

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 728 436**

51 Int. Cl.:

A61K 38/16 (2006.01)

C12N 9/22 (2006.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.04.2012 PCT/US2012/032426**

87 Fecha y número de publicación internacional: **11.10.2012 WO12138927**

96 Fecha de presentación y número de la solicitud europea: **05.04.2012 E 12725896 (0)**

97 Fecha y número de publicación de la concesión europea: **13.03.2019 EP 2694091**

54 Título: **Método para la generación de TALE-nucleasas compactas y usos de la mismas**

30 Prioridad:

05.04.2011 US 201161472065 P

13.06.2011 US 201161496454 P

20.06.2011 US 201161499043 P

20.06.2011 US 201161499047 P

09.09.2011 US 201161533123 P

09.09.2011 US 201161533098 P

22.12.2011 US 201161579544 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
24.10.2019

73 Titular/es:

**CELLECTIS (100.0%)
8, rue de la Croix Jarry
75013 Paris, FR**

72 Inventor/es:

**DUCHATEAU, PHILIPPE;
JULLERAT, ALEXANDRE;
VALTON, JULIEN;
BERTONATI, CLAUDIA;
EPINAT, JEAN-CHARLES;
SILVA, GEORGE, H. y
BEURDELEY, MARINE**

74 Agente/Representante:

ISERN JARA, Jorge

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 728 436 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

Método para la generación de TALE-nucleasas compactas y usos de la mismas

5 Campo de la invención

La presente invención describe un método para la generación de nucleasas de efector de tipo activador de transcripción (TALEN) compactas que pueden abordar y procesar de forma eficaz ADN bicatenario. Más específicamente, la presente invención describe un método para la creación de TALEN, que consiste en un único dominio de unión a ADN de TALE fusionado con al menos un dominio catalítico de modo que la entidad activa está compuesta de una única cadena polipeptídica para una vectorización simple y eficaz y no requiere dimerización para abordar una secuencia diana de ADN bicatenaria individual específica de interés y procesar el ADN cercano a dicha secuencia diana de ADN. La presente invención también se refiere a TALEN compactas, vectores, composiciones y kits usados para implementar el método.

15

Antecedentes de la invención

Los genomas de mamífero sufren constantemente diversos tipos de daños, de los que las roturas bicatenarias (DSB) se consideran los más peligrosos (Haber 2000). La reparación de las DSB pueden producirse mediante diversos mecanismos que pueden depender del contexto celular. La reparación mediante recombinación homóloga (HR) puede restaurar la secuencia original en la rotura. A causa de su estricta dependencia en la homología de secuencia amplia, este mecanismo se sugiere como activo principalmente durante las fases S y G2 del ciclo celular, donde las cromátidas hermanas están en cercana proximidad (Sonoda, Hohegger *et al.* 2006). La hibridación monocatenaria (SSA) es otro proceso dependiente de homología que puede reparar las DSB entre repeticiones directas y promueve de este modo las eliminaciones (Paques y Haber 1999). Finalmente, la unión de extremos no homólogos (NHEJ) del ADN es una ruta principal para la reparación de DSB que puede funcionar en todo el ciclo celular y no depende de la recombinación homóloga (Moore y Haber 1996; Haber 2008). La NHEJ parece comprender al menos dos componentes diferentes: (i) una ruta que consiste principalmente en la reunión directa de los extremos de DSB, y que depende de las proteínas XRCC4, Lig4 y Ku y; (ii) una ruta de NHEJ alternativa, que no depende de XRCC4, Lig4 y Ku, y es especialmente propensa a errores, provocando principalmente eliminaciones, produciéndose las uniones entre microhomologías Frank, Sekiguchi *et al.* 1998; Gao, Sun *et al.* 1998; Guirouilh-Barbat, Huck *et al.* 2004; Guirouilh-Barbat, Rass *et al.* 2007; Haber 2008; McVey y Lee 2008).

20

25

30

35

40

45

La dirección génica homóloga (HGT), descrita por primera vez hace más de 25 años (Hinnen, Hicks *et al.* 1978; Orr-Weaver, Szostak *et al.* 1981; Orr-Weaver, Szostak *et al.* 1983; Rothstein 1983), fue uno de los primeros métodos para la manipulación genómica racional y sigue siendo actualmente un criterio para la generación de células genomanipuladas o ratones knock-out (Capecchi 2001). No obstante, una eficacia inherentemente baja ha evitado que se use como protocolo rutinario en la mayoría de tipos celulares y organismos. Para abordar estas cuestiones, se ha propuesto una variedad amplia de estrategias racionales con la intención de conseguir más de un 1 % de modificaciones dirigidas. Muchos grupos se han centrado en potenciar la eficacia de HGT, habiendo llegado a ser evidentes dos disciplinas principales: (i) los llamados métodos de "optimización de matriz", que consisten esencialmente en modificar la estructura del vector dirigido para conseguir una eficacia máxima y; (ii) métodos que implican efectores adicionales para estimular HR, en general endonucleasas específicas de secuencia. El campo de la optimización de matriz ha cubierto una amplia gama de técnicas, con grados variables de éxito (Russell e Hirata 1998; Inoue, Dong *et al.* 2001; Hirata, Chamberlain *et al.* 2002; Taubes 2002; Gruenert, Bruscia *et al.* 2003; Sangiuolo, Scaldaferrri *et al.* 2008; Bedayat, Abdolmohamadi *et al.* 2010). La estimulación de HR mediante nucleasas, por otro lado, ha demostrado ser eficaz repetidamente (Paques y Duchateau 2007; Carroll 2008).

50

55

60

65

Para DSB inducidas por reactivos biológicos, por ejemplo, meganucleasas, ZFN y TALEN (véase a continuación), que escinden el ADN por hidrólisis de dos enlaces fosfodiéster, el ADN puede reunirse de una manera sin empalmes por religamiento simple de los extremos cohesivos. Como alternativa, pueden producirse inserciones o eliminaciones (indel) perjudiciales de diversos tamaños en las roturas, provocando finalmente una inactivación génica (Liang, Han *et al.* 1998; Lloyd, Plaisier *et al.* 2005; Doyon, McCammon *et al.* 2008; Perez, Wang *et al.* 2008; Santiago, Chan *et al.* 2008; Kim, Lee *et al.* 2009; Yang, Djukanovic *et al.* 2009). La naturaleza de este proceso, que no depende de recombinación específica de sitio u homóloga, da lugar a una tercera estrategia dirigida basada en mutagénesis inducida por endonucleasa. Esta estrategia, así como las aplicaciones relacionadas, puede ser más simple que las basadas en recombinación homóloga por que (a) no se necesita introducir una matriz de reparación y; (b) la eficacia será menos dependiente del tipo celular (en contraste con HR, NHEJ es probablemente activa en todo el ciclo celular (Delacote y Lopez 2008). La mutagénesis dirigida basada en NEHJ se ha usado para desencadenar la inactivación de genes individuales o incluso múltiples genes en líneas celulares inmortalizadas (Cost, Freyvert *et al.* 2010; Liu, Chan *et al.* 2010). Además, este método abre nuevas perspectivas para organismos en que los métodos de inactivación génica basada en HR clásicos han demostrado ser ineficaces, o al menos difíciles de establecer (Doyon, McCammon *et al.* 2008; Geurts, Cost *et al.* 2009; Shukla, Doyon *et al.* 2009; Yang, Djukanovic *et al.* 2009; Gao, Smith *et al.* 2010; Mashimo, Takizawa *et al.* 2010; Menoret, Iscache *et al.* 2010).

Durante los últimos 15 años, se ha documentado bien el uso de meganucleasas para inducir de forma satisfactoria la

dirección génica, partiendo de experimentos directos que implican I-SceI de tipo silvestre para un trabajo más refinado que implica enzimas completamente rediseñadas (Stoddard, Scharenberg *et al.* 2007; Galetto, Duchateau *et al.* 2009; Marcaida, Munoz *et al.* 2010; Arnould, Delenda *et al.* 2011). Las meganucleasas, también llamadas endonucleasas de asentamiento (HE), pueden dividirse en cinco familias basándose en la secuencia y los motivos estructurales: 5 LAGLIDADG, GIY-YIG, HNH, secuencia His-Cys y PD-(D/E)XK (Stoddard 2005; Zhao, Bonocora *et al.* 2007). Hay datos estructurales disponible para al menos un miembro de cada familia. La familia mejor estudiada es la de las proteínas LAGLIDADG, con un conjunto considerable de trabajo bioquímico, genético y estructural que ha establecido que estas endonucleasas podrían usarse como herramientas moleculares (Stoddard, Scharenberg *et al.* 2007; Arnould, Delenda *et al.* 2011). Las proteínas que son miembro están compuestas de dominios que adoptan un plegamiento $\alpha\beta\alpha\beta\alpha$ similar, comprendiendo el motivo LAGLIDADG la región terminal de la primera hélice y contribuyendo no solamente a un centro catalítico de dos partes, sino también formando la interacción de subunidad central/subunidad (Stoddard 2005). Dos de estos dominios α/β se ensamblan para formar la proteína funcional, creando las hebras β en cada uno una región de unión a ADN con forma de silla de montar. La separación espacial del centro catalítico con regiones que interactúan directamente con el ADN ha permitido un rediseño de la especificidad 15 (Seligman, Chisholm *et al.* 2002; Sussman, Chadsey *et al.* 2004; Arnould, Chames *et al.* 2006; Doyon, Pattanayak *et al.* 2006; Rosen, Morrison *et al.* 2006; Smith, Grizot *et al.* 2006; Arnould, Perez *et al.* 2007). Además, mientras que todas las proteínas LAGLIDADG conocidas analizadas hasta la fecha actúan como "cleavasas" que cortan ambas hebras de ADN diana, se ha hecho progreso reciente en la generación de "meganickasas" que escinden únicamente una hebra (Niu, Tenney *et al.* 2008; McConnell Smith, Takeuchi *et al.* 2009). Dichas enzimas pueden proporcionar, en principio, niveles similares de HR inducida diana con una minimización en la frecuencia de NHEJ. 20

Aunque numerosos esfuerzos de diseño se han centrado en HE LAGLIDADG, los miembros de otras dos familias, GIY-YIG y HNH, son de particular interés. Los estudios bioquímicos estructurales han establecido que en ambas familias, las proteínas que son miembros pueden adoptar un plegamiento de dos partes con distintos dominios 25 funcionales: (1) un dominio catalítico responsable principalmente de la escisión del ADN y; (2) un dominio de unión a ADN para proporcionar especificidad diana (Stoddard 2005; Marcaida, Munoz *et al.* 2010). Las HE GIY-YIG relacionadas I-TevI e I-BmI se han explotado para demostrar la capacidad de intercambio de la región de unión a ADN para estas enzimas (Liu, Derbyshire *et al.* 2006). El análisis de la HE I-BasI reveló que aunque el dominio catalítico del extremo N pertenece a la familia HNH, la región de unión a ADN en el extremo C se parece al motivo de repetición de endonucleasa codificado en intrón (IENR1) encontrado en endonucleasas de la familia GIY-YIG (Landthaler y Shub 30 2003). La cabeza catalítica de I-BasI tiene similitud de secuencia con las de las HE HNH I-HmI, I-HmII e I-TwoI, todas las cuales funcionan como nickasas específicas de hebra (Landthaler, Begley *et al.* 2002; Landthaler y Shub 2003; Landthaler, Lau *et al.* 2004; Shen, Landthaler *et al.* 2004; Landthaler, Shen *et al.* 2006).

Aunque las familias anteriores de proteínas contienen nucleasas específicas de secuencia, el motivo HNH también se ha identificado en nucleasas no específicas tales como colicinas de *E. coli* (por ejemplo, ColE9 y ColE7), EndA de *S. pneumoniae*, NucA de *Anabaena* y CAD (Midon, Schafer *et al.* 2011). Además de tener el motivo HNH, varias de estas nucleasas contienen el motivo DRGH distintivo y comparten homología estructural con elementos centrales que forman el motivo de sitio activo de dedo $\beta\beta\alpha$ -Me. Los estudios mutacionales de los restos en los motivos HNH/DRGH han 40 confirmado su función en la actividad de escisión de ácido nucleico (Ku, Liu *et al.* 2002; Doudeva, Huang *et al.* 2006; Eastberg, Eklund *et al.* 2007; Huang y Yuan 2007). Además, la afinidad de unión a ADN y la preferencia de secuencia para ColE7 podría alterarse de forma eficaz (Wang, Wright *et al.* 2009). Dichos estudios detallados ilustran el potencial en el rediseño de nucleasas no específicas con fines de dirección.

Las nucleasas de dedos de cinc (ZFN), generadas por fusión de los dominios de unión a ADN basados en dedos de cinc a un dominio catalítico independiente mediante un conector flexible (Kim, Cha *et al.* 1996; Smith, Berg *et al.* 1999; Smith, Bibikova *et al.* 2000), representa otro tipo de nucleasa genomanipulada habitualmente usada para estimular la dirección génica. Las ZFN arquetípicas se basan en el dominio catalítico de la enzima de restricción de tipo IIS FokI y se han usado satisfactoriamente para inducir la corrección de genes, la inserción de genes y la eliminación de genes. Los dominios de unión a ADN basados en dedos de cinc están hechos de las cadenas de 3 o 4 dedos de cinc individuales, que reconoce en cada una un triplete de ADN (Pabo, Peisach *et al.* 2001). En teoría, una de las ventajas principales de las ZFN es que son fáciles de diseñar, usando ensamblaje combinatorio de dedos de cinc preexistentes con patrones de reconocimiento conocidos (Choo y Klug 1994; Choo y Klug 1994; Kim, Lee *et al.* 2009). Sin embargo, un examen cercano de estructuras de alta resolución muestra que hay realmente comunicaciones entre las unidades 55 (Elrod-Erickson, Rould *et al.* 1996) y se han usado varios métodos para ensamblar proteínas ZF eligiendo dedos de cinc individuales de una manera dependiente del contexto (Greisman y Pabo 1997; Isalan y Choo 2001; Maeder, Thibodeau-Beganny *et al.* 2008; Ramirez, Foley *et al.* 2008) para conseguir mejores tasas de éxito y reactivos de mejor calidad.

Recientemente, se ha descrito una nueva clase de nucleasa química usando un dominio catalítico de FokI (Christian, Cermak *et al.* 2010; Li, Huang *et al.* 2011). El dominio de unión a ADN de estas nucleasas se obtiene de efectores de tipo activador de transcripción (TALE), una familia de proteínas usada en el proceso de infección por patógenos vegetales del género *Xanthomonas*. En estos dominios de unión a ADN, la especificidad de secuencia está dirigida por una serie de repeticiones de 33-35 aminoácidos, que difieren esencialmente en dos posiciones (Boch, Scholze *et al.* 2009; Moscou y Bogdanov 2009). Cada par de bases de la diana de ADN está en contacto con una única repetición, produciéndose la especificidad por los dos aminoácidos variantes de la repetición (el llamado dipéptido 65

variable de repetición, RVD). La modularidad aparente de estos dominios de unión a ADN se ha confirmado a un cierto grado por ensamblaje modular de la proteína derivada de TALE diseñada con nuevas especificidades (Boch, Scholze *et al.* 2009; Moscou y Bogdanove 2009). Sin embargo, aún no puede descartarse un determinado nivel de dependencia de los patrones de repetición individual/reconocimiento de bases, como se observó para las proteínas de dedos de cinc (véase anteriormente). Además, se ha demostrado que los efectores TAL naturales pueden dimerizarse (Gurlebeck, Szurek *et al.* 2005) y la manera en que esto afectaría a una nucleasa derivada de TALE "basada en dimerización" actualmente es desconocida.

La disposición funcional de un TALE-nucleasa (TALEN) basada en FokI es esencialmente el de una ZFN, estando remplazado el dominio de unión a ADN de dedos de cinc por el dominio de TALE (Christian, Cermak *et al.* 2010; Li, Huang *et al.* 2011). Por tanto, la escisión de ADN por una TALEN requiere dos regiones de reconocimiento de ADN que flanquean una región central inespecífica. Esta región de ADN "espaciadora" central es esencial para promover la catálisis por el dominio catalítico de FokI de dimerización, y se ha establecido un amplio esfuerzo en la optimización de la distancia entre los sitios de unión a ADN (Christian, Cermak *et al.* 2010; Miller, Tan *et al.* 2011). La longitud del espaciador se ha variado de 14 a 30 pares de bases, siendo la eficacia en la escisión del ADN interdependiente de la longitud del espaciador, así como la construcción del armazón de TALE (es decir, la naturaleza de la construcción de fusión usada). Aún se desconoce si diferencias en la región repetida (es decir, tipo de RVD y número usado) tiene un impacto sobre los requisitos del "espaciador" de ADN o sobre la eficacia de la escisión del ADN por TALEN. No obstante, las TALE-nucleasas han demostrado ser activas a diversos grados en ensayos basados en células en levaduras, células de mamífero y plantas (Christian, Cermak *et al.* 2010; Li, Huang *et al.* 2011; Mahfouz, Li *et al.* 2011; Miller, Tan *et al.* 2011).

Las arquitecturas de TALEN clásicas basadas en la fusión de un dominio de unión de TALE al dominio catalítico de FokI se han descrito por Miller *et al.* ("A TALE nuclease architecture for efficient genome editing"; *nature biotechnology*, volumen 29, n.º 2, páginas 143-148; 2011).

Las endonucleasas de asentamiento GIY-YIG tales como I-TevI se describen en Liu *et al.* ("Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions"; *Nucleic Acids Research*, volumen 34, n.º 6, páginas 1755-1764; 2006).

El uso de la propiedad modular de la endonucleasa Bmrl y la especificidad de unión a ADN de C.BclI para generar una endonucleasa combinatoria se ha descrito en Siu-hong *et al.* ("Catalytic domain of restriction endonuclease Bmrl as a cleavage module for engineering endonucleases with novel substrate specificities"; *Nucleic Acids Research*, volumen 35, n.º 18, páginas 6238-6248; 2007).

Los autores de la invención han desarrollado un nuevo tipo de TALEN que puede genomanipularse para que reconozca y procese específicamente un ADN diana de forma eficaz. Estas "TALEN compactas" (cTALEN) novedosas no requieren dimerización para la actividad de procesamiento de ADN, aliviando de este modo la necesidad de sitios diana "dobles" con "espaciadores" de ADN intermedios. Además, la invención permite general varios tipos distintos de enzimas que pueden potenciar diferentes rutas de reparación de ADN (HR frente a NHEJ).

Breve resumen de la invención

La presente invención describe un método para generar nucleasas efectoras de tipo activador de transcripción (TALEN) compactas compuestas de una única cadena polipeptídica que no requieren dimerización para abordar una secuencia diana de ADN bicatenario única específica de interés y procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés. La presente invención también describe la creación de proteínas de fusión de un único polipéptido funcionales para la vectorización simple y eficaz. En otro aspecto, la presente invención describe TALEN compactas que comprenden al menos un dominio potenciador, en las que dicho dominio potenciador potencia la eficacia de procesamiento del ADN de dichas TALEN compactas cercano a dicha secuencia diana de ADN bicatenario única de interés. La presente invención también se refiere a TALEN compactas, vectores, composiciones y kits usados para implementar el método. La presente invención también se refiere a métodos para el uso de dichas TALEN compactas de acuerdo con la invención para diversas aplicaciones que varían desde la escisión de ADN diana hasta la regulación génica diana. Los métodos de acuerdo con la presente invención pueden usarse en diversos campos que varían desde la creación de organismos transgénicos hasta el tratamiento de enfermedades genéticas.

Breve descripción de las figuras

Además de las características precedentes, la invención comprende además otras características que surgirán de la siguiente descripción, así como de los dibujos adjuntos. Una apreciación más completa de la invención y muchas de las ventajas consiguientes de la misma se obtendrá fácilmente según llegue a entenderse mejor la misma por referencia a las siguientes figuras junto con la descripción detallada a continuación.

Figura 1: Estrategias de dirección génica inducida por endonucleasa. Tras la escisión, los mecanismos de reparación del ADN pueden producir uno de varios resultados. (A) Cuando se aborda una rotura bicatenaria entre dos repeticiones directas, la HR puede producir la eliminación de una repetición junto con la secuencia intermedia. La inserción génica (B) o corrección (C) pueden conseguirse por la introducción de una matriz de reparación de ADN que contiene secuencias homólogas a la secuencia endógena que rodea la rotura del ADN. Pueden corregirse

mutaciones en o distales a la rotura, disminuyendo la frecuencia de corrección según aumenta la distancia. (D) La reparación incorrecta de extremos de ADN por NHEJ propensa a errores puede provocar inserciones o eliminaciones de diversos tamaños, que dan lugar a inactivación génica.

5 Figura 2: Secuencias de ADN diana reconocidas por I-Crel e I-TevI. C1234 (SEQ ID NO: 3) representa la secuencia de ADN natural parcialmente simétrica reconocida y escindida por la meganucleasa I-Crel de tipo silvestre. C1221 (SEQ ID NO: 2) representa una secuencia de ADN palindrómica artificial, derivada de C1234 (SEQ ID NO: 3), también reconocida y escindida por I-Crel (SEQ ID NO: 1). Los nucleótidos están numerados hacia fuera (-/+)

10 desde el centro de la diana. La escisión del ADN se produce en cualquier lado de la secuencia subrayada para generar extremos salientes 3' de 4 nucleótidos. Para meganucleasas basadas en I-Crel, la naturaleza de los nucleótidos en las posiciones -2 a +2 pueden interferir potencialmente con la actividad de escisión de la proteína. Tev (SEQ ID NO: 105) representa la secuencia de ADN asimétrica reconocida y escindida por la meganucleasa I-TevI de tipo silvestre. La numeración de nucleótidos es relativa al sitio de inserción del intrón de la secuencia diana natural. La escisión por I-TevI se produce en cualquier lado de la secuencia subrayada para generar extremos

15 salientes 3' de 2 nucleótidos.

Figura 3: Secuencias de los ADN diana reconocidos por TALEN y construcciones de TALEN compactas. Los ADN diana para las TALEN compactas genomanipuladas cTN-Avr y cTN-Pth se basan en las secuencias asimétricas de origen natural AvrBs3 (19 pb) [en los armazones proteínicos basales bT1-Avr (SEQ ID NO: 136) y bT2-Avr (SEQ ID NO: 137)] y PthXo1 (25 pb) [en los armazones proteínicos basales bT1-Pth (SEQ ID NO: 138) y bT2-Pth (SEQ ID NO: 139)], respectivamente. Para cada secuencia, los nucleótidos están numerados hacia fuera (-/+)

20 para la T de anclaje (posición -1). Las secuencias mostradas se ponen en contacto directo con la proteína para proporcionar especificidad diana. Los dipéptidos variables de repetición (RVD) de tipo silvestre corresponden a los dipéptidos encontrados en las repeticiones de las proteínas efectoras de origen natural dirigidas a cada secuencia. Los RVD cifrados se basan en el subconjunto de dipéptidos/pares de nucleótidos enumerados. Los RVD artificiales se obtienen por lectura directa de la secuencia de ADN subyacente usando el código RVD cifrado (SEQ ID NO: 245 a 249).

Figura 4: Esquema de configuraciones de fusión de meganucleasa. Las construcciones de fusión se optimizan para abordar o superar distintos problemas. (A) La adición de dos dominios catalíticos a una meganucleasa activa no solamente puede potenciar la actividad de escisión (por ejemplo, tres posibilidades de lograr la escisión de ADN por evento de unión), sino que también puede promover alteraciones de secuencia por NHEJ propensa a errores ya que se escinden pequeñas secciones de ADN por cada par de eventos de escisión. (B) Cuando el rediseño de especificidad imposibilita el mantenimiento de la actividad de escisión de la meganucleasa, los dominios catalíticos adheridos proporcionan la función de escisión de hebra necesaria. (C) y (D) representan los casos de (A) y (B), respectivamente, cuando únicamente un dominio catalítico se tolera por proteína de fusión (por ejemplo, como una fusión en el extremo N o C o en el contexto de una molécula monocatenaria). En todos los casos, el dominio catalítico ideado puede ser una cleavasa (capacidad de escindir ambas hebras del ADN) o una nickasa (escisión de únicamente una única hebra de ADN) dependiendo de la aplicación. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación. Los componentes de las proteínas de fusión se enumeran en la leyenda.

30

35

40

Figura 5: Esquema de configuraciones de cTALEN. Se diseñan TALEN compactas para aliviar la necesidad de múltiples restos proteínicos independientes cuando se aborda un evento de escisión de ADN. De forma importante, la región "espaciadora" necesaria y los sitios diana dobles esenciales para la función de TALEN clásicas actuales son innecesarios. Además, como el dominio catalítico no requiere contactos de ADN específicos, no hay restricciones sobre las regiones que rodean el dominio de unión a ADN de TALE central. (A) Construcción de fusión del extremo N para promover HR mediante una ruta de reparación convencional (dominio cleavasa) o conservativa (dominio nickasa). (B) Construcción de fusión en el extremo C con propiedades como en (A). (C) La adhesión de dos dominios catalíticos a ambos extremos de la TALE permite la escisión doble con potenciación en NHEJ. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación. Los componentes de las proteínas de fusión se enumeran en la leyenda.

45

50

Figura 6: Esquema de configuraciones de cTALEN potenciadas. Las TALEN compactas pueden potenciarse mediante la adición de un dominio para promover actividades existentes o alternativas. Como cada extremo del dominio de unión a ADN de TALE es susceptible a fusión, el orden (en el extremo N frente al C) de adición de los dominios catalíticos y potenciadores puede variar con la aplicación. (A) Una cTALEN convencional con un dominio potenciador en el extremo C. (B) El dominio potenciador está fusionado a la cTALEN mediante el extremo N del dominio catalítico. Dicha configuración puede usarse para ayudar y/o anclar el dominio catalítico cerca del ADN para aumentar la actividad de escisión. (C) El dominio potenciador está emparedado entre el dominio catalítico y el dominio de unión a ADN de TALE. El dominio potenciador puede promover la comunicación entre los dominios flanqueantes (es decir, para ayudar en la catálisis y/o unión a ADN) o puede usarse para superar el nucleótido T necesario en la posición -1 de todas las dianas basadas en TALE. (D) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo N de la proteína TALE natural. (E) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo C de la proteína TALE natural. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación. Los componentes de las proteínas de

55

60

65

fusión se enumeran en la leyenda.

Figura 7: Esquema de configuraciones de cTALEN en trans. Las TALEN compactas pueden combinarse con dominios potenciadores auxiliares para promover actividades alternativas. Los dominios auxiliares proporcionan una función adicional que no es esencial para la actividad de cTALEN. (A) Una cTALEN convencional con un dominio catalítico de nickasa en el extremo N se convierte en una "cleavasa" mediante la adición separada del dominio auxiliar. (B) En algunos casos, la necesidad de abordar la especificidad del dominio auxiliar es necesaria. Dicha configuración puede conseguirse mediante una fusión de TALE y puede usarse para ayudar y/o anclar el dominio auxiliar cerca del ADN para aumentar la actividad de la cTALEN. (C) El dominio auxiliar dirigido se proporciona antes o después de la cTALEN para realizar una tarea independiente. La comunicación entre las proteínas de fusión no es necesaria. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación. Los componentes de las proteínas de fusión se enumeran en la leyenda.

Figura 8: Esquema de escisión del ADN, religamiento *in vivo* y otras rutas de reparación. En células, la escisión por endonucleasas de corte infrecuente peptídico habitualmente produce una rotura bicatenaria del ADN (DSB) con extremos cohesivos. Por ejemplo, las meganucleasas de la familia LAGLIDADG, tales como I-SceI y I-CreI, producen DSB con salientes 3'. Estos extremos cohesivos pueden religarse *in vivo* por NHEJ, produciendo una reparación sin empalmes, y la restauración de la secuencia diana escindible, que a su vez puede procesarse de nuevo por la misma endonucleasa. Por tanto, puede tener lugar una serie de ciclos infructuosos de eventos de escisión y religamiento. La NHEJ imprecisa o recombinación homóloga puede alterar o eliminar el sitio de escisión, provocando la salida del ciclo; esto también puede aplicarse a las TALEN compactas y las TALEN compactas potenciadas (A). Otras dos vías también pueden detener el proceso: (i) Puede producirse pérdida del cromosoma como consecuencia del fracaso en la reparación de DSB; (ii) una pérdida de nucleasa (degradación, dilución, división celular, etc.). B-E: Consecuencias de escisión de enlaces fosfodiéster adicionales. La adición de una única actividad nickasa (B) o de dos actividades nickasa que afectan a la misma hebra (C) provocaría una interrupción monocatenaria, y suprimiría los extremos cohesivos, que a su vez podrían afectar el espectro de eventos. La adición de dos actividades nickasa que afectan a hebras opuestas (D) o de una nueva actividad cleavasa que genera una segunda DSB (E) provocaría una interrupción bicatenaria; como consecuencia, ya no es posible un religamiento perfecto, y podría estimularse uno o varios resultados de reparación alternativos. La figura actual no hace suposiciones respecto a las frecuencias relativas de estos resultados alternativos (NHEJ imprecisa, recombinación homóloga, otras, etc.). Los triángulos rellenos representan hidrólisis de enlaces fosfodiéster.

Figura 9: Actividad de TALE AvrBs3::TevI en levadura (37 °C). El control negativo consiste en una TALEN sin ningún RVD. n.d. indica actividad no detectable, + indica una actividad por encima de 0,3 en ensayo de levadura y +++ indica una actividad por encima de 0,7 en ensayo de levadura (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 10: Actividad de TALE AvrBs3::TevI en células de mamífero. (Ensayo extracromosómico en CHO-K1). pCLS8993 (SEQ ID NO: 194) está representado por una barra negra y pCLS8994 (SEQ ID NO: 195) está representado por una barra gris oscura. Control negativo (vector vacío) por una barra blanca y control positivo (meganucleasa I-SceI) por una barra gris claro. Los datos están normalizados respecto al control positivo.

Figura 11: Actividad de TALE AvrBs3::NucA en levadura (37 °C). El control negativo es una diana que carece de un sitio de reconocimiento (ctrl. neg. : SEQ ID NO: 228). Compacta es una diana que tiene únicamente un sitio de reconocimiento (SEQ ID NO: 224). n.d. indica actividad no detectable, + indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; ++, actividad por encima de 0,5 en ensayo de levadura a 37 °C y +++ actividad por encima de 0,7 en ensayo de levadura a 37 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 12: Actividad de TALE AvrBs3:: ColE7 en levadura (37 °C). El control negativo es una diana que carece de un sitio de reconocimiento (ctrl. neg. : SEQ ID NO: 228). Compacta es una diana que tiene únicamente un sitio de reconocimiento (SEQ ID NO: 224). n.d. indica actividad no detectable, + indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; ++, actividad por encima de 0,5 en ensayo de levadura a 37 °C y +++ actividad por encima de 0,7 en ensayo de levadura a 37 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 13: Esquema y detalles de una TALEN compacta prototípica TALE::Crel. Esta clase de TALEN compacta está dirigida a una secuencia de reconocimiento de dos partes compuesta del sitio de unión a ADN de TALE proximal a un sitio diana de meganucleasa. El sitio de meganucleasa TCRB02-A genomanipulado se muestra junto con detalles del RVD y la secuencias de ADN reconocidas por el resto de TALE. Se presenta una región del gen del receptor B de célula de linfocitos T, que resalta la disposición endógena del sitio diana híbrido de TALEN compacta basado en TALE::Crel.

Figura 14: Actividad de construcciones basadas en TALE::scTB2aD01 en levadura (30 °C). Se muestran las disposiciones de las diversas dianas híbridas, empezando (5') con la región reconocida por el dominio de unión a

ADN de TALE en mayúsculas, la región espaciadora inespecífica en minúsculas y el sitio diana de meganucleasa en mayúscula subrayados. La actividad en levadura se ilustra para construcciones representativas selectivas. n.d. indica actividad no detectable, + indica una actividad por encima de 0,3 en ensayo de levadura a 30 °C; ++, actividad por encima de 0,5 en ensayo de levadura a 30 °C y +++ actividad por encima de 0,7 en ensayo de levadura a 30 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 15: Transferencia de Western de construcciones basadas en TALE::scTB2aD01. Las construcciones se expresaron en células HEK293 y se prepararon extractos de proteínas totales 48 horas después de la transfección. La proteína se detectó usando un anticuerpo policlonal anti-I-Crel.

Figura 16: Toxicidad de construcciones basadas en TALE::scTB2aD01 en células CHOK1. La citotoxicidad se basa en niveles detectables de GFP expresada en células vivas, en el día 1 frente al día 6, respecto a un control convencional (transfección de plásmido vacío).

Figura 17: Actividad de NHEJ de construcciones basadas en TALE::scTB2aD01 en células HEK293. Se usa un análisis basado en PCR después de la transfección de ADN genómico para evaluar la actividad *in vivo*. La escisión de secuencias de ADN emparejadas incorrectamente por la endonucleasa de T7 es indicativa de eventos de NHEJ resultantes de la actividad de la cTALEN o meganucleasa en el locus diana.

Figura 18: Actividad de T_{evl}::TALE-AvrBs3 +/- TALE-RagT2-R::T_{evl} en levadura (37 °C). El control negativo es una diana que carece de un sitio de reconocimiento (ctrl. neg. : SEQ ID NO: 228). n.d. indica actividad no detectable, + indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; ++, actividad por encima de 0,5 en ensayo de levadura a 37 °C y +++ actividad por encima de 0,7 en ensayo de levadura a 37 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 19: Actividad de TALE::SnaseSTAAU en levadura (37 °C). El control negativo es una diana que carece de un sitio de reconocimiento. Compacta es una diana que tiene únicamente un sitio de reconocimiento (SEQ ID NO: 224). n.d. indica actividad no detectable a 37 °C, +/- indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; + indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; ++ indica una actividad por encima de 0,5 en ensayo de levadura a 37 °C; +++ indica una actividad por encima de 0,75 en ensayo de levadura a 37 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Figura 20: Actividad de TALE::CoIE7 con diversos conectores polipeptídicos en levadura (37 °C). Compacta es una diana que tiene únicamente un sitio de reconocimiento (SEQ ID NO: 224). n.d. indica actividad no detectable a 37 °C, + indica una actividad por encima de 0,3 en ensayo de levadura a 37 °C; ++ indica una actividad por encima de 0,5 en ensayo de levadura a 37 °C; +++ indica una actividad por encima de 0,75 en ensayo de levadura a 37 °C (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Tabla 6: Lista de dianas de AvrBs3 con diversas longitudes de espaciador (SEQ ID NO: 157 a 192).

Tabla 7: Lista de dianas AvrBs3 con diversas longitudes de espaciador (SEQ ID NO: 157 a 192) incluyendo una diana con únicamente un sitio de reconocimiento (compacta, SEQ ID NO: 224) y una diana de control negativo (ctrl. neg., SEQ ID NO: 228) que consiste en una diana sin ningún sitio de reconocimiento.

Tabla 13: Lista de dianas de RagT2-R/AvrBs3 híbridas con diversas longitudes de espaciador (SEQ ID NO: 315 a 350).

Descripción detallada de la invención

Salvo que se defina específicamente en este documento, todos los términos técnicos y científicos usados tienen el mismo significado que el habitualmente comprendido por un experto en los campos de genoterapia, bioquímica, genética y biología molecular.

Todos los métodos y materiales similares o equivalentes a los descritos en este documento pueden usarse en la práctica o ensayo de la presente invención, describiéndose en este documento los métodos y materiales adecuados. En caso de conflicto, la presente memoria descriptiva, incluyendo las definiciones, prevalecerá. Además, los materiales, métodos y ejemplos son únicamente ilustrativos y no pretenden ser limitantes, salvo que se especifique de otro modo.

La práctica de la presente invención empleará, salvo que se indique de otro modo, técnicas convencionales de biología celular, cultivo celular, biología molecular, biología transgénica, microbiología, ADN recombinante e inmunología, que están dentro de las habilidades de la técnica. Dichas técnicas se explican completamente en la bibliografía. Véase,

por ejemplo, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley y son Inc, Library of Congress, Estados Unidos); Molecular Cloning: A Laboratory Manual, Tercera Edición, (Sambrook *et al*, 2001, Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis *et al*. patente de Estados Unidos n.º 4683195; Nucleic Acid Hybridization (B. D. Harries y S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); la serie, Methods In ENZYMOLOGY (J. Abelson y M. Simon, eds. principales, Academic Press, Inc., Nueva York), específicamente, vol. 154 y 155 (Wu *et al*. eds.) y vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller y M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer y Walker, eds., Academic Press, Londres, 1987); Handbook Of Experimental Immunology, Volúmenes I-IV (D. M. Weir y C. C. Blackwell, eds., 1986); y Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

En un primer aspecto, la presente invención describe un método para generar nucleasas efectoras de tipo activador de transcripción (cTALEN) compactas compuestas de una única cadena polipeptídica que no requieren dimerización para abordar una secuencia diana de ADN bicatenario única específica de interés y procesar el ADN cercano a dicha secuencia diana de ADN bicatenaria única de interés.

De acuerdo con un primer aspecto, la presente invención describe un método para generar nucleasas efectoras de tipo activador de transcripción (cTALEN) compactas, que comprende la etapas de:

- (i) Genomanipular un armazón de TALE central (a) que comprende diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión al ADN y abordar una secuencia diana de ADN bicatenario única específica de interés, (b) en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN;
- (ii) Determinar o genomanipular al menos un dominio catalítico en el que dicho dominio catalítico puede procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés cuando se fusiona a dicho armazón de TALE central genomanipulado de (i);
- (iii) Opcionalmente determinar o genomanipular un conector peptídico para fusionar dicho dominio catalítico de (ii) a dicho armazón de TALE central genomanipulado de (i);

obteniendo de ese modo una entidad de TALEN compacta compuesta de una única cadena polipeptídica que no requiere dimerización para abordar una secuencia diana de ADN bicatenaria única específica de interés y procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés. En otras palabras, la TALEN compacta de acuerdo con la presente invención es una unidad de entidad activa que puede abordar, por sí misma, únicamente una secuencia diana de ADN bicatenario única específica de interés mediante un dominio de unión a ADN y de procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés.

En otra realización, es un método para abordar y procesar un ADN bicatenario, que comprende:

- (a) Seleccionar una secuencia diana de ADN de interés en una hebra de un ADN bicatenario;
- (b) proporcionar un monómero de TALEN compacta única, que comprende:

- (i) Un armazón de TALE central que comprende regiones de dipéptido variable de repetición (RVD) que tiene especificidad de unión a ADN en dicha secuencia diana de ADN de interés;
- (ii) Al menos un dominio catalítico en el que dicho dominio catalítico puede procesar ADN a unas pocas pares de bases de dicha secuencia diana de ADN de interés cuando se fusiona al extremo C y/o N de dicho armazón de TALE central de (i);
- (iii) Opcionalmente un conector peptídico para fusionar dicho dominio catalítico de (ii) con dicho armazón de TALE central de (i) cuando se necesita;

en el que dicho monómero de TALEN compacta se ensambla para unir y procesar dicha ADN bicatenario sin requerir dimerización;

- (c) Poner en contacto dicho ADN bicatenario con dicho monómero único de modo que la doble hebra se procese a una pocas pares de bases en la dirección 3' y/o 5' de dicha secuencia diana de una hebra.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado puede comprender un dominio del extremo N adicional que produce un armazón de TALE central genomanipulado que comprende secuencialmente un dominio del extremo N y diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado puede comprender un dominio del extremo C adicional que produce un armazón de TALE central genomanipulado que comprende secuencialmente diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés y un dominio del extremo C, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del

ADN.

Como también se divulga en este documento, dicho almacén de TALE central genomanipulado puede comprender dominios en el extremo N y en el extremo C adicionales que producen un almacén de TALE central genomanipulado que comprende secuencialmente un dominio en el extremo N, diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés y un dominio en el extremo C, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN. Como también se divulga en este documento, dicho almacén de TALE central genomanipulado puede comprender las secuencias proteínicas seleccionadas del grupo que consiste en ST1 (SEQ ID NO: 134) y ST2 (SEQ ID NO: 135). Como también se divulga en este documento, dicho almacén de TALE genomanipulado puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 134 y la SEQ ID NO: 135.

Como también se divulga en este documento, dicho almacén de TALE central genomanipulado puede comprender las secuencias proteínicas seleccionadas del grupo que consiste en bT1-Avr (SEQ ID NO: 136), bT2-Avr (SEQ ID NO: 137), bT1-Pth (SEQ ID NO: 138) y bT2-Pth (SEQ ID NO: 139). Como también se divulga en este documento, dicho almacén de TALE genomanipulado puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 136 a la SEQ ID NO: 139.

Como también se divulga en este documento, dichos dominios en el extremo N y el extremo C adicionales del almacén de TALE central genomanipulado se obtienen preferiblemente de TALE natural. Como también se divulga en este documento, dichos dominios del extremo N y el extremo C adicionales del almacén de TALE central genomanipulado se obtienen más preferiblemente de TALE natural como AvrBs3, PthXol, AvrHahl, PthA, Tallc como ejemplos no limitantes. Como también se divulga en este documento, dichos dominios en el extremo N y/o el extremo C adicionales pueden ser más preferiblemente formas truncadas de dominios en el extremo N y/o el extremo C respectivos de TALE natural como AvrBs3, PthXol, AvrHahl, PthA, Tallc como ejemplos no limitantes de los que se obtienen. Como también se divulga en este documento, dichas secuencias de los dominios en el extremo N y el extremo C adicionales del almacén de TALE central genomanipulado pueden seleccionarse más preferiblemente del grupo que consiste en ST1 SEQ ID NO: 134 y ST2 SEQ ID NO: 135 como se ejemplifica respectivamente en los almacenes proteínicos basales bT1-Avr (SEQ ID NO: 136) o bT1-Pth (SEQ ID NO: 138) y bT2-Avr (SEQ ID NO: 137) o bT2-Pth (SEQ ID NO: 139).

Como también se divulga en este documento, cada RVD de dicho almacén central puede estar hecho de 30 a 42 aminoácidos, más preferiblemente 33 o 34, en el que dos aminoácidos críticos ubicados en las posiciones 12 y 13 median el reconocimiento de un nucleótido de dicha secuencia diana de ácido nucleico; dos aminoácidos críticos equivalentes pueden estar ubicados en las posiciones diferentes a 12 y 13 especialmente en RVD más grandes de 33 o 34 aminoácidos de longitud. Preferiblemente, los RVD asociados con el reconocimiento de los diferentes nucleótidos son HD para reconocer C, NG para reconocer T, NI para reconocer A, NN para reconocer G o A, NS para reconocer A, C, G o T, HG para reconocer T, IG para reconocer T, NK para reconocer G, HA para reconocer C, ND para reconocer C, HI para reconocer C, HN para reconocer G, NA para reconocer G, SN para reconocer G o A e YG para reconocer T, TL para reconocer A, VT para reconocer A o G y SW para reconocer A. Más preferiblemente, los RVD asociados con el reconocimientos de los nucleótidos C, T, A, G/A y G respectivamente se seleccionan del grupo que consiste en NN o NK para reconocer G, HD para reconocer C, NG para reconocer T y NI para reconocer A, TL para reconocer A, VT para reconocer A o G y SW para reconocer A. Como también se divulga en este documento, los RVD asociados con el reconocimiento del nucleótido C pueden seleccionarse del grupo que consiste en N* y los RVD asociados con el reconocimiento del nucleótido T pueden seleccionarse del grupo que consiste en N* y H*, donde * indica una interrupción en la secuencia de repetición que corresponde a una ausencia de resto de aminoácido en la segunda posición del RVD. Como también se divulga en este documento, los aminoácidos críticos 12 y 13 pueden mutarse en otros restos de aminoácido para modular su especificidad hacia los nucleótidos A, T, C y G y, en particular, para potenciar su especificidad. Por otros restos de aminoácido se entiende cualquiera de los veinte restos de aminoácido naturales o derivados de aminoácidos no naturales.

Como también se divulga en este documento, dicho almacén central de la presente invención puede comprender entre 8 y 30 RVD. Más preferiblemente, dicho almacén central de la presente invención puede comprender entre 8 y 20 RVD; de nuevo más preferiblemente 15 RVD.

Como también se divulga en este documento, dicho almacén central puede comprender un único RVD truncado adicional hecho de 20 aminoácidos ubicado en el extremo C de dicho conjunto de RVD, es decir, un semi-RVD en el extremo C adicional. En este caso, dicho almacén central de la presente invención puede comprender entre 8,5 y 30,5 RVD, refiriéndose "5" al semi-RVD mencionado previamente (o RVD terminal, o semirrepetición). Más preferiblemente, dicho almacén central de la presente invención puede comprender entre 8,5 y 20,5 RVD, de nuevo más preferiblemente, 15,5 RVD. Como también se divulga en este documento, dicho semi-RVD está preferiblemente en un contexto de almacén central que permite una ausencia de especificidad de dicho semi-RVD hacia los nucleótidos A, C, G, T. Como también se divulga en este documento, dicho semi-RVD más preferiblemente está ausente.

Como también se divulga en este documento, dicho almacén central puede comprender RVD de diferentes orígenes. Como también se divulga en este documento, dicho almacén central puede comprender preferiblemente RVD

originados de diferentes efectores TAL de origen natural. Como también se divulga en este documento, la estructura interna de algunos RVD del almacén central de la presente invención pueden estar constituidos preferiblemente por estructuras o secuencias originadas de diferentes efectores TAL de origen natural. Como también se divulga en este documento, Dicho almacén central de la presente invención puede comprender dominios de tipo RVD. Los dominios de tipo RVD tienen una secuencia diferente de los RVD de origen natural, pero tienen la misma función y/o estructura global dentro de dicho almacén central de la presente invención.

Como también se divulga en este documento, dicho dominio en el extremo N adicional de dicho almacén de TALE central genomanipulado puede ser un dominio potenciador. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en la proteína de unión a ARN Puf o la superfamilia de anquirina, como ejemplos no limitantes. Como también se divulga en este documento, dicha secuencia de dominio potenciador puede seleccionarse del grupo que consiste en los dominios proteínicos de la SEQ ID NO: 4 y la SEQ ID NO: 5, como ejemplos no limitantes enumerados en la tabla 1, un mutante funcional, una variante o un derivado de los mismos.

Como también se divulga en este documento, dicho dominio en el extremo C adicional de dicho almacén de TALE central genomanipulado puede ser un dominio potenciador. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en hidrolasa/transferasa de la familia de *Pseudomonas aeruginosa*, el dominio polimerasa de la familia de ligasa D de *Mycobacterium tuberculosis*, el factor de inicio eIF2 de la familia de *Pyrococcus*, el factor de inicio de la traducción de la familia Aif2, como ejemplos no limitantes. Como también se divulga en este documento, dicha secuencia de dominio potenciador puede seleccionarse del grupo que consiste en los dominios proteínicos de la SEQ ID NO: 6 a la SEQ ID NO: 9, como ejemplos no limitantes enumerados en la tabla 1, un mutante funcional, una variante o un derivado de los mismos.

Tabla 1: Lista de dominios potenciadores para armazón de TALE central genomanipulado.

ID de GENBANK/SWISS-PROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
gi 262368139 pdb 3K5Q	fem-3	4	>gi 262368139 pdb 3K5Q A, estructura cristalina del complejo Fbf-2FBE SNWVLTWSDSNGEMRSLSEVLDSDGLMKFAVDKTCQFLKAVKGSLSYQKQFLFFQVIGRKDD FLKLTNIFGNLYVQSVIGISLATNDGDTYKQEKLNKFISSQMTDMCLDKFACRVIQSSLNMDLSLAC KLVQALPRDARLAIQVQDQMANHVIQYVAVIPLKNWFEIVFVATPEHLRQICSDKYGCRVWQTIIEKL TADSMNVDTLSAQNLRERALQRLMITSVTNRQELATNEVANIYIICHVSNDDLAVYRECIIKELMRNL LSLSEKQFASHVVEKAFHAPLELLAEMMDEIFDGYIPIPTDGTGDALDIMMFHQFGNYYVQCMLTICDDA VSGRRQTKFEGGYDHAISFQDWLKNLHRSVTKERHRLSRFSSGKMIETLANRSTHPIYELQ
gi 308387836 pdb 3LTJ	aRep	5	>gi 308387836 pdb 3LTJ A, estructura de una nueva familia de proteínas de repetición alfa helicoide (alfa-rep) artificiales basadas en repeticiones de tipo termostables al calor MRGSHHHHHHTDPEKVEMYIKNLQDDSYVRRAAVALGKGERAVEPLIKALKDEDAWVRRAAAADALGQIGD ERAVEPLIKALKDEGWVRRQSAVALGQIGDERAVEPLIKALKDEDFVRIAAAFALGEIGDERAVEPLIKALKDEE GWVRRQSAADALGEIGGERVRAAMEKLAETGTGFARKVAVNYLETHKSLIS
gi 109157579 pdb 2FAO	Dominio de polimerasa Ligd de <i>Pseudomonas aeruginosa</i>	6	>gi 109157579 pdb 2FAO A, estructura cristalina de dominio polimerasa de Ligd de <i>Pseudomonas aeruginosa</i> MGARKASAGASRAATAGVRISHPQRLIDPSIQASKLELAEFHARYADILLRDLRERPVSLVRGPDGIGGE LFFQKHAARLIKIPGIVQDLPALDPGHPPLIQIRSAEALVGAIVQMGSIIEFHTVWNASLANLERDRFVLDLD PDPALPWKRMLEATQLSLTLLDELSLRAFELTSGGKGMHLVPLERRHGWDEVKDFQAQANSQHARLMPE RFSVSGPRMRVGVKIFVDYLRNRSRGASTVAAYSVRAREGLPVSPVFFREELDSLQGANQWNLRSLPQRLD ELAGDDPWADYAGTRQRISAAMRRQLGRG
2R9L_A GI: 164519498	ligasa D de <i>Mycobacterium tuberculosis</i>	7	Dominio polimerasa de ligasa D de <i>Mycobacterium tuberculosis</i> en complejo con ADN: acceso: 2R9L_A GI: 164519498 >gi 164519498 pdb 2R9L A, dominio polimerasa de ligasa D de <i>Mycobacterium tuberculosis</i> en complejo con ADN GSHMGSSEQRVTLTNADKLVYPATGTTKSDIFDYVAGVAEVMGLGHIAGRATRKRW/PNGVVDQPAFFEKQ LALSAPPWLSRATVAHRSGLTTTYPIDTSATGLAWIAQQAALEVHV/PQW/RFAE/PGSGELNPGPATRIVFD LDPGEGVMMAQLAEVARAVRDLADNGLVTFPVTSGSKGLHLYTLPDPEVSSRGATVAKRVAQRLEQAM PALVSTMTKSLRAGKVFVDWSONSGSKTTIAPYSLRGRTHPTVAAPRTWAEILDPPALRQLSYDEVETRI ARDGDLLERILDADAPVADRLTRY

(continuación)

ID de GENBANK/SWISS-PROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
1KJZ_A GI:20664108	Subunidad gamma grande del factor de inicio Eif2 de <i>Pyrococcus abyssii</i>	8	<p>Estructura de la subunidad gamma grande del factor de inicio Eif2 de <i>Pyrococcus abyssii</i> G235d mutante: 1KJZ_A GI:20664108. >gi 20664108 pdb 1KJZ A_cadena A, estructura de la subunidad gamma grande del factor de inicio Eif2 de <i>Pyrococcus abyssii</i> G235d mutante</p> <p>GEKRSRQAEVNIQMVGHVDHGKTTLTKALITGVVWTDTHSEELRRGITIKIGFADAEIRRCPCNGRSTSP</p> <p>VCPYCGHETEFVRRVDFDAPGHEALMTTMLAGASLMDGAILVIAANEPCCPRPQTRHEILMALQIGOKNI IIAQNKIEIVDKKALENYRQIKIEFEGTVAENAPIIPISALHGANIDVLYKAIEDFIPTPKRDPNKKPK MLVLRSDVNRKPGTPPEKLVGGVLDGSSIVQGLKLVGDEIRPGVPEEYEHGRIKYEPTITENSLOAGGQ FVEEAYPGGLVGGTKLDPYLTGKDLIMAGNVVGGKPKLPPVWDSLRLEVHLLERVVGTQELKVEPIKRX EVLINVGRTARTMGLVTGLGKDEIEVKLQIPVCAEPGDRVAISRQISRWRLVIGYGIKE</p>
2D74_A GI:112490420	Factor de inicio de la traducción Aif2betagama	9	<p>Estructura cristalina del factor de inicio de la traducción Aif2betagama heterodimérico: 2D74_A GI: 112490420. >gi 112490420 pdb 2D74 A_cadena A, estructura cristalina del factor de inicio de la traducción Aif2betagama heterodimérico</p> <p>MGEKRTQAEVNIQMVGHVDHGKTTLTKALITGVVWTDTHSEELRRGITIKIGFADAEIRCSNCGRYSTS PICTYCCHETEFVRRVDFDAPGHEALMTTMLAGASLMDGAILVIAANEPCCPRPQTRHEILMALQIGOKNI IIAQNKIEIVDKKALENYRQIKIEFEGTVAENAPIIPISALHGANIDVLYKAIEDFIPTPKRDSNKKPP KMLVLRSDVNRKPGTPPEKLVGGVLDGSSIVQGLKLVGDEIRPGVPEEYEHGRIKYEPTITENSLOAGG QFVEEAYPGGLVGGTKLDPYLTGKDLIMAGNVVGGKPKLPPVWTDLRLEVHLLERVVGTQELNVEPIKR KEVLLINVGRTARTMGLVTALGKDEIEKLVQIPVCAEPGERVAISRQISRWRLVIGYGIKELEHHHHHH</p>

Como también se divulga en este documento, el dominio catalítico que tiene capacidad de procesamiento de ADN cerca de la secuencia diana de ADN monocatenario única de interés, cuando se fusiona a dicha armazón de TALE central genomanipulado de acuerdo con el método de la presente invención, se fusiona preferiblemente en la parte del extremo N de dicho armazón de TALE central. Como también se divulga en este documento, dicho dominio catalítico se fusiona preferiblemente en la parte del extremo C de dicho armazón de TALE central. Como también se divulga en este documento, pueden fusionarse dos dominios catalíticos tanto a la parte del extremo N de dicho armazón de TALE central como a la parte del extremo C de dicho armazón de TALE central. Como también se divulga en este documento, dicho dominio catalítico puede tener más preferiblemente una actividad enzimática seleccionada del grupo que consiste en actividad nucleasa, actividad polimerasa, actividad cinasa, actividad fosfatasa, actividad metilasa, actividad topoisomerasa, actividad integrasa, actividad transposasa o actividad ligasa. Como también se divulga en este documento, el dominio catalítico fusionado al armazón de TALE central de la presente invención puede ser un activador o represor de la transcripción (es decir, un regulador de transcripción) o una proteína que interactúa con o modifica otras proteínas tales como histonas. Los ejemplos no limitantes de actividades de procesamiento del ADN de dicha TALEN compacta de la presente invención incluyen, por ejemplo, creación o modificación de elementos reguladores epigenéticos, generación de inserciones, eliminaciones o reparaciones específicas de sitio en el ADN, control de la expresión génica y modificación de la estructura de la cromatina.

