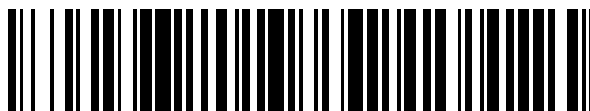


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 590 343**

51 Int. Cl.:

C12Q 1/68 (2006.01)

C12N 9/22 (2006.01)

C12N 15/55 (2006.01)

C12Q 1/34 (2006.01)

G01N 33/573 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.05.2011 PCT/US2011/035775**

87 Fecha y número de publicación internacional: **17.11.2011 WO11143124**

96 Fecha de presentación y número de la solicitud europea: **09.05.2011 E 11781086 (1)**

97 Fecha y número de publicación de la concesión europea: **06.07.2016 EP 2569425**

54 Título: **Composiciones de endorribonucleasas y métodos de uso de las mismas**

30 Prioridad:

12.11.2010 US 413287 P

19.07.2010 US 365627 P

10.05.2010 US 333163 P

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

21.11.2016

73 Titular/es:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (100.0%)

**1111 Franklin Street, 12th Floor
Oakland, CA 94607, US**

72 Inventor/es:

**HAURWITZ, RACHEL E.;
DOUDNA, JENNIFER A.;
WIEDENHEFT, BLAKE y
JINEK, MARTIN**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 590 343 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

Composiciones de endorribonucleasas y métodos de uso de las mismas

5 **Antecedentes**

Las enzimas de restricción de ADN transformaron la biología molecular en la década de 1970 haciendo posible escindir secuencias específicas de ADN a voluntad. La secuenciación de moléculas de ARN actualmente implica copiar el ARN en una hebra de ADN que después se secuencia por métodos convencionales. Este enfoque, también conocido como SecARN, es consistente y puede producir muchos millones de lecturas de secuencia. Sin embargo, la necesidad de generar ADNc introduce un sesgo intrínseco debido a las eficacias dependientes de secuencia de las etapas individuales.

15 Bibliografía

Carte et al. (2008) Genes Dev. 22:3489; publicación de Patente de Estados Unidos n.º 2010/0093026; documento EP 1 947 177; documento WO 2006/123537.

20 **Sumario de la invención**

La presente descripción proporciona endorribonucleasas Csy4 variantes, ácidos nucleicos que codifican las endorribonucleasas Csy4 variantes y células hospedadoras modificadas genéticamente con los ácidos nucleicos de acuerdo con las reivindicaciones. Las endorribonucleasas Csy4 variantes encuentran uso en diversas aplicaciones, que también se proporcionan. La presente descripción también proporciona métodos de detección de una secuencia específica en un polirribonucleótido diana; y métodos de regulación de la producción de un ARN diana en una célula eucariota.

Breve descripción de los dibujos

30 Las Figuras 1A-C representan el reconocimiento específico de un sustrato pre-ARNcr por Pa14Csy4. La secuencia de nucleótidos representada es 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1). Las Figuras 2A-C representan estructuras cristalinas de Csy4 unida al sustrato de ARN. Las Figuras 3A y 3B representan: una vista detallada del centro catalítico de Csy4 (Figura 3A); y la actividad de escisión de Csy4 de tipo silvestre (WT) y mutantes (Figura 3B).
35 La Figura 4 representa aminoácidos invariantes entre 12 secuencias de Csy4. Pa (SEQ ID NO: 8); Yp (SEQ ID NO: 34); Ec89 (SEQ ID NO: 39); Dn (SEQ ID NO: 79); Ab (SEQ ID NO: 84); MP1 (SEQ ID NO: 2); MP01 (SEQ ID NO: 3); SW (SEQ ID NO: 4); Pm (SEQ ID NO: 85); Pw (SEQ ID NO: 13); y Dd (SEQ ID NO: 10). Las Figuras 5A-5BD representan una alineación de secuencia de aminoácidos de diversos polipéptidos de Csy4, así como las secuencias de nucleótidos de secuencias de ARN reconocidas por cada polipéptido Csy4.
40 La Figura 6 representa ejemplos de secuencias de aminoácidos de endorribonucleasas específicas de secuencia, enzimáticamente inactivas. La Figura 7 representa un ejemplo de un método para detectar una secuencia específica en un polirribonucleótido diana. La Figura 8 representa el efecto de imidazol sobre la activación de diversas variantes de Csy4 enzimáticamente inactivas.
45 La Figura 9 representa un método a modo de ejemplo de aislamiento de un ARN diana. Se muestra un tallo-bucle (SEQ ID NO: 103) diana de Csy4. La Figura 10 representa un método a modo de ejemplo de regulación de la expresión de un ARN diana en una célula eucariota. Se muestra una secuencia de sustrato de ARN (SEQ ID NO: 103) de Csy4.

50 **Definiciones**

Como se usa en este documento, "polirribonucleótido" se refiere a una forma polimérica de ribonucleótidos, e incluye ARN, ARN que contiene uno o más desoxirribonucleótidos, y ADN que contiene uno o más ribonucleótidos. Un polirribonucleótido puede incluir, en algunos casos, uno o más nucleótidos modificados (por ejemplo, desoxiinosina, desoxiuridina o hidroximetildesoxiuridina). En algunos casos, un polirribonucleótido consiste en ribonucleótidos solamente (es decir, no incluye ningún desoxirribonucleótido). En algunos casos, un polirribonucleótido comprende ribonucleótidos, y uno o más ribonucleótidos modificados, pero no incluye ningún desoxirribonucleótido. En otros casos, un polirribonucleótido comprende ribonucleótidos, y puede comprender uno o más ribonucleótidos modificados, y uno o más desoxirribonucleótidos (incluyendo desoxirribonucleótidos modificados). En algunos casos, cuando un polirribonucleótido comprende uno o más desoxirribonucleótidos, los desoxirribonucleótidos comprenden de aproximadamente el 50 % a aproximadamente el 40 %, de aproximadamente el 40 % a aproximadamente el 30 %, de aproximadamente el 30 % a aproximadamente el 20 %, de aproximadamente el 20 % a aproximadamente el 10 %, de aproximadamente el 10 % a aproximadamente el 1 % o menos del 1 %, de los nucleótidos totales en el polirribonucleótido.

Las expresiones "ácido nucleico" y "polinucleótido" se usan de forma intercambiable y se refieren a una forma polimérica de nucleótidos de cualquier longitud, desoxirribonucleótidos o ribonucleótidos o análogos de los mismos. Ejemplos no limitantes de polinucleótidos incluyen ácidos nucleicos lineales y circulares, ARN mensajero (ARNm), ADNc, polinucleótidos recombinantes, vectores, sondas y cebadores.

5 Una "muestra biológica" abarca diversos tipos de muestras obtenidos de una célula, materia extracelular, un tejido o un organismo multicelular. La definición abarca sangre y otras muestras líquidas de origen biológico, muestras de tejido sólido tales como una muestra de biopsia o cultivos tisulares o células derivadas de los mismos y la descendencia de los mismos. La definición también incluye muestras que se han manipulado de cualquier modo
10 después de su obtención, tal como por tratamiento con reactivos, solubilización o enriquecimiento para ciertos componentes, tales como polinucleótidos. La expresión "muestra biológica" abarca una muestra clínica y también incluye células en cultivo, sobrenadantes celulares, lisados celulares, suero, plasma, fluido biológico (por ejemplo, fluido cefalorraquídeo, fluido de lavado bronquioalveolar, orina, sangre, una fracción sanguínea (por ejemplo, plasma; suero), esputo y similares), y muestras tisulares. En algunos casos, una muestra biológica comprende
15 células. En otros casos, una muestra biológica está libre de células.

La expresión "unido de forma funcional" se refiere a la unión funcional entre moléculas para proporcionar una función deseada. Por ejemplo, "unido de forma funcional" en el contexto de ácidos nucleicos se refiere a un enlace funcional entre ácidos nucleicos para proporcionar una función deseada tal como transcripción, traducción y similares, por
20 ejemplo, un enlace funcional entre un secuencia de control de la expresión de ácido nucleico (tal como un promotor, secuencia señal o serie de sitios de unión de factores de transcripción) y un segundo polinucleótido, donde la secuencia del control de la expresión afecta a la transcripción y/o traducción del segundo polinucleótido. "Unido de forma funcional" en el contexto de un polipéptido se refiere a un enlace funcional entre las secuencias de aminoácidos (por ejemplo, de diferentes dominios) para proporcionar una actividad deseada del polipéptido.

25 "Aislado" se refiere a una proteína o ácido nucleico que, si existe de forma natural, está en un entorno diferente de aquel en que puede existir de forma natural. Se entiende que "aislado" incluye proteínas o ácidos nucleicos que están dentro de muestras que están sustancialmente enriquecidas para la proteína o ácido nucleico de interés y/o en que la proteína o ácido nucleico de interés está parcial o sustancialmente purificado. Cuando la proteína o ácido nucleico no es de origen natural, "aislado" indica que la proteína o ácido nucleico se ha separado de un entorno en que se preparó por medios de síntesis o recombinantes.

"Sustancialmente puro" indica que una entidad (por ejemplo, polipéptido o un ácido nucleico) compone más de aproximadamente el 50 % del contenido total de la composición (por ejemplo, proteínas totales de la composición) y normalmente, más de aproximadamente el 60 % del contenido de proteínas totales. En algunas realizaciones, "sustancialmente puro" se refiere a composiciones en que al menos el 75 %, al menos el 85 %, al menos el 90 % o más de la composición total es la entidad de interés (por ejemplo, el 95 % de las proteínas totales). En algunas realizaciones, la proteína o ácido nucleico de interés compondrá más de aproximadamente el 90 %, más de aproximadamente el 95 %, más de aproximadamente el 98 % o más de aproximadamente el 99 % de las proteínas
40 totales o ácido nucleico de la composición.

Antes de describir adicionalmente la presente invención, debe entenderse que esta invención no está limitada a realizaciones particulares descritas, ya que estas pueden variar, por su puesto. También debe entenderse que la terminología usada en este documento es con el fin de describir realizaciones particulares solamente, y no pretende ser limitante, ya que el alcance de la presente invención se limitará solamente por las reivindicaciones adjuntas.

50 Cuando se proporciona un intervalo de valores, se entiende que cada valor intermedio, hasta la décima parte de la unidad del límite inferior salvo que el contexto indique claramente lo contrario, entre el límite superior e inferior de ese intervalo y cualquier otro valor indicado o intermedio en ese intervalo indicado, está abarcado dentro de la invención. Los límites superior e inferior de estos intervalos más pequeños pueden incluirse independientemente en los intervalos más pequeños, y también están abarcados dentro de la invención, sujetos a cualquier límite específicamente excluido en el intervalo indicado. Cuando el intervalo indicado incluye uno o ambos límites, los intervalos que excluyen cualquiera o ambos de esos límites incluidos también se incluyen en la invención.

55 Salvo que se defina de otro modo, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comprendido habitualmente por un experto en la materia a la cual pertenece esta invención. Aunque también puede usarse cualquier método y material similar o equivalente a los descritos en este documento en la práctica o ensayo de la presente invención, los métodos y materiales preferidos se describen ahora. Todas las publicaciones mencionadas en este documento desvelan y describen los métodos y/o materiales en relación con las publicaciones que se citan.

65 Debe apreciarse que como se usa en este documento y en las reivindicaciones adjuntas, las formas singulares "uno", "una", y "el", "la" incluyen referentes plurales salvo que el contexto indique claramente lo contrario. Por tanto, por ejemplo, referencia a "una endorribonucleasa específica de sitio" incluye una pluralidad de dichas endorribonucleasas específicas de sitio y referencia a "el polirribonucleótido diana" incluye referencia a uno o más polirribonucleótidos diana y equivalentes de los mismos conocidos para los expertos en la materia, y así en adelante.

Se aprecia adicionalmente que las reivindicaciones pueden redactarse para excluir cualquier elemento opcional. Por tanto, esta afirmación pretende servir como base antecedente para el uso de terminología exclusiva tal como "únicamente", "solamente" y similares en relación con la enumeración de los elementos de la reivindicación o el uso de una limitación "negativa".

5 Las publicaciones analizadas en este documento se proporcionan únicamente para su descripción antes de la fecha de presentación de la presente solicitud. Nada en este documento debe interpretarse como una admisión de que la presente invención no tiene derecho a antedatar dicha publicación en virtud de invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas reales de publicación que puede que se necesiten confirmar independientemente.

Descripción detallada

15 La presente descripción proporciona endorribonucleasas Csy4 variantes, ácidos nucleicos que codifican las endorribonucleasas Csy4 variantes y células hospedadoras modificadas genéticamente con los ácidos nucleicos. Las endorribonucleasas Csy4 variantes encuentran uso en diversas aplicaciones, que también se proporcionan. La presente descripción también proporciona métodos de detección de una secuencia específica en un polirribonucleótido diana; y métodos de regulación de la producción de un ARN diana en una célula eucariota.

20 Métodos de detección de una secuencia en un polirribonucleótido diana

La presente descripción proporciona un método de detección de una secuencia en un polirribonucleótido diana. Los métodos son útiles para detectar la presencia de una secuencia particular en un polirribonucleótido, y pueden usarse, por lo tanto, para detectar un polirribonucleótido que comprenda una secuencia particular. Por ejemplo, el método puede usarse para detectar la presencia de un polirribonucleótido de un patógeno en una muestra (por ejemplo, en una muestra biológica).

Un presente método puede detectar tan poco como 100 copias, hasta una única copia, de un polirribonucleótido diana. Por tanto, por ejemplo, un presente método puede detectar de 1 a aproximadamente 5, de aproximadamente 5 a aproximadamente 10, de aproximadamente 10 a aproximadamente 50 o de aproximadamente 50 a aproximadamente 100 o más de 100 copias de un polirribonucleótido diana en una muestra (por ejemplo, en una única célula, en un único embrión u otra muestra biológica). Un presente método, por tanto, es útil para diversas aplicaciones forenses, de investigación y diagnóstico.

35 En algunas realizaciones, un presente método de detección de una secuencia específica en un polirribonucleótido diana comprende: a) poner en contacto el polirribonucleótido diana con una sonda oligonucleotídica que comprende la secuencia específica y una endorribonucleasa Csy4 específica de secuencia enzimáticamente activa en condiciones que favorecen la formación de dúplex entre la sonda oligonucleotídica y el polirribonucleótido diana, donde el dúplex se escinde por la endorribonucleasa Csy4; y b) detectar la unión específica entre la sonda oligonucleotídica y el polirribonucleótido diana, donde la detección de la formación de dúplex entre la sonda oligonucleotídica y el polirribonucleótido diana indica la presencia de la secuencia específica en el polirribonucleótido diana.

45 En algunos casos, la sonda oligonucleotídica se une a un péptido, y el péptido se libera tras la escisión del dúplex por la endorribonucleasa Csy4; en estos casos, la etapa de detección implica la detección del péptido liberado. Por ejemplo, el péptido liberado se detecta por unión a un anticuerpo específico para el péptido, por ejemplo, donde el anticuerpo está inmovilizado. En algunas realizaciones, el polirribonucleótido diana se inmoviliza en un soporte sólido. Los polirribonucleótidos diana incluyen cualquiera de diversos polinucleótidos, por ejemplo, el polirribonucleótido diana puede ser un polirribonucleótido de un patógeno.

50 Como se ha indicado anteriormente, en algunas realizaciones, el anticuerpo o el polinucleótido diana se inmoviliza en un soporte sólido (soporte insoluble). Los soportes insolubles adecuados incluyen, aunque sin limitación, perlas de agarosa, perlas magnéticas, una tira de ensayo, una placa multipocillo y similares. El soporte insoluble puede comprender diversas sustancias (vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poliacrilamidas, agarosas y magnetita) y puede proporcionarse en diversas formas, incluyendo, por ejemplo, perlas de agarosa, perlas de poliestireno, perlas de látex, perlas magnéticas, partículas de metal coloide, chips y superficies de vidrio y/o de silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de placas de reacción (por ejemplo, placas multipocillo), tubos de plástico, etc.

60 En algunas realizaciones, el método generalmente implica: a) poner en contacto un polirribonucleótido diana con una endorribonucleasa específica de secuencia; y b) detectar fragmentos de escisión producidos por escisión específica de sitio del polirribonucleótido diana, donde la producción de fragmentos de escisión esperados tras la escisión en una secuencia específica en el polirribonucleótido indica la presencia de la secuencia específica.

65

En otras realizaciones, un presente método de detección de una secuencia en un polirribonucleótido diana implica: a) poner en contacto un polirribonucleótido diana con: i) una endorribonucleasa específica de secuencia; y ii) una sonda oligonucleotídica que comprende un resto de detección unido, donde la sonda oligonucleotídica comprende una secuencia específica de nucleótidos conocida; donde la sonda oligonucleotídica forma un dúplex con una secuencia complementaria en el polirribonucleótido diana basado en la unión de la secuencia de nucleótidos conocida presente en la sonda oligonucleotídica con una secuencia complementaria en el polirribonucleótido diana, y donde la endorribonucleasa específica de secuencia escinde el dúplex de un modo específico de secuencia, liberando de ese modo el resto de detección de la sonda oligonucleotídica; y b) detectar el resto de detección liberado, donde la liberación del resto de detección indica la presencia de la secuencia específica. En algunas realizaciones, se usan dos o más sondas oligonucleotídicas diferentes, que comprenden cada una una secuencia específica diferente de nucleótidos conocida.

En algunas realizaciones, el resto de detección es un polipéptido. El polipéptido puede detectarse usando un ensayo inmunológico (por ejemplo, un ensayo inmunoabsorbente ligado a enzimas (ELISA); un radioinmunoensayo (RIA); etc.), usando un anticuerpo específico para el resto de detección polipeptídico. El anticuerpo específico para el resto de detección polipeptídico puede comprender un marcador detectable. El ensayo inmunológico puede realizarse sobre una tira de ensayo (por ejemplo, en un ensayo de flujo lateral) u otro medio adecuado tal como placa multipocillo.

En algunas realizaciones, el resto de detección es una proteína fluorescente, donde las proteínas fluorescentes adecuadas son como se describe en este documento. En otras realizaciones, el resto de detección es luciferina u otro sustrato para luciferasa. Las luciferinas adecuadas u otros sustratos de luciferasa incluyen, por ejemplo, luciferina (por ejemplo, una luciferina de luciérnaga); una aminoluciferina; coelenterazina; una coelenterazina modificada como se describe en la patente de Estados Unidos n.º 7.537.912; un análogo de coelenterazina como se describe en la publicación de patente de Estados Unidos n.º 2009/0081129 (por ejemplo, un análogo de coelenterazina permanente de membrana como se describe en la publicación de patente de Estados Unidos n.º 2009/0081129, por ejemplo, una de las Estructuras II, III, IV, V, y VI de la publicación de patente de Estados Unidos n.º 2009/0081129); aminoluciferina; dihidroluciferina; metiléter de luciferina 6'; o cloroetiléter de luciferina 6'. Véase, por ejemplo, Branchini, B.R. et al. Anal. Biochem. 2010, 396, 290-296; y Mezzanotte, L. et al., In vivo bioluminescence imaging of murine xenograft cancer models with a red-shifted thermostable luciferase. Mol. Imaging Biol. (9 de noviembre de 2009, online; PubMed ID: 19937390).

Un ejemplo no limitante de un presente método de detección se ilustra esquemáticamente en la **Figura 7**. En el ejemplo representado en la Figura 7, se ponen en contacto pequeños oligonucleótidos que se unen a regiones concretas de un polinucleótido diana (por ejemplo, un ARN vírico) con el polinucleótido diana, donde los oligonucleótidos comprenden restos detectables (por ejemplo, ligandos; péptidos; etc.). Se añade una endonucleasa de restricción específica de secuencia, enzimáticamente activa (RRE) que aborda el dúplex de oligonucleótido/ARN vírico. La enzima escinde el dúplex de oligonucleótido/ARN vírico; y se liberan ligandos para la detección. La enzima escinde dúplex adicionales, amplificando de ese modo la señal. Los ligandos liberados se detectan usando un ensayo de flujo lateral (por ejemplo, tira de ensayo) o un ensayo de base inmunológica (por ejemplo, ELISA).

Una endorribonucleasa específica de secuencia adecuada es una endorribonucleasa específica de secuencia, enzimáticamente activa. Las endorribonucleasas que son adecuadas para su uso en un presente método de detección incluyen endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que tienen al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la **Figura 4** (secuencias de aminoácidos de Csy4).

Las endorribonucleasas que son adecuadas para su uso en un presente método de detección incluyen endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que tienen al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la **Figura 5** (secuencias de aminoácidos de Csy4). La **Figura** proporciona secuencias unidas específicamente por las diversas endorribonucleasas. En algunos casos, una endorribonucleasa Csy4 específica de secuencia enzimáticamente activa adecuada puede comprender una secuencia de aminoácidos de una secuencia de aminoácidos de Csy4 representada en la **Figura 5**.

Las endorribonucleasas que son adecuadas para su uso en un presente método de detección incluyen endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que difieren de una secuencia de aminoácidos expuesta en una cualquiera de las **Figuras 4 o 5** en 1 a 20 (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20) sustituciones y/o inserciones y/o deleciones de aminoácidos.

El polirribonucleótido diana a detectarse puede estar presente en una muestra, por ejemplo, una muestra biológica tal como sangre, un producto sanguíneo (por ejemplo, plasma), orina, fluido cefalorraquídeo, fluido de lavado broncoalveolar, saliva, un tejido, células, etc. El polirribonucleótido diana puede estar aislado o purificado. El polirribonucleótido diana puede ser un ARN mensajero (ARNm), un ARN vírico, ARN bacteriano, ARN de parásito u otras especies de ARN. Los ARN víricos incluyen, aunque sin limitación, cualquier miembro de *Flaviviridae*, por ejemplo, virus de la hepatitis C, virus del dengue, virus de la fiebre amarilla, virus del Nilo occidental, etc.; cualquier miembro de *Retroviridae*; un virus de inmunodeficiencia (por ejemplo, virus de la inmunodeficiencia humana); etc.

El polirribonucleótido diana a detectarse puede estar presente en una célula de un organismo multicelular (o puede obtenerse de una célula de un organismo multicelular).

El polirribonucleótido diana a detectarse puede estar presente en u obtenerse de una célula u organismo de cualquiera de los seis reinos, por ejemplo, Bacteria (por ejemplo, Eubacteria); Archaeobacteria; Protista; Fungi; Plantae; y Animalia. Las fuentes adecuadas de polirribonucleótidos diana incluyen miembros de tipo vegetal del reino Protista, incluyendo, aunque sin limitación, algas (por ejemplo, algas verdes, algas rojas, glaucosifitas, cianobacterias); miembros tipo hongos de Protista, por ejemplo, mohos de cieno, mohos acuáticos, etc.; miembros tipo animal de Protista, por ejemplo, flagelados (por ejemplo, Euglena), ameboides (por ejemplo, amebas), esporozoos (por ejemplo, Apicomplexa, Myxozoa, Microsporidia), y ciliados (por ejemplo, Paramecium). Fuentes adecuadas de polirribonucleótidos diana incluyen miembros del reino Fungi, incluyendo, aunque sin limitación, miembros de cualquiera de los filos: Basidiomycota (hongos asociados; por ejemplo, miembros de *Agaricus*, *Amanita*, *Boletus*, *Cantherellus*, etc.); Ascomycota (hongos de saco, incluyendo, por ejemplo, *Saccharomyces*); Mycophycophyta (líquenes); Zygomycota (hongos de conjugación); y Deuteromycota. Fuentes adecuadas de polirribonucleótidos diana incluyen miembros del reino Plantae, incluyendo, aunque sin limitación, miembros de cualquiera de las siguientes divisiones: Bryophyta (por ejemplo, musgos), Anthocerotophyta (por ejemplo, briofitas), Hepaticophyta (por ejemplo, hepáticas), Lycophyta (por ejemplo, musgos asociados), Sphenophyta (por ejemplo, colas de caballo), Psilophyta (por ejemplo, helechos en escobilla), Ophioglossophyta, Pterophyta (por ejemplo, helechos), Cycadophyta, Ginkgophyta, Pinophyta, Gnetophyta, y Magnoliophyta (por ejemplo, plantas florales). Fuentes adecuadas de polirribonucleótidos diana incluyen miembros del reino Animalia, incluyendo, aunque sin limitación, miembros de cualquiera de los filos: Porifera (esponjas); Placozoa; Orthonectida (parásitos de invertebrados marinos); Rhombozoa; Cnidaria (corales, anémonas, medusas, plumas de mar, pensamientos de mar, avispas marinas); Ctenophora (ctenóforos); Platyhelminthes (platelmintos); Nemertina (gusanos planos); Ngathostomulida (gusanos con mandíbula); Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cycliophora; Mollusca (moluscos); Sipuncula (gusanos del cacahuete); Annelida (gusanos segmentados); Tardigrada (osos de agua); Onychophora (gusanos aterciopelados); Arthropoda (incluyendo los subfilos: Chelicerata, Myriapoda, Hexapoda, y Crustacea, donde el Chelicerata incluye, por ejemplo, arácnidos, Merostomata, y Pycnogonida, donde el Myriapoda incluye, por ejemplo, Chilopoda (ciempiés), Diplopoda (milpiés), Paropoda, y Symphyla, donde el Hexapoda incluye insectos, y donde el Crustacea incluye camarones, krill, percebes, etc.; Phoronida; Ectoprocta (animales del musgo); Brachiopoda; Echinodermata (por ejemplo estrella de mar, margaritas de mar, estrellas pluma, erizos de mar, pepinos de mar, estrellas frágiles, cestas frágiles, etc.); Chaetognatha (gusanos flecha); Hemichordata (gusanos bellota); y Chordata. Miembros adecuados de Chordata incluyen cualquier miembro de los siguientes subfilos: Urochordata (ascidias; incluyendo Ascidiacea, Thaliacea, y Larvacea); Cephalochordata (lancelados); Myxini (mixinas); y Vertebrata, donde miembros de Vertebrata incluyen, por ejemplo, miembros de Petromyzontida (lampreas), Chondrichthyes (peces cartilaginosos), Actinopterygii (curimbas), Actinista (celocantos), Dipnoi (pez pulmonado), Reptilia (reptiles, por ejemplo, serpientes, caimanes, cocodrilos, lagartos, etc.), Aves (pájaros); y Mammalian (mamíferos). Plantas adecuadas incluyen cualquier monocotiledónea y cualquier dicotiledónea.

Por tanto, por ejemplo, un polirribonucleótido diana puede estar presente en u obtenerse de células de organismos que incluyen, aunque sin limitación, un protozoo, una planta, un hongo, una célula de alga, una levadura, un reptil, un anfibio, un mamífero, un microorganismo marino, un invertebrado marino, un artrópodo, un isópodo, un insecto, un arácnido, una arqueobacteria y una eubacteria.

Un polirribonucleótido diana puede estar presente en u obtenerse de un embrión no humano, por ejemplo, un embrión de *Drosophila*; un embrión de pez cebra; un embrión de ratón; etc.

Un polirribonucleótido diana puede estar presente en u obtenerse de una célula madre, por ejemplo, una célula madre *in vitro*; una célula madre no humana; etc. Células madre adecuadas incluyen células madre embrionarias, células madre adultas y células madre pluripotentes inducidas (iPS).

En algunas realizaciones, el polirribonucleótido diana se aislará de un tejido recogido de un organismo; de una célula particular o grupo de células aisladas de un organismo; etc. Por ejemplo, cuando el organismo es una planta, el polirribonucleótido diana se aislará, en algunas realizaciones, del xilema, el floema, la capa del cambium, hojas, raíces, etc. Cuando el organismo es un animal, el polirribonucleótido diana se aislará, en algunas realizaciones, de un tejido particular (por ejemplo, pulmón, hígado, corazón, riñón; cerebro, bazo, piel, tejido fetal, etc.) o un tipo celular particular (por ejemplo, células neuronales, células epiteliales, células endoteliales, astrocitos, macrófagos, células de la glia, células insulares, linfocitos T, linfocitos B, etc.).

Métodos de regulación de la producción de un ARN diana

La presente descripción proporciona un método de regulación de la protección de un ARN diana en una célula. El método implica generalmente poner en contacto una célula hospedadora genéticamente modificada con un agente que activa un promotor inducible, donde la célula hospedadora modificada genéticamente se modifica genéticamente con un vector de expresión recombinante que comprende una secuencia de nucleótidos que codifica una enzima que cataliza la escisión en un sitio de escisión específico de secuencia en un polirribonucleótido sustrato, donde la secuencia de nucleótidos codificantes de la enzima está unida de forma funcional al promotor inducible y donde, tras la activación del promotor inducible, la enzima se produce en la célula y escinde dicho ARN diana de un ARN precursor.

La **Figura 10** proporciona una representación esquemática de un método a modo de ejemplo de la regulación de la producción de un ARN diana. En la Figura 10, se modifica un ARN diana endógeno para incluir un sustrato de ARN de Csy4 (por ejemplo, GUUCACUGCCGUUAUAGGCAG (SEQ ID NO: 103); o SEQ ID NO: 1) en la región no traducida 3' (3' UTR). La expresión de Csy4 en la célula hospedadora conduce a la unión y escisión del sustrato de ARN. El ARN escindido ahora carece de su cola poliA y se degradará.

Por ejemplo, en algunas realizaciones, la presente descripción proporciona un método de regulación de la producción de un ARN diana en una célula eucariota, donde el método implica poner en contacto una célula hospedadora modificada genéticamente con un agente que activa un promotor inducible, donde la célula hospedadora modificada genéticamente se modifica genéticamente con un vector de expresión recombinante que comprende una secuencia de nucleótidos que codifica una endorribonucleasa Csy4 específica de secuencia enzimáticamente activa que cataliza la escisión en un sitio de escisión específico de secuencia en un polirribonucleótido sustrato, donde la secuencia de nucleótidos que codifica la enzima está unida de forma funcional al promotor inducible y donde, tras la activación del promotor inducible, la enzima se produce en la célula y escinde dicho ARN diana de un ARN precursor. En algunos casos, la especie de ARN diana es un ARN regulador. En algunos casos, la escisión de dicho ARN diana desde un ARN precursor inactiva el ARN precursor.

Un endorribonucleasa específica de secuencia adecuada es una endorribonucleasa específica de secuencia enzimáticamente activa. Las endorribonucleasas que son adecuadas para su uso en un presente método de regulación de la producción de un ARN diana incluyen endorribonucleasas que se unen y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que tienen el menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la Figura 4 (secuencias de aminoácidos de Csy4).

Las endorribonucleasas que son adecuadas para su uso en un presente método de regulación de la producción de un ARN diana incluyen endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que tiene al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la **Figura 5** (secuencias de aminoácidos de Csy4). La Figura 5 proporciona secuencias unidas específicamente por las diversas endorribonucleasas.

Las endorribonucleasas que son adecuadas para su uso en un presente método de regulación de la producción de un ARN diana incluyen endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyendo polipéptidos enzimáticamente activos y que difieren de una secuencia de aminoácidos expuesta en una cualquiera de las **Figuras 4 o 5** en 1 a 20 (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20) sustituciones y/o inserciones y/o deleciones de aminoácidos.

Un promotor inducible adecuado puede incluir un promotor que es funcional en una célula eucariota. Los promotores inducibles adecuados son conocidos en la técnica. Por ejemplo, los promotores inducibles adecuados incluyen, aunque sin limitación, un promotor GAL1, un promotor GAL10, un promotor ADH2, un promotor PHO5, un promotor CUP1, un promotor GAL7, un promotor MET25, un promotor MET3, un promotor CYC1, un promotor HIS 3, un promotor ADH1, un promotor PGK, un promotor GAPDH, un promotor ADC1, un promotor TRP1, un promotor URA3, un promotor LEU2, un promotor ENO, un promotor TP1, y AOX1. Los promotores inducibles adecuados incluyen promotores inducibles por tetraciclina; un promotor de metalotioneína; promotores inducibles por tetraciclina, promotores inducibles por metionina; y promotores inducibles por galactosa, que son promotores bien conocidos en la técnica. Otros promotores adecuados incluyen el promotor de la alcohol deshidrogenasa ADH2 (reprimido en glucosa, inducido cuando se agota la glucosa y se genera etanol) y el promotor de la metalotioneína CUP1 (inducido en presencia de Cu²⁺, Zn²⁺).

Los agentes que inducen cualquier promotor inducible dado son conocidos en la técnica. Por ejemplo, los promotores regulables por tetraciclina pueden regularse por tetraciclina o doxiciclina; pueden usarse carbohidratos para inducir un promotor inducible por carbohidratos (por ejemplo, galactosa para un promotor inducible por

galactosa); puede usarse metionina para inducir un promotor inducible por metionina; pueden usarse metales para inducir un promotor de metalotioneína.

5 El ARN diana puede ser un ARN regulador. Los ARN reguladores son bien conocidos en la técnica e incluyen, por ejemplo, microARN, ARN de horquilla corta (ARNhc) y similares.

En algunas realizaciones, la escisión del ARN diana a partir de un ARN precursor inactiva el ARN precursor.

10 La célula hospedadora modificada genéticamente puede ser una célula *in vitro*, por ejemplo, una célula procariota o una célula eucariota (por ejemplo, una célula de mamífero, incluyendo células primarias, líneas celulares transformadas y similares). La célula hospedadora modificada genéticamente puede ser una célula *in vivo*. En algunas realizaciones, la célula *in vivo* es una célula no humana.

15 La célula hospedadora modificada genéticamente puede ser una célula de un organismo multicelular (o puede obtenerse de una célula de un organismo multicelular).

