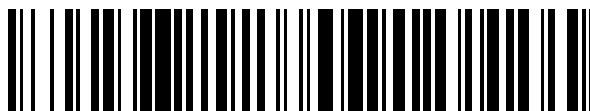


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 440 801**

51 Int. Cl.:

C12N 9/22 (2006.01)

C12N 15/90 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.10.2006** **E 12162075 (1)**

97 Fecha y número de publicación de la concesión europea: **02.10.2013** **EP 2484758**

54 Título: **Meganucleasas racionalmente diseñadas con especificidad de secuencia y afinidad de unión a ADN alteradas**

30 Prioridad:

18.10.2005 US 727512 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.01.2014

73 Titular/es:

PRECISION BIOSCIENCES (100.0%)
302 East Pettigrew Street, Dibrell Building, Suite
A-100
Durham, North Carolina 27701 , US

72 Inventor/es:

HELLINGA, HOMME, W.;
SMITH, JAMES JEFFERSON y
JANTZ, DEREK

74 Agente/Representante:

IZQUIERDO FACES, José

ES 2 440 801 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**Meganucleasas racionalmente diseñadas con especificidad de secuencia y afinidad de unión a ADN alteradas**

5

FINANCIACIÓN GUBERNAMENTAL

La invención fue financiada en parte por las ayudas 2R01-GM-0498712, 5F32-GM072322 y 5 DP1 OD000122 del Instituto Nacional de Ciencias Médicas Generales del Instituto Nacional de la Salud de los Estados Unidos de América. Por tanto, el gobierno de EE.UU. puede tener ciertos derechos en la invención.

CAMPO DE LA INVENCION

La invención se refiere al campo de la biología molecular y tecnología de ácidos nucleicos recombinantes. En particular, la invención se refiere a meganucleasas que no se producen naturalmente racionalmente diseñadas con especificidad de secuencia de reconocimiento de ADN alterada. La invención también se refiere a procedimientos de producción de tales meganucleasas, y a procedimientos de producción de ácidos nucleicos recombinantes y organismos usando tales meganucleasas.

ANTECEDENTES DE LA INVENCION

La ingeniería del genoma requiere la capacidad para insertar, delecionar, sustituir y manipular de otro modo secuencias genéticas específicas dentro de un genoma, y tiene numerosas aplicaciones terapéuticas y biotecnológicas. El desarrollo de medios eficaces para la modificación del genoma sigue siendo un objetivo principal en la terapia génica, agrotecnología y biología sintética (Porteus y col. (2005), Nat. Biotechnol. 23: 967-73; Tzfira y col. (2005), Trends Biotechnol. 23: 567-9; McDaniel y col. (2005), Curr. Opin. Biotechnol. 16: 476-83). Un procedimiento común de inserción o modificación de una secuencia de ADN implica introducir una secuencia de ADN transgénica flanqueada por secuencias homólogas a la diana genómica y seleccionar o cribar un evento de recombinación homólogo satisfactorio. La recombinación con el ADN transgénico se produce raramente, pero puede estimularse por una rotura bicatenaria en el ADN genómico en el sitio diana. Se han empleado numerosos procedimientos para crear roturas de ADN bicatenario, que incluyen tratamientos de irradiación y químicos. Aunque estos procedimientos estimulan eficazmente la recombinación, las roturas bicatenarias están dispersas al azar en el genoma, que puede ser altamente mutagénico y tóxico. Actualmente, la incapacidad para elegir como diana modificaciones génicas para sitios únicos dentro de una referencia cromosómica es un impedimento importante para la satisfactoria ingeniería del genoma.

Un enfoque para alcanzar este objetivo es estimular la recombinación homóloga en una rotura bicatenaria en un sitio diana usando una nucleasa con especificidad por una secuencia que es suficientemente grande para estar presentes en solo un único sitio dentro del genoma (véase, por ejemplo, Porteus y col. (2005), Nat. Biotechnol. 23: 967-73). La eficacia de esta estrategia se ha demostrado en una variedad de organismos usando fusiones quiméricas entre un dominio de unión a ADN de dedo de cinc manipulado y el dominio de nucleasa no específico de la enzima de restricción FokI (Porteus (2006), Mol Ther 13: 438-46; Wright y col. (2005), Plant J. 44: 693-705; Urnov y col. (2005), Nature 435: 646-51). Aunque estas nucleasas de dedo de cinc artificiales estimulan la recombinación específica de sitio, retienen actividad de escisión no específica residual resultante de la regulación por disminución del dominio de nucleasa y frecuentemente se escinden en sitios involuntarios (Smith y col. (2000), Nucleic Acids Res. 28: 3361-9). Tal escisión involuntaria puede producir mutaciones y toxicidad en el organismo tratado (Porteus y col. (2005), Nat. Biotechnol. 23: 967-73).

Un grupo de nucleasas que se producen naturalmente que reconocen sitios de escisión de 15-40 pares de bases comúnmente encontrados en los genomas de plantas y hongos puede proporcionar una alternativa de ingeniería del genoma menos tóxica. Tales "meganucleasas" o "endonucleasas de recirculación" están frecuentemente asociadas a elementos de ADN parasíticos, tales como intrones e inteínas de auto-corte y empalme del Grupo 1. Promueven naturalmente la recombinación homóloga o inserción génica en localizaciones específicas en el genoma del huésped produciendo una rotura bicatenaria en el cromosoma, que recluta la maquinaria de reparación de ADN celular (Stoddard (2006), Q. Rev. Biophys. 38: 49-95). Las meganucleasas se agrupan comúnmente en cuatro familias: la familia LAGLIDADG, la familia GIY-YIG, la familia de la caja His-Cys y la familia HNH. Estas familias se caracterizan por motivos estructurales, que afectan la actividad catalítica y secuencia de reconocimiento. Por ejemplo, miembros de la familia LAGLIDADG se caracterizan por tener tanto una como dos copias del motivo LAGLIDADG conservado (véase Chevalier y col. (2001), Nucleic Acids Res. 29(18): 3757-3774). Las meganucleasas de LAGLIDADG con una única copia del motivo LAGLIDADG forman homodímeros, mientras que miembros con dos copias del motivo LAGLIDADG se encuentran como monómeros. Similarmente, los miembros de la familia GIY-YIG tienen un módulo de GIY-YIG, que tiene 70-100 residuos de longitud e incluye cuatro o cinco motivos de secuencia conservados con cuatro residuos invariantes, dos de los cuales se requieren para la actividad (véase Van Roey y col. (2002), Nature Struct. Biol. 9: 806-811). Las meganucleasas de la caja His-Cys se caracterizan por una serie altamente conservada de histidinas y cisteínas sobre una región que engloba varios cientos de residuos de aminoácidos (véase Chevalier y col. (2001), Nucleic Acids Res. 29(18): 3757-3774). En el caso de la familia de NHN, los miembros se definen por

motivos que contienen dos pares de histidinas conservadas rodeadas por residuos de asparagina (véase Chevalier y col. (2001), *Nucleic Acids Res.* 29(18): 3757-3774). Las cuatro familias de meganucleasas están ampliamente separadas entre sí con respecto a elementos estructurales conservados y, por consiguiente, especificidad de secuencias de reconocimiento de ADN y actividad catalítica.

Las meganucleasas naturales, principalmente de la familia LAGLIDADG, se han usado para promover eficazmente la modificación del genoma específica de sitio en plantas, levadura, *Drosophila*, células de mamífero y ratones, pero este enfoque se ha limitado a la modificación de tanto genes homólogos que conservan la secuencia de reconocimiento de meganucleasas (Monnat y col. (1999), *Biochem. Biophys. Res. Commun.* 255: 88-93) como a genomas previamente manipulados en los que se ha introducido una secuencia de (Rouet y col. (1994), *Mol. Cell. Biol.* 14: 8096-106; Chilton y col. (2003), *Plant Physiol.* 133: 956-65; Puchta y col. (1996), *Proc. Natl. Acad. Sci. USA* 93: 5055-60; Rong y col. (2002), *Genes Dev.* 16: 1568-81; Gouble y col. (2006), *J. Gene Med.* 8(5):616-622).

La implementación sistemática de la modificación génica estimulada por nucleasas requiere el uso de enzimas manipuladas con especificidades personalizadas para elegir como diana roturas de ADN para sitios existentes en un genoma y, por tanto, ha habido gran interés en adaptar las meganucleasas para promover modificaciones génicas en sitios médicamente o biotecnológicamente relevantes (Porteus y col. (2005), *Nat. Biotechnol.* 23: 967-73; Sussman y col. (2004), *J. Mol. Biol.* 342: 31-41; Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62).

La meganucleasa I-Crel de *Chlamydomonas reinhardtii* es un miembro de la familia LAGLIDADG que reconoce y corta una secuencia de reconocimiento de 22 pares de bases en el cromosoma del cloroplasto, y que presenta una diana atractiva para el rediseño de meganucleasas. La enzima natural es un homodímero en el que cada monómero hace contactos directos con 9 pares de bases en la secuencia de reconocimiento de longitud completa. Se han usado técnicas de selección genética para identificar mutaciones en I-Crel que alteran la preferencia de bases en una única posición en esta secuencia de reconocimiento (Sussman y col. (2004), *J. Mol. Biol.* 342: 31-41; Chames y col. (2005), *Nucleic Acids Res.* 33: e178; Seligman y col. (2002), *Nucleic Acids Res.* 30: 3870-9) o, más recientemente, en tres posiciones en la secuencia de reconocimiento (Arnould y col. (2006), *J. Mol. Biol.* 355: 443-58). La superficie de separación de la proteína I-Crel-ADN contiene nueve aminoácidos que se ponen en contacto con las bases de ADN directamente y al menos cinco posiciones adicionales que pueden formar posibles contactos en superficies de separación modificadas. Los documentos WO 2004/067753 y WO 2006 y WO 2006/097853 describen enfoques combinatorios para alterar la especificidad de meganucleasas en aminoácidos directamente poniéndose en contacto con el sitio de unión a ADN. El tamaño de esta superficie de separación impone una complejidad combinatoria que es improbable que se muestre adecuadamente en bibliotecas de secuencias construidas para seleccionar enzimas con sitios de escisión drásticamente alterados.

Sigue existiendo la necesidad de nucleasas que facilitarán la precisa modificación de un genoma. Además, sigue existiendo la necesidad de técnicas para generar nucleasas con secuencias de reconocimiento racionalmente diseñadas predeterminadas que permitirán la manipulación de secuencias genéticas en sitios genéticos específicos y de técnicas que utilicen tales nucleasas para manipular genéticamente organismos con modificaciones de secuencia precisas.

RESUMEN DE LA INVENCIÓN

La presente invención se basa, en parte, en la identificación y caracterización de residuos de aminoácidos específicos en la familia LAGLIDADG de meganucleasas que hacen contactos con bases de ADN y el esqueleto de ADN cuando las meganucleasas se asocian con una secuencia de reconocimiento de ADN bicatenario, y así afectan la especificidad y actividad de las enzimas. Este descubrimiento se ha usado, como se describe en detalle más adelante, para identificar sustituciones de aminoácidos que pueden alterar la especificidad de secuencias de reconocimiento y/o afinidad de unión a ADN de las meganucleasas, y para diseñar racionalmente y desarrollar meganucleasas que pueden reconocer una secuencia de ADN deseada que no reconoce meganucleasas que se producen naturalmente. La invención también proporciona procedimientos que usan tales meganucleasas para producir ácidos nucleicos recombinantes y organismos utilizando las meganucleasas para producir recombinación de una secuencia genética deseada en un número limitado de sitios dentro del genoma del organismo, para terapia génica, para el tratamiento de infecciones patógenas y para aplicaciones *in vitro* en diagnóstico e investigación.

Así, en algunas realizaciones, la invención proporciona meganucleasas recombinantes que tienen especificidad alterada por al menos un medio sitio de la secuencia de reconocimiento con respecto a una meganucleasa I-Crel natural, en la que la meganucleasa incluye un polipéptido que tiene al menos el 85% de similitud de secuencias con los residuos 2-153 de la meganucleasa I-Crel natural de SEC ID N°: 1, pero en la que la meganucleasa recombinante tiene especificidad por un medio sitio de la secuencia de reconocimiento que se diferencia al menos una par de bases de un medio sitio dentro de una secuencia de reconocimiento de la meganucleasa I-Crel seleccionada de SEC ID N°: 2, SEC ID N°: 3, SEC ID N°: 4 y SEC ID N°: 5, y en la que la meganucleasa recombinante incluye al menos una modificación enumerada en la reivindicación 1.

Las meganucleasas de la invención pueden incluir una, dos, tres o más de las modificaciones que se han desvelado en el presente documento, como se ha reivindicado, con el fin de afectar la especificidad de secuencia de las

meganucleasas recombinantes en una, dos, tres o más posiciones dentro de la secuencia de reconocimiento. Las meganucleasas pueden incluir solo las modificaciones novedosas desveladas en el presente documento, o pueden incluir las novedosas modificaciones desveladas en el presente documento en combinación con modificaciones encontradas en la técnica anterior. Sin embargo, se excluyen específicamente meganucleasas recombinantes que comprendan solo las modificaciones de la técnica anterior.

Así, en un aspecto, se desvelan procedimientos de producción de una célula eucariota genéticamente modificada que incluye una secuencia exógena de interés insertada en un cromosoma, transfectando la célula con (i) una primera secuencia de ácidos nucleicos que codifica una meganucleasa de la invención, y (ii) una segunda secuencia de ácidos nucleicos que incluye dichas secuencias de interés, en los que la meganucleasa produce un sitio de escisión en el cromosoma y la secuencia de interés se inserta en el cromosoma en el sitio de escisión tanto por recombinación homóloga como unión de extremos no homólogos.

Alternativamente, en otro aspecto, se desvelan procedimientos para producir una célula eucariota genéticamente modificada que incluye una secuencia exógena de interés insertada en un cromosoma, introduciendo una proteína meganucleasa de la invención en la célula, y transfectando la célula con un ácido nucleico que incluye la secuencia de interés, en los que la meganucleasa produce un sitio de escisión en el cromosoma y la secuencia de interés se inserta en el cromosoma en el sitio de escisión tanto por recombinación homóloga como unión de extremos no homólogos.

En otro aspecto se desvelan procedimientos para producir una célula eucariota genéticamente modificada alterando una secuencia diana en un cromosoma, transfectando la célula con un ácido nucleico que codifica una meganucleasa de la invención, en el que la meganucleasa produce un sitio de escisión en el cromosoma y la secuencia diana es alterada por unión de extremos no homólogos en el sitio de escisión.

Estos y otros aspectos de la invención serán evidentes para un experto habitual en la materia basándose en la siguiente descripción detallada de la invención.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La Figura 1(A) ilustra las interacciones entre el homodímero I-CreI y su secuencia de reconocimiento bicatenaria que se produce naturalmente, basándose en datos cristalográficos. Esta representación esquemática representa la secuencia de reconocimiento (SEC ID N°: 2 y SEC ID N°: 3), mostrada desenrollada para fines de ilustración solo, unida por el homodímero, mostrado como dos óvalos. Las bases de cada medio sitio de ADN están numeradas -1 a -9, y los residuos de aminoácidos de I-CreI que forman la superficie de reconocimiento se indican por las designaciones de aminoácidos de una letra y números que indican la posición del residuo. Líneas negras continuas: enlaces de hidrógeno a bases de ADN. Líneas discontinuas: posiciones de aminoácidos que forman contactos adicionales en diseños de enzima, pero no se ponen en contacto con el ADN en el complejo natural. Flechas: residuos que interactúan con el esqueleto de ADN e influyen en la actividad de escisión.

La Figura 1(B) ilustra los contactos naturales entre el par de bases A-T en la posición -4 del medio sitio de escisión en el lado derecho de la Figura 1(A). Específicamente, se muestra que el residuo Q26 interactúa con la base A. El residuo 177 está en proximidad al par de bases, pero no interactúa específicamente.

La Figura 1(C) ilustra las interacciones entre una variante racionalmente diseñada de la meganucleasa I-CreI en la que el residuo 177 se ha modificado a E77. Como resultado de este cambio se prefiere un par de bases G-C en la posición -4. La interacción entre Q26 y la base G está mediada por una molécula de agua, como se ha observado cristalográficamente para el medio sitio de escisión en el lado izquierdo de la Figura 1(A).

La Figura 1(D) ilustra las interacciones entre una variante racionalmente diseñada de la meganucleasa I-CreI en la que el residuo Q26 se ha modificado a E26 y el residuo 177 se ha modificado a R77. Como resultado de este cambio se prefiere un par de bases C-G en la posición -4.

La Figura 1(E) ilustra las interacciones entre una variante racionalmente diseñada de la meganucleasa I-CreI en la que el residuo Q26 se ha modificado a A26 y el residuo 177 se ha modificado a Q77. Como resultado de este cambio se prefiere un par de bases T-A en la posición -4.

La Figura 2(A) muestra una comparación de una secuencia de reconocimiento para cada una de la meganucleasa natural I-CreI (WT) y 11 heterodímeros de meganucleasas racionalmente diseñadas de la invención. Las bases que están conservadas con respecto a la secuencia de reconocimiento WT están sombreadas. Los medios sitios de 9 pb están en negrita. WT: natural (SEC ID N°: 4); CF: alelo $\Delta F508$ del gen CFTR humano responsable de la mayoría de los casos de fibrosis quística (SEC ID N°: 25); MYD: el gen DM cinasa humano asociado a distrofia miotónica (SEC ID N°: 27); CCR: el gen CCR5 humano (un co-receptor importante del VIH) (SEC ID N°: 26); ACH: el gen FGFR3 humano correlacionado con acondroplasia (SEC ID N°: 23); TAT: el gen TAT/REV del VIH-1 (SEC ID N°: 15); HSV: el gen UL36 del VHS-1 (SEC ID N°: 28); LAM: el gen p05 del bacteriófago λ (SEC ID N°: 22); POX: el gen gp009 del virus variólico (viruela) (SEC ID N°: 30); URA: el gen URA3 de *Saccharomyces cerevisiae* (SEC ID N°: 36); GLA: el

gen GL2 de *Arabidopsis thaliana* (SEC ID N°: 32); BRP: el gen BP-1 de *Arabidopsis thaliana* (SEC ID N°: 33).

La Figura 2(B) ilustra los resultados de la incubación de cada uno de I-CreI natural (WT) y 11 heterodímeros de meganucleasas racionalmente diseñadas con plásmidos que alojan los sitios de reconocimiento para las 12 enzimas durante 6 horas a 37 °C. El porcentaje de escisión se indica en cada recuadro.

La Figura 3 ilustra patrones de escisión de homodímeros de I-CreI naturales y racionalmente diseñados. (A) I-CreI natural. (B) I-CreI K116D. (C-L) Meganucleasas racionalmente diseñadas de la invención. Se incubaron enzimas con un conjunto de plásmidos que alojan palíndromos del medio sitio de escisión previsto las 27 correspondientes a variaciones de pares de bases individuales. Los gráficos de barras muestran escisión fraccionaria (F) en 4 horas a 37 °C. Barras negras: patrones de escisión esperados basados en la Tabla 1. Barras grises: Sitios de ADN que se desvían de los patrones de escisión esperados. Círculos blancos indican bases en el sitio de reconocimiento previsto. También se muestran transcurso de tiempo de escisión durante dos horas. Las representaciones de transcurso de tiempo de círculos blancos en C y L se corresponden con la escisión por las enzimas CCR1 y BRP2 que carecen de la mutación E80Q. Los sitios de escisión se corresponden con los medio sitios de 5' (columna izquierda) y 3' (columna derecha) para las enzimas heterodímeras descritas en la Fig. 2(A).

DESCRIPCIÓN DETALLADA DE LA INVENCION

1.1 Introducción

La invención puede definirse por las reivindicaciones adjuntas. La presente invención se basa, en parte, en la identificación y caracterización de aminoácidos específicos en la familia LAGLIDADG de meganucleasas que hacen contactos específicos con bases de ADN y contactos no específicos con el esqueleto de ADN cuando las meganucleasas se asocian a una secuencia de reconocimiento de ADN bicatenario, y que así afectan la especificidad de secuencias de reconocimiento y afinidad de unión a ADN de las enzimas. Este descubrimiento se ha usado, como se describe en detalle más adelante, para identificar sustituciones de aminoácidos en las meganucleasas que pueden alterar la especificidad de las enzimas, y para diseñar racionalmente y desarrollar meganucleasas que pueden reconocer una secuencia de ADN deseada que no reconoce meganucleasas que se producen naturalmente. Finalmente, la invención proporciona usos para las meganucleasas racionalmente diseñadas en la producción de células recombinantes y organismos, además de en terapia génica, aplicaciones antipatógenas, anticancerígenas e *in vitro*, como se ha desvelado en el presente documento.

Como materia general, la invención proporciona procedimientos para generar meganucleasas LAGLIDADG racionalmente diseñadas que contienen residuos de aminoácidos alterados en sitios dentro de la meganucleasa que son responsables de (1) unión específica de secuencia a bases individuales en la secuencia de reconocimiento de ADN bicatenario, o (2) unión no específica de secuencia al esqueleto de fosfodiéster de una molécula de ADN bicatenario. Similarmente, alterar los aminoácidos que participan en la unión al esqueleto de ADN pueden alterar no solo la actividad de la enzima, sino también el grado de especificidad o degeneración de la unión a la secuencia de reconocimiento aumentando o disminuyendo la afinidad de unión global por el ADN bicatenario.

Como se describe en detalle más adelante, procedimientos de diseñar racionalmente meganucleasas incluyen la identificación de los aminoácidos responsables del reconocimiento/unión de ADN, y la aplicación de una serie de reglas para seleccionar cambios de aminoácidos apropiados. Con respecto a la especificidad de secuencias de meganucleasa, las reglas incluyen tanto consideraciones estéricas referentes a las distancias en un complejo de meganucleasa-ADN entre las cadenas laterales de aminoácidos de la meganucleasa y las bases en las hebras codificantes y no codificantes del ADN, y consideraciones referentes a las interacciones químicas no covalentes entre grupos funcionales de las cadenas laterales de aminoácidos y la base de ADN deseada en la posición relevante.

