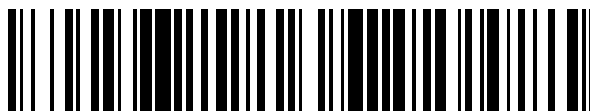


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 732 735**

51 Int. Cl.:

C12N 9/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.10.2008** **E 16163925 (7)**

97 Fecha y número de publicación de la concesión europea: **10.04.2019** **EP 3098309**

54 Título: **Meganucleasas monocatenarias diseñadas racionalmente con secuencias de reconocimiento no palindrómicas**

30 Prioridad:

31.10.2007 US 1247 P

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

25.11.2019

73 Titular/es:

**PRECISION BIOSCIENCES, INC. (100.0%)
302 East Pettigrew Street, Dibrell Building, Suite
A-100
Durham, NC 27701, US**

72 Inventor/es:

**SMITH, JAMES, JEFFERSON y
JANTZ, DEREK**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 732 735 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Meganucleasas monocatenarias diseñadas racionalmente con secuencias de reconocimiento no palindrómicas

5 **Campo de la invención**

La invención se refiere al campo de la biología molecular y tecnología de ácidos nucleicos recombinantes. En particular, la invención se refiere a meganucleasas de origen no natural diseñadas racionalmente en las que un par de subunidades enzimáticas que tienen especificidad para semisitios de secuencia de reconocimiento diferentes se unen en un polipéptido individual para formar un heterodímero funcional con una secuencia de reconocimiento no palindrómica. La invención también se refiere a procedimientos para producir dichas meganucleasas y procedimientos para producir ácidos nucleicos recombinantes y organismos que usan dichas meganucleasas.

15 **Antecedentes de la invención**

La ingeniería genómica requiere la capacidad para insertar, suprimir, sustituir y manipular de cualquier 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 principal en la terapia génica, agrotecnología, y biología sintética (Porteus y col. (2005), Nat. Biotechnol. 23: 967-73; Tzfira y col. (2005), Trends Biotechnol. 23: 567-9; McDaniel y col. (2005), Curr. Opin. Biotechnol. 16: 476-83). Un procedimiento habitual para insertar o modificar una secuencia de ADN implica introducir una secuencia de ADN transgénico flanqueada por secuencias homólogas de la diana genómica y seleccionar o explorar con respecto a un acontecimiento recombinante homólogo exitoso. La recombinación con el ADN transgénico se produce pocas veces pero puede estimularse por una rotura bicatenaria en el ADN genómico en el sitio diana. Se ha empleado numerosos procedimientos para crear roturas bicatenarias de ADN, incluyendo irradiación y tratamientos químicos. Aunque estos procedimientos estimulan eficazmente la recombinación, las roturas bicatenarias se dispersan aleatoriamente en el genoma, lo que puede ser altamente mutagénico y tóxico. En la actualidad, la incapacidad para dirigir modificaciones génicas a sitios únicos dentro de un fondo cromosómico es un impedimento importante para la ingeniería genómica exitosa.

Un enfoque para conseguir este objetivo es estimular la recombinación homóloga en una rotura bicatenaria en un locus diana usando una nucleasa con especificidad por una secuencia que es suficientemente grande para estar presente solamente en un único sitio dentro del genoma (véase, por ejemplo, Porteus y col. (2005), Nat. Biotechnol. 23: 967-73). La eficacia de esta estrategia se ha demostrado en diversos organismos usando fusiones quiméricas entre un dominio de unión a 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: 438-46; Wright y col. (2005), Plant J. 44: 693-705; Urnov y col. (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 que resulta de la infrarregulación del dominio nucleasa y frecuentemente se escinden en sitios no pretendidos (Smith y col. (2000), Nucleic Acids Res. 28: 3361-9). Dicha escisión no pretendida puede provocar mutaciones y toxicidad en el organismo tratado (Porteus y col. (2005), Nat. Biotechnol. 23: 967-73).

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 pueden proporcionar una alternativa de ingeniería genómica menos tóxica. Dichas "meganucleasas" o "endonucleasas buscadoras" se asocian con frecuencia con elementos de ADN parasitarios, tales como intrones de autocorte y empalme de grupo I e inteínas. Promueven de forma natural la recombinación homóloga o inserción génica en localizaciones específicas en el genoma huésped produciendo una rotura bicatenaria en el cromosoma, que recluta la maquinaria de reparación de ADN celular (Stoddard (2006), Q. Rev. Biophys. 38: 49-95). Las meganucleasas se agrupan 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 por tener una o dos copias del motivo LAGLIDADG conservado (véase Chevalier y col. (2001), Nucleic Acids Res. 29(18): 3757-3774). Las meganucleasas LAGLIDADG con una única copia del motivo LAGLIDADG ("meganucleasas mono-LAGLIDADG") forman homodímeros, mientras que los miembros con dos copias del motivo LAGLIDADG ("meganucleasas di-LAGLIDADG") se encuentran como monómeros. Las meganucleasas mono-LAGLIDADG tales como I-Crel, I-Ceul e I-Msol reconocen y escinden sitios de ADN que son palindrómicos o pseudopalindrómicos, mientras que las meganucleasas di-LAGLIDADG tales como I-Scel, I-Anil e I-Dmol reconocen generalmente sitios de ADN que no son palindrómicos (Stoddard (2006), Q. Rev. Biophys. 38: 49-95).

Se han usado meganucleasas naturales de la familia de LAGLIDADG para promover eficazmente la modificación genómica específica de sitio en plantas, levadura, *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 meganucleasas (Monnat y col. (1999), Biochem. Biophys. Res. Commun. 255: 88-93) o a genomas previamente modificados por ingeniería genética en los que se ha introducido una secuencia de reconocimiento (Rouet y col. (1994), Mol. Cell. Biol. 14: 8096-106; Chilton y col. (2003), Plant Physiol. 133: 956-65; Puchta y col. (1996), Proc. Natl. Acad. Sci. USA

93: 5055-60; Rong y col. (2002), *Genes Dev.* 16: 1568-81; Gouble y col. (2006), *J. Gene Med.* 8(5):616-622).

La implementación sistemática de modificación génica estimulada por nucleasa 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 adaptar meganucleasas para estimular modificaciones génicas en sitios médica o biotecnológicamente relevantes (Porteus y col. (2005), *Nat. Biotechnol.* 23: 967-73; Sussman y col. (2004), *J. Mol. Biol.* 342: 31-41; Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62).

I-Crel es un miembro de la familia LAGLIDADG que reconoce y corta una secuencia de reconocimiento de 22 pares de bases en el cromosoma del cloroplasto y que presenta una diana atractiva para rediseño de meganucleasas. La enzima natural 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 modificar la preferencia de sitio de escisión de I-Crel natural (Sussman y col. (2004), *J. Mol. Biol.* 342: 31-41; Chames y col. (2005), *Nucleic Acids Res.* 33: e178; Seligman y col. (2002), *Nucleic Acids Res.* 30: 3870-9, Arnould y col. (2006), *J. Mol. Biol.* 355: 443-58, Rosen y col. (2006), *Nucleic Acids Res.* 34: 4791-4800, Arnould y col. (2007), *J. Mol. Biol.* 371: 49-65, documentos WO 2008/010009, WO 2007/093918, WO 2007/093836, WO 2006/097784, WO 2008/059317, WO 2008/059382, WO 2008/102198, WO 2007/060495, WO 2007/049156, WO 2006/097853, WO 2004/067736). Más recientemente, se describió un procedimiento de diseño racional de meganucleasas mono-LAGLIDADG que es capaz de rediseñar de forma exhaustiva I-Crel y otras de tales meganucleasas para dirigir a sitios de ADN ampliamente divergentes, incluyendo sitios en genomas de mamífero, levadura, planta, bacterianos y virales (documento WO 2007/047859).

Una limitación importante para usar meganucleasas mono-LAGLIDADG tales como I-Crel para la mayoría de aplicaciones de ingeniería genética es el hecho de que estas enzimas se dirigen de forma natural a sitio de reconocimiento de ADN palindrómicos. Dichos sitios de ADN palindrómicos largos (10-40 pb) son escasos en la naturaleza y es improbable que aparezcan por casualidad en un sitio de ADN de interés. Para dirigirse a un sitio de ADN no palindrómico con una meganucleasa mono-LAGLIDADG, se puede producir un par de monómeros que reconocen los dos semisitios diferentes y que heterodimerizan para formar una meganucleasa que escinde el sitio no palindrómico deseado. Puede conseguirse heterodimerización mediante coexpresión de un par de monómeros de meganucleasa en una célula huésped o mezclando un par de meganucleasas homodiméricas purificadas *in vitro* y permitiendo que las subunidades se vuelvan a asociar en heterodímeros (Smith y col. (2006), *Nuc. Acids Res.* 34:149-157; Chames y col. (2005), *Nucleic Acids Res.* 33:178-186; documentos WO 2007/047859, WO 2006/097854, WO 2007/057781, WO 2007/049095, WO 2007/034262). Ambos enfoques padecen dos limitaciones primarias: (1) requieren la expresión de dos genes de meganucleasa para producir la especie heterodimérica deseada (que complica el suministro génico y aplicaciones *in vivo*) y (2) el resultado es una mezcla de aproximadamente 25 % del primer homodímero, 50 % del heterodímero y 25 % del segundo homodímero, mientras que solamente se desea el heterodímero. Esta última limitación puede superarse en gran medida modificando por ingeniería genética las interfaces de dimerización de las dos meganucleasas para promover la heterodimerización frente a la homodimerización como se describe en los documentos WO 2007/047859, WO 2008/093249, WO 2008/093152, y Fajardo-Sanchez y col. (2008), *Nucleic Acids Res.* 36:2163-2173. Aún así, deben expresarse dos genes de meganucleasa y no se evita completamente la homodimerización.

Un enfoque alternativo a la formación de meganucleasas con sitios de reconocimiento no palindrómicos derivados de una o más meganucleasas mono-LAGLIDADG es la producción de un único polipéptido que comprende una fusión de las subunidades LAGLIDADG derivadas de dos meganucleasas. Pueden aplicarse dos procedimientos generales para producir dicha meganucleasa.

En el primer procedimiento, una de las dos subunidades LAGLIDADG de una meganucleasa di-LAGLIDADG pueden reemplazarse por una subunidad LAGLIDADG de una meganucleasa mono-LAGLIDADG. Este enfoque se demostró reemplazando la subunidad C terminal de la meganucleasa di-LAGLIDADG I-Dmol con una subunidad I-Crel (Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62; Chevalier y col. (2002), *Mol. Cell* 10: 895-905; documento WO 2003/078619). El resultado fue una meganucleasa híbrida I-Dmol/I-Crel que reconoció y escindió un sitio de ADN híbrido.

En el segundo procedimiento, puede unirse una par de subunidades mono-LAGLIDADG por un engarce peptídico para crear una "meganucleasa heterodimérica monocatenaria". Se ha presentado un intento de producir dicho derivado monocatenario de I-Crel (Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62; documento WO 2003/078619). Sin embargo, como se analiza en el presente documento así como en Fajardo-Sanchez y col. (2008), *Nucleic Acids Res.* 36: 2163-2173, hay ahora pruebas que sugieren que este procedimiento no produjo una meganucleasa heterodimérica monocatenaria en la que las subunidades I-Crel unidas covalentemente actuaban juntas para reconocer y escindir un sitio de reconocimiento no palindrómico.

Por lo tanto, sigue existiendo una necesidad en la técnica de procedimientos para la producción de meganucleasas heterodiméricas monocatenarias derivadas de enzimas mono-LAGLIDADG tales como I-Crel para reconocer y cortar sitios de ADN no palindrómicos.

Sumario de la invención

La invención se define por las reivindicaciones adjuntas.

5 La presente invención se basa, en parte, en el desarrollo de proteínas de fusión en las que un engarce peptídico une covalentemente dos subunidades de meganucleasa LAGLIDADG heterólogos para formar una “meganucleasa heterodimérica monocatenaria” o “meganucleasa monocatenaria”, en la que al menos la subunidad N terminal deriva de una meganucleasa monoLAGLIDADG y en la que las subunidades actúan juntas para unirse preferentemente a y
 10 escindir un sitio de reconocimiento de ADN no palindrómico que es un híbrido de los semisitios de reconocimiento de las dos subunidades. En particular, la invención puede usarse para modificar por ingeniería genética meganucleasas monocatenarias que reconocen secuencias de ADN no palindrómicas que las meganucleasas de origen natural no reconocen. La invención también proporciona procedimientos que usan dichas meganucleasas para producir ácidos nucleicos y organismos recombinantes utilizando las meganucleasas para provocar recombinación de una secuencia genética deseada en un número limitado de loci dentro del genoma del organismo para, entre otros, ingeniería
 15 genética, terapia génica, tratamiento de infecciones patógenas, y aplicaciones *in vitro* en diagnóstico e investigación.

Por lo tanto, se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades LAGLIDADG unidas covalentemente derivadas de una o más meganucleasas mono-LAGLIDADG que actúan juntas para reconocer y escindir un sitio de reconocimiento no palindrómico. La subunidad mono-LAGLIDADG puede
 20 derivar de una meganucleasa natural seleccionada de I-Crel, I-Msol e I-Ceul.

También se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades mono-LAGLIDADG en las que la subunidad N terminal deriva de una meganucleasa natural seleccionada de I-Crel, I-Msol e I-Ceul, y la subunidad C terminal también deriva de una meganucleasa natural seleccionada de I-Crel, I-Msol e I-Ceul, pero la subunidad N terminal deriva de una meganucleasa natural de una especie diferente que la
 25 subunidad C terminal.

También se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades LAGLIDADG en las que la subunidad N terminal deriva de una meganucleasa natural seleccionada de I-Crel, I-Msol e I-Ceul, y la subunidad C terminal deriva de una única subunidad LAGLIDADG de una meganucleasa di-
 30 LAGLIDADG natural seleccionada de I-Dmol, I-Scel e I-Anil.

Las meganucleasas mono-LAGLIDADG naturales incluyen, sin limitación, la meganucleasa I-Crel de SEQ ID NO: 1, la meganucleasa I-Msol de SEQ ID NO: 2, y la meganucleasa I-Ceul de SEQ ID NO: 3. Las meganucleasas di-
 35 LAGLIDADG naturales incluyen, sin limitación, la meganucleasa I-Dmol de SEQ ID NO: 4, la meganucleasa I-Scel de SEQ ID NO: 5, y la meganucleasa I-Anil de SEQ ID NO: 6.

Los dominios LAGLIDADG naturales incluyen, sin limitación, los restos 9-151 de la meganucleasa I-Crel natural de SEQ ID NO: 1; los restos 11-162 de la meganucleasa I-Msol natural de SEQ ID NO: 2; y los restos 55-210 de la
 40 meganucleasa I-Ceul natural de SEQ ID NO: 3; los restos 9-96 de la I-Dmol natural de SEQ ID NO: 4; los restos 105-178 de la I-Dmol natural de SEQ ID NO: 4; los restos 32-123 de la I-Scel natural de SEQ ID NO: 5; los restos 134-225 de la I-Scel natural de SEQ ID NO: 5; los restos 4-121 de la I-Anil natural de SEQ ID NO: 6; y los restos 136-254 de la I-Anil natural de SEQ ID NO: 6.

Las subunidades LAGLIDADG derivadas de una meganucleasa LAGLIDADG natural incluyen, sin limitación, subunidades que incluyen un dominio LAGLIDADG que tiene al menos 85 % de identidad de secuencia, u 85 %-100 % de identidad de secuencia, con uno cualquiera de los restos 9-151 de la meganucleasa I-Crel natural de SEQ ID NO: 1; los restos 11-162 de la meganucleasa I-Msol natural de SEQ ID NO: 2; y los restos 55-210 de la
 45 meganucleasa I-Ceul natural de SEQ ID NO: 3; los restos 9-96 de la I-Dmol natural de SEQ ID NO: 4; los restos 105-178 de la I-Dmol natural de SEQ ID NO: 4; los restos 32-123 de la I-Scel natural de SEQ ID NO: 5; los restos 134-225 de la I-Scel natural de SEQ ID NO: 5; los restos 4-121 de la I-Anil natural de SEQ ID NO: 6; y los restos 136-254 de la I-Anil natural de SEQ ID NO: 6.
 50

Las subunidades LAGLIDADG derivadas de una meganucleasa LAGLIDADG natural también incluyen, sin limitación, subunidades que comprenden cualquiera de las secuencias polipeptídicas anteriores en las que se han incluido una o más modificaciones de aminoácidos de acuerdo con los procedimientos de diseño racional de meganucleasas LAGLIDADG desvelados en el documento WO 2007/047859, así como otras variantes de meganucleasa de origen no natural conocidas en la técnica.
 55

También se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades LAGLIDADG derivadas de subunidades de LAGLIDADG de origen natural cada una de las cuales reconoce un semisitio de ADN natural seleccionado de SEQ ID NO: 7-30.
 60

También se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades LAGLIDADG modificadas por ingeniería genética con respecto a especificidad de unión a ADN cada una de las cuales reconoce un semisitio de ADN que difiere en al menos una base de un semisitio de ADN natural seleccionado
 65

de SEQ ID NO: 7-30.

También se desvelan meganucleasas monocatenarias recombinantes que comprenden un par de subunidades LAGLIDADG en las que una subunidad es natural y reconoce un semisitio de ADN natural seleccionado de SEQ ID NO: 7-30 y el otro está modificado por ingeniería genética con respecto a especificidad de unión a ADN y reconoce un sitio de ADN que difiere en al menos una base de un semisitio de ADN natural seleccionado de SEQ ID NO: 7-30.

En algunas realizaciones, el engarce polipeptídico que une las subunidades LAGLIDADG es un engarce flexible. En realizaciones particulares, el engarce puede incluir 23-40 restos, 25-31 restos, o cualquier número dentro de esos intervalos. En otras realizaciones, al menos el 50 %, o el 50 %-100 %, de los restos que forman el engarce son restos polares no cargados.

En otras realizaciones, el engarce polipeptídico que une las subunidades LAGLIDADG tiene una estructura secundaria estable. En realizaciones particulares, la estructura secundaria estable comprende al menos dos estructuras hélice α . En otras realizaciones particulares, la estructura secundaria estable comprende del extremo N terminal al C terminal un primer bucle, una primera hélice α , una primera vuelta, una segunda hélice α y un segundo bucle. En algunas realizaciones particulares, el engarce puede incluir 23-56 restos, o cualquier número dentro de ese intervalo.

En otro aspecto, se desvelan diversos procedimientos de uso para las meganucleasas monocatenarias descritas y posibilitadas en el presente documento. Estos procedimientos incluyen producir células y organismos modificados genéticamente, tratar enfermedades por terapia génica, tratar infecciones patógenas y usar las meganucleasas monocatenarias recombinantes para aplicaciones *in vitro* para diagnóstico e investigación.

Por lo tanto, en un aspecto, la invención proporciona procedimientos 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 está insertada en el cromosoma en el sitio de escisión por recombinación homóloga o unión de extremos no homólogos.

Como alternativa, en otro aspecto, la invención proporciona procedimientos 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 en el sitio de escisión por recombinación homóloga o unión de extremos no homólogos.

En otro aspecto, la invención proporciona procedimientos para producir una célula eucariota modificada genéticamente alterando 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 está alterada por unión de extremos no homólogos en el sitio de escisión.

En otro aspecto, la invención proporciona procedimientos para producir un organismo modificado genéticamente no humano produciendo una célula eucariota modificada genéticamente de acuerdo con los procedimientos descritos anteriormente, y cultivando la célula eucariota modificada genéticamente para producir el organismo modificado genéticamente. En estas realizaciones, la célula eucariota se selecciona de un gameto, un cigoto, una célula de blastocitos, una célula madre embrionaria y una célula de protoplasto.

En otro aspecto se desvelan procedimientos para tratar una enfermedad por terapia génica en una eucariota, transfectando al menos una célula del eucariota con uno o más ácidos nucleicos incluyendo (i) una primera secuencia de ácido nucleico que codifica una meganucleasa de la invención 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 desvelan procedimientos para tratar una enfermedad por terapia génica en un eucariota, introduciendo una proteína meganucleasa de la invención en al menos una células 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 a extremos no homólogos, y la inserción de la secuencia de interés proporciona terapia génica para la enfermedad.

En otro aspecto se desvelan procedimientos para tratar una enfermedad por terapia génica en un eucariota alterando una secuencia diana en un cromosoma del eucariota, transfectando al menos una célula del eucariota 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 altera por unión de extremos no homólogos en el sitio de escisión, en los que la alteración de la secuencia diana proporciona la terapia génica para la enfermedad.

En otro aspecto se desvelan procedimientos para tratar una infección patógena viral o procarionta en un huésped eucariota alterando una secuencia diana en un genoma del patógeno, transfectando al menos una célula infectada del huésped con un ácido nucleico que codifica una meganucleasa de la invención, en los que la meganucleasa produce un sitio de escisión en el genoma y la secuencia diana se altera por (1) unión de extremos no homólogos en el sitio de escisión o (2) recombinación homóloga con un segundo ácido nucleico, y en los que la alteración de la secuencia diana proporciona tratamiento para la infección.

Resultarán evidentes aspectos y realizaciones de la invención para un experto en la materia basándose en la siguiente descripción detallada de la invención.

