

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 749 108**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 15/113** (2010.01)

**C12Q 1/6811** (2008.01)

**C12N 9/22** (2006.01)

**C12N 15/90** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.03.2013 PCT/US2013/033106**

87 Fecha y número de publicación internacional: **26.09.2013 WO13142578**

96 Fecha de presentación y número de la solicitud europea: **20.03.2013 E 13715080 (1)**

97 Fecha y número de publicación de la concesión europea: **10.07.2019 EP 2828386**

54 Título: **Escisión de ADN dirigida por ARN mediante el complejo Cas9-ARNcr**

30 Prioridad:

**20.03.2012 US 201261613373 P**

**17.04.2012 US 201261625420 P**

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

**19.03.2020**

73 Titular/es:

**VILNIUS UNIVERSITY (100.0%)**

**Universiteto g. 3**

**01513 Vilnius, LT**

72 Inventor/es:

**SIKSNYS, VIRGINIJUS;  
GASIUNAS, GIEDRIUS;  
KARVELIS, TAUTVYDAS;  
LUBYS, ARVYDAS;  
ZALIAUSKIENE, LOLITA;  
GLEMZAITE, MONIKA y  
SMITH, ANJA**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

ES 2 749 108 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

Escisión de ADN dirigida por ARN mediante el complejo Cas9-ARNcr

**Resumen**

5 Los sistemas CRISPR/Cas proporcionan una inmunidad adaptativa contra virus y plásmidos en las bacterias y las  
 arqueas. El silenciamiento de los ácidos nucleicos invasores es ejecutado por complejos de ribonucleoproteína (RNP)  
 cargados previamente con pequeños ARNcrs interferentes que actúan como guías para el direccionamiento hacia  
 ácidos nucleicos extraños y la degradación de los mismos. En este documento describimos el aislamiento del complejo  
 10 Cas9-ARNcr y demostramos que genera *in vitro* una rotura de la hebra doble en sitios específicos, en moléculas de  
 ADN diana que son complementarias a las secuencias de ARNcr y son portadoras de un motivo adyacente al proto-  
 espaciador (PAM) corto en las inmediaciones de la secuencia que se aparea. Mostramos que la escisión del ADN es  
 ejecutada por dos sitios activos distintos (RuvC y HNH) dentro de Cas9, para generar mellas específicas del sitio en  
 hebras de ADN opuestas. La especificidad de secuencia del complejo Cas9-ARNcr viene dictada por el ARNcr de 42  
 nt que incluye un fragmento de 20 nt complementario a la secuencia del protoespaciador en el ADN diana. El complejo  
 15 se puede ensamblar *in vitro* o *in vivo*. En conjunto, nuestros datos demuestran que el complejo Cas9-ARNcr actúa  
 como una endonucleasa guiada por ARN con reconocimiento y corte del sitio diana específico de la secuencia, a  
 través de dos mellas en hebras distintas.

**Antecedentes**

Las repeticiones palindrómicas cortas agrupadas e intercaladas regularmente (CRISPR, del inglés “Clustered Regu-  
 20 larly Interspaced Short Palindromic Repeats”) junto con los genes *cas* (asociados a CRISPR, del inglés “CRISPR-  
 associated”) comprenden un sistema inmune adaptativo que proporciona una resistencia adquirida frente a ácidos  
 nucleicos extraños invasores en bacterias y arqueas (Barrangou y col., 2007. Science 315:1709-12). CRISPR consiste  
 en disposiciones de secuencias de repeticiones cortas conservadas, intercaladas con secuencias de ADN variables  
 únicas de tamaño similar, denominadas espaciadores, que a menudo provienen del ADN de un fago o plásmido (Ba-  
 25 rrangou y col., 2007. Science 315:1709-12; Bolotin y col., 2005. Microbiology 151:2551-61; Mojica y col., 2005. J Mol  
 Evol 60:174-82). El sistema CRISPR-Cas actúa mediante la adquisición de segmentos cortos de ADN extraño (espa-  
 ciadores) que se insertan en la región CRISPR y proporcionan inmunidad frente a exposiciones posteriores a fagos y  
 plásmidos que son portadores de secuencias que se aparean (Barrangou y col., 2007. Science 315:1709-12; Brouns  
 y col., 2008. Science 321:960-4). La inmunidad CRISPR-Cas generalmente se lleva a cabo a través de tres etapas,  
 30 denominadas i) adaptación/inmunización/adquisición de espaciadores, ii) expresión de CRISPR/biogénesis de ARNcr,  
 iii) interferencia/inmunidad. (Horvath y Barrangou, 2010. Science 327:167-70; Deveau y col., 2010. Annu Rev Micro-  
 biol. 64:475-93; Marraffini y Sontheimer, 2010. Nat Rev Genet 11, 181-90; Bhaya y col., Annu Rev Genet 45:273-97;  
 Wiedenheft y col., 2012. Nature 482:331-338). En este documento, nos centramos específicamente en la etapa de  
 interferencia/inmunidad que permite el silenciamiento mediado por ARNcr de ácidos nucleicos extraños.

El documento US 2010/0076057 A1 desvela la interferencia de una diana con ARNcr.

35 Los sistemas CRISPR-Cas altamente diversos se clasifican en tres tipos principales, que se subdividen adicionalmente  
 en diez subtipos según el contenido y las secuencias del elemento central (Makarova y col., 2011. Nat Rev Microbiol  
 9:467-77). La organización estructural y la función de los complejos de nucleoproteína implicados en el silenciamiento  
 mediado por ARNcr de ácidos nucleicos extraños, difieren entre los distintos tipos de CRISPR/Cas (Wiedenheft y col.,  
 40 2012. Nature 482:331-338). En el sistema de Tipo I-E, tal y como se ejemplifica con *Escherichia coli*, los ARNcrs se  
 incorporan en un complejo efector con multisubunidades denominado Cascade (complejo asociado con CRISPR para  
 la defensa antiviral, del inglés “CRISPR-associated complex for antiviral defence”) (Brouns y col., 2008. Science  
 321:960-4), que se une al ADN diana y desencadena una degradación a través de la proteína Cas3 distintiva (Sinkunas  
 y col., 2011. EMBO J 30:1335-42; Beloglazova y col., 2011. EMBO J 30:616-27). En los sistemas CRISPR/Cas de  
 45 Tipo III de *Sulfolobus solfataricus* y *Pyrococcus furiosus*, el módulo Cas RAMP (Cmr) y el complejo ARNcr reconocen  
 y escinden el ARN sintético *in vitro* (Hale y col., 2012. Mol Cell 45:292-302; Zhang y col., 2012. Mol Cell, 45:303-13)  
 mientras que el sistema CRISPR/Cas de *Staphylococcus epidermidis* se dirige al ADN *in vivo* (Marraffini y Sontheimer,  
 Science. 322:1843-5).

Los complejos de RNP involucrados en el silenciamiento del ADN a través de los sistemas CRISPR/Cas de Tipo II,  
 50 más específicamente en el sistema CRISPR3/Cas de *Streptococcus thermophilus* DGCC7710 (Horvath y Barrangou,  
 2010. Science 327:167-70), consisten en cuatro genes *cas cas9*, *cas1*, *cas2* y *csn2*, que se encuentran corriente arriba  
 de 12 unidades de repetición-espaciador (Figura 1A). *Cas9* (anteriormente denominado *cas5* o *csn1*) es el gen distin-  
 tivo para los sistemas de Tipo II (Makarova y col., 2011. Nat Rev Microbiol 9:467-77). En el sistema CRISPR1/Cas de  
 55 *S. thermophilus* estrechamente relacionado, la alteración de *cas9* anula la interferencia del ADN mediada por ARNcr  
 (Barrangou y col., 2007. Science 315:1709-12). Hemos observado recientemente que el sistema CRISPR3/Cas de *S.*  
*thermophilus* se puede transferir a *Escherichia coli*, y que este sistema heterólogo proporciona una protección contra  
 la transformación con plásmidos y la infección con fagos, *de novo* (Sapranaukas y col., 2011. Nucleic Acids Res  
 39:9275-82). La interferencia contra el ADN de fagos y plásmidos proporcionada por CRISPR3 de *S. thermophilus*  
 requiere la presencia, dentro del ADN diana, de una secuencia de protoespaciador complementaria al ARNcr obtenido  
 a partir del espaciador, y una secuencia PAM (motivo adyacente al protoespaciador) conservada, NGGNG, ubicada

inmediatamente corriente abajo del protoespaciador (Deveau y col., 2008. J Bacteriol 190:1390-400; Horvath y col., 2008. J Bacteriol 190:1401-12; Mojica y col., 2009. Microbiology 155:733-40). Mutaciones puntuales aisladas en PAM o posiciones del protoespaciador definidas permiten que los fagos o plásmidos eviten la inmunidad mediada por CRISPR (Deveau y col., 2008. J Bacteriol 190:1390-400; Garneau y col., 2010. Nature 468:67-71; Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82). Hemos establecido que en el sistema heterólogo, *cas9* es el único gen *cas* necesario para la interferencia codificada por CRISPR (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82), lo que sugiere que esa proteína está involucrada en el procesamiento del ARNcr y/o el silenciamiento mediado por ARNcr del ADN invasor. Cas9 del sistema CRISPR3/Cas de *S. thermophilus* es una proteína con multidominios grande compuesta por 1.409 restos de aa (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82). Contiene dos dominios de nucleasa, un dominio de nucleasa similar a RuvC cerca del extremo amino terminal y un dominio de nucleasa similar a HNH en el medio de la proteína. El análisis mutacional ha establecido que la interferencia proporcionada *in vivo* por Cas9 requiere ambos motivos RuvC y HNH (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82).

Se proporciona el aislamiento del complejo Cas9-ARNcr del sistema CRISPR3/Cas de *S. thermophilus*, así como el ensamblaje del complejo *in vitro* a partir de componentes distintos y la demostración de que escinde tanto un oligodesoxinucleótido sintético como ADN plasmídico que es portador de una secuencia de nucleótidos complementaria al ARNcr, de una manera dependiente de PAM. Además, proporcionamos una evidencia experimental de que el PAM es reconocido en el contexto de un ADN bicatenario y es decisivo para la unión y la escisión de ADN *in vitro*. Finalmente, observamos que los sitios activos RuvC y HNH de Cas9 son responsables de la escisión de las hebras de ADN opuestas. Resumiendo, nuestros datos demuestran que el complejo Cas9-ARNcr actúa como una endonucleasa guiada por ARN que utiliza el ARN para el reconocimiento del sitio diana y Cas9 para la escisión del ADN. La organización modular simple del complejo Cas9-ARNcr, en donde la especificidad hacia las dianas de ADN está codificada por un ARNcr pequeño y la maquinaria para la escisión consiste en una única proteína Cas con multidominios, proporciona una plataforma versátil para la modificación genética de endonucleasas universales de ADN guiadas por ARN. De hecho, proporcionamos una evidencia de que alterando la secuencia de ARN dentro del complejo Cas9-ARNcr, se pueden diseñar endonucleasas programables tanto para aplicaciones *in vitro* como *in vivo*, y proporcionamos una prueba de concepto para esta aplicación novedosa. Estos hallazgos allanan el camino para el desarrollo de nuevas herramientas moleculares para la cirugía del ADN dirigida por ARN.

### **Compendio de la invención**

El procedimiento de la invención se define en la reivindicación 1. Además se desvela un procedimiento para una modificación específica de sitio de una molécula de ADN diana a través de la puesta en contacto en condiciones adecuadas, de una molécula de polidesoxinucleótido diana; y se proporciona una endonucleasa de ADN guiada por ARN que comprende al menos una secuencia de ARN y al menos uno de un motivo de sitio activo RuvC y un motivo de sitio activo HNH; para dar como resultado la molécula de polidesoxinucleótido diana modificada en una región que se determina mediante la unión complementaria de la secuencia de ARN a la molécula de ADN diana. El procedimiento incluye incubar en condiciones adecuadas una composición que incluye un polidesoxinucleótido bicatenario diana o un polidesoxinucleótido monocatenario diana; en el que un polidesoxinucleótido bicatenario contiene un motivo adyacente al protoespaciador corto (PAM), que no es obligatorio para un polidesoxinucleótido monocatenario; y en donde PAM comprende una secuencia 5'NGGNG-3'; un polirribonucleótido (ARNcr) que comprende regiones 3' y 5' en el que la región 3' comprende al menos 22 nt de la repetición presente en un microbio que contiene un locus CRISPR y una región 5' comprende al menos 20 nt de la secuencia del espaciador inmediatamente corriente abajo de la repetición en el locus CRISPR, que es sustancialmente complementaria, opcionalmente complementaria, a una porción del polinucleótido diana, un polipéptido en el que la secuencia de aminoácidos del polipéptido y la secuencia de aminoácidos de la SEQ ID NO: 1 tienen al menos un 80 % de identidad, aislado a partir de *S. thermophilus*, o un microorganismo modificado genéticamente que incluye una *E. coli* modificada genéticamente, o en el que el polipéptido se produce mediante un procedimiento seleccionado de tecnología de ADN recombinante o síntesis química; un ARNtracr de polirribonucleótido con la secuencia de nucleótidos SEQ ID NO: 5 (o que tiene al menos el 80 % de identidad) que comprende regiones 5' y 3', en el que la región 5' que está compuesta por al menos 22 nucleótidos es complementaria a la región 3' de 22 nucleótidos del ARNcr, y una región 3'. En el que los polirribonucleótidos son producidos mediante transcripción *in vitro* o síntesis química. En el que, unas condiciones adecuadas significa condiciones *in vitro* en las que puede tener lugar la reacción.

Se desvela un procedimiento para la conversión de un polipéptido Cas9 en una mellasa (del inglés "nickase"), que corta solo una hebra del ADN bicatenario, inactivando uno de los sitios activos (RuvC o HNH) en el polipéptido mediante al menos una mutación puntual, se proporcionan mutaciones puntuales ejemplificadas por D31A (SEQ ID NO: 2), N891A (SEQ ID NO: 3) y H868A (SEQ ID NO: 4). El mutante del motivo RuvC corta solo la hebra de ADN inferior con respecto al motivo 5'NGGNG-3', mientras que el mutante del motivo de HNH corta la hebra superior.

El complejo polipéptido-polirribonucleótidos se puede aislar a partir de un microbio modificado genéticamente (por ejemplo, *Escherichia coli* o *Streptococcus thermophilus*), o ensamblar *in vitro* a partir de componentes distintos. En el microbio modificado genéticamente, los componentes del complejo se pueden codificar en uno, dos o tres plásmidos distintos que contienen promotores del hospedador del microbio modificado genéticamente o promotores procedentes del genoma de un hospedador natural.

Se desvela un procedimiento para ensamblar un complejo polipéptido-polirribonucleótidos activo *in vitro* que

comprende incubar los componentes del complejo en condiciones adecuadas para proporcionar el ensamblaje del complejo. El complejo se puede ensamblar utilizando tres o cuatro componentes. El procedimiento para ensamblar tres componentes comprende la incubación del polipéptido Cas9, el polirribonucleótido ARNtracr de 78 nt (SEQ ID NO: 5) y el polirribonucleótido ARNcr de 42 nt (5'-NNNNNNNNNNNNNNNNNNNNNNN GUUUUAGA-GCUGUGUUUGUUUCG-3') (SEQ ID NO: 15) en condiciones adecuadas para ensamblar el complejo. El procedimiento para ensamblar cuatro componentes comprende incubar el polipéptido Cas9; el polirribonucleótido ARNtracr de 102 nt (SEQ ID NO: 6); la secuencia que contiene el polirribonucleótido 5'-NNNNNNNNNNNNNNNNNNNNNNN GUUUUAGA-GCUGUGUUUGUUUCG-3' (SEQ ID NO: 15) y regiones flanqueantes y polipéptido RNasa III, que escinde un polinucleótido de ARN bicatenario. Ejemplos para la secuencia que contiene el polirribonucleótido 5'-NNNNNNNNNNNNNNNNNNNNNNN GUUUUAGAGCUGUGUUUGUUUCG-3' (SEQ ID NO: 15) son la SEQ ID NO: 8, la SEQ ID NO: 9, la SEQ ID NO: 10, la SEQ ID NO: 11 y la SEQ ID NO: 12. Ejemplos de una fuente de RNasaIII adecuada incluyen *Escherichia coli* o *Streptococcus thermophilus*.

Se desvela un procedimiento que se proporciona para reprogramar una especificidad del complejo Cas9-ARNcr mezclando componentes distintos o usando un casete que contiene una sola unidad de repetición-espaciador-repetición. Cualquier secuencia se puede insertar entre dos repeticiones en el casete usando endonucleasas de restricción adecuadas. El casete se puede usar para direccionar secuencias *in vivo*, o para producir ARN ribonucleótido adecuado para ensamblar el complejo *in vitro*.

### **Breve descripción de las Figuras**

La Figura 1 muestra que la proteína Cas9 se purifica junto con ARNcr. (A) Representación esquemática del sistema CRISPR3/Cas de *S. thermophilus*. Cuatro genes cas (cas9, cas1, cas2, csn2) se encuentran corriente arriba de la disposición de repetición-espaciador de CRISPR, que consiste en 13 secuencias de repeticiones (R) y 12 espaciadores únicos (S1-S12). El ARNtracr, requerido para la maduración del ARNcr en los sistemas CRISPR de Tipo II (Deltcheva y col., 2011. Nature 471:602-7), se encuentra corriente arriba del gen cas9 y se codifica en la hebra de ADN opuesta (mostrada por una flecha) con respecto a los otros elementos del sistema CRISPR3/Cas. (B) Representación esquemática de loci heterólogos en dos plásmidos utilizados para la coexpresión del complejo Cas9-ARNcr. La cepa de *E. coli* RR1 contenía los plásmidos pCas9(-)1SP (que codifica Cas1, Cas2, Csn2, SP1 y ARNtracr) y pASKIBA-Cas9 (que codifica la versión marcada con Strep de Cas9). (C) Análisis Northern de complejos Cas9-ARNcr usando un oligonucleótido anti-ADNcr como sonda. M1 - oligodesoxinucleótido de 84 nt correspondiente a la unidad de espaciador S1-repetición; M2 - oligorribonucleótido sintético de 42 nt correspondiente al ARNcr CRISPR3 de *S. thermophilus* previsto (véase la Figura 4); ARNcr (wt) - ARNcr aislado a partir del complejo Cas9 wt; K1 - ARNcr (wt) tratado con DNasa I durante 15 min; K2 - ARNcr (wt) tratado con RNasaI durante 15 minutos, D31A - ARNcr purificado a partir del complejo mutante Cas9 D31A; N891A - ARNcr purificado a partir del complejo mutante Cas9 N891A.

La Figura 2 muestra la escisión de ADN a través de complejos Cas9-ARNcr obtenidos mediante la coexpresión de Cas9 con el locus CRISPR de longitud completa. (A) Representación esquemática del locus CRISPR/Cas del plásmido pCas9(-) recombinante que es portador de una disposición autóctona de 12 espaciadores-repeticiones del sistema SthCRISPR3/Cas y el plásmido pASKIBA-Cas9 que es portador del gen cas9 con un marcador Strep en el extremo C-terminal. (B) Ensayo de escisión de oligodúplex. Ambos plásmidos, pCas9(-) y pASKIBA-Cas9 se coexpresaban en *E. coli*, los complejos Cas9-ARNcr se purificaron y se sometieron a un análisis de escisión usando los oligodúplex SP1 (primer protoespaciador) y SP2 (segundo protoespaciador) marcados con 33P en el extremo 5' de la hebra (+). Los productos de la reacción se analizaron en gel de PAA.

La Figura 3 muestra inmunidad frente a una transformación con plásmido en células de *E. coli* proporcionada por el sistema SthCRISPR3/Cas. (A) Representación esquemática del locus CRISPR/Cas del plásmido recombinante pCRISPR3 que es portador de una disposición autóctona de 12 espaciadores-repeticiones del sistema SthCRISPR3/Cas y el plásmido pCRISPR3-SP1 modificado genéticamente que es portador de 1 unidad de espaciador-repetición. (B) Interferencia de la transformación de un plásmido a través del sistema SthCRISPR3/Cas en células de *E. coli*. Las cepas receptoras de *Escherichia coli* RR1 que eran portadoras de los plásmidos pACYC184, pCRISPR3 o pCRISPR3-SP1, se transformaron con el plásmido pSP1 que era portador de protoespaciadores y PAM o pUC18 (1). La eficacia de la transformación se expresa como ufc por nanogramos de ADN plasmídico (media  $\pm$  DE).

La Figura 4 muestra una comparación de los sistemas CRISPR/Cas de Tipo IIA procedentes de las cepas de *S. thermophilus* DGCC7710, LMD-9 y *S. pyogenes* SF370. (A) Organización esquemática de los sistemas CRISPR/Cas. Las secuencias de nucleótidos correspondientes al ARNtracr requerido para la maduración del ARNcr en *S. pyogenes* (2) están presentes en LMD-9 y DGCC7710. Porcentaje de restos idénticos y similares (entre paréntesis) entre las secuencias de proteínas correspondientes que están conectadas por líneas discontinuas. (B) Alineación de las secuencias repetidas conservadas y ARNtracr. Las secuencias correspondientes procedentes de DGCC7710 y LMD-9 son idénticas. Las posiciones de nucleótidos que son idénticas en las tres cepas están marcadas con un asterisco debajo de las secuencias alineadas. La Figura 4(B) desvela las SEQ ID NOS 50, 50-52 y 52-53, respectivamente, por orden de aparición. (C) Comparación de las secuencias de ARNcr. La secuencia y la longitud del ARNcr de *S. pyogenes* se determinaron mediante un análisis de secuenciación profunda (2). La longitud aproximada del ARNcr procedente de las cepas de *S. thermophilus* LMD-9 (2) y DGCC7710 (este

trabajo), se determinaron mediante el análisis de transferencia Northern. La Figura 4(C) desvela las SEQ ID NOS 54-56, respectivamente, por orden de aparición.

La Figura 5 muestra que el complejo Cas9-ARNcr corta *in vitro* ADN bicatenario dentro de un protoespaciador. (A) Sustrato de oligodúplex utilizado en el ensayo de escisión. El oligodúplex SP1 de 55 nt contiene el protoespaciador1 (letras rojas), PAM (letras azules) y secuencias flanqueantes de 10 nt a ambos lados, idénticas a las del plásmido pSP1. En el oligodúplex SP1, la hebra de ADN complementaria al fragmento 5'-terminal del ARNcr (letras rojas) se denomina hebra (+), una hebra de ADN opuesta se denomina hebra (-). La Figura 5(A) desvela las SEQ ID NOS 31, 7 y 34, respectivamente, por orden de aparición. (B) Escisión del oligodúplex SP1. Se incubaron 2,5 nM del complejo Cas9-ARNcr y 1 nM del oligodúplex SP1 marcado con 33P en el extremo 5' de la hebra (+) o (-), en tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C durante intervalos de tiempo variados (de 30 s a 10 min) y los productos de la reacción se analizaron en el gel de PAA al 20%. Los carriles M1 y M2 contienen oligodesoxinucleótidos de 37 nt y 18 nt, sintetizados químicamente, marcados con 33P en el extremo 5' que se corresponden con los productos de escisión de las hebras de ADN (-) y (+), respectivamente. Las posiciones de la escisión se indican con flechas. La Figura 5(B) desvela la SEQ ID NO: 31. (C) Representación esquemática del plásmido pSP1 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) utilizado en el ensayo de escisión del plásmido. La Figura 5(C) desvela la SEQ ID NO: 57. (D) Escisión del plásmido pSP1. Análisis en gel de agarosa de los productos de escisión de pSP1 (panel izquierdo). SC - ADN plasmídico superenrollado, OC - ADN circular abierto mellado en una de las hebras, FLL - ADN lineal de longitud completa cortado en ambas hebras. Las mezclas de reacción finales a 37 °C contenían 2,5 nM de plásmido pSP1 y 2,5 nM de complejo Cas9-ARNcr en el tampón de reacción (sección B). Electroferogramas de secuenciación directa (panel derecho) de las hebras (+) (parte superior) y (-) (parte inferior) del producto de escisión del plásmido pSP1. La adición sin molde de adenina (T en la secuencia complementaria inversa que se muestra en este documento) en el extremo de la secuencia, es un artefacto de la secuenciación causado por la polimerasa. La Figura 5(D) desvela las SEQ ID NOS 57-59, 58 y 60, respectivamente, por orden de aparición.

La Figura 6 muestra un análisis de la unión y la escisión del ADN de la proteína Cas9-Chis que carece de ARNcr. Análisis del cambio de la movilidad electroforética (EMSA) de la proteína Cas9-Chis que se une a (A) el oligodúplex SP1 bicatenario y (B) el oligonucleótido s(+)-SP1 monocatenario. Los experimentos de cambio de la movilidad electroforética se realizaron en el tampón de unión (Tris-acetato 40 mM, pH 8,3 a 25 °C, EDTA 0,1, 0,1 mg/ml de BSA, glicerol al 10% v/v). Las reacciones contenían 0,5 nM del oligodúplex marcado con 33P y la proteína, en las concentraciones indicadas encima de cada carril. (C). Ensayo de escisión de oligonucleótidos. Se incubaron 5 nM de proteína Cas9-Chis en el tampón de reacción (Tris-HCl 10 mM, pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C con oligonucleótido 1 nM. El oligodúplex SP1 se marcó con 33P en el extremo 5' de la hebra (+) o (-). El oligonucleótido monocatenario s(+)-SP1 se marcó con 33P en el extremo 5'.

La Figura 7 muestra una reprogramación del complejo Cas9-ARNcr. (A) Representación esquemática de loci heterólogos en dos plásmidos utilizados para la reprogramación del complejo Cas9-ARNcr. pCas(-)SPN se construyeron a partir del plásmido pCas9(-) (véase la Figura 2A), insertando una secuencia espaciadora (SN) nueva (5'-CC ACC CAG CAA AAT TCG GTT TTC TGG CTG-3' (SEQ ID NO: 16)) e inactivando el gen Cas9 como se describe en (1). (B) Análisis en gel de agarosa de los productos de escisión del ADN plasmídico. pSP1 y pSP1+SPN (plásmido pSP1 con un nuevo protoespaciador y PAM insertados en el sitio AatII, se incubaron a una concentración de 2,5 nM con 2 nM de complejo Cas9-ARNcr en tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C durante intervalos de tiempo variados y los productos de la reacción se analizaron en gel de agarosa. SC - ADN plasmídico superenrollado, OC - ADN circular abierto mellado en una de las hebras de ADN, FLL - ADN lineal de longitud completa cortado en ambas hebras. (C) Escisión del oligodúplex SP1. Se incubaron 2,5 nM del complejo Cas9-ARNcr y oligodúplex SPN 1 nM (Tabla S2) marcado con 33P en el extremo 5' de la hebra (+) o (-), en tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C. M1 - marcador de 18 nt de longitud. Los carriles M1 y M2 contienen oligodesoxinucleótidos de 18 nt y 37 nt marcados con 33P en el extremo 5', sintetizados químicamente, correspondientes a los productos de escisión de las hebras de ADN (+) y (-), respectivamente. (D) Representación esquemática del sustrato del oligodúplex SPN y los productos de escisión. El oligodúplex SPN contiene el nuevo protoespaciador (letras rojas), PAM (letras azules). Las posiciones de la escisión se indican con flechas. La Figura 7(D) desvela la SEQ ID NO: 39.

La Figura 8 muestra el impacto de la longitud del espaciador sobre la inmunidad codificada por CRISPR. (A) Representación esquemática de versiones acortadas de protoespaciadores insertados en los plásmidos transformados. La Figura 8(A) desvela las SEQ ID NOS 7 y 61-66, respectivamente, por orden de aparición. (B) Efecto de la longitud del protoespaciador sobre la eficacia de la transformación del plásmido. La eficacia de la transformación se expresa como ufc por nanogramo de ADN plasmídico (media ± DE). (C). Representación esquemática de oligodúplex utilizados en la escisión *in vitro* y experimentos de unión. La Figura 8(C) desvela las SEQ ID NOS 31 y 38, respectivamente, por orden de aparición. (D) Cursos temporales de la escisión del oligodúplex de 27 pb (protoespaciador SP1 de longitud completa, círculos negros) y el oligodúplex de 20 pb (protoespaciador SP1-20 truncado, cuadrados) a través del complejo Cas9-ARNcr. (E) Ensayo de cambio de la movilidad electroforética de la unión del oligodúplex SP1 y SP1-20 a través del complejo Cas9-ARNcr.

La Figura 9 muestra que se requiere PAM para la unión *in vitro* de ADN y la escisión con el complejo Cas9-ARNcr. (A) Análisis en gel de agarosa de productos de escisión de ADN plasmídico. Tres plásmidos diferentes:

PAM+protoespaciador+ (plásmido pSP1 que contiene el protoespaciador y PAM), PAM-protoespaciador- (plásmido pUC18 que contiene múltiples PAMs pero sin protoespaciador) y PAM-protoespaciador+ (pSP1-pΔ (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) que contiene un protoespaciador sin PAM) se incubaron a una concentración de 2,5 nM con 2 nM de complejo Cas9-ARNcr en el tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C durante intervalos de tiempo variados y los productos de la reacción se analizaron en gel de agarosa. SC - ADN de plásmido superenrollado, OC - ADN circular abierto mellado en una de las hebras de ADN, FLL - ADN lineal de longitud completa cortado en ambas hebras. (B) Cursos temporales de la hidrólisis de la hebra (+) en los oligodesoxinucleótidos monocatenarios y bicatenarios. Las reacciones que contenían 2 nM de Cas9-ARNcr y 1 nM de oligodesoxinucleótido se llevaron a cabo a 37 °C en el tampón de reacción (sección A). Se usaron los oligodúplex SP1 (círculos negros) y SP1-pΔ (cuadrados blancos) como ADNds. s(+)-SP1 (triángulos blancos) y s(+)-SP1-pΔ (cuadrados negros) se usaron como ADNss. (C) y (D) ADNds y ADNss (hebra (+)) que se unen mediante el complejo Cas9-ARNcr. Las reacciones contenían 0,5 nM del oligonucleótido ADNss o ADNds marcado con 33P, y la proteína en las concentraciones indicadas encima de cada carril. Después de 15 minutos a temperatura ambiente, las muestras se sometieron a PAGE durante 2 horas y se analizaron como se describe en "Materiales y procedimientos".

