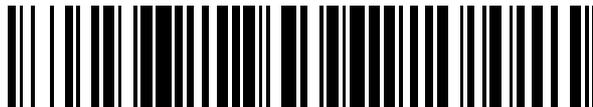


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 393 508**

51 Int. Cl.:

C12N 9/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06842325 .0**

96 Fecha de presentación: **03.10.2006**

97 Número de publicación de la solicitud: **1954807**

97 Fecha de publicación de la solicitud: **13.08.2008**

54 Título: **Variantes de la endonucleasa doméstica I-Crel que tienen nueva especificidad de escisión y uso de las mismas**

30 Prioridad:

25.10.2005 WO PCT/IB2005/003564

45 Fecha de publicación de la mención BOPI:

21.12.2012

45 Fecha de la publicación del folleto de la patente:

21.12.2012

73 Titular/es:

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

72 Inventor/es:

PAQUES, FRÉDÉRIC

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 393 508 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

Variantes de la endonucleasa doméstica I-CreI que tienen nueva especificidad de escisión y uso de las mismas

5 La invención se refiere a un procedimiento para ingeniería de variantes de la endonucleasa doméstica I-CreI capaces de escindir sitios I-CreI mutantes que tienen variaciones en las posiciones ± 8 a ± 10 . La invención también se refiere a una variante de la endonucleasa doméstica I-CreI que se puede obtener mediante dicho procedimiento, a un vector que codifica dicha variante, a una célula, un animal o una planta modificados por dicho vector y al uso de dicha variante de la endonucleasa doméstica I-CreI y productos derivados para ingeniería genética, terapia genómica y terapia antiviral.

Las meganucleasas son, por definición, endonucleasas específicas de secuencia con sitios de escisión grandes (> 14 pb) que pueden producir roturas en el ADN bicatenario (DSBs) en loci específicos en células vivas (Thierry y Dujon, *Nucleic Acids Res.*, 1992, 20, 5625 – 5631). Las meganucleasas se han usado para estimular la recombinación homóloga en las proximidades de sus secuencias diana en células y plantas cultivadas (Rouet y col., *Mol. Cell. Biol.*, 1994, 14, 8096 – 106; Choulika y col., *Mol. Cell. Biol.*, 1995, 15, 1968 – 73; Donoho y col., *Mol. Cell. Biol.*, 1998, 18, 4070 – 8; Elliott y col., *Mol. Cell. Biol.*, 1998, 18, 93 – 101; Sargent y col., *Mol. Cell. Biol.*, 1997, 17, 267 – 77; Puchta y col., *Proc. Natl. Acad. Sci. USA*, 1996, 93, 5055 – 60; Chiurazzi y col., *Plant Cell*, 1996, 8, 2057 – 2066), lo que convierte a la recombinación inducida por meganucleasas en un procedimiento eficiente y sólido para la ingeniería genómica. El uso de la recombinación inducida por meganucleasas siempre ha estado limitado por el repertorio de meganucleasas naturales y la principal limitación de la tecnología actual es el requisito de la introducción previa de un sitio de escisión para meganucleasas en el locus de interés.

Por tanto, la fabricación de meganucleasas artificiales con especificidades de sustrato adaptadas se encuentra en intensa investigación. Dichas proteínas podrían usarse para escindir secuencias cromosómicas genuinas y abren nuevas perspectivas para la ingeniería genómica en un amplio abanico de aplicaciones. Por ejemplo, las meganucleasas podrían usarse para inducir la corrección de mutaciones ligadas a enfermedades monogénicas hereditarias y sortear el riesgo debido a los transgenes insertados aleatoriamente usados en los actuales enfoques de la terapia génica (Hacein-Bey-Abina y col., *Science*, 2003, 302, 415 – 419).

Recientemente, los dominios de unión del ADN del dedo de cinc de las proteínas del dedo de cinc (ZFP) de tipo Cys2-His2 se han podido fusionar con el dominio catalítico de la endonucleasa *FokI* para inducir recombinación en varios tipos celulares, incluidas las células linfoides humanas (Smith y col., *Nucleic Acids Res.*, 1999, 27, 674 – 81; Pabo y col., *Annu. Rev. Biochem.*, 2001, 70, 313 – 40; Porteus and Baltimore, *Science*, 2003, 300, 763; Umov y col., *Nature*, 2005, 435, 646 – 651; Bibikova y col., *Science*, 2003, 300, 764). La especificidad de unión de las ZFP es relativamente fácil de manipular y en la actualidad se dispone de un repertorio de nuevas ZFP artificiales capaces de unirse a muchas secuencias (g/a)nn(g/a)nn(g/a)nn (Pabo y col., citado anteriormente; Segal y Barbas, *Curr. Opin. Biotechnol.*, 2001, 12, 632 – 7; Isalan y col., *Nat. Biotechnol.*, 2001, 19, 656 – 60). No obstante, conservar una especificidad de sustrato muy estrecha es uno de los principales problemas de las aplicaciones de ingeniería genómica y actualmente no está claro si las ZFP satisfarían los requisitos muy estrictos para aplicaciones terapéuticas. Además, estas proteínas de fusión han mostrado una toxicidad elevada en las células (Porteus and Baltimore, citado anteriormente; Bibikova y col., *Genetics*, 2002, 161, 1169 1175)), probablemente a causa de un nivel bajo de especificidad.

En la naturaleza, las meganucleasas están representadas esencialmente por endonucleasas domésticas (ED), una familia de endonucleasas codificadas por elementos genéticos móviles cuya función es iniciar la recombinación inducida por rotura del ADN bicatenario (DSB) en un proceso denominado localización (Chevalier and Stoddard, *Nucleic Acids Res.*, 2001, 29, 3757 – 74; Kostriken y col., *Cell*; 1983, 35, 167 – 74; Jacquier and Dujon, *Cell*, 1985, 41, 383 – 94). Se han identificado varios cientos de ED en bacterias, eucariotas y arqueobacterias (Chevalier and Stoddard, citado anteriormente); no obstante, la probabilidad de encontrar un sitio de escisión para ED en un gen elegido es muy baja.

Dada su función biológica y sus excepcionales proteínas de escisión en términos de eficacia y especificidad, las ED proporcionan armazones ideales para derivar nuevas endonucleasas para ingeniería genómica. Durante la última década se han acumulado datos que caracterizan la familia LAGLIDADG, la más grande de las cuatro familias de ED (Chevalier and Stoddard, citados previamente). LAGLIDADG se refiere a la única secuencia que actualmente está conservada en la familia y se encuentra en una o (con mayor frecuencia) dos copias en la proteína. Las proteínas con un único motivo, como I-CreI, forman homodímeros y escinden secuencias de ADN palindrómicas o pseudopalindrómicas; mientras que las proteínas más grandes con doble motivo, como I-SceI, son monómeros y escinden dianas no palindrómicas. Se han cristalizado siete proteínas LAGLIDADG diferentes y exhiben una conservación muy sorprendente de la estructura del núcleo que contrasta con la ausencia de similitud a nivel de la secuencia primaria (Jurica y col., *Mol. Cell.*, 1998, 2, 469 – 76; Chevalier y col., *Nat. Struct. Biol.*, 2001, 8, 312 – 6 ;

Chevalier y col., *J. Mol. Biol.*, 2003, 329, 253 – 69; Moure y col., *J. Mol. Biol.*, 2003, 334, 685 – 95; Moure y col., *Nat. Struct. Biol.*, 2002, 9, 764 – 70; Ichiyanagi y col., *J. Mol. Biol.*, 2000, 300, 889 – 901; Duan y col., *Cell*, 1997, 89, 555 – 64; Bolduc y col., *Genes Dev.*, 2003, 17, 2875 – 88; Silva y col., *J. Mol. Biol.*, 1999, 286, 1123 – 36). En esta estructura del núcleo dos pliegues $\alpha\beta\alpha\beta\beta\alpha$ característicos, también denominados dominios centrales de la endonucleasa doméstica LAGLIDADG, a los que contribuyen dos monómeros o dos dominios en las proteínas LAGLIDADG dobles, están uno enfrente del otro con una simetría de dos veces. La unión al ADN depende de las cuatro hebras β de cada dominio, plegadas en una lámina β antiparalela y formando una silla en la ranura principal de la hélice de ADN. El análisis de la estructura de I-Crel unida a su diana natural muestra que en cada monómero, ocho residuos Y33, Q38, N30, K28, Q26, Q44, R68 y R70) establecen una interacción directa con siete bases en las posiciones $\pm 3, 4, 5, 6, 7, 9$ y 10 (Jurica y col., 1998, citado previamente). Además, algunos residuos establecen contacto mediado por agua con varias bases, por ejemplo S40, K28 y N30, con el par de bases en la posición $+ 8$ y $- 8$ (Chevalier y col., 2003, citado previamente). El núcleo catalítico es central, con una contribución de ambos monómeros / dominios simétricos. Además de esta estructura del núcleo, se pueden encontrar otros dominios, por ejemplo, PI-Scel, una inteína, tiene un dominio de corte y empalme de proteínas, y un dominio de ADN adicional (Moure y col., 2002, citado anteriormente; Grindl y col., *Nucleic Acids Res.*, 1998, 26, 1857 – 62).

Se están investigando dos abordajes para derivar meganucleasas nuevas de las endonucleasas domésticas.

- proteínas monocatenarias híbridas o quiméricas

Se podrían obtener nuevas meganucleasas mediante intercambio de los dominios centrales de endonucleasa doméstica LAGLIDADG de diferentes monómeros (Epinat y col., *Nucleic Acids Res.*, 2003, 31, 2952 – 62; Chevalier y col., *Mol. Cell.*, 2002, 10, 895 – 905; Steuer y col., *Chembiochem.*, 2004, 5, 206 – 13; las solicitudes internacionales de PCT WO 03/078619 y WO 2004/031346). Estas meganucleasas quiméricas monocatenarias en las que dos dominios centrales de endonucleasa doméstica LAGLIDADG de diferentes meganucleasas están unidos mediante un espaciador, son capaces de escindir la diana híbrida correspondiente a la fusión de las dos semiseuencias diana del ADN parental. Estos resultados significan que los dos dominios de unión a ADN de un dímero de I-Crel se comportan de forma independiente; cada dominio de unión a ADN se une a una mitad diferente del sitio diana del ADN. La construcción de ED artificiales quiméricas y de una cadena ha sugerido que se puede usar un abordaje combinatorio para obtener nuevas meganucleasas que escinden nuevas secuencias diana (no palindrómicas): diferentes monómeros o dominios centrales se pueden fusionar en una única proteína para alcanzar nuevas especificidades.

No obstante, este abordaje no enriquece considerablemente el número de secuencias de ADN que pueden ser las dianas de las endonucleasas domésticas, ya que las nuevas dianas que se generan son el resultado de la combinación de dos semisitios diana de ADN diferentes.

- variantes proteicas

A menudo se ha demostrado que alterar la especificidad de sustrato de las proteínas de unión a ADN mediante mutagénesis y detección selectiva / selección es difícil (Lanio y col., *Protein Eng.*, 2000, 13, 275 – 281; Voziyarov y col., *J. Mol. Biol.*, 2003, 326, 65 – 76; Santoro y col., *P.N.A.S.*, 2002, 99, 4185 – 4190; Buchholz y Stewart, *Nat. Biotechnol.*, 2001, 19, 1047 – 1052), y, más particularmente, la ingeniería del dominio de unión al ADN de las ED siempre se ha considerado una tarea de enormes proporciones (Ashworth y col., *Nature* 2006, 441, 656 – 659; Gimble y col., *J. Mol. Biol.*, 2003, 334, 993 – 1008 ; Arnould y col., *J. Mol. Biol.*, 2006, 355, 443 458; Doyon y col., *J. Am. Chem. Soc.*, 2006, 128, 2477 – 2484; Steuer y col., citado previamente; Seligman y col., *Nucleic Acids Res.*, 2002, 30, 3870 – 3879).

El análisis de la estructura de cristal de I-Crel / ADN indica que los 9 aminoácidos están en contacto directo con el sitio doméstico (Chevalier y col., 2003; Jurica y col., citado previamente) cuya aleatorización tendría como resultado 20^9 combinaciones, un número más allá de toda capacidad de detección selectiva hoy en día.

Por tanto, varios laboratorios han confiado en un abordaje semirracional (Chica y col., *Curr. Opin. Biotechnol.*, 2005, 16, 378 – 384) para limitar la diversidad de las bibliotecas mutantes que se van a manipular: un pequeño conjunto de residuos relevantes se escoge de acuerdo con datos estructurales. No obstante, esto no fue suficiente para crear endonucleasas rediseñadas que escindan secuencias elegidas:

- Seligman y colaboradores usaron un abordaje racional para sustituir residuos individuales específicos del pliegue $\alpha\beta\alpha\beta\beta\alpha$ de I-Crel (Sussman y col., *J. Mol. Biol.*, 2004, 342, 31 – 41; Seligman y col., *Nucleic Acids Res.*, 2002, citado previamente; Seligman y col., *Genetics*, 1997, 147, 1653 – 64). No obstante, se observó una escisión sustancial con pocas variantes de I-Crel (Y33C, Y33H, Y33R, Y33L, Y33S, Y33T, S32K, S32R) y solo para una diana modificada en la posición ± 10 .

- De un modo similar, Gimble y col. (citado previamente) modificaron el dominio de unión a ADN adicional de PI-Scel; obtuvieron una proteína variante con una especificidad de unión alterada pero con una especificidad de sustrato sin alterar, y la mayoría de las proteínas mantenían mucha afinidad por la secuencia diana silvestre.

5 Para llegar a un gran número de secuencias, sería extremadamente valioso poder generar otras variantes de endonucleasas domésticas con nueva especificidad de sustrato, es decir, que puedan escindir dianas de ADN que no son escindidas por la endonucleasa doméstica parental o por las pocas variantes que se han aislado hasta ahora.

En concreto, sería extremadamente valioso generar variantes de endonucleasas domésticas capaces de escindir
10 nuevas dianas de ADN en las que varios nucleótidos de la diana de ADN de la meganucleasa silvestre se han mutado de forma simultánea.

No obstante, este abordaje no es fácil, ya que la interfaz de unión al ADN de las ED es muy compacta y las dos horquillas $\beta\beta$ diferentes que son responsables de casi todas las interacciones específicas de base son parte de un solo pliegue. Por tanto, la mutación de varios aminoácidos colocados en las proximidades que es necesaria para unir
15 una diana mutada en varias posiciones puede romper la estructura de la interfaz de unión.

El inventor ha sometido a ingeniería a cientos de nuevas variantes de I-Crel que, en conjunto, se dirigen a los 64 posibles sitios mutados de I-Crel que difieren en las posiciones ± 10 , ± 9 y ± 8 . Estas variantes que tienen nueva
20 especificidad de sustrato hacia los nucleótidos ± 8 , ± 9 y/o ± 10 incrementan el número de secuencias de ADN que pueden ser diana de las meganucleasas.

Posibles aplicaciones incluyen ingeniería genética, ingeniería genómica, terapia génica y terapia antiviral.

25 Por tanto, la invención se refiere a un procedimiento para realizar ingeniería en una variante de endonucleasa doméstica I-Crel que tiene una especificidad de escisión modificada, que comprende al menos las etapas de:

(a) sustituir al menos uno de los aminoácidos K28, N30, Y33, Q38 y/o S40 de la horquilla $\beta_1\beta_2$ de I-Crel, con un aminoácido seleccionado del grupo constituido por A, C, D, E, G, H, K, N, P, Q, R, S, T, L, V, W e Y.

30

(b) seleccionar las variantes de I-Crel de la etapa (a) que son capaces de escindir una secuencia diana de ADN constituida por un sitio I-Crel mutante en el que al menos el doblete nucleotídico aa en las posiciones - 9 a - 8 y/o el doblete nucleotídico tt en las posiciones + 8 a + 9 se han sustituido por un doblete nucleotídico diferente.

35 **Definiciones**

- Los residuos de aminoácidos en una secuencia polipeptídica se designan en el presente documento de acuerdo con el código de una letra, en el que, por ejemplo, K significa Lys o Lisina, N significa Asn o Asparagina e Y significa Tyr o Tirosina.

40

- Los nucleótidos se designan del siguiente modo: se usa un código de una letra para designar 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
45 representa a t o c y n representa g, a, t o c.

- con "I-Crel" se pretende decir I-Crel silvestre que tiene la secuencia SWISSPROT P05725 o código de acceso pdb 1g9y.

50 - con "variante de I-Crel" o "variante" se pretende decir una proteína obtenida mediante sustitución de al menos un aminoácido de I-Crel con un aminoácido diferente.

- con "variante de I-Crel funcional" se pretende decir una variante de I-Crel que es capaz de escindir un ADN diana, preferentemente un ADN diana que no está escindido por I-Crel. Por ejemplo, dichas variantes tienen una variación
55 de aminoácido en las posiciones en contacto con la secuencia diana de ADN o que interaccionan directa o indirectamente con dicha diana de ADN.

