(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 786 193**

(51) Int. CI.:

C12N 15/63 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **12.12.2013 E 16151694 (3)**

(97) Fecha y número de publicación de la concesión europea: **05.02.2020 EP 3064585**

(54) Título: **Modificación por tecnología genética y optimización de sistemas, métodos y composiciones enzimáticas mejorados para la manipulación de secuencias**

(30) Prioridad:

12.12.2012 US 201261736527 P
02.01.2013 US 201361748427 P
30.01.2013 US 201361758468 P
25.02.2013 US 201361769046 P
15.03.2013 US 201361802174 P
15.03.2013 US 201361791409 P
28.03.2013 US 201361806375 P
20.04.2013 US 201361814263 P
06.05.2013 US 201361819803 P
28.05.2013 US 201361828130 P
17.06.2013 US 201361835931 P
17.06.2013 US 201361836101 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.10.2020

(73) Titular/es:

THE BROAD INSTITUTE, INC. (33.3%)
415 Main Street
Cambridge, MA 02142, US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY (33.3%) y
PRESIDENT AND FELLOWS OF HARVARD COLLEGE (33.3%)

(72) Inventor/es:

ZHANG, FENG;
RAN, FEI y
SHALEM, OPHIR

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

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

Modificación por tecnología genética y optimización de sistemas, métodos y composiciones enzimáticas mejorados para la manipulación de secuencias

Esta solicitud reivindica prioridad de la solicitud de patente provisional de EE.UU. 61/836.101 titulada MODIFICACIÓN POR TECNOLOGÍA GENÉTICA Y OPTIMIZACIÓN DE SISTEMAS, MÉTODOS Y COMPOSICIONES ENZIMÁTICAS MEJORADOS PARA LA MANIPULACIÓN DE SECUENCIAS presentada el 17 de junio de 2013. Esta solicitud también reivindica prioridad de las solicitudes de patente provisionales de EE.UU. 61/758.468; 61/769.046; 61/802.174; 61/806.375; 61/814.263; 61/819.803 y 61/828.130, titulada MODIFICACIÓN POR TECNOLOGÍA GENÉTICA Y OPTIMIZACIÓN DE SISTEMAS, MÉTODOS Y COMPOSICIONES PARA LA MANIPULACIÓN DE SECUENCIAS, presentadas el 30 de enero de 2013; 25 de febrero de 2013; 15 de de marzo de 2013; 28 de marzo de 2013; 20 de abril de 2013; 6 de mayo de 2013 y 28 de mayo de 2013, respectivamente. Esta solicitud también reivindica prioridad de las solicitudes de patente provisionales de EE.UU. 61/736.527 y 61/748.427, ambas con el título SISTEMAS, MÉTODOS Y COMPOSICIONES PARA LA MANIPULACIÓN DE SECUENCIAS, presentadas el 12 de diciembre de 2012 y 2 de enero de 2013, respectivamente. La prioridad también se reivindica de las solicitudes de patente provisionales de EE.UU. 61/791.409 y 61/835.931, presentadas el 15 de marzo de 2013 y 17 de junio de 2013, respectivamente.

Se hace referencia a las solicitudes de patente provisionales de EE.UU. 61/836.127, 61/835.936, 61/836.080, 61/836.123 y 61/835.973, presentadas cada una el 17 de junio de 2013.

CAMPO DE LA INVENCIÓN

La presente invención se refiere, en general, a la modificación por tecnología genética y la optimización de sistemas, métodos y composiciones utilizados para el control de la expresión génica que implican la focalización de la secuencia, tales como la perturbación del genoma o la edición de genes, que se refieren a Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas (CRISPR) y componentes de los mismos.

DECLARACIÓN PATROCINADA POR EL GOBIERNO FEDERAL

Esta invención se hizo con el apoyo del Gobierno bajo la NIH Pioneer Award (1DP1MH100706) adjudicada por el National Institute of Health. El Gobierno tiene ciertos derechos en la invención.

ANTECEDENTES DE LA INVENCIÓN

Recientes progresos en las técnicas de secuenciación del genoma y métodos de análisis han acelerado significativamente la capacidad para catalogar y mapear factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Se necesitan tecnologías de focalización del genoma, precisas para permitir una modificación por tecnología genética inversa sistemática de las variaciones genéticas causales, permitiendo una perturbación selectiva de elementos genéticos individuales, así como para avanzar en la biología de síntesis, biotecnológica y aplicaciones médicas. Aunque las técnicas de edición del genoma, tales como los dedos de zinc diseñadores, efectores del tipo activador de la transcripción (TALEs), o meganucleasas migratorias están disponibles para la producción de perturbaciones del genoma focalizado, sigue habiendo una necesidad de nuevas tecnologías de la ingeniería del genoma que son asequibles, fáciles de instalar, expansibles y susceptibles de fijar como objetivo múltiples posiciones dentro del genoma eucariota.

SUMARIO DE LA INVENCIÓN

El CRISPR/Cas o el sistema CRISPR-Cas (ambos términos se utilizan indistintamente a lo largo de toda esta solicitud) no requieren la generación de proteínas modificadas para fijar como objetivo secuencias diana específicas, sino más bien una sola enzima Cas puede ser programada por una molécula de ARN corta para reconocer una diana de ADN específica, en otras palabras, la enzima Cas puede ser reclutada a una diana de ADN específica utilizando dicha molécula de ARN corta. Añadiendo el sistema CRISPR-Cas al repertorio de técnicas de secuenciación del genoma y métodos de análisis puede simplificar significativamente la metodología y acelerar la capacidad de catalogar y mapear los factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Para utilizar el sistema CRISPR-Cas de manera efectiva para la edición del genoma sin efectos perjudiciales, es fundamental entender aspectos de la modificación genética y la optimización de estas herramientas de modificación genética del genoma, que son aspectos de la invención reivindicada.

Por consiguiente, existe una necesidad urgente de sistemas y técnicas, alternativos y robustos, para la focalización de la secuencia de dirección con una amplia gama de aplicaciones. Aspectos de esta invención satisfacen esta

necesidad y proporcionan ventajas relacionadas. Un complejo de CRISPR a modo de ejemplo comprende una enzima de CRISPR que forma complejo con una secuencia guía hibridada con una secuencia diana dentro del polinucleótido diana, en donde la enzima de CRISPR es un ortólogo Cas, p. ej., un ortólogo Cas9 de un género que incluye, pero no se limita a *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria,Roseburia, Parvibaculum, Staphylococcus, Nitratifactor, Mycoplasma* y *Campilobacter.*. La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia tracr.

En un aspecto, la invención proporciona métodos para usar uno o más elementos de un sistema de CRISPR. El complejo de CRISPR de la invención proporciona un medio eficaz para modificar un polinucleótido diana. El complejo de CRISPR de la invención tiene una amplia diversidad de utilidades incluyendo modificar (p. ej., suprimir, insertar, translocar, inactivar, activar, reprimir, alterar la metilación, transferir restos específicos) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo de CRISPR de la invención tiene un amplio espectro de aplicaciones en, p. ej., la edición del gen o del genoma, la regulación de genes, la terapia génica, el descubrimiento de fármacos, el rastreo de fármacos, el diagnóstico de enfermedades, y el pronóstico. En aspectos preferidos de la invención el complejo CRISPR comprende una enzima Cas, preferiblemente un ortólogo Cas9, de un género que incluye, pero no se limita a *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria,Roseburia, Parvibaculum, Staphylococcus, Nitratifactor, Mycoplasma* y *Campilobacter.* Aspectos de la invención se refieren a enzimas de CRISPR que tienen una función optimizada. Con respecto a la enzima de CRISPR que es una enzima Cas, realizaciones preferidas de la invención se refieren a ortólogos Cas9 que tienen especificidad diana mejorada en un sistema CRISPR-Cas9. Esto se puede realizar mediante enfoques que incluyen, pero no se limitan al diseño y preparación de ARNs guía que tiene una actividad óptima, la selección de enzimas Cas9 de una longitud específica, el truncamiento de la enzima Cas9 haciéndola más pequeña en longitud que la enzima Cas9 de tipo salvaje correspondiente, truncando las moléculas de ácido nucleico que codifican la mismo y generando enzimas Cas9 quiméricas en las que diferentes partes de la enzima se permutan o intercambian entre diferentes ortólogos para llegar a enzimas quiméricas que tienen especificidad adaptada. Aspectos de la invención se refieren también a métodos para mejorar la especificidad de la diana de una enzima ortólogo Cas9 o el diseño de un sistema de CRISPR-Cas9 que comprende diseñar o preparar ARNs guía que tiene actividad óptima y/o seleccionar o preparar una enzima ortólogo Cas9 que tiene un tamaño o una longitud más pequeña que la correspondiente Cas9 de tipo salvaje, con lo cual se hace avanzar el empaquetamiento de un ácido nucleico que codifica la misma en un vector de suministro, ya que hay menos secuencia de codificación para la misma en el vector de suministro que para la correspondiente Cas9 de tipo salvaje y/o la generación de enzimas Cas9 quiméricas.

También se proporcionan usos en medicina de las presentes secuencias, vectores, enzimas o sistemas. También se proporcionan las mismas para uso en la edición de un gen o genoma. También se proporciona el uso de las mismas en la fabricación de un medicamento para la edición de un gen o genoma, por ejemplo el tratamiento mediante la edición del gen o genoma. También se proporcionan las presentes secuencias, vectores, enzimas o sistemas para uso en terapia.

En un aspecto adicional de la invención, una enzima de CRISPR, p. ej., una enzima Cas9 puede comprender una o más mutaciones y se pueden utilizar como una proteína de unión a ADN genérico con o sin fusión a, o estando operativamente enlazada a un dominio funcional. Las mutaciones pueden ser mutaciones introducidas artificialmente y pueden incluir, pero no se limitan a una o más mutaciones en un dominio catalítico. Ejemplos de dominios catalíticos con referencia a una enzima Cas9 pueden incluir, pero no se limitan a dominios RuvC I, RuvC II, RuvC III y HNH. Ejemplos preferidos de mutaciones adecuadas son el o los residuos catalíticos en el extremo N del dominio RuvC I de Cas9 o el o los residuos catalíticos en el dominio HNH interno. En algunas realizaciones, la Cas9 es (o se deriva de) Cas9 de *Streptococcus pyogenes* (SpCas9). En tales realizaciones, las mutaciones preferidas son en cualquiera o en todas las posiciones 10, 762, 840, 854, 863 y/o 986 de SpCas9 o posiciones correspondientes en otros ortólogos Cas9 con referencia a la numeración de la posición de SpCas9 (que puede determinarse, por ejemplo, mediante herramientas de comparación de secuencias estándar, p. ej., ClustalW o MegAlign por Lasergene 10 suite). En particular, cualquiera o todas de las siguientes mutaciones son preferidos en SpCas9: D10A, E762A, H840A, N854A, N863A y/o D986A; así como también se prevé una sustitución conservativa para cualquiera de los aminoácidos de sustitución. También se prefieren las mismas mutaciones (o sustituciones conservativas de estas mutaciones) en posiciones correspondientes con referencia a la numeración de la posición de SpCas9 en otros ortólogos Cas9. Particularmente preferidos son SpCas9 D10 y H840. Sin embargo, en otras Cas9s, también se prefieren los residuos correspondientes a SpCas9 D10 y H840. Estos son ventajosos como cuando mutados individualmente proporcionan actividad nickase, y cuando ambas mutaciones están presentes la Cas9 se convierte en un mutante catalíticamente nulo que es útil para la unión de ADN genérico. Se han identificado y caracterizado mutaciones adicionales. Otros aspectos de la invención se refieren a la enzima Cas9 mutada que está fusionada o enlazada de forma operativa a los dominios que incluyen pero no están limitados a un activador transcripcional, represor transcripcional, una recombinasa, una transposasa, una remodelador de histona, una ADN metiltransferasa, una criptocromo, un dominio inducible/contro!able por la luz o un dominio inducible/contro!able químicamente.

Un aspecto adicional de la invención proporciona proteínas Cas9 quiméricas y métodos de generar proteínas Cas9 quiméricas. Proteínas Cas9 quiméricas son proteínas que comprenden fragmentos que proceden de diferentes ortólogos Cas9. Por ejemplo, el extremo N de un primer ortólogo Cas9 puede estar fusionado con el extremo C de un segundo ortólogo Cas9 para generar una proteína Cas9 quimérica resultante. Estas proteínas Cas9 quiméricas pueden tener una especificidad mayor o una mayor eficiencia que la especificidad o eficiencia original de cualquiera de las enzimas Cas9 individuales a partir de las cuales se generó la proteína quimérica. Estas proteínas quiméricas también pueden comprender una o más mutaciones o pueden estar enlazadas a uno o más dominios funcionales. Por lo tanto, aspectos de la invención se refieren a una enzima Cas quimérica, en donde la enzima comprende uno o más fragmentos de un primer ortólogo Cas y uno o más fragmentos de un segundo ortólogo Cas. En una realización de la invención, los uno o más fragmentos del primer o segundo ortólogo Cas son del extremo C o N del primer o segundo ortólogo Cas. En una realización adicional, el primer o segundo ortólogo Cas se selecciona de un género que pertenece al grupo que consiste en *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma y Campylobacter*,

En una realización adicional, la invención proporciona métodos para generar componentes mutantes del complejo de CRISPR que comprende una enzima Cas, p. ej., un ortólogo Cas9. Los componentes mutantes pueden incluir, pero no se limitan a secuencias de tracrRNA y de emparejamiento tracr mutante o secuencias guías quiméricas mutantes que permiten mejorar el rendimiento de estos ARN en las células. También se proporciona el uso de la presente composición o la enzima en la preparación de un medicamento para la modificación de una secuencia diana.

La invención, en todavía otro aspecto, proporciona composiciones y métodos relacionados con una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende:

A) - I. una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema CRISPR-Cas, en donde la secuencia de polinucleótido comprende:
(a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota,
(b) una secuencia de emparejamiento de tracr, y
(c) una secuencia de tracr, y
II. una secuencia de polinucleótidos que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear,
en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',
en donde, cuando se transcribe, la secuenciade emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y la secuencia de polinucleótidos que codifica una enzima CRISPR es ADN o ARN,
o
(B) I. polinucleótidos que comprenden:
(a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procariota, y
(b) al menos una o más secuencias de emparejamiento tracr,
II. una secuencia de polinucleótido que codifica una enzima de CRISPR, y
III. una secuencia de polinucleótidos que comprende una secuencia tracr,
en donde cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y la secuencia de polinucleótido que codifica una enzima de CRISPR es ADN o ARN, y
en donde la enzima CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Fiaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma y Campylobacter.

La invención, en un aspecto adicional, proporciona: (A) una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende un sistema de vectores que comprende uno o más vectores que comprenden: I. un primer elemento regulador enlazado operativamente a una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema CRISPR-Cas, en donde la secuencia de polinucleótidos comprende (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota, (b) una secuencia de emparejamiento tracr, y (c) una secuencia de tracr, y II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima CRISEPR que comprende al menos una o más secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3', en donde los componentes I y II están situados en el mismo o en diferentes vectores del sistema, en donde, cuando se transcriben, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo

CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma y Campylobacter* o (B) una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que comprende I. un primer elemento regulador enlazado operativamente a (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procariota, y (b) por lo menos una o más secuencias de emparejamiento tracr II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR, y III. un tercer elemento regulador enlazado operativamente a una secuencia tracr, en el que los componentes I, II y III están situados en el mismo o diferentes vectores del sistema, en el que cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que, se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma y Campylobacter*, y en donde se aplica al menos uno de los siguientes criterios.

Los criterios son los siguientes y se apreciará que se puede aplicar cualquier número de ellos, preferiblemente 1 o más, preferiblemente 2 o más, y preferiblemente 3 o más, 4 o más, o 5 o más, o todos se pueden aplicar:
- se selecciona la enzima de CRISPR que tiene un tamaño específico y tiene una longitud de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1119 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos;
- y/o la enzima de CRISPR está truncada en comparación con la enzima de CRISPR de tipo salvaje correspondiente;
- y/o la enzima de CRISPR es una nucleasa que dirige la escisión de las dos cadenas en la ubicación de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en el lugar de la secuencia diana;
- y/o la secuencia guía comprende al menos 10, al menos 15 o al menos 20 nucleótidos;
- y/o la enzima de CRISPR esta optimizada en codones o está optimizada en codones para la expresión en una célula eucariota;
- y/o la enzima de CRISPR comprende una o más mutaciones;
- y/o la enzima de CRISPR comprende una enzima de CRISPR quimérica;
- y/o la enzima de CRISPR tiene uno o más otros atributos comentados en esta memoria.

En algunas realizaciones, la enzima de CRISPR está truncada en comparación con una enzima de CRISPR de tipo salvaje o la enzima de CRISPR se compone de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, a menos 1000-1099 aminoácidos, al menos 1100-1199 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos. En formas de realización preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.

En algunas realizaciones, la enzima de CRISPR es una nucleasa que dirige la escisión de ambas cadenas en la ubicación de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en la ubicación de la secuencia diana. En realizaciones adicionales, la enzima de CRISPR es un mutante catalíticamente nulo que es una proteína de unión a ADN genérica. En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.

En algunas realizaciones, la secuencia guía comprende al menos quince nucleótidos. En algunas realizaciones, la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota. En algunas realizaciones, la enzima de CRISPR comprende una o más mutaciones. En algunas realizaciones, la enzima de CRISPR comprende una enzima de CRISPR quimérica. En algunas realizaciones, la enzima de CRISPR tiene uno o más de otros atributos comentados en esta memoria. En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.

En determinadas realizaciones, la enzima de CRISPR comprende una o más mutaciones. La una o más mutaciones pueden estar en un dominio particular de la enzima. En una realización preferida, la una o más mutaciones pueden estar en un dominio catalítico. En una realización preferida adicional, el dominio catalítico es un dominio RuvC I,

RuvC II, RuvC III o HNH. En una realización más preferida, la una o más mutaciones están en un dominio RuvCI o HNH de la enzima de CRISPR. En una realización preferida adicional, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9 y la mutación puede estar en una o en posiciones que incluyen pero no se limitan a las posiciones que corresponden a D10A, E762A, H840A, N854A, N863A o D986A con referencia a la numeración de la posición de SpCas9 y/o es una mutación como se comenta de otra manera en esta memoria. En algunas realizaciones, la enzima de CRISPR tiene una o más mutaciones en un dominio particular de la enzima, en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida a la secuencia de tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde la enzima comprende, además, un dominio funcional. El dominio funcional puede incluir, pero no se limita a activador transcripcional, represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente.

En algunas realizaciones, el dominio funcional es el dominio activador transcripcional VP64. En algunas realizaciones, el dominio funcional es el dominio represor transcripcional KRAB. En algunas realizaciones, el dominio represor transcripcional es SID, o concatámeros de SID (es decir, SID4X). En algunas realizaciones, se proporciona una enzima de modificación epigenética, p. ej., una proteína de modificación histona o un proteína de modificación de cromatina epigenética. En algunas realizaciones, se proporciona un dominio activador, que puede ser el dominio activador P65.

Un aspecto adicional de la invención comprende métodos de modificar dos o más loci genómicos de interés. En una realización preferida de la invención, dos o más loci genómicos están diferencialmente modulados mediante la utilización de una o más enzimas de CRISPR, p. ej., dos o más ortólogos Cas9, estando enlazado operativamente cada uno de los ortólogos a uno o más dominios funcionales. En un aspecto, la invención proporciona un método para modificar dos o más loci genómicos en una célula eucariota. Por lo tanto, aspectos de la invención proporcionan un método para modular la expresión de dos o más loci genómicos de interés en un organismo, que comprende suministrar una composición que se produce de forma no natural o que se produce mediante ingeniería genética, que comprende un sistema de vector que comprende uno o más vectores que comprenden
I. un primer elemento regulador enlazado operativamente a una primera secuencia de polinucelótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la primera secuencia de polinucleótidos comprende
(i) una primera secuencia guía capaz de hibridarse con una primera secuencia diana en un primer locus genómico en una célula del organismo,
(ii) una primera secuencia de emparejamiento tracr, y
(iii) una primera secuencia de tracr, y
II. un segundo elemento regulador enlazado operativamente a una segunda secuencia de polinucelótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la segunda secuencia de polinucleótidos comprende
(i) una segunda secuencia guía capaz de hibridarse con una segunda secuencia diana en un segundo locus genómico en la célula del organismo,
(ii) una segunda secuencia de emparejamiento tracr, y
(iii) una segunda secuencia de tracr, y
III. un tercer elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una primera enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un primer dominio funcional,

IV. un cuarto elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una segunda enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un segundo dominio funcional, en donde (i), (ii) y (iii) en **I** y **II** están dispuestos en una orientación 5' a 3', en donde los componentes **I**, **II**, **III** y **IV** se encuentran en los mismos o en diferentes vectores del sistema, en donde, cuando se transcriben, cada una de las secuencias de emparejamiento tracr se hibrida con su secuencia tracr correspondiente y la primera y segunda secuencias guía dirigen la unión específica para la secuencia del primer y segundo complejos de CRISPR a la primera y segunda secuencia diana, en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y en donde la expresión de la enzima de CRISPR proporciona la manipulación de la secuencia diana, en donde la primera y segunda enzimas de CRISPR comprenden cada una dos o más mutaciones, en donde la primera y segunda enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor , Mycoplasma y Campylobacter*, y en donde el primer locus genómico es modulado por la actividad del primer dominio funcional y el segundo locus genómico es modulado por la actividad del segundo dominio funcional. En una realización adicional, el primer dominio funcional se selecciona del grupo que consiste en un activador transcripcional, represor transcripcional, una recombinasa, una transposasa, una remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/contro!able por la luz o un dominio inducible/contro!able químicamente. En una realización adicional, el segundo dominio funcional se selecciona del grupo que consiste en un activador transcripcional, represor transcripcional, una recombinasa, una

transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/contro!able por la luz o un dominio inducible/contro!able químicamente. En realizaciones preferidas, la primera o segunda enzima de CRISPR es una Cas9 de Sutterella wadsworthensis, una Cas9 de Filifactor alocis, una Cas9 de Lactobacillus johnsonii, una Cas9 de Campylobacter lari, una Cas9 de Corynebacter diptheriae, una Cas9 de Parvibaculum lavamentivorans, una Cas9 de Mycoplasma gailisepticum, una Cas9 de Staphylococcus aureus subsubespecie Aureus Cas9, una Cas9 de Legionella pneumophila Paris, una Cas9 de Treponema denticola, una Cas9 de Staphylococcus pseudintermedius, una Cas9 de Neisseria cinerea.

En algunas realizaciones, la enzima de CRISPR es una enzima de CRISPR de tipo I, II o III, preferentemente una enzima CRISPR de tipo II. Esta enzima de CRISPR de tipo II puede ser cualquier enzima Cas. Una enzima Cas puede ser identificada como Cas9, ya que esto puede referirse a la clase general de enzimas que comparten homología con el mayor nucleasa con múltiples dominios nucleasa del sistema de CRISPR tipo II. Lo más preferiblemente, la enzima Cas9 es, o se deriva de SpCas9 o *Staphylococcus aureus* subsubespecie Aureus SaCas9. Por se deriva se entiende que la enzima derivada se basa en gran medida, en el sentido de tener un alto grado de homología de secuencia con, una enzima de tipo salvaje, pero que ha sido mutada (modificada) de alguna manera tal como se describe en esta memoria.

Se apreciará que las expresiones enzima Cas y CRISPR se utilizan en general en esta memoria de forma indistinta, a menos que de otro modo resulte evidente. Tal como se mencionó anteriormente, muchas de las numeraciones de residuos utilizadas en esta memoria se refieren a la enzima Cas9 del locus CRISPR de tipo II en *Streptotoccus pyogenes*. Sin embargo, se apreciará que esta invención incluye muchas más Cas9s de otras especies de microbios tales como los pertenecientes al género *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma o Campylobacter*, tal como SpCas9, SaCas9, St1Cas9, St3Cas9 y así sucesivamente, en donde St es *Streptococcus thermophilus*.

Un ejemplo de una secuencia optimizada en codones, en este caso optimizada para seres humanos (es decir, optimizada para la expresión en seres humanos) se proporciona en esta memoria, ver la secuencia optimizada en codones humana SaCas9. Si bien esto se prefiere, se apreciará que son posibles otros ejemplos y se conoce la optimización de codones para una especie huésped.

Aspectos adicionales de la invención se refieren a la especificidad de escisión mejorada, la secuencia tracr optimizada, el ARN guía quimérico optimizado, la estructura co-plegada de tracrRNA y la secuencia de emparejamiento tracr, la estabilización de estructuras secundarias de tracRNA, tracrRNA con una región acortada de apareamiento de bases, tracrRNA con elementos de ARN fusionados, la clonación y el suministro simplificados, sistemas de toxicidad reducida y/o inducibles. Otro aspecto de la invención se refiere a la estabilización de ARN quimérico, y/o secuencia guía y/o una parte del mismo de complejos de CRISPR, en donde la enzima de CRISPR es un ortólogo de CRISPR, en donde el ARN quimérico y/o la secuencia guía y/o una parte del mismo se estabiliza mediante nucleótidos sintéticos o modificados químicamente (p. ej., LNA/BNA; modificación de tiol, modificación de reticulación 2'/3'-OH), se modifica para ser resistentes a la degradación/hidrólisis y al que se han añadido elementos de estabilidad estructural.

La invención comprende, además, en determinadas realizaciones un método de modificar un organismo o un organismo no humano mediante la manipulación de una secuencia diana en un locus genómico de interés, que comprende suministrar una composición que se produce de forma no natural o que se produce mediante ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que codifican operativamente una composición comentada en esta memoria para la expresión de los mismos. Preferiblemente, el vector es un vector viral, tal como vectores virales lenti- o baculo- o preferiblemente adeno-asociados, pero se conocen y se proporcionan otros medios de suministro (tales como sistemas de levadura, microvesículas, pistolas de genes/medios de fijación de vectores a nanopartículas de oro).

En esta memoria se describe diversos medios de suministro y se comentan en esta sección.

Suministro Viral: La enzima de CRISPR, por ejemplo una Cas9, y/o cualquiera de los presentes ARN, por ejemplo, un ARN guía, puede ser suministrada utilizando virus adeno-asociado (AAV), lentivirus, adenovirus u otros tipos de vectores virales, o combinaciones de los mismos. Cas9 y uno o más ARN guía pueden ser empaquetados en uno o más vectores virales. En algunas realizaciones, el vector viral es suministrado al tejido de interés, por ejemplo, mediante inyección intramuscular, mientras que otras veces el suministro viral es a través de la vía intravenosa, transdérmica, intranasal, oral, mucosal, u otros métodos de suministro. Una administración de este tipo puede ser o bien a través de una sola dosis o de múltiples dosis. Un experto en la técnica entiende que la dosis real a suministrar en esta memoria pueden variar en gran medida dependiendo de una diversidad de factores tales como el vector elegido, la célula, organismo o tejido diana, el estado general del sujeto a tratar, el grado de

transformación/modificación buscado, la vía de administración, el modo de administración, el tipo de de transformación/modificación buscado, etc.

Una dosificación de este tipo puede contener, además, por ejemplo, un soporte (agua, solución salina, etanol, glicerol, lactosa, sacarosa, fosfato de calcio, gelatina, dextrano, agar, pectina, aceite de cacahuete, aceite de sésamo, etc.), un diluyente, un soporte farmacéuticamente aceptable (p. ej., solución salina tamponada con fosfato), un excipiente farmacéuticamente aceptable, un adyuvante para potenciar la antigenicidad, un compuesto o molécula inmunoestimulador y/u otros compuestos conocidos en la técnica. El adyuvante en esta memoria puede contener una suspensión de minerales (alumbre, hidróxido de aluminio, fosfato de aluminio) sobre los que se adsorbe el antígeno, o una emulsión de agua-en-aceite en la que la disolución de antígeno se emulsiona en aceite (MF-59, adyuvante incompleto de Freund), a veces con la inclusión de micobacterias muertas (adyuvante completo de Freund) para mejorar adicionalmente la antigenicidad (inhibe la degradación del antígeno y/o provoca la afluencia de macrófagos). Los adyuvantes también incluyen moléculas inmunoestimulantes tales como citoquinas, moléculas coestimulantes y, por ejemplo, moléculas de ADN o de ARN inmunoestimulantes tales como oligonucleótidos CpG. Una formulación de dosificación de este tipo es fácilmente determinable por un experto en la técnica. La dosificación puede contener, además, una o más sales farmacéuticamente aceptables tales como, por ejemplo, una sal de ácido mineral tal como un hidrocloruro, un hidrobromuro, un fosfato, un sulfato, etc.; y las sales de ácidos orgánicos tales como acetatos, propionatos, malonatos, benzoatos, etc. Adicionalmente, también pueden estar presentes sustancias auxiliares tales como agentes humectantes o emulsionantes, sustancias tamponadoras del pH, geles o materiales gelificantes, aromas, colorantes, microesferas, polímeros, agentes de suspensión, etc. Además, también pueden estar presentes uno o más de otros ingredientes farmacéuticos convencionales tales como conservantes, humectantes, agentes de suspensión, tensioactivos, antioxidantes, agentes antiapelmazantes, cargas, agentes quelantes, agentes de revestimiento, estabilizadores químicos, etc., especialmente si la forma de dosificación es una forma reconstituible. Ingredientes a modo de ejemplo adecuados incluyen celulosa microcristalina, carboximetilcelulosa de sodio, polisorbato 80, alcohol feniletílico, clorobutanol, sorbato de potasio, ácido sórbico, dióxido de azufre, galato de propilo, los parabenos, vainillina de etilo, glicerol, fenol, paraclorofenol, gelatina, albúmina y una combinación de los mismos. Una discusión a fondo de excipientes farmacéuticamente aceptables está disponible en REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991).

En una realización en esta memoria, el suministro es a través de un adenovirus, que puede ser en una dosis de refuerzo única que contiene al menos 1 x 10^5 partículas (a la que también se alude como unidades de partículas, pu) de vector adenoviral. En una realización en esta memoria, la dosis es preferiblemente de al menos aproximadamente 1 x 10^6 partículas (por ejemplo, aproximadamente 1 x 10^6-1 x 10^{12} partículas), más preferiblemente al menos aproximadamente 1 x 10^7 partículas, más preferiblemente al menos aproximadamente 1 10^8 partículas (p. ej., aproximadamente 1 x 10^8 -1 x 10^{11} partículas o aproximadamente 1 x 10^8- 1 x 10^{12} partículas), y lo más preferiblemente al menos aproximadamente 1 x 10^0 partículas (p. ej., aproximadamente 1* x 10^9 - 1 x 10^{10} partículas o aproximadamente 1 x 10^9 -1 x 10^{12} partículas), o incluso al menos aproximadamente 1 x 10^{10} partículas (p. ej., aproximadamente 1 x 10^{10} - 1 x 10^{12} partículas) del vector adenoviral. Alternativamente, la dosis comprende no más de aproximadamente 1 x 10^{14} partículas, preferiblemente no más de aproximadamente 1 x 10^{13} partículas, incluso más preferiblemente no más de aproximadamente 1 x 10^{12} partículas, incluso más preferiblemente no más de aproximadamente 1 x 10^{11} partículas, y lo más preferiblemente no más de aproximadamente 1 x 10^{10} partículas (p. ej., no más de aproximadamente 1 x 10^9 partículas). Por lo tanto, la dosis puede contener una sola dosis de vector adenoviral con, por ejemplo, aproximadamente 1 x 10^6 unidades de partículas (pu), aproximadamente 2 x 10^6 pu, aproximadamente 4 x 10^6 pu, aproximadamente 1 x 10^7 pu, aproximadamente 2 x 10^7 pu, aproximadamente 4 x 10^7 pu, aproximadamente 1 x 10^8 pu, aproximadamente 2 x 10^8 pu, aproximadamente 4 x 10^8 pu, aproximadamente 1 x 10^9 pu, aproximadamente 2 x 10^9 pu, aproximadamente 4 x 10^9 pu, aproximadamente 1 x 10^{10} pu, aproximadamente 2 x 10^{10} pu, aproximadamente 4 x 10^{10} pu, aproximadamente 1 x 10^{11} pu, aproximadamente 2 x 10^{11} pu, aproximadamente 4 x 10^{11} pu, aproximadamente 1 x 10^{12} pu, aproximadamente 2 x 10^{12} pu, aproximadamente 4 x 10^{12} pu, del vector adenoviral. Véanse, por ejemplo, los vectores adenovirales en la patente de EE.UU. Nº 8.454.972 B2 expedida a Nabel, et. al., concedida el 4 de junio de 2013, y las dosis en la col. 29, líneas 36-58 de la misma. En una realización en esta memoria, el adenovirus se suministra a través de múltiples dosis.

En una realización en esta memoria, el suministro es a través de un AAV. Una dosis terapéuticamente eficaz para el suministro in vivo del AAV a un ser humano se cree que está en el intervalo de aproximadamente 20 a aproximadamente 50 ml de solución salina que contiene de aproximadamente 1 x 10^{10} a aproximadamente 1 x 10^{10} AAV funcional/ml de disolución. La dosificación puede ser ajustada para equilibrar el beneficio terapéutico contra efectos secundarios. En una realización en esta memoria, la dosis de AAV está generalmente en el intervalo de concentraciones de aproximadamente 1 x 10^5 hasta 1 x 10^{10} genomas de AAV, de aproximadamente 1 x 10^8 a 1 x 10^{20} genomas de AAV, de aproximadamente 1 x 10^{10} a aproximadamente 1 x 10^{16} genomas, o aproximadamente 1 x 10^{11} a aproximadamente 1 x 10^{16} genomas de AAV. Una dosificación humana puede ser de aproximadamente 1 x 10^{13} genomas de AAV. Tales concentraciones pueden ser suministradas en de aproximadamente 0,001 ml a aproximadamente 100 ml, DE aproximadamente 0,05 a aproximadamente 50 ml o de aproximadamente 10 a aproximadamente 25 ml de una disolución de soporte. Otras dosificaciones eficaces pueden establecerse fácilmente por un experto ordinario en la técnica a través de ensayos de rutina que establecen curvas de dosis-respuesta.

Véase, por ejemplo, la patente de EE.UU. Nº 8.404.658 B2 expedida a Hajjar, et al., concedida el 26 de marzo de 2013, en la col. 27, líneas 45-60.

En una realización en esta memoria, el suministro es a través de un plásmido. En tales composiciones de plásmidos, la dosis debe ser una cantidad suficiente de plásmido para inducir una respuesta. Por ejemplo, cantidades adecuadas de ADN de plásmido en composiciones de plásmidos puede ser de aproximadamente 0,1 a aproximadamente 2 mg, o de aproximadamente 1 µg a aproximadamente 10 µg.

Las dosis en esta memoria se basan en una media de un individuo de 70 kg. La frecuencia de administración está dentro del ámbito del profesional médico o veterinario (p. ej., médico, veterinario), o un científico experto en la técnica.

Los vectores virales pueden ser inyectados en el tejido de interés. Para la modificación del genoma específico para el tipo de célula, la expresión de Cas9 puede ser impulsada por un promotor específico de tipo celular. Por ejemplo, la expresión específica para el hígado podría utilizar el promotor de albúmina y la expresión específica para las neuronas podría utilizar el promotor Sinapsina I.

Suministro de ARN: La enzima de CRISPR, por ejemplo, una Cas9, y/o cualquiera de los presentes ARN, por ejemplo un ARN guía, también se puede suministrar en forma de ARN. ARNm de Cas9 se pueden generar utilizando la transcripción in vitro. Por ejemplo, ARNm de Cas9 se puede sintetizar utilizando una casete de PCR que contiene los siguientes elementos: secuencia T7_promotor-kozak (GCCACC)-Cas9-3' UTR de la cola beta globina-poliA (una cadena de 120 o más adeninas). La casete se puede utilizar para la transcripción por la T7polimerasa. ARNs guía también se pueden transcribir utilizando la transcripción in vitro a partir de una casete que contiene la secuencia T7_promotor-GG-ARN guía.

Para potenciar la expresión y reducir la toxicidad, la enzima CRISPR y/o el ARN guía se pueden modificar utilizando pseudo-U o 5-Metil-C.

ARNm de la enzima de CRISPR y ARN guía se pueden administrar simultáneamente utilizando nanopartículas o envueltas lipídicas.

Por ejemplo, Su X, Fricke I, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivey using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm 6 de junio de 2011; 8(3):774-87 doi:10.1021/mp100390w Epub 1 de abril de 2011) describe nanopartículas de núcleo y envoltura estructuradas biodegradables con un núcleo poli (β-amino éster) (PBAE) envuelto por una envoltura bicapa de fosfolípidos. Estos fueron desarrollados para el suministro in vivo de ARNm. El componente PBAE sensible al pH fue escogido para fomentar la interrupción del endosoma, mientras que la capa de la superficie de lípidos se seleccionó para minimizar la toxicidad del núcleo de policationes. Por lo tanto, éstos se prefieren para el suministro de ARN de la presente invención.

Además de ello, Michael S.D. Kormann et al. ("Expression of therapeutic proteins after delivey of chemically modified mRNA in mice: Nature Biotechnology, Volumen: 29, páginas: 154-157 (2011) publicado en línea el 9 de enero de 2011) describen el uso de envolturas de lípidos para suministrar ARN. El uso de envolturas de lípidos también se prefiere en la presente invención.

Métodos de suministro de ARNm son especialmente prometedores para suministrar actualmente al hígado.

ARNm de la enzima de CRISPR ARNguía también podrían ser suministrados por separado. ARNm de la enzima de CRISPR puede ser suministrado antes del ARN guía para dar tiempo a que la enzima de CRISPR se exprese. ARNm de la enzima de CRISPR puede ser administrada 1-12 horas (preferiblemente alrededor de 2-6 horas) antes de la administración de ARNguía.

Alternativamente, ARNm de la enzima de CRISPR y ARN guía se pueden administrar juntos. Ventajosamente, una segunda dosis de refuerzo de ARN guía se puede administrar 1 -12 horas (preferiblemente alrededor de 2-6 horas) después de la administración inicial de ARNm de la enzima de CRISPR + ARNguía.

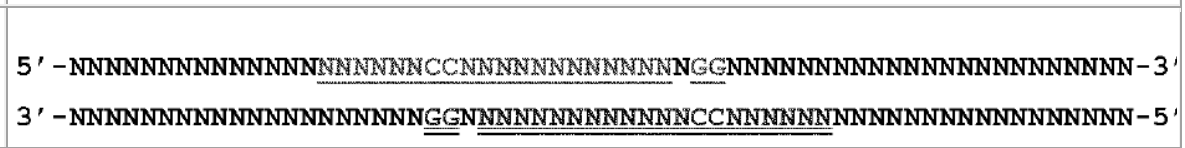
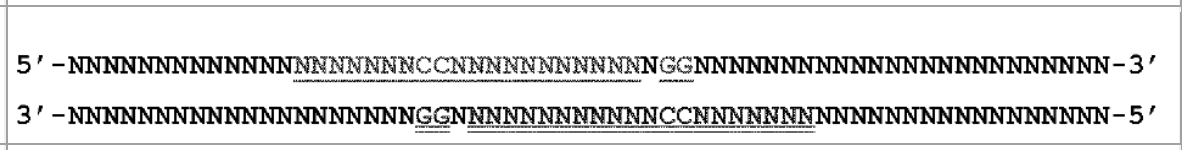
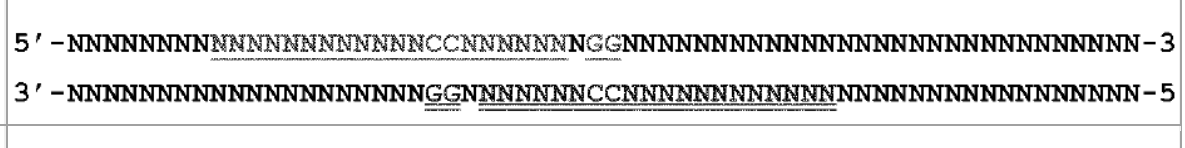
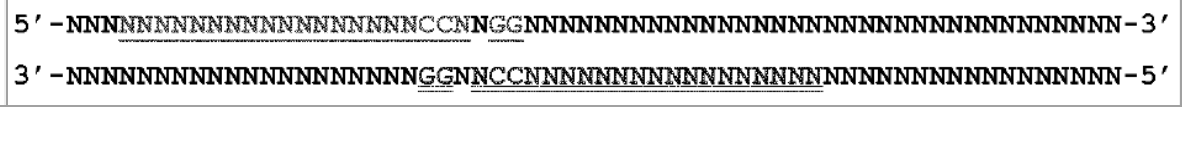
Administraciones adicionales de ARNm de la enzima de CRISPR y/o ARN guía podrían ser útiles para alcanzar los niveles más eficientes de modificación del genoma.

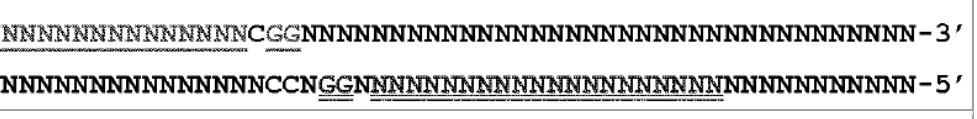
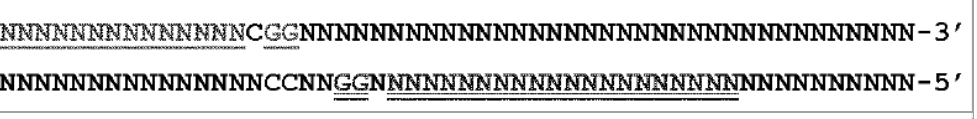
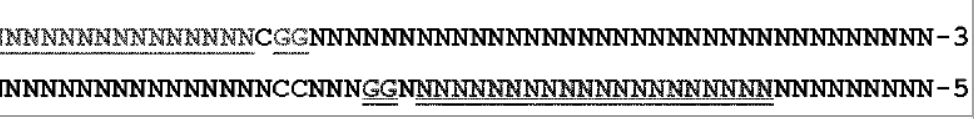
Para la minimización de la toxicidad y los efectos fuera de objetivo, será importante controlar la concentración de ARNm de la enzima de CRISPR y el ARN guía suministrado. Concentraciones óptimas del ARNm de la enzima de CRISPR y ARN guía se pueden determinar mediante el ensayo de diferentes concentraciones en un modelo celular

o animal y se puede utilizar la secuenciación profunda para analizar el grado de modificación en potenciales loci genómicos fuera de objetivo. Por ejemplo, para la secuencia guía que fija como objetivo 5'-GAGTCCGAGCAGAAGAAGAA-3' en el gen EMX1 del genoma humano, la secuenciación profunda se puede utilizar para evaluar el nivel de modificación en los dos loci siguientes fuera de objetivo, 1: 5'-GAGTCCTAGCAGGAGAAGAA-3' y 2: 5'-GAGTCTAAGCAGAAGAAGAA-3'. La concentración que da el mayor nivel de modificación en objetivo al tiempo que minimiza el nivel de modificación fuera de objetivo debe ser elegido para el suministro in vivo.

Alternativamente, para minimizar el nivel de toxicidad y los efectos fuera de objetivo, la enzima de CRISPR nickase ARNm (por ejemplo Cas9 de S pyogenes con la mutación D10A) puede ser suministrada con un par de ARNs guía que fijan como objetivo un sitio de interés. Los dos ARN guías tienen que estar separados de la siguiente manera. Secuencias guía en rojo (subrayado sencillo) y azul (subrayado doble), respectivamente (estos ejemplos se basan en el requisito del PAM para Cas9 de Streptococcus pyogenes).

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
14	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
13	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
12	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
11	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5
10	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5
9	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
8	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
7	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3′ 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5′
6	5′-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN-3 3′-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNN-5

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
5	
4	
3	
2	
1	
romo	
1	
2	
3	
4	
5	
6	

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
7	
8	
12	
13	
14	
15	
16	
17	

La interrogación adicional del sistema ha dado a la solicitante evidencia del colgante 5' (véase, p. ej., Ran et al, Cell. 12 de septiembre de 2013;154(6):1380-9 y Solicitud de Patente Provisional de EE.UU. Nº de Serie 61/871.301 presentada el 28 de agosto de 2013). La solicitante ha identificado, además, parámetros que se refieren a la escisión eficaz por parte del mutante Cas9 nickase cuando se combina con dos ARNs guía y estos parámetros incluyen, pero no se limitan a la longitud del colgante 5'. En realizaciones de la invención, el colgante 5' es a lo sumo de 200 pares de bases, preferiblemente a lo sumo de100 pares de bases, o más preferiblemente a lo sumo de 50 pares de bases. En realizaciones de la invención, el colgante 5' es de al menos 26 pares de bases, preferiblemente de al menos 30 pares de bases o más preferiblemente de 34 a 50 pares de bases o de 1-34 pares de bases. En otros métodos preferidos de la invención, la primera secuencia guía que dirige la escisión de una cadena del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia de guía dirige la escisión de otra cadena cerca de la segunda secuencia diana da como resultado un corte romo o un colgante 3'. En realizaciones de la invención, el colgante 3'

es a lo sumo de 150, 100 ó 25 pares de bases o al menos de 15, 10 ó 1 pares de bases. En realizaciones preferidas, el colgante 3' es de 1-100 pares de bases.

Aspectos de la invención se refieren a la expresión del producto génico que se encuentra disminuido o siendo introducido adicionalmente un polinucleótido molde en la molécula de ADN que codifica el producto génico o siendo escindida una secuencia intermedia precisamente al permitir que los dos colgantes 5' se reasocien o liguen o siendo alterada la actividad o función del producto génico, o siendo aumentada la expresión del producto génico. En una realización de la invención, el producto génico es una proteína.

Sólo los pares de sgRNA que crean colgantes 5' con menos de 8 pb se solapan entre las secuencias guía (compensación mayor que -8 pb) fueron capaces de mediar en la formación de indel detectable. De manera importante, cada una de las guías utilizadas en estos ensayos es capaz de inducir de manera eficiente indeles cuando se combina con Cas9 de tipo salvaje, lo que indica que las posiciones relativas de los pares de guía son los parámetros más importantes en la predicción de la actividad de doble mella.

Dado que Cas9n y Cas9H840A mellan cadenas opuestas de ADN, la sustitución de Cas9n con Cas9H840A con un par de sgRNA dados debería dar lugar a la inversión del tipo de colgante. Por ejemplo, un par de sgRNAs que generará un colgante 5' con Cas9n debería generar, en principio, el correspondiente colgante 3' en su lugar. Por lo tanto, los pares de sgRNA que conducen a la generación de un colgante 3' con Cas9n podrían utilizarse con Cas9H840A para generar un colgante 5'. Inesperadamente, la solicitante sometieron a ensayo Cas9H840A con un conjunto de pares de sgRNA diseñados para generar colgantes tanto 5' como 3' (intervalo de compensaciones de -278 a +58 pb), pero no fueron capaces de observar la formación de indeles. Puede ser necesario un trabajo adicional para identificar las reglas de diseño necesarias para el apareamiento de sgRNA para permitir un doble mellado por Cas9H840A.

Opciones de suministro adicionales al cerebro incluyen la encapsulación de la enzima de CRISPR y ARN guía en forma de ADN o ARN en los liposomas y la conjugación a caballos de Troya moleculares para el suministro a la barrera hematoencefálica (BBB). Caballos de Troya moleculares han demostrado ser eficaces para el suministro de vectores de expresión B-gal en el cerebro de primates no humanos. El mismo enfoque se puede utilizar para el suministro de vectores que contienen enzima de CRISPR y ARN guía. Por ejemplo, Xia CF y Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology" Mol Pharm. mayo-junio de 2009; 6(3):747-51 doi : 10.1021/mp800194) describe cómo es posible la entrega de ARN corto de interferencia (siRNA) a células en cultivo, e in vivo, con el uso combinado de un anticuerpo monoclonal (mAb) específico para el receptor y la tecnología avidina-biotina. Los autores también informan que, debido a que la unión entre el mAb de fijación como objetivo y el siRNA es estable con la tecnología de avidina-biotina, y se observan efectos de RNAi en sitios distantes tales como el cerebro in vivo después de la administración intravenosa del siRNA fijado como objetivo.

Zhang Y, Schlachetzki F, Pardridge WM. ("Global non-viral gene transfer to the primate brain following intravenous administration" Mol Ther. enero de 2003; 7(1):11-8.) describen cómo la expresión de plásmidos que codifican informadores tales como luciferasa fueron encapsulados en el interior de un "virus artificial" constituido por inmunoliposoma 85 nm pegilado, que fue fijado como objetivo al cerebro de mono rhesus in vivo con un anticuerpo monoclonal (MAb) al receptor de la insulina huma (HIR). El HIRMAb permite que el liposoma que porta el gen exógeno se someta a transcitosis a través de la barrera sangre-cerebro y la endocitosis a través de la membrana plasmática neuronal después de la inyección intravenosa. El nivel de expresión del gen de luciferasa en el cerebro era 50 veces más alto en el mono rhesus en comparación con la rata. La expresión neuronal general del gen beta-galactosidasa en el cerebro de los primates fue demostrada tanto por histoquímica como por microscopía confocal. Los autores indican que este enfoque hace factibles transgénicos adultos reversibles en 24 horas. Por consiguiente, se prefiere el uso de inmunoliposoma. Éste se puede utilizar en unión con anticuerpos para fijar como objetivo tejidos diana específicos o proteínas de la superficie celular.

También se prefieren otros medios de suministro o de ARN tales como a través de nanopartículas (Clio, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R, y Anderson, D., Lipid-like nanoparticles for small interfering RNA delivey to endotelial cells, Advanced Functional Materials, 19:3112-31 18, 2010) o exosomas (Schroeder, A., Levins, C., Cortez, C., Langer, R. y Anderson, D., Lipid-based nanotherapeutics fro siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). De hecho, los exozomas han demostrado ser particularmente útiles en el suministro de siRNA, un sistema con algunos paralelismos con el sistema de CRISPR. Por ejemplo, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo" Nat Protoc. diciembre de 2012;7(12):2112-26 doi: 10.1038/nprot.2012.131 Epub 15 de noviembre de 2012) describen cómo los exosomas son herramientas prometedoras para el suministro de fármacos a través de diferentes barreras biológicas y pueden ser aprovechados para el suministro de siRNA in vitro e in vivo. Su enfoque es generar exosomas fijados como objetivo a través de transfección de un vector de expresión, que comprende una proteína exosomai fusionada con un ligando de péptido. Los exosomas son entonces purificados y caracterizados a partir del sobrenadante de células transfectadas y, a continuación, el siRNA se carga en los exosomas.

Un aspecto de la manipulación de una secuencia diana también se refiere a la manipulación epigenética de una secuencia diana. Esto puede ser del estado de cromatina de una secuencia diana tal como por modificación del estado de metilación de la secuencia diana (es decir, adición o eliminación de metilación o patrones de metilación o islas CpG), modificación de histona, el aumento o la reducción de la accesibilidad a la secuencia diana, o mediante la promoción o la reducción de plegado 3D.

Un vector puede significar no sólo un sistema viral o de levadura (por ejemplo, en los casos en los que los ácidos nucleicos de interés pueden estar operativamente enlazados a y bajo el control de (en términos de expresión, tal como para proporcionar en última instancia un ARN procesado) un promotor, sino que también un suministro directo de ácidos nucleicos a una célula huésped.

La invención comprende también, en determinadas realizaciones, un método para tratar o inhibir una afección provocada por un defecto en una secuencia diana en un locus genómico de interés en un sujeto o un sujeto no humano en necesidad del mismo, que comprende la modificación del sujeto o del sujeto no humano mediante la manipulación de la secuencia diana y en el que la afección es susceptible de tratamiento o la inhibición por manipulación de la secuencia diana que comprende proporcionar un tratamiento que comprende: suministrar una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que comprende codificar de manera operativa una composición de la presente memoria comentada para la expresión de los mismos, en el que la secuencia diana es manipulada por la composición cuando se expresa.

En determinadas realizaciones de los métodos de esta memoria, los métodos pueden incluir la expresión de la inducción, que puede ser inducir la expresión de la enzima de CRISPR y/o inducir la expresión de las secuencias guía, tracr o de emparejamiento tracr. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es un eucariota o un eucariota no humano. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es una planta. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es un mamífero o un mamífero no humano. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto son algas.

Mientras que en los métodos de esta memoria el vector puede ser un vector viral y éste es ventajosamente un AAV, se pueden emplear otros vectores virales tal como se describen en esta memoria. Por ejemplo, se pueden utilizar baculovirus para la expresión en células de insectos. Estas células de insectos pueden, a su vez, ser útiles para producir grandes cantidades de otros vectores, tales como vectores de AAV adaptados para el suministro de la presente invención.

Asimismo, se prevé un método para suministrar la presente enzima de CRISPR, que comprende suministrar a la célula un ARNm que codifican la enzima de CRISPR. Se apreciará que la enzima de CRISPR está truncada, es de un tamaño específico según se describe en esta memoria, es una nucleasa o nickase o una proteína de unión a ADN genérico, está optimizada en codones, comprende una o más mutaciones y/o comprende una enzima de CRISPR quimérica, o las otras opciones como se discuten en esta memoria.

Asimismo, se prevé un método para preparar un vector para el suministro de las composiciones o las presentes enzimas de CRISPR de la invención y para su uso en los presentes métodos.

Se prefieren los vectores virales AAV. Por lo tanto, en un aspecto adicional, se proporciona un método de preparar un vector viral AAV, que comprende transfectar un plásmido o plásmidos que contienen o que consisten esencialmente en una molécula o moléculas de ácido nucleico que codifican el AAV en células de infectadas con AAV, y suministrar AAV rep y/o cap obligatorio para la replicación y el empaquetamiento del AAV. A este respecto, se apreciará que la enzima de CRISPR está truncada, está compuesto por menos de mil aminoácidos o menos de cuatro mil aminoácidos, es una nucleasa o nickase, está optimizada en codones, comprende una o más mutaciones y/o comprende una enzima de CRISPR quimérica tal como se comenta en esta memoria. En algunas realizaciones, el AAV rep y/o cap obligatorio para la replicación y el empaquetamiento del AAV se suministra mediante la transfección de las células con plásmido o plásmidos helper (cooperadores). En algunas realizaciones, el virus helper es un virus de la viruela, adenovirus, virus herpes o baculovirus. En algunas realizaciones, el virus de la viruela es un virus vacuna. En algunas realizaciones las células son células de mamíferos. Y en algunas realizaciones las células son células de insectos y el virus helper es baculovirus.

La invención comprende además en determinadas realizaciones una enzima de CRISPR modificada. Diferencias con la enzima de CRISPR de tipo salvaje pueden comprender: la enzima de CRISPR modificada está truncada en comparación con una enzima de CRISPR de tipo salvaje, o la enzima de CRISPR es de un tamaño específico, p. ej., de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1199 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699

aminoácidos o al menos 2000 aminoácidos; y/o la enzima de CRISPR es una nucleasa que dirige la escisión de ambas cadenas en la ubicación de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en el lugar de la secuencia diana, o la enzima de CRISPR es un mutante catalítico nulo que funciona como una proteína de unión de ADN; y/o la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota, y/o la enzima de CRISPR comprende una o más mutaciones y/o la enzima de CRISPR comprende una enzima de CRISPR quimérica y/o la enzima de CRISPR tiene uno o más de otros atributos comentados en la presente memoria. Por consiguiente, en determinadas realizaciones, la enzima de CRISPR comprende una o más mutaciones en residuos catalíticos tales como spCa9 D10A, E762A, H840A, N854A, N863A o D986A o los correspondientes a ellos en otras enzimas Cas (tal como se describe en esta memoria) con referencia a la numeración de la posición de SpCas9, y/o tiene una o más mutaciones se encuentra en un dominio RuvC I, RuvC II, RuvC III, HNH u otro dominio descrito en esta memoria de la enzima de CRISPR y/o la enzima de CRISPR tiene una o más mutaciones en un dominio catalítico, en el que, cuando se transcribe, una secuencia de emparejamiento tracr se hibrida con una secuencia tracr y una secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a una secuencia diana, y en donde la enzima comprende, además, un dominio funcional. El dominio funcional puede incluir, pero no se limita a activador transcripcional, represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente. La enzima de CRISPR en determinadas realizaciones puede tener el dominio funcional que sea un dominio activador transcripcional, p. ej., VP64. En algunas realizaciones, un dominio represor transcripcional es KRAB. En algunas realizaciones, un dominio represor de la transcripción es SID, o concatámeros de SID (es decir SID4X). En algunas realizaciones, se proporciona una enzima modificadora epigenética.

En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.

Aspectos de la invención también comprenden identificar nuevos ortólogos de enzimas de CRISPR. Los métodos para identificar nuevos ortólogos de enzimas CRISPR pueden implicar identificar secuencias tracr en los genomas de interés. La identificación de secuencias tracr puede estar relacionada con los siguientes pasos: Búsqueda de las repeticiones directas o secuencias de emparejamiento tracr en una base de datos para identificar una región CRISPR que comprende una enzima de CRISPR, p. ej., Figura 18. Búsqueda de secuencias homólogas en la región de CRISPR que flanquean la enzima de CRISPR tanto en las direcciones sentido como antisentido. Búsqueda de terminadores de la transcripción y estructuras secundarias. Identificar cualquier secuencia que no sea una repetición directa o una secuencia de emparejamiento tracr, pero que tiene más de un 50% de identidad con la repetición directa o la secuencia de emparejamiento como una secuencia tracr potencial. Tomar la secuencia tracr potencial y analizar las secuencias de terminador de la transcripción asociados con la misma.

La invención comprende en determinadas realizaciones el uso de una composición como se describe en esta memoria o una enzima de CRISPR en la medicina. La invención comprende además, en determinadas realizaciones una composición o enzima de CRISPR de la invención utilizada en un método de la invención. La invención comprende también, en determinadas realizaciones, el uso de una composición o una enzima de CRISPR de la invención, preferiblemente ex vivo, en la edición de un gen o genoma o, preferiblemente un método ex vivo para la edición de un gen o genoma. La invención también comprende en determinadas realizaciones el uso de una composición de acuerdo con o una enzima de CRISPR de la invención en la fabricación de un medicamento para la edición ex vivo de un gen o genoma o para uso en un método tal como se comenta en esta memoria.

Además, la invención, en determinadas realizaciones, comprende una composición o una enzima de CRISPR de la invención, en donde la secuencia diana está flanqueada o es seguida, en su extremo 3', por un PAM adecuado para la enzima de CRISPR, típicamente una Cas y, en particular, una Cas9. Esta secuencia de PAM es específica para cada una de las Cas9, pero se puede determinar fácilmente por métodos descritos en esta memoria.

Por ejemplo, una PAM adecuada es 5'-NRG o 5'-NNGRR para las enzimas (o derivados de enzimas) SpCas9 o SaCas9, respectivamente. Se apreciará que la referencia hecha en esta memoria a *Staphylococcus aureus* incluye preferiblemente *Staphylococcus aureus subespecie aureus*.

Por consiguiente, es un objeto de la invención consiste no abarcar dentro de la invención alguno producto, proceso de fabricar el producto o método de utilizar el producto de forma conocida con anterioridad, de manera que la solicitante se reserva el derecho y, con ello, revela un descargo de responsabilidad de cualquier producto, proceso o método previamente conocido. Se señala, además, que la invención no pretende abarcar dentro del alcance de la invención cualquier producto, proceso o fabricación del producto o método de utilización del producto, que no cumpla con los requisitos de la descripción escrita y de habilitación de USPTO (35 USC § 112, párrafo primero) o la OEP (artículo 83 del CPE), de tal manera que la solicitante se reserva el derecho y con ello revela un descargo de responsabilidad de cualquier producto, proceso de fabricar el producto o el método de utilizar el producto descrito anteriormente.

Hay que señalar que en esta descripción y en particular en las reivindicaciones y/o párrafos, términos y expresiones tales como "comprende", "que comprende", "comprendiendo" y similares pueden tener el significado que se le atribuye en la ley de patentes de Estados Unidos; p. ej., puede significar "incluye", "incluido", "que incluye", y similares; y que términos tales como "que consiste esencialmente en" y "consiste esencialmente en" tienen el significado que se les atribuye en la ley de patentes de Estados Unidos, p. ej., permiten elementos no citados explícitamente, pero excluyen elementos que se encuentran en la técnica anterior o que afectan a una base o característica nueva de la invención.

Estas y otras realizaciones se describen o son evidentes a partir de y quedan abarcadas por la siguiente descripción detallada.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las características nuevas de la invención se exponen con particularidad en las reivindicaciones adjuntas. Una mejor comprensión de las características y ventajas de la presente invención se obtendrá haciendo referencia a la siguiente descripción detallada que recoge realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos de los cuales:

La **Figura 1** muestra un esquema de nucleasa Cas9 guiada por ARN. La nucleasa Cas9 de Streptococcus pyogenes (amarillo) fija como objetivo el ADN genómico mediante un ARN guía sintético (SgRNA) que consiste en una secuencia guía de 20 nt y un armazón. Los pares de bases de la secuencia guía la diana de ADN, directamente aguas arriba de un motivo adyacente al protoespaciador 5'-NGG requerido (PAM) y Cas9 media en una rotura de doble cadena (DSB) ~ 3 pb aguas arriba del PAM (indicado por un triángulo).

Las **Figuras 2A-2F** muestran un sistema CRISPR a modo de ejemplo y un posible mecanismo de acción (A), una adaptación de ejemplo para la expresión en células eucariotas, y los resultados de las pruebas que evalúan la localización nuclear y la actividad de CRISPR (B-F).

La **Figura 3** muestra una representación esquemática de plásmidos de suministro in vivo de AAV utilizando secuencias de repeticiones terminales invertidas (ITRs) y ARNs quiméricos guía para preferiblemente ayudar en el suministro por AAV o sistemas asociados a AAV.

Las **Figuras 4A-D** muestran una representación circular del análisis filogenético revelando cinco familias de Cas9s, incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos de Cas9 pequeñas (~ 1100 aminoácidos).

Las **Figura 5 A-F** muestran una representación lineal del análisis filogenético que revelan cinco familias de Cas9s, incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos de Cas9 pequeñas (~ 1100 aminoácidos).

La **Figura 6** muestra un gráfico que representa la distribución de la longitud de ortólogos Cas9.

La **Figura 7** muestra una representación de la logotipos de la secuencia de los PAMs de los ortólogos Cas9 como complementos inversos.

Las **Figuras 8 A-J** muestran 18 estructuras de ARN quimérico que conservaron la secuencia y estructuras secundarias del dúplex de secuencias emparejamiento tracr:tracr, al tiempo que se acorta la región de apareamiento basal y la fusión de los dos elementos de ARN a través de un bucle artificial.

Las **Figuras 9 A-O** muestran una lista de secuencias Cas9 optimizadas en codones, humanas, a emparejarse con los ARNs guía quiméricos proporcionados en las Figuras 8 A-J.

Las **Figuras 10 A-M** muestran secuencias en las que se ubican los puntos de mutación dentro del gen SpCas9.

La **Figura 11** muestra loci de CRISPR de Tipo II en diferentes organismos.

La **Figura 12** muestra las secuencias guía ARN correspondientes a CRISPR loci en diferentes organismos.

La **Figura 13 A-II** muestra una tabla que lista los ortólogos Cas9 y sus secuencias PAM correspondientes.

La **Figura 14** muestra que el PAM para Staphylococcus aureus subespecie Aureus Cas9 es NNGRR.

Las **Figuras 15A-D** muestran diseños de vectores sencillos y múltiples para SaCas9.

La **Figura 16** muestra un diseño del vector y las imágenes en gel de ortólogos Cas9 y utilizándose respectivos sgRNAs para escindir dos dianas candidatas presentes en un banco basado en pUC19.

La **Figura 17** muestra la escisión in vitro mediante quimeras SpCas9, St3Cas9, Sp_St3 y quimer St3_Sp. Los PAM para Cas9 St3Cas9 y St3_Sp quiméricas son NGG.

La **Figura 18** muestra una imagen del servidor de web de CRISPRs.

Las **Figuras 19 A-L** muestran un alineamiento de secuencias múltiple para 12 ortólogos Cas9. Dos residuos catalíticos están resaltados. El primer residuo resaltado es el residuo Asp catalítico en el dominio RuvCI, y el segundo residuo resaltado es el residuo catalítico His en el dominio HNH. La mutación de uno o de los otros residuos en Ala puede convertir Cas9 en una nickase. La mutación de ambos residuos convierte Cas9 en un mutante catalíticamente nulo - útil para la unión a ADN genérico.

La **Figura 20** muestra una transferencia Western que muestra que ortólogos Cas9 se expresan en células HEK 293FT; plásmidos de ADN que codifican ortólogos Cas9 son transfectados en células HEK 293 FT y los lisados de células se recogen para la transferencia Western. Nada de ADN se transfecta para el ortólogo Cas9 nº 8.

La **Figura 21** muestra la escisión in vitro de dianas candidato en el plásmido pUC19 por parte de 10 ortólogos Cas9. Los PAMs consenso se predicen por logos de secuencia (Figura 7), con base en los cuales se eligen dianas candidato en pUC19. 7 de 9 nuevos ortólogos Cas9 ensayados han escindido con éxito al menos una diana pUC19. SpCas9 también puede escindir NGA in vitro.

La **Figura 22** muestra un Gel Surveyor Representativo que muestra la escisión genómica por SaCas9.

La **Figura 23** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Todas las dianas).

La **Figura 24** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Dianas escindidas).

La **Figura 25** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Todas las dianas, desechar las dianas de baja eficiencia y huérfanas).

La **Figura 26** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Dianas escindidas, desechar las dianas de baja eficiencia y huérfanas).

La **Figura 27** muestra una Secuencia Logo para Tratar Separadores Escindidos y PAMs (Nuevo test del genoma endógeno que muestra que no se requiere T).

Las figuras en esta memoria son sólo para fines ilustrativos y no están necesariamente dibujadas a escala.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

La invención se refiere a la modificación y optimización de sistemas, métodos y composiciones utilizados para el control de la expresión génica que implican la fijación como objetivo de la secuencia tal como la perturbación del genoma o la edición del genoma, que se refieren al sistema CRISPR y componentes del mismo. En realizaciones ventajosas, la enzima CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.La invención utiliza ácidos nucleicos para unir secuencias de ADN diana. Esto es ventajoso, ya que los ácidos nucleicos son mucho más fáciles y económicos de producir que, por ejemplo, péptidos, y la especificidad se puede variar de acuerdo con la longitud del tramo en el que se busca homología. Por ejemplo, no se requiere el posicionamiento en 3-D complejo de múltiples dedos.

Los términos y expresiones "polinucleótido", "nucleótido", "secuencia de nucleótidos", "ácido nucleico" y "oligonucleótido" se utilizan indistintamente. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitativos de polinucleótidos: regiones codificadoras o no codificadoras de un gen o fragmento de gen, loci (locus) definidos a partir de análisis de ligamiento, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosomal, ARN de interferencia corto (siRNA), ARN de horquilla corto (shRNA), micro-ARN

(miRNA), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. El término también abarca estructuras de tipo ácido nucleico con cadenas principales sintéticas, véase, p. ej., Eckstein, 1991: Baserga et al., 1992; Miiligan, 1993; documentos WO97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; y Samstag, 1996. Un polinucleótido puede comprender uno o más nucleótidos modificados tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, modificaciones en la estructura de nucleótidos se pueden impartir antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleótidos. Un polinucleótido puede modificarse adicionalmente después de la polimerización tal como por conjugación con un componente de marcaje.

En aspectos de la invención, las expresiones "ARN quimérico", "ARN guía quimérico", "ARN guía", "ARN guía sencillo" y "ARN guía sintético" se utilizan indistintamente y se refieren a la secuencia de polinucleótidos que comprende la secuencia guía, la secuencia tracr y la secuencia de emparejamiento tracr. La expresión "secuencia guía" se refiere a la secuencia de aproximadamente 20 pb dentro del ARN guía que especifica el sitio diana y puede utilizarse indistintamente con los términos "guía" o "espaciador". La expresión "secuencia de emparejamiento tracr" también se puede utilizar de forma indistinta con el término "repetición o repeticiones directas".

Tal como se utiliza en esta memoria, la expresión “tipo salvaje" es una expresión de la técnica entendida por las personas expertas y significa la forma típica de un organismo, cepa, gen o característica tal como se produce en la naturaleza que se distingue de las formas mutantes o variantes.

Tal como se utiliza en esta memoria, el término "variante" se deben tomar en el sentido de la exhibición de cualidades que tienen un patrón que se desvía de lo que ocurre en la naturaleza.

Las expresiones "que se produce de forma no natural" o "modificado por ingeniería genética" se utilizan indistintamente e indican la implicación de la mano del hombre. Las expresiones, cuando se refieren a moléculas de ácidos nucleicos o polipéptidos significan que la molécula de ácido nucleico o el polipéptido está al menos sustancialmente libre de al menos otro componente con el que están asociados de forma natural en la naturaleza y como se encuentra en la naturaleza.

"Complementariedad" se refiere a la capacidad de un ácido nucleico de formar enlace(s) hidrógeno con otra secuencia de ácido nucleico por apareamiento de bases de Watson-Crick tradicional u otros tipos no tradicionales. A porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que puede formar enlaces hidrógeno (p. ej., apareamiento de bases de Watson-Crick) con una segunda secuencia de ácido nucleico (p. ej., siendo 5, 6, 7, 8, 9, 10 de 10 50%, 60%, 70%, 80%, 90% y 100% complementarias). "Perfectamente complementarios" significa que todos los residuos contiguos de una secuencia de ácidos nucleicos se enlazarán por hidrógeno con el mismo número de residuos contiguos en una segunda secuencia de ácido nucleico. "Sustancialmente complementarios", tal como se utiliza en esta memoria, se refiere a un grado de complementariedad que es al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% o 100% sobre una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, o más nucleótidos, o se refiere a dos ácidos nucleicos que se hibridan en condiciones rigurosas.

Tal como se utiliza en esta memoria, "condiciones rigurosas" para la hibridación se refieren a condiciones en las que un ácido nucleico que tiene complementariedad con una secuencia diana se hibrida predominantemente con la secuencia diana, y sustancialmente no se hibrida con secuencias no diana. Las condiciones rigurosas son generalmente dependientes de la secuencia, y varían dependiendo de un cierto número de factores. En general, cuanto mayor sea la secuencia, más alta será la temperatura a la que la secuencia se hibrida específicamente con su secuencia diana. Ejemplos no limitantes de las condiciones restrictivas se describen en detalle en Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Parte I, Segundo Capítulo "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Cuando se haga referencia a una secuencia de polinucleótidos, entonces también se contemplan secuencias complementarias o parcialmente complementarias. Éstas son preferiblemente capaces de hibridarse a la secuencia de referencia en condiciones muy rigurosas. Generalmente, con el fin de maximizar la tasa de hibridación, se seleccionan condiciones de hibridación relativamente de baja rigurosidad: aproximadamente 20 a 25°C más bajas que el punto de fusión térmico (Tm). El Tm es la temperatura a la que el 50% de la secuencia diana específica se hibrida con una sonda perfectamente complementaria en solución a una fuerza iónica y pH definidos. Generalmente, con el fin de requerir al menos aproximadamente 85% de complementariedad de nucleótidos de las secuencias hibridadas, las condiciones de lavado altamente rigurosas se seleccionan para que sean aproximadamente 5 a 15°C inferiores al Tm. Con el fin de requerir al menos aproximadamente 70% de complementariedad de nucleótidos de las secuencias hibridadas, las condiciones de lavado moderadamente rigurosas se seleccionan para que sean aproximadamente 15 a 30°C inferiores al Tm. Condiciones de lavado altamente permisivas (rigurosidad muy baja) pueden ser tan bajas como 50°C inferiores al Tm, lo que permite un alto nivel de emparejamientos erróneos entre las secuencias hibridadas. Los expertos en la técnica reconocerán que otros parámetros físicos y químicos en las fases de hibridación y lavado también pueden alterarse para afectar el resultado de una señal de hibridación detectable de

un nivel específico de homología entre las secuencias diana y sonda. Condiciones altamente rigurosas preferidas comprenden la incubación en formamida al 50%, 5xSSC y SDS al 1% a 42°C, o la incubación en 5xSSC y SDS al 1% a 65°C, con lavado en 0,2xSSC y SDS al 0,1% SDS a 65°C.

"Hibridación" se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza a través de enlaces hidrógeno entre las bases de los residuos de nucleótidos. El enlace hidrógeno se puede producir por apareamiento de bases de Watson Crick, la unión Hoogstein, o en cualquier otra forma específica para la secuencia. El complejo puede comprender dos cadenas que forman una estructura dúplex, tres o más cadenas que forman un complejo multi-cadena, una sola cadena auto-hibridante, o cualquier combinación de éstas. Una reacción de hibridación puede constituir una etapa de un proceso más extenso tal como el inicio de la PCR, o la escisión de un polinucleótido por una enzima. A una secuencia capaz de hibridarse con una secuencia dada se la alude como el "complemento" de la secuencia dada.

Tal como se utiliza en esta memoria, la expresión "locus genómico" o "locus" (plural loci) es la ubicación específica de un gen o secuencia de ADN en un cromosoma. Un "gen" se refiere a tramos de ADN o ARN que codifican un polipéptido o una cadena de ARN que tiene que jugar un papel funcional en un organismo y, por lo tanto, es la unidad molecular hereditaria en los organismos vivos. Para los fines de esta invención, se puede considerar que los genes incluyen regiones que regulan la producción del producto génico, estén o no estas secuencias reguladoras adyacentes a las secuencias codificadoras y/o transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a secuencias de promotor, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada al ribosoma interno, potenciadores, silenciadores, aisladores, elementos de borde, orígenes de replicación, sitios de unión a la matriz y regiones de control del locus.

Tal como se utiliza en esta memoria, "expresión de un locus genómico" o "expresión génica" es el proceso por el cual la información de un gen se utiliza en la síntesis de un producto génico funcional. Los productos de la expresión génica son a menudo proteínas, pero en genes no codificadores de proteínas tales como genes de ARNr o genes de ARNt, el producto es ARN funcional. El proceso de la expresión génica se utiliza por todos los eucariotas vivos conocidos (incluyendo organismos multicelulares), procariotas (bacterias y arqueas) y virus para generar productos funcionales para sobrevivir. Tal como se utiliza en esta memoria, "expresión" de un gen o ácido nucleico abarca no sólo la expresión del gen celular, sino también la transcripción y la traducción de ácido o ácidos nucleicos en los sistemas de clonación y en cualquier otro contexto. Tal como se utiliza en esta memoria, "expresión" también se refiere al proceso mediante el cual un polinucleótido se transcribe a partir de un molde de ADN (como en y ARNm u otro transcrito de ARN) y/o el proceso mediante el cual un ARNm transcrito se traduce posteriormente en péptidos, polipéptidos o proteínas. A transcritos y polipéptidos codificados se les puede aludir selectivamente como "producto génico". Si el polinucleótido se deriva de ADN genómico, la expresión puede incluir el corte y empalme de ARNm en una célula eucariota.

Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente en esta memoria para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y puede estar interrumpido por no aminoácidos. Los términos también abarcan un polímero de aminoácidos que ha sido modificado; por ejemplo, formación de enlaces disulfuro, glicosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación tal como conjugación con un componente de marcaje. Tal como se utiliza en esta memoria, el término "aminoácido" incluye aminoácidos naturales y/o no naturales o sintéticos, incluyendo glicina y tanto los isómeros ópticos D o L y análogos de aminoácidos y peptidomirnéticos.

Tal como se utiliza en esta memoria, el término "dominio" o la expresión "dominio de proteína" se refiere a una parte de una secuencia de proteína que pueden existir y funcionar independientemente del resto de la cadena de proteína.

Tal como se describe en aspectos de la invención, la identidad de secuencia está relacionada con la homología de secuencia. Las comparaciones de homología pueden realizarse a simple vista, o más habitualmente, con la ayuda de programas de comparación de secuencias fácilmente disponibles. Estos programas informáticos disponibles comercialmente pueden calcular el porcentaje (%) de homología entre dos o más secuencias y también pueden calcular la identidad de secuencia compartida por dos o más secuencias de aminoácidos o de ácido nucleico. Las homologías de secuencia pueden generarse mediante cualquiera de un cierto número de programas de ordenador conocidos en la técnica, por ejemplo BLAST o FASTA, etc. Un programa informático adecuado para llevar a cabo dicha alineación es el paquete GCG Wisconsin Bestfit (Universidad de Wisconsin, EE.UU.; Devereux et al., 1984, Nucleic Acids Research 12:387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 ibid. - Capítulo 18), FASTA (Atschul et al., 1990, J. Mol. Biol, 403-410) y el conjunto GENEWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda fuera de línea y en línea (véase Ausubel et al., 1999 ibid, páginas 7-58 a 7-60). Sin embargo, se prefiere utilizar el programa GCG Bestfit.

El % de homología puede calcularse sobre secuencias contiguas, es decir, una secuencia se alinea con la otra secuencia y cada uno de los aminoácidos o nucleótidos en una secuencia se compara directamente con el aminoácido o nucleótido correspondiente en la otra secuencia, un residuo cada vez. Esto se denomina alineamiento "sin huecos". Típicamente, estos alineamientos sin huecos se realizan sólo sobre un número relativamente corto de residuos.

Aunque este es un método muy simple y consistente, no tiene en consideración que, por ejemplo, en un par de secuencias de otro modo idénticas, una inserción o deleción puede provocar que los siguientes restos de aminoácidos se coloquen fuera de alineamiento, por lo tanto, resultando potencialmente en una gran reducción en el % de homología cuando se realiza un alineamiento global. En consecuencia, la mayoría de los métodos de comparación de secuencias se diseñan para producir alineamientos óptimos que tienen en cuenta posibles inserciones y deleciones sin penalizar indebidamente la puntuación de homología o identidad global. Esto se consigue insertando "huecos" en el alineamiento de secuencias para intentar maximizar la homología o identidad local.

Sin embargo, estos métodos más complejos asignan "penalizaciones por hueco" a cada uno de los huecos que se produce en el alineamiento, de modo que, para el mismo número de aminoácidos idénticos, un alineamiento de la secuencia con tan pocos huecos como sea posible - lo que refleja una mayor relatividad entre las dos secuencias comparadas - puede alcanzar una puntuación más alta que con muchos huecos. Se utilizan típicamente “costos de huecos de afinidad" que cargan un coste relativamente alto para la existencia de un hueco y una penalización más pequeña para cada uno de los residuos posteriores en el hueco. Este es el sistema de puntuación de huecos más comúnmente utilizado. Penalizaciones de huecos elevadas pueden producir, por supuesto, alineamientos optimizados con menos huecos. La mayoría de los programas de alineamiento permiten modificar las penalizaciones por hueco. Sin embargo, se prefiere utilizar los valores por defecto cuando se utiliza un software de este tipo para las comparaciones de secuencias. Por ejemplo, al utilizar el paquete GCG Wisconsin Bestfit la penalización por hueco por defecto para secuencias de aminoácidos es -12 para un hueco y -4 para cada extensión,

El cálculo del % máximo de homología, por tanto, requiere primero la producción de una alineación óptima, teniendo en consideración las penalizaciones por hueco. Un programa informático adecuado para llevar a cabo dicho alineamiento es el paquete GCG Wisconsin Bestfit (Devereux et al., 1984 Nuc. Acids Research 12 p 387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 Short Protocols in Molecular Biology, 4ª Ed. - Capítulo 18), FASTA (Altschul et al., 1990, J. Mol. Biol. 403-410) y el conjunto GENEWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda en línea y fuera de línea (véase Ausubel et al., 1999, Short Protocols in Molecular Biology, páginas 7-58 a 7-60). Sin embargo, para algunas aplicaciones, se prefiere utilizar el programa GCG Bestfit. Una nueva herramienta, denominada BLAST 2 Sequences, está también disponible para la comparación de secuencias de proteínas y nucleótidos (véase FEMS Microbiol Lett. 1999 174(2):247-50; FEMS Microbiol Lett. 1999 177(1):187-8 y el sitio web del Centro Nacional de información sobre Biotecnología en la página web de National Institutes for Health).

Aunque el % de homología final puede medirse en términos de identidad, el proceso de alineamiento en sí no se basa típicamente en una comparación de pares de todo o nada. En su lugar, se utiliza generalmente una matriz de puntuación de similitud a escala que asigna puntuaciones a cada comparación por pares basada en la similitud química o la distancia evolutiva. Un ejemplo de una matriz de este tipo comúnmente utilizada es la matriz BLOSUM62 - la matriz por defecto para el conjunto de programas BLAST. Los programas GCG Wisconsin generalmente utilizan cualquiera de los valores por defecto públicos o una tabla de comparación de símbolos personalizada, si se suministra (véase el manual del usuario para más detalles). Para algunas aplicaciones, se prefiere utilizar los valores por defecto públicos para el paquete GCG, o en el caso de otro software, la matriz por defecto, tal como BLOSUM62.

Alternativamente, las homologías porcentuales pueden calcularse utilizando la característica de alineamiento múltiple en DNASIS™ (Hitachi Software), basado en un algoritmo, análogo a CLUSTAL (Higgins DG y Sharp PM (1988), Gene 73 (1), 237-244). Una vez que el software ha producido un alineamiento óptimo, es posible calcular el % de homología, preferentemente el % de identidad de secuencia. El software hace normalmente esto como parte de la comparación de secuencias y genera un resultado numérico.

Las secuencias también pueden tener deleciones, inserciones o sustituciones de residuos de aminoácidos que producen un cambio silencioso y resultan en una sustancia funcionalmente equivalente. Sustituciones de aminoácidos deliberadas pueden realizarse sobre la base de la similitud en las propiedades de aminoácidos (tales como polaridad, carga, solubilidad, hidrofobicidad, hidrofilicidad y/o la naturaleza anfipática de los residuos) y, por lo tanto, es útil para agrupar juntos aminoácidos en grupos funcionales. Los aminoácidos pueden agruparse basándose en las propiedades de sus cadenas laterales solas. Sin embargo, es más útil incluir los datos de mutación también. Los conjuntos de aminoácidos así obtenidos es probable que se conserven por razones estructurales. Estos conjuntos pueden describirse en forma de un diagrama de Venn (Livingstone CD y Barton GJ (1993) "Protein

sequence alignments: a strategy for the hierarchical analysis of residue conservation" Comput. Appl. Biosci. 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" J. Theor. Biol. 119; 205-218). Sustituciones conservadoras se pueden hacer, por ejemplo, de acuerdo con la tabla que figura a continuación que describe un diagrama de Venn generalmente aceptado de agrupación de aminoácidos.

Conjunto		Sub-conjunto	
Hidrofóbico	F W Y H K M I L V A G C	Aromático	F W Y H
		Alifático	I L V
Polar	W Y H K R E D C S T N Q	Cargado	H K R E D
		Positivamente cargado	H K R
		Negativamente cargado	E D
Pequeño	V C A G S P T N D	Minúsculo	A G S

Realizaciones de la invención incluyen secuencias (tanto de polinucleótidos o polipéptidos) que puede comprender la sustitución homóloga (sustitución y reemplazo se utilizan ambos en esta memoria para dar a entender el intercambio de un residuo de aminoácido o nucleótido existente, con un residuo o nucleótido alternativo) que puede producirse, es decir, una sustitución comparable en el caso de aminoácidos tal como básico por básico, ácido por ácido, polar por polar, etc. También puede producirse una sustitución no homóloga, es decir, de una clase de residuo a otro o, alternativamente, que implica la inclusión de aminoácidos no naturales tales como ornitina (a la que se alude en lo sucesivo como Z), ornitina del ácido diaminobutírico (a la que se alude en lo sucesivo como B), ornitina norleucina (a la que se alude en lo sucesivo como O), pirilalanina, tienilalanina, naftilalanina y fenilglicina.

Secuencias de aminoácidos variantes pueden incluir grupos espaciadores adecuados que pueden insertarse entre cualesquiera dos residuos de aminoácidos de la secuencia, incluyendo grupos alquilo tales como grupos metilo, etilo o propilo, además de espaciadores de aminoácidos tales como residuos glicina o β-alanina. Una forma adicional de variación, que implica la presencia de uno o más residuos aminoácidos en forma peptoide, pueda ser bien comprendida por los expertos en la técnica. Para evitar dudas, "la forma peptoide" se utiliza para referirse a residuos aminoácidos variantes, en donde el grupo de sustituyentes carbono α está en átomo de nitrógeno del residuo en lugar del carbono α. Procedimientos para preparar péptidos en forma peptoide son conocidos en la técnica, por ejemplo Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 y Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134.

La práctica de la presente invención emplea, a menos que se indique lo contrario, técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que están dentro de la experiencia de la técnica. Véase Sambrook, Fritsch y Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2ª edición (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. comps., (1987)); la serie METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames y G.R. Taylor comps. (1995)), Harlow y Lane, comps. (1988) ANTIBODIES, A LABORATORY MANUAL, y ANIMAL CELL CULTURE (R.I. Freshney, comp. (1987)).

En un aspecto, la invención proporciona vectores que se utilizan en la ingeniería y optimización de los sistemas CRISPR/Cas.

Tal como se utiliza en esta memoria, un "vector" es una herramienta que permite o facilita la transferencia de una entidad de un entorno a otro. Es un replicón, tal como un plásmido, fago, o cósmido, en el que se puede insertar otro segmento de ADN con el fin de llevar a cabo la replicación del segmento insertado. Generalmente, un vector es capaz de replicarse cuando se asocia con los elementos de control apropiados. En general, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Vectores incluyen, pero no se limitan a, moléculas de ácido nucleico que son de cadena sencilla, de cadena doble, o parcialmente de doble cadena; moléculas de ácido nucleico que comprenden uno o más extremos libres, sin extremos libres (p. ej., circulares); moléculas de ácido nucleico que comprenden ADN, ARN, o ambos; y otras variedades de polinucleótidos conocidos en la técnica. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de doble cadena circular en el que se pueden insertar segmentos de ADN adicionales tal como por técnicas de clonación molecular estándares. Otro tipo de vector es un vector viral, en el que las secuencias de ADN o ARN derivadas de virus están presentes en el vector para el empaquetamiento en un virus (p. ej., retrovirus, retrovirus defectuosos en la replicación, adenovirus, adenovirus defectuosos en la replicación, y virus adeno-asociados). Los vectores virales también incluyen polinucleótidos portados por un virus para la transfección en una célula huésped. Determinados vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (p. ej., vectores bacterianos que tienen un origen bacteriano de replicación y vectores episomales de mamífero). Otros vectores (p. ej., vectores no episomales de mamífero) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y, con ello, se replican junto con el genoma del huésped. Además, determinados vectores son capaces de dirigir la expresión de genes a los que están enlazados operativamente. Tales vectores se

denominan en esta memoria "vectores de expresión". Vectores de expresión comunes de utilidad en técnicas de ADN recombinante están a menudo en forma de plásmidos. En esta memoria se proporciona una discusión adicional de vectores.

Los vectores de expresión recombinantes pueden comprender un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladores, que pueden seleccionarse sobre la base de las células huésped a utilizar para la expresión, que están operativamente enlazadas a la secuencia de ácido nucleico a expresar. Dentro de un vector de expresión recombinante, "enlazada operativamente" pretende significar que la secuencia de nucleótidos de interés está enlazada al o a los elementos reguladores de una manera que permite la expresión de la secuencia de nucleótidos (p. ej., en un sistema de transcripción/traducción in vitro o en una célula huésped cuando el vector se introduce en la célula huésped). Con respecto a métodos de recombinación y de clonación, se hace mención a la solicitud de patente de EE.UU. US 2004-0171156 A1.

Aspectos y realizaciones de la invención se refieren a vectores bicistrónicos para ARN quimérico y Cas9 o un ortólogo Cas9. En algunas realizaciones, la Cas9 es impulsada por el promotor promotor CBh. En algunas realizaciones, el ARN quimérico es impulsado por un promotor U6. Preferiblemente, el CBh y U6 se utilizan juntos, en el sentido de que Cas9 es impulsada por el promotor CBh y el ARN quimérico es impulsado por un promotor U6. En algunas realizaciones, el ARN guía quimérico consiste en una secuencia guía de 20 pb (Ns) unidos a la secuencia tracr (que va desde la primera "U" de la cadena inferior al extremo del transcrito), que está truncada en diversas posiciones tal como se indica. Las secuencias guía y tracr están separadas preferiblemente por la secuencia de emparejamiento tracr. Un ejemplo preferido de una secuencia de emparejamiento tracr es GUUUUAGAGCUA. Esto es seguido preferiblemente por una secuencia de bucle. El bucle es preferiblemente GAAA, pero no se limita a esta secuencia o, de hecho, sólo es de 4 pb de longitud. De hecho, secuencias formadoras de bucle preferidas para su uso en estructuras de horquilla son de cuatro nucleótidos de longitud, y lo más preferiblemente tienen la secuencia GAAA. Sin embargo, se pueden utilizar secuencias de bucle más largas o más cortas, al igual que secuencias alternativas. Las secuencias incluyen preferiblemente un triplete de nucleótidos (por ejemplo, AAA), y un nucleótido adicional (por ejemplo C o G). Ejemplos de secuencias formadoras de bucle incluyen CAAA y AAAG. A lo largo de esta solicitud, ARN quimérico también se puede denominar ARN guía sencillo o guía sintético (sgRNA).

Por ejemplo, tal como se describe en detalle específico en el Ejemplo 2, ARNs guía quiméricos pueden ser diseñados tal como se muestra en la Figura 8. Los loci de CRISPR en algunas de estas familias se representan en la Figura 11. Las correspondientes secuencias de ARN guía se muestran en la Figura 12. Los autores de la invención analizaron la secuencia de ADN genómico dentro ~ 2kb de las proteínas Cas9 e identificaron repeticiones directas que van desde 35 pb a 50 pb, con separadores intermedios que van desde 29 pb a 35 pb. Sobre la base de la secuencia de repetición directa, realizaron búsquedas de secuencias candidatas de tracrRNA con los siguientes criterios: fuera de la matriz de crRNA, pero que contienen un alto grado de homología con repeticiones directas (tal como se requiere para el apareamiento de bases repetición directa:tracRNA; el análisis personalizado de cálculo), fuera de las regiones codificantes de los componentes de la proteína, que contienen señales de terminación de la transcripción de ~ 60 pb-120 pb aguas abajo de la región de homología con repeticiones directas y el co-plegamiento con repetición directa para formar un dúplex, seguido de dos o más estructuras de horquilla en el extremo distal de la secuencia de tracrRNA. Sobre la base de estos criterios de predicción, los autores de la invención seleccionaron un conjunto inicial de 18 proteínas Cas9 y sus repeticiones y tracrRNAs directos asociados de forma única distribuidos a través de las cinco familias Cas9. La solicitante generó, además, un conjunto de 18 estructuras de ARN quiméricas que conservaron la secuencia y las estructuras secundarias del dúplex repetición directa:tracRNA nativo, al tiempo que acortan la región de apareamiento de bases y fusionan los dos elementos de ARN a través de un bucle artificial (Figuras 6 A-J).

La expresión "elemento regulador" pretende incluir promotores, potenciadores, sitios de entrada al ribosoma interno (IRES), y otros elementos de control de la expresión (p. ej., señales de terminación de la transcripción tales como señales de poliadenilación y secuencias poli-U). Se describen estos elementos reguladores, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Elementos reguladores incluyen los que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huéspedes y los que dirigen la expresión directa de la secuencia de nucleótidos sólo en determinadas células huéspedes (p. ej., secuencias reguladoras específicas para el tejido). Un promotor específico para el tejido puede dirigir la expresión principalmente en un tejido deseado de interés tal como el músculo, neurona, hueso, piel, sangre, órganos específicos (p. ej., hígado, páncreas), o tipos de células particulares (p. ej., linfocitos). Los elementos reguladores también pueden dirigir la expresión de una manera dependiente del tiempo tal como de una manera dependiente del ciclo celular o dependiente de la fase de desarrollo, que puede o puede no ser también específico para el tejido o tipo de célula. En algunas realizaciones, un vector puede comprender uno o más promotores pol III (p. ej., 1, 2, 3, 4, 5 o más promotores pol III), uno o más promotores pol II (p. ej., 1, 2, 3, 4, 5 o más promotores pol II), uno o más promotores pol I (p. ej., 1, 2, 3, 4, 5 o más promotores pol I), o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, pero no se limitan a promotores U6

y H1. Ejemplos de promotores pol II incluyen, pero no se limitan al promotor del virus del sarcoma de Rous retroviral (RSV) LTR (opcionalmente con el potenciador RSV), el promotor de citomegalovirus (CMV) (opcionalmente con el potenciador CMV) [véase, p. ej., Boshart et al., Cell, 41:521-530 (1985)], el promotor SV40, el promotor de dihidrofolato reductasa, el promotor β-actina, el promotor fosfoglicerol quinasa (PGK) y el promotor EF1α. También quedan abarcados por la expresión "elemento regulador" elementos potenciadores tales como WPRE; potenciadores de CMV; el segmento R-U5' en LTR de HTLV-I (Mol. Cell. Biol., Vol 8(1), págs. 466-472, 1988); el potenciador SV40; y la secuencia de intrón entre los exones 2 y 3 de β-globina de conejo (Proc. Natl, Acad. Sci. USA, Vol, 78(3), págs. 1527-31, 1981). Se apreciará por los expertos en la técnica que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped a transformar, el nivel de expresión deseado, etc. Un vector puede ser introducido en células huéspedes para con ello producir transcritos, proteínas o péptidos, incluyendo proteínas o péptidos de fusión, codificados por ácidos nucleicos tal como se describe en esta memoria (p. ej., repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR), transcritos, proteínas, enzimas, formas mutantes de los mismos, proteínas de fusión de los mismos, etc.). Con respecto a secuencias reguladoras, se hace mención de la publicación de la solicitud de patente de EE.UU. 10/491.026. Con respecto a los promotores, se hace mención a la publicación PCT WO 2011/028929 y la solicitud de patente de EE.UU. 12/511.940.

Los vectores pueden ser diseñados para la expresión de transcritos de CRISPR (p. ej., transcritos de ácidos nucleicos, proteínas o enzimas) en células procariotas o eucariotas. Por ejemplo, los transcritos de CRISPR pueden expresarse en células bacterianas tales como Escherichia coli, células de insectos (utilizando vectores de expresión de baculovirus), células de levaduras o células de mamífero. Células huéspedes adecuadas se discuten en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, California (1990). Alternativamente, el vector de expresión recombinante puede transcribirse y traducirse in vitro, por ejemplo utilizando secuencias reguladoras del promotor T7 y T7 polimerasa.

Los vectores se pueden introducir y propagar en una célula procariota o procariótica. En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector que se ha de introducir en una célula eucariota o como un vector intermedio en la producción de un vector a ser introducido en una célula eucariota (por ejemplo, una ampliación de un plásmido como parte de un sistema de empaquetado de un vector viral). En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector y expresar uno o más ácidos nucleicos tal como para proporcionar una fuente de una o más proteínas para el suministro a una célula huésped u organismo huésped. La expresión de proteínas en procariotas se lleva a cabo más a menudo en Escherichia coli con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o de no fusión. Los vectores de fusión añaden un cierto número de aminoácidos a una proteína codificada en el mismo tal como al extremo amino de la proteína recombinante. Tales vectores de fusión pueden cumplir uno o más fines tales como: (i) aumentar la expresión de la proteína recombinante; (ii) aumentar la solubilidad de la proteína recombinante; y (iii) ayudar en la purificación de la proteína recombinante, actuando como un ligando en purificación por afinidad. A menudo, en vectores de expresión de fusión, un sitio de escisión proteolítica se introduce en la unión del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante del resto de fusión después de la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen Factor Xa, trombina y enteroquinasa. Ejemplos de vectores de expresión de fusión incluyen pGEX (Pharmacia Biotech Inc; Smith y Johnson, 1988. Gene 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRIT 5 (Pharmacia, Piscataway, N.J.) que fusionan la glutatión S transferasa (GST), proteína de unión a maltosa, o proteína A, respectivamente, a la proteína recombinante diana.

Ejemplos de vectores de expresión de E. coli sin fusión inducibles adecuados incluyen pTrc (Amrann et al., (1988) Gene 69: 301-315) y pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZIMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

En algunas realizaciones, un vector es un vector de expresión de levadura. Ejemplos de vectores para la expresión en la levadura Saccharomyces cerevisiae incluyen pYepSecl (Baldari, et al., 1987. EMBO J. 6:229-234), pMFa (Kuijan y Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (Invitrogen Corp, San Diego, Calif).

En algunas realizaciones, un vector impulsa la expresión de proteínas en células de insecto utilizando vectores de expresión de baculovirus. Vectores de baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (p. ej., células SF9) incluyen la serie pAc (Smith, et al., 1983. Mol. Cell. Biol. 3:2156-2165) y la serie pVL (Lucklow y Summers, 1989. Virology 170:31-39).

En algunas realizaciones, un vector es capaz de impulsar la expresión de una o más secuencias en células de mamífero utilizando un vector de expresión de mamífero. Ejemplos de vectores de expresión de mamífero incluyen pCDM8 (Seed, 1987. Nature 329:840) y pMT2PC (Kaufman, et al., 1987. EMBO J. 6:187-195). Cuando se utiliza en células de mamífero, las funciones de control del vector de expresión son proporcionadas típicamente por uno o más elementos reguladores. Por ejemplo, promotores comúnmente utilizados se derivan de polioma, adenovirus 2,

citomegalovirus, virus de simio 40, y otros descritos en esta memoria y conocidos en la técnica. Para otros sistemas de expresión adecuados tanto para células procariotas como eucariotas véanse, p ej., los Capítulos 16 y 17 de Sambrook, et al., MOLECULAR CLONING:. A LABORATORY MANUAL. 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989.

En algunas realizaciones, el vector de expresión de mamífero recombinante es capaz de dirigir la expresión del ácido nucleico preferentemente en un tipo de célula particular (p. ej., elementos reguladores específicos para tejidos se utilizan para expresar el ácido nucleico). Elementos reguladores específicos para tejidos son conocidos en la técnica. Ejemplos no limitantes de promotores específicos para tejidos, adecuados, incluyen el promotor de albúmina (específico para el hígado; Pinkert, et al., 1987, Genes Dev. 1: 268-277), promotores linfoide-específicos (Calame y Eaton, 1988. Adv. Immunol. 43:235-275), en particular promotores de receptores de células T (Winoto y Baltimore, 1989. EMBO J. 8:729-733) e inmunoglobulinas (Baneiji, et al., 1983. Cell 33: 729-740; Queen y Baltimore, 1983 Cell 33: 741-748), promotores específicos para neuronas (p. ej., el promotor para neurofilamentos; Byrne y Ruddle, 1989. Proc. Natl. Acad. Sci. USA 86: 5473-5477), promotores específicos para páncreas (Edlund, et al., 1985. Science 230:912-916) y promotores específicos para glándulas mamarias (p. ej., promotor de suero lácteo; patente de EE.UU. Nº. 4.873.316 y la Publicación de Solicitud Europea Nº 264.166.). Promotores regulados en el desarrollo también están comprendidos, p. ej., los promotores hox murinos (Kessel y Gruss, 1990. Science 249:374-379) y el promotor de α-fetoproteína (Campes y Tilghman, 1989. Genes Dev. 3: 537-546). Con respecto a estos vectores procarióticos y eucarióticos, se hace mención a la patente de EE.UU. 6.750.059. Otras realizaciones de la invención pueden estar relacionadas con el uso de vectores virales, en lo que respecta a los que se hace mención en la solicitud de patente de EE.UU. 13/092.085. Elementos reguladores específicos para tejidos son conocidos en la técnica y, en este sentido, se hace mención a la patente de EE.UU. 7.776.321.

En algunas realizaciones, un elemento regulador está enlazado operativamente a uno o más elementos de un sistema de CRISPR con el fin de impulsar la expresión de los uno o más elementos del sistema de CRISPR. En general, las CISPRs (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), también conocidas como SPIDRs (Repeticiones Directas Intercaladas Separadas), constituyen una familia de loci de ADN que habitualmente son específicos para una especie bacteriana particular. El locus de CRISPR puede comprender una clase distinta de Repeticiones de Secuencia Corta Intercaladas (SSRs) que fueron reconocidas en E. coli (Ishino et al, J. Bacteriol., 169:5429-5433 [1987]; y Nakata et al, J. Bacteriol., 171: 3553-3556 [1989]), y genes asociados. SSRs intercaladas similares han sido identificadas en Haloferax mediterranei, Streptococcus pyogenes, Anabaena y Mycobacterium tuberculosis (Véase, Groenen et al, Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; y Mojica et al., Mol. Microbiol., 17:85-93 [1995]). Los loci CRISPR difieren típicamente de otros SSRs por la estructura de las repeticiones, que se han denominado repeticiones cortas regularmente espaciadas (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; y Mojica et. al., Mol Microbiol, 36:244-246 [2000]). En general, las repeticiones son elementos cortos que se producen en grupos que están espaciados regularmente por secuencias intermedias únicas con una longitud sustancialmente constante (Mojica et al., [2000], supra). Aunque las secuencias de repetición están altamente conservadas entre las cepas, el número de repeticiones intercaladas y las secuencias de las regiones del espaciador difieren típicamente de una cepa a otra (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). Loci CRISPR han sido identificados en más de 40 procariotas (Véase, p. ej., Jansen et al, Mol. Microbiol., 43:1565-1575 [2002]; y Mojica et al., [2005]), incluyendo, pero no limitado a Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacterium, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomyces, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema y Thermotoga.

En general, "sistema de CRISPR" se refiere colectivamente a transcritos y otros elementos implicados en la expresión de o dirigir la actividad de genes CRISPR-asociados ("Cas"), incluyendo secuencias que codifican un gen Cas, una secuencia tracr (CRISPR trans-activante) (p. ej., tracrRNA o un tracrRNA parcial activo), una secuencia de emparejamiento tracr (que abarca una "repetición directa" y una repetición directa parcial procesada por tracrRNA en el contexto de un sistema de CRISPR endógeno), una secuencia guía (a la que también se alude como un "espaciador" en el contexto de un sistema de CRISPR endógeno), u otras secuencias y transcritos de un locus CRISPR. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un sistema de CRISPR de tipo I, tipo II o de tipo III. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un organismo particular que puede comprender un sistema de CRISPR endógeno tal como Streptococcus pyogenes. En general, un sistema de CRISPR se caracteriza por elementos que fomentan la formación de un complejo de CRISPR en el sitio de una secuencia diana (al que también se alude como un protoespaciador en el contexto de un sistema de CRISPR endógeno). En el contexto de la formación de un complejo de CRISPR, "secuencia diana" se refiere a una secuencia a la que una secuencia guía está diseñada para que tenga complementariedad, en donde la hibridación entre una secuencia diana y una secuencia guía fomenta la formación de un complejo de CRISPR. Una secuencia diana puede comprender cualquier polinucleótido tal como

polinucleótidos de ADN o ARN. En algunas realizaciones, una secuencia diana se encuentra en el núcleo o citoplasma de una célula.

En realizaciones preferidas de la invención, el sistema de CRISPR es un sistema de CRISPR de tipo II y la enzima Cas es Cas9, que cataliza la escisión de ADN. La acción enzimática por Cas9 derivada de Streptococcus pyogenes o cualquier Cas9 estrechamente relacionada genera roturas de doble cadena en las secuencias del sitio diana que se hibridan a 20 nucleótidos de la secuencia guía y que tienen una secuencia del motivo adyacente al protoespaciador (PAM) NGG/NRG (por ejemplo, tal como se discute en otra parte, un PAM adecuado es 5'-NRG o 5'-NNGRR para SpCas9 o enzimas (o enzimas derivadas) SaCas9, respectivamente) después de los 20 nucleótidos de la secuencia diana. La actividad de CRISPR a través de Cas9 para el reconocimiento de ADN específico para el sitio y la escisión se definen por la secuencia guía, la secuencia tracr que se hibrida en parte a la secuencia guía y la secuencia de PAM. Se describen más aspectos del sistema de CRISPR en Karginov y Hannon. The CRISPR system: small RNA-guided defense in bacteria and archaea, Mole Cell 2010, 15 de enero; 37(1): 7.

El locus de CRISPR de tipo II de Streptococcus pyogenes SF370 contiene un racimo de cuatro genes Cas9, Cas1, Cas2 y Csn1, así como dos elementos de ARN no codificantes, tracrRNA y una matriz característica de secuencias repetitivas (repeticiones directas) interespaciadas por tramos cortos de secuencias no repetitivas (espaciadores, aproximadamente de 30 pb cada uno). En este sistema, la rotura de doble cadena de ADN (DSB) fijada como objetivo se genera en cuatro etapas secuenciales (Figura 2A). En primer lugar, dos ARN no codificantes, la matriz pre-crRNA y tracrRNA se transcriben a partir del locus de CRISPR. En segundo lugar, tracrRNA se hibrida a las repeticiones directas de pre-crRNA, que se procesa luego en crRNA maduros que contienen secuencias de espaciador individuales. En tercer lugar, el complejo crRNA:tracrRNA maduro dirige Cas9 a la diana de ADN que consiste en el protoespaciador y el PAM correspondiente a través de la formación de heteroduplex entre la región del espaciador del crRNA y el ADN del protoespaciador. Por último, Cas9 media en la escisión del ADN diana aguas arriba del PAM para crear una DSB dentro del protoespaciador (Figura 2A). La Figura 2B muestra la ubicación nuclear de Cas9 optimizado en codones. Para fomentar la iniciación de la transcripción precisa, se seleccionó el promotor U6 basado en ARN polimerasa III para impulsar la expresión de tracrRNA (Figura 2C). De manera similar, se desarrolló una construcción basada en el promotor U6 para expresar una matriz de pre-crRNA que consiste en un único espaciador flanqueado por dos repeticiones directas (DRs, también abarcados por la expresión "secuencias de emparejamiento de tracr"; Figura 2C). El espaciador inicial se diseñó para fijar como objetivo un sitio diana de 33 pares de bases (pb) (protoespaciador de 30 pb más una secuencia del motivo de CRISPR (PAM) de 3 pb que satisface el motivo de reconocimiento de NGG de Cas9) en el locus EMX1 humano (Figura 2C), un gen clave en el desarrollo de la corteza cerebral.

La Figura 16 muestra ortólogos Cas9 y respectivos sgRNAs se utilizan para escindir dos dianas candidatas presentes en un banco basado en pUC19. La diana 1 es seguida por un PAM aleatorio que contiene 7 bases degeneradas (5'-NNNNNNN -3'), y la diana 1', que contiene la misma secuencia diana que la diana 1, es seguida por un PAM fijo (5'-TGGAGAAT-3). El sgRNA de cada uno de los ortólogos Cas9 contiene la secuencia guía contra la diana 1 o diana 1'. Imágenes en gel muestran la escisión con éxito por 20 ortólogos Cas9, lo que indica que estos diseños de sgRNA son funcionales con sus respectivas enzimas Cas9.

En algunas realizaciones, repeticiones directas o secuencias de emparejamiento tracr se descargan de la base de datos CRISPRs o se identifican in silico mediante el rastreo de motivos repetitivos que 1. se encuentran en una ventana de 2 kb de secuencia genómica flanqueante del locus CRISPR de tipo II, 2. abarcan de 20 a 50 pb, y 3. están separadas por 20 a 50 pb. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo, 1 y 2, 2 y 3 ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios.

En algunas realizaciones tracrRNA candidato se predice posteriormente por 1. homología de la secuencia con repeticiones directas (búsqueda con motivos en Geneious con apareamientos erróneos de hasta 18 pb), 2. presencia de un terminador de la transcripción independiente de Rho predicho en la dirección de la transcripción, y 3. estructura secundaria de horquilla estable entre tracrRNA y la repetición directa. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones, diseños de ARNs guía quiméricos sintéticos (sgRNAs) incorporan al menos 8 pb de la estructura dúplex entre la repetición directa y tracrRNA.

Varios aspectos del sistema de CRISPR pueden mejorarse aún más para aumentar la eficiencia y versatilidad de la fijación como objetivo de CRISPR. Una actividad de Cas9 óptima puede depender de la disponibilidad de Mg2+ libre a niveles más altos que los presentes en el núcleo de mamíferos (véase, p. ej., Jinek et al., 2012, Science, 337:816), y la preferencia por un motivo NGG/NRG inmediatamente aguas abajo de la protoespaciador restringe la capacidad de orientar por término medio cada 12 pb en el genoma humano.

Típicamente, en el contexto de un sistema de CRISPR endógeno, la formación de un complejo de CRISPR (que comprende una secuencia guía hibridada a una secuencia diana y formando complejo con una o más proteínas Cas)

resulta en la escisión de una o las dos cadenas en o cerca de (p. ej., dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, o más pares de bases) de la secuencia diana. Sin desear estar limitados por la teoría, la secuencia tracr, que puede comprender o consistir en la totalidad o una porción de una secuencia tracr de tipo salvaje (p. ej., aproximadamente o más de aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85, o más nucleótidos de una secuencia tracr de tipo salvaje), también pueden formar parte de un complejo de CRISPR, tal como mediante la hibridación a lo largo de al menos una porción de la secuencia de tracr a la totalidad o una parte de una secuencia de emparejamiento tracr que está enlazada operativamente a la secuencia guía. En algunas realizaciones, uno o más vectores que impulsan la expresión de uno o más elementos de un sistema de CRISPR se introducen en una célula huésped de manera que la expresión de los elementos del sistema de CRISPR dirigen la formación de un complejo de CRISPR en uno o más sitios diana. Por ejemplo, una enzima Cas, una secuencia guía enlazada a una secuencia tracr-emparejamiento, y una secuencia de tracr podrían ser unidas operativamente en cada caso para separar elementos reguladores en vectores separados. Alternativamente, dos o más de los elementos expresados a partir de los mismos o diferentes elementos reguladores, se pueden combinar en un solo vector, con uno o más vectores adicionales que proporcionan cualesquiera componentes del sistema de CRISPR no incluidos en el primer vector. Elementos del sistema de CRISPR que se combinan en un único vector pueden estar dispuestos en cualquier orientación adecuada, tal como un elemento situado en 5' con respecto a ("aguas arriba" de) o 3' con respecto a ("aguas abajo" de) un segundo elemento. La secuencia codificadora de un elemento puede estar situada en la misma o en una cadena diferente de la secuencia codificadora de un segundo elemento, y estar orientada en la misma o en dirección opuesta. En algunas realizaciones, un único promotor dirige la expresión de un transcrito que codifica una enzima de CRISPR y una o más de la secuencia guía, secuencia de emparejamiento tracr (opcionalmente enlazada operativamente a la secuencia de guía) y una secuencia de tracr incrustada dentro de una o más secuencias de intrones (p. ej., cada una en un intrón diferente, dos o más en al menos un intrón, o todas en un solo intrón). En algunas realizaciones, la enzima de CRISPR, la secuencia guía, la secuencia de emparejamiento tracr y la secuencia tracr están enlazadas operativamente a y son expresada a partir del mismo promotor.

