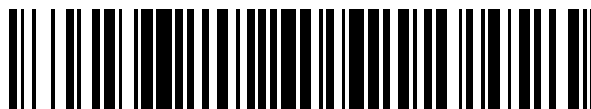


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 689 256**

51 Int. Cl.:

C12N 9/22 (2006.01)

C12N 15/74 (2006.01)

C07K 14/195 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.12.2012** **E 16169431 (0)**

97 Fecha y número de publicación de la concesión europea: **04.07.2018** **EP 3091072**

54 Título: **Ribonucleoproteínas Cascada modificadas y usos de las mismas**

30 Prioridad:

30.12.2011 GB 201122458

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

12.11.2018

73 Titular/es:

CARIBOU BIOSCIENCES, INC. (100.0%)
2929 7th Street, Suite 105
Berkeley, CA 94710, US

72 Inventor/es:

BROUNS, STAN JOHAN JOZEF y
VAN DER OOST, JOHN

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 689 256 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

Ribonucleoproteínas Cascada modificadas y usos de las mismas

La invención se refiere al campo de la ingeniería genética y más particularmente al área de modificación génica y/o genómica de organismos, que incluyen procariotas y eucariotas. La invención también se refiere a procedimientos de fabricación de herramientas específicas de sitio para su uso en procedimientos de análisis genómico y modificación genética, ya sea *in vivo* o *in vitro*. La invención se refiere más particularmente al campo de ribonucleoproteínas que reconocen y asocian secuencias de ácido nucleico de una manera específica de secuencia.

Las bacterias y arqueas tienen una amplia variedad de mecanismos de defensa contra el ADN invasivo. Los denominados sistemas de defensa CRISPR/Cas proporcionan inmunidad adaptativa al integrar los fragmentos de ADN plasmídico y viral en loci de repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) en el cromosoma del huésped. Las secuencias virales o derivadas del plásmido, conocidas como espaciadores, están separadas entre sí mediante la repetición de secuencias derivadas del huésped. Estos elementos repetitivos son la memoria genética de este sistema inmunitario y cada locus CRISPR contiene un repertorio diverso de secuencias únicas "espaciadoras" adquiridas durante encuentros previos con elementos genéticos extraños.

La adquisición de ADN extraño es la primera etapa de la inmunización, pero la protección requiere que el CRISPR se transcriba y que estos transcritos largos se procesen en ARN cortos derivados de CRISPR (ARNcr) que contienen una secuencia espaciadora única complementaria a un provocador de ácido nucleico extraño.

Además del ARNcr, los experimentos genéticos en varios organismos han revelado que se requiere un conjunto único de proteínas (Cas) asociadas a CRISPR para las etapas de adquisición de inmunidad, para la biogénesis de ARNcr y para la interferencia dirigida. Asimismo, se ha demostrado que un subconjunto de proteínas Cas de sistemas CRISPR filogenéticamente distintos se ensamblan en complejos grandes que incluyen un ARNcr.

Una reciente reevaluación de la diversidad de los sistemas CRISPR/Cas ha dado lugar a una clasificación de tres tipos distintos (Makarova K. y col (2011) *Nature Reviews Microbiology* - AOP 9 de mayo de 2011; doi:10.1038/nrmicro2577) que varían en el contenido del gen *cas* y muestran diferencias importantes a lo largo de la vía de defensa de CRISPR. (La clasificación de Makarova y la nomenclatura para genes asociados a CRISPR se adopta en la presente memoria descriptiva) Los transcritos de ARN de los loci de CRISPR (pre-ARNcr) se escinden específicamente en las secuencias repetidas mediante endorribonucleasas (Cas) asociadas a CRISPR en sistemas de tipo I y tipo III o mediante RNasa III en sistemas de tipo II; los ARNcr generados son utilizados por un complejo de proteínas Cas como ARN guía para detectar secuencias complementarias de ADN o ARN invasor. La escisión de los ácidos nucleicos objetivo se ha demostrado *in vitro* para el sistema B tipo III de *Pyrococcus furiosus*, que escinde ARN en un mecanismo anclado a la regla y, más recientemente, *in vivo* para el sistema tipo II de *Streptococcus thermophilus*, que escinde ADN en la secuencia objetivo complementaria (protoespaciadora). Por el contrario, para los sistemas de tipo I, el mecanismo de interferencia de CRISPR sigue siendo en gran parte desconocido.

El organismo modelo de la cepa K12 de *Escherichia coli* posee un CRISPR/Cas tipo I-E (anteriormente conocido como CRISPR subtipo E (Cse)). Contiene ocho genes *cas* (*cas1*, *cas2*, *cas3* y *cse1*, *cse2*, *cas7*, *cas5*, *cas6e*) y un CRISPR corriente abajo (repeticiones tipo 2). En K12 de *Escherichia coli*, los ocho genes *cas* están codificados corriente arriba del locus de CRISPR. *Cas1* y *Cas2* no parecen ser necesarios para la interferencia del objetivo, pero es probable que participen en la adquisición de una nueva secuencia objetivo. Por el contrario, seis proteínas Cas: *Cse1*, *Cse2*, *Cas3*, *Cas7*, *Cas5* y *Cas6e* (anteriormente también conocidas como *CasA*, *CasB*, *Cas3*, *CasC/Cse4*, *CasD* y *CasE/Cse3*, respectivamente) son esenciales para la protección contra la estimulación del fago lambda. Cinco de estas proteínas: *Cse1*, *Cse2*, *Cas7*, *Cas5* y *Cas6e* (anteriormente conocidas como *CasA*, *CasB*, *CasC/Cse4*, *CasD* y *CasE/Cse3*, respectivamente) se ensamblan con un ARNcr para formar una ribonucleoproteína (RNP) de múltiples subunidades denominada Cascada.

En *E. coli*, Cascada es un complejo ribonucleoproteico de 405 kDa compuesto por una estequiometría desigual de cinco proteínas Cas funcionalmente esenciales: *Cse1*₁*Cse2*₂*Cas7*₆*Cas5*₁*Cas6e*₁ (es decir, bajo la nomenclatura previa *CasA*₁*B*₂*C*₆*D*₁*E*₁) y un ARN derivado de CRISPR de 61 nt. Cascada es una RNP forzada que depende del ARNcr para ensamblaje y estabilidad del complejo, y para la identificación de secuencias de ácidos nucleicos invasoras. Cascada es un complejo de vigilancia que encuentra y une ácidos nucleicos extraños que son complementarios a la secuencia espaciadora del ARNcr.

Jore y col. (2011) titulado "*Structural basis for CRISPR RNA-guided DNA recognition by Cascade*" *Nature Structural & Molecular Biology* 18: 529-537 describe cómo se produce una escisión del transcrito de pre-ARNcr por la subunidad *Cas6e* de Cascada, lo que resulta en un ARNcr maduro de 61 nt que es retenido por el complejo CRISPR. El ARNcr sirve como un ARN guía para la unión específica de secuencia de Cascada a moléculas de ADN bicatenario (bc) a través del emparejamiento de bases entre el espaciador de ARNcr y el protoespaciador complementario, formando así el llamado bucle R. Se sabe que esto es un proceso independiente de ATP.

Brouns S.J.J., y col (2008) titulado "*Small CRISPR RNAs guide antiviral defense in prokaryotes*" *Science* 321: 960-964 enseña que Cascada cargada con un ARNcr requiere *Cas3* para la resistencia a fagos *in vivo*.

- 5 Marraffini L. y Sontheimer E. (2010) titulado "*CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea*" *Nature Reviews Genetics* 11: 181-190 es un artículo de revisión que resume el estado del conocimiento en la técnica en el campo. Se han planteado algunas sugerencias sobre las aplicaciones y tecnologías basadas en CRISPR, pero se trata principalmente de generar cepas resistentes a fagos de bacterias domesticadas para la industria láctea. La escisión específica de las moléculas de ARN *in vitro* por un complejo RNPcr en *Pyrococcus furiosus* se sugirió como algo que espera un mayor desarrollo. La manipulación de los sistemas CRISPR también se sugiere como una posible forma de reducir la transmisión de cepas bacterianas resistentes a los antibióticos en los hospitales. Los autores enfatizan que se necesitarán más esfuerzos de investigación para explorar la utilidad potencial de la tecnología en estas áreas.
- 10 El documento US2011236530 A1 (Manoury y col.) titulado "*Genetic cluster of strains of Streptococcus thermophilus having unique rheological properties for dairy fermentation*" desvela ciertas cepas de *S. thermophilus* que fermentan la leche de modo que sea altamente viscosa y débilmente glutinosa. Se desvela un locus de CRISPR específico de secuencia definida.
- 15 El documento US2011217739 A1 (Terns y col.) titulado "*Cas6 polypeptides and methods of use*" desvela polipéptidos que tienen actividad endorribonucleasa Cas6. Los polipéptidos escinden un polinucleótido de ARN objetivo que tiene un dominio de reconocimiento de Cas6 y un sitio de escisión. La escisión se puede llevar a cabo *in vitro* o *in vivo*. Microbios tales como *E. coli* o *Haloferax volcanii* se modifican genéticamente para expresar la actividad de la endorribonucleasa Cas6.
- 20 El documento WO2010054154 (Danisco) titulado "*Bifidobacteria CRISPR sequences*" desvela diversas secuencias CRISPR halladas en bifidobacterias y su uso en la fabricación de cepas genéticamente alteradas de las bacterias que están alteradas en sus características de resistencia a fagos.
- 25 El documento US2011189776 A1 (Terns y col.) titulado "*Prokaryotic RNAi-like system and methods of use*" describe procedimientos de inactivación de polinucleótidos objetivo *in vitro* o en microbios procariotas *in vivo*. Los procedimientos utilizan un ARNsip que tiene una región 5' de 5-10 nucleótidos seleccionados entre una repetición de un locus de CRISPR inmediatamente corriente arriba de un espaciador. La región 3' es esencialmente complementaria a una porción del polinucleótido objetivo. También se describen polipéptidos que tienen actividad endonucleasa en presencia de ARNsip y polinucleótido objetivo.
- 30 El documento EP2341149 A1 (Danisco) titulado "*Use of CRISPR associated genes (CAS)*" describe cómo pueden utilizarse uno o más genes Cas para modular la resistencia de las células bacterianas contra bacteriófagos, particularmente bacterias que proporcionan un cultivo iniciador o cultivo probiótico en productos lácteos.
- El documento WO2010075424 (*The Regents of the University of California*) titulado "*Compositions and methods for downregulating prokaryotic genes*" desvela un polinucleótido aislado que comprende una matriz CRISPR. Al menos un espaciador del CRISPR es complementario a un gen de una procariota, por lo que puede regular negativamente la expresión del gen; particularmente cuando el gen está asociado con la producción de biocombustibles.
- 35 El documento WO2008108989 (Danisco) titulado "*Cultures with improved phage resistance*" desvela la selección de cepas bacterianas resistentes a bacteriófagos y también la selección de las cepas que tienen un espaciador adicional que tiene una identidad del 100 % con una región del ARN del fago. Se describen combinaciones mejoradas de cepas y rotaciones de cultivos iniciadores para su uso en la industria láctea. Ciertos fagos se describen para su uso como agentes de biocontrol.
- 40 El documento WO2009115861 (Institut Pasteur) titulado "*Molecular typing and subtyping of Salmonella by identification of the variable nucleotide sequences of the CRISPR loci*" desvela procedimientos de detección e identificación de bacterias del género *Salmonella* utilizando sus secuencias de nucleótidos variables contenidas en loci de CRISPR.
- 45 El documento WO2006073445 (Danisco) titulado "*Detection and typing of bacterial strains*" describe la detección y tipificación de cepas bacterianas en productos alimenticios, suplementos dietéticos y muestras ambientales. Las cepas de *Lactobacillus* se identifican a través de secuencias específicas de nucleótidos CRISPR.
- 50 Urnov F y col. (2010) titulado "*Genome editing with engineered zinc finger nucleases*" *Nature* 11: 636-646 es un artículo de revisión sobre nucleasas con dedos de cinc y cómo han sido fundamentales en el campo de la genética inversa en una variedad de organismos modelo. Las nucleasas con dedos de cinc se han desarrollado de modo que es posible orientar con precisión la escisión del genoma seguido de modificación génica en el procedimiento de reparación posterior. Sin embargo, las nucleasas con dedos de cinc se generan fusionando una serie de dominios de unión al ADN con dedos de cinc a un dominio de escisión del ADN. La especificidad de la secuencia de ADN se logra mediante el acoplamiento de varios dedos de cinc en serie, cada uno de los cuales reconoce un motivo de tres nucleótidos. Un inconveniente significativo con la tecnología es que se deben desarrollar nuevos dedos de cinc para cada nuevo locus de ADN que requiere ser escindido. Esto requiere una ingeniería de proteínas y una identificación sistemática exhaustiva para garantizar la especificidad de la unión al ADN.
- 55 El documento WO2011097036 A1 describe fusiones TALEN que comprenden variantes FokI. También se desvelan

endonucleasas dirigidas a ZNF.

Wiedenheft y col. (2011) titulado "*Structures of the RNA-guided surveillance complex from a bacterial immune system*" *Nature* 477(7369):486-489 describe un complejo de vigilancia de múltiples subunidades bacterianas llamado Cascada (complejo asociado a CRISPR para defensa antiviral) requerido para la protección contra bacteriófagos.

- 5 En los campos de la ingeniería genética e investigación genómica existe una necesidad continua de agentes mejorados para la detección y/o escisión de ácidos nucleicos específica de secuencia/sitio.

Los inventores han realizado un descubrimiento sorprendente en el sentido de que ciertas bacterias que expresan Cas3, que tienen actividad helicasa-nucleasa, expresan Cas3 como una fusión con Cse1. Los inventores también han podido inesperadamente producir fusiones artificiales de Cse1 con otras enzimas nucleasa.

- 10 Los inventores también han descubierto que el reconocimiento de ADN objetivo independiente de Cas3 por Cascada marca el ADN para la escisión por Cas3, y que la unión al ADN de Cascada se rige por los requisitos topológicos del ADN objetivo.

Los inventores han descubierto además que Cascada es incapaz de unirse a plásmidos objetivo relajados, pero sorprendentemente Cascada muestra una gran afinidad por los objetivos que tienen una topología negativamente superenrollada (nSE).

- 15

La invención se define en las reivindicaciones adjuntas. Por consiguiente, la invención proporciona una proteína de fusión artificial de una subunidad proteica Cse1 asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISP) de tipo I y una endonucleasa FokI; una molécula de ácido nucleico que codifica dicha proteína de fusión; un vector de expresión que comprende una molécula de ácido nucleico; un complejo proteico Cascada que comprende dicha proteína de fusión; un complejo ribonucleoproteico que comprende dicha proteína de fusión y una molécula de ARN CRISP (ARNcr); así como una célula eucariota que comprende dicho complejo ribonucleoproteico, en el que la célula es diferente a una célula madre embrionaria humana o célula germinal humana y en el que la célula no forma parte del cuerpo humano o animal. En otro aspecto, la invención proporciona un procedimiento de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico objetivo *in vitro* que comprende la puesta en contacto del ácido nucleico objetivo con dicho complejo ribonucleoproteico.

- 20
- 25

Se desvela un complejo asociado a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) para la defensa antiviral (Cascada), el complejo proteico Cascada, o una porción del mismo, que comprende al menos las subunidades proteicas asociadas a CRISPR:

- 30 - Cas7 (o COG 1857) que tiene una secuencia de aminoácidos de SEQ ID NO: 3 o una secuencia de al menos un 18 % de identidad con la misma,
- Cas5 (o COG1688) que tiene una secuencia de aminoácidos de SEQ ID NO: 4 o una secuencia de al menos un 17 % de identidad con la misma, y
- 35 - Cas6 (o COG 1583) que tiene una secuencia de aminoácidos de la SEQ ID NO: 5 o una secuencia de al menos un 16 % de identidad con la misma,

y en la que al menos una de las subunidades incluye una secuencia de aminoácidos adicional que proporciona la actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina.

- 40 Una subunidad que incluye una secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina es un ejemplo de lo que se puede denominar "una subunidad ligada a al menos una fracción funcional"; una fracción funcional es el polipéptido o proteína constituido por la secuencia de aminoácidos adicional. La actividad de activación de la transcripción puede ser la que conduce a la activación o regulación positiva de un gen deseado; la actividad de represión de la transcripción conduce a la represión o regulación negativa de un gen deseado. La selección del gen se debe a la orientación del complejo Cascada de la invención con una molécula de ARN, como se describe más adelante.
- 45

La secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina está formada preferentemente por restos de aminoácidos contiguos. Estos aminoácidos adicionales se pueden ver como un polipéptido o proteína que es contiguo y forma parte de la(s) subunidad(es) Cas o Cse afectadas. Dicha secuencia de polipéptido o proteína preferentemente no forma normalmente parte de ninguna secuencia de aminoácidos de la subunidad Cas o Cse. En otras palabras, la secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina puede ser distinta de una secuencia de aminoácidos de la subunidad Cas o Cse, o porción de la misma, es decir, puede ser distinta de una secuencia de aminoácidos de la subunidad Cas3 o porción de la misma.

50

55

La secuencia de aminoácidos adicional con actividad de modificación, visualización, activación de la transcripción o

represión de la transcripción de un ácido nucleico o cromatina puede, según se desee, obtenerse o derivarse del mismo organismo, p. ej., *E. coli*, como la(s) subunidad(es) Cas o Cse.

Adicionalmente y/o alternativamente a lo anterior, la secuencia de aminoácidos adicional que actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina puede ser "heteróloga" a la secuencia de aminoácidos de la(s) subunidad(es) Cas o Cse. Por lo tanto, la secuencia de aminoácidos adicional puede obtenerse o derivarse de un organismo diferente del organismo del cual se deriva(n) o se origina(n) la(s) subunidad(es) Cas y/o Cse.

De principio a fin, la identidad de secuencia se puede determinar por medio de BLAST y posterior alineamiento múltiple de secuencias de Cobalt en el servidor web del Centro Nacional para la Información Biotecnológica, en el que la secuencia en cuestión se compara con una secuencia de referencia (p. ej., SEQ ID NO: 3, 4 o 5). Las secuencias de aminoácidos se pueden definir en términos de porcentaje de similitud de secuencia en base a una matriz BLOSUM62 o identidad porcentual con una secuencia de referencia dada (p. ej., SEQ ID NO: 3, 4 o 5). La similitud o identidad de una secuencia implica una etapa inicial de realizar el mejor alineamiento antes de calcular el porcentaje de conservación con la referencia y refleja una medida de la relación evolutiva de las secuencias.

Cas7 puede tener una similitud de secuencia de al menos 31 % con SEQ ID NO: 3; Cas5 puede tener una similitud de secuencia de al menos 26 % con SEQ ID NO: 4. Cas6 puede tener una similitud de secuencia de al menos 27 % con SEQ ID NO: 5.

Para Cse1/CasA (502 AA):

>gi|16130667|ref|NP_417240.1| ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]

MNLLIDNWIPVRPRNGGKVQIINLQSLYCSRQDQWRLSLPRDDMELAAALALLVCIGQII
APAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHPFMQTKGVKANDVTTPM
EKLLAGVSGATNCAFNQPGQGEALCGGCTAIALFNQANQAPGFGGGFKSGLRGGT
PVTTFVRGIDLRSTVLLNVLTLPRLQKQFPNESHTENQPTWIKPIKSNEIPASSIGFVR
GLFWQPAHIELCDPIGIGKSCCGQESNLRYTGFLKEKFTFTVNGLWPHPHSPCLVTV
KKGEVEEKFLAFTTSAPSWTQISRVVVDKIIQNENGNRVAADVNNQFRNIAPQSPLELI
MGGYRNNQASILERRHDVLMFNQGWQQYGNVINEIVTVGLGYKTALRKALYTFAE
GFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNFSQADEVIADLRDKLHQL
CEMLFNQSVAPYAHHPKLISTLALARATLYKHLRELKPQGGPSNG [SEQ ID NO: 1]

Para Cse2/CasB (160 AA):

>gi|16130666|ref|NP_417239.1| ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]

MADEIDAMALYRAWQQLDNGSCAQIRRVSEPDELRDIPAFYRLVQPFQWENPRHQQ
ALLRMVFCLSAGKNVIRHQDKKSEQTTGISLGRALANSGRINERRIFQLIRADRTADM
VQLRRLTHAEPVLDWPLMARMLTWWGKRERQQLLEDFVLTNNKNA [SEQ ID

NO: 2]

Para Cas7/CasC/Cse4 (363 AA):

>gi|16130665|ref|NP_417238.1| ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]

MSNFINIHVLISHSPSCLNRDDMNMQKDAIFGGKRRVRISQSLKRAMRKSGYYAQN
IGESSLRTHLAQLRDVLRQKLGERFDQKIIDKTLALLSGKSVDEAEKISADAVTPWV
VGEIAWFCEQVAKAEADNLDDKLLKVLKEDIAAIRVNLQQGVDIALSGRMATSGM
MTELKGVDGAMSIAHAITTHQVDSIDWFTA VDDLQEQSAHLGTQEFSSGVFYRY
ANINLAQLQENLGGASREQALEIATHVHMLATEVPGAKQRTYAAFNPADMVMVN
FSDMPLSMANAFEKAVKAKDGFLQPSIQAFNQYWDRVANGYGLNGAAAQFSLSDV
DPITAVKQKQMPTEQLKSWVRNNGEA [SEQ ID NO: 3]

Para Cas5/CasD (224 AA):

>gi|90111483|ref|NP_417237.2| ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]

MRSYLILRLAGPMQAWGQPTFEGTRPTGRFPTRSGLLGLLGACLGIRDDTSSLQAL
 SESVQFAVRCDELILDDRRVSVTGLRDYHTVLGAREDYRGLKSHETIQTWREYLCD
 ASFTVALWLTPHATMWISELEKAVLKPRYTPYLGRRCPLTHPLFLGTCQASDPQKA
 LLNYEPVGGDIYSEESVTGHHLKFTARDEPMITLPRQFASREWYVIKGGMDVSQ
 [SEQ ID NO: 4]

Para Cas6e/CasE (199 AA):

- 5 >gi|16130663|ref|NP_417236.1| Enzima de escisión del precursor de ARN CRISPR; ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]

MYLSKVIIARAWSRDLYQLHQGLWHLFPNRPDAARDFLFHVEKRNTPEGCHVLLQS
 AQMPVSTAVATVIKTKQVEFQLQVGVPLYFRLRANPIKTILDNQKRLDSKGNIKRCR
 VPLIKEAEQIAWLQRKLGNAARVEDVHPISERPQYFSGDGKSGKIQTVCFEGVLTIND
 APALIDLTVQQGIGPAKSMGCGLLSLAPL [SEQ ID NO: 5]

- Al definir el intervalo de variantes de secuencia, para que no quede ninguna duda, cada uno de los siguientes límites opcionales sobre el grado de variación, se aplica a cada SEQ ID NO: 1, 2, 3, 4 o 5 comenzando desde el intervalo de variantes más amplio según se especifica en términos del porcentaje de identidad respectivo anterior. Por lo tanto, el intervalo de variantes puede incluir: al menos 16 %, o al menos 17 %, o al menos 18 %, o al menos 19 %, o al menos 20 %, o al menos 21 %, o al menos 22 %, o al menos 23 %, o al menos 24 %, o al menos 25 %, o al menos 26 %, o al menos 27 %, o al menos 28 %, o al menos 29 %, o al menos 30 %, o al menos 31 %, o al menos 32 %, o al menos 33 %, o al menos 34 %, o al menos 35 %, o al menos 36 %, o al menos 37 %, o al menos 38 %, o al menos 39 %, o al menos 40 %, o al menos 41 %, o al menos 42 %, o al menos 43 %, o al menos 44 %, o al menos 45 %, o al menos 46 %, o al menos 47 %, o al menos 48 %, o al menos 49 %, o al menos 50 %, o al menos 51 %, o al menos 52 %, o al menos 53 %, o al menos 54 %, o al menos 55 %, o al menos 56 %, o al menos 57 %, o al menos 58 %, o al menos 59 %, o al menos 60 %, o al menos 61 %, o al menos 62 %, o al menos 63 %, o al menos 64 %, o al menos el 65 %, o al menos 66 %, o al menos 67 %, o al menos 68 %, o al menos 69 %, o al menos 70 %, o al menos 71 %, o al menos 72 %, o al menos 73 %, o al menos 74 %, o al menos 75 %, o al menos 76 %, o al menos 77 %, o al menos 78 %, o al menos 79 %, o al menos 80 %, o al menos 81 %, o al menos 82 %, o al menos 83 %, o al menos 84 %, o al menos 85 %, o al menos 86 %, o al menos 87 %, o al menos 88 %, o al menos 89 %, o al menos 90 %, o al menos 91 %, o al menos 92 %, o al menos 93 %, o al menos 94 %, o al menos 95 %, o al menos 96 %, o al menos 97 %, o al menos 98 %, o al menos 99 %, o 100 % de identidad de secuencia de aminoácidos.

- 25 De principio a fin, la nomenclatura de Makarova y col. (2011) se está utilizando en la definición de las subunidades proteicas Cas. La Tabla 2 en la página 5 del artículo de Makarova y col. enumera los genes Cas y los nombres de las familias y superfamilias a las que pertenecen. De principio a fin, la referencia a una proteína Cas o subunidad proteica Cse incluye una referencia cruzada a la familia o superfamilia de la cual estas subunidades forman parte.

- 30 De principio a fin, las secuencias de referencia de las subunidades Cas y Cse de la invención se pueden definir como una secuencia de nucleótidos que codifica la secuencia de aminoácidos. Por ejemplo, la secuencia de aminoácidos de la SEQ ID NO: 3 para Cas7 también incluye todas las secuencias de ácidos nucleicos que codifican esa secuencia de aminoácidos. Las variantes de Cas7 incluidas dentro del alcance de la invención incluyen por lo tanto secuencias de nucleótidos de al menos las identidades o similitudes porcentuales de aminoácidos definidas con la secuencia de ácido nucleico de referencia; así como todas las posibles identidades o similitudes porcentuales entre ese límite inferior y 100 %.

Los complejos Cascada de la invención pueden estar constituidos por subunidades derivadas o modificadas a partir de más de una procariota bacteriana o arquea diferente. Además, las subunidades de diferentes subtipos de Cas pueden mezclarse.

- 40 En un aspecto preferente, la subunidad Cas6 es una subunidad Cas6e de SEQ ID NO: 17 a continuación, o una secuencia de al menos un 16 % de identidad con la misma.

La secuencia de una subunidad Cas6e preferente es >gi|16130663|ref|NP_417236.1| enzima de escisión del precursor de ARN CRISPR; ARN CRISP (ARNcr) que contiene una proteína del complejo antiviral Cascada [cepa K-12 de *Escherichia coli* sustr. MG1655]:

MYLSKVIIARAWSRDLYQLHQGLWHLFPNRPDAARDFLFHVEKRNTPEGCHVLLQS
 AQMPVSTAVATVIKTKQVEFQLQVGVPLYFRLRANPIKTILDNQKRLDSKGNIKRCR
 VPLIKEAEQIAWLQRKLGNAARVEDVHPISERPQYFSGDGKSGKIQTVCFEGVLTIND
 APALIDLTVQQGIGPAKSMGCGLLSLAPL [SEQ ID NO: 17]

- Los complejos de Cascada, o porciones de la misma, que comprenden al menos una subunidad que incluye una secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina, pueden comprender además una subunidad Cse2 (o similar a YgcK) que tiene una secuencia de aminoácidos de SEQ ID NO: 2 o una secuencia de al menos 20 % de identidad con la misma, o una porción de la misma. Alternativamente, la subunidad Cse se define como aquella que tiene al menos un 38 % de similitud con SEQ ID NO: 2. Opcionalmente, dentro del complejo proteico de la invención, es la subunidad Cse2 la que incluye la secuencia de aminoácidos adicional que tiene actividad de modificación de un ácido nucleico o cromatina.
- Los complejos Cascada de la invención pueden comprender una subunidad Cse1 (o similar a YgcL) que tiene una secuencia de aminoácidos de SEQ ID NO: 1 o una secuencia de al menos 9 % de identidad con la misma, o una porción de la misma. Opcionalmente dentro del complejo proteico de la invención, es la subunidad Cse1 la que incluye la secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina.
- Un complejo de Cascada es un complejo proteico del sistema CRISPR-Cas tipo I; más preferentemente un complejo proteico de CRISPR-Cas del subtipo I-E o puede estar basado en un complejo tipo I-A o tipo I-B. Un complejo tipo I-C, D o F es posible. En realizaciones particularmente preferentes basadas en el sistema de *E. coli*, las subunidades pueden tener las siguientes estequiometrias: Cse1₁Cse2₂Cas7₆Cas5₁Cas6₁ o Cse1₁Cse2₂Cas7₆Cas5₁Cas6e₁.
- La secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina puede fusionarse de forma traduccional mediante la expresión a través de sistemas de expresión de proteínas naturales o artificiales, o unirse covalentemente mediante una etapa de síntesis química a la al menos una subunidad; preferentemente, a la al menos una fracción funcional está fusionada o unida a al menos la región del extremo N terminal y/o la región del extremo C terminal de al menos una de las subunidades Cse1, Cse2, Cas7, Cas5, Cas6 o Cas6e. En una realización, la secuencia de aminoácidos adicional que tiene actividad de modificación de un ácido nucleico o cromatina está fusionada o unida al extremo N terminal o al extremo C terminal de una subunidad Cse1, preferentemente, el enlace está en la región del extremo N terminal de una subunidad Cse1.
- La secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina es una nucleasa FokI, o un mutante o una porción activa de la misma.
- La proteína en cuestión puede ser una proteína heteróloga de una especie distinta de la especie bacteriana de la cual las subunidades de la proteína Cascada tienen su origen de secuencia.
- Preferentemente, un complejo proteico de la invención se puede fusionar con el dominio N terminal de FokI y otro complejo proteico de la invención se puede fusionar con el dominio C terminal de FokI. Estos dos complejos proteicos pueden utilizarse en conjunto para lograr un corte bicatenario específico de locus ventajoso en un ácido nucleico, por lo que la ubicación del corte en el material genético reside en el diseño y elección del usuario, guiado por el componente de ARN (definido y descrito a continuación) y debido a la presencia de una secuencia denominada "motivo adyacente al protoespaciador" (MAP) en la cadena de ácido nucleico objetivo (también descrita con más detalle a continuación).
- En una realización preferente, un complejo proteico de la invención tiene una secuencia de aminoácidos adicional que es una endonucleasa de restricción modificada, p. ej., FokI. La modificación está preferentemente en el dominio catalítico. En realizaciones preferentes, la FokI modificada es KKR Sharkey o ELD Sharkey que está fusionada a la proteína Cse1 del complejo proteico. En una aplicación preferente de estos complejos de la invención, dos de estos complejos (KKR Sharkey y ELD Sharkey) pueden estar juntos en combinación. Un par de heterodímeros de complejos proteicos que emplean FokI modificada de forma diferente tiene una ventaja particular en el corte bicatenario dirigido de ácido nucleico. Si se utilizan homodímeros, entonces es posible que exista más escisión en sitios no objetivo debido a la actividad no específica. Un enfoque heterodimérico aumenta ventajosamente la fidelidad de la escisión en una muestra de material.
- El complejo Cascada con secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina definido y descrito anteriormente es una parte componente de un sistema general de la invención que permite ventajosamente al usuario seleccionar en una materia predeterminada un locus genético preciso que se desea escindir, marcar o alterar de alguna manera, p. ej., metilación, utilizando cualquiera de las entidades de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina definidas en la presente memoria. La otra parte componente del sistema es una molécula de ARN que actúa como una guía para dirigir el complejo Cascada de la invención al locus correcto en el ADN o ARN que tiene por objeto modificarse, cortarse o marcarse.
- El complejo Cascada de la invención también comprende preferentemente una molécula de ARN que comprende una secuencia de ribonucleótidos de al menos 50 % de identidad con una secuencia de ácido nucleico objetivo

deseada, y en la que el complejo proteico y la molécula de ARN forman un complejo ribonucleoproteico. Preferentemente, el complejo ribonucleoproteico se forma cuando la molécula de ARN se hibrida con su secuencia deseada de ácido nucleico objetivo. El complejo ribonucleoproteico se forma cuando los componentes necesarios de la combinación de la fracción funcional Cascada y la molécula de ARN y el ácido nucleico (ADN o ARN) están presentes en conjunto en condiciones fisiológicas adecuadas, ya sea *in vivo* o *in vitro*. Sin desear quedar ligado a teoría alguna particular, los inventores creen que en el contexto de ADNbc, particularmente ADN negativamente superenrollado, el complejo Cascada que se asocia con el ADNbc provoca un desenrollamiento parcial de las cadenas dúplex que luego permite que el ARN se asocie con una cadena; todo el complejo ribonucleoproteico luego migra a lo largo de la cadena de ADN hasta que se alcanza una secuencia objetivo esencialmente complementaria a al menos una porción de la secuencia de ARN, momento en el que se produce una interacción estable entre ARN y la cadena de ADN y surte efecto la función de la fracción funcional, ya sea modificando, cortando nucleasa o marcando el ADN en ese locus.

En realizaciones preferentes, una porción de la molécula de ARN tiene al menos un 50 % de identidad con la secuencia de ácido nucleico objetivo; más preferentemente al menos un 95 % de identidad con la secuencia objetivo. En realizaciones más preferentes, la porción de la molécula de ARN es esencialmente complementaria a lo largo de su longitud con respecto a la secuencia de ADN objetivo; es decir, solo hay uno, dos, tres, cuatro o cinco desapareamientos que pueden ser contiguos o no contiguos. La molécula de ARN (o porción de la misma) puede tener al menos 51 %, o al menos 52 %, o al menos 53 %, o al menos 54 %, o al menos 55 %, o al menos 56 %, o al menos 57 %, o al menos 58 %, o al menos 59 %, o al menos 60 %, o al menos 61 %, o al menos 62 %, o al menos 63 %, o al menos 64 %, o al menos 65 %, o al menos 66 %, o al menos 67 %, o al menos 68 %, o al menos 69 %, o al menos 70 %, o al menos 71 %, o al menos 72 %, o al menos 73 %, o al menos 74 %, o al menos 75 %, o al menos 76 %, o al menos 77 %, o al menos 78 %, o al menos 79 %, o al menos 80 %, o al menos 81 %, o al menos 82 %, o al menos 83 %, o al menos 84 %, o al menos 85 %, o al menos 86 %, o al menos 87 %, o al menos 88 %, o al menos 89 %, o al menos 90 %, o al menos 91 %, o al menos 92 %, o al menos 93 %, o al menos 94 %, o al menos 95 %, o al menos 96 %, o al menos 97 %, o al menos 98 %, o al menos 99 %, o 100 % de identidad con la secuencia objetivo.

El ácido nucleico objetivo puede ser ADN (mc o bc) o ARN.

En otras realizaciones preferentes, la molécula de ARN o porción de la misma tiene al menos un 70 % de identidad con el ácido nucleico objetivo. En tales niveles de identidad, el ácido nucleico objetivo es preferentemente ADNbc.

La molécula de ARN requerirá preferentemente una alta especificidad y afinidad por la secuencia de ácido nucleico objetivo. Es deseable una constante de disociación (K_d) en el intervalo de 1 pM a μ M, preferentemente 1-100 nM como se determina mediante electroforesis en gel preferentemente nativa, o alternativamente calorimetría de titulación isotérmica, resonancia de plasmón superficial o procedimientos de titulación basados en fluorescencia. La afinidad puede determinarse utilizando un ensayo de desplazamiento de movilidad electroforética (EDME), también llamado ensayo de retardo en gel (véase Semenova E y col. (2011) *Proc. Natl. Acad. Sci. USA* 108: 10098-10103).

La molécula de ARN se modela preferentemente en lo que se conoce en la naturaleza en procariotas como moléculas de ARN CRISPR (ARNcr). La estructura de las moléculas de ARNcr ya está establecida y explicada con más detalle en Jore y col. (2011) *Nature Structural & Molecular Biology* 18: 529-537. En resumen, un ARNcr maduro de tipo I-E a menudo tiene 61 nucleótidos de longitud y consiste en una región de 5' "mango" de 8 nucleótidos, la secuencia "espaciadora" de 32 nucleótidos y una secuencia 3' de 21 nucleótidos que forma una horquilla con un bucle tetranucleotídico. Sin embargo, el ARN utilizado en la invención no tiene que diseñarse estrictamente respecto al diseño de ARNcr de origen natural, ya sea en longitud, regiones o secuencias de ARN específicas. Lo que está claro, sin embargo, es que las moléculas de ARN para su uso en la invención pueden diseñarse basándose en la información de la secuencia génica en las bases de datos públicas o descubrirse recientemente, y luego fabricarse artificialmente, p. ej., por síntesis química en su conjunto o en parte. Las moléculas de ARN de la invención también se pueden diseñar y producir por medio de expresión en células genéticamente modificadas o sistemas de expresión libre de células y esta opción puede incluir la síntesis de parte o la totalidad de la secuencia de ARN.

La estructura y los requisitos de ARNcr también se han descrito en Semenova E y col. (2011) *Proc. Natl. Acad. Sci. EE.UU.* 108: 10098-10103. Hay una porción denominada "SEMILLA" que forma el extremo 5' de la secuencia espaciadora y que está flanqueada en 5' a ella por 5' mango de 8 nucleótidos. Semenova y col. (2011) han hallado que todos los residuos de la secuencia semilla deben ser complementarios a la secuencia objetivo, aunque para el residuo en la posición 6, se puede tolerar un desapareamiento. De forma similar, cuando se diseña y fabrica un componente de ARN de un complejo ribonucleoproteico de la invención dirigido a un locus objetivo (es decir, secuencia), se pueden aplicar las reglas necesarias de coincidencia y falta de coincidencia para la secuencia SEMILLA.

La invención incluye por lo tanto un procedimiento de detección y/o localización de un único cambio de base en una molécula de ácido nucleico objetivo que comprende la puesta en contacto de una muestra de ácido nucleico con un complejo ribonucleoproteico de la invención como se ha descrito anteriormente, o con un complejo Cascada y un componente de ARN distinto de la invención como se ha descrito anteriormente, y en el que la secuencia del componente de ARN (que incluye el complejo ribonucleoproteico) es tal que discrimina entre un alelo normal y un

alelo mutante en virtud de un único cambio de base en la posición 6 de una secuencia contigua de 8 residuos nucleotídicos.

En realizaciones de la invención, la molécula de ARN puede tener una longitud en el intervalo de 35-75 residuos. En realizaciones preferentes, la porción del ARN que es complementaria y se utiliza para dirigir una secuencia de ácido nucleico deseada es de 32 o 33 residuos de longitud. (En el contexto de un ARNcr de origen natural, esto correspondería a la porción espaciadora, como se muestra en la figura 1 de Semenova y col. (2011)).

Un complejo ribonucleoproteico de la invención puede tener adicionalmente un componente de ARN que comprende 8 residuos en 5' en la secuencia de ARN que tiene al menos una complementariedad sustancial con la secuencia objetivo de ácido nucleico. (Se entenderá que la secuencia de ARN que tiene al menos una complementariedad sustancial con la secuencia objetivo de ácido nucleico se corresponde en el contexto de un ARNcr como la secuencia espaciadora. Se consideraría que la secuencia flanqueante 5' del ARN corresponde al 5' mango de un ARNcr. Esto se muestra en la figura 1 de Semenova y col. (2011)).

Un complejo ribonucleoproteico de la invención puede tener una secuencia 3' que forma una horquilla y un bucle tetranucleótido en la secuencia de ARN que tiene al menos una complementariedad sustancial con la secuencia objetivo de ADN. (En el contexto de ARNcr, esto correspondería a un mango 3' que flanquea la secuencia espaciadora como se muestra en la figura 1 de Semenova y col. (2011)).

En algunas realizaciones, el ARN puede ser un ARN CRISPR (ARNcr).

Las proteínas y complejos Cascada de la invención se pueden caracterizar *in vitro* en términos de su actividad de asociación con el componente de guiado de ARN para formar un complejo ribonucleoproteico en presencia del ácido nucleico objetivo (que puede ser ADN o ARN). Puede utilizarse un ensayo de desplazamiento de movilidad electroforética (EMDE) como ensayo funcional para la interacción de complejos de la invención con sus objetivos de ácido nucleico. Básicamente, el complejo Cascada-fracción funcional de la invención se mezcla con objetivos de ácido nucleico y la interacción estable del complejo Cascada-fracción funcional se controla por EMDE o mediante lectura específica de la fracción funcional, por ejemplo, escisión endonucleolítica del ADN objetivo al deseado sitio. Esto puede determinarse mediante un análisis adicional de la longitud del fragmento de restricción utilizando enzimas comercialmente disponibles con especificidades conocidas y sitios de escisión en una molécula de ADN objetivo.

La visualización de la unión de las proteínas o los complejos Cascada de la invención a ADN o ARN en presencia de ARN guía se puede lograr utilizando imágenes de exploración/microscopía de fuerzas atómicas (SFM/AFM) y esto puede proporcionar un ensayo para la presencia de complejos funcionales de la invención.

Se desvela una molécula de ácido nucleico que codifica al menos una subunidad proteica asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISP) seleccionada entre:

- a. una subunidad Cse1 que tiene una secuencia de aminoácidos de SEQ ID NO: 1 o una secuencia de al menos un 9 % de identidad con la misma;
- b. una subunidad Cse2 que tiene una secuencia de aminoácidos de SEQ ID NO: 2 o una secuencia de al menos un 20 % de identidad con la misma;
- c. una subunidad Cas7 que tiene una secuencia de aminoácidos de SEQ ID NO: 3 o una secuencia de al menos un 18 % de identidad con la misma;
- d. una subunidad Cas5 que tiene una secuencia de aminoácidos de SEQ ID NO: 4 o una secuencia de al menos un 17 % de identidad con la misma;
- e. una subunidad Cas6 que tiene una secuencia de aminoácidos de SEQ ID NO: 5 o una secuencia de al menos un 16 % de identidad con la misma; y

en el que al menos a, b, c, d o e incluye una secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina.

La secuencia de aminoácidos adicional que tiene actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico o cromatina se fusiona preferentemente con la subunidad de la proteína asociada a CRISPR.

