ratón, cerdo, conejo, oveja, cabra, vaca) incluyen: Hogan y col., *Manipulating the Mouse Embryo* (Cold Spring Harbor Press 1986); Krimpenfort y col., 1991, *Bio/Technology* 9: 86; Palmiter y col., 1985, *Cell* 41: 343; Kraemer y col., *Genetic Manipulation of the Early Mammalian Embryo* (Cold Spring Harbor Laboratory Press 1985); Hammer y col., 1985, *Nature*, 315: 680; Purcel y col., 1986, *Science*, 244: 1281; Wagner y col., patente de EE.UU. Nº 5.175.385; Krimpenfort y col., patente de EE.UU. Nº 5.175.384, cuyos respectivos contenidos se incorporan por referencia.

La fase final del procedimiento es inyectar las células ES dirigidas en blastocitos y transferir los blastocitos en hembras pseudogestantes. Los animales quiméricos resultantes se crían y la descendencia se analiza mediante transferencia de tipo Southern para identificar los individuos portadores del transgén. Terceros ha comentado los procedimientos para la producción de mamíferos no roedores y otros animales (véase Houdebine y Chourrout, *supra*; Pursel y col., *Science* 244: 1281-1288, 1989; y Simms y col., *Bio/Technology* 6: 179-183, 1988) Los animales portadores del transgén se pueden identificar mediante procedimientos bien conocidos en la técnica, por ejemplo mediante transferencia puntual o transferencia de tipo Southern.

El término transgénico, como se usa en el presente documento, incluye adicionalmente cualquier organismo cuyo genoma se ha alterado mediante manipulación *in vitro* del embrión temprano o huevo fertilizado o mediante tecnología transgénica para inducir deficiencia en un gen específico. La expresión "deficiencia génica", como se usa en el presente documento, hace referencia a la rotura dirigida de un gen *in vivo* con pérdida de función que se ha conseguido mediante el uso del vector de la invención. En una realizaciones, los animales transgénicos con deficiencias génicas son aquellos en los que el gen diana se ha convertido en no funcional mediante una inserción dirigida en el gen que se va a hacer no funcional apuntando a un sitio de seudorecombinación localizado dentro de la secuencia génica.

Terapia génica y trastornos

En el presente documento se describe un procedimiento de tratamiento de un trastorno en un sujeto que necesite dicho tratamiento. En una realización del procedimiento, al menos una célula o tipo de célula (o de tejido etc.) del sujeto tiene un sitio de recombinación. Esta(s) célula(s) se transforma(n) con una construcción de ácido nucleico (una "construcción dirigida") que comprende una segunda secuencia de recombinación y uno o más polinucleótidos de interés (normalmente un gen terapéutico). En la misma célula se introduce una recombinasa que reconoce específicamente las secuencias de recombinación en condiciones tales que la secuencia de ácido nucleico de interés se inserta en el genoma a través de un evento de recombinación entre el primero y el segundo sitios de recombinación. Los sujetos tratables usando los procedimientos de la invención incluyen seres humanos y animales no humanos. Dichos procedimientos usan las construcciones dirigidas y las recombinasas de la presente invención.

Diversos trastornos se pueden tratar usando el procedimiento de la invención, incluidos trastornos monogénicos, enfermedades infecciosas, trastornos adquiridos, cáncer y similares. Trastornos monogénicos de ejemplo incluyen deficiencia de ADA, fibrosis quística, hipercolesterolemia familiar, hemofilia, enfermedad granulomatosa crónica, distrofia muscular de Duchenne, anemia de Fanconi, drepanocitosis, enfermedad de Gaucher, síndrome de Hunter, SCID ligada al cromosoma X y similares.

Entre las enfermedades infecciosas tratables usando los procedimientos descritos en el presente documento se incluyen infección con varios tipos de virus, incluyendo el virus linfotrópico de linfocitos T humano, el virus de la gripe, el virus del papiloma, el virus de la hepatitis, el virus del herpes, el virus de Epstein-Bar, los virus de la inmunodeficiencia (VIH y similares), el citomegalovirus y similares. También se incluyen infecciones con otros organismos patogénicos tales como *Mycobacterium Tuberculosis*, *Mycoplasma pneumoniae* y similares, o parásitos tales como *Plasmodium falciparum* y similares.

La expresión "trastorno adquirido", tal como se usa en el presente documento, se refiere a un trastorno no congénito. Dichos trastornos se consideran, en general, más complejos que los trastornos monogénicos y pueden ser el resultado de actividad inadecuada o indeseada de uno o más genes. Ejemplos de dichos trastornos incluyen arteriopatía periférica, artritis reumatoide, arteriopatía coronaria y similares.

Un grupo concreto de trastornos adquiridos tratables usando los procedimientos de la invención incluyen varios tipos de cáncer, incluyendo cánceres de tumores sólidos y hematopoyéticos tales como leucemias y linfomas. Los tumores sólidos que son tratables usando el procedimiento de la invención incluyen carcinomas, sarcomas, osteomas, fibrosarcomas, condrosarcomas y similares. Cánceres específicos incluyen cáncer de mama, cáncer de cerebro, cáncer de pulmón (microcítico y macrocítico), cáncer de colon, cáncer de páncreas, cáncer de próstata, cáncer gástrico, cáncer de vejiga, cáncer de riñón, cáncer de cabeza y cuello y similares.

La idoneidad de la ubicación concreta en el genoma depende, en parte, del trastorno concreto que se esté tratando. Por ejemplo, si el trastorno es un trastorno monogénico y el tratamiento deseado es la adición de un ácido nucleico terapéutico que codifica una forma no mutada del ácido nucleico que se piensa que es el agente causante del trastorno, una ubicación adecuada puede ser una región del genoma que no codifica ninguna proteína conocida y

que permite un nivel de expresión razonable del ácido nucleico añadido. En la técnica se conocen bien los procedimientos de identificación de ubicaciones adecuadas en el genoma y se describen adicionalmente en los ejemplos que figuran a continuación.

5 La construcción de ácido nucleico útil en la presente realización está compuesta adicionalmente de uno o más fragmentos de ácido nucleico de interés. Fragmentos de ácido nucleico de interés preferidos para usar en la presente realización son los genes terapéuticos y/o las regiones control, como se ha definido anteriormente. La elección de la secuencia de ácido nucleico dependerá de la naturaleza del trastorno que se va a tratar. Por ejemplo, una construcción de ácido nucleico destinada a tratar la hemofilia, que está producida por una deficiencia del factor de la coagulación IX, puede comprender un fragmento de ácido nucleico que codifica el factor IX funcional. Una construcción de ácido nucleico destinada a tratar la arteriopatía periférica obstructiva puede comprender fragmentos de ácido nucleico que codifican proteínas que estimulan el crecimiento de nuevos vasos sanguíneos, tales como, por ejemplo, el factor de crecimiento endotelial vascular, el factor de crecimiento derivado de plaquetas y similares. Los expertos en la técnica deberían reconocer fácilmente qué fragmentos de ácido nucleico de interés serían útiles en el tratamiento de un trastorno concreto.

La construcción de ácido nucleico se puede administrar al sujeto que se está tratando usando diversos procedimientos. La administración puede tener lugar *in vivo* o *ex vivo*. Con "*in vivo*" se quiere decir en el cuerpo vivo de un animal. Con "*ex vivo*" se quiere decir que las células u órganos se modifican fuera del cuerpo y estas células u órganos normalmente se devuelven a un cuerpo vivo.

En la técnica se conocen bien los procedimientos para la administración terapéutica de construcciones de ácido nucleico. Las construcciones de ácido nucleico se pueden liberar con lípidos catiónicos (Goddard y col., Gene Therapy, 4: 1231-1236, 1997; Gorman y col., Gene Therapy 4: 983-992, 1997; Chadwick y col., Gene Therapy 4: 937-942, 1997; Gokhale y col., Gene Therapy 4: 1289-1299, 1997; Gao y Huang, Gene Therapy 2: 710-722, 1995, todos ellos incorporados por referencia en el presente documento) usando vectores virales (Monahan y col., Gene Therapy 4: 40-49, 1997; Onodera y col., Blood 91: 30-36, 1998, todos ellos incorporados por referencia en el presente documento), mediante captación de "ADN desnudo" y similares. Las técnicas bien conocidas en la materia para la transfección de células (véase la discusión anterior) se pueden usar para la administración *ex vivo* de construcciones de ácido nucleico. El médico individual escogerá la formulación exacta, la vía de administración y la dosis según la afección del paciente. (Véase, por ejemplo, Fingl y col., 1975, en "The Pharmacological Basis of Therapeutics", cap. 1 pág. 1).

Cabe destacar que el médico encargado del tratamiento sabrá cómo y cuándo finalizar, interrumpir o ajustar la administración debido a toxicidad, a disfunción orgánica y similares. Por el contrario, el médico encargado del tratamiento también sabrá cómo ajustar el tratamiento a niveles mayores si la respuesta clínica no fuera la adecuada (excluyendo la toxicidad). La magnitud de una dosis administrada en el tratamiento del trastorno variará con la gravedad de la afección que se va a tratar, con la vía de administración y similares. La gravedad de la afección se puede evaluar, por ejemplo, en parte, mediante procedimientos de evaluación pronóstica convencionales. Además, la dosis y quizá la frecuencia de la dosis, también variarán de acuerdo con la edad, el peso corporal y la respuesta del paciente individual.

En general, en el tratamiento de un trastorno se deberá modificar al menos 1-10 % de las células dirigidas para modificación genómica. Por tanto, el procedimiento y la vía de administración se escogerán de forma óptima para modificar al menos 0,1-1 % de las células diana por administración. De este modo, el número de administraciones se puede mantener al mínimo con el fin de aumentar la eficiencia y la conveniencia del tratamiento.

Dependiendo de las afecciones específicas que se estén tratando, estos agentes se pueden formular y administrar por vía sistémica o local. Técnicas para la formulación y administración se pueden encontrar en "Remington's Pharmaceutical Sciences," 1990, 18ª ed., Mack Publishing Co., Easton, Pa. Las vías adecuadas pueden incluir administración oral, rectal, transdérmica, vaginal, transmucosa o intestinal; administración parenteral, incluidas inyecciones intramusculares, subcutáneas e intramedulares, así como inyecciones intratecales, intraventriculares directas, intravenosas, intraperitoneales, intranasales o intraoculares, por citar algunas.

El sujeto que está siendo tratado recibirá adicionalmente una recombinasa que reconoce de forma específica la primera y la segunda secuencias de recombinación que se seleccionan para usar. La recombinasa concreta se puede administrar incluyendo un ácido nucleico que la codifica como parte de una construcción de ácido nucleico o como proteína para su captación por las células cuyo genoma se va a modificar. Los procedimientos y las vías de administración serán similares a los descritos anteriormente para la administración de una construcción diana que comprende una secuencia de recombinación y una secuencia de ácido nucleico de interés. Es probable que la proteína recombinasa solamente se requiera durante un periodo limitado de tiempo para integrar la secuencia de ácido nucleico de interés. Por tanto, si se introduce como un gen de recombinasa, el vector portador el gen de la recombinasa carecerá de secuencias que participan en la retención prolongada. Por ejemplo, el ADN plasmídico convencional se degrada rápidamente en la mayoría de las células de mamífero. Asimismo, el gen de la recombinasa puede estar provisto de secuencias de expresión génica que limiten su expresión. Por ejemplo, se puede usar un promotor inducible, de modo que la expresión de la recombinasa se puede regular temporalmente

mediante exposición limitada al agente inductor. Uno de estos grupos de ejemplo de promotores es el de los promotores respondedores a la ecdisona, cuya expresión se puede regular usando ecdiesteroides u otros agonistas no esteroideos. Otro grupo de promotores son los promotores respondedores a la tetraciclina, cuya expresión se puede regular usando tetraciclina o doxiciclina.

5

Ejemplos

MÉTODOS GENERALES

10 Las técnicas convencionales de ADN recombinante y de clonación molecular usadas en el presente documento son bien conocidas en la materia y se describen en Sambrook, J., Fritsch, E. F. y Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. (1989) y en T. J. Silhavy, M. L. Bannan y L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) y en Ausubel, F. M. y col., *Current Protocols in Molecular Biology*, Greene Publishing Assoc. y Wiley-Interscience, Nueva York, NY (1987). Los materiales y procedimientos adecuados para el mantenimiento y crecimiento de cultivos bacterianos son bien conocidos en la materia. Las técnicas adecuadas para usar en los ejemplos siguientes se pueden encontrar como se indica en Phillip, G. y col., *Manual of Methods for General Bacteriology*, American Society for Microbiology, Washington, DC. (1994) o en Brock, T.D. *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). Todos los reactivos, enzimas de restricción y materiales usados para el crecimiento y el mantenimiento de las células huésped se obtuvieron en New England Biolabs (Beverly, MA), Invitrogen Corporation (Carlsbad, CA), Stratagene Corporation (La Jolla, CA), Promega Corporation (Madison, WI), DIFCO Laboratories (Detroit, MI), o Sigma/Aldrich Chemical Company (San Luis, MO) a menos que se especifique lo contrario.

25 Las manipulaciones de secuencias genéticas y la alineación y comparación de secuencias de polinucleótidos y peptídicas se pueden realizar usando la suite de programas disponibles en Invitrogen Corporation, Carlsbad, CA (Software Vector NTI versión 8.0), DNASTAR, Inc., Madison, WI (Software DNASTAR versión 6.0) o Genetics Computer Group Inc., Madison, WI (Wisconsin Package Versión 9.0).

30 El significado de las abreviaturas es el siguiente: "h" significa hora(s), "µl," significa microlitro(s), "ml" significa mililitro(s), "l" significa litro(s), "µM" significa micromolar, "mM" significa milimolar, "ng" significa nanogramo(s), "µg" significa microgramo(s), "mg" significa miligramo(s), "A" significa adenina o adenosina, "T" significa timina o timidina, "G" significa guanina o guanosina, "C" significa citidina o citosina, "nt" significa nucleótido(s), "aa" significa aminoácido(s), "pb" significa par(es) de bases, "kb" significa kilobase(s), "k" significa kilo, "µ" significa micro, "φ" significa Phi, "β" significa beta, "SE" significa error estándar, "Luc" significa luciferasa de luciérnaga, "RLuc" significa luciferasa de *Renilla* y "°C" significa grados centígrados.

40 Los ejemplos siguientes demuestran que los sistemas de recombinasa específica de sitio derivados del bacteriófago SPβc2 de *Bacillus subtilis*, del bacteriófago SF370.1 de *Streptococcus pyogenes*, del bacteriófago Bxb1 de *Mycobacterium smegmatis*, del bacteriófago A118 de *Listeria monocytogenes* y del bacteriófago φRv1 de *Mycobacterium tuberculosis* funcionan en células eucariotas. Estos ejemplos se ofrecen para ilustrar, pero no limitar, la presente invención.

EJEMPLO 1: DISEÑO, SÍNTESIS Y CLONACIÓN DE GENES DE RECOMBINASA Y PLÁSMIDOS PARA EL ENSAYO DE RECOMBINACIÓN INTRAMOLECULAR

50 Después de analizar la literatura publicada y las secuencias disponibles en Genbank se seleccionaron numerosas recombinasas específicas de sitio y se analizó la integración, escisión, inversión y sustitución del ADN en células de mamífero y vegetales. Las secuencias de aminoácidos para recombinasas específicas de sitio grandes de la familia de las serinas (Smith, M. C. y H. M. Thorpe 2000 Diversity in the serine recombinases. *Mol. Microbiol.*, 44: 299-307) se obtuvieron de GenBank y se sometieron a transcripción inversa a ADN. Dado que las fuentes de recombinasas procedían de bacterias o virus de bacterias, los autores de la invención optimizaron la secuencia de ADN para la expresión de la recombinasa en células de mamífero sin cambiar la secuencia de aminoácidos codificada. Los genes se sintetizaron completamente usando los codones de expresión de alto nivel en seres humanos y ratones y con sitios convenientes de enzimas de restricción para clonación. Además, en lo posible se han evitado regiones de un contenido muy alto (>80 %) o muy bajo (<30 %) en GC. Además, durante la optimización se evitaron los siguientes motivos de secuencias de acción en cis optimizando la estabilidad y la traducción del ARN:

60 - cajas TATA internas, sitios chi y sitios de entrada al ribosoma

- tramos de secuencias ricas en AT o GC

- secuencias de repetición y estructuras secundarias de ARN

65 - sitios donantes y aceptores de corte y empalme (crípticos), puntos de ramificación

- sitios de poli(A)

- La optimización del codón y del ARN tuvo como resultado una diferencia del 20-30 % de la secuencia entre genes nativos (es decir, la secuencia de ADN disponible en Genbank) y sintéticos. Los genes sintéticos que codifican las recombinasas se clonaron en el plásmido de expresión de mamíferos y de *E. coli* pDual obtenido de Stratagene Corporation (La Jolla, CA, nº de catálogo 214501). El vector de expresión pDual dirige la expresión de genes heterólogos en células de mamífero y procariontas. Para la expresión constitutiva en células de mamífero, el vector contiene el promotor/potenciador del gen temprano inmediato del citomegalovirus (CMV) humano. El gen de la recombinasa se clona en el único sitio de enzima de restricción Eam 1104 1 presente entre el promotor del CMV y la secuencia de terminación del SV40. Durante la síntesis de las secuencias génicas, los autores de la invención añadieron el sitio de reconocimiento de la enzima de restricción Eam 1104 1 al principio (antes del codón de iniciación ATG) y al final (después del codón de terminación TAG) del gen facilitando la digestión con la enzima Eam 1104 1 y clonar en el mismo sitio en el plásmido pDual. La clonación de genes sintéticos, la secuenciación de clones confirmando la secuencia génica después de clonar en el vector pDual se realizó usando los procedimientos de clonación de ADN convencionales (Sambrook, J., E. F. Fritsch y col., 1989. Molecular Cloning: A laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY). La descripción de los plásmidos de expresión se proporciona a continuación.
- 20 1.1 Plásmido de expresión de la recombinasa SPβc2: una secuencia de ADN sintética (SEC ID N°: 1) optimizada por codón para la expresión en células animales y que codifica la ADN recombinasa específica de sitio yokA del fago SPβc2 de *Bacillus subtilis* (SEC ID N°: 2, número de acceso en Genbank #T12765, Lazarevic, V., A. Dusterhoft y col., 1999, Nucleotide sequence of the Bacillus subtilis temperate bacteriophage SPβc2. Microbiology 145: 1055-67) se clonó en el vector de expresión pDual en el sitio de restricción Eam 1104 1 siguiendo los procedimientos recomendados por Stratagene (La Jolla, CA).
- 30 1.2 Plásmido de expresión de la recombinasa SF370.1: una secuencia de ADN sintética (SEC ID N°: 3) optimizada por codón para la expresión en células animales y que codifica la aparente recombinasa del bacteriófago SF370.1 de *Streptococcus pyogenes* (SEC ID N°: 4, número de acceso en Genbank #T12765, Canchaya, C., F. Desiere y col., 2002, Genome organization and characterization of an inducible prophage and prophage remnants integrated in the Streptococcus pyogenes strain SF370. Virology 302: 245-58) se clonó en el vector de expresión pDual en el sitio de restricción Eam 1104 1 siguiendo los procedimientos recomendados por Stratagene (La Jolla, CA).
- 35 1.3 Plásmido de expresión de la recombinasa Bxb1: una secuencia de ADN sintética (SEC ID N°: 5) optimizada por codón para la expresión en células animales y que codifica la aparente recombinasa del bacteriófago Bxb1 de *Mycobacterium smegmatis* (SEC ID N°: 6, número de acceso en Genbank AAG59740, Mediavilla, J., S. Jain y col., 2000, Genome organization and characterization of mycobacteriophage Bxb1. Microbiology 38: 955-70) se clonó en el vector de expresión pDual en el sitio de restricción Eam 1104 1 siguiendo los procedimientos recomendados por Stratagene (La Jolla, CA).
- 40 1.4 Plásmido de expresión de la recombinasa A118: una secuencia de ADN sintética (SEC ID N°: 7) optimizada por codón para la expresión en células animales y que codifica la aparente recombinasa del bacteriófago A118 de *Listeria monocytogenes* (SEC ID N°: 8, CAB53817, Loessner, M. J., R. B. Inman y col., 2000, Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of Listeria monocytogenes: implications for phage evolution. Mol. Microbiol. 35: 324-40) se clonó en el vector de expresión pDual en el sitio de restricción Eam 1104 1 siguiendo los procedimientos recomendados por Stratagene (La Jolla, CA).
- 45 1.5 Plásmido de expresión de la recombinasa φRv1: una secuencia de ADN sintética (SEC ID N°: 9) optimizada por codón para la expresión en células animales y que codifica la aparente recombinasa del bacteriófago φRv1 de *Mycobacterium tuberculosis* (SEC ID N°: 10, número de acceso en Genbank CAB09083, Bibb, L. A. y G. F. Hatfull 2002, Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. Microbiology 45: 1515-26) se clonó en el vector de expresión pDual en el sitio de restricción Eam 1104 1 siguiendo los procedimientos recomendados por Stratagene (La Jolla, CA).
- 50 1.6 Plásmido de expresión en plantas de la recombinasa A118: una secuencia de ADN sintética (SEC ID N°: 7) optimizada por codón para la expresión en células animales y que codifica la aparente recombinasa de los bacteriófagos A118 de *Listeria monocytogenes* se clonó en el plásmido de expresión en plantas pILTAB358 entre el promotor del virus del mosaico de la vena de la yuca y la secuencia de terminación NOS (Verdaguer, B., A. Kochko y col., 1998, Functional organization of the cassava vein mosaic virus (CsVMV) promoter. Plant Mol. Biol. 37: 1055-67). El ADN del plásmido pILTAB se obtuvo del Donald Danforth Center for Plant Research, San Luis, MO. Las construcciones son similares al plásmido de expresión A118 usado en células animales a excepción de que el promotor del CMV y el terminador del SV40 estaban sustituidos por el promotor del virus del mosaico de la vena de la yuca y el terminador 35S, respectivamente.
- 60 65 Diseño y construcción de plásmidos para ensayo de recombinación intramolecular

Los plásmidos para ensayo de recombinación intramolecular se construyeron usando el plásmido gWiz™ Luc (Gene Therapy Systems, San Diego, CA). Este plásmido confiere resistencia a kanamicina en *E. coli* y expresa de forma constitutiva un gen de la luciferasa a partir del promotor del CMV cuando se introduce en células de mamífero. El vector también contiene sitios de restricción Sal I y Not I únicos entre el promotor del CMV y el codón de iniciación del gen de la luciferasa. Se crearon sitios de reconocimiento para las enzimas de restricción Apa I y Nhe I insertando un oligonucleótido entre los sitios Sal I y Not I. Los oligonucleótidos que contienen el sitio *attP* de la recombinasa y que tienen sitios de restricción flanqueantes Sal I y Apa I se sintetizaron, hibridaron e insertaron entre los sitios Sal I y Apa I. De un modo similar, los oligonucleótidos que contienen la secuencia *attB* se insertaron entre los sitios Nhe I y Not I. Una secuencia de terminación o FIN de la transcripción de 1.296 pb se amplificó mediante PCR a partir del plásmido pBS302 (número de acceso en Genbank U51223, nucleótidos 193-1488) y se clonó en los sitios Apa I y Nhe I entre los sitios *attP* y *attB*. La construcción final tenía las secuencias *attP*, FIN y *attB* colocadas entre el promotor del CMV y el gen de la luciferasa, como se muestra en la figura 1. El plásmido expresaría el gen de la luciferasa únicamente después de la delección de la secuencia FIN debido a la recombinación entre los sitios *attP* y *attB*. La descripción de los plásmidos para ensayo de recombinación intramolecular se proporciona a continuación.