La Figura 10 muestra el análisis de la unión y la escisión de ARN del complejo Cas9-ARNcr. (A) Análisis del cambio de la movilidad electroforética (EMSA) del complejo Cas9-ARNcr que se une al fragmento de ARN de 84 nt que contiene el protoespaciador-1, PAM y secuencias flanqueantes de 24 nt en ambos lados. Panel izquierdo: hebra (-) de ARN; panel central: hebra (+) de ARN; panel derecho: ARN bicatenario. Los fragmentos de ARN utilizados para el análisis se generaron mediante transcripción *in vitro* (TranscriptAid™ T7 High Yield Transcription Kit, Fermentas) a partir de fragmentos de PCR con el promotor T7 insertado en el extremo frontal de la secuencia codificadora de ARN. Los fragmentos de PCR que codificaban las hebras (+) y (-) de ARN se obtuvieron a partir del plásmido pSP1 (1) con las siguientes parejas de cebadores correspondientes: 5' taatagactactataGggtaccgagctcgaattg 3' (SEQ ID NO: 17)/5' GGAAACAGCTATGACCATGATTACGAATTC -3' (SEQ ID NO: 18) y 5' gggtagcagctcgaattgaaattcTAAACG 3' (SEQ ID NO: 19)/5' taatagactactataGggAAACAGCTATGACCATGATTACG 3' (SEQ ID NO: 20) (promotor de ARN polimerasa T7 subrayado, inicio de la transcripción en negrita). Las reacciones contenían 1 nM del fragmento de ARN marcado con 33P, y la proteína en las concentraciones tal y como se indican encima de cada carril. Después de 15 minutos a temperatura ambiente, las muestras se sometieron a PAGE durante 2 horas y se analizaron como se describe en "Materiales y procedimientos". (B) Ensayo de escisión del ARN. Se incubaron 2,5 nM de complejo Cas9-ARNcr en el tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C en presencia de 1 nM de las hebras (+) y (-) de ARN (panel izquierdo) o ARN bicatenario marcado en la hebra (+) o (-) (panel derecho). Los productos de la reacción se analizaron en gel de PAA desnaturizante.

La Figura 11 muestra que los motivos del sitio activo RuvC y HNH de Cas9 contribuyen a la escisión de hebras de ADN opuestas. (A) Localización de los motivos de un sitio activo conservado dentro de la proteína Cas9. Se indican los restos de aminoácidos identificados como cruciales para la actividad *in vivo* de Cas9 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82). (B) Análisis en gel de agarosa de la escisión del plásmido pSP1 mediante Cas9 y proteínas mutantes. Las reacciones se llevaron a cabo como se describe en "Materiales y procedimientos". (C) Preferencia de hebra del mutante D31A. Las reacciones se realizaron como se describe en la Figura 2A y "Materiales y procedimientos". El mutante D31 escinde solo la hebra (+) del oligodúplex SP1. La Figura 11(C) desvela las SEQ ID NOS 31 y 67, respectivamente, por orden de aparición. (D) Preferencia de hebra del mutante N891A. El mutante N891 escinde solo la hebra (-) del oligodúplex SP1. Las posiciones de la escisión se indican con flechas. La Figura 11(D) desvela las SEQ ID NOS 31 y 68, respectivamente, por orden de aparición.

La Figura 12 muestra las propiedades de los complejos de mutante del sitio activo de Cas9-ARNcr. (A) Secuenciación directa de los productos de reacción obtenidos con el mutante de Cas9, D31A (motivo de sitio activo similar a RuvC). La Figura 12(A) desvela las SEQ ID NOS 58, 59, 58 y 58, respectivamente, por orden de aparición. (B) Secuenciación directa de los productos de reacción obtenidos con el mutante de Cas9 N891A (motivo de sitio activo similar a HNH). La Figura 12(B) desvela las SEQ ID NOS 58, 58, 58 y 60, respectivamente, por orden de aparición. (C) Unión del oligodúplex SP1 mediante los complejos Cas9-ARNcr wt y mutantes del sitio activo. (D) Escisión de la hebra (+)SP1 por complejos mutantes Cas9-ARNcr.

La Figura 13 muestra la masa molecular de la proteína Cas9-Chis wt. Los experimentos de filtración en gel se llevaron a cabo a temperatura ambiente usando una columna Superdex 200 10/300 GL (GE healthcare) equilibrada previamente con tampón fosfato de sodio 10 mM (pH 7,4) que contenía cloruro de sodio 500 mM. Los PMs aparentes de Cas9 (triángulo negro) fueron calculados mediante interpolación a partir de la curva estándar obtenida, usando un conjunto de proteínas de PMs conocidos (círculos negros) (patrones de filtración en gel de Bio-Rad).

La Figura 14 muestra una disposición esquemática y el mecanismo de escisión de ADN dirigido por ARNcr a través del complejo Cas9-ARNcr. La estructura del dominio de Cas9 se muestra esquemáticamente en la parte superior. El complejo Cas9-ARNcr se une al ADNds que contiene PAM. El ARNcr se une a la hebra (+) complementaria dando como resultado la separación de la hebra de ADN y la formación del bucle R. En el complejo ternario, el sitio activo RuvC de Cas9 está ubicado en el fosfato escindible sobre la hebra (-) no apareada, mientras que el sitio activo HNH está ubicado en el fosfato escindible en la hebra (+) del ADN unido al ARNcr. La acción coordinada de ambos sitios activos da como resultado la rotura de la hebra doble a 4 nt de distancia del PAM, lo que genera

un ADN de extremos romos. La Figura 14 desvela las SEQ ID NOS. 31 y 69, respectivamente, por orden de aparición.

La Figura 15 muestra una electroforesis natural de Cas9-ARNcr y los productos de escisión. La proteína en las concentraciones indicadas encima de cada carril, se incubó en tampón de reacción (Tris-HCl 10 mM pH = 7,5, NaCl 10 mM, MgCl<sub>2</sub> 10 mM, 0,1 mg/ml de BSA) a 37 °C durante 30 min en presencia de oligodúplex SP1 0,5 nM. Las muestras se mezclaron con solución de colorante de carga (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y se analizaron mediante PAGE no desnaturante. Los carriles del gel marcados con M - forma mezclada de los productos de reacciones de escisión. Las viñetas a cada lado del gel ilustran los complejos de proteína-ADN y el ADN que se corresponde a cada banda, mientras que las viñetas debajo del gel ilustran la forma principal del sustrato después de la reacción.

La Figura 16 muestra la escisión del ADN plasmídico a través del complejo Cas9-ARNcr. (A) escisión del ADN plasmídico de pSP1 y pUC18. El complejo Cas9-ARNcr se incubó con los plásmidos pSP1 y pUC18 en un tampón de reacción proporcionado en el Ejemplo 1. El plásmido pSP1 contenía una secuencia de protoespaciador1 flanqueada por la secuencia 5'-GGNG-3'PAM. La secuencia del protoespaciador1 no estaba presente en pUC18. Los productos de la reacción se analizaron en el gel de agarosa. En esas condiciones, el plásmido pSP1 se convierte en una forma lineal, mientras que el plásmido pUC18 que carece de la secuencia de protoespaciador1, es resistente a la escisión. (B) Reacciones de escisión de pSP1 en ausencia de uno de los componentes. En las mezclas de reacción que carecen de uno de los componentes (Cas9, ARNcr o ARNtracr, respectivamente), el plásmido pSP1 no se escinde. SC - ADN de plásmido superenrollado, OC - ADN circular abierto mellado en una de las hebras de ADN, FLL - ADN lineal de longitud completa cortado en ambas hebras.

La Figura 17 muestra la escisión del oligodúplex de ADN a través del complejo Cas9-ARNcr. La hebra del oligodúplex que es complementaria a ARNcr está marcada como hebra (+), mientras que la otra hebra - es la hebra (-). Para controlar las reacciones de escisión, la hebra (+) o (-) del oligodúplex se marcó con P33 en el extremo 5'. M1 y M2 son marcadores de oligonucleótidos sintéticos correspondientes a los 37 nt de la hebra (-) y a los 18 nt de la hebra (+), los cuales se usaron para determinar el tamaño de los productos de escisión y para cartografiar la posición de la escisión. La proteína Cas9 corta ambas hebras del oligodúplex dentro del protoespaciador, después del nucleótido 37, 4 nt corriente arriba del PAM (5'-GGNG-3') dejando extremos romos. Ambas hebras de sustrato no específico (K1 y K2) no se escinden cuando se incuban con el complejo Cas9-ARNcr durante 30 minutos. La Figura 17 desvela la SEQ ID NO: 31.

La Figura 18 muestra la escisión del ADN plasmídico a través del complejo Cas9-ARNcr ensamblado en ausencia de RNasalll. El complejo Cas9-ARNcr se incubó con el plásmido pSP1 y los productos de la reacción se analizaron en los geles de agarosa. El plásmido pSP1 es resistente a la escisión en presencia de complejo ensamblado sin ARNcr (panel izquierdo). El plásmido pSP1 se convierte en una forma lineal en presencia del complejo ensamblado usando un ARNcr sintético de 42 nt (sin RNasalll) (panel central). El plásmido pSP1 se convierte en una mezcla de formas de ADN lineales y circulares en presencia de complejo ensamblado usando un transcrito de ARN CRISPR (sin RNasalll) (panel derecho).

La Figura 19 muestra la escisión del oligodúplex de ADN a través del complejo Cas9-ARNcr. La hebra del oligodúplex que es complementaria al ARNcr está marcada como hebra (+), mientras que la otra hebra - es la hebra (-). Para controlar la reacción de escisión, la hebra (+) o (-) del oligodúplex se marcó con P33 en el extremo 5'. M1 y M2 son marcadores de oligonucleótidos sintéticos correspondientes a los 37 nt de la hebra (-) y a los 18 nt de la hebra (+), los cuales se usaron para determinar el tamaño de los productos de escisión y para cartografiar la posición de la escisión. La proteína Cas9 corta ambas hebras del oligodúplex dentro del protoespaciador, después del nucleótido 37 a partir del extremo 5', 4 nt corriente arriba del PAM (5'-GGNG-3') dejando extremos romos. Ambas hebras de sustrato no específico (K1 y K2) no se escinden cuando se incuban con el complejo Cas9-ARNcr durante 30 minutos. La Figura 19 desvela la SEQ ID NO: 31.

La Figura 20 muestra (A) una representación esquemática del sistema CRISPR3/Cas de *S. thermophilus* DGCC7710. Cuatro genes cas (*cas9*, *cas1*, *cas2*, *csn2*) se encuentran corriente arriba de la disposición de espaciador-repetición de CRISPR, que consiste en 13 secuencias de repeticiones (R) y 12 espaciadores únicos (S1-S12). El ARNtracr, requerido para la maduración del ARNcr en los sistemas CRISPR/Cas de Tipo II (Deltcheva y col., 2011. Nature 471, 602-7), se encuentra corriente arriba del gen *cas9* y está codificado en la hebra de ADN opuesta (mostrada por una flecha) con respecto a los otros elementos de ese sistema. (B) Rutas para una nueva inserción de espaciador en la región CRISPR y síntesis de ARN CRISPR. El oligodúplex sintético que codifica una secuencia espaciadora deseada y que contiene extremos compatibles con la restricción SapI y Eco31I, se insertó entre dos repeticiones. La región CRISPR se amplificó usando PCR. El nuevo espaciador que codificaba el ARN CRISPR se obtuvo mediante transcripción *in vitro*. (C) Ensamblaje *in vitro* del complejo Cas9-ARN. Los transcritos de ARN CRISPR y ARNtracr se ensamblaron en un dúplex. La proteína Cas9 se preincubó primero con un dúplex de ARN, seguido de una incubación posterior con RNasalll para generar un complejo Cas9-ARN catalíticamente competente.

La Figura 21 muestra A. Representación esquemática del plásmido pUC18. La distancia entre los sitios de restricción SapI y AatII es de 775 pb, mientras que la distancia entre dos espaciadores es de 612 pb. B. Escisión del

plásmido pUC18 mediante complejos reprogramados de Cas9-ARNcr. "1" - plásmido pUC18; "2" - pUC18 cortado con AatII; "3" - pUC18 cortado con un complejo que contiene un protoespaciador1 que se aparea con el ARNcr; "4" - pUC18 cortado con SapI; "5" - pUC18 cortado con un complejo que contiene un protoespaciador2 que se aparea con el ARNcr; "6" - pUC18 cortado con AatII y SapI; "7" - pUC18 cortado con una mezcla de los complejos utilizados en las pistas 3 y 5.

La Figura 22 muestra la escisión de ADN genómico con el complejo Cas9-ARN ensamblado *in vitro*. (A) Análisis en gel de agarosa de los productos de la escisión de ADN de  $\lambda$  lineal. El ADN del fago  $\lambda$  se incubó con el complejo Cas9-ARN en el tampón de reacción durante varios intervalos de tiempo. El sitio diana para el complejo Cas9-ARN se encuentra a 8 kb de distancia del sitio *cos*. (B). Selección de sondas para experimentos de transferencia Southern. El ADN genómico se fragmentó mediante tratamiento con la enzima PstI. El protoespaciador se encuentra entre dos sitios PstI. Si el ADN genómico se escinde con el complejo Cas9-ARN, se debe detectar un fragmento de 466 pb. De lo contrario, la sonda se hibridará con un fragmento de 1499 pb de longitud. (C) Análisis con transferencia Southern de fragmentos de ADN genómico. Pista C - ADN genómico de *E. coli* fragmentado con PstI. Cas9-ARN - el ADN genómico se incubó con el complejo Cas9-ARN antes de la fragmentación (D). Escisión del ADN genómico humano a través del complejo Cas9-ARNcr. La cantidad relativa de fragmentos de ADN intactos se estimó mediante qPCR.

La Figura 23 ilustra esquemáticamente secuencias de direccionamiento contenidas en el plásmido informador (pMTC-DSR+eGFP). La secuencia que codifica eGFP está separada por un intrón procedente del gen GAPDH. Se indican las secuencias que codifican RFP 5' y 3'. *homol* indica secuencias homólogas en el gen RFP, necesarias para que tenga lugar una recombinación homóloga. A, B, C y D indican cuatro sitios diana distintos para la escisión mediada por Cas9. Las dianas A y B se encuentran en el intrón. Las dianas C y D se encuentran en las regiones codificadoras de eGFP. Cre indica un sitio diana para la endonucleasa Cre y se encuentra en la secuencia intrónica.

La Figura 24 muestra una reducción de las células positivas para eGFP después de la introducción de complejos Cas9/ARN. Las células CHO-K1 se transfectaron con el plásmido informador y los complejos Cas9/ARN que contenían ARNcr dirigido a la secuencia A de eGFP (intrónica), a la secuencia C de eGFP (codificadora) o a una secuencia K no específica. Se determinó el porcentaje de células positivas para eGFP mediante citometría de flujo. Como controles negativos había células no transfectadas (NC) o transfectadas solo con el plásmido informador (ADN) o solo con el plásmido informador y la proteína Cas9, así como con el plásmido informador y el complejo Cas9-ARNcr no específico (ADN+K).

La Figura 25 muestra imágenes de células en las que la aparición de RFP sugería la reparación de la rotura de la hebra doble mediada por Cas9/ARN, mediante una recombinación homóloga (HR). Cuarenta y ocho horas después de la cotransfección con el plásmido informador y los complejos Cas9/ARN dirigidos a la secuencia C de eGFP, se visualizaron células CHO-k1 mediante microscopía de fluorescencia para eGFP y RFP.

La Figura 26 ilustra esquemáticamente secuencias de direccionamiento contenidas en el plásmido informador (pMTC-DSR+eGFP). La secuencia que codifica eGFP está separada por el intrón GAPDH copiado del ADN genómico. Las secuencias que codifican N y C de RFP son las indicadas. Las secuencias homólogas en el gen RFP son necesarias para que tenga lugar una recombinación homóloga. La diana E ubicada dentro del intrón de eGFP, se indica en negrita.

La Figura 27 es un gel que muestra los complejos Cas9/ARN usando ARNcr sintético y la función ARNtracr de manera similar a los complejos Cas9/ARN usando ARNcr sintético y ARNtracr transcrito *in vitro*. Los plásmidos se visualizaron después de una electroforesis en gel de agarosa. Carril C: plásmido sin cortar. Carriles 1-3: plásmidos cortados con Cas9+ARNcr y o bien 1: ARNtracr control, transcrito *in vitro*; 2: ARNtracr sintético sin modificar (89 nt); o 3: ARNtracr sintético sin modificar (74 nt).

Las Figuras 28A-E muestran esquemáticamente secuencias de direccionamiento contenidas en el plásmido informador (pMTC-DSR+eGFP) y resultados de un procesamiento/reordenamiento genético potencial.

La Figura 29 muestra una reducción de las células positivas para eGFP después de la introducción de complejos Cas9/ARN. Las células CHO-K1 se transfectaron con el plásmido informador y los complejos Cas9/ARN que contenían ARNcr dirigido a la secuencia A de eGFP (intrónica), a la secuencia C de eGFP (codificadora) o a una secuencia K no específica. Se determinó el porcentaje de células positivas para eGFP mediante citometría de flujo. Como controles negativos había células no transfectadas (NC) o transfectadas con el plásmido informador solo (ADN) o solo con el plásmido informador y la proteína Cas9, así como con el plásmido informador y el complejo Cas9-ARNcr no específico (ADN+K).

La Figura 30 muestra esquemáticamente secuencias de direccionamiento contenidas en el plásmido informador (pMTC-DSR+eGFP). La secuencia que codifica eGFP se indica en negro y está separada por el intrón GAPDH copiado del ADN genómico. Las secuencias que codifican N y C de RFP se indican en gris. Las secuencias homólogas en el gen RFP (gris claro) son necesarias para que tenga lugar una recombinación homóloga. La diana E ubicada dentro del intrón de eGFP se indica en negrita.

Los siguientes ejemplos no limitantes describen adicionalmente los procedimientos, las composiciones, los usos y las



realizaciones.

### **Descripción detallada de realizaciones ilustrativas**

#### **Ejemplo 1**

5 En este ejemplo, hemos aislado el complejo Cas9-ARNcr del sistema CRISPR3/Cas de *S. thermophilus* y hemos demostrado que corta de manera dependiente de PAM tanto un oligodesoxinucleótido sintético como ADN plasmídico que es portador de una secuencia de nucleótidos complementaria al ARNcr. Además, hemos proporcionado una evidencia experimental de que PAM es reconocido en el contexto de ADN bicatenario y es decisivo para la unión y la escisión de ADN *in vitro*. Finalmente, hemos mostrado que los motivos RuvC y HNH de Cas9 contribuyen a la escisión de hebras de ADN opuestas. En conjunto, nuestros datos demuestran que el complejo Cas9-ARNcr actúa como una endonucleasa guiada por ARN que utiliza un módulo de ARN para el reconocimiento del sitio diana y emplea dos sitios activos distintos en el módulo de la proteína para la escisión del ADN. Estos hallazgos allanan el camino para modificar genéticamente complejos Cas9-ARNcr programables como endonucleasas guiadas por ARN universales.

#### **Materiales y procedimientos**

15 Manipulaciones del ADN. Se empleó ADN genómico de la cepa de *Streptococcus thermophilus* DGCC7710 como molde en reacciones de PCR para clonar cas9. Para generar un plásmido pASKIBA3-Cas9 que se usó para la expresión de la variante de la proteína Cas9 marcada con Strep en el extremo C-terminal, un fragmento de PCR amplificado con los siguientes cebadores: 5'-ACGTCTCAAATGTTGTTTAATAAGTGTATAATAATTTTC-3' (SEQ ID NO: 21) y 5'-ACGTCTCCGCGCTACCCTCTCCTAGTTTG-3' (SEQ ID NO: 22), se clonó en el vector de expresión pASK-IBA3 a través de los sitios Esp3I. Para generar un plásmido pBAD-Cas9 que se utilizó para la expresión de la variante de la proteína Cas9 marcada con 6xHis en el extremo C-terminal ("6xHis" desvelada como la SEQ ID NO: 23), un fragmento de PCR amplificado con la siguiente pareja de cebadores: 5'-ACGTCTCACATGACTAAGCCATACTCAATTGGAC-3' (SEQ ID NO: 24) y 5'-ACTCGAGACCCTCTCCTAGTTTGCAA-3' (SEQ ID NO: 25) se clonó en el vector de expresión pBAD24-Chis a través de los sitios NcoI y XhoI. Una secuenciación completa del gen cas9 en los plásmidos pASKIBA3-Cas9 y pBAD-Cas9 no reveló ninguna diferencia con la secuencia original de cas9. Para obtener los plásmidos pCas9(-)SP1 (Figura 1B) y pCRISPR3-SP1 (Figura 2A), que eran portadores de un solo espaciador1, un fragmento de PCR amplificado procedente del plásmido pCRISPR3 con la siguiente pareja de cebadores: 5' GACCACTTATTGAGGTAATGAG 3' (SEQ ID NO: 26)/5' CAAACCAGGATCCAAGCTAATACAGCAG-3' (SEQ ID NO: 27) (los sitios BamHI (GGATCC) están subrayados), se clonó en los plásmidos pCas9(-) y pCRISPR3 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82), respectivamente.

30 Expresión y purificación de la proteína Cas9 y el complejo Cas9-ARNcr. Una versión marcada con (His)<sub>6</sub> ("His)<sub>6</sub>" desvelada como la SEQ ID NO: 23) de la proteína Cas9 se expresó y se purificó utilizando un esquema descrito para la proteína Cas3 procedente del sistema CRISPR4/Cas de *S. thermophilus* (Sinkunas y col., 2011. EMBO J 30:1335-42). Para la purificación del complejo Cas9-ARNcr, una versión marcada con Strep de la proteína Cas9 se expresó en la cepa de *E. coli* RR1, que era portadora del plásmido pCas9(-)SP1 (Figura 1B). El medio LB se complementó con Ap (100 µg/ml) y Cm (10 µg/ml). Las células de *E. coli* para el aislamiento del complejo Cas9-ARNcr se cultivaron en dos etapas. En primer lugar, se cultivaron 4 ml de cultivo celular a 37 °C hasta una DO<sub>600</sub> de ~0,5, y se indujo la expresión mediante la adición de 0,2 µg/ml de anhidrotetraciclina (AHT) (Sigma). Después de 4 h, se inoculó 1/400 del cultivo preinducido en medio LB de nuevo aporte, complementado con Ap (100 µg/ml), Cm (12 µg/ml) y AHT (0,2 µg/ml) y se cultivó a 37 °C durante la noche. Las células recogidas se rompieron con ultrasonidos y los restos celulares se eliminaron por centrifugación. El material sobrenadante se cargó en la columna StrepTrap HP de 1 ml (GE Healthcare) y se eluyó con 2,5 mM de destiobiotina. Se obtuvieron aproximadamente 1,5 µg de la proteína Cas9 en un solo ciclo a partir de 1 L de cultivo de *E. coli*. Las fracciones que contenían Cas9 se almacenaron a +4 °C durante varios días. La homogeneidad de las preparaciones de proteína se estimó mediante SDS-PAGE. Las concentraciones de proteína en los complejos Cas9-ARNcr se determinaron mediante análisis densitométrico en geles de SDS-PAGE que contenían muestras de proteínas Cas9 purificadas con Strep-Tactin junto con cantidades conocidas de proteína Cas9 marcada con His. La concentración de los complejos Cas9-ARNcr se expresa como la concentración de proteína Cas9, suponiendo que Cas9 es un monómero y se une al ARNcr en un complejo con una estequiometría de 1:1.

50 Análisis por transferencia Northern. El ARN unido a Cas9 se aisló a partir de Cas9 purificada con Strep-Tactin, coexpresada con el plásmido pCas9(-)SP1, usando el mini kit miRNeasy (Qiagen). Las transferencias Northern se realizaron haciendo correr el ARN sobre un gel de poliacrilamida al 10% con urea 7 M en tampón MOPS/NaOH 20 mM pH 8. El ARN se transfirió a una membrana de nailon SensiBlot™ Plus (Fermentas) mediante transferencia semiseca utilizando Trans-blot SD (Bio-Rad). El ARN se reticuló con la membrana con l-etil-3-(3-dimetilaminopropil) carbodiimida 0,16 M (EDC) (Pierce)/1-metilimidazol 0,13 M (Sigma) pH 8 a 60 °C durante 1 h. La membrana se prehibridó con tampón 2 × SSC que contenía SDS al 1% y 0,1 mg/ml de ADN desnaturizado procedente de testículos de peces (Ambion) durante 1 h a 40 °C. Las transferencias se sometieron a ensayo durante 12 h con un oligonucleótido de ADN anti-ARNcr de 42 nt, marcado con <sup>32</sup>P-5' que contenía 20 nt del espaciador1 y 22 nt de la secuencia de repetición (5'-TCGAAACAACACAGCTCTAAAACCTGTCCTCTTCTTTAGC-3' (SEQ ID NO: 28)). Las transferencias se lavaron 3 × durante 15 minutos con tampón 0,2 × SSC que contenía SDS al 0,2%, y se visualizaron usando la formación de imágenes de fósforo. Un oligorribonucleótido sintético de 42 nt (5'-CGCUAAAGAGGAAGAGGACAGUUUUAGA-GCUGUGUUUUUCG-3' (SEQ ID NO: 7)) y un oligonucleótido de ADN de 84 nt.

- 5 Sustratos de oligonucleótidos. Todos los sustratos de oligonucleótidos utilizados en este estudio se proporcionan en la Tabla 1. Los oligodesoxirribonucleótidos se compraron en Metabion (Martinsried, Alemania). Los extremos 5' de los oligonucleótidos se radiomarcaron usando PNK (Fermentas) y [ $\gamma$ -<sup>32</sup>P]ATP (Hartmann Analytic). Los dúplex se prepararon apareando dos oligonucleótidos con secuencias complementarias (SP1, SP1- $\Delta$ p, SP2). El marcador radiactivo se introdujo en el extremo 5' de una hebra de ADN individual antes del apareamiento con la hebra sin marcar.
- 10 Reacciones con sustratos de oligonucleótidos. Las reacciones se llevaron a cabo normalmente añadiendo 2 nM de complejo Cas9-ARNcr a un oligonucleótido 1 nM marcado en Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 10 mM, 0,1 mg/ml de BSA y MgCl<sub>2</sub> 10 mM a 37 °C. Se extrajeron partes alícuotas a intervalos programados y se inactivaron con colorante de carga (formamida al 95% v/v, azul de bromofenol al 0,01%, EDTA 25 mM, pH 9,0) y se sometieron a electroforesis en gel desnaturizante a través de poli(acrilamida) al 20%, seguido de detección mediante formación de imágenes de fósforo con FLA-5100 (Fujifilm).
- 15 Reacciones con sustratos plasmídicos. Las reacciones con el plásmido pUC18 y sus derivados (Sapranaukas y col., 2011. *Nucleic Acids Res* 39:9275-82) se llevaron a cabo a 37 °C en el tampón utilizado para las reacciones con sustratos de oligonucleótidos. Las mezclas de reacción normalmente contenían plásmido superenrollado 2,5 nM y complejo Cas9-ARNcr 2 nM. Las reacciones se iniciaron añadiendo proteína a la mezcla de los otros componentes. Se extrajeron partes alícuotas a intervalos programados y se inactivaron con fenol/cloroformo. La fase acuosa se mezcló con una solución de colorante de carga (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y se analizó mediante electroforesis a través de agarosa.
- 20 Determinación de la posición del corte en el plásmido. Para lograr un corte completo del sustrato plasmídico, se incubaron 8 nM de complejo Cas9-ARNcr con 2,5 nM de plásmido superenrollado en el tampón de reacción a 37 °C durante 10 min. Los productos de la reacción se purificaron y se concentraron usando el kit de purificación de PCR GeneJET (Fermentas). La región que circundaba al espaciador1 de Cas9 linealizada y los plásmidos mellados se secuenciaron directamente con los siguientes cebadores: 5'-ccgcatcaggcgccattcgcc-3' (SEQ ID NO: 29) (secuenciación de la hebra (+)) y 5'-gcgaggaagcggaagagcgccc-3' (SEQ ID NO: 30) (secuenciación de la hebra (-)).
- 25 Ensayo de la unión. Cantidades crecientes de complejo proteína-ARNcr se mezclaron con 0,5 nM de sustratos de ADN bicatenarios y monocatenarios marcados con <sup>32</sup>P (Tabla 1) en el tampón de unión (Tris-acetato 40 mM, pH 8,3 a 25 °C, EDTA 0,1, 0,1 mg/ml de BSA, glicerol al 10% v/v) y se incubaron durante 15 minutos a temperatura ambiente. Los complejos exentos de ADN y proteína-ADN se separaron sobre el gel de poli(acrilamida) al 8% no desnaturizante (relación de acrilamida/N,N'-metilbisacrilamida 29:1) usando Tris-acetato 40 mM (pH 8,3) complementado con EDTA 0,1 mM como tampón de ejecución. La electroforesis se realizó a temperatura ambiente durante 3 h a 6 V/cm.
- 30 Mutagénesis. Los mutantes D31A y N891A se obtuvieron mediante mutagénesis dirigida al sitio como se ha descrito previamente (Tamulaitis y col., 2007. *Nucleic Acids Res* 35:4792-9). La secuenciación del gen completo para cada mutante confirmó que solo se había introducido la mutación diseñada.