La célula hospedadora modificada genéticamente puede ser una célula obtenida de o presente en un organismo de cualquiera de los seis reinos, por ejemplo, Bacteria (por ejemplo, Eubacteria); Archaeobacteria; Protista; Fungi; Plantae; y Animalia. Organismos adecuados incluyen miembros de tipo vegetal del reino Protista, incluyendo, aunque sin limitación, algas (por ejemplo, algas verdes, algas rojas, glaucofitos, cianobacterias); miembros tipo hongos de Protista, por ejemplo, mohos de cieno, mohos acuáticos, etc.; miembros tipo animal de Protista, por ejemplo, flagelados (por ejemplo, Euglena), ameboides (por ejemplo, amebas), esporozoos (por ejemplo, Apicomplexa, Myxozoa, Microsporidia), y ciliados (por ejemplo, Paramecium). Organismos adecuados incluyen miembros del reino Fungi, incluyendo, aunque sin limitación, miembros de cualquiera de los filos: Basidiomycota (hongos asociados; por ejemplo, miembros de Agaricus, Amanita, Boletus, Cantherellus, etc.); Ascomycota (hongos de saco, incluyendo, por ejemplo, Saccharomyces); Mycophycophyta (líquenes); Zygomycota (hongos de conjugación); y Deuteromycota. Organismos adecuados incluyen miembros del reino Plantae, incluyendo, aunque sin limitación, miembros de cualquiera de las siguientes divisiones: Bryophyta (por ejemplo, musgos), Anthocerotophyta (por ejemplo, briofitas), Hepaticophyta (por ejemplo, hepáticas), Lycophyta (por ejemplo, musgos asociados), Sphenophyta (por ejemplo, colas de caballo), Psilophyta (por ejemplo, helechos en escobilla), Ophioglossophyta, Pterophyta (por ejemplo, helechos), Cycadophyta, Ginkgophyta, Pinophyta, Gnetophyta, y Magnoliophyta (por ejemplo, plantas florales). Organismos adecuados incluyen miembros del reino Animalia, incluyendo, aunque sin limitación, miembros de cualquiera de los filos: Porifera (esponjas); Placozoa; Orthonectida (parásitos de invertebrados marinos); Rhombozoa; Cnidaria (corales, anémonas, medusas, plumas de mar, pensamientos de mar, avispa marina); Ctenophora (ctenóforos); Platyhelminthes (platelmintos); Nemertina (gusanos planos); Nematostomulida (gusanos con mandíbula); Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cycliophora; Mollusca (moluscos); Sipuncula (gusanos del cacahuete); Annelida (gusanos segmentados); Tardigrada (osos de agua); Onychophora (gusanos aterciopelados); Arthropoda (incluyendo los subfilos: Chelicerata, Myriapoda, Hexapoda, y Crustacea, donde el Chelicerata incluye, por ejemplo, arácnidos, Merostomata, y Pycnogonida, donde el Myriapoda incluye, por ejemplo, Chilopoda (ciempiés), Diplopoda (milpiés), Paropoda, y Symphyla, donde el Hexapoda incluye insectos, y donde el Crustacea incluye camarones, krill, percebes, etc.; Phoronida; Ectoprocta (animales del musgo); Brachiopoda; Echinodermata (por ejemplo estrella de mar, margaritas de mar, estrellas pluma, erizos de mar, pepinos de mar, estrellas frágiles, cestas frágiles, etc.); Chaetognatha (gusanos flecha); Hemichordata (gusanos bellota); y Chordata. Miembros adecuados de Chordata incluyen cualquier miembro de los siguientes subfilos: Urochordata (ascidias; incluyendo Ascidiacea, Thaliacea, y Larvacea); Cephalochordata (lancelados); Myxini (mixinas); y Vertebrata, donde miembros de Vertebrata incluyen, por ejemplo, miembros de Petromyzontida (lampreas), Chondrichthyes (peces cartilaginosos), Actinopterygii (curimbas), Actinista (celocantos), Dipnoi (pez pulmonado), Reptilia (reptiles, por ejemplo, serpientes, caimanes, cocodrilos, lagartos, etc.), Aves (pájaros); y Mammalian (mamíferos). Plantas adecuadas incluyen cualquier monocotiledónea y cualquier dicotiledónea.

Por tanto, por ejemplo, una célula hospedadora modificada genéticamente puede ser una célula obtenida de o presente en un protozoo, una planta, un hongo, una célula de alga, una levadura, un reptil, un anfibio, un mamífero, un microorganismo marino, un invertebrado marino, un artrópodo, un isópodo, un insecto, un arácnido, una arqueobacteria y una eubacteria.

Las células de mamífero adecuadas incluyen células primarias y líneas celulares inmortalizadas. Las líneas celulares de mamífero adecuadas incluyen líneas celulares humanas, líneas celulares de primate no humano, líneas celulares de roedor (por ejemplo, ratón, rata) y similares. Las líneas celulares de mamífero adecuadas incluyen, aunque sin limitación, células HeLa (por ejemplo, American Type Culture Collection (ATCC) N.º CCL-2), células CHO (por ejemplo, ATCC N.º CRL9618, CCL61, CRL9096), células 293 (por ejemplo, ATCC N.º CRL-1573), células Vero, células NIH 3T3 (por ejemplo, ATCC N.º CRL-1658), células Huh-7, células BHK (por ejemplo, ATCC N.º CCL10), células PC12 (ATCC N.º CRL1721), células COS, células COS-7 (ATCC N.º CRL1651), células RAT1, células de ratón L (ATCC N.º CCLI.3), células renales embrionarias humanas (HEK) (ATCC N.º CRL1573), células HLHepG2, y similares.

La célula hospedadora modificada genéticamente puede ser una célula obtenida de o presente en un embrión no humano, por ejemplo, un embrión de *Drosophila*; un embrión de pez cebra; un embrión de ratón; etc.

5 La célula hospedadora modificada genéticamente puede ser una célula madre, por ejemplo, una célula madre *in vitro*; una célula madre no humana; etc. Las células madre adecuadas incluyen células madre embrionarias, células madre adultas y células madre pluripotentes inducidas (iPS).

Métodos de aislamiento de ácido nucleico diana

10 La presente descripción proporciona métodos de aislamiento de un ácido nucleico diana a partir de una población mixta de ácidos nucleicos. Los métodos generalmente implican: a) poner en contacto una población mixta de ácidos nucleicos con una endorribonucleasa enzimáticamente inactiva específica de secuencia inmovilizada, donde la población mixta de ácidos nucleicos incluye un ácido nucleico diana que comprende una secuencia de nucleótidos "marca" (o "de reconocimiento") que se une específicamente por la endorribonucleasa enzimáticamente inactiva
15 específica de secuencia inmovilizada, de modo que el ácido nucleico diana que comprende la secuencia de nucleótidos marca ("ácido nucleico diana marcado") se une a la endorribonucleasa enzimáticamente inactiva específica de secuencia inmovilizada, formando un complejo de ácido nucleico diana marcado/endorribonucleasa enzimáticamente activa específica de secuencia inmovilizada, donde la etapa de contacto tiene lugar en una solución líquida (una "solución de unión"); y b) añadir imidazol a la solución líquida hasta una concentración final de
20 aproximadamente 100 mM a aproximadamente 500 mM (por ejemplo, de aproximadamente 100 mM a aproximadamente 150 mM, de aproximadamente 150 mM a aproximadamente 200 mM, de aproximadamente 200 mM a aproximadamente 250 mM, de aproximadamente 250 mM a aproximadamente 300 mM, de aproximadamente 300 mM a aproximadamente 350 mM, de aproximadamente 350 mM a aproximadamente 400 mM, de aproximadamente 400 mM a aproximadamente 450 mM o de aproximadamente 450 mM a aproximadamente 500
25 mM), formando de ese modo una solución de reactivación que reactiva enzimáticamente la endorribonucleasa enzimáticamente inactiva de modo que la endorribonucleasa queda enzimáticamente activa y escinde el ácido nucleico diana de la secuencia de nucleótidos "marca", liberando de ese modo el ácido nucleico diana. La **Figura 9** es una representación esquemática de una realización a modo de ejemplo de un presente método de aislamiento de un ARN diana.

30 El método puede incluir adicionalmente una o más etapas de lavado. Por ejemplo, después de la etapa (a) y antes de la etapa (b), la endorribonucleasa enzimáticamente inactiva específica de secuencia inmovilizada que comprende un ácido nucleico diana unido que comprende una secuencia de nucleótidos "marca" puede lavarse una o más veces con la solución de unión, de modo que el ácido nucleico diana permanezca unido a la endorribonucleasa
35 enzimáticamente inactiva específica de secuencia, y cualquier ácido nucleico no unido se retire por lavado.

La población mixta de ácidos nucleicos puede incluir ARN y ADN. El ácido nucleico diana es un ARN que comprende una secuencia de nucleótidos "marca" o "de reconocimiento" que se une específicamente por la endorribonucleasa específica de secuencia. En su estado enzimáticamente inactivo (estado "no inducido"), la
40 endorribonucleasa puede unirse a, pero no puede escindir, el ARN diana marcado. En su estado enzimáticamente activo (estado "inducido") (por ejemplo, en presencia de imidazol en una concentración de aproximadamente 100 mM a aproximadamente 500 mM), la endorribonucleasa puede tanto unirse a como escindir la secuencia de nucleótidos de reconocimiento en el ácido nucleico diana marcado, liberando de ese modo el ácido nucleico diana de la marca.

45 La solución de unión puede incluir un tampón y una sal; y carece de imidazol. La solución de reactivación puede incluir imidazol en una concentración final de aproximadamente 100 mM a aproximadamente 500 mM, por ejemplo, de aproximadamente 100 mM a aproximadamente 150 mM, de aproximadamente 150 mM a aproximadamente 200 mM, de aproximadamente 200 mM a aproximadamente 250 mM, de aproximadamente 250 mM a aproximadamente 300 mM, de aproximadamente 300 mM a aproximadamente 350 mM, de aproximadamente 350 mM a aproximadamente 400 mM o de aproximadamente 400 mM a aproximadamente 500 mM. La presencia de imidazol reactiva la endorribonucleasa enzimáticamente inactiva específica de secuencia de modo que la endorribonucleasa queda enzimáticamente activa, por ejemplo, la endorribonucleasa muestra al menos aproximadamente un 50%, al menos aproximadamente un 60%, al menos aproximadamente un 70%, al menos aproximadamente un 80%, al menos aproximadamente un 90%, al menos aproximadamente un 95% o más de un 95% de la endorribonucleasa
50 específica de secuencia de tipo silvestre (por ejemplo, una secuencia de aminoácidos representada en la **Figura 5** (por ejemplo, SEQ ID NO: 6, 8 o 9)). Como ejemplo no limitante, la endorribonucleasa enzimáticamente inactiva específica de secuencia es un mutante H29A de Csy4 (como se describe a continuación; y como representa en la **Figura 6**); el contacto del mutante de Csy4 (H29A) con imidazol, como se ha descrito anteriormente, reactiva la endorribonucleasa de modo que es capaz de escindir, de un modo específico de secuencia, una secuencia de reconocimiento en un ácido ribonucleico diana. También es adecuado para su uso un doble mutante H29A, S50C de Csy4 (como se describe a continuación). En algunas realizaciones, la secuencia "marca" o de reconocimiento comprende la secuencia de nucleótidos 5'-GUUCACUGCCGUUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1).

65 La secuencia de nucleótidos "marca" o "de reconocimiento" puede introducirse en un ácido nucleico usando métodos recombinantes convencionales. Por tanto, el ácido nucleico diana marcado incluirá una marca que se escinde enzimáticamente, liberando de ese modo el ácido nucleico diana.

En algunas realizaciones, el ácido nucleico diana marcado (ARN) tendrá uno o más polipéptidos unidos al mismo. Un ARN diana marcado que tiene uno o más polipéptidos unidos al mismo se menciona en este documento como complejo proteico de ARN. Por tanto, en algunas realizaciones, el ARN diana que se aísla usando un presente método es un complejo proteico de ARN. En algunas realizaciones, un presente método puede comprender

5

Un presente método proporciona aislamiento de un ARN diana (o complejo proteico de ARN). En algunas realizaciones, un presente método proporciona la purificación de un ARN diana (o complejo proteico de ARN) de modo que el ARN diana (o complejo proteico de ARN) es al menos aproximadamente un 50 % puro, al menos aproximadamente un 60 % puro, al menos aproximadamente un 70 % puro, al menos aproximadamente un 80 % puro, al menos aproximadamente un 90 % puro, al menos aproximadamente un 95 % puro, al menos aproximadamente un 98 % puro o más de un 98 % puro.

10

En algunas realizaciones, una proteína unida a un ARN diana en un complejo de ARN diana/proteína puede eluirse del complejo de ARN/proteína. La proteína eluida puede caracterizarse adicionalmente, por ejemplo, por secuenciación, digestión enzimática, un ensayo funcional, etc.

15

La población mixta de ácidos nucleicos puede estar presente en un lisado celular. Por ejemplo, se introduce un vector de expresión que comprende una secuencia de nucleótidos que codifica un ARN diana marcado en una célula (por ejemplo, *in vitro* o *in vivo*), de modo que la célula sintetice el ARN diana marcado. Se prepara un lisado a partir de la célula y el lisado (opcionalmente sometido a una o más etapas de enriquecimiento para los ácidos nucleicos) se aplica a la endorribonucleasa enzimáticamente inactiva específica de secuencia inmovilizada.

20

La endorribonucleasa enzimáticamente inactiva específica de secuencia puede inmovilizarse en cualquiera de diversos soportes insolubles. Los soportes insolubles adecuados incluyen, aunque sin limitación, perlas de agarosa, perlas magnéticas, una tira de ensayo, una placa multipocillo y similares. El soporte insoluble puede comprender diversas sustancias (vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poliacrilamidas, agarosas, y magnetita) y puede proporcionarse en diversas formas incluyendo, por ejemplo, perlas de agarosa, perlas de poliestireno, perlas de látex, perlas magnéticas, partículas de metal coloide, chips y superficies de vidrio y/o silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de placas de reacción (por ejemplo, placas multipocillo), tubos de plástico, etc.

25

30

La presente descripción también proporciona un método de aislamiento de un polipéptido que se une a un ARN diana, donde el método comprende: a) poner en contacto un complejo inmovilizado con una solución líquida que comprende un polipéptido que se une al ARN diana, donde el complejo inmovilizado comprende la endorribonucleasa Csy4 variante y un ARN diana marcado que comprende una secuencia de nucleótidos de reconocimiento que se une específicamente por la endorribonucleasa Csy4 variante, donde dicho contacto provoca la unión del polipéptido al ARN diana, donde dicho contacto se realiza en una solución de unión que carece de imidazol; y b) eludir el polipéptido unido.

35

40

Endorribonucleasas

La presente descripción proporciona una endorribonucleasa específica de secuencia. En algunas realizaciones, la presente descripción proporciona una endorribonucleasa específica de secuencia que se une a una secuencia de reconocimiento en un polirribonucleótido diana, pero que no escinde el polirribonucleótido diana, es decir, la endorribonucleasa específica de secuencia es enzimáticamente inactiva en la hidrólisis del polirribonucleótido diana. En algunas realizaciones, la presente descripción proporciona una endorribonucleasa específica de secuencia que se une a una secuencia de reconocimiento en un polirribonucleótido diana, y escinde el polirribonucleótido diana dentro o cerca de la secuencia de reconocimiento, es decir, la endorribonucleasa específica de secuencia es enzimáticamente activa en la hidrólisis del polirribonucleótido diana.

45

50

En algunas realizaciones, una presente endorribonucleasa específica de secuencia se inmoviliza sobre un sustrato insoluble. Los sustratos insolubles adecuados incluyen, aunque sin limitación, perlas de agarosa, perlas magnéticas, una tira de ensayo, una placa multipocillo y similares. El sustrato insoluble puede comprender diversas sustancias (vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poliacrilamidas, agarosas, y magnetita) y puede proporcionarse en diversas formas incluyendo, por ejemplo, perlas de agarosas, perlas de poliestireno, perlas de látex, perlas magnéticas, partículas de metal coloide, chips y superficies de vidrio y/o silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de placas de reacción (por ejemplo, placas multipocillo), tubos de plástico, etc.

55

60

Endorribonucleasa específica de secuencia enzimáticamente inactiva

La presente descripción proporciona una endorribonucleasa específica de secuencia enzimáticamente inactiva, donde la endorribonucleasa específica de secuencia específicamente inactiva se une a una secuencia diana en un polirribonucleótido de un modo específico de secuencia. Una presente endorribonucleasa específica de secuencia enzimáticamente inactiva se une a un polirribonucleótido diana de un modo específico de secuencia, pero no

65

escinde el polirribonucleótido diana. Una presente endorribonucleasa específica de secuencia enzimáticamente inactiva es útil para aislar un ARN diana de una población mixta de ácidos nucleicos, como se ha descrito anteriormente.

- 5 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva comprende una o más sustituciones de aminoácidos en comparación con un polipéptido Csy4, CasE o Cas6 enzimáticamente activo, de origen natural.

- 10 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva comprende una sustitución de aminoácido en His-29 de un polipéptido Csy4 o en una posición equivalente en un polipéptido CasE o Cas6. En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva comprende una sustitución de aminoácido en Ser-148 de un polipéptido Csy4 o en una posición equivalente en un polipéptido CasE o Cas6.

- 15 La **Figura 6** representa ejemplos no limitantes de secuencias de aminoácidos de endorribonucleasa específica de secuencia enzimáticamente inactiva adecuadas. En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva comprende una secuencia de aminoácidos que tiene al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos representada en la **Figura 6** donde la secuencia de aminoácidos incluye la sustitución en His-29, Ser-50 o tanto His-29 como Ser-50. Por ejemplo, la endorribonucleasa Csy4 variante puede incluir una sustitución H29A (His-29 a Ala-29), una sustitución S50C (Ser-50 a Cys-50) o tanto una sustitución H29A como S50C.

- 25 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva es una endorribonucleasa Csy4 variante. En algunas realizaciones una presente endorribonucleasa Csy4 variante comprende una secuencia de aminoácidos que tiene al menos aproximadamente un 95 % de identidad de secuencia de aminoácidos con la secuencia de aminoácidos expuesta en la **Figura 6**, donde la endorribonucleasa comprende una sustitución de aminoácidos en His-29, donde la endorribonucleasa Csy4 variante es enzimáticamente inactiva en ausencia de imidazol, y donde la endorribonucleasa Csy4 variante se puede activar en presencia de imidazol. En algunos casos, la sustitución de aminoácidos es una sustitución His-29 a Ala-29. En algunos casos, la endorribonucleasa Csy4 variante también incluye una sustitución Ser-50. En algunos casos, una presente endorribonucleasa Csy4 variante se une a un sustrato de ARN que comprende la secuencia de nucleótidos 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1).

- 35 Una presente endorribonucleasa específica de secuencia enzimáticamente inactiva es enzimáticamente inactiva "de forma condicional", por ejemplo, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, una presente endorribonucleasa Csy4 variante) es enzimáticamente inactiva en ausencia de imidazol; y la endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, la presente endorribonucleasa Csy4 variante) se puede activar por imidazol. Por ejemplo, la endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, la presente endorribonucleasa Csy4 variante) puede activarse enzimáticamente por contacto de la endorribonucleasa con imidazol a una concentración de aproximadamente 100 mM a aproximadamente 500 mM.

- 45 La presencia de imidazol (por ejemplo en un intervalo de concentración de aproximadamente 100 mM a aproximadamente 500 mM) reactiva la endorribonucleasa enzimáticamente inactiva específica de secuencia de modo que la endorribonucleasa queda enzimáticamente activa, por ejemplo, la endorribonucleasa muestra al menos aproximadamente un 50 %, al menos aproximadamente un 60 %, al menos aproximadamente un 70 %, al menos aproximadamente un 80 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 % o más del 95 % de la endorribonucleasa específica de secuencia de tipo silvestre (por ejemplo, una secuencia de aminoácidos presentada en la Figura 5 (por ejemplo, SEQ ID NO: 6, 8 o 9)).

- 55 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, una presente endorribonucleasa Csy4 variante) comprende un marcador detectable, incluyendo un resto que proporciona una señal detectable. Los marcadores y/o restos detectables adecuados que proporcionan una señal detectable incluyen, aunque sin limitación, una enzima, un radioisótopo, un miembro de un par FRET, un miembro de un par de unión específico; un fluoróforo; una proteína fluorescente; un punto cuántico; y similares.

- 60 Los pares FRET (donante/aceptor) adecuados para su uso incluyen, aunque sin limitación, EDANS/fluoresceína, IAEDANS/fluoresceína, fluoresceína/tetrametilrodamina, fluoresceína/Cy 5, IEDANS/DABCYL, fluoresceína/QSY-7, fluoresceína/LC rojo 640, fluoresceína/Cy 5.5 y fluoresceína/LC rojo 705. Además, puede usarse un par donante/aceptor de fluoróforo/punto cuántico.

- 65 Los fluoróforos ("marcador fluorescente") adecuados incluyen cualquier molécula que pueda detectarse mediante sus propiedades fluorescentes inherentes, que incluyen fluorescencia detectable tras excitación. Los marcadores fluorescentes adecuados incluyen, aunque sin limitación, fluoresceína, rodamina, tetrametilrodamina, eosina,

eritrosina, cumarina, metilcumarina, pireno, verde Malaquita, estilbena, amarillo Lucifer, Cascade Blue™, rojo Texas, IAEDANS, EDANS, BODIPY FL, LC rojo 640, Cy 5, Cy 5.5, LC rojo 705 y verde Oregón. Se describen colorantes ópticos adecuados en el 2002 Molecular Probes Handbook, 9ª Ed., de Richard P. Haugland.

- 5 Las enzimas adecuadas incluyen, aunque sin limitación, peroxidasa de rábano rusticano, luciferasa, β-galactosidasa y similares.

10 Las proteínas fluorescentes adecuadas incluyen, aunque sin limitación, una proteína fluorescente verde (GFP), por ejemplo, una GFP de *Aequoria victoria* o un mutante o derivado de la misma, por ejemplo, como se describe en las patentes de Estados Unidos n.º 6.066.476; 6.020.192; 5.985.577; 5.976.796; 5.968.750; 5.968.738; 5.958.713; 5.919.445; 5.874.304; una proteína fluorescente roja; una proteína fluorescente amarilla; cualquiera de diversas proteínas fluorescentes y coloreadas de especies de antozoos, como se describe en, por ejemplo, Matz *et al.* (1999) Nature Biotechnol. 17:969-973; y similares.

- 15 Nanopartículas adecuadas incluyen, por ejemplo, punto cuántico (QD), nanopartículas fluorescentes o luminiscentes y nanopartículas magnéticas. Puede detectarse cualquier propiedad óptica o magnética o característica de la nanopartícula o nanopartículas.

20 Los QD y métodos para su síntesis son bien conocidos en la técnica (véanse, por ejemplo, las patentes de Estados Unidos n.º 6.322.901; 6.576.291; y 6.815.064). Los QD pueden volverse solubles en agua aplicando capas de recubrimiento que comprenden diversos materiales diferentes (véanse, por ejemplo, las patentes de Estados Unidos n.º 6.423.551; 6.251.303; 6.319.426; 6.426.513; 6.444.143; y 6.649.138). Por ejemplo, los QD pueden solubilizarse usando polímeros anfífilos. Los polímeros a modo de ejemplo que se han empleado incluyen ácido poliacrílico de bajo peso molecular modificado con octilamina, fosfolípidos derivatizados con polietilenglicol (PEG), polianhídridos, copolímeros de bloque, etc. Los QD pueden conjugarse a un polipéptido mediante cualquiera de varios grupos funcionales diferentes o agentes de unión que pueden unirse directa o indirectamente a una capa de recubrimiento (véanse, por ejemplo, las patentes de Estados Unidos n.º 5.990.479; 6.207.392; 6.251.303; 6.306.610; 6.325.144; y 6.423.551).

30 Los QD con una amplia diversidad de espectros de absorción y emisión están disponibles en el mercado, por ejemplo, en Quantum Dot Corp. (Hayward Calif.; ahora propiedad de Invitrogen) o en Evident Technologies (Troy, N.Y.). Por ejemplo, están disponibles QD que tienen longitudes de onda de emisión máximas de aproximadamente 525, 535, 545, 565, 585, 605, 655, 705, y 800 nm. Por tanto, los QD pueden tener un intervalo de diferentes colores a través de la parte visible del espectro y en algunos casos incluso más allá.

35 Los radioisótopos adecuados incluyen, aunque sin limitación, ¹⁴C, ³H, ³²P, ³³P, ³⁵S, ¹²⁵I, y ¹³¹I. El uso de radioisótopos como marcadores es bien conocido en la técnica.

40 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, una presente endorribonucleasa Csy4 variante) se inmoviliza sobre un sustrato insoluble. Los sustratos insolubles adecuados incluyen, aunque sin limitación, perlas de agarosa, perlas magnéticas, una tira de ensayo, una placa multipocillo y similares. El sustrato insoluble puede comprender diversas sustancias (vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poli(acrilamidas, agarosas, y magnetita) y puede proporcionarse en diversas formas, incluyendo, por ejemplo, perlas de agarosas, perlas de poliestireno, perlas de látex, perlas magnéticas, partículas de metal coloide, chips y superficies de vidrio y/o silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de placas de reacción (por ejemplo, placas multipocillo), tubos de plástico, etc.

50 En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente inactiva (por ejemplo, una presente endorribonucleasa Csy4 variante) se purifica, por ejemplo, es al menos un 80 % pura, al menos un 85 % pura, al menos un 90 % pura, al menos un 95 % pura, al menos un 98 % pura, al menos un 99 % pura o más del 99 % pura.

55 Composiciones

La presente descripción proporciona composiciones que comprenden una presente endorribonucleasa enzimáticamente inactiva específica de secuencia. Una presente composición puede comprender, además de una presente endorribonucleasa enzimáticamente inactiva específica de secuencia, uno o más de: una sal, por ejemplo, NaCl, MgCl, KCl, MgSO₄, etc.; un agente tamponante, por ejemplo, un tampón Tris, ácido N-(2-hidroxietil)piperazina-N'-(2-etanosulfónico) (HEPES), ácido 2-(N-morfolino)etanosulfónico (MES), sal sódica del ácido 2-(N-morfolino)etanosulfónico (MES), ácido 3-(N-morfolino)propanosulfónico (MOPS), ácido N-tris[hidroximetil]metil-3-aminopropanosulfónico (TAPS), etc.; un agente solubilizante; un detergente, por ejemplo, un detergente no iónico tal como Tween-20, etc.; un inhibidor de proteasa; y similares.

65

Endorribonucleasa específica de secuencia enzimáticamente activa

En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente activa comprende un resto que proporciona detección. Por ejemplo, una presente endorribonucleasa específica de

5 secuencia enzimáticamente activa puede comprender un resto unido de forma covalente o no covalente que proporciona detección.

Dichos marcadores detectores incluyen cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos. Los restos que proporcionan detección

10 incluyen, aunque sin limitación, una molécula fluorescente; un punto cuántico; una enzima (diferente a la endorribonucleasa), donde la enzima cataliza la conversión de un sustrato en un producto detectable, donde el producto es directamente detectable; una nanopartícula; y similares.

Las proteínas fluorescentes adecuadas que pueden unirse a una presente endorribonucleasa específica de secuencia enzimáticamente activa incluyen, aunque sin limitación, una proteína fluorescente verde (GFP), por

15 ejemplo una GFP de *Aequoria victoria* o un mutante o derivado de la misma, por ejemplo, como se describe en las patentes de Estados Unidos n.º 6.066.476; 6.020.192; 5.985.577; 5.976.796; 5.968.750; 5.968.738; 5.958.713; 5.919.445; 5.874.304; una proteína fluorescente roja; una proteína fluorescente amarilla; cualquiera de diversas proteínas fluorescentes y coloreadas de especies de antozoos, como se describe en, por ejemplo, Matz *et al.* (1999)

20 Nature Biotechnol. 17:969-973; y similares.

Las nanopartículas adecuadas incluyen, por ejemplo, punto cuántico (QD), nanopartículas fluorescentes o luminiscentes y nanopartículas magnéticas. Puede detectarse cualquier propiedad óptica o magnética o

25 característica de la nanopartícula o nanopartículas.

Los QD y métodos para su síntesis son bien conocidos en la técnica (véanse, por ejemplo, las patentes de Estados Unidos n.º 6.322.901; 6.576.291; y 6.815.064). Los QD pueden volverse solubles en agua aplicando capas de recubrimiento que comprenden diversos materiales diferentes (véanse, por ejemplo, las patentes de Estados Unidos

30 n.º 6.423.551; 6.251.303; 6.319.426; 6.426.513; 6.444.143; y 6.649.138). Por ejemplo, los QD pueden solubilizarse usando polímeros anfífilos. Los polímeros a modo de ejemplo que se han empleado incluyen ácido poliacrílico de bajo peso molecular modificado con octilamina, fosfolípidos derivatizados con polietilenglicol (PEG), polianhídridos, copolímeros de bloque, etc. Los QD pueden conjugarse a un polipéptido mediante cualquiera de varios grupos

35 funcionales diferentes o agentes de unión que pueden unirse directa o indirectamente a una capa de recubrimiento (véanse, por ejemplo, las patentes de Estados Unidos n.º 5.990.479; 6.207.392; 6.251.303; 6.306.610; 6.325.144; y 6.423.551).

Los QD con una amplia diversidad de espectros de absorción y emisión están disponibles en el mercado, por ejemplo, en Quantum Dot Corp. (Hayward Calif.; ahora propiedad de Invitrogen) o en Evident Technologies (Troy, N.Y.). Por ejemplo, están disponibles QD que tienen longitudes de onda de emisión máximas de aproximadamente

40 525, 535, 545, 565, 585, 605, 655, 705, y 800 nm. Por tanto, los QD pueden tener un intervalo de diferentes colores a través de la parte visible del espectro y en algunos casos incluso más allá.

En algunas realizaciones, una presente endorribonucleasa específica de secuencia enzimáticamente activa se purifica, por ejemplo, es al menos un 80 % pura, al menos un 85 % pura, al menos un 90 % pura, al menos un 95 %

45 pura, al menos un 98 % pura, al menos un 99 % pura o más del 99 % pura.

Composiciones

La presente descripción proporciona composiciones que comprenden una presente endorribonucleasa enzimáticamente activa específica de secuencia. Una presente composición puede comprender, además de una

50 presente endorribonucleasa enzimáticamente activa específica de secuencia, uno o más de: una sal, por ejemplo, NaCl, MgCl, KCl, MgSO₄, etc.; un agente tamponante, por ejemplo, un tampón Tris, ácido N-(2-hidroxietil)piperazina-N'-(2-etanosulfónico) (HEPES), ácido 2-(N-morfolino)etanosulfónico (MES), sal sódica del ácido 2-(N-morfolino)etanosulfónico (MES), ácido 3-(N-morfolino)propanosulfónico (MOPS), ácido N-tris[hidroximetil]metil-3-

55 aminopropanosulfónico (TAPS), etc.; un agente solubilizante; un detergente, por ejemplo, un detergente no iónico tal como Tween-20, etc.; un inhibidor de proteasa; y similares.

La presente descripción proporciona composiciones que comprenden una presente endorribonucleasa enzimáticamente inactiva específica de secuencia (por ejemplo, una presente endorribonucleasa Csy4 variante).

60 Una presente composición puede comprender, además de una presente endorribonucleasa enzimáticamente inactiva específica de secuencia (por ejemplo, una presente endorribonucleasa Csy4 variante), uno o más de: una sal, por ejemplo, NaCl, MgCl, KCl, MgSO₄, etc.; un agente tamponante, por ejemplo, un tampón Tris, ácido N-(2-hidroxietil)piperazina-N'-(2-etanosulfónico) (HEPES), ácido 2-(N-morfolino)etanosulfónico (MES), sal sódica del ácido 2-(N-morfolino)etanosulfónico (MES), ácido 3-(N-morfolino)propanosulfónico (MOPS), ácido N-tris[hidroximetil]metil-

65 3-aminopropanosulfónico (TAPS), etc.; un agente solubilizante; un detergente, por ejemplo, un detergente no iónico tal como Tween-20, etc.; un inhibidor de proteasa; y similares. En algunas realizaciones la composición carece de

imidazol. En algunas realizaciones la composición comprende imidazol en una concentración de aproximadamente 100 nM a aproximadamente 500 mM.

Métodos de producción de una presente endorribonucleasa específica de secuencia

Una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia) puede producirse por cualquier método conocido, por ejemplo, métodos de síntesis convencionales de síntesis de proteínas; métodos de ADN recombinante; etc.

Cuando una presente endorribonucleasa específica de secuencia se sintetiza químicamente, la síntesis puede proceder mediante fase líquida o fase sólida. La síntesis de polipéptidos en fase sólida (SPPS), en que el aminoácido C-terminal de la secuencia se une a un soporte insoluble seguido por adición secuencial del resto de aminoácidos en la secuencia, es un ejemplo de un método adecuado para la síntesis química de una presente endorribonucleasa específica de secuencia. Están disponibles diversas formas de SPPS, tales como Fmoc y Boc, para sintetizar una presente endorribonucleasa específica de secuencia. Se describen técnicas para síntesis en fase sólida por Barany y Merrifield, *Solid-Phase Peptide Synthesis*; pág. 3-284 en *The Peptides: Analysis, Synthesis, Biology*. Vol. 2: *Special Methods in Peptide Synthesis*, Parte A., Merrifield, *et al.*, J. Am. Chem. Soc., 85: 2149-2156 (1963); Stewart *et al.*, *Solid Phase Peptide Synthesis*, 2ª ed. Pierce Chem. Co., Rockford, Ill. (1984); y Ganesan A. 2006 *Mini Rev. Med Chem.* 6:3-10 y Camarero JA *et al.*, 2005 *Protein Pept Lett.* 12:723-8.