1.7 Plásmido para el ensayo de recombinación intramolecular SPβc2: una secuencia de oligonucleótido sintético de 99 pb que contiene el sitio *attP* de la recombinasa SPβc2 (SEC ID N°: 11), una secuencia de FIN de 1.296 pb (SEC ID N°: 12) y una secuencia oligonucleotídica sintética de 96 pb que contiene un sitio *attB* (SEC ID N°: 13) de la recombinasa SPβc2 se clonaron en ese orden entre el promotor del CMV y el gen de la luciferasa del plásmido gWiz™ Luc.

1.8 Plásmido para el ensayo de recombinación intramolecular SF370.1: una secuencia de oligonucleótido sintético de 99 pb que contiene el sitio *attP* de la recombinasa SF370.1 (SEC ID N°: 14), una secuencia de FIN de 1.296 pb (SEC ID N°: 12) y una secuencia oligonucleotídica sintética de 96 pb que contiene un sitio *attB* (SEC ID N°: 15) de la recombinasa SF370.1 se clonaron en ese orden entre el promotor del CMV y el gen de la luciferasa del plásmido gWiz™ Luc.

1.9 Plásmido para el ensayo de recombinación intramolecular Bxb1: una secuencia de oligonucleótido sintético de 52 pb que contiene el sitio *attP* de la recombinasa Bxb1 (SEC ID N°: 16), una secuencia de FIN de 1.296 pb (SEC ID N°: 12) y una secuencia oligonucleotídica sintética de 46 pb que contiene un sitio *attB* (SEC ID N°: 17) de la recombinasa Bxb1 se clonaron en ese orden entre el promotor del CMV y el gen de la luciferasa del plásmido gWiz™ Luc.

1.10 Plásmido para el ensayo de recombinación intramolecular A118: una secuencia de oligonucleótido sintético de 99 pb que contiene el sitio *attP* de la recombinasa A118 (SEC ID N°: 18), una secuencia de FIN de 1.296 pb (SEC ID N°: 12) y una secuencia oligonucleotídica sintética de 96 pb que contiene un sitio *attB* (SEC ID N°: 19) de la recombinasa A118 se clonaron en ese orden entre el promotor del CMV y el gen de la luciferasa del plásmido gWiz™ Luc.

1.11 Plásmido para el ensayo de recombinación intramolecular φRv1: una secuencia de oligonucleótido sintético de 99 pb que contiene el sitio *attP* de la recombinasa φRv1 (SEC ID N°: 20), una secuencia de FIN de 1.296 pb (SEC ID N°: 12) y una secuencia oligonucleotídica sintética de 96 pb que contiene un sitio *attB* (SEC ID N°: 21) de la recombinasa φRv1 se clonaron en ese orden entre el promotor del CMV y el gen de la luciferasa del plásmido gWiz™ Luc.

1.12 Plásmido vegetal para el ensayo de recombinación intramolecular A118: una secuencia de oligonucleótido sintético de 99 pb que contiene el sitio *attP* de la recombinasa A118 (SEC ID N°: 18), una secuencia de FIN de 1.296 pb (SEC ID N°: 12), una secuencia oligonucleotídica sintética de 96 pb que contiene un sitio *attB* (SEC ID N°: 19) de una recombinasa A118 y el gen de la luciferasa se clonaron en este orden entre el promotor del virus del mosaico de la vena de la yuca y la secuencia terminadora de NOS depILTAB358.

EJEMPLO 2: ENSAYOS DE RECOMBINACIÓN INTRAMOLECULAR TRANSITORIA

Con el fin de determinar la actividad de las recombinasas en células de mamífero y vegetales se desarrolló un ensayo transitorio. Brevemente, el ensayo consistía en la clonación del gen de la recombinasa en un plásmido de expresión, creando el correspondiente plásmido para el ensayo de recombinación intramolecular transitoria, introduciendo los ADN plasmídicos en las células mediante transfección y analizando la actividad de la enzima luciferasa. Los plásmidos del ensayo de recombinasa contenían el promotor del CMV, *attP*:FIN:*attB*-gen indicador de la luciferasa-secuencias de terminación. La secuencia FIN es una secuencia señal de la terminación de la transcripción. En ausencia de recombinación, la expresión del gen indicador de la luciferasa se evita mediante la secuencia FIN presente entre el promotor y el gen indicador. La recombinación entre los sitios *attP* y *attB* debido a la recombinasa introducida tiene como resultado la delección de la secuencia FIN y la activación del gen indicador. Este ensayo es sensible y sólido porque es un formato de OFF a ON (apagado a encendido) y la cantidad del indicador de luciferasa se puede analizar fácilmente detectando la luz emitida por la luciferasa con un luminómetro. El formato del ensayo se representa gráficamente en la figura 1.

Transfecciones transitorias y ensayos de luciferasa

Las células se mantuvieron a 37 °C y 5 % de CO₂ en medio DMEM suplementado con 10 % de suero bovino fetal y 1 % de penicilina/estreptomina (obtenidas de Invitrogen, Carlsbad, CA) o en otros medios, según se indique. El día de la transfección se sembraron las placas a diferentes densidades en función del tipo de célula usado. Las células se transfectaron con el plásmido para el ensayo de recombinación intramolecular solo o junto con varias cantidades del ADN del plásmido de expresión de recombinasa usando Lipofectamine 2000™ de acuerdo con las instrucciones de los fabricantes (Invitrogen, Carlsbad, CA). El plásmido indicador de la luciferasa de *Renilla* expresado de forma constitutiva (pRL-CMV de Promega, Madison, WI) se co-transfectó (2 ng/pocillo) y se usó como control interno normalizando los datos. Veinticuatro o cuarenta y ocho horas después de la transfección (según la línea celular) se desechó el medio y se lisaron las células con tampón de lisis pasiva (Promega, Madison, WI). Después, se analizaron los extractos usando el kit Dual Luciferase Assay (Promega, Madison, WI) en un lector de placas equipado con inyectores (Dynex Technologies, Chantilly, VA). Los datos mostrados son las proporciones entre las actividades de luciferasa (Luc) y la luciferasa de *Renilla* (RLuc), a menos que se indique lo contrario. Se observaron resultados similares cuando se compararon las actividades Luc (unidades relativas de luz) (datos no mostrados). Dado que el número de duplicados y de experimentos variaba para diferentes construcciones y líneas celulares se usó el error estándar indicando la variación experimental.

2.1 Ensayo de recombinación intramolecular transitoria en células HEK293 humanas

Células (20.000 células por pocillo en una placa de 96 pocillos) se transfectaron con 25 ng de plásmido para ensayo de recombinación intramolecular y 0, 10, 25 o 75 ng del correspondiente plásmido de recombinasa y se incubaron durante 24 horas. Las células se lisaron con 50 µl de tampón de lisis pasiva y se analizaron extractos de 25 µl. Se realizaron de seis a veinte ensayos duplicados y se representaron las proporciones de Luc/RLuc (valores medios) ± SE. Los valores mostrados encima de las barras en la figura 8 son el número de inducciones (proporción entre la actividad luciferasa en presencia del plásmido de recombinasa y la actividad en ausencia de plásmido de recombinasa).

Como se muestra en la figura 2, la transfección del plásmido de ensayo para recombinación intramolecular solo mostró nada o muy poca actividad luciferasa (proporcionada como la proporción Luc/RLuc). La transfección es de cantidades crecientes del plásmido de expresión de la recombinasa A118. Se observaron resultados similares para SF370.1, SPβc2, βRV1 y Bxb1. Estos resultados indicaron claramente que las recombinasas eran funcionales en las células HEK293. Las recombinasas participaron en la recombinación entre sus sitios *attP* y *attB* y delecionaron la secuencia FIN en el plásmido para ensayo de recombinación intramolecular y activaron la expresión del gen de la luciferasa.

2.2 Ensayo de recombinación intramolecular transitoria en células NIH3T3 de ratón

Células (5.000 células por pocillo en una placa de 96 pocillos) se transfectaron con 25 ng de plásmido para ensayo de recombinación intramolecular y 0, 10, 25 o 75 ng del correspondiente plásmido de expresión de la recombinasa y se incubaron durante 24 horas. Las células se lisaron con 50 µl de tampón de lisis pasiva y se analizaron extractos de 25 µl. Se realizaron de dos a catorce ensayos duplicados y se representaron las proporciones de Luc/RLuc (valores medios) ± SE. Los valores mostrados encima de las barras en la figura 3 son el número de inducciones.

La figura 3 muestra los datos obtenidos de la transfección de NIH3T3 con el plásmido para el ensayo de recombinación intramolecular solo o junto con cantidades crecientes (10, 25 o 75 ng) del plásmido de expresión de la recombinasa. La co-transfección del plásmido de recombinasa y el plásmido para ensayo de recombinación intramolecular aumentó la actividad de la luciferasa muchas veces. Por ejemplo, la transfección de las células con 25 ng del plásmido para el ensayo de recombinación intramolecular con Bxb1ny 75 ng del plásmido de expresión de la recombinasa Bxb1 aumentó 66 veces la actividad de la luciferasa en comparación con la transfección con 25 ng de plásmido para el ensayo de recombinación intramolecular de Bxb1 solo. Similar a la de Bxb1, las recombinasas de A118, SF370.1, SPβc2 y φRV1 también aumentaron la actividad luciferasa (figura 3), lo que muestra que estas recombinasas son funcionales en células NIH3T3 de ratón y son eficaces en la recombinación de sus sitios *attP* y *attB*.

2.3 Ensayo de recombinación intramolecular transitoria en células de ovario de hámster chino (CHO)

Células (15.000 células por pocillo en una placa de 96 pocillos) se transfectaron con 25 ng de plásmido para ensayo de recombinación intramolecular y 0, 10, 25 o 75 ng del correspondiente plásmido de expresión de la recombinasa y se incubaron durante 24 horas. Las células se lisaron con 50 µl de tampón de lisis pasiva y se analizaron extractos de 25 µl. Se realizaron de dos a ocho ensayos duplicados y se representaron las proporciones de Luc/RLuc (valores medios) ± SE. Los valores mostrados encima de las barras en la figura 4 son el número de inducciones.

Como se muestra en la figura 4, la transfección del plásmido de ensayo para recombinación intramolecular A118,

SF370.1 o ϕ RV1 no mostró nada o mostró muy poca actividad luciferasa. La cotransfección con cantidades crecientes de los correspondientes plásmidos de expresión de recombinasa A118, SF370.1 o ϕ RV1 aumentó la actividad luciferasa. Estos resultados indicaron claramente que las recombinasas eran funcionales en las células CHO. Las recombinasas participaron en la recombinación entre sus sitios *attP* y *attB* y deleccionaron la secuencia FIN en el plásmido para ensayo de recombinación intramolecular y activaron la expresión del gen de la luciferasa.

2.4 Ensayo de recombinación intramolecular transitoria en células HeLa humanas

Se transfectaron células (15.000 células por pocillo en una placa de 96 pocillos) con 25 ng de plásmido para ensayo de recombinación intramolecular y 0, 10, 25 o 75 ng del correspondiente plásmido de expresión de la recombinasa y se incubaron durante 24 horas. Se realizaron de dos a ocho ensayos duplicados y se representaron las proporciones de Luc/RLuc (valores medios) \pm SE. Los valores mostrados encima de las barras en la figura 5 son el número de inducciones.

Como se muestra en la figura 5, la transfección del plásmido de ensayo para recombinación intramolecular A118, SF370.1 o ϕ RV1 solo mostró nada o muy poca actividad luciferasa. La cotransfección con cantidades crecientes de los correspondientes plásmidos de expresión de recombinasa A118, SF370.1 o ϕ RV1 aumentó la actividad luciferasa. Estos resultados mostraron que las recombinasas eran funcionales en las células HeLa.

2.5 Ensayo de recombinación intramolecular transitoria en células estromales de médula ósea de rata

Células estromales primarias de médula ósea de ratas se sembraron en placa previamente un día antes de la transfección a una densidad de 4.000 células/cm² y se cultivaron en medio que contiene 50 % de medio alfa de medio mínimo esencial (α MEM), 50 % de HamsF12, 10 % de FBS, 15 de Pen/Estrep. (100 U/ml de penicilina G y 100 mg/ml de sulfato de estreptomina). Las células se transfectaron con 25 ng de plásmido para ensayo de recombinación intramolecular y 0, 50, 100 o 200 ng del correspondiente plásmido de recombinasa y se incubaron durante 48 horas. Las células se lisaron con 50 μ l de tampón de lisis pasiva y se analizaron extractos de 25 μ l. Se realizaron ocho ensayos duplicados y se representaron las proporciones de Luc/RLuc (valores medios) \pm SE. Los valores mostrados encima de las barras en la figura 6 son el número de inducciones.

La figura 6 muestra los datos obtenidos de la transfección de células estromales de médula ósea de rata con el plásmido para ensayo de recombinación intramolecular solo o junto con cantidades crecientes (50, 100 o 200 ng) del correspondiente plásmido de expresión de la recombinasa. La co-transfección del plásmido para el ensayo de recombinación intramolecular y el plásmido de expresión de recombinasa aumentó la actividad de la luciferasa muchas veces. Por ejemplo, la transfección de las células con 25 ng del plásmido para el ensayo de recombinación intramolecular con Bxb1n y 200 ng del plásmido de expresión de la recombinasa Bxb1 aumentó 501 veces la actividad de la luciferasa en comparación con la transfección con 25 ng de plásmido para el ensayo de recombinación intramolecular de Bxb1 solo. Similar a la de Bxb1, las recombinasas de A118, SF370.1, SP β c2 y ϕ RV1 también aumentaron la actividad luciferasa (figura 6), lo que muestra que estas recombinasas son funcionales en células estromales de médula ósea de rata y son eficaces en la recombinación de sus sitios *attP* y *attB*.

2.6 Ensayo de recombinación intramolecular transitoria en células madre neurales de ratón

Las células madre neurales de ratón C 17.2 (mNSC) se obtuvieron del Dr. Evan Snyder del Burnham Research Institute, La Jolla, CA y se mantuvieron usando el protocolo recomendado (Ryder, E.F., E. Y. Snyder y col., 1990. Establishment and characterization of multipotent neural cell lines using retrovirus vector-mediated oncogene transfer. J. Neurobiol., 21: 356-75). Las células se dividieron un día antes de la transfección y se sembraron en placas de 48 pocillos a una densidad de 120.000 células por pocillo. Después de incubar durante la noche, el medio de cultivo se sustituyó con medio sin suero. Las células se transfectaron con 50 ng del plásmido para el ensayo de recombinación intramolecular solo o junto con 0, 25, 50, 100, o 200 ng del ADN del plásmido de expresión de recombinasa usando el reactivo de transfección Lipofectamine 2000™ de acuerdo con las instrucciones de los fabricantes (Invitrogen, Carlsbad, CA). El plásmido indicador de la luciferasa de *Renilla* expresado de forma constitutiva (pRL-CMV de Promega, Madison, WI) se co-transfectó (4 ng/pocillo) como control interno normalizando los datos. Dos días después de la transfección, se desechó el medio y se lisaron las células con 75 μ l tampón de lisis pasiva (Promega, Madison, WI). Se analizaron los extractos (50 μ l) para detectar las actividades luciferasa y luciferasa de *Renilla* usando el kit Dual Luciferase Assay (Promega, Madison, WI) en un lector de placas equipado con inyectores (Dynex Technologies, Chantilly, VA). Los datos mostrados en la figura 7 son las proporciones entre las actividades de luciferasa (Luc) y de luciferasa de *Renilla* (RLuc) y es la media de 4 transfecciones por tratamiento. Las barras de error representan el error estándar.

De forma similar a los resultados observados en las células HEK293, NIH3T3, CHO, HeLa y estromales de médula ósea de rata, las recombinasas A118, SF370.1, SP β c2, ϕ RV1 y Bxb1 eran funcionales en mNSC y aumentaron la actividad luciferasa (figura 7). La co-transfección de cantidades crecientes (25, 50, 100, o 200 ng) del plásmido de expresión de recombinasa con el correspondiente plásmido para el ensayo de recombinación intramolecular (50 ng) tuvo como resultado mayores actividades de luciferasa y el número de inducciones varió de 72 a 5.349.

2.7 Ensayo de recombinación intramolecular transitoria en células BY2 de tabaco

5 Cultivos de suspensiones celulares de BY2 de *Nicotiana tabacum* se mantuvieron en medio MS en oscuridad y se subcultivaron semanalmente (Nagata, T., T. Nemoto y S. Hasezawa.1992. Tobacco BY-2 cell line as the Hela cell in the cell biology of higher plants. Intl. Rev. Cytol., 132: 1-30). Los protoplastos preparados a partir de cultivos de 3 días se resuspendieron en manitol 0, M y se distribuyeron en placas petri de 35 mm en alícuotas de 1 ml ($\sim 5 \times 10^5$ células) Los protoplastos se mezclaron con ADN plasmídico y se sometieron a electroporación a 0,56 K Volts durante 80 μ segundos usando un sistema de electroporación de ondas cuadradas con el electrodo Petripulser (BTX, San Diego, CA, EE.UU.). Las células se transfectaron con 10 μ g del plásmido de ensayo de recombinación intramolecular y 0 o 10 μ g para el plásmido de expresión de recombinasa. Tras la electroporación, los protoplastos se diluyeron con 1 ml de 2 x medio de cultivo de protoplastos (Watanabe Y., T. Meshi y Y. Okada. 1987. Infection of tobacco protoplasts with in vitro transcribed tobacco mosaic virus RNA using an improved electroporation method. Virology, 192: 264-272), se alicuotaron como dos cultivos de 1 ml y se incubaron a 27 °C durante 17 h. Los protoplastos se lisaron mediante descongelación y adición de 250 μ l, 5 x tampón de lisis pasiva (Promega, Madison, WI, EE.UU.). Veinte μ l del extracto celular se analizaron determinando la actividad luciferasa usando el kit Dual Luciferase Assay en un lector de placas equipado con inyectores. Los datos mostrados en la figura 8 son las unidades relativas de luz debida a la actividad de luciferasa. Los valores mostraron una meda de 22 duplicados y las barras de error son el error estándar.

20 Como se muestra en la figura 8, la transfección de células BY2 con el plásmido vegetal del ensayo para recombinación intramolecular solo mostraron muy poca actividad luciferasa. La co-transfección con el plásmido de expresión en plantas de la recombinasa A118 tuvo como resultado un incremento de 364 veces la actividad de la luciferasa. Los datos indicaron claramente que la recombinasa recombinaba los sitios *attP* y *attB* en células vegetales.

EJEMPLO 3: INTEGRACIÓN ESTABLE DEL ADN PLASMÍDICO QUE CONTIENE LA SECUENCIA ATTP O ATTB EN CROMOSOMA DE CÉLULAS HEK293 QUE CONTIENE EL SITIO ATTB O ATTP

30 El ensayo para la integración del ADN plasmídico en el sitio *attP* o *attB* en el cromosoma se realizó en un procedimiento de dos etapas. En la primera etapa se generó una línea celular estable que contiene una única copia del sitio *attP* o *attB* de cada enzima y se caracterizó. En la segunda etapa, un plásmido que contiene el sitio *attP* o *attB* se integró en los sitios *attB* o *attP* del cromosoma, respectivamente, en presencia del plásmido de expresión de la recombinasa.

Generación de clones de HEK293 estables con una secuencia attP o attB en el cromosoma

40 Una única copia de la secuencia *attP* o *attB* de cada recombinasa (SEC ID N°: 11, 13-21) se introdujo en el locus de FRT en células Flp-In™-293 obtenidas de Invitrogen [Carlsbad, CA (n° de catálogo R750-07)] siguiendo el procedimiento recomendado por el fabricante. El locus FRT en células Flp-In™-293 tiene un promotor de CMV, sitio de integración de FRT para la recombinasa Flp y resistencia al gen de fusión zeocina y β -galactosidasa. Estas células crecen en presencia del antibiótico zeocina y expresan el gen marcador de la β -galactosidasa. La secuencia *attP* o *attB* de cada enzima se clonó en el plásmido pcDNA/FRT (Invitrogen, Carlsbad, CA, n° de catálogo V6010-20) en la región de múltiples sitios de clonación presentes entre el promotor del CMV y la secuencia de terminación BGH. El plásmido de clonación pcDNA/FRT tiene un sitio FRT precedente al gen de la higromicina. El gen de la higromicina carece de un promotor y un codón de iniciación ATG. Por tanto, la transfección del plásmido pcDNA/FRT que contiene el sitio *attP* o *attB* en células de mamífero no conferirá resistencia a higromicina. La integración del plásmido pcDNA/FRT se produce en el locus FRT en células Flp-In™-293 únicamente tras la co-transfección con el plásmido de expresión de la recombinasa Flp (pCG44, Invitrogen, Carlsbad, CA). La integración tiene como resultado la obtención de resistencia a higromicina y pérdida de resistencia a zeocina y expresión de β -galactosidasa. El procedimiento se muestra esquemáticamente en la figura 9.