En los ácidos nucleicos, la secuencia de nucleótidos puede ser la que codifica la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5, respectiva o al definir el intervalo de secuencias variantes de la misma, puede ser una secuencia hibridable con esa secuencia de nucleótidos, preferentemente en condiciones rigurosas, más preferentemente en condiciones de rigurosidad muy alta. Una variedad de condiciones de hibridación rigurosa será familiar para el lector experto en el campo. La hibridación de una molécula de ácido nucleico se produce cuando dos moléculas de ácido nucleico complementarias experimentan una cantidad de enlaces de hidrógeno entre sí conocida como emparejamiento de bases Watson-Crick. La rigurosidad de la hibridación puede variar en función de las condiciones ambientales (es decir, químicas/físicas/biológicas) que rodean los ácidos nucleicos, la temperatura, la naturaleza del procedimiento de hibridación y la composición y longitud de las moléculas de ácido

nucleico utilizadas. Los cálculos con respecto a las condiciones de hibridación requeridas para alcanzar grados particulares de rigurosidad se discuten en Sambrook y col., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); y Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes Part I*, Capítulo 2 (Elsevier, Nueva York, 1993). T_m es la temperatura a la que el 50 % de una cadena dada de una molécula de ácido nucleico se hibrida con su cadena complementaria. Lo siguiente es un conjunto a modo de ejemplo de condiciones de hibridación y no es limitante:

Muy alta rigurosidad (permite secuencias que comparten al menos 90 % de identidad para hibridarse)

Hibridación:	5x SSC a 65 °C durante 16 horas
Lavar dos veces:	2x SSC a temperatura ambiente (TA) durante 15 minutos cada vez
Lavar dos veces:	0,5x SSC a 65 °C durante 20 minutos cada vez

Alta rigurosidad (permite secuencias que comparten al menos 80 % de identidad) para hibridarse)

Hibridación:	5x-6x SSC a 65 °C durante 16-20 horas
Lavar dos veces:	2x SSC a TA durante 5-20 minutos cada vez
Lavar dos veces:	1x SSC a 55 °C-70 °C durante 30 minutos cada vez

Baja rigurosidad (permite secuencias que comparten al menos 50 % de identidad para hibridarse)

Hibridación:	6x SSC a TA a 55 °C durante 16-20 horas
Lavar al menos dos veces:	2x-3x SSC a TA durante 20-30 minutos cada vez

La molécula de ácido nucleico puede ser una molécula de ácido nucleico aislada y puede ser un ARN o una molécula de ADN.

La secuencia de aminoácidos adicional puede seleccionarse entre una helicasa, una nucleasa, una nucleasa-helicasa (p. ej., Cas3), una ADN metiltransferasa (p. ej., Dam), una ADN desmetilasa, una histona metiltransferasa, una histona desmetilasa, una acetilasa, una desacetilasa, una fosfatasa, una quinasa, un (co)activador de la transcripción, una subunidad de ARN polimerasa, un represor de la transcripción, una proteína de unión al ADN, una proteína estructural del ADN, una proteína marcadora, una proteína indicadora, una proteína fluorescente, una proteína de unión al ligando (p. ej., mCherry o una proteína de unión a metales pesados), un péptido señal (p. ej., una secuencia de señal Tat), una secuencia de localización subcelular (p. ej., secuencia de localización nuclear) o un epítipo de anticuerpo. La secuencia de aminoácidos adicional puede ser, o de una proteína diferente del organismo del que se deriva(n) la(s) subunidad(es) de proteína Cascada relevante(s).

La invención incluye un vector de expresión que comprende una molécula de ácido nucleico como se ha definido anteriormente. Un vector de expresión puede contener la secuencia de nucleótidos que codifica una única subunidad de la proteína Cascada y también la secuencia de nucleótidos que codifica la secuencia de aminoácidos adicional, por lo que durante la expresión, la subunidad y la secuencia adicional se fusionan. Otros vectores de expresión pueden comprender secuencias de nucleótidos que codifican solo una o más subunidades de la proteína Cascada que no están fusionadas con ninguna secuencia de aminoácidos adicional.

La secuencia de aminoácidos adicional con actividad de modificación de un ácido nucleico o cromatina puede fusionarse a cualquiera de las subunidades Cascada a través de un polipéptido enlazador. El enlazador puede tener cualquier longitud hasta aproximadamente 60 o hasta aproximadamente 100 restos aminoácidos. Preferentemente, el enlazador tiene un número de aminoácidos en el intervalo de 10 a 60, más preferentemente 10-20. Los aminoácidos son preferentemente aminoácidos polares y/o pequeños y/o cargados (p. ej., Gln, Ser, Thr, Pro, Ala, Glu, Asp, Lys, Arg, His, Asn, Cys, Tyr). El péptido enlazador está diseñado preferentemente para obtener el espaciado y el posicionamiento correctos de la fracción funcional fusionada y la subunidad de Cascada a la que se fusiona la fracción para permitir una interacción apropiada con el nucleótido objetivo.

Un vector de expresión de la invención (con o sin secuencia de nucleótidos que codifica restos aminoácidos que en la expresión se fusionará con una subunidad de proteína Cascada) puede comprender además una secuencia que codifica una molécula de ARN como se ha definido anteriormente. Por consiguiente, dichos vectores de expresión se pueden utilizar en un huésped apropiado para generar una ribonucleoproteína de la invención que puede dirigirse a una secuencia de nucleótidos deseada.

Por consiguiente, la invención también proporciona un procedimiento de modificación, visualización o activación o represión de la transcripción de un ácido nucleico objetivo que comprende la puesta en contacto del ácido nucleico con un complejo ribonucleoproteico como se ha definido anteriormente. La modificación puede efectuarse mediante la escisión del ácido nucleico o la unión al mismo.

La invención también incluye un procedimiento de modificación, visualización o activación o represión de la transcripción de un ácido nucleico objetivo que comprende la puesta en contacto del ácido nucleico con un complejo proteico Cascada como se ha definido anteriormente, más una molécula de ARN como se ha definido anteriormente.

Según los procedimientos anteriores, la modificación, visualización o activación o represión de la transcripción de un ácido nucleico objetivo puede por lo tanto llevarse a cabo *in vitro* y en un entorno libre de células; es decir, el procedimiento se lleva a cabo como una reacción bioquímica ya sea libre en solución o si implica una fase sólida. El ácido nucleico objetivo se puede unir a una fase sólida, por ejemplo.

En un entorno libre de células, el orden de adición de cada uno de los ácidos nucleicos objetivo, el complejo proteico Cascada y la molécula de ARN queda a opción del experto en la materia. Los tres componentes se pueden añadir simultáneamente, secuencialmente en cualquier orden deseado, o por separado en diferentes momentos y en el orden deseado. De este modo, es posible añadir el ácido nucleico objetivo y el ARN simultáneamente a una mezcla de reacción y luego añadir el complejo proteico Cascada de la invención por separado y más tarde en una secuencia de etapas de procedimiento específicas.

La modificación, visualización o activación o represión de la transcripción de un ácido nucleico objetivo se puede realizar *in situ* en una célula, ya sea una célula aislada o como parte de un tejido, órgano u organismo multicelular. Por lo tanto, en el contexto de tejidos y órganos completos, y en el contexto de un organismo, el procedimiento puede llevarse a cabo *in vivo* o puede llevarse a cabo aislando una célula del tejido, órgano completo u organismo y luego devolviendo la célula tratada con un complejo ribonucleoproteico a su ubicación anterior, o una ubicación diferente, ya sea dentro del mismo organismo u organismo diferente. Por lo tanto, el procedimiento incluiría aloinjertos, autoinjertos, isoinjertos y xenoinjertos.

En estas realizaciones, el complejo ribonucleoproteico o el complejo proteico Cascada de la invención requiere una forma apropiada de administración a la célula, que será bien conocida por los expertos en la materia, incluyendo la microinyección, ya sea en el citoplasma celular o en el núcleo.

Además, cuando se presenta por separado, la molécula de ARN requiere una forma apropiada de administración a una célula, ya sea simultánea, separada o secuencialmente con el complejo proteico Cascada. Tales formas de introducir ARN en las células son bien conocidas por los expertos en la materia y pueden incluir la administración *in vivo* o *ex vivo* por procedimientos de transfección convencionales. Pueden utilizarse procedimientos físicos, tales como microinyección y electroporación, así como coprecipitación de calcio, y polímeros y lípidos catiónicos disponibles comercialmente, y péptidos que penetran en la célula, partículas que penetran en la célula (pistola de genes). Por ejemplo, los virus pueden utilizarse como vehículos de administración, ya sea al citoplasma y/o al núcleo, p. ej., a través de la fusión (reversible) del complejo proteico Cascada de la invención o un complejo ribonucleoproteico de la invención con la partícula viral. Se puede utilizar administración viral (p. ej., administración de adenovirus) o administración mediada por *Agrobacterium*.

La invención también incluye un procedimiento de modificación, visualización, o activación o represión de la transcripción de un ácido nucleico objetivo en una célula, que comprende transfectar, transformar o transducir la célula con cualquiera de los vectores de expresión como se ha descrito anteriormente. Los procedimientos de transfección, transformación o transducción son de los tipos bien conocidos por un experto en la materia. Cuando existe un vector de expresión utilizado para generar la expresión de un complejo Cascada de la invención y cuando el ARN se añade directamente a la célula, entonces se puede utilizar el mismo procedimiento o un procedimiento diferente de transfección, transformación o transducción. De forma similar, cuando se utiliza un vector de expresión para generar expresión de un complejo Cascada-fusión funcional de la invención y cuando se utiliza otro vector de expresión para generar la expresión del ARN *in situ*, entonces puede ser utilizado el mismo procedimiento o un procedimiento diferente de transfección, transformación o transducción.

En otras realizaciones, el ARNm que codifica el complejo Cascada de la invención se introduce en una célula de modo que el complejo Cascada se expresa en la célula. El ARN que guía el complejo Cascada a la secuencia objetivo deseada también se introduce en la célula, ya sea de manera simultánea, por separado o secuencialmente a partir del ARNm, de manera que se forma el complejo ribonucleoproteico necesario en la célula.

En los procedimientos mencionados anteriormente de modificación o visualización de un ácido nucleico objetivo, la secuencia de aminoácidos adicional puede ser un marcador y el marcador se asocia con el ácido nucleico objetivo; preferentemente en el que el marcador es una proteína; opcionalmente una proteína fluorescente, p. ej., proteína verde fluorescente (VPF) o proteína amarilla fluorescente (PAF) o mCherry. Ya sea *in vitro*, *ex vivo* o *in vitro*, los procedimientos de la invención pueden utilizarse para visualizar directamente un locus objetivo en una molécula de ácido nucleico, preferentemente en la forma de una estructura de orden superior tal como un plásmido o cromosoma superenrollado o un ácido nucleico objetivo monocatenario tal como ARNm. La visualización directa de un locus objetivo puede utilizar micrografía electrónica o microscopía de fluorescencia.

Se pueden utilizar otros tipos de marcadores para marcar el ácido nucleico objetivo incluyendo moléculas de colorante orgánico, marcadores radiactivos y marcadores de espín que pueden ser moléculas pequeñas.

En los procedimientos de la invención descritos anteriormente, el ácido nucleico objetivo es ADN; preferentemente ADNbc aunque el objetivo puede ser ARN; preferentemente ARNm.

En los procedimientos de la invención de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico objetivo en el que el ácido nucleico objetivo es ADNbc, la secuencia de

aminoácidos adicional con actividad de modificación de un ácido nucleico o cromatina es una endonucleasa FokI. De esta forma, el corte de ADN específico de secuencia única se puede modificar por ingeniería utilizando los complejos Cascada-fracciones funcionales. La secuencia elegida del componente de ARN del complejo ribonucleoproteico final proporciona la especificidad de secuencia deseada para la acción de la secuencia de aminoácidos adicional.

- 5 Por lo tanto, la invención también proporciona un procedimiento de unión final no homóloga de una molécula de ADNbc en una célula en un locus deseado para eliminar al menos una parte de una secuencia de nucleótidos de la molécula de ADNbc; opcionalmente para bloquear la función de un gen o genes, en el que el procedimiento comprende realizar roturas bicatenarias utilizando cualquiera de los procedimientos de modificación de un ácido nucleico objetivo como se ha descrito anteriormente.
- 10 La invención proporciona además un procedimiento de recombinación homóloga de un ácido nucleico en una molécula de ADNbc en una célula en un locus deseado para modificar una secuencia de nucleótidos existente o insertar una secuencia de nucleótidos deseada, en el que el procedimiento comprende realizar una rotura monocatenaria o bicatenaria en el locus deseado utilizando cualquiera de los procedimientos de modificación de un ácido nucleico objetivo como se ha descrito anteriormente.
- 15 Por lo tanto, la invención también proporciona un procedimiento de modificación, activación o represión de la expresión génica en un organismo que comprende modificar, activar la transcripción o reprimir la transcripción de una secuencia de ácido nucleico objetivo según cualquiera de los procedimientos descritos anteriormente, en el que el ácido nucleico es ADNbc y la fracción se selecciona entre una enzima modificadora del ADN (p. ej., una desmetilasa o desacetilasa), un activador de la transcripción o un represor de la transcripción.
- 20 La invención proporciona adicionalmente un procedimiento de modificación, activación o represión de la expresión génica en un organismo que comprende modificar, activar la transcripción o reprimir la transcripción de una secuencia de ácido nucleico objetivo según cualquiera de los procedimientos descritos anteriormente, en el que el ácido nucleico es un ARNm y la fracción funcional es una ribonucleasa; opcionalmente seleccionado entre una endonucleasa, una exonucleasa 3' o una exonucleasa 5'.
- 25 En cualquiera de los procedimientos de la invención como se ha descrito anteriormente, la célula que se somete al procedimiento puede ser una procariota. De forma similar, la célula puede ser una célula eucariota, p. ej., una célula vegetal, una célula de insecto, una célula de levadura, una célula fúngica, una célula de mamífero o una célula humana. Cuando la célula es de un mamífero o ser humano, puede ser una célula madre (pero puede no ser una célula madre embrionaria humana). Dichas células madre para su uso en la invención son preferentemente células madre aisladas. Opcionalmente según cualquier procedimiento de la invención, una célula se transfecta *in vitro*.
- 30

Preferentemente, sin embargo, en cualquiera de los procedimientos de la invención, el ácido nucleico objetivo tiene una estructura terciaria específica, opcionalmente superenrollada, más preferentemente en la que el ácido nucleico objetivo está negativamente superenrollado. Ventajosamente, los complejos ribonucleoproteico de la invención, ya sea producidos *in vitro*, o formados dentro de las células, o formados dentro de las células a través de la maquinaria de expresión de la célula, pueden utilizarse para dirigir un locus al que de otro modo sería difícil acceder para aplicar la actividad funcional de un componente deseado, ya sea marcado o etiquetado de una secuencia específica, modificación de la estructura de ácido nucleico, activación o desactivación de la expresión génica o modificación de la secuencia objetivo en sí misma que implica corte monocatenario o bicatenario seguido de inserción de uno o más residuos de nucleótidos o un casete.

- 40 La invención también incluye una composición farmacéutica que comprende un complejo proteico Cascada o un complejo ribonucleoproteico de la invención como se ha descrito anteriormente.

Asimismo se desvela una composición farmacéutica que comprende un ácido nucleico aislado o un vector de expresión de la invención como se ha descrito anteriormente.

- 45 También se desvela un kit que comprende un complejo proteico Cascada de la invención como se ha descrito anteriormente más una molécula de ARN de la invención como se ha descrito anteriormente.

También se desvela un complejo proteico Cascada o un complejo ribonucleoproteico o un ácido nucleico o un vector, como se ha descrito anteriormente para su uso como medicamento.

- 50 La invención permite una variedad de posibilidades para alterar físicamente el ADN de huéspedes procarióticos o eucarióticos en un locus genómico especificado, o cambiar los patrones de expresión de un gen en un locus dado. El ADN genómico del huésped puede escindirse o modificarse mediante metilación, visualizarse mediante fluorescencia, activarse transcripcionalmente o reprimirse por dominios funcionales tales como nucleasas, metilasas, proteínas fluorescentes, activadores o represores de la transcripción, respectivamente, fusionados a subunidades Cascada adecuadas. Además, la capacidad de unión a ARN guiado por ARN de Cascada permite el control del tráfico de ARN en células vivas utilizando proteínas de fusión Cascada fluorescentes, y proporciona formas de
- 55 secuestrar o destruir ARNm del huésped que causa interferencia con los niveles de expresión génica de una célula huésped.

En cualquiera de los procedimientos de la invención, el ácido nucleico objetivo puede definirse, preferentemente si es ADNbc, mediante la presencia de al menos uno de los siguientes tripletes de nucleótidos: 5'-CTT-3', 5'-CAT-3', 5'-CCT-3' o 5'-CTC-3' (o 5'-CUU-3', 5'-CAU-3', 5'-CCU-3' o 5'-CTC-3' si el objetivo es un ARN). La ubicación del triplete se encuentra en la cadena objetivo adyacente a la secuencia a la que se hibrida el componente de la molécula de ARN de una ribonucleoproteína de la invención. El triplete marca el punto en la secuencia de cadena objetivo en la que el emparejamiento de bases con el componente de la molécula de ARN de la ribonucleoproteína no tiene lugar en una dirección 5' a 3' (corriente abajo) del objetivo (mientras se produce corriente arriba de la secuencia objetivo desde ese punto sometido a la longitud preferente de la secuencia de ARN del componente de la molécula de ARN de la ribonucleoproteína de la invención). En el contexto de un sistema CRISPR de tipo I nativo, los tripletes corresponden a lo que se conoce como "MAP" (motivo adyacente al protoespaciador). Para objetivos ADNmc o ARNmc, la presencia de uno de los tripletes no es tan necesaria.

La invención se describirá ahora en detalle y con referencia a los ejemplos específicos y dibujos en los que:

Las Figuras y los Ejemplos que no pertenecen a la invención reivindicada son solo para fines ilustrativos.

La **Figura 1** muestra los resultados de los ensayos de desplazamiento en gel en los que Cascada se une al ADN plasmídico negativamente superenrollado (nSE) pero no al ADN relajado. **A)** Desplazamiento en gel de ADN plasmídico nSE con J3-Cascada, que contiene un ARNcr dirigido (J3). pUC-λ se mezcló con cantidades crecientes 2 veces de J3-Cascada, a partir de una relación molar pUC-λ:Cascada de 1:0,5 hasta una relación molar de 1:256. El primer y el último carril contienen solo pUC-λ. **B)** Desplazamiento en gel como en (A) con R44-Cascada que contiene un ARNcr no dirigido (R44). **C)** Desplazamiento en gel como en (A) con Nt.BspQI escotado con pUC-λ. **D)** Desplazamiento en gel como en (A) con PdmI linearizado con pUC-λ. **E)** El ajuste de la fracción pUC-λ unida a J3-Cascada representada frente a la concentración libre de J3-Cascada proporciona la constante de disociación (Kd) para la unión específica. **F)** El ajuste de la fracción pUC-λ unida a R44-Cascada representada frente a la concentración libre de R44-Cascada proporciona la constante de disociación (Kd) para la unión no específica. **G)** Unión específica de Cascada al protoespaciador controlado por análisis de restricción, utilizando el único sitio de restricción BsmI en la secuencia del protoespaciador. Los carriles 1 y 5 contienen solo pUC-λ. Los carriles 2 y 6 contienen pUC-λ mezclado con Cascada. Los carriles 3 y 7 contienen pUC-λ mezclado con Cascada y la posterior adición de BsmI. Los carriles 4 y 8 contienen pUC-λ mezclado con BsmI. **H)** Desplazamiento en gel de pUC-λ unido a Cascada con posterior escisión Nt.BspQI de una cadena del plásmido. Los carriles 1 y 6 contienen solo pUC-λ. Los carriles 2 y 7 contienen pUC-λ mezclado con Cascada. Los carriles 3 y 8 contienen pUC-λ mezclado con Cascada y posterior escotamiento con Nt.BspQI. Los carriles 4 y 9 contienen pUC-λ mezclado con Cascada, seguido de la adición de una sonda de ADNmc complementario a la cadena desplazada en el bucle R y posterior escotado con Nt.BspQI. Los carriles 5 y 10 contienen pUC-λ escotado con Nt.BspQI. **I)** Desplazamiento en gel de pUC-λ unido a Cascada con subsiguiente escotado con Nt.BspQI del plásmido. Los carriles 1 y 6 contienen solo pUC-λ. Los carriles 2 y 7 contienen pUC-λ mezclado con Cascada. Los carriles 3 y 8 contienen pUC-λ mezclado con Cascada y posterior escisión de Nt.BspQI. Los carriles 4 y 9 contienen pUC-λ mezclado con Cascada, seguido de la adición de una sonda de ADNmc complementario a la cadena desplazada en el bucle R y posterior escisión con Nt.BspQI. Los carriles 5 y 10 contienen pUC-λ escindido con Nt.BspQI. **J)** Desplazamiento en gel de pUC-λ unido a Cascada con posterior escisión EcoRI de ambas cadenas del plásmido. Los carriles 1 y 6 contienen solo pUC-λ. Los carriles 2 y 7 contienen pUC-λ mezclado con Cascada. Los carriles 3 y 8 contienen pUC-λ mezclado con Cascada y posterior escisión de EcoRI. Los carriles 4 y 9 contienen pUC-λ mezclado con Cascada, seguido de la adición de una sonda de ADNmc complementario a la cadena desplazada en el bucle R y posterior escisión con EcoRI. Los carriles 5 y 10 contienen pUC-λ escindido con EcoRI.

La Figura 2 muestra micrografías de fuerzas de barrido que demuestran cómo Cascada induce la flexión del ADN objetivo tras la unión del protoespaciador. **A-P)** Imágenes de microscopía de fuerzas de barrido de ADN plasmídico nSE con J3-Cascada que contiene un ARNcr objetivo (J3). pUC-λ se mezcló con J3-Cascada en una relación pUC-λ:Cascada de 1:7. Cada imagen muestra una superficie de 500 x 500 nm. Los puntos blancos corresponden a Cascada.

La **Figura 3** muestra cómo el análisis de BiFC revela que Cascada y Cas3 interactúan tras el reconocimiento del objetivo. **A)** fluorescencia de Venus de células que expresan CascadaΔCse1 y CRISPR 7Tm, que se dirige a 7 protoespaciadores en el genoma del fago λ, y a las proteínas de fusión Cse1-N155Venus y Cas3-C85Venus. **B)** Imagen de campo claro de las células en (A). **C)** Superposición de (A) y (B). **D)** Fluorescencia de Venus de células infectadas por fago λ que expresan las proteínas de fusión CascadaΔCse1 y CRISPR 7Tm y Cse1-N155Venus y Cas3-C85Venus. **E)** Imagen de campo claro de las células en (G). **F)** Superposición de (G) y (H). **G)** Fluorescencia de Venus de células infectadas por fago λ que expresan las proteínas CascadaΔCse1 y CRISPR R44 no dirigido y N155Venus y C85Venus. **H)** Imagen de campo claro de las células en (J). **I)** Superposición de (J) y (K). **J)** Promedio de la intensidad de fluorescencia de 4-7 células individuales de cada cepa, según se determina utilizando la herramienta de perfil del visor LSM (Carl Zeiss).

La **Figura 4** muestra las actividades de nucleasa y helicasa Cas3 durante la interferencia CRISPR. **A)** Las células BL21-AI competentes que expresan Cascada, un mutante Cas3 y CRISPR J3 se transformaron con pUC-λ. Se representan las unidades formadoras de colonias por microgramo pUC-λ (ufc/μg de ADN) para cada una de

las cepas que expresan un mutante Cas3. Las células que expresan Cas3 y CRISPR J3 o CRISPR R44 ts sirven como controles positivos y negativos, respectivamente. **B)** Las células BL21-AI que portan plásmidos que codifican Cascada, Cas3 y CRISPR así como pUC-λ se cultivan en condiciones que suprimen la expresión de genes cas y CRISPR. En t = 0 la expresión es inducida. Se muestra el porcentaje de células que perdieron pUC-λ con el tiempo, según lo determinado por la relación de células resistentes a la ampicilina y sensibles a la ampicilina.

La **Figura 5** muestra cómo un complejo de fusión Cascada-Cas3 proporciona resistencia *in vivo* y tiene actividad de nucleasa *in vitro*. **A)** SDS-PAGE teñido con azul de Coomassie de Cascada purificada y complejo de fusión Cascada-Cas3. **B)** Eficiencia de formación de placas del fago λ en células que expresan el complejo de fusión Cascada-Cas3 y un CRISPR dirigido (J3) o no dirigido (R44) y en células que expresan Cascada y Cas3 por separado junto con CRISPR dirigido (J3). **C)** Desplazamiento en gel (en ausencia de iones metálicos divalentes) del plásmido objetivo nSE con complejo de fusión J3-Cascada-Cas3. pUC-λ se mezcló con cantidades crecientes 2 veces de J3-Cascada-Cas3, a partir de una relación molar pUC-λ:J3-Cascada-Cas3 de 1:0,5 hasta una relación molar de 1:128. El primer y último carril contienen solo pUC-λ. **D)** Desplazamiento en gel (en ausencia de iones metálicos divalentes) del plásmido no objetivo nSE con complejo de fusión J3-Cascada-Cas3. Se mezcló pUC-p7 con cantidades crecientes 2 veces de J3-Cascada-Cas3, a partir de una relación molar pUC-p7:J3-Cascada-Cas3 de 1:0,5 hasta una relación molar de 1:128. El primer y último carril contienen solo pUC-p7. **E)** Incubación del plásmido objetivo nSE (pUC-λ, izquierda) o plásmido no objetivo nSE (pUC-p7, derecha) con J3-Cascada-Cas3 en presencia de MgCl₂ 10 mM. Los carriles 1 y 7 contienen solo plásmido. **F)** Ensayo al igual que en (E) en presencia de ATP 2 mM. **G)** Ensayo al igual que en (E) con el complejo mutante J3-Cascada-Cas3K320N. **H)** Ensayo al igual que en (G) en presencia de ATP 2 mM.

La **Figura 6** es un diagrama esquemático que muestra un modelo de la vía de tipo I de interferencia CRISPR en *E. coli*.

La **Figura 7** es un diagrama esquemático que muestra cómo se utiliza una realización de fusión Cascada-FokI de la invención para crear dímeros FokI que cortan ADNbc para producir extremos romos como parte de un proceso de unión final no homóloga o recombinación homóloga.

La **Figura 8** muestra cómo el análisis de BiFC revela que Cascada y Cas3 interactúan tras el reconocimiento del objetivo. Superposición de la imagen de campo claro y fluorescencia de Venus de células que expresan Cascada sin Cse1, Cse1-N155Venus y Cas3-C85Venus ni CRISPR 7Tm, que se dirige a 7 protoespaciadores en el genoma del fago Lambda, o al CRISPR R44 sin dirigir. Las células que expresan CRISPR 7Tm son fluorescentes solo cuando son infectadas por el fago Lambda, mientras que las células que expresan CRISPR R44 no son fluorescentes. Los puntos fluorescentes altamente intensos (fuera de las células) se deben a cristales de sal que reflejan la luz. Las barras blancas corresponden a 10 micrómetros.

La **Figura 9** muestra secuencias de pUC-λ de 4 clones [SEQ ID NOs: 39-42] que codifican CRISPR J3, Cascada y Cas3 (ts o S483AT485A) indican que estos son mutantes de escape que llevan delecciones (parciales) del protoespaciador o portan una única mutación puntual en la región semilla, lo que explica la incapacidad para curar estos plásmidos.

La **Figura 10** muestra los alineamientos de secuencia de genes *cas3* de organismos que contienen el sistema CRISPR/Cas de tipo I-E. Alineamiento de genes *cas3-cse1* de *Streptomyces sp.* SPB78 (1ª secuencia, número de acceso: ZP_07272643.1) [SEQ ID NO: 43], en *Streptomyces griseus* (2ª secuencia, número de acceso YP_001825054) [SEQ ID NO: 44], y en *Catenulispora acidiphila* DSM 44928 (3ª secuencia, número de acceso YP_003114638) [SEQ ID NO: 45] y una proteína de fusión Cas3-Cse1 de *E. coli* [SEQ ID NO: 46] que incluye la secuencia de unión de polipéptido de *S. griseus*.

La **Figura 11** muestra el diseño de un par de Cascada^{KKR/ELD} nucleasa en el que los dominios de nucleasa FokI están mutados de modo que solo los heterodímeros que consisten en los dominios de nucleasa KKR y ELD son diferentes y la distancia entre los sitios de unión opuestos puede variarse para determinar la distancia óptima entre un par de nucleasas Cascada.

La **Figura 12** es un diagrama esquemático que muestra la dirección del genoma mediante un par de nucleasas Cascada-FokI.

La **Figura 13** muestra un gel SDS PAGE de complejos Cascada-nucleasa.

La **Figura 14** muestra geles de electroforesis de ensayos de escisión *in vitro* de Cascada^{KKR/ELD} en ADN plasmídico.

La **Figura 15** muestra los patrones de escisión y la frecuencia de Cascada^{KKR/ELD} [SEQ ID NO: 47].

Ejemplos - Materiales y procedimientos utilizados

Cepas, clonación de genes, plásmidos y vectores

Las cepas BL21-AI de *E. coli* y BL21 de *E. coli* (DE3) se utilizaron de principio a fin. La Tabla 1 enumera todos los plásmidos utilizados en este estudio. Los pWUR408, pWUR480, pWUR404 y pWUR547 previamente descritos se utilizaron para la producción de Strep-tag II R44-Cascada, y pWUR408, pWUR514 y pWUR630 se utilizaron para la producción de Strep-tag II J3-Cascada (Jore y col., (2011) *Nature Structural & Molecular Biology* 18, 529-536; Semenova y col., (2011) *Proceedings of the National Academy of Sciences of The United States of America* 108, 10098-10103). pUC-λ (pWUR610) y pUC-p7 (pWUR613) se han descrito en otra parte (Jore y col., 2011; Semenova y col., 2011). La proteína C85Venus está codificada por pWUR647, que corresponde a pET52b (Novagen) que contiene la construcción sintética GA1070943 (Tabla 2) (Geneart) clonada entre los sitios BamHI y NotI. La proteína N155Venus está codificada por pWUR648, que corresponde a pRSF1b (Novagen) que contiene la construcción sintética GA1070941 (Tabla 2) (Geneart) clonada entre los sitios NotI y XhoI. La proteína de fusión Cas3-C85Venus está codificada por pWUR649, que corresponde a pWUR647 que contiene el producto de amplificación Cas3 utilizando los cebadores BG3186 y BG3213 (Tabla 3) entre los sitios NcoI y BamHI. La proteína de fusión CasA-N155Venus está codificada por pWUR650, que corresponde a pWUR648 que contiene el producto de amplificación CasA utilizando los cebadores BG3303 y BG3212 (Tabla 3) entre los sitios NcoI y BamHI. CRISPR 7Tm está codificado por pWUR651, que corresponde a pACYCDuet-1 (Novagen) que contiene la construcción sintética GA1068859 (Tabla 2) (Geneart) clonada entre los sitios NcoI y KpnI. Cascada que codifica pWUR400, CascadaΔCse1 que codifica WUR401 y Cas3 que codifica pWUR397 se han descrito anteriormente (Jore y col., 2011). Cas3H74A que codifica pWUR652 se construyó utilizando mutagénesis dirigida al sitio de pWUR397 con los cebadores BG3093, BG3094 (Tabla 3).

Tabla 1 - Plásmidos utilizados

Plásmidos	Descripción y orden de genes (5'3')	Sitios restricción	de	Cebadores	Fuente
pWUR397	<i>cas3</i> en pRSF-1b, sin marcadores				1
pWUR400	<i>casA-casB-casC-casD-casE</i> en pCDF-1b, sin marcadores				1
pWUR401	<i>casB-casC-casD-casE</i> en pCDF-1b, sin marcadores				1
pWUR404	<i>casE</i> en pCDF-1b, sin marcadores				1
pWUR408	<i>casA</i> en pRSF-1b, sin marcadores				1
pWUR480	<i>casB</i> con Strep-tag II (N-term)- <i>casC-casD</i> en pET52b				1
pWUR514	<i>casB</i> con Strep-tag II (N-term)- <i>casC-casDCasE</i> en pET52b				2
pWUR547	<i>E. coli</i> R44 CRISPR, 7x espaciador n.º 2, en pACYCDuet-1				2
pWUR613	pUC-p7; pUC19 que contiene R44-protospaciador en un amplicón P7 de fago de 350 pb				2
pWUR630	CRISPR poli J3, 5x espaciador J3 en pACYCDuet-1	NcoI/KpnI			Este estudio
pWUR610	pUC-λ; pUC19 que contiene J3-protospaciador en un amplicón de fago λ 350 pb				3
pWUR647	<i>C85Venus</i> ; GA1070943 (Tabla S1) en pET52b	BamHI/NotI			Este estudio
pWUR648	<i>N155Venus</i> ; GA1070941 (Tabla S1) en pRSF1b	NotI/XhoI			Este estudio
pWUR649	<i>cas3-C85Venus</i> ; pWUR647 que contiene amplicón <i>cas3</i>	NcoI/BamHI		BG3186 + BG3213	Este estudio
pWUR650	<i>casA-N155Venus</i> pWUR648 que contiene amplicón <i>casA</i>	NcoI/NotI		BG3303 + BG3212	Este estudio
pWUR651	CRISPR 7Tm; GA1068859 (Tabla S1) en pACYCDuet-1	NcoI/KpnI			Este estudio
	<i>casB</i> con Strep-tag II (N-term)- <i>casC-casDCasE</i> en pCDF-1b				Este estudio
	fusión <i>cas3-casA</i>				Este estudio

(continuación)

Plásmidos	Descripción y orden de genes (5'3')	Sitios restricción	de	Cebadores	Fuente
	fusión <i>cas3H74A-CasA</i>				Este estudio
	fusión <i>cas3D75A-CasA</i>				Este estudio
	fusión <i>cas3K320N-CasA</i>				Este estudio
	fusión <i>cas3D452N-CasA</i>				Este estudio

La fuente 1 en la tabla anterior es de Brouns y col (2008) *Science* 321, 960-964.

La fuente 2 en la tabla anterior es de Jore y col (2008) *Nature Structural & Molecular Biology* 18, 529-537.

5

Tabla 2 - Construcciones sintéticas

GA1070943

ACTGGAAAGCGGGCAGTGAAAGGAAGGCCCATGAGGCCAGTTAATTAAGCGGA
 TCCTGGCGGCGGCAGCGGCGGCGGCAGCGACAAGCAGAAGAACGGCATCAAGG
 CGAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGTCAGCTCGCCGACC
 ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACC
 ACTACCTGAGCTACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
 ACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGA
 GCTGTACAAGTAAGCGGCCGCGGCGCGCCTAGGCCTTGACGGCCTTCCTTCAATT
 CGCCCTATAGTGAG [SEQ ID NO: 6]

GA1070941

CACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATTTAATTAAGCGGCC
 GCAGGCGGCGGCAGCGGCGGCGGCAGCATGGTGAGCAAGGGCGAGGAGCTGTT
 CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
 GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
 GAAGCTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCCTCGTGACC
 ACCCTCGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC
 ACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
 CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTTCGAGGGCGA
 CACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
 CATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCAC
 GGCCTAACTCGAGGGCGCGCCCTGGGCCTCATGGGCCTTCCGCTCACTGCCCGCT
 TTCCAG [SEQ ID NO: 7]

10

GA1068859

CACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATGAGCTCCATGGAAA
 CAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGT
 TGTTTTTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCC
 CGCGCCAGCGGGGATAAACC GGCGCCGATTGAAGGTCCGGTGGATGGCTTAAAAG
 AGTTCCCCGCGCCAGCGGGGATAAACC GGCGCAGGTACAGCAGGTAGCGCAGAT
 CATCAAGAGTTCCCCGCGCCAGCGGGGATAAACC GACTTCTCTCCGAAAAGTCA
 GGACGCTGTGGCAGAGTTCCCCGCGCCAGCGGGGATAAACC GCCTACGCGCTGA
 ACGCCAGCGGTGTGGTGAATGAGTTCCCCGCGCCAGCGGGGATAAACC GGTGTG
 GCCATGCACGCCTTTAACGGTGAAGTGGAGTTCCCCGCGCCAGCGGGGATAAAC
 CGCACGAACTCAGCCAGAACGACAAACAAAAGGCGAGTTCCCCGCGCCAGCGG
 GGATAAACC GGCAACAGTACGCGCCCCACGCTGACGGTTTCTGAGTTCCCCGCGC
 CAGCGGGGATAAACC GCAGCTCCCATTTTCAAACCCAGGTACCCTGGGCCTCATG
 GGCCTTCCGCTCACTGCCCGCTTTCCAG [SEQ ID NO: 8]

GA1047360

GAGCTCCCGGGCTGACGGTAATAGAGGCACCTACAGGCTCCGGTAAAACGGAAA
 CAGCGCTGGCCTATGCTTGGAACCTTATTGATCAACAAATTGCGGATAGTGTTAT
 TTTTGCCCTCCCAACACAAGCTACCGCGAATGCTATGCTTACGAGAATGGAAGCG
 AGCGCGAGCCACTTATTTTCATCCCCAAATCTTATTCTTGCTCATGGCAATTCACG
 GTTTAACCACTCTTTCAATCAATAAAATCACGCGCGATTACTGAACAGGGGGCAA
 GAAGAAGCGTGCGTTTCAGTGTGTGTCAGTGGTTGTACAAAGCAATAAGAAAGTG
 TTTCTTGCGCAAATCGGCGTTTGCACGATTGATCAGGTGTTGATTTCGGTATTGCC
 AGTTAAACACCGCTTTATCCGTGGTTTGGGAATTGGTAGATCTGTTTTAATTGTTA
 ATGAAGTTCATGCTTACGACACCTATATGAACGGCTTGCTCGAGGCAGTGCTCAA
 GGCTCAGGCTGATGTGGGAGGGAGTGTTATTCTTCTTTCCGCAACCCTACCAATG
 AAACAAAAACAGAAGCTTCTGGATACTTATGGTCTGCATACAGATCCAGTGGAA
 AATAACTCCGCATATCCACTCATTAAGTGGCGAGGTGTGAATGGTGCGCAACGTT
 TTGATCTGCTAGCGGATCCGGTACC [SEQ ID NO: 9]

Tabla 3 - Cebadores

BG3186	ATAGCGCCATGGAACCTTTTAAATATATATGCCATTA [SEQ ID NO: 10]
BG3213	ACAGTGGGATCCGCTTTGGGATTTGCAGGGATGACTCTGGT [SEQ ID NO: 11]
BG3303	ATAGCGTCATGAATTTGCTTATTGATACTGGATTCTGTACG [SEQ ID NO: 12]
BG3212	ACAGTGGCGGCCGCGCCATTTGATGGCCCTCCTTGCAGTTTAA [SEQ ID NO: 13]
BG3076	CGTATATCAAACCTTCCAATAGCATGAAGAGCAATGAAAAATAAC [SEQ ID NO: 14]
BG3449	ATGATACCGCGAGACCCACGCTC [SEQ ID NO: 15]
BG3451	CGGATAAAGTTGCAGGACCACTTC [SEQ ID NO: 16]

5

Producción y purificación de proteínas

La cascada se expresó y purificó como se describe (Jore y col., 2011). Durante toda la purificación, se utilizó un tampón que contenía HEPES 20 mM pH 7,5, NaCl 75 mM, DTT 1 mM, EDTA 2 mM para la resuspensión y el lavado. La elución de proteína se realizó en el mismo tampón que contenía destiobiotina 4 mM. El complejo de fusión Cascada-Cas3 se expresó y purificó de la misma manera, realizándose las etapas de lavado con HEPES 20 mM, pH 7,5, NaCl 200 mM y DTT 1 mM, y elución en HEPES 20 mM, pH 7,5, NaCl 75 mM, DTT 1 mM que contenía 4 mM de destiobiotina.

10

Ensayo de desplazamiento de movilidad electroforética

Se mezclaron los subcomplejos Cascada purificada o Cascada con pUC- λ en un tampón que contenía HEPES 20 mM, pH 7,5, NaCl 75 mM, DTT 1 mM, EDTA 2 mM, y se incubaron a 37 °C durante 15 minutos. Las muestras se analizaron durante la noche en un gel de agarosa TAE al 0,8 % y se tiñeron posteriormente con una dilución SyBR segura (Invitrogen) 1:10000 en TAE durante 30 minutos. La escisión con BsmI (Fermentas) o Nt.BspQI (New England Biolabs) se realizó en el tampón de reacción HEPES suplementado con MgCl₂ 5 mM.

Microscopía de fuerzas de barrido

La Cascada purificada se mezcló con pUC- λ (en una relación de 7:1, Cascada 250 nM, ADN 35 nM) en un tampón que contenía HEPES 20 mM, pH 7,5, NaCl 75 mM, DTT 0,2 mM, EDTA 0,3 mM y se incubó a 37 °C durante 15 minutos. Posteriormente, para la preparación de muestra de AFM, la mezcla de incubación se diluyó 10x en agua bidestilada y se añadió MgCl₂ a una concentración final de 1,2 mM. La deposición de los complejos de proteína-ADN y formación de imágenes se llevó a cabo como se ha descrito anteriormente (Dame y col., (2000) *Nucleic Acids Res.* 28: 3504-3510).

Microscopio fluorescente

Las células BL21-AI portadoras de CRISP en plásmidos codificadores de gen *cas* se cultivaron durante la noche a 37 °C en caldo Luria-Bertani (LB) que contenía ampicilina (100 µg/ml), kanamicina (50 µg/ml), estreptomycin (50 µg/ml) y cloranfenicol (34 µg/ml). El cultivo durante la noche se diluyó 1:100 en LB que contenía antibiótico reciente, y se cultivó durante 1 hora a 37 °C. La expresión de los genes *cas* y CRISPR se indujo durante 1 hora añadiendo L-arabinosa a una concentración final de 0,2 % e IPTG hasta una concentración final de 1 mM. Para la infección, las células se mezclaron con el fago Lambda a una multiplicidad de infección (MDI) de 4. Las células se aplicaron a portaobjetos de microscopio cubiertos con poli-L-lisina y se analizaron utilizando un microscopio de escaneo láser confocal Zeiss LSM510 basado en un microscopio invertido Axiovert, con un objetivo de inmersión en aceite 40x (N.A. de 1.3) y un láser de argón como fuente de excitación (514 nm) y detección a 530-600 nm. El estenopeico se ajustó a 203 µm para todas las mediciones.