**Tabla 1.** Sustratos de oligonucleótidos. La secuencia del protoespaciador está subrayada, PAM está en negrita.

Oligonucleótido	Secuencia	Descripción
SP1 (SEQ ID NO: 31)	5' - GCTCGAATTGAAATTTAAACGGCTAAAGAGGAAAGGAGGACATGGTGAATTCGTAAT - 3' 3' - CGAGCTTAACITTAAGATTTGCGATTTCTCCITTCCTCCGTACCACITTAAGCATT - 5'	sustrato del oligodúplex de 55 pb que contiene el protoespaciador1 y PAM
SP1-pΔ (SEQ ID NO: 32)	5' - GCTCGAATTGAAATTTAAACGGCTAAAGAGGAAAGGAGGACAAAATTCGTAAT - 3' 3' - CGAGCTTAACITTAAGATTTGCGATTTCTCCITTCCTCCITTAAGCATT - 5'	sustrato del oligodúplex de 50 pb que contiene el protoespaciador2
SP2 (SEQ ID NO: 33)	5' - GCTCGAATTGACTGCTGTATTAGCTTGGTTGGTTTGGTGAATTCGTAAT - 3' 3' - CGAGCTTAACATGACGACATAATGCAACCAACAACCAACCCACTTAAGCATT - 5'	sustrato del oligodúplex de 55 pb que contiene el protoespaciador2 y PAM (oligodúplex sin protoespaciador1)
s(+) SP1 (SEQ ID NO: 34)	5' - ATTACGAAATTCACCATGTCCCTCTCCCTCTTTAGCGTTTAGAATTTCAATTCGAGC - 3'	sustrato del oligonucleótido de ADNss de 55 nt, hebra (+) del oligodúplex SP1
s(+) SP1-pΔ (SEQ ID NO: 35)	5' - ATTACGAAATTTGTCCTCTCCCTCTTTAGCGTTTAGAATTTCAATTCGAGC - 3'	sustrato del oligonucleótido de ADNss de 50 nt, hebra (+) del oligodúplex SP1-pΔ
s(+) SP2 (SEQ ID NO: 36)	5' - ATTACGAAATTCACCAACAACCAACAACCAAGCTAAATACAGCAGTACAAATTCGAGC - 3'	sustrato del oligonucleótido de ADNss de 55 nt, hebra (+) del oligodúplex SP2
s(-) SP1 (SEQ ID NO: 37)	5' - GCTCGAATTTGAAATTTAAACGCTAAAGAGGAAAGGAGACATGGTGAATTCGTAAT - 3'	sustrato del oligonucleótido de ADNss de 55 nt, hebra (-) del oligodúplex SP1
SP1-20 (SEQ ID NO: 38)	5' - GCTCGAATTTGCGCTAAAGAGGAAAGGAGACATGGTGAATTCGTAAT - 3' 3' - CGAGCTTAACGCGAATTTCTCCITTCCTCCGTACCACITTAAGCATT - 5'	sustrato del oligodúplex de 45 nt que contiene 20 nt del protoespaciador1 y PAM
SPN (SEQ ID NO: 39)	5' - GCTCGAATTTGCCACCCAGCAAAAATTCGGTTTTCTGGCTGATGGTGAATTCGTAAT - 3' 3' - CGAGCTTAACGGTGGGTCGTTTTAAGCCAAAAGACCCGACTACCACITTAAGCATT - 5'	sustrato del oligodúplex de 55 pb que contiene el protoespaciadorN y PAM

## Resultados

Expresión y purificación del complejo Cas9-ARNcr. El gen *cas9* procedente del sistema CRISPR3 de la cepa DGCC7710 de *S. thermophilus* se clonó en el vector pASK-IBA3 para producir una estructura artificial que codificaba una proteína Cas9 de fusión que contenía un marcador Strep(II) en el extremo C-terminal (Figura 1B). Inicialmente, tratamos de purificar el complejo Cas9-ARNcr a partir de la cepa de *E. coli* RR1 que expresaba la proteína Cas9 sobre el vector pASK-IBA3, y otras proteínas Cas (excepto Cas9) sobre el plásmido pCas9(-) (Sapranaukas y col., 2011). pCas9(-) también contenía una disposición CRISPR3 completa compuesta por 12 unidades de espaciador-repetición (Figura 2A). Para lograr una transcripción simultánea de todos los genes diana, realizamos la expresión del gen *cas9* en dos etapas. En primer lugar, indujimos la expresión de Cas9 en un volumen pequeño de cultivo de *E. coli* y después de 4 h transferimos una parte alícuota de cultivo preinducido a un volumen mayor de medio LB de nuevo aporte que ya contenía inductor y se incubó durante la noche. El complejo de proteína Cas9 se purificó a partir del extracto celular bruto usando sefarosa Strep-Tactin. Logramos aislar una pequeña cantidad de complejo Cas9-ARNcr que únicamente mostraba rastros de actividad nucleolítica sobre el oligodúplex SP1 que contenía un protoespaciador1 y PAM. Asumimos que la baja actividad de escisión se podía deber a la heterogeneidad intrínseca de los complejos Cas9-ARNcr resultantes de la transcripción de 12 unidades de espaciador-repetición. Si todas las unidades de espaciador-repetición se transcriben uniformemente en un ARNcr maduro, la concentración del complejo Cas9 que contiene un ARNcr contra el espaciador-1, constituirá una fracción de un 1/12 de la concentración total de Cas9-ARNcr. La actividad de escisión de la preparación de Cas9-ARNcr contra el oligodúplex SP2 que contiene un protoespaciador-2 y PAM, es compatible con la heterogeneidad de los complejos Cas9-ARNcr (Figura 2B). Para aumentar el rendimiento del complejo Cas9-ARNcr específico, diseñamos un plásmido pCas9(-)SP1 que contenía una sola unidad de R-espaciador1-R en la disposición de CRISPR (Figura 1B). El ensayo de interferencia de la transformación del plásmido confirmaba que el sistema CRISPR3/Cas que es portador de un solo espaciador1, evita la transformación del plásmido pSP1 en *E. coli* con la misma eficacia que el sistema CRISPR3/Cas que es portador de una región CRISPR completa (Figura 3B). Hemos aislado un complejo Cas9-ARNcr siguiendo el procedimiento descrito anteriormente y hemos analizado el ARNcr unido a la proteína Cas9.

La proteína Cas9 se purifica conjuntamente con ARNcr. El sistema CRISPR3/Cas de *S. thermophilus* pertenece al subtipo Tipo IIA (antiguo Nmeni o CASS4) de los sistemas CRISPR/Cas (Makarova y col., 2011. Nat Rev Microbiol 9:467-77). Se ha observado que en el sistema CRISPR/Cas de Tipo IIA de *Streptococcus pyogenes*, el ARN pequeño transcrito (ARNtracr) y la RNasaIII bacteriana están implicados en la generación del ARNcr (Deltcheva y col., 2011. Nature 471:602-7). El ARNcr de *Streptococcus pyogenes* tiene solo 42 nt de longitud y no tiene un "asa 5'" que se conserva en los ARNcrs de los sistemas CRISPR de Tipo I y III (Hale y col., 2009. Cell 139:945-56; Jore y col., 2011. Nat Struct Mol Biol 18:529-36). De acuerdo con el análisis de transferencia Northern, se genera un ARNcr de longitud similar en el sistema LMD-9 CRISPR3/Cas de *Streptococcus thermophilus* (Makarova y col., 2011. Nat Rev Microbiol 9:467-77), que es casi idéntico al sistema CRISPR3/Cas de la cepa DGCC7710 (Figuras 4A y B). Asumimos que el ARNcr aislado a partir del complejo Cas9-ARNcr, expresado en la cepa heteróloga de *E. coli* (Figura 1) puede tener la misma longitud (Figura 4). Por lo tanto, para analizar ácidos nucleicos extraídos a partir del complejo Cas9 purificado con Strep-Tactin, utilizamos un oligonucleótido de ADN anti-ARNcr de 42 nt, que comprendía la región de 22 nt correspondiente al extremo 3' de la secuencia de la repetición y 20 nt en el extremo 5' del fragmento SP1. El ácido nucleico presente en el complejo Cas9 se hibridaba con un oligonucleótido anti-ARNcr, y era sensible al tratamiento con RNasa pero no con DNasa (Figura 1C). El tamaño del ARNcr extraído era idéntico al oligorribonucleótido sintético de 42 nt correspondiente al ARNcr putativo del sistema CRISPR3 de la cepa DGCC7710 de *S. thermophilus* (Figura 3A, Figura 4C). Tomados en conjunto, estos datos confirman que la proteína Cas9 marcada con Strep se purifica junto con un ARNcr de 42 nt, que se obtiene a partir de la región CRISPR3.

La proteína Cas9 escinde el ADN bicatenario dentro de un protoespaciador. Para analizar la actividad *in vitro* de un complejo Cas9-ARNcr purificado, primero usamos el oligodúplex SP1 (Tabla 1) que contiene la secuencia del protoespaciador idéntica al espaciador SP1 en la disposición CRISPR3, la secuencia PAM 5'-TGGTG-3' corriente abajo del protoespaciador y secuencias flanqueantes de 10 nt procedentes del plásmido pSP1 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) (Figura 5A). La hebra del oligodúplex complementaria al ARNcr se denomina hebra (+), mientras que la hebra dúplex opuesta se denomina hebra (-). Para controlar la reacción de escisión, la hebra (+) o (-) del oligodúplex SP1 se marcó con P33 en el extremo 5'. Los datos mostrados en la Figura 5B demuestran que el complejo Cas9-ARNcr corta ambas hebras del oligodúplex en una posición fija. El cartografiado de la posición del corte utilizando oligonucleótidos sintéticos como marcadores del tamaño, reveló que el complejo Cas9-ARNcr corta ambas hebras del oligodúplex SP1 dentro del protoespaciador, 4 nt corriente arriba del PAM (Figura 5B) dejando extremos romos. Cabe mencionar que no se observa ninguna escisión después de 2 h de incubación del oligodúplex SP1 con la proteína Cas9 que carece de ARNcr (Figura 6C).

Para someter a ensayo si el complejo Cas9-ARNcr puede localizar el protoespaciador y cortar el ADN *in vitro* en sustratos de ADN largos que imitan un ADN extraño invasor *in vivo*, analizamos la escisión del plásmido pSP1 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) (Figura 5C) que era portador del protoespaciador1 y PAM. En presencia del complejo Cas9-ARNcr, la forma superenrollada del plásmido pSP1 se convertía en una forma lineal (Figura 5D), mientras que el plásmido pUC18 que carecía de protoespaciador1 no se escindía. Esto significa que ambas hebras del plásmido pSC1 se escindían específicamente dentro de la región del protoespaciador. Utilizamos una secuenciación directa para determinar los extremos de la forma de ADN lineal, formada después de la escisión con Cas9-ARNcr. Los resultados de la secuenciación confirmaron que la escisión del ADN plasmídico se producía a

una distancia de 4 nt de la secuencia de PAM, de manera similar a la escisión del oligodúplex SP1 (Figura 5D). Las posiciones de la escisión identificadas en experimentos *in vitro* (Figura 4) para el sistema CRISPR3/Cas de *S. thermophilus* son idénticas a las determinadas en experimentos de escisión *in vivo* para el sistema CRISPR1/Cas en *S. thermophilus* (Garneau y col., 2010. Nature 468:67-71). Para verificar si la escisión inducida por Cas9-ARNcr ocurre en la misma posición en otras secuencias de protoespaciador, analizamos la escisión del oligodúplex SP2 que es portador de un protoespaciador-2 y secuencias PAM a través del complejo heterogéneo Cas9-ARNcr aislado a partir del hospedador que es portador de 12 unidades de espaciador-repetición. Hemos encontrado que ese complejo heterogéneo Cas9-ARNcr corta la hebra (+) del oligodúplex SP2 exactamente en la misma posición que en el oligodúplex SP1. La especificidad de la escisión con Cas9-ARNcr está dirigida por la secuencia de ARNcr. Para demostrar directamente que la especificidad del complejo Cas9-ARNcr se puede reprogramar cambiando el ARNcr en el complejo de ribonucleoproteína, insertamos un nuevo espaciador (SN) en lugar del espaciador S1 en la región CRISPR, generando el plásmido pCas(-)SN que contenía solo una región CRISPR mínima y una secuencia codificadora de ARNcr (Figura 7), coexpresamos este plásmido junto con pASKIBA-Cas9 y purificamos el complejo Cas9-ARNcr. La especificidad de la escisión del complejo Cas9-ARNcr se analizó usando los plásmidos pSP1+SPN y pSP1. El plásmido pSP1+SPN que contiene la secuencia del protoespaciador que se aparea con el espaciador SN en la región CRISPR, se linealizó a través del complejo Cas9-ARNcr, mientras que el plásmido pSP1 que carece de una secuencia complementaria, permaneció intacto (Figura 7B). Para determinar la posición de la escisión dentro de la secuencia del espaciador SPN, realizamos experimentos con oligodúplex SPN, que contenía un protoespaciador complementario al espaciador SN y PAM (Figura 7D). El ensayo de escisión del oligodúplex confirmaba (Figura 7C y D) que el complejo Cas9-ARNcr con una especificidad rediseñada, escinde ambas hebras de ADN dentro del protoespaciador SN, 4 nt corriente arriba de PAM, de forma idéntica a otros complejos Cas9-ARNcr.

La longitud del espaciador en la región CRISPR3 de *S. thermophilus* es de 30 nt. De acuerdo con los datos proporcionados en la Figura 1C, el ARNcr maduro copurificado con la proteína Cas9 está compuesto por 42 nt. Esto significa que solo 20 nt de ARNcr son complementarios a la hebra (+) del protoespaciador. Para evaluar si el extremo 5' del protoespaciador es importante para la interferencia del plásmido a través del sistema CRISPR3 de *S. thermophilus*, diseñamos los plásmidos pSP1-27, pSP1-23, pSP1-19, pSP1-15, pSP1-11 con el protoespaciador1 truncado en 5' (la longitud del protoespaciador era de 27 pb, 23 pb, 19 pb, 15 pb, 11 pb, respectivamente), y analizamos la eficacia de la transformación de la cepa receptora que contenía pCRISPR3 (Figura 8B). Los plásmidos que contenían truncamientos de 4 o 7 pb en el extremo 5' del protoespaciador1, no tenían ningún efecto sobre la capacidad de la cepa receptora para interferir con la transformación del plásmido. Las versiones más cortas de protoespaciador (11, 15, 19 pb) anulaban la capacidad de la cepa receptora para evitar una transformación del plásmido. Estos datos muestran que el extremo 5' del protoespaciador, que no tiene complementariedad con el ARNcr maduro, no es importante para la función de CRISPR3/Cas. Respaldo totalmente los experimentos *in vivo*, el oligodúplex SP1-20 que contiene solo 20 nt del protoespaciador-1, es escindido con eficacia con Cas9-ARNcr (Figura 8D y E).

Se requiere PAM para la unión y la escisión del ADN a través de Cas9-ARNcr. Los plásmidos que son portadores de un protoespaciador pero no de PAM (pSP1-pΔ) o múltiples PAMs, pero no tienen protoespaciador (pUC18), son resistentes a la escisión con Cas9-ARNcr (Figura 8A). Por lo tanto, de acuerdo con los datos *in vivo*, se requiere tanto PAM como un protoespaciador para la escisión de un ADN bicatenario a través del complejo Cas9-ARNcr (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82). Para averiguar si PAM es reconocido en el contexto de un ADN bicatenario o monocatenario, analizamos la unión de Cas9-ARNcr y la escisión de oligodesoxinucleótidos i) SP1 (que contiene tanto protoespaciador como PAM), ii) SP1-Δp (contiene solo protoespaciador) y iii) SP2 (contiene solo PAM). Las hebras (+) de estos oligodesoxinucleótidos se usaron como sustratos de ADN monocatenario (s(+))SP1, s(+))SP1-Δp, s(+))SP2, en consecuencia) (Tabla 1).

De acuerdo con los experimentos de escisión de plásmidos, los oligodúplex que tienen solo protoespaciador, pero no tienen PAM, no se escinden a través de Cas9-ARNcr (Figura 9B). Por otro lado, la hebra (+) en la forma monocatenaria se corta a una velocidad similar, independientemente de si tiene o no PAM (Figura 9B). Estos datos muestran claramente que se requiere PAM solo para una escisión del ADN bicatenario pero no para una hebra monocatenaria.

Para someter a ensayo si PAM es importante para la unión del ADN a través del complejo Cas9-ARNcr, se realizaron experimentos de cambio de la movilidad electroforética. Para evitar una escisión, se realizaron experimentos de unión en ausencia de iones Mg<sup>2+</sup> que son necesarios para la escisión. Cas9-ARNcr mostraba diferentes patrones de unión para oligonucleótidos bicatenarios y monocatenarios. En el caso del oligodúplex SP1, se observa un complejo de baja movilidad ya con una concentración de 1 nM (Figura 9C). Por otro lado, no se observa unión con las mismas condiciones experimentales en los oligodúplex sin PAM (SP1-Δp) o sin protoespaciador (SP2). Además, no se observa un complejo de baja movilidad en el caso de una proteína Cas9 sin ARNcr (Figura 6A), lo que confirma que el ARNcr es importante para la formación del complejo. Por lo tanto, en conjunto, los experimentos de unión muestran claramente que el complejo de la proteína Cas9 no se puede unir al ADN bicatenario en ausencia de PAM, incluso si contiene un ARNcr complementario al protoespaciador. Dicho con otras palabras, los sustratos de ADN bicatenarios que carecen de PAM no se escinden porque se requiere PAM para la unión de Cas9-ARNcr.

Por otro lado, los oligonucleótidos monocatenarios (hebra (+)) están unidos mediante Cas9-ARNcr con la misma afinidad, independientemente de la presencia de PAM (Figura 9D). Nuevamente, no se observó una unión del oligonucleótido de ADN monocatenario sin protoespaciador (Figura 9D), o de la proteína Cas9 que carecía de ARNcr (Figura 6C). En conjunto, estos datos indican que el complejo Cas9-ARNcr discrimina PAM solo en el ADN bicatenario pero

no en el monocatenario.

Dado que algunos sistemas CRISPR de Tipo III proporcionan una interferencia del ARN en lugar del ADN, hemos estudiado la unión y escisión de ARN a través del complejo Cas9-ARNcr. El Cas9-ARNcr no escindía específicamente ni el ARN monocatenario ni el ARN bicatenario que era portador de un protoespaciador y PAM (Figura 10B). Este hallazgo confirma una vez más que el ADN es una diana principal para el sistema CRISPR3/Cas de *S. thermophilus*. El complejo Cas9-ARNcr se une a un ARN complementario que contiene un protoespaciador, pero esa interacción probablemente no es funcionalmente importante, porque un ARN monocatenario no se corta específicamente con Cas9 dentro de un protoespaciador.

Mutagénesis de los motivos RuvC y HNH de la proteína Cas9. Los experimentos de transformación de plásmidos indican que los motivos RuvC y HNH (Figura 11A) son importantes para la función de Cas9 (Sapranauskas y col., 2011. *Nucleic Acids Res* 39:9275-82). Para someter a ensayo si estos motivos están implicados en la escisión del ADN diana, expresamos y purificamos los mutantes D31A y N891A siguiendo un procedimiento descrito para Cas9 wt. Ambos mutantes se purificaron conjuntamente con ARNcr idéntico al ARNcr en el complejo Cas9 wt (Figura 11C). Para someter a ensayo si las proteínas mutantes conservaban la actividad de escisión, realizamos un seguimiento de la escisión del plásmido pSP1 mediante complejos Cas9-ARNcr mutantes. Sorprendentemente, en lugar del producto de reacción lineal observado para la proteína Cas9 wt, ambos mutantes producían una forma de ADN mellada (Figura 11B), lo que indica que ambos mutantes de los sitios activos escinden solo una hebra de ADN del sustrato plasmídico dentro de un protoespaciador.

Para determinar si las proteínas mutantes muestran una preferencia por una hebra, analizamos la escisión del mutante D31A y N891A del oligodúplex SP1. El mutante del sitio activo RuvC (D31A) cortaba la hebra (+) del oligodúplex en la misma posición que la proteína Cas9-ARNcr wt, mientras que la hebra (-) permanecía intacta (Figura 11C). Y viceversa, el mutante del sitio activo HNH (N891A) cortaba solo la hebra (-), pero no la hebra (+) del oligodúplex SP1 (Figura 11D). Tomados en conjunto, estos datos indican que los sitios activos RuvC y HNH actúan sobre hebras de ADN opuestas para generar una rotura de la hebra doble. Para someter a ensayo si se conserva el mismo patrón de corte durante la escisión del ADN plasmídico, secuenciamos regiones de protoespaciador de plásmidos mellados. Los datos de secuencias "run-off" confirmaron que el mutante del sitio activo RuvC cortaba solo la hebra (+) del ADN, mientras que el mutante HNH/McrA solo la hebra (-) (Figura 12A y B). Además, encontramos que el mutante RuvC cortaba la hebra (+) de un ADN monocatenario pero no se detectó tal corte para el mutante HNH (Figura 12D).

Para someter a ensayo si las mutaciones alteraban la afinidad de la unión al ADN de los complejos mutantes de proteína-ARNcr, se estudió la unión del ADN utilizando el ensayo de cambio de la movilidad electroforética. Ambos complejos mutantes de proteína-ARNcr se unen al oligodúplex SP1 con la misma afinidad que la proteína de tipo silvestre (wt, del inglés "wild type") (Figura 12C). Por lo tanto, las mutaciones en los sitios activos putativos de Cas9 no tienen un efecto significativo sobre las propiedades de unión al ADN de hebra doble del complejo Cas9-ARNcr. Dado que el ARNcr de 42 nt estaba presente en los complejos de proteínas mutantes (Figura 12C), llegamos a la conclusión de que los complejos mutantes de Cas9-ARNcr perdían la capacidad de cortar una de las hebras del ADN diana debido a una mutación del sitio activo. Dado que la proteína Cas9-marcador His es un monómero en solución (Figura 13), es probable que la proteína Cas9 sea funcional como monómero y utilice dos sitios activos para el corte de hebras de ADN opuestas. Una estrategia similar es explotada por algunas endonucleasas de restricción (Armalyte y col., 2005. *J Biol Chem* 280 41584-94).

#### 40 Discusión

El complejo Cas9-ARNcr del sistema CRISPR3/Cas de *S. thermophilus* es una endonucleasa guiada por ARNcr. Este trabajo demuestra que el complejo Cas9-ARNcr del sistema CRISPR3/Cas de *S. thermophilus* es una endonucleasa dirigida por ARNcr que corta ambas hebras de ADN en presencia de iones Mg<sup>2+</sup> dentro de un protoespaciador situado 4 nt corriente abajo de la secuencia PAM, para producir productos de escisión con extremos romos. La especificidad de secuencia del complejo Cas9-ARNcr viene dictada por el ARNcr de 42 nt que incluye un fragmento de ~20 nt complementario a la secuencia del protoespaciador en el ADN diana. A este respecto, el ARNcr maduro en el complejo Cas9 del sistema CRISPR3/Cas de *S. thermophilus*, es similar al ARNcr de *Streptococcus pyogenes* que tiene un asa en 3' de una secuencia de repetición, pero carece de parte de la secuencia del espaciador y del asa en 5' correspondiente al fragmento de repetición (Deltcheva y col., 2011). Por lo tanto, el ARNcr presente en el complejo Cas9-ARNcr del sistema CRISPR3/Cas de *S. thermophilus* es complementario solo a la parte de la secuencia de protoespaciador distal de PAM. No es sorprendente que el truncamiento del extremo 3' de la secuencia del protoespaciador de 10 nucleótidos no tenga ningún efecto sobre la escisión con Cas9-ARNcr de oligodúplex sintéticos o ADN plasmídico (Figura 8).

La maquinaria para el corte del complejo Cas9-ARNcr reside en la proteína Cas9 que proporciona dos sitios activos para la escisión del enlace fosfodiéster. Los sitios activos similares a RuvC y HNH de la proteína Cas9 están ubicados en diferentes dominios y actúan independientemente sobre hebras de ADN individuales. La sustitución de la alanina de los restos del sitio activo en los motivos RuvC y HNH, transforma el complejo Cas9-ARNcr en una endonucleasa que realiza una mella específica de la hebra de forma similar a las enzimas mellasas (Chan y col., 2011. *Nucleic Acids Res* 39:1-18). De forma consistente con estudios *in vivo*, una actividad funcional del complejo Cas9-ARNcr *in vitro* es absolutamente dependiente de la presencia del motivo adyacente al protoespaciador NGGNG, corriente arriba de la

secuencia del protoespaciador. Los datos presentados en la Figura 3 muestran que se requiere PAM para la unión de Cas9-ARNcr al ADN bicatenario. Si falta la secuencia de PAM en el ADN bicatenario, el complejo Cas9-ARNcr no se une a ese ADN, incluso si contiene una secuencia de protoespaciador complementaria. Por otro lado, Cas9-ARNcr no muestra una unión al ADN si PAM (o múltiples PAMs) está presente, pero la secuencia de protoespaciador está ausente. Por lo tanto, de forma compatible con los datos *in vivo*, tanto las secuencias de PAM como del protoespaciador son requisitos previos necesarios para la unión al ADN bicatenario y la posterior escisión. Contrariamente a la unión de Cas9-ARNcr al ADN bicatenario, el motivo de la secuencia PAM no tiene ningún efecto sobre la unión al ADN monocatenario ya que: un oligodesoxinucleótido monocatenario que contiene un protoespaciador con o sin secuencia PAM, se une igualmente bien pero con menor afinidad que al ADN bicatenario. En presencia de iones Mg<sup>2+</sup>, Cas9 corta un ADN monocatenario unido al ARNcr usando su sitio activo HNH.

Mecanismo de interferencia del ADN en los sistemas de Tipo II. Nuestros resultados establecen un modelo simple para el mecanismo de corte del ADN bicatenario a través del complejo Cas9-ARNcr en el sistema CRISPR3/Cas de *S. thermophilus* (Figura 14). Los complejos Cas9-ARNcr que utilizan un mecanismo que todavía debe definirse, localizan y se unen a una secuencia de protoespaciador dentro del ADN bicatenario en un procedimiento dependiente de PAM. Es posible que PAM en el ADN bicatenario sirva como un sitio de iniciación (señal) para la separación de hebras y promueva un emparejamiento posterior del ARNcr con la hebra (+) complementaria del ADN. Queda por establecer si un módulo de proteína Cas9 o un ARNcr unido a Cas9 (por ejemplo, usando nucleótidos en el "asa 3'" conservada de la secuencia de repetición conservada) reconoce la secuencia PAM. A pesar de la falta de estos detalles del mecanismo, nuestros datos demuestran claramente que PAM es reconocido por Cas9-ARNcr en el contexto de un ADN bicatenario. La unión de Cas9-ARNcr a la secuencia diana en el ADNds presumiblemente da como resultado la estructura de bucle R, en la que la hebra (-) se desplaza y la hebra de ADN complementaria (+) se empareja con el ARNcr. En presencia de iones Mg<sup>2+</sup>, la escisión del enlace fosfodiéster se produce en ambas hebras, 4 nt corriente arriba en 5' de la secuencia PAM para generar extremos romos en el ADN. El análisis de la escisión del ADN a través de los mutantes de los motivos RuvC o HNH demuestra que los sitios activos similares a RuvC y HNH de la proteína Cas9, actúan sobre las hebras (-) y (+), respectivamente. Por lo tanto, en el complejo Cas9-ARNcr catalíticamente competente, el dominio N-terminal que contiene el resto catalítico D31A del motivo RuvC, está situado en la hebra de ADN (-) desplazada, mientras que la parte central de Cas9 que contiene el motivo HNH, está localizada próxima al enlace fosfodiéster escindible de la hebra de ADN (+) emparejada con el ARNcr. Después de la escisión del ADN, Cas9-ARNcr permanece unido a los productos de reacción (Figura 15). En conjunto, los datos presentados en este documento sugieren un primer mecanismo molecular para la etapa de interferencia del ADN a través del sistema CRISPR3/Cas de *S. thermophilus*. Ya que Cas9 es un gen distintivo (Makarova y col., 2011. *Nat Rev Microbiol* 9:467-77) para los sistemas de Tipo IIA y Tipo IIB, el mecanismo de corte propuesto en este documento es probable que se conserve en otros sistemas de Tipo IIA y Tipo IIB. Versiones independientes de proteínas similares a Cas9 que no forman parte del sistema CRISPR, se han identificado mediante bioinformática (Makarova y col., 2011. *Biol Direct* 6:38). En vista de los datos proporcionados en este documento, sugerimos que esas proteínas pueden proporcionar una interferencia contra un ADN extraño de manera similar a Cas9, si se cargan con moléculas pequeñas de ARNcr que se pueden generar a través de la ruta diferente de CRISPR.