Finalmente, la mayoría de las meganucleasas naturales que se unen a ADN como homodímeros reconocen secuencias de reconocimiento pseudo- o completamente palindrómicas. Debido a que se espera que longitudinalmente los palíndromos sean raros, la probabilidad de encontrar una secuencia palindrómica en un sitio genómico de interés es extremadamente baja. Por consiguiente, si estas enzimas van a rediseñarse para reconocer sitios genómicos de interés, es necesario diseñar dos monómeros de enzima que reconozcan diferentes medios sitios que puedan heterodimerizarse para escindir la secuencia de reconocimiento híbrida no palindrómica. Por tanto, en algunos aspectos se desvelan meganucleasas racionalmente diseñadas en las que los monómeros que se diferencian al menos una posición de aminoácido se dimerizan para formar heterodímeros. En algunos casos, ambos monómeros se diseñan racionalmente para formar un heterodímero que reconoce una secuencia de reconocimiento no palindrómica. Una mezcla de dos monómeros diferentes puede producir hasta tres formas activas de dímero de meganucleasas: los dos homodímeros y el heterodímero. Además, o alternativamente, en algunos casos se alteran residuos de aminoácidos en las superficies de separación en las que los monómeros pueden interaccionar para formar dímeros, con el fin de aumentar o disminuir la probabilidad de formación de homodímeros o heterodímeros.

Así, en un aspecto, se desvelan procedimientos para diseñar racionalmente meganucleasas LAGLIDADG que contienen cambios de aminoácidos que alteran la especificidad y/o actividad de las enzimas. En otro aspecto, la

invención proporciona las meganucleasas racionalmente diseñadas resultantes de estos procedimientos. En otro aspecto, la invención proporciona procedimientos que usan tales meganucleasas racionalmente diseñadas para producir ácidos nucleicos recombinantes y organismos en los que una secuencia de ADN deseada o sitio genético dentro del genoma de un organismo se modifica por la inserción, delección, sustitución u otra manipulación de secuencias de ADN. En otro aspecto, las meganucleasas de la invención puede usarse para reducir la supervivencia de patógenos o células cancerosas usando meganucleasas racionalmente diseñadas que tienen secuencias de reconocimiento específicas para patógeno o específicas para cáncer.

1.2 Referencias y definiciones

Como se usa en el presente documento, el término “meganucleasa” se refiere a una endonucleasa que se une a ADN bicatenario en una secuencia de reconocimiento que es mayor que 12 pares de bases. Meganucleasas que se producen naturalmente pueden ser monoméricas (por ejemplo, I-SceI) o diméricas (por ejemplo, I-CreI). El término meganucleasa, como se usa en el presente documento, puede usarse para referirse a meganucleasas monoméricas, meganucleasas diméricas, o a los monómeros que se asocian para formar una meganucleasa dimérica. El término “endonucleasa de recirculación” es sinónimo del término “meganucleasa”.

Como se usa en el presente documento, el término “meganucleasa LAGLIDADG” se refiere tanto a meganucleasas que incluyen un único motivo LAGLIDADG, que son naturalmente diméricas, como a meganucleasas que incluyen dos motivos LAGLIDADG, que son naturalmente monoméricas. El término “meganucleasa mono-LAGLIDADG” se usa en el presente documento para referirse a meganucleasas que incluyen un único motivo LAGLIDADG, y el término “meganucleasa di-LAGLIDADG” se usa en el presente documento para referirse a meganucleasas que incluyen dos motivos LAGLIDADG, cuando sea necesario distinguir entre las dos. Cada uno de los dos dominios estructurales de una meganucleasa di-LAGLIDADG que incluye un motivo LAGLIDADG puede denominarse una subunidad LAGLIDADG.

Como se usa en el presente documento, el término “racionalmente diseñada” significa que no se produce naturalmente y/o genéticamente manipulada. Las meganucleasas racionalmente diseñadas de la invención se diferencian de meganucleasas naturales o que se producen naturalmente en su secuencia de aminoácidos o estructura primaria, y pueden también diferenciarse en su estructura secundaria, terciaria o cuaternaria. Además, las meganucleasas racionalmente diseñadas de la invención también se diferencian de meganucleasas naturales o que se producen naturalmente en la especificidad y/o actividad de secuencias de reconocimiento.

Como se usa en el presente documento, con respecto a una proteína, el término “recombinante” significa que tiene una secuencia de aminoácidos alterada como resultado de la aplicación de técnicas de ingeniería genética a ácidos nucleicos que codifican la proteína, y células u organismos que expresan la proteína. Con respecto a un ácido nucleico, el término “recombinante” significa que tiene una secuencia de ácidos nucleicos alterada como resultado de la aplicación de técnicas de ingeniería genética. Las técnicas de ingeniería genética incluyen, pero no se limitan a, PCR y tecnologías de clonación de ADN; transfección, transformación y otras tecnologías de transferencia de genes; recombinación homóloga; mutagénesis dirigida al sitio; y fusión de genes. Según esta definición, una proteína que tiene una secuencia de aminoácidos idéntica a una proteína que se produce naturalmente, pero producida por clonación y expresión en un huésped heterólogo, no se considera recombinante.

Como se usa en el presente documento con respecto a proteínas recombinantes, el término “modificación” significa cualquier inserción, delección o sustitución de un residuo de aminoácido en la secuencia recombinante con respecto a una secuencia de referencia (por ejemplo, una natural).

Como se usa en el presente documento, el término “genéticamente modificado” se refiere a una célula u organismo en el que, o en un antepasado del que, una secuencia de ADN genómica se ha modificado deliberadamente por tecnología recombinante. Como se usa en el presente documento, el término “genéticamente modificado” engloba el término “transgénico”.

Como se usa en el presente documento, el término “natural” se refiere a cualquier forma que se produzca naturalmente de una meganucleasa. El término “natural” no pretende significar la variante alélica más común de la enzima en la naturaleza sino, más bien, cualquier variante alélica encontrada en la naturaleza. Las meganucleasas naturales se distinguen de meganucleasas recombinantes o que no se producen naturalmente.

Como se usa en el presente documento, el término “medio sitio de la secuencia de reconocimiento” o simplemente “medio sitio” significa una secuencia de ácidos nucleicos en una molécula de ADN bicatenario que es reconocida por un monómero de una meganucleasa mono-LAGLIDADG o por una subunidad LAGLIDADG de una meganucleasa di-LAGLIDADG.

Como se usa en el presente documento, el término “secuencia de reconocimiento” se refiere a un par de medios sitios que está unido y se escinde por tanto un dímero de meganucleasa mono-LAGLIDADG o un monómero de meganucleasa di-LAGLIDADG. Los dos medios sitios pueden o pueden no separarse por pares de bases que no son específicamente reconocidos por la enzima. En los casos de I-CreI, I-MsoI y I-CeuI, el medio sitio de la secuencia de

reconocimiento de cada monómero se extiende 9 pares de bases, y los dos medios sitios se separan cuatro pares de bases que no son reconocidos específicamente, pero que constituyen el sitio de escisión real (que tiene un nucleótido protuberante de 4 pares de bases). Así, las secuencias de reconocimiento combinadas de dímeros de meganucleasas I-Crel, I-MsoI y I-CeuI normalmente se extienden 22 pares de bases, que incluyen dos medios sitios de 9 pares de bases que flanquean un sitio de escisión de 4 pares de bases. Los pares de bases de cada medio sitio se designan -9 a -1, siendo la posición -9 la más distal desde el sitio de escisión y estando la posición -1 adyacente a los 4 pares de bases centrales, que se designan N₁-N₄. La hebra de cada medio sitio que está orientada 5' a 3' en la dirección de -9 a -1 (es decir, hacia el sitio de escisión) se designa la hebra "codificante" y la hebra opuesta se designa la "hebra no codificante", aunque ninguna hebra puede codificar proteína. Así, la hebra "codificante" de un medio sitio es la hebra no codificante del otro medio sitio. Véase, por ejemplo, la Figura 1 (A). En el caso de la meganucleasa I-SceI, que es un monómero de meganucleasa di-LAGLIDADG, la secuencia de reconocimiento es una secuencia no palindrómica de aproximadamente 18 pb, y no hay pares de bases centrales que no sean específicamente reconocidos. Por comodidad, una de las dos hebras se denomina la hebra "codificante" y la otra la hebra "no codificante", aunque ninguna hebra puede codificar proteína.

Como se usa en el presente documento, el término "especificidad" significa la capacidad de una meganucleasa para reconocer y escindir moléculas de ADN bicatenario solo en una secuencia de pares de bases particular denominada la secuencia de reconocimiento, o solo en un conjunto particular de secuencias de reconocimiento. El conjunto de secuencias de reconocimiento compartirá ciertas posiciones conservadas o motivos de secuencia, pero puede degenerar en una o más posiciones. Una meganucleasa altamente específica puede escindir solo una o muy pocas secuencias de reconocimiento. La especificidad puede determinarse en un ensayo de escisión como se describe en el Ejemplo 1. Como se usa en el presente documento, una meganucleasa tiene especificidad "alterada" si se une a y escinde una secuencia de reconocimiento que no está unida a y se escinde por una meganucleasa de referencia (por ejemplo, una natural) o si la tasa de escisión de una secuencia de reconocimiento aumenta o disminuye una cantidad estadísticamente significativa ($p < 0,05$) con respecto a una meganucleasa de referencia.

Como se usa en el presente documento, el término "degeneración" significa lo opuesto a "especificidad." Una meganucleasa altamente degenerada puede escindir un gran número de secuencias de reconocimiento divergentes. Una meganucleasa puede tener degeneración de secuencias en una única posición dentro de un medio sitio o en múltiples posiciones, incluso todas, dentro de un medio sitio. Tal degeneración de secuencias puede resultar de (i) la incapacidad de cualquier aminoácido en el dominio de unión a ADN de una meganucleasa para hacer un contacto específico con cualquier base en una o más posiciones en la secuencia de reconocimiento, (ii) la capacidad de uno o más aminoácidos en el dominio de unión a ADN de una meganucleasa para hacer contactos específicos con más de una base en una o más posiciones en la secuencia de reconocimiento, y/o (iii) afinidad de unión a ADN suficiente no específica para actividad. Una posición "completamente" degenerada puede estar ocupada por cualquiera de las cuatro bases y puede designarse con "N" en un medio sitio. Una posición "parcialmente" degenerada puede estar ocupada por dos o tres de las cuatro bases (por ejemplo, cualquier purina (Pu), cualquier pirimidina (Py), o no G).

Como se usa en el presente documento con respecto a meganucleasas, el término "afinidad de unión a ADN" o "afinidad de unión" significa la tendencia de una meganucleasa a asociarse no covalentemente con una molécula de ADN de referencia (por ejemplo, una secuencia de reconocimiento o una secuencia arbitraria). La afinidad de unión se mide por una constante de disociación, K_D (por ejemplo, la K_D de I-Crel para la secuencia de reconocimiento WT es aproximadamente 0,1 nM). Como se usa en el presente documento, una meganucleasa tiene afinidad de unión "alterada" si la K_D de la meganucleasa recombinante para una secuencia de reconocimiento de referencia aumenta o disminuye una cantidad estadísticamente significativa ($p < 0,05$) con respecto a una meganucleasa de referencia.

Como se usa en el presente documento con respecto a monómeros de meganucleasa, el término "afinidad por formación de dímeros" significa la tendencia de un monómero de meganucleasa para asociarse no covalentemente con un monómero de meganucleasa de referencia. La afinidad para la formación de dímeros puede medirse con el mismo monómero (es decir, formación de homodímeros) o con un monómero diferente (es decir, formación de heterodímeros) tal como una meganucleasa natural de referencia. La afinidad de unión se mide por una constante de disociación, K_D . Como se usa en el presente documento, una meganucleasa tiene afinidad "alterada" por formación de dímeros si la K_D del monómero de meganucleasa recombinante para un monómero de meganucleasa de referencia aumenta o disminuye una cantidad estadísticamente significativa ($p < 0,05$) con respecto a un monómero de meganucleasa de referencia.

Como se usa en el presente documento, el término "palindrómica" se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de medios sitios idénticos. En este caso, sin embargo, la secuencia palindrómica no necesita ser palindrómica con respecto a los cuatro pares de bases centrales, que no son contactados por la enzima. En el caso de meganucleasas diméricas, las secuencias de ADN palindrómicas son reconocidas por homodímeros en los que los dos monómeros hacen contactos con medios sitios idénticos.

Como se usa en el presente documento, el término "pseudo-palindrómica" se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de medios sitios palindrómicos no idénticos o imperfectamente palindrómicos. En este caso, la secuencia pseudo-palindrómica no solo no necesita ser palindrómica con respecto a los cuatro pares de bases centrales, sino que también puede desviarse de una

secuencia palindrómica entre los dos medios sitios. Las secuencias de ADN pseudo-palindrómicas son típicas de los sitios de ADN naturales reconocidos por meganucleasas homodiméricas naturales en las que dos monómeros de enzima idénticos hacen contactos con diferentes medios sitios.

Como se usa en el presente documento, el término “no palindrómica” se refiere a una secuencia de reconocimiento compuesta por dos medio sitios no relacionados de una meganucleasa. En este caso, la secuencia no palindrómica no necesita ser palindrómica con respecto a ninguno de los cuatro pares de bases centrales o los dos medios sitios monómeros. Las secuencias de ADN no palindrómicas son reconocidas por cualquier meganucleasa di-LAGLIDADG, meganucleasas mono-LAGLIDADG altamente degeneradas (por ejemplo, I-CeuI) o por heterodímeros de monómeros de meganucleasa mono-LAGLIDADG que reconocen medios sitios no idénticos.

Como se usa en el presente documento, el término “actividad” se refiere a la tasa a la que una meganucleasa de la invención escinde una secuencia de reconocimiento particular. Tal actividad es una reacción enzimática medible, que implica la hidrólisis de enlaces fosfodiéster de ADN bicatenario. La actividad de una meganucleasa que actúa sobre un sustrato de ADN particular es afectada por la afinidad o aversez de la meganucleasa para ese sustrato de ADN particular que está, a su vez, afectado por tanto interacciones específicas de secuencia como específicas de no secuencia con el ADN.

Como se usa en el presente documento, el término “recombinación homóloga” se refiere al procedimiento celular natural en el que una rotura de ADN bicatenaria se repara usando una secuencia de ADN homóloga como molde de reparación (véase, por ejemplo, Cahill y col. (2006), *Front. Biosci.* 11:1958-1976). La secuencia de ADN homóloga puede ser una secuencia cromosómica endógena o un ácido nucleico exógeno se administró a la célula. Así, en algunas realizaciones, se usa una meganucleasa racionalmente diseñada para escindir una secuencia de reconocimiento dentro de una secuencia diana y un ácido nucleico exógeno con homología por o similitud de secuencias sustancial con la secuencia diana se administra a la célula y se usa como molde para reparar por recombinación homóloga. La secuencia de ADN del ácido nucleico exógeno, que puede diferenciarse significativamente de la secuencia diana, se incorpora así en la secuencia cromosómica. El procedimiento de recombinación homóloga se produce principalmente en organismos eucariotas. El término “homología” se usa en el presente documento como equivalente a “similitud de secuencias” y no pretende requerir identidad por vinculación del linaje o filogenética.

Como se usa en el presente documento, el término “unión de extremos no homólogos” se refiere al proceso celular natural en el que una rotura de ADN bicatenario se repara por la unión directa de dos segmentos de ADN no homólogos (véase, por ejemplo, Cahill y col. (2006), *Front. Biosci.* 11:1958-1976). La reparación de ADN por unión de extremos no homólogos es propensa a error y frecuentemente produce la adición sin molde o delección de secuencias de ADN en el sitio de reparación. Así, en ciertas realizaciones, una meganucleasa racionalmente diseñada puede usarse para producir una rotura bicatenaria en una secuencia de reconocimiento de meganucleasas dentro de una secuencia diana para alterar un gen (por ejemplo, introduciendo inserciones de bases, delecciones de bases o mutaciones por desplazamiento del marco) por unión de extremos no homólogos. En otras realizaciones, un ácido nucleico exógeno que carece homología con o similitud de secuencias sustancial con la secuencia diana puede capturarse en el sitio de una rotura de ADN bicatenario estimulada por meganucleasa por unión de extremos no homólogos (véase, por ejemplo, Salomon, y col. (1998), *EMBO J.* 17:6086-6095). El procedimiento de unión de extremos no homólogos se produce en tanto eucariotas como procariotas tales como bacterias.

Como se usa en el presente documento, el término “secuencia de interés” significa cualquier secuencia de ácidos nucleicos, tanto si codifica como si no una proteína, ARN o elemento regulador (por ejemplo, una secuencia de potenciador, silenciador o de promotor), que puede insertarse en un genoma o usarse para sustituir una secuencia de ADN genómica usando una proteína meganucleasa. Las secuencias de interés pueden tener secuencias de ADN heterólogas que permiten marcar una proteína o ARN que se expresa de la secuencia de interés. Por ejemplo, una proteína puede marcarse con marcas que incluyen, pero no se limitan a, un epítipo (por ejemplo, c-myc, FLAG) u otro ligando (por ejemplo, poli-His). Además, una secuencia de interés puede codificar una proteína de fusión, según técnicas conocidas en la técnica (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, Wiley 1999). En algunas realizaciones, la secuencia de interés está flanqueada por una secuencia de ADN que es reconocida por la meganucleasa recombinante para la escisión. Así, las secuencias flanqueantes se escinden permitiendo la apropiada inserción de la secuencia de interés en secuencias de reconocimiento genómicas escindidas por la meganucleasa recombinante. En algunas realizaciones, la secuencia entera de interés es homóloga a o tiene similitud de secuencias sustancial con una secuencia diana en el genoma de forma que la recombinación homóloga sustituya eficazmente la secuencia diana con la secuencia de interés. En otras realizaciones, la secuencia de interés está flanqueada por secuencias de ADN con homología a o similitud de secuencias sustancial con la secuencia diana de forma que la recombinación homóloga inserte la secuencia de interés dentro del genoma en el sitio de la secuencia diana. En algunas realizaciones, la secuencia de interés es sustancialmente idéntica a la secuencia diana, excepto por mutaciones u otras modificaciones en la secuencia de reconocimiento de meganucleasa de forma que la meganucleasa no pueda escindir la secuencia diana después de haberse modificado por la secuencia de interés.

Como se usa en el presente documento con respecto a tanto secuencias de aminoácidos como secuencias de

ácidos nucleicos, los términos “porcentaje de similitud” y “similitud de secuencias” se refieren a una medida del grado de similitud de dos secuencias basándose en un alineamiento de las secuencias que maximiza la similitud entre residuos de aminoácidos alineados o nucleótidos, y que es una función del número de residuos idénticos o similares o nucleótidos, el número de residuos o nucleótidos total, y la presencia y longitud de huecos en el alineamiento de secuencias. Está disponible una variedad de algoritmos y programas informáticos para determinar la similitud de secuencias usando parámetros estándar. Como se usa en el presente documento, la similitud de secuencias se mide usando el programa BLASTp para secuencias de aminoácidos y el programa BLASTn para secuencias de ácidos nucleicos, ambas de las cuales están disponibles del Centro Nacional para Información Biotecnológica (www.ncbi.nlm.nih.gov/), y se describen en, por ejemplo, Altschul y col. (1990), J. Mol. Biol. 215:403-410; Gish y States (1993), Nature Genet. 3:266-272; Madden y col. (1996), Meth. Enzymol. 266:131-141; Altschul y col. (1997), Nucleic Acids Res. 25:33 89-3402; Zhang y col. (2000), J. Comput. Biol. 7(1-2):203-14. Como se usa en el presente documento, el porcentaje de similitud de dos secuencias de aminoácidos es la puntuación basada en los siguientes parámetros para el algoritmo BLASTp: tamaño de palabra = 3; penalización por apertura de hueco = -11; penalización por extensión de hueco = -1; y matriz de puntuación = BLOSUM62. Como se usa en el presente documento, el porcentaje de similitud de dos secuencias de ácidos nucleicos es la puntuación basada en los siguientes parámetros para el algoritmo BLASTn: tamaño de palabra = 11; penalización por apertura de hueco = -5; penalización por extensión de hueco = -2; recompensa por coincidencia = 1; y penalización por falta de coincidencia = -3.

Como se usa en el presente documento con respecto a modificaciones de dos proteínas o secuencias de aminoácidos, el término “correspondiente a” se usa para indicar que una modificación especificada en la primera proteína es una sustitución del mismo residuo de aminoácido que en la modificación en la segunda proteína, y que la posición de aminoácido de la modificación en las primeras proteínas se corresponde con o alinea con la posición de aminoácido de la modificación en la segunda proteína cuando las dos proteínas se someten a alineamientos de secuencias convencionales (por ejemplo, usando el programa BLASTp). Así, la modificación del residuo “X” al aminoácido “A” en la primera proteína se corresponderá con la modificación del residuo “Y” al aminoácido “A” en la segunda proteína si los residuos X y Y se corresponden entre sí en un alineamiento de secuencias, y a pesar del hecho de que X y Y puedan ser números diferentes.

Como se usa en el presente documento, la recitación de un intervalo numérico para una variable pretende expresar que la invención puede ponerse en práctica con la variable igual a cualquiera de los valores dentro de ese intervalo. Así, para una variable que es inherentemente discreta, la variable puede ser igual a cualquier valor de número entero dentro del intervalo numérico, que incluye los extremos del intervalo. Similarmente, para una variable que es inherentemente continua, la variable puede ser igual a cualquier valor real dentro del intervalo numérico, que incluye los extremos del intervalo. Como ejemplo, y sin limitación, una variable que se describe como que tiene valores entre 0 y 2 puede tomar los valores 0, 1 ó 2 si la variable es inherentemente discreta, y puede tomar los valores 0,0, 0,1, 0,01, 0,001, o cualquier otro valor real ≥ 0 y ≤ 2 si la variable es inherentemente continua.

Como se usa en el presente documento, a menos que se indique específicamente de otro modo, la palabra “o” se usa en el sentido inclusivo de “y/o” y no el sentido exclusivo de “cualquiera/o.”

2.1 Meganucleasas racionalmente diseñadas con especificidad de secuencias alterada

Se desvelan procedimientos para diseñar racionalmente meganucleasas de la familia LAGLIDADG recombinantes. En este aspecto, las meganucleasas recombinantes se diseñan racionalmente prediciendo primero sustituciones de aminoácidos que pueden alterar la preferencia de bases en cada posición en el medio sitio. Estas sustituciones pueden validarse experimentalmente individualmente o en combinaciones para producir meganucleasas con la especificidad de escisión deseada.