Como también se divulga en este documento, dicho dominio catalítico más preferiblemente tiene una actividad endonucleasa. Como también se divulga en este documento, dicho dominio catalítico más preferiblemente tiene actividad de escisión en dicho ADN bicatenario de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicho dominio catalítico más preferiblemente tiene una actividad nickasa sobre dicho ADN bicatenario de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicho dominio catalítico se selecciona más preferiblemente del grupo que consiste en las proteínas Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP1I, I-Basl, I-Bmol, I-Hmul, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, Bsml, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-CreI, hExoI (EXO1_HUMANA), ExoI de levadura (EXO1_LEVADURA), ExoI de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de los mismos. De acuerdo con el método de la presente invención, dicho dominio catalítico es I-TevI (SEQ ID NO: 20). Como también se divulga en este documento, el dominio catalítico I-TevI (SEQ ID NO: 20), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del extremo N de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 420-432. Como también se divulga en este documento, dicho dominio catalítico es preferiblemente ColE7 (SEQ ID NO: 11), un mutante funcional, una variante o un derivado del mismo. Como también se divulga en este documento, el dominio catalítico ColE7 (SEQ ID NO: 11), un mutante funcional, una variante o un derivado del mismo se fusiona preferiblemente en el dominio del extremo N de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico ColE7 (SEQ ID NO: 11), un mutante funcional, una variante o un derivado del mismo se fusiona preferiblemente en el dominio del extremo C de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención preferiblemente comprende una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 435-438.

Como también se divulga en este documento, dicho dominio catalítico es preferiblemente NucA (SEQ ID NO: 26), un mutante funcional, una variante o un derivado del mismo. Como también se divulga en este documento, el dominio catalítico NucA (SEQ ID NO: 26), un mutante funcional, una variante o un derivado del mismo se fusiona preferiblemente en el dominio del extremo N de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico NucA (SEQ ID NO: 26), un mutante funcional, una variante o un derivado del mismo se fusiona preferiblemente en el dominio del extremo C de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención preferiblemente comprende una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 433-434.

Como también se divulga en este documento, dicho dominio catalítico puede ser I-CreI (SEQ ID NO: 1), un mutante funcional, una variante o un derivado del mismo. Como también se divulga en este documento, el dominio catalítico I-CreI (SEQ ID NO: 1), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del extremo N de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico I-CreI (SEQ ID NO: 1), un mutante funcional, una variante o un

derivado del mismo puede fusionarse al dominio del extremo C de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 439-441 y la SEQ ID NO: 444-446.

Como también se divulga en este documento, dicho dominio catalítico puede ser una enzima de restricción tal como Mmel, R-HinPII, R.MspI, R.MvaI, Nb.BsrDI, BsrDI A, Nt.BspD6I, ss.BspD6I, R.PleI, MlyI y AlwI como ejemplos no limitantes enumerados en la tabla 2. Como también se divulga en este documento, dicho dominio catalítico puede tener una actividad exonucleasa.

Como también se divulga en este documento, cualquier combinación de dos dominios catalíticos seleccionados del grupo que consiste en la proteína Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinPII, I-BasI, I-BmI, I-HmI, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococcos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, Bsml, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-CreI, hExoI (EXO1_HUMANA), ExoI de levadura (EXO1_LEVADURA), ExoI de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de estos dominios proteínicos de los mismos, pueden fusionarse tanto a la parte del extremo N como la parte del extremo C de dicho armazón de TALE central, respectivamente. Por ejemplo, el dominio catalítico I-HmI puede fusionarse a la parte del extremo N de dicho armazón de TALE central y el dominio catalítico ColE7 puede fusionarse a la parte del extremo C de dicho armazón de TALE central. En otro ejemplo, el dominio catalítico I-TevI puede fusionarse a la parte del extremo N de dicho armazón de TALE central y el dominio catalítico ColE7 puede fusionarse a la parte del extremo C de dicho armazón de TALE central. Como también se divulga en este documento, de acuerdo con el método de la presente invención, dicho monómero de TALEN compacta única puede comprender una combinación de dos dominios catalíticos, respectivamente fusionados a la parte del extremo C y a la parte del extremo N de dicho armazón de TALE central seleccionada del grupo que consiste en:

- (i) un dominio Nuc A (SEQ ID NO: 26) en el extremo N y un dominio Nuc A (SEQ ID NO: 26) en el extremo C;
- (ii) un dominio ColE7 (SEQ ID NO: 11) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C;
- (iii) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C;
- (iv) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio NucA (SEQ ID NO: 26) en el extremo C;
- (v) un dominio ColE7 (SEQ ID NO: 11) en el extremo N y un dominio NucA (SEQ ID NO: 26) en el extremo C;
- (vi) un dominio NucA (SEQ ID NO: 26) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C.

Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 448 y 450.

Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una combinación de dos dominios catalíticos, respectivamente fusionados a la parte del extremo C y a la parte del extremo N de dicho armazón de TALE central seleccionada del grupo que consiste en:

- (i) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio FokI (SEQ ID NO: 368) en el extremo C;
- (ii) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio TevI (SEQ ID NO: 20) en el extremo C;
- (iii) un dominio scTrex2 (SEQ ID NO: 451) en el extremo N y un dominio FokI (SEQ ID NO: 368) en el extremo C.

Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 447-450 y la SEQ ID NO: 452.

Puede preverse la inserción de dicho dominio catalítico entre dos partes del armazón de TALE central genomanipulado, comprendiendo cada parte un conjunto de RVD. En este último caso, el número de RVD para cada parte del armazón de TALE central genomanipulado puede ser el mismo o no. En otras palabras, puede preverse dividir dicho armazón de TALE central para insertar un dominio catalítico entre las dos partes resultantes del armazón de TALE central genomanipulado. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 453-455.

Tabla 2: Lista de dominios catalíticos/potenciadores para TALEN compactas o TALEN compactas potenciadas.

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
ACC85607.1	Mmel	10	>gi 186469979 gb ACC85607.1 Mmel [Methylophilus methylotrophus] MALSMNLRKKAIEFSEKWFVASENEQAMPFLIDFFVEVGIINRVAIVTIIHAAVKKFAALKEGSGGQVFD LFWEGILLIENKSGKDLQADYQALDYSGIAEEDLPRVLCDEQERLIDLIETISVEFLKDDLYLN VRSFGEIAGYQIGVKEQDFINIKAAERMGKIHUIALAVGNEGHALIYIVRAIIFLFAEDTILFKSLF QEYIEITLEDGGSLAHHIHTLFFVINTPEKRLNKIDDEHLAFFFYINCHLFEEDLFFAQDFKAMRLALIL DLCSLDWRLSEALFQSLFQSLMEAKKRNAGHYTSEANLKLIEPLDDELWVDFZEVKRNKMLLAF HKRLRGITFEEDGACGGNPLVITYRELRLALIEVLRLGHRGGVLDIEHLIQINVDQFVGIHIELEFFAQ IAQVALMILDHQSNKLSQKFCNTYFARIPKISTPHILNANALQIDWADVLEAKVCFILCNPPFVYKSKQ TFQKADIGLAVFVGVKASRSDIIPWARYTQANRYTQTNRTRGAFVSTNITQESQVSLRMPILMISIGTK INFAHRTESRTNEASGVAAVHVCTIGFGLKDSQFKIIVEYEIINGEDPLAKKNNINLYLGGVDVLACKR QQPISKLPLSMGYKRFYDIDVNFPLTULEKNGCFINISFSSKAYFPHAVGSDLEFINNYSMWLMLAGADLSE IRAMELVILARIKRVQDFRLKESANFTQGSSTPKKFTYIQQPDIYLLIFETSSENEQFIPITGFVDANVI SSNATYHLPBAEPLIYGLLSSTMGKNCNHNNAVCGHLEERVYASALVYVNHPPKIQPNEKQSKALEAAPAI LMAHSNTFNESLAQLYDFRIVTSELAMARQKLNKAVDVSYGFKGFNTEIARIAPLITLSTYQMMISLGRFEK EIKKSGGRN
Q471122	Colicina-E7 (CEAT_ECOLX)	11	>gi 12644448 sp Q47112.2 CEAT_ECOLX Nombre recomendado: Completo=Colicina-E7 MSGGDGRGHNSGABNTGGNINGBFTGLGGNGGASDGGWSSENNEWGGGGGSHWGGGGHNGGNGEN SGGGNSSVAAZMAAFGEZALAAPCAGTLCISVSGEALSAALADIFAAKGPFFKFSAWGIALYCIPLSELA KDDFNMSKI VTSLF AEIVTVNVVSTLFLDQATVSVTKRVTIDVVNDTRQHJAVVAGVPMSPVYVNRKFTR IPQVFRASFTGVP SLTVSTYKGLPVSTLPRGITEDKRTAVPAGFTFGGSGHEAVIREPKESQKPVYV SVIDVLTFAQVKGQDEEKRLQCEWDAHPVEVAENHYEQANLQANIKDVAQNNQERQAKAVGVYNEIK SELDAAKSTLADAKAEIKQFERFARPNMAGHRMCMAGLMAQRAQTDVNNKKAFTDAAAKERDADVAL SSALEKAKQENKEDAKALDKESKRNFQKATGKGFVNNKWLNNAGKDLGSPVEDRLANKLRDKEFK SFDDFRKWFWEVSKDFELSKQFSNNWDRHMVGRKFKTRTQDVSQKRTSFEIHHKFI SQNGGVYDMON ISVVTPKRHIDIRGK
CAA38134.1	EndA	12	>gi 47374 emb CAA38134.1 EndA [Streptococcus pneumoniae] MNRKTRQTLIGLIVLGLLSTGYSYIKQNE SAFPSEKINLSQKQKQSERP SQALAE SVLTDAYKSIKGSLL EWNGSGAP IVNGNKINLDAKVSXFPYADNKKIYVQKSTVFIVANALLSKATROYKMKXETGNGSTISWTFP GWHQVQNLKGSYTHAVDRGHLLGYALIGLDFGDASTSNPKNIAVQTAWANQAQAEYSTGQNYVESKVRK ALDQNKRVRYRVTLYAYASNEDLVPSASQIEAKS SDGELETFNVLVNVPVQKGLQLOYRTGEVTVTQ

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
P25736.1	Endo I (END1_ECOLI)	13	>gi 832515 sp P25736.1 END1_ECOLI Nombre recomendado: Completo=Endonucleasa-1; AliName: Full-Endonuclease I; Short-Endo I; Flags: Precursor NYYVLEIAAVVLEAAAFSGFALAGISLNGFEGAKAAVAVVAVVADAPGTFYGGCXINWCKKQVWVQLGGCGV KNNENASAVTWEVTVVAVWQVGFHDSQRRQDGGKXNCAKELVYTRSEFSQMHNIQVAVGVVDRGQENKMSQ KNSGEGYGGCCAKVYVDFSEKAAAFPARARGAIAATVYVYRQDGNLTLERQQVTFNFANRNYVPTWDECE HUEBLAKVQGNHNEVYVQAAQARKE
Q14249.4	Endo G humana (NUCG_HUMANA)	14	>gi 31737579 sp Q14249.4 NUCG_HUMANA Nombre recomendado: Endonucleasa G completa, mitocondrial; Corta=Endo G; Marcas: Precursor KMAADAGLTLASQAGLGNVVEGWRDREEDAPAAAGLLGRLPVLVVAAMALPPVFGDPAQGEELAKYQLP CLAQLKSEZSYVLCDEMTGCLAKVWVQLRFRLEKLDLRRDCDSELDVSHAVHRAINADYRGGGDRG HLAAANHRMSQKAKNUTTYLSSVAFQVFBUNQANNNLEKYSBLSLBSYQVAVVYCTGELFLEHTEAGH STVYVQVIGEMRVAVFTHEFRVLLIEMARGGQIELRYYVNFNAFVUEALFLEREVLFIESIENASGLLVE NILAGAGSGEALITAGSK
P38447.1	Endo G Bovina (NUCG_BOVINA)	15	>gi 585596 sp P38447.1 NUCG_BOVINA Nombre recomendado: Endonucleasa G completa, mitocondrial; Corta=Endo G; Marcas: Precursor PQLLRAGLTLALGAGLIGAAAEZWRQRADARATPGLLRLLPVLVVAAGAGLPANFAPAGGEGEELKAYG LFGVAQLKSRASVYVLCYDQFVREGALWVVKQLRFRLEGLRGGNSHSCTDEHEDDQVSHAYHHAITNADYHGGGFE RQSLAANHRMSQKAKNUTTYLHRVAFQVFBUNQANNNLEKYSBLSLBSYQVAVVYCTGELFLEHTEAGH GSEYVQVIGEMRVAVFTHEFRVLLIEMARGGQIELRYYVNFNAFVUEALFLEREVLFIESIENASGLLVE VFNLLRAGSGEALITAGSK
AAW33811.1	R-HinP11	16	>gi 57116674 gb AAW33811.1 JRHinP11 endonucleasa de restricción [<i>Haemophilus influenzae</i>] MHLVGGRTADGTFKNEEDIAQRTEWKEDEAGQWVIMGHNLQEIKEVRAVWLGQYRMDINVQVLE YKDALDHNQWVLYGNSGEHCIDKDFLAFYGEANNKTDENLRIHRHFTGELPETHONTKNSNMENTE FSGSEQNVILWMLDNRVLVLTDLRGGGELFALVYVIVRAGVGNNAJWILANLNEVLOHTYGGSLIILGPR GSEFRVTIQRGGNGSEIANNLQFKIDFLELFDI
AAO93095.1	I-Basi	17	>gi 29838473 gb AAO93095.1 I-Basi [lego de <i>Bacillus Bastille</i>] MFQEEWKIVIGEDYVEVSNKQVAKRTGVINAGYMINESYLCIXETVVKRYISLHVRLVAREFCEGY SPELDANIKDTDSMNNYDNLLEWLTADNLRKQVREKGLNTHITAREALAKVSKKAVDVIYKGGSEYIATY PSATEAEALGVQGRKLSIVCHGKRQHTGGYHFKFNSGVDPNRSVSKK
AAK09365.1	I-Bmal	18	>gi 12958590 gb AAK09365.1 AF321518_2 intrón codificado I-Bmol [<i>Bacillus mojavensis</i>] MKSGVTKITNKNTRKPYIGSEEDCESEKLVHFRNLNRRHINRYLNNEFNHGGQVTEGEVILHLELLA IASEQWTLIDNEYEEMNISKAYHGGDLTSYHPDESNILKADUSLAKVYLKMISEEXAKNWCYVQGENN PMEGRHEITTLKLSNHNKLYSTHAMPFKGKHSESESKLSEYASQVGEKKNFFYCKIHSDEFKTYN SKAFKCRKPRANSRQVLIIDGTEYESATEASRQINLVVPAITLHRKSKNKEYSYPYK

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
P34081.1	I-Hmul	19	>gi 465641 sp P34081.1 Hmul_BPSP1 Nombre recomendado; Completo=Endonucleasa de ADN I-Hmul; Nombre alternativo: Completo=endonucleasa de asentamiento HNH I-Hmul MENNDRKCYGSHYQVNTGDEVYGLKGGKTLKRCIFEDGVRHRIIGLFQGGKQRTFQVHRLVALIHFCGVEYDGS LVYDHRKDKNDNNLSTNLRWVTKINVENQMSRGTLANVSMALQQFAXIKNQKFLIVLSPDQIEKEXFSTRC ACEELGLTRSRVTDVIAIGHRIHHKGYTFRYKLAG
P13299.2	I-TevI	20	>gi 6094464 sp P13299.2 TEV1_BPT4 Nombre recomendado; Completo=Endonucleasa 1 asociada a intrón; Nombre alternativo: Completo=I-TevI; Nombre alternativo: Completo=proteína IRF SKSGIVQIRKNTLNKVVYVSGAKDFEERKRFEDLKGKCHSSIELQRFSEVZHGNVZDCAILEEETPYEKDL LIERANWRELNSKINGYNIADATFDTCSHPLAKEIIMKSEETVKAHMLLGFQGRKALYGRHGBAN GRNREETHREFCGCVRLQIQTGAYTCSTCRNRSSENNFFNHEDITKSKISEKDKGRKESTWIKKISGCDGV IFDCAADABERKISGLAVTVKSDKNNFYINA
P07072.2	I-TevII	21	>gi 20141823 sp P07072.2 TEV2_BPT4 Nombre recomendado; Completo=Endonucleasa 2 asociada a intrón; Nombre alternativo: Completo=I-TevII MWEELRSLKIANSWAEYMYVRFDDKSEVYIGERKFKTIYGGKTNWKEVNSSSKLVKELKDYKAKWIIIQ VFUSYELALHHEMIAIRYFNNEFJINKQIIGYKFKMYPOSEHRQWLNARHKKKILSLKHKKREREMLI ERYKNNSSSAHVANNISGETAKKTAVSIALKSCNFRSFRSAARFVPCSEZQVSNHFNVDIKITIIHVEY EYVINDNLYKSEYDAEDLKLHESLIRKQJLCLDADTPNIVSYEVEK
Q38419.1	I-TerIII	22	>gi 11387192 sp Q38419.1 TEV3_BPR03 Nombre recomendado; Completo=Endonucleasa 3 asociada a intrón; Nombre alternativo: Completo=I-TevIII MNYRKIWDANGPFPKQSDGRDDEIHHKQGNRENDLENLCLJIQEHYDIIHLAQDYQACHAIKLRNMY SPEEISYLASKAAKRSIQIFNYFEVRANIASYKSKIEGIFHLIDGELCRKSNLRVALGJHNFQQAEL HLAKVRERNIAAIKEGTHVFCGGMOSSETOSKRVNDGSHHFLSEDHKRTSAKTTLEMVYKNGTHPAGKEIT CDFCGHIGKGFPEYLXENDRCKLNPRIQNLNCPYCDKDKDLSPTYKHWGGDCKARFND
AAM00817.1	I-TwoI	23	>gi 19881200 gb AAM00817.1 AF485080_2 HNH endonucleasa I-TwoI [fago de estafilococosTwort] MEELEKREIDGNSMLGNKGVYGRRNKILALDTEGCKYKRIEISFNNECRPILLGVHKLVLDFEGDINT EKEIFHHKNNLRDQNFLENLEWTVVSENIKHAZDIGALKSERRVTCILYKGEFLSCYDSEFLAKALKV SRSVIESFRNSLVLSIFVAKREPTIQGAPLNKKEIFEHSLIKGLGNPPLKAVTNEDETYFLLIMDISKIFN EESKXQVQRYGYZKWSYIIEHIDFYEYKQIH
P11405.1	R.MspI	24	>gi 135239 sp P11405.1 T2M1_MORSP Nombre recomendado; Completo=Enzima de restricción de tipo 2 MspI; Corto=R.MspI; Nombre alternativo: Completo=Endonucleasa MspI; Nombre alternativo: Completo=Enzima de restricción de tipo II MspI MTEILLKLYDGFIIQQLFHHQGVYSDRLGKLVKRYILDIFKDIELKXNYNMTFGEDLGGKLLKAL NIDLRLIIDVSSVTDGERTIARSSKTDATINFTDMSSELVTEMLKHSSEKVSYLAFTYDVTICTW GYSSEKLELLERKQNDQSAKLFVQKQRLKLELLEFEPERFHWCVTLHARSSGNILLRDLLERFQVI DREYVDTIKNIIDQVNSGLAESREARRPFPOTCLMWTYASGEMARWQKFC

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ. ID NO	SECUENCIA FASTA
R.Mval	R.Mval	25	>gi 119392963 gb AA03024.2 AF472612_1.R.Mval [Kocuria varians] MSEYLNLRKAEIQNVVDSGRREIKRIGNTGIBRIFEDLLEKEEDNMAPDFEDLEIKTRTAARSLDLE TKSPINPHGANINLANKYCKQKQYGNKI LHQVYSGNKTKNSNRYDKFIDIDAKSQUVRLKVFUKQIIM IDMSVYVSDLSQNLDDNMLYIYAVI SAKSKIENRHYFYKNSAMGLFTDLIVQLSCRGIENGDDIIVVDIPI GAYBSKSKKGGKIHORHGAERINENKLELYGEVYIV
CAA45962.1	NucA	26	>gi 39041 emb CAA45962.1 NucA [Nostoc sp. PCC 7120] MGIQGLVVAALVALIVGCSFVQSGVFPFLTEILSFSISVHLLGLNDSGATEFKLITPDMYLVNRYALSLYN NEKGTANWVWCLNSWFLGNFRQCNFRKILFPAWVRVTPSMYSGSGITDRGHIAPESDRTKTIEDNAA IFLMIINNEQTFEDNUNNYWGNLEDYCRELVSGKELYIVAGPNSLSKPLKSKVIVFASVWKLIVVLDSE CSOLEGITANTFVLAVINIENDPELANNDRWAYKVSVELESLTCYDFLSNVSPNIQTSIESKVDN
P37994.2	NucM	27	>gi 313104150 sp P37994.2 NUCM_DICD3 Nombre recomendado: Compto=Nucleasa nucM, Marcas: Precursor MLRNVIVPAVLGAGLITLAAAGDINNFTQAKAAAKIHQCAPGTFYCSCKINQGRKGTDFLAGGGYQV RKDANRASRIEWEHVVEAWQEGHQRCWQGGGRNCTKDDVYRQLTEIDLHLQEAIGEYNGDRGNEMYSQ NWGGERQYGCENKIDKESQLAAPPERRARGAIARTFYMKDRYINLNSHQQTQLFDANNKCYEATIMECT REKRIAAVQGNHNYVYQACQP
AAF19759.1	Vvn	28	>gi 6635279 gb AAF19759.1 AF063303_1 precursor de nucleasa Vvn [Vibrio vulnificus] MKRLFTIASPTAFIQAAPPESFSAKQGVKTYQDRPTSEYGGDLIEWQKKGIPNLETGGYQVREQQ TRASRIEWEHVVEAWQEGHQRCWQGGGRNCTKDDVYRQLTEIDLHLQEAIGEYNGDRGNEMYSQ VDGVSYGGCENQWNEKURKVPFQTELENGSIAHIYLYMSQYDEQLSAQQQLMQUANNKSTPVDWEECTND DRIAKIQGNHNPVYQSQQTQ
AAF19759.1 (referencia)	Vvn_CLS	29	>Vvn_CLS (variante de AAF 19759.1) MAGQAPPSESAKQGVKTYQDRPTSEYGGDLIEWQKKGIPNLETGGYQVRSKQUTRAIRLEWERYVVA WQEGHQRCWQGGGRNCTKDDVYRQLTEIDLHLQEAIGEYNGDRGNEMYSQ FTRKRVPPFRARSGSLAARTIYLYMSQYDEQLSAQQQLMQUANNKSTPVDWEECTHEDRIAKIQGNHNPV SQSFGYTGCSBAD
P00644.1	Nucleasa de estafilococos (NUC_STAAU)	30	>gi 128852 sp P00644.1 NUC_STAAU Nombre recomendado: Completo=Termonucleasa, Corto=TNasa, Nombre alternativo: Completo=Nucleasa de micrococos; Nombre alternativo: Completo=Nucleasa de estafilococos, Contiene: Nombre recomendado: Completo=Nucleasa B, Contiene: Nombre recomendado: Completo=Nucleasa A, Marcas: Precursor MLWNTCYLLDAGICIMAVIYGLIAGMAGISVSGQYAKREFFATSCGLVLTAVVVSGLSGANAGQDQDQV NNSGSEDETVYSATSTKSLRKEPATLIMKIDGDTVKLASHGQPMTEFLGLLVDTFETSRFKGQVKEYGFDA SAFTKRVNNAKKIEVEFDKQARTKYSKGLAYIYADGKAVNEALVAVQGLAKRVAVYVKEKNTHECHLRKES EACAKKELKLNIRSEDNADGGQ

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
P43270.1	Nucleasa de estafilococos (NUC_STAHY)	31	>gi 1171859 sp P43270.1 NUC-STAHY Nombre recomendado: Completo=Termonucleasa; Corto=TNasa; Nombre alternativo: Completo=Nucleasa de micrococcos; Nombre alternativo: Completo=Nucleasa de estafilococos; Marcas: Precursor MKKITTGLIIVVAAALIVLSIGPMTESGFFKRAENLNANEQYKVIIRVIDGGTIIIVKDCRQGNLRNICVD TETVTPNTEVQYGEAEASDETKRHLLINQVRFLETKQEMRIGRIILAYVWLGKEMENFKLAKEGGLAAK EYRPNKYGERIEQAKQAKKRNININ
P29769.1	Nucleasa de micrococcos (NUC_SHIFL)	32	>gi 266681 sp P29769.1 NUC_SHIFL Nombre recomendado: Completo=Nucleasa de micrococcos; Marcas: Precursor MKSALAALRAVAAAVVLIIVSVPAMADFERGEVVRILDCDTIDVIVNRQTIIRVRLADIDAFESGQAFGSRAR QRLADLFRQEVQVTEKVDRYGRILGVVYAPLQYFGGQTQLTINAIMVQEGMAWAYRYGKPTDAQMY EYEKELARRQRLLELWSDPNAQEPFKWRRAAKNATN
P94492.1	Endonucleasa yncB	33	>gi 81345826 sp YNCB_BACSU Nombre recomendado: Completo=Endonucleasa yncB; Marcas: Precursor MKKILLINLAIIVLSITLAAAGSNHAAKNHSDSNGTEQVSQDTHSNEYNGTEQAGTPEKKNQKALVNVTL DRAIDGDTIKVINYNGKQDVRVLLVDTPEFKFNSCVQFYGEDASKANKELVNSGKLELEFDGDRDDKY GRLLAYVYVDGKSVQETLLKGLARVAYVYEPNTKYIDQFRILDEQEAASKDKLSIWSRSGYVTVNRFNGCV K
P00641.1	Endodesoxirribonucleasa I (ENRN_BPT7)	34	>gi 1193701 sp P00641.1 ENRN_BPT7 Nombre recomendado: Completo=Endodesoxirribonucleasa 1; Nombre alternativo: Completo=Endodesoxirribonucleasa I; Corto=Endonucleasa MAGYGAQKIRKVGAEFRSGLEDKYSKQLESKGIKFEYEERKVPYVIPAENHITYTPDZILLPNSGIFVETKGLW ESDDRKKKHLILIREQHPELDRIIVFSSSRKTKLYEGSPTSYGEFCEKKGKIKFADEKLLIPAWIKIPEKREVEFFD RLKRKGGGKK
Q53H47.1	Meinasa	35	>gi 74740552 sp Q53H47.1 SETMR_HUMANA Nombre recomendado: Completo=Histona-lisina N-metiltransferasa; SETMAR; Nombre alternativo: Completo=Dominio SET y proteína que contiene el gen de fusión transposasa mariner; Corto=HeMar1; Corto=Meinasa; Incluye: Nombre recomendado: Completo=Histona-lisina N-metiltransferasa; Incluye: Nombre recomendado: Completo=Transposasa Manner Hsmar1 MAEFKIKPEAPTEQLDVACCCENLVPVGAWFFGAAPFQYTPDQVHVGEGADIDSTQITFFGCIQVKTFCI PGTCSCLRHGENYDINSCLRDIGSGGKAAEPVEFCNVLCRCDHCRNRVAVQKGLQFHQFQVETKHKKGWGL RTLEFIPKGRFVCEYAGEVLEFSEVQRRIHLQTKSDSNYIIAIRREHYNGQVNETFVDFTYIGNIGRFLN HSCENLIMIPVRIDSMVEKALFAAKDIPVEELSYDYSGRYLNLTIVSEDKERLADHGRKLRPCYCGAKS CTAELEFDSSLYCQVEKSNLSCGNEKEP SMCSSAPSVFSCKGLTLETHMMMLDKKQIRALFLFEFKGR KAETIRNINNAFPGGANERTVQWVFKKICRQDESLDEERSGRSEVNDQLRAIIEADFLITIREVA EELVNHSTVVRHLKQIGRVKLDKXWVPELITENQKRRREVS SLLIRNHNEPFLDRIVTCDEKWLIDY NRRRSACWLDQEEAPKHFPPKPI LHPKKVMVTIWSAAAGLIHYSELANGETITSEKVAQEIEMNQKLRRI QLALVNRAGPILLDNARPHVAQPTLQKLANELQVEVLPHPFPYSPDLILFTNYHVKHLNLFLOGKREHNOQ DAENAFQEFVESQSTDFYATGINQLISRWQKVCVDCNGSSTFD

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
ABD15132.1	Nb.BsrDI	36	>gi 86757493 gb ABD15132.1 Nb.BsrDI [Geobacillus stearothermophilus] MTEYDHLIYADSEHEGHWCKENLAKIAGSDGGRHQIDYLGQEIFRHSIFSDLIINWIVFGYSKSRHLIP KQIKDLLEWKEZDFIAKDFPKNDKILFAVEETGAVFTGNOALCRGERIYGSARXQIEFWYLLISEPQCHKIDG GTRRHSIFWLIIMGLKATQVATPSILHYSINNPEDYNSGNGIKFLPNSILAGLIINVCYLIKNFLKQCLP LLSIQYENLSEETNGQWKEQIDDFLAGEELANTNTEKELAFMTASLAISQIVKIPTEELNWFNIDKVNTEGTF CGLIWKYDELQYDLEKAVGSKKAYCLSNWAGKPKLESLKWKINSOKLFDKAPKLTFFKFNANKLDAPF VT-SNNNYIVYITSERNI LKLFYKXDLALIAIEIAFRLKNGKLPIDIDEKPALIYICNENKPKORLFGFFTQQ LSAFSTIFGKKNIMPRIVVAAYFQYVQALPKNNKNGKITLKLKELDFEYIFGQVWVKLNECKAY
ABD15133.1	BsrDI A	37	>gi 86757494 gb ABD15133.1 BsrDI A [Geobacillus stearothermophilus] MTDYRYSELSSEETANWAFEIKTKNIDWFAFSNPTAGPWKRVMADKASNREGEVHRFGREDERPDIIL VNDNISLILILEAKEKLNQLISKSQVOKSVEVFLTLLSSILKEASDNNYKCDRLKYINVLGILKGSERQETS QKQIDNAPFRVYRDSLVKKNLKEINFTPTNICIDILGVGESIKNKKEEISIKLHVSNIIYAEIYFKFKTKHLL EKLAVLN
ABN42182.1	Nt.BapD6I (R.BspD6I subunidad grande)	38	>gi 125396906 gb ABN42182.1 endonucleasa de restricción heterodimérica R.BapD6I subunidad grande [Bacillus sp. D6] MAKVNWYVSGSRSEKIQELKVLANFEGSYKGVKQKAEFAKELAAALPQFLGTTYKKEAFAFSTR DRVAFNKTYGFVFDSEGYLRITEAGKNLANNRPKDVF LKQIVWQYPSFQHGKREYPEREWSINPLVFE VLSLLKAVGGISKLDIAMPCLTATNNQVDELAEEIMQFHNREKLIKQGNKKLSEYENYFFKRYFKLYGN VCKIRECKSDGSHKSKIEITKRNARADVADATIRYFHYTGLFVARCQLVLENEKSDLIDEIISSSKVVKN YTRVEEFHEVYGNPSLPQSFETKEQLDLJAHRIQDENTRLAEQVVEHFFWVKVEIQVLEDIYNSLNKKV LVEI LKDVITHAKELQLELKKKQLQADPNDEKQLEEVIDLLEVYHEKKNVIEEKIKARFLANKNIIVFEWL IWNGFITLGNALYKNNVEIDBELQPVTHAAGNQPMEIYEDFVLSGEVTSKQATQFNMESERVTRHY LNKXKELEKQGVRELYCEFIAPENKNTFEFHMAYNIQVNTIRIIPLSLEQFNMLINQKXGLETSGRRLS SYDIKNLWSELYRTIIECEKRYTQIKAGLETLNNWVDKEVRF
ABN42183.1	ssBspD6I (R.BspD6I subunidad pequeña)	39	>gi 125396907 gb ABN42183.1 endonucleasa de restricción heterodimérica R.BspD6I subunidad pequeña [Bacillus sp. D6] NQDILDFYEEVERTINPKNYFEMNTYRVFKKLGSYKNLVPNFKLDDSGHFICNNAIPGVEDI LVEYEHFSI LIECSLTIGEKQLDYEGDSVWRHLQCYKKGIEAYTILFLGKSIDLSFARHIGENKESERVIPLTVDQFKK LVTQLKGGGGEHFNFNKLSILIKLLRSDLGYDQAEELWLIIFYNLK

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
AAK27215.1	R.PleI	40	<p>>gij13448813 gbiAAK27215.1 AF355461_2 endonucleasa de restricción R.PleI [Paucimonas lemoignei]</p> <p>MAKILDERVAFITIGSPRTYDHWVFIEELLDNFMDVWNRCTOALMHLIREEGEFDHEGRNDYAFGRD RINRAPSZGEVILITPELSDDAVELEKAKKCOIFLRQSLKTLQFSPYERLSOKAALFYVQZYLELFB IVRHFGSLTDELMKIFGLIIDFRLFNQIVDKIEDFVGRKINRGRVYRERKTEELGRVYDELFGL TEAANKLITIHESGHNEDYADACVYLDATGNWVWYQKSLGIVQZKHEVDFTLNTKIEPQTIKDEA SYEVLGNFYKLFVDDVDRXXXLAFDFKTKNKVNALITPELKEELZNEILSRKSLAKSQISDKNF KLVHDIQVAVFKSINRFLSDAPLMLJLWNTWRAKMTLGGKIKANLXFDVFGSFMSTAIKMPDITVCEYD DFGLSEVNTMAGSKCYENRGEFVSRHIGLKKRSZAFVYCLFVAPVHNSVAHFVPSHVAVDIYYVGR SLITPELEVEPKKIEEDTFKASYIPKSDNVHKEFKNEASZADEAGNEAVWYEGYKRTANWMLSLF</p>
AAK39546.1	MlyI	41	<p>>gij13786046 gbiAAK39546.1 AF355462_2 MlyI [Micrococcus lysae]</p> <p>MASLSKTHLFGFTSPRTIEKIIPELDILSQFSGKVVGENQINFDALFNISDFYSGTTFYQDPALAAD RITRAPKALGFIGLKPVIQLTKAGNQLVNRKRLPELFTKQLLKQLFSPYHQSPYRITQSPVNEVNRPSYLELLR INEIGSISKTEIALFVQLVWVNYKFDKRNKILKFRNTRKNRVSQWKTYYQEFKQISIFADNEVTAK NFRITSGSDGSPKAVTEEGNMDYADAFRYIRGTQLVTIDONHLKISGLKQDQVDFLLKNTDRNAL NLRLMEYNYLDFDQLVLEDNSGLNSKINQIDDSINVESLKIDAKDLNLDLXIQRKAKTLEDTVNH LKLRSDIEDILOVFAKIKRQVQVDFLLEWNTWRAFALNHTQAIEGNFTVDLQMPANTAPGKRPDIE INYGSPFCIVEVTMSGGTQFNMEGSSVPRHYGDVLRKVDHDAVCIFLAPVAPQTRAFFNLNRLSTAH YGGKTKIIPMSLDGFCELOVGTIHNFGDINKLKNLNLINLESEDEIWPFEIISKIETWAI</p>
YP_004134094.1	AmI	42	<p>>gij1319768594 ref YP_004134094.1 endonucleasa de restricción de tipo II, AmI [Geobacillus sp. Y412MC52]</p> <p>MNRKNTKRYWFITREERDRFHQEA LLALQKATDDFR LKKA GNRVHKRYEELANMGIKRNNVSHDGGG GRTWAMMLKTFSYCYVDDDDGYIRLTQVYGEKLIQGEKVYENTRQVLT LQYDNAVFLQFGRPKYDEGFR I RFLVFLIKLANDESLDFYVTKEEITVFANTAQRDQLDEIVHRILAFRAKAFEREEMKQDIAAFDHR E RSDKGRDRFYEAHSDVAHTFMLSDYTGVEYIRGKALKGDSKINEIKQEAIEIEKRYFNTRYNISLE RNNENSGLDVDSYKASRYGNTKPAANSKLRAMAERI LAGFPSIEGSKEEIAGALOKYLSPRDIEKVIH EIVENKDDFEGINSDFVEYI LNEKDNLAFEDKTGGIFSA LQFDVAMRPAKNCERTEIEIIARYCGSKFC IIDAKNYAGKFP LSSSLVSHMASEYIPNYTGYEGKELTFGYVTANDFVSGNNEKISDKAKRITGNPIS GFLVTA RLLGLFDYCIENDVFL EDRALFLVAVKNGKYSLEALLREKETI</p>

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
AA Y97906.1	Mva12691	43	<p>>gi 68480350 gb AA Y97906.1 Mva12691 endonucleasa de restricción [<i>Kocuria varians</i>]</p> <p>MYLNTAVENIYGDNI VECRAFVHILEGFKLANISITGEYDLQNIITPKFCIVIDKRYIFIFIPGTSAS RWNEDIYKELVLNNGCE LKECADAILTRIFSEDEELVLSMFSAAALPACNNTWQRS GRAYSLZAAITPY FYIVLQGGKKEIKKGGKSKKFAKLPNPFALSFSVILNTIKKFAFSLIVYDCAFEDGALSIDLYSNCYGI DDFSLYLFKDLIIEENWLRHLANNIYNKVVFFLQLRASVDENGKNE SGGKTYAIFERKDPYRGLTVVVRERKI PWKRTATKIFENF LKNAQAFIFRDLIFLSTKSYGVSKDSLP LTFIFSEHREVEVANYICNDLYIDKVVSD EFVWMIYKXEDLALCIIINGFKP GGGDSKDFURGLPFFTRKLLILDLIILIMEGPAPF IQMDYLDSDDEKLNK INGLWQSIFAFSDAILVDSSTRWNKFTVYNAYLKRWVQREKESNTPISYFPRKSVGCHVDVTSLSHILE TYIGKHFEACNPFPGGWSGVSLKNNIIVKRWTSKYRVSQGGTKRPDHIYQIVYNSTDTLLIILSKGIKN DLLESKEANYGIMINYLKMLMARDYTAVKKGGERKNLHGQMTLDKFLTESAVAYLFTTDEFDNEYTSAAE LLVHSNTQLAFALSIKEXNSVNHIFTANTVAYNFAEYILEIIRNSHLP LKIKIYKPI</p> <p>>gi 313667100 gb ADR72996.1 BsrI [<i>Geobacillus stearothermophilus</i>]</p> <p>MNRIYSEVKEGGIIFKEVIQSVLEKANEVUVVNSANLOYSDVSVLSLIRNKKEFOLLASVVRDKEEI P IVNVEFSTAVTDDHELQRADAMFWAYKYPYLKISFMEKKSQTADDFGGRLDASVNDQIIRNYRVD GVVYHIEWESMENSAYVNAELYPSCFDCAPELASSLFLCILEIEKCNIEDYRILLDKLGRQKVAVW GNFREKTLIQWKHEXFDLLERF SKSSRMEXDKNWELKIKVNRVYGHAMDFERTIAFWKVLWLGDEWKI VAEFQLRKTLKGRQSYQSLEDEVSQEKLMNLA SEJIKNGWVISPDKAIEHKLAISSTMI STIDLQTP ERKYITDPSLKGYLQHGILITNIIYKNLLYVDEIRFDIQRKTIASLTWNKEIVNDYYKSLMPLQLDKNLR VLP LTSI KNI SEDLI TWESKE ILLINIGYKILAA SYPEAQQGRC ILVQFTGKTKERKEIDLIALSPKGGV ILLECRDKLSKSDDCCEANNDLINHHYDKVTKLINVLNINNYNNNI IYTGWAGLIGRKNVENLFPVDFVI KFKYDANKMLKLNWEI NSDI LGKHSGSISMEDVAVVRKRS</p>
ADR72996.1	BsrI	44	
AAL86024.1	BsmI	45	<p>>gi 19347662 gb AAL86024.1 BsmI [<i>Geobacillus stearothermophilus</i>]</p> <p>MWVFRJHGDNIIECERVVIDLISKINPQKVRGFI SLSCPF IEIIFKEGHDYFHWRFQMF GENKNTNR WNSNILLDLQKGSFLXETPDVITISLNNKEEILMAREFC SALQAGNQAWQRS GRAYSVGRIGYFYIYI VDFVKYELNNSDRSRKNLRFPPNPAIPYSYISHSKNTSNEIVQAYFGRGEEYQPKYDKLKFDETFI AEDD LADYILAKLQHRDTSNIEQLL IKNKLNKVEFLSKNTKNDNFTYSEWESIYNGTYRITNLP SLGRFKERK KLAEKSLSGKVKEFNIVQRY SVGLASSDLP FGVI RRESRNDFINVCKLXNINDMMLI KELEADLIV CMLGEFKRGDNDREDRGALFIVAMLAGENACI FT FYIYGLIKGAINLIDQINKLAKNGLWKSEVSL DFIVLDQPIIGESYNEFRLLIKNKNKESILRKT SKQONLAVDPFNHYQENDVDIVYSIFKYIVENCFS GMENPPGGDWSGLSITANGHEFRWLSLPAVSENGKRFDHVQIIDLQFKP LILLISIESKEKPNDFE KIGV QLIKYIEYLFDFTPSVQQRKLAGGNWFGNKSIVPNDFILLSAGAPIDYDNLITENDYEKIFEYTCDDLIIA IKNQNPQKWWIKFKPKNTLAEKLVNYIKLNFKSNIFEDIGFFHIEG</p>

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
AD124225.1	Nb.BisCI	46	<p>>gij297185870 gb AD124225.1 BisCI variante de enzima melladora de hebra inferior [construcción sintética]</p> <p>MKRILYLLTEEREKINI IHQI INLEYKATLHEGAKIVFVMEENKFTFIYHVKGI EVEGFDVLIKIVSG HSSFVYDLVFDSDNLAERKNTITIFDLQDYEJLQSYFEGKGIWRIIPSPSDLPKYVVFETKIDDDHESRNI NAYQRSKFEVCELYGKEVKVMYLDISDGRILSGDTHNEGMRMLVTNNVNLVGVFNMYLFFD IKEF INEKNRIADNGPSHNVEIRLKLDEKNIYI SAKLDKNGKKNKNI SNDPNI GAVAI ISATLRNLNWKGD IELLNHLFSSI SSRSGNKLLYIMKLGVRFNINWNNIKNININFFYNIITSEKIVS IYHLLYVED KLSNARVIFDNHAGCGKSYFRTLNKKI IPVCKEIP LPALVIFEDSDQNIVKVI AAKAENVYNGVEQLSTF DKFTESYINKYYF GAAVCESVITWCKSNFYVSFYLDKDGSAVFL</p> <p>>gij297185868 gb AD124224.1 BisCI variante de enzima melladora de hebra superior [construcción sintética]</p> <p>MKRILYLLTEEREKINI IHQI INLEYKATLHEGAKIVFVMEENKFTFIYHVKGI EVEGFDVLIKIVSG HSSFVYDLVFDSDNLAERKNIITIFDLQDYEJLQSYFEGKGIWRIIPSPSDLPKYVVFETKIDDDHESRNI NAYQRSKFEVCELYGKEVKVMYLDISDGRILSGDTHNEGMRMLVTNNVNLVGVFNMYLFFD IKEF INEKNRIADNGPSHNVEIRLKLDEKNIYI SAKLDKNGKKNKNI SNDPNI GAVAI ISATLRNLNWKGD IELLNHLFSSI SSRSGNKLLYIMKLGVRFNINWNNIKNININFFYNIITSEKIVS IYHLLYVED KLSNARVIFDNHAGCGKSYFRTLNKKI IPVCKEIP LPALVIFEDSDQNIVKVI AAKAENVYNGVEQLSTF DKFTESYINKYYF GAAVCESVITWCKSNFYVSFYLDKDGSAVFL</p>
AD124224.1	Nl.BisCI	47	<p>>gij85720924 gb ABC75874.1 R1.BisI [Geobacillus thermoglucosidasius]</p> <p>MKITEGIVHVAMRHEFLKSNQWKLIAQOYFGSDDELTAALNI VDFVARDNGSDPDRRHBLQKIVFDLIAVK NDDLAVTEAKPKYSGJDRDKLLYL LLSERKKHDFYAALEKFA TEENHPELLFVSKLNIIIFGLAFSASENKFK KDPGEVYIRVSGI EFAFMEGYDWG</p>
ABC75874.1	R1.BisI	48	<p>>gij85720924 gb ABC75874.1 R1.BisI [Geobacillus thermoglucosidasius]</p> <p>MKITEGIVHVAMRHEFLKSNQWKLIAQOYFGSDDELTAALNI VDFVARDNGSDPDRRHBLQKIVFDLIAVK NDDLAVTEAKPKYSGJDRDKLLYL LLSERKKHDFYAALEKFA TEENHPELLFVSKLNIIIFGLAFSASENKFK KDPGEVYIRVSGI EFAFMEGYDWG</p>
ABC75876.1	R2.BisI	49	<p>>gij85720926 gb ABC75876.1 R2.BisI [Geobacillus thermoglucosidasius]</p> <p>MQTEQLMKSLTIYFDDI QEGLWFRNLRH LLESASLEA ITGSLKRNPNLADVLVKYDREDI IILINQTFPI LV IERTIEVPSGSHNVGQRYGR LAAASEAGVFLVYF GPYAARKHGGATEGERYMNLRLEFYALDVMQKNGSAI TTINWFDQNFELIQDF SKDKMKRYLEMFONLKYGIAG NLAIRNSFQAEQLAERKTFVETITNE EQYDVPDQVQILNAREAFFNELGISENKALI CDEVVLYQVGMHYVRSDFITGMALLYKLYLILGSERNRC LILKTFNITIDMWRKYVAFGSRERKDVRIYRSVSDGILFADGYSKEEL</p>

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
AAX14652.1	BbvCI subunidad 1	50	>gj 60202520 gb AAX14652.1 BbvCI subunidad 1 de endonucleasa [Brevibacillus brevis] MINEDEFFIYEQLSHKKNLEQKGNKAEDEETEELVQAKSCYHAFIEGINYDEVTKIDINSVVALEDYIS IAKEIEKHKHKNENRSDYAGSIIPEFLYRIVHVAIVKAGLKPISFTHNIIIEISGAHREGLQIRKMEQ FALGERHEVDVKIASESHRVTISLAVACEVKINIDKNKNGIDFSAENKRTYPSAYTIIITETLDFSPDEN HSSGLIDEIIVLAKQVTRANKVQKALCPVFAELLEIIIEISYRASNVKGVHDRLQEGCKLIRV
AAX14653.1	BbvCI subunidad 2	51	>gj 60202521 gb AAX14653.1 BbvCI subunidad 2 de endonucleasa [Brevibacillus brevis] MENQENPLVYTGCKLERAKKDKTAKKVFEEFCVMEAYNCWKEALCICQRPDKDSVLKLVAAINTYKDA VEPIEDSRINSQEVLPQPSLLEEFFEYLSRIDISIVGVNIPIRHPANGYLSLSFNPHNIEILIQSEFYTV RAKDHDFIIGGSAKLTIQGHGGEGETTINIVFAVAIECKRYLERNNMDEACGTAENLGRAATPYCLAFFVA EYLIKDDGAPELTEIDEIIVILRHQRNSERKFKPFPIDGELIIWDLIQEVNHLGKIWDDPNSALQRGR VFNRE
CAA74998.1	Bpu10I subunidad alfa	52	>gj 2894388 emb CAA74998.1 Bpu10I subunidad alfa de endonucleasa de restricción [Bacillus pumilus] MGVEQEWIKNIIDMYQSPLEIFSHASNLIHQKREKNEKLLKALEIITPNYISYISILLNNHMTKKEI VILVDALNEYMNTLRHPSVKSVESHQADFYSSVLEFFNLAFRNLIKOLNEKIKVNSQKDIIDICIFDPY NEGRVVFKKRVDVALIENKVFVNNVEISDEAFIIVAIETIKINLDKNNLSGIEQEVDSLAKETFFLCLYY CITELADFALEKQNYASTHIDEVFIILRKKRQGVRRGTFLEVVHADLIIEVVEQVGEHLKFKDPIKTLK RHWTEGYLIIKGGK
CAA74999.1	Bpu10I subunidad beta	53	>gj 2894389 emb CAA74999.1 Bpu10I subunidad beta de endonucleasa de restricción [Bacillus pumilus] MTQIDLSNTHKGSILFEKQNVKKEYLQAYRBYLYFRRSIDGLEITNDEAIFKLTQAANNVQDNLVLE ESRPNSQEAFRYTIIEEFYHLFKDLVKKKFCQEPSSIVNCKANSVLSFSPE.SFLGLENPIPIYIHT KQDFVLCQAVDILKISPKNELKNETELVVFVIAIECKTYIERNNLDSCAATASHLKAAMPYCLYIVAS EYMKMDQAYPELTDIDEVFIILCKASVGENTALKKGLFFRHKIDENLWVELEFHVVERHLNRVWRKSPNEALS RGRVIGRP
ABM69266.1	Bmrl	54	>gj 12316737 gb ABM69266.1 Bmrl [Bacillus megaterium] MNYFSLHPNVVYATGRPKGLINMLESVWISNQKFGDGTNYLISGFANYNGGIRPYETFFIHHNHGKVIAT LGGSTQRLSQKQVAELVSRGVQVYIINRKLILHAKLYGSSSNSSGE.SLVVSSGNFTGFCMSQNVESLIL LDNNITSSNGESWNGMNSKLDQXWCIHNLSNSNFTSESNLLYDERTINLTDQTKVILILILGHADT ARIQAPKSKAGEGSQYFWLSKDSYCFPPITIRNKRGTKATYSLIINMNYD.IKYIDSQCRVTFEANN FDFRLGTGKLRVTNVAASDDLAATRVQDSQYELRIIKKGS.SNYDALDSAAVNFIGNRGRKRYGYTPNDEF GRIIGAKF

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
CAC12783.1	BflI	55	<p>>gi 10798463 emb CAC12783.1 endonucleasa de restricción BflI [<i>Bacillus firmus</i>]</p> <p>MNEEGLHENVYATGRPKGLICLNLENVWVSNHTPGEGLYLIISGFENYNGVRFYETFFIEHINQGRVIAI LGGSTIQRLSSRQVVEELNRGVEVHIINRKRILHAKLYGTSNNLGEGLVWSSGNTTGFQMSQNIASLL LDNNTIQSMGE SWNDMI SENLNONWHI HMNTNATDASEGMWLLYDERINLTLDLTERVTLIVTLGHADT ARIQAAPGTTAQQGTQYFWLSKDSYDFPFLTRRRGTKATVSSLLNNYIDINVTOTQCRVTFEAENN FDFRLGTGKLRVTCVAKSNDIAAITHVGDSDYELRIIKQQTPEHSQLDPPYAVSFLCNRGKRFQYISNEEF GRIIGVTE</p>
P05725.1	I-CreI	1	<p>>gi 140470 sp P05725.1 DNE1_CHLRE Nombre recomendado: Completo=Endonucleasa de ADN I-CreI; Nombre alternativo: Completo=23S:rRNA, proteína de intrón</p> <p>MNRYNKEFLILYLAFVTDGCGSIIAQIKPNQSYKFKHQLSLAFQVTKTQRWFLDKLVDEIGVGYVDR GSVSDYILSEIKPLHNFLLQLQPLKLLKQKQANLVKLIWKLP SAKESFDKFLVUCIWWVDQIAALNDSKI RNTTSETVRAVLDLSLSEKXKSSP</p>
Q8UC84.2	hExoI (EXO1_HUMANA)	56	<p>>gi 85700954 sp Q8UC84.2 EXO1_HUMAN Nombre recomendado: Completo=Endonucleasa 1; Corto=hExo1; Nombre alternativo: Completo=Endonucleasa I; Corto=hExoI</p> <p>MGIQLLQFTKEASEPTHVRKYKGVAVDTYCWLHKGALACAEKILAKGEPTRDYVGFQMKFVWMLLSHG IKPILVFDGCTLP SKKEVERSRERROANLLKQKLLREKGVSEARECFTRSNITHAMAEKVIKAARSQ GVDCLVAPYEADAQLAYLNKAGIVQALITTEDSDLLAFGCKKVIILKMDQFGNLEIDQAHIGMCRGLGDFV TEEFRVMCILSGCDYLSLRGIGLAKACKVIALANNPDIVKVIKKGILKNNITVFEEDYINGFLRANN IFLYQLVEDEIKRKLIPLNAYEDDVPETLSYAGQYVDDSIALQIALGNKDNIFEQDDYNDPTAMEAH SRSHWDDKTCRKSANVSSIWHRNYSRPEESTVSDAPQLKENFSTVGVSRVISTKGLNLPKRSSTVYKRP RSAELGEDDLLSQYSLFTKTKRNSBEGNKSLSFSEVTFVDLVNGETNKKSVSTPPTRNKKEATFLQPK NEESGAVVVFGRSRFFCSSDSDCVSNKVISQLDETAVIDKNNLHSEYDQEGXSLAVDIDVARNSS DDIPNNHIFGDHIPDKATVTFDEESYGFESKFTRTISFPTLGHASCFNSWGGGLGDFSRKPSFSESTAL QQFRKSDSPTSLPENNSDVSQIKSESDDESHPLEACSSQESGSEFLQSSNASKLSQCSKDS DSEESDCNLIKLDSDQTSKLRLSHFSSKDDTPLRNKVFGLYKSSADSLSTTKIKPLGPARASGLSKKP ASTQKRKHNAENKPGIQLKLNELHKNFGEKNDSEKLPFCRKP LSPVRDNIQLTPEAEEDIFNKPFGGRV QRAIFQ</p>

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
P39875.2	Exol de levadura (EXO1_LEVADURA)	57	>gi 1706421 sp P39875.2 EXO1_YEAST Nombre recomendado: Completo=Exodesoxirribonucleasa 1; Nombre alternativo: Completo=Exodesoxirribonucleasa I; Corto=EXO I; Corto=Exonucleasa I; Nombre alternativo: Completo=Proteína DHS1 MGIQGLLPQLRPIQNPVSLRRYRGEVLAIDGYAWLHRAACSCAYELAMGRFIDKYLQFFIKRESLLKTFK VEPYLVFDGDAIPVKKSTESKARDARKENKAIAERLMACGEKKNANDYFKCVDIITP EMAKCIIICYCKLN GIRYIVAFLEADSNVYLEQKNIVQGIISDSDLLVEGCRRLITWLANDYGECLCIEI CRDNFIKLPKXFFLG SLTNEEIIITWVCLSGDYINGIEKVLGITANKLVRNTIERIILS IQREGKLMIPDTYINEYEAFLAF QFQRFVCFIHKKIIVSINEIPLYLKDTEKRRLYACI GVIHRETQMKQIVHFDODDIHHLHLKLAQGD NPYDFHQLANREHKLQLA SKSNIEFGKTNINSEAKVKPIESFQKMTKLDHNP KVANNIHS LKQAE DK LTMAIKRRKLSNANVQETLKDTRSKFNKPSMTVVENFKKGSIQDFKEDTNSQSLEEFVSEQLSTQ IPSSFITTNLEDDDLSEEVSEVVDIEEDRKNSECKTICNEIYNIDDDCGDITSEDIETAEASVPTSS TTSFPGSSQHSISGCTKVLQKFRYSSESGVYANRQPLFRHVYKQSRGMVYVYVQNRUDDCCDIDNDGKNQI TQRP SLRKSMTIGARSQRAIVIDMSVDERKSFNSPILHEESKMDIETTESQARPAVRSISLLSQFVYK GK
BAJ43803.1	E. coli Exol	58	>gi 315136644 dbj BAJ43803.1 exonucleasa I [Escherichia coli DH1] MNSGQQRTELEFDVETSTREALEERLQFAALFIDSEENVIQEPVFEYCRPAQVLPQIGVALIIVIT FQGRANQRENEAFAFISELMTVTNTCLGQWVHFDEKVTNFTYENYDEYAMSNQSDNERHLLDY MRACTALRSEETNRPENQDGLRFRLEKLTAKNGTIRHENAHRMADVYATAMGKIVKTSQPLDFVLT HNSKHLMLIDVPSQMSJAVAVRDMISAKRQNTSKVAPJAMBFAKNAVIMWLAQDISELLELDITLR EELVATATLQDQNAVFEVLEVLKNCCEVLGAMTILREDEAGLGTNGSQCCLQELKLRHNSQVREKVAL FAEAEFFTSRNVDAQDINQETDQGRANMKSIVLITFENRLLFDLITFQRRERKLLNFRHADRPFQGITD YASQQWLERHDDVTFTEFEGGFAHELQKVVQVYADQKERVALLKALMCAEELV
Q9BQ50.1	TREX2 humana	59	>gi 47608206 sp Q9BQ50.1 TREX2_HUMAN Nombre recomendado: Completo=Exonucleasa 2 de tres reparaciones principales; Nombre alternativo: Completo=Exonucleasa 3'-5' TREX2 MGRAGSP LFRGSPWRMDCCGSRGRCSPTLCSLRTCYPRGNITMSEAPRAEITVFLDLEATGLPGEVEEI AELSLFAVHRSSLEINPEHDESGALVLRVLDKXLLKCMCPERPTAKASEITGLSSEGLARCKAGEDGAV VRTLQAFLSRQAGPICLVAHNGEDYDFPLLCAEILRRLGARLPRTVCLDITLFAIRGLDRHSSIGTRARGR QGYSLCSLFHRYFRNAEP SAHSAEGDVHTLLLI FLHRAEELLAWADEQARGWAHIEPMYLPFDQDPSLEA
Q91XB0.2	TREX1 de ratón	60	>gi 47608196 sp Q91XB0.2 TREX1_MOUSE Nombre recomendado: Completo=Exonucleasa 1 de tres reparaciones principales; Nombre alternativo: Completo=Exonucleasa 3'-5' TREX1 MGSQTLP HGHMOTLLI FLQLEATGLPSSRPEVTELC LLAVHRRALENTISIQGHPDVPVPRPPVVDKLSLC IAPGKACSPGASRITGLSKAFLEVCGRFDENLAILLRAF LQRQGFCCILVAHNDYRDFLLQITLAR LSTPDLDCGTF CVDSIAALKALEQASDSCNGSKRAYSLSIYTRLYWQAFDSHTAEGDVILLSICQW KPCALLQWDEHARPFSTVXPMYGIPTATGTTNLRPHAAATATIPLATANGSPSNGRSRRPKSPPEKXVPE APSQECLLAPLSLITLLTALATLTYCLFLASPCQ