Comparación con otros complejos de interferencia de ARN. El mecanismo propuesto en este documento para la escisión de ADN bicatenario a través del complejo Cas9-ARNcr difiere significativamente del sistema de Tipo I-E (antiguo *E. coli* o CASS2) (Jore y col., 2011. *Nat Struct Mol Biol* 18:529-36). En el sistema de *E. coli*, ARNcr y las proteínas Cas se ensamblan en un gran complejo de ribonucleoproteína denominado Cascade que facilita el reconocimiento de la diana al mejorar la hibridación específica de secuencia entre el ARN CRISPR y las secuencias diana complementarias (Jore y col., 2011. *Nat Struct Mol Biol* 18:529-36). El reconocimiento de la diana depende de PAM y se rige por la secuencia de ARNcr "semilla" ubicada en el extremo 5' de la región del espaciador (Semenova y col., 2011. *Proc Natl Acad Sci USA* 108:10098-103) Sin embargo, aunque el complejo Cascade-ARNcr solo es capaz de unirse a ADN bicatenario que contiene PAM y un protoespaciador, requiere una proteína Cas3 accesoria para la escisión del ADN. Cas3 es una nucleasa y helicasa de ADN monocatenario que es capaz de escindir el ADN monocatenario produciendo múltiples cortes (Sinkunas y col., 2011. *EMBO J* 30:1335-42). Los detalles del mecanismo de acción de Cas3 sobre un sustrato biológico adecuado (por ejemplo, Cascade-ARNcr unido al ADN bicatenario en el complejo de tipo bucle R) aún no se han establecido. Sin embargo, se ha demostrado recientemente que Cas3 de *M. jannaschii* solo es capaz de cortar ambas hebras de ADN en el sustrato sintético que imita el bucle R (Beloglazova y col., 2011. *EMBO J* 30:616-27). Se ha propuesto que Cas3 puede seguir un mecanismo similar para la escisión del ADN en presencia del complejo Cascade-ARNcr. Por lo tanto, los datos actuales muestran claramente que los detalles del mecanismo de la etapa de interferencia para el sistema de Tipo I-E, difieren de los del sistema CRISPR3 tanto por la maquinaria catalítica como por el mecanismo y la complejidad.

En los sistemas CRISPR de subtipo III-B presentes en muchas arqueas y algunas bacterias, las proteínas RAMP de módulo Cas (Cmr) y el ARNcr se ensamblan en el complejo efector que se dirige hacia el ARN invasor (Hale y col., 2009. *Cell* 139:945-56; Hale y col., 2012. *Mol Cell* 45:292-302). En *Pyrococcus furiosus*, el complejo silenciador de ARN está compuesto por seis proteínas Cmr1-6 y el ARNcr se une al ARN diana y lo corta a una distancia fija con respecto al extremo 3' del psiARN. La actividad de corte depende de los iones Mg<sup>2+</sup>, sin embargo, la o las proteínas Cmr individuales responsables de la escisión del ARN diana aún no se han identificado. El complejo efector de *Sulfolobus solfataricus* compuesto por siete proteínas Cmr1-7 y ARNcr, corta el ARN invasor en una reacción endonucleolítica en los dinucleótidos UA (Zhang y col., 2012. *Mol Cell* 45:303-13). Es importante destacar que ambos complejos Cmr-

ARNcr realizan la escisión del ARN de manera independiente de PAM.

Los datos proporcionados en esta memoria muestran que el complejo Cas9-ARNcr del sistema CRISPR3 es hasta ahora el sistema de interferencia de ADN más simple, compuesto por una sola proteína Cas9 unida a la molécula de ARNcr. La organización modular simple del complejo Cas9-ARNcr, en donde la especificidad hacia la diana de ADN está codificada por el ARNcr y la maquinaria para la escisión es aportada por la proteína Cas, proporciona una plataforma versátil para la modificación genética de endonucleasas universales de ADN guiadas por ARN.

## Ejemplo 2

Ensamblado *in vitro* del complejo Cas9-ARNcr a partir de 4 componentes

En este ejemplo demostramos que el complejo Cas9-ARNcr catalíticamente activo se puede ensamblar *in vitro* mezclando 4 componentes individuales: la variante de la proteína Cas9 marcada en el extremo C-terminal con (His)6 ("His)6" desvelada como la SEQ ID NO: 23), el transcrito de ARNtracr (SEQ ID NO: 5), el transcrito de ARN CRISPR (SEQ ID NO: 8) y la RNasalll de *E. coli* (Abgene). La proteína Cas9 se incubó previamente con transcritos de ARNtracr y de ARN CRISPR, seguido por una incubación adicional con RNasalll para generar un complejo Cas9-ARNcr catalíticamente competente que se usa para la escisión de ADN específica de sitio.

Más específicamente, los fragmentos de ARN necesarios para ensamblar el complejo fueron producidos mediante transcripción *in vitro* (kit de transcripción TranscriptAid™ T7 High Yield, Fermentas) del fragmento generado por PCR que contenía un promotor T7 en el extremo proximal de la secuencia que codifica el ARN. Los fragmentos de ADN generados con PCR que codificaban el ARN CRISPR y ARNtracr se produjeron usando el plásmido pCas9(-)SP1 como molde con la siguiente pareja de cebadores: 5'-taatacgaactactataGgtagaaaagatatacctacagg-3' (SEQ ID NO: 40)/5'-CAACAACCAAGCTAATACAGCAG-3' (SEQ ID NO: 41) y 5'-aaaacaccgaatcggtgccac-3' (SEQ ID NO: 42)/5'-taatacgaactactataGggTAATAATAATTGTGGTTTGAACCATTC-3' (SEQ ID NO: 43) (promotor de la polimerasa de ARN T7 subrayado, inicio de la transcripción en negrita). El transcrito de ARN CRISPR de 150 nt está compuesto por secuencias de Repetición-Espaciador1-Repetición de 102 nt, flanqueadas por las regiones de 23 nt corriente arriba y 25 nt corriente abajo, necesarias para aparear el cebador. El transcrito de 105 nt de ARNtracr se compone de un segmento de 38 nt parcialmente complementario al fragmento de la secuencia de la repetición de CRISPR3 de DCGG7710 de *S. thermophilus* (secuencia anti-repetición), flanqueado por la región de 16 nt corriente arriba y 51 nt corriente abajo. Los fragmentos de ARN producidos mediante transcripción *in vitro* se purificaron usando el kit de limpieza RNeasy MinElute Cleanup (Qiagen).

Para ensamblar *in vitro* el complejo Cas9-ARNcr catalíticamente competente, la proteína Cas9 marcada con (His)6 ("His)6" desvelada como la SEQ ID NO: 23), se mezcló con transcritos de ARN CRISPR y de ARNtracr en una relación molar de 1:0,5:1 y se incubó previamente en un tampón que contenía Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM a 37 °C durante 30 minutos, seguido por la adición de RNasalll (Ambion), MgCl<sub>2</sub> y DTT y una incubación posterior durante 30 minutos adicionales. Las concentraciones finales de los componentes en la mezcla del ensamblado eran las siguientes: 100 nM de proteína Cas9 marcada con (His)6 ("His)6" desvelada como la SEQ ID NO: 23), 50 nM de ARN CRISPR, 100 nM de ARNtracr, RNasalll 50 nM, MgCl<sub>2</sub> 10 mM y DTT 1 mM.

A continuación proporcionamos evidencias experimentales de que el complejo Cas9-ARNcr ensamblado *in vitro* guiado por la secuencia de ARNcr, escinde el ADN en el sitio específico para generar extremos romos. A este respecto, el complejo Cas9-ARNcr se puede usar como una alternativa para una endonucleasa de restricción o meganucleasa para la escisión de ADN específica de sitio *in vitro*. La especificidad de secuencia del complejo viene dictada por la secuencia de ARNcr que se puede modificar genéticamente para dirigirla a una diana de ADN conveniente.

En primer lugar, la actividad de escisión del ADN del complejo Cas9-ARNcr ensamblado *in vitro*, se sometió a ensayo sobre los sustratos plasmídicos pSP1 y pUC18. El plásmido pSP1 contenía una secuencia de protoespaciador1 flanqueada por la secuencia 5'-GGNG-3'PAM. La secuencia del protoespaciador1 no estaba presente en pUC18. Se llevaron a cabo reacciones en los plásmidos pUC18 y pSP1 (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) a 37 °C en Tris HCl 10 mM (pH 7,5 a 37 °C), NaCl 50 mM, 0,05 mg/ml de BSA, DTT 0,5 mM y MgCl<sub>2</sub> 10 mM. Las mezclas de reacción normalmente contenían 3,0 nM de ADN plasmídico superenrollado. Las reacciones se iniciaron mezclando volúmenes de 50 µl de complejo Cas9-ARNcr y ADN plasmídico (relación 1:1 v/v) en un tampón de reacción. Se extrajeron partes alícuotas a intervalos programados y se inactivaron con fenol/cloroformo. La fase acuosa se mezcló con una solución de colorante de carga (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y los productos de la reacción se analizaron mediante electroforesis a través de agarosa (Figura 16). Para verificar si el plásmido pSP1 previamente escindido por el complejo Cas9-ARNcr se podía volver a ligar, purificamos el producto de la escisión de pSP1 lineal a partir del gel de agarosa usando el kit de extracción en gel GeneJET (Fermentas) y lo ligamos de nuevo usando ADN ligasa T4 (Fermentas). Después de la transformación de células de *E. coli* con la mezcla de ligación, se seleccionaron cinco clones individuales a partir de los transformantes resultantes, se purificó el ADN plasmídico y se sometió a secuenciación usando los siguientes cebadores: 5'-ccgcatcaggcgccattcgcc-3' (SEQ ID NO: 29) (secuenciación de la hebra (+)) y 5'-gcgaggaagcggaagagcgccc-3' (SEQ ID NO: 30) (secuenciación de la hebra (-)). El análisis de la secuencia reveló que la secuencia de ADN del plásmido pSP1 en el locus que se había escindido con el complejo Cas9-ARNcr y ligado de nuevo, era idéntica a la secuencia del plásmido que no se había tratado. La transformación de *E. coli* con la mezcla de ligación en ausencia de la ADN ligasa T4, no produjo transformantes



indicando que no se purifican conjuntamente trazas de plásmido superenrollado con el producto de reacción lineal.

A continuación, se sometió a ensayo la actividad de corte del complejo Cas9-ARNcr ensamblado *in vitro* sobre un dúplex de oligodesoxinucleótido de SP1 de 55 pb sintético que contenía una secuencia de protoespaciador que se apareaba con la secuencia del espaciador de ARNcr (Figura 17). Las condiciones de la reacción eran idénticas a las descritas anteriormente para la escisión del ADN plasmídico, excepto que se usó 1 nM de oligodúplex. El análisis del producto de la reacción reveló que el complejo Cas9-ARNcr ensamblado *in vitro* cortaba ambas hebras del oligodúplex en una posición fija, dentro del protoespaciador, después del 37° nucleótido desde el extremo 5', 4 nt corriente arriba de la secuencia PAM 5'-GGNG-3', dejando extremos romos (Figura 17).

### Ejemplo 3

#### 10 Complejo Cas9-ARNcr ensamblado *in vitro* a partir de 3 componentes

En este ejemplo, demostramos que el complejo Cas9-ARNcr activo se puede ensamblar *in vitro* mezclando 3 componentes individuales: la variante de la proteína Cas9 marcada en el extremo C-terminal con (His)6 ("His)6" desvelada como la SEQ ID NO: 23), el transcrito de ARNtracr proporcionado en el Ejemplo 1 (SEQ ID NO: 5 y SEQ ID NO: 6) y el transcrito de ARN CRISPR (SEQ ID NO: 8) proporcionado en el Ejemplo 1 o ARNcr sintético (SEQ ID NO: 8) que se corresponde con el ARNcr putativo del sistema CRISPR3/Cas de la cepa DGCC7710 de *S. thermophilus*. El oligorribonucleótido sintético de 42 nt está compuesto por 20 nt idénticos al espaciador1 de la región CRISPR3 en el extremo 5'-terminal y 22 nt de la secuencia de repetición en el extremo 3'. Más específicamente, se obtuvieron los transcritos de ARNtracr y ARN CRISPR como se ha descrito en el Ejemplo 1. Para generar el complejo Cas9-ARNcr, la proteína Cas9 marcada con (His)6 ("His)6" desvelada como la SEQ ID NO: 23) se mezcló con el transcrito de ARNtracr y de ARN CRISPR, o ARNcr sintético de 42 nt, en una relación molar de 1:0,5:1, y se incubó en un tampón que contenía Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM a 37 °C durante 1 h. Las concentraciones finales de los componentes en la mezcla del ensamblado eran las siguientes: 100 nM de proteína Cas9 marcada con (His)6 ("His)6" desvelada como la SEQ ID NO: 23), 50 nM de ARN CRISPR o ARNcr sintético de 42 nt, 100 nM de ARNtracr.

A continuación proporcionamos evidencias experimentales de que el complejo Cas9-ARNcr ensamblado *in vitro* guiado por la secuencia de ARNcr escinde el ADN en el sitio específico para generar extremos romos. A este respecto, el complejo Cas9-ARNcr se puede usar como una alternativa para una endonucleasa de restricción o meganucleasa para la escisión de ADN específica de sitio *in vitro*. La especificidad de secuencia del complejo viene dictada por la secuencia de ARNcr que se puede modificar genéticamente para dirigirla a una diana de ADN conveniente.

En primer lugar, la actividad de escisión del ADN del complejo Cas9-ARNcr ensamblado *in vitro*, se sometió a ensayo sobre los sustratos plasmídicos pSP1 y pUC18. El plásmido pSP1 contenía una secuencia de protoespaciador1 flanqueada por la secuencia 5'-GGNG-3'PAM. La secuencia del protoespaciador1 no estaba presente en pUC18. Se llevaron a cabo reacciones en los sustratos de plásmidos (Sapranaukas y col., 2011. Nucleic Acids Res 39:9275-82) a 37 °C en Tris HCl 10 mM (pH 7,5 a 37 °C), NaCl 50 mM, 0,05 mg/ml de BSA, DTT 0,5 mM y MgCl2 10 mM. Las mezclas de reacción normalmente contenían 3,0 nM de ADN plasmídico superenrollado. Las reacciones se iniciaron mezclando volúmenes de 50 µl de complejo Cas9-ARNcr y ADN plasmídico (relación 1:1 v/v) en un tampón de reacción. Se extrajeron partes alícuotas a intervalos programados y se inactivaron con fenol/cloroformo. La fase acuosa se mezcló con una solución de colorante de carga (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y los productos de la reacción se analizaron mediante electroforesis a través de agarosa (Figura 18).

A continuación, se sometió a ensayo la actividad de corte del complejo Cas9-ARNcr ensamblado *in vitro* sobre un dúplex de oligodesoxinucleótido de SP1 de 55 pb sintético que contenía una secuencia de protoespaciador que se apareaba con la secuencia del espaciador de ARNcr (Figura 19). Las condiciones de la reacción eran idénticas a las descritas anteriormente para la escisión del ADN plasmídico, excepto que se usó 1 nM de oligodúplex. El análisis del producto de la reacción reveló que el complejo Cas9-ARNcr ensamblado *in vitro* cortaba ambas hebras del oligodúplex en una posición fija, dentro del protoespaciador, después del 37° nucleótido desde el extremo 5', 4 nt corriente arriba de la secuencia PAM 5'-GGNG-3', dejando extremos romos (Figura 19).

### Ejemplo 4

Casete de espaciador intercambiable para la reprogramación de la especificidad del complejo Cas9-ARNcr

En este ejemplo describimos un casete de espaciador intercambiable que permite producir ARNcr que es portador de una secuencia de nucleótidos contra cualquier diana de ADN conveniente, para ser utilizado para ensamblar el complejo Cas9-ARNcr descrito en los Ejemplos 1 y 2 (Figura 20B). El casete es portador de una única unidad de repetición-espaciador-repetición que permite la inserción del oligodúplex que es portador de la nueva secuencia espaciadora requerida para generar un ARNcr deseado. Para diseñar un casete, primero construimos un casete que contenía una secuencia líder, una secuencia de repetición y un sitio de reconocimiento SapI único, próximo a la secuencia de la repetición, seguido por el sitio BamHI (Figura 20C). Para generar una región CRISPR que contenía el espaciador único deseado, insertamos un oligodúplex sintético que contenía una secuencia espaciadora única y una unidad de repetición en el plásmido cortado previamente con las enzimas de restricción SapI y BamHI. Usando este casete produjimos transcritos de ARNcr que contenían secuencias de nucleótidos complementarias a los protoespaciadores N1 y N2 presentes en el plásmido pUC18 (véase más abajo).

Como prueba de la demostración del principio, utilizamos un casete de espaciador intercambiable para generar ARNcr1 y ARNcr2 que estaban diseñados para dirigirse al plásmido pUC18 en el protoespaciador1 y el protoespaciador2, respectivamente, incorporamos ARNcr1 y ARNcr2 en el complejo Cas9 como se ha descrito en el Ejemplo 1 y utilizamos esos complejos para la escisión del plásmido pUC18. El protoespaciador N1 está ubicado cerca del sitio de la endonucleasa de restricción SapI, mientras que el protoespaciador N2 está cerca del sitio AatII. La distancia entre los sitios de restricción SapI y AatII es de 775 pb, mientras que la distancia entre los sitios de escisión del complejo Cas9-ARNcr putativos, ubicados en los espaciadores N1 y N2, es de 612 pb (Figura 21A). Los fragmentos de PCR de ARNcr1 y ARNcr2 que contenían el promotor T7 en el extremo proximal, se obtuvieron a partir de los plásmidos del casete de espaciador intercambiable correspondiente y se usaron para producir mediante transcripción *in vitro*, transcritos de ARN CRISPR que eran portadores de secuencias que se emparejaban con las secuencias del espaciador N1 o el espaciador N2. Los complejos catalíticamente activos de Cas9 con ARNcr1 y ARNcr2 se ensamblaron para la escisión del ADN como se ha descrito en el Ejemplo 1. Los complejos ensamblados *in vitro* que contenían ARNcr1 o ARNcr2 linealizaban el plásmido pUC18 (Figura 21B). Cuando ambos complejos se incubaban con el plásmido pUC18, se obtenían dos fragmentos de ADN (2074 y 612 pb) (Figura 21B), lo que indica que la escisión del plásmido se producía en sitios dirigidos por las moléculas de ARNcr presentes en los complejos.

### Ejemplo 5

Procedimiento de clonación que emplea el complejo Cas9-ARNcr

En este ejemplo demostramos que el complejo Cas9-ARNcr se puede usar para preparar un vector para el procedimiento de clonación. Primero demostramos que los productos de escisión obtenidos a través del complejo Cas9-ARNcr se pueden volver a ligar mediante una ADN ligasa. Purificamos el producto de escisión de pSP1 lineal a partir de un gel de agarosa y lo volvimos a ligar usando ADN ligasa. Después de la transformación de células de *E. coli* con la mezcla de ligación, se seleccionaron cinco clones individuales a partir de los transformantes resultantes, se purificó el ADN plasmídico y se sometió a secuenciación. El análisis de la secuencia reveló que la secuencia de ADN del plásmido pSP1 en el locus que se había escindido con el complejo Cas9-ARNcr y vuelto a ligar, era idéntica a la secuencia del plásmido sin tratar. La transformación de *E. coli* con la mezcla de ligación en ausencia de ADN ligasa T4 no producía transformantes, lo que indica que no se purificaban conjuntamente trazas de plásmido superenrollado con el producto de reacción lineal. Este resultado ilustra que los extremos del ADN generados por la escisión con Cas9, son sustratos para la ADN ligasa T4 y, por lo tanto, deben contener un fosfato en el extremo 5'-terminal y un grupo OH libre en el extremo 3'-terminal (Lehman, 1974).

A continuación, analizamos la escisión del plásmido pUC18 con el complejo Cas9 cargado con ARNcr1 y ARNcr2 descrito en el Ejemplo 5 (Figura 21A). En primer lugar, pUC18 se escindió con un complejo, se purificó y se volvió a ligar. La secuenciación de 10 clones en cada caso confirmaba que la secuencia del plásmido escindido y vuelto a ligar era idéntica a la secuencia del plásmido sin tratar (Figura 21C). Este experimento sugiere que no se introducen mutaciones adicionales después de la escisión a través del complejo Cas9-ARNcr y la ligación, y que el complejo Cas9-ARNcr se puede usar para experimentos de clonación. Cuando se incubaban ambos complejos con el plásmido pUC18, se obtuvieron dos fragmentos de ADN (2074 y 612 pb) (Figura 21B), lo que indica que se había producido una escisión del plásmido en sitios seleccionados por las moléculas de ARNcr presentes en los complejos. Para demostrar que el plásmido pUC18 escindido con los complejos Cas9-ARNcr es adecuado para una modificación genética, clonamos un fragmento de PCR que contenía un promotor y un gen de resistencia a la tetraciclina procedente del plásmido pACYC184, con el vector pUC18 cortado previamente con la mezcla de complejo Cas9 que contenía ARNcr1 o ARNcr2. Los clones se seleccionaron en medio enriquecido con tetraciclina y ampicilina. La secuenciación de 4 clones seleccionados confirmaba que el fragmento de PCR intacto se había insertado en una posición deseada ((Figura 21C).

Más específicamente, los 2 µg de pUC18 se incubaron con la mezcla de complejos de Cas9-ARNcr ensamblados por separado (250 nM cada uno) que contenían diferentes ARNcrs, durante 1 hora a 37 °C en 100 µl de volumen de reacción (Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM, DTT 1 mM y MgCl<sub>2</sub> 10 mM). El fragmento de vector obtenido se purificó desde el gel de agarosa usando el kit de extracción en gel GeneJET (Thermo Fisher Scientific) y se dividió en dos partes iguales. Una parte del vector cortado previamente se desfosforiló con la fosfatasa alcalina FastAP mientras que la otra parte no se trató. Se obtuvo un inserto de 1282 pb que contenía un promotor y un gen de resistencia a la tetraciclina a partir del plásmido pACYC184 mediante PCR. Después de la purificación usando el kit de purificación de PCR GeneJET (Thermo Fisher Scientific), una solución que contenía el fragmento de PCR se dividió en dos partes. Una parte se fosforiló con una polinucleótido cinasa T4 (Thermo Fisher Scientific) mientras que la otra parte no se trató. El vector sin tratar se ligó con el fragmento de PCR sin tratar, mientras que un vector desfosforilado se ligó con un fragmento fosforilado usando la ADN ligasa T4 (Thermo Fisher Scientific). Los clones se seleccionaron en un medio complementado con 100 µg/ml de Ap y 25 µg/ml de Tc.

### Ejemplo 6

Escisión de sustratos de ADN largos mediante el complejo Cas9 ARNcr

En este ejemplo, demostramos que Cas9-ARNcr se puede dirigir para escindir dianas en moléculas de ADN largas, incluyendo el fago λ, *E. coli* y ADNs genómicos humanos.

Más específicamente, dirigimos el complejo Cas9-ARN para escindir sitios específicos en el bacteriófago  $\lambda$  (48 kb), la cepa de *E. coli* BL-21 (4,6 Mb) y ADNs genómicos humanos (3,2 Gb). El complejo Cas9-ARNcr se ensambló como se ha descrito en los Ejemplos 2 y 3. Utilizamos ARNcrs sintéticos de 42 nt de longitud, preARNcrs de 150 nt y ARNtracr sintetizados usando transcripción *in vitro* a partir de moldes generados como se ha descrito en el Ejemplo 4.

5 Las reacciones de escisión del ADN de  $\lambda$  se iniciaron mezclando el ADN de  $\lambda$  (Thermo Fisher Scientific) con el complejo Cas9-ARN ensamblado (relación 1:1 v/v) e incubando a 37 °C. La mezcla de reacción final contenía 2  $\mu$ g de ADN de  $\lambda$ , complejo Cas9-ARN 50 nM, Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM, DTT 1 mM y MgCl<sub>2</sub> 10 mM en 100  $\mu$ l de volumen de reacción. Se extrajeron partes alícuotas a intervalos programados y se inactivaron con fenol/cloroformo. La fase acuosa se mezcló con una solución de tinte de carga 3X (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y los productos de la reacción se analizaron mediante electroforesis a través de geles de agarosa y tinción con bromuro de etidio. El análisis de los productos de la escisión del ADN genómico del fago  $\lambda$  lineal en gel de agarosa, confirmaba que el ADN de ~40 pb de longitud se escinde de forma eficaz en un solo sitio (Figura 22A).

15 El ADN procedente de la cepa de *E. coli* BL21 (DE3) se aisló usando el kit de purificación de ADN genómico (Thermo Fisher Scientific). Para el ensayo de escisión, el ADN genómico de *E. coli* se combinó con el complejo Cas9-ARN ensamblado (relación 1:1 v/v) y se incubó durante 3 horas a 37 °C. La mezcla de reacción final contenía 30  $\mu$ g de ADN genómico, complejo Cas9-ARN 1  $\mu$ M, Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM, DTT 1 mM y MgCl<sub>2</sub> 10 mM en 300  $\mu$ l de volumen de reacción. Después de la incubación, se añadieron 30  $\mu$ l de PstI FastDigest (Thermo Fisher Scientific) y la mezcla de reacción se incubó durante 16 horas adicionales a 37 °C. La reacción terminó calentando la mezcla de reacción durante 30 minutos a 55 °C con Proteinasa K (0,5 mg/ml; Thermo Fisher Scientific) y SDS (0,5%, p/v), seguido de 30 minutos de incubación a temperatura ambiente con RNasa A (0,25 mg/ml; Thermo Fisher Scientific). Después de la extracción con fenol/cloroformo, el ADN precipitó con isopropanol y se disolvió en tampón TE (Tris-HCl 10 mM, pH 8,0 y EDTA 1 mM). Se mezclaron 10  $\mu$ g de ADN con solución de colorante de carga 3X (azul de bromofenol al 0,01% y EDTA 75 mM en glicerol al 50% v/v) y se sometieron a electroforesis en gel de agarosa al 1%.

25 Para analizar los productos de la escisión con Cas9-ARNcr de ADN genómico de *E. coli*, diseñamos una sonda contra un fragmento de ADN que contenía una diana del complejo Cas9-ARN (un protoespaciador) (Figura 22B) y realizamos un análisis de transferencia Southern. El análisis de transferencia Southern se realizó como se describe en (Sambrook y col, 1989. Molecular Cloning: A Laboratory Manual) con las siguientes modificaciones. El ADN fraccionado se transfirió desde el gel de agarosa a una membrana de nailon SensiBlot Plus (Thermo Fisher Scientific) mediante una transferencia semiseca. El ADN se desnaturalizó y se fijó sobre la membrana colocándolo sobre papel absorbente saturado con NaOH 0,4 M durante 10 minutos, se aclaró con 2X SSC y se secó al aire. La membrana se prehibridó con tampón 6X SSC que contenía SDS al 0,5% y 100  $\mu$ g/ml de ADN de esperma de salmón desnaturalizado (Amresco) durante 1 hora a 65 °C. La sonda de hibridación se generó mediante PCR usando ADN genómico de BL21(DE3) de *E. coli* como molde, obteniendo un producto de 397 pb. Los extremos 5' se desfosforilaron con fosfatasa FastAP (Thermo Fisher Scientific) y se radiomarcaron mediante incubación con [ $\gamma$ -<sup>32</sup>P]ATP (Hartmann Analytic) y T4 PNK (Thermo Fisher Scientific). La sonda marcada se purificó usando el kit de purificación de PCR GeneJET (Thermo Fisher Scientific), se desnaturalizó calentando a 95 °C durante 5 minutos, se enfrió rápidamente sobre hielo y se añadió directamente a la solución de prehibridación. La membrana se hibridó durante 16 horas a 65 °C y se lavó dos veces con 2X SSC, SDS al 0,5% y dos veces con 2X SSC, SDS al 0,1% a temperatura ambiente, se secó al aire y se visualizó mediante formación de imágenes de fósforo (FLA-5100; Fujifilm).

40 La sonda se diseñó para que se dirigiera al fragmento de ADN que contenía una diana (un protoespaciador) para el complejo Cas9-ARN (Figura 22B). La distancia entre dos dianas de PstI es de ~1500 pb, mientras que la distancia entre el protoespaciador y la diana de PstI izquierda es de 466 pb. Después de la escisión con el complejo Cas9, detectamos solo un fragmento de ADN de 466 pb (Figura 22C), lo que significa que todas las dianas de ADN se escindían con la proteína Cas9 en la posición deseada. Estos datos demuestran claramente que la proteína Cas9 encuentra de forma eficaz dianas en moléculas muy largas y complejas, como el ADN vírico y bacteriano.