- con "dominio de endonucleasa doméstica" o "dominio" se pretende decir la región que interacciona con la mitad del ADN diana de una endonucleasa doméstica y que puede asociarse con el otro dominio de la misma endonucleasa
60 doméstica que interacciona con la otra mitad de la diana de ADN para formar una endonucleasa funcional capaz de escindir dicha diana de ADN.

- 5 - Con "dominio central" se quiere decir "dominio central de la endonucleasa doméstica LAGLIDADG" que es el pliegue $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ característico de las endonucleasas domésticas de la familia de LAGLIDADG, correspondiente a una secuencia de aproximadamente cien residuos de aminoácidos. Dicho dominio comprende cuatro hebras beta ($\beta_1, \beta_2, \beta_3, \beta_4$) plegadas en una lámina beta antiparalela que interacciona con una mitad de la diana de ADN. Por ejemplo, en el caso de la endonucleasa doméstica dimérica I-Crel (163 aminoácidos, el dominio central de la endonucleasa doméstica LAGLIDADG corresponde a los residuos 6 a 94. En el caso de la endonucleasa doméstica monomérica, dos de estos dominios se encuentran en la secuencia de la endonucleasa; por ejemplo, en I-Dmol (194 aminoácidos), el primer dominio (residuos 7 a 99) y el segundo dominio (residuos 104 a 194) están separados por un ligador corto (residuos 100 a 103).
- 10 - con "sitio de I-Crel" se quiere decir una secuencia de ADN bicatenario de 22 a 24 pb que es escindida por I-Crel. Los sitios de I-Crel incluyen el sitio doméstico de I-Crel no palindrómico silvestre (natural) y las secuencias palindrómicas derivadas que están presentes en la figura 1A, tal como la secuencia 5'- $t_{12}c_{-11}a_{-10}a_{-9}a_{-8}a_{-7}-c_{-6}g_{-5}t_{-4}c_{-3}g_{-2}t_{-1}a_{+1}c_{+2}g_{+3}a_{+4}c_{+5}g_{+6}t_{+7}t_{+8}t_{+9}t_{+10}g_{+11}a_{+12}$ también denominada C1221 (SEQ ID NO: 3).
- 15 - con "diana de ADN", "secuencia de la diana de ADN", "secuencia diana", "diana", "sitio de reconocimiento", "secuencia de reconocimiento", "sitio de reconocimiento doméstico", "sitio doméstico", "sitio de escisión" se pretende decir una secuencia polinucleotídica bicatenaria de 22 a 24 pb palindrómica, parcialmente palindrómica (seudopalindrómica) o no palindrómica que es reconocida y escindida por una meganucleasa, por ejemplo una endonucleasa doméstica LAGLIDADG tal como I-Crel, o una variante o una meganucleasa quimérica monocatenaria derivada de dicha meganucleasa. Estos términos se refieren a una localización de ADN distinta, preferentemente una localización genómica, en la que una rotura bicatenaria (escisión) se va a inducir por la endonucleasa. La diana de ADN está definida por la secuencia 5' a 3' de una hebra del polinucleótido bicatenario.
- 20 - con "semisitio diana de ADN" se pretende decir la porción de la diana de ADN que está unida por cada dominio central de la endonucleasa doméstica LAGLIDADG.
- con "diana de ADN quimérica" o "diana de ADN híbrida" se pretende decir la fusión de una mitad diferente de cada secuencia diana de ADN de la meganucleasa parental.
- 30 - con "variante de endonucleasa doméstica con nueva especificidad" se pretende decir una variante que tiene un patrón de dianas escindidas diferente del de la endonucleasa doméstica parental. Las expresiones "nueva especificidad", "especificidad modificada", "nueva especificidad de escisión", "nueva especificidad de sustrato" que son equivalentes y se usan de forma intercambiable, se refieren a la especificidad de las variantes hacia los nucleótidos de la secuencia diana de ADN.
- 35 - con "vector", se pretende decir una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido.
- 40 - con "homólogo" se pretende decir una secuencia con suficiente identidad con otra para conducir a una recombinación homóloga entre secuencias, más particularmente que tiene una identidad de al menos 95 %, preferentemente una identidad de 97 % y, más preferentemente, de 99 %.
- "identidad", se refiere a la identidad de secuencia entre dos moléculas de ácido nucleico o polipéptidos. La identidad se puede determinar comparando una posición en cada secuencia que se puede alinear a efectos de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base, las moléculas son idénticas en dicha posición. Un grado de similitud o identidad entre las secuencias de ácido nucleico o de aminoácidos es una función del número de nucleótidos idénticos o equivalentes en las posiciones compartidas por las secuencias de ácidos nucleicos. Se pueden usar varios algoritmos y/o programas de alineación para calcular la identidad entre dos secuencias, incluidos FASTA o BLAST, que están disponibles como parte del paquete de análisis de secuencia GCG (University of Wisconsin, Madison, Wis.), y se pueden usar con, por ejemplo, parámetros predeterminados.
- 45 - "individuo" incluye mamíferos, así como otros vertebrados (p. ej., aves, peces y reptiles). El término "mamífero", como se usa en el presente documento, se refiere a cualquier animal vertebrado, incluidos los monotremas; marsupiales y placentarios, que maman cuando son jóvenes y dan a luz a jóvenes vivos, mamíferos eutarianos o placentarios) o son ponedores de huevos (mamíferos metatarios o no placentarios). Ejemplos de especies de mamíferos incluyen seres humanos y otros primates (p. ej., monos, chimpancés), roedores (p. ej., ratas, ratones, cobayas) y rumiantes (p. ej., vacas, cerdos, caballos).
- 50 - "enfermedad genética" se refiere a cualquier enfermedad, parcial o completamente, directa o indirectamente,
- 55
- 60

debido a una anomalía en uno o varios genes. Dicha anomalía puede ser una mutación, una inserción o una delección. Dicha mutación puede ser una mutación puntual. Dicha anomalía puede afectar a la secuencia de codificación del gen o su secuencia reguladora. Dicha anomalía puede afectar a la estructura de la secuencia genómica o a la estructura o estabilidad del ARNm codificado. Dicha enfermedad genética puede ser recesiva o dominante. Dicha enfermedad genética podría ser, entre otras, fibrosis quística, corea de Huntington, hipercolesterolemia familiar (defecto en los receptores de LDL), hepatoblastoma, enfermedad de Wilson, profirias hepáticas congénitas, trastornos hereditarios del metabolismo hepático, síndrome de Lesch Nyhan, drepanocitosis, talasemias, serodermia pigmentosa, anemia de Fanconi, retinitis pigmentosa, ataxia telangiectasia, síndrome de Bloom, retinoblastoma, distrofia muscular de Duchenne y enfermedad de Tay-Sachs.

10

De acuerdo con el procedimiento de la invención, la(s) mutación(es) de aminoácidos en la etapa (a) se introducen en I-Crel silvestre o en una variante funcional de la misma. La etapa a) puede comprender la introducción de mutaciones adicionales, en particular en otras posiciones en contacto con la secuencia diana de ADN o que interaccionan directa o indirectamente con dicha diana de ADN. Las variantes funcionales comprenden mutaciones que no afectan a la estructura de la proteína. Por ejemplo, las mutaciones de aminoácidos de la etapa (a) se pueden introducir en una variante de I-Crel que comprende una o más mutaciones seleccionadas del grupo constituido por:

15

- la mutación de isoleucina en la posición 24 en una valina (I24V),

20

- la mutación de la arginina en la posición 70 en una serina (R70S), y

- la mutación del ácido aspártico en la posición 75, en un aminoácido sin carga, preferentemente una asparagina (D75N) o una valina (D75V).

25

La etapa a) se puede realizar generando una biblioteca de variantes como se describe en la solicitud de PCT internacional WO 2004/067736.

La detección selectiva en la etapa (b) se puede realizar usando un ensayo de escisión *in vitro* o *in vivo*, como se describe en la solicitud de PCT internacional WO 2004/067736.

30

De acuerdo con una realización ventajosa de dicho procedimiento, la diana de ADN en la etapa b) deriva de un sitio I-Crel que se selecciona de C1234, C4334 y C1221 (SEQ ID NO: 1 a 3, figura 1A).

35

De acuerdo con otra realización ventajosa de dicho procedimiento, la diana de ADN en la etapa b) comprende una secuencia que tiene la fórmula:

$$C^{-11}N^{-10}N^{-9}N^{-8}M^{-7}Y^{-6}N^{-5}N^{-4}N^{-3}K_2Y^{-1}r + 1m + 2n + 3n + 4n + 5r + 6k + 7n + 8n + 9n + 10g + 11$$

(I),

40

en la que n es a, t, c o g, m es a o c, y es c o t, k es g o t, r es a o g (SEQ ID NO: 75), que proporciona que cuando n_{9n-8} es aa, n_{+8n+9} es diferente de tt y cuando n_{+8n+9} es tt, n_{8n-9} es diferente a un aa.

45

De acuerdo con una realización preferida de dicho procedimiento, n_{5n-4n-3} es gtc y/o n_{+3n+4n+5} es gac.

La diana de ADN en la etapa b) puede ser palindrómica, no palindrómica o seudo palindrómica. Preferentemente, la secuencia de nucleótidos desde las posiciones - 11 a - 8 y + 8 a + 11 y/o la secuencia de nucleótidos desde las posiciones - 5 a - 3 y/o + 3 a + 5 son palindrómicas.

50

De acuerdo con otra realización ventajosa de dicho procedimiento, la diana de ADN en la etapa b) comprende un doblete nucleotídico en las posiciones - 9 a - 8, que se selecciona del grupo constituido por: ag, at, ac, ga, gg, gt, gc, ta, tg, tt, cg, ct o cc, y/o un doblete nucleotídico en las posiciones + 8 a + 9, que es la secuencia complementaria inversa de dicho doblete nucleotídico en las posiciones - 9 a - 8, es decir ct, ta, gt, tc, cc, ac, gc, at, ca, aa, cg, ag o gg.

55

De acuerdo con otra realización ventajosa de dicho procedimiento, la diana de ADN en la etapa b) comprende además la sustitución del nucleótido a en la posición - 10 y/o del nucleótido t en la posición + 10 del sitio I-Crel, con un nucleótido diferente.

60

Preferentemente, dicha diana de ADN comprende un triplete nucleotídico en las posiciones - 10 a - 8, que se selecciona del grupo constituido por: aac, aag, aat, acc, acg, act, aga, agc, agg, agt, ata, atg, cag, cga, cgg, ctg, gac, gag, gat, gcc, gga, ggc, ggg, ggt, gta, gtg, gtt, tac, tag, tat, tcc, tga, tgc, tgg, tgt o ttg, y/o un triplete nucleotídico en

las posiciones + 8 a + 10, que es la secuencia complementaria inversa de dicho triplete nucleotídico en las posiciones - 10 a - 8.

De acuerdo con otra realización ventajosa de dicho procedimiento, la etapa (b) se realiza *in vivo* en las condiciones en las que la rotura de la doble cadena en la secuencia diana de ADN mutada que se genera mediante dicha variante conduce a la activación de un marcador de selección positiva o un gen indicador, o la inactivación de un marcador de selección negativa o un gen indicador, mediante reparación mediada por recombinación de dicha rotura de la doble cadena de ADN.

Por ejemplo, la actividad de escisión de la variante de I-Crel de la invención se puede medir mediante un ensayo de recombinación de repetición directa, en levaduras o células de mamífero, usando un vector indicador, como se describe en la solicitud de PCT WO 2004/067736. El vector indicador comprende dos copias no funcionales truncadas de un gen indicador (repeticiones directas) y una secuencia diana de ADN dentro de la secuencia intermedia, clonadas en un vector de expresión de levaduras o de mamífero (Figura 2). La secuencia diana de ADN deriva de un sitio I-Crel tal como C1221, mediante sustitución de uno a tres nucleótidos en las posiciones ± 8 a 10 (Figura 1B). La expresión de una variante de I-Crel funcional que es capaz de escindir la secuencia diana de ADN induce recombinación homóloga entre las repeticiones directas resultante en un gen indicador funcional, cuya expresión se puede monitorizar mediante un ensayo adecuado.

De acuerdo con otra realización ventajosa de dicho procedimiento, comprende una etapa adicional c₁) de expresión de una variante obtenida en la etapa b) para permitir la formación de homodímeros.

De acuerdo con otra realización ventajosa de dicho procedimiento, comprende una etapa adicional c₂) de coexpresión de una variante obtenida en la etapa b) y I-Crel o una variante funcional de la misma para permitir la formación de heterodímeros. Preferentemente, dos variantes diferentes obtenidas en la etapa b) se coexpresan.

Por ejemplo, las células huésped se pueden modificar mediante uno o dos vectores de expresión recombinante que codifica dicha(s) variante(s). Las células se cultivan después en condiciones que permiten la expresión de la(s) variante(s) y los homodímeros / heterodímeros que se forman y, después, se recuperan del cultivo celular.

De acuerdo con el procedimiento de la invención, las endonucleasas quiméricas monocatenarias se pueden construir mediante la fusión de una variante obtenida en la etapa b) con un dominio / monómero de endonucleasa doméstica. Dicho dominio / monómero puede ser de una endonucleasa doméstica silvestre o una variante funcional de la misma.

Los procedimientos para construir moléculas quiméricas monocatenarias derivadas de endonucleasas domésticas son bien conocidos en la técnica (Epinat y col., *Nucleic Acids Res.*, 2003, 31, 2952 – 62; Chevalier y col., *Mol. Cell.*, 2002, 10, 895 – 905; Steuer y col., *ChemBiochem.*, 2004, 5, 206 – 13; las solicitudes internacionales de PCT WO 03/078619 y WO 2004/031346). Cualquiera de estos procedimientos se puede aplicar para construir moléculas quiméricas monocatenarias derivadas de las variantes como se define en la presente invención.

También se describe una variante de la meganucleasa I-Crel que se puede obtener mediante el procedimiento como se ha definido anteriormente, pudiendo dicha variante escindir una secuencia diana de ADN constituida por un sitio I-Crel mutante en el que al menos uno de los nucleótidos a en la posición - 9 a - 8 o uno de los nucleótidos t en la posición + 8 y + 9 se han sustituido por nucleótido diferente.

Dicha variante de I-Crel tiene una arginina (R) o una lisina (K) en la posición 38; las variantes que tienen R o K en la posición 38 puede escindir una diana de ADN que comprende una guanina en la posición - 9 o una citosina en la posición + 9.

Dicha variante se puede seleccionar de las variantes que tienen residuos de aminoácidos en las posiciones 28, 30, 33, 38 y 40 respectivamente, que se seleccionan del grupo constituido por:

Q28/N30/Y33/K38/R40, R28/N30/K33/R38/Q40, Q28/N30/R33/R38/R40, Q28/N30/Y33/K38/K40, K28/N30/T33/R38/Q40, K28/N30/S33/R38/E40, S28/N30/Y33/R38/K40, K28/N30/S33/R38/D40, K28/N30/S33/R38/S40, Q28/N30/Y33/R38/K40, Q28/N30/K33/R38/T40, N28/N30/S33/R38/K40, N28/N30/S33/R38/R40, E28/N30/R33/R38/K40, R28/N30/T33/R38/A40, Q28/N30/Y33/R38/A40, Q28/N30/Y33/R38/S40, K28/N30/R33/K38/A40, R28/N30/A33/K38/540, A28/N30/N33/R38/K40, Q28/N30/S33/R38/K40, K28/A30/H33/R38/S40, K28/H30/H33/R38/S40, K28/E30/S33/R38/S40, K28/N30/H33/R38/540, K28/D30/H33/K38/S40, K28/K30/H33/R38/S40, K28/S30/H33/R38/S40 y K28/G30/V33/R38/540.

