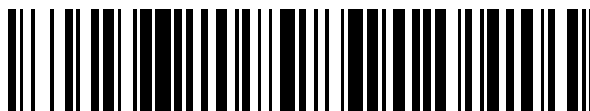


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 582 091**

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 13164628 (3)**

97 Fecha y número de publicación de la concesión europea: **04.05.2016 EP 2628794**

54 Título: **Meganucleasas diseñadas racionalmente 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:

**09.09.2016**

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:

**VALLEJO LÓPEZ, Juan Pedro**

**Observaciones :**

**Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes**

**ES 2 582 091 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).

**DESCRIPCION**

Meganucleasas diseñadas racionalmente con especificidad de secuencia y afinidad de unión a ADN alteradas

5 Esta solicitud reivindica el beneficio de prioridad de la Solicitud Provisional de Patente de los Estados Unidos N° 60/727.512, presentada el 18 de octubre de 2005.

**Apoyo gubernamental**

10 La invención fue apoyada en parte por las subvenciones 2R01-GM-0498712, 5F32-GM072322 y 5 DP1 OD000122 del National Institute of General Medical Sciences de los National Institutes of Health de los Estados Unidos de América. Por lo tanto, el Gobierno de los Estados Unidos tiene determinados derechos sobre la invención.

**Campo de la invención**

15 La invención se refiere al campo de la biología molecular y de tecnología de ácidos nucleicos recombinantes. En particular, la invención se refiere a meganucleasas de origen no natural diseñadas racionalmente con especificidad de secuencia de reconocimiento de ADN alterada y/o afinidad alterada. La invención también se refiere a métodos para producir tales meganucleasas y métodos para producir ácidos nucleicos recombinantes y organismos que usan tales meganucleasas.

**Antecedentes de la invención**

25 La ingeniería genómica requiere la capacidad de insertar, suprimir, 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 modificación del genoma sigue siendo un objetivo importante en la terapia génica, agrotecnología y biología sintética (Porteus *et al* (2005), Nat. Biotechnol. 23:967-73; Tzfira *et al* (2005), Trends Biotechnol. 23:567-9; McDaniel *et al* (2005), Curr. Opin. Biotechnol. 16:476-83). Un método habitual para insertar o modificar una secuencia de ADN implica introducir una secuencia de ADN transgénico flanqueada por secuencias homólogas a la diana genómica y seleccionar o explorar con respecto a un acontecimiento de recombinación homóloga exitoso. La recombinación con el ADN transgénico se produce en pocas ocasiones pero puede estimularse por una rotura de doble cadena en el ADN genómico en el sitio diana. Se han empleado numerosos métodos para crear roturas de doble cadena de ADN, incluyendo irradiación y tratamientos químicos. Aunque estos métodos estimulan eficazmente la recombinación, las roturas de doble cadena se dispersan de forma aleatoria en el genoma, lo que puede ser altamente mutagénico y tóxico. En la actualidad, la incapacidad de dirigir modificaciones génicas a sitios únicos dentro de un fondo cromosómico es un impedimento importante para la ingeniería genómica exitosa.

40 Un enfoque para conseguir este objetivo es estimular la recombinación homóloga en una rotura de doble cadena en un locus diana usando una nucleasa con especificidad para una secuencia que es suficientemente grande para estar presente solamente en un sitio sencillo dentro del genoma (véase, por ejemplo, Porteus *et al* (2005), Nat. Biotechnol. 23:967-73). La eficacia de esta estrategia se ha demostrado en diversos organismos usando fusiones quiméricas entre un dominio de unión de ADN de dedo de cinc obtenido por ingeniería genética y el dominio de nucleasa no específico de la enzima de restricción FokI (Porteus (2006), Mol Ther 13:43846; Wright *et al* (2005), Plant J. 44:693-705; Umov *et al* (2005), Nature 435:646-51). Aunque estas nucleasas de dedos de cinc artificiales estimulan la recombinación específica de sitio, conservan actividad de escisión no específica residual resultante de la baja regulación del dominio nucleasa y escinden frecuentemente en sitios no pretendidos (Smith *et al.* (2000), Nucleic Acids Res. 28:3361-9). Tal escisión no pretendida puede provocar mutaciones y toxicidad en el organismo tratado (Porteus *et al* (2005), Nat. Biotechnol. 23:967-73).

50 Un grupo de nucleasas de origen natural que reconocen sitios de escisión de 15-40 pares de bases habitualmente hallados en los genomas de plantas y hongos puede proporcionar una alternativa de ingeniería genómica menos tóxica. Tales "meganucleasas" o "endonucleasas de búsqueda" se asocian frecuentemente con elementos de ADN parasitarios, tales como intrones de auto-corte y empalme de grupo 1 e inteínas. Promueven de forma natural la recombinación homóloga o inserción génica en localizaciones específicas en el genoma hospedador produciendo una rotura de doble cadena en el cromosoma, que recluta la maquinaria de reparación de ADN celular (Stoddard (2006), Q. Rev. Biophys. 38:49-95). Las meganucleasas se agrupan habitualmente en cuatro familias: la familia LAGLIDADG, la familia GIY-YIG, la familia de caja His-Cys y la familia HNH. Estas familias se caracterizan por motivos estructurales, que afectan a la actividad catalítica y secuencia de reconocimiento. Por ejemplo, los miembros de la familia LAGLIDADG se caracterizan porque tienen una o dos copias del motivo LAGLIDADG conservado (véase Chevalier *et al* (2001), Nucleic Acids Res 29(18): 3757-3774). Las meganucleasas LAGLIDADG con una copia sencilla del motivo LAGLIDADG forman homodímeros, mientras que los miembros con dos copias del motivo LAGLIDADG se encuentran como monómeros. De forma similar, los miembros de la familia GIY-YIG tienen un módulo GIY-YIG, que es de 70-100 restos de longitud e incluye cuatro o cinco motivos de secuencia conservados con cuatro restos invariantes, dos de los cuales se requieren para actividad (véase Van Roey *et al* (2002), Nature Struct. Biol. 9:806-811). Las meganucleasas de caja His-Cys se caracterizan por una serie altamente conservada de histidinas y cisteínas sobre una región que abarca varios cientos de restos de aminoácidos (véase Chevalier *et al* (2001), Nucleic Acids Res 29(18): 3757-3774).

En el caso de la familia NHN, los miembros se definen por motivos que contiene dos pares de histidinas conservadas rodeadas por restos de asparagina (véase Chevalier *et al* (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, en consecuencia, la especificidad de secuencia de reconocimiento de ADN y actividad catalítica.

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

La implementación sistemática de modificación génica estimulada por nucleasas requiere el uso de enzimas modificadas por ingeniería genética con especificidades adaptadas para dirigir roturas de ADN a sitios existentes en un genoma y, por lo tanto, ha habido gran interés en la adaptación de meganucleasas para promover las modificaciones génicas en sitios médica o biotecnológicamente relevantes (Porteus *et al* (2005), *Nat. Biotechnol.* 23:967-73; Sussman *et al* (2004), *J. Mol. Biol.* 342:31-41; Epinat *et al* (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 de cloroplastos y que presenta una diana atractiva para rediseño de meganucleasas. La enzima de tipo silvestre es un homodímero en el que cada monómero entra en contacto directo 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 posición sencilla en esta secuencia de reconocimiento (Sussman *et al* (2004), *J. Mol. Biol.* 342:31-41; Chames *et al.* (2005), *Nucleic Acids Res* 33: e178; Seligman *et al* (2002), *Nucleic Acids Res* 30:3870-9) o, más recientemente, en tres posiciones en la secuencia de reconocimiento (Arnould *et al* (2006), *J. Mol Biol.* 355:443-58). La interfaz de ADN-proteína I-Crel contiene nueve aminoácidos que entran en contacto con las bases de ADN directamente y al menos cinco posiciones adicionales que pueden formar contactos potenciales en interfaces modificadas. El tamaño de esta interfaz impone una complejidad combinatoria de la que es poco probable tomar muestras de forma adecuada en bibliotecas de secuencias construidas para seleccionar con respecto a enzimas con sitios de escisión drásticamente alterados..

Sigue existiendo la necesidad de nucleasas que faciliten la modificación precisa de un genoma. Además, sigue existiendo la necesidad de técnicas para generar nucleasas con secuencias de reconocimiento diseñadas racionalmente predeterminadas que permitan la manipulación de secuencias genéticas en loci genéticos específicos y de técnicas que utilicen tales nucleasas para modificar por ingeniería genética organismos con modificaciones de secuencia precisas.

#### Sumario de la invención

La invención se define por la materia-objeto de las reivindicaciones adjuntas. La invención se refiere a un monómero de meganucleasa recombinante que comprende:

un polipéptido que tiene al menos un 85 % de similitud de secuencia con los restos 2-153 de la meganucleasa I-Crel de SEQ ID NO: 1; en el que la afinidad de unión a ADN se ha aumentado mediante al menos una modificación de un resto que forma parte de la superficie de contacto de la estructura de la meganucleasa I-Crel de SEQ ID NO: 1, correspondiendo dicha modificación a una sustitución seleccionada del grupo que consiste en

E80 con H, N, Q, S, K o R. Lo que se divulga es la identificación y caracterización de restos de aminoácidos específicos en la familia de meganucleasas LAGLIDADG que entran en contacto con bases de ADN y la cadena principal de ADN cuando las meganucleasas se asocian con una secuencia de reconocimiento de ADN bicatenaria, y de este modo afectan a la especificidad y actividad de las enzimas. Este descubrimiento se ha usado, como se describe en detalle posteriormente, para identificar sustituciones de aminoácidos que pueden alterar la especificidad de secuencia de reconocimiento y/o afinidad de unión a ADN de las meganucleasas, y para diseñar y desarrollar racionalmente meganucleasas que puedan reconocer una secuencia de ADN deseada que las meganucleasas de origen natural no reconocen. La invención también proporciona métodos que usan tales meganucleasas para producir ácidos nucleicos recombinantes y organismos utilizando las meganucleasas para provocar recombinación de una secuencia genética deseada en un número limitado de loci dentro del genoma del organismo, para terapia génica, para tratamiento de infecciones patógenas y para aplicaciones *in vitro* en diagnóstico e investigación.

Por lo tanto, se divulgan meganucleasas recombinantes que tienen afinidad de unión alterada por el ADN bicatenario en relación a una meganucleasa I-Crel de tipo silvestre, en la que la meganucleasa incluye un polipéptido que tiene al menos 85 % de similitud de secuencia con los restos 2-153 de la meganucleasa I-Crel de SEQ ID NO: 1 y en la que la afinidad de unión a ADN se ha (1) aumentado por al menos una modificación correspondiente a una sustitución

seleccionada entre (a) sustitución de E80, D137, 181, L112, P29, V64 o Y66 con H, N, Q, S, T, K o R o (b) sustitución de T46, T140 o T143 con K o R; o, por el contrario, (2) reducido mediante al menos una modificación correspondiente a una sustitución seleccionada entre (a) sustitución de K34, K48, R51, K82, K116 o K139 con H, N, Q, S, T, D o E o (b) sustitución de I81, L112, P29, V64, Y66, T46, T140 o T143 con D o E.

Las meganucleasas divulgadas pueden incluir una, dos tres o más de las modificaciones de restos de contacto con la estructura que se han divulgado en el presente documento para afectar a la afinidad de unión a ADN. Además, estas modificaciones que afectan a la afinidad de unión a ADN pueden combinarse con una o más de las nuevas modificaciones de los restos en contacto con la base descritos anteriormente que alteran la especificidad de secuencia de las meganucleasas recombinantes en posiciones específicas dentro de la secuencia de reconocimiento o con las modificaciones de la técnica anterior descritas anteriormente o con una combinación de las nuevas modificaciones y las modificaciones de la técnica anterior. En particular, al combinar modificaciones de contacto con la estructura y modificaciones de contacto con la base, las meganucleasas recombinantes pueden diseñarse racionalmente con una especificidad y actividad deseadas. Por ejemplo, pueden diseñarse aumentos en la afinidad de unión a ADN que pueden compensar las pérdidas de afinidad resultantes de los cambios diseñados para los restos de contacto con la base o pueden diseñarse reducciones en la afinidad que también pueden reducir la especificidad de secuencias y ampliar el conjunto de secuencias de reconocimiento para una enzima.

Se divulgan monómeros de meganucleasa recombinante o heterodímeros con afinidad alterada para formación de dímeros también pueden incluir una, dos, tres o más de las modificaciones de restos de contacto de base; una, dos, tres o más de las modificaciones de restos de contacto de cadena principal descritos anteriormente; o combinaciones de ambos. Por lo tanto, por ejemplo, los contactos de base de un monómero pueden modificarse para alterar la especificidad de secuencia, los contactos de cadena principal de un monómero pueden modificarse para alterar la afinidad de unión a ADN y el interfaz proteína-proteína puede modificarse para afectar a la formación de dímeros. Un monómero recombinante tal puede combinarse con un monómero modificado de forma similar para producir un heterodímero de meganucleasa diseñado racionalmente con actividad y especificidad de secuencia deseadas.

En otro aspecto, la invención proporciona diversos métodos de uso para las meganucleasas diseñadas racionalmente descritas y posibilitadas en el presente documento. Estos métodos incluyen producir células madre y organismos modificados genéticamente, tratar enfermedades por terapia génica, tratar infecciones por patógenos y usar las meganucleasas recombinantes para aplicaciones *in vitro* para diagnóstico e investigación.

Por lo tanto, en un aspecto, la invención proporciona métodos para producir una célula eucariota modificada genéticamente que incluye una secuencia exógena de interés insertada en un cromosoma, transfectando la célula con (i) una primera secuencia de ácido nucleico que codifica una meganucleasa de la invención y (ii) una segunda secuencia de ácido nucleico que incluye dicha secuencia de interés, en la 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 por recombinación homóloga o por unión de extremos no homólogos, en la que dicha célula eucariota no es una célula madre humana.

Como alternativa, en otro aspecto, la invención proporciona métodos para producir una célula eucariota modificada genéticamente 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 del sitio de escisión por recombinación homóloga o por unión de extremos no homóloga, en la que dicha célula eucariota no es una célula madre humana.

En otro aspecto, la invención proporciona métodos para producir una célula eucariota modificada genéticamente rompiendo una secuencia diana en un cromosoma, transfectando la célula con un ácido nucleico que codifica una meganucleasa de la invención, en los que la meganucleasa produce un sitio de escisión en el cromosoma y la secuencia diana se rompe por unión de extremos no homólogos en el sitio de escisión, en la que dicha célula eucariota no es una célula madre humana.

En otro aspecto, la invención proporciona métodos para producir un organismo no humano modificado genéticamente produciendo una célula eucariota modificada genéticamente de acuerdo con los métodos descritos anteriormente, y cultivando la célula eucariota modificada genéticamente para producir el organismo modificado genéticamente, en la que dicha célula eucariota no es una célula madre humana. En estas realizaciones, la célula eucariota puede seleccionarse de un gameto, un cigoto, un blastocisto, una célula madre embrionaria y un protoplasto.

En otro aspecto, la invención proporciona métodos para tratar una enfermedad por terapia génica en un eucariota, transfectando al menos una célula del eucariota con uno o más ácidos nucleicos que incluyen (i) una primera secuencia de ácido nucleico que codifica una meganucleasa divulgada en el presente documento y (ii) una segunda secuencia de ácido nucleico que incluye una 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 por recombinación homóloga o unión de extremos no homólogos, y la inserción de la secuencia de interés proporciona terapia génica para la enfermedad.

Como alternativa, en otro aspecto, se describen métodos para tratar una enfermedad por terapia génica en un eucariota, introduciendo una proteína meganucleasa divulgada en el presente documento en al menos una célula del eucariota, y transfectando la célula con un ácido nucleico que incluye una 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 por recombinación homóloga o unión de extremos no homólogos, y la inserción de la secuencia de interés proporciona terapia génica para la enfermedad.

En otro aspecto, se describen métodos para tratar una enfermedad por terapia génica en un eucariota rompiendo una secuencia diana en un cromosoma del eucariota, transfectando al menos una célula del eucariota con un ácido nucleico que codifica una meganucleasa divulgada en el presente documento, en los que la meganucleasa produce un sitio de escisión en el cromosoma y la secuencia diana se rompe por unión de extremos no homólogos en el sitio de escisión, en los que la rotura de la secuencia diana proporciona la terapia génica para la enfermedad.

En otro aspecto, se describen métodos para tratar una infección por patógenos virales o procariotas en un hospedador eucariota rompiendo una secuencia diana en un genoma del patógeno, transfectando al menos una célula infectada del hospedador con un ácido nucleico que codifica una meganucleasa divulgada en el presente documento, en los que la meganucleasa produce un sitio de escisión en el genoma y la secuencia diana se rompe por (1) unión de extremos no homólogos en el sitio de escisión o (2) por recombinación homóloga con un segundo ácido nucleico, y en los que la rotura de la secuencia diana proporciona tratamiento para la infección.

Estos y otros aspectos de la invención resultarán evidentes para un experto 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 de I-Crel y su secuencia de reconocimiento bicatenaria de origen natural, basándose en datos cristalográficos. Esta representación esquemática representa la secuencia de reconocimiento (SEQ ID NO: 2 y SEQ ID NO: 3), mostrada desplegada solamente para fines de ilustración, unida por el homodímero, mostrada como dos óvalos. Las bases para cada semi-sitio de ADN se numeran de -1 a -9, y los restos de aminoácidos de I-Crel que forman la superficie de reconocimiento se indican por designaciones de aminoácidos de una letra y números que indican la posición del resto. Líneas negras sólidas: enlaces de hidrógeno con bases de ADN. Líneas discontinuas: posiciones de aminoácidos que forman contactos adicionales en diseños de enzimas pero que no entran en contacto con el ADN en el complejo de tipo silvestre. Flechas: restos que interactúan con la cadena principal de ADN e influyen en la actividad de escisión.

La Figura 1 (B) ilustra los contactos de tipo silvestre entre el par de bases A-T en la posición -4 del semi-sitio de escisión en el lado derecho de la Figura 1 (A). Específicamente, se muestra que el resto Q26 interactúa con la base A. El resto 177 está próximo al par de bases pero no interactúa de forma específica.

La Figura 1 (C) ilustra las interacciones entre una variante diseñada racionalmente de la meganucleasa I-Crel en la que el resto 177 se ha modificado a E77. Como resultado de este cambio, se prefieren 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 semi-sitio de escisión en el lado izquierdo de la Figura 1 (A).

La Figura 1 (D) ilustra las interacciones entre una variante diseñada racionalmente de la meganucleasa I-Crel en la que el resto Q26 se ha modificado a E26 y el resto 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 diseñada racionalmente de la meganucleasa I-Crel en la que el resto Q26 se ha modificado a A26 y el resto 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 meganucleasa I-Crel de tipo silvestre (WT) y 11 heterodímeros de meganucleasa diseñados racionalmente. Las bases que se conservan en relación con la secuencia de reconocimiento WT están sombreadas. Los semi-sitios de 9 pb están en negrita. WT: tipo silvestre (SEQ ID NO: 4); CF: alelo AF508 del gen CFTR humano responsable de la mayoría de los casos de fibrosis quística (SEQ ID NO: 25); MYD: el gen de DM quinasa humano asociado con distrofia miotónica (SEQ ID NO: 27); CCR: el gen CCR5 humano (un co-receptor de VIH principal) (SEQ ID NO: 26); ACH: el gen FGFR3 humano correlacionado con la acondroplasia (SEQ ID NO: 23); TAT: el gen  $\tau$ AT/REV de VIH-1 (SEQ ID NO: 15); VHS: el gen UL36 de VHS-1 (SEQ ID NO: 28); LAM: el gen p05 de bacteriófago X (SEQ ID NO: 22); POX: el gen gp009 de virus Variola (viruela) (SEQ ID NO: 30); URA: el gen *URA3* de *Saccharomyces cerevisiae* (SEQ ID NO: 36); GLA: el gen GL2 de *Arabidopsis thaliana* (SEQ ID NO: 32); BRP: el gen *BP-1* de *Arabidopsis thaliana* (SEQ ID NO: 33).