50 Para analizar los productos de la escisión con Cas9-ARNcr del ADN genómico humano, usamos ADN extraído a partir de cerebro humano. El ADN genómico humano se combinó con el complejo Cas9-ARNcr ensamblado (relación 1:1 v/v) y se incubó durante 30 minutos a 37 °C. La mezcla de reacción final contenía 1  $\mu$ g de ADN genómico, Cas9 100 nM, Tris-HCl 10 mM (pH 7,5 a 37 °C), NaCl 100 mM, DTT 1 mM y MgCl<sub>2</sub> 10 mM en 100  $\mu$ l de volumen de reacción. Los complejos Cas9-ARNcr-HS1 (SeqID n° 13) y Cas9-ARNcr-HS2 (SeqID n° 14) se ensamblaron para dirigirse a los loci RASGEF1C o ARL15, respectivamente. Los productos de la escisión se analizaron usando qPCR (Figura 22D). Después del tratamiento con el complejo Cas9-ARNcr, la cantidad de dianas de ADN intactas disminuía más de 25 veces. El análisis de los resultados obtenidos a partir de los datos de la qPCR reveló que los complejos Cas9-ARN escinden el ADN genómico humano de manera eficaz en los loci deseados. Estos datos demuestran claramente que la proteína Cas9 encuentra las dianas de forma eficaz en moléculas muy largas y complejas, tales como el ADN vírico, bacteriano y de mamífero.

### Ejemplo 7

Evidencia para la edición de genes de un plásmido informador en células de mamífero después de la transfección de complejos Cas9/ARN.

Se construyó un plásmido informador para controlar la reparación de roturas de la hebra doble, ya fuera mediante unión de extremos no homólogos (NHEJ) o por recombinación homóloga (HR). El plásmido contenía GFP con un intrón y flanqueando las secuencias eGFP hay secuencias RFP 5' y 3', así como sitios de homología (Figura 23). La reducción de la fluorescencia de eGFP usando este plásmido informador era una indicación de NHEJ, en donde se había reparado de forma imperfecta mediante NHEJ una rotura de la hebra doble mediada por Cas9/ARN en las dianas C o D, alterando de este modo la secuencia codificadora de eGFP. El direccionamiento hacia las dianas intrónicas A y B y la reparación mediante NHEJ probablemente no da lugar a una reducción de la fluorescencia de eGFP porque las mutaciones inducidas por NHEJ generalmente deletan o insertan <20 bps y, por lo tanto, no afectarán a las regiones que codifican eGFP o a las uniones en sitios de empalme. La aparición de fluorescencia de RFP, por otro lado, era una indicación de HR, en donde la rotura de la hebra doble mediada por Cas9/ARN se repara mediante HR usando las secuencias homólogas de RFP indicadas.

El direccionamiento del ARNcr empleaba moléculas de ARN de 42 nucleótidos, como se ha descrito anteriormente, que tenían 22 nucleótidos que son la secuencia de repetición, y 20 nucleótidos (secuencia espaciadora) son para la diana específica. Como se ha descrito anteriormente, el ADN diana necesita el motivo de *S. thermophilus* o PAM que es "NGGNG", corriente abajo del protoespaciador en la diana. GFP no se había "modificado genéticamente" para contener ese motivo PAM; varias secuencias diana dentro de eGFP se presentan de forma natural con la secuencia PAM y los ARNcrs se diseñaron para que se dirigieran a las secuencias de espaciadores adyacentes. RFP era un marcador de la recombinación homóloga después de que Cas9/ARN creara una rotura de la hebra doble en eGFP.

La Figura 28A muestra la estructura artificial de un gen informador para el análisis de la actividad de la proteína Cas9 en células eucariotas *in vivo*. La secuencia del intrón contiene tres sitios diana de cas9 (A, E, B); el gen GFP contiene dos (C, D) sitios diana de Cas9. El gen RFP se divide en la posición Y196, en donde se anula la fluorescencia de RFP. La Figura 28B muestra que se observa fluorescencia de GFP después del procesamiento del intrón *in vivo*. La Figura 28C muestra que las roturas del ADNds facilitadas por el complejo Cas9/ARNcr, en cualquiera de los sitios diana de nucleasa mencionados anteriormente, pueden inducir HR, dando como resultado el reensamblamiento del gen RFP y la aparición de fluorescencia RFP. Las Figuras 28D y E muestran que las roturas del ADNds facilitadas por el complejo Cas9/ARNcr, en cualquiera de los sitios diana de nucleasa mencionados anteriormente, pueden inducir NHEJ. Las mutaciones en la secuencia del gen GFP darán como resultado una pérdida o disminución de la fluorescencia de GFP; las mutaciones en el intrón puede que no afecten a la fluorescencia de GFP, sin embargo, en otros casos pueden producir un ARN mensajero maduro con secuencias de intrón empalmadas incorrectamente y provocar una pérdida o disminución de la fluorescencia de GFP.

La proteína Cas9 de *S. thermophilus*, purificada a partir de *E. coli*, formaba un complejo con ARNtracr transcrito *in vitro* y ARNcr sintético no modificado dirigido a la secuencia A (intrónica) o la secuencia C (codificadora) de eGFP. Para la transfección, los complejos Cas9/ARN (ya sea dirigidos a A o a C) se incubaron con el reactivo de transfección TurboFECT y el ADN plasmídico informador también se incubó con TurboFECT en tubos separados y ambos se añadieron a las células CHO-K1. El porcentaje de células positivas para eGFP se determinó por citometría de flujo. Como se muestra en las Figuras 24 y 29, cuando las células se transfectaban con el plásmido informador solo o con el plásmido informador con proteína Cas9 sola, el porcentaje de células positivas para GFP era de aproximadamente un 40-50%, lo que es indicativo de la eficacia global de la transfección. Sin embargo, cuando los complejos Cas9/ARN que se dirigían a la secuencia C de eGFP, se añadían a las células junto con el plásmido informador, el porcentaje de células positivas para eGFP se reducía a aproximadamente un 15%. Esta disminución de las células positivas para eGFP se observaba solo con los complejos Cas9/ARN dirigidos a la secuencia C y no había una disminución significativa en las células positivas para eGFP observadas con los complejos Cas9/ARN dirigidos a la secuencia A o con un ARN no específico. Este resultado indicaba que Cas9/ARN que se dirigía a la secuencia C de eGFP daba como resultado una edición génica de eGFP mediante la introducción de una rotura de la hebra doble y una corrección imperfecta mediante NHEJ, creando una delección en la secuencia codificadora de eGFP.

Además de analizar el porcentaje de células positivas para eGFP, las células transfectadas también se visualizaron mediante microscopía fluorescente para controlar la aparición de células positivas para RFP, una indicación de reparación mediante HR de la rotura de la hebra doble mediada por Cas9, en lugar de mediante NHEJ. Como se observa en la Figura 25, se observa RFP en algunas células después de la transfección con el plásmido informador y los complejos Cas9/ARN dirigidos a la secuencia C de eGFP, lo que sugiere una reparación de la rotura de la hebra doble mediante HR.

### Ejemplo 8

Los complejos Cas9/ARN preparados empleando ARNtracs y ARNcrs sintéticos no modificados son funcionales *in vitro*

Los experimentos descritos en el Ejemplo 7 anterior empleaban complejos Cas9/ARN que comprendían Cas9 purificada, ARNcrs sintéticos y ARNtracr transcrito *in vitro*. Para determinar si los complejos Cas9/ARN eran funcionales cuando se preparaban usando componentes de ARN totalmente sintéticos (ARNcr y ARNtracr), se sintetizaron ARNtracs sin modificar de *S. thermophilus* (tanto la versión de 89 meros endógena como una versión más corta de 74 meros que se esperaba que conservara la funcionalidad). Los ARNcrs sintéticos sin modificar se generaron contra la diana E (véanse las Figuras 26 y 30) localizada dentro del intrón de eGFP en el plásmido informador descrito

anteriormente y se generaron complejos Cas9/ARN (ARNcr y ARNtracr). Para analizar estos complejos, el plásmido informador utilizado anteriormente se incubó con los complejos *in vitro* y se realizó un seguimiento de la restricción mediante electroforesis en gel.

5 Como se observa en la Figura 27, los complejos Cas9/ARN compuestos por ARNs totalmente sintéticos eran igualmente funcionales en el ensayo *in vitro* que los complejos Cas9/ARN compuestos por ARNcr sintético y ARNtracr transcrito *in vitro*.

Secuencias

**SEQ ID NO: 1**

Cepa WT\_Cas9\_S. *thermophilus* DGCC7710 CRISPR3-Cas

Una letra:

10

m l f n k c i i i s n l d f s n e k e c m t k p y s i g l d i g t n s v g w a v i t d n y k v p s k k m k v l g n t s k k y i k n l l g v l l f d s g i t a e g r r l k r t a r r r y t r r r  
n r i l y l q e i f s t e m a t l d d a f f q r l d d s f l v p d d k r d s k y p i f g n l v e e k v y h d e f p t i y h r k y l a d s t k k a d l r l v y l a l a h m i k y r g h f l i e g e f  
n s k n n d i q n f q d f l d t y n a i f e s d l s l e n s k q l e e i v k d k i s k l e k k d r i l k l f p g e k n s g i f s e f l k l i v n q a d f r k e f n l d e k a s l h f s k e s y  
d e d l e t l l g y i g d d y s d v f l k a k k y d a i l l s g f l t v d n e a p l s s a m i k r y n e h k e d l a l l k e y i r n i s l k t y n e v f k d d t k n g y a g y i d g k t n  
q e d f y v y l k n l l a e f e g a d y f l e k i d r e d f l r k q r t f d n s i p y q i h l q e m r a i l d k q a k f y p f l a k n k e r i e k i l t f r i p y y v g p l a r g n s d f a w s  
i r k n e k i t p w n f e d v i d k e s s a e a f i n r m t s f d l y l p e e k v l p k h s l l y e t f n v n e l t k v r f i a e s m r d y q f l d s k q k k d i v r l y f k d k r k v t d  
k d i e y l h a i y g y d g i e l k g i e k q f n s s l s t y h d l l n i i n d k e f l d d s s n e a i i e e i h t l t i f e d r e m i k q r l s k f e n i f d k s v l k k l s r r h y t g w g k l  
s a k l i n g i r d e k s g n t i l d y l i d d g i s n r n f m q l i h d d a l s f k k i q k a q i g d e d k n i e v v k s l p g s p a i k k g i l q s i k i v d e l v k v m g g r k  
p e s i v v e m a r e n q y t n q g k s n s q r l k r l e k s l e k e l g s k i l k e n i p a k l s k i d n n a l q n d r l y l y l q n g k d m y t g d d l d i d r l s n y d i d i h i p  
q a f l k d n s i d n k v l v s s a n r g k s d d f p s l e v v k k r k t f w y q l l k s k l i s q r k f d n l t k a e r g g l l p e d k a g f i q r l v e t r q i t k h v a r l l d e k f n  
n k k d e n n r a v r t v k i i t k l s t l v s q f r k d f e l y k v r e i n d f h h a d a y n a v i a s a l l k k y p k l e p e f v y g d y p k y n s f r e r k s a t e k v y f y s n i  
m n i f k k s i s l a d g r v i e r p l i e v n e e t g e s v w n k e s d l a t v r r v l s y p q v n v k k v e e q n h g l d r g k p k g l f n a n l s s k p k p n s n e n l v g a k  
e y l d p k y g y a g i s n s f a v l v k g t i e k g a k k k i t n v l e f q g i s i l d r i n y r k d k l n f l e k y k d i e l i e l p k y s l f e l s d g s r m l a s i l s t n n k r  
g e i h k g n q i f l s q k f v k l y h a k r i s t i n e n h r k y v e n h k k e f e l f y i l e f n e n y v g a k k n g k l l n s a f q s w q n h s i d e l e s s f i g p t g s e r  
k g l f e l t s r g s a a d f e l v g k i p r y r d y t p s s l l k d a t l i h q s v t g l y e t r i d l a k l g e g

Tres letras:

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys  
CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla  
ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer  
LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla  
GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle  
LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln  
ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe  
GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg  
LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla  
HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn  
AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp  
LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu

GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer  
 GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp  
 GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu  
 GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla  
 IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer  
 AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle  
 ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr  
 AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu  
 LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu  
 ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet  
 ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg  
 IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn  
 SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu  
 AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp  
 LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn  
 ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe  
 LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal  
 ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu  
 LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle  
 IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle  
 HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu  
 AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly  
 LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu  
 AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp  
 AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly  
 AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu  
 GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer  
 IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln  
 GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu  
 AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr  
 LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg  
 LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle  
 AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer  
 LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle  
 SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp  
 LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla

ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr  
ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu  
TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal  
IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp  
TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr  
SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu  
ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp  
LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu  
GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer  
SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro  
LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr  
IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle  
LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys  
AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer  
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly  
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer  
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu  
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys  
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer  
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer  
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer  
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg  
IleAspLeuAlaLysLeuGlyGluGly

**SEQ ID NO: 2**  
**Mutante D31A**  
Una letra:

mlfnciislnldfsnkekcmtkpsysiglaigtmsvgwavitdnykvpskkmkvlgnstskkyiknllgvlldfgitaegrllkrtrrrytrrr  
nrilylqeifstematlddaffqrlddsflvpddkrdsypifgnlveekvyhdefptiyhrkyladstkkadrlrvylalahmikyrghfliegef  
nsknndiqknfqdfltdynaifedslslenskqleeivkdkisklekkdrilklfpgekngifseflklivgnqadfrkcfnldekaslhfskesy  
dedletllgyigddysdvflkakklydailsgflvtvdnetepssamikrynehkedlallkeyirnislktynevfkddtkngyagyidgktn  
qedfyvylknllaefegadyflekidredflrkqrtfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws  
irkrnektpwnfedvidkessaeafinmmtsfdlylpeekvlpkhslyetfnvyneltkvrfaesmrdyqfldskqkkdivrlyfkdkrkvt

kdiieylhaiygydgielkgiekqfnsslstyhdllniindkeflddssneaieeihtltifedremikqrlskfenifdksvlkklsrrhytgwgl  
 saklingirdeksntildyliddgismnfmqlihddalsfkkkiqkaqiigdedkgnikevkslpgspaikkgilqsikivdelvkvmggrk  
 pesivvemarenqytnqgksnsqqlrkrlekslkelgskilkenipaklskidnmalqndrlylylqngkdmtygdlddidrlsnydidhiip  
 qafldnsidnkvlvssasnrksddfpslevvkrkrktfwyqlkksklsqrkfdnlkaerggllpedkagfiqrqlvetrqitkhvarlldekfn  
 nkkdenravrtvkiitlkslvsqfrkdfelykvreindfhahadaylnaviasallkkypklepefygdypkynsfrerksatekvfyfysni  
 mnifkksisladgrvierplievneetgesvwnkesdlatvrrvlsypqvnvkkveeqnhglrdgkpkglfnanlsskpkpnsnenlvgak  
 eyldpkkygyyagisnsfavlvkgtiekgakkkitnvlefgqisildrinyrkdklnflekykdielielpkyslfelsdgsrrmlasilstmnkr  
 geihkgnqiflsqkfvklylhakrisntinenhrkyvenhkkefeelfyilefnyevgakkngklinsafqswqnhsidecssfigptgser  
 kglfeltsrgsaadfevlgykipryrdytpssllkdatlihqsvtglyetridlaklgeg

Tres letras:

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys  
 CysMetThrLysProTyrSerIleGlyLeuAlaIleGlyThrAsnSerValGlyTrpAla  
 ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer  
 LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla  
 GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle  
 LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln  
 ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe  
 GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg  
 LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla  
 HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn  
 AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp  
 LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu  
 GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer  
 GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp  
 GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu  
 GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla  
 IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer  
 AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle  
 ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr  
 AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu  
 LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu  
 ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet  
 ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg  
 IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn  
 SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu



AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp  
 LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn  
 ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe  
 LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal  
 ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu  
 LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle  
 IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle  
 HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu  
 AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly  
 LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu  
 AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp  
 AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly  
 AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu  
 GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer  
 IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln  
 GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu  
 AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr  
 LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg  
 LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle  
 AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer  
 LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle  
 SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp  
 LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla  
 ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr  
 ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu  
 TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal  
 IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp  
 TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr  
 SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu  
 ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp  
 LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu  
 GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer  
 SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro  
 LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr  
 IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle  
 LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys

AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer  
ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly  
AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer  
AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu  
LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys  
LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer  
PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer  
AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer  
SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg  
IleAspLeuAlaLysLeuGlyGluGly

**SEQ ID NO: 3**  
**Mutante N891A**  
Una letra:

m l f n k c i i s i n l d f s n k e k c m t k p y s i g l d i g t m s v g w a v i t d n y k v p s k k m k v l g n t s k k y i k k n l l g v l l f d s g i t a e g r r l k r t a r r r y t r r r  
n r i l y l q e i f s t e m a t l d d a f f q r l d d s f l v p d d k r d s k y p i f g n l v e e k v y h d e f p t i y h r k y l a d s t k k a d l r l v y l a l a h m i k y r g h f l i e g e f  
n s k n n d i q n f q d f l d t y n a i f e s d l s l e n s k q l e e i v k d k i s k l e k k d r i l k l f p g e k n s g i f s e f l k l i v n q a d f r k e f n l d e k a s l h f s k e s y  
d e d l e t l l g y i g d d y s d v f l k a k k l y d a i l l s g f l t v d n e a p l s s a m i k r y n e h k e d l a l l k e y i m i s l k t y n e v f k d d t k n g y a g y i d g k t n  
q e d f y v y l k n l a e f e g a d y f l e k i d r e d f l r k q r t f d n g s i p y q i h l q e m r a i l d k q a k f y p f l a k n k e r i e k i l t f r i p y y v g p l a r g n s d f a w s  
i r k n e k i t p w n f e d v i d k e s s a e a f i n r m t s f d l y l p e e k v l p k h s l l y e t f n v y n e l t k v r f i a e s m r d y q f l d s k q k k d i v r l y f k d k r k v t d  
k d i e y l h a i y g y d g i e l k g i e k q f n s s l s t y h d l n i i n d k e f l d d s s n e a i e e i i h t l t i f e d r e m i k q r l s k f e n i f d k s v l k k l s r r h y t g w g k l  
s a k l i n g i r d e k s g n t i l d y l i d d g i s n r n f m q l i h d d a l s f k k k i q k a q i i g d e d k n i e v v k s l p g s p a i k k g i l q s i k i v d e l v k v m g g r k  
p e s i v v e m a r e n q y t n q g k s n s q q r l k r l e k s l k e l g s k i l k e n i p a k l s k i d n n a l q n d r l y l y l q n g k d m y t g d d l d i d r l s n y d i d h i i p  
q a f l k d n s i d n k v l v s s a s a r g k s d d f p s l e v v k k r k t f w y q l l k s k l i s q r k f d n l t k a e r g g l l p e d k a g f i q r q l v e t r q i t k h v a r l l d e k f n  
n k k d e n n r a v r t v k i i t l k s t l v s q f r k d f e l y k v r e i n d f h h a d a y l n a v i a s a l l k k y p k l e p e f v y g d y p k y n s f r e r k s a t e k v y f y s n i  
m n i f k k s i s l a d g r v i e r p l i e v n e e t g e s v w n k e s d l a t v r r v l s y p q v n v v k k v e e q n h g l d r g k p k g l f n a n l s s k p k p n s n e n l v g a k  
e y l d p k y g g y a g i s n s f a v l v k g t i e k g a k k k i t n v l e f q g i s i l d r i n y r k d k n f l e k g y k d i e l i i e l p k y s l f e l s d g s r r m l a s i l s t n k r  
g e i h k g n q i f l s q k f v k l l y h a k r i s t i n e n h r k y v e n h k k e f e l f y i l e f n e n y v g a k k n g k l l n s a f q s w q n h s i d e l e s s f i g p t g s e r  
k g l f e l t s r g s a a d f e l g v k i p r y r d y t p s s l l k d a t l i h q s v t g l y e t r i d l a k l g e g

Tres letras:

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys  
CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla

ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer  
 LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla  
 GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle  
 LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln  
 ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe  
 GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg  
 LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla  
 HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn  
 AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp  
 LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu  
 GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer  
 GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp  
 GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu  
 GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla  
 IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer  
 AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle  
 ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr  
 AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu  
 LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu  
 ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet  
 ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg  
 IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn  
 SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu  
 AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp  
 LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn  
 ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe  
 LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal  
 ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu  
 LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle  
 IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle  
 HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu  
 AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly  
 LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu  
 AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp  
 AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly  
 AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu

GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer  
 IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln  
 GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu  
 AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr  
 LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg  
 LeuSerAsnTyrAspIleAspHisIleIleProGlnAlaPheLeuLysAspAsnSerIle  
 AspAsnLysValLeuValSerSerAlaSerAlaArgGlyLysSerAspAspPheProSer  
 LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle  
 SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp  
 LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla  
 ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr  
 ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu  
 TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal  
 IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp  
 TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr  
 SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu  
 ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp  
 LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu  
 GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer  
 SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro  
 LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr  
 IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle  
 LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys  
 AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer  
 ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly  
 AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer  
 AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu  
 LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys  
 LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer  
 PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer  
 AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer  
 SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg  
 IleAspLeuAlaLysLeuGlyGluGly

**SEQ ID NO: 4**  
**Mutante H868A**  
 Una letra

mfnkciisinldfsnkecmtpysigldigtntsvgwavitdnykvpskkmkvlngntskkyikknllgvllfdsgitaegrllkrtarryrrr  
 nrilylqeifstematlddaffqrlrddsfvppdkrdskypifgnlveekvyhdefptiyhlrkyladstkkadrlrvylalahmikyrghfliegef  
 nsknndiqknfqdldtynafesdlslenkqleevkdkisklekkdrillkfpgeknsgefklivgnqadfrkcfnldekaslhfskesy  
 dedletllgyigddysdvflkakklydaillsgflvtvdnetepssamikrynehkedlallkeyirnislkynevfkdtdtkngyagyidgktn  
 qedfyvylknllaefegadyflekidredflrkqrfdngsipyqihlqemraildkqakfypflaknkeriekiltfripyyvgplargnsdfaws  
 irkrnekitypwnfedvidkessaeafinrmtsfldlylpeekvlpkhslyetfnvyneltkvrfaesmrdyqfldskqkkdivrlyfkdkrvtd  
 kdiiylhaiygydielgkqfnsststyhdlhniindkeflddssneaieeihtltifedremikqrlskfenifdksvlkklsrrhytgwgl  
 saklingirdeksgntildyliddgisnrnfmqlihddalsfkkkiqkaqiigdedkgnikevkslpgspaikkgilqsikivdelvkvmmgrk  
 pesivvemarenqytnqgksnsqqrkrlekslkelgskilkenipaklskidnnaqndrlylylqngkdmtygdlddidrlnsydidaiipq  
 aflkdnidnkvlvssasnrgksddfpslevvkrktfwyqllksklisqrkfdnltkaerggllpedkagfiqrqlvetrqitkhvarlldekfnn  
 kkdenravrtvkiitlksltvsqfrkdfelykvreindfhahdaylnaviasallkkyplefvygdypkynsfrerksatekvyfysnim  
 nifkksisladgrvierplievneetgesvwnkesdlatvrrrvsypqvnvkvveeqnhgldrgkpkglfnanlsskpkpnsnenlvakey  
 ldpkkyggyagisnsfavlvkgtickgakkkitnvlefqgisildrinyrkdlnfllekgykdielielpkyslfelsdgsrrmlasilstnnkrge  
 ihkgnqiflsqkfvklyhakrisntinenhrkyvenhkkefeelfyylefnyvgyakkngklinsafqswqnhsidecslssfigtgserkgl  
 feltsrgsaadfevlgykipryrdytpssllkdatlihqsvtglyetridlaklgeg

**Tres letras:**

MetLeuPheAsnLysCysIleIleIleSerIleAsnLeuAspPheSerAsnLysGluLys  
 CysMetThrLysProTyrSerIleGlyLeuAspIleGlyThrAsnSerValGlyTrpAla  
 ValIleThrAspAsnTyrLysValProSerLysLysMetLysValLeuGlyAsnThrSer  
 LysLysTyrIleLysLysAsnLeuLeuGlyValLeuLeuPheAspSerGlyIleThrAla  
 GluGlyArgArgLeuLysArgThrAlaArgArgArgTyrThrArgArgArgAsnArgIle  
 LeuTyrLeuGlnGluIlePheSerThrGluMetAlaThrLeuAspAspAlaPhePheGln  
 ArgLeuAspAspSerPheLeuValProAspAspLysArgAspSerLysTyrProIlePhe  
 GlyAsnLeuValGluGluLysValTyrHisAspGluPheProThrIleTyrHisLeuArg  
 LysTyrLeuAlaAspSerThrLysLysAlaAspLeuArgLeuValTyrLeuAlaLeuAla  
 HisMetIleLysTyrArgGlyHisPheLeuIleGluGlyGluPheAsnSerLysAsnAsn  
 AspIleGlnLysAsnPheGlnAspPheLeuAspThrTyrAsnAlaIlePheGluSerAsp  
 LeuSerLeuGluAsnSerLysGlnLeuGluGluIleValLysAspLysIleSerLysLeu  
 GluLysLysAspArgIleLeuLysLeuPheProGlyGluLysAsnSerGlyIlePheSer  
 GluPheLeuLysLeuIleValGlyAsnGlnAlaAspPheArgLysCysPheAsnLeuAsp  
 GluLysAlaSerLeuHisPheSerLysGluSerTyrAspGluAspLeuGluThrLeuLeu

GlyTyrIleGlyAspAspTyrSerAspValPheLeuLysAlaLysLysLeuTyrAspAla  
 IleLeuLeuSerGlyPheLeuThrValThrAspAsnGluThrGluAlaProLeuSerSer  
 AlaMetIleLysArgTyrAsnGluHisLysGluAspLeuAlaLeuLeuLysGluTyrIle  
 ArgAsnIleSerLeuLysThrTyrAsnGluValPheLysAspAspThrLysAsnGlyTyr  
 AlaGlyTyrIleAspGlyLysThrAsnGlnGluAspPheTyrValTyrLeuLysAsnLeu  
 LeuAlaGluPheGluGlyAlaAspTyrPheLeuGluLysIleAspArgGluAspPheLeu  
 ArgLysGlnArgThrPheAspAsnGlySerIleProTyrGlnIleHisLeuGlnGluMet  
 ArgAlaIleLeuAspLysGlnAlaLysPheTyrProPheLeuAlaLysAsnLysGluArg  
 IleGluLysIleLeuThrPheArgIleProTyrTyrValGlyProLeuAlaArgGlyAsn  
 SerAspPheAlaTrpSerIleArgLysArgAsnGluLysIleThrProTrpAsnPheGlu  
 AspValIleAspLysGluSerSerAlaGluAlaPheIleAsnArgMetThrSerPheAsp  
 LeuTyrLeuProGluGluLysValLeuProLysHisSerLeuLeuTyrGluThrPheAsn  
 ValTyrAsnGluLeuThrLysValArgPheIleAlaGluSerMetArgAspTyrGlnPhe  
 LeuAspSerLysGlnLysLysAspIleValArgLeuTyrPheLysAspLysArgLysVal  
 ThrAspLysAspIleIleGluTyrLeuHisAlaIleTyrGlyTyrAspGlyIleGluLeu  
 LysGlyIleGluLysGlnPheAsnSerSerLeuSerThrTyrHisAspLeuLeuAsnIle  
 IleAsnAspLysGluPheLeuAspAspSerSerAsnGluAlaIleIleGluGluIleIle  
 HisThrLeuThrIlePheGluAspArgGluMetIleLysGlnArgLeuSerLysPheGlu  
 AsnIlePheAspLysSerValLeuLysLysLeuSerArgArgHisTyrThrGlyTrpGly  
 LysLeuSerAlaLysLeuIleAsnGlyIleArgAspGluLysSerGlyAsnThrIleLeu  
 AspTyrLeuIleAspAspGlyIleSerAsnArgAsnPheMetGlnLeuIleHisAspAsp  
 AlaLeuSerPheLysLysLysIleGlnLysAlaGlnIleIleGlyAspGluAspLysGly  
 AsnIleLysGluValValLysSerLeuProGlySerProAlaIleLysLysGlyIleLeu  
 GlnSerIleLysIleValAspGluLeuValLysValMetGlyGlyArgLysProGluSer  
 IleValValGluMetAlaArgGluAsnGlnTyrThrAsnGlnGlyLysSerAsnSerGln  
 GlnArgLeuLysArgLeuGluLysSerLeuLysGluLeuGlySerLysIleLeuLysGlu  
 AsnIleProAlaLysLeuSerLysIleAspAsnAsnAlaLeuGlnAsnAspArgLeuTyr  
 LeuTyrTyrLeuGlnAsnGlyLysAspMetTyrThrGlyAspAspLeuAspIleAspArg  
 LeuSerAsnTyrAspIleAspAlaIleIleProGlnAlaPheLeuLysAspAsnSerIle  
 AspAsnLysValLeuValSerSerAlaSerAsnArgGlyLysSerAspAspPheProSer  
 LeuGluValValLysLysArgLysThrPheTrpTyrGlnLeuLeuLysSerLysLeuIle  
 SerGlnArgLysPheAspAsnLeuThrLysAlaGluArgGlyGlyLeuLeuProGluAsp  
 LysAlaGlyPheIleGlnArgGlnLeuValGluThrArgGlnIleThrLysHisValAla  
 ArgLeuLeuAspGluLysPheAsnAsnLysLysAspGluAsnAsnArgAlaValArgThr  
 ValLysIleIleThrLeuLysSerThrLeuValSerGlnPheArgLysAspPheGluLeu  
 TyrLysValArgGluIleAsnAspPheHisHisAlaHisAspAlaTyrLeuAsnAlaVal