Breve descripción del dibujo

La Figura 1 es un diagrama de los componentes estructurales de una realización de un engarce de la invención (Engarce 9) y los restos N terminal y C terminal de las subunidades de endonucleasa unidas por el revestimiento.

Descripción detallada de la invención

La invención se define por las reivindicaciones adjuntas.

1.1 Introducción

La presente invención se basa, en parte, en el desarrollo de proteínas de fusión en las que un engarce peptídico une covalentemente dos subunidades de meganucleasa LAGLIDADG heterólogas para formar una "meganucleasa heterodimérica monocatenaria" en la que las subunidades actúan juntas para unirse preferentemente a y escindir un sitio de reconocimiento de ADN no palindrómico que es un híbrido de los semisitios de reconocimiento de las dos subunidades. En particular, la invención puede usarse para modificar por ingeniería genética meganucleasas monocatenarias que reconocen secuencias de ADN no palindrómicas que las meganucleasas de origen natural no reconocen.

Este descubrimiento se ha usado, como se describe en detalle posteriormente, para unir meganucleasas mono-LAGLIDADG, que actúan de forma natural como homodímeros, en meganucleasas monocatenarias. Además, el descubrimiento se ha usado para unir meganucleasas mono-LAGLIDADG que se han vuelto a modificar por ingeniería genética con respecto a especificidad de reconocimiento de ADN en heterodímeros monocatenarios que reconocen y escinden secuencias de ADN que son un híbrido de los sitios palindrómicos reconocidos por los dos homodímeros de meganucleasa. La invención proporciona secuencias de engarce peptídico ilustrativas para unir subunidades de LAGLIDADG en polipéptidos individuales. Resulta importante que la invención proporciona un procedimiento general para la producción de secuencias de engarce y la selección de puntos de fusión para ligar diferentes subunidades LAGLIDADG para producir meganucleasas monocatenarias diseñadas racionalmente funcionales.

Se desvelan procedimientos que usan dichas meganucleasas para producir ácidos nucleicos, células y organismos recombinantes utilizando las meganucleasas para provocar recombinación de una secuencia genética deseada en un número limitado de loci dentro del genoma del organismo para, entre otros, ingeniería genética, terapia génica, tratamiento de infecciones patógenas y cáncer, y aplicaciones *in vitro* en diagnóstico e investigación.

Como un asunto general, se desvelan procedimientos para generar meganucleasas monocatenarias que comprenden dos subunidades LAGLIDADG en las que la subunidad N terminal deriva de una meganucleasa mono-LAGLIDADG natural tal como I-Crel, I-Msol o I-Ceul o una variante de la misma y la subunidad C terminal deriva de una meganucleasa mono-LAGLIDADG o uno de los dos dominios de una meganucleasa di-LAGLIDADG tal como I-Scel, I-Dmol, o I-Anil. El procedimiento es distinto de los descritos previamente (Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62; Chevalier y col. (2002), *Mol. Cell* 10:895-905; documento WO 2003/078619) porque requiere el uso de secuencias de engarce específicas y nuevas y puntos de fusión para producir meganucleasas monocatenarias recombinantes en las que la subunidad N terminal deriva de una meganucleasa mono-LAGLIDADG.

Como se describe en detalle posteriormente, el procedimiento para producir una meganucleasa monocatenaria recombinante incluye el uso de puntos de fusión definidos en las dos subunidades LAGLIDADG para unir así como el uso de secuencias de engarce definidas para unir las en un único polipéptido. Además, se proporciona un conjunto de reglas para identificar puntos de fusión no descritos de forma explícita en el presente documento así como para producir secuencias de engarce funcionales que no se describen de forma explícita en el presente documento.

Por lo tanto, en un aspecto, se desvelan procedimientos para producir meganucleasas LAGLIDADG monocatenarias recombinantes. En otro aspecto, se desvelan meganucleasas monocatenarias recombinantes que resultan de estos procedimientos. En otro aspecto, se desvelan procedimientos que usan dichas meganucleasas monocatenarias para

producir ácidos nucleicos, células y organismos recombinantes en los que se modifica una secuencia de ADN deseada o locus genético dentro del genoma de la célula u organismo mediante la inserción, delección, sustitución u otra manipulación de secuencias de ADN. En otro aspecto, se desvelan procedimientos para reducir la supervivencia de patógenos o células cancerosas usando meganucleasas monocatenarias que tienen secuencias de reconocimiento específicas de patógeno o específicas de cáncer.

1.2 Referencias y definiciones

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

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 de longitud. Las meganucleasas de origen natural pueden ser monoméricas (por ejemplo, I-SceI) o diméricas (por ejemplo, I-CreI). El término meganucleasa, como se usa en el presente documento, puede usarse para hacer referencia a meganucleasas monoméricas, meganucleasas diméricas, a los monómeros que se asocian para formar una meganucleasa dimérica, o a una meganucleasa monocatenaria recombinante de la invención. La expresión “endonucleasa buscadora” es sinónima del término “meganucleasa”.

Como se usa en el presente documento la expresión “meganucleasa LAGLIDADG” se refiere a meganucleasas que incluyen un único motivo LAGLIDADG, que son de forma natural diméricas, o a meganucleasas que incluyen dos motivos LAGLIDADG, que son de forma natural monoméricas. La expresión “meganucleasa mono-LAGLIDADG” se usa en el presente documento para hacer referencia a meganucleasas que incluyen un único motivo LAGLIDADG, y la expresión “meganucleasa di-LAGLIDADG” se usa en el presente documento para hacer referencia 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 y tiene actividad enzimática, y cada uno de los monómeros individuales de una meganucleasa mono-LAGLIDADG, puede denominarse una subunidad LAGLIDADG, o simplemente “subunidad”.

Como se usa en el presente documento, y en referencia a una secuencia peptídica, “final” se refiere al extremo C terminal y “comienzo” se refiere al extremo N terminal. Por lo tanto, por ejemplo, “el comienzo del motivo LAGLIDADG” se refiere al aminoácido más N terminal en la secuencia peptídica que comprende el motivo LAGLIDADG mientras que “el final del motivo LAGLIDADG” se refiere al aminoácido más C terminal en la secuencia peptídica que comprende el motivo LAGLIDADG.

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

Como se usa en el presente documento, con respecto a una proteína, el término “recombinante” significa tener 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 tener 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, PCR y técnicas de clonación de ADN; transfección, transformación u otras tecnologías de transferencia génica; recombinación homóloga; mutagénesis dirigida; y fusión génica. De acuerdo con esta definición, una proteína que tenga una secuencia de aminoácidos idéntica a una proteína de origen natural, pero producida por clonación y expresión en un huésped heterólogo, no se considera recombinante.

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

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”.

Como se usa en el presente documento, el término “natural” se refiere a cualquier forma de origen natural de una meganucleasa. El término “natural” no pretende significar la variante alélica más común de la enzima en la naturaleza sino, en su lugar, cualquiera variante alélica hallada en la naturaleza. Las meganucleasas naturales se distinguen de meganucleasas recombinantes o de origen no natural.

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

5 Como se usa en el presente documento, la expresión “secuencia de reconocimiento” se refiere a un par de semisitios que se unen y escinden por un dímero de meganucleasa mono-LAGLIDADG o un monómero de meganucleasa di-LAGLIDADG. Los dos semisitios pueden estar o no 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 semisitio de secuencia de reconocimiento de cada monómero abarca 9 pares de bases, y los dos semisitios están separados por cuatro pares de bases que no entran en contacto directamente por unión de la enzima sino que constituyen el propio sitio de escisión (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 abarcan normalmente 22 pares de bases, incluyendo dos semisitios de 9 pares de bases que flanquean un sitio de escisión de 4 pares de bases. En el caso de la meganucleasa I-SceI, que es un monómero de meganucleasa di-LAGLIDADG, la secuencia de reconocimiento es una secuencia no palindrómica de aproximadamente 18 pb, y no hay pares de bases centrales que no se reconocen específicamente. Por convención, una de las dos cadenas se denomina la cadena “sentido” y la otra la cadena “antisentido”, aunque ninguna de las cadenas puede codificar proteína.

20 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 la 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 puede estar degradado 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, una natural) en condiciones fisiológicas, o si la tasa de escisión de una secuencia de reconocimiento aumenta o se reduce en una cantidad biológicamente significativa (por ejemplo al menos 2x, o 2x-10x) en relación con una meganucleasa de referencia.

30 Como se usa en el presente documento, el término “palindrómico” se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de semisitios idénticos. Sin embargo, no es necesario que la secuencia palindrómica sea palindrómica con respecto a los pares de bases centrales que no están directamente en contacto por unión de la enzima (por ejemplo, los cuatro pares de bases centrales de un sitio de reconocimiento de I-Crel). En el caso de meganucleasas diméricas de origen natural, las secuencias de ADN palindrómicas se reconocen por homodímeros en los que dos monómeros entran en contacto con semisitios idénticos.

35 Como se usa en el presente documento, la expresión “seudopalindrómico” se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de semisitios palindrómicos de forma imperfecta o no idénticos. Además de los pares de bases centrales que no están directamente en contacto por unión de la enzima, la secuencia seudo palindrómica puede desviarse de una secuencia palindrómica entre los dos semisitios de reconocimiento en 1-3 pares de bases en cada uno de los dos semisitios. Las secuencias de ADN seudopalindrómicas son típicas de los sitios de ADN naturales reconocidos por meganucleasas homodiméricas naturales en las que dos monómeros enzimáticos idénticos entran en contacto con semisitios ligeramente diferentes.

40 Como se usa en el presente documento, el término “no palindrómico” se refiere a una secuencia de reconocimiento compuesta de dos semisitios no relacionados de una meganucleasa. En este caso, no es necesario que la secuencia no palindrómica sea palindrómica con respecto a los pares de bases centrales o 4 o más pares de bases en cada uno de los dos semisitios. Las secuencias de ADN no palindrómicas se reconocen por meganucleasas di-LAGLIDADG, meganucleasas mono-LAGLIDADG altamente degradadas (por ejemplo I-Ceul) o por heterodímeros de monómeros de meganucleasa mono-LAGLIDADG que reconocen semisitios no idénticos. En el último caso, una secuencia de reconocimiento no palindrómica puede denominarse “secuencia híbrida” debido a que el heterodímero de dos monómeros mono-LAGLIDADG diferentes, si están fusionados o no en un único polipéptido, escindirán una secuencia de reconocimiento que comprende un semisitio reconocido por cada monómero. Por lo tanto, la secuencia de reconocimiento heterodimérica es un híbrido de las dos secuencias de reconocimiento homodiméricas.

45 Como se usa en el presente documento, el término “engarce” se refiere a una secuencia peptídica exógena usada para unir dos subunidades LAGLIDADG en un único polipéptido. Un engarce puede tener una secuencia que se encuentra en proteínas naturales, o puede ser una secuencia artificial que no se encuentra en ninguna proteína natural. Un engarce puede ser flexible y carecer de estructura secundaria o puede tener propensión a formar una estructura tridimensional específica en condiciones fisiológicas.

50 Como se usa en el presente documento, la expresión “punto de fusión” se refiere al punto de unión entre una subunidad LAGLIDADG y un engarce. Específicamente, el “punto de fusión N terminal” es el último aminoácido (más C terminal) de la subunidad LAGLIDADG N terminal antes de la secuencia de engarce y el “punto de fusión C

terminal” es el primer aminoácido (más N terminal) de la subunidad de LAGLIDADG C terminal después de la secuencia de engarce.

5 Como se usa en el presente documento, la expresión “meganucleasa monocatenaria” se refiere a un polipéptido que comprende un par de subunidades LAGLIDADG unidas por un engarce. Una meganucleasa monocatenaria tiene la organización: subunidad N terminal-engarce-subunidad C terminal. Una meganucleasa monocatenaria se distingue de una meganucleasa di-LAGLIDADG natural porque la subunidad N terminal debe derivar de una meganucleasa mono-LAGLIDADG y, por lo tanto, el engarce debe ser exógeno de la subunidad N terminal.

10 Como se usa en el presente documento, la expresión “recombinación homóloga” se refiere al proceso celular natural en el que se repara rotura de ADN bicatenaria usando una secuencia de ADN homóloga como el molde de reparación (véase, por ejemplo, Cahill y col. (2006), *Front. Biosci.* 11:1958-1976). La secuencia de ADN homóloga puede ser una secuencia cromosómica endógena o un ácido nucleico exógeno que se suministró a la célula. Por lo tanto, en algunas realizaciones, una meganucleasa diseñada racionalmente se usa para escindir una secuencia de reconocimiento dentro de una secuencia diana y un ácido nucleico exógeno con homología para o similitud de secuencia sustancial con la secuencia diana se suministra a la célula 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 por lo tanto en 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.

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

35 Como se usa en el presente documento, la expresión “secuencia de interés” significa cualquiera secuencia de ácido 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 genómica 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 de la secuencia de interés. Por ejemplo, una proteína puede marcarse con marcadores que incluyen, pero sin limitación, un epítipo, (por ejemplo, c-myc, FLAG) u otro ligando (por ejemplo, poli-His). Además, una secuencia de interés puede codificar una proteína de fusión, de acuerdo con técnicas conocidas en la materia (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, Wiley 1999). En algunas realizaciones, la secuencia de interés está flanqueada por una secuencia de ADN que 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 genómicas escindidas por la meganucleasa recombinante. En algunas realizaciones, la secuencia completa de interés es homóloga de o tiene similitud de secuencia sustancial con una 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 está flanqueada por secuencias de ADN con homología con o similitud de secuencia sustancial con la secuencia diana de modo que la recombinación homóloga inserte la secuencia de interés dentro del genoma en el locus de la secuencia diana. En algunas realizaciones, la secuencia de interés es sustancialmente idéntica a la secuencia diana excepto por mutaciones u otras modificaciones en la secuencia de reconocimiento de meganucleasas de modo que la meganucleasa no puede escindir la secuencia diana después de que se ha modificado por la secuencia de interés.

60 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 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. Está disponible 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 ácido nucleico, ambos de los cuales están disponibles a través del Centro Nacional para la Información Biotecnológica (www.ncbi.nlm.nih.gov/), y se describen, por ejemplo, en Altschul y col. (1990), *J. Mol. Biol.* 215: 403 -

410; Gish y States (1993), *Nature Genet.* 3: 266-272; Madden y col. (1996), *Meth. Enzymol.* 266: 131-141; Altschul y col. (1997), *Nucleic Acids Res.* 25: 33 89-3402); Zhang y col. (2000), *J. Comput. Biol.* 7(1-2): 203-14. Como se usa en el presente documento, el porcentaje de similitud de dos secuencias de aminoácidos es la puntuación basada en los siguientes parámetros para el algoritmo BLASTp: tamaño de palabra = 3; penalización 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 “correspondiente a” se usa para indicar que una modificación especificada en la primera proteína es una sustitución del mismo resto de aminoácido que 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 aminoácidos 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” al aminoácido “A” en la segunda proteína si los restos X y Y se corresponden entre sí en un alineamiento de secuencia, y a pesar del hecho de que X y 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 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 o 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 sentido exclusivo de “bien/o”.

2. Meganucleasas monocatenarias derivadas de subunidades de LAGLIDADG

Las comparaciones estructurales de meganucleasas mono- y di-LAGLIDADG revelan que la subunidades N terminales de meganucleasas di-LAGLIDADG tienden a ser menores que los monómeros mono-LAGLIDADG. La consecuencia de esto es que, en el caso de meganucleasas di-LAGLIDADG, el final (extremo C terminal) de la subunidad N terminal está mucho más cerca del inicio (extremo N terminal) de la subunidad C terminal. Esto significa que un engarce relativamente corto (por ejemplo, 5-20 aminoácidos) es suficiente para unir las dos subunidades. En el caso de meganucleasas mono-LAGLIDADG, el extremo C terminal de un monómero está generalmente muy lejos (aproximadamente 48 Å en el caso de I-Crel) del extremo N terminal del segundo monómero. Por lo tanto, fusionar un par de meganucleasas mono-LAGLIDADG en un único polipéptido requiere un engarce peptídico más largo (por ejemplo, >20 aminoácidos) que puede abarcar esta distancia. Un procedimiento alternativo, en el que la subunidad N terminal está truncada en un punto espacialmente más cercano al comienzo de la subunidad C terminal se ha presentado previamente (Epinat y col. (2003), *Nucleic Acids Res.* 31: 2952-62; documento WO 2003/078619), pero produce poco, o ningún, heterodímero funcional, como se muestra en el Ejemplo 1 posteriormente. Un análisis exhaustivo con respecto a la dificultad asociada con la producción de una meganucleasa monocatenaria funcional derivada de I-Crel puede encontrarse en Fajardo-Sanchez y col. (2008), *Nucleic Acids Res.* 36: 2163-2173.

2.1 Puntos de fusión para I-Crel

Se realizó una serie de mutantes de truncamiento en los que I-Crel natural o una variante modificada por ingeniería genética de I-Crel que se había alterado con respecto a su preferencia de sitio de escisión de ADN (designado “CCR2”, SEQ ID NO: 31; véase documento WO 2007/047859) se terminaron antes del aminoácido C terminal natural, Pro 163 (Tabla 1). Los homodímeros mutantes se expresaron en *E. coli*, se purificaron y se incubaron con la secuencia de reconocimiento natural (SEQ ID NO: 34-35) o la secuencia de reconocimiento de CCR2 (SEQ ID NO: 32-33) para ensayar con respecto a actividad de escisión.

TABLA 1
Mutantes de Truncamiento de I-Crel

Aminoácido C terminal	Actividad Natural	Actividad CCR2
Asp-153	+	+
Val-151	+	+
Val-148	+	-
Arg-141	-	-
Asn-136	-	-
Val-129	-	-
Ile-109	-	-
Leu-95	-	-

5 Se descubrió que I-Crel natural estaba activada cuando se truncaba en el resto 148 o restos más C terminales, pero estaba inactiva cuando se truncaba en el resto 141 o restos más N terminales. Por lo tanto, al menos algunos de los restos 141 a 147, o sustituciones conservativas de estos restos, se requerían para actividad natural. De forma similar, se descubrió que CCR2 estaba activa cuando se truncaba en el resto 151 o restos más C terminales, pero estaba inactiva cuando se terminaba en el resto 148 o restos más N terminales. Por lo tanto, al menos algunos de los restos 148 a 150, o sustituciones conservativas de estos restos, se requieren para la actividad CCR2. La diferencia entre I-Crel natural y la meganucleasa CCR2 diseñada racionalmente se debe supuestamente a una reducción de la estabilidad estructural de la meganucleasa CCR2 de modo que sea más sensible a desestabilización adicional por un truncamiento C terminal prematuro. Estos resultados de truncamiento son coherentes con una publicación de Prieto y col. en la que se descubrió que el bucle C terminal de I-Crel (aminoácidos 138-142) es esencial para la actividad de escisión (Prieto y col. (2007), Nucl. Acids Res. 35: 3262-3271). Tomados juntos, estos resultados indican que algunos restos cerca del extremo C terminal de I-Crel son esenciales para la unión de ADN y/o actividad catalítica y es poco probable que los procedimientos para producción de meganucleasa monocatenaria que truncan una subunidad I-Crel antes de aproximadamente el resto 142 (por ejemplo, Epinat y col. (2003), Nucl. Acids Res. 31: 2952-62; documento WO 2003/078619) produzcan una meganucleasa monocatenaria en la que ambas subunidades LAGLIDADG sean catalíticamente activas.

Por lo tanto, de acuerdo con la presente invención, el punto de fusión N terminal (es decir, entre la subunidad I-Crel N terminal y el engarce) debería quedar en o en dirección C terminal del resto 142 de la subunidad N terminal, incluyendo cualquiera de las posiciones 142-151, o cualquier posición C terminal del resto 151. Los restos 154-163 de I-Crel no están estructurados (Jurica y col. (1998), Mol. Cell 2: 469-476) y, por lo tanto, la inclusión de estos restos aumentará la flexibilidad y, posiblemente, la inestabilidad estructural de la meganucleasa monocatenaria resultante. Por el contrario, si se determina que se desean o requieren menos flexibilidad y mayor estabilidad estructural, pueden seleccionarse los puntos de fusión en los restos 142-153.

30 Cuando la subunidad LAGLIDADG C terminal de una meganucleasa monocatenaria deriva de I-Crel, el punto de fusión C terminal del engarce estará hacia el extremo N terminal de la secuencia I-Crel. Los restos 7, 8 y 9 son de interés particular como puntos de fusión C terminales en I-Crel debido a que estos restos (1) están estructuralmente conservados entre los miembros de la familia de meganucleasa LAGLIDADG y, por lo tanto, pueden proporcionar mayor compatibilidad en la formación de heterodímeros con otros miembros de la familia LAGLIDADG y (2) inician una hélice alfa que contiene el motivo LAGLIDADG conservado que está implicado en la función catalítica. Sin embargo, también pueden emplearse puntos de fusión N terminales del resto 7, incluyendo cualquiera de los restos 1-6 de acuerdo con la invención.