55 Los ADN plasmídicos de pcDNA/FRT que contienen *attP* o *attB* se integraron en células Flp-In™-293 y las líneas clonales para cada sitio *attP* o *attB* se seleccionaron en medio que contenía higromicina. Como cabía esperar, las células perdieron la actividad de β -galactosidasa y eran sensibles a zeocina. La presencia del plásmido pcDNA/FRT con una secuencia *attP* o *attB* en el locus FRT también se confirmó mediante PCR (figura 10). En el análisis de PCR, los autores de la invención detectaron integración de la secuencia *attP* o *attB* en el locus FRT en el genoma usando un cebador que se une a *attP* o *attB* y otro cebador que se une a la secuencia del locus de FRT adyacente. Por tanto, el clon sería positivo para PCR únicamente si el sitio *attP* o *attB* está integrado en el cromosoma. Como cabe esperar, las líneas seleccionadas son positivas para *attP* o *attB*. La PCR no amplificó una banda específica del ADN genómico aislado de las células Flp-In™-293 parentales (calles P, Panel C en la figura 10) sino que amplificó una banda del ADN aislado de células íntegro con el plásmido pcDNA/FRT que contenía *attP* o *attB* (calles I, panel C en la figura 10) para cada recombinasa analizada. Las células 293 estables con sitios *attP* o *attB* se usaron integrando el plásmido que contiene los sitios *attB* o *attP*, respectivamente.

65

Integración del ADN plasmídico en el sitio attP o attB del cromosoma

Los plásmidos del ensayo de integración se construyeron colocando la secuencia attP o attB de cada recombinasa inmediatamente antes del gen de la resistencia a puomicina. En este plásmido, el gen de la puomicina no tiene su propio promotor. No obstante, la recombinación entre el attP en el cromosoma y el attB en el plásmido del ensayo de integración (o attB en el cromosoma y attP en el plásmido) integraría el gen de la puomicina al lado del promotor del CMV presente inmediatamente antes del sitio attP o attB en las células Flp-In™-293 generadas antes (figura 9). La integración dará lugar a la expresión del gen de puomicina y el crecimiento de dichas células en presencia del antibiótico puomicina. No se espera que la integración aleatoria del plásmido del ensayo proporcione resistencia a puomicina. La línea celular estable Flp-In™-293 que contiene la secuencia attP se transfectó con el plásmido del ensayo de integración que contiene el sitio attB y con o sin el correspondiente plásmido de expresión de la recombinasa usando los protocolos convencionales. En otro caso, se generó la línea celular estable Flp-In™-293 con la secuencia attB integrada de forma estable y se usó integrando el plásmido del ensayo de integración que contiene attP. Las células Flp-In™-293 que contienen un sitio cromosómico attP o attB (150.000 a 30.000 células) se transfectaron con 100 ng del plásmido del ensayo de integración y 400 ng del plásmido de expresión de la recombinasa. Después, se lisaron las células en medio que contiene el antibiótico puomicina. Si la recombinasa es funcional, cabe esperar que la secuencia attB que contiene el plásmido se integre en el sitio attP en el cromosoma y al contrario.

El número de colonias resistentes a puomicina obtenidas de las células Flp-In™-293 que contienen el sitio attB o attP después de la co-transfección con el plásmido del ensayo de integración que contiene attP o attB y el correspondiente plásmido de expresión de la recombinasa en 3 experimentos independientes se muestra en las tablas 1 y 2 siguientes. En ausencia del plásmido de la recombinasa no se observaron colonias resistentes a puomicina. Estos resultados mostraron claramente que las recombinasas facilitaban la recombinación entre el sitio attP o attB del cromosoma y el sitio attP o attB del plásmido, lo que tiene como resultado la integración del ADN del plásmido en el cromosoma. Los autores de la invención también han confirmado la integración del plásmido mediante el aislamiento del ADN genómico de los clones resistentes a puomicina y detectaron la presencia de sitios attL y attR en el cromosoma. La recombinación entre attB y attP da lugar a la creación de los sitios attL y attR, que son sitios híbridos entre attB y attP. La amplificación por PCR usando los cebadores específicos de attL o attR amplificaron la banda específica prevista únicamente en los clones resistentes a puomicina tras la integración del plásmido del ensayo (calles I, paneles A y B en la figura 10) pero no en las células parentales que contienen attP o attB que se usaron para la integración (calles P, paneles A y B en la figura 10).

Tabla 1. Integración del plásmido que contiene attP en el cromosoma con un sitio attB.

Recombinasa	Sitio en el cromosoma	Sitio en el plásmido del ensayo	Número de clones puomicina ^R		
			Exp. Nº 1	Exp. Nº 2	Exp. Nº 3
A118	attB	attP	28	12	0
SF370.1	attB	attP	No realizado	48	148
SPβc2	attB	attP	77	303	270
φRv1	attB	attP	4	9	0
Bxb1	attB	attP	4	3	12

Tabla 2: integración del plásmido que contiene attB en el cromosoma con un sitio attP.

Recombinasa	Sitio en el cromosoma	Sitio en el plásmido del ensayo	Número de clones puomicina ^R		
			Exp. Nº 1	Exp. Nº 2	Exp. Nº 3
A118	attP	attB	34	55	26
SF370.1	attP	attB	0	2	2
SPβc2	attP	attB	268	293	445
Bxb1	attP	attB	12	8	No realizado

EJEMPLO 4: DELECIÓN DEL ADN CROMOSÓMICO FLANQUEADO POR LOS SITIOS ATTP Y ATTB

El ensayo para la delección de la secuencia attP:FIN:attB localizada en el cromosoma se realizó en un procedimiento de dos etapas. En la primera etapa se generaron líneas celulares estables que contienen una única copia de la construcción promotor CMV-attP:FIN:attB-gen de la luciferasa-Terminador para cada recombinasa y se caracterizaron. En la segunda etapa, el plásmido de expresión de la recombinasa se transfectó de forma transitoria en células estables con promotor CMV-attP:FIN:attB-gen de la luciferasa-Terminador y las células se analizaron determinando la actividad luciferasa. Si la recombinasa es activa en células de mamífero, la recombinación entre los

sitios cromosómicos *attP* y *attB* dará lugar a la delección de la secuencia FIN y la activación de la expresión de la luciferasa. El formato del ensayo se representa gráficamente en la figura 11.

5 Generación de clones estables de HEK293 estables con la construcción promotor del CMV-*attP*-FIN-*attB*-gen de la luciferasa en el cromosoma

Una única copia de la construcción promotor del CMV-*attP*:FIN:*attB*-gen de la luciferasa-Terminador se introdujo en el locus de FRT de células Flp-In™-293 obtenidas de Invitrogen, Carlsbad, CA (nº de catálogo R750-07) como se ha descrito anteriormente. La secuencia *attP*:STOP:*attB*-gen de la luciferasa de cada recombinasa que estaba presente en plásmidos de ensayo de recombinación intramolecular (véase Diseño y construcción de plásmidos del ensayo de recombinación intramolecular y la figura 1) se clonó en el plásmido pcDNA/FRT (Invitrogen, Carlsbad, CA, nº de catálogo V6010-20) en la región de múltiples sitios de clonación presentes entre el promotor del CMV y la secuencia de terminación BGH. El plásmido pcDNA/FRT construido con el promotor del CMV-*attP*:STOP:*attB*-gen de la luciferasa-Terminador se insertó en el locus de FRT de células Flp-In™-293 usando la Flp recombinasa. La integración de este plásmido tiene como resultado la obtención de resistencia a higromicina y pérdida de resistencia a zeocina y expresión de β-galactosidasa.

Las células Flp-In™-293 se transfectaron con el plásmido pcDNA/FRT que contiene el promotor del CMV-*attP*:STOP:*attB*-gen de la luciferasa-Terminador junto con el plásmido de expresión de Flp (pCG44, Invitrogen, Carlsbad, CA). Se seleccionaron los clones resistentes a higromicina y se expandieron (figura 11). La inserción del plásmido pcDNA/FRT también se confirmó analizando los clones seleccionados determinando la actividad de β-galactosidasa. Los clones seleccionados perdieron la actividad de β-galactosidasa. Los clones aislados se usaron para la transfección con los plásmidos de expresión de recombinasa,

25 Delección de una secuencia FIN del cromosoma y activación de la luciferasa en líneas celulares estables

En la segunda etapa, las células resistentes a higromicina que contienen la construcción promotor del CMV-*attP*:FIN:*attB*-gen de la luciferasa-Terminador para cada recombinasa se transfectaron de forma transitoria con el correspondiente plásmido de expresión de recombinasa. Se transfectaron células (15.000 por pocillo, formato de 96 pocillos) con 0, 25, 50, 100 o 200 ng de plásmidos de expresión de recombinasa y se incubaron durante 24 horas. Las células se lisaron con 50 µl de tampón de lisis pasiva y se analizaron extractos de 25 µl. Se realizaron dieciséis ensayos duplicados y se representó su actividad luciferasa (media de unidades relativas de luz) ± SE.

Como se muestra en la figura 12, la transfección de cantidades crecientes (0, 25, 50, 100, o 200 ng) de cada plásmido de expresión de la recombinasa en su correspondiente clon de células Flp-In™-293 que contienen *attP*:FIN:*attB* aumentaba la actividad luciferasa. Estos resultados mostraron que las recombinasas pueden recombinar las secuencias *attP* y *attB* colocadas en el cromosoma. La recombinación tuvo como resultado la delección de la secuencia flanqueada por los sitios *attP* y *attB* y la activación del gen de la luciferasa.

40 EJEMPLO 5: INTEGRACIÓN DE ADN EN LOS SEUDOSITIOS DE UNIÓN EN EL CROMOSOMA EN LAS CÉLULAS HEK293

El ensayo para la inserción o integración de un plásmido que contiene un sitio de integración *attP* o *attB* en el seudositio *attB* nativo o el seudositio *attP* presente en las células HEK293 se realizó mediante co-transfección de las células con el plásmido de expresión de la recombinasa y el correspondiente plásmido dirigido que contiene el sitio *attP* o *attB* y el gen de resistencia a higromicina y seleccionar las células estables en medio que contiene el antibiótico higromicina. El procedimiento se representa esquemáticamente en la figura 13. Las células HEK293 se mantuvieron a 37 °C y con 5 % de CO₂ en medio DMEM suplementado con 10 % de suero bovino fetal y 1 % de penicilina/estreptomina (obtenidas de Invitrogen, Carlsbad, CA). El día de la transfección, las células se sembraron en placas a una densidad de 750.000 células por placa petri de 35 mm. Las células se transfectaron con 50 ng del plásmido dirigido que contiene el sitio *attP* o *attB* y un gen de resistencia a higromicina dirigido por el promotor de la Ubiquitina C (figura 13) solo o junto con 4 µg del plásmido para la expresión de la recombinasa usando Lipofectamine 2000™ de acuerdo con las instrucciones de los fabricantes (Invitrogen, Carlsbad, CA). La integración cromosómica del plásmido dará lugar a la expresión del gen de higromicina y el crecimiento de dichas células en presencia del antibiótico higromicina. Cabe destacar que la integración aleatoria del plásmido dirigido (es decir, en sitios que no son seudositios) también podría resultar en la generación de clones resistentes a higromicina. No obstante, cuando el plásmido diana se introduce en células junto con el plásmido de expresión de recombinasa, cabe esperar que el número de clones HEK293 resistentes a higromicina sea mayor si el genoma contiene seudositios de unión. Asimismo, por ejemplo, si la integración se debe a la recombinación entre el seudositio *attB* sobre el genoma y el sitio *attP* sobre el plásmido dirigido, el sitio *attP* sobre el plásmido dirigido se corta de un modo preciso y el plásmido se inserta en los seudositios *attB* en el genoma, lo que tiene como resultado la creación del seudositio *attL* y el seudositio *attR* que se pueden identificar mediante secuenciación del ADN de plásmidos rescatados. Por el contrario, las integraciones aleatorias generalmente conservan el sitio *attP* intacto tras la integración.

Los clones de HEK293 resistentes a higromicina obtenidos en presencia del plásmido de expresión de la recombinasa se combinaron, se realizó una preparación del ADN genómico y se digirió con las enzimas de restricción que cortaron el plásmido integrado (es decir, fuera de la región de pUC ori y el gen del marcador seleccionable bacteriano), el ADN digerido se auto-ligó y el ADN ligado se transformó en *E. coli* rescatando el plásmido integrado que contiene el ADN genómico adyacente, siguiendo los procedimientos comunes en este campo (Thyagarajan, B. y col., (2001) Site-specific genomic integration in mammalian cells mediated by phage ϕ C31 integrase. Mol. Cell. Biol. 21: 3926-3934). El ADN genómico preparado para los clones resistentes a higromicina (10 μ g) se digirió con las enzimas de restricción Bgl II, Xba I, Eco 0109I, Ban II, Sty I, Bso BI, o Btg I en un volumen total de 40 μ l durante 3 horas a 37 °C. 20 μ l de cada digestión se ligaron en un volumen total de 200 μ l durante la noche a 4 °C y después se purificaron. El ADN ligado se introdujo en *E. coli* mediante electroporación y después se seleccionaron las colonias de *E. coli* resistentes a ampicilina en una placa que contiene el antibiótico. Los ADN plasmídicos se prepararon a partir de colonias bacterianas y después se secuenciaron los ADN plasmídicos rescatados. La secuencia del ADN genómico recuperado se usó identificando la localización en el cromosoma alineando la secuencia genómica recuperada con la secuencia del genoma humano en el documento Genbank, NIH Library of Medicine usando el programa BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>).

Cuando el seudo sitio del plásmido dirigido que contiene el sitio *attP* de la recombinasa SF370.1 o SP β c2 se introdujo en células HEK293 se obtuvieron 9 y 0 clones resistentes a higromicina, respectivamente (tabla 3). Por el contrario, cuando el ADN plasmídico dirigido se co-introdujo en las células HEK293 junto con los respectivos plásmidos de expresión de la recombinasa SF370.1 o SP β c2 se recuperaron más de 100 clones resistentes a higromicina en cada caso (tabla 3). Estos resultados indican claramente que la integración mediada por la recombinasa en los seudositios *attB* en el cromosoma era altamente eficiente y la integración en los seudositios era mucho mayor que la integración aleatoria del plásmido dirigido (es decir, la integración en ausencia de recombinasa). El ADN genómico se aisló de clones de células HEK293 combinados resistentes a higromicina obtenidos con la recombinasa SF370.1, los plásmidos se rescataron del genoma y se identificaron las seudosecuencias de *attB* mediante secuenciación de 100 ADN plasmídicos como se ha descrito en lo que antecede. De los 100 plásmidos rescatados secuenciados, 41 eran seudositios *attB* diferentes, ya que había más integraciones en algún seudositio que en otros seudositios. Por ejemplo, 35 de 100 integraciones recuperadas era en un solo sitio. La secuencia de nucleótidos de este seudositio *attB* se proporciona en la figura 14. Estos resultados sugieren que la recombinasa SF370.1 integraba preferentemente el ADN plasmídico en este sitio en comparación con otros sitios.

Tabla 3: integración del plásmido que contiene *attP* en seudositios *attB* del cromosoma de células HEK293

Recombinasa	Seudositio en el cromosoma	Sitio en el plásmido dirigido	Número de clones higromicina ^R	
			Sin recombinasa	Con recombinasa
SF370.1	<i>attB</i>	<i>attP</i>	9	>100
SP β c2	<i>attB</i>	<i>attP</i>	0	>100

Se realizó un análisis similar con clones de HEK293 resistentes a higromicina obtenidos después de dirigir el plásmido que contiene *attP* de SP β c2 usando la recombinasa SP β c2 y se secuenciaron ADN de 109 plásmidos recuperados. El análisis de la secuencia mostró que 105 de 107 integraciones se produjeron en seudositios *attB* y 2 integraciones se encontraban en sitios aleatorios. Había 54 seudositios *attB* diferentes entre los 105 sitios de integración recuperados. Quince de las integraciones se produjeron en una secuencia de un seudositio mostrado en la figura 14. Estos resultados muestran que los cromosomas eucariotas y humanos sirven como dianas eficientes para una integración específica de sitio y precisa en los seudositios *att* usando las enzimas descubiertas. Estos sitios forman dianas naturales para la integración que se pueden usar en muchas aplicaciones de biotecnología y médicas.

Listado de secuencias

- <110> RheoGene, Inc.
Padidam, Malla
- <120> Serina recombinasas específicas de sitio y procedimientos de su uso
- <130> A01505-PCT
- <140> No asignado todavía
- <141> 8-2-2.005
- <160> 21
- <170> PatentIn versión 3.3

ES 2 422 424 T3

<210> 1
 <211> 1638
 <212> ADN
 <213> Artificial

5 <220>
 <223> gen sintético

<400> 1

10 atggaactga agaacatcgt gaacagctac aacatcacca acatcctggg ctacctgcgg 60
 agaagcaggc aggacatgga gagagagaag cggaccggcg aggacaccct caccgagcag 120
 aaggaactca tgaacaagat cctcaccgcc atcgagatcc cctacgagct gaagatggag 180
 atcggcagcg gcgagagcat cgacggcaga cccgtgttca aggagtgcct gaaggatctg 240
 gaggagggca agtaccaggc catcgccgtg aaggagatca ccaggctgag cagaggcagc 300
 tacagcgacg ccggccagat cgtgaacctg ctgcagagca agcggctcat catcatcacc 360
 ccctacaagg tgtacgaccc cagaaacccc gtcgacatgc ggcagatccg gttcgagctg 420
 ttcattggcca gggaggagt cggatgacc cgggagagaa tgaccggcgc caagtacacc 480
 tacgccgccc agggcaagtg gatcagcggc ctggccccct acggctacca gctgaacaag 540
 aaaaccagca agctggacc cgtggaggac gaggccaagg tggcgcagct catcttcaac 600
 atcttctga acgggctgaa cggcaaggac tacagctaca cagccatcgc cagccacctc 660
 accaatctgc agatccctac ccccagcggc aagaagcggg ggaaccagta caccatcaag 720
 gccatcctgc agaacgaggt gtacatcggc accgtgaagt acaaggtgcg ggagaaaacc 780
 aaggacggca agcggacat caggcctgag aaggagcaga tcgtggtgca ggacgcccac 840
 gccctatca tcgacaagga gcagttccag cagagccagg tgaagatcgc caacaaggtg 900
 cccctgctgc ccaacaagga cgagtccgag ctgagcgagc tggccggagt gtgcacctgc 960
 agcaagtgcg gcgagcctct gagcaagtac gagagcaagc gcatccgaa gaacaaggat 1020
 ggcaccgaga gcgtgtacca cgtgaagtcc ctcacctgca agaagaacaa gtgcacctac 1080
 gtgcggtaca acgacgtgga gaacgccatc ctggattacc tgagcagcct gaacgacctg 1140
 aatgacagca ccctcaciaa gcacatcaac agcatgctct ccaagtacga ggacgacaac 1200
 agcaacatga aaaccaagaa gcagatgagc gagcacctga gccagaagga gaaggagctt 1260
 aagaataagg agaacttcat cttcgacaag tacgagtccg gcatctactc cgacgagctg 1320

ttcctgaagc ggaaggccgc cctggacgag gagttcaagg agctgcagaa cgccaagaac 1380
 gagctgaatg gcctgcagga taccagagc gagatcgaca gcaacaccgt gcggaacaac 1440
 atcaacaaga tcatcgacca gtaccacatc gagagcagca gcgagaagaa gaatgagctg 1500
 ctgaggatgg tgctgaagga cgtgatcgtg aacatgacct agaagcga gggccccatc 1560
 cccgcccagt tcgagatcac acccatcctg cggttcaact ttatcttcga tctcaccgcc 1620
 accaacagct tccactag 1638