Pueden usarse métodos recombinantes convencionales para la producción de una presente endorribonucleasa específica de secuencia. Por ejemplo, se insertan ácidos nucleicos que codifican una presente endorribonucleasa específica de secuencia en vectores de expresión. Los segmentos de ADN que codifican una presente endorribonucleasa específica de secuencia se unen de forma funcional a secuencias de control en el vector o vectores de expresión que aseguran la expresión de los polipéptidos codificados. Las secuencias de control de la expresión incluyen, aunque sin limitación, promotores (por ejemplo, promotores asociados de forma natural o heterólogos), secuencias señal, elementos potenciadores y secuencias de terminación de la transcripción. Las secuencias de control de la expresión pueden ser sistemas promotores eucariotas en vectores capaces de transformar o transfectar células hospedadoras eucariotas (por ejemplo, células COS o CHO). Una vez se ha incorporado el vector en el hospedador apropiado, el hospedador se mantiene en condiciones adecuadas para expresión de alto nivel de las secuencias de nucleótidos, y la recogida y purificación de la endorribonucleasa.

Ácidos nucleicos y células hospedadoras

La presente descripción proporciona un ácido nucleico que comprende una secuencia de nucleótidos que codifica una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia). En algunas realizaciones, el ácido nucleico es un vector de expresión, donde el vector de expresión puede proporcionar la producción de la endorribonucleasa específica de secuencia, como ejemplo, en una célula.

Una secuencia de nucleótidos que codifica una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia) puede unirse de forma funcional a uno o más elementos reguladores, tales como un promotor y potenciador que permiten la expresión de la secuencia de nucleótidos en las células diana pretendidas (por ejemplo, una célula que está modificada genéticamente para sintetizar la endorribonucleasa codificada).

En algunas realizaciones, un presente ácido nucleico comprende una secuencia de nucleótidos que codifica un polipéptido que tiene al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la Figura 4 o Figura 5. En algunas realizaciones, un presente ácido nucleico comprende una secuencia de nucleótidos que codifica un polipéptido Csy4 variante, como se ha descrito anteriormente.

Una secuencia de nucleótidos que codifica una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia) puede unirse de forma funcional a un elemento de control de la transcripción (por ejemplo, un promotor, un potenciador, etc.). Los elementos promotores y potenciadores adecuados son conocidos en la técnica. Para su expresión en una célula bacteriana, los promotores adecuados incluyen, aunque sin limitación, lacI, lacZ, T3, T7, gpt, lambda P y trc. Para la expresión en una célula eucariota, los promotores adecuados incluyen, aunque sin limitación, el promotor temprano inmediato de citomegalovirus; el promotor de la timidina quinasa del virus del herpes simple; los promotores temprano y tardío de SV40; el promotor presente en repeticiones terminales largas de un retrovirus; el promotor de la metalotioneína-I de ratón; y diversos

promotores específicos de tejido conocidos en la técnica.

En algunas realizaciones, por ejemplo, para la expresión en una célula de levadura, un promotor adecuado es un promotor constitutivo tal como un promotor ADH1, un promotor PGK1, un promotor ENO, un promotor PYK1 y similares; o un promotor regulable tal como un promotor GAL1, un promotor GAL10, un promotor ADH2, un promotor PHO5, un promotor CUP1, un promotor GAL7, un promotor MET25, un promotor MET3, un promotor CYC1, un promotor HIS3, un promotor ADH1, un promotor PGK, un promotor GAPDH, un promotor ADC1, un promotor TRP1, un promotor URA3, un promotor LEU2, un promotor ENO, un promotor TP1, y AOX1 (por ejemplo, para su uso en *Pichia*). La selección del vector y promotor apropiado está acorde al nivel de habilidad habitual en la técnica.

Los promotores adecuados para su uso en células hospedadoras procariotas incluyen, aunque sin limitación, un promotor de ARN polimerasa de bacteriófago T7; un promotor *trp*; un promotor del operón *lac*; un promotor híbrido, por ejemplo, un promotor híbrido *lac/tac*, un promotor híbrido *tac/trc*, un promotor *trp/lac*, un promotor *T7/lac*; un promotor *trc*; un promotor *tac*, y similares; un promotor *araBAD*; promotores regulados *in vivo*, tales como un promotor *ssaG* o un promotor relacionado (véase, por ejemplo, la publicación de patente de Estados Unidos n.º 20040131637), un promotor *pagC* (Pulkkinen y Miller, J. Bacteriol., 1991: 173(1): 86-93; Alpuche-Aranda et al., PNAS, 1992; 89(21): 10079-83), un promotor *nirB* (Harborne et al. (1992) Mol. Micro. 6:2805-2813), y similares (véanse, por ejemplo, Dunstan et al. (1999) Infect. Immun. 67:5133-5141; McKelvie et al. (2004) Vaccine 22:3243-3255; y Chatfield et al. (1992) Biotechnol. 10:888-892); un promotor *sigma70*, por ejemplo, un promotor *sigma70* consenso (véanse, por ejemplo, los números de acceso a GenBank AX798980, AX798961, y AX798183); un promotor de fase estacionaria, por ejemplo, un promotor *dps*, un promotor *spv*, y similares; un promotor derivado de la isla de patogenicidad SPI-2 (véase, por ejemplo, el documento WO96/17951); un promotor *actA* (véase, por ejemplo, Shetron-Rama et al. (2002) Infect. Immun. 70:1087-1096); un promotor *rpsM* (véase, por ejemplo, Valdivia y Falkow (1996). Mol. Microbiol. 22:367); un promotor *tet* (véase, por ejemplo, Hillen, W. y Wissmann, A. (1989) In Saenger, W. y Heinemann, U. (eds), Topics in Molecular and Structural Biology, Protein-Nucleic Acid Interaction. Macmillan, Londres, RU, Vol. 10, pág. 143-162); un promotor SP6 (véase, por ejemplo, Melton et al. (1984) Nucl. Acids Res. 12:7035); y similares. Los promotores fuertes adecuados para su uso en procariotas tales como *Escherichia coli* incluyen, aunque sin limitación, *Trc*, *Tac*, *T5*, *T7*, y *P_{Lambda}*. Ejemplos no limitantes de operadores para su uso en células hospedadoras bacterianas incluyen un operador del promotor de lactosa (la proteína represora *Lacl* cambia de conformación cuando se pone en contacto con lactosa, evitando de ese modo que la proteína represora *Lacl* se una al operador), un operador del promotor de triptófano (cuando está en complejo con triptófano, la proteína represora *TrpR* tiene una conformación que se une al operador; en ausencia de triptófano, la proteína represora *TrpR* tiene una conformación que no se une al operador), y un operador del promotor *tac* (véase, por ejemplo, deBoer et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80:21-25).

Una secuencia de nucleótidos que codifica una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia puede estar presente en un vector de expresión y/o un vector de clonación. Un vector de expresión puede incluir un marcador de selección, un origen de replicación y otras características que proporcionan replicación y/o mantenimiento del vector.

Son conocidas grandes cantidades de vectores y promotores adecuados para los expertos en la materia; muchos están disponibles en el mercado para generar una presente construcción recombinante. Los siguientes vectores se proporcionan a modo de ejemplo. Bacterianos: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene, La Jolla, Calif., EE.UU.); pTrc99A, pKK223-3, pKK233-3, pDR540, y pRIT5 (Farmacia, Uppsala, Suecia). Eucarióticos: pWLneo, pSV2cat, pOG44, PXR1, pSG (Stratagene) pSVK3, pBPV, pMSG y pSVL (Farmacia).

Los vectores de expresión generalmente tienen sitios de restricción convenientes localizados cerca de la secuencia promotora para proporcionar la inserción de secuencias de ácido nucleico que codifican proteínas heterólogas. Puede estar presente un marcador de selección funcional en el hospedador de expresión. Los vectores de expresión adecuados incluyen, aunque sin limitación, vectores víricos (por ejemplo, vectores víricos basados en el virus vaccinia; poliovirus; adenovirus (véanse, por ejemplo, Li et al., Invest Ophthalmol Vis Sci 35:2543 2549, 1994; Borrás et al., Gene Ther 6:515 524, 1999; Li y Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; documento WO 94/12649, documento WO 93/03769; documento WO 93/19191; documento WO 94/28938; documento WO 95/11984 y documento WO 95/00655); virus adenoasociados (véanse, por ejemplo, Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Ophthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava en el documento WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; y Flotte et al., PNAS (1993) 90:10613-10617); SV40; virus del herpes simple; virus de la inmunodeficiencia humana (véanse, por ejemplo, Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); un vector retrovírico (por ejemplo, el Virus de la Leucemia Murina, virus de necrosis del bazo, y vectores derivados de retrovirus tales como Virus del Sarcoma de Rous, Virus del Sarcoma de Harvey, virus de la leucosis aviar, virus de la inmunodeficiencia humana, virus del sarcoma mieloproliferativo, y virus de tumor mamario); y similares.

La presente descripción proporciona células hospedadoras modificadas genéticamente aisladas (por ejemplo, células *in vitro*) que se modifican genéticamente con un presente ácido nucleico. En algunas realizaciones, una presente célula hospedadora modificada genéticamente aislada puede producir una presente endorribonucleasa específica de secuencia (por ejemplo, una presente endorribonucleasa enzimáticamente activa específica de secuencia; una presente endorribonucleasa enzimáticamente inactiva específica de secuencia).

Células hospedadoras adecuadas incluyen células hospedadoras eucariotas, tales como una célula de mamífero, una célula hospedadora de insecto, una célula de levadura; y células procariotas, tales como una célula bacteriana. La introducción de un presente ácido nucleico en la célula hospedadora puede lograrse, por ejemplo, por precipitación con fosfato cálcico, transfección mediada por DEAE dextrano, transfección mediada por liposomas, electroporación u otro método conocido.

Células de mamífero adecuadas incluyen células primarias y líneas celulares inmortalizadas. Las líneas celulares de mamífero adecuadas incluyen líneas celulares humanas, líneas celulares de primate no humano, líneas celulares de roedor (por ejemplo, ratón, rata) y similares. Las líneas celulares de mamífero adecuadas incluyen, aunque sin limitación, células HeLa (por ejemplo, American Type Culture Collection (ATCC) N.º CCL-2), células CHO (por ejemplo, ATCC N.º CRL9618, CCL61, CRL9096), células 293 (por ejemplo, ATCC N.º CRL-1573), células Vero, células NIH 3T3 (por ejemplo, ATCC N.º CRL-1658), células Huh-7, células BHK (por ejemplo, ATCC N.º CCL10), células PC12 (ATCC N.º CRL1721), células COS, células COS-7 (ATCC N.º CRL1651), células RAT1, células L de ratón (ATCC N.º CCLI.3), células renales embrionarias humanas (HEK) (ATCC N.º CRL1573), células HLHepG2, y similares.

Células de levadura adecuadas incluyen, aunque sin limitación, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyceslactis*, *Candida albicans*, *Aspergillusnidulans*, *Aspergillusniger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, y similares.

Células procariotas adecuadas incluyen, aunque sin limitación, cualquiera de diversas cepas de laboratorio de *Escherichia coli*, *Lactobacillus sp.*, *Salmonella sp.*, *Shigella sp.*, y similares. Véanse, por ejemplo, Carrier et al. (1992) J. Immunol. 148:1176-1181; patente de Estados Unidos n.º 6.447.784; y Sizemore et al. (1995) Science 270:299-302. Ejemplos de cepas de *Salmonella* que pueden emplearse en la presente invención incluyen, aunque sin limitación, *Salmonella typhi* y *S. typhimurium*. Cepas de *Shigella* adecuadas incluyen, aunque sin limitación, *Shigella flexneri*, *Shigella sonnei*, y *Shigella dysenteriae*. Normalmente, la cepa de laboratorio es una que es no patogénica. Ejemplos no limitantes de otras bacterias adecuadas incluyen, aunque sin limitación, *Bacillus subtilis*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus sp.*, y similares. En algunas realizaciones, la célula hospedadora es *Escherichia coli*.

Kits

La presente descripción también proporciona kits para determinar la secuencia de nucleótidos de un polirribonucleótido diana. La presente descripción proporciona kits para realizar escisión específica de secuencia de un polirribonucleótido sustrato. La presente descripción proporciona kits para realizar la detección de una secuencia de ARN en un polirribonucleótido diana. La presente descripción proporciona kits para realizar el aislamiento de un ARN diana. La presente descripción proporciona kits para realizar el aislamiento de un polipéptido que se une a un ARN diana.

Kits para realizar secuenciación directa de un polirribonucleótido

Un presente kit para realizar secuenciación directa de un polirribonucleótido incluye al menos una presente endorribonucleasa enzimáticamente inactiva específica de secuencia, donde la endorribonucleasa enzimáticamente inactiva específica de secuencia está purificada. En algunas realizaciones, la endorribonucleasa específica de secuencia enzimáticamente inactiva se une a una molécula aceptora o a una molécula donante, para detección FRET.

Un presente kit para realizar secuenciación directa de un polirribonucleótido incluye al menos una presente endorribonucleasa enzimáticamente inactiva específica de secuencia; y puede incluir uno o más componentes adicionales, donde el uno o más componentes adicionales pueden ser: 1) un tampón; 2) una sonda oligonucleotídica que comprende una secuencia definida; 3) una sonda oligonucleotídica que comprende una secuencia definida, donde la sonda oligonucleotídica está unida a una molécula aceptora o a una molécula donante, para detección FRET; 4) un soporte insoluble, para unir a un polirribonucleótido diana; 5) un polirribonucleótido de control positivo, donde el polirribonucleótido de control positivo comprende una secuencia conocida de nucleótidos; 6) una sonda oligonucleotídica de control positivo que se une a y forma un dúplex con la secuencia conocida del polirribonucleótido de control positivo.

Además de los componentes mencionados anteriormente, un presente kit puede incluir adicionalmente instrucciones para usar los componentes del kit para poner en práctica los presentes métodos. Las instrucciones para poner en práctica los presentes métodos están generalmente registradas en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden estar impresas sobre un sustrato, tal como papel o plástico, etc. Por tanto, las instrucciones pueden estar presentes en los kits como un prospecto, en la etiqueta del recipiente del kit o en componentes del mismo (es decir, asociadas con el envase o subenvase) etc. En otras realizaciones, las instrucciones están presentes como un archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, etc. En otras realizaciones más, las instrucciones exactas no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota, por ejemplo, mediante internet. Un ejemplo de esta realización es un kit que incluye una dirección web donde pueden verse las instrucciones y/o a partir de la cual pueden descargarse las instrucciones. Como con las instrucciones, este medio para obtener las instrucciones está registrado en un sustrato adecuado.

15 Kits para realizar escisión específica de secuencia de un polirribonucleótido sustrato

Un presente kit para realizar escisión específica de secuencia de un polirribonucleótido sustrato incluye al menos una endorribonucleasa específica de secuencia purificada y/o un ácido nucleico que comprende una secuencia de nucleótidos que codifica la endorribonucleasa específica de secuencia. Un presente kit para realizar la escisión específica de secuencia de un polirribonucleótido sustrato puede incluir, además de una endorribonucleasa específica de secuencia purificada (y/o un ácido nucleico que comprende una secuencia de nucleótidos que codifica la endorribonucleasa específica de secuencia), uno o más componentes adicionales. Los componentes adicionales adecuados incluyen, por ejemplo, un tampón; un polirribonucleótido sustrato que sirve como control positivo; patrones de tamaño de polirribonucleótido; un sustrato de control negativo; y similares. Los componentes pueden estar cada uno en recipientes diferentes. El kit puede incluir adicionalmente uno o más controles positivos y negativos.

Además de los componentes mencionados anteriormente, un presente kit puede incluir adicionalmente instrucciones para usar los componentes del kit para poner en prácticas los presentes métodos. Las instrucciones para poner en práctica los presentes métodos están generalmente registradas en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden estar impresas sobre un sustrato, tal como papel o plástico, etc. Por tanto, las instrucciones pueden estar presentes en los kits como un prospecto, en la etiqueta del recipiente del kit o componentes del mismo (es decir, asociada con el envase o subenvase), etc. En otras realizaciones, las instrucciones están presentes como un archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, etc. En otras realizaciones más, las instrucciones exactas no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota, por ejemplo, mediante internet. Un ejemplo de esta realización es un kit que incluye una dirección web donde pueden verse las instrucciones y/o a partir de la cual pueden descargarse las instrucciones. Como con las instrucciones, este medio para obtener las instrucciones está registrado en un sustrato adecuado.

40 Kits para realizar la detección de una secuencia en un polirribonucleótido diana

Un presente kit para realizar la detección de una secuencia en un polirribonucleótido diana (por ejemplo, para realizar la detección de un polirribonucleótido) puede incluir una sonda oligonucleotídica que comprende una secuencia conocida. En algunas realizaciones, el kit incluirá una sonda oligonucleotídica que comprende una secuencia conocida y que comprende un resto detectable, por ejemplo, un polipéptido que puede detectarse usando un ensayo inmunológico; una proteína fluorescente; un luciferina; etc. El kit puede incluir adicionalmente un polirribonucleótido de control positivo que comprende una secuencia de nucleótidos capaz de formar un dúplex con la sonda oligonucleotídica. El kit puede incluir adicionalmente una endorribonucleasa específica de secuencia enzimáticamente activa que detecta específicamente y escinde un dúplex formado la sonda oligonucleotídica y un polirribonucleótido diana. El kit puede incluir adicionalmente uno o más de un tampón; componentes para detectar el resto de detectable; una tira de ensayo; y similares. El kit puede incluir adicionalmente uno o más controles positivos y negativos.

Además de los componentes mencionados anteriormente, un presente kit puede incluir adicionalmente instrucciones para usar los componentes del kit para poner en práctica los presentes métodos. Las instrucciones para poner en práctica los presentes métodos están generalmente registrados en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden estar impresas sobre un sustrato, tal como papel o plástico, etc. Por tanto, las instrucciones pueden estar presentes en los kits como un prospecto, en la etiqueta del recipiente del kit o componentes del mismo, (es decir, asociada con el envase o subenvase), etc. En otras realizaciones, las instrucciones están presentes como un archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, etc. En otras realizaciones más, las instrucciones exactas no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota, por ejemplo, mediante internet. Un ejemplo de esta realización es un kit que incluye una dirección web donde pueden verse las instrucciones y/o a partir de la cual pueden descargarse las instrucciones. Como con las instrucciones, este medio para obtener las instrucciones está registrado en un sustrato adecuado.

Kits para realizar el aislamiento de un ARN diana

Un presente kit para realizar el aislamiento (por ejemplo, purificación) de un ARN diana puede incluir uno o más de:

5 1) una presente endorribonucleasa enzimáticamente inactiva específica de secuencia; 2) una construcción de expresión que comprende una secuencia de nucleótidos "marca", es decir, una secuencia de nucleótidos que se une específicamente por la endorribonucleasa enzimáticamente inactiva específica de secuencia, donde puede insertarse una secuencia de nucleótidos que codifica un ARN diana de elección 3' de la secuencia de nucleótidos "marca"; y 3) imidazol. La endorribonucleasa enzimáticamente inactiva específica de secuencia puede inmovilizarse

10 en un soporte insoluble. El kit puede incluir adicionalmente una composición líquida para poner en contacto una población mixta de ácidos nucleicos con la endorribonucleasa enzimáticamente inactiva específica de secuencia inmovilizada. El kit puede incluir adicionalmente un tampón de lavado. El kit puede incluir adicionalmente uno o más controles positivos y negativos. Un control positivo incluiría un vector de expresión que comprende una secuencia de nucleótidos por codifica un ARN diana marcado, donde la marca se une específicamente por la endorribonucleasa enzimáticamente inactiva específica de secuencia. Los componentes pueden estar cada uno en recipientes diferentes.

Por ejemplo, un presente kit puede incluir una presente endorribonucleasa enzimáticamente inactiva específica de secuencia. Un presente kit puede incluir adicionalmente un vector de expresión recombinante que comprende, en

20 orden de 5' a 3' y en unión funcional; a) una secuencia de nucleótidos que codifica un sustrato de ARN que se une específicamente por una presente endorribonucleasa Csy4 variante; y b) un sitio de clonación múltiple adecuado para la inserción de un ácido nucleico que codifica el ARN diana. La secuencia de nucleótidos que codifica el sustrato de ARN puede estar unida de forma funcional a un promotor. En algunos casos, el promotor es un promotor inducible. El sustrato de ARN puede comprender la secuencia de nucleótidos 5'-

25 GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1). En algunos casos, el vector de expresión recombinante comprende, insertado en el sitio de clonación múltiple, una secuencia de nucleótidos que codifica el ARN diana. El kit puede incluir adicionalmente un tampón que carece de imidazol. El kit puede incluir adicionalmente imidazol o una solución de imidazol. El kit puede incluir adicionalmente uno o más tampones de lavado. En algunos casos, el kit incluirá un vector de expresión de control positivo. La endorribonucleasa Csy4 variante puede

30 inmovilizarse en un soporte insoluble, donde el soporte insoluble adecuado incluye, aunque sin limitación, perlas de agarosas, perlas magnéticas, una tira de ensayo, una placa multipocillo y similares. El soporte insoluble puede comprender diversas sustancias (vidrio, poliestireno, cloruro de polivinilo, polipropileno, polietileno, policarbonato, dextrano, nylon, amilosa, celulosas naturales y modificadas, poliacrilamidas, agarosas, y magnetita) y puede proporcionarse en diversas formas incluyendo, por ejemplo, perlas de agarosa, perlas de poliestireno, perlas de

35 látex, perlas magnéticas, partículas de metal coloide, chips y superficies de vidrio y/o silicio, tiras de nitrocelulosa, membranas de nylon, láminas, pocillos de placas de reacción (por ejemplo, placas multipocillo), tubos de plástico, etc.

Además de los componentes mencionados anteriormente, un presente kit puede incluir adicionalmente instrucciones para usar los componentes del kit para poner en práctica los presentes métodos. Las instrucciones para poner en práctica los presentes métodos están generalmente registrados en un medio de grabación adecuado. Por ejemplo, las instrucciones pueden estar impresas sobre un sustrato, tal como papel o plástico, etc. Por tanto, las instrucciones pueden estar presentes en los kits como un prospecto, en la etiqueta del recipiente del kit o componentes del mismo, (es decir, asociada con el envase o subenvase), etc. En otras realizaciones, las instrucciones están presentes como

40 un archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, etc. En otras realizaciones más, las instrucciones exactas no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota, por ejemplo, mediante internet. Un ejemplo de esta realización es un kit que incluye una dirección web donde pueden verse las instrucciones y/o a partir de la cual pueden descargarse las instrucciones. Como con las instrucciones, este medio

50 para obtener las instrucciones está registrado en un sustrato adecuado.

Métodos de secuenciación directa de un polirribonucleótido diana

La presente descripción proporciona un método de determinación directa de la secuencia de nucleótidos de un polirribonucleótido diana. Por tanto, por ejemplo, el método no requiere síntesis de un polidesoxirribonucleótido equivalente de un polirribonucleótido diana para determinar la secuencia de nucleótidos de polirribonucleótido diana.

El diagnóstico vírico, la medicina personalizada, el análisis de transcrito de una única célula y el perfilado traduccional son todos los campos en que encuentra uso la detección y secuenciación directa de ARN. Un presente método de secuenciación de polirribonucleótido, y un presente método de detección de una secuencia específica en un polirribonucleótido, encuentran uso en estos diversos campos.

60

Un presente método de secuenciación de polirribonucleótido generalmente implica: a) poner en contacto un polirribonucleótido diana con una sonda oligonucleotídica que comprende una secuencia conocida específica y una endorribonucleasa específica de secuencia enzimáticamente inactiva en condiciones que favorecen la formación de dúplex entre la sonda oligonucleotídica y el polirribonucleótido diana, donde la endorribonucleasa específica de

65

secuencia enzimáticamente inactiva se une a la secuencia específica en el dúplex; y b) detectar la unión específica entre la sonda nucleotídica y el polirribonucleótido diana, donde la unión específica de la endorribonucleasa específica de secuencia enzimáticamente inactiva al dúplex indica la presencia de la secuencia específica en el polirribonucleótido diana.

5 En algunos casos, la endorribonucleasa específica de secuencia enzimáticamente inactiva se une (de forma covalente o no covalente) a un marcador de emisión. Por "marcador de emisión" se entiende cualquier molécula que pueda detectarse mediante sus propiedades inherentes de emisión, que incluyen emisión detectable tras excitación. Los marcadores de emisión adecuados incluyen, aunque sin limitación, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metilcumarinas, pireno, verde Malaquita, estilbena, amarillo Lucifer, Cascade Blue™, rojo Texas, IAEDANS, EDANS, BODIPY FL, LC rojo 640, Cy 5, Cy 5.5, LC rojo 705 y verde Oregón. Se describen colorantes ópticos adecuados en el 2002 Molecular Probes Handbook, 9ª Ed., de Richard P. Haugland.

15 En algunos casos, la sonda oligonucleotídica usada en un presente método de secuenciación de polirribonucleótido se une a una molécula donante, la endorribonucleasa específica de secuencia enzimáticamente inactiva se une a una molécula aceptora, y la detección de la formación de dúplex es por transferencia de energía de resonancia de fluorescencia (también mencionada como "transferencia de energía de resonancia de Förster" o "FRET").

20 La transferencia de energía de resonancia de Förster (FRET) es un fenómeno conocido en la técnica donde la excitación de un colorante de emisión se transfiere a otro con emisión de un fotón. Un par FRET consiste en un cromóforo donante y un cromóforo aceptor (donde el cromóforo aceptor puede ser una molécula inactivadora). El espectro de emisión del donante y el espectro de absorción del aceptor deben solaparse, y las moléculas deben estar en cercana proximidad. La distancia entre el donante y el aceptor en que el 50 % de los donantes se desactivan (transferencia de energía al aceptor) se define por el radio de Förster, que es normalmente de 10-100 angstrom. Pueden detectarse cambios en el espectro de emisión que comprenden pares FRET, lo que indica cambios en la cantidad de los que están en cercana proximidad (es decir, en 100 angstrom entre sí). Esto normalmente resulta de la unión o disociación de dos moléculas, una de las cuales está marcada con un donante FRET y la otra de las cuales está marcada con un aceptor FRET, donde dicha unión pone el par FRET en cercana proximidad.

30 La unión de dichas moléculas provocará una emisión aumentada del aceptor y/o la inactivación de la emisión de fluorescencia del donante. Los pares FRET (donante/aceptor) adecuados para su uso incluyen, aunque sin limitación, EDANS/fluoresceína, IAEDANS/fluoresceína, fluoresceína/tetrametilrodamina, fluoresceína/Cy 5, EDANS/DABCYL, fluoresceína/QSY-7, fluoresceína/LC rojo 640, fluoresceína/Cy 5.5 y fluoresceína/LC rojo 705. Además, puede usarse un par donante/aceptor de fluoróforo/punto cuántico. EDANS es ácido (5-((2-aminoetil)amino)naftaleno-1-sulfónico); IAEDANS es ácido 5-((2-(yodoacetil)amino)etil)amino)naftaleno-1-sulfónico; DABCYL es ácido 4-(4-dimetilaminofenil)diacenilbenzoico.

40 Cy3, Cy5, Cy 5.5 y similares son cianinas. Por ejemplo, Cy3 y Cy5 son colorantes fluorescentes solubles en agua reactivos de la familia de colorantes cianina. Los colorantes Cy3 son rojos (excitación ~550 nm, emisión ~570 nm y por lo tanto aparece verde), mientras que Cy5 es fluorescente en la región roja (~650/670 nm) pero absorbe en la región naranja (~649 nm). También pueden usarse colorantes Alexa Fluor, Dylight, IRIS Dyes, colorantes Seta, colorantes SeTau, colorantes SRfluor y colorantes Square.

45 En otro aspecto de FRET, puede emplearse una molécula donante de emisión y una molécula aceptora no de emisión ("inactivadora"). En esta aplicación, la emisión del donante aumentará cuando el inactivador se desplaza de la cercana proximidad al donante y se disminuirá la emisión cuando el inactivador se pone en cercana proximidad al donante. Los inactivadores útiles incluyen, aunque sin limitación, DABCYL, QSY 7 y QSY 33. Los pares útiles de donante fluorescente/inactivador incluyen, aunque sin limitación, EDANS/DABCYL, rojo Texas/DABCYL, BODIPY/DABCYL, amarillo Lucifer/DABCYL, cumarina/DABCYL y fluoresceína/QSY 7 dye.

50 En algunos casos, la endorribonucleasa específica de secuencia enzimáticamente inactiva se une (de forma covalente o no covalente) a una enzima marcadora. Por "enzima marcadora" se entiende una enzima que puede hacerse reaccionar en presencia de un sustrato de enzima marcadora que produce un producto detectable. Las enzimas marcadoras adecuadas también incluyen marcadores ópticamente detectables (por ejemplo, en el caso de peroxidasa de rábano rústico (HRP)). Las enzimas marcadoras adecuadas incluyen, aunque sin limitación, HRP, fosfatasa alcalina, luciferasa, β-galactosidasa y glucosa oxidasa. Los métodos para el uso de dichos sustratos son bien conocidos en la técnica. La presencia de la enzima marcadora se revela generalmente a través de la catálisis de la enzima de una reacción con un sustrato de enzima marcadora, produciendo un producto identificable. Dichos productos pueden ser opacos, tales como la reacción de peroxidasa de rábano rústico con tetrametilbencidina, y pueden tener diversos colores. Se han desarrollado otros sustratos de enzima marcadora, tales como Luminol (disponible en Pierce Chemical Co.), que producen productos de reacción fluorescentes. Los métodos para identificar enzimas marcadoras con sustratos de enzimas marcadoras son bien conocidos en la técnica y están disponibles muchos kits comerciales. Se describen ejemplos y métodos para el uso de diversas enzimas marcadoras en Savage *et al.*, *Previews* 247:6-9 (1998), Young, J. *Virol. Methods* 24:227-236 (1989).

65 En algunos casos, la endorribonucleasa específica de secuencia enzimáticamente inactiva comprende un

radioisótopo. Por "radioisótopo" se entiende cualquier molécula radiactiva. Los radioisótopos adecuados para su uso en la invención incluyen, aunque sin limitación, ^{14}C , ^3H , ^{32}P , ^{33}P , ^{35}S , ^{125}I , y ^{131}I . El uso de radioisótopos como marcadores es bien conocido en la técnica.

5 En algunos casos, la endorribonucleasa específica de secuencia enzimáticamente inactiva se une (de forma covalente o no covalente) a un miembro de un par de unión específico ("compañero de un par de unión"). Por "compañero de un par de unión" o "miembro de un par de unión" se entiende uno de un primer y un segundo resto, donde el primer y el segundo resto tiene una afinidad de unión específica por el otro. Los pares de unión adecuados incluyen, aunque sin limitación, antígeno/anticuerpos (por ejemplo, digoxigenina/antidigoxigenina, dinitrofenilo (DNP)/anti-DNP, dansilo-X/antidansilo, fluoresceína/antifluoresceína, amarillo Lucifer/anti amarillo Lucifer y rodamina/antirrodamina), biotina/avidina (o biotina/estreptavidina) y proteína de unión a calmodulina (CBP)/calmodulina.

10 En algunas realizaciones, la sonda oligonucleotídica comprende una modificación que proporciona resistencia aumentada a hidrólisis no específica. Dichas modificaciones son bien conocidas en la técnica e incluye, por ejemplo, enlaces internucleosídicos resistentes a nucleasa, estructuras modificadas o modificaciones de bases, sustituciones de bases, modificaciones de azúcares y similares.

20 Las estructuras oligonucleotídicas modificadas adecuadas que contienen un átomo de fósforo en las mismas incluyen, por ejemplo, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, fosfonatos de metilo y otros fosfonatos de alquilo incluyendo 3'-alquileo fosfonatos, 5'-alquileo fosfonatos, y fosfonatos quirales, fosfinatos, fosforamidatos incluyendo 3'-aminofosforamidato y aminoalquilfosforamidatos, fosforodiamidatos, tionofosforamidatos, tionalquilfosfonatos, tionalquilfosfotriésteres, selenofosfatos y boranofosfatos que tienen enlaces 3'-5' normales, análogos unidos 2'-5' de estos, y aquellos que tienen polaridad invertida donde uno o más enlaces internucleotídicos son un enlace 3' a 3', 5' a 5' o 2' a 2'. Los oligonucleótidos adecuados que tienen polaridad invertida comprenden un único enlace 3' a 3' en el enlace internucleotídico más 3', es decir, un único resto nucleosídico invertido que puede ser uno básico (la nucleobase se pierde o tiene un grupo hidroxilo en lugar de la misma). También se incluyen diversas sales (tales como, por ejemplo, potasio o sodio), sales mixtas y formas de ácido libre.

30 Un oligonucleótido modificado puede comprender uno o más fosforotioatos y/o enlaces internucleosídico de heteroátomos, en particular $-\text{CH}_2\text{-NH-O-CH}_2-$, $-\text{CH}_2\text{-N}(\text{CH}_3)\text{-O-CH}_2-$ (conocido como metileno (metilimino) o estructura MMI), $-\text{CH}_2\text{-O-N}(\text{CH}_3)\text{-CH}_2-$, $-\text{CH}_2\text{-N}(\text{CH}_3)\text{-N}(\text{CH}_3)\text{-CH}_2-$ y $-\text{O-N}(\text{CH}_3)\text{-CH}_2\text{-CH}_2-$ (donde el enlace internucleotídico fosfodiéster nativo está representado como $-\text{O-P}(=\text{O})(\text{OH})\text{-O-CH}_2-$). Los enlaces internucleosídicos de tipo MMI se describen en la patente de Estados Unidos n.º 5.489.677 referenciada anteriormente. Los enlaces internucleosídicos amida adecuados se describen en la patente de Estados Unidos n.º 5.602.240.