La Figura 2 (B) ilustra los resultados de incubación de cada I-Crel de tipo silvestre (WT) y 11 heterodímeros de meganucleasa diseñados racionalmente con plásmidos que albergan los sitios de reconocimiento para las 12 enzimas durante 6 horas a 37 °C. El porcentaje de escisión se indica en cada caja.

La Figura 3 ilustra patrones de escisión de homodímeros de I-Crel diseñados racionalmente y de tipo silvestre. (A) I-Crel de tipo silvestre. (B) I-Crel K116D. (C-L) meganucleasas diseñadas racionalmente. Las enzimas se incubaron con un conjunto de plásmidos que albergaban palíndromos del semi-sitio de escisión pretendido, las 27 variaciones de pares de bases sencillos. Los gráficos de barras muestran escisión fraccionada (F) en 4 horas a 37 °C. Barras negras: patrones de escisión esperados basándose en la Tabla 1. Barras grises: sitios de ADN que se

desvían de los patrones de escisión esperados. Los círculos blancos indican bases en el sitio de reconocimiento pretendido. También se muestran ciclos temporales de escisión durante dos horas. Las representaciones de ciclo temporal de círculo abierto en C y L corresponden a escisión por las enzimas CCR1 y BpR2 que carecen de la mutación E80Q. Los sitios de escisión corresponden a los semi-sitios 5' (columna izquierda) y 3' (columna derecha) para las enzimas heterodiméricas descritas en la Figura 2 (A).

## Descripción detallada de la invención

### 1.1 Introducción

La invención se define por las reivindicaciones adjuntas. La divulgación se basa, en parte, en la identificación y caracterización de aminoácidos específicos en la familia LAGLIDADG de meganucleasas que realizan contactos específicos con bases de ADN y contactos no específicos con la cadena principal de ADN cuando las meganucleasas se asocian con una secuencia de reconocimiento de ADN bicatenario, y que afectan por lo tanto a la especificidad de secuencia de reconocimiento y afinidad de unión a ADN de las enzimas. Este descubrimiento se ha usado, como se describe en detalle posteriormente, para identificar sustituciones de aminoácidos en las meganucleasas que pueden alterar la especificidad y/o afinidad de las enzimas, y para diseñar y desarrollar racionalmente meganucleasas que puedan reconocer una secuencia de ADN deseada que las meganucleasas de origen natural no reconocen y/o que tienen especificidad y/o afinidad aumentada o reducida en relación con las meganucleasas de origen natural. Además, debido a que la afinidad de unión a ADN afecta a la actividad enzimática así como a la especificidad de secuencia, se divulgan meganucleasas diseñadas racionalmente con actividad alterada en relación con meganucleasas de origen natural. Además, la invención proporciona meganucleasas diseñadas racionalmente en las que los restos en el interfaz entre los monómeros asociados para formar un dímero se han modificado para promover la formación de heterodímeros. Finalmente, se divulgan usos para las meganucleasas diseñadas racionalmente en la producción de células recombinantes y organismos, así como en terapia génica, antipatogénica, anti-neoplásica y aplicaciones *in vitro*, como se describe en el presente documento.

En general, se describen métodos para generar meganucleasas LAGLIDADG diseñadas racionalmente que contienen restos 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 con la cadena principal de fosfodiéster de una molécula de ADN bicatenaria. Debido a que la actividad enzimática se correlaciona con la afinidad de unión a ADN, sin embargo, la alteración de los aminoácidos implicados en la unión con la secuencia de reconocimiento de ADN puede alterar no solamente la especificidad de la meganucleasa a través de interacciones de pares de bases específicos, sino también la actividad de la meganucleasa aumentando o reduciendo la afinidad de unión global para el ADN bicatenario. De forma similar, la alteración de los aminoácidos implicados en la unión con la cadena principal de ADN puede alterar no solamente la actividad de la enzima, sino también el grado de especificidad o degeneración de unión con la secuencia de reconocimiento aumentando o reduciendo la afinidad de unión global para el ADN bicatenario.

Como se describe en detalle posteriormente, los métodos para diseñar racionalmente meganucleasas incluyen la identificación de los aminoácidos responsables de reconocimiento/unión de ADN, y la aplicación de una serie de reglas para seleccionar cambios de aminoácidos apropiados. Con respecto a especificidad de secuencia de meganucleasa, las reglas incluyen tanto consideraciones estéricas en relación con las distancias en un complejo de meganucleasa-ADN entre las cadenas laterales de aminoácidos de la meganucleasa y las bases en las cadenas sentido y antisentido del ADN como consideraciones en relación con 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, una mayoría de meganucleasas naturales que se unen a ADN como homodímeros reconocen secuencias de reconocimiento pseudo o completamente palindrómicas. Debido a que se espera que los palíndromos largos sean poco comunes, la probabilidad de encontrar una secuencia palindrómica en un sitio genómico de interés es extremadamente baja. En consecuencia, si estas enzimas deben rediseñarse para reconocer sitios genómicos de interés, es necesario diseñar dos monómeros enzimáticos que reconozcan diferentes semi-sitios que puedan heterodimerizar para escindir la secuencia de reconocimiento de híbridos no palindrómica. Por lo tanto, en algunos aspectos, la invención proporciona meganucleasas diseñadas racionalmente en las que los monómeros que difieren en al menos una posición de aminoácido dimerizan para formar heterodímeros. En algunos casos, ambos monómeros se diseñan racionalmente para formar un heterodímero que reconozca una secuencia de reconocimiento no palindrómica. Una mezcla de dos monómeros diferentes puede dar como resultado hasta tres formas activas de dímero de meganucleasa: los dos homodímeros y el heterodímero. Además, o como alternativa, en algunos casos, los restos de aminoácidos se alteran en las interfaces en las que los monómeros pueden interaccionar para formar dímeros, para aumentar o reducir la probabilidad de formación de homodímeros o heterodímeros.

Por lo tanto, en un aspecto, se describen métodos para diseñar de forma racional meganucleasas LAGLIDADG que contienen cambios de aminoácidos que alteran la especificidad y/o actividad de las enzimas. En otro aspecto, se describen las meganucleasas diseñadas racionalmente resultantes de estos métodos. En otro aspecto, se divulgan métodos que usan tales meganucleasas diseñadas racionalmente para producir ácidos nucleicos recombinantes y organismos en los que una secuencia de ADN deseada o locus genético dentro del genoma de un organismo se

modifica por la inserción, delección, sustitución u otra manipulación de secuencias de ADN. En otro aspecto, se describen métodos para reducir la supervivencia de patógenos o células cancerosas usando meganucleasas diseñadas racionalmente que tiene secuencias de reconocimiento específicas de patógeno o específicas de cáncer.

5 1.2 Referencias y Definiciones

La bibliografía científica y de patentes indicada en el presente documento establece el conocimiento que está disponible para los expertos en la materia.

10 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 de 12 pares de bases. Las meganucleasas de origen natural pueden ser monoméricas (por ejemplo, I-Crel) o diméricas (por ejemplo, I-Crel). 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 un meganucleasa dimérica. La expresión “endonucleasa de búsqueda” es sinónima del término “meganucleasa”.

15 Como se usa en el presente documento, la expresión “meganucleasa LAGLIDADG” se refiere a meganucleasas que incluyen un motivo LAGLIDADG sencillo, que son diméricas de forma natural, o a meganucleasas que incluyen dos motivos LAGLIDADG, que son monoméricas de forma natural. La expresión “meganucleasa mono-LAGLIDADG” se usa en el presente documento para referirse a meganucleasas que incluyen un motivo LAGLIDADG sencillo, y la expresión “meganucleasa di-LAGLIDADG” se usa en el presente documento para referirse a meganucleasas que incluyen dos motivos LAGLIDADG, cuando es necesario distinguir entre los dos. Cada uno de los dos dominios estructurales de una meganucleasa di-LAGLIDADG que incluye un motivo LAGLIDADG puede denominarse una subunidad LAGLIDADG.

20 Como se usa en el presente documento, la expresión “diseñado racionalmente” significa de origen no natural y/o obtenido por ingeniería genética. Las meganucleasas diseñadas racionalmente de la invención difieren de las meganucleasas de origen natural o de tipo silvestre en su secuencia de aminoácidos o estructura primaria y también pueden diferir en su estructura secundaria, terciaria o cuaternaria. Además, las meganucleasas diseñadas racionalmente divulgadas en el presente documento también difieren de meganucleasas de tipo silvestre o de origen natural en especificidad de secuencia de reconocimiento y/o actividad.

25 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 ácido nucleico alterada como resultado de la aplicación de técnicas de ingeniería genética. Las técnicas de ingeniería genética incluyen, pero sin limitación, técnicas de clonación de ADN y PCR; transfección, transformación y otras tecnologías de transferencia génica; recombinación homóloga; mutagénesis dirigida; y fusión génica. De acuerdo con esta definición, una proteína que tiene una secuencia de aminoácidos idéntica a una proteína de origen natural, pero que se produce por clonación y expresión en un hospedador heterólogo, no se considera recombinante.

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

35 Como se usa en el presente documento, la expresión “modificado genéticamente” se refiere a una célula u organismo en el que, o en un ancestro del que, una secuencia de ADN genómico se ha modificado deliberadamente por tecnología recombinante. Como se usa en el presente documento, la expresión “modificado genéticamente” abarca el término “transgénico”.

40 Como se usa en el presente documento, la expresión “tipo silvestre” se refiere a cualquier forma de origen natural de una meganucleasa. La expresión “tipo silvestre” no pretende significar la variante alélica más común de la enzima en la naturaleza sino, en su lugar, cualquier variante alélica encontrada en la naturaleza. Las meganucleasas de tipo silvestre se distinguen de meganucleasas recombinantes o de origen no natural.

45 Como se usa en el presente documento, la expresión “semi-sitio de secuencia de reconocimiento” o simplemente “semi-sitio” significa una secuencia de ácido nucleico en una molécula de ADN bicatenario que se reconoce por un monómero de una meganucleasa mono-LAGLIDADG o por una subunidad de LAGLIDADG de una meganucleasa di-LAGLIDADG.

50 Como se usa en el presente documento, la expresión “secuencia de reconocimiento” se refiere a un par de semi-sitios que se unen y se escinden por un dímero de meganucleasa mono-LAGLIDADG o un monómero de meganucleasa di-LAGLIDADG. Los dos semi-sitios pueden o no estar separados por pares de bases que no se reconocen específicamente por la enzima. En los casos de I-Crel, I-Msol e I-Ceul, el semi-sitio de secuencia de reconocimiento de cada monómero abarca 9 pares de bases, y los dos semi-sitios se separan por cuatro pares de bases que no se

reconocen específicamente pero que constituyen el sitio de escisión real (que tiene un saliente de 4 pares de bases). Por lo tanto, las secuencias de reconocimiento combinadas de los dímeros de meganucleasa I-Crel, I-Msol e I-Ceul normalmente abarcan 22 pares de bases, incluyendo dos semi-sitios de 9 pares de bases que flanquean un sitio de escisión de 4 pares de bases. Los pares de bases de cada semi-sitio se designan de -9 a -1, siendo la posición -9 la más distal del sitio de escisión y siendo la posición -1 adyacente a los 4 pares de bases centrales, que se designan N1-N4. La hebra de cada semi-sitio que se orienta de 5' a 3' en la dirección de -9 a -1 (es decir, hacia el sitio de escisión), se designa la hebra "sentido" y la hebra opuesta se designa la "hebra antisentido", aunque ninguna de las hebras puede codificar proteína. Por lo tanto, la hebra "sentido" de un semi-sitio es la hebra antisentido del otro semi-sitio. Véase, por ejemplo, la Figura 1 (A). En el caso de la meganucleasa I-Scel, 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 se reconozcan específicamente. Por convención, una de las dos hebras se denomina la hebra "sentido" y la otra la hebra "antisentido", aunque ninguna de las hebras 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 bicatenarias solamente en una secuencia particular de pares de bases denominada secuencia de reconocimiento, o solamente en un conjunto particular de secuencias de reconocimiento. El conjunto de secuencias de reconocimiento compartirá ciertas posiciones conservadas o motivos de secuencia, pero pueden degenerarse en una o más posiciones. Una meganucleasa altamente específica es capaz de escindir solamente 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 escindida por una meganucleasa de referencia (por ejemplo, de tipo silvestre) o si la tasa de escisión de una secuencia de reconocimiento aumenta o se reduce por una cantidad estadísticamente significativa ( $p < 0,05$ ) en relación con una meganucleasa de referencia.

Como se usa en el presente documento, el término "degeneración" significa lo opuesto de "especificidad". Una meganucleasa altamente degenerada es capaz de escindir un gran número de secuencias de reconocimiento divergentes. Una meganucleasa puede tener degeneración de secuencia en una posición sencilla dentro de un semi-sitio o en múltiples, incluso todas las, posiciones dentro de un semi-sitio. Dicha degeneración de secuencia puede resultar de (i) la incapacidad de cualquier aminoácido en el dominio de unión a ADN de una meganucleasa para realizar 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 realizar un contacto específico con más de una base en una o más posiciones en la secuencia de reconocimiento y/o (iii) suficiente afinidad de unión a ADN no específico para actividad. Una posición degenerada "completamente" puede ocuparse por cualquiera de las cuatro bases y puede designarse con una "N" en un semi-sitio. Una posición degenerada "parcialmente" puede ocuparse por dos o tres de las cuatro bases (por ejemplo, purina (Pu), pirimidina (Py) o no G).

Como se usa en el presente documento con respecto a meganucleasas, la expresión "afinidad de unión a ADN" o "afinidad de unión" significa la tendencia de una meganucleasa a asociarse de forma no covalente 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 de 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 se reduce en una cantidad estadísticamente significativa ( $p < 0,05$ ) en relación con una meganucleasa de referencia.

Como se usa en el presente documento con respecto a monómeros de meganucleasa, la expresión "afinidad para formación de dímeros" significa la tendencia de un monómero de meganucleasa para asociarse de forma no covalente con un monómero de meganucleasa de referencia. La afinidad para 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 de tipo silvestre 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" para formación de dímeros si la  $K_D$  del monómero de meganucleasa recombinante para un monómero de meganucleasa de referencia aumenta o se reduce en una cantidad estadísticamente significativa ( $p < 0,05$ ) en relación con un monómero de meganucleasa de referencia.

Como se usa en el presente documento, el término "palindrómico" se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de semi-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 entran en contacto con la enzima. En el caso de meganucleasas diméricas, las secuencias de ADN palindrómicas se reconocen por homodímeros en los que los dos monómeros realizan contactos con semi-sitios idénticos.

Como se usa en el presente documento, la expresión "pseudo-palindrómico" se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de semi-sitios palindrómicos no idénticos o imperfectos. En este caso, la secuencia pseudo-palindrómica no solamente no necesita ser palindrómica con respecto a los cuatro pares de base centrales, sino que también puede desviarse de una secuencia palindrómica entre los dos semi-sitios.



Las secuencias de ADN pseudo-palindrómicas son típicas de los sitios de ADN naturales reconocidos por meganucleasas homodiméricas de tipo silvestre en las que dos monómeros enzimáticos idénticos realizan contactos con semi-sitios diferentes.

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

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

Como se usa en el presente documento, la expresión “recombinación homóloga” se refiere al proceso natural, celular  
20 en el que una rotura de ADN de doble cadena se repara usando una secuencia de ADN homóloga como el molde de reparación (véase, por ejemplo, Cahill *et al* (2006), *Front. Biosci.* 11:1958-1976). La secuencia de ADN homóloga puede ser una secuencia cromosómica endógena o un ácido nucleico exógeno que se suministró a la célula. Por lo tanto, en algunas realizaciones, se usa una meganucleasa diseñada racionalmente para escindir una secuencia de reconocimiento dentro de una secuencia diana y se suministra a la célula un ácido nucleico exógeno con homología o  
25 similitud de secuencia sustancial con la secuencia diana y se usa como un molde para reparación por recombinación homóloga. La secuencia de ADN del ácido nucleico exógeno, que puede diferir significativamente de la secuencia diana, se incorpora de este modo a la secuencia cromosómica. El proceso de recombinación homóloga se produce principalmente en organismos eucariotas. El término “homología” se usa en el presente documento como equivalente a “similitud de secuencia” y no se pretende que requiera identidad por descendencia o relación filogenética.

Como se usa en el presente documento, la expresión “unión de extremos no homólogos” se refiere al proceso natural,  
30 celular en el que una rotura de ADN de doble cadena se repara por la unión directa de dos segmentos de ADN no homólogos (véase, por ejemplo, Cahill *et al* (2006), *Front. Biosci.* 11:1958-1976). La reparación de ADN por unión de extremos no homólogos es propensa a errores y da como resultado frecuentemente la adición o delección sin molde de secuencias de ADN en el sitio de reparación. Por lo tanto, en ciertas realizaciones, una meganucleasa diseñada  
35 racionalmente puede usarse para producir una rotura de doble cadena en una secuencia de reconocimiento de meganucleasa dentro de una secuencia diana para romper un gen (por ejemplo, introduciendo inserciones de bases, delecciones de bases o mutaciones de desplazamiento de fase) por unión de extremos no homólogos. En otras realizaciones, un ácido nucleico exógeno sin homología o similitud de secuencia sustancial con la secuencia diana puede capturarse en el sitio de una rotura de ADN de doble cadena estimulada por meganucleasa por unión de  
40 extremos no homólogos (véase, por ejemplo, Salomon, *et al* (1998), *EMBO J.* 17:60866095). El proceso de unión de extremos no homólogos se produce tanto en eucariotas como en procariotas tales como bacterias.

Como se usa en el presente documento, la expresión “secuencia de interés” significa cualquier secuencia de ácido  
45 nucleico, tanto si codifica una proteína, ARN o elemento regulador (por ejemplo, una secuencia potenciadora, silenciadora o promotora), que puede insertarse en un genoma o usarse para reemplazar una secuencia de ADN genómico usando una proteína meganucleasa. Las secuencias de interés pueden tener secuencias de ADN heterólogas que posibilitan el marcaje de una proteína o ARN que se expresa a partir de la secuencia de interés. Por ejemplo, una proteína puede marcarse con marcadores incluyendo, pero sin limitación, un epítipo (por ejemplo, c-  
50 myc, FLAG) u otro ligando (por ejemplo, poli-His). Además, una secuencia de interés puede codificar una proteína de fusión, de acuerdo con técnicas conocidas en la materia (véase, por ejemplo, Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley 1999). En algunas realizaciones, la secuencia de interés está flanqueada por una secuencia de ADN que se reconoce por la meganucleasa recombinante para escisión. Por lo tanto, las secuencias flanqueantes se escinden posibilitando la inserción apropiada de la secuencia de interés en secuencias de reconocimiento  
55 genómicas escindidas por la meganucleasa recombinante. En algunas realizaciones, la secuencia completa de interés es homóloga o tiene similitud de secuencia sustancial con la secuencia diana en el genoma de modo que la recombinación homóloga reemplaza eficazmente la secuencia diana con la secuencia de interés. En otras realizaciones, la secuencia de interés se flanquea por secuencias de ADN con homología o similitud de secuencia sustancial con la secuencia diana de modo que la recombinación homóloga inserta la secuencia de interés dentro del  
60 genoma en el locus de la secuencia diana. En algunas realizaciones, la secuencia de interés es sustancialmente idéntica a la secuencia diana excepto para mutaciones u otras modificaciones en la secuencia de reconocimiento de meganucleasa de modo que la meganucleasa no pueda escindir la secuencia diana después de que se haya modificado por la secuencia de interés.

65 Como se usa en el presente documento con respecto tanto a secuencias de aminoácidos como a secuencias de ácido nucleico, las expresiones “porcentaje de similitud” y “similitud de secuencia” se refieren a una medida del grado de

similitud de dos secuencias basándose en un alineamiento de las secuencias que maximiza la similitud entre los restos de aminoácidos o nucleótidos alineados, y que está en función del número de restos o nucleótidos idénticos o similares, el número de restos o nucleótidos totales, y la presencia y longitud de huecos en el alineamiento de secuencia. Están disponibles diversos algoritmos y programas informáticos para determinar la similitud de secuencia usando parámetros convencionales. Como se usa en el presente documento, la similitud de secuencia se mide usando el programa BLASTp para secuencias de aminoácidos y el programa BLASTn para secuencias de ácidos nucleicos, ambos de los cuales están disponibles a través del Centro Nacional para la Información Biotecnológica ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) y se describen en, por ejemplo, Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-410; Gish y States (1993), *Nature Genet.* 3:266-272; Madden *et al.* (1996), *Meth. Enzymol.* 266:131-141; Altschul *et al.* (1997), *Nucleic Acids Res.* 25:33 89-3402); Zhang *et al.* (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 la palabra = 3; penalización de apertura de hueco = -11; penalización de extensión de hueco = -1; y matriz de puntuación = BLOSUM62. Como se usa en el presente documento, el porcentaje de similitud de dos secuencias de ácido nucleico es la puntuación basada en los siguientes parámetros para el algoritmo BLASTn: tamaño de palabra = 11; penalización de apertura de hueco = -5; penalización de extensión de hueco = -2; recompensa de coincidencia = 1; y penalización de emparejamiento erróneo = -3.