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
Q9NSU2.1	TREX1 humana	61	>gi 47606216 sp Q9NSU2.1 TREX1_HUMAN Nombre recomendado: Completo=Exonucleasa 1 de tres reparaciones principales; Nombre alternativo: Completo=Exonucleasa 3'-5' TREX1; Nombre alternativo: Completo=DNasa III MGRGARNRQSRIVQGRPEMFCFPPFPLFLRILITLGTHTFPCSSFGSAGTYPTMGSQALPFGFMQTLI FFDMEATGLFSGPKVTELCILLAVHRCALSPFISQGFPPVTFPPFRVVDKLSICVAFGKACSPFAASBITI GLSTAVLRAHGQCFDDNLANLLAFLLRQPCFWCLVAHNGDRYDFPLLQAELENLGLISALDGAFCVDS ITALKALERASPSSEHGPKSYSLSISYTRLYGQSPDSHTAESGDVLLALLSLCQWRPQALLRWVDAHARP FGTIRPMYCVTASARTKPPSAVTTTAHLATIRANTSPRLGESRCGTVDLPPVVDKPCALEREGLLAPLGLLLA ILLAVATLYGLSLATPGE
Q9BG99.1	TREX1 bovina	62	>gi 47606205 sp Q9BG99.1 TREX1_BOVIN Nombre recomendado: Completo=Exonucleasa 1 de tres reparaciones principales; Nombre alternativo: Completo=Exonucleasa 3'-5' TREX1 MGSRALPFGPVTLIFLDLEATGLPFSQPKITELCLLAVHRYALEGLSAFQQGSPFAPVPPRVLDKLSLCL VAPGKVCSPFAASEITGLSTAVLAHGRKAFDADLVNLLNTFLQRQPCFWCLVAHNGDRYDFPLRAELAL LGLASALDDAFCVDSIAALKALEPTGSSSEHGPKSYSLSISYTRLYGQAPDSHTAEGDVLALLSVCCQW RPRALLRWVDAHAKPFSTVYKPMYVITITSTGPNRPSAVTAVIPLARASDTGENLRGDRSPKPAFSPKMPK CAPPCGLLAPLGLLAFLLAVAMLYCLSLAMPQQ
AAH91242.1	TREX1 de rata	63	>gi 60888197 gb AAH91242.1 proteina Trex1 [<i>Rattus norvegicus</i>] MGGKLPFSDMGGTLFLERLGTLPYSGFRTTECLLAVHRIALDNKSSRSGQFFPFRFHVYKSLIC LAVGKPCSGASBITGUTTAGLEAHGSGQFNQMLATILQWYLRQSPQPCCLVAHNGDRYDFLLQALLAE LWITSEFQSTGVDGTAALATLEQASSTFEGFRMYSYLSGTYTLYL5QAFTEGRTAESDNLALLSLCCW RQALLQWYKWRASFTIWRKSHGGKATIGIARPELQCMATISELALANLSEFSNRSRSGKRRUTREPEKXV PEAFPRGLLAPLGLLAFLLAVAMLYCLSLAMPQQ
AAH63664.1	DNA2 humana	64	>gi 39793966 gb AAH63664.1 proteina DNA2 [<i>Homo sapiens</i>] FALPASPMEQLNELELLNEKSFWEAELEAELEFQKXVVASFFRTVLSLCMDNRYLVLAVNTVQNKEGNCE KRLVITASQLENKEICILRNDWCSPVPEFGDIHLEGDCTSDTWIIDDFGYLIDYDMLISGTSIASS IECNRRAVLSETFRSDPATRQMLIGTVLHEVFOQALNNSYAPKQLQELAFQTYQEIHLKEMYHNLNSQ DEIKQEVEDYLPSPCEKACDFMHNKTSIDPPEQMQLSLESDNERNSTICNIEVVKPMDIEESIWSRPFGLK GRIDVTVGVKIHRYKTKYKINPLELKTGKESNIEHRSQVVLYTLLSQERRADPEAGLLLYLKTGQMYF VHANHDKRKLKLANQMAFSDFHRISKSATRQKTQLASLPQIIEKERTCKYCSQIGNCALYSRAVEQGM DCSVP IVMFLKIEETQHLQTHLEYEJLWCLMLETESQKDNKHNQNIWLMASEMERKSGSCIIGNLI RMEHVKIVCDGQYLENEQCKGCAIPVTLNLAGDRVIVSGEERSLFAISGIVYKELNMTTIVICLDRNLSV LPESTLFRADQEKNCIDITFLGNLSKUMENTFVSKKLRDLIDFREEQFI SYLSVLPHEAKDTVACIL KGLNKPQQAANKVLLSKDYTLIVGPNGTGKTTITLVPAPQVEKXGQSVNTEAKLIVEITLSEIVKAG CSESDIGIIPYRQQLKIINDLLARSIGMVEYNTVDYKQGRUKSIVLVSFVRSNRDGTVSGEILLKDWRRIN VAITRAKRLILLCCVPSLNCYPPLEKILLNHNSEKLIIDLPSSREHESICHILGDFQRE

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
P38859.1	DNA2 de levadura (DNA2_LEVADURA)	65	<p>>gll731738 sp P38859.1 DNA2_YEAST Nombre recomendado: Completo=Helicasa dependiente de ATP de replicación de ADN DNA2</p> <p>MEGIPQNKRSAS I SVSPAKKTRKEI I QND-EXA I LSKQTKRKKKYAFAP I NNLNKNTKVSNA5VLR5I AVSQCVRNTSNTKD INKAVSESVKQLENQVRFKREMSNLSRHDFDTQDEGMEVIVKYEPLQDRMSDK TTSAAEYSDUYEDVQNPSSPTIVFNRLKTVLSFTN IQVFNADVNLQIQENGENE DVRE KPAE I STRE SLRN IDD I LDD IEGD LTIKPI I I I K F S D L F S S P I K A P N V E K K A E V N A E V D K M D S I G D S N D G D D S L I D I L T O K Y V E R K S E S Q I T I O G N T N C K S G A Q E S C G K N D N T K S G E I I D H E W D N Q A I T G N A F Y E N E D S N C R I K N E I K H E Y N S D S E S D D S A I E L L N E T Q T Q E P N T I E Q D L D K V E K M V S D D L A T A T D S I L S A V A L R A K S G A P R D G W R I V I V S L R S V E L F K I G I Q K I I E C I D G K G E Q S V V V N R H P V Y I E F E A G D V I H I I E G K N I E N K R L L S D D K N E K T Q L A N D N I L V I N F D V L F S A I S V G S S V G L R S I I Q M Q F Q P R G E P S I V N T L G N I V H E L L Q D S I K Y K L S H N N I S M E I I I Q K L D S L E T Y S F I I I C N E E I Q Y V K E L V M K E A E N I L Y F V N K F V S K S N Y C Y T S I S G T A N I Q P I S I S N V I D I E E N I W S F I Y G L K G F I D A T E A N V E N N K K H I V P L E V K T O K S R S V S Y E V Q G L I Y T L L L N D R Y E I P I E F L L Y F T R U K N N T K F S V L R S I K H I L M S H R M S N F K R Q L Q E V F G Q A Q S R F E L P L L R D S S C D S C F I K E S C M V I N K L I E D G T P E S G L V E G E F I I L T N H L S O N L A N Y K E F F T R Y N D L I T K E S S I T C V N K E I L F L L D G S T R E S R S R C L S G L S V S E V V E H E K T I G A Y I Y C F S R R R D N N S Q S M L S S Q I A A N D F V I I S D E G H F C L C Q S S V Q F I N P A K I G I S V K R K L I N N K L L O K E K G V T I I Q S V V E S E L E Q S S L I A I Q N L V T Y R I D K N D I Q S L S L A R F N L L S I F L P A V S F G V D I V D R S K L C R K T K R S D G N E I I R S L L V D N R A P K F R H A N D P V I P Y K L S K D T T I N I N Q K E A I D K V M R A E D Y A L I I G M P G T K T V T A E I I K I I V S E G K R V I L T S Y T H S A V D N I I K L R A N T N I G I M E L G M K H V H E D I Q Y V F N Y A S V R G Y N D Y L S K I N G T S V V A I T C L G I N D I L F I L N E K D F D V I L D E A S Q I S M E V A L G P L R Y G N R F I M V G D H Y Q L P P L V K N D A A R L G G L E S L F K T F C E M H P E S V A S L T L Q Y R M O G D I V I L S N F L I Y D N K L K C G N N E V F A Q S L E L P M P E A L S R Y N E S A N S K Q W L E I L E P T R K Y V F L N Y I N C P D I I E Q S E R D N I T N E G E A E L T I Q C V E G M L L S G V P C E D I G V M T I Y R A Q L L L K K I E N K N V Y D G L E I L L T A D Q F O G R A K K C I I I S M V R E N S Q L N G A L L K E L R R V N V A M T R A K S K L I I I G S K T I G S V P E I K S F V N L E E R M W V Y T T N C K D A L Y K Y R F P R S N A I D E A R K G C K R I G A K F I T S K S F V S D K P I I K E I I Q E Y E S</p>
AA445863.1	VP16	66	<p>>gll330318 gb AA45863.1 VP16 [Human herpesvirus 2]</p> <p>MDLLEHDDLFADROCVSEFFPFAAGCGHNTFAPDFIATVTELESGQLMSEFPEFVFAALFNHLLDGLGE AGEALDTLQIWNDDLFQGFPTFRDQDREDTLSTLESDVLDQSCAVDVENDFDIQRANDVAEFTILFAE RDELESTYEMAAFTFRELRRREEDYKTVLANFGALTELRARVNSQHRQARRRGRMGLRSHRTIA DRYVETARLARYLFRLYLFERE IIMRAYASQMSDFPDLCDLESWQLACLPLPQIMFINGELTY SGVFEARLRLELHIREHLNLFVNRMAAEFGAFATTFVLCQKARSSGTEKLLIRAKLIDGTSVAT SEGEVMEHAYRHTFRNNGGLTSGLLDQDQDAFMAGLVAFPMGFGRAGDRE RRLSTIARITDWE LQDELLELGEVQMTFADALDQDFLSDLDVESSPSEGMCTHFPVBYGALDVGDFEFGNE TDAMGIDITDQ</p>

(continuación)

ID de GENBANK/SWISSPROT	NOMBRE	SEQ ID NO	SECUENCIA FASTA
ACM07430.1	Colicina E9	366	<p>>g 221185856 gb ACM07430.1 colicina E9 [Escherichia coli]</p> <p>MSGDGRGHNTGAHSTSGNINGGPTGIGVSGGASDGGGWSSENNPWSGGGGSIHWGGSGRNGGGNGN SGGSSGIGGNLQSAVAFAVAFGFFALSTFGAGGLAVSISASELSAALAGI IAKLKVNLKFIFFGVVLSLSL IPSEIAKDDPNMMSKIVTSLPADDI TESPYSILP LDKAIVNVVAVVDDVKDERQNI SVVSGVPMSEVFW DAKPTERPGWF TASFPEAPVLANI SVNDSTFAVQTLSPGVNTNTDKDVRP AGFTGGNTRDAVIRFE KDSSG HNAVYVSQVLSLSPQVKKQRQDEENRRQEWDA THPVEAERNYERAPAEINQANEDVARNQERQAKAVQ VYNSKSELDAAANKILLADIAIAELKQFNRF AHDFFMAGGHHMKQMAGLKAKQRAQTIDVNNKLAAFDAARAEKS DADAALSAACERRKQKENKEKDAKDKLDESKENKPKGKATGKGVGDKWLDGKXDSGAP IPDRIADXL RDKFEFKSPDDFRKAVWEIEVSKDFELSKNLNP SNKSSVSKGYSPFTEKNQVGGKRVVYELHHDKPIISGGGE VYDMDNIRVITPKRHIDIHGK</p>
NP_775816.1	APFL	367	<p>>g 135233 sp P14870.1 T2F1_PLAOK Nombre recomendado: Completo=Enzima de restricción de tipo 2 FokI; Corto=R FokI; Nombre alternativo: Completo=Endonucleasa FokI; Nombre alternativo: Completo=Enzima de restricción de tipo II FokI; Nombre alternativo: Completo=Enzima de restricción de tipo IIS FokI</p> <p>MELSNVSKIRCFGWQNPQKTEENLKRIVQVFDKNSKVENVRNIKIPIITLVZESKIQEILVAIMNIGSDLIF TYRELIVGTISIRSEAP CDALIGATIADCGNNKQYIDNNSDGE LKNHALGCEISYIKNKSGSEVITDVGQ AYKSDADSAI EKEILL EALISYFPALNILLLEDQGHUTKFDLGRNNGE SGESEFTSLPESGILLDTLAN AMPKDKGSETRNNWESSEGRYRMTGGWLDKLAAPQKRRKFIITFKRPTDRNEFIIRAFKITTGERLKVIR RARGGINCTAVRRAVYKELATNLTIDREYVATRBALLLELLIKAGSLIEQIQDMLKALGDSVSTIEN DIEGLINTGIE IE ERGSH TQLMDHLLQVIE NBOVTRALVSELEKESLPHKLLAYVE HSYIELIELAR RSTCDRI LEMRVMEEFMAYVGYRGRHLGGSEKFDALITVVGSP IDYGVIVDINKAYSGGINKLPIGGADENC RYVERNQTRNKHINPNEWKKYYP SEVTEFEKFLVRSGRFKQNYKAQLITRLNHTNKGAVLSEVRELLIGDE MIKAGILLLEEVRRKFNCEINP</p>
P14870.1	FokI	368	<p>>g 221185857 gb ACM07431.1 colicina E9 proteina inmunitaria [Escherichia coli]</p> <p>MELKHSISDYTEAREFLQLVITICNADTSEELVLKLVTHFEEMTzHP SGSDLIYYPKEGDDDSPSGIVNT VYQWRARANGKSGFKQG</p>

5 Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio catalítico al armazón de TALE central de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1, NFS2, CFS1, RM2, BQY, QGPSG, LGPDGRKA, 1a8h_1, 1dnpA_1, 1d8cA_2, 1ckqA_3, 1sbp_1, 1ev7A_1, 1alo_3, 1amf_1, 1adjA_3, 1fcdC_1, 1a13_2, 1g3p_1, 1acc_3, 1ahjB_1, 1acc_1, 1af7_1, 1heiA_1, 1bia_2, 1igtB_1, 1nfkA_1, 1au7A_1, 1bpoB_1, 1b0pA_2, 1c05A_2, 1gcb_1, 1bt3A_1, 1b3oB_2, 16vpA_6, 1dhx_1, 1b8aA_1 y 1qu6A_1, como se enumera en la tabla 3 (de la SEQ ID NO: 67 a la SEQ ID NO: 104 y de la SEQ ID NO: 372 a la SEQ ID NO: 415). Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio catalítico al armazón de TALE central de acuerdo con el método de la presente invención puede seleccionarse del grupo que
10 donde no se necesita un conector peptídico para fusionar un dominio catalítico al armazón de TALE para obtener una cTALEN.

Tabla 3: Lista de conectores peptídicos que pueden usarse en TALEN compactas o TALEN compactas potenciadas.

Nombre (PDB)	Aminoácidos	Tamaño	Longitud	Secuencia	SEQ ID NO
1a8h_1	285 - 287	3	6,636	NVG	67
1dnpA_1	130 - 133	4	7,422	DSVI	68
1d8cA_2	260 - 263	4	8,782	IVEA	69
1ckqA_3	169 - 172	4	9,91	LEGS	70
1sbp_1	93 - 96	4	10,718	YTST	71
1ev7A_1	169 - 173	5	11,461	LQENL	72
1alo_3	360 - 364	5	12,051	VGRQP	73
1amf_1	81 - 85	5	13,501	LGNSL	74
1adjA_3	323 - 328	6	14,835	LPEEKG	75
1fcdC_1	76 - 81	6	14,887	QTYQPA	76
1ai3_2	265 - 270	6	15,485	FSHSTT	77
1g3p_1	99 - 105	7	17,903	GYTYINP	78
1acc_3	216 - 222	7	19,729	LTKYKSS	79
1ahjB_1	106 - 113	8	17,435	SRPSESEG	80
1acc_1	154 - 161	8	18,776	PELKQKSS	81
1aif7_1	89 - 96	8	22,502	LTTNLTAF	82
1heaA_1	322 - 330	9	13,534	TATPPGSVT	83
1bia_2	268 - 276	9	16,089	LDNFINRPV	84
1igtB_1	111 - 119	9	19,737	VSSAKTTAP	85
1nfkA_1	239 - 248	10	13,228	DSKAPNASNL	86
1au7A_1	103 - 112	10	20,486	KRRTTISIAA	87
1bpoB_1	138 - 148	11	21,645	PVKMFRHSSL	88
1b0pA_2	625 - 635	11	26,462	APAETKAEPMT	89
1c05A_2	135 - 148	14	23,819	YTRLPERSELP AEI	90
1gcb_1	57 - 70	14	27,39	VSTDSTPVTNQKSS	91
1bt3A_1	38 - 51	14	28,818	YKLP AVTTMKV RPA	92
1b3oB_2	222 - 236	15	20,054	IARTDLKKNRDYPLA	93
16vpA_6	312 - 332	21	23,713	TEEPGAPLTPPTLHGNCARA	94
1dhx_1	81 - 101	21	42,703	ARFTLAVGDNRVLDMASTYFD	95
1b8aA_1	95 - 120	26	31,305	IVL NRAETPLDPTGKVKAEIDTR	96

(continuación)

Nombre (PDB)	Aminoácidos	Tamaño	Longitud	Secuencia	SEQ ID NO
1qu6A_1	79 - 106	28	51,301	ILNKEKAVSPLLLTTTNSSEGLSMGNY	97
NFS1	-	20	-	GSDITKSKISEKMKGGQFPG	98
NFS2	-	23	-	GSDITKSKISEKMKGLGPDGRKA	99
CFS1	-	10	-	SLTKSKISGS	100
RM2	-	32	-	AAGGSALTAGALSLTAGALSLTAGALSGGGGS	101
BQY	-	25	-	AAGASSVSASGHIAPLSLPPSPPSVGS	102
QGPSG	-	5	-	QGPSG	103
LGPDGRKA	-	8	-	LGPDGRKA	104
TAL1	-	15	-	SGGSGSNVSGSGSG	372
TAL2	-	20	-	SGGSGSLTTLTAFSGSGG	373
TAL3	-	22	-	SGGSGSKRRTTISIAASGSGG	374
TAL4	-	17	-	SGGSGSVGRQPSGSGG	375
TAL5	-	26	-	SGGSGSYTRLPERSELPAEISGSGG	376
TAL6	-	38	-	SGGSGSIVLNRAETPLDPTGKVKAEIDTRSGSGG	377
TAL7	-	21	-	SGGSGSTATPPGSVTSGSGG	378
TAL8	-	21	-	SGGSGSLDNFINRPVSGSGG	379
TAL9	-	21	-	SGGSGSVSSAKTTAPSGSGG	380
TAL10	-	22	-	SGGSGSDSKAPNASNLGSGSG	381
TAL11	-	23	-	SGGSGSPVKMFDHRHSLGSGSG	382
TAL12	-	23	-	SGGSGSAPAEKAEPMTSGSGG	383
TAL13	-	26	-	SGGSGSVSTDSTPVTNOKSSSGSGG	384
TAL14	-	16	-	SGGSGSDSVISGSGG	385
TAL15	-	33	-	SGGSGSARFTLAVGDNRVLDMASTYFDSGSGG	386
TAL16	-	17	-	SGGSGSLQENLGGSGG	387
TAL17	-	19	-	SGGSGSGTYINPSSGSGG	388
TAL18	-	26	-	SGGSGSYKLPVTTMKVRPASGSGG	389
TAL19	-	16	-	SGGSGSLEGGSGSGG	390
TAL20	-	16	-	SGGSGSIVEASGSGG	391
TAL21	-	18	-	SGGSGSQTYOPASGSGG	392
TAL22	-	27	-	SGGSGSIARTDLKKNRDYPLASGSGG	393

(continuación)

Nombre (PDB)	Aminoácidos	Tamaño	Longitud	Secuencia	SEQ ID NO
TAL23	-	18	-	SGGSGSLPEEKGGSGSG	394
TAL24	-	16	-	SGGSGSYTSTSGSG	395
TAL25	-	20	-	SGGSGSRPSESEGGSGSG	396
TAL26	-	17	-	SGGSGSLGNLSGGSG	397
TAL27	-	19	-	SGGSGSLTKYKSSSGSG	398
TAL28	-	33	-	SGGSGSTEPEGAPLTPPTLHGNQARASGGSG	399
TAL29	-	18	-	SGGSGSFSHTTSGSG	400
TAL30	-	20	-	SGGSGPELKQKSSSGSG	401
TAL31	-	40	-	SGGSGSILNKEKKAVSPLLLTTNSSEGLSMGNYSGGSG	402
TAL32	-	31	-	ELAEFHARYADLLLRDLRERPVSLVRGPDG	403
TAL33	-	31	-	ELAEFHARPDPLLRDLRERPVSLVRGLGSG	404
TAL34	-	26	-	ELAEFHARYADLLLRDLRERSGGSG	405
TAL35	-	31	-	DIFDYYAGVAEVMGLGHAGRPA TRKRWPNSG	406
TAL36	-	31	-	DIFDYYAGDPVVMGLGHAGRPA TRKRWLGGSG	407
TAL37	-	26	-	DIFDYYAGVAEVMGLGHAGRSGSG	408
Conector A		37		SIVAQLSRDPALVDFQKLAACLGRRPALDAVKKGL	409
Conector B		37		SIVAQLSRDPAAVSAQKAKAACLGRRPALDAVKKGL	410
Conector C		37		SIVAQLSRDPAVVTFHKLKLAACLGRRPALDAVKKGL	411
Conector D		44		SIVAQLSRDPAQSLAQELSLNESQIKIACLGRRPALDAVKKGL	412
Conector E		40		SIVAQLSRDPALQLPPLERLTDACLGRRPALDAVKKGL	413
Conector F		38		SIVAQLSRDPAIHKKFSSIQMACLGRRPALDAVKKGL	414
Conector G		40		SIVAQLSRDPAAAAAATNDHAAAAACLGRRPALDAVKKGL	415

Dependiendo de su composición estructural [tipo de almacén de TALE central, tipo de dominio o dominios catalíticos con actividades enzimáticas asociadas y finalmente tipo de conector o conectores], una TALEN compacta puede comprender diferentes niveles de actividades enzimáticas diferentes capaces de procesar ADN de forma diferente, produciendo una eficacia de procesamiento de ADN global para dicha TALEN compacta, teniendo cada una de dichas actividades enzimáticas diferentes su propia eficacia de procesamiento de ADN.

Como también se divulga en este documento, el método de acuerdo con la presente invención puede comprender además la etapas de:

- 10 (i) Genomanipulación de al menos un dominio potenciador;
 (ii) Opcionalmente determinación o genomanipulación de un conector peptídico para fusionar dicho dominio potenciador a una parte dicha entidad de TALEN compacta;

15 obteniendo de ese modo una entidad de TALEN compacta con eficacia de procesamiento de ADN potenciada cerca de una secuencia diana de ADN bicatenario única de interés, es decir, una TALEN compacta potenciada. En otras palabras, de acuerdo con el método de la presente invención, dicho monómero de TALEN compacta única comprende además:

- 20 (i) Al menos un dominio potenciador;
 (ii) Opcionalmente un conector peptídico para fusionar dicho dominio potenciador con una parte de dicha entidad activa de monómero de TALEN compacta única.

25 Como también se divulga en este documento, dicho dominio potenciador puede fusionarse al extremo N de la parte de almacén de TALE central de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse al extremo C de la parte de almacén de TALE central de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse a la parte de dominio catalítico de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse entre el extremo N de la parte de almacén de TALE central y la parte catalítica de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse entre el extremo C de la parte de almacén de TALE central y la parte catalítica de dicha entidad de TALEN compacta. Puede preverse la inserción de dicho dominio catalítico y/o dominio potenciador entre dos partes del almacén de TALE central genomanipulado, comprendiendo cada parte un conjunto de RVD. En este último caso, el número de RVD para cada almacén de TALE central genomanipulado puede ser el mismo o no. En otras palabras, puede preverse dividir dicho almacén de TALE central para insertar un dominio catalítico y/o un dominio potenciador entre las dos partes resultantes de dicho almacén de TALE central genomanipulado.

35 Como también se divulga en este documento, dicho dominio potenciador puede ser catalíticamente activo o no, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un dominio proteínico seleccionado de grupo que consiste en Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP1I, I-BasI, I-BmoI, I-Hmul, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu 10I subunidad alfa, Bpu 10I subunidad beta, Bmrl, Bfil, I-CreI, hExoI (EXO1_HUMANA), ExoI de levadura (EXO1_LEVADURA), ExoI de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y la SEQ ID NO: 1, 366 y 367, un mutante funcional, una variante o un derivado de los mismos. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un derivado catalíticamente activo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un derivado catalíticamente inactivo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en I-TevI (SEQ ID NO: 20), ColE7 (SEQ ID NO: 11) y NucA (SEQ ID NO: 26).

50 Como también se divulga en este documento, dicha TALEN compacta potenciada de acuerdo con el método de la presente invención puede comprender un segundo dominio potenciador. Como también se divulga en este documento, dicho segundo dominio potenciador puede tener las mismas característica que el primer dominio potenciador. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte estructural a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte funcional a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte estructural y funcional a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender un dominio catalítico y un dominio potenciador. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender un dominio catalítico y dos

dominios potenciadores. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender dos dominios catalíticos y un dominio potenciador. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender dos dominios catalíticos y dos dominios potenciadores. Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un dominio proteínico derivado de una proteína seleccionada del grupo que consiste en Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP1I, I-Basl, I-Bmol, I-Hmul, I-TeVl, I-TeVll, I-TeVlll, I-Twol, R.Mspl, R.Mval, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococcos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCl, Nt.BtsCl, R1.BtsI, R2.BtsI, BbvCl subunidad 1, BbvCl subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-Crel, hExoI (EXO1_HUMANA), ExoI de levadura (EXO1_LEVADURA), ExoI de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y la SEQ ID NO: 1, 366 y 367, un mutante funcional, una variante o un derivado de los mismos. Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un derivado catalíticamente activo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un derivado catalíticamente inactivo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte estructural a dicha entidad de TALEN compacta potenciada. Como también se divulga en este documento, cualquier combinación de dominios catalíticos y/o potenciadores enumerados anteriormente, como ejemplos no limitantes, puede preverse para fusionarse a dicho armazón de TALE central proporcionando soporte estructural y/o funcional a dicha entidad de TALEN compacta. Más preferiblemente, pueden preverse combinaciones de dominios catalíticos seleccionados del grupo de TevI (SEQ ID NO: 20), ColE7 (SEQ ID NO: 11) y NucA (SEQ ID NO: 26). Opcionalmente, FokI (SEQ ID NO: 368) puede usarse en combinación con otro dominio catalítico de acuerdo con la lista de la tabla 2. Dichas combinaciones de dominios catalíticos y/o potenciadores puede preverse respecto a las aplicaciones prevista para usar el método de la presente invención.

Dependiendo de su composición estructural [tipo de armazón de TALE central, tipo de dominio o dominios catalíticos con actividades enzimáticas asociadas, finalmente tipo de conector o conectores y tipo de dominio o dominios potenciadores], una TALEN compacta potenciada puede presentar diferentes niveles de actividades enzimáticas diferentes capaces de procesar ADN de forma diferente, produciendo una eficacia de procesamiento de ADN global para dicha TALEN compacta potenciada, teniendo cada una de dichas actividades enzimáticas diferentes su propia eficacia de procesamiento de ADN.

Como también se divulga en este documento, la eficacia de procesamiento de ADN de la entidad de TALEN compacta de acuerdo con el método de la presente invención puede potenciarse genomanipulando al menos un dominio potenciador y un conector peptídico, obteniendo de ese modo una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada cerca de una secuencia diana de ADN bicatenario única de interés, es decir, una TALEN compacta potenciada.

Dependiendo de su composición estructural, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta, potenciada, puede tener una actividad enzimática dominante seleccionada del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. Como también se divulga en este documento, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta potenciada de acuerdo con la presente invención puede ser una combinación de diferentes actividades enzimáticas seleccionadas del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. Como también se divulga en este documento, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta potenciada puede ser una de sus diferentes actividades enzimáticas seleccionadas del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. En este caso, la eficacia de procesamiento de ADN global es equivalente a una actividad de procesamiento de ADN entre las actividades enzimáticas mencionadas anteriormente. Como también se divulga en este documento, dicha actividad de procesamiento de ADN de la entidad de TALEN compacta que está potenciada por el potenciador puede ser una actividad cleavasa o una actividad nickasa o una combinación tanto de una actividad cleavasa como de una actividad nickasa.

La potenciación de la eficacia de procesamiento de ADN de una entidad de TALEN compacta puede ser una consecuencia de un soporte estructural por al menos un dominio potenciador. Como también se divulga en este documento, dicho soporte estructural puede potenciar la unión de una entidad de TALEN compacta por dicha secuencia diana de ADN en comparación con la unión de dicha entidad de TALEN compacta de partida por la misma secuencia diana de ADN, ayudando indirectamente de este modo a que el dominio o los dominios catalíticos obtengan una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada. Como también se divulga en

este documento, dicho soporte estructural puede potenciar la actividad catalítica existente de una entidad de TALEN compacta por una secuencia diana de ADN en comparación con la unión de una entidad de TALEN compacta de partida por la misma secuencia diana de ADN para obtener una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada.

5 Como también se divulga en este documento, dicho potenciador de acuerdo con el método de la presente invención puede potenciar tanto la unión de la entidad de TALEN compacta por dicha secuencia diana de ADN como la actividad catalítica del dominio o dominios catalíticos para obtener una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada. Todos estos ejemplos no limitantes dan lugar a una entidad de TALEN compacta con eficacia de procesamiento de ADN potenciada por una secuencia diana de ADN en un locus genómico de interés,
10 es decir, una TALEN compacta potenciada.

La potenciación de la eficacia de procesamiento de ADN de una entidad de TALEN compacta, en comparación con una entidad de TALEN compacta de partida, también puede ser una consecuencia de un soporte funcional por al menos un dominio potenciador. En una realización preferida, dicho soporte funcional puede ser la consecuencia de la hidrólisis de enlaces fosfodiéster adicionales. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una nucleasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una endonucleasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una cleavasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una nickasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una exonucleasa.

En experimentos de manipulación de genoma, la eficacia de endonucleasa de corte infrecuente, por ejemplo, su capacidad de inducir un evento deseado (dirección a gen homólogo, mutagénesis dirigida, eliminación o escisión de secuencia) en un locus, depende de varios parámetros, incluyendo la actividad específica de la nucleasa, probablemente la accesibilidad de la diana y la eficacia y resultado de la ruta o rutas de reparación que provocan el evento deseado (reparación homóloga para dirección génica, rutas de NHEJ para mutagénesis dirigida).

30 La escisión por endonucleasas de corte infrecuente peptídico habitualmente genera extremos cohesivos, con salientes 3' para meganucleasas LAGLIDADG (Chevalier y Stoddard 2001) y salientes 5' para nucleasas de dedos de cinc (Smith, Bibikova *et al.* 2000). Estos extremos, que son el resultado de la hidrólisis de enlaces fosfodiéster, pueden volver a ligarse *in vivo* por NHEJ de una manera sin empalmes (es decir, un religamiento sin marcas). La restauración de una secuencia diana escindible permite un nuevo evento de escisión por la misma endonucleasa y, por tanto, puede tener lugar una serie de ciclos infructuosos de eventos de escisión y religamiento. Las evidencias indirectas han demostrado que incluso en la levadura *Saccharomyces cerevisiae*, dichos ciclos podrían tener lugar tras la escisión continua por la endonucleasa HO (Lee, Paques *et al.* 1999). En células de mamífero, varios experimentos han demostrado que un perfecto religamiento de extremos cohesivos compatibles resultantes de dos DSB inducidos por I-SceI independientes, pero cercanos es un proceso eficaz (Guirouilh-Barbat, Huck *et al.* 2004; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Cheng *et al.* 2008; Bennardo, Gunn *et al.* 2009). La ausencia de la proteína de reparación de ADN Ku no afecta significativamente a la frecuencia global de eventos de NHEJ que vuelve a unir los extremos de las dos DSB; sin embargo, potencia muy fuertemente la contribución de NHEJ imprecisa al proceso de reparación en células inmortalizadas CDO y células ES de ratón (Guirouilh-Barbat, Huck *et al.* 2004; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Cheng *et al.* 2008). Además, la ausencia de Ku estimula los eventos inducidos por I-SceI tales como NHEJ imprecisa (Bennardo, Cheng *et al.* 2008), hibridación monocatenaria (Bennardo, Cheng *et al.* 2008) y conversión génica (Pierce, Hu *et al.* 2001; Bennardo, Cheng *et al.* 2008) en células ES de ratón. Se han hecho observaciones similares con células deficientes para la proteína de reparación XRCC4 (Pierce, Hu *et al.* 2001; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Gunn *et al.* 2009) (aunque la deficiencia de XRCC4 afecta al nivel global de NHEJ en células CHO (Guirouilh-Barbat, Rass *et al.* 2007)) o para DNA-PK (Pierce, Hu *et al.* 2001). Por el contrario, la atenuación de
50 CtlP ha demostrado suprimir "alt-NHEJ" (una forma independiente de Ku y XRCC4 de NHEJ más propensa a producir NHEJ imprecisa), hibridación monocatenaria y conversión génica, mientras que no afecta al nivel global de reunión de dos extremos compatibles generados por I-SceI (Bennardo, Cheng *et al.* 2008). Por tanto, la competición entre las diferentes rutas de reparación DSB puede afectar al espectro o los eventos de reparación resultantes de una DSB inducida por nucleasa.

55 Además, la resección de DSB es importante para determinadas rutas de DSB. La resección de DSB extensiva, que produce la generación de regiones monocatenarias grandes (unos pocos cientos de nucleótidos al menos), se ha demostrado en levaduras que inicia la hibridación monocatenaria (Sugawara y Haber 1992) y la invasión de hebra, la etapa dependiente de ATP que inicia muchos eventos de recombinación homólogos de invasión de dúplex de ADN por una hebra homóloga a esa (White y Haber 1990; Sun, Treco *et al.* 1991) (para una revisión de los mecanismos, véase (Paques y Haber 1999)). En células eucariotas, la resección de DSB depende de varias proteínas incluyendo BLM/Sgs1 y DNA2, EXO1 y el complejo MRN (Mre11, Rad50, Nbs1/Xrs2) y se cree que es el resultado de diferentes rutas. MRN está implicado en un proceso de resección a pequeña escala, mientras que dos rutas redundantes dependientes de BLM y DNA2 por un lado, y de EXO1 por otro lado, estarían implicadas en la resección extensiva (Mimitou y Symington 2008; Nimmonkar, Genschel *et al.* 2011). Además, el procesamiento de los extremos que implica un nucleótido dañado (resultante de la escisión química o de un aducto voluminoso), requiere la proteína CtIP/Sae2

junto con RMN (Sartori, Lukas *et al.* 2007; Buis, Wu *et al.* 2008; Hartsuiker, Mizuno *et al.* 2009). La sobreexpresión de la exonucleasa Trex2 demostró estimular fuertemente NHEJ imperfecta asociada con la pérdida de únicamente unas pocas pares de bases (Bennardo, Gunn *et al.* 2009), aunque inhibía diversos tipos de eventos de reparación de ADN entre secuencias distantes (tal como hibridación monocatenaria, NHEJ entre extremos de diferentes roturas o

5 reparación por NHEJ de una DSB individual que implica microhomologías remotas). En el mismo estudio, se sugirió que Trex2 reseca los salientes 3' dejados por I-SceI de una manera no procesiva. Por tanto, el tipo de ruta estimulada podría depender, a su vez, del tipo de resección (longitud de resección, monocatenario frente a bicatenario, resección de hebra 5' frente a hebra 3').

10 Por tanto, la eficacia de una TALEN compacta, por ejemplo, su capacidad de producir un evento deseado tal como mutagénesis dirigida o dirección a gen homólogo (véase la definición para la definición completa de "eficacia de TALEN compacta"), puede potenciarse mediante una potenciación o modificación de su eficacia de procesamiento de ADN global (véase la definición para la definición completa de "eficacia de procesamiento de ADN global"), por ejemplo, el

15 resultante global o el resultado general de diferentes actividades enzimáticas separadas de dicha TALEN compacta.

Como también se divulga en este documento, la potenciación de la eficacia de procesamiento de ADN global de una entidad de TALEN compacta, en comparación con una entidad de TALEN compacta de partida, puede ser la hidrólisis de enlaces fosfodiéster adicionales en el sitio de escisión.

Dicha hidrólisis de enlaces fosfodiéster adicionales en el sitio de escisión por dicho al menos un potenciador de acuerdo con la invención puede dar lugar a diferentes tipos de resección de DSB que afecta a dicho sitio de escisión de DSB, una única hebra de ADN o ambas hebras de ADN, que afecta a cualquier extremo saliente 5', cualquier extremo saliente 3' o ambos extremos y dependiendo de la longitud de dicha resección. Por tanto, añadir nuevas actividades nickasa o cleavasa a la actividad cleavasa existente de una entidad de TALEN compacta puede potenciar la eficacia de la TALEN compacta potenciada resultante, en un locus genómico de interés (figura 8B-8E). Como un ejemplo no

20 limitante, la adición de dos actividades nickasa en hebras opuestas (figura 8D) o de una nueva actividad cleavasa que genera una segunda DSB (figura 8E) puede provocar una interrupción bicatenaria. Como consecuencia, el religamiento perfecto ya no es posible, y puede estimularse uno o varios resultados de reparación alternativos tales como NHEJ imprecisa, recombinación homóloga o SSA por ejemplo. Como otro ejemplo no limitante, la adición de una única actividad nickasa puede provocar una interrupción monocatenaria, y suprimir la cohesividad de los extremos,

25 que también puede potenciar la eficacia de la TALEN compacta potenciada resultante en un locus genómico de interés, mediante estimulación de uno o varios resultados de reparación alternativos mencionados anteriormente.

En este aspecto descrito de la presente invención, la potenciación de la eficacia de procesamiento de ADN de una TALEN compacta se refiere al aumento en el nivel detectado de dicha eficacia de procesamiento de ADN, frente a una

35 secuencia de ADN diana, o de TALEN compacta potenciada en comparación con la actividad de una primera TALEN compacta frente a la misma secuencia de ADN diana. Dicha primera TALEN compacta puede ser una TALEN compacta de partida o una TALEN compacta que ya se ha genomanipulado o una TALEN compacta potenciada. Puede preverse varias rondas de potenciación a partir de una TALEN compacta de partida o a partir de una TALEN compacta potenciada de partida.

40 En este aspecto descrito del método de la presente invención, la potenciación de la eficacia de procesamiento de ADN de la entidad de TALEN compacta (o TALEN compacta potenciada) se refiere al aumento en el nivel detectado de dicha eficacia de procesamiento de ADN frente a una secuencia de ADN diana de interés o cerca de dicha secuencia de ADN de interés en comparación con la eficacia de una primera TALEN compacta o TALEN compacta de partida

45 frente a o cerca de la misma secuencia de ADN diana. En este caso, la TALEN compacta de partida se toma como el armazón de referencia para medir la eficacia de procesamiento de ADN. Dicha TALEN compacta potenciada es una TALEN compacta genomanipulada que comprende un dominio potenciador de acuerdo con este aspecto descrito de la invención. Dicha TALEN compacta potenciada también puede tomarse como armazón de referencia para potenciación adicional de dicha eficacia de procesamiento de ADN. Como un ejemplo no limitante, dicha eficacia de procesamiento de ADN puede producirse a partir de una recombinación inducida por escisión generada por dicha TALEN compacta potenciada. En este caso, dicho nivel de recombinación inducida por escisión puede determinarse, por ejemplo, por un ensayo de recombinación basado en células como se describe en la solicitud PCT internacional

50 WO 2004/067736. De forma importante, la potenciación de la eficacia en células (generación potenciada de mutagénesis dirigida o recombinación dirigida) puede estar asociada, aunque no necesariamente, con una potenciación de la actividad de escisión que podría detectarse en determinados ensayos *in vitro*. Por ejemplo, actividades fosfodiesterasa adicionales como se describe en la figura 8 podrían afectar escasamente al perfil de escisión, que se detecta por escisión *in vitro* y separación de los productos de escisión en un gel de electroforesis. Sin embargo, como se explica anteriormente, y en leyenda de la figura 8, los extremos de DSB generados de esta manera podrían ser más propensos para inducir reordenamientos genómicos detectables tales como mutagénesis dirigida (por NHEJ imprecisa) o recombinación homóloga. Dicha potenciación en la recombinación inducida por escisión de dicha TALEN compacta potenciada es una potenciación de al menos un 5 % en comparación con el armazón de partida o TALEN compacta de partida, más preferiblemente una potenciación de al menos un 10 %, de nuevo más preferiblemente una potenciación de al menos un 15 %, de nuevo más preferiblemente una potenciación de al menos un 20 %, de nuevo más preferiblemente una potenciación de al menos un 25 %, de nuevo más preferiblemente una

55 potenciación de un 50 %, de nuevo más preferiblemente una potenciación mayor de un 50 %, que provoca una potenciación de la eficacia de procesamiento de ADN de dicha TALEN compacta potenciada de al menos un 5 % en

60

65

comparación con el almacén de partida o TALEN compacta de partida, más preferiblemente una potenciación de al menos un 10 %, de nuevo más preferiblemente una potenciación de al menos un 15 %, de nuevo más preferiblemente una potenciación de al menos un 20 %, de nuevo más preferiblemente una potenciación de al menos un 25 %, de nuevo más preferiblemente una potenciación de un 50 %, de nuevo más preferiblemente una potenciación mayor de un 50 %.

Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio potenciador a una parte de dicha entidad de TALEN compacta de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1, NFS2, CFS1, RM2, BQY, QGPSG, LGPDGRKA, 1a8h_1, 1dnpA_1, 1d8cA_2, 1ckqA_3, 1sbp_1, 1ev7A_1, 1alo_3, 1amf_1, 1adjA_3, 1fcdC_1, 1al3_2, 1g3p_1, 1acc_3, 1ahjB_1, 1acc_1, 1af7_1, 1heiA_1, 1bia_2, 1igtB_1, 1nfkA_1, 1au7A_1, 1bpoB_1, 1b0pA_2, 1c05A_2, 1gcb_1, 1bt3A_1, 1b3oB_2, 16vpA_6, 1dhx_1, 1b8aA_1 y 1qu6A_1 enumerados en la tabla 3 (de la SEQ ID NO: 67 a la SEQ ID NO: 104 y de la SEQ ID NO: 372 a la SEQ ID NO: 415). Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio potenciador a una parte de dicha entidad de TALEN compacta de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1 (SEQ ID NO: 98), NFS2 (SEQ ID NO: 99) y CFS1 (SEQ ID NO: 100). También se describe el caso donde no se necesita un conector peptídico para fusionar un dominio potenciador a una parte de dicha entidad de TALEN compacta para obtener una TALEN compacta potenciada.

Dependiendo de su composición estructural [tipo de almacén de TALE central, tipo de dominio o dominios catalíticos con actividades enzimáticas asociadas, tipo de potenciadores y finalmente tipo de conector o conectores], una TALEN compacta o una TALEN compacta potenciada puede comprender diferentes niveles de actividades enzimáticas diferentes que pueden procesar de forma diferente el ADN como se menciona anteriormente. Añadiendo nuevas actividades enzimáticas a dicha TALEN compacta o dicha TALEN compacta potenciada o potenciando la eficacia de procesamiento de ADN de una o varias de sus actividades enzimáticas constitutivas, se puede potenciar la eficacia de procesamiento de ADN global de una TALEN compacta o TALEN compacta potenciada en comparación con una TALEN compacta de partida o TALEN compacta potenciada.

Se diseñan TALEN compactas para aliviar la necesidad de múltiples restos proteínicos independientes cuando se aborda un evento de procesamiento de ADN. De forma importante, la región "espaciadora" necesaria y los sitios diana dobles esenciales para la función de las TALEN actuales con innecesarios. Como cada extremo del almacén de TALE central es susceptible a fusión, el orden (en el extremo N frente al C) de adición de los dominios catalíticos y de potenciación puede variar con la aplicación. Además, como el dominio catalítico no requiere contactos de ADN específicos, no hay restricciones sobre las regiones que rodean el almacén de TALE central, como ejemplos no limitantes representados en la figura 5: (A) Construcción de fusión en el extremo N para promover la recombinación homóloga inducida por un dominio cleavasa o por un dominio nickasa. (B) Construcción de fusión en el extremo C con propiedades como en (A). (C) La adhesión de dos dominios catalíticos a ambos extremos del almacén de TALE central permite la escisión doble con potenciación en NHEJ. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación.

Las TALEN compactas pueden potenciarse mediante la adición de un dominio para promover actividades existentes o alternativas como ejemplos no limitantes representados en la figura 6: (A) Una TALEN compacta convencional con un dominio potenciador fusionado al extremo C de su parte de almacén de TALE central. (B) El dominio potenciador está fusionado a la TALEN compacta mediante el extremo N de su parte de dominio catalítico. Dicha configuración puede usarse para ayudar y/o anclar la parte de dominio catalítico cerca del ADN para aumentar la actividad de procesamiento de ADN. (C) El dominio potenciador se empareda entre la parte de dominio catalítico y la parte de almacén de TALE central. El dominio potenciador puede promover la comunicación entre los dominios flanqueantes (es decir, para ayudar en la catálisis y/o unión a ADN) o puede usarse para superar el nucleótido T necesario en la posición -1 de todas las dianas basadas en TALE. (D) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo N del almacén de TALE central genomanipulado. (E) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo C del almacén de TALE central genomanipulado. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación.

La naturaleza del dominio o dominios catalíticos comprendidos en la TALEN compacta y la TALEN compacta potenciada depende de la aplicación. Como un ejemplo no limitante, un dominio nickasa permitiría una relación de HR/NHEJ mayor que un dominio cleavasa, siendo de este modo más conveniente para aplicaciones terapéuticas (McConnell Smith, Takeuchi *et al.* 2009; Metzger, McConnell-Smith *et al.* 2011). Por ejemplo, el acoplamiento de un dominio cleavasa en un lado con un dominio nickasa en el otro podría provocar la escisión de una hebra simple de ADN que abarca la región de unión de una TALEN compacta. La generación dirigida de salientes monocatenarios prolongados podría aplicarse en aplicaciones que abordan mecanismos de reparación de ADN. Para inactivación génica dirigida, entonces se prefiere el uso de dos dominios cleavasa. Como también se divulga en este documento, puede favorecerse el uso de dos dominios nickasa. Además, se describe un método para generar varios tipos diferentes de TALEN compactas que pueden aplicarse a aplicaciones que varían desde escisión de ADN dirigida a regulación génica dirigida.

En otro aspecto, La presente invención describe una TALEN compacta, que comprende:

- (i) Un almacén de TALE central que comprende diferentes conjuntos de regiones de dipéptido variable de

repetición (RVD) para cambiar la especificidad de unión al ADN y abordar una secuencia diana de ADN bicatenario única específica de interés, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN;

(ii) Al menos un dominio catalítico en el que dicho dominio catalítico puede procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés cuando se fusiona a dicho armazón de TALE central genomanipulado de (i);

(iii) Opcionalmente un conector peptídico para fusionar dicho dominio catalítico de (ii) a dicho armazón de TALE central genomanipulado de

(i) cuando se necesita;

de modo que dicha TALEN compacta no requiere dimerización para abordar una secuencia diana de ADN bicatenario única específica de interés y procesar el ADN cerca de dicha secuencia diana de ADN bicatenario única de interés. En otras palabras, la TALEN compacta de acuerdo con la presente invención es una unidad de entidad activa que puede abordar, por sí misma, únicamente una secuencia diana de ADN bicatenario única específica de interés mediante un dominio de unión a ADN y de procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés.

La presente invención se refiere a un monómero de TALEN compacta, que comprende:

(i) Un armazón de TALE central que comprende regiones de dipéptido variable de repetición (RVD) que tiene especificidad de unión a ADN en una secuencia diana de ADN bicatenario específica de interés;

(ii) Al menos un dominio catalítico en el que dicho dominio catalítico puede procesar ADN a unas pocas pares de bases de dicha secuencia diana de ADN bicatenario de interés cuando se fusiona al extremo C o N de dicho armazón de TALE central de (i);

(iii) Opcionalmente un conector peptídico para fusionar dicho dominio catalítico de (ii) con dicho armazón de TALE central genomanipulado de (i) cuando se necesita;

en el que dicho monómero de TALEN compacta se ensambla para unirse a dicha secuencia de ADN diana y procesar ADN bicatenario sin requerir dimerización.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado de TALEN compacta de acuerdo con la presente invención puede comprender un dominio del extremo N adicional que produce un armazón de TALE central genomanipulado que comprende secuencialmente un dominio del extremo N y diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado de la TALEN compacta de acuerdo con la presente invención puede comprender un dominio del extremo C adicional que produce un armazón de TALE central genomanipulado que comprende secuencialmente diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés y un dominio en el extremo C, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado de la TALEN compacta de acuerdo con la presente invención puede comprender dominios en el extremo N y en el extremo C adicionales que producen un armazón de TALE central genomanipulado que comprende secuencialmente un dominio en el extremo N, diferentes conjuntos de regiones de dipéptido variable de repetición (RVD) para cambiar la especificidad de unión a ADN y abordar una secuencia diana de ADN bicatenario única específica de interés y un dominio en el extremo C, en que una selección de dominios catalíticos puede adherirse para lograr el procesamiento del ADN.

Como también se divulga en este documento, dicho armazón de TALE central genomanipulado de acuerdo con la presente invención puede comprender las secuencias proteínicas seleccionadas del grupo que consiste en ST1 (SEQ ID NO: 134) y ST2 (SEQ ID NO: 135). Como también se divulga en este documento, dicho armazón de TALE central genomanipulado puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 134 y la SEQ ID NO: 135. Como también se divulga en este documento, dicho armazón de TALE central genomanipulado de acuerdo con la presente invención puede comprender la secuencias proteínicas seleccionadas del grupo que consiste en bT1-Avr (SEQ ID NO: 136), bT2-Avr (SEQ ID NO: 137), bT1-Pth (SEQ ID NO: 138) y bT2-Pth (SEQ ID NO: 139). Como también se divulga en este documento, dicho armazón de TALE genomanipulado puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 136 a la SEQ ID NO: 139.

Como también se divulga en este documento, dichos dominios en el extremo N y el extremo C adicionales del almacén de TALE central genomanipulado pueden obtenerse preferiblemente de TALE natural. Como también se divulga en este documento, dichos dominios en el extremo N y el extremo C adicionales del almacén de TALE central genomanipulado pueden obtenerse más preferiblemente de TALE natural seleccionada del grupo que consiste en AvrBs3, PthXol, AvrHahl, PthA, Tallc como ejemplos no limitantes. Como también se divulga en este documento, dichos dominios en el extremo N y/o dichos dominios en el extremo C pueden ser formas truncadas de dominios en el extremo N y/o el extremo C respectivos de TALE naturales como AvrBs3, PthXol, AvrHahl, PthA, Tallc como ejemplos no limitantes, de las que se obtienen. Como también se divulga en este documento, dichas secuencias de los dominios en el extremo N y el extremo C adicionales del almacén de TALE central genomanipulado pueden seleccionarse del grupo que consiste en ST1 SEQ ID NO: 134 y ST2 SEQ ID NO: 135 como se ejemplifica respectivamente en los armazones proteínicos basales bT1-Avr (SEQ ID NO: 136) o bT1-Pth (SEQ ID NO: 138) y bT2-Avr (SEQ ID NO: 137) o bT2-Pth (SEQ ID NO: 139).

Como también se divulga en este documento, cada RVD de dicho almacén central puede estar hecho de 30 a 42 aminoácidos, más preferiblemente 33 o 34, en el que dos aminoácidos críticos ubicados en las posiciones 12 y 13 median el reconocimiento de un nucleótido de dicha secuencia diana de ácido nucleico; dos aminoácidos críticos equivalentes pueden estar ubicados en las posiciones diferentes a 12 y 13 especialmente en RVD más grandes de 33 o 34 aminoácidos de longitud. Preferiblemente, los RVD asociados con el reconocimiento de los diferentes nucleótidos son HD para reconocer C, NG para reconocer T, NI para reconocer A, NN para reconocer G o A, NS para reconocer A, C, G o T, HG para reconocer T, IG para reconocer T, NK para reconocer G, HA para reconocer C, ND para reconocer C, HI para reconocer C, HN para reconocer G, NA para reconocer G, SN para reconocer G o A e YG para reconocer T, TL para reconocer A, VT para reconocer A o G y SW para reconocer A. Más preferiblemente, los RVD asociados con el reconocimientos de los nucleótidos C, T, A, G/A y G respectivamente se seleccionan del grupo que consiste en NN o NK para reconocer G, HD para reconocer C, NG para reconocer T y NI para reconocer A, TL para reconocer A, VT para reconocer A o G y SW para reconocer A. Como también se divulga en este documento, los RVD asociados con el reconocimiento del nucleótido C pueden seleccionarse del grupo que consiste en N* y los RVD asociados con el reconocimiento del nucleótido T se seleccionan del grupo que consiste en N* y H*, donde * indica una interrupción en la secuencia de repetición que corresponde a una ausencia de resto de aminoácido en la segunda posición del RVD. Como también se divulga en este documento, los aminoácidos críticos 12 y 13 pueden mutarse en otros restos de aminoácido para modular su especificidad hacia los nucleótidos A, T, C y G y, en particular, para potenciar su especificidad. Por otros restos de aminoácido se entiende cualquiera de los veinte restos de aminoácido naturales o derivados de aminoácidos no naturales.

Como también se divulga en este documento, dicho almacén central de la presente invención puede comprender entre 8 y 30 RVD. Más preferiblemente, dicho almacén central de la presente invención puede comprender entre 8 y 20 RVD; de nuevo más preferiblemente 15 RVD.