40 Un oligonucleótido modificado puede comprender una o más estructuras de esqueleto morfolino como se describe en, por ejemplo, la patente de Estados Unidos n.º 5.034.506. Por ejemplo, en algunas realizaciones, un oligonucleótido modificado comprende un anillo morfolino de 6 miembros en lugar de un anillo de ribosa. En algunas de estas realizaciones, un fosforodiamidato u otro enlace internucleosídico no fosfodiéster reemplaza un enlace fosfodiéster. Los ácidos morfolinonucleicos ("morfolinos") incluyen bases unidas a anillos de morfolina en lugar de anillos de desoxirribosa; además, la estructura fosfato puede incluir un grupo no fosfato, por ejemplo, un grupo fosforodiamidato en lugar de fosfatos. Summerton (1999) *Biochim. Biophys. Acta* 1489:141; Heasman (2002) *Dev. Biol.* 243:209; Summerton y Weller (1997) *Antisense & Nucl. Acid Drug Dev.* 7:187; Hudziak *et al.* (1996) *Antisense & Nucl. Acid Drug Dev.* 6:267; Partridge *et al.* (1996) *Antisense & Nucl. Acid Drug Dev.* 6:169; Amantana *et al.* (2007) *Bioconj. Chem.* 18:1325; Morcos *et al.* (2008) *BioTechniques* 45:616.

50 Un oligonucleótido modificado puede comprender una estructura modificada. Las estructuras polinucleotídicas modificadas que no incluyen un átomo de fósforo en la misma tienen estructuras que se forman por cortos enlaces internucleosídicos de alquilo o cicloalquilo de cadena corta, enlaces internucleosídicos de alquilo o cicloalquilo y heteroátomos mixtos o uno o más enlaces internucleosídicos heteroatómicos o heterocíclicos de cadena corta. Estos incluyen aquellos que tienen enlaces morfolino (formados en parte, a partir de la parte de azúcar de un nucleósido); estructuras siloxano; estructuras sulfuro, sulfóxido y sulfona; estructuras formacetilo y tioformacetilo; estructuras metileno formacetilo y tioformacetilo; estructuras riboacetilo; estructuras que contienen alqueno; estructuras sulfamato; estructuras metilnimino y metilhidrazino; estructuras sulfonato y sulfonamida; estructuras amida; y otras que tienen partes mixtas de componente N, O, S y CH_2 .

60 Un oligonucleótido modificado puede comprender uno o más restos de azúcar sustituidos. Los oligonucleótidos adecuados comprenden un grupo sustituyente de azúcar seleccionado de: OH; F; O-, S- o N-alquilo; O-, S- o N-alqueno; O-, S- o N-alquino; u O-alquil-O-alquilo, donde el alquilo, alqueno y alquino pueden ser alquilo C_1 a C_{10} sustituido o sin sustituir o alqueno y alquino C_2 a C_{10} . También son adecuados $\text{O}((\text{CH}_2)_n\text{O})_m\text{CH}_3$, $\text{O}(\text{CH}_2)_n\text{OCH}_3$, $\text{O}(\text{CH}_2)_n\text{NH}_2$, $\text{O}(\text{CH}_2)_n\text{CH}_3$, $\text{O}(\text{CH}_2)_n\text{ONH}_2$, y $\text{O}(\text{CH}_2)_n\text{ON}((\text{CH}_2)_n\text{CH}_3)_2$, donde n y m son de 1 a aproximadamente 10. Otros oligonucleótidos adecuados comprenden un grupo sustituyente de azúcar seleccionado de alquilo inferior C_1 a C_{10} , alquilo, alqueno, alquino, alcarilo, aralquilo, O-alcarilo u O-aralquilo inferior sustituido, SH, SCH_3 , OCN, Cl, Br, CN, CF_3 , OCF_3 , SOCH_3 , SO_2CH_3 , ONO_2 , NO_2 , N_3 , NH_2 , heterocicloalquilo, heterocicloalcarilo, aminoalquilamino,

polialquilamino, sililo sustituido, un grupo de escisión de ARN, un grupo indicador, un intercalador y similares. Una modificación adecuada incluye 2'-metoxietoxi (2'-O-CH₂CH₂OCH₃, también conocido como 2'-O-(2-metoxietilo) o 2'-MOE) (Martin *et al.*, Helv. Chim. Acta, 1995, 78, 486-504), es decir, un grupo alcoialcoxi. Una modificación adecuada adicional incluye 2'-dimetilaminoetoxietoxi, es decir, un grupo O(CH₂)₂ON(CH₃)₂, también conocido como 2'-DMAOE, y 2'-dimetilaminoetoxietoxi (también conocido en la técnica como 2'-O-dimetil-amino-etoxi-etilo o 2'-DMAEOE), es decir, 2'-O-CH₂-O-CH₂-N(CH₃)₂.

Un oligonucleótido modificado puede comprender una o más modificaciones o sustituciones de nucleobase (a menudo mencionada en la técnica simplemente como "base"). Como se usa en este documento, las nucleobases "no modificadas" o "naturales" incluyen las bases de purina adenina (A) y guanina (G), y las bases de pirimidina timina (T), citosina (C) y uracilo (U). Las nucleobases modificadas incluyen otras nucleobases sintéticas y naturales tales como 5-metilcitosina (5-me-C), 5 hidroximetilcitosina, xantina, hipoxantina, 2-aminoadenina, derivados 6-metilo y otros derivados de alquilo de adenina y guanina, derivados 2-propilo y otros derivados de alquilo de adenina y guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina, 5-propinil (-C≡C-CH₃) uracilo y citosina y otros derivados alquilino de bases pirimidina, 6-azo uracilo, citosina y timina, 5-uracilo (pseudouracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tioalquil, 8-hidroxil adenina u otras adeninas y guaninas 8-sustituidas, 5-halo particularmente 5-bromo, 5-trifluorometil uracilos y citosinas y otros uracilos y citosinas 5-sustituidas, 7-metilguanina y 7-metiladenina, 2-F-adenina, 2-amino-adenina, 8-azaguanina y 8-azaadenina, 7-desazaguanina y 7-desazaadenina y 3-desazaguanina y 3 desazaadenina. Nucleobases modificadas adicionales incluyen pirimidinas tricíclicas tales como fenoxazina citidina (1H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), fenotiazina citidina (1H-pirimido(5,4-b)(1,4)benzotiazin-2(3H)-ona), pinzas G tales como fenoxazina citidina (por ejemplo, 9-(2-aminoetoxi)-H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), carbazol citidina (2H-pirimido(4,5-b)indol-2-ona), piridoindol citidina (H-pirido(3',2':4,5)pirrolo(2,3-d)pirimidin-2-ona).

Los restos de base heterocíclica también pueden incluir aquellos en que la base de purina o pirimidina está remplazada con otros heterociclos, por ejemplo, 7-desaza-adenina, 7-desazaguanosina, 2-aminopiridina y 2-piridona.

Una endorribonucleasa específica de secuencia enzimáticamente inactiva adecuada incluye una endorribonucleasa específica de secuencia enzimáticamente inactiva descrita a continuación en este documento. Por ejemplo, puede usarse una endorribonucleasa específica de secuencia enzimáticamente inactiva como se representa en la Figura 6.

En algunas realizaciones, el polirribonucleótido diana se une (de forma covalente o no covalente) a un soporte sólido (un soporte insoluble). Los soportes insolubles adecuados incluyen, aunque sin limitación, perlas, placas (por ejemplo, placas multipocillo), tiras, etc., donde el soporte insoluble puede comprender diversos materiales incluyendo, aunque sin limitación, poliestireno, polipropileno, agarosa y similares.

Las sondas oligonucleotídicas ("oligonucleótido de detección") pueden ser ARN, ADN o cualquier versión modificada químicamente de un ARN o ADN, por ejemplo, ácido peptidnucleico (PNA), ácidos nucleicos cerrados (LNA), y similares.

Un presente método de secuenciación de polirribonucleótido puede incluir una o más etapas de lavado, por ejemplo, para retirar componentes unidos no específicamente tales como sondas oligonucleotídicas unidas no específicamente, cualquier resto detectable unido no específicamente y similares.

Un ejemplo no limitante del modo de realizar un presente método de secuenciación de polirribonucleótido es el siguiente. Un polirribonucleótido diana se une a un soporte sólido. El polirribonucleótido diana es de secuencia desconocida y es el "ARN a secuenciar". Cuatro sondas oligonucleotídicas de cuatro secuencias diferentes conocidas de nucleótidos que comprenden cada una un diferente fluoróforo (fluoróforos 1-4). Los fluoróforos son miembros de pares FRET. Los miembros equivalentes de los pares FRET son puntos cuánticos. El punto cuántico está unido a una endorribonucleasa específica de secuencia enzimáticamente inactiva. La endorribonucleasa específica de secuencia enzimáticamente inactiva se une, pero no escinde, al dúplex formado entre una sonda oligonucleotídica y el polirribonucleótido diana. Solamente una de las cuatro sondas oligonucleotídicas se une a y forma un dúplex con el polirribonucleótido diana. Una etapa de lavado retira cualquier sonda oligonucleotídica no unida. La unión de la sonda oligonucleotídica-fluoróforo2 provoca la formación de dúplex con el polirribonucleótido diana. El fluoróforo2 por tanto se pone en proximidad al punto cuántico unido a la endorribonucleasa específica de secuencia enzimáticamente inactiva, y se inactiva la fluorescencia.

Métodos de escisión de un polirribonucleótido

La presente descripción proporciona un método de escisión de un polirribonucleótido de un modo específico de secuencia. El método generalmente implica poner en contacto un polirribonucleótido sustrato con una endorribonucleasa específica de secuencia enzimáticamente activa (por ejemplo, una endorribonucleasa Csy4) en condiciones que favorecen la escisión específica de secuencia del sustrato polirribonucleotídico. Un presente método de escisión de un polirribonucleótido de un modo específico de secuencia puede usarse para: 1) retirar una marca de afinidad de un polirribonucleótido sustrato; 2) generar una población de polirribonucleótidos producto que

tienen homogeneidad en el extremo 5', por ejemplo, donde los polirribonucleótidos sustrato son ARNm transcritos *in vitro*; y 3) para regular la expresión génica en una célula *in vitro* o *in vivo*.

Polirribonucleótidos sustrato

5 Las expresiones "polirribonucleótidos sustrato" y "polirribonucleótido diana" se usan de forma intercambiable en este documento para hacer referencia a un polirribonucleótido que se une por una endorribonucleasa específica de secuencia de un modo específico de secuencia. Un polirribonucleótido sustrato puede ser monocatenario. En algunos casos, un polirribonucleótido sustrato es bicatenario.

10 Una endorribonucleasa se une a y escinde un polirribonucleótido sustrato de un modo específico de secuencia. Por tanto, por ejemplo, una endorribonucleasa se une a y escinde un polirribonucleótido sustrato en una secuencia específica, mencionada en este documento como "secuencia de reconocimiento" o un "sitio de reconocimiento".

15 Una secuencia de reconocimiento puede ser una secuencia tetranucleotídica, una secuencia pentanucleotídica, una secuencia hexanucleotídica, una secuencia heptanucleotídica, una secuencia octanucleotídica o más larga de un octanucleótido. Por ejemplo, en algunas realizaciones, la secuencia de reconocimiento es de 9 ribonucleótidos, 10 ribonucleótidos, 11 ribonucleótidos, 12 ribonucleótidos, 13 ribonucleótidos, 14 ribonucleótidos, 15 ribonucleótidos, 16 ribonucleótidos, 17 ribonucleótidos, 18 ribonucleótidos, 19 ribonucleótidos o 20 ribonucleótidos de longitud. En algunas realizaciones, una endorribonucleasa específica de secuencia escinde inmediatamente 5' de una secuencia de reconocimiento. En algunas realizaciones, una endorribonucleasa específica de secuencia escinde dentro de una secuencia de reconocimiento. En algunos casos, una secuencia de reconocimiento está inmediatamente 5' de una estructura secundaria. En algunos casos, una secuencia de reconocimiento está localizada 5' de una estructura secundaria y en 1 nucleótido (nt), 2 nt, 3 nt, 4 nt, 5 nt o 5 nt a 10 nt de la estructura secundaria. En algunos casos, una secuencia de reconocimiento está inmediatamente 3' de una estructura secundaria. En algunos casos, una secuencia de reconocimiento está localizada 3' de una estructura secundaria y en 1 nucleótido (nt), 2 nt, 3 nt, 4 nt, 5 nt o 5 nt a 10 nt de la estructura secundaria.

30 En algunas realizaciones, un polirribonucleótido sustrato comprende la estructura $X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}X_{14}X_{15}$, donde los nucleótidos X_1 - X_5 forman pares de bases con X_{11} - X_{15} de modo que X_1 y X_{15} forman la base de una estructura de tronco, y de modo que X_6 , X_7 , X_8 , X_9 , y X_{10} forman un bucle; la estructura es una estructura helicoidal de forma A regular.

35 En algunas realizaciones, el polirribonucleótido sustrato comprende una marca e afinidad; y un presente método proporciona la retirada de la marca de afinidad del polirribonucleótido sustrato.

Endorribonucleasas específicas de secuencia

40 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyen polipéptidos enzimáticamente activos que escinden (hidrolizan) un polirribonucleótido sustrato de un modo independiente de iones metálicos.

45 Las características estructurales de una endorribonucleasa que se une a y escinde un polirribonucleótido sustrato de un modo específico de secuencia e independiente de iones metálicos pueden incluir una o más de las siguientes: 1) una hélice alfa altamente básica para el reconocimiento no específico de secuencia de la estructura fosfato del ARN a través del surco mayor del ARN; por ejemplo, R114, R115, R118, R119 o equivalentes de los mismos; 2) R102 y/o Q104 o equivalentes de los mismos, que hacen contactos de unión de hidrógeno con el surco mayor del tronco de ARN; 3) y uno o más de His29, Ser148, y Tyr176 o equivalentes de los mismos, implicados en la catálisis; y 4) F155 o un equivalente del mismo.

50 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyen polipéptidos enzimáticamente activos que tienen al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la **Figura 4** (secuencias de aminoácidos de Csy4).

60 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyen polipéptidos enzimáticamente activos que tiene al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos expuesta en la **Figura 5** (secuencias de aminoácidos de Csy4).

65 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de

secuencia incluyen polipéptidos enzimáticamente activos que tiene al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos de Cas6.

5 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyen polipéptidos enzimáticamente activos que tiene al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, al menos aproximadamente un 85 %, al menos aproximadamente un 90 %, al menos aproximadamente un 95 %, al menos aproximadamente un 98 %, al menos aproximadamente un 99 % o un 100 % de identidad de secuencia de aminoácidos con una secuencia de aminoácidos de CasE.

10 Las endorribonucleasas que se unen a y escinden un polirribonucleótido sustrato de un modo específico de secuencia incluyen polipéptidos enzimáticamente activos que difieren de una secuencia de aminoácidos expuesta en una cualquiera de las Figuras 4 o 5 en 1 a 20 (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 15 19 o 20) sustituciones y/o inserciones y/o deleciones de aminoácidos.

Condiciones de reacción

20 Una endorribonucleasa específica de secuencia puede hidrolizar un polirribonucleótido sustrato de un modo específico de secuencia a una temperatura en un intervalo de aproximadamente 15 °C a aproximadamente 100 °C, por ejemplo, en un intervalo de aproximadamente 15 °C a aproximadamente 17 °C, de aproximadamente 17 °C a aproximadamente 20 °C, de aproximadamente 20 °C a aproximadamente 25 °C, de aproximadamente 25 °C a aproximadamente 30 °C, de aproximadamente 30 °C a aproximadamente 40 °C, de aproximadamente 40 °C a aproximadamente 50 °C, de aproximadamente 50 °C a aproximadamente 60 °C, de aproximadamente 60 °C a aproximadamente 70 °C, de aproximadamente 70 °C a aproximadamente 80 °C, de aproximadamente 80 °C a aproximadamente 90 °C o de aproximadamente 90 °C a aproximadamente 100 °C.

30 Una endorribonucleasa específica de secuencia puede hidrolizar un polirribonucleótido sustrato de un modo específico de secuencia en un intervalo de pH de aproximadamente 4,0 a aproximadamente 8,0, por ejemplo, de aproximadamente pH 4,0 a aproximadamente 4,5, de aproximadamente pH 4,5 a aproximadamente 5,0, de aproximadamente pH 5,0 a aproximadamente 5,5, de aproximadamente pH 5,5 a aproximadamente 6,0, de aproximadamente pH 6,0 a aproximadamente 6,5, de aproximadamente pH 6,5 a aproximadamente 7,0, de aproximadamente pH 7,0 a aproximadamente 7,5, de aproximadamente pH 6,5 a aproximadamente 7,5, de aproximadamente pH 7,5 a aproximadamente 8,0, de aproximadamente pH 6,5 a aproximadamente 8,0 o de aproximadamente pH 5,5 a aproximadamente 7,5.

Ejemplos

40 Los siguientes ejemplos se exponen para proporcionar a los expertos en la materia una divulgación y descripción completa del modo de preparar y usar la presente invención, y no pretenden limitar el alcance de lo que los inventores consideran su invención ni pretenden representar que los siguientes experimentos son todos o los únicos experimentos realizados. Se han hecho esfuerzos por asegurar la precisión con respecto a los números usados (por ejemplo, cantidades, temperatura, etc.) pero deben contabilizarse algunos errores y desviaciones experimentales. Salvo que se indique lo contrario, las partes son partes en peso, el peso molecular es peso molecular promedio en peso, la temperatura es en grados Celsius y la presión es en o cerca de la atmosférica. Pueden usarse abreviaturas convencionales, por ejemplo, pb, pares de bases; kb, kilobases; pl, picolitros; s, segundos; min, minutos; h, horas; aa, aminoácidos; kb, kilobases; pb, pares de bases; nt, nucleótidos; i.m., intramuscular; i.p., intraperitoneal; s.c., subcutánea; y similares.

50 **Ejemplo 1:** Detección directa de ARN y secuenciación usando proteínas de la familia Csy4

Materiales y métodos

55 Se expresaron Csy4 de tipo silvestre, mutantes puntuales y Csy4 con selenometionina (SeMet) en células Rosetta 2(DE3) como una fusión His6-proteína de unión a maltosa (MBP) o una proteína de fusión a His6 y se purificaron por cromatografía de afinidad por Ni, seguido por retirada proteolítica de la marca His(MBP), una etapa adicional de afinidad por Ni, y cromatografía por exclusión de tamaño. Se transcribieron los pre-ARNcr *in vitro* con polimerasa T7 y se purificaron en un gel desnaturalizante. El complejo se formó incubando ARN con Csy4 una relación 2:1 durante 30 minutos a 30 °C seguido por cromatografía por exclusión de tamaño. El complejo se cristalizó usando el método de gota colgante en citrato sódico 200 mM pH 5,0, cloruro de magnesio 100 mM, poli(etilenglicol) (PEG)-4000 al 20 % (p/v) (complejo de tipo silvestre (WT)) o acetato sódico 150 mM, pH 4,6, PEG4000 al 17 % o acetato sódico 160 mM, pH 4,6, PEG4000 al 18 % (complejo que contiene S22C). La estructura del complejo Csy4 WT-ARN se determinó por el método de dispersión anómalo de múltiples longitudes de onda (MAD) usando cristales SeMet-sustituidos. La estructura del complejo Csy4(S22C)-ARN se determinó por remplazo molecular.

65 **Anotación génica, clonación, expresión proteica y purificación.** El análisis de secuencia comparativa de los

genes de Csy4 entre especies identificó una región conservada 20 codones cadena arriba del codón de inicio anotado en el genoma PA14. Lee, et al. *Genome Biol* 7, R90 (2006). La secuencia de Csy4 conservada (PA14_33300) se amplificó por PCR a partir del ADN genómico de *Pseudomonas aeruginosa* UCBPP-PA14 usando Pa14Csy4_fwd: caccatggaccactacctcgacattcg y Pa14Csy4_rev: gaaccaggaacgaaacctcc. El producto de reacción en cadena de la polimerasa (PCR) se clonó usando el sistema Gateway en el vector de entrada pENTR/TEV/D-TOPO (Invitrogen), seguido por recombinación específica de sitio en el vector de expresión pHGWA o pHMGWA. Busso, et al. *Analytical Biochemistry* 343, 313-321, (2005). Se introdujeron mutaciones puntuales en Csy4 usando el kit de Mutagénesis Dirigida al Sitio QuikChange (Stratagene). El plásmido de expresión Pa14Csy4 se transformó en células *E. coli* Rosetta 2 (DE3) (Novagen) o se co-transformó con el vector pMK que expresa ARN CRISPR sintetizado por Geneart (Regensburg, Alemania). Las células Rosetta 2 (DE3) se cultivaron en Caldo Luria (LB) suplementado con ampicilina y cloranfenicol. La expresión proteica se indujo con β -D-1-tiogalactopiranosido de isopropilo (IPTG) 0,5 mM (Affymetrix) a una densidad celular de \sim 0,50D seguido por agitación a 18 °C durante 16 horas. Las células se sedimentaron y re-suspendieron en tampón de lisis (hidrogenofosfato disódico 15,5 mM, dihidrogenofosfato sódico 4,5 mM, cloruro sódico 500 mM, imidazol 10 mM, inhibidores de proteasa, glicerol al 5 %, Tritón X-100 al 0,01 %, 100 μ l/ml DNaseI, clorhidrato de Tris[2-carboxietil] fosfina (TCEP) 1 mM, fluoruro de fenilmetilsulfonilo 0,5 mM, pH 7,4) y se sonicaron en hielo durante dos minutos en ráfagas de 10 segundos. El lisado se aclaró por centrifugación (24.000 x g, 30 minutos) y se incubó con resina de afinidad por níquel-ácido nitrilotriacético (Ni-NTA) en lotes (Qiagen). La proteína unida se eluyó con tampón de alto contenido de imidazol (dihidrogenofosfato disódico 15,5 mM, dihidrogenofosfato sódico 4,5 mM, cloruro sódico 500 mM, imidazol 300 mM, TCEP 1 mM, glicerol al 5 %, pH 7,4) y se dializó durante una noche en tampón de diálisis (tampón de elución con solamente 20 mM de imidazol) en presencia del virus del grabado del tabaco (TEV) para escindir la marca His₆ o His₆MBP. La proteína se concentró (Amicon) y se purificó en una columna de afinidad por níquel (GE) seguido por columnas en tándem Sup75 (16/60) en tampón de filtración en gel (HEPES 100 mM pH 7,5, KCl 500 mM, glicerol al 5 %, TCEP 1 mM). La muestra después se dializó frente a tampón de filtración en gel que contenía solamente 150 mM de cloruro potásico. Se usó un protocolo similar para la preparación de la proteína derivatizada con selenometionina (SeMet) y la única diferencia notable fue el medio de expresión. En resumen, se cultivaron células BL21(DE3) transformadas con el vector de expresión de Csy4(pH-GWA) en medio mínimo M9 suplementado con ampicilina, como se ha descrito previamente. Wiedenheft, et al. *Structure* 17, 904-912 (2009).

Ensayos de actividad nucleasa. Se incubaron 75 pmol de Csy4 de tipo silvestre o mutante con 5 pmol de Pa14 pre-ARNcr transcrito *in vitro* (preparado como se ha descrito; Wiedenheft (2009) *supra*) en reacciones de 10 μ l que contenían HEPES 20 mM pH 7,5, tampón cloruro potásico 100 mM a 25 °C durante cinco minutos. Las reacciones se interrumpieron con la adición de 50 μ l de fenol ácido-cloroformo (Ambion). Se añadieron 100 μ l de tampón de reacción adicional y las muestras se centrifugaron (16.000 x g 30 minutos) y se retiraron 16 μ l de muestra acuosa, se mezclaron 1:1 con tampón de carga de formamida 2X, y se separaron en gel de poliacrilamida desnaturalizante al 15 %. El ARN se visualizó con tinción con SYBR Gold (Invitrogen).

Cristalización. Todos los experimentos de cristalización se realizaron a 18 °C usando el método de difusión de vapor de gota colgante mezclando volúmenes iguales (1 μ l + 1 μ l) del complejo y las soluciones de reserva. Los cristales con forma de placa del complejo de Csy4 de tipo silvestre-ARN se cultivaron en citrato sódico 200 mM pH 5,0, cloruro de magnesio 100 mM, polietilenglicol-4000 (PEG4000) al 20 % (p/v). Estos cristales pertenecían al grupo espacial C2, contenían una copia del complejo en la unidad asimétrica y difractaban a 2,3 Å de resolución en fuentes de sincrotrón de rayos-X. Usando el complejo reconstituido con el mutante puntual Csy4S22C, se obtuvieron dos formas cristalinas adicionales en acetato sódico 150 mM pH 4,6, PEG4000 al 17 % (p/v) y acetato sódico 160 mM pH 4,6, PEG4000 al 18 %. Inicialmente, aparecieron cristales hexagonales en 24 h. Estos cristales difractaban a 2,6 Å de resolución, pertenecían al grupo espacial *P6*₁ y contenían una copia del complejo en la unidad asimétrica. Después de 48 h, la misma condición de cristalización produjo cristales con forma de aguja que pertenecían al grupo espacial P212121, contenían dos copias del complejo y difractaban hasta 1,8 Å de resolución. Para la recogida de datos, todas las formas cristalinas se crioprotegieron por impregnación en su agua madre respectiva suplementada con glicerol al 30 % para enfriar rápidamente en nitrógeno líquido.

Determinación de la estructura. Todos los datos de difracción se recogieron a 100 K en líneas de luz 8.2.2 y 8.3.1 del Advanced Light Source (Lawrence Berkeley National Laboratory). Los datos se procesaron usando XDS. Kabsch, *Acta Crystallogr D Biol Crystallogr* 66, 125-132 (2010). Las fases experimentales se determinaron a partir de un experimento de dispersión anómala de múltiples longitudes de onda (MAD) de tres longitudes de onda (conjuntos de datos de pico, inflexión y remotos) usando los cristales monoclinicos de Csy4-ARN que contenían Csy4 de tipo silvestre sustituido con selenometionina. Se localizaron dos sitios de selenio usando el módulo Hybrid Substructure Search (HySS) del paquete Phenix. Grosse-Kunstleve, y Adams. *Acta Crystallogr D Biol Crystallogr* 59, 1966-1973 (2003). El refinado de la estructura, el ajuste de fase y la modificación de densidad se realizaron usando AutoSHARP. Vonnrhein, et al. *Methods Mol Biol* 364, 215-230 (2007). El mapa de densidad de electrones resultante mostraba capas claras de densidad atribuibles a la proteína y el ARN alternando a lo largo del eje-c, con la capa de ARN compuesta de dos hélices de ARN apiladas de forma coaxial acopladas en una interacción de "bucle de beso". Se obtuvo un modelo atómico inicial para la proteína Csy4 por construcción automática usando el módulo Phenix AutoBuild. Terwilliger, et al. *Acta Crystallogr D Biol Crystallogr* 64, 61-69, (2008). El modelo del complejo se completó por ciclos iterativos de construcción manual en COOT (Emsley, y Cowtan, *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132 (2004)) y refinamiento usando Phenix.refine³⁶ (Adams, et al. *Acta Crystallogr D Biol Crystallogr* 66, 213-

221 (2010)) frente a un conjunto de datos de resolución de 2,33 Å nativo, produciendo un modelo final con un factor R_{trabajo} cristalográfico del 21,4 % y un factor R_{libre} del 26,4 % (Tabla 1).

Tabla 1 Recogida de datos ajuste de fase y estadística de refinamiento

	WT Nativo	S22C Nativo	S22C Nativo	WT SeMet		
Recogida de datos						
Grupo espacial	C2	$P2_12_12_1$	$P6_1$	C2		
Dimensiones de celda						
a. b. c (Å)	62,37, 72,77, 86,82	40,1, 78,9, 145,9	39,25, 39,25, 297,37	62,33, 47,23, 87,26		
α, β, γ (°)	90,0, 108,2, 90,0	90,0, 90,0, 90,0	90,0, 90,0, 120,0	90,0, 108,3, 90,0		
				<i>Pico</i>	<i>Inflexión</i>	<i>Remoto</i>
Longitud de onda (Å)	1,11159	0,99992	1,11588	0,97949	0,97971	0,97204
	19,68-2,33	69,4-1,80	22,38-2,60	82,86-2,80	82,86-2,80	82,86-2,80
Resolución (Å)*	(2,50-2,33)	(1,90-1,80)	(2,70-2,60)	(2,90-2,80)	(2,90-2,80)	(2,90-2,80)
R_{sim} (%)*	5,8 (44,6)	7,0 (52,8)	3,3 (31,1)	9,4 (38,3)	8,9 (38,5)	9,0 (38,1)
I/σ^*	18,9 (3,35)	31,1 (3,1)	29,8 (3,8)	17,0 (4,4)	14,5(3,7)	14,5(3,6)
Complejidad (%)*	96,6 (98,3)	98,7(91,0)	99,4 (98,5)	99,6 (96,7)	99,5 (99,3)	99,3 (96,4)
Redundancia*	4,4 (4,4)	19,8 (6,5)	6,1 (5,4)	5,7 (5,3)	3,8 (3,7)	3,8 (3,7)
Refinado						
Resolución (Å)	19,70-2,33	69,4-1,80	19,60-2,60			
N.º de reflexiones	9974	43284	7798			
$R_{\text{trabajo}}/R_{\text{libre}}$	0,214/ 0,265	0,187/0,220	0,255/ 0,279			
N.º de átomos						
Proteína	1273	2975	1364			
ARN	313	642	321			
Agua/ligandos	41	386	5			
factores-B						

5

Tabla 1 Recogida de datos ajuste de fase y estadística de refinamiento

	WT Nativo	S22C Nativo	S22C Nativo	WT SeMet
Proteína	47,7	29,1	101,5	
ARN	109,3	35,3	103,0	
Agua/ligandos	44,9	33,5	74,5	
Desviaciones r.m.s.				
Longitudes de enlace (Å)	0,007	0,011	0,002	
Ángulos de enlace (°)	1,0	1,5	0,7	

* Valores entre paréntesis indican el esqueleto de resolución más alta

El modelo incluye los nucleótidos de ARN C1-G15 y el grupo fosfato del nucleótido de C16 y los restos proteicos 1-104, 109-120 y 139-187. Debido a la disposición estratificada de la proteína y el ARN en la red cristalina y la ausencia de contactos cristalinos laterales con la capa de ARN, el ARN muestra un desorden significativo, como se evidencia por los factores de temperatura notablemente elevados ($>100 \text{ \AA}^2$) y la ausencia de densidad interpretable para la base nucleotídica de U9. El desorden también es evidente en los restos proteicos 109-120, correspondientes a la hélice rica en arginina insertada en el surco mayor del ARN, para la cual solamente la estructura polipeptídica podía construirse (excepto para los restos Arg115 y Arg118).

10

15

Las estructuras del complejo Csy4(S22C)-ARN en las formas cristalinas hexagonal y ortorrómbica se determinaron por remplazo molecular en Phaser (McCoy, *et al.* J Appl Crystallogr 40, 658-674 (2007)), usando la proteína Csy4 (que carece de la hélice rica en arginina) y modelos de ARN de la forma cristalina monoclinica como conjuntos de búsqueda diferentes. En ambas formas cristalinas, la densidad de electrones para la hélice rica en arginina y la región enlazadora que comprende los restos de Csy4 105-108 era inmediatamente apreciable en mapas $2F_o - F_c$ obtenidos a partir de soluciones de remplazo molecular. La estructura del complejo Csy4(S22C)-ARN en la forma hexagonal se refinó hasta un factor R_{trabajo} del 25,5 % y R_{libre} de 27,9 a resolución de 2,6 Å. El modelo final incluye los restos de Csy4 1-120 y 139-187 y los nucleótidos de ARN C1-G15 más el grupo fosfato del nucleótido C16. La forma cristalina ortorrómbica del complejo de Csy4(S22C)-ARN se ha resuelto a 1,8 Å de resolución y se refinó hasta un factor R_{trabajo} del 18,7 % y R_{libre} del 22,0 %, con excelente estereoquímica. De los dos complejos en la unidad asimétrica, el complejo 1 (cadenas A y C) contiene los restos de Csy4 1-187 y los nucleótidos de ARN C1-G15 más el grupo fosfato del nucleótido C16, aunque el complejo menos ordenado 2 (cadenas B y D) comprende los restos de Csy4 1-187 con la excepción de los restos 13-15 y 135-138, que no muestran densidad electrónica ordenada, y los nucleótidos de ARN C1-G15 y el grupo fosfato del nucleótido C16. Las dos copias de Csy4 se superponen con una rmsd de 1,15 Å sobre 179 átomos Ca, proviniendo las mayores diferencias de las posiciones ligeramente

20

25

30

diferentes de la hélice rica en arginina. Las dos moléculas de ARN en la unidad asimétrica se superponen con una rmsd de 1,49 Å, debiéndose la mayor desviación al nucleótido que sobresale U9, que asume diferentes conformaciones en los dos ARN. Nuestro análisis e ilustraciones en todo el manuscrito se basan en el complejo 1 de la forma cristalina ortorrómbica. Todas las ilustraciones estructurales se generaron usando Pymol ([http://www\(dot\)pymol\(dot\)org](http://www(dot)pymol(dot)org)).