IleAlaSerAlaLeuLeuLysLysTyrProLysLeuGluProGluPheValTyrGlyAsp  
 TyrProLysTyrAsnSerPheArgGluArgLysSerAlaThrGluLysValTyrPheTyr  
 SerAsnIleMetAsnIlePheLysLysSerIleSerLeuAlaAspGlyArgValIleGlu  
 ArgProLeuIleGluValAsnGluGluThrGlyGluSerValTrpAsnLysGluSerAsp  
 LeuAlaThrValArgArgValLeuSerTyrProGlnValAsnValValLysLysValGlu  
 GluGlnAsnHisGlyLeuAspArgGlyLysProLysGlyLeuPheAsnAlaAsnLeuSer  
 SerLysProLysProAsnSerAsnGluAsnLeuValGlyAlaLysGluTyrLeuAspPro  
 LysLysTyrGlyGlyTyrAlaGlyIleSerAsnSerPheAlaValLeuValLysGlyThr  
 IleGluLysGlyAlaLysLysLysIleThrAsnValLeuGluPheGlnGlyIleSerIle  
 LeuAspArgIleAsnTyrArgLysAspLysLeuAsnPheLeuLeuGluLysGlyTyrLys  
 AspIleGluLeuIleIleGluLeuProLysTyrSerLeuPheGluLeuSerAspGlySer  
 ArgArgMetLeuAlaSerIleLeuSerThrAsnAsnLysArgGlyGluIleHisLysGly  
 AsnGlnIlePheLeuSerGlnLysPheValLysLeuLeuTyrHisAlaLysArgIleSer  
 AsnThrIleAsnGluAsnHisArgLysTyrValGluAsnHisLysLysGluPheGluGlu  
 LeuPheTyrTyrIleLeuGluPheAsnGluAsnTyrValGlyAlaLysLysAsnGlyLys  
 LeuLeuAsnSerAlaPheGlnSerTrpGlnAsnHisSerIleAspGluLeuCysSerSer  
 PheIleGlyProThrGlySerGluArgLysGlyLeuPheGluLeuThrSerArgGlySer  
 AlaAlaAspPheGluPheLeuGlyValLysIleProArgTyrArgAspTyrThrProSer  
 SerLeuLeuLysAspAlaThrLeuIleHisGlnSerValThrGlyLeuTyrGluThrArg  
 IleAspLeuAlaLysLeuGlyGluGly

**SEQ ID NO: 5**

ARNtra-cr, inmaduro (102 nt):

uaauaauuuugugguuuugaaaccauucgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcac  
 cgauucgguguuuuu

5 **SEQ ID NO: 6**

ARNtracr maduro de 78 nt:

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauucgguguuuuu

Variantes más cortas:

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauucggug (SEQ ID NO: 44)

10 gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcaccgauu (SEQ ID NO: 45)

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaagguggcac (SEQ ID NO: 46)

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaaaaggu (SEQ ID NO: 47)

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaacuugaa (SEQ ID NO: 48)

gggcgaaacaacacagcgaguuaaaauaaggcuuaguccguacucaac (SEQ ID NO: 49)

15 **SEQ ID NO: 7**

ARNcr de 42 nt procedente del espaciador 1:

5'-CGCUAAAGAGGAAGAGGACAGUUUUAGAGCUGUGUUGUUUCG-3'

**SEQ ID NO: 8**

pre-ARNcr de 150 nt

## ES 2 749 108 T3

5' -

ggguagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaacaaauucuaa  
**acgcuaaagaggaagaggaca**guuuuagagcuguguguuuucgaaugguuccaaaacuacugcuguau  
uagcuugguuguug-3'

### SEQ ID NO: 9

ARNcr1

5' -

ggguagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaactTGTCATGA  
TAATAATGGTTTCTTAGACGTCguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug  
uauuagcuugguuguug-3'

5

### SEQ ID NO: 10

ARNcr2

5' -

ggguagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaacacgagccg  
gaagcataaagtgtaaagcctgguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug  
uauuagcuugguuguug-3'

### SEQ ID NO: 11

ARN CRISPR anti-fago  $\lambda$

5' -

ggguagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaactcaagggga  
gaatagaggctctcgttgcatgguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug  
uauuagcuugguuguug-3'

10

### SEQ ID NO: 12

ARN CRISPR anti *E. coli*

5' -

ggguagaaaagauauccuacgagguuuuagagcuguguguuuucgaaugguuccaaaaccgggaggg  
aagctgcatgatgcatggttatgguuuuagagcuguguguuuucgaaugguuccaaaacuacugcug  
uauuagcuugguuguug-3'

15

### SEQ ID NO: 13

ARNcr-HS1

5'-GCUCCGGGGCUCGAUGAAGGUUUUAGAGCUGUGUUGUUUCG-3'

### SEQ ID NO: 14

ARNcr-HS2

UGAAUCGUGAAUCUGCUCAGUUUUAGAGCUGUGUUGUUUCG

20

La solicitud contiene una Lista de Secuencias que se ha presentado en formato ASCII y se incorpora a este documento como referencia en su totalidad. La copia ASCII, creada el 20 de marzo de 2013, se denomina 078981\_6\_SL.txt y tiene un tamaño de 64,4 kilobytes.



Las realizaciones mostradas y descritas en la memoria descriptiva son solo realizaciones específicas de inventores que son expertos en la materia y no son limitantes de ninguna manera.

LISTA DE SECUENCIAS

- <110> VILNIUS UNIVERSITY
  - 5 <120> ESCISIÓN DE AND DIRIGIDA POR ARN MEDIANTE EL COMPLEJO CAS9-ARNCR
  - <130> 078981.6
  - <140>
  - <141>
  - 10 <150> 61/625,420
  - <151> 2012-04-17
  - <150> 61/613,373
  - <151> 2012-03-20
  - <160> 69
  - <170> PatentIn versión 3.5
  - 15 <210> 1
  - <211> 1409
  - <212> PRT
  - <213> Streptococcus thermophilus
  - <400> 1
- ```

Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
1           5           10           15

Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile
          20           25           30

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
          35           40           45

Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile
          50           55           60

Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala
65           70           75           80

Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
          85           90           95

Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala
          100          105          110

Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val
          115          120          125

Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val
          130          135          140
    
```

ES 2 749 108 T3

Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg  
 145 150 155 160  
 Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr  
 165 170 175  
 Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu  
 180 185 190  
 Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp  
 195 200 205  
 Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu  
 210 215 220  
 Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu  
 225 230 235 240  
 Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser  
 245 250 255  
 Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp  
 260 265 270  
 Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser  
 275 280 285  
 Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly  
 290 295 300  
 Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala  
 305 310 315 320  
 Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala  
 325 330 335  
 Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp  
 340 345 350  
 Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr  
 355 360 365  
 Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile  
 370 375 380  
 Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu  
 385 390 395 400

ES 2 749 108 T3

Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg  
 405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro  
 420 425 430

Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala  
 435 440 445

Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile  
 450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn  
 465 470 475 480

Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro  
 485 490 495

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe  
 500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val  
 515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu  
 530 535 540

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe  
 545 550 555 560

Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp  
 565 570 575

Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile  
 580 585 590

Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn  
 595 600 605

Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys  
 610 615 620

Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile  
 625 630 635 640

His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu  
 645 650 655

ES 2 749 108 T3

Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser  
660 665 670

Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn  
675 680 685

Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile  
690 695 700

Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp  
705 710 715 720

Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp  
725 730 735

Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser  
740 745 750

Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu  
755 760 765

Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu  
770 775 780

Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln  
785 790 795 800

Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys  
805 810 815

Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn  
820 825 830

Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys  
835 840 845

Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr  
850 855 860

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile  
865 870 875 880

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp  
885 890 895

Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr

ES 2 749 108 T3

|     |      |     |     |     |     |      |      |     |     |     |     |      |      |     |     |     |
|-----|------|-----|-----|-----|-----|------|------|-----|-----|-----|-----|------|------|-----|-----|-----|
|     |      |     | 900 |     |     |      |      |     | 905 |     |     |      |      |     |     | 910 |
| Gln | Leu  | Leu | Lys | Ser | Lys | Leu  | Ile  | Ser | Gln | Arg | Lys | Phe  | Asp  | Asn | Leu |     |
|     |      | 915 |     |     |     |      | 920  |     |     |     |     | 925  |      |     |     |     |
| Thr | Lys  | Ala | Glu | Arg | Gly | Gly  | Leu  | Leu | Pro | Glu | Asp | Lys  | Ala  | Gly | Phe |     |
|     | 930  |     |     |     |     | 935  |      |     |     |     | 940 |      |      |     |     |     |
| Ile | Gln  | Arg | Gln | Leu | Val | Glu  | Thr  | Arg | Gln | Ile | Thr | Lys  | His  | Val | Ala |     |
| 945 |      |     |     |     | 950 |      |      |     |     | 955 |     |      |      |     | 960 |     |
| Arg | Leu  | Leu | Asp | Glu | Lys | Phe  | Asn  | Asn | Lys | Lys | Asp | Glu  | Asn  | Asn | Arg |     |
|     |      |     |     | 965 |     |      |      |     | 970 |     |     |      |      |     | 975 |     |
| Ala | Val  | Arg | Thr | Val | Lys | Ile  | Ile  | Thr | Leu | Lys | Ser | Thr  | Leu  | Val | Ser |     |
|     |      |     | 980 |     |     |      |      | 985 |     |     |     |      |      |     | 990 |     |
| Gln | Phe  | Arg | Lys | Asp | Phe | Glu  | Leu  | Tyr | Lys | Val | Arg | Glu  | Ile  | Asn | Asp |     |
|     |      | 995 |     |     |     |      | 1000 |     |     |     |     |      | 1005 |     |     |     |
| Phe | His  | His | Ala | His | Asp | Ala  | Tyr  | Leu | Asn | Ala | Val | Ile  | Ala  | Ser |     |     |
|     | 1010 |     |     |     |     | 1015 |      |     |     |     |     | 1020 |      |     |     |     |
| Ala | Leu  | Leu | Lys | Lys | Tyr | Pro  | Lys  | Leu | Glu | Pro | Glu | Phe  | Val  | Tyr |     |     |
|     | 1025 |     |     |     |     | 1030 |      |     |     |     |     | 1035 |      |     |     |     |
| Gly | Asp  | Tyr | Pro | Lys | Tyr | Asn  | Ser  | Phe | Arg | Glu | Arg | Lys  | Ser  | Ala |     |     |
|     | 1040 |     |     |     |     | 1045 |      |     |     |     |     | 1050 |      |     |     |     |
| Thr | Glu  | Lys | Val | Tyr | Phe | Tyr  | Ser  | Asn | Ile | Met | Asn | Ile  | Phe  | Lys |     |     |
|     | 1055 |     |     |     |     | 1060 |      |     |     |     |     | 1065 |      |     |     |     |
| Lys | Ser  | Ile | Ser | Leu | Ala | Asp  | Gly  | Arg | Val | Ile | Glu | Arg  | Pro  | Leu |     |     |
|     | 1070 |     |     |     |     | 1075 |      |     |     |     |     | 1080 |      |     |     |     |
| Ile | Glu  | Val | Asn | Glu | Glu | Thr  | Gly  | Glu | Ser | Val | Trp | Asn  | Lys  | Glu |     |     |
|     | 1085 |     |     |     |     | 1090 |      |     |     |     |     | 1095 |      |     |     |     |
| Ser | Asp  | Leu | Ala | Thr | Val | Arg  | Arg  | Val | Leu | Ser | Tyr | Pro  | Gln  | Val |     |     |
|     | 1100 |     |     |     |     | 1105 |      |     |     |     |     | 1110 |      |     |     |     |
| Asn | Val  | Val | Lys | Lys | Val | Glu  | Glu  | Gln | Asn | His | Gly | Leu  | Asp  | Arg |     |     |
|     | 1115 |     |     |     |     | 1120 |      |     |     |     |     | 1125 |      |     |     |     |
| Gly | Lys  | Pro | Lys | Gly | Leu | Phe  | Asn  | Ala | Asn | Leu | Ser | Ser  | Lys  | Pro |     |     |
|     | 1130 |     |     |     |     | 1135 |      |     |     |     |     | 1140 |      |     |     |     |

ES 2 749 108 T3

Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu  
 1145 1150 1155

Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe  
 1160 1165 1170

Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys  
 1175 1180 1185

Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg  
 1190 1195 1200

Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly  
 1205 1210 1215

Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu  
 1220 1225 1230

Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu  
 1235 1240 1245

Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile  
 1250 1255 1260

Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg  
 1265 1270 1275

Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn  
 1280 1285 1290

His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe  
 1295 1300 1305

Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn  
 1310 1315 1320

Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys  
 1325 1330 1335

Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe  
 1340 1345 1350

Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly  
 1355 1360 1365

Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu  
 1370 1375 1380

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu  
 1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly  
 1400 1405

<210> 2  
 <211> 1409  
 <212> PRT  
 <213> Streptococcus thermophilus

ES 2 749 108 T3

<400> 2

Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser  
 1 5 10 15

Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Ala Ile  
 20 25 30

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val  
 35 40 45

Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile  
 50 55 60

Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala  
 65 70 75 80

Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg  
 85 90 95

Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala  
 100 105 110

Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val  
 115 120 125

Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val  
 130 135 140

Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg  
 145 150 155 160

Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr  
 165 170 175

Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu  
 180 185 190

ES 2 749 108 T3

Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp  
 195 200 205

Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu  
 210 215 220

Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu  
 225 230 235 240

Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser  
 245 250 255

Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp  
 260 265 270

Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser  
 275 280 285

Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly  
 290 295 300

Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala  
 305 310 315 320

Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala  
 325 330 335

Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp  
 340 345 350

Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr  
 355 360 365

Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile  
 370 375 380

Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu  
 385 390 395 400

Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg  
 405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro  
 420 425 430

Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala  
 435 440 445



ES 2 749 108 T3

Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile  
450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn  
465 470 475 480

Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro  
485 490 495

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe  
500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val  
515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu  
530 535 540

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe  
545 550 555 560

Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp  
565 570 575

Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile  
580 585 590

Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn  
595 600 605

Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys  
610 615 620

Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile  
625 630 635 640

His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu  
645 650 655

Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser  
660 665 670

Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn  
675 680 685

Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile  
690 695 700

ES 2 749 108 T3

Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp  
 705 710 715 720  
 Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp  
 725 730 735  
 Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser  
 740 745 750  
 Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu  
 755 760 765  
 Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu  
 770 775 780  
 Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln  
 785 790 795 800  
 Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys  
 805 810 815  
 Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn  
 820 825 830  
 Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys  
 835 840 845  
 Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr  
 850 855 860  
 Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile  
 865 870 875 880  
 Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp  
 885 890 895  
 Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr  
 900 905 910  
 Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu  
 915 920 925  
 Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe  
 930 935 940  
 Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala



ES 2 749 108 T3

Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg  
 1190 1195 1200

Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly  
 1205 1210 1215

Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu  
 1220 1225 1230

Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu  
 1235 1240 1245

Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile  
 1250 1255 1260

Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg  
 1265 1270 1275

Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn  
 1280 1285 1290

His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe  
 1295 1300 1305

Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn  
 1310 1315 1320

Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys  
 1325 1330 1335

Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe  
 1340 1345 1350

Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly  
 1355 1360 1365

Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu  
 1370 1375 1380

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu  
 1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly  
 1400 1405

<210> 3  
 < 211> 1409  
 < 212> PRT  
 < 213> Streptococcus thermophilus  
 <400> 3

5

ES 2 749 108 T3

Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser  
1 5 10 15

Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile  
20 25 30

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val  
35 40 45

Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile  
50 55 60

Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala  
65 70 75 80

Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg  
85 90 95

Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala  
100 105 110

Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val  
115 120 125

Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val  
130 135 140

Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg  
145 150 155 160

Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr  
165 170 175

Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu  
180 185 190

Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp  
195 200 205

Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu  
210 215 220

Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu  
225 230 235 240

ES 2 749 108 T3

Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser  
 245 250 255  
 Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp  
 260 265 270  
 Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser  
 275 280 285  
 Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly  
 290 295 300  
 Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala  
 305 310 315 320  
 Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala  
 325 330 335  
 Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp  
 340 345 350  
 Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr  
 355 360 365  
 Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile  
 370 375 380  
 Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu  
 385 390 395 400  
 Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg  
 405 410 415  
 Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro  
 420 425 430  
 Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala  
 435 440 445  
 Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile  
 450 455 460  
 Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn  
 465 470 475 480  
 Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro  
 485 490 495

ES 2 749 108 T3

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe  
 500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val  
 515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu  
 530 535 540

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe  
 545 550 555 560

Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp  
 565 570 575

Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile  
 580 585 590

Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn  
 595 600 605

Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys  
 610 615 620

Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile  
 625 630 635 640

His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu  
 645 650 655

Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser  
 660 665 670

Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn  
 675 680 685

Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile  
 690 695 700

Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp  
 705 710 715 720

Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp  
 725 730 735

Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser  
 740 745 750

ES 2 749 108 T3

Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu  
           755                                  760                                  765

Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu  
       770                                  775                                  780

Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln  
 785                                  790                                  795                                  800

Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys  
                                   805                                  810                                  815

Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn  
                                   820                                  825                                  830

Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys  
                                   835                                  840                                  845

Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr  
       850                                  855                                  860

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile  
 865                                  870                                  875                                  880

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Ala Arg Gly Lys Ser Asp  
                                   885                                  890                                  895

Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr  
                                   900                                  905                                  910

Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu  
                                   915                                  920                                  925

Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe  
       930                                  935                                  940

Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala  
 945                                  950                                  955                                  960

Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg  
                                   965                                  970                                  975

Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser  
                                   980                                  985                                  990

Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp



ES 2 749 108 T3

| 995  |     |     |     | 1000 |     |      |     | 1005 |     |     |      |     |     |     |
|------|-----|-----|-----|------|-----|------|-----|------|-----|-----|------|-----|-----|-----|
| Phe  | His | His | Ala | His  | Asp | Ala  | Tyr | Leu  | Asn | Ala | Val  | Ile | Ala | Ser |
| 1010 |     |     |     |      |     | 1015 |     |      |     |     | 1020 |     |     |     |
| Ala  | Leu | Leu | Lys | Lys  | Tyr | Pro  | Lys | Leu  | Glu | Pro | Glu  | Phe | Val | Tyr |
| 1025 |     |     |     |      |     | 1030 |     |      |     |     | 1035 |     |     |     |
| Gly  | Asp | Tyr | Pro | Lys  | Tyr | Asn  | Ser | Phe  | Arg | Glu | Arg  | Lys | Ser | Ala |
| 1040 |     |     |     |      |     | 1045 |     |      |     |     | 1050 |     |     |     |
| Thr  | Glu | Lys | Val | Tyr  | Phe | Tyr  | Ser | Asn  | Ile | Met | Asn  | Ile | Phe | Lys |
| 1055 |     |     |     |      |     | 1060 |     |      |     |     | 1065 |     |     |     |
| Lys  | Ser | Ile | Ser | Leu  | Ala | Asp  | Gly | Arg  | Val | Ile | Glu  | Arg | Pro | Leu |
| 1070 |     |     |     |      |     | 1075 |     |      |     |     | 1080 |     |     |     |
| Ile  | Glu | Val | Asn | Glu  | Glu | Thr  | Gly | Glu  | Ser | Val | Trp  | Asn | Lys | Glu |
| 1085 |     |     |     |      |     | 1090 |     |      |     |     | 1095 |     |     |     |
| Ser  | Asp | Leu | Ala | Thr  | Val | Arg  | Arg | Val  | Leu | Ser | Tyr  | Pro | Gln | Val |
| 1100 |     |     |     |      |     | 1105 |     |      |     |     | 1110 |     |     |     |
| Asn  | Val | Val | Lys | Lys  | Val | Glu  | Glu | Gln  | Asn | His | Gly  | Leu | Asp | Arg |
| 1115 |     |     |     |      |     | 1120 |     |      |     |     | 1125 |     |     |     |
| Gly  | Lys | Pro | Lys | Gly  | Leu | Phe  | Asn | Ala  | Asn | Leu | Ser  | Ser | Lys | Pro |
| 1130 |     |     |     |      |     | 1135 |     |      |     |     | 1140 |     |     |     |
| Lys  | Pro | Asn | Ser | Asn  | Glu | Asn  | Leu | Val  | Gly | Ala | Lys  | Glu | Tyr | Leu |
| 1145 |     |     |     |      |     | 1150 |     |      |     |     | 1155 |     |     |     |
| Asp  | Pro | Lys | Lys | Tyr  | Gly | Gly  | Tyr | Ala  | Gly | Ile | Ser  | Asn | Ser | Phe |
| 1160 |     |     |     |      |     | 1165 |     |      |     |     | 1170 |     |     |     |
| Ala  | Val | Leu | Val | Lys  | Gly | Thr  | Ile | Glu  | Lys | Gly | Ala  | Lys | Lys | Lys |
| 1175 |     |     |     |      |     | 1180 |     |      |     |     | 1185 |     |     |     |
| Ile  | Thr | Asn | Val | Leu  | Glu | Phe  | Gln | Gly  | Ile | Ser | Ile  | Leu | Asp | Arg |
| 1190 |     |     |     |      |     | 1195 |     |      |     |     | 1200 |     |     |     |
| Ile  | Asn | Tyr | Arg | Lys  | Asp | Lys  | Leu | Asn  | Phe | Leu | Leu  | Glu | Lys | Gly |
| 1205 |     |     |     |      |     | 1210 |     |      |     |     | 1215 |     |     |     |
| Tyr  | Lys | Asp | Ile | Glu  | Leu | Ile  | Ile | Glu  | Leu | Pro | Lys  | Tyr | Ser | Leu |
| 1220 |     |     |     |      |     | 1225 |     |      |     |     | 1230 |     |     |     |

ES 2 749 108 T3

Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu  
 1235 1240 1245

Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile  
 1250 1255 1260

Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg  
 1265 1270 1275

Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn  
 1280 1285 1290

His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe  
 1295 1300 1305

Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn  
 1310 1315 1320

Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys  
 1325 1330 1335

Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe  
 1340 1345 1350

Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly  
 1355 1360 1365

Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu  
 1370 1375 1380

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu  
 1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly  
 1400 1405

<210> 4  
 <211> 1409  
 <212> PRT  
 <213> Streptococcus thermophilus  
 <400> 4

5

Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser  
 1 5 10 15

Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile  
 20 25 30

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val



ES 2 749 108 T3

Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly  
 290 295 300

Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala  
 305 310 315 320

Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala  
 325 330 335

Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp  
 340 345 350

Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr  
 355 360 365

Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile  
 370 375 380

Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu  
 385 390 395 400

Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg  
 405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro  
 420 425 430

Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala  
 435 440 445

Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile  
 450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn  
 465 470 475 480

Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro  
 485 490 495

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe  
 500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val  
 515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu  
 530 535 540

ES 2 749 108 T3

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe  
 545 550 555 560

Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp  
 565 570 575

Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile  
 580 585 590

Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn  
 595 600 605

Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys  
 610 615 620

Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile  
 625 630 635 640

His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu  
 645 650 655

Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser  
 660 665 670

Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn  
 675 680 685

Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile  
 690 695 700

Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp  
 705 710 715 720

Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp  
 725 730 735

Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser  
 740 745 750

Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu  
 755 760 765

Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu  
 770 775 780

Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln  
 785 790 795 800

ES 2 749 108 T3

Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys  
805 810 815

Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn  
820 825 830

Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys  
835 840 845

Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr  
850 855 860

Asp Ile Asp Ala Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile  
865 870 875 880

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp  
885 890 895

Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr  
900 905 910

Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu  
915 920 925

Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe  
930 935 940

Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala  
945 950 955 960

Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg  
965 970 975

Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser  
980 985 990

Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp  
995 1000 1005

Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser  
1010 1015 1020

Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr  
1025 1030 1035

Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala



ES 2 749 108 T3

Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn  
 1280 1285 1290

His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe  
 1295 1300 1305

Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn  
 1310 1315 1320

Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys  
 1325 1330 1335

Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe  
 1340 1345 1350

Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly  
 1355 1360 1365

Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu  
 1370 1375 1380

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu  
 1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly  
 1400 1405

5 <210> 5  
 <211> 102  
 <212> ARN  
 <213> Streptococcus thermophilus

<400> 5

uaaauaauu ugugguuga aaccauucga aacaacacag cgaguuaaaa uaaggcuuag 60

uccguacuca acuugaaaag guggcaccga uucgguguuu uu 102

10 <210> 6  
 <211> 78  
 <212> ARN  
 <213> Streptococcus thermophilus

<400> 6

gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaaaaggugg 60

caccgauucg guguuuuu 78

15 <210> 7  
 <211> 42  
 <212> ARN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 7

cgcuaaagag gaagaggaca guuuuagagc uguguuuuu cg 42

<210> 8



ES 2 749 108 T3

< 211> 150  
 < 212> ARN  
 < 213> Secuencia artificial

5

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Polinucleótido sintético"

<400> 8

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaaca | 60  |
| aaauucuaaac gcuaaagagg aagaggacag uuuuagagcu guguuguuuc gauggguucc | 120 |
| aaaacuacug cuguauuagc uugguuguug                                   | 150 |

10

<210> 9  
 < 211> 150  
 < 212> ADN  
 < 213> Secuencia artificial

15

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la molécula combinada de ADN/ARN: Oligonucleótido sintético"

<400> 9

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact | 60  |
| gtcatgataa taatggtttc ttagacgtcg uuuuagagcu guguuguuuc gauggguucc | 120 |
| aaaacuacug cuguauuagc uugguuguug                                  | 150 |

20

<210> 10  
 < 211> 150  
 < 212> ADN  
 < 213> Secuencia artificial

25

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la molécula combinada de ADN/ARN: Oligonucleótido sintético"

30

<400> 10

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaaaca | 60  |
| cgagccggaa gcataaagtg taaagcctgg uuuuagagcu guguuguuuc gauggguucc  | 120 |
| aaaacuacug cuguauuagc uugguuguug                                   | 150 |

35

<210> 11  
 < 211> 150  
 < 212> ADN  
 < 213> Secuencia artificial

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

40

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la molécula combinada de ADN/ARN: Oligonucleótido sintético"

ES 2 749 108 T3

<400> 11

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaact 60

caagggagaa tagaggctct cgttgcatg uuuuagagcu guguuguuuc gaugguucc 120

aaaacuacug cuguuuuagc uugguuguug 150

<210> 12

< 211> 150

< 212> ADN

< 213> Secuencia artificial

<220>

< 221> fuente

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

10 <220>

< 221> fuente

< 223> /nota="Descripción de la molécula combinada de ADN/ARN: Oligonucleótido sintético"

<400> 12

ggguagaaaa gauauccuac gagguuuuag agcuguguug uuucgaaugg uuccaaaacc 60

gggaggggaag ctgcatgatg cgatgttatg uuuuagagcu guguuguuuc gaugguucc 120

aaaacuacug cuguuuuagc uugguuguug 150

15 <210> 13

< 211> 42

< 212> ARN

< 213> Secuencia artificial

<220>

< 221> fuente

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

20 <400> 13

gcucccgggg cucgaugaag guuuuagagc uguguuguuu cg 42

25 <210> 14

< 211> 42

< 212> ARN

< 213> Secuencia artificial

<220>

< 221> fuente

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

30 <400> 14

ugaaucguga aaucugcuca guuuuagagc uguguuguuu cg 42

<210> 15

< 211> 42

< 212> ARN

< 213> Secuencia artificial

<220>

< 221> fuente

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

40 <220>

< 221> modified\_base

< 222> (1)..(20)

< 223> a, c, u, g, desconocidas o distintas

<400> 15

nnnnnnnnnn nnnnnnnnnn guuuuagagc uguguuuuu cg 42

<210> 16  
 < 211> 29  
 < 212> ADN  
 5 < 213> Secuencia artificial

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 16  
 10 ccaccagca aaattcggtt ttctggctg 29

<210> 17  
 < 211> 36  
 < 212> ADN  
 < 213> Secuencia artificial

15 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 17  
 20 taatcgact cactataggg taccgagctc gaattg 36

<210> 18  
 < 211> 31  
 < 212> ADN  
 < 213> Secuencia artificial

25 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 18  
 30 gggaaacagc tatgaccatg attacgaatt c 31

<210> 19  
 < 211> 31  
 < 212> ADN  
 < 213> Secuencia artificial

35 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 19  
 gggtagcgag ctggaattga aattctaac g 31

<210> 20  
 < 211> 43  
 40 < 212> ADN  
 < 213> Secuencia artificial

<220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

45 <400> 20  
 taatcgact cactataggg aaacagctat gaccatgatt acg 43

<210> 21  
 < 211> 38  
 < 212> ADN  
 50 < 213> Secuencia artificial

<220>  
 < 221> fuente

# ES 2 749 108 T3

< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 21  
acgtctcaaa tgggttttaa taagtgata ataattc 38

5 <210> 22  
< 211> 29  
< 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
10 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 22  
acgtctccgc gctaccctct cctagttg 29

15 <210> 23  
< 211> 6  
< 212> PRT  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Marcador 6xHis sintético"

20 <400> 23  
His His His His His His

1 5

<210> 24  
< 211> 34  
25 < 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

30 <400> 24  
acgtctcaca tgactaagcc atactcaatt ggac 34

<210> 25  
< 211> 27  
35 < 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

40 <400> 25  
actcgagacc ctctcctagt ttgcaa 27

<210> 26  
< 211> 23  
< 212> ADN  
< 213> Secuencia artificial

45 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 26  
gaccattat tgaggtaaat gag 23

50 <210> 27  
< 211> 28

# ES 2 749 108 T3

< 212> ADN  
< 213> Secuencia artificial

5 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 27  
caaaccagga tccaagctaa tacagcag 28

10 <210> 28  
< 211> 42  
< 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