Como se usa en el presente documento con respecto a modificaciones de dos proteínas o secuencias de aminoácidos, la expresión "que corresponde a" se usa para indicar que una modificación especificada en la primera proteína es una sustitución del mismo resto de aminoácido que en la modificación en la segunda proteína, y que la posición de aminoácido de la modificación en las primeras proteínas corresponde a o se alinea con la posición de aminoácido de la modificación en la segunda proteína cuando las dos proteínas se someten a alineamientos de secuencia convencionales (por ejemplo, usando el programa BLASTp). Por lo tanto, la modificación del resto "X" al aminoácido "A" en la primera proteína se corresponderá con la modificación del resto "Y" a aminoácido "A" en la segunda proteína si los restos X e Y se corresponden entre sí en un alineamiento de secuencia, y a pesar del hecho de que X e Y pueden ser números diferentes.

Como se usa en el presente documento, la enumeración de un intervalo numérico para una variable pretende transmitir que la invención puede practicarse con la variable igual a cualquiera de los valores dentro de ese intervalo. Por lo tanto, para una variable que es inherentemente discreta, la variable puede ser igual a cualquier valor de número entero dentro del intervalo numérico, incluyendo los puntos finales del intervalo. De forma similar, para una variable que es inherentemente continua, la variable puede ser igual a cualquier valor real dentro del intervalo numérico, incluyendo los puntos finales del intervalo. Como ejemplo, y sin limitación, una variable que se describe 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 no ser que se indique específicamente de otro modo, la palabra "o" se usa en el sentido inclusivo de "y/o" y no en el sentido exclusivo de "bien/o"

## 2.1 Meganucleasas diseñadas racionalmente con especificidad de secuencia alterada

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

De acuerdo con la divulgación, se predicen sustituciones de aminoácidos que pueden provocar un cambio deseado en la preferencia de bases determinando las cadenas laterales de aminoácidos de una meganucleasa de referencia (por ejemplo, una meganucleasa de tipo silvestre o una meganucleasa de referencia de origen no natural) que son capaces de participar en la realización de contactos con las bases de ácidos nucleicos de la secuencia de reconocimiento de ADN de la meganucleasa y la cadena principal de fosfodiéster de ADN, y la naturaleza espacial y química de esos contactos. Estos aminoácidos incluyen pero sin limitación cadenas laterales implicadas en la puesta en contacto con el semi-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 describe por datos de coordenadas atómicas, de un polipéptido o complejo de dos o más polipéptidos pueden obtenerse de varias maneras. Por ejemplo, pueden realizarse determinaciones de estructura proteica usando técnicas que incluyen, pero sin limitación, cristalografía de rayos X RMN, y espectrometría de masas. Otro enfoque es analizar bases de datos de coordenadas estructurales existentes para la meganucleasa de interés o una meganucleasa relacionada. Tales datos estructurales están con frecuencia disponibles de bases de datos en forma de coordenadas tridimensionales. Con frecuencia se puede acceder a estos datos a través de bases de datos en línea (por ejemplo, el Banco de Datos de Proteínas RCSB en [www.rcsb.org/pdb](http://www.rcsb.org/pdb)).

Puede obtenerse información estructural experimentalmente analizando los patrones de difracción de, por ejemplo, rayos X o electrones, creados por series bi o tridimensionales regulares (por ejemplo, cristales) de proteínas o complejos proteicos. Se usan métodos computacionales para transformar los datos de difracción en coordenadas atómicas tridimensionales en el espacio. Por ejemplo, se ha usado el campo de cristalografía de rayos X para generar información estructural tridimensional en muchos complejos de proteína-ADN, incluyendo meganucleasas (véase, por ejemplo., Chevalier *et al* (2001), *Nucleic Acids Res* 29(18): 3757-3774).

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

Como alternativa, puede usarse formación de modelos computacionales aplicando algoritmos basados en las estructuras primarias conocidas y, cuando estén disponibles, estructuras secundarias, terciarias y/o cuaternarias de la proteína/ADN, así como la naturaleza fisicoquímica conocida de las cadenas laterales de aminoácidos, bases de ácidos nucleicos e interacciones de enlace. Tales métodos pueden incluir opcionalmente enfoques por iteraciones, o restricciones derivadas experimentalmente. Un ejemplo de dicho software computacional es el programa CNS descrito en Adams *et al* (1999), *Acta Crystallogr. D. Biol. Crystallogr.* 55 (Pt 1): 181-90. Se ha desarrollado otros diversos programas computacionales que predicen el ordenamiento espacial de aminoácidos en una estructura proteica y predicen la interacción de las cadenas laterales de aminoácidos de la proteína con diversas moléculas diana (véase, por ejemplo, Patente de Estados Unidos. N° 6.988.041).

Por lo tanto, se divulgan modelos computacionales usados para identificar restos de aminoácidos específicos que interaccionan específicamente con bases de ácido nucleico ADN y/o facilitan interacciones de cadena principal de fosfodiéster no específica. Por ejemplo, pueden producirse modelos informáticos de la totalidad de las interacciones meganucleasa-ADN potenciales usando un programa de software adecuado, incluyendo, pero sin limitación, MOLSCRIPT™ 2.0 (Avatar Software AB, Stockhol, Suecia), el programa de presentación gráfica O (Jones *et al.* (1991), *Acta Crystallography*, A47: 110), el programa de presentación gráfica GRASP™ (Nicholls *et al* (1991), *PROTEINS, Structure, Function and Genetics* 11 (4): 281ff), o el programa de presentación gráfica INSIGHT™ (TSI, Inc., Shoreview, MN). Está disponible en el mercado hardware informático adecuado para producir, ver y manipular representaciones estructurales tridimensionales de complejos de proteína- ADN y se conoce bien en la técnica (por ejemplo, Silicon Graphics Workstation, Silicon Graphics, Inc., Mountainview, CA).

Específicamente, pueden resolverse interacciones entre una meganucleasa y sus secuencias de reconocimiento de ADN bicatenario usando métodos conocidos en la técnica. Por ejemplo, puede determinarse una representación, o modelo, de la estructura tridimensional de una estructura de complejo multicomponente, para la que se ha producido un cristal, usando técnicas que incluyen el reemplazo molecular o SIR/MIR (reemplazo isomorfo sencillo/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 pueden 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 *et al.* (1992), *X-PLOR Version 3.1.A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT).

La determinación de la estructura proteica e interacción de meganucleasa-ADN potencial posibilita selecciones racionales con respecto a los aminoácidos que pueden cambiarse para afectar a la actividad y especificidad enzimática. Las decisiones se basan en varios factores con respecto a interacciones de cadena laterales de aminoácidos con una base particular o cadena principal de fosfodiéster de ADN. Las interacciones químicas usadas para determinar las sustituciones de aminoácidos apropiadas incluyen, pero sin limitación, fuerzas de van der Waals, impedimento estérico, enlaces iónicos, enlaces de hidrógeno e interacciones hidrófobas. Pueden seleccionarse sustituciones de aminoácidos que favorecen o desfavorecen interacciones específicas de la meganucleasa con una base particular en un semi-sitio de secuencia de reconocimiento potencial para aumentar o reducir la especificidad para esa secuencia y, en cierto grado, la actividad y afinidad de unión global. Además, pueden seleccionarse sustituciones de aminoácidos que aumentan o reducen la afinidad de unión para la cadena principal de fosfodiéster de ADN bicatenario para aumentar o reducir la actividad global y, en cierto grado, reducir o aumentar la especificidad.

Por lo tanto, 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 bases en un semi-sitio de secuencia de reconocimiento de ADN. En algunas divulgaciones, la superficie de contacto comprende los aminoácidos de la enzima con carbonos-β a menos de 9,0 Å de un donador o aceptor con enlaces de hidrógeno en el surco principal en una de las bases en el par, y con cadenas laterales orientadas hacia el ADN, independientemente de si los restos realizan contactos con las bases en el complejo de ADN-meganucleasa de tipo silvestre. En otras divulgaciones, pueden excluirse restos si los restos no entran en contacto en el complejo de ADN-meganucleasa de tipo silvestre, o pueden incluirse o excluirse restos a discreción del diseñador para alterar el número o identidad de los restos considerados. En un ejemplo, como se describe posteriormente, para las posiciones de base -2, -7, -8 y -9 del semi- sitio de I-Crel de tipo silvestre, las superficies de contacto se limitaron a las posiciones de aminoácidos que realmente interaccionan en el complejo de ADN-enzima de tipo silvestre. 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 estaban implicadas en contactos de tipo silvestre pero que podrían potencialmente entrar en contacto con una base si se sustituyeran con un aminoácido diferente.

5 Debería observarse que, aunque un semi-sitio de secuencia de reconocimiento se representa normalmente con respecto a solamente una hebra de ADN, las meganucleasas se unen en el surco principal de ADN bicatenario, y entran en contacto con bases de ácidos nucleicos en ambas cadenas. Además, las designaciones de hebras "sentido" y "antisentido" son completamente arbitrarias con respecto a la unión y reconocimiento de meganucleasas. La especificidad de secuencia en una posición puede conseguirse a través de interacciones con un miembro de un par de bases, o por una combinación de interacciones con ambos miembros de un par de bases. Por lo tanto, por ejemplo, para favorecer la presencia de un par de bases A/T en la posición X, en la que la base A está en la hebra "sentido" y la base T está en la hebra "antisentido", se seleccionan restos que están suficientemente cerca para entrar en contacto con la hebra sentido en la posición X y que favorece la presencia de una A y/o se seleccionan restos que están suficientemente cerca para entrar en contacto con la hebra antisentido en la posición X y que favorecen la presencia de una T. De acuerdo con la invención, un resto se considera suficientemente cerca si el carbono-β del resto está a una distancia de 9 Å del átomo más cercano de la base relevante.

Por lo tanto, por ejemplo, un aminoácido con un carbono-β a una distancia de 9 Å de la hebra sentido de ADN pero mayor de 9 Å de la hebra antisentido se considera para interacciones potenciales solamente con la hebra sentido. De forma similar, un aminoácido con un carbono-β a una distancia de 9 Å de la hebra antisentido de ADN pero mayor de 9 Å de la hebra sentido se considera para interacciones potenciales solamente con la hebra antisentido. Los aminoácidos con carbonos-β que están a una distancia de 9 Å de ambas hebras de ADN se consideran para interacciones potenciales con cualquiera de las hebras.

25 Para cada superficie de contacto, las sustituciones de aminoácidos potenciales se seleccionan basándose en su capacidad predicha para interactuar favorablemente con una o más de las cuatro bases de ADN. El proceso de selección se basa en dos criterios primarios: (i) el tamaño de las cadenas laterales de aminoácidos, que afectará a sus interacciones estéricas con diferentes bases de ácido nucleico, y (ii) la naturaleza química de las cadenas laterales de aminoácidos, que afectarán a sus interacciones electrostáticas y de enlace con las diferentes bases de ácido nucleico.

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

Los aminoácidos con cadenas laterales relativamente más cortas y más pequeñas pueden asignarse al Grupo 1, incluyendo glicina (G), alanina (A), serina (S), treonina (T), cisteína (C), valina (V), leucina (L), isoleucina (I), aspartato (D), asparagina (N) y prolina (P). Se espera, sin embargo, que la prolina se use menos frecuentemente debido a su inflexibilidad relativa. Además, se espera que la glicina se use menos frecuentemente debido a que introduce flexibilidad no deseada en la cadena principal peptídica y su tamaño muy pequeño reduce la probabilidad de contactos eficaces cuando reemplaza un resto más grande. Por otro lado, la glicina puede usarse en algunos casos para promover una posición degenerada. Los aminoácidos con cadenas laterales de longitud y tamaño relativamente intermedio pueden asignarse al Grupo 2, incluyendo 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, incluyendo lisina (K), metionina (M), arginina (R), histidina (H), fenilalanina (F), tirosina (Y) y triptófano (W). Se espera, sin embargo, que el triptófano se use menos frecuentemente debido a su inflexibilidad relativa. Además, la flexibilidad de la cadena lateral de lisina, arginina y metionina permite que estos aminoácidos realicen contactos de base desde distancias largas o intermedias, garantizando su inclusión en los Grupos tanto 2 como 3. Estos grupos también se muestran en forma tabular 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 aminoácidos diferentes se evalúan con respecto a sus interacciones potenciales con las diferentes bases de ácido nucleico (por ejemplo, fuerzas de van der Waals, enlaces iónicos, enlaces de hidrógeno e interacciones hidrófobas) y se seleccionan restos que favorecen o desfavorecen a las interacciones específicas de la meganucleasa con una base particular en una posición particular en el semi-sitio de la secuencia de reconocimiento de ADN bicatenario. En algunos casos, puede desearse crear un semi-sitio con una o más posiciones degeneradas completas o parciales. En algunos casos, pueden seleccionarse restos que favorecen la presencia de dos o más bases, o restos que desfavorecen a una o más bases. Por ejemplo, puede conseguirse reconocimiento de bases degeneradas parciales por impedimento estérico de una pirimidina en una posición sentido o antisentido.

Se consigue el reconocimiento de bases de guanina (G) usando aminoácidos con cadenas laterales básicas que forman enlaces de hidrógeno con N7 y O6 de la base. Se confiere especificidad de citosina (C) por cadenas laterales cargadas negativamente que interactúan de forma desfavorable con los grupos electronegativos del surco principal en todas las bases excepto C. Se diseña racionalmente reconocimiento de timina (T) usando interacciones hidrófobas y de van der Waals entre cadenas laterales hidrófobas y el grupo metilo del surco principal en la base. Finalmente, se reconocen bases de alanina (A) usando las cadenas laterales de carboxamida Asn y Gln o la cadena lateral de hidroxilo de Tyr a través de un par de enlaces de hidrógeno con N7 y N6 de la base. Finalmente, puede usarse His para conferir especificidad para una base de purina (A o G) donando un enlace de hidrógeno con N7. Estas reglas sencillas para reconocimiento de ADN pueden aplicarse para predecir superficies de contacto en las que una o ambas de las bases en una posición de par de bases particular se reconocen a través de un contacto diseñado racionalmente.

Por lo tanto, basándose en sus interacciones de unión con las diferentes bases de ácido nucleico, y las bases que favorecen en una posición con la que entran en contacto, cada resto de aminoácido puede asignarse a uno o más grupos diferentes correspondientes a las diferentes bases que favorecen (es decir, G, C, T o A). Por tanto, 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 (Q), tirosina (Y) e histidina (H). Obsérvese que la histidina aparece tanto en el Grupo G como en el Grupo A; que la serina (S) no se incluye en ningún grupo pero puede usarse para favorecer una posición degenerada; y que prolina, glicina y triptófano no se incluyen en ningún grupo particular debido a consideraciones estéricas predominantes. Estos grupos también se muestran en forma tabular 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)	

Por lo tanto, de acuerdo con la divulgación, para efectuar un cambio deseado en el semi-sitio de secuencia de reconocimiento de una meganucleasa en una posición dada X, (1) determinar al menos la parte relevante de la estructura tridimensional del complejo ADN-meganucleasa de referencia o de tipo silvestre y las cadenas laterales de resto de aminoácido que definen la superficie de contacto en la posición X; (2) determinar la distancia entre el carbono-β de al menos un resto que comprende la superficie de contacto y al menos una base del par de bases de la posición X; y (3) (a) para un resto que está <6 Å de la base, seleccionar un resto del Grupo 1 y/o Grupo 2 que es un miembro apropiado de Grupo G, Grupo C, Grupo T o Grupo A para promover el cambio deseado y/o (b) para un resto que está >6 Å de la base, seleccionar un resto del Grupo 2 y/o Grupo 3 que es un miembro apropiado del Grupo G, Grupo C, Grupo T o Grupo A para promover el cambio deseado. Puede seleccionarse más de un resto tal que comprenda la superficie de contacto para análisis y modificación y, en algunas realizaciones, cada resto tal se analiza y se modifican restos múltiples. De forma similar, puede determinarse la distancia entre el carbono-β de un resto incluido en la superficie de contacto y cada una de las dos bases del par de bases en la posición X y, si el resto está a una distancia de 9 Å de ambas bases, entonces pueden realizarse diferentes sustituciones para afectar a las dos bases del par (por ejemplo, un resto del Grupo 1 para afectar a una base próxima en una hebra o un resto del Grupo 3 para afectar a una base distal en la otra hebra). Como alternativa, una combinación de sustituciones de restos capaces de interactuar con ambas bases en un par puede afectar a la especificidad (por ejemplo, un resto del Grupo T que

entra en contacto con la hebra sentido combinado con un resto del Grupo A que entra en contacto con la hebra antisentido para seleccionar con respecto a T/A). Finalmente, pueden validarse múltiples modificaciones alternativas de los restos empíricamente (por ejemplo, produciendo la meganucleasa recombinante y ensayando su reconocimiento de secuencia) o de forma computacional (por ejemplo, por realización de modelos informáticos del complejo de meganucleasa-ADN de la enzima modificada) para seleccionar entre alternativas.

Una vez que se seleccionan una o más modificaciones de aminoácidos deseadas de la meganucleasa de tipo silvestre o de referencia, la meganucleasa diseñada racionalmente puede producirse por métodos y técnicas recombinantes bien conocidos en la materia. En algunas realizaciones, se usan técnicas de mutagénesis dirigida o no aleatoria para crear modificaciones específicas de secuencia. Los ejemplos no limitantes de técnicas de mutagénesis no aleatoria incluyen PCR de cebador solapante (véase, por ejemplo, Wang *et al.* (2006), *Nucleic Acids Res.* 34(2): 517-527), mutagénesis dirigida (véase, por ejemplo, Patente de Estados Unidos N<sup>o</sup> 7.041.814), mutagénesis de casete (véase, por ejemplo, Patente de Estados Unidos N<sup>o</sup> 7.041.814) y el protocolo del fabricante para el kit de Sistemas de Mutagénesis Altered Sites<sup>®</sup> II disponible en el mercado de Promega Biosciences, Inc. (San Luis Obispo, CA).

El reconocimiento y escisión de una secuencia de ADN específica por una meganucleasa diseñada racionalmente puede ensayarse por cualquier método conocido por un experto en la materia (véase, por ejemplo, Publicación de Patente de Estados Unidos N<sup>o</sup> 2006/0078552). En ciertas realizaciones, la determinación de la escisión de meganucleasa se determina por ensayos de escisión *in vitro*. Tales ensayos usan escisión *in vitro* de un sustrato polinucleotídico que comprende la secuencia de reconocimiento pretendida de la meganucleasa ensayada y, en ciertas realizaciones, variaciones de la secuencia de reconocimiento pretendida en la que una o más bases en uno o ambos semi-sitios se ha cambiado a una base diferente. Normalmente, el sustrato polinucleotídico es una molécula de ADN bicatenaria que comprende un sitio diana que se ha sintetizado y clonado en un vector. El sustrato polinucleotídico puede ser lineal o circular, y normalmente comprende solamente una secuencia de reconocimiento. La meganucleasa se incuba con el sustrato polinucleotídico en condiciones apropiadas y los polinucleótidos resultantes se analizan por métodos conocidos para identificar productos de escisión (por ejemplo, electroforesis o cromatografía). Si hay una secuencia de reconocimiento sencilla 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 meganucleasa puede ensayarse como se describe en, por ejemplo Wang *et al.* (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, Publicación de Patente de Estados Unidos N<sup>o</sup> 2006/0078552). En realizaciones particulares, el ensayo *in vivo* es un ensayo de recombinación de hibridación monocatenaria (SSA). Este tipo de ensayo se conoce por los expertos en la materia (Rudin *et al.* (1989), *Genetics* 122: 519-534; Fishman-Lobell *et al.* (1992), *Science* 258: 480-4).