Resultados

Se cree que la inmunidad mediada por CRISPR sucede en aproximadamente el 90 % de genomas de arqueobacterias y el 40 % de genomas de bacterias basándose en la presencia de loci CRISPR en genomas secuenciados. Horvath y Barrangou, *Science* 327, 167-170 (2010); Jansen, *et al.* *Molecular Microbiology* 43, 1565-1575 (2002); Sorek, *et al.* *Nat Rev Microbiol* 6, 181-186 (2008); Marraffini, y Sontheimer, *Nat Rev Genet* 11, 181-190 (2010). Las proteínas asociadas a CRISPR (Cas) que pertenecen a los ocho subtipos CRISPR/Cas conocidos son muy divergentes a nivel de estructura primaria, oscureciendo la identificación de homólogos funcionales. Haft, *et al.* *PLoS Comput Biol* 1, e60 (2005); Makarova, *et al.* *Biology Direct* 1, 1-26 (2006). *Pseudomonas aeruginosa* UCBPP-PA14 (a partir de ahora Pa14), un patógeno oportunista Gram-negativo que alberga un sistema CRISPR/Cas del subtipo *Yersinia*, contiene seis genes Cas flanqueados por dos elementos CRISPR (**Fig. 1A**). Aunque Cas1 se encuentra de forma universal entre organismos que contienen CRISPR, y Cas3 es evidente en la mayoría de subtipos Csy 1-4 son únicos para el subtipo *Yersinia*. Ambos elementos CRISPR comprenden una disposición característica de repeticiones de 28 nucleótidos idénticas en ambos CRISPR (salvo por un nucleótido) dispersos con espaciadores únicos de ~32 nucleótidos, algunos de los cuales acoplan con secuencias encontradas en bacteriófagos o plásmidos. Grissa, *et al.* *BMC Bioinformatics* 8, 172 (2007). En muchos organismos se ha demostrado que los loci CRISPR se transcriben como unidades únicas largas y se procesan de forma postranscripcional para producir ARNcr que contienen cada uno una secuencia única flanqueada por secuencias derivadas del elemento de repetición. Brouns *et al.* *Science* 321, 960-964 (2008); Carte, *et al.* *Genes y Development* 22, 3489-3496 (2008); Tang, *et al.* *Proc. Natl. Acad. Sci. USA* 99, 7536-7541 (2002); Lillestol, *et al.* *Archaea* 2, 59-72 (2006); Lillestol, *et al.* *Mol Microbiol* 72, 259-272 (2009); Tang, *et al.* *Molecular Microbiology* 55, 469-481 (2005).

Para identificar la proteína o proteínas responsables de producir ARNcr a partir de los largos transcritos CRISPR (pre-ARNcr) en el subtipo *Yersinia*, se expresó de forma recombinante cada una de las seis proteínas Cas de Pa14, y se ensayaron las proteínas expresadas de forma recombinante para la función endonucleolítica usando un pre-ARNcr transcrito *in vitro*. Basándose en la actividad de procesamiento del pre-ARNcr específica de secuencia, se descubrió que Csy4 es la endorribonucleasa responsable de la biogénesis de ARNcr. Como se observa para el procesamiento de ARNcr dentro de otros dos subtipos de CRISPR/Cas (Brouns *et al.* (2008) *supra*; Carte *et al.* (2008) *supra*), la escisión del transcrito CRISPR es una reacción rápida independiente de iones metálicos. Csy4 escinde el pre-ARNcr dentro del elemento de repetición en la base de una estructura de tallo-bucle predicha, generando ARNcr de ~60 nucleótidos que consisten en una secuencia única de 32 nucleótidos (derivada de fagos) flanqueada en los extremos 5' y 3' por ocho y 20 nucleótidos, respectivamente, de secuencia de repetición (**Fig. 1A**).

Para que Csy4 sea eficaz, se hipotetizó que su mecanismo de reconocimiento de ARN debe ser muy específico para abordar solamente transcritos derivados de CRISPR y no otros ARN celulares que contienen horquillas y/o secuencias relacionadas. Para ensayar esto, se expresó Csy4 en *E. coli* solamente o se coexpresó con un ARN de Pa14 CRISPR. A pesar de su alto punto isoeléctrico (PI=10,2), Csy4 no se asocia con ácidos nucleicos celulares; sin embargo, cuando se coexpresa con un Pa14 CRISPR, la proteína se asocia con un ARNcr (**Fig. 1B, C**). Estas observaciones subrayan la especificidad del reconocimiento de Csy4, conduciéndonos a explorar las interacciones proteína/ARN necesarias para el reconocimiento el sustrato de Csy4 y la escisión. Se realizaron ensayos de unión y actividad de Csy4 *in vitro* usando oligonucleótidos de ARN correspondientes a diferentes regiones de la secuencia de repetición de Pa14 CRISPR de 28 nucleótidos. Usando este enfoque, se identificó un fragmento mínimo de ARN reconocido por Csy4 que consistía en el tallo-bucle derivado de la repetición y un nucleótido cadena abajo. Los ensayos de escisión utilizando este ARN mínimo como sustrato mostraron que la actividad de Csy4 requiere un 2'OH en la ribosa inmediatamente cadena arriba del sitio de escisión. Una 2'-desoxirribosa en esta posición anula completamente la escisión, pero no altera la unión de Csy4.

Para obtener percepciones estructurales en el reconocimiento y escisión de ARNcr, se cocrystalizó Csy4 en complejo con un sustrato mínimo de ARN. Para generar un complejo estable para análisis estructural, se unió Csy4 al sustrato mínimo de ARN de 16 nucleótidos no escindible descrito anteriormente en que el nucleótido que precede el sitio de escisión es un 2'-desoxinucleótido. Se obtuvieron cristales del complejo en tres grupos espaciales únicos, mostrando cada uno diferente compactación cristalina; una contenía Csy4 de tipo silvestre y dos contenían un mutante puntual de Csy4. La estructura cristalina del complejo de Csy4-ARN se resolvió hasta una resolución de 1,8 Å (**Fig. 2A, Tabla 1**), revelando un mecanismo no anticipado por el cual el ARN CRISPR se reconoce y procesa para su uso por la maquinaria de silenciamiento mediada por CRISPR. Csy4 hace contactos específicos de secuencia en el surco mayor del tallo-bucle de la secuencia de repetición de CRISPR y contactos adicionales no específicos de secuencia con la estructura fosfato del tronco de ARN. La mayoría de las interacciones proteína/ARN caracterizadas están mediadas por el surco menor de una hélice de ARN; el reconocimiento del surco mayor de ARN por Csy4 es un mecanismo muy inusual de interacción proteína/ARN.

A nivel de estructura primaria, Csy4 es muy distinto de las otras endorribonucleasas conocidas implicadas en

biogénesis de ARNcr (CasE de *Thermus thermophilus* (Ebihara, *et al.*, Protein Sci 15, 1494-1499 (2006)) y Cas6 de *Pyrococcus furiosus* Carte *et al.*, (2008) supra), compartiendo solamente ~10% de identidad. Las estructuras cristalinas tanto de CasE como de Cas6 indican que estas proteínas adoptan plegamientos tipo ferredoxina en tándem. De forma destacable, Csy4 comparte este plegamiento con estas enzimas; en el complejo Csy4-ARN, el dominio N-terminal (restos 1-94) de Csy4 de hecho adopta un plegamiento tipo ferredoxina. Sin embargo, aunque el dominio C-terminal (restos 95-187) comparte la misma conectividad de estructura secundaria que un plegamiento tipo ferredoxina, su conformación es marcadamente diferente. Sorprendentemente, una hélice rica en arginina (restos 108-120) del dominio de ferredoxina C-terminal putativo se inserta en el surco mayor de ARN en horquilla. Las superposiciones estructurales usando el servidor DALI (Holm, y Sander, J Mol Biol 233, 123-138 (1993)) indican que Csy4 en su conformación de unión a ARN se superpone con CasE y Cas6 con una desviación cuadrática media (rmsd) de 3,8 Å (sobre 111 átomos Ca) y 3,9 Å (sobre 104 átomos Ca), respectivamente. Csy4, CasE y Cas6 podrían ser descendientes de una única endorribonucleasa ancestral que ha divergido de forma notable a nivel de secuencia según coevolucionó con la secuencia de repetición del locus CRISPR, manteniendo al mismo tiempo un plegamiento proteico similar.

El sustrato ARNcr forma una estructura de horquilla predicha para esta subclase de repeticiones de ARNcr (Kunin, *et al.* Genome Biol 8, R61 (2007)), formando pares de bases los nucleótidos 1-5 y 11-15 para producir un tronco helicoidal de forma A regular. El pentabucle GUAUA contiene un par de bases G6-A10 compartido y un nucleótido que sobresale U9, su estructura reminiscente de pentabucle es GNR(N)A encontrados en el tronco-bucle intramolecular de ARN nuclear pequeño U6 de levadura (Huppler, *et al.* Nat Struct Biol 9, 431-435 (2002)) y en el ARN de BoxB del bacteriófago lambda (Legault, *et al.* Cell 93, 289-299 (1998)). En el complejo de Csy4-ARN, el tronco-bucle de ARN cabalga sobre la horquilla β formada por las hebras β 7- β 8 de Csy4, con el par de bases C1-G15 apilándose directamente en la cadena lateral aromática de Phe155 (Fig. 2B). Esto ancla el tronco de ARN y lo orienta al ángulo apropiado para permitir interacciones específicas de secuencia en el surco mayor.

Dos restos en un segmento enlazador que conecta el cuerpo de Csy4 a la hélice rica en arginina, Arg102 y Gln104, hacen contactos de unión de hidrógeno en el surco mayor del tronco de ARN, reconociendo de forma específica de secuencia G15 y A14, respectivamente (Fig. 2B). La interacción Csy4-ARNcr se estabiliza adicionalmente por la inserción de la hélice rica en arginina en el surco mayor de la horquilla de ARN en proximidad del nucleótido que sobresale U9 (Fig. 2C). Las cadenas laterales de Arg114, Arg115, Arg118 y Arg119 contactan con los grupos fosfato de los nucleótidos 2-6. Adicionalmente, la cadena lateral de Arg115 acopla con la base de G6 como única interacción específica de secuencia entre la hélice rica en arginina y la horquilla de ARN. De forma interesante, esta interacción es muy reminiscente del modo en que ciertas proteínas víricas interactúan con el surco mayor de moléculas de ARNcr, por ejemplo, la interacción Tat/Tar en el virus de la inmunodeficiencia humana (VIH)²³ y el complejo lambda-N/boxB en fagos lambdaoides (Cai, *et al.* Nature Structural Biology 5, 203-212 (1998)). En ambos casos, se emplea una hélice α muy básica para el reconocimiento no específico de secuencia con la estructura fosfato del ARN a través de surco mayor de ARN.

Csy4 reconoce el elemento de horquilla de la secuencia de repetición CRISPR y escinde inmediatamente cadena abajo de la misma. La estructura descrita en este ejemplo contiene un ARN que imita el sustrato, que no es competente para la escisión. En el sitio activo, se observó densidad solamente para el grupo fosfato 3' del nucleótido penúltimo, pero no se observó densidad para el azúcar o base terminal, supuestamente debido a la flexibilidad de este nucleótido (Fig. 3A). El fosfato escindible se une en un bolsillo localizado entre el giro β y la horquilla β 7- β 8 en un lado y la hélice α 1 y un bucle rico en glicina, previamente identificado en Cas6 y CasE, en el otro. Tres restos próximos a ese grupo fosfato probablemente participan en la catálisis, His29, Ser148 y Tyr176. Estos restos son invariantes entre 12 secuencias de Csy4 que se identificaron usando una búsqueda BLAST (Altschul, *et al.* Nucleic Acids Research 25, 3389-3402 (1997)) acoplada con la verificación manual de un locus CRISPR adyacente (Grissa, *et al.* BMC Bioinformatics 8, 172 (2007)) (Fig. 4).

La estructura sugiere que varios restos en Csy4 son importantes para mediar el reconocimiento/unión del sustrato y catálisis. Se generaron mutantes puntuales de cada uno de estos restos; se ensayó su actividad de escisión bioquímicamente (Fig. 3B). La mutación de los restos del sitio catalítico putativo His29 o Ser148 anula la actividad de escisión. Sin embargo, la mutación de Tyr176 en fenilalanina no altera la actividad, indicando que Tyr176 puede desempeñar un papel crucial en la orientación de His29, aunque no participa directamente en la catálisis. La mutación de Arg102 en alanina anula la acumulación de ARNcr, mientras que la mutación de Gln104 en alanina no altera significativamente la actividad, lo que sugiere que Arg102 que reconoce el par de bases terminal, es importante para orientar apropiadamente el sustrato de ARN, pero Gln104 no es necesario para la actividad *in vitro*. Parece que Phe155 desempeña un gran papel en orientar apropiadamente el sustrato de ARN, ya que una mutación de alanina en este resto altera gravemente la biogénesis de ARNcr.

La identificación de una serina implicada en la mediación de la escisión del ARN es inesperada. Aunque la mutación de His29 en alanina provoca una Csy4 catalíticamente inactiva, la mutación en lisina restaura parcialmente la actividad, sugiriendo fuertemente que His29 actúa como donante de protones, no inicia la escisión mediante un ataque nucleófilo.

Los CRISPR son la memoria genética de un sistema inmunitario basado en ácidos nucleicos que depende de

pequeños ARN derivados de CRISPR para guiar al sistema inmunitario a secuencias afines asociadas con elementos genéticos invasores. El análisis filogenético de las secuencias de repetición CRISPR ha identificado distintas categorías de CRISPR (Kunin, *et al.* Genome Biol 8, R61 (2007)) que se correlacionan con un conjunto particular de genes Cas. La covariación de los genes Cas con tipos de secuencia específicos de repetición CRISPR sugiere que las repeticiones CRISPR han coevolucionado con los genes Cas que son responsables de la adaptación de CRISPR, la generación de ARNcr y el silenciamiento de elementos genéticos invasores. La estructura descrita aquí detalla un mecanismo de reconocimiento inusual que discrimina sustratos de ARNcr basado en especificidad tanto de secuencia como de estructura, proporcionando una enorme percepción de la capacidad de Csy4 y sus homólogos de distinguir fácilmente el ARN sustrato entre todos los ARN celulares.

Figuras 1A-C. Pa14Csy4 reconoce específicamente solamente su sustrato pre-ARNcr. **a**, Esquema del locus CRISPR/Cas en Pa14. Los seis genes Cas están flanqueados en ambos lados por loci CRISPR. Lo ampliado es un esquema que muestra el tronco-bucle predicho en la repetición directa de 28 nucleótidos (letras negras) separada por secuencias espaciadoras de 32 nucleótidos (azul). Las flechas rojas indican el enlace escindido por Csy4. **b, c** Comparación de proteína (**b**) y contenido de ARN (**c**) después de la expresión de Pa14Csy4 en *E. coli* con (+) y sin (-) un plásmido que contiene un locus Pa14 CRISPR. La Csy4 purificada de ambas preparaciones se separó en dos combinaciones. La mitad se resolvió en SDS-PAGE y se visualizó con tinción con azul de Coomassie; la mitad se extrajo con fenol ácido-cloroformo, se resolvió en UREA-PAGE y se visualizó con SYBR Gold (Invitrogen).

Figuras 2A-C. La estructura cristalina de Csy4 unida al sustrato de ARN. **a**, Vistas frontal y posterior del complejo. Csy4 está coloreada en azul y la estructura de ARN está coloreada en naranja. **b**, Interacciones detalladas entre los restos R102 y Q104 y los nucleótidos A14 y G15. El enlace de hidrógeno se representa con líneas discontinuas. **c**, Interacciones detalladas entre una hélice alfa rica en arginina y la estructura de ARN y G6.

Figuras 3A y 3B. Sitio activo putativo. **a**, Vista detallada del centro catalítico. **b**, Actividad de escisión de Csy4. Se incubó Csy4 de tipo silvestre (WT) y una serie de mutantes puntuales individuales con pre-ARNcr transcrito *in vitro* durante 5 minutos a 25 °C. Los productos se extrajeron con fenol ácido-cloroformo y se resolvieron en UREA-PAGE y se visualizaron por tinción con SYBR Gold.

Ejemplo 2: Secuenciación directa de ARN

Un ARN puede sintetizarse a nivel de molécula individual usando transferencia de energía de resonancia de Förster (FRET). El ARN a secuenciarse se unirá a una superficie sólida a través de su ribosa 3'. El ARN debe estar espaciado lo suficiente de las moléculas adyacentes de ARN sobre la superficie para permitir la detección a nivel de molécula individual. El espaciado se dictamina por métodos limitados por difracción, que dependen de la longitud de onda de la luz emitida. Como alternativa, el espaciado del ARN puede ser más cercano que el límite de difracción, si se usan métodos de imágenes de súper-resolución. En la primera etapa de detección de secuencia, se añade una proteína de la familia Csy4, de actividad de unión a ácido nucleico conocida, al ARN a secuenciarse, junto con una combinación de oligonucleótidos de detección. La proteína Csy4 solamente se unirá al ARN a secuenciarse si uno de los oligonucleótidos de detección puede formar una doble hélice de 4 pares de bases con el ARN a secuenciarse. Además, el nucleótido de detección debe formar pares de bases con 3 nucleótidos adicionales 3' de la secuencia de reconocimiento de 4 pares de bases en el ARN a secuenciarse, para que la proteína Csy4 se una de forma estable. Los oligonucleótidos de detección contendrán una extensión de 3 nucleótidos 3' de la secuencia de reconocimiento de 4-nucleótidos. En la combinación de oligonucleótidos de detección, la extensión de 3 nucleótidos tendrá un nucleótido 5' definido seguido de dos posiciones nucleotídicas aleatorias; o un nucleótido aleatorio en la posición 5' seguido de un nucleótido definido y un nucleótido aleatorio; o 2 nucleótidos aleatorios en el extremo 5', seguido de un nucleótido definido. En cualquiera de estas combinaciones, el nucleótido definido se conoce basándose en la molécula fluorescente unida, cuyo espectro de emisión o excitación está definido por el nucleótido. La proteína Csy4 se unirá a un punto cuántico cuyo espectro de excitación solapa con el espectro de emisión de la molécula fluorescente unida al oligonucleótido de detección. Después de la unión de los oligonucleótidos de detección y Csy4, se retira por lavado el exceso de reactivos. Se detecta un evento de unión positivo solamente si el nucleótido de detección forma una doble hélice de 7 nucleótidos con el ARN a secuenciarse. Si se produce la unión, el complejo ternario resultante del ARN a secuenciarse, el oligonucleótido de detección y la proteína Csy4 puede detectarse por FRET a partir de la molécula fluorescente unida al oligonucleótido de detección para el punto cuántico unido a la proteína Csy4. Después de cada ciclo de unión, la proteína Csy4 y los oligonucleótidos de detección se retirarán de la muestra usando desnaturalización química y/o térmica y lavado. En posteriores etapas de secuenciación, se incubarán otras proteínas Csy4 de diferente especificidad de secuencia y sus correspondientes oligonucleótidos de detección con el ARN a secuenciarse, de una manera similar. Pueden idearse otras variaciones de la extensión de 3-nucleótidos en el oligonucleótido de detección, tales como extensiones de diferentes longitudes, en el extremo 5' o el extremo 3' del oligonucleótido de detección. El oligonucleótido de detección podría ser ARN, ADN o cualquier versión modificada químicamente de estos polímeros, tales como PNA o LNA.

Ejemplo 3: Endorribonucleasa específica de secuencia inducible

Mediante técnicas bioquímicas y estructurales, se han generado mutantes puntuales de Csy4 que carecen de actividad de escisión, reteniendo al mismo tiempo la actividad de unión al sustrato. Un ejemplo es el mutante

Csy4(H29A) descrito anteriormente. El mutante Csy4(H29A) por lo demás catalíticamente inactivo puede reactivarse en presencia de imidazol exógeno. La adición de imidazol entre 150 mM y 300 mM al tampón de reacción es suficiente para estimular la actividad de escisión casi de tipo silvestre. Los resultados se muestran en la **Figura 8**. La **Figura 8** muestra un ensayo de actividad de escisión que representa el rescate de imidazol. Csy4H29A es un mutante catalíticamente inactivo de Csy4 que retiene la capacidad de unirse a su sustrato con una k_d of < 1 nM.

Detalles de reacción: Cada 10 μ l de reacción contiene 5 pmol del sustrato pre-ARNcr transcrito *in vitro*, 100 pmol de Csy4 (WT o H29A, según se indica en la **Figura 8**), HEPES 20 mM pH 7,5, KCl 100 mM e imidazol 150-300 mM, según se indica. Las reacciones se realizaron durante 30 minutos a 25 °C. Los productos se extrajeron con fenol ácido-cloroformo, se separaron en un gel desnaturalizante al 15 % y se visualizaron con SYBR Gold. La caracterización bioquímica de Csy4(H29A) muestra que se une a su sustrato de ARN con < 1 nM de afinidad.

Csy4(H29A) es útil para aplicaciones tanto *in vivo* como *in vitro*, para las que no existe enfoque alternativo actual.

Csy4(H29A) (también mencionado en este documento como Csy4 "inducible"), es útil para purificar un complejo particular de ARN/proteína (RNP) a partir de una mezcla compleja de ARN y RNP (complejos de ARN/proteína). Por ejemplo, los investigadores pueden estar interesados en comprender qué proteínas se unen a un transcrito de ARN particular. Usando este sistema, los investigadores podrían diseñar una construcción de expresión para su ARN de elección que incluiría una marca 5' que consiste en la secuencia diana de Csy4 de tallo-bucle. Los investigadores entonces introducirían por transfección esta construcción de expresión en su tipo celular de elección, conduciendo a la generación de muchos ARN y RNP. Las células entonces se lisarían y el lisado se aplicaría a una columna que contiene Csy4 inducible inmovilizado en perlas de agarosa. Los ARN o RNP que tienen una secuencia diana de Csy4 se unirán. Una etapa posterior de lavado retirará los ARN unidos de forma no específica. Un lavado con imidazol (~300 mM) activará Csy4 inducible, que escindirá la secuencia diana y liberará el ARN/RNP unido. Este método se ilustra esquemáticamente en la Figura 9.

Un método similar podría ser útil para ensamblar RNP *in vitro*. Por ejemplo, podría transcribirse un ARN de elección *in vitro* usando una construcción similar al plásmido de expresión diseñado para el experimento anterior. (La construcción debe introducir la secuencia diana de tallo-bucle de Csy4 en el extremo 5' del ARN transcrito.) Este producto transcrito *in vitro* podría entonces incubarse con proteínas que se sabe que o se sospecha que se unen al transcrito particular. La columna que contiene Csy4 inducible podría usarse para purificar los RNP formados *in vitro* de la proteína libre.

Ejemplo 4

Se ha determinado el mecanismo para el reconocimiento de sustrato específico por la endorribonucleasa CasE, un componente esencial del sistema inmunitario CRISPR encontrado en la mayoría de bacterias y arqueobacterias (van der Oost et al., Trends Biochem Sci. 34, 401-7 (2009)). Usando métodos estructurales y bioquímicos, se identificó la secuencia de ARN mínima requerida para la escisión óptima del sustrato, una secuencia de 20 nucleótidos (5-24 de la secuencia de repetición CRISPR) que incluye un tallo-bucle de siete pares de bases seguido de dos nucleótidos desapareados. La estructura de este ARN unido a CasE de *Thermus thermophilus* se resolvió a 2,0 Å de resolución usando cristalografía de rayos X. Esta estructura revela numerosos contactos específicos de secuencia entre la proteína y el ARN, incluyendo varias interacciones en el surco mayor del ARN. El par de bases terminal en el tallo-bucle está alterado, con A22 invertido hacia fuera de la hélice y apilado en la base con U23. Esta conformación está parcialmente estabilizada por las interacciones con S34 y E38, que también confieren especificidad de secuencia para el reconocimiento de sustrato. La estabilización adicional de la formación de A22 y U23 se consigue mediante el posicionamiento del nucleótido terminal, G24, que se invierte de nuevo en registro con el tallo-bucle, pero reside muy por debajo de la hélice en un bolsillo de unión compuesto por los restos D18, E24, y K31, estando en contacto R27 con la estructura entre U23 y G24.

El posicionamiento de A22 alarga la estructura del ARN en el fosfato escindible, estirándolo entre dos restos del sitio activo, Y23 y H26. Basado en esta observación y la estabilización aparente de esta conformación de ARN por la unión de G24, se hipotetizó que G24 puede ser necesario para posicionar el ARN en una conformación catalítica. Coherente con esta hipótesis, la delección o mutación de G24 reduce significativamente la actividad de escisión, también las mutaciones de los restos proteicos implicados en la unión de G24. Para confirmar el papel de G24 en la inducción de la conformación catalítica del ARN, se determinó la estructura de CasE unida a un ARN de 19 nucleótidos que carecía del resto G24 terminal. Este complejo cristalizó en dos formas diferentes, que revelaron dos conformaciones diferentes de ARN en el sitio activo de la proteína. En una forma cristalina ($P2_1$), la estructura de 2,5 Å contenía 8 moléculas en la unidad asimétrica. Las 8 moléculas revelaron que la base A22 se apila con G21, manteniendo la geometría de forma A con el resto del tallo-bucle. Además de los cambios en la estructura del ARN, la estructura proteica también difiere de la conformación catalítica observada en la estructura de 2,0 Å. En esa estructura, un bucle que contenía R158 y K160 se yuxtapone con el sitio activo, lo que sugiere que estos restos pueden desempeñar un papel en la catálisis o en la estabilización del intermedio en estado de transición. En la estructura de 2,5 Å, este bucle está distal del sitio activo y parcialmente desordenado, lo que sugiere que el posicionamiento del bucle es flexible. De forma interesante, este bucle también está desordenado en la estructura apo de CasE (Ebihara et al., Protein Sci. 15, 1494-9 (2006)), lo que sugiere que la conformación correcta del ARN es

necesaria para la estabilización de este bucle.

La segunda forma cristalina ($P2_12_12_1$) obtenida para el complejo de CasE/ARN de 19 nucleótidos se usó para determinar una estructura de 1,5 Å, que reveló que el ARN unido en la conformación catalítica con A22 y U23 se invertía hacia fuera de la hélice. Sin embargo, el bucle que contiene R158 y K160 permanece desordenado en esta estructura, lo que sugiere que la unión de G24 también puede ser necesaria para la estabilización de esta estructura proteica. La observación de dos conformaciones diferentes de ARN para el mismo complejo sugiere que el ARN puede mostrar varios estados estructurales, y que puede requerir que G24 lo bloquee en la conformación catalíticamente competente.

Listado de secuencias

<110> The Regents of the University of California
Haurwitz, Rachel E.
Doudna, Jennifer A.
Wiedenheft, Blake
Jinek, Martin

<120> Composiciones de endorribonucleasa y métodos de uso de las mismas

<130> BERK-118WO

<150> US 61/413.287
<151> 12-11-2010

<150> US 61/365.627
<151> 19-07-2010

<150> US 61/333.163
<151> 10-05-2010

<160> 103

<170> FastSEQ para Windows Versión 4.0

<210> 1
<211> 28
<212> ARN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 1
guucacugcc guauaggcag cuaagaaa 28

<210> 2
<211> 204
<212> PRT
<213> *Moritella* sp. PE36

<400> 2

```

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

      165      170      175
Tyr Ala Ala Asp Phe Leu Asp Tyr Gly Phe Ala Thr Asn Glu Gln His
      180      185      190
Arg Gly Thr Val Pro Lys Leu Ser Ser Leu Met Lys
      195      200

```

<210> 3
 <211> 192
 <212> PRT
 <213> *Moritella* sp. PE36

5

<400> 3

```

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

```

10

ES 2 590 343 T3

<210> 4
 <211> 224
 <212> PRT
 <213> *Shewanella* sp. W3-18-1

5
 <400> 4

```

Met Lys Tyr Tyr Leu Asp Ile Thr Leu Leu Pro Asp Ala Glu Ala Asn
 1   5   5   5   5   5   5   5   5   5   5   5   5   5   5
Leu Gly Phe Leu Trp His Lys Val Tyr Gln Gln Ile His Leu Met Leu
 20  20  20  20  20  20  20  20  20  20  20  20  20  20  20
Val Glu His Lys Val Ser Val Glu Asn Ser Ala Ile Gly Leu Ser Phe
 35  35  35  35  35  35  35  35  35  35  35  35  35  35  35
Pro Lys Tyr Asp Ala Lys Ser Phe Ser Asp Asn Thr Lys Phe Pro Leu
 50  50  50  50  50  50  50  50  50  50  50  50  50  50  50
Gly Asp Lys Leu Arg Leu Phe Ala Gly Thr Glu Gln Gln Leu Ala Asp
 65  65  65  65  65  65  65  65  65  65  65  65  65  65  65
Leu Lys Val Ala Gln Trp Leu Ala Arg Leu Ala Asp Tyr Val His Ile
 85  85  85  85  85  85  85  85  85  85  85  85  85  85  85
Lys Ala Ile Lys Ala Val Pro Asp Asn Val Ser Glu Tyr Ala Tyr Phe
100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
Lys Arg Arg His Phe Lys Ser Pro Asp Lys Leu Arg Arg Asn Ile Asp
115 115 115 115 115 115 115 115 115 115 115 115 115 115 115
Ala Arg Ala Ile Val Ile Ala Gln Lys Asn Gly Phe Ala Ile Asn Glu
130 130 130 130 130 130 130 130 130 130 130 130 130 130 130
Val Lys Thr Arg Leu Leu Ala Ser Ile Asp Asn Leu Asp Thr Lys Ser
145 145 145 145 145 145 145 145 145 145 145 145 145 145 145
Lys Leu Pro Phe Ile Asn Leu Arg Ser Leu Ser Thr Glu Lys Asp Val
165 165 165 165 165 165 165 165 165 165 165 165 165 165 165
Ser Pro Ala Asp Arg Arg Lys Phe Leu Leu Phe Ile Glu Cys Glu Lys
180 180 180 180 180 180 180 180 180 180 180 180 180 180 180
Val Thr Lys Pro Ser Gln Asn Asn Gly Leu Phe Asn Cys Tyr Gly Leu
195 195 195 195 195 195 195 195 195 195 195 195 195 195 195
Ser Arg Arg Ala Gln Thr Glu Gln Ala Ala Val Pro Trp Phe Glu Gly
210 210 210 210 210 210 210 210 210 210 210 210 210 210 210
    
```

10
 <210> 5
 <211> 167
 <212> PRT
 <213> *Pseudomonas aeruginosa*

15
 <400> 5


```

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

```

<210> 6
 <211> 187
 <212> PRT
 <213> *Pseudomonas aeruginosa*

5

<400> 6

```

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
 1          5          10          15
Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala Leu
 20          25          30
Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu Asp
 35          40          45
Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
 50          55          60
Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg
 65          70          75          80
Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr Pro
 85          90          95
Tyr Arg Gln Val Ser Arg Val Gln Val Lys Ser Asn Pro Glu Arg Leu
 100          105          110
Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Ala Arg
 115          120          125
Lys Arg Ile Pro Asp Thr Val Ala Arg Ala Leu Asp Leu Pro Phe Val
 130          135          140
Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg

 145          150          155          160
His Gly Pro Leu Gln Val Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
 165          170          175
Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe
 180          185

```

10

<210> 7
 <211> 28
 <212> ARN
 <213> Secuencia artificial

15

ES 2 590 343 T3

<220>

<223> Oligonucleótido sintético

<400> 7

5 guucacugcc guguaggcag cuaagaaa 28

<210> 8

<211> 187

<212> PRT

10 <213> *Pseudomonas aeruginosa*

<400> 8

```

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

15

<210> 9

<211> 187

<212> PRT

<213> *Pseudomonas aeruginosa*

20

<400> 9

ES 2 590 343 T3

```

Met Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe Pro
1      5      10      15
Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala Leu
      20      25      30
Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu Asp
      35      40      45
Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser Ala
50      55      60
Asp Asp Leu His Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu Arg

65      70      75      80
Asp His Leu Gln Phe Gly Glu Ala Ala Val Val Pro His Pro Thr Pro
      85      90      95
Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg Leu
      100      105      110
Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala Arg
115      120      125
Lys Arg Ile Pro Asp Thr Val Ala Arg Thr Leu Asp Leu Pro Phe Val
130      135      140
Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile Arg
145      150      155
His Gly Pro Leu Gln Ala Thr Ala Glu Glu Gly Gly Phe Thr Cys Tyr
      165      170      175
Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe
      180      185

```

<210> 10
 <211> 184
 5 <212> PRT
 <213> *Dickeya dadantii*

<400> 10

```

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

```

10
 <210> 11
 <211> 28
 <212> ARN

<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

5 <400> 11
guucacugcc gcguaggcag cuuagaaa 28

10 <210> 12
<211> 28
<212> ARN
<213> Secuencia artificial

15 <220>
<223> Oligonucleótido sintético

<400> 12
guucacugcc gaguaggcag cuuagaaa 28

20 <210> 13
<211> 184
<212> PRT
<213> *Pectobacterium wasabiae*

25 <400> 13

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

30 <210> 14
<211> 28
<212> ARN
<213> Secuencia artificial

35 <220>
<223> Oligonucleótido sintético

<400> 14
guucacugcc guauaggcag cuuagaaa 28

40 <210> 15
<211> 184

ES 2 590 343 T3

<212> PRT
 <213> *Photorhabdus asymbiotica*

<400> 15

5

```

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

<210> 16
 <211> 29
 <212> ARN
 <213> Secuencia artificial

10

<220>
 <223> Oligonucleótido sintético

15

<400> 16
 guucacuguc guacaggcag cuuagaaaa 29

20

<210> 17
 <211> 28
 <212> ARN
 <213> Secuencia artificial

25

<220>
 <223> Oligonucleótido sintético

<400> 17
 gugcacugcc guacaggcag cuuagaaa 28

30

<210> 18
 <211> 24
 <212> ARN
 <213> Secuencia artificial

35

<220>
 <223> Oligonucleótido sintético

<400> 18
 acugccguac aggcaguuuu gaaa 24

40

<210> 19
 <211> 28
 <212> ARN
 <213> Secuencia artificial
 5
 <220>
 <223> Oligonucleótido sintético
 <400> 19
 10 guucacugcc gcacaggcag cuuagaaa 28
 <210> 20
 <211> 28
 <212> ARN
 15 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido sintético
 20 <400> 20
 guguacugcc guacaggcag cuuagaaa 28
 <210> 21
 <211> 28
 25 <212> ARN
 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido sintético
 30 <400> 21
 guucacugcc guacaggcag cuuagaaa 28
 <210> 22
 <211> 184
 <212> PRT
 <213> *Dickeya dadantii*
 35
 <400> 22
 40