En algunas realizaciones, un vector comprende uno o más sitios de inserción, tales como una secuencia de reconocimiento de endonucleasa de restricción (al que también se alude como un "sitio de clonación"). En algunas realizaciones, uno o más sitios de inserción (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más sitios de inserción) están situados aguas arriba y/o aguas abajo de uno o más elementos de secuencia de uno o más vectores. En algunas realizaciones, un vector comprende un sitio de inserción aguas arriba de una secuencia de emparejamiento tracr y, opcionalmente, aguas abajo de un elemento regulador enlazado operativamente a la secuencia de emparejamiento tracr, de tal manera que después de la inserción de una secuencia guía en el sitio de inserción y tras la expresión de la secuencia guía dirige la unión específica para la secuencia de un complejo CRISPR a una secuencia diana en una célula eucariota. En algunas realizaciones, un vector comprende dos o más sitios de inserción, estando situado cada uno de los sitios de inserción entre dos secuencias de emparejamiento tracr con el fin de permitir la inserción de una secuencia guía en cada uno de los sitios. En una disposición de este tipo, las dos o más secuencias guía pueden comprender dos o más copias de una única secuencia guía, dos o más secuencias guía diferentes, o combinaciones de éstas. Cuando se utilizan múltiples secuencias guía diferentes, se puede utilizar una única construcción de expresión para fijar como objetivo la actividad de CRISPR a múltiples secuencias diana diferentes, correspondiente, dentro de una célula. Por ejemplo, un único vector puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, o más secuencias guía. En algunas realizaciones, pueden proporcionarse aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más de dichos vectores que contienen la secuencia guía, y opcionalmente suministrarse a una célula.

En algunas realizaciones, un vector comprende un elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR tal como una proteína Cas. Ejemplos no limitantes de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas 10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homólogos de las mismas, o versiones modificadas de las mismas. En algunas realizaciones, la enzima de CRISPR no modificada tiene actividad de escisión de ADN, tal como Cas9. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas en la ubicación de una secuencia diana, tal como dentro de la secuencia diana y/o dentro del complemento de la secuencia diana. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas dentro de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, o más pares de bases desde el primer o el último nucleótidos de una secuencia diana. En algunas realizaciones, un vector codifica una enzima de CRISPR que está mutada con respecto a una enzima de tipo salvaje correspondiente, de tal manera que la enzima de CRISPR mutada carece de la capacidad de escindir una o ambas cadenas de un polinucleótido diana que contiene una secuencia diana. Por ejemplo, una sustitución-aspartato-a alanina (D10A) en el dominio catalítico RuvC I de Cas9 de S. pyogenes convierte Cas9 de una nucleasa que escinde ambas hebras en una nicasa (escinde una sola cadena). Otros ejemplos de mutaciones que hacen de Cas9 una nicasa incluyen, sin limitación, H840A, N854A y N863A en SpCas9. Tal como se comenta en esta memoria, posiciones correspondientes puede estar conservadas en otras Cas9s, es decir, en ortólogos Cas9 de o derivados de otras especies bacterianas, con referencia a la numeración de la posición de SpCas9. (Figura 19) muestra un alineamiento de secuencias múltiples de 12 ortólogos Cas9 e indica el residuo

Asp catalítico conservado en el dominio RuvC I y el residuo His catalítico conservado en el dominio HNH. La mutación de uno u otro residuo en Ala pueden convertir el ortólogo Cas9 en una nickase. La mutación de los dos residuos puede convertir el ortólogo Cas9 en un mutante catalíticamente nulo - útil para la unión a ADN genérico. Como un ejemplo adicional, dos o más dominios catalíticos de Cas9 (RuvC I, RuvC II y RuvC III o el dominio HNH) pueden mutarse para producir un ortólogo Cas9 mutado que carece sustancialmente de toda actividad de escisión de ADN. En algunas realizaciones, una mutación D10A se combina con uno o más de mutaciones H840A, N854A o N863A para producir una enzima Cas9 que carece sustancialmente de toda actividad de escisión de ADN. En algunas realizaciones, una enzima de CRISPR se considera que carece sustancialmente de toda actividad de escisión de ADN cuando la actividad de escisión de ADN de la enzima mutada es menor que aproximadamente 25%, 10%, 5%, 1%, 0,1%., 0,01%, o inferior con respecto a su forma no mutada.

Una sustitución aspartato-a-alanina (D10A) en el dominio catalítico RuvC I de SpCas9 fue diseñada para convertir la nucleasa en una nickase (SpCas9n) (véase, p. ej., Sapranauskas et al., 2011, Nucleic Acid Research, 39: 9275; Gasiunas et al., 2012, Proc. Natl. Acad. Sci. USA, 109:E2579), de manera que el ADN genómico mellado sufre la reparación dirigida a homología de alta fidelidad (HDR). La solicitante utilizó el ensayo SURVEYOR para confirmar que SpCas9n no genera indeles en la diana protoespaciador EMX1. Se observó que la co-expresión de crRNA quimérico que fija como objetivo EMX1 (que tiene también el componente tracrRNA) con SpCas9 produjo indeles en el sitio diana, mientras que la co-expresión con SpCas9n no lo hizo (n = 3). Además de ello, la secuenciación de 327 amplicones no detectó indeles inducidos por SpCas9n. El mismo locus fue seleccionado para testar HR mediada por CRISPR al co-transfectar células HEK 293FT con EMX1, hSpCas9 o hSpCas9n que fija como objetivo ARN quimérico, así como un molde de HR para introducir un par de sitios de restricción (HindIII y NheI) cerca del protoespaciador. De hecho, SpCas9 y SpCas9n catalizaban la integración del molde HR en el locus EMX1. La amplificación por PCR de la región diana, seguido de la digestión por restricción con HindIII reveló productos de escisión que corresponden a tamaños de los fragmentos esperados, mediando SpCas9 y SpCas9n niveles similares de las eficiencias de HR. La solicitante verificó, además, HR utilizando la secuenciación de Sanger de amplicones genómicos y demostró la utilidad de CRISPR para facilitar la inserción de genes fijada como objetivo en el genoma de mamíferos. Dada la especificidad diana de 14 pb (12 pb desde el espaciador y 2-pb desde el PAM) de la SpCas9 de tipo salvaje, la disponibilidad de un nickase puede reducir significativamente la probabilidad de modificaciones fuera de objetivo, ya que roturas de una cadena sencilla no son sustratos para la vía NHEJ propenso a errores. La Figura 10A-M proporciona un esquema que indica las posiciones de las mutaciones en SpCas9 y ortólogos Cas9 comparten típicamente la organización general de los dominios 3-4 RuvC y un dominio HNH. El dominio RuvC más próximo a 5’ escinde la cadena no complementaria, y el dominio HNH escinde la cadena complementaria.

El residuo catalítico en el dominio RuvC 5' se identifica a través de la comparación de homología de la Cas9 de interés con otros ortólogos Cas9 (de locus de CRISPR de tipo II de S. pyogenes, locus 1 de CRISPR de S. thermophilus, locus 3 de CRISPR de S. thermophilus y locus de CRISPR de tipo II de Franciscilla novicida), y el residuo Asp conservada está mutado a alanina para convertir Cas9 en una enzima de mellado de cadena complementaria. De manera similar, los residuos His y Asn conservados en los dominios HNH están mutados a Alanina para convertir Cas9 en una enzima de mellado de cadena no complementaria.

En algunas realizaciones, una secuencia codificadora de la enzima que codifica una enzima CRISPR está optimizada en codones para la expresión en células particulares, tales como células eucariotas. Las células eucariotas pueden ser las de o derivadas de un organismo particular, tal como un mamífero, incluyendo, pero no limitado a, ser humano, ratón, rata, conejo, perro o mamífero no humano, incluyendo primate no humano. En algunas realizaciones, se pueden excluir procesos para modificar la identidad genética de la línea germinal de seres humanos y/o procesos para modificar la identidad genética de animales que es probable que les provoque sufrimiento sin beneficio médico sustancial alguno para el hombre o los animales, y también animales que resultan de estos procesos.

En general, la optimización de codones se refiere a un proceso de modificar una secuencia de ácido nucleico para la expresión potenciada en las células huéspedes de interés mediante el reemplazo de al menos un codón (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 10, 15 , 20, 25, 50, o más codones) de la secuencia nativa con codones que son con más frecuencia o lo más frecuentemente utilizados en los genes de esa célula huésped, al tiempo que se mantiene la secuencia de aminoácidos nativa. Diversas especies exhiben una desviación particular para ciertos codones de un aminoácido particular. Desviación del codón (diferencias en el uso de codones entre organismos) se correlaciona a menudo con la eficacia de la traducción de ARN mensajero (ARNm) que, es a su vez, se cree que depende, entre otros, de las propiedades de los codones que están siendo traducidos y la disponibilidad de moléculas de ARN de transferencia (ARNt) particulares. El predominio de los ARNt seleccionados en una célula es, por lo general, un reflejo de los codones más frecuentemente utilizados en la síntesis de péptidos. Por consiguiente, los genes pueden adaptarse para la expresión génica óptima en un organismo dado sobre la base de la optimización de codones. Tablas de uso de codones están fácilmente disponibles, por ejemplo, en la base de datos "Codon Usage Database", disponible en www.kazusa.orjp/codon/(visitada el 9 de julio de 2002), y estas tablas pueden ser adaptados en un cierto número de maneras. Véase Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence database status for the year 2000" Nucl. Acids Res. 28:292 (2000). También están

disponibles algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en una célula huésped particular, tal como Gene Forge (Aptagen; Jacobus, PA). En algunas realizaciones, uno o más codones (p. ej., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, o más, o todos los codones) en una secuencia que codifica una enzima de CRISPR corresponden al codón utilizado con mayor frecuencia para un aminoácido particular.

En algunas realizaciones, un vector codifica una enzima de CRISPR que comprende una o más secuencias de localización nuclear (NLSs), tal como aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLSs. En algunas realizaciones, la enzima de CRISPR puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo amino, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo carboxi, o una combinación de éstas (p. ej., una o más NLS en el extremo amino y una o más NLS en el extremo carboxi). Cuando está presente más de una NLS, cada una se puede seleccionar independientemente de las otras, de manera que una sola NLS puede estar presente en más de una copia y/o en combinación con una o más de otras NLSs presentes en una o más copias. En una realización preferida de la invención, la enzima de CRISPR comprende a lo sumo 6 NLSs. En algunas realizaciones, una NLS se considera cerca del extremo N o C cuando el aminoácido más cercano de la NLS está dentro de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 o más aminoácidos a lo largo de la cadena de polipéptido desde el extremo N o C. Ejemplos no limitantes de NLSs incluyen una secuencia NLS derivada de: la NLS del antígeno T grande del virus SV40, que tiene la secuencia de aminoácidos PKKKRKV; la NLS del nucleoplasma (p. ej., la NLS bipartita nucleoplasmina con la secuencia KRPAATKKAGQAKKKK); las NLS c-myc que tienen la secuencia de aminoácidos PAAKRVKLD o RQRRNELKRSP; la NLS hRNPAI M9 que tiene la secuencia NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; la secuencia RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV del dominio IBB de importina-alfa; las secuencias VSRKRPRP y PPKKARED de la proteína T mioma; la secuencia POPKKKPL de p53 humana; la secuencia SALIKKKKKMAP de ratón c-ab1 IV; las secuencias DRLRR y PQKKRK del virus NS1 de la gripe; la secuencia RKLKKKIKKL del antígeno delta del virus de la hepatitis; la secuencia REKKKFLKRR de la proteína Mx1 de ratón; la secuencia KRKGDEVDGVDEVAKKKSKK de la poli(ADP-ribosa)polimerasa humana; y la secuencia RKCLQAGMNLEARKTKK del glucocorticoide de los receptores de hormonas esteroides (humanas).

En general, la una o más NLSs son de una fuerza suficiente para impulsar la acumulación de la enzima de CRISPR en una cantidad detectable en el núcleo de una célula eucariota. En general, la fuerza de la actividad de localización nuclear puede derivar del número de NLSs en la enzima de CRISPR, la o las NLSs particulares utilizadas, o una combinación de estos factores. La detección de la acumulación en el núcleo puede ser realizada por cualquier técnica adecuada. Por ejemplo, un marcador detectable puede estar fusionado a la enzima CRISPR, de manera que puede ser visualizada la ubicación dentro de una célula, tal como en combinación con medios para detectar la ubicación del núcleo (p. ej., un colorante específico para el núcleo, tal como DAPI). Los núcleos de las células también pueden aislarse de las células, cuyos contenidos pueden luego ser analizados mediante cualquier procedimiento adecuado para la detección de proteínas, tal como inmunohistoquímica, transferencia Western o un ensayo de actividad enzimática. La acumulación en el núcleo también puede determinarse indirectamente tal como mediante un ensayo para el efecto de la formación de complejos CRISPR (p. ej., ensayo para la escisión o mutación de ADN en la secuencia diana, o ensayo para la actividad de la expresión génica alterada, afectada por la formación de complejos de CRISPR y/o actividad de la enzima de CRISPR), en comparación con un control no expuesto a la enzima o complejo CRISPR, o expuesto a una enzima de CRISPR que carece de la una o más NLSs.

En general, una secuencia guía es cualquier secuencia de polinucleótidos que tiene suficiente complementariedad con una secuencia de polinucleótidos diana para hibridarse con la secuencia diana y dirigir la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana. En algunas realizaciones, el grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean óptimamente utilizando un algoritmo de alineación adecuado, se trata de o más de aproximadamente 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99%, o más. Un alineamiento óptimo puede ser determinado con el uso de cualquier algoritmo adecuado para el alineamiento de secuencias, ejemplos no limitativos de los cuales incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsch, algoritmos basados en la Transformada de Burrows-Wheeler (p. ej., el Alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; disponible en www.novocmft.com), ELAND (Illumina, San Diego, CA), SOAP (disponible en soap.genomics.org.cn) y Maq (disponible en maq.sourceforge.net). En algunas realizaciones, una secuencia guía está relacionada en aproximadamente o más de aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 o más nucleótidos de longitud. En algunas realizaciones, una secuencia guía es menor que aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, o menos nucleótidos de longitud. La capacidad de una secuencia de guía para dirigir unión específica para la secuencia de un complejo CRISPR a una secuencia diana se puede determinar mediante cualquier ensayo adecuado. Por ejemplo, los componentes de un sistema de CRISPR, suficientes para formar un complejo de CRISPR, incluyendo la secuencia guía a ensayar, se pueden proporcionar a una célula huésped que tiene la secuencia diana correspondiente, tal como por transfección con vectores que codifican los componentes de la secuencia CRISPR, seguido de un estudio de escisión preferencial dentro de la secuencia diana tal como el ensayo Surveyor como se describe en esta memoria. De manera similar, la escisión de una secuencia de polinucleótido diana puede ser evaluada en un tubo de ensayo proporcionando la secuencia diana, componentes de un complejo de CRISPR, que incluye la secuencia guía

a ensayar y una secuencia guía de control diferente de la secuencia guía de ensayo, y comparando la unión o la tasa de escisión en la secuencia diana entre las reacciones de ensayo y de la secuencia guía de control. Son posibles otros ensayos, y se les ocurrirán a los expertos en la técnica.

Nickase multiplezda: Aspectos de optimización y las enseñanzas de Cas9 detallada en esta solicitud también se pueden utilizar para generar nickases Cas9. Nickases Cas9 se pueden utilizar ventajosamente en combinación con pares de ARNs guía para generar roturas de la doble cadena de ADN con colgantes definidos. Cuando se utilizan dos pares de ARNs guía, es posible escindir un fragmento de ADN intermedio. Si un trozo exógeno de ADN es escindido por los dos pares de ARNs guía para generar colgantes compatibles con el ADN genómico, entonces el fragmento de ADN exógeno se puede ligar en el ADN genómico para reemplazar el fragmento escindido. Por ejemplo, esto se puede utilizar para eliminar la expansión de repetición de trinucleótidos en el gen huntingtina (HIT para el tratamiento de la enfermedad de Huntington.

Cas9 y su ARN guía quimérico, o combinación de tracrRNA y crRNA, se pueden suministrar como ADN o ARN. El suministro de Cas9 y ARN guía ambos como moléculas de ARN (normal o con un contenido de bases o modificaciones de la cadena principal) se puede utilizar para reducir la cantidad de tiempo que la proteína Cas9 persiste en la célula. Esto puede reducir el nivel de la actividad de escisión fuera de objetivo en la célula diana. Dado que el suministro de Cas9 como ARNm requiere tiempo para ser traducido en proteína, podría ser ventajoso suministrar el ARN guía varias horas después del suministro de ARNm de Cas9, para maximizar el nivel de ARN guía disponible para la interacción con la proteína Cas9.

En situaciones en las que la cantidad de ARN guía es limitante, puede ser deseable introducir Cas9 como ARNm y ARN guía en forma de una casete de expresión de ADN con un promotor que impulsa la expresión del ARN guía. De esta manera la cantidad de ARN guía disponible será amplificada a través de la transcripción.

Se puede introducir una diversidad de sistemas de suministro para introducir Cas9 (ADN o ARN) y ARN guía (ADN o ARN) en la célula huésped. Éstos incluyen el uso de liposomas, vectores virales, electroporación, nanopartículas, nanohilos (Shalek et al., Nano Letters, 2012), exosomas. Liposomas de caballos de Troya moleculares (Pardridge et al., Cold Spring Harb Protoc; 2010; doi: 10.1101/pdb.prot5407) se pueden utilizar para suministrar Cas9 y ARN guía a través de la barrera hematoencefálica.

Discusión de ARNs guía para ortólogos: En algunas realizaciones, se selecciona una secuencia guía para reducir el grado de estructura secundaria dentro del ARN guía. En algunas realizaciones, aproximadamente o menos de aproximadamente 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, o menos de los nucleótidos del ARN guía participa en el apareamiento de bases auto-complementario cuando se pliegan de manera óptima. El plegamiento óptimo se puede determinar por cualquier algoritmo de plegamiento de polinucleótidos adecuado. Algunos programas se basan en el cálculo de la energía libre de Gibbs mínima. Un ejemplo de un algoritmo de este tipo es mFold, tal como se describe por Zuker y Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Otro ejemplo de algoritmo de plegamiento es el servidor web en línea RNAfold, desarrollado en el Instituto de Química Teórica de la Universidad de Viena, utilizando el algoritmo de predicción de estructura centroide (véase, p. ej. A.R. Gruber et al., 2008, Cell 106(1):23-24, y PA Carr y GM Church 2009, Nature Biotechnology 27(12): 1151-62). Un método de optimización del ARN guía de un ortólogo Cas9 comprende romper tramos poliU en el ARN guía. Tramos poliU que pueden romperse puede comprender una serie de 4, 5, 6, 7, 8, 9 ó 10 Us.

En general, una secuencia de emparejamiento tracr incluye cualquier secuencia que tiene una suficiente complementariedad con una secuencia tracr para fomentar uno o más de: (1) la escisión de una secuencia guía flanqueada por secuencias de emparejamiento tracr en una célula que contiene la secuencia tracr correspondiente; y (2) la formación de un complejo de CRISPR en una secuencia diana, en donde el complejo CRISPR comprende la secuencia de emparejamiento tracr hibridada a la secuencia de tracr. En general, el grado de complementariedad es con referencia al alineamiento óptimo de la secuencia de emparejamiento tracr y la secuencia tracr, a lo largo de la longitud de la más corta de las dos secuencias. El alineamiento óptimo puede ser determinado por cualquier algoritmo de alineamiento adecuado, y puede representar, además, estructuras secundarias tales como la auto-complementariedad dentro de la secuencia de tracr o la secuencia de emparejamiento tracr. En algunas realizaciones, el grado de complementariedad entre la secuencia tracr y la secuencia de emparejamiento tracr a lo largo de la longitud de la más corta de las dos cuando se alinean óptimamente es de aproximadamente o más de aproximadamente 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99%, o superior. En algunas realizaciones, la secuencia de tracr es aproximadamente o más de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, o más nucleótidos de longitud. En algunas realizaciones, la secuencia tracr y la secuencia de emparejamiento tracr están contenidas dentro de un único transcrito, de modo que la hibridación entre los dos produce un transcrito que tiene una estructura secundaria, tal como una horquilla. En una realización de la invención, el transcrito o secuencia de polinucleótido transcrita tiene al menos dos o más horquillas. En realizaciones preferidas, el transcrito tiene dos, tres, cuatro o cinco horquillas. En una realización adicional de la invención, el transcrito tiene a lo sumo cinco horquillas. En una estructura de horquilla, la parte de la secuencia 5' del final "N" y aguas arriba del bucle corresponde a la secuencia de emparejamiento de tracr, y la parte de la secuencia 3' del

bucle corresponde a la secuencia de tracr. Ejemplos no limitantes adicionales de polinucleótidos sencillos que comprenden una secuencia guía, una secuencia de emparejamiento de tracr y una secuencia de tracr son como sigue (listados de 5' a 3'), en donde "N" representa una base de una secuencia guía, el primer bloque de letras en minúscula representa la secuencia de emparejamiento de tracr y el segundo bloque de letras en minúscula representa la secuencia tracr, y la secuencia poli-T final representa el terminador de la transcripción: 1) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggctt catgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT; (2) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca acaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT; (3) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca acaccctgtcattttatggcagggtgtTTTTTT; (4) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaa agtggcaccgagtcggtgcTTTTTT; (5) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaa aaagtgTTTTTTT; y (6) NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTT TTT. En algunas realizaciones, las secuencias (1) a (3) se utilizan en combinación con Cas9 de S. thermophilius CRISPR1. En algunas realizaciones, las secuencias (4) a (6) se utilizan en combinación con Cas9 de *S. pyogenes*. En algunas realizaciones, la secuencia de tracr es un transcrito separado de un transcrito que puede comprender la secuencia de emparejamiento de tracr.

Discusión de los emparejamientos de tracr para ortólogos: En algunas realizaciones, también se proporciona un molde de recombinación. Un molde de recombinación puede ser un componente de otro vector tal como se describe en esta memoria, contenido en un vector separado, o proporcionado como un polinucleótido separado. En algunas realizaciones, un molde de recombinación está diseñado para servir como un molde en la recombinación homóloga tal como dentro o cerca de una secuencia diana mellada o escindida por una enzima de CRISPR como parte de un complejo de CRISPR. Un polinucleótido molde puede ser de cualquier longitud adecuada, tal como aproximadamente o más de aproximadamente 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, o más nucleótidos de longitud. En algunas realizaciones, el polinucleótido molde es complementario a una parte de un polinucleótido que puede comprender la secuencia diana. Cuando se alinean óptimamente, un polinucleótido molde podría solaparse con uno o más nucleótidos de una secuencia diana (p. ej., aproximadamente o más de aproximadamente 1, 5, 10, 15, 20, o más nucleótidos). En algunas realizaciones, cuando una secuencia de un molde y un polinucleótido, que comprende una secuencia diana, se alinean de manera óptima, el nucleótido más próximo del polinucleótido del molde está dentro de aproximadamente 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, o más nucleótidos de la secuencia diana.

En algunas realizaciones, la enzima de CRISPR es parte de una proteína de fusión que comprende uno o más dominios de proteínas heterólogos (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más dominios, además de la enzima de CRISPR). Una proteína de fusión enzima de CRISPR puede comprender cualquier secuencia de la proteína adicional, y opcionalmente una secuencia de unión entre cualquiera de dos dominios. Ejemplos de dominios de proteínas que pueden fusionarse con una enzima de CRISPR incluyen, sin limitación, macadores de epítopos, secuencias de gen indicador y dominios de proteínas que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de la modificación de histonas, actividad de escisión del ARN y actividad de unión a ácido nucleico. Ejemplos no limitantes de marcadores de epítopos incluyen marcadores de histidina (His), marcadores V5, marcadores FLAG, marcadores de hemaglutinina de la gripe (HA), marcadores Myc, marcadores VSV-G y marcadores de tiorredoxina (Trx). Ejemplos de genes informadores incluyen, pero no se limitan a, glutatión-S-transferasa (GST), peroxidasa de rábano picante (HRP), cloranfenicol acetiltransferasa (CAT), beta-galactosidasa, beta-glucuronidasa, luciferasa, proteína fluorescente verde (GFP), HcRed, DsRed, proteína fluorescente cian (CFP), proteína amarilla fluorescente (YFP) y proteínas autofluorescentes incluyendo la proteína fluorescente azul (BFP). Una enzima de CRISPR puede fusionarse a una secuencia de gen que codifica una proteína o un fragmento de una proteína que une moléculas de ADN o une otras moléculas celulares, incluyendo pero no limitada a la proteína de unión a maltosa (MBP), S-tag, fusiones del dominio de unión a ADN (DBD) de Lex A, fusiones del dominio de unión a ADN de GAL4 y fusiones de la proteína BP16 del virus herpes simplex (HSV). Dominios adicionales que pueden formar parte de una proteína de fusión que puede comprender una enzima de CRISPR se describen en el documento US20110059502. En algunas realizaciones, se utiliza una enzima de CRISPR marcada para identificar la ubicación de una secuencia diana.

En algunas realizaciones, una enzima de CRISPR puede formar un componente de un sistema inducible. La naturaleza inducible del sistema permitiría un control de espacio-temporal de la edición gen o la expresión génica utilizando una forma de energía. La forma de energía puede incluir, pero no se limita a la radiación electromagnética, energía sonora, energía química y energía térmica. Ejemplos de sistema inducible incluyen promotores de tetraciclina inducibles (Tet-On o Tet-Off), sistemas de activaciones de transcripción de dos híbridos de moléculas pequeñas (FKBP, ABA, etc.), o sistemas inducibles por la luz (fitocromo, dominios LOV, o criptocromo). En una realización, la enzima CRISPR puede ser una parte de un Efector Transcripcional Inducible por la Luz (LITE) para

dirigir cambios en la actividad transcripcional de una manera específica para la secuencia. Los componentes de una luz pueden incluir una enzima de CRISPR, un heterodímero citocromo sensible a la luz (p. ej., de Arabidopsis thaliana), y un dominio de activación/represión transcripcional. Ejemplos adicionales de proteínas de unión a ADN inducibles y métodos para su uso se proporcionan en los documentos US 61/736465 y US 61/721.283.

En algunos aspectos, la invención proporciona métodos que comprenden suministrar uno o más polinucleótidos, tales como o uno o más vectores tal como se describen en esta memoria, uno o más transcritos de los mismos, y/o una o más proteínas transcritas de los mismos, a una célula huésped. En algunos aspectos, la invención proporciona, además, células producidas por tales métodos, y animales que comprenden o son producidos a partir de tales células. En algunas realizaciones, una enzima de CRISPR en combinación con (y opcionalmente formando complejo con) una secuencia guía se suministra a una célula. Métodos de transferencia génica basados en virus y no virales convencionales pueden utilizarse para introducir ácidos nucleicos en células de mamífero o tejidos diana. Tales métodos se pueden utilizar para administrar ácidos nucleicos que codifican componentes de un sistema de CRISPR para células en cultivo, o en un organismo huésped. Sistemas de suministro de vectores no virales incluyen plásmidos de ADN, ARN (p. ej., un transcrito de un vector descrito en esta memoria), ácido nucleico desnudo y ácido nucleico formando complejo con un vehículo de suministro, tal como un liposoma. Sistemas de suministro de vectores virales incluyen virus de ADN y ARN, que tienen genomas episomales o integrados después dl suministro a la célula. Para una revisión de procedimientos de terapia génica, véase Anderson, Science 256: 808- 813 (1992); Nabel y Felgner, TIBTECH 11: 211-217 (1993); Mitani y Caskey, TIBTECH 11;162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357: 455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer y Perricaudet. British Medical Bulletin 51(1):31-44 (1995); Haddada et al., en Current Topics in Microbiology and Immunology Doerfler y Böhm (comps.) (1995); y Yu et al, Gene Therapy 1:13-26 (1994).

Métodos de suministro no viral de ácidos nucleicos incluyen lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, policatión o conjugados de lípido:ácido nucleico, ADN desnudo, viriones artificiales y absorción potenciada por agente de ADN. La lipofección se describe, p. ej., en las patente de EE.UU. Nºs. 5.049.386, 4.946.787; y 4.897.355) y reactivos de lipofección se venden comercialmente (p. ej., Transfectam™ y Lipofectin™). Lípidos catiónicos y neutros que son adecuados para una lipofección de reconocimiento de receptores eficiente de polinucleótidos incluyen los de Felgner, documento WO 91/17424; documento WO 91/16024. El suministro puede ser a células (p. ej., administración in vitro o ex vivo) o tejidos diana (p. ej., la administración in vivo).

La preparación de complejos de lípido:ácido nucleico, incluyendo liposomas fijados como objetivo tales como complejos de inmunolípidos, es bien conocida para un experto en la técnica (véase, p. ej., Crystal, Science 270: 404-410 (1995); Blaese et al , Cancer Gene Ther. 2: 291-297 (1995); Behr et al, Bioconjugate Chem. 5: 382-389 (1994); Remy et al, Bioconjugate Chem 5: 647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52: 4817-4820 (1992); patentes de EE.UU. Nºs 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028 y 4.946.787).

El uso de sistemas basados en virus de ARN o ADN para el suministro de ácidos nucleicos aprovecha los procesos altamente evolucionados para fijar como objetivo un virus a células específicas en el cuerpo y el tráfico de la carga útil viral al núcleo. Vectores virales pueden administrarse directamente a los pacientes (in vivo) o pueden utilizarse para tratar células in vitro y las células modificadas se pueden opcionalmente administrar a pacientes (ex vivo). Sistemas basados en virus convencionales podrían incluir vectores retroviral, de lentivirus, de virus adeno-asociados y de virus del herpes simplex para la transferencia de genes. La integración en el genoma huésped es posible con métodos de transferencia de genes de retrovirus, lentivirus y virus adeno-asociados, a menudo resulta en la expresión a largo plazo del transgén insertado. Adicionalmente, las elevadas eficiencias de la transducción se han observado en muchos tipos de células y tejidos diana diferentes.

El tropismo de un retrovirus puede alterarse por incorporación de proteínas de la envoltura extraña y expansión de la población diana potencial de células diana. Los vectores lentivirales son vectores retrovirales que son capaces de transducir o infectar células que no se dividen y que producen típicamente altos títulos virales. Por lo tanto, la selección de un sistema de transferencia de genes retroviral dependería del tejido diana. Los vectores retrovirales se componen de repeticiones cis-terminales de larga acción con capacidad de empaquetamiento para hasta 6-10 kb de secuencia extraña. LTRs que actúan en cis mínimos son suficientes para la replicación y el empaquetamiento de los vectores, que se utilizan luego para integrar el gen terapéutico en la célula objetivo para proporcionar una expresión del transgen permanente. Vectores retrovirales ampliamente utilizados incluyen los basados en virus de la leucemia murina (MuLV), virus de la leucemia del mono gibón (GaLV), virus de la inmunodeficiencia de simios (SIV), virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véase, p. ej., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al, J. Virol. 66: 1635-1640 (1992); Sommnerfelt et al, Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63: 2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); documento PCT/US94/05700).

En aplicaciones en las que se prefiere la expresión transitoria, se pueden utilizar sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una muy alta eficiencia de transducción en muchos tipos de células y no requieren la división celular. Con este tipo de vectores se han obtenido un título y niveles de expresión altos. Este vector puede producirse en grandes cantidades en un sistema relativamente simple.

Vectores de virus adeno-asociados ("AAV") también pueden utilizarse para transducir células con ácidos nucleicos diana, p. ej., en la producción in vitro de ácidos nucleicos y péptidos, y para procedimientos de terapia génica in vivo y ex vivo (véase, p. ej., West et al., Virology 160:38-47 (1987); patente de EE.UU. Nº. 4.797.368; documento WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). La construcción de vectores AAV recombinantes se describe en un cierto número de publicaciones, incluyendo la patente de EE.UU. Nº. 5.173.414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat y Muzyczka, PNAS 81:6466-6470 (1984); y Samulski et al., J. Virol. 63: 03822-3828 (1989).

Células de empaquetamiento se utilizan típicamente para formar partículas de virus que son capaces de infectar una célula huésped. Tales células incluyen células 293, que empaquetan adenovirus, y células ψ2 o células PA317, que empaquetan retrovirus. Los vectores virales utilizados en terapia génica se generan habitualmente por el productor de una línea celular que empaqueta un vector de ácido nucleico en una partícula viral. Los vectores contienen típicamente las secuencias virales mínimas requeridas para el empaquetamiento y la posterior integración en un huésped, siendo reemplazadas otras secuencias virales por una casete de expresión para el o los polinucleótidos a expresar. Las funciones virales que faltan se suministran normalmente en trans por la línea celular de empaquetamiento. Por ejemplo, los vectores AAV utilizados en terapia génica poseen típicamente sólo secuencias ITR del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma huésped. El ADN viral se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, a saber rep y cap, pero que carece de secuencias ITR. La línea celular también puede estar infectada con adenovirus como un helper (cooperador). El virus helper fomenta la replicación del vector AAV y la expresión de genes AAV a partir del plásmido helper. El plásmido helper no está empaquetado en cantidades significativas debido a una carencia de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, p. ej., tratamiento térmico al que el adenovirus es más sensible que AAV. Métodos adicionales para el suministro de ácidos nucleicos a las células son conocidos por los expertos en la técnica. Véase, por ejemplo, el documento US20030087817.

Por consiguiente, el AAV se considera un candidato ideal para su uso como un vector de transducción. Tales vectores de transducción AAV pueden comprender suficientes funciones que actúan en cis para replicar en presencia de adenovirus o virus herpes o virus de la viruela (p. ej., virus vacuna) funciones helper proporcionadas en trans. AAV recombinante (rAAV) puede utilizarse para transportar genes exógenos a células de una diversidad de linajes. En estos vectores, los genes cap y/o rep de AAV se eliminan del genoma viral y se reemplazan por un segmento de ADN de elección. Vectores AAV actuales pueden alojar hasta 4300 bases de ADN insertado.

Existe un cierto número de maneras de producir rAAV, y la invención proporciona rAAV y métodos para preparar rAAV. Por ejemplo, un plásmido o plásmidos que contienen o consisten esencialmente en la construcción viral deseada se transfectan en células infectadas con AAV. Además, un segundo o adicional plásmido helper se co-transfecta en estas células para proporcionar los genes cap y/o rep de AAV que son obligatorias para replicación y el empaquetamiento de la construcción viral recombinante. Bajo estas condiciones, las proteínas rep y/o cap de AAV actúan en trans para estimular la replicación y el empaquetamiento de la construcción de rAAV. Dos a tres días después de la transfección, se recoge el rAAV. Tradicionalmente el rAAV se recoge de las células junto con adenovirus. El adenovirus contaminante se inactiva luego por tratamiento térmico. En la presente invención, el rAAV se recoge ventajosamente no de las propias células, sino a partir de sobrenadante celular. Por consiguiente, en un aspecto inicial, la invención proporciona preparar rAAV, y además de lo anterior, rAAV se puede preparar por un método que comprende o consiste esencialmente en: infectar células susceptibles con un ADN exógeno que contiene rAAV, incluyendo ADN para la expresión, y virus helper (p. ej., adenovirus, virus herpes, virus de la viruela tales como el virus vacuna) en el que el rAAV carece de cap y/o rep funcionales (y el virus helper (p. ej., adenovirus, virus herpes, virus de la viruela tales como el virus vacuna) proporciona la función cap y/o rep de la que carece el rAAV); o infectar células susceptibles con un rAAV que contiene ADN exógeno incluyendo ADN para la expresión, en donde el recombinante carece de cap y/o rep funcional, y transfectar dichas células con un plásmido que suministra la función cap y/o rep de la que carece el rAAV; o infectar células susceptibles con un rAAV que contiene ADN exógeno incluyendo ADN para la expresión, en donde el recombinante carece de cap y/o rep funcional, en donde dichas células suministran la función cap y/o rep de la que carece el recombinante carece; o transfectar las células susceptibles con un AAV que carecen de cap y/o rep funcional y plásmidos para la inserción de ADN exógeno en el recombinante, de modo que el ADN exógeno se expresa por el recombinante y para suministrar las funciones rep y/o cap, con lo que la transfección resulta en un rAAV que contiene el ADN exógeno incluyendo ADN para la expresión que carece de cap y/o rep funcional.

El rAAV puede ser de un AAV tal como se describe en esta memoria, y ventajosamente puede ser un rAAV1, rAAV2, AAV5 o rAAV que tiene un híbrido o una cápside que puede comprender AAV1, AAV2, AAV5 o cualquier combinación de los mismos. Se puede seleccionar el AAV del rAAV con respecto a las células a ser fijadas como

objetivo por el rAAV; p. ej., se puede seleccionar AAV serotipos 1, 2, 5 o un híbrido o la cápside de AAV 1, AAV2, AAV5 o cualquier combinación de los mismos para fijar como objetivo células del cerebro o neuronales; y se puede seleccionar AAV4 para fijar como objetivo tejido cardíaco, AAV8 para fijar como objetivo tejido del hígado.

Además de las células 293, en la práctica de la invención se pueden utilizar otras células y la infectividad relativa de determinados serotipos de AAV in vitro en cuanto a estas células (véase Grimm, D. et al, J. Virol 82.: 5887-5911 (2008)) es la siguiente:

Línea Celular	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2,5	0,0	0,1	10	0,7	0,0
HEK293	25	100	2,5	0,1	0,1	5	0,7	0,1
HeLa	3	100	2,0	0,1	6,7	1	0,2	0,1
HepG2	3	100	16,7	0,3	1,7	5	0,3	ND
Hep1A	20	100	0,2	1,0	0,1	1	0,2	0,0
911	17	100	11	0,2	0,1	17	0,1	ND
CHO	100	100	14	1,4	333	50	10	1,0
COS	33	100	33	3,3	5,0	14	2,0	0,5
MeWo	10	100	20	0,3	6,7	10	1,0	0,2
NIH3T3	10	100	2,9	2,9	0,3	10	0,3	ND
A549	14	100	20	ND	0,5	10	0,5	0,1
HT1180	20	100	10	0,1	0,3	33	0,5	0,1
Monocitos	1111	100	ND	ND	125	1429	ND	ND
DC inmadura	2500	100	ND	ND	222	2857	ND	ND
DC madura	2222	100	ND	ND	333	3333	ND	ND

La invención proporciona rAAV que contiene o consiste esencialmente en una molécula de ácido nucleico exógeno que codifica un sistema de CRISPR (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), p. ej., una pluralidad de casetes que comprenden o consisten en una primera casete que comprende o consiste esencialmente en un promotor, una molécula de ácido nucleico que codifica una proteína CRISPR-asociada (Cas) (proteínas nucleasa o helicasa putativas), p. ej., Cas9 y un terminador, y dos, o más, de manera ventajosa hasta el límite de tamaño de empaquetamiento del vector, p. ej., en total (incluyendo la primera casete) de cinco casetes que comprenden o consisten esencialmente en un promotor, una molécula de ácido nucleico que codifica ARN guía (gRNA) y un terminador (p. ej., cada una de las casetes representada esquemáticamente como Promotor-gRNA1-terminador, Promotor-gRNA2-terminador…. Promotor-gRNA(N)-terminador (en donde N es un número que puede ser insertado que se encuentra en un límite superior del límite de tamaño de empaquetamiento del vector), o dos o más rAAVs individuales, cada uno de los cuales contiene una o más de una casete de un sistema de CRISPR, p. ej., un primer rAAV que contiene la primera casete que comprende o que consiste esencialmente en un promotor, una molécula de ácido nucleico que codifica Cas, p. ej., Cas9 y un terminador y un segundo rAAV que contiene una pluralidad, cuatro, casetes que comprende o consiste esencialmente en un promotor, molécula de ácido nucleico que codifica ARN guía (gRNA) y un terminador (p. ej., cada una de las casetes representada esquemáticamente como Promotor-gRNA1-terminador, Promotor-gRNA2-terminador, Promotor-gRNA(N)-terminador (en donde N es un número que puede ser insertado que se encuentra en un límite superior del límite de tamaño de empaquetamiento del vector). Dado que el rAAV es un virus de ADN, las

moléculas de ácido nucleico en la discusión en esta memoria en relación con AAV o rAAV son ventajosamente ADN. El promotor es en algunas realizaciones ventajosamente promotor de Sinapsina 1 humana (hSyn).

Se prefieren dos formas de empaquetar moléculas de ácido nucleico codificadoras de Cas9, p. ej., ADN, en vectores virales para mediar en la modificación del genoma in vivo:
Para lograr una inactivación del mediada por NHEJ:
Vector de virus sencillo:
Vector que contiene dos o más casetes de expresión:
Promotor-molécula de ácido nucleico que codifica Cas9-terminador
Promotor-gRNA1-terminador
Promotor-gRNA2-terminador
Promotor-gRNA(N)-terminador (hasta el límite de tamaño del vector)
Vector de virus doble:
Vector 1 que contiene una casete de expresión para impulsar la expresión de Cas9
Promotor-molécula de ácido nucleico que codifica Cas9-terminador
Vector 2 que contiene uno o más casetes de expresión para impulsar la expresión de uno o más RNAs guía
Promotor-gRNA1-terminador
Promotor-gRNA(N)-terminador (hasta el límite de tamaño del vector)

Mediar en la reparación directa de la homología. Además de los enfoques de vector de virus sencillo y doble descritos anteriormente, se utiliza un vector adicional para suministrar un molde de reparación directa de la homología.

El promotor utilizado para impulsar la expresión de la molécula de ácido nucleico que codifica Cas9 puede incluir:
AAV ITR puede servir como un promotor: esto es ventajoso para eliminar la necesidad de un elemento promotor adicional (que puede ocupar espacio en el vector). El espacio adicional liberado se puede utilizar para impulsar la expresión de elementos adicionales (gRNA, etc.). Además, la actividad ITR es relativamente más débil, por lo que se puede utilizar para reducir la toxicidad debido a la sobre-expresión de Cas9.

Para la expresión ubicua, se pueden utilizar promotores: CMV, CAG, CBh, PGK, SV40, cadenas pesada o ligera de ferritina, etc.

Para la expresión en el cerebro, se pueden utilizar promotores: Sinapslna I para todas las neuronas, CaMKIIalpha para las neuronas excitadoras, GAD67 o GAD65 o VGAT para las neuronas GABAérgicas, etc.
Para la expresión en el hígado, se puede utilizar promotor de albúmina
Para la expresión en pulmón, se puede utilizar SP-B
Para células endoteliales, se puede utilizar ICAM
Para células hematopoyéticas, se puede utilizar IFNbeta o CD45,
Para osteoblastos, se puede utilizar OG-2
El promotor utilizado para impulsar ARN guía puede incluir:
Promotores Pol III tales como U6 o H1
Uso de promotor Pol II y casetes intrónicas para expresar gRNA

En cuanto a AAV, el AAV puede ser AAV1, AAV2, AAV5 o cualquier combinación de los mismos. Se puede seleccionar el AAV del AAV en relación con las células a fijar como objetivo; p. ej., se puede seleccionar AAV serotipos 1, 2, 5 o un AAV1, AAV2, AAV5 híbrido o de la cápside o cualquier combinación de los mismos para fijar como objetivo el cerebro o células neuronales; y se puede seleccionar AAV4 para fijar como objetivo tejido cardíaco. AAV8 es útil para el suministro al hígado. Los promotores y vectores anteriores se prefieren individualmente.

En algunas realizaciones, una célula huésped es transfectada de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria. En algunas realizaciones, una célula se transfecta tal como se produce de forma natural en un sujeto. En algunas realizaciones, una célula que se transfecta se toma de un sujeto. En algunas realizaciones, la célula se deriva de células tomadas de un sujeto tal como una línea celular. En la técnica se conoce una amplia diversidad de líneas celulares para el cultivo de tejidos. Ejemplos de líneas celulares incluyen, pero no se limitan a, C8161 CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCeII, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 epitelial del riñón de mono, fibroblastos de embrión de ratón BALB/ 3T3, 3T3 Swiss, 3T3-L1, fibroblastos fetales humanos 132-d5; fibroblastos de ratón 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3,

EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, células JY, células K562, Ku812, KCL22, KG1, KEYO1, LNCap, Ma-ME1 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SkBr3, T2, T-47D, T84, línea celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, YAR, y variedades transgénicas de las mismas. Las líneas celulares están disponibles de una diversidad de fuentes conocidas por los expertos en la técnica (véase, p. ej., American Type Culture Collection (ATCC) (Manassus, Va.)). En algunas realizaciones, una célula transfectada con uno o más vectores descritos en esta memoria se utiliza para establecer una nueva línea celular que puede comprender una o más secuencias derivadas del vector. En algunas realizaciones, una célula transfectadas transitoriamente con los componentes de un sistema de CRISPR, tal como se describe en esta memoria (por ejemplo, por transfección transitoria de uno o más vectores, o la transfección con ARN), y modificadas a través de la actividad de un complejo de CRISPR, se utiliza para establecer una nueva línea celular que puede comprender células que contienen la modificación, pero que carecen de cualquier otra secuencia exógena. En algunas realizaciones, las células transfectadas de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria, o líneas celulares derivadas de estas células se utilizan en la evaluación de uno o más compuestos de ensayo.

En algunas realizaciones, uno o más vectores descritos en esta memoria se utilizan para producir un animal transgénico no humano o planta transgénica. En algunas realizaciones, el animal transgénico es un mamífero, tal como un ratón, rata o conejo. Métodos para producir plantas y animales transgénicos son conocidos en la técnica, y generalmente comienzan con un método de transfección celular tal como se describe en esta memoria.

Con los recientes progresos en genómica de cultivos, la capacidad de utilizar los sistemas de CRISPR-Cas para llevar a cabo la edición y la manipulación de genes eficiente y rentable permitirá la rápida selección y la comparación de manipulaciones genéticas individuales y multiplexadas para transformar dichos genomas para una producción mejorada y rasgos potenciados. A este respecto se hace referencia a las patentes y publicaciones de EE.UU. Patente de EE.UU. Nº 6.603.061 - Agrobacterium-Mediated Plant Transformation Method; patente de EE.UU. Nº. 7.868.149 - Plant Genome Sequences and Uses Thereof y de EE.UU. 2009/0100536 – Transgenic Plants with Enhanced Agronomic Traits. En la práctica de la invención, también se citan en la misma el contenido y la divulgación de Morrell et al “Crop genomics: advances and applications” Nat Rev Genet. 29 de diciembre de 2011; 13(2):85-96. En una forma de realización ventajosa de la invención, el sistema CRISPR/Cas9 se utiliza para modificar genéticamente microalgas (Ejemplo 7). Por consiguiente, referencia en esta memoria a células animales también se puede aplicar, mutatis mutandis, a células vegetales, a menos que de otro modo resulte evidente.

En un aspecto, la invención proporciona métodos de modificar un polinucleótido diana en una célula eucariota, que puede ser in vivo, ex vivo o in vitro. En algunas realizaciones, el método comprende el muestreo de una célula o población de células de un animal humano o no humano o planta (incluyendo micro-algas), y la modificación de la célula o células. El cultivo puede producirse en cualquier fase ex vivo. La célula o las células pueden incluso ser re-introducidos en el animal no humano o la planta (incluyendo micro-algas). Para las células reintroducidas se prefiere particularmente que las células sean células madre.

En algunas realizaciones, el método comprende permitir que un complejo de CRISPR se una al polinucleótido diana para efectuar la escisión de dicho polinucleótido diana, modificando así el polinucleótido diana, en el que el complejo de CRISPR puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido diana, en el que dicha secuencia guía está unida a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr.

En un aspecto, la invención proporciona un método de modificar la expresión de un polinucleótido en una célula eucariota. En algunas realizaciones, el método comprende permitir que un complejo de CRISPR se una al polinucleótido de manera que dicha unión resulte en una expresión incrementada o reducida de dicho polinucleótido; en el que el complejo de CRISPR comprende una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido, en el que dicha secuencia guía está enlazada a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr. Consideraciones similares se aplican como antes a métodos para modificar un polinucleótido diana. De hecho, estas opciones de muestreo, cultivo y re-introducción se aplican a lo largo de los aspectos de la presente invención.

En un aspecto, la invención proporciona kits que contienen uno o más de los elementos descritos en los métodos y las composiciones anteriores. Los elementos pueden proporcionarse individualmente o en combinaciones, y se pueden proporcionar en cualquier recipiente adecuado, tal como un vial, una botella, o un tubo. En algunas realizaciones, el kit incluye instrucciones en uno o más idiomas, por ejemplo, en más de un idioma.

En algunas realizaciones, un kit comprende uno o más reactivos para su uso en un proceso que utiliza uno o más de los elementos descritos en esta memoria. Los reactivos pueden proporcionarse en cualquier recipiente adecuado.

Por ejemplo, un kit puede proporcionar uno o más tampones de reacción o de almacenamiento. Los reactivos se pueden proporcionar en una forma que sea utilizable en un ensayo particular, o en una forma que requiere la adición de uno o más de otros componentes antes de su uso (p. ej., en forma de concentrado o liofilizada). Un tampón puede ser cualquier tampón, incluyendo, pero no limitado a un tampón carbonato de sodio, un tampón bicarbonato de sodio, un tampón borato, un tampón Tris, un tampón MOPS, un tampón HEPES, y combinaciones de los mismos. En algunas realizaciones, el tampón es alcalino. En algunas realizaciones, el tampón tiene un pH de aproximadamente 7 a aproximadamente 10. En algunas realizaciones, el kit comprende uno o más oligonucleótidos correspondientes a una secuencia guía para la inserción en un vector con el fin de enlazar operativamente la secuencia guía y un elemento regulador. En algunas realizaciones, el kit comprende un polinucleótido molde de recombinación homóloga.

En un aspecto, la invención proporciona métodos para usar uno o más elementos de un sistema de CRISPR. El complejo de CRISPR de la invención proporciona un medio eficaz para la modificación de un polinucleótido diana. El complejo de CRISPR de la invención tiene una amplia variedad de utilidad, incluyendo modificar (por ejemplo, suprimir, insertar, translocar, inactivar, activar) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo de CRISPR de la invención tiene un amplio espectro de aplicaciones en, p. ej., la terapia génica, el rastreo de fármacos, el diagnóstico de enfermedades y el pronóstico. Un complejo de CRISPR a modo de ejemplo puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro del polinucleótido diana. La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia de tracr.

En una realización, esta invención proporciona un método para escindir un polinucleótido diana. El método puede comprender modificar un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido diana y efectúa la escisión de dicho polinucleótido diana. Típicamente, el complejo de CRISPR de la invención, cuando se introduce en una célula, crea una ruptura (p. ej., una rotura de la cadena sencilla o doble) en la secuencia del genoma. Por ejemplo, el método puede utilizarse para escindir un gen de la enfermedad en una célula.

La ruptura creada por el complejo de CRISPR puede ser reparada por unos procesos de reparación tales como la vía de unión de extremos no homólogos (NHEJ) propensa a errores o la reparación dirigida a la homología de alta fidelidad (HDR). Durante estos procesos de reparación, un molde de polinucleótido exógeno puede ser introducido en la secuencia del genoma. En algunos métodos, el proceso de HDR se utiliza para modificar la secuencia del genoma. Por ejemplo, un molde de polinucleótido exógeno que puede comprender una secuencia que se ha de integrar flanqueada por una secuencia aguas arriba y una secuencia aguas abajo se introduce en una célula. Las secuencias aguas arriba y aguas abajo comparten similitud de secuencia con cualquiera de los lados del sitio de integración en el cromosoma.

Cuando se desee, un polinucleótido donante puede ser ADN, p. ej., un plásmido de ADN, un cromosoma bacteriano artificial (BAC), un cromosoma artificial de levadura (YAC), un vector viral, un trozo lineal de ADN, un fragmento de la PCR, un ácido nucleico desnudo, o un ácido nucleico formando complejo con un vehículo de suministro tal como un liposoma o poloxámero.

El molde de polinucleótido exógeno comprende una secuencia para ser integrada (p. ej., un gen mutado). La secuencia de integración puede ser una secuencia endógena o exógena a la célula. Ejemplos de una secuencia a ser integrada incluyen polinucleótidos que codifican una proteína o un ARN no codificante (p. ej., un microARN). Por lo tanto, la secuencia de la integración puede estar operativamente enlazada a una secuencia o secuencias de control apropiadas. Alternativamente, la secuencia a ser integrada puede proporcionar una función reguladora.

Las secuencias de aguas arriba y aguas abajo del molde de polinucleótido exógeno se seleccionan para fomentar la recombinación entre la secuencia cromosómica de interés y el polinucleótido donante. La secuencia de aguas arriba es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia del genoma de aguas arriba del sitio diana para la integración. De manera similar, la secuencia de aguas abajo es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia cromosómica aguas abajo del sitio diana de la integración. Las secuencias aguas arriba y aguas abajo en el molde de polinucleótido exógeno puede tener 75%, 80%, 85%, 90%, 95% o 100% de identidad de secuencia con la secuencia del genoma de destino. Preferiblemente, las secuencias de aguas arriba y aguas abajo en la plantilla de polinucleótido exógeno tienen aproximadamente 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia del genoma fijada como objetivo. En algunos métodos, las secuencias de aguas arriba y aguas abajo en el molde de polinucleótido exógeno tienen aproximadamente el 99% o el 100% de identidad de la secuencia con la secuencia del genoma fijada como objetivo.

Una secuencia de aguas arriba o aguas abajo puede comprender de aproximadamente 20 pb a aproximadamente 2500 pb, por ejemplo, aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 ó 2500 pb. En algunos métodos, la secuencia aguas arriba o aguas abajo a modo de ejemplo tiene aproximadamente 200 pb a aproximadamente 2000 pb,

aproximadamente 600 pb a aproximadamente 1000 pb, o más particularmente aproximadamente 700 pb a aproximadamente 1000 pb.

En algunos métodos, el molde de polinucleótido exógeno puede comprender, además, un marcador. Este marcador puede hacer más fácil rastrear integraciones fijadas como objetivo. Ejemplos de marcadores adecuados incluyen sitios de restricción, proteínas fluorescentes o marcadores seleccionables. El molde de polinucleótido exógeno de la invención puede construirse utilizando técnicas recombinantes (véase, por ejemplo, Sambrook et al., 2001 y Ausubel et al., 1996).

En un método a modo de ejemplo, para la modificación de un polinucleótido diana mediante la integración de un molde de polinucleótido exógeno, se introduce en la secuencia del genoma una rotura de la cadena doble por el complejo de CRISPR, la ruptura se repara mediante recombinación homóloga de un molde de polinucleótido exógeno de tal forma que el molde se integra en el genoma. La presencia de una rotura de cadena doble facilita la integración del molde.

En otras realizaciones, esta invención proporciona un método para modificar la expresión de un polinucleótido en una célula eucariota. El método puede comprender aumentar o disminuir la expresión de un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido,

En algunos métodos, un polinucleótido diana puede ser inactivado para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo de CRISPR a una secuencia diana en una célula, el polinucleótido diana se inactiva de manera que la secuencia no se transcribe, la proteína codificada no se produce o la secuencia no funciona como la secuencia de tipo salvaje. Por ejemplo, una proteína o secuencia codificadora de microARN puede inactivarse de manera que no se produzca la proteína.

En algunos métodos, una secuencia de control puede ser inactivada de manera que ya no funcione como una secuencia de control. Tal como se utiliza en esta memoria, "secuencia de control" se refiere a cualquier secuencia de ácido nucleico que efectúa la transcripción, traducción o la accesibilidad de una secuencia de ácido nucleico. Ejemplos de una secuencia de control incluyen un promotor, un terminador de la transcripción, y un promotor son secuencias de control.

La secuencia diana inactivada puede incluir una mutación por deleción (es decir, la deleción de uno o más nucleótidos), una mutación por inserción (es decir, la inserción de uno o más nucleótidos), o una mutación sin sentido (es decir, la sustitución de un único nucleótido con otro nucleótido de modo que se introduce un codón de parada). En algunos métodos, la inactivación de una secuencia diana da como resultado la "inactivación" de la secuencia diana.

Una expresión alterada de una o más secuencias del genoma asociadas con una vía de señalización bioquímica puede ser determinada por ensayo de una diferencia en los niveles de ARNm de los genes correspondientes entre la célula modelo de ensayo y una célula control, cuando están en contacto con un agente candidato. Alternativamente, la expresión diferencial de las secuencias asociadas con una vía de señalización bioquímica se determina mediante la detección de una diferencia en el nivel del producto génico o polipéptido codificado.

Para ensayar para una alteración inducida por el agente en el nivel de transcritos de ARNm o polinucleótidos correspondientes, ácido nucleico contenido en una muestra se extrae primero de acuerdo con métodos estándares en la técnica. Por ejemplo, el ARNm puede aislarse utilizando diversas enzimas líticas o disoluciones químicas de acuerdo con los procedimientos establecidos en Sambrook et al. (1989), o puede extraerse mediante resinas de unión a ácido nucleico siguiendo las instrucciones adjuntas proporcionadas por los fabricantes. El ARNm contenido en la muestra de ácido nucleico extraído se detecta a continuación mediante procedimientos de amplificación o ensayos de hibridación convencionales (p. ej., análisis de transferencia Northern) de acuerdo con métodos ampliamente conocidos en la técnica o en base a los métodos ejemplificados en esta memoria.

Para los fines de esta invención, amplificación significa cualquier método que emplea un cebador y una polimerasa capaz de replicar una secuencia diana con fidelidad razonable. La amplificación puede ser portada por el ADN-polimerasas naturales o recombinantes tales como TaqGold™, T7 ADN polimerasa, el fragmento Klenow de la ADN polimerasa de E. coli y transcriptasa inversa. Un método de amplificación preferido es la PCR. En particular, el ARN aislado se puede someter a un ensayo de transcripción inversa que está acoplado con una reacción en cadena de la polimerasa cuantitativa (RT-PCR) con el fin de cuantificar el nivel de expresión de una secuencia asociada con una vía de señalización bioquímica.

La detección del nivel de expresión génica puede llevarse a cabo en tiempo real en un ensayo de amplificación. En un aspecto, los productos amplificados pueden ser directamente visualizados con agentes de unión a ADN

fluorescentes, incluyendo, pero no limitados a, intercaladores de ADN y aglomerantes de de la muesca de ADN. Debido a que la cantidad de los intercaladores incorporados en las moléculas de ADN de doble cadena es típicamente proporcional a la cantidad de los productos de ADN amplificados, se puede determinar convenientemente la cantidad de los productos amplificados mediante la cuantificación de la fluorescencia del colorante intercalado utilizando sistemas ópticos convencionales en la técnica. Colorante de unión a ADN adecuado para esta aplicación incluyen verde SYBR, azul SYBR, DAPI, yoduro de propidio, Hoeste, oro SYBR, bromuro de etidio, acridinas, proflavina, naranja de acridina, acriflavina, fluorcoumanina, elipticina, daunomicina, cloroquina, distamicina D, cromomicina, homidio, mitramicina, polipiridilos rutenio, antramicina, y similares.

En otro aspecto, otros marcadores fluorescentes tales como sondas específicas para la secuencia se pueden emplear en la reacción de amplificación para facilitar la detección y cuantificación de los productos amplificados. La amplificación cuantitativa basada en la sonda se basa en la detección específica para la secuencia de un producto amplificado deseado. Utiliza, sondas específicas para la diana, fluorescentes (p. ej., sondas TaqMan®) resultando en una mayor especificidad y sensibilidad. Métodos para realizar la amplificación cuantitativa basada en la sonda están bien establecidos en la técnica y se enseña en la Patente de EE.UU. Nº 5.210.015.

Todavía en otro aspecto, se pueden realizar ensayos de hibridación convencionales utilizando sondas de hibridación que comparten homología de secuencia con las secuencias asociadas con una vía de señalización bioquímica. Típicamente, se deja que las sondas formen complejos estables con las secuencias asociadas con una vía de señalización bioquímica contenida dentro de la muestra biológica derivada del sujeto de ensayo en una reacción de hibridación. Se apreciará por un experto en la técnica que, cuando se utilice antisentido como el ácido nucleico sonda, los polinucleótidos diana proporcionados en la muestra se eligen para que sean complementarios a las secuencias de los ácidos nucleicos antisentido. En cambio, cuando la sonda de nucleótidos es un ácido nucleico sentido, el polinucleótido diana se selecciona para que sea complementario a las secuencias del ácido nucleico sentido.

La hibridación puede realizarse en condiciones de diferentes rigurosidad, por ejemplo tal como se describen en esta memoria. Condiciones de hibridación adecuadas para la práctica de la presente invención son tales que la interacción de reconocimiento entre la sonda y secuencias asociadas con una vía de señalización bioquímica es a la vez lo suficientemente específica y lo suficientemente estable. Condiciones que aumentan la rigurosidad de una reacción de hibridación son ampliamente conocidas y están publicadas en la técnica. Véase, por ejemplo, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, segunda edición). El ensayo de hibridación se puede formar utilizando sondas inmovilizadas sobre cualquier soporte sólido, incluyendo, pero no limitadas a nitrocelulosa, vidrio, silicio, y una diversidad de matrices de genes. Un ensayo de hibridación preferido se lleva a cabo en chips de genes de alta densidad tal como se describe en la Patente de EE.UU. Nº 5.445.934.

Para una detección conveniente de los complejos de sonda-diana que se forman durante el ensayo de hibridación, las sondas de nucleótidos se conjugan a un marcador detectable. Marcadores detectables adecuados para su uso en la presente invención incluyen cualquier composición detectable por medios fotoquímicos, bioquímicos, espectroscópicos, inmunoquímicos, eléctricos, ópticos o químicos. Una amplia variedad de marcadores detectables adecuados son conocidos en la técnica, que incluyen marcadores fluorescentes o quimioluminiscentes, marcadores de isótopos radiactivos, enzimáticos u otros ligandos. En realizaciones preferidas, probablemente se deseará emplear un marcador fluorescente o una etiqueta de enzima tal como digoxigenina, β-galactosidasa, ureasa, fosfatasa alcalina o peroxidasa, complejo de avidina/biotina.

Los métodos de detección utilizados para detectar o cuantificar la intensidad de hibridación dependerán típicamente del marcador arriba seleccionado. Por ejemplo, los marcadores radiactivos se pueden detectar utilizando una película fotográfica o un phosphoimager. Los marcadores fluorescentes pueden ser detectados y cuantificados utilizando un fotodetector para detectar la luz emitida. Los marcadores enzimáticos se detectan típicamente proveyendo a la enzima de un sustrato y midiendo el producto de reacción producido por la acción de la enzima sobre el sustrato; y finalmente marcadores colorimétricos son detectadas simplemente visualizando el marcador coloreado.

Un cambio inducido por agente en la expresión de secuencias asociadas con una vía de señalización bioquímica también se puede determinar mediante el examen de los productos génicos correspondientes. La determinación del nivel de proteína implica típicamente a) poner en contacto la proteína contenida en una muestra biológica con un agente que se une específicamente a una proteína asociada con una vía bioquímica de señalización; y (b) identificar cualquier complejo de agente:proteína así formado. En un aspecto de esta realización, el agente que une específicamente una proteína asociada con una vía de señalización bioquímica es un anticuerpo, preferiblemente un anticuerpo monoclonal.

La reacción se lleva a cabo poniendo en contacto el agente con una muestra de las proteínas asociadas con una vía de señalización bioquímica derivada de las muestras de ensayo en condiciones que permitan que se forme un complejo entre el agente y las proteínas asociadas con una vía de señalización bioquímica. La formación del complejo se puede detectar directa o indirectamente de acuerdo con procedimientos estándares en la técnica. En el método de detección directo, los agentes se suministran con un marcador detectable y agentes que no han reaccionado pueden separarse del complejo; la cantidad de marcador restante indicando con ello la cantidad de complejo formado. Para un método de este tipo, es preferible seleccionar los marcadores que permanecer unidos a los agentes, incluso durante condiciones de lavado rigurosas. Es preferible que el marcador no interfiera con la reacción de unión. En la alternativa, un procedimiento de detección indirecta puede utilizar un agente que contiene un marcador introducido química o eiizimáticamente. Un marcador deseable no interfiere generalmente con la unión o la estabilidad del complejo agente:polipéptido resultante. Sin embargo, el marcador se diseña típicamente para que sea accesible a un anticuerpo para una unión eficaz y, por lo tanto, para generar una señal detectable.

Una amplia diversidad de marcadores adecuados para la detección de los niveles de proteína son conocidos en la técnica. Ejemplos no limitantes incluyen radioisótopos, enzimas, metales coloidales, compuestos fluorescentes, compuestos bioluminiscentes y compuestos quimioluminiscentes.

La cantidad de complejos agente:polipéptido formados durante la reacción de unión se puede cuantificar mediante ensayos cuantitativos estándares. Tal como se ilustra anteriormente, la formación de complejo agente:polipéptido se puede medir directamente por la cantidad de marcador permaneció en el sitio de unión. En una alternativa, la proteína asociada con una vía de señalización bioquímica se testa en cuanto a su capacidad para competir con un análogo marcado por sitios de unión en el agente específico. En este ensayo competitivo, la cantidad de marcador capturado es inversamente proporcional a la cantidad de secuencias de proteínas asociadas con una vía de señalización bioquímica presente en una muestra de ensayo.

En la técnica está disponible un cierto número de técnicas para el análisis de proteínas sobre la base de los principios generales esbozados anteriormente. Estos incluyen, pero no se limitan a radioinmunensayos, ELISA (ensayos inmunorradiométricos ligados a enzima), inmunoensayos "sándwich", ensayos inmunorradiométricos, inmunoensayos in situ (utilizando, p. ej., oro coloidal, enzimas o radioisótopos etiquetas), análisis de transferencia Western, ensayos de inmunoprecipitación, ensayos de inmunofluorescencia y SDS-PAGE.

Anticuerpos que reconocen específicamente o se unen a proteínas relacionadas con una vía de señalización bioquímica son preferibles para la realización de los análisis de proteínas mencionados anteriormente. En los casos en los que se desee, se pueden utilizar los anticuerpos que reconocen un tipo específico de modificaciones post-traduccionales (p. ej., modificaciones inducibles en la vía bioquímica de señalización). Modificaciones post-traduccionales incluyen, pero no se limitan a glicosilación, lipidación, acetilación y fosforilación. Estos anticuerpos pueden adquirirse de proveedores comerciales. Por ejemplo, los anticuerpos anti-fosfotirosina que reconocen específicamente las proteínas tirosina-fosforiladas están disponibles de un número de proveedores, incluyendo Invitrogen y Perkin Elmer. Anticuerpos anti-fosfotirosina son particularmente útiles en la detección de proteínas que son fosforiladas diferencialmente en sus residuos de tirosina en respuesta a un estrés ER. Tales proteínas incluyen, pero no se limitan al factor 2 alfa de iniciación de la traducción eucariótica (eIF-2α). Alternativamente, estos anticuerpos pueden generarse utilizando tecnologías de anticuerpos policlonales o monoclonales convencionales mediante la inmunización de un animal huésped o una célula productora de anticuerpos con una proteína diana que exhibe la modificación post-traduccional deseada.

En la práctica del método en cuestión, puede ser deseable discernir el patrón de expresión de una proteína asociada con una vía de señalización bioquímica en diferentes tejidos corporales, en diferentes tipos de células y/o en diferentes estructuras subcelulares. Estos estudios se pueden realizar con el uso de anticuerpos específicos para el tejido, específicos para la célula o específicos para la estructura subcelular, capaces de unirse a marcadores de proteínas que se expresan preferentemente en determinados tejidos, tipos de célula o estructuras subcelulares.

Una expresión alterada de un gen asociado con una vía de señalización bioquímica también se puede determinar examinando un cambio en la actividad del producto génico con relación a una célula de control. El ensayo para un cambio inducido por agente en la actividad de una proteína asociada con una vía de señalización bioquímica dependerá de la actividad biológica y/o la vía de transducción de señales que está bajo investigación. Por ejemplo, en los casos en los que la proteína es una quinasa, un cambio en su capacidad para fosforilar el o los sustratos aguas abajo puede ser determinado por una diversidad de ensayos conocidos en la técnica. Ensayos representativos incluyen, pero no se limitan a inmunotransferencia e inmunoprecipitación con anticuerpos tales como anticuerpos anti-fosfotirosina que reconocen las proteínas fosforiladas. Además, la actividad quinasa se puede detectar por ensayos quimioluminiscentes de alto rendimiento tales como los ensayos AlphaScrea™ (disponible de Perkin Elmer) y eTag™ (Chan-Hui, et al. (2003) Clinical Immunology 111:162-174).

En los casos en los que la proteína asociada con una vía de señalización bioquímica es parte de una cascada de señalización que conduce a una fluctuación de la condición de pH intracelular, moléculas sensibles al pH tales como colorantes de pH fluorescentes se pueden utilizar como moléculas informadoras. En otro ejemplo en el que la proteína asociada a una vía de señalización bioquímica es un canal de iones, se pueden vigilar las fluctuaciones en el potencial de membrana y/o en la concentración intracelular de iones. Un cierto número de kits comerciales y dispositivos de alto rendimiento son particularmente adecuados para un rastreo rápido y robusto para moduladores de canales de iones. Instrumentos representativos incluyen FLIPRTM (Molecular Devices, Inc.) y VIPR (Aurora Biosciences). Estos instrumentos son capaces de detectar reacciones en pocillos de más de 1000 muestras de una microplaca de forma simultánea, y proporcionan una medición en tiempo real y datos funcionales en el espacio de un segundo o incluso un milisegundo.

En la práctica de cualquiera de los métodos descritos en esta memoria, un vector adecuado se puede introducir en una célula o un embrión a través de uno o más métodos conocidos en la técnica, incluyendo, sin limitación, microinyección, eiectroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección de liposomas, transfección de dendrímeros, transfección de choque térmico, transfección por nucleofección, magnetofección, lipofección, impalefección, transfección óptica, captación de propiedad reforzada por el agente de ácidos nucleicos y el suministro a través de liposomas, inmunoliposomas, virosomas o viriones artificiales. En algunos métodos, el vector se introduce en un embrión mediante microinyección. El vector o los vectores se pueden microinyectar en el núcleo o en el citoplasma del embrión. En algunos métodos, el vector o los vectores se pueden introducir en una célula por nucleofección.

El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura).

Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido "asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de transcripción o traducción en un nivel anormal o en una forma anormal en células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo . Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o el progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee mutación o mutaciones o una variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que es responsable de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.

El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora de un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura).

El polinucleótido diana de un complejo CRISPR puede incluir un cierto número de genes y polinucleótidos asociados a enfermedades, así como genes y polinucleótidos asociados a la vía de señalización bioquímica tal como se enumera en las solicitudes de patente provisional de EE.UU. 61/736.527 y 61/748.427 con referencia Broad BI-2011/008/WSGR Nº de Expediente 44063-701.101 y BI-2011/008/WSGR Nº de Expediente 44063-701.102, respectivamente, ambos con el título SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATIONS, presentadas el 12 de diciembre de 2012 y el 2 de enero de 2013, respectivamente.

Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido “asociado a la enfermedad” se refiere a cualquier gen o polinucleótido que está proporcionando productos de la transcripción o traducción en un nivel anormal o en una forma anormal a células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo. Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee la o las mutaciones o variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que son responsables de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.

Ejemplos de genes y polinucleótidos asociados a una enfermedad están disponibles en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web.

Ejemplos de genes y polinucleótidos asociados con la enfermedad se enumeran en las Tablas A y B. Información específica de la enfermedad está disponible en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web. Ejemplos de señalización de genes y polinucleótidos asociados a la vía de señalización bioquímica se enumeran en la Tabla C.

Mutaciones en estos genes y estas vías pueden resultar en la producción de proteínas inadecuadas o proteínas en cantidades inadecuadas que afectan a la función. Ejemplos adicionales de genes, enfermedades y proteínas son de las solicitudes Provisionales de EE.UU. 61/736.527 presentada el 12 de diciembre de 2012 y 61/748.427 presentada el 2 de enero de 2013. Genes, proteínas y vías de este tipo pueden ser el polinucleótido diana de un complejo de CRISPR.

Tabla A

ENFERMEDAD/TRASTORNOS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIFIa; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR
	gamma; WT1 (Tumor de Wilms); miembros de la Familia de receptors FGF
	(5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Receptor de Andrógeno); TSG101; IGF; Receptor de IGF1; Igf1 (4
	variantes); Igf2 (3 variantes); Receptor de Igf1; Receptor de Igf2;
	Bax; Bcl2; familia de caspasas (9 miembros:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Macular relacionada con la edad	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsinaD;
Degeneración	Vldlr; Ccr2
Esquizofrenia	Neuregulina1 (Nrg1); Erb4 (receptor para Neuregulina);
	Complexina1 (Cplx1); Tph1 Triptofanohidroxilasa; Tph2
Triptófano	hidroxilasa 2; Neurexina 1; GSK3; GSK3a;
	GSK3b

ENFERMEDAD/TRASTORNOS	GENE(S)
Trastornos	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Repetición de Trinucleótidos	HTT (Enfermedad de Huntington); SBMA/SMAX1/AR (Enfermedad de Kennedy);
Trastornos	FXN/X25 (Ataxia de Friedrich); ATX3 (Enfermedad de Machado-
	Joseph); ATXN1 and ATXN2 (ataxias espinocerebelares);
	DMPK (distrofia miotónica); Atrofins-1 y Atnl
	(Enfermedad DRPLA); CBP (Creb-BP – inestabilidad global); VLDLR
	(Alzheimer); Atxn7; Atxn10
Síndrome X Frágil	FMR2; FXR1; FXR2; mGLUR5
Relacionado con Secretasa	APH-1 (alfa y beta); Presenilina (Psen1); nicastrina
Trastornos	(Ncstn); PEN-2
Otros	Nos1; Parp1; Nat1; Nat2
Trastornos relacionados con Priones	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;
	VEGF-b; VEGF-c)
Adicción a fármacos	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2;
	Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autismo	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexina 1; Frágil X
	(FMR2 (AFF2); FXR1; FXR2; Mglur5)
Enfermedad de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS1;
	SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1,
	Aquaporina 1); Uchl1; Uchl3; APP
Inflamación	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a(CTLA8); IL-

ENFERMEDAD/TRASTORNOS	GENE(S)
	17b; IL-17c; IL,-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa;
	OD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3c11
Enfermedad de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkin; PINK1

Tabla B:

Enfermedades y trastornos de la sangre y coagulación	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); síndrome de linfocito desnudo (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Trastornos de hemorragia (TBXA2R, P2RX1, P2X1); Factor H y tipo factor H 1 (HF 1, CFH, HUS); Factor V y factor VIII (MCFD2); Deficiencia de factor VII (F7); Deficiencia de factor X (F10); Deficiencia de factor XI (F11); Deficiencia de factor XII (F 12, HAF); Deficiencia de factor XIIIA (F13A1, F13A); Deficiencia de factor XIIIB (F13B); anemia de Fanconi (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Trastornos de linfohistiocitosis hemofagocítica (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB), Trastornos hemorrágicos (PI, ATT, F5); Deficiencias y trastornos de leucocitos (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); anemia de células de Sickle (HBB); Talasemia (HBA2, HBB, HBD, LCRB, HBA1).
Enfermedades y trastornos de disregulación celular y oncología	Linfoma no-Hodgkin B-cel1 (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL 11, ARLTS 1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Enfermedades y trastornos relacionados con la inflamación y el sistema inmune	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Síndrome linfoproliferativo autoinmune (TNFRSF6, APT1, FAS, CD95, ALPS1A); Inmunodeficiencia combinada, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), Susceptibilidad o infección por VIH (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Inmunodeficiencias (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflamación (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, II,-17d, IL-17f), II-23, Cx3crl, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL- 12a, IL-12b), CTLA4, Cx3cl1); Inmunodeficiencias combinadas graves (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).

Enfermedades y trastornos metabólicos, del hígado, riñon y proteínas	Neuropatía amiloide (TTR, PALB); Amiloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Fibrosis quística (CFTR, ABCC7, CF, MRP7); Enfermedades de almacenamiento de glucógeno (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3), Insuficiencia hepática, brote temprano y trastorno neurológico (SCOD1, SCO1), Deficiencia de lipasa hepática (LIPC), Hepatoblastoma, cáncer y carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Enfermedad de riñón quístico medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); fenilcetonuria (PAH, PKU1, QDPR, DHPR, PTS); Riñón poliquístico y enfermedad hepática (FCYT, PKHD 1, ARPKD, PKD 1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Enfermedades y trastornos Musculares/ del Esqueleto	Distrofia muscular de Becker (DMD, BMD, MYF6), Distrofia muscular de Duchenne (DMD, BMD); Distrofia muscular de Emery-Dreifuss (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Distrofia muscular facioescapulohumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT 1, CAV3, LGMD1C, SEPN 1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Enfermedades y trastornos neurológicos y neuronales	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); enfermedad de Alzheimer (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP 1, ACE 1, MPO, PACIP 1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autismo (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexina 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Síndrome X frágil (FMR2, FXR1, FXR2, mGLUR55); Enfermedad y trastornos tipo enfermedad de Huntington (HD, IT 15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Enfermedad de Parkinson (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Síndrome de Rett (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Sinucleína, DJ-1); Esquizofrenia (Neuregulina1 (Nrg1), Erb4 (receptor para Neuregulina), Complexina1 (Cplx1), Tph1 Triptófano hidroxilasa, Tph2, Triptófano hidroxilasa 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Trastornos relacionados con secretasa (APH-1 (alfa y beta), Presenilina (Psen1), nicastrina, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trastornos de la repetición de trinucleótidos (HTT (Enfermedad de Huntington), SBMA/SMAX1/AR (Enfermedad de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Enfermedad de Machado-Joseph), ATXN1 y ATXN2 (ataxias espinocerebelares), DMPK (distrofia miotónica), Atrofina-1 y Atn1 (DRPLA Dx), CBP (Creb-BP – inestabilidad global), VLDLR (Alzheimer), Atxn7, Atxn10).
Enfermedades y trastornos oculares	Degeneración macular relacionada con la edad (Abcr, Cc12, Cc2, cp (ceruloplasmina), Timp3, catepsinaD, Vldlr, Ccr2); Cataratas (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2,

	CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Opacidad y distrofia de la córnea (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Córnea plana congénita (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Amaurosis congénita del hígado (CRB1, RP 12, CRX, CORD2, CRD, RPGRIP 1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Distrofia macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Tabla C:

FUNCIÓN CELULAR	GENES
Señalización PI3K/AKT	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; FPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	IPRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
Señalización ERK/MAPK	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;

FUNCIÓN CELULAR	GENES
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Señalización del Receptor de Glucocorticoides	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINA1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Señalización de Guía Axonal	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF 1; RAC 1; RAP1A; EIF4E; PRKCZ; NRP 1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;

FUNCIÓN CELULAR	GENES
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Señalización del Receptor Efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Señalización del Citoesqueleto de Actina	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
	IPRKAA2; FIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; 1TGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF 1; GSN; DYRK1A; ITGB 1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;

FUNCIÓN CELULAR	GENES
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Señalización de la Enfermedad de Huntington	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CLASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Señalización de la Apoptosis	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; CARP1
Señalización del Receptor de Células B	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	IMAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;

FUNCIÓN CELULAR	GENES
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Señalización de la Extravasación de Leucocitos	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	P1K3C3; MAPK; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MUMP9
Señalización de Integrina	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	IRLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Señalización de la Respuesta de Fase Aguda	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
	AKT2; IKBKB; PIK3GA; FOS; NFKB2; MAP3K14;

FUNCIÓN CELULAR	GENES
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL 1R1; IL6
Señalización de PTEN	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CLASP3; RPS6KB1
Señalización de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIQC3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP 1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Señalización del Receptor de Aril Hidrocarbono	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;

FUNCIÓN CELULAR	GENES
	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Señalización del Metabolismo Xenobiótico	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB 1; KEAP 1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP94AA1
Señalización de SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPS1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;

FUNCIÓN CELULAR	GENES
	CRKL; BRAF; SGK
Señalización de PPAr/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
Señalización de NF-KB	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAY6;
	TBKL1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL 1R1
Señalización de Neuregulina	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1

FUNCIÓN CELULAR	GENES
Señalización de Wnt y Beta catenina	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Señalización del Receptor de Insulina	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	AP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
Señalización de IL-6	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Colestasis Hepática	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABC1;

FUNCIÓN CELULAR	GENES
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
Señalización de IGF-1	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
Respuesta de Estrés Oxidativo mediada por NRF2	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Fibrosis Hepática/Activación de Células Estelares	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Señalización de PPAR	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;

FUNCIÓN CELULAR	GENES
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Señalización de Fc Epsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
Señalización del Receptor Acoplado a Proteína G	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; REB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Metabolismo de Inositol Fosfato	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
	MAPK8; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
MAPK8	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3G2A; DYRK1A; MAP2K2; PIP5KIA; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
Señalización de PDGF	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;

FUNCIÓN CELULAR	GENES
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
Señalización de VEGF	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
Señalización de Células Asesinas Naturales	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	P1K3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Ciclo Celular: GI/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Regulación del punto de verificación	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	IGSK3B; RBL1; HDAC6
Señalización del Receptor de Células T	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Señalización del Receptor de Muerte	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;

FUNCIÓN CELULAR	GENES
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPS1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
Señalización de FGF	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
Señalización de GM-GSF	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Señalización de la Esclerosis Lateral Amiotrófica	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGF; BIRC2; BAX; AKT3; CASP3; BIRC3
Señalización de JAK/Stat	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1

FUNCIÓN CELULAR	GENES
Metabolismo de Nicotinato y Nicotinamida	PRKCE; IRAK; PRKAA2; EIF2AK2; GRK6; MAPK1;
	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAF2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Señalización de Quimioquinas	CXCR4; ROCK2; MAPS1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
Señalización de IL-2	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Depresión Sináptica a Largo Plazo	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Señalización del Receptor de Estrógenos	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Vía de la Ubiquitinación de Proteínas	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;

FUNCIÓN CELULAR	GENES
	USP1; VHL; HSP90AA1; BIRC3
Señalización de IL-10	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
Activación de VDR/RXR	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
Señalización de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGEBR1; SMAD4; JUN; SMAD5
Señalización del Receptor de tipo Toll	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
Señalización de p38 MAPK	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K.7; TGFBR1; MYC; ATF4; IL 1R1;
	SRF; STAT1
Señalización de Neurotrofina/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;

FUNCIÓN CELULAR	GENES
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
Activación de FXR/RXR	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Potenciación Sináptica a Largo Plazo	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Señalización de Calcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
Señalización de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Señalización de Hipoxia en el	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
Sistema Cardiovascular	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
Inhibición Mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
de la Función RXR	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
Activación de LXR/RXR	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;

FUNCIÓN CELULAR	GENES
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
Procesamiento Amiloide	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
Señalización de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Ciclo Celular: G2/M Regulación del Punto	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
de Verificación de Daño a ADN	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
	PRKDC; ATM; SFN; CDKN2A
Señalización de Óxido Nítrico en el	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
Sistema Cardiovascular	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Metabolismo de Purina	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD 1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
Señalización mediada por cAMP	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunción Mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Señalización de Notch	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4

FUNCIÓN CELULAR	GENES
Vía del Estrés del Retículo Endoplásmico	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
	EIF2AK3; CASP3
Metabolismo de Pirimidina	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Señalización de Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Señalización Cardiaca y Beta Adrenérgica	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
	PPP2R5C
Glicolisis/Gluconeogénesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Señalización de Interferón	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Señalización Sonic Hedgehog	ARRB2; SMO; GL12; DYRK1A; GLI1; GSK3B; DYRK1B
Metabolismo de Glicerofosfolípidos	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Degradación de Fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo de triptófano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradación de Lisina	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Vía de la Reparación de la Escisión de Nucleótidos	ERCC5; ERCC4; XPA; XPC; ERCC1
Metabolismo de Almidón y Sacarosa	UCHL1; HK2; GCK; GPI; HK1
Metabolismo de Aminoazúcares	NQO1; HK2; GCK; HK1
Metabolismo del Ácido Araquidónico	PRDX6; GRN; YWHAZ; CYP1B1

FUNCIÓN CELULAR	GENES
Señalización del Ritmo Circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de Coagulación	BDKRB1; F2R; SERPINE1; F3
Señalización del Receptor de Dopamina	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Metabolismo de Glutation	IDH2; GSTP1; ANPEP; IDH1
Metabolismo de Glicerolípidos	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo del Ácido Linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo de Metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo de Piruvato	GLO1; ALDH1A1; PKM2; LDHA
Metabolismo de Arginina y Prolina	ALDH1A1; NOS3; NOS2A
Señalización de Eicosanoides	PRDX6; GRN; YWHAZ
Metabolismo de Fructosa y Manosa	HK2; GCK; HK1
Metabolismo de Galactosa	HK2; GCK; HK1
Biosíntesis de Estilbeno, Coumarina y	PRDX6; PRDX1; TYR
Lignina	
Vía de la Presentación de Antígenos	CALR; B2M
Biosíntesis de Esteroides	NQO1; DHCR7
Metabolismo de Butanoato	ALDH1A1; NLGN1
Ciclo del Citrato	IDH2; IDH1
Metabolismo de Ácidos Grasos	ALDH1A1; CYP1B1

FUNCIÓN CELULAR	GENES
Metabolismo de Glicerofosfolípidos	PRDX6; CHKA
Metabolismo de Histidina	PRMT5; ALDH1A1
Metabolismo de Inositol	ERO1L; APEX1
Metabolismo de Xenobióticos	GSTP1; CYP1B1
por el Citocromo p450	
Metabolismo del Metano	PRDX6; PRDX1
Metabolismo de Fenilalanina	PRDX6; PRDX1
Metabolismo de Propanoato	LDH1A1; LDHA
Metabolismo de Selenoaminoácido	PRMT5; AHCY
Metabolismo de Esfingolípidos	SPHK1; SPHK2
Metabolismo de Aminofosfonato	PRMT5
Metabolismo de Andrógenos y Estrógenos	PRMT5
Metabolismo de Ascorbato y Aldarato	ALDH1A1
Biosíntesis de Ácidos Biliares	ALDH1A1
Metabolismo de Cisteína	LDHA
Biosíntesis de Ácidos Grasos	FASN
Señalización del Receptor de Glutamato	GNB2L1

FUNCIÓN CELULAR	GENES
Respuesta al Estrés Oxidativo medida por NRF2	PRDX1
Vía del Fosfato de Pentosa	GPI
Interconversiones de Pentosa y Glucuronato	UCHL1
Metabolism de Retinol	ALDH1A1
Metabolismo de Riboflavina	TYR
Metabolismo de Tirosina	PRMT5, TYR
Biosíntesis de Ubiquinona	PRMT5
Degradación de Valina, Leucina e	ALDH1A1
Isoleucina	
Metabolismo de Glicina, Serina y	CHKA
Treonina	
Degradación de Lisina	ALDH1A1
Dolor/Sabor	TRPM5; TRPA1
Dolor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Función Mitocondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurología del desarrollo	BMP-4; Cordina (Chrd); Nogina (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenina;

FUNCIÓN CELULAR	GENES
	Dkk-1; Proteínas relacionadas con Frizzled; Otx-2; Gbx2; FGF-8;
	Reelina; Dab1; unc-86 (Pou4f1 o Brn3a); Numb; Reln

Realizaciones de la invención también se refieren a métodos y composiciones relacionados con la inactivación de genes, la amplificación de genes y la reparación de mutaciones particulares asociadas con la inestabilidad en la repetición de ADN y trastornos neurológicos (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Segunda Edición, Academic Press 13 de oct. de 2011 - Medical). Se ha encontrado que aspectos específicos de secuencias repetidas en tándem son los responsables de más de una veintena de enfermedades humanas (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep.-Oct.;7(5):551-8). El sistema de CRISPR-Cas puede ser aprovechado para corregir estos defectos de la inestabilidad genómica.

Un aspecto adicional de la invención se refiere a la utilización del sistema de CRISPR-Cas para la corrección de defectos en los genes EMP2A y EMP2B que han sido identificados que están asociados con la enfermedad de Lafora. La enfermedad de Lafora es una afección autosómica recesiva que se caracteriza por la epilepsia mioclónica progresiva que puede comenzar como convulsiones epilépticas en la adolescencia. Unos pocos casos de la enfermedad pueden ser provocados por mutaciones en genes aún por identificar. La enfermedad provoca convulsiones, espasmos musculares, dificultad para caminar, demencia y, finalmente, la muerte. Actualmente no existe una terapia que haya demostrado ser eficaz contra el progreso de la enfermedad. Otras anomalías genéticas asociadas con la epilepsia también pueden ser fijadas como objetivo por el sistema de CRISPR-Cas y la genética subyacente se describe con más detalle en Genetics of Epilepsy and Genetic Epilepsies, editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

Todavía en otro aspecto de la invención, el sistema de CRISPR-Cas puede utilizarse para corregir defectos oculares que surgen de varias mutaciones genéticas descritas adicionalmente en Genetic Diseases of the Eye, Segunda Edición, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Varios aspectos adicionales de la invención se refieren a la corrección de los defectos asociados con una amplia gama de enfermedades genéticas que se describe con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Disorders (sitio web en health.nih.gov/topic/GeneticDisorders). Las enfermedades genéticas del cerebro pueden incluir, pero no se limitan a adrenoleucodistrofia, agenesia del cuerpo calloso, síndrome de Aicardi, enfermedad de Alper, enfermedad de Alzheimer, síndrome de Barth, enfermedad de Batten, CADASIL, Degeneración cerebelosa, enfermedad de Fabry, enfermedad de Gerstmann-Straussier-Scheinker, enfermedad de Huntington y otros trastornos de la repetición de tripletes, enfermedad de Leigh, síndrome de Lesch-Nyhan, enfermedad de Menkes, miopatías mitocondriales y colpocefalía NINDS. Estas enfermedades se describen con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Brain Disorders.

En algunas realizaciones, la afección puede ser la neoplasia. En algunas realizaciones, en los casos en los que la afección es la neoplasia, los genes a ser fijados como objetivo son cualquiera de los enumerados en la Tabla A (en este caso PTEN, etcétera). En algunas realizaciones, la afección puede ser la degeneración macular relacionada con la edad. En algunas realizaciones, la afección puede ser un trastorno esquizofrénico. En algunas realizaciones, la afección puede ser un trastorno de repeticiones de trinucleótidos. En algunas realizaciones, la afección puede ser el síndrome X frágil. En algunas realizaciones, la afección puede ser un trastorno relacionado con secretasa. En algunas realizaciones, la afección puede ser un trastorno relacionado con Priones. En algunas realizaciones, la afección puede ser ALS. En algunas realizaciones, la afección puede ser una adicción a los fármacos. En algunas realizaciones, la afección puede ser el autismo. En algunas realizaciones, la afección puede ser la enfermedad de Alzheimer. En algunas realizaciones, la afección puede ser la inflamación. En algunas realizaciones, la afección puede ser la enfermedad de Parkinson.

Ejemplos de proteínas asociadas con la enfermedad de Parkinson incluyen, pero no se limitan a α-sinucleína, DJ-1, LRRK2, Pink1, Parkin, UCHL1, Sinfilina-1 y NURR1.

Ejemplos de proteínas relacionadas con la adicción pueden incluir ABAT, por ejemplo.

Ejemplos de proteínas relacionadas con la inflamación pueden incluir la proteína quimioatrayente de monocitos 1 (MCP1) codificada por el gen Ccr2, el receptor de quimioquina C-C tipo 5 (CCR5) codificado por el gen Ccr5, el

receptor IIB de IgG (FCGR2b, también denominado CD32) codificado por el gen Fcgr2b o la proteína Fc epsilon R1g (FCER1g) codificada por el gen FcerIg, por ejemplo.

Ejemplos de proteínas asociadas a enfermedades cardiovasculares pueden incluir IL1B (interleuquina 1, beta), XDH (xantina deshidrogenasa), TP53 (proteína tumoral p53), PTGIS (prostaglandina 12 (prostaciclina) sintasa), MB (mioglobina), IL4 (interleuquina 4), ANGPT1 (angiopoyetina 1), ABCG8 (casete de unión a ATP, subfamilia G (WHITE) miembro, 8), o CTSK (catepsina K), por ejemplo.

Ejemplos de proteínas asociadas a la enfermedad de Alzheimer pueden incluir la proteína receptor de lipoproteínas de muy baja densidad (VLDLR) codificada por el gen VLDLR, la enzima activante del modificador de tipo ubiquitina 1 (UBA1) codificada por el gen UBA1 o la proteína de la subunidad catalítica E1 de la enzima activante de NEDD8 (UBE1C) codificada por el gen UBA3, por ejemplo.

Ejemplos de proteínas asociadas al trastorno del espectro del autismo pueden incluir la proteína 1 asociada al receptor de benzodiazepina (periférico) (BZRAP1) codificada por el gen BZRAP1, la proteína del miembro 2 de la familia AF4/FMR2 (AFF2) codificada por el gen AFF2 (también denominada MFR2), la proteína homóloga 1 autosómica del retardo mental X frágil (FXR1) codificada por el gen FXR1 o la proteína homóloga 2 autosómica del retardo mental X frágil (FXR2) codificada por el gen FXR2, por ejemplo.

Ejemplos de proteínas asociadas con la degeneración macular puede incluir la proteína de la casete de unión a ATP, sub-familia A (ABC1) miembro 4 (ABCA4) codificada por el gen ABCR, la proteína apolipoproteína E (APOE) codificada por el gen APOE o la proteína quimioquina (motivo C-C) Ligando 2 (CCL2) codificada por el gen CCL2, por ejemplo.

Ejemplos de proteínas asociadas con la esquizofrenia pueden incluir NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, y combinaciones de las mismas.

Ejemplos de proteínas implicadas en la supresión de tumores pueden incluir ATM (ataxia telangiectasia mutada), ATR (ataxia telangiectasia y relacionada con Rad3), EGFR (receptor del factor de crecimiento epidérmico), ERBB2 (homólogo 2 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB3 (homólogo 3 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB4 (homólogo 4 del oncogén de la leucemia eritroblástica v-erb-b2), Notch 1, Notch2, Notch 3 o Notch 4, por ejemplo.

Ejemplos de proteínas asociadas con un trastorno secretasa pueden incluir PSENEN (homólogo de potenciador de presenilina 2 (C. elegans)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (precursor de la proteína amiloide beta (A4)), APH1B (homólogo B defectuoso de la faringe anterior 1 (C. elegans)), PSEN2 (presenilina 2 (enfermedad de Alzheimer 4)) o BACE1 (enzima de escisión de APP del sitio beta 1), por ejemplo.

Ejemplos de proteínas asociadas con la esclerosis lateral amiotrófica pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.

Ejemplos de proteínas asociadas con las enfermedades priónicas pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.

Ejemplos de proteínas relacionadas con enfermedades neurodegenerativas en los trastornos por priones pueden incluir A2M (alfa-2-macroglobulina), AATF (factor de transcripción antagonizante de la apoptosis), ACPP (fosfatasa ácida prostática), ACTA2 (actina alfa 2 aorta músculo liso), ADAM22 (dominio de metalopeptidasa ADAM), ADORA3 (receptor de adenosina A3) o ADRA1D (receptor adrenérgico alfa-1D para el adrenorreceptor alfa-1D), por ejemplo.

Ejemplos de proteínas asociadas con inmunodeficiencia puede incluir A2M [alfa-2- macroglobulina]; AANAT [arilalquilamina N-acetiltransferasa]; ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1]; ABCA2 [casete de unión a ATP, subfamilia A (ABC1), miembro 2]; o ABCA3 [casete de unión a ATP, subfamilia A (ABC1), miembro 3]; por ejemplo.

Ejemplos de proteínas asociadas con Trastornos por Repetición de Trinucleótidos incluyen AR (receptor de andrógenos), FMR1 (retardo mental X frágil 1), HTT (huntingtina) o DMPK (distrofia miotónica-proteína quinasa), FXN (frataxina), ATXN2 (ataxina 2), por ejemplo.

Ejemplos de proteínas asociadas con los Trastornos de Neurotransmisión incluyen SST (somatostatina), NOS1 (óxido nítrico sintasa 1 (neuronal)), ADRA2A (receptor adrenérgico alfa -2A), ADRA2C (receptor adrenérgico alfa-2C), TACR1 (receptor de taquiquinina 1) o HTR2C (receptor de 5-hidroxitriptamina (serotonina) 2C), por ejemplo.

Ejemplos de secuencias asociadas al neurodesarrollo incluyen A2BP1 [proteína de unión 1 ataxina 2], AADAT [aminoadipato aminotransferasa], AANAT [arilalquilamina N-acetiltransferasa], ABAT [4-aminobutirato aminotransferasa], ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1] o ABCA13 [casete de unión a ATP, subfamilia A (ABC1), miembro 13], por ejemplo.

Ejemplos adicionales de afecciones preferidas tratables con el presente sistema se pueden seleccionar de: Síndrome de Aicardi-Goutieres; Enfermedad de Alexander; Síndrome de Allan-Herndon-Dudley; Trastornos Relacionados con POLG; Alfa-Manosidosis (Tipos I y III); Síndrome de Alström; Síndrome de Angelman; Ataxia-Telangiectasia; Lipofuscinosis Ceroide Neuronal; Beta-Talasemia; Atrofia Óptica Bilateral y Atrofia Óptica (Infantil) Tipo 1; Retinoblastoma (bilateral); Enfermedad de Canavan; Síndrome Cerebro-oculofacioesquelético 1 [COFS1]; Xantomatosis Cerebrotendinosa: Síndrome de Cornelia de Lange; Trastornos Relacionados con MAPT; Enfermedades Priónicas Genéticas; Síndrome de Dravet; Enfermedad de Alzheimer Familiar de Aparición Temprana; Ataxia de Friedreich [FRDA]; Síndrome de Fryns; Fucosidosis; Distrofia Muscular Congénita de Fukuyama; Galactosialidosis; Enfermedad de Gaucher; Acidemias Orgánicas; Linfohistiocitosis Hemofagocítica; Síndrome de Progeria de Hutchinson-Gilford; Mucolipidosis II; Enfermedad de Almacenamiento de Ácido Siálico Libre Infantil; Neurodegeneración Asociada a PLA2G6; Síndrome de Jervell y Lange-Nielsen; Epidermolisis Bullosa de la Unión; Enfermedad de Huntington; Enfermedad de Krabbe (Infantil); Síndrome de Leigh y NARP asociado a ADN mitocondrial; Síndrome de Lesch-Nyhan; Lisencefalia Asociada a LIS1; Síndrome de Lowe; Enfermedad de la Orina de Jarabe de Arce; Síndrome de Duplicación de MECP2; Trastornos de Transporte de Cobre Relacionados con ATP7A; Distrofia Muscular Relacionada con LAMA2; Deficiencia de Arilsulfatasa A; Mucopolisacaridosis Tipos I, II o III; Trastornos de la Biogénesis de Peroxisomas, Espectro del Síndrome de Zellweger; Neurodegeneración con Trastornos de Acumulación de Hierro en el Cerebro; Deficiencia de Esfingomielinasa Ácida; Enfermedad de Niemann-Pick Tipo C; Encefalopatía de Glicina; Trastornos Relacionados con ARX; Trastornos del Ciclo de la Urea; Osteogénesis Imperfecta Relacionada con COL1A1/2; Síndromes Mitocondriales de Deleción de ADN; Trastornos relacionados con PLP1; Síndrome de Perry; Síndrome de Phelan-McDermid; Enfermedad del Almacenamiento de Glucógeno Tipo II (Enfermedad de Pompe) (Infantil); Trastornos Relacionados con MAPT; Trastornos Relacionados con MECP2; Condrodisplasia Punctata Rizomélica Tipo 1; Síndrome de Roberts; Enfermedad de Sandhoff; Enfermedad de Schindler - Tipo 1; Deficiencia de Adenosina Desaminasa; Síndrome de Smith-Lemli-Opitz; Atrofia Muscular Espinal; Ataxia Espinocerebelosa de Brote Infantil; Deficiencia de Hexosaminidasa A; Displasia Tanatofórica Tipo 1; Trastornos Relacionados con Colágeno Tipo VI; Síndrome de Usher Tipo I; Distrofia Muscular Congénita; Síndrome de Wolf-Hirschhorn; Deficiencia Lisosomal de Lipasa Ácida; y Xenoderma Pigmentosum.

Como resultará evidente, se prevé que el presente sistema se pueda utilizar para fijar como objetivo a cualquier secuencia de polinucleótidos de interés. Algunos ejemplos de afecciones o enfermedades que podrían ser tratadas útilmente utilizando el presente sistema se incluyen en las Tablas anteriores y también se proporcionan allí ejemplos de genes asociados actualmente con esas afecciones. Sin embargo, los genes ejemplificados no son exhaustivos.

EJEMPLOS

Los siguientes ejemplos se dan con el propósito de ilustrar diversas realizaciones de la invención y no pretenden limitar la presente invención de modo alguno. Los presentes ejemplos, junto con los métodos descritos en esta memoria son actualmente representativos de realizaciones preferidas, son ilustrativos y no pretenden ser limitaciones del alcance de la invención.

Ejemplo 1: Mejora del sistema Cas9 para la aplicación in vivo

Las solicitantes llevarón a cabo una búsqueda metagenómica para una Cas9 con pequeño peso molecular. La mayoría de los homólogos de Cas9 son bastante grandes. Muchos homólogos de Cas9 conocidos son grandes y contienen más de 1300 amino ácidos. Por ejemplo, la SpCas9 es de alrededor de 1368aa de larga, que es demasiado grande para ser empaquetada fácilmente en vectores virales para el suministro. En la Figura 6 se muestra un gráfico que representa la distribución de longitud de homólogos de Cas9: El gráfico se genera a partir de secuencias depositadas en GenBank. Algunas de las secuencias pueden haber sido anotadas erróneamente y, por lo tanto, la frecuencia exacta para cada uno de los tramos no tiene que ser necesariamente fiable. No obstante, proporciona un esbozo de la distribución de proteínas Cas9 y sugiere que existen homólogos de Cas9 más cortos.

A través de análisis computacional, la solicitante encontró que en la cepa bacteriana *Campylobacter* existen dos proteínas Cas9 con menos de 1000 aminoácidos. La secuencia para una Cas9 de *Campylobacter jejuni* se presenta a continuación. En esta longitud, CjCas9 puede empaquetarse fácilmente en AAV, lentivirus, adenovirus y otros

vectores virales para el suministro sólido en células primarias y en modelos animales *in vivo*. En una realización preferida de la invención, se utiliza la proteína Cas9 de *S. aureus*.

Cas9 de Campylobacter jejuni (CjCas9)

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARR
KARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFAR
VILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSK
EFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFS
HLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLK
NGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDIT
LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNE
LNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIE
LAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCA
YSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDS
AKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYL
DFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVI
IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIF
VSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFR
VDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYK
DSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKS
IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK.

El elemento tracrRNA putativo para esta CjCas9 es:

TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGGG
ACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT

La secuencia de Repetición Directa es:

GTTTTAGTCCCTTTTTAAATTTCTTTATGGTAAAAT

Un ejemplo de un ARN quía quimérico para CjCas9 es:

NNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUU
UGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUU

Ejemplo 2: Diversidad de Cas9 y ARNs quiméricos

El sistema de CRISPR-Cas de tipo salvaje es un mecanismo inmune adaptativo contra la invasión de ADN exógeno empleado por diversas especies a través de bacterias y arqueas. El sistema de CRISPR-Cas de tipo II consiste en un conjunto de genes que codifican proteínas responsables de la "adquisición" de ADN extraño en el locus de CRISPR, así como un conjunto de genes que codifican la "ejecución" del mecanismo de escisión de ADN; éstos incluyen la ADN nucleasa (Cas9), un cr-RNA transactivador no codificador (tracrRNA) y una serie de espaciadores derivados de ADN extraños flanqueados por repeticiones directas (crRNAs). Tras la maduración por Cas9, el dúplex tracRNA y crRNA guía la nucleasa Cas9 a una secuencia de ADN diana especificada por las secuencias guía del espaciador, y media en las roturas de la doble cadena en el ADN cerca de un corto motivo de secuencia en el ADN diana que se requiere para la escisión y que es específico para cada uno de los sistemas CRISPR-Cas. Los sistemas de CRISPR-Cas de tipo II se encuentran en todo el reino bacteriano y son muy diversos en la secuencia y el tamaño de proteínas Cas9, y la secuencia de repetición directa de tracrRNA y crRNA, la organización del genoma de estos elementos y el requisito del motivo para la escisión diana. Una especie puede tener múltiples sistemas de CRISPR-Cas distintos.

La solicitante evaluó 207 Cas9s putativos de especies bacterianas identificadas sobre la base de homología de secuencia con Cas9s y estructuras ortólogas conocidas a subdominios conocidos, incluyendo el dominio HNH de endonucleasa y los dominios RuvC endonucleasa [información de Eugene Koonin y Kira Makarova]. El análisis filogenético basado en la conservación de secuencias de proteínas de este conjunto reveló cinco familias de Cas9s,

incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos Cas9s pequeños (~ 1100 aminoácidos) (Figuras 4 y 5A-F).

En algunas realizaciones, las secuencias de emparejamiento tracr o las repeticiones directas se descargan ya sea desde la base de datos CRISPRs o se identifican in silico mediante el rastreo de motivos repetitivos que 1. se encuentran en una ventana de 2 kb de secuencia genómica flanqueante del locus de CRISPR de tipo II, 2. abarcan de 20 a 50 pb, y 3. están separados por 20 a 50 pb. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo, 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones tracrRNA candidatas son predichas posteriormente predichas por 1. homología de la secuencia con repeticiones directas (rastreo de motivos en Geneious con hasta apareamientos erróneos de hasta 18 pb), 2. presencia de un terminador transcripcional independiente de Rho predicho en la dirección de la transcripción y 3. la estructura secundaria de horquilla estable entre tracrRNA y la repetición directa. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones, diseños de ARNs quiméricos guía sintéticos (sgRNAs) incorporan al menos 12 pb de la estructura dúplex entre la repetición directa y tracrRNA.

La lista de las secuencias de ortólogos Cas9 humanas, optimizadas en codones, a emparejarse con los ARNs quiméricos proporcionados en las Figuras 8 A-J se proporciona en las Figuras 9 A-O. La solicitante también ha demostrado que los ortólogos Cas9 pueden escindir sus dianas en ensayos de escisión in vitro (Figura 16). Los loci de CRISPR en algunas de estas familias se representa en la Figura 11. Las secuencias de ARN guía correspondientes se muestran en la Figura 12. La solicitante analizó sistemáticamente la secuencia de ADN genómico dentro de ~ 2 kb de las proteínas Cas9 utilizando el código de análisis computacional habitual e identificó repeticiones directas que oscilaban entre 35 pb y 50 pb, con espaciadores intermedios que oscilaban entre 29 pb y 35 pb. Sobre la base de la secuencia de repetición directa, la solicitante rastreó de forma computacional secuencias candidatas tracrRNA con los siguientes criterios: fuera de la matriz de crRNA, pero que contienen un alto grado de homología con repeticiones directas (tal como se requiere para el apareamiento de bases repetición directa:tracRNA; análisis computacional habitual), fuera de las regiones codificadoras de los componentes de proteínas, que contienen señales de terminación de la transcripción Rho-independientes, ~ 60 pb-120 pb aguas abajo de la región de homología de repeticiones directas, y co-plegamiento con repeticiones directas para formar un dúplex, seguido por dos o más estructuras de horquilla en el extremo distal de la secuencia de tracrRNA. Sobre la base de estos criterios de predicción, la solicitante seleccionó un conjunto inicial de 18 proteínas Cas9 y sus repeticiones directas asociadas de forma única y tracrRNAs distribuidos en todas las cinco familias Cas9. La solicitante generó, además, un conjunto de 18 estructuras de ARN quimérico que conservaron la secuencia y estructuras secundarias del dúplex repetición directa nativa:tracrRNA al tiempo que acortaron la región de apareamiento de bases y fusionaron los dos elementos de ARN a través de un bucle artificial (Figuras 8A-J).

Ejemplo 3: Ortólogos Cas9

La solicitante ha generado ortólogos Cas9 optimizados en codones para progresar en la expresión en células eucariotas.

Un ejemplo de una secuencia humana optimizada en codones (es decir, optimizada para la expresión en seres humanos.) : SaCas9: se proporciona a continuación

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAA
GGTCGAAGCGTCCATGAAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTG
GGGTATGGGATTATTGACTATGAAACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCA
AGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGA
AACGACGGAGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCT
GACCGACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAAGGCCTGAGTC
AGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGT
GCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATC
TCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGCTGA
AGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGCGACTACGTCAAAGA
AGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGAT
ACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCC
CCTTCGGATGGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTT
CCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCTGAATG
ACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTT
CCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAG
GAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAAAACCAG
AGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAAATCAT
TGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGG
ACATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGAT
TAGTAATCTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTC
TGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTGGTC
CCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTCA
TTCTGTCACCCGTGGTCAAGCGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATC
AAGAAGTACGGCCTGCCCAATGATATCATTATCGAGCTGGCTAGGGAGAAGAACAGCAAGG
ACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTG
AAGAGATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCT
GCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGC
TGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAAT
TCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTC
CTTTCCAGTACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTC
TGAATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAG

AGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTGGACACA
AGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGA
TGTGAAAGTCAAGTCCATCAACGGCGGGTTCACATCTTTTCTGAGGCGCAAATGGAAGTTTA
AAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTATCGCAAATGC
CGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAG
ATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGAGACAGAACAGGAGTACAAGG
AGATTTTCATCACTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCT
CACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAA
AAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAA
TGACAAGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGAT
CCTCAGACATATCAGAAACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCCAC
TGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAGGATAATGG
CCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACA
GACGATTACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTCG
ATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAA
AAGGAGAACTACTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAG
ATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGG
CGAACTGTATAGGGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATG
ATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAGCGCCCCCCTCGAATTAT
CAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAAC
CTGTATGAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATTC

La solicitante analizó ortólogos Cas9 para identificar las secuencias de PAM relevantes y el correspondiente ARN guía quimérico tal como se indica en la Figura 13A-II. Este conjunto expandido de PAMs proporciona una fijación como objetivo más amplia en todo el genoma y también aumenta significativamente el número de sitios diana únicos y proporciona un potencial para identificar nuevas Cas9s con niveles incrementados de especificidad en el genoma. La solicitante determinó que el PAM para subespecies de Staphylococcus aureus subespecie Aureus Cas9 era NNGRR (Figura 14). Staphylococcus aureus subespecie Aureus Cas9 también se conoce como SaCas9. Las Figuras 15 a-d proporcionan diseños únicos o múltiples de vectores para SaCas9.

La Figura 7 muestra logotipos de secuencia para PAMs putativos tal como se indica mediante complementos inversos. Ortólogos Cas9 y sus respectivos sgRNAs se utilizaron para escindir un banco de dianas que portan un PAM aleatorio (secuencia de 7 pb inmediatamente 3' de la secuencia diana). Los productos escindidos se aislaron y secuenciaron profundamente para proporcionar secuencias candidatas de 7 pb que eran permisivas a la escisión para cada uno de los ortólogos Cas9. Para cada uno de los ortólogos Cas9 se determinaron PAMs consenso alineando todas las secuencias candidatas de 7 pb (Figuras 7 y 21).

El trabajo adicional de examinar la termodinámica y la estabilidad *in vivo* de dúplex de sgRNA y ADN proporcionará probablemente un poder predictivo adicional para actividades fuera de objetivo, mientras que la exploración de mutantes y ortólogos SpCas9 puede proporcionar nuevas variantes con una especificidad mejorada. La especificidad de ortólogos Cas9 puede evaluarse adicionalmente, examinando la capacidad de cada una de las Cas9 de tolerar apareamiento erróneos entre el ARN guía y su diana de ADN.