15 <210> 2
 <211> 545
 <212> PRT
 <213> Fago SPBc2 de Bacillus subtilis

20 <400> 2

ES 2 422 424 T3

Met Glu Leu Lys Asn Ile Val Asn Ser Tyr Asn Ile Thr Asn Ile Leu
1 5 10 15

Gly Tyr Leu Arg Arg Ser Arg Gln Asp Met Glu Arg Glu Lys Arg Thr
20 25 30

Gly Glu Asp Thr Leu Thr Glu Gln Lys Glu Leu Met Asn Lys Ile Leu
35 40 45

Thr Ala Ile Glu Ile Pro Tyr Glu Leu Lys Met Glu Ile Gly Ser Gly
50 55 60

Glu Ser Ile Asp Gly Arg Pro Val Phe Lys Glu Cys Leu Lys Asp Leu
65 70 75 80

Glu Glu Gly Lys Tyr Gln Ala Ile Ala Val Lys Glu Ile Thr Arg Leu
85 90 95

Ser Arg Gly Ser Tyr Ser Asp Ala Gly Gln Ile Val Asn Leu Leu Gln
100 105 110

Ser Lys Arg Leu Ile Ile Ile Thr Pro Tyr Lys Val Tyr Asp Pro Arg
115 120 125

Asn Pro Val Asp Met Arg Gln Ile Arg Phe Glu Leu Phe Met Ala Arg
130 135 140

Glu Glu Phe Glu Met Thr Arg Glu Arg Met Thr Gly Ala Lys Tyr Thr
145 150 155 160

Tyr Ala Ala Gln Gly Lys Trp Ile Ser Gly Leu Ala Pro Tyr Gly Tyr
165 170 175

Gln Leu Asn Lys Lys Thr Ser Lys Leu Asp Pro Val Glu Asp Glu Ala
180 185 190

Lys Val Val Gln Leu Ile Phe Asn Ile Phe Leu Asn Gly Leu Asn Gly

ES 2 422 424 T3

195					200					205					
Lys	Asp	Tyr	Ser	Tyr	Thr	Ala	Ile	Ala	Ser	His	Leu	Thr	Asn	Leu	Gln
	210					215					220				
Ile	Pro	Thr	Pro	Ser	Gly	Lys	Lys	Arg	Trp	Asn	Gln	Tyr	Thr	Ile	Lys
225					230					235					240
Ala	Ile	Leu	Gln	Asn	Glu	Val	Tyr	Ile	Gly	Thr	Val	Lys	Tyr	Lys	Val
				245					250					255	
Arg	Glu	Lys	Thr	Lys	Asp	Gly	Lys	Arg	Thr	Ile	Arg	Pro	Glu	Lys	Glu
			260					265					270		
Gln	Ile	Val	Val	Gln	Asp	Ala	His	Ala	Pro	Ile	Ile	Asp	Lys	Glu	Gln
		275					280					285			
Phe	Gln	Gln	Ser	Gln	Val	Lys	Ile	Ala	Asn	Lys	Val	Pro	Leu	Leu	Pro
	290					295					300				
Asn	Lys	Asp	Glu	Phe	Glu	Leu	Ser	Glu	Leu	Ala	Gly	Val	Cys	Thr	Cys
305					310					315					320
Ser	Lys	Cys	Gly	Glu	Pro	Leu	Ser	Lys	Tyr	Glu	Ser	Lys	Arg	Ile	Arg
				325					330					335	
Lys	Asn	Lys	Asp	Gly	Thr	Glu	Ser	Val	Tyr	His	Val	Lys	Ser	Leu	Thr
			340					345					350		
Cys	Lys	Lys	Asn	Lys	Cys	Thr	Tyr	Val	Arg	Tyr	Asn	Asp	Val	Glu	Asn
		355					360					365			
Ala	Ile	Leu	Asp	Tyr	Leu	Ser	Ser	Leu	Asn	Asp	Leu	Asn	Asp	Ser	Thr
	370					375					380				
Leu	Thr	Lys	His	Ile	Asn	Ser	Met	Leu	Ser	Lys	Tyr	Glu	Asp	Asp	Asn
385					390					395					400
Ser	Asn	Met	Lys	Thr	Lys	Lys	Gln	Met	Ser	Glu	His	Leu	Ser	Gln	Lys
				405					410					415	
Glu	Lys	Glu	Leu	Lys	Asn	Lys	Glu	Asn	Phe	Ile	Phe	Asp	Lys	Tyr	Glu
			420					425					430		
Ser	Gly	Ile	Tyr	Ser	Asp	Glu	Leu	Phe	Leu	Lys	Arg	Lys	Ala	Ala	Leu
		435					440					445			
Asp	Glu	Glu	Phe	Lys	Glu	Leu	Gln	Asn	Ala	Lys	Asn	Glu	Leu	Asn	Gly
	450					455					460				
Leu	Gln	Asp	Thr	Gln	Ser	Glu	Ile	Asp	Ser	Asn	Thr	Val	Arg	Asn	Asn
465					470					475					480

ES 2 422 424 T3

Ile Asn Lys Ile Ile Asp Gln Tyr His Ile Glu Ser Ser Ser Glu Lys
 485 490 495

Lys Asn Glu Leu Leu Arg Met Val Leu Lys Asp Val Ile Val Asn Met
 500 505 510

Thr Gln Lys Arg Lys Gly Pro Ile Pro Ala Gln Phe Glu Ile Thr Pro
 515 520 525

Ile Leu Arg Phe Asn Phe Ile Phe Asp Leu Thr Ala Thr Asn Ser Phe
 530 535 540

His
 545

<210> 3
 <211> 1416
 5 <212> ADN
 <213> Artificial

<220>
 <223> Gen sintético

10 <400> 3

atgcggaagg tggccatcta cagcaggggtg agcaccatca atcaggccga ggagggctac 60
 agcatccagg gccagatcga ggcctcacc aagtactgcg aggccatgga gtggaagatc 120
 tacaagaact acagcgacgc cggcttcagc ggcggcaagc tggagagacc cgccatcacc 180
 gagttgatcg aggacggcaa gaacaacaag ttcgacacca tcctggtgta caagctggac 240
 cggctgagca gaaacgtgaa ggacaccctg tacctggtga aggacgtggt caccgccaac 300
 aacatccact tcgtgagcct gaaggagaac atcgacacca gcagcgccat gggcaatctg 360
 ttcctcacac tgctgagcgc aattgccgag ttcgagcggg agcagatcaa ggaacggatg 420
 cagttcggcg tgatgaacag agccaagagc ggcaagacca ccgcctggaa aaccctcca 480
 tacggctacc ggtacaaca ggacgagaaa accctgagcg tgaacgagct ggaggccgcc 540
 aatgtgaggc agatgttcga catgatcatc agcggctgca gcatcatgag catcaccaac 600
 tacgcccggg acaacttcgt gggcaacacc tggaccacg tgaaggtgaa gcggatcctg 660
 gagaacgaga cctacaaagg cctggtgaag taccgggagc agaccttag cggcgatcac 720
 caggccatca tcgacgaaa gacctacaac aaggcccaga tcgccctggc ccacagaacc 780
 gacaccaaga ccaacaccag acccttcag ggcaagtaca tgctgagcca catcgccaag 840
 tgcggtact gtggcgcccc tctgaaggtg tgcaccggca gggccaagaa tgacggcacc 900
 cggagacaga cctacgtgtg cgtgaacaag accgagagcc tggccagaag gagcgtgaac 960
 aactacaaca accagaagat ctgcaacacc ggccggtacg agaagaagca catcgagaag 1020
 tacgtgatcg acgtgctgta taagctgag caccagaagg agtacctgaa gaagatcaag 1080
 aaggacgaca acatcatcga tatcaccccc ctgaagaagg agatcgagat catcgacaag 1140
 aagattaacc ggctgaacga cctgtacatc aacgacctca tcgacctgcc caagctgaag 1200
 aaagacatcg aggagctgaa ccacctgaag gacgactaca ataaggccat caagctgaac 1260
 tacctggaca agaagaacga ggacagcctg ggcatgctca tggacaacct ggacatccgc 1320
 aagagcagct acgacgtgca gagccggatc gtgaagcagc tcatcgacag ggtggaggtg 1380
 15 accatggaca atatcgacat catcttcaag ttctag 1416

ES 2 422 424 T3

<210> 4

<211> 471

<212> PRT

5 <213> Aparente recombinasa del bacteriófago SF370.1

<400> 4

```

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

```

10

ES 2 422 424 T3

Asn Thr Trp Thr His Val Lys Val Lys Arg Ile Leu Glu Asn Glu Thr
 210 215 220

Tyr Lys Gly Leu Val Lys Tyr Arg Glu Gln Thr Phe Ser Gly Asp His
 225 230 235 240

Gln Ala Ile Ile Asp Glu Lys Thr Tyr Asn Lys Ala Gln Ile Ala Leu
 245 250 255

Ala His Arg Thr Asp Thr Lys Thr Asn Thr Arg Pro Phe Gln Gly Lys
 260 265 270

Tyr Met Leu Ser His Ile Ala Lys Cys Gly Tyr Cys Gly Ala Pro Leu
 275 280 285

Lys Val Cys Thr Gly Arg Ala Lys Asn Asp Gly Thr Arg Arg Gln Thr
 290 295 300

Tyr Val Cys Val Asn Lys Thr Glu Ser Leu Ala Arg Arg Ser Val Asn
 305 310 315 320

Asn Tyr Asn Asn Gln Lys Ile Cys Asn Thr Gly Arg Tyr Glu Lys Lys
 325 330 335

His Ile Glu Lys Tyr Val Ile Asp Val Leu Tyr Lys Leu Gln His Asp
 340 345 350

Lys Glu Tyr Leu Lys Lys Ile Lys Lys Asp Asp Asn Ile Ile Asp Ile
 355 360 365

Thr Pro Leu Lys Lys Glu Ile Glu Ile Ile Asp Lys Lys Ile Asn Arg
 370 375 380

Leu Asn Asp Leu Tyr Ile Asn Asp Leu Ile Asp Leu Pro Lys Leu Lys
 385 390 395 400

Lys Asp Ile Glu Glu Leu Asn His Leu Lys Asp Asp Tyr Asn Lys Ala
 405 410 415

Ile Lys Leu Asn Tyr Leu Asp Lys Lys Asn Glu Asp Ser Leu Gly Met
 420 425 430

Leu Met Asp Asn Leu Asp Ile Arg Lys Ser Ser Tyr Asp Val Gln Ser
 435 440 445

Arg Ile Val Lys Gln Leu Ile Asp Arg Val Glu Val Thr Met Asp Asn
 450 455 460

Ile Asp Ile Ile Phe Lys Phe
 465 470

<210> 5
 <211> 1503
 <212> ADN
 5 <213> Gen sintético

<400> 5

ES 2 422 424 T3

```

atgctgggctc tgggtggtgat caggctgagc agagtgaccg atgccaccac aagccctgag 60
agacagctgg agagctgcca gcagctgtgt gcccagagag gatgggacgt ggtgggagtg 120
gccgaggatc tggatgtgag cggagccgtg gacccttcg acagaaagcg gagacccaac 180
ctggccagat ggctggcctt tgaggagcag cccttcgatg tgatcgtggc ctacagagtg 240
gacaggctga cccggagcat tagacacctc cagcagctgg tgactggggc cgaggaccac 300
aagaaactgg tggtagcgc cacagaggcc cacttcgata ccaccacccc ctttgctgca 360
gtggtgatcg ccctgatggg cacagtggcc cagatggagc tggaggccat caaggagagg 420
aatcggctcg ccgccactt caatcagag gccggcaagt acagaggaag cctgcctcct 480
tggggctacc tgcccacaag agtggatggc gagtggagac tggcctga ccctgtgcag 540
agggagagaa tcctggaagt gtatcacccg gtggtggaca atcacgagcc tctgcacctg 600
gtggcccacg acctgaatag gagaggcgtg ctgtcccca aggattactt cgcccagctc 660
cagggcagag agcctcaggg cagagagtgg tctgccaccg ccctgaaaag atctatgatc 720
agcgaggcca tgctgggcta cgccacctg aatggcaaga ccgtgagggg tgatgatgga 780
gcccctctgg tgagagccga gccatcctg acaagggaa acgctggaggc tctgagagcc 840
gaactggtga aaaccagcag agccaagcct gccgtgagca cacctagcct gctgctgaga 900
gtgctgttct gtgccgtgtg tggcgagcct gcctacaagt ttgccggcgg aggcagaaag 960
caccctcggg acagatgtag gagcatgggc ttccctaagc actgaggcaa tggcaccgtg 1020
gccatggccg aatgggacgc cttttgcgag gagcaagtgc tggatctgct gggagatgcc 1080
gagaggctgg agaaagtgtg ggtggccgga tccgattctg ccgtggaact ggccgaagtg 1140
aatgctgaac tggtagacct gaccagcctg atcggcagcc ctgcctatag agccggaagc 1200
cctcagagag aagccctgga cgccagaatt gccgacctg ccgagagaca ggaggaactg 1260
gagggactgg aggccagacc ttctggctgg gagtggagag agaccggcca gagattcggc 1320
gattggtgga gggagcagga taccgccgcc aagaacacct ggctgaggag catgaacgtg 1380
aggctgacct tcgacgtgag aggcggcctg accagaacca tcgacttcgg cgacctccag 1440
gagtatgagc agcacctgag actgggaagc gtggtggaga gactgcacac aggcattgtcc 1500
tag 1503

```

<210> 6

<211> 500

5 <212> PRT

<213> Aparente recombinasa del micobacteriófago Bxb1

<400> 6

10 Met Arg Ala Leu Val Val Ile Arg Leu Ser Arg Val Thr Asp Ala Thr
1 5 10 15

ES 2 422 424 T3

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

ES 2 422 424 T3

Lys Pro Ala Val Ser Thr Pro Ser Leu Leu Leu Arg Val Leu Phe Cys
 290 295 300

Ala Val Cys Gly Glu Pro Ala Tyr Lys Phe Ala Gly Gly Gly Arg Lys
 305 310 320

His Pro Arg Tyr Arg Cys Arg Ser Met Gly Phe Pro Lys His Cys Gly
 325 330 335

Asn Gly Thr Val Ala Met Ala Glu Trp Asp Ala Phe Cys Glu Glu Gln
 340 345 350

Val Leu Asp Leu Leu Gly Asp Ala Glu Arg Leu Glu Lys Val Trp Val
 355 360 365

Ala Gly Ser Asp Ser Ala Val Glu Leu Ala Glu Val Asn Ala Glu Leu
 370 375 380

Val Asp Leu Thr Ser Leu Ile Gly Ser Pro Ala Tyr Arg Ala Gly Ser
 385 390 395 400

Pro Gln Arg Glu Ala Leu Asp Ala Arg Ile Ala Ala Leu Ala Ala Arg
 405 410 415

Gln Glu Glu Leu Glu Gly Leu Glu Ala Arg Pro Ser Gly Trp Glu Trp
 420 425 430

Arg Glu Thr Gly Gln Arg Phe Gly Asp Trp Trp Arg Glu Gln Asp Thr
 435 440 445

Ala Ala Lys Asn Thr Trp Leu Arg Ser Met Asn Val Arg Leu Thr Phe
 450 455 460

Asp Val Arg Gly Gly Leu Thr Arg Thr Ile Asp Phe Gly Asp Leu Gln
 465 470 475 480

Glu Tyr Glu Gln His Leu Arg Leu Gly Ser Val Val Glu Arg Leu His
 485 490 495

Thr Gly Met Ser
 500

<210> 7
 <211> 1359
 5 <212> ADN
 <213> Gen sintético

<400> 7

atgaaggccg ccatctacat cagagtgagc acccaggagc aggtggagaa ctacagcatc 60
 caggcccaga ccgagaagct caccgccctg tgcagaagca aggactggga cgtgtacgac 120
 atcttcatcg acggcggcta cagcggcagc aacatgaaca gacccgccct gaacgagatg 180
 10 ctgagcaagc tgcacgagat cgatgccgtg gtggtgtaca ggctggacag gctgagcaga 240

ES 2 422 424 T3

agccagaggg acaccatcac cctcatcgag gagtacttcc tgaagaacaa cgtggagttc 300
 gtgagcctga gcgagaccct ggacaccagc agccccttcg gcagagccat gatcggcatc 360
 ctgagcgtgt tcgccagct cgagagagag accatccggg acaggatggt gatgggcaag 420
 atcaagagga tcgaggccgg cctgcccctc accaccgcca agggcagaac cttcggctac 480
 gacgtgatcg acaccaagct gtacatcaac gaggaggagg ccaagcagct gcagctcatc 540
 tacgatatct tcgaggagga gcagagcatc accttcctgc agaagcggct gaagaagctg 600
 ggcttaagg tgcggaccta caaccggtac aacaactggc tcaccaacga cctgtactgc 660
 ggctacgtga gctacaagga caaggtgcac gtgaagggga tccacgagcc catcatcagc 720
 gaggagcagt tctaccgggt gcaggagatc ttcacccgca tgggcaagaa cccaacatg 780
 aaccgggaca gcgccagcct gctgaacaat ctgggtggtg gcagcaagtg cggcctgggc 840
 ttcgtgcaca ggagaaagga caccatgagc cggggcaaga agtaccacta ccggtactac 900
 agctgcaaga cctacaagca caccacgag ctggagaagt gcggcaacaa gatctggagg 960
 gccgacaagc tggaggagtt gatcatcaac cgggtgaaca actacagctt cgccagccgg 1020
 aacgtggata aggaggacga gctggacagc ctgaatgaga agcttaagat cgagcacgcc 1080
 aagaagaagc gcctgttcga cctgtacatt aacggcagct acgaggtgag cgagctggac 1140
 tccatgatga acgacatcga cgcccagatc aactactacg agagccagat cgaggccaac 1200
 gaggagctga agaagaacaa gaagatccag gagaacctgg ccgacctggc caccgtggat 1260
 ttcgacagcc tggagttcag ggagaagcag ctgtacctga agtccctcat caataagatc 1320
 tacatcgacg gggagcaggt gaccatcgag tggtctgtag 1359

<210> 8

<211> 452

5 <212> PRT

<213> Aparente recombinasa del bacteriófago A118

<400> 8

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

10

ES 2 422 424 T3

Asn Val Glu Phe Val Ser Leu Ser Glu Thr Leu Asp Thr Ser Ser Pro
 100 105 110
 Phe Gly Arg Ala Met Ile Gly Ile Leu Ser Val Phe Ala Gln Leu Glu
 115 120 125
 Arg Glu Thr Ile Arg Asp Arg Met Val Met Gly Lys Ile Lys Arg Ile
 130 135 140
 Glu Ala Gly Leu Pro Leu Thr Thr Ala Lys Gly Arg Thr Phe Gly Tyr
 145 150 155 160
 Asp Val Ile Asp Thr Lys Leu Tyr Ile Asn Glu Glu Glu Ala Lys Gln
 165 170 175
 Leu Gln Leu Ile Tyr Asp Ile Phe Glu Glu Glu Gln Ser Ile Thr Phe
 180 185 190
 Leu Gln Lys Arg Leu Lys Lys Leu Gly Phe Lys Val Arg Thr Tyr Asn
 195 200 205
 Arg Tyr Asn Asn Trp Leu Thr Asn Asp Leu Tyr Cys Gly Tyr Val Ser
 210 215 220
 Tyr Lys Asp Lys Val His Val Lys Gly Ile His Glu Pro Ile Ile Ser
 225 230 235 240
 Glu Glu Gln Phe Tyr Arg Val Gln Glu Ile Phe Thr Arg Met Gly Lys
 245 250 255
 Asn Pro Asn Met Asn Arg Asp Ser Ala Ser Leu Leu Asn Asn Leu Val
 260 265 270
 Val Cys Ser Lys Cys Gly Leu Gly Phe Val His Arg Arg Lys Asp Thr
 275 280 285
 Met Ser Arg Gly Lys Lys Tyr His Tyr Arg Tyr Tyr Ser Cys Lys Thr
 290 295 300
 Tyr Lys His Thr His Glu Leu Glu Lys Cys Gly Asn Lys Ile Trp Arg
 305 310 315 320
 Ala Asp Lys Leu Glu Glu Leu Ile Ile Asn Arg Val Asn Asn Tyr Ser
 325 330 335
 Phe Ala Ser Arg Asn Val Asp Lys Glu Asp Glu Leu Asp Ser Leu Asn
 340 345 350
 Glu Lys Leu Lys Ile Glu His Ala Lys Lys Lys Arg Leu Phe Asp Leu
 355 360 365

ES 2 422 424 T3

Tyr Ile Asn Gly Ser Tyr Glu Val Ser Glu Leu Asp Ser Met Met Asn
 370 375 380

Asp Ile Asp Ala Gln Ile Asn Tyr Tyr Glu Ser Gln Ile Glu Ala Asn
 385 390 395 400

Glu Glu Leu Lys Lys Asn Lys Lys Ile Gln Glu Asn Leu Ala Asp Leu
 405 410 415

Ala Thr Val Asp Phe Asp Ser Leu Glu Phe Arg Glu Lys Gln Leu Tyr
 420 425 430

Leu Lys Ser Leu Ile Asn Lys Ile Tyr Ile Asp Gly Glu Gln Val Thr
 435 440 445

Ile Glu Trp Leu
 450

- <210> 9
- <211> 1410
- <212> ADN
- <213> Gen sintético

5

<400> 9

```

atgCGgtaca ccacccccgt gagagccgcc gtgtacctga gaatcagcga ggacagaagc 60
ggcGagcagc tgggcgtggc cagacagaga gaggactgcc tgaagctgtg cggccagaga 120
aagtgggtgc cctggagta cctggacaac gatgtgagcg ccagcaccgg caagaggaga 180
cccgcctacg agcagatgct ggccgacatc accgcccgca agatcgccgc cgtggtggcc 240
tgggacctgg ataggctgca caggagacct atcgagctgg aggccttcat gagcctggcc 300
gatgagaaaa gactggccct ggccaccgtg gccggcgacg tggacctggc cccccccag 360
ggcagactgg tggccagact taagggcagc gtggccgccc acgagaccga gcacaagaag 420
gccagacagc ggagagccgc cagacagaag gccgagagag gccaccccaa ctggagcaag 480
gccttcggct acctgcctgg ccccaacggc cccgagcccg accctagaac cgcccctctg 540
gtgaagcagg cctacgccga catcctggcc ggagccagcc tgggcgacgt gtgcagacag 600
tggaatgacg ccggagcctt caccatcacc ggagaccctt ggaccaccac caccctgagc 660
aagttcctgc ggaagcccag aaacgccggc ctgagagcct acaagggcgc cagatacggc 720
cccgtcgaca gagatgccat cgtgggcaag gccagtgga gccccctggt ggacgaggcc 780
accttctggg ccgctcaggc cgtgctggac gccctggca gagccccagg cagaaagagc 840
gtgCGgagac acctgctcac cggcctggcc ggctgcggca agtgcggcaa ccacctggcc 900
ggcagctaca gaaccgatgg gcaggtggtg tacgtgtgca aggcctgcca cggcgtggcc 960
attctggccg acaacatcga gccatcctg taccatcctg tggccgagag actggccatg 1020
cccGacgccg tggatctgct gaggaggag atccacgacg ccgccgaggc cgagaccatc 1080
agactcgagc tggaaaccct gtacggcgag ctggacagag tggccgtgga gagagccgag 1140
    
```

10

ES 2 422 424 T3

ggcctgctca cagccagaca ggtgaagatc agcaccgaca tcgtgaacgc caagatcacc 1200
 aagctgcagg ccaggcagca ggaccaggag aggctgagag tgttcgacgg catccccctg 1260
 ggcaccacctc aggtggccgg catgattgcc gagctgagcc ccgatagatt cagggtctgtg 1320
 ctggatgtgc tggccgaggt ggtggtgcag cccgtgggca agagcggcag aatcttcaac 1380
 cccgagcggg tgcaggtgaa ctggagatag 1410