Los siguientes puntos de fusión N terminal y C terminal de I-Crel se seleccionaron para experimentación adicional, pero no deberían interpretarse como limitantes del alcance de la invención:

TABLA 2
Puntos de Fusión de I-Crel

Punto de fusión N terminal	Punto de fusión C terminal
Val-151	Lys-7
Leu-152	Asp-8
Asp-153	Phe-9

2.2 Engarces para meganucleasas monocatenarias derivadas de I-Crel

Para el fin de ligar un par de monómeros de I-Crel en un único polipéptido, se consideraron dos clases generales de engarce: (1) un engarce desestructurado que carece de estructura secundaria; y (2) un engarce estructurado que

5 tiene estructura secundaria. Se conocen bien en la técnica ejemplos de engarces desestructurados, e incluyen secuencias artificiales con alto contenido de Gly y Ser, o repeticiones. También se conocen bien en la técnica engarces estructurados, e incluyen los diseñados usando principios básicos de plegamiento de proteínas (por ejemplo, Aurora y Rose (1998), Protein Sci. 7: 21-38; Fersht, Structure and Mechanism in Protein Science, W. H. Freeman 1998).

10 La invención se validó usando un par de monómeros I-Crel diseñados racionalmente llamados "LAM1" (SEQ ID NO: 36) y "LAM2" (SEQ ID NO: 37). Estas endonucleasas diseñadas racionalmente se produjeron usando los procedimientos descritos en el documento WO 2007/047859 y se caracterizan en el mismo. Como resultará evidente para los expertos en la materia, sin embargo, los monómeros LAM1 y LAM2 son meramente ilustrativos de los muchos monómeros que pueden emplearse, incluyendo subunidades mono-LAGLIDADG naturales, subunidades mono-LAGLIDADG naturales truncadas en el extremo N terminal y/o C terminal, subunidades di-LAGLIDADG naturales truncadas en el extremo N terminal y/o C terminal, y modificaciones diseñadas racionalmente de cualquiera de los anteriores.

20 Un monómero a modo de ejemplo, LAM1, difiere en 7 aminoácidos de I-Crel natural y reconocen el semisitio:

5'-TGCGGTGTC-3' (SEQ ID NO: 38)
3'-ACGCCACAG-5' (SEQ ID NO: 39)

25 Por lo tanto, el homodímero LAM 1 reconoce la secuencia de reconocimiento palindrómica (en la que cada N no está restringido):

5'-TGCGGTGTCNNNNGACACCGCA-3' (SEQ ID NO: 40)
3'-ACGCCACAGNNNNCTGTGGCGT-5' (SEQ ID NO: 41)

30 El otro monómero a modo de ejemplo, LAM2, difiere en 5 aminoácidos de I-Crel natural y reconoce el semisitio:

5'-CAGGCTGTC-3' (SEQ ID NO: 42)
3'-GTCCGACAG-5' (SEQ ID NO: 43)

35 Por lo tanto, el homodímero LAM 2 reconoce la secuencia de reconocimiento palindrómica (en la que cada N no está restringido):

40 5'-CAGGCTGTCNNNNGACAGCCTG-3' (SEQ ID NO: 44)
3'-GTCCGACAGNNNNCTGTCCGAC-5' (SEQ ID NO: 45)

Un heterodímero que comprende un monómero LAM 1 y un monómero LAM2 ("heterodímero LAM1/LAM2") reconoce por lo tanto la secuencia de reconocimiento híbrida:

45 5'-TGCGGTGTCNNNNGACAGCCTG-3' (SEQ ID NO: 40)
3'-ACGCCACAGNNNNCTGTCCGAC-5' (SEQ ID NO: 41)

2.2.1 Engarces flexibles para meganucleasas monocatenarias

50 Se conoce diversos engarces peptídicos altamente flexibles en la técnica y pueden usarse de acuerdo con la invención. Por ejemplo, y sin limitación, se sabe que los engarces peptídicos que comprenden repeticiones de unidades Gly-Ser-Ser no están estructurados y son flexibles (Fersht, Structure and Mechanism in Protein Science, W. H. Freeman 1998). Los engarces con esta y composiciones similares se usan frecuentemente para fusionar dominios proteicos entre sí (por ejemplo, anticuerpos monocatenarios (Mack y col. (1995), Proc. Nat. Acad. Sci. 92: 7021-7025); receptores de factor de crecimiento (Ueda y col. (2000), J. Immunol. Methods 241: 159-170); enzimas (Brodelius y col. (2002), 269: 3570-3577); y dominios de nucleasa y de unión a ADN (Kim y col. (1996), Proc. Nat. Acad. Sci. 93: 1156-1160).

60 En general, el engarce flexible puede incluir cualquier secuencia polipeptídica que no forme estructuras secundarias estables en condiciones fisiológicas. En algunas realizaciones, los engarces incluyen un alto porcentaje (por ejemplo, > 50 %, 60 %, 70 %, 80 % o 90 %, o generalmente, 50 %-100 %) de restos no cargados polares (es decir, Gly, Ser, Cys, Asn, Gln, Tyr, Thr). Además, en algunas realizaciones, los engarces incluyen un bajo porcentaje de restos hidrófobos grandes (es decir, Phe, Trp, Met). Los engarces pueden incluir repeticiones de diversas longitudes (por ejemplo, (SG)_n, (GSS)_n, (SGGS)_n), pueden incluir secuencias aleatorias o pueden incluir combinaciones de las dos.

65

Por lo tanto, de acuerdo con la invención, se produjo un conjunto de fusiones monocatenarias entre LAM1 y LAM2 en las que un engarce peptídico altamente flexible unía covalentemente la subunidad N terminal (LAM1) con la subunidad C terminal (LAM2) usando Val-151 o Asp-153 como el punto de fusión N terminal y Phe-9 como el punto de fusión C terminal. Las proteínas monocatenarias se expresaron en *E. coli*, se purificaron y se ensayaron con respecto a la capacidad para escindir un sitio de ADN híbrido que comprendía un semisitio LAM 1 y un semisitio LAM2 (SEQ ID NO: 46 y 47). La actividad de escisión se clasificó en una escala de cuatro puntos: - sin actividad detectable; + actividad mínima; ++ actividad media; +++ actividad comparable al heterodímero LAM1/LAM2 producido por coexpresión de los dos monómeros en *E. coli* antes de la purificación de endonucleasa. Las proteínas también se evaluaron por SDS-PAGE para determinar el grado en que la región de engarce se proteolizaba durante la expresión o purificación para liberar las dos subunidades.

TABLA 3
Meganucleasas I-Crel Monocatenarias con Engarces Gly-Ser

Número de engarce	Parte de fusión N terminal	Parte de fusión C terminal	Secuencia de engarce	Actividad	Proteólisis del engarce
1	Val-151	Phe-9	(GSS) ₇ G	-	-
2	Val-151	Phe-9	(GSS) ₈ G	-	-
3	Val-151	Phe-9	(GSS) ₉ G	+	+
4	Val-151	Phe-9	(GSS) ₁₀ G	ND	+++
5	Val-151	Phe-9	(GSS) ₁₁ G	ND	+++
6	Val-151	Phe-9	(GSS) ₉ GG	+	+
7	Val-151	Phe-9	(GSS) ₉ GSG	+	+
8	Asp-153	Phe-9	(GSS) ₉ G	+	+

Los resultados indicaron que los engarces flexibles, tales como los engarces Gly-Ser en la Tabla 3, son adecuados para producción de meganucleasas monocatenarias siempre que la longitud sea apropiada (véase también Ejemplo 2). Por ejemplo, haciendo referencia a la Tabla 3, las meganucleasas monocatenarias que incluían engarces 1 y 2, que comprendían 22 y 25 aminoácidos totales, respectivamente, no mostraban ninguna actividad de escisión detectable con los puntos de fusión ensayados. SDS-PAGE indicó que estas meganucleasas estaban intactas y no se degradaron por proteasas, lo que condujo a la conclusión de que estas meganucleasas monocatenarias eran estructuralmente estables pero estaban restringidas funcionalmente por engarces que eran demasiado cortos para permitir que las subunidades LAGLIDADG individuales adoptaran la conformación necesaria para unión y/o catálisis de ADN. Los engarces 3, 6, 7 y 8, que comprendían 28, 29, 30 y 28 aminoácidos, respectivamente, mostraron todos niveles bajos de actividad de escisión. SDS-PAGE indicó que una cantidad pequeña (5 %-10 %) de cada uno se proteolizó en subunidades individuales mientras que la mayoría tenía un peso molecular correspondiente a las meganucleasas monocatenarias de longitud completa (~40 kilodalton). Los números 3 y 8 tienen la misma secuencia de engarce pero puntos de fusión N terminales en Val-151 y Asp-153, respectivamente. Ambas meganucleasas monocatenarias mostraron niveles comparables de actividad, lo que indica que el punto de fusión preciso no es crítico en este caso. Finalmente, los engarces 4 y 5, que comprendían 31 y 34 aminoácidos, respectivamente, no produjeron meganucleasa monocatenaria detectable cuando se purificaron de *E. coli*. Estos engarces se proteolizaron completamente en las subunidades LAM1 y/o LAM2 individuales como se detecta por SDS-PAGE y, por lo tanto, la actividad de escisión de estas meganucleasas no se investigó adicionalmente.

Estos resultados condujeron a los inventores a concluir que los engarces Gly-Ser son aceptables para la producción de meganucleasas monocatenarias basándose en la subunidad LAGLIDADG de la meganucleasa mono-LAGLIDADG I-Crel y los puntos de fusión particulares empleados, siempre que los engarces fueran mayores de 25 y menores de 31 aminoácidos de longitud. Para meganucleasas monocatenarias basadas en I-Crel con estos puntos de fusión, los engarces más cortos evitan la catálisis mientras que los engarces más largos son inestables y propensos a corte por proteasas.

Los efectos de variar los puntos de fusión en las longitudes de engarce aceptables pueden determinarse empíricamente por experimentación rutinaria y/o predecirse basándose en la realización de modelos tridimensionales de las estructuras proteicas. Significativamente, a medida que un punto de fusión se mueve en dirección N terminal o C terminal, puede moverse más cerca o más lejos del otro punto de fusión dependiendo de la estructura secundaria y terciaria de la proteína cerca del punto de fusión. Por lo tanto, por ejemplo, mover el punto de fusión N terminal en dirección C terminal (por ejemplo, del resto 150 al resto 155 para una subunidad N terminal) no da como resultado necesariamente que el punto de fusión N terminal esté físicamente más cerca del punto de fusión C terminal debido a que, por ejemplo, los restos N terminales en esa región pueden ser parte de una estructura secundaria/terciaria que apunta en el sentido o el sentido contrario del punto de fusión C terminal. Por lo tanto, mover un punto de fusión N terminal en dirección N terminal o C terminal, o mover un punto de fusión C

terminal en dirección N terminal a C terminal, puede dar como resultado un desplazamiento en el intervalo de longitudes de engarce aceptables hacia engarces más largos o más cortos. Ese desplazamiento, sin embargo, se determina fácilmente, como se muestra por los experimentos presentados en el presente documento, por experimentación rutinaria y/o realización de modelos tridimensionales.

5 Por lo tanto, en algunas realizaciones, los engarces flexibles útiles tienen longitudes de más de 25 restos y menos de 31 restos (incluyendo todos los valores entre medias), como se muestra en la Tabla 3 para una meganucleasa monocatenaria basada en dos subunidades LAGLIDADG I-Crel. En otras realizaciones, sin embargo, empleando diferentes subunidades LAGLIDADG y/o diferentes puntos de fusión, los engarces flexibles útiles pueden tener longitudes mayores de 15 y menores de 40 restos (incluyendo todos los valores entre medias), siempre que los engarces no se proteolisen exhaustivamente y que la meganucleasa monocatenaria conserve su actividad de unión a ADN y escisión como se determina por los ensayos sencillos descritos en el presente documento.

2.2.2 Engarces diseñados, estructurados para meganucleasas monocatenarias

15 En un intento de producir meganucleasas basadas en I-Crel monocatenarias con actividad nucleasa comparable a la enzima dimérica natural que son tanto suficientemente estables para almacenamiento a largo plazo como resistentes a proteólisis, pueden usarse engarces que tengan estructuras secundarias estables para unir subunidades de forma covalente. Una búsqueda en el banco de Datos de Proteínas (www.rcsb.org) no reveló ninguna proteína LAGLIDADG caracterizada estructuralmente con engarces adecuados para abarcar la gran distancia (aproximadamente 48 Å) entre los puntos de fusión N y C terminal identificados en I-Crel. Por lo tanto, se emplearon principios básicos conocidos que gobiernan la estructura proteica (por ejemplo, Aurora y Rose (1998), Protein Sci. 7: 21-38; Fersht, Structure and Mechanism in Protein Science, W. H. Freeman 1998) para producir un conjunto de engarces que se espera que tengan elementos estructurales adecuados para unir las dos subunidades. Específicamente, se postuló que un engarce adecuado comprendería (enumerado desde el punto de fusión N terminal al punto de fusión C terminal):

(1) **Bucle 1.** Este elemento estructural comienza en el punto de fusión N terminal e invierte la dirección de la cadena peptídica sobre sí misma (un giro de 180°). La secuencia puede ser de 3-8 aminoácidos y puede incluir al menos un resto de glicina o, en algunas realizaciones, 2-3 glicinas. Este elemento estructural puede estabilizarse introduciendo un motivo de "recubrimiento C" para terminar la hélice α C terminal de I-Crel e iniciar la vuelta posterior. Se describe normalmente que el motivo de recubrimiento de hélice comienza con un aminoácido hidrófobo en la vuelta final de la hélice (Aurora y Rose (1998), Protein Sci. 7: 21-38). El recubrimiento C puede tomar cualquiera de las formas enumeradas en la Tabla 4:

35

TABLA 4
Motivos de recubrimiento C

Número	Motivo
1	$h_1xpx-Gh$
2	$h_1xpx-nxhx$
3	$h_1xpx-nxph$
4	$h_1xxx-Gphx$
5	$h_1xxx-Gpph$
6	$h_1xxx-Pppph$
7	$h_1xxx-Ppph$

en la que h = un aminoácido hidrófobo (Ala, Val, Leu, Ile, Met, Phe, Trp, Thr o Cys); p = un aminoácido polar (Gly, Ser, Thr, Asn, Gln, Asp, Glu, Lys, Arg); n = un aminoácido no β ramificado (no Val, Ile, Thr o Pro); x = cualquier aminoácido del grupo h o p; G = glicina; y P = prolina. Obsérvese que Thr aparece en ambos grupos h y p debido a que su cadena lateral tiene grupos funcionales tanto hidrófobos (grupo metilo) como polares (hidroxilo). El guion designa el final de la hélice α y h_1 es un aminoácido hidrófobo en la vuelta final de la hélice (es decir, un aminoácido hidrófobo 0-4 aminoácidos antes del punto de fusión N terminal). En el caso de I-Crel, h_1 es normalmente Val-151 o Leu-152. Por lo tanto, un ejemplo de motivo 7 es la secuencia $V_{151}L_{152}D_{153}S$ -PGSV (véase, por ejemplo, Tabla 6, Engarce 9).

(2) **hélice α 1.** Después del bucle 1, esta primera hélice α en el engarce se diseña para quedar antiparalela a la hélice C terminal en I-Crel (aminoácidos 144-153) en la cara externa de la proteína durante una distancia de aproximadamente 30 Å. Este segmento debería ser de 10-20 aminoácidos de longitud, no debería contener ningún aminoácido glicina o prolina fuera de los motivos de recubrimiento N y C (posteriormente), y aminoácidos hidrófobos y polares alternativos con periodicidad de 3-4 aminoácidos para internar una cara de la hélice (la cara hidrófoba) contra la superficie de la subunidad de I-Crel N terminal exponiendo a la vez la otra cara al disolvente. La hélice podría, por ejemplo, tomar la forma pphpphphpph en la que p es cualquier aminoácido polar y h es cualquier

50

aminoácido hidrófobo pero no es glicina o prolina tal como la secuencia SQASSAASSASS (véase, por ejemplo, Tabla 6, Engarce 9). Están disponibles numerosos algoritmos para determinar la propensión helicoidal de una secuencia peptídica (por ejemplo, BMERC-PSA, <http://bmerc-www.bu.edu/psa/>; NNPREDICT, <http://alexander.compbio.ucsf.edu/~nomi/nnpredict.html>; PredictProtein, <http://www.predictprotein.org>) y cualquiera de estos puede usarse para producir una secuencia de la longitud apropiada que puede esperarse que adopte estructura secundaria α helicoidal. Como alternativa, esta secuencia de hélice podría derivar de una secuencia peptídica que se sabe que adopta una estructura secundaria α helicoidal en una proteína natural o diseñada. Están disponibles numerosos ejemplos de dichas secuencias peptídicas en el banco de Datos de Proteínas (www.rcsb.org).

Además, puede ser deseable iniciar la hélice α con un motivo de recubrimiento N para estabilizar su estructura (Aurora y Rose (1998), *Protein Sci.* 7:21-38). Este motivo abarca el punto de unión de bucle-hélice α y normalmente tiene una de las formas mostradas en la Tabla 5:

TABLA 5
Motivos de recubrimiento N

Número	Motivo
1	h-xpxhx
2	h-xxpph
3	hp-xpxhx
4	hp-xxpph
5	hpp-xpxhx
6	hpp-xxpph

en la que las designaciones son como en la Tabla 4 anterior y el guion representa la unión entre el bucle y la hélice. Un ejemplo del número de motivo 2 es la secuencia L-SPSQA (véase, por ejemplo, Tabla 6, Engarce 9).

(3) Vuelta 1. Después de la hélice α 1, se introduce una secuencia peptídica flexible, corta para girar la orientación completa de la cadena peptídica aproximadamente 90° en relación con la orientación de la hélice α 1. Esta secuencia puede ser de 3-8 aminoácidos de longitud y puede contener una o, en algunas realizaciones, 2-3 glicinas. Esta secuencia también puede contener un recubrimiento C tal como uno de los motivos en la Tabla 4 para estabilizar la hélice α 1 e iniciar la vuelta. Un ejemplo es la secuencia ASSS-PGSGI (véase, por ejemplo, Tabla 6, Engarce 9) que se adapta al motivo de recubrimiento C número 6. En este caso, la secuencia ASSS es la vuelta final de la hélice α 1 mientras que la secuencia PGSGI es la vuelta 1.

(4) hélice α 2. Esta hélice sigue al Vuelta 1 y se diseña para quedar en la superficie de I-Crel en un surco creado en la interfaz entre las subunidades LAGLIDADG. La superficie de este surco comprende principalmente aminoácidos 94-100 y 134-139 de la subunidad N terminal y los aminoácidos 48-61 de la subunidad C terminal.

La hélice α 2 puede diseñarse para que sea más corta que la hélice α 1 y puede comprender 1-3 vueltas de la hélice (4-12 aminoácidos). La hélice α 2 puede tener la misma composición de aminoácidos global que la hélice α 1 y también puede estabilizarse mediante la adición de un motivo de recubrimiento N de la Tabla 5. La secuencia I-SEALR es un ejemplo (véase, por ejemplo, Tabla 6, Engarce 9) que se adapta al motivo de recubrimiento N número 1. El engarce 9 incorpora una hélice α 2 relativamente corta que comprende la secuencia SEALRA que se espera que realice aproximadamente dos vueltas. Experimentos con diferentes secuencias de hélice α 2 de engarce han demostrado la importancia del registro helicoidal en esta región del engarce. La adición de un único aminoácido (por ejemplo, un Engarce 11), dos aminoácidos (por ejemplo, AS, Engarce 12) o tres aminoácidos (por ejemplo ASS, Engarce 13) antes de la terminación de la hélice α 2 con un aminoácido glicina puede dar como resultado que las proteínas I-Crel monocatenarias sean inestables y precipiten unos momentos después de la purificación de *E. coli* (Tabla 6). Por el contrario, la adición de cuatro aminoácidos (por ejemplo, ASSA, engarce 14), que se espera que realicen una vuelta adicional completa y restauran el registro helicoidal al del Engarce 1 es estable y activa.

(5) Bucle 2. Este bucle termina la hélice α 2 y gira la cadena peptídica sobre sí misma para unirse con la subunidad I-Crel C terminal en el punto de fusión C terminal. Como con el bucle 1, esta secuencia puede ser de 3-8 aminoácidos de longitud y puede contener una o más glicinas. También puede contener un motivo de recubrimiento C de la Tabla 4 para estabilizar la hélice α 2. Por ejemplo, la secuencia ALRA-GA del engarce 9 se adapta al motivo de recubrimiento C número 1. Además, este segmento puede comenzar un recubrimiento N en la hélice α N terminal (aminoácidos 7-20) de la subunidad I-Crel C terminal. Por ejemplo, la secuencia T-KSK₇E₈F₉ del Engarce 9 se adapta al motivo de recubrimiento N número 2. En este caso, el punto de fusión C terminal es Lys-7. En otros casos, el punto de fusión puede moverse más lejos en la segunda subunidad (por ejemplo a los aminoácidos 8 o 9), opcionalmente con la adición de 1-2 aminoácidos al bucle 2 para compensar el cambio en el registro helicoidal a medida que se mueve el punto de fusión C terminal. Por ejemplo, los engarces 15-23 en la Tabla 6 posterior tienen Glu-8 como el punto de fusión C terminal y todos tienen un aminoácido adicional en el Bucle 2 en relación con los engarces 1-6.