Como resultará evidente para un experto en la materia, pueden realizarse sustituciones, inserciones o deleciones de aminoácidos adicionales a dominios de las enzimas meganucleasa distintas de las implicadas en el reconocimiento y unión de ADN sin pérdida completa de actividad. Las sustituciones pueden ser sustituciones conservativas de restos de aminoácidos similares en posiciones restringidas estructural o funcionalmente, o pueden ser sustituciones no conservativas en posiciones que están menos restringidas estructural o funcionalmente. Tales sustituciones, inserciones y deleciones pueden identificarse por un experto en la materia mediante experimentación rutinaria sin esfuerzo indebido. Por lo tanto, las meganucleasas recombinantes divulgadas en el presente documento incluyen proteínas que tienen cualquiera de 85 % a 99 % de similitud de secuencia (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 de tipo silvestre I-CreI, I-MsoI, I-SceI e I-CeuI, las secuencias más N-terminales y C-terminales no son claramente visibles en estudios de cristalografía de rayos X, lo que sugiere que esas posiciones no están estructural o funcionalmente restringidas. Por lo tanto, estos restos pueden excluirse del cálculo de la similitud de secuencia y pueden usarse las siguientes secuencias de meganucleasa de referencia: restos 2-153 de SEQ ID NO: 1 para I-CreI, restos 6-160 de SEQ ID NO: 6 para I-MsoI, restos 3-186 de SEQ ID NO: 9 para I-SceI y restos 5-211 de SEQ ID NO: 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 hospedadores. Todos los miembros de esta familia tienen una o dos copias de un motivo LAGLIDADG altamente conservado junto con otros motivos estructurales implicados en escisión de secuencias de ADN específicas. Las enzimas que tienen una copia sencilla del motivo LAGLIDADG (es decir meganucleasas mono-LAGLIDADG) actúan como dímeros, mientras que las enzimas que tienen dos copias de este motivo (es decir, meganucleasas di-LAGLIDADG) actúan como monómeros.

Todos los miembros de la familia LAGLIDADG reconocen y escinden secuencias relativamente largas (> 12 pb), dejando salientes 3' de cuatro nucleótidos. Estas enzimas también comparten varios motivos estructurales además del motivo LAGLIDADG, incluyendo un ordenamiento similar de hebras p-anti-paralelas en la interfaz de ADN-proteína. Los aminoácidos dentro de estos motivos estructurales conservados son responsables de la interacción con las bases de ADN para conferir especificidad de secuencia de reconocimiento. La similitud estructural global entre algunos

miembros de la familia (por ejemplo, I-Crel, I-Msol, I-Scel y I-Ceul) se ha elucidado por cristalografía de rayos X. En consecuencia, los miembros de esta familia pueden modificarse en aminoácidos particulares dentro de tales motivos estructurales para cambiar la actividad global o especificidad de secuencia de las enzimas, y puede esperarse razonablemente que las modificaciones correspondientes tengan resultados similares en otros miembros de la familia. Véase, en general, Chevalier *et al.* (2001), *Nucleic Acid Res.* 29(18): 37573774).

### 2.2.1 Meganucleasas Derivadas de I-Crel

En un aspecto, se describen meganucleasas diseñadas racionalmente que se basan en o derivan de la meganucleasa I-Crel de *Chlamydomonas reinhardtii*. La secuencia de aminoácidos de tipo Silvestre de la meganucleasa I-Crel se muestra en SEQ ID NO: 1, que corresponde al N° de Acceso de Genbank PO5725. Se muestra a continuación dos semi-sitios de secuencia de reconocimiento de la meganucleasa I-Crel de tipo silvestre de la estructura cristalina PDB N° 1BP7:

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' SEQ ID NO: 2  
3'-C T T T G A C A G A G T G C T G C A A A A C-5' SEQ ID NO: 3

Posición -1-2-3-4-5-6-7-8-9

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

I-Crel de tipo silvestre también reconoce y corta la siguiente secuencia palindrómica (excepto para las bases centrales N1-N4):

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' SEQ ID NO: 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' SEQ ID NO: 5

Posición -1-2-3-4-5-6-7-8-9

Se considera que la secuencia palindrómica de SEQ ID NO: 4 y SEQ ID NO: 5 es un mejor sustrato para la I-Crel de tipo silvestre debido a que la enzima se une a este sitio con mayor afinidad y lo escinde más eficazmente que la secuencia de ADN natural. Para los fines de la presente descripción, y particularmente con respecto a los resultados experimentales presentados en el presente documento, esta secuencia palindrómica escindida por I-Crel de tipo silvestre se denomina "WT" (véase, por ejemplo, Figura 2 (A)). Los dos semi-sitios de secuencia de reconocimiento se muestran en negrita en sus hebras sentido respectivas.

La Figura 1 (A) representa las interacciones de un homodímero de meganucleasa I-Crel de tipo silvestre con una secuencia de reconocimiento de ADN bicatenario, la Figura 1(B) muestra las interacciones específicas entre restos de aminoácidos de la enzima y bases en la posición -4 de un semi-sitio para una enzima de tipo silvestre y una secuencia de reconocimiento de tipo silvestre, y las Figuras 1(C)-(E) muestran las interacciones específicas entre los restos de aminoácidos de la enzima y bases en la posición -4 de un semi-sitio para tres meganucleasas diseñadas racionalmente divulgadas en el presente documento con especificidad alterada en la posición -4 del semi-sitio.

Por lo tanto, la preferencia de bases en cualquier posición de base especificada del semi-sitio puede alterarse racionalmente a cada uno de los otros tres pares de bases usando los métodos descritos en el presente documento. En primer lugar, se determina la superficie de reconocimiento de tipo silvestre en la posición de base especificada (por ejemplo, analizando estructuras co-cristalinas del complejo de meganucleasa-ADN; o por realización de modelos informáticos de los complejos de meganucleasa-ADN). En segundo lugar, se determinan los restos de contacto existentes y potenciales basándose en las distancias entre los carbonos p de las posiciones de aminoácidos circundantes y las bases de ácido nucleico 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 restos de contacto de ADN-meganucleasa de tipo silvestre I-Crel en la posición -4 implican una glutamina en la posición 26 que realiza un enlace de hidrógeno con una base A en la hebra de ADN antisentido. El resto 77 también se identificó como potencialmente capaz de entrar en contacto con la base -4 en la hebra sentido de ADN. El carbono p del resto 26 está a 5,9 Å de distancia de N7 de la base A de la hebra de ADN antisentido, y el carbono p del resto 77 está a 7,15 Å de distancia del metilo C5 de la T en la hebra sentido. De acuerdo con las reglas químicas de base y distancia descritas en el presente documento, una C en la hebra sentido podría realizar un enlace de hidrógeno con un ácido glutámico en la posición 77 y una G en la hebra antisentido podría enlazar con glutamina en la posición 26 (mediada por una molécula de agua, como se observa en la estructura cristalina de I-Crel de tipo silvestre) (véase Figura 1 (C)); una G en la hebra sentido podría realizar un enlace de hidrógeno con una arginina en la posición 77 y una C en la hebra antisentido podría realizar un enlace de hidrógeno

con un ácido glutámico en la posición 26 (véase Figura 1(D)); una A en la hebra sentido podría realizar un enlace de hidrógeno con una glutamina en la posición 77 y una T en la hebra antisentido podría formar contactos hidrófobos con una alanina en la posición 26 (véase Figura 1 (E)). Si el contacto específico de base se proporciona por la posición 77, entonces el contacto de tipo silvestre, Q26, puede sustituirse (por ejemplo, con un resto de serina) para reducir o retirar su influencia en la especificidad. Como alternativa, pueden combinarse mutaciones complementarias en las posiciones 26 y 77 para especificar un par de bases particular (por ejemplo, A26 especifica una T en la hebra antisentido y Q77 especifica una A en la hebra sentido (Figura 1 (E)). Todas estas sustituciones de restos predichas se han validado experimentalmente.

- 5
- 10 Por lo tanto, de acuerdo con la invención, se ha identificado una variedad sustancial de modificaciones de aminoácidos para el dominio de reconocimiento de ADN de la meganucleasa I-Crel que, individualmente o en combinación, dan como resultado meganucleasas recombinantes con especificidades alteradas en bases individuales dentro del semi-sitio de secuencia de reconocimiento de ADN, de modo que estas meganucleasas diseñadas racionalmente tienen semi-sitios diferentes de la enzima de tipo silvestre. Las modificaciones de aminoácidos de I-Crel y el cambio resultante en la especificidad del semi-sitio de secuencia de reconocimiento se muestran en la Tabla 1:
- 15

TABLA 1

Posn.	Base de hebra sentido favorecida										
	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*	<b>Q44*</b>	C44*						
-3	Q68 C24* <b>I24*</b>	E68 F68 K24* R24*	<b>R68</b>	M68 C68 L68 F68		H68		Y68	K68		
-4	A26* Q77	E77 K26*	R77 E26*					S77 <b>Q26*</b>			S26*
-5		E42	R42			<b>K28*</b>	C28* Q42				M66 K66
-6	Q40 C28*	E40 R28*	R40	C40 I40 V40 C79 I79 V79 Q28*	A40 A79 A28* H28*						S40 S28*
-7	<b>N30*</b> <b>Q38</b>	E38 K30* R30*	K38 R38 E30*	I38 L38			C38				H38 N38 Q30*
-8	F33 Y33	E33 D33	F33 H33	L33 V33 I33 F33 C33		R32*	R33				



-9		E32	R32 K32	L32 V32 A32 C32				D32 132			S32 N32 H32 Q32 T32
----	--	-----	------------	--------------------------	--	--	--	------------	--	--	---------------------------------

### 2.2.2 Meganucleasas recombinante específicamente excluidas

5 La presente invención no pretende adoptar ciertas meganucleasas recombinantes que se han descrito en la técnica anterior y que se han desarrollado por métodos alternativos. Estas meganucleasas excluidas incluyen las descritas por Arnould *et al* (2006), J. Mol Biol. 355:443-58; Sussman *et al* (2004), J. Mol Biol. 342:31-41; Chames *et al* (2005), Nucleic Acids Res 33: e178; Seligman *et al* (2002), Nucleic Acids Res 30:3870-9; y Ashworth *et al* (2006), Nature 441 (7093): 656-659; las descripciones completas de las cuales se incorporan por la presente por referencia, incluyendo meganucleasas recombinantes basadas en I-Crel con sustituciones sencillas 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 "modificaciones excluidas".

### 2.2.3 Meganucleasas con múltiples cambios en el semi-sitio de secuencia de reconocimiento

En otro aspecto, se describen meganucleasas diseñadas racionalmente que se producen combinando dos o más modificaciones de aminoácidos como se describe en la sección 2.2.1 anterior, para alterar la preferencia de semi-sitio en dos o más posiciones en un semi-sitio de secuencia de reconocimiento de ADN. Por ejemplo, sin limitación, y como se describe más completamente posteriormente, la enzima DJ1 derivó de ICrel 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 degeneración completa en la posición -9). La meganucleasa DJ1 diseñada racionalmente reconoce de forma invariable C-7 G-6 G-5 en comparación con la preferencia de tipo silvestre por A.7 A-6 C.5, y tiene tolerancia aumentada para A en la posición -9.

La capacidad para combinar sustituciones de restos que afectan a diferentes posiciones de bases se debe en parte a la naturaleza modular de las meganucleasas LAGLIDADG. Una mayoría de los contactos de base en las interfaces de reconocimiento de LAGLIDADG se realizan por cadenas laterales de aminoácidos individuales, y la interfaz está relativamente sin interconectividad o redes de enlaces de hidrógeno entre cadenas laterales que interaccionan con bases adyacentes. Esto generalmente permite la manipulación de restos que interaccionan con una posición de base sin afectar a 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 método usado para identificar estas mutaciones. El método predice sustituciones de cadenas laterales que interaccionan directamente con una base sencilla. Generalmente se evita la interconectividad o redes de enlaces de hidrógenos entre cadenas laterales para mantener la independencia de las sustituciones dentro de la interfaz de reconocimiento.

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

Pueden usarse combinaciones de sustituciones de amino, identificadas como se describe en el presente documento, para alterar racionalmente la especificidad de una meganucleasa de tipo silvestre (o una meganucleasa previamente modificada) a partir 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 "sentido" de la secuencia de reconocimiento de meganucleasa I-Crel WT (SEQ ID NO: 4) así como varias otras secuencias para las que sería útil una meganucleasa diseñada racionalmente. Las bases conservadas entre la secuencia de reconocimiento WT y la secuencia de reconocimiento deseada están sombreadas. De acuerdo con la invención, pueden diseñarse racionalmente meganucleasas recombinantes basadas en la meganucleasa I-Crel para cada una de estas secuencias de reconocimiento deseadas, así como cualquier otra, por sustituciones de aminoácidos adecuadas como se describe en el presente documento.

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

Como se ha descrito anteriormente, la afinidad de unión a ADN de las meganucleasas recombinantes descritas en el presente documento pueden modularse alterando ciertos aminoácidos que forman la superficie de contacto con la cadena principal de fosfodiéster de ADN. La superficie de contacto comprende los aminoácidos en la enzima con carbonos-β a menos de 9 Å de la cadena principal de ADN y con cadenas laterales orientadas hacia el ADN, independientemente de si los restos realizan contactos con la cadena principal de ADN en el complejo de ADN-meganucleasa de tipo silvestre. Debido a que la unión con ADN es un precursor necesario para la actividad enzimática, se ha mostrado que los aumentos/reducciones de la afinidad de unión con ADN provocan aumentos/reducciones, respectivamente, en la actividad enzimática. Sin embargo, también se ha mostrado que los aumentos/reducciones de la afinidad de unión con ADN provocan reducciones/aumentos en la especificidad de secuencia de meganucleasa. Por lo tanto, tanto la actividad como la especificidad pueden modularse modificando los contactos de cadena principal de fosfodiéster.

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

(i) Retirar repulsión electrostática entre la enzima y cadena principal de ADN. Si un aminoácido identificado tiene una cadena lateral cargada negativamente (por ejemplo, ácido aspártico, ácido glutámico) que se esperaría que repeliera la cadena principal de ADN cargada negativamente, la repulsión puede eliminarse sustituyendo un aminoácido con una cadena lateral cargada positivamente o no cargada, sometida a efectos de interferencia estérica. Un ejemplo verificado experimentalmente es la mutación de ácido glutámico 80 en I-Crel a glutamina.

(ii) Introducir interacción de atracción electrostática entre la enzima y la cadena principal de ADN. En cualquiera de las posiciones de la superficie de contacto, se espera que la introducción de un aminoácido con una cadena lateral cargada positivamente (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 la cadena principal del ADN. Si un aminoácido de la superficie de contacto no realiza un enlace de hidrógeno con la cadena principal de ADN debido a que carece de una funcionalidad de enlaces de hidrógeno apropiada o tiene una cadena lateral que es demasiado corta, demasiado larga y/o demasiado inflexible para interactuar con la cadena principal de ADN, puede introducirse un aminoácido polar capaz de donar un enlace de hidrógeno (por ejemplo, serina, treonina, tirosina, histidina, glutamina, asparagina, lisina, cisteína o arginina) con la longitud y flexibilidad apropiadas, sujeto a efectos de interferencia estérica.

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

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

(ii) Retirar la atracción electrostática entre la enzima y el ADN. Si cualquier aminoácido de la superficie de contacto tiene una cadena lateral cargada positivamente (por ejemplo, lisina o arginina) que interactúa con la cadena principal de ADN cargada negativamente, esta interacción favorable puede eliminarse sustituyendo un aminoácido con una cadena lateral cargada negativamente o no cargada, sujeta a efectos de interferencia estérica. Un ejemplo verificado experimentalmente es la mutación de lisina 116 en I-Crel a ácido aspártico.

(iii) Retirar un enlace de hidrógeno entre la enzima y la cadena principal de ADN. Si cualquier aminoácido de la superficie de contacto realiza un enlace de hidrógeno con la cadena principal de ADN, este puede sustituirse por un aminoácido que no se esperaría que hiciera un enlace de hidrógeno similar debido a que su cadena lateral no está funcionalizada de forma apropiada o carece de las características de flexibilidad/longitud necesarias.

Por ejemplo, en algunas meganucleasas recombinantes basadas en I-Crel, el ácido glutámico en la posición 80 en la meganucleasa I-Crel se altera a una lisina o una glutamina para aumentar la actividad. En otras meganucleasas recombinantes, 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 se reduce cambiando la lisina en la posición 34 de I-Crel a ácido aspártico, cambiando la tirosina en la posición 66 a ácido aspártico y/o cambiando la lisina en la posición 116 a ácido aspártico.

Las actividades de las meganucleasas recombinantes pueden modularse de modo que la enzima recombinante tenga cualquiera de ninguna actividad a actividad muy alta con respecto a una secuencia de reconocimiento particular. Por ejemplo, la meganucleasa recombinante DJ1 cuando porta una mutación de ácido glutámico en la posición 26 pierde completamente su actividad. 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 semi-sitio de secuencia de reconocimiento (véase Figura 1(D)).

De acuerdo con la divulgación, los aminoácidos en diversas posiciones próximas a la cadena principal de ADN de fosfodiéster pueden cambiarse para afectar simultáneamente tanto a la actividad como a la especificidad de la meganucleasa. Este "ajuste" de la especificidad de actividad enzimática se consigue aumentando o reduciendo el

número de contactos realizados por aminoácidos con la cadena principal de fosfodiéster. Puede facilitarse diversos contactos con la cadena principal de fosfodiéster por cadenas laterales de aminoácidos. En algunos contactos, los enlaces iónicos, puentes salinos, enlaces de hidrógeno e impedimento estérico afectan a la asociación de cadenas laterales de aminoácidos con la cadena principal de fosfodiéster. Por ejemplo, para la meganucleasa ICrel, la alteración de la lisina en la posición 116 a un ácido aspártico retira un puente salino entre pares de bases de ácido nucleico en las posiciones -8 y -9, reduciendo la tasa de escisión enzimática pero aumentando la especificidad.

Los restos que forman la superficie de contacto de cadena principal de la meganucleasa de tipo silvestre I-Crel (SEQ ID NO: 1), I-Msol (SEQ ID NO: 6), I-Scel (SEQ ID NO: 9) and I-Ceul (SEQ ID NO: 12) se identifican en la Tabla 5 a continuación:

TABLA 5

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

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

- (1) Seleccionar un aminoácido de la Tabla 5 para la enzima correspondiente que está cargado negativamente (D o E), es hidrófobo (A, C, F, G, I, L, M, P, V, W, Y) o es no cargado/polar (H, N, Q, S, T).
- (2) Si el aminoácido está cargado negativamente o es hidrófobo, mutarlo a no cargado/polar (menos efecto) o cargado positivamente (K o R, más efecto).
- (3) Si el aminoácido es no cargado/polar, mutarlo a cargado positivamente.

Para reducir la afinidad de una enzima y hacerla de este modo menos activa/más específica:

- (1) Seleccionar un aminoácido de la Tabla 5 para la enzima correspondiente que esté cargado positivamente (K o R), sea hidrófobo (A, C, F, G, I, L, M, P, V, W, Y) o no cargado/polar (H, N, Q, S, T).
- (2) Si el aminoácido está cargado positivamente, mutarlo a no cargado/polar (menos efecto) o cargado negativamente (más efecto).
- (3) Si el aminoácido es hidrófobo o no cargado/polar, mutarlo a cargado negativamente.