Dicha variante I-Crel tiene residuos de aminoácidos en las posiciones 28, 30, 33, 38 y 40 respectivamente, que se seleccionan del grupo constituido por:

ES 2 393 508 T3

K28/R30/G33/T38/S40, K28/N30/N33/Q38/S40, K28/H30/M33/A38/S40, R28/D30/V33/Q38/S40, 5 K28/G30/V33/G38/S40.	K28/N30/H33/S38/S40, K28/N30/D33/Q38/S40, K28/S30/S33/H38/S40, K28/G30/V33/T38/S340,	K28/Q30/H33/Q38/S40, K28/D30/R33/G38/S40, K28/G30N33/A38/S40, K28/G30/V33/H38/S40,	K28/N30/H33/G38/S40, K28/N30/H33/A38/S40, K28/D30N33/H38/S40, K28/G30/V33/R38/S40 y
--	---	---	--

10 Dicha variante I-Crel es una variante capaz de escindir al menos una secuencia diana de ADN constituida por un sitio I-Crel mutante en el que al menos la a en posición - 8 y/o la t en posición + 8 se han sustituido con un nucleótido diferente, teniendo dicha variante los residuos de aminoácidos en las posiciones 28, 30, 33, 38 y 40 respectivamente, que se seleccionan del grupo constituido por:

15 Q28/N30/Y33/K38/R40, Q28/N30/T33/Q38/K40, N28/N30/Y33/Q38/R40, K28/N30/S33/R38/D40, A28/N30/S33/Q38/R40, K28/N30/R33/Q38/E40, Q28/N30/Y33/Q38/R40, K28/N30/N33/Q38/A40, K28/N30/R33/T38/R40, R28/N30/T33/R38/A40, S28/N30/Q33/A38/A40, K28/N30/R33/Q38/R40, K28/N30/R33/A38/R40, K28/N30/R33/N38/A40, K28/N30/Y33/Q38/N40, K28/N30/R33/A38/T40, R28/N30/Y33/Q38/S40, K28/N30/R33/Q38/A40, K28/D30/N33/H38/S40, K28/A30/R33/Q38/S40, K28/R30/R33/E38/S40, K28/D30/H33/K38/S40, K28/G30/R33/Q38/S40 35 K28/N30/H33/N38/S40, K28/T30/D33/H38/S40, K28/N30/R33/A38/S40, K28/D30/R33/T38/S40, K28/N30/H33/S38/S40, K28/N30/D33/Q38/S40, K28/S30/S33/H38/S40, R28/D30/V33/Q38/S40, K28/G30/V33/R38/S40, R28/N30/V33/Q38/S40 y N28/T30/V33/D38/S40.	R28/N30/K33/R38/Q40, Q28/N30/R33/R38/K40, K28/N30/S33/R38/E40, K28/N30/R33/E38/R40, Q28/N30/Y33/R38/K40, N28/N30/S33/R38/K40, S28/N30/R33/Q38/R40, S28/N30/Y33/Q38/K40, Q28/N30/E33/D38/H40, H28/N30/Y33/D38/S40, Q28/N30/Y33/E38/K40, Q28/N30/R33/A38/R40, K28/N30/T33/A38/A40, T28/N30/E33/S38/D40, K28/N30/R33/S38/S40, T28/N30/T33/Q38/R40, Q28/N30/R33/Q38/R40, K28/N30/T33/A38/S40, K28/E30/S33/R38/S40, K28/S30/R33/G38/S40, K28/D30/G33/H38/S40, K28/K30/H33/R38/S40, K28/R30/P33/G38/S40, K28/H30/H33/A38/S40, K28/H30M33/Q38/S40, K28/S30/H33/Q38/S40, K28/D30/R33/S38/S40, K28/Q30/H33/Q38/S40, K28/D30/R33/G38/S40, K28/G30/V33/A38/S40, K28/G30/V33/Q38/S40, K28/G30/V33/G38/S40,	Q28/N30/R33/R38/R40, K28/N30/T33/Q38/R40, Q28/N30/N33/Q38/K40, K28/N30/S33/R38/S40, Q28/N30/K33/R38/T40, N28/N30/S33/R38/R40, Q28/N30/R33/Q38/K40, T28/N30/R33/Q38/R40, R28/N30/Y33/N38/A40, Q28/N30/Y33/R38/A40, T28/N30/N33/Q38/R40, Q28/N30/N33/Q38/R40, K28/N30/R33/K38/A40, R28/N30/N33/Q38/D40, K28/N30/R33/Y38/A40, K28/N30/R33/Q38/Y40, A28/N30/R33/Q38/R40, K28/A30/R33/R38/S40, K28/H30/T33/P38/S40, K28/S30/H33/H38/S40, K28/R30/H33/G38/S40, K28/Q30/N33/Q38/S40, K28/R30/G33/N38/S40, K28/R30/G33/S38/S40, K28/A30/D33/H38/S40, K28/D30/A33/H38/S40, K28/A30/H33/Q38/S40, K28/N30/H33/G38/S40, K28/N30/H33/A38/S40, K28/S30/V33/Q38/S40, K28/G30/V33/T38/S40, R28/A30/V33/G38/S40,	Q28/N30/Y33/K38/K40, S28/N30/R33/S38/R40, S28/N30/Y33/R38/K40, R28/N30/R33/D38/R40, R28/N30/A33/Y38/Q40, Q28/N30/Y33/Q38/K40, E28/N30/R33/R38/K40, Q28/N30/T33/Q38/R40, Q28/N30/Y33/T38/R40, Q28/N30/Y33/A38/R40, Q28/N30/Y33/R38/S40, R28/N30/R33/E38/R40, R28/N30/A33/K38/S40, R28/N30/R33/Y38/Q40, A28/N30/N33/R38/K40, Q28/N30/S33/R38/K40, K28/N30/R33/Q38/Q40, K28/H30/H33/R38/S40, K28/G30/H33/Y38/S40, K28/N30/H33/R38/S40, K28/A30/N33/Q38/S40, K28/Q30/T33/Q38/S40, K28/N30/A33/Q38/S40, K28/S30/R33/Q38/S40, K28/S30/H33/R38/S40, K28/D30/A33/H38/S40, K28/A30/H33/Q38/S40, K28/N30/H33/G38/S40, K28/N30/H33/A38/S40, K28/D30/V33/H38/S40, K28/G30/V33/H38/S40, R28/D30/V33/R38/S40,
---	---	--	--

45 Dicha variante I-Crel es una variante capaz de escindir al menos una secuencia diana de ADN constituida por un sitio I-Crel mutante en el que al menos la a en la posición - 9 y/o la t en posición + 9 se han sustituido con un nucleótido diferente, teniendo dicha variante los residuos de aminoácidos en las posiciones 28, 30, 33, 38 y 40 respectivamente, que se seleccionan del grupo constituido por:

50 Q28/N30/Y33/K38/R40, Q28/N30/T33/Q38/K40, K28/N30/T33/R38/Q40, K28/N30/R33/E38/R40, R28/N30/A33/Y38/Q40, 55 K28/N30/N33/Q38/A40, R28/N30/T33/R38/A40, R28/N30/R33/E38/R40, K28/N30/R33/N38/A40, K28/N30/R33/A38/T40, 60 Q28/N30/S33/R38/K40, K28/N30/T33/A38/S40, K28/E30/S33/R38/S40,	R28/N30/K33/R38/Q40, Q28/N30/R33/R38/K40, K28/N30/S33/R38/E40, K28/N30/S33/R38/S40, N28/N30/S33/R38/K40, K28/N30/R33/T38/Q40, H28/N30/Y33/D38/S40, K28/N30/R33/A38/R40, R28/N30/R33/Y38/Q40, K28/N30/R33/N38/Q40, K28/N30/R33/A38/Q40, K28/A30/H33/R38/S40, K28/S30/R33/G38/S40,	Q28/N30/R33/R38/R40, K28/N30/T33/Q38/R40, S28/N30/Y33/R38/K40, R28/N30/R33/D38/R40, N28/N30/S33/R38/R40, K28/N30/R33/T38/R40, K28/N30/R33/Q38/R40, K28/N30/T33/A38/A40, K28/N30/R33/S38/S40, T28/N30/T33/Q38/R40, A28/N30/R33/Q38/R40, K28/H30/H33/R38/S40, K28/S30/H33/H38/S40,	Q28/N30/Y33/K38/K40, S28/N30/R33/S38/R40, K28/N30/S33/R38/D40, Q28/N30/Y33/R38/K40, E28/N30/R33/R38/K40, Q28/N30/E33/D38/H40, Q28/N30/R33/A38/R40, K28/N30/R33/K38/A40, A28/N30/N33/R38/K40, K28/N30/R33/Q38/Y40, K28/N30/R33/Q38/A40, K28/D30/N33/H38/S40, K28/N30/H33/R38/S40,
---	--	--	--

5	<p>K28/R30/R33/E38/S40, K28/Q30/N33/Q38/S40, K28/R30/G33/S38/S40, K28/N30/R33/A38/S40, K28/R30/G33/T38/S40, K28/D30/R33/G38/S40, K28/G30/V33/A38/S40, K28/G30/V33/R38/S40 y K28/G30/V33/G38/S40.</p>	<p>K28/D30/G33/H38/S40, K28/R30/G33/N38/S40, K28/T30/D33/H38/S40, K28/D30/A33/H38/S40, K28/N30/H33/S38/S40, K28/N30/H33/A38/S40, K28/D30/V33/H38/S40,</p>	<p>K28/D30/H33/K38/S40, K28/N30/H33/N38/S40, K28/A30/D33/H38/S40, K28/D30/R33/T38/S40, K28/N30/H33/G38/S40, K28/H30/M33/A38/S40, K28/G30/V33/T38/S340,</p>	<p>K28/K30/H33/R38/S40, K28/H30/H33/A38/S40, K28/S30/H33/R38/S40, K28/D30/R33/S38/S40, K28/N30/N33/Q38/S40, K28/S30/S33/H38/S40, K28/G30/V33/H38/S40,</p>
---	--	---	--	---

10 Dicha variante I-Crel puede comprender una o más mutación(es) adicionales.

Los residuos que están mutados pueden estar, de forma ventajosa, en las posiciones en contacto con la secuencia diana de ADN o que interactúan directa o indirectamente con dicha diana de ADN. Preferentemente, dichas mutaciones están en las posiciones seleccionadas del grupo constituido por: I24, Q26, S32, Q44, R68, R70, D75, I77 y T140. Preferentemente, dicha variante de I-Crel comprende una o más mutaciones seleccionadas del grupo constituido por:

- la mutación de isoleucina en la posición 24 en una valina (I24V),

20 - la mutación de la arginina en la posición 70 en una serina (IR70S), y

- la mutación del ácido aspártico en la posición 75, en un aminoácido sin carga, preferentemente una asparagina (D75N) o una valina (D75V).

25 Además, otros residuos se pueden mutar en toda la secuencia de I-Crel y, en particular, en la mitad C-terminal de I-Crel (posiciones 80 a 163). Las sustituciones en la mitad C-terminal de I-Crel están, preferentemente, en las posiciones: 80, 82, 85, 86, 87, 94, 96, 100, 103, 114, 115, 117, 125, 129, 131, 132, 147, 151, 153, 154, 155, 157, 159 y 160 de I-Crel.

30 Además, dicha variante puede incluir uno o más residuos insertados en el extremo NH₂ y/o el extremo COOH. Por ejemplo, un residuo de metionina se introduce en el extremo NH₂, un marcador (epítipo o secuencia de polihistidina) se introduce en el extremo NH₂ y/o el extremo COOH; dicho marcador es útil para la detección y/o purificación de dicha variante.

35 La variante I-Crel puede ser un homodímero o un heterodímero.

Dicha variante I-Crel también puede ser un heterodímero que comprende monómeros de dos variantes diferentes.

También se describe una endonucleasa quimérica monocatenaria que comprende un monómero de una variante I-Crel como se ha definido anteriormente.

40

También se describe un fragmento polinucleotídico que codifica una variante I-Crel o una endonucleasa quimérica monocatenaria derivada de dicha variante, como se ha definido anteriormente.

45 También se describe un vector recombinante que comprende al menos un fragmento polinucleotídico que codifica una variante o una endonucleasa quimérica monocatenaria derivada de dicha variante, como se ha definido anteriormente. Dicho vector puede comprender un fragmento polinucleotídico, un monómero de una variante homodimérica, dos monómeros o un monómero y un dominio de una molécula monocatenaria. Como alternativa, dicho vector puede comprender dos fragmentos polinucleotídicos, cada uno de los cuales codifica uno de los monómeros de una variante heterodimérica.

50

Un tipo de vector preferido es un episoma, es decir un ácido nucleico capaz de replicación extracromosómica. Vectores preferidos son aquellos capaces de replicarse de forma autónoma y/o de expresar los ácidos nucleicos a los que están unidos. Los vectores capaces de dirigir la expresión de los genes a los que están unidos operativamente se denominan en el presente documento "vectores de expresión".

55

Un vector comprende, entre otros, un YAC (cromosoma artificial de levadura), un BAC (artificial bacteriano), un vector baculovirus, un fago, un fagemido, un cósmido, un vector viral, 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 técnicas de ADN recombinante a menudo están en forma de "plásmidos", que, en general, se denomina bucles de ADN bicatenario circular que, en su forma de vector, no están unidos al cromosoma. Los expertos en la técnica conocen un número elevado de vectores adecuados.

60

Los vectores virales incluyen retrovirus, adenovirus, parvovirus (p. ej., virus adenoasociados), coronavirus, virus de ARN de hebra negativa tales como ortomixovirus (p. ej., el virus de la gripe), rhabdovirus (p. ej., virus de la rabia y de la estomatitis vesicular), paramixovirus (p. ej., del sarampión y Sendai), virus de ARN de hebra positiva tales como picornavirus y alfavirus, y virus de ADN bicatenario, incluidos adenovirus, herpesvirus (p. ej., virus del herpes simple de tipos 1 y 2, virus de Epstein-Barr, citomegalovirus), y poxvirus (p. ej., vaccinia, ciruela aviar y viruela del canario). Otros virus incluyen, por ejemplo, el virus Norwalk, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus y el virus de la hepatitis.

Los vectores pueden comprender marcadores seleccionables, por ejemplo: Neomicina fosfotransferasa, hisitidinol deshidrogenasa, dihidrofolato reductasa, higromicina fosfotransferasa, timidina quinasa del virus del herpes simple, adenosina desaminasa, glutamina sintetasa e hipoxantina-guanina fosforibosil transferasa para cultivo de células eucariotas, TRP1 para *S. cerevisiae*; resistencia a tetraciclina, rifampicina o ampicilina en *E. coli*.

Preferentemente, dichos vectores son vectores de expresión en los que la(s) secuencia(s) que codifican la variante de la invención se colocan bajo el control de los elementos de control de la transcripción y traducción adecuados para permitir la producción o síntesis de dicha variante. Por tanto, dicho polinucleótido está comprendido en el casete de expresión. Más particularmente, le vector comprende un origen de replicación, un promotor unido operativamente a dicho polinucleótido de codificación, un sitio de unión a ribosoma, un sitio de corte y empalme en ARN (cuando se usa ADN genómico), un sitio de poliadenilación y un sitio de terminación de la transcripción. También puede comprender un potenciador. La selección de promotor dependerá de la célula en la que se exprese el polipéptido. Preferentemente, cuando dicha variante es un heterodímero, los dos polinucleótidos que codifican cada uno de los monómeros se incluyen en un vector que es capaz de dirigir la expresión de ambos polinucleótidos de forma simultánea.

Dicho vector puede también incluir una construcción dirigida que comprende secuencias que comparten homologías con la región que rodea a la secuencia diana de ADN como se ha definido anteriormente.

Más preferentemente, dicha construcción de ADN dirigida comprende:

- a) secuencias que comparten homologías con la región que rodea a la secuencia diana de ADN como se ha definido anteriormente, y
- b) secuencias que se van a introducir flanqueadas por la secuencia como se define en a).

También se describe una célula huésped procariota o eucariota que se modifica mediante un polinucleótido o un vector como se ha definido anteriormente, preferentemente un vector de expresión.

También se describe un animal transgénico no humano o una planta transgénica, que se caracteriza porque todas o parte de sus células están modificadas por un polinucleótido o un vector como se ha definido anteriormente.

Como se usa en el presente documento, una célula se refiere a una célula procariota, tal como una célula bacteriana, o célula eucariota, tal como una célula de animal, planta o levadura.

También se describe una composición que comprende al menos una variante I-*Crel*, una endonucleasa química monocatenaria derivada de dicha variante, uno o dos polinucleótido(s), preferentemente incluidos en los vectores de expresión, como se ha definido anteriormente.

Dicha composición puede contener una construcción de ADN dirigida que comprende la secuencia que repara el sitio de interés flanqueado por secuencias que comparten homologías con el locus diana.

La(s) secuencia(s) polinucleotídicas que codifican la variante o la endonucleasa química monocatenaria derivada de dicha variante, como se ha definido anteriormente, se puede(n) preparar mediante cualquier procedimiento conocido por los expertos en la técnica. Por ejemplo, se amplifican a partir de un molde de ADNc mediante reacción en cadena de la polimerasa con cebadores específicos. Preferentemente, los codones de dicho ADNc se escogen para favorecer la expresión de dicha proteína en el sistema de expresión deseado.

El vector recombinante que comprende dichos polinucleótidos se puede obtener e introducir en una célula huésped mediante ADN recombinante bien conocido y técnicas de ingeniería genética.

La variante se produce expresando el(los) polipéptido(s) como se ha definido anteriormente; preferentemente dicho(s) polipéptido(s) se expresan o coexpresan en una célula huésped modificada por uno o dos vector(es) de expresión en condiciones adecuadas para la expresión o coexpresión de los polipéptidos y la variante se recupera

del cultivo de células huésped.