15 <400> 28  
tcgaaacaac acagctctaa aactgtcctc ttctcttta gc 42

<210> 29  
< 211> 21  
< 212> ADN  
20 < 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

25 <400> 29  
ccgcatcagg cgccattcg c 21

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

30 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"

<400> 30  
gcgaggaagc ggaagagcgc cc 22

35 <210> 31  
< 211> 55  
< 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

40 <400> 31  
gctcgaattg aaattctaaa cgctaaagag gaagaggaca tggatgaattc gtaat 55

45 <210> 32  
< 211> 50  
< 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

50 <400> 32  
gctcgaattg aaattctaaa cgctaaagag gaagaggaca aattcgtaat 50

ES 2 749 108 T3

<210> 33  
< 211> 55  
< 212> ADN  
< 213> Secuencia artificial

5 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 33  
gctcgaattg tactgctgta ttagcttggg ttgtggttg tgggaattc gtaat 55

10 <210> 34  
< 211> 55  
< 212> ADN  
< 213> Secuencia artificial

15 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 34  
attacgaatt caccatgtcc tctcctctt tagcgttag aattcaatt cgagc 55

20 <210> 35  
< 211> 50  
< 212> ADN  
< 213> Secuencia artificial

25 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 35  
attacgaatt tgcctctc ctcttagcg ttagaattt caattcgagc 50

30 <210> 36  
< 211> 55  
< 212> ADN  
< 213> Secuencia artificial

35 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 36  
attacgaatt caccacaaac caacaacaa gctaatacag cagtacaatt cgagc 55

40 <210> 37  
< 211> 55  
< 212> ADN  
< 213> Secuencia artificial

45 <220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 37  
gctcgaattg aaattctaaa cgctaaagag gaagaggaca tgggaattc gtaat 55

50 <210> 38  
< 211> 45  
< 212> ADN  
< 213> Secuencia artificial

<220>  
< 221> fuente  
< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

ES 2 749 108 T3

<400> 38  
 gctcgaattg cgctaaagag gaagaggaca tgggaattc gtaat 45  
  
 <210> 39  
 < 211> 55  
 5 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 10 <400> 39  
 gctcgaattg ccaccagca aaattcggtt ttctggctga tgggaattc gtaat 55  
  
 <210> 40  
 < 211> 41  
 < 212> ADN  
 15 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"  
  
 20 <400> 40  
 taatacgact cactataggg tagaaaagat atctacgag g 41  
  
 <210> 41  
 < 211> 23  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 25 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"  
  
 <400> 41  
 caacaaccaa gctaatacag cag 23  
  
 30 <210> 42  
 < 211> 22  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"  
  
 35 <400> 42  
 aaaaacaccg aatcggtgcc ac 22  
  
 <210> 43  
 < 211> 48  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Cebador sintético"  
  
 45 <400> 43  
 taatacgact cactataggg taataataat tgggttga aaccattc 48  
  
 <210> 44  
 < 211> 73  
 < 212> ARN  
 < 213> Streptococcus thermophilus  
  
 50 <400> 44

ES 2 749 108 T3

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
|    | gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaaaaggugg        | 60 |
|    | caccgauucg gug                                                           | 73 |
| 5  | <210> 45<br>< 211> 68<br>< 212> ARN<br>< 213> Streptococcus thermophilus |    |
|    | <400> 45                                                                 |    |
|    | gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaaaaggugg        | 60 |
|    | caccgauu                                                                 | 68 |
| 10 | <210> 46<br>< 211> 63<br>< 212> ARN<br>< 213> Streptococcus thermophilus |    |
|    | <400> 46                                                                 |    |
|    | gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaaaaggugg        | 60 |
|    | cac                                                                      | 63 |
| 15 | <210> 47<br>< 211> 58<br>< 212> ARN<br>< 213> Streptococcus thermophilus |    |
|    | <400> 47                                                                 |    |
| 20 | gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaaaaggu          | 58 |
|    | <210> 48<br>< 211> 53<br>< 212> ARN<br>< 213> Streptococcus thermophilus |    |
| 25 | <400> 48<br>gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaacuu gaa   | 53 |
|    | <210> 49<br>< 211> 48<br>< 212> ARN<br>< 213> Streptococcus thermophilus |    |
| 30 | <400> 49<br>gggcgaaaca acacagcgag uaaaaaauag gcuuaguccg uacucaac         | 48 |
|    | <210> 50<br>< 211> 36<br>< 212> ADN<br>< 213> Streptococcus thermophilus |    |
| 35 | <400> 50<br>gttttagagc tgtgtgttt cgaatggtc caaaac                        | 36 |
|    | <210> 51<br>< 211> 36<br>< 212> ADN<br>< 213> Streptococcus pyrogenes    |    |
| 40 | <400> 51<br>gttttagagc tatgctgttt tgaatggtcc caaaac                      | 36 |
|    | <210> 52<br>< 211> 36<br>< 212> ADN<br>< 213> Streptococcus thermophilus |    |
| 45 |                                                                          |    |



<400> 52  
 ttaactcgc tgtgtggtt cgaatggttt caaacc 36  
  
 <210> 53  
 < 211> 36  
 5 < 212> ADN  
 < 213> Streptococcus pyrogenes  
  
 <400> 53  
 ttaactcgc tatgctggtt tgaatggttc caacaa 36  
  
 <210> 54  
 10 < 211> 42  
 < 212> ARN  
 < 213> Streptococcus pyrogenes  
  
 <400> 54  
 gauuucuucu ugcgcuuuuu guuuuagagc uaugcuguuu ug 42  
 15 <210> 55  
 < 211> 42  
 < 212> ARN  
 < 213> Streptococcus thermophilus  
  
 <400> 55  
 20 guucacugua cgaguacuua guuuuagagc uguguuuuu cg 42  
  
 <210> 56  
 < 211> 42  
 < 212> ARN  
 < 213> Streptococcus thermophilus  
  
 <400> 56  
 25 cgcuaaagag gaagaggaca guuuuagagc uguguuuuu cg 42  
  
 <210> 57  
 < 211> 35  
 30 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 57  
 35 aaattctaaa cgctaaagag gaagaggaca tgggtg 35  
  
 <210> 58  
 < 211> 18  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 40 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 58  
 45 aggaagagga catggtga 18  
  
 <210> 59  
 < 211> 10  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 50 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 59

ES 2 749 108 T3

aggaagagga            10  
 <210> 60  
 < 211> 10  
 < 212> ADN  
 5    < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 60  
 10    tacatggtga            10  
  
 <210> 61  
 < 211> 35  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 15    <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 61  
 20    caccatgtcc tcttcctctt tagcgtttag aattt            35  
  
 <210> 62  
 < 211> 32  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 25    <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 62  
 30    caccatgtcc tcttcctctt tagcgtttag aa            32  
  
 <210> 63  
 < 211> 28  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 35    <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 63  
 40    caccatgtcc tcttcctctt tagcggtt            28  
  
 <210> 64  
 < 211> 24  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 45    <220>  
 < 221> fuente  
 < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"  
  
 <400> 64  
 50    caccatgtcc tcttcctctt tagc            24  
  
 <210> 65  
 < 211> 20  
 < 212> ADN  
 < 213> Secuencia artificial  
  
 <220>  
 < 221> fuente

# ES 2 749 108 T3

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 65  
caccatgtcc tcttctctt        20

5        <210> 66  
         < 211> 16  
         < 212> ADN  
         < 213> Secuencia artificial

<220>  
< 221> fuente

10        < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 66  
caccatgtcc tcttcc        16

15        <210> 67  
         < 211> 55  
         < 212> ADN  
         < 213> Secuencia artificial

<220>  
< 221> fuente

20        < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 67  
attacgaatt ctctgtgcc tcttctctt tagcgtttag aattcaatt cgagc        55

25        <210> 68  
         < 211> 55  
         < 212> ADN  
         < 213> Secuencia artificial

<220>  
< 221> fuente

30        < 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 68  
attacgaatt gtggttgcc tcttctctt tagcgtttag aattcaatt cgagc        55

35        <210> 69  
         < 211> 42  
         < 212> ARN  
         < 213> Secuencia artificial

<220>  
< 221> fuente

< 223> /nota="Descripción de la Secuencia artificial: Oligonucleótido sintético"

<400> 69  
gcuaaagagg aagaggacag uuuuagagcu guguuguuuc ga        42

## REIVINDICACIONES

1. Un procedimiento *in vitro* para la modificación específica de sitio de una molécula de ADN diana, comprendiendo el procedimiento poner en contacto, en condiciones adecuadas,
- 5 una molécula de ADN diana; y  
una endonucleasa de ADN guiada por ARN que comprende al menos una secuencia de ARN y al menos uno de un motivo de sitio activo RuvC y un motivo de sitio activo HNH;
- en el que  
la endonucleasa de ADN guiada por ARN es un complejo Cas9-ARNcr  
comprendiendo dicho complejo Cas9-ARNcr una Cas9, ARNcr y ARNtracr  
10 para dar como resultado la molécula de ADN diana modificada en una región que está determinada por la unión complementaria del ARNcr a la molécula de ADN diana,  
comprendiendo el procedimiento adicionalmente ensamblar el complejo polipéptido-polirribonucleótidos *in vitro*  
incubando dichos componentes del complejo en condiciones adecuadas para el ensamblaje del complejo, en el que  
el ADN diana es bicatenario o monocatenario, y en el que un ADN diana bicatenario contiene un motivo adyacente al  
15 protoespaciador.
2. El procedimiento de la reivindicación 1, en el que el complejo comprende
- a) un polipéptido que tiene al menos el 80 % de identidad con la SEQ ID NO: 1,  
b) un primer ARNcr polirribonucleotídico que comprende una región 3' y una región 5', en la que la región 3'  
20 comprende al menos 22 nucleótidos de una repetición presente en un locus CRISPR y la región 5' comprende al menos 20 nucleótidos de una secuencia espaciadora inmediatamente corriente abajo de la repetición en el locus CRISPR, y  
c) un segundo ARNtracr polirribonucleotídico que comprende una región 5' y 3' en las que al menos una porción de la región 5' es complementaria a la región 3' del ARNcr.
3. El procedimiento de la reivindicación 2, en el que la región 5' del ARNtracr es al menos de 22 nucleótidos y es  
25 complementaria a la región 3' de 22 nucleótidos del ARNcr.
4. El procedimiento de la reivindicación 2, en el que los componentes del complejo se aíslan de un microorganismo modificado genéticamente.
5. El procedimiento de la reivindicación 2, en el que el polipéptido se aísla de una *Escherichia coli* modificada genéticamente.
- 30 6. El procedimiento de la reivindicación 2, en el que el polipéptido se produce por un procedimiento seleccionado de tecnología de ADN recombinante o síntesis química.
7. El procedimiento de la reivindicación 2, en el que el polipéptido se expresa como un polipéptido de fusión que incluye al menos una secuencia adicional de aminoácidos.
8. El procedimiento de la reivindicación 7, en el que la secuencia adicional se utiliza para la purificación del polipéptido  
35 de fusión mediante cromatografía de afinidad.
9. El procedimiento de la reivindicación 2, en el que el polipéptido contiene una mutación puntual en el motivo de sitio activo RuvC.
10. El procedimiento de la reivindicación 9, en el que la mutación puntual es D31A.
11. El procedimiento de la reivindicación 2, en el que el polipéptido contiene una mutación puntual en el motivo de sitio  
40 activo HNH.
12. El procedimiento de la reivindicación 11, en el que la mutación puntual es N891A.
13. El procedimiento de la reivindicación 2, en el que la primera secuencia polirribonucleotídica (ARNcr) comprende 5'-NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUUUUCG-3' (SEQ ID NO: 15).
14. El procedimiento de la reivindicación 2, en el que el complejo se genera *in vitro*, comprendiendo el procedimiento  
45 incubar los componentes del complejo en condiciones adecuadas para el ensamblaje del complejo,  
en el que el complejo comprende el primer polirribonucleótido que tiene una secuencia que comprende 5'-NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUUUUCG-3' (SEQ ID NO: 15) con cualquier  
secuencia espaciadora conveniente.
15. El procedimiento de la reivindicación 14, en el que el primer polirribonucleótido se obtiene mediante transcripción  
50 *in vitro* a partir de un fragmento de ADN que contiene una sola unidad de repetición-espaciador-repetición, donde el

espaciador tiene cualquier secuencia conveniente, o se sintetiza químicamente.

16. El procedimiento de la reivindicación 14, en el que los componentes comprenden polipéptido Cas9 (SEQ ID NO: 1), polirribonucleótido ARNtracr (SEQ ID NO: 5) y polirribonucleótido ARNcr (5'-NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUGUGUUUUUCG-3') (SEQ ID NO: 15) con cualquier secuencia espaciadora conveniente.

5

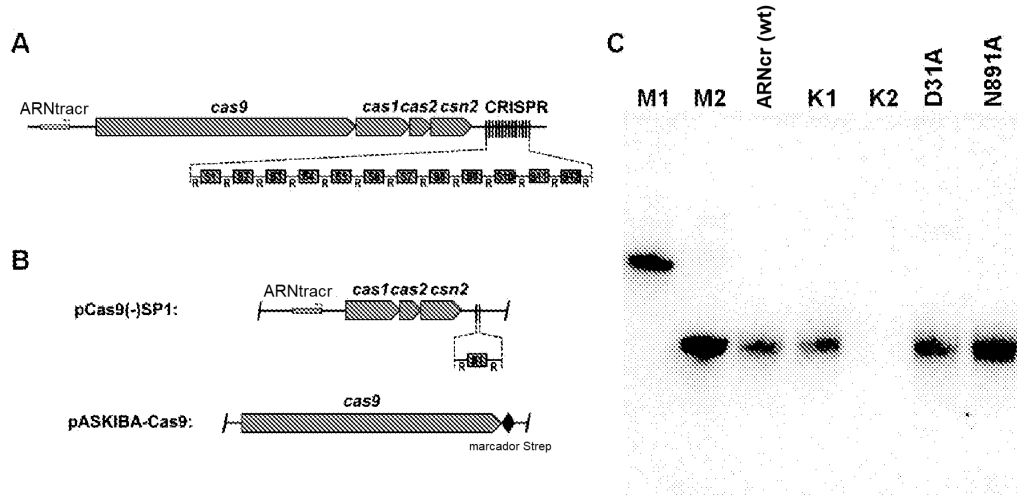


Figura 1.

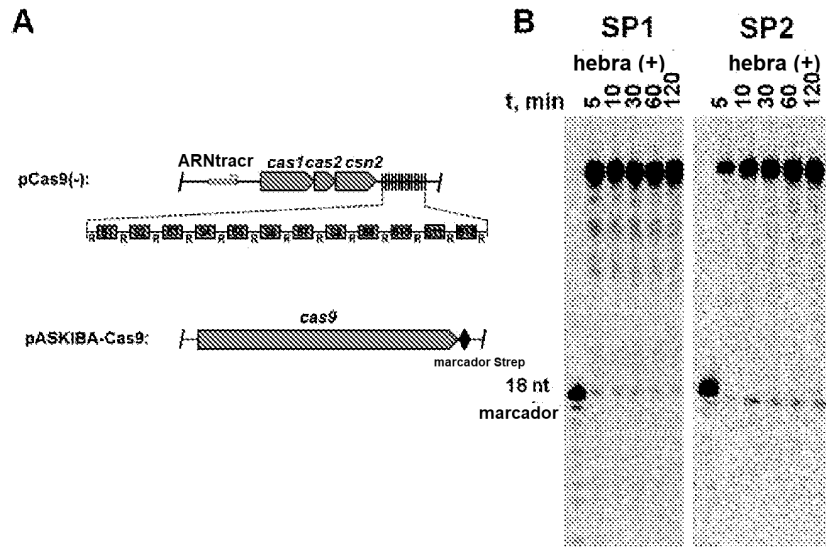


Figura 2.

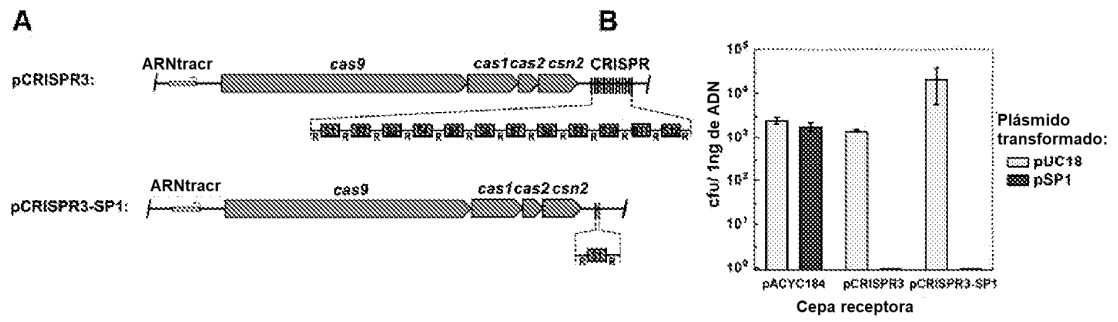


Figura 3.







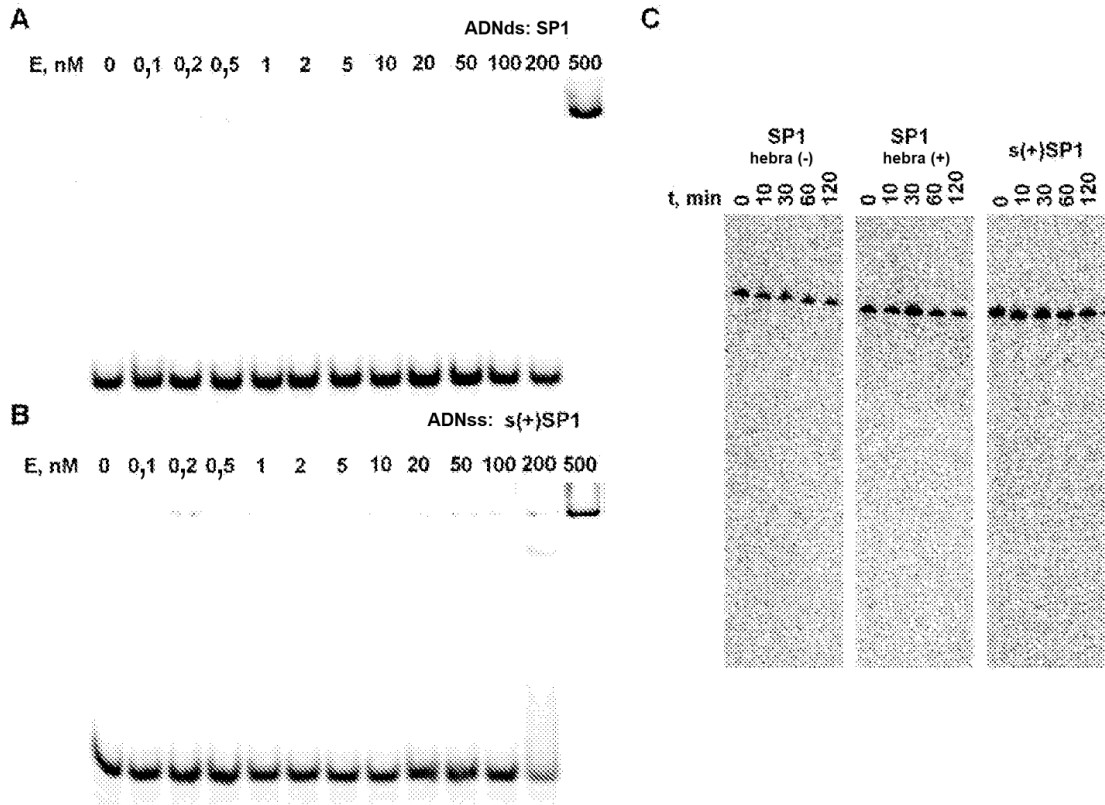


Figura 6.

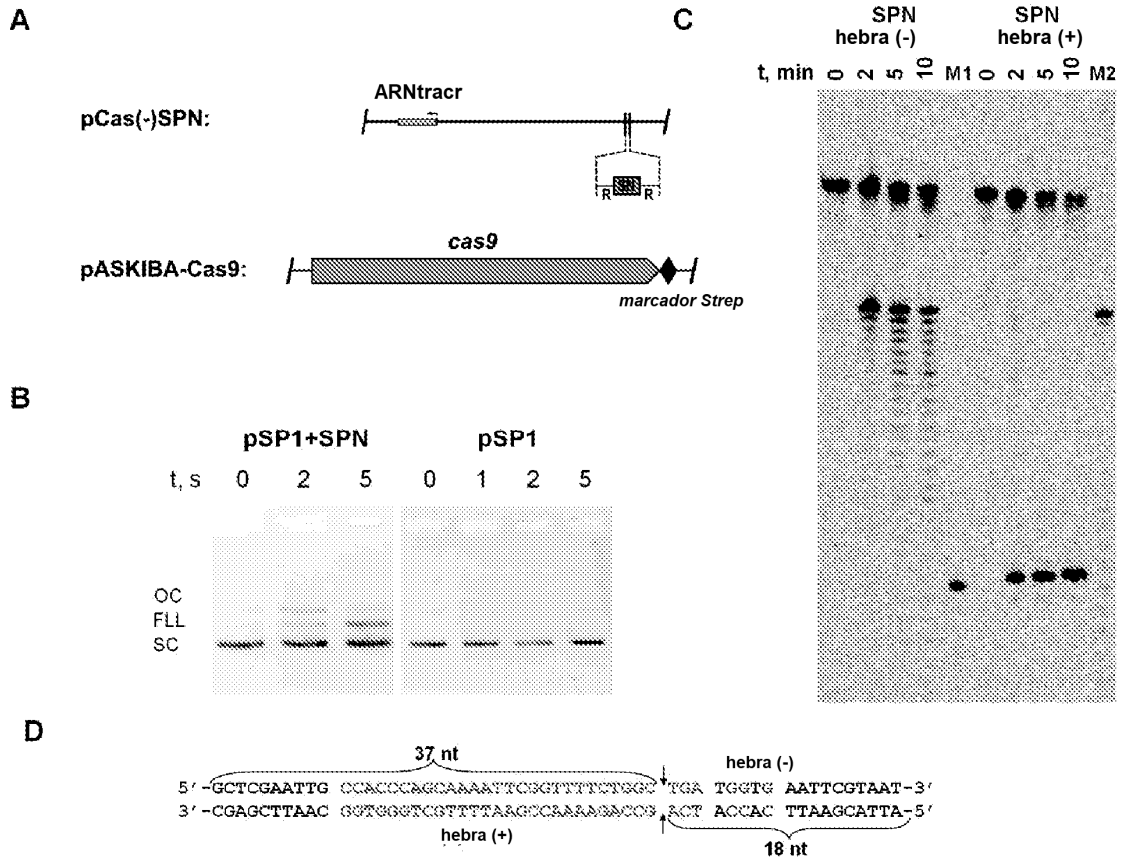


Figura 7.



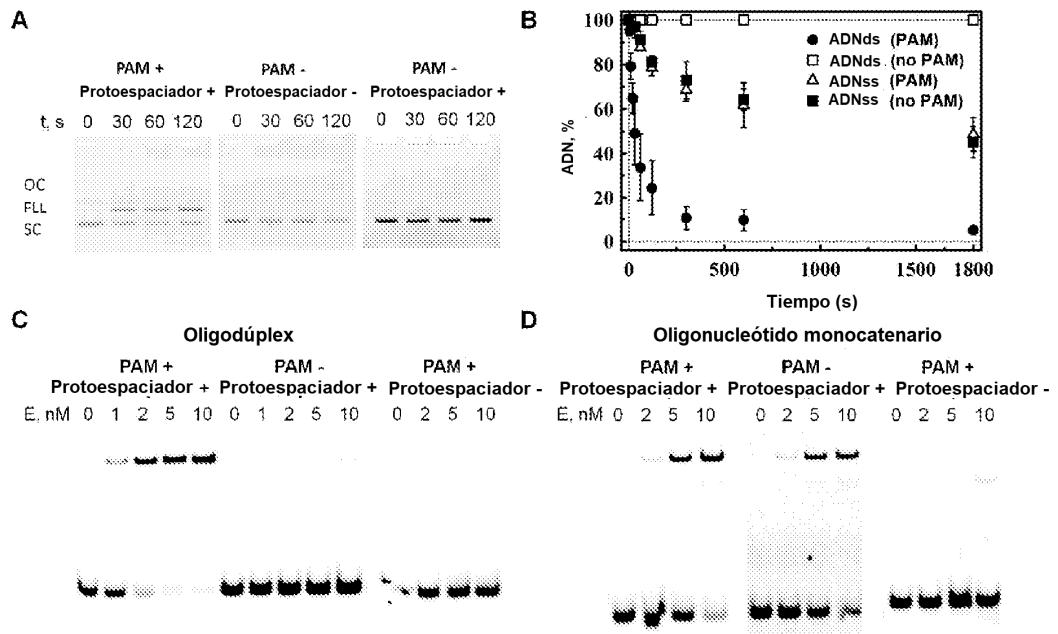


Figura 9.



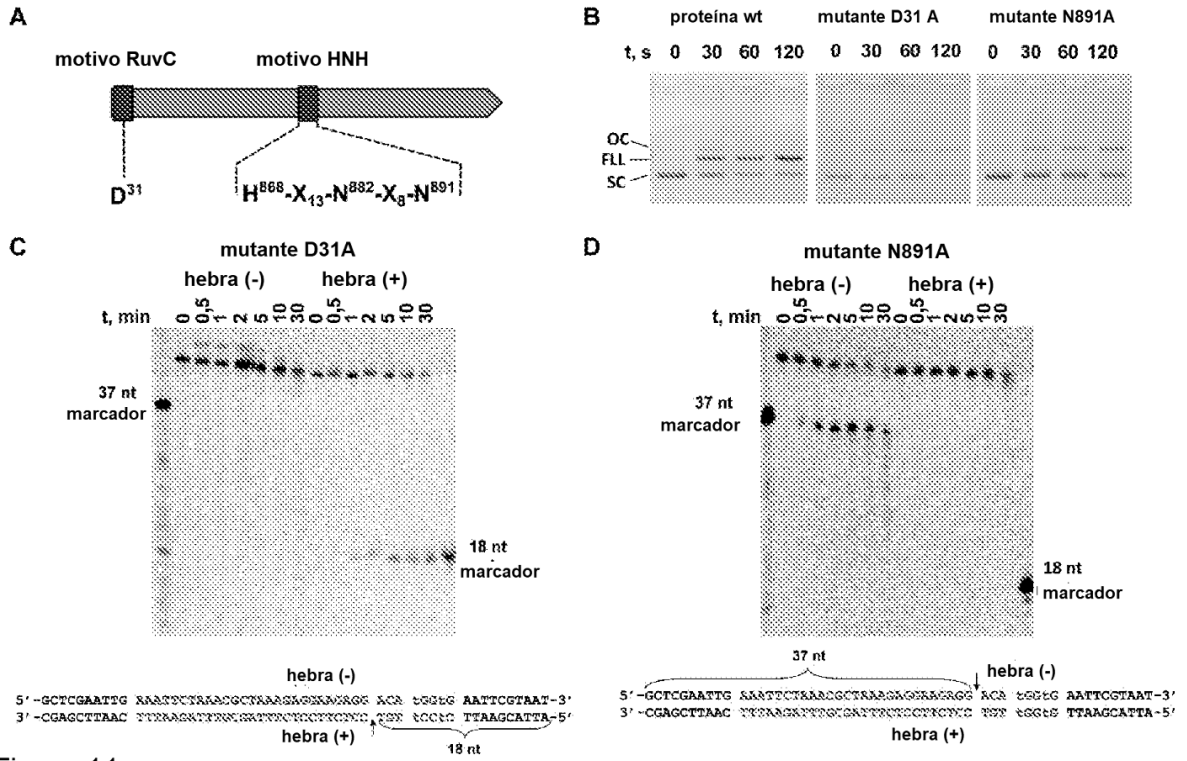


Figura 11.





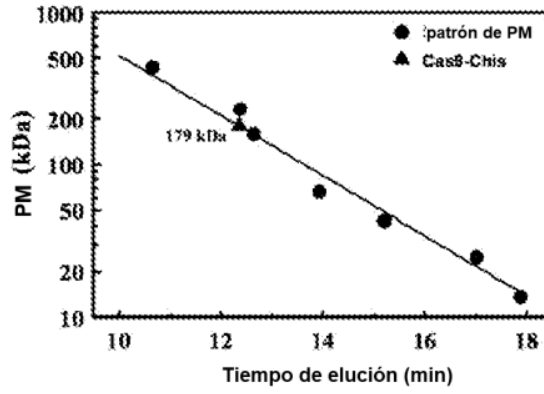


Figura 13.

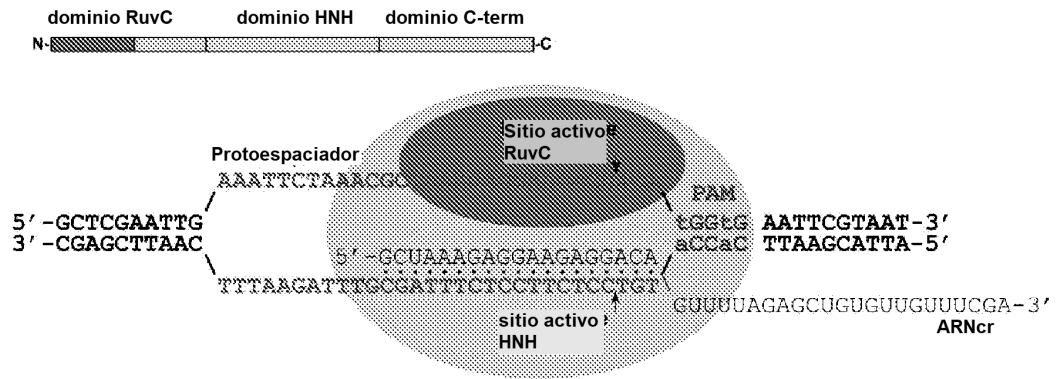


Figura 14.