#### 4. Meganucleasas Heterodiméricas

En otro aspecto, la invención proporciona meganucleasas que son heterodímeros formados por la asociación de dos monómeros, uno de los cuales puede ser una forma de tipo silvestre y uno o ambos de los cuales pueden ser una forma de origen no natural o recombinante. Por ejemplo, la meganucleasa I-Crel de tipo silvestre es normalmente un homodímero compuesto de dos monómeros que se unen cada uno a un semi-sitio en la secuencia de reconocimiento pseudo-palindrómica. Una meganucleasa recombinante heterodimérica puede producirse combinando dos meganucleasas que reconocen semi-sitios diferentes, por ejemplo co-expresando las dos meganucleasas en una célula o mezclando dos meganucleasas en solución. La formación de heterodímeros puede favorecerse frente a la formación de homodímeros alterando aminoácidos en cada uno de los dos monómeros que afectan a su asociación en dímeros. En realizaciones particulares, ciertos aminoácidos en la interfaz de los dos monómeros se alteran a partir de aminoácidos cargados negativamente (D o E) a aminoácidos cargados positivamente (K o R) en un primer monómero y de aminoácidos cargados positivamente a aminoácidos cargados negativamente en un segundo monómero (Tabla 6). Por ejemplo en el caso de meganucleasas derivadas de I-Crel, las lisinas de 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 proceso es un par de monómeros en el que el primer monómero tiene un exceso de restos cargados positivamente en la interfaz del dímero y el segundo monómero tiene un exceso de restos cargados negativamente en la interfaz del dímero. El primer y segundo monómero se asociarán, por lo tanto, preferentemente sobre sus pares de monómeros idénticos debido a las interacciones electrostáticas entre los aminoácidos alterados en la interfaz.

TABLA 6

I-Crel: Sustituciones de Primer Monómero	I-Crel: Sustituciones de Segundo Monómero
K7 a E7 o D7	E8 a K8 o R8
K57 a E57 o D57	E61 a K61 o R61
K96 a E96 o D96	

Como alternativa, o además, ciertos aminoácidos en la interfaz de los dos monómeros pueden alterarse para provocar impedimento estérico de la formación de homodímeros. Específicamente, los aminoácidos en la interfaz del dímero de un monómero se sustituyen con restos más grandes o más voluminosos que evitarán de forma estérica el homodímero. Los aminoácidos en la interfaz del dímero del segundo monómero pueden sustituirse opcionalmente con restos más pequeños para compensar los restos más voluminosos del primer monómero y retirar cualquier conflicto en

el heterodímero o pueden no estar modificados.

En otra alternativa, puede internarse un puente iónico o enlace de hidrógeno en el núcleo hidrófobo de una interfaz heterodimérica. Específicamente, un resto hidrófobo en un monómero en el núcleo de la interfaz puede sustituirse con un resto cargado positivamente. Además, un resto hidrófobo en el segundo monómero, que interacciona en el homodímero de tipo silvestre con el resto hidrófobo sustituido en el primer monómero, puede sustituirse con un resto cargado negativamente. Por lo tanto, los dos restos sustituidos pueden formar un puente iónico o enlace de hidrógeno. Al mismo tiempo, la repulsión electrostática de una carga satisfecha internada en una interfaz hidrófoba debería desfavorecer la formación de homodímeros.

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

## 5. Métodos para Producir Células y Organismos Recombinantes

Aspectos de la presente invención proporcionan adicionalmente métodos para producir células y organismos recombinantes, transgénicos o modificados genéticamente de otro modo usando meganucleasas diseñadas racionalmente. Por tanto, en ciertas realizaciones, se desarrollan meganucleasas recombinantes para provocar específicamente una rotura de doble cadena en un sitio sencillo o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para posibilitar la inserción o inserciones precisas de una secuencia de interés por recombinación homóloga. En otras realizaciones, se desarrollan meganucleasas recombinantes para provocar específicamente una rotura de doble cadena en un sitio sencillo o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para (a) posibilitar la inserción o las inserciones poco comunes de una secuencia de interés por unión de extremos no homólogos o (b) posibilitar la rotura 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 modo que la secuencia de interés se integra en el cromosoma. En el caso de recombinación homóloga, una secuencia insertada puede reemplazar una secuencia endógena, de modo que el ADN original se reemplaza por ADN exógeno de igual longitud, pero con una secuencia de nucleótidos alterada. Como alternativa, una secuencia insertada puede incluir más o menos bases que la secuencia que reemplaza.

Por lo tanto, de acuerdo con este aspecto de la invención, los organismos recombinantes incluyen, pero sin limitación, especies vegetales monocotiledóneas tales como arroz, trigo, maíz y centeno, y especies dicotiledóneas tales como legumbres (por ejemplo, alubias, soja, lentejas, cacahuetes, guisantes), alfalfa, trébol, tabaco y especies de *Arabidopsis*. Además, los organismos recombinantes pueden incluir, pero sin limitación, 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 métodos de la invención implican la introducción de una secuencia de interés en una célula tal como una célula germinal o célula madre que pueda convertirse en un organismo recombinante maduro o permita que el organismo modificado genéticamente resultante dé lugar a descendencia que porta la secuencia insertada de interés en su genoma, en el que dicha célula eucariota no es una célula madre.

Pueden suministrarse proteínas de meganucleasa a células para escindir ADN genómico, lo que posibilita la recombinación homóloga o unión a extremos no homólogos en el sitio de escisión con una secuencia de interés, por diversos mecanismos diferentes conocidos en la técnica. Por ejemplo, la proteína meganucleasa recombinante puede introducirse en una célula por técnicas que incluyen, pero sin limitación, microinyección o transfecciones de liposomas (véase, por ejemplo, Lipofectamine, Invitrogen Corp., Carlsbad, CA). La formulación de liposomas puede usarse para facilitar la fusión de bicapas lipídicas con una célula diana, permitiendo de este modo que los contenidos del liposoma o proteínas asociadas con su superficie se introduzcan en la célula. Como alternativa, la enzima puede fusionarse con un péptido de captación apropiado tal como el de la proteína TAT de VIH para dirigir la captación celular (véase, por ejemplo, Mudez *et al.* (2005), *Med. Res. Rev.* 25: 679-736).

Como alternativa, se insertan secuencias génicas que codifican la proteína meganucleasa en un vector y se transfectan en una célula eucariota usando técnicas conocidas en la materia (véase, por ejemplo, Ausubel *et al.*, *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 materia.

Los ejemplos no limitantes de vectores para transfección de ADN incluyen vectores de virus, plásmidos, cósmidos y vectores YAC. La transfección de secuencias de ADN puede conseguirse por diversos métodos conocidos por los expertos en la materia. Por ejemplo, se usan liposomas e inmunoliposomas para suministrar secuencias de ADN a células (véase, por ejemplo, Lasic *et al.* (1995), *Science* 267: 1275-76). Además, pueden utilizarse virus para introducir vectores en células (véase, por ejemplo, Patente de Estados Unidos N° 7.037.492). Como alternativa, pueden utilizarse estrategias de transfección de modo que los vectores se introduzcan como ADN desnudo (véase, por

ejemplo, Rui *et al.* (2002), *Life Sci.* 71(15): 1771-8).

Los métodos generales para suministrar ácidos nucleicos a células incluyen: (1) métodos químicos (Graham *et al.* (1973), *Virology* 54(2): 536-539; Zatloukal *et al.* (1992), *Ann. N.Y. Acad. Sci.*, 660:136-153); (2) métodos físicos tales como microinyección (Capecchi (1980), *Cell* 22(2): 479-488), electroporación (Wong *et al.* (1982), *Biochim. Biophys. Res. Commun.* 107(2): 584-587; Fromm *et al.* (1985), *Proc. Nat'l Acad. Sci. USA* 82(17): 5824-5828; Patente de Estados Unidos N° 5.384.253) e inyección balística (Johnston *et al.* (1994), *Methods Cell. Biol.* 43(A): 353-365; Fynan *et al.* (1993), *Proc. Nat'l Acad. Sci. USA* 90 (24): 11478-11482); (3) vectores virales (Clapp (1993), *Clin. Perinatol.* 20(1): 155-168; Lu *et al.* (1993), *J. Exp. Med.* 178 (6):2089-2096; Eglitis *et al.* (1988), *Avd. Exp. Med. Biol.* 241: 19-27; Eglitis *et al.* (1988), *Biotechniques* 6(7): 608-614); y (4) mecanismos mediados por receptor (Curiel *et al.* (1991), *Proc. Nat'l Acad. Sci. USA* 88(19): 8850-8854; Curiel *et al.* (1992), *Hum. Gen. Ther.* 3(2): 147-154; Wagner *et al.* (1992), *Proc. Nat'l Acad. Sci. USA* 89 (13): 6099-6103).

En ciertas realizaciones, se produce una planta modificada genéticamente, que contiene la secuencia de interés insertada en el genoma. En ciertas realizaciones, la planta modificada genéticamente se produce transfectando la célula vegetal con secuencias de ADN correspondientes a la meganucleasa recombinante y la secuencia de interés, que puede estar o no flanqueada por las secuencias de reconocimiento de meganucleasa y/o secuencias sustancialmente idénticas a la secuencia diana. En otras realizaciones, la planta modificada genéticamente se produce transfectando la célula vegetal con secuencias de ADN correspondientes a la meganucleasa recombinante solamente, de modo que la escisión promueva unión de extremos no homólogos y rompa la secuencia diana que contiene la secuencia de reconocimiento. En tales realizaciones, las secuencias de meganucleasa están bajo el control de secuencias reguladoras que posibilitan la expresión de la meganucleasa en las células vegetales hospedadoras. Estas secuencias reguladoras incluyen, pero sin limitación, promotores vegetales constitutivos tales como el promotor NOS, promotores génicos inducibles químicamente tales como el promotor inducible por dexametasona (véase, por ejemplo, Gremillon *et al.* (2004), *Plant J.* 37: 218-228), y promotores específicos de tejido vegetal tales como el promotor LGC1 (véase, por ejemplo, Singh *et al.* (2003), *FEBS Lett.* 542: 47-52).

Los métodos adecuados para introducir ADN en células vegetales incluyen prácticamente cualquier método por el que pueda introducirse ADN en una célula, incluyendo pero sin limitación una infección de *Agrobacterium*, transformación mediada por PEG de protoplastos (Omirulleh *et al.* (1993), *Plant Molecular Biology*, 21: 415-428), captación de ADN mediada por inhibición/deseccación, electroporación, agitación con fibras de carburo de silicio, inyección balística o bombardeo de microproyectiles y similares.

En otras realizaciones, se produce un animal modificado genéticamente usando una meganucleasa recombinante. Como con células vegetales, las secuencias de ácido nucleico pueden introducirse en una célula germinal o una célula que con el tiempo se convertirá en un organismo transgénico, en el que dicha célula eucariota no es una célula madre humana. En algunas realizaciones, la célula es un huevo fertilizado, y pueden inyectarse moléculas de ADN exógeno en el pro-núcleo del huevo fertilizado. Los huevos microinyectados se transfieren a los oviductos de madres adoptivas pseudoembarazadas y se permite que se desarrollen. La meganucleasa recombinante se expresa en el huevo fertilizado (por ejemplo, bajo el control de un promotor constitutivo tal como 3-fosfoglicerato quinasa) y facilita la recombinación homóloga de la secuencia de interés en uno o varios sitios discretos en el genoma. Como alternativa, los animales modificados genéticamente pueden obtenerse utilizando células madre embrionarias recombinantes ("ES") para la generación de los organismos transgénicos, como se describe en Gossler *et al.* (1986), *Proc. Natl. Acad. Sci. USA* 83: 9065 9069.

En ciertas realizaciones, un vector de expresión de mamíferos recombinante es capaz de dirigir la expresión específica de tejido del ácido nucleico preferentemente en un tipo celular particular. Se conocen en la técnica elementos reguladores específicos de tejido. Los ejemplos no limitantes de promotores específicos de tejido adecuados incluyen el promotor de albúmina (específico de hígado; Pinkert *et al.* (1987), *Genes Dev.* 1: 268-277), promotores específicos linfoides (Calame y Eaton (1988), *Adv. Immunol.* 43: 235-275), en promotores particulares de receptores de linfocitos T (Winoto y Baltimore (1989), *EMBO J.* 8: 729-733) e inmunoglobulinas (Banerji *et al.* (1983), *Cell* 33: 729-740; Queen y Baltimore (1983), *Cell* 33: 741-748), promotores específicos de neuronas (por ejemplo, el promotor de neurofilamentos; Byrne y Ruddle (1989), *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), promotores específicos del páncreas (Edlund *et al.* (1985), *Science* 230: 912-916) y promotores específicos de glándula mamaria (por ejemplo, promotor de suero de la leche; Patente de Estados Unidos N° 4.873.316 y Publicación de Patente Europea EP 0 264 166). También se abarcan promotores regulados por el desarrollo, por ejemplo, los promotores de hox murinos (Kessel y Gruss (1990), *Science* 249: 374-379) y el promotor de a-fetoproteína (Campes y Tilghman (1989), *Genes Dev.* 3: 537-546).

En ciertas realizaciones, una meganucleasa diseñada racionalmente puede marcarse con un epítipo peptídico (por ejemplo, un epítipo HA, FLAG o Myc) para controlar los 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 de SV40) o señales de localización de cloroplastos o mitocondrias. En otras realizaciones, la meganucleasa puede fusionarse con una señal de exportación nuclear para situarla en el citoplasma. La meganucleasa también puede fusionarse con una proteína no relacionada o dominio proteico tal como una proteína que estimula la recombinación homóloga o reparación de ADN (por ejemplo, recA, RAD51, RAD52,

RAD54, RAD57 o BRCA2).

6. Métodos para Terapia Génica

5 Se describe 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 génica tal como un promotor, potenciador o silenciador para reemplazar un gen o región reguladora génica que sea defectuosa en su estructura y/o función. La expresión "terapia génica" también puede referirse a modificaciones hechas a un gen deletéreo o elemento regulador (por ejemplo, oncogenes) que reduce o elimina la expresión del gen. Puede realizarse terapia génica para tratar afecciones congénitas, afecciones resultantes de mutaciones o daño a loci genéticos específicos durante la vida del paciente, o afecciones resultantes de organismos infecciosos.

15 En algunos aspectos, se reemplazan o deshabilitan genes disfuncionales mediante la inserción de secuencias de ácido nucleico exógeno en una región del genoma que afecta a la expresión génica. En ciertas realizaciones, la meganucleasa recombinante se dirige a una secuencia particular en la región del genoma para modificar de modo que alivie la afección. La secuencia puede ser una región dentro de un exón, intrón, promotor u otra región reguladora que esté provocando expresión disfuncional del gen. Como se usa en el presente documento, la expresión "expresión disfuncional" significa expresión aberrante de un producto génico porque la célula produce muy poco del producto génico, demasiado del producto génico, o produce un producto génico que tiene una función diferente tal como sin la función necesaria o que tiene más funciones que la necesaria.

25 Pueden usarse secuencias de ácido nucleico exógeno insertadas en la región modificada para proporcionar secuencias "reparadas" que normalizan el gen. Puede conseguirse reparación génica por la introducción de secuencias génicas apropiadas en el gen lo que permite que se restablezca la función apropiada. En estas realizaciones, la secuencia de ácido nucleico para insertar puede ser la secuencia codificante completa de una proteína o, en ciertas realizaciones, un fragmento del gen que comprende solamente la región para reparar. En otras realizaciones la secuencia de ácido nucleico para insertar comprende una secuencia promotora u otros elementos reguladores de modo que se reparen mutaciones que provocan expresión o regulación anómala. En otras realizaciones, la secuencia de ácido nucleico para insertar contiene el codón de parada de la traducción apropiado que falta en un gen mutado. La secuencia de ácido nucleico también puede tener secuencias para detener la transcripción en un gen recombinante sin señales de parada de la transcripción apropiadas.

35 Como alternativa, las secuencias de ácido nucleico pueden eliminar la función génica completamente rompiendo la secuencia reguladora del gen o proporcionando un silenciador para eliminar la función génica. En algunas realizaciones, la secuencia de ácido nucleico exógena proporciona un codón de parada de la traducción para evitar la expresión del producto génico. En otras realizaciones, las secuencias de ácido nucleico exógeno proporcionan elemento de parada de la transcripción para evitar la expresión de una molécula de ARN de longitud completa. En otras realizaciones más, la función génica se rompe directamente por la meganucleasa introduciendo inserciones de bases, supresiones de bases y/o mutaciones de desplazamiento de fase a través de unión de extremos no homólogos.

45 En muchos casos, es deseable dirigir las secuencias genéticas apropiadas a una célula o población de células diana que es la causa de la patología. Dicha dirección de agentes terapéuticos evita que los agentes terapéuticos se dirijan a células sanas. Esto aumenta la eficacia del tratamiento, reduciendo a la vez los efectos potencialmente negativos que el tratamiento podría tener en células sanas.

50 El suministro de genes de meganucleasa recombinantes y la secuencia de interés para insertar en el genoma de las células de interés puede conseguirse por diversos mecanismos. En algunas realizaciones, los ácidos nucleicos se suministran a las células por medio de virus con genes virales particulares inactivados para evitar la reproducción del virus. Por lo tanto, un virus puede alterarse de modo que sea capaz solamente de suministro y mantenimiento dentro de una célula diana, pero no conserve la capacidad para replicarse dentro de la célula o tejido diana. Pueden introducirse una o más secuencias de ADN en el genoma viral alterado, para producir un genoma viral que actúe como un vector, y puede o no insertarse en un genoma hospedador y expresarse posteriormente. Más específicamente, ciertas realizaciones incluyen emplear un vector retroviral tal como, pero sin limitación, los vectores MFG o pLJ. Un vector MFG es un vector de virus de leucemia murina de Moloney simplificado (MoMLV) en el que las secuencias de ADN que codifican las proteínas pol y env se han suprimido para hacerlo defectuoso para replicación. Un vector retroviral pLJ también es una forma del MoMLV (véase, por ejemplo, Korman *et al.* (1987), Proc. Nat'l Acad. Sci., 84: 2150-2154). En otras realizaciones, puede usarse un adenovirus recombinante o virus adeno-asociado como un vector de suministro.

60 En otros aspectos, el suministro de proteína meganucleasa recombinante y/o secuencias génicas de meganucleasa recombinante a una célula diana se consigue mediante el uso de liposomas. La producción de liposomas que contienen cargas de ácido nucleico y/o proteína se conoce en la técnica (véase, por ejemplo, Lasic *et al.* (1995), Science 267: 1275-76). Los inmunoliposomas incorporan anticuerpos contra antígenos asociados a células en liposomas, y pueden suministrar secuencias de ADN para la meganucleasa o la meganucleasa en sí misma a tipos celulares específicos (véase, por ejemplo, Lasic *et al.* (1995), Science 267: 1275-76; Young *et al.* (2005), J. Calif. Dent.

Assoc. 33(12): 967-71; Pfeiffer *et al.* (2006), J. Vasc. Surg. 43(5):1021-7). Se conocen bien en la técnica métodos para producir y usar formulaciones de liposomas (véase, por ejemplo, Patente de Estados Unidos N<sup>o</sup> 6.316.024, Patente de Estados Unidos N<sup>o</sup> 6.379.699, Patente de Estados Unidos N<sup>o</sup> 6.387.397, Patente de Estados Unidos N<sup>o</sup> 6.511.676 y Patente de Estados Unidos N<sup>o</sup> 6,593,308 y referencias citadas en las mismas). En algunas realizaciones, se usan liposomas para suministrar las secuencias de interés así como la proteína meganucleasa recombinante o secuencias génicas de meganucleasa recombinante.

#### 7. Métodos para Tratar Infección por Patógenos

Se describen métodos para tratar infección por un patógeno. Los organismos patógenos incluyen virus tales como, pero sin limitación, virus del herpes simple 1, virus del herpes simple 2, virus de la inmunodeficiencia humana 1, virus de la inmunodeficiencia humana 2, virus variola, virus de la polio, virus de Epstein-Barr y virus del papiloma humano, y organismos bacterianos tales como, pero sin limitación, *Bacillus anthracis*, especies de *Haemophilus*, especies de *Pneumococcus*, *Staphylococcus aureus*, especies de *Streptococcus*, *Staphylococcus aureus* resistente a meticilina y *Mycoplasma tuberculosis*. Los organismos patógenos también incluyen organismos fúngicos tales como, pero sin limitación, especies de *Candida*, *Blastomyces*, *Cryptococcus* y especies de *Histoplasma*.