Ejemplo 4: Mutaciones Cas9

En este ejemplo, la solicitante demuestra que las siguientes mutaciones pueden convertir SpCas9 en una enzima que produce mella: D10A, E762A, H840A, N854A, N863A, D986A.

La solicitante proporciona secuencias que muestran donde se encuentran los puntos de mutación en el gen SpCas9 (Figuras 10A-M). La solicitante también muestra que los nickases todavía son capaces de mediar en la recombinación homóloga. Además de ello, SpCas9 con estas mutaciones reduce (individualmente) el nivel de ruptura de la doble cadena. Ortólogos Cas9 comparten todos la organización general de los dominios 3-4 RuvC y un dominio HNH (Figura 19). El dominio RuvC más próximo a 5' escinde la cadena no complementaria y el dominio HNH escinde la cadena complementaria. Todas las notaciones son en referencia a la secuencia guía.

El residuo catalítico en el dominio 5' RuvC se identifica a través de la comparación de homología de la Cas9 de interés con otros ortólogos Cas9 (de locus de CRISPR tipo 11 de S. pyogenes, locus 1 de CRISPR de S. thermophilus, locus 3 de CRISPR de S. thermophilus y el locus de CRISPR tipo II de Franciscilla novicida), y el residuo Asp conservado está mutado a alanina para convertir Cas9 en una enzima que produce mella de la cadena complementaria. DE manera similar, los residuos His y Asn conservados en los dominios HNH están mutados a alanina para convertir Cas9 en una enzima que produce mella de cadena no complementaria.

Ejemplo 5: Optimización funcional de Cas9

Para la función mejorada o para desarrollar nuevas funciones, la solicitante genera proteínas quiméricas Cas9 mediante la combinación de fragmentos de diferentes ortólogos Cas9.

Por ejemplo, la solicitante fusiona el extremo N de St1Cas9 (el fragmento de esta proteína está en negrita) con extremo C de SpCas9 (el fragmento de esta proteína está subrayado).

> Sti(N)Sp(C)Cas9

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGR
RLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELF
IALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER
YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDE
FINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP
DEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAM
GPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQ
MDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANS
SIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYID
EKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY
LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

> Sp(N)St1(C)Cas9

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARETNEDD
EKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRL
WHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATA
NQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISK
FDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLR
RHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIET
GELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIY
ATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHD
PQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGP
EIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILG
LKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKD
TETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSG
QCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF

La solicitantes también ha generado proteínas quiméricas Sp_St3 y ha demostrado la escisión *in vitro* por SpCas9, St3Cas9, Sp_St3 quimera y St3_Sp quimera (Figura 17).

El beneficio de hacer Cas9 quimérica incluye:

a. reducir la toxicidad

b. mejorar la expresión en células eucariotas

c. mejorar la especificidad

d. reducir el peso molecular de la proteína, hacer la proteína más pequeña que mediante la combinación de los dominios más pequeños de diferentes homólogos Cas9.

e. Alterar el requisito de la secuencia de PAM

Ejemplo 6: Suministro de Cas9 in vivo utilizando partículas o vectores AAV

Suministro *in vivo* – método AAV

AAV es ventajoso frente a otros vectores virales por un par de razones:

- Baja toxicidad (esto puede ser debido al método de purificación que no requiere ultracentrifugación de partículas celulares que pueden activar la respuesta inmune)
- Baja probabilidad de provocar una mutagénesis insercional porque no se integra en el genoma del huésped.

Mientras que determinados vectores AAV actuales pueden alojar hasta 4300 bases de ADN insertado, como un límite superior o un límite de empaquetamiento, el AAV puede tener ADN insertado de 4,5 o 4,75 KB. Esto significa que el ADN que codifica una enzima Cas9 así como un promotor y terminador de la transcripción tiene que encajar

todos en el mismo vector viral. Construcciones mayores que 4,5 o 4,75 KB conducirá a una producción de virus significativamente reducida. SpCas9 es bastante grande, el gen por sí es de más de 4,1 kb, lo que hace que sea difícil de empaquetar en un AAV. Por lo tanto, realizaciones de la invención incluyen utilizar ortólogos de Cas9 que sean más cortos. Por ejemplo:

Especie	Tamaño de Cas9
Corynebacter diphtheria	3252
Fubacterium ventriosum	3321
Streptococcus pasteurianus	3390
Lactobacillus farciminis	3378
Sphaerochaeta globus	3537
Azospirillum B510	3504
Gluconacetobacter diazotrophicus	3150
Neisseria cinerea	3246
Roseburia intestinalis	3420
Parvibaculum lavamentivorans	3111
Staphylococcus aureus	3159
Nitratifractor salsuginis DSM 16511	3396
Campylobacter lari CF89-12	3009
Streptococcus thermophilus LMD-9	3396

La Figura 3 proporciona representaciones esquemáticas de vectores de AAV que se pueden utilizar en los métodos y composiciones de la invención. El empaquetamiento se comentó anteriormente.

Ejemplo 7: Modificación por Ingeniería Genética de Microalgas utilizando Cas9

Métodos de suministrar Cas9

Método 1: La solicitante suministra Cas9 y ARN guía utilizando un vector que expresa Cas9 bajo el control de un promotor constitutivo tal como Hsp70A-Rbc S2 o beta2-tubulina.

Método 2: La solicitante proporciona Cas9 y T7 polimerasa utilizando vectores que expresan Cas9 y T7 polimerasa bajo el control de un promotor constitutivo tal como Hsp70A-Rbc S2 o beta2-tubulina. ARN guía será suministrada utilizando un promotor T7 que contiene vector que impulsa el ARN guía.

Método 3: La solicitante suministra ARNm de Cas9 y ARN guía transcrito in vitro a las células de algas. ARN puede transcribirse in vitro. ARNm de Cas9 consistirá en la región codificadora de Cas9 así como 3' UTR de Copl para asegurar la estabilización del ARNm de Cas9.

Para la recombinación homóloga, la solicitante proporciona un molde de reparación dirigido a una homología adicional.

Secuencia para una casete que impulsa la expresión de Cas9 bajo el control del promotor de beta2-tubulina, seguido por la 3' UTR de Cop1.

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTC
CCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTG
CATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCG
ATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTAC
CACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCGTTTCAGTCACAACCCGCAAACATGTACCCATACGATGTTCCAGATTACG

CTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCGGC
CTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGT
GCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAG
AACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCT
GAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTG
CAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACT
GGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTG
AGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGC
CCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCC
CGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGC
TGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCT
GCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGA
GAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCA
ACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGAC
ACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGA
CCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAG
AGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACG
ACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCT
GAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACAT
TGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAA
AGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCT
GCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGG
AAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCC
CTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGC
GGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGC
CTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACC
GAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGG
ACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTAC

TTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC
AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTT
CCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACAC
TGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTC
GACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC
TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGAC
GACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGA
TAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCA
TCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAG
CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGAC
AGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
CAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAG
CTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGA
CATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAA
GGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAG
AGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCA
GCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCG
AGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACAC
TAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGT
CCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCA
ACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTA
CGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCC
AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGA
TCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCC
CAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC
CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG

CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAG
CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCC
CTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCA
GCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTT
GTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTC
CAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGG
AAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCAT
CACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCA
AGAAGAAGAGAAAGGTGGAGGCCAGCTAAGGATCCGGCAAGACTGGCCCCGCTTG
GCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTG
TGTTGGCCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAG
AATGTAACGTCAGTTGATGGTACT

accggaatggttagcttacacegccaaaatgctggcgtagtaggtcaagactctgagactatcgaactcgcacctgaatacgctgaggctat
cgcaaccegtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgtagttcctcctaagccgtggactggcattactggtggtgg
ctattgggctaacggtcgtcgtcctctggcgctggtgcgtactcacagtaagaaagcactgatgcgctacgaagacgtttacatgcctgaggt
gtacaaagcgattaacattgcgcaaaacaccgcatggaaaatcaacaagaaagtcctagcggtcgccaacgtaatcaccaagtggaagca
ttgtccggtcgaggacatccctgcgattgagcgtgaagaactcccgatgaaaccggaagacatcgacatgaatcctgaggctctcaccgcg
tggaaacgtgctgccgctgctgtgtaccgcaaggacaaggctcgcaagtctcgccgtatcagccttgagttcatgcttgagcaagccaataa
gtttgctaaccataaggccatctggttcccttacaacatggactggcgcggtcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgat
atgaccaaaggactgcttacgctggcgaaaggtaaaccaatcggtaaggaaggttactactggctgaaaatccacggtgcaaactgtgcg
ggtgtcgacaaggttccgttcccgagcgcatcaagttcattgaggaaaaccacgagaacatcatggcttgcgctaagtctccactggagaa
cacttggtgggctgagcaagattctccgttctgcttccttgcgttctgctttgagtacgctggggtacagcaccacggcctgagctataactgct
ccctccgctggcgtttgacgggtcttgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtggtcgcgcggttaacttgc
ttcctagtgaaaccgttcaggacatctacgggattgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatgggaccgataac
gaagtagttaccgtgaccgatgagaacactggtgaaatctctgagaaagtcaagctgggcactaaggcactggctggtcaatggctggctt
acggtgttactcgcagtgtgactaagcgttcagtcatgacgctggcttacgggtccaaagagttcggcttccgtcaacaagtgctggaagata
ccattcagccagctattgattccggcaagggtctgatgttcactcagccgaatcaggctgctggatacatggctaagctgatttgggaatctgt
gagcgtgacggtggtagctgcggttgaagcaatgaactggcttaagtctgctgctaagctgctggctgctgaggtcaaagataagaagact
ggagagattcttcgcaagcgttgcgctgtgcattgggtaactcctgatggtttccctgtgtggcaggaatacaagaagcctattcagacgcgc
ttgaacctgatgttcctcggtcagttccgcttacagcctaccattaacaccaacaaagatagcgagattgatgcacacaaacaggagtctggt
atcgctcctaactttgtacacagccaagacggtagccaccttcgtaagactgtagtgtgggcacacgagaagtacggaatcgaatcttttgca
ctgattcacgactccttcggtacgattccggctgacgctgcgaacctgttcaaagcagtgcgcgaaactatggttgacacatatgagtcttgtg
atgtactggctgatttctacgaccagttcgctgaccagttgcacgagtctcaattggacaaaatgccagcacttccggctaaaggtaacttgaa
cctccgtgacatcttagagtcggacttcgcgttcgcgtaaGGATCCGGCAAGACTGGCCCCGCTTGGCAACG
CAACAGTGAGCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGG
CCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTA
ACGTCAGTTGATGGTACT

Secuencia de un casete que impulsa la expresión de T7 polimerasa bajo el control del promotor beta-2 tubulina, seguido por la 3' UTR de Cop1:

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTC
CCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTG
CATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCG
ATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTAC
CACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCGTTTCAGTCACAACCCGCAAACatgcctaagaagaagaggaaggttaacacgattaacatcgctaag
aacgacttctctgacatcgaactggctgctatcccgttcaacactctggctgaccattacggtgagcgtttagctcgcgaacagttggccctg
agcatgagtcttacgagatgggtgaagcacgcttccgcaagatgtttgagcgtcaacttaaagctggtgaggttgcggataacgctgccgcc
aagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactggtttgaggaagtgaaagctaagcgcggcaagcgcccga
cagccttccagttcctgcaagaaatcaagccggaagccgtagcgtacatcaccattaagaccactctggcttgcctaaccagtgctgacaat
acaaccgttcaggctgtagcaagcgcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgtgaccttgaagctaagcact
tcaagaaaaacgttgaggaacaactcaacaagcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctgacatgctctct
aagggtctactcggtggcgaggcgtggtcttcgtggcataaggaagactctattcatgtaggagtacgctgcatcgagatgctcattgagtca

Secuencia de ARN guía impulsado por el promotor T7 (promotor T7, Ns representa secuencia de fijación de objetivo):

gaaat**TAATACGACTCACTATA**NNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaa

gttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttttt

Suministro del gen:

La cepa Chlamydomonas reinhardtii CC-124 y CC-125 del Chlamydomonas Resource Center se utilizará para la electroporación. El protocolo de electroporación sigue el protocolo recomendado estándar del kit GeneArt Chlamydomonas Engineering (información de sitio web en tools.invitrogen.com/content/sfs/manuals/geneart_chlamy_kits_man.pdf).

También, la solicitantes generó una línea de Chlamydomonas reinhardtii que expresa Cas9 de forma constitutiva. Esto se puede hacer utilizando pChlamyl (linearizado utilizando PvuI) y seleccionando las colonias resistentes a higromicina. La secuencia para pChlamyl que contiene Cas9 figura debajo. De este modo, para lograr una inactivación de genes simplemente se tiene que suministrar ARN para el ARNguía. Para la recombinación homóloga, la soilicitante suministra ARNguía, así como un molde de recombinación homóloga linearizado.

pChlamyl -Cas9:

TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATC
TCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAA
CTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC
CCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGA
GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCG
GGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGC
TACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCC
TTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG
CTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG
AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC

CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT
CTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGT
TTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG
CAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAC
TCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGTTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTT
TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCT
ATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTT
GCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCT
TTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG
AGCGAGGAAGCGGTCGCTGAGGCTTGACATGATTGGTGCGTATGTTTGTATGAAGCT
ACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGAAGCTCTGGAAGGGCCGCGATG
GGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGGCACCCATCCGGTAT
AAAAGCCCGCGACCCCGAACGGTGACCTCCACTTTCAGCGACAAACGAGCACTTAT
ACATACGCGACTATTCTGCCGCTATACATAACCACTCAGCTAGCTTAAGATCCCATC
AAGCTTGCATGCCGGGCGCGCCAGAAGGAGCGCAGCCAAACCAGGATGATGTTTGA
TGGGGTATTTGAGCACTTGCAACCCTTATCCGGAAGCCCCCTGGCCCACAAAGGCTA
GGCGCCAATGCAAGCAGTTCGCATGCAGCCCCTGGAGCGGTGCCCTCCTGATAAAC
CGGCCAGGGGGCCTATGTTCTTTACTTTTTTACAAGAGAAGTCACTCAACATCTTAA
AATGGCCAGGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCGTGCTTGCAGAT
TTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCCTCTGTCGCTGTCT
CAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGGAGATTCGAGGTACCAT
GTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAG
CGTCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAA
CACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCG
GCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAG

ACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGG
TGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAG
AAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGA
GAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGG
CCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACT
TCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATC
CAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGG
CGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAA
ATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATT
GCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGG
CCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACG
CCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTG
AGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAA
AGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGA
GCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTAC
AAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAA
GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCC
CCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTT
ACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATC
CCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAG
AAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGC
GCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAAC
GAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGA
GCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCG
GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGA
AATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCT
GAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG
AAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGG
CTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCG

GAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACA
AGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGA
AACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAA
AGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCG
GCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTC
GTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGA
GAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATC
GAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAA
ACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGAT
ATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCA
TATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAG
AAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCG
GCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAG
ATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGA
AGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCA
GTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTC
GTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCA
GGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTT
CAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGA
CAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTG
CGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGAC
AGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTG
GCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAA
GAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGA
ATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATC
ATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCT
GGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATG

TGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGAT
AATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCAT
CGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACA
AAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAG
AATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGC
CACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCA
GCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAA
CATATGATTCGAATGTCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGG
CTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAA
GCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAA
TAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAA
CACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCT
AAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACATGACACAAGAAT
CCCTGTTACTTCTCGACCGTATTGATTCGGATGATTCCTACGCGAGCCTGCGGAACG
ACCAGGAATTCTGGGAGGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCGTGC
TTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCCTCTGT
CGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGCCGCTGGCC
CGCCGAGCCCTGGAGGAGCTCGGGCTGCCGGTGCCGCCGGTGCTGCGGGTGCCCGG
CGAGAGCACCAACCCCGTACTGGTCGGCGAGCCCGGCCCGGTGATCAAGCTGTTCG
GCGAGCACTGGTGCGGTCCGGAGAGCCTCGCGTCGGAGTCGGAGGCGTACGCGGTC
CTGGCGGACGCCCCGGTGCCGGTGCCCCGCCTCCTCGGCCGCGGCGAGCTGCGGCC
CGGCACCGGAGCCTGGCCGTGGCCCTACCTGGTGATGAGCCGGATGACCGGCACCA
CCTGGCGGTCCGCGATGGACGGCACGACCGACCGGAACGCGCTGCTCGCCCTGGCC
CGCGAACTCGGCCGGGTGCTCGGCCGGCTGCACAGGGTGCCGCTGACCGGGAACAC
CGTGCTCACCCCCCATTCCGAGGTCTTCCCGGAACTGCTGCGGGAACGCCGCGCGGC
GACCGTCGAGGACCACCGCGGGTGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCC
TGGAGGACTGGCTGCCGGACGTGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTC
GTCCACGGCGACCTGCACGGGACCAACATCTTCGTGGACCTGGCCGCGACCGAGGT
CACCGGGATCGTCGACTTCACCGACGTCTATGCGGGAGACTCCCGCTACAGCCTGGT
GCAACTGCATCTCAACGCCTTCCGGGGCGACCGCGAGATCCTGGCCGCGCTGCTCGA

CGGGGCGCAGTGGAAGCGGACCGAGGACTTCGCCCGCGAACTGCTCGCCTTCACCT
TCCTGCACGACTTCGAGGTGTTCGAGGAGACCCCGCTGGATCTCTCCGGCTTCACCG
ATCCGGAGGAACTGGCGCAGTTCCTCTGGGGGCCGCCGGACACCGCCCCCGGCGCC
TGATAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCC
TAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAA
CAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

Para todas las células de Chlamydomonas reinhardtii modificados, la solicitante utiliza PCR, ensayo de nucleasa SURVEYOR y secuenciación de ADN para verificar la modificación con éxito.

Ejemplo 8: Reconocimiento de SaCas9 y PAM para aplicaciones in vivo

El proyecto comenzó ya que la solicitante quería explorar adicionalmente la diversidad del sistema de CRISPR/Cas de tipo II tras la identificación del sistema de CRISPR/Cas de *Streptococcus pyogenes* (Sp) y *Streptococcus thermophilus* (St) CRISPR/Cas como una herramienta de modificación por ingeniería genética del genoma funcional en células de mamífero.

Al definir nuevos sistemas funcionales de CRISPR/Cas de tipo II para su aplicación en células de mamífero, la solicitante será capaz, en potencia, de encontrar:

(1) Un sistema de CRISPR/Cas con una mayor eficiencia y/o especificidad

(2) Un sistema de CRISPR/Cas con diferente Motivo Adyacente al Protoespaciador (PAM) que permite la fijación como objetivo de una gama más amplia de loci genómicos

(3) Un sistema de CRISPR/Cas con un tamaño más pequeño, de modo que la solicitante podría suministrarlos in vivo en un único vector con el sistema de suministro vírico en mamíferos tales como vectores de virus adeno-asociados (AAV) que tiene un límite de tamaño de empaquetamiento (el actual sistema Sp o St excede este límite de 4,7kb)

y otros rasgos deseables.

Identificación y Diseño del sistema de CRISPR/Cas en Sa para aplicación in vivo

La solicitante sometió a ensayo un nuevo sistema de CRISPR/Cas de tipo II en *Staphylococcus aureus* (Sa) que funciona in vitro en un ensayo de escisión de ADNds e identificó un PAM putativo de NNGRRT. Los componentes de este sistema son una proteína Cas9 de Sa, un ARN guía de CRISPR con repeticiones directas (DR) de Sa que formará un complejo de ARN guía funcional con tracrRNA de Sa. Este sistema de tres componentes es similar a todos los demás sistemas de CRISPR/Cas de tipo II. Por lo tanto, la solicitante diseñó un sistema de dos componentes, en donde la solicitante fusionó el tracrRNA de Sa al ARN de CRISPR guía de Sa a través de un corto bucle de tallo para formar un ARN guía quimérico, exactamente como lo hizo la solicitante con el sistema de CRISPR/Cas de *Streptococcus pyogenes* (Sp). Esta ARN guía quimérico fue capaz de apoyar la escisión de ADNds in vitro. Por lo tanto, la solicitante decidió clonar el sistema completo de dos componentes: cas9 y el ARN guía quimérico, en un vector de AAV para probar su funcionalidad en organismos vivos.

La solicitante eligió el sistema de AAV, ya que es un virus de mamífero no inmunogénico, basado en ADNss y no integrante, que tiene un amplio espectro de tropismo en diferentes tejidos/órganos dependiendo del serotipo, que se ha demostrado que es seguro para la aplicación in vivo y también apoya la expresión a largo plazo del transgén en los organismos vivos.

El diseño del vector de AAV inicial tiene (1) un promotor de CMV que impulsa la proteína SaCas9 con un solo NLS y un marcador de epítopo HA. (2), el promotor U6 humano que impulsa el ARN quimérico. Éstos se colocan entre dos Repeticiones Terminales Invertidas (ITRs) del serotipo 2 de AAV mejor estudiado que sirven como la señal de empaquetamiento viral.

El ensayo de la secuencia de PAM en genoma de mamíferos endógeno es el siguiente: Se seleccionaron espaciadores de la diana SaCas9 a través de múltiples genes para cubrir diferentes secuencias de PAM potenciales. Diferentes espaciadores se clonaron en la casete de ADNds de expresión de U6-sgRNA (ARN guía sencillo). Los espaciadores en la casete de ADNds de expresión fueron co-transfectados en líneas de células de mamíferos (293FT para dianas humanas, N2a y Hepa para dianas de ratón). 72 horas después de la transfección, se extrajo todo el ADN genómico y se sometió al ensayo de nucleasa surveyor. Se hizo pasar a través de Gel TBEPage para detectar la escisión genómica. Se cuantifica la eficacia de escisión de ADN genómico y se representa.

Sumarios de la Eficiencia de Escisión del Genoma y otras Estadísticas en Todas las Dianas Testadas

Secuencias SaCas9 PAM	Recuentos Dianas	Recuentos Dianas Cambiadas	Porcentaje Dianas Escindidas (%)	Eficiencia de Escisión Acumulativa (%)	Contenido Medio GC de Espaciador (%)
GAAA	1	1	100.0	5.4	65.0
GAAC	2	2	100.0	6.1	55.0
GAAG	8	8	100.0	47.1	65.0
GAAT	9	8	88.9	138.4	66.1
GAGA	3	3	100.0	17.5	63.3
GAGC	8	8	100.0	44.2	60.0
GAGG	12	12	100.0	93.3	58.3
GAGT	44	20	45.5	434.0	56.9
GGAA	2	2	100.0	4.7	60.0
GGAC	3	2	66.7	39.9	60.0
GGAG	12	9	75.0	36.9	59.6
GGAT	20	10	50.0	166.2	59.0
GGGA	7	5	71.4	39.1	63.6
GGGC	11	9	81.8	70.3	65.5
GGGG	8	5	62.5	53.3	70.0
GGGT	45	19	40.0	402.3	58.2
Total General	156	120	61.2	1618.6	59.4

Sumario de la Eficiencia de Escisión del Genoma y otras Estadísticas en Todas las Dianas Testadas

Secuenias de Sa Cas9 PAM	Recuento de Dianas	Recuento de Dianas Cambiadas	Porcentaje de Dianas Escindidas (%)	Eficiencia de Escisión Acumulativa (%)	Eficiencia de Escisión Media (%)	Contenido GC de Espaciador Medio (%)
GAAA	1	1	100.0	5.4	5.4	65.0
GAAC	2	2	100.0	6.1	3.0	55.0
GAAG	8	8	100.0	47.1	5.9	65.0
GAAT	4	4	100.0	68.4	17.1	65.0
GAGA	2	2	100.0	12.5	6.3	67.5
GAGC	5	5	100.0	39.2	7.8	61.0
GAGG	11	11	100.0	88.3	8.0	58.2
GAGT	13	10	76.9	199.0	15.3	58.2
GGAA	2	2	100.0	4.7	2.3	60.0
GGAC	3	2	66.7	39.9	13.3	60.0
GGAG	12	9	75.0	36.9	3.1	59.6
GGAT	13	9	69.2	161.2	12.4	59.8
GGGA	7	5	71.4	39.1	5.6	63.6
GGGC	11	9	81.8	70.3	6.4	65.5
GGGG	8	5	62.5	53.3	6.7	70.0
GGGT	14	8	57.1	182.3	13.0	54.6
Total General	116	92	79.3	1053.6	9.1	60.5

Los resultados del ensayo de PAM se muestran en las Figs. 22-27. Un ensayo completo de más de 100 dianas identificó que el PAM para SaCas9 podría describirse como NNGRR (pero no NNGRRT como se indicó anteriormente).

Sumario del Ensayo de PAM: (1) NNGRR para SaCas9 PAM general – de ayuda para el diseño de nuevas dianas, (2) Testado de doble nickase con nuevas dianas, (3) NNGRG puede ser un PAM más potente.

Referencias:

1. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013)

2. Mali, P. et al. RNA-Guided Human Genome Engineering via Cas9. Science 339, 823-826 (2013).

3. Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013).

4. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31, 230-232 (2013).

5. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471,602-607 (2011).

6. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).

7. Wang, H. et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell 153, 910-918 (2013).

8. Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649,247-256 (2010).

9. Bogenhagen, D.F. y Brown, D.D. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell 24, 261-270 (1981).

10. Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

11. Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res 40, 5368-5377 (2012).

12. Valton, J. et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem 287, 38427-38432 (2012).

13. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761 (2010).

14. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29, 143-148 (2011).

15. Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids research 39,9283-9293 (2011).

16. Hsu, P.D. y Zhang, F. Dissecting neural function using targeted genome engineering technologies. ACS chemical neuroscience 3, 603-610 (2012).

17. Sanjana, N.E. et al. A transcription activator-like effector toolbox for genome engineering. Nature protocols 7, 171-192 (2012).

18. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. Science 300, 763 (2003).

19. Miller, J.C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25, 778-785 (2007).

20. Sander, J.D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8, 67-69 (2011).

21. Wood, A.J. et al. Targeted genome editing across species using ZFNs and TALENs. Science 333, 307 (2011).

22. Bobis-Wozowicz, S., Osiak, A., Rahman, S.H. y Cathomen, T. Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. Methods 53, 339-346 (2011).

23. Jiang, W., Bikard, D., Cox, D., Zhang, F. y Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233-239 (2013).

24. Qi, L.S. et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152, 1173-1183 (2013).

25. Michaelis, L.M., Maud "Die kinetik der invertinwirkung.". Biochem. z (1913).

26. Mahfouz, M.M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA. binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108, 2623-2628 (2011).

27. Wilson, E.B. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22, 209-212 (1927).

28. Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12, 238-251 (2013).

29. Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318-331 (2011).

30. Carlson, D.F. et al. Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A 109, 17382-17387 (2012).

31. Geurts, A.M. et al. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. Science 325, 433-433 (2009).

32. Takasu, Y. et al. Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. Insect Biochem Molec 40, 759-765 (2010).

33. Watanabe, T. et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. Nat Commun 3 (2012).

34. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012).

35. Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509-1512 (2009).

36. Moscou, M.J. y Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. Science 326,1501 (2009).

37. Deveau, H., Garneau, J.E. y Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64,475-493 (2010).

38. Horvath, P. y Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167-170 (2010).

39. Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9, 467-477 (2011).

40. Bhaya, D., Davison, M. y Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45, 273-297 (2011).

41. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71 (2010).

42. Gasiunas, G., Barrangou, R., Horvath, P. y Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109, E2579-2586 (2012).

43. Umov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. y Gregory, P.D. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11, 636-646 (2010).

44. Perez, E.E. et al. Establishment of HIV-1 resistance in CD4(+) T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26, 808-816 (2008).

45. Chen, F.Q. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8, 753-U796 (2011).

46. Bedell, V.M. et al. In vivo genome editing using a high-efficiency TALEN system. Nature 491, 114-U 133 (2012).

47. Saleh-Gohari, N. y Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic Acids Res 32, 3683-3688 (2004).

48. Sapranauskas, R. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res 39, 9275-9282 (2011).

49. Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res 23, 720-723 (2013).

50. Tuschl, T. Expanding small RNA interference. Nat Biotechnol 20, 446-448 (2002).

51. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. y Kucherlapati, R.S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317, 230-234 (1985).

52. Thomas, K.R., Folger, K.R. y Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. Cell 44, 419-428 (1986).

53. Hasty, P., Rivera-Perez, J. y Bradley, A. The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol 11, 5586-5591 (1991).

54. Wu, S., Ying, G.X., Wu, Q. y Capecchi, M.R. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. Nat Protoc 3, 1056-1076 (2008).

55. Oliveira, T.Y. et al. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods 375, 176-181 (2012).

56. Tremblay et al., Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene; Human Gene Therapy. agosto de 2012, 23(8): 883-890.

57. Shalek et al. Nanowire-mediated delivery enables functional interrogation of primary immune cells: application to the analysis of chronic lymphocytic leukemia. Nano Letters, 12 de diciembre de 2012;12(12):6498-504.

58. Pardridge et al. Preparation of Trojan horse liposomes (THLs) for gene transfer across the blood-brain barrier; Cold Spring Harb Protoc; abril de 2010 (4)

59. Plosker GL et al. Fluvastatin: a review of its pharmacology and use in the management of hyper cholesterolaemia; Drugs 1996, 51(3):433-459).

60. Trapani et al. Potential role of nonstatin cholesterol lowering agents; IUBMB Life, Volumen 63, Cuaderno 11, páginas 964-971, noviembre de 2011

61. Birch AM et al. DGAT1 inhibitors as anti-obesity and anti-diabetic agents; Current Opinion in Drug Discovery & Development, 2010, 13(4):489-496

62. Fuchs et al. Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi), Oncogene 2002,21(37):5716-5724.

63. McManaman JL et al. Perilipin-2 Null Mice are Protected Against Diet-Induced Obesity, Adipose Inflammation and Fatty Liver Disease; The Journal of Lipid Research, jlr.M035063. Publicado por primera vez el 12 de febrero de 2013.

64. Tang J et al. Inhibition of SREBP by a Small Molecule, Betulin, Improves Hyperlipidemia and Insulin Resistance and Reduces Atherosclerotic Plaques; Cell Metabolism, Volumen 13, Cuaderno 1, 44-56, 5 de enero de 2011.

65. Dumitrache et al. Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair; Genetics. agosto de 2011; 188(4): 787-797

Aun cuando realizaciones preferidas de la presente invención se han mostrado y descrito en esta memoria, resultará obvio para los expertos en la técnica que estas realizaciones se proporcionan a modo de ejemplo.

LISTADO DE SECUENCIAS

<110> THE BROAD INSTITUTE, INC. MASSACHUSETTS INSTITUTE OF TECHNOLOGY PRESIDENT AND FELLOWS OF HARVARD COLLEGE

<120> MODIFICACIÓN Y OPTIMIZACIÓN DE SISTEMAS, MÉTODOS Y COMPOSICIONES ENZIMÁTICAS MEJORADOS PARA LA MANIPULACIÓN DE SECUENCIAS

<130> 1199-1001EP Divl

<140> DIVISIONAL EP DERIVADA DE LA SOLIC. EP Nº 13814364.9 (PCT/US2013/074691)
<141> 12-12-2013

<150> 61/836.101
<151> 17-06-2013

<150> 61/835.931
<151> 17-06-2013

<150> 61/828.130
<151> 28-05-2013

<150> 61/819.803
<151> 06-05-2013

<150> 61/814.263
<151> 20-04-2013

<150> 61/806.375
<151> 28-03-2013

<150> 61/802.174
<151> 15-03-2013

<150> 61/791.409
<151> 15-03-2013

<150> 61/769.046
<151> 25-02-2013

<150> 61/758.468
<151> 30-01-2013

<150> 61/748.427
<151> 02-01-2013

<150> 61/736.527
<151> 12-12-2012

<160> 228

<170> PatentIn version 3.5

<210> 1
<211> 20
<212> ADN
<213> Homo sapiens

<400> 1
gagtccgagc agaagaagaa 20

<210> 2
<211> 20
<212>ADN
<213> Homo sapiens

<400> 2
gagtcctagc aggagaagaa 20

<210> 3
<211> 20
<212> ADN
<213> Homo sapiens

<400> 3
gagtctaagc agaagaagaa 20

<210> 4
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(44)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (47)..(60)
<223> a, c, u, g, desconocida u otra

<400> 4
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnnggnnnn nnnnnnnnnn 60

<210> 5
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(14)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (17)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 5
nnnnnnnnnn nnnnccnnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 6
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(43)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (46)..(60)
<223> a, c, u, g, desconocida u otra

<400> 6
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn 60

<210> 7
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(15)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (18)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 7
nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 8
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(42)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (45)..(60)
<223> a, c, u, g, desconocida u otra

<400> 8
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnggnnnnnn nnnnnnnnnn 60

<210> 9
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(15)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (18)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 9
nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 10
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(41)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (44)..(60)
<223> a, c, u, g, desconocida u otra

<400> 10
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nggnnnnnnn nnnnnnnnnn 60

<210> 11
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(17)
<223> a, c, u, g, desconocida u otra

<220>
<221 base_modificada
<222> (20)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 11
nnnnnnnnnn nnnnnnnccn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 12
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(40)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (43)..(60)
<223> a, c, u, g, desconocida u otra

<400> 12
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn ggnnnnnnnn nnnnnnnnnn 60

<210> 13
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(18)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (21)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada

<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 13
nnnnnnnnnn nnnnnnnncc nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 14
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(39)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (42)..(60)
<223> a, c, u, g, desconocida u otra

<400> 14
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnng gnnnnnnnnn nnnnnnnnnn 60

<210> 15
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada

<222> (1)..(19)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (22)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 15
nnnnnnnnnn nnnnnnnnnc cnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 16
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 16
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 17
<211> 60

<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 17
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 18
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(37)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (40)..(60)
<223> a, c, u, g, desconocida u otra

<400> 18
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnggn nnnnnnnnnn nnnnnnnnnn 60

<210> 19
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(21)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (24)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 19
nnnnnnnnnn nnnnnnnnnn nccnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 20
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23).. (36)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (39)..(60)
<223> a, c, u, g, desconocida u otra

<400> 20
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnggnn nnnnnnnnnn nnnnnnnnnn 60

<210> 21
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(22)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (25)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 21
nnnnnnnnnn nnnnnnnnnn nnccnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 22
<211> 60

<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 22
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnggnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 23
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(23)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (26)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 23
nnnnnnnnnn nnnnnnnnnn nnnccnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 24
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(34)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (37)..(60)
<223> a, c, u, g, desconocida u otra

<400> 24
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnggnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 25
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(24)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (27)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 25
nnnnnnnnnn nnnnnnnnnn nnnnccnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 26
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(33)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (36)..(60)
<223> a, c, u, g, desconocida u otra

<400> 26
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 27
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(25)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (28)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 27
nnnnnnnnnn nnnnnnnnnn nnnnnccnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 28
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(32)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (35)..(60)
<223> a, c, u, g, desconocida u otra

<400> 28
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnggnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 29
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(26)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (29)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 29
nnnnnnnnnn nnnnnnnnnn nnnnnnccnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 30
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(31)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (34)..(60)
<223> a, c, u, g, desconocida u otra

<400> 30
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nggnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 31
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(27)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (30)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 31
nnnnnnnnnn nnnnnnnnnn nnnnnnnccn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 32
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(30)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (33)..(60)
<223> a, c, u, g, desconocida u otra

<400> 32
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn ggnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 33
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(28)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (31)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 33
nnnnnnnnnn nnnnnnnnnn nnnnnnnncc nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 34
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(29)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (32)..(60)
<223> a, c, u, g, desconocida u otra

<400> 34
nnnnnnnnnn nnnnnnnnnn ccnnnnnnng gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 35
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(29)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (32)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 35
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnc cnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 36
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(28)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (31)..(60)
<223> a, c, u, g, desconocida u otra

<400> 36
nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 37
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(30)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (33)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 37
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 38
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada

<222> (23)..(27)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (30)..(60)
<223> a, c, u, g, desconocida u otra

<400> 38
nnnnnnnnnn nnnnnnnnnn ccnnnnnggn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 39
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(31)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (34)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 39
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nccnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 40
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(26)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (29)..(60)
<223> a, c, u, g, desconocida u otra

<400> 40
nnnnnnnnnn nnnnnnnnnn ccnnnnggnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 41
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(32)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (35)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 41
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnccnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 42
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(25)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (28)..(60)
<223> a, c, u, g, desconocida u otra

<400> 42
nnnnnnnnnn nnnnnnnnnn ccnnnggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 43
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(33)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (36)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 43
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnccnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 44
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(24)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (27)..(60)
<223> a, c, u, g, desconocida u otra

<400> 44
nnnnnnnnnn nnnnnnnnnn ccnnggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 45
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(34)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (37)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 45
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnccnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 46
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (23)..(23)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (26)..(60)
<223> a, c, u, g, desconocida u otra

<400> 46
nnnnnnnnnn nnnnnnnnnn ccnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 47
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(38)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (41)..(60)
<223> a, c, u, g, desconocida u otra

<400> 47
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccngg nnnnnnnnnn nnnnnnnnnn 60

<210> 48
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(21)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (26)..(60)
<223> a, c, u, g, desconocida u otra

<400> 48
nnnnnnnnnn nnnnnnnnnn nccggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 49
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (40)..(60)
<223> a, c, u, g, desconocida u otra

<400> 49
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccggn nnnnnnnnnn nnnnnnnnnn 60

<210> 50
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(23)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (26)..(60)
<223> a, c, u, g, desconocida u otra

<400> 50
nnnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 51
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(33)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 51
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnggccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 52
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(22)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (26)..(60)
<223> a, c, u, g, desconocida u otra

<400> 52
nnnnnnnnnn nnnnnnnnnn nncggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 53
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(32)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (35)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 53
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnggnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 54
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(31)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (34)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 54
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nggnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 55
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(30)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (33)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 55
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ggnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 56
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(29)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (32)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 56
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng gnnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 57
<211> 60
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(28)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (31)..(35)
<223> a, c, u, g, desconocida u otra

<220>
<221> base_modificada
<222> (38)..(60)
<223> a, c, u, g, desconocida u otra

<400> 57
nnnnnnnnnn nnnnnnnnnn nnnnnnnngg nnnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 58
<211> 12
<212> ARN
<213> Secuencia Artificial

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

<400> 58
guuuuagagc ua 12

<210> 59
<211> 7
<212> PRT
<213> Virus del simio 40

<400> 59
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 60
<211> 16
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción de Desconocido: Secuencia de NLS bipartita de nucleoplasmina"

<400> 60
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 61
<211> 9
<212> PRT
<213> Desconocido

<220>

<221> fuente
<223> /nota="Descripción de Desconocido: secuencia de NLS C-myc"

<400> 61
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 62
<211> 11
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> / nota="Descripción de Desconocido: secuencia de NLS C-myc"

<400> 62
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> 63
<211> 38
<212> PRT
<213> Homo sapiens

<400> 63
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 64
<211> 42

<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción de Desconocido: dominio IBB de secuencia de importina-alfa"

<400> 64
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> 65
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción de Desconocido: secuencia de proteína T de mioma"

<400> 65
Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 66
<211> 8
<212> PRT
<213> Desconocido

<220>
<221> fuente
<223> / nota="Descripción de Desconocido: secuencia de proteína T de mioma"

<400> 66
Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 67
<211> 8

<212> PRT
<213> Homo sapiens

<400> 67
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 68
<211> 12
<212> PRT
<213> Mus musculus

<400> 68
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 69
<211> 5
<212> PRT
<213> Virus de la gripe

<400> 69
Asp Arg Leu Arg Arg
1 5

<210> 70
<211> 7
<212> PRT
<213> Virus de la gripe

<400> 70
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 71
<211> 10
<212> PRT
<213> Virus de Hepatitis delta

<400> 71
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 72
<211> 10

<212> PRT
<213> Mus musculus

<400> 72
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 73
<211> 20
<212> PRT
<213> Homo sapiens

<400> 73
Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys
20

<210> 74
<211> 17
<212> PRT
<213> Homo sapiens

<400> 74
Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 75
<211> 137
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 75
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

tcgttattta atttttt 137

<210> 76
<211> 123
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 76
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 77
<211> 110
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 77
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt 110

<210> 78
<211> 102
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 78
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt 102

<210> 79
<211> 88
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada

<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 79
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gttttttt 88

<210> 80
<211> 76
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, desconocida u otra

<400> 80
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcatt tttttt 76

<210> 81
<211> 984
<212> PRT
<213> Campylobacter jejuni

<400> 81
Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
1 5 10 15

Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
20 25 30

Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
35 40 45

Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg
50 55 60

Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
65 70 75 80

Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
85 90 95

Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu
100 105 110

Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg
115 120 125

Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala

```
    130                 135                 140

Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln
145                 150                 155                 160

Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu
                165                 170                 175

Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu
            180                 185                 190

Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe
        195                 200                 205

Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu
    210                 215                 220

Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser
225                 230                 235                 240

His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro
                245                 250                 255

Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile
            260                 265                 270

Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys
        275                 280                 285

Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu
    290                 295                 300

Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu
305                 310                 315                 320

Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys
                325                 330                 335

Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu
            340                 345                 350

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
        355                 360                 365

Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
    370                 375                 380
```

```
Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
385                 390                 395                 400

Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
                405                 410                 415

Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
            420                 425                 430

Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr
        435                 440                 445

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

Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
465                 470                 475                 480

Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
                485                 490                 495

Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys
            500                 505                 510

Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
        515                 520                 525

Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
    530                 535                 540

Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
545                 550                 555                 560

Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu
                565                 570                 575

Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
            580                 585                 590

Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala
        595                 600                 605

Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
    610                 615                 620

Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
625                 630                 635                 640
```

```
Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp
                645                 650                 655

Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln
            660                 665                 670

Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser
        675                 680                 685

Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His
    690                 695                 700

Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser
705                 710                 715                 720

Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser
                725                 730                 735

Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys
            740                 745                 750

Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp
        755                 760                 765

Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser
    770                 775                 780

Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln
785                 790                 795                 800

Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys
                805                 810                 815

Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg
            820                 825                 830

Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro
        835                 840                 845

Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
    850                 855                 860

Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865                 870                 875                 880

Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
                885                 890                 895
```

```
Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
            900                 905                 910

Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
        915                 920                 925

Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
    930                 935                 940

Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
945                 950                 955                 960

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
                965                 970                 975

Arg Gln Arg Glu Asp Phe Lys Lys
            980
```

<210> 82
<211> 91
<212> ADN
<213> Secuencia Artificial

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

<400> 82
tataatctca taagaaattt aaaaagggac taaaataaag agtttgcggg actctgcggg 60

gttacaatcc cctaaaaccg cttttaaaat t 91

<210> 83
<211> 36
<212> ADN
<213> Secuencia Artificial

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

<400> 83
gttttagtcc ctttttaaat ttctttatgg taaaat 36

<210> 84
<211> 95
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 84
nnnnnnnnnn nnnnnnnnnn guuuuagucc cgaaagggac uaaaauaaag aguuugcggg 60

acucugcggg guuacaaucc ccuaaaaccg cuuuu 95

<210> 85
<211> 3243
<212> ADN
<213> Secuencia Artificial

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

<400> 85

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc 120

gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg 180

ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc 240

ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac 300

ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc 360

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg 540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac 600

tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720

ggagaaggga gcccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga 780

cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg 840

tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg 900

gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca 960

ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg 1020

acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc 1080

acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg 1140

actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg 1200

acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg 1260

tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt 1320

gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag 1380

atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc 1440

cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt 1500

atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag 1560

aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag 1620

aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg 1680

tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc 1740

gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc 1800

aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca 1860

gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag 1920

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc 1980

tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc 2040

ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc 2100

atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac 2160

aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt 2220

aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag 2280

aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc 2340

actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg 2400

gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat 2460

aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag 2520

ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag 2580

acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat 2640

aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc 2700

gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac 2760

gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat 2820

gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa 2880

aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag 2940

attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat 3000

ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat 3060

atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga 3120

attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg 3180

ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa 3240

ttc 3243

<210> 86
<211> 1115
<212> PRT
<213> Secuencia Artificial

<220>
<221> fuente
<223> / nota="Descripción de la Secuencia Artificial: Polipéptido sintético”

<400> 86

```
Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1               5                  10                  15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
            20                  25                  30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
        35                  40                  45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
    50                  55                  60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
65                  70                  75                  80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
                85                  90                  95

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
            100                 105                 110

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

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
    130                 135                 140

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
145                 150                 155                 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
```

```
                165                 170                 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
            180                 185                 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
        195                 200                 205

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

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
225                 230                 235                 240

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
                245                 250                 255

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
            260                 265                 270

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
        275                 280                 285

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
    290                 295                 300

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
305                 310                 315                 320

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
                325                 330                 335

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
            340                 345                 350

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
        355                 360                 365

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
    370                 375                 380

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
385                 390                 395                 400

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
                405                 410                 415
```

```
Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
            420                 425                 430

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
        435                 440                 445

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
    450                 455                 460

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
465                 470                 475                 480

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
                485                 490                 495

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
            500                 505                 510

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
        515                 520                 525

Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
    530                 535                 540

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
545                 550                 555                 560

Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
                565                 570                 575

Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser
            580                 585                 590

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
        595                 600                 605

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
    610                 615                 620

Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
625                 630                 635                 640

Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
                645                 650                 655

Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
            660                 665                 670
```

```
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
        675                 680                 685

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

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
705                 710                 715                 720

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
                725                 730                 735

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
            740                 745                 750

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met
        755                 760                 765

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
    770                 775                 780

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
785                 790                 795                 800

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
                805                 810                 815

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
            820                 825                 830

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln
        835                 840                 845

Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp
    850                 855                 860

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly
865                 870                 875                 880

Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
                885                 890                 895

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
            900                 905                 910

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe
        915                 920                 925
```

```
Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
    930                 935                 940

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
945                 950                 955                 960

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
                965                 970                 975

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
            980                 985                 990

Lys Gly Ser Pro Glu Asp Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln
        995                 1000                1005

His Lys  His Tyr Leu Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe
    1010                1015                1020

Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu
    1025                1030                1035

Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala
    1040                1045                1050

Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr Asn Leu  Gly Ala Pro
    1055                1060                1065

Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile Asp Arg  Lys Arg Tyr
    1070                1075                1080

Thr Ser  Thr Lys Glu Val Leu  Asp Ala Thr Leu Ile  His Gln Ser
    1085                1090                1095

Ile Thr  Gly Leu Tyr Glu Thr  Arg Ile Asp Leu Ser  Gln Leu Gly
    1100                1105                1110

Gly Asp
    1115
```

<210> 87
<211> 1374
<212> PRT
<213> Secuencia Artificial

<220>
<221> fuente
<223> / nota="Descripción de la Secuencia Artificial: Polipéptido sintético”

<400> 87

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

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

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
```

```
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

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

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
```

```
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

```
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Thr Asn
        755                 760                 765

Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys Ala Asn Lys
    770                 775                 780

Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln Tyr Asn Gly
785                 790                 795                 800

Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys Gln Leu Ala
                805                 810                 815

Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys Leu Tyr Thr
            820                 825                 830

Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser Asn Gln Phe
        835                 840                 845

Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp Asp Ser Leu
    850                 855                 860

Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu Lys Gly Gln
865                 870                 875                 880

Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala Trp Ser Phe
                885                 890                 895

Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu Ser Asn Lys
            900                 905                 910

Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys Phe Asp Val
        915                 920                 925

Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg Tyr Ala Ser
    930                 935                 940

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

Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser Gln Leu Arg
                965                 970                 975

Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His His His Ala
            980                 985                 990

Val Asp Ala Leu Ile Ile Ala Ala  Ser Ser Gln Leu Asn  Leu Trp Lys
        995                 1000                1005
```

```
Lys Gln  Lys Asn Thr Leu Val  Ser Tyr Ser Glu Asp  Gln Leu Leu
    1010                 1015                 1020

Asp Ile  Glu Thr Gly Glu Leu  Ile Ser Asp Asp Glu  Tyr Lys Glu
    1025                 1030                 1035

Ser Val  Phe Lys Ala Pro Tyr  Gln His Phe Val Asp  Thr Leu Lys
    1040                 1045                 1050

Ser Lys  Glu Phe Glu Asp Ser  Ile Leu Phe Ser Tyr  Gln Val Asp
    1055                 1060                 1065

Ser Lys  Phe Asn Arg Lys Ile  Ser Asp Ala Thr Ile  Tyr Ala Thr
    1070                 1075                 1080

Arg Gln  Ala Lys Val Gly Lys  Asp Lys Ala Asp Glu  Thr Tyr Val
    1085                 1090                 1095

Leu Gly  Lys Ile Lys Asp Ile  Tyr Thr Gln Asp Gly  Tyr Asp Ala
    1100                 1105                 1110

Phe Met  Lys Ile Tyr Lys Lys  Asp Lys Ser Lys Phe  Leu Met Tyr
    1115                 1120                 1125

Arg His  Asp Pro Gln Thr Phe  Glu Lys Val Ile Glu  Pro Ile Leu
    1130                 1135                 1140

Glu Asn  Tyr Pro Asn Lys Gln  Ile Asn Glu Lys Gly  Lys Glu Val
    1145                 1150                 1155

Pro Cys  Asn Pro Phe Leu Lys  Tyr Lys Glu Glu His  Gly Tyr Ile
    1160                 1165                 1170

Arg Lys  Tyr Ser Lys Lys Gly  Asn Gly Pro Glu Ile  Lys Ser Leu
    1175                 1180                 1185

Lys Tyr  Tyr Asp Ser Lys Leu  Gly Asn His Ile Asp  Ile Thr Pro
    1190                 1195                 1200

Lys Asp  Ser Asn Asn Lys Val  Val Leu Gln Ser Val  Ser Pro Trp
    1205                 1210                 1215

Arg Ala  Asp Val Tyr Phe Asn  Lys Thr Thr Gly Lys  Tyr Glu Ile
    1220                 1225                 1230

Leu Gly  Leu Lys Tyr Ala Asp  Leu Gln Phe Glu Lys  Gly Thr Gly
```

```
    1235                1240                1245

Thr Tyr  Lys Ile Ser Gln Glu  Lys Tyr Asn Asp Ile  Lys Lys Lys
    1250                1255                1260

Glu Gly  Val Asp Ser Asp Ser  Glu Phe Lys Phe Thr  Leu Tyr Lys
    1265                1270                1275

Asn Asp  Leu Leu Leu Val Lys  Asp Thr Glu Thr Lys  Glu Gln Gln
    1280                1285                1290

Leu Phe  Arg Phe Leu Ser Arg  Thr Met Pro Lys Gln  Lys His Tyr
    1295                1300                1305

Val Glu  Leu Lys Pro Tyr Asp  Lys Gln Lys Phe Glu  Gly Gly Glu
    1310                1315                1320

Ala Leu  Ile Lys Val Leu Gly  Asn Val Ala Asn Ser  Gly Gln Cys
    1325                1330                1335

Lys Lys  Gly Leu Gly Lys Ser  Asn Ile Ser Ile Tyr  Lys Val Arg
    1340                1345                1350

Thr Asp  Val Leu Gly Asn Gln  His Ile Ile Lys Asn  Glu Gly Asp
    1355                1360                1365

Lys Pro  Lys Leu Asp Phe
    1370
```

<210> 88
<211> 4677
<212> ADN
<213> Secuencia Artificial

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

<400> 88

```
tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg       60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg      120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga      180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag      240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc      300

acaacccgca aacatgtacc catacgatgt tccagattac gcttcgccga agaaaaagcg      360

caaggtcgaa gcgtccgaca agaagtacag catcggcctg gacatcggca ccaactctgt      420
```

gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc aagaaattca aggtgctggg 480

caacaccgac cggcacagca tcaagaagaa cctgatcgga gccctgctgt tcgacagcgg 540

cgaaacagcc gaggccaccc ggctgaagag aaccgccaga agaagataca ccagacggaa 600

gaaccggatc tgctatctgc aagagatctt cagcaacgag atggccaagg tggacgacag 660

cttcttccac agactggaag agtccttcct ggtggaagag gataagaagc acgagcggca 720

ccccatcttc ggcaacatcg tggacgaggt ggcctaccac gagaagtacc ccaccatcta 780

ccacctgaga aagaaactgg tggacagcac cgacaaggcc gacctgcggc tgatctatct 840

ggccctggcc cacatgatca agttccgggg ccacttcctg atcgagggcg acctgaaccc 900

cgacaacagc gacgtggaca agctgttcat ccagctggtg cagacctaca accagctgtt 960

cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag gccatcctgt ctgccagact 1020

gagcaagagc agacggctgg aaaatctgat cgcccagctg cccggcgaga agaagaatgg 1080

cctgttcggc aacctgattg ccctgagcct gggcctgacc cccaacttca agagcaactt 1140

cgacctggcc gaggatgcca aactgcagct gagcaaggac acctacgacg acgacctgga 1200

caacctgctg gcccagatcg gcgaccagta cgccgacctg tttctggccg ccaagaacct 1260

gtccgacgcc atcctgctga gcgacatcct gagagtgaac accgagatca ccaaggcccc 1320

cctgagcgcc tctatgatca agagatacga cgagcaccac caggacctga ccctgctgaa 1380

agctctcgtg cggcagcagc tgcctgagaa gtacaaagag attttcttcg accagagcaa 1440

gaacggctac gccggctaca ttgacggcgg agccagccag gaagagttct acaagttcat 1500

caagcccatc ctggaaaaga tggacggcac cgaggaactg ctcgtgaagc tgaacagaga 1560

ggacctgctg cggaagcagc ggaccttcga caacggcagc atccccccacc agatccacct 1620

gggagagctg cacgccattc tgcggcggca ggaagatttt tacccattcc tgaaggacaa 1680

ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc tactacgtgg gccctctggc 1740

caggggaaac agcagattcg cctggatgac cagaaagagc gaggaaacca tcacccccctg 1800

gaacttcgag gaagtggtgg acaagggcgc ttccgcccag agcttcatcg agcggatgac 1860

caacttcgat aagaacctgc ccaacgagaa ggtgctgccc aagcacagcc tgctgtacga 1920

gtacttcacc gtgtataacg agctgaccaa agtgaaatac gtgaccgagg gaatgagaaa 1980

gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg gacctgctgt tcaagaccaa 2040

ccggaaagtg accgtgaagc agctgaaaga ggactacttc aagaaaatcg agtgcttcga 2100

ctccgtggaa atctccggcg tggaagatcg gttcaacgcc tccctgggca cataccacga 2160

tctgctgaaa attatcaagg acaaggactt cctggacaat gaggaaaacg aggacattct 2220

ggaagatatc gtgctgaccc tgacactgtt tgaggacaga gagatgatcg aggaacggct 2280

gaaaacctat gcccacctgt tcgacgacaa agtgatgaag cagctgaagc ggcggagata 2340

caccggctgg ggcaggctga gccggaagct gatcaacggc atccgggaca agcagtccgg 2400

caagacaatc ctggatttcc tgaagtccga cggcttcgcc aacagaaact tcatgcagct 2460

gatccacgac gacagcctga cctttaaaga ggacatccag aaagcccagg tgtccggcca 2520

gggcgatagc ctgcacgagc acattgccaa tctggccggc agccccgcca ttaagaaggg 2580

catcctgcag acagtgaagg tggtggacga gctcgtgaaa gtgatgggcc ggcacaagcc 2640

cgagaacatc gtgatcgaaa tggccagaga gaaccagacc acccagaagg gacagaagaa 2700

cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa gagctgggca gccagatcct 2760

gaaagaacac cccgtggaaa acacccagct gcagaacgag aagctgtacc tgtactacct 2820

gcagaatggg cgggatatgt acgtggacca ggaactggac atcaaccggc tgtccgacta 2880

cgatgtggac catatcgtgc ctcagagctt tctgaaggac gactccatcg acaacaaggt 2940

gctgaccaga agcgacaaga accggggcaa gagcgacaac gtgccctccg aagaggtcgt 3000

gaagaagatg aagaactact ggcggcagct gctgaacgcc aagctgatta cccagagaaa 3060

gttcgacaat ctgaccaagg ccgagagagg cggcctgagc gaactggata aggccggctt 3120

catcaagaga cagctggtgg aaacccggca gatcacaaag cacgtggcac agatcctgga 3180

ctcccggatg aacactaagt acgacgagaa tgacaagctg atccgggaag tgaaagtgat 3240

caccctgaag tccaagctgg tgtccgattt ccggaaggat ttccagtttt acaaagtgcg 3300

cgagatcaac aactaccacc acgcccacga cgcctacctg aacgccgtcg tgggaaccgc 3360

cctgatcaaa aagtacccta agctggaaag cgagttcgtg tacggcgact acaaggtgta 3420

cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc ggcaaggcta ccgccaagta 3480

cttcttctac agcaacatca tgaacttttt caagaccgag attaccctgg ccaacggcga 3540

gatccggaag cggcctctga tcgagacaaa cggcgaaacc ggggagatcg tgtgggataa 3600

gggccgggat tttgccaccg tgcggaaagt gctgagcatg ccccaagtga atatcgtgaa 3660

aaagaccgag gtgcagacag gcggcttcag caaagagtct atcctgccca agaggaacag 3720

cgataagctg atcgccagaa agaaggactg ggaccctaag aagtacggcg gcttcgacag 3780

ccccaccgtg gcctattctg tgctggtggt ggccaaagtg gaaaagggca agtccaagaa 3840

actgaagagt gtgaaagagc tgctggggat caccatcatg gaaagaagca gcttcgagaa 3900

gaatcccatc gactttctgg aagccaaggg ctacaaagaa gtgaaaaagg acctgatcat 3960

caagctgcct aagtactccc tgttcgagct ggaaaacggc cggaagagaa tgctggcctc 4020

tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc tccaaatatg tgaacttcct 4080

gtacctggcc agccactatg agaagctgaa gggctccccc gaggataatg agcagaaaca 4140

gctgtttgtg gaacagcaca agcactacct ggacgagatc atcgagcaga tcagcgagtt 4200

ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa gtgctgtccg cctacaacaa 4260

gcaccgggat aagcccatca gagagcaggc cgagaatatc atccacctgt ttaccctgac 4320

caatctggga gcccctgccg ccttcaagta ctttgacacc accatcgacc ggaagaggta 4380

caccagcacc aaagaggtgc tggacgccac cctgatccac cagagcatca ccggcctgta 4440

cgagacacgg atcgacctgt ctcagctggg aggcgacagc cccaagaaga agagaaaggt 4500

ggaggccagc taaggatccg gcaagactgg ccccgcttgg caacgcaaca gtgagcccct 4560

ccctagtgtg tttggggatg tgactatgta ttcgtgtgtt ggccaacggg tcaacccgaa 4620

cagattgata cccgccttgg catttcctgt cagaatgtaa cgtcagttga tggtact 4677

<210> 89
<211> 3150
<212> ADN
<213> Secuencia Artificial

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

<400> 89
tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgccta agaagaagag gaaggttaac acgattaaca tcgctaagaa 360

cgacttctct gacatcgaac tggctgctat cccgttcaac actctggctg accattacgg 420

tgagcgttta gctcgcgaac agttggccct tgagcatgag tcttacgaga tgggtgaagc 480

acgcttccgc aagatgtttg agcgtcaact taaagctggt gaggttgcgg ataacgctgc 540

cgccaagcct ctcatcacta ccctactccc taagatgatt gcacgcatca acgactggtt 600

tgaggaagtg aaagctaagc gcggcaagcg cccgacagcc ttccagttcc tgcaagaaat 660

caagccggaa gccgtagcgt acatcaccat taagaccact ctggcttgcc taaccagtgc 720

tgacaataca accgttcagg ctgtagcaag cgcaatcggt cgggccattg aggacgaggc 780

tcgcttcggt cgtatccgtg accttgaagc taagcacttc aagaaaaacg ttgaggaaca 840

actcaacaag cgcgtagggc acgtctacaa gaaagcattt atgcaagttg tcgaggctga 900

catgctctct aagggtctac tcggtggcga ggcgtggtct tcgtggcata aggaagactc 960

tattcatgta ggagtacgct gcatcgagat gctcattgag tcaaccggaa tggttagctt 1020

acaccgccaa aatgctggcg tagtaggtca agactctgag actatcgaac tcgcacctga 1080

atacgctgag gctatcgcaa cccgtgcagg tgcgctggct ggcatctctc cgatgttcca 1140

accttgcgta gttcctccta agccgtggac tggcattact ggtggtggct attgggctaa 1200

cggtcgtcgt cctctggcgc tggtgcgtac tcacagtaag aaagcactga tgcgctacga 1260

agacgtttac atgcctgagg tgtacaaagc gattaacatt gcgcaaaaca ccgcatggaa 1320

aatcaacaag aaagtcctag cggtcgccaa cgtaatcacc aagtggaagc attgtccggt 1380

cgaggacatc cctgcgattg agcgtgaaga actcccgatg aaaccggaag acatcgacat 1440

gaatcctgag gctctcaccg cgtggaaacg tgctgccgct gctgtgtacc gcaaggacaa 1500

ggctcgcaag tctcgccgta tcagccttga gttcatgctt gagcaagcca ataagtttgc 1560

taaccataag gccatctggt tcccttacaa catggactgg cgcggtcgtg tttacgctgt 1620

gtcaatgttc aacccgcaag gtaacgatat gaccaaagga ctgcttacgc tggcgaaagg 1680

taaaccaatc ggtaaggaag gttactactg gctgaaaatc cacggtgcaa actgtgcggg 1740

tgtcgacaag gttccgttcc ctgagcgcat caagttcatt gaggaaaacc acgagaacat 1800

catggcttgc gctaagtctc cactggagaa cacttggtgg gctgagcaag attctccgtt 1860

ctgcttcctt gcgttctgct ttgagtacgc tggggtacag caccacggcc tgagctataa 1920

ctgctccctt ccgctggcgt ttgacgggtc ttgctctggc atccagcact tctccgcgat 1980

gctccgagat gaggtaggtg gtcgcgcggt taacttgctt cctagtgaaa ccgttcagga 2040

catctacggg attgttgcta agaaagtcaa cgagattcta caagcagacg caatcaatgg 2100

gaccgataac gaagtagtta ccgtgaccga tgagaacact ggtgaaatct ctgagaaagt 2160

caagctgggc actaaggcac tggctggtca atggctggct tacggtgtta ctcgcagtgt 2220

gactaagcgt tcagtcatga cgctggctta cgggtccaaa gagttcggct tccgtcaaca 2280

agtgctggaa gataccattc agccagctat tgattccggc aagggtctga tgttcactca 2340

gccgaatcag gctgctggat acatggctaa gctgatttgg gaatctgtga gcgtgacggt 2400

ggtagctgcg gttgaagcaa tgaactggct taagtctgct gctaagctgc tggctgctga 2460

ggtcaaagat aagaagactg gagagattct tcgcaagcgt tgcgctgtgc attgggtaac 2520

tcctgatggt ttccctgtgt ggcaggaata caagaagcct attcagacgc gcttgaacct 2580

gatgttcctc ggtcagttcc gcttacagcc taccattaac accaacaaag atagcgagat 2640

tgatgcacac aaacaggagt ctggtatcgc tcctaacttt gtacacagcc aagacggtag 2700

ccaccttcgt aagactgtag tgtgggcaca cgagaagtac ggaatcgaat cttttgcact 2760

gattcacgac tccttcggta cgattccggc tgacgctgcg aacctgttca aagcagtgcg 2820

cgaaactatg gttgacacat atgagtcttg tgatgtactg gctgatttct acgaccagtt 2880

cgctgaccag ttgcacgagt ctcaattgga caaaatgcca gcacttccgg ctaaaggtaa 2940

cttgaacctc cgtgacatct tagagtcgga cttcgcgttc gcgtaaggat ccggcaagac 3000

tggccccgct tggcaacgca acagtgagcc cctccctagt gtgtttgggg atgtgactat 3060

gtattcgtgt gttggccaac gggtcaaccc gaacagattg atacccgcct tggcatttcc 3120

tgtcagaatg taacgtcagt tgatggtact 3150

<210> 90
<211> 125
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (23)..(42)
<223> a, c, t, g, desconocida u otra

<400> 90
gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaata 60

gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 120

ttttt 125

<210> 91
<211> 8452
<212> ADN
<213> Secuencia Artificial

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

<400> 91
tgcggtattt cacaccgcat caggtggcac ttttcgggga aatgtgcgcg gaacccctat 60

ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagattat caaaaaggat 120

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 180

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 240

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 300

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 360

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 420

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 480

agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc 540

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 600

catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 660

ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 720

atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 780

tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg cgccacatag 840

cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 900

cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 960

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 1020

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 1080

ttgaagcatt tatcagggtt attgtctcat gaccaaaatc ccttaacgtg agttttcgtt 1140

ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 1200

gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 1260

ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 1320

aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 1380

gcctacatac ctcgctctgc taatcctgtt accagtggct gttgccagtg gcgataagtc 1440

gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 1500

aacgggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 1560

cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 1620

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 1680

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 1740

atgctcgtca gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 1800

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 1860

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 1920

gcgcagcgag tcagtgagcg aggaagcggt cgctgaggct tgacatgatt ggtgcgtatg 1980

tttgtatgaa gctacaggac tgatttggcg ggctatgagg gcgggggaag ctctggaagg 2040

gccgcgatgg ggcgcgcggc gtccagaagg cgccatacgg cccgctggcg gcacccatcc 2100

ggtataaaag cccgcgaccc cgaacggtga cctccacttt cagcgacaaa cgagcactta 2160

tacatacgcg actattctgc cgctatacat aaccactcag ctagcttaag atcccatcaa 2220

gcttgcatgc cgggcgcgcc agaaggagcg cagccaaacc aggatgatgt ttgatggggt 2280

atttgagcac ttgcaaccct tatccggaag ccccctggcc cacaaaggct aggcgccaat 2340

gcaagcagtt cgcatgcagc ccctggagcg gtgccctcct gataaaccgg ccaggggcc 2400

tatgttcttt actttttac aagagaagtc actcaacatc ttaaaatggc caggtgagtc 2460

gacgagcaag cccggcggat caggcagcgt gcttgcagat ttgacttgca acgcccgcat 2520

tgtgtcgacg aaggcttttg gctcctctgt cgctgtctca agcagcatct aaccctgcgt 2580

cgccgtttcc atttgcagga gattcgaggt accatgtacc catacgatgt tccagattac 2640

gcttcgccga agaaaaagcg caaggtcgaa gcgtccgaca agaagtacag catcggcctg 2700

gacatcggca ccaactctgt gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc 2760

aagaaattca aggtgctggg caacaccgac cggcacagca tcaagaagaa cctgatcgga 2820

gccctgctgt tcgacagcgg cgaaacagcc gaggccaccc ggctgaagag aaccgccaga 2880

agaagataca ccagacggaa gaaccggatc tgctatctgc aagagatctt cagcaacgag 2940

atggccaagg tggacgacag cttcttccac agactggaag agtccttcct ggtggaagag 3000

gataagaagc acgagcggca ccccatcttc ggcaacatcg tggacgaggt ggcctaccac 3060

gagaagtacc ccaccatcta ccacctgaga aagaaactgg tggacagcac cgacaaggcc 3120

gacctgcggc tgatctatct ggccctggcc cacatgatca agttccgggg ccacttcctg 3180

atcgagggcg acctgaaccc cgacaacagc gacgtggaca agctgttcat ccagctggtg 3240

cagacctaca accagctgtt cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag 3300

gccatcctgt ctgccagact gagcaagagc agacggctgg aaaatctgat cgcccagctg 3360

cccggcgaga agaagaatgg cctgttcggc aacctgattg ccctgagcct gggcctgacc 3420

cccaacttca agagcaactt cgacctggcc gaggatgcca aactgcagct gagcaaggac 3480

acctacgacg acgacctgga caacctgctg gcccagatcg gcgaccagta cgccgacctg 3540

tttctggccg ccaagaacct gtccgacgcc atcctgctga gcgacatcct gagagtgaac 3600

accgagatca ccaaggcccc cctgagcgcc tctatgatca agagatacga cgagcaccac 3660

caggacctga ccctgctgaa agctctcgtg cggcagcagc tgcctgagaa gtacaaagag 3720

attttcttcg accagagcaa gaacggctac gccggctaca ttgacggcgg agccagccag 3780

gaagagttct acaagttcat caagcccatc ctggaaaaga tggacggcac cgaggaactg 3840

ctcgtgaagc tgaacagaga ggacctgctg cggaagcagc ggaccttcga caacggcagc 3900

atcccccacc agatccacct gggagagctg cacgccattc tgcggcggca ggaagatttt 3960

tacccattcc tgaaggacaa ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc 4020

tactacgtgg gccctctggc caggggaaac agcagattcg cctggatgac cagaaagagc 4080

gaggaaacca tcaccccctg gaacttcgag gaagtggtgg acaagggcgc ttccgcccag 4140

agcttcatcg agcggatgac caacttcgat aagaacctgc ccaacgagaa ggtgctgccc 4200

aagcacagcc tgctgtacga gtacttcacc gtgtataacg agctgaccaa agtgaaatac 4260

gtgaccgagg gaatgagaaa gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg 4320

gacctgctgt tcaagaccaa ccggaaagtg accgtgaagc agctgaaaga ggactacttc 4380

aagaaaatcg agtgcttcga ctccgtggaa atctccggcg tggaagatcg gttcaacgcc 4440

tccctgggca cataccacga tctgctgaaa attatcaagg acaaggactt cctggacaat 4500

gaggaaaacg aggacattct ggaagatatc gtgctgaccc tgacactgtt tgaggacaga 4560

gagatgatcg aggaacggct gaaaacctat gcccacctgt tcgacgacaa agtgatgaag 4620

cagctgaagc ggcggagata caccggctgg ggcaggctga gccggaagct gatcaacggc 4680

atccgggaca agcagtccgg caagacaatc ctggatttcc tgaagtccga cggcttcgcc 4740

aacagaaact tcatgcagct gatccacgac gacagcctga cctttaaaga ggacatccag 4800

aaagcccagg tgtccggcca gggcgatagc ctgcacgagc acattgccaa tctggccggc 4860

agccccgcca ttaagaaggg catcctgcag acagtgaagg tggtggacga gctcgtgaaa 4920

gtgatgggcc ggcacaagcc cgagaacatc gtgatcgaaa tggccagaga gaaccagacc 4980

acccagaagg gacagaagaa cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa 5040

gagctgggca gccagatcct gaaagaacac cccgtggaaa acacccagct gcagaacgag 5100

aagctgtacc tgtactacct gcagaatggg cgggatatgt acgtggacca ggaactggac 5160

atcaaccggc tgtccgacta cgatgtggac catatcgtgc ctcagagctt tctgaaggac 5220

gactccatcg acaacaaggt gctgaccaga agcgacaaga accggggcaa gagcgacaac 5280

gtgccctccg aagaggtcgt gaagaagatg aagaactact ggcggcagct gctgaacgcc 5340

aagctgatta cccagagaaa gttcgacaat ctgaccaagg ccgagagagg cggcctgagc 5400

gaactggata aggccggctt catcaagaga cagctggtgg aaacccggca gatcacaaag 5460

cacgtggcac agatcctgga ctcccggatg aacactaagt acgacgagaa tgacaagctg 5520

atccgggaag tgaaagtgat caccctgaag tccaagctgg tgtccgattt ccggaaggat 5580

ttccagtttt acaaagtgcg cgagatcaac aactaccacc acgcccacga cgcctacctg 5640

aacgccgtcg tgggaaccgc cctgatcaaa aagtacccta agctggaaag cgagttcgtg 5700

tacggcgact acaaggtgta cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc 5760

ggcaaggcta ccgccaagta cttcttctac agcaacatca tgaacttttt caagaccgag 5820

attaccctgg ccaacggcga gatccggaag cggcctctga tcgagacaaa cggcgaaacc 5880

ggggagatcg tgtgggataa gggccgggat tttgccaccg tgcggaaagt gctgagcatg 5940

ccccaagtga atatcgtgaa aaagaccgag gtgcagacag gcggcttcag caaagagtct 6000

atcctgccca agaggaacag cgataagctg atcgccagaa agaaggactg ggaccctaag 6060

aagtacggcg gcttcgacag ccccaccgtg gcctattctg tgctggtggt ggccaaagtg 6120

gaaaagggca agtccaagaa actgaagagt gtgaaagagc tgctggggat caccatcatg 6180

gaaagaagca gcttcgagaa gaatcccatc gactttctgg aagccaaggg ctacaaagaa 6240

gtgaaaaagg acctgatcat caagctgcct aagtactccc tgttcgagct ggaaaacggc 6300

cggaagagaa tgctggcctc tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc 6360

tccaaatatg tgaacttcct gtacctggcc agccactatg agaagctgaa gggctccccc 6420

gaggataatg agcagaaaca gctgtttgtg gaacagcaca agcactacct ggacgagatc 6480

atcgagcaga tcagcgagtt ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa 6540

gtgctgtccg cctacaacaa gcaccgggat aagcccatca gagagcaggc cgagaatatc 6600

atccacctgt ttaccctgac caatctggga gcccctgccg ccttcaagta ctttgacacc 6660

accatcgacc ggaagaggta caccagcacc aaagaggtgc tggacgccac cctgatccac 6720

cagagcatca ccggcctgta cgagacacgg atcgacctgt ctcagctggg aggcgacagc 6780

cccaagaaga agagaaaggt ggaggccagc taacatatga ttcgaatgtc tttcttgcgc 6840

tatgacactt ccagcaaaag gtagggcggg ctgcgagacg gcttcccggc gctgcatgca 6900

acaccgatga tgcttcgacc ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc 6960

gctccagggc gagcgctgtt taaatagcca ggccccccgat tgcaaagaca ttatagcgag 7020

ctaccaaagc catattcaaa cacctagatc actaccactt ctacacaggc cactcgagct 7080

tgtgatcgca ctccgctaag ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa 7140

catgacacaa gaatccctgt tacttctcga ccgtattgat tcggatgatt cctacgcgag 7200

cctgcggaac gaccaggaat tctgggaggt gagtcgacga gcaagcccgg cggatcaggc 7260

agcgtgcttg cagatttgac ttgcaacgcc cgcattgtgt cgacgaaggc ttttggctcc 7320

tctgtcgctg tctcaagcag catctaaccc tgcgtcgccg tttccatttg cagccgctgg 7380

cccgccgagc cctggaggag ctcgggctgc cggtgccgcc ggtgctgcgg gtgcccggcg 7440

agagcaccaa ccccgtactg gtcggcgagc ccggcccggt gatcaagctg ttcggcgagc 7500

actggtgcgg tccggagagc ctcgcgtcgg agtcggaggc gtacgcggtc ctggcggacg 7560

ccccggtgcc ggtgccccgc ctcctcggcc gcggcgagct gcggcccggc accggagcct 7620

ggccgtggcc ctacctggtg atgagccgga tgaccggcac cacctggcgg tccgcgatgg 7680

acggcacgac cgaccggaac gcgctgctcg ccctggcccg cgaactcggc cgggtgctcg 7740

gccggctgca cagggtgccg ctgaccggga acaccgtgct caccccccat tccgaggtct 7800

tcccggaact gctgcgggaa cgccgcgcgg cgaccgtcga ggaccaccgc gggtggggct 7860

acctctcgcc ccggctgctg gaccgcctgg aggactggct gccggacgtg gacacgctgc 7920

tggccggccg cgaaccccgg ttcgtccacg gcgacctgca cgggaccaac atcttcgtgg 7980

acctggccgc gaccgaggtc accgggatcg tcgacttcac cgacgtctat gcgggagact 8040

cccgctacag cctggtgcaa ctgcatctca acgccttccg gggcgaccgc gagatcctgg 8100

ccgcgctgct cgacggggcg cagtggaagc ggaccgagga cttcgcccgc gaactgctcg 8160

ccttcacctt cctgcacgac ttcgaggtgt tcgaggagac cccgctggat ctctccggct 8220

tcaccgatcc ggaggaactg gcgcagttcc tctgggggcc gccggacacc gcccccggcg 8280

cctgataagg atccggcaag actggccccg cttggcaacg caacagtgag cccctcccta 8340

gtgtgtttgg ggatgtgact atgtattcgt gtgttggcca acgggtcaac ccgaacagat 8400

tgatacccgc cttggcattt cctgtcagaa tgtaacgtca gttgatggta ct 8452

<210> 92
<211> 102
<212> ADN
<213> Secuencia Artificial

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

<400> 92
gttttagagc tatgctgttt tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa 60

gaagaagttt tagagctatg ctgttttgaa tggtcccaaa ac 102

<210> 93
<211> 100
<212> ADN
<213> Homo sapiens

<400> 93
cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag 60

aagaagggct cccatcacat caaccggtgg cgcattgcca 100

<210> 94
<211> 50
<212> ADN
<213> Homo sapiens

<400> 94
agctggagga ggaagggcct gagtccgagc agaagaagaa gggctcccac 50

<210> 95
<211> 30
<212> ARN
<213> Secuencia Artificial

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

<400> 95
gaguccgagc agaagaagaa guuuuagagc 30

<210> 96
<211> 49
<212> ADN
<213> Secuencia Artificial

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

<400> 96
agctggagga ggaagggcct gagtccgagc agaagagaag ggctcccat 49

<210> 97
<211> 53
<212> ADN
<213> Homo sapiens

<400> 97
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccatca cat 53

<210> 98
<211> 52
<212> ADN
<213> Homo sapiens

<400> 98
ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at 52

<210> 99
<211> 54
<212> ADN
<213> Homo sapiens

<400> 99
ctggaggagg aagggcctga gtccgagcag aagaaagaag ggctcccatc acat 54

<210> 100
<211> 50
<212> ADN
<213> Homo sapiens

<400> 100
ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat 50

<210> 101
<211> 47
<212> ADN
<213> Homo sapiens

<400> 101
ctggaggagg aagggcctga gcccgagcag aagggctccc atcacat 47

<210> 102
<211> 155
<212> ARN
<213> Secuencia Artificial

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

<400> 102
guuucagugc uauaggaaac uauaggaaau caccuucggg ugagcugaaa uccccuaaag 60

cuaagauuga auccggccac uaucuauuag uagauauccg gauauucuga uauaaaaccu 120

cauucuuuga uuagaccaaa ggaugagguu uuuuu 155

<210> 103
<211> 107
<212> ARN
<213> Secuencia Artificial

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

<400> 103
guuugagagu uggaaacaac gaguucaaau aagaauucau caaaaucguc ccuuuuggga 60

ccgcucauug uggagcauca aggcuuaaca ugguuaagcc uuuuuuu 107

<210> 104
<211> 99
<212> ARN
<213> Secuencia Artificial

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

<400> 104
guuugagagu aggaaacuac acgguucaaa uaaagaauuu uucuaaucgc ccaaugggcc 60

cauauugaua uggaugaaac ucgcuuagcg aguuuuuuu 99

<210> 105
<211> 91
<212> ARN
<213> Secuencia Artificial

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

<400> 105
guuuuagcac uagaaauagu uaaguuaaaa caagcuuaaa gcgucaaugu aauauuuuau 60

uaacacccua cugugucagu gggguuuuuu u 91

<210> 106
<211> 94
<212> ARN
<213> Secuencia Artificial

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

<400> 106
guuuuagaug gugaaaacca gauuuaaaau caagcaaugc aucuuuugau gcaaaguuuc 60

aauauuuguc ccacguuauc gagggacuuu uuuu 94

<210> 107
<211> 78
<212> ARN
<213> Secuencia Artificial

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

<400> 107
auuuuaguac cuggaaacag aucuacuaaa acaaggcuuu augccgaaau caagagcacc 60

gacgggugcu cuuuuuuu 78

<210> 108
<211> 84
<212> ARN
<213> Secuencia Artificial

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

<400> 108
guuuuuguac ucgaaagagc cuacaaagau aaggcuuuau gccgaauuca agcaccccau 60

guuuugacau gaggugcuuu uuuu 84

<210> 109
<211> 85
<212> ARN
<213> Secuencia Artificial

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

<400> 109
guuuuuguac cuuagaaaua agaucuacaa aaauaaggau uuauuccgaa uuuaccaccu 60

auuuuaauua auaggugguu uuuuu 85

<210> 110
<211> 100
<212> ARN
<213> Secuencia Artificial

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

<400> 110
guugugauuu gcuuucauuu gaaaaauuga agcaaaucac aauaaggauu auuccguugu 60

gaaaacaauu aaagcggucu ugcaaaaggu cgcuuuuuuu 100

<210> 111
<211> 102
<212> ARN
<213> Secuencia Artificial

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

<400> 111
guugugauuc gcuuucaaug aaaauugaag cgaaucacaa uaaggauuau uccguuguga 60

aaacauuuac uacggggcau cgaaagacug ccucguuuuu uu 102

<210> 112
<211> 90
<212> ARN
<213> Secuencia Artificial

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

<400> 112
guugugguuu gauuagaaau aaucuuauca caauaaggcu auaugccgua gacgaaaguc 60

uuuagucccg cuucgguggg acuuuuuuuu 90

<210> 113
<211> 78
<212> ARN
<213> Secuencia Artificial

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

<400> 113
guuggggaug accgcgaaag cgauuaucuc uaauaagacu uaagucgcaa aaugcucccu 60

auuuugggag cuuuuuuu 78

<210> 114
<211> 85
<212> ARN
<213> Secuencia Artificial

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

<400> 114
guugcggcug gagaaaucca gccguuaaca uguucccuuc ggggagcacg aaaugcgggg 60

cgggccacgg uccgccccuu uuuuu 85

<210> 115
<211> 126
<212> ARN
<213> Secuencia Artificial

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

<400> 115
gccgugguuu ccgaaaggaa accacgguaa cagaauuacc guaagguuuu uucugugaag 60

gaucaucccu cgcuugggca accaggcggg ggaaauuccu cguucgggcc aaucagcccu 120

uuuuuu 126

<210> 116
<211> 82
<212> ARN
<213> Secuencia Artificial

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

<400> 116
guuguaauuc ccugaaaagg uuauuacaau aagguaagaa accuaaaagc ucuaauccca 60

uucuucggaa ugggauuuuu uu 82

<210> 117
<211> 83
<212> ARN
<213> Secuencia Artificial

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

<400> 117
guuuuaguac ucuggaaaca gaaucuacua aaacaaggca aaaugccgug uuuaucucgu 60

caacuuguug gcgagauuuu uuu 83

<210> 118
<211> 94
<212> ARN
<213> Secuencia Artificial

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

<400> 118
guuuuagcac uguacaagaa auugucgugc uaaaauaagg cgcuguuaau gcagcugccg 60

cauccgccag agcauuuaug cucuggcuuu uuuu 94

<210> 119
<211> 4350
<212> ADN
<213> Secuencia Artificial

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

<400> 119

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgac tcagagcgag cgacgatttt cttgcagcat tggcattgac 120

atgggggcta agtacactgg ggtgttctac gcactgttcg accgggagga actgcccaca 180

aacctgaaca gcaaggccat gaccctggtc atgcctgaga cagggccaag atacgtgcag 240

gcacagagaa ctgccgtcag acacaggctg cgcggacaga agagatatac cctggctagg 300

aaactggcat ttctggtggt cgacgatatg atcaagaaac aggaaaagag gctgactgat 360

gaggaatgga aacgaggacg ggaggccctg tccggcctgc tgaagcggag agggtactct 420

cggcccaacg ctgacggcga agatctgacc cctctggaga atgtgagagc agacgtgttc 480

gccgctcatc ctgccttcag cacatatttt tccgaagtgc gctctctggc tgagcagtgg 540

gaggagttca ccgcaaacat cagcaatgtc gagaagtttc tgggcgaccc aaacatcccc 600

gccgataaag agttcattga atttgccgtg gctgaagggc tgattgacaa gaccgagaag 660

aaagcctacc agtcagctct gagcaccctg agggcaaacg ccaatgtgct gacaggactg 720

cggcagatgg gccacaagcc tagatcagaa tattttaaag caatcgaggc cgacctgaag 780

aaagatagcc gcctggccaa gattaacgaa gcattcggag gagcagagcg cctggctcga 840

ctgctgggaa acctgtccaa tctgcagctg cgggcagaaa gatggtactt caatgccccc 900

gacatcatga aggatagggg ctgggagcct gatcgcttca agaaaacact ggtgcgggct 960

tttaagttct ttcacccagc aaaggaccag aacaaacagc atctggaact gatcaaacag 1020

attgagaaca gcgaagatat cattgagact ctgtgcaccc tggacccaaa cagaaccatc 1080

ccccccttacg aggatcagaa caataggcgc ccacccctgg accagactct gctgctgagt 1140

cccgaaaagc tgacccggca gtatggcgag atctggaaaa catggagcgc cagactgacc 1200

tccgctgaac ccacactggc acctgcagcc gagattctgg aaagatctac cgacaggaag 1260

agtcgcgtgg cagtcaacgg acacgagcca ctgcctacac tggcttacca gctgagttat 1320

gcactgcaga gagccttcga caggtcaaaa gccctggatc catatgctct gagggcactg 1380

gctgcaggct caaaaagcaa taagctgaca tccgcccgca ctgctctgga gaactgcatc 1440

ggaggccaga atgtgaaaac cttcctggac tgtgcccgac ggtactatcg ggaagcagac 1500

gatgccaaag tcgggctgtg gttcgacaac gccgatggac tgctggagag atctgacctg 1560

catcctccaa tgaagaaaaa gatcctgccc ctgctggtgg ccaatattct gcagacagat 1620

gaaaccacag gccagaagtt tctggacgag atctggcgaa aacagattaa ggggcgggaa 1680

actgtggcta gccgatgtgc acggatcgag acagtgcgga aatccttcgg gggaggcttt 1740

aacattgcct acaataccgc tcagtatagg gaggtgaaca agctgccccg caatgcccag 1800

gataaagaac tgctgacaat cagagatagg gtggctgaga ctgcagactt cattgccgct 1860

aacctggggc tgtctgacga gcagaaaaga aagttcgcca atcctttag tctggctcag 1920

ttctacaccc tgatcgagac agaagtgtcc ggattttctg caactaccct ggccgtccac 1980

ctggagaacg cctggaggat gacaatcaag gatgctgtga ttaatgggga aactgtcaga 2040

gcagcacagt gcagcaggct gcctgcagag acagctcgcc cattcgatgg actggtgaga 2100

aggctggtcg acagacaggc ttgggagatc gcaaagaggg tgtcaactga cattcagagc 2160

aaagtcgatt tctccaacgg catcgtggac gtcagcattt ttgtggagga aaataagttc 2220

gagttttccg catctgtggc cgatctgaaa aagaacaaac gggtcaaaga caagatgctg 2280

tccgaggccg aaaagctgga aaccagatgg ctgatcaaaa atgagcggat caagaaggcc 2340

agccggggaa cttgtcccta caccggcgat aggctggctg agggggggaga aatcgaccac 2400

attctgcccc gaagcctgat caaggatgcc cggggaattg tgtttaacgc tgagcctaat 2460

ctgatctatg caagctcccg cggcaaccag ctgaaaaaga atcagcgata cagtctgtca 2520

gatctgaagg ccaactatcg gaatgagatc ttcaaaacta gcaacatcgc tgcaattacc 2580

gccgagattg aggacgtggt cactaagctg cagcagaccc atagactgaa attctttgat 2640

ctgctgaatg agcacgaaca ggactgcgtg cggcacgccc tgttcctgga cgatggcagc 2700

gaagctcgcg acgcagtgct ggagctgctg gcaacacagc gccgaactcg cgtcaacggg 2760

acacagatct ggatgattaa gaacctggcc aacaagatcc gagaggaact gcagaattgg 2820

tgtaagacaa ctaacaatag actgcacttt caggccgctg caactaacgt gtccgatgca 2880

```
aagaatctga ggctgaaact ggcccagaac cagcccgact tcgagaagcc agatatccag     2940
cccattgcca gccattccat cgacgccctg tgctctttcg ctgtggggag tgctgacgca     3000
gaacgcgatc agaatggatt tgactacctg gatggcaaga ccgtgctggg actgtatcca     3060
cagagctgtg aggtcattca cctgcaggcc aagccccagg aggaaaaaag tcatttcgat     3120
tcagtggcta tctttaagga aggcatctac gcagagcagt tcctgcctat ctttaccctg     3180
aacgaaaaga tctggattgg atatgagaca ctgaatgcca aaggcgaaag atgcggggct     3240
attgaggtga gtggcaaaca gccaaaggag ctgctggaaa tgctggcccc cttctttaac     3300
aagcctgtgg gcgacctgtc agcccacgct acttaccgga tcctgaaaaa gcctgcatat     3360
gagtttctgg caaaggcagc tctgcagcca ctgagcgcag aggaaaaaag actggcagcc     3420
ctgctggatg ctctgcgcta ctgtaccagt cgaaagtcac tgatgagcct gttcatggct     3480
gcaaacggga aatccctgaa aaagcgggag gacgtgctga aacccaagct gttccagctg     3540
aaggtcgagc tgaaaggcga aaagagcttc aagctgaacg ggagcctgac cctgcctgtg     3600
aagcaggact ggctgagaat ctgcgatagc ccagaactgg cagacgcctt tggcaaaccc     3660
tgttccgccg atgagctgac atctaagctg gctcgcattt ggaaacgacc tgtgatgcgg     3720
gatctggctc atgcaccagt ccggagagag ttcagcctgc ccgcaatcga caacccaagt     3780
ggagggttca ggattaggcg caccaacctg tttggcaatg agctgtacca ggtgcacgcc     3840
atcaacgcta aaaagtatcg cggcttcgcc tccgctgggt ctaatgtcga ctggtccaag     3900
gggatcctgt ttaacgagct gcagcatgaa aatctgaccg agtgcggagg caggttcatt     3960
acaagcgccg atgtgactcc tatgtccgaa tggcgcaagg tggtcgcaga ggacaacctg     4020
tctatctgga ttgctccagg gacagaagga cgacggtacg tgagggtcga gacaacattc     4080
atccaggcca gtcactggtt tgaacagtca gtggagaatt gggccattac tagtcctctg     4140
tcactgccag cttccttcaa ggtggacaaa ccagctgagt ttcagaaggc agtcggaacc     4200
gagctgtcag aactgctggg ccagcccagg agcgaaatct tcattgagaa cgtgggcaat     4260
gccaagcata tccgcttttg gtacattgtg gtgagcagca acaaaaagat gaacgagtct     4320
tacaacaatg tgtctaagag ttaagaattc                                      4350
```

<210> 120
<211> 4269
<212> ADN
<213> Secuencia Artificial

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

<400> 120

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgaa aaaagaaatc aaagactact tcctggggct ggatgtgggg 120

actgggagcg tggggtgggc tgtgaccgat actgactaca aactgctgaa ggctaaccga 180

aaagacctgt ggggcatgag atgcttcgag acagccgaaa ctgctgaggt gcggagactg 240

cacaggggag ccaggcgccg aatcgagcgg agaaagaaac gcattaagct gctgcaggag 300

ctgttctctc aggaaatcgc caaaaccgat gagggcttct ttcagagaat gaaggaaagc 360

ccctttacg ctgaggacaa aacaatcctg caggaaaaca ctctgttcaa tgacaaggat 420

tttgctgata agacttacca caaagcatat cctaccatta atcatctgat caaggcttgg 480

attgagaaca aggtgaaacc agacccccga ctgctgtacc tggcatgtca caacatcatt 540

aagaaaaggg gacatttcct gtttgaaggc gacttcgatt cagagaatca gtttgatacc 600

agcatccagg cactgttcga gtatctgcgc gaggacatgg aggtggacat cgatgccgac 660

agccagaagg tcaaagagat tctgaaggat agctccctga agaactctga aaaacagagt 720

cggctgaata agatcctggg gctgaagcct tccgacaaac agaagaaagc catcacaaac 780

ctgatttctg gaaacaagat caatttcgcc gatctgtacg acaatccaga tctgaaggac 840

gctgagaaaa actcaatcag cttctccaag gacgattttg atgcactgag tgacgatctg 900

gcctcaattc tgggcgacag ctttgaactg ctgctgaagg ccaaagctgt ctataactgc 960

tctgtgctga gtaaggtcat cggggacgag cagtacctga gcttcgccaa ggtgaaaatc 1020

tacgaaaagc acaaaaccga tctgacaaag ctgaaaaacg tgatcaagaa acatttcccc 1080

aaggactaca agaaggtctt tggatacaac aagaacgaga aaaacaacaa caattactcc 1140

ggctatgtgg gagtctgtaa gaccaagagt aagaaactga tcattaacaa ctcagtcaac 1200

caggaagatt tctacaagtt tctgaaaact atcctgtcag ccaagagcga gatcaaggaa 1260

gtgaatgaca tcctgaccga gattgaaact ggcaccttc tgccaaagca gatctctaaa 1320

agtaacgcag agattcccta tcagctgagg aaaatggagc tggaaaagat cctgtccaat 1380

gccgaaaagc acttctcttt tctgaagcag aaagacgaaa aaggactgtc acatagcgag 1440

aagatcatta tgctgctgac cttcaagatc ccttactata ttggcccaat caacgataat 1500

cacaagaaat tctttcccga cagatgctgg gtggtcaaga aagagaaatc cccttctggc 1560

aagaccacac catggaactt ctttgatcat atcgacaagg aaaaaacagc agaggccttc 1620

attacttcta ggaccaattt ttgcacttac ctggtgggag agagcgtcct gcctaagtct 1680

agtctgctgt actccgaata taccgtgctg aacgagatca acaatctgca gatcattatc 1740

gatggcaaga atatttgtga catcaagctg aaacagaaga tctacgagga cctgttcaag 1800

aagtacaaga aaattaccca gaagcagatc agcaccttca tcaagcacga aggcatctgc 1860

aacaaaaccg atgaagtgat catcctgggg attgacaagg aatgtacatc aagcctgaaa 1920

agctacatcg agctgaaaaa cattttcggc aagcaggtgg atgagatctc cactaagaat 1980

atgctggagg aaattatcag atgggctacc atctacgacg agggggaagg aaagaccatc 2040

ctgaaaacaa agatcaaggc agaatacgga aagtattgct ccgacgagca gattaagaaa 2100

atcctgaacc tgaagttctc cggctggggg cgactgtctc ggaaatttct ggagacagtg 2160

actagtgaaa tgcccggctt ctcagaacct gtcaatatta tcaccgccat gagggagaca 2220

cagaacaatc tgatggagct gctgtcctct gagttcacct tcaccgagaa cattaagaaa 2280

atcaattctg gattcgaaga tgccgagaag cagtttagtt acgacggcct ggtgaaacca 2340

ctgtttctga gtccctcagt caagaaaatg ctgtggcaga ccctgaagct ggtgaaagag 2400

attagccata tcacacaggc ccccctaag aaaattttca tcgaaatggc aaaggggcc 2460

gagctggaac ctgctcggac taagaccaga ctgaaaatcc tgcaggatct gtataacaat 2520

tgtaagaacg atgctgacgc cttcagctca gagatcaaag acctgagcgg aaagattgag 2580

aacgaagata atctgaggct gcgctccgac aagctgtacc tgtactatac tcagctgggg 2640

aaatgcatgt attgtggaaa gccaattgag atcggccacg tgttcgatac ctcaaactac 2700

gatattgacc atatctatcc ccagagcaag atcaaagacg atagcatttc caatcgggtg 2760

ctggtctgca gctcctgtaa caagaacaag gaggacaagt acccactgaa atcagagatc 2820

cagagcaagc agcgcggctt ctggaacttt ctgcagcgaa acaatttcat ttctctggag 2880

aagctgaata gactgacaag ggccactcca atcagtgacg atgagacagc caagtttatt 2940

gctaggcagc tggtggaaac tcgccaggct accaaggtgg ccgctaaagt cctggaaaag 3000

atgttccccg agacaaaaat cgtgtacagc aaggccgaga ctgtctccat gttccggaac 3060

aagtttgata tcgtgaagtg cagagaaatt aacgattttc accatgctca cgacgcatac 3120

ctgaatatcg tggtcggcaa cgtgtataat accaagttca caaacaatcc ttggaacttt 3180

atcaaggaga aaagagataa tccaaagatt gctgacacct acaactacta taaggtgttc 3240

gattatgacg tcaaaaggaa caatatcaca gcatgggaga aggggaaaac tattatcacc 3300

gtgaaagaca tgctgaagag aaacacacca atctacacta ggcaggcagc ctgtaagaaa 3360

ggggagctgt tcaatcagac cattatgaag aaaggactgg gccagcaccc cctgaagaaa 3420

gaaggacctt tttccaatat ctctaaatac ggcgggtata acaaggtgag cgctgcatac 3480

tatacactga ttgagtatga ggaaaagggc aacaaaatcc gctccctgga aactattccc 3540

ctgtacctgg tgaaagatat ccagaaggat caggacgtcc tgaagtctta tctgacagac 3600

ctgctgggga agaaagagtt caagatcctg gtgcccaaga tcaagatcaa cagcctgctg 3660

aagatcaatg ggtttccttg ccatattaca ggaaaaacta acgatagttt cctgctgcgc 3720

cctgccgtgc agttttgctg ttcaaacaat gaggtcctgt acttcaagaa aattatccgg 3780

ttttccgaaa tccgctctca gcgagagaag atcgggaaaa caattagccc atacgaggac 3840

ctgagcttcc ggtcatatat caaggagaac ctgtggaaga aaactaagaa cgatgaaatc 3900

ggagagaagg aattttacga cctgctgcag aagaaaaacc tggagatcta tgatatgctg 3960

ctgactaagc acaaagacac catctacaag aaacgcccta attctgccac cattgatatc 4020

ctggtgaagg ggaaagagaa gttcaaaagc ctgattatcg aaaaccagtt tgaagtgatc 4080

ctggagatcc tgaagctgtt ttctgcaaca cggaatgtca gtgacctgca gcatatcgga 4140

ggcagcaagt actccggcgt ggccaaaatc gggaacaaga tctctagtct ggataactgt 4200

atcctgatct atcagtccat caccggcatc ttcgagaaac ggatcgacct gctgaaggtg 4260

taagaattc 4269

<210> 121
<211> 4179
<212> ADN
<213> Secuencia Artificial

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

<400> 121
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgac caaggagtat tacctggggc tggatgtggg gaccaattcc 120

gtgggatggg cagtgaccga ttctcagtac aacctgtgca agtttaagaa aaaggatatg 180

tggggcatcc ggctgttcga aagcgccaac acagcaaagg accggagact gcagagaggg 240

aataggcgcc gactggagcg gaaaaagcag agaattgatc tgctgcagga aatcttctcc 300

ccagagatct gcaagattga ccccactttc tttatccgac tgaacgaatc ccggctgcac 360

ctggaggaca agtctaacga tttcaaatac ccactgttta ttgagaagga ctattctgat 420

atcgagtact ataaagagtt ccccaccatt tttcacctga ggaagcatct gatcgagagt 480

gaggaaaaac aggatatccg gctgatctac ctggccctgc acaacatcat taagacccga 540

ggacattttc tgattgacgg cgatctgcag agcgccaagc agctgaggcc catcctggat 600

acattcctgc tgtccctgca ggaggaacag aacctgtcag tgagcctgtc cgaaaatcag 660

aaggacgagt atgaggaaat tctgaaaaac cgcagcatcg ccaagtccga aaaagtgaaa 720

aagctgaaga atctgtttga gatctcagac gagctggaaa aagaggagaa gaaggcccag 780

agcgccgtga tcgagaactt ctgcaagttt atcgtgggaa ataagggcga tgtctgtaaa 840

ttcctgcggg tgtctaagga ggaactggag attgactctt tcagtttttc agagggcaag 900

tacgaggacg acatcgtgaa aaacctggag gaaaaagtgc ctgaaaaggt ctacctgttt 960

gagcagatga aggcaatgta tgattggaat attctggtcg acatcctgga aaccgaggaa 1020

tacatcagct tcgccaaagt gaagcagtat gagaaacaca agactaacct gcggctgctg 1080

agagacatca ttctgaaata ctgcaccaag gatgagtata atcggatgtt taacgacgag 1140

aaggaagctg gcagctacac cgcatatgtg gggaaactga aaaagaacaa caagaagtac 1200

tggatcgaga aaaagagaaa tcccgaggag ttctacaaat ccctgggcaa gctgctggat 1260

aaaattgagc ctctgaagga ggacctggaa gtgctgacta tgatgatcga ggagtgtaag 1320

aaccacaccc tgctgccaat tcagaaaaat aaggacaacg gcgtgatccc ccaccaggtg 1380

catgaggtcg aactgaaaaa gatcctggaa aatgccaaaa agtactattc cttcctgacc 1440

gagacagaca aggatgggta ctcagtggtc cagaaaatcg agagcatttt caggtttcgc 1500

atcccctact atgtggggcc tctgagtacc cggcaccagg aaaagggatc aaacgtgtgg 1560

atggtcagaa aacctggcag ggaggatcgc atctacccat ggaatatgga ggaaatcatt 1620

gactttgaga agagcaacga aaatttcatt acacggatga ctaacaaatg tacatatctg 1680

atcggggaag atgtcctgcc caagcattct ctgctgtaca gtaaatatat ggtgctgaat 1740

gagctgaaca atgtgaaggt cagaggaaaa aagctgccta catctctgaa acagaaggtg 1800

ttcgaggacc tgtttgaaaa caaatccaaa gtgactggaa agaatctgct ggagtacctg 1860

cagatccagg acaaagatat ccagattgac gatctgtctg gcttcgacaa ggacttcaag 1920

accagcctga agagctatct ggacttcaaa aagcagattt ttggggagga aatcgagaag 1980

gaaagcattc agaacatgat cgaagatatc attaagtgga tcaccatcta cggcaatgac 2040

aaggagatgc tgaaacgagt gattcgggct aattatagca accagctgac agaggaacag 2100

atgaaaaaga tcactggatt tcagtacagt ggctggggga acttctcaaa gatgtttctg 2160

aaagggatca gcggatccga cgtgagcacc ggcgaaacat tcgacatcat taccgcaatg 2220

tgggagacag acaacaatct gatgcagatc ctgtcaaaaa agttcacctt tatggacaac 2280

gtcgaggact tcaacagcgg caaggtcggg aaaatcgaca agattactta cgatagcacc 2340

gtgaaggaaa tgttcctgtc ccctgagaac aaaagggccg tctggcagac cattcaggtg 2400

gctgaggaga tcaagaaagt gatgggctgc gagccaaaaa agatctttat tgaaatggca 2460

cggggcgggg agaaggtgaa aaagaggaca aaatctcgca aggcccagct gctggagctg 2520

tacgccgctt gcgaggaaga ttgtagagaa ctgatcaagg agattgagga ccgggacgag 2580

agggacttca atagcatgaa gctgtttctg tactataccc agttcgggaa atgtatgtat 2640

tccggcgacg acatcgatat taacgagctg attcgcggca attctaagtg ggaccgagat 2700

cacatctacc cccagagcaa aattaaggac gattccatcg ataacctggt gctggtcaat 2760

aagacatata atgccaaaaa gtccaatgag ctgctgtctg aggacatcca gaaaaagatg 2820

cattcattct ggctgagcct gctgaacaaa aagctgatca ctaaaagcaa gtacgaccgc 2880

ctgactcgaa agggcgactt taccgatgag gaactgagtg ggttcatcgc tagacagctg 2940

gtggaaacaa ggcagtcaac taaggcaatc gccgatatct tcaagcagat ctacagctcc 3000

gaggtggtct atgtgaagag cagcctggtg agcgacttca ggaaaaagcc actgaactac 3060

ctgaagtctc ggagagtcaa tgattaccac catgcaaaag acgcctatct gaacattgtg 3120

gtcgggaacg tgtacaacaa aaagtttacc agtaatccca tccagtggat gaaaaagaat 3180

cgcgatacaa actatagcct gaacaaggtg ttcgaacacg acgtggtcat taacggagaa 3240

gtgatctggg aaaagtgcac ataccatgag gacactaata cctatgatgg aggcactctg 3300

gaccgaatcc ggaagattgt ggaacgcgat aacattctgt acaccgagta cgcttattgt 3360

gagaagggcg aactgtttaa tgcaaccatc cagaacaaaa atggaaactc cacagtctct 3420

ctgaaaaagg gcctggacgt gaaaaagtac gggggatact tcagcgccaa cacaagttac 3480

ttctcactga tcgagtttga ggacaagaag ggggatagag caaggcacat cattggagtg 3540

cctatctata ttgcaaacat gctggagcat tctccaagtg ccttcctgga gtactgcgaa 3600

cagaaggggt atcagaatgt gcggattctg gtcgagaaaa tcaaaaagaa cagcctgctg 3660

atcattaatg gataccctct gcgcattcga ggcgagaacg aagtggatac ttcctttaag 3720

agggccatcc agctgaagct ggaccagaaa aactatgagc tggtccgcaa tatcgagaag 3780

ttcctggaaa aatacgtgga gaaaaaggga aactatccaa ttgacgagaa tagagatcac 3840

atcacacatg aaaagatgaa ccagctgtac gaggtgctgc tgtccaaaat gaaaaagttc 3900

aacaagaagg gcatggccga cccctctgat aggatcgaaa agagtaagcc taaattcatc 3960

aagctggagg acctgatcga taagattaat gtgatcaaca aaatgctgaa cctgctgcgc 4020

tgtgacaatg atactaaggc cgacctgtct ctgattgagc tgcccaaaaa cgctgggagt 4080

ttcgtggtca aaaagaatac catcggaaag tcaaaaatca tcctggtgaa tcagagcgtg 4140

actggactgt acgagaatag acgggaactg taagaattc 4179

<210> 122
<211> 4086
<212> ADN
<213> Secuencia Artificial

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

<400> 122
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatggg gaggaaacct tacattctgt ctctggatat tggaactggg 120

tccgtcggct acgcttgcat ggataaagga ttcaacgtgc tgaagtacca cgacaaagat 180

gccctgggag tgtatctgtt cgacggcgct ctgactgcac aggagcggag acagtttagg 240

```
acctccaggc gccgaaagaa ccggagaatc aaacgcctgg gcctgctgca ggaactgctg      300
gcacccctgg tgcagaaccc taatttctac cagtttcagc ggcagttcgc ctggaagaac      360
gacaatatgg attttaagaa caagagcctg tctgaggtgc tgagcttcct gggatatgaa      420
tccaagaaat accctaccat ctaccacctg caggaggctc tgctgctgaa agacgagaag      480
tttgatccag aactgatcta catggcactg tatcatctgg tgaaatacag aggccacttt      540
ctgttcgatc atctgaagat cgagaacctg actaacaatg acaatatgca cgatttcgtg      600
gagctgattg aaacctatga gaacctgaac aatatcaagc tgaatctgga ctacgagaaa      660
accaaagtga tctatgagat tctgaaagac aacgaaatga ctaagaatga tagagccaaa      720
agggtcaaga acatggagaa gaaactggaa cagttctcta tcatgctgct ggggctgaag      780
ttcaatgagg gaaaactgtt taaccacgcc gataatgctg aggaactgaa gggggctaac      840
cagagccata catttgcaga caactacgag gaaaatctga ctcccttcct gaccgtggaa      900
cagtcagagt ttattgaaag ggccaacaaa atctatctga gcctgactct gcaggatatc      960
ctgaagggca agaaatcaat ggctatgagc aaagtggccg cttacgacaa gttcagaaat     1020
gagctgaaac aggtgaagga cattgtctat aaggctgatt ctaccaggac acagttcaag     1080
aaaatctttg tgagctccaa gaaaagtctg aagcagtacg acgcaactcc caacgatcag     1140
accttctcta gtctgtgcct gtttgaccag tacctgattc gcccaaagaa acagtatagc     1200
ctgctgatca aggagctgaa gaaaatcatt ccccaggact ccgaactgta ctttgaggca     1260
gaaaatgata ccctgctgaa ggtgctgaac accacagaca atgctagcat ccctatgcag     1320
attaacctgt acgaggcaga aaccatcctg cgaaatcagc agaaatatca cgccgagatc     1380
acagatgaga tgattgaaaa ggtgctgtct ctgatccagt tccgcattcc atactatgtg     1440
gggcccctgg tcaacgacca tacagccagt aagtttggat ggatggagcg caaaagtaac     1500
gaatcaatca agccttggaa tttcgacgag gtggtcgatc gaagtaaatc agccactcag     1560
tttattaggc gcatgaccaa caagtgttcc tacctgatca atgaggatgt gctgccaaaa     1620
aactctctgc tgtatcagga gatggaagtc ctgaacgaac tgaatgccac acagatcagg     1680
ctgcagactg acccaaaaaa ccgcaagtac cgaatgatgc cccagattaa gctgttcgct     1740
gtggagcaca tctttaagaa atataaaacc gtcagccatt ccaagttcct ggaaattatg     1800
ctgaacagca atcacaggga gaactttatg aatcatggag aaaagctgag tatcttcggc     1860
acacaggacg ataagaaatt tgcatcaaag ctgtcaagct accaggacat gactaaaatc     1920
ttcggggata ttgagggaaa gcgcgcccag attgaggaaa tcattcagtg gatcaccatt     1980
tttgaggaca agaaaatcct ggtgcagaag ctgaaagagt gctatcctga actgacatcc     2040
aagcagatca accagctgaa gaaactgaat tactctggct gggggaggct gagtgagaag     2100
```

ctgctgactc acgcctatca gggccatagc atcattgaac tgctgcgcca ctccgatgag 2160

aatttcatgg aaattctgac caacgacgtg tacgggttcc agaattttat caaagaggaa 2220

aaccaggtcc agagcaataa gatccagcat caggatattg ccaacctgac tacctctccc 2280

gctctgaaga aaggcatctg gagtacaatt aagctggtgc gggagctgac ttccattttc 2340

ggggagcctg aaaagatcat tatggagttt gctaccgagg accagcagaa aggcaagaaa 2400

cagaaatcaa gaaagcagct gtgggacgat aacatcaaga aaaataagct gaaaagcgtg 2460

gacgagtaca aatatatcat tgatgtcgcc aataagctga acaatgagca gctgcagcag 2520

gaaaaactgt ggctgtacct gagccagaac ggcaagtgta tgtatagcgg gcagtccatc 2580

gacctggatg ccctgctgtc ccccaatgct accaagcact acgaggtgga tcatattttc 2640

cctcggagct tcatcaagga cgatagcatt gacaacaagg tgctggtcat caagaaaatg 2700

aatcagacaa agggcgatca ggtgcccctg cagttcattc agcagcctta cgagagaatc 2760

gcatattgga agagcctgaa caaagccggg ctgatctctg atagtaaact gcacaagctg 2820

atgaaaccag agttcacagc tatggacaag gaaggcttca tccagcggca gctggtggag 2880

actagacaga tcagcgtgca tgtccgggat tttctgaaag aggaataccc taataccaaa 2940

gtgatcccaa tgaaggccaa aatggtgagc gagttccgga agaaatttga catcccaaag 3000

attagacaga tgaacgacgc acaccatgcc atcgatgctt acctgaatgg cgtggtctat 3060

cacggggcac agctggccta ccccaacgtg gacctgtttg atttcaattt taagtgggag 3120

aaagtccgag aaaagtggaa agccctggga gagttcaaca caaagcagaa atctcgggaa 3180

ctgttctttt tcaagaaact ggagaagatg gaagtgtccc agggcgagcg gctgatctct 3240

aagatcaagc tggacatgaa ccacttcaag atcaactact ccagaaagct ggccaacatc 3300

cctcagcagt tttataatca gaccgcagtg tctccaaaga cagccgagct gaaatacgaa 3360

tctaacaaga gtaatgaggt ggtctataag ggactgacac cataccagac ttatgtggtc 3420

gccatcaaga gcgtgaacaa gaaaggcaag gagaaaatgg aataccagat gatcgaccac 3480

tacgtgttcg atttttataa attccagaac ggcaatgaga aggaactggc tctgtacctg 3540

gcacagaggg agaacaagga cgaagtgctg gatgctcaga ttgtctatag tctgaataag 3600

ggggatctgc tgtacatcaa caatcatccc tgctatttcg tgtcacgcaa agaggtcatc 3660

aacgcaaagc agtttgagct gaccgtggaa cagcagctgt ctctgtacaa cgtgatgaac 3720

aacaaggaga caaatgtcga aaagctgctg atcgagtatg acttcattgc cgagaaagtg 3780

atcaacgaat accaccatta tctgaatagc aagctgaaag aaaagcgagt ccggaccttt 3840

ttctcagaga gcaaccagac acacgaggac ttcatcaagg ccctggacga gctgtttaag 3900

gtggtcaccg catccgccac aaggtctgat aaaatcggga gtcgcaagaa cagcatgact 3960

catcgagcct tcctgggaaa aggcaaggac gtgaagattg cttacacctc catctctgga 4020

ctgaaaacaa ctaaacctaa gagtctgttt aagctggccg agtcaagaaa cgaactgtaa 4080

gaattc 4086

<210> 123
<211> 4209
<212> ADN
<213> Secuencia Artificial

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

<400> 123

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc      60
aaggtcgaag cgtccatgac aaaaatcaaa gacgactaca tcgtgggact ggacatcggc     120
acagactcct gcgggtgggt ggctatgaac agcaataatg acattctgaa actgcagggc     180
aagaccgcaa tcgggtcacg cctgttcgag ggagggaaga gcgcagctga acggagactg     240
tttcgcacca cacacaggcg catcaaacga cggagatggc gactgaagct gctggaggag     300
ttcttcgacc cctacatggc agaggtggat ccttatttct ttgcccggct gaaggaatct     360
ggcctgagtc cactggacaa aagaaagacc gtgagctcca ttgtgttccc cacatccgcc     420
gaggataaga agttctacga cgattaccct acaatctacc atctgaggta taaactgatg     480
actgaggacg aaaagttcga tctgcgcgaa gtgtacctgg ctatccacca tatcattaag     540
taccgaggaa acttcctgta taataccagt gtgaaagact tcaaggcatc aaagatcgat     600
gtcaaatcta gtatcgagaa gctgaacgag ctgtatgaaa atctgggcct ggacctgaac     660
gtggagttca acattagcaa tactgccgag atcgaaaagg tgctgaaaga caagcagatc     720
ttcaagcggg ataaagtcaa gaaaattgcc gagctgtttg ctatcaaaac cgacaacaag     780
gaacagagca agagaatcaa agatatttcc aaacaggtgg ccaatgctgt cctggggtac     840
aagaccaggt tcgacacaat cgctctgaaa gagatttcca aggacgaact gtctgattgg     900
aacttcaaac tgtcagacat cgatgcagac agcaagtttg aggccctgat gggaaacctg     960
gatgagaatg aacaggccat cctgctgact attaaggagc tgtttaacga agtgaccctg    1020
aatggaattg tcgaggacgg caacaccctg agcgaatcca tgatcaacaa gtacaatgat    1080
caccgggacg atctgaagct gctgaaagaa gtgatcgaaa atcatattga cagaaagaaa    1140
gccaaggagc tggcactggc ctacgatctg tatgtcaaca ataggcacgg acagctgctg    1200
caggctaaga aaaagctggg caaaatcaag ccccgctcta aggaggactt ctacaaagtg    1260
gtcaacaaga atctggacga ttcacgggca agcaaggaga tcaaaaagaa aattgaactg    1320
gacagcttta tgcctaagca gagaaccaac gccaatggcg tgatcccata ccagctgcag    1380
```

cagctggagc tggataagat catcgaaaac cagtctaagt actatccatt cctgaaggag 1440

attaatcccg tgtcaagcca cctgaaagag gccccctata agctggacga actgatccga 1500

tttcgggtgc cttactatgt cggcccccctg atttctccta acgagagtac caaggatatc 1560

cagacaaaga aaaaccagaa tttcgcctgg atgattcgca aagaggaagg gcgaatcaca 1620

ccttggaact ttgaccagaa ggtggatcga attgagagcg ccaataagtt catcaaacgg 1680

atgactacca aggacactta cctgtttggg gaggatgtgc tgccagctaa cagcctgctg 1740

tatcagaagt tcaccgtcct gaacgaactg aacaacatcc ggattaatgg aaaaagaatc 1800

tccgtggacc tgaagcagga gatctacgaa aacctgttta agaaacacac aactgtgacc 1860

gtcaagaaac tggagaatta tctgaaggaa aaccataatc tggtgaaagt cgagatcaag 1920

gggctggccg atgaaaagaa attcaacagc ggactgacca catacaatag attcaagaac 1980

ctgaacatct ttgacaacca gattgacgat ctgaagtaca ggaacgattt cgagaagatc 2040

atcgaatggt ctacaatttt tgaggacaag agtatctaca aagaaaagct gaggagcatc 2100

gattggctga acgagaagca gattaacgct ctgtctaata tcagactgca ggggtgggga 2160

aggctgagta agaaactgct ggcacagctg cacgaccata atggccagac catcattgag 2220

cagctgtggg attcccagaa caatttcatg cagattgtga cacaggccga ctttaaagat 2280

gctatcgcaa aggccaacca gaatctgctg gtggctacct cagtcgagga cattctgaac 2340

aatgcataca caagccccgc aaacaagaaa gccatcagac aggtcatcaa ggtggtcgac 2400

gatatcgtga aggcagcctc cggaaaggtc ccaaaacaga tcgccattga gttcactagg 2460

gatgctgacg aaaatcccaa gagaagtcag accaggggct caaagctgca gaaagtgtac 2520

aaggacctga gcactgagct ggcctccaag accattgctg aggaactgaa cgaagcaatc 2580

aaagacaaga aactggtgca ggataagtac tatctgtact ttatgcagct ggggcgggac 2640

gcctatacag gagagcctat caatatcgat gaaatccaga agtacgatat cgaccacatt 2700

ctgccacagt ctttcatcaa ggacgatgcc ctggacaaca gggtgctggt gagccgggct 2760

gtgaacaatg gcaaatctga taatgtgcct gtcaagctgt ttggcaacga gatggctgca 2820

aatctgggga tgactatcag gaaaatgtgg gaggaatgga agaacatcgg cctgattagc 2880

aaaacaaagt acaacaatct gctgactgat cccgaccaca ttaacaagta taagagtgcc 2940

gggttcatca ggcgccagct ggtggagaca tcacagatca tcaagctggt gagcactatc 3000

ctgcagagtc gctaccctaa cactgaaatc attaccgtga aggctaagta caatcattat 3060

ctgcgggaga aatttgacct gtataagagc agagaagtca acgactacca ccatgctatt 3120

gatgcatatc tgtccgccat ctgcggaaat ctgctgtacc agaactatcc aaatctgcgg 3180

cccttctttg tgtacggcca gtataagaaa ttctcctctg atcctgacaa agagaaggcc 3240

atttttaaca aaacccgcaa gttctccttt atctctcagc tgctgaaaaa caagagtgag 3300