Empleando los principios descritos anteriormente, se desarrolló el conjunto de engarces descritos en la Tabla 6. Se construyó un conjunto de meganucleasas I-Crel monocatenarias que incorporan los engarces entre las subunidades LAM 1 y LAM2 y cada una se ensayó con respecto a actividad frente a la secuencia de reconocimiento híbrida LAM1/LAM2. En todos los casos, el punto de fusión N terminal fue Asp-153 de LAM 1 y el punto de fusión C terminal fue Lys-7 o Glu-8 (indicados en la columna "CFP") de LAM2. La actividad de escisión se clasificó en una escala de cuatro puntos: - actividad no detectable; + actividad mínima; ++ actividad media; +++ actividad comparable al heterodímero LAM1/LAM2 producido por coexpresión de los dos monómeros en *E. coli* antes de la purificación de endonucleasa. Inmediatamente después de la purificación, las meganucleasas monocatenarias se centrifugaron (2100 g durante 10 minutos) para sedimentar proteína precipitada (indicativa de inestabilidad estructural) y la cantidad de precipitado (ppt) observado se puntuó: - no precipitado; + poco precipitado; ++ precipitado significativo. Las muestras proteicas que precipitaron en un grado significativo no pudieron ensayarse con respecto a actividad de escisión.

TABLA 6
Engarces para I-Crel monocatenaria

Nº	CFP	Secuencia de Engarce	Actividad	ppt
9	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGATKS	+++	-
10	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGGATKS	+++	-
11	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAAGGATKS	ND	++
12	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASGGATKS	ND	++
13	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASSGGATKS	ND	++
14	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASSAGGATKS	+++	-
15	E8	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGATKEF	++	+
16	E8	SLPGSVGGI _I SPSQASSAASSASSSPGSGI _I SEAPRAGATKEF	++	-
17	E8	SLPGSVGGI _I SPSQASSAASSASSSPGSGI _I SEATRAGATKEF	++	+
18	E8	SLPGSI _I GGI _I SPSQASSAASSASSSPGSGI _I SEALRAGATKEF	++	+
19	E8	SLPGSVGGV _V SPSQASSAASSASSSPGSGV _V SEASRAGATKEF	++	+
20	E8	SLPGSVGGI _I SPSQASSAASSASSSPGSGI _I SEALRAGATKEF	++	+
21	E8	SLPGSI _I GGI _I SPSQASSAASSASSSPGSGI _I SEASRAGATKEF	++	-
22	E8	SPGSI _I GGV _V SPSQASSAASSASSSPGSGI _I SEATRAGATKEF	++	-
23	E8	SLPGSI _I GGV _V SPSQASSAASSASSSPGSGI _I SEAPRAGATKEF	ND	++
24	E8	SLPGSVGGI _I SPSQASSAASSASSSPGSGI _I SEALRAGATKEF	++	-
25	E8	SLPGSI _I GGV _V SPSQASSAASSASSASSPAGSGI _I SEASRAGATKEF	++	-

Las meganucleasas monocatenarias de cada uno de estos engarces excepto para 11-13 y 23 (que no se investigaron) corrieron como una banda única del peso molecular deseado (~40 kilodalton) en un gel de SDS-PAGE, lo que indica resistencia a escisión proteolítica de la secuencia de engarce. En al menos un caso (Engarce 9), la meganucleasa LAM monocatenaria pudo almacenarse a 4 °C durante más de 4 semanas sin ninguna prueba de degradación o pérdida de actividad de escisión. Además, varias endonucleasas LAM monocatenarias (9, 10 y 14) escindieron la secuencia de reconocimiento LAM1/LAM2 híbrida con eficacia comparable al heterodímero LAM1/LAM2 purificado, lo que indica que la fusión de subunidades I-Crel usando estos engarces no altera significativamente la actividad endonucleasa (véase Ejemplo 2).

En claro contraste con el heterodímero LAM1/LAM2 purificado (que es, de hecho, una mezcla de homo y heterodímeros), las meganucleasas LAM monocatenarias que incorporan los engarces en la Tabla 6 escinden el sitio híbrido mucho más eficazmente que una de las secuencias palindrómicas (véase Ejemplo 2). Las secuencias palindrómicas normalmente se cortan con eficacia <5 % en relación con el sitio híbrido. Esta escisión no pretendida de los sitios de ADN palindrómicos podrían deberse a (1) homodimerización de subunidades LAM 1 o LAM2 de un par de proteínas monocatenarias diferentes, (2) cortes secuencial de ambas cadenas de la secuencia palindrómica por una única subunidad (LAM1 o LAM2) dentro de la meganucleasa monocatenaria o (3) cantidades pequeñas de LAM 1 o LAM2 homodiméricas que se forman después de la escisión proteolítica de la meganucleasa monocatenaria en sus subunidades individuales (aunque los resultados de SDS-PAGE hacen esta última explicación improbable). Aunque las meganucleasas monocatenarias I-Crel mantienen cierta actividad hacia los sitios de ADN

palindrómicos, la actividad está tan desproporcionadamente sesgada a favor del sitio híbrido que este enfoque representa una mejora muy significativa sobre los procedimientos existentes.

3. Meganucleasas monocatenarias derivadas de I-Msol

5 I-Msol es un homólogo estructural cercano de I-Crel y se han presentado procedimientos similares para rediseñar la especificidad de unión de ADN de esta meganucleasa (documento WO 2007/047859). El procedimiento presentado anteriormente para la producción de una meganucleasa I-Crel monocatenaria puede aplicarse directamente a I-Msol. Los aminoácidos Phe-160, Leu-161 y Lys-162 de I-Msol son estructuralmente homólogos de, respectivamente, Val-151, Leu-152 y Asp-153 de I-Crel. Estos aminoácidos, por lo tanto, pueden seleccionarse como los puntos de fusión N terminales para I-Msol. Además, la estructura cristalina de rayos X de I-Msol revela que los aminoácidos 161-166 actúan de forma natural como un recubrimiento C e inician una vuelta en el extremo C terminal de la proteína que invierte la dirección de la cadena peptídica. Por lo tanto, Ile-66 puede seleccionarse como el punto de fusión N terminal siempre que el engarce esté acortado en su extremo N terminal para retirar la parte de recubrimiento C del Bucle 1. Pro-9, Thr-10 y Glu-11 de I-Msol son estructuralmente homólogos de, respectivamente, Lys-7, Glu-8 y Phe-9 de I-Crel y pueden seleccionarse como puntos de fusión C terminales para I-Msol (Tabla 7). Además, debido a que la secuencia L₇Q₈P₉T₁₀E₁₁A₁₂ de I-Msol forma un recubrimiento N natural (motivo 2 de la Tabla 5), Leu-7 puede incluirse como un punto de fusión.

TABLA 7
Puntos de Fusión de I-Msol

Puntos de fusión N terminales	Puntos de fusión C terminales
Phe-160	Leu-7
Leu-161	Pro-9
Lys-162	Thr-10
Ile-166	Glu-11

25 Cualquiera de los engarces en las Tablas 3 o 6 pueden usarse para la producción de endonucleasas I-Msol monocatenarias. Por ejemplo, el Engarce 9 de la Tabla 6 puede usarse para unir un par de subunidades de I-Msol en una meganucleasa monocatenaria funcional usando Lys-162 y Pro-9 como puntos de fusión. En una realización, Pro-9 se cambia a un aminoácido diferente (por ejemplo, alanina o glicina) debido a que la prolina es estructuralmente restrictiva. Esto es análogo a seleccionar Thr-10 como el punto de fusión C terminal y añadir un aminoácido terminal al extremo C terminal de los engarces enumerados en las Tablas 3 o 6. Por ejemplo los Engarces 26 y 27 de la Tabla 8 son idénticos al Engarce 9 de la Tabla 6 excepto por la adición de un único aminoácido en el extremo C terminal para explicar un cambio en el punto de fusión C terminal de Pro-9 (estructuralmente homólogo a Lys-7 de I-Crel) a Thr-10 (estructuralmente homólogo a Glu-8 de I-Crel).

35 En otra realización, como se describe en el Ejemplo 4, una meganucleasa monocatenaria derivada de I-Mso también puede producirse de forma exitosa usando una secuencia de engarce seleccionada del Engarce 28-30 de la Tabla 8 en el que I-166 se selecciona como el punto de fusión N terminal y Leu-7 se selecciona como el punto de fusión C terminal. Debido a que I-166 se selecciona como el punto de fusión N terminal, la parte de recubrimiento C del Bucle 1 (correspondiente a los primeros 6 aminoácidos de cada uno de los engarces de la Tabla 6) puede retirarse. Además, la hélice α 1 de los engarces 28-30 se alarga en 3 aminoácidos (AAS, subrayados en la Tabla 8) en relación con los engarces enumerados en la Tabla 6, correspondientes a una vuelta adicional de la hélice. Usando los engarces 28-30 y los puntos de fusión especificados, es posible producir meganucleasas monocatenarias de alta actividad resistentes a proteasa que comprenden un par de subunidades derivadas de I-Mso (véase Ejemplo 4).

TABLA 8
Engarces para I-Msol monocatenaria

Nº	NFP	CFP	Secuencia de engarce	Actividad	ppt
26	K162	T10	PGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKSA	++	-
27	K162	T10	PGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKSG	++	-
28	I166	L7	GGASPSQASSAASSASSA <u>AASS</u> PGSGISEALRAASSLASKPGST	+++	-
29	I166	L7	GGASPSQASSAASSASSA <u>AASS</u> PGSGISEALRAASSPGST	+++	-
30	I166	L7	GGASPSQASSAASSASSA <u>AASS</u> PGSGPSEALRAASSFASKPGST	+++	-

4. Meganucleasas monocatenarias derivadas de I-Ceul

I-Ceul es un homólogo estructural cercano de I-Crel y se han presentado procedimientos similares para rediseñar la especificidad de unión a ADN de esta meganucleasa (documento WO 2007/047859). El procedimiento presentado anteriormente para la producción de una meganucleasa I-Crel monocatenaria puede aplicarse directamente a I-Ceul. Los aminoácidos Ala-210, Arg-211 y Asn-212 de I-Ceul son estructuralmente homólogos de, respectivamente, Val-151, Leu-152 y Asp-153 de I-Crel. Estos aminoácidos, por lo tanto, pueden seleccionarse como los puntos de fusión N terminal para I-Ceul. Ser-53, Glu-54 y Ser-55 de I-Ceul son estructuralmente homólogos de, respectivamente, Lys-7, Glu-8 y Phe-9 de I-Crel y pueden seleccionarse como puntos de fusión C terminales para I-Ceul (Tabla 9).

TABLA 9
Puntos de Fusión de I-Ceul

Puntos de fusión N terminales	Puntos de fusión C terminales
Ala-210	Ser-53
Arg-211	Glu-54
Asn-212	Ser-55

Cualquiera de los engarces en las Tablas 3 o 6 pueden ser eficaces para la producción de endonucleasas I-Ceul monocatenarias. Por ejemplo, las subunidades de I-Ceul pueden unirse por el engarce 9 de la Tabla 6 usando Asn-212 como el punto de fusión N terminal y Ser-53 como el punto de fusión C terminal.

Los puntos de fusión C terminales seleccionados para I-Ceul dan como resultado la retirada de los aminoácidos 1-52 de la subunidad I-Ceul C terminal. Los análisis estructurales (Spiegel y col. (2006), Structure 14: 869-880) revelan que estos aminoácidos forman un dominio estructurado que descansa sobre la superficie de I-Ceul e interna una cantidad sustancial de área de superficie hidrófoba a la que contribuyen los aminoácidos 94-123. Es posible, por lo tanto, que la retirada de este dominio N terminal desestabilice la subunidad I-Ceul C terminal en la meganucleasa monocatenaria. Para mitigar esta posibilidad, los aminoácidos hidrófobos que se expondrían por la retirada de este dominio terminal pueden mutarse a aminoácidos polares (por ejemplo, aminoácidos hidrófobos no ramificados β , pueden mutarse a Ser mientras que los aminoácidos hidrófobos ramificados β pueden mutarse a Thr). Por ejemplo, Leu-101, Tyr-102, Leu-105, Ala-121 y Leu-123 pueden mutarse a Ser mientras que Val-95, Val-98 e Ile-113 pueden mutarse a Thr.

Como alternativa, el dominio N terminal de la subunidad I-Ceul C terminal puede dejarse en gran medida intacto y unirse a la subunidad N terminal mediante un engarce truncado. Esto puede conseguirse usando Lys-7, Pro-8, Gly-9 o Glu-10 como un punto de fusión C terminal. El engarce puede ser un engarce Gly-Ser flexible (por ejemplo, Engarce 3 de la Tabla 3) truncado en aproximadamente el 50 % de su longitud (es decir, $(GSS)_4G$ o $(GSS)_5G$). Como alternativa, el engarce puede ser cualquiera de los engarces de la Tabla 6 truncado dentro de la vuelta 1. Por lo tanto, usando el engarce 9 de la Tabla 6 como ejemplo, puede realizarse una meganucleasa I-Ceul monocatenaria con la siguiente composición:

Subunidad N terminal N₂₁₂-SLPGSVGGLSPSQASSAASSASSSPGS-G₉ subunidad C terminal

5. Meganucleasas monocatenarias derivadas de dos miembros de la familia LAGLIDADG diferentes

La presente invención también permite la producción de meganucleasas monocatenarias en las que cada una de las subunidades deriva de un dominio LAGLIDADG natural diferente. "Diferente", como se usa en la presente descripción se refiere a subunidades de LAGLIDADG que no derivan del mismo miembro de la familia LAGLIDADG natural. Por lo tanto, como se usa en la presente descripción, las subunidades LAGLIDADG diseñadas racionalmente del mismo miembro de la familia (por ejemplos, dos subunidades I-Crel que se han modificado por ingeniería genética con respecto a especificidad de escisión de ADN) no se consideran "diferentes". Específicamente, la invención permite la producción de meganucleasas monocatenarias que comprenden una subunidad N-terminal derivada de una meganucleasa mono-LAGLIDADG (por ejemplo, I-Crel, I-Msol o I-Ceul) ligada a una subunidad C terminal derivada de una meganucleasa mono-LAGLIDADG diferente o uno de los dos dominios LAGLIDADG de una meganucleasa di-LAGLIDADG. Por ejemplo, puede producirse una meganucleasa monocatenaria que comprende una subunidad I-Crel N terminal que puede haberse diseñado racionalmente o no con respecto a especificidad de sitio de reconocimiento de ADN, ligada a una subunidad I-Msol C terminal que también puede haberse diseñado racionalmente o no con respecto a especificidad de sitio de reconocimiento de ADN.

En los casos de I-Crel, I-Msol e I-Ceul, los puntos de fusión deseables y engarces son como se ha descrito anteriormente. Por ejemplo, puede producirse una fusión de I-Crel con I-Msol monocatenaria usando el Engarce 9 de la Tabla 6 para unir Asp-153 de I-Crel con Thr-10 de I-Msol. La Tabla 9 enumera puntos de fusión C terminales potenciales para dominios LAGLIDADG individuales de las meganucleasas di-LAGLIDADG I-Scel, I-Dmol e I-Anil.

TABLA 10
Puntos de Fusión C terminales para Subunidades de Meganucleasa di-LAGLIDADG

I-Scel N-terminal (31-123)	I-Scel C-terminal (132-225)	I-Anil N-terminal (3-125)	I-Anil C-terminal (135-254)	I-Dmol N-terminal (8-98)	I-Dmol C-terminal (104-178)
I-31	Y-132	D3	S-135	S-8	R-104
E-32	L-133	L4	Y-136	G-9	E-105
Q-33	T-134	Y6	F-137	I-10	Q-106

5 Los puntos de fusión enumerados en las Tablas 7, 9 y 10 se basan en comparaciones de estructura entre la meganucleasa en cuestión e I-Crel en las que se identificaron posiciones de aminoácidos que son estructuralmente homólogas de los puntos de fusión de I-Crel. Los puntos de fusión también pueden identificarse en subunidades LAGLIDADG que no se han caracterizado estructuralmente usando alineamientos de secuencia proteica con I-Crel. Esto es particularmente cierto de los puntos de fusión C terminales que pueden identificarse fácilmente en cualquier subunidad LAGLIDADG basándose en la localización del motivo LAGLIDADG conservado. Los aminoácidos que
 10 están a 4-6 restos N terminal del comienzo del motivo LAGLIDADG son puntos de fusión C terminales aceptables.

Debido a que las interfaces de dimerización entre subunidades de diferentes endonucleasas LAGLIDADG varían, las subunidades pueden no asociarse en "heterodímeros" funcionales a pesar de estar unidas covalentemente como un
 15 único polipéptido. Para promover asociación, la interfaz entre las dos subunidades puede diseñarse racionalmente, como se describe en el documento WO 2007/047859. En su forma más sencilla, esto implica sustituir restos de la interfaz de una subunidad a otra. Por ejemplo, I-Crel e I-Msol difieren en la región de interfaz principalmente en Glu-8 de I-Crel (que es un Thr en la posición homóloga de I-Msol, aminoácido 10) y Leu-11 (que es una Ala en la posición homóloga de I-Msol, aminoácido 13). Por lo tanto, las subunidades I-Crel e I-Msol pueden hacerse interactuar eficazmente cambiando Glu-8 y Leu-11 de la subunidad de I-Crel a Thr y Ala, respectivamente, o
 20 cambiando Thr-10 y Ala-13 de la subunidad I-Msol a Glu y Leu, respectivamente.

También pueden usarse técnicas tales como algoritmos de diseño de proteínas informáticos para diseñar racionalmente las interfaces de subunidades. Dichos procedimientos se conocen en la técnica. Por ejemplo, Chevalier y col. usaron un algoritmo informático para rediseñar la interfaz entre I-Crel y el dominio LAGLIDADG N terminal de I-Dmol para permitir que los dos interactúen (Chevalier y col. (2002), Mol. Cell 10: 895-905). Teniendo
 25 en cuenta estos resultados, puede producirse una meganucleasa monocatenaria que comprende una subunidad N terminal derivada de I-Crel y una subunidad C terminal derivada del dominio LAGLIDADG N terminal de I-Dmol (1) seleccionando un punto de fusión I-Crel de la Tabla 2, (2) seleccionando un punto de fusión C terminal en I-Dmol de la Tabla 10, (3) seleccionando un engarce de la Tabla 6 (o diseñando un engarce similar basándose en las reglas proporcionadas) y (4) incorporando las mutaciones L11A, F16I, K96N y L97F en la subunidad I-Crel y las mutaciones I19W, H51F y L55R en la subunidad I-Dmol como proponen Chevalier y col.

Como alternativa, pueden usarse procedimientos empíricos tales como evolución dirigida para modificar por ingeniería genética la interfaz entre dos subunidades LAGLIDADG diferentes. Dichos procedimientos se conocen en
 35 la técnica. Por ejemplo, pueden producirse bibliotecas genéticas en las que se seleccionan aleatoriamente aminoácidos específicos en la interfaz de la subunidad, y miembros de bibliotecas que permiten que se explore experimentalmente la interacción entre las dos subunidades. Dichos procedimientos de exploración se conocen en la técnica (por ejemplo, Sussman y col. (2004), J. Mol. Biol. 342: 31-41; Chames y col. (2005), Nucl. Acids Res. 33: e178; Seligman y col. (2002), Nucl. Acids Res. 30: 3870-9; Arnould y col. (2006), J. Mol. Biol. 355: 443-58) y pueden realizarse para ensayar con respecto a la capacidad de una meganucleasa monocatenaria que comprende dos subunidades LAGLIDADG diferentes para escindir un sitio de ADN híbrido dentro de una célula de levadura o bacteriana.

45 6. Meganucleasas monocatenarias con actividad de especificidad de escisión de ADN, y/o afinidad de unión a ADN alteradas

La invención puede usarse para producir meganucleasas monocatenarias que comprenden subunidades LAGLIDADG individuales que se han modificado por ingeniería genética con respecto a especificidad de escisión de ADN usando diversos procedimientos. Dichos procedimientos incluyen diseño racional (por ejemplo, documento WO
 50 2007/047859), diseño informático (por ejemplo, Ashworth y col. (2006), Nature 441: 656-659), y selección genética (Sussman y col. (2004), J. Mol. Biol. 342: 31-41; Chames y col. (2005), Nucl. Acids Res. 33: e178; Seligman y col. (2002), Nucl. Acids Res. 30: 3870-9; Arnould y col. (2006), J. Mol. Biol. 355: 443-58). Dichas meganucleasas pueden dirigirse a sitios de ADN que difieren de los sitios reconocidos por meganucleasas naturales. La invención también puede usarse para unir subunidades de LAGLIDADG que se han diseñado racionalmente para tener actividad alterada (por ejemplo, documento WO 2007/047859; Arnould y col. (2007), J. Mol. Biol. 371(1): 49-65) o afinidad de
 55 unión a ADN como se describe en el documento WO 2007/047859.