Como también se divulga en este documento, dicho almacén central puede comprender un único RVD truncado adicional hecho de 20 aminoácidos ubicado en el extremo C de dicho conjunto de RVD, es decir, un semi-RVD en el extremo C adicional. En este caso, dicho almacén central comprende entre 8,5 y 30,5 RVD, refiriéndose "5" al semi-RVD mencionado previamente (o RVD terminal, o semirrepetición). Más preferiblemente, dicho almacén central puede comprender entre 8,5 y 20,5 RVD, de nuevo más preferiblemente, 15,5 RVD. Como también se divulga en este documento, dicho semi-RVD puede estar en un contexto de almacén central que permite una ausencia de especificidad de dicho semi-RVD hacia los nucleótidos A, C, G, T. Como también se divulga en este documento, dicho semi-RVD puede estar ausente.

Como también se divulga en este documento, dicho almacén central puede comprender RVD de diferentes orígenes. Como también se divulga en este documento, dicho almacén central puede comprender RVD originados de diferentes efectores TAL de origen natural. Como también se divulga en este documento, la estructura interna de algunos RVD del almacén central puede estar constituida por estructuras o secuencias originadas de diferentes efectores TAL de origen natural. Como también se divulga en este documento, dicho almacén central puede comprender dominios de tipo RVD. Los dominios de tipo RVD tienen una secuencia diferente de los RVD de origen natural, pero tienen la misma función y/o estructura global dentro de dicho almacén central.

Como también se divulga en este documento, dicho dominio en el extremo N adicional de dicho almacén de TALE central genomanipulado de dicha TALEN compacta de acuerdo con la presente invención puede ser un dominio potenciador. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en la proteína de unión a ARN Puf o la superfamilia de anquirina, como ejemplos no limitantes. Como también se divulga en este documento, dicha secuencia de dominio potenciador puede seleccionarse del grupo que consiste en dominios proteínicos de la SEQ ID NO: 4 y la SEQ ID NO: 5 como ejemplos no limitantes enumerados en la tabla 1, un mutante funcional, una variante o un derivado de los mismos. Como también se divulga en este documento, dicho dominio en el extremo C adicional de dicho almacén de TALE central genomanipulado puede ser un dominio potenciador. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en hidrolasa/transferasa de la familia de *Pseudomonas aeruginosa*, el dominio polimerasa de la familia de ligasa D de *Mycobacterium tuberculosis*, el factor de inicio eIF2 de la familia de *Pyrococcus*, el factor de inicio de la traducción de la familia Aif2 como ejemplos no limitantes. Como también se divulga en este documento,

dicha secuencia de dominio potenciador puede seleccionarse del grupo que consiste en dominios proteínicos de la SEQ ID NO: 6 la SEQ ID NO: 9 como ejemplos no limitantes enumerados en la tabla 1.

- 5 En otra realización preferida, el dominio catalítico que tiene capacidad de procesamiento de ADN cerca de la secuencia
diana de ADN bicatenario única de interés, cuando se fusiona a dicha armazón de TALE central genomanipulado de
acuerdo con la presente invención, se fusiona en la parte del extremo N de dicho armazón de TALE central. En otra
realización preferida, dicho dominio catalítico se fusiona en la parte del extremo C de dicho armazón de TALE central.
Como también se divulga en este documento, pueden fusionarse dos dominios catalíticos tanto a la parte del extremo
N de dicho armazón de TALE central como a la parte del extremo C de dicho armazón de TALE central. Como también
10 se divulga en este documento, dicho dominio catalítico puede tener una actividad enzimática seleccionada del grupo
que consiste en actividad nucleasa, actividad polimerasa, actividad cinasa, actividad fosfatasa, actividad metilasa,
actividad topoisomerasa, actividad integrasa, actividad transposasa o actividad ligasa. Como también se divulga en
este documento, el dominio catalítico fusionado al armazón de TALE central de la presente invención puede ser un
activador o represor de la transcripción (es decir, un regulador de transcripción) o una proteína que interactúa con o
15 modifica otras proteínas tales como histonas. Los ejemplos no limitantes de actividades de procesamiento del ADN de
dicha TALEN compacta de la presente invención incluyen, por ejemplo, creación o modificación de elementos
reguladores epigenéticos, generación de inserciones, eliminaciones o reparaciones específicas de sitio en el ADN,
control de la expresión génica y modificación de la estructura de la cromatina.
- 20 Como también se divulga en este documento, dicho dominio catalítico puede tener una actividad endonucleasa. Como
también se divulga en este documento, dicho dominio catalítico de la TALEN compacta de acuerdo con la presente
invención puede tener actividad de escisión sobre dicho ADN bicatenario de acuerdo con método de la presente
invención. Como también se divulga en este documento, dicho dominio catalítico puede tener una actividad nickasa
sobre dicho ADN bicatenario de acuerdo con el método de la presente invención. Como también se divulga en este
25 documento, dicho dominio catalítico puede seleccionarse del grupo que consiste en las proteínas Mmel, Colicina-E7
(CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G
bovina (NUCG_BOVINA), R.HinP11, I-Basl, I-Bmol, I-Hmul, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA,
NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa
de micrococos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI,
30 BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI,
Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad
alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-CreI, hExoI (EXO1_HUMANA), ExoI de levadura (EXO1_LEVADURA), ExoI
de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2
de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y
35 las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de los mismos. El dominio catalítico
de la TALEN compacta de acuerdo con la presente invención es I-TevI (SEQ ID NO: 20). En otra realización preferida,
el dominio catalítico I-TevI (SEQ ID NO: 20), un mutante funcional, una variante o un derivado del mismo se fusiona
en el dominio del extremo N de dicho armazón de TALE central de acuerdo con la TALEN compacta de la presente
invención. En otra realización preferida, dicha TALEN compacta de acuerdo con la presente invención comprende una
40 secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un
95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ
ID NO: 426-432.
- Como también se divulga en este documento, dicho dominio catalítico de la TALEN compacta de acuerdo con la
presente invención puede ser ColE7 (SEQ ID NO: 11), un mutante funcional, una variante o un derivado del mismo.
45 Como también se divulga en este documento, el dominio catalítico ColE7 (SEQ ID NO: 11), un mutante funcional, una
variante o un derivado del mismo puede fusionarse al dominio del extremo N de dicho armazón de TALE central de
acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico
ColE7 (SEQ ID NO: 11), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del
extremo C de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se
50 divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede
comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más
preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del
grupo de la SEQ ID NO: 435-438.
- Como también se divulga en este documento, dicho dominio catalítico de la TALEN compacta de acuerdo con la
presente invención puede ser NucA (SEQ ID NO: 26), un mutante funcional, una variante o un derivado del mismo.
55 Como también se divulga en este documento, el dominio catalítico NucA (SEQ ID NO: 26), un mutante funcional, una
variante o un derivado del mismo puede fusionarse al dominio del extremo N de dicho armazón de TALE central de
acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico
NucA (SEQ ID NO: 26), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del
extremo C de dicho armazón de TALE central de acuerdo con el método de la presente invención. Como también se
60 divulga en este documento, dicha TALEN compacta de acuerdo con el método de la presente invención puede
comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más
preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del
grupo de la SEQ ID NO: 433-434.
- 65 Como también se divulga en este documento, dicho dominio catalítico puede ser I-CreI (SEQ ID NO: 1), un mutante
funcional, una variante o un derivado del mismo. Como también se divulga en este documento, el dominio catalítico I-

Crel (SEQ ID NO: 1), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del extremo N de dicho almacén de TALE central de acuerdo con el método de la presente invención. Como también se divulga en este documento, el dominio catalítico I-Crel (SEQ ID NO: 1), un mutante funcional, una variante o un derivado del mismo puede fusionarse al dominio del extremo C de dicho almacén de TALE central de acuerdo con la presente invención. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 439-441 y la SEQ ID NO: 444-446.

Como también se divulga en este documento, dicho dominio catalítico puede ser una enzima de restricción tal como Mmel, R-HinPII, R.MspI, R.MvaI, Nb.BsrDI, BsrDI A, Nt.BspD6I, ss.BspD6I, R.PleI, MlyI y AlwI como ejemplos no limitantes enumerados en la tabla 2. Como también se divulga en este documento, dicho dominio catalítico puede tener una actividad exonucleasa. Como también se divulga en este documento, cualquier combinación de dos dominios catalíticos seleccionados del grupo que consiste en la proteínas Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP1I, I-BasI, I-BmoI, I-HmuI, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococcos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-Crel, hExol (EXO1_HUMANA), Exol de levadura (EXO1_LEVADURA), Exol de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de estos dominios proteínicos de los mismos, pueden fusionarse tanto a la parte del extremo N como la parte del extremo C de dicho almacén de TALE central, respectivamente. Por ejemplo, el dominio catalítico I-HmuI puede fusionarse a la parte del extremo N de dicho almacén de TALE central y el dominio catalítico ColE7 puede fusionarse a la parte del extremo C de dicho almacén de TALE central. En otro ejemplo, el dominio catalítico I-TevI puede fusionarse a la parte del extremo N de dicho almacén de TALE central y el dominio catalítico ColE7 puede fusionarse a la parte del extremo C de dicho almacén de TALE central. La tabla 14 a continuación da ejemplos no limitantes de combinaciones de dominios catalíticos que pueden estar comprendidos en un monómero de TALEN compacto. Opcionalmente, FokI (SEQ ID NO: 368) puede usarse en combinación con otro dominio catalítico de acuerdo con la lista de la tabla 2.

Tabla 14: Ejemplos de combinaciones de dominios catalíticos respectivamente fusionados a la parte de extremo N y C de armazones centrales de TALEN compacta que dan lugar a TALEN de escisión doble.

Dominio catalítico fusionado a la parte de extremo N del almacén de TALE central	Dominio catalítico fusionado a la parte de extremo C del almacén de TALE central	TALEN de escisión doble
I-TevI	I-TevI	TevI-TevI
ColE7	ColE7	ColE7-ColE7
NucA	NucA	NucA-NucA
I-TevI	ColE7	TevI-ColE7
I-TevI	NucA	TevI-NucA
ColE7	I-TevI	ColE7-TevI
ColE7	NucA	ColE7-NucA
NucA	I-TevI	NucA-TevI
NucA	ColE7	NucA-ColE7

Como también se divulga en este documento, dicho monómero de TALEN compacta única puede comprender una combinación de dos dominios catalíticos, respectivamente fusionados a la parte del extremo N y a la parte del extremo C de dicho almacén de TALE central seleccionados del grupo que consiste en:

- (i) un dominio Nuc A (SEQ ID NO: 26) en el extremo N y un dominio Nuc A (SEQ ID NO: 26) en el extremo C;
- (ii) un dominio ColE7 (SEQ ID NO: 11) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C;
- (iii) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C;
- (iv) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio NucA (SEQ ID NO: 26) en el extremo C;
- (v) un dominio ColE7 (SEQ ID NO: 11) en el extremo N y un dominio NucA (SEQ ID NO: 26) en el extremo C;
- (vi) un dominio NucA (SEQ ID NO: 26) en el extremo N y un dominio ColE7 (SEQ ID NO: 11) en el extremo C.

Como también se divulga en este documento, dicha TALEN compacta de acuerdo con la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 448 y 450.

Como también se divulga en este documento, dicha TALEN compacta de acuerdo con la presente invención puede comprender una combinación de dos dominios catalíticos, respectivamente fusionados a la parte del extremo C y a la parte del extremo N de dicho armazón de TALE central seleccionados del grupo que consiste en:

- 5 (i) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio FokI (SEQ ID NO: 368) en el extremo C;
- (ii) un dominio TevI (SEQ ID NO: 20) en el extremo N y un dominio TevI (SEQ ID NO: 20) en el extremo C;
- (iii) un dominio scTrex2 (SEQ ID NO: 451) en el extremo N y un dominio FokI (SEQ ID NO: 368) en el extremo C.

10 Como también se divulga en este documento, dicha TALEN compacta de acuerdo con la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 447-450 y la SEQ ID NO: 452.

15 Puede preverse la inserción de dicho dominio catalítico y/o dicho dominio potenciador entre dos partes del armazón de TALE central genomanipulado, comprendiendo cada parte un conjunto de RVD. En este último caso, el número de RVD para cada parte del armazón de TALE central genomanipulado puede ser el mismo o no. En otras palabras, puede preverse dividir dicho armazón de TALE central para insertar un dominio catalítico y/o un dominio potenciador entre las dos partes resultantes de dicho armazón de TALE central genomanipulado. Como también se divulga en este documento, dicha TALEN compacta de acuerdo con la presente invención puede comprender una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 453-455.

20 En otras palabras, se describe que el monómero de TALEN compacta comprende una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo que consiste en la SEQ ID NO: 420-450 y 452 -455.

25 En otra realización preferida de acuerdo con el método de la presente invención, el conector peptídico que puede unir dicho dominio catalítico al armazón de TALE central de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1, NFS2, CFS1, RM2, BQY, QGPSG, LGPDGRKA, 1a8h_1, 1dnpA_1, 1d8cA_2, 1ckqA_3, 1sbp_1, 1ev7A_1, 1alo_3, 1amf_1, 1adjA_3, 1fcdC_1, 1a13_2, 1g3p_1, 1acc_3, 1ahjB_1, 1acc_1, 1af7_1, 1heiA_1, 1bia_2, 1igtB_1, 1nfkA_1, 1au7A_1, 1bpoB_1, 1b0pA_2, 1c05A_2, 1gcb_1, 1bt3A_1, 1b3oB_2, 16vpA_6, 1dhx_1, 1b8aA_1 y 1qu6A_1, como se enumera en la tabla 3 (de la SEQ ID NO: 67 a la SEQ ID NO: 104 y de la SEQ ID NO: 372 a la SEQ ID NO: 415). Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio catalítico al armazón de TALE central de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1 (SEQ ID NO: 98), NFS2 (SEQ ID NO: 99) y CFS1 (SEQ ID NO: 100). También se describe el caso donde no se necesita un conector peptídico para fusionar un dominio catalítico al armazón de TALE para obtener una cTALEN.

30 Dependiendo de su composición estructural [tipo de armazón de TALE central, tipo de dominio o dominios catalíticos con actividades enzimáticas asociadas y finalmente tipo de conector o conectores], una TALEN compacta de acuerdo con la presente invención puede comprender diferentes niveles de actividades enzimáticas diferentes que pueden procesar de forma diferente el ADN, produciendo una eficacia de procesamiento de ADN global para dicha TALEN compacta, teniendo cada una de dichas actividades enzimáticas diferentes su propia eficacia de procesamiento de ADN.

35 Como también se divulga en este documento, la TALEN compacta de acuerdo con la presente invención puede comprender además:

- 50 (i) Al menos un dominio potenciador;
- (ii) Opcionalmente un conector peptídico para fusionar dicho dominio potenciador con una parte de dicha entidad activa de TALEN compacta;

55 obteniendo de ese modo una entidad de TALEN compacta con eficacia de procesamiento de ADN potenciada cerca de una secuencia diana de ADN bicatenario única de interés, es decir, una TALEN compacta potenciada.

En otras palabras, dicho monómero de TALEN compacta única puede comprender además:

- 60 (i) Al menos un dominio potenciador;
- (ii) Opcionalmente un conector peptídico para fusionar dicho dominio potenciador con una parte de dicha entidad activa de monómero de TALEN compacta única.

65 Como también se divulga en este documento, dicho dominio potenciador puede fusionarse al extremo N de la parte de armazón de TALE central de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse al extremo C de la parte de armazón de TALE central de dicha entidad

de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse a la parte de dominio catalítico de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse entre la parte de extremo N del armazón de TALE central y la parte catalítica de dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede fusionarse entre la parte de extremo C del armazón de TALE central y la parte catalítica de dicha entidad de TALEN compacta. Puede preverse la inserción de dicho dominio catalítico y/o dominio potenciador entre dos partes del armazón de TALE central genomanipulado, comprendiendo cada parte un conjunto de RVD. En este último caso, el número de RVD para cada armazón de TALE central genomanipulado puede ser el mismo o no. En otras palabras, puede preverse dividir dicho armazón de TALE central para insertar un dominio catalítico y/o un dominio potenciador entre las dos partes resultantes de dicho armazón de TALE central genomanipulado.

Como también se divulga en este documento, dicho dominio potenciador puede ser catalíticamente activo o no, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un dominio proteínico seleccionado de grupo que consiste en Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP11, I-Basl, I-Bmol, I-Hmul, I-TevI, I-TevII, I-TevIII, I-TwoI, R.Mspl, R.Mval, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-Crel, hExol (EXO1_HUMANA), Exol de levadura (EXO1_LEVADURA), Exol de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de estos dominios proteínicos de los mismos. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un derivado catalíticamente activo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede consistir en un derivado catalíticamente inactivo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte estructural a dicha entidad de TALEN compacta. Como también se divulga en este documento, dicho dominio potenciador puede seleccionarse del grupo que consiste en I-TevI (SEQ ID NO: 20), CoIE7 (SEQ ID NO: 11) y NucA (SEQ ID NO: 26).

Como también se divulga en este documento, dicha TALEN compacta puede comprender un segundo dominio potenciador. Como también se divulga en este documento, dicho segundo dominio potenciador puede tener las mismas características que el primer dominio potenciador. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte estructural a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte funcional a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede proporcionar soporte estructural y funcional a la entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender un dominio catalítico y un dominio potenciador. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender un dominio catalítico y dos dominios potenciadores. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender dos dominios catalíticos y un dominio potenciador. Como también se divulga en este documento, dicha entidad de TALEN compacta potenciada puede comprender dos dominios catalíticos y dos dominios potenciadores.

Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un dominio proteínico derivado de una proteína seleccionada del grupo que consiste en Mmel, Colicina-E7 (CEA7_ECOLX), Colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMANA), Endo G bovina (NUCG_BOVINA), R.HinP11, I-Basl, I-Bmol, I-Hmul, I-TevI, I-TevII, I-TevIII, I-TwoI, R.Mspl, R.Mval, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococos (NUC_SHIFL), endonucleasa yncB, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I subunidad grande), ss.BspD6I (R.BspD6I subunidad pequeña), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-Crel, hExol (EXO1_HUMANA), Exol de levadura (EXO1_LEVADURA), Exol de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_LEVADURA) y VP16, como se enumera en la tabla 2 (de la SEQ ID NO: 10 a la SEQ ID NO: 66 y las SEQ ID NO: 1, 366 y 367), un mutante funcional, una variante o un derivado de estos dominios proteínicos de los mismos. Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un derivado catalíticamente activo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte funcional y/o estructural a dicha entidad de TALEN compacta potenciada. Como también se divulga en este documento, dicho segundo dominio potenciador puede consistir en un derivado catalíticamente inactivo de los dominios proteínicos enumerados anteriormente y en la tabla 2, proporcionando soporte estructural a dicha entidad de TALEN compacta potenciada.

Como también se divulga en este documento, cualquier combinación de dominios catalíticos y/o potenciadores enumerados anteriormente, como ejemplos no limitantes, puede preverse para fusionarse a dicho armazón de TALE central proporcionando soporte estructural y/o funcional a dicha entidad de TALEN compacta. Más preferiblemente, combinaciones de dominios catalíticos enumeradas en la tabla 14. De nuevo más preferiblemente, pueden preverse combinaciones de dominios catalíticos seleccionados del grupo de TevI (SEQ ID NO: 20), CoIE7 (SEQ ID NO: 11) y

NucA (SEQ ID NO: 26). Opcionalmente, FokI (SEQ ID NO: 368) puede usarse en combinación con otro dominio catalítico de acuerdo con la lista de la tabla 2. Dichas combinaciones de dominios catalíticos y/o potenciadores puede preverse respecto a las aplicaciones prevista para usar el método de la presente invención. Dependiendo de su composición estructural [tipo de armazón de TALE central, tipo de dominio o dominios catalíticos con actividades enzimáticas asociadas, tipo de conector o conectores y tipo de dominio o dominios potenciadores], una TALEN compacta potenciada puede presentar diferentes niveles de actividades enzimáticas diferentes capaces de procesar ADN de forma diferente, produciendo una eficacia de procesamiento de ADN global para dicha TALEN compacta potenciada, teniendo cada una de dichas actividades enzimáticas diferentes su propia eficacia de procesamiento de ADN.

Como también se divulga en este documento, la eficacia de procesamiento de ADN de la entidad de TALEN compacta de acuerdo con la presente invención puede potenciarse genomanipulando al menos un dominio potenciador y un conector peptídico, obteniendo de ese modo una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada cerca de una secuencia diana de ADN bicatenario única de interés, es decir, una TALEN compacta potenciada.

Dependiendo de su composición estructural, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta, potenciada, puede tener una actividad enzimática dominante seleccionada del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. Como también se divulga en este documento, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta potenciada puede ser una combinación de diferentes actividades enzimáticas seleccionadas del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. Como también se divulga en este documento, la eficacia de procesamiento de ADN global que está potenciada en dicha TALEN compacta potenciada puede ser una de sus diferentes actividades enzimáticas seleccionadas del grupo que consiste en una actividad nucleasa, una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. En este caso, la eficacia de procesamiento de ADN global es equivalente a una actividad de procesamiento de ADN entre las actividades enzimáticas mencionadas anteriormente. Como también se divulga en este documento, dicha actividad de procesamiento de ADN de la entidad de TALEN compacta que está potenciada por el potenciador puede ser una actividad cleavasa o una actividad nickasa o una combinación tanto de una actividad cleavasa como de una actividad nickasa.

La potenciación de la eficacia de procesamiento de ADN de una entidad de TALEN compacta de acuerdo con la presente invención puede ser una consecuencia de un soporte estructural por dicho al menos un dominio potenciador. Como también se divulga en este documento, dicho soporte estructural puede potenciar la unión de una entidad de TALEN compacta de acuerdo con la invención por dicha secuencia diana de ADN en comparación con la unión de una entidad de TALEN compacta de partida por la misma secuencia diana de ADN, ayudando indirectamente de este modo a que el dominio o los dominios catalíticos obtengan una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada. Como también se divulga en este documento, dicho soporte estructural puede potenciar la actividad catalítica existente de una entidad de TALEN compacta por una secuencia diana de ADN en comparación con la unión de una entidad de TALEN compacta de partida por la misma secuencia diana de ADN para obtener una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada.

Como también se divulga en este documento, dicho potenciador puede potenciar tanto la unión de la entidad de TALEN compacta por dicha secuencia diana de ADN como la actividad catalítica del dominio o dominios catalíticos para obtener una entidad de TALEN compacta con actividad de procesamiento de ADN potenciada. Todos estos ejemplos no limitantes dan lugar a una entidad de TALEN compacta con eficacia de procesamiento de ADN potenciada por una secuencia diana de ADN en un locus genómico de interés, es decir, una TALEN compacta potenciada.

La potenciación de la eficacia de procesamiento de ADN de una entidad de TALEN compacta de acuerdo con la presente invención, en comparación con una entidad de TALEN compacta de partida, también puede ser una consecuencia de un soporte funcional por dicho al menos un dominio potenciador. Como también se divulga en este documento, dicho soporte funcional puede ser la consecuencia de la hidrólisis de enlaces fosfodiéster adicionales. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una nucleasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una endonucleasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una cleavasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una nickasa. Como también se divulga en este documento, dicho soporte funcional puede ser la hidrólisis de enlaces fosfodiéster adicionales por un dominio proteínico derivado de una exonucleasa.

En experimentos de manipulación de genoma, la eficacia de endonucleasa de corte infrecuente, por ejemplo, su capacidad de inducir un evento deseado (dirección a un gen homólogo, mutagénesis dirigida, eliminación o escisión de secuencia) en un locus, depende de varios parámetros, incluyendo la actividad específica de la nucleasa,

probablemente la accesibilidad de la diana y la eficacia y resultado de la ruta o rutas de reparación que provocan el evento deseado (reparación homóloga para dirección génica, rutas de NHEJ para mutagénesis dirigida).

5 La escisión por endonucleasas de corte infrecuente peptídico habitualmente genera extremos cohesivos, con salientes 3' para meganucleasas LAGLIDADG (Chevalier y Stoddard 2001) y salientes 5' para nucleasas de dedos de cinc (Smith, Bibikova *et al.* 2000). Estos extremos, que son el resultado de la hidrólisis de enlaces fosfodiéster, pueden volver a ligarse *in vivo* por NHEJ de una manera sin empalmes (es decir, un religamiento sin marcas). La restauración de una secuencia diana escindible permite un nuevo evento de escisión por la misma endonucleasa y, por tanto, puede tener lugar una serie de ciclos infructuosos de eventos de escisión y religamiento. Las evidencias indirectas han demostrado que incluso en la levadura *Saccharomyces cerevisiae*, dichos ciclos podrían tener lugar tras la escisión continua por la endonucleasa HO (Lee, Paques *et al.* 1999). En células de mamífero, varios experimentos han demostrado que un perfecto religamiento de extremos cohesivos compatibles resultantes de dos DSB inducidos por I-SceI independientes, pero cercanos es un proceso eficaz (Guirouilh-Barbat, Huck *et al.* 2004; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Cheng *et al.* 2008; Bennardo, Gunn *et al.* 2009). La ausencia de la proteína de reparación de ADN Ku no afecta significativamente a la frecuencia global de eventos de NHEJ que vuelve a unir los extremos de las dos DSB; sin embargo, potencia muy fuertemente la contribución de NHEJ imprecisa al proceso de reparación en células inmortalizadas CDO y células ES de ratón (Guirouilh-Barbat, Huck *et al.* 2004; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Cheng *et al.* 2008). Además, la ausencia de Ku estimula los eventos inducidos por I-SceI tales como NHEJ imprecisa (Bennardo, Cheng *et al.* 2008), hibridación monocatenaria (Bennardo, Cheng *et al.* 2008) y conversión génica (Pierce, Hu *et al.* 2001; Bennardo, Cheng *et al.* 2008) en células ES de ratón. Se han hecho observaciones similares con células deficientes para la proteína de reparación XRCC4 (Pierce, Hu *et al.* 2001; Guirouilh-Barbat, Rass *et al.* 2007; Bennardo, Gunn *et al.* 2009) (aunque la deficiencia de XRCC4 afecta al nivel global de NHEJ en células CHO (Guirouilh-Barbat, Rass *et al.* 2007)) o para DNA-PK (Pierce, Hu *et al.* 2001). Por el contrario, la atenuación de CtIP ha demostrado suprimir "alt-NHEJ" (una forma independiente de Ku y XRCC4 de NHEJ más propensa a producir NHEJ imprecisa), hibridación monocatenaria y conversión génica, mientras que no afecta al nivel global de reunión de dos extremos compatibles generados por I-SceI (Bennardo, Cheng *et al.* 2008). Por tanto, la competición entre las diferentes rutas de reparación DSB puede afectar al espectro o los eventos de reparación resultantes de una DSB inducida por nucleasa.

30 Además, la resección de DSB es importante para determinadas rutas de DSB. La resección de DSB extensiva, que produce la generación de regiones monocatenarias grandes (unos pocos cientos de nucleótidos al menos), se ha demostrado en levaduras que inicia la hibridación monocatenaria (Sugawara y Haber 1992) y la invasión de hebra, la etapa dependiente de ATP que inicia muchos eventos de recombinación homólogos de invasión de dúplex de ADN por una hebra homóloga a esa (White y Haber 1990; Sun, Treco *et al.* 1991) (para una revisión de los mecanismos, véase (Paques y Haber 1999)). En células eucariotas, la resección de DSB depende de varias proteínas incluyendo BLM/Sgs1 y DNA2, EXO1 y el complejo MRN (Mre11, Rad50, Nbs1/Xrs2) y se cree que es el resultado de diferentes rutas. MRN está implicado en un proceso de resección a pequeña escala, mientras que dos rutas redundantes dependientes de BLM y DNA2 por un lado, y de EXO1 por otro lado, estarían implicadas en la resección extensiva (Mimitou y Symington 2008; Nimonkar, Genschel *et al.* 2011). Además, el procesamiento de los extremos que implica un nucleótido dañado (resultante de la escisión química o de un aducto voluminoso), requiere la proteína CtIP/Sae2 junto con RMN (Sartori, Lukas *et al.* 2007; Buis, Wu *et al.* 2008; Hartsuiker, Mizuno *et al.* 2009). La sobreexpresión de la exonucleasa Trex2 demostró estimular fuertemente NHEJ imperfecta asociada con la pérdida de únicamente unas pocas pares de bases (Bennardo, Gunn *et al.* 2009), aunque inhibía diversos tipos de eventos de reparación de ADN entre secuencias distantes (tal como hibridación monocatenaria, NHEJ entre extremos de diferentes roturas o reparación por NHEJ de una DSB individual que implica microhomologías remotas). En el mismo estudio, se sugirió que Trex2 reseca los salientes 3' dejados por I-SceI de una manera no procesiva. Por tanto, el tipo de ruta estimulada podría depender, a su vez, del tipo de resección (longitud de resección, monocatenario frente a bicatenario, resección de hebra 5' frente a hebra 3').

50 Por tanto, la eficacia de una TALEN compacta, por ejemplo, su capacidad de producir un evento deseado tal como mutagénesis dirigida o dirección a gen homólogo (véase la definición para la definición completa de "eficacia de TALEN compacta"), puede potenciarse mediante una potenciación o modificación de su eficacia de procesamiento de ADN global (véase la definición para la definición completa de "eficacia de procesamiento de ADN global"), por ejemplo, el resultado global o el resultado general de diferentes actividades enzimáticas separadas de dicha TALEN compacta.

55 Como también se divulga en este documento, la potenciación de la eficacia de procesamiento de ADN global de una entidad de TALEN compacta de acuerdo con la presente invención, en comparación con una entidad de TALEN compacta de partida, puede ser la hidrólisis de enlaces fosfodiéster adicionales en el sitio de escisión. Dicha hidrólisis de enlaces fosfodiéster adicionales en el sitio de escisión por dicho al menos un potenciador de acuerdo con la invención puede dar lugar a diferentes tipos de resección de DSB que afecta a dicho sitio de escisión de DSB, una única hebra de ADN o ambas hebra de ADN, que afecta a cualquier extremo saliente 5', cualquier extremo saliente 3' o ambos extremos y dependiendo de la longitud de dicha resección. Por tanto, añadir nuevas actividades nickasa o cleavasa a la actividad cleavasa existente de una entidad de TALEN compacta puede potenciar la eficacia de la TALEN compacta potenciada resultante, en un locus genómico de interés (figura 8B-8E). Como un ejemplo no limitante, la adición de dos actividades nickasa en hebras opuestas (figura 8D) o de una nueva actividad cleavasa que genera una segunda DSB (figura 8E) puede provocar una interrupción bicatenaria. Como consecuencia, el

religamiento perfecto ya no es posible, y puede estimularse uno o varios resultados de reparación alternativos tales como NHEJ imprecisa, recombinación homóloga o SSA por ejemplo. Como otro ejemplo no limitante, la adición de una única actividad nickasa puede provocar una interrupción monocatenaria, y suprimir la cohesividad de los extremos, que también puede potenciar la eficacia de la TALEN compacta potenciada resultante en un locus genómico de interés, mediante estimulación de uno o varios resultados de reparación alternativos mencionados anteriormente.

En este aspecto descrito de la presente invención, la potenciación de la eficacia de procesamiento de ADN de una TALEN compacta se refiere al aumento en el nivel detectado de dicha eficacia de procesamiento de ADN, frente a una secuencia de ADN diana, de una TALEN compacta en comparación con la actividad de una primera TALEN compacta frente a la misma secuencia de ADN diana. Dicha primera TALEN compacta puede ser una TALEN compacta de partida o una TALEN compacta que ya se ha genomanipulado o una TALEN compacta potenciada. Puede preverse varias rondas de potenciación a partir de una TALEN compacta de partida o a partir de una TALEN compacta potenciada de partida.

En este aspecto descrito de la presente invención, la potenciación de la eficacia de procesamiento de ADN de la entidad de TALEN compacta (o TALEN compacta potenciada) se refiere al aumento en el nivel detectado de dicha eficacia de procesamiento de ADN frente a una secuencia de ADN diana de interés o cerca de dicha secuencia de ADN de interés en comparación con la eficacia de una primera TALEN compacta o TALEN compacta de partida frente a o cerca de la misma secuencia de ADN diana. En este caso, la TALEN compacta de partida se toma como el almacén de referencia para medir la eficacia de procesamiento de ADN. Dicha TALEN compacta potenciada es una TALEN compacta genomanipulada que comprende un dominio potenciador. Dicha TALEN compacta potenciada también puede tomarse como almacén de referencia para potenciación adicional en dicha eficacia de procesamiento de ADN. Como un ejemplo no limitante, dicha eficacia de procesamiento de ADN puede producirse a partir de una recombinación inducida por escisión generada por dicha TALEN compacta potenciada. En este caso, dicho nivel de recombinación inducida por escisión puede determinarse, por ejemplo, por un ensayo de recombinación basado en células como se describe en la solicitud PCT internacional WO 2004/067736. De forma importante, la potenciación de la eficacia en células (generación potenciada de mutagénesis dirigida o recombinación dirigida) puede estar asociada, aunque no necesariamente, con una potenciación de la actividad de escisión que podría detectarse en determinados ensayos *in vitro*. Por ejemplo, actividades fosfodiesterasa adicionales como se describe en la figura 8 podrían afectar escasamente al perfil de escisión, que se detecta por escisión *in vitro* y separación de los productos de escisión en un gel de electroforesis. Sin embargo, como se explica anteriormente, y en leyenda de la figura 8, los extremos de DSB generados de esta manera podrían ser más propensos para inducir reordenamientos genómicos detectables tales como mutagénesis dirigida (por NHEJ imprecisa) o recombinación homóloga. Dicha potenciación en la recombinación inducida por escisión de dicha TALEN compacta potenciada es una potenciación de al menos un 5 % en comparación con el almacén de partida o TALEN compacta de partida, más preferiblemente una potenciación de al menos un 10 %, de nuevo más preferiblemente una potenciación de al menos un 15 %, de nuevo más preferiblemente una potenciación de al menos un 20 %, de nuevo más preferiblemente una potenciación de al menos un 25 %, de nuevo más preferiblemente una potenciación de un 50 %, de nuevo más preferiblemente una potenciación mayor de un 50 %, que provoca una potenciación de la eficacia de procesamiento de ADN de dicha TALEN compacta potenciada de al menos un 5 % en comparación con el almacén de partida o TALEN compacta de partida, más preferiblemente una potenciación de al menos un 10 %, de nuevo más preferiblemente una potenciación de al menos un 15 %, de nuevo más preferiblemente una potenciación de al menos un 20 %, de nuevo más preferiblemente una potenciación de al menos un 25 %, de nuevo más preferiblemente una potenciación de un 50 %, de nuevo más preferiblemente una potenciación mayor de un 50 %.

En otra realización preferida de acuerdo con el método de la presente invención, el conector peptídico que puede unir dicho dominio potenciador a una parte de dicha entidad de TALEN compacta de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1, NFS2, CFS1, RM2, BQY, QGPSG, LGPDGRKA, 1a8h_1, 1dnpA_1, 1d8cA_2, 1ckqA_3, 1sbp_1, 1ev7A_1, 1alo_3, 1amf_1, 1adjA_3, 1fcdC_1, 1al3_2, 1g3p_1, 1acc_3, 1ahjB_1, 1acc_1, 1af7_1, 1heiA_1, 1bia_2, 1igtB_1, 1nfkA_1, 1au7A_1, 1bpoB_1, 1b0pA_2, 1c05A_2, 1gcb_1, 1bt3A_1, 1b3oB_2, 16vpA_6, 1dhx_1, 1b8aA_1 y 1qu6A_1 como se enumera en la tabla 3 (de la SEQ ID NO: 67 a la SEQ ID NO: 104 y de la SEQ ID NO: 372 a la SEQ ID NO: 415). Como también se divulga en este documento, el conector peptídico que puede unir dicho dominio potenciador a una parte de dicha entidad de TALEN compacta de acuerdo con el método de la presente invención puede seleccionarse del grupo que consiste en NFS1 (SEQ ID NO: 98), NFS2 (SEQ ID NO: 99) y CFS1 (SEQ ID NO: 100). También se describe el caso donde no se necesita un conector peptídico para fusionar un dominio potenciador a una parte de dicha entidad de TALEN compacta para obtener una TALEN compacta potenciada.

Dependiendo de su composición estructural [tipo de almacén de TALE central, tipo de domino o dominios catalíticos con actividades enzimáticas asociadas, tipo de potenciadores y finalmente tipo de conector o conectores], una TALEN compacta o una TALEN compacta potenciada puede comprender diferentes niveles de actividades enzimáticas diferentes que pueden procesar de forma diferente el ADN como se menciona anteriormente. Añadiendo nuevas actividades enzimáticas a dicha TALEN compacta o TALEN compacta potenciada o potenciando la eficacia de procesamiento de ADN de una o varias de sus actividades enzimáticas constitutivas, se puede potenciar la eficacia de procesamiento de ADN global de una TALEN compacta o TALEN compacta potenciada en comparación con una TALEN compacta de partida o TALEN compacta potenciada.

De acuerdo con la presente invención, se diseñan TALEN compactas para aliviar la necesidad de múltiples restos proteínicos independientes cuando se aborda un evento de procesamiento de ADN. De forma importante, la región "espaciadora" necesaria y los sitios diana dobles esenciales para la función de las TALEN actuales con innecesarios, ya que las TALEN compactas de acuerdo con la invención comprenden un almacén de TALE central que contiene únicamente un dominio de unión a ADN para abordar una secuencia diana de ADN bicatenario única específica de interés y procesar el ADN cerca de dicha secuencia diana de ADN bicatenario única de interés. Como cada extremo del almacén de TALE central es susceptible a fusión, el orden (en el extremo N frente al C) de adición de los dominios catalíticos y de potenciación puede variar con la aplicación. Además, como el dominio catalítico no requiere contactos de ADN específicos, no hay restricciones sobre las regiones que rodean el almacén de TALE central, como ejemplos no limitantes representados en la figura 5: (A) Construcción de fusión en el extremo N para promover la recombinación homóloga inducida por un dominio cleavasa o por un dominio nickasa. (B) Construcción de fusión en el extremo C con propiedades como en (A). (C) La adhesión de dos dominios catalíticos a ambos extremos del almacén de TALE central permite la escisión doble con potenciación en NHEJ. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación.

Las TALEN compactas pueden potenciarse mediante la adición de un dominio para promover actividades existentes o alternativas como ejemplos no limitantes representados en la figura 6: (A) Una TALEN compacta convencional con un dominio potenciador fusionado al extremo C de su parte de almacén de TALE central. (B) El dominio potenciador está fusionado a la TALEN compacta mediante el extremo N de su parte de dominio catalítico. Dicha configuración puede usarse para ayudar y/o anclar la parte de dominio catalítico cerca del ADN para aumentar la actividad de procesamiento de ADN. (C) El dominio potenciador se empareda entre la parte de dominio catalítico y la parte de almacén de TALE central. El dominio potenciador puede promover la comunicación entre los dominios flanqueantes (es decir, para ayudar en la catálisis y/o unión a ADN) o puede usarse para superar el nucleótido T necesario en la posición -1 de todas las dianas basadas en TALE. (D) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo N del almacén de TALE central genomanipulado. (E) El dominio potenciador se usa para reemplazar funcionalmente la región del extremo C del almacén de TALE central genomanipulado. Las uniones de fusión (en el extremo N frente al C) y los diseños de conector pueden variar con la aplicación.

La naturaleza del dominio o dominios catalíticos comprendidos en la TALEN compacta y la TALEN compacta potenciada depende de la aplicación. Como un ejemplo no limitante, un dominio nickasa permitiría una relación de HR/NHEJ mayor que un dominio cleavasa, siendo de este modo más conveniente para aplicaciones terapéuticas (McConnell Smith, Takeuchi *et al.* 2009; Metzger, McConnell-Smith *et al.* 2011). Por ejemplo, el acoplamiento de un dominio cleavasa en un lado con un dominio nickasa en el otro podría provocar la escisión de una hebra simple de ADN que abarca la región de unión de una TALEN compacta. La generación dirigida de salientes monocatenarios prolongados podría aplicarse en aplicaciones que abordan mecanismos de reparación de ADN. Para inactivación génica dirigida, entonces se prefiere el uso de dos dominios cleavasa. Como también se divulga en este documento, puede favorecerse el uso de dos dominios nickasa. Además, se describe un método para generar varios tipos diferentes de TALEN compactas que pueden aplicarse a aplicaciones que varían desde escisión de ADN dirigida a regulación génica dirigida.

La presente invención también describe métodos para el uso de dichas TALEN compactas de acuerdo con la invención para diversas aplicaciones que varían desde la escisión de ADN diana hasta la regulación génica diana. También se describe en este documento un método para aumentar HR dirigida (y NHEJ) cuando se promueve la actividad de rotura bicatenaria en una TALEN compacta dirigida a una secuencia diana de ADN de acuerdo con la invención. Como también se divulga en este documento, la adición de al menos dos dominios potenciadores cleavasa catalíticamente activos puede permitir el aumento de la mutagénesis inducida por rotura bicatenaria dando lugar a una pérdida de información genética y evitando cualquier religamiento sin marcas de locus genómico dirigida de interés por NHEJ.

También se describe en este documento un método para aumentar HR dirigida con menos NHEJ (es decir, de un modo más conservativo) cuando se promueve actividad de rotura monocatenaria en una TALEN compacta dirigida a una secuencia diana de ADN de acuerdo con la invención.

También se describe en este documento un método para aumentar la escisión de una hebra individual de ADN que abarca la región de unión de una TALEN compacta cuando tanto un dominio potenciador cleavasa como un dominio potenciador nickasa, respectivamente, se fusionan tanto al extremo N como al extremo C de un almacén de TALE central.

También se describe en este documento un método para el tratamiento de una enfermedad genética causada por una mutación en una secuencia diana de ADN bicatenario única específica en un gen, que comprende administrar a un sujeto que lo necesita una cantidad eficaz de una variante de una TALEN compacta de acuerdo con la presente invención.

En otra realización preferida, la presente invención se refiere a un método para insertar un transgén en una secuencia diana de ADN bicatenario única específica de un locus genómico de una célula, tejido o animal no humano, o una planta en la que al menos una TALEN compacta de la presente invención es transitoria o no se introduce en dicha

célula, tejido, animal no humano o planta.

La presente invención también describe un método para modular la actividad de una TALEN compacta cuando se expresa en una célula, en el que dicho método comprende la etapa de introducir en dicha célula un dominio auxiliar que modula la actividad de dicha TALEN compacta. La presente invención también describe un método que permite tener un control temporal de la actividad de una TALEN compacta cuando se expresa en una célula introduciendo en dicha célula un dominio auxiliar que modula la actividad de dicha TALEN compacta una vez que dicha TALEN compacta ha conseguido su actividad (escisión de ADN, mellado de ADN u otras actividades de procesamiento de ADN). La presente invención también describe un método para inhibir la actividad de una TALEN compacta cuando se expresa en una célula, en el que dicho método comprende la etapa de introducir en dicha célula un dominio auxiliar que inhibe la actividad de dicha TALEN compacta. Como también se divulga en este documento, el dominio catalítico de la TALEN compacta puede ser NucA (SEQ ID NO: 26) y dicho dominio auxiliar es NuiA (SEQ ID NO: 229), un mutante funcional, una variante o un derivado del mismo. Como también se divulga en este documento, el dominio catalítico de dicha TALEN compacta es ColE7 (SEQ ID NO: 11) y dicho dominio auxiliar puede ser Im7 (SEQ ID NO: 230), un mutante funcional, una variante o un derivado del mismo.

También se abarca en el alcance de la presente invención un polinucleótido recombinante que codifica una TALEN compacta, de acuerdo con la presente invención. También se abarca en el alcance de la presente invención, un vector que comprende un polinucleótido recombinante que codifica una TALEN compacta de acuerdo con la presente invención.

También se abarca en el alcance de la presente invención, un célula hospedadora que comprende un polinucleótido recombinante que codifica una TALEN compacta de acuerdo con la presente invención.

También se abarca en el alcance de la presente invención, un animal transgénico no humano que comprende un polinucleótido recombinante que codifica una TALEN compacta de acuerdo con la presente invención.

También se abarca en el alcance de la presente invención, una planta transgénica que comprende un polinucleótido recombinante que codifica una TALEN compacta de acuerdo con la presente invención.

La presente invención también se refiere a kits usados para implementar el método de acuerdo con la presente invención. Más preferiblemente, se abarca en el alcance de la presente invención, un kit que comprende una TALEN compacta de acuerdo con la presente invención e instrucciones para usar dicho kit en la potenciación de la eficacia de procesamiento de ADN de una secuencia diana de ADN bicatenario única de interés.

Para fines de tratamiento, las TALEN compactas de la presente invención y un excipiente farmacéuticamente aceptable se administran en una cantidad terapéuticamente eficaz. Se dice que tal combinación se administra en una "cantidad terapéuticamente eficaz" si la cantidad administrada es fisiológicamente significativa. Un agente es fisiológicamente significativo si su presencia provoca un cambio detectable en la fisiología del destinatario. En el presente contexto, un agente es fisiológicamente significativo si su presencia provoca una disminución en la gravedad de uno o más síntomas de la enfermedad diana y una corrección del genoma de la lesión o anomalía. Los vectores que comprenden ADN y/o ácido nucleico de dirección que codifica una TALEN compacta pueden introducirse en una célula mediante una diversidad de métodos (por ejemplo, inyección, captación directa, bombardeo con proyectiles, liposomas, electroporación). Las TALEN compactas puede expresarse de forma estable o transitoria en células usando vectores de expresión. Las técnicas de expresión en células eucariotas son bien conocidas para los expertos en la técnica. (Véase *Current Protocols in Human Genetics*: Capítulo 12 "Vectors For Gene Therapy" y Capítulo 13 "Delivery Systems for Gene Therapy").

En un aspecto descrito adicional de la presente invención, la TALEN compacta de la presente invención es sustancialmente no inmunógena, es decir, genera poca o ninguna respuesta inmunológica adversa. Puede usarse una diversidad de métodos para mejorar o eliminar las reacciones inmunológicas perjudiciales de este tipo de acuerdo con la invención. Como también se divulga en este documento, la TALEN compacta puede estar sustancialmente libre de N-formilmetionina. Otra manera para evitar las reacciones inmunológicas indeseadas es conjugar la TALEN compacta con polietilenglicol ("PEG") o polipropilenglicol ("PPG") (preferiblemente de 500 a 20 000 Dalton de peso molecular promedio (PM)). La conjugación con PEG o PPG, como se describe por Davis *et al.* (documento US 4179337) por ejemplo, puede proporcionar conjugados de TALEN compactas solubles en agua, fisiológicamente activos, no inmunógenos con actividad antiviral. Se describen métodos similares que también usan un copolímero de polietileno-polipropilenglicol en Saifer *et al.* (documento US 5006333).

En otro aspecto de la presente invención, es una composición que comprende una TALEN compacta de acuerdo con la presente invención y un vehículo. Más preferiblemente, es una composición farmacéutica que comprende una TALEN compacta de acuerdo con la presente invención y un vehículo farmacéuticamente activo conocido en el estado de la técnica.

En el alcance de la presente invención y para todas las aplicaciones mencionadas anteriormente, puede preverse el uso de más de una TALEN compacta (es decir, una entidad activa de TALEN compacta) (es decir, una entidad activa de TALEN compacta potenciada) para procesar ADN de acuerdo con la invención. Como también se divulga en este documento, pueden usarse dos TALEN compactas diferentes. Como también se divulga en este documento, como ejemplos no limitantes, dichas dos TALEN compactas diferentes pueden comprender el mismo armazón de TALE central o no; dichas dos TALEN compactas diferentes pueden comprender el mismo conjunto de dipéptidos variables de repetición o no; dichas dos TALEN compactas diferentes pueden comprender el mismo dominio catalítico o no. Cuando se usan dos entidades activas de TALEN compacta idénticas para el procesamiento de ADN de acuerdo con la invención, pueden considerarse un par homodimérico de entidades activas de TALEN compacta. Cuando se usan dos entidades activas de TALEN compacta no idénticas para el procesamiento de ADN de acuerdo con la invención, pueden considerarse un par heterodimérico de entidades activas de TALEN compacta. Como ejemplo no limitante, cuando se usan dos TALEN compactas de acuerdo la presente invención, una de las TALEN compactas pueden modular la actividad de la otra, dando lugar, por ejemplo, a un evento de procesamiento de ADN potenciado en comparación con el mismo evento de procesamiento de ADN conseguido por únicamente una TALEN compacta; en este ejemplo no limitante, una Trans-TALEN modula y potencia la actividad catalítica de una TALEN compacta inicial. Como también se divulga en este documento, pueden usarse tres TALEN compacta o tres TALEN compactas potenciadas. Como también se divulga en este documento, puede usarse más de tres TALEN compactas o tres TALEN compactas potenciadas para el procesamiento de ADN de acuerdo con la invención. Como también se divulga en este documento, puede usarse una combinación de TALEN compactas y TALEN compactas potenciadas para el procesamiento de ADN de acuerdo con la invención. Como un ejemplo no limitante, puede usarse una TALEN compacta y una TALEN compacta potenciada. Como otro ejemplo no limitante, puede usarse una TALEN compacta y una TALEN compacta de escisión doble. Como también se divulga en este documento, puede usarse una combinación de TALEN compactas, TALEN compactas potenciadas y TALEN compactas de escisión doble, comprendiendo dichas TALEN compactas el mismo dominio catalítico o no, el mismo armazón de TALE central o no. Cuando tienen que usarse varias TALEN compactas, la secuencia diana de ADN para cada TALEN compacta de la combinación a usar puede estar ubicada en el mismo locus de ADN genómico endógeno de interés o no. Dichas secuencias diana de ADN puede estar ubicadas a una distancia aproximada de 1000 pares de bases (pb). Más preferiblemente, dichas secuencias diana de ADN pueden estar ubicadas a una distancia aproximada de 500 pb o 200 pb, o 100 pb o 50 pb o 20 pb, 19 pb, 18 pb, 17 pb, 16 pb, 15 pb, 14 pb, 13 pb, 12 pb, 11 pb, 10 pb, 9 pb, 8 pb, 7 pb, 6 pb, 5 pb, 4 pb, 3 pb, 2 pb, 1 pb. Dichas secuencias diana de ADN ubicadas a distancias mencionadas anteriormente son secuencias de ADN "cercanas" en referencia a la secuencia de ADN diana para el procesamiento de ADN de acuerdo con la presente invención.

Como también se divulga en este documento, dos entidades activas de TALEN compacta pueden usarse como una manera de conseguir dos actividades de procesamiento de ADN diferentes cerca de una secuencia diana de ADN de acuerdo con la invención. Como un ejemplo no limitante, pueden usarse dos TALEN compactas dirigidas a dicha secuencia de ADN o secuencias de ADN cerca de dicha secuencia de ADN diana y que comprende cada una un dominio catalítico derivado de nickasa; en este caso, este uso de dos entidades activas de TALEN compacta puede representar una manera alternativa de conseguir una rotura bicatenaria cerca de dicha secuencia diana de ADN, en comparación con el uso de una TALEN compacta dirigida a dicha secuencia de ADN que comprende un dominio catalítico derivado de cleavasa, o no. Como otro ejemplo no limitante, puede usarse una TALEN compacta que comprende un dominio derivado de cleavasa y una TALEN compacta que comprende un dominio derivado de exonucleasa para generar una rotura bicatenaria y crear una interrupción, respectivamente, para conseguir un evento de NHEJ impreciso en el locus genómico de interés que comprende dicha secuencia diana de ADN. En este caso, incluso si cada TALEN compacta que forma este par heterodimérico de TALEN compactas es activo por sí mismo, cada una de estas entidades activas depende de la otra para conseguir la actividad de procesamiento de ADN resultante deseada. De hecho, en este caso particular, la actividad resultante deseada es una interrupción creada por la actividad exonucleasa, siendo posible dicha actividad exonucleasa únicamente a partir de la rotura bicatenaria conseguida por el dominio cleavasa de las otra TALEN compactas. En el alcance de la presente invención, también se prevé el caso donde dos entidades activas de TALEN compacta idénticas (un par homodimérico de TALEN compactas) son dependientes entre sí para conseguir una actividad de procesamiento de ADN resultante deseada.

Definiciones

- 55 - Los restos de aminoácido en una secuencia polipeptídica se denominan en este documento de acuerdo con el código de una letra, en que, por ejemplo, Q significa Gln o resto de glutamina, R significa Arg o resto de arginina y D significa Asp o resto de ácido aspártico.
- Sustitución de aminoácido significa el remplazo de un resto de aminoácido con otro, por ejemplo, el remplazo de un resto de arginina con un resto de glutamina en una secuencia peptídica es una sustitución de aminoácido.
- 60 - Actividad de procesamiento de ADN potenciada/aumentada/mejorada, se refiere a un aumento en el nivel detectado de una TALEN compacta o TALEN compacta potenciada dada asociada con actividad de procesamiento de ADN frente a una secuencia de ADN diana o secuencia diana de ADN por una segunda TALEN compacta o TALEN compacta potenciada en comparación con la actividad de una primera TALEN compacta o TALEN compacta potenciada frente a la secuencia de ADN diana. La segunda TALEN compacta o TALEN compacta potenciada puede ser una variante de la primera y puede comprender uno o más restos de aminoácidos sustituidos en comparación con la primera TALEN compacta o TALEN compacta potenciada. La segunda TALEN compacta o

TALEN compacta potenciada puede ser una variante de la primera y puede comprender uno o más dominios catalíticos y/o potenciadores en comparación con dicha primera TALEN compacta o TALEN compacta potenciada. Esta definición se aplica más ampliamente para otras endonucleasas y endonucleasas de corte infrecuente.