Estudios de transformación de pUC- λ

Se inoculó LB que contenía kanamicina (50 µg/ml), estreptomycin (50 µg/ml) y cloranfenicol (34 µg/ml) a partir de un preinóculo de una noche y se cultivó hasta una DO₆₀₀ de 0,3. La expresión de genes *cas* y CRISPR se indujo durante 45 minutos con 0,2 % de L-arabinosa e IPTG 1 mM. Las células se recogieron por centrifugación a 4 °C y se hicieron competentes por resuspensión en tampón enfriado con hielo que contenía RbCl₂ 100 mM, MnCl₂ 50 mM, acetato de potasio 30 mM, CaCl₂ 10 mM y 15 % de glicerol, pH 5,8. Después de una incubación de 3 horas, las células se recogieron y se volvieron a suspender en un tampón que contenía MOPS 10 mM, RbCl 10 mM, CaCl₂ 75 mM, glicerol al 15 %, pH 6,8. La transformación se realizó añadiendo 80 ng de pUC- λ , seguido de un choque térmico de 1 minuto a 42 °C, y 5 minutos de choque frío en hielo. Las células siguientes se cultivaron en LB durante 45 minutos a 37 °C antes de la siembra en placas de LB-agar que contenían 0,2 % de L-arabinosa, IPTG 1 mM, ampicilina (100 µg/ml), kanamicina (50 µg/ml), estreptomycin (50 µg/ml) y cloranfenicol (34 µg/ml).

El curado del plásmido se analizó transformando las células BL21-AI que contienen el gen *cas* y los plásmidos que codifican CRISPR con pUC- λ , mientras se cultivan las células en presencia de glucosa al 0,2 % para suprimir la expresión del gen de la polimerasa T7. La expresión de los genes *cas* y CRISPR se indujo mediante la recogida de las células y la resuspensión en LB que contenía arabinosa al 0,2 % e IPTG 1 mM. Las células se sembraron en placas en agar-LB que contenían estreptomycin, kanamicina y cloranfenicol (no selectivo para pUC- λ) o ampicilina, estreptomycin, kanamicina y cloranfenicol (selectivo para pUC- λ). Después del crecimiento durante la noche, el porcentaje de pérdida de plásmidos se puede calcular a partir de la relación de unidades formadoras de colonias en las placas selectivas y no selectivas.

Estudios de infección por fago Lambda

La sensibilidad del huésped a la infección por fago se ensayó utilizando un fago virulento Lambda (λ_{vir}), como en (Brouns y col (2008) *Science* 321, 960-964). La sensibilidad del huésped a la infección se calculó como la eficacia de la formación de placas (la relación de recuento de placas de una cepa que contiene un CRISPR anti- λ a la de la cepa que contiene un CRISPR R44 no dirigido) como se describe en Brouns y col (2008).

Ejemplo 1 - Cascada se une exclusivamente al ADN objetivo negativamente superenrollado

El plásmido derivado de pUC19 de 3 kb denotado pUC- λ , contiene un fragmento de ADN de 350 pb correspondiente a parte del gen J del fago λ , que está dirigido por J3-Cascada (Cascada asociada con ARNcr que contiene un espaciador J3 (Westra y col (2010) *Molecular Microbiology* 77, 1380-1393). Los ensayos de desplazamiento de movilidad electroforética muestran que Cascada tiene una alta afinidad solo por el plásmido objetivo negativamente superenrollado (nSE). En una relación molar de J3-Cascada a pUC- λ de 6:1, todo el plásmido nSE estaba unido por Cascada (véase la Fig. 1A), mientras Cascada que porta el ARNcr R44 no dirigido (R44-Cascada) mostró una unión no específica a una relación molar de 128:1 (véase la Fig. 1B). La constante de disociación (Kd) de pUC- λ nSE se determinó que era $13 \pm 1,4$ nM para J3-Cascada (véase la Fig. 1E) y 429 ± 152 nM para R44-Cascada (véase la Fig. 1F).

J3-Cascada no pudo unir el ADN objetivo relajado con afinidad medible, como pUC- λ escotado (véase la Fig. 1C) o lineal (véase la Fig. 1D), que muestra que Cascada tiene una gran afinidad por sustratos de ADN más grandes con una topología nSE.

Para distinguir la unión no específica de la unión específica, se utilizó el sitio de restricción BsmI localizado dentro del protoespaciador. Si bien la adición de la enzima BsmI a pUC- λ da un producto lineal en presencia de R44-Cascada (véase la Fig. 1G, carril 4), pUC- λ está protegido de la escisión de BsmI en presencia de J3-Cascada (véase la Fig. 1G, carril 7), que indica unión específica al protoespaciador. Esto muestra que Cas3 no se requiere para la unión específica de secuencia *in vitro* de Cascada a una secuencia de protoespaciador en un plásmido nSE.

La Cascada que se une a pUC- λ nSE fue seguido por escotamiento con Nt.BspQI, dando lugar a una topología CA. La Cascada se libera del plásmido después del escotamiento de la cadena, como puede verse por la ausencia de un desplazamiento de la movilidad (véase la Fig. 1H, comparar el carril 8 con el carril 10). Por el contrario, Cascada permanece unida a su objetivo de ADN cuando se añade una sonda de ADNmc complementario a la cadena desplazada con la reacción antes de la escisión del ADN por Nt.BspQI (véase la Fig. 1H, carril 9). La sonda estabiliza artificialmente el bucle R de Cascada en el ADN objetivo relajado. Se realizan observaciones similares cuando ambas cadenas de ADN de pUC- λ se escinden después de la unión a Cascada (véase la Fig. 1I, carril 8 y carril 9).

Ejemplo 2 - Cascada induce la flexión del ADN objetivo unido

Se visualizaron complejos formados entre Cascada purificada y pUC- λ . Se formaron complejos específicos que contienen un solo complejo de J3-Cascada unido, mientras que R44-Cascada inespecífico no produce complejos unidos a ADN en este ensayo en condiciones idénticas. De las 81 moléculas de ADN observadas, se descubrió que el 76 % tenían J3-Cascada unida (véase la Fig. 2A-P). De estos complejos, en la mayoría de los casos, se descubrió Cascada en el vértice de un bucle (86 %), mientras que solo se encontró una fracción pequeña en posiciones no apicales (14 %). Estos datos muestran que la unión de Cascada causa la flexión y posiblemente la envoltura del ADN, probablemente para facilitar la fusión local del dúplex de ADN.

Ejemplo 3: Fusiones naturales de Cas3 y Cse1: Cas3 interactúa con Cascada tras el reconocimiento del protoespaciador

La Figura S3 muestra el análisis de la secuencia de genes *cas* de organismos que contienen el tipo I-E. El sistema CRISPR/Cas revela que Cas3 y Cse1 aparecen como proteínas de fusión en *Streptomyces* sp. SPB78 (número de acceso: ZP_07272643.1), en *Streptomyces griseus* (número de acceso YP_001825054), y en *Catenulispota acidiphila* DSM 44928 (número de acceso YP_003114638).

Ejemplo 4: La complementación de fluorescencia bimolecular (BiFC) muestra cómo una proteína de fusión Cse1 que forma parte de Cascada continúa interactuando con Cas3.

Los experimentos de BiFC se utilizaron para controlar las interacciones entre Cas3 y Cascada *in vivo* antes y después de la infección por fago λ . Los experimentos de BiFC se basan en la capacidad de las mitades no fluorescentes de una proteína fluorescente, p. ej., la proteína amarilla fluorescente (PAF) para replegarse y formar una molécula fluorescente cuando las dos mitades se encuentran muy cerca. Como tal, proporciona una herramienta para revelar las interacciones proteína-proteína, ya que la eficacia del replegado aumenta mucho si las concentraciones locales son altas, p. ej., cuando las dos mitades de la proteína fluorescente se fusionan con los pares de interacción. Cse1 se fusionó en el extremo C terminal con los 155 aminoácidos N-terminales de Venus (Cse1-N155Venus), una versión mejorada de PAF (Nagai y col. (2002) *Nature Biotechnology* 20, 87-90). Cas3 se fusionó C-terminalmente con los 85 aminoácidos C-terminales de Venus (Cas3-C85Venus).

El análisis de BiFC revela que Cascada no interactúa con Cas3 en ausencia de ADN invasor (Fig. 3ABC, Fig. 3P y Fig. 8). Tras la infección por el fago λ , sin embargo, las células que expresan Cascada Δ Cse1, Cse1-N155Venus y Cas3-C85Venus son fluorescentes si coexpresan el anti- λ CRISPR 7Tm (Fig. 3DEF, Fig. 3P y Fig. 8). Cuando coexpresan un CRISPR R44 sin dirigir (Fig. 3GHI, Fig. 3P y Fig. 8), las células permanecen no fluorescentes. Esto muestra que Cascada y Cas3 interactúan específicamente durante la infección en el reconocimiento del protoespaciador y que Cse1 y Cas3 están muy cerca el uno del otro en el complejo efector binario Cascada-Cas3.

Estos resultados también muestran claramente que una fusión de Cse1 con una proteína heteróloga no interrumpe la formación ribonucleoproteica de Cascada y ARNcr, ni interrumpe la interacción de Cascada y Cas3 con el ADN del fago objetivo, incluso cuando el Cas3 mismo es también una proteína de fusión.

Ejemplo 5: La preparación de una fusión diseñada Cas3-Cse1 proporciona una proteína dentro de la actividad funcional *in vivo*

Proporcionar una evidencia *in vitro* para la actividad de escisión del ADN de Cas3 requirió Cas3 purificado y activo. A pesar de varias estrategias de solubilización, Cas3 se sobreprodujo (Howard y col (2011) *Biochem. J.* 439, 85-95) en BL21de *E. coli* está presente principalmente en agregados inactivos y cuerpos de inclusión. Por lo tanto, Cas3 se produjo como una proteína de fusión Cas3-Cse1, que contiene un enlazador idéntico al de la proteína de fusión

Cas3-Cse1 en *S. griseus* (véase la Fig. 10). Cuando se coexpresó con Cascada Δ Cse1 y CRISPR J3, el complejo de fusión era soluble y se obtuvo en alta pureza con la misma estequiometría aparente que Cascada (Fig. 5A). Cuando se ensayó la funcionalidad de este complejo para proporcionar resistencia contra la infección por fago λ , la eficacia de la formación de placas (efp) en las células que expresan el complejo de fusión J3-Cascada-Cas3 fue idéntica a las células que expresan las proteínas distintas (Fig. 5B).

Dado que el complejo de fusión J3-Cascada-Cas3 era funcional *in vivo*, los ensayos de escisión de ADN *in vitro* se llevaron a cabo utilizando este complejo. Cuando se incubó J3-Cascada-Cas3 con pUC- λ en ausencia de metales divalentes, se observó la unión del plásmido con relaciones molares similares a las observadas para Cascada (Fig. 5C), mientras que una unión específica a un plásmido no objetivo (pUC-p7, un plásmido derivado de pUC19 del mismo tamaño que pUC- λ , pero que carece de un protoespaciador) se produjo solo con altas relaciones molares (Fig. 5D), lo que indica que la unión específica al ADN del complejo es también similar a la de Cascada solo.

Curiosamente, el complejo de fusión J3-Cascada-Cas3 muestra actividad de endonucleasa dependiente de magnesio en plásmidos objetivo nSE. En presencia de Mg^{2+} 10 mM, J3-Cascada-Cas3 fragmenta pUC- λ nSE (Fig. 5E, carril 3-7), pero no se observa escisión para los sustratos que no contienen la secuencia objetivo (Fig. 5E, carril 9-13), o que tienen una topología relajada. No se observa desplazamiento de la banda CA resultante, en línea con las observaciones previas de que Cascada se disocia espontáneamente después de la escisión, sin requerir actividad de helicasa Cas3 dependiente de ATP. En cambio, la actividad helicasa de Cas3 parece estar implicada en la degradación del plásmido exonucleolítico. Cuando se añaden tanto magnesio como ATP a la reacción, se produce la degradación completa del plásmido (Fig. 5H).

Los inventores han descubierto que Cascada solo es incapaz de unirse a protoespaciadores en el ADN relajado. Por el contrario, los inventores han descubierto que Cascada localiza de forma eficiente los objetivos en ADN negativamente superenrollado, y posteriormente recluta Cas3 a través de la subunidad Cse1. La escisión endonucleolítica por el dominio de nucleasa HD de Cas3 causa la liberación espontánea de Cascada del ADN a través de la pérdida del superenrollamiento, lo que permite la reubicación de Cascada para localizar nuevos objetivos. El objetivo se desenrolla progresivamente y se escinde mediante la unión de actividad helicasa dependiente de ATP y la actividad de nucleasa HD de Cas3, lo que conduce a la degradación completa del ADN objetivo y a la neutralización del invasor.

Con referencia a la Figura 6 y sin desear quedar ligado a teoría particular alguna, un mecanismo de operación para la vía de interferencia CRISPR tipo I en *E. coli* puede implicar (1) en primer lugar, que Cascada que lleva un ARNcr barra el ADN plasmídico nSE para un protoespaciador, con MAP adyacente. Si durante esta etapa se produce la separación de cadena, esto se desconoce. (2) La unión del protoespaciador específico de la secuencia se logra mediante el emparejamiento de bases entre el ARNcr y la cadena complementaria del ADN, formando un bucle R. Al unirse, Cascada induce la flexión del ADN. (3) La subunidad Cse1 de Cascada recluta Cas3 al unirse al ADN. Esto se puede lograr mediante cambios conformacionales de Cascada que tienen lugar tras la unión del ácido nucleico. (4) El dominio HD (parte más oscura) de Cas3 cataliza el escotamiento dependiente de Mg^{2+} de la cadena desplazada del bucle R, alterando así la topología del plásmido objetivo de nSE a CA relajado. (5a y 5b) La relajación del plásmido provoca la disociación espontánea de Cascada. Mientras tanto, Cas3 muestra actividad de exonucleasa dependiente de ATP en el plásmido objetivo, requiriendo el dominio helicasa para el desenrollamiento de ADNbc objetivo y el dominio de nucleasa HD para la actividad de escisión sucesiva. (6) Cas3 degrada el plásmido completo de una manera dependiente de ATP a medida que avanza, se desenrolla y se escinde el ADNbc objetivo.

Ejemplo 6: Preparación de proteínas artificiales de fusión Cas-strep tag y ensamblaje de complejos Cascada

Los complejos Cascada se producen y purifican como se describe en Brouns y col (2008) *Science* 321: 960-4 (2008), utilizando los plásmidos de expresión enumerados en la Tabla suplementaria 3 de Jore y col (2011) *Nature Structural & Molecular Biology* 18: 529-537. Cascada se purifica rutinariamente con un Strep-tag II N-terminal fusionado con CasB (o CasC en CasCDE). La cromatografía de exclusión por tamaños (Superdex 200 HR 10/30 (GE)) se realiza utilizando Tris-HCl 20 mM (pH 8,0), NaCl 0,1 M, ditiotretol 1 mM. Las preparaciones de Cascada (~0,3 mg) se incuban con DNasa I (Invitrogen) en presencia de $MgCl_2$ 2,5 mM durante 15 min a 37 °C antes del análisis de exclusión por tamaños. Los ácidos nucleicos co-purificados se aíslan por extracción utilizando un volumen igual de fenol:cloroformo:alcohol isoamílico (25:24:1) pH 8,0 (Fluka), y se incuban con DNasa I (Invitrogen) suplementado con $MgCl_2$ 2,5 mM o RNasa A (Fermentas) durante 10 min a 37 °C. Se producen proteínas de la subunidad de Cas fusionadas a la secuencia de aminoácidos de Strep-Tag.

Los ensayos de placa que muestran la actividad biológica de las subunidades Strep-Tag Cascada se realizan utilizando el bacteriófago Lambda y la eficacia de la formación de placas (EFP) se calculó como se describe en Brouns y col (2008).

Para la purificación de ARNcr, las muestras se analizan mediante HPLC de fase inversa de par de iones en un Agilent 1100 HPLC con detector UV_{260nm} (Agilent) utilizando una columna DNasep de 50 mm x 4,6 mm I. D. (Transgenomic, San Jose, CA). El análisis cromatográfico se realiza utilizando las siguientes condiciones de tampón: A) acetato de trietilamonio 0,1 M (TEAA) (pH 7,0) (Fluka); B) tampón A con acetonitrilo de grado MS LC 25 % (v/v)

(Fisher). El ARNcr se obtiene inyectando Cascada intacta purificada a 75 °C utilizando un gradiente lineal que comienza con 15 % de tampón B y se extiende hasta 60 % B en 12,5 min, seguido de una extensión lineal hasta 100 % B durante 2 min a un caudal de 1,0 ml/min. La hidrólisis del extremo terminal fosfato cíclico se realizó incubando el ARNcr purificado por HPLC en una concentración final de HCl 0,1 M a 4 °C durante 1 hora. Las muestras se concentran a 5-10 µl en un concentrador de vacío (Eppendorf) antes del análisis ESI-MS.

El análisis de espectrometría de masas e ionización por electrospray de ARNcr se realiza en modo negativo utilizando un espectrómetro de masas UHR-TOF (maXis) o un instrumento HCT Ultra PTM Discovery (ambos de Bruker Daltonics), acoplado a un sistema de cromatografía líquida capilar en línea (Ultimate 3000, Dionex, R.U.). Las separaciones de ARN se realizan utilizando una columna capilar monolítica (PS-DVB) (200 µm x 50 mm I.D., Dionex, R.U.). La cromatografía se realiza utilizando las siguientes condiciones de tampón: C) 1,1,1,3,3,3-hexafluoro-2-propanol 0,4 M (HFIP, Sigma-Aldrich) ajustado con trietilamina (TEA) a pH 7,0 y TEAA 0,1 mM, y D) tampón C con 50 % de metanol (v/v) (Fisher). El análisis de ARN se realiza a 50 °C con 20 % de tampón D, extendiéndose a 40 % D en 5 min seguido de una extensión lineal a 60 % D durante 8 min a un caudal de 2 µl/min.

La proteína de Cascada se analiza por espectrometría de masas nativa en acetato de amonio 0,15 M (pH 8,0) a una concentración de proteína de 5 µM. La preparación de proteína se obtiene mediante cinco etapas secuenciales de concentración y dilución a 4 °C utilizando un filtro centrífugo con un límite de corte de 10 kDa (Millipore). Las proteínas se pulverizan desde los capilares de vidrio de borosilicato y se analizan en un tiempo de vuelo de electrospray LCT o instrumentos de tiempo de vuelo cuadrupolo modificados (ambos Waters, R.U.) ajustados para un rendimiento óptimo en la detección de alta masa (véase Tahallah N y col (2001) *Rapid Commun Mass Spectrom* 15: 596-601 (2001) y van den Heuvel, R.H. y col. *Anal Chem* 78: 7473-83 (2006). Las mediciones de masa exactas de las proteínas Cas individuales se adquirieron en condiciones desnaturizantes (50 % de acetonitrilo, 50 % de MQ, 0,1 % de ácido fórmico). Los subcomplejos en solución se generaron mediante la adición de 2-propanol a la solución de pulverización hasta una concentración final del 5 % (v/v). La configuración del instrumento fue la siguiente; voltaje de la aguja ~1,2 kV, voltaje del cono ~175 V, presión de la fuente 9 mbar. Se utilizó xenón como el gas de colisión para el análisis espectrométrico de masas en tándem a una presión de 1,510⁻² mbar. El voltaje de colisión varió entre 10-200 V.

Los ensayos de desplazamiento de movilidad electroforética (EDME) se utilizan para demostrar la actividad funcional de los complejos Cascada para ácidos nucleicos objetivo. EDME se realiza incubando Cascada, CasBCDE o CasCDE con ácido nucleico marcado 1 nM en Tris-Cl 50 mM, pH 7,5, NaCl 100 mM. Se utiliza ADN de esperma de salmón (Invitrogen) como competidor. Las reacciones de EDME se incuban a 37 °C durante 20-30 minutos antes de la electroforesis en geles de poliacrilamida al 5 %. Los geles se secan y se analizan utilizando pantallas de almacenamiento de fósforo y un generador de imágenes de fósforo PMI (Bio-Rad). La actividad de escisión y unión al ADN objetivo de Cascada se ensaya en presencia de Ca 1-10 mM, iones Mg o Mn.

Los objetivos de ADN son oligonucleótidos largos purificados en gel (Isogen Life Sciences o Biolegio), enumerados en la Tabla suplementaria 3 de Jore y col (2011). Los oligonucleótidos se marcan en el extremo utilizando γ³²P-ATP (PerkinElmer) y T4 quinasa (Fermentas). Los objetivos de ADN bicatenario se preparan hibridando oligonucleótidos complementarios y digiriendo ADNmc restante con exonucleasa I (Fermentas). Los objetivos de ARN marcados se transcriben *in vitro* utilizando kits T7 Maxiscript o T7 Mega Shortscript (Ambion) con α³²P-CTP (PerkinElmer) y eliminando el molde por digestión con DNasa I (Fermentas). Los objetivos de ARN bicatenario se preparan hibridando ARNs complementarios y digiriendo ARNmc excedente con RNasa T1 (Fermentas), seguido de extracción con fenol.

Los ensayos de desplazamiento de movilidad del plásmido se realizan utilizando el plásmido pWUR613 que contiene el protoespaciador R44. El fragmento que contiene el protoespaciador se amplifica por PCR a partir del ADN genómico del bacteriófago P7 utilizando los cebadores BG3297 y BG 3298 (véase la Tabla 3 complementaria de Jore y col. (2011). El plásmido (0,4 µg) y Cascada se mezclaron en una relación molar de 1:10 en un tampón que contenía Tris-HCl 5 mM (pH 7,5) y NaCl 20 mM y se incubaron a 37 °C durante 30 minutos. Las proteínas de Cascada se eliminaron mediante tratamiento con proteinasa K (Fluka) (0,15 U, 15 min, 37 °C) seguido de extracción con fenol/cloroformo. Los complejos de ARN-ADN se trataron posteriormente con RNasaH (Promega) (2 U, 1 h, 37 °C).

Las fusiones de la subunidad proteica Strep-Tag-Cas que forman complejos proteicos Cascada o subcomplejos activos con el componente de ARN (equivalente a un ARNcr) tienen la actividad biológica y funcional esperada de barrido y unión específica y escisión de objetivos de ácido nucleico. Las fusiones de las subunidades Cas con las cadenas de aminoácidos de colorantes fluorescentes también forman complejos Cascada y subcomplejos con el componente ARN (equivalente a ARNcr) que retienen la actividad biológica y funcional y permiten la visualización de la ubicación de una secuencia de ácido nucleico objetivo en ADNbc por ejemplo.

Ejemplo 7 - Un par de Cascada-nucleasa y ensayo de actividad de nucleasa *in vitro*

Se han introducido seis mutaciones denominadas "Sharkey" por mutagénesis aleatoria e identificación sistemática para mejorar la actividad de la nucleasa y la estabilidad del dominio de nucleasa no específico de la enzima de restricción FokI de *Flavobacterium okeanoikoites* (véase Guo, J., y col. 2010) *J. Mol. Biol.* 400: 96-107). Se han

introducido otras mutaciones que reducen la actividad de escisión fuera del objetivo. Esto se logra mediante ingeniería de interacciones electrostáticas en la interfaz del dímero FokI de un par ZFN, creando una variante FokI con una interfaz cargada positivamente (KKR, E490K, I538K, H537R) y otra con una interfaz cargada negativamente (ELD, Q486E, I499L, N496D) (véase Doyon, Y., y col. (2011) *Nature Methods* 8: 74-9). Cada una de estas variantes es catalíticamente inactiva como homodímero, reduciendo así la frecuencia de escisión fuera del objetivo.

Diseño de la Cascada-nucleasa

Se ha fusionado de forma traduccional nucleasas FokI mejoradas al extremo N-terminal de Cse1 para generar variantes de Cse1 que son FokI^{KKR}-Cse1 y FokI^{ELD}-Cse1, respectivamente. Estas dos variantes se coexpresan con las subunidades de Cascada (Cse2, Cas7, Cas5 y Cas6e) y uno de dos plásmidos CRISPR distintos con espaciadores uniformes. Esto carga el complejo Cascada^{KKR} con P7-ARNcr uniforme y el complejo Cascada^{ELD} con M13 g8-ARNcr uniforme. Estos complejos se purifican utilizando Cse2 marcada con StreptII N-terminalmente como se describe en Jore, M.M., y col., (2011) *Nat. Struct. Mol. Biol.* 18(5): 529-536. Además, se puede llevar a cabo una etapa de purificación adicional utilizando una FokI marcada con HIS en el extremo N-terminal, para asegurar la purificación de los complejos de fusión Cascada-nucleasa completos e intactos.

Las secuencias de nucleótidos y aminoácidos de las proteínas de fusión utilizadas en este ejemplo fueron las siguientes:

>secuencia de nucleótidos de FokI-(Sharkey-ELD)-Cse1

ATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAAAAGTGAAGTGCGCCAC
 AAAGTGAATATGTGCCGCATGAATATATCGAGCTGATTGAAATTGCACGTAATC
 CGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGAATTTTTTATGAAAGTGTA
 CGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAAACCGGATGGTGCAATTTAT
 ACCGTTGGTAGCCCGATTGATTATGGTGTTATTGTTGATACCAAAGCCTATAGCG
 GTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAATGGAACGTTATGTGGAAG
 AAAATCAGACCCGTGATAAACATCTGAATCCGAATGAATGGTGGAAAGTTTATC
 CGAGCAGCGTTACCGAGTTTAAATTCTGTTTGTAGCGGTCACCTCAAAGGCAA
 CTATAAAGCACAGCTGACCCGTCTGAATCATATTACCAATTGTAATGGTGCAGTT
 CTGAGCGTTGAAGAACTGCTGATTGGTGGTGAAATGATTAAAGCAGGCACCCTG
 ACCCTGGAAGAAGTTCGTCGCAAATTTAACAATGGCGAAATCAACTTTGCGGAT
CCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAGCGTGGCGAGCatgaatttgct
 tattgataactggattcctgtacgcccgcgaaacggggggaaagtccaaatcataaatctgcaatcgctatactgcagtagagatcagt
 ggcgattaagtttccccgtgacgatatggaactggcgcgttagcactgctggttgcaattgggcaaattatcgccccggcaaaagatg
 acgttgaatttcgacatcgcataatgaatccgctcactgaagatgagtttcaacaactcatcgccgctggatagatatgttctaccta
 cacgcagaacatccctttatgcagaccaaaggtgtcaagcaaatgatgtgactccaatggaaaaactgttggtggggaagcggcg
 cgacgaattgtcatttgtcaatcaaccggggcaggggtgaagcattatgtggtggatgcactgcgattgcgttattcaaccaggcgaat
 caggcaccaggttttggtggtggttttaaagcgggttacgtggaggaacacctgtaacaacgttcgtacgtgggatcgatcttcgtcaa
 cgggtgtactcaatgtcctcacattacctgcttcaaaaacaatttctaataatcacatacggaacaaacctacctggattaaacct
 atcaagtccaatgagtcataacctgctcgtcaattgggttgcctggtctattctggcaaccagcgcatattgaattatgcgatccatt
 gggattggtaaatgttctgtgtgacaggaagcaatttgcgttataccggttttctaaggaaaaatttacctttacagttaatgggctat
 ggcccatccgcattccccctgtctggttaacagtcagaaaggggaggttgaggaaaaatttctgttccaccacctccgcaccatcat
 ggacacaaatcagccgagttgtggtagataagattattcaaatgaaatggaaatcgctggcggcggttgatcaattcagaat
 attgcgcgcaaaagtcctctgaattgattatggggggatcgtataaatcaagcatctattctgaacggcgcatgatgtgtgatgtt
 aatcaggggtggcaacaatcggcaatgtgataaacgaaatagtactgttgggttggatataaacagccttacgcaaggcgttata
 tacctttgcagaaggggttaaaaataaagactcaaaagggcgagtcctgttcagactgcagaaaggcatttctatcgacagag
 tgaattattaattcccgatgtactggcgaatgttaattttccaggctgatgagtaataagctgatttacgagacaaactcatcaattgtgt
 gaaatgctatttaataatctgtagtcctatgcacatcatcctaataaataagcacattagcgcttgcgcgcccagcgtatacaaa
 ttacgggagttaaaaccgcaaggaggccatcaaatggctga [SEQ ID NO: 18]

>secuencia proteica de FokI-(Sharkey-ELD)-Cse1

MAQLVKSELEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGY
RGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTR
DKHLNPNEWVKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEEL
LIGGEMIKAGTLTLEEVRKFNNGEINFADPTNRAKGLEAVSVASMNLLIDNWIPVRP
RNGGKVQIINLQSLYCSRQWRLSLPRDDMELAAALLVCIGQIIAPAKDDVEFRHRI
MNPLTEDEFQQLIAPWIDMFYLNHAEHPFMQTKGVKANDVTPMEKLLAGVSGATN
CAFVNQPGQGEALCGGCTAIALFNQANQAPGFGGGFKSGLRGGTPVTTFVRGIDLRS
TVLLNVLTLPRLQKQFPNESHTENQPTWIKPIKSNEIPASSIGFVRGLFWQPAHIELC
DPIGIGKSCCGQESNLRYTGFLKEKFTFTVNLWPHPHSPCLVTVKKGEVEEKFLAF
TTSAPSWTQISRVVVDKIIQNENGNRVAAVVNQFRNIAPQSLELIMGGYRNNQASIL
ERRHDVLMFNQGWQQYGNVINEIVTVGLGYKTALRKALYTFAEGFKNKDFKGAGV
SVHETAERHFYRQSELLIPDVLANVNFSQADEVIADLRDKLHQLCEMLFNQSVAPYA
HHPKLISTLALARATLYKHLRELKPQGGPSNG*[SEQ ID NO: 19]

>secuencia de nucleótidos de FokI-(Sharkey-KKR)-Cse1

ATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAAAAGTGAAGTGCGCCAC
AAACTGAAATATGTGCCGCATGAATATATCGAGCTGATTGAAATTGCACGTAATC
CGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGAATTTTTTATGAAAGTGTA
CGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAAACCGGATGGTGCAATTTAT
ACCGTTGGTAGCCCGATTGATTATGGTGTTATTGTTGATACCAAAGCCTATAGCG
GTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAATGCAGCGTTATGTGAAAG
AAAATCAGACCCGCAACAAACATATTAACCCGAATGAATGGTGGAAGTTTATC
CGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTAGCGGTCACCTCAAAGGCAA
CTATAAAGCACAGCTGACCCGTCTGAATCGTAAAACCAATTGTAATGGTGACGTT
CTGAGCGTTGAAGAACTGCTGATTGGTGGTGAAATGATTAAAGCAGGCACCCTG
ACCCTGGAAGAAGTTCGTCGCAAATTTAACAATGGCGAAATCAACTTTGCGGAT
CCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAGCGTGGCGAGC atgaatttgct
tattgataactggattcctgtacgcccgcgaaacgggggaaagtccaaatcataaatctgcaatcgctatactgcagtagagatcagt
ggcgattaagtttccccgtgacgatatggaactggccgcttagcactgctggttgattgggcaaattatcgccccggcaaaagatg
acgttgaatttcgacatcgcataatgaatccgctcactgaagatgagtttcaacaactcatcgccgctggatagatatgttctacctaat
cacgcagaacatccctttatgcagaccaaaggtgtcaaagcaaagatgtgactccaatggaaaaactgttgctggggtaagcggcg
cgacgaattgtgcatttgtaataaccggggcagggtgaagcattatgtggtggatgcactgcgattgcgttattcaaccaggcgaat

caggcaccagggttttggtggtggttttaaaagcggtttacgtggaggaacacctgtaacaacgttcgtacgtgggatcgatcttcgttcaa
cggtgttactcaatgtcctcacattacctcgtcttcaaaaacaatttctaataatcacatacggaaaaccaacctacgttgattaaacct
atcaagtccaatgagtcatactgcttcgtcaattgggtttgtccgtggtctattctggcaaccagcgcatattgaattatgcgatccatt
gggattggtaaatgttcttctgtgtggacaggaaagcaatttgctgtataccgggttttctaaggaaaaatttacctttacagttaatgggctat
ggcccatccgcattccccttctgtgtgtaacagtaagaaaggggagggttgaggaaaaatttcttgccttcaccacctccgcaccatcat
ggacacaaatcagccgagttgtgtagataagattattcaaaatgaaatggaaatcgctggcgcggttgatcaattcagaaat
attgcgcgcgcaaagtctcttgaattgattatggggggatcgtataatcaagcatctattctgaacggcgctcatgatgtgtgatgtt
aatcaggggtggcaacaatcggcaatgtgataaacgaaatagtactgttggttgggatataaacagccttacgcaaggcggtata
taccttgcagaagggttataaaataaagacttcaaagggggcggagtcctgttcatgagactgcagaaggcatttctatcgacagag
tgaattattaattccgatgtactggcgaatgttaattttccaggctgatgaggaatagctgatttacgagacaaactcatcaattgtgt
gaaatgctatttaataatctgtactccctatgcacatcatcctaataataagcacattagcgcttgcccgcgccacgctatacaaaaca
tttacgggagttaaaaccgcaaggagggccatcaaatggctga [SEQ ID NO: 20]

>secuencia proteica de FokI-(Sharkey-KKR)-Cse1

MAQLVKSELEKKSELRHKLKYPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGY
RGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQT
RNKHINPNEWWKVYPSSVTEFKFLVSGHFKGNYKAQLTRLNRKTNCGAVLSVEE
LLIGGEMIKAGTLTLEEVRRKFNNGEINFADPTNRAKGLEAVSVASMNLLIDNWIPV
RPRNGGKVQIINLQSLYCSRQWRLSLPRDDMELALALLVCIGQIIAPAKDDVEFR
HRIMNPLTEDEFQQLIAPWIDMFYLNHAEHPFMQTKGVKANDVTPMEKLLAGVSGA
TNCAFVNQPGQGEALCGGCTAIALFNQANQAPGFGGFKSGLRGGTPVTTFVRGIDL
RSTVLLNVLTLPRLQKQFPNESHTENQPTWIKPIKSNEIPASSIGFVRGLFWQPAHIEL
CDPIGIGKSCCGQESNLRYTGFLKEKFTFTVNGLWPHPHSPCLVTVKKGEVEEKFL
AFTTSAPSWTQISRVVVDKIIQENGNRVAADVNNQFRNIAPQSPLELIMGGYRNNQA
SILERRHDVLMFNQGWQQYGNVINEIVTVGLGYKTALRKALYTFAEGFKNKDFKGA
GVSVHETAERHFYRQSELLIPDVLANVNFSQADEVIADLRDKLHQLCEMLFNQSVAP
YAHHPKLISLALARATLYKHLRELKPQGGPSNG* [SEQ ID NO: 21]

>secuencia de nucleótidos de His₆-SLN monopartita dual SV40-FokI-(Sharkey-KKR)-Cse1

ATGcatcaccatcatcaccacCCGAAAAAAAAAGCGCAAAGTGGATCCGAAGAAAAAACGTAAAG
TTGAAGATCCGAAGACATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAA

AAAGTGAAGTGCGCCACAACTGAAATATGTGCCGCATGAATATATCGAGCTGA
 TTGAAATTGCACGTAATCCGACCCAGGATCGTATTCTGGAAATGAAAGTGATGG
 AATTTTTTATGAAAGTGTACGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAA
 ACCGGATGGTGCAATTTATACCGTTGGTAGCCCGATTGATTATGGTGTATTGTT
 GATACCAAAGCCTATAGCGGTGGTTATAATCTGCCGATTGGTCAGGCAGATGAA
 ATGCAGCGTTATGTGAAAGAAAATCAGACCCGCAACAAACATATTAACCCGAAT
 GAATGGTGGAAAGTTTATCCGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTTA
 GCGGTCACTTCAAAGGCAACTATAAAGCACAGCTGACCCGTCTGAATCGTAAAA
 CCAATTGTAATGGTGCAGTTCTGAGCGTTGAAGAACTGCTGATTGGTGGTGAAT
 GATTAAAGCAGGCACCCTGACCCTGGAAGAAGTTCGTCGCAAATTTAACAATGG
 CGAAATCAACTTTGCGGATCCCAACCAACCGCGCGAAAGGCCTGGAAGCGGTG
AGCGTGGCGAGCatgaatttgcttattgataactggattcctgtacgcccgcaaacgggggaaagtccaaatcataaat
 ctgcaatcgctatactgcagtagagatcagtgccgattaagtgtgccccgtgacgatatggaactggcgcttagcactgctggttgc
 attgggcaaatatcgccccggcaaaagatgacgttgaaattcgacatcgcataatgaatccgctcactgaagatgagtttcaacaactc
 atcgcgccgtggatagatatgttctacctaatacagcagaacatccctttatgcagaccaaagggtgcaagcaaatgatgtgactcca
 atggaaaaactgttggtgggtaagcggcgacgaattgtcattgtcaatcaaccggggcagggtgaagcattatgtggtggat
 gcactgcgattgcgttattcaaccaggcgaatcaggcaccagggtttggtggtggttttaaagcggtttacgtggaggaacacctgtaa
 caacgttcgtacgtgggatcgatcttcgttcaacgggtgttactcaatgtcctcacattacctcgtcttcaaaaacaatttctaataatcac
 atacggaaaaccaacctacgtgattaaacctatcaagtccaatgagtcatacctgcttcgtcaattgggtttgtccgtggtctattctgg
 caaccagcgcatattgaattatgcgatccattgggattggttaattgtctgtgacaggaaagcaatttgcgttataccggtttctt
 aaggaaaaatttacctttacagttaatgggctatggcccatccgcattcccttctgtgtaacagtcaagaaaggggaggttgaggaa
 aaatttctgtttcaccacctccgcacatcatggacacaaatcagccgagttgtggtagataagattattcaaaatgaaatggaaatc
 gcgtggcgggcggtgtgaatcaattcagaatattgcgcgcgaaagtcccttgaattgattatgggggatatcgtaataatcaagcat
 ctattctgaacggcgatgatgtgtgtgtttaatcagggtggcaacaatacggcaatgtgataaacgaaatagtactgttggtttg
 ggatataaacagccttacgcaaggcggttatatacctttgcagaagggtttaaaataaagacttcaaggggcccggagtctctgttcat
 gagactgcagaaaggcatttctatcgacagagtgaattattaattcccgatgtactggcgaatgttaattttccaggctgatgaggtaat
 agctgatttacgagacaaactcatcaattgtgtgaaatgctatttaataatctgtagctccctatgcacatcactctaaattaataagcaca
 ttgagcgttggccgcgcacgctatacaaacatttacgggaggttaaaaccgcaaggaggggccatcaaatggctga [SEQ ID
 NO: 22]

>secuencia proteica de His₆-SLN monopartita dual SV40-FokI-(Sharkey-KKR)-Cse1

MHHHHHHPKKKRKVDPKKKRKVEDPKDMAQLVKSELEEKKSELRHKLKYVPHEYI
 ELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVD
 TKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGH
 FKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFA
 DPTNRAKGLEAVSVASMNLLIDNWIPVRPRNGGKVQIINLQSLYCSRQDQWRLSLPRD
 DMELAALALLVCIGQIIAPAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHP
 FMQTKGVKANDVTPMEKLLAGVSGATNCAFNQPGQGEALCGGCTAIALFNQANQ
 APGFGGGFKSGLRGGTPVTTFVRGIDLRSTVLLNVLTLPRLQKQFPNESHTENQPTWI
 KPIKSNEIPASSIGFVRGLFWQPAHIELCDPIGIGKCSCCGQESNLRYTGFLKEKFTFT
 VNGLWPHPHSPCLVTVKKGEVEEKFLAFTTSAPSWTQISRVVVDKIIQENGNRVA
 VVNQFRNIAPQSPLELIMGGYRNNQASILERRHDVLMFNQGWQQYGNVINEIVTVGL
 GYKTALRKALYTFAEGFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNFSQ
 ADEVIADLRDKLHQLCEMLFNQSVAPYAHHPKLISTLALARATLYKHLRELKPQGGP
 SNG* [SEQ ID NO: 23]

>secuencia de nucleótidos de His₆-SLN monopartita dual SV40-FokI (Sharkey-ELD)-CseI

ATGcatcaccatcatcaccacCCGAAAAAAAAAGCGCAAAGTGGATCCGAAAGAAAAACGTAAAG
 TTGAAGATCCGAAAGACATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAAA
 AAGTGAAGTGCGCCACAACTGAAATATGTGCCGCATGAATATATCGAGCTGAT
 TGAAATTGCACGTAATCCGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGA
 ATTTTTTATGAAAGTGTACGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAAA
 CCGGATGGTGCAATTTATACCGTTGGTAGCCCGATTGATTATGGTGTTATTGTTG
 ATACCAAAGCCTATAGCGGTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAA
 TGGAACGTTATGTGGAAGAAAATCAGACCCGTGATAAACATCTGAATCCGAATG
 AATGGTGGAAGTTTATCCGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTAG
 CGGTCACCTCAAAGGCAACTATAAAGCACAGCTGACCCGTCTGAATCATATTACC
 AATTGTAATGGTGCAGTTCTGAGCGTTGAAGAACTGCTGATTGGTGGTGAAATGA
 TTAAGCAGGCACCCTGACCCTGGAAGAAGTTCGTGCGAAATTTAACAATGGCG
 AAATCAACTTTGCGGATCCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAG
CGTGGCGAGCatgaatttgcttattgataactggattcctgtacgcccgcgaaacggggggaaagtccaaatcataatctg
 caatcgctatactgcagtagagatcagtggcgattaagtttgccccgtgacgatatggaactggccgcttagcactgctggttgattg
 ggcaaatatcgccccggcaaaagatgacgttgaatttcgacatcgcataatgaatccgctcactgaagatgagttcaacaactcatcg

cgccgtggatagatatgttctaccttaatcacgcagaacatcccttatgcagaccaaaggtgtcaaagcaaagatgtgactccaatgg
 aaaaactgttggctgggtaagcggcgacgaattgtgcattgtcaatcaaccggggcagggtgaagcattatgtgggtgacac
 tgcgattgcgttattcaaccaggcgaatcaggcaccagggttttggtggtggttttaaaagcggtttacgtggaggaacacctgtaaac
 gttcgtacgtgggatcgatcttcgttaacgggtgttactcaatgtcctcacattacctcgtcttcaaaaacaatttctaataatcacatac
 gaaaaccaacctacctggtattaaacctatcaagtccaatgagtcctatactgcttcgtcaattgggtttgtccgtggtctattctggcaacc
 agcgcatattgaattatgcgatccattgggattggtaaatgttctgtgtggacaggaaagcaatttgcgttataccgggtttcttaagga
 aaaatttacctttacagttaatgggctatggcccatccgcattcccttgcgtgtaacagtcagaaaggggaggttgaggaaaaattt
 cttgctttcaccacctccgcaccatcatggacacaaatcagccgagttgtggttagataagattattcaaaatgaaaatggaaatcgcggtg
 gcggcggtgtgaatcaattcagaaatattgcgccgaaagtccttgaattgattatggggggatcgtataatcaagcatctattct
 tgaacggcgctcatgatgtgtgatgtttaatcaggggtggcaacaatacggcaatgtgataaacgaaatagtactgttggttgggat
 aaaacagccttacgcaaggcggttatatactttgcagaagggtttaaaaaataagacttcaaagggggccggagtctctgttcagact
 gcagaaaggcatttctatcgacagagtgaattattaattcccgatgtactggcgaatgttaattttccaggctgatgaggaatagctga
 ttacgagacaaactcatcaattgtgtgaaatgctatttaataatctgtagctccctatgcacatcatcctaaattaataagcacattagcg
 cttgcccgcgccacgtatatacaaacatttacgggagttaaaaccgcaaggaggggccatcaaatggctga [SEQ ID NO: 24]

>secuencia proteica de His₆-SLN monopartita dual SV40-FokI-(Sharkey-ELD)-Cse1

MHHHHHHHPKKKRKVDPKKKRKVEDPKDMAQLVKSELEEKKSELRHKLKYVPHEYI
 ELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVD
 TKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWVKVYPSSVTEFKFLFVSGH
 FKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNGEINFA
 DPTNRAKGLEAVSVASMNLLIDNWIPVRPRNGGKVQIINLQSLYCSRQWRLSLPRD
 DMELAALALLVCIGQIIAPAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHP
 FMQTKGVKANDVTPMEKLLAGVSGATNCAFVNQPGQGEALCGGCTAIALFNQANQ
 APGFGGGFKSGLRGGTPVTTFFVRGIDLRSTVLLNVLTLPRLQKQFPNESHTENQPTWI
 KPIKSNEIPASSIGFVRGLFWQPAHIELCDPIGIGKSCCGQESNLRYTGFLKEKFTFT
 VNGLWPHPHSPCLVTVKKGEVEEKFLAFTTSAPSWTQISRVVVDKIIQNENGNRVAA
 VVNQFRNIAPQSPELIMGGYRNNQASILERHDVLMFNQGWQQYGNVINEIVTVGL
 GYKTALRKALYTFAEGFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNFSQ
 ADEVIADLRDKLHQLCEMLFNQSVAPYAHHPKLISLTLARATLYKHLRELKPQGGP
 SNG* [SEQ ID NO: 25]

Ensayo de escisión de ADN

- 5 La especificidad y la actividad de los complejos se ensayaron utilizando un plásmido objetivo artificialmente construido como sustrato. Este plásmido contiene sitios de unión M13 y P7 en cadenas opuestas de manera que ambos dominios FokI se enfrentan entre sí (véase la Figura 11). La distancia entre los sitios de unión de Cascada varía entre 25 y 50 pares de bases con incrementos de 5 pb. Como los sitios de unión de Cascada necesitan estar flanqueados por cualquiera de las cuatro secuencias de MAP conocidas (5'-protoespaciador-CTT/CAT/CTC/CCT-3'
- 10 este intervalo de distancia proporciona suficiente flexibilidad para diseñar dicho par para casi cualquier secuencia dada.