En algunas realizaciones, una meganucleasa diseñada racionalmente puede dirigirse a una secuencia de reconocimiento dentro del genoma del patógeno, por ejemplo a un gen o elemento regulador que sea 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 meganucleasa de una secuencia de reconocimiento en un genoma de patógeno puede estimular la mutación dentro de un gen diana esencial en forma de una inserción, delección o desplazamiento de fase, estimulando la unión de extremos no homólogos. Como alternativa, la escisión de un plásmido bacteriano puede dar como resultado pérdida del plásmido junto con cualquier gen codificado en él, tal como genes de toxinas (por ejemplo, gen de Factor Letal de *B. anthracis*) o genes de resistencia a antibióticos. Como se ha observado anteriormente, la meganucleasa puede suministrarse al paciente infectado, animal o planta en forma proteica o de ácido nucleico usando técnicas que son habituales en la materia. En ciertas realizaciones, el gen de meganucleasa puede incorporarse en un genoma de bacteriófago para suministro a bacterias patógenas.

Se describen agentes terapéuticos para el tratamiento de ciertas formas de cáncer. Debido a que los virus humanos se asocian con frecuencia con formación de tumores (por ejemplo, Virus de Epstein-Barr y carcinomas nasofaríngeos; Virus del Papiloma Humano y cáncer cervical) la inactivación de estos patógenos virales puede evitar el desarrollo o progresión de cáncer. Como alternativa, pueden usarse roturas de doble cadena dirigidas a los genomas de estos virus asociados con tumores usando meganucleasas diseñadas racionalmente para desencadenar apoptosis a través de la ruta de respuesta a daño de ADN. De esta manera, es posible inducir selectivamente apoptosis en células tumorales que albergan el genoma viral.

#### 8. Métodos para Genotipación e Identificación de Patógenos

Se describen herramientas para investigación y desarrollo de biología molecular *in vitro*. Es habitual en la técnica usar endonucleasas específicas de sitio (por ejemplo, enzimas de restricción) para el aislamiento, clonación y manipulación de ácidos nucleicos tales como plásmidos, productos de PCR, secuencias BAC, secuencias YAC, virus y secuencias genómicas de organismos eucariotas y procariotas (véase, por ejemplo, Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley 1999). Por lo tanto, en algunas realizaciones, puede usarse una meganucleasa diseñada racionalmente para manipular secuencias de ácido nucleico *in vitro*. Por ejemplo, pueden usarse meganucleasas diseñadas racionalmente que reconocen un par de secuencias de reconocimiento dentro de la misma molécula de ADN para aislar el segmento de ADN intermedio para manipulación posterior tal como ligación en un plásmido bacteriano, BAC o YAC.

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

9. Métodos para la Producción de Dominios de Unión a ADN Adaptados

En otro aspecto, se describen proteínas de unión a ADN diseñadas racionalmente sin actividad de escisión de endonucleasas. La actividad catalítica de una meganucleasa diseñada racionalmente puede eliminarse mutando aminoácidos implicados en catálisis (por ejemplo, la mutación de Q47 a E en I-Crel, véase Chevalier *et al.* (2001), *Biochemistry*. 43: 14015-14026); la mutación de D44 o D145 a N en I-SceI; la mutación de E66 a Q en I-Ceul; la mutación de D22 a N en I-Msol). La meganucleasa inactivada puede fusionarse después con un dominio efector de otra proteína incluyendo, pero sin limitación, un activador de la transcripción (por ejemplo, el dominio de transactivación de GAL4 o el dominio de transactivación de VP16), un represor de la transcripción (por ejemplo, el dominio KRAB de la proteína Kruppel), un dominio de ADN metilasa (por ejemplo, M.CviPI o M.SssI) o un dominio de acetiltransferasa de histonas (por ejemplo, HDAC1 o HDAC2). Se conocen en la técnica proteínas quiméricas que consisten en un dominio de unión a ADN modificado por ingeniería genética, más notablemente un dominio de dedos de zinc modificado por ingeniería genética, y un dominio efector (véase, por ejemplo, Papworth *et al.* (2006), *Gene* 366: 27-38).

**Ejemplos**

La presente invención se ilustra adicionalmente mediante los siguientes ejemplos, que no deben entenderse como limitantes. Los ejemplos 1-4 a continuación se refieren específicamente a meganucleasas diseñadas racionalmente basadas en I-Crel. Los ejemplos que no pertenecen a la invención son solo con fines ilustrativos.

**EJEMPLO 1**

Diseño Racional de Meganucleasas que Reconocen el Gen TAT de VIH-1

1. Diseño de Meganucleasas.

Se diseñó un par de meganucleasas para reconocer y escindir el sitio de ADN 5'-GAAGAGCTCATCAGAACAGTCA-3' (SEQ ID NO: 15) hallado en el Gen TAT de VIH-1. De acuerdo con la Tabla 1, se diseñaron dos meganucleasas, TAT1 y TAT2, para unir los semi-sitios 5'-GAAGAGCTC-3' (SEQ ID NO: 16) y 5'-TGACTGTTC-3' (SEQ ID NO: 17), respectivamente, usando los siguientes contactos de bases (los contactos no WT están en negrita):

TAT1:

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

TAT2:

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

Las dos enzimas se clonaron, se expresaron en *E. coli* y se ensayaron con respecto a actividad enzimática frente a la secuencia de reconocimiento de ADN correspondiente como se describe posteriormente. En ambos casos, se descubrió que las meganucleasas diseñadas racionalmente estaban inactivas. Se produjo después una segunda generación de cada una en la que E80 estaba mutada a Q para mejorar los contactos con la cadena principal de ADN. Se descubrió que la enzima TAT2 de segunda generación era activa contra su secuencia de reconocimiento pretendida mientras que la enzima TAT1 de segunda generación permanecía inactiva. La inspección visual de la estructura co-cristalina de I-Crel de tipo silvestre sugirió que TAT1 estaba inactiva debido a un conflicto estérico entre R40 y K28. Para aliviar este conflicto, 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) manteniendo a la vez la mutación Q80. Cuando estas enzimas se produjeron en *E. coli* y se ensayaron se descubrió que las variantes TAT1 con S28 y T28 eran ambas activas contra la secuencia de reconocimiento pretendida manteniendo a la vez la preferencia de base deseada en la posición -7.

2. Construcción de Meganucleasas Recombinantes.

Se introdujeron mutaciones para las enzimas I-Crel rediseñadas usando cebadores mutagénicos en una estrategia de PCR solapante. Se unieron fragmentos de ADN recombinantes de I-Crel generados en una PCR primaria 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 un marcador de seis histidinas fusionado en el extremo 3' del gen para purificación (Novagen Corp., San Diego, CA). Todas las secuencias de ácido nucleico se confirmaron usando secuenciación de Didesoxinucleótido de Sanger (véase Sanger *et al.* (1977), *Proc. Natl. Acad. Sci. USA*. 74 (12):



5463-7).

I-Crel de tipo Silvestre y todas las meganucleasas obtenidas por ingeniería genética se expresaron y se purificaron usando el siguiente método. Las construcciones clonadas en un vector pET21a se transformaron en pLysS BL21 (DE3) químicamente competentes, y se sembraron en placas 2xYT que contenían carbanilina 200 µg/ml. Después de crecimiento durante una noche, las colonias bacterianas transformadas se retiraron por raspado de las placas y se usaron para inocular 50 ml de caldo 2XYT. Las células se cultivaron a 37 °C con agitación hasta que alcanzaron una densidad óptica de 0,9 a una longitud de onda de 600 nm. La temperatura de crecimiento se redujo después de 37 °C a 22 °C. Se indujo expresión proteica mediante la adición de IPTG 1 mM y las células se incubaron con agitación durante dos horas y media. Las células se sedimentaron después por centrifugación durante 10 minutos 6.000 x g. 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 en vórtex. Las células se rompieron después con 12 pulsos de sonicación a 50 % de potencia y los residuos celulares se sedimentaron por centrifugación durante 15 minutos a 14.000 x g. Los sobrenadantes celulares se diluyeron en 4 ml de tampón de unión y se cargaron en una columna de Sepharose quelante metálica 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 meganucleasa 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 a tampón SA (Tris-HCl 25 mM, pH 8,0, NaCl 100 mM, MgCl<sub>2</sub> 5 mM, EDTA 5mM) para ensayos y almacenamiento usando columnas de desalación de centrifugación Zeba (Pierce Biotechnology, Inc., Rockford, IL). La concentración enzimática se determinó por absorbancia a 280 nm usando un coeficiente de extinción de 23.590 M<sup>-1</sup>cm<sup>-1</sup>. La pureza y el peso molecular de las enzimas se confirmaron después por espectrometría de masas MALDI-TOF.

Se produjeron enzimas heterodiméricas por purificación de las dos proteínas de forma independiente y mezclándolas *in vitro* o construyendo un operón artificial para expresión en tándem de las dos proteínas en *E. coli*. En el caso anterior, las meganucleasas purificadas se mezclaron 1:1 en solución y se pre-incubaron a 42 °C durante 20 minutos antes de la adición de sustrato de ADN. En el segundo caso, los dos genes se clonaron secuencialmente en el vector de expresión pET-21a usando *Mdel/EcoRI* y *EcoRI/HindIII*. El primer gen en el operón termina con dos codones de parada para evitar errores de lectura durante la transcripción. Un espaciador de ácido nucleico de 12 pares de bases y una secuencia Shine-Dalgarno del vector pET21 separaron el primer y segundo genes en el operón artificial.

### 3. Ensayos de Escisión.

Todas las enzimas purificadas como se han descrito anteriormente se ensayaron con respecto a actividad por incubación con sustratos de ADN lineales, bicatenarios que contenían la secuencia de reconocimiento de meganucleasas. Se hibridaron oligonucleótidos genéticos correspondientes a las hebras tanto sentido como antisentido de la secuencia de reconocimiento 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 didesoxinucleótidos Sanger. Todos los sustratos plasmídicos se linealizaron con *XmnI*, *Scal* o *BpmI* simultáneamente con el producto de digestión de la meganucleasa. Los productos de digestión enzimática contenían 5 µl de sustrato de ADN 0,05 µM, 2,5 µl de meganucleasa I-Crel recombinante 5 µM, 9,5 µl de tampón SA y 0,5 µl de *XmnI*, *Scal* o *BpmI*. Los productos de digestión se incubaron a 37 °C o 42 °C para ciertas enzimas meganucleasas, durante 4 horas. Las digestiones se detuvieron añadiendo Proteínasa K 0,3 µg/ml y SDS 0,5 %, y se incubaron durante una hora a 37 °C. Las digestiones se analizaron en agarosa 1,5 % y se visualizaron por tinción con bromuro de etidio.

Para evaluar la preferencia de semi-sitios de meganucleasa, se incubaron meganucleasas diseñadas racionalmente con un conjunto de sustratos de ADN correspondientes a un palíndromo perfecto del semi-sitio pretendido así como cada una de las 27 sustituciones de pares de bases sencillas posibles en el semi-sitio. De esta manera, fue posible determinar cuán tolerante es cada enzima a desviaciones de su semi-sitio pretendido.

### 4. Especificidad de Secuencia de Reconocimiento.

Las meganucleasas recombinantes purificadas TAT1 y TAT2 reconocieron secuencias de ADN que eran distintas de la secuencia de reconocimiento de meganucleasa de tipo silvestre (Figura 2 (B)). La meganucleasa I-Crel de tipo silvestre escinde la secuencia de reconocimiento WT, pero no corta la secuencia pretendida para TAT1 ni la secuencia pretendida para TAT2. TAT1 y TAT2, de forma similar, cortan sus secuencias de reconocimiento pretendidas pero no la secuencia de tipo silvestre. Las meganucleasas se evaluaron después con respecto a preferencia de semi-sitio y especificidad global (Figura 3). Se descubrió que I-Crel de tipo silvestre era altamente tolerante a sustituciones de par de bases sencillo en su semi-sitio natural. Por el contrario, se descubrió que TAT1 y TAT2 eran altamente específicas y completamente intolerantes a sustituciones de bases en las posiciones -1, -2, -3, -6 y -8 en el caso de TAT1 y las posiciones -1, -2 y -6 en el caso de TAT2.

**EJEMPLO 2**Diseño Racional de Meganucleasas con Afinidad de Unión a ADN Alterada5 1. Meganucleasas con afinidad aumentada y actividad aumentada.

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

10

CCR1:

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

BRP2:

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

15 Ambas enzimas se expresaron en *E. coli*, se purificaron y se ensayaron como en el Ejemplo 1. Se descubrió que ambas enzimas de primera generación escindían sus secuencias de reconocimiento pretendidas con tasas que eran considerablemente más bajas que las de I-Crel de tipo silvestre con su secuencia de reconocimiento natural. Para aliviar esta pérdida de actividad, la afinidad de unión a ADN de CCR1 y BRP2 se aumentó mutando E80 a Q en ambas enzimas. Se descubrió que estas versiones de segunda generación de CCR1 y BRP2 escindían sus secuencias de reconocimiento pretendidas con tasas catalíticas sustancialmente aumentadas.

20

2. Meganucleasas con afinidad de unión a ADN reducida y actividad reducida pero especificidad aumentada.

25 Se descubrió que I-Crel de tipo silvestre era altamente tolerante a sustituciones en su semi-sitio (Figura 3(A)). En un intento de hacer la enzima más específica, la lisina en la posición 116 de la enzima, que normalmente realiza un puente salino con un fosfato de la cadena principal de ADN se mutó a ácido aspártico para reducir la afinidad de unión a ADN. Se descubrió que esta enzima diseñada racionalmente escindía la secuencia de reconocimiento de tipo silvestre con actividad sustancialmente reducida pero la enzima recombinante era considerablemente más específica que la de tipo silvestre. La preferencia de semi-sitio de la variante K116D se evaluó como en el Ejemplo 1 y se descubrió que la enzima era completamente intolerante a la desviación de su semi-sitio natural en las posiciones -1, -2 y -3 y presentaba al menos preferencia de base parcial en las 6 posiciones restantes en el semi-sitio ((Figura 3(B)).

30

**EJEMPLO 3**35 Heterodímeros de Meganucleasa Diseñada Racionalmente1. Escisión de sitios de ADN no palindrómicos por heterodímeros de meganucleasa formados en solución.

40 Se diseñaron dos meganucleasas, LAM1 y LAM2, para escindir los semi-sitios 5'-TGCGGTGTC-3' (SEQ ID NO: 20) y 5'-CAGGCTGTC-3' (SEQ ID NO: 21), respectivamente. Se esperaba que el heterodímero de estas dos enzimas reconociera la secuencia de ADN 5'-TGCGGTGCCGGCGACAGCCTG-3' (SEC ID ° 22) hallada en el gen p05 del bacteriófago X.

45

LAM1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	T	G	C	G	G	T	G	T	C
Restos de Contacto	<b>C32</b>	<b>R33</b>	<b>R30/E38</b>	<b>D28/R40</b>	<b>R42</b>	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
Restos de Contacto	S32	Y33	<b>E30/R38</b>	<b>R40</b>	<b>K28/E42</b>	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 después 1:1 y se incubaron a 42 °C durante 20 minutos para permitirles intercambiar subunidades y reequilibrarse. La solución de enzima resultante, que se espera que sea una mezcla de homodímero LAM1, homodímero LAM2 y heterodímero LAM1/LAM2, se incubó con tres secuencias de reconocimiento diferentes correspondientes al palíndromo perfecto del semi-sitio LAM1, el palíndromo perfecto del semi-sitio LAM2 y el sitio híbrido no palindrómico hallado en el genoma del bacteriófago X. La enzima LAM1 purificada sola corta el sitio palindrómico de LAM1, pero no el sitio palindrómico LAM2 ni el sitio híbrido de LAM1/LAM2. De forma similar, la enzima LAM2 purificada sola corta el sitio palindrómico LAM2 pero no el sitio palindrómico LAM1 ni el sitio híbrido de LAM1/LAM2. La mezcla 1:1 de LAM1 y LAM2, sin embargo, escinde los tres sitios de ADN. La escisión del sitio híbrido LAM1/LAM2 indica que pueden mezclarse dos meganucleasas rediseñadas distintas en solución para formar una enzima heterodimérica capaz de escindir un sitio de ADN no palindrómico.

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

Los genes que codifican las enzimas LAM1 y LAM 2 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 secuencias de reconocimiento potenciales descritas anteriormente. Se descubrió que la mezcla de enzimas co-expresada escindía los tres sitios, incluyendo sitio híbrido LAM1/LAM2, lo que indica que pueden co-expresarse dos meganucleasas diseñadas racionalmente distintas para formar una enzima heterodimérica capaz de escindir un sitio de ADN no palindrómico.

3. Escisión preferente de sitios de ADN no palindrómicos por heterodímeros de meganucleasa con interfaces proteína-proteína modificadas.

Para aplicaciones que requieran la escisión de sitios de ADN no palindrómicos, es deseable promover la formación de heterodímeros de enzima minimizando a la vez 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 las lisinas en las posiciones 7, 57 y 96 se cambiaron a ácidos glutámicos. Esta enzima se co-expresó después y se purificó como anteriormente con una variante de LAM2 en la que los ácidos glutámicos en las posiciones 8 y 61 se cambiaron a lisina. En este caso, se esperaba que la formación del homodímero de LAM1 se redujera debido a la repulsión electrostática entre E7, E57 y E96 en un monómero y E8 y E61 en el otro monómero. De forma similar, se esperaba que la formación del homodímero de LAM2 se redujera debido a repulsión electrostática entre K7, K57 y K96 en un monómero y K8 y K61 en el otro monómero. Por el contrario, se esperaba que el heterodímero 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 interfaces modificadas se co-expresaron y ensayaron como se ha descrito anteriormente, se descubrió que el sitio híbrido LAM1/LAM2 se escindía preferentemente frente a los dos sitios palindrómicos, lo que indica que las sustituciones en la interfaz proteína-proteína de meganucleasa pueden dirigir la formación preferente de heterodímeros.