- 5 - Actividad de procesamiento de ADN se refiere a una actividad enzimática particular / dada de dicha TALEN compacta o TALEN compacta potenciada o más ampliamente para calificar la actividad enzimática de una endonucleasa de corte infrecuente. Dicha actividad de procesamiento de ADN puede referirse a una actividad de escisión, una actividad cleavasa, una actividad nickasa, más ampliamente una actividad nucleasa, pero también una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. En el alcance de esta definición, dicha actividad de procesamiento de ADN dada de una actividad enzimática particular también puede describir como eficacia de procesamiento de ADN de dicha actividad enzimática particular. Los métodos para potenciar la actividad de procesamiento de ADN de la TALEN compacta o TALEN compacta potenciada de acuerdo con esta definición son ilustrativos de la presente invención.
- 10 - La eficacia de procesamiento de ADN global describe, para una TALEN compacta o una TALEN compacta potenciada, el resultante global o el efecto general de diferentes actividades enzimáticas diferentes que puede comprender dicha TALEN compacta o TALEN compacta potenciada. De acuerdo con estas actividades enzimáticas separadas diferentes, una TALEN compacta o una TALEN compacta potenciada presenta una capacidad global de procesar ADN cerca de una secuencia diana en un locus genómico de interés, es decir, una eficacia de procesamiento de ADN global. Dicha eficacia de procesamiento de ADN global puede calificar o clasificar una segunda TALEN compacta o TALEN compacta potenciada dada en comparación con una primera TALEN compacta o TALEN compacta potenciada dada. Dependiendo de dicha composición de TALEN compactas o TALEN compactas potenciadas [tipo de armazón de TALE central, tipo de domino o dominios catalíticos con actividades enzimáticas asociadas, finalmente tipo de conector o conectores y tipo de dominio o dominios potenciadores], dicha eficacia de procesamiento de ADN global puede referirse a únicamente una actividad enzimática, dos actividades enzimáticas, tres actividades enzimáticas, cuatro actividades enzimáticas o más de cuatro actividades enzimáticas. Dicha eficacia de procesamiento de ADN global puede referirse a la suma de actividades enzimáticas individuales. Dicha eficacia de procesamiento de ADN global puede referirse a la sinergia o efecto combinado de diferentes actividades enzimáticas comprendidas en una TALEN compacta o TALEN compacta potenciada dada. Una potenciación de la eficacia de procesamiento de ADN de una TALEN compacta de acuerdo con la presente invención puede reflejar una sinergia, un efecto combinado potenciado, que produce una TALEN compacta potenciada caracterizada por una eficacia de procesamiento de ADN global que es mayor que la suma de las eficacias de procesamiento de ADN respectivas de TALEN compactas de partica diferentes o que la suma de eficacias de procesamiento de ADN respectivas de actividades enzimáticas diferentes comprendidas en una misma TALEN compacta de partida.
- 15 - La eficacia de una endonucleasa de corte infrecuente de acuerdo con la presente invención es la propiedad de dicha endonucleasa de corte infrecuente de producir un evento deseado. Este evento deseado puede ser, por ejemplo, dirección génica homóloga, mutagénesis dirigida o eliminación o escisión de secuencias. La eficacia del evento deseado depende de varios parámetros, incluyendo la actividad específica de la nucleasa y la ruta o rutas de reparación que provocan el efecto deseado (eficacia de reparación homóloga para dirección génica, eficacia y resultado de rutas de NHEJ para mutagénesis dirigida. La eficacia de una endonucleasa de corte infrecuente para un locus pretende indicar su capacidad de producir un evento deseado en este locus. La eficacia de una endonucleasa de corte infrecuente para una diana pretende indicar su capacidad de producir un evento deseado como consecuencia de la escisión de esta diana.
- 20 - Los nucleótidos se denominan de la siguiente manera: se usa el código de una letra para denominar la base de un nucleósido: a es adenina, t es timina, c es citosina y g es guanina. Para los nucleótidos degenerados, r representa g o a (nucleótidos de purina), k representa g o t, s representa g o c, w representa a o t, m representa a o c, y representa t o c (nucleótidos de pirimidina), d representa g, a o t, v representa g, a o c, b representa g, t o c, h representa a, t o c, y n representa g, a, t o c.
- 25 - Por "meganucleasa", se pretende un subtipo de endonucleasa de corte infrecuente que tiene una secuencia diana de ADN bicatenario mayor de 12 pb. Dicha meganucleasa es una enzima dimérica, en la que cada dominio está en un monómero o una enzima monomérica que comprende los dos dominios en un único polipéptido.
- 30 - Por "dominio de meganucleasa" se pretende la región que interactúa con una mitad de la diana de ADN de una meganucleasa y puede asociarse con el otro dominio de la misma meganucleasa que interactúa con la otra mitad de la diana de ADN para formar una meganucleasa funcional que puede escindir dicha diana de ADN.
- 35 - Por "variante de endonucleasa", "variante de endonucleasa de corte infrecuente", "variante de endonucleasa de corte infrecuente quimérica" o "variante de meganucleasa" o "variante de TALEN compacta", o "variante de TALEN compacta potenciada" o "variante de TALEN compacta de escisión doble" o "variante" se pretende una endonucleasa, endonucleasa de corte infrecuente, endonucleasa de corte infrecuente quimérica, meganucleasa o TALEN compacta, TALEN compacta potenciada, TALEN compacta de escisión doble obtenida por remplazo de al menos un resto en la secuencia de aminoácidos de la endonucleasa precursora, endonucleasa de corte infrecuente, endonucleasa de corte infrecuente quimérica, meganucleasa o TALEN compacta, TALEN compacta potenciada, TALEN compacta de escisión doble con al menos un aminoácido diferente. La denominación "variante" también se aplica, por ejemplo, a una TALEN compacta potenciada que comprende al menos un dominio proteínico complementario (dominio catalítico o potenciador) en comparación con la entidad de TALEN compacta de partida.
- 40 - También se abarcan en el alcance de la presente definición, variantes y dominios proteínicos comprendidos en estas variantes que presentan una secuencia con un alto porcentaje de identidad o un alto porcentaje de homología

con secuencias de TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos y polipéptidos, a nivel nucleotídico o polipeptídico. Por alto porcentaje de identidad o alto porcentaje de homología se pretende un 60 %, más preferiblemente un 70 %, más preferiblemente un 75 %, más preferiblemente un 80 %, más preferiblemente un 85 %, más preferiblemente un 90 %, más preferiblemente un 95, más preferiblemente un 97 %, más preferiblemente un 99 % o cualquier número entero comprendido entre un 60 % y un 99 %.

- Por "conector de péptido", "conector peptídico" o "espaciador peptídico" se pretende indicar una secuencia peptídica que permite la conexión de diferentes monómeros en una proteína de fusión y la adopción de la conformación correcta para la actividad de dicha proteína de fusión y que no altera la especificidad de ninguno de los monómeros por sus dianas. Los conectores peptídicos pueden ser de diversos tamaños, de 3 aminoácidos a 50 aminoácidos como intervalo indicativo no limitante. Los conectores peptídicos también pueden estar estructurados o no estructurados.
- Por "relacionado con", particularmente en la expresión "un tipo celular relacionado con el tipo celular u organismo elegido", se pretende un tipo celular o un organismo que comparte características con dicho tipo celular elegido o dicho organismo elegido; este tipo celular u organismo relacionado con el tipo celular u organismo elegido, puede obtenerse de dicho tipo celular u organismo elegido o no.
- Por "subdominio" se pretende la región de un dominio central de endonucleasa de asentamiento LAGLIDADG que interactúa con una parte distinta de un semisitio diana de ADN de endonucleasa de asentamiento.
- Por "construcción de ADN de dirección/matriz de reparación mínima/matriz de reparación" se pretende indicar una construcción de ADN que comprende una primera y una segunda parte que son homólogas a las regiones 5' y 3' de la diana de ADN *in situ*. La construcción de ADN también comprende una tercera parte colocada entre la primera y la segunda parte que comprende algo de homología con la secuencia de ADN correspondiente *in situ* o como alternativa no comprende homología con la regiones 5' y 3' de la diana de ADN *in situ*. Después de la escisión de la diana de ADN, se estimula un evento de recombinación homóloga entre el genoma que contiene el gen diana comprendido en el locus de interés y la matriz de reparación, en la que la secuencia genómica que contiene la diana de ADN se reemplaza por la tercera parte de la matriz de reparación y una parte variable de la primera y la segunda parte de la matriz de reparación.
- Por "variante funcional" se pretende una variante catalíticamente activa de una proteína, pudiendo tener dicha variante propiedad adicionales en comparación con su proteína precursora. Como un ejemplo no limitante, una variante funcional de una meganucleasa puede tener la capacidad de escindir una secuencia diana de ADN, siendo preferiblemente dicha diana una nueva diana que no se escinde por la meganucleasa precursora. Esta definición también se aplica a TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos que constituyen dichas TALEN. También se abarcan en el alcance de la presente definición, variantes funcionales, polipéptidos y dominios proteínicos comprendidos en estas moléculas que presentan una secuencia con un alto porcentaje de identidad o un alto porcentaje de homología con secuencias de TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos y polipéptidos, a nivel nucleotídico o polipeptídico. Por alto porcentaje de identidad o alto porcentaje de homología se pretende un 60 %, más preferiblemente un 70 %, más preferiblemente un 75 %, más preferiblemente un 80 %, más preferiblemente un 85 %, más preferiblemente un 90 %, más preferiblemente un 95, más preferiblemente un 97 %, más preferiblemente un 99 % o cualquier número entero comprendido entre un 60 % y un 99 %.
- Por "derivado de" o "derivado o derivados" se pretende indicar, por ejemplo, una variante de meganucleasa que se crea a partir de una meganucleasa precursora y, por tanto, la secuencia peptídica de la variante meganucleasa está relacionada con (nivel de secuencia principal) pero deriva de (mutaciones) la secuencia peptídica de la meganucleasa precursora. En esta definición, las mutaciones abarcan eliminaciones o inserciones de varios restos de aminoácidos; como ejemplo no limitante, una variante truncada de una meganucleasa I-CreI se considera un armazón derivado de la meganucleasa I-CreI. Esta expresión también puede aplicarse a TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos que constituyen dichas TALEN. También se abarcan en el alcance de la presente definición, derivados de TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos y derivados de polipéptidos que presentan una secuencia con un alto porcentaje de identidad o un alto porcentaje de homología con secuencias de TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble o dominios proteínicos y polipéptidos, a nivel nucleotídico o polipeptídico. Por alto porcentaje de identidad o alto porcentaje de homología se pretende un 60 %, más preferiblemente un 70 %, más preferiblemente un 75 %, más preferiblemente un 80 %, más preferiblemente un 85 %, más preferiblemente un 90 %, más preferiblemente un 95, más preferiblemente un 97 %, más preferiblemente un 99 % o cualquier número entero comprendido entre un 60 % y un 99 %.
- Por "I-CreI" se pretende la I-CreI de tipo silvestre que tiene la secuencia del código de acceso de pdb 1g9y, correspondiente a la secuencia SEQ ID NO: 1 en el listado de secuencias. En la presente solicitud de patente, las variantes de I-CreI descritas pueden comprender una alanina adicional después de la primera metionina de la secuencia de I-CreI de tipo silvestre (SEQ ID NO: 1). Estas variantes también pueden comprender dos restos de alanina adicionales y un resto de ácido aspártico después de la prolina final de la secuencia de I-CreI de tipo silvestre como se muestra en la SEQ ID NO: 106. Estos restos adicionales no afectan a las propiedades de la

enzima y para evitar confusiones, estos restos adicionales no afectan a la numeración de los restos en *I-Crel* o una variante menciona en la presente solicitud de patente, ya que estas referencias se refieren exclusivamente a restos de la enzima *I-Crel* de tipo silvestre (SEQ ID NO: 1) presente en la variante, de modo que, por ejemplo, el resto 2 de *I-Crel* es de hecho el resto 3 de una variante que comprende una alanina adicional después de la primera metionina.

- 5
- Por TALEN compacta, TALEN compacta potenciada, TALEN compacta de escisión doble con especificidad novedosa se pretende una variante de estas proteínas que tiene un patrón de dianas escindidas diferentes de las de sus TALEN compactas, TALEN compactas potenciadas, TALEN compactas de escisión doble precursoras respectivas. Las expresiones "especificidad novedosa", "especificidad modificada", "especificidad de escisión
 - 10 novedosa", "especificidad de sustrato novedosa" que son equivalentes y se usan indistintamente, se refieren a la especificidad de la variante hacia los nucleótidos de la secuencia diana de ADN.
 - Por "sitio *I-Crel*" se pretende una secuencia de ADN bicatenario de 22 a 24 pb que se escinde por *I-Crel*. Los sitios *I-Crel* incluyen el sitio de asentamiento de *I-Crel* no palindrómico de tipo silvestre y las secuencias palindrómicas derivadas tales como la secuencia 5'-t₁₂C-₁₁a-₁₀a-₉a-₈a-₇C-₆g-₅t-₄C-₃g-₂t-₁a-₁C-₂g-₃a-₄C-₅g-₆t-₇t-₈t-₉t-₁₀g-₁₁a-₁₂ (SEQ ID NO: 2), también llamada C1221 o diana C1221.
 - 15 - Por "dominio" o "dominio central" se pretende el "dominio central de endonucleasa de asentamiento LAGLIDADG" que es el plegamiento $\alpha\beta\beta\alpha\beta\beta\alpha$ característico de las endonucleasas de asentamiento de la familia LAGLIDADG, correspondiente a una secuencia de aproximadamente cien restos de aminoácidos. Dicho dominio comprende cuatro hebras ($\beta_1\beta_2\beta_3\beta_4$) plegadas en una lámina beta antiparalela que interactúa con una mitad de la diana de
 - 20 ADN. Este dominio puede asociarse con otro dominio central de endonucleasa de asentamiento LAGLIDADG que interactúa con la otra mitad de la diana de ADN para formar una endonucleasa funcional con capacidad de escindir dicha diana de ADN. Por ejemplo, en el caso de la endonucleasa de asentamiento dimérica *I-Crel* (163 aminoácidos), el dominio central de endonucleasa de asentamiento LAGLIDADG corresponde a los restos 6 a 94.
 - Por "horquilla beta" se pretende dos hebras beta consecutivas de la lámina beta antiparalela de un dominio central de endonucleasa de asentamiento LAGLIDADG ($\beta_1\beta_2$ o $\beta_3\beta_4$) que están conectadas por un bucle o un giro.
 - 25 - Por "meganucleasa de una cadena", "meganucleasa quimérica de una cadena", "derivado de meganucleasa de una cadena", "derivado de meganucleasa quimérica de una cadena" o "derivado de una cadena" se pretende una meganucleasa que comprende dos dominios de endonucleasa de asentamiento LAGLIDADG o dominios centrales ligados por un espaciador peptídico. La meganucleasa de una cadena puede escindir una secuencia diana de ADN quimérica que comprende una mitad diferente de cada secuencia diana de meganucleasa precursora.
 - 30 - Por "diana de ADN", "secuencia diana de ADN", "secuencia de ADN diana", "secuencia diana", "sitio diana", "diana", "sitio", "sitio de interés", "sitio de reconocimiento", "sitio de reconocimiento polinucleotídico", "secuencia de reconocimiento", "sitio de reconocimiento de asentamiento", "sitio de asentamiento", "sitio de escisión" se pretende una secuencia polinucleotídica bicatenaria palindrómica, parcialmente palindrómica (seudopalindrómica) o no palindrómica que se reconoce y puede escindirse por una endonucleasa de asentamiento LAGLIDADG tal como *I-Crel*, o una variante, o una meganucleasa quimérica de una cadena derivada de *I-Crel*. Dicha secuencia diana de ADN puede calificarse como "escindible" por una endonucleasa, endonucleasa de corte infrecuente,
 - 35 endonucleasa de corte infrecuente quimérica o meganucleasa cuando se reconoce dentro de una secuencia genómica y se sabe que corresponde a la secuencia diana de ADN de una endonucleasa, endonucleasa de corte infrecuente, endonucleasa de corte infrecuente quimérica o meganucleasa dada o una variante de dicha endonucleasa, endonucleasa de corte infrecuente, endonucleasa de corte infrecuente quimérica o meganucleasa. Estos términos se refieren a una ubicación de ADN específica, preferiblemente una ubicación genómica, pero también una parte de material genético que puede existir independientemente del cuerpo principal de material genético tal como plásmidos, episomas, virus, transposones o en orgánulos tales como mitocondrias o cloroplastos
 - 40 como ejemplos no limitantes, en que puede inducirse una rotura bicatenaria (escisión) por la endonucleasa, endonucleasa de corte infrecuente, endonucleasa de corte infrecuente quimérica o meganucleasa. Para la subfamilia de LAGLIDADG de endonucleasas de corte infrecuente, la diana de ADN se define por la secuencia 5' a 3' de una hebra del polinucleótido bicatenario, como se indica anteriormente para C1221 (SEQ ID NO: 2). La escisión de la diana de ADN puede producirse en los nucleótidos en las posiciones +2 y -2, respectivamente para la hebra codificante y la no codificante. Salvo que se indique de otro modo, la posición en que puede producirse la escisión de la diana de ADN por una variante derivada de *I-Crel*, corresponde al sitio de escisión en la hebra codificante de la diana de ADN. En el caso particular de TALEN compactas, una subclase de endonucleasas de corte infrecuente quiméricas, las siguientes expresiones "diana de ADN", "secuencia diana de ADN", "secuencia de ADN diana", "secuencia diana", "sitio diana", "diana", "sitio", "sitio de interés", "sitio de reconocimiento", "sitio de
 - 45 reconocimiento polinucleotídico" y "secuencia de reconocimiento" pueden aplicarse para calificar su secuencia diana de ADN específica con la particularidad de que dicha secuencia diana de ADN específica reconocida por la TALEN compacta de acuerdo con la invención es la que se procesa y/o corta o no por la TALEN compacta. Una TALEN compacta, una TALEN compacta potenciada o una TALEN compacta de escisión doble puede procesar y/o cortar ADN dentro de dicha secuencia diana de ADN específica. Una TALEN compacta, una TALEN compacta potenciada o una TALEN compacta de escisión doble también puede procesar y/o cortar ADN fuera de dicha secuencia diana de ADN específica.
 - 50 - Por "ADN cerca de dicha secuencia diana de ADN específica" o por "ADN cercano" se pretende la secuencia o secuencias de ADN ubicadas dentro o fuera de dicha secuencia diana de ADN específica. También se pretende una secuencia o secuencias de ADN unidas por una TALEN compacta o una TALEN compacta potenciada en dicha ubicación de secuencia diana de ADN específica o ADN ubicado a una distancia 5' o 3' de 1-100 pares de bases (pb), 1-50 pares de bases (pb) o 1-25 pares de bases (pb) desde dicha secuencia diana de ADN específica.
 - 55
 - 60
 - 65

Cuando tienen que usarse varias TALEN compactas en una aplicación de manipulación de genoma particular, la secuencia diana de ADN para cada TALEN compacta de la combinación a usar puede estar ubicada en el mismo locus de ADN genómico endógeno de interés o no. Dichas secuencias diana de ADN puede estar ubicadas a una distancia aproximada de 1-1000 pares de bases (pb), más preferiblemente de 1-500 pb, más preferiblemente de 1-100 pb, más preferiblemente de 1-100 pb, más preferiblemente de 1-50 pb, más preferiblemente de 1-25 pb, más preferiblemente de 1-10 pb. En otra realización, dicha secuencia diana de ADN para cada TALEN compacta de la combinación a usar puede estar ubicada en la misma hebra de ADN o no. Dichas secuencias diana de ADN ubicadas a distancias mencionadas anteriormente son secuencias de ADN "cercanas" en referencia a la secuencia de ADN diana para el procesamiento de ADN de acuerdo con la presente invención.

- Por "secuencia diana de ADN bicatenario única" se pretende un sitio de unión de TALEN compacta o TALEN compacta potenciada o TALEN compacta de escisión doble. El sitio de unión a ADN de reconocimiento de una TALEN compacta o TALEN compacta potenciada o TALEN compacta de escisión doble puede ser de 12 a 100 pares de bases (pb) de longitud, habitualmente de más de 12 pb de longitud.

- Por "semisitio diana de ADN", "semisitio de escisión o semisitio" se pretende la parte de la diana de ADN que se une por cada dominio central de endonucleasa de asentamiento LAGLIDADG.

- El término "endonucleasa" se refiere a cualquier enzima de tipo silvestre o variante que puede catalizar la hidrólisis (escisión) de enlaces entre ácidos nucleicos dentro de una molécula de ADN o ARN, preferiblemente una molécula de ADN. Las endonucleasas pueden clasificarse como endonucleasas de corte infrecuente que tienen típicamente un reconocimiento polinucleotídico de más de 12 pares de bases (pb) de longitud, más preferiblemente 14-45 pb. Las endonucleasas de corte infrecuente aumentan significativamente la HR induciendo roturas bicatenarias (DSB) de ADN en un locus definido (Rouet, Smih *et al.* 1994; Rouet, Smih *et al.* 1994; Choulika, Perrin *et al.* 1995; Pingoud y Silva 2007). Las endonucleasas de corte infrecuente pueden ser, por ejemplo, una endonucleasa de asentamiento (Paques y Duchateau 2007), una nucleasa de dedos de cinc (ZFN) quimérica resultante de la fusión de dominios de dedos de cinc genomanipulados con el dominio catalítico de una enzima de restricción tal como FokI (Porteus y Carroll 2005) o una endonucleasa química (Eisenschmidt, Lanio *et al.* 2005; Arimondo, Thomas *et al.* 2006; Simon, Cannata *et al.* 2008). En endonucleasas químicas, se conjuga un agente de escisión químico o peptídico con un polímero de ácidos nucleicos o con otro ADN que reconoce una secuencia diana específica, dirigiendo de ese modo la actividad de escisión a una secuencia específica. Las endonucleasas químicas también abarcan nucleasas sintéticas como conjugados de ortofenantrolina, una molécula de escisión de ADN y oligonucleótidos que forman tripe hélice (TFO), que se sabe que unen secuencias de ADN específicas (Kalish y Glazer 2005). Dichas endonucleasas químicas están comprendidas en el término "endonucleasa" de acuerdo con la presente invención.

Las endonucleasas de corte infrecuente también pueden ser, por ejemplo, TALEN, una nueva clase de nucleasas quiméricas que usan un dominio catalítico FokI y un dominio de unión a ADN derivado del efector de tipo activador de transcripción (TALE), una familia de proteínas usadas en el proceso de infección por patógenos vegetales del género *Xanthomonas genus* (Boch, Scholze *et al.* 2009; Boch, Scholze *et al.* 2009; Moscou y Bogdanove 2009; Moscou y Bogdanove 2009; Christian, Cermak *et al.* 2010; Christian, Cermak *et al.* 2010; Li, Huang *et al.* 2010; Li, Huang *et al.* 2011). La disposición funcional de un TALE-nucleasa (TALEN) basada en FokI es esencialmente la de una ZFN, con el dominio de unión a ADN de dedos de cinc que se reemplaza por el dominio de TALE. Por tanto, la escisión de ADN por una TALEN requiere dos regiones de reconocimiento de ADN que flanquean una región central inespecífica. Las endonucleasas de corte infrecuente abarcadas en la presente invención también pueden obtenerse de TALEN. Los autores de la invención han desarrollado un nuevo tipo de TALEN que pueden genomanipularse para que reconozcan y procesen específicamente un ADN diana de forma eficaz. Estas "TALEN compactas" (cTALEN) novedosas no requieren dimerización para la actividad de procesamiento de ADN, aliviando de este modo la necesidad de sitios diana "dobles" con "espaciadores" de ADN intermedios; estas TALEN compactas pueden verse como una subclase de endonucleasas de corte infrecuente o endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención.

La endonucleasa de corte infrecuente puede ser una endonucleasa de asentamiento, también conocida con el nombre de meganucleasa. Dichas endonucleasas de asentamiento son bien conocidas en la técnica (Stoddard 2005). Las endonucleasas de asentamiento reconocen una secuencia diana de ADN y generan una rotura monocatenaria o bicatenaria. Las endonucleasas de asentamiento son altamente específicas, reconociendo sitios diana de ADN que varían de 12 a 45 pares de bases (pb) de longitud, habitualmente que varían de 14 a 40 pb de longitud. La endonucleasa de asentamiento de acuerdo con la invención puede corresponder, por ejemplo, a una endonucleasa LAGLIDADG, a una endonucleasa HNH o a una endonucleasa GIY-YIG.

- En la naturaleza, las meganucleasas están representadas esencialmente por las endonucleasas de asentamiento. Las endonucleasas de asentamiento (HE) son una familia generalizada de meganucleasas naturales que incluye cientos de familias de proteínas (Chevalier y Stoddard 2001). Estas proteínas están codificadas por elementos genéticos móviles que se propagan por un proceso llamado "de asentamiento": la endonucleasa escinde un alelo afín del que hay un elemento móvil ausente, estimulando de ese modo un evento de recombinación homóloga que duplica el ADN móvil en el locus destinatario. Dadas sus propiedades de escisión excepcionales en términos de eficacia y especificidad, podrían representar armazones ideales para obtener endonucleasas novedosas altamente específicas.
- Las HE pertenecen a cuatro familias principales. La familia LAGLIDADG, denominada después de un motivo peptídico conservado implicado en el centro catalítico, es el grupo más generalizado y mejor caracterizado. Ahora hay disponibles siete estructuras. Mientras que la mayoría de proteínas de esta familia son monoméricas y presentan dos motivos LAGLIDADG, unas pocas tienen únicamente un motivo y, por tanto, dimerizan para escindir secuencias diana palindrómicas o pseudopalindrómicas.
- Aunque el péptido LAGLIDADG es la única región conservada entre los miembros de la familia, estas proteínas comparten una arquitectura muy similar. El centro catalítico está flanqueado por dos dominios de unión a ADN con una simetría doble perfecta para homodímeros tales como I-Crel (Chevalier, Monnat *et al.* 2001), I-Msol (Chevalier, Turmel *et al.* 2003) y I-Ceul (Spiegel, Chevalier *et al.* 2006) y con una seudosimetría para monómeros tales como I-Scel (Moure, Gimble *et al.* 2003), I-Dmol (Silva, Dalgaard *et al.* 1999) o I-Anil (Bolduc, Spiegel *et al.* 2003). Ambos monómeros y ambos dominios (para proteínas monoméricas) contribuyen al centro catalítico, organizado alrededor de cationes divalentes. Justo por encima del centro catalítico, los dos péptidos LAGLIDADG también desempeñan una función esencial en la superficie de contacto de dimerización. La unión a ADN depende de dos plegamientos $\alpha\beta\alpha\beta\beta\alpha$ con forma de silla de montar típicos, que se asientan en el surco mayor del ADN. Otros dominios pueden encontrarse, por ejemplo, en inteínas tales como PI-Pful (Ichihyanagi, Ishino *et al.* 2000) y PI-Scel (Moure, Gimble *et al.* 2002), cuyo dominio de corte y empalme proteínico también está implicado en la unión a ADN.
- La generación de meganucleasas quiméricas funcionales, por fusión del dominio I-Dmol en el extremo N con un monómero I-Crel (Chevalier, Kortemme *et al.* 2002; Epinat, Arnould *et al.* 2003); solicitud PCT internacional WO 03/078619 (Collectis) y documento WO 2004/031346 (Fred Hutchinson Cancer Research Center, Stoddard *et al.*) han demostrado la plasticidad de las proteínas LAGLIDADG.
- Diferentes grupos también han usado una estrategia de semirrotación para alterar de forma local la especificidad de I-Crel (Seligman, Stephens *et al.* 1997; Sussman, Chadsey *et al.* 2004); Solicitudes PCT internacionales WO 2006/097784, WO 2006/097853, WO 2007/060495 y WO 2007/049156 (Collectis); (Arnould, Chames *et al.* 2006; Rosen, Morrison *et al.* 2006; Smith, Grizot *et al.* 2006), I-Scel (Doyon, Pattanayak *et al.* 2006), PI-Scel (Gimble, Moure *et al.* 2003) y I-Msol (Ashworth, Havranek *et al.* 2006).
- Además, se genomanipularon cientos de derivados de I-Crel con especificidad localmente alterada combinando la estrategia de semirrotación y el cribado de alto rendimiento:
- Los restos Q44, R68 y R70 o Q44, R68, D75 y 177 de I-Crel se mutagenizaron y se identificó una colección de variantes con especificidad alterada en la posición ± 3 a 5 de la diana de ADN (diana de ADN 5NNN) por cribado (solicitudes PCT internacionales WO 2006/097784 y WO 2006/097853 (Collectis); (Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).
 - Los restos K28, N30 y Q38 o N30, Y33 y Q38 o K28, Y33, Q38 y S40, de I-Crel se mutagenizaron y se identificó una colección de variantes con especificidad alterada en las posiciones ± 8 a 10 de la diana de ADN (diana de ADN 10NNN) por cribado (Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006); solicitudes PCT internacionales WO 2007/060495 y WO 2007/049156 (Collectis)).
- Se combinaron dos variantes diferentes y se ensamblaron en una endonucleasa heterodimérica funcional que puede escindir una diana quimérica resultante de la fusión de dos mitades diferentes de cada secuencia diana de ADN variante ((Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006); solicitudes PCT internacionales WO 2006/097854 y WO 2007/034262).
- Además, los restos 28 a 40 y 44 a 77 de I-Crel demostraron formar dos subdominios funcionales parcialmente separables, que pueden unirse a distintas partes del semisitio diana de endonucleasa de asentamiento Smith, Grizot *et al.* 2006); solicitudes PCT internacionales WO 2007/049095 y WO 2007/057781 (Collectis)).
- La combinación de mutaciones de los dos subdominios de I-Crel dentro del mismo monómero permitió el diseño de moléculas quiméricas novedosas (homodímeros) que pueden escindir una secuencia diana de ADN combinada palindrómica que comprende los nucleótidos en las posiciones ± 3 a 5 y ± 8 a 10 que se unen por cada subdominio ((Smith, Grizot *et al.* 2006); solicitudes PCT internacionales WO 2007/049095 y WO 2007/057781 (Collectis)).
- El método para producir variantes de meganucleasa y los ensayos basados en recombinación inducida por escisión en células de mamífero o levadura, que se usan para cribar variantes con especificidad alterada se describen en la solicitud PCT internacional WO 2004/067736; (Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould,

Chames *et al.* 2006). Estos ensayos producen un gen indicador LacZ funcional que puede controlarse por métodos convencionales.

5 La combinación de las dos etapas anteriores permite una estrategia combinatoria mayor, que implica cuatro subdominios diferentes. Los diferentes subdominios pueden modificarse por separado y combinarse para obtener una variante de meganucleasa completamente rediseñada (molécula heterodimérica o monocatenaria) con especificidad elegida. En una primera etapa, se combinan pares de meganucleasas novedosas en nuevas moléculas ("semimeganucleasas") que escinden dianas palindrómicas derivadas de la diana que se quiere escindir. Después, la combinación de dichas "semimeganucleasas" puede producir una especie heterodimérica que escinde la diana de interés. El ensamblaje de cuatro conjuntos de mutaciones en endonucleasas heterodiméricas que escinden una secuencia diana modelo o una secuencia de diferentes genes se ha descrito en las siguientes solicitudes de patente internacionales de Collectis: gen XPC (documento WO2007/093918), gen RAG (documento WO2008/010093), gen HPRT (documento WO2008/059382), gen de beta 2 microglobulina (documento WO2008/102274), gen Rosa26 (documento WO2008/152523), gen de la hemoglobina beta humana (documento WO2009/13622) y gen de la cadena gamma del receptor de interleucina 2 humano (documento WO2009019614).

Estas variantes pueden usarse para escindir secuencias cromosómicas genuinas y han abierto camino para perspectivas novedosa en varios campos, incluyendo genoterapia.

20 Ejemplos de dicha endonucleasa incluyen *I-Sce I*, *I-Chu I*, *I-Cre I*, *I-Csm I*, *PI-Sce I*, *PI-Tli I*, *PI-Mtu I*, *I-Ceu I*, *I-Sce II*, *I-Sce III*, *HO*, *PI-Civ I*, *PI-Ctr I*, *PI-Aae I*, *PI-Bsu I*, *PI-Dha I*, *PI-Dra I*, *PI-Mav I*, *PI-Mch I*, *PI-Mfu I*, *PI-Mfl I*, *PI-Mga I*, *PI-Mgo I*, *PI-Min I*, *PI-Mka I*, *PI-Mle I*, *PI-Mma I*, *PI-Msh I*, *PI-Msm I*, *PI-Mth I*, *PI-Mtu I*, *PI-Mxe I*, *PI-Npu I*, *PI-Pfu I*, *PI-Rma I*, *PI-Spb I*, *PI-Ssp I*, *PI-Fac I*, *PI-Mja I*, *PI-Pho I*, *PI-Tag I*, *PI-Thy I*, *PI-Tko I*, *PI-Tsp I*, *I-Msol*.

25 Una endonucleasa de asentamiento puede ser una endonucleasa LAGLIDADG tal como *I-SceI*, *I-Crel*, *I-Ceul*, *I-Msol*, y *I-Dmol*.

30 Dicha endonucleasa LAGLIDADG puede ser *I-Sce I*, un miembro de la familia que contiene dos motivos LAGLIDADG y funciona como monómero, siendo su masa molecular aproximadamente dos veces la masa de otros miembros de la familia como *I-Crel* que contiene únicamente un motivo LAGLIDADG y funciona como homodímeros.

35 Las endonucleasa mencionadas en la presente solicitud, abarcan endonucleasas tanto de tipo silvestre (de origen natural) como variantes. Las endonucleasas de acuerdo con la invención pueden ser una endonucleasa "variante", es decir, una endonucleasa que no existe de forma natural en la naturaleza y que se obtiene por ingeniería genética o por mutagénesis aleatoria, es decir, una endonucleasa genomanipulada. Esta endonucleasa variante puede obtenerse, por ejemplo, por sustitución de al menos un resto en la secuencia de aminoácidos de una endonucleasa de origen natural de tipo silvestre con un aminoácido diferente. Dicha sustitución o sustituciones pueden introducirse, por ejemplo, por mutagénesis dirigida al sitio y/o por mutagénesis aleatoria. En el marco de la presente invención, dichas endonucleasas variantes permanecen funcionales, es decir, retienen la capacidad de reconocer (función de unión) y opcionalmente escindir específicamente una secuencia diana para iniciar el proceso de dirección génica.

40 La endonucleasa variante de acuerdo con la invención escinde una secuencia diana que es diferente de la secuencia diana de la endonucleasa de tipo silvestre correspondiente. Los métodos para obtener dichas endonucleasas variantes con especificidades novedosas son bien conocidos en la técnica.

45 Las endonucleasas variantes pueden ser homodímeros (meganucleasa que comprende dos monómeros idénticos) o heterodímeros (meganucleasa que comprende dos monómeros no idénticos). Se entiende que el alcance de la presente invención también abarca endonucleasas variantes *per se*, incluyendo heterodímeros (documento WO2006097854), heterodímeros obligados (documento WO2008093249) y meganucleasas monocatenarias (documento WO03078619 y documento WO2009095793) como ejemplos no limitantes, que pueden escindir una diana de interés en una secuencia polinucleotídica o en un genoma. La invención también abarca una variante híbrida *per se* compuesta de dos monómeros de diferentes orígenes (documento WO03078619).

50 Las endonucleasas con especificidades novedosas pueden usarse en el método de acuerdo con la presente invención para dirección génica y de ese modo integrar un transgén de interés en un genoma en una ubicación predeterminada.

55 - Por "meganucleasa precursora" se pretende indicar una meganucleasa de tipo silvestre o una variante de dicha meganucleasa de tipo silvestre con propiedades idénticas o, como alternativa, una meganucleasa con algunas características alteradas en comparación con la versión de tipo silvestre de la misma meganucleasa. Esta expresión también puede trasladarse a una endonucleasa, una endonucleasa de corte infrecuente, una endonucleasa de corte infrecuente quimérica, una TALEN o una TALEN compacta y derivados.

60 - Por "vector de suministro" o "vectores de suministro" se pretende cualquier vector de suministro que pueda usarse en la presente invención para poner en contacto celular (es decir, "contactar") o suministrar el interior de las células o compartimentos subcelulares agentes/productos químicos y moléculas (proteínas y ácidos nucleicos) necesarios en la presente invención. Incluye, aunque sin limitación, vectores liposómicos de suministro, vectores víricos de suministro, vectores de suministro de fármacos, vehículos químicos, vehículos poliméricos, lipoplexos, poliplexos,

65

dendrimeros, microburbujas (agentes de contraste de ultrasonidos), nanopartículas, emulsiones u otros vectores de transferencia apropiados. Estos vectores de suministro permiten el suministro de moléculas, agentes químicos, macromoléculas (genes, proteínas) u otros vectores tales como plásmidos, péptidos desarrollados por Diatos. En estos casos, los vectores de suministro son vehículos moleculares. Por "vector de suministro" o "vectores de

- 5 suministro" también se pretenden métodos de suministro para realizar transfección.
- Los términos "vector" o "vectores" se refieren a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Un "vector" en la presente invención incluye, aunque sin limitación, un vector vírico, un plásmido, un vector de ARN o una molécula de ADN o ARN lineal o circular que puede consistir en un ácido nucleico cromosómico, no cromosómico, semisintético o sintético. Los vectores preferidos son aquellos con capacidad de replicación autónoma (vector episómico) y/o expresión de ácidos nucleicos a los que están ligados (vectores de expresión). Se conocen grande cantidad de vectores adecuados por los expertos en la materia y están disponibles en el mercado.

15 Los vectores víricos incluyen retrovirus, adenovirus, parvovirus (por ejemplo, virus adenoasociados), coronavirus, virus de ARN de hebra negativa tales como ortomixovirus (por ejemplo, virus de la gripe), rabdovirus (por ejemplo, virus de la rabia y de la estomatitis vesicular), paramixovirus (por ejemplo, sarampión y Sendai), virus de ARN de hebra positiva tales como picornavirus y alfavirus y virus de ADN bicatenario incluyendo adenovirus, herpesvirus (por ejemplo, virus del herpes simple de tipo 1 y 2, virus de Epstein-Barr, citomegalovirus) y poxvirus (por ejemplo, vaccinia, viruela aviar y viruela del canario). Otros virus incluyen el virus Norwalk, togavirus, flavivirus, reovirus, papovavirus, hepadnavirus y virus de la hepatitis, por ejemplo. Los ejemplos de retrovirus incluyen: leucosis-sarcoma aviar, virus de tipo C, de tipo B de mamífero, virus de tipo D, grupo HTLV-BLV, lentivirus, espumavirus (Coffin, J. M., *Retroviridae: The viruses and their replication*, In *Fundamental Virology*, Tercera Edición, B. N. Fields, *et al.*, Eds., Lippincott-Raven Publishers, Filadelfia, 1996).

- 25 - Por "vector lentivírico" se entiende vectores lentivíricos basados en VIH que son muy prometedores para suministro génico a causa de su capacidad de empaquetado relativamente grande, inmunogenicidad reducida y su capacidad de transducir de forma estable con alta eficacia una gran gama de diferentes tipos celulares. Los vectores lentivíricos se generan habitualmente después de transfección transitoria de tres (empaquetamiento, envuelta y transferencia) o más plásmidos en célula productoras. Como el VIH, los vectores lentivíricos entran en la célula diana a través de la interacción de glucoproteínas de la superficie vírica con receptores en la superficie celular. Al entrar, el ARN vírico experimenta transcripción inversa, que está mediada por el complejo de transcriptasa inversa vírica. El producto de transcripción inversa es un ADN vírico lineal bicatenario, que es el sustrato para el integración vírica del ADN de células infectadas.
- 30 - Por "vectores lentivíricos de integración (o LV)", se entiende dichos vectores, como ejemplo no limitante, que pueden integrarse en el genoma de una célula diana.
- 35 - En oposición a "vectores lentivíricos no integrantes (o NILV)" se entiende vectores de suministro génico eficaces que no se integran en el genoma de una célula diana mediante la acción de la integrasa vírica.

40 Un tipo de vector preferido es un episoma, es decir, un ácido nucleico capaz de replicación extracromosómica. Los vectores preferidos son aquellos capaces de replicación autónoma y/o expresión de ácidos nucleicos a los que están unidos. Los vectores capaces de dirigir la expresión de los genes a los que están unidos de forma funcional se mencionan en este documento como "vectores de expresión". Un vector de acuerdo con la presente invención comprende, aunque sin limitación, un YAC (cromosoma artificial de levadura), un BAC (artificial bacteriano), un vector baculovírico, un fago, un fagómido, un cósmido, un vector vírico, un plásmido, un vector de ARN o una molécula de ADN o ARN lineal o circular que puede consistir en ADN cromosómico, no cromosómico, semisintético o sintético. En general, los vectores de expresión de utilidad en las técnicas de ADN recombinante a menudo están en forma de "plásmidos" que se refieren generalmente a bucles de ADN bicatenario circulares que, en su forma de vector, no están unidos al cromosoma. Se conocen grandes cantidades de vectores adecuados por los expertos en la materia. Los vectores pueden comprender marcadores de selección, por ejemplo: neomicina fosfotransferasa, histidinol deshidrogenasa, dihidrofolato reductasa, higromicina fosfotransferasa, timidina cinasa del virus del herpes simple, adenosina desaminasa, glutamina sintetasa e hipoxantina-guanina fosforribosiltransferasa para cultivo de células eucariotas; TRP1 para *S. cerevisiae*; resistencia a tetraciclina, rifampicina o ampicilina en *E. coli*. Preferiblemente dichos vectores son vectores de expresión, en los que una secuencia que codifica un polipéptido de interés se coloca bajo el control de elementos de control transcripcional y traduccional apropiados para permitir la producción o síntesis de dicho polipéptido. Por lo tanto, dicho polinucleótido está comprendido en un casete de expresión. Más particularmente, el vector comprende un origen de replicación, un promotor unido de forma funcional a dicho polinucleótido codificante, un sitio de unión al ribosoma, un sitio de corte y empalme de ARN (cuando se usa ADN genómico), un sitio de poliadenilación y un sitio de terminación de la transcripción. También puede comprender elementos potenciadores o silenciadores. La selección del promotor dependerá de la célula en que se expresa el polipéptido. Los promotores adecuados incluyen promotores específicos de tejido y/o inducibles. Los ejemplos de promotores inducibles son: promotor de la metalotioneína eucariota que se induce por niveles aumentados de metales pesados, promotor lacZ procarriota que se induce en respuesta a isopropil-β-D-tiogalacto-piranósido (JPTG) y promotor de choque térmico eucariota que se induce por temperatura aumentada. Ejemplos de promotores específicos de tejido son genes de la creatina cinasa de músculo esquelético, antígeno específico de la próstata (PSA), α antitripsina proteasa, proteína A y B tensioactiva (SP) humana, β caseína y proteína de lactosuero ácida.

- Los promotores inducibles pueden inducirse por patógenos o sobrecarga, más preferiblemente por sobrecarga como frío, calor, luz UV o altas concentraciones iónicas (revisado en Potenza C *et al.* 2004, *in vitro* Cell Dev Biol 40:1-22). El promotor inducible puede inducirse por agentes químicos (revisado en Moore, Samalova *et al.* 2006); (Padidam 2003); (Wang, Zhou *et al.* 2003); (Zuo y Chua 2000).

5 Los vectores de suministro y vectores pueden asociarse o combinarse con cualquier técnica de permeabilización celular tal como sonoporación o electroporación o derivados de estas técnicas.

- Por célula o células se pretende cualquier célula viva procariota o eucariota, líneas celulares derivadas de estos organismos para cultivos *in vitro*, células primarias de origen animal o vegetal.
- Por "célula primaria" o "células primarias" se pretenden células recogidas directamente de tejido vivo (es decir material biopsia) y establecidas para cultivo *in vitro*, que han experimentado muy pocas duplicaciones de la población y, por lo tanto, son más representativas de los componentes funcionales principales y características de tejidos de los que se obtienen, en comparación con líneas celulares tumorigénicas o inmortalizadas de forma artificial continuas. Estas células, por tanto, representan un modelo más valioso para el estado *in vivo* al que se refieren.
- En el marco de la presente invención, "células eucariotas" se refieren a una célula fúngica, vegetal o animal o una línea celular derivada de los organismos enumerados a continuación y establecida para cultivo *in vitro*. Más preferiblemente, el hongo es del género *Aspergillus*, *Penicillium*, *Acremonium*, *Trichoderma*, *Chrysosporium*, *Mortierella*, *Kluyveromyces* o *Pichia*; más preferiblemente, el hongo es de la especie *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus oryzae*, *Aspergillus terreus*, *Penicillium chrysogenum*, *Penicillium citrinum*, *Acremonium chrysogenum*, *Trichoderma reesei*, *Mortierella alpine*, *Chrysosporium lucknowense*, *Kluyveromyces lactis*, *Pichia pastoris* o *Pichia ciferrii*.

25 Más preferiblemente la planta es del género *Arabidopsis*, *Nicotiana*, *Solanum*, *Lactuca*, *Brassica*, *Oryza*, *Asparagus*, *Pisum*, *Medicago*, *Zea*, *Hordeum*, *Secale*, *Triticum*, *Capsicum*, *Cucumis*, *Cucurbita*, *Citrullis*, *Citrus*, *Sorghum*; más preferiblemente, la planta es de la especie *Arabidopsis thaliana*, *Nicotiana tabaccum*, *Solanum lycopersicum*, *Solanum tuberosum*, *Solanum melongena*, *Solanum esculentum*, *Lactuca saliva*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Oryza glaberrima*, *Oryza sativa*, *Asparagus officinalis*, *Pisum sativum*, *Medicago sativa*, *Zea mays*, *Hordeum vulgare*, *Secale cereal*, *Triticum aestivum*, *Triticum durum*, *Capsicum sativus*, *Cucurbita pepo*, *Citrullus lanatus*, *Cucumis melo*, *Citrus aurantifolia*, *Citrus maxima*, *Citrus medica*, *Citrus reticulata*.

30 Más preferiblemente la célula animal es del género *Homo*, *Rattus*, *Mus*, *Sus*, *Bos*, *Danio*, *Canis*, *Felis*, *Equus*, *Salmo*, *Oncorhynchus*, *Gallus*, *Meleagris*, *Drosophila*, *Caenorhabditis*; más preferiblemente, la célula animal es de la especie *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Sus scrofa*, *Bos taurus*, *Danio rerio*, *Canis lupus*, *Felis catus*, *Equus caballus*, *Salmo salar*, *Oncorhynchus mykiss*, *Gallus gallus*, *Meleagris gallopavo*, *Drosophila melanogaster*, *Caenorhabditis elegans*.

35 En la presente invención, la célula puede ser una célula vegetal, una célula de mamífero, una célula de pez, una célula de insecto o líneas celulares derivadas de estos organismos para cultivos *in vitro* o células primarias recogidas directamente de tejido vivo y establecidas para cultivo *in vitro*. Como ejemplos no limitantes, la célula puede ser protoplastos obtenidos de organismos vegetales enumerados anteriormente. Como ejemplos no limitantes, pueden seleccionarse líneas celulares del grupo que consiste en células CHO-K1; células HEK293; células Caco2; células U2-OS; células NIH 3T3; células NSO; células SP2; células CHO-S; células DG44; células K-562, células U-937; células MRC5; células IMR90; células Jurkat; células HepG2; células HeLa; células HT-1080; células HCT-116; células Hu-h7; células Huvec; células Molt 4.

40 Todas estas líneas celulares pueden modificarse por el método de la presente invención para proporcionar modelos de línea celulares para producir, expresar, cuantificar, detectar, estudiar un gen o una proteína de interés; estos métodos también pueden usarse para cribar moléculas biológicamente activas de interés en investigación y producción y diversos campos tales como de agentes químicos, biocombustibles, agentes terapéuticos y agronomía como ejemplos no limitantes. La inmunoterapia adoptiva usando linfocitos T manipulados genéticamente es una estrategia prometedor para el tratamiento de neoplasias y enfermedades infecciosas. Las estrategias más actuales dependen de la transferencia génica por integración aleatoria de un receptor de linfocitos T (TCR) apropiado o receptor de antígenos quimérico (CAR). La estrategia dirigida usando endonucleasa de corte infrecuente es un método alternativo eficaz y seguro para transferir genes a linfocitos T y generar linfocitos T manipulados genéticamente.

- Por "homóloga" se pretende una secuencia con suficiente identidad con otra para dar lugar a recombinación homóloga entre secuencias, más particularmente que tienen al menos un 95 % de identidad, preferiblemente un 97 % de identidad y más preferiblemente un 99 %.
- Identidad" se refiere a identidad de secuencia entre dos moléculas de ácido nucleico o polipéptidos. La identidad puede determinarse comparando una composición en cada secuencia que puede alinearse con fines de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base, entonces las moléculas son idénticas en esa posición. Un grado de similitud o identidad entre secuencias de ácido nucleico o aminoácidos es una función del número de nucleótidos idénticos o coincidentes en posiciones compartidas por las secuencias de ácido nucleico. Pueden usarse diversos algoritmos y/o programas de alineación para calcular la

identidad entre dos secuencias, incluyendo FASTA o BLAST, que están disponibles como parte del paquete de análisis de secuencias GCG (University of Wisconsin, Madison, Wis.), y pueden usarse con, por ejemplo, ajustes por defecto.

- 5 - Por "mutación" se pretende la sustitución, eliminación, inserción de uno o más nucleótidos/aminoácidos en un polinucleótido (ADNc, gen) o una secuencia polipeptídica. Dicha mutación puede afectar a la secuencia codificante de un gen o su secuencia reguladora. También puede afectar a la estructura de la secuencia genómica o la estructura/estabilidad del ARNm codificado.
- 10 - En el marco de la presente invención, la expresión "mutagénesis inducida por rotura bicatenaria" (mutagénesis inducida por DSB) se refiere a un evento de mutagénesis consecutivo a un evento de NHEJ después de una DSB inducida por endonucleasa, que da lugar a inserción/eliminación en el sitio de escisión de una endonucleasa.
- 15 - Por "gen" se entiende la unidad básica de herencia, que consiste en un segmento de ADN dispuesto de una manera lineal a lo largo de un cromosoma, que codifica una proteína específica o segmento de proteína. Un gen típicamente incluye un promotor, una región no traducida 5', una o más secuencias codificantes (exones), opcionalmente intrones, una región no traducida 3'. El gen puede comprender además un terminador, potenciadores y/o silenciadores.
- 20 - Como se usa en este documento, el término "transgén" se refiere a una secuencia que codifica un polipéptido. Preferiblemente, el polipéptido codificado por el transgén no se expresa o se expresa, pero no es biológicamente activo, en la célula, tejido o individuo en que se inserta el transgén. Mucho más preferiblemente, el transgén codifica un polipéptido terapéutico útil para el tratamiento de un individuo.
- 25 - La expresión "gen de interés" o "GOI" se refiere a cualquier secuencia de nucleótidos que codifica un producto génico conocido o putativo.
- 30 - Como se usa en este documento, el término "locus" es la ubicación física específica de una secuencia de ADN (por ejemplo, de un gen) en un cromosoma. El término "locus" habitualmente se refiere a la ubicación física específica de una secuencia diana de endonucleasa en un cromosoma. Dicho locus, que comprende una secuencia diana que se reconoce y escinde por una endonucleasa de acuerdo con la invención, se menciona como "locus de acuerdo con la invención". Además, la expresión "locus genómico de interés" se usa para calificar una secuencia de ácido nucleico en un genoma que puede ser una diana putativa para una rotura bicatenaria de acuerdo con la invención. Se entiende que el locus genómico considerado de interés de la presente invención no solamente puede calificar una secuencia de ácido nucleico que existe en el cuerpo principal de material genético (es decir, en un cromosoma) de una célula, sino también una parte de material genético que puede existir independientemente a dicho cuerpo principal de material genético tal como plásmidos, episomas, virus, transposones o en orgánulos tales como mitocondrias o cloroplastos como ejemplos no limitantes.
- 35 - Por la expresión "pérdida de información genética" se entiende la eliminación o adición de al menos un fragmento de ADN dado (al menos un nucleótido) o secuencia, que limita los sitios de reconocimiento de las endonucleasas, endonucleasas de corte infrecuente quiméricas, TALEN compactas o TALEN compactas potenciada de la presente invención o la secuencia intermedia entre al menos dos sitios de procesamiento de las endonucleasas, endonucleasas de corte infrecuente quiméricas, TALEN compacta o TALEN compacta potenciada y que da lugar a un cambio de la secuencia original alrededor de dichos sitios de corte por endonucleasa, sitios de corte por endonucleasa de corte infrecuente quimérica, sitio de unión a ADN de reconocimiento de TALEN compacta o TALEN compacta potenciada dentro del locus genómico de interés. Esta pérdida de información genética puede ser, como ejemplo no limitante, la eliminación de una secuencia intermedia entre dos sitios de corte por endonucleasa o entre dos sitios de procesamiento de una TALEN compacta o TALEN compacta potenciada. Como otro ejemplo no limitante, esta pérdida de información genética también puede ser una escisión de una única hebra de ADN que abarca la región de unión de una TALEN compacta o una TALEN compacta potenciada.
- 40 - Por "religamiento sin marcas" se pretende el evento de religamiento perfecto, sin pérdida de información genética (sin eventos de inserción/eliminación) de los extremos rotos de ADN mediante un proceso de NHEJ después de la creación de un evento de rotura bicatenaria.
- 45 - Por "NHEJ imprecisa" se pretende el religamiento de extremos de ácidos nucleicos generados por una DSB, con inserciones o eliminaciones de nucleótidos. La NHEJ imprecisa es un resultado y no una ruta de reparación y puede estar producida por diferentes rutas de NHEJ (dependientes de Ku o independientes de Ku como ejemplos no limitantes).
- 50 - Por "proteína de fusión" se pretende el resultado de un proceso bien conocido en la técnica que consiste en la unión de dos o más genes que originalmente codifican proteínas separas o parte de las mismas, reduciendo la traducción de dicho "gen de fusión" un único polipéptido con propiedades funcionales derivadas de cada una de las proteínas originales.
- 55 - Por "endonucleasa de corte infrecuente quimérica" se entiende cualquiera proteína de fusión que comprenda una endonucleasa de corte infrecuente. Dicha endonucleasa de corte infrecuente podría estar en la parte del extremo N de dicha endonucleasa de corte infrecuente quimérica; en oposición, dicha endonucleasa de corte infrecuente podría estar en la parte del extremo C de dicha endonucleasa de corte infrecuente quimérica. Una "endonucleasa de corte infrecuente quimérica" de acuerdo con la presente invención que comprende dos dominios catalíticos puede describirse como "bifuncional" o como "meganucleasa bifuncional". Una "endonucleasa de corte infrecuente quimérica" de acuerdo con la presente invención que comprende más de dos dominios catalíticos puede describirse como "multifuncional" o como "meganucleasa multifuncional". Como ejemplos no limitantes, las endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención pueden ser una proteína de fusión entre una endonucleasa de corte infrecuente y un dominio catalítico; las endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención también pueden ser una proteína de fusión entre una endonucleasa de corte
- 60
- 65

- infrecuente y dos dominios catalíticos. Como se menciona previamente, la parte de endonucleasa de corte infrecuente de endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención puede ser una meganucleasa que comprende dos monómeros idénticos, dos monómeros no idénticos o una meganucleasa de una cadena. La parte de endonucleasa de corte infrecuente de endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención también puede ser el dominio de unión a ADN de una endonucleasa de corte infrecuente inactiva. En otros ejemplos no limitantes, las endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención pueden derivar de una TALE nucleasa (TALEN), es decir, una fusión entre un dominio de unión a ADN derivado de un efector de tipo activador de transcripción (TALE) y uno o dos dominios catalíticos. En otros ejemplos no limitantes, una subclase de endonucleasas de corte infrecuente quiméricas de acuerdo con la presente invención puede ser una "TALE nucleasa compacta" (cTALEN), es decir, una fusión entre un armazón de TALE central genomanipulado que comprende al menos un dominio de regiones de dipéptido variable de repetición y al menos un dominio catalítico, constituyendo dicha proteína de fusión una entidad activada de TALEN compacta que no requiere dimerización para la actividad de procesamiento de ADN. Dicho dominio catalítico puede ser una endonucleasa enumerada en la tabla 2 como ejemplos no limitantes; dicho dominio catalítico puede ser una endonucleasa de corte frecuente tal como una enzima de restricción seleccionada del grupo que consiste en Mmel, R-HinPII, R.MspI, R.MvaI, Nb.BsrDI, BsrDI A, Nt.BspD6I, ss.BspD6I, R.PleI, MlyI y AlwI como ejemplos de enzimas de restricción no limitantes enumerados en la tabla 2.
- Por "dominio o dominios potenciadores" o "potenciador o potenciadores" se entiende dominios proteínicos que proporcionan soporte funcional y/o estructural a un armazón proteínico, una TALEN compacta como ejemplo no limitante, por lo tanto, que permite una potenciación en la eficacia de procesamiento de ADN global de la proteína de fusión resultante, es decir, una TALEN compacta potenciada, respecto a la eficacia de procesamiento de ADN de la TALEN compacta de partida. Un dominio particular es un dominio potenciador cuando proporciona al menos un 5 % de potenciación en la eficacia del armazón de partida, más preferiblemente un 10 %, de nuevo más preferiblemente un 15 %, de nuevo más preferiblemente un 20 %, de nuevo más preferiblemente un 25 %, de nuevo más preferiblemente un 50 %, de nuevo más preferiblemente mayor de un 50 %. Ejemplos no limitantes de dichos dominios potenciadores se dan en las tablas 1 y 2. Por "dominios potenciadores auxiliares" o "potenciadores auxiliares" o "dominios auxiliares" se entienden dominios proteínicos que actúan en trans con una TALEN compacta o una TALEN compacta potenciada para proporcionar una función adicional que no es esencial para dicha actividad de TALEN compacta básica o dicha actividad de TALEN compacta potenciada. Cuando dichos potenciador auxiliares se usan, la TALEN compacta o TALEN compacta potenciada se convierte en "TALEN trans", respectivamente, TALEN compacta trans y TALEN compacta potenciada trans.
 - Por "dominio catalítico" se pretende el dominio proteínico o módulo de una enzima que contiene el sitio activo de dicha enzima; por sitio activo se pretende la parte de dicha enzima en que se produce la catálisis del sustrato. Las enzimas, pero también sus dominios catalíticos, se clasifican y nombran de acuerdo con la reacción que catalizan. El número de la Comisión de Enzimas (número FC) es una esquema de clasificación numérica para las enzimas, basado en las reacciones químicas que catalizan (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Como se describe en la presente invención, cualquier dominio catalítico puede fusionarse a un armazón de TALE central genomanipulado para generar una entidad activa de TALEN compacta con una eficacia de procesamiento de ADN proporcionada por al menos dicha actividad de dominio catalítico. Dicho dominio catalítico puede proporcionar una actividad nucleasa (actividades endonucleasa o exonucleasa, cleavasa o nickasa), una actividad polimerasa, una actividad cinasa, una actividad fosfatasa, una actividad metilasa, una actividad topoisomerasa, una actividad integrasa, una actividad transposasa o una actividad ligasa como ejemplos no limitantes. Ejemplos no limitantes de dichos dominios catalíticos se dan en las tablas 1 y 2. Como también se divulga en este documento, dicho dominio catalítico puede considerarse un dominio potenciador. Si es catalíticamente activo, dicho dominio potenciador puede proporcionar soporte funcional y/o estructural al armazón de TALEN compacta que da lugar a una TALEN compacta potenciada cuando se fusiona con la misma. Si es catalíticamente inactivo, dicho dominio potenciador proporciona soporte estructural al armazón de TALEN compacta que da lugar a una TALEN compacta potenciada cuando se fusiona con la misma. Se describe la fusión de dominios catalíticos de acuerdo con la presente invención a una parte de una TALEN clásica para dar a estas TALEN clásicas nuevas propiedades catalíticas proporcionadas por al menos dicha actividad de dominio catalítico o para mejorar su eficacia de procesamiento de ADN.
 - Por "dominio catalítico nucleasa" se pretende el dominio proteínico que comprende el sitio activo de una enzima endonucleasa o exonucleasa. Dicho dominio catalítico nucleasa puede ser, por ejemplo, un "dominio cleavasa" o un "dominio nickasa". Por "dominio cleavasa" se pretende un dominio proteínico cuya actividad catalítica genera una rotura bicatenaria (DSB) en una diana de ADN. Por "dominio nickasa" se pretende un dominio proteínico cuya actividad catalítica genera una rotura monocatenaria en una secuencia diana de ADN. Ejemplos no limitantes de dichos dominios catalíticos se dan en las tablas 1 y 2 con un números de GenBank o NCBI o UniProtKB/Swiss-Prot como referencia.
 - Por una "TALE nucleasa" (TALEN) o una "TALEN clásica" se pretende una proteína de fusión que consiste en una dominio de unión a ADN derivado de un efector de tipo activador de transcripción (TALE) y un dominio catalítico FokI, que tienen que dimerizar para formar una entidad activa que pueda escindir una secuencia diana de ad.
 - Por "TALE nucleasa compacta" (cTALEN) se pretende una denominación general de acuerdo con la presente invención para una proteína de fusión entre un armazón de TALE central genomanipulado que comprende al menos un dominio de dipéptido variable de repetición y al menos un dominio catalítico, constituyendo dicha proteína de fusión una entidad activa de TALEN compacta (o cTALEN) y que no requiere dimerización para la

actividad de procesamiento de ADN. Se diseñan TALEN compactas para aliviar la necesidad de múltiples restos proteínicos independientes cuando se aborda un evento de escisión de ADN. De forma importante, la región "espaciadora" necesaria y los sitios diana dobles esenciales para la función de las TALEN actuales con innecesarios. En otras palabras, la TALEN compacta de acuerdo con la presente invención es una unidad de entidad activa que puede abordar, por sí misma, únicamente una secuencia diana de ADN bicatenario única específica de interés mediante un dominio de unión a ADN y de procesar el ADN cercano a dicha secuencia diana de ADN bicatenario única de interés. Además, como el dominio catalítico no requiere contacto de ADN específico, no hay restricciones sobre las regiones que rodean el dominio de unión a ADN de TALE central. En el alcance de la presente invención, también puede preverse alguna preferencia de secuencia en el dominio catalítico. Cuando una cTALEN comprende únicamente un dominio catalítico, la cTALEN puede calificarse como "cTALEN básica" o "cTALEN". Cuando una cTALEN comprende además al menos un "dominio potenciador", la cTALEN puede calificarse como cTALEN potenciada o una "eTALEN". Una cTALEN o una eTALEN que comprende al menos un dominio catalítico cleavasa y un dominio catalítico nickasa o al menos dos dominios catalíticos cleavasa puede calificarse específicamente como cTALEN de escisión doble o "dcTALEN". Una cTALEN o una eTALEN que actúa con un dominio auxiliar en trans se califica como TALEN compacta trans o TALEN compacta potenciada trans, siendo ambas "TALEN trans".