Las secuencias de los plásmidos objetivo utilizadas son las siguientes. El número indicó la distancia entre los sitios objetivo M13 y P7. Los protoespaciadores se muestran en negrita, MAPs están subrayados:

Las secuencias de los plásmidos objetivo. El número indica la distancia entre los sitios objetivo M13 y P7. (Protoespaciadores en negrita, MAPs subrayados).

5 >50 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTGCTTTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGTTCCGA
AGCTGTCTTTTCGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 26]

>45 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGTTCAAGCTG
TCTTTTCGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 27]

>40 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGAAGCTGTCTTT
CGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 28]

10

>35 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTTCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTAAGCTGTCTTTTCGCT
GCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 29]

>30 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTGCTAGCTCTAGAACTAGTCCTCAGCCTAGGAAGCTGTCTTTTCGCTGCTGA
GGGTGACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 30]

15

>25 pb

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG
TCTTCTCTAGAACTAGTCCTCAGCCTAGGAAGCTGTCTTTTCGCTGCTGAGGGTG
ACGATCCCGCATAGGCGGCCTTTAACTCg
 gatcc [SEQ ID NO: 31]

20

La escisión de los plásmidos objetivo se analizó en geles de agarosa, cuando el plásmido negativamente superenrollado (nSE) se puede distinguir del plásmido linealizado o escotado. El sitio de escisión del par de Cascada^{KKR/ELD} en un vector objetivo se determinó aislando los productos de escisión lineal de un gel de agarosa y rellenando los extremos 3' rebajados dejados por la escisión de FokI con el fragmento Klenow de ADN polimerasa de *E. coli* para crear extremos romos. El vector lineal se autoligó, transformó, amplificó, aisló y secuenció. El relleno de los extremos 3' rebajados y el religamiento conducirán a nucleótidos adicionales en la secuencia que representa el saliente dejado por la escisión de FokI. Al alinear las lecturas de secuencia con la secuencia original, los sitios de escisión pueden encontrarse en un nivel clonal y mapearse. A continuación, las bases adicionales incorporadas en

la secuencia después del relleno en los extremos 3' rebajados dejados por la escisión con FokI están subrayadas:

Escisión plegada

5' CTTGCGCTAGCTCTAGAA  CTAGTCCTCAGCCTAGGCCTAAG 3'
3' GAACGCGATCGAGATCTTGATC AGGAGTCGGATCCGGATTC 5'

5 **3' relleno en ligamiento**

5' CTTGCGCTAGCTCTAGAACTAG - CTAGTCCTCAGCCTAGGCCTAAG 3'
3' GAACGCGATCGAGATCTTGATC - GATCAGGAGTCGGATCCGGATTC 5'

Leyendo de arriba a abajo, las secuencias 5'- 3' anteriores son SEQ ID NOs: 32-35, respectivamente.

Escisión de un locus objetivo en células humanas

- 10 El gen CCR5 humano codifica la proteína del receptor 5 de quimiocina C-C, que sirve como el receptor para el virus de la inmunodeficiencia humana (VIH) en la superficie de los glóbulos blancos. El gen *CCR5* se dirige utilizando un par de nucleasas Cascada^{KKR/ELD} además de un locus artificial de PVF. Un par de sitios de unión adecuados se selecciona en la región de codificación de *CCR5*. Dos matrices CRISPR distintas que contienen espaciadores uniformes dirigidos a cada uno de los sitios de unión se construyen utilizando síntesis de ADN (Geneart).
- 15 La selección del gen objetivo CCR5 humano y los diseños CRISPR utilizados son los siguientes:

>Parte de la secuencia genómica humana CCR5, que contiene MLA completo (posición 347-1446).

GGTGGAAACAAGATGGATTATCAAGTGTCAAGTCCAATCTATGACATCAATTATTA
TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCAAATCGCAGCCCGCCTCCT
GCCTCCGCTCTACTCACTGGTGTTCATCTTTGGTTTTGTGGGCAACATGCTGGTCA
TCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCT
CAACCTGGCCATCTCTGACCTGTTTTTCCTTCTTACTGTCCCCTTCTGGGCTCACT
ATGCTGCCGCCCAGTGGGACTTTGGAAATACAATGTGTCAACTCTTGACAGGGCT

CTATTTTATAGGCTTCTTCTCTGGAATCTTCTTCATCATCCTCCTGACAATCGATA
GGTACCTGGCTGTCGTCCATGCTGTGTTTGCTTTAAAAGCCAGGACGGTCACCTT
TGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCA
 GGAATCATCTTTACCAGATCTCAAAAAGAAGGTCTTCATTACACCTGCAGCTCTC
 ATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGT
 CATCTTGGGGCTGGTCCTGCCGCTGCTTGTTCATGGTCATCTGCTACTCGGGAATC
 CTA AAAACTCTGCTTCGGTGTGAAATGAGAAGAAGAGGCACAGGGCTGTGAGG
 CTTATCTTCACCATCATGATTGTTTATTTTCTCTTCTGGGCTCCCTACAACATTGTC
 CTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATTGCAGTAGCTCTA
 ACAGGTTGGACCAAGCTATGCAGGTGACAGAGACTCTTGGGATGACGCACTGCT
 GCATCAACCCCATCATCTATGCCTTTGTGCGGGGAGAAGTTCAGAAACTACCTCTT
 AGTCTTCTTCCAAAAGCACATTGCCAAACGCTTCTGCAAATGCTGTTCTATTTTCC
 AGCAAGAGGCTCCCGAGCGAGCAAGCTCAGTTTACACCCGATCCACTGGGGAGC
 AGGAAATATCTGTGGGCTTGTGACACGGACTCAAGTGGGCTGGTGACCCAGTC

[SEQ ID NO: 36]

Red1/2: sitios objetivo elegidos (distancia: 34 pb, MAP 5'-CTT-3'). "Red 1 aparece primero en la secuencia subrayada en el cuadro anterior. Red2 es la segunda secuencia subrayada.

>Matriz CRISPR red1 (cursiva = espaciadores, negrita = repeticiones)

ccatggTAATACGACTCACTATAGGGAGAATTAGCTGATCTTTAATAATAAGGAAAT
 GTTACATTAAGGTTGGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAAAT
 GTATACTTTT**AGAGAGTTCCCCGCGCCAGCGGGGATAAAACCGCAAACACAGCA**
TGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGGGGATAAAACCGCAA
CACAGCATGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGGGGATAAA
CCGCAAACACAGCATGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGG
GGATAAAACCGAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTTAACC
 CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGggtacc [SEQ ID NO: 37]

>Matriz CRISPR red2 (cursiva: espaciadores, negrita: repeticiones)

ccatggTAATACGACTCACTATAGGGAGAATTAGCTGATCTTTAATAATAAGGAAAT
 GTTACATTAAGGTTGGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAAAT
 GTATACTTTT**AGAGAGTTCCCCGCGCCAGCGGGGATAAAACCGTGTGATCACTTG**
GGTGGTGGCTGTGTTTGCGTGAGTTCCCCGCGCCAGCGGGGATAAAACCGTGTGA
TCACTTGGGTGGTGGCTGTGTTTGCGTGAGTTCCCCGCGCCAGCGGGGATAAAC
CGTGTGATCACTTGGGTGGTGGCTGTGTTTGCGTGAGTTCCCCGCGCCAGCGGGG
ATAAACCGAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTTAACCCC
 TTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGggtacc [SEQ ID NO: 38]

Administración de Cascada^{KKR/ELD} en el núcleo de las células humanas

Cascada es muy estable como complejo proteína-ARN de múltiples subunidades y se produce fácilmente en cantidades de mg en *E. coli*. La transfección o microinyección del complejo en su forma intacta como purificado a partir de *E. coli* se utiliza como procedimientos de administración (véase la Figura 12). Como se muestra en la figura 12, las nucleasas Cascada-FokI se purifican a partir de *E. coli* y se encapsulan en vesículas de transfección de proteínas. Estas se fusionan luego con la membrana celular de células HepG2 humanas liberando las nucleasas en el citoplasma (etapa 2). Las secuencias de SLN luego se reconocen por las proteínas importinas, que facilitan el paso de los nucleoporos (etapa 3). Cascada^{KKR} (rectángulo abierto) y Cascada^{ELD} (rectángulo lleno) encontrarán y escindirán su sitio objetivo (etapa 4), induciendo vías de reparación del ADN que alterarán el sitio objetivo y llevarán a los cambios deseados. Las nucleasas Cascada^{KKR/ELD} necesitan actuar solo una vez y no requieren presencia permanente en la célula codificada en el ADN.

Para administrar Cascada en células humanas, los reactivos de transfección de proteínas se utilizan de varias fuentes, incluyendo Pierce, NEB, Fermentas y Clontech. Estos reactivos se han desarrollado recientemente para la administración de anticuerpos, y son útiles para transfectar un amplio intervalo de estirpes celulares humanas con eficiencias de hasta el 90 %. Las células HepG2 humanas se transfectan. Además, se transfectan otras estirpes celulares que incluyen CHO-K1, COS-7, HeLa y células madre no embrionarias.

Para importar el par de nucleasas Cascada^{KKR/ELD} en el núcleo, se fusiona una señal de localización nuclear (SLN) monopartita en tándem del gran antígeno T del virus simio 40 (SV40) al extremo N-terminal de FokI. Esto garantiza la importación de Cascada^{ELD/KKR} solo intacta en el núcleo. (El complejo de poro nuclear transloca las ARN polimerasas (550 kDa) y otros complejos proteicos grandes). Como comprobación previa a las transformaciones, la actividad de nucleasa del par de nucleasas Cascada^{KKR/ELD} se comprueba *in vitro* utilizando complejos purificados y amplicones por PCR de *CCR5* para excluir la transfección de pares de nucleasas Cascada^{KKR/ELD} no productivas.

Ensayo Surveyor

Las células transfectadas se cultivan y se someten a pases durante varios días. La eficacia de la escisión del ADN objetivo *in vivo* se evalúa a continuación utilizando el ensayo Surveyor de Guschin, D.Y., y col. (2010) *Methods Mol. Biol.*, 649: 247-256. Brevemente, los amplicones de PCR del locus de ADN objetivo se mezclarán 1:1 con amplicones de PCR de células no tratadas. Estos se calientan y se les permite recocer, dando lugar a desapareamientos en los sitios objetivo que han sido reparados erróneamente por NHEJ. A continuación, se utiliza una nucleasa de desapareamiento erróneo para escindir solo moléculas de ADN desapareadas, dando un máximo de 50% de escisión cuando se completa la escisión de ADN objetivo por Cascada^{KKR/ELD}. Este procedimiento fue luego seguido por la secuenciación de los amplicones de ADN objetivo de las células tratadas. El ensayo permite una evaluación y optimización rápidas del procedimiento de administración.

Producción de pares de Cascada-nucleasa

Los complejos Cascada-nucleasa se construyeron como se ha explicado anteriormente. La purificación por afinidad de *E. coli* utilizando la subunidad Cse2 marcada con StrepII produce un complejo con la estequiometría esperada cuando se compara con Cascada nativa. Con referencia a la figura 13, esto muestra la estequiometría de la Cascada nativa (1), Cascada^{KKR} con P7 ARNcr y Cascada^{ELD} con M13 ARNcr 24 h después de la purificación utilizando solo estreptactina. Las bandas en Cascada nativa (1) son de arriba a abajo: Cse1, Cas7, Cas5, Cas6e, Cse2. Cascada^{KKR/ELD} muestra la banda de fusión FokI-Cse1 y una banda adicional que representa Cse1 con una pequeña parte de FokI como resultado de la degradación proteolítica.

Además de una proteína de fusión FokI-Cse1 intacta, se observó que una fracción de la proteína de fusión FokI-Cse1 se escinde proteolíticamente, dando como resultado una proteína Cse1 con solo el enlazador y una pequeña parte de FokI unida a ella (según lo confirmado por espectrometría de masas, datos no mostrados). En la mayoría de los aislamientos de proteínas, la fracción de proteína de fusión degradada es aproximadamente del 40 %. La proteína aislada se almacena de forma estable en el tampón de elución (HEPES 20 mM, pH 7,5, NaCl 75 mM, DTT 1 mM, destiobiotina 4 mM) con Tween 20 al 0,1 % adicional y glicerol al 50 % a -20 °C. En estas condiciones de almacenamiento, se ha descubierto que la integridad y la actividad del complejo son estables durante al menos tres semanas (datos no mostrados).

Introducción de un etiqueta His₆ y SLN en la Cascada-nucleasa

El diseño de fusión Cascada nucleasa se modificó para incorporar una señal de localización nucleolar (SLN) para permitir el transporte al núcleo células eucariotas. Para esto, una SLN monopartita en tándem del gran antígeno T del virus de simio SV40 (secuencia: PKKKRKVDPKKKRKV) se fusionó por traducción al extremo N terminal de la proteína de fusión FokI-Cse1, directamente precedido por una etiqueta His₆ en el extremo N-terminal. La etiqueta His₆ (secuencia: MHHHHHH) permite una etapa de purificación de afinidad de resina adicional de Ni²⁺ después de la purificación de StrepII. Esta etapa adicional asegura el aislamiento del complejo de fusión de Cascada-nucleasa de longitud completa, y aumenta la eficacia de la escisión al eliminar la unión de complejos Cascada no intactos al sitio objetivo formando un par de nucleasas improductivo.

Ensayo de escisión *in vitro*

La actividad y especificidad de Cascada^{KKR/ELD} se ensayó *in vitro* como se ha descrito anteriormente. La Figura 14A muestra plásmidos con distancias entre protoespaciadores de 25-50 pb (incrementos de 5 pb, carriles 1-6) incubados con Cascada^{KKR/ELD} durante 30 minutos a 37 °C. El carril 10 contiene el plásmido objetivo en sus tres topologías posibles: la banda más baja representa la forma inicial, negativamente superenrollada (nSE) del plásmido, la banda media representa la forma linealizada (escindida por XbaI), mientras que la banda superior representa la forma circular abierta (CA) (después de escotamiento con Nt.BbrCI). El carril 7 muestra la incubación de un plásmido con ambos sitios de unión eliminados (control negativo). Por lo tanto, la figura 14A muestra un ensayo de escisión típico utilizando varios plásmidos objetivo en los que los sitios de unión están separados por 25 a 50 pares de bases en incrementos de 5 pb (carriles 1 a 6). Estos plásmidos con distancias de 25-50 pb se incubaron con Cascada^{KKR/ELD} que llevaba anti P7 y M13 ARNcr, respectivamente. Un plásmido que no contiene sitios de unión sirvió como control (carril 7). El plásmido original existe en forma negativamente superenrollada (nSE, carril de control 8), y los productos escotados o linealizados son claramente distinguibles. Tras la incubación, se forma un producto de escisión lineal cuando los sitios de unión se separaron en 30, 35 y 40 pares de bases (carriles 2, 3, 4). A una distancia de 25, 45 y 50 pares de bases (carriles 1, 5, 6), el plásmido objetivo parecía estar escindido de manera incompleta conduciendo a la forma fragmentada (CA). Estos resultados muestran la mejor escisión en plásmidos con distancias entre 30 y 40 pb, lo que proporciona suficiente flexibilidad al diseñar un par de ARNcr para cualquier locus dado. Las distancias más cortas y más largas dan como resultado una mayor actividad de escotamiento al crear menos RBCs. Hay muy poca actividad en un plásmido en el que se han eliminado dos protoespaciadores, mostrando la especificidad del objetivo (carril 7).

Condiciones de escisión

Para evaluar las condiciones de tampón óptimas para los ensayos de escisión, y para estimar si se espera actividad del complejo en condiciones fisiológicas, se seleccionaron los siguientes dos tampones: (1) NEB4 (New England Biolabs, acetato de potasio 50 mM, Tris-acetato 20 mM, acetato de magnesio 10 mM, ditiotreititol 1 mM, pH 7,9) y (2) tampón O (Fermentas, Tris-HCl 50 mM, MgCl₂ 10 mM, NaCl 100 mM, 0,1 mg/ml de ASB, pH 7,5). De los dos, se recomienda NEB4 para la actividad óptima de la enzima FokI intacta comercial. Tampón O fue seleccionado de un muestreo rápido para dar buena actividad y especificidad (datos no mostrados). La Figura 14B muestra la incubación con diferentes tampones y diferentes tiempos de incubación. Los carriles 1-4 se incubaron con tampón O Fermentas (carril 1, 2 durante 15 minutos, carril 3, 4 durante 30 minutos), los carriles 5, 6 se incubaron con NEB4 (30 minutos). Los carriles 1, 3, 5 utilizaban el plásmido objetivo con espaciamento de 35 pb, los carriles 2, 4, 6 utilizaban el plásmido no objetivo (sin sitios de unión). Los carriles 7, 8 se han incubado solo con Cascada^{KKR} o Cascada^{ELD} respectivamente (tampón O). El carril 9 es el marcador de topología como en (A). Los carriles 10 y 11 muestran el plásmido objetivo y el no objetivo incubados sin adición de Cascada. Por lo tanto, en la Figura 14B, se ensayó la actividad en el plásmido objetivo con una distancia de 35 pares de bases (carril 1, 3, 5) y un plásmido de control no objetivo (carril 2, 4, 6). Hubo una gran cantidad de escotamiento inespecífico y menos escisión en NEB4 (carril 5, 6), mientras que el tampón O solo muestra actividad en el plásmido objetivo con una gran cantidad de escisión específica y poco escotamiento (carril 1-4). La diferencia es probablemente causada por la concentración de NaCl en tampón O, la mayor fuerza iónica debilita las interacciones proteína-proteína, lo que lleva a una actividad menos inespecífica. La incubación de 15 o 30 minutos muestra poca diferencia en el plásmido objetivo y el no objetivo (carril 1, 2 o 3, 4 respectivamente). La adición de un solo tipo de Cascada (P7^{KKR} o M13^{ELD}) no da como resultado la actividad de escisión (carril 7, 8) como se esperaba. Este experimento muestra que la actividad nucleasa Cascada específica por un par diseñado se produce cuando la concentración de NaCl es al menos 100 mM, que está cerca de la concentración salina fisiológica dentro de las células (NaCl 137 mM). Se espera que el par de nucleasas Cascada sea totalmente activo *in vivo*, en células eucariotas, mientras que muestra una actividad de escisión fuera de objetivo insignificante.

Sitio de escisión

Se determinó el sitio de escisión en el plásmido objetivo con un espaciamento de 35 pb (pObjetivo35). La Figura 15 muestra cómo la secuenciación revela sitios de escisión corriente arriba y corriente abajo por Cascada^{KKR/ELD} en el plásmido objetivo con espaciado de 35 pares de bases. En la Figura 15A) se muestra la región objetivo dentro de pObjetivo35 con sitios de escisión potenciales anotados. Las partes de los protoespaciadores están indicadas en rojo y azul. B) El diagrama de barras muestra cuatro patrones de escisión diferentes y su abundancia relativa dentro de clones secuenciados. Las barras azules representan el saliente generado, mientras que el borde izquierdo y derecho de cada barra representa el sitio de escisión izquierdo y derecho (véase B para la anotación).

La Figura 15A muestra la secuencia original de pObjetivo35, con sitios de escisión numerados de -7 a +7, en el que 0 se encuentra en el medio entre los dos protoespaciadores (indicados en rojo y azul). Se secuenciaron diecisiete clones y estos muestran escisión alrededor de la posición 0, creando salientes variables entre 3 y 5 pb (véase la Figura 15B). Los salientes de 4 son los más abundantes (acumulativamente 88 %), mientras que los salientes de 3 y 5 ocurren solo una vez (6 % cada uno). La escisión se produjo exactamente como se esperaba sin que los clones muestren escisión fuera del objetivo.

Escisión de un locus objetivo en células humanas.

Las nucleasas Cascada^{KKR/ELD} se modificaron con éxito para contener una etiqueta His₆ N-terminal seguida de una señal de localización nucleolar monopartita dual. Estas proteínas de fusión de Cascada nucleasa modificadas se coexpresaron con cualquiera de las dos matrices CRISPR sintéticamente construidas, cada una dirigida a un sitio de unión en el gen CCR5 humano. En primer lugar, la actividad de este nuevo par de nucleasas se valida *in vitro* mediante el ensayo de la actividad en un plásmido que contiene esta región del gen CCR5. El par nucleasa se transfecta a una estirpe celular humana, p. ej., estirpe celular HeLa. La eficiencia de la escisión del objetivo se evalúa utilizando el ensayo Surveyor como se ha descrito anteriormente.

LISTADO DE SECUENCIAS

- 10 <110> CARIBOU BIOSCIENCES, INC.
- <120> RIBONUCLEOPROTEÍNAS CASCADA MODIFICADAS Y USOS DE LAS MISMAS
- <130> SMW/FP7203417
- 15 <140> EP
- <141> 21-12-2012
- <150> EP 12812999.6
- 20 <151> 21-12-2012
- <150> PCT/EP2012/076674
- <151> 21-12-2012
- 25 <150> GB 1122458.1
- <151> 30-12-2011
- <160> 47
- 30 <170> PatentIn versión 3.5
- <210> 1
- <211> 502
- <212> PRT
- 35 <213> *Escherichia coli*
- <400> 1

ES 2 689 256 T3

Met	Asn	Leu	Leu	Ile	Asp	Asn	Trp	Ile	Pro	Val	Arg	Pro	Arg	Asn	Gly	1	5	10	15
Gly	Lys	Val	Gln	Ile	Ile	Asn	Leu	Gln	Ser	Leu	Tyr	Cys	Ser	Arg	Asp	20	25	30	
Gln	Trp	Arg	Leu	Ser	Leu	Pro	Arg	Asp	Asp	Met	Glu	Leu	Ala	Ala	Leu	35	40	45	
Ala	Leu	Leu	Val	Cys	Ile	Gly	Gln	Ile	Ile	Ala	Pro	Ala	Lys	Asp	Asp	50	55	60	
Val	Glu	Phe	Arg	His	Arg	Ile	Met	Asn	Pro	Leu	Thr	Glu	Asp	Glu	Phe	65	70	75	80
Gln	Gln	Leu	Ile	Ala	Pro	Trp	Ile	Asp	Met	Phe	Tyr	Leu	Asn	His	Ala	85	90	95	
Glu	His	Pro	Phe	Met	Gln	Thr	Lys	Gly	Val	Lys	Ala	Asn	Asp	Val	Thr	100	105	110	
Pro	Met	Glu	Lys	Leu	Leu	Ala	Gly	Val	Ser	Gly	Ala	Thr	Asn	Cys	Ala	115	120	125	

ES 2 689 256 T3

Phe Val Asn Gln Pro Gly Gln Gly Glu Ala Leu Cys Gly Gly Cys Thr
 130 135 140
 Ala Ile Ala Leu Phe Asn Gln Ala Asn Gln Ala Pro Gly Phe Gly Gly
 145 150 155 160
 Gly Phe Lys Ser Gly Leu Arg Gly Gly Thr Pro Val Thr Thr Phe Val
 165 170 175
 Arg Gly Ile Asp Leu Arg Ser Thr Val Leu Leu Asn Val Leu Thr Leu
 180 185 190
 Pro Arg Leu Gln Lys Gln Phe Pro Asn Glu Ser His Thr Glu Asn Gln
 195 200 205
 Pro Thr Trp Ile Lys Pro Ile Lys Ser Asn Glu Ser Ile Pro Ala Ser
 210 215 220
 Ser Ile Gly Phe Val Arg Gly Leu Phe Trp Gln Pro Ala His Ile Glu
 225 230 235 240
 Leu Cys Asp Pro Ile Gly Ile Gly Lys Cys Ser Cys Cys Gly Gln Glu
 245 250 255
 Ser Asn Leu Arg Tyr Thr Gly Phe Leu Lys Glu Lys Phe Thr Phe Thr
 260 265 270
 Val Asn Gly Leu Trp Pro His Pro His Ser Pro Cys Leu Val Thr Val
 275 280 285
 Lys Lys Gly Glu Val Glu Glu Lys Phe Leu Ala Phe Thr Thr Ser Ala
 290 295 300
 Pro Ser Trp Thr Gln Ile Ser Arg Val Val Val Asp Lys Ile Ile Gln
 305 310 315 320
 Asn Glu Asn Gly Asn Arg Val Ala Ala Val Val Asn Gln Phe Arg Asn
 325 330 335
 Ile Ala Pro Gln Ser Pro Leu Glu Leu Ile Met Gly Gly Tyr Arg Asn
 340 345 350
 Asn Gln Ala Ser Ile Leu Glu Arg Arg His Asp Val Leu Met Phe Asn
 355 360 365
 Gln Gly Trp Gln Gln Tyr Gly Asn Val Ile Asn Glu Ile Val Thr Val
 370 375 380

ES 2 689 256 T3

Gly Leu Gly Tyr Lys Thr Ala Leu Arg Lys Ala Leu Tyr Thr Phe Ala
385 390 395 400

Glu Gly Phe Lys Asn Lys Asp Phe Lys Gly Ala Gly Val Ser Val His
405 410 415

Glu Thr Ala Glu Arg His Phe Tyr Arg Gln Ser Glu Leu Leu Ile Pro
420 425 430

Asp Val Leu Ala Asn Val Asn Phe Ser Gln Ala Asp Glu Val Ile Ala
435 440 445

Asp Leu Arg Asp Lys Leu His Gln Leu Cys Glu Met Leu Phe Asn Gln
450 455 460

Ser Val Ala Pro Tyr Ala His His Pro Lys Leu Ile Ser Thr Leu Ala
465 470 475 480

Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu Lys Pro Gln
485 490 495

Gly Gly Pro Ser Asn Gly
500

<210> 2
<211> 160
<212> PRT
<213> *Escherichia coli*

<400> 2

Met Ala Asp Glu Ile Asp Ala Met Ala Leu Tyr Arg Ala Trp Gln Gln
1 5 10 15

Leu Asp Asn Gly Ser Cys Ala Gln Ile Arg Arg Val Ser Glu Pro Asp
20 25 30

Glu Leu Arg Asp Ile Pro Ala Phe Tyr Arg Leu Val Gln Pro Phe Gly
35 40 45

Trp Glu Asn Pro Arg His Gln Gln Ala Leu Leu Arg Met Val Phe Cys
50 55 60

Leu Ser Ala Gly Lys Asn Val Ile Arg His Gln Asp Lys Lys Ser Glu
65 70 75 80

Gln Thr Thr Gly Ile Ser Leu Gly Arg Ala Leu Ala Asn Ser Gly Arg
85 90 95

5

10

Ile Asn Glu Arg Arg Ile Phe Gln Leu Ile Arg Ala Asp Arg Thr Ala
 100 105 110

Asp Met Val Gln Leu Arg Arg Leu Leu Thr His Ala Glu Pro Val Leu
 115 120 125

Asp Trp Pro Leu Met Ala Arg Met Leu Thr Trp Trp Gly Lys Arg Glu
 130 135 140

Arg Gln Gln Leu Leu Glu Asp Phe Val Leu Thr Thr Asn Lys Asn Ala
 145 150 155 160

<210> 3
 <211> 363
 <212> PRT
 <213> *Escherichia coli*

<400> 3

Met Ser Asn Phe Ile Asn Ile His Val Leu Ile Ser His Ser Pro Ser
 1 5 10 15

Cys Leu Asn Arg Asp Asp Met Asn Met Gln Lys Asp Ala Ile Phe Gly
 20 25 30

Gly Lys Arg Arg Val Arg Ile Ser Ser Gln Ser Leu Lys Arg Ala Met
 35 40 45

Arg Lys Ser Gly Tyr Tyr Ala Gln Asn Ile Gly Glu Ser Ser Leu Arg
 50 55 60

Thr Ile His Leu Ala Gln Leu Arg Asp Val Leu Arg Gln Lys Leu Gly
 65 70 75 80

Glu Arg Phe Asp Gln Lys Ile Ile Asp Lys Thr Leu Ala Leu Leu Ser
 85 90 95

Gly Lys Ser Val Asp Glu Ala Glu Lys Ile Ser Ala Asp Ala Val Thr
 100 105 110

Pro Trp Val Val Gly Glu Ile Ala Trp Phe Cys Glu Gln Val Ala Lys
 115 120 125

Ala Glu Ala Asp Asn Leu Asp Asp Lys Lys Leu Leu Lys Val Leu Lys
 130 135 140

Glu Asp Ile Ala Ala Ile Arg Val Asn Leu Gln Gln Gly Val Asp Ile
 145 150 155 160

ES 2 689 256 T3

Ala Leu Ser Gly Arg Met Ala Thr Ser Gly Met Met Thr Glu Leu Gly
165 170 175

Lys Val Asp Gly Ala Met Ser Ile Ala His Ala Ile Thr Thr His Gln
180 185 190

Val Asp Ser Asp Ile Asp Trp Phe Thr Ala Val Asp Asp Leu Gln Glu
195 200 205

Gln Gly Ser Ala His Leu Gly Thr Gln Glu Phe Ser Ser Gly Val Phe
210 215 220

Tyr Arg Tyr Ala Asn Ile Asn Leu Ala Gln Leu Gln Glu Asn Leu Gly
225 230 235 240

Gly Ala Ser Arg Glu Gln Ala Leu Glu Ile Ala Thr His Val Val His
245 250 255

Met Leu Ala Thr Glu Val Pro Gly Ala Lys Gln Arg Thr Tyr Ala Ala
260 265 270

Phe Asn Pro Ala Asp Met Val Met Val Asn Phe Ser Asp Met Pro Leu
275 280 285

Ser Met Ala Asn Ala Phe Glu Lys Ala Val Lys Ala Lys Asp Gly Phe
290 295 300

Leu Gln Pro Ser Ile Gln Ala Phe Asn Gln Tyr Trp Asp Arg Val Ala
305 310 315 320

Asn Gly Tyr Gly Leu Asn Gly Ala Ala Ala Gln Phe Ser Leu Ser Asp
325 330 335

Val Asp Pro Ile Thr Ala Gln Val Lys Gln Met Pro Thr Leu Glu Gln
340 345 350

Leu Lys Ser Trp Val Arg Asn Asn Gly Glu Ala
355 360

<210> 4
<211> 224
<212> PRT
<213> *Escherichia coli*

<400> 4

Met Arg Ser Tyr Leu Ile Leu Arg Leu Ala Gly Pro Met Gln Ala Trp
1 5 10 15

ES 2 689 256 T3

Gly	Gln	Pro	Thr 20	Phe	Glu	Gly	Thr	Arg 25	Pro	Thr	Gly	Arg	Phe	Pro	Thr
Arg	Ser	Gly 35	Leu	Leu	Gly	Leu	Leu	Gly	Ala	Cys	Leu	Gly 45	Ile	Gln	Arg
Asp	Asp 50	Thr	Ser	Ser	Leu	Gln 55	Ala	Leu	Ser	Glu	Ser	Val	Gln	Phe	Ala
Val 65	Arg	Cys	Asp	Glu	Leu 70	Ile	Leu	Asp	Asp	Arg	Arg	Val	Ser	Val	Thr 80
Gly	Leu	Arg	Asp	Tyr 85	His	Thr	Val	Leu	Gly 90	Ala	Arg	Glu	Asp	Tyr 95	Arg
Gly	Leu	Lys	Ser 100	His	Glu	Thr	Ile	Gln 105	Thr	Trp	Arg	Glu	Tyr 110	Leu	Cys
Asp	Ala	Ser 115	Phe	Thr	Val	Ala	Leu 120	Trp	Leu	Thr	Pro	His 125	Ala	Thr	Met
Val	Ile	Ser	Glu	Leu	Glu	Lys 135	Ala	Val	Leu	Lys	Pro	Arg	Tyr	Thr	Pro
Tyr 145	Leu	Gly	Arg	Arg	Ser 150	Cys	Pro	Leu	Thr	His 155	Pro	Leu	Phe	Leu	Gly 160
Thr	Cys	Gln	Ala	Ser 165	Asp	Pro	Gln	Lys	Ala 170	Leu	Leu	Asn	Tyr	Glu 175	Pro
Val	Gly	Gly	Asp 180	Ile	Tyr	Ser	Glu	Glu 185	Ser	Val	Thr	Gly	His 190	His	Leu
Lys	Phe	Thr 195	Ala	Arg	Asp	Glu	Pro 200	Met	Ile	Thr	Leu	Pro 205	Arg	Gln	Phe
Ala	Ser 210	Arg	Glu	Trp	Tyr	Val 215	Ile	Lys	Gly	Gly	Met 220	Asp	Val	Ser	Gln

<210> 5
<211> 199
<212> PRT
<213> *Escherichia coli*

<400> 5

Met Tyr Leu Ser Lys Val Ile Ile Ala Arg Ala Trp Ser Arg Asp Leu
1 5 10 15

ES 2 689 256 T3

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

<210> 6
 <211> 392
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> GA1070943

<400> 6

ES 2 689 256 T3

```

actggaaagc gggcagtgaa aggaaggccc atgaggccag ttaattaagc ggatcctggc      60
ggcggcagcg gcggcggcag cgacaagcag aagaacggca tcaaggcgaa cttcaagatc      120
cgccacaaca tcgaggacgg cggcgtgcag ctgcgcgacc actaccagca gaacaccccc      180
atcggcgacg gccccgtgct gctgcccagc aaccactacc tgagctacca gtccgccctg      240
agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc tggagtctgt gaccgcccgc      300
gggatcactc tcggcatgga cgagctgtac aagtaagcgg ccgcggcgcg cctaggcctt      360
gacggccttc cttcaattcg ccctatagtg ag                                     392

```

<210> 7
 <211> 603
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> GA1070941

<400> 7

```

cactataggg cgaattggcg gaaggccgtc aaggccgcat ttaattaagc ggccgcaggc      60
ggcggcagcg gcggcggcag catggtgagc aaggcgagg agctgttcac cggggtggtg      120
cccacctctg tcgagctgga cggcgacgta aacggccaca agttcagcgt gtccggcgag      180
ggcgagggcg atgccaccta cggcaagctg accctgaagc tcactctgac caccggcaag      240
ctgcccgtgc cctggcccac cctcgtgacc accctcggtt acggcctgca gtgcttcgcc      300
cgctaccccg accacatgaa gcagcacgac ttcttcaagt ccgccatgcc cgaaggctac      360
gtccaggagc gcaccatctt cttcaaggac gacggcaact acaagaccg cgccgaggtg      420
aagttcgagg gcgacaccct ggtgaaccgc atcgagctga agggcatcga cttcaaggag      480
gacggcaaca tcctggggca caagctggag tacaactaca acagccaca cgtctatatc      540
acggcctaac tcgagggcgc gccctgggac tcatgggcct tccgctcact gcccgcttcc      600
cag                                                                                   603

```

<210> 8
 <211> 679
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> GA1068859

<400> 8

cactataggg cgaattggcg gaagggcgtc aaggccgcat gagctccatg gaaacaaaga 60
 attagctgat ctttaataat aaggaaatgt tacattaagg ttggtgggtt gtttttatgg 120
 gaaaaaatgc ttttaagaaca aatgtatact tttagagagt tccccgcgc agcggggata 180
 aaccggggccg attgaaggtc cgggtggatgg cttaaaagag ttccccgcgc cagcggggat 240
 aaaccgcgc aggtacagca ggtagcgcag atcatcaaga gttccccgcgc ccagcgggga 300
 taaaccgact tctctccgaa aagtcaggac gctgtggcag agttccccgc gccagcgggg 360
 ataaaccgcc tacgcgctga acgccagcgg tgtggtgaat gagttccccg cgcagcggg 420
 gataaaccgg tgtggccatg cacgccttta acggtgaact ggagttcccc gcgccagcgg 480
 ggataaaccg cacgaactca gccagaacga caaacaaaag gcgagttccc cgcgccagcg 540
 gggataaacc ggcaccagta cgcgccccac gctgacggtt tctgagttcc cgcgcccagc 600
 ggggataaac cgcagctccc attttcaaac ccaggtaccc tgggcctcat gggccttccg 660
 ctcaactgccc gctttccag 679

<210> 9
 <211> 685
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> GA1047360

<400> 9

gagctccccg gctgacggta atagaggcac ctacaggctc cggtaaaacg gaaacagcgc 60
 tggcctatgc ttggaaactt attgatcaac aaattgcgga tagtggtatt tttgccctcc 120
 caacacaagc taccgcgaat gctatgctta cgagaatgga agcgagcgcg agccacttat 180
 tttcatcccc aaatcttatt cttgctcatg gcaattcacg gtttaaccac ctctttcaat 240
 caataaaatc acgcgcgatt actgaacagg ggcaagaaga agcgtgggtt cagtgttgtc 300
 agtggttgtc acaaagcaat aagaaagtgt ttcttgggca aatcggcggt tgcacgattg 360
 atcaggtggt gatttcggta ttgccagtta aacaccgctt tatccgtggt ttgggaattg 420
 gtagatctgt ttttaattgtt aatgaagttc atgcttacga cacctatatg aacggcttgc 480
 tcgaggcagt gctcaaggct caggctgatg tgggagggag tgttattctt ctttccgcaa 540
 ccctaccaat gaaacaaaaa cagaagcttc tggatactta tgggtctgcat acagatccag 600
 tggaaaataa ctccgcatat ccactcatta actggcgagg tgtgaatggt gcgcaacggt 660
 ttgatctgct agcggatccg gtacc 685

<210> 10
 <211> 37
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador BG3186

 <400> 10
 5 atagcgccat ggaacctttt aaatatatat gccatta 37

 <210> 11
 <211> 41
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador BG3213

 15 <400> 11
 acagtgggat ccgctttggg attgcaggg atgactctgg t 41

 <210> 12
 <211> 43
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador BG3303
 25
 <400> 12
 atagcgcat gaattgctt attgataact ggattcctgt acg 43

 <210> 13
 <211> 44
 30 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador BG3212
 35
 <400> 13
 acagtggcgg ccgcgccatt tgatggccct ccttgcggtt ttaa 44

 40 <210> 14
 <211> 45
 <212> ADN
 <213> Secuencia artificial

 45 <220>
 <223> Cebador BG3076

 <400> 14
 50 cgtatatcaa actttccaat agcatgaaga gcaatgaaaa ataac 45

 <210> 15
 <211> 23
 <212> ADN
 55 <213> Secuencia artificial

 <220>
 <223> Cebador BG3449
 60
 <400> 15
 atgataccgc gagacccaag ctg 23

 <210> 16
 <211> 24

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> Cebador BG3451	
	<400> 16	
	cggataaagt tgcaggacca cttc	24
10	<210> 17	
	<211> 199	
	<212> PRT	
	<213> <i>Escherichia coli</i>	
15	<400> 17	