<210> 10

<211> 469

5 <212> PRT

<213> Aparente recombinasa del bacteriófago PhiRv1

<400> 10

Met Arg Tyr Thr Thr Pro Val Arg Ala Ala Val Tyr Leu Arg Ile Ser
 1 5 10 15

Glu Asp Arg Ser Gly Glu Gln Leu Gly Val Ala Arg Gln Arg Glu Asp
 20 25 30

Cys Leu Lys Leu Cys Gly Gln Arg Lys Trp Val Pro Val Glu Tyr Leu
 35 40 45

Asp Asn Asp Val Ser Ala Ser Thr Gly Lys Arg Arg Pro Ala Tyr Glu
 50 55 60

Gln Met Leu Ala Asp Ile Thr Ala Gly Lys Ile Ala Ala Val Val Ala
 65 70 75 80

Trp Asp Leu Asp Arg Leu His Arg Arg Pro Ile Glu Leu Glu Ala Phe
 85 90 95

Met Ser Leu Ala Asp Glu Lys Arg Leu Ala Leu Ala Thr Val Ala Gly
 100 105 110

Asp Val Asp Leu Ala Thr Pro Gln Gly Arg Leu Val Ala Arg Leu Lys
 115 120 125

Gly Ser Val Ala Ala His Glu Thr Glu His Lys Lys Ala Arg Gln Arg
 130 135 140

Arg Ala Ala Arg Gln Lys Ala Glu Arg Gly His Pro Asn Trp Ser Lys
 145 150 155 160

Ala Phe Gly Tyr Leu Pro Gly Pro Asn Gly Pro Glu Pro Asp Pro Arg
 165 170 175

Thr Ala Pro Leu Val Lys Gln Ala Tyr Ala Asp Ile Leu Ala Gly Ala
 180 185 190

Ser Leu Gly Asp Val Cys Arg Gln Trp Asn Asp Ala Gly Ala Phe Thr
 195 200 205

10

ES 2 422 424 T3

Ile Thr Gly Arg Pro Trp Thr Thr Thr Thr Leu Ser Lys Phe Leu Arg
 210 215 220

Lys Pro Arg Asn Ala Gly Leu Arg Ala Tyr Lys Gly Ala Arg Tyr Gly
 225 230 235 240

Pro Val Asp Arg Asp Ala Ile Val Gly Lys Ala Gln Trp Ser Pro Leu
 245 250 255

Val Asp Glu Ala Thr Phe Trp Ala Ala Gln Ala Val Leu Asp Ala Pro
 260 265 270

Gly Arg Ala Pro Gly Arg Lys Ser Val Arg Arg His Leu Leu Thr Gly
 275 280 285

Leu Ala Gly Cys Gly Lys Cys Gly Asn His Leu Ala Gly Ser Tyr Arg
 290 295 300

Thr Asp Gly Gln Val Val Tyr Val Cys Lys Ala Cys His Gly Val Ala
 305 310 315 320

Ile Leu Ala Asp Asn Ile Glu Pro Ile Leu Tyr His Ile Val Ala Glu
 325 330 335

Arg Leu Ala Met Pro Asp Ala Val Asp Leu Leu Arg Arg Glu Ile His
 340 345 350

Asp Ala Ala Glu Ala Glu Thr Ile Arg Leu Glu Leu Glu Thr Leu Tyr
 355 360 365

Gly Glu Leu Asp Arg Leu Ala Val Glu Arg Ala Glu Gly Leu Leu Thr
 370 375 380

Ala Arg Gln Val Lys Ile Ser Thr Asp Ile Val Asn Ala Lys Ile Thr
 385 390 395 400

Lys Leu Gln Ala Arg Gln Gln Asp Gln Glu Arg Leu Arg Val Phe Asp
 405 410 415

Gly Ile Pro Leu Gly Thr Pro Gln Val Ala Gly Met Ile Ala Glu Leu
 420 425 430

Ser Pro Asp Arg Phe Arg Ala Val Leu Asp Val Leu Ala Glu Val Val
 435 440 445

Val Gln Pro Val Gly Lys Ser Gly Arg Ile Phe Asn Pro Glu Arg Val
 450 455 460

Gln Val Asn Trp Arg
 465

<210> 11
 <211> 99
 5 <212> ADN
 <213> Artificial

<220>
 <223> Sitio attP de SPBc2

10 <400> 11

ES 2 422 424 T3

	acggcagagt aagcttcttt ttttcgtagg atatgtagta agtatcttaa tatacagctt	60
	tatctgtttt ttaagatact tactactttt cttagtggg	99
	<210> 12	
	<211> 1315	
5	<212> ADN	
	<213> Artificial	
	<220>	
	<223> Secuencia de FIN	
10	<400> 12	
	aagcttactt accatgtcag atccagacat gataagatac attgatgagt ttggacaaac	60
	cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt	120
	atttgtaac attataagct gcaataaaca agttaacaac aacaattgca ttcattttat	180
	gtttcaggtt cagggggagg tgtgggagg tttttaaac aagtaaac tctacaaatg	240
	tggtatggct gattatgatc tctagtcaag gcactataca tcaaatattc cttattaacc	300
	cctttacaaa ttaaaaagct aaaggtacac aatttttgag catagttatt aatagcagac	360
	actctatgcc tgtgtggagt aagaaaaaac agtatgttat gattataact gttatgccta	420
	cttataaagg ttacagaata tttttccata attttcttgt atagcagtgc agctttttcc	480
	tttgtggtgt aaatagcaaa gcaagcaaga gttctattac taaacacagc atgactcaaa	540
	aaacttagca attctgaagg aaagtccttg gggcttctc ctttctctt cttttttgga	600
	ggagtagaat gttgagagtc agcagtagcc tcatcatcac tagatggcat ttcttctgag	660
	caaacaggt tttcctcatt aaaggcattc caccactgct cccttctc agttccatag	720
	gttggaatct aaaatacaca aacaattaga atcagtagtt taacacatta tactttaa	780
	aattttatat ttaccttaga gctttaaatc tctgtaggta gtttgtccaa ttatgtcaca	840
	ccacagaagt aaggttcctt cacaagatc cctcgagaaa aaaaatataa aagagatgga	900
	ggaacgggaa aaagttagtt gtggtgatag gtggcaagt gtattccgta agaacaaca	960
	gaaaagcatt tcatattatg gctgaactga gcgaacaagt gcaaaattta agcatcaacg	1020
	acaacaacga gaatggttat gttcctcctc acttaagagg aaaaccaaga agtgccagaa	1080
	ataacatgag caactacaat aacaacaacg gcggctacaa cgggtggccgt ggcggtggca	1140
	gcttctttag caacaaccgt cgtggtggtt acggcaacgg tggtttctt ggtggaaca	1200
	acggtggcag cagatctaac ggccgttctg gtggttagatg gatcgatggc aaacatgtcc	1260
	cagctccaag aaacgaaaag gccgagatcg ccatatttgg tgtccccgag gatcc	1315
15	<210> 13	
	<211> 96	
	<212> ADN	
	<213> Artificial	
20	<220>	
	<223> Sitio attB de SPBc2	
	<400> 13	
	tcagataaca gcttgggtggc acccattgtg ttcacaggag atacagcttt atctgtactg	60
25	atattaatga catgctgcac tcggtgtgaa agggca	96
	<210> 14	

ES 2 422 424 T3

<211> 99
 <212> ADN
 <213> Artificial

5 <220>
 <223> Sitio attP de SF370.1

<400> 14

acgaaaggag gtcgtgaaat ggataaaaaa atacagcgtt tttcatgtac aactatacta 60

10 gttgtagtgc ctaaataatg cttttaaac ttaaaaata 99

<210> 15
 <211> 96
 <212> ADN
 <213> Artificial

15 <220>
 <223> Sitio attB de SF370.1

20 <400> 15

taaaagggat aataacgttt gtaaaggaga ctgataatgg catgtacaac tatactcgtc 60

ggtaaaaagg catcttatga tggctcaacc atggtt 96

<210> 16
 <211> 52
 <212> ADN
 <213> Artificial

25 <220>
 <223> Sitio attP de Bxb1

30 <400> 16

gtggttgc tggtaacca ccgcggtctc agtgggtgac ggtacaaacc ca 52

35 <210> 17
 <211> 46
 <212> ADN
 <213> Artificial

40 <220>
 <223> Sitio attB de Bxb1

<400> 17

45 ggccggcttg tcgacgacgg cggctccgt cgtcaggatc atccgg 46

<210> 18
 <211> 99
 <212> ADN
 <213> Artificial

50 <220>
 <223> Sitio attP de A118

55 <400> 18

acgctagtag cttgtttatt tagattgttt agttcctcgt tttctctcgt tggaagaaga 60

agaaacgaga aactaaaatt ataaataaaa agtaaccta 99

60 <210> 19
 <211> 96

<212> ADN
 <213> Artificial

 <220>
 5 <223> Sitio attB de A118

 <400> 19

 ttgagctaat taaaaccagc tgtaactttt tcggatcaag ctatgaagga cgcaaagagg 60
 gaactaaaca cttaatggg gttaccata agccac 96
 10
 <210> 20
 <211> 99
 <212> ADN
 <213> Artificial
 15 <220>
 <223> Sitio attP de PhiRv1
 <400> 20

 acgagacagc agcacgcaca ggtgtagtgt atctcacagg tccacggtg gccgtggact 60
 gctgaagaac attccacgcc aggagatcaa ccatgacca 99
 20
 <210> 21
 <211> 96
 <212> ADN
 <213> Artificial
 25
 <220>
 <223> Sitio attB de PhiRv1

 <400> 21
 30
 tggcgtagca gcttctcgtg gtggtggaag gtggtggtgc ggggttgcc gtggtcgagg 60
 tggggtggtg gtagccattc ggtgtggccg tggggtg 96

Listado de secuencias

35 <110> RheoGene, Inc.

 <120> Serina recombinasas específicas de sitio y procedimientos de su uso

 <130> IN05P003WOEP
 40
 <140> EP05722807.4
 <141> 8-2-2.005

 <150> US 11/049.552
 45 <151> 2-2-2.005

 <160> 21

 <170> PatentIn versión 3.3
 50
 <210> 1
 <211> 1638
 <212> ADN
 <213> Artificial
 55
 <220>
 <223> gen sintético

 <400> 1
 60

ES 2 422 424 T3

atggaactga agaacatcgt gaacagctac aacatcacca acatcctggg ctacctgcgg 60
 agaagcaggc aggacatgga gagagagaag cggaccggcg aggacaccct caccgagcag 120
 aaggaactca tgaacaagat cctcaccgcc atcgagatcc cctacgagct gaagatggag 180
 atcggcagcg gcgagagcat cgacggcaga cccgtgttca aggagtgcct gaaggatctg 240
 gaggagggca agtaccaggc catcgcctgt aaggagatca ccaggctgag cagaggcagc 300
 tacagcgacg ccggccagat cgtgaacctg ctgcagagca agcggctcat catcatcacc 360
 ccctacaagg tgtacgaccc cagaaacccc gtcgacatgc ggcagatccg gttcgagctg 420
 ttcatggcca gggaggagtt cgagatgacc cgggagagaa tgaccggcgc caagtacacc 480
 tacgccgccc agggcaagtg gatcagcggc ctggccccct acggctacca gctgaacaag 540
 aaaaccagca agctggaccc cgtggaggac gaggccaagg tgggtgcagct catcttcaac 600
 atcttctga acgggctgaa cggcaaggac tacagctaca cagccatcgc cagccacctc 660
 accaatctgc agatccctac ccccagcggc aagaagcggg ggaaccagta caccatcaag 720
 gccatcctgc agaacgaggt gtacatcggc accgtgaagt acaaggtgcg ggagaaaacc 780
 aaggacggca agcggaccat caggcctgag aaggagcaga tcgtggtgca ggacgcccac 840
 gccctatca tcgacaagga gcagttccag cagagccagg tgaagatcgc caacaagggtg 900
 cccctgctgc ccaacaagga cgagttcgag ctgagcagc tggccggagt gtgcacctgc 960
 agcaagtgcg gcgagcctct gagcaagtac gagagcaagc gcatccggaa gaacaaggat 1020
 ggcaccgaga gcgtgtacca cgtgaagtcc ctcacctgca agaagaaca gtgcacctac 1080
 gtgcggtaca acgacgtgga gaacgccatc ctggattacc tgagcagcct gaacgacctg 1140
 aatgacagca ccctcacaaa gcacatcaac agcatgctct ccaagtacga ggacgacaac 1200
 agcaacatga aaaccaagaa gcagatgagc gacacactga gccagaagga gaaggagctt 1260

 aagaataagg agaacttcat cttcgacaag tacgagtccg gcatctactc cgacgagctg 1320
 ttctgaagc ggaaggccgc cctggacgag gaggttcaagg agctgcagaa cgccaagaac 1380
 gagctgaatg gcctgcagga taccagagc gagatcgaca gcaacaccgt gcggaacaac 1440
 atcaacaaga tcatcgacca gtaccacatc gagagcagca gcgagaagaa gaatgagctg 1500
 ctgcggatgg tgctgaagga cgtgatcgtg aacatgacc agaagcgcaa gggccccatc 1560
 ...cccgccagt tcgagatcac acccatcctg .cggttcaact ttatcttca tctcaccgcc 1620
 accaacagct tccactag 1638

5 <210> 2
 <211> 545
 <212> PRT
 <213> Fago SPBc2 de Bacillus subtilis

10 <400> 2

ES 2 422 424 T3

Met Glu Leu Lys Asn Ile Val Asn Ser Tyr Asn Ile Thr Asn Ile Leu
1 5 10 15

Gly Tyr Leu Arg Arg Ser Arg Gln Asp Met Glu Arg Glu Lys Arg Thr
20 25 30

Gly Glu Asp Thr Leu Thr Glu Gln Lys Glu Leu Met Asn Lys Ile Leu
35 40 45

Thr Ala Ile Glu Ile Pro Tyr Glu Leu Lys Met Glu Ile Gly Ser Gly
50 55 60

Glu Ser Ile Asp Gly Arg Pro Val Phe Lys Glu Cys Leu Lys Asp Leu
65 70 75 80

Glu Glu Gly Lys Tyr Gln Ala Ile Ala Val Lys Glu Ile Thr Arg Leu
85 90 95

Ser Arg Gly Ser Tyr Ser Asp Ala Gly Gln Ile Val Asn Leu Leu Gln
100 105 110

Ser Lys Arg Leu Ile Ile Ile Thr Pro Tyr Lys Val Tyr Asp Pro Arg
115 120 125

Asn Pro Val Asp Met Arg Gln Ile Arg Phe Glu Leu Phe Met Ala Arg
130 135 140

Glu Glu Phe Glu Met Thr Arg Glu Arg Met Thr Gly Ala Lys Tyr Thr
145 150 155 160

Tyr Ala Ala Gln Gly Lys Trp Ile Ser Gly Leu Ala Pro Tyr Gly Tyr
165 170 175

Gln Leu Asn Lys Lys Thr Ser Lys Leu Asp Pro Val Glu Asp Glu Ala
180 185 190

ES 2 422 424 T3

Lys Val Val Gln Leu Ile Phe Asn Ile Phe Leu Asn Gly Leu Asn Gly
 195 200 205
 Lys Asp Tyr Ser Tyr Thr Ala Ile Ala Ser His Leu Thr Asn Leu Gln
 210 215 220
 Ile Pro Thr Pro Ser Gly Lys Lys Arg Trp Asn Gln Tyr Thr Ile Lys
 225 230 235
 Ala Ile Leu Gln Asn Glu Val Tyr Ile Gly Thr Val Lys Tyr Lys Val
 245 250 255
 Arg Glu Lys Thr Lys Asp Gly Lys Arg Thr Ile Arg Pro Glu Lys Glu
 260 265 270
 Gln Ile Val Val Gln Asp Ala His Ala Pro Ile Ile Asp Lys Glu Gln
 275 280 285
 Phe Gln Gln Ser Gln Val Lys Ile Ala Asn Lys Val Pro Leu Leu Pro
 290 295 300
 Asn Lys Asp Glu Phe Glu Leu Ser Glu Leu Ala Gly Val Cys Thr Cys
 305 310 315 320
 Ser Lys Cys Gly Glu Pro Leu Ser Lys Tyr Glu Ser Lys Arg Ile Arg
 325 330 335
 Lys Asn Lys Asp Gly Thr Glu Ser Val Tyr His Val Lys Ser Leu Thr
 340 345 350
 Cys Lys Lys Asn Lys Cys Thr Tyr Val Arg Tyr Asn Asp Val Glu Asn
 355 360 365
 Ala Ile Leu Asp Tyr Leu Ser Ser Leu Asn Asp Leu Asn Asp Ser Thr
 370 375 380
 Leu Thr Lys His Ile Asn Ser Met Leu Ser Lys Tyr Glu Asp Asp Asn
 385 390 395 400
 Ser Asn Met Lys Thr Lys Lys Gln Met Ser Glu His Leu Ser Gln Lys
 405 410 415
 Glu Lys Glu Leu Lys Asn Lys Glu Asn Phe Ile Phe Asp Lys Tyr Glu
 420 425 430
 Ser Gly Ile Tyr Ser Asp Glu Leu Phe Leu Lys Arg Lys Ala Ala Leu
 435 440 445
 Asp Glu Glu Phe Lys Glu Leu Gln Asn Ala Lys Asn Glu Leu Asn Gly
 450 455 460

ES 2 422 424 T3

Leu Gln Asp Thr Gln Ser Glu Ile Asp Ser Asn Thr Val Arg Asn Asn
465 470 475 480

Ile Asn Lys Ile Ile Asp Gln Tyr His Ile Glu Ser Ser Ser Glu Lys
485 490 495

Lys Asn Glu Leu Leu Arg Met Val Leu Lys Asp Val Ile Val Asn Met
500 505 510

Thr Gln Lys Arg Lys Gly Pro Ile Pro Ala Gln Phe Glu Ile Thr Pro
515 520 525

Ile Leu Arg Phe Asn Phe Ile Phe Asp Leu Thr Ala Thr Asn Ser Phe
530 535 540

His
545

<210> 3
<211> 1416
5 <212> ADN
<213> Artificial

<220>
<223> Gen sintético

10 <400> 3

```

atgcggaagg tggccatcta cagcagggtg agcaccatca atcaggccga ggagggctac    60
agcatccagg gccagatcga ggcctcacc aagtactgcg aggccatgga gtggaagatc    120
tacaagaact acagcgacgc cggcttcagc ggcggcaagc tggagagacc cgccatcacc    180
gagttgatcg aggacggcaa gaacaacaag ttcgacacca tcctggtgta caagctggac    240
cggctgagca gaaacgtgaa ggacaccctg tacctggtga aggacgtgtt caccgccaac    300
aacatccact tcgtgagcct gaaggagaac atcgacacca gcagcgccat gggcaatctg    360
ttcctcacac tgctgagcgc aattgccgag ttcgagcggg agcagatcaa ggaacggatg    420
cagttcggcg tgatgaacag agccaagagc ggcaagacca cgcctggaa aaccctcca    480
tacggctacc ggtacaacaa ggacgagaaa accctgagcg tgaacgagct ggaggccgcc    540
aatgtgaggc agatgttcga catgatcatc agcggctgca gcatcatgag catcaccaac    600
tacgcccggg acaacttcgt gggcaacacc tggaccacg tgaaggtgaa gcggatcctg    660
gagaacgaga cctacaaagg cctggtgaag taccgggagc agaccttag cggcgatcac    720
caggccatca tcgacgaaaa gacctacaac aaggcccaga tcgccctggc ccacagaacc    780
gacaccaaga ccaacaccag acccttcag ggcaagtaca tgctgagcca catcgccaag    840
tgcggtact gtggcgcctc tctgaagggtg tgcaccggca gggccaagaa tgacggcacc    900
cggagacaga cctacgtgtg cgtgaacaag accgagagcc tggccagaag gagcgtgaac    960
aactacaaca accagaagat ctgcaacacc ggccggtacg agaagaagca catcgagaag   1020
tacgtgatcg acgtgctgta taagctgcag cacgacaagg agtacctgaa gaagatcaag   1080
    
```

ES 2 422 424 T3

aaggacgaca acatcatcga taccaccccc ctgaagaagg agatcgagat catcgacaag 1140
 aagattaacc ggctgaacga cctgtacatc aacgacctca tcgacctgcc caagctgaag 1200
 aaagacatcg aggagctgaa ccacctgaag gacgactaca ataaggccat caagctgaac 1260
 tacctggaca agaagaacga ggacagcctg ggcattgctca tggacaacct ggacatccgc 1320
 aagagcagct acgacgtgca gagccggatc gtgaagcagc tcatcgacag ggtggagggtg 1380
 accatggaca atatcgacat catcttcaag ttctag 1416