```
aacagcaagg aaatcgctaa gaaactgaaa cgggcatacc agttcaagta tatgctggtg     3360

tctcgagaga ctgaaacccg ggaccaggag atgttcaaaa tgaccgtgta cccccggttc     3420

agccacgata cagtcaaggc tcctaggaac ctgattccaa agaaaatggg catgtcccct     3480

gacatctacg gaggctatac aaacaattct gacgcataca tggtcatcgt ccgcattgat     3540

aagaaaaagg gaactgagta taagatcctg ggcattccaa cccgggaact ggtgaatctg     3600

aaaaaggccg agaaggagga ccattacaaa agctatctga aggagatcct gacaccaagg     3660

attctgtaca acaaaaatgg gaagcgcgat aaaaagatca cttccttcga aattgtgaaa     3720

tctaagatcc cctataagca ggtcatccag gatggggaca aaaagtttat gctgggaagt     3780

tcaacatacg tgtataacgc aaagcagctg acactgagca ctgagtccat gaaagccatc     3840

actaacaatt tcgataagga cagcgatgag aacgacgctc tgattaaggc atacgatgaa     3900

atcctggaca aagtggataa gtatctgcca ctgttcgaca tcaacaagtt ccgggagaag     3960

ctgcacagtg ggcgagaaaa gttcatcaag ctgagcctgg aggacaaaaa ggataccatc     4020

ctgaaagtgc tggaaggact gcatgataac gctgtcatga caaagatccc tactattggc     4080

ctgtccacac cactggggtt catgcagttt cccaacggcg tgattctgag cgagaatgcc     4140

aaactgatct accagtcccc caccgggctg ttcaaaaagt cagtgaagat cagcgacctg     4200

taagaattc                                                             4209
```

<210> 124
<211> 3405
<212> ADN
<213> Secuencia Artificial

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

<400> 124

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatggg ctacactgtg ggactggata ttggggtggc ttccgtcggg      120

gtggctgtgc tggatgagaa tgacaacatc gtggaggctg tgtcaaacat ctttgatgaa      180

gccgacacaa gcaacaataa ggtgcggaga actctgaggg agggcaggcg cacaaagcgg      240

cggcagaaaa cccgcattga ggacttcaag cagctgtggg agacttcagg ctacatcatt      300

cctcacaagc tgcatctgaa tatcattgag ctgcgcaaca aagggctgac cgaactgctg      360

agcctggatg agctgtattg cgtgctgctg tccatgctga agcaccgggg gatctcctac      420

ctggaggacg ccgacgatgg cgagaagggg aatgcctata agaaaggact ggcttttaac      480

gaaaaacagc tgaaggagaa aatgccatgt gagatccagc tggaacgcat gaagaaatac      540
```

gggaagtacc atggagagtt catcatcgaa attaatgatg agaaggaata ccagagcaac 600

gtgttcacca caaaggctta taagaaggag ctggaaaaga tcttcgagac acagcggtgc 660

aacggcaaca agatcaacac aaagttcatt aagaaataca tggagatcta cgaacgaaag 720

cgggaatact atatcggacc aggcaatgag aaaagcagaa cagactacgg catctatact 780

accaggactg atgaggaagg gaatttcatc gacgagaaga acatttttgg caaactgatc 840

gggaagtgta gtgtgtaccc cgaggaatat agagcaagct ccgcctcata caccgcccag 900

gagttcaatc tgctgaacga tctgaacaat ctgaaaatca acaatgagaa gctgacagaa 960

tttcagaaga aagagattgt cgaaatcatt aaggacgctt ctagtgtgaa catgaggaaa 1020

atcattaaga aagtcatcga tgaggacatt gaacagtaca gcggagcacg aatcgataag 1080

aaaggcaagg aaatctacca caccttcgag atctatcgga agctgaagaa agagctgaaa 1140

acaatcaatg tggatatcga ctcttttact agagaggaac tggataagac catggacatc 1200

ctgaccctga acacagagag ggaaagtatt gtgaaggcct tcgacgaaca gaaatttgtc 1260

tacgaggaaa atctgatcaa gaaactgatt gagtttcgga agaacaatca gagactgttc 1320

agcggctggc atagtttttc atacaaggct atgctgcagc tgatcccagt gatgtacaag 1380

gagcccaaag aacagatgca gctgctgacc gaaatgaacg tgttcaaaag taagaaagag 1440

aagtacgtca actacaagta catcccagag aacgaagtgg tcaaggagat ctataacccc 1500

gtggtcgtga agagcattag aacaactgtg aaaattctga atgcactgat caagaaatac 1560

gggtatcctg aatccgtcgt gatcgagatg ccaagggata agaactctga cgatgagaag 1620

gaaaagatcg acatgaacca gaagaaaaac caggaggaat acgagaaaat cctgaacaag 1680

atctacgatg agaagggaat cgaaattacc aacaaggact acaagaaaca gaagaaactg 1740

gtgctgaagc tgaaactgtg gaacgagcag gaaggactgt gcctgtattc cggcaagaaa 1800

atcgctattg aggatctgct gaatcacccc gagttctttg aaattgacca tatcattcct 1860

aagagcatct ccctggacga ttctcgcagt aacaaggtcc tggtgtacaa aacagaaaat 1920

tctatcaagg agaacgatac cccctaccac tatctgacac ggattaacgg aaagtggggc 1980

tttgacgaat ataaagctaa tgtgctggag ctgagaaggc gcggcaagat cgacgataag 2040

aaagtgaaca atctgctgtg catggaggat atcactaaga ttgacgtcgt gaaagggttc 2100

attaaccgca atctgaacga caccagatac gcatccaggg tggtgctgaa cgaaatgcag 2160

tccttctttg agtctcgaaa gtactgtaat actaaggtca aagtgatccg aggctctctg 2220

acctatcaga tgcggcagga tctgcacctg aagaaaaaca gagaggaatc atacagccac 2280

catgctgtgg acgcaatgct gatcgcattc tcccagaagg ggtacgaggc ctataggaag 2340

atccagaaag attgctacga ctttgagaca ggcgaaattc tggacaagga aaaatggaat 2400

aagtacattg acgatgacga gtttgatgac atcctgtata aagagaggat gaacgaaatc 2460

cgcaagaaaa tcattgaggc cgaggaaaag gtgaagtaca actacaagat cgataagaag 2520

tgcaatcgcg ggctgtgtaa ccagactatc tacgggaccc gagaaaagga cggaaaaatc 2580

cacaagattt caagctacaa catctatgat gacaaggagt gtaattccct gaagaaaatg 2640

attaacagtg ggaaaggatc agatctgctg atgtacaaca atgatcctaa gacatatcgc 2700

gacatgctga aaatcctgga aacttactcc tctgagaaga atccattcgt ggcatataac 2760

aaagagacag gagactactt tcggaaatat tctaagaatc acaacggacc caaggtcgag 2820

aaggtgaaat actatagcgg ccagatcaac tcctgcatcg atatttctca caagtacggc 2880

catgccaaaa atagtaagaa agtcgtgctg gtgtcactga acccttatag aaccgacgtc 2940

tactatgata atgacacagg caagtactat ctggtcgggg tgaagtacaa tcatatcaaa 3000

tgtgtcggaa acaagtacgt gattgatagc gagacatata acgaactgct gaggaaggag 3060

ggcgtgctga acagcgatga gaacctggag gacctgaaca gcaaaaacat cacttacaag 3120

ttcagtctgt acaagaacga catcatccag tacgagaagg gcggggaata ctatacagag 3180

cgctttctga gccgaatcaa agaacagaag aacctgattg agactaaacc catcaataag 3240

cctaacttcc agcggaagaa taagaaaggc gagtgggaaa ataccagaaa ccagatcgcc 3300

ctggctaaga ctaaatacgt ggggaagctg gtcaccgatg tgctgggaaa ctgttacatc 3360

gtgaacatgg agaagttctc cctggtcgtg gacaaataag aattc 3405

<210> 125
<211> 3474
<212> ADN
<213> Secuencia Artificial

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

<400> 125
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgac taacggcaag attctggggc tggacattgg catcgcaagc 120

gtgggggtgg ggattattga ggcaaaaact ggaaaggtgg tgcatgccaa ttcccggctg 180

ttctctgccg ctaacgctga gaacaatgca gaacggagag ggtttagggg atctaggcgc 240

ctgaatcgac ggaagaaaca ccgcgtgaag cgagtccggg atctgttcga gaaatacgga 300

atcgtcaccg actttcgcaa cctgaatctg aacccttatg agctgcgagt gaagggcctg 360

accgaacagc tgaaaaacga ggaactgttc gcagccctga gaacaatctc taagagaagg 420

gggattagtt acctggacga tgccgaggac gatagtaccg gatcaacaga ctatgctaag 480

agcatcgatg agaatcgccg actgctgaaa aacaagacac caggccagat tcagctggag 540

aggctggaaa agtacggcca gctgcgcggg aatttcaccg tctatgacga gaacggggaa 600

gcccatcgcc tgatcaatgt gtttagtaca tcagattacg agaaagaagc acggaagatc 660

ctggagacac aggccgacta caacaagaaa atcacagctg agttcattga cgattatgtg 720

gaaatcctga cccagaaacg aaagtactat cacggccccg ggaacgaaaa gagccggact 780

gactacggac ggttccggac cgatgggacc acactggaga atattttcgg aatcctgatt 840

ggcaagtgca acttttaccc tgatgaatat cgagcaagca aggccagcta caccgcacag 900

gagtataatt tcctgaacga cctgaacaat ctgaaggtga gcaccgaaac agggaagctg 960

tcaacagagc agaaagaaag cctggtggag tttgccaaga atactgctac cctgggaccc 1020

gctaaactgc tgaaggagat cgcaaaaatt ctggactgta aggtggatga gatcaaagga 1080

tacagagagg acgataaagg caagccagat ctgcatacct tcgagcccta taggaaactg 1140

aagtttaatc tggaaagcat caacattgac gatctgtccc gcgaagtgat cgacaagctg 1200

gctgatattc tgactctgaa caccgagaga gaaggaatcg aggacgcaat taagaggaat 1260

ctgccaaacc agttcacaga ggaacagatc agcgagatca tcaaggtgcg gaagagccag 1320

tccactgctt tcaataaggg ctggcactct tttagtgcaa aactgatgaa cgagctgatc 1380

cccgaactgt acgccacctc cgacgagcag atgacaattc tgactcggct ggaaaaattc 1440

aaggtcaata agaaaagctc caaaaacaca aagactatcg acgagaagga agtcactgat 1500

gagatctaca atcctgtggt cgccaagagc gtgagacaga ccatcaaaat cattaacgct 1560

gcagtcaaga aatatggcga cttcgataag atcgtgattg aaatgccacg ggataaaaat 1620

gctgacgatg agaagaagtt catcgacaag agaaataagg agaacaagaa ggaaaaggac 1680

gatgccctga aaagggccgc ttacctgtat aattctagtg acaagctgcc cgatgaggtg 1740

ttccacggca acaagcagct ggaaaccaaa atccgactgt ggtatcagca gggggagcgg 1800

tgcctgtata gtggaaagcc catctcaatt caggagctgg tgcataactc taacaatttc 1860

gaaatcgatc acattctgcc tctgtcactg agctttgacg atagtctggc caataaggtg 1920

ctggtctacg cttggacaaa ccaggagaaa ggccagaaaa ccccttatca ggtcatcgac 1980

tccatggatg cagcctggtc tttcagggag atgaaggact acgtgctgaa acagaaggga 2040

ctgggcaaga aaaagcgcga ctatctgctg actaccgaga acatcgataa gattgaagtg 2100

aagaagaagt tcatcgagag gaatctggtg gatactcgct acgcatctcg agtggtcctg 2160

aactctctgc agagtgccct gagagagctg gggaaagaca ctaaggtgtc tgtggtcagg 2220

ggacagttca ccagtcagct gcggagaaaa tggaagatcg ataagagccg cgagacatac 2280

caccatcacg cagtggacgc cctgatcatt gctgcatcaa gccagctgaa actgtgggag 2340

aagcaggaca atcccatgtt tgtggattat ggcaagaacc aggtggtcga caaacagact 2400

ggggagatcc tgtccgtgtc tgacgatgag tacaaggaac tggtgttcca gccccttat 2460

cagggctttg tgaataccat ctcctctaaa gggttcgagg acgaaattct gtttagctac 2520

caggtggatt ccaaatataa ccggaaggtc agtgacgcaa ccatctactc aacaagaaaa 2580

gccaagattg gcaaggataa gaaagaggaa acctacgtgc tgggaaaaat caaggacatc 2640

tactcccaga atggcttcga taccttcatc aagaagtaca acaaagataa gactcagttc 2700

ctgatgtatc agaaggactc tctgacatgg gagaacgtga tcgaagtcat tctgagggac 2760

tacccaacaa ctaagaaaag cgaggacggc aaaaatgatg tgaagtgcaa ccccttgag 2820

gaatacaggc gcgagaatgg gctgatctgt aagtattcca agaaagggaa aggaactccc 2880

atcaagagcc tgaagtacta tgacaagaaa ctggggaact gcatcgatat taccccagag 2940

gaatcacgca ataaggtcat cctgcagagc attaaccctt ggcgagccga cgtgtacttc 3000

aatccagaga cactgaagta cgaactgatg ggcctgaaat attcagatct gagctttgaa 3060

aagggcactg ggaactacca tatcagccag gagaaatatg acgctatcaa agagaaggaa 3120

ggaattggca agaaatccga gttcaagttt acactgtacc gcaacgacct gatcctgatc 3180

aaggatatcg ccagtggcga gcaggaaatc tacagattcc tgtcaagaac tatgcccaat 3240

gtgaaccact acgtcgagct gaagccttac gacaaggaaa agttcgataa cgtgcaggag 3300

ctggtcgaag cactgggaga ggcagataaa gtgggacgat gtatcaaagg actgaataag 3360

ccaaacatca gcatctacaa ggtgagaacc gacgtcctgg gaaacaaata tttcgtgaag 3420

aaaaagggcg acaaacccaa gctggatttt aagaacaaca agaagtaaga attc 3474

<210> 126
<211> 3462
<212> ADN
<213> Secuencia Artificial

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

<400> 126
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatggc cgaccgaatc tctctggggc tggacattgg ggtggcaagc 120

gtggggttct cagtgctgga cattgataag ggaaaagtca ttgagctggg cgccaggctg 180

ttctctgcta ctgtggccgc tggcaaccag gatcgaagag acatgcgagg agccaggcgc 240

ctgctgaacc ggaacaagca gcgacggcag gataccggaa agctgttcaa gaaatttggc 300

ctgatcgacg attttgataa gggcagcttc tacgacaact ttaatcagaa cctgaatcct 360

tatgagctga gagtgaaagg cctgacagaa cagctgacta aggaggaact ggccgagtct 420

ctgtaccaga tcgtgaaaca tagggggatt agttatgcac tgaaggacgc cgatgtggac 480

gaaggcggga cagactactc agtcagcctg aaaatcaaca gccaggagct ggcagaaaag 540

actccagccc agattcagct gcagagactg aatgattatg gaaaggtgag gggccaggtg 600

gtcatcggcg acgatccaga caaccagaag gtgctgctga atgtgttccc cacatcagct 660

tacgagaaag aagcaaagca gatcattgcc actcagcagc agttctatcc tgagagcctg 720

accgacaagt tcaccgagga atactgccag atcctgactc gcaagcgaga ttattttgtg 780

gggccaggaa acgagaaaag ccggaccgac tacgggatct acaagactga tggaagaacc 840

ctggacaatc tgttcgagga actgatcggc cacgataaga tctaccccga ggaactgcgg 900

gcatctgcag ccagttatac cgcccagctg tttaacgtgc tgaatgacct gaacaatctg 960

agaatcctga actacgagga tgggaaactg acaaaggagg acaaggaaaa gatcatcgct 1020

gaaattaaga acaacaccac aactatcaac atgctgaatg tgattaagaa agtcgccggg 1080

tgttccaagg acgatatcaa aggattccga gtgaatgaga aggataaacc cgaaatcagc 1140

tccatgcctg tgtaccgcaa aattcataag gacctgctga aggccggcgt ggatatctca 1200

gactggcccg tcgagttcat cgacgaactg agctttattc tgacactgaa cactgagaat 1260

ggggaaattc gcaaacagct gaacaatcga ctggccccta agttcgattt tctgaacgct 1320

gacctgatcc agctgatcat tgataataag gactcctttg agattaagac taacaacaag 1380

tggcacagat tcagcgtgaa aaccatgaac aaactgatcc cagagatgat ggaaagaccc 1440

gtggagcaga tgaccctgct gaatgaaatg ggactggtca agaaagataa gaaacgcttt 1500

gagaacaata agtacctgcc ttacaaggaa atcgcaaagg acattttcaa cccagtggcc 1560

tccaaatctg tccgcgaggc cctgaagatc gtgaatgctg tcctgaagaa atacggccac 1620

attgattatc tggtggtcga aatgcctcgg gataaaaacc tgaaggagga acaggacaat 1680

atcaaggagt tccagaacaa aaataagaaa gctaaggacg ctgcattcga agcatttgtg 1740

aaatcagtcg ggagcgagca gagagtgaag gaagccctgt ctaaaaaccg gaagctgcag 1800

atgaagatga gactgtggta tcagcagcag gagatcgatc catataatgg aaagacaatc 1860

gatgccactg acctgattaa caatcctgat aagttcgaga ttgaccatat cattccacag 1920

agtatctcat tttacgacag tattaacaat aagaccctgt gcttcgcctc aatgaaccag 1980

gtgaaaggac agaaaacccc ctacgagttt atgctggaag gccacgggca gtcctatgac 2040

aagttcaaag ctacagtgat ggcaaacaag aattttggca aggctaaaag ggcaaactac 2100

ctgttcgagg aaaatgtgag cgatatcgag actcggaaga gattcctgtc ccgcaacctg 2160

gtggacaccc gatattctag tcgggtggtg ctgaacagcc tgcaggattt ctttcgggag 2220

aaatctgccg acaccaaggt gacagtcatt cgcggcaagt ttacctccaa catgcgaaaa 2280

cattggcaca tcgataagac tagggagaca ttccaccatc acgccattga cgcttctatc 2340

attgccgcta caccatttct gcgcatgtgg aagaaaggag gcactatctt ccccgtgaag 2400

```
gtcggagaag aaagtatcga tattgagaca ggcgaaattc tggacgataa gaattttgac     2460

aaagcaatgt acgaggaacc ctatagtggc ttcgtgtcag agatcatgaa cgccgacgat     2520

cggatcaagt tcagccacca ggtggataag aaaatgaata ggaaggtgag cgacgccacc     2580

atctacagta ctcgcaccgg gaaactggct aaggataaga aagacgctga gtacatcgtg     2640

gcaaaggtca aagatatcta cagcgtggac ggattcaaga agttcaagaa agtctacgat     2700

aaggacaaaa ccaagtttct gctgtacaaa tatgatccta ggacattctc aaagctggag     2760

cgcatcatta gcgattgccc agacaaagtg gaaaaggtcc agacaaacgg caaagtgaag     2820

gctgtcgata tcagtccatt cgagatgtac agaagggacc atgggatgat caagaaatac     2880

tcaaagaaag gaaacggccc cgccatcaag cagctgaagt acctggataa gaaactgggc     2940

agccacatcg acattacccc cgcaaacgcc aatggaaaac acgtgatcct gcagagtctg     3000

aagccttgga gaaccgacgt ctatctgaac cacgagacag gcgagtacga aatcatgggg     3060

attaagtata gcgatctgaa gttcaacaag aatgaggggt acggaatcaa gaaagacaag     3120

tatctggaaa ttaagaaagt ggaggaagtc tccgagaagt ctgagttcat gtttagcctg     3180

tacaggaagg atcgcgtgaa agtccaggac atgaaaaccg gcgagtccgt ggaactgctg     3240

ttctggagca ggaacttttc caataagaaa tacgctgagc tgaagcccat ctcccaggca     3300

gaaaacgaca agaaactgcc tgtgtatggc aaagggagac tgatcaagag gctgattccc     3360

aaaaactgta agatctggaa agtgaatacc acaattctgg gcgatcccta ctatctggag     3420

aaagaaagcg actcccctaa ggatatcctg gactaagaat tc                        3462
```

<210> 127
<211> 4611
<212> ADN
<213> Secuencia Artificial

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

<400> 127

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgaa gaaaacactg ggactggacc tggggactaa cagcattggg      120

tgggccgtca tcaactcaaa catcgactca gaaggaaaag aaaagctggt ggggatcagc      180

tcctgcggaa gccggatcat tcctatggac gctaccacac tgggagattt cggaaagggc      240

aacactaaaa gtccagtggc agagcgaacc cgactgcggg gcattcggag actgctggaa      300

aggtcactgc tgaggcgcga gcgactgcac cgagtgctgt cagtcatggg gttcctgcca      360

gagcattacg ctagccagct ggaccgctat ggaaagtttc tgccagaaac cgagcccaag      420
```

ctggcatggt acaaagacga tagcggcagg tatcagttcc tgtttcagaa gtccttccac 480

gagatgctgg aggactttcg acagcatcag ccagaactgg tggcaggaga gaagaaaatc 540

ccttacgatt ggacaatcta ctatctgcgg aagaaagcac tgtctcagga gatcactaag 600

gaggaactgg cctggattct gctgaatttc aaccagaaac gcggctacta tcagctgcga 660

ggggaggaag agcaggaaga gaacaataag agcgtggagt atcacgccct gaaagtggtg 720

agcgtcgagg actctggaga aagaaagggc aaagatatct ggtacaatgt gactctggag 780

aacggatggg tctatcgacg ggctagcaac attcccctgg actggaccgg caaggtgaaa 840

gagttcgtgg tcactaccga gctggacgat gccgggaacc caaagaaaga taaggaagga 900

aatgtgaaaa gatcctttag ggcaccaaag gaggacgatt gggggctgct gaaaactagg 960

acccaggctc agatcgacga atccggcaag accgtgggaa cttacatcta cgagtctctg 1020

ctgtgtatgc ccaaccagaa gatccgggga aaactggtga ggaccattga acgcaagtac 1080

tataaagacg agctgagaca gatcctggtg aagcagagcg agttccacgc cgctctgcag 1140

gatcataatc tgctgctgtc ctgtatcgaa gagctgtacc ctaacaatga ggcccacaga 1200

aggctgctga gcgccagcag cttcatctac ttcctgatcg aggacattct gttttatcag 1260

cgcccactga agtcccagaa agggctgatc gataactgcc cctacgagtc tcacatctac 1320

aaggataaga aagatggaag tctgcaccat gtgcctctga agtgtgtcag caaatcccat 1380

ccactgttcc aggaatttcg cctgtggcag ttcctgtcta acctgcgaat ctaccagaga 1440

gagaggatga tggacggcag tctgaaactg gacgtggatg tcacccggga gttcctgccc 1500

tcagaagagg actacgtgaa gctgtttgat tggctgaatg agagaaagga aatctctcag 1560

aaattcctgc tggcttataa acctttttggg ctgaagaaaa acgaacaggc aaattacaga 1620

tggaactatg tggaggacaa gagctacccc tgcaacgaga cacgggcaga aatcaagagc 1680

agactgtcca aagccggagt gcctgaagag tttctgactg aagagaagga agaggccctg 1740

tggcacatcc tgtattctat tagtgataag aaagagctga ctaaggctct gggcaccttc 1800

gcagccaaaa actgtctgaa tgagtctttc gtggaagtct ttgccaagat ccccccttttt 1860

gagtcaaact acgctgcata tagcctgaag gctattagga aactgctggc actgatgcgc 1920

atggggaagt actggaatga acaggccatc gacaggcaga ctcgcgatcg aatcgagaaa 1980

attctgaccg gagagtatga cgaaacaatc cggagcagag tgagggagaa ggcaatgctg 2040

ctgaccgata ttagcagctt ccggggcctg cctctgtggc tggcctgtta catcgtgtat 2100

gaccgccact cagagagcaa ggaactggtc aaatgggaga caccagccga catcgatcat 2160

ttcctgtcca agtttaaaca gaacagcctg cggaacccca tcgtggaaca ggtcattaca 2220

gagtccctgc gcactgtgcg agacatttgg aagcaggagg gaaaaatcga tgagattcac 2280

gtggaactgg gccgggagat gaagaaccct gcaaaagagc gcgcccgaat tacagctcag 2340

gtgcaggaaa atgagaacac taatctgaga atcaaggctc tgctggcaga gttcatgaac 2400

cccgaatttg agattgaaaa tgtgcatcca tactcacccg gccagcagga aatcctgcgg 2460

atctacgagg acggcgtgct gagcgggatc gctgagaagg atctgcctga cgatatcaca 2520

gcaattctga agaaattccg agaaaacgac gtgaagaaac ggccaacaac tagcgaagtc 2580

ctgcggtaca aactgtggct ggagcagcgg tacagatccc catataccgg aagagtgatc 2640

cccctgggca agctgttcac acctgcttac gagatcgaac acgtgattcc ccagagccgg 2700

tattttgacg attccatctc taacaaagtg atctgcgaaa gtgccgtcaa taagctgaaa 2760

gataactgtc tgggctatga gttcatcaag aaacattccg gggagatggt ggaactgggg 2820

aatggagaga cagtgcccgt gttcagcgtg gaagagtacg aacggttcgt caaggagtct 2880

tacttcggca acagtaagaa aatgaagaaa ctgctgctgg aggacatccc agatagcttc 2940

attgagagac agctgaatga cagtcgatac atctcacggg tggtcacatc tctgctgagt 3000

aacctggtgt gcgaagaggg agagcaggat ggcctgtcca agaatgtgat cgtctgtacc 3060

ggcgggatta cagacaggct gaagaaagat tggggagtgc aggaagtctg gaaccgcatc 3120

attctgcctc ggttcctgag actgaatgag atcaccggac ggacagactt tacaagtact 3180

tcagtgaacg gccacctgct gcctgccctg ccactgtacc tgcagaaggg ctttaataag 3240

aaaagaattg accataggca ccatgccatg gatgctatcg tgattgcctg cgctaaccgg 3300

aatatcgtca actacctgaa caattcctct gctagaaaga acagcgaaat tagccgatat 3360

gacctgcagc ggctgctgtg tgagaaggtg aaaaccgatg ccaacggcaa ttacaagtgg 3420

atcctgagga aaccatggga gacattcacc caggatgtgt atgccgctct gacaaacatc 3480

gtggtcagct ttaagcagaa tctgcgcgtg attaatcgaa ccacaaacta ctatcagcac 3540

tacaacgagg caggggaaaa gcgcatgatc cctcagacca aaggcgacag atgggccatt 3600

agaaagccaa tgcataaaga tactgtgtat ggcgaagtca atctgagaaa ggagaaaacc 3660

ctgccactga gggacgtggt caagaacccc agtatcgtgg tcgataaatc actgaagaac 3720

aagctgtacg agctgctgaa gagccagtat gacctgaagg caatcgccaa atacttcgag 3780

acacaccagg acgtgtgggc agatgtcaac ctgaagaaaa ttaaggtgta ctacttcact 3840

aaggagacaa acgaaagatt ctttgccact aggaagagcc tggacccatc ctttgatcag 3900

aagaaaatcg aagaggaagt gactgatacc ggcgtccaga agattctgct gcaccatctg 3960

cagcagaaca ataacgaccc tgatatggcc ttttccccag acggcatcga taggatgaac 4020

cagaatatga ccattctgaa tgacgggaag tggcaccagc ccatctacaa agtgcgcaca 4080

tatgaaaagg cagataaatt cgccgtcggc gagtctggga acaaggccaa gaaatttgtg 4140

gaagcagcca aaggcaccaa tctgttcttt gccgtgtacg agagtgtcca ggaggacgaa 4200

gctatcggga agcaggtctg caaacggaca ttcgccacta tccccctgaa cgaagtgatc 4260

aagagaaaga aacagggcct gcccgctgca cctgaggacc tgaacgggaa tctgcccaag 4320

tttgtgctga gccctaacga tctggtctac ctgcccaccg aggaagagag gaatagttca 4380

cgcatcattc agcctctgga cagggagcgc atctataaga tggtgagctc ctctgggagt 4440

cagtgcttct ttatcaaagt gttcgtcgcc aattcaattt gggataagaa cgaatacagc 4500

agcctgaaca agatggagag ggctattaca aacgaaatga tcaaggagat ttgtgtgcct 4560

atcaaaattg accgcctggg caatgtcagc ctgatccaga tttaagaatt c 4611

<210> 128
<211> 4458
<212> ADN
<213> Secuencia Artificial

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

<400> 128
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgac aaagactatt ctgggactgg atctggggac taacagcatt 120

ggatgggcac tgattaacca ggacttcgat aacaaaaagg gggagatcct gggcatgggc 180

agccggatca ttcctatgac acaggatatt aaggacgagt tcggaaaagg caactctatc 240

agtcagacag ctgagcgaac tcgactgcgg ggagtccgga gactgatcca cagaactctg 300

ctgaggcgcg agcggctgca tagagtgctg aatatcattg gcttcctgcc agaacactat 360

gccaaccaga ttgactttga gaagaggttc ggcaagttca agcctgaatc tgagccaaaa 420

atcagtttcg atggaaatga cgtgttcctg tttgagaaga gctaccagga aatgctggtc 480

gactttaaaa tccaccagcc acagctggtg agcaacggca agaaaattcc ccatgactgg 540

acaatctact atctgagaaa gaaagctctg tctaagaaaa ttgagaagga ggaactggca 600

tggatcctgc tgaacttcaa tcagaaacgg ggatactatc agctgagagg cgaggaagag 660

gaagagacac caaataagct ggtggagttt catagcctga aaattgtgga ggtcaacgcc 720

gatgaacccc agaaaggaaa gcctgagatc tggtactcac tggtgctgga aaacggctgg 780

gtctatcgac gggctagcaa gactcccctg ttcgattgga aagacaagat tagggagttt 840

atcgtgacca cagaaatcaa caatgatgga actgtcaaaa ccgacaagga gggcaccgaa 900

aagaggagct tccgcgcacc aaaacccgag gactggactc tgcagaagaa aaagaccgag 960

ttcgagctgt caaagagcgg gaatgaagtg ggagccttca tctacgaaag tatcctgcag 1020

aaacccaacc agaagattag aggcaaactg atttcaacca tcgagaggaa attttataag 1080

gaagagctga aaaccatcct gaaaacacag ctgttctttc acaaagaact gaaggatgag 1140

aaactgtaca atgcctgcat cgaagagctg tataagaaca atgaagctca ccggagcctg 1200

ctgtccaaca aggggttcga gcatctgttt attaacgaca tcctgttcta ccagcgacct 1260

ctgcggtcta aaaagagtca gatctcaaac tgcccactgg agaagcgcac atataaaaag 1320

gaggggattg aaatcactga gggcatcaaa gtgatctcca aatctaatcc aatctaccag 1380

gagttccggc tgtggcagtg gattagcaac ctgtccctgt attgtatcga acccaccgag 1440

acaaatgtga cttcaacctt tctgaacagc attgaagatt acgagaatct gttcgaattt 1500

ctgaacaatc gcaaggaaat cgagcagaag cacctgctga aatatctgct ggagaaccag 1560

gggtttaaag gaaagctgct gacaaacgaa ctggagaagt tccgctggaa ttttgtcgct 1620

gacaaaaagt acccctgtaa tgagacaggc agcctgctgc atactcggct gagcaaagtg 1680

aagaacattt cccctgattt cctgaccaag gaaatcgagc accagctgtg gcatatcatc 1740

tacagcgtga ccgacaagat tgaatatgag caggccctga aaacctttgc tcggaaaaac 1800

aatctggatg tggactcctt ctttgagcac ttcaaaaaga tccccccttt tgagtctacc 1860

tacggagcat atagtgaaaa ggccatcaag aagctgctgc cactgatcag actgggcaaa 1920

tactggaact gggaggccat tgatagtatc tcaaaggaca ggattagtaa aatcctgtca 1980

ggggaatacg atgagaacat taagaacaga gtgagggaga aagcagtcca cctgacctcc 2040

gaaaacaatt tccagggact gcaggagtgg ctggccaagt acatcgtcta tgatcgccat 2100

tctgagggca atgacctggg gaagtggact agcgtgtccg acctggagac atacctgaag 2160

gagttcaagc agcatagcct gcggaaccct attgtggagc aggtcatcac agaaactctg 2220

agagtggtca aggatatttg gatcaagcac gggaaaggaa ccgaaaattt ctttgacgaa 2280

atccatgtgg agctgggccg ggaaatgaag aacaattccg aggatcgcaa acgactgacc 2340

aacacaatta ctgaaaacga gaatacaaac ctgagaatca aggccctgct gatggaaatg 2400

atgaatgata acgacgtgga gaacgtcagg ccttactctc caagtcagca ggagattctg 2460

aagatctatg aggacggagc tctgaatagc aacatcgagc tggacgatga aattgtgaag 2520

atctccaaaa aggcagagcc caccaaatct gaactgcagc gctacaagct gtggctggag 2580

cagaaatacc gatcccctta tactggccag gtcatcccac tgaacaagct gttcacctct 2640

gaatatgaga tcgaacacgt ggtccctcag agtcgcttct ttgacgatag cttcagcaac 2700

aaagtgatct gcgagtcagc cgtcaacaag cggaaggata accagctggg gctgcagttc 2760

atcaagaacc atagcggaga aaaagtggag ctgggcttcg ggaaggtggt ccaggtcttt 2820

acagaagagc agtacctgga ttttgtgaag gagcactata gcaaaaatcg ctccaagcat 2880

aacaaactgc tgctggaaga gattcccgag aagatgatcg aaaggcagct gaatgacact 2940

cgctacatca gtaagttcgt gagctccatt ctgtccaaca tcgtcagatc tgagaaggac 3000

gatgacggcc tgaatagcaa aaacattctg cctggaaatg gcaagatcac taccgaactg 3060

aaaagggact gggggctgaa tgatgtgtgg aacgacctga ttctgccaag attcgagagg 3120

atgaatacca tcacaaacag cgatctgttt acaacttaca acgacaagta tcagaaacac 3180

ctgcccaccg tgcctttcga gtactccaag ggctttcaga aaaagcgcat tgatcaccga 3240

caccatgcta tggacgcact ggtcatcgca tgtgccacac gggatcatct gaatctgatg 3300

aacaatcagt ctgccaagag tgaactgaaa cgatacgacc tgcagaacaa gctgcggaaa 3360

aaggagcctt acttcaacca gaaggagaac aaacagaagg aagccttcaa ggattttatc 3420

aaaccatggg gcactttcac cgaggacagc aagaatgctc tggaaaaaat cattatctcc 3480

tttaagcaga acctgagagt gatcaacaag gcaacaaact catacgagag ctataaggat 3540

gaggacggga atctgaggat cggcaaggat gggaaaccag agaagggcct gatcaagcag 3600

aaggggctga actgggcaat cagaaagccc ctgcacaaag acaccgtgtc aggccagatt 3660

aacctgagca ggatcaagct gcccaatggg aaaatcctga cagccactcg caagaatctg 3720

gataccagtt tcgacctgaa aacaattgag aactcaatca ccgacacagg cattcagaaa 3780

atcctgaaga attacctgct gtccaaggaa tctccagagc tggccttctc tcccgagggc 3840

ctggaagaga tgaacaatga gatcgaaaag tacaacaatg ggaaatttca ccatcctatt 3900

aacaaggcta ggatctatga actggggagc aaattcaatg tcggacagac aggcaacaga 3960

agggataagt ttgtggaggc cgctaaagga actaatctgt tctttgctat ctaccaggac 4020

gagaataaga accgctctta tgaaactatt cccctgaacg aagtgatcga gcaccagaag 4080

tggcgagcaa gtctgcctaa ggaagagcag gagaaaattc cactggtgcc cgtcaacaag 4140

ctgaagggga ccttcatctt ttccctgtct cccaacgatc tggtgtacgt cccttccaat 4200

gacgagctgg aacgaagtgc ttcaattgac ttctctaagc tgaaaaagga acagatcaac 4260

cggctgtata aaatggtgtc tagttcagga tcccagtgct tctttgtgaa gagtgaggtc 4320

gcaacctcag tggtcaacaa aatggaatac agcagcctga acaagatgga gaagtctatc 4380

gataacctga tggtgaagga gatttgtatc aaactgaaga ttgacaggct gggcaacatc 4440

agcaaggcct aagaattc 4458

<210> 129
<211> 4689
<212> ADN
<213> Secuencia Artificial

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

<400> 129
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgac caaaatcctg gggctggacc tggggactaa tagcatcggg 120

tgggcaatcc gcgacacaga aaatgagggc atcaatcaga tcctggacaa gggcgcccgc 180

attttcagcg agggagtgaa gtccgagaat ggcaaagaaa tcagtagggc agctgaacga 240

accgcttacc ggagcgcaag aaagatcaaa tatcggagaa aactgcggaa gtacgagaca 300

ctgaaggtgc tgtcaattaa cggaatgtgc cccctgagcg tggaggaagt cgaacagtgg 360

aagaaatctg gcttcaagga atatcccctg aatcctgagt ttctggattg gctgcggacc 420

aacgaggaca agaacatcaa tccttacctg ttcagagata aagctagcaa gcagaaagtg 480

acactgtttg aactgggaag agctctgtat cacatcgcac agaggcgcgg cttcctgagt 540

aacaggctgg accagtcagc cgagggcgtc tttgaggaac ataatcctca gatccagaac 600

ctgattgagg acctggacag ctccaacaca attctgaatg agctgaagga atactatatc 660

aatctgggga tcattgacga gactgaaaag tccggcttca agaaagacct ggatgagggg 720

gaaaagaaac tgaagtcact gtacaacagc ctggtggcca tcacaaagaa aaacgctaac 780

gatatcgaaa cttgcaagca ggagctgatc gctagactga ataagaaaga ggacctgggc 840

aaggtcaaag ggaagatcaa agatatttct caggcaatgc tggacaggaa ctttaagact 900

ctgggacagt atttcttttc actgtacaac aaggagcgaa tccggaacca gtataccagc 960

cgcgaggaac actacctgga ggagttcatc attatctgtc agacccaggg gatcgaggga 1020

attaacacaa atgagaagct gcctgagaag aagttcaccg gcctggcaaa agatctgtat 1080

cgagccattt tctttcagcg gccactgaaa tcccagaagg gcctgatcgg gaagtgctct 1140

ttcgagaaga acaagtctcg ctgtgccgtg agtcacccag actatgagga atttcgaatg 1200

tggagctacc tgaacacaat caaaattggc actcagtccg agaaaaccct gcgcttcctg 1260

acactggagg aaaagctgaa actggtgccc aagttctacc gaaagagtga tttcaacttc 1320

gaggtgctgg ctaaggagct ggtcgaaaaa ggagcatcat tcggctacta caagtctagt 1380

aagaaaaatg agttctttta ctggttcaac tataagccca ccgattcagt gagcgcctgc 1440

gtggtgagcg cttctctgga gaacgcaatc ggcaaggact ggaagatcaa gactttcgaa 1500

tatcagacta gaaacaccga gaagaatgaa gtgaccaaat ccgtcgatta caaggacctg 1560

tggcacctgc tgtccgtggc aacatctgat acttacctgt atgactttgc catcgaaaag 1620

ctgaaactgg agcctaaaaa cgcaaaggcc ttcagcaaga caaaactgaa gaaagacttt 1680

gccagtctgt cactggcagc catcgctaaa attctgccat atctgaagca gggcctgctg 1740

tactcccacg ccgtgttcat ggctaacatc gagaatattg tcgacgccga tatctggaag 1800

gatgaggaac agcagaagtt catccagtcc aagattgtgg agctggtcga caattatatc 1860

gtggaaaagt ctaaactgga gctgatcaat gggctgctga agatctacaa caccgaggat 1920

aaggaaggac ggaaagtgta ctattcaaag gaggctgaaa gcctgttcga ggcagacctg 1980

agaaagaaac tggtcccctt ctacaaggct aacatcatca tcgaggaaca cgagcaggaa 2040

atcattttcc aggatctgtt tcctatcctg atggaccagc tgaagaaaca ggagttcatc 2100

aaaattggca gactggataa gcagattgaa gccttcctgg agggggaaaa tgaggaagga 2160

cagatctttt gtaaccacac agataagctg aagaaactgt accatccaag cgacatcgag 2220

gtgtttaaga aaaagactat caaagatgag tgggggaatg aaaaggtggt cctgggatcc 2280

ccactgacca catctatcaa gaaccccatg gcaatgagag ccctgcacca gctgaggaag 2340

gtgctgaata cactgattgc caacgaccag atcgacgagg atactcggat ccatattgag 2400

atggccagag aactgaacga tgctaataag cgaaaaggca tccaggactt ccagaacgag 2460

aacaagaagt ttcgggagga agccatcaag gagatcaaga agctgtacct ggaggaatgc 2520

cacaaagacg tgaaccccac cgaggacgat atcctgaggt atcagctgtg gctggaacag 2580

ggaaagtgcg agatctacga ggaaggcaac aatatcagca tttgtgatat cattggcagc 2640

aatccctcct atgacatcga gcataccatt cctcggagca tctcccagga taacagccag 2700

atgaacaaga cactgtgtag cctgaagttc aacagatcca tcaaaaagca gaagatgcca 2760

gtggagctgt ccaactacaa tgacattctg cccaggatcg cacactggaa aaaggaggcc 2820

gaggaactga ctaggcagat cgaaaccatt tctcgcagca tcaagagcgc tgcaaccaag 2880

gtggccaagg ataagaacat cagaaagagg cattacctga cactgaagcg agactatatt 2940

ctggggaaat acgagcggtt cacttgggag gaacctaaag tgggctttaa aaactcccag 3000

atcccagaca ctgggatcat taccaagtac gctcaggcat atctgaaatc ttacttcaaa 3060

agggtggaga gtgtcaaggg aggagcagtg gctgagttcc ggaaactgtg gggcatccag 3120

gaatctttta tcgatgagaa ctggtggaag cactataagg acaaagatag agacaaacat 3180

acccaccata caatcgacgc aatcactatt gcctgtatgc ccaaggataa atacgacctg 3240

ctggcacacg cctggaggct ggaggatgaa caggacaaaa aggccgctaa ggtgctgatc 3300

gagcaggcca aaccatggaa aaccttcaag gaggatatcg aaaagattga gactgaaatc 3360

ctggtgagcc attttacccc cgacaacgtc aaaaagcagt caaagagcat catcaagaat 3420

cgcggcaaaa aggtgtacgt cctgaagaac gagctgcctg tgaacttcaa gaacaagatc 3480

gaagggaagg attatttcaa gctgaaattt gacagcaaga ttctgtacaa aatccccaaa 3540

aagaaagaga agcagaccga tacattctat gaggaactgc ctaaaaacta cctgaatgga 3600

gtggaaggca aggactactt caaaatcaat actaccggca agaccttcta caaaatccca 3660

atttttaacc agggcgacac aatccggggg agcctgcacc aggagacaac ttacggagcc 3720

attaagctgc cagatatcga cattgaaaca aagaaacccc tgcatactga taagggaggc 3780

ttcattctga agaaagacat caaaggcaat gagattgtgt tctttgtggt ccgcaaggaa 3840

atctctaaaa ttagtgagaa cgatgtgcag aatatcgtcg acaacgtggt ccggaagaaa 3900

attgagaatg ctatcgcaaa ctccctgatc acttttaaga tcgtgaagaa aaagaaagtg 3960

gccgtcatca aagagggggt caccatttgg atgagggagc ccaacattga aaaggggatc 4020

gagggaatcc ctattaagaa agtgcgcatc attaccaatt ctgtcaaaaa ccctatcgag 4080

attaaggtgc acagtccact gtccaagtct cgccacaagc ataaacagaa ggtctacggc 4140

cagaacgatg agaattatgc catggctctg tacgaactgg acgggaagcg ggagttcgaa 4200

ctgatcaaca attttaatct ggccaagctg ctgaaacaga gtcagtcata ctatccactg 4260

cataaagaga aggaaatcaa gggaaagaaa attctggtgc ccatcgagaa gagaaacaac 4320

aaggatgtca tcctgaagag gggccagcag gtggtgttct acgacaaaac cgtggagaac 4380

cccaaggata tctctgaaat cattgacttt cgcgagcgaa tctacatcat tgaagggctg 4440

accatccaga gacagaaaga taagaaaaca tccaaagtga atgagtacgg aatcattcag 4500

ctgcgccact tcaaggaagc tcgaaaaagt gaggaaatct caaaggataa cttcaaacct 4560

gacggcgagt tcaagatcaa cgagaataag ccaactagga aaatgaacca taatcagttc 4620

accgcctttg tggaggggat cgacttcaag gtcctgccta gcggaaagtt tcagaaaatt 4680

taagaattc 4689

<210> 130
<211> 3621
<212> ADN
<213> Secuencia Artificial

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

<400> 130
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgag caagaaagtc agtagacgct atgaagaaca ggcacaggag 120

atttgtcaga ggctggggag tagaccttat tccatcgggc tggacctggg agtgggatct 180

atcggagtgg cagtcgccgc ttacgaccca attaagaaac agccctccga tctggtgttc 240

gtcagctcca ggatttttat cccttctacc ggcgcagccg agcggagaca gaagagagga 300

cagaggaaca gcctgcgcca ccgagcaaat cgcctgaaat tcctgtggaa gctgctggct 360

gaacgaaacc tgatgctgtc atatagcgag caggacgtgc cagatcctgc acggctgaga 420

tttgaggacg ctgtggtccg agcaaaccca tacgagctgc ggctgaaggg cctgaatgaa 480

cagctgaccc tgagcgagct ggggtatgca ctgtaccaca tcgccaatca taggggatct 540

agttcagtgc gcacattcct ggatgaggaa aagagctccg acgataagaa actggaggaa 600

cagcaggcta tgacagaaca gctggcaaaa gagaagggaa tttccacttt catcgaagtg 660

ctgacagcct ttaacactaa tggcctgatc gggtacagga actccgagtc tgtgaagagt 720

aagggcgtgc cagtccccac tcgcgacatc atttcaaatg agattgatgt gctgctgcag 780

acccagaagc agttctatca ggaaatcctg tcagacgagt actgcgatcg gattgtcagc 840

gcaatcctgt ttgaaaacga gaagatcgtg ccagaagccg gctgctgtcc ctatttccct 900

gacgagaaga aactgcccag atgtcacttt ctgaatgagg aaaggcgcct gtgggaagcc 960

attaacaatg ctaggatcaa gatgcccatg caggagggcg ctgcaaaacg ctaccagagt 1020

gcttcattca gcgacgagca gagacacatt ctgtttcata tcgcaaggag cgggactgat 1080

atcaccccta aactggtgca gaaggagttc ccagccctga aaacctccat cattgtgctg 1140

cagggaaaag agaaggctat tcagaagatc gcaggcttcc gatttcgacg gctggaggaa 1200

aaatcttttt ggaagagact gagtgaggaa cagaaggacg atttctttag cgcctggaca 1260

aacactcctg acgataaaag actgtccaag tacctgatga aacacctgct gctgacagaa 1320

aatgaggtgg tcgacgccct gaaaaccgtg agcctgattg gagattatgg cccaatcggc 1380

aagaccgcaa cacagctgct gatgaaacat ctggaggatg gcctgactta caccgaagcc 1440

ctggagcggg gaatggaaac cggcgagttc caggaactgt cagtgtggga gcagcagagc 1500

ctgctgccct actatgggca gatcctgaca ggatctactc aggccctgat ggggaagtat 1560

tggcactctg cttttaaaga aaagagagac agtgagggat tctttaagcc taacacaaat 1620

agcgatgagg aaaaatacgg caggattgcc aacccagtgg tccatcagac tctgaacgaa 1680

ctgcgcaagc tgatgaatga gctgattacc atcctgggag ctaaacctca ggagatcaca 1740

gtggaactgg cacgagagct gaaggtcgga gctgagaaaa gagaggacat cattaagcag 1800

cagaccaaac aggaaaagga ggctgtgctg gcatatagca agtactgcga gcccaacaat 1860

ctggacaaaa ggtatattga aaggttccgc ctgctggagg atcaggcctt tgtgtgccct 1920

tactgtctgg agcacattag cgtcgcagat atcgcagctg gaagggcaga cgtggatcat 1980

atcttcccac gcgacgatac agctgacaac tcctacggga ataaggtggt cgcacaccga 2040

cagtgtaacg atatcaaggg aaagcggacc ccctatgcag ccttcagtaa tacatcagcc 2100

tggggcccca tcatgcatta cctggacgaa acccctggga tgtggcgcaa aagaaggaag 2160

tttgagacaa acgaggaaga gtatgctaag tacctgcagt caaaaggctt cgtgagcagg 2220

tttgaaagcg ataactccta tatcgcaaaa gctgcaaagg agtacctgcg ctgcctgttc 2280

aatccaaaca atgtgactgc cgtcgggtcc ctgaagggaa tggagacatc tatcctgcgg 2340

aaggcctgga atctgcaggg aattgacgat ctgctgggca gccggcactg gagtaaggac 2400

gccgatacca gccccacaat gcgcaaaaac cgggacgaca atcggcacca tggcctggac 2460

gccatcgtgg ctctgtattg ttccagatct ctggtccaga tgattaacac catgtccgag 2520

```
cagggcaagc gagcagtgga aatcgaggct atgattccta tcccagggta cgcatccgaa     2580
ccaaatctgt ctttcgaagc ccagcgggag ctgtttagaa agaaaatcct ggagttcatg     2640
gacctgcacg cctttgtgag tatgaaaacc gacaacgatg caaatggcgc cctgctgaaa     2700
gatactgtgt attcaattct gggagcagac acccagggag aggatctggt gttcgtggtc     2760
aagaaaaaga ttaaggacat cggcgtgaaa atcggggatt atgaagaggt cgcatctgcc     2820
attcgaggcc ggatcaccga caaacagcca aagtggtatc ccatggaaat gaaagataag     2880
atcgagcagc tgcagtctaa gaacgaagcc gctctgcaga aatacaagga gagtctggtg     2940
caggcagccg ctgtcctgga agagagtaat aggaagctga ttgagtcagg caaaaagccc     3000
atccagctga gtgaaaaaac aatttcaaaa aaggccctgg agctggtggg cgggtactat     3060
tacctgatta gcaacaacaa gcgcacaaag actttcgtgg tcaaggaacc ttcaaacgag     3120
gtgaaagggt tcgcatttga cactggaagc aatctgtgcc tggactttta tcacgatgcc     3180
cagggaaagc tgtgtggcga gatcattaga aaaatccagg ctatgaaccc ttcctataag     3240
ccagcataca tgaaacaggg gtattctctg tacgtgagac tgtaccaggg cgacgtgtgc     3300
gagctgaggg caagcgatct gactgaagca gagtccaacc tggccaagac cacacatgtc     3360
cgcctgccca atgctaaacc tgggcgaacc ttcgtgatca ttatcacctt tacagagatg     3420
gggtctggat atcagatcta cttcagcaac ctggccaagt ccaaaaaggg acaggacact     3480
agttttaccc tgactaccat caagaattat gatgtgcgga aagtccagct gtctagtgcc     3540
gggctggtga gatacgtcag ccctctgctg gtggacaaaa tcgagaagga tgaagtcgct     3600
ctgtgtggag agtaagaatt c                                               3621
```

<210> 131
<211> 3588
<212> ADN
<213> Secuencia Artificial

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

<400> 131

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60
aaggtcgaag cgtccatggc aagacctgca tttcgggcac ctcggagaga acacgtcaac      120
ggctggaccc ctgacccaca tcggattagc aaacccttt tcatcctggt gagctggcac       180
ctgctgtccc gggtggtcat tgacagctcc tctggatgct tcccaggcac cagccgggac      240
cacaccgaca agtttgccga gtgggaatgt gctgtgcagc cctacaggct gagtttcgac      300
ctggggacca actcaatcgg atggggcctg ctgaatctgg atcgccaggg aaaaccaagg      360
```

gagatccgag cactggggtc ccgcatcttc agcgacggac gggatcccca ggacaaggct 420

tctctggctg tggcacggag actggccaga cagatgaggc gccgacggga cagatatctg 480

actagaagga ccaggctgat gggagctctg gtgcgcttcg gcctgatgcc agcagaccct 540

gcagctagga agcgcctgga agtggccgtc gatccatacc tggcacgaga gcgggccaca 600

agagaaaggc tggagccctt cgaaatcggg agggccctgt ttcacctgaa ccagcgccga 660

ggatataaac ccgtgcgcac cgccacaaag cctgatgagg aagccggcaa ggtgaaagag 720

gctgtcgaaa ggctggaggc agcaatcgct gcagccggag cacctactct gggagcttgg 780

ttcgcatggc gaaaaacacg aggagaaact ctgcgagcac gactggctgg gaagggaaaa 840

gaggctgcat acccattcta tcccgcacgg agaatgctgg aggccgaatt tgacactctg 900

tgggcagagc aggccaggca ccatccagat ctgctgaccg ccgaagctcg cgagatcctg 960

cggcacagaa tttttcatca gcggcccctg aagccacctc cagtgggaag atgcactctg 1020

taccctgacg atgggagagc tcctagggca ctgccaagcg ctcagaggct gcgcctgttc 1080

caggagctgg ccagcctgag agtgatccac ctggacctgt ccgaacgccc tctgacccca 1140

gctgagcgag atcggattgt ggcatttgtc cagggcagac cccctaaagc cggaaggaag 1200

cctggcaaag tgcagaagag cgtcccattc gaaaagctga gggggctgct ggagctgcca 1260

ccaggcactg ggttttctct ggagagtgac aaacggcctg aactgctggg cgacgagaca 1320

ggcgccagaa tcgcaccagc attcggacct ggatggacag ctctgcctct ggaggaacag 1380

gacgccctgg tggaactgct gctgacagag gcagaaccag agagggcaat tgcagctctg 1440

actgcacgat gggctctgga cgaggcaact gcagcaaagc tggctggagc aaccctgcca 1500

gattttcacg gacgatatgg caggcgcgca gtggctgaac tgctgcctgt cctggaacgc 1560

gagacacgag gcgacccaga tgggagagtg aggcccatcc ggctggacga ggcagtgaaa 1620

ctgctgagag gcgggaagga tcactcagac ttcagccggg aaggagctct gctggacgca 1680

ctgccctact atggagccgt gctggagaga catgtcgctt ttgggacagg aaaccccgca 1740

gatcctgagg aaaagcgggt gggaagagtc gccaatccca ctgtgcacat cgctctgaac 1800

cagctgcggc atctggtgaa tgcaattctg gccaggcacg gccgccctga ggaaatcgtg 1860

attgagctgg cacgggacct gaaaaggtct gccgaagatc gacggagaga ggacaagcgg 1920

caggccgata accagaaaag aaatgaggaa cgcaagcgac tgatcctgag tctgggagag 1980

cgcccaaccc cacgaaacct gctgaagctg cggctgtggg aggaacaggg cccagtggaa 2040

aataggcgct gcccctactc tggggagaca attagtatga gaatgctgct gagcgagcag 2100

gtggacatcg atcacattct gccattcagc gtgtccctgg acgattccgc tgcaaacaag 2160

gtggtctgtc tgcgggaggc caacagaatc aagcggaata gatctccctg ggaggccttc 2220

ggccatgaca gtgagagatg ggcagggatt ctggcacgag cagaagctct gcccaagaac 2280

aaaaggtggc gctttgctcc tgacgcactg gagaaactgg aaggagaggg aggcctgcga 2340

gcacgacacc tgaatgatac aaggcatctg agtcgcctgg ccgtggagta tctgcggtgc 2400

gtctgtccta aggtgcgggt gagcccaggc cgactgactg cactgctgcg acggagatgg 2460

ggcatcgacg ccattctggc agaagcagat ggacctccac cagaagtgcc cgcagagaca 2520

ctggaccctt ccccagctga gaagaaccga gcagaccacc ggcaccatgc cctggatgct 2580

gtggtcatcg gctgtattga tcgctcaatg gtgcagcgag tccagctggc cgctgcaagc 2640

gcagaaagag aggccgctgc aagggaggac aatatcaggc gcgtgctgga gggattcaaa 2700

gaggaacctt gggatggctt tagagctgaa ctggagcgac gggcacggac catcgtggtg 2760

agccacagac cagaacatgg gattggggga gccctgcata aggagacagc ttacgggcct 2820

gtggaccctc cagaggaagg attcaacctg gtggtcagga aaccaatcga cggcctgtca 2880

aaggatgaga ttaatagcgt gcgggacccc cggctgagaa gggcactgat cgatcgcctg 2940

gccattcgcc gacgggatgc taacgaccct gctaccgcac tggccaaagc agctgaggat 3000

ctggcagcac agccagcctc ccgcggcatc agaagggtgc gggtcctgaa gaaagaatct 3060

aaccccatta gggtggagca cggcgggaat ccaagtggac cccgctcagg aggccctttt 3120

cataagctgc tgctggcagg agaggtgcac catgtggacg tcgcactgcg agcagatggc 3180

cgccgatggg tgggacactg ggtcacactg ttcgaggcac atgggggacg gggagcagac 3240

ggagctgcag ccccacctag actgggcgat ggggaaagat ttctgatgag gctgcacaag 3300

ggagactgcc tgaaactgga gcataagggc agagtgaggg tcatgcaggt ggtcaaactg 3360

gaacctagtt caaatagcgt ggtcgtggtc gagccacacc aggtgaaaac cgacagatcc 3420

aaacatgtca agatctcttg tgatcagctg cgcgctcgag gagcacggag agtgaccgtc 3480

gatccactgg gacgggtgag agtccacgcc ccaggagcta gggtgggaat cggaggggac 3540

gccggacgaa ccgctatgga acccgcagag gatattagct aagaattc 3588

<210> 132
<211> 3234
<212> ADN
<213> Secuencia Artificial

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

<400> 132
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatggg agagaatatg attgacgaga gtctgacctt cggcattgac 120

ctggggattg ggagttgtgg gtgggcagtg ctgaggcggc caagcgcctt cggacggaag 180

ggcgtgatcg agggaatggg ctcctggtgc tttgacgtcc ccgagacatc taaagaacgg 240

actcctacca accagatccg gagatccaat agactgctga ggcgcgtgat ccgacggaga 300

aggaaccgga tggccgctat tcgccgactg ctgcacgcag caggcctgct gccatcaacc 360

gacagcgatg ccctgaagcg gcccggacat gatccttggg aactgagggc acgaggcctg 420

gacaagccac tgaaacccgt ggagttcgct gtggtcctgg gccacatcgc aaagcggaga 480

gggtttaaat ccgctgcaaa gagaaaagcc acaaacatta gctccgacga taagaaaatg 540

ctgacagccc tggaggctac tcgagaacgg ctggggagat acaggaccgt gggagaaatg 600

ttcgccaggg accctgattt tgcttctagg cgccgaaatc gcgagggcaa gtatgatagg 660

accacagctc gcgacgatct ggagcacgaa gtgcacgccc tgttcgcagc tcagcggaga 720

ctgggacagg gatttgccag tccagaactg gaggaggcct tcaccgcttc agcatttcat 780

cagaggccca tgcaggacag cgagcgcctg gtgggattct gccctttga gcgaaccgaa 840

aagcgggcag ccaaactgac accctctttc gagcgctttc gactgctggc ccggctgctg 900

aacctgagaa tcactacccc agacggagag cggcccctga cagtggatga aattgctctg 960

gtcacccggg acctgggcaa gaccgcaaaa ctgagtatca agcgggtgag aactctgatt 1020

ggactggagg acaatcagag gttcacaact atccgccccg aggacgaaga tcgagacatt 1080

gtggctcgga caggcggggc aatgacaggg actgccaccc tgaggaaggc actgggagag 1140

gccctgtgga ctgatatgca ggagcgccct gaacagctgg acgctatcgt gcaggtcctg 1200

agcttctttg aggccaacga aacaatcact gagaagctga gggaaattgg cctgactctg 1260

gccgtgctgg acgtcctgct gaccgcactg gatgccggag tgttcgccaa gtttaaaggc 1320

gctgcacaca tcagcaccaa agccgctagg aatctgctgc cacatctgga gcagggcagg 1380

cgctacgatg aggcctgcac aatggcaggg tatgaccacg cagcctcccg cctgtctcac 1440

catggccaga tcgtggcaaa gacacagttc aacgccctgg tcactgagat cggcgaatcc 1500

attgccaatc caatcgctcg gaaggcactg atcgaggggc tgaaacagat ttgggccatg 1560

agaaaccact gggggctgcc cggaagtatc catgtggagc tggcccggga tgtcggcaac 1620

tcaattgaaa agcgacggga gattgaaaag cacatcgaga aaaatactgc cctgagggct 1680

cgcgagagaa gggaggtgca tgatctgctg gacctggaag atgtcaatgg cgacaccctg 1740

ctgcgatacc ggctgtggaa ggagcaggga ggcaaatgcc tgtatacagg gaaggccatc 1800

cacattcggc agatcgctgc aactgacaac tccgtgcagg tcgatcatat tctgccttgg 1860

agccggttcg gcgacgatag ttttaacaac aagaccctgt gtctggcctc tgctaatcag 1920

cagaagaaaa ggtcaacacc atacgagtgg ctgagcggcc agactgggga tgcatggaac 1980

gccttcgtgc agcgcatcga gacaaataag gaactgagag ggtttaagaa aaggaactat 2040

ctgctgaaga atgctaaaga ggcagaggaa aaattcagaa gcaggaacct gaatgacacc 2100

```
agatacgccg ctaggctgtt cgcagaggcc gtgaagctgc tgtatgcctt tggggagaga     2160

caggaaaaag ggggaaaccg ccgagtgttt actcggcctg gagcactgac cgcagcactg     2220

agacaggctt ggggagtgga gagcctgaag aaacaggatg ggaagcgcat caatgacgat     2280

cgacaccatg ccctggatgc tctgaccgtg gctgcagtcg acgaggccga aattcagagg     2340

ctgacaaaat cattccacga gtgggaacag cagggcctgg ggcggcctct gcggagagtg     2400

gagccacctt gggagagctt ccgggcagac gtcgaggcta cctaccctga agtgtttgtc     2460

gcacggccag agaggcgccg agcaagagga gaaggccatg ccgctaccat ccggcaggtg     2520

aaggagagag aatgcacacc aattgtgttt gagagaaagg ctgtctctag tctgaaagag     2580

gcagacctgg aacgaatcaa agatggcgag cgcaacgaag caattgtgga ggccatcagg     2640

agctggattg ccactggacg cccagctgat gcaccaccac gctccccccg aggcgacatc     2700

attaccaaga tcaggctggc caccaccatc aaggcagccg tgcctgtccg cggagggacc     2760

gcaggaaggg gagaaatggt gcgcgcagat gtgttcagca agccaaaccg gagagggaaa     2820

gacgagtggt atctggtgcc cgtgtatcca caccagatca tgaacaggaa ggcttggccc     2880

aaacctccaa tgcgctcaat tgtggccaat aaggatgagg acgaatggac cgaagtggga     2940

cctgaacacc agttccggtt tagcctgtac cctagatcca atatcgagat cattaggcca     3000

tctggagaag tgatcgaagg atatttcgtc ggcctgcatc gcaacactgg cgctctgacc     3060

atcagtgcac acaatgatcc caagagtatc cattcaggca ttgggaccaa gacactgctg     3120

gccatttcca aataccaggt ggacagattc ggcagaaagt ctccagtgcg caaagaggtc     3180

cgaacttggc acggggaagc ctgtatctct cccacccccc ctggataaga attc           3234
```

<210> 133
<211> 3504
<212> ADN
<213> Secuencia Artificial

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

<400> 133

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgag agagaacggg agtgacgaaa ggcggaggaa tatggacgaa      120

aaaatggatt acaggattgg gctggacatc ggcatcgcat cagtggggtg ggccgtcctg      180

cagaacaatt cagacgatga gcctgtgagg attgtcgacc tgggagtgcg cattttcgat      240

accgcagaga tcccaaagac aggcgaaagc ctggcaggac cccggagagc agctcgaacc      300

acaaggcgcc gactgcggag aaggaaacac cgcctggacc gaatcaagtg gctgttcgag      360
```

```
aaccaggggc tgatcaatat tgacgatttt ctgaagagat acaacatggc cggactgcca      420
gatgtgtacc agctgcggta tgaggctctg gacagaaaac tgaccgatga ggaactggcc      480
caggtgctgc tgcacatcgc taagcatcgg ggcttcagaa gcaccaggaa agccgagaca      540
gcagccaagg aaaacggggc agtgctgaag gccacagacg agaatcagaa acgaatgcag      600
gaaaagggat acaggacagt gggcgagatg atctacctgg acgaggcctt ccggactggc      660
tgctcttgga gtgagaaagg gtacatcctg acccccccgca acaaggctga aaattatcag      720
cacacaatgc tgcgggcaat gctggtggag gaagtcaagg agattttcag ctcccagcgc      780
cgactgggca acgaaaaagc cactgaggaa ctggaggaaa agtacctgga gatcatgacc      840
tcccagcgct cttttgacct ggggcctgga atgcagccag atgggaagcc ctccccttat      900
gcaatggagg gcttctctga cagagtgggg aaatgtactt ttctggggga tcagggagag      960
ctgaggggcg ctaaggggac ctacacagcc gaatatttcg tggctctgca gaaaatcaac     1020
cacacaaagc tggtcaatca ggacggcgag acaaggaatt tcactgagga agagcggaga     1080
gccctgactc tgctgctgtt tacccagaaa gaggtgaagt acgctgcagt ccgcaagaaa     1140
ctgggcctgc ctgaggacat cctgttctac aacctgaact acaagaaggc cgctactaaa     1200
gaagagcagc agaaggagaa ccagaatacc gaaaaagcca agtttatcgg gatgccatac     1260
tatcacgatt acaagaaatg cctggaagag agagtgaagt atctgaccga gaacgaagtc     1320
agggacctgt ttgatgagat cggaatgatt ctgacttgtt acaaaaatga cgattcccgc     1380
accgaacgac tggccaagct gggactggtg cccatcgaga tggaaggcct gctggcttat     1440
actcctacca aattccagca tctgtctatg aaggcaatgc ggaacatcat tcccttttctg     1500
gagaaaggga tgacctacga caaggcttgc gaagaggcag gatatgactt caaagccgat     1560
agcaagggga ctaaacagaa gctgctgacc ggagagaacg tgaatcagac aatcaacgaa     1620
attactaatc ctgtggtcaa acgctcagtg agccagacag tgaaggtcat taacgccatc     1680
attcggactt acggcagtcc acaggctatc aatattgagc tggcaagaga aatgtcaaag     1740
acctttgaag agaggcgcaa aatcaagggg gacatggaga aacggcagaa gaacaatgaa     1800
gatgtgaaga aacagattca ggagctggga aaactgtctc ctacaggcca ggacatcctg     1860
aagtacagac tgtggcagga gcagcagggg atttgtatgt atagtggaaa aaccatccca     1920
ctggaagagc tgttcaagcc cggctacgac atcgatcaca ttctgcccta ttcaattaca     1980
ttcgacgata gctttaggaa caaagtgctg gtcacatccc aggagaacag acagaagggc     2040
aataggactc cttacgagta tatggggaac gacgaacagc gctggaatga gtttgaaacc     2100
agggtgaaaa ctaccatccg cgattacaag aaacagcaga agctgctgaa gaaacatttc     2160
tctgaagagg aaaggagtga gtttaaagaa cggaacctga cagacactaa gtacatcaca     2220
accgtgatct acaacatgat cagacagaat ctggagatgg ccccccctgaa ccgccctgaa     2280
```

```
aagaaaaagc aggtgcgggc tgtcaatggc gcaattaccg cctacctgcg aaaacggtgg     2340

gggctgccac agaagaatcg ggagacagac acacaccatg ctatggatgc agtggtcatc     2400

gcctgctgta ccgacggcat gatccagaaa attagtagat acacaaaggt gagagagagg     2460

tgctattcaa agggaacaga gttcgtcgat gcagagactg gcgaaatctt tagacccgag     2520

gactacagca gggccgagtg ggatgaaatt ttcggcgtgc acatcccaaa gccctgggag     2580

acatttcgcg ccgaactgga cgtccgaatg ggggacgatc caaagggatt cctggacact     2640

catagcgatg tggctctgga gctggattat cccgagtaca tctacgaaaa cctgcggcct     2700

atcttcgtga gcagaatgcc aaatcacaag gtcaccggag cagcccatgc tgacacaatt     2760

cggtccccaa gacactttaa agatgagggc atcgtgctga ctaagaccgc actgaccgac     2820

ctgaaactgg acaaggatgg ggagatcgac ggatactata acccccagtc cgatctgctg     2880

ctgtacgaag cactgaaaaa gcagctgctg ctgtatggca atgatgccaa aaaggccttc     2940

gctcaggact ttcataaacc caaggccgat ggaactgagg gccctgtggt caggaaggtg     3000

aagatccaga aaaagcagac catgggagtg ttcgtcgact ctggcaacgg gattgccgag     3060

aatggcggga tggtgcgcat cgatgtgttc cgagtcaacg gcaagtacta ttttgtgccc     3120

gtctacaccg ctgacgtggt caaaaaggtg ctgcctaata gggccagtac agctcacaag     3180

ccatacggcg agtggaaagt gatggaggac aaggatttcc tgtttagtct gtattcacgc     3240

gacctgatcc atatcaagtc taaaaaggat atccctatta agatggtgaa cggaggcatg     3300

gaggggatca aggaaaccta cgcatactat attggagccg acatcagcgc tgcaaatatc     3360

cagggcattg cccacgattc caggtataaa ttccgcggac tgggcattca gtctctggac     3420

gtgctggaga agtgtcagat cgatgtgctg ggacatgtca gcgtggtccg atccgaaaag     3480

cggatgggct ttagctaaga attc                                            3504
```

<210> 134
<211> 3243
<212> ADN
<213> Secuencia Artificial

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

<400> 134

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc      120

gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg      180

ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc      240
```

ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac 300

ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc 360

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg 540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac 600

tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720

ggagaaggga gcccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga 780

cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg 840

tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg 900

gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca 960

ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg 1020

acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc 1080

acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg 1140

actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg 1200

acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg 1260

tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt 1320

gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag 1380

atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc 1440

cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt 1500

atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag 1560

aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag 1620

aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg 1680

tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc 1740

gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc 1800

aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca 1860

gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag 1920

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc 1980

tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc 2040

ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc 2100

atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac 2160

```
aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt     2220
aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag     2280
aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc     2340
actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg     2400
gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat     2460
aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag     2520
ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag     2580
acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat     2640
aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc     2700
gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac     2760
gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat     2820
gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa     2880
aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag     2940
attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat     3000
ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat     3060
atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga     3120
attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg     3180
ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa     3240
ttc                                                                  3243
```

<210> 135
<211> 3243
<212> ADN
<213> Secuencia Artificial

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

<400> 135

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60
aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc      120
gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg      180
ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc      240
ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac      300
ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc      360
```

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg 540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac 600

tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720

ggagaaggga gcccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga 780

cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg 840

tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg 900

gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca 960

ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg 1020

acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc 1080

acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg 1140

actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg 1200

acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg 1260

tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt 1320

gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag 1380

atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc 1440

cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt 1500

atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag 1560

aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag 1620

aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg 1680

tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc 1740

gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc 1800

aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca 1860

gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag 1920

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc 1980

tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc 2040

ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc 2100

atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac 2160

aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt 2220

aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag 2280

```
aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc     2340

actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg     2400

gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat     2460

aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag     2520

ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag     2580

acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat     2640

aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc     2700

gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac     2760

gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat     2820

gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa     2880

aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag     2940

attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat     3000

ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat     3060

atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga     3120

attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg     3180

ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa     3240

ttc                                                                   3243
```

<210> 136
<211> 3891
<212> ADN
<213> Secuencia Artificial

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

<400> 136

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgaa caatagcatc aaatctaaac ctgaagtgac catcgggctg      120

gacctgggag tgggaagcgt ggggtgggca atcgtggata acgaaacaaa catcattcac      180

catctgggct ccaggctgtt ttctcaggcc aagactgctg aggatcggag atctttccgc      240

ggggtgaggc gcctgatccg acggagaaaa tacaagctga aacgattcgt caatctgatt      300

tggaagtaca acagctattt cggcttcaag aacaaagagg acatcctgaa caattatcag      360

gagcagcaga agctgcacaa taccgtgctg aacctgaaat cagaggcact gaatgccaag      420

atcgatccta aagcactgag ctggattctg cacgactacc tgaagaacag aggccatttt      480
```

tatgaggaca atagggattt caacgtgtac ccaacaaagg agctggccaa gtacttcgat 540

aagtacgggt actacaaggg aatcattgac agcaaggagg acaatgataa caaactggag 600

gaagagctga caaagtacaa attctccaat aagcactggc tggaagaggt gaagaaagtc 660

ctgtctaacc agactggcct gccagaaaag tttaaagaag agtatgagtc actgttcagc 720

tacgtgagaa attattcaga gggcccaggg agcatcaact ctgtcagtcc ctacgggatc 780

taccatctgg acgaaaaaga gggaaaggtg gtccagaagt acaacaacat ctgggataag 840

acaatcggaa agtgcaacat cttccctgac gagtatagag ctcccaagaa cagtcctatc 900

gcaatgattt tcaatgaaat caacgagctg tccacaatca ggtcatacag catctacctg 960

actggctggt tcattaatca ggagttcaag aaagcctacc tgaacaagct gctggatctg 1020

ctgatcaaaa ccaacggaga gaagccaatt gacgcaaggc agttcaagaa actgcgcgaa 1080

gagacaatcg ccgaaagcat tggcaaagag acactgaagg atgtggagaa tgaagagaaa 1140

ctggaaaagg aggaccacaa gtggaaactg aagggactga agctgaatac caacggcaaa 1200

atccagtaca acgatctgag ctccctggct aagtttgtgc acaaactgaa gcagcatctg 1260

aaactggatt tcctgctgga ggaccagtat gcaacactgg acaagatcaa tttcctgcag 1320

tccctgtttg tgtacctggg caagcacctg agatattcca atagggtcga ttctgccaac 1380

ctgaaggaat tttccgactc taacaaactg ttcgagcgca tcctgcagaa acagaaggat 1440

gggctgttca agctgtttga acagactgac aaagacgatg agaagatcct ggcccagaca 1500

catagtctgt caactaaggc catgctgctg gctattaccc ggatgacaaa tctggacaac 1560

gatgaggaca accagaaaaa caatgacaag ggctggaatt ttgaggccat caaaaacttc 1620

gatcagaagt ttatcgacat caccaagaaa aacaacaacc tgagcctgaa acagaataag 1680

cgctacctgg acgatcgatt catcaacgat gctattctgt cccctggggt gaagcgaatc 1740

ctgcgggagg caaccaaggt ctttaatgcc attctgaaac agttctctga agagtacgac 1800

gtgacaaagg tggtcatcga actggctcgc gagctgagcg aagagaagga actggagaac 1860

acaaagaact acaagaaact gatcaagaaa aacggcgaca agattagtga gggcctgaaa 1920

gcactgggga tctcagaaga tgagatcaaa gacattctga agagtcccac taaatcatac 1980

aagtttctgc tgtggctgca gcaggaccac atcgatcctt atagcctgaa ggagatcgcc 2040

ttcgacgata tttttaccaa aacagaaaag ttcgagatcg accatatcat tccctacagc 2100

atttccttcg acgattctag ttcaaacaag ctgctggtgc tggctgaaag taatcaggca 2160

aagtcaaacc agactcctta tgagttcatc agctccggaa acgcaggcat taagtgggaa 2220

gattacgagg cctattgccg caagttcaag gatggggact ctagtctgct ggacagcacc 2280

cagcggtcca agaaattcgc caaaatgatg aaaaccgata cctcaagcaa gtacgacatc 2340

ggatttctgg ctcgaaatct gaacgatact cggtacgcaa ccattgtgtt ccgggacgcc 2400

ctggaggact atgctaataa ccacctggtc gaggacaaac ccatgtttaa ggtggtctgt 2460

atcaatgggt ccgtgacctc tttcctgcgg aagaactttg acgattcctc ttacgccaag 2520

aaagatagag acaagaatat ccaccatgct gtggatgcaa gtatcatctc aattttcagc 2580

aacgagacaa agactctgtt caaccagctg actcagtttg ctgactataa actgttcaag 2640

aacaccgatg gcagctggaa gaaaatcgac cctaagacag gggtggtcac tgaagtgacc 2700

gacgagaatt ggaagcagat tagggtgcgc aaccaggtga gcgaaatcgc caaagtcatt 2760

gagaagtaca tccaggatag caacatcgaa agaaaggcta ggtattcccg caaaatcgag 2820

aataagacta acatttccct gtttaatgac accgtgtact ctgccaagaa agtcggctat 2880

gaggatcaga tcaaaagaaa gaacctgaaa accctggaca ttcacgaatc tgctaaagag 2940

aataagaaca gtaaagtgaa gcggcagttt gtctacagaa agctggtgaa tgtcagcctg 3000

ctgaataacg ataagctggc agacctgttc gccgaaaaag aggatatcct gatgtatagg 3060

gccaatccat gggtcatcaa cctggctgag cagattttca atgaatacac tgagaacaag 3120

aaaatcaagt cccagaacgt gtttgaaaaa tatatgctgg acctgaccaa agagttcccc 3180

gagaagttca gcgagtttct ggtgaagtcc atgctgagaa acaagaccgc catcatctac 3240

gacgataaga aaaacattgt ccatcgaatc aaacggctga agatgctgag ttcagaactg 3300

aaagagaata agctgtctaa cgtgatcatt aggtctaaga atcagagtgg gaccaaactg 3360

tcataccagg atacaatcaa cagcctggcc ctgatgatta tgcgcagcat cgaccctact 3420

gctaagaaac agtatattcg agtgccactg aataccctga acctgcacct gggagatcat 3480

gactttgatc tgcacaatat ggatgcttac ctgaagaaac caaaattcgt gaagtatctg 3540

aaagcaaacg aaatcggcga cgagtacaag ccctggaggg tcctgacatc tggcactctg 3600

ctgatccata agaaggataa gaaactgatg tacatcagct ccttccagaa tctgaacgac 3660

gtgatcgaaa ttaagaatct gatcgaaacc gagtataaag agaacgacga ttctgatagt 3720

aagaaaaaga aaaaggcaaa ccgctttctg atgaccctga gcacaatcct gaatgactac 3780

attctgctgg acgccaagga taacttcgac atcctggggc tgtctaaaaa tcggatcgat 3840

gagattctga acagtaagct gggactggac aagattgtga ataagaatt c 3891

<210> 137
<211> 3093
<212> ADN
<213> Secuencia Artificial

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

<400> 137

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgag gattctgggg tttgacattg gcattaacag catcgggtgg 120

gcttttgtgg agaacgacga actgaaggac tgcggagtgc ggatcttcac aaaggccgag 180

aacccaaaaa ataaggaaag cctggcactg ccccggagaa atgcacgcag ctccaggcgc 240

cgactgaaac ggagaaaggc ccggctgatc gctattaaga gaatcctggc caaagagctg 300

aagctgaact acaaggacta tgtcgcagct gatggagagc tgccaaaggc ctacgaagga 360

tccctggcat ctgtgtacga gctgcggtat aaggccctga cacagaacct ggaaactaaa 420

gatctggcca gagtgatcct gcacattgct aagcataggg ggtacatgaa caagaacgag 480

aagaaatcaa acgacgctaa gaaaggaaag atcctgagcg ctctgaaaaa caatgcactg 540

aagctggaga actaccagag cgtgggcgaa tacttctaca aggagttctt tcagaaatac 600

aagaaaaaca caaagaactt catcaagatc cgcaacacta aggataatta caacaattgc 660

gtgctgtcta gtgacctgga aaaagagctg aagctgatcc tggaaaaaca gaaggagttc 720

ggctacaact actctgaaga tttcatcaac gagattctga aggtcgcctt ctttcagcgg 780

cccctgaagg acttcagtca cctggtgggg gcctgcactt tctttgagga agagaaaagg 840

gcctgtaaga acagctactc tgcctgggag tttgtggctc tgaccaagat cattaacgag 900

atcaagagcc tggagaagat cagcggcgaa attgtgccaa cccagacaat caacgaggtc 960

ctgaatctga tcctggacaa ggggtctatc acctacaaga aattcagaag ttgtatcaat 1020

ctgcatgaga gtatcagctt caagagcctg aagtatgata aagaaaacgc cgagaatgct 1080

aaactgatcg acttccgcaa gctggtggag tttaagaaag ccctgggagt ccacagcctg 1140

tcccggcagg aactggatca gatctccact catatcaccc tgattaagga caacgtgaag 1200

ctgaaaaccg tcctggagaa atacaacctg agtaatgaac agatcaacaa tctgctggaa 1260

attgagttca acgattatat caacctgagc ttcaaggccc tgggaatgat tctgccactg 1320

atgcgcgagg gcaaacgata cgacgaggcc tgcgagatcg ccaatctgaa acctaagacc 1380

gtggacgaga agaaagattt cctgccagca ttttgtgatt ccattttcgc ccacgagctg 1440

tctaaccccg tggtcaatag ggctatcagc gaataccgca aggtgctgaa cgcactgctg 1500

aagaaatatg gaaaggtcca caaaattcat ctggagctgg ctcgcgacgt gggcctgtcc 1560

aagaaagcac gagagaagat cgaaaaagag cagaaggaaa accaggccgt gaatgcatgg 1620

gccctgaagg aatgcgagaa tattggcctg aaggccagcg caaagaacat cctgaaactg 1680

aagctgtgga aagaacagaa ggagatctgt atctactccg gaaataagat ctctattgag 1740

cacctgaaag atgaaaaggc cctggaggtg gaccatatct acccctattc taggagtttc 1800

gacgattctt ttatcaacaa agtgctggtg ttcaccaagg aaaatcagga gaaactgaac 1860

aagacacctt tcgaggcctt tggcaagaat attgaaaaat ggagcaagat ccagaccctg 1920

```
gctcagaacc tgccatacaa gaaaaagaat aagattctgg acgagaactt caaagataag     1980

cagcaggagg actttatctc tcgaaatctg aacgacaccc ggtatatcgc tacactgatt     2040

gcaaaataca caaaggagta tctgaacttc ctgctgctga gcgaaaatga gaacgccaat     2100

ctgaagagtg gcgaaaaagg gtcaaagatc cacgtgcaga ctattagcgg gatgctgacc     2160

tccgtcctga ggcacacatg ggggtttgac aaaaaggatc gcaacaatca tctgcaccat     2220

gcactggatg ccatcattgt ggcctacagt acaaattcaa tcattaaggc tttcagcgat     2280

ttccggaaaa accaggagct gctgaaggcc agattctacg ctaaagaact gacttccgat     2340

aactataaac atcaggtcaa gttctttgag cctttcaaga gttttagaga aaaaatcctg     2400

tcaaagatcg acgagatttt cgtgtccaaa ccacctcgaa agcgagctag gcgcgcactg     2460

cacaaggata cctttcattc tgagaacaag atcattgaca agtgcagcta caactccaag     2520

gaaggcctgc agattgccct gagctgtgga agagtgagga aaatcggcac taagtatgtc     2580

gagaatgata ccatcgtgag ggtcgacatt ttcaaaaagc agaacaagtt ttacgctatc     2640

ccaatctacg caatggattt tgccctgggg atcctgccca ataagatcgt gattactgga     2700

aaagataaga acaataaccc caaacagtgg cagaccattg acgaatcata cgagttctgc     2760

tttagcctgt ataagaatga cctgatcctg ctgcagaaaa agaacatgca ggaacctgag     2820

ttcgcctact ataacgattt ttcaatcagc acatcaagca tttgtgtgga gaaacacgac     2880

aacaagttcg aaaatctgac tagcaaccag aagctgctgt tttccaatgc aaaagagggc     2940

tctgtgaagg tcgaaagtct ggggatccag aacctgaaag tgttcgagaa gtacatcatt     3000

acccccctgg gagataaaat taaggctgac tttcagcctc gagaaaacat cagcctgaaa     3060

accagtaaaa agtatggcct gaggtaagaa ttc                                  3093
```

<210> 138
<211> 4104
<212> ADN
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(4104)

<400> 138

```
atg gac aag aag tac agc atc ggc ctg gac atc ggc acc aac tct gtg          48
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15

ggc tgg gcc gtg atc acc gac gag tac aag gtg ccc agc aag aaa ttc          96
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

aag gtg ctg ggc aac acc gac cgg cac agc atc aag aag aac ctg atc         144
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
```

```
gga gcc ctg ctg ttc gac agc ggc gaa aca gcc gag gcc acc cgg ctg        192
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

aag aga acc gcc aga aga aga tac acc aga cgg aag aac cgg atc tgc        240
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

tat ctg caa gag atc ttc agc aac gag atg gcc aag gtg gac gac agc        288
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

ttc ttc cac aga ctg gaa gag tcc ttc ctg gtg gaa gag gat aag aag        336
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

cac gag cgg cac ccc atc ttc ggc aac atc gtg gac gag gtg gcc tac        384
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

cac gag aag tac ccc acc atc tac cac ctg aga aag aaa ctg gtg gac        432
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

agc acc gac aag gcc gac ctg cgg ctg atc tat ctg gcc ctg gcc cac        480
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

atg atc aag ttc cgg ggc cac ttc ctg atc gag ggc gac ctg aac ccc        528
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

gac aac agc gac gtg gac aag ctg ttc atc cag ctg gtg cag acc tac        576
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

aac cag ctg ttc gag gaa aac ccc atc aac gcc agc ggc gtg gac gcc        624
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

aag gcc atc ctg tct gcc aga ctg agc aag agc aga cgg ctg gaa aat        672
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

ctg atc gcc cag ctg ccc ggc gag aag aag aat ggc ctg ttc ggc aac        720
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

ctg att gcc ctg agc ctg ggc ctg acc ccc aac ttc aag agc aac ttc        768
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

gac ctg gcc gag gat gcc aaa ctg cag ctg agc aag gac acc tac gac        816
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

gac gac ctg gac aac ctg ctg gcc cag atc ggc gac cag tac gcc gac        864
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

ctg ttt ctg gcc gcc aag aac ctg tcc gac gcc atc ctg ctg agc gac        912
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
```

```
    290                 295                 300

atc ctg aga gtg aac acc gag atc acc aag gcc ccc ctg agc gcc tct        960
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

atg atc aag aga tac gac gag cac cac cag gac ctg acc ctg ctg aaa       1008
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

gct ctc gtg cgg cag cag ctg cct gag aag tac aaa gag att ttc ttc       1056
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

gac cag agc aag aac ggc tac gcc ggc tac att gac ggc gga gcc agc       1104
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

cag gaa gag ttc tac aag ttc atc aag ccc atc ctg gaa aag atg gac       1152
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

ggc acc gag gaa ctg ctc gtg aag ctg aac aga gag gac ctg ctg cgg       1200
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

aag cag cgg acc ttc gac aac ggc agc atc ccc cac cag atc cac ctg       1248
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

gga gag ctg cac gcc att ctg cgg cgg cag gaa gat ttt tac cca ttc       1296
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

ctg aag gac aac cgg gaa aag atc gag aag atc ctg acc ttc cgc atc       1344
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

ccc tac tac gtg ggc cct ctg gcc agg gga aac agc aga ttc gcc tgg       1392
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

atg acc aga aag agc gag gaa acc atc acc ccc tgg aac ttc gag gaa       1440
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

gtg gtg gac aag ggc gct tcc gcc cag agc ttc atc gag cgg atg acc       1488
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

aac ttc gat aag aac ctg ccc aac gag aag gtg ctg ccc aag cac agc       1536
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

ctg ctg tac gag tac ttc acc gtg tat aac gag ctg acc aaa gtg aaa       1584
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

tac gtg acc gag gga atg aga aag ccc gcc ttc ctg agc ggc gag cag       1632
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

aaa aag gcc atc gtg gac ctg ctg ttc aag acc aac cgg aaa gtg acc       1680
```

```
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

gtg aag cag ctg aaa gag gac tac ttc aag aaa atc gag tgc ttc gac       1728
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

tcc gtg gaa atc tcc ggc gtg gaa gat cgg ttc aac gcc tcc ctg ggc       1776
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

aca tac cac gat ctg ctg aaa att atc aag gac aag gac ttc ctg gac       1824
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

aat gag gaa aac gag gac att ctg gaa gat atc gtg ctg acc ctg aca       1872
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

ctg ttt gag gac aga gag atg atc gag gaa cgg ctg aaa acc tat gcc       1920
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

cac ctg ttc gac gac aaa gtg atg aag cag ctg aag cgg cgg aga tac       1968
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

acc ggc tgg ggc agg ctg agc cgg aag ctg atc aac ggc atc cgg gac       2016
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

aag cag tcc ggc aag aca atc ctg gat ttc ctg aag tcc gac ggc ttc       2064
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

gcc aac aga aac ttc atg cag ctg atc cac gac gac agc ctg acc ttt       2112
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

aaa gag gac atc cag aaa gcc cag gtg tcc ggc cag ggc gat agc ctg       2160
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

cac gag cac att gcc aat ctg gcc ggc agc ccc gcc att aag aag ggc       2208
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

atc ctg cag aca gtg aag gtg gtg gac gag ctc gtg aaa gtg atg ggc       2256
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

cgg cac aag ccc gag aac atc gtg atc gcc atg gcc aga gag aac cag       2304
Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
        755                 760                 765

acc acc cag aag gga cag aag aac agc cgc gag aga atg aag cgg atc       2352
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

gaa gag ggc atc aaa gag ctg ggc agc cag atc ctg aaa gaa cac ccc       2400
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
gtg gaa aac acc cag ctg cag aac gag aag ctg tac ctg tac tac ctg       2448
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

cag aat ggg cgg gat atg tac gtg gac cag gaa ctg gac atc aac cgg       2496
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

ctg tcc gac tac gat gtg gac gcc atc gtg cct cag agc ttt ctg aag       2544
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

gac gac tcc atc gac gcc aag gtg ctg acc aga agc gac aag gcc cgg       2592
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860

ggc aag agc gac aac gtg ccc tcc gaa gag gtc gtg aag aag atg aag       2640
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

aac tac tgg cgg cag ctg ctg aac gcc aag ctg att acc cag aga aag       2688
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

ttc gac aat ctg acc aag gcc gag aga ggc ggc ctg agc gaa ctg gat       2736
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

aag gcc ggc ttc atc aag aga cag ctg gtg gaa acc cgg cag atc aca       2784
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

aag cac gtg gca cag atc ctg gac tcc cgg atg aac act aag tac gac       2832
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

gag aat gac aag ctg atc cgg gaa gtg aaa gtg atc acc ctg aag tcc       2880
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

aag ctg gtg tcc gat ttc cgg aag gat ttc cag ttt tac aaa gtg cgc       2928
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

gag atc aac aac tac cac cac gcc cac gcc gcc tac ctg aac gcc gtc       2976
Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val
            980                 985                 990

gtg gga acc gcc ctg atc aaa aag  tac cct aag ctg gaa  agc gag ttc     3024
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

gtg tac  ggc gac tac aag gtg  tac gac gtg cgg aag  atg atc gcc        3069
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

aag agc  gag cag gaa atc ggc  aag gct acc gcc aag  tac ttc ttc        3114
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

tac agc  aac atc atg aac ttt  ttc aag acc gag att  acc ctg gcc        3159
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050
```

```
aac ggc  gag atc cgg aag cgg  cct ctg atc gag aca  aac ggc gaa        3204
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                 1060                 1065

acc ggg  gag atc gtg tgg gat  aag ggc cgg gat ttt  gcc acc gtg        3249
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                 1075                 1080

cgg aaa  gtg ctg agc atg ccc  caa gtg aat atc gtg  aaa aag acc        3294
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                 1090                 1095

gag gtg  cag aca ggc ggc ttc  agc aaa gag tct atc  ctg ccc aag        3339
Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                 1105                 1110

agg aac  agc gat aag ctg atc  gcc aga aag aag gac  tgg gac cct        3384
Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                 1120                 1125

aag aag  tac ggc ggc ttc gac  agc ccc acc gtg gcc  tat tct gtg        3429
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                 1135                 1140

ctg gtg  gtg gcc aaa gtg gaa  aag ggc aag tcc aag  aaa ctg aag        3474
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                 1150                 1155

agt gtg  aaa gag ctg ctg ggg  atc acc atc atg gaa  aga agc agc        3519
Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                 1165                 1170

ttc gag  aag aat ccc atc gac  ttt ctg gaa gcc aag  ggc tac aaa        3564
Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                 1180                 1185

gaa gtg  aaa aag gac ctg atc  atc aag ctg cct aag  tac tcc ctg        3609
Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                 1195                 1200

ttc gag  ctg gaa aac ggc cgg  aag aga atg ctg gcc  tct gcc ggc        3654
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                 1210                 1215

gaa ctg  cag aag gga aac gaa  ctg gcc ctg ccc tcc  aaa tat gtg        3699
Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                 1225                 1230

aac ttc  ctg tac ctg gcc agc  cac tat gag aag ctg  aag ggc tcc        3744
Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245

ccc gag  gat aat gag cag aaa  cag ctg ttt gtg gaa  cag cac aag        3789
Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260

cac tac  ctg gac gag atc atc  gag cag atc agc gag  ttc tcc aag        3834
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275

aga gtg  atc ctg gcc gac gct  aat ctg gac aaa gtg  ctg tcc gcc        3879
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
```

```
    1280                1285                1290

tac aac  aag cac cgg gat aag  ccc atc aga gag cag  gcc gag aat        3924
Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

atc atc  cac ctg ttt acc ctg  acc aat ctg gga gcc  cct gcc gcc        3969
Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

ttc aag  tac ttt gac acc acc  atc gac cgg aag agg  tac acc agc        4014
Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

acc aaa  gag gtg ctg gac gcc  acc ctg atc cac cag  agc atc acc        4059
Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

ggc ctg  tac gag aca cgg atc  gac ctg tct cag ctg  gga ggc gac        4104
Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365
```

<210> 139
<211> 1368
<212> PRT
<213> Homo sapiens

<400> 139

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140
```

```
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

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

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

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

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
```

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

```
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
```

```
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140
```

```
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                 1150                 1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                 1165                 1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                 1180                 1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                 1195                 1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                 1210                 1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                 1225                 1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365
```

<210> 140
<211> 155
<212> ARN
<213> Suterella wadsworthensis

<400> 140
guuucagugc uauaggaaac uauaggaaau caccuucggg ugagcugaaa uccccuaaag 60

cuaagauuga auccggccac uaucuauuag uagauauccg gauauucuga uauaaaaccu 120

cauucuuuga uuagaccaaa ggaugagguu uuuuu 155

<210> 141
<211> 107
<212> ARN
<213> Treponema denticola

<400> 141
guuugagagu uggaaacaac gaguucaaau aagaauucau caaaaucguc ccuuuuggga 60

ccgcucauug uggagcauca aggcuuaaca ugguuaagcc uuuuuuu 107

<210> 142
<211> 99
<212> ARN
<213> Filifactor alocis

<400> 142
guuugagagu aggaaacuac acgguucaaa uaaagaauuu uucuaaucgc ccaaugggcc 60

cauauugaua uggaugaaac ucgcuuagcg aguuuuuuu 99

<210> 143
<211> 91
<212> ARN
<213> Staphylococcus pseudintermedius

<400> 143
guuuuagcac uagaaauagu uaaguuaaaa caagcuuaaa gcgucaaugu aauauuuuau 60

uaacacccua cugugucagu gggguuuuuu u 91

<210> 144
<211> 94
<212> ARN
<213> Lactobacillus johnsonii

<400> 144
guuuuagaug gugaaaacca gauuuaaaau caagcaaugc aucuuuugau gcaaaguuuc 60

aauauuuguc ccacguuauc gagggacuuu uuuu 94

<210> 145
<211> 78
<212> ARN
<213> Eubacterium ventriosum

<400> 145
auuuuaguac cuggaaacag aucuacuaaa acaaggcuuu augccgaaau caagagcacc 60

gacgggugcu cuuuuuuu 78

<210> 146
<211> 84
<212> ARN
<213> Streptococcus pasteurianus

<400> 146
guuuuuguac ucgaaagagc cuacaaagau aaggcuuuau gccgaauuca agcaccccau 60

guuuugacau gaggugcuuu uuuu 84

<210> 147
<211> 85
<212> ARN
<213> Lactobacillus farciminis

<400> 147
guuuuuguac cuuagaaaua agaucuacaa aaauaaggau uuauuccgaa uuuaccaccu 60

auuuuaauua auaggugguu uuuuu 85

<210> 148
<211> 100
<212> ARN
<213> Bacteroides coprophilus

<400> 148
guugugauuu gcuuucauuu gaaaaauuga agcaaaucac aauaaggauu auuccguugu 60

gaaaacaauu aaagcggucu ugcaaaaggu cgcuuuuuuu 100

<210> 149
<211> 102
<212> ARN
<213> Flaviivola taffensis

<400> 149
guugugauuc gcuuucaaug aaaauugaag cgaaucacaa uaaggauuau uccguuguga 60

aaacauuuac uacggggcau cgaaagacug ccucguuuuu uu 102

<210> 150
<211> 90
<212> ARN
<213> Flavobacterium columnare

<400> 150
guuguggguuu gauuagaaau aaucuuauca caauaaggcu auaugccgua gacgaaaguc 60

uuuagucccg cuucgguggg acuuuuuuuu 90

<210> 151
<211> 78
<212> ARN
<213> Sphaerochaeta globus

<400> 151
guuggggaug accgcgaaag cgauuaucuc uaauaagacu uaagucgcaa aaugcucccu 60

auuuugggag cuuuuuuu 78

<210> 152
<211> 85
<212> ARN
<213> Azospirillum sp.

<400> 152
guugcggcug gagaaaucca gccguuaaca uguucccuuc ggggagcacg aaaugcgggg 60

cgggccacgg uccgccccuu uuuuu 85

<210> 153
<211> 126
<212> ARN
<213> Gluconacetobacter diazotrophicus

<400> 153
gccgugguuu ccgaaaggaa accacgguaa cagaauuacc guaagguuuu uucugugaag 60

gaucaucccu cgcuugggca accaggcggg ggaaauuccu cguucgggcc aaucagcccu 120

uuuuuu 126

<210> 154
<211> 82
<212> ARN
<213> Roseburia intestinalis

<400> 154
guuguaauuc ccugaaaagg uuauuacaau aagguaagaa accuaaaagc ucuaauccca 60

uucuucggaa ugggauuuuu uu 82

<210> 155
<211> 83
<212> ARN
<213> Staphylococcus aureus

<400> 155
guuuuaguac ucuggaaaca gaaucuacua aaacaaggca aaaugccgug uuuaucucgu 60

caacuuguug gcgagauuuu uuu 83

<210> 156
<211> 94
<212> ARN
<213> Mycoplasma gallisepticum

<400> 156
guuuuagcac uguacaagaa auugucgugc uaaaauaagg cgcuguuaau gcagcugccg 60

cauccgccag agcauuuaug cucuggcuuu uuuu 94

<210> 157
<211> 117
<212> ARN
<213> Campylobacter lari

<400> 157
guuuuagucu cugaaaagag acuaaaauaa gugguuuuug gucauccacg caggguuaca 60

aucccuuuaa aaccauuaaa auucaaauaa acuagguugu aucaacuuag uuuuuuu 117

<210> 158
<211> 119
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 158
nnnnnnnnnn nnnnnnnnnn acugggguuc aggaaacuga accucaguaa gcauuggcuc 60

guuuccaaug uugauugcuc cgccggugcu ccuuauuuuu aagggcgccg gcuuuuuuu 119

<210> 159
<211> 1084
<212> PRT
<213> Corynebacterium diphtheriae

<400> 159

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

```
Pro Leu Tyr Pro Trp Lys Val Arg Ala Glu Leu Ala Ala Ser Tyr Ile
            100                 105                 110

Ala Asp Glu Lys Glu Arg Gly Glu Lys Leu Ser Val Ala Leu Arg His
        115                 120                 125

Ile Ala Arg His Arg Gly Trp Arg Asn Pro Tyr Ala Lys Val Ser Ser
    130                 135                 140

Leu Tyr Leu Pro Asp Gly Pro Ser Asp Ala Phe Lys Ala Ile Arg Glu
145                 150                 155                 160

Glu Ile Lys Arg Ala Ser Gly Gln Pro Val Pro Glu Thr Ala Thr Val
                165                 170                 175

Gly Gln Met Val Thr Leu Cys Glu Leu Gly Thr Leu Lys Leu Arg Gly
            180                 185                 190

Glu Gly Gly Val Leu Ser Ala Arg Leu Gln Gln Ser Asp Tyr Ala Arg
        195                 200                 205

Glu Ile Gln Glu Ile Cys Arg Met Gln Glu Ile Gly Gln Glu Leu Tyr
    210                 215                 220

Arg Lys Ile Ile Asp Val Val Phe Ala Ala Glu Ser Pro Lys Gly Ser
225                 230                 235                 240

Ala Ser Ser Arg Val Gly Lys Asp Pro Leu Gln Pro Gly Lys Asn Arg
                245                 250                 255

Ala Leu Lys Ala Ser Asp Ala Phe Gln Arg Tyr Arg Ile Ala Ala Leu
            260                 265                 270

Ile Gly Asn Leu Arg Val Arg Val Asp Gly Glu Lys Arg Ile Leu Ser
        275                 280                 285

Val Glu Glu Lys Asn Leu Val Phe Asp His Leu Val Asn Leu Thr Pro
    290                 295                 300

Lys Lys Glu Pro Glu Trp Val Thr Ile Ala Glu Ile Leu Gly Ile Asp
305                 310                 315                 320

Arg Gly Gln Leu Ile Gly Thr Ala Thr Met Thr Asp Asp Gly Glu Arg
                325                 330                 335

Ala Gly Ala Arg Pro Pro Thr His Asp Thr Asn Arg Ser Ile Val Asn
            340                 345                 350
```

```
Ser Arg Ile Ala Pro Leu Val Asp Trp Trp Lys Thr Ala Ser Ala Leu
        355                 360                 365

Glu Gln His Ala Met Val Lys Ala Leu Ser Asn Ala Glu Val Asp Asp
    370                 375                 380

Phe Asp Ser Pro Glu Gly Ala Lys Val Gln Ala Phe Phe Ala Asp Leu
385                 390                 395                 400

Asp Asp Asp Val His Ala Lys Leu Asp Ser Leu His Leu Pro Val Gly
                405                 410                 415

Arg Ala Ala Tyr Ser Glu Asp Thr Leu Val Arg Leu Thr Arg Arg Met
            420                 425                 430

Leu Ser Asp Gly Val Asp Leu Tyr Thr Ala Arg Leu Gln Glu Phe Gly
        435                 440                 445

Ile Glu Pro Ser Trp Thr Pro Pro Thr Pro Arg Ile Gly Glu Pro Val
    450                 455                 460

Gly Asn Pro Ala Val Asp Arg Val Leu Lys Thr Val Ser Arg Trp Leu
465                 470                 475                 480

Glu Ser Ala Thr Lys Thr Trp Gly Ala Pro Glu Arg Val Ile Ile Glu
                485                 490                 495

His Val Arg Glu Gly Phe Val Thr Glu Lys Arg Ala Arg Glu Met Asp
            500                 505                 510

Gly Asp Met Arg Arg Arg Ala Ala Arg Asn Ala Lys Leu Phe Gln Glu
        515                 520                 525

Met Gln Glu Lys Leu Asn Val Gln Gly Lys Pro Ser Arg Ala Asp Leu
    530                 535                 540

Trp Arg Tyr Gln Ser Val Gln Arg Gln Asn Cys Gln Cys Ala Tyr Cys
545                 550                 555                 560