ES 2 689 256 T3

```

Met Tyr Leu Ser Lys Val Ile Ile Ala Arg Ala Trp Ser Arg Asp Leu
1          5          10          15

Tyr Gln Leu His Gln Gly Leu Trp His Leu Phe Pro Asn Arg Pro Asp
          20          25          30

Ala Ala Arg Asp Phe Leu Phe His Val Glu Lys Arg Asn Thr Pro Glu
          35          40          45

Gly Cys His Val Leu Leu Gln Ser Ala Gln Met Pro Val Ser Thr Ala
50          55          60

Val Ala Thr Val Ile Lys Thr Lys Gln Val Glu Phe Gln Leu Gln Val
65          70          75          80

Gly Val Pro Leu Tyr Phe Arg Leu Arg Ala Asn Pro Ile Lys Thr Ile
          85          90          95

Leu Asp Asn Gln Lys Arg Leu Asp Ser Lys Gly Asn Ile Lys Arg Cys
          100          105          110

Arg Val Pro Leu Ile Lys Glu Ala Glu Gln Ile Ala Trp Leu Gln Arg
          115          120          125

Lys Leu Gly Asn Ala Ala Arg Val Glu Asp Val His Pro Ile Ser Glu
          130          135          140

Arg Pro Gln Tyr Phe Ser Gly Asp Gly Lys Ser Gly Lys Ile Gln Thr
145          150          155          160

Val Cys Phe Glu Gly Val Leu Thr Ile Asn Asp Ala Pro Ala Leu Ile
          165          170          175

Asp Leu Val Gln Gln Gly Ile Gly Pro Ala Lys Ser Met Gly Cys Gly
          180          185          190

Leu Leu Ser Leu Ala Pro Leu
          195

```

<210> 18
 <211> 2154
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Proteína de fusión

 <400> 18

atggctcaac tggttaaaag cgaactggaa gagaaaaaaa gtgaactgog ccacaaactg	60
aaatatgtgc cgcataaata tatcgagctg attgaaattg cacgtaatcc gaccaggat	120
cgtattcttg aaatgaaagt gatggaattt tttatgaaag tgtacggcta tcgcggtgaa	180
catctgggtg gtagccgtaa accggatggt gcaatttata ccgttggtag cccgattgat	240
tatggtgtta ttgttgatac caaagcctat agcgggtggtt ataactctgc gattggtcag	300
gcagatgaaa tggaacgtta tgtggaagaa aatcagaccc gtgataaaca tctgaatccg	360
aatgaatggt ggaaagtta tccgagcagc gttacogagt ttaaattcct gtttgttagc	420
ggtcacttca aaggcaacta taaagcacag ctgacccgtc tgaatcatat taccaattgt	480
aatggtgcag ttctgagcgt tgaagaactg ctgattggtg gtgaaatgat taaagcaggc	540
accctgaccc tggaagaagt tcgtcgcaaa ttaacaatg gcgaaatcaa ctttgcggat	600
cccaccaacc gcgcgaaagg cctggaagcg gtgagcgttg cgagcatgaa tttgcttatt	660
gataactgga ttctgtacg cccgcgaaac ggggggaaag tccaaatcat aaactctcaa	720
tcgtataact gcagtagaga tcagtggcga ttaagtttgc cccgtgacga tatggaactg	780
gccgctttag cactgctggt ttgcattggg caaattatcg ccccgcaaaa agatgacgtt	840
gaatttcgac atgcataat gaatccgctc actgaagatg agtttcaaca actcatcgcg	900
ccgtggatag atatgttcta ccttaatcac gcagaacatc cctttatgca gaccaaaggt	960
gtcaaagcaa atgatgtgac tccaatggaa aaactgttgg ctggggtaag cggcgcgacg	1020
aattgtgcat ttgtcaatca accggggcag ggtgaagcat tatgtggtgg atgcactgcg	1080
attgcgttat tcaaccaggc gaatcaggca ccagggtttg gtggtggttt taaaagcggg	1140
ttacgtggag gaacacctgt aacaacgttc gtacgtggga tcgatcttcg ttcaacggtg	1200
ttactcaatg tcctcacatt acctcgtctt caaaaacaat ttctaatga atcacatacg	1260
gaaaaccaac ctacctggat taaacctatc aagtccaatg agtctatacc tgcttcgtca	1320
attgggtttg tccgtggtct attctggcaa ccagcgcata ttgaattatg cgatccatt	1380
gggattggta aatgttcttg ctgtggacag gaaagcaatt tgcgttatac cggttttctt	1440
aaggaaaaat ttacctttac agttaatggg ctatggcccc atccgcattc ccctgtctg	1500
gtaacagtca agaaagggga ggttgaggaa aaatttcttg ctttcaccac ctccgcacca	1560
tcattggacac aaatcagccg agttgtggta gataagatta ttcaaatga aaatggaaat	1620

ES 2 689 256 T3

cgcggtggcgg	cggttgtgaa	tcaattcaga	aatattgcgc	cgcaaagtcc	tcttgaattg	1680
attatggggg	gatatcgtaa	taatcaagca	tctattcttg	aacggcgtca	tgatgtgttg	1740
atgtttaatc	aggggtggca	acaatacggc	aatgtgataa	acgaaatagt	gactgttgg	1800
ttgggatata	aaacagcctt	acgcaaggcg	ttatatacct	ttgcagaagg	gttataaaat	1860
aaagacttca	aaggggccgg	agtctctgtt	catgagactg	cagaaaggca	tttctatcga	1920
cagagtgaat	tattaattcc	cgatgtactg	gcgaatgtta	atttttccca	ggctgatgag	1980
gtaatagctg	atttacgaga	caaacttcat	caattgtgtg	aaatgctatt	taatcaatct	2040
gtagctccct	atgcacatca	tcctaaatta	ataagcacat	tagcgcttgc	cgcgccacg	2100
ctatacaaac	atttacggga	gttataaacg	caaggagggc	catcaaatgg	ctga	2154

<210> 19
 <211> 717
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Proteína de fusión

<400> 19

ES 2 689 256 T3

Met	Ala	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	1	5	10	15
Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	20	25	30	
Ile	Ala	Arg	Asn	Pro	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	35	40	45	
Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Glu	His	Leu	Gly	Gly	50	55	60	
Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	65	70	75	80
Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	85	90	95	
Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Glu	Arg	Tyr	Val	Glu	Glu	Asn	Gln	100	105	110	
Thr	Arg	Asp	Lys	His	Leu	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	115	120	125	
Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys				

ES 2 689 256 T3

130						135						140				
Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	
145					150					155					160	
Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	
				165					170					175		
Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	
			180					185					190			
Asn	Gly	Glu	Ile	Asn	Phe	Ala	Asp	Pro	Thr	Asn	Arg	Ala	Lys	Gly	Leu	
		195					200					205				
Glu	Ala	Val	Ser	Val	Ala	Ser	Met	Asn	Leu	Leu	Ile	Asp	Asn	Trp	Ile	
	210						215				220					
Pro	Val	Arg	Pro	Arg	Asn	Gly	Gly	Lys	Val	Gln	Ile	Ile	Asn	Leu	Gln	
225					230					235					240	
Ser	Leu	Tyr	Cys	Ser	Arg	Asp	Gln	Trp	Arg	Leu	Ser	Leu	Pro	Arg	Asp	
				245					250					255		
Asp	Met	Glu	Leu	Ala	Ala	Leu	Ala	Leu	Leu	Val	Cys	Ile	Gly	Gln	Ile	
			260					265					270			
Ile	Ala	Pro	Ala	Lys	Asp	Asp	Val	Glu	Phe	Arg	His	Arg	Ile	Met	Asn	
		275					280					285				
Pro	Leu	Thr	Glu	Asp	Glu	Phe	Gln	Gln	Leu	Ile	Ala	Pro	Trp	Ile	Asp	
	290					295					300					
Met	Phe	Tyr	Leu	Asn	His	Ala	Glu	His	Pro	Phe	Met	Gln	Thr	Lys	Gly	
305					310					315					320	
Val	Lys	Ala	Asn	Asp	Val	Thr	Pro	Met	Glu	Lys	Leu	Leu	Ala	Gly	Val	
				325					330					335		
Ser	Gly	Ala	Thr	Asn	Cys	Ala	Phe	Val	Asn	Gln	Pro	Gly	Gln	Gly	Glu	
			340					345					350			
Ala	Leu	Cys	Gly	Gly	Cys	Thr	Ala	Ile	Ala	Leu	Phe	Asn	Gln	Ala	Asn	
		355					360					365				
Gln	Ala	Pro	Gly	Phe	Gly	Gly	Gly	Phe	Lys	Ser	Gly	Leu	Arg	Gly	Gly	
	370					375					380					

ES 2 689 256 T3

Thr	Pro	Val	Thr	Thr	Phe	Val	Arg	Gly	Ile	Asp	Leu	Arg	Ser	Thr	Val	385	390	395	400
Leu	Leu	Asn	Val	Leu	Thr	Leu	Pro	Arg	Leu	Gln	Lys	Gln	Phe	Pro	Asn	405	410	415	
Glu	Ser	His	Thr	Glu	Asn	Gln	Pro	Thr	Trp	Ile	Lys	Pro	Ile	Lys	Ser	420	425	430	
Asn	Glu	Ser	Ile	Pro	Ala	Ser	Ser	Ile	Gly	Phe	Val	Arg	Gly	Leu	Phe	435	440	445	
Trp	Gln	Pro	Ala	His	Ile	Glu	Leu	Cys	Asp	Pro	Ile	Gly	Ile	Gly	Lys	450	455	460	
Cys	Ser	Cys	Cys	Gly	Gln	Glu	Ser	Asn	Leu	Arg	Tyr	Thr	Gly	Phe	Leu	465	470	475	480
Lys	Glu	Lys	Phe	Thr	Phe	Thr	Val	Asn	Gly	Leu	Trp	Pro	His	Pro	His	485	490	495	
Ser	Pro	Cys	Leu	Val	Thr	Val	Lys	Lys	Gly	Glu	Val	Glu	Glu	Lys	Phe	500	505	510	
Leu	Ala	Phe	Thr	Thr	Ser	Ala	Pro	Ser	Trp	Thr	Gln	Ile	Ser	Arg	Val	515	520	525	
Val	Val	Asp	Lys	Ile	Ile	Gln	Asn	Glu	Asn	Gly	Asn	Arg	Val	Ala	Ala	530	535	540	
Val	Val	Asn	Gln	Phe	Arg	Asn	Ile	Ala	Pro	Gln	Ser	Pro	Leu	Glu	Leu	545	550	555	560
Ile	Met	Gly	Gly	Tyr	Arg	Asn	Asn	Gln	Ala	Ser	Ile	Leu	Glu	Arg	Arg	565	570	575	
His	Asp	Val	Leu	Met	Phe	Asn	Gln	Gly	Trp	Gln	Gln	Tyr	Gly	Asn	Val	580	585	590	
Ile	Asn	Glu	Ile	Val	Thr	Val	Gly	Leu	Gly	Tyr	Lys	Thr	Ala	Leu	Arg	595	600	605	
Lys	Ala	Leu	Tyr	Thr	Phe	Ala	Glu	Gly	Phe	Lys	Asn	Lys	Asp	Phe	Lys	610	615	620	
Gly	Ala	Gly	Val	Ser	Val	His	Glu	Thr	Ala	Glu	Arg	His	Phe	Tyr	Arg	625	630	635	640

Gln Ser Glu Leu Leu Ile Pro Asp Val Leu Ala Asn Val Asn Phe Ser
645 650 655

Gln Ala Asp Glu Val Ile Ala Asp Leu Arg Asp Lys Leu His Gln Leu
660 665 670

Cys Glu Met Leu Phe Asn Gln Ser Val Ala Pro Tyr Ala His His Pro
675 680 685

Lys Leu Ile Ser Thr Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His
690 695 700

Leu Arg Glu Leu Lys Pro Gln Gly Gly Pro Ser Asn Gly
705 710 715

<210> 20
<211> 2154
<212> ADN
<213> Secuencia artificial

<220>
<223> Proteína de fusión

<400> 20

atggctcaac tgggttaaaag cgaactggaa gagaaaaaaa gtgaactgcg ccacaaactg	60
aaatatgtgc cgcatagaata tatcgagctg attgaaattg cagtaaatcc gaccacaggat	120
cgtattctgg aatgaaagt gatggaattt tttatgaaag tgtacggcta tcgcggtgaa	180
catctgggtg gtagccgtaa accggatggt gcaatttata ccgttggtag ccgattgat	240
tatggtgtta ttgttgatac caaagcctat agcgggtggt ataactctgc gattggtcag	300
gcagatgaaa tgcagcgtta tgtgaaagaa aatcagaccc gcaacaaaca tattaaoccg	360
aatgaatggt ggaaagtta tcogagcagc gttaccgagt ttaaattcct gtttggttagc	420
ggtcacttca aaggcaacta taaagcacag ctgaccgcgc tgaatcgtaa aaccaattgt	480
aatggtgcag ttctgagcgt tgaagaactg ctgattggtg gtgaaatgat taaagcaggc	540
accctgaacc tggaagaagt tcgtcgcaaa tttaacaatg gcgaaatcaa ctttgcggtat	600
cccaccaacc gcgcgaaagg cctggaagcg gtgagcgtgg cgagcatgaa tttgcttatt	660
gataactgga ttctgtacg ccgcgaaac ggggggaaag tccaaatcat aaatctgcaa	720
tcgctatact gcagtagaga tcagtggcga ttaagtttgc ccgtgacga tatggaactg	780
gccgcttttag cactgctggt ttgcattggg caaattatcg ccccgcaaa agatgacggt	840
gaatttcgac atcgcataat gaatccgctc actgaagatg agtttcaaca actcatcgcg	900
ccgtggatag atatgttcta ccttaatcac gcagaacatc cctttatgca gaccaaaggt	960
gtcaaagcaa atgatgtgac tccaatggaa aaactgttggt ctggggtaag cggcgcgacg	1020

aattgtgcat ttgtcaatca accggggcag ggtgaagcat tatgtggtgg atgcactgcg 1080
attgcgttat tcaaccaggc gaatcaggca ccaggttttg gtggtggttt taaaagcggc 1140
ttacgtggag gaacacctgt aacaacgttc gtacgtggga tcgatcttcg ttcaacggtg 1200
ttactcaatg tcctcacatt acctcgtcctt caaaaacaat ttcctaataga atcacatacg 1260
gaaaaccaac ctacctggat taaacctatc aagtccaatg agtctatacc tgcttcgtca 1320
attgggtttg tccgtggtct attctggcaa ccagcgcata ttgaattatg cgatccatt 1380
gggattggta aatgttcttg ctgtggacag gaaagcaatt tgcgttatac cggttttctt 1440
aaggaaaaat ttacctttac agttaatggg ctatggcccc atccgcattc cccttgctcg 1500
gtaacagtca agaaagggga ggttgaggaa aaatttcttg ctttcaccac ctccgcacca 1560
tcatggacac aaatcagccg agttgtggta gataagatta ttcaaatga aaatggaaat 1620
cgcgtggcgg cggttgtgaa tcaattcaga aatattgcgc cgcaaagtcc tcttgaattg 1680
attatggggg gatatcgtaa taatcaagca tctattcttg aacggcgtca tgatgtgttg 1740
atgtttaatc aggggtggca acaatacggc aatgtgataa acgaaatagt gactgttggt 1800
ttgggatata aaacagcctt acgcaaggcg ttatatacct ttgcagaagg gtttaaaaat 1860
aaagacttca aaggggccgg agtctctgtt catgagactg cagaaaggca tttctatcga 1920
cagagtgaat tattaattcc cgatgtactg gcgaatgtta atttttccca ggctgatgag 1980
gtaatagctg atttacgaga caaacttcat caattgtgtg aaatgctatt taatcaatct 2040
gtagctccct atgcacatca tctaaatta ataagcacat tagcgcttgc ccgcgccacg 2100
ctatacaaac atttacggga gttaaaaccg caaggagggc catcaaattg ctga 2154

<210> 21
<211> 717
5 <212> PRT
<213> Secuencia artificial
<220>
10 <223> Proteína de fusión
<400> 21

Met Ala Gln Leu Val Lys Ser Glu Leu Glu Lys Lys Ser Glu Leu
1 5 10 15

Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
20 25 30

Ile Ala Arg Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
35 40 45

Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly

ES 2 689 256 T3

50		55		60
Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp				
65		70		75
				80
Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu				
		85		90
				95
Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Lys Glu Asn Gln				
		100		105
				110
Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro				
		115		120
				125
Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys				
		130		135
				140
Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn Arg Lys Thr Asn Cys				
		145		150
				155
				160
Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met				
		165		170
				175
Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn				
		180		185
				190
Asn Gly Glu Ile Asn Phe Ala Asp Pro Thr Asn Arg Ala Lys Gly Leu				
		195		200
				205
Glu Ala Val Ser Val Ala Ser Met Asn Leu Leu Ile Asp Asn Trp Ile				
		210		215
				220
Pro Val Arg Pro Arg Asn Gly Gly Lys Val Gln Ile Ile Asn Leu Gln				
		225		230
				235
				240
Ser Leu Tyr Cys Ser Arg Asp Gln Trp Arg Leu Ser Leu Pro Arg Asp				
		245		250
				255
Asp Met Glu Leu Ala Ala Leu Ala Leu Leu Val Cys Ile Gly Gln Ile				
		260		265
				270
Ile Ala Pro Ala Lys Asp Asp Val Glu Phe Arg His Arg Ile Met Asn				
		275		280
				285
Pro Leu Thr Glu Asp Glu Phe Gln Gln Leu Ile Ala Pro Trp Ile Asp				
		290		295
				300

Met 305	Phe	Tyr	Leu	Asn	His 310	Ala	Glu	His	Pro	Phe 315	Met	Gln	Thr	Lys	Gly 320
Val	Lys	Ala	Asn	Asp 325	Val	Thr	Pro	Met	Glu 330	Lys	Leu	Leu	Ala	Gly 335	Val
Ser	Gly	Ala	Thr 340	Asn	Cys	Ala	Phe	Val 345	Asn	Gln	Pro	Gly	Gln 350	Gly	Glu
Ala	Leu	Cys 355	Gly	Gly	Cys	Thr	Ala 360	Ile	Ala	Leu	Phe	Asn 365	Gln	Ala	Asn
Gln	Ala 370	Pro	Gly	Phe	Gly	Gly 375	Gly	Phe	Lys	Ser	Gly 380	Leu	Arg	Gly	Gly
Thr 385	Pro	Val	Thr	Thr	Phe 390	Val	Arg	Gly	Ile	Asp 395	Leu	Arg	Ser	Thr	Val 400
Leu	Leu	Asn	Val	Leu 405	Thr	Leu	Pro	Arg	Leu 410	Gln	Lys	Gln	Phe	Pro 415	Asn
Glu	Ser	His	Thr 420	Glu	Asn	Gln	Pro	Thr 425	Trp	Ile	Lys	Pro	Ile 430	Lys	Ser
Asn	Glu	Ser 435	Ile	Pro	Ala	Ser	Ser 440	Ile	Gly	Phe	Val	Arg 445	Gly	Leu	Phe
Trp	Gln 450	Pro	Ala	His	Ile	Glu 455	Leu	Cys	Asp	Pro	Ile 460	Gly	Ile	Gly	Lys
Cys 465	Ser	Cys	Cys	Gly	Gln 470	Glu	Ser	Asn	Leu	Arg 475	Tyr	Thr	Gly	Phe	Leu 480
Lys	Glu	Lys	Phe	Thr 485	Phe	Thr	Val	Asn	Gly 490	Leu	Trp	Pro	His	Pro 495	His
Ser	Pro	Cys	Leu 500	Val	Thr	Val	Lys	Lys 505	Gly	Glu	Val	Glu	Glu 510	Lys	Phe
Leu	Ala	Phe	Thr	Thr	Ser	Ala	Pro 520	Ser	Trp	Thr	Gln	Ile 525	Ser	Arg	Val
Val	Val 530	Asp	Lys	Ile	Ile	Gln 535	Asn	Glu	Asn	Gly	Asn 540	Arg	Val	Ala	Ala
Val 545	Val	Asn	Gln	Phe	Arg 550	Asn	Ile	Ala	Pro	Gln 555	Ser	Pro	Leu	Glu	Leu 560

ES 2 689 256 T3

Ile Met Gly Gly Tyr Arg Asn Asn Gln Ala Ser Ile Leu Glu Arg Arg
565 570 575

His Asp Val Leu Met Phe Asn Gln Gly Trp Gln Gln Tyr Gly Asn Val
580 585 590

Ile Asn Glu Ile Val Thr Val Gly Leu Gly Tyr Lys Thr Ala Leu Arg
595 600 605

Lys Ala Leu Tyr Thr Phe Ala Glu Gly Phe Lys Asn Lys Asp Phe Lys
610 615 620

Gly Ala Gly Val Ser Val His Glu Thr Ala Glu Arg His Phe Tyr Arg
625 630 635 640

Gln Ser Glu Leu Leu Ile Pro Asp Val Leu Ala Asn Val Asn Phe Ser
645 650 655

Gln Ala Asp Glu Val Ile Ala Asp Leu Arg Asp Lys Leu His Gln Leu
660 665 670

Cys Glu Met Leu Phe Asn Gln Ser Val Ala Pro Tyr Ala His His Pro
675 680 685

Lys Leu Ile Ser Thr Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His
690 695 700

Leu Arg Glu Leu Lys Pro Gln Gly Gly Pro Ser Asn Gly
705 710 715

<210> 22
<211> 2235
<212> ADN
<213> Secuencia artificial

<220>
<223> Proteína de fusión

<400> 22

atgcatcacc atcatcacca cccgaaaaaa aagcgcaaag tggatccgaa gaaaaaacgt 60
aaagttgaag atccgaaaga catggctcaa ctggttaaaa gccaactgga agagaaaaaa 120
agtgaactgc gccacaaact gaaatatgtg ccgcatgaat atacgagct gattgaaatt 180
gcacgtaatc cgacccagga tcgtattctg gaaatgaaag tgatggaatt ttttatgaaa 240
gtgtacgggt atcgcggtga acatctgggt ggtagccgta aaccggatgg tgcaatttat 300
accggttgga gcccgattga ttatggtgtt attgttgata ccaaagccta tagcgggtgt 360
tataatctgc cgattggtca ggcagatgaa atgcagcggt atgtgaaaga aaatcagacc 420

ES 2 689 256 T3

```

cgcaacaaac atattaaccc gaatgaatgg tggaaagttt atccgagcag cgttaccgag      480
tttaaatcc tgtttgttag cggtcacttc aaaggcaact ataaagcaca gctgacccgt      540
ctgaatcgta aaaccaattg taatggtgca gttctgagcg ttgaagaact gctgattggt      600
ggtgaaatga ttaaagcagg caccctgacc ctggaagaag ttcgtcgcaa atttaacaat      660
ggcgaaatca actttgcgga tcccaccaac cgcgcgaaa ggcctggaagc ggtgagcgtg      720
gcgagcatga atttgcttat tgataactgg attcctgtac gcccgcgaaa cggggggaaa      780
gtccaaatca taaatctgca atcgctatac tgcagtagag atcagtggcg attaagtttg      840
ccccgtgacg atatggaact ggccgcttta gcactgctgg tttgcattgg gcaaattatc      900
gccccggcaa aagatgacgt tgaatttoga catcgcataa tgaatccgct cactgaagat      960
gagtttcaac aactcatogc gccgtggata gatatgttct accttaatca cgcagaacat     1020
ccctttatgc agaccaaagg tgtcaaagca aatgatgtga ctccaatgga aaaactggtg     1080
gctggggtaa gcggcgcgac gaattgtgca tttgtcaatc aaccggggca ggggtgaagca     1140
ttatgtggtg gatgcactgc gattgctgta ttcaaccagg cgaatcaggc accagggttt     1200
ggtggtggtt ttaaaagcgg tttacgtgga ggaacacctg taacaacgtt cgtacgtggg     1260
atcgatcttc gttcaacggt gttactcaat gtccctcacat tacctcgtct tcaaaaacaa     1320
tttcctaattg aatcacatac ggaaaaaccaa cotacctgga ttaaacctat caagtccaat     1380
gagtctatac ctgcttcgtc aattgggttt gtccgtggtc tattctggca accagcgcat     1440
attgaattat gcgatcccat tgggattggt aaatgttctt gctgtggaca ggaagcaat     1500
ttgcgttata ccggttttct taaggaaaaa tttaccttta cagttaatgg gctatggccc     1560
catccgcatt cccctgtctt ggtaacagtc aagaaagggg aggttgagga aaaaattctt     1620
gctttcacca cctccgcacc atcatggaca caaatcagcc gagttgtggt agataagatt     1680
attcaaatg aaaatggaaa tcgcgtggcg gcggttgtga atcaattcag aaatattgcg     1740
ccgcaaagtc ctcttgaatt gattatgggg ggatatcgta ataataagc atctattctt     1800
gaacggcgtc atgatgtgtt gatgtttaat caggggtggc aacaatacgg caatgtgata     1860
aacgaaatag tgactgttgg tttgggatat aaaacagcct tacgcaaggc gttatatacc     1920
tttgcagaag ggtttaaaaa taaagacttc aaaggggccg gagtctctgt tcatgagact     1980
gcagaaaggc atttctatcg acagagtga ttattaattc ccgatgtact ggcgaatgtt     2040
aatttttccc aggctgatga ggtaatagct gatttacgag acaaacttca tcaattgtgt     2100
gaaatgctat ttaatcaatc tgtagctccc tatgcacatc atcctaaatt aataagcaca     2160
ttagcgcttg ccgcgccac gctatacaaa catttacggg agttaaaacc gcaaggaggg     2220
ccatcaaatg gctga                                         2235

```

<210> 23
 <211> 744
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Proteína de fusión

<400> 23

ES 2 689 256 T3

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

ES 2 689 256 T3

Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	210	215	220
Phe	Ala	Asp	Pro	Thr	Asn	Arg	Ala	Lys	Gly	Leu	Glu	Ala	Val	Ser	Val	225	230	235
Ala	Ser	Met	Asn	Leu	Leu	Ile	Asp	Asn	Trp	Ile	Pro	Val	Arg	Pro	Arg	245	250	255
Asn	Gly	Gly	Lys	Val	Gln	Ile	Ile	Asn	Leu	Gln	Ser	Leu	Tyr	Cys	Ser	260	265	270
Arg	Asp	Gln	Trp	Arg	Leu	Ser	Leu	Pro	Arg	Asp	Asp	Met	Glu	Leu	Ala	275	280	285
Ala	Leu	Ala	Leu	Leu	Val	Cys	Ile	Gly	Gln	Ile	Ile	Ala	Pro	Ala	Lys	290	295	300
Asp	Asp	Val	Glu	Phe	Arg	His	Arg	Ile	Met	Asn	Pro	Leu	Thr	Glu	Asp	305	310	315
Glu	Phe	Gln	Gln	Leu	Ile	Ala	Pro	Trp	Ile	Asp	Met	Phe	Tyr	Leu	Asn	325	330	335
His	Ala	Glu	His	Pro	Phe	Met	Gln	Thr	Lys	Gly	Val	Lys	Ala	Asn	Asp	340	345	350
Val	Thr	Pro	Met	Glu	Lys	Leu	Leu	Ala	Gly	Val	Ser	Gly	Ala	Thr	Asn	355	360	365
Cys	Ala	Phe	Val	Asn	Gln	Pro	Gly	Gln	Gly	Glu	Ala	Leu	Cys	Gly	Gly	370	375	380
Cys	Thr	Ala	Ile	Ala	Leu	Phe	Asn	Gln	Ala	Asn	Gln	Ala	Pro	Gly	Phe	385	390	395
Gly	Gly	Gly	Phe	Lys	Ser	Gly	Leu	Arg	Gly	Gly	Thr	Pro	Val	Thr	Thr	405	410	415
Phe	Val	Arg	Gly	Ile	Asp	Leu	Arg	Ser	Thr	Val	Leu	Leu	Asn	Val	Leu	420	425	430
Thr	Leu	Pro	Arg	Leu	Gln	Lys	Gln	Phe	Pro	Asn	Glu	Ser	His	Thr	Glu	435	440	445
Asn	Gln	Pro	Thr	Trp	Ile	Lys	Pro	Ile	Lys	Ser	Asn	Glu	Ser	Ile	Pro	450	455	460

ES 2 689 256 T3

Ala	Ser	Ser	Ile	Gly	Phe	Val	Arg	Gly	Leu	Phe	Trp	Gln	Pro	Ala	His	465	470	475	480
Ile	Glu	Leu	Cys	Asp	Pro	Ile	Gly	Ile	Gly	Lys	Cys	Ser	Cys	Cys	Gly	485	490	495	
Gln	Glu	Ser	Asn	Leu	Arg	Tyr	Thr	Gly	Phe	Leu	Lys	Glu	Lys	Phe	Thr	500	505	510	
Phe	Thr	Val	Asn	Gly	Leu	Trp	Pro	His	Pro	His	Ser	Pro	Cys	Leu	Val	515	520	525	
Thr	Val	Lys	Lys	Gly	Glu	Val	Glu	Glu	Lys	Phe	Leu	Ala	Phe	Thr	Thr	530	535	540	
Ser	Ala	Pro	Ser	Trp	Thr	Gln	Ile	Ser	Arg	Val	Val	Val	Asp	Lys	Ile	545	550	555	560
Ile	Gln	Asn	Glu	Asn	Gly	Asn	Arg	Val	Ala	Ala	Val	Val	Asn	Gln	Phe	565	570	575	
Arg	Asn	Ile	Ala	Pro	Gln	Ser	Pro	Leu	Glu	Leu	Ile	Met	Gly	Gly	Tyr	580	585	590	
Arg	Asn	Asn	Gln	Ala	Ser	Ile	Leu	Glu	Arg	Arg	His	Asp	Val	Leu	Met	595	600	605	
Phe	Asn	Gln	Gly	Trp	Gln	Gln	Tyr	Gly	Asn	Val	Ile	Asn	Glu	Ile	Val	610	615	620	
Thr	Val	Gly	Leu	Gly	Tyr	Lys	Thr	Ala	Leu	Arg	Lys	Ala	Leu	Tyr	Thr	625	630	635	640
Phe	Ala	Glu	Gly	Phe	Lys	Asn	Lys	Asp	Phe	Lys	Gly	Ala	Gly	Val	Ser	645	650	655	
Val	His	Glu	Thr	Ala	Glu	Arg	His	Phe	Tyr	Arg	Gln	Ser	Glu	Leu	Leu	660	665	670	
Ile	Pro	Asp	Val	Leu	Ala	Asn	Val	Asn	Phe	Ser	Gln	Ala	Asp	Glu	Val	675	680	685	
Ile	Ala	Asp	Leu	Arg	Asp	Lys	Leu	His	Gln	Leu	Cys	Glu	Met	Leu	Phe	690	695	700	
Asn	Gln	Ser	Val	Ala	Pro	Tyr	Ala	His	His	Pro	Lys	Leu	Ile	Ser	Thr	705	710	715	720

Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu Lys
725 730 735

Pro Gln Gly Gly Pro Ser Asn Gly
740

<210> 24
<211> 2235
<212> ADN
<213> Secuencia artificial

<220>
<223> Proteína de fusión

<400> 24

```

atgcatcacc atcatcacca cccgaaaaaa aagcgcaaag tggatccgaa gaaaaaacgt      60
aaagttgaag atcogaaaga catggctcaa ctggttaaaa gcgaactgga agagaaaaaa      120
agtgaactgc gccacaaact gaaatatgtg cgcgatgaat atatcgagct gattgaaatt      180
gcacgtaatc cgaccagga tcgtattctg gaaatgaaag tgatggaatt ttttatgaaa      240
gtgtacggct atcgcgggtga acatctgggt ggtagccgta aaccggatgg tgcaatttat      300
accgttggtg gcccgattga ttatggtggt attgttgata ccaaagccta tagcgggtgg      360
tataatctgc cgattggtca ggcagatgaa atggaacgtt atgtggaaga aaatcagacc      420
cgtgataaac atctgaatcc gaatgaatgg tggaaagttt atccgagcag cgttacccgag      480
tttaaattcc tgtttggttag cggtcacttc aaaggcaact ataaagcaca gctgaccogt      540
ctgaatcata ttaccaattg taatggtgca gttctgagcg ttgaagaact gctgattggt      600
ggtgaaatga ttaaagcagg caccctgacc ctggaagaag ttcgtcgcaa atttaacaat      660
ggcgaaatca actttgcgga tcccaccaac cgcgcgaaag gcctggaagc ggtgagcgtg      720
gcgagcatga atttgcttat tgataactgg attcctgtac gcccgcgaaa cggggggaaa      780
gtccaaatca taaatctgca atcgctatac tgcagtagag atcagtggcg attaagtttg      840
ccccgtgacg atatggaact ggccgcttta gcactgctgg tttgcattgg gcaaattatc      900
gccccggcaa aagatgacgt tgaatttcga catcgcataa tgaatccgct cactgaagat      960
gagtttcaac aactcatcgc gccgtggata gatatgttct accttaatca cgcagaacat     1020
ccctttatgc agaccaaagg tgtcaaagca aatgatgtga ctccaatgga aaaactgttg     1080
gctggggtaa gcggcgcgac gaattgtgca tttgtcaatc aaccggggca gggtgaaagca     1140
ttatgtggtg gatgcactgc gattgcgtta ttcaaccagg cgaatcaggc accaggtttt     1200
ggtggtggtt ttaaaagcgg ttacgtgga ggaacacctg taacaacgtt cgtacgtggg     1260
atcgatcttc gttcaacggt gttactcaat gtcctcacat tacctcgtct tcaaaaacaa     1320

```

```

tttcctaattg aatcacatac ggaaaaccaa cctacctgga ttaaacctat caagtccaat 1380
gagtctatac ctgcttcgtc aattgggttt gtccgtgggc tattctggca accagcgcat 1440
attgaattat gcgatcccat tgggattggg aaatgttctt gctgtggaca ggaaagcaat 1500
ttgcgttata ccggttttct taaggaaaaa tttaccttta cagttaatgg gctatggccc 1560
catccgcatt ccccttgtct ggtaacagtc aagaaagggg aggttgagga aaaatttctt 1620
gctttcacca cctccgcacc atcatggaca caaatcagcc gagttgtggg agataagatt 1680
attcaaatg aaaatggaaa tcgcgtggcg gcggttgtga atcaattcag aaatattgcg 1740
ccgcaaagtc ctcttgaatt gattatgggg ggatatcgta ataatcaagc atctattctt 1800
gaacggcgtc atgatgtgtt gatgtttaat caggggtggc aacaatacgg caatgtgata 1860
aacgaaatag tgactgttgg tttgggatat aaaacagcct tacgcaaggc gttatatacc 1920
tttgcagaag ggtttaaaaa taaagacttc aaaggggccc gagtctctgt tcatgagact 1980
gcagaaaggc atttctatcg acagagtga tattaattc ccgatgtact ggcgaatgtt 2040
aatttttccc aggtgatga ggtaatagct gatttacgag acaaacttca tcaattgtgt 2100
gaaatgctat ttaatcaatc ttagctccc tatgcacatc atcctaaatt aataagcaca 2160
ttagcgcttg cccgcgccac gctatacaaa catttacggg agttaaacc gcaaggaggg 2220
ccatcaaatg gctga 2235

```

<210> 25
 <211> 744
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Proteína de fusión

<400> 25

```

Met His His His His His His Pro Lys Lys Lys Arg Lys Val Asp Pro
1          5          10          15

Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Asp Met Ala Gln Leu Val
20          25          30

Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys
35          40          45

Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Pro
50          55          60

Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys
65          70          75          80

```

Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro Asp
85 90 95

Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val
100 105 110

Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala
115 120 125

Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln Thr Arg Asp Lys His
130 135 140

Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu
145 150 155 160

Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala
165 170 175

Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu
180 185 190

Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr
195 200 205

Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn
210 215 220

Phe Ala Asp Pro Thr Asn Arg Ala Lys Gly Leu Glu Ala Val Ser Val
225 230 235 240

Ala Ser Met Asn Leu Leu Ile Asp Asn Trp Ile Pro Val Arg Pro Arg
245 250 255

Asn Gly Gly Lys Val Gln Ile Ile Asn Leu Gln Ser Leu Tyr Cys Ser
260 265 270

Arg Asp Gln Trp Arg Leu Ser Leu Pro Arg Asp Asp Met Glu Leu Ala
275 280 285

Ala Leu Ala Leu Leu Val Cys Ile Gly Gln Ile Ile Ala Pro Ala Lys
290 295 300

Asp Asp Val Glu Phe Arg His Arg Ile Met Asn Pro Leu Thr Glu Asp
305 310 315 320

Glu Phe Gln Gln Leu Ile Ala Pro Trp Ile Asp Met Phe Tyr Leu Asn
325 330 335

ES 2 689 256 T3

His	Ala	Glu	His	Pro	Phe	Met	Gln	Thr	Lys	Gly	Val	Lys	Ala	Asn	Asp	340	345	350	
Val	Thr	Pro	Met	Glu	Lys	Leu	Leu	Ala	Gly	Val	Ser	Gly	Ala	Thr	Asn	355	360	365	
Cys	Ala	Phe	Val	Asn	Gln	Pro	Gly	Gln	Gly	Glu	Ala	Leu	Cys	Gly	Gly	370	375	380	
Cys	Thr	Ala	Ile	Ala	Leu	Phe	Asn	Gln	Ala	Asn	Gln	Ala	Pro	Gly	Phe	385	390	395	400
Gly	Gly	Gly	Phe	Lys	Ser	Gly	Leu	Arg	Gly	Gly	Thr	Pro	Val	Thr	Thr	405	410	415	
Phe	Val	Arg	Gly	Ile	Asp	Leu	Arg	Ser	Thr	Val	Leu	Leu	Asn	Val	Leu	420	425	430	
Thr	Leu	Pro	Arg	Leu	Gln	Lys	Gln	Phe	Pro	Asn	Glu	Ser	His	Thr	Glu	435	440	445	
Asn	Gln	Pro	Thr	Trp	Ile	Lys	Pro	Ile	Lys	Ser	Asn	Glu	Ser	Ile	Pro	450	455	460	
Ala	Ser	Ser	Ile	Gly	Phe	Val	Arg	Gly	Leu	Phe	Trp	Gln	Pro	Ala	His	465	470	475	480
Ile	Glu	Leu	Cys	Asp	Pro	Ile	Gly	Ile	Gly	Lys	Cys	Ser	Cys	Cys	Gly	485	490	495	
Gln	Glu	Ser	Asn	Leu	Arg	Tyr	Thr	Gly	Phe	Leu	Lys	Glu	Lys	Phe	Thr	500	505	510	
Phe	Thr	Val	Asn	Gly	Leu	Trp	Pro	His	Pro	His	Ser	Pro	Cys	Leu	Val	515	520	525	
Thr	Val	Lys	Lys	Gly	Glu	Val	Glu	Glu	Lys	Phe	Leu	Ala	Phe	Thr	Thr	530	535	540	
Ser	Ala	Pro	Ser	Trp	Thr	Gln	Ile	Ser	Arg	Val	Val	Val	Asp	Lys	Ile	545	550	555	560
Ile	Gln	Asn	Glu	Asn	Gly	Asn	Arg	Val	Ala	Ala	Val	Val	Asn	Gln	Phe	565	570	575	
Arg	Asn	Ile	Ala	Pro	Gln	Ser	Pro	Leu	Glu	Leu	Ile	Met	Gly	Gly	Tyr	580	585	590	

ES 2 689 256 T3

Arg Asn Asn Gln Ala Ser Ile Leu Glu Arg Arg His Asp Val Leu Met
595 600 605

Phe Asn Gln Gly Trp Gln Gln Tyr Gly Asn Val Ile Asn Glu Ile Val
610 615 620

Thr Val Gly Leu Gly Tyr Lys Thr Ala Leu Arg Lys Ala Leu Tyr Thr
625 630 635 640

Phe Ala Glu Gly Phe Lys Asn Lys Asp Phe Lys Gly Ala Gly Val Ser
645 650 655

Val His Glu Thr Ala Glu Arg His Phe Tyr Arg Gln Ser Glu Leu Leu
660 665 670

Ile Pro Asp Val Leu Ala Asn Val Asn Phe Ser Gln Ala Asp Glu Val
675 680 685

Ile Ala Asp Leu Arg Asp Lys Leu His Gln Leu Cys Glu Met Leu Phe
690 695 700

Asn Gln Ser Val Ala Pro Tyr Ala His His Pro Lys Leu Ile Ser Thr
705 710 715 720

Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu Lys
725 730 735

Pro Gln Gly Gly Pro Ser Asn Gly
740

<210> 26
<211> 168
<212> ADN
<213> Secuencia artificial

<220>
<223> Plásmido objetivo

<400> 26

gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttg 60
ctttcgagcg ctagctctag aactagtctt cagcctaggc ctggttccga agctgtcttt 120
cgctgctgag ggtgacgata ccgcataggc ggcctttaac tcggatcc 168

<210> 27
<211> 163
<212> ADN
<213> Secuencia artificial

<220>
<223> Plásmido objetivo

<400> 27

ES 2 689 256 T3

		gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttt	60
		tcgagcgcta gctctagaac tagtcctcag cctaggcctc gttcaagctg tctttcgctg	120
		ctgaggggtga cgatcccgca taggcggcct ttaactcgga tcc	163
5	<210> 28 <211> 158 <212> ADN <213> Secuencia artificial		
10	<220> <223> Plásmido objetivo		
	<400> 28		
		gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttc	60
		gagcgctagc tctagaacta gtcctcagcc taggcctcga agctgtcttt cgctgctgag	120
		ggtgacgata ccgcataggc ggcctttaac tcggatcc	158
15	<210> 29 <211> 153 <212> ADN <213> Secuencia artificial		
20	<220> <223> Plásmido objetivo		
	<400> 29		
25		gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttg	60
		cgctagctct agaactagtc ctcagcctag gcctaagctg tctttcgctg ctgaggggtga	120
		cgatcccgca taggcggcct ttaactcgga tcc	153
30	<210> 30 <211> 148 <212> ADN <213> Secuencia artificial		
35	<220> <223> Plásmido objetivo		
	<400> 30		
		gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttg	60
		ctagctctag aactagtcct cagcctagga agctgtcttt cgctgctgag ggtgacgata	120
		ccgcataggc ggcctttaac tcggatcc	148
40	<210> 31 <211> 143 <212> ADN <213> Secuencia artificial		
45	<220> <223> Plásmido objetivo		