ES 2 590 343 T3

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

<210> 23
 <211> 28
 5 <212> ARN
 <213> Secuencia artificial

 <220>
 <223> Oligonucleótido sintético
 10
 <400> 23
 guucacugcc guguaggcag cuuagaaa 28

 <210> 24
 <211> 188
 <212> PRT
 <213> *Xanthomonas albilineans*
 15
 <400> 24
 20

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

- 5 <210> 25
- <211> 28
- <212> ARN
- <213> Secuencia artificial

- 10 <220>
- <223> Oligonucleótido sintético

- <400> 25
- guucacugcc guguaggcag cucagaaa 28

- 15 <210> 26
- <211> 27
- <212> ARN
- <213> Secuencia artificial

- 20 <220>
- <223> Oligonucleótido sintético

- <400> 26
- uucacugccg uguaggcagc ucagaaa 27

- 25 <210> 27
- <211> 28
- <212> ARN
- <213> Secuencia artificial

- 30 <220>
- <223> Oligonucleótido sintético

- <400> 27
- guucacugcc guauaggcag cucagaaa 28

- 35 <210> 28
- <211> 184
- <212> PRT
- <213> *Pectobacterium atrosepticum*

- 40

<400> 28

```

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

5 <210> 29
 <211> 184
 <212> PRT
 <213> *Photorhabdus luminescens*

10 <400> 29

```

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

15 <210> 30
 <211> 188
 <212> PRT

ES 2 590 343 T3

<213> *DesulfuriVibrio alkaliphilus*

<400> 30

```

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

```

5

<210> 31

<211> 28

<212> ARN

10 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

15 <400> 31

guucacugcc gcacaggcag cucagaaa 28

<210> 32

<211> 184

20 <212> PRT

<213> *Dickeya zeae*

<400> 32

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

Ala Thr Ala Thr Ile Pro Trp Phe
 180

5 <210> 33
 <211> 28
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 33
 gugcacugcc guauaggcag cuuagaaa 28

15 <210> 34
 <211> 184
 <212> PRT
 <213> *Yersinia pestis*

20 <400> 34

ES 2 590 343 T3

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

5 <210> 35
 <211> 30
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 35
 uguucacugc cgcacaggca gcuuagaaaa 30

15 <210> 36
 <211> 184
 <212> PRT
 <213> *Dickeya dadantii*

20 <400> 36

Met Asp His Tyr Ile Glu Ile Arg Val Leu Pro Asp Pro Glu Phe Ser

ES 2 590 343 T3

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

<210> 37

<211> 28

5 <212> ARN

<213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

10

<400> 37

gugaacugcc gcauaggcag cuagaaa 28

<210> 38

15 <211> 198

<212> PRT

<213> *Escherichia coli*

<400> 38

20

```

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

          180          185          190
Ala Thr Ile Pro Trp Phe
          195

```

<210> 39
 <211> 184
 <212> PRT
 <213> *Escherichia coli*

5

<400> 39

```

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

```

10

<210> 40

ES 2 590 343 T3

<211> 184
 <212> PRT
 <213> *Tolomonas auensis*

5 <400> 40

```

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

Ser Ile Ala Thr Ile Pro Trp Phe
 180
    
```

10 <210> 41
 <211> 28
 <212> ARN
 <213> Secuencia artificial

15 <220>
 <223> Oligonucleótido sintético

<400> 41
 cuucacugcc gcacaggcag cuuagaaa 28

20 <210> 42
 <211> 184
 <212> PRT
 <213> *Erwinia pyrifoliae*

25 <400> 42

ES 2 590 343 T3

```

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

```

5 <210> 43
 <211> 28
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 43
 uucacugccg uacaggcagc uuagaaaa 28

15 <210> 44
 <211> 198
 <212> PRT
 <213> *Escherichia coli*

20 <400> 44

Met Ala Val Ser Leu Val Arg Asn Arg Asn Lys Glu Leu Pro Met Asp

ES 2 590 343 T3

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

<210> 45
 <211> 190
 <212> PRT
 <213> *Verminephrobacter eiseniae*

5

<400> 45

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

10

<210> 46
 <211> 28
 <212> ARN
 <213> Secuencia artificial

15

ES 2 590 343 T3

<220>

<223> Oligonucleótido sintético

5 <400> 46
guucacugcc ggauaggcag cucagaaa 28

<210> 47

<211> 195

10 <212> PRT

<213> *Chromobacterium violaceum*

<400> 47

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

15

<210> 48

<211> 28

<212> ARN

20 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

25 <400> 48
guucacugcc ggauaggcag cuuagaaa 28

<210> 49

<211> 184

30 <212> PRT

<213> *Erwinia tasmaniensis*

<400> 49

ES 2 590 343 T3

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

5 <210> 50
 <211> 28
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 50
 guucacugcc guacaggcag cuuagaag 28

15 <210> 51
 <211> 190
 <212> PRT
 <213> *Delftia acidovorans*

20 <400> 51

ES 2 590 343 T3

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

<210> 52

<211> 28

5 <212> ARN

<213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

10

<400> 52

guucgcugcc gcguaggcag cucagaaa 28

<210> 53

<211> 184

<212> PRT

<213> *Enterobacter* sp. 638

15

<400> 53

20

ES 2 590 343 T3

```

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

```

<210> 54
 <211> 186
 5 <212> PRT
 <213> *ThioalkaliVibrio* sp. K90mix
 <400> 54

```

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

```

10
 <210> 55
 <211> 28
 <212> ARN
 15 <213> Secuencia artificial

ES 2 590 343 T3

<220>

<223> Oligonucleótido sintético

5 <400> 55
guuagcugcc gcacaggcag cucagaaa 28

<210> 56

<211> 193

10 <212> PRT

<213> *Zymomonas mobilis*

<400> 56

```

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

```

15

<210> 57

<211> 193

<212> PRT

20 <213> *Zymomonas mobilis*

<400> 57

```

Met Leu Ala Asn Pro Val Asp Ser Tyr Gln Asp Ile Tyr Ile Leu Pro
 1          5          10          15
Asn Gln Glu Ile Ala Pro His Ile Ile Met Glu Lys Leu Phe Ser Leu
 20          25          30
Leu His Leu Glu Leu Val Arg Leu Gly Ser Gln His Ile Gly Ile Ser
 35          40          45
Phe Pro Glu His Asp Asn Asn Lys Pro Cys Leu Gly Ser Arg Leu Arg
 50          55          60
Leu His Gly Ala Gly Ala Asp Leu His Glu Leu Ala Leu Ser Gly Trp

```

ES 2 590 343 T3

```

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

```

<210> 58
 <211> 191
 <212> PRT
 <213> *Acidovorax* sp. JS42

5

<400> 58

```

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

```

10

<210> 59
 <211> 28
 <212> ARN
 <213> Secuencia artificial

15

<220>
 <223> Oligonucleótido sintético

20

<400> 59
 guucacugcc gcuaaggcag cucagaaa 28

<210> 60
 <211> 182

<212> PRT
 <213> *Desulfurispirillum indicum*

<400> 60

5

```

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

<210> 61
 <211> 184
 <212> PRT
 <213> *Photobacterium profundum*

10

<400> 61

```

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

15

ES 2 590 343 T3

<210> 62
 <211> 184
 <212> PRT
 <213> *Legionella pneumophila*

5

<400> 62

```

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

10 <210> 63
 <211> 184
 <212> PRT
 <213> *Legionella pneumophila*

15 <400> 63

ES 2 590 343 T3

```

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

```

<210> 64
 <211> 183
 5 <212> PRT
 <213> *Shewanella putrefaciens*

<400> 64

```

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

```

10
 <210> 65
 <211> 28
 <212> ARN
 15 <213> Secuencia artificial

ES 2 590 343 T3

<220>

<223> Oligonucleótido sintético

<400> 65

5 guucaccgcc gcacaggcgg cuuagaaa 28

<210> 66

<211> 184

<212> PRT

10 <213> *Legionella pneumophila*

<400> 66

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

165

170

175

Thr Ala Thr Val Pro Trp Phe Asp
180

15

<210> 67

<211> 28

<212> ARN

<213> Secuencia artificial

20

<220>

<223> Oligonucleótido sintético

<400> 67

25 guuaacugcc gcacaggcag cuuagaag 28

<210> 68

<211> 183

<212> PRT

30 <213> *Vibrio metschnikovii*

<400> 68

ES 2 590 343 T3

```

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

```

5 <210> 69
 <211> 28
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 69
 guucacugcc gcagaggcag cuuagaaa 28

15 <210> 70
 <211> 102
 <212> PRT
 <213> *Citrobacter koseri*

20 <400> 70

```

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

```

25 <210> 71
 <211> 183
 <212> PRT
 <213> *Vibrio* sp. RC586

<400> 71

```

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

5 <210> 72
 <211> 28
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 72
 aguguucugc cgaaaggca gcuaagaa 28

15 <210> 73
 <211> 184
 <212> PRT
 <213> *Vibrio cholerae*

20 <400> 73

ES 2 590 343 T3

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

5 <210> 74
 <211> 29
 <212> ARN
 <213> Secuencia artificial

10 <220>
 <223> Oligonucleótido sintético

<400> 74
 guucacugcc gcacaggcag cuuagaaau 29

15 <210> 75
 <211> 197
 <212> PRT
 <213> *Oxalobacter formigenes*

20 <400> 75

ES 2 590 343 T3

```

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

Asn Ala Ser Leu Lys Val Phe Thr Thr Tyr Gly Leu Ser Ala Glu Ser
      180      185      190
Thr Ile Pro Glu Phe
      195
  
```

- <210> 76
- <211> 28
- 5 <212> ARN
- <213> Secuencia artificial

- <220>
- <223> Oligonucleótido sintético
- 10
- <400> 76
- guucacugcc guauaggcag cuuagaag 28

- <210> 77
- <211> 193
- <212> PRT
- <213> *Psychromonas ingrahamii*

- <400> 77
- 20

ES 2 590 343 T3

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

<210> 78

<211> 29

5 <212> ARN

<213> Secuencia artificial

<220>

10 <223> Oligonucleótido sintético

<400> 78

guguuucccg ugcccacggg gaugaaccg 29

<210> 79

15 <211> 202

<212> PRT

<213> *Dichelobacter nodosus*

<400> 79

20

ES 2 590 343 T3

```

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

```

<210> 80
 <211> 194
 5 <212> PRT
 <213> *Shewanella baltica*

<400> 80

```

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

```

10

<210> 81
 <211> 194

ES 2 590 343 T3

<212> PRT
 <213> *Shewanella baltica*

<400> 81

5

```

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

<210> 82
 <211> 204
 <212> PRT
 <213> *Acinetobacter baumannii*

10

<400> 82

ES 2 590 343 T3

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

195

200

- <210> 83
- <211> 28
- 5 <212> ARN
- <213> Secuencia artificial

- <220>
- <223> Oligonucleótido sintético
- 10
- <400> 83
- guucauggcg gcacacgcca uuuagaaa 28

- <210> 84
- <211> 203
- 15 <212> PRT
- <213> *Acinetobacter baumannii*

- <400> 84
- 20

ES 2 590 343 T3

```

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

```

<210> 85

<211> 186

5 <212> PRT

<213> *Pasteurella multocida*

<400> 85

```

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

```

10

ES 2 590 343 T3

<210> 86
 <211> 28
 <212> ARN
 <213> Secuencia artificial
 5
 <220>
 <223> Oligonucleótido sintético
 <400> 86
 10 guucaccauc guguagaugg cuuagaaa 28
 <210> 87
 <211> 28
 <212> ARN
 15 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido sintético
 20 <400> 87
 guuaacugcc guauaggcag cuuagaaa 28
 <210> 88
 <211> 187
 25 <212> PRT
 <213> *Aggregatibacter actinomycetemcomitans*
 <400> 88
 Met Thr Val Gln Thr His Tyr Ile Glu Ile Lys Ala Ile Pro Gln Val
 1 5 10 15
 Asp Met Leu Gln Thr Glu Val Ile Gly Phe Cys Leu Gln Lys Leu His
 20 25 30
 Gln Ile Leu Pro His Phe Glu Gly Arg Ile Gly Leu Ala Phe Pro Ala
 35 40 45
 Tyr Gly Asn Asp Lys Thr Leu Gly Gly Ile Ile Arg Leu Phe Gly Thr
 50 55 60
 Glu Asn Asp Cys Gly Phe Ile His Phe Lys Leu Gln Ser Leu Arg Asp
 65 70 75 80
 Tyr Ala Leu Ile Ser Glu Val Met Pro Ile Pro Glu Lys Val Arg Ser
 85 90 95
 Tyr Arg Ile Tyr Gln Arg Ile Gln Pro Lys Gly Gln Ser Ser Ile Arg
 100 105 110
 Arg Ala Glu Lys Arg Leu Thr Ala Gln Gly Lys Trp Asn Glu Glu Val
 115 120 125
 Leu Gln Asn Met Leu Gln Lys Gln Ala Thr Gln Arg Ile Tyr Pro His
 130 135 140
 Ala His Leu Lys Ser Ser Ser Thr Lys Gln Gln Phe Ile Leu Ala Ile
 145 150 155 160
 Lys Ser Val His Gln Thr Lys Ala Val Glu Gly Val Phe Ser Ala Tyr
 165 170 175
 Gly Leu Ser Gln Thr Thr Thr Val Pro His Phe
 180 185
 30
 <210> 89
 <211> 28
 <212> ARN
 35 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido sintético
 40 <400> 89

ES 2 590 343 T3

cuucacugcc gaauaggcag cuuagaaa 28

<210> 90

<211> 199

5 <212> PRT

<213> *Marinomonas* sp. MWYL1

<400> 90

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

10

<210> 91

<211> 28

<212> ARN

15 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

20 <400> 91

guucgccgcc gagcacgagg cuuagaaa 28

<210> 92

<211> 190

25 <212> PRT

<213> *Dialister invisus*

<400> 92

ES 2 590 343 T3

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

- 5 <210> 93
- <211> 28
- <212> ARN
- <213> Secuencia artificial

- 10 <220>
- <223> Oligonucleótido sintético

- <400> 93
- guuaacugcc gcuaagguag uuuagaaa 28

- 15 <210> 94
- <211> 28
- <212> ARN
- <213> Secuencia artificial

- 20 <220>
- <223> Oligonucleótido sintético

- <400> 94
- guuaucugcc guauaggcag cuuagaaa 28

- 25 <210> 95
- <211> 189
- <212> PRT
- <213> *Actinobacillus pleuropneumoniae*

- 30 <400> 95

ES 2 590 343 T3

```

Met Ser Glu Leu Thr His Tyr Ile Glu Leu Lys Ala Ile Pro Gln Val
 1          5          10          15
Asp Ile Leu Gln Thr Asp Val Ile Ala His Gly Leu Gln Ile Leu His
 20          25          30
Lys Phe Leu Pro Leu Tyr Gln Gly Glu Ile Gly Leu Ser Phe Pro Ala
 35          40          45
Tyr Gly Leu Gly Arg Thr Leu Gly Gly Ile Ile Arg Val Phe Gly Asn
 50          55          60
Glu Gln His Cys Thr Gln Ile Lys Thr Gln Leu Ile Gly Glu Gly Leu
 65          70          75          80

Gln Asp Tyr Val Leu Ile Thr Ser Val Thr Pro Val Pro Glu Glu Ile
 85          90          95
Val Glu Tyr His Arg Tyr Gln Arg Val His Arg Lys Gly Gln Ser Ala
 100         105         110
Ile Arg Arg Thr Glu Gln Phe Leu Val Gln Gln Gly Lys Trp Thr Glu
 115         120         125
Glu Ile Arg Gln Glu Met Leu Ile His Gln Gln Asn Gln Lys Val Phe
 130         135         140
Pro Tyr Val Lys Leu Lys Ser Gly Ser Thr Lys Gln His Phe Val Leu
 145         150         155         160
Ala Ile Arg Gln Leu Arg Leu Ala Glu Pro Ala Ser Gly Leu Phe Asn
 165         170         175
Ala Tyr Gly Leu Ser Gln Ala Ala Thr Val Pro His Phe
 180         185

```

<210> 96

<211> 28

5 <212> ARN

<213> Secuencia artificial

<220>

10 <223> Oligonucleótido sintético

<400> 96

cuucacugcc guauaggcag cuuagaaa 28

<210> 97

15 <211> 189

<212> PRT

<213> *Actinobacillus pleuropneumoniae*

<400> 97

20

ES 2 590 343 T3

```

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

```

<210> 98
 <211> 189
 5 <212> PRT
 <213> *Actinobacillus pleuropneumoniae*
 <400> 98

```

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

```

10
 <210> 99
 <211> 189
 <212> PRT
 15 <213> *Actinobacillus pleuropneumoniae*

<400> 99

```

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

```

5 <210> 100
 <211> 201
 <212> PRT
 <213> *Streptococcus thermophilus*

10 <400> 100

```

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

```

ES 2 590 343 T3

<210> 101
 <211> 187
 <212> PRT
 <213> *Pseudomonas aeruginosa*

5

<400> 101

```

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

10

<210> 102
 <211> 187
 <212> PRT
 <213> *Pseudomonas aeruginosa*

15

<400> 102

ES 2 590 343 T3

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

<210> 103

<211> 20

5 <212> ARN

<213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

10

<400> 103

guucacugcc guauaggcag

20

REIVINDICACIONES

- 5 1. Una endorribonucleasa Csy4 variante que comprende una secuencia de aminoácidos que tiene al menos aproximadamente un 95 % de identidad de secuencia de aminoácidos con la secuencia de aminoácidos de las SEQ ID NO: 101 o la SEQ ID NO: 102 expuestas en la Figura 6, en donde la endorribonucleasa comprende una sustitución de aminoácido en His29, en donde la endorribonucleasa Csy4 variante es enzimáticamente inactiva en ausencia de imidazol, y en donde la endorribonucleasa Csy4 variante se puede activar en presencia de imidazol.
- 10 2. La endorribonucleasa Csy4 variante de la reivindicación 1, en la que la sustitución de aminoácido es una sustitución de His29 a Ala29.
- 15 3. La endorribonucleasa Csy4 variante de la reivindicación 1, en donde la endorribonucleasa Csy4 variante comprende un resto que proporciona una señal detectable.
- 20 4. La endorribonucleasa Csy4 variante de la reivindicación 3, en la que dicho resto se selecciona entre un fluoróforo, un punto cuántico, una enzima diferente a la endorribonucleasa o una nanopartícula.
- 25 5. La endorribonucleasa Csy4 variante de la reivindicación 1, en donde la endorribonucleasa está inmovilizada en un soporte insoluble.
- 30 6. La endorribonucleasa Csy4 variante de la reivindicación 5, en la que dicho soporte insoluble es una perla.
- 35 7. La endorribonucleasa Csy4 variante de la reivindicación 1, en donde la endorribonucleasa Csy4 variante se une a un sustrato de ARN que comprende la secuencia de nucleótidos 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1).
- 40 8. Un ácido nucleico que comprende una secuencia de nucleótidos que codifica la endorribonucleasa Csy4 variante de la reivindicación 1.
- 45 9. Un vector de expresión recombinante que comprende una secuencia de nucleótidos que codifica la endorribonucleasa Csy4 variante de la reivindicación 1.
- 50 10. El vector de expresión recombinante de la reivindicación 9, en el que la secuencia de nucleótidos que codifica la endorribonucleasa Csy4 variante está unida de forma funcional a un promotor.
- 55 11. El vector de expresión recombinante de la reivindicación 10, en el que dicho promotor es un promotor inducible.
- 60 12. Una célula hospedadora modificada genéticamente *in vitro* que comprende el vector de expresión recombinante de la reivindicación 9.
- 65 13. Un kit para purificar un ARN diana presente en una población mixta de ácidos nucleicos, comprendiendo el kit:
la endorribonucleasa Csy4 variante de la reivindicación 1.
- 70 14. El kit de la reivindicación 13, en el que la endorribonucleasa está inmovilizada en un soporte insoluble.
- 75 15. El kit de las reivindicaciones 13 o 14, que comprende adicionalmente un vector de expresión recombinante que comprende, en orden de 5' a 3' y en unión funcional:
a) una secuencia de nucleótidos que codifica un sustrato de ARN que está unido específicamente por la endorribonucleasa Csy4 variante de la reivindicación 1;
b) un sitio de clonación múltiple adecuado para la inserción de un ácido nucleico que codifica el ARN diana.
- 80 16. El kit de la reivindicación 15, en el que la secuencia de nucleótidos que codifica el sustrato de ARN está unida de forma funcional a un promotor.
- 85 17. El kit de la reivindicación 16, en el que el promotor es un promotor inducible.
- 90 18. El kit de la reivindicación 15, en el que el sustrato de ARN comprende la secuencia de nucleótidos 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO: 1).
- 95 19. El kit de la reivindicación 15, en el que el vector de expresión recombinante comprende, insertada en el sitio de clonación múltiple, una secuencia de nucleótidos que codifica el ARN diana.
- 100 20. El kit de la reivindicación 13, que comprende adicionalmente uno o más de (a) imidazol, (b) uno o más tampones de lavado y (c) un vector de expresión de control positivo.

21. Un método de aislamiento de un ARN diana presente en una población mixta de ácidos nucleicos, comprendiendo el método:

- 5 a) poner en contacto una población mixta de ácidos nucleicos con la endorribonucleasa Csy4 variante de la reivindicación 1, en donde la endorribonucleasa Csy4 variante está inmovilizada en un soporte insoluble, en donde la población mixta de ácidos nucleicos comprende un ARN diana marcado que comprende una secuencia de nucleótidos de reconocimiento que está unida específicamente por la endorribonucleasa Csy4 variante inmovilizada, formando un complejo de ARN diana marcado-endorribonucleasa Csy4 variante inmovilizada, en donde dicho contacto se realiza en una solución de unión que carece de imidazol;
- 10 b) añadir imidazol a la solución de unión hasta una concentración final de aproximadamente 100 mM a aproximadamente 500 mM, formando de ese modo una solución de reactivación que reactiva enzimáticamente la endorribonucleasa Csy4 variante inmovilizada, en donde la endorribonucleasa Csy4 variante inmovilizada reactivada escinde el ARN diana de la marca; y
- 15 c) recoger el ARN diana liberado.

22. El método de la reivindicación 21, que comprende adicionalmente una etapa de lavado realizada después de la etapa (a) y antes de la etapa (b).

23. Un método de aislamiento de un polipéptido que se une a un ARN diana, comprendiendo el método:

- 20 a) poner en contacto un complejo inmovilizado con una solución líquida que comprende un polipéptido que se une al ARN diana, en donde el complejo inmovilizado comprende la endorribonucleasa Csy4 variante de la reivindicación 1 y un ARN diana marcado que comprende una secuencia de nucleótidos de reconocimiento que está unida específicamente por la endorribonucleasa Csy4 variante, donde dicho contacto da como resultado la
- 25 unión del polipéptido al ARN diana, en el que dicho contacto se realiza en una solución de unión que carece de imidazol; y
- b) eluir el polipéptido unido.

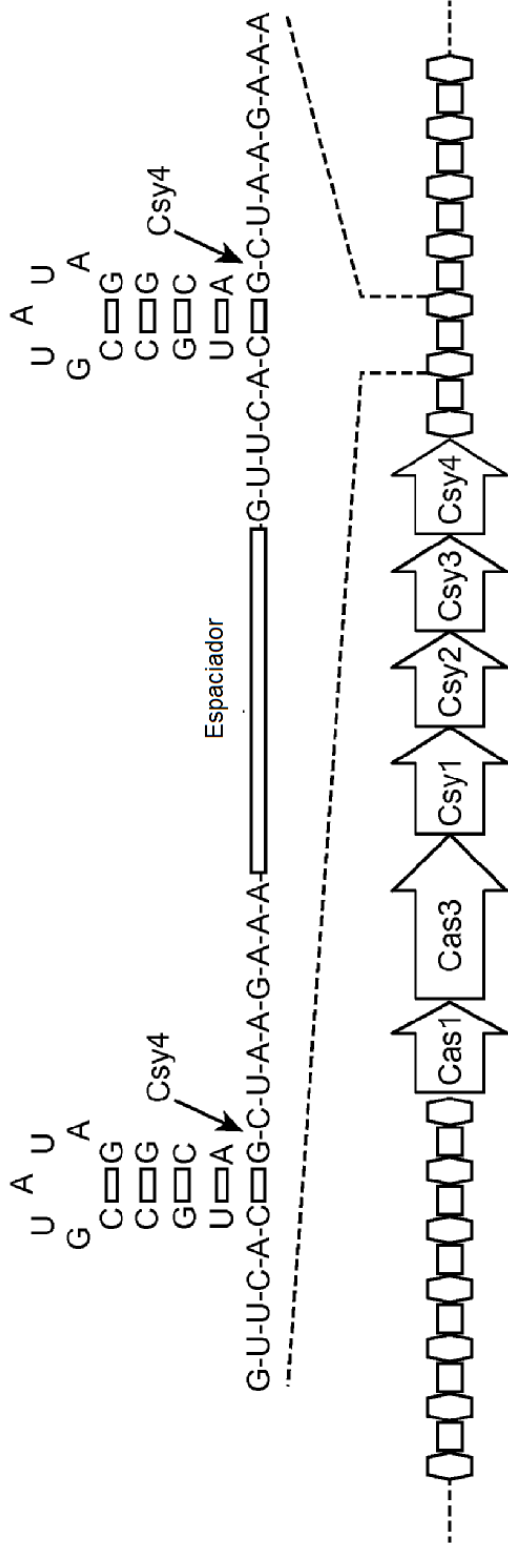


FIG. 1A

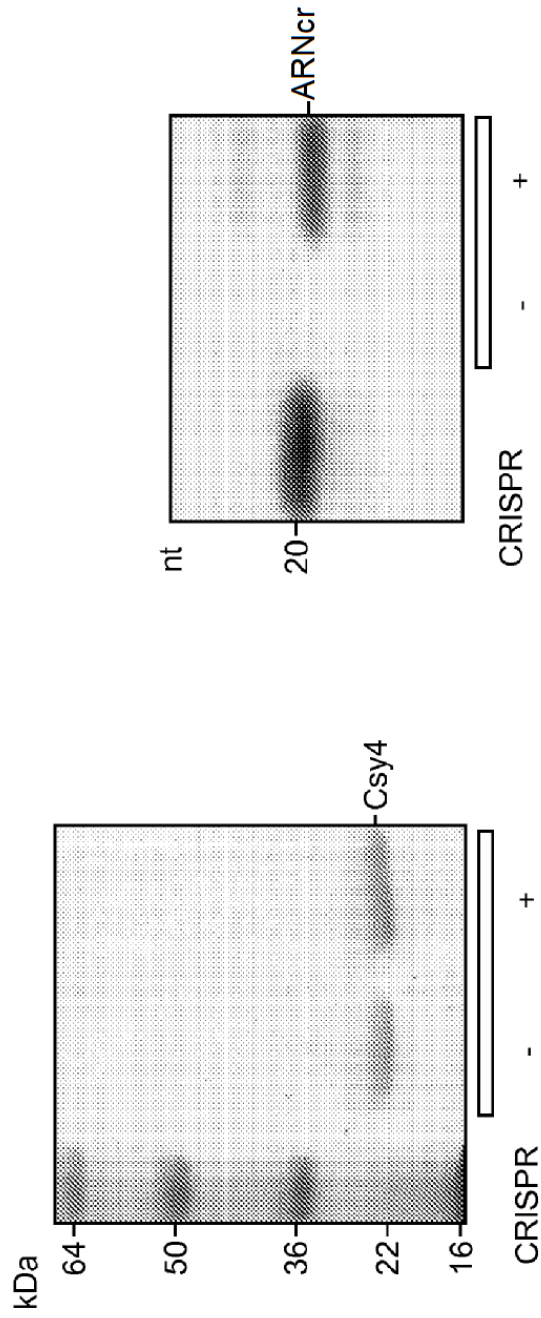


FIG. 1B

FIG. 1C

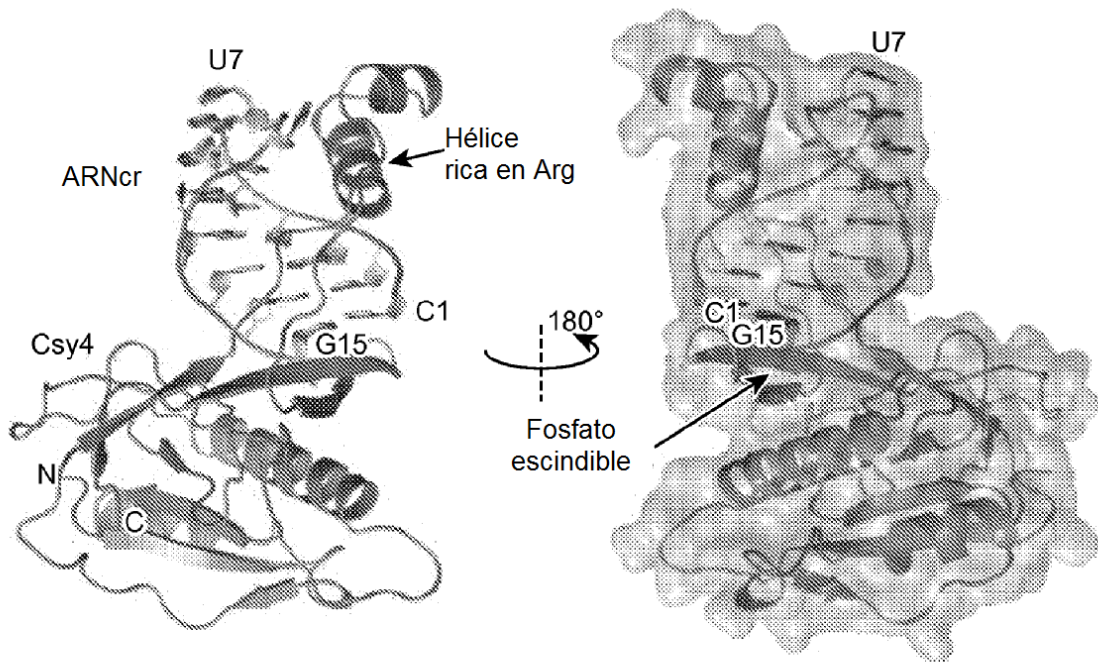


FIG. 2A

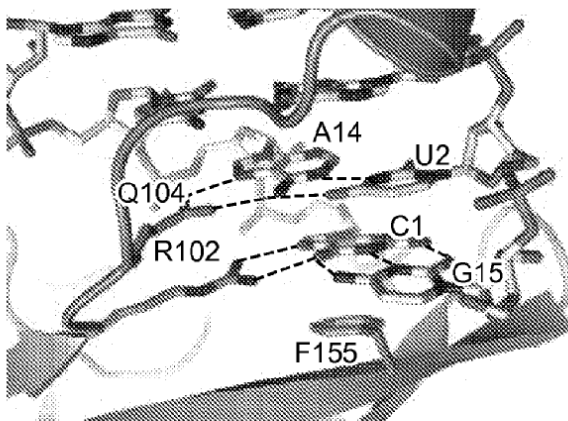


FIG. 2B

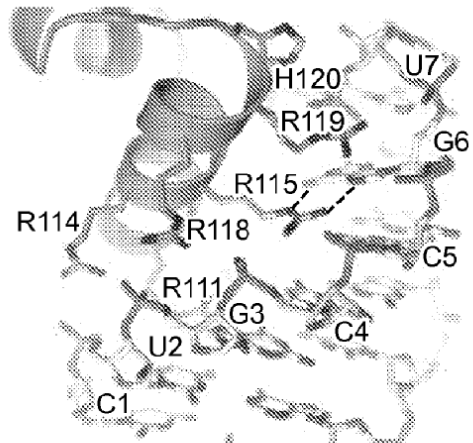
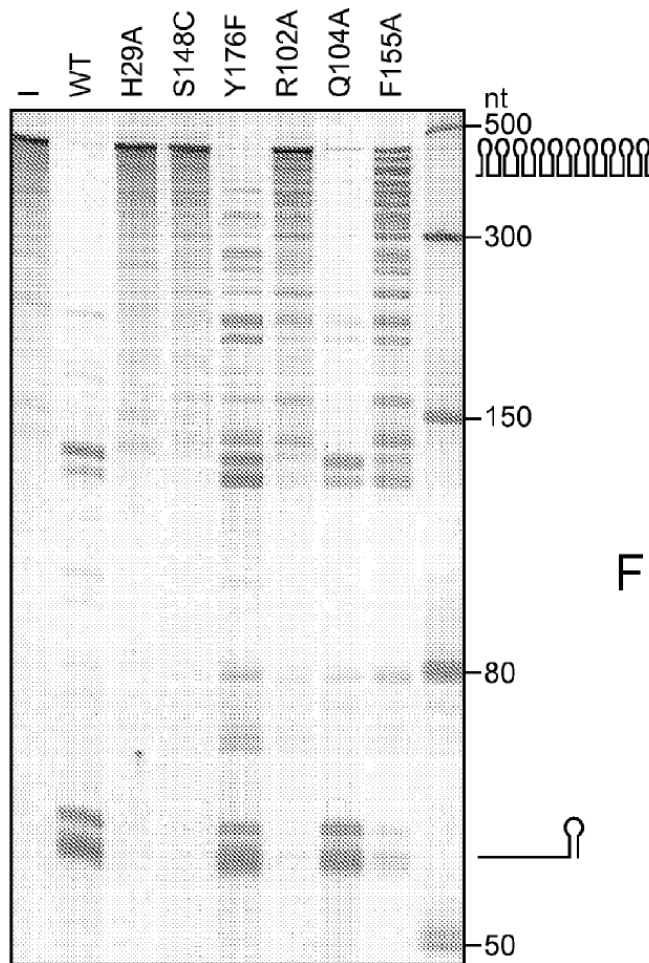
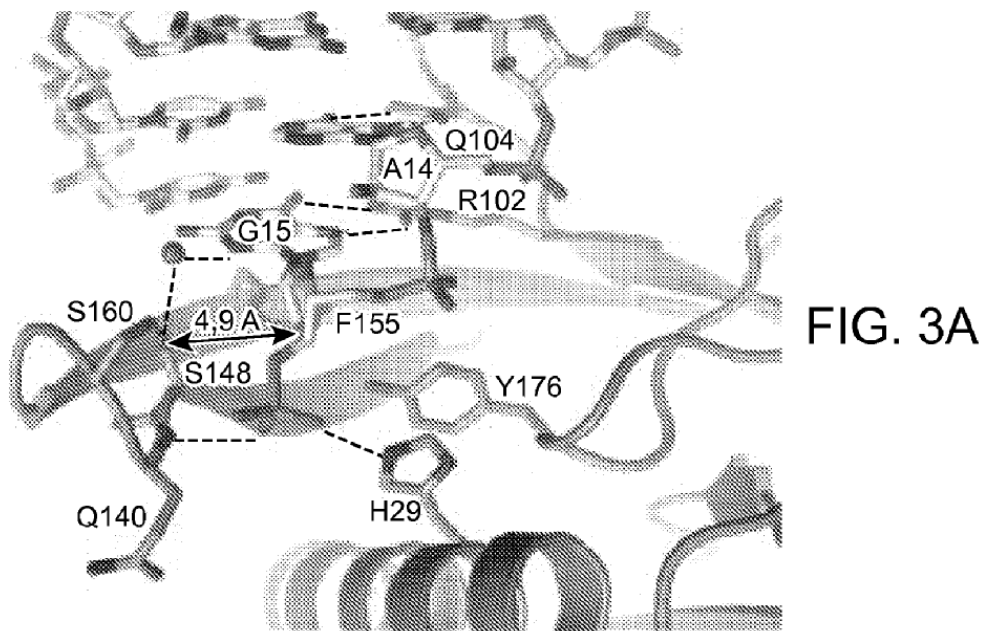