Gly Ser Pro Ile Thr Phe Ser Asn Ser Glu Met Asp His Ile Val Pro
                565                 570                 575

Arg Ala Gly Gln Gly Ser Thr Asn Thr Arg Glu Asn Leu Val Ala Val
            580                 585                 590

Cys His Arg Cys Asn Gln Ser Lys Gly Asn Thr Pro Phe Ala Ile Trp
```

595 600 605

Ala Lys Asn Thr Ser Ile Glu Gly Val Ser Val Lys Glu Ala Val Glu
610 615 620

Arg Thr Arg His Trp Val Thr Asp Thr Gly Met Arg Ser Thr Asp Phe
625 630 635 640

Lys Lys Phe Thr Lys Ala Val Val Glu Arg Phe Gln Arg Ala Thr Met
645 650 655

Asp Glu Glu Ile Asp Ala Arg Ser Met Glu Ser Val Ala Trp Met Ala
660 665 670

Asn Glu Leu Arg Ser Arg Val Ala Gln His Phe Ala Ser His Gly Thr
675 680 685

Thr Val Arg Val Tyr Arg Gly Ser Leu Thr Ala Glu Ala Arg Arg Ala
690 695 700

Ser Gly Ile Ser Gly Lys Leu Lys Phe Phe Asp Gly Val Gly Lys Ser
705 710 715 720

Arg Leu Asp Arg Arg His His Ala Ile Asp Ala Ala Val Ile Ala Phe
725 730 735

Thr Ser Asp Tyr Val Ala Glu Thr Leu Ala Val Arg Ser Asn Leu Lys
740 745 750

Gln Ser Gln Ala His Arg Gln Glu Ala Pro Gln Trp Arg Glu Phe Thr
755 760 765

Gly Lys Asp Ala Glu His Arg Ala Ala Trp Arg Val Trp Cys Gln Lys
770 775 780

Met Glu Lys Leu Ser Ala Leu Leu Thr Glu Asp Leu Arg Asp Asp Arg
785 790 795 800

Val Val Val Met Ser Asn Val Arg Leu Arg Leu Gly Asn Gly Ser Ala
805 810 815

His Lys Glu Thr Ile Gly Lys Leu Ser Lys Val Lys Leu Ser Ser Gln
820 825 830

Leu Ser Val Ser Asp Ile Asp Lys Ala Ser Ser Glu Ala Leu Trp Cys
835 840 845

```
Ala Leu Thr Arg Glu Pro Gly Phe Asp Pro Lys Glu Gly Leu Pro Ala
    850                 855                 860

Asn Pro Glu Arg His Ile Arg Val Asn Gly Thr His Val Tyr Ala Gly
865                 870                 875                 880

Asp Asn Ile Gly Leu Phe Pro Val Ser Ala Gly Ser Ile Ala Leu Arg
                885                 890                 895

Gly Gly Tyr Ala Glu Leu Gly Ser Ser Phe His His Ala Arg Val Tyr
            900                 905                 910

Lys Ile Thr Ser Gly Lys Lys Pro Ala Phe Ala Met Leu Arg Val Tyr
        915                 920                 925

Thr Ile Asp Leu Leu Pro Tyr Arg Asn Gln Asp Leu Phe Ser Val Glu
    930                 935                 940

Leu Lys Pro Gln Thr Met Ser Met Arg Gln Ala Glu Lys Lys Leu Arg
945                 950                 955                 960

Asp Ala Leu Ala Thr Gly Asn Ala Glu Tyr Leu Gly Trp Leu Val Val
                965                 970                 975

Asp Asp Glu Leu Val Val Asp Thr Ser Lys Ile Ala Thr Asp Gln Val
            980                 985                 990

Lys Ala Val Glu Ala Glu Leu Gly  Thr Ile Arg Arg Trp  Arg Val Asp
        995                 1000                1005

Gly Phe  Phe Ser Pro Ser Lys  Leu Arg Leu Arg Pro  Leu Gln Met
    1010                1015                1020

Ser Lys  Glu Gly Ile Lys Lys  Glu Ser Ala Pro Glu  Leu Ser Lys
    1025                1030                1035

Ile Ile  Asp Arg Pro Gly Trp  Leu Pro Ala Val Asn  Lys Leu Phe
    1040                1045                1050

Ser Asp  Gly Asn Val Thr Val  Val Arg Arg Asp Ser  Leu Gly Arg
    1055                1060                1065

Val Arg  Leu Glu Ser Thr Ala  His Leu Pro Val Thr  Trp Lys Val
    1070                1075                1080

Gln
```

<210> 160
<211> 3336
<212> ADN
<213> Secuencia Artificial

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

<400> 160

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgaa gtaccatgtc ggaatcgatg tcggaacctt ttctgtgggg 120

ctggctgcta ttgaagtgga tgacgctgga atgcctatta agaccctgag tctggtgtca 180

cacattcatg actcaggact ggatcctgac gagatcaaga gcgctgtgac caggctggca 240

agctccggaa tcgcccggag aacaaggcgc ctgtaccgac ggaagagaag gcgcctgcag 300

cagctggata agttcatcca gaggcagggc tggccagtga tcgagctgga agattacagc 360

gacccccctgt atccttggaa ggtgcgcgcc gaactggccg cttcttatat tgctgacgag 420

aaggaacggg gggagaaact gagtgtggct ctgagacaca tcgcaaggca tcgcggatgg 480

aggaaccctt acgccaaggt gtctagtctg tatctgccag atggcccctc agacgccttc 540

aaggctatta gggaggaaat caaacgcgct agcggccagc ctgtgccaga gactgcaacc 600

gtcgggcaga tggtgaccct gtgcgaactg ggcacactga agctgcgagg agagggagga 660

gtgctgagtg cacggctgca gcagtcagat tacgcccgcg agatccagga aatttgtcga 720

atgcaggaga tcggccagga actgtatcgc aagatcattg acgtggtgtt cgcagccgag 780

tccccaaagg gctctgcctc aagccgggtg gggaaagatc ctctgcagcc aggaaagaac 840

agagcactga aagccagcga cgcttttcag cgataccgga ttgctgcact gatcggcaat 900

ctgagagtca gggtggatgg ggagaagagg attctgagcg tggaggagaa gaacctggtg 960

ttcgaccacc tggtgaatct gactccaaag aaagagcccg aatgggtgac catcgccgaa 1020

attctgggca tcgatcgcgg gcagctgatc ggaacagcta ctatgaccga cgatggagag 1080

cgagcaggag cccgaccccc tacacacgat actaacagaa gtattgtgaa cagccggatc 1140

gcaccactgg tcgactggtg gaaaacagct agcgcactgg agcagcacgc catggtgaag 1200

gcactgtcca acgccgaagt cgacgatttt gattctcccg agggagcaaa agtgcaggca 1260

ttctttgccg atctggacga tgacgtccac gccaagctgg acagcctgca tctgcctgtg 1320

ggacgagcag cttactccga ggacactctg gtcagactga cccgacggat gctgagtgat 1380

ggggtggacc tgtataccgc ccggctgcag gagttcggaa ttgaacctag ctggacccca 1440

cccacaccaa gaatcggaga gcctgtcggc aatccagccg tcgaccgggt gctgaaaaca 1500

gtgagcagat ggctggaatc cgcaacaaag acttggggcg ccccagagag ggtcatcatt 1560

gagcacgtgc gcgaaggctt cgtcactgag aaacgcgctc gagaaatgga tggggacatg 1620

agaaggcgcg cagcccggaa cgccaagctg tttcaggaga tgcaggaaaa gctgaatgtg 1680

cagggcaaac ccagtcgagc cgatctgtgg agataccagt cagtgcagag acagaactgc 1740

cagtgtgcct attgcgggtc cccaattacc ttttctaata gtgaaatgga ccacatcgtg 1800

cccagagcag ggcagggatc caccaacaca agggagaatc tggtcgccgt gtgccatcgc 1860

tgtaaccagt ctaagggcaa tacacccttc gctatttggg caaaaaacac ttctatcgaa 1920

ggggtcagtg tgaaggaggc cgtggaacgg accagacatt gggtcactga taccggcatg 1980

agaagcactg acttcaagaa gttcaccaag gctgtggtcg agcggtttca gagagcaaca 2040

atggatgagg aaatcgacgc cagaagcatg gaatccgtcg cctggatggc taatgagctg 2100

aggagccgcg tggctcagca cttcgcatcc catggaacca cagtcagggt gtaccgaggc 2160

agcctgacag cagaggctcg acgggcatct gggatcagtg gaaagctgaa attctttgat 2220

ggcgtgggga agtccaggct ggatagaagg caccatgcta ttgacgctgc agtgatcgca 2280

ttcacctctg actatgtggc cgaaacactg gctgtccgct caaacctgaa acagagccag 2340

gcccaccgac aggaggctcc tcagtggaga gagttcaccg gcaaggatgc agagcatcga 2400

gcagcttgga gagtgtggtg ccagaagatg gaaaaactga gcgccctgct gaccgaggac 2460

ctgcgagatg accgggtggt cgtgatgtct aacgtgcgac tgcggctggg aaatggcagt 2520

gcccacaagg aaaccattgg caaactgtca aaggtgaaac tgtcctctca gctgtcagtc 2580

agcgatatcg acaaagcaag ttcagaggcc ctgtggtgtg ctctgaccag agagcccgga 2640

ttcgatccta aggaaggcct gcccgctaac cctgagagac acatcagggt gaatgggaca 2700

catgtctacg ccggggacaa tattggactg tttccagtgt cagcaggaag catcgcactg 2760

aggggaggat acgcagagct gggcagctcc ttccaccatg ctcgcgtgta taaaattact 2820

tccggcaaga aacccgcatt tgccatgctg agggtgtaca ccatcgatct gctgccttat 2880

cgcaaccagg acctgtttag cgtggaactg aagccacaga caatgtccat gaggcaggct 2940

gagaagaaac tgcgcgacgc tctggcaact gggaatgcag aatatctggg atggctggtc 3000

gtggatgacg agctggtcgt ggatacatct aagattgcca ctgaccaggt caaagcagtg 3060

gaggccgaac tggggactat ccgccgatgg cgggtggatg gattctttc cccctctaaa 3120

ctgagactga ggcctctgca gatgtccaag gaggggatca agaaagagtc cgctcccgaa 3180

ctgtctaaaa tcattgacag accaggatgg ctgcccgccg tgaacaagct gttctctgat 3240

ggaaatgtca ccgtcgtgcg gagagactct ctgggacgcg tgcgactgga gagtacagcc 3300

cacctgcctg tcacttggaa ggtgcagtaa gaattc 3336

<210> 161
<211> 175
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 161
nnnnnnnnnn nnnnnnnnnn guuucagugc uauaggaaac uauaggaaau caccuucggg 60

ugagcugaaa uccccuaaag cuaagauuga auccggccac uaucuauuag uagauauccg 120

gauauucuga uauaaaaccu cauucuuuga uuagaccaaa ggaugagguu uuuuu 175

<210> 162
<211> 1422
<212> PRT
<213> Sutterella wadsworthensis

<400> 162
Met Thr Gln Ser Glu Arg Arg Phe Ser Cys Ser Ile Gly Ile Asp Met
1 5 10 15

Gly Ala Lys Tyr Thr Gly Val Phe Tyr Ala Leu Phe Asp Arg Glu Glu
20 25 30

Leu Pro Thr Asn Leu Asn Ser Lys Ala Met Thr Leu Val Met Pro Glu
35 40 45

Thr Gly Pro Arg Tyr Val Gln Ala Gln Arg Thr Ala Val Arg His Arg
50 55 60

Leu Arg Gly Gln Lys Arg Tyr Thr Leu Ala Arg Lys Leu Ala Phe Leu
65 70 75 80

Val Val Asp Asp Met Ile Lys Lys Gln Glu Lys Arg Leu Thr Asp Glu
85 90 95

Glu Trp Lys Arg Gly Arg Glu Ala Leu Ser Gly Leu Leu Lys Arg Arg
100 105 110

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

Asn Val Arg Ala Asp Val Phe Ala Ala His Pro Ala Phe Ser Thr Tyr
130 135 140

```
Phe Ser Glu Val Arg Ser Leu Ala Glu Gln Trp Glu Glu Phe Thr Ala
145                 150                 155                 160

Asn Ile Ser Asn Val Glu Lys Phe Leu Gly Asp Pro Asn Ile Pro Ala
                165                 170                 175

Asp Lys Glu Phe Ile Glu Phe Ala Val Ala Glu Gly Leu Ile Asp Lys
            180                 185                 190

Thr Glu Lys Lys Ala Tyr Gln Ser Ala Leu Ser Thr Leu Arg Ala Asn
        195                 200                 205

Ala Asn Val Leu Thr Gly Leu Arg Gln Met Gly His Lys Pro Arg Ser
    210                 215                 220

Glu Tyr Phe Lys Ala Ile Glu Ala Asp Leu Lys Lys Asp Ser Arg Leu
225                 230                 235                 240

Ala Lys Ile Asn Glu Ala Phe Gly Gly Ala Glu Arg Leu Ala Arg Leu
                245                 250                 255

Leu Gly Asn Leu Ser Asn Leu Gln Leu Arg Ala Glu Arg Trp Tyr Phe
            260                 265                 270

Asn Ala Pro Asp Ile Met Lys Asp Arg Gly Trp Glu Pro Asp Arg Phe
        275                 280                 285

Lys Lys Thr Leu Val Arg Ala Phe Lys Phe Phe His Pro Ala Lys Asp
    290                 295                 300

Gln Asn Lys Gln His Leu Glu Leu Ile Lys Gln Ile Glu Asn Ser Glu
305                 310                 315                 320

Asp Ile Ile Glu Thr Leu Cys Thr Leu Asp Pro Asn Arg Thr Ile Pro
                325                 330                 335

Pro Tyr Glu Asp Gln Asn Asn Arg Arg Pro Pro Leu Asp Gln Thr Leu
            340                 345                 350

Leu Leu Ser Pro Glu Lys Leu Thr Arg Gln Tyr Gly Glu Ile Trp Lys
        355                 360                 365

Thr Trp Ser Ala Arg Leu Thr Ser Ala Glu Pro Thr Leu Ala Pro Ala
    370                 375                 380

Ala Glu Ile Leu Glu Arg Ser Thr Asp Arg Lys Ser Arg Val Ala Val
```

385 390 395 400

Asn Gly His Glu Pro Leu Pro Thr Leu Ala Tyr Gln Leu Ser Tyr Ala
405 410 415

Leu Gln Arg Ala Phe Asp Arg Ser Lys Ala Leu Asp Pro Tyr Ala Leu
420 425 430

Arg Ala Leu Ala Ala Gly Ser Lys Ser Asn Lys Leu Thr Ser Ala Arg
435 440 445

Thr Ala Leu Glu Asn Cys Ile Gly Gly Gln Asn Val Lys Thr Phe Leu
450 455 460

Asp Cys Ala Arg Arg Tyr Tyr Arg Glu Ala Asp Asp Ala Lys Val Gly
465 470 475 480

Leu Trp Phe Asp Asn Ala Asp Gly Leu Leu Glu Arg Ser Asp Leu His
485 490 495

Pro Pro Met Lys Lys Lys Ile Leu Pro Leu Leu Val Ala Asn Ile Leu
500 505 510

Gln Thr Asp Glu Thr Thr Gly Gln Lys Phe Leu Asp Glu Ile Trp Arg
515 520 525

Lys Gln Ile Lys Gly Arg Glu Thr Val Ala Ser Arg Cys Ala Arg Ile
530 535 540

Glu Thr Val Arg Lys Ser Phe Gly Gly Gly Phe Asn Ile Ala Tyr Asn
545 550 555 560

Thr Ala Gln Tyr Arg Glu Val Asn Lys Leu Pro Arg Asn Ala Gln Asp
565 570 575

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

Ile Ala Ala Asn Leu Gly Leu Ser Asp Glu Gln Lys Arg Lys Phe Ala
595 600 605

Asn Pro Phe Ser Leu Ala Gln Phe Tyr Thr Leu Ile Glu Thr Glu Val
610 615 620

Ser Gly Phe Ser Ala Thr Thr Leu Ala Val His Leu Glu Asn Ala Trp
625 630 635 640

```
Arg Met Thr Ile Lys Asp Ala Val Ile Asn Gly Glu Thr Val Arg Ala
                645                 650                 655

Ala Gln Cys Ser Arg Leu Pro Ala Glu Thr Ala Arg Pro Phe Asp Gly
            660                 665                 670

Leu Val Arg Arg Leu Val Asp Arg Gln Ala Trp Glu Ile Ala Lys Arg
        675                 680                 685

Val Ser Thr Asp Ile Gln Ser Lys Val Asp Phe Ser Asn Gly Ile Val
    690                 695                 700

Asp Val Ser Ile Phe Val Glu Glu Asn Lys Phe Glu Phe Ser Ala Ser
705                 710                 715                 720

Val Ala Asp Leu Lys Lys Asn Lys Arg Val Lys Asp Lys Met Leu Ser
                725                 730                 735

Glu Ala Glu Lys Leu Glu Thr Arg Trp Leu Ile Lys Asn Glu Arg Ile
            740                 745                 750

Lys Lys Ala Ser Arg Gly Thr Cys Pro Tyr Thr Gly Asp Arg Leu Ala
        755                 760                 765

Glu Gly Gly Glu Ile Asp His Ile Leu Pro Arg Ser Leu Ile Lys Asp
    770                 775                 780

Ala Arg Gly Ile Val Phe Asn Ala Glu Pro Asn Leu Ile Tyr Ala Ser
785                 790                 795                 800

Ser Arg Gly Asn Gln Leu Lys Lys Asn Gln Arg Tyr Ser Leu Ser Asp
                805                 810                 815

Leu Lys Ala Asn Tyr Arg Asn Glu Ile Phe Lys Thr Ser Asn Ile Ala
            820                 825                 830

Ala Ile Thr Ala Glu Ile Glu Asp Val Val Thr Lys Leu Gln Gln Thr
        835                 840                 845

His Arg Leu Lys Phe Phe Asp Leu Leu Asn Glu His Glu Gln Asp Cys
    850                 855                 860

Val Arg His Ala Leu Phe Leu Asp Asp Gly Ser Glu Ala Arg Asp Ala
865                 870                 875                 880

Val Leu Glu Leu Leu Ala Thr Gln Arg Arg Thr Arg Val Asn Gly Thr
                885                 890                 895
```

```
Gln Ile Trp Met Ile Lys Asn Leu Ala Asn Lys Ile Arg Glu Glu Leu
            900                 905                 910

Gln Asn Trp Cys Lys Thr Thr Asn Asn Arg Leu His Phe Gln Ala Ala
        915                 920                 925

Ala Thr Asn Val Ser Asp Ala Lys Asn Leu Arg Leu Lys Leu Ala Gln
    930                 935                 940

Asn Gln Pro Asp Phe Glu Lys Pro Asp Ile Gln Pro Ile Ala Ser His
945                 950                 955                 960

Ser Ile Asp Ala Leu Cys Ser Phe Ala Val Gly Ser Ala Asp Ala Glu
                965                 970                 975

Arg Asp Gln Asn Gly Phe Asp Tyr Leu Asp Gly Lys Thr Val Leu Gly
            980                 985                 990

Leu Tyr Pro Gln Ser Cys Glu Val  Ile His Leu Gln Ala  Lys Pro Gln
        995                 1000                1005

Glu Glu  Lys Ser His Phe Asp  Ser Val Ala Ile Phe  Lys Glu Gly
    1010                1015                1020

Ile Tyr  Ala Glu Gln Phe Leu  Pro Ile Phe Thr Leu  Asn Glu Lys
    1025                1030                1035

Ile Trp  Ile Gly Tyr Glu Thr  Leu Asn Ala Lys Gly  Glu Arg Cys
    1040                1045                1050

Gly Ala  Ile Glu Val Ser Gly  Lys Gln Pro Lys Glu  Leu Leu Glu
    1055                1060                1065

Met Leu  Ala Pro Phe Phe Asn  Lys Pro Val Gly Asp  Leu Ser Ala
    1070                1075                1080

His Ala  Thr Tyr Arg Ile Leu  Lys Lys Pro Ala Tyr  Glu Phe Leu
    1085                1090                1095

Ala Lys  Ala Ala Leu Gln Pro  Leu Ser Ala Glu Glu  Lys Arg Leu
    1100                1105                1110

Ala Ala  Leu Leu Asp Ala Leu  Arg Tyr Cys Thr Ser  Arg Lys Ser
    1115                1120                1125

Leu Met  Ser Leu Phe Met Ala  Ala Asn Gly Lys Ser  Leu Lys Lys
    1130                1135                1140
```

```
Arg Glu  Asp Val Leu Lys Pro  Lys Leu Phe Gln Leu  Lys Val Glu
    1145                 1150                 1155

Leu Lys  Gly Glu Lys Ser Phe  Lys Leu Asn Gly Ser  Leu Thr Leu
    1160                 1165                 1170

Pro Val  Lys Gln Asp Trp Leu  Arg Ile Cys Asp Ser  Pro Glu Leu
    1175                 1180                 1185

Ala Asp  Ala Phe Gly Lys Pro  Cys Ser Ala Asp Glu  Leu Thr Ser
    1190                 1195                 1200

Lys Leu  Ala Arg Ile Trp Lys  Arg Pro Val Met Arg  Asp Leu Ala
    1205                 1210                 1215

His Ala  Pro Val Arg Arg Glu  Phe Ser Leu Pro Ala  Ile Asp Asn
    1220                 1225                 1230

Pro Ser  Gly Gly Phe Arg Ile  Arg Arg Thr Asn Leu  Phe Gly Asn
    1235                 1240                 1245

Glu Leu  Tyr Gln Val His Ala  Ile Asn Ala Lys Lys  Tyr Arg Gly
    1250                 1255                 1260

Phe Ala  Ser Ala Gly Ser Asn  Val Asp Trp Ser Lys  Gly Ile Leu
    1265                 1270                 1275

Phe Asn  Glu Leu Gln His Glu  Asn Leu Thr Glu Cys  Gly Gly Arg
    1280                 1285                 1290

Phe Ile  Thr Ser Ala Asp Val  Thr Pro Met Ser Glu  Trp Arg Lys
    1295                 1300                 1305

Val Val  Ala Glu Asp Asn Leu  Ser Ile Trp Ile Ala  Pro Gly Thr
    1310                 1315                 1320

Glu Gly  Arg Arg Tyr Val Arg  Val Glu Thr Thr Phe  Ile Gln Ala
    1325                 1330                 1335

Ser His  Trp Phe Glu Gln Ser  Val Glu Asn Trp Ala  Ile Thr Ser
    1340                 1345                 1350

Pro Leu  Ser Leu Pro Ala Ser  Phe Lys Val Asp Lys  Pro Ala Glu
    1355                 1360                 1365

Phe Gln  Lys Ala Val Gly Thr  Glu Leu Ser Glu Leu  Leu Gly Gln
```

```
    1370                1375                1380

Pro Arg  Ser Glu Ile Phe Ile  Glu Asn Val Gly Asn  Ala Lys His
    1385                1390                1395

Ile Arg  Phe Trp Tyr Ile Val  Val Ser Ser Asn Lys  Lys Met Asn
    1400                1405                1410

Glu Ser  Tyr Asn Asn Val Ser  Lys Ser
    1415                1420
```

<210> 163
<211> 4350
<212> ADN
<213> Secuencia Artificial

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

<400> 163

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgac tcagagcgag cgacgatttt cttgcagcat tggcattgac      120

atgggggcta agtacactgg ggtgttctac gcactgttcg accgggagga actgcccaca      180

aacctgaaca gcaaggccat gaccctggtc atgcctgaga cagggccaag atacgtgcag      240

gcacagagaa ctgccgtcag acacaggctg cgcggacaga agagatatac cctggctagg      300

aaactggcat ttctggtggt cgacgatatg atcaagaaac aggaaaagag gctgactgat      360

gaggaatgga aacgaggacg ggaggccctg tccggcctgc tgaagcggag agggtactct      420

cggcccaacg ctgacggcga agatctgacc cctctggaga atgtgagagc agacgtgttc      480

gccgctcatc ctgccttcag cacatatttt tccgaagtgc gctctctggc tgagcagtgg      540

gaggagttca ccgcaaacat cagcaatgtc gagaagtttc tgggcgaccc aaacatcccc      600

gccgataaag agttcattga atttgccgtg gctgaagggc tgattgacaa gaccgagaag      660

aaagcctacc agtcagctct gagcaccctg agggcaaacg ccaatgtgct gacaggactg      720

cggcagatgg gccacaagcc tagatcagaa tattttaaag caatcgaggc cgacctgaag      780

aaagatagcc gcctggccaa gattaacgaa gcattcggag gagcagagcg cctggctcga      840

ctgctgggaa acctgtccaa tctgcagctg cgggcagaaa gatggtactt caatgccccc      900

gacatcatga aggatagggg ctgggagcct gatcgcttca agaaaacact ggtgcgggct      960

tttaagttct ttcacccagc aaaggaccag aacaaacagc atctggaact gatcaaacag     1020

attgagaaca gcgaagatat cattgagact ctgtgcaccc tggacccaaa cagaaccatc     1080

ccccccttacg aggatcagaa caataggcgc ccacccctgg accagactct gctgctgagt     1140
```

cccgaaaagc tgacccggca gtatggcgag atctggaaaa catggagcgc cagactgacc 1200

tccgctgaac ccacactggc acctgcagcc gagattctgg aaagatctac cgacaggaag 1260

agtcgcgtgg cagtcaacgg acacgagcca ctgcctacac tggcttacca gctgagttat 1320

gcactgcaga gagccttcga caggtcaaaa gccctggatc catatgctct gagggcactg 1380

gctgcaggct caaaaagcaa taagctgaca tccgcccgca ctgctctgga gaactgcatc 1440

ggaggccaga atgtgaaaac cttcctggac tgtgcccgac ggtactatcg ggaagcagac 1500

gatgccaaag tcgggctgtg gttcgacaac gccgatggac tgctggagag atctgacctg 1560

catcctccaa tgaagaaaaa gatcctgccc ctgctggtgg ccaatattct gcagacagat 1620

gaaaccacag gccagaagtt tctggacgag atctggcgaa aacagattaa ggggcgggaa 1680

actgtggcta gccgatgtgc acggatcgag acagtgcgga aatccttcgg gggaggcttt 1740

aacattgcct acaataccgc tcagtatagg gaggtgaaca agctgccccg caatgcccag 1800

gataaagaac tgctgacaat cagagatagg gtggctgaga ctgcagactt cattgccgct 1860

aacctggggc tgtctgacga gcagaaaaga aagttcgcca atccttttag tctggctcag 1920

ttctacaccc tgatcgagac agaagtgtcc ggattttctg caactaccct ggccgtccac 1980

ctggagaacg cctggaggat gacaatcaag gatgctgtga ttaatgggga aactgtcaga 2040

gcagcacagt gcagcaggct gcctgcagag acagctcgcc cattcgatgg actggtgaga 2100

aggctggtcg acagacaggc ttgggagatc gcaaagaggg tgtcaactga cattcagagc 2160

aaagtcgatt tctccaacgg catcgtggac gtcagcattt ttgtggagga aaataagttc 2220

gagttttccg catctgtggc cgatctgaaa aagaacaaac gggtcaaaga caagatgctg 2280

tccgaggccg aaaagctgga aaccagatgg ctgatcaaaa atgagcggat caagaaggcc 2340

agccggggaa cttgtcccta caccggcgat aggctggctg agggggggaga aatcgaccac 2400

attctgcccc gaagcctgat caaggatgcc cggggaattg tgtttaacgc tgagcctaat 2460

ctgatctatg caagctcccg cggcaaccag ctgaaaaaga atcagcgata cagtctgtca 2520

gatctgaagg ccaactatcg gaatgagatc ttcaaaacta gcaacatcgc tgcaattacc 2580

gccgagattg aggacgtggt cactaagctg cagcagaccc atagactgaa attctttgat 2640

ctgctgaatg agcacgaaca ggactgcgtg cggcacgccc tgttcctgga cgatggcagc 2700

gaagctcgcg acgcagtgct ggagctgctg gcaacacagc gccgaactcg cgtcaacggg 2760

acacagatct ggatgattaa gaacctggcc aacaagatcc gagaggaact gcagaattgg 2820

tgtaagacaa ctaacaatag actgcacttt caggccgctg caactaacgt gtccgatgca 2880

aagaatctga ggctgaaact ggcccagaac cagcccgact tcgagaagcc agatatccag 2940

cccattgcca gccattccat cgacgccctg tgctctttcg ctgtggggag tgctgacgca 3000

```
gaacgcgatc agaatggatt tgactacctg gatggcaaga ccgtgctggg actgtatcca     3060
cagagctgtg aggtcattca cctgcaggcc aagccccagg aggaaaaaag tcatttcgat     3120
tcagtggcta tctttaagga aggcatctac gcagagcagt tcctgcctat ctttaccctg     3180
aacgaaaaga tctggattgg atatgagaca ctgaatgcca aaggcgaaag atgcggggct     3240
attgaggtga gtggcaaaca gccaaaggag ctgctggaaa tgctggcccc cttctttaac     3300
aagcctgtgg gcgacctgtc agcccacgct acttaccgga tcctgaaaaa gcctgcatat     3360
gagtttctgg caaaggcagc tctgcagcca ctgagcgcag aggaaaaaag actggcagcc     3420
ctgctggatg ctctgcgcta ctgtaccagt cgaaagtcac tgatgagcct gttcatggct     3480
gcaaacggga aatccctgaa aaagcgggag gacgtgctga aacccaagct gttccagctg     3540
aaggtcgagc tgaaaggcga aaagagcttc aagctgaacg ggagcctgac cctgcctgtg     3600
aagcaggact ggctgagaat ctgcgatagc ccagaactgg cagacgcctt tggcaaaccc     3660
tgttccgccg atgagctgac atctaagctg gctcgcattt ggaaacgacc tgtgatgcgg     3720
gatctggctc atgcaccagt ccggagagag ttcagcctgc ccgcaatcga caacccaagt     3780
ggagggttca ggattaggcg caccaacctg tttggcaatg agctgtacca ggtgcacgcc     3840
atcaacgcta aaaagtatcg cggcttcgcc tccgctgggt ctaatgtcga ctggtccaag     3900
gggatcctgt ttaacgagct gcagcatgaa aatctgaccg agtgcggagg caggttcatt     3960
acaagcgccg atgtgactcc tatgtccgaa tggcgcaagg tggtcgcaga ggacaacctg     4020
tctatctgga ttgctccagg gacagaagga cgacggtacg tgagggtcga gacaacattc     4080
atccaggcca gtcactggtt tgaacagtca gtggagaatt gggccattac tagtcctctg     4140
tcactgccag cttccttcaa ggtggacaaa ccagctgagt ttcagaaggc agtcggaacc     4200
gagctgtcag aactgctggg ccagcccagg agcgaaatct tcattgagaa cgtgggcaat     4260
gccaagcata tccgcttttg gtacattgtg gtgagcagca acaaaaagat gaacgagtct     4320
tacaacaatg tgtctaagag ttaagaattc                                     4350
```

<210> 164
<211> 112
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 164
nnnnnnnnnn nnnnnnnnnn guuucagugg ugaaaacccu gaaaucaaca aaauuaaaga 60

uugaaucguu uuucuaugcu cgucuuaaua gcgagcauau aacgauuuuu uu 112

<210> 165
<211> 1372
<212> PRT
<213> Legionella pneumophila

<400> 165

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

```
Gln Leu Asp Ser Ile Lys Lys Lys Ile Pro Ser Val Cys Leu Ser Asn
    210                 215                 220

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

Leu Ala Lys Asn Pro Lys Gln Phe Asp Glu Gln Thr Phe Gly Asn Glu
                245                 250                 255

Phe Leu Arg Met Leu Lys Asn Phe Arg His Leu Lys Gly Ser Gln Glu
            260                 265                 270

Ser Leu Ala Val Arg Asn Leu Ile Gln Gln Leu Glu Gln Ser Gln Asp
        275                 280                 285

Tyr Ile Ser Ile Leu Glu Lys Thr Pro Pro Glu Ile Thr Ile Pro Pro
    290                 295                 300

Tyr Glu Ala Arg Thr Asn Thr Gly Met Glu Lys Asp Gln Ser Leu Leu
305                 310                 315                 320

Leu Asn Pro Glu Lys Leu Asn Asn Leu Tyr Pro Asn Trp Arg Asn Leu
                325                 330                 335

Ile Pro Gly Ile Ile Asp Ala His Pro Phe Leu Glu Lys Asp Leu Glu
            340                 345                 350

His Thr Lys Leu Arg Asp Arg Lys Arg Ile Ile Ser Pro Ser Lys Gln
        355                 360                 365

Asp Glu Lys Arg Asp Ser Tyr Ile Leu Gln Arg Tyr Leu Asp Leu Asn
    370                 375                 380

Lys Lys Ile Asp Lys Phe Lys Ile Lys Lys Gln Leu Ser Phe Leu Gly
385                 390                 395                 400

Gln Gly Lys Gln Leu Pro Ala Asn Leu Ile Glu Thr Gln Lys Glu Met
                405                 410                 415

Glu Thr His Phe Asn Ser Ser Leu Val Ser Val Leu Ile Gln Ile Ala
            420                 425                 430

Ser Ala Tyr Asn Lys Glu Arg Glu Asp Ala Ala Gln Gly Ile Trp Phe
        435                 440                 445

Asp Asn Ala Phe Ser Leu Cys Glu Leu Ser Asn Ile Asn Pro Pro Arg
    450                 455                 460
```

```
Lys Gln Lys Ile Leu Pro Leu Leu Val Gly Ala Ile Leu Ser Glu Asp
465                 470                 475                 480

Phe Ile Asn Asn Lys Asp Lys Trp Ala Lys Phe Lys Ile Phe Trp Asn
                485                 490                 495

Thr His Lys Ile Gly Arg Thr Ser Leu Lys Ser Lys Cys Lys Glu Ile
            500                 505                 510

Glu Glu Ala Arg Lys Asn Ser Gly Asn Ala Phe Lys Ile Asp Tyr Glu
        515                 520                 525

Glu Ala Leu Asn His Pro Glu His Ser Asn Asn Lys Ala Leu Ile Lys
    530                 535                 540

Ile Ile Gln Thr Ile Pro Asp Ile Ile Gln Ala Ile Gln Ser His Leu
545                 550                 555                 560

Gly His Asn Asp Ser Gln Ala Leu Ile Tyr His Asn Pro Phe Ser Leu
                565                 570                 575

Ser Gln Leu Tyr Thr Ile Leu Glu Thr Lys Arg Asp Gly Phe His Lys
            580                 585                 590

Asn Cys Val Ala Val Thr Cys Glu Asn Tyr Trp Arg Ser Gln Lys Thr
        595                 600                 605

Glu Ile Asp Pro Glu Ile Ser Tyr Ala Ser Arg Leu Pro Ala Asp Ser
    610                 615                 620

Val Arg Pro Phe Asp Gly Val Leu Ala Arg Met Met Gln Arg Leu Ala
625                 630                 635                 640

Tyr Glu Ile Ala Met Ala Lys Trp Glu Gln Ile Lys His Ile Pro Asp
                645                 650                 655

Asn Ser Ser Leu Leu Ile Pro Ile Tyr Leu Glu Gln Asn Arg Phe Glu
            660                 665                 670

Phe Glu Glu Ser Phe Lys Lys Ile Lys Gly Ser Ser Ser Asp Lys Thr
        675                 680                 685

Leu Glu Gln Ala Ile Glu Lys Gln Asn Ile Gln Trp Glu Glu Lys Phe
    690                 695                 700

Gln Arg Ile Ile Asn Ala Ser Met Asn Ile Cys Pro Tyr Lys Gly Ala
```

705 710 715 720

Ser Ile Gly Gly Gln Gly Glu Ile Asp His Ile Tyr Pro Arg Ser Leu
725 730 735

Ser Lys Lys His Phe Gly Val Ile Phe Asn Ser Glu Val Asn Leu Ile
740 745 750

Tyr Cys Ser Ser Gln Gly Asn Arg Glu Lys Lys Glu Glu His Tyr Leu
755 760 765

Leu Glu His Leu Ser Pro Leu Tyr Leu Lys His Gln Phe Gly Thr Asp
770 775 780

Asn Val Ser Asp Ile Lys Asn Phe Ile Ser Gln Asn Val Ala Asn Ile
785 790 795 800

Lys Lys Tyr Ile Ser Phe His Leu Leu Thr Pro Glu Gln Gln Lys Ala
805 810 815

Ala Arg His Ala Leu Phe Leu Asp Tyr Asp Asp Glu Ala Phe Lys Thr
820 825 830

Ile Thr Lys Phe Leu Met Ser Gln Gln Lys Ala Arg Val Asn Gly Thr
835 840 845

Gln Lys Phe Leu Gly Lys Gln Ile Met Glu Phe Leu Ser Thr Leu Ala
850 855 860

Asp Ser Lys Gln Leu Gln Leu Glu Phe Ser Ile Lys Gln Ile Thr Ala
865 870 875 880

Glu Glu Val His Asp His Arg Glu Leu Leu Ser Lys Gln Glu Pro Lys
885 890 895

Leu Val Lys Ser Arg Gln Gln Ser Phe Pro Ser His Ala Ile Asp Ala
900 905 910

Thr Leu Thr Met Ser Ile Gly Leu Lys Glu Phe Pro Gln Phe Ser Gln
915 920 925

Glu Leu Asp Asn Ser Trp Phe Ile Asn His Leu Met Pro Asp Glu Val
930 935 940

His Leu Asn Pro Val Arg Ser Lys Glu Lys Tyr Asn Lys Pro Asn Ile
945 950 955 960

```
Ser Ser Thr Pro Leu Phe Lys Asp Ser Leu Tyr Ala Glu Arg Phe Ile
                965                 970                 975

Pro Val Trp Val Lys Gly Glu Thr Phe Ala Ile Gly Phe Ser Glu Lys
            980                 985                 990

Asp Leu Phe Glu Ile Lys Pro Ser  Asn Lys Glu Lys Leu  Phe Thr Leu
        995                 1000                1005

Leu Lys  Thr Tyr Ser Thr Lys  Asn Pro Gly Glu Ser  Leu Gln Glu
    1010                1015                1020

Leu Gln  Ala Lys Ser Lys Ala  Lys Trp Leu Tyr Phe  Pro Ile Asn
    1025                1030                1035

Lys Thr  Leu Ala Leu Glu Phe  Leu His His Tyr Phe  His Lys Glu
    1040                1045                1050

Ile Val  Thr Pro Asp Asp Thr  Thr Val Cys His Phe  Ile Asn Ser
    1055                1060                1065

Leu Arg  Tyr Tyr Thr Lys Lys  Glu Ser Ile Thr Val  Lys Ile Leu
    1070                1075                1080

Lys Glu  Pro Met Pro Val Leu  Ser Val Lys Phe Glu  Ser Ser Lys
    1085                1090                1095

Lys Asn  Val Leu Gly Ser Phe  Lys His Thr Ile Ala  Leu Pro Ala
    1100                1105                1110

Thr Lys  Asp Trp Glu Arg Leu  Phe Asn His Pro Asn  Phe Leu Ala
    1115                1120                1125

Leu Lys  Ala Asn Pro Ala Pro  Asn Pro Lys Glu Phe  Asn Glu Phe
    1130                1135                1140

Ile Arg  Lys Tyr Phe Leu Ser  Asp Asn Asn Pro Asn  Ser Asp Ile
    1145                1150                1155

Pro Asn  Asn Gly His Asn Ile  Lys Pro Gln Lys His  Lys Ala Val
    1160                1165                1170

Arg Lys  Val Phe Ser Leu Pro  Val Ile Pro Gly Asn  Ala Gly Thr
    1175                1180                1185

Met Met  Arg Ile Arg Arg Lys  Asp Asn Lys Gly Gln  Pro Leu Tyr
    1190                1195                1200
```

```
Gln Leu  Gln Thr Ile Asp Asp  Thr Pro Ser Met Gly  Ile Gln Ile
    1205                 1210                 1215

Asn Glu  Asp Arg Leu Val Lys  Gln Glu Val Leu Met  Asp Ala Tyr
    1220                 1225                 1230

Lys Thr  Arg Asn Leu Ser Thr  Ile Asp Gly Ile Asn  Asn Ser Glu
    1235                 1240                 1245

Gly Gln  Ala Tyr Ala Thr Phe  Asp Asn Trp Leu Thr  Leu Pro Val
    1250                 1255                 1260

Ser Thr  Phe Lys Pro Glu Ile  Ile Lys Leu Glu Met  Lys Pro His
    1265                 1270                 1275

Ser Lys  Thr Arg Arg Tyr Ile  Arg Ile Thr Gln Ser  Leu Ala Asp
    1280                 1285                 1290

Phe Ile  Lys Thr Ile Asp Glu  Ala Leu Met Ile Lys  Pro Ser Asp
    1295                 1300                 1305

Ser Ile  Asp Asp Pro Leu Asn  Met Pro Asn Glu Ile  Val Cys Lys
    1310                 1315                 1320

Asn Lys  Leu Phe Gly Asn Glu  Leu Lys Pro Arg Asp  Gly Lys Met
    1325                 1330                 1335

Lys Ile  Val Ser Thr Gly Lys  Ile Val Thr Tyr Glu  Phe Glu Ser
    1340                 1345                 1350

Asp Ser  Thr Pro Gln Trp Ile  Gln Thr Leu Tyr Val  Thr Gln Leu
    1355                 1360                 1365

Lys Lys  Gln Pro
    1370
```

<210> 166
<211> 4200
<212> ADN
<213> Secuencia Artificial

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

<400> 166

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgga gtcctctcag atcctgtcac ctatcggcat cgacctgggc      120
```

ggaaagttta ctggagtctg cctgtcacac ctggaagcat ttgctgagct gcctaaccac 180

gcaaatacaa agtactcagt gatcctgatt gatcataaca atttccagct gagccaggca 240

cagcggagag ccactagaca cagggtgcgc aacaagaaaa gaaatcagtt cgtgaagagg 300

gtcgccctgc agctgtttca gcacatcctg tcccgggacc tgaacgccaa ggaggaaacc 360

gctctgtgcc attacctgaa caatagaggc tacacctatg tggacacaga tctggacgag 420

tatatcaaag atgaaaccac aattaacctg ctgaaggagc tgctgccaag cgagtccgaa 480

cataatttca ttgactggtt tctgcagaag atgcagagct ccgagttccg gaagatcctg 540

gtgagcaaag tcgaggaaaa gaaagacgat aaagaactga agaacgccgt gaagaacatc 600

aagaacttca tcaccggatt cgagaaaaat agcgtggaag gccaccgaca tcggaaggtc 660

tactttgaga acatcaagtc cgatattaca aaagacaatc agctggattc tatcaagaaa 720

aagattccta gcgtgtgtct gtccaacctg ctgggccacc tgtccaacct gcagtggaag 780

aatctgcata ggtatctggc caaaaatcca aagcagttcg acgagcagac cttcgggaac 840

gaatttctgc ggatgctgaa gaattttcga cacctgaaag gatctcagga gagtctggct 900

gtgcgcaacc tgatccagca gctggaacag tctcaggatt acatcagtat cctggagaaa 960

acccccccctg aaatcaccat tccaccctat gaggcccgga caaacactgg catggaaaaa 1020

gaccagagcc tgctgctgaa ccccgagaag ctgaacaatc tgtacccaaa ctggagaaat 1080

ctgatccccg ggatcattga cgcccaccct ttcctggaga aggatctgga acatacaaag 1140

ctgagagaca ggaaacgcat catttctccc agtaaacagg acgagaagcg ggatagctac 1200

atcctgcaga gatatctgga cctgaacaaa aagatcgata agttcaagat caagaagcag 1260

ctgagctttc tgggacaggg caaacagctg cctgctaacc tgatcgagac acagaaggag 1320

atggaaactc acttcaattc tagtctggtg tctgtcctga tccagattgc tagtgcatac 1380

aacaaggaga gggaagatgc cgctcagggg atctggttcg acaatgcctt ttcactgtgc 1440

gagctgagca acatcaatcc tccacgcaaa cagaagattc tgcccctgct ggtgggagca 1500

atcctgagcg aggacttcat taacaacaag gataagtggg ccaaattcaa gatcttttgg 1560

aacacccaca agattgggcg aacatcactg aaaagcaagt gtaaagagat cgaggaagcc 1620

cggaagaaca gtggaaacgc ttttaaaatc gactacgagg aagctctgaa tcacccagag 1680

cattcaaaca acaaggccct gatcaagatc attcagacca ttcccgatat cattcaggcc 1740

atccagtcac acctgggaca taacgacagc caggctctga tctaccacaa tcctttctca 1800

ctgagccagc tgtatactat cctggagaca aagagggatg gctttcataa aaactgcgtg 1860

gccgtcactt gtgaaaatta ctggcggagc cagaaaaccg agatcgaccc agaaatttcc 1920

tatgcatcta ggctgccagc agacagtgtg cgccccttcg atggcgtcct ggcacgaatg 1980

atgcagcggc tggcctacga gatcgccatg gctaagtggg aacagatcaa acacattcct 2040

gataactcaa gcctgctgat cccaatctac ctggagcaga atcggttcga atttgaggag 2100

agcttcaaga agatcaaggg gtcctctagt gacaaaaccc tggagcaggc catcgaaaag 2160

cagaacattc agtgggagga aaagttccag agaatcatta acgcaagtat gaatatctgc 2220

ccttacaagg gcgcctcaat tggcgggcag ggggagatcg accacatcta cccaaggtcc 2280

ctgtctaaaa agcatttcgg cgtgatcttt aactccgaag tcaatctgat ctactgttca 2340

agccagggga atcgcgagaa aaaggaggaa cactacctgc tggaacatct gtctccactg 2400

tatctgaaac accagttcgg cactgacaac gtgtccgata tcaagaattt tatttctcag 2460

aacgtcgcta atatcaaaaa gtacatttcc ttccacctgc tgaccccaga gcagcagaag 2520

gcagcacggc acgccctgtt tctggattat gacgatgaag cattcaaaac cattacaaag 2580

tttctgatgt ctcagcagaa ggccagagtg aacggcacac agaaattcct ggggaagcag 2640

atcatggagt ttctgtccac tctggcagat tctaaacagc tgcagctgga gttcagcatc 2700

aagcagatta ccgccgagga agtgcacgac catagagagc tgctgtctaa gcaggaaccc 2760

aaactggtca agagtaggca gcagagtttc ccttcacacg ctatcgacgc aactctgacc 2820

atgtctattg ggctgaagga gttcccacag tttagtcagg aactggataa ctcatggttt 2880

atcaatcacc tgatgccaga cgaggtgcat ctgaaccccg tccggagcaa ggaaaagtac 2940

aacaaaccta acatctcctc tactccactg ttcaaggatt ccctgtatgc tgagcggttc 3000

atccccgtgt gggtcaaggg agaaaccttc gcaatcggct ttagcgagaa ggacctgttc 3060

gagatcaagc cctccaacaa ggagaaactg tttacactgc tgaaaaccta cagtactaaa 3120

aatcctggcg agtcactgca ggaactgcag gctaagagca aagcaaagtg gctgtacttc 3180

ccaatcaaca aaaccctggc cctggagttc ctgcaccatt attttcacaa ggaaatcgtg 3240

acacccgacg atactaccgt ctgccatttc atcaactcac tgagatacta cactaaaaag 3300

gagagcatca ccgtgaaaat tctgaaggaa cccatgcctg tgctgtccgt caagttcgag 3360

agttcaaaaa agaacgtgct gggatctttt aaacacacaa tcgccctgcc tgctactaag 3420

gattgggaga ggctgttcaa ccatccaaat tttctggcac tgaaggccaa cccagctccc 3480

aatcctaaag agttcaatga gttcatccgc aagtacttcc tgagcgacaa caatcccaac 3540

tccgatatcc ctaacaatgg ccacaatatc aagccccaga aacataaggc cgtgcgaaag 3600

gtctttagcc tgccagtgat ccccgggaac gctggaacca tgatgcgcat taggcgcaaa 3660

gacaataagg gacagccact gtatcagctg cagacaatcg acgatactcc cagcatgggc 3720

atccagatta acgaggatcg cctggtgaaa caggaagtcc tgatggacgc ctacaagaca 3780

cgaaatctga gcactatcga tgggattaac aattccgagg gacaggcata tgccacattc 3840

gacaactggc tgaccctgcc cgtgagcacc ttcaagcctg agatcatcaa gctggaaatg 3900

aagcctcact ctaaaacccg acggtacatc agaattacac agagtctggc cgacttcatc 3960

aaaactattg atgaggctct gatgatcaag cccagtgact caattgacga tcctctgaac 4020

atgccaaatg agatcgtgtg taaaaacaag ctgttcggga atgaactgaa gcctagggat 4080

ggaaaaatga agatcgtgag cactggcaag attgtcacct acgagtttga aagcgactcc 4140

acccccagt ggatccagac cctgtatgtg acacagctga aaaagcagcc ttaagaattc 4200

<210> 167
<211> 127
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 167
nnnnnnnnnn nnnnnnnnnn guuugagagu uggaaacaac gaguucaaau aagaauucau 60

caaaaucguc ccuuuuggga ccgcucauug uggagcauca aggcuuaaca ugguuaagcc 120

uuuuuuu 127

<210> 168
<211> 1395
<212> PRT
<213> Treponema denticola

<400> 168
Met Lys Lys Glu Ile Lys Asp Tyr Phe Leu Gly Leu Asp Val Gly Thr
1 5 10 15

Gly Ser Val Gly Trp Ala Val Thr Asp Thr Asp Tyr Lys Leu Leu Lys
20 25 30

Ala Asn Arg Lys Asp Leu Trp Gly Met Arg Cys Phe Glu Thr Ala Glu
35 40 45

Thr Ala Glu Val Arg Arg Leu His Arg Gly Ala Arg Arg Arg Ile Glu
50 55 60

Arg Arg Lys Lys Arg Ile Lys Leu Leu Gln Glu Leu Phe Ser Gln Glu
65 70 75 80

Ile Ala Lys Thr Asp Glu Gly Phe Phe Gln Arg Met Lys Glu Ser Pro

```
                85                  90                  95

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

Asp Lys Asp Phe Ala Asp Lys Thr Tyr His Lys Ala Tyr Pro Thr Ile
        115                 120                 125

Asn His Leu Ile Lys Ala Trp Ile Glu Asn Lys Val Lys Pro Asp Pro
    130                 135                 140

Arg Leu Leu Tyr Leu Ala Cys His Asn Ile Ile Lys Lys Arg Gly His
145                 150                 155                 160

Phe Leu Phe Glu Gly Asp Phe Asp Ser Glu Asn Gln Phe Asp Thr Ser
                165                 170                 175

Ile Gln Ala Leu Phe Glu Tyr Leu Arg Glu Asp Met Glu Val Asp Ile
            180                 185                 190

Asp Ala Asp Ser Gln Lys Val Lys Glu Ile Leu Lys Asp Ser Ser Leu
        195                 200                 205

Lys Asn Ser Glu Lys Gln Ser Arg Leu Asn Lys Ile Leu Gly Leu Lys
    210                 215                 220

Pro Ser Asp Lys Gln Lys Lys Ala Ile Thr Asn Leu Ile Ser Gly Asn
225                 230                 235                 240

Lys Ile Asn Phe Ala Asp Leu Tyr Asp Asn Pro Asp Leu Lys Asp Ala
                245                 250                 255

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

Asp Asp Leu Ala Ser Ile Leu Gly Asp Ser Phe Glu Leu Leu Leu Lys
        275                 280                 285

Ala Lys Ala Val Tyr Asn Cys Ser Val Leu Ser Lys Val Ile Gly Asp
    290                 295                 300

Glu Gln Tyr Leu Ser Phe Ala Lys Val Lys Ile Tyr Glu Lys His Lys
305                 310                 315                 320

Thr Asp Leu Thr Lys Leu Lys Asn Val Ile Lys Lys His Phe Pro Lys
                325                 330                 335
```

```
Asp Tyr Lys Lys Val Phe Gly Tyr Asn Lys Asn Glu Lys Asn Asn Asn
            340                 345                 350

Asn Tyr Ser Gly Tyr Val Gly Val Cys Lys Thr Lys Ser Lys Lys Leu
        355                 360                 365

Ile Ile Asn Asn Ser Val Asn Gln Glu Asp Phe Tyr Lys Phe Leu Lys
    370                 375                 380

Thr Ile Leu Ser Ala Lys Ser Glu Ile Lys Glu Val Asn Asp Ile Leu
385                 390                 395                 400

Thr Glu Ile Glu Thr Gly Thr Phe Leu Pro Lys Gln Ile Ser Lys Ser
                405                 410                 415

Asn Ala Glu Ile Pro Tyr Gln Leu Arg Lys Met Glu Leu Glu Lys Ile
            420                 425                 430

Leu Ser Asn Ala Glu Lys His Phe Ser Phe Leu Lys Gln Lys Asp Glu
        435                 440                 445

Lys Gly Leu Ser His Ser Glu Lys Ile Ile Met Leu Leu Thr Phe Lys
    450                 455                 460

Ile Pro Tyr Tyr Ile Gly Pro Ile Asn Asp Asn His Lys Lys Phe Phe
465                 470                 475                 480

Pro Asp Arg Cys Trp Val Val Lys Lys Glu Lys Ser Pro Ser Gly Lys
                485                 490                 495

Thr Thr Pro Trp Asn Phe Phe Asp His Ile Asp Lys Glu Lys Thr Ala
            500                 505                 510

Glu Ala Phe Ile Thr Ser Arg Thr Asn Phe Cys Thr Tyr Leu Val Gly
        515                 520                 525

Glu Ser Val Leu Pro Lys Ser Ser Leu Leu Tyr Ser Glu Tyr Thr Val
    530                 535                 540

Leu Asn Glu Ile Asn Asn Leu Gln Ile Ile Ile Asp Gly Lys Asn Ile
545                 550                 555                 560

Cys Asp Ile Lys Leu Lys Gln Lys Ile Tyr Glu Asp Leu Phe Lys Lys
                565                 570                 575

Tyr Lys Lys Ile Thr Gln Lys Gln Ile Ser Thr Phe Ile Lys His Glu
            580                 585                 590
```

```
Gly Ile Cys Asn Lys Thr Asp Glu Val Ile Ile Leu Gly Ile Asp Lys
        595                 600                 605

Glu Cys Thr Ser Ser Leu Lys Ser Tyr Ile Glu Leu Lys Asn Ile Phe
    610                 615                 620

Gly Lys Gln Val Asp Glu Ile Ser Thr Lys Asn Met Leu Glu Glu Ile
625                 630                 635                 640

Ile Arg Trp Ala Thr Ile Tyr Asp Glu Gly Glu Gly Lys Thr Ile Leu
                645                 650                 655

Lys Thr Lys Ile Lys Ala Glu Tyr Gly Lys Tyr Cys Ser Asp Glu Gln
            660                 665                 670

Ile Lys Lys Ile Leu Asn Leu Lys Phe Ser Gly Trp Gly Arg Leu Ser
        675                 680                 685

Arg Lys Phe Leu Glu Thr Val Thr Ser Glu Met Pro Gly Phe Ser Glu
    690                 695                 700

Pro Val Asn Ile Ile Thr Ala Met Arg Glu Thr Gln Asn Asn Leu Met
705                 710                 715                 720

Glu Leu Leu Ser Ser Glu Phe Thr Phe Thr Glu Asn Ile Lys Lys Ile
                725                 730                 735

Asn Ser Gly Phe Glu Asp Ala Glu Lys Gln Phe Ser Tyr Asp Gly Leu
            740                 745                 750

Val Lys Pro Leu Phe Leu Ser Pro Ser Val Lys Lys Met Leu Trp Gln
        755                 760                 765

Thr Leu Lys Leu Val Lys Glu Ile Ser His Ile Thr Gln Ala Pro Pro
    770                 775                 780

Lys Lys Ile Phe Ile Glu Met Ala Lys Gly Ala Glu Leu Glu Pro Ala
785                 790                 795                 800

Arg Thr Lys Thr Arg Leu Lys Ile Leu Gln Asp Leu Tyr Asn Asn Cys
                805                 810                 815

Lys Asn Asp Ala Asp Ala Phe Ser Ser Glu Ile Lys Asp Leu Ser Gly
            820                 825                 830

Lys Ile Glu Asn Glu Asp Asn Leu Arg Leu Arg Ser Asp Lys Leu Tyr
        835                 840                 845
```

```
Leu Tyr Tyr Thr Gln Leu Gly Lys Cys Met Tyr Cys Gly Lys Pro Ile
    850                 855                 860

Glu Ile Gly His Val Phe Asp Thr Ser Asn Tyr Asp Ile Asp His Ile
865                 870                 875                 880

Tyr Pro Gln Ser Lys Ile Lys Asp Asp Ser Ile Ser Asn Arg Val Leu
                885                 890                 895

Val Cys Ser Ser Cys Asn Lys Asn Lys Glu Asp Lys Tyr Pro Leu Lys
            900                 905                 910

Ser Glu Ile Gln Ser Lys Gln Arg Gly Phe Trp Asn Phe Leu Gln Arg
        915                 920                 925

Asn Asn Phe Ile Ser Leu Glu Lys Leu Asn Arg Leu Thr Arg Ala Thr
    930                 935                 940

Pro Ile Ser Asp Asp Glu Thr Ala Lys Phe Ile Ala Arg Gln Leu Val
945                 950                 955                 960

Glu Thr Arg Gln Ala Thr Lys Val Ala Ala Lys Val Leu Glu Lys Met
                965                 970                 975

Phe Pro Glu Thr Lys Ile Val Tyr Ser Lys Ala Glu Thr Val Ser Met
            980                 985                 990

Phe Arg Asn Lys Phe Asp Ile Val  Lys Cys Arg Glu Ile  Asn Asp Phe
        995                 1000                1005

His His  Ala His Asp Ala Tyr  Leu Asn Ile Val Val  Gly Asn Val
    1010                1015                1020

Tyr Asn  Thr Lys Phe Thr Asn  Asn Pro Trp Asn Phe  Ile Lys Glu
    1025                1030                1035

Lys Arg  Asp Asn Pro Lys Ile  Ala Asp Thr Tyr Asn  Tyr Tyr Lys
    1040                1045                1050

Val Phe  Asp Tyr Asp Val Lys  Arg Asn Asn Ile Thr  Ala Trp Glu
    1055                1060                1065

Lys Gly  Lys Thr Ile Ile Thr  Val Lys Asp Met Leu  Lys Arg Asn
    1070                1075                1080

Thr Pro  Ile Tyr Thr Arg Gln  Ala Ala Cys Lys Lys  Gly Glu Leu
```

```
    1085                1090                1095

Phe Asn  Gln Thr Ile Met Lys  Lys Gly Leu Gly Gln  His Pro Leu
    1100                1105                1110

Lys Lys  Glu Gly Pro Phe Ser  Asn Ile Ser Lys Tyr  Gly Gly Tyr
    1115                1120                1125

Asn Lys  Val Ser Ala Ala Tyr  Tyr Thr Leu Ile Glu  Tyr Glu Glu
    1130                1135                1140

Lys Gly  Asn Lys Ile Arg Ser  Leu Glu Thr Ile Pro  Leu Tyr Leu
    1145                1150                1155

Val Lys  Asp Ile Gln Lys Asp  Gln Asp Val Leu Lys  Ser Tyr Leu
    1160                1165                1170

Thr Asp  Leu Leu Gly Lys Lys  Glu Phe Lys Ile Leu  Val Pro Lys
    1175                1180                1185

Ile Lys  Ile Asn Ser Leu Leu  Lys Ile Asn Gly Phe  Pro Cys His
    1190                1195                1200

Ile Thr  Gly Lys Thr Asn Asp  Ser Phe Leu Leu Arg  Pro Ala Val
    1205                1210                1215

Gln Phe  Cys Cys Ser Asn Asn  Glu Val Leu Tyr Phe  Lys Lys Ile
    1220                1225                1230

Ile Arg  Phe Ser Glu Ile Arg  Ser Gln Arg Glu Lys  Ile Gly Lys
    1235                1240                1245

Thr Ile  Ser Pro Tyr Glu Asp  Leu Ser Phe Arg Ser  Tyr Ile Lys
    1250                1255                1260

Glu Asn  Leu Trp Lys Lys Thr  Lys Asn Asp Glu Ile  Gly Glu Lys
    1265                1270                1275

Glu Phe  Tyr Asp Leu Leu Gln  Lys Lys Asn Leu Glu  Ile Tyr Asp
    1280                1285                1290

Met Leu  Leu Thr Lys His Lys  Asp Thr Ile Tyr Lys  Lys Arg Pro
    1295                1300                1305

Asn Ser  Ala Thr Ile Asp Ile  Leu Val Lys Gly Lys  Glu Lys Phe
    1310                1315                1320
```

```
Lys Ser  Leu Ile Ile Glu Asn  Gln Phe Glu Val Ile  Leu Glu Ile
    1325                 1330                 1335

Leu Lys  Leu Phe Ser Ala Thr  Arg Asn Val Ser Asp  Leu Gln His
    1340                 1345                 1350

Ile Gly  Gly Ser Lys Tyr Ser  Gly Val Ala Lys Ile  Gly Asn Lys
    1355                 1360                 1365

Ile Ser  Ser Leu Asp Asn Cys  Ile Leu Ile Tyr Gln  Ser Ile Thr
    1370                 1375                 1380

Gly Ile  Phe Glu Lys Arg Ile  Asp Leu Leu Lys Val
    1385                 1390                 1395
```

<210> 169
<211> 4269
<212> ADN
<213> Secuencia Artificial

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

<400> 169

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60
aaggtcgaag cgtccatgaa aaaagaaatc aaagactact tcctggggct ggatgtgggg      120
actgggagcg tggggtgggc tgtgaccgat actgactaca aactgctgaa ggctaaccga      180
aaagacctgt ggggcatgag atgcttcgag acagccgaaa ctgctgaggt gcggagactg      240
cacaggggag ccaggcgccg aatcgagcgg agaaagaaac gcattaagct gctgcaggag      300
ctgttctctc aggaaatcgc caaaaccgat gagggcttct ttcagagaat gaaggaaagc      360
ccctttacg ctgaggacaa aacaatcctg caggaaaaca ctctgttcaa tgacaaggat      420
tttgctgata agacttacca caaagcatat cctaccatta atcatctgat caaggcttgg      480
attgagaaca aggtgaaacc agacccccga ctgctgtacc tggcatgtca caacatcatt      540
aagaaaaggg gacatttcct gtttgaaggc gacttcgatt cagagaatca gtttgatacc      600
agcatccagg cactgttcga gtatctgcgc gaggacatgg aggtggacat cgatgccgac      660
agccagaagg tcaaagagat tctgaaggat agctccctga agaactctga aaaacagagt      720
cggctgaata agatcctggg gctgaagcct tccgacaaac agaagaaagc catcacaaac      780
ctgatttctg gaaacaagat caatttcgcc gatctgtacg acaatccaga tctgaaggac      840
gctgagaaaa actcaatcag cttctccaag gacgattttg atgcactgag tgacgatctg      900
gcctcaattc tgggcgacag ctttgaactg ctgctgaagg ccaaagctgt ctataactgc      960
```

tctgtgctga gtaaggtcat cggggacgag cagtacctga gcttcgccaa ggtgaaaatc 1020

tacgaaaagc acaaaaccga tctgacaaag ctgaaaaacg tgatcaagaa acatttcccc 1080

aaggactaca agaaggtctt tggatacaac aagaacgaga aaaacaacaa caattactcc 1140

ggctatgtgg gagtctgtaa gaccaagagt aagaaactga tcattaacaa ctcagtcaac 1200

caggaagatt tctacaagtt tctgaaaact atcctgtcag ccaagagcga gatcaaggaa 1260

gtgaatgaca tcctgaccga gattgaaact ggcacctttc tgccaaagca gatctctaaa 1320

agtaacgcag agattcccta tcagctgagg aaaatggagc tggaaaagat cctgtccaat 1380

gccgaaaagc acttctcttt tctgaagcag aaagacgaaa aaggactgtc acatagcgag 1440

aagatcatta tgctgctgac cttcaagatc ccttactata ttggcccaat caacgataat 1500

cacaagaaat tctttcccga cagatgctgg gtggtcaaga aagagaaatc cccttctggc 1560

aagaccacac catggaactt ctttgatcat atcgacaagg aaaaaacagc agaggccttc 1620

attacttcta ggaccaattt ttgcacttac ctggtgggag agagcgtcct gcctaagtct 1680

agtctgctgt actccgaata taccgtgctg aacgagatca acaatctgca gatcattatc 1740

gatggcaaga atatttgtga catcaagctg aaacagaaga tctacgagga cctgttcaag 1800

aagtacaaga aaattaccca gaagcagatc agcaccttca tcaagcacga aggcatctgc 1860

aacaaaaccg atgaagtgat catcctgggg attgacaagg aatgtacatc aagcctgaaa 1920

agctacatcg agctgaaaaa cattttcggc aagcaggtgg atgagatctc cactaagaat 1980

atgctggagg aaattatcag atgggctacc atctacgacg agggggaagg aaagaccatc 2040

ctgaaaacaa agatcaaggc agaatacgga aagtattgct ccgacgagca gattaagaaa 2100

atcctgaacc tgaagttctc cggctggggg cgactgtctc ggaaatttct ggagacagtg 2160

actagtgaaa tgcccggctt ctcagaacct gtcaatatta tcaccgccat gagggagaca 2220

cagaacaatc tgatggagct gctgtcctct gagttcacct tcaccgagaa cattaagaaa 2280

atcaattctg gattcgaaga tgccgagaag cagtttagtt acgacggcct ggtgaaacca 2340

ctgtttctga gtccctcagt caagaaaatg ctgtggcaga ccctgaagct ggtgaaagag 2400

attagccata tcacacaggc ccccctaag aaaattttca tcgaaatggc aaaggggcc 2460

gagctggaac ctgctcggac taagaccaga ctgaaaatcc tgcaggatct gtataacaat 2520

tgtaagaacg atgctgacgc cttcagctca gagatcaaag acctgagcgg aaagattgag 2580

aacgaagata atctgaggct gcgctccgac aagctgtacc tgtactatac tcagctgggg 2640

aaatgcatgt attgtggaaa gccaattgag atcggccacg tgttcgatac ctcaaactac 2700

gatattgacc atatctatcc ccagagcaag atcaaagacg atagcatttc caatcgggtg 2760

ctggtctgca gctcctgtaa caagaacaag gaggacaagt acccactgaa atcagagatc 2820

cagagcaagc agcgcggctt ctggaacttt ctgcagcgaa acaatttcat ttctctggag 2880

```
aagctgaata gactgacaag ggccactcca atcagtgacg atgagacagc caagtttatt     2940
gctaggcagc tggtggaaac tcgccaggct accaaggtgg ccgctaaagt cctggaaaag     3000
atgttccccg agacaaaaat cgtgtacagc aaggccgaga ctgtctccat gttccggaac     3060
aagtttgata tcgtgaagtg cagagaaatt aacgattttc accatgctca cgacgcatac     3120
ctgaatatcg tggtcggcaa cgtgtataat accaagttca caaacaatcc ttggaacttt     3180
atcaaggaga aaagagataa tccaaagatt gctgacacct acaactacta taaggtgttc     3240
gattatgacg tcaaaaggaa caatatcaca gcatgggaga aggggaaaac tattatcacc     3300
gtgaaagaca tgctgaagag aaacacacca atctacacta ggcaggcagc ctgtaagaaa     3360
ggggagctgt tcaatcagac cattatgaag aaaggactgg gccagcaccc cctgaagaaa     3420
gaaggacctt tttccaatat ctctaaatac ggcgggtata acaaggtgag cgctgcatac     3480
tatacactga ttgagtatga ggaaaagggc aacaaaatcc gctccctgga aactattccc     3540
ctgtacctgg tgaaagatat ccagaaggat caggacgtcc tgaagtctta tctgacagac     3600
ctgctgggga agaaagagtt caagatcctg gtgcccaaga tcaagatcaa cagcctgctg     3660
aagatcaatg ggtttccttg ccatattaca ggaaaaacta acgatagttt cctgctgcgc     3720
cctgccgtgc agttttgctg ttcaaacaat gaggtcctgt acttcaagaa aattatccgg     3780
ttttccgaaa tccgctctca gcgagagaag atcgggaaaa caattagccc atacgaggac     3840
ctgagcttcc ggtcatatat caaggagaac ctgtggaaga aaactaagaa cgatgaaatc     3900
ggagagaagg aattttacga cctgctgcag aagaaaaacc tggagatcta tgatatgctg     3960
ctgactaagc acaaagacac catctacaag aaacgcccta attctgccac cattgatatc     4020
ctggtgaagg ggaaagagaa gttcaaaagc ctgattatcg aaaaccagtt tgaagtgatc     4080
ctggagatcc tgaagctgtt ttctgcaaca cggaatgtca gtgacctgca gcatatcgga     4140
ggcagcaagt actccggcgt ggccaaaatc gggaacaaga tctctagtct ggataactgt     4200
atcctgatct atcagtccat caccggcatc ttcgagaaac ggatcgacct gctgaaggtg     4260
taagaattc                                                            4269
```

<210> 170
<211> 119
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 170
nnnnnnnnnn nnnnnnnnnn guuugagagu aggaaacuac acgguucaaa uaaagaauuu 60

uucuaaucgc ccaaugggcc cauauugaua uggaugaaac ucgcuuagcg aguuuuuuu 119

<210> 171
<211> 1365
<212> PRT
<213> Filifactor alocis

<400> 171

```
Met Thr Lys Glu Tyr Tyr Leu Gly Leu Asp Val Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Thr Asp Ser Gln Tyr Asn Leu Cys Lys Phe Lys Lys
            20                  25                  30

Lys Asp Met Trp Gly Ile Arg Leu Phe Glu Ser Ala Asn Thr Ala Lys
        35                  40                  45

Asp Arg Arg Leu Gln Arg Gly Asn Arg Arg Arg Leu Glu Arg Lys Lys
    50                  55                  60

Gln Arg Ile Asp Leu Leu Gln Glu Ile Phe Ser Pro Glu Ile Cys Lys
65                  70                  75                  80

Ile Asp Pro Thr Phe Phe Ile Arg Leu Asn Glu Ser Arg Leu His Leu
                85                  90                  95

Glu Asp Lys Ser Asn Asp Phe Lys Tyr Pro Leu Phe Ile Glu Lys Asp
            100                 105                 110

Tyr Ser Asp Ile Glu Tyr Tyr Lys Glu Phe Pro Thr Ile Phe His Leu
        115                 120                 125

Arg Lys His Leu Ile Glu Ser Glu Glu Lys Gln Asp Ile Arg Leu Ile
    130                 135                 140

Tyr Leu Ala Leu His Asn Ile Ile Lys Thr Arg Gly His Phe Leu Ile
145                 150                 155                 160

Asp Gly Asp Leu Gln Ser Ala Lys Gln Leu Arg Pro Ile Leu Asp Thr
                165                 170                 175

Phe Leu Leu Ser Leu Gln Glu Glu Gln Asn Leu Ser Val Ser Leu Ser
            180                 185                 190
```

```
Glu Asn Gln Lys Asp Glu Tyr Glu Glu Ile Leu Lys Asn Arg Ser Ile
        195                 200                 205

Ala Lys Ser Glu Lys Val Lys Lys Leu Lys Asn Leu Phe Glu Ile Ser
    210                 215                 220

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

Asn Phe Cys Lys Phe Ile Val Gly Asn Lys Gly Asp Val Cys Lys Phe
                245                 250                 255

Leu Arg Val Ser Lys Glu Glu Leu Glu Ile Asp Ser Phe Ser Phe Ser
            260                 265                 270

Glu Gly Lys Tyr Glu Asp Asp Ile Val Lys Asn Leu Glu Glu Lys Val
        275                 280                 285

Pro Glu Lys Val Tyr Leu Phe Glu Gln Met Lys Ala Met Tyr Asp Trp
    290                 295                 300

Asn Ile Leu Val Asp Ile Leu Glu Thr Glu Glu Tyr Ile Ser Phe Ala
305                 310                 315                 320

Lys Val Lys Gln Tyr Glu Lys His Lys Thr Asn Leu Arg Leu Leu Arg
                325                 330                 335

Asp Ile Ile Leu Lys Tyr Cys Thr Lys Asp Glu Tyr Asn Arg Met Phe
            340                 345                 350

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

Lys Lys Asn Asn Lys Lys Tyr Trp Ile Glu Lys Lys Arg Asn Pro Glu
    370                 375                 380

Glu Phe Tyr Lys Ser Leu Gly Lys Leu Leu Asp Lys Ile Glu Pro Leu
385                 390                 395                 400

Lys Glu Asp Leu Glu Val Leu Thr Met Met Ile Glu Glu Cys Lys Asn
                405                 410                 415

His Thr Leu Leu Pro Ile Gln Lys Asn Lys Asp Asn Gly Val Ile Pro
            420                 425                 430

His Gln Val His Glu Val Glu Leu Lys Lys Ile Leu Glu Asn Ala Lys
        435                 440                 445
```

```
Lys Tyr Tyr Ser Phe Leu Thr Glu Thr Asp Lys Asp Gly Tyr Ser Val
    450                 455                 460

Val Gln Lys Ile Glu Ser Ile Phe Arg Phe Arg Ile Pro Tyr Tyr Val
465                 470                 475                 480

Gly Pro Leu Ser Thr Arg His Gln Glu Lys Gly Ser Asn Val Trp Met
                485                 490                 495

Val Arg Lys Pro Gly Arg Glu Asp Arg Ile Tyr Pro Trp Asn Met Glu
            500                 505                 510

Glu Ile Ile Asp Phe Glu Lys Ser Asn Glu Asn Phe Ile Thr Arg Met
        515                 520                 525

Thr Asn Lys Cys Thr Tyr Leu Ile Gly Glu Asp Val Leu Pro Lys His
    530                 535                 540

Ser Leu Leu Tyr Ser Lys Tyr Met Val Leu Asn Glu Leu Asn Asn Val
545                 550                 555                 560

Lys Val Arg Gly Lys Lys Leu Pro Thr Ser Leu Lys Gln Lys Val Phe
                565                 570                 575

Glu Asp Leu Phe Glu Asn Lys Ser Lys Val Thr Gly Lys Asn Leu Leu
            580                 585                 590

Glu Tyr Leu Gln Ile Gln Asp Lys Asp Ile Gln Ile Asp Asp Leu Ser
        595                 600                 605

Gly Phe Asp Lys Asp Phe Lys Thr Ser Leu Lys Ser Tyr Leu Asp Phe
    610                 615                 620

Lys Lys Gln Ile Phe Gly Glu Glu Ile Glu Lys Glu Ser Ile Gln Asn
625                 630                 635                 640

Met Ile Glu Asp Ile Ile Lys Trp Ile Thr Ile Tyr Gly Asn Asp Lys
                645                 650                 655

Glu Met Leu Lys Arg Val Ile Arg Ala Asn Tyr Ser Asn Gln Leu Thr
            660                 665                 670

Glu Glu Gln Met Lys Lys Ile Thr Gly Phe Gln Tyr Ser Gly Trp Gly
        675                 680                 685

Asn Phe Ser Lys Met Phe Leu Lys Gly Ile Ser Gly Ser Asp Val Ser
    690                 695                 700
```

```
Thr Gly Glu Thr Phe Asp Ile Ile Thr Ala Met Trp Glu Thr Asp Asn
705                 710                 715                 720

Asn Leu Met Gln Ile Leu Ser Lys Lys Phe Thr Phe Met Asp Asn Val
                725                 730                 735

Glu Asp Phe Asn Ser Gly Lys Val Gly Lys Ile Asp Lys Ile Thr Tyr
            740                 745                 750

Asp Ser Thr Val Lys Glu Met Phe Leu Ser Pro Glu Asn Lys Arg Ala
        755                 760                 765

Val Trp Gln Thr Ile Gln Val Ala Glu Glu Ile Lys Lys Val Met Gly
    770                 775                 780

Cys Glu Pro Lys Lys Ile Phe Ile Glu Met Ala Arg Gly Gly Glu Lys
785                 790                 795                 800

Val Lys Lys Arg Thr Lys Ser Arg Lys Ala Gln Leu Leu Glu Leu Tyr
                805                 810                 815

Ala Ala Cys Glu Glu Asp Cys Arg Glu Leu Ile Lys Glu Ile Glu Asp
            820                 825                 830

Arg Asp Glu Arg Asp Phe Asn Ser Met Lys Leu Phe Leu Tyr Tyr Thr
        835                 840                 845

Gln Phe Gly Lys Cys Met Tyr Ser Gly Asp Asp Ile Asp Ile Asn Glu
    850                 855                 860

Leu Ile Arg Gly Asn Ser Lys Trp Asp Arg Asp His Ile Tyr Pro Gln
865                 870                 875                 880

Ser Lys Ile Lys Asp Asp Ser Ile Asp Asn Leu Val Leu Val Asn Lys
                885                 890                 895

Thr Tyr Asn Ala Lys Lys Ser Asn Glu Leu Leu Ser Glu Asp Ile Gln
            900                 905                 910

Lys Lys Met His Ser Phe Trp Leu Ser Leu Leu Asn Lys Lys Leu Ile
        915                 920                 925

Thr Lys Ser Lys Tyr Asp Arg Leu Thr Arg Lys Gly Asp Phe Thr Asp
    930                 935                 940

Glu Glu Leu Ser Gly Phe Ile Ala Arg Gln Leu Val Glu Thr Arg Gln
```

945 950 955 960

Ser Thr Lys Ala Ile Ala Asp Ile Phe Lys Gln Ile Tyr Ser Ser Glu
965 970 975

Val Val Tyr Val Lys Ser Ser Leu Val Ser Asp Phe Arg Lys Lys Pro
980 985 990

Leu Asn Tyr Leu Lys Ser Arg Arg Val Asn Asp Tyr His His Ala Lys
995 1000 1005

Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Val Tyr Asn Lys Lys
1010 1015 1020

Phe Thr Ser Asn Pro Ile Gln Trp Met Lys Lys Asn Arg Asp Thr
1025 1030 1035

Asn Tyr Ser Leu Asn Lys Val Phe Glu His Asp Val Val Ile Asn
1040 1045 1050

Gly Glu Val Ile Trp Glu Lys Cys Thr Tyr His Glu Asp Thr Asn
1055 1060 1065

Thr Tyr Asp Gly Gly Thr Leu Asp Arg Ile Arg Lys Ile Val Glu
1070 1075 1080

Arg Asp Asn Ile Leu Tyr Thr Glu Tyr Ala Tyr Cys Glu Lys Gly
1085 1090 1095

Glu Leu Phe Asn Ala Thr Ile Gln Asn Lys Asn Gly Asn Ser Thr
1100 1105 1110

Val Ser Leu Lys Lys Gly Leu Asp Val Lys Lys Tyr Gly Gly Tyr
1115 1120 1125

Phe Ser Ala Asn Thr Ser Tyr Phe Ser Leu Ile Glu Phe Glu Asp
1130 1135 1140

Lys Lys Gly Asp Arg Ala Arg His Ile Ile Gly Val Pro Ile Tyr
1145 1150 1155

Ile Ala Asn Met Leu Glu His Ser Pro Ser Ala Phe Leu Glu Tyr
1160 1165 1170

Cys Glu Gln Lys Gly Tyr Gln Asn Val Arg Ile Leu Val Glu Lys
1175 1180 1185

```
Ile Lys  Lys Asn Ser Leu Leu  Ile Ile Asn Gly Tyr  Pro Leu Arg
    1190                 1195                 1200

Ile Arg  Gly Glu Asn Glu Val  Asp Thr Ser Phe Lys  Arg Ala Ile
    1205                 1210                 1215

Gln Leu  Lys Leu Asp Gln Lys  Asn Tyr Glu Leu Val  Arg Asn Ile
    1220                 1225                 1230

Glu Lys  Phe Leu Glu Lys Tyr  Val Glu Lys Lys Gly  Asn Tyr Pro
    1235                 1240                 1245

Ile Asp  Glu Asn Arg Asp His  Ile Thr His Glu Lys  Met Asn Gln
    1250                 1255                 1260

Leu Tyr  Glu Val Leu Leu Ser  Lys Met Lys Lys Phe  Asn Lys Lys
    1265                 1270                 1275

Gly Met  Ala Asp Pro Ser Asp  Arg Ile Glu Lys Ser  Lys Pro Lys
    1280                 1285                 1290

Phe Ile  Lys Leu Glu Asp Leu  Ile Asp Lys Ile Asn  Val Ile Asn
    1295                 1300                 1305

Lys Met  Leu Asn Leu Leu Arg  Cys Asp Asn Asp Thr  Lys Ala Asp
    1310                 1315                 1320

Leu Ser  Leu Ile Glu Leu Pro  Lys Asn Ala Gly Ser  Phe Val Val
    1325                 1330                 1335

Lys Lys  Asn Thr Ile Gly Lys  Ser Lys Ile Ile Leu  Val Asn Gln
    1340                 1345                 1350

Ser Val  Thr Gly Leu Tyr Glu  Asn Arg Arg Glu Leu
    1355                 1360                 1365
```

<210> 172
<211> 4179
<212> ADN
<213> Secuencia Artificial

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

<400> 172

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgac caaggagtat tacctggggc tggatgtggg gaccaattcc      120
```

gtgggatggg cagtgaccga ttctcagtac aacctgtgca agtttaagaa aaaggatatg 180

tggggcatcc ggctgttcga aagcgccaac acagcaaagg accggagact gcagagaggg 240

aataggcgcc gactggagcg gaaaaagcag agaattgatc tgctgcagga aatcttctcc 300

ccagagatct gcaagattga ccccactttc tttatccgac tgaacgaatc ccggctgcac 360

ctggaggaca agtctaacga tttcaaatac ccactgttta ttgagaagga ctattctgat 420

atcgagtact ataaagagtt ccccaccatt tttcacctga ggaagcatct gatcgagagt 480

gaggaaaaac aggatatccg gctgatctac ctggccctgc acaacatcat taagacccga 540

ggacattttc tgattgacgg cgatctgcag agcgccaagc agctgaggcc catcctggat 600

acattcctgc tgtccctgca ggaggaacag aacctgtcag tgagcctgtc cgaaaatcag 660

aaggacgagt atgaggaaat tctgaaaaac cgcagcatcg ccaagtccga aaaagtgaaa 720

aagctgaaga atctgtttga gatctcagac gagctggaaa aagaggagaa gaaggcccag 780

agcgccgtga tcgagaactt ctgcaagttt atcgtgggaa ataagggcga tgtctgtaaa 840

ttcctgcggg tgtctaagga ggaactggag attgactctt tcagtttttc agagggcaag 900

tacgaggacg acatcgtgaa aaacctggag gaaaaagtgc ctgaaaaggt ctacctgttt 960

gagcagatga aggcaatgta tgattggaat attctggtcg acatcctgga aaccgaggaa 1020

tacatcagct tcgccaaagt gaagcagtat gagaaacaca agactaacct gcggctgctg 1080

agagacatca ttctgaaata ctgcaccaag gatgagtata atcggatgtt taacgacgag 1140

aaggaagctg gcagctacac cgcatatgtg gggaaactga aaaagaacaa caagaagtac 1200

tggatcgaga aaaagagaaa tcccgaggag ttctacaaat ccctgggcaa gctgctggat 1260

aaaattgagc ctctgaagga ggacctggaa gtgctgacta tgatgatcga ggagtgtaag 1320

aaccacaccc tgctgccaat tcagaaaaat aaggacaacg gcgtgatccc ccaccaggtg 1380

catgaggtcg aactgaaaaa gatcctggaa aatgccaaaa agtactattc cttcctgacc 1440

gagacagaca aggatgggta ctcagtggtc cagaaaatcg agagcatttt caggtttcgc 1500

atcccctact atgtggggcc tctgagtacc cggcaccagg aaaagggatc aaacgtgtgg 1560

atggtcagaa aacctggcag ggaggatcgc atctacccat ggaatatgga ggaaatcatt 1620

gactttgaga agagcaacga aaatttcatt acacggatga ctaacaaatg tacatatctg 1680

atcggggaag atgtcctgcc caagcattct ctgctgtaca gtaaatatat ggtgctgaat 1740

gagctgaaca atgtgaaggt cagaggaaaa aagctgccta catctctgaa acagaaggtg 1800

ttcgaggacc tgtttgaaaa caaatccaaa gtgactggaa agaatctgct ggagtacctg 1860

cagatccagg acaaagatat ccagattgac gatctgtctg gcttcgacaa ggacttcaag 1920

accagcctga agagctatct ggacttcaaa aagcagattt ttggggagga aatcgagaag 1980

gaaagcattc agaacatgat cgaagatatc attaagtgga tcaccatcta cggcaatgac 2040

aaggagatgc tgaaacgagt gattcgggct aattatagca accagctgac agaggaacag 2100

atgaaaaaga tcactggatt tcagtacagt ggctggggga acttctcaaa gatgtttctg 2160

aaagggatca gcggatccga cgtgagcacc ggcgaaacat tcgacatcat taccgcaatg 2220

tgggagacag acaacaatct gatgcagatc ctgtcaaaaa agttcacctt tatggacaac 2280

gtcgaggact tcaacagcgg caaggtcggg aaaatcgaca agattactta cgatagcacc 2340

gtgaaggaaa tgttcctgtc ccctgagaac aaaagggccg tctggcagac cattcaggtg 2400

gctgaggaga tcaagaaagt gatgggctgc gagccaaaaa agatctttat tgaaatggca 2460

cggggcgggg agaaggtgaa aaagaggaca aaatctcgca aggcccagct gctggagctg 2520

tacgccgctt gcgaggaaga ttgtagagaa ctgatcaagg agattgagga ccgggacgag 2580

agggacttca atagcatgaa gctgtttctg tactataccc agttcgggaa atgtatgtat 2640

tccggcgacg acatcgatat taacgagctg attcgcggca attctaagtg ggaccgagat 2700

cacatctacc cccagagcaa aattaaggac gattccatcg ataacctggt gctggtcaat 2760

aagacatata atgccaaaaa gtccaatgag ctgctgtctg aggacatcca gaaaaagatg 2820

cattcattct ggctgagcct gctgaacaaa aagctgatca ctaaaagcaa gtacgaccgc 2880

ctgactcgaa agggcgactt taccgatgag gaactgagtg ggttcatcgc tagacagctg 2940

gtggaaacaa ggcagtcaac taaggcaatc gccgatatct tcaagcagat ctacagctcc 3000

gaggtggtct atgtgaagag cagcctggtg agcgacttca ggaaaaagcc actgaactac 3060

ctgaagtctc ggagagtcaa tgattaccac catgcaaaag acgcctatct gaacattgtg 3120

gtcgggaacg tgtacaacaa aaagtttacc agtaatccca tccagtggat gaaaaagaat 3180

cgcgatacaa actatagcct gaacaaggtg ttcgaacacg acgtggtcat taacggagaa 3240

gtgatctggg aaaagtgcac ataccatgag gacactaata cctatgatgg aggcactctg 3300

gaccgaatcc ggaagattgt ggaacgcgat aacattctgt acaccgagta cgcttattgt 3360

gagaagggcg aactgtttaa tgcaaccatc cagaacaaaa atggaaactc cacagtctct 3420

ctgaaaaagg gcctggacgt gaaaaagtac gggggatact tcagcgccaa cacaagttac 3480

ttctcactga tcgagtttga ggacaagaag ggggatagag caaggcacat cattggagtg 3540

cctatctata ttgcaaacat gctggagcat tctccaagtg ccttcctgga gtactgcgaa 3600

cagaaggggt atcagaatgt gcggattctg gtcgagaaaa tcaaaaagaa cagcctgctg 3660

atcattaatg gataccctct gcgcattcga ggcgagaacg aagtggatac ttcctttaag 3720

agggccatcc agctgaagct ggaccagaaa aactatgagc tggtccgcaa tatcgagaag 3780

ttcctggaaa aatacgtgga gaaaaaggga aactatccaa ttgacgagaa tagagatcac 3840

atcacacatg aaaagatgaa ccagctgtac gaggtgctgc tgtccaaaat gaaaaagttc 3900

aacaagaagg gcatggccga cccctctgat aggatcgaaa agagtaagcc taaattcatc 3960

aagctggagg acctgatcga taagattaat gtgatcaaca aaatgctgaa cctgctgcgc 4020

tgtgacaatg atactaaggc cgacctgtct ctgattgagc tgcccaaaaa cgctgggagt 4080

ttcgtggtca aaaagaatac catcggaaag tcaaaaatca tcctggtgaa tcagagcgtg 4140

actggactgt acgagaatag acgggaactg taagaattc 4179

<210> 173
<211> 111
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 173
nnnnnnnnnn nnnnnnnnnn guuuuagcac uagaaauagu uaaguuaaaa caagcuuaaa 60

gcgucaaugu aauauuuuau uaacacccua cugugucagu gggguuuuuu u 111

<210> 174
<211> 1334
<212> PRT
<213> Staphylococcus pseudintermedius

<400> 174
Met Gly Arg Lys Pro Tyr Ile Leu Ser Leu Asp Ile Gly Thr Gly Ser
1 5 10 15

Val Gly Tyr Ala Cys Met Asp Lys Gly Phe Asn Val Leu Lys Tyr His
20 25 30

Asp Lys Asp Ala Leu Gly Val Tyr Leu Phe Asp Gly Ala Leu Thr Ala
35 40 45

Gln Glu Arg Arg Gln Phe Arg Thr Ser Arg Arg Arg Lys Asn Arg Arg
50 55 60

Ile Lys Arg Leu Gly Leu Leu Gln Glu Leu Leu Ala Pro Leu Val Gln
65 70 75 80

Asn Pro Asn Phe Tyr Gln Phe Gln Arg Gln Phe Ala Trp Lys Asn Asp
85 90 95

```
Asn Met Asp Phe Lys Asn Lys Ser Leu Ser Glu Val Leu Ser Phe Leu
            100                 105                 110

Gly Tyr Glu Ser Lys Lys Tyr Pro Thr Ile Tyr His Leu Gln Glu Ala
        115                 120                 125

Leu Leu Leu Lys Asp Glu Lys Phe Asp Pro Glu Leu Ile Tyr Met Ala
    130                 135                 140

Leu Tyr His Leu Val Lys Tyr Arg Gly His Phe Leu Phe Asp His Leu
145                 150                 155                 160

Lys Ile Glu Asn Leu Thr Asn Asn Asp Asn Met His Asp Phe Val Glu
                165                 170                 175

Leu Ile Glu Thr Tyr Glu Asn Leu Asn Asn Ile Lys Leu Asn Leu Asp
            180                 185                 190

Tyr Glu Lys Thr Lys Val Ile Tyr Glu Ile Leu Lys Asp Asn Glu Met
        195                 200                 205

Thr Lys Asn Asp Arg Ala Lys Arg Val Lys Asn Met Glu Lys Lys Leu
    210                 215                 220

Glu Gln Phe Ser Ile Met Leu Leu Gly Leu Lys Phe Asn Glu Gly Lys
225                 230                 235                 240

Leu Phe Asn His Ala Asp Asn Ala Glu Glu Leu Lys Gly Ala Asn Gln
                245                 250                 255

Ser His Thr Phe Ala Asp Asn Tyr Glu Glu Asn Leu Thr Pro Phe Leu
            260                 265                 270

Thr Val Glu Gln Ser Glu Phe Ile Glu Arg Ala Asn Lys Ile Tyr Leu
        275                 280                 285

Ser Leu Thr Leu Gln Asp Ile Leu Lys Gly Lys Lys Ser Met Ala Met
    290                 295                 300

Ser Lys Val Ala Ala Tyr Asp Lys Phe Arg Asn Glu Leu Lys Gln Val
305                 310                 315                 320

Lys Asp Ile Val Tyr Lys Ala Asp Ser Thr Arg Thr Gln Phe Lys Lys
                325                 330                 335

Ile Phe Val Ser Ser Lys Lys Ser Leu Lys Gln Tyr Asp Ala Thr Pro
            340                 345                 350
```

```
Asn Asp Gln Thr Phe Ser Ser Leu Cys Leu Phe Asp Gln Tyr Leu Ile
        355                 360                 365

Arg Pro Lys Lys Gln Tyr Ser Leu Leu Ile Lys Glu Leu Lys Lys Ile
    370                 375                 380

Ile Pro Gln Asp Ser Glu Leu Tyr Phe Glu Ala Glu Asn Asp Thr Leu
385                 390                 395                 400

Leu Lys Val Leu Asn Thr Thr Asp Asn Ala Ser Ile Pro Met Gln Ile
                405                 410                 415

Asn Leu Tyr Glu Ala Glu Thr Ile Leu Arg Asn Gln Gln Lys Tyr His
            420                 425                 430

Ala Glu Ile Thr Asp Glu Met Ile Glu Lys Val Leu Ser Leu Ile Gln
        435                 440                 445

Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Asn Asp His Thr Ala
    450                 455                 460

Ser Lys Phe Gly Trp Met Glu Arg Lys Ser Asn Glu Ser Ile Lys Pro
465                 470                 475                 480

Trp Asn Phe Asp Glu Val Val Asp Arg Ser Lys Ser Ala Thr Gln Phe
                485                 490                 495

Ile Arg Arg Met Thr Asn Lys Cys Ser Tyr Leu Ile Asn Glu Asp Val
            500                 505                 510

Leu Pro Lys Asn Ser Leu Leu Tyr Gln Glu Met Glu Val Leu Asn Glu
        515                 520                 525

Leu Asn Ala Thr Gln Ile Arg Leu Gln Thr Asp Pro Lys Asn Arg Lys
    530                 535                 540

Tyr Arg Met Met Pro Gln Ile Lys Leu Phe Ala Val Glu His Ile Phe
545                 550                 555                 560

Lys Lys Tyr Lys Thr Val Ser His Ser Lys Phe Leu Glu Ile Met Leu
                565                 570                 575

Asn Ser Asn His Arg Glu Asn Phe Met Asn His Gly Glu Lys Leu Ser
            580                 585                 590

Ile Phe Gly Thr Gln Asp Asp Lys Lys Phe Ala Ser Lys Leu Ser Ser
        595                 600                 605
```

```
Tyr Gln Asp Met Thr Lys Ile Phe Gly Asp Ile Glu Gly Lys Arg Ala
    610                 615                 620

Gln Ile Glu Glu Ile Ile Gln Trp Ile Thr Ile Phe Glu Asp Lys Lys
625                 630                 635                 640

Ile Leu Val Gln Lys Leu Lys Glu Cys Tyr Pro Glu Leu Thr Ser Lys
                645                 650                 655

Gln Ile Asn Gln Leu Lys Lys Leu Asn Tyr Ser Gly Trp Gly Arg Leu
            660                 665                 670

Ser Glu Lys Leu Leu Thr His Ala Tyr Gln Gly His Ser Ile Ile Glu
        675                 680                 685

Leu Leu Arg His Ser Asp Glu Asn Phe Met Glu Ile Leu Thr Asn Asp
    690                 695                 700

Val Tyr Gly Phe Gln Asn Phe Ile Lys Glu Glu Asn Gln Val Gln Ser
705                 710                 715                 720

Asn Lys Ile Gln His Gln Asp Ile Ala Asn Leu Thr Thr Ser Pro Ala
                725                 730                 735

Leu Lys Lys Gly Ile Trp Ser Thr Ile Lys Leu Val Arg Glu Leu Thr
            740                 745                 750

Ser Ile Phe Gly Glu Pro Glu Lys Ile Ile Met Glu Phe Ala Thr Glu
        755                 760                 765

Asp Gln Gln Lys Gly Lys Lys Gln Lys Ser Arg Lys Gln Leu Trp Asp
    770                 775                 780

Asp Asn Ile Lys Lys Asn Lys Leu Lys Ser Val Asp Glu Tyr Lys Tyr
785                 790                 795                 800

Ile Ile Asp Val Ala Asn Lys Leu Asn Asn Glu Gln Leu Gln Gln Glu
                805                 810                 815

Lys Leu Trp Leu Tyr Leu Ser Gln Asn Gly Lys Cys Met Tyr Ser Gly
            820                 825                 830

Gln Ser Ile Asp Leu Asp Ala Leu Leu Ser Pro Asn Ala Thr Lys His
        835                 840                 845

Tyr Glu Val Asp His Ile Phe Pro Arg Ser Phe Ile Lys Asp Asp Ser
```

```
    850                 855                 860

Ile Asp Asn Lys Val Leu Val Ile Lys Lys Met Asn Gln Thr Lys Gly
865                 870                 875                 880

Asp Gln Val Pro Leu Gln Phe Ile Gln Gln Pro Tyr Glu Arg Ile Ala
                885                 890                 895

Tyr Trp Lys Ser Leu Asn Lys Ala Gly Leu Ile Ser Asp Ser Lys Leu
            900                 905                 910

His Lys Leu Met Lys Pro Glu Phe Thr Ala Met Asp Lys Glu Gly Phe
        915                 920                 925

Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Ser Val His Val Arg
    930                 935                 940

Asp Phe Leu Lys Glu Glu Tyr Pro Asn Thr Lys Val Ile Pro Met Lys
945                 950                 955                 960

Ala Lys Met Val Ser Glu Phe Arg Lys Lys Phe Asp Ile Pro Lys Ile
                965                 970                 975

Arg Gln Met Asn Asp Ala His His Ala Ile Asp Ala Tyr Leu Asn Gly
            980                 985                 990

Val Val Tyr His Gly Ala Gln Leu  Ala Tyr Pro Asn Val  Asp Leu Phe
        995                 1000                1005

Asp Phe  Asn Phe Lys Trp Glu  Lys Val Arg Glu Lys  Trp Lys Ala
    1010                1015                1020

Leu Gly  Glu Phe Asn Thr Lys  Gln Lys Ser Arg Glu  Leu Phe Phe
    1025                1030                1035

Phe Lys  Lys Leu Glu Lys Met  Glu Val Ser Gln Gly  Glu Arg Leu
    1040                1045                1050

Ile Ser  Lys Ile Lys Leu Asp  Met Asn His Phe Lys  Ile Asn Tyr
    1055                1060                1065

Ser Arg  Lys Leu Ala Asn Ile  Pro Gln Gln Phe Tyr  Asn Gln Thr
    1070                1075                1080

Ala Val  Ser Pro Lys Thr Ala  Glu Leu Lys Tyr Glu  Ser Asn Lys
    1085                1090                1095
```

```
Ser Asn  Glu Val Val Tyr Lys  Gly Leu Thr Pro Tyr  Gln Thr Tyr
    1100                 1105                 1110

Val Val  Ala Ile Lys Ser Val  Asn Lys Lys Gly Lys  Glu Lys Met
    1115                 1120                 1125

Glu Tyr  Gln Met Ile Asp His  Tyr Val Phe Asp Phe  Tyr Lys Phe
    1130                 1135                 1140

Gln Asn  Gly Asn Glu Lys Glu  Leu Ala Leu Tyr Leu  Ala Gln Arg
    1145                 1150                 1155

Glu Asn  Lys Asp Glu Val Leu  Asp Ala Gln Ile Val  Tyr Ser Leu
    1160                 1165                 1170

Asn Lys  Gly Asp Leu Leu Tyr  Ile Asn Asn His Pro  Cys Tyr Phe
    1175                 1180                 1185

Val Ser  Arg Lys Glu Val Ile  Asn Ala Lys Gln Phe  Glu Leu Thr
    1190                 1195                 1200

Val Glu  Gln Gln Leu Ser Leu  Tyr Asn Val Met Asn  Asn Lys Glu
    1205                 1210                 1215

Thr Asn  Val Glu Lys Leu Leu  Ile Glu Tyr Asp Phe  Ile Ala Glu
    1220                 1225                 1230

Lys Val  Ile Asn Glu Tyr His  His Tyr Leu Asn Ser  Lys Leu Lys
    1235                 1240                 1245

Glu Lys  Arg Val Arg Thr Phe  Phe Ser Glu Ser Asn  Gln Thr His
    1250                 1255                 1260

Glu Asp  Phe Ile Lys Ala Leu  Asp Glu Leu Phe Lys  Val Val Thr
    1265                 1270                 1275

Ala Ser  Ala Thr Arg Ser Asp  Lys Ile Gly Ser Arg  Lys Asn Ser
    1280                 1285                 1290

Met Thr  His Arg Ala Phe Leu  Gly Lys Gly Lys Asp  Val Lys Ile
    1295                 1300                 1305

Ala Tyr  Thr Ser Ile Ser Gly  Leu Lys Thr Thr Lys  Pro Lys Ser
    1310                 1315                 1320

Leu Phe  Lys Leu Ala Glu Ser  Arg Asn Glu Leu
    1325                 1330
```

<210> 175
<211> 4086
<212> ADN
<213> Secuencia Artificial

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

<400> 175

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatggg gaggaaacct tacattctgt ctctggatat tggaactggg 120

tccgtcggct acgcttgcat ggataaagga ttcaacgtgc tgaagtacca cgacaaagat 180

gccctgggag tgtatctgtt cgacggcgct ctgactgcac aggagcggag acagtttagg 240

acctccaggc gccgaaagaa ccggagaatc aaacgcctgg gcctgctgca ggaactgctg 300

gcacccctgg tgcagaaccc taatttctac cagtttcagc ggcagttcgc ctggaagaac 360

gacaatatgg attttaagaa caagagcctg tctgaggtgc tgagcttcct gggatatgaa 420

tccaagaaat accctaccat ctaccacctg caggaggctc tgctgctgaa agacgagaag 480

tttgatccag aactgatcta catggcactg tatcatctgg tgaaatacag aggccacttt 540

ctgttcgatc atctgaagat cgagaacctg actaacaatg acaatatgca cgatttcgtg 600

gagctgattg aaacctatga gaacctgaac aatatcaagc tgaatctgga ctacgagaaa 660

accaaagtga tctatgagat tctgaaagac aacgaaatga ctaagaatga tagagccaaa 720

agggtcaaga acatggagaa gaaactggaa cagttctcta tcatgctgct ggggctgaag 780

ttcaatgagg gaaaactgtt taaccacgcc gataatgctg aggaactgaa gggggctaac 840

cagagccata catttgcaga caactacgag gaaaatctga ctcccttcct gaccgtggaa 900

cagtcagagt ttattgaaag ggccaacaaa atctatctga gcctgactct gcaggatatc 960

ctgaagggca agaaatcaat ggctatgagc aaagtggccg cttacgacaa gttcagaaat 1020

gagctgaaac aggtgaagga cattgtctat aaggctgatt ctaccaggac acagttcaag 1080

aaaatctttg tgagctccaa gaaaagtctg aagcagtacg acgcaactcc caacgatcag 1140

accttctcta gtctgtgcct gtttgaccag tacctgattc gcccaaagaa acagtatagc 1200

ctgctgatca aggagctgaa gaaaatcatt ccccaggact ccgaactgta ctttgaggca 1260

gaaaatgata ccctgctgaa ggtgctgaac accacagaca atgctagcat ccctatgcag 1320

attaacctgt acgaggcaga aaccatcctg cgaaatcagc agaaatatca cgccgagatc 1380

acagatgaga tgattgaaaa ggtgctgtct ctgatccagt tccgcattcc atactatgtg 1440

gggcccctgg tcaacgacca tacagccagt aagtttggat ggatggagcg caaaagtaac 1500

gaatcaatca agccttggaa tttcgacgag gtggtcgatc gaagtaaatc agccactcag 1560

```
tttattaggc gcatgaccaa caagtgttcc tacctgatca atgaggatgt gctgccaaaa     1620
aactctctgc tgtatcagga gatggaagtc ctgaacgaac tgaatgccac acagatcagg     1680
ctgcagactg acccaaaaaa ccgcaagtac cgaatgatgc cccagattaa gctgttcgct     1740
gtggagcaca tctttaagaa atataaaacc gtcagccatt ccaagttcct ggaaattatg     1800
ctgaacagca atcacaggga gaactttatg aatcatggag aaaagctgag tatcttcggc     1860
acacaggacg ataagaaatt tgcatcaaag ctgtcaagct accaggacat gactaaaatc     1920
ttcggggata ttgagggaaa gcgcgcccag attgaggaaa tcattcagtg gatcaccatt     1980
tttgaggaca agaaaatcct ggtgcagaag ctgaaagagt gctatcctga actgacatcc     2040
aagcagatca accagctgaa gaaactgaat tactctggct gggggaggct gagtgagaag     2100
ctgctgactc acgcctatca gggccatagc atcattgaac tgctgcgcca ctccgatgag     2160
aatttcatgg aaattctgac caacgacgtg tacgggttcc agaattttat caaagaggaa     2220
aaccaggtcc agagcaataa gatccagcat caggatattg ccaacctgac tacctctccc     2280
gctctgaaga aaggcatctg gagtacaatt aagctggtgc gggagctgac ttccattttc     2340
ggggagcctg aaaagatcat tatggagttt gctaccgagg accagcagaa aggcaagaaa     2400
cagaaatcaa gaaagcagct gtgggacgat aacatcaaga aaaataagct gaaaagcgtg     2460
gacgagtaca aatatatcat tgatgtcgcc aataagctga acaatgagca gctgcagcag     2520
gaaaaactgt ggctgtacct gagccagaac ggcaagtgta tgtatagcgg gcagtccatc     2580
gacctggatg ccctgctgtc ccccaatgct accaagcact acgaggtgga tcatattttc     2640
cctcggagct tcatcaagga cgatagcatt gacaacaagg tgctggtcat caagaaaatg     2700
aatcagacaa agggcgatca ggtgcccctg cagttcattc agcagcctta cgagagaatc     2760
gcatattgga agagcctgaa caaagccggg ctgatctctg atagtaaact gcacaagctg     2820
atgaaaccag agttcacagc tatggacaag gaaggcttca tccagcggca gctggtggag     2880
actagacaga tcagcgtgca tgtccgggat tttctgaaag aggaataccc taataccaaa     2940
gtgatcccaa tgaaggccaa aatggtgagc gagttccgga agaaatttga catcccaaag     3000
attagacaga tgaacgacgc acaccatgcc atcgatgctt acctgaatgg cgtggtctat     3060
cacggggcac agctggccta ccccaacgtg gacctgtttg atttcaattt taagtgggag     3120
aaagtccgag aaaagtggaa agccctggga gagttcaaca caaagcagaa atctcgggaa     3180
ctgttctttt tcaagaaact ggagaagatg gaagtgtccc agggcgagcg gctgatctct     3240
aagatcaagc tggacatgaa ccacttcaag atcaactact ccagaaagct ggccaacatc     3300
cctcagcagt tttataatca gaccgcagtg tctccaaaga cagccgagct gaaatacgaa     3360
tctaacaaga gtaatgaggt ggtctataag ggactgacac cataccagac ttatgtggtc     3420
```

gccatcaaga gcgtgaacaa gaaaggcaag gagaaaatgg aataccagat gatcgaccac 3480

tacgtgttcg atttttataa attccagaac ggcaatgaga aggaactggc tctgtacctg 3540

gcacagaggg agaacaagga cgaagtgctg gatgctcaga ttgtctatag tctgaataag 3600

ggggatctgc tgtacatcaa caatcatccc tgctatttcg tgtcacgcaa agaggtcatc 3660

aacgcaaagc agtttgagct gaccgtggaa cagcagctgt ctctgtacaa cgtgatgaac 3720

aacaaggaga caaatgtcga aaagctgctg atcgagtatg acttcattgc cgagaaagtg 3780

atcaacgaat accaccatta tctgaatagc aagctgaaag aaaagcgagt ccggaccttt 3840

ttctcagaga gcaaccagac acacgaggac ttcatcaagg ccctggacga gctgtttaag 3900

gtggtcaccg catccgccac aaggtctgat aaaatcggga gtcgcaagaa cagcatgact 3960

catcgagcct tcctgggaaa aggcaaggac gtgaagattg cttacacctc catctctgga 4020

ctgaaaacaa ctaaacctaa gagtctgttt aagctggccg agtcaagaaa cgaactgtaa 4080

gaattc 4086

<210> 176
<211> 114
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 176
nnnnnnnnnn nnnnnnnnnn guuuuagaug gugaaaacca gauuuaaaau caagcaaugc 60

aucuuuugau gcaaaguuuc aauauuuguc ccacguuauc gagggacuuu uuuu 114

<210> 177
<211> 1375
<212> PRT
<213> Lactobacillus johnsonii

<400> 177

```
Met Thr Lys Ile Lys Asp Asp Tyr Ile Val Gly Leu Asp Ile Gly Thr
1               5                  10                  15

Asp Ser Cys Gly Trp Val Ala Met Asn Ser Asn Asn Asp Ile Leu Lys
            20                  25                  30

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

Ser Ala Ala Glu Arg Arg Leu Phe Arg Thr Thr His Arg Arg Ile Lys
    50                  55                  60

Arg Arg Arg Trp Arg Leu Lys Leu Leu Glu Glu Phe Phe Asp Pro Tyr
65                  70                  75                  80

Met Ala Glu Val Asp Pro Tyr Phe Phe Ala Arg Leu Lys Glu Ser Gly
                85                  90                  95

Leu Ser Pro Leu Asp Lys Arg Lys Thr Val Ser Ser Ile Val Phe Pro
            100                 105                 110

Thr Ser Ala Glu Asp Lys Lys Phe Tyr Asp Asp Tyr Pro Thr Ile Tyr
        115                 120                 125

His Leu Arg Tyr Lys Leu Met Thr Glu Asp Glu Lys Phe Asp Leu Arg
    130                 135                 140

Glu Val Tyr Leu Ala Ile His His Ile Ile Lys Tyr Arg Gly Asn Phe
145                 150                 155                 160

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

Lys Ser Ser Ile Glu Lys Leu Asn Glu Leu Tyr Glu Asn Leu Gly Leu
            180                 185                 190

Asp Leu Asn Val Glu Phe Asn Ile Ser Asn Thr Ala Glu Ile Glu Lys
        195                 200                 205

Val Leu Lys Asp Lys Gln Ile Phe Lys Arg Asp Lys Val Lys Lys Ile
    210                 215                 220

Ala Glu Leu Phe Ala Ile Lys Thr Asp Asn Lys Glu Gln Ser Lys Arg
225                 230                 235                 240

Ile Lys Asp Ile Ser Lys Gln Val Ala Asn Ala Val Leu Gly Tyr Lys
                245                 250                 255

Thr Arg Phe Asp Thr Ile Ala Leu Lys Glu Ile Ser Lys Asp Glu Leu
            260                 265                 270

Ser Asp Trp Asn Phe Lys Leu Ser Asp Ile Asp Ala Asp Ser Lys Phe
        275                 280                 285