**EJEMPLO 4**

Heterodímeros de Meganucleasa Adicionales que Escinden Secuencias de ADN Fisiológicas

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

Puede producirse un heterodímero de meganucleasa diseñado racionalmente (ACH1/ACH2) que escinde la secuencia 5'-CTGGGAGTCTCAGGACAGCCTG-3' (SEQ ID NO: 23) en el gen FGFR3 humano, en el que mutaciones provocan acondroplasia. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Cre1, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

ACH1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	I	G	G	G	A	G	I	C
Restos de Contacto	<b>D32</b>	<b>C33</b>	<b>E30/R38</b>	<b>R40/D28</b>	<b>R42</b>	<b>A26/Q77</b>	<b>R68</b>	<b>Q44</b>	<b>R70</b>

ACH2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	C	A	G	G	C	I	G	I	C
Restos de Contacto	<b>D32</b>	<b>Y33</b>	<b>E30/R38</b>	<b>R40</b>	<b>K28/E42</b>	<b>Q26</b>	<b>R68</b>	<b>Q44</b>	<b>R70</b>

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (HGH1/HGH2) que escinda la secuencia 5'-CCAGGTGTCTCTGGACTCCTCC-3' (SEQ ID NO: 24) en el promotor del gen de Hormona del

Crecimiento Humana. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

HGH1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>C</u>	<u>A</u>	<u>G</u>	<u>G</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>D32</b>	<b>C33</b>	N30/Q38	<b>R40/D28</b>	<b>R42</b>	Q26	R68	Q44	R70

5

HGH2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>G</u>	<u>G</u>	<u>A</u>	<u>G</u>	<u>G</u>	<u>A</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>K32</b>	<b>R33</b>	N30/Q38	<b>R40/D28</b>	<b>R42</b>	<b>A26</b>	R68	Q44	R70

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (CF1/CF2) que escinda la secuencia 5'-GAAAATCATTGGTGTTCCT-3' (SEQ ID NO: 25) en el alelo ΔF508 del gen CFTR humano. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

10

CF1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>G</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>I</u>	<u>A</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>S32</b>	<b>Y33</b>	N30/Q38	<b>Q40</b>	<b>K28</b>	<b>Q26</b>	<b>H68/C24</b>	Q44	R70

15

CF2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>A</u>	<u>G</u>	<u>G</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>C</u>	<u>A</u>	<u>C</u>
Restos de Contacto	<b>N32</b>	<b>R33</b>	<b>E30/R38</b>	<b>Q40</b>	<b>K28</b>	<b>A26</b>	<b>Y68/C24</b>	T44	R70

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (CCR1/CCR2) que escinde la secuencia 5'-AACCTCTCCAGTGAGATGCCT-3' (SEQ ID NO: 26) en el del gen CcR5 humano (y correceptor de VIH). Por ejemplo, se diseñó una meganucleasa basándose en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

20

CCR1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>A</u>	<u>A</u>	<u>C</u>	<u>C</u>	<u>C</u>	<u>I</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>N32</b>	<b>Y33</b>	<b>R30/E38</b>	<b>E40/R28</b>	<b>E42</b>	Q26	<b>Y68/K24</b>	Q44	R70

CCR2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>A</u>	<u>G</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>I</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>N32</b>	<b>R33</b>	<b>E30/R38</b>	<b>E40</b>	<b>K28</b>	Q26	<b>Y68/K24</b>	Q44	R70

25

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (MYD1/MYD2) que escinde la secuencia 5'-GACCTCGTCCCTCCGACTCGCTG-3' (SEC ID N° 27) en la región no traducida 3' del gen de DM quinasa humana. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

30

MYD1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>G</u>	<u>A</u>	<u>C</u>	<u>C</u>	<u>I</u>	<u>C</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>S32</b>	<b>Y33</b>	<b>R30/E38</b>	<b>E40/R28</b>	<b>K66</b>	Q26/E77	R68	Q44	R70

MYD1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>C</u>	<u>A</u>	<u>G</u>	<u>C</u>	<u>G</u>	<u>A</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>Y33</u>	<u>E30/R38</u>	<u>E40/R28</u>	<u>R42</u>	<u>A26/Q77</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

2. Heterodímeros de meganucleasa que escinden secuencias de ADN en genomas patógenos.

5 Puede producirse un heterodímero de meganucleasa diseñada racionalmente (VHS1/VHS2) que escinde la secuencia 5'-GACCTCGTCTCCGACTCGCTG-3' (SEC ID N° 28) en el gen UL36 de Virus del Herpes Simple-1 y Virus de Herpes Simple-2. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

10

VHS1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>C</u>	<u>I</u>	<u>C</u>	<u>G</u>	<u>A</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>C33</u>	<u>R30/E38</u>	<u>R40/</u>	<u>Q42/K28</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

VHS2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>I</u>	<u>G</u>	<u>C</u>	<u>C</u>	<u>G</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>C32</u>	<u>R33</u>	<u>R30/E38</u>	<u>E40/R28</u>	<u>R42</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

15

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (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 basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

ANT1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>A</u>	<u>C</u>	<u>A</u>	<u>A</u>	<u>G</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>N32</u>	<u>C33</u>	<u>N30/Q38</u>	<u>Q40/A28</u>	<u>R42</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

20

ANT2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>I</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>C</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>C32</u>	<u>Y33</u>	<u>N30/Q38</u>	<u>Q40</u>	<u>E42</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

25

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (POX1/POX2) que escinde la secuencia 5'-AAAAGTGTCAAATGACATCGCA-3' (SEC ID N° 30) en el gen gp009 de virus Variola (viruela). Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

POX1:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>C</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>N32</u>	<u>C33</u>	<u>N30/Q38</u>	<u>Q40</u>	<u>K28</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

30

POX2:

Posición	-9	-8	-7	-6	-5	-4	-3	-2	-1
Base	<u>I</u>	<u>G</u>	<u>C</u>	<u>G</u>	<u>A</u>	<u>I</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>C32</u>	<u>R33</u>	<u>R30/E38</u>	<u>R40</u>	<u>C28/Q42</u>	<u>Q26</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

Puede producirse un homodímero de meganucleasa diseñada racionalmente (EBB1/EBB1) 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 basada en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

5

EBB1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>G</u>	<u>G</u>	<u>G</u>	<u>G</u>	<u>I</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>R33</u>	<u>D30/O38</u>	<u>R40/D28</u>	<u>R42</u>	<u>Q26</u>	<u>Y68/K24</u>	<u>Q44</u>	<u>R70</u>

EBB1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>A</u>
Base	<u>G</u>	<u>G</u>	<u>A</u>	<u>G</u>	<u>G</u>	<u>C</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>R33</u>	<u>D30/Q38</u>	<u>R40/D28</u>	<u>R42</u>	<u>Q26</u>	<u>Y68/K24</u>	<u>Q44</u>	<u>R70</u>

3. Heterodímeros de meganucleasa que escinden secuencias de ADN en genomas vegetales

10

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (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 basada en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

15

GLA1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>A</u>	<u>C</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>Y33</u>	<u>R30/E38</u>	<u>S40/C79</u>	<u>K28</u>	<u>A26/Q77</u>	<u>Y68/K24</u>	<u>Q44</u>	<u>R70</u>

GLA2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>A</u>	<u>C</u>	<u>C</u>	<u>G</u>	<u>A</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>Y33</u>	<u>R30/E38</u>	<u>E40/R28</u>	<u>R42</u>	<u>A26/Q77</u>	<u>Y68/K24</u>	<u>Q44</u>	<u>R70</u>

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (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 basada en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

25

BRP1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>I</u>	<u>G</u>	<u>C</u>	<u>C</u>	<u>I</u>	<u>C</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>C32</u>	<u>R33</u>	<u>R30/E38</u>	<u>R28/E40</u>	<u>K66</u>	<u>Q26/E77</u>	<u>Y68/K24</u>	<u>Q44</u>	<u>R70</u>

BRP2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>I</u>	<u>C</u>	<u>C</u>	<u>G</u>	<u>G</u>	<u>G</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<u>S32</u>	<u>C33</u>	<u>R30/E38</u>	<u>E40/R28</u>	<u>R42</u>	<u>S26/R77</u>	<u>R68</u>	<u>Q44</u>	<u>R70</u>

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (MGC1/MGC2) que escinde la secuencia 5'-TAAAATCTCTAAGGTCTGTGCA-3' (SEC ID N° 34) en el gen de Magnesio Quelatasa de *Nicotiana tabacum*. Por ejemplo, se diseñó una meganucleasa basada en la meganucleasa I-CreI, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

30

MGC1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>I</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>I</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>C32</b>	<b>Y33</b>	<b>N30/Q38</b>	<b>Q40/</b>	<b>K28</b>	<b>Q26</b>	<b>Y68/K24</b>	<b>Q44</b>	<b>R70</b>

MGC2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>I</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>C</u>	<u>A</u>	<u>G</u>	<u>A</u>	<u>C</u>
Restos de Contacto	<b>S32</b>	<b>R33</b>	<b>R30/E38</b>	<b>Q40</b>	<b>K28</b>	<b>A26/Q77</b>	<b>R68</b>	<b>T44</b>	<b>R70</b>

- 5 Puede producirse un heterodímero de meganucleasa diseñada racionalmente (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 basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

10

CYP:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>C</u>	<u>A</u>	<u>A</u>	<u>G</u>	<u>A</u>	<u>A</u>	<u>I</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>D32</b>	<b>Y33</b>	<b>N30/Q38</b>	<b>R40/</b>	<b>K28</b>	<b>Q77/A26</b>	<b>Y68</b>	<b>Q44</b>	<b>R70</b>

HGH2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>I</u>	<u>I</u>	<u>A</u>	<u>A</u>	<u>I</u>	<u>G</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>S32</b>	<b>C33</b>	<b>N30/Q38</b>	<b>Q40</b>	<b>K66</b>	<b>R77/S26</b>	<b>Y68/K24</b>	<b>Q44</b>	<b>R70</b>

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

15

Puede producirse un heterodímero de meganucleasa diseñada racionalmente (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 basada en la meganucleasa I-Crel, como se ha descrito anteriormente, con los siguientes restos de contacto y semi-sitios de secuencia de reconocimiento:

20

URA1:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>I</u>	<u>I</u>	<u>A</u>	<u>G</u>	<u>A</u>	<u>I</u>	<u>G</u>	<u>A</u>	<u>C</u>
Restos de Contacto	<b>S32</b>	<b>C33</b>	<b>N30/Q38</b>	<b>R40</b>	<b>K28</b>	<b>Q26</b>	<b>R68</b>	<b>T44</b>	<b>R70</b>

URA2:

Posición	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Base	<u>A</u>	<u>I</u>	<u>G</u>	<u>C</u>	<u>G</u>	<u>I</u>	<u>C</u>	<u>I</u>	<u>C</u>
Restos de Contacto	<b>N32</b>	<b>C33</b>	<b>E30/R38</b>	<b>E40/R28</b>	<b>R42</b>	<b>Q26</b>	<b>Y68/K24</b>	<b>Q44</b>	<b>R70</b>

25 5. Especificidad de Secuencia de Reconocimiento

30

Las meganucleasas diseñadas racionalmente perfiladas anteriormente en el presente Ejemplo se clonaron, se expresaron en *E. coli* y se purificaron como en el Ejemplo 1. Cada meganucleasa purificada se mezcló después 1:1 con su compañero 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 pretendida para cada heterodímero de meganucleasa. Como se muestra en la Figura 3, cada heterodímero de meganucleasa diseñada racionalmente escinde su sitio de ADN pretendido.

LISTADO DE SECUENCIAS

SEQ ID NO: 1 (I-Crel de tipo silvestre, n.º de referencia de GenBank PO5725)

5  
 1 MNTKYNKEFL LYLAFVDGD GSIIAQIKPN QSYKFKHQLS LAFQVTQKTQ RRWFLDKLVD  
 61 EIGVGYVRDR GSVSDYILSE IKPLHNFLTQ LQPFLKQK QANLVLKI IW RLPSAKESPD  
 121 KFLEVCTWVD QIAALNDSKT RKTSETVRA VLDSLSEKKK SSP

SEQ ID NO: 2 (secuencia de reconocimiento de I-Crel de tipo silvestre)  
 1 GAAACTGTCT CACGACGTTT TG

10  
 SEQ ID NO: 3 (secuencia de reconocimiento de I-Crel de tipo silvestre)  
 1 GAAAACGTCTG TGAGACAGTT CC

SEQ ID NO: 4 (secuencia de reconocimiento de I-Crel de tipo silvestre)  
 1 CAAACTGTCTG TGAGACAGTT TG

15  
 SEQ ID NO: 5 (secuencia de reconocimiento de I-Crel de tipo silvestre)  
 1 CAAACTGTCT CACGACAGTT TG

20  
 SEQ ID NO: 6 (I-Msol de tipo silvestre, n.º de referencia de GenBank AAL34387)

1 MTTKNTLQPT EAAYIAGFLD GDGSIYAKLI PRPDYKDIKY QVSLAISFIQ RKDKFPYLQD  
 61 IYDQLGKRGD LRKDRGDGIA DYTIIGSTHL SIILPDLVPY LRIKKKQANR ILHIINLYPQ  
 121 AQKNPSKFLD LVKIVDDVQN LNKRADELKS TNYDRLLLEEF LKAGKIESSP

25  
 SEQ ID NO: 7 (I-Msol de tipo silvestre, secuencia de reconocimiento)  
 1 CAGAACGTCTG TGAGACAGTT CC

SEQ ID NO: 8 (I-Msol de tipo silvestre, secuencia de reconocimiento)  
 1 GGAACGTCT CACGACGTTT TG

30  
 SEQ ID NO: 9 (I-Scel de tipo silvestre, n.º de referencia de GenBank CAA09843)

1 MKNIKKNQVM NLGPNSKLLK EYKSQLELN IEQFEAGIGL ILGDAYIRSR DEGKTYCMQF  
 61 EWKNKAYMDH VCLLYDQWVL SPPHKKERVN HLGNLVITWG AQTFKHQAFN KLANLFIVNN  
 121 KKTIPNNLVE NYLTPMSLAY WFMDDGGKWD YNKNSTNKS I VLNTQSFTFE EVEYLVKGLR  
 181 NKFQNLNVCYVK INKNKPIIYI DSMSYLIFYN LIKPYLIPQM MYKLPNTISS ETEFLK

35  
 SEQ ID NO: 10 (I-Scel de tipo silvestre, secuencia de reconocimiento)  
 1TTACCCTGTT ATCCCTAG

SEQ ID NO: 11 (I-Scel de tipo silvestre, secuencia de reconocimiento)  
 1 CTAGGGATAA CAGGGTAA

40  
 SEQ ID NO: 12 (I-Ceul de tipo silvestre, n.º de referencia de GenBank P32761)

1 MSNFILKPGE KLPQDKLEEL KKINDAVKKT KNFSKYLIDL RKLQIDEVQ VTSESKLFLA  
 61 GFLEGEASLN ISTKKLATSK FGLVVDPEFN VTQHVNGVKV LYLALFVFKT GRIRHKSGSN  
 121 ATLVLTIDNR QSLEEKVIPF YEQYVAFSS PEKVVRVANF KALLELFNND AHQDLEQLVN  
 181 KILPIWDQMR KQQQSNEG F PNLEAAQDFA RNYKKGKIK

45  
 SEQ ID NO: 13 (I-Ceul de tipo silvestre, secuencia de reconocimiento)  
 1 ATAACGGTCC TAAGGTAGCG AA

SEQ ID NO: 14 (I-Ceul de tipo silvestre, secuencia de reconocimiento)  
 1 TTCGCTACCT TAGGACCGTT AT

50  
 SEQ ID NO: 15 (gen TAT de VIH-1, secuencia parcial)  
 1 GAAGAGCTCA TCAGAACAGT CA



## ES 2 582 091 T3

- 5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65
- SEQ ID NO: 16 (semisito de secuencia de reconocimiento de TAT 1 diseñado racionalmente)  
1 GAAGAGCTC
- SEQ ID NO: 17 (semisito de secuencia de reconocimiento de TAT2 diseñado racionalmente)  
1 TGA CTGTTC
- SEQ ID NO: 18 (semisito de secuencia de reconocimiento de CCR1 diseñado racionalmente)  
1 AACCTCTC
- SEQ ID NO: 19 (semisito de secuencia de reconocimiento de BRP2 diseñado racionalmente)  
1CTCCGGGTC
- SEQ ID NO: 20 (semisito de secuencia de reconocimiento de LAM1 diseñado racionalmente)  
1TGCGGTGTC
- SEQ ID NO: 21 (semisito de secuencia de reconocimiento de LAM2 diseñado racionalmente)  
1 CAGGCTGTC
- SEQ ID NO: 22 (secuencia de reconocimiento de LAM1/LAM2 en el gen p05 de bacteriófago  $\lambda$ )  
1 TGCGGTGTCC GGCGACAGCC TG
- SEQ ID NO: 23 (secuencia de reconocimiento potencial en el gen FGFR3 humano)  
1 CTGGGAGTCT CAGGACAGCC TG
- SEQ ID NO: 24 (secuencia de reconocimiento potencial en promotor de la hormona del crecimiento humana)  
1 CCAGGTGTCT CTGGACTCCT CC
- SEQ ID NO: 25 (secuencia de reconocimiento potencial en el alelo  $\Delta$ F508 del gen CFTR humano)  
1 GAAAATATCA TTGGTGTTC CT
- SEQ ID NO: 26 (secuencia de reconocimiento potencial el gen CCR5 humano)  
1 AACCTCTCC AGTGAGATGC CT
- SEQ ID NO: 27 (secuencia de reconocimiento potencial en la región 3' UTR del gen de DM cinasa humana)  
1 GACCTCGTCC TCCGACTCGC TG
- SEQ ID NO: 28 (secuencia de reconocimiento potencial en el gen UL36 del Herpes Simplex Virus-1 y el Herpes Simplex Virus-2)  
1 CTCGATGTCTG GACGACACGG CA
- SEQ ID NO: 29 (secuencia de reconocimiento potencial en el genoma de *Bacillus anthracis*)  
1 ACAAGTGTCT ATGGACAGTT TA
- SEQ ID NO: 30 (secuencia de reconocimiento potencial en el gen gp009 del virus Variola (varicela))  
1 AAAACTGTCA AATGACATCG CA
- SEQ ID NO: 31 (secuencia de reconocimiento potencial en el gen BALF2 del virus de Epstein-Barr)  
1 CGGGGTCTCG TGCGAGGCCT CC
- SEQ ID NO: 32 (secuencia de reconocimiento potencial en el gen GL2 de *Arabidopsis thaliana*)  
1 CACTAACTCG TATGAGTCGG TG
- SEQ ID NO: 33 (secuencia de reconocimiento potencial en el gen BP1 de *Arabidopsis thaliana*)  
1 TGCCTCCTCT AGAGACCCGG AG
- SEQ ID NO: 34 (secuencia de reconocimiento potencial en el gen de quelastasa de magnesio de *Nicotiana tabacum*)  
1 TAAAATCTCT AAGGTCTGTG CA
- SEQ ID NO: 35 (secuencia de reconocimiento potencial en el gen CYP82E4 de *Nicotiana tabacum*)  
1 CAAGAATTCA AGCGAGCATT AA
- SEQ ID NO: 36 (secuencia de reconocimiento potencial en el gen URA3 de *Saccharomyces cerevisiae*)  
1 TTAGATGACA AGGGAGACGC AT

LISTADO DE SECUENCIAS

- <110> Precision Biosciences
- 5 <120> MEGANUCLEASAS DISEÑADAS RACIONALMENTE CON ESPECIFICIDAD DE SECUENCIA Y AFINIDAD DE UNIÓN A ADN ALTERADAS
- <130> 111-046T10
- 10 <150> EP 06 826 293.0  
<151> 18-10-2006
- <150> PCT/US2006/040919  
<151> 18-10-2006
- 15 <150> 60/727.512  
<151> 18-10-2005
- <160> 37
- 20 <170> PatentIn Ver. 3.3
- <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 <212> ADN
- <213> *Chlamydomonas reinhardtii*

# ES 2 582 091 T3

<400>2  
gaaactgtct cacgacgttt tg      22

5      <210> 3  
      <211> 22  
      <212> ADN  
      <213> *Chlamydomonas reinhardtii*

10     <400> 3  
      caaaacgtcg tgagacagtt tc      22

15     <210> 4  
      <211> 22  
      <212> ADN  
      <213> *Chlamydomonas reinhardtii*

20     <400> 4  
      caaactgtcg tgagacagtt tg      22

25     <210>5  
      <211> 22  
      <212> ADN  
      <213> *Chlamydomonas reinhardtii*

30     <400> 5  
      caaactgtct cacgacagtt tg      22

      <210> 6  
      <211> 170  
      <212> PRT  
      <213> *Monomastix sp.*

      <400>6

ES 2 582 091 T3

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

5 <210>7  
 <211> 22  
 <212> ADN  
 <213> *Monomastix sp.*

10 <400> 7  
 cagaacgtcg tgagacagtt cc 22

15 <210> 8  
 <211> 22  
 <212> ADN  
 <213> *Monomastix sp.*

20 <400> 8  
 ggaactgtct cacgacgttc tg 22

25 <210> 9  
 <211> 235  
 <212> PRT  
 <213> *Saccharomyces cerevisiae*

<400>9

ES 2 582 091 T3

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

<210> 10  
 <211> 18  
 <212> ADN  
 <213> *Saccharomyces cerevisiae*

<400> 10  
 ttaccctgtt atccctag 18

<210> 11  
 <211> 18  
 <212> ADN  
 <213> *Saccharomyces cerevisiae*

<400> 11  
 ctaggataa cagggtaa 18

<210> 12

ES 2 582 091 T3

<211> 218  
 <212> PRT  
 <213> *Chlamydomonas eugametos*

5 <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

10 <210> 13  
 <211> 22  
 <212> ADN  
 <213> *Chlamydomonas eugametos*

15 <400> 13  
 ataacggtcc taaggtagcg aa 22

<210> 14  
 <211> 22  
 <212> ADN  
 20 <213> *Chlamydomonas eugametos*

# ES 2 582 091 T3

<400> 14  
ttcgctacct taggaccggt at 22

5 <210> 15  
<211> 22  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

<400> 15  
gaagagctca tcagaacagt ca 22

15 <210> 16  
<211> 9  
<212> ADN  
<213> Secuencia artificial

20 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

<400> 16  
gaagagctc 9

25 <210> 17  
<211> 9  
<212> ADN  
<213> Secuencia artificial

30 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

<400> 17  
tgactgttc 9

35 <210> 18  
<211> 9  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

45 <400> 18  
aaccctctc 9

<210> 19  
<211> 9

50 <212> ADN  
<213> Secuencia artificial

<220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

55 <400> 19  
ctccgggtc 9

<210> 20  
<211> 9  
<212> ADN  
<213> Secuencia artificial

60 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

65 <220>  
<223> Descripción de Secuencia artificial: Oligonucleótido sintético

# ES 2 582 091 T3

<400> 20  
 tgcggtgtc 9

5 <210> 21  
 <211> 9  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> Descripción de Secuencia artificial: Oligonucleótido sintético