La descripción escrita anteriormente de la invención proporciona una manera y proceso para generarla y usarla de modo que cualquier experto en la materia esté capacitado para generar y usar la misma, estando provista esta habilitación en particular para el objeto de las reivindicaciones adjuntas, que componen una parte de la descripción original.

Como se usa anteriormente, las expresiones "seleccionado del grupo que consiste en", "elegido de", y similares incluyen mezclas de los materiales especificados.

Cuando se indica en este documento un límite o intervalo numérico, los puntos finales están incluidos. Además, todos los valores y subintervalos dentro de un límite o intervalo numérico se incluyen específicamente como escritos explícitamente.

Habiendo descrito en general esta invención, puede obtenerse una comprensión mayor por referencia a determinados ejemplos específicos, que se proporcionan en este documento con fines de ilustración únicamente, y no pretenden ser limitantes salvo que se especifique de otro modo.

Ejemplos

Ejemplo 1

Se eligió la meganucleasa I-Crel de tipo silvestre (SEQ ID NO: 106) como armazón precursor en el que fusionar el dominio catalítico de I-TevI (SEQ ID NO: 107). La I-TevI de tipo silvestre funciona como cleavasa monomérica de la familia GIY-YIG para generar una rotura bicatenaria escalonada en su ADN diana. Guiados por los datos bioquímicos y estructurales, se diseñaron construcciones de longitud variable de la región del extremo N de I-TevI que abarca el dominio catalítico completo y la región intolerante a eliminaciones de su conector (de la SEQ ID NO: 109 a la SEQ ID NO: 114). En todos salvo un caso, los fragmentos se fusionaron al extremo N de I-Crel con un conector polipeptídico de 5 restos intermedio (-QGPSG-; SEQ ID NO: 103). La construcción de fusión sin conector contenía restos (-LGPDGRKA-; SEQ ID NO: 104) similares a los del conector artificial. Como I-Crel es un homodímero, todas las construcciones de fusión contienen tres centros catalíticos (figura 4, donde "dominio catalítico" = cleavasa): el sitio activo de I-Crel natural en la superficie de contacto del dímero y un sitio activo I-TevI por monómero.

La actividad de cada meganucleasa "trifuncional" se evaluó usando nuestro ensayo de levaduras previamente descrito en las solicitudes PCT internacionales WO 2004/067736 y en (Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006). Todas las construcciones podían escindir el ADN diana C1221 con una actividad comparable a la de I-Crel de tipo silvestre (tabla 4).

Para validar la actividad del dominio catalítico I-TevI independiente del núcleo catalítico I-Crel, se hicieron mutaciones puntuales D20N para inactivar el armazón de I-Crel [SEQ ID NO: 108, de la SEQ ID NO: 115 a la SEQ ID NO: 120; Chevalier, Sussman *et al.* 2004]. Las pruebas en nuestros ensayos de levadura no mostraron actividad visible a partir de la proteína mutante de I-Crel inactivada (D20N) en solitario (tabla 4). Sin embargo, pudo observarse actividad de escisión para fusiones que tenían el dominio catalítico de I-TevI (tabla 4).

60

Tabla 4: Actividad en ensayo de levadura para fusiones de I-TevI/I-Crel. Se muestra la actividad relativa de las proteínas de tipo silvestre y de fusión en las dos dianas de proteína precursoras (C1221 para I-Crel y Tev para I-TevI). Se observa actividad máxima (++++) con cada proteína dada en su diana de ADN nativa. I-Crel N20 es una variante inactiva del armazón de I-Crel de tipo silvestre. En todos los otros casos, la actividad se detecta únicamente en la diana C1221 ya que el reconocimiento de ADN está dirigido por el armazón de I-Crel. Las variantes de fusión "N20" ilustran la actividad de escisión debida al dominio catalítico de I-TevI.

Construcción proteínica	Actividad relativa en ensayo de levadura (37 °C)	
	Diana C1221	Diana Tev
I-Crel	++++	-
I-TevI	-	++++
I-Crel_N20	-	-
hTevCre_D01	++++	-
hTevCre_D02	++++	-
hTevCre_D03	++++	-
hTevCre_D04	++++	-
hTevCre_D05	++++	-
hTevCre_D06	++++	-
hTevCre_D01_N20	++	-
hTevCre_D02_N20	++	-
hTevCre_D03_N20	++	-
hTevCre_D04_N20	++	-
hTevCre_D05_N20	-	-
hTevCre_D06_N20	-	-

La actividad relativa se clasifica como: -, sin actividad detectable; +, <25 % de actividad; ++, de 25 % a <50 % de actividad; +++, de 50 % a <75 % de actividad; +++++, de 75 % a 100 % de actividad.

Ejemplo 2

Se diseñaron armazones de fusión de proteína basándose en una forma truncada de I-Crel (SEQ ID NO: 106, I-Crel_X; SEQ ID NO: 121) y tres polipéptidos conectores diferentes (NFS1 = SEQ ID NO: 98; NFS2 = SEQ ID NO: 99; CFS1 = SEQ ID NO: 100) fusionados al extremo N o C de la proteína. Se generaron modelos estructurales en todos los casos, con el objetivo de diseñar un conector de fusión "basal" que atravesara la superficie de armazón precursor de I-Crel con poco o ningún efecto sobre su unión a ADN o actividades de escisión. Para los dos armazones de fusión en el extremo N, se usó el polipéptido que abarca los restos 2 a 153 de I-Crel, con una mutación K82A para permitir la colocación del conector. El armazón de fusión en el extremo C contiene los restos 2 a 155 de I-Crel de tipo silvestre. Para ambos tipos de armazón de fusión, el extremo "libre" del conector (es decir, en el que puede unirse un polipéptido) se diseña para que sea proximal al ADN, como se determina a partir de modelos construidos usando las estructuras de complejo de I-Crel/ADN como punto de partida (PDB id: 1g9z). Los dos armazones de fusión del extremo N de I-Crel (I-Crel_NFS1 = SEQ ID NO: 122 y I-Crel_NFS2 = SEQ ID NO: 123) y el único armazón de fusión del extremo C (I-Crel_CFS1 = SEQ ID NO: 124) se ensayaron en nuestro ensayo de levadura (véase el ejemplo 1) y se descubrió que tenía actividad similar a la de I-Crel de tipo silvestre (tabla 5).

Tabla 5: Actividad en ensayo de levadura para fusiones de ColE7/I-Crel. Se muestra la actividad relativa de las proteínas de tipo silvestre de fusión sobre la diana C1221. I-Crel_X representa una versión truncada de I-Crel basada en la estructura cristalina y se usó como base para los armazones de fusión (I-Crel_NFS1, I-Crel_NFS2 y I-Crel_CFS1). Las construcciones "N20" son variantes inactivas de los armazones basados en I-Crel respectivos. La actividad se detecta en todos los casos en los que el armazón I-Crel es inactivo o cuando se proporciona la catálisis de ADN por el dominio ColE7.

Construcción proteínica	Actividad relativa en ensayo de levadura (37 °C)
	Diana C1221
I-Crel	++++
I-Crel_X	++++
I-Crel_NFS1	++++

(continuación)

Construcción proteínica	Actividad relativa en ensayo de levadura (37 °C)
	Diana C1221
I-Crel_NFS2	++++
I-Crel-CFS1	++++
I-Crel_NFS1_N20	-
I-Crel_NFS2_N20	-
I-Crel_CFS1_N20	-
hColE7Cre_D0101	++++
hColE7Cre_D0102	++++
hCreColE7_D0101	++++
hColE7Cre_D0101_N20	+++
hColE7Cre_D0102_N20	+++
hCreColE7_D0101_N20	++

La actividad relativa se clasifica como: -, sin actividad detectable; +, <25 % de actividad; ++, de 25 % a <50 % de actividad; +++, de 50 % a <75 % de actividad; +++++, de 75 % a 100 % de actividad.

La colicina E7 es una nucleasa no específica de la familia HNH que puede procesar ADN monocatenario y bicatenario (Hsia, Chak *et al.* 2004). Guiados por los datos bioquímicos y estructurales, se seleccionó la región de ColE7 que abarca el dominio catalítico completo SEQ ID NO: 140; (Hsia, Chak *et al.* 2004). Este dominio de ColE7 se fusionó al extremo N de I-Crel_NFS1 (SEQ ID NO: 122) o I-Crel_NFS2 (SEQ ID NO: 123) para crear hColE7Cre_D0101 (SEQ ID NO: 128) o hColE7Cre_D0102 (SEQ ID NO: 129), respectivamente. Además, se generó una construcción de fusión en el extremo C, hCreColE7_D0101 (SEQ ID NO: 130), usando I-Crel_CFS1 (SEQ ID NO: 124). Como I-Crel es un homodímero, todas las construcciones de fusión contienen tres centros catalíticos (figura 4, donde "dominio catalítico" = cleavasa): el sitio activo de I-Crel natural en la superficie de contacto del dímero y un sitio activo ColE7 por monómero.

La actividad de cada meganucleasa "trifuncional" se evaluó usando nuestro ensayo de levadura (véase el ejemplo 1). Todas las construcciones podían escindir el ADN diana C1221 con una actividad comparable a la de I-Crel de tipo silvestre (tabla 5). Para validar la actividad del dominio catalítico de ColE7 independiente del núcleo catalítico de I-Crel, se hicieron mutaciones puntuales D20N para inactivar el armazón I-Crel (SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 131, SEQ ID NO: 132, la SEQ ID NO: 133; (Chevalier, Sussman *et al.* 2004)). Las pruebas en nuestros ensayos de levadura no mostraron actividad visible a partir de las proteínas mutantes I-Crel (D20N) inactivadas en solitario (tabla 5). Sin embargo, pudo observarse actividad de escisión para fusiones que tenían el dominio catalítico de ColE7 (tabla 5).

Ejemplo 3

Se generan dos armazones de TALE central en que (a) podían insertarse diferentes conjuntos de dominios RVD para cambiar la especificidad de unión a ADN y; (b) pudo adherirse una selección de dominios catalíticos, en el extremo N o C, para lograr la escisión de ADN (o mellado). Los armazones centrales (sT1: SEQ ID NO: 134 y sT2: SEQ ID NO: 135) difieren en las regiones del extremo N y C, donde sT2 es una variante truncada que carece de 152 restos de aminoácido del extremo N (Szurek, Rossier *et al.* 2002) y los últimos 220 restos del extremo C en comparación con sT1. En sT1, la región del extremo C es un truncamiento con respecto a los dominios TALE de tipo silvestre, que finaliza en un sitio de restricción definido de forma fortuita (BamHI) en la secuencia codificante de ADN.

Usando los dos armazones centrales, se generan cuatro proteínas de unión a ADN de TALE "basales" (bT1-Avr = SEQ ID NO: 136, bT2-Avr = SEQ ID NO: 137, bT1-Pth = SEQ ID NO 138 y bT2-Pth = SEQ ID NO 139) por inserción del conjunto correspondiente de dominios de repetición que reconocen las secuencias asimétricas de origen natural AvrBs3 (19 pb) y PthXo1 (25 pb) (figura 3). Se enumeran secuencias proteínicas ejemplares de los armazones basales en la SEQ ID NO: 136 a la SEQ ID NO: 139. Como tales, estos armazones pueden ensayarse *in vitro* para la capacidad de unión a ADN en dianas que tienen únicamente una sola secuencia de reconocimiento. Para la comparación con TALEN existentes, el dominio catalítico de la nucleasa FokI (SEQ ID NO: 368 y particularmente los restos P381 a F583 como ejemplo no limitante) puede fusionarse al extremo N o C de los armazones basales. La escisión eficaz usando estos controles requiere ADN de sitio diana que contengan dos secuencias de unión a TALE.

Además de verificar la actividad usando secuencias de origen natural, se generaron cinco construcciones de RVD artificiales que reconocen secuencias relevantes (figura 3): RagT2-R, NptIIIT5-L, NptIIIT5-R, NptIIIT6-L, NptIIIT6-R. Se enumeran secuencias proteínicas ejemplares de las RVD insertadas en la SEQ ID NO: 253 a la SEQ ID NO: 257. Se usan secuencias de RVD artificiales como se indica anteriormente dentro del armazón sT1 o sT2 para generar las TALEN compactas dirigidas deseadas.

Se generan TALEN compactas básicas (cTALEN) mediante fusión de dominios catalíticos al extremo N o C de los armazones basales (figura 5, A o B, respectivamente). Se presenta una lista no exhaustiva de dominios catalíticos susceptibles a fusión con dominios de unión a ADN de TALE en la tabla 2. Se presenta una lista no exhaustiva de conectores que pueden usarse en la tabla 3. Es apreciable que el diseño del conector puede depender de la naturaleza del dominio catalítico adherido y su aplicación dada. También puede preverse que pueden construirse conectores especialmente diseñados para controlar mejor o regular la actividad de alguno o ambos dominios. Los ejemplos 5, 6 y 7 a continuación analizan métodos adicionales y alternativos en que pueden definirse los conectores. Todos los diseños de cTALEN se evalúan usando nuestro ensayo de levadura (véase el ejemplo 1) y proporcionan actividad detectable comparable con las meganucleasas diseñadas existentes.

Ejemplo 3a: TALEN compacta TALE::TevI

El dominio catalítico de I-TevI (SEQ ID NO: 20), un miembro de la familia de endonucleasa GIY-YIG, se fusionó a un armazón derivado de TALE (compuesto de un dominio en el extremo N, un núcleo central compuesto de RVD y un dominio del extremo C) para crear una nueva clase de cTALEN (TALE::TevI). Para distinguir la orientación (extremo N frente a extremo C) de las fusiones del dominio catalítico (CD), los nombres de las construcciones se escriben como CD::TALE-RVD (el dominio catalítico está fusionado en el extremo N al dominio de TALE) o TALE-RVD::CD (el dominio catalítico está fusionado en el extremo C al dominio de TALE), donde "-RVD" indica opcionalmente la secuencia reconocida por el dominio de TALE y "CD" es el tipo de dominio catalítico. En este documento, se describen construcciones TALE::TevI novedosas que abordan la secuencia AvrBs3, por ejemplo, llamada, por tanto, TALE-AvrBs3::TevI.

Actividad de TALE::TevI en levadura

Se seleccionó un armazón de TALE central, sT2 (SEQ ID NO: 135) en que (a) podían insertarse diferentes conjuntos de dominios de RVD para cambiar la especificidad de unión a ADN, y; (b) podía adherirse una selección de dominios catalíticos derivados de I-TevI, en el extremo N o C, para lograr la escisión de ADN (o mellado). El armazón truncado sT2 mencionado previamente se generó por PCR de un molde de armazón de TALEN central de longitud completa (pCLS7183, SEQ ID NO: 141) usando los cebadores CMP_G061 (SEQ ID NO: 142) y CMP_G065 (SEQ ID NO: 143) y se clonó en el vector pCLS7865 (SEQ ID NO: 144) para generar pCLS7865-cTAL11_CFS1 (pCLS9009, SEQ ID NO: 145), donde CFS1 indica la secuencia de aminoácidos -GSSG-(con sitios de restricción subyacentes BamHI y KpnII en el ADN codificante para facilitar la clonación). Se amplificaron tres variantes del dominio catalítico I-TevI (SEQ ID NO: 20) por PCR en los moldes TevCreD01 [proteína de la SEQ ID NO: 109 en el plásmido pCLS6614 (SEQ ID NO: 146)] usando el par de cebadores CMP_G069 (SEQ ID NO: 147) y CMP_G070 (SEQ ID NO: 148), TevCreD02 [proteína de la SEQ ID NO: 110 en el plásmido pCLS6615 (SEQ ID NO: 203)] usando el par de cebadores CMP_G069 (SEQ ID NO: 147) y CMP_G071 (SEQ ID NO: 149) o TevCreD05 [proteína de la SEQ ID NO: 113 en el plásmido pCLS6618 (SEQ ID NO: 258)] usando el par de cebadores CMP_G069 (SEQ ID NO: 147) y CMP_G115 (SEQ ID NO: 259) y se subclonó en la estructura pCLS9009 por restricción y ligamiento usando los sitios de restricción BamHI y EagI, produciendo pCLS7865-cT11_TevD01 (pCLS9010, SEQ ID NO: 150), pCLS7865-cT11_TevD02 (pCLS9011, SEQ ID NO: 151) y pCLS7865-cT11_TevD05 (pCLS15775, SEQ ID NO: 260), respectivamente. Todas las fusiones contienen el dipéptido -GS- que une el dominio de unión a ADN derivado de TALE y el dominio catalítico derivado de I-TevI.

La secuencia de ADN que codifica los RVD para abordar el sitio AvrBs3 (SEQ ID NO: 152) se subclonó en ambos plásmidos pCLS9010 (SEQ ID NO: 150, que codifica la proteína de la SEQ ID NO: 420), pCLS9011 (SEQ ID NO: 151, que codifica la proteína de la SEQ ID NO: 421) y pCLS15775 (SEQ ID NO: 260, que codifica la proteína de la SEQ ID NO: 422) usando enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia de RVD insertada para crear las construcciones TALE-AvrBs3::TevI posteriores cT11Avr_TevD01 (pCLS9012, SEQ ID NO: 218, que codifica la proteína de la SEQ ID NO: 423), cT11Avr_TevD02 (pCLS9013, SEQ ID NO: 153, que codifica la proteína de la SEQ ID NO: 424) y cT11Avr_TevD05 (pCLS15776, SEQ ID NO: 261, que codifica la proteína de la SEQ ID NO: 425), respectivamente. Estas construcciones TALE-AvrBs3::TevI se secuenciaron y el inserto se transfirió a vectores adicionales según lo necesario (véase a continuación).

Los plásmidos de expresión de levadura de TALE-AvrBs3::TevI finales, pCLS8523 (SEQ ID NO: 154), pCLS8524 (SEQ ID NO: 155) y pCLS12092 (SEQ ID NO: 262), se prepararon por clonación *in vivo* en levadura usando los plásmidos pCLS9012, pCLS9013 y pCLS15776, respectivamente. Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de cada plásmido linealizado por digestión con BssHII y 1 ng del ADN del plásmido pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar, respectivamente, la levadura *S. cerevisiae* cepa FYC2-6A (*MATa, trp1Δ63, leu2Δ1, his3Δ200*) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones TALE-AvrBs3::TevI se ensayaron en un ensayo de SSA de levadura como se describe previamente

(solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6). Los niveles de actividad de TALE-AvrBs3 en sus respectivas dianas en células de levadura se muestran en la figura 9. Los datos resumidos en la figura 9 muestran que TALE-AvrBs3::Tev es activa frente a varias dianas en levadura.

Actividad de TALE::Tevl en células de mamífero

Se subclonó el ADN que codifica la construcción TALE-AvrBs3::Tevl a partir de pCLS9012 (SEQ ID NO: 218) o pCLS9013 (SEQ ID NO: 153) en el plásmido de expresión de mamífero pCLS1853 (SEQ ID NO: 193) usando las enzimas de restricción *Ascl* y *XhoI* para el plásmido receptor y las enzimas de restricción *Bss*III y *XhoI* para el inserto TALE-AvrBs3::Tevl, que da lugar a los plásmidos de expresión de mamífero pCLS8993 y pCLS8994 (SEQ ID NO: 194 y 195), respectivamente.

Todos plásmidos indicadores diana de mamífero que contenían las secuencias diana de ADN de TALEN se construyeron usando el protocolo de Gateway convencional (INVITROGEN) en un vector indicador CHO (Arnould, Chames *et al.* 2006, Grizot, Epinat *et al.* 2010). Las construcciones TALE-AvrBs3::Tevl se ensayaron en un ensayo extracromosómico en células de mamífero (CHO K1) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI, que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6).

Para este ensayo, se transfectaron células CHO K1 en un formato de placa de 96 pocillos con 75 ng de vector diana y una cantidad creciente de cada ADN variante de 0,7 a 25 ng, en presencia de reactivo PolyFect (1 μ l por pocillo). La cantidad total de ADN transfectado se completó a 125 ng (ADN diana, ADN variante, ADN transportador) usando un vector vacío. Setenta y dos horas después de la transfección, se retiró el medio de cultivo y se añadieron 150 μ l de tampón de lisis/revelación para ensayo líquido de β -galactosidasa. Después de incubación a 37 °C, se midió la densidad óptica a 420 nm. El proceso completo se realiza en una plataforma automatizada Velocity11 BioCel (Grizot, Epinat *et al.* 2009).

Se muestran los niveles de actividad en células de mamífero para las construcciones TALE-AvrBs3::Tevl (12,5 ng de ADN transfectado) en la diana Avr15 (SEQ ID NO: 167) en la figura 10. Parece que TALE-AvrBs3::Tevl es eficaz para escindir la secuencia diana.

Actividad nickasa de TALE::Tevl

Los resultados descritos en los ejemplos anteriores ilustran dos fusiones TALE::Tevl, que contienen cada una un dominio de unión a ADN basado en TALE y un dominio catalítico basado en I-Tevl, que funcionan generando actividad detectable. Los ensayos usados miden la recombinación de repeticiones en tándem por hibridación monocatenaria, un proceso que se activa esencialmente por una DSB (Sugawara y Haber 1992; Paques y Duchateau 2007). Las fusiones TALE::Tevl pueden tener una actividad nickasa insuficiente para activar en solitario una señal en el ensayo basado en células. Sin embargo, dos proteínas TALE::Tevl que se unen a dos sitios cercanos a veces pueden generar dos mellas independientes, que cuando están proximales o en hebras de ADN diferentes pueden crear una DSB. En este caso, cada TALE::Tevl es una cTALEN que puede generar una mella.

Se configuran diferentes experimentos para medir la actividad nickasa de TALE::Tevl:

50 Ensayo de mellado y/o linealización de plásmido circular superenrollado

Las secuencias que codifican las construcciones de TALE-AvrBs3::Tevl cT11Avr_TevD01 y cT11Avr_TevD02 se clonan en un vector de expresión basado en T7 usando sitios de restricción de *NcoI*/*EagI* para producir los plásmidos pCLS9021 (SEQ ID NO: 201) y pCLS9022 (SEQ ID NO: 202), respectivamente. Esta etapa de clonación produce las proteínas TALE-AvrBs3::Tevl que tienen una marca de hexa-His adicional para la purificación. Los plásmidos pCLS9021 y pCLS9022 entonces se usan para producir proteínas activas por uno de dos métodos:

1. Se usan plásmidos en un sistema de transcripción/traducción *in vitro* convencional; los lisados de la traducción se usan directamente sin purificación adicional.

2. Se usan plásmidos para transformar células de *E. coli* L21(DE3) para su expresión usando protocolos convencionales, concretamente: crecimiento hasta fase logarítmica, inducción con IPTG, recolección, lisis celular y purificación mediante métodos de afinidad para proteínas marcadas con His.

Las proteínas activas se ensayan frente a las dianas de ADN que tienen ninguna, una o dos secuencias de sitio de reconocimiento de AvrBs3. Cuando hay más de un sitio presente, se yuxtaponen secuencias de reconocimiento

idénticas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb.

Se realiza un ensayo de mellado y/o linealización de plásmido circular superenrollado. Los plásmidos que albergan las dianas de ADN descritas anteriormente se preparan por métodos convencionales y se purifican en columna para producir plásmido superenrollado de >98 % de pureza. Se incuban cantidades crecientes de las proteínas TALE-AvrBs3::TevI (preparadas como se describe anteriormente) con cada plásmido en condiciones para promover la escisión de ADN durante 1 h a 37 °C. Los productos de reacción se separan en geles de agarosa y se visualizan por tinción con EtBr.

10 *Ensayo de mellado y/o escisión de ADN lineal*

Se realiza un ensayo de mellado y/o escisión de ADN lineal. Se preparan productos de PCR que contienen las secuencias diana descritas anteriormente por métodos convencionales y se purifican en columna para producir el sustrato lineal de >98 % de pureza. Entonces se incuban cantidades crecientes de las proteínas TALE-AvrBs3::TevI (preparadas como se describe anteriormente) con cada sustrato de PCR en condiciones para promover la escisión del ADN durante 1 h a 37 °C. Los productos de reacción se separan en un gel de acrilamida desnaturalizante y se visualiza el ADN monocatenario.

20 Genomanipulación de la TALE::TevI

Se eligen variantes que difieren por truncamientos del dominio del extremo C de la TALEN derivado de AvrBs3 (SEQ ID NO: 196) como armazones de partida. Un subconjunto de estas variantes incluye el truncamiento después de las posiciones E886 (CO), P897 (C11), G914 (C28), L926 (C40), D950 (C64), R1000 (C115), D1059 (C172) (los dominios proteínicos de los dominios en el extremo C truncados C11 a C172 se dan respectivamente en la SEQ ID NO: 204 a 209) y P1117 [también mencionado como Cter wt o WT Cter (SEQ ID NO: 210) que carece del dominio de activación del dominio del extremo C de AvrBs3 natural (SEQ ID NO: 220)]. Los plásmidos que codifican los armazones variantes que contienen el dominio del extremo N derivado de AvrBs3, el conjunto derivado de AvrBs3 de dominios de repetición y el dominio en el extremo C derivado de AvrBs3 truncado [pCLS7821, pCLS7803, pCLS7807, pCLS7809, pCLS7811, pCLS7813, pCLS7817 (SEQ ID NO: 211 a 217) que se basan en pCLS7184 (SEQ ID NO: 196)] permiten la clonación de cualquier dominio catalítico en fusión con el dominio del extremo C, usando los sitios de restricción BamHI y EagI.

Se diseñan variantes del dominio catalítico de I-TevI (SEQ ID NO: 20) a partir de la región del extremo N de I-TevI. Un subconjunto de estas variantes incluye truncamientos del dominio catalítico, como la región intolerante a eliminación de su conector, la región tolerante a eliminación de su conector y su dedo de cinc (SEQ ID NO: 197 a 200) nombrado en Liu *et al.*, 2008 (Liu, Dansereau *et al.* 2008).

El ADN correspondiente a estas variantes de I-TevI se amplifica por PCR para introducir, a nivel de ADN, un sitio de restricción de BamHI (en el 5' de la hebra codificante) y uno EagI (en el 3' de la hebra codificante) y, a nivel de proteína, un conector (por ejemplo tramo -SGGSGS-, SEQ ID NO: 219) entre el dominio del extremo C de TALE y la variante del dominio catalítico de I-TevI. Las construcciones TALE::TevI finales se generan por inserción de la variante de los dominios catalíticos de I-TevI en las variantes de armazón usando BamHI y EagI y procedimientos de biología molecular convencionales.

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones TALE-AvrBs3::TevI se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianasseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6).

55 Ejemplo 3b: TALEN compacta TevI::TALE

El armazón de TALE central sT2 (SEQ ID NO: 135) descrito en el ejemplo 3a se seleccionó para generar pCLS7865-cTAL11_NFS1 (pCLS9008, SEQ ID NO: 234), donde NFS1 denomina la secuencia de aminoácidos -GSSG- (con los sitios de restricción subyacentes BamHI y Kpn2I en el DNA codificante para facilitar la clonación). Se amplificaron cuatro variantes del dominio catalítico de I-TevI (SEQ ID NO: 20) por PCR en los moldes TevCreD01 [proteína de la SEQ ID NO: 109 en el plásmido pCLS6614 (SEQ ID NO: 146)] usando los pares de cebadores CMP_G001 (SEQ ID NO: 239) y CMP_G067 (SEQ ID NO: 263) o CMP_G152 (SEQ ID NO: 264), TevCreD02 [proteína de la SEQ ID NO: 110 en el plásmido pCLS6615 (SEQ ID NO: 203)] usando el par de cebadores CMP_G001 (SEQ ID NO: 239) y CMP_G068 (SEQ ID NO: 240) o TevCreD05 [proteína de la SEQ ID NO: 113 en el plásmido pCLS6618 (SEQ ID NO: 258)] usando el par de cebadores CMP_G001 (SEQ ID NO: 239) y CMP_G114 (SEQ ID NO: 265) y se subclonó en la estructura pCLS9008 por restricción y ligamiento usando los sitios de restricción NcoI y Kpn2I, produciendo pCLS7865-

TevW01_cT11 (pCLS15777, SEQ ID NO: 266, que codifica la pretina de la SEQ ID NO: 426), pCLS7865-TevD01_cT11 (pCLS15778, SEQ ID NO: 267, que codifica la pretina de la SEQ ID NO: 427), pCLS7865-TevD02_cT11 (pCLS12730, SEQ ID NO: 235, que codifica la proteína de la SEQ ID NO: 428) y pCLS7865-TevD05_cT11 (pCLS15779, SEQ ID NO: 268, que codifica la proteína de la SEQ ID NO: 429), respectivamente. Mientras que la fusión basada en TevW01_cT11 contiene el dipéptido -SG- que une el dominio de unión a ADN derivado de TALE y el dominio catalítico derivado de I-TevI, todas las demás construcciones incorporan una pentapéptido más largo -QGPSG- para unir los dominios.

Actividad de TevI::TALE en levadura

La secuencia de ADN que codifica las RVD para abordar el sitio AvrBs3 (SEQ ID NO: 152) se subclonó en los plásmidos pCLS15777 (SEQ ID NO: 266), pCLS15778 (SEQ ID NO: 267) y pCLS12730 (SEQ ID NO: 235) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia de RVD insertada para crear las construcciones TevI::TALE-AvrBs3 posteriores TevW01_cT11Avr (pCLS15780, SEQ ID NO: 269, que codifica la proteína de la SEQ ID NO: 430), TevD01_cT11Avr (pCLS15781, SEQ ID NO: 270, que codifica la proteína de la SEQ ID NO: 431) y TevD02_cT11Avr (pCLS12731, SEQ ID NO: 236, que codifica la proteína de la SEQ ID NO: 432), respectivamente. Se usó una técnica de clonación similar para introducir las RVD para abordar el sitio RagT2-R (SEQ ID NO: 271) en el plásmido pCLS15779 (SEQ ID NO: 268) para crear la construcción posterior TevD05_cT11RagT2-R (pCLS15782, SEQ ID NO: 272). Todas las construcciones TevI::TALE se secuenciaron y los insertos se transfirieron a vectores adicionales según lo necesario (véase a continuación).

Los plásmidos de expresión de levadura basados en TevI::TALE finales, pCLS11979 (SEQ ID NO: 273), pCLS8521 (SEQ ID NO: 274), pCLS8522 (SEQ ID NO: 237) y pCLS12100 (SEQ ID NO: 275), se prepararon por clonación *in vivo* en levadura usando el plásmido pCLS15780 (SEQ ID NO: 269), pCLS15781 (SEQ ID NO: 270), pCLS12731 (SEQ ID NO: 236) y pCLS15782 (SEQ ID NO: 272), respectivamente. Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de cada plásmido linealizado por digestión con BssHII y 1 ng del ADN del plásmido pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar, respectivamente, la levadura *S. cerevisiae* cepa FYC2-6A (*MAT α* , *trp1 Δ 63*, *leu2 Δ 1*, *his3 Δ 200*) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones TevI::TALE-AvrBs3 y TevI::TALE-RagT2-R se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento de AvrBs3 o RagT2-R (SEQ ID NO: 238, tabla 8). El nivel de actividad de TevI::TALE-AvrBs3 fue comparable al de TALE-AvrBs3::TevI (pCLS8524, SEQ ID NO: 155) en dianas adecuadas. Se ilustra la actividad significativa en la tabla 8 para una diana de sitio único de muestra, de acuerdo con la cTALEN de la presente invención.

Tabla 8: Actividad de TevI::TALE-AvrBs3 y TevI::TALE-RagT2-R en dianas de ADN de sitio doble y único. La actividad relativa se clasifica como: n.d., sin actividad detectable; +, <25 % de actividad; ++, de 25 % a <50 % de actividad; +++, de 50 % a <75 % de actividad; +++, de 75 % a 100 % de actividad.

ADN diana	Construcción de TALEN		
	TevI::TALE-AvrBs3	TevI::TALE-RagT2-R	TALE-AvrBs3::FokI
Avr25 (sitio doble) (SEQ ID NO: 177)	++++	n.d.	++++
Avr25RAGT2R (sitio único) (SEQ ID NO: 238)	++	++	n.d.

Actividad de TevI::TALE en plantas

La secuencia de ADN que codifica las RVD para abordar los sitios NptIIIT5-L y NptIIIT6-L (de la SEQ ID NO: 276 a 279) se subclonaron en el plásmido pCLS12730 (SEQ ID NO: 235) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para las secuencias de RVD insertadas para crear las construcciones TevI::TALE posteriores TevD02_cT11NptIIIT5-L (pCLS15783, SEQ ID NO: 280) y TevD02_cT11NptIIIT6-L (pCLS15784, SEQ ID NO: 281), respectivamente. Las construcciones se secuenciaron y los insertos de TevI::TALE se transfirieron por técnicas de clonación convencionales al plásmido pCLS14529 (SEQ ID NO: 282) para generar los plásmidos de expresión de TevI::TALE-NptIIIT5-L y TevI::TALE-NptIIIT6-L finales, pCLS14579 (SEQ ID NO: 283) y pCLS14581 (SEQ ID NO: 284), respectivamente. El plásmido pCLS14529 permite clonar secuencias del gen de interés después de un promotor que confiere altos niveles de expresión constitutiva en células vegetales.

Para ensayar la actividad en células vegetales, se empleó un ensayo de hibridación monocatenaria (SSA) basado en YFP. El gen indicador YFP tiene una corta duplicación de la secuencia codificante que está interrumpida por un sitio diana de TALEN NptII5 o NptII6. La escisión del sitio diana estimula la recombinación entre las repeticiones, provocando la reconstitución de un gen YFP funcional. Para cuantificar la escisión, el indicador se introduce junto con una construcción que codifica una TALEN o TALEN compacta basada en FokI en protoplastos de tabaco por transformación mediada por PEG (como se conoce u obtiene del estado de la técnica). Se obtuvieron eficacias de transformación uniformes usando la misma cantidad de plásmido en cada transformación, es decir, 15 µg de cada uno de los plásmidos que codifican YFP y la TALEN o cTALEN. Después de 24 horas, los protoplastos se sometieron a citometría de flujo para cuantificar el número de células positivas a YFP. Los niveles de actividad de TevI::TALE, usando las cTALEN de acuerdo con la presente invención, en plantas fueron comparables a los de las construcciones de control de TALEN basada en FokI en las dianas ensayadas (tabla 9).

Tabla 9: Actividad de TevI::TALE-NptII5-L y TevI::TALE-NptII6-L en dianas de ADN apropiadas. La actividad relativa se clasifica respecto a las construcciones de control como: n.a., no aplicable; +, 100 % de actividad del control (2 % de células positivas a YFP)

ADN diana	Construcción de TALEN			
	TevI::TALE-NptII5-L	NptII5.1 de control	TevI::TALE-NptII6-L	NptII6.1 de control
NptII5.1	+++	+	n.a.	n.a.
NptII6.1	n.a.	n.a.	+	+

Ejemplo 3c: TALEN compacta TALE::NucA

NucA (SEQ ID NO: 26), una endonucleasa no específica de *Anabaena* sp., se fusionó a un armazón derivado de TALE (compuesto de un dominio en el extremo N, un núcleo central compuesto de RVD y un dominio en el extremo C) para crear una nueva clase de cTALEN (TALE::NucA). Para distinguir la orientación (extremo N frente a extremo C) de las fusiones del dominio catalítico (CD), los nombres de las construcciones se escriben como CD::TALE-RVD (el dominio catalítico está fusionado en el extremo N al dominio de TALE) o TALE-RVD::CD (el dominio catalítico está fusionado en el extremo C al dominio de TALE), donde "-RVD" indica opcionalmente la secuencia reconocida por el dominio de TALE y "CD" es el tipo de dominio catalítico. En este documento, se describen construcciones de TALE::NucA novedosas que abordan, por ejemplo, la secuencia AvrBs3 y, por tanto, se llaman TALE-AvrBs3::NucA. De forma notable, la endonucleasa NucA de tipo silvestre puede inhibirse mediante la formación de complejos con la proteína NuiA (SEQ ID NO: 229). En un contexto de TALEN compacta, la proteína NuiA puede funcionar como dominio auxiliar para modular la actividad nucleasa de las construcciones TALE::NucA.

Actividad de TALE::NucA en levadura

Se seleccionó un armazón de TALE central, sT2 (SEQ ID NO: 135) en que (a) podían insertarse diferentes conjuntos de dominios de RVD para cambiar la especificidad de unión a ADN, y; (b) podía adherirse una selección de dominios catalíticos derivados de NucA, en el extremo N o C, para lograr la escisión de ADN (o mellado). Como se menciona previamente, el armazón truncado sT2 se generó por PCR a partir de un molde de armazón de TALEN central de longitud completa (pCLS7183, SEQ ID NO: 141) usando los cebadores CMP_G061 (SEQ ID NO: 142) y CMP_G065 (SEQ ID NO: 143) y se clonó en el vector pCLS7865 (SEQ ID NO: 144) para generar pCLS7865-cTAL11_CFS1 (pCLS9009, SEQ ID NO: 145), donde CFS1 indica la secuencia de aminoácidos -GSSG- (con sitios de restricción subyacentes BamHI y Kpn2I en el ADN codificante para facilitar la clonación). El dominio catalítico NucA (SEQ ID NO: 26) correspondiente a los restos de aminoácido 25 a 274, se subclonó en la estructura pCLS9009 (SEQ ID NO: 145) por restricción y ligamiento usando los sitios de restricción BamHI y EagI, produciendo pCLS7865-cT11_NucA (pCLS9937, SEQ ID NO: 221, que codifica la proteína de la SEQ ID NO: 433). La fusión contiene el dipéptido -GS- que une el dominio de unión a ADN derivado de TALE y el dominio catalítico derivado de NucA. La etapa de clonación también se produce a nivel de aminoácidos en una secuencia AAD en Cter del dominio catalítico NucA.

La secuencia de ADN que codifica las RVD para abordar el sitio AvrBs3 (SEQ ID NO: 152) se subclonó en el plásmido pCLS9937 (SEQ ID NO: 221) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia de RVD insertada para crear la construcción de TALE-AvrBs3::NucA posterior cT11Avr_NucA (pCLS9938, SEQ ID NO: 222, que codifica la proteína de la SEQ ID NO: 434). La construcción TALE-AvrBs3::NucA se secuenció y el inserto se transfirió a vectores adicionales según lo necesario (véase a continuación).

El plásmido de expresión de levadura de TALE-AvrBs3::NucA final, pCLS9924 (SEQ ID NO: 223), se preparó por clonación *in vivo* en levadura usando el plásmido pCLS9938 (SEQ ID NO: 222). Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de plásmido (pCLS9938) linealizado por digestión con BssHII y 1 ng del ADN plasmídico pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar la levadura *S. cerevisiae* cepa FYC2-6A (MAT α , trp1 Δ 63, *leu2* Δ 1, *his3* Δ 200) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould

et al. 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

La construcción de TALE-AvrBs3::NucA se ensayó en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224; tabla 7).

Genomanipulación de la TALE::NucA

Se eligen variantes que difieren por truncamientos del dominio del extremo C de la TALEN derivado de AvrBs3 (SEQ ID NO: 196) como armazones de partida. Un subconjunto de estas variantes incluye el truncamiento después de las posiciones E886 (CO), P897 (C11), G914 (C28), L926 (C40), D950 (C64), R1000 (C115), D1059 (C172) (los dominios proteínicos de los dominios en el extremo C truncados C11 a C172 se dan respectivamente en la SEQ ID NO: 204 a 209) y P1117 [también mencionado como Cter wt o WT Cter (SEQ ID NO: 210) que carece del dominio de activación del dominio del extremo C de AvrBs3 natural (SEQ ID NO: 220)]. Los plásmidos que codifican los armazones variantes que contienen el dominio del extremo N derivado de AvrBs3, el conjunto derivado de AvrBs3 de dominios de repetición y el dominio en el extremo C derivado de AvrBs3 truncado [pCLS7821, pCLS7803, pCLS7807, pCLS7809, pCLS7811, pCLS7813, pCLS7817 (SEQ ID NO: 211 a 217) que se basan en pCLS7184 (SEQ ID NO: 196)] permiten la clonación de cualquier dominio catalítico en fusión con el dominio del extremo C, usando los sitios de restricción BamHI y EagI.

El ADN correspondiente a los restos de aminoácido 25 a 274 de NucA se amplifica por PCR para introducir, a nivel de ADN, un sitio de restricción de BamHI (en el 5' de la hebra codificante) y uno EagI (en el 3' de la hebra codificante) y, a nivel de proteína, un conector (por ejemplo tramo -SGGSGS-, SEQ ID NO: 219) entre el dominio del extremo C del TALE y el dominio catalítico de NucA. Las construcciones TALE::NucA finales se generan por inserción del dominio catalítico NucA en las variantes de armazón usando BamHI y EagI y procedimientos de biología molecular convencionales. Por ejemplo, las variantes de armazón truncadas después de las posiciones P897 (C11), G914 (C28) y D950 (C64), respectivamente codificadas por pCLS7803, pCLS7807, pCLS7811, (SEQ ID NO: 212, 213 y 215), se fusionaron con el dominio catalítico NucA (SEQ ID NO: 26), dando lugar a pCLS9596, pCLS9597 y pCLS9599 (de la SEQ ID NO: 225 a 227). La etapa de clonación también se produce a nivel de aminoácidos en una secuencia AAD en Cter del dominio catalítico NucA.

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones TALE-AvrBs3::NucA se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones TALE-AvrBs3::NucA se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224). Los datos resumidos en la figura 11 muestran que las construcciones TALE-AvrBs3::NucA son activas en todas las dianas que tienen al menos un sitio de reconocimiento AvrBs3, de acuerdo con la cTALEN de la presente invención.

Ejemplo 3d: TALEN compacta TALE::CoIE7

El dominio catalítico de CoIE7 (SEQ ID NO: 140), una endonucleasa no específica de *E.coli*, se fusionó a un armazón derivado de TALE (compuesto de un dominio en el extremo N, un núcleo central compuesto de RVD y un dominio en el extremo C) para crear una nueva clase de cTALEN (TALE::CoIE7). Para distinguir la orientación (extremo N frente a extremo C) de las fusiones del dominio catalítico (CD), los nombres de las construcciones se escriben como CD::TALE-RVD (el dominio catalítico está fusionado en el extremo N al dominio de TALE) o TALE-RVD::CD (el dominio catalítico está fusionado en el extremo C al dominio de TALE), donde "-RVD" indica opcionalmente la secuencia reconocida por el dominio de TALE y "CD" es el tipo de dominio catalítico. En este documento, se describen construcciones de TALE::CoIE7 novedosas que abordan, por ejemplo, la secuencia AvrBs3 y, por tanto, se llaman TALE-AvrBs3::CoIE7. De forma notable, la endonucleasa CoIE7 de tipo silvestre puede inhibirse mediante la formación de complejos con la proteína de inmunidad Im7 (SEQ ID NO: 230). En un contexto de TALEN compacta, la proteína Im7 puede funcionar como dominio auxiliar para modular la actividad nucleasa de las construcciones TALE::CoIE7.

65 Actividad de TALE::CoIE7 en levadura

Se seleccionó un armazón de TALE central, sT2 (SEQ ID NO: 135) en que (a) podían insertarse diferentes conjuntos de dominios de RVD para cambiar la especificidad de unión a ADN, y; (b) podía adherirse una selección de dominios catalíticos derivados ColE7, en el extremo N o C, para lograr la escisión de ADN (o mellado). Como se menciona previamente, el armazón truncado sT2 se generó por PCR a partir de un molde de armazón de TALEN central de longitud completa (pCLS7183, SEQ ID NO: 141) usando los cebadores CMP_G061 (SEQ ID NO: 142) y CMP_G065 (SEQ ID NO: 143) y se clonó en el vector pCLS7865 (SEQ ID NO: 144) para generar pCLS7865-cTAL11_CFS1 (pCLS9009, SEQ ID NO: 145), donde CFS1 indica la secuencia de aminoácidos -GSSG- (con sitios de restricción subyacentes BamHI y Kpn2I en el ADN codificante para facilitar la clonación). El dominio catalítico ColE7 (SEQ ID NO: 140) se subclonó en la estructura pCLS9009 por restricción y ligamiento usando los sitios de restricción Kpn2I y EagI, produciendo pCLS7865-cT11_ColE7 (pCLS9939, SEQ ID NO: 231, que codifica la proteína de la SEQ ID NO: 435). La fusión contiene el dipéptido -GSSG- que une el dominio de unión a ADN derivado de TALE y el dominio catalítico derivado de ColE7.

La secuencia de ADN que codifica las RVD para abordar el sitio AvrBs3 (SEQ ID NO: 152) se subclonó en el plásmido pCLS9939 (SEQ ID NO: 231) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia de RVD insertada para crear la construcción de TALE-AvrBs3::ColE7 posterior cT11Avr_ColE7 (pCLS9940, SEQ ID NO: 232, que codifica la proteína de la SEQ ID NO: 436). La construcción TALE-AvrBs3::ColE7 se secuenció y el inserto se transfirió a vectores adicionales según lo necesario (véase a continuación).

El plásmido de expresión de levadura de TALE-AvrBs3::ColE7 final, pCLS8589 (SEQ ID NO: 233), se preparó por clonación *in vivo* en levadura usando el plásmido pCLS9940 (SEQ ID NO: 232). Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de plásmido (pCLS9940) linealizado por digestión con BssHII y 1 ng del ADN del plásmido pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar la levadura *S. cerevisiae* cepa FYC2-6A (*MAT α* , *trp1 Δ 63*, *leu2 Δ 1*, *his3 Δ 200*) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

La construcción de TALE-AvrBs3::ColE7 se ensayó en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224, tabla 7). Los niveles de actividad de TALE-AvrBs3:ColE7 en las respectivas dianas en células de levadura se muestran en la figura 12.

Actividad de TALE::ColE7 en plantas

La secuencia de ADN que codifica las RVD para abordar los sitios NptIIIT5-L y NptIIIT6-L (de la SEQ ID NO: 276 a 279) se subclonaron en el plásmido pCLS15785 (SEQ ID NO: 285, un mutante K497A de ColE7 modificado en el extremo N del plásmido pCLS9939, SEQ ID NO: 231) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para las secuencias RVD insertadas para crear las construcciones TALE::ColE7_A497 posteriores cT11NptIIIT5-L_ColE7_A497 (pCLS15786, SEQ ID NO: 286) y cT11NptIIIT6-L_ColE7_A497 (pCLS15787, SEQ ID NO: 287), respectivamente. Las construcciones se secuenciaron y los insertos de TALE::ColE7_A497 se transfirieron por técnicas de clonación convencionales al plásmido pCLS14529 (SEQ ID NO: 282) para generar los plásmidos de expresión de TALE-NptIIIT5-L::ColE7_A497 y TALE-NptIIIT6-L::ColE7_A497 finales, pCLS14584 (SEQ ID NO: 288, que codifica la proteína de la SEQ ID NO: 437) y pCLS14587 (SEQ ID NO: 289, que codifica la proteína de la SEQ ID NO: 438), respectivamente. El plásmido pCLS14529 permite clonar secuencias del gen de interés después de un promotor que confiere altos niveles de expresión constitutiva en células vegetales.

Para ensayar la actividad en células vegetales, se empleó un ensayo de hibridación monocatenaria (SSA) basado en YFP. El gen indicador YFP tiene una corta duplicación de la secuencia codificante que está interrumpida por un sitio diana de TALEN NptIIIT5 o NptIIIT6. La escisión del sitio diana estimula la recombinación entre las repeticiones, provocando la reconstitución de un gen YFP funcional. Para cuantificar la escisión, el indicador se introduce junto con una construcción que codifica una TALEN o TALEN compacta basada en FokI en protoplastos de tabaco por transformación mediada por PEG. Se obtuvieron eficacias de transformación uniformes usando la misma cantidad de plásmido en cada transformación, es decir, 15 μ g de cada uno de los plásmidos que codifican YFP y la TALEN o cTALEN. Después de 24 horas, los protoplastos se sometieron a citometría de flujo para cuantificar el número de células positivas a YFP. Los niveles de actividad de TALE::ColE7_A497, usando las cTALEN de acuerdo con la presente invención, en plantas fueron comparables a los de las construcciones de control de TALEN basada en FokI en las dianas ensayadas (tabla 10).

Tabla 10: Actividad de TALE-NptII5-L::ColE7_A497 y TALE-NptII6-L::ColE7_A497 en dianas de ADN apropiadas. La actividad relativa se clasifica respecto a las construcciones de control como: n.a., no aplicable; +, 100 % de actividad del control (8 % de células positivas a YFP).

Construcción de TALEN				
ADN diana	TALE-NptII5-L:: ColE7_A497	NptII5.1 de control	TALE-NptII6-L:: ColE7_A497	NptII6.1 de control
NptII5.1	+	+	n.a.	n.a.
NptII6.1	n.a.	n.a.	+	+

5 Genomanipulación de la TALE::ColE7

Se eligen variantes que difieren por truncamientos del dominio del extremo C de la TALEN derivado de AvrBs3 (SEQ ID NO: 196) como armazones de partida. Un subconjunto de estas variantes incluye el truncamiento después de las posiciones E886 (CO), P897 (C11), G914 (C28), L926 (C40), D950 (C64), R1000 (C115), D1059 (C172) (los dominios proteínicos de los dominios en el extremo C truncados C11 a C172 se dan respectivamente en la SEQ ID NO: 204 a 209) y P1117 [también mencionado como Cter wt o WT Cter (SEQ ID NO: 210) que carece del dominio de activación del dominio del extremo C de AvrBs3 natural (SEQ ID NO: 220)]. Los plásmidos que codifican los armazones variantes que contienen el dominio del extremo N derivado de AvrBs3, el conjunto derivado de AvrBs3 de dominios de repetición y el dominio en el extremo C derivado de AvrBs3 truncado [pCLS7821 pCLS7803 pCLS7807, pCLS7809, pCLS7811, pCLS7813, pCLS7817 (SEQ ID NO: 211 a 217) que se basan en pCLS7184 (SEQ ID NO: 196)] permiten la clonación de cualquier dominio catalítico en fusión con el dominio del extremo C, usando los sitios de restricción BamHI y EagI.

El ADN correspondiente al dominio catalítico de ColE7 se amplifica por PCR para introducir, a nivel de ADN, un sitio de restricción de BamHI (en el 5' de la hebra codificante) y uno EagI (en el 3' de la hebra codificante) y, a nivel de proteína, un conector (por ejemplo tramo -SGGSGS-, SEQ ID NO: 219) entre el dominio del extremo C del TALE y el dominio catalítico ColE7. Adicionalmente, pueden generarse variantes del dominio de endonucleasa ColE7 que modula la actividad catalítica que tiene cambios (individualmente o combinados) en las siguientes posiciones: K446, R447, D493, R496, K497, H545, N560 y H573 [las posiciones se refieren a la secuencia de aminoácidos de la proteína ColE7 completa (SEQ ID NO: 11)]. Las construcciones TALE::ColE7 finales se generan por inserción del dominio catalítico ColE7 en las variantes de armazón usando BamHI y EagI y procedimientos de biología molecular convencionales.