<400> 31	
	gaattcacaa cggtagagcaa gtcactgttg gcaagccagg atctgaacaa taccgtcttc 60
	tctagaacta gtcctcagcc taggaagctg tctttcgctg ctgaggggtga cgatcccga 120
	taggcggcct ttaactcgga tcc 143
5	<210> 32 <211> 41 <212> ADN <213> Secuencia artificial
10	<220> <223> Producto de escisión
15	<400> 32 ctggcgctag ctctagaact agtcctcagc ctaggcctaa g 41
20	<210> 33 <211> 41 <212> ADN <213> Secuencia artificial
25	<220> <223> Producto de escisión
30	<400> 33 cttaggccta ggctgaggac tagttctaga gctagcgcaa g 41
35	<210> 34 <211> 45 <212> ADN <213> Secuencia artificial
40	<220> <223> Llenado en ligamiento
45	<400> 34 ctggcgctag ctctagaact agctagtcct cagcctaggc ctaag 45
50	<210> 35 <211> 45 <212> ADN <213> Secuencia artificial
55	<220> <223> Llenado en ligamiento
	<400> 35 cttaggccta ggctgaggac tagctagttc tagagctagc gcaag 45
	<210> 36 <211> 1100 <212> ADN <213> <i>Homo sapiens</i>
	<400> 36

ES 2 689 256 T3

```

ggtggaacaa gatggattat caagtgtcaa gtccaatcta tgacatcaat tattatacat      60
cggagccctg ccaaaaaatc aatgtgaagc aaatcgagc cgcctcctg cctccgctct      120
actcactggt gttcatcttt ggttttgtgg gcaacatgct ggtcatcctc atcctgataa      180
actgcaaaaag gctgaagagc atgactgaca tctacctgct caacctggcc atctctgacc      240
tgtttttcct tcttactgtc cccttctggg ctcaactatgc tgccgccag tgggactttg      300
gaaatacaat gtgtcaactc ttgacagggc tctattttat aggcttcttc totggaatct      360
tcttcatcat cctcctgaca atcgataggt acctggctgt cgtccatgct gtgtttgctt      420
taaaagccag gacggtcacc tttggggtgg tgacaagtgt gatcacttgg gtggtggctg      480
tgtttgcgtc tctcccagga atcatcttta ccagatctca aaaagaaggt cttcattaca      540
cctgcagctc tcattttcca tacagtcagt atcaattctg gaagaatttc cagacattaa      600
agatagtcat cttggggctg gtccctgccg tgcttgtcat ggtcatctgc tactcgggaa      660
tcctaaaaac tctgcttcgg tgtcgaaatg agaagaagag gcacagggct gtgaggctta      720
tcttcaccat catgattgtt tattttctct tctgggctcc ctacaacatt gtccttctcc      780
tgaacacctt ccaggaattc tttggcctga ataattgcag tagctctaac aggttgacc      840
aagctatgca ggtgacagag actcctggga tgacgcactg ctgcatcaac cccatcatct      900
atgcctttgt cggggagaag ttcagaaact acctcttagt cttcttccaa aagcacattg      960
ccaaacgctt ctgcaaatgc tgttctatct tccagcaaga ggctcccgag cgagcaagct     1020
cagtttacac ccgatccact ggggagcagg aaatatctgt gggcttgtga cccggactca     1080
agtgggctgg tgaccagtc                                     1100

```

<210> 37
 <211> 424
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Matriz CRISPR red1

<400> 37

```

ccatggtaat acgactcact ataggagaa ttagctgatc ttaataata aggaaatgtt      60
acattaaggt tggtaggttg tttttatggg aaaaaatgct ttaagaacaa atgtatactt      120
ttagagagtt ccccgcgcca gcggggataa accgcaaaca cagcatggac gacagccagg      180
tacctagagt tccccgccc agcggggata aaccgcaaac acagcatgga cgacagccag      240
gtacctagag ttccccgccc cagcggggat aaaccgcaaa cacagcatgg acgacagcca      300
ggtacctaga gttccccgcg ccagcgggga taaaccgaaa acaaaaggct cagtcggaag      360
actgggcctt ttgttttaac cccttggggc ctctaaacgg gtcttgaggg gttttttggg      420
tacc                                     424

```

<210> 38
 <211> 424
 <212> ADN
 <213> Secuencia artificial

ES 2 689 256 T3

	<220>	
	<223> Matriz CRISPR red2	
	<400> 38	
5		
	ccatggtaat acgactcact atagggagaa ttagctgac ttttaataata aggaaatggt	60
	acattaaggt tgggtgggttg tttttatggg aaaaaatgct ttaagaacaa atgtatactt	120
	ttagagaggt ccccgcgcca gcggggataa accgtgtgat cacttgggtg gtggctgtgt	180
	ttgcgtgagt tccccgcgcc agcggggata aaccgtgtga tcacttgggt ggtggctgtg	240
	tttgcgtgag ttccccgcgc cagcggggat aaaccgtgtg atcacttggg tgggtggctgt	300
	gtttgcgtga gttccccgcg ccagcgggga taaaccgaaa acaaaaggct cagtcggaag	360
	actgggcctt ttgttttaac cccttggggc ctctaaacgg gtcttgaggg gttttttggg	420
	tacc	424
	<210> 39	
	<211> 43	
10	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> plásmido	
15	<400> 39	
	aaggatgcca gtgataagt gaatgccatg tgggctgtca aaa 43	
	<210> 40	
20	<211> 43	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
25	<223> mutante plásmido 1	
	<400> 40	
	aaggatgcca gtgataagt gaatgccatg tgggctgtca aaa 43	
30	<210> 41	
	<211> 19	
	<212> ADN	
	<213> Secuencia artificial	
35	<220>	
	<223> mutante plásmido 2	
	<400> 41	
	gccatgtggg ctgtcaaaa 19	
40	<210> 42	
	<211> 23	
	<212> ADN	
	<213> Secuencia artificial	
45	<220>	
	<223> mutante plásmido 3	
	<400> 42	

gaatgccatg tgggctgtca aaa 23

<210> 43

<211> 1604

<212> PRT

<213> *Streptomyces* sp. SPB78

<400> 43

```

Met  Pro  Asp  Gln  Leu  Asn  Ala  Pro  Thr  Pro  Leu  Gly  Asp  Arg  Leu  Thr
1              5              10              15

Gly  Ala  Val  Arg  Thr  Val  Trp  Ala  Lys  His  Asp  Arg  Asp  Thr  Gly  Lys
                20              25              30

Trp  Leu  Pro  Leu  Trp  Arg  His  Met  Thr  Asp  Ser  Ala  Ala  Val  Ala  Gly
          35              40              45

Leu  Leu  Trp  Asp  His  Trp  Leu  Pro  Arg  Asn  Ile  Lys  Asp  Leu  Ile  Ala
50              55              60

Glu  Pro  Leu  Pro  Gly  Gly  Val  Ala  Asp  Ala  Arg  Ser  Leu  Cys  Val  Trp
65              70              75              80

Leu  Ala  Gly  Thr  His  Asp  Ile  Gly  Lys  Ala  Thr  Pro  Ala  Phe  Ala  Cys
                85              90              95

Gln  Val  Asp  Glu  Leu  Ala  Gly  Val  Met  Thr  Ala  Ala  Gly  Leu  Asp  Met
                100              105              110

Arg  Thr  Ser  Lys  Gln  Leu  Gly  Glu  Asp  Arg  Arg  Met  Ala  Pro  His  Gly
          115              120              125

Leu  Ala  Gly  Gln  Val  Leu  Leu  Gln  Glu  Trp  Leu  Glu  Glu  Arg  Arg  Gly
130              135              140

Trp  Thr  His  Arg  Ala  Ser  Ala  Gln  Phe  Ala  Val  Val  Ala  Gly  Gly  His
145              150              155              160

```

ES 2 689 256 T3

His Gly Val Pro Pro Asp His Met Gln Leu His Asn Leu Asp Ala His
 165 170 175
 Pro Glu Leu Leu Arg Thr Gln Gly Leu Ala Glu Ala Gln Trp Arg Ala
 180 185 190
 Val Gln Asp Glu Leu Leu Asp Ala Cys Ala Leu Val Phe Gly Val Glu
 195 200 205
 Glu Arg Leu Asp Ala Trp Arg Thr Val Lys Leu Pro Gln Thr Val Gln
 210 215 220
 Val Leu Leu Thr Ala Thr Val Ile Val Ser Asp Trp Ile Ala Ser Asn
 225 230 235 240
 Pro Asp Leu Phe Pro Tyr Phe Pro Glu Glu His Pro Arg Glu Glu Ala
 245 250 255
 Glu Arg Val Ala Ala Ala Trp Gln Gly Leu Leu Leu Pro Ala Pro Trp
 260 265 270
 Glu Pro Glu Glu Pro Ser Ala Pro Ala Ala Glu Phe Tyr Ala Ser Arg
 275 280 285
 Phe Ala Leu Pro Pro Gly Ala Val Val Arg Pro Val Gln Glu Gln Ala
 290 295 300
 Leu Ala Met Ala Arg Asp Met Glu Arg Pro Gly Met Leu Ile Ile Glu
 305 310 315 320
 Ala Pro Met Gly Glu Gly Lys Thr Glu Ala Ala Leu Ala Val Ala Glu
 325 330 335
 Val Phe Ala Ala Arg Ser Gly Ala Gly Gly Cys Tyr Val Ala Leu Pro
 340 345 350
 Thr Met Ala Thr Ser Asn Ala Met Phe Pro Arg Leu Leu Arg Trp Leu
 355 360 365
 Asp Arg Leu Pro Arg Ala Asp Val Ser Gly Gly Arg Asp His Glu Gln
 370 375 380
 Arg Ser Val Leu Leu Ala His Ala Lys Ser Ala Leu Gln Glu Asp Tyr
 385 390 395 400
 Ala Thr Leu Met Arg Glu Ser His Arg Thr Ile Ala Ala Val Asp Ala
 405 410 415

ES 2 689 256 T3

Tyr Gly Asp Asp Ser Arg Pro Arg Lys Gly Arg Pro Ala Ala Asp Gly
 420 425 430
 Val Arg Arg Lys Ala Pro Ala Glu Leu Val Ala His Gln Trp Leu Arg
 435 440 445
 Gly Arg Lys Lys Gly Leu Leu Ala Ser Phe Ala Val Gly Thr Ile Asp
 450 455 460
 Gln Leu Leu Met Ala Gly Leu Lys Ser Arg His Leu Ala Leu Arg His
 465 470 475 480
 Leu Ala Met Ala Gly Lys Val Val Val Ile Asp Glu Val His Ala Tyr
 485 490 495
 Asp Thr Tyr Met Asn Ala Tyr Leu Asp Arg Val Leu Ala Trp Leu Gly
 500 505 510
 Glu Tyr Arg Val Pro Val Val Val Leu Ser Ala Thr Leu Pro Ala Arg
 515 520 525
 Arg Arg Gly Glu Leu Ala Ala Ala Tyr Thr Gly Glu Asp Ala Gln Ala
 530 535 540
 Leu Thr Glu Ala Thr Gly Tyr Pro Leu Leu Thr Ala Val Val Pro Gly
 545 550 555 560
 Arg Glu Ala Val Gln Phe Val Ala Ala Ala Ser Gly Arg Gly Ser Asp
 565 570 575
 Val Leu Leu Glu Lys Leu Asp Asp Asp Asp Glu Ala Leu Ala Asp Arg
 580 585 590
 Leu Asp Thr Asp Leu Ala Asp Gly Gly Cys Ala Leu Val Val Arg Asn
 595 600 605
 Thr Val Asp Arg Val Met Asp Thr Ala Ser Val Leu Arg Glu Arg Phe
 610 615 620
 Gly Ala Asp His Val Thr Val Ala His Ala Arg Phe Val Asp Leu Asp
 625 630 635 640
 Arg Ala Arg Lys Asp Ser Glu Leu Leu Ala Arg Phe Gly Pro Pro Asp
 645 650 655
 Pro Asp Gly Gly Ser Pro Gln Arg Pro Arg Asn Ala His Ile Val Val
 660 665 670

ES 2 689 256 T3

Ala	Ser	Gln	Val	Ala	Glu	Gln	Ser	Leu	Asp	Val	Asp	Phe	Asp	Leu	Leu	675	680	685
Val	Ser	Asp	Leu	Cys	Pro	Val	Asp	Leu	Leu	Leu	Gln	Arg	Met	Gly	Arg	690	695	700
Leu	His	Arg	His	Pro	Arg	Gly	Arg	Asp	Gln	Glu	Arg	Arg	Pro	Ala	Arg	705	710	715
Leu	Arg	Gln	Ala	Arg	Cys	Leu	Val	Thr	Gly	Val	Gly	Trp	Asp	Thr	Ser	725	730	735
Pro	Ala	Pro	Glu	Ala	Asp	Glu	Gly	Ser	Arg	Ala	Ile	Tyr	Gly	Ala	Tyr	740	745	750
Ser	Leu	Leu	Arg	Ser	Leu	Ala	Val	Leu	Ala	Pro	His	Leu	Gly	Thr	Ala	755	760	765
Gly	Ala	Ala	Gly	His	Pro	Leu	Arg	Leu	Pro	Glu	Asp	Ile	Ser	Pro	Leu	770	775	780
Val	Arg	Arg	Ala	Tyr	Gly	Glu	Glu	Asp	Pro	Cys	Pro	Pro	Glu	Trp	Glu	785	790	795
Pro	Val	Leu	Ala	Pro	Ala	Arg	Asp	Lys	Tyr	Arg	Thr	Ala	Arg	Glu	Arg	805	810	815
Gln	Ser	Gln	Lys	Ala	Glu	Val	Phe	Arg	Leu	Asp	Glu	Val	Arg	Lys	Ala	820	825	830
Gly	Arg	Pro	Leu	Ile	Gly	Trp	Ile	Asp	Ala	Gly	Val	Gly	Asp	Ala	Asp	835	840	845
Asp	Thr	Pro	Val	Gly	Arg	Ala	Gln	Val	Arg	Asp	Thr	Lys	Glu	Gly	Leu	850	855	860
Glu	Val	Leu	Val	Val	Arg	Arg	Arg	Ala	Asp	Gly	Ser	Leu	Cys	Thr	Leu	865	870	875
Pro	Trp	Leu	Asp	Lys	Gly	Arg	Gly	Gly	Leu	Glu	Leu	Pro	Val	Asp	Ala	885	890	895
Val	Pro	Ser	Ala	Leu	Ala	Ala	Arg	Ala	Val	Ala	Ala	Ser	Gly	Leu	Arg	900	905	910
Leu	Pro	Tyr	His	Phe	Thr	Ser	Ser	Pro	Gln	Thr	Leu	Asp	Arg	Thr	Leu			

ES 2 689 256 T3

915	920	925
Ala Glu Leu Glu Glu Leu Tyr Val Pro Ala Trp Gln Glu Lys Glu Ser 930 935 940		
His Trp Ile Ala Gly Glu Leu Ile Leu Ala Leu Asp Glu Glu Gly Arg 945 950 955 960		
Ala Ala Leu Ala Gly Gln Gln Leu Val Tyr Asn Pro Glu Glu Gly Leu 965 970 975		
Leu Val Ala Ser Ala Asp Ala Asn Thr Glu Ala Thr Ser Gly Arg Val 980 985 990		
Met Asp Gly Lys Pro Ser Ser Ala Gly Asp Gly Lys Pro Gly His Ala 995 1000 1005		
Ala Asp Gly Asn Arg Ala Arg Thr Thr Val Gly Gln Ser Pro Ala 1010 1015 1020		
Asp Arg Gln Thr His Gln Pro Pro Glu Gly Glu Arg His Pro Val 1025 1030 1035		
Pro Pro Ser Ala Ala Pro Pro Pro Ala Arg Pro Ser Phe Asp Leu 1040 1045 1050		
Thr Ser Arg Pro Trp Leu Pro Val Leu Leu Lys Asp Gly Ser Glu 1055 1060 1065		
Arg Glu Leu Ser Leu Pro Glu Val Phe Asp Gln Ala Arg Asp Ile 1070 1075 1080		
Arg Arg Leu Val Gly Asp Leu Pro Thr Gln Asp Phe Ala Leu Thr 1085 1090 1095		
Arg Met Leu Leu Ala Leu Leu Tyr Asp Ala Leu Ser Glu Pro Gly 1100 1105 1110		
Gly Asp Met Ala Pro Ala Asp Thr Asp Ala Trp Glu Glu Leu Trp 1115 1120 1125		
Leu Ser Gln Ser Ala Tyr Ala Ala Pro Val Ala Ala Tyr Leu His 1130 1135 1140		
Arg Tyr Arg Glu Arg Phe Asp Leu Leu His Pro Glu Ser Pro Phe 1145 1150 1155		

Phe	Gln	Thr	Pro	Gly	Leu	Arg	Thr	Ala	Lys	Asn	Glu	Val	Phe	Ser
1160						1165					1170			
Leu	Asn	Arg	Leu	Val	Ala	Asp	Val	Pro	Asn	Gly	Asp	Pro	Phe	Phe
1175						1180					1185			
Ser	Met	Arg	Arg	Pro	Gly	Val	Asp	Arg	Leu	Gly	Phe	Ala	Glu	Ala
1190						1195					1200			
Ala	Arg	Trp	Leu	Val	His	Ala	Gln	Ala	Tyr	Asp	Thr	Ser	Gly	Ile
1205						1210					1215			
Lys	Thr	Gly	Ala	Val	Gly	Asp	Pro	Arg	Val	Lys	Ala	Gly	Lys	Gly
1220						1225					1230			
Tyr	Pro	Gln	Gly	Pro	Ala	Trp	Ala	Gly	Asn	Leu	Gly	Gly	Val	Leu
1235						1240					1245			
Leu	Glu	Gly	Asp	Asn	Leu	His	Glu	Thr	Leu	Leu	Leu	Asn	Leu	Ile
1250						1255					1260			
Ala	Gly	Asp	Thr	Pro	Gly	Val	His	Ala	Ala	Glu	Val	Asp	Arg	Pro
1265						1270					1275			
Ala	Trp	Arg	Ala	Glu	Pro	Ser	Gly	Pro	Ala	Pro	Ala	Pro	Asp	Leu
1280						1285					1290			
Gly	Leu	Arg	Pro	Tyr	Gly	Leu	Arg	Asp	Leu	Tyr	Thr	Trp	Gln	Ser
1295						1300					1305			
Arg	Arg	Ile	Arg	Leu	His	His	Asp	Ala	Asp	Gly	Val	His	Gly	Val
1310						1315					1320			
Val	Leu	Ala	Tyr	Gly	Asp	Ser	Leu	Glu	Pro	His	Asn	Arg	His	Gly
1325						1330					1335			
His	Glu	Pro	Met	Thr	Ser	Trp	Arg	Arg	Ser	Pro	Thr	Gln	Glu	Lys
1340						1345					1350			
Lys	Arg	Gln	Glu	Asn	Leu	Val	Tyr	Leu	Pro	Arg	Glu	His	Asp	Pro
1355						1360					1365			
Ser	Arg	Leu	Ala	Trp	Arg	Gly	Met	Asp	Gly	Leu	Leu	Ala	Gly	Arg
1370						1375					1380			
Glu	Thr	Gly	Ser	Ala	Gln	Gly	Pro	Asp	Gly	Ala	Asp	Arg	Leu	Ala
1385						1390					1395			

Pro	Lys	Val	Val	Gln	Trp	Ala	Ala	Gln	Leu	Thr	Thr	Glu	Gly	Leu
1400						1405					1410			
Leu	Pro	Arg	Gly	Tyr	Leu	Ile	Arg	Thr	Arg	Val	Ile	Gly	Ala	Arg
1415						1420					1425			
Tyr	Gly	Thr	Gln	Gln	Ser	Val	Ile	Asp	Glu	Val	Val	Asp	Asp	Gly
1430						1435					1440			
Val	Leu	Met	Pro	Ala	Val	Leu	Leu	His	Glu	Ala	Asp	Arg	Arg	Tyr
1445						1450					1455			
Gly	Asp	Lys	Ala	Val	Asp	Ala	Leu	His	Asp	Ala	Glu	Lys	Ala	Val
1460						1465					1470			
Gly	Ala	Leu	Ala	Gln	Leu	Ala	Ala	Asp	Leu	Ala	Leu	Ala	Val	Gly
1475						1480					1485			
Thr	Asp	Pro	Glu	Pro	Gly	Arg	Asn	Thr	Ala	Arg	Asp	Leu	Gly	Phe
1490						1495					1500			
Gly	Thr	Leu	Asp	Thr	His	Tyr	Arg	Arg	Trp	Leu	Arg	Glu	Leu	Gly
1505						1510					1515			
Gly	Thr	Ser	Asp	Pro	Glu	Glu	His	Arg	Asp	Arg	Trp	Lys	Gln	Glu
1520						1525					1530			
Val	Arg	Arg	Leu	Val	Ala	Glu	Leu	Gly	Glu	Arg	Leu	Leu	Asp	Gly
1535						1540					1545			
Ala	Gly	Pro	Ala	Ala	Trp	Glu	Gly	Arg	Leu	Val	Glu	Thr	Gly	Lys
1550						1555					1560			
Gly	Thr	Arg	Trp	Leu	Asn	Asp	Ala	Ala	Ala	Glu	Leu	Arg	Phe	Arg
1565						1570					1575			
Thr	Arg	Leu	Arg	Glu	Phe	Leu	Thr	Thr	Ala	Pro	Asp	Thr	Pro	Thr
1580						1585					1590			
Ser	Pro	Arg	Pro	Ala	Pro	Val	Glu	Ser	Pro	Ala				
1595						1600								

<210> 44
 <211> 1559
 <212> PRT
 <213> *Streptomyces griseus*

<400> 44

ES 2 689 256 T3

Met	Ser	Asn	Thr	Pro	Met	Ser	Arg	Asp	His	Pro	Glu	Ser	Leu	Ser	Ala	1	5	10	15
Tyr	Ala	Arg	Leu	Ser	Pro	Val	Ser	Arg	Thr	Ala	Trp	Gly	Lys	His	Asp	20	25	30	
Arg	Gln	Thr	Glu	Gln	Trp	Leu	Pro	Leu	Trp	Arg	His	Met	Ala	Asp	Ser	35	40	45	
Ala	Ala	Val	Ala	Glu	Arg	Leu	Trp	Asp	Gln	Trp	Val	Pro	Asp	Asn	Val	50	55	60	
Lys	Ala	Leu	Ile	Ala	Asp	Ala	Phe	Pro	Gln	Gly	Ala	Gln	Asp	Ala	Arg	65	70	75	80
Arg	Val	Ala	Val	Phe	Leu	Ala	Cys	Val	His	Asp	Ile	Gly	Lys	Ala	Thr	85	90	95	
Pro	Ala	Phe	Ala	Cys	Gln	Val	Asp	Gly	Leu	Ala	Asp	Arg	Met	Arg	Ala	100	105	110	
Ala	Gly	Leu	Ser	Met	Pro	Tyr	Leu	Lys	Gln	Phe	Gly	Leu	Asp	Arg	Arg	115	120	125	
Met	Ala	Pro	His	Gly	Leu	Ala	Gly	Gln	Leu	Leu	Leu	Gln	Glu	Trp	Leu	130	135	140	
Ala	Glu	Arg	Phe	Gly	Trp	Ser	Glu	Arg	Ala	Ser	Gly	Gln	Phe	Ala	Val	145	150	155	160
Val	Ala	Gly	Gly	His	His	Gly	Thr	Pro	Pro	Asp	His	Gln	His	Ile	His	165	170	175	
Asp	Leu	Gly	Leu	Arg	Pro	His	Leu	Leu	Arg	Thr	Ala	Gly	Glu	Ser	Gln	180	185	190	
Asp	Thr	Trp	Arg	Ser	Val	Gln	Asp	Glu	Leu	Met	Asp	Ala	Cys	Ala	Val	195	200	205	
Arg	Ala	Gly	Val	Gly	Gly	Arg	Phe	Gly	Ala	Trp	Arg	Ser	Val	Arg	Leu	210	215	220	
Pro	Gln	Pro	Val	Gln	Val	Val	Leu	Thr	Ala	Ile	Val	Ile	Val	Ser	Asp	225	230	235	240
Trp	Ile	Ala	Ser	Ser	Ser	Glu	Leu	Phe	Pro	Tyr	Asp	Pro	Ala	Ser	Trp	245	250	255	

ES 2 689 256 T3

Ser Pro Val Gly Pro Glu Gly Glu Gly Arg Arg Leu Thr Ala Ala Trp
 260 265 270
 Gly Gly Leu Asp Leu Pro Gly Pro Trp Arg Ala Asp Gln Pro Asp Cys
 275 280 285
 Thr Ala Ala Glu Leu Phe Gly Lys Arg Phe Asp Leu Pro Glu Gly Ala
 290 295 300
 Gly Val Arg Pro Val Gln Glu Glu Ala Val Arg Val Ala Gln Glu Leu
 305 310 315 320
 Pro Gly Pro Gly Leu Leu Ile Ile Glu Ala Pro Met Gly Glu Gly Lys
 325 330 335
 Thr Glu Ala Ala Phe Ala Ala Ala Glu Ile Leu Ala Ala Arg Thr Gly
 340 345 350
 Ala Gly Gly Cys Leu Val Ala Leu Pro Thr Arg Ala Thr Gly Asp Ala
 355 360 365
 Met Phe Pro Arg Leu Leu Arg Trp Leu Glu Arg Leu Pro Ser Asp Gly
 370 375 380
 Pro Arg Ser Val Val Leu Ala His Ala Lys Ala Ala Leu Asn Glu Val
 385 390 395 400
 Trp Ala Gly Met Thr Lys Ala Asp Arg Arg Lys Ile Thr Ala Val Asp
 405 410 415
 Leu Asp Ser Gln Val Glu Asp Val Ser Ser Ala Gly Gly Ala Arg Arg
 420 425 430
 Ala Asn Pro Ala Ser Leu His Ala His Gln Trp Leu Arg Gly Arg Lys
 435 440 445
 Lys Ala Leu Leu Ser Ser Phe Ala Val Gly Thr Val Asp Gln Val Leu
 450 455 460
 Phe Ala Gly Leu Lys Ser Arg His Leu Ala Leu Arg His Leu Ala Val
 465 470 475 480
 Ala Gly Lys Val Val Ile Val Asp Glu Val His Ala Tyr Asp Ala Tyr
 485 490 495
 Met Ser Ala Tyr Leu Asp Arg Val Leu Glu Trp Leu Ala Ala Tyr Arg

ES 2 689 256 T3

500					505					510					
Val	Pro	Val	Val	Met	Leu	Ser	Ala	Thr	Leu	Pro	Ala	His	Arg	Arg	Arg
		515					520					525			
Glu	Leu	Ala	Ala	Ala	Tyr	Ala	Gly	Glu	Glu	Thr	Pro	Glu	Leu	Ala	Asp
	530					535					540				
Ala	Leu	Ala	Leu	Pro	Asp	Asp	Ala	Tyr	Pro	Leu	Ile	Thr	Ala	Val	Ala
545					550					555					560
Pro	Gly	Gly	Leu	Val	Leu	Thr	Ala	Arg	Pro	Glu	Pro	Ala	Ser	Gly	Arg
				565					570					575	
Arg	Thr	Glu	Val	Val	Leu	Glu	Arg	Leu	Gly	Asp	Gly	Pro	Ala	Leu	Leu
			580					585						590	
Ala	Ala	Arg	Leu	Asp	Glu	Glu	Leu	Arg	Asp	Gly	Gly	Cys	Ala	Leu	Val
		595					600					605			
Val	Arg	Asn	Thr	Val	Asp	Arg	Val	Leu	Glu	Ala	Ala	Glu	His	Leu	Arg
	610					615						620			
Ala	His	Phe	Gly	Ala	Glu	Ala	Val	Thr	Val	Ala	His	Ser	Arg	Phe	Val
625					630					635					640
Ala	Ala	Asp	Arg	Ala	Arg	Asn	Asp	Thr	Val	Leu	Arg	Glu	Arg	Phe	Gly
				645					650					655	
Pro	Gly	Gly	Asp	Arg	Pro	Ala	Gly	Pro	His	Ile	Val	Val	Ala	Ser	Gln
			660					665					670		
Val	Val	Glu	Gln	Ser	Leu	Asp	Ile	Asp	Phe	Asp	Leu	Leu	Val	Thr	Asp
		675					680					685			
Leu	Ala	Pro	Val	Asp	Leu	Val	Leu	Gln	Arg	Met	Gly	Arg	Leu	His	Arg
	690					695					700				
His	Pro	Arg	Thr	Arg	Pro	Pro	Arg	Leu	Ser	Arg	Ala	Arg	Cys	Leu	Ile
705					710					715					720
Thr	Gly	Val	Glu	Asp	Trp	His	Ala	Glu	Arg	Pro	Val	Pro	Val	Arg	Gly
				725					730					735	
Ser	Leu	Ala	Val	Tyr	Gln	Gly	Pro	His	Thr	Leu	Leu	Arg	Ala	Leu	Ala
			740					745					750		

Val Leu Gly Pro His Leu Asp Gly Val Pro Leu Val Leu Pro Asp His
 755 760 765
 Ile Ser Pro Leu Val Gln Ala Ala Tyr Asp Glu Arg Pro Val Gly Pro
 770 775 780
 Ala His Trp Ala Pro Val Leu Asp Glu Ala Arg Arg Gln Tyr Leu Thr
 785 790 795 800
 Arg Leu Ala Glu Lys Arg Glu Arg Ala Asp Val Phe Arg Leu Gly Pro
 805 810 815
 Val Arg Arg Pro Gly Arg Pro Leu Phe Gly Trp Leu Asp Gly Asn Ala
 820 825 830
 Gly Asp Ala Asp Asp Ser Arg Thr Gly Arg Ala Gln Val Arg Asp Ser
 835 840 845
 Glu Glu Ser Leu Glu Val Leu Val Val Gln Arg Arg Ala Asp Gly Arg
 850 855 860
 Leu Thr Thr Val Ser Trp Leu Asp Gly Gly Arg Gly Gly Leu Asp Leu
 865 870 875 880
 Pro Glu His Ala Pro Pro Pro Pro Arg Ala Ala Glu Val Val Ala Ala
 885 890 895
 Cys Ala Leu Thr Leu Pro Arg Ser Leu Thr His Pro Gly Val Ile Asp
 900 905 910
 Arg Thr Ile Ala Glu Leu Glu Arg Phe Val Val Pro Ala Trp Gln Val
 915 920 925
 Lys Glu Cys Pro Trp Leu Ala Gly Glu Leu Leu Leu Val Leu Asp Glu
 930 935 940
 Asp Cys Gln Thr Arg Leu Ser Gly Leu Glu Val His Tyr Ser Thr Asp
 945 950 955 960
 Gln Gly Leu Arg Val Gly Ser Val Gly Thr Arg Ser Thr Asn Arg Ala
 965 970 975
 Lys Gly Leu Glu Ala Val Ser Val Ala Ser Phe Asp Leu Val Ser Arg
 980 985 990
 Pro Trp Leu Pro Val Gln Tyr Glu Asp Gly Ala Thr Gly Glu Leu Ser
 995 1000 1005

ES 2 689 256 T3

Leu Arg 1010	Glu Val Phe Ala Arg 1015	Ala Gly Glu Val Arg 1020	Arg Leu Val
Gly Asp 1025	Leu Pro Thr Gln Glu 1030	Leu Ala Leu Leu Arg 1035	Leu Leu Leu
Ala Ile 1040	Leu Tyr Asp Ala Tyr 1045	Asp Glu Ala Pro Gly 1050	Arg Ser Gly
Gly Ala 1055	Pro Ala Gln Leu Glu 1060	Asp Trp Glu Ala Leu 1065	Trp Asp Glu
Pro Asp 1070	Ser Phe Ala Val Val 1075	Ala Gly Tyr Leu Asp 1080	Arg His Arg
Asp Arg 1085	Phe Asp Leu Leu His 1090	Pro Glu Arg Pro Phe 1095	Phe Gln Val
Ala Gly 1100	Leu His Thr Gln Lys 1105	His Glu Val Ala Ser 1110	Leu Asn Arg
Ile Val 1115	Ala Asp Val Pro Asn 1120	Gly Glu Ala Phe Phe 1125	Ser Met Arg
Arg Pro 1130	Gly Val His Arg Leu 1135	Gly Leu Ala Glu Ala 1140	Ala Arg Trp
Leu Val 1145	His Thr His Ala Tyr 1150	Asp Ala Ser Gly Ile 1155	Lys Ser Gly
Met Glu 1160	Gly Asp Ala Arg Val 1165	Lys Gly Gly Lys Val 1170	Tyr Pro Gln
Gly Val 1175	Gly Trp Val Gly Gly 1180	Leu Gly Gly Val Phe 1185	Ala Glu Gly
Ala Ser 1190	Leu Arg Glu Thr Leu 1195	Leu Leu Asn Leu Ile 1200	Pro Thr Asp
Glu Asp 1205	Ile Leu Thr Ser Glu 1210	Pro Lys Ala Asp Leu 1215	Pro Val Trp
Arg Arg 1220	Glu Thr Pro Pro Gly 1225	Pro Gly Val Val Glu 1230	Gly Asp Pro
Ser Ala 1235	Pro Arg Pro Ala Gly 1240	Pro Arg Asp Leu Tyr 1245	Thr Trp Gln

ES 2 689 256 T3

Ser	Arg	Arg	Leu	Leu	Leu	His	Thr	Glu	Gly	Ser	Asp	Ala	Ile	Gly
1250						1255					1260			
Val	Val	Leu	Gly	Tyr	Gly	Asp	Pro	Leu	Ser	Pro	Ala	Asn	Arg	Gln
1265						1270					1275			
Lys	Thr	Glu	Pro	Met	Thr	Gly	Trp	Arg	Arg	Ser	Pro	Ala	Gln	Glu
1280						1285					1290			
Lys	Lys	Leu	Gly	Arg	Pro	Leu	Val	Tyr	Leu	Pro	Arg	Gln	His	Asp
1295						1300					1305			
Pro	Gly	Arg	Ala	Ala	Trp	Arg	Gly	Leu	Ala	Ser	Leu	Leu	Tyr	Pro
1310						1315					1320			
Gln	Gly	Glu	Asp	Gly	Asp	Thr	Thr	Gly	Arg	Gly	Thr	Asp	Arg	Ser
1325						1330					1335			
Arg	Pro	Ala	Gly	Ile	Val	Arg	Trp	Leu	Ala	Leu	Leu	Ser	Thr	Glu
1340						1345					1350			
Gly	Val	Leu	Pro	Lys	Gly	Ser	Leu	Ile	Arg	Thr	Arg	Leu	Val	Gly
1355						1360					1365			
Ala	Val	Tyr	Gly	Thr	Gln	Gln	Ser	Val	Val	Asp	Asp	Val	Val	Asp
1370						1375					1380			
Asp	Ser	Ile	Ala	Leu	Pro	Val	Val	Leu	Leu	His	Gln	Asp	Arg	Arg
1385						1390					1395			
Leu	His	Gly	Ala	Val	Ala	Val	Asp	Ala	Val	Ala	Asp	Ala	Glu	Arg
1400						1405					1410			
Ala	Val	Ser	Ala	Leu	Gly	His	Leu	Ala	Gly	Asn	Leu	Ala	Arg	Ala
1415						1420					1425			
Ser	Gly	Ser	Glu	Ala	Gly	Pro	Ala	Thr	Ala	Thr	Ala	Arg	Asp	Gln
1430						1435					1440			
Gly	Phe	Gly	Ala	Leu	Asp	Gly	Pro	Tyr	Arg	Arg	Trp	Leu	Val	Asp
1445						1450					1455			
Leu	Ala	Glu	Asp	Thr	Asp	Leu	Glu	Arg	Ala	Arg	Ala	Ala	Trp	Arg
1460						1465					1470			
Asp	Thr	Val	Arg	Leu	Val	Val	Leu	Gly	Ile	Gly	Arg	Glu	Leu	Leu

ES 2 689 256 T3

1475		1480		1485
Asp Ala Ala Gly Arg Ala Ala Ala Glu Gly Arg Val Ile Glu Leu				
1490		1495		1500
Pro Gly Val Gly Lys Arg Trp Ile Asp Ser Ser Arg Ala Asp Leu				
1505		1510		1515
Trp Phe Arg Thr Arg Ile Asn Arg Val Leu Pro Arg Pro Leu Pro				
1520		1525		1530
Glu Ala His Ala Pro Thr Ala Asp Ile His Ala Gly His Ala Val				
1535		1540		1545
Arg Ala Asp Glu Ala Leu Ser Glu Glu Thr Val				
1550		1555		

<210> 45
 <211> 1540
 <212> PRT
 <213> *Catenulispora acidiphila*

<400> 45

ES 2 689 256 T3

Met	Phe	Asn	Val	Gly	Ser	Thr	Arg	Cys	Trp	Gly	Asp	Gly	Gly	Leu	Arg
1				5					10					15	
Asn	Ala	Ala	Glu	Asp	Leu	Ser	Ala	Ala	Thr	Arg	Ser	Ala	Trp	Ala	Lys
			20					25					30		
Ser	Asp	Pro	Asp	Ser	Gly	Gln	Ser	Leu	Ser	Leu	Ile	Arg	His	Leu	Ala
		35					40					45			
Asp	Ser	Ala	Ala	Ile	Ala	Glu	His	Leu	Trp	Asp	Gln	Trp	Leu	Pro	Asp
	50					55					60				
His	Val	Lys	Ser	Leu	Ile	Ala	Glu	Gly	Leu	Pro	Glu	Gly	Leu	Val	Asp
65					70					75					80
Gly	Arg	Thr	Leu	Ala	Val	Trp	Leu	Ala	Gly	Thr	His	Asp	Ile	Gly	Lys
			85						90					95	
Leu	Thr	Pro	Ala	Phe	Ala	Cys	Gln	Cys	Glu	Pro	Leu	Ala	Gln	Ala	Met
			100					105					110		
Arg	Glu	Cys	Gly	Leu	Asp	Met	Pro	Thr	Arg	Thr	Gln	Phe	Gly	Asp	Asp
		115					120					125			
Arg	Arg	Val	Ala	Pro	His	Gly	Leu	Ala	Gly	Gln	Val	Leu	Leu	Arg	Glu

ES 2 689 256 T3

130															
Trp	Leu	Met	Glu	Arg	His	Gly	Trp	Ser	Gly	Arg	Ser	Ala	Asp	Ala	Phe
145					150					155					160
Thr	Val	Ile	Ala	Gly	Gly	His	His	Gly	Val	Pro	Pro	Ser	Tyr	Ser	Gln
				165					170					175	
Leu	His	Asp	Leu	Asp	Ala	Tyr	Pro	Glu	Leu	Leu	Arg	Thr	Pro	Gly	Ala
			180					185					190		
Ser	Glu	Gly	Ile	Trp	Lys	Ser	Ser	Gln	His	Glu	Leu	Leu	Asp	Ala	Cys
		195					200						205		
Ala	Val	Met	Thr	Gly	Ala	Ser	Ser	Arg	Leu	Ala	His	Trp	Arg	Gly	Leu
	210					215					220				
Arg	Leu	Ser	Gln	Gln	Ala	Gln	Val	Leu	Leu	Thr	Gly	Leu	Val	Ile	Val
225					230					235					240
Ala	Asp	Trp	Ile	Ala	Ser	Asn	Thr	Asp	Leu	Phe	Pro	Tyr	Pro	Ala	Leu
				245					250					255	
Gly	Thr	Gly	Glu	Ala	Ala	Ile	Asp	Pro	Gly	Lys	Arg	Val	Glu	Leu	Ala
			260					265					270		
Trp	Arg	Gly	Leu	Glu	Leu	Pro	Ala	Pro	Trp	Ala	Pro	Lys	Tyr	Leu	Met
		275					280					285			
Pro	Gly	Met	Gln	Gly	Leu	Leu	Ala	Ser	Arg	Phe	Gly	Leu	Pro	Ala	Asp
	290					295					300				
Ala	Gln	Leu	Arg	Pro	Val	Gln	Gln	Met	Ala	Val	Gln	Leu	Ala	Ser	Ala
305					310					315					320
Asn	Ala	Ala	Pro	Gly	Leu	Leu	Val	Ile	Glu	Ala	Pro	Met	Gly	Glu	Gly
				325					330					335	
Lys	Thr	Glu	Ala	Ala	Leu	Leu	Ala	Ala	Glu	Ile	Leu	Ala	Ala	Arg	Ser
			340					345					350		
Gly	Ala	Gly	Gly	Val	Phe	Leu	Ala	Leu	Pro	Thr	Gln	Ala	Thr	Ser	Asn
		355					360					365			
Ala	Met	Phe	Ala	Arg	Val	Val	Asn	Trp	Leu	Arg	Gln	Val	Pro	Arg	Glu
	370					375					380				