Según la invención, las sustituciones de aminoácidos que pueden producir un cambio deseado en la preferencia de bases se predicen determinando las cadenas laterales de aminoácidos de una meganucleasa de referencia (por ejemplo, una meganucleasa natural, o una meganucleasa de referencia que no se produce naturalmente) que puede participar en hacer contactos con las bases de ácido nucleico de la secuencia de reconocimiento de meganucleasa de ADN y el esqueleto de fosfodiéster de ADN, y la naturaleza espacial y química de aquellos contactos. Estos aminoácidos incluyen, pero no se limitan a, cadenas laterales que facilitan la puesta en contacto del medio sitio de ADN de referencia. Generalmente, esta determinación requiere tener conocimiento de la estructura del complejo entre la meganucleasa y su secuencia de reconocimiento de ADN bicatenario, o conocimiento de la estructura de un complejo altamente similar (por ejemplo, entre la misma meganucleasa y una secuencia de reconocimiento de ADN alternativa, o entre una variante alélica o filogenética de la meganucleasa y su secuencia de reconocimiento de ADN).

Las estructuras tridimensionales, como se describen por datos de coordenadas atómicas, de un polipéptido o complejo de dos o más polipéptidos pueden obtenerse de varias formas. Por ejemplo, las determinaciones de la estructura de proteínas pueden prepararse usando técnicas que incluyen, pero no se limitan a, cristalografía de rayos X, RMN y espectrometría de masas. Otro enfoque es para analizar bases de datos de coordenadas estructurales existentes para la meganucleasa de interés o una meganucleasa relacionada. Tales datos

estructurales están frecuentemente disponibles de bases de datos en forma de coordenadas tridimensionales. Frecuentemente, estos datos están accesibles mediante bases de datos en línea (por ejemplo, la base de datos de proteínas RCSB en www.rcsb.org/pdb).

La información estructural puede obtenerse experimentalmente analizando los patrones de difracción de, por ejemplo, rayos X o electrones, creados por matrices bi o tridimensionales regulares (por ejemplo, cristales) de proteínas o complejos de proteína. Se usan procedimientos computacionales para transformar los datos de difracción en coordenadas atómicas tridimensionales en el espacio. Por ejemplo, el campo de la cristalografía de rayos X se ha usado para generar información estructural tridimensional sobre muchos complejos de proteína-ADN, que incluyen meganucleasas (véase, por ejemplo, Chevalier y col. (2001), *Nucleic Acids Res.* 29(18): 3757-3774).

También se ha usado resonancia magnética nuclear (RMN) para determinar distancias interatómicas de moléculas en disolución. Los procedimientos de RMN multidimensional combinados con procedimientos computacionales han tenido éxito en determinar las coordenadas atómicas de polipéptidos de tamaño creciente (véase, por ejemplo, Tzakos y col. (2006), *Annu. Rev. Biophys. Biomol. Struct.* 35:19-42.).

Alternativamente, puede usarse modelado computacional aplicando algoritmos basados en las estructuras primarias conocidas y, cuando estén disponibles, estructuras secundarias, terciarias y/o cuaternarias de la proteína/ADN, además de la naturaleza fisicoquímica conocida de las cadenas laterales de aminoácidos, bases de ácidos nucleicos e interacciones de enlace. Tales procedimientos pueden incluir opcionalmente enfoques iterativos, o limitaciones experimentalmente derivadas. Un ejemplo de tal software computacional es el programa CNS descrito en Adams y col. (1999), *Acta Crystallogr. D. Biol. Crystallogr.* 55 (Pt 1): 181-90. Se ha desarrollado una variedad de otros programas computacionales que predicen la disposición espacial de aminoácidos en una estructura de proteína y predicen la interacción de las cadenas laterales de aminoácidos de la proteína con diversas moléculas diana (véase, por ejemplo, la patente de EE.UU. n° 6.988.041).

Así, en algunas realizaciones de la invención, se usan modelos computacionales para identificar residuos de aminoácidos específicos que interaccionan específicamente con bases de ácidos nucleicos de ADN y/o facilitan interacciones de esqueleto fosfodiéster no específicas. Por ejemplo, pueden producirse modelos informáticos de la totalidad de la posible interacción de meganucleasa-ADN usando un programa de software adecuado que incluye, pero no se limita a, MOLSCRIPT™ 2.0 (Avatar Software AB, Estocolmo, Suecia), el programa de representación gráfica O (Jones y col. (1991), *Acta Crystallography*, A47: 110), el programa de representación gráfica GRASP™ (Nicholls y col., (1991), *PROTEINS, Structure, Function and Genetics* 11 (4): 281ff), o el programa de representación gráfica INSIGHT™ (TSI, Inc., Shoreview, MN). Hardware informático adecuado para producir, visualizar y manipular representaciones de estructuras tridimensionales de complejos de proteína-ADN están comercialmente disponibles y son muy conocidos en la técnica (por ejemplo, Silicon Graphics Workstation, Silicon Graphics, Inc., Mountainview, CA).

Específicamente, las interacciones entre una meganucleasa y sus secuencias de reconocimiento de ADN bicatenario pueden resolverse usando procedimientos conocidos en la técnica. Por ejemplo, una representación, o modelo, de la estructura tridimensional de una estructura compleja multicomponente, para la que se ha producido un cristal, puede determinarse usando técnicas que incluyen sustitución molecular o SIR/MIR (sustitución isomorfa individual/múltiple) (véase, por ejemplo, Brunger (1997), *Meth. Enzym.* 276: 558-580; Navaza y Saludjian (1997), *Meth. Enzym.* 276: 581-594; Tong y Rossmann (1997), *Meth. Enzym.* 276: 594-611; y Bentley (1997), *Meth. Enzym.* 276: 611-619) y puede realizarse usando un programa de software tal como AMoRe/Mosflm (Navaza (1994), *Acta Cryst.* A50: 157-163; CCP4 (1994), *Acta Cryst.* D50: 760-763) o XPLOR (véase, Brünger y col. (1992), *X-PLOR versión 3.1. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT).

La determinación de estructura de proteínas y posible interacción de meganucleasa-ADN permite elecciones racionales referentes a los aminoácidos que pueden cambiarse para afectar la actividad y especificidad enzimática. Las decisiones se basan en varios factores referentes a las interacciones de cadenas laterales de aminoácidos con una base particular o esqueleto de fosfodiéster de ADN. Las interacciones químicas usadas para determinar sustituciones de aminoácidos apropiadas incluyen, pero no se limitan a, fuerzas de van der Waals, impedimento estérico, enlace iónico, enlace de hidrógeno e interacciones hidrófobas. Pueden seleccionarse sustituciones de aminoácidos que tanto favorecen como desfavorecen interacciones específicas de la meganucleasa con una base particular en un posible medio sitio de la secuencia de reconocimiento con el fin de aumentar o disminuir la especificidad por esa secuencia y, a algún grado, afinidad de unión global y actividad. Además, pueden seleccionarse sustituciones de aminoácidos que tanto aumentan como disminuyen la afinidad de unión por el esqueleto de fosfodiéster de ADN bicatenario con el fin de aumentar o disminuir la actividad global y, a algún grado, para reducir o aumentar la especificidad.

Así, en realizaciones específicas, se determina una estructura tridimensional de un complejo de meganucleasa-ADN y se define una "superficie de contacto" para cada par de base en un medio sitio de la secuencia de reconocimiento de ADN. En algunas realizaciones, la superficie de contacto comprende aquellos aminoácidos en la enzima con β -carbonos inferiores a 9,0 Å de un donante o aceptor de hidrógeno del surco mayor en cualquier base en el par, y con cadenas laterales orientadas hacia el ADN, independientemente de si los residuos hacen contactos de base o no en

el complejo de meganucleasa natural-ADN. En otras realizaciones, los residuos pueden excluirse si los residuos no hacen contacto en el complejo de meganucleasa natural-ADN, o pueden incluirse o excluirse residuos a criterio del diseñador para alterar el número o identidad de los residuos considerados. En un ejemplo, como se describe más adelante, para posiciones de base -2, -7, -8 y -9 del medio sitio de I-Crel natural, las superficies de contacto se limitaron a las posiciones de aminoácido que en realidad interactúan en el complejo de enzima natural-ADN. Para las posiciones -1, -3, -4 -5, y -6, sin embargo, se definió que las superficies de contacto contenían posiciones de aminoácidos adicionales que no participaban en contactos naturales, pero que podrían contactar potencialmente con una base si se sustituyeran con un aminoácido diferente.

Debe observarse, aunque un medio sitio de la secuencia de reconocimiento se representa normalmente con respecto a solo una hebra de ADN, que las meganucleasas se unen en el surco mayor de ADN bicatenario, y hacen contacto con bases de ácidos nucleicos en ambas hebras. Además, las designaciones de hebra "codificante" y "no codificante" son completamente arbitrarias con respecto a la unión y reconocimiento de meganucleasas. Las especificidad por secuencia en una posición puede lograrse tanto mediante interacciones con un miembro de un par de bases como por una combinación de interacciones con ambos miembros de un par de bases. Así, por ejemplo, con el fin de favorecer la presencia de un par de bases A/T en la posición X, en la que la base A está sobre la hebra "codificante" y la base T está sobre la hebra "no codificante", se seleccionan residuos que están suficientemente próximos para poner en contacto la hebra codificante en la posición X y que favorecen la presencia de una A, y/o se seleccionan residuos que están suficientemente próximos para poner en contacto la hebra no codificante en la posición X y que favorecen la presencia de una T. Según la invención, un residuo se considera suficientemente próximo si el β -carbono del residuo está dentro de 9 Å del átomo más próximo de la base relevante.

Así, por ejemplo, un aminoácido con un β -carbono dentro de 9 Å de la hebra codificante de ADN, pero superior a 9 Å de la hebra no codificante, se considera para posibles interacciones con solo la hebra codificante. Similarmente, un aminoácido con un β -carbono dentro de 9 Å de la hebra no codificante de ADN, pero superior a 9 Å de la hebra codificante, se considera para posibles interacciones con solo la hebra no codificante. Los aminoácidos con β -carbonos que están dentro de 9 Å de ambas hebras de ADN se consideran para posibles interacciones con cualquier hebra.

Para cada superficie de contacto, las posibles sustituciones de aminoácidos se seleccionan basándose en su capacidad predicha para interactuar favorablemente con una o más de las cuatro bases de ADN. El procedimiento de selección se basa en dos criterios primarios: (i) el tamaño de las cadenas laterales del aminoácido, que afectará sus interacciones estéricas con diferentes bases de ácidos nucleicos, y (ii) la naturaleza química de las cadenas laterales del aminoácido, que afectará sus interacciones electrostáticas y de unión con las diferentes bases de ácidos nucleicos.

Con respecto al tamaño de las cadenas laterales, pueden seleccionarse aminoácidos con cadenas laterales más cortas y/o más pequeñas si un β -carbono de aminoácido en una superficie de contacto está <6 Å de una base, y pueden seleccionarse aminoácidos con cadenas laterales más largas y/o mayores si un β -carbono de aminoácido en una superficie de contacto está >6 Å de una base. Los aminoácidos con cadenas laterales que son intermedias en tamaño pueden seleccionarse si un β -carbono de aminoácido en una superficie de contacto es 5-8 Å de una base.

Los aminoácidos con cadenas laterales relativamente más cortas y más pequeñas pueden asignarse al Grupo 1, que incluye glicina (G), alanina (A), serina (S), treonina (T), cisteína (C), valina (V), leucina (L), isoleucina (I), aspartato (D), asparagina (N) y prolina (P). Sin embargo, se espera que la prolina vaya a usarse menos frecuentemente debido a su inflexibilidad relativa. Además, se espera que la glicina vaya a usarse menos frecuentemente debido a que introduce flexibilidad no deseada en el esqueleto de péptido y su tamaño muy pequeño reduce la probabilidad de contactos eficaces cuando sustituye un residuo mayor. Por otra parte, puede usarse glicina en algunos casos para promover una posición degenerada. Los aminoácidos con cadenas laterales de longitud y tamaño relativamente intermedio pueden asignarse al Grupo 2, que incluye lisina (K), metionina (M), arginina (R), glutamato (E) y glutamina (Q). Los aminoácidos con cadenas laterales relativamente más largas y/o mayores pueden asignarse al grupo 3, que incluyen lisina (K), metionina (M), arginina (R), histidina (H), fenilalanina (F), tirosina (Y) y triptófano (W). Sin embargo, se espera que el triptófano vaya a usarse menos frecuentemente debido a su inflexibilidad relativa. Además, la flexibilidad de la cadena lateral de lisina, arginina y metionina permite que estos aminoácidos hagan contactos de bases de distancias largas o intermedias, garantizando su inclusión en ambos grupos 2 y 3. Estos grupos también se muestran en forma tabulada a continuación:

Grupo 1	Grupo 2	Grupo 3
glicina (G)	glutamina (Q)	arginina (R)
alanina (A)	glutamato (E)	histidina (H)
serina (S)	lisina (K)	fenilalanina (F)
treonina (T)	metionina (M)	tirosina (Y)
cisteína (C)	arginina (R)	triptófano (W)
valina (V)		lisina (K)
leucina (L)		metionina (M)
isoleucina (I)		
aspartato (D)		
asparagina (N)		
prolina (P)		

Con respecto a la naturaleza química de las cadenas laterales, los diferentes aminoácidos se evalúan para sus posibles interacciones con las diferentes bases de ácidos nucleicos (por ejemplo, fuerzas de van der Waals, enlace iónico, enlace de hidrógeno e interacciones hidrófobas) y se seleccionan residuos que tanto favorecen como desfavorecen interacciones específicas de la meganucleasa con una base particular en una posición particular en el medio sitio de la secuencia de reconocimiento de ADN bicatenario. En algunos casos puede desearse crear un medio sitio con una o más posiciones degeneradas completas o parciales. En tales casos pueden elegirse residuos que favorecen la presencia de dos o más bases, o residuos que desfavorecen una o más bases. Por ejemplo, puede lograrse reconocimiento de bases degeneradas parciales impidiendo estéricamente una pirimidina en una posición sentido o antisentido.

El reconocimiento de bases de guanina (G) se logra usando aminoácidos con cadenas laterales básicas que forman enlaces de hidrógeno con N7 y O6 de la base. La especificidad de citosina (C) se confiere por cadenas laterales negativamente cargadas que interaccionan desfavorablemente con los grupos electronegativos del surco mayor presentes en todas las bases, excepto C. El reconocimiento de la timina (T) se diseña racionalmente usando interacciones hidrófobas y de van der Waals entre cadenas laterales hidrófobas y el grupo metilo del surco mayor en la base. Finalmente, se reconocen bases de adenina (A) usando las cadenas laterales de carboxamida Asn y Gln o la cadena lateral de hidroxilo de Tyr mediante un par de enlaces de hidrógeno a N7 y N6 de la base. Por último lugar, la His puede usarse para conferir especificidad por una base de purina (A o G) donando un enlace de hidrógeno a N7. Estas reglas directas para el reconocimiento de ADN pueden aplicarse a superficies de contacto predichas en las que una o ambas de las bases en una posición de par de bases particular son reconocidas mediante un contacto racionalmente diseñado.

Así, basándose en sus interacciones de unión con las diferentes bases de ácidos nucleicos, y las bases que se favorecen en una posición con la que hacen contacto, cada residuo de aminoácido puede asignarse a uno o más grupos diferentes correspondientes a las diferentes bases que favorecen (es decir, G, C, T o A). Así, el grupo G incluye arginina (R), lisina (K) e histidina (H); el grupo C incluye aspartato (D) y glutamato (E); el grupo T incluye alanina (A), valina (V), leucina (L), isoleucina (I), cisteína (C), treonina (T), metionina (M) y fenilalanina (F); y el grupo A incluye asparagina (N), glutamina (N), tirosina (Y) e histidina (H). Obsérvese que la histidina aparece en tanto el grupo G como el grupo A; que la serina (S) no está incluida en ningún grupo, pero que puede usarse para favorecer una posición degenerada; y que la prolina, glicina y triptófano no están incluidos en ningún grupo particular debido a consideraciones estéricas predominantes. Estos grupos también se muestran en forma tabulada a continuación:

Grupo G	Grupo C	Grupo T	Grupo A
arginina (R)	aspartato (D)	alanina (A)	asparagina (N)
lisina (K)	glutamato (E)	valina (V)	glutamina (Q)
histidina (H)		leucina (L)	tirosina (Y)
		isoleucina (I)	histidina (H)
		cisteína (C)	
		treonina (T)	
		metionina (M)	
		fenilalanina (F)	

Así, según la invención, con el fin de efectuar un cambio deseado en el medio sitio de la secuencia de reconocimiento de una meganucleasa en una posición X dada, (1) determinar al menos la porción relevante de la estructura tridimensional del complejo de meganucleasa natural o de referencia-ADN y las cadenas laterales de residuos de aminoácidos que definen la superficie de contacto en la posición X; (2) determinar la distancia entre el β -carbono de al menos un residuo que comprende la superficie de contacto y al menos una base del par de bases en la posición X; y (3) (a) para un residuo que está < 6 Å de la base, seleccionar un residuo del Grupo 1 y/o Grupo 2 que es un miembro de uno apropiado del Grupo G, Grupo C, Grupo T o Grupo A para promover el cambio deseado, y/o (b) para un residuo que está > 6 Å de la base, seleccionar un residuo del Grupo 2 y/o Grupo 3 que es un miembro de uno apropiado del Grupo G, Grupo C, Grupo T o Grupo A para promover el cambio deseado. Puede seleccionarse más de un residuo tal que comprenda la superficie de contacto para análisis y modificación y, en

algunas realizaciones, cada residuo tal se analiza y se modifican múltiples residuos. Similarmente, puede determinarse la distancia entre el β -carbono de un residuo incluido en la superficie de contacto y cada una de las dos bases del par de bases en la posición X y, si el residuo está dentro de 9 Å de ambas bases, entonces pueden hacerse diferentes sustituciones para afectar las dos bases del par (por ejemplo, un residuo del Grupo 1 para afectar una base proximal en una hebra, o un residuo del Grupo 3 para afectar una base distal en la otra hebra). Alternativamente, una combinación de sustituciones de residuos que pueden interaccionar con ambas bases en un par puede afectar la especificidad (por ejemplo, un residuo del Grupo T que se pone en contacto con la hebra codificante combinado con un residuo del Grupo A que se pone en contacto con la hebra no codificante para seleccionar T/A). Finalmente, pueden validarse múltiples modificaciones alternativas de los residuos tanto empíricamente (por ejemplo, produciendo la meganucleasa recombinante y probando su reconocimiento de secuencias) o computacionalmente (por ejemplo, por modelado informático del complejo de meganucleasa-ADN de la enzima modificada) para elegir entre alternativas.

Una vez se han seleccionado una o más modificaciones de aminoácidos deseadas de la meganucleasa natural o de referencia, la meganucleasa racionalmente diseñada puede producirse mediante procedimientos recombinantes y técnicas muy conocidas en la técnica. En algunas realizaciones se usan técnicas de mutagénesis no al azar o dirigida al sitio para crear modificaciones de secuencias específicas. Ejemplos no limitantes de técnicas de mutagénesis no al azar incluyen PCR de cebadores que se solapan (véase, por ejemplo, Wang y col. (2006), *Nucleic Acids Res.* 34(2): 517-527), mutagénesis dirigida al sitio (véase, por ejemplo, la patente de EE.UU. n° 7.041.814), mutagénesis de casete (véase, por ejemplo, la patente de EE.UU. n° 7.041.814), y el protocolo del fabricante para el kit Altered Sites® II Mutagenesis Systems comercialmente disponible de Promega Biosciences, Inc. (San Luis Obispo, CA).

El reconocimiento y escisión de una secuencia de ADN específica por una meganucleasa racionalmente diseñada puede ensayarse mediante cualquier procedimiento conocido por un experto en la materia (véase, por ejemplo, la publicación de patente de EE.UU. n° 2006/0078552). En ciertas realizaciones, la determinación de la escisión de meganucleasas se determina por ensayos de escisión *in vitro*. Tales ensayos usan escisión *in vitro* de un sustrato de polinucleótido que comprende la secuencia de reconocimiento prevista de la meganucleasa ensayada y, en ciertas realizaciones, variaciones de la secuencia de reconocimiento prevista en la que una o más bases en uno o ambos medios sitios se han cambiado a una base diferente. Normalmente, el sustrato de polinucleótido es una molécula de ADN bicatenario que comprende un sitio diana que se ha sintetizado y clonado en un vector. El sustrato de polinucleótido puede ser lineal o circular, y normalmente comprende solo una secuencia de reconocimiento. La meganucleasa se incuba con el sustrato de polinucleótido bajo condiciones apropiadas, y los polinucleótidos resultantes se analizan mediante procedimientos conocidos para identificar productos de escisión (por ejemplo, electroforesis o cromatografía). Si hay una única secuencia de reconocimiento en un sustrato de ADN bicatenario lineal, la actividad de meganucleasa se detecta por la aparición de dos bandas (productos) y la desaparición de la banda de sustrato de longitud completa inicial. En una realización, la actividad de meganucleasa puede ensayarse como se describe en, por ejemplo, Wang y col. (1997), *Nucleic Acid Res.*, 25: 3767-3776.

En otras realizaciones, el patrón de escisión de la meganucleasa se determina usando ensayos de escisión *in vivo* (véase, por ejemplo, la publicación de patente de EE.UU. n° 2006/0078552). En realizaciones particulares, la prueba *in vivo* es una prueba de recombinación por hibridación monocatenaria (SSA). Este tipo de prueba es conocida para aquellos expertos en la materia (Rudin y col. (1989), *Genetics* 122: 519-534; Fishman-Lobell y col. (1992), *Science* 258: 480-4).

Como será evidente para un experto en la materia, pueden hacerse sustituciones, inserciones o deleciones de aminoácidos adicionales a dominios de las enzimas meganucleasa distintos de aquellos que participan en el reconocimiento y unión de ADN sin pérdida de actividad completa. Las sustituciones pueden ser sustituciones conservativas de residuos de aminoácidos similares en posiciones estructuralmente o funcionalmente limitadas, o pueden ser sustituciones no conservativas en las posiciones que son menos estructuralmente o funcionalmente limitadas. Tales sustituciones, inserciones y deleciones pueden identificarse por un experto habitual en la materia por experimentación rutinaria sin excesivo esfuerzo. Así, en algunas realizaciones, las meganucleasas recombinantes de la invención incluyen proteínas que tienen en cualquier parte del 85% al 99% de similitud de secuencias (por ejemplo, 85%, 87,5%, 90%, 92,5%, 95%, 97,5%, 99%) con una secuencia de meganucleasa de referencia. Con respecto a cada una de las proteínas I-CreI, I-MsoI, I-SceI y I-CeuI naturales, las secuencias de casi el extremo N y el extremo C no son claramente visibles en estudios de cristalografía de rayos X, sugiriendo que estas posiciones no están estructuralmente o funcionalmente limitadas. Por tanto, estos residuos pueden excluirse del cálculo de similitud de secuencias, y pueden usarse las siguientes secuencias de meganucleasa de referencia: residuos 2-153 de SEC ID N°: 1 para I-CreI, residuos 6-160 de SEC ID N°: 6 para I-MsoI, residuos 3-186 de SEC ID N°: 9 para I-SceI, y residuos 5-211 de SEC ID N°: 12 para I-CeuI.