Glu Ala Leu Met Gly Asn Leu Asp Glu Asn Glu Gln Ala Ile Leu Leu
```

290 295 300

Thr Ile Lys Glu Leu Phe Asn Glu Val Thr Leu Asn Gly Ile Val Glu
305 310 315 320

Asp Gly Asn Thr Leu Ser Glu Ser Met Ile Asn Lys Tyr Asn Asp His
325 330 335

Arg Asp Asp Leu Lys Leu Leu Lys Glu Val Ile Glu Asn His Ile Asp
340 345 350

Arg Lys Lys Ala Lys Glu Leu Ala Leu Ala Tyr Asp Leu Tyr Val Asn
355 360 365

Asn Arg His Gly Gln Leu Leu Gln Ala Lys Lys Lys Leu Gly Lys Ile
370 375 380

Lys Pro Arg Ser Lys Glu Asp Phe Tyr Lys Val Val Asn Lys Asn Leu
385 390 395 400

Asp Asp Ser Arg Ala Ser Lys Glu Ile Lys Lys Lys Ile Glu Leu Asp
405 410 415

Ser Phe Met Pro Lys Gln Arg Thr Asn Ala Asn Gly Val Ile Pro Tyr
420 425 430

Gln Leu Gln Gln Leu Glu Leu Asp Lys Ile Ile Glu Asn Gln Ser Lys
435 440 445

Tyr Tyr Pro Phe Leu Lys Glu Ile Asn Pro Val Ser Ser His Leu Lys
450 455 460

Glu Ala Pro Tyr Lys Leu Asp Glu Leu Ile Arg Phe Arg Val Pro Tyr
465 470 475 480

Tyr Val Gly Pro Leu Ile Ser Pro Asn Glu Ser Thr Lys Asp Ile Gln
485 490 495

Thr Lys Lys Asn Gln Asn Phe Ala Trp Met Ile Arg Lys Glu Glu Gly
500 505 510

Arg Ile Thr Pro Trp Asn Phe Asp Gln Lys Val Asp Arg Ile Glu Ser
515 520 525

Ala Asn Lys Phe Ile Lys Arg Met Thr Thr Lys Asp Thr Tyr Leu Phe
530 535 540

```
Gly Glu Asp Val Leu Pro Ala Asn Ser Leu Leu Tyr Gln Lys Phe Thr
545                 550                 555                 560

Val Leu Asn Glu Leu Asn Asn Ile Arg Ile Asn Gly Lys Arg Ile Ser
                565                 570                 575

Val Asp Leu Lys Gln Glu Ile Tyr Glu Asn Leu Phe Lys Lys His Thr
            580                 585                 590

Thr Val Thr Val Lys Lys Leu Glu Asn Tyr Leu Lys Glu Asn His Asn
        595                 600                 605

Leu Val Lys Val Glu Ile Lys Gly Leu Ala Asp Glu Lys Lys Phe Asn
    610                 615                 620

Ser Gly Leu Thr Thr Tyr Asn Arg Phe Lys Asn Leu Asn Ile Phe Asp
625                 630                 635                 640

Asn Gln Ile Asp Asp Leu Lys Tyr Arg Asn Asp Phe Glu Lys Ile Ile
                645                 650                 655

Glu Trp Ser Thr Ile Phe Glu Asp Lys Ser Ile Tyr Lys Glu Lys Leu
            660                 665                 670

Arg Ser Ile Asp Trp Leu Asn Glu Lys Gln Ile Asn Ala Leu Ser Asn
        675                 680                 685

Ile Arg Leu Gln Gly Trp Gly Arg Leu Ser Lys Lys Leu Leu Ala Gln
    690                 695                 700

Leu His Asp His Asn Gly Gln Thr Ile Ile Glu Gln Leu Trp Asp Ser
705                 710                 715                 720

Gln Asn Asn Phe Met Gln Ile Val Thr Gln Ala Asp Phe Lys Asp Ala
                725                 730                 735

Ile Ala Lys Ala Asn Gln Asn Leu Leu Val Ala Thr Ser Val Glu Asp
            740                 745                 750

Ile Leu Asn Asn Ala Tyr Thr Ser Pro Ala Asn Lys Lys Ala Ile Arg
        755                 760                 765

Gln Val Ile Lys Val Val Asp Asp Ile Val Lys Ala Ala Ser Gly Lys
    770                 775                 780

Val Pro Lys Gln Ile Ala Ile Glu Phe Thr Arg Asp Ala Asp Glu Asn
785                 790                 795                 800
```

```
Pro Lys Arg Ser Gln Thr Arg Gly Ser Lys Leu Gln Lys Val Tyr Lys
                805                 810                 815

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

Glu Ala Ile Lys Asp Lys Lys Leu Val Gln Asp Lys Tyr Tyr Leu Tyr
        835                 840                 845

Phe Met Gln Leu Gly Arg Asp Ala Tyr Thr Gly Glu Pro Ile Asn Ile
    850                 855                 860

Asp Glu Ile Gln Lys Tyr Asp Ile Asp His Ile Leu Pro Gln Ser Phe
865                 870                 875                 880

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

Asn Asn Gly Lys Ser Asp Asn Val Pro Val Lys Leu Phe Gly Asn Glu
            900                 905                 910

Met Ala Ala Asn Leu Gly Met Thr Ile Arg Lys Met Trp Glu Glu Trp
        915                 920                 925

Lys Asn Ile Gly Leu Ile Ser Lys Thr Lys Tyr Asn Asn Leu Leu Thr
    930                 935                 940

Asp Pro Asp His Ile Asn Lys Tyr Lys Ser Ala Gly Phe Ile Arg Arg
945                 950                 955                 960

Gln Leu Val Glu Thr Ser Gln Ile Ile Lys Leu Val Ser Thr Ile Leu
                965                 970                 975

Gln Ser Arg Tyr Pro Asn Thr Glu Ile Ile Thr Val Lys Ala Lys Tyr
            980                 985                 990

Asn His Tyr Leu Arg Glu Lys Phe  Asp Leu Tyr Lys Ser  Arg Glu Val
        995                 1000                1005

Asn Asp  Tyr His His Ala Ile  Asp Ala Tyr Leu Ser  Ala Ile Cys
    1010                1015                1020

Gly Asn  Leu Leu Tyr Gln Asn  Tyr Pro Asn Leu Arg  Pro Phe Phe
    1025                1030                1035

Val Tyr  Gly Gln Tyr Lys Lys  Phe Ser Ser Asp Pro  Asp Lys Glu
    1040                1045                1050
```

```
Lys Ala  Ile Phe Asn Lys Thr  Arg Lys Phe Ser Phe  Ile Ser Gln
    1055                 1060                 1065

Leu Leu  Lys Asn Lys Ser Glu  Asn Ser Lys Glu Ile  Ala Lys Lys
    1070                 1075                 1080

Leu Lys  Arg Ala Tyr Gln Phe  Lys Tyr Met Leu Val  Ser Arg Glu
    1085                 1090                 1095

Thr Glu  Thr Arg Asp Gln Glu  Met Phe Lys Met Thr  Val Tyr Pro
    1100                 1105                 1110

Arg Phe  Ser His Asp Thr Val  Lys Ala Pro Arg Asn  Leu Ile Pro
    1115                 1120                 1125

Lys Lys  Met Gly Met Ser Pro  Asp Ile Tyr Gly Gly  Tyr Thr Asn
    1130                 1135                 1140

Asn Ser  Asp Ala Tyr Met Val  Ile Val Arg Ile Asp  Lys Lys Lys
    1145                 1150                 1155

Gly Thr  Glu Tyr Lys Ile Leu  Gly Ile Pro Thr Arg  Glu Leu Val
    1160                 1165                 1170

Asn Leu  Lys Lys Ala Glu Lys  Glu Asp His Tyr Lys  Ser Tyr Leu
    1175                 1180                 1185

Lys Glu  Ile Leu Thr Pro Arg  Ile Leu Tyr Asn Lys  Asn Gly Lys
    1190                 1195                 1200

Arg Asp  Lys Lys Ile Thr Ser  Phe Glu Ile Val Lys  Ser Lys Ile
    1205                 1210                 1215

Pro Tyr  Lys Gln Val Ile Gln  Asp Gly Asp Lys Lys  Phe Met Leu
    1220                 1225                 1230

Gly Ser  Ser Thr Tyr Val Tyr  Asn Ala Lys Gln Leu  Thr Leu Ser
    1235                 1240                 1245

Thr Glu  Ser Met Lys Ala Ile  Thr Asn Asn Phe Asp  Lys Asp Ser
    1250                 1255                 1260

Asp Glu  Asn Asp Ala Leu Ile  Lys Ala Tyr Asp Glu  Ile Leu Asp
    1265                 1270                 1275

Lys Val  Asp Lys Tyr Leu Pro  Leu Phe Asp Ile Asn  Lys Phe Arg
```

```
    1280                1285                1290

Glu Lys  Leu His Ser Gly Arg  Glu Lys Phe Ile Lys  Leu Ser Leu
    1295                1300                1305

Glu Asp  Lys Lys Asp Thr Ile  Leu Lys Val Leu Glu  Gly Leu His
    1310                1315                1320

Asp Asn  Ala Val Met Thr Lys  Ile Pro Thr Ile Gly  Leu Ser Thr
    1325                1330                1335

Pro Leu  Gly Phe Met Gln Phe  Pro Asn Gly Val Ile  Leu Ser Glu
    1340                1345                1350

Asn Ala  Lys Leu Ile Tyr Gln  Ser Pro Thr Gly Leu  Phe Lys Lys
    1355                1360                1365

Ser Val  Lys Ile Ser Asp Leu
    1370                1375
```

<210> 178
<211> 4209
<212> ADN
<213> Secuencia Artificial

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

<400> 178

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgac aaaaatcaaa gacgactaca tcgtgggact ggacatcggc      120

acagactcct gcgggtgggt ggctatgaac agcaataatg acattctgaa actgcagggc      180

aagaccgcaa tcgggtcacg cctgttcgag ggagggaaga gcgcagctga acggagactg      240

tttcgcacca cacacaggcg catcaaacga cggagatggc gactgaagct gctggaggag      300

ttcttcgacc cctacatggc agaggtggat ccttatttct ttgcccggct gaaggaatct      360

ggcctgagtc cactggacaa aagaaagacc gtgagctcca ttgtgttccc cacatccgcc      420

gaggataaga agttctacga cgattaccct acaatctacc atctgaggta taaactgatg      480

actgaggacg aaaagttcga tctgcgcgaa gtgtacctgg ctatccacca tatcattaag      540

taccgaggaa acttcctgta taataccagt gtgaaagact tcaaggcatc aaagatcgat      600

gtcaaatcta gtatcgagaa gctgaacgag ctgtatgaaa atctgggcct ggacctgaac      660

gtggagttca acattagcaa tactgccgag atcgaaaagg tgctgaaaga caagcagatc      720

ttcaagcggg ataaagtcaa gaaaattgcc gagctgtttg ctatcaaaac cgacaacaag      780
```

gaacagagca agagaatcaa agatatttcc aaacaggtgg ccaatgctgt cctggggtac 840

aagaccaggt tcgacacaat cgctctgaaa gagatttcca aggacgaact gtctgattgg 900

aacttcaaac tgtcagacat cgatgcagac agcaagtttg aggccctgat gggaaacctg 960

gatgagaatg aacaggccat cctgctgact attaaggagc tgtttaacga agtgaccctg 1020

aatggaattg tcgaggacgg caacaccctg agcgaatcca tgatcaacaa gtacaatgat 1080

caccgggacg atctgaagct gctgaaagaa gtgatcgaaa atcatattga cagaaagaaa 1140

gccaaggagc tggcactggc ctacgatctg tatgtcaaca ataggcacgg acagctgctg 1200

caggctaaga aaaagctggg caaaatcaag ccccgctcta aggaggactt ctacaaagtg 1260

gtcaacaaga atctggacga ttcacgggca agcaaggaga tcaaaaagaa aattgaactg 1320

gacagcttta tgcctaagca gagaaccaac gccaatggcg tgatcccata ccagctgcag 1380

cagctggagc tggataagat catcgaaaac cagtctaagt actatccatt cctgaaggag 1440

attaatcccg tgtcaagcca cctgaaagag gccccctata agctggacga actgatccga 1500

tttcgggtgc cttactatgt cggccccctg atttctccta acgagagtac caaggatatc 1560

cagacaaaga aaaaccagaa tttcgcctgg atgattcgca aagaggaagg gcgaatcaca 1620

ccttggaact ttgaccagaa ggtggatcga attgagagcg ccaataagtt catcaaacgg 1680

atgactacca aggacactta cctgtttggg gaggatgtgc tgccagctaa cagcctgctg 1740

tatcagaagt tcaccgtcct gaacgaactg aacaacatcc ggattaatgg aaaaagaatc 1800

tccgtggacc tgaagcagga gatctacgaa aacctgttta agaaacacac aactgtgacc 1860

gtcaagaaac tggagaatta tctgaaggaa aaccataatc tggtgaaagt cgagatcaag 1920

gggctggccg atgaaaagaa attcaacagc ggactgacca catacaatag attcaagaac 1980

ctgaacatct ttgacaacca gattgacgat ctgaagtaca ggaacgattt cgagaagatc 2040

atcgaatggt ctacaatttt tgaggacaag agtatctaca aagaaaagct gaggagcatc 2100

gattggctga acgagaagca gattaacgct ctgtctaata tcagactgca ggggtgggga 2160

aggctgagta agaaactgct ggcacagctg cacgaccata atggccagac catcattgag 2220

cagctgtggg attcccagaa caatttcatg cagattgtga cacaggccga ctttaaagat 2280

gctatcgcaa aggccaacca gaatctgctg gtggctacct cagtcgagga cattctgaac 2340

aatgcataca caagccccgc aaacaagaaa gccatcagac aggtcatcaa ggtggtcgac 2400

gatatcgtga aggcagcctc cggaaaggtc ccaaaacaga tcgccattga gttcactagg 2460

gatgctgacg aaaatcccaa gagaagtcag accaggggct caaagctgca gaaagtgtac 2520

aaggacctga gcactgagct ggcctccaag accattgctg aggaactgaa cgaagcaatc 2580

aaagacaaga aactggtgca ggataagtac tatctgtact ttatgcagct ggggcgggac 2640

gcctatacag gagagcctat caatatcgat gaaatccaga agtacgatat cgaccacatt 2700

ctgccacagt ctttcatcaa ggacgatgcc ctggacaaca gggtgctggt gagccgggct 2760

gtgaacaatg gcaaatctga taatgtgcct gtcaagctgt ttggcaacga gatggctgca 2820

aatctgggga tgactatcag gaaaatgtgg gaggaatgga agaacatcgg cctgattagc 2880

aaaacaaagt acaacaatct gctgactgat cccgaccaca ttaacaagta taagagtgcc 2940

gggttcatca ggcgccagct ggtggagaca tcacagatca tcaagctggt gagcactatc 3000

ctgcagagtc gctaccctaa cactgaaatc attaccgtga aggctaagta caatcattat 3060

ctgcgggaga aatttgacct gtataagagc agagaagtca acgactacca ccatgctatt 3120

gatgcatatc tgtccgccat ctgcggaaat ctgctgtacc agaactatcc aaatctgcgg 3180

cccttctttg tgtacggcca gtataagaaa ttctcctctg atcctgacaa agagaaggcc 3240

atttttaaca aaacccgcaa gttctccttt atctctcagc tgctgaaaaa caagagtgag 3300

aacagcaagg aaatcgctaa gaaactgaaa cgggcatacc agttcaagta tatgctggtg 3360

tctcgagaga ctgaaacccg ggaccaggag atgttcaaaa tgaccgtgta ccccggttc 3420

agccacgata cagtcaaggc tcctaggaac ctgattccaa agaaaatggg catgtcccct 3480

gacatctacg gaggctatac aaacaattct gacgcataca tggtcatcgt ccgcattgat 3540

aagaaaaagg gaactgagta taagatcctg ggcattccaa cccgggaact ggtgaatctg 3600

aaaaaggccg agaaggagga ccattacaaa agctatctga aggagatcct gacaccaagg 3660

attctgtaca acaaaaatgg gaagcgcgat aaaaagatca cttccttcga aattgtgaaa 3720

tctaagatcc cctataagca ggtcatccag gatggggaca aaaagtttat gctgggaagt 3780

tcaacatacg tgtataacgc aaagcagctg acactgagca ctgagtccat gaaagccatc 3840

actaacaatt tcgataagga cagcgatgag aacgacgctc tgattaaggc atacgatgaa 3900

atcctggaca aagtggataa gtatctgcca ctgttcgaca tcaacaagtt ccgggagaag 3960

ctgcacagtg ggcgagaaaa gttcatcaag ctgagcctgg aggacaaaaa ggataccatc 4020

ctgaaagtgc tggaaggact gcatgataac gctgtcatga caaagatccc tactattggc 4080

ctgtccacac cactggggtt catgcagttt cccaacggcg tgattctgag cgagaatgcc 4140

aaactgatct accagtcccc caccgggctg ttcaaaaagt cagtgaagat cagcgacctg 4200

taagaattc 4209

<210> 179
<211> 178
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 179
nnnnnnnnnn nnnnnnnnnn guuguagcuc ccauucucga aagagaaccg uugcuacaau 60

aaggccgucu gaaaagaugu gccgcaacgc ucugccccuu aaagcuucug cuuuaagggg 120

caucguuuau uucgguuaaa aaugccgucu gaaaccgguu uuuagguuuc agacggca 178

<210> 180
<211> 1082
<212> PRT
<213> Neisseria cinerea

<400> 180
Met Ala Ala Phe Lys Pro Asn Pro Met Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Ile Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Ala Ala Arg Arg Leu
50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
130 135 140

Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys
145 150 155 160

Gly Val Ala Asp Asn Thr His Ala Leu Gln Thr Gly Asp Phe Arg Thr

```
                165                 170                 175

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

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Asn Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Asn Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Asp Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Thr Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Val Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Asp Leu Asp Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
                405                 410                 415
```

```
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Asn Arg Tyr Asp Glu Ala Cys Thr Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ser Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Ala Leu Gly
        595                 600                 605

Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
    610                 615                 620

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670
```

```
Ile Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Met Leu Leu Thr
        675                 680                 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Ile Ala
                725                 730                 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
        755                 760                 765

Lys Ala His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Lys Tyr Val Thr Pro Leu Phe Ile Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845

Ser Ala Lys Arg Leu Asp Glu Gly Ile Ser Val Leu Arg Val Pro Leu
    850                 855                 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925
```

```
Gln Lys Thr Gly Val Trp Val His Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940

Ala Thr Ile Val Arg Val Asp Val Phe Glu Lys Gly Gly Lys Tyr Tyr
945                 950                 955                 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
                965                 970                 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Thr Val Met Asp
            980                 985                 990

Asp Ser Phe Glu Phe Lys Phe Val  Leu Tyr Ala Asn Asp  Leu Ile Lys
        995                 1000                1005

Leu Thr  Ala Lys Lys Asn Glu  Phe Leu Gly Tyr Phe  Val Ser Leu
    1010                1015                1020

Asn Arg  Ala Thr Gly Ala Ile  Asp Ile Arg Thr His  Asp Thr Asp
    1025                1030                1035

Ser Thr  Lys Gly Lys Asn Gly  Ile Phe Gln Ser Val  Gly Val Lys
    1040                1045                1050

Thr Ala  Leu Ser Phe Gln Lys  Tyr Gln Ile Asp Glu  Leu Gly Lys
    1055                1060                1065

Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg Pro Pro  Val Arg
    1070                1075                1080
```

<210> 181
<211> 3330
<212> ADN
<213> Secuencia Artificial

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

<400> 181

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc      60

aaggtcgaag cgtccatggc tgccttcaaa cctaatccta tgaactacat cctgggcctg     120

gacattggaa tcgcttctgt cgggtgggct atcgtggaaa tcgacgagga agagaaccct     180

atcagactga ttgatctggg agtcagagtg tttgaaaggg cagaggtgcc aaagaccggc     240

gactccctgg ccgctgcacg gagactggct cggtctgtca ggcgcctgac acgacggaga     300
```

gcacacaggc tgctgcgagc taggcgcctg ctgaagagag agggcgtgct gcaggccgct 360

gacttcgatg aaaacggcct gatcaagagc ctgcccaata ctccttggca gctgagagca 420

gccgctctgg acaggaagct gaccccactg gagtggtctg ccgtgctgct gcacctgatc 480

aagcatcgcg gctacctgag tcagcgaaaa aatgaagggg agacagcaga taaggagctg 540

ggagcactgc tgaaaggagt ggccgacaac actcatgctc tgcagaccgg cgattttagg 600

acacccgctg agctggcact gaataagttc gaaaaagaga gtggacacat tcgaaaccag 660

cggggcgact attcacatac cttcaaccgc aaggatctgc aggccgagct gaatctgctg 720

tttgaaaagc agaaagagtt cgggaatccc cacgtgtccg acgggctgaa agaaggaatc 780

gagacactgc tgatgactca gaggcctgca ctgtctggcg atgccgtgca gaagatgctg 840

gggcattgca cctttgaacc aacagagccc aaggcagcca aaaacaccta cacagccgag 900

aggttcgtgt ggctgacaaa gctgaacaat ctgcgcatcc tggaacaggg cagtgagcgg 960

cccctgactg acaccgaaag agccacactg atggatgagc cttacaggaa gtctaaactg 1020

acttatgccc aggctcgcaa gctgctggac ctggacgata ctgccttctt taagggcctg 1080

aggtacggga aagataatgc agaagccagc accctgatgg agatgaaggc ctatcacgct 1140

atctcccgcg ccctggaaaa agagggcctg aaggacaaga aatctccccc gaacctgagt 1200

cctgaactgc aggatgagat tgggaccgct tttagcctgt tcaagactga cgaggatatc 1260

accggacgcc tgaaagaccg agtgcagccc gaaattctgg aggcactgct gaagcacatc 1320

agttttgata aattcgtgca gatttcactg aaggccctgc gacggatcgt ccctctgatg 1380

gagcagggca atcggtacga cgaggcctgc accgagatct acggagatca ttatggcaag 1440

aaaaacacag aagagaaaat ctatctgccc cctattcctg ccgacgagat ccggaatcca 1500

gtggtcctga gagctctgtc acaggcaaga aaagtgatca acggagtggt cagaaggtac 1560

ggcagccctg ctaggatcca cattgaaacc gcacgcgaag tgggaaagtc ctttaaagac 1620

cgcaaggaaa tcgagaagcg acaggaagag aatagaaaag atagggaaaa gtctgctgca 1680

aaattcaggg agtactttcc aaacttcgtg ggcgaaccca agagtaaaga catcctgaag 1740

ctgcgcctgt acgagcagca gcacgggaag tgtctgtata gcggaaaaga aattaacctg 1800

ggccggctga atgaaaaggg ctatgtggag atcgatcatg cactgccctt ttccagaaca 1860

tgggacgatt ctttcaacaa taaggtcctg gctctgggga gcgagaacca gaacaaggga 1920

aatcagactc cttacgaata tttcaacggg aaggacaata gccgagaatg gcaggagttt 1980

aaagcccgcg tggagacaag ccggttccca cgaagcaaga aacagcggat tctgctgcag 2040

aagtttgacg aagatggatt caaagagaga aacctgaatg acacccggta catcaacaga 2100

tttctgtgcc agttcgtggc tgatcacatg ctgctgaccg gaaagggcaa acgccgagtc 2160

tttgcaagca acggccagat cacaaatctg ctgaggggct tctgggggct gcggaaggtg 2220

agagccgaga atgaccgcca ccatgcactg gatgccgtgg tcgtggcttg ttccactatt 2280

gcaatgcagc agaagatcac caggtttgtg cgctataaag agatgaacgc cttcgacgga 2340

aagacaattg ataaagaaac tggcgaggtg ctgcaccaga aggcacattt tcctcagcca 2400

tgggagttct tcgcccagga agtgatgatc cgggtctttg ggaagcctga cggaaaacca 2460

gagttcgaag aggccgatac cccagaaaag ctgcggacac tgctggctga aaaactgagc 2520

tccagacccg aggcagtgca caagtacgtc acccccctgt tcattagcag ggcccctaat 2580

cgcaaaatgt ccgggcaggg acatatggag actgtgaaat cagctaagcg gctggacgaa 2640

ggcatcagcg tgctgagagt cccactgacc cagctgaagc tgaaagatct ggagaagatg 2700

gtgaaccggg aaagagagcc caagctgtat gaagctctga aagcaagact ggaggcccac 2760

aaggacgatc cagctaaagc atttgccgag cccttctaca aatatgacaa ggccggcaat 2820

cggacacagc aggtgaaggc tgtcagagtg gagcaggtcc agaaaactgg ggtctgggtg 2880

cacaaccata atggaattgc cgacaacgct acaatcgtcc gggtggatgt gttcgagaaa 2940

ggcgggaagt actatctggt gcctatctac tcctggcagg tcgccaaggg aatcctgcca 3000

gatagagctg tcgtgcaggg caaagacgaa gaggattgga ctgtgatgga cgattctttc 3060

gagtttaagt tcgtcctgta cgcaaacgac ctgatcaagc tgacagccaa gaaaaatgaa 3120

tttctggggt atttcgtgtc actgaacagg gcaactggag ccatcgatat tcgcacacat 3180

gacactgata gcaccaaggg aaaaaacggc atctttcagt ctgtgggcgt caagaccgcc 3240

ctgagtttcc agaaatatca gattgacgaa ctggggaagg agatccgacc ctgtcggctg 3300

aagaaacgac cacccgtgcg gtaagaattc 3330

<210> 182
<211> 145
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 182
nnnnnnnnnn nnnnnnnnnn gcugcggauu gcgggaaauc gcuuuucgca agcaaauuga 60

ccccuugugc gggcucggca ucccaagguc agcugccggu uauuaucgaa aaggcccacc 120

gcaagcagcg cgugggccuu uuuuu 145

<210> 183
<211> 1037
<212> PRT
<213> Parvibaculum lavamentivorans

<400> 183

```
Met Glu Arg Ile Phe Gly Phe Asp Ile Gly Thr Thr Ser Ile Gly Phe
1               5                  10                  15

Ser Val Ile Asp Tyr Ser Ser Thr Gln Ser Ala Gly Asn Ile Gln Arg
            20                  25                  30

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

Leu Asn Gln Gln Arg Arg Gln Lys Arg Met Met Arg Arg Gln Leu Arg
    50                  55                  60

Arg Arg Arg Ile Arg Arg Lys Ala Leu Asn Glu Thr Leu His Glu Ala
65                  70                  75                  80

Gly Phe Leu Pro Ala Tyr Gly Ser Ala Asp Trp Pro Val Val Met Ala
                85                  90                  95

Asp Glu Pro Tyr Glu Leu Arg Arg Arg Gly Leu Glu Glu Gly Leu Ser
            100                 105                 110

Ala Tyr Glu Phe Gly Arg Ala Ile Tyr His Leu Ala Gln His Arg His
        115                 120                 125

Phe Lys Gly Arg Glu Leu Glu Glu Ser Asp Thr Pro Asp Pro Asp Val
    130                 135                 140

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

Leu Lys Asn Glu Gln Thr Thr Leu Gly Ala Trp Leu Ala Arg Arg Pro
                165                 170                 175

Pro Ser Asp Arg Lys Arg Gly Ile His Ala His Arg Asn Val Val Ala
            180                 185                 190

Glu Glu Phe Glu Arg Leu Trp Glu Val Gln Ser Lys Phe His Pro Ala
        195                 200                 205

Leu Lys Ser Glu Glu Met Arg Ala Arg Ile Ser Asp Thr Ile Phe Ala
    210                 215                 220
```

```
Gln Arg Pro Val Phe Trp Arg Lys Asn Thr Leu Gly Glu Cys Arg Phe
225                 230                 235                 240

Met Pro Gly Glu Pro Leu Cys Pro Lys Gly Ser Trp Leu Ser Gln Gln
                245                 250                 255

Arg Arg Met Leu Glu Lys Leu Asn Asn Leu Ala Ile Ala Gly Gly Asn
            260                 265                 270

Ala Arg Pro Leu Asp Ala Glu Glu Arg Asp Ala Ile Leu Ser Lys Leu
        275                 280                 285

Gln Gln Gln Ala Ser Met Ser Trp Pro Gly Val Arg Ser Ala Leu Lys
    290                 295                 300

Ala Leu Tyr Lys Gln Arg Gly Glu Pro Gly Ala Glu Lys Ser Leu Lys
305                 310                 315                 320

Phe Asn Leu Glu Leu Gly Gly Glu Ser Lys Leu Leu Gly Asn Ala Leu
                325                 330                 335

Glu Ala Lys Leu Ala Asp Met Phe Gly Pro Asp Trp Pro Ala His Pro
            340                 345                 350

Arg Lys Gln Glu Ile Arg His Ala Val His Glu Arg Leu Trp Ala Ala
        355                 360                 365

Asp Tyr Gly Glu Thr Pro Asp Lys Lys Arg Val Ile Ile Leu Ser Glu
    370                 375                 380

Lys Asp Arg Lys Ala His Arg Glu Ala Ala Ala Asn Ser Phe Val Ala
385                 390                 395                 400

Asp Phe Gly Ile Thr Gly Glu Gln Ala Ala Gln Leu Gln Ala Leu Lys
                405                 410                 415

Leu Pro Thr Gly Trp Glu Pro Tyr Ser Ile Pro Ala Leu Asn Leu Phe
            420                 425                 430

Leu Ala Glu Leu Glu Lys Gly Glu Arg Phe Gly Ala Leu Val Asn Gly
        435                 440                 445

Pro Asp Trp Glu Gly Trp Arg Arg Thr Asn Phe Pro His Arg Asn Gln
    450                 455                 460

Pro Thr Gly Glu Ile Leu Asp Lys Leu Pro Ser Pro Ala Ser Lys Glu
465                 470                 475                 480
```

```
Glu Arg Glu Arg Ile Ser Gln Leu Arg Asn Pro Thr Val Val Arg Thr
                485                 490                 495

Gln Asn Glu Leu Arg Lys Val Val Asn Asn Leu Ile Gly Leu Tyr Gly
            500                 505                 510

Lys Pro Asp Arg Ile Arg Ile Glu Val Gly Arg Asp Val Gly Lys Ser
        515                 520                 525

Lys Arg Glu Arg Glu Glu Ile Gln Ser Gly Ile Arg Arg Asn Glu Lys
    530                 535                 540

Gln Arg Lys Lys Ala Thr Glu Asp Leu Ile Lys Asn Gly Ile Ala Asn
545                 550                 555                 560

Pro Ser Arg Asp Asp Val Glu Lys Trp Ile Leu Trp Lys Glu Gly Gln
                565                 570                 575

Glu Arg Cys Pro Tyr Thr Gly Asp Gln Ile Gly Phe Asn Ala Leu Phe
            580                 585                 590

Arg Glu Gly Arg Tyr Glu Val Glu His Ile Trp Pro Arg Ser Arg Ser
        595                 600                 605

Phe Asp Asn Ser Pro Arg Asn Lys Thr Leu Cys Arg Lys Asp Val Asn
    610                 615                 620

Ile Glu Lys Gly Asn Arg Met Pro Phe Glu Ala Phe Gly His Asp Glu
625                 630                 635                 640

Asp Arg Trp Ser Ala Ile Gln Ile Arg Leu Gln Gly Met Val Ser Ala
                645                 650                 655

Lys Gly Gly Thr Gly Met Ser Pro Gly Lys Val Lys Arg Phe Leu Ala
            660                 665                 670

Lys Thr Met Pro Glu Asp Phe Ala Ala Arg Gln Leu Asn Asp Thr Arg
        675                 680                 685

Tyr Ala Ala Lys Gln Ile Leu Ala Gln Leu Lys Arg Leu Trp Pro Asp
    690                 695                 700

Met Gly Pro Glu Ala Pro Val Lys Val Glu Ala Val Thr Gly Gln Val
705                 710                 715                 720

Thr Ala Gln Leu Arg Lys Leu Trp Thr Leu Asn Asn Ile Leu Ala Asp
                725                 730                 735
```

```
Asp Gly Glu Lys Thr Arg Ala Asp His Arg His His Ala Ile Asp Ala
            740                 745                 750

Leu Thr Val Ala Cys Thr His Pro Gly Met Thr Asn Lys Leu Ser Arg
        755                 760                 765

Tyr Trp Gln Leu Arg Asp Asp Pro Arg Ala Glu Lys Pro Ala Leu Thr
    770                 775                 780

Pro Pro Trp Asp Thr Ile Arg Ala Asp Ala Glu Lys Ala Val Ser Glu
785                 790                 795                 800

Ile Val Val Ser His Arg Val Arg Lys Lys Val Ser Gly Pro Leu His
                805                 810                 815

Lys Glu Thr Thr Tyr Gly Asp Thr Gly Thr Asp Ile Lys Thr Lys Ser
            820                 825                 830

Gly Thr Tyr Arg Gln Phe Val Thr Arg Lys Lys Ile Glu Ser Leu Ser
        835                 840                 845

Lys Gly Glu Leu Asp Glu Ile Arg Asp Pro Arg Ile Lys Glu Ile Val
    850                 855                 860

Ala Ala His Val Ala Gly Arg Gly Gly Asp Pro Lys Lys Ala Phe Pro
865                 870                 875                 880

Pro Tyr Pro Cys Val Ser Pro Gly Gly Pro Glu Ile Arg Lys Val Arg
                885                 890                 895

Leu Thr Ser Lys Gln Gln Leu Asn Leu Met Ala Gln Thr Gly Asn Gly
            900                 905                 910

Tyr Ala Asp Leu Gly Ser Asn His His Ile Ala Ile Tyr Arg Leu Pro
        915                 920                 925

Asp Gly Lys Ala Asp Phe Glu Ile Val Ser Leu Phe Asp Ala Ser Arg
    930                 935                 940

Arg Leu Ala Gln Arg Asn Pro Ile Val Gln Arg Thr Arg Ala Asp Gly
945                 950                 955                 960

Ala Ser Phe Val Met Ser Leu Ala Ala Gly Glu Ala Ile Met Ile Pro
                965                 970                 975

Glu Gly Ser Lys Lys Gly Ile Trp Ile Val Gln Gly Val Trp Ala Ser
```

```
            980                 985                 990

Gly Gln Val Val Leu Glu Arg Asp  Thr Asp Ala Asp His  Ser Thr Thr
        995                 1000                1005

Thr Arg  Pro Met Pro Asn Pro  Ile Leu Lys Asp Asp  Ala Lys Lys
    1010                1015                1020

Val Ser  Ile Asp Pro Ile Gly  Arg Val Arg Pro Ser  Asn Asp
    1025                1030                1035
```

<210> 184
<211> 3195
<212> ADN
<213> Secuencia Artificial

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

<400> 184

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60
aaggtcgaag cgtccatgga gaggattttc ggctttgaca tcggcacaac aagtatcgga      120
ttcagcgtga ttgattacag tagcacccag tccgcaggca acatccagag gctgggcgtg      180
cgcattttcc ctgaggcaag ggacccagat gggacccccc tgaaccagca gcggagacag      240
aaacgcatga tgaggcgcca gctgcgacgg agaaggattc gccgaaaggc actgaatgag      300
acactgcacg aagccggctt tctgccagct tacgggtctg cagattggcc cgtggtcatg      360
gccgacgagc cttatgaact gcggagaagg ggactggagg aaggcctgag tgcttacgag      420
ttcggacggg caatctatca tctggcccag caccggcatt ttaaaggcag agaactggag      480
gaatccgata cacccgaccc tgatgtggac gatgagaagg aagccgctaa cgagagagca      540
gccactctga aggccctgaa aaatgaacag accacactgg gagcatggct ggcccgccga      600
ccccttctg accgcaagcg aggaatccac gcccatagga acgtggtcgc tgaggagttc      660
gagcgcctgt gggaagtgca gtccaagttt cacccgctc tgaaatctga ggaaatgcgg      720
gcaagaatca gtgatacaat tttcgcccag aggcctgtgt tttggcgcaa gaacactctg      780
ggagagtgca gattcatgcc tggcgaacca ctgtgtccca aggggtcctg gctgtctcag      840
cagcggagaa tgctggagaa actgaacaat ctggctatcg caggcgggaa tgctaggcca      900
ctggatgcag aggaacgcga cgccattctg agtaagctgc agcagcaggc cagcatgtcc      960
tggccaggcg tgcggtcagc tctgaaggca ctgtacaaac agagaggcga gcccggggct     1020
gaaaagagcc tgaaattcaa cctggagctg ggaggcgaat ccaagctgct gggaaatgcc     1080
ctggaggcta aactggcaga tatgtttggc cctgactggc cagctcaccc ccgaaagcag     1140
```

gagatccggc acgcagtgca tgaacggctg tgggctgcag attacggcga gacacccgac 1200

aagaaaagag tcatcattct gtccgagaag gatcgaaaag ctcatcggga agccgctgca 1260

aactctttcg tggcagactt tggaattact ggcgagcagg cagctcagct gcaggccctg 1320

aagctgccaa ccggctggga accttatagc atcccagcac tgaacctgtt cctggccgag 1380

ctggaaaagg gggagaggtt tggagccctg gtgaatggac ctgattggga aggctggagg 1440

cgcacaaact tcccccaccg caatcagcct actggggaga tcctggacaa gctgccaagt 1500

cccgcctcaa aagaggaaag ggaacgcatt agccagctgc gcaacccaac cgtggtccga 1560

acacagaatg agctgagaaa ggtggtcaac aatctgatcg ggctgtatgg aaaacccgat 1620

cgaatccgga ttgaagtggg ccgggacgtc gggaagtcca aaagagaaag ggaggaaatc 1680

cagtctggca ttcgacggaa cgagaagcag agaaagaaag ccactgaaga tctgatcaaa 1740

aacggaattg ctaatcctag ccgggacgat gtggagaagt ggatcctgtg gaaagagggc 1800

caggaaagat gcccatacac cggcgaccag attggcttca atgccctgtt tagagaaggc 1860

agatatgagg tggaacacat ctggcctcgc tctcgaagtt ttgataacag cccaaggaat 1920

aagacactgt gtcgcaaaga cgtgaacatc gagaagggaa ataggatgcc tttcgaggca 1980

tttggccatg acgaagatcg gtggagcgcc atccagatta gactgcaggg catggtgtca 2040

gccaaagggg gaactgggat gagccccgga aaggtcaaac gcttcctggc taagaccatg 2100

cctgaggatt ttgcagcccg gcagctgaac gacacaagat acgctgcaaa gcagatcctg 2160

gcccagctga aaaggctgtg gccagacatg ggacctgagg ctccagtgaa ggtcgaagca 2220

gtgactggac aggtcaccgc ccagctgcgc aaactgtgga ctctgaacaa tattctggct 2280

gacgatgggg agaaaaccag agcagatcac aggcaccatg ccatcgacgc tctgacagtg 2340

gcctgcactc atcctggaat gaccaacaag ctgagcaggt attggcagct gcgcgacgat 2400

ccacgagcag agaagccagc tctgactcca ccctgggata ccatccgcgc cgacgctgag 2460

aaagccgtgt ctgaaattgt ggtcagtcac cgggtgagaa agaaagtcag cggcccactg 2520

cataaggaga ctacctacgg cgatacaggg actgacatta agaccaaatc cggcacatat 2580

agacagttcg tgaccaggaa gaaaatcgag tcactgagca agggggagct ggatgaaatt 2640

cgcgaccccc gaatcaaaga aattgtggca gctcacgtcg caggacgagg aggcgacccc 2700

aagaaggcct tccctccata cccctgtgtg tctcccggag gccctgagat ccggaaggtc 2760

agactgacca gtaaacagca gctgaacctg atggcccaga cagggaatgg atacgctgac 2820

ctgggctcca accaccatat cgcaatctac cggctgcccg atgggaaggc cgacttcgag 2880

attgtgtcac tgtttgatgc tagcagaagg ctggcacaga gaaatccaat cgtgcagagg 2940

acacgagcag acggagccag cttcgtcatg tccctggcag ccggagaggc catcatgatt 3000

cccgaaggct caaagaaagg gatctggatt gtgcagggag tctgggcaag cggacaggtg 3060

gtcctggaga gggacaccga tgctgaccac tctacaacta cccgccctat gccaaacccc 3120

atcctgaagg acgatgccaa gaaagtgagt atcgatccta ttggccgagt ccggccatca 3180

aatgactaag aattc 3195

<210> 185
<211> 103
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 185
nnnnnnnnnn nnnnnnnnnn guuuuaguac ucuggaaaca gaaucuacua aaacaaggca 60

aaaugccgug uuuaucucgu caacuuguug gcgagauuuu uuu 103

<210> 186
<211> 1053
<212> PRT
<213> Staphylococcus aureus

<400> 186

Met	Lys	Arg	Asn	Tyr	Ile	Leu	Gly	Leu	Asp	Ile	Gly
1				5				10			
Ile	Thr	Ser	Val	Gly	Tyr	Gly	Ile	Ile	Asp	Tyr	Glu
				20				25			
Thr	Arg	Asp	Val	Ile	Asp	Ala	Gly	Val	Arg	Leu	Phe
				35				40			
Lys	Glu	Ala	Asn	Val	Glu	Asn	Asn	Glu	Gly	Arg	Arg
				50				55			
Ser	Lys	Arg	Gly	Ala	Arg	Arg	Leu	Lys	Arg	Arg	Arg
				65				70			
Arg	His	Arg	Ile	Gln	Arg	Val	Lys	Lys	Leu	Leu	Phe
				85				90			
Asp	Tyr	Asn	Leu	Leu	Thr	Asp	His	Ser	Glu	Leu	Ser
				100				105			
Gly	Ile	Asn	Pro	Tyr	Glu	Ala	Arg	Val	Lys	Gly	Leu
				115				120			
Ser	Gln	Lys	Leu	Ser	Glu	Glu	Glu	Phe	Ser	Ala	Ala
				130				135			
Leu	Leu	His	Leu	Ala	Lys	Arg	Arg	Gly	Val	His	Asn
				145				150			
Val	Asn	Glu	Val	Glu	Glu	Asp	Thr	Gly	Asn	Glu	Leu
				165				170			
Ser	Thr	Lys	Glu	Gln	Ile	Ser	Arg	Asn	Ser	Lys	Ala
				180				185			
Leu	Glu	Glu	Lys	Tyr	Val	Ala	Glu	Leu	Gln	Leu	Glu
				195				200			
Arg	Leu	Lys	Lys	Asp	Gly	Glu	Val	Arg	Gly	Ser	Ile
				210				215			
Asn	Arg	Phe	Lys	Thr	Ser	Asp	Tyr	Val	Lys	Glu	Ala
				225				230			
Lys	Gln	Leu	Leu	Lys	Val	Gln	Lys	Ala	Tyr	His	Gln
				245				250			
Leu	Asp	Gln	Ser	Phe	Ile	Asp	Thr	Tyr	Ile	Asp	Leu
				260				265			
Leu	Glu	Thr	Arg	Arg	Thr	Tyr	Tyr	Glu	Gly	Pro	Gly
				275				280			
Glu	Gly	Ser	Pro	Phe	Gly	Trp	Lys	Asp	Ile	Lys	Glu
				290				295			
Trp	Tyr	Glu	Met	Leu	Met	Gly	His	Cys	Thr	Tyr	Phe
				305				310			
Pro	Glu	Glu	Leu	Arg	Ser	Val	Lys	Tyr	Ala	Tyr	Asn
				325				330			
Ala	Asp	Leu	Tyr	Asn	Ala	Leu	Asn	Asp	Leu	Asn	Asn
				340				345			
Leu	Val	Ile	Thr	Arg	Asp	Glu	Asn	Glu	Lys	Leu	Glu
				355				360			
Tyr	Tyr	Glu	Lys	Phe	Gln	Ile	Ile	Glu	Asn	Val	Phe
				370				375			
Lys	Gln	Lys	Lys	Lys	Pro	Thr	Leu	Lys	Gln	Ile	Ala
				385				390			
Lys	Glu	Ile	Leu	Val	Asn	Glu	Glu	Asp	Ile	Lys	Gly
				405				410			
Tyr	Arg	Val	Thr	Ser	Thr	Gly	Lys	Pro	Glu	Phe	Thr
				420				425			
Asn	Leu	Lys	Val	Tyr	His	Asp	Ile	Lys	Asp	Ile	Thr
				435				440			
Ala	Arg	Lys	Glu	Ile	Ile	Glu	Asn	Ala	Glu	Leu	Leu
				450				455			
Asp	Gln	Ile	Ala	Lys	Ile	Leu	Thr	Ile	Tyr	Gln	Ser
				465				470			
Ser	Glu	Asp	Ile	Gln	Glu	Glu	Leu	Thr	Asn	Leu	Asn
				485				490			
Ser	Glu	Leu	Thr	Gln	Glu	Glu	Ile	Glu	Gln	Ile	Ser
				500				505			
Asn	Leu	Lys	Gly	Tyr	Thr	Gly	Thr	His	Asn	Leu	Ser
				515				520			
Leu	Lys	Ala	Ile	Asn	Leu	Ile	Leu	Asp	Glu	Leu	Trp
				530				535			
His	Thr	Asn	Asp	Asn	Gln	Ile	Ala	Ile	Phe	Asn	Arg
				545				550			
Leu	Lys	Leu	Val	Pro	Lys	Lys	Val	Asp	Leu	Ser	Gln
				565				570			
Gln	Lys	Glu	Ile	Pro	Thr	Thr	Leu	Val	Asp	Asp	Phe
				580				585			
Ile	Leu	Ser	Pro	Val	Val	Lys	Arg	Ser	Phe	Ile	Gln
				595				600			
Ser	Ile	Lys	Val	Ile	Asn	Ala	Ile	Ile	Lys	Lys	Tyr
				610				615			
Gly	Leu	Pro	Asn	Asp	Thr	Ile	Ile	Glu	Leu	Ala	Arg
				625				630			
Glu	Lys	Asn	Ser	Lys	Asp	Ala	Gln	Lys	Met	Ile	Asn
				645				650			
Glu	Met	Gln	Lys	Arg	Asn	Arg	Gln	Thr	Asn	Glu	Arg
				660				665			
Ile	Glu	Glu	Ile	Ile	Arg	Thr	Thr	Gly	Lys	Glu	Asn
				675				680			
Ala	Lys	Tyr	Leu	Ile	Glu	Lys	Ile	Lys	Leu	His	Asp
				690				695			
Met	Gln	Glu	Gly	Lys	Cys	Leu	Tyr	Ser	Leu	Glu	Ala
				705				710			
Ile	Pro	Leu	Glu	Asp	Leu	Leu	Asn	Asn	Pro	Phe	Asn
				725				730			
Tyr	Glu	Val	Asp	His	Ile	Ile	Pro	Arg	Ser	Val	Ser
				740				745			
Phe	Asp	Asn	Ser	Phe	Asn	Asn	Lys	Val	Leu	Val	Lys
				755				760			
Gln	Glu	Glu	Asn	Ser	Lys	Lys	Gly	Asn	Arg	Thr	Pro
				770				775			
Phe	Gln	Tyr	Leu	Ser	Ser	Ser	Asp	Ser	Lys	Ile	Ser
				785				790			
Tyr	Glu	Thr	Phe	Lys	Lys	His	Ile	Leu	Asn	Leu	Ala
				805				810			
Lys	Gly	Lys	Gly	Arg	Ile	Ser	Lys	Thr	Lys	Lys	Glu
				820				825			
Tyr	Leu	Leu	Glu	Glu	Arg	Asp	Ile	Asn	Arg	Phe	Ser
				835				840			
Val	Gln	Lys	Asp	Phe	Ile	Asn	Arg	Asn	Leu	Val	Asp
				850				855			
Thr	Arg	Tyr	Ala	Thr	Arg	Gly	Leu	Met	Asn	Leu	Leu
				865				870			
Arg	Ser	Tyr	Phe	Arg	Val	Asn	Asn	Leu	Asp	Val	Lys
				885				890			
Val	Lys	Ser	Ile	Asn	Gly	Gly	Phe	Thr	Ser	Phe	Leu
				900				905			
Arg	Arg	Lys	Trp	Lys	Phe	Lys	Lys	Glu	Arg	Asn	Lys
				915				920			
Gly	Tyr	Lys	His	His	Ala	Glu	Asp	Ala	Leu	Ile	Ile
				930				935			
Ala	Asn	Ala	Asp	Phe	Ile	Phe	Lys	Glu	Trp	Lys	Lys
				945				950			
Leu	Asp	Lys	Ala	Lys	Lys	Val	Met	Glu	Asn	Gln	Met
				965				970			
Phe	Glu	Glu	Lys	Gln	Ala	Glu	Ser	Met	Pro	Glu	Ile
				980				985			
Glu	Thr	Glu	Ile	Glu	Gly	Glu	Tyr	Lys	Asp	Ile	Phe
				995				1000			
Ile	Thr	Pro	His	Gln	Ile	Met	His	Ile	Lys	Asp	Phe
				1010				1015			
Lys	Asp	Tyr	Lys	Tyr	Ser	His	Arg	Val	Asp	Lys	Lys
				1025				1030			
Pro	Asn	Arg	Glu	Leu	Ile	Asn	Asp	Thr	Leu	Tyr	Ser
				1040				1045			
Thr	Arg	Lys	Asp	Asp	Lys	Gly	Asn	Thr	Leu	Ile	Val
				1050							

```
Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
        115                 120                 125

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
    130                 135                 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
145                 150                 155                 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
                165                 170                 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
            180                 185                 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
        195                 200                 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
    210                 215                 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
225                 230                 235                 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
                245                 250                 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
            260                 265                 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
        275                 280                 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
    290                 295                 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
305                 310                 315                 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
                325                 330                 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
            340                 345                 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
```

```
        355                 360                 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
    370                 375                 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385                 390                 395                 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
                405                 410                 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
            420                 425                 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
        435                 440                 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
    450                 455                 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465                 470                 475                 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
                485                 490                 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
            500                 505                 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
        515                 520                 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
    530                 535                 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
545                 550                 555                 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
                565                 570                 575

Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
            580                 585                 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
        595                 600                 605
```

```
Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
    610                 615                 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625                 630                 635                 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
                645                 650                 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
            660                 665                 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
        675                 680                 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
    690                 695                 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705                 710                 715                 720

Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
                725                 730                 735

Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
            740                 745                 750

Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
        755                 760                 765

Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
    770                 775                 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
785                 790                 795                 800

Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
                805                 810                 815

Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
            820                 825                 830

Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
        835                 840                 845

Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
    850                 855                 860
```

```
Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
865                 870                 875                 880

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
                885                 890                 895

Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
            900                 905                 910

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
        915                 920                 925

Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
    930                 935                 940

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945                 950                 955                 960

Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
                965                 970                 975

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
            980                 985                 990

Glu Val Asn Met Ile Asp Ile Thr  Tyr Arg Glu Tyr Leu  Glu Asn Met
        995                 1000                1005

Asn Asp  Lys Arg Pro Pro Arg  Ile Ile Lys Thr Ile  Ala Ser Lys
    1010                1015                1020

Thr Gln  Ser Ile Lys Lys Tyr  Ser Thr Asp Ile Leu  Gly Asn Leu
    1025                1030                1035

Tyr Glu  Val Lys Ser Lys Lys  His Pro Gln Ile Ile  Lys Lys Gly
    1040                1045                1050
```

<210> 187
<211> 3243
<212> ADN
<213> Secuencia Artificial

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

<400> 187

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc      120
```

gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg 180

ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc 240

ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac 300

ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc 360

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg 540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac 600

tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720

ggagaaggga gcccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga 780

cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg 840

tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg 900

gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca 960

ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg 1020

acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc 1080

acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg 1140

actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg 1200

acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg 1260

tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt 1320

gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag 1380

atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc 1440

cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt 1500

atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag 1560

aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag 1620

aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg 1680

tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc 1740

gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc 1800

aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca 1860

gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag 1920

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc 1980

tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc 2040

ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc 2100

atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac 2160

aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt 2220

aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag 2280

aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc 2340

actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg 2400

gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat 2460

aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag 2520

ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag 2580

acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat 2640

aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc 2700

gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac 2760

gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat 2820

gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa 2880

aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag 2940

attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat 3000

ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat 3060

atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga 3120

attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg 3180

ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa 3240

ttc 3243

<210> 188
<211> 114
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 188
nnnnnnnnnn nnnnnnnnnn guuuuagcac uguacaagaa auugucgugc uaaaauaagg 60

cgcuguuaau gcagcugccg cauccgccag agcauuuaug cucuggcuuu uuuu 114

<210> 189
<211> 1269
<212> PRT
<213> Mycoplasma gallisepticum

<400> 189

Met Asn Asn Ser Ile Lys Ser Lys Pro Glu Val Thr Ile Gly Leu Asp
1 5 10 15

Leu Gly Val Gly Ser Val Gly Trp Ala Ile Val Asp Asn Glu Thr Asn
20 25 30

Ile Ile His His Leu Gly Ser Arg Leu Phe Ser Gln Ala Lys Thr Ala
35 40 45

Glu Asp Arg Arg Ser Phe Arg Gly Val Arg Arg Leu Ile Arg Arg Arg
50 55 60

Lys Tyr Lys Leu Lys Arg Phe Val Asn Leu Ile Trp Lys Tyr Asn Ser
65 70 75 80

Tyr Phe Gly Phe Lys Asn Lys Glu Asp Ile Leu Asn Asn Tyr Gln Glu
85 90 95

Gln Gln Lys Leu His Asn Thr Val Leu Asn Leu Lys Ser Glu Ala Leu
100 105 110

Asn Ala Lys Ile Asp Pro Lys Ala Leu Ser Trp Ile Leu His Asp Tyr
115 120 125

Leu Lys Asn Arg Gly His Phe Tyr Glu Asp Asn Arg Asp Phe Asn Val
130 135 140

Tyr Pro Thr Lys Glu Leu Ala Lys Tyr Phe Asp Lys Tyr Gly Tyr Tyr
145 150 155 160

Lys Gly Ile Ile Asp Ser Lys Glu Asp Asn Asp Asn Lys Leu Glu Glu
165 170 175

Glu Leu Thr Lys Tyr Lys Phe Ser Asn Lys His Trp Leu Glu Glu Val
180 185 190

Lys Lys Val Leu Ser Asn Gln Thr Gly Leu Pro Glu Lys Phe Lys Glu
195 200 205

Glu Tyr Glu Ser Leu Phe Ser Tyr Val Arg Asn Tyr Ser Glu Gly Pro

```
    210                 215                 220

Gly Ser Ile Asn Ser Val Ser Pro Tyr Gly Ile Tyr His Leu Asp Glu
225                 230                 235                 240

Lys Glu Gly Lys Val Val Gln Lys Tyr Asn Asn Ile Trp Asp Lys Thr
                245                 250                 255

Ile Gly Lys Cys Asn Ile Phe Pro Asp Glu Tyr Arg Ala Pro Lys Asn
            260                 265                 270

Ser Pro Ile Ala Met Ile Phe Asn Glu Ile Asn Glu Leu Ser Thr Ile
        275                 280                 285

Arg Ser Tyr Ser Ile Tyr Leu Thr Gly Trp Phe Ile Asn Gln Glu Phe
    290                 295                 300

Lys Lys Ala Tyr Leu Asn Lys Leu Leu Asp Leu Leu Ile Lys Thr Asn
305                 310                 315                 320

Gly Glu Lys Pro Ile Asp Ala Arg Gln Phe Lys Lys Leu Arg Glu Glu
                325                 330                 335

Thr Ile Ala Glu Ser Ile Gly Lys Glu Thr Leu Lys Asp Val Glu Asn
            340                 345                 350

Glu Glu Lys Leu Glu Lys Glu Asp His Lys Trp Lys Leu Lys Gly Leu
        355                 360                 365

Lys Leu Asn Thr Asn Gly Lys Ile Gln Tyr Asn Asp Leu Ser Ser Leu
    370                 375                 380

Ala Lys Phe Val His Lys Leu Lys Gln His Leu Lys Leu Asp Phe Leu
385                 390                 395                 400

Leu Glu Asp Gln Tyr Ala Thr Leu Asp Lys Ile Asn Phe Leu Gln Ser
                405                 410                 415

Leu Phe Val Tyr Leu Gly Lys His Leu Arg Tyr Ser Asn Arg Val Asp
            420                 425                 430

Ser Ala Asn Leu Lys Glu Phe Ser Asp Ser Asn Lys Leu Phe Glu Arg
        435                 440                 445

Ile Leu Gln Lys Gln Lys Asp Gly Leu Phe Lys Leu Phe Glu Gln Thr
    450                 455                 460
```

```
Asp Lys Asp Asp Glu Lys Ile Leu Ala Gln Thr His Ser Leu Ser Thr
465                 470                 475                 480

Lys Ala Met Leu Leu Ala Ile Thr Arg Met Thr Asn Leu Asp Asn Asp
                485                 490                 495

Glu Asp Asn Gln Lys Asn Asn Asp Lys Gly Trp Asn Phe Glu Ala Ile
            500                 505                 510

Lys Asn Phe Asp Gln Lys Phe Ile Asp Ile Thr Lys Lys Asn Asn Asn
        515                 520                 525

Leu Ser Leu Lys Gln Asn Lys Arg Tyr Leu Asp Asp Arg Phe Ile Asn
    530                 535                 540

Asp Ala Ile Leu Ser Pro Gly Val Lys Arg Ile Leu Arg Glu Ala Thr
545                 550                 555                 560

Lys Val Phe Asn Ala Ile Leu Lys Gln Phe Ser Glu Glu Tyr Asp Val
                565                 570                 575

Thr Lys Val Val Ile Glu Leu Ala Arg Glu Leu Ser Glu Glu Lys Glu
            580                 585                 590

Leu Glu Asn Thr Lys Asn Tyr Lys Lys Leu Ile Lys Lys Asn Gly Asp
        595                 600                 605

Lys Ile Ser Glu Gly Leu Lys Ala Leu Gly Ile Ser Glu Asp Glu Ile
    610                 615                 620

Lys Asp Ile Leu Lys Ser Pro Thr Lys Ser Tyr Lys Phe Leu Leu Trp
625                 630                 635                 640

Leu Gln Gln Asp His Ile Asp Pro Tyr Ser Leu Lys Glu Ile Ala Phe
                645                 650                 655

Asp Asp Ile Phe Thr Lys Thr Glu Lys Phe Glu Ile Asp His Ile Ile
            660                 665                 670

Pro Tyr Ser Ile Ser Phe Asp Asp Ser Ser Ser Asn Lys Leu Leu Val
        675                 680                 685

Leu Ala Glu Ser Asn Gln Ala Lys Ser Asn Gln Thr Pro Tyr Glu Phe
    690                 695                 700

Ile Ser Ser Gly Asn Ala Gly Ile Lys Trp Glu Asp Tyr Glu Ala Tyr
705                 710                 715                 720
```

```
Cys Arg Lys Phe Lys Asp Gly Asp Ser Ser Leu Leu Asp Ser Thr Gln
                725                 730                 735

Arg Ser Lys Lys Phe Ala Lys Met Met Lys Thr Asp Thr Ser Ser Lys
            740                 745                 750

Tyr Asp Ile Gly Phe Leu Ala Arg Asn Leu Asn Asp Thr Arg Tyr Ala
        755                 760                 765

Thr Ile Val Phe Arg Asp Ala Leu Glu Asp Tyr Ala Asn Asn His Leu
    770                 775                 780

Val Glu Asp Lys Pro Met Phe Lys Val Val Cys Ile Asn Gly Ser Val
785                 790                 795                 800

Thr Ser Phe Leu Arg Lys Asn Phe Asp Asp Ser Ser Tyr Ala Lys Lys
                805                 810                 815

Asp Arg Asp Lys Asn Ile His His Ala Val Asp Ala Ser Ile Ile Ser
            820                 825                 830

Ile Phe Ser Asn Glu Thr Lys Thr Leu Phe Asn Gln Leu Thr Gln Phe
        835                 840                 845

Ala Asp Tyr Lys Leu Phe Lys Asn Thr Asp Gly Ser Trp Lys Lys Ile
    850                 855                 860

Asp Pro Lys Thr Gly Val Val Thr Glu Val Thr Asp Glu Asn Trp Lys
865                 870                 875                 880

Gln Ile Arg Val Arg Asn Gln Val Ser Glu Ile Ala Lys Val Ile Glu
                885                 890                 895

Lys Tyr Ile Gln Asp Ser Asn Ile Glu Arg Lys Ala Arg Tyr Ser Arg
            900                 905                 910

Lys Ile Glu Asn Lys Thr Asn Ile Ser Leu Phe Asn Asp Thr Val Tyr
        915                 920                 925

Ser Ala Lys Lys Val Gly Tyr Glu Asp Gln Ile Lys Arg Lys Asn Leu
    930                 935                 940

Lys Thr Leu Asp Ile His Glu Ser Ala Lys Glu Asn Lys Asn Ser Lys
945                 950                 955                 960

Val Lys Arg Gln Phe Val Tyr Arg Lys Leu Val Asn Val Ser Leu Leu
                965                 970                 975
```

```
Asn Asn Asp Lys Leu Ala Asp Leu Phe Ala Glu Lys Glu Asp Ile Leu
            980                 985                 990

Met Tyr Arg Ala Asn Pro Trp Val  Ile Asn Leu Ala Glu  Gln Ile Phe
        995                1000                1005

Asn Glu  Tyr Thr Glu Asn Lys  Lys Ile Lys Ser Gln  Asn Val Phe
    1010                1015                1020

Glu Lys  Tyr Met Leu Asp Leu  Thr Lys Glu Phe Pro  Glu Lys Phe
    1025                1030                1035

Ser Glu  Phe Leu Val Lys Ser  Met Leu Arg Asn Lys  Thr Ala Ile
    1040                1045                1050

Ile Tyr  Asp Asp Lys Lys Asn  Ile Val His Arg Ile  Lys Arg Leu
    1055                1060                1065

Lys Met  Leu Ser Ser Glu Leu  Lys Glu Asn Lys Leu  Ser Asn Val
    1070                1075                1080

Ile Ile  Arg Ser Lys Asn Gln  Ser Gly Thr Lys Leu  Ser Tyr Gln
    1085                1090                1095

Asp Thr  Ile Asn Ser Leu Ala  Leu Met Ile Met Arg  Ser Ile Asp
    1100                1105                1110

Pro Thr  Ala Lys Lys Gln Tyr  Ile Arg Val Pro Leu  Asn Thr Leu
    1115                1120                1125

Asn Leu  His Leu Gly Asp His  Asp Phe Asp Leu His  Asn Met Asp
    1130                1135                1140

Ala Tyr  Leu Lys Lys Pro Lys  Phe Val Lys Tyr Leu  Lys Ala Asn
    1145                1150                1155

Glu Ile  Gly Asp Glu Tyr Lys  Pro Trp Arg Val Leu  Thr Ser Gly
    1160                1165                1170

Thr Leu  Leu Ile His Lys Lys  Asp Lys Lys Leu Met  Tyr Ile Ser
    1175                1180                1185

Ser Phe  Gln Asn Leu Asn Asp  Val Ile Glu Ile Lys  Asn Leu Ile
    1190                1195                1200

Glu Thr  Glu Tyr Lys Glu Asn  Asp Asp Ser Asp Ser  Lys Lys Lys
```

```
    1205                1210                1215

Lys Lys  Ala Asn Arg Phe Leu  Met Thr Leu Ser Thr  Ile Leu Asn
    1220                1225                1230

Asp Tyr  Ile Leu Leu Asp Ala  Lys Asp Asn Phe Asp  Ile Leu Gly
    1235                1240                1245

Leu Ser  Lys Asn Arg Ile Asp  Glu Ile Leu Asn Ser  Lys Leu Gly
    1250                1255                1260

Leu Asp  Lys Ile Val Lys
    1265
```

<210> 190
<211> 3891
<212> ADN
<213> Secuencia Artificial

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

<400> 190

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60
aaggtcgaag cgtccatgaa caatagcatc aaatctaaac ctgaagtgac catcgggctg      120
gacctgggag tgggaagcgt ggggtgggca atcgtggata acgaaacaaa catcattcac      180
catctgggct ccaggctgtt ttctcaggcc aagactgctg aggatcggag atctttccgc      240
ggggtgaggc gcctgatccg acggagaaaa tacaagctga aacgattcgt caatctgatt      300
tggaagtaca acagctattt cggcttcaag aacaaagagg acatcctgaa caattatcag      360
gagcagcaga agctgcacaa taccgtgctg aacctgaaat cagaggcact gaatgccaag      420
atcgatccta aagcactgag ctggattctg cacgactacc tgaagaacag aggccatttt      480
tatgaggaca atagggattt caacgtgtac ccaacaaagg agctggccaa gtacttcgat      540
aagtacgggt actacaaggg aatcattgac agcaaggagg acaatgataa caaactggag      600
gaagagctga caaagtacaa attctccaat aagcactggc tggaagaggt gaagaaagtc      660
ctgtctaacc agactggcct gccagaaaag tttaaagaag agtatgagtc actgttcagc      720
tacgtgagaa attattcaga gggcccaggg agcatcaact ctgtcagtcc ctacgggatc      780
taccatctgg acgaaaaaga gggaaaggtg gtccagaagt acaacaacat ctgggataag      840
acaatcggaa agtgcaacat cttccctgac gagtatagag ctcccaagaa cagtcctatc      900
gcaatgattt tcaatgaaat caacgagctg tccacaatca ggtcatacag catctacctg      960
actggctggt tcattaatca ggagttcaag aaagcctacc tgaacaagct gctggatctg     1020
```

ctgatcaaaa ccaacggaga gaagccaatt gacgcaaggc agttcaagaa actgcgcgaa 1080

gagacaatcg ccgaaagcat tggcaaagag acactgaagg atgtggagaa tgaagagaaa 1140

ctggaaaagg aggaccacaa gtggaaactg aagggactga agctgaatac caacggcaaa 1200

atccagtaca acgatctgag ctccctggct aagtttgtgc acaaactgaa gcagcatctg 1260

aaactggatt tcctgctgga ggaccagtat gcaacactgg acaagatcaa tttcctgcag 1320

tccctgtttg tgtacctggg caagcacctg agatattcca atagggtcga ttctgccaac 1380

ctgaaggaat tttccgactc taacaaactg ttcgagcgca tcctgcagaa acagaaggat 1440

gggctgttca agctgtttga acagactgac aaagacgatg agaagatcct ggcccagaca 1500

catagtctgt caactaaggc catgctgctg gctattaccc ggatgacaaa tctggacaac 1560

gatgaggaca accagaaaaa caatgacaag ggctggaatt ttgaggccat caaaaacttc 1620

gatcagaagt ttatcgacat caccaagaaa aacaacaacc tgagcctgaa acagaataag 1680

cgctacctgg acgatcgatt catcaacgat gctattctgt cccctggggt gaagcgaatc 1740

ctgcgggagg caaccaaggt ctttaatgcc attctgaaac agttctctga agagtacgac 1800

gtgacaaagg tggtcatcga actggctcgc gagctgagcg aagagaagga actggagaac 1860

acaaagaact acaagaaact gatcaagaaa aacggcgaca agattagtga gggcctgaaa 1920

gcactgggga tctcagaaga tgagatcaaa gacattctga agagtcccac taaatcatac 1980

aagtttctgc tgtggctgca gcaggaccac atcgatcctt atagcctgaa ggagatcgcc 2040

ttcgacgata tttttaccaa aacagaaaag ttcgagatcg accatatcat tccctacagc 2100

atttccttcg acgattctag ttcaaacaag ctgctggtgc tggctgaaag taatcaggca 2160

aagtcaaacc agactcctta tgagttcatc agctccggaa acgcaggcat taagtgggaa 2220

gattacgagg cctattgccg caagttcaag gatggggact ctagtctgct ggacagcacc 2280

cagcggtcca agaaattcgc caaaatgatg aaaaccgata cctcaagcaa gtacgacatc 2340

ggatttctgg ctcgaaatct gaacgatact cggtacgcaa ccattgtgtt ccgggacgcc 2400

ctggaggact atgctaataa ccacctggtc gaggacaaac ccatgtttaa ggtggtctgt 2460

atcaatgggt ccgtgacctc tttcctgcgg aagaactttg acgattcctc ttacgccaag 2520

aaagatagag acaagaatat ccaccatgct gtggatgcaa gtatcatctc aattttcagc 2580

aacgagacaa agactctgtt caaccagctg actcagtttg ctgactataa actgttcaag 2640

aacaccgatg gcagctggaa gaaaatcgac cctaagacag gggtggtcac tgaagtgacc 2700

gacgagaatt ggaagcagat tagggtgcgc aaccaggtga gcgaaatcgc caaagtcatt 2760

gagaagtaca tccaggatag caacatcgaa agaaaggcta ggtattcccg caaaatcgag 2820

aataagacta acatttccct gtttaatgac accgtgtact ctgccaagaa agtcggctat 2880

gaggatcaga tcaaaagaaa gaacctgaaa accctggaca ttcacgaatc tgctaaagag 2940

aataagaaca gtaaagtgaa gcggcagttt gtctacagaa agctggtgaa tgtcagcctg 3000

ctgaataacg ataagctggc agacctgttc gccgaaaaag aggatatcct gatgtatagg 3060

gccaatccat gggtcatcaa cctggctgag cagattttca atgaatacac tgagaacaag 3120

aaaatcaagt cccagaacgt gtttgaaaaa tatatgctgg acctgaccaa agagttcccc 3180

gagaagttca gcgagtttct ggtgaagtcc atgctgagaa acaagaccgc catcatctac 3240

gacgataaga aaaacattgt ccatcgaatc aaacggctga agatgctgag ttcagaactg 3300

aaagagaata agctgtctaa cgtgatcatt aggtctaaga atcagagtgg gaccaaactg 3360

tcataccagg atacaatcaa cagcctggcc ctgatgatta tgcgcagcat cgaccctact 3420

gctaagaaac agtatattcg agtgccactg aataccctga acctgcacct gggagatcat 3480

gactttgatc tgcacaatat ggatgcttac ctgaagaaac caaaattcgt gaagtatctg 3540

aaagcaaacg aaatcggcga cgagtacaag ccctggaggg tcctgacatc tggcactctg 3600

ctgatccata agaaggataa gaaactgatg tacatcagct ccttccagaa tctgaacgac 3660

gtgatcgaaa ttaagaatct gatcgaaacc gagtataaag agaacgacga ttctgatagt 3720

aagaaaaaga aaaaggcaaa ccgctttctg atgaccctga gcacaatcct gaatgactac 3780

attctgctgg acgccaagga taacttcgac atcctggggc tgtctaaaaa tcggatcgat 3840

gagattctga acagtaagct gggactggac aagattgtga ataagaatt c 3891

<210> 191
<211> 137
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 191
nnnnnnnnnn nnnnnnnnnn guuuuagucu cugaaaagag acuaaaauaa gugguuuuug 60

gucauccacg caggguuaca aucccuuuaa aaccauuaaa auucaaauaa acuagguugu 120

aucaacuuag uuuuuuu 137

<210> 192
<211> 1003

<212> PRT
<213> Campylobacter lari

<400> 192

```
Met Arg Ile Leu Gly Phe Asp Ile Gly Ile Asn Ser Ile Gly Trp Ala
1               5                  10                  15

Phe Val Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe Thr
            20                  25                  30

Lys Ala Glu Asn Pro Lys Asn Lys Glu Ser Leu Ala Leu Pro Arg Arg
        35                  40                  45

Asn Ala Arg Ser Ser Arg Arg Arg Leu Lys Arg Arg Lys Ala Arg Leu
    50                  55                  60

Ile Ala Ile Lys Arg Ile Leu Ala Lys Glu Leu Lys Leu Asn Tyr Lys
65                  70                  75                  80

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

Leu Ala Ser Val Tyr Glu Leu Arg Tyr Lys Ala Leu Thr Gln Asn Leu
            100                 105                 110

Glu Thr Lys Asp Leu Ala Arg Val Ile Leu His Ile Ala Lys His Arg
        115                 120                 125

Gly Tyr Met Asn Lys Asn Glu Lys Lys Ser Asn Asp Ala Lys Lys Gly
    130                 135                 140

Lys Ile Leu Ser Ala Leu Lys Asn Asn Ala Leu Lys Leu Glu Asn Tyr
145                 150                 155                 160

Gln Ser Val Gly Glu Tyr Phe Tyr Lys Glu Phe Phe Gln Lys Tyr Lys
                165                 170                 175

Lys Asn Thr Lys Asn Phe Ile Lys Ile Arg Asn Thr Lys Asp Asn Tyr
            180                 185                 190

Asn Asn Cys Val Leu Ser Ser Asp Leu Glu Lys Glu Leu Lys Leu Ile
        195                 200                 205

Leu Glu Lys Gln Lys Glu Phe Gly Tyr Asn Tyr Ser Glu Asp Phe Ile
    210                 215                 220

Asn Glu Ile Leu Lys Val Ala Phe Phe Gln Arg Pro Leu Lys Asp Phe
225                 230                 235                 240

Ser His Leu Val Gly Ala Cys Thr Phe Phe Glu Glu Glu Lys Arg Ala
```

```
                245                 250                 255

Cys Lys Asn Ser Tyr Ser Ala Trp Glu Phe Val Ala Leu Thr Lys Ile
            260                 265                 270

Ile Asn Glu Ile Lys Ser Leu Glu Lys Ile Ser Gly Glu Ile Val Pro
        275                 280                 285

Thr Gln Thr Ile Asn Glu Val Leu Asn Leu Ile Leu Asp Lys Gly Ser
    290                 295                 300

Ile Thr Tyr Lys Lys Phe Arg Ser Cys Ile Asn Leu His Glu Ser Ile
305                 310                 315                 320

Ser Phe Lys Ser Leu Lys Tyr Asp Lys Glu Asn Ala Glu Asn Ala Lys
                325                 330                 335

Leu Ile Asp Phe Arg Lys Leu Val Glu Phe Lys Lys Ala Leu Gly Val
            340                 345                 350

His Ser Leu Ser Arg Gln Glu Leu Asp Gln Ile Ser Thr His Ile Thr
        355                 360                 365

Leu Ile Lys Asp Asn Val Lys Leu Lys Thr Val Leu Glu Lys Tyr Asn
    370                 375                 380

Leu Ser Asn Glu Gln Ile Asn Asn Leu Leu Glu Ile Glu Phe Asn Asp
385                 390                 395                 400

Tyr Ile Asn Leu Ser Phe Lys Ala Leu Gly Met Ile Leu Pro Leu Met
                405                 410                 415

Arg Glu Gly Lys Arg Tyr Asp Glu Ala Cys Glu Ile Ala Asn Leu Lys
            420                 425                 430

Pro Lys Thr Val Asp Glu Lys Lys Asp Phe Leu Pro Ala Phe Cys Asp
        435                 440                 445

Ser Ile Phe Ala His Glu Leu Ser Asn Pro Val Val Asn Arg Ala Ile
    450                 455                 460

Ser Glu Tyr Arg Lys Val Leu Asn Ala Leu Leu Lys Lys Tyr Gly Lys
465                 470                 475                 480

Val His Lys Ile His Leu Glu Leu Ala Arg Asp Val Gly Leu Ser Lys
                485                 490                 495
```

```
Lys Ala Arg Glu Lys Ile Glu Lys Glu Gln Lys Glu Asn Gln Ala Val
            500                 505                 510

Asn Ala Trp Ala Leu Lys Glu Cys Glu Asn Ile Gly Leu Lys Ala Ser
        515                 520                 525

Ala Lys Asn Ile Leu Lys Leu Lys Leu Trp Lys Glu Gln Lys Glu Ile
    530                 535                 540

Cys Ile Tyr Ser Gly Asn Lys Ile Ser Ile Glu His Leu Lys Asp Glu
545                 550                 555                 560

Lys Ala Leu Glu Val Asp His Ile Tyr Pro Tyr Ser Arg Ser Phe Asp
                565                 570                 575

Asp Ser Phe Ile Asn Lys Val Leu Val Phe Thr Lys Glu Asn Gln Glu
            580                 585                 590

Lys Leu Asn Lys Thr Pro Phe Glu Ala Phe Gly Lys Asn Ile Glu Lys
        595                 600                 605

Trp Ser Lys Ile Gln Thr Leu Ala Gln Asn Leu Pro Tyr Lys Lys Lys
    610                 615                 620

Asn Lys Ile Leu Asp Glu Asn Phe Lys Asp Lys Gln Gln Glu Asp Phe
625                 630                 635                 640

Ile Ser Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ala Thr Leu Ile Ala
                645                 650                 655

Lys Tyr Thr Lys Glu Tyr Leu Asn Phe Leu Leu Leu Ser Glu Asn Glu
            660                 665                 670

Asn Ala Asn Leu Lys Ser Gly Glu Lys Gly Ser Lys Ile His Val Gln
        675                 680                 685

Thr Ile Ser Gly Met Leu Thr Ser Val Leu Arg His Thr Trp Gly Phe
    690                 695                 700

Asp Lys Lys Asp Arg Asn Asn His Leu His His Ala Leu Asp Ala Ile
705                 710                 715                 720

Ile Val Ala Tyr Ser Thr Asn Ser Ile Ile Lys Ala Phe Ser Asp Phe
                725                 730                 735

Arg Lys Asn Gln Glu Leu Leu Lys Ala Arg Phe Tyr Ala Lys Glu Leu
            740                 745                 750
```

```
Thr Ser Asp Asn Tyr Lys His Gln Val Lys Phe Phe Glu Pro Phe Lys
        755                 760                 765

Ser Phe Arg Glu Lys Ile Leu Ser Lys Ile Asp Glu Ile Phe Val Ser
    770                 775                 780

Lys Pro Pro Arg Lys Arg Ala Arg Arg Ala Leu His Lys Asp Thr Phe
785                 790                 795                 800

His Ser Glu Asn Lys Ile Ile Asp Lys Cys Ser Tyr Asn Ser Lys Glu
                805                 810                 815

Gly Leu Gln Ile Ala Leu Ser Cys Gly Arg Val Arg Lys Ile Gly Thr
            820                 825                 830

Lys Tyr Val Glu Asn Asp Thr Ile Val Arg Val Asp Ile Phe Lys Lys
        835                 840                 845

Gln Asn Lys Phe Tyr Ala Ile Pro Ile Tyr Ala Met Asp Phe Ala Leu
    850                 855                 860

Gly Ile Leu Pro Asn Lys Ile Val Ile Thr Gly Lys Asp Lys Asn Asn
865                 870                 875                 880

Asn Pro Lys Gln Trp Gln Thr Ile Asp Glu Ser Tyr Glu Phe Cys Phe
                885                 890                 895

Ser Leu Tyr Lys Asn Asp Leu Ile Leu Leu Gln Lys Lys Asn Met Gln
            900                 905                 910

Glu Pro Glu Phe Ala Tyr Tyr Asn Asp Phe Ser Ile Ser Thr Ser Ser
        915                 920                 925

Ile Cys Val Glu Lys His Asp Asn Lys Phe Glu Asn Leu Thr Ser Asn
    930                 935                 940

Gln Lys Leu Leu Phe Ser Asn Ala Lys Glu Gly Ser Val Lys Val Glu
945                 950                 955                 960

Ser Leu Gly Ile Gln Asn Leu Lys Val Phe Glu Lys Tyr Ile Ile Thr
                965                 970                 975

Pro Leu Gly Asp Lys Ile Lys Ala Asp Phe Gln Pro Arg Glu Asn Ile
            980                 985                 990

Ser Leu Lys Thr Ser Lys Lys Tyr  Gly Leu Arg
        995                 1000
```

<210> 193
<211> 3093
<212> ADN
<213> Secuencia Artificial

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

<400> 193

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgag gattctgggg tttgacattg gcattaacag catcgggtgg 120

gcttttgtgg agaacgacga actgaaggac tgcggagtgc ggatcttcac aaaggccgag 180

aacccaaaaa ataaggaaag cctggcactg ccccggagaa atgcacgcag ctccaggcgc 240

cgactgaaac ggagaaaggc ccggctgatc gctattaaga gaatcctggc caaagagctg 300

aagctgaact acaaggacta tgtcgcagct gatggagagc tgccaaaggc ctacgaagga 360

tccctggcat ctgtgtacga gctgcggtat aaggccctga cacagaacct ggaaactaaa 420

gatctggcca gagtgatcct gcacattgct aagcataggg ggtacatgaa caagaacgag 480

aagaaatcaa acgacgctaa gaaaggaaag atcctgagcg ctctgaaaaa caatgcactg 540

aagctggaga actaccagag cgtgggcgaa tacttctaca aggagttctt tcagaaatac 600

aagaaaaaca caaagaactt catcaagatc cgcaacacta aggataatta caacaattgc 660

gtgctgtcta gtgacctgga aaaagagctg aagctgatcc tggaaaaaca gaaggagttc 720

ggctacaact actctgaaga tttcatcaac gagattctga aggtcgcctt ctttcagcgg 780

cccctgaagg acttcagtca cctggtgggg gcctgcactt tctttgagga agagaaaagg 840

gcctgtaaga acagctactc tgcctgggag tttgtggctc tgaccaagat cattaacgag 900

atcaagagcc tggagaagat cagcggcgaa attgtgccaa cccagacaat caacgaggtc 960

ctgaatctga tcctggacaa ggggtctatc acctacaaga aattcagaag ttgtatcaat 1020

ctgcatgaga gtatcagctt caagagcctg aagtatgata aagaaaacgc cgagaatgct 1080

aaactgatcg acttccgcaa gctggtggag tttaagaaag ccctgggagt ccacagcctg 1140

tcccggcagg aactggatca gatctccact catatcaccc tgattaagga caacgtgaag 1200

ctgaaaaccg tcctggagaa atacaacctg agtaatgaac agatcaacaa tctgctggaa 1260

attgagttca acgattatat caacctgagc ttcaaggccc tgggaatgat tctgccactg 1320

atgcgcgagg gcaaacgata cgacgaggcc tgcgagatcg ccaatctgaa acctaagacc 1380

gtggacgaga agaaagattt cctgccagca ttttgtgatt ccattttcgc ccacgagctg 1440

tctaaccccg tggtcaatag ggctatcagc gaataccgca aggtgctgaa cgcactgctg 1500

```
aagaaatatg gaaaggtcca caaaattcat ctggagctgg ctcgcgacgt gggcctgtcc     1560
aagaaagcac gagagaagat cgaaaaagag cagaaggaaa accaggccgt gaatgcatgg     1620
gccctgaagg aatgcgagaa tattggcctg aaggccagcg caaagaacat cctgaaactg     1680
aagctgtgga aagaacagaa ggagatctgt atctactccg gaaataagat ctctattgag     1740
cacctgaaag atgaaaaggc cctggaggtg gaccatatct acccctattc taggagtttc     1800
gacgattctt ttatcaacaa agtgctggtg ttcaccaagg aaaatcagga gaaactgaac     1860
aagacacctt tcgaggcctt tggcaagaat attgaaaaat ggagcaagat ccagaccctg     1920
gctcagaacc tgccatacaa gaaaaagaat aagattctgg acgagaactt caaagataag     1980
cagcaggagg actttatctc tcgaaatctg aacgacaccc ggtatatcgc tacactgatt     2040
gcaaaataca caaaggagta tctgaacttc ctgctgctga gcgaaaatga gaacgccaat     2100
ctgaagagtg gcgaaaaagg gtcaaagatc cacgtgcaga ctattagcgg gatgctgacc     2160
tccgtcctga ggcacacatg ggggtttgac aaaaaggatc gcaacaatca tctgcaccat     2220
gcactggatg ccatcattgt ggcctacagt acaaattcaa tcattaaggc tttcagcgat     2280
ttccggaaaa accaggagct gctgaaggcc agattctacg ctaaagaact gacttccgat     2340
aactataaac atcaggtcaa gttctttgag cctttcaaga gttttagaga aaaaatcctg     2400
tcaaagatcg acgagatttt cgtgtccaaa ccacctcgaa agcgagctag gcgcgcactg     2460
cacaaggata cctttcattc tgagaacaag atcattgaca agtgcagcta caactccaag     2520
gaaggcctgc agattgccct gagctgtgga agagtgagga aaatcggcac taagtatgtc     2580
gagaatgata ccatcgtgag ggtcgacatt ttcaaaaagc agaacaagtt ttacgctatc     2640
ccaatctacg caatggattt tgccctgggg atcctgccca ataagatcgt gattactgga     2700
aaagataaga acaataaccc caaacagtgg cagaccattg acgaatcata cgagttctgc     2760
tttagcctgt ataagaatga cctgatcctg ctgcagaaaa agaacatgca ggaacctgag     2820
ttcgcctact ataacgattt ttcaatcagc acatcaagca tttgtgtgga gaaacacgac     2880
aacaagttcg aaaatctgac tagcaaccag aagctgctgt tttccaatgc aaaagagggc     2940
tctgtgaagg tcgaaagtct ggggatccag aacctgaaag tgttcgagaa gtacatcatt     3000
acccccctgg gagataaaat taaggctgac tttcagcctc gagaaaacat cagcctgaaa     3060
accagtaaaa agtatggcct gaggtaagaa ttc                                  3093
```

<210> 194
<211> 103
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 194
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 195
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<400> 195
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

```
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

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

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

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

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430
```

```
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
```

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

```
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170
```

```
Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                  1180                  1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                  1195                  1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                  1210                  1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                  1225                  1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                  1240                  1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                  1255                  1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                  1270                  1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                  1285                  1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                  1300                  1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                  1315                  1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                  1330                  1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                  1345                  1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                  1360                  1365
```

<210> 196
<211> 4218
<212> ADN
<213> Secuencia Artificial

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

<400> 196

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccgacaa gaagtacagc atcggcctgg acatcggcac caactctgtg 120

ggctgggccg tgatcaccga cgagtacaag gtgcccagca agaaattcaa ggtgctgggc 180

aacaccgacc ggcacagcat caagaagaac ctgatcggag ccctgctgtt cgacagcggc 240

gaaacagccg aggccacccg gctgaagaga accgccagaa gaagatacac cagacggaag 300

aaccggatct gctatctgca agagatcttc agcaacgaga tggccaaggt ggacgacagc 360

ttcttccaca gactggaaga gtccttcctg gtggaagagg ataagaagca cgagcggcac 420

cccatcttcg gcaacatcgt ggacgaggtg gcctaccacg agaagtaccc caccatctac 480

cacctgagaa agaaactggt ggacagcacc gacaaggccg acctgcggct gatctatctg 540

gccctggccc acatgatcaa gttccggggc cacttcctga tcgagggcga cctgaacccc 600

gacaacagcg acgtggacaa gctgttcatc cagctggtgc agacctacaa ccagctgttc 660

gaggaaaacc ccatcaacgc cagcggcgtg gacgccaagg ccatcctgtc tgccagactg 720

agcaagagca gacggctgga aaatctgatc gcccagctgc ccggcgagaa gaagaatggc 780

ctgttcggca acctgattgc cctgagcctg ggcctgaccc ccaacttcaa gagcaacttc 840

gacctggccg aggatgccaa actgcagctg agcaaggaca cctacgacga cgacctggac 900

aacctgctgg cccagatcgg cgaccagtac gccgacctgt ttctggccgc caagaacctg 960

tccgacgcca tcctgctgag cgacatcctg agagtgaaca ccgagatcac caaggccccc 1020

ctgagcgcct ctatgatcaa gagatacgac gagcaccacc aggacctgac cctgctgaaa 1080

gctctcgtgc ggcagcagct gcctgagaag tacaaagaga ttttcttcga ccagagcaag 1140

aacggctacg ccggctacat tgacggcgga gccagccagg aagagttcta caagttcatc 1200

aagcccatcc tggaaaagat ggacggcacc gaggaactgc tcgtgaagct gaacagagag 1260

gacctgctgc ggaagcagcg gaccttcgac aacggcagca tccccccacca gatccacctg 1320

ggagagctgc acgccattct gcggcggcag gaagattttt acccattcct gaaggacaac 1380

cgggaaaaga tcgagaagat cctgaccttc cgcatccccct actacgtggg ccctctggcc 1440

aggggaaaca gcagattcgc ctggatgacc agaaagagcg aggaaaccat cacccctgg 1500

aacttcgagg aagtggtgga caagggcgct tccgcccaga gcttcatcga gcggatgacc 1560

aacttcgata agaacctgcc caacgagaag gtgctgccca agcacagcct gctgtacgag 1620

tacttcaccg tgtataacga gctgaccaaa gtgaaatacg tgaccgaggg aatgagaaag 1680

cccgccttcc tgagcggcga gcagaaaaag gccatcgtgg acctgctgtt caagaccaac 1740

cggaaagtga ccgtgaagca gctgaaagag gactacttca agaaaatcga gtgcttcgac 1800

tccgtggaaa tctccggcgt ggaagatcgg ttcaacgcct ccctgggcac ataccacgat 1860

ctgctgaaaa ttatcaagga caaggacttc ctggacaatg aggaaaacga ggacattctg 1920

gaagatatcg tgctgaccct gacactgttt gaggacagag agatgatcga ggaacggctg 1980

aaaacctatg cccacctgtt cgacgacaaa gtgatgaagc agctgaagcg gcggagatac 2040

accggctggg gcaggctgag ccggaagctg atcaacggca tccgggacaa gcagtccggc 2100

aagacaatcc tggatttcct gaagtccgac ggcttcgcca acagaaactt catgcagctg 2160

atccacgacg acagcctgac ctttaaagag gacatccaga aagcccaggt gtccggccag 2220

ggcgatagcc tgcacgagca cattgccaat ctggccggca gccccgccat taagaagggc 2280

atcctgcaga cagtgaaggt ggtggacgag ctcgtgaaag tgatgggccg gcacaagccc 2340

gagaacatcg tgatcgaaat ggccagagag aaccagacca cccagaaggg acagaagaac 2400

agccgcgaga gaatgaagcg gatcgaagag ggcatcaaag agctgggcag ccagatcctg 2460

aaagaacacc ccgtggaaaa cacccagctg cagaacgaga agctgtacct gtactacctg 2520

cagaatgggc gggatatgta cgtggaccag gaactggaca tcaaccggct gtccgactac 2580

gatgtggacc atatcgtgcc tcagagcttt ctgaaggacg actccatcga caacaaggtg 2640

ctgaccagaa gcgacaagaa ccggggcaag agcgacaacg tgccctccga agaggtcgtg 2700

aagaagatga agaactactg gcggcagctg ctgaacgcca agctgattac ccagagaaag 2760

ttcgacaatc tgaccaaggc cgagagaggc ggcctgagcg aactggataa ggccggcttc 2820

atcaagagac agctggtgga aacccggcag atcacaaagc acgtggcaca gatcctggac 2880

tcccggatga acactaagta cgacgagaat gacaagctga tccgggaagt gaaagtgatc 2940

accctgaagt ccaagctggt gtccgatttc cggaaggatt tccagtttta caaagtgcgc 3000

gagatcaaca actaccacca cgcccacgac gcctacctga acgccgtcgt gggaaccgcc 3060

ctgatcaaaa agtaccctaa gctggaaagc gagttcgtgt acggcgacta caaggtgtac 3120

gacgtgcgga agatgatcgc caagagcgag caggaaatcg gcaaggctac cgccaagtac 3180

ttcttctaca gcaacatcat gaacttttтc aagaccgaga ttaccctggc caacggcgag 3240

atccggaagc ggcctctgat cgagacaaac ggcgaaaccg gggagatcgt gtgggataag 3300

ggccgggatt ttgccaccgt gcggaaagtg ctgagcatgc cccaagtgaa tatcgtgaaa 3360

aagaccgagg tgcagacagg cggcttcagc aaagagtcta tcctgcccaa gaggaacagc 3420

gataagctga tcgccagaaa gaaggactgg gaccctaaga agtacggcgg cttcgacagc 3480

cccaccgtgg cctattctgt gctggtggtg gccaaagtgg aaaagggcaa gtccaagaaa 3540

ctgaagagtg tgaaagagct gctggggatc accatcatgg aaagaagcag cttcgagaag 3600

aatcccatcg actttctgga agccaagggc tacaaagaag tgaaaaagga cctgatcatc 3660

aagctgccta agtactccct gttcgagctg gaaaacggcc ggaagagaat gctggcctct 3720

gccggcgaac tgcagaaggg aaacgaactg gccctgccct ccaaatatgt gaacttcctg 3780

tacctggcca gccactatga gaagctgaag ggctcccccg aggataatga gcagaaacag 3840

ctgtttgtgg aacagcacaa gcactacctg gacgagatca tcgagcagat cagcgagttc 3900

tccaagagag tgatcctggc cgacgctaat ctggacaaag tgctgtccgc ctacaacaag 3960

caccgggata agcccatcag agagcaggcc gagaatatca tccacctgtt taccctgacc 4020

aatctgggag cccctgccgc cttcaagtac tttgacacca ccatcgaccg gaagaggtac 4080

accagcacca aagaggtgct ggacgccacc ctgatccacc agagcatcac cggcctgtac 4140

gagacacgga tcgacctgtc tcagctggga ggcgacagcc ccaagaagaa gagaaaggtg 4200

gaggccagct aagaattc 4218

<210> 197
<211> 116
<212> ARN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, desconocida u otra

<400> 197
nnnnnnnnnn nnnnnnnnnn guuuuuguac ucucaagauu uagaaacuug cagaagcuac 60

aaagauaagg cuucaugccg aaaucaacac ccugucauuu uauggcaggg uguuuu 116

<210> 198
<211> 1120
<212> PRT
<213> Streptococcus thermophilus

<400> 198

```
Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1               5                  10                  15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
            20                  25                  30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
        35                  40                  45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
    50                  55                  60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
```

65 70 75 80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
85 90 95

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
100 105 110

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

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
130 135 140

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
145 150 155 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
165 170 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
180 185 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
195 200 205

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

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
225 230 235 240

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
245 250 255

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
260 265 270

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
275 280 285

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
290 295 300

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
305 310 315 320

```
Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
                325                 330                 335

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
            340                 345                 350

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
        355                 360                 365

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
    370                 375                 380

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
385                 390                 395                 400

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
                405                 410                 415

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
            420                 425                 430

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
        435                 440                 445

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
    450                 455                 460

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
465                 470                 475                 480

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
                485                 490                 495

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
            500                 505                 510

Glu Thr Asn Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys
        515                 520                 525

Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln
    530                 535                 540

Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys
545                 550                 555                 560

Gln Leu Ala Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys
                565                 570                 575
```

```
Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser
            580                 585                 590

Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp
        595                 600                 605

Asp Ser Leu Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu
    610                 615                 620

Lys Gly Gln Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala
625                 630                 635                 640

Trp Ser Phe Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu
                645                 650                 655

Ser Asn Lys Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys
            660                 665                 670

Phe Asp Val Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg
        675                 680                 685

Tyr Ala Ser Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala
    690                 695                 700

His Lys Ile Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser
705                 710                 715                 720

Gln Leu Arg Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His
                725                 730                 735

His His Ala Val Asp Ala Leu Ile Ile Ala Ala Ser Ser Gln Leu Asn
            740                 745                 750

Leu Trp Lys Lys Gln Lys Asn Thr Leu Val Ser Tyr Ser Glu Asp Gln
        755                 760                 765

Leu Leu Asp Ile Glu Thr Gly Glu Leu Ile Ser Asp Asp Glu Tyr Lys
    770                 775                 780

Glu Ser Val Phe Lys Ala Pro Tyr Gln His Phe Val Asp Thr Leu Lys
785                 790                 795                 800

Ser Lys Glu Phe Glu Asp Ser Ile Leu Phe Ser Tyr Gln Val Asp Ser
                805                 810                 815

Lys Phe Asn Arg Lys Ile Ser Asp Ala Thr Ile Tyr Ala Thr Arg Gln
            820                 825                 830
```

```
Ala Lys Val Gly Lys Asp Lys Ala Asp Glu Thr Tyr Val Leu Gly Lys
        835                 840                 845

Ile Lys Asp Ile Tyr Thr Gln Asp Gly Tyr Asp Ala Phe Met Lys Ile
    850                 855                 860

Tyr Lys Lys Asp Lys Ser Lys Phe Leu Met Tyr Arg His Asp Pro Gln
865                 870                 875                 880

Thr Phe Glu Lys Val Ile Glu Pro Ile Leu Glu Asn Tyr Pro Asn Lys
                885                 890                 895

Gln Ile Asn Glu Lys Gly Lys Glu Val Pro Cys Asn Pro Phe Leu Lys
            900                 905                 910

Tyr Lys Glu Glu His Gly Tyr Ile Arg Lys Tyr Ser Lys Lys Gly Asn
        915                 920                 925

Gly Pro Glu Ile Lys Ser Leu Lys Tyr Tyr Asp Ser Lys Leu Gly Asn
    930                 935                 940

His Ile Asp Ile Thr Pro Lys Asp Ser Asn Asn Lys Val Val Leu Gln
945                 950                 955                 960

Ser Val Ser Pro Trp Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly
                965                 970                 975

Lys Tyr Glu Ile Leu Gly Leu Lys Tyr Ala Asp Leu Gln Phe Glu Lys
            980                 985                 990

Gly Thr Gly Thr Tyr Lys Ile Ser  Gln Glu Lys Tyr Asn  Asp Ile Lys
        995                 1000                1005

Lys Lys  Glu Gly Val Asp Ser  Asp Ser Glu Phe Lys  Phe Thr Leu
    1010                1015                1020

Tyr Lys  Asn Asp Leu Leu Leu  Val Lys Asp Thr Glu  Thr Lys Glu
    1025                1030                1035

Gln Gln  Leu Phe Arg Phe Leu  Ser Arg Thr Met Pro  Lys Gln Lys
    1040                1045                1050

His Tyr  Val Glu Leu Lys Pro  Tyr Asp Lys Gln Lys  Phe Glu Gly
    1055                1060                1065

Gly Glu  Ala Leu Ile Lys Val  Leu Gly Asn Val Ala  Asn Ser Gly
```

```
    1070                1075                1080

Gln Cys  Lys Lys Gly Leu Gly  Lys Ser Asn Ile Ser  Ile Tyr Lys
    1085                1090                1095

Val Arg  Thr Asp Val Leu Gly  Asn Gln His Ile Ile  Lys Asn Glu
    1100                1105                1110

Gly Asp  Lys Pro Lys Leu Met
    1115                1120
```

<210> 199
<211> 3477
<212> ADN
<213> Secuencia Artificial

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

<400> 199

```
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc       60

aaggtcgaag cgtcctccga cctggtactt ggactggata ttggtatcgg ttcggtggga      120

gtcggaatcc tcaacaaggt cacgggggag atcattcaca agaactcgcg gatcttcccc      180

gcagctcagg ctgagaacaa cttggtgcgg agaacgaata ggcagggcag gcgactggcg      240

aggaggaaga aacacaggag agtccgattg aaccggctgt tcgaggagtc cggtttgatc      300

accgacttta cgaaaatctc gattaacctt aatccctatc agcttcgggt gaaaggcctg      360

acagacgaac tttcgaatga ggaacttttc atcgcgctga aaaacatggt caagcacaga      420

gggatttcct acctcgatga cgcctcggat gacggaaatt cctcagtagg agattatgca      480

cagatcgtga aagagaactc aaagcaactg gaaacaaaga caccggggca gatccaactt      540

gaaagatacc agacatacgg acagctcaga ggagatttta cggtggagaa ggacggtaaa      600

aagcacagac tcattaacgt atttcccacg tcggcgtaca gatccgaagc gctccgcatc      660

cttcagactc aacaggagtt caacccgcaa attactgatg agttcatcaa ccgctatttg      720

gaaatcttga ccggaaagcg caagtattat catgggccgg gtaatgagaa atccagaaca      780

gattacggcc gatacagaac ttcgggggaa accttggata acatctttgg tattttgatt      840

ggaaagtgca ccttttaccc ggacgagttt cgagcggcca aggcgtcata cacagcacaa      900

gagtttaatc tcttgaatga tttgaacaac ttgacggtcc ccacggagac aaagaagctc      960

tccaaagagc aaaagaacca aatcatcaac tacgtcaaga acgagaaggc tatggggcca     1020

gcgaagctgt tcaagtatat cgctaaactt ctcagctgtg atgtggcgga catcaaaggg     1080

taccgaatcg acaagtcggg aaaagcggaa attcacacgt ttgaagcata tcgaaagatg     1140
```

aaaacgttgg aaacactgga cattgagcag atggaccggg aaacgctcga caaactggca 1200

tacgtgctca cgttgaatac tgaacgagag ggaatccaag aggcccttga acatgagttc 1260

gccgatggat cgttcagcca gaagcaggtc gacgaacttg tgcaattccg caaggcgaat 1320

agctccatct tcgggaaggg atggcacaac ttttcggtca aactcatgat ggagttgatc 1380

ccagaacttt atgagacttc ggaggagcaa atgacgatct tgacgcgctt ggggaaacag 1440

aaaacgacaa gctcatcgaa caaaactaag tacattgatg agaaattgct gacggaagaa 1500

atctataatc cggtagtagc gaaatcggta agacaagcga tcaaaatcgt gaacgcggcg 1560

atcaaggaat atggtgactt tgataacatc gtaattgaaa tggctagaga gacgaacgaa 1620

gatgacgaga aaaaggcaat ccagaagatc cagaaggcca acaaggatga aaaagatgca 1680

gcgatgctta aagcggccaa ccaatacaat ggaaaggcgg agctgcccca ttcagtgttt 1740

cacggtcata aacagttggc gaccaagatc cgactctggc atcagcaggg tgagcggtgt 1800

ctctacaccg gaaagactat ctccatccat gacttgatta acaattcgaa ccagtttgaa 1860

gtggatcata ttctgccccc gtcaatcacc tttgacgact cgcttgcgaa caaggtgctc 1920

gtgtacgcaa cggcaaatca ggagaaaggc cagcggactc cgtatcaggc gctcgactca 1980

atggacgatg cgtggtcatt ccgggagctg aaggcgttcg tacgcgagag caagacactg 2040

agcaacaaaa agaaagagta tctgctgaca gaggaggaca tctcgaaatt cgatgtcagg 2100

aagaagttca tcgagcggaa tcttgtcgac actcgctacg cttccagagt agtactgaac 2160

gcgctccagg aacactttag agcgcacaaa attgacacga aggtgtcagt ggtgagaggg 2220

cagttcacat cccaactccg ccgacattgg ggcatcgaaa agacgcggga cacatatcac 2280

catcatgcgg tggacgcgct gattattgcc gcttcgtccc agttgaatct ctggaaaaag 2340

cagaagaaca cgctggtgtc gtattcggag gatcagcttt tggacatcga aaccggggag 2400

ctgatttccg acgatgaata caaagaatcg gtgtttaagg caccatatca gcatttcgtg 2460

gacacgctga agagcaaaga gtttgaggac agcatcctct tttcgtacca agtggactcg 2520

aagtttaatc gcaagatttc agacgccaca atctacgcga cgaggcaggc gaaggtgggc 2580

aaagataaag cagatgaaac ctacgtcctt ggtaaaatca aggacatcta cactcaggac 2640

gggtacgatg cgttcatgaa aatctacaag aaggataagt cgaagtttct catgtaccgc 2700

cacgatccac agactttcga aaaagtcatt gagcctattt tggagaacta ccctaacaag 2760

caaatcaacg agaaagggaa agaagtcccg tgcaacccct ttctgaagta caaggaagag 2820

cacggttata tccgcaaata ctcgaagaaa ggaaatgggc ctgagattaa gtcgcttaag 2880

tattacgact caaagttggg taaccacatc gacattaccc cgaaagactc caacaacaaa 2940

gtcgtgttgc agtccgtctc gccctggcga gcagatgtgt attttaataa gacgaccggc 3000

aaatatgaga tccttggact caaatacgca gaccttcaat tcgaaaaggg gacgggcact 3060

tataagattt cacaagagaa gtacaacgac atcaagaaaa aggaaggggt cgattcagat 3120

tcggagttca aattcaccct ctacaaaaac gacctcctgc ttgtgaagga cacagaaacg 3180

aaggagcagc agctctttcg gttcctctca cgcacgatgc ccaaacaaaa acattacgtc 3240

gaacttaaac cttacgataa gcaaaagttt gaaggggggag aggcactgat caaagtattg 3300

ggtaacgtag ccaatagcgg acagtgtaag aaagggctgg gaaagtccaa tatctcgatc 3360

tataaagtac gaacagatgt attgggaaac cagcatatca tcaaaaatga gggggataaa 3420

cccaaactcg atttcagccc caagaagaag agaaaggtgg aggccagcta agaattc 3477

<210> 200
<211> 27
<212> ADN
<213> Secuencia Artificial

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

<400> 200
gcccgggtgg aactggtagc catgaat 27

<210> 201
<211> 26
<212> ADN
<213> Secuencia Artificial

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

<400> 201
gttgaagatg aagcccagag cggagt 26

<210> 202
<211> 27
<212> ADN
<213> Secuencia Artificial

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

<400> 202
gcttccgacg aggtggccat caaggat 27

<210> 203
<211> 27
<212> ADN
<213> Secuencia Artificial

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

<400> 203
gcaccatctc tccgtggtac cccgggt 27

<210> 204
<211> 26
<212> ADN
<213> Secuencia Artificial

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

<400> 204
ggtggaactg gtagccatga atgaga 26

<210> 205
<211> 27
<212> ADN
<213> Secuencia Artificial

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

<400> 205
gccatgaatg agaccgaccc aaagagc 27

<210> 206
<211> 27

<212> ADN
<213> Secuencia Artificial

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

<400> 206
gcatcctcgt gggcacttcc gacgagg 27

<210> 207
<211> 27
<212> ADN
<213> Secuencia Artificial

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

<400> 207
gcagagcgga gtgctgttct cccaagt 27

<210> 208
<211> 26
<212> ADN
<213> Secuencia Artificial

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

<400> 208
ggtcggtctc attcatggct accagt 26

<210> 209
<211> 26
<212> ADN
<213> Secuencia Artificial

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

ç<400> 209
gcaataaaag gtgctattgc tatagt 26

<210> 210
<211> 26
<212> ADN
<213> Secuencia Artificial

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

<400> 210
gtactcaacc aagtcattct gagaat 26

<210> 211
<211> 26
<212> ADN
<213> Secuencia Artificial

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

<220>
<221> base_modificada
<222> (1)..(7)
<223> a, c, t, g, desconocida u otra

<400> 211
nnnnnnngta ctcaaccaag tcattc 26

<210> 212
<211> 1121
<212> PRT
<213> Streptococcus thermophilus

<400> 212

```
Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1               5                  10                  15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
            20                  25                  30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
        35                  40                  45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
    50                  55                  60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
65                  70                  75                  80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
                85                  90                  95

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
            100                 105                 110

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

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
    130                 135                 140

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
145                 150                 155                 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
                165                 170                 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
            180                 185                 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
        195                 200                 205

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
    210                 215                 220
```

```
Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
225                 230                 235                 240

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
                245                 250                 255

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
            260                 265                 270

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
        275                 280                 285

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
    290                 295                 300

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
305                 310                 315                 320

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
                325                 330                 335

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
            340                 345                 350

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
        355                 360                 365

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
    370                 375                 380

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
385                 390                 395                 400

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
                405                 410                 415

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
            420                 425                 430

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
        435                 440                 445

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
    450                 455                 460

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
465                 470                 475                 480
```

```
Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
                485                 490                 495

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
            500                 505                 510

Glu Thr Asn Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys
        515                 520                 525

Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln
    530                 535                 540

Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys
545                 550                 555                 560

Gln Leu Ala Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys
                565                 570                 575

Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser
            580                 585                 590

Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp
        595                 600                 605

Asp Ser Leu Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu
    610                 615                 620

Lys Gly Gln Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala
625                 630                 635                 640

Trp Ser Phe Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu
                645                 650                 655

Ser Asn Lys Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys
            660                 665                 670

Phe Asp Val Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg
        675                 680                 685

Tyr Ala Ser Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala
    690                 695                 700

His Lys Ile Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser
705                 710                 715                 720

Gln Leu Arg Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His
```

```
                725                 730                 735

His His Ala Val Asp Ala Leu Ile Ile Ala Ala Ser Ser Gln Leu Asn
            740                 745                 750

Leu Trp Lys Lys Gln Lys Asn Thr Leu Val Ser Tyr Ser Glu Asp Gln
        755                 760                 765

Leu Leu Asp Ile Glu Thr Gly Glu Leu Ile Ser Asp Asp Glu Tyr Lys
    770                 775                 780

Glu Ser Val Phe Lys Ala Pro Tyr Gln His Phe Val Asp Thr Leu Lys
785                 790                 795                 800

Ser Lys Glu Phe Glu Asp Ser Ile Leu Phe Ser Tyr Gln Val Asp Ser
                805                 810                 815

Lys Phe Asn Arg Lys Ile Ser Asp Ala Thr Ile Tyr Ala Thr Arg Gln
            820                 825                 830

Ala Lys Val Gly Lys Asp Lys Ala Asp Glu Thr Tyr Val Leu Gly Lys
        835                 840                 845

Ile Lys Asp Ile Tyr Thr Gln Asp Gly Tyr Asp Ala Phe Met Lys Ile
    850                 855                 860

Tyr Lys Lys Asp Lys Ser Lys Phe Leu Met Tyr Arg His Asp Pro Gln
865                 870                 875                 880

Thr Phe Glu Lys Val Ile Glu Pro Ile Leu Glu Asn Tyr Pro Asn Lys
                885                 890                 895

Gln Ile Asn Glu Lys Gly Lys Glu Val Pro Cys Asn Pro Phe Leu Lys
            900                 905                 910

Tyr Lys Glu Glu His Gly Tyr Ile Arg Lys Tyr Ser Lys Lys Gly Asn
        915                 920                 925

Gly Pro Glu Ile Lys Ser Leu Lys Tyr Tyr Asp Ser Lys Leu Gly Asn
    930                 935                 940

His Ile Asp Ile Thr Pro Lys Asp Ser Asn Asn Lys Val Val Leu Gln
945                 950                 955                 960

Ser Val Ser Pro Trp Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly
                965                 970                 975
```

```
Lys Tyr Glu Ile Leu Gly Leu Lys Tyr Ala Asp Leu Gln Phe Glu Lys
            980                 985                 990

Gly Thr Gly Thr Tyr Lys Ile Ser  Gln Glu Lys Tyr Asn  Asp Ile Lys
        995                 1000                1005

Lys Lys  Glu Gly Val Asp Ser  Asp Ser Glu Phe Lys  Phe Thr Leu
    1010                1015                1020

Tyr Lys  Asn Asp Leu Leu Leu  Val Lys Asp Thr Glu  Thr Lys Glu
    1025                1030                1035

Gln Gln  Leu Phe Arg Phe Leu  Ser Arg Thr Met Pro  Lys Gln Lys
    1040                1045                1050

His Tyr  Val Glu Leu Lys Pro  Tyr Asp Lys Gln Lys  Phe Glu Gly
    1055                1060                1065

Gly Glu  Ala Leu Ile Lys Val  Leu Gly Asn Val Ala  Asn Ser Gly
    1070                1075                1080

Gln Cys  Lys Lys Gly Leu Gly  Lys Ser Asn Ile Ser  Ile Tyr Lys
    1085                1090                1095

Val Arg  Thr Asp Val Leu Gly  Asn Gln His Ile Ile  Lys Asn Glu
    1100                1105                1110

Gly Asp  Lys Pro Lys Leu Asp  Phe
    1115                1120
```

<210> 213
<211> 20
<212> ADN
<213> Corynebacter diphtheriae

<400> 213
aaaagggaat aagggcgaca 20

<210> 214
<211> 27
<212> ADN
<213> Sutterella wadsworthensis

<220>
<221> base_modificada
<222> (1)..(1)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (4)..(7)
<223> a, c, t, g, desconocida u otra

<400> 214
nggnnnncat tctgagaata gtgtatg 27

<210> 215
<211> 20
<212> ADN
<213> Sutterella wadsworthensis

<400> 215
gcgaccgagt tgctcttgcc 20

<210> 216
<211> 27
<212> ADN
<213> Legionella pneumophila

<220>
<221> base_modificada
<222> (1)..(1)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (4)..(7)
<223> a, c, t, g, desconocida u otra

<400> 216
nggnnnncat tctgagaata gtgtatg 27

<210> 217
<211> 20
<212> ADN
<213> Legionella pneumophila

<400> 217
gcgaccgagt tgctcttgcc 20

<210> 218
<211> 27

<212> ADN
<213> Filifactor alocis

<220>
<221> base_modificada
<222> (1)..(3)
<223> a, c, t, g, desconocida u otra

<400> 218
nnnaagcttg agtactcacc agtcaca 27

<210> 219
<211> 20
<212> ADN
<213> Filifactor alocis

<400> 219
ctcttccttt ttcaatatta 20

<210> 220
<211> 27
<212> ADN
<213> Lactobacillus johnsonii

<220>
<221> base_modificada
<222> (1)..(2)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (6)..(7)
<223> a, c, t, g, desconocida u otra

<400> 220
nnaaannntaa atgcttcaat aatattg 27

<210> 221
<211> 20
<212> ADN
<213> Lactobacillus johnsonii

<400> 221
aaagggaata agggcgacac 20

<210> 222
<211> 27
<212> ADN
<213> Parvibaculum lavamentivorans

<220>
<221> base_modificada
<222> (1)..(3)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (7)..(7)
<223> a, c, t, g, desconocida u otra

<400> 222
nnncatnaaa ggaagagtat gagtatt 27

<210> 223
<211> 20
<212> ADN
<213> Staphylococcus aureus

<400> 223
ctattctcag aatgacttgg 20

<210> 224
<211> 27
<212> ADN
<213> Mycoplasma gallisepticum

<220>
<221> base_modificada
<222> (1)..(2)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (5)..(5)
<223> a, c, t, g, desconocida u otra

<400> 224
nnganggagt cacagaaaag catctta 27

<210> 225
<211> 20
<212> ADN
<213> Mycoplasma gallisepticum

<400> 225
gtttctgggt gagcaaaaac 20

<210> 226
<211> 27
<212> ADN
<213> Campylobacter lari

<220>
<221> base_modificada
<222> (1)..(2)
<223> a, c, t, g, desconocida u otra

<220>
<221> base_modificada
<222> (6)..(6)
<223> a, c, t, g, desconocida u otra

<400> 226
nngggnagaa ggcaaaatgc cgcaaaa 27

<210> 227
<211> 20
<212> ADN
<213> Campylobacter lari

<400> 227
ggtattatcc cgtattgacg 20

<210> 228
<211> 20
<212> ADN
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción de Desconocido: oligonucleótido de SpCas9"

<400> 228
gggactcaac caagtcattc 20

REIVINDICACIONES

1. Una composición que se produce de forma no natural o modificada, que comprende un sistema de vectores que comprende uno o más vectores que comprenden

I. un primer elemento regulador enlazado operativamente a un ARN guía de un sistema CRISPR-Cas, que comprende :

(a) un polinucleótido de ARN guía capaz de hibridarse con una secuencia diana en una célula eucariótica,

(b) un polinucleótido de ARN de apareamiento tracr, y

(c) un polinucleótido de ARN tracr, y

II. un segundo elemento regulador enlazado operativamente a una secuencia codificante de proteínas que codifica una proteína Cas9 que comprende al menos una o más secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3',

en donde los componentes I y II están ubicados en los mismos o en diferentes vectores del sistema, en donde, cuando se transcriben, el polinucleótido de ARN de apareamiento tracr se hibrida con el polinucleótido de ARN tracr y el polinucleótido de ARN guía dirige la unión específica para la secuencia de un complejo CRISPR a la secuencia diana, y

en donde el complejo CRISPR comprende la proteína Cas9 complejada con (1) el polinucleótido de ARN guía que se hibrida con la secuencia diana, y (2) el polinucleótido de ARN de apareamiento tracr que se hibrida al polinucleótido de ARN tracr,

en donde la proteína Cas9 es Cas9 de *Staphylococcus aureus* (SaCas9).