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones de TALE-AvrBs3::ColE7 se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224, tabla 7).

40 Ejemplo 3e: TALEN compacta TALE::Crel

La meganucleasa I-Crel de tipo silvestre (SEQ ID NO: 106) se eligió como molde proteínico para derivar un dominio catalítico específico de secuencia que cuando se fusiona a un armazón derivado de TALE (compuesto de un dominio en el extremo N, un núcleo central compuesto de RVD y un dominio en el extremo C) generará una nueva clase de cTALEN (TALE::Crel). Para distinguir la orientación (extremo N frente a extremo C) de las fusiones del dominio catalítico (CD), los nombres de las construcciones se escriben como CD::TALE-RVD (el dominio catalítico está fusionado en el extremo N al dominio de TALE) o TALE-RVD::CD (el dominio catalítico está fusionado en el extremo C al dominio de TALE), donde "-RVD" indica opcionalmente la secuencia reconocida por el dominio de TALE y "CD" es el tipo de dominio catalítico. En este documento, se describen construcciones basadas en TALE::Crel novedosas que abordan, por ejemplo, la secuencia del gen del receptor B de linfocitos T (gen TCRB, SEQ ID NO: 290, figura 13), tanto mediante el dominio de unión a ADN de TALE como el dominio de I-Crel rediseñado. De forma notable, La especificidad de la TALEN compacta TALE::Crel está dirigida tanto por el dominio de unión a ADN de TALE, así como el dominio catalítico derivado de I-Crel. En un contexto de TALEN compacta, dichas proteínas pueden proporcionar, con una proteína monomérica de tamaño razonable, la alta especificidad necesaria demandada por las aplicaciones terapéuticas.

Actividad de TALE::Crel en levadura

60 Se seleccionó un armazón de TALE central, sT2 (SEQ ID NO: 135) en que (a) podían insertarse diferentes conjuntos

de dominios de RVD para cambiar la especificidad de unión a ADN, y; (b) podía adherirse una selección de dominios catalíticos derivados de I-Crel, en el extremo N o C, para lograr la escisión de ADN (o mellado). Como se menciona previamente, el armazón truncado sT2 se generó por PCR a partir de un molde de armazón de TALEN central de longitud completa (pCLS7183, SEQ ID NO: 141) usando los cebadores CMP_G061 (SEQ ID NO: 142) y CMP_G065 (SEQ ID NO: 143) y se clonó en el vector pCLS7865 (SEQ ID NO: 144) para generar pCLS7865-cTAL11_CFS1 (pCLS9009, SEQ ID NO: 145), donde CFS1 indica la secuencia de aminoácidos -GSSG- (con sitios de restricción subyacentes BamHI y Kpn2I en el ADN codificante para facilitar la clonación). Un dominio catalítico I-Crel rediseñado, diseñado para abordar una secuencia en el gen del receptor B de linfocitos T (gen TCRB, SEQ ID NO: 290, figura 13) se subclonó en dos etapas. En primer lugar, el armazón I-Crel_NFS1 (SEQ ID NO: 122), donde NFS1 (SEQ ID NO: 98) comprende un conector de 20 aminoácidos -GSDITKSKISEKMKGGQPSG- (con sitios de restricción subyacentes BamHI y Kpn2I en el ADN codificante para facilitar la clonación), se fusionó al armazón pCLS7865-cTAL11_CFS1 (usando los sitios de restricción BamHI y EagI) para insertar en el mismo marco de la secuencia codificante el conector NFS1. La meganucleasa I-Crel se reemplazó posteriormente por la construcción de meganucleasa TCRB02-A genommanipulada (pCLS6857, SEQ ID NO: 291) usando los sitios de restricción Kpn2I y XhoI, produciendo pCLS7865-cT11_scTB2aD01 (pCLS15788, SEQ ID NO: 292, que codifica la proteína de la SEQ ID NO: 439). También se subclonaron dos variantes de mutación puntual de la meganucleasa TCRB02-A, TCRB02-A_148C (pCLS12083, SEQ ID NO: 293, que codifica la proteína de la SEQ ID NO: 442) y TCRB02-A_333C (pCLS12195, SEQ ID NO: 294, que codifica la proteína de la SEQ ID NO: 443), como dominios catalíticos fusionados al núcleo de unión de TALE, produciendo las construcciones pCLS7865-cT11_scTB2aD01_148C (pCLS15789, SEQ ID NO: 295, que codifica la proteína de la SEQ ID NO: 440) y pCLS7865-cT11_scTB2aD01_333C (pCLS15790, SEQ ID NO: 296, que codifica la proteína de la SEQ ID NO: 441).

Se diseñaron tres secuencias de ADN que codifican RVD que abordan el gen TCRB a diferentes distancias desde el sitio de meganucleasa, que dan lugar a las RVD TCRB02A1 (SEQ ID NO: 297), TCRB02A2 (SEQ ID NO: 298) y TCRB02A3 (SEQ ID NO: 299), que abordan secuencias ubicadas a 7 pb, 12 pb y 16 pb, respectivamente, delante del sitio TCRB de meganucleasa (figura 13). Las secuencias de ADN para cada RVD se subclonaron independientemente en el plásmido pCLS15788 (SEQ ID NO: 292) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia RVD insertada para crear las construcciones de TALE::scTB2aD01 posteriores cT11TB2A1_scTB2aD01 (pCLS15791, SEQ ID NO: 300), cT11TB2A2_scTB2aD01 (pCLS15792, SEQ ID NO: 301) y cT11TB2A3_scTB2aD01 (pCLS15793, SEQ ID NO: 302). Adicionalmente, las RVD TCRB02A2 (SEQ ID NO: 298) se clonaron de forma similar en pCLS15789 (SEQ ID NO: 295) para crear cT11TB2A2_scTB2aD01_148C (pCLS15794, SEQ ID NO: 303). Todas las construcciones se secuenciaron y los diversos insertos se transfirieron a vectores adicionales según lo necesario (véase a continuación).

Los plásmidos de expresión de levadura de TALE::scTB2aD01 finales, pCLS13449 (SEQ ID NO: 304, que codifica la proteína de la SEQ ID NO: 444), pCLS13450 (SEQ ID NO: 305, que codifica la proteína de la SEQ ID NO: 445), pCLS13451 (SEQ ID NO: 306, que codifica la proteína de la SEQ ID NO: 446) y pCLS15148 (SEQ ID NO: 307, que codifica la proteína de la SEQ ID NO: 455), se prepararon por clonación *in vivo* en levadura usando los plásmidos pCLS15791 (SEQ ID NO: 300), pCLS15792 (SEQ ID NO: 301), pCLS15793 (SEQ ID NO: 302) y pCLS15794 (SEQ ID NO: 303), respectivamente. Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de cada plásmido linealizado por digestión con BssHII y 1 ng del ADN del plásmido pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar, respectivamente, la levadura *S. cerevisiae* cepa FYC2-6A (MATa, *trp1Δ63*, *leu2Δ1*, *his3Δ200*) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN o meganucleasa se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones basadas en TALE::scTB2aD01 se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas híbridas TCRB02Tsp7 (SEQ ID NO: AC4), TCRB02Tsp12 (SEQ ID NO: AC5) y TCRB02Tsp16 (SEQ ID NO: AC6), ilustradas en la figura 14. La diana únicamente TCRB02.1 se incluyó para comparar la actividad con la meganucleasa TCRB02-A genommanipulada (pCLS6857, SEQ ID NO: 291), que no requiere los sitios de unión a ADN de TALE para la actividad. Los niveles de actividad en las dianas respectivas en células de levadura para las construcciones basadas en TALE::scTB2aD01 indicadas se muestran en la figura 14. De forma notable, en las condiciones *in vivo* ensayadas, la construcción TALE-TB2A2::scTB2aD01_148C (pCLS15794, SEQ ID NO: 303) ya no escindía dianas que carecen de la secuencia de ADN reconocida por el resto de unión a ADN de TALE.

Actividad de TALE::Crel en células de mamífero

se subclonó el DNA que codifica las construcciones TALE-TB2A2::scTB2aD01 y TALE-TB2A3::scTB2aD01 a partir de pCLS15792 (SEQ ID NO: 301) y pCLS15793 (SEQ ID NO: 302) en el plásmido de expresión de mamífero pCLS1853 (SEQ ID NO: 193) usando las enzimas de restricción AscI y XhoI para el plásmido receptor y las enzimas de restricción BssHII y XhoI para los insertos basados en TALE::scTB2aD01, que da lugar a los plásmidos de expresión de mamífero

pCLS14894 y pCLS14895 (SEQ ID NO: 308 y 309), respectivamente.

Todos plásmidos indicadores diana de mamífero que contenían las secuencias diana de ADN de TALEN se construyeron usando el protocolo de Gateway convencional (INVITROGEN) en un vector indicador CHO (Arnould, Chames *et al.* 2006, Grizot, Epinat *et al.* 2010).

Para controlar los niveles de expresión de proteína, las construcciones basadas en TALE::scTB2aD01 se transfectaron en células de mamífero (HEK293) junto a la meganucleasa TCRB02-A genomanipulada (pCLS6857, SEQ ID NO: 291). En resumen, las células se transfectaron, respectivamente, con 300 ng de cada plásmido codificante de proteína en presencia de lipofectamina. Cuarenta y ocho horas después de la transfección, se analizaron 20 µg de extracto de proteína total para cada muestra por transferencia de Western usando un anticuerpo policlonal anti-I-Crel. Se muestra una transferencia de Western típica en la figura 15.

Se evaluó la toxicidad relativa de las construcciones basadas en TALE::scTB2aD01 usando un ensayo de supervivencia celular. Se usaron células CHOK1 para sembrar placas a una densidad de $2,5 \times 10^3$ células por pocillo. El siguiente día, se usaron cantidades variables de plásmido que codifica las construcciones basadas en TALE::scTB2aD01 (pCLS14894 y pCLS14895; SEQ ID NO: 308 y 309) o la meganucleasa TCRB02-A genomanipulada (pCLS6857, SEQ ID NO: 291) y una cantidad constante de plásmido que codifica GFP (10 ng) para transfectar las células con una cantidad total de 200 ng usando reactivo Polyfect. Los niveles de GFP se controlaron por citometría de flujo (Guava EasyCyte, Guava technologies) en los días 1 y 6 después de la transfección. La supervivencia celular se expresa como un porcentaje, calculado como una relación (células transfectadas con TALEN y meganucleasa que expresan GFP en día 6/células transfectadas con control que expresan GFP en el día 6) corregida para eficacia de transfección determinada en el día 1. Los datos del ensayo de supervivencia celular típicos se muestran en la figura 16.

La actividad de escisión *in vivo* se controló mediante detección de eventos de NHEJ en presencia de exonucleasa TREX2. Se usó el plásmido (3 µg) que codifica las construcciones basadas en TALE::scTB2aD01 (pCLS14894 y pCLS14895; SEQ ID NO: 308 y 309) o la meganucleasa TCRB02-A genomanipulada (pCLS6857, SEQ ID NO: 291) y 2 µg de plásmido que codifica scTrex2 (pCLS8982, SEQ ID NO: 310) para transfectar las células HEK293 en presencia de lipofectamina. Se extrajo el ADN genómico 2 y 7 días después de la transfección con el kit de sangre y tejido DNeasy (Qiagen) y se amplificó la región que abarca el sitio TCRB02 (figura 13) usando PCR con los oligos TRBC2F3 (SEQ ID NO: 311) y TRBC2R3B (SEQ ID NO: 312) en el día 2 después de la transfección y con los oligos TRBC2F4 (SEQ ID NO: 315) y TRBC2R4B (SEQ ID NO: 314) en el día 7 después de la transfección. Los productos de PCR respectivos (100 ng) se desnaturalizaron por calor, se permitió que rehidraran por enfriamiento lento, después se trataron con endonucleasa 1 de T7 (NEB) durante 15 minutos a 37 °C. Los productos de PCR digeridos se separan en geles de acrilamida al 10 % y se visualizan con tinción de SYBRgreen (Invitrogen). La escisión de secuencias de ADN emparejadas incorrectamente por la endonucleasa de T7 es indicativa de eventos de NHEJ resultantes de la actividad de la cTALEN o meganucleasa en el locus diana. La figura 17 ilustra la actividad de NHEJ detectable de las construcciones basadas en TALE::scTB2aD01 (pCLS14894 y pCLS14895; SEQ ID NO: 308 y 309) en comparación con la meganucleasa TCRB02-A genomanipulada (pCLS6857, SEQ ID NO: 291). Mientras en que en el día 2 los resultados de NHEJ son comparables para todas las construcciones, la actividad de NHEJ en el día 7 pueden detectarse únicamente para las construcciones basadas en TALE::scTB2aD01, lo que sugiere que estas TALEN compactas no inducen citotoxicidad.

45 Genomanipulación de la TALE::Crel

Una propiedad novedosa significativa de la TALEN compacta TALE::Crel reside en la capacidad de manipular independientemente la especificidad "híbrida" de la molécula final. Por tanto, la relación de actividad inherente/especificidad puede modularse dentro de las construcciones derivadas de TALE::Crel, lo que permite una dirección específica sin precedentes con retención de una alta actividad de escisión de ADN. En su forma más simple, la redirección satisfactoria del dominio de unión de ADN de TALE se consigue mediante el RVD cifrado (figura 3) con una seudocorrespondencia uno a uno a la base de ADN subyacente. La manipulación del resto I-Crel, sin embargo, presente más problemas en la medida en que existe una codependencia potencial de contactos de proteína-ADN necesarios para la unión a ADN adecuada y actividad de escisión. Se han descrito métodos (documentos WO 2006/097854, WO 2008/093249, WO 03078619, WO 2009/095793, WO 2007/049095, WO 2007/057781, WO 2006/097784, WO 2006/097853, WO 2007/060495, WO 2007/049156 y WO 2004/067736) para rediseñar satisfactoriamente la meganucleasa I-Crel para abordar secuencias de ADN novedosas. Como algunos de estos métodos dependen de una estrategia agrupada, puede preverse que usando dicha estrategia pueda reducirse la especificidad "absoluta" del resto I-Crel en una manera por etapas. Por ejemplo, la descomposición de la superficie de interacción de ADN con I-Crel en las regiones concretas 10NNN, 7NN, 5NNN y 2NN (por semisubunidad monomérica) permite una manipulación novedosa en la que se mantiene una alta especificidad en la región 5NNN-2NN central a expensas de "pérdida" o ampliación de especificidad en las regiones 10NNN-7NN externas. En esencia, dicha estrategia podría reducir la complejidad del rediseño del armazón derivado de I-Crel para un contexto de TALEN compacta ya que únicamente la "selectividad" en la escisión es necesaria para el dominio catalítico, con posterior especificidad proporcionada por la parte de unión a ADN de TALE de la fusión proteínica. Tomado conjuntamente, la facilidad de manipulación combinada con la alta especificidad potencial y la alta actividad de escisión de ADN hacen

que las TALEN compactas derivadas de TALE::Crel sean herramientas ideales para aplicaciones terapéuticas. Finalmente, debe apreciarse que el resto I-Crel podría en principio remplazarse con una gran cantidad de dominios catalíticos derivados de endonucleasa de asentamiento que existen de forma natural existentes o rediseñados.

5 Ejemplo 3f: Actividad de TALE::SnaseSTAAU

Actividad de TALE::SnaseSTAAU en levadura

10 Se eligen variantes que difieren por truncamientos del dominio del extremo C de la TALEN derivado de AvrBs3 (SEQ ID NO: 196) como armazones de partida. Un subconjunto de estas variante incluye el truncamiento después de las posiciones G914 (C28) y L926 (C40) (los dominios proteínicos de los dominios del extremo C truncados C28 y C40 se dan respectivamente en la SEQ ID NO: 205 y 206). Los plásmidos que codifican los armazones variantes que contienen el dominio del extremo N derivado de AvrBs3, el conjunto derivado de AvrBs3 de dominios de repetición y el dominio en el extremo C derivado de AvrBs3 truncado [pCLS7807 y pCLS7809, (SEQ ID NO: 213 y 214) que se basan en pCLS7184 (SEQ ID NO: 196)] permite la clonación de cualquier domino catalítico en fusión con el dominio del extremo C, usando los sitios de restricción BamHI y EagI.

20 El ADN correspondiente a los restos de aminoácido 83 a 231 de SnaseSTAAU (SEQ ID NO: 30) se amplifica por PCR para introducir, a nivel de ADN, un sitio de restricción de BamHI (en el 5' de la hebra codificante) y uno EagI (en el 3' de la hebra codificante) y, a nivel de proteína, un conector (por ejemplo tramo -SGGSGS-, SEQ ID NO: 219) entre el dominio del extremo C del TALE y el dominio catalítico SnaseSTAAU. Las construcción TALE::SnaseSTAAU finales se generan por inserción del dominio catalítico SnaseSTAAU en las variantes de armazón usando BamHI y EagI y procedimientos de biológica molecular convencionales. Las variantes de armazón truncadas después de las posiciones G914 (C28) y L926 (C40), respectivamente, codificadas por pCLS7807 y pCLS7809, (SEQ ID NO: 213 y 214), se fusionaron al dominio catalítico SnaseSTAAU (SEQ ID NO: 30), que da lugar a pCLS9082 y pCLS9081 (SEQ ID NO: 370 y 371). La etapa de clonación también se produce a nivel de aminoácidos en una secuencia AAD en Cter del dominio catalítico SnaseSTAAU.

30 Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

35 Las construcciones TALE-AvrBs3::SnaseSTAAU se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI, que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones TALE-AvrBs3::SnaseSTAAU se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224). Los datos resumidos en la figura 19 muestran que las construcciones TALE-AvrBs3::SnaseSTAAU son activas en dianas que tienen dos sitio de reconocimiento AvrBs3, de acuerdo con la proteína quimérica de la presente invención, pero también en dianas que contienen únicamente un sitio de reconocimiento AvrBs3.

45 Ejemplo 4

50 Las cTALEN básicas están compuestas de un único dominio de unión a ADN fusionado a un único dominio catalítico y se diseñan para estimular la HR mediante una escisión de ADN bicatenario única o evento de mellado monocatenario. Para determinadas aplicaciones (por ejemplo, inactivación génica), es favorable potenciar el nivel de NHEJ. Este ejemplo ilustra la creación de una cTALEN de escisión doble (dcTALEN) que puede lograr la escisión de ADN bicatenario en dos sitios distintos que flanquean el dominio de unión a ADN de TALE (figura 5C). Se espera que la escisión simultánea del ADN en los dos sitios elimine la secuencia intermedia y, por lo tanto, anule el religamiento "sin marcas" por NHEJ (figura 1).

Los armazones basales (de la SEQ ID NO: 136 a la SEQ ID NO: 139) descritos en el ejemplo 3 se usan como puntos de partida para los diseños de fusión. Se presenta una lista no exhaustiva de dominios catalíticos susceptibles a fusión con dominios de unión a ADN de TALE en la tabla 2. Se presenta una lista no exhaustiva de conectores que pueden usarse en la tabla 3. Véanse los ejemplos 3, 5, 6 y 7 para detalles adicionales respecto a la elección del conector o dominio de potenciación. Para los diseños de dcTALEN, se fusiona al menos un dominio cleavasa (en el extremo N o C) al dominio de unión a ADN de TALE. El dominio catalítico adicional puede ser un dominio nickasa o cleavasa (endonucleasa o exonucleasa) y depende de la naturaleza de la aplicación. Por ejemplo, el acoplamiento de un dominio cleavasa en un lado con un dominio nickasa en el otro podría provocar la escisión de una hebra simple de ADN que abarca la región de unión a ADN de TALE. La generación dirigida de salientes monocatenarios prolongados podría aplicarse en aplicaciones que abordan mecanismos de reparación de ADN. Para inactivación génica dirigida, se prefiere el uso de dos dominios cleavasa en la dcTALEN.

Todos los diseños de dcTALEN se evalúan usando nuestro ensayo de levadura (véase el ejemplo 1) y proporcionan actividad detectable comparable con las meganucleasas genomanipuladas existentes. Además, las posibles potenciaciones de NHEJ se controlan usando el ensayo basado en células de mamífero como se describe en el ejemplo 3.

Ejemplo 4a: Actividad de las TALEN de escisión doble Tevl::TALE::FokI y Tevl::TALE::TevI

Se generaron TALEN de escisión doble (CD::TALE::CD), que poseen un dominio catalítico derivado de I-TevI en el extremo N y un dominio catalítico en el extremo C derivado de FokI (SEQ ID NO: 368) o I-TevI (SEQ ID NO: 20) en el armazón bT2-Avr basal (SEQ ID NO: 137). El fragmento de dominio catalítico de I-TevI se escindió del plásmido pCLS12731 (SEQ ID NO: 236) y se subclonó en los vectores pCLS15795 (SEQ ID NO: 351) y pCLS9013 (SEQ ID NO: 153) por restricción y ligamiento usando los sitios de restricción NcoI y NsiI, produciendo TevD02_cT11Avr_FokI-L (pCLS15796, SEQ ID NO: 352, que codifica la proteína de la SEQ ID NO: 447) y TevD02_cT11Avr_TevD02 (pCLS15797, SEQ ID NO: 353, que codifica la proteína de la SEQ ID NO: 448), respectivamente. Todas las construcciones se secuenciaron y el inserto se transfirió a vectores adicionales según lo necesario (véase a continuación).

Las plásmidos de expresión en levadura de Tevl::TALE-AvrBs3::FokI y Tevl::TALE-AvrBs3::TevI finales, pCLS13299 (SEQ ID NO: 354, que codifica la proteína de la SEQ ID NO: 449) y pCLS13301 (SEQ ID NO: 355, que codifica la proteína de la SEQ ID NO: 450), se prepararon por clonación *in vivo* en levadura usando los plásmidos CLS15796 (SEQ ID NO: 352) y pCLS15797 (SEQ ID NO: 353), respectivamente. Para generar una secuencia codificante intacta por recombinación homóloga *in vivo*, se usaron aproximadamente 40 ng de cada plásmido linealizado por digestión con BssHII y 1 ng del ADN del plásmido pCLS0542 (SEQ ID NO: 156) linealizado por digestión con NcoI y EagI para transformar, respectivamente, la levadura *S. cerevisiae* cepa FYC2-6A (*MAT α* , *trp1 Δ 63*, *leu2 Δ 1*, *his3 Δ 200*) usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones de Tevl::TALE-AvrBs3::FokI y Tevl::TALE-AvrBs3::TevI se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento de AvrBs3 o RagT2-R (SEQ ID NO: 238, tabla 11). En dianas adecuadas, los niveles de actividad de Tevl::TALE-AvrBs3::FokI y Tevl::TALE-AvrBs3::TevI en levadura eran comparables con los de sus moléculas precursoras que carecen del dominio catalítico derivado de I-TevI en el extremo N. Se ilustra la actividad significativa en la tabla 11 para una diana de sitio único de muestra, de acuerdo con la dcTALEN de la presente invención.

Tabla 11: Actividad de diversas cTALEN y dcTALEN en dianas de ADN de sitio doble e individual. La actividad relativa se clasifica como: n.d., sin actividad detectable; +, <25 % de actividad; ++, de 25 % a <50 % de actividad; +++, de 50 % a <75 % de actividad; +++, de 75 % a 100 % de actividad.

Construcción de TALEN	ADN diana	
	Avr25 (sitio doble) (SEQ ID NO: 177)	Avr25RAGT2R (sitio único) (SEQ ID NO: 238)
TevI::TALE-AvrBs3	++++	+++
TALE-AvrBs3::FokI	++++	n.d.
TALE-AvrBs3::TevI	++++	n.d.
TevI::TALE-AvrBs3::FokI	++++	+++
TevI::TALE-AvrBs3::TevI	++++	+++

Ejemplo 4b: TALEN de escisión doble scTrex2::TALE::FokI

Una TALEN de escisión doble (CD::TALE::CD), que posee un dominio catalítico scTrex2 en el extremo N y un dominio catalítico en el extremo C derivado de FokI, se generó sobre el armazón bT2-Avr basal (SEQ ID NO: 137). El fragmento de dominio catalítico de scTrex2 se escindió del plásmido pCLS15798 (SEQ ID NO: 356, que codifica la proteína de la SEQ ID NO: 451) y se subclonó en el vector pCLS15795 (SEQ ID NO: 351) por restricción y ligamiento usando los sitios de restricción NcoI y NsiI, produciendo scTrex2_cT11Avr_FokI-L (pCLS15799, SEQ ID NO: 357, que codifica la proteína de la SEQ ID NO: 452). La construcción se secuenció y el inserto se transfirió a vectores adicionales según lo necesario (véase a continuación).

El ADN que codifica las construcciones de TALE-AvrBs3::FokI o scTrex2::TALE-AvrBs3::FokI a partir de pCLS15795 (SEQ ID NO: 351) o pCLS15799 (SEQ ID NO: 357), respectivamente, se subclonó en el plásmido de expresión de mamífero pCLS1853 (SEQ ID NO: 193) usando las enzimas de restricción AscI y XhoI para el plásmido receptor y las enzimas de restricción BssHII y XhoI para los insertos, que da lugar a los plásmidos de expresión de mamífero pCLS14972 y pCLS14971 (SEQ ID NO: 358 y 359), respectivamente.

Todos plásmidos indicadores diana de mamífero que contenían las secuencias diana de ADN de TALEN se construyeron usando el protocolo de Gateway convencional (INVITROGEN) en un vector indicador CHO (Arnould, Chames *et al.* 2006, Grizot, Epinat *et al.* 2010). Las construcciones de TALE-AvrBs3::FokI y scTrex2::TALE-AvrBs3::FokI se ensayaron en un ensayo extracromosómico en células de mamífero (CHO K1) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI, que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 6).

Para este ensayo, se transfectaron células CHO K1 en un formato de placa de 96 pocillos con 75 ng de vector diana y una cantidad creciente de cada ADN variante de 0,7 a 25 ng, en presencia de reactivo PolyFect (1 µl por pocillo). La cantidad total de ADN transfectado se completó a 125 ng (ADN diana, ADN variante, ADN transportador) usando un vector vacío. Setenta y dos horas después de la transfección, se retiró el medio de cultivo y se añadieron 150 µl de tampón de lisis/revelación para ensayo líquido de β-galactosidasa. Después de incubación a 37 °C, se midió la densidad óptica a 420 nm. El proceso completo se realiza en una plataforma automatizada Velocity11 BioCel (Grizot, Epinat *et al.* 2009).

Los niveles de actividad en células de mamífero dianas adecuadas para la construcción scTrex2::TALE-AvrBs3::FokI fueron comparable con los de la molécula TALE-AvrBs3::FokI precursora, que indica que el resto scTrex2 adicional no altera la función de escisión de ADN de TALEN. La evaluación de la función de scTrex2 se realiza en ensayos adecuados para la detección de eventos de NHEJ.

Ejemplo 5

Los diseños basales para los armazones de cTALEN se basan en los dominios de unión a ADN de TALE establecidos. Las TALEN compactas se diseñan para que sean lo más pequeñas y eficaces posibles. Para obtener este objetivo, por lo tanto, puede ser necesario recurrir a dominios "potenciadores" para unir el hueco funcional entre el dominio de unión a ADN de TALE compacta y los diversos dominios catalíticos. La figura 6 (A-E) ilustra diversas configuraciones no exhaustivas en las que pueden aplicarse dichos dominios potenciadores. Obsérvese que la figura es únicamente ilustrativa, y están implicadas variaciones del extremo N frente al C (es decir, la figura 6A también puede tener un dominio potenciador en el extremo N y un dominio catalítico en el extremo C). Las tablas 1 y 2 enumeran posibles dominios potenciadores que podrían ayudar en la unión al ADN (contactos específicos y no específicos).

Las TALEN potenciadas (eTALEN) se crean usando cTALEN funcionales del ejemplo 3. La adición del dominio potenciador se evalúa en nuestro ensayo de levadura (véase el ejemplo 1). Un dominio potenciador particular se considera útil si proporciona una potenciación mínima de un 5 % en la eficacia de la cTALEN de partida, más preferiblemente una potenciación mínima de un 10 %, más preferiblemente un 20 %, más preferiblemente un 30 %, más preferiblemente un 40 %, más preferiblemente un 50 %, de nuevo más preferiblemente una potenciación mayor de un 50 %.

Ejemplo 5a: TALEN potenciadas TALE::ColE7::TALE

Las TALEN potenciadas (TALE::CD::TALE), que poseen los dominios de unión a ADN de TALE en el extremo N y C limitando un dominio de escisión de ADN central, se generaron usando el armazón central sT2 (SEQ ID NO: 135). La disposición de esta clase de TALEN compacta se ilustra en la figura 6B, en la que el propio "dominio potenciador" en el extremo N es un dominio de unión a ADN de TALE. Un derivado de mutación puntual del dominio catalítico ColE7 (pCLS15785, SEQ ID NO: 285) se eligió para el núcleo catalítico de la eTALEN. Dos construcciones finales, TALE-AvrBs3::ColE7_A497::TALE-RagT2-R (pCLS15800, SEQ ID NO: 360, que codifica la proteína de la SEQ ID NO: 453) y TALE-RagT2-R::ColE7_A497::TALE-AvrBs3 (pCLS15801, SEQ ID NO: 361, que codifica la proteína de la SEQ ID

NO: 454) se obtuvieron usando técnicas de clonación molecular convencionales con secuencias de ADN de sT2 (SEQ ID NO: 135), pCLS15785 (SEQ ID NO: 285), AvrBs3 (SEQ ID NO: 152) y RagT2-R (SEQ ID NO: 271) como moldes. Todas las construcciones TALE::CD::TALE se secuenciaron y los insertos se transfirieron a vectores adicionales según lo necesario (véase a continuación).

5 Los plásmidos de expresión de levadura basados en TALE::CD::TALE finales, pCLS12106 (SEQ ID NO: 362) y pCLS12110 (SEQ ID NO: 363), se prepararon por restricción y ligamiento usando los sitios de restricción NcoI y EagI para subclonar en el plásmido pCLS0542 (SEQ ID NO: 156). La levadura *S. cerevisiae* cepa FYC2-6A (*MAT α* , *trp1 Δ 63*, *leu2 Δ 1*, *his3 Δ 200*) se transformó usando un protocolo de transformación con LiAc de alta eficacia (Arnould *et al.* 2007).

10 Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

15 Las construcciones de TALE::CD::TALE se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas híbridas AvrBs3/RagT2-R asimétricas para comparar la actividad con una TALEN compacta precursora (por ejemplo, pCLS8589, SEQ ID NO: 233), que tiene actividad en dianas con un único sitio de unión. Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento de AvrBs3 o RagT2-R.

Ejemplo 6

25 Hasta la fecha, todos efectores TAL conocidos y derivados de los mismos parecen requerir una base T en las posiciones -1 (figura 3) en la secuencia de reconocimiento. Para superar esta limitación, se usa un dominio potenciador para reemplazar la región del extremo N de la proteína TALE. El modelado de homología basado en secuencia y estructura de la región TALE del extremo N de derivados bT2 ha producido tres proteínas candidatas potenciales (tabla 1): (i) factor de unión a Fem-3, SEQ ID NO: 4, (FBF1, familia Puf de proteínas de unión a ARN) de *C. elegans*; (ii) proteínas de repetición alfa helicoides artificiales (α Rep), SEQ ID NO: 5 y; (iii) proteínas de la superfamilia de anquirina.

30 El contenido y disposición de los elementos estructurales secundarios permite usar estos modelos como puntos de partida para dominios potenciadores que reemplazan la región del extremo N de la proteína TALE.

Las proteínas quiméricas se construyen usando las regiones análogas de uno de los 3 candidatos mencionados para reemplazar la región de la proteína TALE del extremo N hasta el primer dominio de repetición canónica. La nueva superficie de contacto se rediseña *in silico*, usando los modelos de homología como guías. Esta estrategia puede usarse para determinar con precisión los determinantes de especificidad para la T necesaria en la posición -1 de la secuencia diana. El dominio potenciador de reemplazo debe proporcionar como mínimo integridad estructural a la proteína cTALEN. Las construcciones se evalúan en nuestro ensayo de levadura (véase el ejemplo 1). Un dominio potenciador particular se considera útil si proporciona una retención mínima de un 5 % en la actividad de la cTALEN de partida en ausencia de una T en la posición diana -1, más preferiblemente una retención mínima de un 10 %, más preferiblemente un 20 %, más preferiblemente un 30 %, más preferiblemente un 40 %, más preferiblemente un 50 %, de nuevo más preferiblemente una retención en la actividad mayor de un 50 %.

Ejemplo 7

45 Para generar armazones más adecuados y compactos para cTALEN, se ha analizado la naturaleza de la región del extremo C (más allá del semidominio de repetición final) de la proteína TALE. El modelado de homología basado en secuencia y estructura de la región TALE del extremo C de derivados bT2 ha producido tres proteínas candidatas potenciales (tabla 1): (i) la hidrolasa/transferasa de *Pseudomonas aeuriginosa*, SEQ ID NO: 6; (ii) el dominio polimerasa de la ligasa D de *Mycobacterium tuberculosis*, la SEQ ID NO: 7; (iii) el factor de inicio eIF2 de *Pyrococcus*, la SEQ ID NO: 8; (iv) factor de inicio de la traducción Aif2betagamma, la SEQ ID NO: 9. Como en el ejemplo 6, se usan modelos de homología para determinar con precisión regiones para generar posibles truncamientos en el extremo C; las posibles posiciones de truncamiento incluyen 28, 40, 64, 118, 136, 169, 190 restos que quedan más allá del último semidominio de repetición. Adicionalmente, pueden usarse regiones homólogas de las proteínas mencionadas anteriormente para reemplazar el dominio del extremo C completamente. Pueden usarse programas de predicción de contacto para identificar, partiendo de la secuencia primaria de una proteína, los pares de restos que están probablemente proximales en el espacio 3D. Dichas proteínas quiméricas deben proporcionar armazones más estables en que construir las cTALEN.

60 Las construcciones se evalúan en nuestro ensayo de levadura (véase el ejemplo 1). Un dominio potenciador particular se considera útil si proporciona una retención mínima de un 5 % en la actividad de la cTALEN de partida, más preferiblemente una retención mínima de un 10 %, más preferiblemente un 20 %, más preferiblemente un 30 %, más preferiblemente un 40 %, más preferiblemente un 50 %, de nuevo más preferiblemente una retención en la actividad mayor de un 50 %.

65 Ejemplo 8

Para generar TALEN compactas con actividades alternativas, se generan cTALEN trans mediante (a) el uso de un dominio catalítico con actividades separables (figura 7A, B) o; (b) provisión de una actividad auxiliar como una fusión de TALE (figura 7C). Modelado basado en secuencia y estructura de la clase III (Chan, Stoddard *et al.* 2011). Se usaron endonucleasas de restricción de tipo IIS (REasas) para crear TALEN trans (véase la tabla 2 para una lista no exhaustiva). La TALEN trans inicial se genera mediante fusión de un dominio catalítico independientemente activo (por ejemplo, la nickasa Nt.BspD6I como se describe en los ejemplos 3 y 4. En principio, esta TALEN trans puede usarse dependiendo de la aplicación. Para convertir la cTALEN en una TALEN trans funcional, el dominio auxiliar (en este caso, ss.BspD6I) se proporciona *en trans* (figura 8A). Dichas proteínas opcionalmente en trans y/o heterodiméricas pueden permitir que los armazones de cTALEN con actividad puedan modularse para una aplicación dada.

Las construcciones se evalúan en nuestro ensayo de levadura (véase el ejemplo 1). Un dominio auxiliar particular se considera útil si proporciona una actividad alternativa a la de la cTALEN de partida.

Si el dominio auxiliar usado muestra actividad independiente de la cTALEN inicial (es decir, en un contexto de TALEN que no es trans), puede fusionarse también a un dominio TALE para dirección específica (figura 7B). Los dominios auxiliares también puede proporcionarse *en trans* como entidades dirigidas para proporcionar funciones no relaciones con la cTALEN (figura 7C).

Ejemplo 8a: Inhibición específica de actividad catalítica de TALEN

Como se menciona en los ejemplos 3c y 3d, tanto NucA (SEQ ID NO: 26) como CoIE7 (SEQ ID NO: 140) pueden inhibirse por la formación de complejos con sus proteínas inhibidoras respectivas, NuiA (SEQ ID NO: 229) y Im7 (SEQ ID NO: 230). Colicina-E9 (SEQ ID NO: 366) es otro ejemplo no limitante de proteína que puede inhibirse por su inhibidor respectivo Im9 (SEQ ID NO: 369). Con las respecto a las TALEN derivadas de los dominios catalíticos NucA (TALE::NucA) o CoIE7 (TALE::CoIE7), los inhibidores sirven como dominios auxiliares (figura 7A) que modulan la actividad evitando la escisión de ADN.

Las proteínas inhibidoras Im7 (SEQ ID NO: 230) y NuiA (SEQ ID NO: 229) se subclonaron en la estructura pCLS7763 (SEQ ID NO: 241) por restricción y ligamiento usando los sitios de restricción NcoI y EagI, produciendo pCLS9922 (SEQ ID NO: 242) y pCLS9923 (SEQ ID NO: 243), respectivamente. Estos plásmidos entonces se usaron en experimentos de cotransformación en el ensayo de SSA de levadura convencional como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Las construcciones TALE-AvrBs3::NucA (pCLS9924, SEQ ID NO: 223) y TALE-AvrBs3::CoIE7 (pCLS8589, SEQ ID NO: 233) se ensayaron en un ensayo de SSA de levadura en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI (pCLS8590, SEQ ID NO: 244), que requiere dos sitios de unión para la actividad. Las dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que varía de 5 a 40 pb (de la SEQ ID NO: 157 a 192, tabla 7). Además, las construcciones se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224, tabla 7). La modulación de la actividad de las TALEN se evaluó en presencia o ausencia de proteína inhibidora específica o inespecífica, usando la TALEN TALE-AvrBs3::FokI como control.

Los datos resumidos en la tabla 12 indican que las construcciones de TALE-AvrBs3::NucA y TALE-AvrBs3::CoIE7 se inactivan específicamente por la presencia de sus proteínas inhibidoras respectivas NuiA y Im7, de acuerdo con la presente invención.

Tabla 12: Actividad de construcciones de TALEN en presencia de proteína inhibidora. La actividad relativa se clasifica como: n.d., sin actividad detectable; +, <25 % de actividad; ++, de 25 % a <50 % de actividad; +++, de 50 % a <75 % de actividad; +++++, de 75 % a 100 % de actividad.

Construcción de TALEN	Proteína inhibidora		
	Ninguna	NuiA	Im7
TALE-AvrBs3::NucA (SEQ ID NO: 223)	++++	n.d.	++++
TALE-AvrBs3::CoIE7 (SEQ ID NO: 233)	++++	++++	n.d.
TALE-AvrBs3::FokI (SEQ ID NO: 244)	++++	++++	++++

Ejemplo 8b: Potenciación de la actividad catalítica de TALEN mediante una TALEN trans

El ejemplo 3b ilustra que la TALE::TevI funciona sin ayuda como una TALEN compacta (pCLS8522, SEQ ID NO: 237). Para potenciar adicionalmente la actividad, se diseñó una TALEN trans usando una construcción de TALE::TevI en

una disposición representada en la figura 7C. La secuencia de ADN que codifica las RVD para abordar el sitio RagT2-R (SEQ ID NO: 271) se subclonó en el plásmido pCLS7865-cT11_TevD02 (pCLS9011, SEQ ID NO: 151) usando las enzimas de restricción de tipo IIS BsmBI para el plásmido receptor y BbvI y SfaNI para la secuencia de RVD insertada para crear la construcción de TALE-RagT2-R::TevI posterior cT11RagT2- R_TevD02 (pCLS15802, SEQ ID NO: 364).
 5 La construcción se secuenció y el inserto se subclonó en la estructura pCLS7763 (SEQ ID NO: 241) por restricción y ligamiento usando los sitios de restricción NcoI y EagI, produciendo pCLS8990 (SEQ ID NO: 365). Los pares de plásmidos pCLS8522 (SEQ ID NO: 237) y pCLS7763 (SEQ ID NO: 241) o pCLS8522 (SEQ ID NO: 237) y pCLS8990 (SEQ ID NO: 365) entonces se usaron en experimentos de cotransformación en el ensayo de SSA de levadura convencional como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould
 10 *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006).

Todos los plásmidos indicadores diana de levadura que contenían las secuencias diana de ADN de TALEN se construyeron como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould
 15 *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006). Los pares de construcciones TALE-RagT2-R::TevI/TevI::TALE-AvrBs3 se ensayaron en un ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas híbridas RagT2-R/AvrBs3 asimétricas para comparar la actividad con una TALEN compacta precursora (por ejemplo, pCLS8522, SEQ ID NO: 237), que tiene actividad en dianas con un único sitio de unión. Las dianas híbridas RagT2-R/AvrBs3 contienen dos
 20 secuencias de reconocimiento diferentes yuxtapuestas con el extremo 3' del primero (RagT2-R) proximal al extremo 5' del segundo (AvrBs3) y separadas por ADN "espaciador" que varía de 5 a 40 pb (SEQ ID NO: G064 a G099, tabla 13). La figura 18 ilustra la modulación en la actividad de TevI::TALE-AvrBs3 proporcionada por la construcción TALE-RagT2-R::TevI, de acuerdo con la cTALEN trans de la presente invención.

25 Ejemplo 9: Remplazo del dominio del extremo C por un conector polipeptídico, actividad con dominio catalítico CoIE7.

Se generó una primera colección de 37 conectores diferentes. Muchos de ellos tienen una estructura común que comprende una región variable que codifica de 3 a 28 restos de aminoácido y flanqueada por las regiones que codifican el tramo SGGSGS (SEQ ID NO: 219) tanto en el extremo 5' como en el 3' (SEQ ID NO: 372 a 408). Estos conectores
 30 contienen sitios de restricción XmaI y BamHI en sus extremos 5' y 3', respectivamente. La colección de conectores entonces se subclona en pCLS7183 (SEQ ID NO: 141) mediante los sitios de restricción XmaI y BamHI para remplazar el dominio del extremo C de la TALEN derivada de AvrBs3 (pCLS7184, SEQ ID NO: 196). El conjunto derivado de AvrBs3 de dominios de repetición (RVD) o cualquier otra secuencia RVD que tenga o carezca del semi-RVD terminal se clona en esta colección de estructuras. El ADN de la colección se obtiene, después de raspado de las colonias de
 35 las placas de Petri, usando técnicas de minipreparación convencionales. La cabeza catalítica FokI se elimina usando las enzimas de restricción BamHI y EagI, purificándose la estructura restante usando técnicas de extracción de gel convencionales.

El ADN que codifica el dominio catalítico CoIE7 (SEQ ID NO: 11) se amplificó por PCR para introducir, a nivel de ADN,
 40 un sitio de restricción de BamHI (en el 5' de la hebra codificante) y uno EagI (en el 3' de la hebra codificante) y, a nivel de proteína, un conector (por ejemplo tramo -SGGSGS-, SEQ ID NO: 219) entre la colección de dominios de extremo C y la cabeza catalítica. Después de digestión con BamHI y EagI y la purificación, el ADN que codifica las diferentes cabezas catalíticas se subclonó individualmente en el armazón de la colección previamente preparado.

El ADN de la colección final se obtiene, después de raspado de las colonias de placas de Petri, usando técnicas de
 45 minipreparación convencionales y las colecciones resultantes se criban en nuestro ensayo de SSA de levadura como se describe previamente (solicitudes PCT internacionales WO 2004/067736 y en Epinat, Arnould *et al.* 2003; Chames, Epinat *et al.* 2005; Arnould, Chames *et al.* 2006; Smith, Grizot *et al.* 2006) en dianas pseudopalindrómicas para comparar la actividad con una TALEN convencional TALE-AvrBs3::FokI, que requiere dos sitios de unión para la actividad. Las
 50 dianas AvrBs3 contienen dos secuencias de reconocimiento idénticas yuxtapuestas con los extremos 3' proximales y separadas por ADN "espaciador" que contiene 15, 18, 21 y 24 pb (SEQ ID NO: 167, 170, 173 y 176, tabla 7). Además, las construcciones (SEQ ID NO: 416-419) se ensayaron en una diana que tiene únicamente, un único sitio de reconocimiento AvrBs3 (SEQ ID NO: 224). Los datos resumidos en la figura 20 muestran secuencias del conector de una fracción de construcciones de CoIE7 que son activas en dianas que tienen dos sitios de reconocimiento AvrBs3 o
 55 únicamente un sitio de reconocimiento AvrBs3.

Lista de referencias citadas

60 Arimondo, P. B., C. J. Thomas, *et al.* (2006). "Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates". *Mol Cell Biol* 26(1): 324-33.
 Arnould, S., P. Chames, *et al.* (2006). Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *Journal of Molecular Biology*. 355:443-58.
 Arnould, S., P. Chames, *et al.* (2006). "Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets". *J Mol Biol* 355(3): 443-58.
 65 Arnould, S., C. Delenda, *et al.* (2011). "The I-Crel meganuclease and its engineered derivatives: applications from cell modification to gene therapy

- gzq083 [pii] 10.1093/protein/gzq083". Protein engineering, design & selection: PEDS 24(1-2): 27-31.
- Arnould, S., C. Perez, *et al.* (2007). "Engineered I-Crel derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells". Journal of Molecular Biology 371(1): 49-65.
- 5 Ashworth, J., J. J. Havranek, *et al.* (2006). "Computational redesign of endonuclease DNA binding and cleavage specificity". Nature 441(7093): 656-9.
- Bedayat, B., A. Abdolmohamadi, *et al.* (2010). "Sequence-specific correction of genomic hypoxanthine-guanine phosphoribosyl transferase mutations in lymphoblasts by small fragment homologous replacement 10.1089/oli.2009.0205". Oligonucleotides 20(1): 7-16.
- 10 Bennardo, N., A. Cheng, *et al.* (2008). "Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair". PLoS Genet 4(6): e1000110.
- Bennardo, N., A. Gunn, *et al.* (2009). "Limiting the persistence of a chromosome break diminishes its mutagenic potential". PLoS Genet 5(10): e1000683.
- Boch, J., H. Scholze, *et al.* (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors". Science 326(5959): 1509-12.
- 15 Boch, J., H. Scholze, *et al.* (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors 1178811 [pii] 10.1126/science.1178811". Science 326(5959): 1509-12.
- Bolduc, J. M., P. C. Spiegel, *et al.* (2003). "Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor". Genes Dev 17(23): 2875-88.
- 20 Buis, J., Y. Wu, *et al.* (2008). "Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation". Cell 135(1): 85-96.
- Capecchi, M. R. (2001). "Generating mice with targeted mutations 10.1038/nm1001-1086 nm1001-1086 [pii]". Nature Medicine 7(10): 1086-90.
- Carroll, D. (2008). "Progress and prospects: zinc-finger nucleases as gene therapy agents gt2008145 [pii] 10.1038/gt.2008.145". Gene therapy 15(22): 1463-8.
- 25 Chames, P., J. C. Epinat, *et al.* (2005). "*In vivo* selection of engineered homing endonucleases using double-strand break induced homologous recombination". Nucleic Acids Res 33(20): e178.
- Chames, P., J. C. Epinat, *et al.* (2005). "*In vivo* selection of engineered homing endonucleases using double-strand break induced homologous recombination". Nucleic Acids Research 33(20): e178.
- 30 Chan, S. H., B. L. Stoddard, *et al.* (2011). "Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity". Nucleic Acids Research 39: 1-18.
- Chevalier, B., M. Turmel, *et al.* (2003). "Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-Crel and I-Msol". J Mol Biol 329(2): 253-69.
- Chevalier, B. S., T. Kortemme, *et al.* (2002). "Design, activity, and structure of a highly specific artificial endonuclease". Mol Cell 10(4): 895-905.
- 35 Chevalier, B. S., R. J. Monnat, Jr., *et al.* (2001). "The homing endonuclease I-Crel uses three metals, one of which is shared between the two active sites". Nat Struct Biol 8(4): 312-6.
- Chevalier, B. S. y B. L. Stoddard (2001). "Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility". Nucleic Acids Res 29(18): 3757-74.
- 40 Choo, Y. y A. Klug (1994). "Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions". Proceedings of the National Academy of Sciences of the United States of America 91(23): 11168-72.
- Choo, Y. y A. Klug (1994). "Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage". Proceedings of the National Academy of Sciences of the United States of America 91(23): 11163-7.
- 45 Choulika, A., A. Perrin, *et al.* (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*". Mol Cell Biol 15(4): 1968-73.
- Christian, M., T. Cermak, *et al.* (2010). "Targeting DNA double-strand breaks with TAL effector nucleases". Genetics 186(2): 757-61.
- 50 Christian, M., T. Cermak, *et al.* (2010). "Targeting DNA double-strand breaks with TAL effector nucleases genetics.110.120717 [pii] 10.1534/genetics.110.120717". Genetics 186(2): 757-61.
- Cost, G. J., Y. Freyvert, *et al.* (2010). "BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells 10.1002/bit.22541". Biotechnology and Bioengineering 105(2): 330-40.
- 55 Delacote, F. y B. S. Lopez (2008). "Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model 5149 [pii]". Cell Cycle 7(1): 33-8.
- Doudeva, L. G., H. Huang, *et al.* (2006). "Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn²⁺ or inhibitor/Ni²⁺ 15/2/269 [pii] 10.1110/ps.051903406". Protein science: a publication of the Protein Society 15(2): 269-80.
- 60 Doyon, J. B., V. Pattanayak, *et al.* (2006). "Directed evolution and substrate specificity profile of homing endonuclease I-SceI". Journal of the American Chemical Society 128(7): 2477-84.
- Doyon, J. B., V. Pattanayak, *et al.* (2006). "Directed evolution and substrate specificity profile of homing endonuclease I-SceI". J Am Chem Soc 128(7): 2477-84.
- 65 Doyon, Y., J. M. McCammon, *et al.* (2008). "Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases nbt1409 [pii] 10.1038/nbt1409". Nature Biotechnology 26(6): 702-8.

- Eastberg, J. H., J. Eklund, *et al.* (2007). "Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism". *Biochemistry* 46(24): 7215-25.
- Eisenschmidt, K., T. Lanio, *et al.* (2005). "Developing a programmed restriction endonuclease for highly specific DNA cleavage". *Nucleic Acids Res* 33(22): 7039-47.
- 5 Elrod-Erickson, M., M. A. Rould, *et al.* (1996). "Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions". *Structure* 4(10): 1171-80.
- Epinat, J. C., S. Arnould, *et al.* (2003). "A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells". *Nucleic Acids Research* 31(11): 2952-62.
- 10 Epinat, J. C., S. Arnould, *et al.* (2003). "A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells". *Nucleic Acids Res* 31(11): 2952-62.
- Frank, K. M., J. M. Sekiguchi, *et al.* (1998). "Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV
10.1038/24172". *Nature* 396(6707): 173-7.
- 15 Galetto, R., P. Duchateau, *et al.* (2009). "Targeted approaches for gene therapy and the emergence of engineered meganucleases
10.1517/14712590903213669". *Expert opinion on biological therapy* 9(10): 1289-303.
- Gao, H., J. Smith, *et al.* (2010). "Heritable targeted mutagenesis in maize using a designed endonuclease TPJ4041 [pii] 10.1111/j.1365-313X.2009.04041.x". *The Plant journal: for cell and molecular biology* 61(1): 176-87. Gao, Y., Y Sun, *et al.* (1998). "A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis S0092-8674(00)81714-6 [pii]". *Cell* 95(7): 891-902.
- 20 Geurts, A. M., G. J. Cost, *et al.* (2009). "Knockout rats via embryo microinjection of zinc-finger nucleases 325/5939/433 [pii] 10.1126/science.1172447". *Science* 325(5939): 433.
- Gimble, F. S., C. M. Moure, *et al.* (2003). "Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system". *J Mol Biol* 334(5): 993-1008.
- 25 Greisman, H. A. y C. O. Pabo (1997). "A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites". *Science* 275(5300): 657-61.
- Gruenert, D. C., E. Bruscia, *et al.* (2003). "Sequence-specific modification of genomic DNA by small DNA fragments 10.1172/JCI19773 112/5/637 [pii]". *The Journal of clinical investigation* 112(5): 637-41.
- 30 Guirouilh-Barbat, J., S. Huck, *et al.* (2004). "Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells". *Mol Cell* 14(5): 611-23.
- Guirouilh-Barbat, J., S. Huck, *et al.* (2004). "Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells
10.1016/j.molcel.2004.05.008 S1097276504002916 [pii]". *Molecular Cell* 14(5): 611-23.
- 35 Guirouilh-Barbat, J., E. Rass, *et al.* (2007). "Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends". *Proc Natl Acad Sci U S A* 104(52): 20902-7.
- Guirouilh-Barbat, J., E. Rass, *et al.* (2007). "Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends
0708541104 [pii] 10.1073/pnas.0708541104". *Proceedings of the National Academy of Sciences of the United States of America* 104(52): 20902-7.
- 40 Gurlebeck, D., B. Szurek, *et al.* (2005). "Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import
TPJ2370 [pii] 10.1111/j.1365-313X.2005.02370.x". *The Plant journal: for cell and molecular biology* 42(2): 175-87.
- Haber, J. (2000). "Partners and pathways repairing a double-strand break". *Trends Genet.* 16(6): 259-264.
- 45 Haber, J. E. (2008). "Alternative endings
0711334105 [pii] 10.1073/pnas.0711334105". *Proceedings of the National Academy of Sciences of the United States of America* 105(2): 405-6.
- Hartsuiker, E., K. Mizuno, *et al.* (2009). "Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions". *Mol Cell Biol* 29(7): 1671-81.
- 50 Hinnen, A., J. B. Hicks, *et al.* (1978). "Transformation of yeast". *Proceedings of the National Academy of Sciences of the United States of America* 75(4): 1929-33.
- Hirata, R., J. Chamberlain, *et al.* (2002). "Targeted transgene insertion into human chromosomes by adeno-associated virus vectors
10.1038/nbt0702-735 nbt0702-735 [pii]". *Nature Biotechnology* 20(7): 735-8.
- 55 Huang, H. y H. S. Yuan (2007). "The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases". *Journal of Molecular Biology* 368(3): 812-21.
- Ichiyanagi, K., Y. Ishino, *et al.* (2000). "Crystal structure of an archaeal intein-encoded homing endonuclease PIPful". *J Mol Biol* 300(4): 889-901.
- 60 Inoue, N., R. Dong, *et al.* (2001). "Introduction of single base substitutions at homologous chromosomal sequences by adeno-associated virus vectors
10.1006/mthe.2001.0283 S1525-0016(01)90283-7 [pii]". *Molecular therapy: the journal of the American Society of Gene Therapy* 3(4): 526-30.
- Isalan, M. y Y. Choo (2001). "Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins
S0076-6879(01)40444-7 [pii]". *Methods in Enzymology* 340: 593-609.
- 65 Kalish, J. M. y P. M. Glazer (2005). "Targeted genome modification via triple helix formation". *Ann N Y Acad Sci*

- 1058: 151-61.
- Kim, H. J., H. J. Lee, *et al.* (2009). "Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly gr.089417.108 [pii] 10.1101/gr.089417.108". *Genome Research* 19(7): 1279-88.
- 5 Kim, Y. G., J. Cha, *et al.* (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain". *Proceedings of the National Academy of Sciences of the United States of America* 93(3): 1156-60.
- Ku, W. Y., Y. W. Liu, *et al.* (2002). "The zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis". *Nucleic Acids Research* 30(7): 1670-8.
- 10 Landthaler, M., U. Begley, *et al.* (2002). "Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of *Staphylococcus aureus* phage Twort". *Nucleic Acids Research* 30(9): 1935-43.
- Landthaler, M., N. C. Lau, *et al.* (2004). "Group I intron homing in *Bacillus* phages SPO1 and SP82: a gene conversion event initiated by a nicking homing endonuclease". *Journal of Bacteriology* 186(13): 4307-14.
- Landthaler, M., B. W. Shen, *et al.* (2006). "I-BasI and I-Hmul: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities". *Journal of Molecular Biology* 358(4): 1137-51.
- 15 Landthaler, M. y D. A. Shub (2003). "The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille". *Nucleic Acids Research* 31(12): 3071-7.
- Lee, S. E., F. Paques, *et al.* (1999). "Role of yeast SIR genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths". *Curr Biol* 9(14): 767-70.
- Li, T., S. Huang, *et al.* (2010). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain". *Nucleic Acids Res* 39(1): 359-72.
- 20 Li, T., S. Huang, *et al.* (2011). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain gkq704 [pii] 10.1093/nar/gkq704". *Nucleic Acids Research* 39(1): 359-72.
- Liang, F., M. Han, *et al.* (1998). "Homology-directed repair is a major double-strand break repair pathway in mammalian cells". *Proceedings of the National Academy of Sciences of the United States of America* 95(9): 5172-7.
- 25 Liu, P. Q., E. M. Chan, *et al.* (2010). "Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases 10.1002/bit.22654". *Biotechnology and Bioengineering* 106(1): 97-105.
- Liu, Q., J. T. Dansereau, *et al.* (2008). "Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI". *J Mol Biol* 379(5): 1094-106.
- Liu, Q., V. Derbyshire, *et al.* (2006). "Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions". *Nucleic Acids Research* 34(6): 1755-64.
- 30 Lloyd, A., C. L. Plaisier, *et al.* (2005). "Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis* 0409339102 [pii] 10.1073/pnas.0409339102". *Proceedings of the National Academy of Sciences of the United States of America* 102(6): 2232-7.
- Maeder, M. L., S. Thibodeau-Beganny, *et al.* (2008). "Rapid".open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification S1097-2765(08)00461-9 [pii] 10.1016/j.molcel.2008.06.016". *Molecular Cell* 31(2): 294-301.
- 35 Mahfouz, M. M., L. Li, *et al.* (2011). "De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks 1019533108 [pii] 10.1073/pnas.1019533108". *Proceedings of the National Academy of Sciences of the United States of America* 108(6): 2623-8.
- 40 Marcaida, M. J., I. G. Munoz, *et al.* (2010). "Homing endonucleases: from basics to therapeutic applications 10.1007/s00018-009-0188-y." *Cellular and molecular life sciences: CMLS* 67(5): 727-48.
- Mashimo, T., A. Takizawa, *et al.* (2010). "Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases 10.1371/journal.pone.0008870". *PloS one* 5(1): e8870.
- 45 McConnell Smith, A., R. Takeuchi, *et al.* (2009). "Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil LAGLIDADG homing endonuclease 0810588106 [pii] 10.1073/pnas.0810588106". *Proceedings of the National Academy of Sciences of the United States of America* 106(13): 5099-104.
- McVey, M. y S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings S0168-9525(08)00229-1 [pii] 10.1016/j.tig.2008.08.007". *Trends in genetics: TIG* 24(11): 529-38.
- 50 Menoret, S., A. L. Iscache, *et al.* (2010). "Characterization of immunoglobulin heavy chain knockout rats 10.1002/eji.201040939". *European Journal of Immunology* 40(10): 2932-41.
- Metzger, M. J., A. McConnell-Smith, *et al.* (2011). "Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template gkq826 [pii] 10.1093/nar/gkq826". *Nucleic Acids Research* 39(3): 926-35.
- 55 Midon, M., P. Schafer, *et al.* (2011). "Mutational and biochemical analysis of the DNA-entry nuclease End A from *Streptococcus pneumoniae* gkq802 [pii] 10.1093/nar/gkq802". *Nucleic Acids Research* 39(2): 623-34.
- Miller, J. C., S. Tan, *et al.* (2011). "A TALE nuclease architecture for efficient genome editing nbt.1755 [pii] 10.1038/nbt.1755". *Nature Biotechnology* 29(2): 143-8.
- 60 Mimitou, E. P. y L. S. Symington (2008). "Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing". *Nature* 455(7214): 770-4.
- Moore, I., M. Samalova, *et al.* (2006). "Transactivated and chemically inducible gene expression in plants". *Plant J* 45(4): 651-83.
- 65 Moore, J. K. y J. E. Haber (1996). "Cell cycle and genetic requirements of two pathways of nonhomologous endjoining repair of double-strand breaks in *Saccharomyces cerevisiae*". *Mol Cell Biol* 16(5): 2164-73.