2.2 Meganucleasas de la familia LAGLIDADG

La familia de meganucleasas LAGLIDADG está compuesta por más de 200 miembros de un grupo filogenético diverso de organismos huésped. Todos los miembros de esta familia tienen una o dos copias de un motivo LAGLIDADG altamente conservado junto con otros motivos estructurales que participan en la escisión de secuencias

de ADN específicas. Las enzimas que tienen una única copia del motivo LAGLIDADG (es decir, meganucleasas mono-LAGLIDADG) sirven de dímeros, mientras que las enzimas que tienen dos copias de este motivo (es decir, meganucleasas di-LAGLIDADG) sirven de monómeros.

5 Todos los miembros de la familia LAGLIDADG reconocen y escinden secuencias relativamente largas (> 12 pb), dejando cuatro nucleótidos protuberantes en 3'. Estas enzimas también comparten varios motivos estructurales, además del motivo LAGLIDADG, que incluye una disposición similar de hebras β antiparalelas en la superficie de separación de proteína-ADN. Los aminoácidos dentro de estos motivos estructurales conservados son responsables de la interacción con las bases de ADN para conferir especificidad por secuencias de reconocimiento. La similitud estructural global entre algunos miembros de la familia (por ejemplo, I-Crel, I-Msol, I-Scel y I-CeuI) se ha esclarecido por cristalografía de rayos X. Por consiguiente, los miembros de esta familia pueden modificarse en aminoácidos particulares dentro de tales motivos estructurales para cambiar la actividad global o especificidad por secuencia de las enzimas, y puede esperarse razonablemente que modificaciones correspondientes tengan resultados similares en otros miembros de la familia. Véase, generalmente, Chevalier y col. (2001), Nucleic Acid Res. 29(18): 3757-3774).

2.2.1 Meganucleasas derivadas de I-Crel

En un aspecto, la presente invención se refiere a meganucleasas racionalmente diseñadas que se basan en o se derivan de la meganucleasa I-Crel de *Chlamydomonas reinhardtii*. La secuencia de aminoácidos natural de la meganucleasa I-Crel se muestra en SEC ID N°: 1, que se corresponde con el n° de acceso de GenBank PO5725. Dos medios sitios de la secuencia de reconocimiento de la meganucleasa I-Crel natural de PDB de estructura cristalina n° 1BP7 se muestran a continuación:

25 Posición -9-8-7-6-5-4-3-2-1
5'-**G A A A C T G T C T C A C G A C G T T T T G**-3' SEC ID N°: 2
3'-**C T T T G A C A G A G T G C T G C A A A C**-5' SEC ID N°: 3
Posición -1-2-3-4-5-6-7-8-9

30 Obsérvese que esta secuencia de reconocimiento natural no es perfectamente palindrómica, incluso fuera de los cuatro pares de bases centrales. Los dos medios sitios de la secuencia de reconocimiento se muestran en negrita en sus hebras codificantes respectivas.

35 I-Crel natural también reconoce y corta perfectamente la siguiente secuencia palindrómica (excepto las bases N₁-N₄ centrales):

40 Posición -9-8-7-6-5-4-3-2-1
5'-**C A A A C T G T C G T G A G A C A G T T T G**-3' SEC ID N°: 4
3'-**G T T T G A C A G C A C T C T G T C A A A C**-5' SEC ID N°: 5
Posición -1-2-3-4-5-6-7-8-9

45 La secuencia palindrómica de SEC ID N°: 4 y SEC ID N°: 5 se considera que es un mejor sustrato para I-Crel natural debido a que la enzima se une a este sitio con mayor afinidad y se escinde más eficientemente que la secuencia de ADN natural. Para los fines de la siguiente divulgación, y con particular respecto a los resultados experimentales presentados en el presente documento, esta secuencia palindrómica escindida por I-Crel natural se denomina "WT" (véase, por ejemplo, la Figura 2(A)). Los dos medios sitios de la secuencia de reconocimiento se muestran en negrita sobre sus hebras codificantes respectivas.

50 La Figura 1(A) representa las interacciones de un homodímero de meganucleasa I-Crel natural con una secuencia de reconocimiento de ADN bicatenario, las Figuras 1(B) muestran las interacciones específicas entre residuos de aminoácidos de la enzima y bases en la posición -4 de un medio sitio para una enzima natural y una secuencia de reconocimiento natural, y las Figuras 1(C)-(E) muestran las interacciones específicas entre residuos de aminoácidos de la enzima y bases en la posición -4 de un medio sitio para tres meganucleasas racionalmente diseñadas de la invención con especificidad alterada en la posición -4 del medio sitio.

60 Así, la preferencia de bases en cualquier posición de base especificada del medio sitio puede alterarse racionalmente para cada uno de los otros tres pares de bases usando los procedimientos desvelados en el presente documento. Primero, la superficie de reconocimiento natural en la posición de base especificada se determina (por ejemplo, analizando estructuras de co-cristales de complejos de meganucleasa-ADN; o por modelado informático de los complejos de meganucleasa-ADN). Segundo, se determinan residuos de contacto existentes y posibles basados en las distancias entre los β -carbonos de las posiciones de aminoácidos de alrededor y las bases de ácidos nucleicos en cada hebra de ADN en la posición de base especificada. Por ejemplo, y sin limitación, como se muestra en la Figura 1(A), los residuos de contacto de la meganucleasa natural I-Crel-ADN en la posición -4 implican una

glutamina en la posición 26 que se une con hidrógeno a una base A sobre una hebra de ADN antisentido. El residuo 77 también se identificó como que posiblemente podía poner en contacto la base de -4 en la hebra codificante de ADN. El β -carbono del residuo 26 está 5,9 Å alejado de N7 de la base de A en la hebra de ADN no codificante, y el β -carbono del residuo 77 está separado 7,15 Å del C5-metilo de T en la hebra codificante. Según la distancia y reglas químicas de bases descritas en el presente documento, una C en la hebra codificante podría enlazarse con hidrógeno con un ácido glutámico en la posición 77 y una G en la hebra no codificante podría enlazarse con glutamina en la posición 26 (mediada por una molécula de agua, como se observa en la estructura cristalina de I-Crel natural) (véase la Fig. 1(C)); una G en la hebra codificante podría enlazarse con hidrógeno con una arginina en la posición 77 y una C en la hebra no codificante podría enlazarse con hidrógeno con un ácido glutámico en la posición 26 (véase la Fig. 1(D)); una A en la hebra codificante podría enlazarse con hidrógeno con una glutamina en la posición 77 y una T en la hebra no codificante podría formar contactos hidrófobos con una alanina en la posición 26 (véase la Fig. 1(E)). Si se proporciona contacto específico de base por la posición 77, entonces el contacto natural, Q26, puede estar sustituido (por ejemplo, con un residuo de serina) para reducir o eliminar su influencia sobre la especificidad. Alternativamente, mutaciones complementarias en las posiciones 26 y 77 pueden combinarse para especificar un par de bases particular (por ejemplo, A26 especifica una T en la hebra no codificante y Q77 especifica una A en la hebra codificante (Fig. 1(E)). Estas sustituciones de residuos predichas se han validado todas experimentalmente.

Así, se ha identificado un número sustancial de modificaciones de aminoácidos al dominio de reconocimiento de ADN de la meganucleasa I-Crel que, individualmente o en combinación, producen meganucleasas recombinantes con especificidades alteradas en bases individuales dentro del medio sitio de la secuencia de reconocimiento de ADN, de forma que estas meganucleasas racionalmente diseñadas tienen medios sitios diferentes de la enzima natural. Las modificaciones de aminoácidos de I-Crel y el cambio resultante en la especificidad del medio sitio de la secuencia de reconocimiento se muestran en la Tabla 1:

TABLA 1

Base de la hebra codificante favorecida											
Posn.	A	C	G	T	A/T	A/C	A/G	C/T	G/T	A/G/T	A/C/G/T
-1	Y75 L75* C75* Y139* C46* A46*	R70* H75* R75* H46* K46* R46*	K70 E70* E75* E46* D46*	Q70* C70 L70 Y75* Q75* H75* H139 Q46* H46*				T46*			G70 A70 S70 G46*
-2	Q70 T44* A44* V44* I44* L44* N44*	E70 D70 K44* R44*	H70 D44* E44*	Q44*	C44*						
-3	Q68 C24* I24*	E68 F68 K24* R24*	R68	M68 C68 L68 F68		H68		Y68	K68		
-4	A26* Q77	E77 K26*	R77 E26*					S77 Q26*			S26*
-5		E42	R42			K28*	C28* Q42				M66 K66
-6	Q40 C28*	E40 R28*	R40	C40 140 V40 C79 179 V79 Q28*	A40 A79 A28* H28*						S40 S28*
-7	N30* Q38	E38 K30* R30*	K38 R38 E30*	138 L38			C38				H38 N38 Q30*

Base de la hebra codificante favorecida											
Posn.	A	C	G	T	A/T	A/C	A/G	C/T	G/T	A/G/T	A/C/G/T
-8	F33 Y33	E33 D33	F33 H33	L33 V33 133 F33 C33		R32*	R33				
-9		E32	R32 K32	L32 V32 A32 C32				D32 I32			S32 N32 H32 Q32 T32
Las entradas en negrita son residuos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento. Un asterisco indica que el residuo contacta la base en la hebra no codificante.											

2.2.2 Meganucleasas derivadas de I-Msol

En otro aspecto, la presente invención se refiere a meganucleasas racionalmente diseñadas que se basan en o se derivan de la meganucleasa de I-Msol de *Monomastix sp.* La secuencia de aminoácidos natural de la meganucleasa I-Msol se muestra en SEC ID N°: 6, que se corresponde con el n° de acceso de GenBank AAL34387. Dos medios sitios de la secuencia de reconocimiento de la meganucleasa I-Msol natural de PDB de estructura cristalina n° 1M5X se muestran a continuación:

Posición -9-8-7-6-5-4-3-2-1
 5'-**C A G A A C G T C G T G A G A C A G T T C C**-3' SEC ID N°: 7
 3'-**G T C T T G C A G C A C T C T G T C A A G G**-5' SEC ID N°: 8
 Posición -1-2-3-4-5-6-7-8-9

Obsérvese que la secuencia de reconocimiento no es perfectamente palindrómica, incluso fuera de los cuatro pares de bases centrales. Los dos medios sitios de la secuencia de reconocimiento se muestran en negrita en sus hebras codificantes respectivas.

Según la invención, se ha identificado un número sustancial de modificaciones de aminoácidos al dominio de reconocimiento de ADN de la meganucleasa I-Msol que, individualmente o en combinación, pueden producir meganucleasas recombinantes con especificidades alteradas en bases individuales dentro de los medios sitios de la secuencia de reconocimiento de ADN, de forma que estas meganucleasas racionalmente diseñadas tienen secuencias de reconocimiento diferentes de la enzima natural. Las modificaciones de aminoácidos de I-Msol y el cambio predicho en la especificidad del medio sitio de la secuencia de reconocimiento se muestran en la Tabla 2:

TABLA 2

Base de la hebra codificante favorecida				
Posición	A	C	G	T
-1	K75* Q77 A49* C49* K79*	D77 E77 K49* R75* K75* R79* K79*	K77 R77 E49* E79*	C77 L77 Q79*
-2	Q75 K81 C47* I47* L47*	E75 D75 R47* K47* K81* R81*	K75 E47* E81*	A75 C75 V75 I75 T75 Q47* Q81*
-3	Q72 C26* L26* V26* A26* I26*	E72 Y72 H26* K26* R26*	R72 K72 Y26* F26*	K72 Y72 H26*

Base de la hebra codificante favorecida				
Posición	A	C	G	T
-4	K28 Q83	K28* R28* E83	R83 K83	K28 K83 Q28*
-5	K28 C28* L28* I28*	K28* R28*	R45 E28*	Q28*
-6	I30* V30* S30* L30* Q43	E43 E85 K30* R30*	R43 K43 K85 R85 E30* D30*	K43 185 V85 L85 Q30*
-7	Q41	E32 E41	R32 R41 K41	K32 M41 L41 I41
-8	Y35 K35	E32	R32 K32 K35 R35	K32 K35
-9	N34 H34	D34 E34 S34	K34 R34 H34	S34 C34 V34 T34 A34
Las entradas en negrita representan residuos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento. Un asterisco indica que el residuo contacta la base en la hebra no codificante.				

En otro aspecto, la presente invención se refiere a meganucleasas racionalmente diseñadas que se basan en o se derivan de la meganucleasa I-SceI de *Saccharomyces cerevisiae*. La secuencia de aminoácidos natural de la meganucleasa I-SceI se muestra en SEC ID N°: 9, que se corresponde con el n° de acceso de GenBank CAA09843. La secuencia de reconocimiento de la meganucleasa natural I-SceI de PDB de estructura cristalina n° 1R7M se muestra a continuación:

Codificante	5'-T T A C C C T G T T A T C C C T A G-3'	SEC ID N°:
10		
No codificante	3'-A A A T G G G A C A A T A G G G A T C-5'	SEC ID N°:
11		
Posición	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	

Obsérvese que la secuencia de reconocimiento es no palindrómica y no hay cuatro pares de bases separando los medios sitios.

Según la invención, se ha identificado un número sustancial de modificaciones de aminoácidos al dominio de reconocimiento de ADN de la meganucleasa I-SceI que, individualmente o en combinación, pueden producir meganucleasas recombinantes con especificidades alteradas en bases individuales dentro de la secuencia de reconocimiento de ADN, de forma que estas meganucleasas racionalmente diseñadas tienen secuencias de reconocimiento diferentes de la enzima natural. La modificaciones de aminoácidos de I-SceI y el cambio predicho en la especificidad de la secuencia de reconocimiento se muestran en la Tabla 3:

TABLA 3

Base de la hebra codificante favorecida				
Posición	A	C	G	T
4	K50	R50* K50* E57	E50* R57 K57	K57 M57 Q50*
5	K48 Q102	R48* K48* E102 E59	E48* K102 R102	Q48* C102 L102 V102
6	K59	R59* K59*	K84 E59*	Q59* Y46
7	C46* L46* V46*	R46* K46* E86	K86 R86 E46*	K68 C86 L86 Q46*
8	K61* S61* V61* A61* L61*	E88 R61* H61*	E61* R88 K88	K88 Q61* H61*
9	T98* C98* V98* L98*	R98* K98*	E98* D98*	Q98*
10	V96* C96* A96*	K96* R96*	D96* E96*	Q96*
11	C90* L90*	K90* R90*	E90*	Q90*
12	Q193	E165 E193 D193	K165 R165	C165 L165 C193 V193 A193 T193 S193
13	C193* L193*	K193* R193* D192	E193* D193* K163 R192	Q193* C163 L163
14	L192* C192*	E161 R192* K192*	K147 K161 R161 R197 D192* E192*	K161 Q192*
15		E151	K151	C151 L151 K151
17	N152* S152* C150* L150* V150* T150*	K152* K150*	N152* S152* D152* D150* E150*	Q152* Q150*
18	K155* C155*	R155* K155*	E155*	H155* Y155*
Las entradas en negrita son residuos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento. Un asterisco indica que el residuo contacta la base en la hebra no codificante.				

2.2.4 Meganucleasas derivadas de I-CeuI

En otro aspecto, la presente invención se refiere a meganucleasas racionalmente diseñadas que se basan en o se

derivan de la meganucleasa I-Ceul de *Chlamydomonas eugametos*. La secuencia de aminoácidos natural de la meganucleasa I-Ceul se muestra en SEC ID N°: 12, que se corresponde con el n° de acceso de GenBank P32761. Dos medios sitios de la secuencia de reconocimiento de la meganucleasa I-Ceul natural de PDB de estructura cristalina n° 2EX5 se muestran a continuación:

Posición -9-8-7-6-5-4-3-2-1
5'-**A T A A C G G T C C T A A G G T A G C G A A**-3' SEC ID N°:
13
3'-**T A T T G C C A G G A T T C C A T C G C T T**-5' SEC ID N°:
14
Posición -1-2-3-4-5-6-7-8-9

Obsérvese que la secuencia de reconocimiento es no palindrómica, incluso fuera de los cuatro pares de bases centrales, a pesar del hecho de que I-Ceul es un homodímero, debido a la degeneración natural en la superficie de separación de reconocimiento de I-Ceul (Spiegel y col. (2006), Structure 14:869-80). Los dos medios sitios de la secuencia de reconocimiento se muestran en negrita en sus hebras codificantes respectivas.

Según la invención, se ha identificado un número sustancial de modificaciones de aminoácidos al dominio de reconocimiento de ADN de la meganucleasa I-Ceul que, individualmente o en combinación, producen meganucleasas recombinantes con especificidades alteradas en bases individuales dentro del medio sitio de la secuencia de reconocimiento de ADN, de forma que estas meganucleasas racionalmente diseñadas pueden tener secuencias de reconocimiento diferentes de la enzima natural. Las modificaciones de aminoácidos de I-Ceul y el cambio predicho en la especificidad de la secuencia de reconocimiento se muestran en la Tabla 4:

TABLA 4

Base de la hebra codificante favorecida				
Posición	A	C	G	T
-1	C92* A92* V92*	K116* R116* D116* K92*	E116* E92*	Q116* Q92*
-2	Q117 C90* L90* V90*	E117 D117 R174* K124* K90* R90* K68*	K117 R124 K124 E124* E90* D90*	C117 V117 T117 Q90*
-3	C70* V70* T70* L70* K70*	K70*	E70* E88*	Q70*
-4	Q126 N126 K88* L88* C88* C72* L72* V72*	E126 D126 R88* K88* K72*	R126 K126 E88* D88*	K126 L126 Q88*
-5	C74* L74* V74* T74*	K74*	E74* K128 R128 E128	C128 L128 V128 T128
-6	Q86	D86 E86 R84* K84*	K128 R128 R86 K86 E84*	K86 C86 L86
-7	L76* C76* K76*	R76* K76* H76*	E76* R84	H76* Q76*
-8	Y79 R79 Q76	D79 E79 D76	R79 K79 K76	C79 L79 V79

2.2.5 Meganucleasas recombinantes específicamente excluidas

La presente invención no pretende englobar ciertas meganucleasas recombinantes que se han descrito en la técnica anterior, y que se han desarrollado mediante procedimientos alternativos. Estas meganucleasas excluidas incluyen aquellas descritas por Arnould y col. (2006), J. Mol. Biol. 355: 443-58; Sussman y col. (2004), J. Mol. Biol. 342: 31-41; Chames y col. (2005), Nucleic Acids Res. 33: e178; Seligman y col. (2002), Nucleic Acids Res. 30: 3870-9; y Ashworth y col. (2006), Nature 441 (7093):656-659; que incluyen meganucleasas recombinantes basadas en I-Crel con sustituciones individuales seleccionadas de C33, R33, A44, H33, K32, F33, R32, A28, A70, E33, V33, A26 y R66. También se excluyen meganucleasas recombinantes basadas en I-Crel con tres sustituciones seleccionadas de A68/N70/N75 y D44/D70/N75, o con cuatro sustituciones seleccionadas de K44/T68/G60/N75 y R44/A68/T70/N75. Estas sustituciones o combinaciones de sustituciones se denominan en el presente documento las "modificaciones excluidas".

2.2.6 Meganucleasas con múltiples cambios en el medio sitio de la secuencia de reconocimiento

En otro aspecto, la presente invención se refiere a meganucleasas racionalmente diseñadas que se producen combinando dos o más modificaciones de aminoácidos como se describe en la Sección 2.2.1 anterior, con el fin de alterar la preferencia de medios sitios en dos o más posiciones en un medio sitio de la secuencia de reconocimiento de ADN. Por ejemplo, sin limitación, y como se ha descrito más completamente más adelante, la enzima DJ1 se derivó de I-Crel incorporando las modificaciones R30/E38 (que favorecen C en la posición -7), R40 (que favorece G en la posición -6), R42 (que favorece G en la posición -5) y N32 (que favorece la degeneración completa en la posición -9). La meganucleasa DJ1 racionalmente diseñada reconoce invariamente C₋₇ G₋₆ G₋₅ en comparación con la preferencia natural para A₋₇ A₋₆ C₋₅, y tiene elevada tolerancia por A en la posición -9.

La capacidad para combinar sustituciones de residuos que afectan diferentes posiciones de base es debida en parte a la naturaleza modular de las meganucleasas LAGLIDADG. La mayoría de los contactos de bases en las superficies de separación de reconocimiento de LAGLIDADG están hechos por cadenas laterales de aminoácidos individuales, y la superficie de separación está relativamente libre de interconectividad o redes de enlace de hidrógeno entre cadenas laterales que interaccionan con bases adyacentes. Esto permite generalmente la manipulación de residuos que interaccionan con una posición de base sin afectar las interacciones de cadenas laterales en bases adyacentes. La naturaleza aditiva de las mutaciones enumeradas en la Sección 2.2.1 anterior también es un resultado directo del procedimiento usado para identificar estas mutaciones. El procedimiento predice sustituciones de cadenas laterales que interaccionan directamente con una única base. La interconectividad o redes de enlace de hidrógeno entre cadenas laterales se evitan generalmente para mantener la independencia de las sustituciones dentro de la superficie de separación de reconocimiento.