También se describe un procedimiento de ingeniería genética que comprende una etapa de rotura de ácido nucleico bicatenario en un sitio de interés localizado en un vector que comprende una diana de ADN como se ha definido anteriormente, poniendo en contacto dicho vector con una variante I-CreI o una endonucleasa quimérica monocatenaria que comprende dicha variante como se ha definido anteriormente, de modo que se induce una recombinación homóloga con otro vector que presenta homología con la secuencia que rodea el sitio de escisión de dicha variante.

10 También se describe un procedimiento de ingeniería genómica que comprende las etapas de: 1) rotura bicatenaria de un locus genómico que comprende al menos una secuencia diana de ADN como se ha definido anteriormente, poniendo en contacto dicha diana con una variante I-CreI o una endonucleasa quimérica monocatenaria que comprende dicha variante como se ha definido anteriormente; 2) mantener dicho locus genómico roto en condiciones adecuadas para la recombinación homóloga con una construcción de ADN dirigida que comprende la secuencia que se va a introducir en dicho locus, flanqueado por secuencias que comparten homologías con el locus diana.

También se describe un procedimiento de ingeniería genómica caracterizado porque comprende las etapas siguientes: 1) rotura bicatenaria de un locus genómico que comprende al menos una secuencia diana de ADN como se ha definido anteriormente, poniendo en contacto dicho sitio de escisión con una variante I-CreI o una endonucleasa quimérica monocatenaria que comprende dicha variante como se ha definido anteriormente; 2) mantener dicho locus genómico roto en condiciones adecuadas para la recombinación homóloga con ADN cromosómica que comparte homologías con las regiones que rodean el sitio de escisión.

25 También se describe el uso de una variante de endonucleasa I-CreI que se puede obtener mediante el procedimiento como se ha descrito anteriormente para biología molecular, ingeniería genética *in vivo* o *in vitro* e ingeniería genómica *in vivo* o *in vitro* para fines no terapéuticos, para escindir una secuencia diana de ADN como se ha definido anteriormente.

30 Biología molecular incluye, sin limitaciones, restricción de ADN y mapeo de ADN. Ingeniería genética y genómica para fines no terapéuticos incluyen, por ejemplo, (i) genes diana de loci específicos en líneas de empaquetamiento celular para producción de proteínas, (ii) genes diana de loci específicos en plantas de cosechas para mejorar cepas e ingeniería metabólica, (iii) recombinación dirigida para la eliminación de marcadores en cosechas modificadas genéticamente, (iv) recombinación dirigida para la eliminación de marcadores en cepas de microorganismos modificadas genéticamente (para, por ejemplo, la producción de antibióticos).

Dicho uso puede ser para inducir una rotura bicatenaria en un sitio de interés que comprende una secuencia diana de ADN escindida por una variante como se ha definido anteriormente, de modo que se induce un acontecimiento de recombinación de ADN, una pérdida de ADN o muerte celular.

40 En una realización concreta, se usa una variante de I-CreI que tiene una arginina (R) o una lisina (K) en la posición 38 para la escisión de una diana de ADN que comprende una guanina en la posición - 9 o una citosina en la posición + 9.

45 Dicha rotura bicatenaria es para: reparar una secuencia específica, modificar una secuencia específica, restablecer un gen funcional en lugar de uno mutado, atenuar o activar un gen endógeno de interés, introducir una mutación en un sitio de interés, introducir un gen exógeno o una parte del mismo, inactivar o detectar un gen endógeno o una parte del mismo, translocar un brazo cromosómico o dejar el ADN sin reparar y degradado.

50 También se describe el uso de al menos una variante de I-CreI como se ha definido anteriormente para la preparación de un medicamento para prevenir, mejorar o curar una enfermedad genética en un individuo que lo necesite, administrándose dicho medicamento por cualquier medio a dicho individuo.

También se describe un procedimiento para prevenir, mejorar o curar una enfermedad genética en un individuo que lo necesite, comprendiendo dicho procedimiento al menos la etapa de administrar a dicho individuo una composición como se ha definido anteriormente por cualquier medio.

60 En una realización concreta, se usa una variante de I-CreI que tiene una arginina (R) o una lisina (K) en la posición 38 para la escisión de una diana de ADN genómico que comprende una guanina en la posición - 9 o una citosina en la posición + 9.

También se describe el uso de al menos una variante de I-CreI como se ha definido anteriormente para la

preparación de un medicamento para prevenir, mejorar o curar una enfermedad causada por un agente infeccioso que presenta un intermedio de ADN, en un individuo que lo necesite, administrándose dicho medicamento por cualquier medio a dicho individuo.

5 También se describe un procedimiento para prevenir, mejorar o curar una enfermedad causada por un agente infeccioso que presenta un intermedio de ADN, en un individuo que lo necesite, comprendiendo dicho procedimiento al menos la etapa de administrar a dicho individuo una composición como se ha definido anteriormente por cualquier medio.

10 También se describe el uso de al menos una variante de I-Crel como se ha definido anteriormente, *in vitro*, para inhibir la propagación, inactivar o eliminar un agente infeccioso que presenta un ADN intermedio, en productos biológicos derivados o productos destinados para usos biológicos o para desinfectar un objeto.

15 También se describe un procedimiento para descontaminar un producto o material de un agente infeccioso que presenta un ADN intermedio, comprendiendo dicho procedimiento al menos la etapa de poner en contacto un producto derivado biológico, un producto destinados para usos biológicos o un objeto, con una composición como se ha definido anteriormente, durante un tiempo suficiente para inhibir la propagación, inactivar o eliminar dicho agente infeccioso.

20 En una realización concreta, se usa una variante de I-Crel que tiene una arginina (R) o una lisina (K) en la posición 38 para la escisión de una diana de ADN de dicho agente infeccioso que comprende una guanina en la posición - 9 o una citosina en la posición + 9.

En otra realización concreta, dicho agente infeccioso es un virus. Por ejemplo, dicho virus es un adenovirus (Ad11, Ad21), herpesvirus (HSV, VZV, EBV, CMV, herpesvirus 6, 7 o 8), hepadnavirus (HBV), papovavirus (HPV), poxvirus o retrovirus (HTLV, HIV).

25 También se describe el uso de al menos una variante I-Crel, como se ha definido anteriormente, como armazí para fabricar otras meganucleasas. Por ejemplo, una segunda ronda de mutagénesis y detección selectiva se puede realizar con dicha variante de I-Crel, con el fin de fabricar nuevas endonucleasas domésticas de segunda generación.

De acuerdo con otra realización ventajosa de dichos usos, dicha variante de I-Crel se asocia con una construcción de ADN dirigida como se ha definido anteriormente.

35 De acuerdo con otra realización ventajosa de dichos usos, dicha variante I-Crel tiene residuos de aminoácidos en las posiciones 28, 30, 33, 38 y 40 respectivamente, que se seleccionan del grupo constituido por: KNSQS, KNRQS, KNTQS, KNHQS.

40 El uso de la variante I-Crel meganucleasa y los procedimientos de usar dicha variante I-Crel meganucleasa incluyen también el uso de la endonucleasa química monocatenaria derivada de dicha variante, en el que el polinucleótido(s), vector, célula, planta transgénica o mamífero transgénico no humano codifica dicha variante o endonucleasa química monocatenaria, como se ha definido anteriormente.

Otras características que aparecerán a partir de la descripción que sigue, que se refiere a ejemplos que ilustran las variantes de endonucleasa doméstica I-Crel y a sus usos, así como a las figuras adjuntas en las que:

50 - la figura 1 representa dianas de ADN. A. Dos dianas de I-Crel palindrómicas derivadas del sitio doméstico natural I-Crel. La diana natural I-Crel contiene dos palíndromos, en recuadros en gris: los nucleótidos - 8 a - 12 y + 8 a + 12 por un lado y los nucleótidos - 5 a - 3 y + 3 a + 5 por otro lado. De la diana natural, denominada C1234 (SEQ ID NO: 1) se puede derivar las secuencias palindrómicas C1221 y C4334 (SEQ ID NO: 2, 3). Ambas se cortan con I-Crel, *in vitro* y en levadura. B. Las 64 dianas de ADN. Las 64 dianas derivan de C1221 (SEQ ID NO: 4 a 67). Corresponden a los palíndromos perfectos de 24 pb resultantes de las sustituciones en las posiciones - 10, - 9, - 8, + 8, + 9 y + 10.

55 - la figura 2 ilustra el principio del ensayo de detección selectiva en levaduras. Una cepa que expresa la meganucleasa que se va a someter a ensayo (MEGA), marcada con el gen LEU2, se aparea con una cepa que elija un plásmido indicador que contiene la diana elegida, marcada por el gen TRP1. La diana está flanqueada por genes LacZ truncados solapantes (LAC y ACZ). En diploides (LEU2 TRP1), la escisión del sitio diana mediante la meganucleasa induce recombinación homóloga entre las dos repeticiones lacZ, resultante en un gen de beta-galactosidasa funcional, que se puede monitorizar mediante tinción con X-gal.

- la figura 3 representa las secuencias de la proteína armazón I-Crel N75 y los cebadores degenerados usados para la construcción de las bibliotecas Ulib4 y Ulib5. A. El armazón (SEQ ID NO: 68) es el ORF de I-Crel que incluye la sustitución del codón D75N y tres codones adicionales (AAD) en el extremo 3'. B. Cebadores (SEQ ID NO: 69, 70, 71),

5

- la figura 4 representa el vector de expresión de meganucleasa pCLS0542. pCLS0542 es un vector de replicación de micrómetros marcado con un gen auxótrofo de LEU2 y un promotor Gal10 inducible para dirigir la expresión de las variantes I-Crel,

10 - la figura 5 representa el vector indicador pCLS0042. El vector indicador está marcado con TRP1 y URA3. Las repeticiones en tándem de LacZ comparten 800 pb de homología y están separados por 1,3 kb de ADN. Están rodeados por el promotor de ADH y secuencias de terminación. Los sitios diana se clonan en el sitio *Sma*I.

- la figura 6 ilustra el fundamento de las bibliotecas de variantes I-Crel. A. Estructura del homodímero de I-Crel unido a su diana de ADN, de acuerdo con Chevalier y col., J. Mol. Biol., 2003, 329, 253 – 269, y localización del área de la interfaz de unión elegida para aleatorización en este estudio; los residuos 28, 30, 33, 38 y 40 están marcados en negro en el monómero de la izquierda. B. Zoom que muestra los residuos 28, 30, 33, 38 y 40 elegidos para aleatorización. C. Resumen de la interacción I-Crel/ADN en la región externa de la diana de ADN de I-Crel (en negro en la figura 6A). La diana representada, C1221 (SEQ ID NO: 3) es una diana palindrómica escindida por Crel (Chevalier y col., 2003, citado previamente). Solo se indican los contactos específicos de base. La región 10NNN ($\pm 8, \pm 9, \pm 10$ nucleótidos, en negro en la Figura 6A) de la diana está en un recuadro. D. Posición de los residuos aleatorizados en Ulib4 (30, 33, 38). E. Posición de los residuos aleatorizados en Ulib5 (28, 30, 38). F. Posición de los residuos aleatorizados en Lib4 (28, 33, 38, 40).

- la figura 7 representa los patrones de escisión de las 141 variantes de I-Crel que escinden 37 nuevas dianas de ADN. Para cada una de las variantes de I-Crel obtenidas tras la detección selectiva y definidas por los residuos en las posiciones 28, 30, 33, 38, 40, 70 y 75, la escisión se monitorizó en levaduras con las 64 dianas descritas en la Figura 1B. Las dianas se designan con tres letras correspondientes a los nucleótidos en las posiciones - 10, - 9 y - 8. Por ejemplo, GGG corresponde a la diana tcgggacgtcgtacgacgtcccga (SEQ ID NO: 4; véase la Figura 1). Los valores corresponden a la intensidad de la escisión, evaluados mediante un software adecuado tras el escaneo del filtro.

- la figura 8 ilustra ejemplos de patrones y los números de mutantes que escinden cada diana. A. Ejemplos de perfiles. Cada nueva endonucleasa se perfila en levaduras en una serie de 64 dianas palindrómicas descritas en la Figura 1B, que difieren de la secuencia mostrada en la figura 1A en las posiciones $\pm 8, \pm 9$ y ± 10 . Estas dianas están dispuestas como en la Figura 8B. Cada secuencia diana se denomina por el triplete - 10, - 9, - 8 (10NNN). Por ejemplo, GGG corresponde a la diana tcgggacgtcgtacgacgtcccga (SEQ ID NO: 4; véase la Figura 1B). Las meganucleasas se prueban 4 veces contra las 64 dianas. Las dianas escindidas por -Crel (D75), I-Crel N75 o diez variantes derivadas se visualizan como puntos negros o grises. B. Los números de mutantes que escinden cada diana y la intensidad media de la escisión. Cada secuencia se denomina por el triplete - 10, - 9, - 8 (10NNN). El número de proteínas que escinden cada diana se muestra más adelante y el nivel de coloración en gris es proporcional a la intensidad media de la señal obtenida con estos cortadores en levaduras.

Ejemplo 1: Endonucleasas Funcionales con nueva especificidad hacia los nucleótidos $\pm 8, \pm 9$ y ± 10 (10NNN)

45 El procedimiento de producir variantes meganucleasas y los ensayos basados en recombinación inducida por escisión en células de levadura, que se usan para detección selectiva de variantes con especificidad alterada, se describen en la solicitud de patente internacional PCT WO 2004/067736 y Epinat y col., N.A.R., 2003, 31, 2952 – 2962. Estos ensayos tienen como resultado un gen indicador LacZ funcional que se puede monitorizar mediante procedimientos estándar (Figura 2).

50

A) Materiales y procedimientos

a) Construcción de bibliotecas mutantes

55 Se sintetizaron los marcos de lectura abiertos de I-Crel wt (I-Crel D75), I-Crel D75N (I-Crel N75) y I-Crel S70 N75 como se ha descrito anteriormente (Epinat y col., N.A.R., 2003, 31, 2952 – 2962). Las bibliotecas combinatorias derivaron del armazón I-Crel N75, I-Crel D75 o I-Crel S70 N75 sustituyendo dos o tres combinaciones diferentes de residuos, potencialmente implicados en las interacciones con las bases en las posiciones ± 8 a 10 de un semisito diana de ADN. La diversidad de las bibliotecas de meganucleasas se generó mediante PCR usando cebadores degenerados que alojan un único codón degenerado en cada una de las posiciones seleccionadas.

60

La mutación D75N se introdujo sustituyendo el codón 75 con aac. Después, tres codones en las posiciones N30, Y33 y Q38 (biblioteca Ulib4) o K28, N30 y Q38 (biblioteca Ulib5) se sustituyeron con un codón degenerado VVK (18 codones) que codifica 12 aminoácidos diferentes. A,D,E,G,H,K,N,P,Q,R,S,T). En consecuencia, la diversidad máxima (teórica) de estas bibliotecas de proteínas fue 12^3 o 1728. No obstante, en términos de ácidos nucleicos, la 5 diversidad fue 18^3 o 5832.

En Lib4, solicitada de BIOMETHODES, una arginina en la posición 70 del armazón I-Crel N75 se sustituyó primero con una serina (R70S). Después se aleatorizaron las posiciones 28, 33, 38, 40. Los aminoácidos regulares (K28, Y33, Q38 y S40) fueron sustituidos con uno de 10 aminoácidos (A,D,E,K,N,Q,R,S,T,Y). La biblioteca resultante tiene 10 una complejidad teórica de 10.000 en términos de proteínas.

Además, las bibliotecas pequeñas de complejidad 225 (15^2) resultantes de la aleatorización de solo dos posiciones se construyeron en un armazón -Crel N75 o I-Crel D75 usando el codón degenerado NVK (24 codones, aminoácidos ACDEGHKNPQRSTWY).

15 Los fragmentos portadores de combinaciones de las mutaciones deseadas se obtuvieron mediante PCR usando un par de cebadores degenerados que codifican 10, 12 o 15 aminoácidos diferentes, y como molde de ADN, los marcos de lectura abierta (ORF) I-Crel N75 (Figure 3A), I-Crel D75 o I-Crel S70 N75. Por ejemplo, la figura 3B ilustra los dos pares de cebadores (Ulib456for y Ulib4rev; Ulib456for y Ulib5rev) usados para generar las bibliotecas Ulib4 y Ulib5, 20 respectivamente. Los correspondientes productos de PCR se clonaron de nuevo en el ORF de I-Crel N75 o I-Crel D75 en el vector de expresión con replicación de levaduras pCLS0542 (Epinat y col., citado anteriormente, Figura 4), portador de un gen marcador auxótrofo de LEU2. En este vector de replicación basado en micrómetros, las variantes I-Crel se encuentran bajo el control del promotor inducible de la galactosa.