FIG. 2C



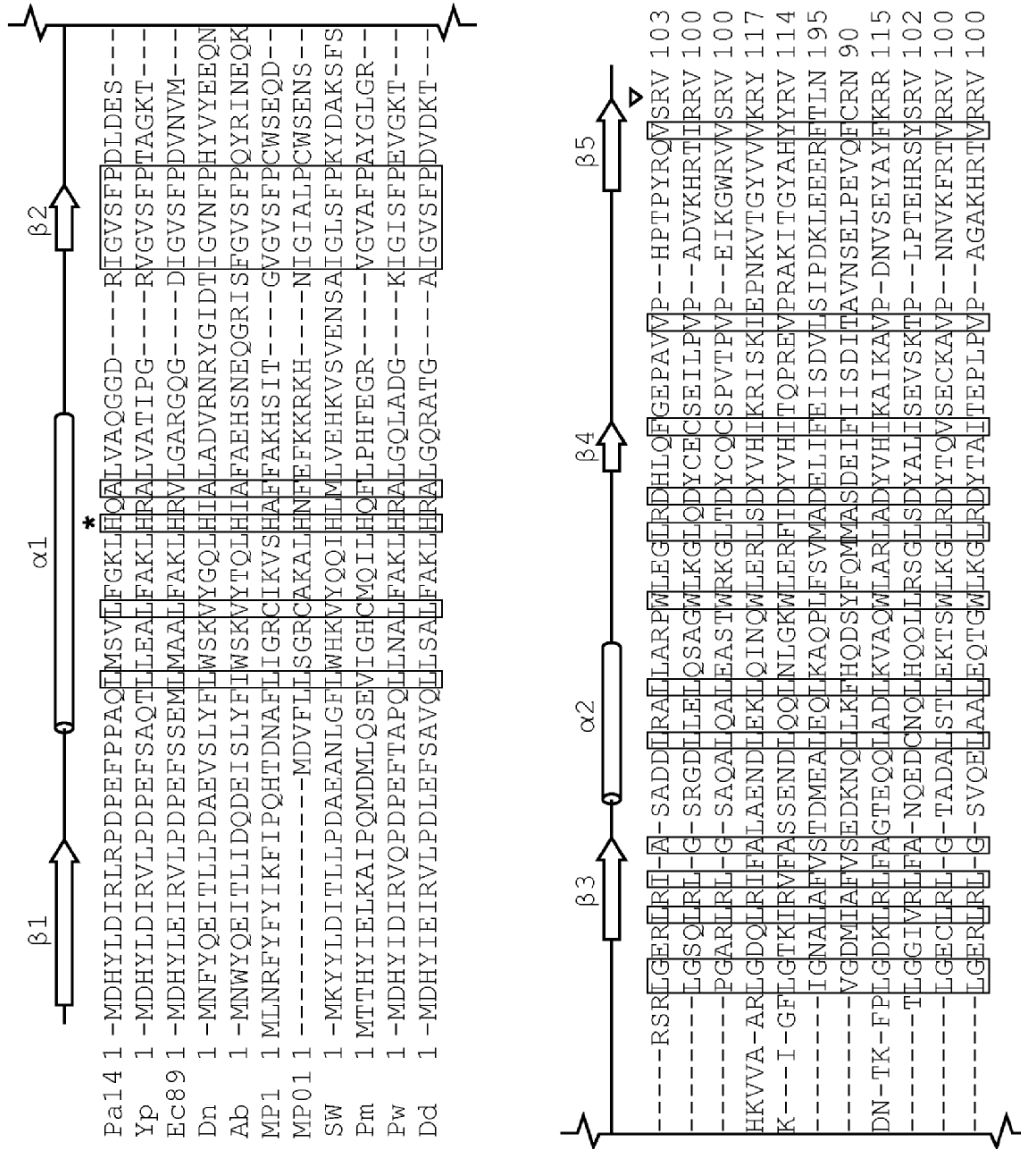


FIG. 4

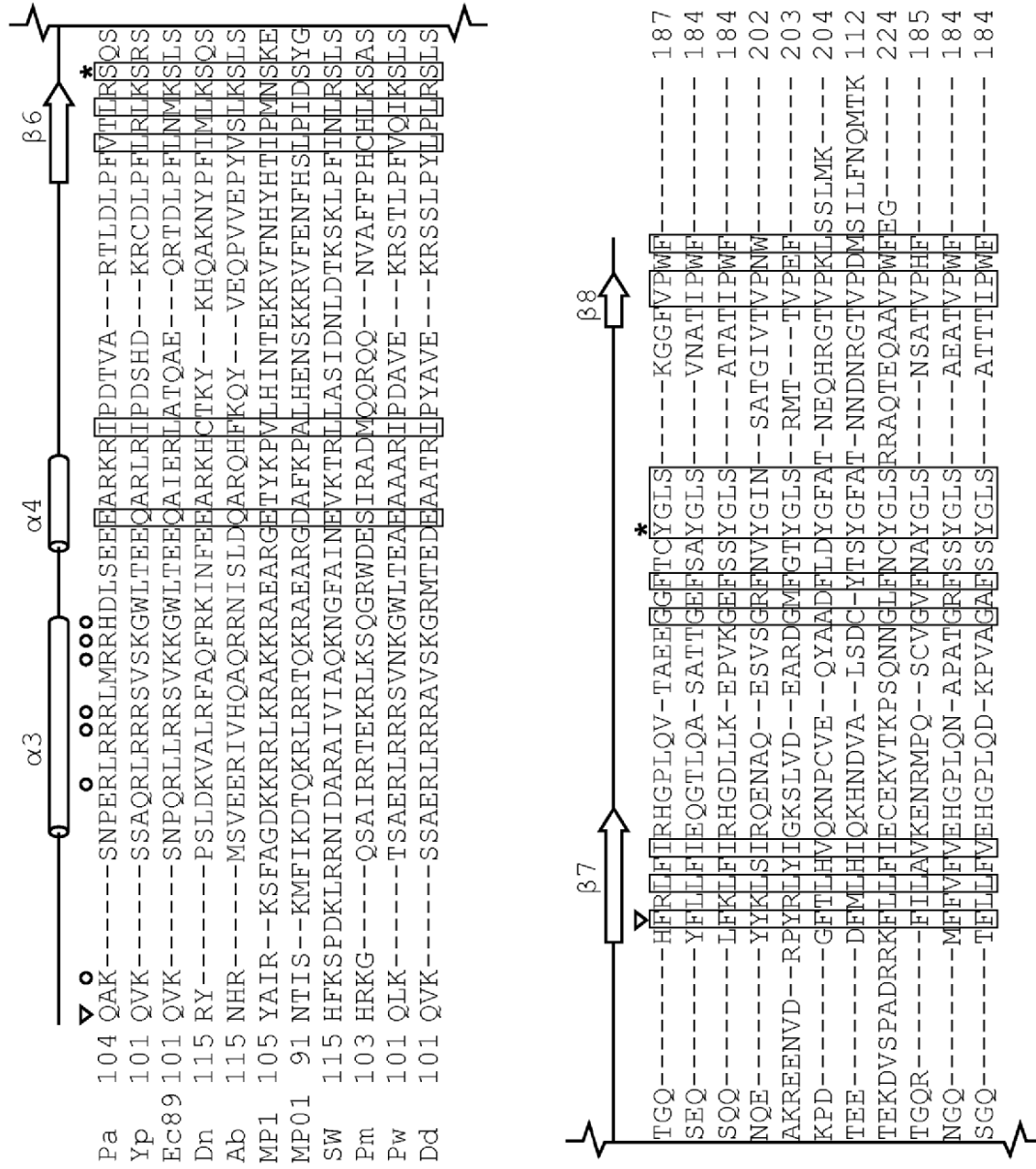


FIG. 4 (Cont.)

FIG. 5A**Secuencia de Csy4**

>gi|116050369|ref|YP_790814.1| proteína hipotética PA14 33300 [Pseudomonas aeruginosa UCBPP-PA14]
 MSVLFGLHOALVAOGGDRIGVSPDLDESRSLGERLRIHASADDLRAILARPWLFLGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRH
 DLSEEEARKRIPDTPVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:5)

Secuencia de reconocimiento de ARN

GUUCACUGCCGUUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5B**Secuencia de Csy4**

>gi|107101871|ref|ZP_01365789.1| proteína hipotética PaerPA 01002916 [Pseudomonas aeruginosa PACS2]
 MDHYLDIRLRPDPFPPAQIMSVLFGKLHQALVÁQGGDRIGVSPDLDESRSLGERLRIHASADDLRAILARPWLFLGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSN
 PERLRRRLMRRHDLSEEEARKRIPDTPVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:6)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUAAGAAA (SEQ ID NO:7)

GUUCACUGCCGUUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5C

>gi|254235433|ref|ZP_04928756.1| proteína hipotética PACG 01340 [Pseudomonas aeruginosa C3719]
 MDHYLDIRLRPDPFPPAQIMSVLFGKLHQALVÁQGGDRIGVSPDLDESRSLGERLRIHASADDLRAIL
 LARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTPVARTLD
 LPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF
 (SEQ ID NO:8)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUAAGAAA (SEQ ID NO:7)

GUUCACUGCCGUUAUAGGCAGCUAAGAAA (SEQ ID NO:1)

FIG. 5D

>gi|254240857|ref|ZP_04934179.1| proteína hipotética PA2G 01531 [Pseudomonas aeruginosa 2192]
 MDHYLDIRLRPDPFPPAQIMSVLFGKLHQALVÁQGGDRIGVSPDLDESRSLGERLRIHASADDLRAIL
 LARPWLEGLRDHLQFGEAAVPHPTPYRQVSRVQAKSNPERLRRRLMRRHDLSEEEARKRIPDTPVARTLD
 LPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSKGGFVPWF (SEQ ID NO:9)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:7)
 GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:1)

FIG. 5E

>gi|242238181|ref|YP_002986362.1| proteína asociada a CRISPR, familia Csy4 [Dickeya dadantii Ech703]
 MDHYIEIRVLPDLEFSAVQLSALFAKLHRAIGORATGAIGVSFPDVKTLGERIRLHSGSVQELAALEQT
 GWLKGLRDYTAITEPLVPFAGAKHRTVRRVQVKSSAERLRRRAVSKGRMTEDEAATRIPIYAVEKRSSLPY
 LPLRSLSSGQTFLIFVEHGPIQDKPVAGAFSSYGLSATITIPWF (SEQ ID NO:10)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCGUAGGCAGCUUAGAAA (SEQ ID NO:11)
 GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)

FIG. 5F

>gi|261822890|ref|YP_003260996.1| proteína asociada a CRISPR, familia Csy4 [Pectobacterium wasabiae WPP163]
 MDHYIDIRVQPDPEFTAPQLINALFAKLHRAIGLADKIGISFPEVKGTLGECRLRHGHTADALSTLEKT
 SWLKGLRDYTOVSECKAVPNNVKFRTVRRVQLKTSAERLRRRSVNGWLTAEAAARIIPDAVEKRSTLPP
 VQIKSLSNQMFVVEHGPIQNAFATGRFSSYGLSAEATVPWF (SEQ ID NO:13)

Secuencia de reconocimiento de ARN

GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:14)

FIG. 5G

>gi|253990195|ref|YP_003041551.1| proteína hipotética PAU 02718 [Photorhabdus asymbiotica subsp. asymbiotica ATCC 43949]
 MDYFFEILVLPDPPEFSKQSIMLFAKLHRAIGVNGRIGVSFPCARKTLGDKLRIHGASEALNDLQAL
 PWLKGLRDYTEIMDIQVPQDTQYRRVSRVQVKSSAERLRRRSIKKGLWLTAEQARQRIPIISKEQRTHLPP
 LLVKSLSSRQTFFLFIHQPIEDKPTPGVSSYGLSASATIPWF (SEQ ID NO:15)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:16)
 GUGCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:17)
 ACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:18)
 GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:19)
 GUGUACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:20)
 GUUCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5H

>gi|307132482|ref|YP_003884498.1| proteína hipotética Dda3937_03453 [Dickeya dadantii 3937]
 MDHYIEIRVLPDPFSGVQLLSALFAKLHRLGQATGALGVSPDAGKTLGERLRLHGSVOELAALEOT
 GWLRGLRDYTAITEPLVPAGVKHRTVRRVQVKSAAERLRRRAVKNKGRMTVDEADARIPYIVVEKRTSLPY
 LPLRSLNGQTFLLFVEHGPIQDKPVAGAFSSYGLSAVATIPWF (SEQ ID NO:22)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:23)
 GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)

FIG. 5I

>gi|285019813|ref|YP_003377524.1| proteína asociada a CRISPR, familia Csy4 [Xanthomonas albilineans GPE_PC73]
 MØHYLDLHLRDPPELAPYQLIGALYARLHRSVLTNTFRIGVSFPDGHDRVPTLGTLLRLHGGDDSTLHHL
 MATTWLHGVRDHTVITISIGAVPSEAVHRQVTRVOAKSSPFLRRAMRRHGISEDLAVQRIIPDSAAEQLR
 LPFVVLGSRSTGQTAFVYFVVRHGPVQPEPVPDGFSSYGLSRGATVPWF (SEQ ID NO:24)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:25)
 UUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:26)
 GUUCACUGCCGUUAGGCAGCUUAGAAA (SEQ ID NO:27)

FIG. 5J

>gi|50122605|ref|YP_051772.1| proteína hipotética ECA3684 [Pectobacterium atrosepticum SCRII043]
 MDHYIDIRVQPDPEFTASQLLNALFAKLHRLVQLANGKIGTSPFVGGKTLGECRLRHGTEDALSTLEKT
 SWLKGLRDYTQVSEKVVVNGVKFRVRRVQLKSSAERLRRRSVSKGWLTAEEAAARIPDAVEKRSALPF
 VQIKLSNGQMFVVEHGPIQNAPTAGRFSSYGLSTEATVPWF (SEQ ID NO:28)

Secuencia de reconocimiento de ARN ...

GUUCACUGCCGUCACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5K

>gi|37525730|ref|NP_929074.1| proteína hipotética plu1796 [Photorhabdus luminescens subsp. laumondii TT01]
 MDYYLEIRVLPDLFESQSQSIFEALFAKLHRLALQLSNGQVGVSPPCARKTLGDTLRIHGSSEALNDLQAL
 PWLKGLRDYTEVIDIQPIPOETKYRCVSRVQVKSAAERLRRRAIKKGWLTGEQARQRIPIISKEQRTHLPPF
 LFLKLSNGQSFLLFVKQGPQDKPTSGIFSSYGLSSSATIPWF (SEQ ID NO:29)

Secuencias de reconocimiento de ARN

GUGCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:17)
 GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5L

>gi|297569494|ref|YP_003690838.1| proteína asociada a CRISPR, familia Csy4 [Desulfovibrio alkaliphilus AHT2]
 MVMAMDCYVEISLLPDPPEFPDSIIMNALFAKLRALAKNGKQELGVSPPEFGKKLNSKLRIHGSEESLKR
 LMDLNWIQGMKDYTRVSGIAKVPDSCQIRIVKRVQAKSSVDRLYRPSVKKGWLSSEENAEQQKERAREGRL
 KLPFVQLKSQTTGQQFRFLFIQHGSIQEKPVTRFSSYGLSNEATVPWF (SEQ ID NO:30)

Secuencia de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:31)

FIG. 5M

>gi|251788340|ref|YP_003003061.1| proteína asociada a CRISPR de la familia Csy4 [Dickeya zeae Ech1591]
 MDHYIEIRVLPDPEFSAVQLLSALFAKLHRAIVATIPGRVGVSFPTAGKTLGSQRLRHGSRGDLLELQSA
 GWRTGLRDYSTITDVLTVPTGAQYRTVRRVQVKSSEERLRRRAVSKGWLTADEAAARIPIYAAEKRTSLPY
 LPLRSLSSGGQTFLLFVEHGFLQDKPVAGTFSSYGLSATATIPWF (SEQ ID NO:32)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUGUAGGCAGCUUAGAAA (SEQ ID NO:23)
 GUGCACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:33)

FIG. 5N

>gi|22125621|ref|NP_669044.1| proteína hipotética y1727 [Yersinia pestis KIM 10]
 MDHYLDIRVLPDPEFSAQTLLEALFAKLHRAIVATIPGRVGVSFPTAGKTLGSQRLRHGSRGDLLELQSA
 GWLKGLDYCECSEIILPVPADVVKHRTIRRVQVKSQAQRLRRRSVSKGWLTEEQARLIRIPDSDHKRCDLPP
 LRLKRSSEQYFLLFIEQGTIQASATTGFSAYGLSVNATIPWF (SEQ ID NO:34)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
 UGUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:35)

FIG. 5O

>gi|271501952|ref|YP_003334978.1| proteína asociada a CRISPR de la familia Csy4 [Dickeya dadantii Ech586]
 MDHYIEIRVLPDPEFSAVQLLSALFAKLHRAIVATIPGRVGVSFPTAGKTLGERLRLHGSVQALAALEQT
 GWLKGLDYSTITDVLTVPTGAQYRTVRRVQVKSSEERLRRRAVSKGRLTADEAAARIPIYAAEKRTSLPY
 LPLRSLSSGGQTFLLFVEHGFLQDKPVAGVFSYGLSAIATIPWF (SEQ ID NO:36)

Secuencias de reconocimiento de ARN

GUGAACUGCCGCAUAGGCAGCUUAGAAA (SEQ ID NO:37)
 GUUCACUGCCGAGUAGGCAGCUUAGAAA (SEQ ID NO:12)

FIG. 5P

>gi|117623067|ref|YP_851980.1| proteína hipotética APEC01_1206 [Escherichia coli APEC 01]
 MAVSLVPRNKNKELPMDHYLEIRVLPDPEFSSEMMAALFAKLHRAIVATIPGRVGVSFPTAGKTLGERLRLHGSVQALAALEQT
 LHGSAQALQALEASTWRKGLTDYCCSPVTPVPEIKGWRVSVRVQVKSQKSNPQRLRRSVKKGWLTFFEQALIE
 RLATQAEQRTDLPFLNMMKSLSSQQLFKLFIHGDLLKPEVKGEFSSYGLSATATIPWF (SEQ ID NO:38)

Secuencia de reconocimiento de ARN
 GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5Q

>gi|91209927|ref|YP_539913.1| proteína hipotética UTI89 C0896 [Escherichia coli UTI89]
 MDHYLEIRVLPDPEFSSEMMAALFAKLHRVVGARGQDIGVSPDVMVMPGARLRTHGSAQALQALEAS
 TWRKGLTDYCCSPVTPPEIKGWRVSRVQKSNPQRLRRSVKKGWLTTEEQAIERLATQAEQRTDLPF
 LNMKLSLSSQQLFKLFIIRHGDLLEKPEVKGEFSSYGLSATATIPWF (SEQ ID NO:39)

Secuencia de reconocimiento de ARN
 GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5R

>gi|237808124|ref|YP_002892564.1| proteína asociada a CRISPR, familia Csy4 [Tolomonas auensis DSM 9187]
 MDHYLDIRLLPEEPEVSEFLINALFAKLHVRIGQQAQGRVGVSPDHHKRIQDILLRLHGQRTDILQAIMA
 DDWLQGLKGYTQCSEVLPITPATVSYRAVKRVQAKSAHNKRQRSIAKGWLTSESAQIRIPDTQQKELHLPF
 VQLKRSRNGQMRVYVEHGPVLAVPVSGYFNAYGLSSIATIPWF (SEQ ID NO:40)

Secuencia de reconocimiento de ARN
 CUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:41)

FIG. 5S

>gi|259907505|ref|YP_002647861.1| proteína Csy4 asociada a CRISPR [Erwinia pyrifoliae Ep1/96]
 MDHYQDIRVVRDEEENGEAVLLAQVFMHLHQVLMRAANGRIGISFFNVKRTIGDRIRLHGTLDDLALQOS
 GWNKCLFRDYIACSDIAPVFKGAAMKFRVRVQVKSSAEKLRKRKSVNKGWLSAQEAAERI SVLNEQRSNLPF
 LQIKSGSNGQAWRLFIEHGSLSVAPSDGFSFSSYGLSAAATIPWF (SEQ ID NO:42)

Secuencias de reconocimiento de ARN

GUUCACUGCCGACAGGCAGCUUAGAAA (SEQ ID NO:19)
 UUCACUGCCGUACAGGCAGCUUAGAAAA (SEQ ID NO:43)

FIG. 5T

>gi|218688670|ref|YP_002396882.1| proteína hipotética ECED1_0855 [Escherichia coli ED1a]
 MAYSLVFRNRKELPMDHYLEITRVLPDPEFSSEMIMAAALFAKLRVVGARGQDYGVSFPDVMMPGTHLR
 LFGSAQALQELFASTWRKGLTDYCCSPVTPVPEIKGWRVSRVQVKSNPQRLRRSVKKGWLTTEEQAIE
 RLATQAEQRITDLPFLNMKSLSSQQQFKLFIIRHGDLIKKIPVKGEFSSYGLSATATIPWF (SEQ ID NO:44)

Secuencia de reconocimiento de ARN

GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5U

>gi|121608426|ref|YP_996233.1| proteína de la familia Csy4 asociada a CRISPR [Verminephrobacter eiseniae EF01-2]
 MSTHYITITLRPPDPEFSPAHLNALHAQLHALVOLGTGVDVGVSPFGFILLRGEHSHLGTTLRLHGATSAL
 QRLQALSWLRGMRDHWKTISEVAPVPTHQHRVWRVQAKSSPERSRRRLMRRLIIDEAQAALQRIIPDQEGR
 RLALPYLRILQSAKQGVFRFLFIEHGFLDTPSPGSGFTYGLSTQATIPWF (SEQ ID NO:45)

Secuencia de reconocimiento de ARN

GUUCACUGCCGGAUAGGCAGCUUAGAAA (SEQ ID NO:46)

FIG. 5V

>gi|34497206|ref|NP_901421.1| proteína hipotética CV 1751 [Chromobacterium violaceum ATCC 12472]
 MDHYLDIRLLPDADFPPVIMNALYAKLHRALAAQQQDIDGVFFGYDPAPSSHDKPLPPTLGLTLRLH
 GSAALDGLMARRWLSGFADHAIIVGDIRPVPAGASAVSVRRROAKSSPARARDLIMRROGISAEFEARRRI
 PDETAQRNLNLPYLTVDSASTGQCFRLFVEQQAAFPSIAAGSFNAYGLSAAAAALPAW (SEQ ID NO:47)

Secuencia de reconocimiento de ARN

GUUCACUGCCGGUAGGCAGCUUAGAAA (SEQ ID NO:48)

FIG. 5W

>gi|188532992|ref|YP_001906789.1| proteína hipotética ETA 08450 [Erwinia tasmaniensis Et1/99]
 MDRYQDIRVRVDAEMTAPVLLAQVFMRLHQVLMRAANGRIGISFFDVKLTIGDRIRLHGTLDDLSLIQQS
 GWDKGLTDYIACSAIDPVPFGAAMRTVRRVQVKSSAERLLRFRSVNKGWLNFAEAAERINVLSEQRSDLPY
 LQIKSGSNHAWRLFIEHGPLVSVFVNGGFSYGLSATATVPWF (SEQ ID NO:49)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
 GUUCACUGCCGUACAGGCAGCUUAGAAAG (SEQ ID NO:50)

FIG. 5X

>gi|160896663|ref|YP_001562245.1| proteína de la familia Csy4 asociada a CRISPR [Delftia acidovorans SPH-1]
 MAMTSHYIDTTLPPDPEFSAHLLGALVAKLHRAIVQLGSTDIGISFPGYSILRPRTLGTILRLHGSEAAAL
 RGLMEQPLQGRMDHVHCTPPALVPEGAVPCLVQRRQFKTSPDLRRRRMRKGETAEQAAAAIPDSVER
 TPDLFVYQLRSASTGQPFCLFVEQKAVQGTAGQEGFNTYGLSLGTAVPWF (SEQ ID NO:51)

Secuencia de reconocimiento de ARN

GUUCGUGCCCGGUAGGCAGCUCAGAAA (SEQ ID NO:52)

FIG. 5Y

>gi|146311064|ref|YP_001176138.1| proteína de la familia Csy4 asociada a CRISPR [Enterobacter sp. 638]
 MDHYLEIRVLSDFEFSEETLMAALFAKLHRAIGARGOGDIGVSFPRYSLKPGDTLRLHGSACQSLDELEKM
 AWKGLSDYCLCKGVLPAPDVNAWRCVSRVQVKSSPQRIIMRRSVKKGWLTETEEAQQRLNLNQEARTDLIPW
 LNLQSLSTGQSFRLFI RHGDIVDMPMCGEFSSYGLSATATIPWF (SEQ ID NO:53)

Secuencia de reconocimiento de ARN

GUUCACUGCCGACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5Z

>gi|289209612|ref|YP_003461678.1| proteína asociada a CRISPR, familia Csy4 [Thioalkalivibrio sp. K90mix]
 MDHYLDLRVMPDFEFKETTLLGALVSKLHRRLLVSMADDIGISLPDHEQEPDLGRRLLRVHGTQGRNLILM
 QDEWLGMSQSLVDATPVPVPOVTVYRPVRRRQKTNAERLRRRMRHGESYEAFQHI PDTVERRVNT
 PFLSVQSASTGGQRFSLFIEHGPPQQHASPGRNTYGLSQDATVPWF (SEQ ID NO:54)

Secuencia de reconocimiento de ARN

GUUAGCUGCCGACAGGCAGCUUAGAAA (SEQ ID NO:55)

FIG. 5AA

>gi|283856310|ref|YP_162420.2| proteína asociada a CRISPR de la familia Csy4 [Zymomonas mobilis subsp. mobilis ZM4]
 MLANPVDSYQDIYLLPNQEIAPHI IMEKLFSLLHLELVRLGSOHIGISFPEHDNKNKPCGLGSRLLHGTGA
 DLHELALSGWITRLDDLYCEDIKSVPEIRQYCVVSRVQAKSSPARLRRRAIRRHGFHDEEAKKVIPTDTA
 FERLELFFIMTGCSTKQPRFFVFI SHKIIQKLMNGNFNYSYGLSIGASVPWF (SEQ ID NO:56)

Secuencia de reconocimiento de ARN

GUUCACUGCCGACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AB

>gi|260752821|ref|YP_003225714.1|proteína asociada a CRISPR, familia Csy4 [Zymomonas mobilis subsp. mobilis NCIMB 11163] MLANPVDSYQDIYLLPNOEIAPIHIMEKLFSLHLELVRLGSOHIGISFPEHNNKPKLGSRLRLHGAGA DLHELALSGWITRLDDYLYCEDIKSVPEIRQYCVVSRVQAKSSPARLRRRAIRRHGFHDEEAKKVIPTDA FERLELFFIMTGTSCSTKQPRFPVFI SHKLIQDKLMNGNFNSYGLSIGASVPWF (SEQ ID NO:57)

Secuencia de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AC

>gi|121592915|ref|YP_984811.1|proteína de la familia Csy4 asociada a CRISPR [Acidovorax sp. JS42] MTHYINITILLPDPFEFSAHILGALVAKLHRAIVQGHHTDIGVSYPOHVSOPLTKRTIGAVIRLHGTPFA LORIMEEDWLKGMRDHTOVGELLVPEANAQHRIVRRRQFKTNADRLRRRRMQRKGETAEQAAAAIPDTVE RRPDLFPVQLRSSSTGQSFCLVVEHGFLLQPLFVAGAFNAYGLGHDATVPWF (SEQ ID NO:58)

Secuencias le reconocimiento de ARN

GUUCACUGCCGCAUAGGCAGCUCAGAAA (SEQ ID NO:59)

FIG. 5AD

>gi|317051103|ref|YP_004112219.1|proteína asociada a CRISPR de la familia Csy4 [Desulfurispirillum indicum S5] MDSYIEIRIILPDQEFEEATILMSTVFAKLHRAIVESGRSDIGVSFPEAGKTPGALLRLHGSIAALESIMTIL SWLTGLQDYTOTSGIILQVPAQAAVQVARVOSKMTASRIRRALKRGSLSEERALELLQSRDQLNQPFTRL LSASTAQKFFLFFIEQRNAEKAGQSVYSAYGLSVGGSTVPWF (SEQ ID NO:60)

Secuencia de reconocimiento de ARN

GUUCACUGCCGCAUAGGCAGCUCAGAAA (SEQ ID NO:59)

FIG. 5AE

>gi|54303643|ref|YP_133636.1| proteína hipotética PBPRB1991 [Photobacterium profundum SS9]
 MDSYVDIQLKPDAAEMREAEELSSKVFTKFKALATLNTNKIGISFPQMNILKGLRILFRHGNASLLKDLQG
 IKWLGALAGYCVQVGEITVVPDQVQYRVISVKRSNLSKAKLKRLLIARGLDKDGEKRYKVKMLSQGFDPFY
 LIDLFSSTGQVYRKFPEFGDIQATSVSDEFDSYGLSNTATIPWF (SEQ ID NO:61)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AF

>gi|54292953|ref|YP_122340.1| proteína hipotética plp10047 [Legionella pneumophila str. Lens]
 MDHYLDISILPDSEFTTPVLMNAIYTNLHKALHTLASTNIGVSFPKYSSTLGNILRIHGKKEALQELQNL
 NWIGMIGYCEASLTKTVPADTKFRTVSRKQPTMSQSKLRRLLIKRNSLLEDEIRQYKAKMFSKGLDNPYI
 ELVSVSNGQRHRRYIEFGELFNEPIPGFLFDQFGLSNSATVPWFD (SEQ ID NO:62)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)
 GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AG

>gi|54295752|ref|YP_128167.1| proteína hipotética lpl2842 [Legionella pneumophila str. Lens]
 MDHYLEISILPDSEFTTPVLMNAIYTNLHKALHTLASTSIGVSFPKYSSTLGNILRIHGKKEVLQDLQNL
 NWIGMIGYCEASLTKTVPADTKFRTVSRKQPTMSQSKLRRLLIKRNTLLEDEIRQYKAKMFSKGLDNPYI
 ELVSVSNGQRHRRYIEFGELFNEPSPGLFDQFGLSNSATVPWFD (SEQ ID NO:63)

Secuencias de reconocimiento de ARN

GUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)
 GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)

FIG. 5AH

>gi|146292921|ref|YP_001183345.1|proteína de la familia Csy4 asociada a CRISPR[Shewanella putrefaciens CN-32]
 MNSYIDIRLKPDAEMREAEELSSKVFTKFKALVTLNSHKIGISFPQMKLSLQGLFRIHGDAALLHDLQGL
 DWLGLAGYCVTAVSAPVDHVQYRIVSVKRSNLSKAKLKRLLIARGLDKDGEKRYKVKMLGQGFDPYLL
 DLFSSSTGQVYRKFVFEFSDIQAHPLIDGEFDSYGLSKTATVPWF (SEQ ID NO:64)

Secuencia de reconocimiento de ARN

GUUACACGCCGCACAGCGCCUUAGAAA (SEQ ID NO:65)