ES 2 689 256 T3

Gly Val Ala Ser Val His Leu Ala His Gly Lys Ala Ala Leu Asp Asp
 385 390 395 400
 Ala Phe Ala Ser Phe Leu Arg Ala Ala Pro Arg Leu Thr Ser Ile Asp
 405 410 415
 Ala Asp Gly Tyr Ala Gly Glu Ala Asn Val Arg Arg Asp Arg Arg Ala
 420 425 430
 Gly Ser Ala Asp Met Val Ala His Gln Trp Leu Arg Gly Arg Lys Lys
 435 440 445
 Gly Ile Leu Ser Pro Phe Val Val Gly Thr Ile Asp Gln Leu Leu Phe
 450 455 460
 Thr Gly Leu Lys Ser Arg His Leu Ala Leu Arg His Leu Ala Val Ala
 465 470 475 480
 Gly Lys Val Val Val Ile Asp Glu Val His Ala Tyr Asp Ala Tyr Met
 485 490 495
 Ser Val Tyr Leu Glu Arg Val Leu Ser Trp Leu Gly Ala Tyr Arg Val
 500 505 510
 Pro Val Val Leu Leu Ser Ala Thr Leu Pro Ala Asp Arg Arg Gln Ala
 515 520 525
 Leu Val Glu Ala Tyr Gly Gly Ile Thr Ser Glu Ala Leu Arg Asp Ala
 530 535 540
 Arg Glu Ala Tyr Pro Val Leu Thr Ala Val Thr Ile Gly Ala Pro Ala
 545 550 555 560
 Gln Ala Val Gly Thr Glu Pro Ala Glu Gly Arg Arg Val Asp Val Asn
 565 570 575
 Val Glu Ala Phe Asp Asp Asp Leu Gly Arg Leu Ala Asp Arg Leu Glu
 580 585 590
 Ala Glu Leu Val Asp Gly Gly Cys Ala Leu Ile Ile Arg Asn Thr Val
 595 600 605
 Gly Arg Val Leu Gln Thr Ala Gln Gln Leu Arg Glu Arg Phe Gly Ala
 610 615 620
 Gly Gln Val Thr Val Ala His Ser Arg Phe Ile Asp Leu Asp Arg Ala
 625 630 635 640

Arg Lys Asp Ala Asp Leu Leu Ala Arg Phe Gly His Asp Gly Ala Arg
 645 650 655
 Pro Arg Arg His Ile Val Val Ala Ser Gln Val Ala Glu Gln Ser Leu
 660 665 670
 Asp Ile Asp Phe Asp Leu Leu Val Thr Asp Leu Ala Pro Ile Asp Leu
 675 680 685
 Val Leu Gln Arg Met Gly Arg Val His Arg His His Arg Gly Gly Pro
 690 695 700
 Glu Gln Ser Glu Arg Pro Pro Ser Leu Arg Thr Ala Arg Cys Leu Val
 705 710 715 720
 Thr Gly Val Asp Trp Ala Gly Ile Pro Ser Ala Pro Ile Ala Gly Ser
 725 730 735
 Val Ala Val Tyr Gly Leu His Pro Leu Leu Arg Ser Leu Ala Val Leu
 740 745 750
 Gln Pro Tyr Leu Thr Gly Ser Ala Leu Thr Leu Pro Gly Asp Ile Asn
 755 760 765
 Pro Leu Val Gln Cys Ala Tyr Ala Gln Ser Phe Val Ala Pro Thr Gly
 770 775 780
 Trp Gly Glu Ala Met Asp Ala Ala Gln Ala Glu His Met Ala His Ile
 785 790 795 800
 Val Gln Gln Arg Glu Gly Ala Met Ala Phe Cys Leu Asp Glu Val Arg
 805 810 815
 Gly Pro Gly Arg Ser Leu Ile Gly Trp Ile Asp Gly Gly Val Gly Asp
 820 825 830
 Ala Asp Asp Thr Arg Ala Gly Arg Ala Gln Val Arg Asp Ser Pro Glu
 835 840 845
 Thr Ile Glu Val Leu Val Val Gln Arg Gly Ser Asp Gly Val Leu Arg
 850 855 860
 Thr Leu Pro Trp Leu Asp Arg Gly Arg Gly Gly Leu Glu Leu Pro Thr
 865 870 875 880
 Glu Ala Val Pro Pro Pro Arg Ala Ala Arg Ala Ala Ala Ala Ser Ala
 885 890 895

ES 2 689 256 T3

Leu Arg Leu Pro Gly Leu Phe Ala Lys Pro Trp Met Phe Asp Arg Val
 900 905 910
 Leu Arg Glu Leu Glu Arg Glu Tyr His Glu Ala Trp Gln Ala Lys Glu
 915 920 925
 Ser Ser Trp Leu Gln Gly Glu Leu Leu Leu Val Leu Asp Glu Glu Cys
 930 935 940
 Arg Thr Val Leu Ala Gly Tyr Glu Leu Ser Tyr Asn Pro Asp Asp Gly
 945 950 955 960
 Leu Glu Met Val Met Pro Gly Glu Pro His Ala Ala Val Val Arg Asp
 965 970 975
 Lys Glu Ala Ser Asp Asp Lys Thr Ala Ser Phe Asp Leu Thr Ser Ala
 980 985 990
 Pro Trp Leu Pro Val Leu Tyr Ala Asp Gly Met Gln Gly Val Leu Ser
 995 1000 1005
 Leu Arg Asp Val Phe Ala Gln Ser Asn Leu Ile Arg Arg Leu Val
 1010 1015 1020
 Gly Asp Leu Pro Thr Gln Asp Phe Ala Leu Leu Arg Leu Leu Leu
 1025 1030 1035
 Ala Val Leu Tyr Asp Ala Val Asp Gly Pro Arg Asp Gly Gln Asp
 1040 1045 1050
 Trp Glu Asp Leu Trp Thr Ser Asp Asp Pro Phe Ala Ala Val Pro
 1055 1060 1065
 Ala Tyr Leu Asp Ser His Arg Glu Arg Phe Asp Leu Leu His Pro
 1070 1075 1080
 Ala Thr Pro Phe Tyr Gln Val Pro Gly Leu Gln Thr Ala Lys Gly
 1085 1090 1095
 Glu Val Gly Pro Leu Asn Lys Ile Val Ala Asp Val Pro Asp Gly
 1100 1105 1110
 Asp Pro Phe Leu Thr Met Arg Met Pro Gly Val Glu Gln Leu Ser
 1115 1120 1125
 Phe Ala Glu Ala Ala Arg Trp Leu Val His Thr Gln Ala Phe Asp

ES 2 689 256 T3

1130	1135	1140
Thr Ser Gly Ile Lys Ser Gly Val Val Gly Asp Pro Lys Ala Val 1145 1150 1155		
Asn Gly Lys Arg Tyr Pro Gln Gly Val Ala Trp Leu Gly Asn Leu 1160 1165 1170		
Gly Gly Val Phe Ala Glu Gly Asp Thr Leu Arg Gln Thr Leu Leu 1175 1180 1185		
Leu Asn Leu Ile Pro Ala Asp Thr Thr Asn Leu Gln Val Thr Ser 1190 1195 1200		
Ala Gln Asp Val Pro Ala Trp Arg Gly Thr Asn Gly Arg Ala Gly 1205 1210 1215		
Ser Asp His Ala Asp Ala Glu Pro Arg Val Pro Ala Gly Leu Arg 1220 1225 1230		
Asp Leu Tyr Thr Trp Gln Ser Arg Arg Ile Arg Leu Glu Tyr Asp 1235 1240 1245		
Thr Arg Gly Val Thr Gly Ala Val Leu Thr Tyr Gly Asp Glu Leu 1250 1255 1260		
Thr Ala His Asn Lys His Gly Val Glu Pro Met Thr Gly Trp Arg 1265 1270 1275		
Arg Ser Lys Pro Gln Glu Lys Lys Leu Gly Leu Ser Thr Val Tyr 1280 1285 1290		
Met Pro Gln Gln His Asp Pro Thr Arg Ala Ala Trp Arg Gly Ile 1295 1300 1305		
Glu Ser Leu Leu Ala Gly Ser Ala Gly Ser Gly Ser Ser Gln Thr 1310 1315 1320		
Gly Glu Pro Ala Ser His Tyr Arg Pro Lys Ile Val Asp Trp Leu 1325 1330 1335		
Gly Glu Leu Ala His His Gly Asn Leu Pro Ser Arg Gly Leu Ile 1340 1345 1350		
Arg Val Arg Thr Ser Gly Ala Val Tyr Gly Thr Gln Gln Ser Ile 1355 1360 1365		

ES 2 689 256 T3

Ile Asp Glu Val Val Ser Asp Glu Leu Thr Met Ala Val Val Leu
1370 1375 1380

Leu His Glu Asp Asp Pro Arg Phe Gly Lys Ala Ala Val Thr Ala
1385 1390 1395

Val Lys Asp Ala Asp Ser Ala Val Ala Ala Leu Gly Asp Leu Ala
1400 1405 1410

Ser Asp Leu Ala Arg Ala Ala Gly Leu Asp Pro Glu Pro Glu Arg
1415 1420 1425

Val Thr Ala Arg Asp Arg Ala Phe Gly Ala Leu Asp Gly Pro Tyr
1430 1435 1440

Arg Arg Trp Leu Leu Asp Leu Gly Asn Ser Thr Asp Pro Ala Ala
1445 1450 1455

Met Arg Ala Val Trp Gln Gly Arg Val Tyr Asp Ile Ile Ala Val
1460 1465 1470

Gln Gly Gln Met Leu Leu Asp Ser Ala Gly Ser Ala Ala Ala Gln
1475 1480 1485

Gly Arg Met Val Lys Thr Thr Arg Gly Glu Arg Trp Met Asp Asp
1490 1495 1500

Ser Leu Ala Asp Leu Tyr Phe Lys Gly Arg Ile Ala Lys Ala Leu
1505 1510 1515

Ser Ser Arg Leu Gly Lys Lys Pro Thr Asp Pro Gly Glu Pro Val
1520 1525 1530

Gly Ile Gln Glu Asp Pro Ala
1535 1540

<210> 46
<211> 1407
<212> PRT
<213> Secuencia artificial

<220>
<223> Fusión Cas3-Cse1

<400> 46

Met Glu Pro Phe Lys Tyr Ile Cys His Tyr Trp Gly Lys Ser Ser Lys
1 5 10 15

Ser Leu Thr Lys Gly Asn Asp Ile His Leu Leu Ile Tyr His Cys Leu

ES 2 689 256 T3

20				25				30							
Asp	Val	Ala	Ala	Val	Ala	Asp	Cys	Trp	Trp	Asp	Gln	Ser	Val	Val	Leu
		35					40					45			
Gln	Asn	Thr	Phe	Cys	Arg	Asn	Glu	Met	Leu	Ser	Lys	Gln	Arg	Val	Lys
	50					55					60				
Ala	Trp	Leu	Leu	Phe	Phe	Ile	Ala	Leu	His	Asp	Ile	Gly	Lys	Phe	Asp
65					70					75					80
Ile	Arg	Phe	Gln	Tyr	Lys	Ser	Ala	Glu	Ser	Trp	Leu	Lys	Leu	Asn	Pro
			85						90					95	
Ala	Thr	Pro	Ser	Leu	Asn	Gly	Pro	Ser	Thr	Gln	Met	Cys	Arg	Lys	Phe
			100					105					110		
Asn	His	Gly	Ala	Ala	Gly	Leu	Tyr	Trp	Phe	Asn	Gln	Asp	Ser	Leu	Ser
		115					120					125			
Glu	Gln	Ser	Leu	Gly	Asp	Phe	Phe	Ser	Phe	Phe	Asp	Ala	Ala	Pro	His
	130					135					140				
Pro	Tyr	Glu	Ser	Trp	Phe	Pro	Trp	Val	Glu	Ala	Val	Thr	Gly	His	His
145					150					155					160
Gly	Phe	Ile	Leu	His	Ser	Gln	Asp	Gln	Asp	Lys	Ser	Arg	Trp	Glu	Met
				165					170					175	
Pro	Ala	Ser	Leu	Ala	Ser	Tyr	Ala	Ala	Gln	Asp	Lys	Gln	Ala	Arg	Glu
			180					185					190		
Glu	Trp	Ile	Ser	Val	Leu	Glu	Ala	Leu	Phe	Leu	Thr	Pro	Ala	Gly	Leu
		195					200					205			
Ser	Ile	Asn	Asp	Ile	Pro	Pro	Asp	Cys	Ser	Ser	Leu	Leu	Ala	Gly	Phe
	210					215					220				
Cys	Ser	Leu	Ala	Asp	Trp	Leu	Gly	Ser	Trp	Thr	Thr	Thr	Asn	Thr	Phe
225					230					235					240
Leu	Phe	Asn	Glu	Asp	Ala	Pro	Ser	Asp	Ile	Asn	Ala	Leu	Arg	Thr	Tyr
				245					250					255	
Phe	Gln	Asp	Arg	Gln	Gln	Asp	Ala	Ser	Arg	Val	Leu	Glu	Leu	Ser	Gly
			260					265					270		

ES 2 689 256 T3

Leu Val Ser Asn Lys Arg Cys Tyr Glu Gly Val His Ala Leu Leu Asp
 275 280 285
 Asn Gly Tyr Gln Pro Arg Gln Leu Gln Val Leu Val Asp Ala Leu Pro
 290 295 300
 Val Ala Pro Gly Leu Thr Val Ile Glu Ala Pro Thr Gly Ser Gly Lys
 305 310 315 320
 Thr Glu Thr Ala Leu Ala Tyr Ala Trp Lys Leu Ile Asp Gln Gln Ile
 325 330 335
 Ala Asp Ser Val Ile Phe Ala Leu Pro Thr Gln Ala Thr Ala Asn Ala
 340 345 350
 Met Leu Thr Arg Met Glu Ala Ser Ala Ser His Leu Phe Ser Ser Pro
 355 360 365
 Asn Leu Ile Leu Ala His Gly Asn Ser Arg Phe Asn His Leu Phe Gln
 370 375 380
 Ser Ile Lys Ser Arg Ala Ile Thr Glu Gln Gly Gln Glu Glu Ala Trp
 385 390 395 400
 Val Gln Cys Cys Gln Trp Leu Ser Gln Ser Asn Lys Lys Val Phe Leu
 405 410 415
 Gly Gln Ile Gly Val Cys Thr Ile Asp Gln Val Leu Ile Ser Val Leu
 420 425 430
 Pro Val Lys His Arg Phe Ile Arg Gly Leu Gly Ile Gly Arg Ser Val
 435 440 445
 Leu Ile Val Asp Glu Val His Ala Tyr Asp Thr Tyr Met Asn Gly Leu
 450 455 460
 Leu Glu Ala Val Leu Lys Ala Gln Ala Asp Val Gly Gly Ser Val Ile
 465 470 475 480
 Leu Leu Ser Ala Thr Leu Pro Met Lys Gln Lys Gln Lys Leu Leu Asp
 485 490 495
 Thr Tyr Gly Leu His Thr Asp Pro Val Glu Asn Asn Ser Ala Tyr Pro
 500 505 510
 Leu Ile Asn Trp Arg Gly Val Asn Gly Ala Gln Arg Phe Asp Leu Leu
 515 520 525

ES 2 689 256 T3

Ala	His	Pro	Glu	Gln	Leu	Pro	Pro	Arg	Phe	Ser	Ile	Gln	Pro	Glu	Pro	530	535	540	
Ile	Cys	Leu	Ala	Asp	Met	Leu	Pro	Asp	Leu	Thr	Met	Leu	Glu	Arg	Met	545	550	555	560
Ile	Ala	Ala	Ala	Asn	Ala	Gly	Ala	Gln	Val	Cys	Leu	Ile	Cys	Asn	Leu	565	570	575	
Val	Asp	Val	Ala	Gln	Val	Cys	Tyr	Gln	Arg	Leu	Lys	Glu	Leu	Asn	Asn	580	585	590	
Thr	Gln	Val	Asp	Ile	Asp	Leu	Phe	His	Ala	Arg	Phe	Thr	Leu	Asn	Asp	595	600	605	
Arg	Arg	Glu	Lys	Glu	Asn	Arg	Val	Ile	Ser	Asn	Phe	Gly	Lys	Asn	Gly	610	615	620	
Lys	Arg	Asn	Val	Gly	Arg	Ile	Leu	Val	Ala	Thr	Gln	Val	Val	Glu	Gln	625	630	635	640
Ser	Leu	Asp	Val	Asp	Phe	Asp	Trp	Leu	Ile	Thr	Gln	His	Cys	Pro	Ala	645	650	655	
Asp	Leu	Leu	Phe	Gln	Arg	Leu	Gly	Arg	Leu	His	Arg	His	His	Arg	Lys	660	665	670	
Tyr	Arg	Pro	Ala	Gly	Phe	Glu	Ile	Pro	Val	Ala	Thr	Ile	Leu	Leu	Pro	675	680	685	
Asp	Gly	Glu	Gly	Tyr	Gly	Arg	His	Glu	His	Ile	Tyr	Ser	Asn	Val	Arg	690	695	700	
Val	Met	Trp	Arg	Thr	Gln	Gln	His	Ile	Glu	Glu	Leu	Asn	Gly	Ala	Ser	705	710	715	720
Leu	Phe	Phe	Pro	Asp	Ala	Tyr	Arg	Gln	Trp	Leu	Asp	Ser	Ile	Tyr	Asp	725	730	735	
Asp	Ala	Glu	Met	Asp	Glu	Pro	Glu	Trp	Val	Gly	Asn	Gly	Met	Asp	Lys	740	745	750	
Phe	Glu	Ser	Ala	Glu	Cys	Glu	Lys	Arg	Phe	Lys	Ala	Arg	Lys	Val	Leu	755	760	765	
Gln	Trp	Ala	Glu	Glu	Tyr	Ser	Leu	Gln	Asp	Asn	Asp	Glu	Thr	Ile	Leu	770	775	780	

ES 2 689 256 T3

Ala	Val	Thr	Arg	Asp	Gly	Glu	Met	Ser	Leu	Pro	Leu	Leu	Pro	Tyr	Val	785	790	795	800
Gln	Thr	Ser	Ser	Gly	Lys	Gln	Leu	Leu	Asp	Gly	Gln	Val	Tyr	Glu	Asp	805	810	815	
Leu	Ser	His	Glu	Gln	Gln	Tyr	Glu	Ala	Leu	Ala	Leu	Asn	Arg	Val	Asn	820	825	830	
Val	Pro	Phe	Thr	Trp	Lys	Arg	Ser	Phe	Ser	Glu	Val	Val	Asp	Glu	Asp	835	840	845	
Gly	Leu	Leu	Trp	Leu	Glu	Gly	Lys	Gln	Asn	Leu	Asp	Gly	Trp	Val	Trp	850	855	860	
Gln	Gly	Asn	Ser	Ile	Val	Ile	Thr	Tyr	Thr	Gly	Asp	Glu	Gly	Met	Thr	865	870	875	880
Arg	Val	Ile	Pro	Ala	Asn	Pro	Lys	Gly	Asp	Pro	Thr	Asn	Arg	Ala	Lys	885	890	895	
Gly	Leu	Glu	Ala	Val	Ser	Val	Ala	Ser	Met	Asn	Leu	Leu	Ile	Asp	Asn	900	905	910	
Trp	Ile	Pro	Val	Arg	Pro	Arg	Asn	Gly	Gly	Lys	Val	Gln	Ile	Ile	Asn	915	920	925	
Leu	Gln	Ser	Leu	Tyr	Cys	Ser	Arg	Asp	Gln	Trp	Arg	Leu	Ser	Leu	Pro	930	935	940	
Arg	Asp	Asp	Met	Glu	Leu	Ala	Ala	Leu	Ala	Leu	Leu	Val	Cys	Ile	Gly	945	950	955	960
Gln	Ile	Ile	Ala	Pro	Ala	Lys	Asp	Asp	Val	Glu	Phe	Arg	His	Arg	Ile	965	970	975	
Met	Asn	Pro	Leu	Thr	Glu	Asp	Glu	Phe	Gln	Gln	Leu	Ile	Ala	Pro	Trp	980	985	990	
Ile	Asp	Met	Phe	Tyr	Leu	Asn	His	Ala	Glu	His	Pro	Phe	Met	Gln	Thr	995	1000	1005	
Lys	Gly	Val	Lys	Ala	Asn	Asp	Val	Thr	Pro	Met	Glu	Lys	Leu	Leu		1010	1015	1020	
Ala	Gly	Val	Ser	Gly	Ala	Thr	Asn	Cys	Ala	Phe	Val	Asn	Gln	Pro					

ES 2 689 256 T3

1025		1030		1035
Gly Gln Gly Glu Ala Leu Cys	Gly Gly Cys Thr Ala	Ile Ala Leu		
1040	1045	1050		
Phe Asn Gln Ala Asn Gln Ala	Pro Gly Phe Gly Gly	Gly Phe Lys		
1055	1060	1065		
Ser Gly Leu Arg Gly Gly Thr	Pro Val Thr Thr Phe	Val Arg Gly		
1070	1075	1080		
Ile Asp Leu Arg Ser Thr Val	Leu Leu Asn Val Leu	Thr Leu Pro		
1085	1090	1095		
Arg Leu Gln Lys Gln Phe Pro	Asn Glu Ser His Thr	Glu Asn Gln		
1100	1105	1110		
Pro Thr Trp Ile Lys Pro Ile	Lys Ser Asn Glu Ser	Ile Pro Ala		
1115	1120	1125		
Ser Ser Ile Gly Phe Val Arg	Gly Leu Phe Trp Gln	Pro Ala His		
1130	1135	1140		
Ile Glu Leu Cys Asp Pro Ile	Gly Ile Gly Lys Cys	Ser Cys Cys		
1145	1150	1155		
Gly Gln Glu Ser Asn Leu Arg	Tyr Thr Gly Phe Leu	Lys Glu Lys		
1160	1165	1170		
Phe Thr Phe Thr Val Asn Gly	Leu Trp Pro His Pro	His Ser Pro		
1175	1180	1185		
Cys Leu Val Thr Val Lys Lys	Gly Glu Val Glu Glu	Lys Phe Leu		
1190	1195	1200		
Ala Phe Thr Thr Ser Ala Pro	Ser Trp Thr Gln Ile	Ser Arg Val		
1205	1210	1215		
Val Val Asp Lys Ile Ile Gln	Asn Glu Asn Gly Asn	Arg Val Ala		
1220	1225	1230		
Ala Val Val Asn Gln Phe Arg	Asn Ile Ala Pro Gln	Ser Pro Leu		
1235	1240	1245		
Glu Leu Ile Met Gly Gly Tyr	Arg Asn Asn Gln Ala	Ser Ile Leu		
1250	1255	1260		

ES 2 689 256 T3

Glu Arg	Arg His Asp Val Leu	Met Phe Asn Gln Gly	Trp Gln Gln
1265	1270	1275	
Tyr Gly	Asn Val Ile Asn Glu	Ile Val Thr Val Gly	Leu Gly Tyr
1280	1285	1290	
Lys Thr	Ala Leu Arg Lys Ala	Leu Tyr Thr Phe Ala	Glu Gly Phe
1295	1300	1305	
Lys Asn	Lys Asp Phe Lys Gly	Ala Gly Val Ser Val	His Glu Thr
1310	1315	1320	
Ala Glu	Arg His Phe Tyr Arg	Gln Ser Glu Leu Leu	Ile Pro Asp
1325	1330	1335	
Val Leu	Ala Asn Val Asn Phe	Ser Gln Ala Asp Glu	Val Ile Ala
1340	1345	1350	
Asp Leu	Arg Asp Lys Leu His	Gln Leu Cys Glu Met	Leu Phe Asn
1355	1360	1365	
Gln Ser	Val Ala Pro Tyr Ala	His His Pro Lys Leu	Ile Ser Thr
1370	1375	1380	
Leu Ala	Leu Ala Arg Ala Thr	Leu Tyr Lys His Leu	Arg Glu Leu
1385	1390	1395	
Lys Pro	Gln Gly Gly Pro Ser	Asn Gly	
1400	1405		

<210> 47

<211> 49

5 <212> ADN

<213> Secuencia artificial

<220>

10 <223> Escisión de secuencia CASCADA

<400> 47

ccgtcttgcg ctagctctag aactagtcct cagcctaggc ctaagctgt 49

REIVINDICACIONES

1. Una proteína de fusión artificial de una subunidad proteica Cse1 asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) de tipo I y una endonucleasa FokI.
- 5 2. La proteína de fusión según la reivindicación 1, en la que la proteína de fusión comprende además una señal de localización nuclear.
3. Una molécula de ácido nucleico que codifica una proteína de fusión según la reivindicación 1 o 2.
4. Un vector de expresión que comprende una molécula de ácido nucleico según la reivindicación 3.
5. Un complejo proteico Cascada que comprende una proteína de fusión según la reivindicación 1 o 2.
- 10 6. Un complejo ribonucleoproteico que comprende la proteína de fusión según la reivindicación 1 o 2 y una molécula de ARN CRISPR (ARNcr).
7. El complejo ribonucleoproteico según la reivindicación 6, que comprende además una subunidad proteica Cas6, una subunidad proteica Cas5, una subunidad proteica Cse2 y una subunidad proteica Cas7.
- 15 8. Una célula eucariota que comprende un complejo ribonucleoproteico según la reivindicación 6 o la reivindicación 7, en la que la célula es distinta de una célula germinal humana y en la que la célula no forma parte del cuerpo humano o animal.
9. Un procedimiento de modificación, visualización, activación de la transcripción o represión de la transcripción de un ácido nucleico objetivo *in vitro* que comprende poner en contacto el ácido nucleico objetivo con un complejo ribonucleoproteico según la reivindicación 6 o la reivindicación 7.
- 20 10. Un procedimiento según la reivindicación 9, en el que el procedimiento de modificación es la unión y/o la escisión del ácido nucleico objetivo.
11. Un procedimiento según la reivindicación 9 o la reivindicación 10, en el que el ácido nucleico objetivo es un ADN bicatenario (ADNbc).