25 b) Construcción de clones diana

El palíndromo de 24 pb C1221 (tcaaaaacgtcgtacgacgttttga, SEQ ID NO: 3) es una repetición del semisitio de la diana de I-Crel natural casi palindrómica (tcaaaaacgtcgtgagacagtttgg, SEQ ID NO: 1). C1221 se escinde con tanta eficiencia como la diana natural de I-Crel *in vitro* y *ex vivo* tanto en células de levadura como de mamífero. Las 64 dianas 30 palindrómicas se derivaron del siguiente modo: 64 pares de oligonucleótidos (ggcatacaagttcnnnacgtcgtacgacgttngacaatcgtctgtca (SEQ ID NO: 72) y secuencias complementarias inversas) se solicitaron de Sigma, hibridaron y clonaron en pGEM-T Easy (PROMEGA) en la misma orientación. Después, un fragmento PvuII de 400 pb se escindió y clonó en el vector de levaduras pFL39-ADHLACURAZ, también denominado pCLS0042, descrito anteriormente (Epinat et al citado previamente, Figura 5) que tiene como resultado 35 64 vectores indicadores de levadura (plásmidos diana).

c) Cepas de levadura

Las tres bibliotecas de variantes de expresión de meganucleasas se transformaron en la cepa de levadura haploide 40 mutante leu2 FYC2-6A: *MATalpha*, *trp1* Δ 63, *leu2* Δ 1, *his3* Δ 200. Para la transformación se usó un protocolo de sustancia química / shock térmico clásico que habitualmente nos da 10^6 transformantes independientes por μ g de ADN derivado de (Gietz and Woods, Methods Enzymol., 2002, 350, 87 – 96). Los clones transformantes individuales (Leu⁺) se escogieron individualmente en microplacas de 96 pocillos. Los 65 plásmidos diana se transformaron usando el mismo protocolo en la cepa de levadura haploide FYBL2-7B: *MATa*, *ura3* Δ 851, *trp1* Δ 63, *leu2* Δ 1, *lys2* 45 Δ 202, resultante en 64 cepas probadoras.

d) Apareamiento de clones que expresan meganucleasa y detección selectiva en levaduras

Los clones que expresan meganucleasa se aparearon con cada una de las 64 cepas diana y los diploides se 50 analizaron para determinar la actividad beta-galactosidasa usando el ensayo de detección selectiva ilustrado en la figura 2.

Los clones de variantes de I-Crel así como las cepas indicadoras de levadura se almacenaron en glicerol (al 20 %) y se replicaron en nuevas microplacas. El apareamiento se realizó usando un filtrador de colonias (QpixII, GENETIX). 55 Los mutantes se vieron en los filtros de nylon que cubren las placas YPD usando una densidad elevada (aproximadamente 20 puntos/cm²). Un segundo proceso se realizó con los mismos filtros para ver una segunda capa consistente en 64 cepas de levadura que alojan indicadores diferentes para cada variante. Las membranas se colocaron en un medio sólido rico en YEPD con agarosa y se incubaron a 30 °C durante una noche para permitir el apareamiento. Después, los filtros se transfirieron a medio sintético sin leucina y triptófano, con galactosa (2 %) 60 como fuente de carbono (y con G418 para los experimentos de coexpresión) y se incubaron durante cinco días a 37 °C para seleccionar diploides, permitir la expresión de meganucleasas, escisión y recombinación del plásmido indicador y expresión de beta-galactosidasa. Tras 5 días, los filtros se colocaron en medio sólido con agarosa con

0,02 % de X-Gal en tampón fosfato sódico 0,5 M, pH 7,0, 0,1 % de SDS, 6 % de dimetil formamida (DMF), β -mercaptoetanol 7 mM, 1 % de agarosa y se incubaron a 37 °C para monitorizar la actividad de β -galactosidasa. Tras dos días de incubación, los clones positivos se identificaron mediante escaneo. La actividad de β -galactosidasa de los clones se cuantificó usando un software adecuado.

5

Los clones que muestran una actividad contra al menos una diana se aislaron (primera detección selectiva). La densidad de mancha se redujo después a 4 manchas/cm² y cada clon positivo se analizó contra las 64 cepas indicadoras por cuadruplicado, creando de este modo perfiles completos (detección selectiva secundaria).

10 e) Secuencia

El marco de lectura abierto (ORF) de los clones positivos identificados durante la primera detección selectiva y/o la secundaria en lavaduras se amplificó mediante PCR con colonias de levadura usando cebadores: PCR-Gal10-F (gcaacttagtgctgacacatacagg, SEQ ID NO: 73) y PCR-Gal10-R (acaaccttgattgcagacttgacc, SEQ ID NO: 74) de PROLIGO. En resumen, la colonia de levaduras se escoge y resuspende en 100 μ l de medio líquido con L-Glu y se cultiva durante la noche. Tras la centrifugación, el sedimento con la levadura se resuspende en 10 μ l de agua estéril y se usa para la reacción de PCR en un volumen final de 50 μ l que contiene 1,5 μ l de cada cebador específico (100 pmol/ μ l). Las condiciones de PCR fueron un ciclo de desnaturalización durante 10 minutos a 94 °C, 35 ciclos de desnaturalización durante 30 s a 94 °C, hibridación durante 1 min a 55 °C, extensión durante 1,5 min a 72 °C y una extensión final de 5 minutos. A continuación, los productos resultantes de la PCR se secuenciaron.

20

f) Análisis de la estructura

Todos los análisis de las estructuras proteicas se realizaron usando PyMOL. Las estructuras de I-Crel corresponden a la entrada en pdb1g9y. La numeración de residuos en el texto siempre se refiere a estas estructuras, excepto por los residuos en el segundo dominio proteico de I-Crel del homodimero en el que los números de residuos se fijan como para el primero dominio.

25

B) Resultados

30

I-Crel es una endonucleasa doméstica dimérica que esconde una diana pseudopalindrómica de 22 pb. El análisis de la estructura de I-Crel unida a su diana natural ha demostrado que en cada monómero, ocho residuos establecieron interacciones directas con siete bases (Jurica y col., 1998, citado previamente). De acuerdo con estos datos estructurales, las bases de los nucleótidos en las posiciones ± 8 a 10 establecen contactos directos con los aminoácidos de I-Crel N30, Y33, Q38 y los contactos indirectos con los aminoácidos de I-Crel K28 and S40 (Figura 6A, 6B, 6C). Por tanto, las nuevas proteínas con mutaciones en las posiciones 30, 33 and 38 podrían mostrar nuevos perfiles de escisión con las 64 dianas resultantes de sustituciones en las posiciones ± 8 , ± 9 y ± 10 de una diana palindrómica escindida por I-Crel (diana 10NNN). Además, las mutaciones podrían alterar el número y las posiciones de los residuos implicados en contacto directo con las bases del ADN. Más específicamente, posiciones distintas a 30, 33, 38, pero localizadas en las cercanías de la proteína clonada, podrían estar implicadas en la interacción con los mismos pares de bases.

35

40

Una biblioteca exhaustiva de proteínas frente a un abordaje de bibliotecas diana se realizó para efectuar ingeniería localmente en esta parte de la interfaz de unión al ADN. La aleatorización de 5 posiciones de aminoácidos conduciría a una diversidad teórica de $20^5 = 3,2 \times 10^6$. No obstante, las bibliotecas con menor diversidad se generaron aleatorizando 2, 3 o 4 residuos a la vez, lo que tiene como resultado una diversidad de 225 (15^2), 1728 (12^3) o 10.000 (10^4). Esta estrategia ha permitido una extensa detección selectiva de cada una de estas bibliotecas contra las 64 dianas palindrómicas de ADN 10NNN usando un ensayo basado en levaduras descrito anteriormente (Epinat y col., 2003, citado previamente y la solicitud de PCR internacional WO 2004/067736) y cuyo principio se describe en la Figura 2.

50

En primer lugar se mutó el armazón de I-Crel de D75 a N. La mutación D75N no afectó a la estructura de la proteína pero disminuyó la toxicidad de I-Crel en los experimentos de sobreexpresión.

55

Después se construyó la biblioteca Ulib4: los residuos 30, 33 y 38 (Figura 6A – C y 6D) se aleatorizaron y los aminoácidos regulares (N30, Y33 y Q38) fueron sustituidos con uno de 12 aminoácidos (A,D,E,G,H,K,N,P,Q,R,S,T). La biblioteca resultante tiene una complejidad de 1728 en términos de proteínas (5832 en términos de ácidos nucleicos).

60

Se construyeron otras dos bibliotecas: Ulib5 y Lib4. En Ulib5, los residuos 28, 30 y 38 (Figura 6A – C y 6D) se aleatorizaron y los aminoácidos regulares (K28, N30 y Q38) fueron sustituidos con uno de 12 aminoácidos (ADEGHKPNQRST). La biblioteca resultante tiene una complejidad de 1728 en términos de proteínas (5832 en términos de ácidos nucleicos). En Lib4, la arginina en la posición 70 se sustituyó primero con una serina. Después,

las posiciones 28, 33, 38 y 40 (Figura 6A – C y 6F) se aleatorizaron y los aminoácidos regulares (K28, Y33, Q38 y S40) fueron sustituidos con uno de 10 aminoácidos (A,D,E,K,N,Q,R,S,T,Y). La biblioteca resultante tiene una complejidad de 10.000 en términos de proteínas.

5 En un experimento de detección selectiva primaria, 20.000 clones de Ulib4, 10.000 clones de Ulib5 y 20.000 clones de Lib4 se aparearon con cada una de las 64 cepas probadoras y las diploides se analizaron para determinar la actividad beta-galactosidasa. Todos los clones que mostraron actividad de escisión con al menos una de las 64 dianas se analizaron en una segunda ronda de detección selectiva contra las 64 dianas, por cuadruplicado, y se estableció cada perfil de escisión, como se muestra en la Figura 8. Después, el ORF de la meganucleasa se
10 amplió de cada cepa mediante PCR y se secuenció.

Tras la detección selectiva secundaria y la secuenciación de las positivas sobre toda la región de codificación, se aisló un total de 1484 mutantes únicos que mostraban actividad de escisión contra al menos una diana. Se pudieron observar diferentes patrones. La Figura 7 ilustra 37 dianas nuevas escindidas mediante una colección de 141
15 variantes, incluidas 34 dianas que no son escindidas por I-Crel y 3 dianas que son escindidas por I-Crel (aag, aat y aac). En la Figura 8A se muestran doce ejemplos de perfil, incluidos I-Crel N75 y I-Crel D75. Algunos de estos nuevos perfiles comparten alguna similitud con el armazón silvestre, mientras que muchos otros eran totalmente diferentes. Las endonucleasas domésticas normalmente pueden acomodar alguna degeneración en sus secuencias diana y las propias proteínas I-Crel y I-Crel N75 escinden una serie de dieciséis y tres dianas, respectivamente. La
20 degeneración de la escisión se encontró para muchas de las endonucleasas nuevas, con una media de 9,9 dianas escindidas por mutante (desviación estándar: 11). No obstante, entre los 1484 mutantes identificados, se encontró que 219 (15 %) solo escinden una diana de ADN, 179 (12 %) escinden dos y 169 (11 %) y 120 (8 %) pudieron escindir 3 y 4 dianas, respectivamente. Por tanto, con independencia de su diana preferida, un número significativo de derivados de I-Crel muestran un nivel de especificidad que es similar, si no mayor, al del mutante I-Crel N75
25 (tres secuencias diana 10NNN escindidas) o I-Crel (dieciséis secuencias diana 10NNN escindidas). Asimismo, la mayoría de los mutantes aislados para especificidad alterada por las secuencias 10NNNN ya no escinde la secuencia diana C1221 original descrita en la Figura 6C (61 % y 59 %, respectivamente).

En conjunto, esta gran colección de mutantes permitió dirigir todas las 64 posibles secuencias de ADN diferentes en
30 las posiciones ± 10 , ± 9 y ± 8 (Figura 8B). No obstante se produjeron enormes variaciones en el número de mutantes que escinde cada diana (Figura 8B), estas cifras variaron de 3 a 396, con una media de 228,5 (desviación estándar: 201.5). La escisión con frecuencia se observó para las dianas con una guanina en ± 8 o una adenina en ± 9 , mientras que una citosina en ± 10 o ± 8 se correlacionó con un número bajo de escisores. Además, no todas las dianas se escindieron con la misma eficiencia. Dado que se pudieron observar variaciones significativas de la señal
35 para una misma diana, en función del mutante (comparar las eficiencias de escisión para la diana 10AAA silvestre en la Figura 8B, por ejemplo), se midió una eficiencia media de escisión para cada diana como se ha indicado anteriormente (Arnould y col., J. Mol. Biol., 2006, 355, 443 – 458). Estas eficiencias medias se representan con niveles de grises en la Figura 8B. El análisis de los resultados muestra una clara correlación entre esta eficiencia media y el número de escisores, siendo la diana de corte más frecuente también el corte más eficiente (compárese,
40 por ejemplo, las dianas 10TCN, 10CTN y 10CCN con 10GAN, 10AAN y 10TAN en la Figura 8B).

Por tanto, se obtuvieron cientos de nuevas variantes, incluidos mutantes con nueva especificidad de sustrato; estas variantes pueden conservar niveles altos de actividad y la especificidad de las proteínas nuevas puede ser más estrecha que la de la proteína silvestre por su diana.
45

Ejemplo 2: Análisis estadístico de las interacciones entre variantes de I-Crel y sus dianas

A) Materiales y procedimientos

50 Se usó la agrupación jerárquica para establecer las posibles correlaciones entre residuos proteicos específicos y las bases diana, como se ha descrito anteriormente (Arnould y col., J. Mol. Biol., 2006, 355, 443 – 458). El agrupamiento se realizó con los datos cuantitativos de la detección selectiva secundaria usando hclust del paquete R. Las variantes se agruparon usando agrupación jerárquica estándar con la distancia euclídea y el método de Ward (Ward, J.H., American Statist. Assoc., 1963, 58, 236 – 244). El dendograma de la mutante se cortó a la altura de 17 para
55 definir los grupos. Para el análisis, las intensidades acumuladas de la escisión de una diana dentro de un grupo se calcularon como la suma de las intensidades de escisión de todos los mutantes del grupo con esta diana, normalizadas a la suma de las intensidades de escisión de todos los mutantes del grupo con todas las dianas.

B) Resultados

60 Se identificaron diez grupos mutantes diferentes (Tabla 1).

ES 2 393 508 T3

Tabla 1: Análisis de grupos

grupo (effectif)	dianas preferidas		nucleótido			aminoácidos preferidos (%)							
	10NNN	(%)	-10 (%)	-9 (%)	-8 (%)	28	30	33	38	40			
1 (44)	GGG	7,1	A	27,0	A	45,9	A	21,3	K 100	N45,5	H	Q	S
	GAG	6,9	C	4,7	C	14,5	C	14,7	100	45,5	38,6	70,5	86,4
	GAT	6,4	G	63,0	G	27,9	G	37,0		K	R		
		S=20.4	T	5,3	T	11,8	T	27,0		15,9	25,0		
										R			
										15,9			
2 (82)	AAG	6,1	A	33,4	A	52,2	A	20,4	K	N	G	Q	S
	TAG	5,6	C	11,7	C	9,4	C	13,8	100	64,6	23,2	68,3	79,3
	GAG	5,2	G	23,7	G	19,9	G	41,7					
		S=16.9	T	31,2	T	18,5	T	24,0					
3 (36)	TAG	4,5	A	24,7	A	45,2	A	19,5	K	N	T	Q	S
	TAC	4,4	C	13,9	C	6,7	C	16,2	100	75,0	61,1	86,1	75,0
	TGG	4,3	G	15,8	G	26,9	G	37,6			C		
		S=13.2	T	45,6	T	21,3	T	26,7			22,2		
4 (74)	GGG	30,6	A	33,1	A	22,6	A	10,2	K	N	R	R	S
	AGG	15,0	C	2,6	C	1,3	C	1,5	93,2	82,4	17,6	26,0	83,6
	AAG	7,6	G	56,3	G	66,2	G	77,7			Y	K	
		S=53.2	T	8,1	T	9,9	T	10,6			16,2	19,2	
5 (115)	GAG	12,0	A	30,0	A	71,6	A	20,6	K	N	R	Q	S
	GAT	11,8	C	5,1	C	5,1	C	15,5	98,3	64,4	23,5	94,7	66,1
	GAA	8,8	G	58,5	G	18,4	G	35,6			H		
		S=32.6	T	6,4	T	5,0	T	28,2			20,9		
										Y			
										19,1			
6 (110)	AAG	9,1	A	40,6	A	59,9	A	15,7	K	N	P	Q	S
	TAG	8,7	C	9,6	C	11,0	C	12,2	100	87,3	22	68,8	61,5
	GAG	8,0	G	23,6	G	15,7	G	49,0					
		S=25.8	T	26,2	T	13,4	T	23,2					
7	AAT	23,7	A	74,9	A	85,2	A	25,0	K	N	Y	Q	S

(continuación)

grupo (effectif)	dianas preferidas		nucleótido			aminoácidos preferidos (%)							
	10NNN	(%)	-10 (%)	-9 (%)	-8 (%)	28	30	33	38	40			
(106)	AAA	16,8	C	17,2	C	2,5	C	11,3	86,8	41,5	92,5	95,2	70,5
	AAG	16,6	G	5,9	G	11,5	G	30,0		T			
		S=57.1	T	2,0	T	0,8	T	33,7		24,5			
8 (384)	GGG	14,0	A	35,5	A	41,2	A	12,6	K	N	Y	Q	S
	TAG	10,2	C	12,4	C	9,1	C	12,2	89,8	63,3	45,8	43,4	62,9
	AAT	6,9	G	30,7	G	37,3	G	52,4					
		S=31.1	T	21,4	T	12,4	T	22,8					
9 (134)	TAG	17,6	A	21,1	A	62,3	A	12,0	K	N	C	Q	S
	TAT	9,9	C	13,0	C	2,2	C	13,9	92,5	76,1	17,9	74,6	72,7
	AAG	7,5	G	9,6	G	22,1	G	51,3			S		
		S=35.0	T	56,3	T	13,4	T	22,7			16,4		
10 (399)	AAG	20,2	A	64,3	A	78,9	A	18,1	K	N	Y	Q	S
	AAT	14,7	C	5,4	C	6,7	C	10,6	96,0	59,5	53,6	69,2	70,1
	AAA	10,7	G	25,2	G	10,4	G	44,8					
		S=45.6	T	5,1	T	4,0	T	26,5					

¹ Frecuencias de la diana y la base corresponden a la intensidad acumulada de la escisión como se describe en Materiales y Procedimientos)

² En cada posición, se indican los residuos presentes en más del 15 % de los grupos.