7. Procedimientos para producir células y organismos recombinantes

Los aspectos de la presente invención proporcionan además procedimientos para producir células y organismos recombinantes, transgénicos o modificados genéticamente de otro modo usando meganucleasas monocatenarias. Por lo tanto, en ciertas realizaciones, se desarrollan meganucleasas monocatenarias recombinantes para provocar específicamente una rotura bicatenaria en un único sitio o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para posibilitar la inserción o las inserciones precisas de una secuencia de interés por recombinación homóloga. En otras realizaciones, se desarrollan meganucleasas recombinantes para provocar específicamente una rotura bicatenaria en un único sitio o en relativamente pocos sitios en el ADN genómico de una célula o un organismo para (a) posibilitar una inserción o inserciones poco habituales de una secuencia de interés por unión de extremos no homólogos o (b) posibilitar la alteración de la secuencia diana por unión de extremos no homólogos. Como se usa en el presente documento con respecto a recombinación homóloga o unión de extremos no homólogos de secuencias de interés, el término "inserción" significa la ligación de una secuencia de interés en un cromosoma de modo que la secuencia de interés se integre en el cromosoma. En el caso de recombinación homóloga, una secuencia insertada puede reemplazar una secuencia endógena, de modo que el ADN original se reemplace por ADN exógeno de igual longitud, pero con una secuencia de nucleótidos alterada. Como alternativa, una secuencia inyectada 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 no humanos tales como primates no humanos, caballos, vacas, cabras, cerdos, ovejas, perros, gatos, cobayas, ratas, ratones, lagartos, peces e insectos tales como especies de *Drosophila*. En otras realizaciones, el organismo es un hongo tal como una especie de *Candida*, *Neurospora* o *Saccharomyces*.

En algunas realizaciones, los procedimientos de la invención implican la introducción de una secuencia de interés en una célula tal como una célula germinal no humana o célula madre no humana que puede convertirse en un organismo recombinante maduro o permitir que el organismo modificado genéticamente resultante dé lugar a una descendencia que porte la secuencia de interés insertada en su genoma.

Pueden suministrarse proteínas de meganucleasa en células para escindir ADN genómico, lo que posibilita la recombinación homóloga o unión de 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 liposoma (véase, por ejemplo, Lipofectamine™, Invitrogen Corp., Carlsbad, CA). La formulación de liposoma puede usarse para facilitar la fusión de bicapa lipídica con una célula diana, permitiendo de este modo que los contenidos del liposoma o proteínas asociadas con su superficie se lleven a la célula. Como alternativa, la enzima puede fusionarse con un péptido de captación apropiado tal como el de la proteína TAT del VIH para captación celular directa (véase, por ejemplo, Hudecz y col. (2005), Med. Res. Rev. 25: 679-736).

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

Los ejemplos no limitantes de los vectores para transfección de ADN incluyen vectores de virus, plásmidos, cósmidos y vectores YAC. La transfección de secuencias de ADN puede conseguirse por diversos procedimientos 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 y col. (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 y col. (2002), Life Sci. 71(15): 1771-8).

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

En ciertas realizaciones, se produce una planta 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 meganucleasas 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 que corresponden a la meganucleasa recombinante solamente, de modo que la escisión promueva unión de extremos no homólogos y altere la secuencia diana que contiene la secuencia de reconocimiento. En dichas realizaciones, las secuencias de meganucleasa están bajo el control de secuencias reguladoras que posibilitan la expresión de la meganucleasa en las células vegetales huésped. 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 y col. (2004), *Plant J.* 37: 218-228), y promotores específicos de tejido vegetal tales como el promotor LGC1 (véase, por ejemplo, Singh y col. (2003), *FEBS Lett.* 542: 47-52).

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

En otras realizaciones, se produce un animal no humano modificado genéticamente usando una meganucleasa recombinante. Como con las células vegetales, las secuencias de ácido nucleico pueden introducirse en una célula germinal no humana o una célula no humana que con el tiempo se convertirá en un organismo transgénico no humano. En algunas realizaciones, la célula es un huevo fertilizado no humano, y pueden inyectarse moléculas de ADN exógeno en el pronúcleo del huevo fertilizado. Los huevos microinyectados se transfieren después a los oviductos de madres de acogida 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 más sitios discretos en el genoma. Como alternativa, los animales modificados genéticamente pueden obtenerse utilizando células madre embrionarias no humanas recombinantes ("ES") para la generación de los transgénicos, como se describe en Gossler y col. (1986), *Proc. Natl. Acad. Sci. USA* 83: 9065-9069.

En ciertas realizaciones, un vector de expresión de mamífero recombinante es capaz de dirigir la expresión específica de tejido del ácido nucleico preferentemente en un tipo celular particular. Se conocen elementos reguladores específicos de tejido en la técnica. Los ejemplos no limitantes de promotores específicos de tejido adecuados incluyen el promotor de albúmina (específico de hígado, Pinkert y col. (1987), *Genes Dev.* 1: 268-277), promotores específicos linfoides (Calame y Eaton (1988), *Adv. Immunol.* 43: 235-275), en particular promotores del receptor de linfocitos T (Winoto y Baltimore (1989), *EMBO J.* 8: 729-733) e inmunoglobulinas (Banerji y col. (1983), *Cell* 33: 729-740; Queen y Baltimore (1983), *Cell* 33: 741-748), promotores específicos de neuronas (por ejemplo, el promotor de neurofilamento; Byrne y Ruddle (1989), *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), promotores específicos de páncreas (Edlund y col. (1985), *Science* 230: 912-916), y promotores específicos de glándulas mamarias (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 están abarcados promotores regulados por el desarrollo, por ejemplo los promotores hox murinos (Kessel y Gruss (1990), *Science* 249: 374-379) y el promotor de α -fetoproteína (Campes y Tilghman (1989), *Genes Dev.* 3: 537-546).

En ciertas realizaciones, una meganucleasa monocatenaria puede marcarse con un epítipo peptídico (por ejemplo, un epítipo HA, FLAG o Myc) para supervisar 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 localizarla en el citoplasma. La meganucleasa puede fusionarse también con una proteína no relacionada o dominio proteico tal como una proteína que estimula la reparación de ADN o recombinación homóloga (por ejemplo, recA, RAD51, RAD52, RAD54, RAD57 o BRCA2).

8. Procedimientos para terapia génica

Aspectos de la invención posibilitan 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 es defectuoso en su estructura y/o función. La expresión "terapia génica" también puede referirse a modificaciones realizadas a un gen deletéreo o elemento regulador (por ejemplo, oncogenes) que reducen o eliminan 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 que resultan de organismos infecciosos.

- En algunos aspectos de la invención, los genes disfuncionales se reemplazan o desactivan mediante la inserción de secuencias de ácido nucleico exógenas 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 para aliviar la afección. La secuencia puede ser una región con un exón, intrón, promotor u otra región reguladora que provoque 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 carecer de la función necesaria o tener más de la función necesaria.
- Pueden usarse secuencias de ácidos nucleicos exógenas insertadas en la región modificada para proporcionar secuencias "reparadas" que normalizan el gen. Puede conseguirse reparación génica mediante la introducción de secuencias génicas apropiadas en el gen que posibilite 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 las mutaciones que provocan expresión o regulación anómala se reparan. En otras realizaciones, la secuencia de ácido nucleico para insertar contiene el codón de parada de 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.
- Como alternativa, las secuencias de ácido nucleico pueden eliminar la función génica por completo alterando 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 exógenas 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 altera directamente por la meganucleasa introduciendo inserciones de bases, delecciones de bases y/o mutaciones de desplazamiento de fase a través de unión de extremos no homólogos.
- En muchos casos, es deseable dirigir las secuencias genéticas apropiadas a una célula o población de células diana que es la causa de la enfermedad. Dicha dirección de compuestos terapéuticos evita que las células sanas sean dianas de los compuestos terapéuticos. Esto aumenta la eficacia del tratamiento, reduciendo a la vez los efectos potencialmente adversos que el tratamiento podría tener en células sanas.
- El suministro de genes de meganucleasa recombinante y la secuencia de interés para insertar en el genoma a 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 conserva su capacidad para replicar dentro de la célula o tejido diana. Una o más secuencias de ADN pueden introducirse en el genoma viral alterado, para producir un genoma viral que actúe como un vector, y pueden insertarse o no en un genoma huésped y expresarse posteriormente. Más específicamente, ciertas realizaciones incluyen emplear un vector retroviral tal como, pero sin limitación, los vectores pLJ o MFG. Un vector MFG es un vector de virus de leucemia murina de Moloney (MoMLV) simplificado en el que las secuencias de ADN que codifican las proteínas pol y env se han suprimido para hacerlo defectuoso en replicación. Un vector retroviral pLJ también es una forma del MoMLV (véase, por ejemplo, Korman y col. (1987), Proc. Nat'l Acad. Sci., 84: 2150-2154). En otras realizaciones, un adenovirus recombinante o virus adenoasociado puede usarse como un vector de suministro.
- En otras realizaciones, el suministro de la 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 cargamento de ácido nucleico y/o proteínas se conoce en la técnica (véase, por ejemplo, Lasic y col. (1995), Science 267: 1275-76). Los inmunoliposomas incorporan anticuerpos contra antígenos asociados a células en liposomas, y pueden suministrar secuencias de ADN para la meganucleasa o la meganucleasa en sí misma a tipos celulares específicos (véase, por ejemplo, Lasic y col. (1995), Science 267: 1275-76; Young y col. (2005), J. Calif. Dent. Assoc. 33(12): 967-71; Pfeiffer y col. (2006), J. Vasc. Surg. 43(5):1021-7). Se conocen bien en la técnica procedimientos para producir y usar formulaciones de liposomas (véase, por ejemplo, Patente de Estados Unidos Nº 6.316.024, Patente de Estados Unidos Nº 6.379.699, Patente de Estados Unidos Nº 6.387.397, Patente de Estados Unidos Nº 6.511.676 y Patente de Estados Unidos Nº 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.

9. Procedimientos para tratar infección por patógenos

- Aspectos de la invención proporcionan también procedimientos 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 metilicina y *Mycoplasma tuberculosis*. Los organismos patógenos también incluyen organismos fúngicos tales como, pero sin limitación, especies de *Candida*, *Blastomyces*, *Cryptococcus* e *Histoplasma*.

En algunas realizaciones, una meganucleasa monocatenaria puede dirigirse a una secuencia de reconocimiento dentro del genoma patógeno, por ejemplo, a un gen o elemento regulador que es esencial para el crecimiento, reproducción o toxicidad del patógeno. En ciertas realizaciones, la secuencia de reconocimiento puede estar en un plásmido bacteriano. La escisión mediada por meganucleasa de una secuencia de reconocimiento en un genoma patógeno puede estimular la mutación dentro de un gen esencial diana en forma de una inserción, deleción o desplazamiento de fase, estimulando la unión de extremos no homólogos. Como alternativa, la escisión de un plásmido bacteriano puede dar como resultado la 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, animal o planta infectado en forma de proteína o ácido nucleico usando técnicas que son habituales en este campo. En ciertas realizaciones, el gen de la meganucleasa puede incorporarse en un genoma de bacteriófago para suministro a bacterias patógenas.

Aspectos de la invención proporcionan también compuestos terapéuticos para el tratamiento de ciertas formas de cáncer. Debido a que los virus humanos con frecuencia se asocian con la 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 del cáncer. Como alternativa, las roturas bicatenarias dirigidas a los genomas de estos virus asociados a tumor usando meganucleasas monocatenarias pueden usarse para desencadenar la apoptosis mediante la ruta de respuesta a daño de ADN. De esta manera, puede ser posible inducir selectivamente apoptosis en células tumorales que albergan el genoma viral.

10. Procedimientos para genotipación e identificación de patógenos

Aspectos de la invención también proporcionan herramientas para investigación y desarrollo de biología molecular *in vitro*. Es habitual en la técnica usar endonucleasas específicas (por ejemplo, enzimas de restricción) para el aislamiento, clonación y manipulación de ácidos nucleicos tales como plásmidos, productos de PCR, secuencias de BAC, secuencias de YAC, virus y secuencias genómicas de organismos eucariotas y procariontes (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, Wiley 1999). Por lo tanto, en algunas realizaciones, puede usarse una meganucleasa monocatenaria para manipular las secuencias de ácido nucleico *in vitro*. Por ejemplo, pueden usarse meganucleasas monocatenarias 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, la presente invención proporciona herramientas para la identificación de genes patógenos y organismos. En una realización, pueden usarse meganucleasas monocatenarias para escindir sitios de reconocimiento que corresponden a regiones genéticas polimórficas correlacionadas con enfermedad para distinguir alelos que provocan enfermedad de alelos sanos (por ejemplo, una meganucleasa monocatenaria que reconoce el alelo ΔF -508 del gen CFTR humano, véase ejemplo 4). En esta realización, las secuencias de ADN aisladas de un paciente humano u otro organismo se digieren con una meganucleasa monocatenaria, posiblemente junto con nucleasas específicas de sitio adicionales, y el patrón de fragmentos de ADN resultantes se analiza por electroforesis en gel, electroforesis capilar, espectrometría de masas u otros procedimientos conocidos en la técnica. Este patrón de fragmentación y, específicamente, la presencia o ausencia de escisión por la meganucleasa monocatenaria, indica el genotipo del organismo revelando si la secuencia de reconocimiento está presente o no en el genoma. En otra realización, se dirige una meganucleasa monocatenaria 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 monocatenaria escinde una secuencia de reconocimiento que es única del patógeno (por ejemplo, la región espaciadora entre los genes de ARNr 16S y 23S en una bacteria; véase, por ejemplo, van der Giessen y col. (1994), *Microbiology* 140: 1103-1108) y puede usarse para distinguir el patógeno de otros organismos estrechamente relacionados después de digestión con endonucleasas del genoma y posterior análisis del patrón de fragmentación por electroforesis, espectrometría de masas u otros procedimientos conocidos en la técnica.

11. Procedimientos para la producción de dominios de unión a ADN adaptados

En otro aspecto, la invención proporciona proteínas de unión a ADN monocatenarias que carecen de actividad de escisión endonucleasa. La actividad catalítica de una meganucleasa monocatenaria puede eliminarse mutando los aminoácidos implicados en la catálisis (por ejemplo, la mutación de Q47 a E en I-CreI, véase Chevalier y col. (2001), *Biochemistry*. 43: 14015-14026); la mutación de D44 o D145 a N en I-SceI; la mutación de E66 a Q en I-CeuI; 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 acetiltransferasa de histona (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 cinc modificado por ingeniería genética y un dominio efector (véase, por ejemplo, Papworth y col. (2006), Gene 366:27-38).

Ejemplos

La invención se ilustra adicionalmente por los siguientes ejemplos, que no deberían interpretarse como limitantes. El ejemplo 1 presenta pruebas de que un procedimiento previamente desvelado para la producción de meganucleasas I-CreI monocatenarias (Epinat y col. (2003), Nucleic Acids Res. 31: 2952-62; documento WO 2003/078619) no es suficiente para la producción de meganucleasas que reconozcan sitios de ADN no palindrómicos. Los ejemplos 2 y 3 presentan pruebas de que el procedimiento descrito en el presente documento es suficiente para producir meganucleasas I-CreI monocatenarias que reconocen sitios de ADN no palindrómicos usando un engarce Gly-Ser flexible (ejemplo 2) o un engarce estructurado diseñado (ejemplo 3). Aunque los ejemplos 2 y 3 posteriores se refieren específicamente a meganucleasas monocatenarias basadas en I-CreI, pueden producirse y usarse de forma similar meganucleasas monocatenarias comprendidas de subunidades derivadas de I-SceI, I-MsoI, I-CeuI, I-AniI y otras meganucleasas LAGLIDADG, como se describe en el presente documento. Los ejemplos que no quedan dentro del alcance de las reivindicaciones son solamente para fines ilustrativos.

Ejemplo 1

Evaluación del procedimiento de Epinat y col.

1. Meganucleasas monocatenarias que usan el procedimiento de Epinat y col.

Epinat y col. (2003), Nucleic Acids Res. 31: 2952-62 y documento WO 2003/078619 presentan la producción de una meganucleasa monocatenaria derivada de la meganucleasa I-CreI. Específicamente, los autores usaron un engarce peptídico de 11 aminoácidos derivado de I-DmI (aminoácidos 94-104 de I-DmI, secuencia MLERIRLFNMR) para unir una subunidad I-CreI N terminal (aminoácidos 1-93 de I-CreI) con una subunidad I-CreI C terminal (aminoácidos 8-163). Este ordenamiento particular de subunidad N terminal-engarce-subunidad C terminal se seleccionó porque imitaba más estrechamente la organización de dominio de la meganucleasa I-DmI di-LAGLIDADG. Los autores evaluaron la meganucleasa I-CreI monocatenaria experimentalmente y descubrieron que escindía una secuencia de reconocimiento de I-CreI natural de forma eficaz, aunque a una tasa significativamente reducida en relación con el homodímero de I-CreI natural.

Debido a que la proteína de fusión producida por estos autores comprendía dos subunidades de otro modo naturales, ambas de las cuales reconocen semisitios de ADN idénticos, fue necesario ensayar la meganucleasa monocatenaria usando el sitio de ADN natural seudopalindrómico. Como tal, no fue posible para los autores descartar la posibilidad de que la actividad de escisión observada no se debiera a escisión por una meganucleasa monocatenaria individual sino, en su lugar, por un dímero intermolecular de dos meganucleasas monocatenarias en las que un dominio de cada una se asociaba para formar una meganucleasa funcional que se comporta de hecho como el homodímero natural. De hecho, una parte sustancial de la subunidad I-CreI N terminal (aminoácidos 94-163) se retiró en la producción de la meganucleasa monocatenaria presentada en Epinat y col. Una inspección de la estructura cristalina tridimensional de I-CreI (Jurica y col. (1998), Mol. Cell 2: 469-476) revela que este truncamiento da como resultado la retirada de tres hélices alfa de la superficie de la subunidad N terminal y la posterior exposición a disolvente de una cantidad significativa de área de superficie hidrófoba. Como tal, los presentes inventores presentaron la hipótesis de que la subunidad N terminal de la meganucleasa I-CreI monocatenaria de Epinat y col. es inestable e inactiva y que la actividad de escisión de ADN observada se debe, de hecho, a la dimerización de las subunidades C terminales de dos proteínas monocatenarias. Los problemas de estabilidad de la proteína que resultan de la aplicación del procedimiento de Epinat y col. también se analizan en Fajardo-Sanchez y col. (2008), Nucleic Acids Res. 36: 2163-2173.

2. Diseño de meganucleasas LAM monocatenarias usando el procedimiento de Epinat y col.

Para evaluar de forma más crítica el procedimiento para producción de meganucleasa I-CreI monocatenaria presentado por Epinat y col. (Epinat y col. (2003), Nucleic Acids Res. 31: 2952-62; documento WO 2003/078619), se produjo una meganucleasa monocatenaria en la que los dominios de I-CreI N y C terminal reconocen semisitios de ADN diferentes. El procedimiento presentado en Epinat y col. se usó para producir un par de meganucleasas monocatenarias que comprenden un dominio LAM1 y un dominio LAM2. Esta meganucleasa "LAM1 epLAM2" (SEQ ID NO: 48) comprende un dominio LAM1 N terminal y un dominio LAM2 C terminal mientras que "LAM2epLAM1" (SEQ ID NO: 49) comprende un dominio LAM2 N terminal y un dominio LAM1 C terminal. En total, ambas meganucleasas monocatenarias difieren en 11 aminoácidos de la presentada por Epinat y col. y todos los cambios de aminoácidos son en regiones de la enzima responsable de reconocimiento de ADN que no se espera que afecten a la interacción de subunidades.

3. Construcción de meganucleasas monocatenarias

LAM1epLAM2 y LAM2epLAM1 se produjeron por PCR de genes LAM1 y LAM2 existentes con cebadores que introducen la secuencia de engarce **I-Dmol** (que se traduce a MLERIRLFNMR) así como sitios de enzimas de restricción para clonación. Las dos subunidades LAM se clonaron secuencialmente en pET-21, un vector con un marcador de seis histidinas fusionado en el extremo 3' del gen monocatenario de longitud completa para purificación (Novagen Corp., San Diego, CA). Todos las secuencias de ácido nucleico se confirmaron usando secuenciación de Didesoxinucleótidos Sanger (véase, Sanger y col. (1977), Proc. Natl. Acad. Sci. USA. 74(12): 5463-7).

Las meganucleasas LAMep se expresaron y purificaron usando el siguiente procedimiento. Las construcciones clonadas en un vector pET21a se transformaron en pLysS de BL21 químicamente competentes (DE3), y se sembraron en placas 2xYT convencionales que contenían carbanicilina 200 µg/ml. Después de crecimiento durante una noche, las colonias bacterianas transformadas se separaron 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 del 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 mediante centrifugación durante 10 minutos a 6000x 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) mediante agitación con vórtex. Las células se rompieron después con 12 pulsos de sonicación al 50 % de potencia y el residuo celular se sedimentó por centrifugación durante 15 minutos a 14.000x 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, MgCl₂ 5 mM, EDTA 5 mM) para ensayos y almacenamiento usando columnas de desalación Zeba spin (Pierce Biotechnology, Inc., Rockford, IL). La concentración de enzima se determinó mediante absorbancia a 280 nm usando un coeficiente de extinción de 23.590 M⁻¹cm⁻¹. La pureza y peso molecular de las enzimas se confirmó después por espectrometría de masas MALDI-TOF.