Figura 1

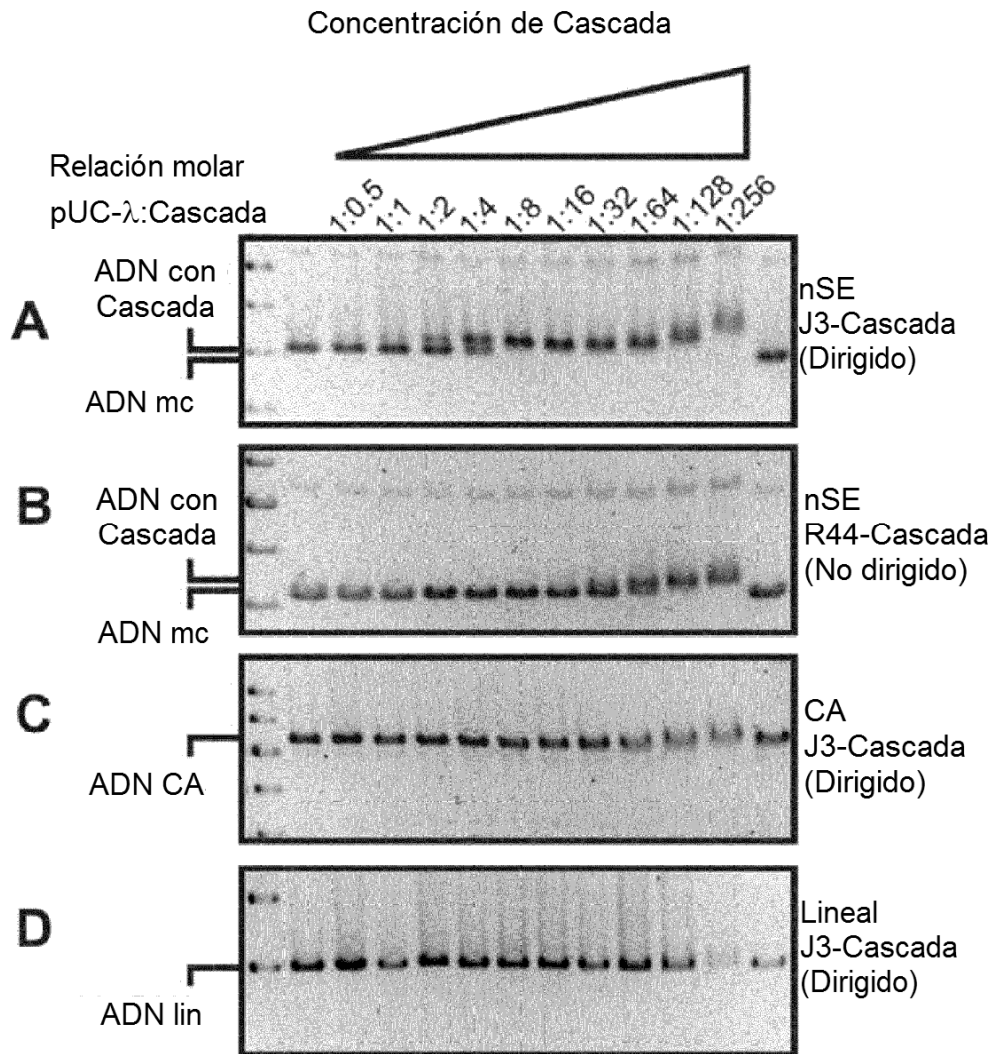


Figura 1 continuación

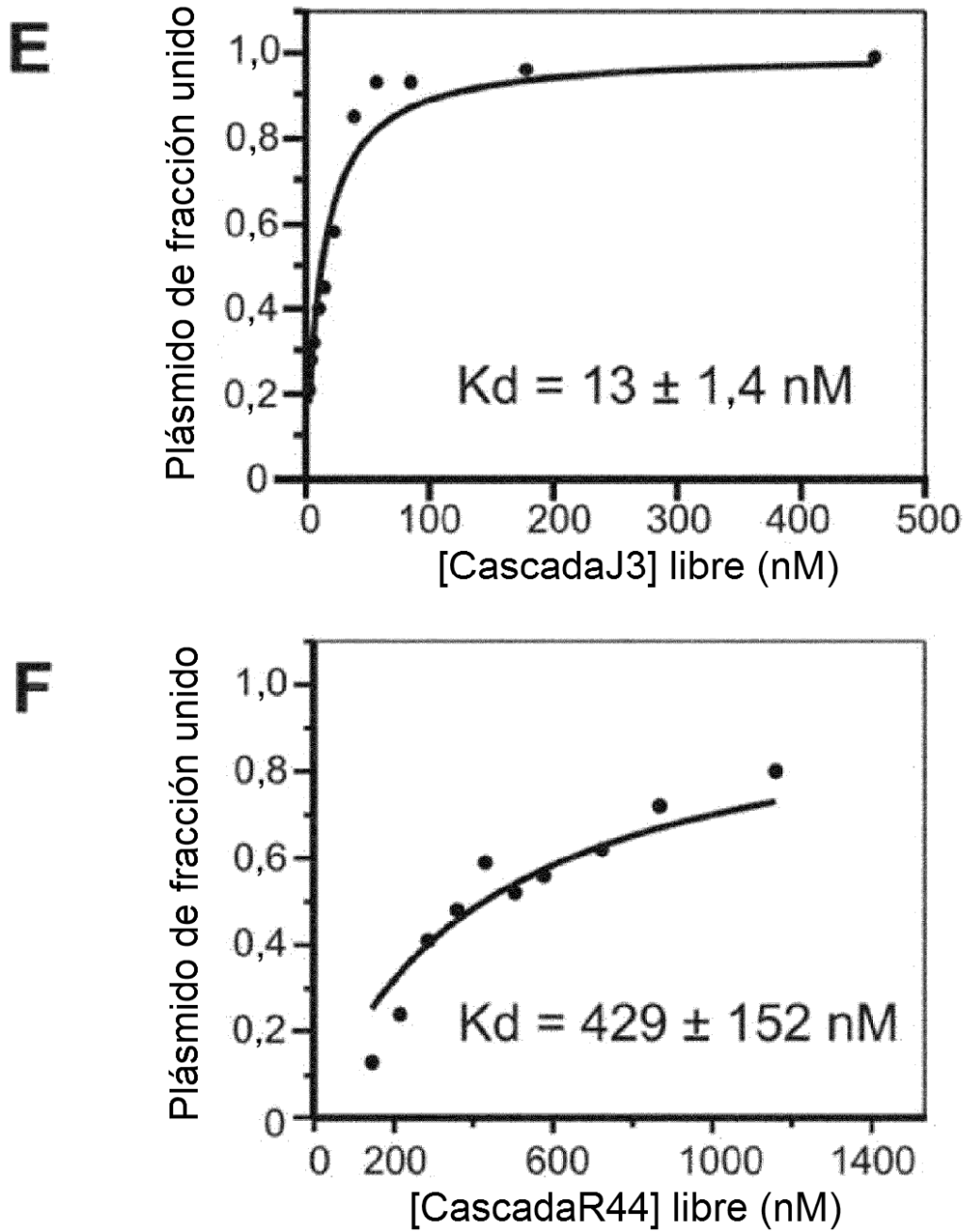


Figura 1 continuación

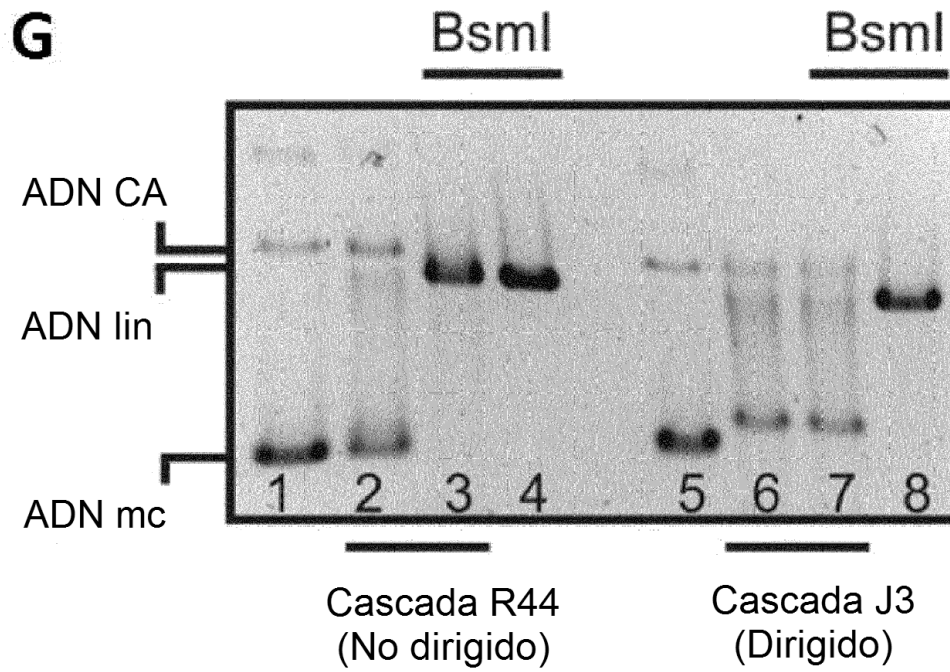


Figura 1 continuación

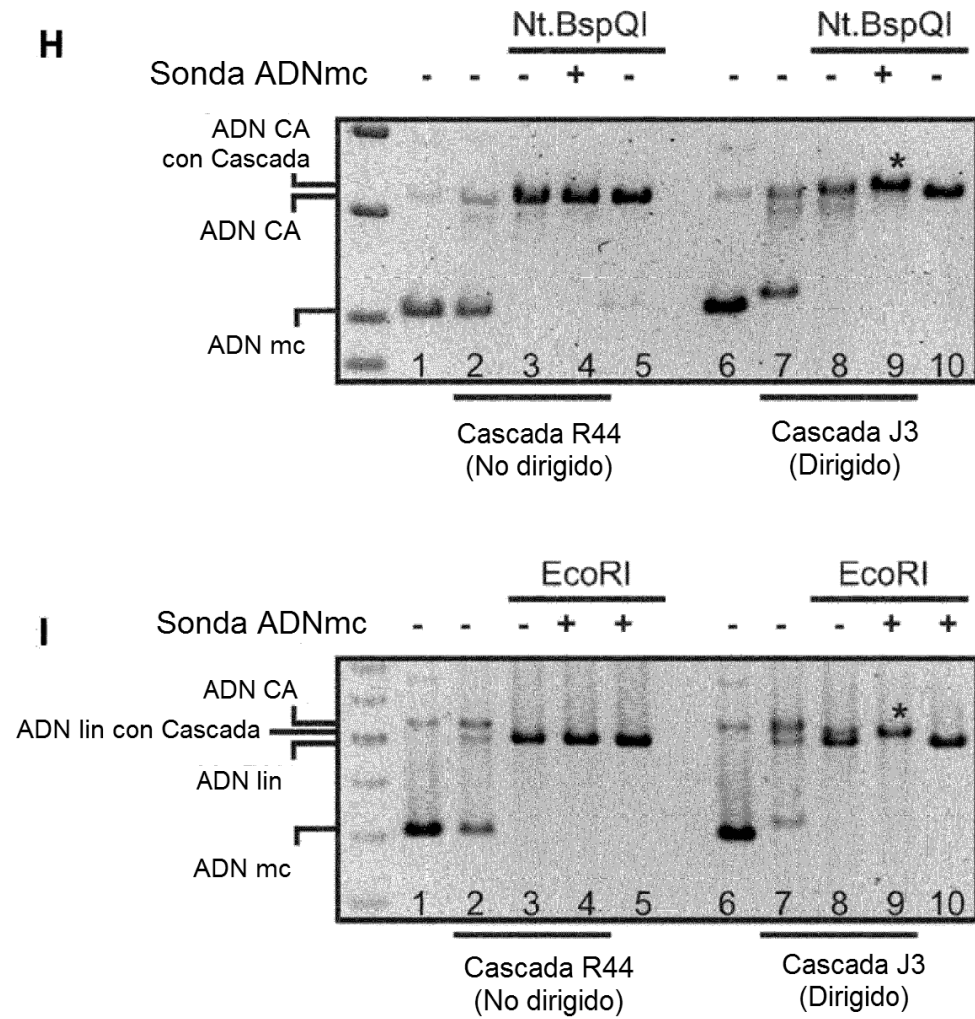


Figura 2

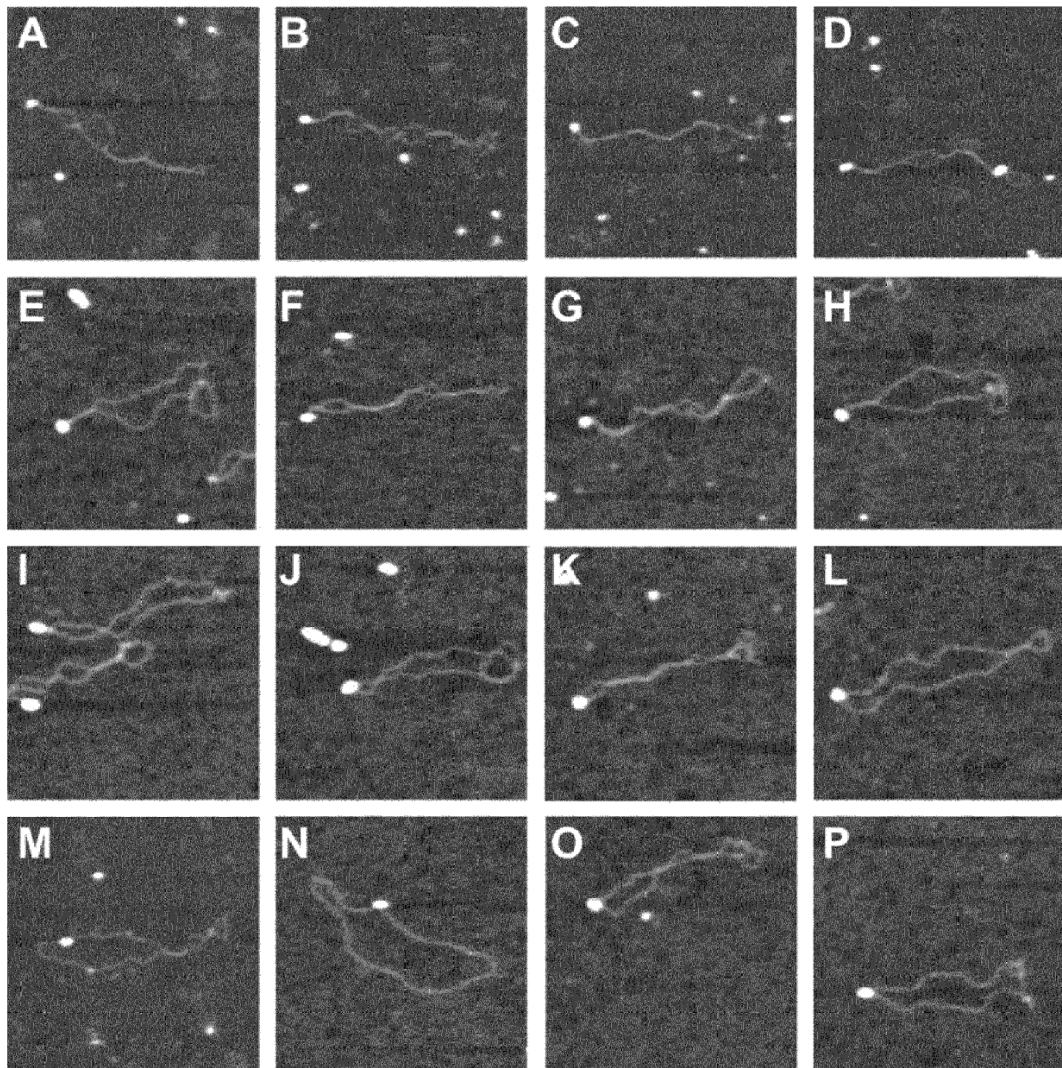


Figura 3

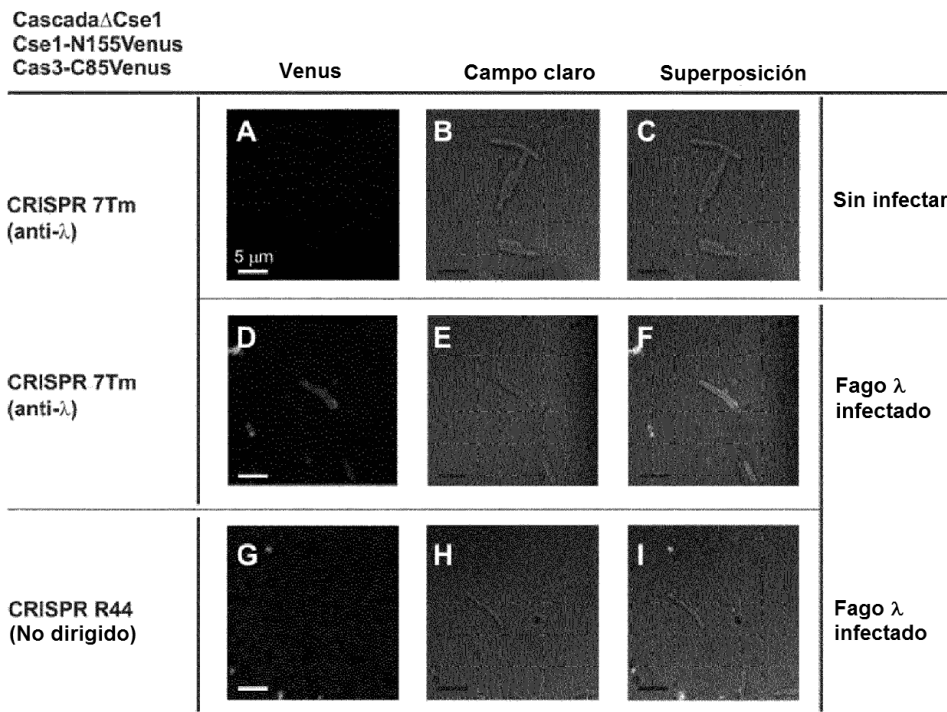


Figura 3 continuación

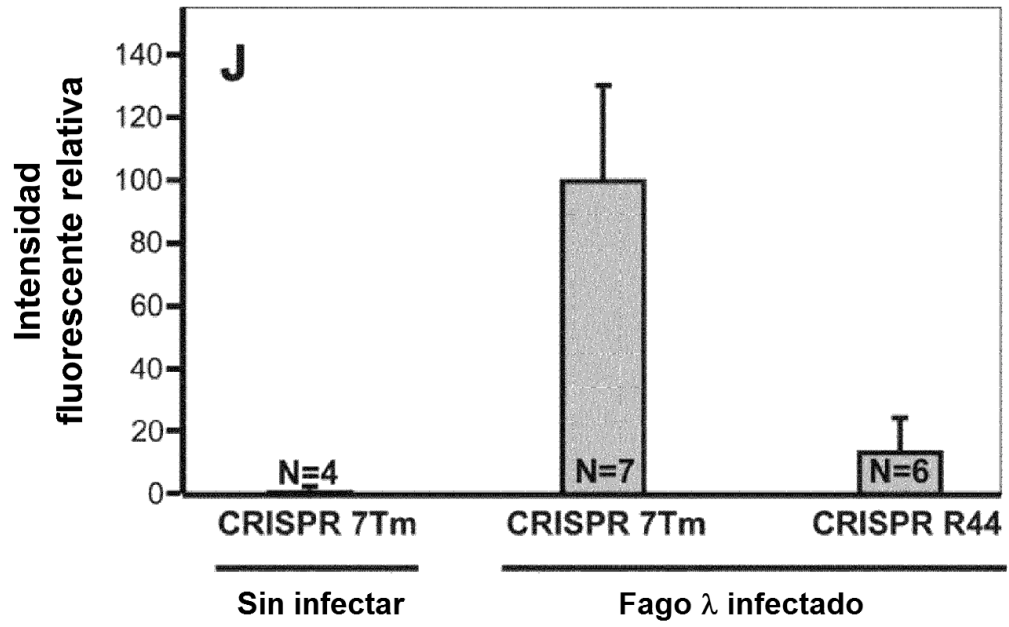


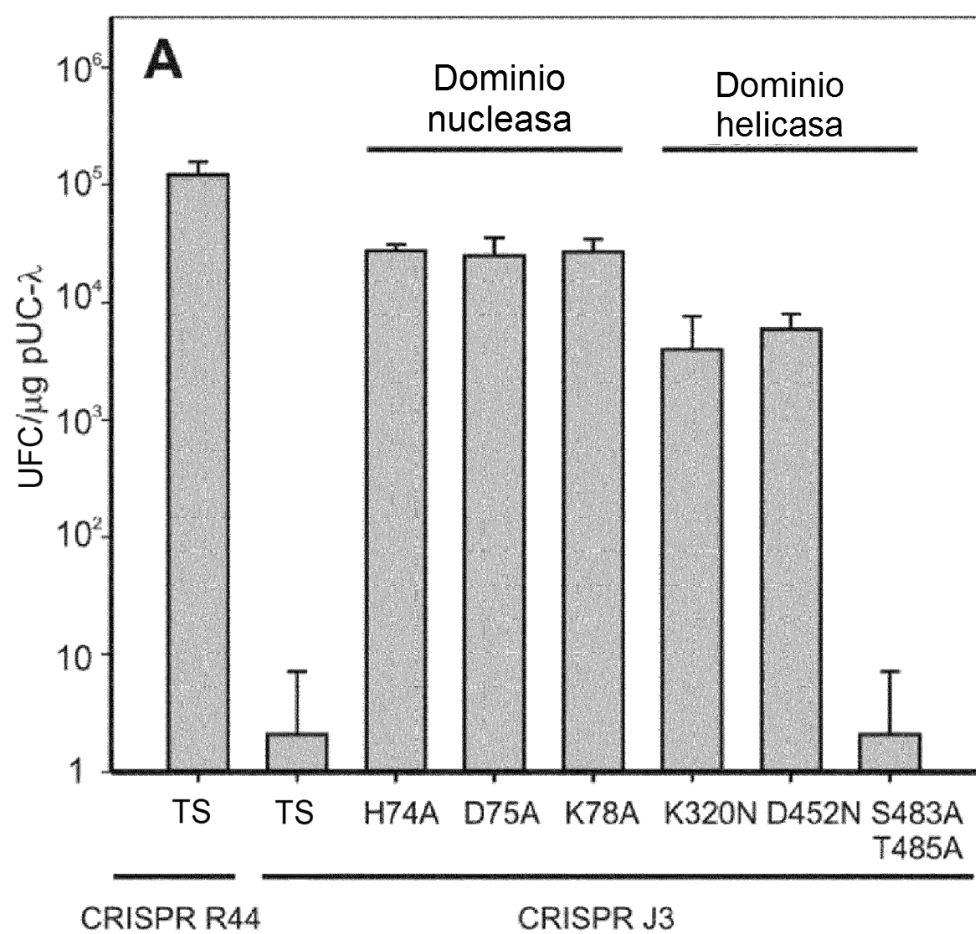
Figura 4

Figura 4 continuación

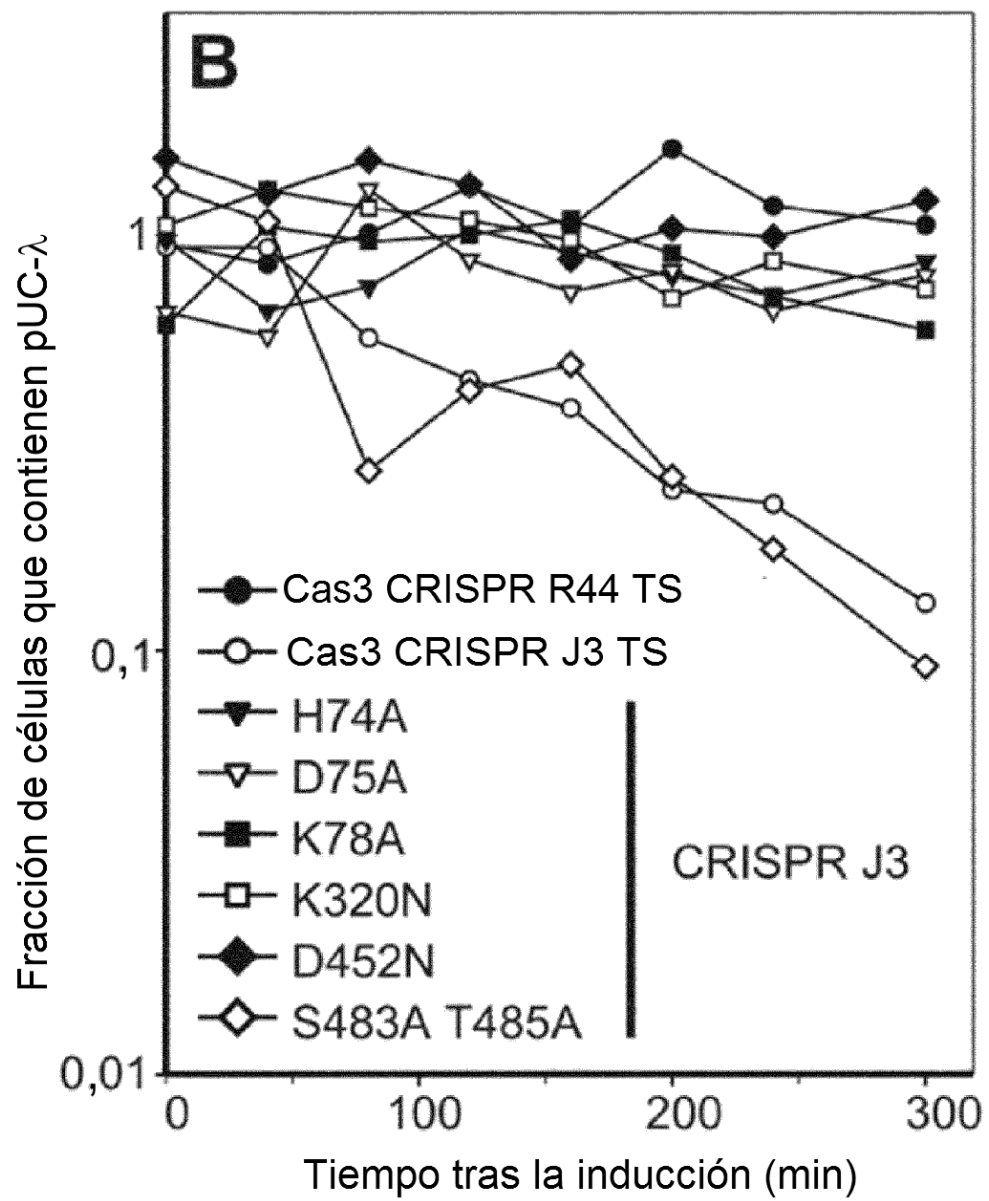


Figura 5

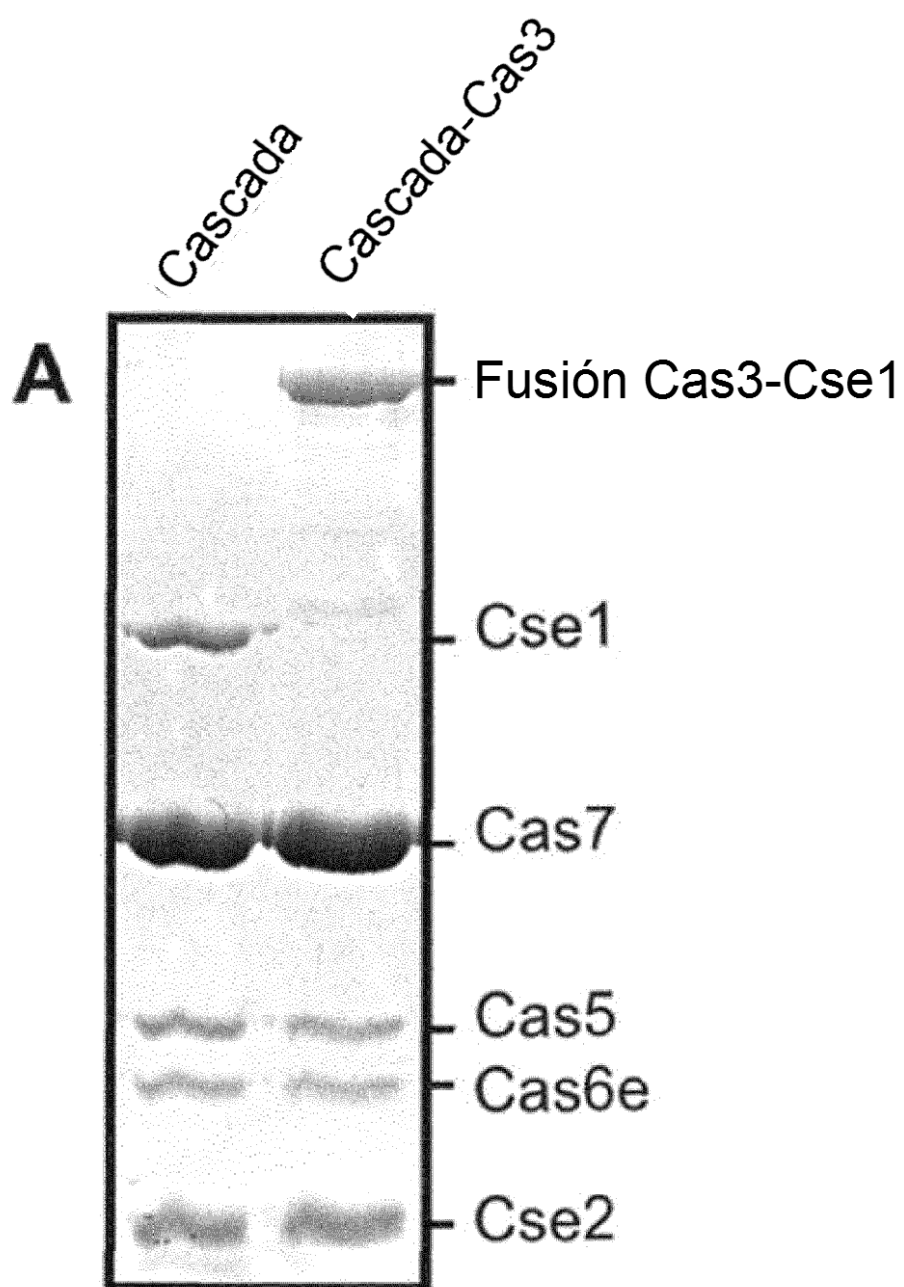


Figura 5 continuación

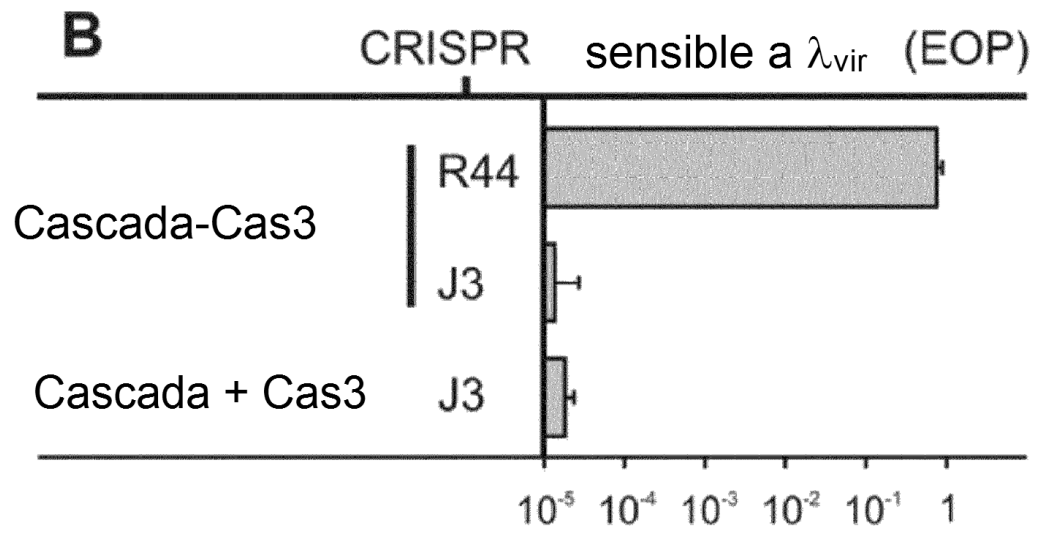


Figura 5 continuación

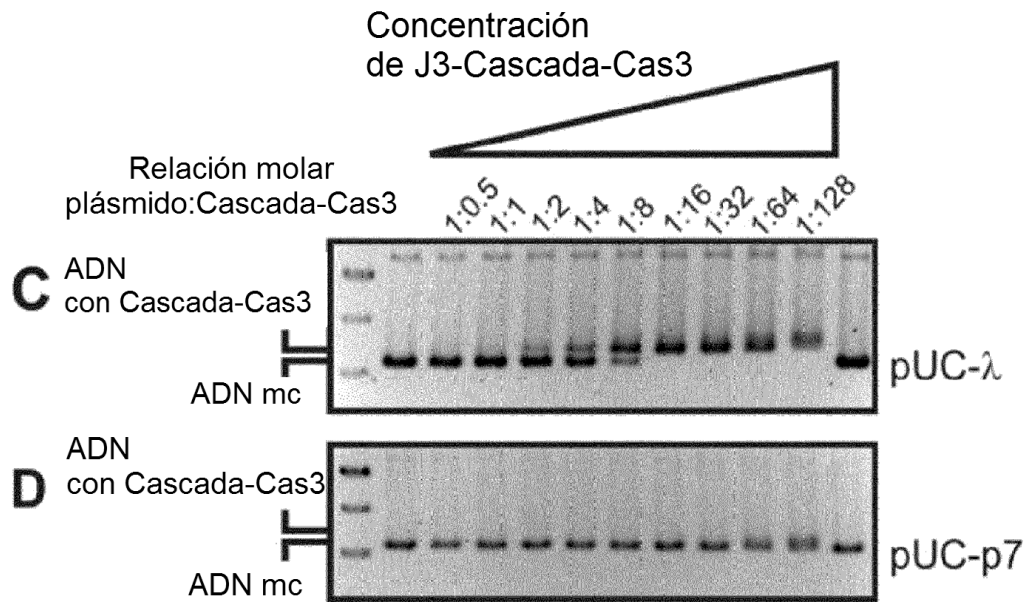


Figura 5 continuación

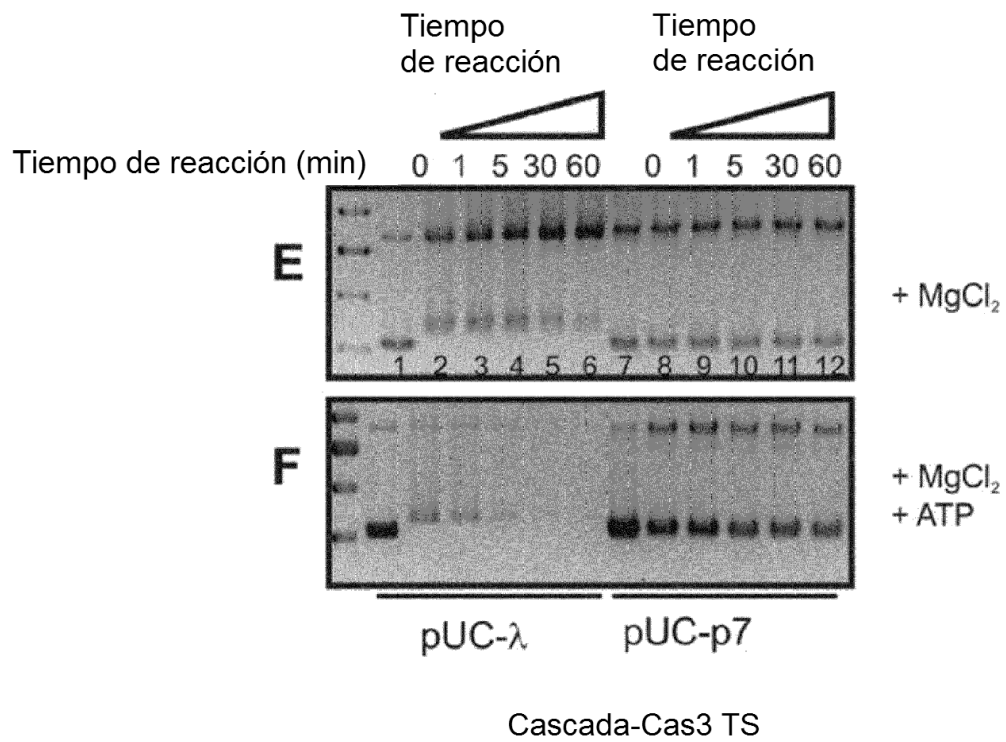


Figura 5 continuación

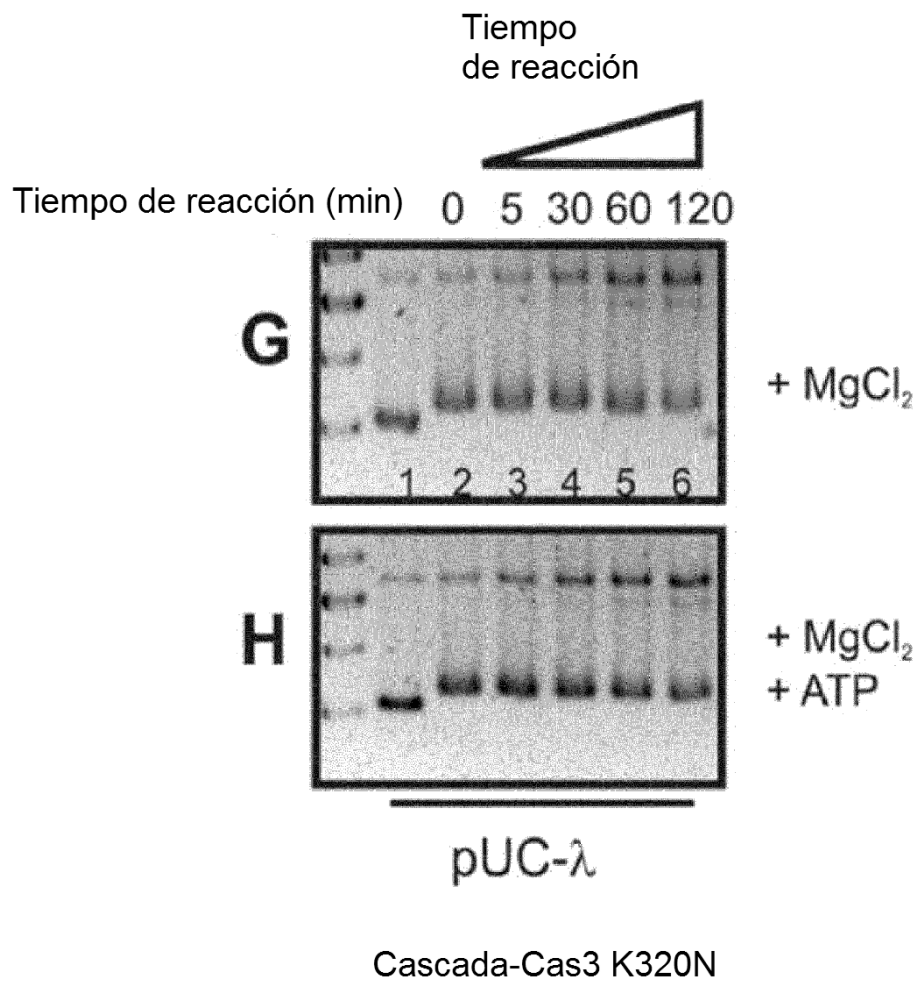
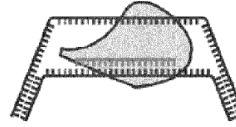


Figura 6

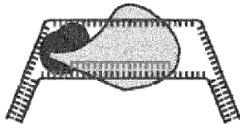
1: Barrido del protoespaciador



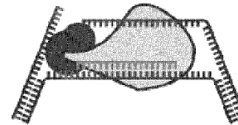
2: Unión del protoespaciador



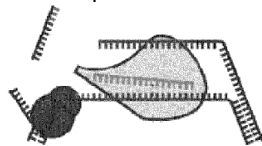
3: Formación del complejo Cas-Cascada3



4: Escotamiento por dominio Cas HD



5a: Actividad de la exo nucleasa dependiente de ATP de Cas3



5b: Disociación de Cascada



6: Degradación completa del plásmido por Cas3

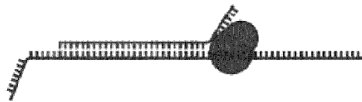


Figura 7

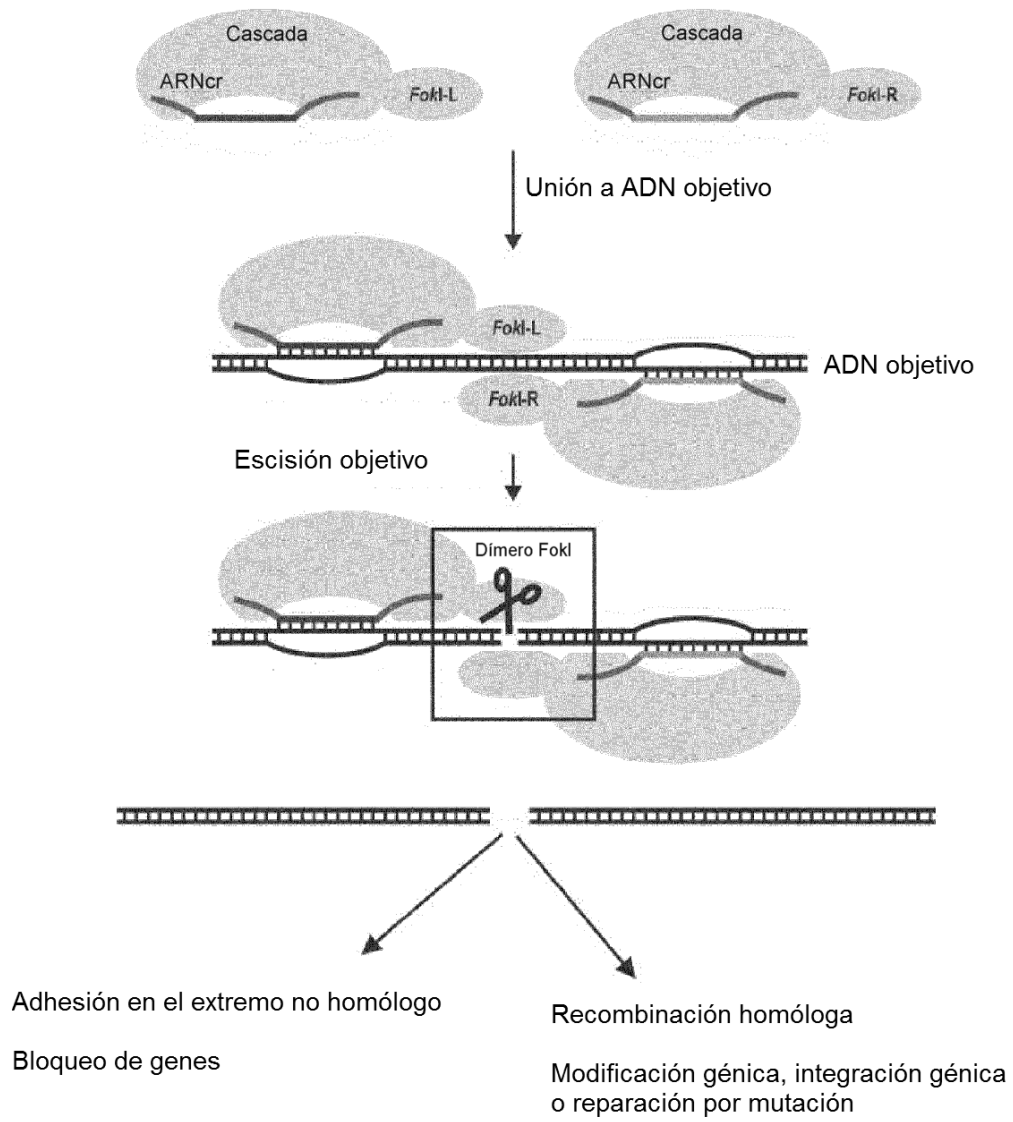


Figura 8

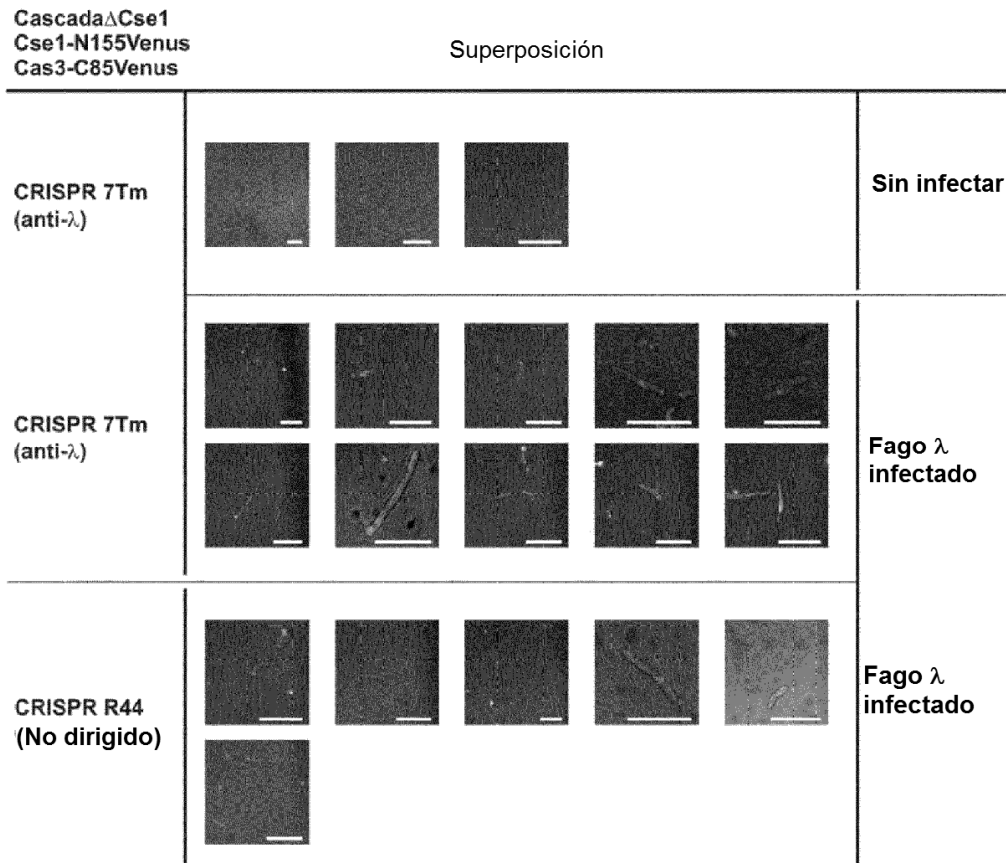


Figura 9

	MAP	<u>semilla</u>
tipo silvestre	AAGGATG	CCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAA
mutante 1	AAGGATG	CGAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAA
mutante 2	-----	-----GCCATGTGGGCTGTCAAAA
mutante 3	-----	-----GAATGCCATGTGGGCTGTCAAAA
mutante 4	-----	-----

Figura 10

[illegible]

Figura 10 continuación

S.sp_5pb78	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537	1538	1539	1540	1541	1542	1543	1544	1545	1546	1547	1548	1549	1550	1551	1552	1553	1554	1555	1556	1557	1558	1559	1560	1561	1562	1563	1564	1565	1566	1567	1568	1569	1570	1571	1572	1573	1574	1575	1576	1577	1578	1579	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589	1590	1591	1592	1593	1594	1595	1596	1597	1598	1599	1600	1601	1602	1603	1604	1605	1606	1607	1608	1609	1610	1611	1612	1613	1614	1615	1616	1617	1618	1619	1620	1621	1622	1623	1624	1625	1626	1627	1628	1629	1630	1631	1632	1633	1634	1635	1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707
------------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Figura 11

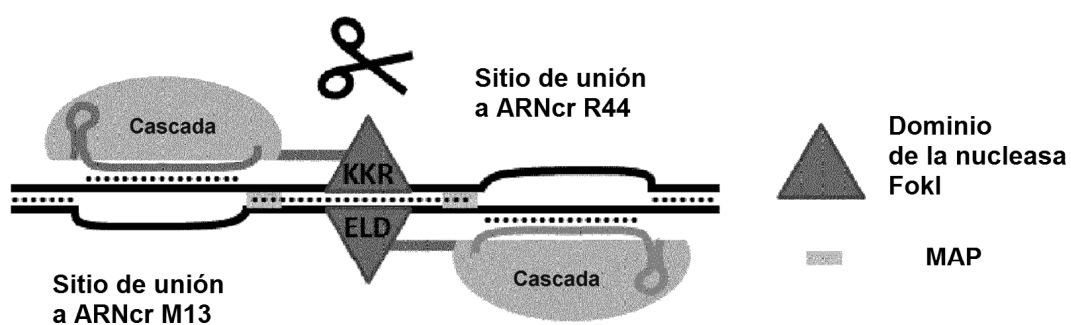


Figura 12

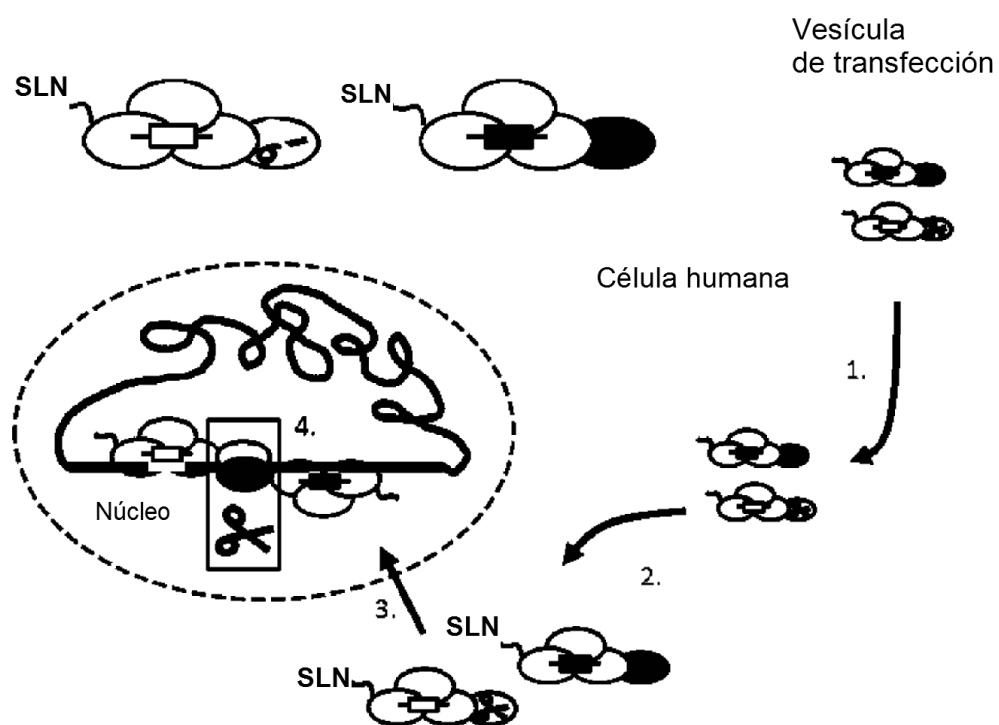


Figura 13

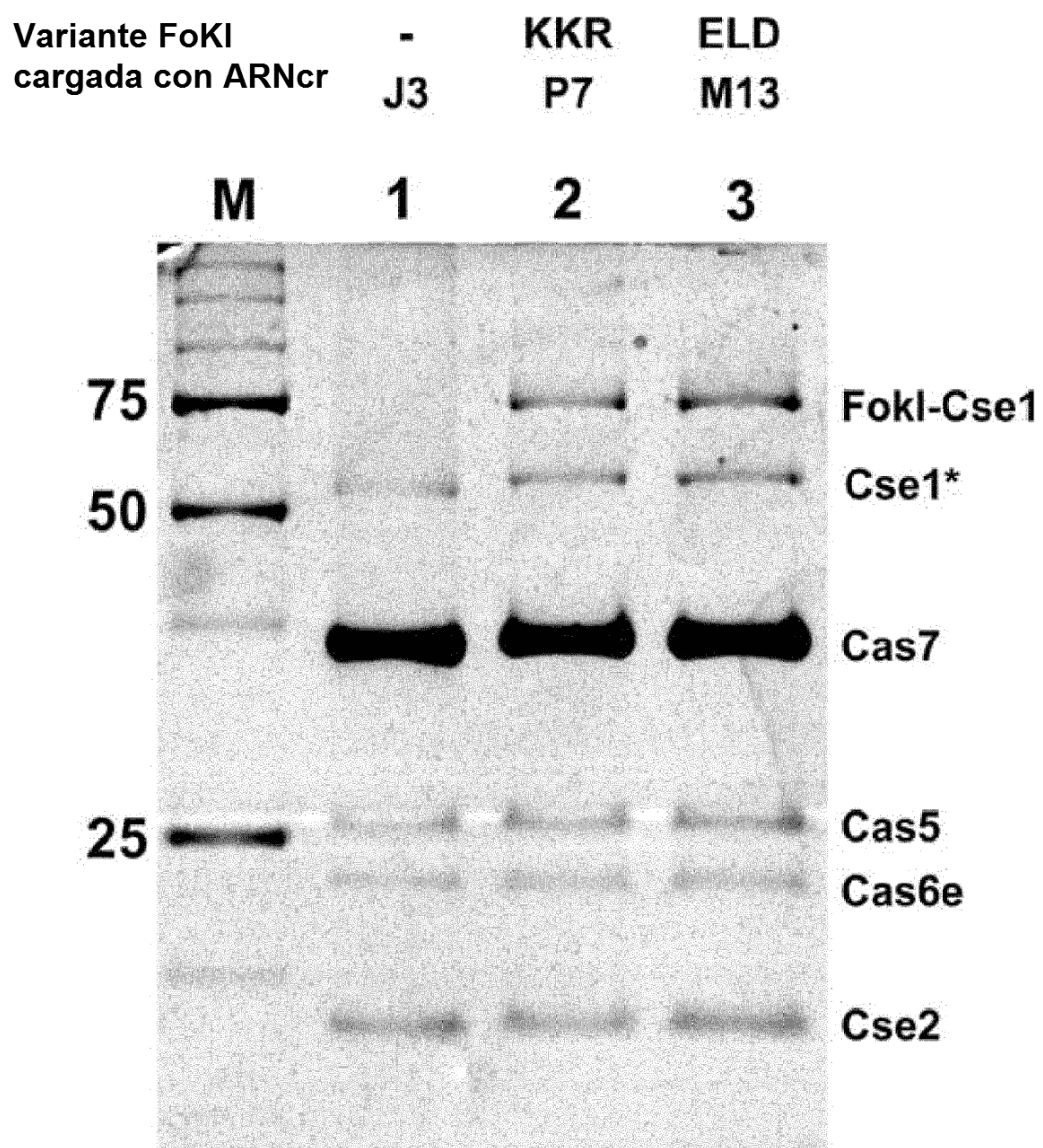


Figura 14

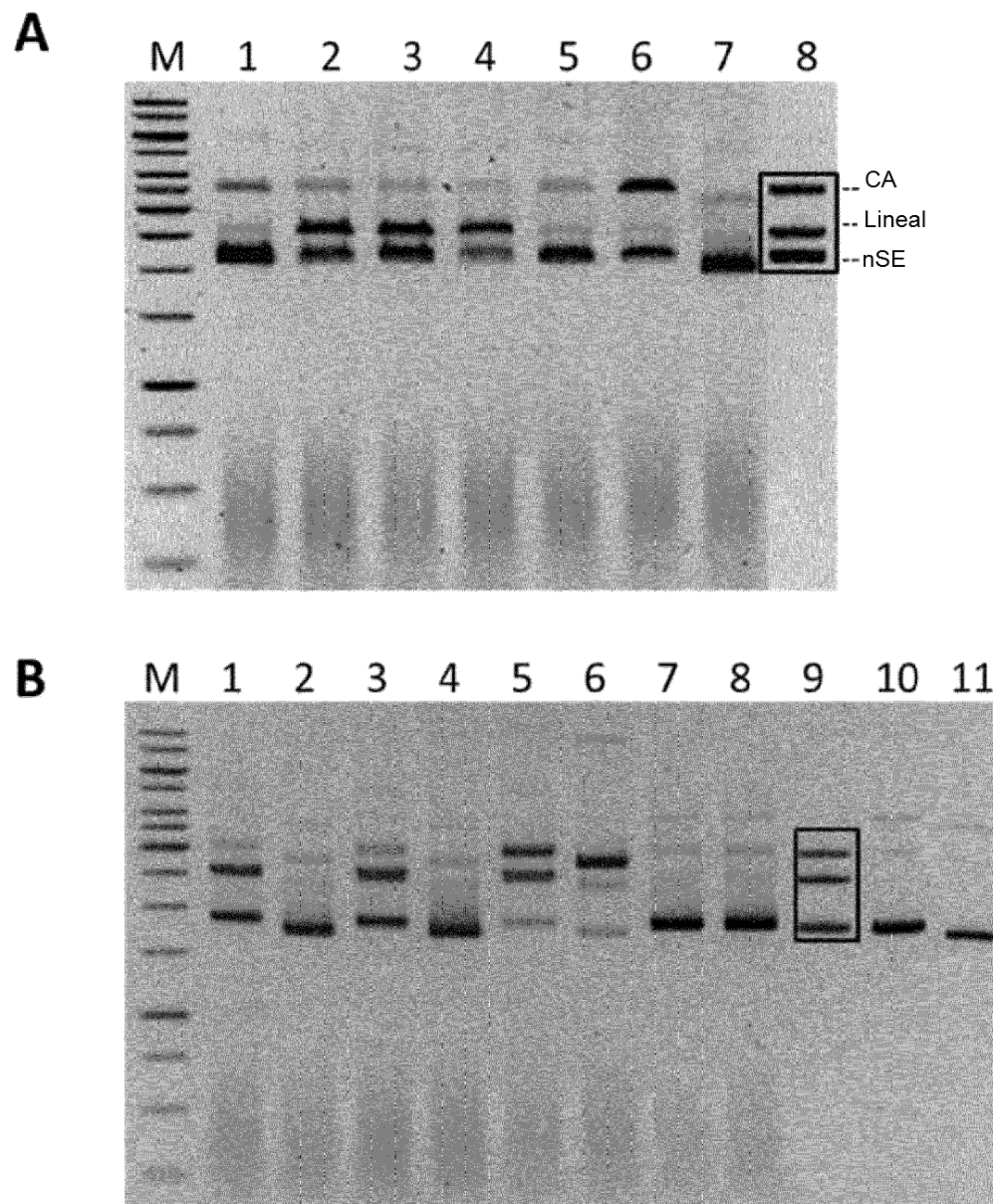
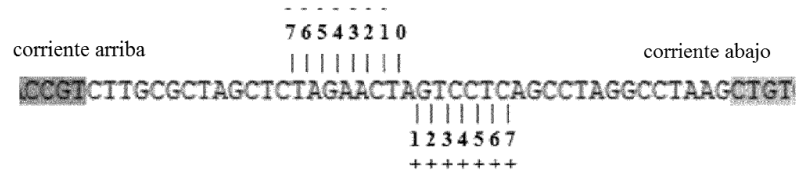


Figura 15

A



B