El análisis de los residuos encontrados en cada grupo mostró sesgos fuertes para todas las posiciones 5 aleatorizadas. Ninguno de los residuos está mutado en todas las bibliotecas usadas en este estudio y cabe esperar

que los residuos encontrados en el armazón I-Crel estén sobrerrepresentados. De hecho, K28, N30 y S40 fueron los residuos más frecuentes en los 10 grupos y realmente no se puede deducir ninguna conclusión para interacciones de ADN / proteínas. No obstante, Y33 fue el residuo más representado únicamente en los grupos 7, 8 y 10, mientras que una fuerte aparición de otros residuos, tales como H, R, G, T, C, P o S, se observó en los otros siete grupos. El residuo Q38 silvestre se sobrerrepresentó en todos los grupos menos en uno, siendo R y K más frecuentes en el grupo 4.

Entre tanto, se observaron fuertes correlaciones entre la naturaleza de los residuos 33 y 38 y la discriminación por sustrato en las posiciones ± 10 y ± 9 de la diana.

10

La prevalencia de Y33 se asoció con frecuencias elevadas de adenina (74.9 % and 64.3 % en los grupos 7 y 10, respectivamente) y esta correlación también se observó, aunque en menor medida, en los grupos 4, 5 y 8. H33 o R33 se correlacionaron con una guanina (63,0 %, 56,3 % y 58,5 %, en los grupos 1, 4 y 5, respectivamente) y T33, C33 o S33 con una timina (45,6 % y 56,3 % en los grupos 3 y 9, respectivamente). G33 fue relativamente frecuente en el grupo 2, el grupo con la representación de bases más uniforme en ± 10 . Estos resultados son coherentes con las observaciones de Seligman y col. (Nucleic Acids Res., 2002, 30, 3870 – 3879), que mostraron anteriormente que una mutación Y33R o Y33H desplazó la especificidad de I-Crel hacia una guanina y Y33C, Y33T, Y33S (y también Y33L) por una timina en la posición ± 10 .

20 Además, R38 and K38 se asociaron con una frecuencia alta excepcional de guanina en el grupo 4, mientras que en todos los demás grupos, el residuo Q38 silvestre se sobrerrepresentó, así como una adenina en ± 9 de la diana.

La estructura de I-Crel unida a su diana (Chevalier y col., 2003, citado anteriormente; Jurica y col., 1998, citado anteriormente) ha mostrado que Y33 y Q38 contactan dos adeninas en -10 y -9 (Figura 6C), y los resultados sugieren que estas interacciones probablemente se mantengan en muchos de los mutantes. Anteriormente se han descrito resultados similares para el residuo 44 y la posición ± 4 (Arnould y col., citado anteriormente). No obstante, al comparar los resultados obtenidos para los pares $33 / \pm 10$, $38 / \pm 9$ y $44 / \pm 4$ muestra que una base dada se puede correlacionar con diferentes residuos de aminoácidos, en función de la posición. Para una guanina, los residuos mayoritarios son R y H en la posición 33, R o K en 38 y K en 44, para adenina, Y en 33 y Q en 38 y 44, y para timina, S, C o T en 33 y A en 44. En los tres casos no se observa ningún patrón claro para la citosina. Por tanto, no hay ningún "código" universal sino una serie de soluciones para contactar cada base, dependiendo la mejor solución de un contexto más general, muy similar al que se ha observado con las proteínas en dedo de cinc (Pabo y col., citado anteriormente).

35 LISTADO DE SECUENCIAS

<110> CELLECTIS
PAQUES, Frederic

40 <120> Variantes de la endonucleasa doméstica I-Crel que tiene nueva especificidad de escisión y uso de las mismas

<130> 1546PCT8

45 <160> 75

<170> PatentIn version 3.3

<210> 1

50 <211> 24

<212> ADN

<213> SECUENCIA ARTIFICIAL

<220>

55 <223> OLIGONUCLEÓTIDO

<400> 1

tcaaaacgctc gtgagacagt ttgg 24

60

<210> 2

<211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL
 5 <220>
 <223> OLIGONUCLEÓTIDO
 <400> 2
 10 ccaaacgtc togagacagt ttgg 24
 <210> 3
 <211> 24
 <212> ADN
 15 <213> SECUENCIA ARTIFICIAL
 <220>
 <223> OLIGONUCLEÓTIDO
 20 <400> 3
 tcaaacgtc gtaogacgtt ttga 24
 <210> 4
 25 <211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL
 <220>
 30 <223> OLIGONUCLEÓTIDO
 <400> 4
 tgggacgtc gtacgacgtc ccga 24
 35 <210> 5
 <211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL
 40 <220>
 <223> OLIGONUCLEÓTIDO
 <400> 5
 45 tgggaacgtc gtacgacgtt ccga 24
 <210> 6
 <211> 24
 50 <212> ADN
 <213> SECUENCIA ARTIFICIAL
 <220>
 <223> OLIGONUCLEÓTIDO
 55 <400> 6
 tgggtacgtc gtaogaagta ccga 24
 60 <210> 7

<211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

5 <400> 7

tcggcacgtc gtaogacgtg coga 24

<210> 8
 10 <211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

<220>
 15 <223> OLIGONUCLEÓTIDO

<400> 8

tcgagacgtc gtaogacgtc toga 24

20

<210> 9
 <211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

25

<220>
 <223> OLIGONUCLEÓTIDO

<400> 9

30

togaaacgtc gtaogacgtt toga 24

<210> 10
 <211> 24
 35 <212> ADN
 <213> secuencia artificial

<220>
 <223> oligonucleótido

40

<400> 10

tcgatacgtc gtaogacgta toga 24

45 <210> 11
 <211> 24
 <212> ADN
 <213> secuencia artificial

50 <220>
 <223> oligonucleótido

<400> 11

55 tcgacacgtc gtaogacgtg toga 24

<210> 12
 <211> 24
 <212> ADN
 60 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 12
 tcgtgacgtc gtaacgacgtc acga 24
 <210> 13
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 13
 tcgtaacgtc gtaacgacggt acga 24
 20 <210> 14
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 14
 30 <210> 15
 <211> 24
 <212> ADN
 <213> secuencia artificial
 35 <220>
 <223> oligonucleótido
 <400> 15
 tcgtcacgtc gtaacgacgtg acga 24
 45 <210> 16
 <211> 24
 <212> ADN
 <213> secuencia artificial
 50 <220>
 <223> oligonucleótido
 <400> 16
 55 <210> 17
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60 <210> 18
 <211> 24
 <212> ADN
 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 17
 tcgcaacgtc gtaogaogtt goga 24
 <210> 18
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 18
 tcgctacgtc gtaogaogta goga 24
 20 <210> 19
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 19
 30 <210> 20
 <211> 24
 <212> ADN
 <213> secuencia artificial
 35 <220>
 <223> oligonucleótido
 <400> 20
 tcaggaogtc gtacgaogtc ctga 24
 45 <210> 21
 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 50 <223> oligonucleótido
 <400> 21
 tcagaacgtc gtaogaogtt ctga 24
 55 <210> 22
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60

<220>
 <223> oligonucleótido

 <400> 22
 5 tcagtagctc gtagtagctc ctga 24

 <210> 23
 <211> 24
 10 <212> ADN
 <213> secuencia artificial

 <220>
 <223> oligonucleótido
 15 <400> 23

 tcagcagctc gtagtagctg ctga 24
 20 <210> 24
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido

 <400> 24
 30 tcaagagctc gtagtagctc ttga 24

 <210> 25
 <211> 24
 <212> ADN
 35 <213> secuencia artificial

 <220>
 <223> oligonucleótido
 40 <400> 25

 tcaaaagctc gtagtagctt ttga 24

 <210> 26
 45 <211> 24
 <212> ADN
 <213> secuencia artificial

 <220>
 50 <223> oligonucleótido

 <400> 26

 tcaatagctc gtagtagctc ttga 24
 55 <210> 27
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60

<220>
 <223> oligonucleótido

 <400> 27
 5 tcaacacgtc gtaogacgtg ttga 24

 <210> 28
 <211> 24
 10 <212> ADN
 <213> secuencia artificial

 <220>
 <223> oligonucleótido
 15 <400> 28

 tcaacacgtc gtaogacgtg ttga 24

 20 <210> 29
 <211> 24
 <212> ADN
 <213> secuencia artificial

 25 <220>
 <223> oligonucleótido

 <400> 29

 30 tcataacgtc gtaogacggtt atga 24

 <210> 30
 <211> 24
 <212> ADN
 35 <213> secuencia artificial

 <220>
 <223> oligonucleótido

 40 <400> 30

 tcattacgtc gtaogacgta atga 24

 <210> 31
 45 <211> 24
 <212> ADN
 <213> secuencia artificial

 <220>
 50 <223> oligonucleótido

 <400> 31

 tcalcacgtc gtaogacgtg atga 24
 55
 <210> 32
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60

<220>
 <223> oligonucleótido

 <400> 32
 5 tcatcaogtc gtaogaogtg atga 24

 <210> 33
 <211> 24
 10 <212> ADN
 <213> secuencia artificial

 <220>
 <223> oligonucleótido
 15 <400> 33

 tcatcaogtc gtaogaogtg atga 24

 20 <210> 34
 <211> 24
 <212> ADN
 <213> secuencia artificial

 25 <220>
 <223> oligonucleótido

 <400> 34

 30 tcactaogtc gtaogaogta gtga 24

 <210> 35
 <211> 24
 <212> ADN
 35 <213> secuencia artificial

 <220>
 <223> oligonucleótido

 40 <400> 35

 tcaccacgtc gtaogaogtg gtga 24

 <210> 36
 45 <211> 24
 <212> ADN
 <213> secuencia artificial

 <220>
 50 <223> oligonucleótido

 <400> 36

 tctggacgtc gtaogaogtc caga 24
 55 <210> 37
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60

<220>
 <223> oligonucleótido

 <400> 37
 5 tctgaacgtc gtaacgaagtt caga 24

 <210> 38
 <211> 24
 10 <212> ADN
 <213> secuencia artificial

 <220>
 <223> oligonucleótido
 15 <400> 38

 tctgtaacgtc gtaacgaagta caga 24

 20 <210> 39
 <211> 24
 <212> ADN
 <213> secuencia artificial

 25 <220>
 <223> oligonucleótido

 <400> 39

 30 tctgcaacgtc gtaacgaagtg caga 24

 <210> 40
 <211> 24
 <212> ADN
 35 <213> secuencia artificial

 <220>
 <223> oligonucleótido

 40 <400> 40

 tctgcaacgtc gtaacgaagtg caga 24

 <210> 41
 45 <211> 24
 <212> ADN
 <213> secuencia artificial

 <220>
 50 <223> oligonucleótido

 <400> 41

 tctaaacgtc gtaacgaagtt taga 24
 55
 <210> 42
 <211> 24
 <212> ADN
 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 42
 tctatacgtc gtaogaogta taga 24
 <210> 43
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 43 24
 tctacacgtc gtaogaogtg taga 24
 20 <210> 44
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 44 24
 30 <210> 45
 <211> 24
 <212> ADN
 <213> secuencia artificial
 35 <220>
 <223> oligonucleótido
 <400> 45
 tcttaacgtc gtaogaogtt aaga 24
 45 <210> 46
 <211> 24
 <212> ADN
 <213> secuencia artificial
 50 <220>
 <223> oligonucleótido
 <400> 46
 55 <210> 47
 <211> 24
 <212> ADN

<213> secuencia artificial
 <220>
 <223> oligonucleótido
 5 <400> 47
 tctcaogtc gtacgaogtg aaga 24
 10 <210> 48
 <211> 24
 <212> ADN
 <213> secuencia artificial
 15 <220>
 <223> oligonucleótido
 <400> 48
 20 tctogaogtc gtaogaogtc gaga 24
 <210> 49
 <211> 24
 <212> ADN
 25 <213> secuencia artificial
 <220>
 <223> oligonucleótido
 <400> 49
 30 tctcaogtc gtaogaogtt gaga 24
 <210> 50
 <211> 24
 35 <212> ADN
 <213> secuencia artificial
 <220>
 <223> oligonucleótido
 40 <400> 50
 tctctaogtc gtacgaogta gaga 24
 45 <210> 51
 <211> 24
 <212> ADN
 <213> secuencia artificial
 50 <220>
 <223> oligonucleótido
 <400> 51
 55 tctccaogtc gtaogaogtg gaga 24
 <210> 52
 <211> 24
 <212> ADN
 60 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 52
 tcggacgtc gtaacgagtc cgga 24
 <210> 53
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 53
 tcogaacgtc gtaacgagtt cgga 24
 20 <210> 54
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 54
 30 <210> 55
 <211> 24
 <212> ADN
 <213> secuencia artificial
 35 <220>
 <223> oligonucleótido
 <400> 55
 tcgcacgtc gtaacgagtg cgga 24
 45 <210> 56
 <211> 24
 <212> ADN
 <213> secuencia artificial
 50 <220>
 <223> oligonucleótido
 <400> 56
 55 <210> 57
 <211> 24
 <212> ADN
 60 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 57
 tccaaacgtc gta cga cgtt tgga 24
 <210> 58
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 58
 tcataacgtc gta cga cgtg tgga 24
 20 <210> 59
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 59
 30 <210> 60
 <211> 24
 <212> ADN
 <213> secuencia artificial
 35 <220>
 <223> oligonucleótido
 <400> 60
 tccacacgtc gta cga cgtg tgga 24
 40 <210> 61
 <211> 24
 <212> ADN
 <213> secuencia artificial
 45 <220>
 <223> oligonucleótido
 <400> 61
 tctgacgtc gta cga cgtc agga 24
 50 <210> 62
 <211> 24
 <212> ADN
 <213> secuencia artificial
 55 <220>
 <223> oligonucleótido
 <400> 61
 tcctaacgtc gta cga cgtt agga 24
 60 <210> 62
 <211> 24
 <212> ADN
 <213> secuencia artificial

<220>
 <223> oligonucleótido
 5 <400> 62
 tcctaacgtc gtaacgaagta agga 24
 <210> 63
 10 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 15 <223> oligonucleótido
 <400> 63
 tcctaacgtc gtaacgaagtg agga 24
 20 <210> 64
 <211> 24
 <212> ADN
 <213> secuencia artificial
 25 <220>
 <223> oligonucleótido
 <400> 64
 30 tcccgacgtc gtaacgaagtc gggg 24
 <210> 65
 <211> 24
 <212> ADN
 35 <213> secuencia artificial
 <220>
 <223> oligonucleótido
 40 <400> 65
 tcccaacgtc gtaacgaagtt gggg 24
 <210> 66
 45 <211> 24
 <212> ADN
 <213> secuencia artificial
 <220>
 50 <223> secuencia artificial
 <400> 66
 tccctaacgtc gtaacgaagta gggg 24
 55 <210> 67
 <211> 24
 <212> ADN
 <213> secuencia artificial
 60