FIG. 5AI

>gi|296106567|ref|YP_003618267.1|proteína hipotética lpa 01476 [Legionella pneumophila 2300/99 Alcoy]
 MDYYVDILIKPDSEKSLNFLSLTYTKLHKVLHDMAS²TNIGVSPKYNITLGNILRI¹HSKKVVLDEL¹GM
 NFLSGINNYEVSP¹IKSVPADSKFRI¹LSRKQ¹TMSQSKMRRL¹FKRGSMTVGD¹IRQYKAKMFAKSIDNPYL
 ELVSGSNGYR¹RYRIE¹FGELL¹DQ¹PVYGE¹FRFGL¹SKTATVPWF (SEQ ID NO:66)

Secuencia de reconocimiento de ARN

GUUAAACUGCCGCACAGCGCCUUAGAAAG (SEQ ID NO:67)

FIG. 5AJ

>gi|260772736|ref|ZP_05881652.1|proteína hipotética VIB 001192 [Vibrio metschnikovii CIP 69.14]
 MDSYIEIRIQPDAEMPEAEELSSKVFTKFKALVILHSNQIGISFPEVNVKIGRLFRLHGEASFLHDLQGL
 NWLGPLSGYCOVSEILAIPEOVQYRIVSVKRSNLSOAKLRRLLIARGLDKDGEKRYKVKMLSQGFDPYLL
 DLFSSSTKQVHRKFFVFEIGEIQPLPVS¹GK¹FDSYGLSH¹TTVPWF (SEQ ID NO:68)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCACAGCGCCUUAGAAA (SEQ ID NO:19)
 GUUCACUGCCGC¹CAUAG¹GCAG¹CCUUAGAAA (SEQ ID NO:69)

FIG. 5AK

>gi|157146437|ref|YP_001453756.1| proteína hipotética CKO_02197 [Citrobacter koseri ATCC BAA-895]
 MAITPVPVAVKGMRTVSRVQVKSSPQRLLRSVRKGMWLTTEEQAQLRVLVESTEQHSDDLPLYLNVKSLSNQQQF
 RVFIRHSELRSEPVSGTFTSYGLSSTATIPWF (SEQ ID NO:70)

Secuencia de reconocimiento de ARN

UGUUCACUGCCGUACAGGCAGCUUAGAAA (SEQ ID NO:21)

FIG. 5AL

>gi|262402803|ref|ZP_06079364.1| proteína hipotética VOA_000785 [Vibrio sp. RC586]
 MDAYIDIRLMPDAEMREAEISSLKVFIKFHKALVKLRNKGISFPEANIKLGRFLRHGEMSAIHDLQGL
 NWLGLAGYCKITTVTHVPDQVQYRIISVKRSNLSKAKLTRLIARGSIDKDGKRYKVKMMLSQGFDNPYL
 DLSSSSTGQVYRKFFEFSDIQADPVDGFEFDSYGLSKTATVPWF (SEQ ID NO:71)

Secuencias de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:19)
 AGUGUUCUGCCGAAUAGGCAGCUAAGAA (SEQ ID NO:72)

FIG. 5AM

>gi|229523353|ref|ZP_04412760.1| proteína hipotética VIF_000211 [Vibrio cholerae TM 11079-80]
 MMDAYIDIRLMPDAEMREAEISSLKVFIKFHKALVKLQSNKIGISFPEANIKLGRFLRHGEMSAIHDLQGL
 LNWLGLAGYCKITTVTHVPDQVQYRIISVKRSNLSKAKLARLIARGSIDKDGKRYKVKMMLRQGFDNPY
 LDLSSSSTGQVYRKFFEFSDIQAEPPVDGFEFDSYGLSKTATVPWF (SEQ ID NO:73)

Secuencia de reconocimiento de ARN

GUUCACUGCCGCACAGGCAGCUUAGAAA (SEQ ID NO:74)

FIG. 5AN

>gi|237748015|ref|ZP_04578495.1|proteína asociada a CRISPR[Oxalobacter formigenes OXCC13]
 MKHYEITLTGSPDFLYHLWMSKLYTQLHLALVENRDASDOVNIQVGFPEYFNEEKGMGFLGTKLRLEFA
 EDETSLQKIDIQKWFVRLNDCIHITPYCRVPLNEITGYAIFSRKHIKSNAERLARQMKRHKDLDSFHEIV
 QRYQKNLAKSPLEPFTIQLESLTNSHPFKLFIKKPAINASLKVFTTYGLSAESTIPEF (SEQ ID NO:75)

Secuencia de reconocimiento de ARN

GUUCACUGCCGUAVAGGCGAGCUUAGAAAG (SEQ ID NO:76)

FIG. 5AO

>gi|119945137|ref|YP_942817.1|proteína de la familia Csy4 asociada a CRISPR[Psychromonas ingrahamii 37]
 MKYYLDITLLPDIETPLGFTWQVYFQOVHIALADNKVGENESDIALSLPNYGDKAFLGNKLRLEFSVSEQ
 ALERLAI TKWLKRFTHDHTITSVKAVPESANEYACFTRKQFNTNISRLARRAKRHEMTEFEKALQYYDNF
 AEEQTKLPHMNIKSLTNNAQFRIFTERSITTKIPKQGTFCNYGLSQAIATVPWF (SEQ ID NO:77)

Secuencia de reconocimiento de ARN

GUGUUCGCCGUGCCACGGGGAUGAACCG (SEQ ID NO:78)

FIG. 5AP

>gi|146328647|ref|YP_001209099.1|proteína hipotética DNO 0170 [Dichelobacter nodosus VCS1703A]
 MNFYQEI TLLPDAEVSLYFLWSKVYQLHLIALADVNRNYGIDTIGVNFPHYVYEEQNHKVVAARLGDQLR
 I FALAENDLEKLIQINQWLERLSDVYVHLKRTSKIEPNKVITGYVVKRYRPSLDKVALRFAQFRKINFEFA
 RKHCTKYKHQAKNYPFIMLKSQSNQYYKLSITRQENAQESVSGRENVYGINSATGIVTVPNW (SEQ ID NO:79)

Secuencia de reconocimiento de ARN

GUUCACCGCCGACAGGGGCUUAGAAA (SEQ ID NO:65)

FIG. 5AQ

>qi|160876478|ref|YP_001555794.1| proteína de la familia Csy4 asociada a CRISPR [Shewanella baltica OS195]
 MNHYLDITLLPNEEVGHYFLWEKLYHQVHLALVEHKNRVGGQFEIAAAPPQFNEMDNLSGSKRLRLATQPQ
 HLEDLKVSNWLRHFTDYLHISIRPVDPKIEVYVAYSRRPAIRANKAREIARRMKRHNETLVEQATAHFEGF
 KPKKTKAPFVYMQSYTKDSRFFLFIQQTHSAVWKEGVSFDSYGLSSRGYLPKF (SEQ ID NO:80)

Secuencia de reconocimiento de ARN

GUUCACCGCGCACAGGGCGCUUAGAAA (SEQ ID NO:65)

FIG. 5AR

>qi|153001745|ref|YP_001367426.1| proteína de la familia Csy4 asociada a CRISPR [Shewanella baltica OS185]
 MNHYLDITLLPNEEVGHYFLWEKLYHQVHLALVEHKNRVGGQFEIAAAPPQFNEMDNLSGSKRLRLATQPQ
 HLEDLKVSNWLRHFTDYLHISIRPVDPKIEVYVAYSRRPAIRANKAREIARRMKRHNETLVEQATAHFEGF
 KPKKTKAPFVYMQSYTKDSRFFLFIQQTHSAVWKEGVSFDSYGLSSRGYLPKF (SEQ ID NO:81)

Secuencia de reconocimiento de ARN

GUUCACCGCGCACAGGGCGCUUAGAAA (SEQ ID NO:65)

FIG. 5AS

>qi|169795154|ref|YP_001712947.1| proteína hipotética ABAYE1000 [Acinetobacter baumannii AYE]
 MNWYQEIITLIDQDEISLYFIWSKVYQLHIAFAEHSNEQGRISFGVSPQYRINEQKKIGFLGTKIRVF
 ASSENDLQQLNGKWLRFIDYVHITQPREVPRAKITGYAHYRNVHRMSVEERIVHOAORRNISLDQAR
 QHFQYVEQPVVEPYVSLKSLSAKRENVDRPYRLYIGKSLVDEARDGMFGTYGLSRMTTVPEF (SEQ ID NO:82)

Secuencia de reconocimiento de ARN

GUUCAUGGGCGCAUACGCCAUUUAGAAA (SEQ ID NO:83)

FIG. 5AT

>gi|213158184|ref|YP_002320235.1| proteína asociada a CRISPR, familia Csy4 [Acinetobacter baumannii AB0057]
 MNWYQEI TLIDQDEISLYFWSKVYQLHIAFAEHSNEQGRISFGVSFPQYRINEQKKIGFLGTKIRVFA
 SSENDLQQLNLGKWLERFIDYVHLTQEREVPRAKITGYAHYRVNHRMSVEERIVHCAQRNLSLDOARQ
 HFKQYVEQPVVEPYVSLKLSAKREENVDPRPYRLYIGKSLVDEARDGMFGIYGLSRMTTVPEF (SEQ ID NO:84)

Secuencia de reconocimiento de ARN

GUUCAUGGGCGCAUACGCCCAUUUAGAAA (SEQ ID NO:83)

FIG. 5AU

>gi|15602173|ref|NP_245245.1| proteína hipotética PM0308 [Pasteurella multocida subsp. multocida str. Pm70]
 MTHYIELKAIPOMDMLQSEVIGHCMQLLHQFLPHFEGRVGVAFPAYGLGRILGGIVRLIFANQEDCNQLH
 QQLLRSGLSDYALISEVSKTPLPTEHRSYSRVHRKQQSARRTEKRLKSQGRWDESIRADMQRQONVAF
 FPCHLKSASTGQRFILAVKENRMPQSCVGFNAYGLSNSATVPHF (SEQ ID NO:85)

Secuencias de reconocimiento de ARN

GUUCACCAÜCGUGUAGAUAGGCUUAGAAA (SEQ ID NO:86)
 GUUAAACUGCCGUAUAGGCAGCUUAGAAA (SEQ ID NO:87)

FIG. 5AV

>gi|293390434|ref|ZP_06634768.1| proteína asociada a CRISPR de la familia Csy4 [Aggregatibacter actinomycetemcomitans D7S-1]
 MIVQTHYIEIKAIPOVDMLOTEVIGFCIQKLLHQILPHFEGRIGLAFPAYGNDKTLGGIIRLIFGTENDCGF
 IHFKLQSLRDYALISEVMPIPEKVSRYRYQRIQPKGQSSIRRAEKRLTAQGRWNEEVLQNLMLQKQATQR
 IYPHAILKSSSTKQQFILLAIKSVSHQTKAVEGVFSAYGLSQTTTVPHF (SEQ ID NO:88)

Secuencia de reconocimiento de ARN

CUUCACUGCCGAAUAGGCAGCUUAGAAA (SEQ ID NO:89)

FIG. 5AW

>gi|152996699|ref|YP_001341534.1| proteína de la familia Csy4 asociada a CRISPR [Marinomonas sp. MWYL1]
 MKHYDITLLPSDDIGVHFLWSKLMMQVHLALVLEIQNEQKQVPVAVSFPKYQPRÉNEKLGFGVGNKLRLLFA
 NDKTDLERLNFGLWLRLEDYHIKSIADVPNDVLSYEFNRRSKSGSPDKHIKRRMQRHNETWEQAAAF
 FKGYSMEKADKDLFFIRMKSLHSDNEFCMSIIRKEAAPSNKHIMFNTYGLSAEGVLPKF (SEQ ID NO:90)

Secuencia de reconocimiento de ARN
 GUUCGCCGCCGAGCAGCGGUUAGAAA (SEQ ID NO:91)

FIG. 5AY

>gi|258645690|ref|ZP_05733159.1| proteína asociada a CRISPR, familia Csy4 [Dialister invisus DSM 15470]
 MEYQEIITLLPCAEVSLAFLWTKVFTQHLIAFADEKMKSGHNLVAVSFPYRETGLGKIRVFAEQEILE
 RLNLKVLGRLLDYVHCTSIRKVPERKLRGYAVSRYPQEGSIWVKARYAKRHPGVTTIEEARLLQGKR
 KSVRLPYIQMKLSLRGGTFLFIKRRVEKESALTECGTYGLSNRRTVPEF (SEQ ID NO:92)

Secuencias de reconocimiento de ARN
 GUUAAUCUGCCCGCAUAGGUUAGAAA (SEQ ID NO:93)
 GUUAUCUGCCGUUAGGCAGCUUAGAAA (SEQ ID NO:94)

FIG. 5AZ

>gi|165975671|ref|YP_001651264.1| proteína hipotética APJL_0216 [Actinobacillus pleuropneumoniae serovar 3 str. JL03]
 MSELTHYIELKAIPOVDILQTDVIAHGLQILHKFLPLYQGEIGLSFPAYGLGRTLGGIIRVFGNEQHCITQ
 IKTQLIGELQDYVLITSVTPPEEIVEYHRVQRVHRKQQSARRTEQFLVQGGKWTTEEIRQEMLIHQQN
 QKVPFVVKLSGSKQHFVLAIRQLRAEPASGLFNAYGLSQAATVPHF (SEQ ID NO:95)

Secuencia de reconocimiento de ARN
 CUUCACUGCCGUUAGGCAGCUUAGAAA (SEQ ID NO:96)

FIG. 5BA

>gi|190149486|ref|YP_001968011.1|proteína hipotética APP7_0217 [Actinobacillus pleuropneumoniae serovar 7 str. AP76]
 MSELTHYIELKAIPOVDILQTDVIAHGLQILHKFLPLYOGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
 IKTLIGEGLQDYVLI TSVTPVPEEIVEYHRYQRVHRKQGSARRTEQFLVQQGKWTEEIROEMLIHQON
 QKVFFHVKLSGSTKQHFVLAIRQLRAEPLVSLFNAFGLSKIAIVPHF (SEQ ID NO:97)

Secuencia de reconocimiento de ARN

CUUCACUGCCGUUAUAGGCAGCUUAGAAA (SEQ ID NO:96)

FIG. 5BB

>gi|303253029|ref|ZP_07339182.1|proteína hipotética APP2_1978 [Actinobacillus pleuropneumoniae serovar 2 str. 4226]
 MSELTHYIELKAIPOVDILQTDVIAHGLQILHKFLPLYOGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
 IKTLIGEGLQDYVLI TSVTPVPEEIVEYHRYQRVHRKQGSARRTEQFLVQQGKWTEEIROEMLIHQON
 QKVFFHVKLSGSTKQHFVLAIRQLRAEPLVSLFNAFGLSKIAIVPHF (SEQ ID NO:98)

FIG. 5BC

>gi|303251662|ref|ZP_07337835.1|proteína hipotética APP6_0866 [Actinobacillus pleuropneumoniae serovar 6 str. Femc]
 MSELTHYIELKAIPOVDILQTDVIAHGLQILHKFLPLYOGEIGLSFPAYGLGRTLGGIIRVFGNEQHCTQ
 IKTLIGEGLQDYVLI TSVTPVPEEIVEYHRYQRVHRKQGSARRTEQFLVQQGKWTEEIROEMLIHQON
 QKVFFHVKLSGSTKQHFVLAIRQLRAEPLVSLFNAFGLSKIAIVPHF (SEQ ID NO:99)

FIG. 5BD

>gi|116627507|ref|YP_820126.1|Desfosfo-CoA quinasa [Streptococcus thermophilus IMD-9]
 MSKTMIGLTGGLASGKSTVVEIKDAGYKVIDADQLVHDMQVKGRRLYQALLDWLGDGILLPNGELNRP
 KIGQLIFSSSEMYQSALQGIKRIEELAAKRDCLAKEEDVFFMDIPLLFENDYQDFDQIWLVAVSPQV
 QQRIMKRNHLSAFEAGMRLASQMLAEKLPYASLVIDNNGNIDDLKKKVKGAIKDLANLV (SEQ ID NO:100)

FIG. 6

His29Ala

1 mdhyldirlr pdpefppaql msvlfgkl**la**q alvaqggdri gvsfpdlides rslgerlri
 61 hasaddlral larpwlegrl dhlqfgepav vphptpyrqv srvqvknspe rllrrlmrrh
 121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfirfir hgplqataee ggftcyglsk
 181 ggfvpwf (SEQ ID NO:101)

His29Ala/Ser50Cys

1 mdhyldirlr pdpefppaql msvlfgkl**la**q alvaqggdri gvsfpdl**dec** rslgerlri
 61 hasaddlral larpwlegrl dhlqfgepav vphptpyrqv srvqvknspe rllrrlmrrh
 121 dlseeearkr ipdtvarald lpfvtlrsqs tgqhfirfir hgplqataee ggftcyglsk
 181 ggfvpwf (SEQ ID NO:102)

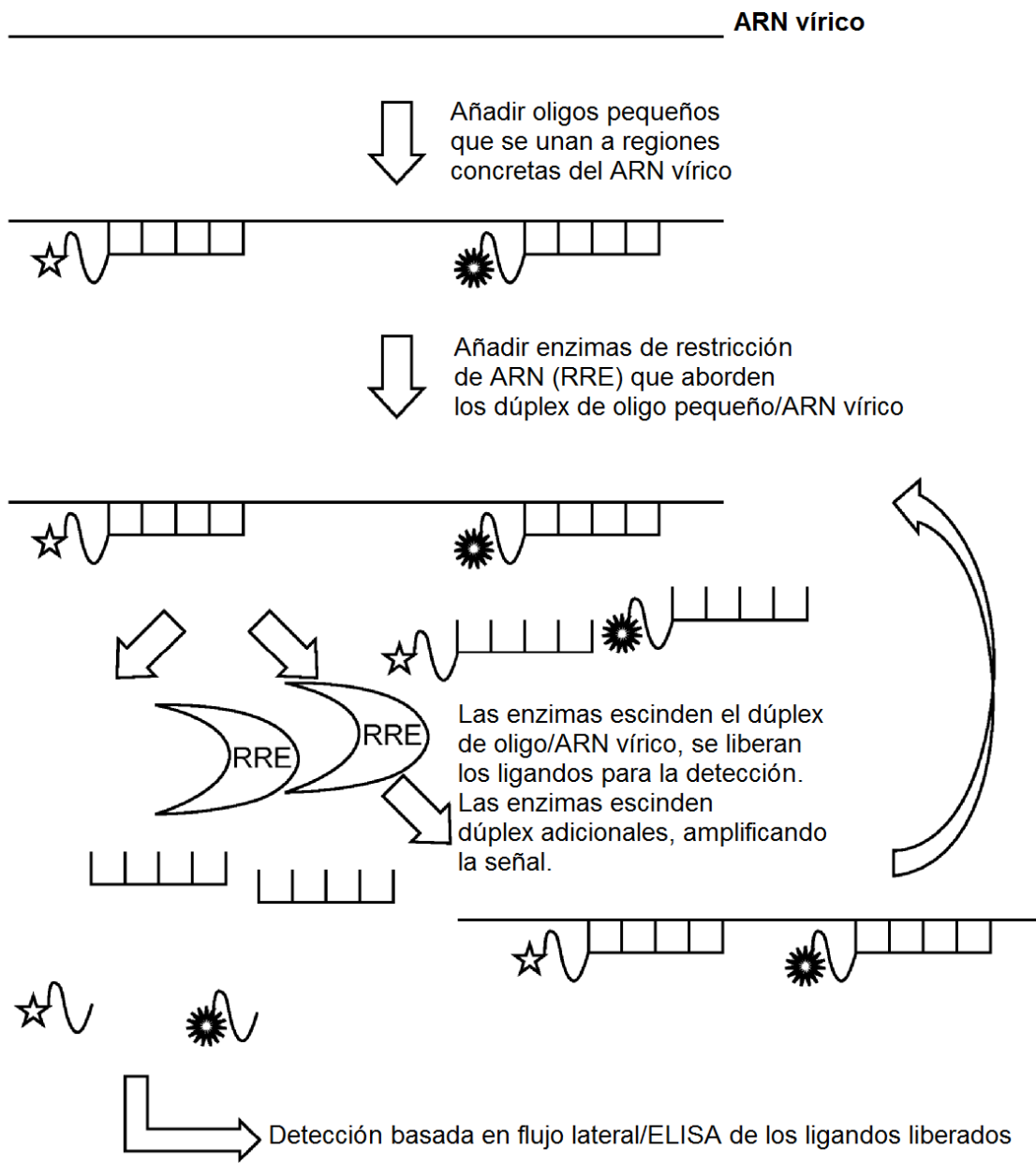


FIG. 7

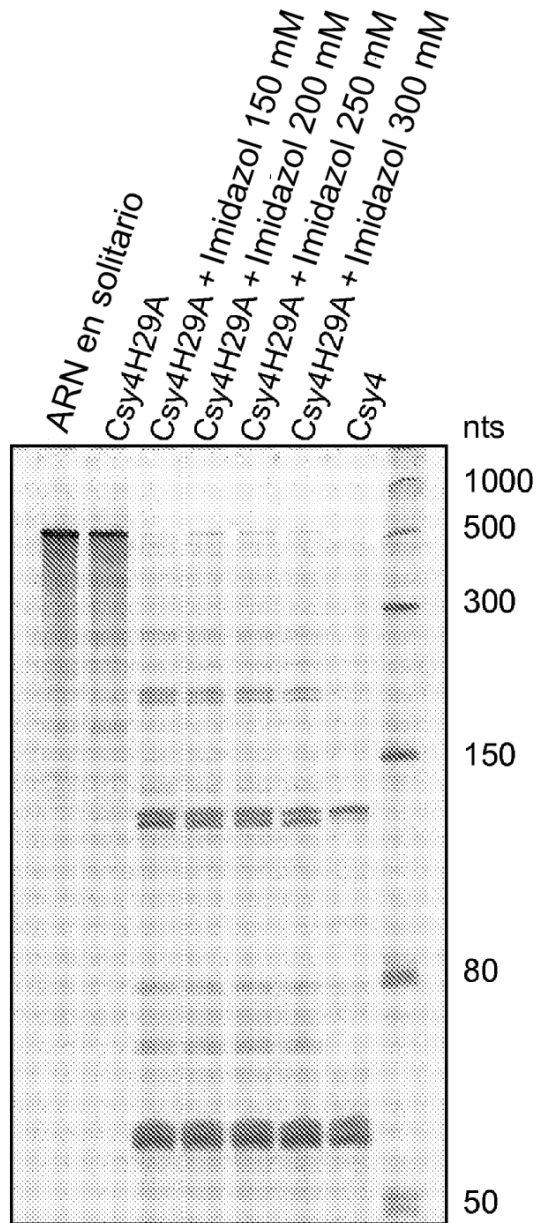


FIG. 8

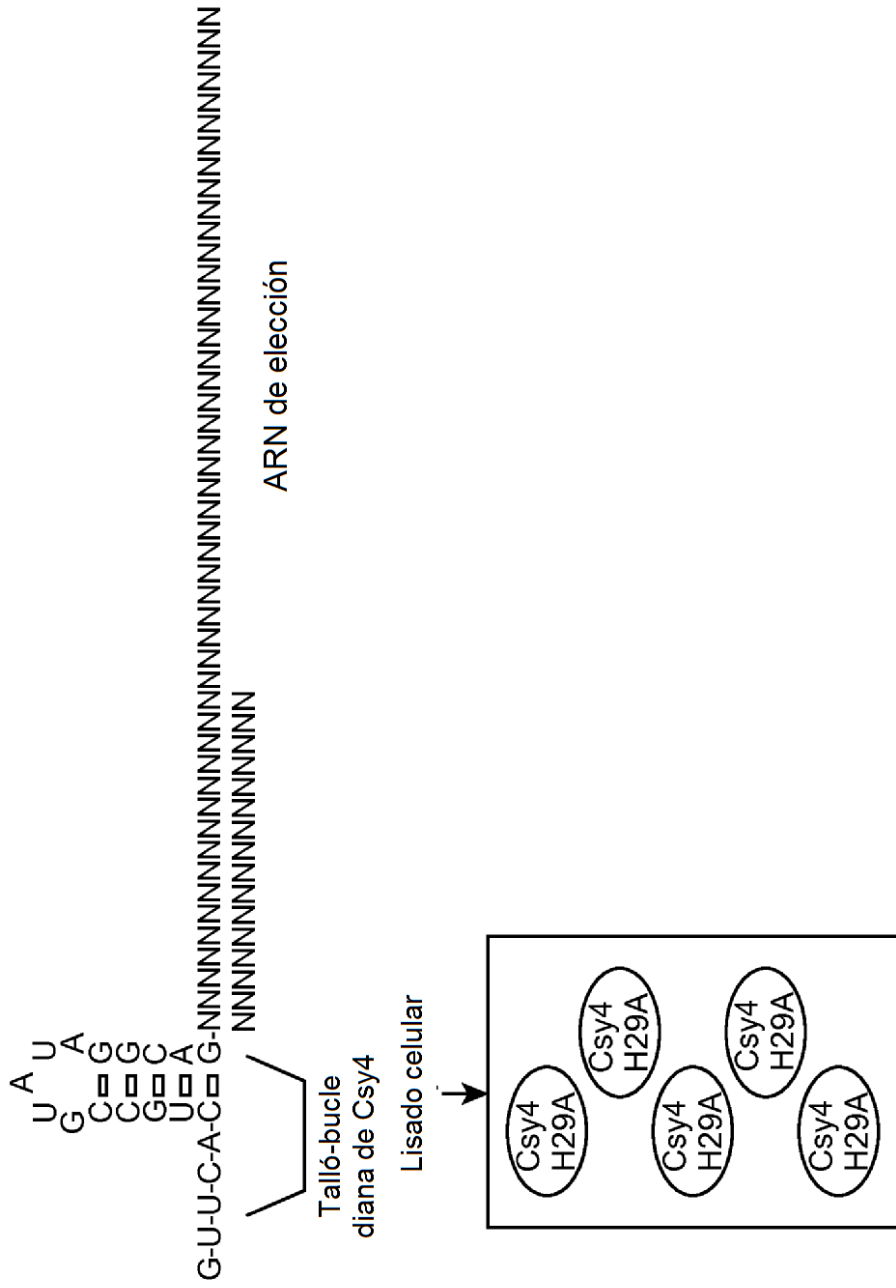


FIG. 9

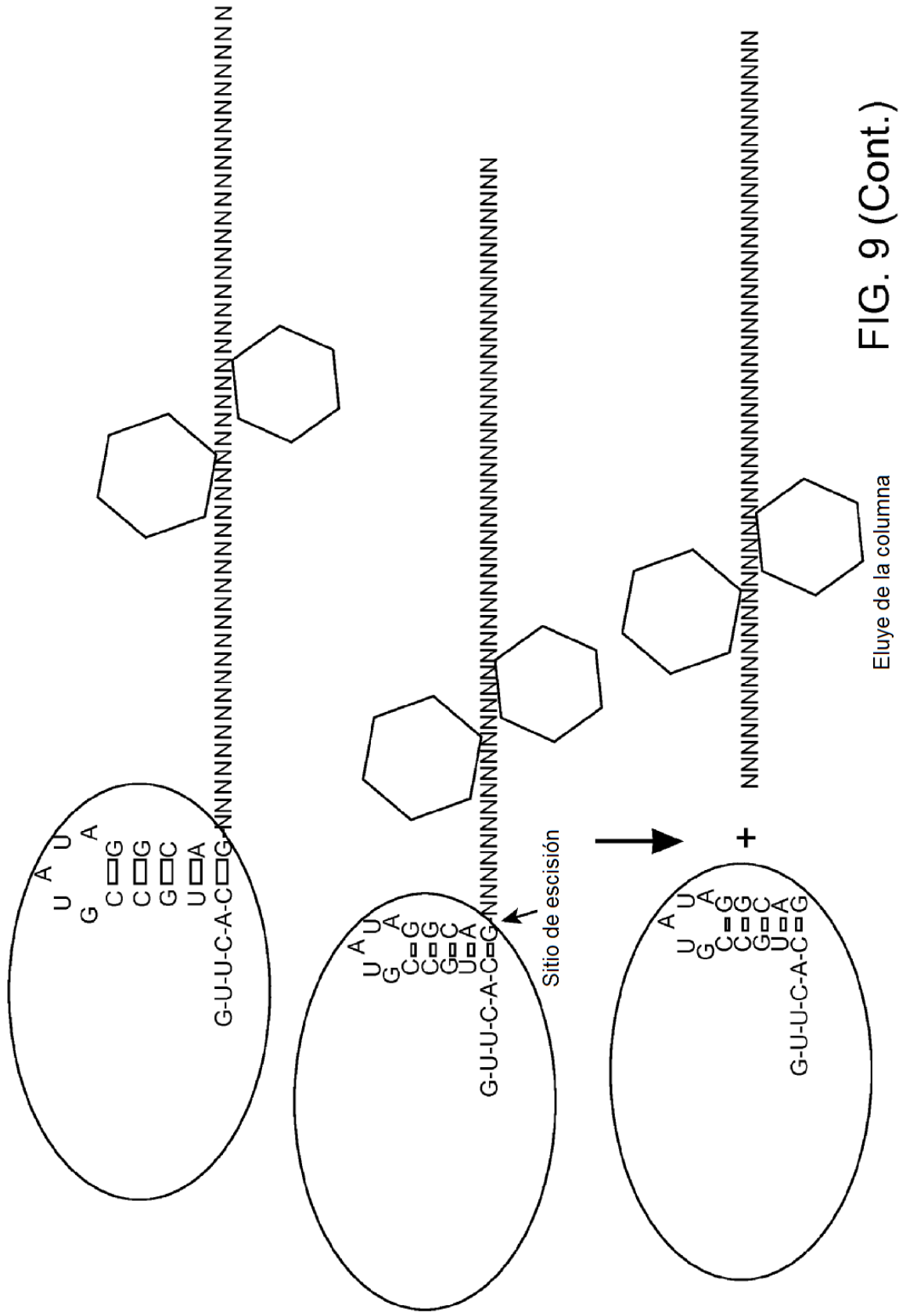


FIG. 9 (Cont.)

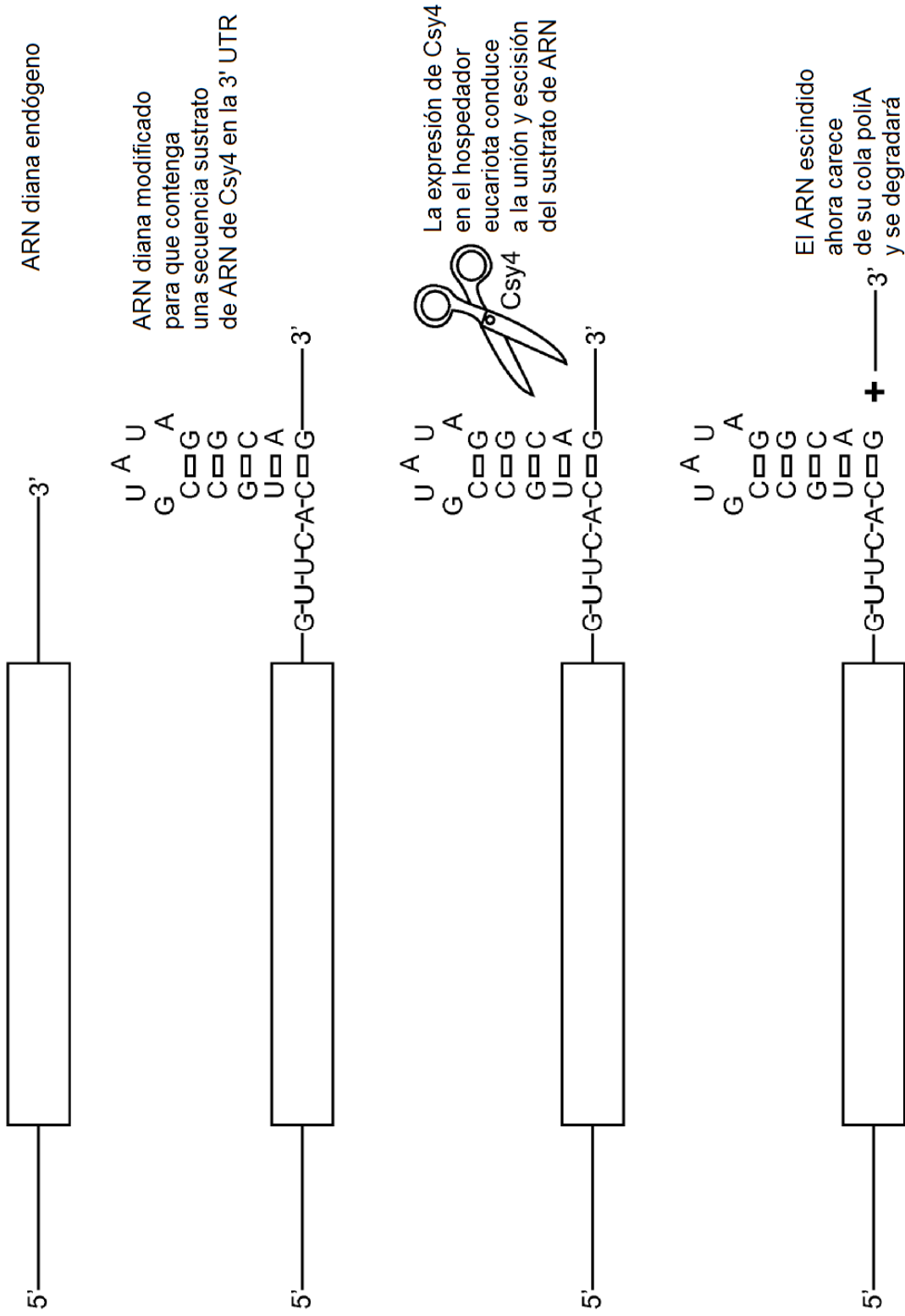


FIG. 10