Ciertas combinaciones de sustituciones de la cadena lateral son completamente o parcialmente incompatibles entre sí. Cuando un par o conjunto incompatible de aminoácidos se incorpora en una meganucleasa racionalmente diseñada, la enzima resultante tendrá actividad catalítica reducida o eliminada. Normalmente, estas incompatibilidades son debidas a interferencia estérica entre las cadenas laterales de los aminoácidos introducidos y la actividad puede restaurarse identificando y eliminando esta interferencia. Específicamente, cuando dos aminoácidos con cadenas laterales grandes (por ejemplo, aminoácidos del Grupo 2 ó 3) se incorporan en posiciones de aminoácidos que son adyacentes entre sí en la estructura de meganucleasa (por ejemplo, posiciones 32 y 33, 28 y 40, 28 y 42, 42 y 77, o 68 y 77 en el caso de meganucleasas derivadas de I-Crel), es probable que estos dos aminoácidos interfieran entre sí y reduzcan la actividad enzimática. Esta interferencia se elimina sustituyendo uno o ambos aminoácidos incompatibles con un aminoácido con una cadena lateral más pequeña (por ejemplo, Grupo 1 o Grupo 2). Por ejemplo, en meganucleasas racionalmente diseñadas derivadas de I-Crel, K28 interfiere con tanto R40 como R42. Para maximizar la actividad enzimática, R40 y R42 pueden combinarse con una serina o ácido aspártico en la posición 28.

Las combinaciones de sustituciones de amino, identificadas como se ha descrito en el presente documento, pueden usarse para alterar racionalmente la especificidad de una meganucleasa natural (o una meganucleasa previamente modificada) de una secuencia de reconocimiento original a una secuencia de reconocimiento deseada que puede estar presente en un ácido nucleico de interés (por ejemplo, un genoma). La Figura 2A, por ejemplo, muestra la hebra "codificante" de la secuencia de reconocimiento de la meganucleasa I-Crel WT (SEC ID N°: 4), además de varias otras secuencias para las que sería útil una meganucleasa racionalmente diseñada. Las bases conservadas entre la secuencia de reconocimiento WT y la secuencia de reconocimiento deseada están sombreadas. Según la invención, las meganucleasas recombinantes basadas en la meganucleasa I-Crel pueden diseñarse racionalmente para cada una de estas secuencias de reconocimiento deseadas, además de cualquier otra, por sustituciones de aminoácidos adecuadas como se describe en el presente documento.

3. Meganucleasas racionalmente diseñadas con afinidad de unión a ADN alterada

Como se ha descrito anteriormente, la afinidad de unión a ADN de las meganucleasas recombinantes de la invención puede modularse alterando ciertos aminoácidos que forman la superficie de contacto con el esqueleto de fosfodiéster de ADN. La superficie de contacto comprende aquellos aminoácidos en la enzima con β -carbonos

inferiores a 9 Å del esqueleto de ADN, y con cadenas laterales orientadas hacia el ADN, independientemente de si los residuos hacen contactos o no con el esqueleto de ADN en el complejo de meganucleasa natural-ADN. Debido a que la unión al ADN es un precursor necesario para la actividad enzimática, se ha mostrado que aumentos/disminuciones en la afinidad de unión a ADN producen aumentos/disminuciones, respectivamente, en la actividad enzimática. Sin embargo, también se ha mostrado que los aumentos/disminuciones en la afinidad de unión a ADN producen disminuciones/aumentos en la especificidad de la secuencia de meganucleasa. Por tanto, tanto la actividad como la especificidad pueden modularse modificando los contactos del esqueleto de fosfodiéster.

Específicamente, para aumentar la actividad enzimática/disminuir la especificidad enzimática:

(i) Eliminar repulsión electrostática entre la enzima y el esqueleto de ADN. Si un aminoácido identificado tiene una cadena lateral negativamente cargada (por ejemplo, ácido aspártico, ácido glutámico) que se esperaría que repeliera el esqueleto de ADN negativamente cargado, la repulsión puede eliminarse sustituyendo un aminoácido con una cadena lateral sin cargar o positivamente cargada, sujeto a efectos de interferencia estérica. Un ejemplo experimentalmente verificado es la mutación de ácido glutámico 80 en I-Crel a glutamina.

(ii) Introducir interacción por atracción electrostática entre la enzima y el esqueleto de ADN. En cualquiera de las posiciones de la superficie de contacto, se espera que la introducción de un aminoácido con una cadena lateral positivamente cargada (por ejemplo, lisina o arginina) aumente la afinidad de unión, sujeta a efectos de interferencia estérica.

(iii) Introducir un enlace de hidrógeno entre la enzima y el esqueleto de ADN. Si un aminoácido de la superficie de contacto no hace un enlace de hidrógeno con el esqueleto de ADN debido a que carece de una funcionalidad para enlace de hidrógeno apropiada o tiene una cadena lateral que es demasiado corta, demasiado larga y/o demasiado inflexible para interactuar con el esqueleto de ADN, puede introducirse un aminoácido polar que puede donar un enlace de hidrógeno (por ejemplo, serina, treonina, tirosina, histidina, glutamina, asparagina, lisina, cisteína o arginina) con la longitud y flexibilidad apropiada, sujeto a efectos de interferencia estérica.

Específicamente, para reducir la actividad enzimática/aumentar la especificidad enzimática:

(i) Introducir repulsión electrostática entre la enzima y el esqueleto de ADN. En cualquiera de las posiciones de la superficie de contacto, se espera que la introducción de un aminoácido con una cadena lateral negativamente cargada (por ejemplo, ácido glutámico, ácido aspártico) reduzca la afinidad de unión, sujeto a efectos de interferencia estérica.

(ii) Eliminar atracción electrostática entre la enzima y el ADN. Si cualquier aminoácido de la superficie de contacto tiene una cadena lateral positivamente cargada (por ejemplo, lisina o arginina) que interacciona con el esqueleto de ADN negativamente cargado, esta interacción favorable puede eliminarse sustituyendo un aminoácido con una cadena lateral sin cargar o negativamente cargada, sujeto a efectos de interferencia estérica. Un ejemplo experimentalmente verificado es la mutación de lisina 116 en I-Crel a ácido aspártico.

(iii) Eliminar un enlace de hidrógeno entre la enzima y el esqueleto de ADN. Si cualquier aminoácido de la superficie de contacto hace un enlace de hidrógeno con el esqueleto de ADN, puede estar sustituido con un aminoácido que no se esperaría que hiciera un enlace de hidrógeno similar debido a que su cadena lateral no está apropiadamente funcionalizada o carece de las características de longitud/flexibilidad necesarias.

Por ejemplo, en algunas meganucleasas recombinantes basadas en I-Crel, el ácido glutámico en la posición 80 en la meganucleasa I-Crel está alterado a tanto una lisina como una glutamina para aumentar la actividad. En otra realización, la tirosina en la posición 66 de I-Crel se cambia a arginina o lisina, que aumenta la actividad de la meganucleasa. En otra realización más, la actividad enzimática disminuye cambiando la lisina en la posición 34 de I-Crel con ácido aspártico, cambiando la tirosina en la posición 66 con ácido aspártico, y/o cambiando la lisina en la posición 116 con ácido aspártico.

Las actividades de las meganucleasas recombinantes pueden modularse de forma que la enzima recombinante tenga en cualquier sitio de no actividad a actividad muy alta con respecto a una secuencia de reconocimiento particular. Por ejemplo, la meganucleasa recombinante DJ1 cuando lleva mutación de ácido glutámico en la posición 26 pierde actividad completamente. Sin embargo, la combinación de la sustitución de ácido glutámico en la posición 26 y una sustitución de glutamina en la posición 80 crea una meganucleasa recombinante con alta especificidad y actividad hacia una guanina en -4 dentro del medio sitio de la secuencia de reconocimiento (véase la Figura 1(D)).

Se desvela que, aminoácidos en diversas posiciones en proximidad al esqueleto de fosfodiéster de ADN, pueden cambiarse para afectar simultáneamente tanto la actividad como la especificidad de meganucleasa. Este "ajuste" de la especificidad y actividad enzimática se lleva a cabo aumentando o disminuyendo el número de contactos hechos por aminoácidos con el esqueleto de fosfodiéster. Una variedad de contactos con el esqueleto de fosfodiéster puede facilitarse por cadenas laterales de aminoácidos. En algunas realizaciones, enlaces iónicos, puentes de sales, enlaces de hidrógeno e impedimento estérico afectan la asociación de cadenas laterales de aminoácidos con el

esqueleto de fosfodiéster. Por ejemplo, para la meganucleasa I-Crel, la alteración de la lisina en la posición 116 a un ácido aspártico elimina un puente de sal entre pares de bases de ácidos nucleicos en las posiciones -8 y -9, reduciendo la tasa de escisión enzimática, pero aumentando la especificidad.

- 5 Los residuos que forman la superficie de contacto del esqueleto de cada una de la meganucleasa I-Crel natural (SEC ID N°: 1), (SEC ID N°: 12) se identifican en la Tabla 5 a continuación:

TABLA 5

	I-Crel
10	P29, K34, T46, K48, R51, V64, Y66, E80, I81, K82, L112, K116, D137, K139, T140, T143

Para aumentar la afinidad de una enzima y así hacerla más activa/menos específica:

- 15 (1) Seleccionar un aminoácido de la Tabla 5 para la enzima correspondiente que esté tanto negativamente cargado (D o E), hidrófobo (A, C, F, G, I, L, M, P, V, W, Y) como sin carga/polar (H, N, Q, S, T).
 (2) Si el aminoácido está negativamente cargado o es hidrófobo, mutarlo a sin carga/polar (menos efecto) o positivamente cargado (K o R, más efecto).
 (3) Si el aminoácido está sin carga/polar, mutarlo a positivamente cargado.

20 Para reducir la afinidad de una enzima y así hacerla menos activa/más específica:

- (1) Seleccionar un aminoácido de la Tabla 5 para la enzima correspondiente que esté tanto positivamente cargado (K o R), hidrófobo (A, C, F, G, I, L, M, P, V, W, Y) como sin carga/polar (H, N, Q, S, T).
 25 (2) Si el aminoácido está positivamente cargado, mutarlo a sin carga/polar (menos efecto) o negativamente cargado (más efecto).
 (3) Si el aminoácido es hidrófobo o sin carga/polar, mutarlo a negativamente cargado.

4. Meganucleasas heterodiméricas

30 En otro aspecto se desvelan meganucleasas que son heterodímeros formados por la asociación de dos monómeros, uno de los cuales puede ser natural y uno o ambos de los cuales puede ser una forma que no se produce naturalmente o recombinante. Por ejemplo, la meganucleasa I-Crel natural es normalmente un homodímero compuesto de dos monómeros que se unen cada uno a medio sitio en la secuencia de reconocimiento pseudo-palindrómica. Una meganucleasa recombinante heterodimérica puede producirse combinando dos meganucleasas que reconocen diferentes medio sitios, por ejemplo, co-expresando las dos meganucleasas en una célula o mezclando dos meganucleasas en disolución. La formación de heterodímeros puede favorecerse con respecto a la formación de homodímeros alterando aminoácidos en cada uno de los dos monómeros que afectan su asociación en dímeros. En realizaciones particulares, ciertos aminoácidos en la superficie de separación de los dos monómeros se alteran de aminoácidos negativamente cargados (D o E) a aminoácidos positivamente cargados (K o R) en un primer monómero y de aminoácidos positivamente cargados a aminoácidos negativamente cargados en un segundo monómero (Tabla 6). Por ejemplo, en el caso de meganucleasas derivadas de I-Crel, las lisinas en las posiciones 7 y 57 se mutan a ácidos glutámicos en el primer monómero y los ácidos glutámicos en las posiciones 8 y 61 se mutan a lisinas en el segundo monómero. El resultado de este procedimiento es un par de monómeros en los que el primer monómero tiene un exceso de residuos positivamente cargados en la superficie de separación del dímero y el segundo monómero tiene un exceso de residuos negativamente cargados en la superficie de separación del dímero. El primer y segundo monómero se asociarán, por tanto, preferencialmente con respecto a sus pares de monómeros idénticos debido a las interacciones electrostáticas entre los aminoácidos alterados en la superficie de separación.

TABLA 6

	I-Crel: Sustituciones del primer monómero	I-Crel: Sustituciones del segundo monómero
	K7 a E7 o D7	E8 a K8 o R8
	K57 a E57 o D57	E61 a K61 o R61
55	K96 a E96 o D96	

Alternativamente, o además, ciertos aminoácidos en la superficie de separación de los dos monómeros pueden alterarse para impedir estéricamente la formación de homodímeros. Específicamente, los aminoácidos en la superficie de separación del dímero de un monómero están sustituidos con residuos más grandes o más voluminosos que evitarán estéricamente el homodímero. Los aminoácidos en la superficie de separación del dímero del segundo monómero opcionalmente pueden estar sustituidos con residuos más pequeños para compensar los residuos más voluminosos en el primer monómero y eliminar cualquier discrepancia en el heterodímero, o pueden estar sin modificar.

- 65 Provisionalmente o además, un puente iónico o enlace de hidrógeno puede enterrarse en el núcleo hidrófobo de una superficie de separación heterodimérica. Específicamente, un residuo hidrófobo en un monómero en el núcleo de la

superficie de separación puede estar sustituido con un residuo positivamente cargado. Además, un residuo hidrófobo en el segundo monómero, que interacciona en el homodímero natural con el residuo hidrófobo sustituido en el primer monómero, puede estar sustituido con un residuo negativamente cargado. Así, los dos residuos sustituidos pueden formar un puente iónico o enlace de hidrógeno. Al mismo tiempo, la repulsión electrostática de una carga sin satisfacer enterrada en una superficie de separación hidrófoba debe desfavorecer la formación de homodímeros.

Finalmente, como se observa anteriormente, cada monómero del heterodímero puede tener diferentes aminoácidos sustituidos en la región de reconocimiento de ADN de forma que cada uno tenga un medio sitio de ADN diferente y la secuencia de reconocimiento de ADN dimérica combinada no sea palindrómica.

5. Procedimientos de producción de células recombinantes y organismos

Los aspectos de la presente invención proporcionan adicionalmente procedimientos para producir células y organismos recombinantes, transgénicos o genéticamente modificados de otro modo usando meganucleasas racionalmente diseñadas. Así, en ciertas realizaciones, las meganucleasas recombinantes se desarrollan para producir específicamente una rotura bicatenaria en un único sitio o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para permitir la(s) precisa(s) inserción (inserciones) de una secuencia de interés por recombinación homóloga. En otras realizaciones se desarrollan meganucleasas recombinantes para producir específicamente una rotura bicatenaria en un único sitio o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para tanto (a) permitir inserción (inserciones) rara(s) de una secuencia de interés por unión de extremos no homólogos como (b) permitir la alteración de la secuencia diana por unión de extremos no homólogos. Como se usa en el presente documento con respecto a recombinación homóloga o unión de extremos no homólogos de secuencias de interés, el término "inserción" significa la ligación de una secuencia de interés en un cromosoma de forma que la secuencia de interés se integre en el cromosoma. En el caso de recombinación homóloga, una secuencia insertada puede sustituir una secuencia endógena, de forma que el ADN original esté sustituido por ADN exógeno de igual longitud, pero con una secuencia de nucleótidos alterada. Alternativamente, una secuencia insertada puede incluir más o menos bases que la secuencia que sustituye.

Por tanto, según este aspecto de la invención, los organismos recombinantes incluyen, pero no se limitan a, especies de plantas monocotiledóneas tales como arroz, trigo, maíz y centeno, y especies dicotiledóneas tales como legumbres (por ejemplo, alubias, soja, lentejas, cacahuets, guisantes), alfalfa, trébol, tabaco y especies de *Arabidopsis*. Además, los organismos recombinantes pueden incluir, pero no se limitan a, animales tales como seres humanos y primates no humanos, caballos, vacas, cabras, cerdos, ovejas, perros, gatos, cobayas, ratas, ratones, lagartos, peces e insectos tales como especies de *Drosophila*. En otras realizaciones, el organismo es un hongo tal como una especie de *Candida*, *Neurospora* o *Saccharomyces*.

En algunas realizaciones, los procedimientos de la invención implican la introducción de una secuencia de interés en una célula tal como una célula germinal o citoblasto que puede convertirse en un organismo recombinante maduro o permitir que el organismo genéticamente modificado resultante dé lugar a progenie que lleva la secuencia de interés insertada en su genoma.

Las proteínas meganucleasas pueden administrarse en células para escindir ADN genómico, que permite recombinación homóloga o unión de extremos no homólogos en el sitio de escisión con una secuencia de interés, mediante una variedad de diferentes mecanismos conocidos en la técnica. Por ejemplo, la proteína meganucleasa recombinante puede introducirse en una célula por técnicas que incluyen, pero no se limitan a, microinyección o transfecciones de liposomas (véase, por ejemplo, Lipofectamine™, Invitrogen Corp., Carlsbad, CA). La formulación de liposoma puede usarse para facilitar la fusión de bicapas de lípidos con una célula diana, permitiendo así que el contenido del liposoma o proteínas asociadas a su superficie se lleve a la célula. Alternativamente, la enzima puede fusionarse con un péptido de captación apropiado tal como el de la proteína TAT del VIH para dirigir la captación celular (véase, por ejemplo, Hudecz y col. (2005), Med. Res. Rev. 25: 679-736).

Alternativamente, las secuencias de genes que codifican la proteína meganucleasa se insertan en un vector y se transfectan en una célula eucariota usando técnicas conocidas en la técnica (véase, por ejemplo, Ausubel y col., Current Protocols in Molecular Biology, Wiley 1999). La secuencia de interés puede introducirse en el mismo vector, un vector diferente, o por otros medios conocidos en la técnica.

Ejemplos no limitantes de vectores para la transfección de ADN incluyen vectores de virus, plásmidos, cósmidos y vectores YAC. La transfección de secuencias de ADN puede llevarse a cabo mediante una variedad de procedimientos conocidos para aquellos expertos en la materia. Por ejemplo, liposomas e inmunoliposomas se usan para administrar secuencias de ADN a células (véase, por ejemplo, Lasic y col. (1995), Science 267: 1275-76). Además, pueden utilizarse virus para introducir vectores en células (véase, por ejemplo, la patente de EE.UU. nº 7.037.492). Alternativamente pueden utilizarse estrategias de transfección de forma que los vectores se introduzcan como ADN desnudo (véase, por ejemplo, Rui y col. (2002), Life Sci. 71(15): 1771-8).

Procedimientos generales para administrar ácidos nucleicos en células incluyen: (1) procedimientos químicos (Graham y col. (1973), Virology 54(2):536-539; Zatloukal y col. (1992), Ann. N.Y. Acad. Sci., 660:136-153; (2)

procedimientos físicos tales como microinyección (Capecchi (1980), Cell 22(2):479-488, electroporación (Wong y col. (1982), Biochim. Biophys. Res. Commun. 107(2):584-587; Fromm y col. (1985), Proc. Nat'l Acad. Sci. USA 82(17):5824-5828; la patente de EE.UU. nº 5.384.253) e inyección balística (Johnston y col. (1994), Methods Cell. Biol. 43(A): 353-365; Fynan y col. (1993), Proc. Nat'l Acad. Sci. USA 90(24): 11478-11482); (3) vectores virales (Clapp (1993), Clin. Perinatol. 20(1): 155-168; Lu y col. (1993), J. Exp. Med. 178(6):2089-2096; Eglitis y col. (1988), Avd. Exp. Med. Biol. 241:19-27; Eglitis y col. (1988), Biotechniques 6(7):608-614); y (4) mecanismos mediados por receptor (Curiel y col. (1991), Proc. Nat'l Acad. Sci. USA 88(19):8850-8854; Curiel y col. (1992), Hum. Gen. Ther. 3(2):147-154; Wagner y col. (1992), Proc. Nat'l Acad. Sci. USA 89 (13):6099-6103).

En ciertas realizaciones se produce una planta genéticamente modificada, que contiene la secuencia de interés insertada en el genoma. En ciertas realizaciones, la planta genéticamente modificada se produce transflectando la célula de planta con secuencias de ADN correspondientes a la meganucleasa recombinante y la secuencia de interés, que puede o puede no flanquearse por la secuencias de reconocimiento de meganucleasa y/o secuencias sustancialmente idénticas a la secuencia diana. En otras realizaciones, la planta genéticamente modificada se produce transflectando la célula de planta con secuencias de ADN correspondientes a la meganucleasa recombinante sola, de forma que la escisión promueva la unión de extremos no homólogos y altere la secuencia diana que contiene la secuencia de reconocimiento. En tales realizaciones, las secuencias de meganucleasa están bajo el control de secuencias reguladoras que permiten la expresión de la meganucleasa en las células vegetales huésped. Estas secuencias reguladoras incluyen, pero no se limitan a, promotores de planta constitutivos tales como el promotor NOS, promotores de genes químicamente inducibles tales como el promotor inducible de dexametasona (véase, por ejemplo, Gremillon y col. (2004), Plant J. 37:218-228) y promotores específicos de tejido de planta tales como el promotor LGC1 (véase, por ejemplo, Singh y col. (2003), FEBS Lett. 542:47-52).

Procedimientos adecuados para introducir ADN en células vegetales incluyen prácticamente cualquier procedimiento por el que el ADN pueda introducirse en una célula, que incluye, pero no se limitan a, infección de *Agrobacterium*, transformación mediada por PEG de protoplastos (Omirulleh y col. (1993), Plant Molecular Biology, 21:415-428), captación de ADN mediada por desecación/inhibición, electroporación, agitación con fibras de carburo de silicio, inyección balística o bombardeo con microproyectiles, y similares.

En otras realizaciones, un animal genéticamente modificado se produce usando una meganucleasa recombinante. Al igual que con células vegetales, las secuencias de ácidos nucleicos pueden introducirse en una célula germinal o una célula que eventualmente se convertirá en organismo transgénico. En algunas realizaciones, la célula es un óvulo fecundado, y moléculas de ADN exógeno pueden inyectarse en el pronúcleo del ovulo fecundado. Los óvulos micro-inyectados se transfieren entonces a los oviductos de madres de cría pseudopreñadas y se deja que se desarrollen. La meganucleasa recombinante se expresa en el óvulo fecundado (por ejemplo, bajo el control de un promotor constitutivo, tal como 3-fosfoglicerato cinasa), y facilita la recombinación homóloga de la secuencia de interés en uno o algunos sitios discretos en el genoma. Alternativamente, los animales genéticamente modificados pueden obtenerse utilizando embriocitoblastos ("ES") recombinantes para la generación de los transgénicos, como se describe por Gossler y col. (1986), Proc. Natl. Acad. Sci. USA 83:9065 9069.