4. Ensayos de escisión

Todas las enzimas purificadas como se ha descrito anteriormente se ensayaron con respecto a actividad mediante incubación con sustratos de ADN bicatenarios lineales que contenían secuencias de reconocimiento de meganucleasas. Los oligonucleótidos sintéticos correspondientes a cadenas tanto sentido como antisentido de las secuencias de reconocimiento se hibridaron y se clonaron en el sitio *SmaI* del plásmido pUC19 mediante ligación de extremos romos. Las secuencias de los sitios de unión clonados se confirmaron por secuenciación de didesoxinucleótidos de Sanger. Todos los sustratos plasmídicos se linealizaron con *XmnI* o *ScaI* simultáneamente con el producto de digestión de meganucleasa. Los productos de digestión enzimáticos contenían 5 µl de sustrato de ADN 0,05 µM, 2,5 µl de meganucleasa monocatenaria 5 µM, 9,5 µl de tampón SA, y 0,5 µl de *XmnI* o *ScaI*. Los productos de digestión se incubaron a 37 °C durante 4 horas. Los productos de digestión se detuvieron añadiendo 0,3 mg/ml de Proteinasa K y SDS 0,5 %, y se incubaron durante una hora a 37 °C. Los productos de digestión se analizaron en agarosa al 1,5 % y se visualizaron mediante tinción con bromuro de etidio.

5. Resultados

Las meganucleasas LAMep producidas usando el procedimiento de Epinat y col. se incubaron con sustratos de ADN que comprendían el palíndromo LAM1 (SEQ ID NO: 40 y 41), el palíndromo LAM2 (SEQ ID NO: 44 y 45) o el sitio híbrido LAM1/LAM2 (SEQ ID NO: 46 y 47). Se descubrió que la meganucleasa monocatenaria LAM1epLAM2 escindía principalmente el palíndromo LAM2 mientras que se descubrió que la meganucleasa monocatenaria LAM2epLAM1 escindía principalmente el palíndromo LAM1. Ninguna de las meganucleasas monocatenarias escindía el sitio híbrido en un grado significativo. Estos resultados sugieren que, de hecho, el procedimiento de Epinat y col. produce meganucleasas monocatenarias que son incapaces de escindir secuencias de ADN no palindrómicas. Se descubrió que ambas meganucleasas monocatenarias escindían principalmente la secuencia de reconocimiento correspondiente a un palíndromo del semisitio reconocido por la subunidad C terminal, lo que sugiere que la subunidad N terminal está inactiva. Por lo tanto, la especie de meganucleasa activa caracterizada por Epinat y col. parece ser principalmente un dímero entre las subunidades C terminales de un par de meganucleasas I-CreI monocatenarias. Como alternativa, la escisión del sitio de ADN palindrómico puede deberse a corte monocatenario secuencial por las subunidades C terminales de meganucleasas I-CreI monocatenarias diferentes. En cualquiera de los casos, a diferencia de las reivindicaciones realizadas por Epinat y col., el procedimiento no produce un heterodímero de I-CreI monocatenario sustancialmente funcional y generalmente no es útil para el reconocimiento y escisión de sitios de ADN no palindrómicos.

Ejemplo 2Meganucleasas I-Crel monocatenarias producidas usando un engarce Gly-Ser flexible5 1. Diseño de meganucleasas LAM monocatenarias usando un engarce Gly-Ser

Las meganucleasas LAM1 y LAM2 diseñadas se fusionaron en un único polipéptido usando el Engarce 3 de la Tabla 3. Val-151 se usó como el punto de fusión N terminal (para la subunidad LAM1) mientras que Phe-9 fue el punto de fusión C terminal (para la subunidad LAM2). La meganucleasa monocatenaria resultante, "LAM1gsLAM2" (SEQ ID NO: 50) se clonó en pET21, se expresó en *E. coli* y se purificó como se ha descrito en el Ejemplo 1.

10 2. Resultados

LAM1gsLAM2 se ensayó con respecto a actividad de escisión usando los mismos sustratos de ADN y condiciones de incubación que se describen en el Ejemplo 1. A diferencia de los resultados con las meganucleasas LAMep, se descubrió que LAM1gsLAM2 escindía principalmente la secuencia de reconocimiento LAM1/LAM2 híbrida (SEQ ID NO: 46 y 47). El alcance de la escisión se redujo significativamente en relación con el heterodímero LAM1/LAM2 producido coexpresando los monómeros LAM 1 y LAM2 en *E. coli*. En las mismas condiciones de reacción, el heterodímero escinde la secuencia de reconocimiento LAM1/LAM2 hasta su compleción, lo que sugiere que el engarce Gly-Ser altera la actividad de escisión en cierto grado. No obstante, LAM1gsLAM2 muestra una preferencia mucho más fuerte por el sitio híbrido frente a los sitios LAM1 o LAM2 palindrómicos y por lo tanto tiene utilidad para aplicaciones en las que la especificidad es de mayor importancia que la actividad.

25 **Ejemplo 3**Meganucleasas I-Crel monocatenarias producidas usando un engarce estructurado30 1. Diseño de meganucleasas LAM monocatenarias usando un engarce estructurado diseñado

Las endonucleasas LAM1 y LAM2 diseñadas se fusionaron en un único polipéptido usando el Engarce 9 de la Tabla 6. Asp-153 se usó como el punto de fusión N terminal (para la subunidad LAM1) mientras que Lys-7 fue el punto de fusión C terminal (para la subunidad LAM2). La meganucleasa monocatenaria resultante, "LAM1desLAM2" (SEQ ID NO: 51) se clonó en pET21a, se expresó en *E. coli* y se purificó como se ha descrito en el Ejemplo 1.

35 2. Resultados

LAM1desLAM2 se ensayó con respecto a actividad de escisión usando los mismos sustratos de ADN y condiciones de incubación que se han descrito en el Ejemplo 1. A diferencia de los resultados con las meganucleasas LAMep, se descubrió que LAM1desLAM2 escindía principalmente la secuencia de reconocimiento LAM1/LAM2 híbrida (SEQ ID NO: 46 y 47). El alcance de la escisión es comparable al heterodímero LAM1/LAM2 producido coexpresando los monómeros LAM 1 y LAM2 en *E. coli*. Estos resultados sugieren que los engarces estructurados diseñados, tales como el Engarce 9, no interfieren significativamente con la actividad de escisión. Además, LAM1desLAM2 es estructuralmente estable y mantiene la actividad catalítica durante > 3 semanas cuando se almacena en tampón SA a 4 °C. Resulta importante que LAM1desLAM2 muestra actividad mínima hacia los sitios LAM1 y LAM2 palindrómicos (SEQ ID NO: 40 y 41 y 44 y 45), lo que indica que la especie funcional producida por el procedimiento desvelado en el presente documento es principalmente un heterodímero monocatenario.

50 **Ejemplo 4**Meganucleasas I-Msol monocatenarias producidas usando un engarce estructurado55 1. Diseño de meganucleasas I-Msol monocatenarias usando un engarce estructurado diseñado

Se fusionó un par de subunidades de endonucleasa I-Msol (no modificadas con respecto a especificidad de escisión de ADN) en un único polipéptido usando el Engarce 30 de la Tabla 8. Se usó Ile-166 como el punto de fusión N terminal mientras que Leu-7 era el punto de fusión C terminal. La meganucleasa monocatenaria resultante, "MSOdesMSO" (SEQ ID NO: 52) se clonó en pET21a con un marcador 6xHis C terminal para facilitar la purificación. La meganucleasa se expresó después en *E. coli* y se purificó como se ha descrito en el Ejemplo 1.

60 2. Resultados

Se ensayó MSOdesMSO purificado con respecto a la capacidad para escindir un sustrato plasmídico que alberga la secuencia de reconocimiento de I-Msol natural (SEQ ID NO: 53 y SEQ ID NO: 54 y 54) en las condiciones de incubación como se ha descrito en el Ejemplo 1. Se descubrió que la enzima tenía actividad de escisión comparable al homodímero de I-Msol (que, en este caso, se esperaba que reconociera y cortara la misma secuencia de reconocimiento que MSOdesMSO). Los análisis de SDS-PAGE revelaron que MSOdesMSO tiene un peso molecular

aparente de ~40 kilodalton, coherente con que sea un par de subunidades I-Msol unidas covalentemente, y no fueron evidentes productos de degradación proteica. Estos resultados indican que la invención es adecuada para la producción de meganucleasas monocatenarias de alta actividad, estables, derivadas de I-Msol.

5

TABLA 11
Modificaciones de I-Crel del documento WO 2007/047859

Base de Cadena Sentido Favorecida											
Posn.	A	C	G	T	A/T	A/C	A/G	C/T	G/T	A/G/T	A/C/G/T
-1	Y75 L75* C75* Y139* C46* A46*	R70* H75* R75* H46* K46* R46*	K70 E70* E75* E46* D46*	Q70* C70 L70 Y75* Q75* H75* H139 Q46* H46*				T46*			G70 A70 S70 G46*
-2	Q70 T44* A44* V44* I44* L44* N44*	E70 D70 K44* R44*	H70 D44* E44*	Q44*	C44*						
-3	Q68 C24* I24*	E68 F68 K24* R24*	R68	M68 C68 L68 F68		H68		Y68	K68		
-4	A26* Q77	E77 K26*	R77 E26*					S77 Q26*			S26*
-5		E42	R42			K28*	C28* Q42				M66 K66
-6	Q40 C28*	E40 R28*	R40	C40 I40 V40 C79 I79 V79 Q28*	A40 A79 A28* H28*						S40 S28*
-7	N30* Q38	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 I32			S32 N32 H32 Q32 T32

Las entradas en negrita son restos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento.

10 Un asterisco indica que el resto entra en contacto con la base en la cadena antisentido.

TABLA 12
Modificaciones de I-Msol del documento WO 2007/047859

Posición	Base de Cadena Sentido Favorecida			
	A	C	G	T
-1	K75* Q77 A49* C49* K79*	D77 E77 K49* R75* K75* R79* K79*	K77 R77 E49* E79*	C77 L77 Q79*
-2	Q75 K81 C47* I47* L47*	E75 D75 R47* K47* K81* R81*	K75 E47* E81*	A75 C75 V75 I75 T75 Q47* Q81*
-3	Q72 C26* L26* V26* A26* I26*	E72 Y72 H26* K26* R26*	R72 K72 Y26* F26*	K72 Y72 H26*
-4	K28 Q83	K28* R28* E83	R83 K83	K28 K83 Q28*
-5	K28 C28* L28* I28*	K28* R28*	R45 E28*	Q28*
-6	I30* V30* S30* L30* Q43	E43 E85 K30* R30*	R43 K43 K85 R85 E30* D30*	K43 I85 V85 L85 Q30*
-7	Q41	E32 E41	R32 R41 K41	K32 M41 L41 I41
-8	Y35 K35	E32	R32 K32 K35 R35	K32 K35
-9	N34 H34	D34 E34 S34	K34 R34 H34	S34 C34 V34 T34 A34

5 Las entradas en negrita representan restos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento. Un asterisco indica que el resto entra en contacto con la base en la cadena antisentido.

TABLA 13
Modificaciones de I-CeU del documento WO 2007/047859

Posición	Base de Cadena Sentido Favorecida			
	A	C	G	T
-1	C92* A92* V92*	K116* R116* D116* K92*	E116* E92*	Q116* Q92*

(continuación)

Base de Cadena Sentido Favorecida				
Posición	A	C	G	T
-2	Q117 C90* L90* V90*	E117 D117 R174* K124* K90* R90* K68*	K117 R124 K124 E124* E90* D90*	C117 V117 T117 Q90*
-3	C70* V70* T70* L70* K70*	K70*	E70* E88*	Q70*
-4	Q126 N126 K88* L88* C88* C72* L72* V72*	E126 D126 R88* K88* K72*	R126 K126 E88* D88*	K126 L126 Q88*
-5	C74* L74* V74* T74*	K74*	E74* K128 R128 E128	C128 L128 V128 T128
-6	Q86	D86 E86 R84* K84*	K128 R128 R86 K86 E84*	K86 C86 L86
-7	L76* C76* K76*	R76* K76* H76*	E76* R84	H76* Q76*
-8	Y79 R79 Q76	D79 E79 D76 E76	R79 K79 K76 R76	C79 L79 V79 L76
-9	Q78 N78 H78 K78	D78 E78	R78 K78 H78	K78 V78 L78 C78 T78

Las entradas en negrita son restos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento.

- 5 Un asterisco indica que el resto entra en contacto con la base en la cadena antisentido.

TABLA 14
Modificaciones de I-Scel del documento WO 2007/047859

Base de Cadena Sentido Favorecida				
Posición	A	C	G	T
4	K50	R50* K50* E57	E50* R57 K57	K57 M57 Q50*
5	K48 Q102	R48* K48* E102 E59	E48* K102 R102	Q48* C102 L102 V102
6	K59	R59* K59*	K84 E59*	Q59* Y46

(continuación)

Posición	Base de Cadena Sentido Favorecida			
	A	C	G	T
7	C46* L46* V46*	R46* K46* E86	K86 R86 E46*	K68 C86 L86 Q46*
8	K61* S61* V61* A61* L61*	E88 R61* H61*	E61* R88 K88	K88 Q61* H61*
9	T98* C98* V98* L98*	R98* K98*	E98* D98*	Q98*
10	V96* C96* A96*	K96* R96*	D96* E96*	Q96*
11	C90* L90*	K90* R90*	E90*	Q90*
12	Q193	E165 E193 D193	K165 R165	C165 L165 C193 V193 A193 T193 S193
13	C193* L193*	K193* R193* D192	E193* D193* K163 R192	Q193* C163 L163
14	L192* C192*	E161 R192* K192*	K147 K161 R161 R197 D192* E192*	K161 Q192*
15		E151	K151	C151 L151 K151
17	N152* S152* C150* L150* V150* T150*	K152* K150*	N152* S152* D152* D150* E150*	Q152* Q150*
18	K155* C155*	R155* K155*	E155*	H155* Y155*

Las entradas en negrita son restos de contacto naturales y no constituyen "modificaciones" como se usa en el presente documento.

5 Un asterisco indica que el resto entra en contacto con la base en la cadena antisentido.

Se desvelan los siguientes artículos:

10 1. Una meganucleasa monocatenaria recombinante que comprende:

una primera subunidad LAGLIDADG derivada de una primera meganucleasa mono-LAGLIDADG, teniendo dicha primera subunidad LAGLIDADG un primer semisitio de reconocimiento;
una segunda subunidad LAGLIDADG derivada de una segunda meganucleasa mono-LAGLIDADG o una meganucleasa di-LAGLIDADG, teniendo dicha segunda subunidad LAGLIDADG un segundo semisitio de reconocimiento;

15

en la que dichas primera y segunda subunidades LAGLIDADG se unen covalentemente por un engarce polipeptídico de modo que dicho primer dominio LAGLIDADG está en sentido N-terminal con respecto a dicho engarce y dicho segundo dominio LAGLIDADG está en sentido C-terminal con respecto a dicho engarce; y en la que dichas primera y segunda subunidades LAGLIDADG son capaces de actuar juntos para reconocer y escindir una secuencia de ADN no palindrómica que es un híbrido de dicho primer semisitio de reconocimiento y dicho segundo semisitio de reconocimiento.

2. La meganucleasa monocatenaria recombinante del artículo 1 en la que:

la primera subunidad LAGLIDADG deriva de una meganucleasa mono-LAGLIDADG seleccionada del grupo que consiste en I-Crel, I-Msol y I-Ceul; y la segunda subunidad LAGLIDADG deriva de (1) una meganucleasa mono-LAGLIDADG seleccionada del grupo que consiste en I-Crel, I-Msol y I-Ceul, o (2) una meganucleasa di-LAGLIDADG seleccionada del grupo que consiste en I-Dmol, I-Scel y I-Anil.

3. La meganucleasa monocatenaria recombinante del artículo 1 en la que:

la primera subunidad LAGLIDADG deriva de una especie diferente que la segunda subunidad LAGLIDADG.

4. La meganucleasa monocatenaria recombinante del artículo 1 en la que:

dicha primera subunidad LAGLIDADG comprende una secuencia polipeptídica que tiene al menos 85 % de identidad de secuencia con un primer dominio LAGLIDADG seleccionado del grupo que consiste en los restos 9-151 de una meganucleasa I-Crel de tipo silvestre de SEQ ID NO: 1; los restos 11-162 de una meganucleasa I-Msol de tipo silvestre de SEQ ID NO: 2; y restos 55-210 de una meganucleasa I-Ceul de tipo silvestre de SEQ ID NO: 3.

5. La meganucleasa monocatenaria recombinante del artículo 2 en la que:

dicha segunda subunidad LAGLIDADG comprende una secuencia polipeptídica que tiene al menos 85 % de identidad de secuencia con un segundo dominio LAGLIDADG seleccionado del grupo que consiste en los restos 9-151 de una meganucleasa I-Crel de tipo silvestre de SEQ ID NO: 1; restos 11-162 de una meganucleasa I-Msol de tipo silvestre de SEQ ID NO: 2; restos 55-210 de una meganucleasa I-Ceul de tipo silvestre de SEQ ID NO: 3; restos 9-96 de una I-Dmol de tipo silvestre de SEQ ID NO: 4; restos 105-178 de una I-Dmol de tipo silvestre de SEQ ID NO: 4; restos 32-123 de una I-Scel de tipo silvestre de SEQ ID NO: 5; restos 134-225 de una I-Scel de tipo silvestre de SEQ ID NO: 5; restos 4-121 de una I-Anil de tipo silvestre de SEQ ID NO: 6; y restos 136-254 de una I-Anil de tipo silvestre de SEQ ID NO: 6.

6. La meganucleasa monocatenaria recombinante del artículo 2 en la que:

cada una de dichas subunidades LAGLIDADG comprende al menos 85 % de identidad con un dominio LAGLIDADG seleccionado de forma independiente del grupo que consiste en los restos 9-151 de una meganucleasa I-Crel de tipo silvestre de SEQ ID NO: 1; restos 11-162 de una meganucleasa I-Msol de tipo silvestre de SEQ ID NO: 2; restos 55-210 de una meganucleasa I-Ceul de tipo silvestre de SEQ ID NO: 3; restos 9-96 de una I-Dmol de tipo silvestre de SEQ ID NO: 4; restos 105-178 de una I-Dmol de tipo silvestre de SEQ ID NO: 4; restos 32-123 de una I-Scel de tipo silvestre de SEQ ID NO: 5; restos 134-225 de una I-Scel de tipo silvestre de SEQ ID NO: 5; restos 4-121 de una I-Anil de tipo silvestre de SEQ ID NO: 6; y restos 136-254 de una I-Anil de tipo silvestre de SEQ ID NO: 6; y

al menos uno de dichos dominios LAGLIDADG comprende al menos una modificación de aminoácidos desvelada en cualquiera de las Tablas 11, 12, 13 y 14.

7. La meganucleasa monocatenaria recombinante del artículo 6 en la que:

al menos un dominio LAGLIDADG deriva de I-Crel y al menos una modificación se selecciona de la Tabla 1 de cualquiera de las Tablas 11, 12, 13 y 14;

al menos un dominio LAGLIDADG deriva de I-Msol y al menos una modificación se selecciona de la Tabla 12;

al menos un dominio LAGLIDADG deriva de I-Ceul y al menos una modificación se selecciona de la Tabla 13;

o

al menos un dominio LAGLIDADG deriva de I-Scel y al menos una modificación se selecciona de la Tabla 14.

8. La meganucleasa monocatenaria recombinante del artículo 2 en la que:

cada una de dichas subunidades LAGLIDADG tiene un semisitio de reconocimiento seleccionado del grupo que consiste en SEQ ID NO: 7-30.

9. La meganucleasa monocatenaria recombinante del artículo 8 en la que:

al menos una de dichas subunidades LAGLIDADG tiene un semisitio de reconocimiento seleccionado del

grupo que consiste en SEQ ID NO: 7-30; y la otra de dichas subunidades LAGLIDADG tiene un semisito de reconocimiento que difiere en una modificación de al menos un par de bases de un semisito de reconocimiento seleccionado del grupo que consiste en SEQ ID NO: 7-30.

- 5
10. La meganucleasa monocatenaria recombinante de una cualquiera de los artículos 1-9 en la que: dicho engarce polipeptídico es un engarce flexible.
- 10
11. La meganucleasa monocatenaria recombinante del artículo 10 en la que: dicho engarce comprende 15-40 restos.
12. La meganucleasa monocatenaria recombinante del artículo 10 en la que: dicho engarce comprende 25-31 restos.
- 15
13. La meganucleasa monocatenaria recombinante del artículo 10 en la que: al menos 50 % de dicho engarce comprende restos con carga no polares.
14. La meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-9 en la que: dicho engarce polipeptídico tiene una estructura secundaria estable.
- 20
15. La meganucleasa monocatenaria recombinante del artículo 14 en la que: dicha estructura secundaria estable comprende al menos dos estructuras de hélice α .
- 25
16. La meganucleasa monocatenaria recombinante del artículo 14 en la que: dicha estructura secundaria estable comprende del extremo N al C terminal un primer bucle, una primera hélice α , una primera vuelta, una segunda hélice α y un segundo bucle.
- 30
17. La meganucleasa monocatenaria recombinante del artículo 14 en la que: dicho engarce comprende 23-56 restos.
- 35
18. Un método para producir una célula eucariota modificada genéticamente que incluye una secuencia exógena de interés insertada en un cromosoma de dicha célula eucariota, que comprende:
- 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 de interés;
- en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión; y
- en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.
- 45
19. Un método como en el artículo 18 en el que: dicho segundo ácido nucleico comprende además secuencias homólogas de secuencias que flanquean dicho sitio de escisión y dicha secuencia de interés se inserta en dicho sitio de escisión por recombinación homóloga.
- 50
20. Un método como en el artículo 18 en el que: dicho segundo ácido nucleico carece de homología sustancial con dicho sitio de escisión y dicha secuencia de interés se inserta en dicho cromosoma por unión de extremos no homólogos.
- 55
21. Un método para producir una célula eucariota modificada genéticamente que incluye una secuencia exógena de interés insertada en un cromosoma de dicha célula eucariota, que comprende:
- introducir una proteína meganucleasa en una célula eucariota; y
- transfectar dicha célula eucariota con un ácido nucleico que incluye dicha secuencia de interés;
- en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión; y
- en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.
- 60
- 65
22. Un método como en el artículo 21 en el que: dicho ácido nucleico comprende además secuencias homólogas de secuencias que flanquean dicho sitio de escisión y dicha secuencia de interés se inserta en dicho sitio de escisión por recombinación homóloga.