<400> 21  
 caggctgtc 9

15 <210> 22  
 <211> 22  
 <212> ADN  
 <213> Bacteriophage lamda-p05

20 <400> 22  
 tgcggtgtcc ggcgacagcc tg 22

25 <210> 23  
 <211> 22  
 <212> ADN  
 <213> *Homo sapiens*

30 <400> 23  
 ctgggagtct caggacagcc tg 22

35 <210> 24  
 <211> 22  
 <212> ADN  
 <213> *Homo sapiens*

<400> 24  
 ccaggtgtct ctggactcct cc 22

40 <210> 25  
 <211> 22  
 <212> ADN  
 <213> *Homo sapiens*

45 <400> 25  
 gaaaatatca ttggtgttc ct 22

50 <210> 26  
 <211> 22  
 <212> ADN  
 <213> *Homo sapiens*

<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 Herpes Simplex 2



<400> 28  
 ctcgatgtcg gacgacacgg ca 22

5 <210> 29  
 <211> 22  
 <212> ADN  
 <213> *Bacillus anthracis*

10 <400> 29  
 acaagtgtct atggacagtt ta 22

15 <210> 30  
 <211> 22  
 <212> ADN  
 <213> Virus Variola

20 <400> 30  
 aaaactgtca aatgacatcg ca 22

25 <210> 31  
 <211> 22  
 <212> ADN  
 <213> Virus de *Epstein-Barr*

30 <400> 31  
 cggggtctcg tgcgaggcct cc 22

35 <210> 32  
 <211> 22  
 <212> ADN  
 <213> *Arabidopsis thaliana*

40 <400> 32  
 cactaactcg tatgagtcgg tg 22

45 <210> 33  
 <211> 22  
 <212> ADN  
 <213> *Arabidopsis thaliana*

50 <400> 33  
 tgcctcctct agagaccgg ag 22

55 <210> 34  
 <211> 22  
 <212> ADN  
 <213> *Nicotiana tabacum*

60 <400> 34  
 taaaatctct aaggtctgtg ca 22

65 <210> 35  
 <211> 22  
 <212> ADN  
 <213> *Nicotiana tabacum*

70 <400> 35  
 caagaattca agcgagcatt aa 22

75 <210> 36  
 <211> 22  
 <212> ADN  
 <213> *Saccharomyces cerevisiae*

80 <400> 36  
 ttagatgaca agggagacgc at 22

ES 2 582 091 T3

<210> 37  
<211> 9  
<212> PRT  
<213> Secuencia artificial

5

<220>  
<223> Descripción de Secuencia artificial: Motivo peptídico sintético

10

<400> 37

**Leu Ala Gly Leu Ile Asp Ala Asp Gly**  
**1 5**

**REIVINDICACIONES**

1. Una meganucleasa recombinante que comprende:

5 un polipéptido que tiene una similitud de al menos el 85 % con los restos 2-153 de la meganucleasa I-Crel de SEQ ID NO: 1;  
 en la que la afinidad de unión a ADN se ha aumentado mediante al menos una modificación de un resto que es parte de la superficie de contacto de la estructura de la meganucleasa I-Crel de SEQ ID NO: 1, correspondiendo dicha modificación a una sustitución seleccionada entre el grupo que consiste en:  
 10 sustitución de E80 por H, N, Q, S, T, K o R.

2. Un método para producir una célula eucariota genéticamente modificada que incluye una secuencia exógena insertada en un cromosoma de dicha célula eucariota, que comprende:

15 transfectar una célula eucariota con uno o más ácidos nucleicos que incluyen  
 (i) una primera secuencia de ácido nucleico que codifica una meganucleasa, y  
 (ii) una segunda secuencia de ácido nucleico que incluye dicha secuencia exógena;  
 20 en donde dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia exógena se inserta en dicho cromosoma en dicho sitio de escisión;  
 en donde dicha meganucleasa es una meganucleasa recombinante de la reivindicación 1; y  
 en donde dicha célula eucariota no es una célula madre humana.

3. Un método para producir una célula eucariota genéticamente modificada que incluye una secuencia exógena insertada en un cromosoma de dicha célula eucariota, que comprende:

30 introducir una proteína meganucleasa en una célula eucariota; y  
 transfectar dicha célula eucariota con un ácido nucleico que incluye dicha secuencia exógena;  
 en donde dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia exógena se inserta en dicho cromosoma en dicho sitio de escisión;  
 en donde dicha meganucleasa es una meganucleasa recombinante de la reivindicación 1; y  
 en donde dicha célula eucariota no es una célula madre humana.

4. Un método para producir una célula eucariota genéticamente modificada interrumpiendo una secuencia diana en un cromosoma de dicha célula eucariota, que comprende:

40 transfectar una célula eucariota con un ácido nucleico que codifica una meganucleasa;  
 en donde dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia exógena se interrumpe mediante unión de extremos no homólogos en dicho sitio de escisión;  
 en donde dicha meganucleasa es una meganucleasa recombinante de la reivindicación 1; y  
 en donde dicha célula eucariota no es una célula madre humana.

5. Un método para producir un organismo no humano genéticamente modificado que comprende:

45 producir una célula eucariota genéticamente modificada de acuerdo con el método de una cualquiera de las reivindicaciones 2-4; y  
 50 cultivar dicha célula eucariota genéticamente modificada para producir dicho organismo genéticamente modificado.

6. Un método como en la reivindicación 5, en el que:

55 dicha célula eucariota se selecciona entre el grupo que consiste en un gameto, un cigoto, una célula de blastocisto, una célula madre embrionaria y un protoplasto.

7. Una meganucleasa recombinante de la reivindicación 1 para su uso como medicamento.

8. Un ácido nucleico aislado que codifica una meganucleasa recombinante de la reivindicación 1 para su uso como medicamento.

9. Un método como en una cualquiera de las reivindicaciones 2 o 3, en el que:

65 dicho ácido nucleico que incluye dicha secuencia exógena comprende además secuencias homólogas a las secuencias que flanquean a dicho sitio de escisión y dicha secuencia exógena se inserta en sitio dicho de escisión mediante recombinación de homólogos; o dicho ácido nucleico que incluye dicha secuencia exógena carece de

homología sustancial respecto de dicho sitio de escisión y dicha secuencia exógena se inserta en dicho cromosoma mediante unión de extremos no homólogos.

5 10. Un medicamento como en una cualquiera de las reivindicaciones 7 u 8, en donde dicho medicamento comprende además:

10 un ácido nucleico aislado que incluye una secuencia exógena que se va a insertar en un sitio de escisión en un cromosoma y secuencias que flanquean a dicha secuencia exógena que son homólogas a las secuencias que flanquean dicho sitio de escisión,  
10 en donde dicho sitio de escisión se va a escindir mediante dicha meganucleasa recombinante.

15 11. Uso de una meganucleasa recombinante de la reivindicación 1 en la fabricación de un medicamento para tratar una infección por un patógeno vírico en un hospedador eucariota interrumpiendo una secuencia diana en un genoma de dicho patógeno vírico.

15 12. Uso de una meganucleasa recombinante de la reivindicación 1 en la fabricación de un medicamento para tratar una infección por un patógeno procariota en un hospedador eucariota interrumpiendo una secuencia diana en un genoma de dicho patógeno procariota.

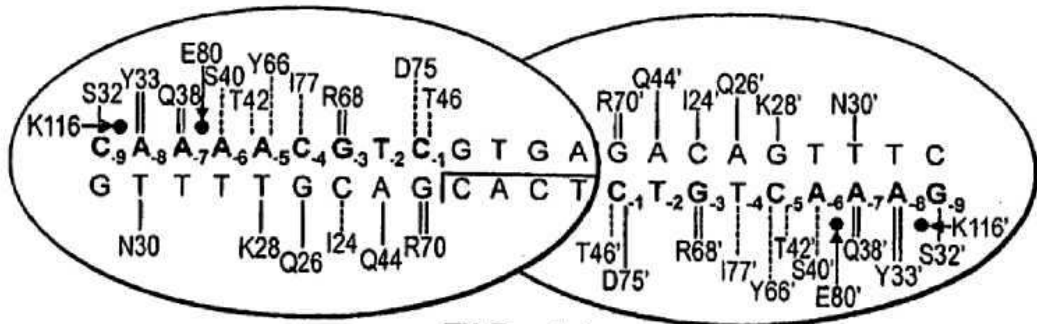


FIG. 1A

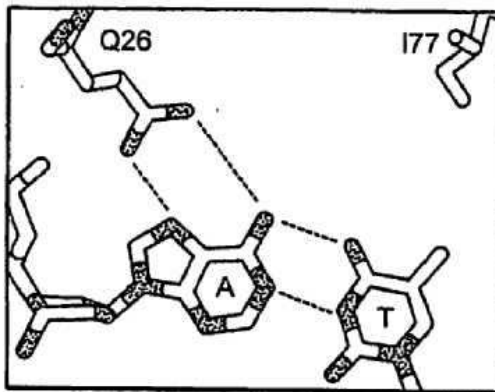


FIG. 1B

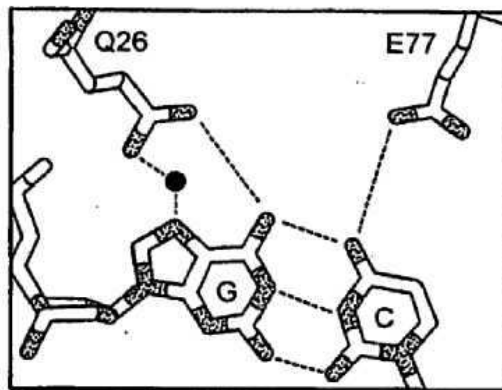


FIG. 1C

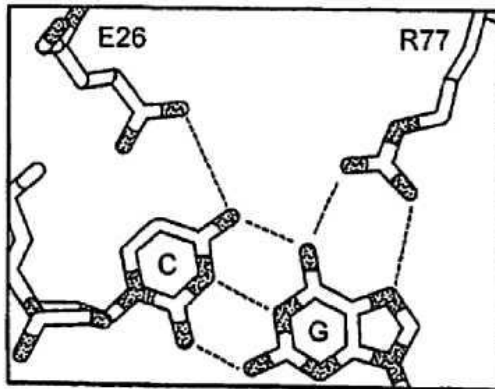


FIG. 1D

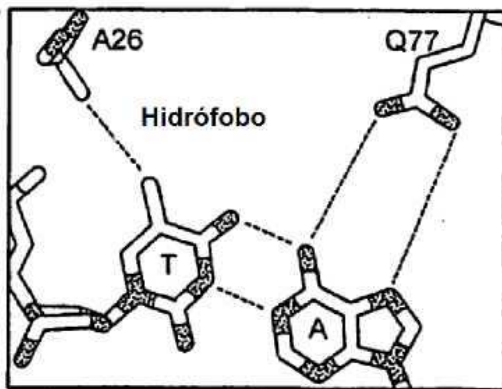


FIG. 1E

WT **CAA**ACTGTCGTGAGACAGTTG

CF1/CF2 G**AAA**A**TAT**CA**TTG**GTGTT**TC**CT

MYD1/MYD2 G**A**CC**TC**G**TC**CT**C**C**GAC**TCG**CTG**

CCR1/CCR2 A**A**CC**CT**C**TC**CA**G**T**GAG**A**TGC**CT

ACH1/ACH2 **C**TGGG**AG**T**CT**CA**G****GAC**AG**CCTG**

TAT1/TAT2 G**AA**GAG**CT**CA**T**CA**GAA**CA**GT**CA

HSV1/HSV2 **C**TCGA**TG**TC**G**GAC**GACA**CGGCA

LAM1/LAM2 TGCGG**TG**TC**CG**G**C****GAC**AG**CCTG**

POX1/POX2 A**AA**ACTG**TC**AAAT**GACA**TCGCA

URA1/URA2 TT**AGA****TG**AC**AA**G**G****GAG**A**CG**CAT

GLA1/GLA2 **CA**CTAAC**TC**GT**AT****GAG**TCGG**TG**

BRP1/BRP2 TG**CCT**CC**TC**TA**GAG**AC**CC**GGAG**G**

FIG. 2A

Enzima	Sitio													
	WT	CF		CCR		TAT		LAM		URA		BRP		
	WT	MYD	ACH	HSV	POX	GLA								
WT	100	0	0	0	0	0	47	0	100	100	0	0		Tipo silvestre
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 variola
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

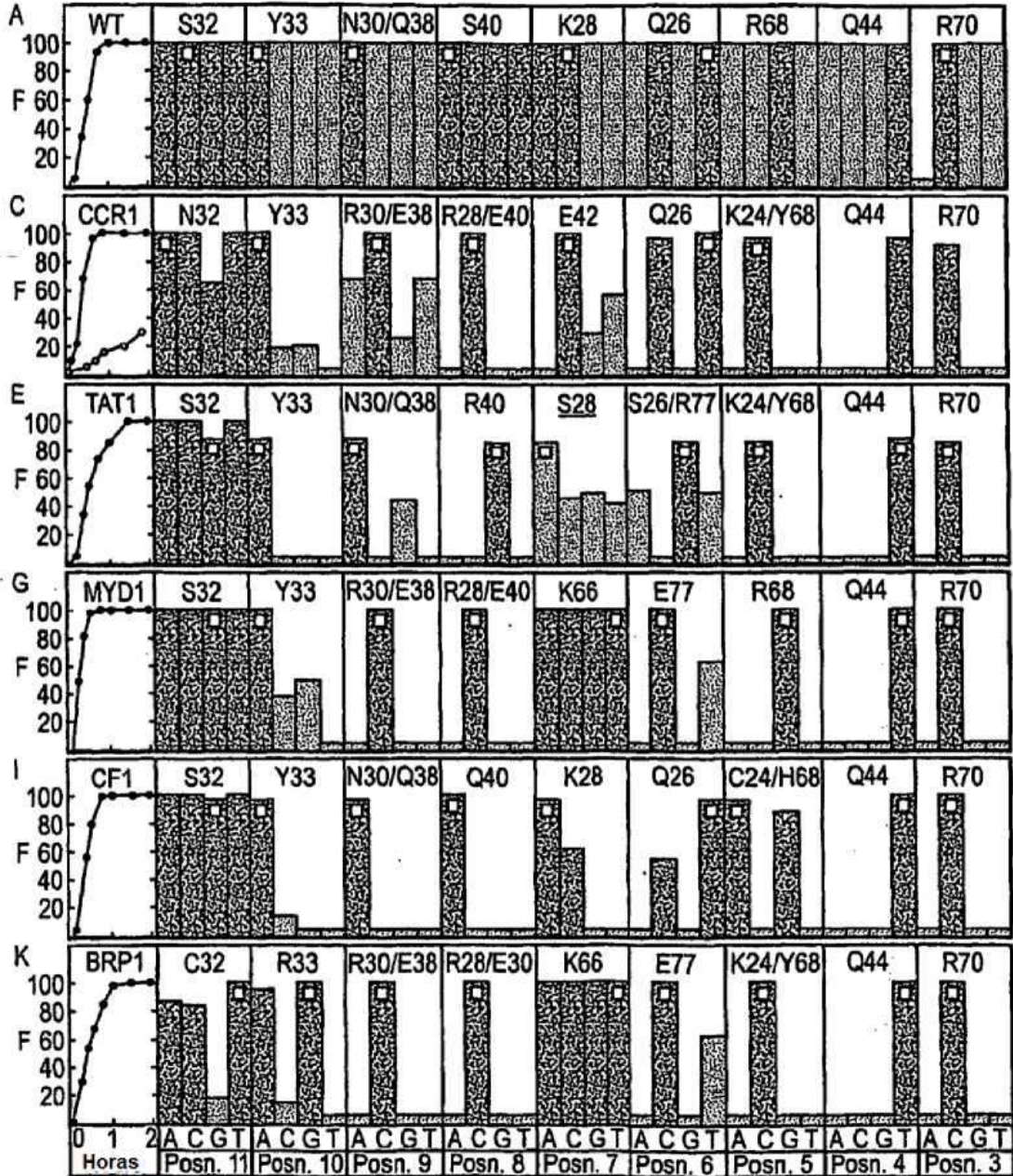


FIG. 3A

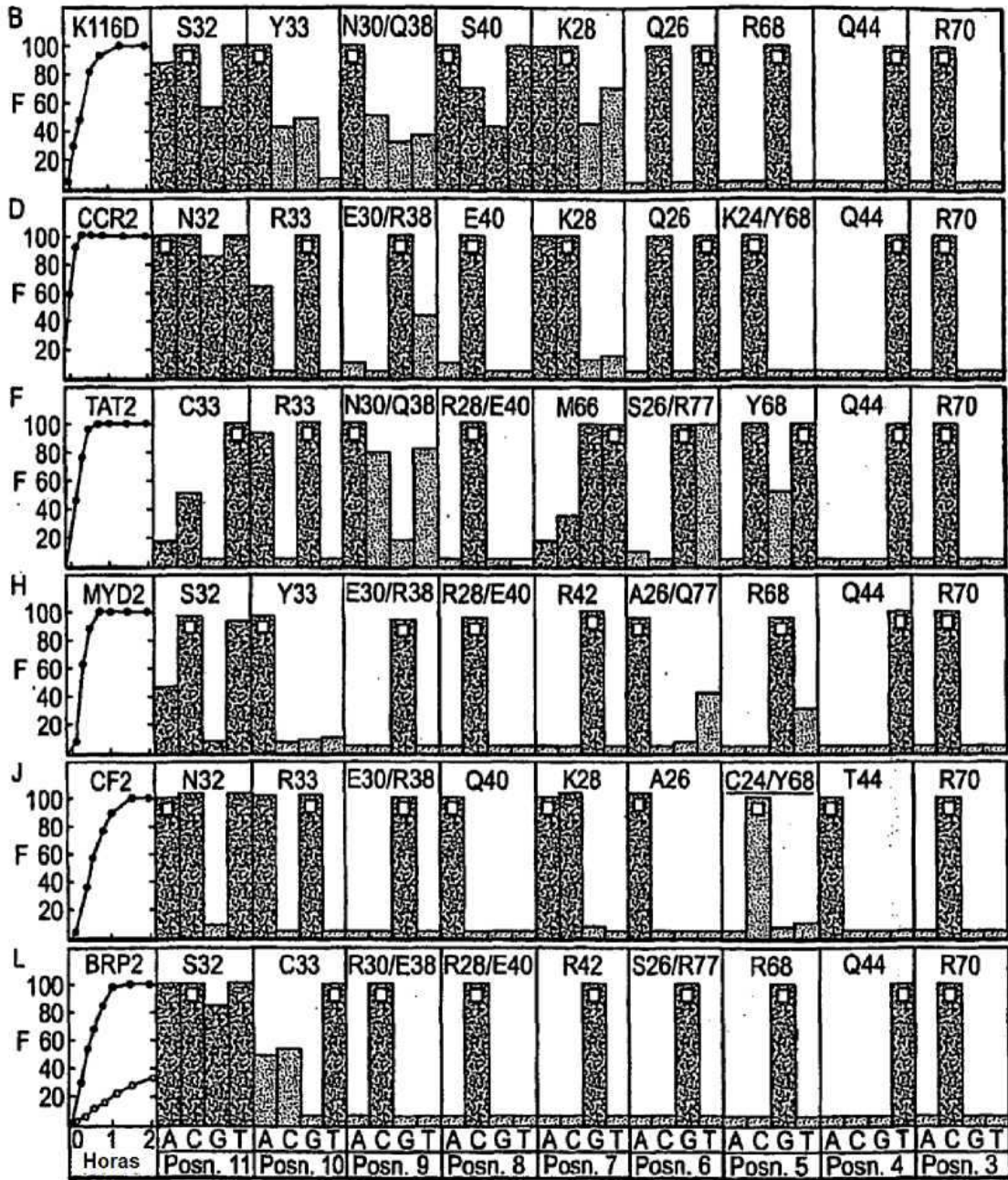


FIG. 3B