En ciertas realizaciones, un vector recombinante de expresión de mamífero puede dirigir la expresión específica de tejido del ácido nucleico preferencialmente en un tipo de célula particular. Los elementos reguladores específicos de tejido se conocen en la técnica. Ejemplos no limitantes de promotores específicos de tejido adecuados incluyen el promotor de albúmina (específico de hígado; Pinkert y col. (1987), Genes Dev. 1: 268-277), promotores específicos de linfocitos (Calame y Eaton (1988), Adv. Immunol. 43: 235-275), en particular promotores de receptores de linfocitos T (Winoto y Baltimore (1989), EMBO J. 8: 729-733) e inmunoglobulinas (Banerji y col. (1983), Cell 33: 729-740; Queen y Baltimore (1983), Cell 33: 741-748), promotores específicos de neuronas (por ejemplo, el promotor del neurofilamento; Byrne y Ruddle (1989), Proc. Natl. Acad. Sci. USA 86: 5473-5477), promotores específicos del páncreas (Edlund y col. (1985), Science 230: 912-916) y promotores específicos de las glándulas mamarias (por ejemplo, promotor de suero de la leche; la patente de EE.UU. nº 4.873.316 y publicación de patente europea EP 0 264 166). También están englobados promotores regulados por el desarrollo, por ejemplo, los promotores de hox murino (Kessel y Gruss (1990), Science 249: 374-379) y el promotor de α -fetoproteína (Campes y Tilghman (1989), Genes Dev. 3: 537-546).

En ciertas realizaciones, una meganucleasa racionalmente diseñada puede marcarse con un epítipo de péptido (por ejemplo, un epítipo de HA, FLAG o Myc) para monitorizar niveles de expresión o localización. En algunas realizaciones, la meganucleasa puede fusionarse con una señal de localización subcelular tal como una señal de localización nuclear (por ejemplo, la señal de localización nuclear del SV40) o señales de localización de cloroplastos o mitocondriales. En otras realizaciones, la meganucleasa puede fusionarse con una señal de exportación nuclear para localizarla en el citoplasma. La meganucleasa también puede fusionarse con una proteína o dominio de proteína sin relacionar tal como una proteína que estimula la reparación de ADN o recombinación homóloga (por ejemplo, recA, RAD51, RAD52, RAD54, RAD57 o BRCA2).

6. Procedimientos para terapia génica

Aspectos de la invención permiten el uso de meganucleasa recombinante para terapia génica. Como se usa en el

presente documento, "terapia génica" significa tratamientos terapéuticos que comprenden introducir en un paciente una copia funcional de al menos un gen, o secuencia reguladora de genes tal como un promotor, potenciador o silenciador para sustituir un gen o región reguladora de genes que es defectuosa en su estructura y/o función. El término "terapia génica" también puede referirse a modificaciones hechas a un gen perjudicial o elemento regulador (por ejemplo, oncogenes) que reduce o elimina la expresión del gen. La terapia génica puede realizarse para tratar afecciones congénitas, afecciones resultantes de mutaciones o lesión a sitios genéticos específicos durante la vida del paciente, o afecciones resultantes de organismos infecciosos.

En algunos aspectos de la invención, genes disfuncionales se sustituyen o inutilizan por la inserción de secuencias de ácidos nucleicos exógenos en una región del genoma que afecta la expresión génica. En ciertas realizaciones, la meganucleasa recombinante es elegida como diana por una secuencia particular en la región del genoma que va a modificarse de manera que se alivie la afección. La secuencia puede ser una región dentro de un exón, intrón, promotor, u otra región reguladora que esté causando expresión disfuncional del gen. Como se usa en el presente documento, el término "expresión disfuncional" significa expresión anómala de un producto génico tanto por la célula que produce demasiado poco producto génico, demasiado producto génico, o que produce un producto génico que tiene una función diferente tal como que carece de la función necesaria o que tiene más de la función necesaria.

Las secuencias de ácidos nucleicos exógenas insertadas en la región modificada pueden usarse para proporcionar secuencias "reparadas" que normalizan el gen. La reparación del gen puede llevarse a cabo por la introducción de secuencias de genes apropiadas en el gen permitiendo restablecer la función apropiada. En estas realizaciones, la secuencia de ácidos nucleicos que va a insertarse puede ser la secuencia codificante entera para una proteína o, en ciertas realizaciones, un fragmento del gen que comprende solo la región que va a repararse. En otras realizaciones, la secuencia de ácidos nucleicos que va a insertarse comprende una secuencia promotora u otros elementos reguladores de forma que se reparen mutaciones que producen expresión o regulación anormal. En otras realizaciones, la secuencia de ácidos nucleicos que va a insertarse contiene el codón de terminación de la traducción apropiado que carece en un gen mutado. La secuencia de ácidos nucleicos también puede tener secuencias para parar la transcripción en un gen recombinante que carece de señales de parada de la transcripción apropiadas.

Alternativamente, las secuencias de ácidos nucleicos pueden eliminar la función del gen completamente alterando la secuencia reguladora del gen o proporcionando un silenciador para eliminar la función génica. En algunas realizaciones, la secuencia de ácidos nucleicos exógenos proporciona un codón de terminación de la traducción para prevenir la expresión del producto génico. En otras realizaciones, las secuencias de ácidos nucleicos exógenas proporcionan elemento de parada de la transcripción para prevenir la expresión de una molécula de ARN de longitud completa. En otras realizaciones más, la función génica es directamente alterada por la meganucleasa introduciendo inserciones de bases, deleciones de bases y/o mutaciones de desplazamiento del marco mediante unión de extremos no homólogos.

En muchos casos se desea dirigir las secuencias genéticas apropiadas a una célula diana o población de células que es la causa de la condición de enfermedad. Tal elección como diana de terapéuticos previene que las células sanas sean elegidas como diana por los agentes terapéuticos. Esto aumenta la eficacia del tratamiento, a la vez que disminuye los posibles efectos adversos que el tratamiento podría tener sobre células sanas.

La administración de genes recombinantes de meganucleasa y la secuencia de interés que va a insertarse en el genoma a las células de interés puede llevarse a cabo mediante una variedad de mecanismos. En algunas realizaciones, los ácidos nucleicos se administran a las células a modo de virus con genes virales particulares inactivados para prevenir la reproducción del virus. Así, un virus puede alterarse de manera que sea capaz solo de administración y mantenimiento dentro de una célula diana, pero no retenga la capacidad para replicarse dentro de la célula diana o tejido. Una o más secuencias de ADN pueden introducirse al genoma viral alterado, de manera que se produzca un genoma viral que actúe de vector, y pueda o pueda no insertarse en un genoma del huésped y posteriormente expresarse. Más específicamente, ciertas realizaciones incluyen emplear un vector retroviral tal como, pero no se limita a, los vectores MFG o pLJ. Un vector MFG es un vector de virus de la leucemia murina de Moloney (MoMLV) simplificado en el que las secuencias de ADN que codifican las proteínas pol y env se han delecionado para hacer defectuosa la replicación. Un vector retroviral pLJ también es una forma del MoMLV (véase, por ejemplo, Korman y col. (1987), Proc. Nat'l Acad. Sci., 84:2150-2154). En otras realizaciones, un adenovirus recombinante o virus adeno-asociado puede usarse como vector de administración.

En otras realizaciones, la administración de proteína de meganucleasa recombinante y/o secuencias de genes de meganucleasa recombinante a una célula diana se lleva a cabo usando liposomas. La producción de liposomas que contienen ácido nucleico y/o carga de proteína se conoce en la técnica (véase, por ejemplo, Lasic y col. (1995), Science 267: 1275-76). Los inmunoliposomas incorporan anticuerpos contra antígenos asociados a células en liposomas, y pueden administrar secuencias de ADN para la meganucleasa o la propia meganucleasa para tipos de células específicos (véase, por ejemplo, Lasic y col. (1995), Science 267: 1275-76; Young y col. (2005), J. Calif. Dent. Asoc. 33(12): 967-71; Pfeiffer y col. (2006), J. Vasc. Surg. 43(5):1021-7). Los procedimientos para producir y usar formulaciones de liposomas son muy conocidos en la técnica, (véase, por ejemplo, la patente de EE.UU. n° 6.316.024, la patente de EE.UU. n° 6.379.699, la patente de EE.UU. n° 6.387.397, la patente de EE.UU. n°

6.511.676 y la patente de EE.UU. nº 6.593.308, y referencias citadas en su interior). En algunas realizaciones, los liposomas se usan para administrar las secuencias de interés, además de la proteína meganucleasa recombinante o secuencias de genes de meganucleasa recombinante.

5 7. Procedimientos para tratar infección por patógeno.

Aspectos de la invención también proporcionan procedimientos para tratar infección por un patógeno. Los organismos patógenos incluyen virus tales como, pero no se limitan a, virus del herpes simple 1, virus del herpes simple 2, virus de la inmunodeficiencia humana 1, virus de la inmunodeficiencia humana 2, virus de la viruela, virus de la poliomielitis, virus de Epstein-Barr y virus del papiloma humano y organismos bacterianos tales como, pero no se limitan a, *Bacillus anthracis*, especies de *Haemophilus*, especies de *Pneumococcus*, *Staphylococcus aureus*, especies de *Streptococcus*, *Staphylococcus aureus* resistente a meticilina y *Mycoplasmun tuberculosis*. Los organismos patógenos también incluyen organismos fúngicos tales como, pero no se limitan a, especies de *Candida*, *Blastomyces*, *Cryptococcus* e *Histoplasma*.

En algunas realizaciones, una meganucleasa racionalmente diseñada puede ser elegida como diana para una secuencia de reconocimiento dentro del genoma del patógeno, por ejemplo, para un gen o elemento regulador que es esencial para el crecimiento, reproducción o toxicidad del patógeno. En ciertas realizaciones, la secuencia de reconocimiento puede estar en un plásmido bacteriano. La escisión mediada por meganucleasas de una secuencia de reconocimiento en un genoma del patógeno puede estimular la mutación dentro de un gen esencial elegido como diana en forma de una inserción, delección o desplazamiento de marco, estimulando la unión de extremos no homólogos. Alternativamente, la escisión de un plásmido bacteriano puede producir pérdida del plásmido junto con cualquier gen codificado por él, tal como genes de toxina (por ejemplo, gen del factor letal de *B. anthracis*) o genes de resistencia a antibióticos. Como se observa anteriormente, la meganucleasa puede administrarse al paciente, animal o planta infectada en tanto en forma de proteína como de ácido nucleico usando técnicas que son comunes en la materia. En ciertas realizaciones, el gen de meganucleasa puede incorporarse en un genoma de bacteriófago para la administración a bacterias patógenas.

Los aspectos de la invención también proporcionan terapéuticos para el tratamiento de ciertas formas de cáncer. Debido a que los virus humanos frecuentemente están asociados a la formación de tumores (por ejemplo, virus de Epstein-Barr y carcinomas nasofaríngeos; virus del papiloma humano y cáncer de cuello uterino), la inactivación de estos patógenos virales puede prevenir desarrollo o progresión del cáncer. Alternativamente, las roturas bicatenarias elegidas como diana para los genomas de estos virus asociados a tumor usando meganucleasas racionalmente diseñadas pueden usarse para desencadenar apoptosis mediante la ruta de respuesta al daño de ADN. De este modo, puede ser posible inducir selectivamente apoptosis en células tumorales que alojan el genoma viral.

8. Procedimientos de genotipado e identificación de patógenos

Aspectos de la invención también proporcionan herramientas para investigación y desarrollo de biología molecular *in vitro*. Es común en la materia usar endonucleasas específicas para sitio (por ejemplo, enzimas de restricción) para el aislamiento, clonación y manipulación de ácidos nucleicos tales como plásmidos, productos de PCR, secuencias de BAC, secuencias de YAC, virus y secuencias genómicas de organismos eucariotas y procariotas (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, Wiley 1999). Así, en algunas realizaciones, una meganucleasa racionalmente diseñada puede usarse para manipular secuencias de ácidos nucleicos *in vitro*. Por ejemplo, las meganucleasas racionalmente diseñadas que reconocen un par de secuencias de reconocimiento dentro de la misma molécula de ADN pueden usarse para aislar el segmento de ADN interviniente para la posterior manipulación tal como ligación en un plásmido bacteriano, BAC o YAC.

En otro aspecto, la presente invención proporciona herramientas para la identificación de genes y organismos patógenos. En una realización, meganucleasas racionalmente diseñadas pueden usarse para escindir sitios de reconocimiento correspondientes a regiones genéticas polimórficas correlacionados con enfermedad para distinguir alelos causantes de enfermedad de alelos sanos (por ejemplo, una meganucleasa racionalmente diseñada que reconoce el alelo AF-508 del gen CFTR humano, véase el Ejemplo 4). En esta realización, las secuencias de ADN aisladas de un paciente humano u otro organismo se digieren con una meganucleasa racionalmente diseñada, posiblemente conjuntamente con nucleasas específicas para sitio adicionales, y el patrón de fragmento de ADN resultante se analiza por electroforesis en gel, electroforesis capilar, espectrometría de masas, u otros procedimientos conocidos en la técnica. Este patrón de fragmentación y, específicamente, la presencia o ausencia de escisión por la meganucleasa racionalmente diseñada, indica el genotipo del organismo revelando si la secuencia de reconocimiento está presente o no en el genoma. En otra realización, una meganucleasa racionalmente diseñada es elegida como diana por una región polimórfica en el genoma de un virus patógeno, hongo o bacteria y se usa para identificar el organismo. En esta realización, la meganucleasa racionalmente diseñada escinde una secuencia de reconocimiento que es única para el patógeno (por ejemplo, la región espaciadora entre los genes ARNr 16S y 23S en una bacteria; véase, por ejemplo, van der Giessen y col. (1994), *Microbiology* 140:1103-1108) y puede usarse para distinguir el patógeno de otros organismos estrechamente relacionados tras la digestión con endonucleasa del genoma y posterior análisis del patrón de fragmentación por electroforesis, espectrometría de masas u otros procedimientos conocidos en la técnica.

9. Procedimientos para la producción de dominios de unión a ADN a medida.

En otro aspecto, la invención proporciona proteínas de unión a ADN racionalmente diseñadas que carecen de actividad de escisión de endonucleasas. La actividad catalítica de una meganucleasa racionalmente diseñada puede eliminarse mutando aminoácidos implicados en la catálisis

EJEMPLOS

La presente invención se ilustra adicionalmente por los siguientes ejemplos. Los ejemplos que no se encuentran dentro del alcance de la invención son solo para fines ilustrativos. Aquellos expertos en la materia reconocerán, o serán capaces de determinar, usando nada más que experimentación rutinaria, numerosos equivalentes a las sustancias específicas y procedimientos descritos en el presente documento. Tales equivalentes pretenden estar englobados en el alcance de las reivindicaciones que siguen a los ejemplos más adelante. Los Ejemplos 1-4 a continuación se refieren específicamente a meganucleasas racionalmente diseñadas basadas en I-Crel, pero similarmente pueden producirse y usarse meganucleasas racionalmente diseñadas basadas en I-SceI, I-MsoI, I-CeuI, y otras meganucleasas LAGLIDADG, como se describe en el presente documento.

EJEMPLO 1

Diseño racional de meganucleasas que reconocen el gen TAT del VIH-1

1. Diseño de meganucleasas.

Se diseñaron un par de meganucleasas para reconocer y escindir el sitio de ADN 5'-GAAGAGCTCATCAGAACAGTCA-3' (SEC ID N°: 15) encontrado en el gen TAT del VIH-1. Según la Tabla 1 se diseñaron dos meganucleasas, TAT1 y TAT2, para unirse a los medios sitios 5'-GAAGAGCTC-3' (SEC ID N°: 16) y 5'-TGACTGTTC-3' (SEC ID N°: 17), respectivamente, usando los siguientes contactos de bases (los contactos no naturales están en negrita):

TAT1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	G	A	A	G	A	G	C	T	C
Residuos de contacto	<u>S32</u>	<u>Y33</u>	<u>N30/ Q38</u>	<u>R40</u>	<u>K28</u>	<u>S26/ R77</u>	<u>K24/ Y68</u>	<u>Q44</u>	<u>R70</u>

TAT2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	A	C	T	G	T	T	C
Residuos de contacto	<u>C32</u>	<u>R33</u>	<u>N30/ Q38</u>	<u>R28/ E40</u>	<u>M66</u>	<u>S26/ R77</u>	<u>Y68</u>	<u>Q44</u>	<u>R70</u>

Las dos enzimas se clonaron, se expresaron en *E. coli* y se ensayaron para actividad enzimática contra la secuencia de reconocimiento de ADN correspondiente como se describe más adelante. En ambos casos se encontró que las meganucleasas racionalmente diseñadas eran inactivas. Entonces se produjo una segunda generación de cada una en la que E80 se mutó a Q para mejorar contactos con el esqueleto de ADN. Se encontró que la enzima TAT2 de segunda generación era activa contra su secuencia de reconocimiento prevista, mientras que la enzima TAT1 de segunda generación siguió siendo inactiva. La inspección visual de la estructura cristalina de I-Crel natural sugirió que TAT1 era inactiva debido a una discrepancia estérica entre R40 y K28. Para aliviar esta discrepancia se produjeron variantes de TAT1 en las que K28 se mutó a un aminoácido con una cadena lateral más pequeña (A, S, T o C) mientras que se mantenía la mutación Q80. Cuando estas enzimas se produjeron en *E. coli* y se ensayaron, se encontró que ambas variantes de TAT1 con S28 y T28 eran activas contra la secuencia de reconocimiento prevista, a la vez que se mantenía la preferencia de bases deseada en la posición -7.

2. Construcción de meganucleasas recombinantes.

Las mutaciones para las enzimas I-Crel rediseñadas se introdujeron usando cebadores mutagénicos en una estrategia de PCR de solapamiento. Los fragmentos de ADN recombinantes de I-Crel generados en una PCR primaria se unieron en una PCR secundaria para producir ácidos nucleicos recombinantes de longitud completa. Todas las construcciones de I-Crel recombinantes se clonaron en vectores pET21a con una marca de seis histidinas fusionadas en el extremo 3' del gen para la purificación (Novagen Corp., San Diego, CA). Todas las secuencias de ácidos nucleicos se confirmaron usando secuenciación de didesoxinucleótido de Sanger (véase Sanger y col. (1977), Proc. Natl. Acad. Sci. USA. 74(12): 5463-7).

Las meganucleasas I-Crel naturales y todas las manipuladas se expresaron y purificaron usando el siguiente procedimiento. Las construcciones clonadas en un vector pET21a se transformaron en BL21 (DE3) pLysS químicamente competentes, y se sembraron en placas 2xYT convencionales que contenían 200 µg/ml de carbanicilina. Tras el crecimiento durante la noche, las colonias bacterianas transformadas se rasparon de las placas y se usaron para inocular 50 ml de caldo 2XYT. Las células se cultivaron a 37 °C agitando hasta que alcanzó una

densidad óptica de 0,9 a una longitud de onda de 600 nm. Entonces, la temperatura de crecimiento se redujo de 37 °C a 22 °C. La expresión de proteínas se indujo mediante la adición de IPTG 1 mM, y las células se incubaron con agitación durante dos horas y media. Entonces, las células se sedimentaron por centrifugación durante 10 min a 6000 xg. Los sedimentos se resuspendieron en 1 ml de tampón de unión (Tris-HCl 20 mM, pH 8,0, NaCl 500 mM, imidazol 10 mM) por agitación con vórtex. Entonces, las células se rompieron con 12 pulsos de sonicación al 50% de potencia y el residuo de células se sedimentó por centrifugación durante 15 min a 14.000 xg. Los sobrenadantes de células se diluyeron en 4 ml de tampón de unión y se cargaron sobre una columna Sepharose quelante de metales cargada con níquel de 200 µl (Pharmacia).

La columna se lavó posteriormente con 4 ml de tampón de lavado (Tris-HCl 20 mM, pH 8,0, NaCl 500 mM, imidazol 60 mM) y con 0,2 ml de tampón de elución (Tris-HCl 20 mM, pH 8,0, NaCl 500 mM, imidazol 400 mM). Las enzimas meganucleasas se eluyeron con 0,6 ml adicionales de tampón de elución y se concentraron a 50-130 µl usando concentradores desechables Vivospin (ISC, Inc., Kaysville, UT). Las enzimas se intercambiaron en tampón SA (Tris-HCl 25 mM, pH 8,0, NaCl 100 mM, MgCl₂ 5 mM, EDTA 5 mM) para ensayos y almacenamiento usando columnas de desalación por centrifugación Zeba (Pierce Biotechnology, Inc., Rockford, IL). La concentración de enzima se determinó por absorbancia a 280 nm usando un coeficiente de extinción de 23.590 M⁻¹cm⁻¹. La pureza y peso molecular de las enzimas se confirmó entonces por espectrometría de masas MALDI-TOF.

Las enzimas heterodiméricas se produjeron tanto purificando las dos proteínas independientemente como mezclándolas *in vitro* o construyendo un operón artificial para expresión en tándem de las dos proteínas en *E. coli*. En el primer caso, las meganucleasas purificadas se mezclaron 1:1 en disolución y se pre-incubaron a 42 °C durante 20 minutos antes de la adición de sustrato de ADN. En el último caso, los dos genes se clonaron secuencialmente en el vector de expresión pET-21a usando *NdeI/EcoRI* y *EcoRI/HindIII*. En primer gen termina con dos codones de terminación para prevenir errores de ultralectura durante la transcripción. Un espaciador de ácidos nucleicos de 12 pares de bases y una secuencia de Shine-Dalgarno del vector pET21 separó el primer y segundo genes en el operón artificial.

3. Ensayos de escisión.

Todas las enzimas purificadas como se ha descrito anteriormente se ensayaron para actividad por incubación con sustratos de ADN bicatenario lineal que contenían la secuencia de reconocimiento de meganucleasa. Los oligonucleótidos sintéticos correspondientes a hebras tanto codificantes como no codificantes de la secuencia de reconocimiento se hibridaron y se clonaron en el sitio *SmaI* del plásmido pUC19 por ligación de extremos romos. Las secuencias de los sitios de unión clonados se confirmaron por secuenciación de didesoxinucleótido de Sanger. Todos los sustratos de plásmido se linealizaron con *XmnI*, *Scal* o *BpmI* simultáneamente con la digestión de meganucleasa. Las digestiones de enzima contuvieron 5 µl de sustrato de ADN 0,05 µM, 2,5 µl de meganucleasa recombinante I-Crel 5 µM, 9,5 µl de tampón SA y 0,5 µl de *XmnI*, *Scal* o *BpmI*. Las digestiones se incubaron a tanto 37 °C como a 42 °C para ciertas enzimas meganucleasas, durante cuatro horas. Las digestiones se detuvieron añadiendo 0,3 mg/ml de proteinasa K y 0,5% de SDS, y se incubaron durante una hora a 37 °C. Las digestiones se analizaron sobre 1,5% de agarosa y se visualizaron por tinción con bromuro de etidio.