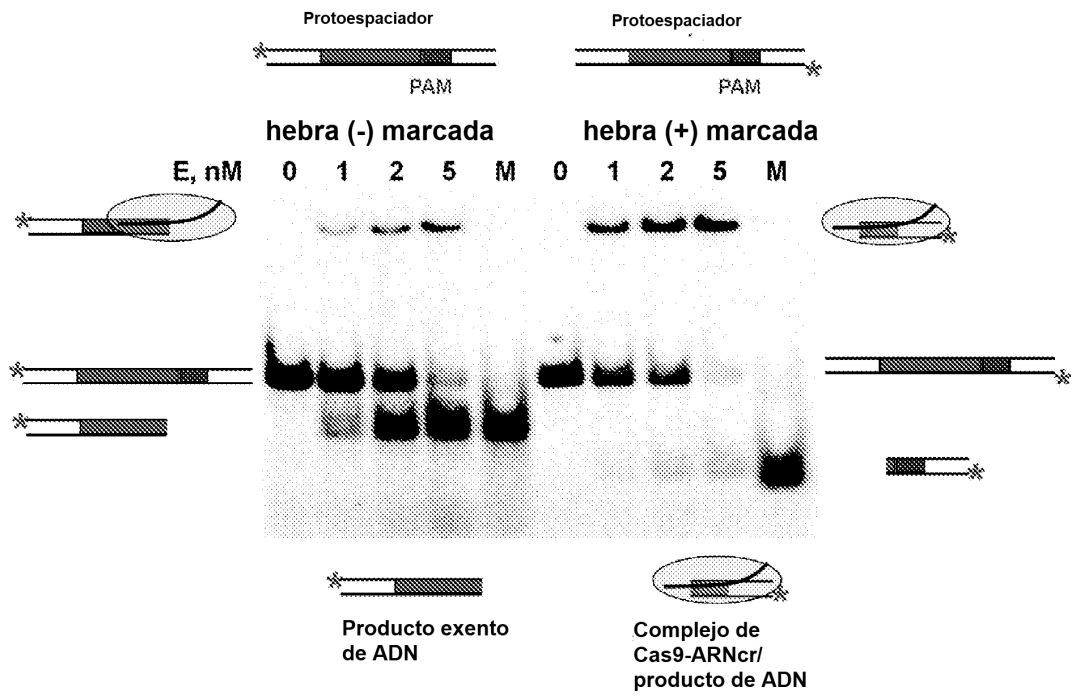


Figura 15.

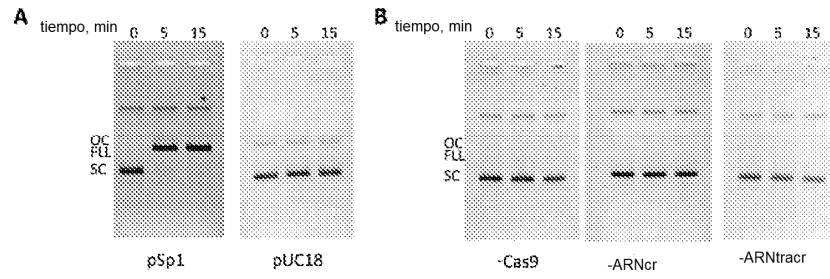


Figura 16.

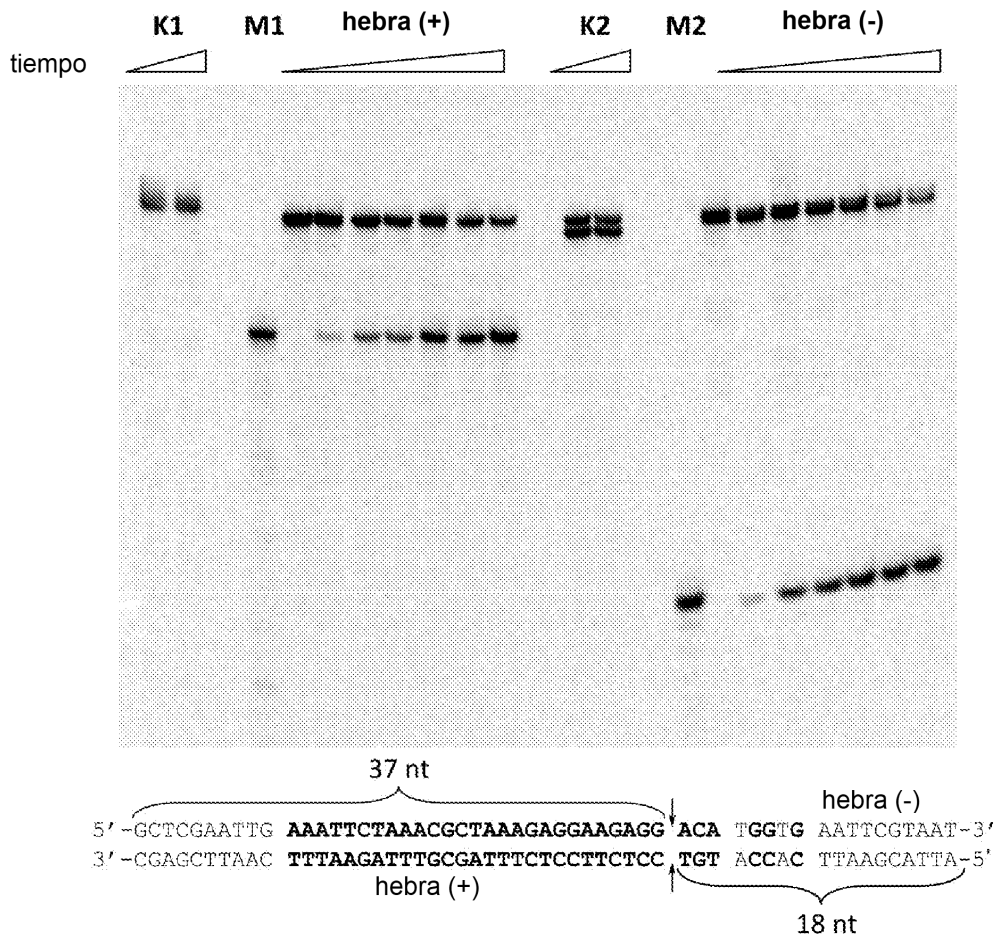


Figura 17.

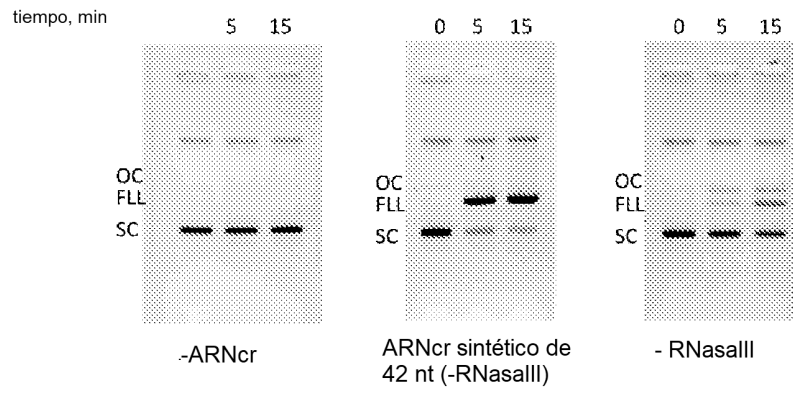


Figura 18.

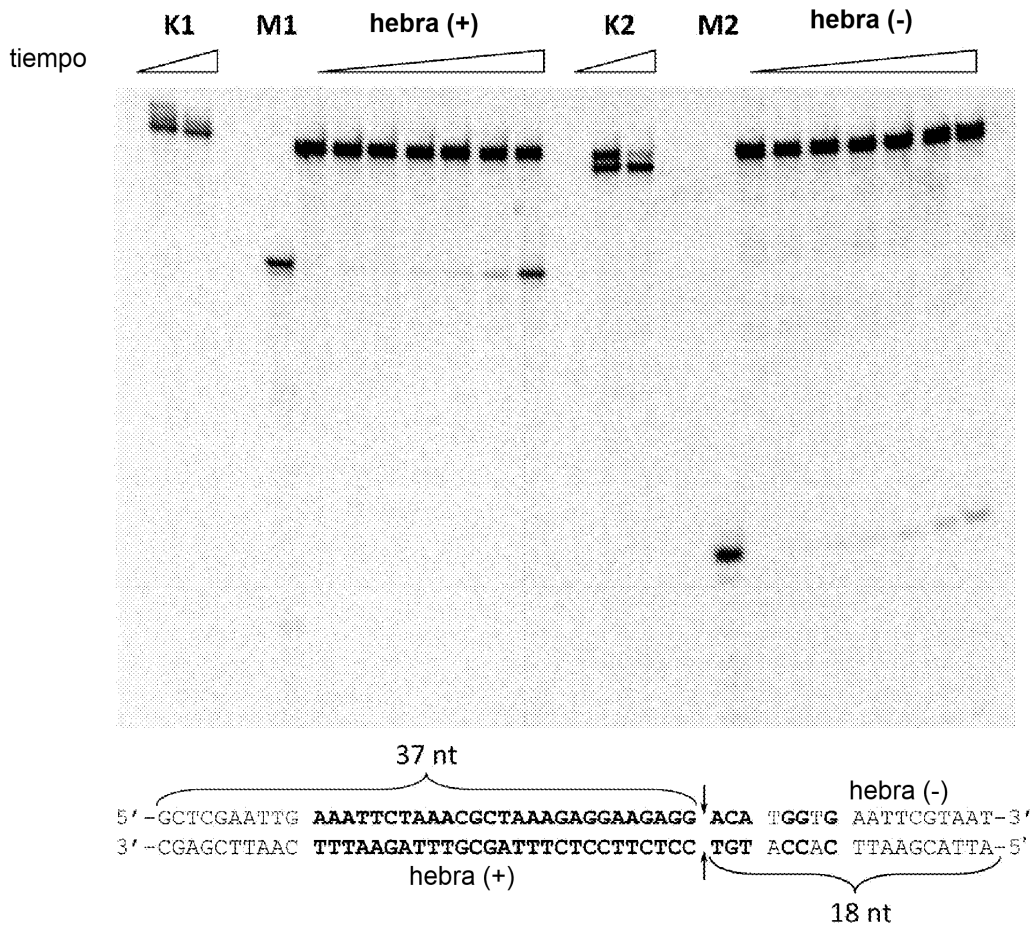


Figura 19.



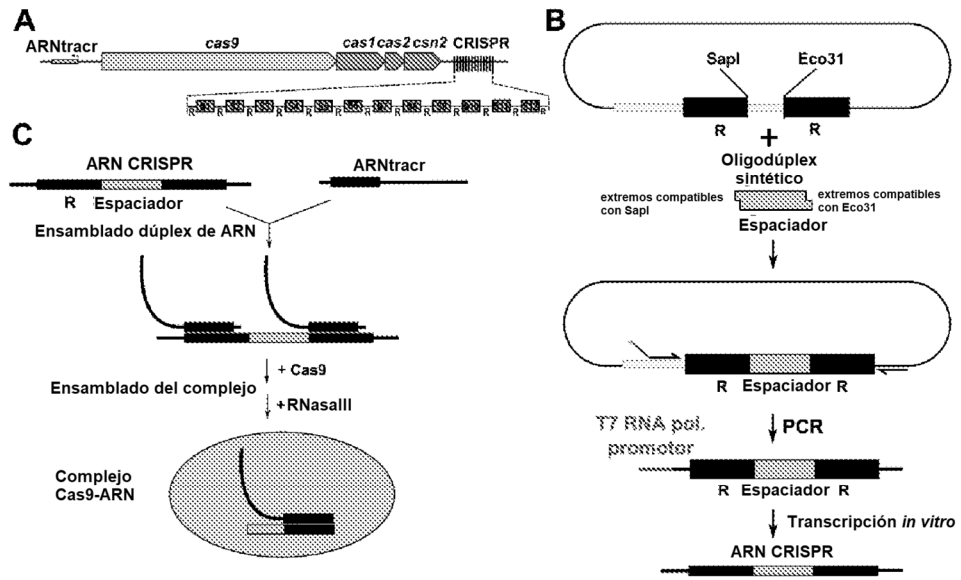


Figura 20.

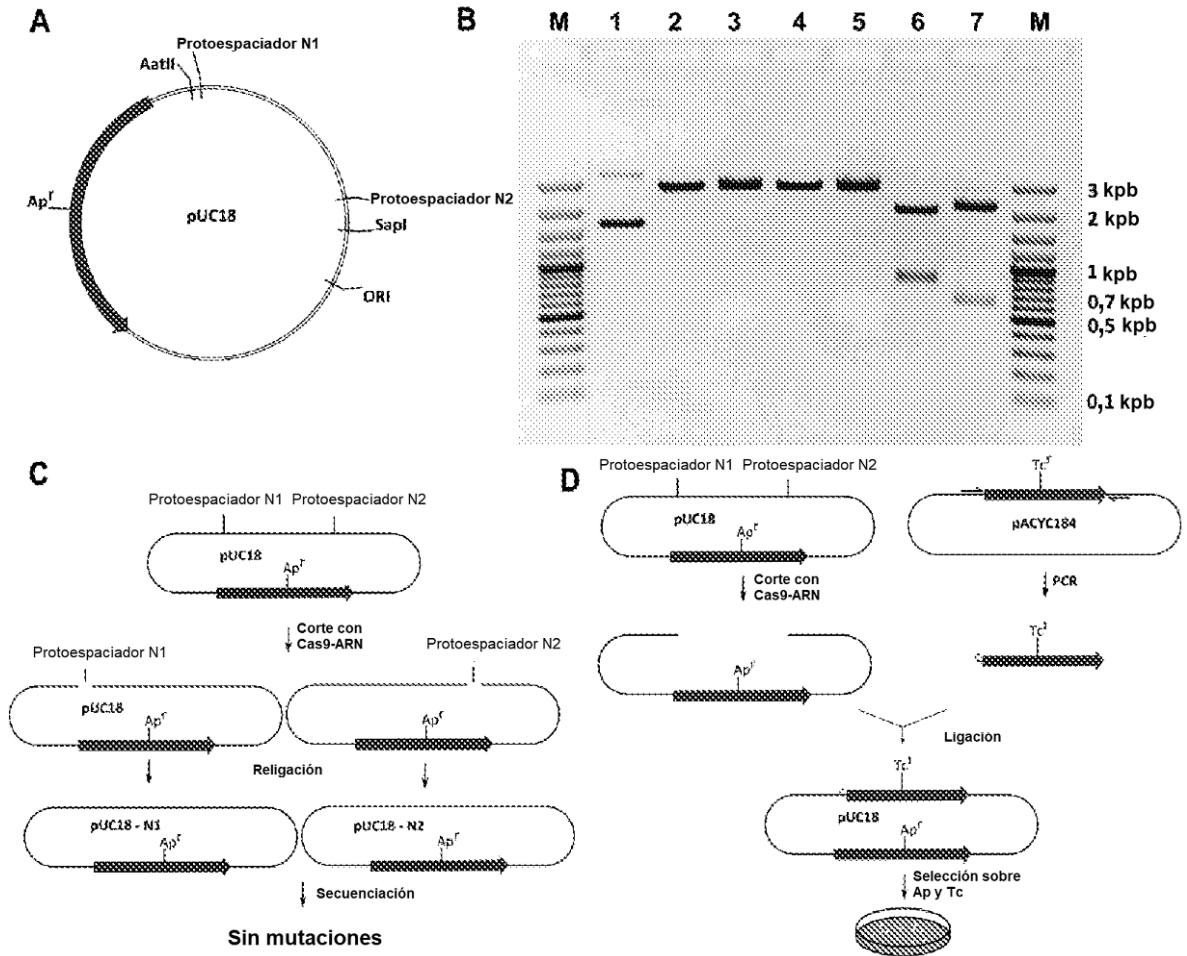


Figura 21.

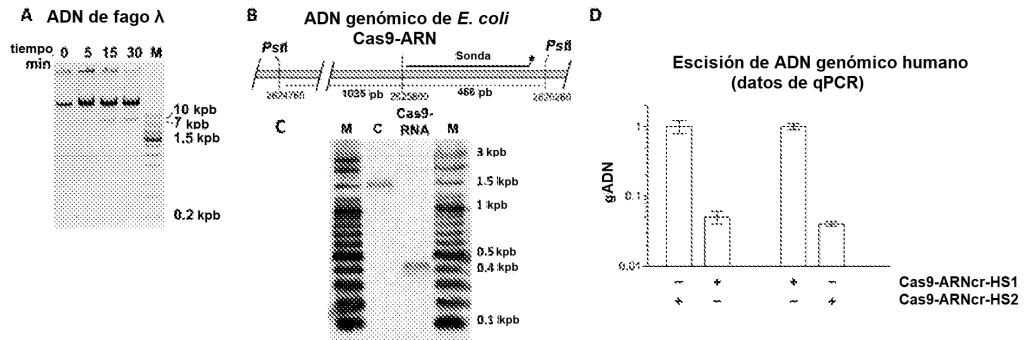


Figura 22.

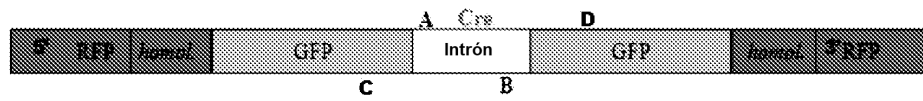


Figura 23.

Contransfección de pMTC-DSR+eGFP y Cas/ARNcr,  
CHO-k1

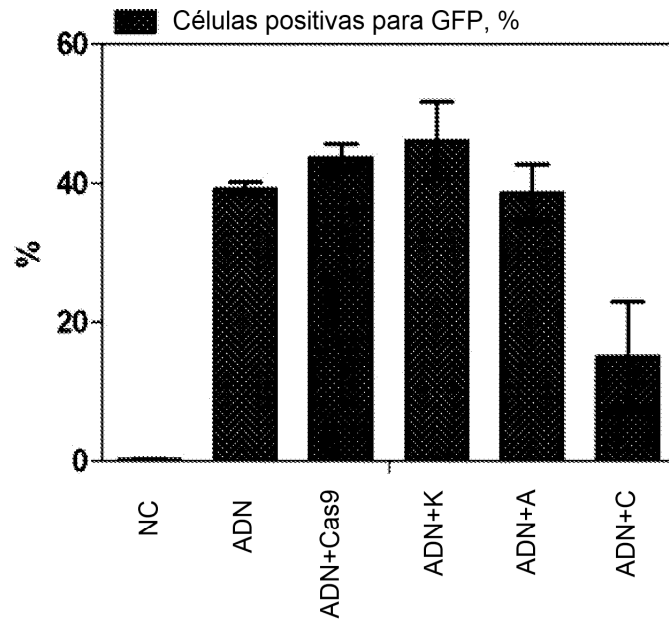


Figura 24.

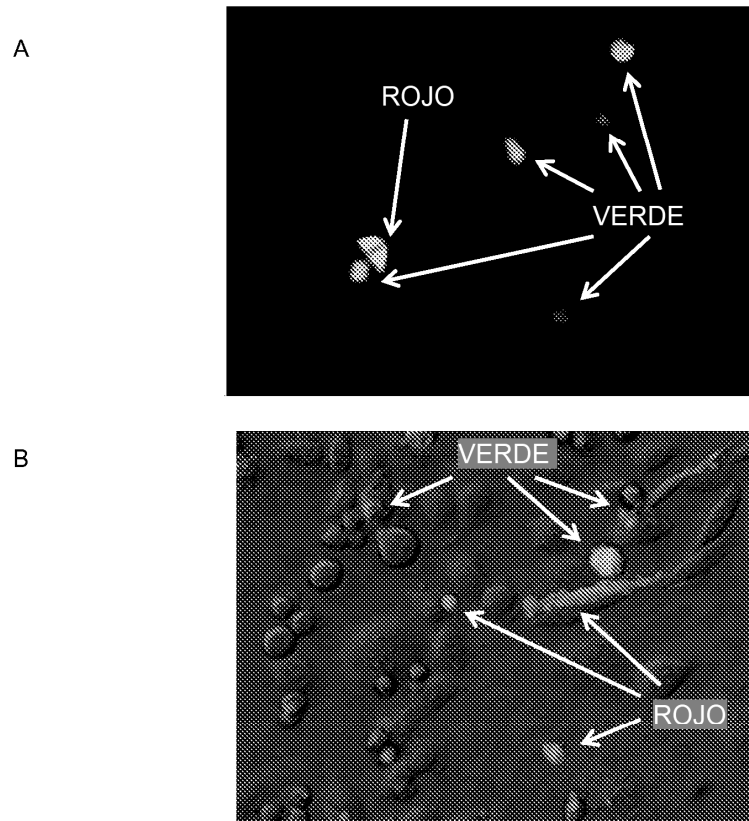


Figura 25.

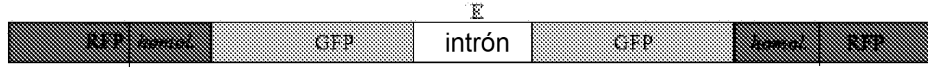


Figura 26.

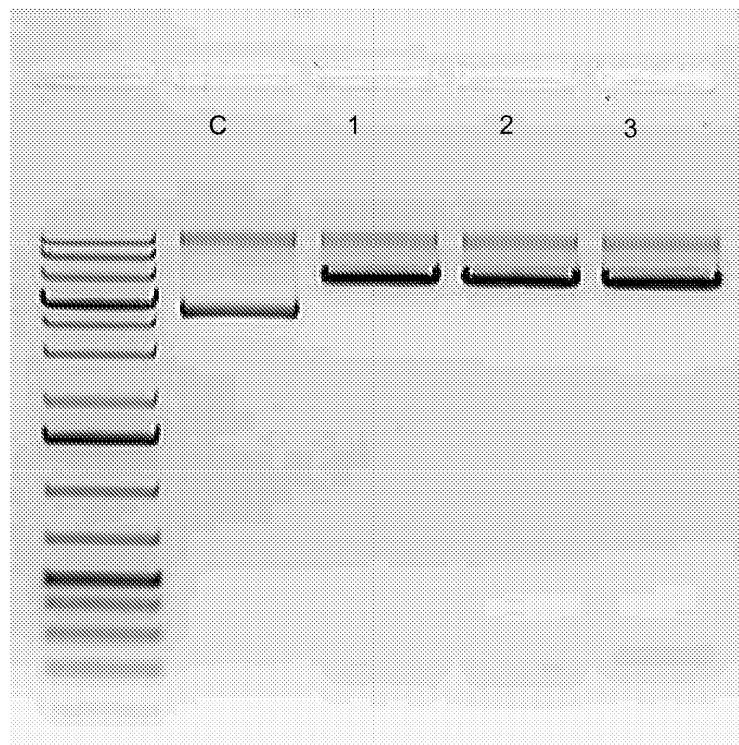


Figura 27.

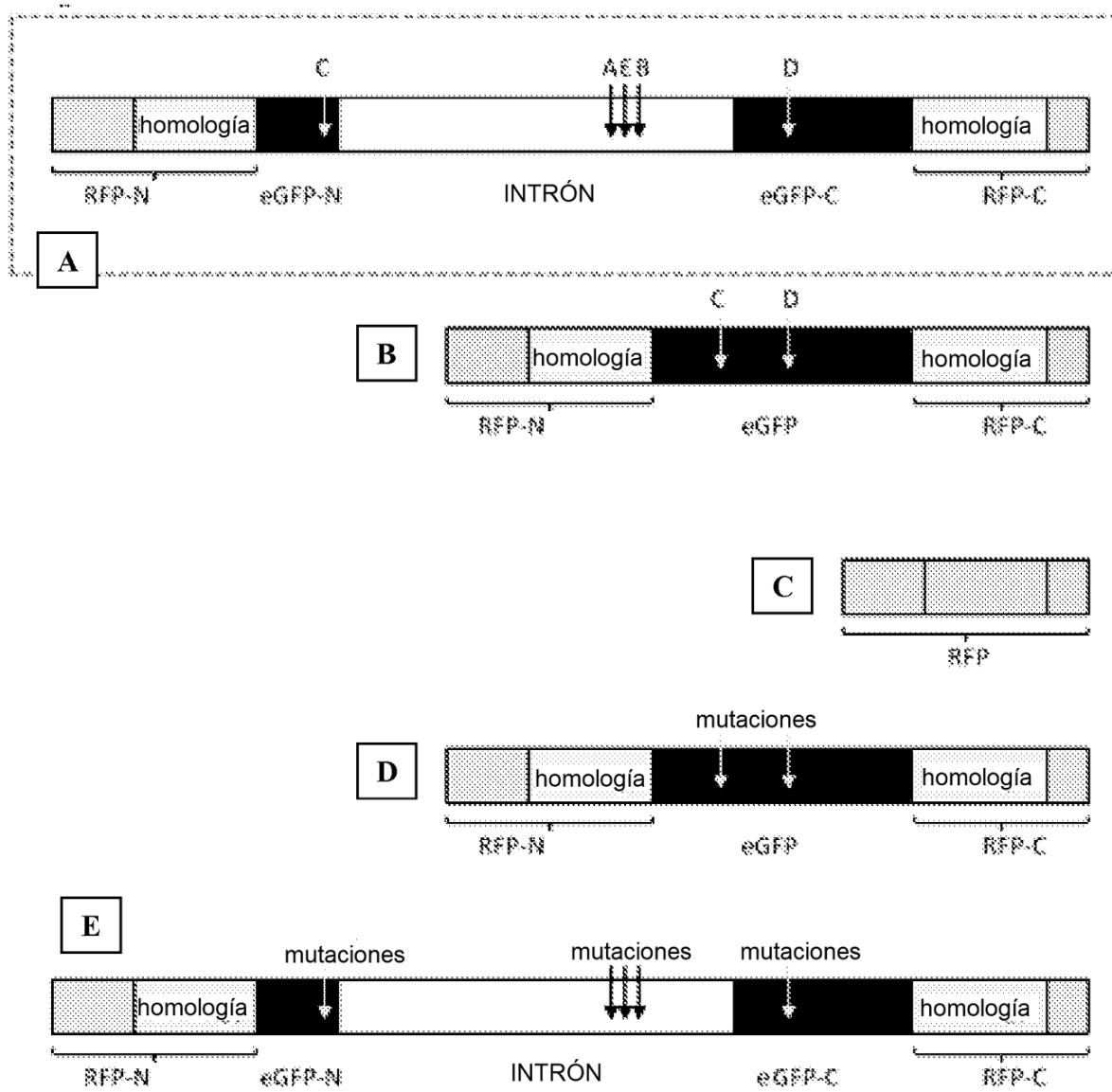


Figura 28.

Cotransfección de pMTC-DSR+eGFP y Cas9/ARNcr,  
CHO-k1

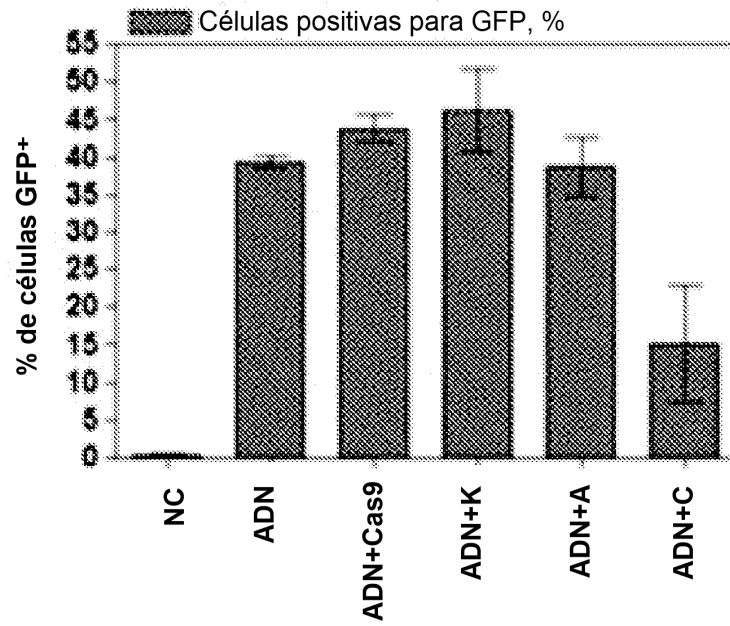


Figura 29.

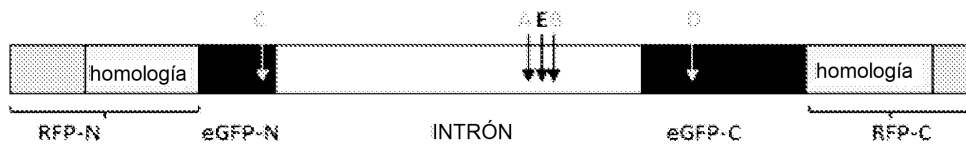


Figura 30.