2. Una composición de acuerdo con la reivindicación 1, en donde la proteína Cas9 tiene una o más mutaciones en cualquier dominio de la proteína.

3. Una composición de acuerdo con la reivindicación 2, en donde la proteína Cas9 tiene una o más mutaciones en un dominio catalítico.

4. Una composición de acuerdo con la reivindicación 2 o 3, en donde la una o más mutaciones están en un dominio RuvCI, RuvCII, RuvCIII o HNH de la proteína Cas9.

5. Una composición de acuerdo con la reivindicación 3 o 4, en donde la proteína Cas9 comprende una o más mutaciones correspondientes a la numeración de la posición de SpCas9 en D10A, H840A, N854A o N863A.

6. Una composición de acuerdo con cualquier reivindicación precedente, en donde la proteína Cas9 comprende, además, un dominio funcional.

7. Una composición de acuerdo con la reivindicación 6, en donde el dominio funcional es un dominio activador transcripcional.

8. Una composición de acuerdo con la reivindicación 7, en donde el dominio activador transcripcional es VP64.

9. Una composición de acuerdo con la reivindicación 6, en donde el dominio funcional es un dominio represor transcripcional.

10. Una composición de acuerdo con la reivindicación 9, en donde el dominio represor transcripcional es KRAB, SID o SID4X.

11. Una composición de acuerdo con la reivindicación 6, en donde el dominio funcional es una recombinasa, una transposasa, un remodelador de histona, una metiltransferasa de ADN, un criptocromo y un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente.

12. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia de polinucleótidos que codifica la proteína Cas9 está optimizada en codones para la expresión en una célula eucariótica.

13. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que comprende dos ARN guía.

14. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que comprende un sistema de vectores que comprende uno o más vectores, y en donde la secuencia de polinucleótidos que comprende o codifica dichos componentes I y II están situadas en el mismo o en diferentes vectores del sistema.

15. Una composición de acuerdo con la reivindicación 14, en donde el sistema de vectores comprende un único vector.

16. Una composición de acuerdo con la reivindicación 14, en donde el uno o más vectores comprenden uno o más vectores virales.

17. Una composición de acuerdo con la reivindicación 16, en donde el uno o más vectores virales comprenden uno o más vectores de retrovirus, lentivirus, adenovirus, virus adeno-asociados o virus herpes simplex.

18. Una composición de acuerdo con una cualquiera de las reivindicaciones 1 a 15, que comprende una nanopartícula, liposoma, exosoma, sistema de levadura o microvesícula.

19. Una célula huésped o una línea celular *ex vivo* o *in vitro* que comprende la composición de acuerdo con una cualquiera de las reivindicaciones precedentes o progenie de las mismas.

20. Una célula huésped, línea celular o progenie de la misma *ex vivo* o *in vivo* de acuerdo con la reivindicación 19, que es una célula madre o una línea celular madre.

21. Un método de modificar un organismo mediante manipulación de una o más secuencias dianas en loci genómicos de interés, que comprende suministrar al organismo la composición de acuerdo con una cualquiera de las reivindicaciones 1-18, en donde el organismo es un organismo no animal.

22. Un método *ex vivo* de modificar una célula o un organismo mediante manipulación de una o más secuencias dianas en loci genómicos de interés, que comprende suministrar a la célula una composición de acuerdo con una cualquiera de las reivindicaciones 1-20, en el que el método no comprende un proceso para modificar la identidad genética de la línea germinal de un ser humano.

23. Un método de acuerdo con la reivindicación 21 o 22, en el que el organismo es una planta o alga.

24. Uso de la composición de acuerdo con una cualquiera de las reivindicaciones 1-18 para la edición de un gen o genoma *ex vivo*, en donde el uso no comprende un proceso para modificar la identidad genética de la línea germinal de un ser humano.

25. Uso de una composición de acuerdo con una cualquiera de las reivindicaciones 1-18, para la preparación de un medicamento.

26. Una composición de acuerdo con una cualquiera de las reivindicaciones 1-18, para uso en corregir defectos oculares que surgen de mutaciones genéticas.

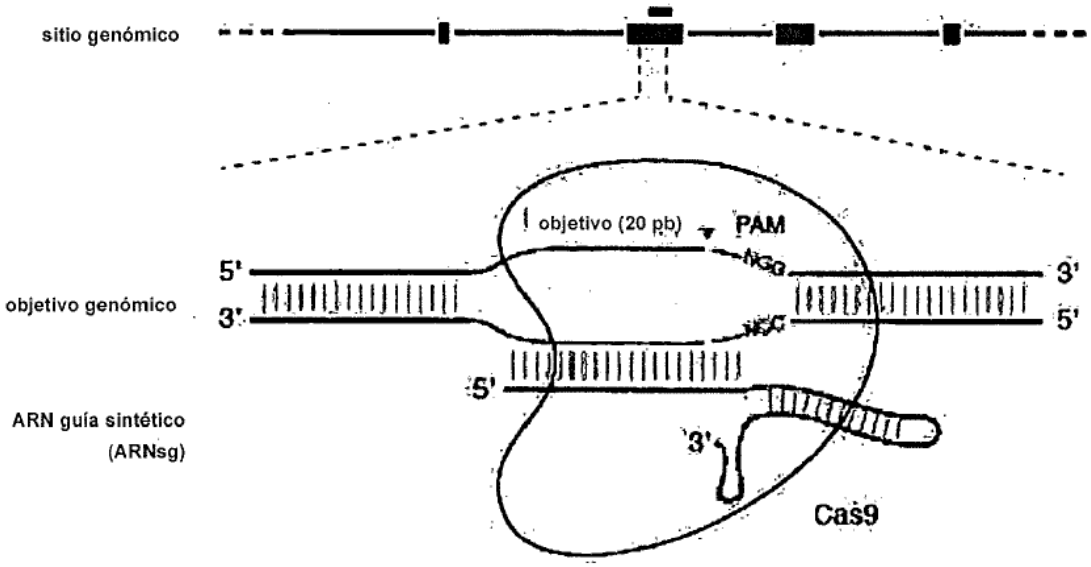

sitio genómico
objetivo (20 pb)
PAM
NGG
5'
3'
objetivo genómico
3'
5'
NCC
5'
ARN guía sintético
(ARNsg)
3'
Cas9

FIG. 1

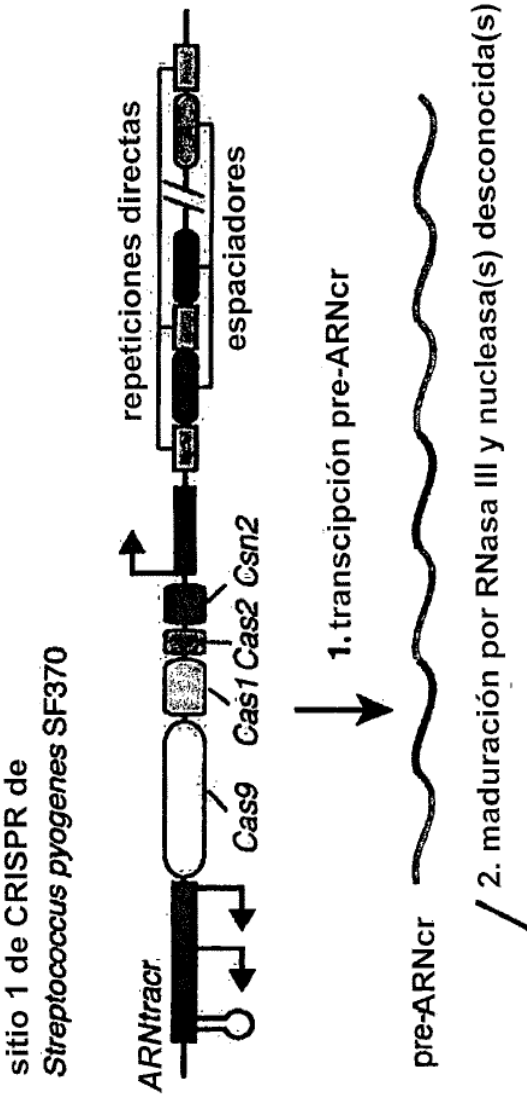

A
sitio 1 de CRISPR de
Streptococcus pyogenes SF370
repeticiones directas
espaciadores
ARNtracr
Cas9
Cas1
Cas2
Csn2
1. transcipción pre-ARNcr
pre-ARNcr
2. maduración por RNasa III y nucleasa(s) desconocida(s)
protoespaciador
PAM
4. RDC mediada por Cas9
ARNcr maduro
3. reconocim. objetivo
Cas9
ARNtrac tratado

FIG. 2A

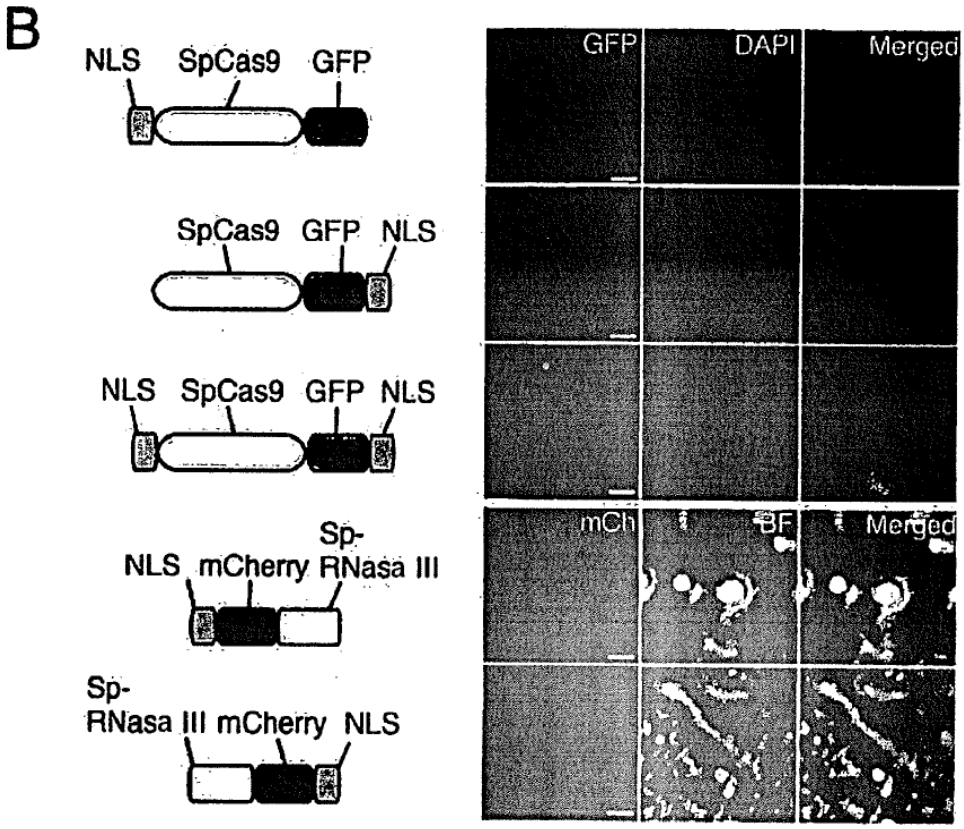

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
NLS mCherry Sp-RNasa III
Sp-RNasa III mCherry NLS
GFP
DAPI
Merged
mCh
BF
Merged

FIG. 2B

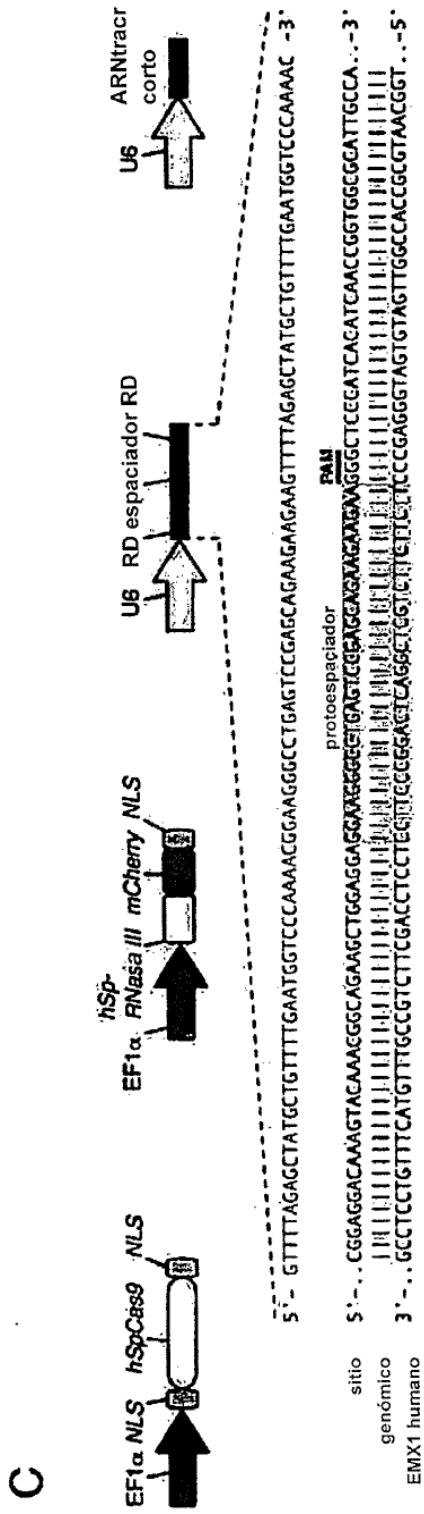

C
EF1α NLS
hSpCas9
NLS
EF1α
hSp-RNasa III
mCherry
NLS
U6
RD espaciador RD
U6
ARNtracr corto
5'- GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC -3'
protoespaciador
PAM
sitio genómico EMX1 humano
5'-...CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGAAGGAGGAGAAGTCCGAGTCCGAGCAGAAGAAGAAGGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA...-3'
3'-...GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCTTCCTCCTCAGGCTCAGGCTCGTCTTCTTCCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT...-5'

FIG. 2C

D

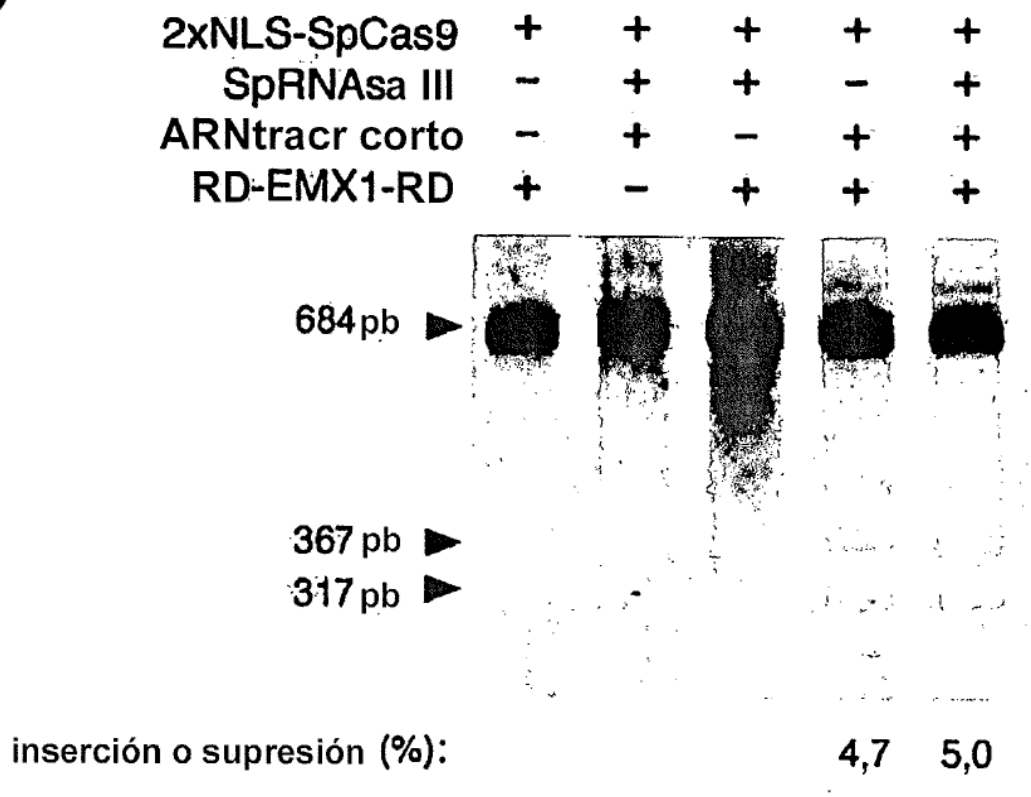

2xNLS-SpCas9 + + + + +
SpRNAsa III − + + − +
ARNtracr corto − + − + +
RD-EMX1-RD + − + + +
684 pb
367 pb
317 pb
inserción o supresión (%):
4,7
5,0

FIG. 2D

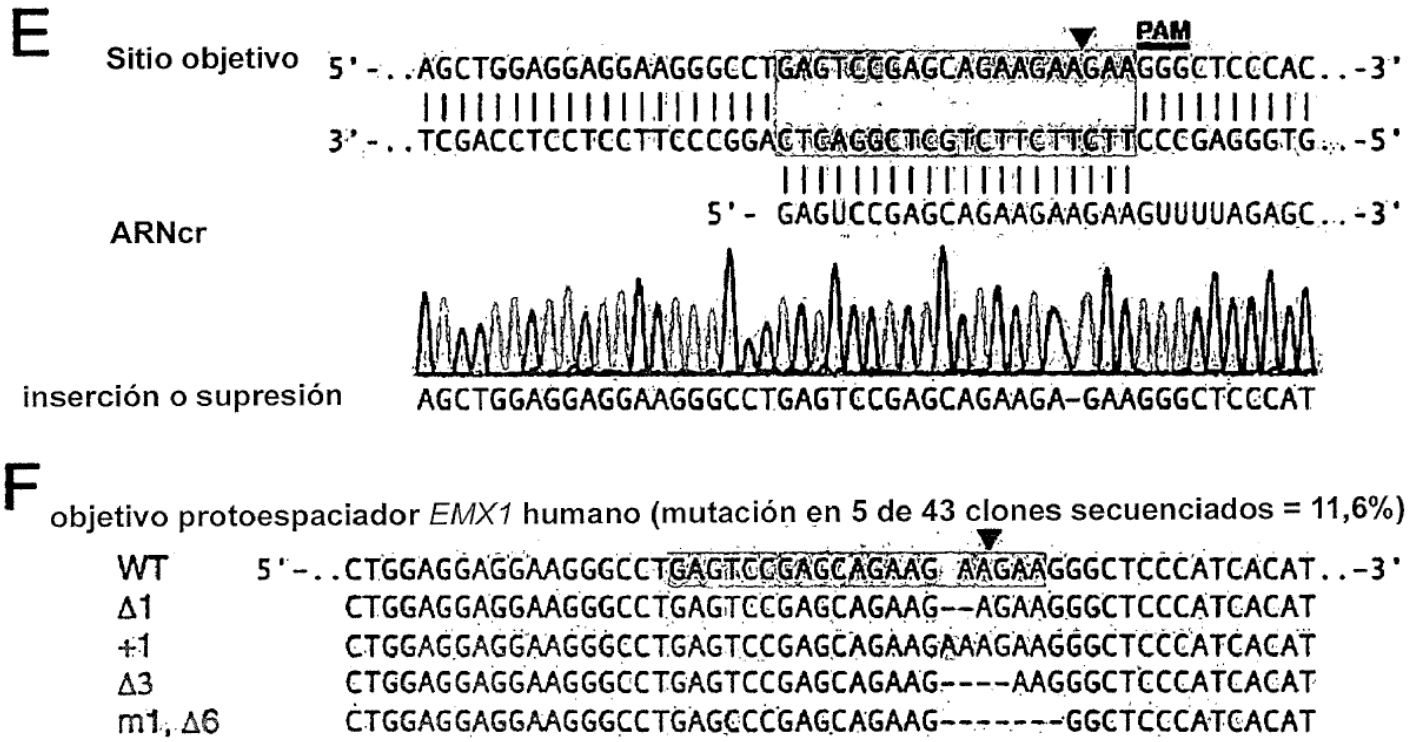
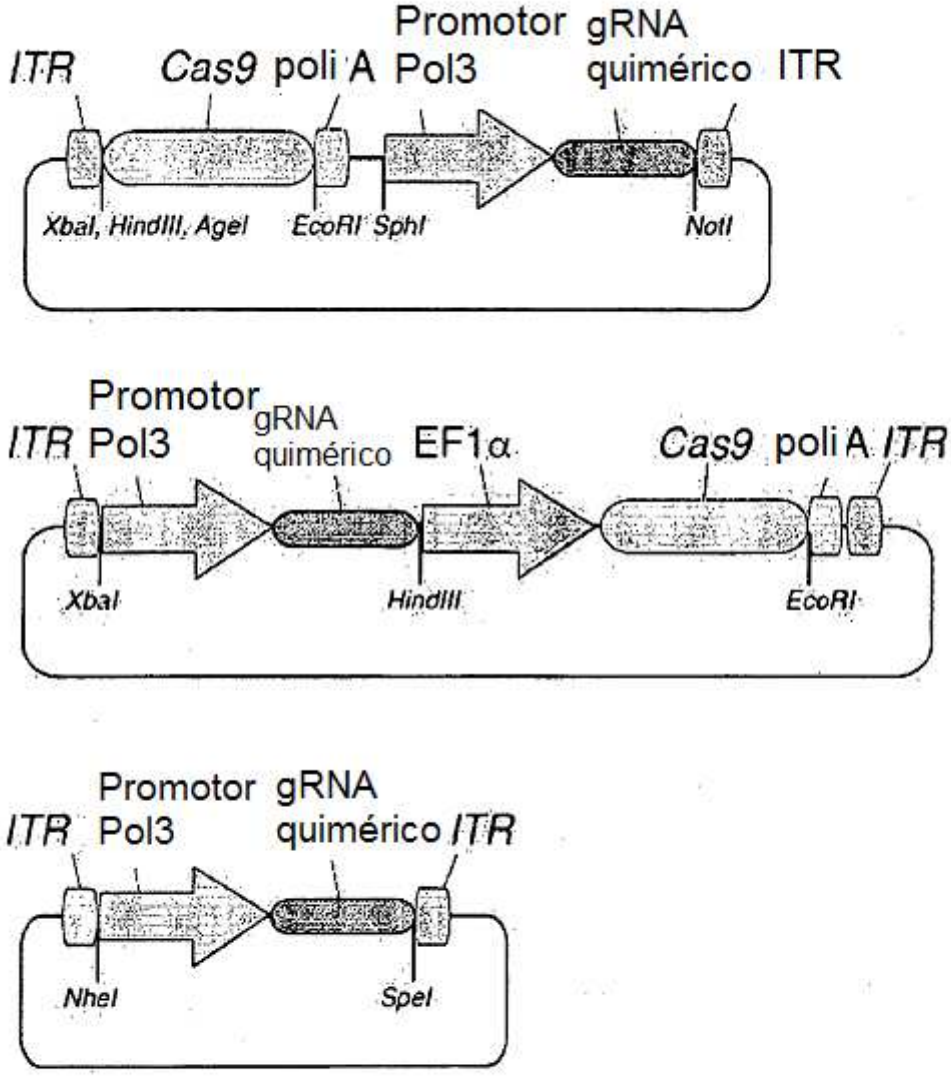

E
Sitio objetivo
PAM
5'-..AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAC..-3'
3'-..TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTG..-5'
ARNcr
5'- GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
inserción o supresión
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA-GAAGGGCTCCCAT
F
objetivo protoespaciador EMX1 humano (mutación en 5 de 43 clones secuenciados = 11,6%)
WT 5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT..-3'
Δ1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG--AGAAGGGCTCCCATCACAT
+1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAAGAAGGGCTCCCATCACAT
Δ3 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG----AAGGGCTCCCATCACAT
m1, Δ6 CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG-------GGCTCCCATCACAT

FIG. 2E-F



- dependiendo del tamaño de la construcción completa, puede contener una o más casetes de ARN guía (gRNA quiméricas Pol3

FIG.3

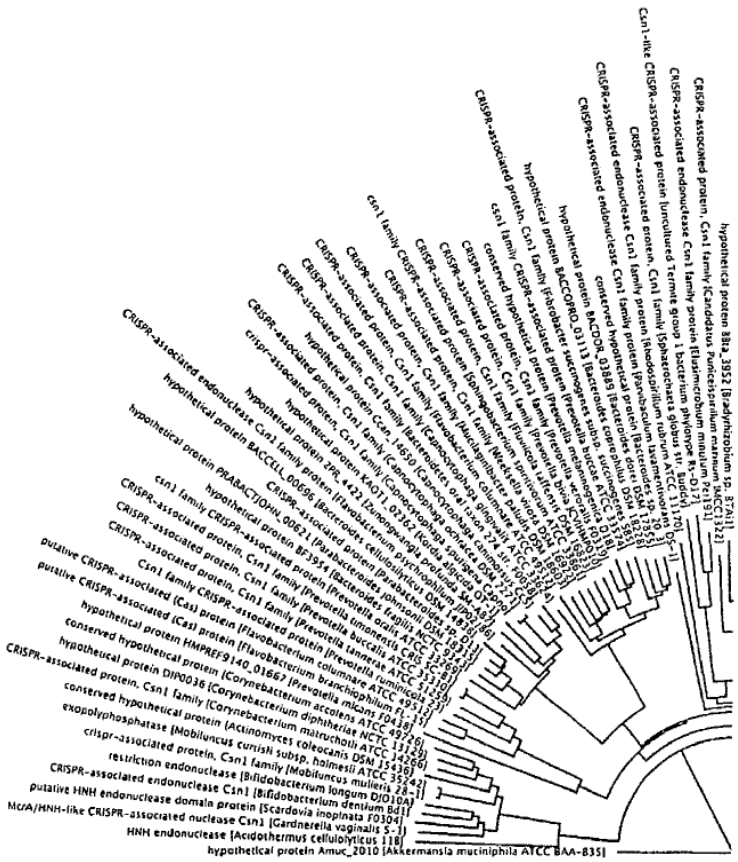


FIG. 4A

hypothetical protein = proteína hipotética; HNH endonuclease = endonucleasa HNH; Mcra/HNH-like CRISPR-associated nuclease Csn1 = Csn1 de endonucleasa asociada a CRISPR de tipo Mcra/HNH; putative HNH endonuclease domain protein = proteína de dominio de la endonucleasa HNH putativa; CRISPR-associated endonuclease Csn1 = Csn1 de endonucleasa asociada a CRISPR; Restriction endonuclease = endonucleasa de restricción; exopolyphosphatase = exopolifosfatasa; conserved hypothetical protein = proteína hipotética conservada; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; putative CRISPR-associated (Cas) Protein = Proteína (Cas) asociada a CRISPR putativa;

hypothetical protein Nham_4054 [Nitrobacter hamburgensis X14]
CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
hypothetical protein [Dinoroseobacter shibae DFL 12]
conserved hypothetical protein SU517_3938 [Sphingomonas sp. S17]
hypothetical protein HDN1F_34120 [gamma proteobacterium HdN1]
CRISPR-associated protein, Csn1 family [Ralstonia syzygii R24]
CRISPR-associated endonuclease Csn1 family [Nitrosomonas sp. AL212]
CRISPR-associated protein, Csn1 family [Verminephrobacter eiseniae EF01-2]
hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]
CRISPR-associated protein, Csn1 family [Acidovorax avenae subsp. avenae ATCC 19860]
CRISPR-associated protein [Helicobacter cinaedi CCUG 18818]
CRISPR-associated protein, Csn1 family [Alicycliphilus denitrificans BC]
CRISPR-associated protein C8J_1455 [Campylobacter jejuni subsp. jejuni 81116]
CRISPR-associated protein, Csn1 family [Helicobacter mustelae 12198]
hypothetical protein CJJ81176_1423 [Campylobacter jejuni subsp. jejuni 81116]
CRISPR-associated protein, Csn1 family [Alicycliphilus denitrificans K601]
CRISPR-associated protein CJE_1455 [Campylobacter jejuni subsp. doylei 269.97]
CRISPR-associated protein, Csn1 family [Campylobacter curvus 525.92]
CRISPR-associated protein CJE_2123 [Campylobacter jejuni subsp. jejuni 81116]
CRISPR-associated protein RPC_4489 [Rhodopseudomonas palustris BisB18]
CRISPR-associated protein, Csn1 family [Glaciecola sp. HTCC2999

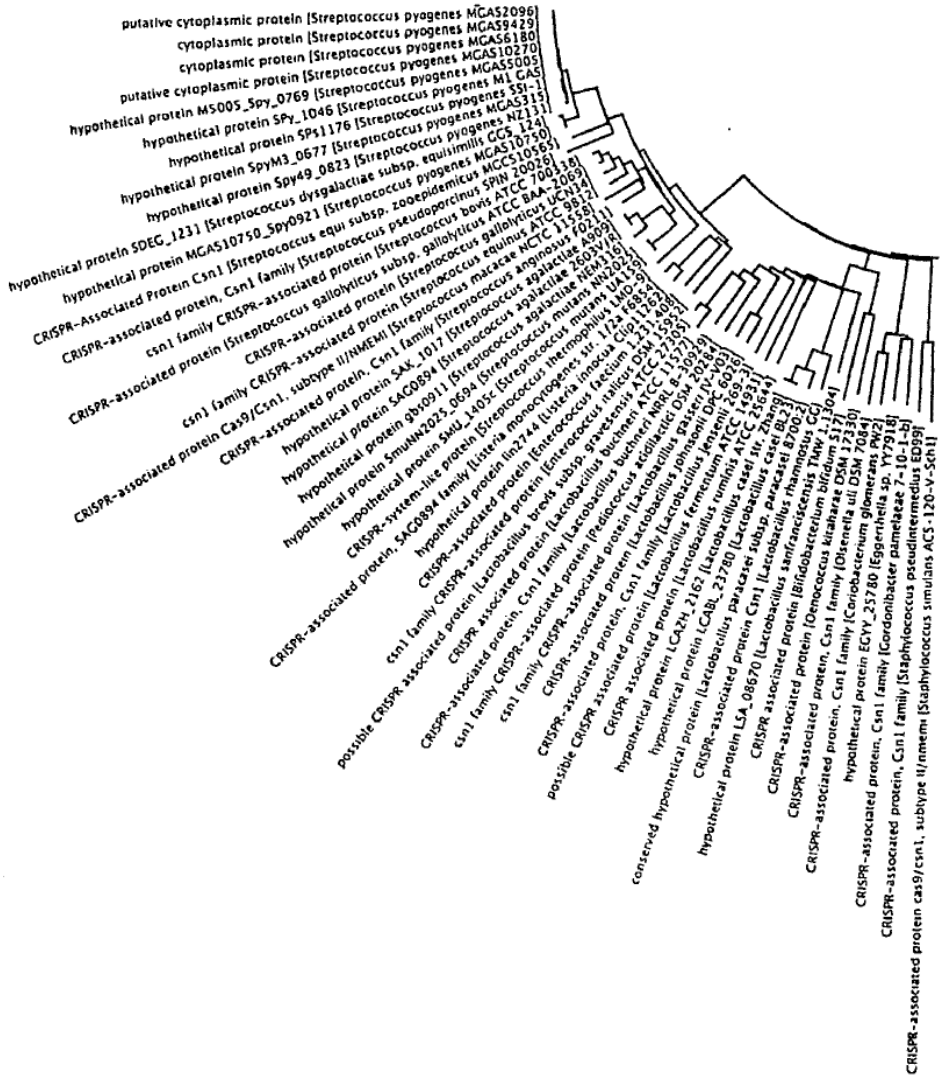


FIG. 4C

Putative cytoplasmic protein = proteína citoplasmática putativa; cytoplasmic protein = proteína citoplasmática; hypothetical protein = proteína hipotética; CRISPR-associated Protein Csn1 = Csn1 Proteína asociada a CRISPR; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; possible CRISPR-associated protein = Proteína asociada a CRISPR posible; conserved hypothetical protein = proteína hipotética conservada;

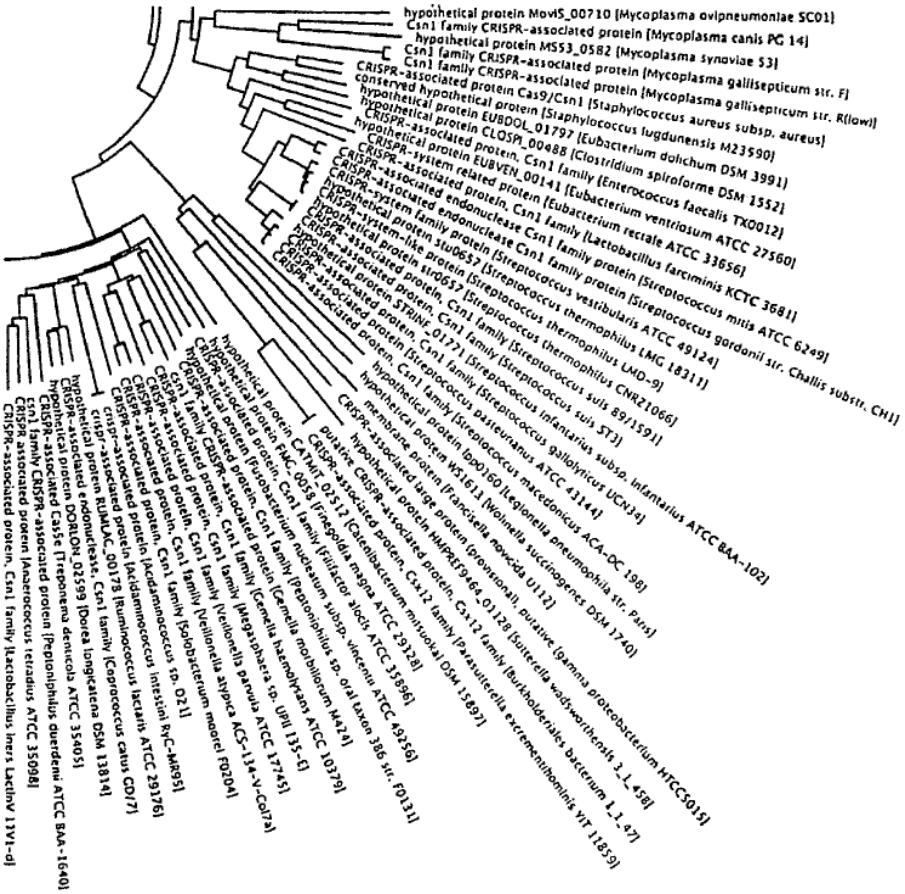


FIG. 4D

Putative CRISPR-associated Protein Csx12 family = familia Csx12 de proteína asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; conserved hypothetical protein = proteína hipotética conservada; CRISPR-system related protein = proteína relacionada con el sistema CRISPR; CRISPR-system like protein = proteína de tipo sistema CRISPR; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; membrane protein = proteína de membrana; CRISPR-associated large protein (provisional), putative = proteína grande asociada a CRISPR (provisional) putativa; CRISPR-associated CasSe = CasSe asociada a CRISPR.

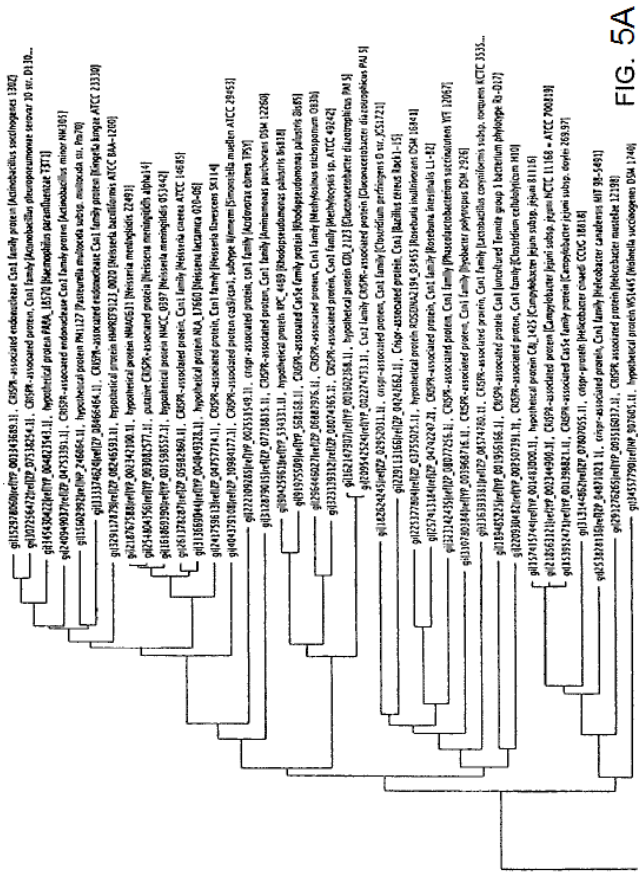


FIG. 5A

Putative CRISPR-associated Protein = proteína asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; CRISPR-associated CasSe family protein = proteína de la familia CasSe asociada a CRISPR

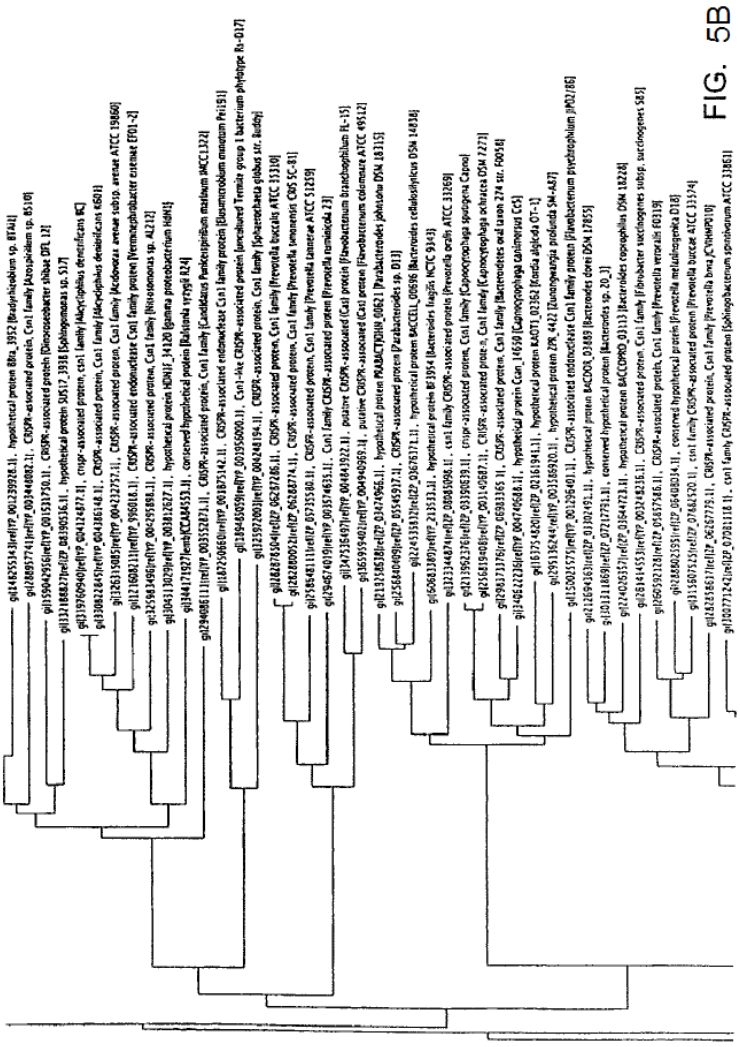


FIG. 5B

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

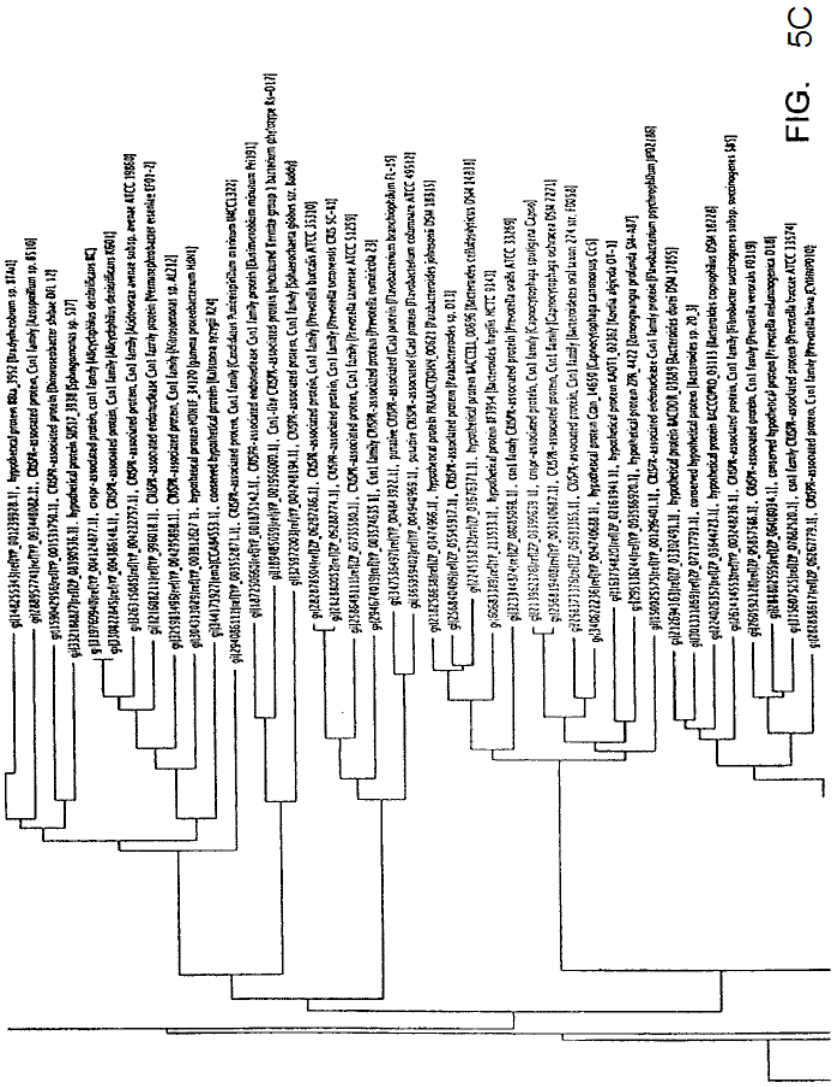


FIG. 5C

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

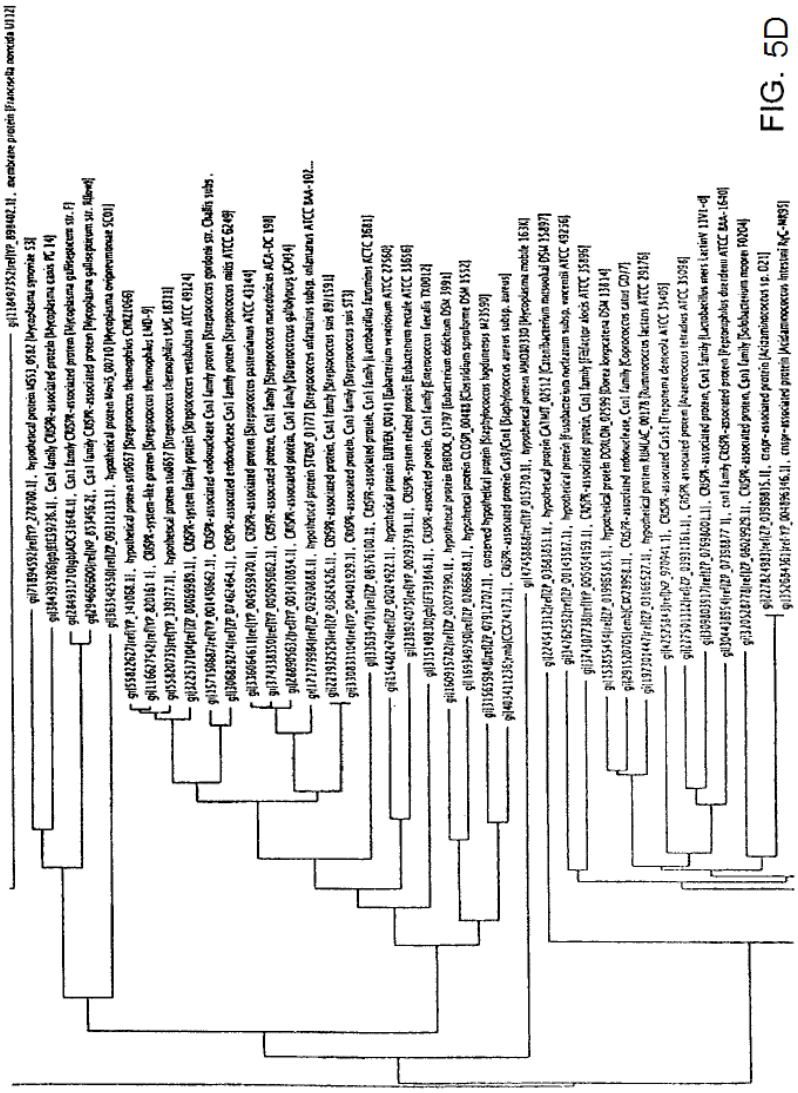


FIG. 5D

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1; CRISPR-system-like protein = Proteína de tipo sistema CRISPR; CRISPR-associated protein Cas9/Csn1= proteína Cas9/Csn1 asociada a CRISPR

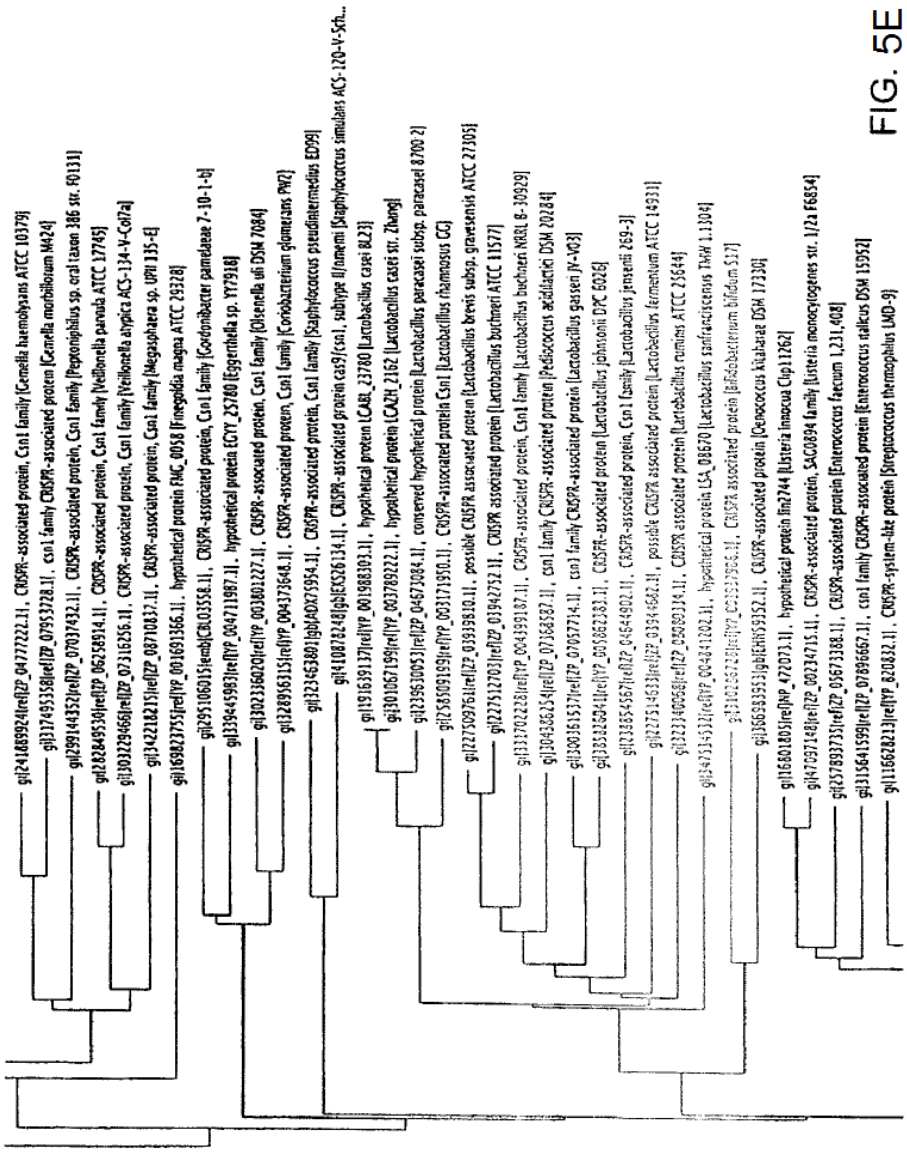


FIG. 5E

Hypothetical protein = proteína hipotética; CRISPR-associated Protein Csn1 family = familia Csn1 Proteína asociada a CRISPR; possible CRISPR-associated protein = Proteína asociada a CRISPR posible; conserved hypothetical protein = proteína hipotética conservada.

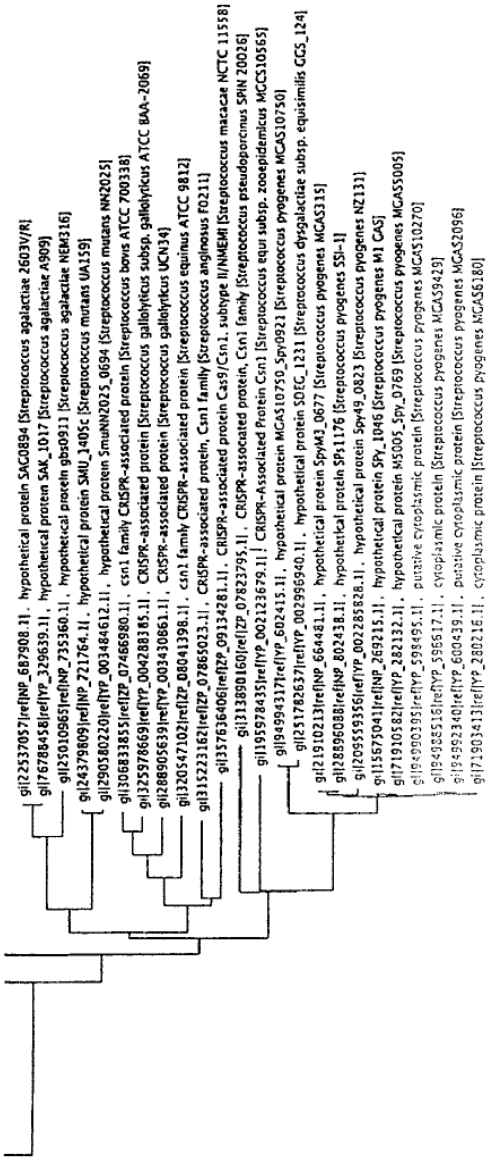


FIG. 5F

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated protein Cas9/Csn1, subtype II= proteína Cas9/Csn1 asociada a CRISPR, subtipo II; Putative cytoplasmic protein = proteína citoplasmática putativa; cytoplasmic protein = proteína citoplasmática

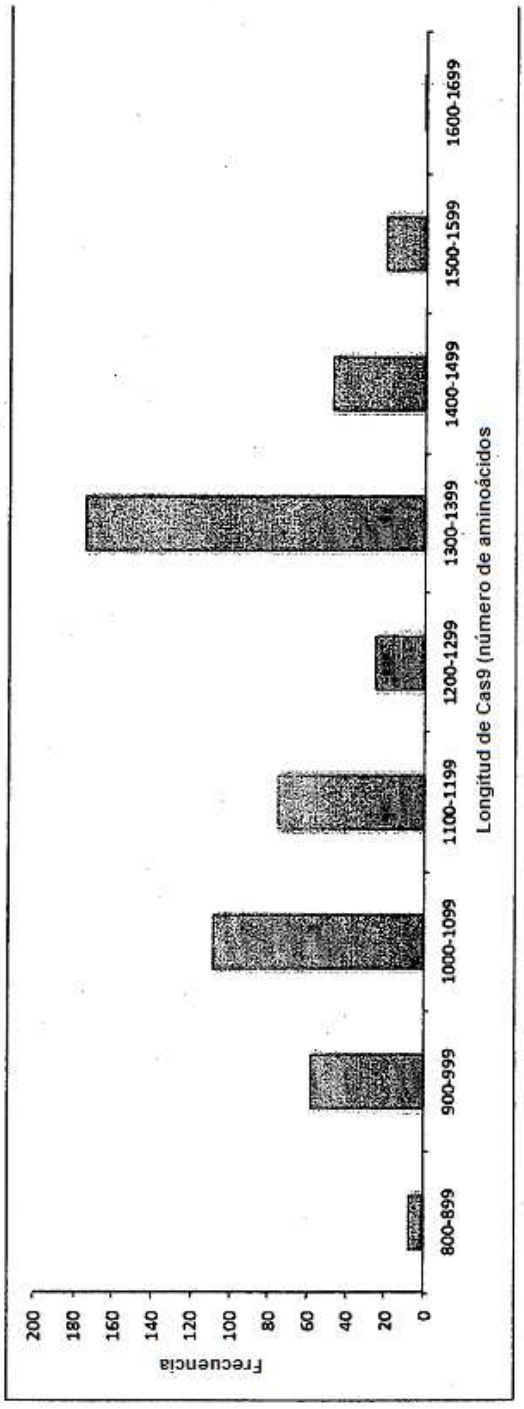

Frecuencia
200
180
160
140
120
100
80
60
40
20
0
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Longitud de Cas9 (número de aminoácidos

FIG. 6

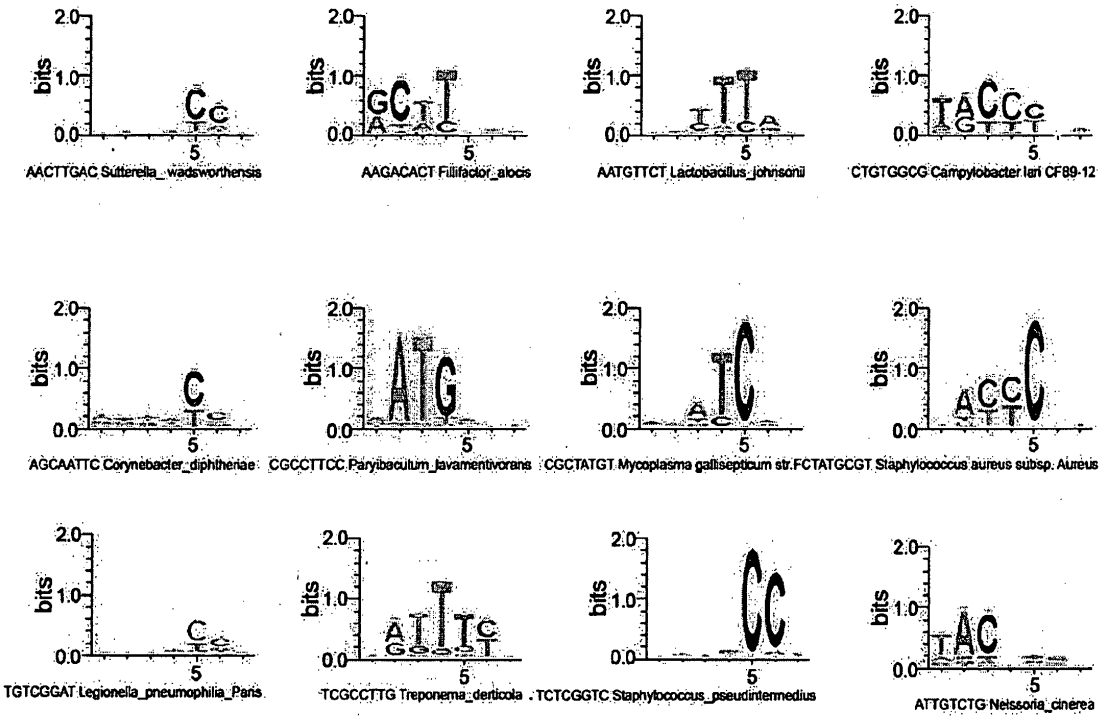
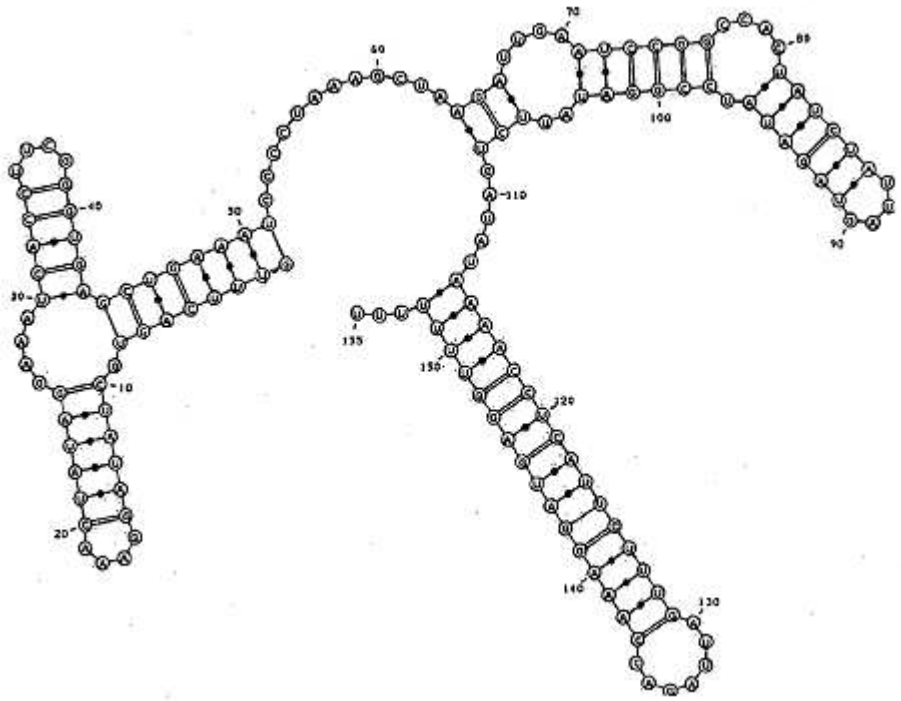
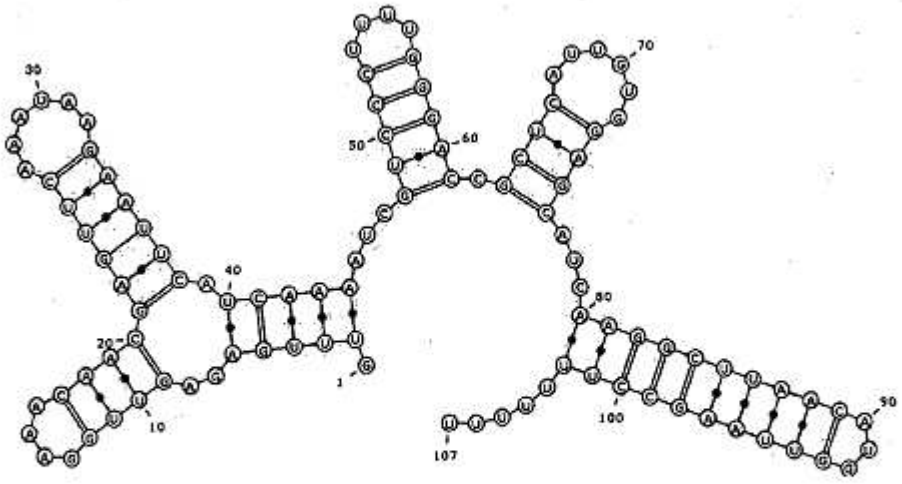

AACTTGAC Sutterella_wadsworthensis
AAGACACT Fillifactor_alocis
AATGTTCT Lactobacillus_johnsonii
CTGTGGCG Campylobacter lari CF89-12
AGCAATTC Corynebacter_diphtheriae
CGCCTTCC Paryibaculum_lavamentivorans
CGCTATGT Mycoplasma gallisepticum str.FCTATGCGT Staphylococcus aureus subsp. Aureus
TGTCGGAT Legionella_pneumophilia_Paris
TCGCCTTG Treponema_denticola
TCTCGGTC Staphylococcus_pseudintermedius
ATTGTCTG Neisseria_cinerea
bits
2.0
1.0
0.0
5

FIG. 7

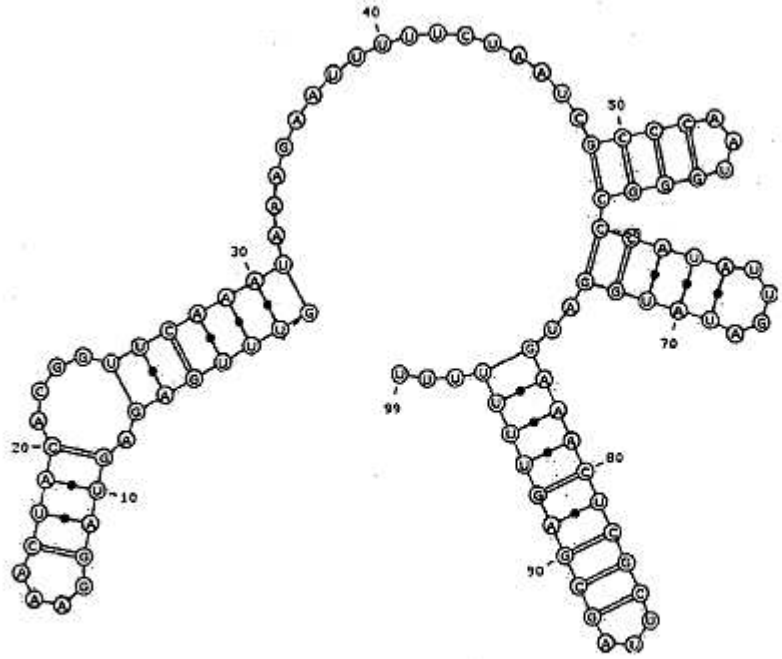


ARN guía quimérico para homólogo 2 de Cas9

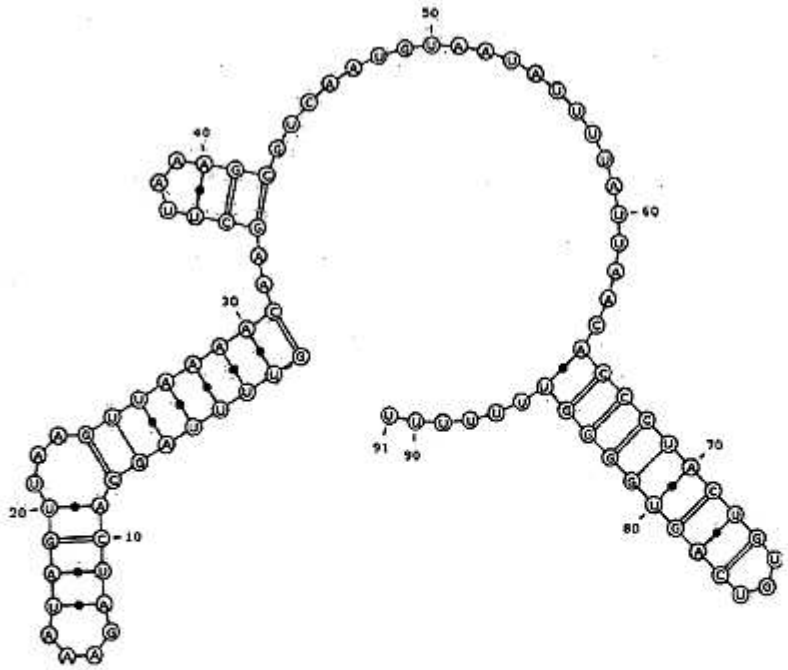


ARN guía quimérico para homólogo 4 de Cas9

FIG. 8A

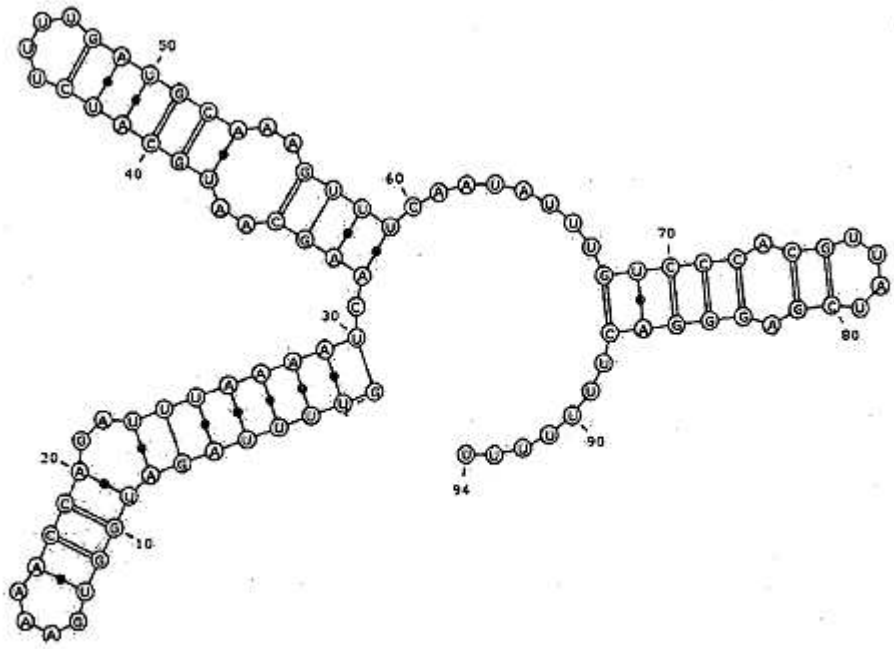


ARN guía quimérico para homólogo 5 de Cas9

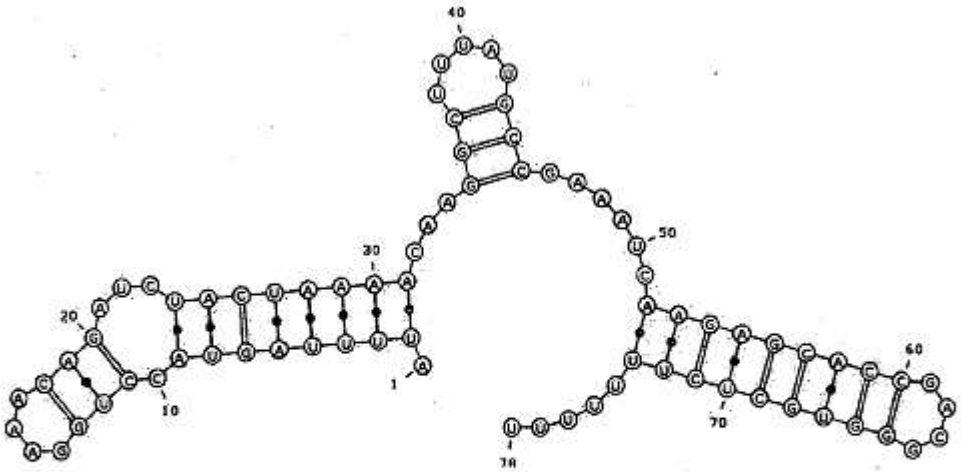


ARN guía quimérico para homólogo 6 de Cas9

FIG. 8B

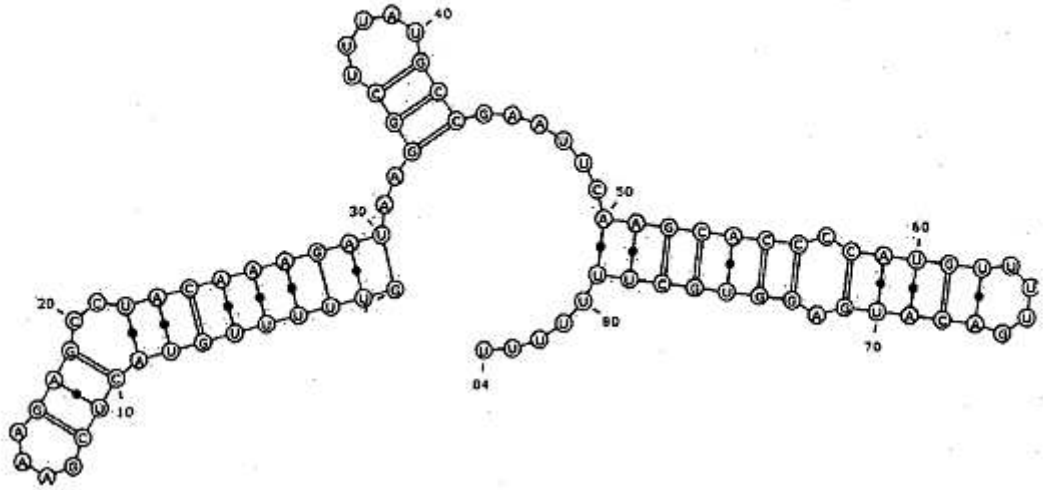


ARN guía quimérico para homólogo 7 de Cas9

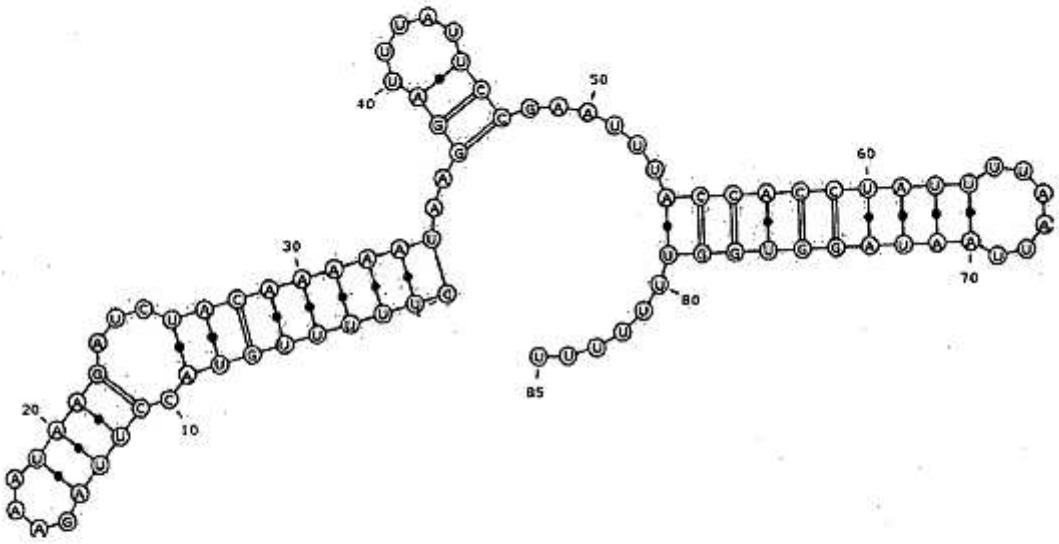


ARN guía quimérico para homólogo 8 de Cas9

FIG. 8C

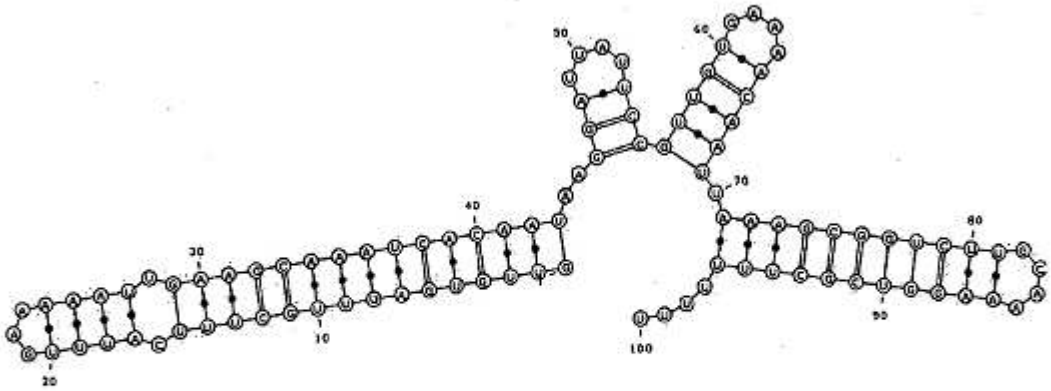


ARN guía quimérico para homólogo 9 de Cas9

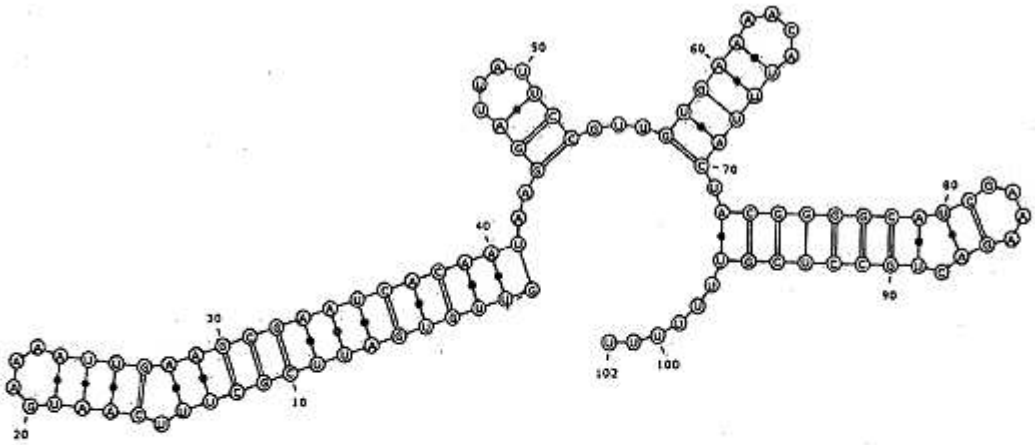


ARN guía quimérico para homólogo 10 de Cas9

FIG. 8D

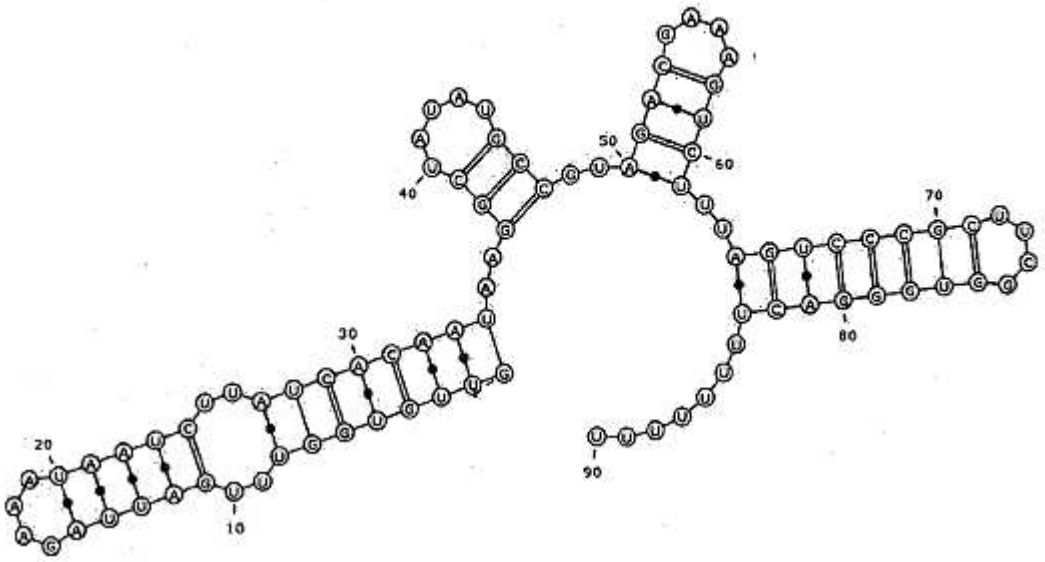


ARN guía quimérico para homólogo 12 de Cas9



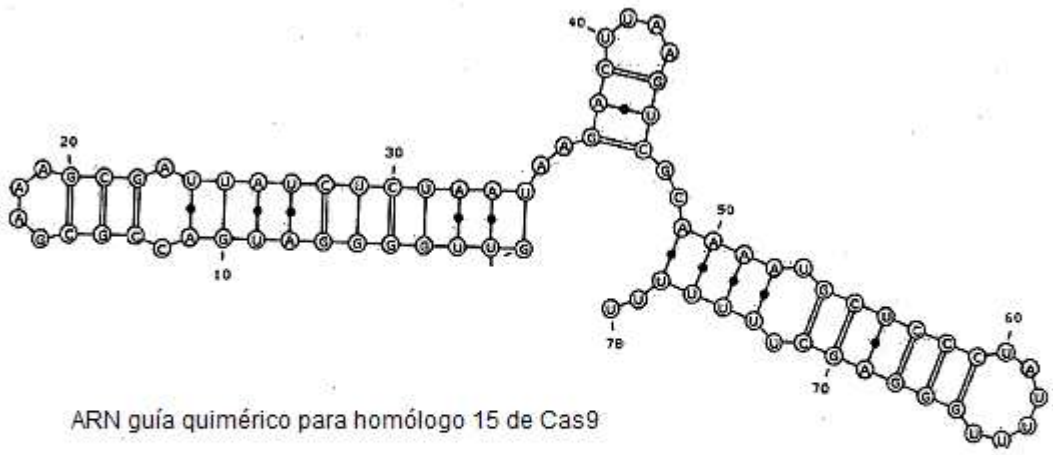
ARN guía quimérico para homólogo 13 de Cas9

FIG. 8E

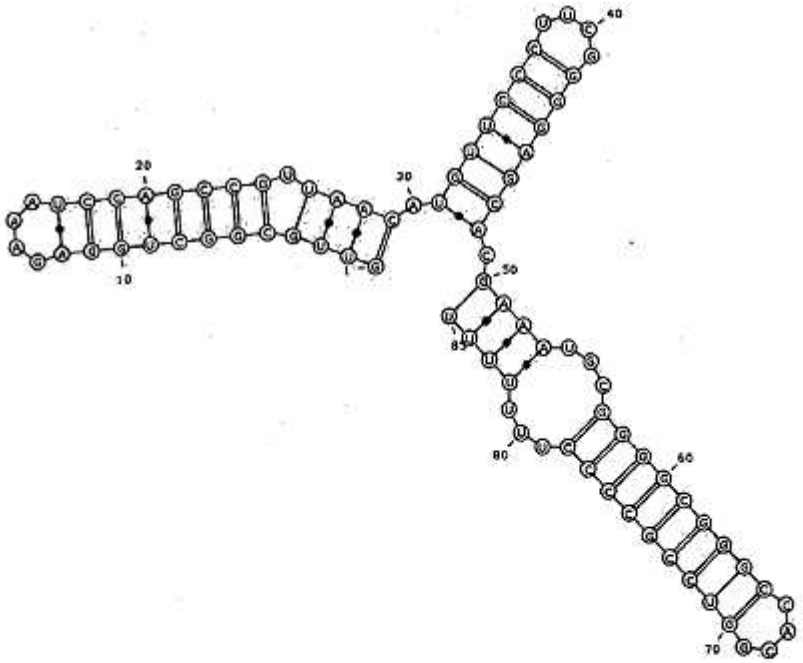


ARN guía quimérico para homólogo 14 de Cas9

FIG. 8F

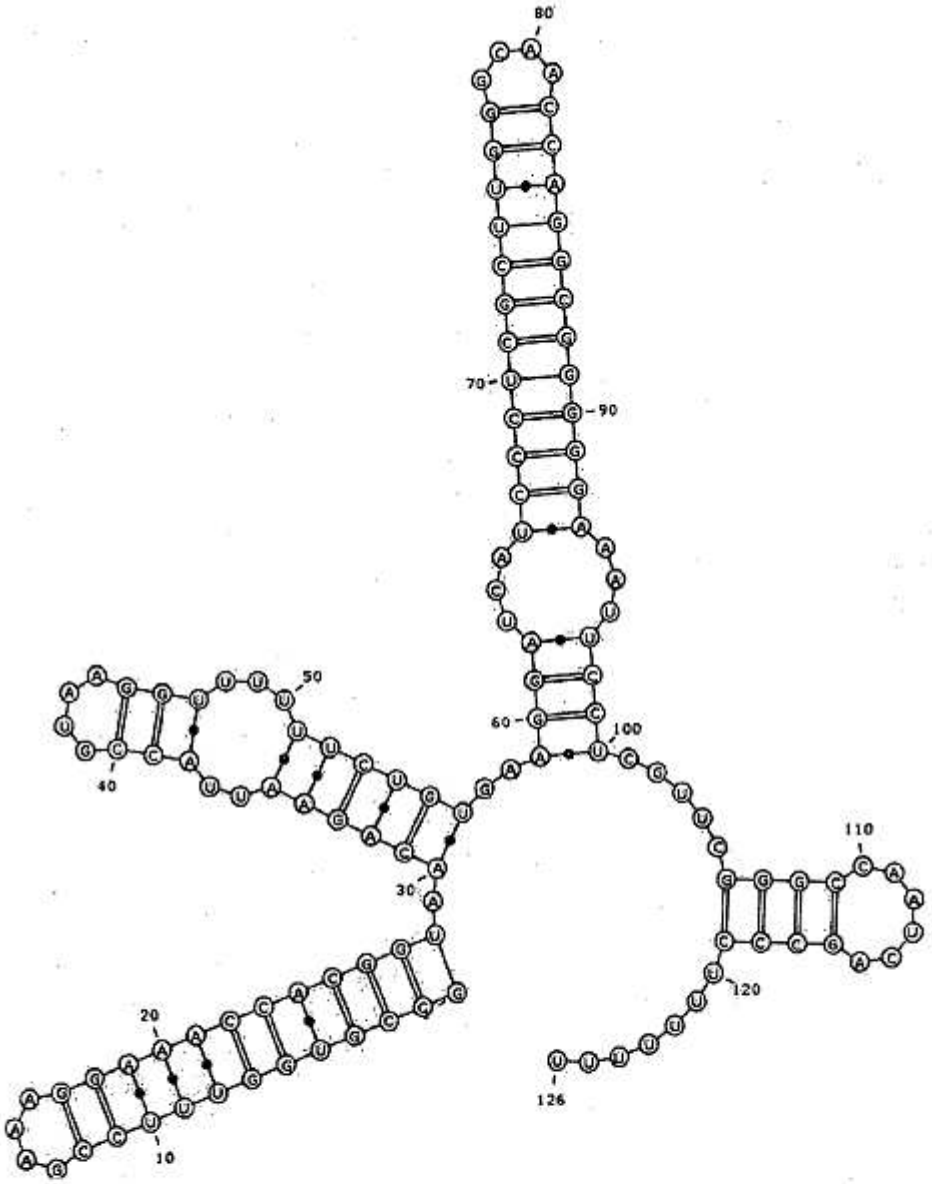


ARN guía quimérico para homólogo 15 de Cas9



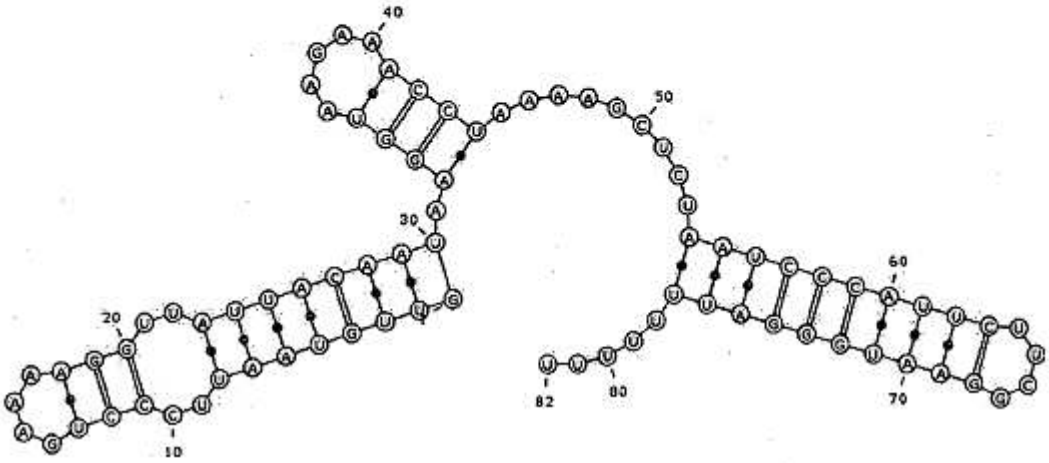
ARN guía quimérico para homólogo 16 de Cas9

FIG. 8G



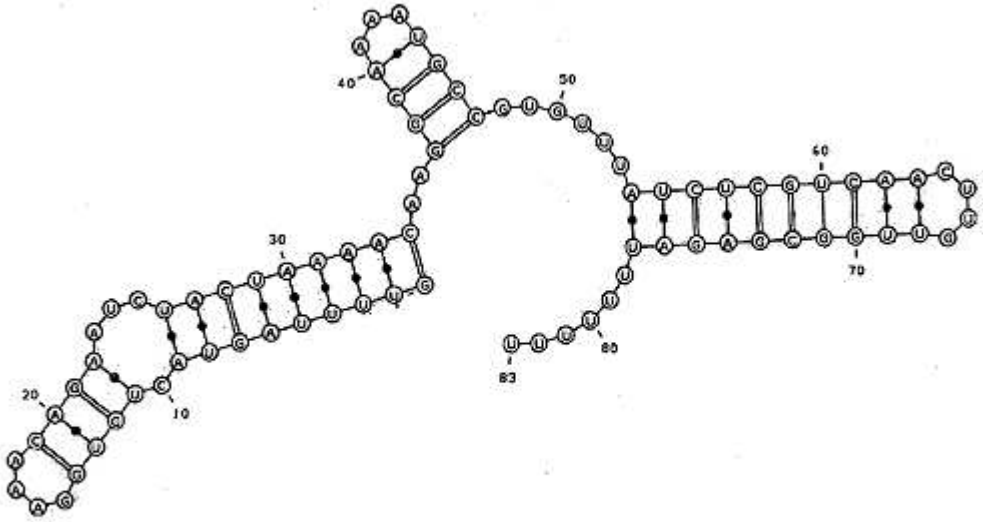
ARN guía quimérico para homólgo 17 de Cas9

FIG. 8H

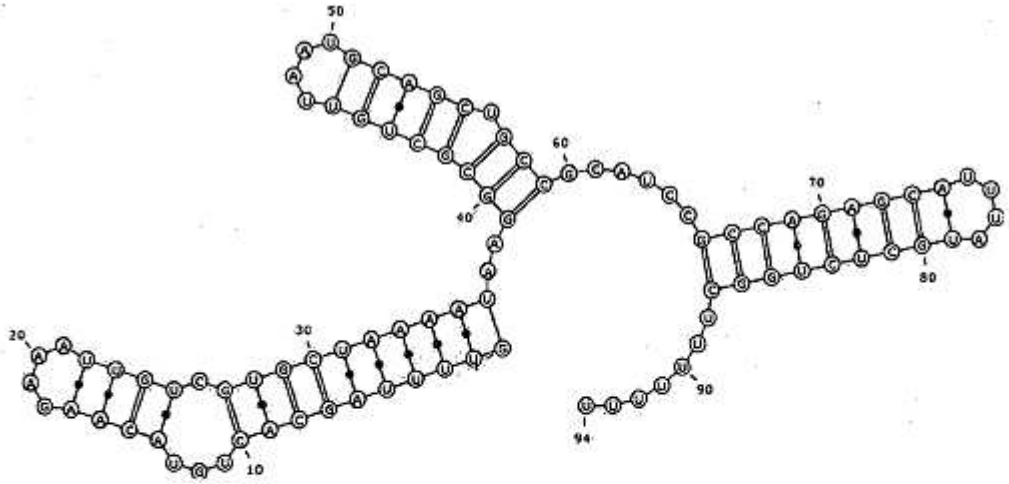


ARN guía quimérico para homólogo 19 de Cas9

FIG. 8I



ARN guía quimérico para homólogo 21 de Cas9



ARN guía quimérico para homólogo 23 de Cas9

FIG. 8J

FIG. 9A

secuencias Cas9 optimizadas en el codón humano

#2

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACTCAG
AGCGAGCGACGATTTTCTTGCAGCATTGGCATTGACATGGGGGCTAAGTACACTGGGGTGTTCTACGCACTGTTCGACCGGGAG
GAACTGCCCACAAACCTGAACAGCAAGGCCATGACCCTGGTCATGCCTGAGACAGGGCCAAGATACGTGCAGGCACAGAGAACT
GCCGTCAGACACAGGCTGCGCGGACAGAAGAGATATACCCTGGCTAGGAAACTGGCATTTCTGGTGGTCGACGATATGATCAAG
AAACAGGAAAAGAGGCTGACTGATGAGGAATGGAAACGAGGACGGGAGGCCCTGTCCGGCCTGCTGAAGCGGAGAGGGTACTCT
CGGCCCAACGCTGACGGCGAAGATCTGACCCCTCTGGAGAATGTGAGAGCAGACGTGTTCGCCGCTCATCCTGCCTTCAGCACA
TATTTTTCCGAAGTGCGCTCTCTGGCTGAGCAGTGGGAGGAGTTCACCGCAAACATCAGCAATGTCGAGAAGTTTCTGGGCGAC
CCAAACATCCCCGCCGATAAAGAGTTCATTGAATTTGCCGTGGCTGAAGGGCTGATTGACAAGACCGAGAAGAAAGCCTACCAG
TCAGCTCTGAGCACCCTGAGGGCAAACGCCAATGTGCTGACAGGACTGCGGCAGATGGGCCACAAGCCTAGATCAGAATATTTT
AAAGCAATCGAGGCCGACCTGAAGAAAGATAGCCGCCTGGCCAAGATTAACGAAGCATTCGGAGGAGCAGAGCGCCTGGCTCGA
CTGCTGGGAAACCTGTCCAATCTGCAGCTGCGGGCAGAAAGATGGTACTTCAATGCCCCCGACATCATGAAGGATAGGGGCTGG
GAGCCTGATCGCTTCAAGAAAACACTGGTGCGGGCTTTTAAGTTCTTTCACCCAGCAAAGGACCAGAACAAACAGCATCTGGAA
CTGATCAAACAGATTGAGAACAGCGAAGATATCATTGAGACTCTGTGCACCCTGGACCCAAACAGAACCATCCCCCCTTACGAG
GATCAGAACAATAGGCGCCCACCCCTGGACCAGACTCTGCTGCTGAGTCCCGAAAAGCTGACCCGGCAGTATGGCGAGATCTGG
AAAACATGGAGCGCCAGACTGACCTCCGCTGAACCCACACTGGCACCTGCAGCCGAGATTCTGGAAAGATCTACCGACAGGAAG
AGTCGCGTGGCAGTCAACGGACACGAGCCACTGCCTACACTGGCTTACCAGCTGAGTTATGCACTGCAGAGAGCCTTCGACAGG
TCAAAAGCCCTGGATCCATATGCTCTGAGGGCACTGGCTGCAGGCTCAAAAAGCAATAAGCTGACATCCGCCCGCACTGCTCTG
GAGAACTGCATCGGAGGCCAGAATGTGAAAACCTTCCTGGACTGTGCCCGACGGTACTATCGGGAAGCAGACGATGCCAAAGTC
GGGCTGTGGTTCGACAACGCCGATGGACTGCTGGAGAGATCTGACCTGCATCCTCCAATGAAGAAAAAGATCCTGCCCCTGCTG
GTGGCCAATATTCTGCAGACAGATGAAACCACAGGCCAGAAGTTTCTGGACGAGATCTGGCGAAAACAGATTAAGGGGCGGGAA
ACTGTGGCTAGCCGATGTGCACGGATCGAGACAGTGCGGAAATCCTTCGGGGGAGGCTTTAACATTGCCTACAATACCGCTCAG
TATAGGGAGGTGAACAAGCTGCCCCGCAATGCCCAGGATAAAGAACTGCTGACAATCAGAGATAGGGTGGCTGAGACTGCAGAC
TTCATTGCCGCTAACCTGGGGCTGTCTGACGAGCAGAAAAGAAAGTTCGCCAATCCTTTTAGTCTGGCTCAGTTCTACACCCTG
ATCGAGACAGAAGTGTCCGGATTTTCTGCAACTACCCTGGCCGTCCACCTGGAGAACGCCTGGAGGATGACAATCAAGGATGCT
GTGATTAATGGGGAAACTGTCAGAGCAGCACAGTGCAGCAGGCTGCCTGCAGAGACAGCTCGCCCATTCGATGGACTGGTGAGA
AGGCTGGTCGACAGACAGGCTTGGGAGATCGCAAAGAGGGTGTCAACTGACATTCAGAGCAAAGTCGATTTCTCCAACGGCATC
GTGGACGTCAGCATTTTTGTGGAGGAAAATAAGTTCGAGTTTTCCGCATCTGTGGCCGATCTGAAAAAGAACAAACGGGTCAAA
GACAAGATGCTGTCCGAGGCCGAAAAGCTGGAAACCAGATGGCTGATCAAAAATGAGCGGATCAAGAAGGCCAGCCGGGGAACT
TGTCCCTACACCGGCGATAGGCTGGCTGAGGGGGGAGAAATCGACCACATTCTGCCCCGAAGCCTGATCAAGGATGCCCGGGGA
ATTGTGTTTAACGCTGAGCCTAATCTGATCTATGCAAGCTCCCGCGGCAACCAGCTGAAAAAGAATCAGCGATACAGTCTGTCA
GATCTGAAGGCCAACTATCGGAATGAGATCTTCAAAACTAGCAACATCGCTGCAATTACCGCCGAGATTGAGGACGTGGTCACT
AAGCTGCAGCAGACCCATAGACTGAAATTCTTTGATCTGCTGAATGAGCACGAACAGGACTGCGTGCGGCACGCCCTGTTCCTG
GACGATGGCAGCGAAGCTCGCGACGCAGTGCTGGAGCTGCTGGCAACACAGCGCCGAACTCGCGTCAACGGGACACAGATCTGG
ATGATTAAGAACCTGGCCAACAAGATCCGAGAGGAACTGCAGAATTGGTGTAAGACAACTAACAATAGACTGCACTTTCAGGCC
GCTGCAACTAACGTGTCCGATGCAAAGAATCTGAGGCTGAAACTGGCCCAGAACCAGCCCGACTTCGAGAAGCCAGATATCCAG
CCCATTGCCAGCCATTCCATCGACGCCCTGTGCTCTTTCGCTGTGGGGAGTGCTGACGCAGAACGCGATCAGAATGGATTTGAC
TACCTGGATGGCAAGACCGTGCTGGGACTGTATCCACAGAGCTGTGAGGTCATTCACCTGCAGGCCAAGCCCCAGGAGGAAAAA
AGTCATTTCGATTCAGTGGCTATCTTTAAGGAAGGCATCTACGCAGAGCAGTTCCTGCCTATCTTTACCCTGAACGAAAAGATC
TGGATTGGATATGAGACACTGAATGCCAAAGGCGAAAGATGCGGGGCTATTGAGGTGAGTGGCAAACAGCCAAAGGAGCTGCTG
GAAATGCTGGCCCCCTTCTTTAACAAGCCTGTGGGCGACCTGTCAGCCCACGCTACTTACCGGATCCTGAAAAAGCCTGCATAT
GAGTTTCTGGCAAAGGCAGCTCTGCAGCCACTGAGCGCAGAGGAAAAAAGACTGGCAGCCCTGCTGGATGCTCTGCGCTACTGT
ACCAGTCGAAAGTCACTGATGAGCCTGTTCATGGCTGCAAACGGGAAATCCCTGAAAAAGCGGGAGGACGTGCTGAAACCCAAG
CTGTTCCAGCTGAAGGTCGAGCTGAAAGGCGAAAAGAGCTTCAAGCTGAACGGGAGCCTGACCCTGCCTGTGAAGCAGGACTGG
CTGAGAATCTGCGATAGCCCAGAACTGGCAGACGCCTTTGGCAAACCCTGTTCCGCCGATGAGCTGACATCTAAGCTGGCTCGC
ATTTGGAAACGACCTGTGATGCGGGATCTGGCTCATGCACCAGTCCGGAGAGAGTTCAGCCTGCCCGCAATCGACAACCCAAGT
GGAGGGTTCAGGATTAGGCGCACCAACCTGTTTGGCAATGAGCTGTACCAGGTGCACGCCATCAACGCTAAAAAGTATCGCGGC
TTCGCCTCCGCTGGGTCTAATGTCGACTGGTCCAAGGGGATCCTGTTTAACGAGCTGCAGCATGAAAATCTGACCGAGTGCGGA
GGCAGGTTCATTACAAGCGCCGATGTGACTCCTATGTCCGAATGGCGCAAGGTGGTCGCAGAGGACAACCTGTCTATCTGGATT
GCTCCAGGGACAGAAGGACGACGGTACGTGAGGGTCGAGACAACATTCATCCAGGCCAGTCACTGGTTTGAACAGTCAGTGGAG
AATTGGGCCATTACTAGTCCTCTGTCACTGCCAGCTTCCTTCAAGGTGGACAAACCAGCTGAGTTTCAGAAGGCAGTCGGAACC
GAGCTGTCAGAACTGCTGGGCCAGCCCAGGAGCGAAATCTTCATTGAGAACGTGGGCAATGCCAAGCATATCCGCTTTTGGTAC
ATTGTGGTGAGCAGCAACAAAAAGATGAACGAGTCTTACAACAATGTGTCTAAGAGTTAAGAATTC

#4

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAAA
GAAATCAAAGACTACTTCCTGGGGCTGGATGTGGGGACTGGGAGCGTGGGGTGGGCTGTGACCGATACTGACTACAAACTGCTG
AAGGCTAACCGAAAAGACCTGTGGGGCATGAGATGCTTCGAGACAGCCGAAACTGCTGAGGTGCGGAGACTGCACAGGGGAGCC
AGGCGCCGAATCGAGCGGAGAAAGAAACGCATTAAGCTGCTGCAGGAGCTGTTCTCTCAGGAAATCGCCAAAACCGATGAGGGC
TTCTTTCAGAGAATGAAGGAAAGCCCCTTTTACGCTGAGGACAAAACAATCCTGCAGGAAAACACTCTGTTCAATGACAAGGAT
TTTGCTGATAAGACTTACCACAAAGCATATCCTACCATTAATCATCTGATCAAGGCTTGGATTGAGAACAAGGTGAAACCAGAC
CCCCGACTGCTGTACCTGGCATGTCACAACATCATTAAGAAAAGGGGACATTTCCTGTTTGAAGGCGACTTCGATTCAGAGAAT
CAGTTTGATACCAGCATCCAGGCACTGTTCGAGTATCTGCGCGAGGACATGGAGGTGGACATCGATGCCGACAGCCAGAAGGTC

FIG. 9B

AAAGAGATTCTGAAGGATAGCTCCCTGAAGAACTCTGAAAAACAGAGTCGGCTGAATAAGATCCTGGGGCTGAAGCCTTCCGAC
AAACAGAAGAAAGCCATCACAAACCTGATTTCTGGAAACAAGATCAATTTCGCCGATCTGTACGACAATCCAGATCTGAAGGAC
GCTGAGAAAAACTCAATCAGCTTCTCCAAGGACGATTTTGATGCACTGAGTGACGATCTGGCCTCAATTCTGGGCGACAGCTTT
GAACTGCTGCTGAAGGCCAAAGCTGTCTATAACTGCTCTGTGCTGAGTAAGGTCATCGGGGACGAGCAGTACCTGAGCTTCGCC
AAGGTGAAAATCTACGAAAAGCACAAAACCGATCTGACAAAGCTGAAAAACGTGATCAAGAAACATTTCCCCAAGGACTACAAG
AAGGTCTTTGGATACAACAAGAACGAGAAAAACAACAACAATTACTCCGGCTATGTGGGAGTCTGTAAGACCAAGAGTAAGAAA
CTGATCATTAACAACTCAGTCAACCAGGAAGATTTCTACAAGTTTCTGAAAACTATCCTGTCAGCCAAGAGCGAGATCAAGGAA
GTGAATGACATCCTGACCGAGATTGAAACTGGCACCTTTCTGCCAAAGCAGATCTCTAAAAGTAACGCAGAGATTCCCTATCAG
CTGAGGAAAATGGAGCTGGAAAAGATCCTGTCCAATGCCGAAAAGCACTTCTCTTTTCTGAAGCAGAAAGACGAAAAAGGACTG
TCACATAGCGAGAAGATCATTATGCTGCTGACCTTCAAGATCCCTTACTATATTGGCCCAATCAACGATAATCACAAGAAATTC
TTTCCCGACAGATGCTGGGTGGTCAAGAAAGAGAAATCCCCTTCTGGCAAGACCACACCATGGAACTTCTTTGATCATATCGAC
AAGGAAAAAACAGCAGAGGCCTTCATTACTTCTAGGACCAATTTTTGCACTTACCTGGTGGGAGAGAGCGTCCTGCCTAAGTCT
AGTCTGCTGTACTCCGAATATACCGTGCTGAACGAGATCAACAATCTGCAGATCATTATCGATGGCAAGAATATTTGTGACATC
AAGCTGAAACAGAAGATCTACGAGGACCTGTTCAAGAAGTACAAGAAAATTACCCAGAAGCAGATCAGCACCTTCATCAAGCAC
GAAGGCATCTGCAACAAAACCGATGAAGTGATCATCCTGGGGATTGACAAGGAATGTACATCAAGCCTGAAAAGCTACATCGAG
CTGAAAAACATTTTCGGCAAGCAGGTGGATGAGATCTCCACTAAGAATATGCTGGAGGAAATTATCAGATGGGCTACCATCTAC
GACGAGGGGGAAGGAAAGACCATCCTGAAAACAAAGATCAAGGCAGAATACGGAAAGTATTGCTCCGACGAGCAGATTAAGAAA
ATCCTGAACCTGAAGTTCTCCGGCTGGGGGCGACTGTCTCGGAAATTTCTGGAGACAGTGACTAGTGAAATGCCCGGCTTCTCA
GAACCTGTCAATATTATCACCGCCATGAGGGAGACACAGAACAATCTGATGGAGCTGCTGTCCTCTGAGTTCACCTTCACCGAG
AACATTAAGAAAATCAATTCTGGATTCGAAGATGCCGAGAAGCAGTTTAGTTACGACGGCCTGGTGAAACCACTGTTTCTGAGT
CCCTCAGTCAAGAAAATGCTGTGGCAGACCCTGAAGCTGGTGAAAGAGATTAGCCATATCACACAGGCCCCCCCTAAGAAAATT
TTCATCGAAATGGCAAAGGGGGCCGAGCTGGAACCTGCTCGGACTAAGACCAGACTGAAAATCCTGCAGGATCTGTATAACAAT
TGTAAGAACGATGCTGACGCCTTCAGCTCAGAGATCAAAGACCTGAGCGGAAAGATTGAGAACGAAGATAATCTGAGGCTGCGC
TCCGACAAGCTGTACCTGTACTATACTCAGCTGGGGAAATGCATGTATTGTGGAAAGCCAATTGAGATCGGCCACGTGTTCGAT
ACCTCAAACTACGATATTGACCATATCTATCCCCAGAGCAAGATCAAAGACGATAGCATTTCCAATCGGGTGCTGGTCTGCAGC
TCCTGTAACAAGAACAAGGAGGACAAGTACCCACTGAAATCAGAGATCCAGAGCAAGCAGCGCGGCTTCTGGAACTTTCTGCAG
CGAAACAATTTCATTTCTCTGGAGAAGCTGAATAGACTGACAAGGGCCACTCCAATCAGTGACGATGAGACAGCCAAGTTTATT
GCTAGGCAGCTGGTGGAAACTCGCCAGGCTACCAAGGTGGCCGCTAAAGTCCTGGAAAAGATGTTCCCCGAGACAAAAATCGTG
TACAGCAAGGCCGAGACTGTCTCCATGTTCCGGAACAAGTTTGATATCGTGAAGTGCAGAGAAATTAACGATTTTCACCATGCT
CACGACGCATACCTGAATATCGTGGTCGGCAACGTGTATAATACCAAGTTCACAAACAATCCTTGGAACTTTATCAAGGAGAAA
AGAGATAATCCAAAGATTGCTGACACCTACAACTACTATAAGGTGTTCGATTATGACGTCAAAAGGAACAATATCACAGCATGG
GAGAAGGGGAAAACTATTATCACCGTGAAAGACATGCTGAAGAGAAACACACCAATCTACACTAGGCAGGCAGCCTGTAAGAAA
GGGGAGCTGTTCAATCAGACCATTATGAAGAAAGGACTGGGCCAGCACCCCCTGAAGAAAGAAGGACCTTTTTCCAATATCTCT
AAATACGGCGGGTATAACAAGGTGAGCGCTGCATACTATACACTGATTGAGTATGAGGAAAAGGGCAACAAAATCCGCTCCCTG
GAAACTATTCCCCTGTACCTGGTGAAAGATATCCAGAAGGATCAGGACGTCCTGAAGTCTTATCTGACAGACCTGCTGGGGAAG
AAAGAGTTCAAGATCCTGGTGCCCAAGATCAAGATCAACAGCCTGCTGAAGATCAATGGGTTTCCTTGCCATATTACAGGAAAA
ACTAACGATAGTTTCCTGCTGCGCCCTGCCGTGCAGTTTTGCTGTTCAAACAATGAGGTCCTGTACTTCAAGAAAATTATCCGG
TTTTCCGAAATCCGCTCTCAGCGAGAGAAGATCGGGAAAACAATTAGCCCATACGAGGACCTGAGCTTCCGGTCATATATCAAG
GAGAACCTGTGGAAGAAAACTAAGAACGATGAAATCGGAGAGAAGGAATTTTACGACCTGCTGCAGAAGAAAAACCTGGAGATC
TATGATATGCTGCTGACTAAGCACAAAGACACCATCTACAAGAAACGCCCTAATTCTGCCACCATTGATATCCTGGTGAAGGGG
AAAGAGAAGTTCAAAAGCCTGATTATCGAAAACCAGTTTGAAGTGATCCTGGAGATCCTGAAGCTGTTTTCTGCAACACGGAAT
GTCAGTGACCTGCAGCATATCGGAGGCAGCAAGTACTCCGGCGTGGCCAAAATCGGGAACAAGATCTCTAGTCTGGATAACTGT
ATCCTGATCTATCAGTCCATCACCGGCATCTTCGAGAAACGGATCGACCTGCTGAAGGTGTAAGAATTC