- Moscou, M. J. y A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors". *Science* 326(5959): 1501.
- Moscou, M. J. y A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors 1178817 [pii] 10.1126/science.1178817". *Science* 326(5959): 1501.
- 5 Moure, C. M., F. S. Gimble, *et al.* (2002). "Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence". *Nat Struct Biol* 9(10): 764-70.
- Moure, C. M., F. S. Gimble, *et al.* (2003). "The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity". *J Mol Biol* 334(4): 685-95.
- 10 Nimonkar, A. V., J. Genschel, *et al.* (2011). "BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair". *Genes Dev* 25(4): 350-62.
- Niu, Y., K. Tenney, *et al.* (2008). "Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity S0022-2836(08)00840-1 [pii] 10.1016/j.jmb.2008.07.010". *Journal of Molecular Biology* 382(1): 188-202.
- Orr-Weaver, T. L., J. W. Szostak, *et al.* (1981). "Yeast transformation: a model system for the study of recombination". *Proceedings of the National Academy of Sciences of the United States of America* 78(10): 6354-8.
- 15 Orr-Weaver, T. L., J. W. Szostak, *et al.* (1983). "Genetic applications of yeast transformation with linear and gapped plasmids". *Methods in Enzymology* 101: 228-45.
- Pabo, C. O., E. Peisach, *et al.* (2001). "Design and selection of novel Cys2His2 zinc finger proteins 70/1/313 [pii] 10.1146/annurev.biochem.70.1.313". *Annual Review of Biochemistry* 70: 313-40.
- 20 Padidam, M. (2003). "Chemically regulated gene expression in plants". *Curr Opin Plant Biol* 6(2): 169-77.
- Paques, F. y P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy". *Current Gene Therapy* 7(1): 49-66.
- Paques, F. y P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy". *Curr Gene Ther* 7(1): 49-66.
- 25 Paques, F. y J. E. Haber (1999). "Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*". *Microbiology and molecular biology reviews: MMBR* 63(2): 349-404.
- Perez, E. E., J. Wang, *et al.* (2008). "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases nbt1410 [pii] 10.1038/nbt1410". *Nature Biotechnology* 26(7): 808-16.
- 30 Pierce, A. J., P. Hu, *et al.* (2001). "Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells". *Genes Dev* 15(24): 3237-42.
- Pingoud, A. y G. H. Silva (2007). "Precision genome surgery". *Nat Biotechnol* 25(7): 743-4.
- Porteus, M. H. y D. Carroll (2005). "Gene targeting using zinc finger nucleases". *Nat Biotechnol* 23(8): 967-73.
- Ramirez, C. L., J. E. Foley, *et al.* (2008). "Unexpected failure rates for modular assembly of engineered zinc fingers nmeth0508-374 [pii] 10.1038/nmeth0508-374". *Nature Methods* 5(5): 374-5.
- 35 Rosen, L. E., H. A. Morrison, *et al.* (2006). "Homing endonuclease I-Crel derivatives with novel DNA target specificities". *Nucleic Acids Research* 34(17): 4791-800.
- Rosen, L. E., H. A. Morrison, *et al.* (2006). "Homing endonuclease I-Crel derivatives with novel DNA target specificities". *Nucleic Acids Res.*
- 40 Rothstein, R. J. (1983). "One-step gene disruption in yeast". *Methods in Enzymology* 101: 202-11.
- Rouet, P., F. Smih, *et al.* (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells". *Proc Natl Acad Sci USA* 91(13): 6064-8.
- Rouet, P., F. Smih, *et al.* (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease". *Mol Cell Biol* 14(12): 8096-106.
- 45 Russell, D. W. y R. K. Hirata (1998). "Human gene targeting by viral vectors 10.1038/ng0498-325". *Nature Genetics* 18(4): 325-30.
- Sangiulolo, F., M. L. Scaldaferrri, *et al.* (2008). "Cfr gene targeting in mouse embryonic stem cells mediated by Small Fragment Homologous Replacement (SFHR) 2904 [pii]". *Frontiers in bioscience: a journal and virtual library* 13: 2989-99.
- 50 Santiago, Y., E. Chan, *et al.* (2008). "Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases 0800940105 [pii] 10.1073/pnas.0800940105". *Proceedings of the National Academy of Sciences of the United States of America* 105(15): 5809-14.
- Sartori, A. A., C. Lukas, *et al.* (2007). "Human CtIP promotes DNA end resection". *Nature* 450(7169): 509-14.
- 55 Seligman, L. M., K. M. Chisholm, *et al.* (2002). "Mutations altering the cleavage specificity of a homing endonuclease". *Nucleic Acids Research* 30(17): 3870-9.
- Seligman, L. M., K. M. Stephens, *et al.* (1997). "Genetic analysis of the *Chlamydomonas reinhardtii* I-Crel mobile intron homing system in *Escherichia coli*". *Genetics* 147(4): 1653-64.
- Shen, B. W., M. Landthaler, *et al.* (2004). "DNA binding and cleavage by the HNH homing endonuclease I-Hmul". *Journal of Molecular Biology* 342(1): 43-56.
- 60 Shukla, V. K., Y. Doyon, *et al.* (2009). "Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases nature07992 [pii] 10.1038/nature07992". *Nature* 459(7245): 437-41.
- Silva, G. H., J. Z. Dalgaard, *et al.* (1999). "Crystal structure of the thermostable archaeal intron-encoded endonuclease I-Dmol". *J Mol Biol* 286(4): 1123-36.
- 65 Simon, P., F. Cannata, *et al.* (2008). "Sequence-specific DNA cleavage mediated by bipyridine polyamide

- conjugates". *Nucleic Acids Res* 36(11): 3531-8.
- Smith, J., J. M. Berg, *et al.* (1999). "A detailed study of the substrate specificity of a chimeric restriction enzyme gkc139 [pii]". *Nucleic Acids Research* 27(2): 674-81.
- 5 Smith, J., M. Bibikova, *et al.* (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains". *Nucleic Acids Research* 28(17): 3361-9.
- Smith, J., S. Grizot, *et al.* (2006). "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences". *Nucleic Acids Research* 34(22): e149.
- Smith, J., S. Grizot, *et al.* (2006). "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences". *Nucleic Acids Res* 34(22): e149.
- 10 Sonoda, E., H. Hohegger, *et al.* (2006). "Differential usage of non-homologous end-joining and homologous recombination in double strand break repair". *DNA Repair (Amst)* 5(9-10): 1021-9.
- Spiegel, P. C., B. Chevalier, *et al.* (2006). "The structure of I-CeuI homing endonuclease: Evolving asymmetric DNA recognition from a symmetric protein scaffold". *Structure* 14(5): 869-80.
- Stoddard, B. L. (2005). "Homing endonuclease structure and function". *Quarterly Reviews of Biophysics* 38(1): 49-95.
- 15 Stoddard, B. L. (2005). "Homing endonuclease structure and function". *Q Rev Biophys* 38(1): 49-95.
- Stoddard, B. L., A. M. Scharenberg, *et al.* (2007). *Advances in Engineering Homing Endonucleases for Gene Targeting: Ten Years After Structures. Progress in Gene Therapy: Autologous and Cancer Stem Cell Gene Therapy.* R. Bertolotti y K. Ozawa, World Scientific Publishing Co. Pte. Ltd. 3: 135-68.
- Sugawara, N. y J. E. Haber (1992). "Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation". *Mol Cell Biol* 12(2): 563-75.
- 20 Sun, H., D. Treco, *et al.* (1991). "Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site". *Cell* 64(6): 1155-61.
- Sussman, D., M. Chadsey, *et al.* (2004). "Isolation and characterization of new homing endonuclease specificities at individual target site positions". *Journal of Molecular Biology* 342(1): 31-41.
- 25 Sussman, D., M. Chadsey, *et al.* (2004). "Isolation and characterization of new homing endonuclease specificities at individual target site positions". *J Mol Biol* 342(1): 31-41.
- Taubes, G. (2002). "Gene therapy. The strange case of chimeraplasty 10.1126/science.298.5601.2116 298/5601/2116 [pii]". *Science* 298(5601): 2116-20.
- Wang, R., X. Zhou, *et al.* (2003). "Chemically regulated expression systems and their applications in transgenic plants". *Transgenic Res* 12(5): 529-40.
- 30 Wang, Y. T., J. D. Wright, *et al.* (2009). "Redesign of high-affinity nonspecific nucleases with altered sequence preference 10.1021/ja907160r". *Journal of the American Chemical Society* 131(47): 17345-53.
- White, C. I. y J. E. Haber (1990). "Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*". *Embo J* 9(3): 663-73.
- 35 Yang, M., V. Djukanovic, *et al.* (2009). "Targeted mutagenesis in the progeny of maize transgenic plants 10.1007/s11103-009-9499-5". *Plant Molecular Biology* 70(6): 669-79.
- Zhao, L., R. P. Bonocora, *et al.* (2007). "The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif". *The EMBO Journal* 26(9): 2432-42.
- 40 Zuo, J. y N. H. Chua (2000). "Chemical-inducible systems for regulated expression of plant genes". *Curr Opin Biotechnol* 11(2): 146-51.

45

REIVINDICACIONES

1. Un monómero de TALEN compacta que comprende:
 - 5 (i) Un almacén de TALE central que comprende regiones de dipéptido variable de repetición (RVD) que tiene especificidad de unión a ADN en una secuencia diana de ADN bicatenario específica de interés;
 - (ii) Al menos un dominio catalítico de I-TevI que puede escindir ADN adyacente a dicha secuencia diana de ADN bicatenario de interés cuando se fusiona al extremo C o N de dicho almacén de TALE central de (i);
 - 10 en el que dicho monómero de TALEN compacta se une a dicha secuencia de ADN diana y escinde ADN bicatenario.
2. Una TALEN compacta de la reivindicación 1, en la que dicho monómero de TALEN compacta escinde ADN bicatenario a una distancia 5' o 3' de 1-25 pares de bases de la secuencia diana de ADN bicatenario específica de interés.
- 15 3. Un monómero de TALEN compacta de la reivindicación 1 o 2, que comprende además un conector peptídico para fusionar dicho dominio catalítico de (ii) con dicho almacén de TALE central de (i).
4. Un monómero de TALEN compacta de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que dicho dominio catalítico se fusiona al dominio del extremo C de dicho almacén de TALE central.
- 20 5. Un monómero de TALEN compacta de acuerdo con cualquiera de las reivindicaciones 1 a 3, en el que dicho dominio catalítico se fusiona al dominio del extremo N de dicho almacén de TALE central.
- 25 6. Un monómero de TALEN compacta de acuerdo con cualquiera de las reivindicaciones 1 a 5, que comprende una secuencia proteínica que tiene al menos un 80 %, más preferiblemente un 90 %, de nuevo más preferiblemente un 95 % de identidad de secuencia de aminoácidos con las secuencias proteínicas seleccionadas del grupo de la SEQ ID NO: 426-432.
- 30 7. Un monómero de TALEN compacta de acuerdo con una cualquiera de las reivindicaciones 3 a 6, en el que dicha secuencia de conector peptídico puede seleccionarse del grupo que consiste en la SEQ ID NO: 67-104 y de la SEQ ID NO: 372 a la SEQ ID NO: 415.
- 35 8. Un polinucleótido recombinante que codifica una TALEN compacta de acuerdo con una cualquiera de las reivindicaciones 1 a 7.
9. Un vector que comprende un polinucleótido recombinante de acuerdo con la reivindicación 8.
- 40 10. Una composición farmacéutica que comprende una TALEN compacta de acuerdo con una cualquiera de las reivindicaciones 1 a 7 y un vehículo farmacéuticamente activo.
11. Una célula hospedadora que comprende un polinucleótido recombinante de la reivindicación 8.
- 45 12. Un animal transgénico no humano que comprende un polinucleótido recombinante de la reivindicación 8.
13. Una planta transgénica que comprende un polinucleótido recombinante de la reivindicación 8.
14. Un kit que comprende un monero de TALEN compacta de acuerdo con una cualquiera de las reivindicaciones 1 a 7 e instrucciones para su uso en la potenciación de la eficacia de procesamiento de ADN de una secuencia diana de ADN bicatenario única de interés.
- 50 15. Un método para abordar y escindir un ADN bicatenario, que comprende:
 - (a) Seleccionar una secuencia diana de ADN de interés en una hebra de un ADN bicatenario;
 - (b) Proporcionar un monómero de TALEN compacta de acuerdo con una cualquiera de las reivindicaciones 1 a 7;
 - 55 (c) Poner en contacto dicho ADN bicatenario con dicho monómero de modo que dicho monómero de TALEN compacta escinda dicho ADN bicatenario.
- 60 16. Un método para insertar *in vitro* o *ex vivo* un transgén en una secuencia diana de ADN bicatenario única específica de un locus genómico de una célula o tejido, en el que se introduce al menos un monómero de TALEN compacta de una cualquiera de las reivindicaciones 1 a 7 en dicha célula o tejido, y en el que el método no es un proceso para modificar la identidad genética de la línea germinal de los seres humanos.
- 65 17. Un método para insertar un transgén en una secuencia diana de ADN bicatenario única específica de un locus genómico de un animal no humano, en el que se introduce al menos un monómero de TALEN compacta de una cualquiera de las reivindicaciones 1 a 7 en dicho animal no humano, en el que el método no es un método de tratamiento del organismo animal.

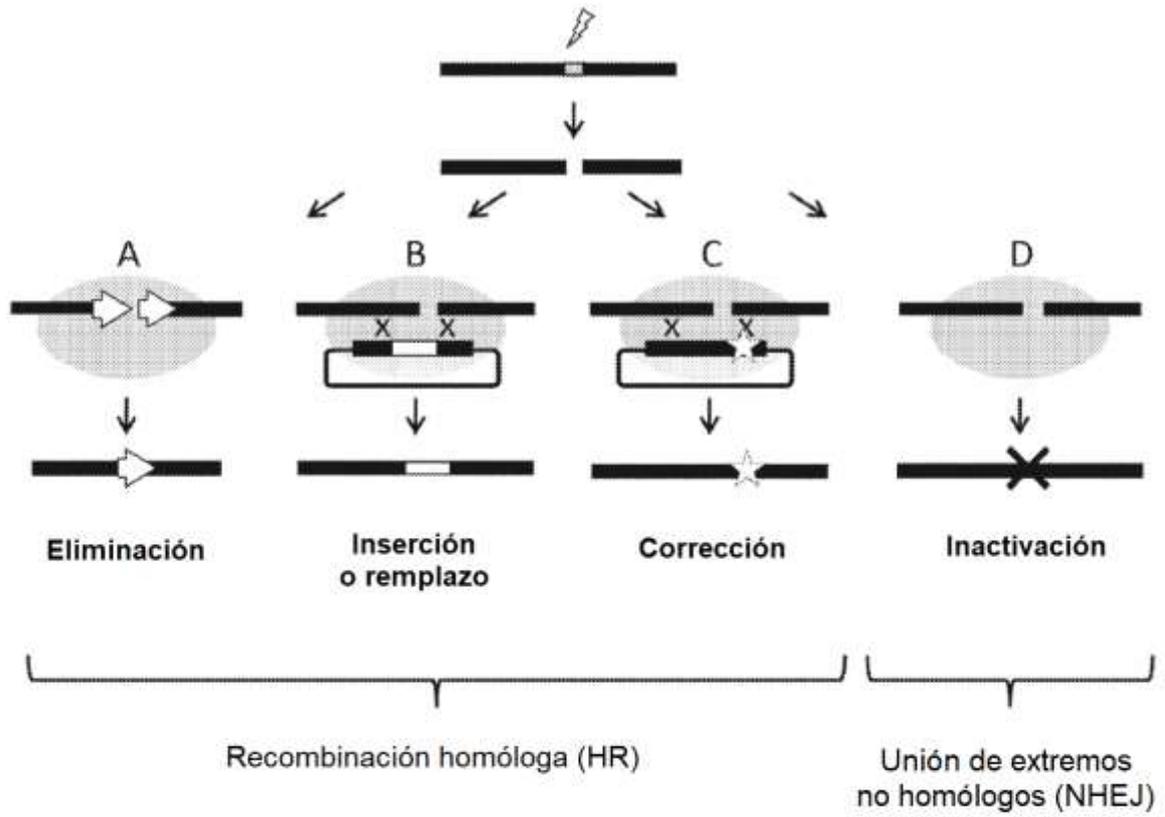
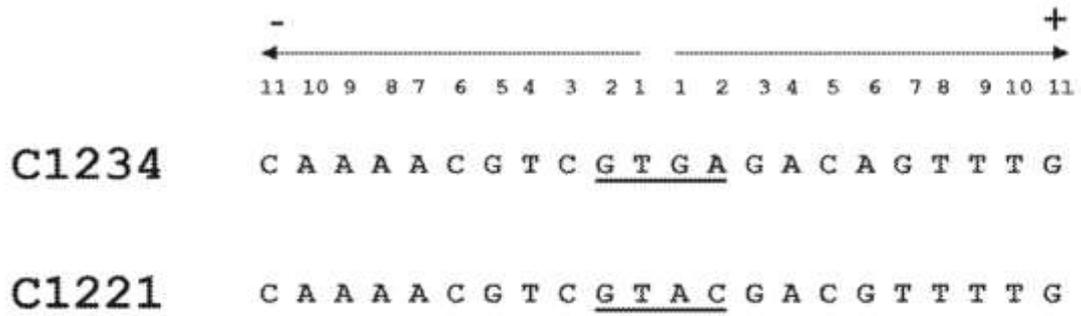


Figura 1

I-CreI:



I-TevI:

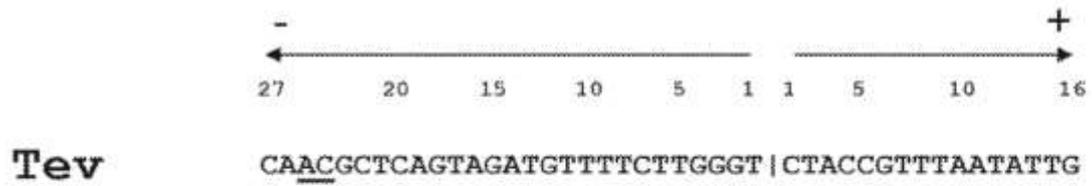
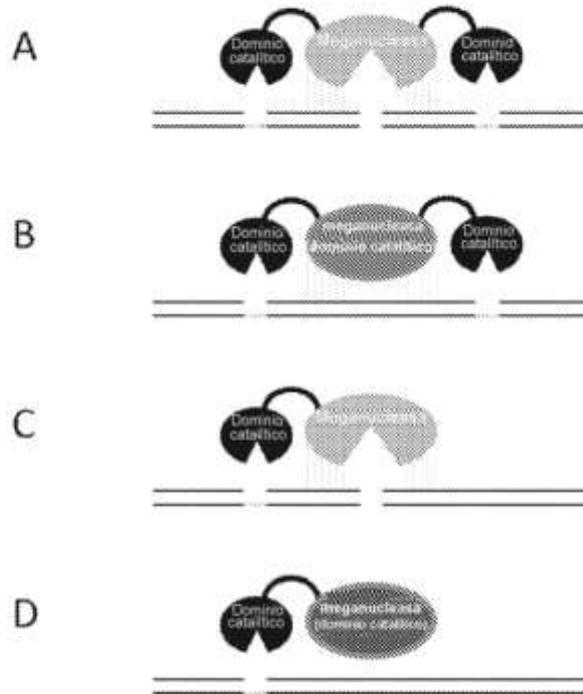


Figura 2

EP 2 694 091 B1



Leyenda

- | | | | |
|---|--|---|---|
|  | - Meganucleasa
(puede ser dimérica o monomérica) |  | - Meganucleasa catalíticamente inactiva;
proporciona especificidad de unión a ADN
(puede ser dimérica o monomérica) |
|  | - Región de contacto con ADN | | |
|  | - Dominio catalítico para escisión de hebra de ADN
(puede ser dominio cleavasa o nickasa) |  | - Región de conector proteínico |
|  | - ADN bicatenario |  | - Hebra individual de ADN opcionalmente escindida
dependiendo del dominio catalítico adherido |

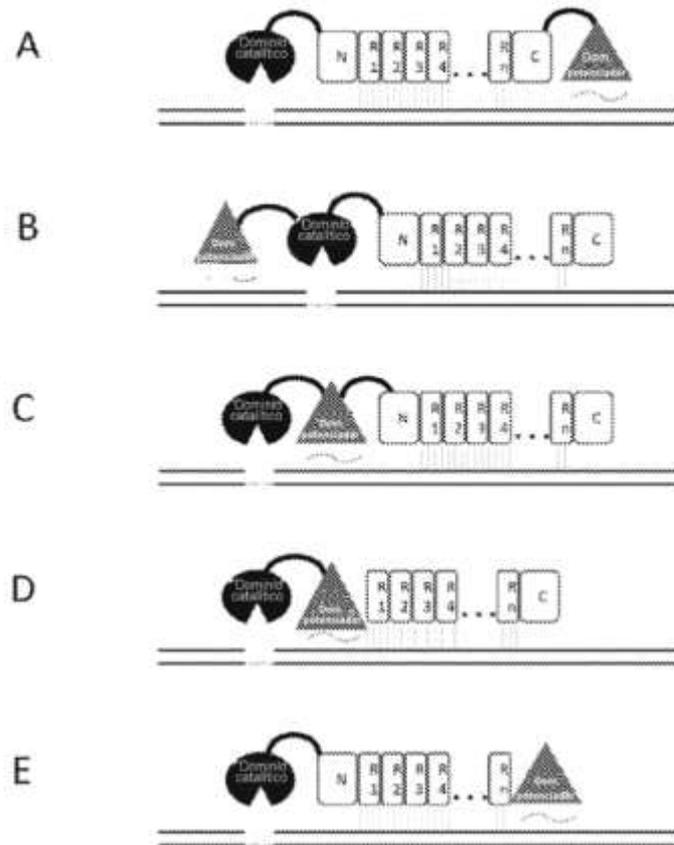
Figura 4



Leyenda

- | | | | |
|---|---|---|---|
|  | - Región del extremo N de la proteína TALE (incluyendo truncamientos de la misma) |  | - Región del extremo C de la proteína TALE (incluyendo truncamientos de la misma) |
|  | - Región de repetición de la proteína TALE (n = repetición final; depende de la construcción) |  | - Regiones de repetición de TALE adicionales (por brevedad, no se muestran) |
|  | - Región de repetición de TALE de contacto con ADN | | |
|  | - Dominio catalítico para escisión de hebra de ADN (puede ser dominio cleavasa o nickasa) |  | - Región de conector proteínico |
|  | - ADN bicatenario |  | - Hebra individual de ADN opcionalmente escindida dependiendo del dominio catalítico adherido |

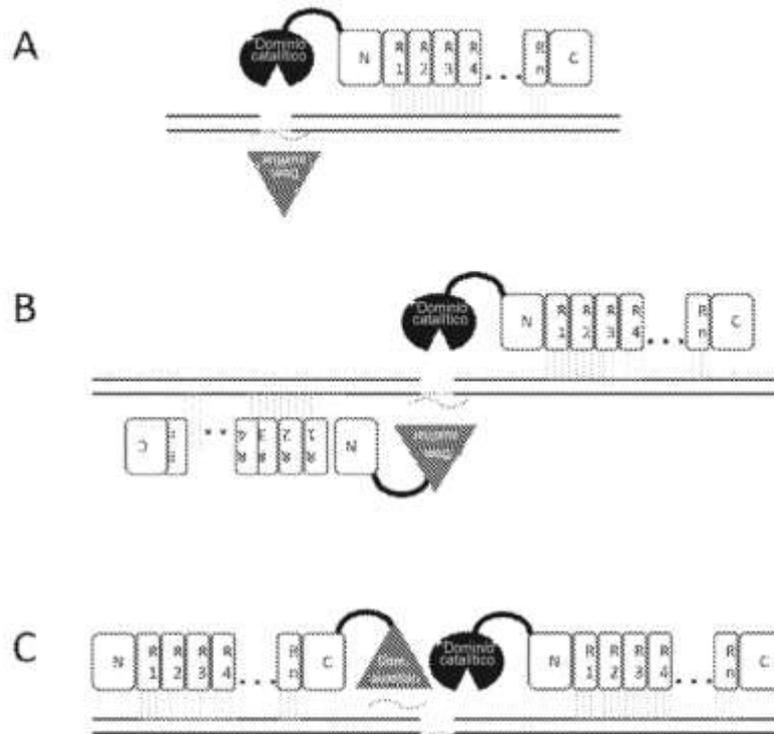
Figura 5



Leyenda

- | | | | |
|---|---|---|---|
|  | - Región del extremo N de la proteína TALE (incluyendo truncamientos de la misma) |  | - Región del extremo C de la proteína TALE (incluyendo truncamientos de la misma) |
|  | - Región de repetición de la proteína TALE (n = repetición final; depende de la construcción) | ... | - Regiones de repetición de TALE adicionales (por brevedad, no se muestran) |
|  | - Región de repetición de TALE de contacto con ADN |  | - Dominio potenciador; proporciona soporte funcional y/o estructural al armazón |
|  | - Dominio catalítico para escisión de hebra de ADN (puede ser dominio cleavasa o nickasa) |  | - Región de conector proteínico |
|  | - ADN bicatenario |  | - Hebra individual de ADN opcionalmente escindida dependiendo del dominio catalítico adherido |

Figura 6



Leyenda

- | | | | |
|---|---|---|---|
|  | - Región del extremo N de la proteína TALE (incluyendo truncamientos de la misma) |  | - Región del extremo C de la proteína TALE (incluyendo truncamientos de la misma) |
|  | - Región de repetición de la proteína TALE (n = repetición final; depende de la construcción) | ... | - Regiones de repetición de TALE adicionales (por brevedad, no se muestran) |
|  | - Región de repetición de TALE de contacto con ADN |  | - Dominio auxiliar; proporciona soporte funcional y/o estructural externo al armazón |
|  | - Dominio catalítico para escisión de hebra de ADN (puede ser dominio cleavasa o nickasa) |  | - Región de conector proteínico |
|  | - ADN bicatenario |  | - Hebra individual de ADN opcionalmente escindida dependiendo del dominio catalítico adherido |

Figura 7

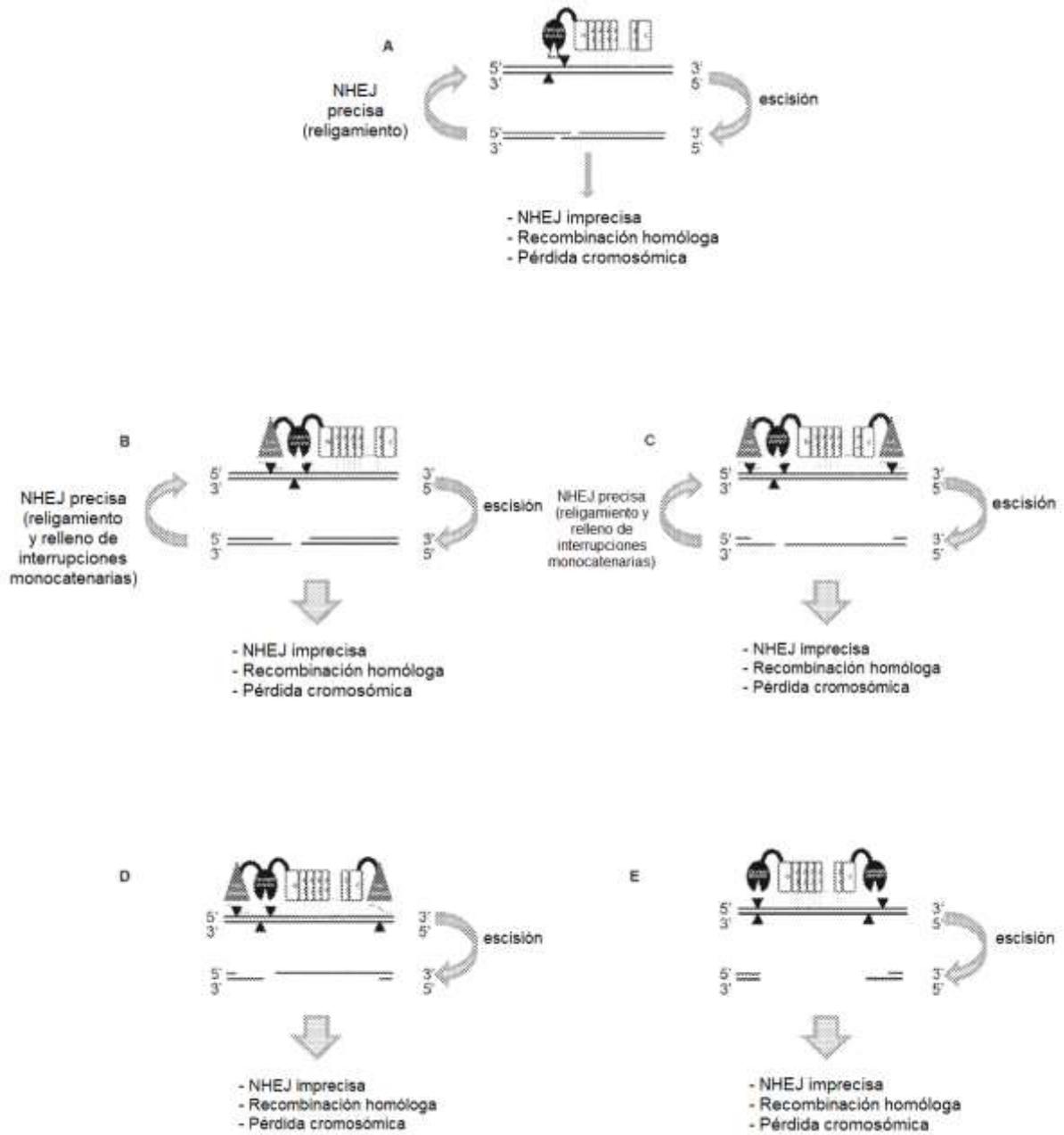


Figura 8

Longitud de espaciador [pb]	Control negativo	cT11Avr_TevD01	cT11Avr_TevD02	cT11Avr_TevD05
5	n.d.	n.d.	n.d.	n.d.
6	n.d.	n.d.	n.d.	n.d.
7	n.d.	n.d.	n.d.	n.d.
8	n.d.	n.d.	n.d.	n.d.
9	n.d.	n.d.	n.d.	n.d.
10	n.d.	+	+	n.d.
11	n.d.	+++	+	n.d.
12	n.d.	+	+	n.d.
13	n.d.	+	+	n.d.
14	n.d.	n.d.	n.d.	n.d.
15	n.d.	+++	+++	+++
16	n.d.	+	+	+
17	n.d.	n.d.	n.d.	n.d.
18	n.d.	n.d.	n.d.	n.d.
19	n.d.	+	n.d.	n.d.
20	n.d.	n.d.	n.d.	n.d.
21	n.d.	+	+	+
22	n.d.	+++	+++	+++
23	n.d.	+++	+++	+++
24	n.d.	+++	+++	+++
25	n.d.	+++	+++	+++
26	n.d.	+++	+++	+++
27	n.d.	+++	+++	+++
28	n.d.	+	+	n.d.
29	n.d.	n.d.	n.d.	n.d.
30	n.d.	n.d.	n.d.	n.d.

Figura 9

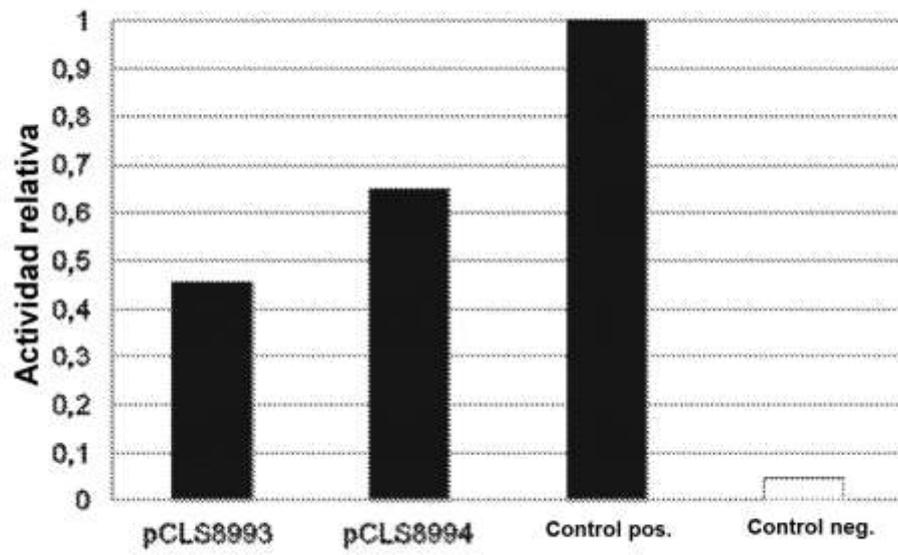


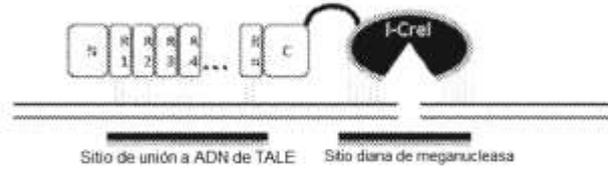
Figura 10

Longitud de espaciador [pb]	pCLS9596	pCLS9597	pCLS9599
5	+++	+++	+++
6	+++	+++	+++
7	+++	+++	+++
8	+++	+++	+++
9	+++	+++	+++
10	+++	+++	+++
11	+++	+++	+++
12	+++	+++	+++
13	+++	+++	+++
14	+++	+++	+++
15	+++	+++	+++
16	+++	+++	+++
17	+++	+++	+++
18	+++	+++	+++
19	+++	+++	+++
20	+++	+++	+++
21	+++	+++	+++
22	+++	+++	+++
23	+++	+++	+++
24	+++	+++	+++
25	+++	+++	+++
26	+++	+++	+++
27	+++	+++	+++
28	+++	+++	+++
29	+++	+++	+++
30	+++	+++	+++
31	+++	+++	+++
32	+++	+++	+++
33	+++	+++	+++
34	+++	+++	+++
35	+++	+++	+++
36	+++	+++	+++
37	+++	+++	+++
38	+++	+++	+++
39	+++	+++	+++
40	++	++	+++
compacta	+++	+++	+++
control neg.	n.d.	n.d.	n.d.

Figura 11

Longitud de espaciador [pb]	pCLS8589
5	+++
6	+++
7	+++
8	+++
9	+++
10	++
11	+++
12	+++
13	+++
14	++
15	++
16	+++
17	+++
18	+++
19	+++
20	+++
21	++
22	+++
23	+++
24	+++
25	+++
26	+++
27	+++
28	+++
29	+++
30	+++
31	+++
32	+++
33	+++
34	+++
35	+++
36	+++
37	++
38	++
39	++
40	++
compacta	++
control neg.	n.d.

Figura 12



TALE::Crel prototipo

Diana de meganucleasa TCRB02-A: GATGGCCATGGTAAGCAGGAGGGC

Nombre de RVD	Secuencia diana	Distancia desde el sitio TCRB02-A	Motivo RVD
TCRB2A1	<u>TGCTGGTCAGCGCC</u>	7 pb	NI-HD-NG-NN-NN-NG-HD-NI-NN-HD-NN-HD-HD-NG
TCRB2A2	<u>TGCTGTGCTGGTCAG</u>	12 pb	NN-HD-NG-NN-NG-NN-HD-NG-NN-NN-NG-HD-NI-NG
TCRB2A3	<u>TGTATGCTGTGCTGG</u>	16 pb	NI-NG-NI-NG-NN-HD-NG-NN-NG-NN-HD-NG-NN-NG

Región génica de TCRB02*	<p>GCCAACAGTGCCTACCAGCAAGGGGTCTGTCTGCCACCATCCTCT ATGAGATCCTGCTAGGGAAGGCCACCCTGTATGCTGT<u>GCTGGTCAGC</u> <u>GCCCTTGTGTTGATGGCCATGGTAAGCAGGAGGGCAGGATGGGGCCA</u> GCAGGCTGGAGGTGACACACTGACACCAAGCACCCAGAAGTATAGAG TCCCTGCCAGGATTGGAGCTGGGCAGTAGGGA</p>
--------------------------	---

*Secuencias diana de TCRB2A1 y TCRB02 subrayadas por referencia

Figura 13

Dianas	Secuencia
TCRB02.1	<u>GATGGCCATGGTAAGCAGGAGGGC</u>
TCRB02Tsp07	TGCTGGTCAGCGCCCttgtggtt <u>GATGGCCATGGTAAGCAGGAGGGC</u>
TCRB02Tsp12	TGCTGTGCTGGTCAGcgcccttgtggtt <u>GATGGCCATGGTAAGCAGGAGGGC</u>
TCRB02Tsp16	TGTATGCTGTGCTGGtcagcgcccttgtggtt <u>GATGGCCATGGTAAGCAGGAGGGC</u>

Diana	pCLS7185	pCLS13450	pCLS15148
TCRB02.1	+++	+	n.d.
TCRB02Tsp07	+++	+	+++
TCRB02Tsp12	+++	++	+++
TCRB02Tsp16	+++	+++	+++

Figura 14

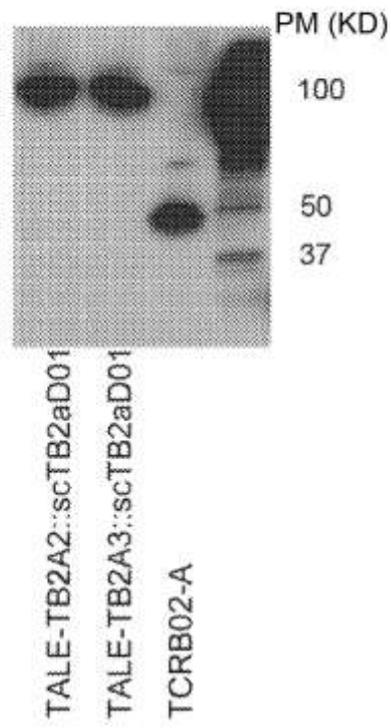


Figura 15

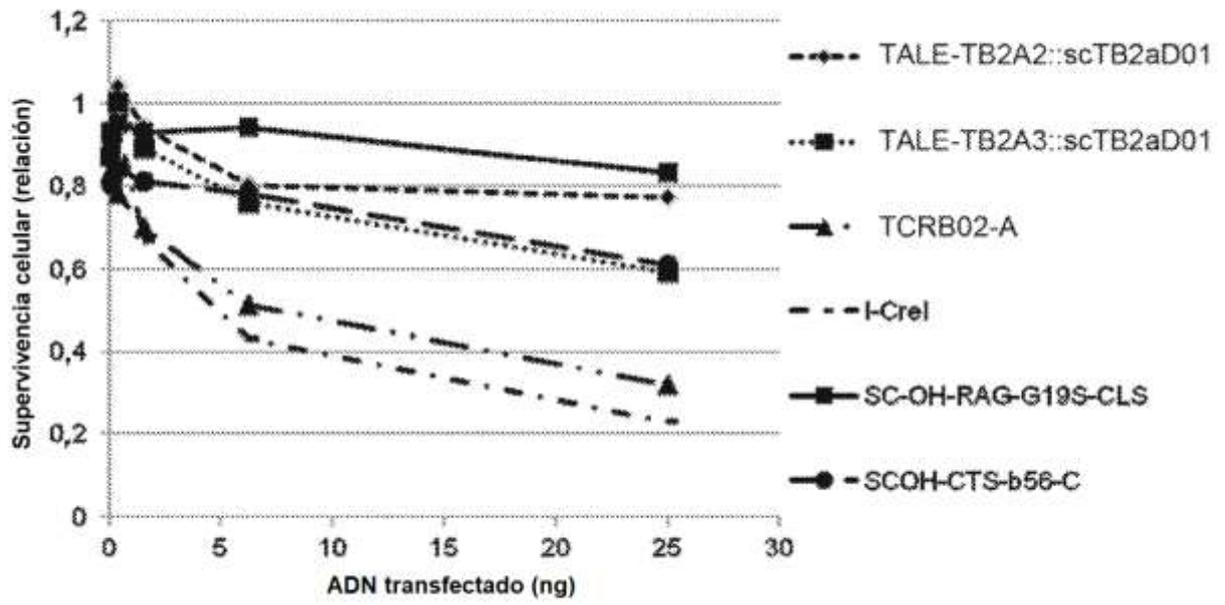


Figura 16

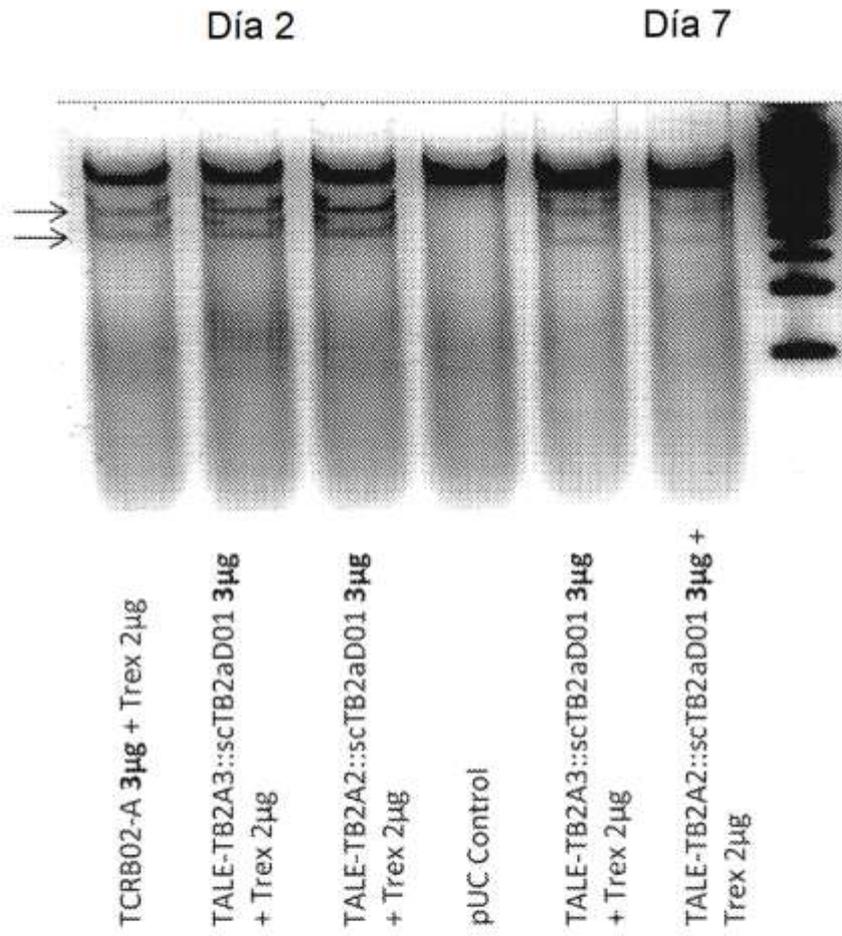


Figura 17

Diana	pCLS8522+pCLS7763	pCLS8522+pCLS8990
C_N_RAGAvr05	+++	++
C_N_RAGAvr06	+++	+
C_N_RAGAvr07	++	+
C_N_RAGAvr08	+	n.d.
C_N_RAGAvr09	+	+
C_N_RAGAvr10	+	n.d.
C_N_RAGAvr11	+	n.d.
C_N_RAGAvr12	+	+
C_N_RAGAvr13	+	n.d.
C_N_RAGAvr14	+	+
C_N_RAGAvr15	+	++
C_N_RAGAvr16	n.d.	n.d.
C_N_RAGAvr17	n.d.	+
C_N_RAGAvr18	n.d.	++
C_N_RAGAvr19	n.d.	++
C_N_RAGAvr20	n.d.	++
C_N_RAGAvr21	n.d.	++
C_N_RAGAvr22	+	+++
C_N_RAGAvr23	+	+++
C_N_RAGAvr24	++	+++
C_N_RAGAvr25	+	+++
C_N_RAGAvr26	+	+++
C_N_RAGAvr27	+	+++
C_N_RAGAvr28	+	++
C_N_RAGAvr29	+	++
C_N_RAGAvr30	+++	++
C_N_RAGAvr31	+++	+++
C_N_RAGAvr32	++	+++
C_N_RAGAvr33	++	++
C_N_RAGAvr34	+++	+++
C_N_RAGAvr35	+++	+++
C_N_RAGAvr36	+++	+++
C_N_RAGAvr37	+++	+++
C_N_RAGAvr38	+	++
C_N_RAGAvr39	+	+
C_N_RAGAvr40	n.d.	n.d.
control neg.	n.d.	n.d.

Figura 18

longitud de espaciador [pb]	pCLS9081	pCLS9082
40	+	+
39	+	+
38	+	n.d.
37	+	n.d.
36	+	+/-
35	+	+
34	+	+
33	+	+
32	+	+
31	+	+
30	+	+
29	+	+
28	+	+/-
27	+	+
26	+	+
25	+	+
24	+	+
23	+	+
22	+	+
21	+	+
20	+	+
19	+	+
18	+	+
17	+	+
16	+	+
15	+	+
14	n.d.	+/-
13	+	+
12	+	+/-
11	+	+/-
10	+	n.d.
09	+	n.d.
08	+	+/-
07	+	+
06	+	+
05	+	+
Compacta	+	+/-

Figura 19

ID de secuencia de cabeza catalítica	ID de secuencia de conector polipeptídico	Longitud de espaciador [pb]				Compacta
		15	18	21	24	
11	385	++	++	+++	+++	++
11	388	+++	+++	++	+++	++
11	394	++	+++	++	+++	++
11	400	+++	+++	+++	+++	++

Figura 20

Nombre de la diana	Secuencia diana de ácido nucleico
Avr05	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr06	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr07	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr08	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr09	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr10	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr11	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr12	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr13	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr14	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr15	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr16	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr17	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr18	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr19	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr20	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr21	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr22	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr23	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr24	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr25	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr26	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr27	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr28	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr29	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr30	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr31	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr32	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr33	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr34	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr35	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr36	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr37	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr38	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr39	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA
Avr40	TATATAAAGCTAAGCCCTCT <u>AGGTAAGAGGTTA</u> AGTTTATATA

Tabla 6

ES 2 728 436 T3

Nombre de la diana	Secuencia diana de ácido nucleico
Avr05	TATATAAACCTAACCCCTCT <u>AGGTAAGAGGGTTAGGTTTATATA</u>
Avr06	TATATAAACCTAACCCCTCT <u>AAGGTACAGAGGGTTAGGTTTATATA</u>
Avr07	TATATAAACCTAACCCCTCT <u>AAGGTACAGAGGGTTAGGTTTATATA</u>
Avr08	TATATAAACCTAACCCCTCT <u>GAAGGTACAGAGGGTTAGGTTTATATA</u>
Avr09	TATATAAACCTAACCCCTCT <u>GAAGGTACAGAGGGTTAGGTTTATATA</u>
Avr10	TATATAAACCTAACCCCTCT <u>TGAAGGTACAGAGGGTTAGGTTTATATA</u>
Avr11	TATATAAACCTAACCCCTCT <u>TGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr12	TATATAAACCTAACCCCTCT <u>TATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr13	TATATAAACCTAACCCCTCT <u>TATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr14	TATATAAACCTAACCCCTCT <u>CATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr15	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACAGAGGGTTAGGTTTATATA</u>
Avr16	TATATAAACCTAACCCCTCT <u>GCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr17	TATATAAACCTAACCCCTCT <u>GCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr18	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr19	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr20	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr21	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr22	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr23	TATATAAACCTAACCCCTCT <u>CTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr24	TATATAAACCTAACCCCTCT <u>CTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr25	TATATAAACCTAACCCCTCT <u>ACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr26	TATATAAACCTAACCCCTCT <u>ACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr27	TATATAAACCTAACCCCTCT <u>CACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr28	TATATAAACCTAACCCCTCT <u>CACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr29	TATATAAACCTAACCCCTCT <u>CCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr30	TATATAAACCTAACCCCTCT <u>CCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr31	TATATAAACCTAACCCCTCT <u>ACCCTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr32	TATATAAACCTAACCCCTCT <u>ACCCTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr33	TATATAAACCTAACCCCTCT <u>GACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr34	TATATAAACCTAACCCCTCT <u>GACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr35	TATATAAACCTAACCCCTCT <u>TGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr36	TATATAAACCTAACCCCTCT <u>TGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr37	TATATAAACCTAACCCCTCT <u>CTGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr38	TATATAAACCTAACCCCTCT <u>CTGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr39	TATATAAACCTAACCCCTCT <u>CTGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Avr40	TATATAAACCTAACCCCTCT <u>CTGACCACTAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
compacta	TATATAAACCTAACCCCTCT <u>TAGCATGAAGGTACCTAGAGGGTTAGGTTTATATA</u>
Control neg.	TTGTTCTCAGGTACCTCAGCCAGA

Tabla 7

ES 2 728 436 T3

Nombre de la diana	Secuencia diana de ácido nucleico
C_N_RAGAvr05	TGTTTATGGTTACTTAT <u>aggta</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr06	TGTTTATGGTTACTTAT <u>aaggta</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr07	TGTTTATGGTTACTTAT <u>aaggta</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr08	TGTTTATGGTTACTTAT <u>gaaggtac</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr09	TGTTTATGGTTACTTAT <u>gaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr10	TGTTTATGGTTACTTAT <u>gaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr11	TGTTTATGGTTACTTAT <u>gaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr12	TGTTTATGGTTACTTAT <u>atgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr13	TGTTTATGGTTACTTAT <u>atgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr14	TGTTTATGGTTACTTAT <u>catgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr15	TGTTTATGGTTACTTAT <u>tagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr16	TGTTTATGGTTACTTAT <u>catgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr17	TGTTTATGGTTACTTAT <u>catgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr18	TGTTTATGGTTACTTAT <u>agcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr19	TGTTTATGGTTACTTAT <u>agcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr20	TGTTTATGGTTACTTAT <u>agcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr21	TGTTTATGGTTACTTAT <u>tagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr22	TGTTTATGGTTACTTAT <u>tagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr23	TGTTTATGGTTACTTAT <u>tagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr24	TGTTTATGGTTACTTAT <u>tagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr25	TGTTTATGGTTACTTAT <u>actagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr26	TGTTTATGGTTACTTAT <u>actagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr27	TGTTTATGGTTACTTAT <u>cactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr28	TGTTTATGGTTACTTAT <u>cactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr29	TGTTTATGGTTACTTAT <u>cactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr30	TGTTTATGGTTACTTAT <u>cactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr31	TGTTTATGGTTACTTAT <u>accactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr32	TGTTTATGGTTACTTAT <u>accactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr33	TGTTTATGGTTACTTAT <u>gaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr34	TGTTTATGGTTACTTAT <u>gaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr35	TGTTTATGGTTACTTAT <u>gaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr36	TGTTTATGGTTACTTAT <u>gaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr37	TGTTTATGGTTACTTAT <u>ctgaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr38	TGTTTATGGTTACTTAT <u>ctgaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr39	TGTTTATGGTTACTTAT <u>tctgaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT
C_N_RAGAvr40	TGTTTATGGTTACTTAT <u>tctgaccactagcatgaaggtacc</u> TATATAAACCTAACCCCTCT

Tabla 13