<210> 4
 <211> 471
 5 <212> PRT

<213> Aparente recombinasa del bacteriófago SF370.1

<400> 4

10

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

ES 2 422 424 T3

Cys Ser Ile Met Ser Ile Thr Asn Tyr Ala Arg Asp Asn Phe Val Gly
 195 200 205
 Asn Thr Trp Thr His Val Lys Val Lys Arg Ile Leu Glu Asn Glu Thr
 210 215 220
 Tyr Lys Gly Leu Val Lys Tyr Arg Glu Gln Thr Phe Ser Gly Asp His
 225 230 235 240
 Gln Ala Ile Ile Asp Glu Lys Thr Tyr Asn Lys Ala Gln Ile Ala Leu
 245 250 255
 Ala His Arg Thr Asp Thr Lys Thr Asn Thr Arg Pro Phe Gln Gly Lys
 260 265 270
 Tyr Met Leu Ser His Ile Ala Lys Cys Gly Tyr Cys Gly Ala Pro Leu
 275 280 285
 Lys Val Cys Thr Gly Arg Ala Lys Asn Asp Gly Thr Arg Arg Gln Thr
 290 295 300
 Tyr Val Cys Val Asn Lys Thr Glu Ser Leu Ala Arg Arg Ser Val Asn
 305 310 315 320
 Asn Tyr Asn Asn Gln Lys Ile Cys Asn Thr Gly Arg Tyr Glu Lys Lys
 325 330 335
 His Ile Glu Lys Tyr Val Ile Asp Val Leu Tyr Lys Leu Gln His Asp
 340 345 350
 Lys Glu Tyr Leu Lys Lys Ile Lys Lys Asp Asp Asn Ile Ile Asp Ile
 355 360 365
 Thr Pro Leu Lys Lys Glu Ile Glu Ile Ile Asp Lys Lys Ile Asn Arg
 370 375 380
 Leu Asn Asp Leu Tyr Ile Asn Asp Leu Ile Asp Leu Pro Lys Leu Lys
 385 390 395 400
 Lys Asp Ile Glu Glu Leu Asn His Leu Lys Asp Asp Tyr Asn Lys Ala
 405 410 415
 Ile Lys Leu Asn Tyr Leu Asp Lys Lys Asn Glu Asp Ser Leu Gly Met
 420 425 430
 Leu Met Asp Asn Leu Asp Ile Arg Lys Ser Ser Tyr Asp Val Gln Ser
 435 440 445
 Arg Ile Val Lys Gln Leu Ile Asp Arg Val Glu Val Thr Met Asp Asn
 450 455 460
 Ile Asp Ile Ile Phe Lys Phe

465

470

5 <210> 5
 <211> 1503
 <212> ADN
 <213> Gen sintético

10 <400> 5

ES 2 422 424 T3

```

atgcgggctc tgggtgtgat caggctgagc agagtgaccg atgccaccac aagccctgag    60
agacagctgg agagctgcca gcagctgtgt gcccagagag gatgggacgt ggtgggagtg    120
gccgaggatc tggatgtgag cggagccgtg gaccccttcg acagaaagcg gagaccaaac    180
ctggccagat ggctggcctt tgaggagcag cccttcgatg tgatcgtggc ctacagagtg    240
gacaggctga cccggagcat tagacacctc cagcagctgg tgcactgggc cgaggaccac    300
aagaaactgg tggtgagcgc cacagaggcc cacttcgata ccaccacccc ctttgctgca    360
gtggtgatcg ccctgatggg cacagtggcc cagatggagc tggaggccat caaggagagg    420
aatcggctcg ccgccactt caatatcagg gccggcaagt acagaggaag cctgcctcct    480
tggggctacc tgcccacaag agtggatggc gagtggagac tgggtcctga ccctgtgag    540
agggagagaa tcctggaagt gtatcaccgc gtggtggaca atcacgagcc tctgacactg    600
gtggcccacg acctgaatag gagaggcgtg ctgtcccca aggattactt cgcccagctc    660
cagggcagag agcctcaggg cagagagtgg tctgccaccg ccctgaaaag atctatgatc    720
agcgaggcca tgctgggcta cgccaccctg aatggcaaga ccgtagggga tgatgatgga    780
gcccctctgg tgagagccga gcccatcctg acaagggaac agctggaggc tctgagagcc    840
gaactggtga aaaccagcag agccaagcct gccgtgagca cacctagcct gctgctgaga    900
gtgctgttct gtgccgtgtg tggcgagcct gcctacaagt ttgccggcgg aggcagaaaag    960
caccctcggt acagatgtag gagcatgggc ttccctaagc actgcggcaa tggcaccgtg   1020
gccatggccg aatgggacgc cttttgagag gagcaagtgc tggatctgct gggagatgcc   1080
gagaggctgg agaaagtgtg ggtggccgga tccgattctg ccgtggaact ggccgaagtg   1140
aatgctgaac tgggtggacct gaccagcctg atcggcagcc ctgcctatag agccggaagc   1200
cctcagagag aagccctgga cgccagaatt gccgccctgg ccgccagaca ggaggaactg   1260
gagggactgg aggccagacc ttctggctgg gagtggagag agaccggcca gagattcggc   1320
gattggtgga gggagcagga taccgccgcc aagaacacct ggctgcggag catgaacgtg   1380
aggctgacct tcgacgtgag aggcggcctg accagaacca tcgacttcgg cgacctccag   1440
gagtatgagc agcacctgag actgggaagc gtggtggaga gactgcacac aggcattgtcc   1500
tag                                                                    1503

```

<210> 6

<211> 500

5 <212> PRT

<213> Aparente recombinasa del micobacteriófago Bxb1

<400> 6

ES 2 422 424 T3

Met Arg Ala Leu Val Val Ile Arg Leu Ser Arg Val Thr Asp Ala Thr
1 5 10 15

Thr Ser Pro Glu Arg Gln Leu Glu Ser Cys Gln Gln Leu Cys Ala Gln
20 25 30

Arg Gly Trp Asp Val Val Gly Val Ala Glu Asp Leu Asp Val Ser Gly
35 40 45

Ala Val Asp Pro Phe Asp Arg Lys Arg Arg Pro Asn Leu Ala Arg Trp
50 55 60

Leu Ala Phe Glu Glu Gln Pro Phe Asp Val Ile Val Ala Tyr Arg Val
65 70 75 80

Asp Arg Leu Thr Arg Ser Ile Arg His Leu Gln Gln Leu Val His Trp
85 90 95

Ala Glu Asp His Lys Lys Leu Val Val Ser Ala Thr Glu Ala His Phe
100 105 110

Asp Thr Thr Thr Pro Phe Ala Ala Val Val Ile Ala Leu Met Gly Thr
115 120 125

Val Ala Gln Met Glu Leu Glu Ala Ile Lys Glu Arg Asn Arg Ser Ala
130 135 140

Ala His Phe Asn Ile Arg Ala Gly Lys Tyr Arg Gly Ser Leu Pro Pro
145 150 155 160

Trp Gly Tyr Leu Pro Thr Arg Val Asp Gly Glu Trp Arg Leu Val Pro
165 170 175

Asp Pro Val Gln Arg Glu Arg Ile Leu Glu Val Tyr His Arg Val Val
180 185 190

Asp Asn His Glu Pro Leu His Leu Val Ala His Asp Leu Asn Arg Arg
195 200 205

Gly Val Leu Ser Pro Lys Asp Tyr Phe Ala Gln Leu Gln Gly Arg Glu
210 215 220

Pro Gln Gly Arg Glu Trp Ser Ala Thr Ala Leu Lys Arg Ser Met Ile
225 230 235 240

Ser Glu Ala Met Leu Gly Tyr Ala Thr Leu Asn Gly Lys Thr Val Arg
245 250 255

Asp Asp Asp Gly Ala Pro Leu Val Arg Ala Glu Pro Ile Leu Thr Arg
260 265 270

Glu Gln Leu Glu Ala Leu Arg Ala Glu Leu Val Lys Thr Ser Arg Ala

ES 2 422 424 T3

275	280	285
Lys Pro Ala Val Ser Thr	Pro Ser Leu Leu Leu	Arg Val Leu Phe Cys
290	295	300
Ala Val Cys Gly Glu Pro Ala Tyr Lys Phe Ala Gly Gly Gly Arg Lys		
305	310	315
His Pro Arg Tyr Arg Cys Arg Ser Met Gly Phe Pro Lys His Cys Gly		
	325	330
Asn Gly Thr Val Ala Met Ala Glu Trp Asp Ala Phe Cys Glu Glu Gln		
	340	345
Val Leu Asp Leu Leu Gly Asp Ala Glu Arg Leu Glu Lys Val Trp Val		
	355	360
Ala Gly Ser Asp Ser Ala Val Glu Leu Ala Glu Val Asn Ala Glu Leu		
	370	375
Val Asp Leu Thr Ser Leu Ile Gly Ser Pro Ala Tyr Arg Ala Gly Ser		
385	390	395
Pro Gln Arg Glu Ala Leu Asp Ala Arg Ile Ala Ala Leu Ala Ala Arg		
	405	410
Gln Glu Glu Leu Glu Gly Leu Glu Ala Arg Pro Ser Gly Trp Glu Trp		
	420	425
Arg Glu Thr Gly Gln Arg Phe Gly Asp Trp Trp Arg Glu Gln Asp Thr		
	435	440
Ala Ala Lys Asn Thr Trp Leu Arg Ser Met Asn Val Arg Leu Thr Phe		
	450	455
Asp Val Arg Gly Gly Leu Thr Arg Thr Ile Asp Phe Gly Asp Leu Gln		
465	470	475
Glu Tyr Glu Gln His Leu Arg Leu Gly Ser Val Val Glu Arg Leu His		
	485	490
		495
Thr Gly Met Ser		
	500	

<210> 7
 <211> 1359
 5 <212> ADN
 <213> Gen sintético

<400> 7

	atgaaggccg ccatctacat cagagtgagc acccaggagc aggtggagaa ctacagcatc	60
10	caggcccaga ccgagaagct caccgcctg tgcagaagca aggactggga cgtgtacgac	120

ES 2 422 424 T3

atcttcatcg acggcggcta cagcggcagc aacatgaaca gacccgccct gaacgagatg 180
ctgagcaagc tgcacgagat cgatgccgtg gtggtgtaca ggctggacag gctgagcaga 240
agccagaggg acaccatcac cctcatcgag gagtacttcc tgaagaacaa cgtggagttc 300
gtgagcctga gcgagaccct ggacaccagc agccccttcg gcagagccat gatcggcatc 360
ctgagcgtgt tcgccagct cgagagagag accatccggg acaggatggt gatgggcaag 420
atcaagagga tcgaggccgg cctgcccctc accaccgcca agggcagaac cttcggctac 480
gacgtgatcg acaccaagct gtacatcaac gaggaggagg ccaagcagct gcagctcatc 540
tacgatatct tcgaggagga gcagagcatc accttcctgc agaagcggct gaagaagctg 600
ggcttcaagg tgcggacct caaccggtac aacaactggc tcaccaacga cctgtactgc 660
ggctacgtga gctacaagga caaggtgcac gtgaagggga tccacgagcc catcatcagc 720
gaggagcagt tctaccgggt gcaggagatc ttcacccgca tgggcaagaa cccaacatg 780
aacggggaca gcgccagcct gctgaacaat ctggtggtgt gcagcaagtg cggcctgggc 840
ttcgtgcaca ggagaaagga caccatgagc cggggcaaga agtaccacta ccggtactac 900
agctgcaaga cctacaagca caccacgag ctggagaagt gcggcaacaa gatctggagg 960
gccgacaagc tggaggagtt gatcatcaac cgggtgaaca actacagctt cgccagccgg 1020
aacgtggata aggaggacga gctggacagc ctgaatgaga agcttaagat cgagcacgcc 1080
aagaagaagc gcctgttcga cctgtacatt aacggcagct acgaggtgag cgagctggac 1140
tccatgatga acgacatcga cgcccagatc aactactacg agagccagat cgaggccaac 1200
gaggagctga agaagaacaa gaagatccag gagaacctgg ccgacctggc caccgtggat 1260
ttcgacagcc tggagttcag ggagaagcag ctgtacctga agtccctcat caataagatc 1320
tacatcgacg gggagcaggt gaccatcgag tggctgtag 1359

<210> 8

<211> 452

5 <212> PRT

<213> Aparente recombinasa del bacteriófago A118

<400> 8

Met Lys Ala Ala Ile Tyr Ile Arg Val Ser Thr Gln Glu Gln Val Glu
1 5 10 15

Asn Tyr Ser Ile Gln Ala Gln Thr Glu Lys Leu Thr Ala Leu Cys Arg
20 25 30

Ser Lys Asp Trp Asp Val Tyr Asp Ile Phe Ile Asp Gly Gly Tyr Ser
35 40 45

Gly Ser Asn Met Asn Arg Pro Ala Leu Asn Glu Met Leu Ser Lys Leu
50 55 60

10 His Glu Ile Asp Ala Val Val Val Tyr Arg Leu Asp Arg Leu Ser Arg
65 70 75 80

ES 2 422 424 T3

Ser Gln Arg Asp Thr Ile Thr Leu Ile Glu Glu Tyr Phe Leu Lys Asn
85 90 95

Asn Val Glu Phe Val Ser Leu Ser Glu Thr Leu Asp Thr Ser Ser Pro
100 105 110

Phe Gly Arg Ala Met Ile Gly Ile Leu Ser Val Phe Ala Gln Leu Glu
115 120 125

Arg Glu Thr Ile Arg Asp Arg Met Val Met Gly Lys Ile Lys Arg Ile
130 135 140

Glu Ala Gly Leu Pro Leu Thr Thr Ala Lys Gly Arg Thr Phe Gly Tyr
145 150 155 160

Asp Val Ile Asp Thr Lys Leu Tyr Ile Asn Glu Glu Glu Ala Lys Gln
165 170 175

Leu Gln Leu Ile Tyr Asp Ile Phe Glu Glu Glu Gln Ser Ile Thr Phe
180 185 190

Leu Gln Lys Arg Leu Lys Lys Leu Gly Phe Lys Val Arg Thr Tyr Asn
195 200 205

Arg Tyr Asn Asn Trp Leu Thr Asn Asp Leu Tyr Cys Gly Tyr Val Ser
210 215 220

Tyr Lys Asp Lys Val His Val Lys Gly Ile His Glu Pro Ile Ile Ser
225 230 235 240

Glu Glu Gln Phe Tyr Arg Val Gln Glu Ile Phe Thr Arg Met Gly Lys
245 250 255

Asn Pro Asn Met Asn Arg Asp Ser Ala Ser Leu Leu Asn Asn Leu Val
260 265 270

Val Cys Ser Lys Cys Gly Leu Gly Phe Val His Arg Arg Lys Asp Thr
275 280 285

Met Ser Arg Gly Lys Lys Tyr His Tyr Arg Tyr Tyr Ser Cys Lys Thr
290 295 300

Tyr Lys His Thr His Glu Leu Glu Lys Cys Gly Asn Lys Ile Trp Arg
305 310 315 320

Ala Asp Lys Leu Glu Glu Leu Ile Ile Asn Arg Val Asn Asn Tyr Ser
325 330 335

Phe Ala Ser Arg Asn Val Asp Lys Glu Asp Glu Leu Asp Ser Leu Asn
340 345 350

ES 2 422 424 T3

Glu Lys Leu Lys Ile Glu His Ala Lys Lys Lys Arg Leu Phe Asp Leu
 355 360 365

Tyr Ile Asn Gly Ser Tyr Glu Val Ser Glu Leu Asp Ser Met Met Asn
 370 375 380

Asp Ile Asp Ala Gln Ile Asn Tyr Tyr Glu Ser Gln Ile Glu Ala Asn
 385 390 395 400

Glu Glu Leu Lys Lys Asn Lys Lys Ile Gln Glu Asn Leu Ala Asp Leu
 405 410 415

Ala Thr Val Asp Phe Asp Ser Leu Glu Phe Arg Glu Lys Gln Leu Tyr
 420 425 430

Leu Lys Ser Leu Ile Asn Lys Ile Tyr Ile Asp Gly Glu Gln Val Thr
 435 440 445

Ile Glu Trp Leu
 450

<210> 9
 <211> 1410
 <212> ADN
 <213> Gen sintético

5

<400> 9

atgCGgtaca ccacccccgt gagagccgcc gtgtacctga gaatcagcga ggacagaagc 60
 ggcgagcagc tgggcgtggc cagacagaga gaggactgcc tgaagctgtg cggccagaga 120
 aagtgggtgc cctgggagta cctggacaac gatgtgagcg ccagcaccgg caagaggaga 180
 cccgcctacg agcagatgct ggccgacatc accgccggca agatcgccgc cgtggtggcc 240
 tgggacctgg ataggctgca caggagacct atcgagctgg aggccttcat gagcctggcc 300
 gatgagaaaa gactggccct ggccaccgtg gccggcgacg tggacctggc cacccccag 360
 ggcagactgg tggccagact taagggcagc gtggccgccc acgagaccga gcacaagaag 420
 gccagacagc ggagagccgc cagacagaag gccgagagag gccaccccaa ctggagcaag 480
 gccttcggct acctgctgg cccaacggc cccgagccc accctagaac cggccctctg 540
 gtgaagcagg cctacgccga catcctggcc ggagccagcc tgggcgacgt gtgcagacag 600
 tggaatgacg ccggagcctt caccatcacc ggcagaccct ggaccaccac caccctgagc 660
 aagttcctgc ggaagcccag aaacgccggc ctgagagcct acaagggcgc cagatacggc 720
 cccgtcgaca gagatgccat cgtgggcaag gccagtgga gcccctggt ggacgaggcc 780
 accttctggg ccgctcaggc cgtgctggac gccctggca gagccccagg cagaaagagc 840
 gtgCGgagac acctgctcac cggcctggcc ggctgcggca agtgcggcaa ccacctggcc 900
 ggcagctaca gaaccgatgg gcaggtggtg tacgtgtgca aggcctgcca cggcgtggcc 960
 attctggccg acaacatcga gcccatcctg taccatcctg tggccgagag actggccatg 1020
 cccgacgccg tggatctgct gaggaggag atccacgacg ccgccgaggc cgagaccatc 1080

10

ES 2 422 424 T3

agactcgagc tggaaacct gtacggcgag ctggacagac tggccgtgga gagagccgag 1140
 ggctgtctca cagccagaca ggtgaagatc agcaccgaca tcgtgaacgc caagatcacc 1200
 aagctgcagg ccaggcagca ggaccaggag aggctgagag tgttcgacgg catccccctg 1260
 ggcaccctc aggtggccgg catgattgcc gagctgagcc ccgatagatt cagggtgtg 1320
 ctggatgtgc tggccgaggt ggtggtgcag cccgtgggca agagcggcag aatcttcaac 1380
 cccgagcggg tgcaggtgaa ctggagatag 1410

<210> 10

<211> 469

5 <212> PRT

<213> Aparente recombinasa del bacteriófago PhiRv1

<400> 10

Met Arg Tyr Thr Thr Pro Val Arg Ala Ala Val Tyr Leu Arg Ile Ser
 1 5 10 15

Glu Asp Arg Ser Gly Glu Gln Leu Gly Val Ala Arg Gln Arg Glu Asp
 20 25 30

Cys Leu Lys Leu Cys Gly Gln Arg Lys Trp Val Pro Val Glu Tyr Leu
 35 40 45

Asp Asn Asp Val Ser Ala Ser Thr Gly Lys Arg Arg Pro Ala Tyr Glu
 50 55 60

Gln Met Leu Ala Asp Ile Thr Ala Gly Lys Ile Ala Ala Val Val Ala
 65 70 75 80

Trp Asp Leu Asp Arg Leu His Arg Arg Pro Ile Glu Leu Glu Ala Phe
 85 90 95

Met Ser Leu Ala Asp Glu Lys Arg Leu Ala Leu Ala Thr Val Ala Gly
 100 105 110

Asp Val Asp Leu Ala Thr Pro Gln Gly Arg Leu Val Ala Arg Leu Lys
 115 120 125

Gly Ser Val Ala Ala His Glu Thr Glu His Lys Lys Ala Arg Gln Arg
 130 135 140

Arg Ala Ala Arg Gln Lys Ala Glu Arg Gly His Pro Asn Trp Ser Lys
 145 150 155 160

Ala Phe Gly Tyr Leu Pro Gly Pro Asn Gly Pro Glu Pro Asp Pro Arg
 165 170 175

Thr Ala Pro Leu Val Lys Gln Ala Tyr Ala Asp Ile Leu Ala Gly Ala
 180 185 190

10

ES 2 422 424 T3

Ser Leu Gly Asp Val Cys Arg Gln Trp Asn Asp Ala Gly Ala Phe Thr
 195 200 205

Ile Thr Gly Arg Pro Trp Thr Thr Thr Thr Leu Ser Lys Phe Leu Arg
 210 215 220

Lys Pro Arg Asn Ala Gly Leu Arg Ala Tyr Lys Gly Ala Arg Tyr Gly
 225 230 235 240

Pro Val Asp Arg Asp Ala Ile Val Gly Lys Ala Gln Trp Ser Pro Leu
 245 250 255

Val Asp Glu Ala Thr Phe Trp Ala Ala Gln Ala Val Leu Asp Ala Pro
 260 265 270

Gly Arg Ala Pro Gly Arg Lys Ser Val Arg Arg His Leu Leu Thr Gly
 275 280 285

Leu Ala Gly Cys Gly Lys Cys Gly Asn His Leu Ala Gly Ser Tyr Arg
 290 295 300

Thr Asp Gly Gln Val Val Tyr Val Cys Lys Ala Cys His Gly Val Ala
 305 310 315 320

Ile Leu Ala Asp Asn Ile Glu Pro Ile Leu Tyr His Ile Val Ala Glu
 325 330 335

Arg Leu Ala Met Pro Asp Ala Val Asp Leu Leu Arg Arg Glu Ile His
 340 345 350

Asp Ala Ala Glu Ala Glu Thr Ile Arg Leu Glu Leu Glu Thr Leu Tyr
 355 360 365

Gly Glu Leu Asp Arg Leu Ala Val Glu Arg Ala Glu Gly Leu Leu Thr
 370 375 380

Ala Arg Gln Val Lys Ile Ser Thr Asp Ile Val Asn Ala Lys Ile Thr
 385 390 395 400

Lys Leu Gln Ala Arg Gln Gln Asp Gln Glu Arg Leu Arg Val Phe Asp
 405 410 415

Gly Ile Pro Leu Gly Thr Pro Gln Val Ala Gly Met Ile Ala Glu Leu
 420 425 430

Ser Pro Asp Arg Phe Arg Ala Val Leu Asp Val Leu Ala Glu Val Val
 435 440 445

Val Gln Pro Val Gly Lys Ser Gly Arg Ile Phe Asn Pro Glu Arg Val
 450 455 460

Gln Val Asn Trp Arg

465

5 <210> 11
 <211> 99
 <212> ADN
 <213> Artificial

10 <220>
 <223> Sitio attP de SPBc2

ES 2 422 424 T3

<400> 11

acggcagagt aagcttcttt ttttcgtag atatgtagta agtatcttaa tatacagctt 60
tatctgtttt ttaagatact tactactttt cttagtgga 99