<220>
<223> oligonucleótido

<400> 67

5 tccccagtc gta cga cgtg ggga 24

<210> 68
<211> 501
10 <212> ADN
<213> secuencia artificial

<220>
<223> Armazón de I-Crel N75

15 <400> 68

atggccaata ccaaataaa caaagagttc ctgctgtacc tggccggctt tgtggacggt 60
gacggtagca tcatcgtca gattaaacca aaccagtctt ataagttaa acatcagcta 120
agcttgacct ttcaggtgac tcaaaagacc cagcggcgtt ggtttctgga caaactagtg 180
gatgaaattg gcgttggtta cgtacgtgat cgcggatccg tttccaacta catcttaagc 240
gaaatcaagc cgctgcacaa cttcctgact caactgcagc cgtttctgaa actgaaacag 300
aaacaggcaa acctggttct gaaaattatc gaacagctgc cgtctgcaaa agaatccccg 360
gacaaattcc tggaagtttg tacctgggtg gatcagattg cagctctgaa cgattctaag 420
acgcgtaaaa ccacttctga aaccgttcgt gctgtgctgg acagcctgag cgagaagaag 480
aaatcctccc cggcggccga c 501

20 <210> 69
<211> 68
<212> ADN
<213> secuencia artificial

25 <220>
<223> oligonucleótido

<400> 69

ctgaaaggtc aagcttagmb batgtttaaa cttmbbagac tgmbbtggtt taatctgagc 60
gatgatgc 68

30 <210> 70
<211> 68
<212> ADN
35 <213> secuencia artificial

<220>
<223> oligonucleótido

40 <400> 70

ctgaaaggtc aagcttagmb batgtttaa cttataagac tgmbbtggmb baatctgagc 60
gatgatgc 68

<210> 71
 <211> 18
 5 <212> ADN
 <213> secuencia artificial

<400> 71
 ctaagcttga ccttcag 18

10 <210> 72
 <211> 49
 <212> ADN
 <213> secuencia artificial

15 <220>
 <223> oligonucleótido

<220>
 20 <221> misc_feature
 <222> (15) .. (17)
 <223> n es a, c, g o t

<220>
 25 <221> misc_feature
 <222> (32) .. (34)
 <223> n es a, c, g, ó t

<400> 72
 30 ggcatacaag ttcnnaag tcgtacgacg tnnngacaat cgtctgca - 49

<210> 73
 <211> 26
 35 <212> ADN
 <213> secuencia artificial

<220>
 <223> oligonucleótido

40 <400> 73
 gcaacttag tgctgacaca tacagg 26

45 <210> 74
 <211> 24
 <212> ADN
 <213> secuencia artificial

50 <220>
 <223> oligonucleótido

<400> 74

55 acaacctga ttgcagact gacc 24

<210> 75

<211> 21
<212> ADN
<213> secuencia artificial

5 <220>
<223> oligonucleótido

<220>
<221> misc_feature
10 <222> (2) .. (4)
<223> n es a, c, g, ó t

<220>
<221> misc_feature
15 <222> (6) .. (8)
<223> n es a, c, g, ó t

<220>
<221> misc_feature
20 <222> (13) .. (15)
<223> n es a, c, g, ó t

<220>
<221> misc_feature
25 <222> (18) .. (20)
<223> n es a, c, g, ó t

<400> 75

30 c n n n m n n n k y r m n n n r k n n n g 21

REIVINDICACIONES

1. Un procedimiento para someter a ingeniería una variante de endonucleasa doméstica I-Crel que tiene una especificidad de escisión modificada en las posiciones ± 8 y ± 9 del sitio diana de I-Crel (SEQ ID NO: 1) que comprende las etapas de:
- (a) sustituir al menos uno de los aminoácidos K28, N30, Y33, Q38 y/o S40 de la horquilla $\beta_1\beta_2$ de I-Crel, para modificar la especificidad de escisión en la posición ± 8 de dicho sitio diana de I-Crel (SEQ ID NO: 1) con un aminoácido seleccionado del grupo constituido por A, C, D, E, G, H, K, N, P, Q, R, S, T, L, V, W e Y, y/o
- (b) sustituir al menos uno de los aminoácidos K28, Y33 y/o S40 de la horquilla $\beta_1\beta_2$ de I-Crel, para modificar la especificidad de escisión en la posición ± 9 de dicho sitio diana de I-Crel (SEQ ID NO: 1) con un aminoácido seleccionado del grupo constituido por A, C, D, E, G, H, K, N, P, Q, R, S, T, L, V, W e Y,
- (c) seleccionar las variantes de I-Crel de la etapa (a) y/o (b) que son capaces de escindir una secuencia diana de ADN constituida por un sitio I-Crel mutante en el que al menos el doblete nucleotídico aa en las posiciones -9 a -8 y/o el doblete nucleotídico tt en las posiciones $+8$ a $+9$ se han sustituido por un doblete nucleotídico diferente.
2. El procedimiento de acuerdo con la reivindicación 1, en el que en la etapa (b) al menos uno de los residuos N30 y Q38 también está sustituido con un aminoácido seleccionado del grupo constituido por A, C, D, E, G, H, K, N, P, Q, R, S, T, L, V, W e Y.
3. El procedimiento de acuerdo con la reivindicación 1 o 2, en el que la diana de ADN en la etapa c) deriva de un sitio I-Crel que se selecciona de las SEQ ID NO: 1 a 3.
4. El procedimiento de acuerdo con la reivindicación 1, la reivindicación 2 o la reivindicación 3, en el que la diana de ADN en la etapa c) comprende una secuencia que tiene la fórmula:
- $$C^{-11}N^{-10}N^{-9}N^{-8}M^{-7}Y^{-6}N^{-5}N^{-4}N^{-3}K_2Y^{-1}r + 1m + 2n + 3n + 4n + 5r + 6k + 7n + 8n + 9n + 10g + 11$$
- en la que n es a, t, c o g, m es a o c, y es c o t, k es g o t, r es a o g (SEQ ID NO: 75), que proporciona que cuando n_{-9} es aa, n_{+8} a n_{+9} es diferente de tt y cuando n_{+8} a n_{+9} es tt, n_{-9} es diferente a un aa.
5. El procedimiento de acuerdo con la reivindicación 4, en el que $n_{-5}n_{-4}n_{-3}$ es gtc y/o $n_{+3}n_{+4}n_{+5}$ es gac.
6. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que la secuencia de nucleótidos de las posiciones -11 a -8 y $+8$ a $+11$ y/o la secuencia nucleotídica de las posiciones -5 a -3 y/o $+3$ a $+5$ de dicha diana de ADN en la etapa c) son palindrómicas.
7. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en el que la diana de ADN en la etapa c) comprende un doblete nucleotídico en las posiciones -9 a -8 , que se selecciona del grupo constituido por: ag, at, ac, ga, gg, gt, gc, ta, tg, tt, cg, ct o cc, y/o un doblete nucleotídico en las posiciones $+8$ a $+9$, que es la secuencia complementaria inversa de dicho doblete nucleotídico en las posiciones -9 a -8 .
8. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 7, en el que la diana de ADN en la etapa c) comprende además la sustitución del nucleótido a en la posición -10 y/o del nucleótido t en la posición $+10$ del sitio I-Crel, con un nucleótido diferente.
9. El procedimiento de acuerdo con la reivindicación 8, en el que dicha diana de ADN comprende un triplete nucleotídico en las posiciones -10 a -8 , que se selecciona del grupo constituido por: aac, aag, aat, acc, acg, act, aga, agc, agg, agt, ata, atg, cag, cga, cgg, ctg, gac, gag, gat, gcc, gga, ggc, ggg, ggt, gta, gtg, gtt, tac, tag, tat, tcc, tga, tgc, tgg, tgt o ttg, y/o un triplete nucleotídico en las posiciones $+8$ a $+10$, que es la secuencia complementaria inversa de dicho triplete nucleotídico en las posiciones -10 a -8 .
10. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en el que la etapa (c) se realiza *in vivo* en las condiciones en las que la rotura de la doble cadena en la secuencia diana de ADN mutada que se genera mediante dicha variante conduce a la activación de un marcador de selección positiva o un gen indicador, o la inactivación de un marcador de selección negativa o un gen indicador, mediante reparación mediada por recombinación de dicha rotura de la doble cadena de ADN.
11. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 10, que comprende una

etapa adicional d₁) de expresión de una variante obtenida en la etapa c) para permitir la formación de homodímeros.

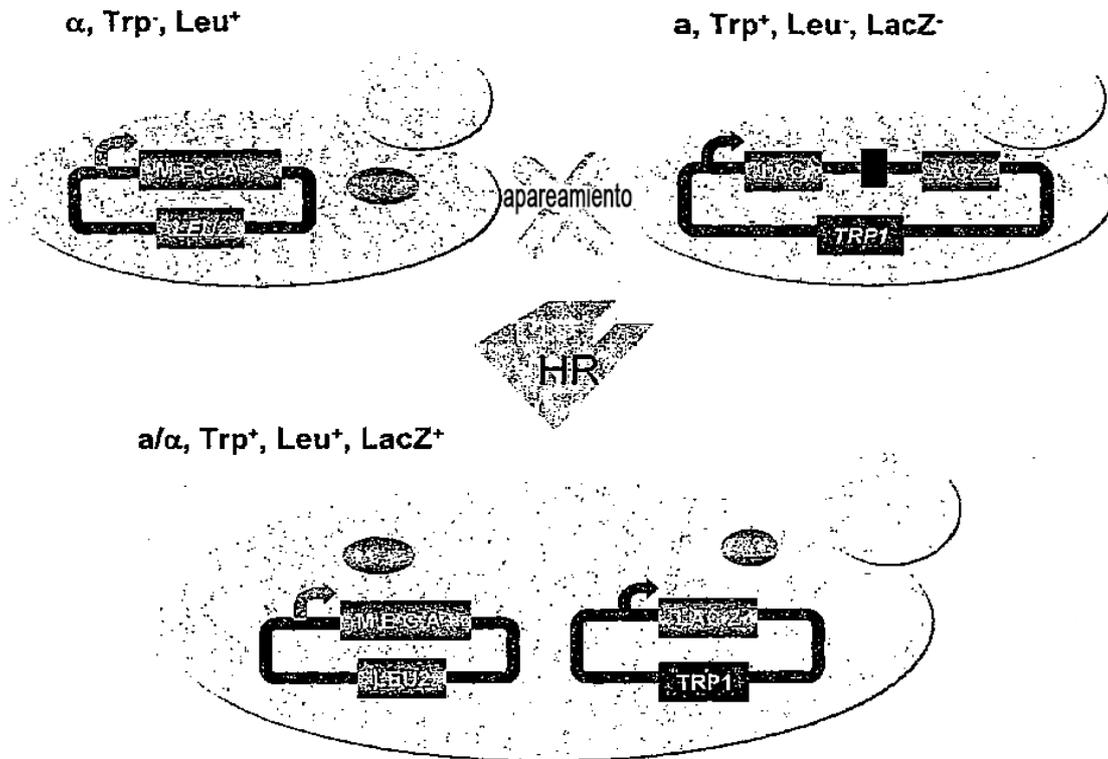
12. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 11, que comprende además una etapa adicional d₂) de coexpresión de una variante obtenida en la etapa c) y I-Crel o una variante funcional de la misma para permitir la formación de heterodímeros.

13. El procedimiento de la reivindicación 12, en el que dos variantes diferentes obtenidas en la etapa c) se coexpresan.

SEQ ID NO 1-3

		-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	
C1234	T	C	A	A	A	A	A	C	G	T	C	G	T	G	A	C	A	T	C	A	G	T	T	T	C	G
C4334	C	C	A	A	A	A	C	T	G	T	C	T	C	G	A	G	A	C	A	G	T	T	T	C	G	G
C1221	T	C	A	A	A	A	A	C	G	T	C	G	T	A	C	C	A	C	G	T	T	T	T	C	A	A

FIGURA 1A



Selección de diploides, prueba de X-Gal

FIGURA 2

SEQ ID NO: 68

```

1   ATGGCCAATA CCAAATATAA CAAAGAGTTC CTGCTGTACC TGGCCGGCTT
51  TGTGGACGGT GACGGTAGCA TCATCGCTCA GATTAAACCA AACCAGTCTT
101 ATAAGTTTAA ACATCAGCTA AGCTTGACCT TTCAGGTGAC TCAAAAGACC
151 CAGCGCCGTT GTTTCTGGA CAAACTAGTG GATGAAATTG GCGTTGGTTA
201 CGTACGTGAT CGCGGATCCG TTTCCAATA CATCTTAAGC GAAATCAAGC
251 CGCTGCACAA CTTCTGACT CAACTGCAGC CGTTTCTGAA ACTGAAACAG
301 AACAGGCAA ACCTGGTTCT GAAAATTATC GAACAGCTGC CGTCTGCAAA
351 AGAATCCCCG GACAAATTCC TGGAAGTTTG TACCTGGGTG GATCAGATTG
401 CAGCTCTGAA CGATTCTAAG ACGCGTAAA CCACTTCTGA AACCGTTCGT
451 GCTGTGCTGG ACAGCCTGAG CGAGAAGAAG AAATCCTCCC CGGCGGCCGA
501 C
    
```

FIGURA 3A

```

Ulib4rev CTGAAAGGTC AAGCTTAGMB BATGTTTAAA CTTMBBAGAC
SEQ ID NO:69 TGMBBTGGTT TAATCTGAGC GATGATGC
    
```

```

Ulib5rev CTGAAAGGTC AAGCTTAGMB BATGTTTAAA CTTATAAGAC
SEQ ID NO:70 TGMBBTGGMB BAATCTGAGC GATGATGC
    
```

```

Ulib456for CTAAGCTTGA CCTTTCAG
SEQ ID NO:71
    
```