Para evaluar la preferencia de medios sitios de meganucleasa, las meganucleasas racionalmente diseñadas se incubaron con un conjunto de sustratos de ADN correspondientes para un palíndromo perfecto del medio sitio previsto, además de cada una de las 27 sustituciones de pares de bases individuales posibles en el medio sitio. De este modo fue posible determinar cómo de tolerante es cada enzima a desviaciones de su medio sitio previsto.

4. Especificidad de secuencias de reconocimiento.

Las meganucleasas TAT1 y TAT2 recombinantes purificadas reconocieron secuencias de ADN que fueron distintas de la secuencia de reconocimiento de meganucleasa natural (Fig. 2(B)). La meganucleasa I-Crel natural escinde la secuencia de reconocimiento WT, pero no corta ni la secuencia prevista para TAT1 ni la secuencia prevista para TAT2. TAT1 y TAT2, asimismo, cortan sus secuencias de reconocimiento previstas, pero no la secuencia natural. Las meganucleasas se evaluaron entonces para preferencia de medios sitios y especificidad global (Fig 3). Se encontró que I-Crel natural era altamente tolerante con sustituciones de pares de bases individuales en su medio sitio natural. A diferencia, se encontró que TAT1 y TAT2 eran altamente específicas y completamente intolerantes con sustituciones de bases en las posiciones -1, -2, -3, -6 y -8 en el caso de TAT1, y posiciones -1, -2 y -6 en el caso de TAT2.

EJEMPLO 2

Diseño racional de meganucleasas con afinidad de unión a ADN alterada

1. Meganucleasas con elevada afinidad y elevada actividad.

Las meganucleasas CCR1 y BRP2 se diseñaron para escindir los medios sitios 5'-AACCCCTCTC-3' (SEC ID N°: 18) y 5'-CTCCGGGTC-3' (SEC ID N°: 19), respectivamente. Estas enzimas se produjeron según la Tabla 1 como en el

Ejemplo 1:

CCR1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	A	C	C	C	T	C	T	C
Residuos de contacto	N32	Y33	R30/ E38	R28/ E40	E42	Q26	K24/ Y68	Q44	R70

BRP2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	T	C	C	G	G	G	T	C
Residuos de contacto	S32	C33	R30/ E38	R28/ E40	R42	S26/ R77	R68	Q44	R70

Ambas enzimas se expresaron en *E. coli*, se purificaron y se ensayaron como en el Ejemplo 1. Se encontró que ambas enzimas de la primera generación escondían sus secuencias de reconocimiento previstas con tasas que fueron considerablemente inferiores a las de I-Crel natural con su secuencia de reconocimiento natural. Para aliviar esta pérdida en actividad, la afinidad de unión a ADN de CCR1 y BRP2 aumentó mutando E80 a Q en ambas enzimas. Se encontró que estas versiones de segunda generación de CCR1 y BRP2 escindieron sus secuencias de reconocimiento con tasas catalíticas sustancialmente elevadas.

2. Meganucleasas con afinidad de unión a ADN disminuida y actividad disminuida, pero especificidad elevada.

Se encontró que I-Crel natural era altamente tolerante con sustituciones en su medio sitio (Fig. 3(A)). En un esfuerzo por hacer la enzima más específica, la lisina en la posición 116 de la enzima, que normalmente hace un puente de sal con un fosfato en el esqueleto de ADN, se mutó a ácido aspártico para reducir la afinidad de unión a ADN. Se encontró que esta enzima racionalmente diseñada escindió la secuencia de reconocimiento natural con actividad sustancialmente reducida, pero la enzima recombinante fue considerablemente más específica que la natural. La preferencia de medios sitios de la variante K116D se evaluó como en el Ejemplo 1 y se encontró que la enzima era totalmente intolerante a desviación de su medio sitio natural en las posiciones -1, -2 y -3, y mostró preferencia de bases al menos parcial en las 6 posiciones restantes en el medio sitio (Fig. 3(B)).

EJEMPLO 3

Heterodímeros de meganucleasas racionalmente diseñadas

1. Escisión de sitios de ADN no palindrómicos por heterodímeros de meganucleasas formados en disolución.

Se diseñaron dos meganucleasas, LAM1 y LAM2, para escindir los medios sitios 5'-TGCGGTGTC-3' (SEC ID N°: 20) y 5'-CAGGCTGTC-3' (SEC ID N°: 21), respectivamente. Era de esperar que el heterodímero de estas dos enzimas reconociera la secuencia de ADN 5'-TGCGGTGTCCGCGACAGCCTG-3' (SEC ID N°: 22) encontrada en el gen p05 del bacteriófago λ.

LAM1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	G	G	T	G	T	C
Residuos de contacto	C32	R33	R30/ E38	D28/ R40	R42	Q26	R68	Q44	R70

LAM2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	G	G	C	T	G	T	C
Residuos de contacto	S32	Y33	E30/ R38	R40	K28/ E42	Q26	R68	Q44	R70

LAM1 y LAM2 se clonaron, se expresaron en *E. coli* y se purificaron individualmente como se describe en el Ejemplo 1. Las dos enzimas se mezclaron entonces 1:1 y se incubaron a 42 °C durante 20 minutos para permitir que intercambiaran subunidades y se reequilibraran. La disolución de enzimas resultante, que se esperaba que fuera una mezcla de homodímero LAM1, homodímero LAM2 y heterodímero LAM1/LAM2, se incubó con tres secuencias de reconocimiento diferentes correspondientes al palíndromo perfecto del medio sitio de LAM1, el palíndromo perfecto del medio sitio de LAM2 y el sitio híbrido no palindrómico encontrado en el genoma del bacteriófago λ. La enzima LAM1 purificada sola corta el sitio palindrómico de LAM1, pero ni el sitio palindrómico de LAM2 ni el sitio híbrido de LAM1/LAM2. Asimismo, la enzima LAM2 purificada sola corta el sitio palindrómico de LAM2, pero ni el sitio palindrómico de LAM1 ni el sitio híbrido de LAM1/LAM2. Sin embargo, la mezcla 1:1 de LAM1 y LAM2 escinde los tres sitios de ADN. La escisión del sitio híbrido de LAM1/LAM2 indica que dos meganucleasas rediseñadas distintas pueden mezclarse en disolución para formar una enzima heterodimérica que puede escindir un sitio de ADN no palindrómico.

2. Escisión de sitios de ADN no palindrómicos por heterodímeros de meganucleasas formados por co-expresión.

Los genes que codifican las enzimas LAM1 y LAM2 descritas anteriormente se dispusieron en un operón para expresión simultánea en *E. coli* como se describe en el Ejemplo 1. Las enzimas co-expresadas se purificaron como en el Ejemplo 1 y la mezcla de enzimas se incubó con las tres posibles secuencias de reconocimiento descritas anteriormente. Se encontró que la mezcla de enzimas co-expresada escindió los tres sitios, que incluyen el sitio híbrido de LAM1/LAM2, que indica que dos meganucleasas racionalmente diseñadas distintas pueden co-expresarse para formar una enzima heterodimérica que puede escindir un sitio de ADN no palindrómico.

3. Escisión preferencial de sitios de ADN no palindrómicos por heterodímeros de meganucleasas con superficies de separación proteína-proteína modificadas.

Para aplicaciones que requieren la escisión de sitios de ADN no palindrómicos se desea promover la formación de heterodímeros de enzima mientras que se minimizan la formación de homodímeros que reconocen y escinden diferentes sitios de ADN (palindrómicos). Para este fin, se produjeron variantes de la enzima LAM1 en las que lisinas en las posiciones 7, 57 y 96 se cambiaron a ácidos glutámicos. Entonces esta enzima se co-expresó y se purificó como antes con una variante de LAM2 en la que ácidos glutámicos en las posiciones 8 y 61 se cambiaron a lisina. En este caso, era de esperar que la formación del homodímero LAM1 se redujera debido a repulsión electrostática entre E7, E57 y E96 en un monómero y E8 y E61 en el otro monómero. Asimismo, era de esperar que la formación del homodímero LAM2 se redujera debido a repulsión electrostática entre K7, K57 y K96 en un monómero y K8 y K61 en el otro monómero. En cambio, era de esperar que el heterodímero de LAM1/LAM2 se favoreciera debido a atracción electrostática entre E7, E57 y E96 en LAM1 y K8 y K61 en LAM2. Cuando las dos meganucleasas con superficies de separación modificadas se co-expresaron y se ensayaron como se ha descrito anteriormente, se encontró que el sitio híbrido de LAM1/LAM2 se escindió preferencialmente con respecto a los dos sitios palindrómicos, que indica que sustituciones en la superficie de separación proteína-proteína de meganucleasa pueden accionar la formación preferencial de heterodímeros.

EJEMPLO 4

Heterodímeros de meganucleasas adicionales que escinden secuencias de ADN fisiológicas

1. Heterodímeros de meganucleasas que escinden secuencias de ADN relevantes para terapia génica.

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (ACH1/ACH2) que escinde la secuencia 5'-CTGGGAGTCTCAGGACAGCCTG-3' (SEC ID N°: 23) en el gen FGFR3 humano, mutaciones en las que produce acondroplasia. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

ACH1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	T	G	G	G	A	G	T	C
Residuos de contacto	D32	C33	E30/ R38	R40/ D28	R42	A26/ Q77	R68	Q44	R70

ACH2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	G	G	C	T	G	T	C
Residuos de contacto	D32	Y33	E30/ R38	R40	K28/ E42	Q26	R68	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (HGH1/HGH2) que escinde la secuencia 5'-CCAGGTGTCTCTGGACTCCTCC-3' (SEC ID N°: 24) en el promotor del gen de la hormona de crecimiento humana. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

HGH1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	C	A	G	G	T	G	T	C
Residuos de contacto	D32	C33	N30/ Q38	R40/ D28	R42	Q26	R68	Q44	R70

HGH2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	G	G	A	G	G	A	G	T	C
Residuos de contacto	K32	R33	N30/ Q38	R40/ D28	R42	A26	R68	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (CF1/CF2) que escinde la secuencia 5'-GAAAATATCATTGGTGTTCCT-3' (SEC ID N°: 25) en el alelo ΔF508 del gen CFTR humano. Por

ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

CF1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	G	A	A	A	A	T	A	T	C
Residuos de contacto	S32	Y33	N30/ Q38	Q40	K28	Q26	H68/ C24	Q44	R70

CF2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	G	G	A	A	A	C	A	C
Residuos de contacto	N32	R33	E30/ R38	Q40	K28	A26	Y68/ C24	T44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (CCR1/CCR2) que escinde la secuencia 5'-AACCTCTCCAGTGAGATGCCT-3' (SEC ID N°: 26) en el gen CCR5 humano (un co-receptor del VIH). Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

CCR1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	A	C	C	C	T	C	T	C
Residuos de contacto	N32	Y33	R30/ E38	E40/ R28	E42	Q26	Y68/ K24	Q44	R70

CCR2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	G	G	C	A	T	C	T	C
Residuos de contacto	N32	R33	E30/ R38	E40	K28	Q26	Y68/ K24	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (MYD1/MYD2) que escinde la secuencia 5'-GACCTCGTCCTCCGACTCGCTG-3' (SEC ID N°: 27) en la región sin traducir de 3' del gen DM cinasa humana. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

MYD1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	G	A	C	C	T	C	G	T	C
Residuos de contacto	S32	Y33	R30/ E38	E40/ R28	K66	Q26/ E77	R68	Q44	R70

MYD2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	G	C	G	A	G	T	C
Residuos de contacto	S32	Y33	E30/ R38	E40/ R28	R42	A26 Q77	R68	Q44	R70

2. Heterodímeros de meganucleasas que escinden secuencias de ADN en genoma del patógenos.

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (HSV1/HSV2) que escinde la secuencia 5'-CTCGATGTCGACGACACGGCA-3' (SEC ID N°: 28) en el gen UL36 del virus del herpes simple 1 y el virus del herpes simple 2. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

HSV1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	T	C	G	A	T	G	T	C
Residuos de contacto	S32	C33	R30/ E38	R40/	Q42/ K28	Q26	R68	Q44	R70

HSV2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	C	G	T	G	T	C
Residuos de contacto	C32	R33	R30/ E38	E40/ R28	R42	Q26	R68	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (ANT1/ANT2) que escinde la secuencia 5'-ACAAGTGTCTATGGACAGTTTA-3' (SEC ID N°: 29) en el genoma de *Bacillus anthracis*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

ANT1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	C	A	A	G	T	G	T	C
Residuos de contacto	N32	C33	N30/ Q38	Q40/ A28	R42	Q26	R68	Q44	R70

ANT2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	A	A	A	C	T	G	T	C
Residuos de contacto	C32	Y33	N30/Q38	Q40	E42	Q26	R68	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (POX1/POX2) que escinde la secuencia 5'-AAACTGTCAAATGACATCGCA-3' (SEC ID N°: 30) en el gen gp009 del virus variólico (viruela). Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

POX1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	A	A	A	C	T	G	T	C
Residuos de contacto	N32	C33	N30/ Q38	Q40	K28	Q26	R68	Q44	R70

POX2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	G	A	T	G	T	C
Residuos de contacto	C32	R33	R30/ E38	R40	C28/ Q42	Q26	R68	Q44	R70

Puede producirse un homodímero de meganucleasas racionalmente diseñadas (EBB1/BBB1) que escinde la secuencia pseudo-palindrómica 5'-CGGGGTCTCGTGCGAGGCCTCC-3' (SEC ID N°: 31) en el gen *BALF2* del virus de Epstein-Barr. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

EBB1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	G	G	G	G	T	C	T	C
Residuos de contacto	S32	R33	D30/ Q38	R40/ D28	R42	Q26	Y68/ K24	Q44	R70

EBB1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	G	G	A	G	G	C	C	T	C
Residuos de contacto	S32	R33	D30/ Q38	R40/ D28	R42	Q26	Y68/ K24	Q44	R70

3. Heterodímeros de meganucleasas que escinden secuencias de ADN en genomas de planta.

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (GLA1/GLA2) que escinde la secuencia 5'-CACTAACTCGTATGAGTCGGTG-3' (SEC ID N°: 32) en el gen *GL2* de *Arabidopsis thaliana*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

GLA1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	C	T	A	A	C	T	C
Residuos de contacto	S32	Y33	R30/ E38	S40/ C79	K28	A26/ Q77	Y68/ K24	Q44	R70

GLA2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	C	C	G	A	C	T	C
Residuos de contacto	S32	Y33	R30/ E38	E40/ R28	R42	A26	Y68/ K24	Q44	R70
						Q77			

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (BRP1/BRP2) que escinde la secuencia 5'-TGCCTCCTCTAGAGACCCGGAG-3' (SEC ID N°: 33) en el gen *BP1* de *Arabidopsis thaliana*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

BRP1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	C	T	C	C	T	C
Residuos de contacto	C32	R33	R30/ E38	R28/ E40	K66	Q26/ E77	Y68/ K24	Q44	R70

BRP2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	T	C	C	G	G	G	T	C
Residuos de contacto	S32	C33	R30/ E38	E40/ R28	R42	S26 R77	R68	Q44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (MGC1/MGC2) que escinde la secuencia 5'-TAAAATCTCTAAGGTCTGTGCA-3' (SEC ID N°: 34) en el gen magnesio quelatasa de *Nicotiana tabacum*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

MGC1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	A	A	A	A	T	C	T	C
Residuos de contacto	C32	Y33	N30/ Q38	Q40/	K28	Q26	Y68/ K24	Q44	R70

MGC2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	A	C	A	G	A	C
Residuos de contacto	S32	R33	R30/ E38	Q40	K28	A26 Q77	R68	T44	R70

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (CYP/HGH2) que escinde la secuencia 5'-CAAGAATTCAAGCGAGCATTAA-3' (SEC ID N°: 35) en el gen CYP82E4 de *Nicotiana tabacum*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

CYP:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	A	G	A	A	T	T	C
Residuos de contacto	D32	Y33	N30/ Q38	R40/	K28	Q77/ A26	Y68	Q44	R70

HGH2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	T	A	A	T	G	C	T	C
Residuos de contacto	S32	C33	N30/ Q38	Q40	K66	R77/ S26	Y68 K24	Q44	R70

4. Heterodímeros de meganucleasas que escinden secuencias de ADN en genomas de levadura.

Puede producirse un heterodímero de meganucleasas racionalmente diseñadas (URA1/URA2) que escinde la secuencia 5'-TTAGATGACAAGGGAGACGCAT-3' (SEC ID N°: 36) en el gen *URA3* de *Saccharomyces cerevisiae*. Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes residuos de contacto y medios sitios de la secuencia de reconocimiento:

URA1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	T	A	G	A	T	G	A	C
Residuos de contacto	S32	C33	N30/ Q38	R40	K28	Q26	R68	T44	R70

URA2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	A	T	G	C	G	T	C	T	C
Residuos de contacto	N32	C33	E30/ R38	E40/ R28	R42	Q26	Y68/ K24	Q44	R70

5. Especificidad de secuencias de reconocimiento.

Las meganucleasas racionalmente diseñadas explicadas brevemente anteriormente en este ejemplo se clonaron, se expresaron en *E. coli* y se purificaron como en el Ejemplo 1. Cada meganucleasa purificada se mezcló entonces 1:1 con su componente de heterodimerización correspondiente (por ejemplo, ACH1 con ACH2, HGH1 con HGH2, etc.) y se incubaron con un sustrato de ADN linealizado que contenía la secuencia de reconocimiento de ADN no palindrómica prevista para cada heterodímero de meganucleasas. Como se muestra en la Figura 3, cada heterodímero de meganucleasas racionalmente diseñadas escinde su sitio de ADN previsto.

LISTADO DE SECUENCIAS

	<110> Precision Biosciences
5	<120> MEGANUCLEASAS RACIONALMENTE DISEÑADAS CON ESPECIFICIDAD DE SECUENCIA ALTERADA Y AFINIDAD DE UNIÓN A ADN
	<130> 111-046T6
10	<140> 60/727.512 <141> 18-10-2005
	<150> 06 826 293.0 <151> 18-10-2006
15	<160> 37
	<170> Patent In Ver. 3.3
20	<210> 1 <211> 163 <212> PRT <213> Chlamydomonas reinhardtii
25	<400> 1
	Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe 1 5 10 15
	Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln Ser 20 25 30
	Tyr Lys Phe Lys His Gln Leu Ser Leu Ala Phe Gln Val Thr Gln Lys 35 40 45
	Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val 50 55 60
	Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu 65 70 75 80
	Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys 85 90 95
	Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu 100 105 110
	Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp 115 120 125
	Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr 130 135 140
	Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys 145 150 155 160
	Ser Ser Pro
30	<210> 2 <211> 22 <212> ADN <213> Chlamydomonas reinhardtii
35	<400> 2 gaaactgtct cagcagcttt tg 22
	<210> 3 <211> 22 <212> ADN

ES 2 440 801 T3

<213> Chlamydomonas reinhardtii
 <400> 3
 caaacgtcg tgagacagtt tc 22
 5
 <210> 4
 <211> 22
 <212> ADN
 <213> Chlamydomonas reinhardtii
 10
 <400> 4
 caaactgtcg tgagacagtt tg 22
 15
 <210> 5
 <211> 22
 <212> ADN
 <213> Chlamydomonas reinhardtii
 20
 <400> 5
 caaactgtct cagacagtt tg 22
 25
 <210> 6
 <211> 170
 <212> PRT
 <213> Monomastix sp.
 <400> 6
 Met Thr Thr Lys Asn Thr Leu Gln Pro Thr Glu Ala Ala Tyr Ile Ala
 1 5 10 15
 Gly Phe Leu Asp Gly Asp Gly Ser Ile Tyr Ala Lys Leu Ile Pro Arg
 20 25 30
 Pro Asp Tyr Lys Asp Ile Lys Tyr Gln Val Ser Leu Ala Ile Ser Phe
 35 40 45
 Ile Gln Arg Lys Asp Lys Phe Pro Tyr Leu Gln Asp Ile Tyr Asp Gln
 50 55 60
 Leu Gly Lys Arg Gly Asn Leu Arg Lys Asp Arg Gly Asp Gly Ile Ala
 65 70 75 80
 Asp Tyr Thr Ile Ile Gly Ser Thr His Leu Ser Ile Ile Leu Pro Asp
 85 90 95
 Leu Val Pro Tyr Leu Arg Ile Lys Lys Lys Gln Ala Asn Arg Ile Leu
 100 105 110
 His Ile Ile Asn Leu Tyr Pro Gln Ala Gln Lys Asn Pro Ser Lys Phe
 115 120 125
 Leu Asp Leu Val Lys Ile Val Asp Asp Val Gln Asn Leu Asn Lys Arg
 130 135 140
 Ala Asp Glu Leu Lys Ser Thr Asn Tyr Asp Arg Leu Leu Glu Glu Phe
 145 150 155 160
 Leu Lys Ala Gly Lys Ile Glu Ser Ser Pro
 165 170
 30
 <210> 7
 <211> 22
 <212> ADN
 <213> Monomastix sp.
 35
 <400> 7
 cagaacgtcg tgagacagtt cc 22

ES 2 440 801 T3

<210> 8
 <211> 22
 <212> ADN
 <213> Monomastix sp.
 5
 <400> 8
 ggaactgtct cagcagcttc tg 22
 10
 <210> 9
 <211> 235
 <212> PRT
 <213> Saccharomyces cerevisiae
 15
 <400> 9
 Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser
 1 5 10 15
 Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu
 20 25 30
 Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg
 35 40 45
 Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn
 50 55 60
 Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu
 65 70 75 80
 Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val
 85 90 95
 Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu
 100 105 110
 Ala Asn Leu Phe Ile Val Asn Asn Lys Lys Thr Ile Pro Asn Asn Leu
 115 120 125
 Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp
 130 135 140
 Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys Ser Ile
 145 150 155 160
 Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Glu Tyr Leu Val
 165 170 175
 Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys Ile Asn
 180 185 190
 Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe
 195 200 205
 Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr Lys Leu
 210 215 220
 Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
 225 230 235
 20
 <210> 10
 <211> 18
 <212> ADN
 <213> Saccharomyces cerevisiae
 25
 <400> 10
 ttaccctgtt atccctag 18
 <210> 11
 <211> 18