5 <210> 12
<211> 1315
<212> ADN
<213> Artificial

10 <220>
<223> Secuencia FIN

<400> 12

aagcttactt accatgtcag atccagacat gataagatac attgatgagt ttggacaaac 60
cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt 120
at ttgtgaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat 180
gtttcagggt cagggggagg tgtgggagg tttttaaagc aagtaaaacc tctacaaatg 240
tggatggct gattatgatc tctagtcaag gcactataca tcaaataatc cttattaacc 300
cctttacaaa ttaaaaagct aaaggtacac aat ttttgag catagttatt aatagcagac 360
actctatgcc tgtgtggagt aagaaaaaac agtatgttat gattataact gttatgccta 420
cttataaagg ttacagaata tttttccata attttcttgt atagcagtgc agctttttcc 480
tttgtggtgt aatagcaaa gcaagcaaga gttctattac taaacacagc atgactcaaa 540
aaacttagca attctgaagg aaagtccttg gggcttctta ctttctctt cttttttgga 600
ggagtagaat gttgagagtc agcagtagcc tcatcatcac tagatggcat ttcttctgag 660
caaacaggt tttctcatt aaaggcattc caccactgct cccattcatc agttccatag 720
gttggaatct aaaatacaca aacaattaga atcagtagtt taacacatta tacacttaa 780
aat tttatattac ttaccttaga gctttaaatc tctgtaggta gtttgcctaa ttatgacaca 840
ccacagaagt aaggttcctt cacaaagatc cctcgagaaa aaaaatataa aagagatgga 900
ggaacgggaa aaagttagtt gtggtgatag gtggcaagt gtattccgta agaacaaca 960
gaaaagcatt tcatattatg gctgaactga gcgaacaagt gcaaaattta agcatcaacg 1020
acaacaacga gaatggttat gttcctcctc acttaagagg aaaaccaaga agtgccagaa 1080
ataacatgag caactacaat aacaacaacg gcggctacaa cgggtggcctg ggcgggtggca 1140
gcttcttttag caacaaccgt cgtggtggtt acggcaacgg tggtttcttc ggtggaaaca 1200
acgggtggcag cagatctaac ggccgttctg gtggtagatg gatcgatggc aaacatgtcc 1260
15 cagctccaag aaacgaaaag gccgagatcg ccatatttgg tgtccccgag gatcc 1315

<210> 13
<211> 96
<212> ADN
20 <213> Artificial

<220>
<223> Sitio attB de SPBc2

25 <400> 13

tcagataaca gcttgggtggc acccattgtg ttcacaggag atacagcttt atctgtactg 60
atattaatga catgctgcac tcggtgtgaa agggca 96

ES 2 422 424 T3

<210> 14
 <211> 99
 <212> ADN
 5 <213> Artificial

 <220>
 <223> Sitio attP de SF370.1

 10 <400> 14

acgaaaggag gtcgtgaaat ggataaaaa atacagcgtt tttcatgtac aactatacta 60
gttgtagtgc ctaaataatg cttttaaac ttaaaaata 99

 <210> 15
 <211> 96
 <212> ADN
 15 <213> Artificial

 <220>
 20 <223> Sitio attB de SF370.1

 <400> 15

taaaagggat aataacgttt gtaaaggaga ctgataatgg catgtacaac tatactcgtc 60
ggtaaaaagg catcttatga tggctcaacc atggtt 96

 25
 <210> 16
 <211> 52
 <212> ADN
 30 <213> Artificial

 <220>
 <223> Sitio attP de Bxb1

 <400> 16
 35
 gtggtttgc tggcaacca ccgcggtctc agtgggttac ggtacaaacc ca 52

 <210> 17
 <211> 46
 40 <212> ADN
 <213> Artificial
 <220>
 <223> Sitio attB de Bxb1
 <400> 17
 45
 ggccggcttg tcgacgacgg cggctccgt cgtcaggatc atccgg 46

 <210> 18
 <211> 99
 50 <212> ADN
 <213> Artificial

 <220>
 <223> Sitio attP de A118
 55
 <400> 18

acgctagtag cttgtttatt tagattgttt agttcctcgt tttctctcgt tgaagaaga 60
agaaacgaga aactaaaatt ataaataaaa agtaaccta 99

 60 <210> 19

ES 2 422 424 T3

<211> 96
 <212> ADN
 <213> Artificial

5 <220>
 <223> Sitio attB de A118

 <400> 19

ttgagctaat taaaaccagc tgtaactttt tcggatcaag ctatgaagga cgcaaagagg 60

10 **gaactaaaca ctttaattggt gttaccata agccac 96**

<210> 20
 <211> 99
 <212> ADN
 15 <213> Artificial

<220>
 <223> Sitio attP de PhiRv1

20 <400> 20

acgagacagc agcacgcaca ggtgtagtgt atctcacagg tccacggtg gccgtggact 60

gctgaagaac attccacgcc aggagatcaa ccatgacca 99

<210> 21
 25 <211> 96
 <212> ADN
 <213> Artificial

<220>
 30 <223> Sitio attB de PhiRv1

 <400> 21

tggcgtagca gcttctcgtg gtggtggaag gtgttggtgc ggggttgcc gtggtcgagg 60

tggggtggtg gtagcattc ggtgtggccg tgggtg 96

REIVINDICACIONES

1. Un método para obtener una recombinación específica de sitio en una célula eucariota aislada, comprendiendo el método: proporcionar una célula eucariota aislada que comprende un primer sitio de recombinación y un segundo sitio de recombinación con una secuencia de polinucleótidos flanqueada por un tercer sitio de recombinación y un cuarto sitio de recombinación; poner en contacto los sitios de recombinación con un polipéptido de recombinasa procarionta, que tiene como resultado la recombinación entre los sitios de recombinación, en el que el polipéptido de recombinasa puede participar en la recombinación entre el primero y el tercer sitios de recombinación y el segundo y el cuarto sitios de recombinación, el primero y el segundo sitios de recombinación son *attP* o *attB*, el tercero y el cuarto sitios de recombinación son *attB* o *attP* y la recombinasa es una recombinasa del fago de *Mycobacterium tuberculosis* y una recombinasa del fago de *Mycobacterium smegmatis*, en el que la recombinasa se selecciona del grupo que consiste en una recombinasa ϕ Rv1 y una recombinasa Bxb1, siempre que cuando el primero y el segundo sitios de unión para recombinación son *attB*, el tercero y el cuarto sitios de unión para recombinación son *attP* y cuando el primero y el segundo sitios de unión para recombinación son *attP* el tercero y el cuarto sitios de unión para recombinación son *attB*.
2. El método de la reivindicación 1, en el que el polipéptido de recombinasa se introduce en la célula eucariota mediante la expresión de un polinucleótido que codifica el polipéptido de recombinasa.
3. El método de la reivindicación 1, en el que el polipéptido de recombinasa se introduce en la célula eucariota como un polipéptido.
4. El método de la reivindicación 1, en el que el polipéptido de recombinasa se introduce en la célula eucariota mediante ARN mensajero que codifica el polipéptido de recombinasa.
5. Un método para obtener recombinaciones específicas de múltiples sitios en una célula eucariota aislada, comprendiendo el método: proporcionar una célula eucariota aislada que comprende un primer sitio de recombinación y un segundo sitio de recombinación con una secuencia de polinucleótidos flanqueada por un tercer sitio de recombinación y un cuarto sitio de recombinación; poner en contacto el primero y el segundo sitios de recombinación con un primer polipéptido de recombinasa procarionta, poner en contacto el tercero y el cuarto sitios de recombinación con un segundo polipéptido de recombinasa procarionta, que tiene como resultado la recombinación entre el primero y el segundo sitios de recombinación y la recombinación entre el tercero y el cuarto sitios de recombinación, en el que el primer polipéptido de recombinasa puede participar en la recombinación entre el primero y el segundo sitio de recombinación y el segundo polipéptido de recombinasa puede participar en la recombinación entre el tercero y el cuarto sitios de recombinación, la primera y la segunda recombinasa son una recombinasa de fago de *Mycobacterium tuberculosis* o una recombinasa de fago de *Mycobacterium smegmatis*, en el que la recombinasa se selecciona del grupo que consiste en recombinasa ϕ Rv1 y recombinasa Bxb1, siempre que el primer polipéptido de recombinasa y el segundo polipéptido de recombinasa sean diferentes.
6. El método de la reivindicación 5, que además comprende un quinto sitio de recombinación y un sexto sitio de recombinación y un tercer polipéptido de recombinasa, en el que el tercer polipéptido de recombinasa puede participar en la recombinación entre el quinto y el sexto sitios de recombinación, siempre que el tercer polipéptido de recombinasa sea diferente al primero y al segundo polipéptidos de recombinasa.
7. Un vector para la integración específica de sitio de una secuencia de polinucleótidos en el genoma de una célula eucariota aislada, comprendiendo dicho vector un polinucleótido de interés y un segundo sitio de recombinación *attB* o *attP*, en el que dicho segundo sitio de recombinación *attB* o *attP* comprende una secuencia de polinucleótidos que se recombina con un primer sitio de recombinación *attP* o *attB* o seudositio *attP* o seudositio *attB* en el genoma de dicha célula eucariota aislada, en el que dicho vector comprende además un polinucleótido que codifica una recombinasa del fago de *Mycobacterium tuberculosis* o una recombinasa del fago de *Mycobacterium smegmatis*, en el que la recombinasa se selecciona del grupo que consiste en una recombinasa ϕ Rv1 y una recombinasa Bxb1, con la condición de que cuando el primer sitio de recombinación es *attB* o seudo *attB*, el segundo sitio de recombinación es *attP* y cuando el primer sitio de recombinación es *attP* o seudo *attP*, el segundo sitio de recombinación es *attB*.
8. El vector de la reivindicación 7, en el que el polinucleótido de interés está unido operativamente a un promotor que participa en la expresión del polinucleótido en la célula eucariota.
9. Una secuencia de polinucleótidos aislada que comprende un ácido nucleico que tiene una identidad de al menos un 90 % con la secuencia de ácido nucleico SEC ID N°: 5 o SEC ID N° 9, en la que el polipéptido codificado por el ácido nucleico tiene actividad recombinasa.
10. La secuencia de polinucleótidos aislada de la reivindicación 9, en el que la secuencia de ácido nucleico es la SEC ID N° 5 o la SEC ID N° 9.