V = A o G o C

M = A o C

B = G o C o T

K = G o T

FIGURA 3B

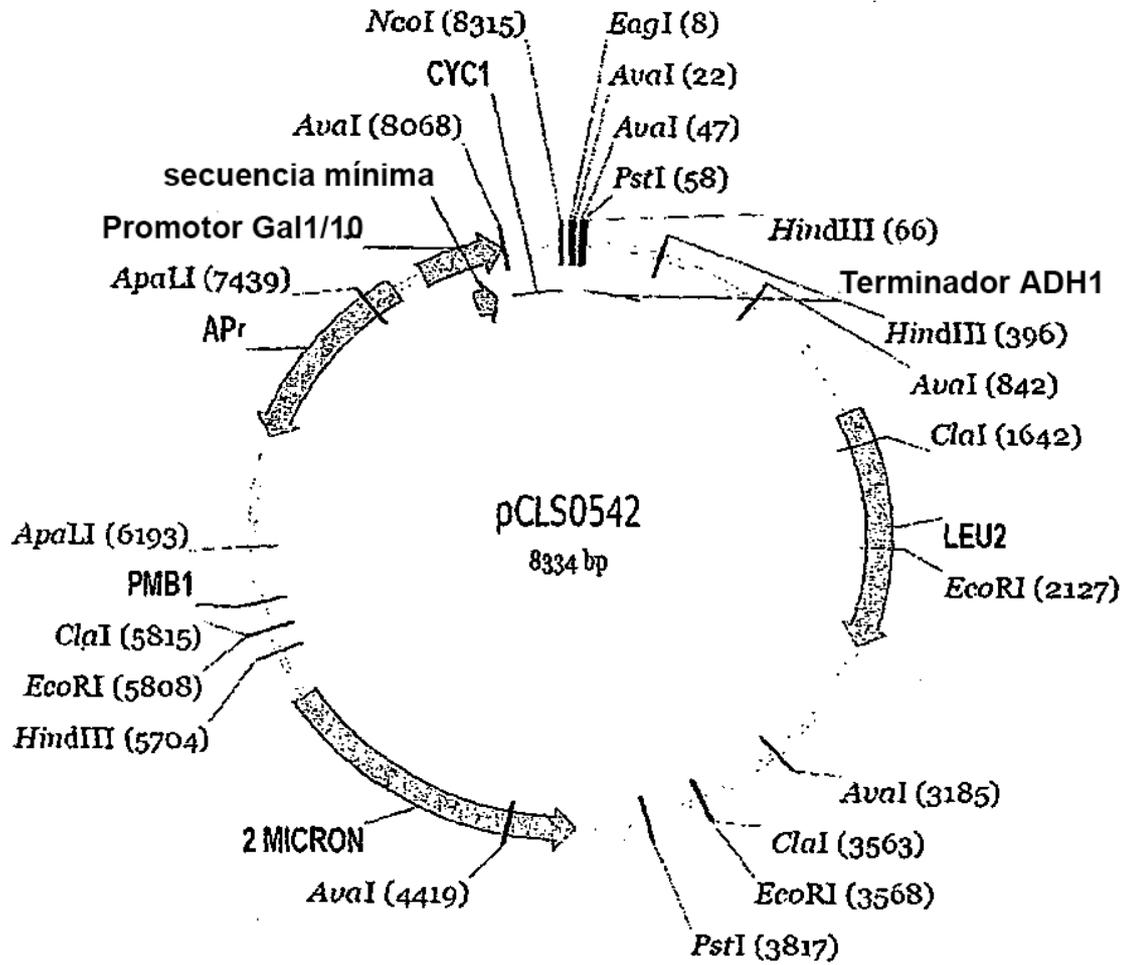


FIGURA 4

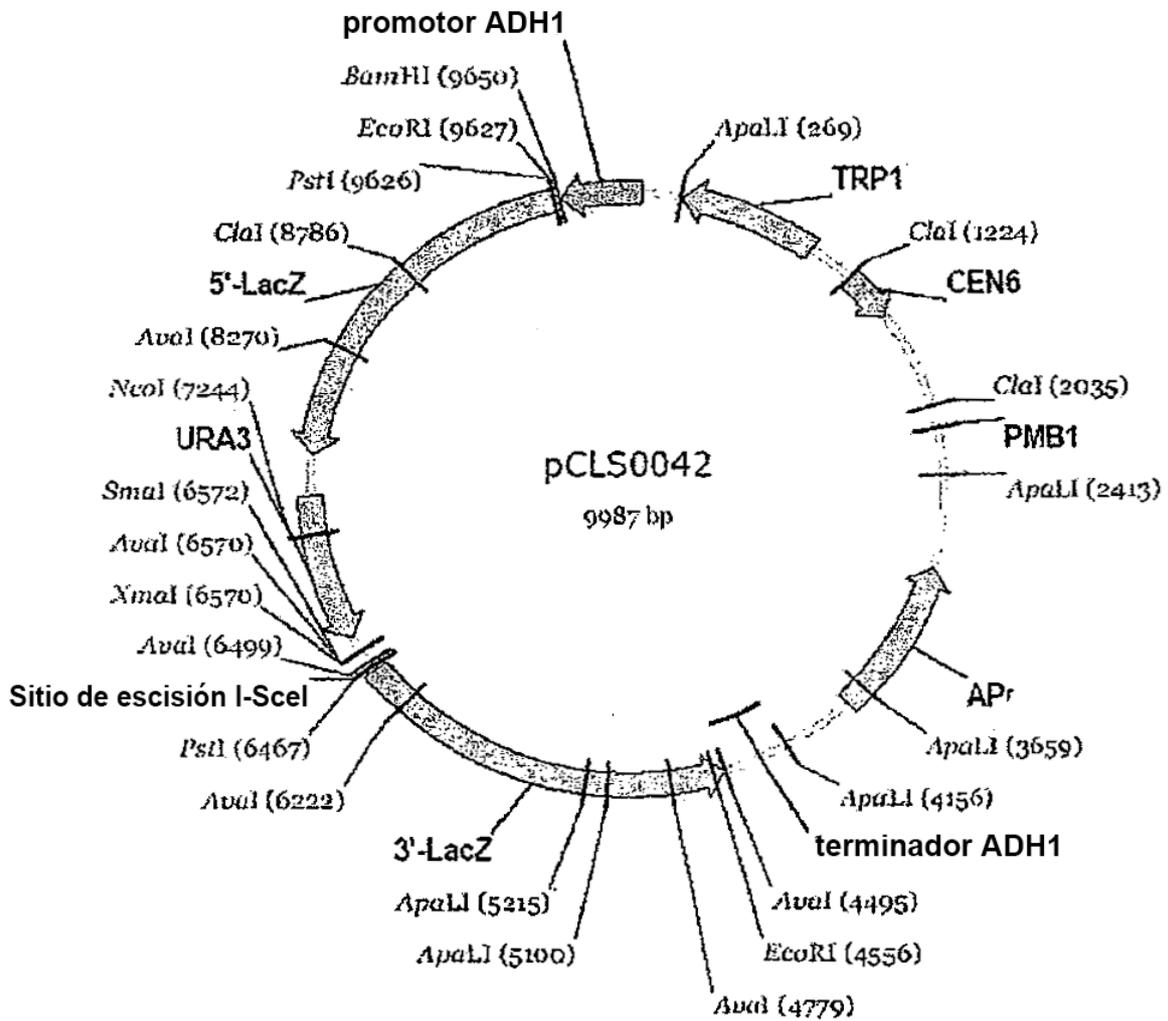


FIGURA 5

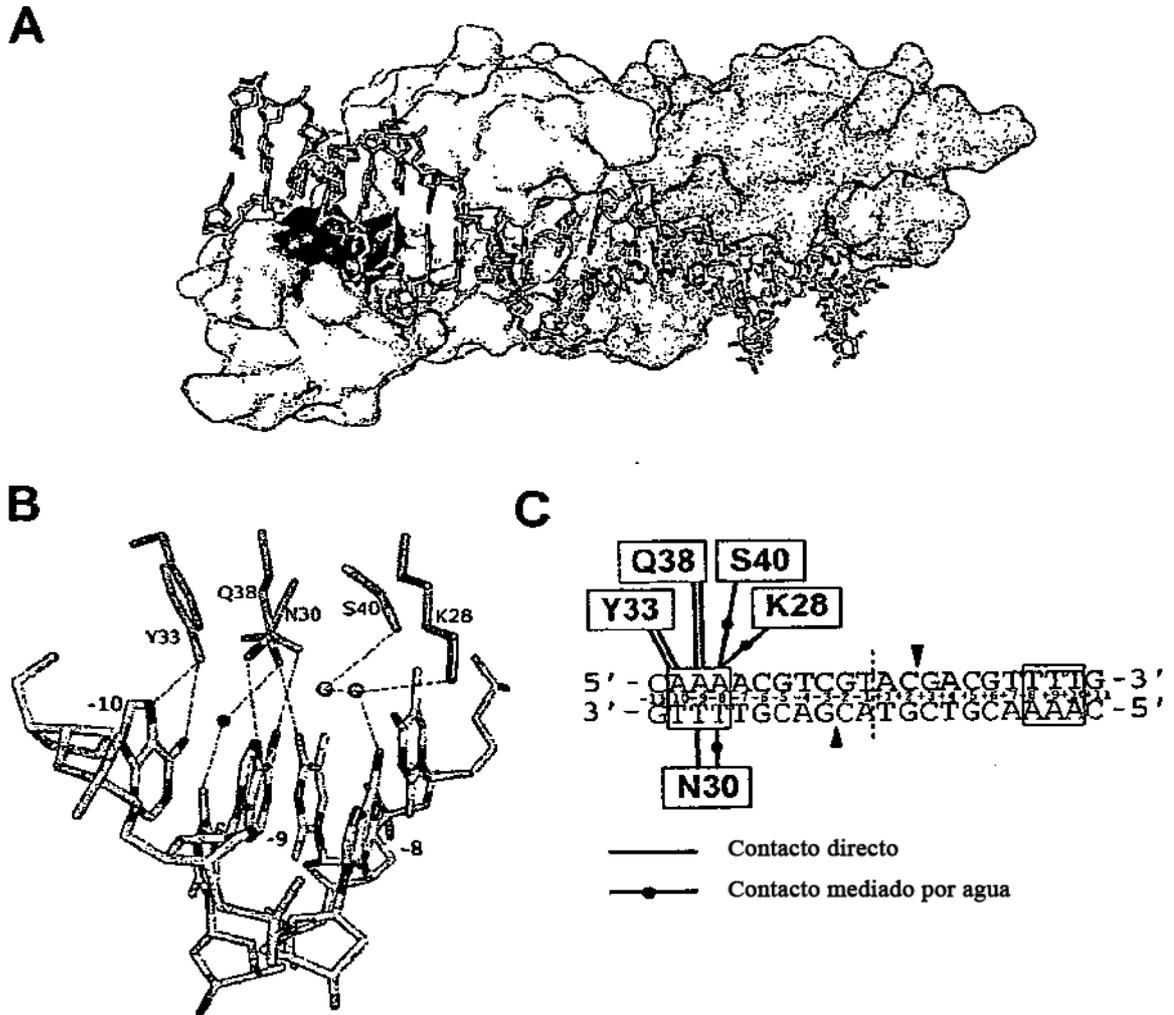


FIGURA 6A B Y C

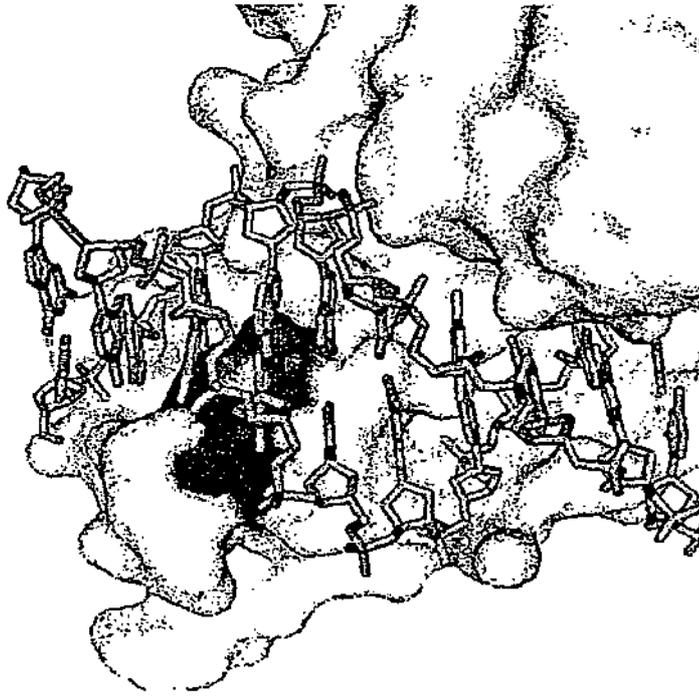


FIGURA 6D

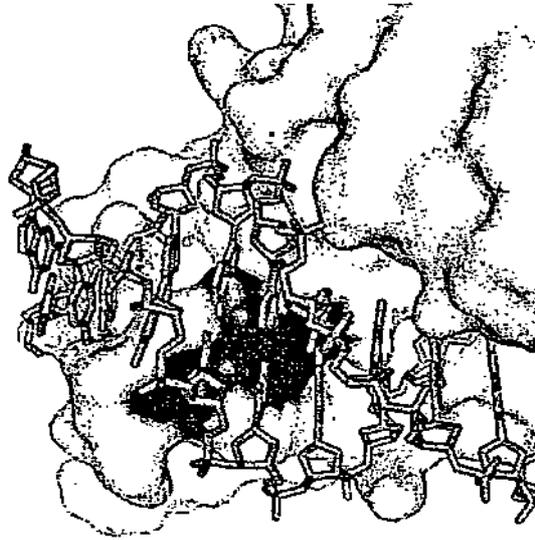


FIGURA 6E

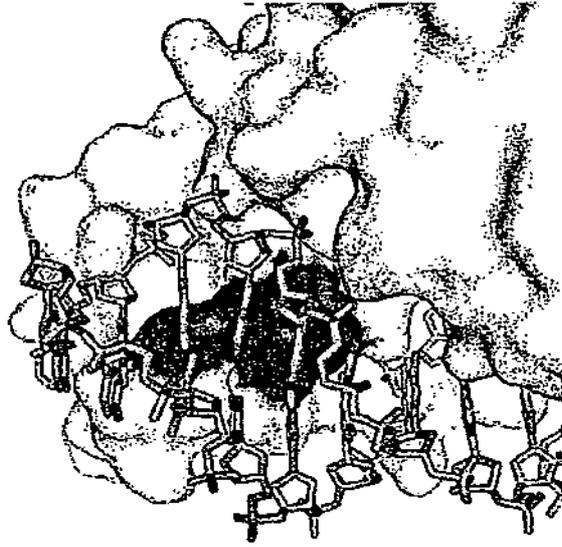


FIGURA 6F

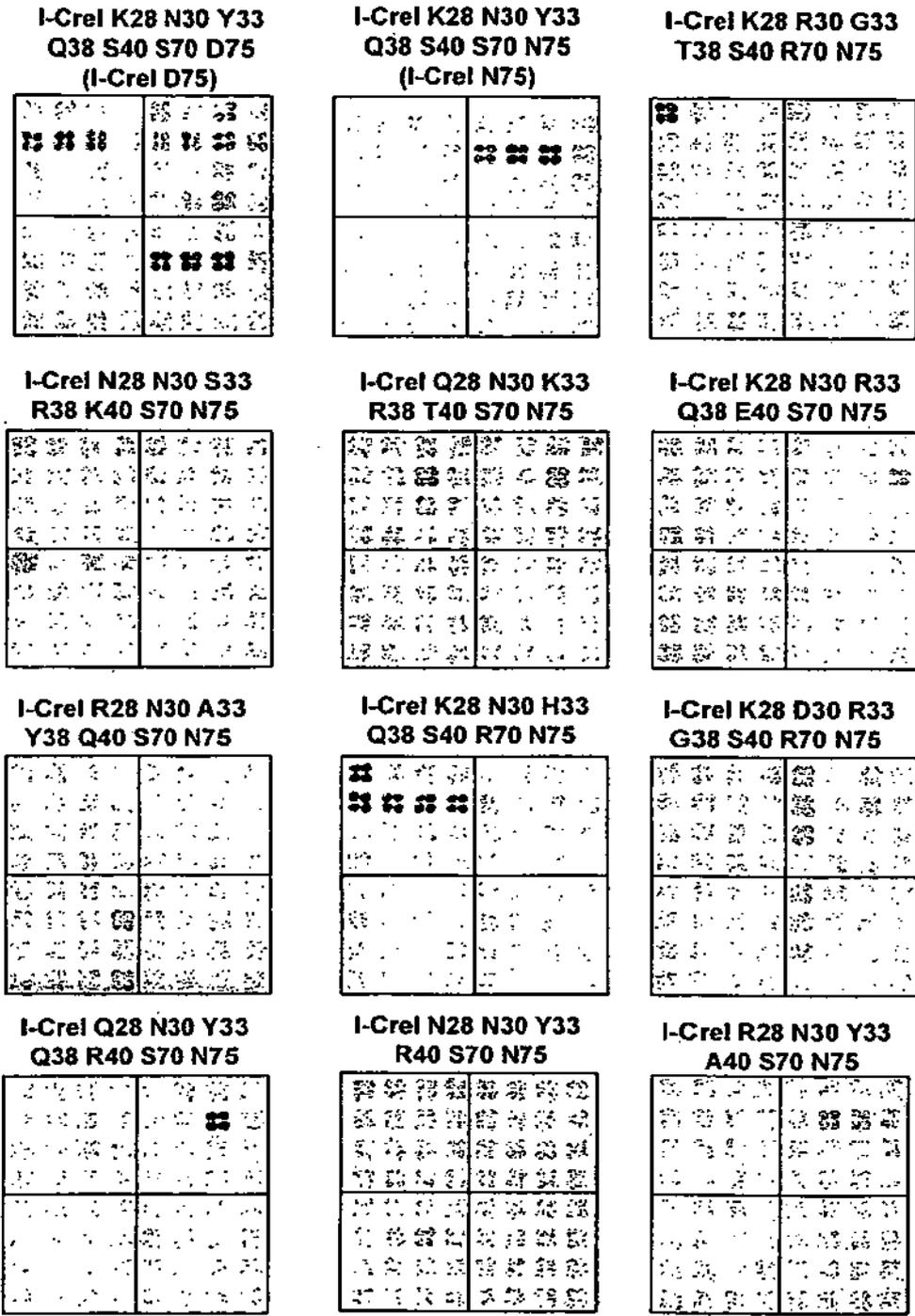


FIGURA 8A

GCG	GGA	GCT	GGC	AGG	AGA	AGT	AGC
618	245	290	31	502	194	295	135
GAG	GAA	GAT	GAC	AAG	AAA	AAT	AAC
686	422	524	313	936	616	821	478
GTG	GTA	GTT	GTC	ATG	ATA	ATT	ATC
249	119	134	78	348	116	190	54
GCG	GCA	GCT	GCC	ACG	ACA	ACT	ACC
142	124	130	103	163	178	201	202
TGG	TGA	TGT	TGC	CGG	CGA	CGT	CGC
322	141	203	165	288	64	107	76
TAG	TAA	TAT	TAC	CAG	CAA	CAT	CAC
589	282	372	318	396	230	224	121
TTG	TTA	TTT	TTC	CTG	CTA	CTT	CTC
201	124	133	125	145	47	41	3
TCG	TCA	TCT	TCC	CCG	CCA	CCT	CCC
68	23	59	38	28	15	33	51

FIGURA 8B