<212> ADN
 <213> *Saccharomyces cerevisiae*

 5 <400> 11
 ctagggataa cagggtaa 18

 <210> 12
 <211> 218
 <212> PRT
 10 <213> *Chlamydomonas eugametos*

 <400> 12

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

 <210> 13
 <211> 22
 <212> ADN
 20 <213> *Chlamydomonas eugametos*

 <400> 13
 ataacgtcc taagtagcg aa 22

 25 <210> 14
 <211> 22
 <212> ADN
 <213> *Chlamydomonas eugametos*

 30 <400> 14
 ttcgtacct taggaccgtt at 22

5
 <210> 15
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

10
 <400> 15
 gaagagctca tcagaacagt ca 22

15
 <210> 16
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

20
 <400> 16
 gaagagctc 9

25
 <210> 17
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

30
 <400> 17
 tgactgttc 9

35
 <210> 18
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

40
 <400> 18
 aaccctctc 9

45
 <210> 19
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

50
 <400> 19
 ctccgggtc 9

55
 <210> 20
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

60
 <400> 20
 tgcggtgtc 9

65
 <210> 21
 <211> 9
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

	<400> 21 caggctgtc	9
5	<210> 22 <211> 22 <212> ADN <213> Bacteriófago lamda-p05	
10	<400> 22 tgcggtgtcc ggcgacagcc tg	22
15	<210> 23 <211> 22 <212> ADN <213> Homo sapiens	
20	<400> 23 ctgggagtct caggacagcc tg	22
25	<210> 24 <211> 22 <212> ADN <213> Homo sapiens	
30	<400> 24 ccaggtgtct ctggactcct cc	22
35	<210> 25 <211> 22 <212> ADN <213> Homo sapiens	
40	<400> 25 gaaaatatca ttggtgttc ct	22
45	<210> 26 <211> 22 <212> ADN <213> Homo sapiens	
50	<400> 26 aaccctctcc agtgagatgc ct	22
55	<210> 27 <211> 22 <212> ADN <213> Homo sapiens	
60	<400> 27 gacctcgtcc tccgactcgc tg	22
65	<210> 28 <211> 22 <212> ADN <213> Virus del herpes simple 2	
	<400> 28 ctcgatgtcg gacgacacgg ca	22
	<210> 29 <211> 22 <212> ADN <213> Bacillus anthracis	
	<400> 29 acaagtgtct atggacagtt ta	22

5	<210> 30 <211> 22 <212> ADN <213> Virus de la viruela	
	<400> 30 aaaactgtca aatgacatcg ca	22
10	<210> 31 <211> 22 <212> ADN <213> Virus de Epstein-Barr	
15	<400> 31 cggggtctcg tgcgaggcct cc	22
20	<210> 32 <211> 22 <212> ADN <213> Arabidopsis thaliana	
	<400> 32 cactaactcg tatgagtcgg tg	22
25	<210> 33 <211> 22 <212> ADN <213> Arabidopsis thaliana	
30	<400> 33 tgcctctct agagacccgg ag	22
35	<210> 34 <211> 22 <212> ADN <213> Nicotiana tabacum	
40	<400> 34 taaaatctct aaggtctgtg ca	22
45	<210> 35 <211> 22 <212> ADN <213> Nicotiana tabacum	
	<400> 35 caagaattca agcgagcatt aa	22
50	<210> 36 <211> 22 <212> ADN <213> Saccharomyces cerevisiae	
55	<400> 36 ttagatgaca agggagacgc at	22
60	<210> 37 <211> 9 <212> PRT <213> Secuencia artificial <220> <223> Descripción de secuencia artificial: motivo de péptido sintético	
65	<400> 37	

ES 2 440 801 T3

Leu Ala Gly Leu Ile Asp Ala Asp Gly
1 5

5

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Un procedimiento de escisión de una secuencia de reconocimiento diana de ADN bicatenario en una célula eucariota que comprende:

introducir en dicha célula eucariota
un ácido nucleico que codifica una meganucleasa o
una proteína meganucleasa,
en el que dicha meganucleasa escinde dicha secuencia de reconocimiento diana de ADN bicatenario, y

en el que dicha meganucleasa es una meganucleasa recombinante que comprende:

un polipéptido que tiene al menos el 85% de similitud de secuencias con los residuos 2-153 de la meganucleasa I-Crel de SEC ID N°: 1,
y que tiene especificidad por un medio sitio de la secuencia de reconocimiento que se diferencia al menos un par de bases de un medio sitio dentro de una secuencia de reconocimiento de la meganucleasa I-Crel seleccionada del grupo que consiste en SEC ID N°: 2, SEC ID N°: 3, SEC ID N°: 4 y SEC ID N°: 5, en el que la especificidad en la posición -4 ha sido alterada:

- (i) a una C en una hebra codificante por una modificación seleccionada del grupo que consiste en I77E y I77S;
- (ii) a una G en una hebra codificante por una modificación seleccionada del grupo que consiste en I77R;
- (iii) a una T en una hebra codificante por una modificación que comprende I77S; o
- (iv) a una A en una hebra codificante por una modificación que comprende I77Q,

con la condición de que no se use embrión humano para fines industriales o comerciales.

2. El procedimiento de la reivindicación 1, en el que la meganucleasa recombinante comprende al menos una segunda sustitución que altera la especificidad de aminoácidos correspondiente a una sustitución seleccionada del grupo que consiste en:

I24C, I24K, I24R, Q26A, K28A, K28C, K28H, K28Q, K28R, K28S, N30E, N30K, N30Q, N30R, S32A, S32C, S32D, S32E, S32H, S32I, S32K, S32L, S32N, S32Q, S32R, S32T, S32V, Y33C, Y33D, Y33E, Y33F, Y33H, Y33I, Y33L, Y33R, Y33V, Q38C, Q38E, Q38H, Q38I, Q38K, Q38L, Q38N, Q38R, S40A, S40C, S40E, S40I, S40Q, S40R, S40V, A42E, A42Q, A42R, Q44A, Q44C, Q44D, Q44E, Q44I, Q44K, Q44L, Q44N, Q44R, Q44T, Q44V, T46A, T46C, T46D, T46E, T46G, T46H, T46K, T46Q, T46R, Y66K, Y66M, R68C, R68E, R68F, R68H, R68K, R68L, R68M, R68Q, R68Y, R70A, R70C, R70D, R70E, R70G, R70H, R70K, R70L, R70Q, R70S, D75C, D75E, D75H, D75L, D75Q, D75R, D75Y, S79A, S79C, S79I, S79V, K139H y K139Y.

3. El procedimiento de la reivindicación 1 o la reivindicación 2, en el que dicha secuencia de reconocimiento diana de ADN bicatenario está comprendida en un cromosoma de dicha célula eucariota.

4. El procedimiento de la reivindicación 3, que comprende además:

transfectar dicha célula eucariota con una secuencia de ácidos nucleicos que incluye una secuencia exógena, en el que dicha secuencia exógena se inserta en el cromosoma en el sitio de escisión, para así producir una célula eucariota genéticamente modificada que incluye dicha secuencia exógena insertada en dicho cromosoma de dicha célula eucariota,

opcionalmente
en el que dicho ácido nucleico comprende además secuencias homólogas a secuencias que flanquean el sitio de escisión y dicha secuencia exógena se inserta en dicho sitio de escisión por recombinación homóloga, o
en el que dicho ácido nucleico carece de homología sustancial con el sitio de escisión y dicha secuencia exógena se inserta en dicho cromosoma por unión de extremos no homólogos.

5. El procedimiento de la reivindicación 3, en el que dicha secuencia diana se altera por unión de extremos no homólogos en el sitio de escisión para así producir una célula eucariota genéticamente modificada que altera la secuencia diana en dicho cromosoma de dicha célula eucariota.

6. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que el nucleótido correspondiente a la posición -4 del sitio de reconocimiento de I-Crel ha sido alterado a una C y la primera sustitución de aminoácidos que altera la especificidad se corresponde con una sustitución I77E.

7. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que el nucleótido correspondiente a la posición -4 del sitio de reconocimiento de I-Crel ha sido alterado a una C y la primera sustitución de aminoácidos que altera la especificidad se corresponde con una sustitución I77S.

8. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que el nucleótido correspondiente a la posición -4 del sitio de reconocimiento de I-CreI ha sido alterado a una G y la primera sustitución de aminoácidos que altera la especificidad se corresponde con una sustitución I77R.

5 9. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que el nucleótido correspondiente a la posición -4 del sitio de reconocimiento de I-CreI ha sido alterado a una T y la primera sustitución de aminoácidos que altera la especificidad se corresponde con una sustitución I77S.

10 10. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que el nucleótido correspondiente a la posición -4 del sitio de reconocimiento de I-CreI ha sido alterado a una A y la primera sustitución de aminoácidos que altera la especificidad se corresponde con una sustitución I77Q.

11. El procedimiento de una cualquiera de las reivindicaciones 1-10, en el que dicha célula es una célula de planta.

15 12. El procedimiento de la reivindicación 11, en el que dicho ácido nucleico se introduce en dicha célula por un vector de *Agrobacterium*, inyección balística, bombardeo con microproyectiles o electroporación.

20 13. El procedimiento de la reivindicación 11, en el que dicha célula es un protoplasto y dicho ácido nucleico se introduce en dicha célula por transformación mediada por PEG.

14. El procedimiento de una cualquiera de las reivindicaciones 1-10, en el que dicha célula es una célula de mamífero.

25 15. El procedimiento de la reivindicación 14, en el que el ácido nucleico se introduce en dicha célula por un adenovirus, un vector de virus adeno-asociado, inyección balística, electroporación, microinyección o transfección de liposomas.

30

35

40

45

50

55

60

65

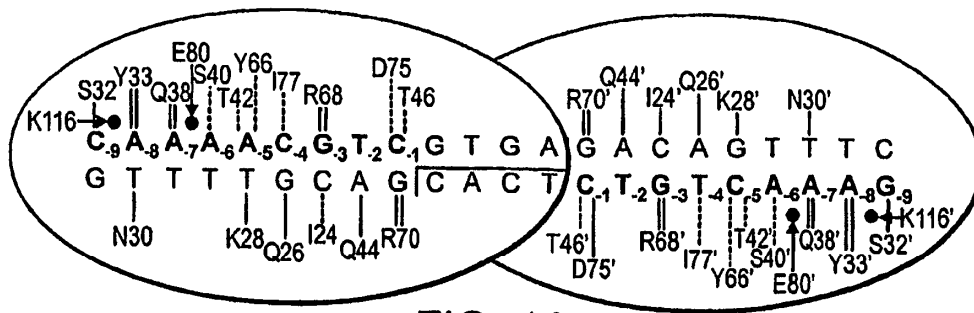


FIG. 1A

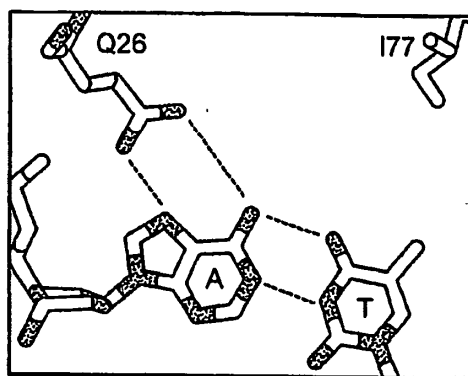


FIG. 1B

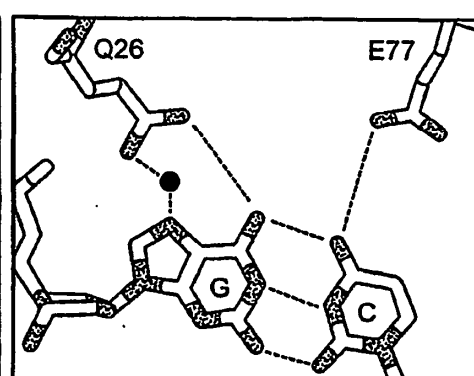


FIG. 1C

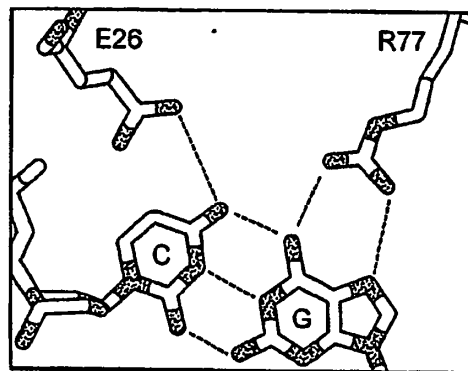


FIG. 1D

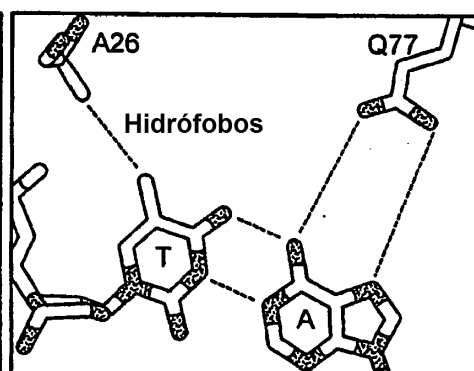


FIG. 1E

WT CAAACTGTCGTGAGACAGTTTG

CF1/CF2 GAAAATATCATTGGTGTTTCCT

MYD1/MYD2 GACCTCGTCCGACTCGCTG

CCR1/CCR2 AACCCCTCTCCAGTGAGATGCCT

ACH1/ACH2 CTGGGAGTCTCAGGACAGCCCTG

TAT1/TAT2 GAGAGCTCATCAGAACAGTCA

HSV1/HSV2 CTCGATGTCTGGACGACACGGCA

LAM1/LAM2 TCGGGTGTCGGCGGACAGCCCTG

POX1/POX2 AAAACTGTCAAATGACATCGCA

URA1/URA2 TTAGATGACAAAGGAGACGCAT

GLA1/GLA2 CACTAACTCGTATGAGTCGGTG

BRP1/BRP2 TGCCTCCCTCTAGAGACCCGGAG

FIG. 2A

Enzima														
Sitio	WT	CF		CCR		TAT		LAM		URA		BRP		
		WT	MYD	ACH	HSV	POX	GLA	WT	MYD	ACH	HSV	POX	GLA	
WT	100	0	0	0	0	0	47	0	100	100	0	0	Natural	
CF1/CF2	0	63	0	0	0	0	0	0	0	0	0	0	Fibrosis quística	
MYD1/MYD2	0	0	94	0	0	0	16	0	0	0	0	0	Distrofia miotónica	
CCR1/CCR2	0	0	0	100	0	0	0	0	0	0	0	0	CCR5	
ACH1/ACH2	0	0	0	0	87	0	0	0	0	0	0	0	Acondroplasia	
TAT1/TAT2	0	0	0	0	0	81	0	0	0	0	0	0	VIH-1	
HSV1/HSV2	0	0	0	0	0	0	67	0	0	0	0	0	Herpes simple 1	
LAM1/LAM2	0	0	0	0	0	0	0	87	0	0	0	0	Fago lambda	
POX1/POX2	100	0	0	0	0	0	0	0	100	0	0	0	Virus de la viruela	
URA1/URA2	0	0	0	0	0	0	0	0	0	70	0	0	<i>S. cerevisiae</i> URA3	
GLA1/GLA2	0	0	0	0	0	0	0	0	0	0	88	0	<i>A. thaliana</i> GL2	
BRP1/BRP2	0	0	0	0	0	0	0	0	0	0	0	100	<i>A. thaliana</i> BP1	

FIG. 2B

FIG. 3A
FIG. 3B

FIG. 3

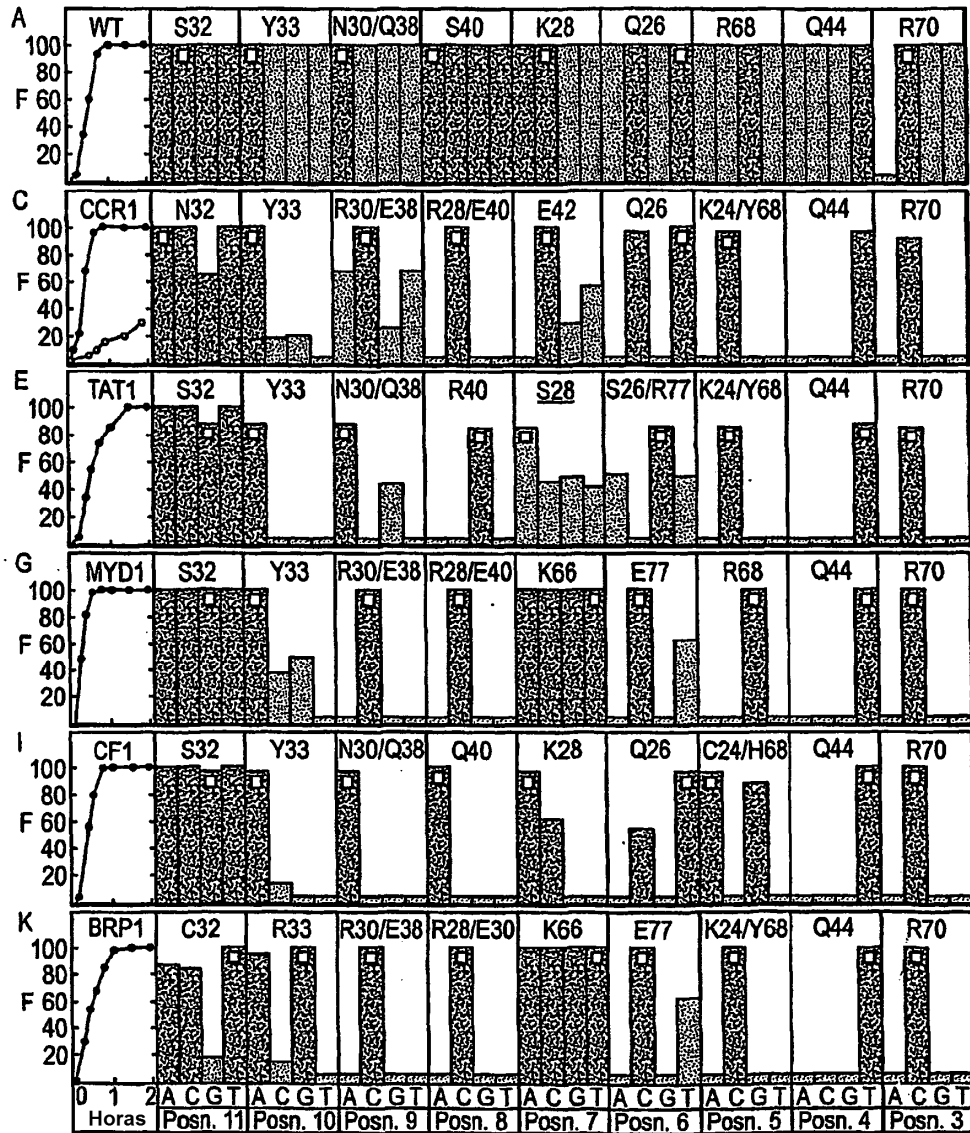


FIG. 3A

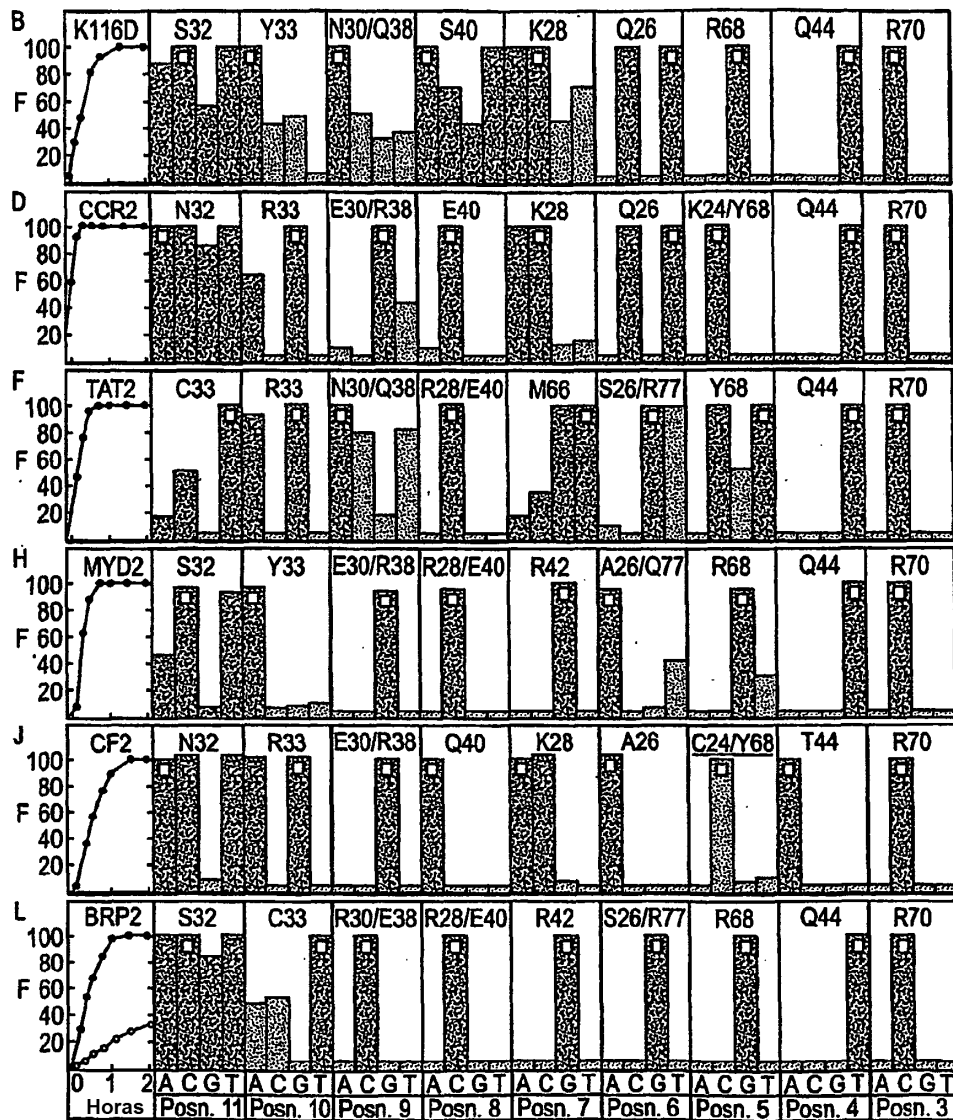


FIG. 3B