23. Un método como en el artículo 21 en el que:
dicho ácido nucleico carece de homología sustancial con dicho sitio de escisión y dicha secuencia de interés se inserta en dicho cromosoma por unión de extremos no homólogos.

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

10 transfectar una célula eucariota con un ácido nucleico que codifica una meganucleasa;
en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia diana se altera por unión de extremos no homólogos en dicho sitio de escisión; y
10 en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

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

15 producir una célula eucariota modificada genéticamente de acuerdo con el método de uno cualquiera de los artículos 18-24; y
20 cultivar dicha célula eucariota modificada genéticamente para producir dicho organismo modificado genéticamente.

25 26. Un método como en el artículo 25 en el que:
dicha célula eucariota se selecciona del grupo que consiste en un gameto, un cigoto, un blastocisto, una célula madre embrionaria y una célula protoplástica.

25 27. Un método para tratar una enfermedad por terapia génica en un eucariota, que comprende:

transfectar al menos una célula de dicho eucariota con uno o más ácidos nucleicos que incluyen

30 (i) una primera secuencia de ácido nucleico que codifica una meganucleasa, y
 (ii) una segunda secuencia de ácido nucleico que incluye una secuencia de interés;

35 en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión;
en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
en el que la inserción de dicha secuencia de interés proporciona dicha terapia génica para dicha enfermedad, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

40 28. Un método como en el artículo 27 en el que:

dicha segunda secuencia de ácido nucleico comprende además secuencias homólogas de secuencias que flanquean dichos sitios de escisión y dicha secuencia de interés se inserta en dicho sitio de escisión por recombinación homóloga.

45 29. Un método como en el artículo 27 en el que:
dicha segunda secuencia de ácido nucleico carece de homología sustancial con dicho sitio de escisión y dicha secuencia de interés se inserta en dicho cromosoma por unión de extremos no homólogos.

50 30. Un método para tratar una enfermedad por terapia génica en un eucariota, que comprende:

55 introducir una proteína meganucleasa en al menos una célula de dicho eucariota; y
transfectar dicha célula eucariota con un ácido nucleico que incluye una secuencia de interés;
en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión;
en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
en el que la inserción de dicha secuencia de interés proporciona dicha terapia génica para dicha enfermedad,
60 en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

31. Un método como en el artículo 30 en el que:
dicho ácido nucleico comprende además secuencias homólogas de secuencias que flanquean dicho sitio de escisión y dicha secuencia de interés se inserta en dicho sitio de escisión por recombinación homóloga.

65 32. Un método como en el artículo 30 en el que:
dicho ácido nucleico carece de homología sustancial con dicho sitio de escisión y dicha secuencia de interés se

inserta en dicho cromosoma por unión de extremos no homólogos.

33. Un método para tratar una enfermedad por terapia génica en un eucariota alterando una secuencia diana en un cromosoma de dicha célula eucariota, que comprende:

5 transfectar al menos una célula de dicho eucariota con un ácido nucleico que codifica una meganucleasa; en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia diana se altera por unión de extremos no homólogos en dicho sitio de escisión; en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
10 en el que la alteración de dicha secuencia diana proporciona dicha terapia génica para dicha enfermedad, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

34. Un método para tratar una infección por patógeno viral en un hospedador eucariota alterando una secuencia diana en un genoma de dicho patógeno viral, que comprende:

15 transfectar al menos una célula infectada de dicho hospedador eucariota con un ácido nucleico que codifica una meganucleasa; en el que dicha meganucleasa produce un sitio de escisión en dicho genoma viral y dicha secuencia diana se altera por unión de extremos no homólogos en dicho sitio de escisión; en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
20 en el que la alteración de dicha secuencia diana proporciona tratamiento para dicha infección, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

35. Un método para tratar una infección por patógeno viral en un hospedador eucariota alterando una secuencia diana en un genoma de dicho patógeno viral, que comprende:

30 transfectar al menos una célula infectada de dicho hospedador eucariota con un primer ácido nucleico que codifica una meganucleasa y un segundo ácido nucleico; en el que dicha meganucleasa produce un sitio de escisión en dicho genoma viral y dicha secuencia diana se altera por recombinación homóloga de dicho genoma viral y dicho segundo ácido nucleico en dicho sitio de escisión; en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
35 en el que dicho segundo ácido nucleico comprende secuencias homólogas de secuencias que flanquean dicho sitio de escisión; y en el que la alteración dicha secuencia diana proporciona tratamiento para dicha infección, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

36. Un método para tratar una infección por patógeno procariota en un hospedador eucariota alterando una secuencia diana en un genoma de dicho patógeno procariota, que comprende:

45 transfectar al menos una célula de dicho patógeno procariota que infecta dicho hospedador eucariota con un ácido nucleico que codifica una meganucleasa; en el que dicha meganucleasa produce un sitio de escisión en dicho genoma procariota y dicha secuencia diana se altera por unión de extremos no homólogos en dicho sitio de escisión; en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; y
50 en el que la alteración de dicha secuencia diana proporciona tratamiento para dicha infección, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

37. Un método para tratar una infección por patógeno procariota en un hospedador eucariota alterando una secuencia diana en un genoma de dicho patógeno procariota, que comprende:

55 transfectar al menos una célula de dicho patógeno procariota que infecta dicho hospedador eucariota con un primer ácido nucleico que codifica una meganucleasa y un segundo ácido nucleico; en el que dicha meganucleasa produce un sitio de escisión en dicho genoma procariota y dicha secuencia diana se altera por recombinación homóloga de dicho genoma procariota y dicho segundo ácido nucleico en dicho sitio de escisión; en el que dicha meganucleasa es una meganucleasa monocatenaria recombinante de uno cualquiera de los artículos 1-17; en el que dicho segundo ácido nucleico comprende secuencias homólogas de secuencias que flanquean dicho sitio de escisión; y
60 en el que la alteración de dicha secuencia diana proporciona tratamiento para dicha infección, en el que el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

LISTADO DE SECUENCIAS

<110> PRECISION BIOSCIENCES, Inc.

5 <120> MEGANUCLEASAS MONOCATENARIAS DISEÑADAS RACIONALMENTE CON SECUENCIAS DE RECONOCIMIENTO NO PALINDRÓMICAS

<130> 111-062T1T1

10 <140>

<141> 31-10-2008

<150> EP 13 165 733.0

<151> 31-10-2008

15

<150> EP 08 845 549.8

<151> 31-10-2008

<150> PCT/US2008/82072

20

<151> 31-10-2008

<150> 61/001,247

<151> 31-10-2007

25

<160> 110

<170> PatentIn versión 3.5

<210> 1

30

<211> 163

<212> PRT

<213> *Chlamydomonas reinhardtii*

<400> 1

35

ES 2 732 735 T3

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 Thr 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

- 5 <210> 2
- <211> 170
- <212> PRT
- <213> *Monomastix sp.*

- 10 <400> 2

ES 2 732 735 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

<210> 3

<211> 218

5 <212> PRT

<213> *Chlamydomonas moewusii*

<400> 3

ES 2 732 735 T3

Met Ser Asn Phe Ile Leu Lys Pro Gly Glu Lys Leu Pro Gln Asp Lys
 1 5 10 15

Leu Glu Glu Leu Lys Lys Ile Asn Asp Ala Val Lys Lys Thr Lys Asn
 20 25 30

Phe Ser Lys Tyr Leu Ile Asp Leu Arg Lys Leu Phe Gln Ile Asp Glu
 35 40 45

Val Gln Val Thr Ser Glu Ser Lys Leu Phe Leu Ala Gly Phe Leu Glu
 50 55 60

Gly Glu Ala Ser Leu Asn Ile Ser Thr Lys Lys Leu Ala Thr Ser Lys
 65 70 75 80

Phe Gly Leu Val Val Asp Pro Glu Phe Asn Val Thr Gln His Val Asn
 85 90 95

Gly Val Lys Val Leu Tyr Leu Ala Leu Glu Val Phe Lys Thr Gly Arg
 100 105 110

Ile Arg His Lys Ser Gly Ser Asn Ala Thr Leu Val Leu Thr Ile Asp
 115 120 125

Asn Arg Gln Ser Leu Glu Glu Lys Val Ile Pro Phe Tyr Glu Gln Tyr
 130 135 140

Val Val Ala Phe Ser Ser Pro Glu Lys Val Lys Arg Val Ala Asn Phe
 145 150 155 160

Lys Ala Leu Leu Glu Leu Phe Asn Asn Asp Ala His Gln Asp Leu Glu
 165 170 175

Gln Leu Val Asn Lys Ile Leu Pro Ile Trp Asp Gln Met Arg Lys Gln
 180 185 190

Gln Gly Gln Ser Asn Glu Gly Phe Pro Asn Leu Glu Ala Ala Gln Asp
 195 200 205

Phe Ala Arg Asn Tyr Lys Lys Gly Ile Lys
 210 215

<210> 4

<211> 194

5 <212> PRT

<213> *Desulfurococcus mobilis*

<400> 4

ES 2 732 735 T3

Met His Asn Asn Glu Asn Val Ser Gly Ile Ser Ala Tyr Leu Leu Gly
 1 5 10 15

Leu Ile Ile Gly Asp Gly Gly Leu Tyr Lys Leu Lys Tyr Lys Gly Asn
 20 25 30

Arg Ser Glu Tyr Arg Val Val Ile Thr Gln Lys Ser Glu Asn Leu Ile
 35 40 45

Lys Gln His Ile Ala Pro Leu Met Gln Phe Leu Ile Asp Glu Leu Asn
 50 55 60

Val Lys Ser Lys Ile Gln Ile Val Lys Gly Asp Thr Arg Tyr Glu Leu
 65 70 75 80

Arg Val Ser Ser Lys Lys Leu Tyr Tyr Tyr Phe Ala Asn Met Leu Glu
 85 90 95

Arg Ile Arg Leu Phe Asn Met Arg Glu Gln Ile Ala Phe Ile Lys Gly
 100 105 110

Leu Tyr Val Ala Glu Gly Asp Lys Thr Leu Lys Arg Leu Arg Ile Trp
 115 120 125

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

Leu Gly Val Arg Asn Thr Ile His Leu Asp Asp His Arg His Gly Val
 145 150 155 160

Tyr Val Leu Asn Ile Ser Leu Arg Asp Arg Ile Lys Phe Val His Thr
 165 170 175

Ile Leu Ser Ser His Leu Asn Pro Leu Pro Pro Glu Arg Ala Gly Gly
 180 185 190

Tyr Thr

<210> 5

<211> 235

5 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 5

ES 2 732 735 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> 6
 <211> 254
 5 <212> PRT

<213> *Emericella nidulans*

<400> 6

Met Ser Asp Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe Glu Gly Asp
 1 5 10 15

Gly Tyr Phe Ser Ile Thr Lys Lys Gly Lys Tyr Leu Thr Tyr Glu Leu
 20 25 30

Gly Ile Glu Leu Ser Ile Lys Asp Val Gln Leu Ile Tyr Lys Ile Lys
 35 40 45

Lys Ile Leu Gly Ile Gly Ile Val Ser Phe Arg Lys Ile Asn Glu Ile
 50 55 60

Glu Met Val Ala Leu Arg Ile Arg Asp Lys Asn His Leu Lys Ser Phe
 65 70 75 80

Ile Leu Pro Ile Phe Glu Lys Tyr Pro Met Phe Ser Asn Lys Gln Tyr
 85 90 95

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

Glu Asp Leu Pro Asp Tyr Thr Arg Ser Asp Glu Pro Leu Asn Ser Ile
 115 120 125

Glu Ser Ile Ile Asn Thr Ser Tyr Phe Ser Ala Trp Leu Val Gly Phe
 130 135 140

Ile Glu Ala Glu Gly Cys Phe Ser Val Tyr Lys Leu Asn Lys Asp Asp
 145 150 155 160

Asp Tyr Leu Ile Ala Ser Phe Asp Ile Ala Gln Arg Asp Gly Asp Ile
 165 170 175

Leu Ile Ser Ala Ile Arg Lys Tyr Leu Ser Phe Thr Thr Lys Val Tyr
 180 185 190

Leu Asp Lys Thr Asn Cys Ser Lys Leu Lys Val Thr Ser Val Arg Ser
 195 200 205

Val Glu Asn Ile Ile Lys Phe Leu Gln Asn Ala Pro Val Lys Leu Leu
 210 215 220

Gly Asn Lys Lys Leu Gln Tyr Leu Leu Trp Leu Lys Gln Leu Arg Lys
 225 230 235 240

Ile Ser Arg Tyr Ser Glu Lys Ile Lys Ile Pro Ser Asn Tyr
 245 250

<210> 7
 <211> 9
 5 <212> ADN
 <213> *Chlamydomonas reinhardtii*

 <400> 7
 gaaactgctc 9
 10
 <210> 8
 <211> 9
 <212> ADN
 <213> *Chlamydomonas reinhardtii*
 15
 <400> 8
 gacagtttc 9

 <210> 9
 <211> 9
 <212> ADN
 <213> *Chlamydomonas reinhardtii*
 20
 <400> 9
 caaaacgctc 9
 25
 <210> 10
 <211> 9
 <212> ADN
 <213> *Chlamydomonas reinhardtii*
 30
 <400> 10
 gacgttttg 9
 35
 <210> 11
 <211> 9
 <212> ADN
 <213> *Monomastix sp.*
 40
 <400> 11
 cagaacgctc 9

 <210> 12
 <211> 9
 <212> ADN
 <213> *Monomastix sp.*
 45
 <400> 12
 gacgttctg 9
 50
 <210> 13
 <211> 9
 <212> ADN
 <213> *Monomastix sp.*
 55
 <400> 13
 ggaactgctc 9

 <210> 14
 <211> 9
 <212> ADN
 <213> *Monomastix sp.*
 60

	<400> 14 gacagttcc	9
5	<210> 15 <211> 9 <212> ADN <213> <i>Chlamydomonas moewusii</i>	
10	<400> 15 ataacggtc	9
15	<210> 16 <211> 9 <212> ADN <213> <i>Chlamydomonas moewusii</i>	
20	<400> 16 gaccgttat	9
25	<210> 17 <211> 9 <212> ADN <213> <i>Chlamydomonas moewusii</i>	
30	<400> 17 ttcgctacc	9
35	<210> 18 <211> 9 <212> ADN <213> <i>Chlamydomonas moewusii</i>	
40	<400> 18 ggtagcgaa	9
45	<210> 19 <211> 5 <212> ADN <213> <i>Saccharomyces cerevisiae</i>	
50	<400> 19 taggg	5
55	<210> 20 <211> 5 <212> ADN <213> <i>Saccharomyces cerevisiae</i>	
60	<400> 20 cccta	5
65	<210> 21 <211> 9 <212> ADN <213> <i>Saccharomyces cerevisiae</i>	
	<400> 21 taatgggac	9
	<210> 22 <211> 9 <212> ADN <213> <i>Saccharomyces cerevisiae</i>	
	<400> 22 gtcccatta	9

	<210> 23	
	<211> 8	
	<212> ADN	
5	<213> <i>Desulfurococcus mobilis</i>	
	<400> 23	
	gccggaac	8
10	<210> 24	
	<211> 8	
	<212> ADN	
	<213> <i>Desulfurococcus mobilis</i>	
15	<400> 24	
	gttccggc	8
	<210> 25	
	<211> 7	
20	<212> ADN	
	<213> <i>Desulfurococcus mobilis</i>	
	<400> 25	
	aacggcc	7
25	<210> 26	
	<211> 7	
	<212> ADN	
	<213> <i>Desulfurococcus mobilis</i>	
30	<400> 26	
	ggccggt	7
	<210> 27	
	<211> 8	
	<212> ADN	
	<213> <i>Emericella nidulans</i>	
35	<400> 27	
	tttacaga	8
	<210> 28	
	<211> 8	
	<212> ADN	
	<213> <i>Emericella nidulans</i>	
45	<400> 28	
	tctgtaa	8
	<210> 29	
	<211> 9	
	<212> ADN	
	<213> <i>Emericella nidulans</i>	
50	<400> 29	
	ctgaggagg	9
	<210> 30	
	<211> 9	
	<212> ADN	
	<213> <i>Emericella nidulans</i>	
60	<400> 30	
	cctcctcag	9
65	<210> 31	

<211> 163
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 31

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 Lys Ala Gln Ile Lys Pro Glu Gln Asn
 20 25 30

Arg Lys Phe Lys His Arg Leu Glu Leu Thr 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 Tyr Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
 65 70 75 80

10 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

15 <210> 32
 <211> 22
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<400> 32

aggcatctca ttagatgc ct

22

<210> 33
 <211> 22
 <212> ADN
 5 <213> Secuencia artificial

 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

 10 <400> 33
 aggcattctct aatgagatgc ct 22

 <210> 34
 <211> 22
 15 <212> ADN
 <213> *Chlamydomonas reinhardtii*

 <400> 34
 20 gaaactgtct caccgacgttt tg 22

 <210> 35
 <211> 22
 <212> ADN
 <213> *Chlamydomonas reinhardtii*
 25
 <400> 35
 caaaaacgtcg tgagacagtt tc 22

 <210> 36
 30 <211> 163
 <212> PRT
 <213> Secuencia artificial

 <220>
 35 <223> Descripción de la secuencia artificial: Polipéptido sintético

 <400> 36

ES 2 732 735 T3

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 Asp Pro Arg Gln Cys
 20 25 30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg 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

<210> 37

<211> 163

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

10

<400> 37

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 Glu Gln Ser
 20 25 30

Tyr Lys Phe Lys His Arg Leu Arg Leu Glu 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

- <210> 38
- <211> 9
- 5 <212> ADN
- <213> Secuencia artificial
- <220>
- 10 <223> Descripción de la secuencia artificial: Oligonucleótido sintético
- <400> 38
- tgcggtgc 9
- <210> 39
- 15 <211> 9
- <212> ADN
- <213> Secuencia artificial
- <223> Descripción de la secuencia artificial: Oligonucleótido sintético
- 20 <400> 39
- gacaccgca 9
- <210> 40

<211> 22
 <212> ADN
 <213> Secuencia artificial

5 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<220>
 <221> modified_base
 10 <222> (10)..(13)
 <223> a, c, t, g, u, desconocido u otro

<400> 40
 15 tgcggtgctn nnngacaccg ca 22

<210> 41
 <211> 22
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<220>
 25 <221> modified_base
 <222> (10)..(13)
 <223> a, c, t, g, u, desconocido u otro

<400> 41
 30 tgcggtgctn nnngacaccg ca 22

<210> 42
 <211> 9
 <212> ADN
 35 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

40 <400> 42
 caggctgctc 9

<210> 43
 <211> 9
 45 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

50 <400> 43
 gacagcctg 9

<210> 44
 <211> 22
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<220>
 <221> modified_base
 <222> (10)..(13)
 65 <223> a, c, t, g, u, desconocido u otro

<400> 44
 caggctgtcn nnngacagcc tg 22

5 <210> 45
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

15 <220>
 <221> modified_base
 <222> (10)..(13)
 <223> a, c, t, g, u, desconocido u otro

<400> 45
 caggctgtcn nnngacagcc tg 22

20 <210> 46
 <211> 22
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

30 <400> 46
 tgcggtgtca ttagacagcc tg 22

35 <210> 47
 <211> 22
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

40 <400> 47
 caggctgtct aatgacaccg ca 22

45 <210> 48
 <211> 260
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

50 <400> 48

ES 2 732 735 T3

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 Asp Pro Arg Gln Cys
 20 25 30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg 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 Met Leu Glu
 85 90 95

Arg Ile Arg Leu Phe Asn Met Arg Glu Phe Leu Leu Tyr Leu Ala Gly
 100 105 110

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln
 115 120 125

Ser Tyr Lys Phe Lys His Arg Leu Arg Leu Glu Phe Gln Val Thr Gln
 130 135 140

Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
 145 150 155 160

Val Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser
 165 170 175

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 180 185 190

Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg
 195 200 205

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 210 215 220

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 225 230 235 240

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 245 250 255

Lys Ser Ser Pro
 260

<210> 49

<211> 260

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

10

<400> 49

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 Glu Gln Ser
 20 25 30

Tyr Lys Phe Lys His Arg Leu Arg Leu Glu 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 Met Leu Glu
 85 90 95