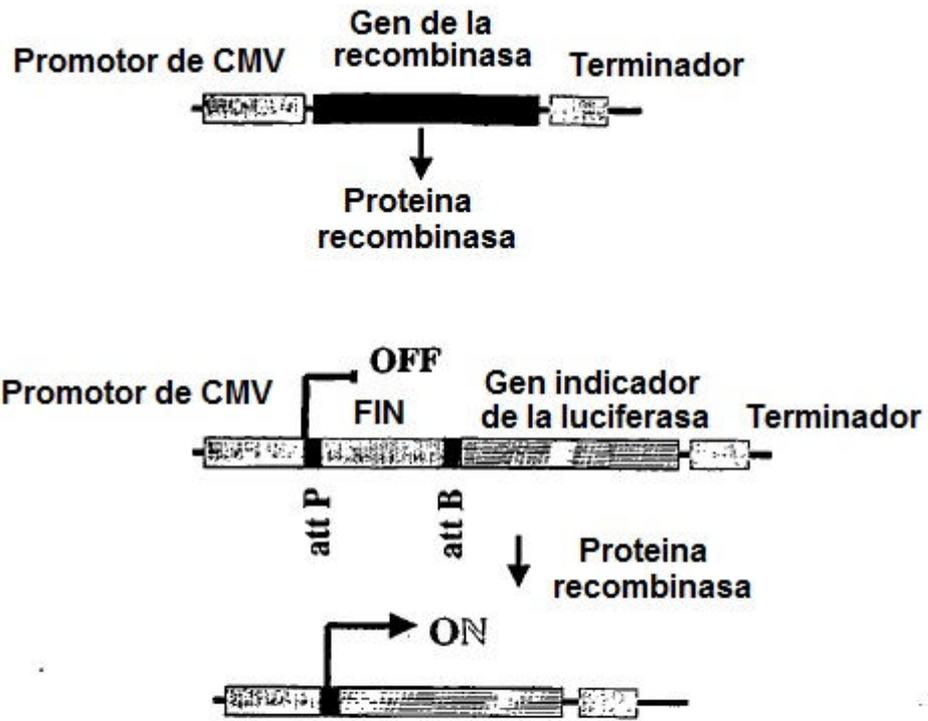


Figura 1

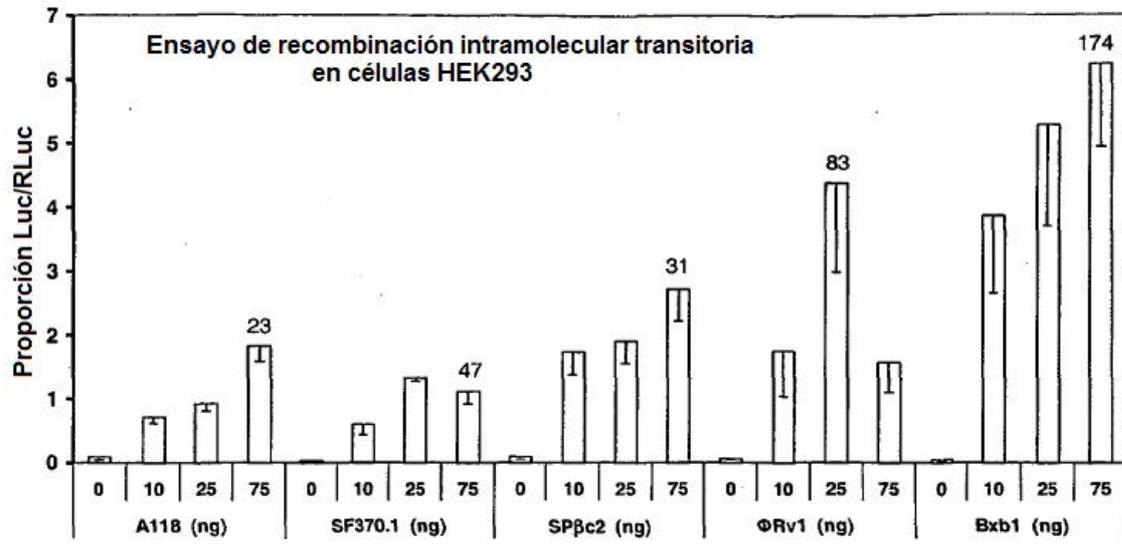


Figura 2

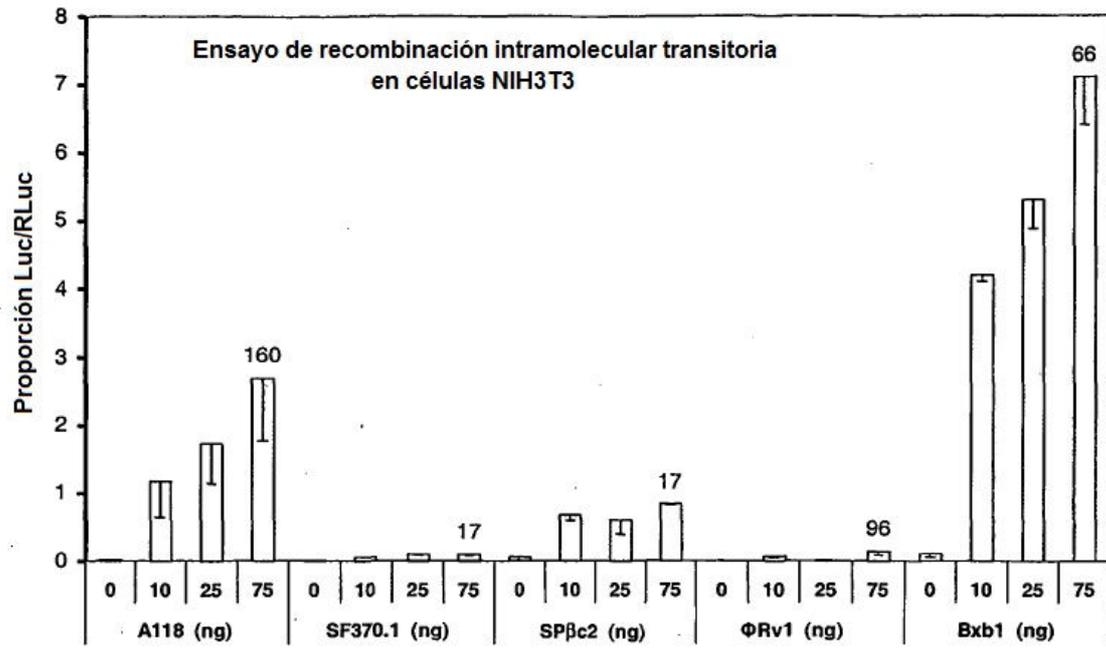


Figura 3

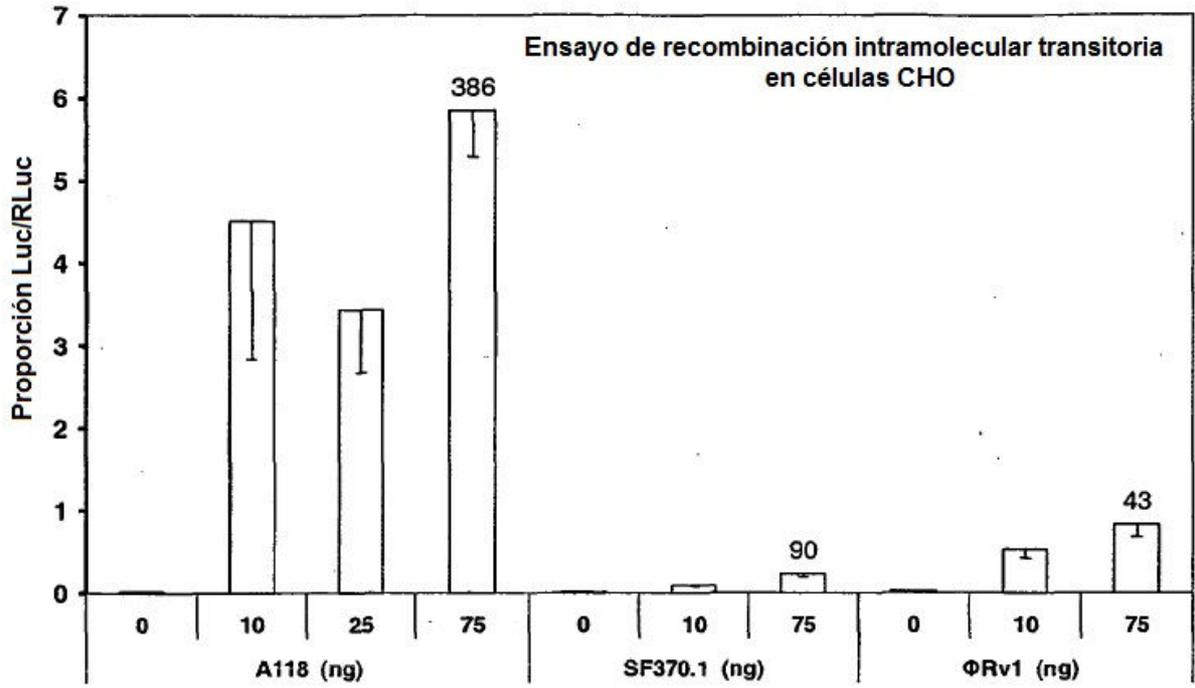


Figura 4

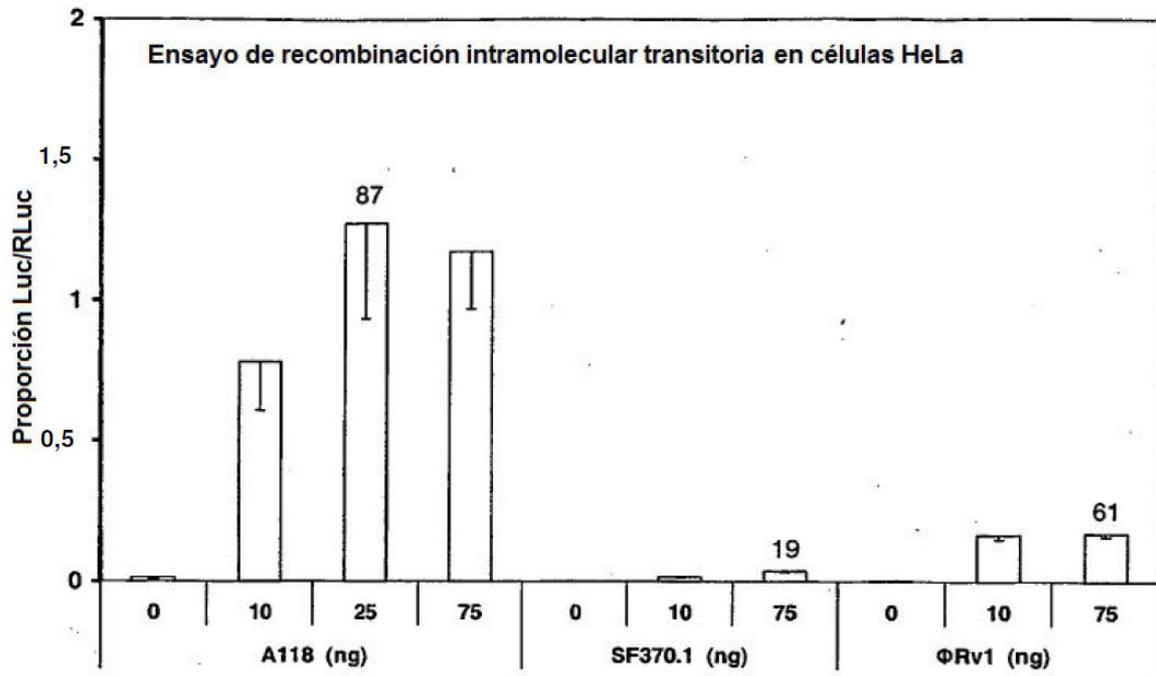


Figura 5

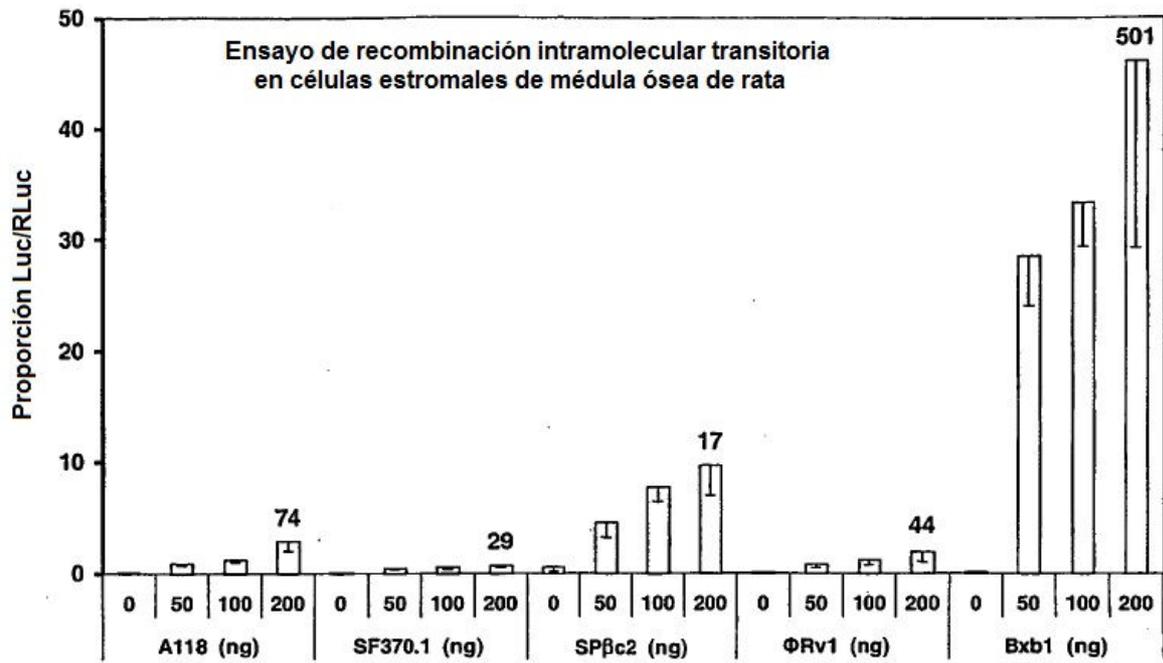


Figura 6

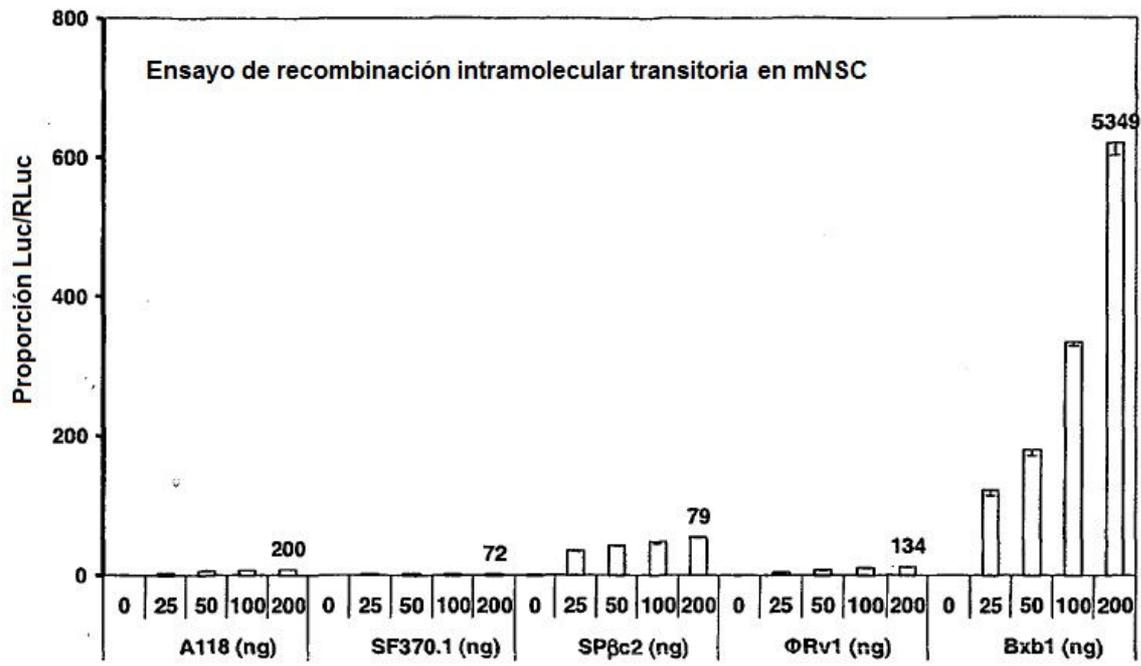


Figura 7

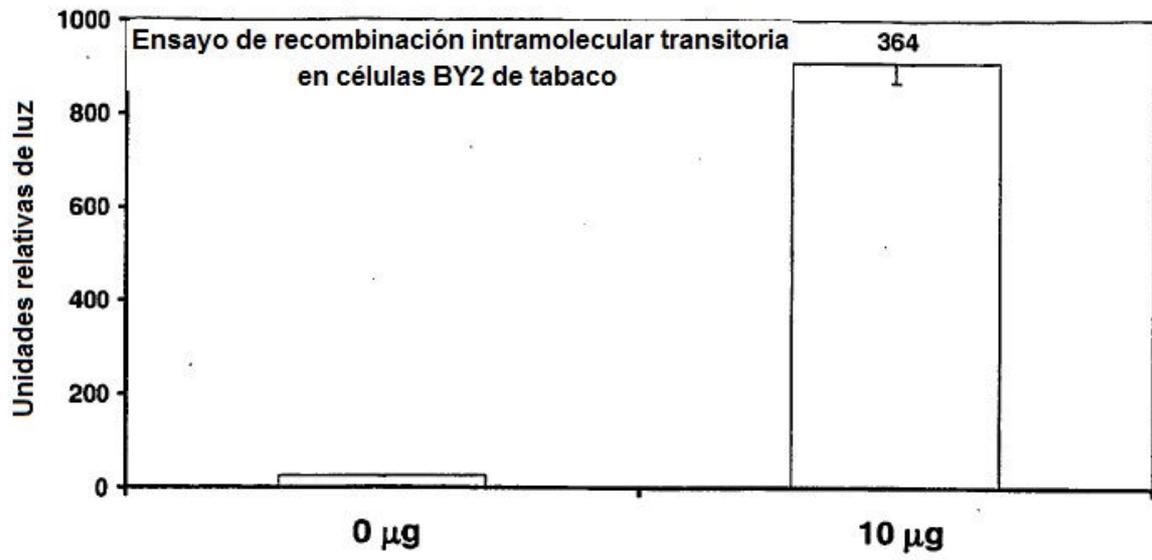


Figura 8

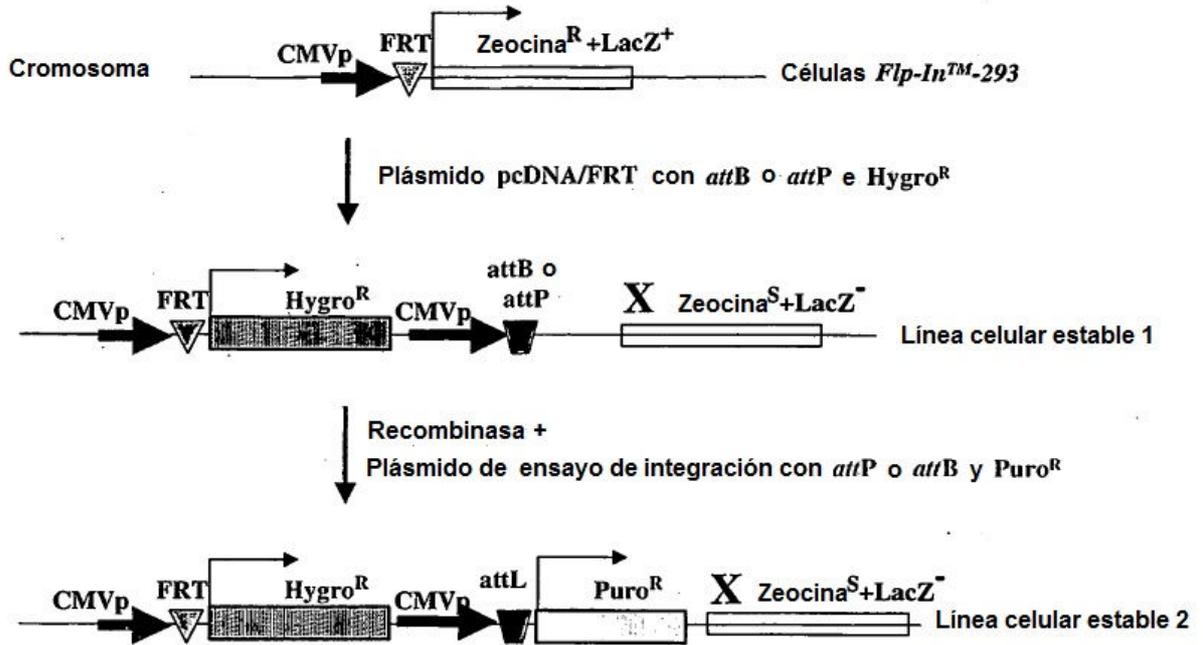


Figura 9

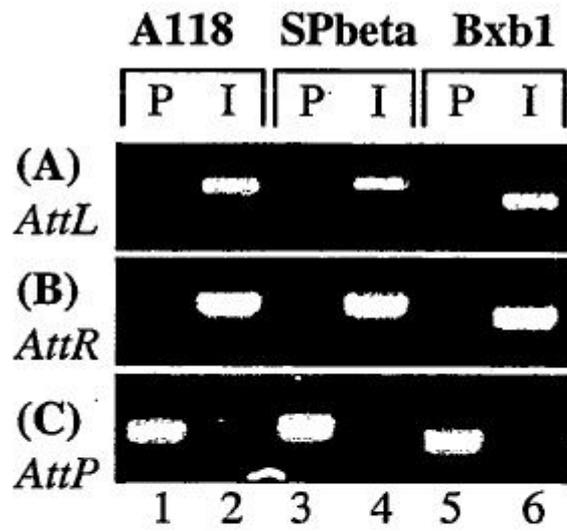
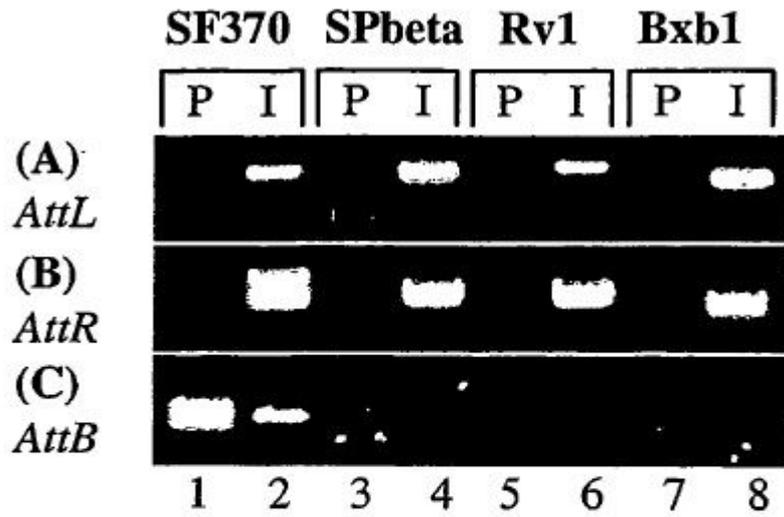


Figura 10

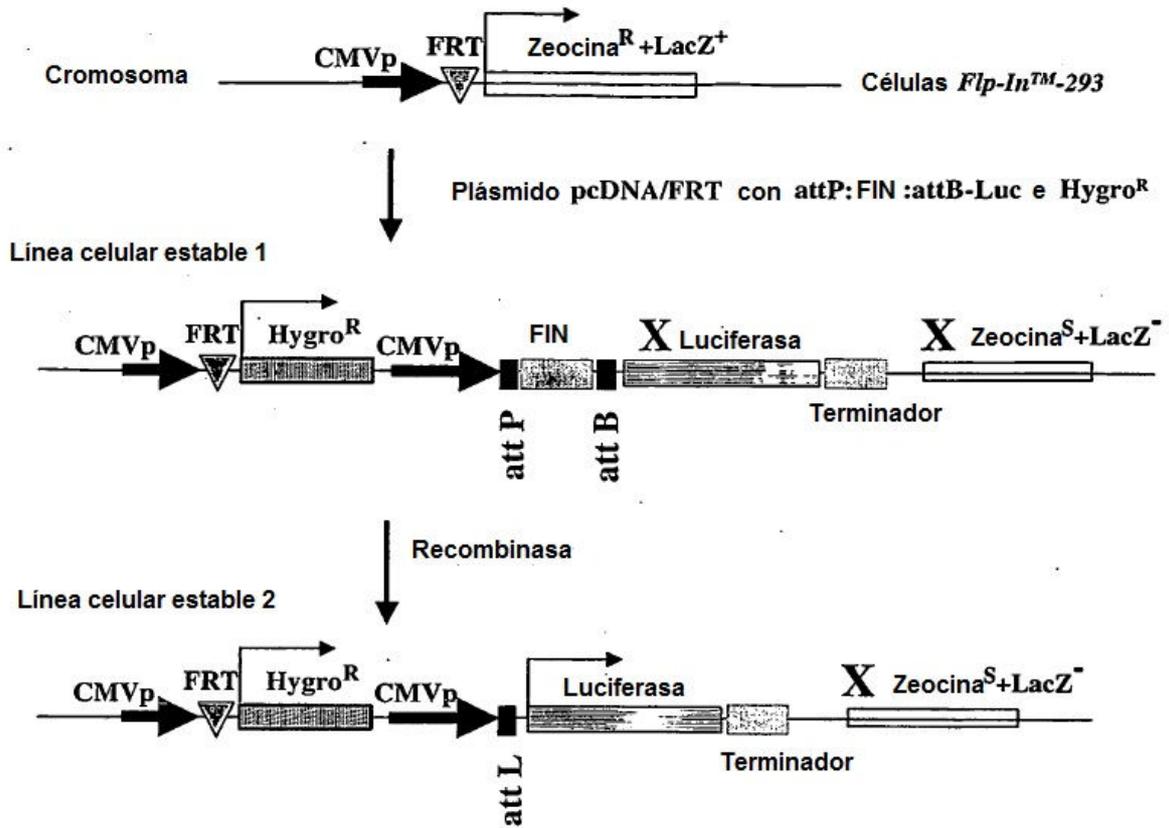


Figura 11

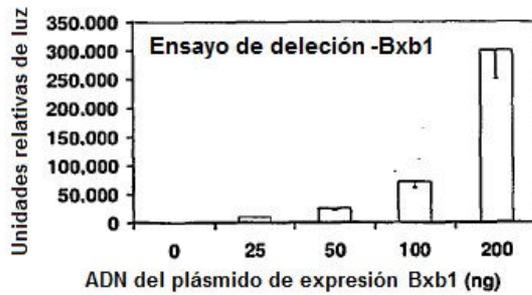
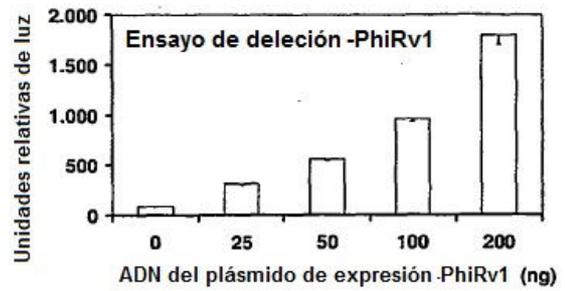
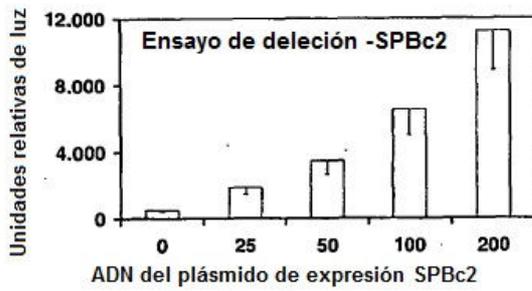
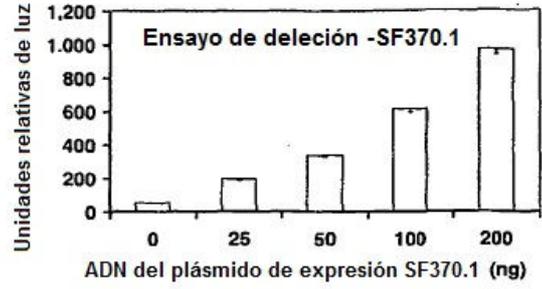
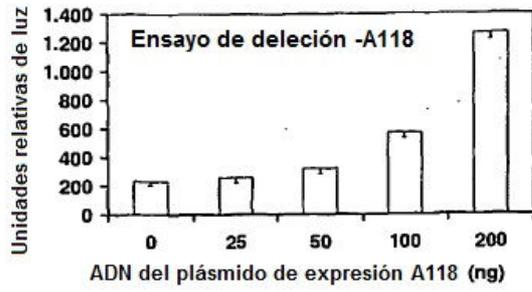


Figura 12

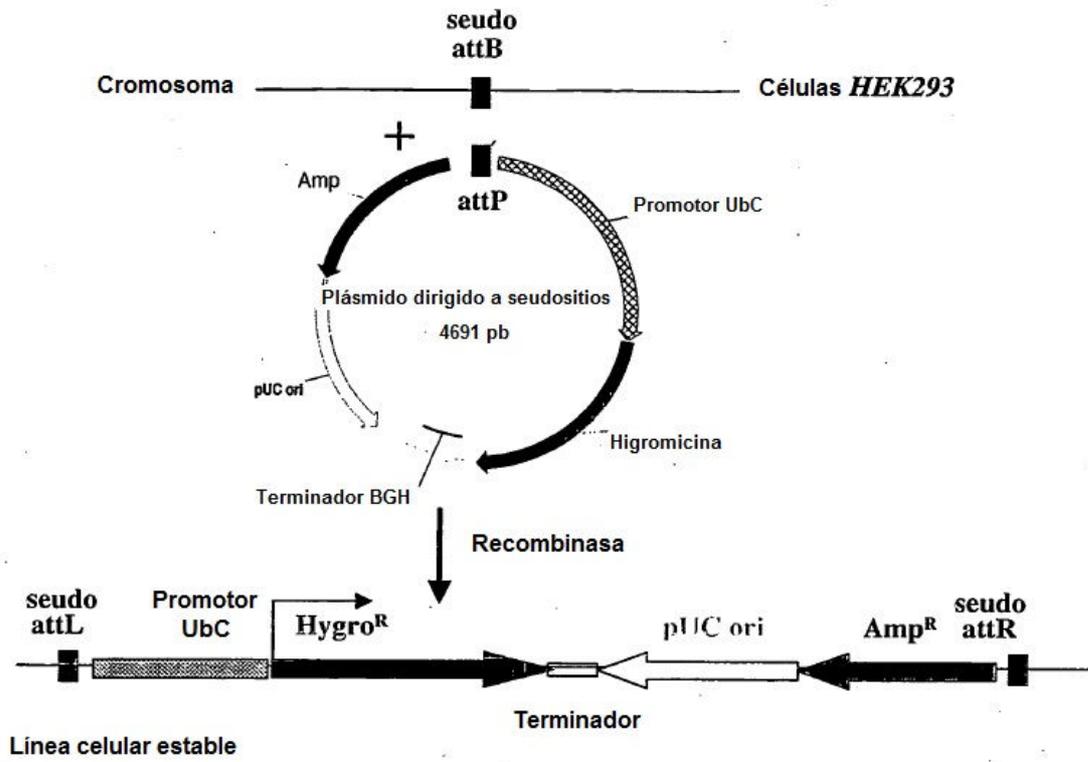


Figura 13

Sitio attP de SF370.1 , 99 pb

ACGAAAGGAGGTCGTGAAATGGATAAAAAAATACAGCGTTTTTCATGTACAACCTATACTAGTTGTAGTGCCTAAATAATGC'TTTTAAAACTTAAAAATA

X

G'TTTTGAAAAACTCTAGGCAGTTTCCTGAATCCCAAGCAGGCTTGTTCAAGGCTTACTATTTAGAGAAAATGGGTCTGACCTGGAGAGTCAGTATTTA

Seudositio attB en el cromosoma 10 humano

↓ **Recombinasa**

ACGAAAGGAGGTCGTGAAATGGATAAAAAAATACAGCGTTTTTCATGTACAACCTATACTATTTAGAGAAAATGGGTCTGACCTGGAGAGTCAGTATTTA

Seudositio attR tras la integración del plásmido diana

Sitio attP de SPβc2, 99 pb

ACGGCAGAGTAAGCTTCTTTTTTCGTTAGATATGTAGTAAGTATCTTAATATACAGCTTTATCTGTTTTTTAAGATACTTACTACTTTTTCTTAGTGGA

X

ATAAGCACAGGAACAACTCATAAGAGCCTGCAATGAGATCATCAGTGTCAAGCACTCATTTAGTGTCTTGGCATAACCCAAATGTTTCAGGAGAGATCT

Seudositio attB en el cromosoma 15 humano

↓ **Recombinasa**

ACGGCAGAGTAAGCTTCTTTTTTCGTTAGATATGTAGTAAGTATCTTAATATACAGCTTTATAGTGTCTTGGCATAACCCAAATGTTTCAGGAGAGATCT

Seudositio attR tras la integración del plásmido diana

Figura 14