Arg Ile Arg Leu Phe Asn Met Arg Glu Phe Leu Leu Tyr Leu Ala Gly
 100 105 110

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln
 115 120 125

Cys Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln
 130 135 140

Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
 145 150 155 160

Val Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser
 165 170 175

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 180 185 190

Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg
 195 200 205

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 210 215 220

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 225 230 235 240

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 245 250 255

Lys Ser Ser Pro
 260

<210> 50

<211> 334

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

10

<400> 50

ES 2 732 735 T3

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 Asp Pro Arg Gln Cys
 20 25 30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg 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 Gly Ser Ser Gly Ser Ser Gly Ser Ser
 145 150 155 160

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 165 170 175

Ser Ser Gly Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly
 180 185 190

Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser Tyr Lys Phe Lys His
 195 200 205

Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp
 210 215 220

Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp
 225 230 235 240

Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His
 245 250 255

Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln
 260 265 270

Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu Pro Ser Ala Lys Glu
 275 280 285

Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala
 290 295 300

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

Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
 325 330

<210> 51
 <211> 350
 <212> PRT
 <213> Secuencia artificial

5

<220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 51

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 Asp Pro Arg Gln Cys
 20 25 30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg 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 Pro Gly Ser Val Gly
 145 150 155 160

Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser
 165 170 175

Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala Gly Ala Thr Lys
 180 185 190

Ser Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly
 195 200 205

Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser Tyr Lys Phe Lys His
 210 215 220

Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp
 225 230 235 240

Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp
 245 250 255

Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His
 260 265 270

Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln
 275 280 285

Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu Pro Ser Ala Lys Glu
 290 295 300

Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala
 305 310 315 320

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

Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
 340 345 350

<210> 52
 <211> 373
 5 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

10 <400> 52

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

ES 2 732 735 T3

			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	Gly	Gly	Ala	Ser	Pro	Ser	Gln	Ala	Ser	Ser
				165					170					175	
Ala	Ala	Ser	Ser	Ala	Ser	Ser	Ala	Ala	Ser	Ser	Pro	Gly	Ser	Gly	Pro
			180						185				190		
Ser	Glu	Ala	Leu	Arg	Ala	Ala	Ser	Ser	Phe	Ala	Ser	Lys	Pro	Gly	Ser
		195					200					205			
Thr	Leu	Gln	Pro	Thr	Glu	Ala	Ala	Tyr	Ile	Ala	Gly	Phe	Leu	Asp	Gly
	210					215					220				
Asp	Gly	Ser	Ile	Tyr	Ala	Lys	Leu	Ile	Pro	Arg	Pro	Asp	Tyr	Lys	Asp
225					230					235					240
Ile	Lys	Tyr	Gln	Val	Ser	Leu	Ala	Ile	Ser	Phe	Ile	Gln	Arg	Lys	Asp
				245					250					255	
Lys	Phe	Pro	Tyr	Leu	Gln	Asp	Ile	Tyr	Asp	Gln	Leu	Gly	Lys	Arg	Gly
			260					265					270		
Asn	Leu	Arg	Lys	Asp	Arg	Gly	Asp	Gly	Ile	Ala	Asp	Tyr	Thr	Ile	Ile
		275					280					285			
Gly	Ser	Thr	His	Leu	Ser	Ile	Ile	Leu	Pro	Asp	Leu	Val	Pro	Tyr	Leu
	290					295					300				
Arg	Ile	Lys	Lys	Lys	Gln	Ala	Asn	Arg	Ile	Leu	His	Ile	Ile	Asn	Leu
305					310					315					320
Tyr	Pro	Gln	Ala	Gln	Lys	Asn	Pro	Ser	Lys	Phe	Leu	Asp	Leu	Val	Lys
				325					330					335	
Ile	Val	Asp	Asp	Val	Gln	Asn	Leu	Asn	Lys	Arg	Ala	Asp	Glu	Leu	Lys
			340					345					350		

Ser Thr Asn Tyr Asp Arg Leu Leu Glu Glu Phe Leu Lys Ala Gly Lys
 355 360 365

Ile Glu Ser Ser Pro
 370

5 <210> 53
 <211> 22
 <212> ADN
 <213> *Monomastix sp.*

10 <400> 53
 ggaactgtct cagcagttc tg 22

15 <210> 54
 <211> 22
 <212> ADN
 <213> *Monomastix sp.*

20 <400> 54
 cagaacgtcg tgagacagtt cc 22

25 <210> 55
 <211> 9
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 55

Leu Ala Gly Leu Ile Asp Ala Asp Gly
 1 5

30 <210> 56
 <211> 22
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

40 <220>
 <221> modified_base
 <222> (10)..(13)
 <223> a, c, t, g, u, desconocido u otro

45 <400> 56
 tgcggtgctn nngacagcc tg 22

50 <210> 57
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

55 <220>
 <221> modified_base
 <222> (10)..(13)
 <223> a, c, t, g, u, desconocido u otro

<400> 57

caggctgtcn nnngacaccg ca

22

<210> 58

<211> 4

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

10

<400> 58

Ser Gly Gly Ser

1

15 <210> 59

<211> 22

<212> PRT

<213> Secuencia artificial

20 <220>

<223> Descripción de la secuencia artificial: Péptido sintético

<400> 59

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1 5 10 15

Ser Ser Gly Ser Ser Gly

20

25

<210> 60

<211> 25

<212> PRT

30 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

35 <400> 60

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1 5 10 15

Ser Ser Gly Ser Ser Gly Ser Ser Gly

20

25

40 <210> 61

<211> 28

<212> PRT

<213> Secuencia artificial

<220>

45 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 61

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1 5 10 15

Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly

20

25

<210> 62
 <211> 31
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

10 <400> 62
 Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10 15
 Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 20 25 30

<210> 63
 <211> 34
 <212> PRT
 <213> Secuencia artificial

15 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

20 <400> 63
 Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10 15
 Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser
 20 25 30
 Ser Gly

25 <210> 64
 <211> 29
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 64
 Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10 15
 Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Gly
 20 25

40 <210> 65
 <211> 30
 <212> PRT
 <213> Secuencia artificial

45 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 65

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10 15

Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Gly
 20 25 30

5 <210> 66
 <211> 8
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético
 <400> 66

Val Leu Asp Ser Pro Gly Ser Val
 1 5

15 <210> 67
 <211> 12
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético
 <400> 67

Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser
 1 5 10

25 <210> 68
 <211> 6
 <212> PRT
 30 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Péptido sintético
 35 <400> 68

Leu Ser Pro Ser Gln Ala
 1 5

40 <210> 69
 <211> 9
 <212> PRT
 <213> Secuencia artificial

<220>
 45 <223> Descripción de la secuencia artificial: Péptido sintético
 <400> 69

Ala Ser Ser Ser Pro Gly Ser Gly Ile
 1 5

50 <210> 70
 <211> 4
 <212> PRT
 <213> Secuencia artificial

55 <220>

<223> Descripción de la secuencia artificial: Péptido sintético

<400> 70

5 **Ala Ser Ser Ser**
1

<210> 71

<211> 5

<212> PRT

10 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

15 <400> 71

Pro Gly Ser Gly Ile
1 5

<210> 72

20 <211> 6

<212> PRT

<213> Secuencia artificial

<220>

25 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 72

Ile Ser Glu Ala Leu Arg
1 5

30 <210> 73

<211> 6

<212> PRT

<213> Secuencia artificial

35 <220>

<223> Descripción de la secuencia artificial: Péptido sintético

<400> 73

40 <210> 74

<211> 4

<212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

<400> 74

45 <210> 75

<211> 6

<212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

<400> 75

50 <210> 76

<211> 6

<212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Péptido sintético

Ala Ser Ser Ala
1

<400> 75

Ala Leu Arg Ala Gly Ala
1 5

5 <210> 76
<211> 6
<212> PRT
<213> Secuencia artificial

10 <220>
<223> Descripción de la secuencia artificial: Péptido sintético

<400> 76

Thr Lys Ser Lys Glu Phe
1 5

15 <210> 77
<211> 40
<212> PRT
20 <213> Secuencia artificial

<220>
<223> Descripción de la secuencia artificial: Polipéptido sintético

25 <400> 77

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Gly Ala Thr Lys Ser
35 40

30 <210> 78
<211> 41
<212> PRT
<213> Secuencia artificial

<220>
35 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 78

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Gly Gly Ala Thr Lys Ser
35 40

40 <210> 79
<211> 42
<212> PRT
45 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 79

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Ala Gly Gly Ala Thr Lys Ser
35 40

5

<210> 80

<211> 43

<212> PRT

10 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

15 <400> 80

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Ala Ser Gly Gly Ala Thr Lys Ser
35 40

20

<210> 81

<211> 44

<212> PRT

<213> Secuencia artificial

<220>

25 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 81

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Ala Ser Ser Gly Gly Ala Thr Lys Ser
35 40

30

<210> 82

<211> 45

<212> PRT

<213> Secuencia artificial

35

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 82

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Ala Ser Ser Ala Gly Gly Ala Thr Lys Ser
35 40 45

5 <210> 83

<211> 41

<212> PRT

<213> Secuencia artificial

10 <220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 83

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
20 25 30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
35 40

15

<210> 84

<211> 41

<212> PRT

20 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

25 <400> 84

Ser Leu Pro Gly Ser Val Gly Gly Ile Ser Pro Ser Gln Ala Ser Ser
1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala
20 25 30

Pro Arg Ala Gly Ala Thr Lys Glu Phe
35 40

30 <210> 85

<211> 41

<212> PRT

<213> Secuencia artificial

<220>

35 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 85

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala
 20 25 30

Thr Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

5 <210> 86
 <211> 41
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 86

Ser Leu Pro Gly Ser Leu Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Pro Ser Glu Ala
 20 25 30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

15 <210> 87
 <211> 41
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 87

25 Ser Leu Pro Gly Ser Val Gly Gly Val Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Val Ser Glu Ala
 20 25 30

Ser Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

30 <210> 88
 <211> 41
 <212> PRT
 <213> Secuencia artificial

35 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 88

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Leu Ser Glu Ala
 20 25 30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

<210> 89

<211> 41

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 89

Ser Leu Pro Gly Ser Leu Gly Gly Ile Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ser Ser Glu Ala
 20 25 30

Ser Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

15 <210> 90

<211> 40

<212> PRT

<213> Secuencia artificial

20 <220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 90

Ser Pro Gly Ser Val Gly Gly Val Ser Pro Ser Gln Ala Ser Ser Ala
 1 5 10 15

Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Thr
 20 25 30

Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

25

<210> 91

<211> 37

<212> PRT

30 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

35 <400> 91

Ser Leu Pro Gly Ser Leu Gly Gly Val Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala Pro Arg Ala Gly
 20 25 30

Ala Thr Lys Glu Phe
 35

<210> 92

<211> 37

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 92

Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Ile Arg Ala Gly
 20 25 30

Ala Thr Lys Glu Phe
 35

15 <210> 93

<211> 44

<212> PRT

<213> Secuencia artificial

20 <220>

<223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 93

Ser Leu Pro Gly Ser Leu Gly Gly Val Ser Pro Ser Gln Ala Ser Ser
 1 5 10 15

Ala Ala Ser Ser Ala Ser Ser Ala Ala Ser Ser Pro Gly Ser Gly Ala
 20 25 30

Ser Glu Ala Ser Arg Ala Gly Ala Thr Lys Glu Phe
 35 40

25

<210> 94

<211> 6

<212> PRT

30 <213> *Monomastix sp.*

<400> 94

Leu Gln Pro Thr Glu Ala
 1 5

35

<210> 95
 <211> 39
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

10 <400> 95

Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala
 1 5 10 15

Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg
 20 25 30

Ala Gly Ala Thr Lys Ser Ala
 35

<210> 96
 <211> 39
 <212> PRT
 <213> Secuencia artificial

15 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

20 <400> 96

Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala
 1 5 10 15

Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg
 20 25 30

Ala Gly Ala Thr Lys Ser Gly
 35

25 <210> 97
 <211> 43
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

35 <400> 97

Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
 1 5 10 15

Ser Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala
 20 25 30

Ala Ser Ser Leu Ala Ser Lys Pro Gly Ser Thr
 35 40

<210> 98
 <211> 39
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 98

10 Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
 1 5 10 15
 Ser Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala
 20 25 30
 Ala Ser Ser Pro Gly Ser Thr
 35

<210> 99
 <211> 43
 <212> PRT
 <213> Secuencia artificial

15 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

<400> 99

20 Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
 1 5 10 15
 Ser Ala Ala Ser Ser Pro Gly Ser Gly Pro Ser Glu Ala Leu Arg Ala
 20 25 30

Ala Ser Ser Phe Ala Ser Lys Pro Gly Ser Thr
 35 40

25 <210> 100
 <211> 4
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 100

Leu Pro Gly Glu
 1

40 <210> 101
 <211> 13
 <212> PRT
 <213> Secuencia artificial

45 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 101

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10

5 <210> 102
 <211> 16
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Descripción de la secuencia artificial: Péptido sintético

<400> 102

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
 1 5 10 15

15 <210> 103
 <211> 29
 <212> PRT
 20 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Péptido sintético

25 <400> 103

Asn Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser
 1 5 10 15

Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly
 20 25

30 <210> 104
 <211> 11
 <212> PRT
 <213> *Desulfurococcus mobilis*

<400> 104

35 Met Leu Glu Arg Ile Arg Leu Phe Asn Met Arg
 1 5 10

40 <210> 105
 <211> 22
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

45 <400> 105
 tgcggtgtca ttagacaccg ca 22

50 <210> 106
 <211> 22
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<400> 106
 tgcggtgtct aatgacaccg ca 22

5 <210> 107
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<400> 107
 caggctgtca ttagacagcc tg 22

15 <210> 108
 <211> 22
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Descripción de la secuencia artificial: Oligonucleótido sintético

<400> 108
 caggctgtca ttagacagcc tg 22

25 <210> 109
 <211> 46
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> Descripción de la secuencia artificial: Polipéptido sintético

35 <400> 109

Val	Leu	Asp	Ser	Leu	Pro	Gly	Ser	Val	Gly	Gly	Leu	Ser	Pro	Ser	Gln
1				5				10					15		

Ala	Ser	Ser	Ala	Ala	Ser	Ser	Ala	Ser	Ser	Ser	Pro	Gly	Ser	Gly	Ile
			20				25					30			

Ser	Glu	Ala	Leu	Arg	Ala	Gly	Ala	Thr	Lys	Ser	Lys	Glu	Phe
		35				40					45		

40 <210> 110
 <211> 6
 <212> PRT
 <213> Secuencia artificial

45 <220>
 <223> Descripción de la secuencia artificial: Marcador 6xHis sintético

<400> 110

His	His	His	His	His	His
1				5	

REIVINDICACIONES

1. Una meganucleasa monocatenaria recombinante que comprende:

- 5 (a) una primera subunidad LAGLIDADG que comprende una secuencia polipeptídica que tiene al menos un 85 % de identidad de secuencia con los restos 9-151 de una meganucleasa I-Crel de tipo silvestre de SEQ ID NO: 1 y que tiene un primer semisitio de reconocimiento; y
 10 (b) una segunda subunidad LAGLIDADG que comprende una secuencia polipeptídica que tiene al menos un 85 % de identidad de secuencia con los restos 9-151 de una meganucleasa I-Crel de tipo silvestre de SEQ ID NO: 1 y que tiene un segundo semisitio de reconocimiento;

en la que dichas primera y segunda subunidades LAGLIDADG están unidas covalentemente mediante un engarce polipeptídico;

en la que dicho engarce polipeptídico comprende 23-56 restos;

- 15 en la que dicha primera subunidad está unida covalentemente a dicho engarce polipeptídico en un resto correspondiente a una posición seleccionada del grupo que consiste en las posiciones 142-163 de SEQ ID NO: 1; en la que dicha segunda subunidad está unida covalentemente a dicho engarce polipeptídico en un resto correspondiente a una posición seleccionada del grupo que consiste en las posiciones 1-9 de SEQ ID NO: 1; y en la que dichas primera y segunda subunidades LAGLIDADG son capaces de actuar juntas para reconocer y escindir una secuencia de ADN no palindrómica que es un híbrido de dicho primer semisitio de reconocimiento y dicho segundo semisitio de reconocimiento.
- 20

2. La meganucleasa monocatenaria recombinante de la reivindicación 1, en la que dicho engarce polipeptídico tiene una estructura secundaria estable.

- 25 3. La meganucleasa monocatenaria recombinante de la reivindicación 2, en la que dicho engarce polipeptídico comprende al menos dos estructuras de hélice α .

- 30 4. La meganucleasa monocatenaria recombinante de la reivindicación 1, en la que dicho engarce polipeptídico comprende 23-40 restos.

5. La meganucleasa monocatenaria recombinante de la reivindicación 1 o la reivindicación 4, en la que dicho engarce polipeptídico comprende 23-40 restos y es un engarce flexible.

- 35 6. La meganucleasa monocatenaria recombinante de la reivindicación 1 o las reivindicaciones 4-5, en la que dicho engarce polipeptídico comprende 25-31 restos.

7. La meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-6, en la que:

- 40 (a) dicha primera subunidad está unida covalentemente a dicho engarce polipeptídico en un resto correspondiente a una posición seleccionada del grupo que consiste en las posiciones 152-163 de SEQ ID NO: 1; o
 (b) dicha primera subunidad está unida covalentemente a dicho engarce polipeptídico en un resto correspondiente a una posición seleccionada del grupo que consiste en las posiciones 151-153 de SEQ ID NO: 1,
 45 1,

opcionalmente en la que dicha segunda subunidad está unida covalentemente a dicho engarce polipeptídico en un resto correspondiente a una posición seleccionada del grupo que consiste en las posiciones 7-9 de SEQ ID NO: 1.

- 50 8. La meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-7, en la que al menos uno de dichos dominios LAGLIDADG comprende al menos una modificación de aminoácido seleccionada del grupo que consiste en: Y75, L75, C75, Y139, C46, A46, H75, R75, H46, K46, R46, K70, E70, E75, E46, D46, Q70, C70, L70, Q75, H139, Q46, G70, A70, S70, G46, T44, A44, V44, I44, L44, N44, D70, K44, R44, H70, D44, E44, C44, Q68, C24, E68, F68, K24, R24, M68, C68, L68, H68, Y68, K68, A26, Q77, E77, K26, R77, E26, S77, Q26, S26, E42, R42, K28, C28, Q42, M66, K66, Q40, E40, R28, R40, C40, I40, V40, C79, I79, V79, Q28, A40, A79, A28, H28, S40, S28, E38, K30, R30, K38, R38, E30, I38, L38, C38, H38, N38, Q30, F33, E33, D33, H33, L33, V33, I33, C33, R32, R33, E32, K32, L32, V32, A32, C32, D32, I32, N32, H32, Q32 y T32.
- 55

9. La meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-8, en la que cada una de dichas subunidades LAGLIDADG tiene un semisitio de reconocimiento seleccionado del grupo que consiste en SEQ ID NO: 7-30, opcionalmente en donde al menos una de dichas subunidades LAGLIDADG tiene un semisitio de reconocimiento seleccionado del grupo que consiste en SEQ ID NO: 7-30; y la otra de dichas subunidades LAGLIDADG tiene un semisitio de reconocimiento que difiere en una modificación de al menos un par de bases de un semisitio de reconocimiento seleccionado del grupo que consiste en SEQ ID NO: 7-30.
- 60

65

10. La meganucleasa monocatenaria recombinante de la reivindicación 1, en la que al menos un 50 % de los restos que forman el engarce son restos polares sin carga.

11. Un procedimiento *ex vivo* para producir una célula eucariota modificada genéticamente que incluye una secuencia exógena de interés insertada en un cromosoma de dicha célula eucariota, que comprende:

(a) 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 de interés;

en donde dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión; y en donde dicha meganucleasa es una meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-10, o

(b) introducir una proteína meganucleasa en una célula eucariota; y transfectar dicha célula eucariota con un ácido nucleico que incluye dicha secuencia de interés; en el que dicha meganucleasa produce un sitio de escisión en dicho cromosoma y dicha secuencia de interés se inserta en dicho cromosoma en dicho sitio de escisión; y en donde dicha meganucleasa es una meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-10,

en donde el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

12. Un procedimiento *ex vivo* para producir una célula eucariota modificada genéticamente alterando una secuencia diana en un cromosoma de dicha célula eucariota, que comprende: 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 diana se altera por unión de extremos no homólogos en dicho sitio de escisión; y en donde dicha meganucleasa es una meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-10

en donde el método no comprende modificar la identidad genética de la línea germinal de seres humanos.

13. Un procedimiento para producir un organismo modificado genéticamente no humano, que comprende: producir una célula eucariota modificada genéticamente de acuerdo con el procedimiento de una cualquiera de las reivindicaciones 11 o 12,

en el que la célula eucariota es un protoplasto, un gameto no humano, un cigoto no humano, un blastocisto no humano o una célula madre embrionaria no humana; y cultivar dicha célula eucariota modificada genéticamente para producir dicho organismo modificado genéticamente, opcionalmente en donde:

(a) el organismo modificado no humano es una planta y la célula eucariota modificada genéticamente es un protoplasto; o

(b) el organismo modificado genéticamente no humano es un animal y la célula eucariota modificada genéticamente es un cigoto no humano o una célula madre embrionaria no humana.

14. Una meganucleasa monocatenaria recombinante de una cualquiera de las reivindicaciones 1-10 para su uso en (a) terapia génica; o (b) el tratamiento de una infección por patógeno viral en un huésped eucariota o una infección por patógeno procariota en un huésped eucariota.

Figura 1

Componentes estructurales del engarce 9