#5

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACCAAG
GAGTATTACCTGGGGCTGGATGTGGGGACCAATTCCGTGGGATGGGCAGTGACCGATTCTCAGTACAACCTGTGCAAGTTTAAG
AAAAAGGATATGTGGGGCATCCGGCTGTTCGAAAGCGCCAACACAGCAAAGGACCGGAGACTGCAGAGAGGGAATAGGCGCCGA
CTGGAGCGGAAAAAGCAGAGAATTGATCTGCTGCAGGAAATCTTCTCCCCAGAGATCTGCAAGATTGACCCCACTTTCTTTATC
CGACTGAACGAATCCCGGCTGCACCTGGAGGACAAGTCTAACGATTTCAAATACCCACTGTTTATTGAGAAGGACTATTCTGAT
ATCGAGTACTATAAAGAGTTCCCCACCATTTTTCACCTGAGGAAGCATCTGATCGAGAGTGAGGAAAAACAGGATATCCGGCTG
ATCTACCTGGCCCTGCACAACATCATTAAGACCCGAGGACATTTTCTGATTGACGGCGATCTGCAGAGCGCCAAGCAGCTGAGG
CCCATCCTGGATACATTCCTGCTGTCCCTGCAGGAGGAACAGAACCTGTCAGTGAGCCTGTCCGAAAATCAGAAGGACGAGTAT
GAGGAAATTCTGAAAAACCGCAGCATCGCCAAGTCCGAAAAAGTGAAAAAGCTGAAGAATCTGTTTGAGATCTCAGACGAGCTG
GAAAAAGAGGAGAAGAAGGCCCAGAGCGCCGTGATCGAGAACTTCTGCAAGTTTATCGTGGGAAATAAGGGCGATGTCTGTAAA
TTCCTGCGGGTGTCTAAGGAGGAACTGGAGATTGACTCTTTCAGTTTTTCAGAGGGCAAGTACGAGGACGACATCGTGAAAAAC
CTGGAGGAAAAAGTGCCTGAAAAGGTCTACCTGTTTGAGCAGATGAAGGCAATGTATGATTGGAATATTCTGGTCGACATCCTG
GAAACCGAGGAATACATCAGCTTCGCCAAAGTGAAGCAGTATGAGAAACACAAGACTAACCTGCGGCTGCTGAGAGACATCATT
CTGAAATACTGCACCAAGGATGAGTATAATCGGATGTTTAACGACGAGAAGGAAGCTGGCAGCTACACCGCATATGTGGGGAAA
CTGAAAAAGAACAACAAGAAGTACTGGATCGAGAAAAAGAGAAATCCCGAGGAGTTCTACAAATCCCTGGGCAAGCTGCTGGAT
AAAATTGAGCCTCTGAAGGAGGACCTGGAAGTGCTGACTATGATGATCGAGGAGTGTAAGAACCACACCCTGCTGCCAATTCAG
AAAAATAAGGACAACGGCGTGATCCCCCACCAGGTGCATGAGGTCGAACTGAAAAAGATCCTGGAAAATGCCAAAAAGTACTAT
TCCTTCCTGACCGAGACAGACAAGGATGGGTACTCAGTGGTCCAGAAAATCGAGAGCATTTTCAGGTTTCGCATCCCCTACTAT

FIG. 9C

GTGGGGCCTCTGAGTACCCGGCACCAGGAAAAGGGATCAAACGTGTGGATGGTCAGAAAACCTGGCAGGGAGGATCGCATCTAC
CCATGGAATATGGAGGAAATCATTGACTTTGAGAAGAGCAACGAAAATTTCATTACACGGATGACTAACAAATGTACATATCTG
ATCGGGGAAGATGTCCTGCCCAAGCATTCTCTGCTGTACAGTAAATATATGGTGCTGAATGAGCTGAACAATGTGAAGGTCAGA
GGAAAAAAGCTGCCTACATCTCTGAAACAGAAGGTGTTCGAGGACCTGTTTGAAAACAAATCCAAAGTGACTGGAAAGAATCTG
CTGGAGTACCTGCAGATCCAGGACAAAGATATCCAGATTGACGATCTGTCTGGCTTCGACAAGGACTTCAAGACCAGCCTGAAG
AGCTATCTGGACTTCAAAAAGCAGATTTTTGGGGAGGAAATCGAGAAGGAAAGCATTCAGAACATGATCGAAGATATCATTAAG
TGGATCACCATCTACGGCAATGACAAGGAGATGCTGAAACGAGTGATTCGGGCTAATTATAGCAACCAGCTGACAGAGGAACAG
ATGAAAAAGATCACTGGATTTCAGTACAGTGGCTGGGGGAACTTCTCAAAGATGTTTCTGAAAGGGATCAGCGGATCCGACGTG
AGCACCGGCGAAACATTCGACATCATTACCGCAATGTGGGAGACAGACAACAATCTGATGCAGATCCTGTCAAAAAAGTTCACC
TTTATGGACAACGTCGAGGACTTCAACAGCGGCAAGGTCGGGAAAATCGACAAGATTACTTACGATAGCACCGTGAAGGAAATG
TTCCTGTCCCCTGAGAACAAAAGGGCCGTCTGGCAGACCATTCAGGTGGCTGAGGAGATCAAGAAAGTGATGGGCTGCGAGCCA
AAAAAGATCTTTATTGAAATGGCACGGGGCGGGGAGAAGGTGAAAAAGAGGACAAAATCTCGCAAGGCCCAGCTGCTGGAGCTG
TACGCCGCTTGCGAGGAAGATTGTAGAGAACTGATCAAGGAGATTGAGGACCGGGACGAGAGGGACTTCAATAGCATGAAGCTG
TTTCTGTACTATACCCAGTTCGGGAAATGTATGTATTCCGGCGACGACATCGATATTAACGAGCTGATTCGCGGCAATTCTAAG
TGGGACCGAGATCACATCTACCCCCAGAGCAAAATTAAGGACGATTCCATCGATAACCTGGTGCTGGTCAATAAGACATATAAT
GCCAAAAAGTCCAATGAGCTGCTGTCTGAGGACATCCAGAAAAAGATGCATTCATTCTGGCTGAGCCTGCTGAACAAAAAGCTG
ATCACTAAAAGCAAGTACGACCGCCTGACTCGAAAGGGCGACTTTACCGATGAGGAACTGAGTGGGTTCATCGCTAGACAGCTG
GTGGAAACAAGGCAGTCAACTAAGGCAATCGCCGATATCTTCAAGCAGATCTACAGCTCCGAGGTGGTCTATGTGAAGAGCAGC
CTGGTGAGCGACTTCAGGAAAAAGCCACTGAACTACCTGAAGTCTCGGAGAGTCAATGATTACCACCATGCAAAAGACGCCTAT
CTGAACATTGTGGTCGGGAACGTGTACAACAAAAAGTTTACCAGTAATCCCATCCAGTGGATGAAAAAGAATCGCGATACAAAC
TATAGCCTGAACAAGGTGTTCGAACACGACGTGGTCATTAACGGAGAAGTGATCTGGGAAAAGTGCACATACCATGAGGACACT
AATACCTATGATGGAGGCACTCTGGACCGAATCCGGAAGATTGTGGAACGCGATAACATTCTGTACACCGAGTACGCTTATTGT
GAGAAGGGCGAACTGTTTAATGCAACCATCCAGAACAAAAATGGAAACTCCACAGTCTCTCTGAAAAAGGGCCTGGACGTGAAA
AAGTACGGGGGATACTTCAGCGCCAACACAAGTTACTTCTCACTGATCGAGTTTGAGGACAAGAAGGGGGATAGAGCAAGGCAC
ATCATTGGAGTGCCTATCTATATTGCAAACATGCTGGAGCATTCTCCAAGTGCCTTCCTGGAGTACTGCGAACAGAAGGGGTAT
CAGAATGTGCGGATTCTGGTCGAGAAAATCAAAAAGAACAGCCTGCTGATCATTAATGGATACCCTCTGCGCATTCGAGGCGAG
AACGAAGTGGATACTTCCTTTAAGAGGGCCATCCAGCTGAAGCTGGACCAGAAAAACTATGAGCTGGTCCGCAATATCGAGAAG
TTCCTGGAAAAATACGTGGAGAAAAAGGGAAACTATCCAATTGACGAGAATAGAGATCACATCACACATGAAAAGATGAACCAG
CTGTACGAGGTGCTGCTGTCCAAAATGAAAAAGTTCAACAAGAAGGGCATGGCCGACCCCTCTGATAGGATCGAAAAGAGTAAG
CCTAAATTCATCAAGCTGGAGGACCTGATCGATAAGATTAATGTGATCAACAAAATGCTGAACCTGCTGCGCTGTGACAATGAT
ACTAAGGCCGACCTGTCTCTGATTGAGCTGCCCAAAAACGCTGGGAGTTTCGTGGTCAAAAAGAATACCATCGGAAAGTCAAAA
ATCATCCTGGTGAATCAGAGCGTGACTGGACTGTACGAGAATAGACGGGAACTGTAAGAATTC

#6
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGGGGAGG
AAACCTTACATTCTGTCTCTGGATATTGGAACTGGGTCCGTCGGCTACGCTTGCATGGATAAAGGATTCAACGTGCTGAAGTAC
CACGACAAAGATGCCCTGGGAGTGTATCTGTTCGACGGCGCTCTGACTGCACAGGAGCGGAGACAGTTTAGGACCTCCAGGCGC
CGAAAGAACCGGAGAATCAAACGCCTGGGCCTGCTGCAGGAACTGCTGGCACCCCTGGTGCAGAACCCTAATTTCTACCAGTTT
CAGCGGCAGTTCGCCTGGAAGAACGACAATATGGATTTTAAGAACAAGAGCCTGTCTGAGGTGCTGAGCTTCCTGGGATATGAA
TCCAAGAAATACCCTACCATCTACCACCTGCAGGAGGCTCTGCTGCTGAAAGACGAGAAGTTTGATCCAGAACTGATCTACATG
GCACTGTATCATCTGGTGAAATACAGAGGCCACTTTCTGTTCGATCATCTGAAGATCGAGAACCTGACTAACAATGACAATATG
CACGATTTCGTGGAGCTGATTGAAACCTATGAGAACCTGAACAATATCAAGCTGAATCTGGACTACGAGAAAACCAAAGTGATC
TATGAGATTCTGAAAGACAACGAAATGACTAAGAATGATAGAGCCAAAAGGGTCAAGAACATGGAGAAGAAACTGGAACAGTTC
TCTATCATGCTGCTGGGGCTGAAGTTCAATGAGGGAAAACTGTTTAACCACGCCGATAATGCTGAGGAACTGAAGGGGGCTAAC
CAGAGCCATACATTTGCAGACAACTACGAGGAAAATCTGACTCCCTTCCTGACCGTGGAACAGTCAGAGTTTATTGAAAGGGCC
AACAAAATCTATCTGAGCCTGACTCTGCAGGATATCCTGAAGGGCAAGAAATCAATGGCTATGAGCAAAGTGGCCGCTTACGAC
AAGTTCAGAAATGAGCTGAAACAGGTGAAGGACATTGTCTATAAGGCTGATTCTACCAGGACACAGTTCAAGAAAATCTTTGTG
AGCTCCAAGAAAAGTCTGAAGCAGTACGACGCAACTCCCAACGATCAGACCTTCTCTAGTCTGTGCCTGTTTGACCAGTACCTG
ATTCGCCCAAAGAAACAGTATAGCCTGCTGATCAAGGAGCTGAAGAAAATCATTCCCCAGGACTCCGAACTGTACTTTGAGGCA
GAAAATGATACCCTGCTGAAGGTGCTGAACACCACAGACAATGCTAGCATCCCTATGCAGATTAACCTGTACGAGGCAGAAACC
ATCCTGCGAAATCAGCAGAAATATCACGCCGAGATCACAGATGAGATGATTGAAAAGGTGCTGTCTCTGATCCAGTTCCGCATT
CCATACTATGTGGGGCCCCTGGTCAACGACCATACAGCCAGTAAGTTTGGATGGATGGAGCGCAAAAGTAACGAATCAATCAAG
CCTTGGAATTTCGACGAGGTGGTCGATCGAAGTAAATCAGCCACTCAGTTTATTAGGCGCATGACCAACAAGTGTTCCTACCTG
ATCAATGAGGATGTGCTGCCAAAAAACTCTCTGCTGTATCAGGAGATGGAAGTCCTGAACGAACTGAATGCCACACAGATCAGG
CTGCAGACTGACCCAAAAAACCGCAAGTACCGAATGATGCCCCAGATTAAGCTGTTCGCTGTGGAGCACATCTTTAAGAAATAT
AAAACCGTCAGCCATTCCAAGTTCCTGGAAATTATGCTGAACAGCAATCACAGGGAGAACTTTATGAATCATGGAGAAAAGCTG
AGTATCTTCGGCACACAGGACGATAAGAAATTTGCATCAAAGCTGTCAAGCTACCAGGACATGACTAAAATCTTCGGGGATATT
GAGGGAAAGCGCGCCCAGATTGAGGAAATCATTCAGTGGATCACCATTTTTGAGGACAAGAAAATCCTGGTGCAGAAGCTGAAA
GAGTGCTATCCTGAACTGACATCCAAGCAGATCAACCAGCTGAAGAAACTGAATTACTCTGGCTGGGGGAGGCTGAGTGAGAAG
CTGCTGACTCACGCCTATCAGGGCCATAGCATCATTGAACTGCTGCGCCACTCCGATGAGAATTTCATGGAAATTCTGACCAAC
GACGTGTACGGGTTCCAGAATTTTATCAAAGAGGAAAACCAGGTCCAGAGCAATAAGATCCAGCATCAGGATATTGCCAACCTG
ACTACCTCTCCCGCTCTGAAGAAAGGCATCTGGAGTACAATTAAGCTGGTGCGGGAGCTGACTTCCATTTTCGGGGAGCCTGAA
AAGATCATTATGGAGTTTGCTACCGAGGACCAGCAGAAAGGCAAGAAACAGAAATCAAGAAAGCAGCTGTGGGACGATAACATC

FIG. 9D

AAGAAAAATAAGCTGAAAAGCGTGGACGAGTACAAATATATCATTGATGTCGCCAATAAGCTGAACAATGAGCAGCTGCAGCAG
GAAAAACTGTGGCTGTACCTGAGCCAGAACGGCAAGTGTATGTATAGCGGGCAGTCCATCGACCTGGATGCCCTGCTGTCCCCC
AATGCTACCAAGCACTACGAGGTGGATCATATTTTCCCTCGGAGCTTCATCAAGGACGATAGCATTGACAACAAGGTGCTGGTC
ATCAAGAAAATGAATCAGACAAAGGGCGATCAGGTGCCCCTGCAGTTCATTCAGCAGCCTTACGAGAGAATCGCATATTGGAAG
AGCCTGAACAAAGCCGGGCTGATCTCTGATAGTAAACTGCACAAGCTGATGAAACCAGAGTTCACAGCTATGGACAAGGAAGGC
TTCATCCAGCGGCAGCTGGTGGAGACTAGACAGATCAGCGTGCATGTCCGGGATTTTCTGAAAGAGGAATACCCTAATACCAAA
GTGATCCCAATGAAGGCCAAAATGGTGAGCGAGTTCCGGAAGAAATTTGACATCCCAAAGATTAGACAGATGAACGACGCACAC
CATGCCATCGATGCTTACCTGAATGGCGTGGTCTATCACGGGGCACAGCTGGCCTACCCCAACGTGGACCTGTTTGATTTCAAT
TTTAAGTGGGAGAAAGTCCGAGAAAAGTGGAAAGCCCTGGGAGAGTTCAACACAAAGCAGAAATCTCGGGAACTGTTCTTTTTC
AAGAAACTGGAGAAGATGGAAGTGTCCCAGGGCGAGCGGCTGATCTCTAAGATCAAGCTGGACATGAACCACTTCAAGATCAAC
TACTCCAGAAAGCTGGCCAACATCCCTCAGCAGTTTTATAATCAGACCGCAGTGTCTCCAAAGACAGCCGAGCTGAAATACGAA
TCTAACAAGAGTAATGAGGTGGTCTATAAGGGACTGACACCATACCAGACTTATGTGGTCGCCATCAAGAGCGTGAACAAGAAA
GGCAAGGAGAAAATGGAATACCAGATGATCGACCACTACGTGTTCGATTTTTATAAATTCCAGAACGGCAATGAGAAGGAACTG
GCTCTGTACCTGGCACAGAGGGAGAACAAGGACGAAGTGCTGGATGCTCAGATTGTCTATAGTCTGAATAAGGGGATCTGCTG
TACATCAACAATCATCCCTGCTATTTCGTGTCACGCAAAGAGGTCATCAACGCAAAGCAGTTTGAGCTGACCGTGGAACAGCAG
CTGTCTCTGTACAACGTGATGAACAACAAGGAGACAAATGTCGAAAAGCTGCTGATCGAGTATGACTTCATTGCCGAGAAAGTG
ATCAACGAATACCACCATTATCTGAATAGCAAGCTGAAAGAAAAGCGAGTCCGGACCTTTTTCTCAGAGAGCAACCAGACACAC
GAGGACTTCATCAAGGCCCTGGACGAGCTGTTTAAGGTGGTCACCGCATCCGCCACAAGGTCTGATAAAATCGGGAGTCGCAAG
AACAGCATGACTCATCGAGCCTTCCTGGGAAAAGGCAAGGACGTGAAGATTGCTTACACCTCCATCTCTGGACTGAAAACAACT
AAACCTAAGAGTCTGTTTAAGCTGGCCGAGTCAAGAAACGAACTGTAAGAATTC

#7

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACAAAA
ATCAAAGACGACTACATCGTGGGACTGGACATCGGCACAGACTCCTGCGGGTGGGTGGCTATGAACAGCAATAATGACATTCTG
AAACTGCAGGGCAAGACCGCAATCGGGTCACGCCTGTTCGAGGGAGGGAAGAGCGCAGCTGAACGGAGACTGTTTCGCACCACA
CACAGGCGCATCAAACGACGGAGATGGCGACTGAAGCTGCTGGAGGAGTTCTTCGACCCCTACATGGCAGAGGTGGATCCTTAT
TTCTTTGCCCGGCTGAAGGAATCTGGCCTGAGTCCACTGGACAAAAGAAAGACCGTGAGCTCCATTGTGTTCCCCACATCCGCC
GAGGATAAGAAGTTCTACGACGATTACCCTACAATCTACCATCTGAGGTATAAACTGATGACTGAGGACGAAAAGTTCGATCTG
CGCGAAGTGTACCTGGCTATCCACCATATCATTAAGTACCGAGGAAACTTCCTGTATAATACCAGTGTGAAAGACTTCAAGGCA
TCAAAGATCGATGTCAAATCTAGTATCGAGAAGCTGAACGAGCTGTATGAAAATCTGGGCCTGGACCTGAACGTGGAGTTCAAC
ATTAGCAATACTGCCGAGATCGAAAAGGTGCTGAAAGACAAGCAGATCTTCAAGCGGGATAAAGTCAAGAAAATTGCCGAGCTG
TTTGCTATCAAAACCGACAACAAGGAACAGAGCAAGAGAATCAAAGATATTTCCAAACAGGTGGCCAATGCTGTCCTGGGGTAC
AAGACCAGGTTCGACACAATCGCTCTGAAAGAGATTTCCAAGGACGAACTGTCTGATTGGAACTTCAAACTGTCAGACATCGAT
GCAGACAGCAAGTTTGAGGCCCTGATGGGAAACCTGGATGAGAATGAACAGGCCATCCTGCTGACTATTAAGGAGCTGTTTAAC
GAAGTGACCCTGAATGGAATTGTCGAGGACGGCAACACCCTGAGCGAATCCATGATCAACAAGTACAATGATCACCGGGACGAT
CTGAAGCTGCTGAAAGAAGTGATCGAAAATCATATTGACAGAAAGAAAGCCAAGGAGCTGGCACTGGCCTACGATCTGTATGTC
AACAATAGGCACGGACAGCTGCTGCAGGCTAAGAAAAAGCTGGGCAAAATCAAGCCCCGCTCTAAGGAGGACTTCTACAAAGTG
GTCAACAAGAATCTGGACGATTCACGGGCAAGCAAGGAGATCAAAAAGAAAATTGAACTGGACAGCTTTATGCCTAAGCAGAGA
ACCAACGCCAATGGCGTGATCCCATACCAGCTGCAGCAGCTGGAGCTGGATAAGATCATCGAAAACCAGTCTAAGTACTATCCA
TTCCTGAAGGAGATTAATCCCGTGTCAAGCCACCTGAAAGAGGCCCCCTATAAGCTGGACGAACTGATCCGATTTCGGGTGCCT
TACTATGTCGGCCCCCTGATTTCTCCTAACGAGAGTACCAAGGATATCCAGACAAAGAAAAACCAGAATTTCGCCTGGATGATT
CGCAAAGAGGAAGGGCGAATCACACCTTGGAACTTTGACCAGAAGGTGGATCGAATTGAGAGCGCCAATAAGTTCATCAAACGG
ATGACTACCAAGGACACTTACCTGTTTGGGGAGGATGTGCTGCCAGCTAACAGCCTGCTGTATCAGAAGTTCACCGTCCTGAAC
GAACTGAACAACATCCGGATTAATGGAAAAAGAATCTCCGTGGACCTGAAGCAGGAGATCTACGAAAACCTGTTTAAGAAACAC
ACAACTGTGACCGTCAAGAAACTGGAGAATTATCTGAAGGAAAACCATAATCTGGTGAAAGTCGAGATCAAGGGGCTGGCCGAT
GAAAAGAAATTCAACAGCGGACTGACCACATACAATAGATTCAAGAACCTGAACATCTTTGACAACCAGATTGACGATCTGAAG
TACAGGAACGATTTCGAGAAGATCATCGAATGGTCTACAATTTTTGAGGACAAGAGTATCTACAAAGAAAAGCTGAGGAGCATC
GATTGGCTGAACGAGAAGCAGATTAACGCTCTGTCTAATATCAGACTGCAGGGGTGGGGAAGGCTGAGTAAGAAACTGCTGGCA
CAGCTGCACGACCATAATGGCCAGACCATCATTGAGCAGCTGTGGGATTCCCAGAACAATTTCATGCAGATTGTGACACAGGCC
GACTTTAAAGATGCTATCGCAAAGGCCAACCAGAATCTGCTGGTGGCTACCTCAGTCGAGGACATTCTGAACAATGCATACACA
AGCCCCGCAAACAAGAAAGCCATCAGACAGGTCATCAAGGTGGTCGACGATATCGTGAAGGCAGCCTCCGGAAAGGTCCCAAAA
CAGATCGCCATTGAGTTCACTAGGGATGCTGACGAAAATCCCAAGAGAAGTCAGACCAGGGGCTCAAAGCTGCAGAAAGTGTAC
AAGGACCTGAGCACTGAGCTGGCCTCCAAGACCATTGCTGAGGAACTGAACGAAGCAATCAAAGACAAGAAACTGGTGCAGGAT
AAGTACTATCTGTACTTTATGCAGCTGGGGCGGGACGCCTATACAGGAGAGCCTATCAATATCGATGAAATCCAGAAGTACGAT
ATCGACCACATTCTGCCACAGTCTTTCATCAAGGACGATGCCCTGGACAACAGGGTGCTGGTGAGCCGGGCTGTGAACAATGGC
AAATCTGATAATGTGCCTGTCAAGCTGTTTGGCAACGAGATGGCTGCAAATCTGGGGATGACTATCAGGAAAATGTGGGAGGAA
TGGAAGAACATCGGCCTGATTAGCAAAACAAAGTACAACAATCTGCTGACTGATCCCGACCACATTAACAAGTATAAGAGTGCC
GGGTTCATCAGGCGCCAGCTGGTGGAGACATCACAGATCATCAAGCTGGTGAGCACTATCCTGCAGAGTCGCTACCCTAACACT
GAAATCATTACCGTGAAGGCTAAGTACAATCATTATCTGCGGGAGAAATTTGACCTGTATAAGAGCAGAGAAGTCAACGACTAC
CACCATGCTATTGATGCATATCTGTCCGCCATCTGCGGAAATCTGCTGTACCAGAACTATCCAAATCTGCGGCCCTTCTTTGTG
TACGGCCAGTATAAGAAATTCTCCTCTGATCCTGACAAAGAGAAGGCCATTTTTAACAAAACCCGCAAGTTCTCCTTTATCTCT
CAGCTGCTGAAAAACAAGAGTGAGAACAGCAAGGAAATCGCTAAGAAACTGAAACGGGCATACCAGTTCAAGTATATGCTGGTG
TCTCGAGAGACTGAAACCCGGGACCAGGAGATGTTCAAAATGACCGTGTACCCCCGGTTCAGCCACGATACAGTCAAGGCTCCT

FIG. 9E

AGGAACCTGATTCCAAAGAAAATGGGCATGTCCCCTGACATCTACGGAGGCTATACAAACAATTCTGACGCATACATGGTCATC
GTCCGCATTGATAAGAAAAAGGGAACTGAGTATAAGATCCTGGGCATTCCAACCCGGGAACTGGTGAATCTGAAAAAGGCCGAG
AAGGAGGACCATTACAAAAGCTATCTGAAGGAGATCCTGACACCAAGGATTCTGTACAACAAAAATGGGAAGCGCGATAAAAAG
ATCACTTCCTTCGAAATTGTGAAATCTAAGATCCCCTATAAGCAGGTCATCCAGGATGGGGACAAAAAGTTTATGCTGGGAAGT
TCAACATACGTGTATAACGCAAAGCAGCTGACACTGAGCACTGAGTCCATGAAAGCCATCACTAACAATTTCGATAAGGACAGC
GATGAGAACGACGCTCTGATTAAGGCATACGATGAAATCCTGGACAAAGTGGATAAGTATCTGCCACTGTTCGACATCAACAAG
TTCCGGGAGAAGCTGCACAGTGGGCGAGAAAAGTTCATCAAGCTGAGCCTGGAGGACAAAAAGGATACCATCCTGAAAGTGCTG
GAAGGACTGCATGATAACGCTGTCATGACAAAGATCCCTACTATTGGCCTGTCCACACCACTGGGGTTCATGCAGTTTCCCAAC
GGCGTGATTCTGAGCGAGAATGCCAAACTGATCTACCAGTCCCCCACCGGGCTGTTCAAAAAGTCAGTGAAGATCAGCGACCTG
TAAGAATTC

#8
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGGGCTAC
ACTGTGGGACTGGATATTGGGGTGGCTTCCGTCGGGGTGGCTGTGCTGGATGAGAATGACAACATCGTGGAGGCTGTGTCAAAC
ATCTTTGATGAAGCCGACACAAGCAACAATAAGGTGCGGAGAACTCTGAGGGAGGGCAGGCGCACAAAGCGGCGGCAGAAAACC
CGCATTGAGGACTTCAAGCAGCTGTGGGAGACTTCAGGCTACATCATTCCTCACAAGCTGCATCTGAATATCATTGAGCTGCGC
AACAAAGGGCTGACCGAACTGCTGAGCCTGGATGAGCTGTATTGCGTGCTGCTGTCCATGCTGAAGCACCGGGGGATCTCCTAC
CTGGAGGACGCCGACGATGGCGAGAAGGGGAATGCCTATAAGAAAGGACTGGCTTTTAACGAAAAACAGCTGAAGGAGAAAATG
CCATGTGAGATCCAGCTGGAACGCATGAAGAAATACGGGAAGTACCATGGAGAGTTCATCATCGAAATTAATGATGAGAAGGAA
TACCAGAGCAACGTGTTCACCACAAAGGCTTATAAGAAGGAGCTGGAAAAGATCTTCGAGACACAGCGGTGCAACGGCAACAAG
ATCAACACAAAGTTCATTAAGAAATACATGGAGATCTACGAACGAAAGCGGGAATACTATATCGGACCAGGCAATGAGAAAAGC
AGAACAGACTACGGCATCTATACTACCAGGACTGATGAGGAAGGGAATTTCATCGACGAGAAGAACATTTTTGGCAAACTGATC
GGGAAGTGTAGTGTGTACCCCGAGGAATATAGAGCAAGCTCCGCCTCATACACCGCCCAGGAGTTCAATCTGCTGAACGATCTG
AACAATCTGAAAATCAACAATGAGAAGCTGACAGAATTTCAGAAGAAAGAGATTGTCGAAATCATTAAGGACGCTTCTAGTGTG
AACATGAGGAAAATCATTAAGAAAGTCATCGATGAGGACATTGAACAGTACAGCGGAGCACGAATCGATAAGAAAGGCAAGGAA
ATCTACCACACCTTCGAGATCTATCGGAAGCTGAAGAAAGAGCTCAATGTGGATATCGACTCTTTTACTAGAGAG
GAACTGGATAAGACCATGGACATCCTGACCCTGAACACAGAGAGGGAAAGTATTGTGAAGGCCTTCGACGAACAGAAATTTGTC
TACGAGGAAAATCTGATCAAGAAACTGATTGAGTTTCGGAAGAACAATCAGAGACTGTTCAGCGGCTGGCATAGTTTTTCATAC
AAGGCTATGCTGCAGCTGATCCCAGTGATGTACAAGGAGCCCAAAGAACAGATGCAGCTGCTGACCGAAATGAACGTGTTCAAA
AGTAAGAAAGAGAAGTACGTCAACTACAAGTACATCCCAGAGAACGAAGTGGTCAAGGAGATCTATAACCCCGTGGTCGTGAAG
AGCATTAGAAACAACTGTGAAAATTCTGAATGCACTGATCAAGAAATACGGGTATCCTGAATCCGTCGTGATCGAGATGCCAAGG
GATAAGAACTCTGACGATGAGAAGGAAAAGATCGACATGAACCAGAAGAAAAACCAGGAGGAATACGAGAAAATCCTGAACAAG
ATCTACGATGAGAAGGGAATCGAAATTACCAACAAGGACTACAAGAAACAGAAGAAACTGGTGCTGAAGCTGAAACTGTGGAAC
GAGCAGGAAGGACTGTGCCTGTATTCCGGCAAGAAAATCGCTATTGAGGATCTGCTGAATCACCCCGAGTTCTTTGAAATTGAC
CATATCATTCCTAAGAGCATCTCCCTGGACGATTCTCGCAGTAACAAGGTCCTGGTGTACAAAACAGAAAATTCTATCAAGGAG
AACGATACCCCCTACCACTATCTGACACGGATTAACGGAAAGTGGGGCTTTGACGAATATAAAGCTAATGTGCTGGAGCTGAGA
AGGCGCGGCAAGATCGACGATAAGAAAGTGAACAATCTGCTGTGCATGGAGGATATCACTAAGATTGACGTCGTGAAAGGGTTC
ATTAACCGCAATCTGAACGACACCAGATACGCATCCAGGGTGGTGCTGAACGAAATGCAGTCCTTCTTTGAGTCTCGAAAGTAC
TGTAATACTAAGGTCAAAGTGATCCGAGGCTCTCTGACCTATCAGATGCGGCAGGATCTGCACCTGAAGAAAAACAGAGAGGAA
TCATACAGCCACCATGCTGTGGACGCAATGCTGATCGCATTCTCCCAGAAGGGGTACGAGGCCTATAGGAAGATCCAGAAAGAT
TGCTACGACTTTGAGACAGGCGAAATTCTGGACAAGGAAAAATGGAATAAGTACATTGACGATGACGAGTTTGATGACATCCTG
TATAAAGAGAGGATGAACGAAATCCGCAAGAAAATCATTGAGGCCGAGGAAAAGGTGAAGTACAACTACAAGATCGATAAGAAG
TGCAATCGCGGGCTGTGTAACCAGACTATCTACGGGACCCGAGAAAAGGACGGAAAAATCCACAAGATTTCAAGCTACAACATC
TATGATGACAAGGAGTGTAATTCCCTGAAGAAAATGATTAACAGTGGGAAAGGATCAGATCTGCTGATGTACAACAATGATCCT
AAGACATATCGCGACATGCTGAAAATCCTGGAAACTTACTCCTCTGAGAAGAATCCATTCGTGGCATATAACAAAGAGACAGGA
GACTACTTTCGGAAATATTCTAAGAATCACAACGGACCCAAGGTCGAGAAGGTGAAATACTATAGCGGCCAGATCAACTCCTGC
ATCGATATTTCTCACAAGTACGGCCATGCCAAAAATAGTAAGAAAGTCGTGCTGGTGTCACTGAACCCTTATAGAACCGACGTC
TACTATGATAATGACACAGGCAAGTACTATCTGGTCGGGGTGAAGTACAATCATATCAAATGTGTCGGAAACAAGTACGTGATT
GATAGCGAGACATATAACGAACTGCTGAGGAAGGAGGGCGTGCTGAACAGCGATGAGAACCTGGAGGACCTGAACAGCAAAAAC
ATCACTTACAAGTTCAGTCTGTACAAGAACGACATCATCCAGTACGAGAAGGGCGGGGAATACTATACAGAGCGCTTTCTGAGC
CGAATCAAAGAACAGAAGAACCTGATTGAGACTAAACCCATCAATAAGCCTAACTTCCAGCGGAAGAATAAGAAAGGCGAGTGG
GAAAATACCAGAAACCAGATCGCCCTGGCTAAGACTAAATACGTGGGGAAGCTGGTCACCGATGTGCTGGGAAACTGTTACATC
GTGAACATGGAGAAGTTCTCCCTGGTCGTGGACAAATAAGAATTC

#9
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACTAAC
GGCAAGATTCTGGGGCTGGACATTGGCATCGCAAGCGTGGGGGTGGGGATTATTGAGGCAAAAACTGGAAAGGTGGTGCATGCC
AATTCCCGGCTGTTCTCTGCCGCTAACGCTGAGAACAATGCAGAACGGAGAGGGTTTAGGGGATCTAGGCGCCTGAATCGACGG
AAGAAACACCGCGTGAAGCGAGTCCGGGATCTGTTCGAGAAATACGGAATCGTCACCGACTTTCGCAACCTGAATCTGAACCCT
TATGAGCTGCGAGTGAAGGGCCTGACCGAACAGCTGAAAAACGAGGAACTGTTCGCAGCCCTGAGAACAATCTCTAAGAGAAGG
GGGATTAGTTACCTGGACGATGCCGAGGACGATAGTACCGGATCAACAGACTATGCTAAGAGCATCGATGAGAATCGCCGACTG
CTGAAAAACAAGACACCAGGCCAGATTCAGCTGGAGAGGCTGGAAAAGTACGGCCAGCTGCGCGGGAATTTCACCGTCTATGAC
GAGAACGGGGAAGCCCATCGCCTGATCAATGTGTTTAGTACATCAGATTACGAGAAAGAAGCACGGAAGATCCTGGAGACACAG

FIG. 9F

GCCGACTACAACAAGAAAATCACAGCTGAGTTCATTGACGATTATGTGGAAATCCTGACCCAGAAACGAAAGTACTATCACGGC
CCCGGGAACGAAAAGAGCCGGACTGACTACGGACGGTTCCGGACCGATGGGACCACACTGGAGAATATTTTCGGAATCCTGATT
GGCAAGTGCAACTTTTACCCTGATGAATATCGAGCAAGCAAGGCCAGCTACACCGCACAGGAGTATAATTTCCTGAACGACCTG
AACAATCTGAAGGTGAGCACCGAAACAGGGAAGCTGTCAACAGAGCAGAAAGAAAGCCTGGTGGAGTTTGCCAAGAATACTGCT
ACCCTGGGACCCGCTAAACTGCTGAAGGAGATCGCAAAAATTCTGGACTGTAAGGTGGATGAGATCAAAGGATACAGAGAGGAC
GATAAAGGCAAGCCAGATCTGCATACCTTCGAGCCCTATAGGAAACTGAAGTTTAATCTGGAAAGCATCAACATTGACGATCTG
TCCCGCGAAGTGATCGACAAGCTGGCTGATATTCTGACTCTGAACACCGAGAGAGAAGGAATCGAGGACGCAATTAAGAGGAAT
CTGCCAAACCAGTTCACAGAGGAACAGATCAGCGAGATCATCAAGGTGCGGAAGAGCCAGTCCACTGCTTTCAATAAGGGCTGG
CACTCTTTTAGTGCAAAACTGATGAACGAGCTGATCCCCGAACTGTACGCCACCTCCGACGAGCAGATGACAATTCTGACTCGG
CTGGAAAAATTCAAGGTCAATAAGAAAAGCTCCAAAAACACAAAGACTATCGACGAGAAGGAAGTCACTGATGAGATCTACAAT
CCTGTGGTCGCCAAGAGCGTGAGACAGACCATCAAAATCATTAACGCTGCAGTCAAGAAATATGGCGACTTCGATAAGATCGTG
ATTGAAATGCCACGGGATAAAAATGCTGACGATGAGAAGAAGTTCATCGACAAGAGAAATAAGGAGAACAAGAAGGAAAAGGAC
GATGCCCTGAAAAGGGCCGCTTACCTGTATAATTCTAGTGACAAGCTGCCCGATGAGGTGTTCCACGGCAACAAGCAGCTGGAA
ACCAAAATCCGACTGTGGTATCAGCAGGGGGAGCGGTGCCTGTATAGTGGAAAGCCCATCTCAATTCAGGAGCTGGTGCATAAC
TCTAACAATTTCGAAATCGATCACATTCTGCCTCTGTCACTGAGCTTTGACGATAGTCTGGCCAATAAGGTGCTGGTCTACGCT
TGGACAAACCAGGAGAAAGGCCAGAAAACCCCTTATCAGGTCATCGACTCCATGGATGCAGCCTGGTCTTTCAGGGAGATGAAG
GACTACGTGCTGAAACAGAAGGGACTGGGCAAGAAAAAGCGCGACTATCTGCTGACTACCGAGAACATCGATAAGATTGAAGTG
AAGAAGAAGTTCATCGAGAGGAATCTGGTGGATACTCGCTACGCATCTCGAGTGGTCCTGAACTCTCTGCAGAGTGCCCTGAGA
GAGCTGGGGAAAGACACTAAGGTGTCTGTGGTCAGGGGACAGTTCACCAGTCAGCTGCGGAGAAAATGGAAGATCGATAAGAGC
CGCGAGACATACCACCATCACGCAGTGGACGCCCTGATCATTGCTGCATCAAGCCAGCTGAAACTGTGGGAGAAGCAGGACAAT
CCCATGTTTGTGGATTATGGCAAGAACCAGGTGGTCGACAAACAGACTGGGGAGATCCTGTCCGTGTCTGACGATGAGTACAAG
GAACTGGTGTTCCAGCCCCCTTATCAGGGCTTTGTGAATACCATCTCCTCTAAAGGGTTCGAGGACGAAATTCTGTTTAGCTAC
CAGGTGGATTCCAAATATAACCGGAAGGTCAGTGACGCAACCATCTACTCAACAAGAAAAGCCAAGATTGGCAAGGATAAGAAA
GAGGAAACCTACGTGCTGGGAAAAATCAAGGACATCTACTCCCAGAATGGCTTCGATACCTTCATCAAGAAGTACAACAAAGAT
AAGACTCAGTTCCTGATGTATCAGAAGGACTCTCTGACATGGGAGAACGTGATCGAAGTCATTCTGAGGGACTACCCAACAACT
AAGAAAAGCGAGGACGGCAAAAATGATGTGAAGTGCAACCCCTTTGAGGAATACAGGCGCGAGAATGGGCTGATCTGTAAGTAT
TCCAAGAAAGGGAAAGGAACTCCCATCAAGAGCCTGAAGTACTATGACAAGAAACTGGGGAACTGCATCGATATTACCCCAGAG
GAATCACGCAATAAGGTCATCCTGCAGAGCATTAACCCTTGGCGAGCCGACGTGTACTTCAATCCAGAGACACTGAAGTACGAA
CTGATGGGCCTGAAATATTCAGATCTGAGCTTTGAAAAGGGCACTGGGAACTACCATATCAGCCAGGAGAAATATGACGCTATC
AAAGAGAAGGAAGGAATTGGCAAGAAATCCGAGTTCAAGTTTACACTGTACCGCAACGACCTGATCCTGATCAAGGATATCGCC
AGTGGCGAGCAGGAAATCTACAGATTCCTGTCAAGAACTATGCCCAATGTGAACCACTACGTCGAGCTGAAGCCTTACGACAAG
GAAAAGTTCGATAACGTGCAGGAGCTGGTCGAAGCACTGGGAGAGGCAGATAAAGTGGGACGATGTATCAAAGGACTGAATAAG
CCAAACATCAGCATCTACAAGGTGAGAACCGACGTCCTGGGAAACAAATATTTCGTGAAGAAAAAGGGCGACAAACCCAAGCTG
GATTTTAAGAACAACAAGAAGTAAGAATTC

#10
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGGCCGAC
CGAATCTCTCTGGGGCTGGACATTGGGGTGGCAAGCGTGGGGTTCTCAGTGCTGGACATTGATAAGGGAAAAGTCATTGAGCTG
GGCGCCAGGCTGTTCTCTGCTACTGTGGCCGCTGGCAACCAGGATCGAAGAGACATGCGAGGAGCCAGGCGCCTGCTGAACCGG
AACAAGCAGCGACGGCAGGATACCGGAAAGCTGTTCAAGAAATTTGGCCTGATCGACGATTTTGATAAGGGCAGCTTCTACGAC
AACTTTAATCAGAACCTGAATCCTTATGAGCTGAGAGTGAAAGGCCTGACAGAACAGCTGACTAAGGAGGAACTGGCCGAGTCT
CTGTACCAGATCGTGAAACATAGGGGGATTAGTTATGCACTGAAGGACGCCGATGTGGACGAAGGCGGGACAGACTACTCAGTC
AGCCTGAAAATCAACAGCCAGGAGCTGGCAGAAAAGACTCCAGCCCAGATTCAGCTGCAGAGACTGAATGATTATGGAAAGGTG
AGGGGCCAGGTGGTCATCGGCGACGATCCAGACAACCAGAAGGTGCTGCTGAATGTGTTCCCCACATCAGCTTACGAGAAAGAA
GCAAAGCAGATCATTGCCACTCAGCAGCAGTTCTATCCTGAGAGCCTGACCGACAAGTTCACCGAGGAATACTGCCAGATCCTG
ACTCGCAAGCGAGATTATTTTGTGGGGCCAGGAAACGAGAAAAGCCGGACCGACTACGGGATCTACAAGACTGATGGAAGAACC
CTGGACAATCTGTTCGAGGAACTGATCGGCCACGATAAGATCTACCCCGAGGAACTGCGGGCATCTGCAGCCAGTTATACCGCC
CAGCTGTTTAACGTGCTGAATGACCTGAACAATCTGAGAATCCTGAACTACGAGGATGGGAAACTGACAAAGGAGGACAAGGAA
AAGATCATCGCTGAAATTAAGAACAACACCACAACTATCAACATGCTGAATGTGATTAAGAAAGTCGCCGGGTGTTCCAAGGAC
GATATCAAAGGATTCCGAGTGAATGAGAAGGATAAACCCGAAATCAGCTCCATGCCTGTGTACCGCAAAATTCATAAGGACCTG
CTGAAGGCCGGCGTGGATATCTCAGACTGGCCCGTCGAGTTCATCGACGAACTGAGCTTTATTCTGACACTGAACACTGAGAAT
GGGGAAATTCGCAAACAGCTGAACAATCGACTGGCCCCTAAGTTCGATTTTCTGAACGCTGACCTGATCCAGCTGATCATTGAT
AATAAGGACTCCTTTGAGATTAAGACTAACAACAAGTGGCACAGATTCAGCGTGAAAACCATGAACAAACTGATCCCAGAGATG
ATGGAAAGACCCGTGGAGCAGATGACCCTGCTGAATGAAATGGGACTGGTCAAGAAAGATAAGAAACGCTTTGAGAACAATAAG
TACCTGCCTTACAAGGAAATCGCAAAGGACATTTTCAACCCAGTGGCCTCCAAATCTGTCCGCGAGGCCCTGAAGATCGTGAAT
GCTGTCCTGAAGAAATACGGCCACATTGATTATCTGGTGGTCGAAATGCCTCGGGATAAAAACCTGAAGGAGGAACAGGACAAT
ATCAAGGAGTTCCAGAACAAAAATAAGAAAGCTAAGGACGCTGCATTCGAAGCATTTGTGAAATCAGTCGGGAGCGAGCAGAGA
GTGAAGGAAGCCCTGTCTAAAAACCGGAAGCTGCAGATGAAGATGAGACTGTGGTATCAGCAGCAGGAGATCGATCCATATAAT
GGAAAGACAATCGATGCCACTGACCTGATTAACAATCCTGATAAGTTCGAGATTGACCATATCATTCCACAGAGTATCTCATTT
TACGACAGTATTAACAATAAGACCCTGTGCTTCGCCTCAATGAACCAGGTGAAAGGACAGAAAACCCCCTACGAGTTTATGCTG
GAAGGCCACGGGCAGTCCTATGACAAGTTCAAAGCTACAGTGATGGCAAACAAGAATTTTGGCAAGGCTAAAAGGGCAAACTAC
CTGTTCGAGGAAAATGTGAGCGATATCGAGACTCGGAAGAGATTCCTGTCCCGCAACCTGGTGGACACCCGATATTCTAGTCGG
GTGGTGCTGAACAGCCTGCAGGATTTCTTTCGGGAGAAATCTGCCGACACCAAGGTGACAGTCATTCGCGGCAAGTTTACCTCC

FIG. 9G

AACATGCGAAAACATTGGCACATCGATAAGACTAGGGAGACATTCCACCATCACGCCATTGACGCTTCTATCATTGCCGCTACA
CCATTTCTGCGCATGTGGAAGAAAGGAGGCACTATCTTCCCCGTGAAGGTCGGAGAAGAAAGTATCGATATTGAGACAGGCGAA
ATTCTGGACGATAAGAATTTTGACAAAGCAATGTACGAGGAACCCTATAGTGGCTTCGTGTCAGAGATCATGAACGCCGACGAT
CGGATCAAGTTCAGCCACCAGGTGGATAAGAAAATGAATAGGAAGGTGAGCGACGCCACCATCTACAGTACTCGCACCGGGAAA
CTGGCTAAGGATAAGAAAGACGCTGAGTACATCGTGGCAAAGGTCAAAGATATCTACAGCGTGGACGGATTCAAGAAGTTCAAG
AAAGTCTACGATAAGGACAAAACCAAGTTTCTGCTGTACAAATATGATCCTAGGACATTCTCAAAGCTGGAGCGCATCATTAGC
GATTGCCCAGACAAAGTGGAAAAGGTCCAGACAAACGGCAAAGTGAAGGCTGTCGATATCAGTCCATTCGAGATGTACAGAAGG
GACCATGGGATGATCAAGAAATACTCAAAGAAAGGAAACGGCCCCGCCATCAAGCAGCTGAAGTACCTGGATAAGAAACTGGGC
AGCCACATCGACATTACCCCCGCAAACGCCAATGGAAAACACGTGATCCTGCAGAGTCTGAAGCCTTGGAGAACCGACGTCTAT
CTGAACCACGAGACAGGCGAGTACGAAATCATGGGGATTAAGTATAGCGATCTGAAGTTCAACAAGAATGAGGGGTACGGAATC
AAGAAAGACAAGTATCTGGAAATTAAGAAAGTGGAGGAAGTCTCCGAGAAGTCTGAGTTCATGTTTAGCCTGTACAGGAAGGAT
CGCGTGAAAGTCCAGGACATGAAAACCGGCGAGTCCGTGGAACTGCTGTTCTGGAGCAGGAACTTTTCCAATAAGAAATACGCT
GAGCTGAAGCCCATCTCCCAGGCAGAAAACGACAAGAAACTGCCTGTGTATGGCAAAGGGAGACTGATCAAGAGGCTGATTCCC
AAAAACTGTAAGATCTGGAAAGTGAATACCACAATTCTGGGCGATCCCTACTATCTGGAGAAAGAAAGCGACTCCCCTAAGGAT
ATCCTGGACTAAGAATTC

#12
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAGAAA
ACACTGGGACTGGACCTGGGGACTAACAGCATTGGGTGGGCCGTCATCAACTCAAACATCGACTCAGAAGGAAAAGAAAAGCTG
GTGGGGATCAGCTCCTGCGGAAGCCGGATCATTCCTATGGACGCTACCACACTGGGAGATTTCGGAAAGGGCAACACTAAAAGT
CCAGTGGCAGAGCGAACCCGACTGCGGGGCATTCGGAGACTGCTGGAAAGGTCACTGCTGAGGCGCGAGCGACTGCACCGAGTG
CTGTCAGTCATGGGGTTCCTGCCAGAGCATTACGCTAGCCAGCTGGACCGCTATGGAAAGTTTCTGCCAGAAACCGAGCCCAAG
CTGGCATGGTACAAAGACGATAGCGGCAGGTATCAGTTCCTGTTTCAGAAGTCCTTCCACGAGATGCTGGAGGACTTTCGACAG
CATCAGCCAGAACTGGTGGCAGGAGAGAAGAAAATCCCTTACGATTGGACAATCTACTATCTGCGGAAGAAAGCACTGTCTCAG
GAGATCACTAAGGAGGAACTGGCCTGGATTCTGCTGAATTTCAACCAGAAACGCGGCTACTATCAGCTGCGAGGGGAGGAAGAG
CAGGAAGAGAACAATAAGAGCGTGGAGTATCACGCCCTGAAAGTGGTGAGCGTCGAGGACTCTGGAGAAAGAAAGGGCAAAGAT
ATCTGGTACAATGTGACTCTGGAGAACGGATGGGTCTATCGACGGGCTAGCAACATTCCCCTGGACTGGACCGGCAAGGTGAAA
GAGTTCGTGGTCACTACCGAGCTGGACGATGCCGGGAACCCAAAGAAAGATAAGGAAGGAAATGTGAAAAGATCCTTTAGGGCA
CCAAAGGAGGACGATTGGGGGCTGCTGAAAACTAGGACCCAGGCTCAGATCGACGAATCCGGCAAGACCGTGGGAACTTACATC
TACGAGTCTCTGCTGTGTATGCCCAACCAGAAGATCCGGGGAAAACTGGTGAGGACCATTGAACGCAAGTACTATAAAGACGAG
CTGAGACAGATCCTGGTGAAGCAGAGCGAGTTCCACGCCGCTCTGCAGGATCATAATCTGCTGCTGTCCTGTATCGAAGAGCTG
TACCCTAACAATGAGGCCCACAGAAGGCTGCTGAGCGCCAGCAGCTTCATCTACTTCCTGATCGAGGACATTCTGTTTTATCAG
CGCCCACTGAAGTCCCAGAAAGGGCTGATCGATAACTGCCCCTACGAGTCTCACATCTACAAGGATAAGAAAGATGGAAGTCTG
CACCATGTGCCTCTGAAGTGTGTCAGCAAATCCCATCCACTGTTCCAGGAATTTCGCCTGTGGCAGTTCCTGTCTAACCTGCGA
ATCTACCAGAGAGAGAGGATGATGGACGGCAGTCTGAAACTGGACGTGGATGTCACCCGGGAGTTCCTGCCCTCAGAAGAGGAC
TACGTGAAGCTGTTTGATTGGCTGAATGAGAGAAAGGAAATCTCTCAGAAATTCCTGCTGGCTTATAAACCTTTTGGGCTGAAG
AAAAACGAACAGGCAAATTACAGATGGAACTATGTGGAGGACAAGAGCTACCCCTGCAACGAGACACGGGCAGAAATCAAGAGC
AGACTGTCCAAAGCCGGAGTGCCTGAAGAGTTTCTGACTGAAGAGAAGGAAGAGGCCCTGTGGCACATCCTGTATTCTATTAGT
GATAAGAAAGAGCTGACTAAGGCTCTGGGCACCTTCGCAGCCAAAAACTGTCTGAATGAGTCTTTCGTGGAAGTCTTTGCCAAG
ATCCCCCCTTTTGAGTCAAACTACGCTGCATATAGCCTGAAGGCTATTAGGAAACTGCTGGCACTGATGCGCATGGGGAAGTAC
TGGAATGAACAGGCCATCGACAGGCAGACTCGCGATCGAATCGAGAAAATTCTGACCGGAGAGTATGACGAAACAATCCGGAGC
AGAGTGAGGGAGAAGGCAATGCTGCTGACCGATATTAGCAGCTTCCGGGGCCTGCCTCTGTGGCTGGCCTGTTACATCGTGTAT
GACCGCCACTCAGAGAGCAAGGAACTGGTCAAATGGGAGACACCAGCCGACATCGATCATTTCCTGTCCAAGTTTAAACAGAAC
AGCCTGCGGAACCCCATCGTGGAACAGGTCATTACAGAGTCCCTGCGCACTGTGCGAGACATTTGGAAGCAGGAGGGAAAAATC
GATGAGATTCACGTGGAACTGGGCCGGGAGATGAAGAACCCTGCAAAAGAGCGCGCCCGAATTACAGCTCAGGTGCAGGAAAAT
GAGAACACTAATCTGAGAATCAAGGCTCTGCTGGCAGAGTTCATGAACCCCGAATTTGAGATTGAAAATGTGCATCCATACTCA
CCCGGCCAGCAGGAAATCCTGCGGATCTACGAGGACGGCGTGCTGAGCGGGATCGCTGAGAAGGATCTGCCTGACGATATCACA
GCAATTCTGAAGAAATTCCGAGAAAACGACGTGAAGAAACGGCCAACAACTAGCGAAGTCCTGCGGTACAAACTGTGGCTGGAG
CAGCGGTACAGATCCCCATATACCGGAAGAGTGATCCCCCTGGGCAAGCTGTTCACACCTGCTTACGAGATCGAACACGTGATT
CCCCAGAGCCGGTATTTTGACGATTCCATCTCTAACAAAGTGATCTGCGAAAGTGCCGTCAATAAGCTGAAAGATAACTGTCTG
GGCTATGAGTTCATCAAGAAACATTCCGGGGAGATGGTGGAACTGGGGAATGGAGAGACAGTGCCCGTGTTCAGCGTGGAAGAG
TACGAACGGTTCGTCAAGGAGTCTTACTTCGGCAACAGTAAGAAAATGAAGAAACTGCTGCTGGAGGACATCCCAGATAGCTTC
ATTGAGAGACAGCTGAATGACAGTCGATACATCTCACGGGTGGTCACATCTCTGCTGAGTAACCTGGTGTGCGAAGAGGGAGAG
CAGGATGGCCTGTCCAAGAATGTGATCGTCTGTACCGGCGGGATTACAGACAGGCTGAAGAAAGATTGGGGAGTGCAGGAAGTC
TGGAACCGCATCATTCTGCCTCGGTTCCTGAGACTGAATGAGATCACCGGACGGACAGACTTTACAAGTACTTCAGTGAACGGC
CACCTGCTGCCTGCCCTGCCACTGTACCTGCAGAAGGGCTTTAATAAGAAAAGAATTGACCATAGGCACCATGCCATGGATGCT
ATCGTGATTGCCTGCGCTAACCGGAATATCGTCAACTACCTGAACAATTCCTCTGCTAGAAAGAACAGCGAAATTAGCCGATAT
GACCTGCAGCGGCTGCTGTGTGAGAAGGTGAAAACCGATGCCAACGGCAATTACAAGTGGATCCTGAGGAAACCATGGGAGACA
TTCACCCAGGATGTGTATGCCGCTCTGACAAACATCGTGGTCAGCTTTAAGCAGAATCTGCGCGTGATTAATCGAACCACAAAC
TACTATCAGCACTACAACGAGGCAGGGGAAAAGCGCATGATCCCTCAGACCAAAGGCGACAGATGGGCCATTAGAAAGCCAATG
CATAAAGATACTGTGTATGGCGAAGTCAATCTGAGAAAGGAGAAAACCCTGCCACTGAGGGACGTGGTCAAGAACCCCAGTATC
GTGGTCGATAAATCACTGAAGAACAAGCTGTACGAGCTGCTGAAGAGCCAGTATGACCTGAAGGCAATCGCCAAATACTTCGAG
ACACACCAGGACGTGTGGGCAGATGTCAACCTGAAGAAAATTAAGGTGTACTACTTCACTAAGGAGACAAACGAAAGATTCTTT

FIG. 9H

GCCACTAGGAAGAGCCTGGACCCATCCTTTGATCAGAAGAAAATCGAAGAGGAAGTGACTGATACCGGCGTCCAGAAGATTCTG
CTGCACCATCTGCAGCAGAACAATAACGACCCTGATATGGCCTTTTCCCCAGACGGCATCGATAGGATGAACCAGAATATGACC
ATTCTGAATGACGGGAAGTGGCACCAGCCCATCTACAAAGTGCGCACATATGAAAAGGCAGATAAATTCGCCGTCGGCGAGTCT
GGGAACAAGGCCAAGAAATTTGTGGAAGCAGCCAAAGGCACCAATCTGTTCTTTGCCGTGTACGAGAGTGTCCAGGAGGACGAA
GCTATCGGGAAGCAGGTCTGCAAACGGACATTCGCCACTATCCCCCTGAACGAAGTGATCAAGAAAAGAAACAGGGCCTGCCC
GCTGCACCTGAGGACCTGAACGGGAATCTGCCCAAGTTTGTGCTGAGCCCTAACGATCTGGTCTACCTGCCCACCGAGGAAGAG
AGGAATAGTTCACGCATCATTCAGCCTCTGGACAGGGAGCGCATCTATAAGATGGTGAGCTCCTCTGGGAGTCAGTGCTTCTTT
ATCAAAGTGTTCGTCGCCAATTCAATTTGGGATAAGAACGAATACAGCAGCCTGAACAAGATGGAGAGGGCTATTACAAACGAA
ATGATCAAGGAGATTTGTGTGCCTATCAAAATTGACCGCCTGGGCAATGTCAGCCTGATCCAGATTTAAGAATTC

#13
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACAAAG
ACTATTCTGGGACTGGATCTGGGGACTAACAGCATTGGATGGGCACTGATTAACCAGGACTTCGATAACAAAAAGGGGGAGATC
CTGGGCATGGGCAGCCGGATCATTCCTATGACACAGGATATTAAGGACGAGTTCGGAAAAGGCAACTCTATCAGTCAGACAGCT
GAGCGAACTCGACTGCGGGGAGTCCGGAGACTGATCCACAGAACTCTGCTGAGGCGCGAGCGGCTGCATAGAGTGCTGAATATC
ATTGGCTTCCTGCCAGAACACTATGCCAACCAGATTGACTTTGAGAAGAGGTTCGGCAAGTTCAAGCCTGAATCTGAGCCAAAA
ATCAGTTTCGATGGAAATGACGTGTTCCTGTTTGAGAAGAGCTACCAGGAAATGCTGGTCGACTTTAAAATCCACCAGCCACAG
CTGGTGAGCAACGGCAAGAAAATTCCCCATGACTGGACAATCTACTATCTGAGAAAGAAAGCTCTGTCTAAGAAAATTGAGAAG
GAGGAACTGGCATGGATCCTGCTGAACTTCAATCAGAAACGGGGATACTATCAGCTGAGAGGCGAGGAAGAGGAAGAGACACCA
AATAAGCTGGTGGAGTTTCATAGCCTGAAAATTGTGGAGGTCAACGCCGATGAACCCCAGAAAGGAAAGCCTGAGATCTGGTAC
TCACTGGTGCTGGAAAACGGCTGGGTCTATCGACGGGCTAGCAAGACTCCCCTGTTCGATTGGAAAGACAAGATTAGGGAGTTT
ATCGTGACCACAGAAATCAACAATGATGGAACTGTCAAAACCGACAAGGAGGGCACCGAAAAGAGGAGCTTCCGCGCACCAAAA
CCCGAGGACTGGACTCTGCAGAAGAAAAAGACCGAGTTCGAGCTGTCAAAGAGCGGGAATGAAGTGGGAGCCTTCATCTACGAA
AGTATCCTGCAGAAACCCAACCAGAAGATTAGAGGCAAACTGATTTCAACCATCGAGAGGAAATTTTATAAGGAAGAGCTGAAA
ACCATCCTGAAAACACAGCTGTTCTTTCACAAAGAACTGAAGGATGAGAAACTGTACAATGCCTGCATCGAAGAGCTGTATAAG
AACAATGAAGCTCACCGGAGCCTGCTGTCCAACAAGGGGTTCGAGCATCTGTTTATTAACGACATCCTGTTCTACCAGCGACCT
CTGCGGTCTAAAAAGAGTCAGATCTCAAACTGCCCACTGGAGAAGCGCACATATAAAAAGGAGGGGATTGAAATCACTGAGGGC
ATCAAAGTGATCTCCAAATCTAATCCAATCTACCAGGAGTTCCGGCTGTGGCAGTGGATTAGCAACCTGTCCCTGTATTGTATC
GAACCCACCGAGACAAATGTGACTTCAACCTTTCTGAACAGCATTGAAGATTACGAGAATCTGTTCGAATTTCTGAACAATCGC
AAGGAAATCGAGCAGAAGCACCTGCTGAAATATCTGCTGGAGAACCAGGGGTTTAAAGGAAAGCTGCTGACAAACGAACTGGAG
AAGTTCCGCTGGAATTTTGTCGCTGACAAAAAGTACCCCTGTAATGAGACAGGCAGCCTGCTGCATACTCGGCTGAGCAAAGTG
AAGAACATTTCCCCTGATTTCCTGACCAAGGAAATCGAGCACCAGCTGTGGCATATCATCTACAGCGTGACCGACAAGATTGAA
TATGAGCAGGCCCTGAAAACCTTTGCTCGGAAAAACAATCTGGATGTGGACTCCTTCTTTGAGCACTTCAAAAAGATCCCCCCT
TTTGAGTCTACCTACGGAGCATATAGTGAAAAGGCCATCAAGAAGCTGCTGCCACTGATCAGACTGGGCAAATACTGGAACTGG
GAGGCCATTGATAGTATCTCAAAGGACAGGATTAGTAAAATCCTGTCAGGGGAATACGATGAGAACATTAAGAACAGAGTGAGG
GAGAAAGCAGTCCACCTGACCTCCGAAAACAATTTCCAGGGACTGCAGGAGTGGCTGGCCAAGTACATCGTCTATGATCGCCAT
TCTGAGGGCAATGACCTGGGGAAGTGGACTAGCGTGTCCGACCTGGAGACATACCTGAAGGAGTTCAAGCAGCATAGCCTGCGG
AACCCTATTGTGGAGCAGGTCATCACAGAAACTCTGAGAGTGGTCAAGGATATTTGGATCAAGCACGGGAAAGGAACCGAAAAT
TTCTTTGACGAAATCCATGTGGAGCTGGGCCGGGAAATGAAGAACAATTCCGAGGATCGCAAACGACTGACCAACACAATTACT
GAAAACGAGAATACAAACCTGAGAATCAAGGCCCTGCTGATGGAAATGATGAATGATAACGACGTGGAGAACGTCAGGCCTTAC
TCTCCAAGTCAGCAGGAGATTCTGAAGATCTATGAGGACGGAGCTCTGAATAGCAACATCGAGCTGGACGATGAAATTGTGAAG
ATCTCCAAAAAGGCAGAGCCCACCAAATCTGAACTGCAGCGCTACAAGCTGTGGCTGGAGCAGAAATACCGATCCCCTTATACT
GGCCAGGTCATCCCACTGAACAAGCTGTTCACCTCTGAATATGAGATCGAACACGTGGTCCCTCAGAGTCGCTTCTTTGACGAT
AGCTTCAGCAACAAAGTGATCTGCGAGTCAGCCGTCAACAAGCGGAAGGATAACCAGCTGGGGCTGCAGTTCATCAAGAACCAT
AGCGGAGAAAAAGTGGAGCTGGGCTTCGGGAAGGTGGTCCAGGTCTTTACAGAAGAGCAGTACCTGGATTTTGTGAAGGAGCAC
TATAGCAAAAATCGCTCCAAGCATAACAAACTGCTGCTGGAAGAGATTCCCGAGAAGATGATCGAAAGGCAGCTGAATGACACT
CGCTACATCAGTAAGTTCGTGAGCTCCATTCTGTCCAACATCGTCAGATCTGAGAAGGACGATGACGGCCTGAATAGCAAAAAC
ATTCTGCCTGGAAATGGCAAGATCACTACCGAACTGAAAAGGGACTGGGGGCTGAATGATGTGTGGAACGACCTGATTCTGCCA
AGATTCGAGAGGATGAATACCATCACAAACAGCGATCTGTTTACAACTTACAACGACAAGTATCAGAAACACCTGCCCACCGTG
CCTTTCGAGTACTCCAAGGGCTTTCAGAAAAAGCGCATTGATCACCGACACCATGCTATGGACGCACTGGTCATCGCATGTGCC
ACACGGGATCATCTGAATCTGATGAACAATCAGTCTGCCAAGAGTGAACTGAAACGATACGACCTGCAGAACAAGCTGCGGAAA
AAGGAGCCTTACTTCAACCAGAAGGAGAACAAACAGAAGGAAGCCTTCAAGGATTTTATCAAACCATGGGGCACTTTCACCGAG
GACAGCAAGAATGCTCTGGAAAAAATCATTATCTCCTTTAAGCAGAACCTGAGAGTGATCAACAAGGCAACAAACTCATACGAG
AGCTATAAGGATGAGGACGGGAATCTGAGGATCGGCAAGGATGGGAAACCAGAGAAGGGCCTGATCAAGCAGAAGGGGCTGAAC
TGGGCAATCAGAAAGCCCCTGCACAAAGACACCGTGTCAGGCCAGATTAACCTGAGCAGGATCAAGCTGCCCAATGGGAAAATC
CTGACAGCCACTCGCAAGAATCTGGATACCAGTTTCGACCTGAAAACAATTGAGAACTCAATCACCGACACAGGCATTCAGAAA
ATCCTGAAGAATTACCTGCTGTCCAAGGAATCTCCAGAGCTGGCCTTCTCTCCCGAGGGCCTGGAAGAGATGAACAATGAGATC
GAAAAGTACAACAATGGGAAATTTCACCATCCTATTAACAAGGCTAGGATCTATGAACTGGGGAGCAAATTCAATGTCGGACAG
ACAGGCAACAGAAGGGATAAGTTTGTGGAGGCCGCTAAAGGAACTAATCTGTTCTTTGCTATCTACCAGGACGAGAATAAGAAC
CGCTCTTATGAAACTATTCCCCTGAACGAAGTGATCGAGCACCAGAAGTGGCGAGCAAGTCTGCCTAAGGAAGAGCAGGAGAAA
ATTCCACTGGTGCCCGTCAACAAGCTGAAGGGGACCTTCATCTTTTCCCTGTCTCCCAACGATCTGGTGTACGTCCCTTCCAAT
GACGAGCTGGAACGAAGTGCTTCAATTGACTTCTCTAAGCTGAAAAAGGAACAGATCAACCGGCTGTATAAAATGGTGTCTAGT
TCAGGATCCCAGTGCTTCTTTGTGAAGAGTGAGGTCGCAACCTCAGTGGTCAACAAAATGGAATACAGCAGCCTGAACAAGATG

FIG. 9I

GAGAAGTCTATCGATAACCTGATGGTGAAGGAGATTTGTATCAAACTGAAGATTGACAGGCTGGGCAACATCAGCAAGGCCTAA
GAATTC

#14
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACCAAA
ATCCTGGGGCTGGACCTGGGGACTAATAGCATCGGGTGGGCAATCCGCGACACAGAAAATGAGGGCATCAATCAGATCCTGGAC
AAGGGCGCCCGCATTTTCAGCGAGGGAGTGAAGTCCGAGAATGGCAAAGAAATCAGTAGGGCAGCTGAACGAACCGCTTACCGG
AGCGCAAGAAAGATCAAATATCGGAGAAAACTGCGGAAGTACGAGACACTGAAGGTGCTGTCAATTAACGGAATGTGCCCCCTG
AGCGTGGAGGAAGTCGAACAGTGGAAGAAATCTGGCTTCAAGGAATATCCCCTGAATCCTGAGTTTCTGGATTGGCTGCGGACC
AACGAGGACAAGAACATCAATCCTTACCTGTTCAGAGATAAAGCTAGCAAGCAGAAAGTGACACTGTTTGAACTGGGAAGAGCT
CTGTATCACATCGCACAGAGGCGCGGCTTCCTGAGTAACAGGCTGGACCAGTCAGCCGAGGGCGTCTTTGAGGAACATAATCCT
CAGATCCAGAACCTGATTGAGGACCTGGACAGCTCCAACACAATTCTGAATGAGCTGAAGGAATACTATATCAATCTGGGGATC
ATTGACGAGACTGAAAAGTCCGGCTTCAAGAAAGACCTGGATGAGGGGGAAAAGAAACTGAAGTCACTGTACAACAGCCTGGTG
GCCATCACAAAGAAAAACGCTAACGATATCGAAACTTGCAAGCAGGAGCTGATCGCTAGACTGAATAAGAAAGAGGACCTGGGC
AAGGTCAAAGGGAAGATCAAAGATATTTCTCAGGCAATGCTGGACAGGAACTTTAAGACTCTGGGACAGTATTTCTTTTCACTG
TACAACAAGGAGCGAATCCGGAACCAGTATACCAGCCGCGAGGAACACTACCTGGAGGAGTTCATCATTATCTGTCAGACCCAG
GGGATCGAGGGAATTAACACAAATGAGAAGCTGCCTGAGAAGAAGTTCACCGGCCTGGCAAAAGATCTGTATCGAGCCATTTTC
TTTCAGCGGCCACTGAAATCCCAGAAGGGCCTGATCGGGAAGTGCTCTTTCGAGAAGAACAAGTCTCGCTGTGCCGTGAGTCAC
CCAGACTATGAGGAATTTCGAATGTGGAGCTACCTGAACACAATCAAAATTGGCACTCAGTCCGAGAAAACCCTGCGCTTCCTG
ACACTGGAGGAAAAGCTGAAACTGGTGCCCAAGTTCTACCGAAAGAGTGATTTCAACTTCGAGGTGCTGGCTAAGGAGCTGGTC
GAAAAAGGAGCATCATTCGGCTACTACAAGTCTAGTAAGAAAAATGAGTTCTTTTACTGGTTCAACTATAAGCCCACCGATTCA
GTGAGCGCCTGCGTGGTGAGCGCTTCTCTGGAGAACGCAATCGGCAAGGACTGGAAGATCAAGACTTTCGAATATCAGACTAGA
AACACCGAGAAGAATGAAGTGACCAAATCCGTCGATTACAAGGACCTGTGGCACCTGCTGTCCGTGGCAACATCTGATACTTAC
CTGTATGACTTTGCCATCGAAAAGCTGAAACTGGAGCCTAAAAACGCAAAGGCCTTCAGCAAGACAAAACTGAAGAAAGACTTT
GCCAGTCTGTCACTGGCAGCCATCGCTAAAATTCTGCCATATCTGAAGCAGGGCCTGCTGTACTCCCACGCCGTGTTCATGGCT
AACATCGAGAATATTGTCGACGCCGATATCTGGAAGGATGAGGAACAGCAGAAGTTCATCCAGTCCAAGATTGTGGAGCTGGTC
GACAATTATATCGTGGAAAAGTCTAAACTGGAGCTGATCAATGGGCTGCTGAAGATCTACAACACCGAGGATAAGGAAGGACGG
AAAGTGTACTATTCAAAGGAGGCTGAAAGCCTGTTCGAGGCAGACCTGAGAAAGAAACTGGTCCCCTTCTACAAGGCTAACATC
ATCATCGAGGAACACGAGCAGGAAATCATTTTCCAGGATCTGTTTCCTATCCTGATGGACCAGCTGAAGAAACAGGAGTTCATC
AAAATTGGCAGACTGGATAAGCAGATTGAAGCCTTCCTGGAGGGGGAAAATGAGGAAGGACAGATCTTTTGTAACCACACAGAT
AAGCTGAAGAAACTGTACCATCCAAGCGACATCGAGGTGTTTAAGAAAAAGACTATCAAAGATGAGTGGGGGAATGAAAAGGTG
GTCCTGGGATCCCCACTGACCACATCTATCAAGAACCCCATGGCAATGAGAGCCCTGCACCAGCTGAGGAAGGTGCTGAATACA
CTGATTGCCAACGACCAGATCGACGAGGATACTCGGATCCATATTGAGATGGCCAGAGAACTGAACGATGCTAATAAGCGAAAA
GGCATCCAGGACTTCCAGAACGAGAACAAGAAGTTTCGGGAGGAAGCCATCAAGGAGATCAAGAAGCTGTACCTGGAGGAATGC
CACAAAGACGTGAACCCCACCGAGGACGATATCCTGAGGTATCAGCTGTGGCTGGAACAGGGAAAGTGCGAGATCTACGAGGAA
GGCAACAATATCAGCATTTGTGATATCATTGGCAGCAATCCCTCCTATGACATCGAGCATACCATTCCTCGGAGCATCTCCCAG
GATAACAGCCAGATGAACAAGACACTGTGTAGCCTGAAGTTCAACAGATCCATCAAAAAGCAGAAGATGCCAGTGGAGCTGTCC
AACTACAATGACATTCTGCCCAGGATCGCACACTGGAAAAAGGAGGCCGAGGAACTGACTAGGCAGATCGAAACCATTTCTCGC
AGCATCAAGAGCGCTGCAACCAAGGTGGCCAAGGATAAGAACATCAGAAAGAGGCATTACCTGACACTGAAGCGAGACTATATT
CTGGGGAAATACGAGCGGTTCACTTGGGAGGAACCTAAAGTGGGCTTTAAAAACTCCCAGATCCCAGACACTGGGATCATTACC
AAGTACGCTCAGGCATATCTGAAATCTTACTTCAAAAGGGTGGAGAGTGTCAAGGGAGGAGCAGTGGCTGAGTTCCGGAAACTG
TGGGGCATCCAGGAATCTTTTATCGATGAGAACTGGTGGAAGCACTATAAGGACAAAGATAGAGACAAACATACCCACCATACA
ATCGACGCAATCACTATTGCCTGTATGCCCAAGGATAAATACGACCTGCTGGCACACGCCTGGAGGCTGGAGGATGAACAGGAC
AAAAAGGCCGCTAAGGTGCTGATCGAGCAGGCCAAACCATGGAAAACCTTCAAGGAGGATATCGAAAAGATTGAGACTGAAATC
CTGGTGAGCCATTTTACCCCCGACAACGTCAAAAAGCAGTCAAAGAGCATCATCAAGAATCGCGGCAAAAAGGTGTACGTCCTG
AAGAACGAGCTGCCTGTGAACTTCAAGAACAAGATCGAAGGGAAGGATTATTTCAAGCTGAAATTTGACAGCAAGATTCTGTAC
AAAATCCCCAAAAAGAAAGAGAAGCAGACCGATACATTCTATGAGGAACTGCCTAAAAACTACCTGAATGGAGTGGAAGGCAAG
GACTACTTCAAAATCAATACTACCGGCAAGACCTTCTACAAAATCCCAATTTTTAACCAGGGCGACACAATCCGGGGGAGCCTG
CACCAGGAGACAACTTACGGAGCCATTAAGCTGCCAGATATCGACATTGAAACAAAGAAACCCCTGCATACTGATAAGGGAGGC
TTCATTCTGAAGAAAGACATCAAAGGCAATGAGATTGTGTTCTTTGTGGTCCGCAAGGAAATCTCTAAAATTAGTGAGAACGAT
GTGCAGAATATCGTCGACAACGTGGTCCGGAAGAAAATTGAGAATGCTATCGCAAACTCCCTGATCACTTTTAAGATCGTGAAG
AAAAAGAAAGTGGCCGTCATCAAAGAGGGGGTCACCATTTGGATGAGGGAGCCCAACATTGAAAAGGGGATCGAGGGAATCCCT
ATTAAGAAAGTGCGCATCATTACCAATTCTGTCAAAAACCCTATCGAGATTAAGGTGCACAGTCCACTGTCCAAGTCTCGCCAC
AAGCATAAACAGAAGGTCTACGGCCAGAACGATGAGAATTATGCCATGGCTCTGTACGAACTGGACGGGAAGCGGGAGTTCGAA
CTGATCAACAATTTTAATCTGGCCAAGCTGCTGAAACAGAGTCAGTCATACTATCCACTGCATAAAGAGAAGGAAATCAAGGGA
AAGAAAATTCTGGTGCCCATCGAGAAGAGAAACAACAAGGATGTCATCCTGAAGAGGGGCCAGCAGGTGGTGTTCTACGACAAA
ACCGTGGAGAACCCCAAGGATATCTCTGAAATCATTGACTTTCGCGAGCGAATCTACATCATTGAAGGGCTGACCATCCAGAGA
CAGAAAGATAAGAAAACATCCAAAGTGAATGAGTACGGAATCATTCAGCTGCGCCACTTCAAGGAAGCTCGAAAAAGTGAGGAA
ATCTCAAAGGATAACTTCAAACCTGACGGCGAGTTCAAGATCAACGAGAATAAGCCAACTAGGAAAATGAACCATAATCAGTTC
ACCGCCTTTGTGGAGGGGATCGACTTCAAGGTCCTGCCTAGCGGAAAGTTTCAGAAAATTTAAGAATTC

#15
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAGCAAG

FIG. 9J

AAAGTCAGTAGACGCTATGAAGAACAGGCACAGGAGATTTGTCAGAGGCTGGGGAGTAGACCTTATTCCATCGGGCTGGACCTG
GGAGTGGGATCTATCGGAGTGGCAGTCGCCGCTTACGACCCAATTAAGAAACAGCCCTCCGATCTGGTGTTCGTCAGCTCCAGG
ATTTTTATCCCTTCTACCGGCGCAGCCGAGCGGAGACAGAAGAGAGGACAGAGGAACAGCCTGCGCCACCGAGCAAATCGCCTG
AAATTCCTGTGGAAGCTGCTGGCTGAACGAAACCTGATGCTGTCATATAGCGAGCAGGACGTGCCAGATCCTGCACGGCTGAGA
TTTGAGGACGCTGTGGTCCGAGCAAACCCATACGAGCTGCGGCTGAAGGGCCTGAATGAACAGCTGACCCTGAGCGAGCTGGGG
TATGCACTGTACCACATCGCCAATCATAGGGGATCTAGTTCAGTGCGCACATTCCTGGATGAGGAAAAGAGCTCCGACGATAAG
AAACTGGAGGAACAGCAGGCTATGACAGAACAGCTGGCAAAAGAGAAGGGAATTTCCACTTTCATCGAAGTGCTGACAGCCTTT
AACACTAATGGCCTGATCGGGTACAGGAACTCCGAGTCTGTGAAGAGTAAGGGCGTGCCAGTCCCCACTCGCGACATCATTTCA
AATGAGATTGATGTGCTGCTGCAGACCCAGAAGCAGTTCTATCAGGAAATCCTGTCAGACGAGTACTGCGATCGGATTGTCAGC
GCAATCCTGTTTGAAAACGAGAAGATCGTGCCAGAAGCCGGCTGCTGTCCCTATTTCCCTGACGAGAAGAAACTGCCCAGATGT
CACTTTCTGAATGAGGAAAGGCGCCTGTGGGAAGCCATTAACAATGCTAGGATCAAGATGCCCATGCAGGAGGGCGCTGCAAAA
CGCTACCAGAGTGCTTCATTCAGCGACGAGCAGAGACACATTCTGTTTCATATCGCAAGGAGCGGGACTGATATCACCCCTAAA
CTGGTGCAGAAGGAGTTCCCAGCCCTGAAAACCTCCATCATTGTGCTGCAGGGAAAAGAGAAGGCTATTCAGAAGATCGCAGGC
TTCCGATTTCGACGGCTGGAGGAAAAATCTTTTTGGAAGAGACTGAGTGAGGAACAGAAGGACGATTTCTTTAGCGCCTGGACA
AACACTCCTGACGATAAAAGACTGTCCAAGTACCTGATGAAACACCTGCTGCTGACAGAAAATGAGGTGGTCGACGCCCTGAAA
ACCGTGAGCCTGATTGGAGATTATGGCCCAATCGGCAAGACCGCAACACAGCTGCTGATGAAACATCTGGAGGATGGCCTGACT
TACACCGAAGCCCTGGAGCGGGGAATGGAAACCGGCGAGTTCCAGGAACTGTCAGTGTGGGAGCAGCAGAGCCTGCTGCCCTAC
TATGGGCAGATCCTGACAGGATCTACTCAGGCCCTGATGGGGAAGTATTGGCACTCTGCTTTTAAAGAAAAGAGAGACAGTGAG
GGATTCTTTAAGCCTAACACAAATAGCGATGAGGAAAAATACGGCAGGATTGCCAACCCAGTGGTCCATCAGACTCTGAACGAA
CTGCGCAAGCTGATGAATGAGCTGATTACCATCCTGGGAGCTAAACCTCAGGAGATCACAGTGGAACTGGCACGAGAGCTGAAG
GTCGGAGCTGAGAAAAGAGAGGACATCATTAAGCAGCAGACCAAACAGGAAAAGGAGGCTGTGCTGGCATATAGCAAGTACTGC
GAGCCCAACAATCTGGACAAAAGGTATATTGAAAGGTTCCGCCTGCTGGAGGATCAGGCCTTTGTGTGCCCTTACTGTCTGGAG
CACATTAGCGTCGCAGATATCGCAGCTGGAAGGGCAGACGTGGATCATATCTTCCCACGCGACGATACAGCTGACAACTCCTAC
GGGAATAAGGTGGTCGCACACCGACAGTGTAACGATATCAAGGGAAAGCGGACCCCCTATGCAGCCTTCAGTAATACATCAGCC
TGGGGCCCCATCATGCATTACCTGGACGAAACCCCTGGGATGTGGCGCAAAAGAAGGAAGTTTGAGACAAACGAGGAAGAGTAT
GCTAAGTACCTGCAGTCAAAAGGCTTCGTGAGCAGGTTTGAAAGCGATAACTCCTATATCGCAAAAGCTGCAAAGGAGTACCTG
CGCTGCCTGTTCAATCCAAACAATGTGACTGCCGTCGGGTCCCTGAAGGGAATGGAGACATCTATCCTGCGGAAGGCCTGGAAT
CTGCAGGGAATTGACGATCTGCTGGGCAGCCGGCACTGGAGTAAGGACGCCGATACCAGCCCCACAATGCGCAAAAACCGGGAC
GACAATCGGCACCATGGCCTGGACGCCATCGTGGCTCTGTATTGTTCCAGATCTCTGGTCCAGATGATTAACACCATGTCCGAG
CAGGGCAAGCGAGCAGTGGAAATCGAGGCTATGATTCCTATCCCAGGGTACGCATCCGAACCAAATCTGTCTTTCGAAGCCCAG
CGGGAGCTGTTTAGAAAGAAAATCCTGGAGTTCATGGACCTGCACGCCTTTGTGAGTATGAAAACCGACAACGATGCAAATGGC
GCCCTGCTGAAAGATACTGTGTATTCAATTCTGGGAGCAGACACCCAGGGAGAGGATCTGGTGTTCGTGGTCAAGAAAAAGATT
AAGGACATCGGCGTGAAAATCGGGGATTATGAAGAGGTCGCATCTGCCATTCGAGGCCGGATCACCGACAAACAGCCAAAGTGG
TATCCCATGGAAATGAAAGATAAGATCGAGCAGCTGCAGTCTAAGAACGAAGCCGCTCTGCAGAAATACAAGGAGAGTCTGGTG
CAGGCAGCCGCTGTCCTGGAAGAGAGTAATAGGAAGCTGATTGAGTCAGGCAAAAAGCCCATCCAGCTGAGTGAAAAAACAATT
TCAAAAAAGGCCCTGGAGCTGGTGGGCGGGTACTATTACCTGATTAGCAACAACAAGCGCACAAAGACTTTCGTGGTCAAGGAA
CCTTCAAACGAGGTGAAAGGGTTCGCATTTGACACTGGAAGCAATCTGTGCCTGGACTTTTATCACGATGCCCAGGGAAAGCTG
TGTGGCGAGATCATTAGAAAAATCCAGGCTATGAACCCTTCCTATAAGCCAGCATACATGAAACAGGGGTATTCTCTGTACGTG
AGACTGTACCAGGGCGACGTGTGCGAGCTGAGGGCAAGCGATCTGACTGAAGCAGAGTCCAACCTGGCCAAGACCACACATGTC
CGCCTGCCCAATGCTAAACCTGGGCGAACCTTCGTGATCATTATCACCTTTACAGAGATGGGGTCTGGATATCAGATCTACTTC
AGCAACCTGGCCAAGTCCAAAAAGGGACAGGACACTAGTTTTACCCTGACTACCATCAAGAATTATGATGTGCGGAAAGTCCAG
CTGTCTAGTGCCGGGCTGGTGAGATACGTCAGCCCTCTGCTGGTGGACAAAATCGAGAAGGATGAAGTCGCTCTGTGTGGAGAG
TAAGAATTC

#16

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGGCAAGA
CCTGCATTTCGGGCACCTCGGAGAGAACACGTCAACGGCTGGACCCCTGACCCACATCGGATTAGCAAACCCTTTTTCATCCTG
GTGAGCTGGCACCTGCTGTCCCGGGTGGTCATTGACAGCTCCTCTGGATGCTTCCCAGGCACCAGCCGGGACCACACCGACAAG
TTTGCCGAGTGGGAATGTGCTGTGCAGCCCTACAGGCTGAGTTTCGACCTGGGGACCAACTCAATCGGATGGGGCCTGCTGAAT
CTGGATCGCCAGGGAAAACCAAGGGAGATCCGAGCACTGGGGTCCCGCATCTTCAGCGACGGACGGGATCCCCAGGACAAGGCT
TCTCTGGCTGTGGCACGGAGACTGGCCAGACAGATGAGGCGCCGACGGGACAGATATCTGACTAGAAGGACCAGGCTGATGGGA
GCTCTGGTGCGCTTCGGCCTGATGCCAGCAGACCCTGCAGCTAGGAAGCGCCTGGAAGTGGCCGTCGATCCATACCTGGCACGA
GAGCGGGCCACAAGAGAAAGGCTGGAGCCCTTCGAAATCGGGAGGGCCCTGTTTCACCTGAACCAGCGCCGAGGATATAAACCC
GTGCGCACCGCCACAAAGCCTGATGAGGAAGCCGGCAAGGTGAAAGAGGCTGTCGAAAGGCTGGAGGCAGCAATCGCTGCAGCC
GGAGCACCTACTCTGGGAGCTTGGTTCGCATGGCGAAAAACACGAGGAGAAACTCTGCGAGCACGACTGGCTGGGAAGGGAAAA
GAGGCTGCATACCCATTCTATCCCGCACGGAGAATGCTGGAGGCCGAATTTGACACTCTGTGGGCAGAGCAGGCCAGGCACCAT
CCAGATCTGCTGACCGCCGAAGCTCGCGAGATCCTGCGGCACAGAATTTTTCATCAGCGGCCCCTGAAGCCACCTCCAGTGGGA
AGATGCACTCTGTACCCTGACGATGGGAGAGCTCCTAGGGCACTGCCAAGCGCTCAGAGGCTGCGCCTGTTCCAGGAGCTGGCC
AGCCTGAGAGTGATCCACCTGGACCTGTCCGAACGCCCTCTGACCCCAGCTGAGCGAGATCGGATTGTGGCATTTGTCCAGGGC
AGACCCCCTAAAGCCGGAAGGAAGCCTGGCAAAGTGCAGAAGAGCGTCCCATTCGAAAAGCTGAGGGGGCTGCTGGAGCTGCCA
CCAGGCACTGGGTTTTCTCTGGAGAGTGACAAACGGCCTGAACTGCTGGGCGACGAGACAGGCGCCAGAATCGCACCAGCATTC
GGACCTGGATGGACAGCTCTGCCTCTGGAGGAACAGGACGCCCTGGTGGAACTGCTGCTGACAGAGGCAGAACCAGAGAGGGCA
ATTGCAGCTCTGACTGCACGATGGGCTCTGGACGAGGCAACTGCAGCAAAGCTGGCTGGAGCAACCCTGCCAGATTTTCACGGA

FIG. 9K

CGATATGGCAGGCGCGCAGTGGCTGAACTGCTGCCTGTCCTGGAACGCGAGACACGAGGCGACCCAGATGGGAGAGTGAGGCCC
ATCCGGCTGGACGAGGCAGTGAAACTGCTGAGAGGCGGGAAGGATCACTCAGACTTCAGCCGGGAAGGAGCTCTGCTGGACGCA
CTGCCCTACTATGGAGCCGTGCTGGAGAGACATGTCGCTTTTGGGACAGGAAACCCCGCAGATCCTGAGGAAAAGCGGGTGGGA
AGAGTCGCCAATCCCACTGTGCACATCGCTCTGAACCAGCTGCGGCATCTGGTGAATGCAATTCTGGCCAGGCACGGCCGCCCT
GAGGAAATCGTGATTGAGCTGGCACGGGACCTGAAAAGGTCTGCCGAAGATCGACGGAGAGAGGACAAGCGGCAGGCCGATAAC
CAGAAAAGAAATGAGGAACGCAAGCGACTGATCCTGAGTCTGGGAGAGCGCCCAACCCCACGAAACCTGCTGAAGCTGCGGCTG
TGGGAGGAACAGGGCCCAGTGGAAAATAGGCGCTGCCCCTACTCTGGGGAGACAATTAGTATGAGAATGCTGCTGAGCGAGCAG
GTGGACATCGATCACATTCTGCCATTCAGCGTGTCCCTGGACGATTCCGCTGCAAACAAGGTGGTCTGTCTGCGGGAGGCCAAC
AGAATCAAGCGGAATAGATCTCCCTGGGAGGCCTTCGGCCATGACAGTGAGAGATGGGCAGGGATTCTGGCACGAGCAGAAGCT
CTGCCCAAGAACAAAAGGTGGCGCTTTGCTCCTGACGCACTGGAGAAACTGGAAGGAGAGGGAGGCCTGCGAGCACGACACCTG
AATGATACAAGGCATCTGAGTCGCCTGGCCGTGGAGTATCTGCGGTGCGTCTGTCCTAAGGTGCGGGTGAGCCCAGGCCGACTG
ACTGCACTGCTGCGACGGAGATGGGGCATCGACGCCATTCTGGCAGAAGCAGATGGACCTCCACCAGAAGTGCCCGCAGAGACA
CTGGACCCTTCCCCAGCTGAGAAGAACCGAGCAGACCACCGGCACCATGCCCTGGATGCTGTGGTCATCGGCTGTATTGATCGC
TCAATGGTGCAGCGAGTCCAGCTGGCCGCTGCAAGCGCAGAAAGAGAGGCCGCTGCAAGGGAGGACAATATCAGGCGCGTGCTG
GAGGGATTCAAAGAGGAACCTTGGGATGGCTTTAGAGCTGAACTGGAGCGACGGGCACGGACCATCGTGGTGAGCCACAGACCA
GAACATGGGATTGGGGGAGCCCTGCATAAGGAGACAGCTTACGGGCCTGTGGACCCTCCAGAGGAAGGATTCAACCTGGTGGTC
AGGAAACCAATCGACGGCCTGTCAAAGGATGAGATTAATAGCGTGCGGGACCCCCGGCTGAGAAGGGCACTGATCGATCGCCTG
GCCATTCGCCGACGGGATGCTAACGACCCTGCTACCGCACTGGCCAAAGCAGCTGAGGATCTGGCAGCACAGCCAGCCTCCCGC
GGCATCAGAAGGGTGCGGGTCCTGAAGAAAGAATCTAACCCCATTAGGGTGGAGCACGGCGGGAATCCAAGTGGACCCCGCTCA
GGAGGCCCTTTTCATAAGCTGCTGCTGGCAGGAGAGGTGCACCATGTGGACGTCGCACTGCGAGCAGATGGCCGCCGATGGGTG
GGACACTGGGTCACACTGTTCGAGGCACATGGGGGACGGGGAGCAGACGGAGCTGCAGCCCCACCTAGACTGGGCGATGGGGAA
AGATTTCTGATGAGGCTGCACAAGGGAGACTGCCTGAAACTGGAGCATAAGGGCAGAGTGAGGGTCATGCAGGTGGTCAAACTG
GAACCTAGTTCAAATAGCGTGGTCGTGGTCGAGCCACACCAGGTGAAAACCGACAGATCCAAACATGTCAAGATCTCTTGTGAT
CAGCTGCGCGCTCGAGGAGCACGGAGAGTGACCGTCGATCCACTGGGACGGGTGAGAGTCCACGCCCCAGGAGCTAGGGTGGGA
ATCGGAGGGGACGCCGGACGAACCGCTATGGAACCCGCAGAGGATATTAGCTAAGAATTC

#17
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGGGAGAG
AATATGATTGACGAGAGTCTGACCTTCGGCATTGACCTGGGGATTGGGAGTTGTGGGTGGGCAGTGCTGAGGCGGCCAAGCGCC
TTCGGACGGAAGGGCGTGATCGAGGGAATGGGCTCCTGGTGCTTTGACGTCCCCGAGACATCTAAAGAACGGACTCCTACCAAC
CAGATCCGGAGATCCAATAGACTGCTGAGGCGCGTGATCCGACGGAGAAGGAACCGGATGGCCGCTATTCGCCGACTGCTGCAC
GCAGCAGGCCTGCTGCCATCAACCGACAGCGATGCCCTGAAGCGGCCCGGACATGATCCTTGGGAACTGAGGGCACGAGGCCTG
GACAAGCCACTGAAACCCGTGGAGTTCGCTGTGGTCCTGGGCCACATCGCAAAGCGGAGAGGGTTTAAATCCGCTGCAAAGAGA
AAAGCCACAAACATTAGCTCCGACGATAAGAAAATGCTGACAGCCCTGGAGGCTACTCGAGAACGGCTGGGGAGATACAGGACC
GTGGGAGAAATGTTCGCCAGGGACCCTGATTTTGCTTCTAGGCGCCGAAATCGCGAGGGCAAGTATGATAGGACCACAGCTCGC
GACGATCTGGAGCACGAAGTGCACGCCCTGTTCGCAGCTCAGCGGAGACTGGGACAGGGATTTGCCAGTCCAGAACTGGAGGAG
GCCTTCACCGCTTCAGCATTTCATCAGAGGCCCATGCAGGACAGCGAGCGCCTGGTGGGATTCTGCCCTTTTGAGCGAACCGAA
AAGCGGGCAGCCAAACTGACACCCTCTTTCGAGCGCTTTCGACTGCTGGCCCGGCTGCTGAACCTGAGAATCACTACCCCAGAC
GGAGAGCGGCCCCTGACAGTGGATGAAATTGCTCTGGTCACCCGGGACCTGGGCAAGACCGCAAAACTGAGTATCAAGCGGGTG
AGAACTCTGATTGGACTGGAGGACAATCAGAGGTTCACAACTATCCGCCCCGAGGACGAAGATCGAGACATTGTGGCTCGGACA
GGCGGGGCAATGACAGGGACTGCCACCCTGAGGAAGGCACTGGGAGAGGCCCTGTGGACTGATATGCAGGAGCGCCCTGAACAG
CTGGACGCTATCGTGCAGGTCCTGAGCTTCTTTGAGGCCAACGAAACAATCACTGAGAAGCTGAGGGAAATTGGCCTGACTCTG
GCCGTGCTGGACGTCCTGCTGACCGCACTGGATGCCGGAGTGTTCGCCAAGTTTAAAGGCGCTGCACACATCAGCACCAAAGCC
GCTAGGAATCTGCTGCCACATCTGGAGCAGGGCAGGCGCTACGATGAGGCCTGCACAATGGCAGGGTATGACCACGCAGCCTCC
CGCCTGTCTCACCATGGCCAGATCGTGGCAAAGACACAGTTCAACGCCCTGGTCACTGAGATCGGCGAATCCATTGCCAATCCA
ATCGCTCGGAAGGCACTGATCGAGGGGCTGAAACAGATTTGGGCCATGAGAAACCACTGGGGGCTGCCCGGAAGTATCCATGTG
GAGCTGGCCCGGGATGTCGGCAACTCAATTGAAAAGCGACGGGAGATTGAAAAGCACATCGAGAAAAATACTGCCCTGAGGGCT
CGCGAGAGAAGGGAGGTGCATGATCTGCTGGACCTGGAAGATGTCAATGGCGACACCCTGCTGCGATACCGGCTGTGGAAGGAG
CAGGGAGGCAAATGCCTGTATACAGGGAAGGCCATCCACATTCGGCAGATCGCTGCAACTGACAACTCCGTGCAGGTCGATCAT
ATTCTGCCTTGGAGCCGGTTCGGCGACGATAGTTTTAACAACAAGACCCTGTGTCTGGCCTCTGCTAATCAGCAGAAGAAAAGG
TCAACACCATACGAGTGGCTGAGCGGCCAGACTGGGGATGCATGGAACGCCTTCGTGCAGCGCATCGAGACAAATAAGGAACTG
AGAGGGTTTAAGAAAAGGAACTATCTGCTGAAGAATGCTAAAGAGGCAGAGGAAAAATTCAGAAGCAGGAACCTGAATGACACC
AGATACGCCGCTAGGCTGTTCGCAGAGGCCGTGAAGCTGCTGTATGCCTTTGGGGAGAGACAGGAAAAAGGGGGAAACCGCCGA
GTGTTTACTCGGCCTGGAGCACTGACCGCAGCACTGAGACAGGCTTGGGGAGTGGAGAGCCTGAAGAAACAGGATGGGAAGCGC
ATCAATGACGATCGACACCATGCCCTGGATGCTCTGACCGTGGCTGCAGTCGACGAGGCCGAAATTCAGAGGCTGACAAAATCA
TTCCACGAGTGGGAACAGCAGGGCCTGGGGCGGCCTCTGCGGAGAGTGGAGCCACCTTGGGAGAGCTTCCGGGCAGACGTCGAG
GCTACCTACCCTGAAGTGTTTGTCGCACGGCCAGAGAGGCGCCGAGCAAGAGGAGAAGGCCATGCCGCTACCATCCGGCAGGTG
AAGGAGAGAGAATGCACACCAATTGTGTTTGAGAGAAAGGCTGTCTCTAGTCTGAAAGAGGCAGACCTGGAACGAATCAAAGAT
GGCGAGCGCAACGAAGCAATTGTGGAGGCCATCAGGAGCTGGATTGCCACTGGACGCCCAGCTGATGCACCACCACGCTCCCCC
CGAGGCGACATCATTACCAAGATCAGGCTGGCCACCACCATCAAGGCAGCCGTGCCTGTCCGCGGAGGGACCGCAGGAAGGGGA
GAAATGGTGCGCGCAGATGTGTTCAGCAAGCCAAACCGGAGAGGGAAAGACGAGTGGTATCTGGTGCCCGTGTATCCACACCAG
ATCATGAACAGGAAGGCTTGGCCCAAACCTCCAATGCGCTCAATTGTGGCCAATAAGGATGAGGACGAATGGACCGAAGTGGGA
CCTGAACACCAGTTCCGGTTTAGCCTGTACCCTAGATCCAATATCGAGATCATTAGGCCATCTGGAGAAGTGATCGAAGGATAT

FIG. 9L

TTCGTCGGCCTGCATCGCAACACTGGCGCTCTGACCATCAGTGCACACAATGATCCCAAGAGTATCCATTCAGGCATTGGGACC
AAGACACTGCTGGCCATTTCCAAATACCAGGTGGACAGATTCGGCAGAAAGTCTCCAGTGCGCAAAGAGGTCCGAACTTGGCAC
GGGGAAGCCTGTATCTCTCCCACCCCCCCTGGATAAGAATTC

#19
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAGAGAG
AACGGGAGTGACGAAAGGCGGAGGAATATGGACGAAAAAATGGATTACAGGATTGGGCTGGACATCGGCATCGCATCAGTGGGG
TGGGCCGTCCTGCAGAACAATTCAGACGATGAGCCTGTGAGGATTGTCGACCTGGGAGTGCGCATTTTCGATACCGCAGAGATC
CCAAAGACAGGCGAAAGCCTGGCAGGACCCCGGAGAGCAGCTCGAACCACAAGGCGCCGACTGCGGAGAAGGAAACACCGCCTG
GACCGAATCAAGTGGCTGTTCGAGAACCAGGGGCTGATCAATATTGACGATTTTCTGAAGAGATACAACATGGCCGGACTGCCA
GATGTGTACCAGCTGCGGTATGAGGCTCTGGACAGAAAACTGACCGATGAGGAACTGGCCCAGGTGCTGCTGCACATCGCTAAG
CATCGGGGCTTCAGAAGCACCAGGAAAGCCGAGACAGCAGCCAAGGAAAACGGGGCAGTGCTGAAGGCCACAGACGAGAATCAG
AAACGAATGCAGGAAAAGGGATACAGGACAGTGGGCGAGATGATCTACCTGGACGAGGCCTTCCGGACTGGCTGCTCTTGGAGT
GAGAAAGGGTACATCCTGACCCCCCGCAACAAGGCTGAAAATTATCAGCACACAATGCTGCGGGCAATGCTGGTGGAGGAAGTC
AAGGAGATTTTCAGCTCCCAGCGCCGACTGGGCAACGAAAAAGCCACTGAGGAACTGGAGGAAAAGTACCTGGAGATCATGACC
TCCCAGCGCTCTTTTGACCTGGGGCCTGGAATGCAGCCAGATGGGAAGCCCTCCCCTTATGCAATGGAGGGCTTCTCTGACAGA
GTGGGGAAATGTACTTTTCTGGGGGATCAGGGAGAGCTGAGGGGCGCTAAGGGGACCTACACAGCCGAATATTTCGTGGCTCTG
CAGAAAATCAACCACACAAAGCTGGTCAATCAGGACGGCGAGACAAGGAATTTCACTGAGGAAGAGCGGAGAGCCCTGACTCTG
CTGCTGTTTACCCAGAAAGAGGTGAAGTACGCTGCAGTCCGCAAGAAACTGGGCCTGCCTGAGGACATCCTGTTCTACAACCTG
AACTACAAGAAGGCCGCTACTAAAGAAGAGCAGCAGAAGGAGAACCAGAATACCGAAAAAGCCAAGTTTATCGGGATGCCATAC
TATCACGATTACAAGAAATGCCTGGAAGAGAGAGTGAAGTATCTGACCGAGAACGAAGTCAGGGACCTGTTTGATGAGATCGGA
ATGATTCTGACTTGTTACAAAAATGACGATTCCCGCACCGAACGACTGGCCAAGCTGGGACTGGTGCCCATCGAGATGGAAGGC
CTGCTGGCTTATACTCCTACCAAATTCCAGCATCTGTCTATGAAGGCAATGCGGAACATCATTCCCTTTCTGGAGAAAGGGATG
ACCTACGACAAGGCTTGCGAAGAGGCAGGATATGACTTCAAAGCCGATAGCAAGGGGACTAAACAGAAGCTGCTGACCGGAGAG
AACGTGAATCAGACAATCAACGAAATTACTAATCCTGTGGTCAAACGCTCAGTGAGCCAGACAGTGAAGGTCATTAACGCCATC
ATTCGGACTTACGGCAGTCCACAGGCTATCAATATTGAGCTGGCAAGAGAAATGTCAAAGACCTTTGAAGAGAGGCGCAAAATC
AAGGGGGACATGGAGAAACGGCAGAAGAACAATGAAGATGTGAAGAAACAGATTCAGGAGCTGGGAAAACTGTCTCCTACAGGC
CAGGACATCCTGAAGTACAGACTGTGGCAGGAGCAGCAGGGGATTTGTATGTATAGTGGAAAAACCATCCCACTGGAAGAGCTG
TTCAAGCCCGGCTACGACATCGATCACATTCTGCCCTATTCAATTACATTCGACGATAGCTTTAGGAACAAAGTGCTGGTCACA
TCCCAGGAGAACAGACAGAAGGGCAATAGGACTCCTTACGAGTATATGGGGAACGACGAACAGCGCTGGAATGAGTTTGAAACC
AGGGTGAAAACTACCATCCGCGATTACAAGAAACAGCAGAAGCTGCTGAAGAAACATTTCTCTGAAGAGGAAAGGAGTGAGTTT
AAAGAACGGAACCTGACAGACACTAAGTACATCACAACCGTGATCTACAACATGATCAGACAGAATCTGGAGATGGCCCCCCTG
AACCGCCCTGAAAAGAAAAAGCAGGTGCGGGCTGTCAATGGCGCAATTACCGCCTACCTGCGAAAACGGTGGGGGCTGCCACAG
AAGAATCGGGAGACAGACACACACCATGCTATGGATGCAGTGGTCATCGCCTGCTGTACCGACGGCATGATCCAGAAAATTAGT
AGATACACAAAGGTGAGAGAGAGGTGCTATTCAAAGGGAACAGAGTTCGTCGATGCAGAGACTGGCGAAATCTTTAGACCCGAG
GACTACAGCAGGGCCGAGTGGGATGAAATTTTCGGCGTGCACATCCCAAAGCCCTGGGAGACATTTCGCGCCGAACTGGACGTC
CGAATGGGGGACGATCCAAAGGGATTCCTGGACACTCATAGCGATGTGGCTCTGGAGCTGGATTATCCCGAGTACATCTACGAA
AACCTGCGGCCTATCTTCGTGAGCAGAATGCCAAATCACAAGGTCACCGGAGCAGCCCATGCTGACACAATTCGGTCCCCAAGA
CACTTTAAAGATGAGGGCATCGTGCTGACTAAGACCGCACTGACCGACCTGAAACTGGACAAGGATGGGGAGATCGACGGATAC
TATAACCCCCAGTCCGATCTGCTGCTGTACGAAGCACTGAAAAAGCAGCTGCTGCTGTATGGCAATGATGCCAAAAAGGCCTTC
GCTCAGGACTTTCATAAACCCAAGGCCGATGGAACTGAGGGCCCTGTGGTCAGGAAGGTGAAGATCCAGAAAAAGCAGACCATG
GGAGTGTTCGTCGACTCTGGCAACGGGATTGCCGAGAATGGCGGGATGGTGCGCATCGATGTGTTCCGAGTCAACGGCAAGTAC
TATTTTGTGCCCGTCTACACCGCTGACGTGGTCAAAAAGGTGCTGCCTAATAGGGCCAGTACAGCTCACAAGCCATACGGCGAG
TGGAAAGTGATGGAGGACAAGGATTTCCTGTTTAGTCTGTATTCACGCGACCTGATCCATATCAAGTCTAAAAAGGATATCCCT
ATTAAGATGGTGAACGGAGGCATGGAGGGGATCAAGGAAACCTACGCATACTATATTGGAGCCGACATCAGCGCTGCAAATATC
CAGGGCATTGCCCACGATTCCAGGTATAAATTCCGCGGACTGGGCATTCAGTCTCTGGACGTGCTGGAGAAGTGTCAGATCGAT
GTGCTGGGACATGTCAGCGTGGTCCGATCCGAAAAGCGGATGGGCTTTAGCTAAGAATTC

#21
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAGG
AACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAACAAGGGACGTGATCGACGCA
GGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGG
AGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAAT
CCTTATGAAGCCAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGC
CGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTCACGCAATAGCAAA
GCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTC
AAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGAT
ACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGATGGAAAGACATCAAG
GAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTG
TACAACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTTCCAGATC
ATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAG
GGCTACCGGGTGACAAGCACTGGAAAACCAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAA
GAAATCATTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAG

FIG. 9M

CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAACCTG
TCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTG
GTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTCATTCTGTCACCCGTGGTCAAG
CGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATTATCGAGCTGGCT
AGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAG
ATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTG
TATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTG
TCCTTCGACAATTCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTTCCAGTAC
CTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGC
AAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTG
GACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCC
ATCAACGGCGGGTTCACATCTTTTCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAA
GATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAG
ATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAG
ATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACC
CTGTATAGTACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACAAG
CTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAAACTGAAGCTGATT
ATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAG
GATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGACGATTACCCTAAC
AGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACT
GTCAAGAATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAG
ATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGGTCATCGGG
GTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAG
CGCCCCCCTCGAATTATCAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTAT
GAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATTC

#22
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAGG
AACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAACAAGGGACGTGATCGACGCA
GGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGG
AGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAAT
CCTTATGAAGCCAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGC
CGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTCACGCAATAGCAAA
GCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTC
AAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGAT
ACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGATGGAAAGACATCAAG
GAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTG
TACAACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTTCCAGATC
ATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAG
GGCTACCGGGTGACAAGCACTGGAAAACCAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAA
GAAATCATTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAG
CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAACCTG
TCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTG
GTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTCATTCTGTCACCCGTGGTCAAG
CGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATTATCGAGCTGGCT
AGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAG
ATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTG
TATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTG
TCCTTCGACAATTCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTTCCAGTAC
CTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGC
AAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTG
GACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCC
ATCAACGGCGGGTTCACATCTTTTCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAA
GATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAG
ATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAG
ATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACC
CTGTATAGTACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACAAG
CTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAAACTGAAGCTGATT
ATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAG
GATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGACGATTACCCTAAC
AGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACT
GTCAAGAATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAG
ATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGGTCATCGGG

FIG. 9N

GTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAG
CGCCCCCCTCGAATTATCAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTAT
GAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATTC

#23
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAACAAT
AGCATCAAATCTAAACCTGAAGTGACCATCGGGCTGGACCTGGGAGTGGGAAGCGTGGGGTGGGCAATCGTGGATAACGAAACA
AACATCATTCACCATCTGGGCTCCAGGCTGTTTTCTCAGGCCAAGACTGCTGAGGATCGGAGATCTTTCCGCGGGGTGAGGCGC
CTGATCCGACGGAGAAAATACAAGCTGAAACGATTCGTCAATCTGATTTGGAAGTACAACAGCTATTTCGGCTTCAAGAACAAA
GAGGACATCCTGAACAATTATCAGGAGCAGCAGAAGCTGCACAATACCGTGCTGAACCTGAAATCAGAGGCACTGAATGCCAAG
ATCGATCCTAAAGCACTGAGCTGGATTCTGCACGACTACCTGAAGAACAGAGGCCATTTTTATGAGGACAATAGGGATTTCAAC
GTGTACCCAACAAAGGAGCTGGCCAAGTACTTCGATAAGTACGGGTACTACAAGGGAATCATTGACAGCAAGGAGGACAATGAT
AACAAACTGGAGGAAGAGCTGACAAAGTACAAATTCTCCAATAAGCACTGGCTGGAAGAGGTGAAGAAAGTCCTGTCTAACCAG
ACTGGCCTGCCAGAAAAGTTTAAAGAAGAGTATGAGTCACTGTTCAGCTACGTGAGAAATTATTCAGAGGGCCCAGGGAGCATC
AACTCTGTCAGTCCCTACGGGATCTACCATCTGGACGAAAAAGAGGGAAAGGTGGTCCAGAAGTACAACAACATCTGGGATAAG
ACAATCGGAAAGTGCAACATCTTCCCTGACGAGTATAGAGCTCCCAAGAACAGTCCTATCGCAATGATTTTCAATGAAATCAAC
GAGCTGTCCACAATCAGGTCATACAGCATCTACCTGACTGGCTGGTTCATTAATCAGGAGTTCAAGAAAGCCTACCTGAACAAG
CTGCTGGATCTGCTGATCAAAACCAACGGAGAGAAGCCAATTGACGCAAGGCAGTTCAAGAAACTGCGCGAAGAGACAATCGCC
GAAAGCATTGGCAAAGAGACACTGAAGGATGTGGAGAATGAAGAGAAACTGGAAAAGGAGGACCACAAGTGGAAACTGAAGGGA
CTGAAGCTGAATACCAACGGCAAAATCCAGTACAACGATCTGAGCTCCCTGGCTAAGTTTGTGCACAAACTGAAGCAGCATCTG
AAACTGGATTTCCTGCTGGAGGACCAGTATGCAACACTGGACAAGATCAATTTCCTGCAGTCCCTGTTTGTGTACCTGGGCAAG
CACCTGAGATATTCCAATAGGGTCGATTCTGCCAACCTGAAGGAATTTTCCGACTCTAACAAACTGTTCGAGCGCATCCTGCAG
AAACAGAAGGATGGGCTGTTCAAGCTGTTTGAACAGACTGACAAAGACGATGAGAAGATCCTGGCCCAGACACATAGTCTGTCA
ACTAAGGCCATGCTGCTGGCTATTACCCGGATGACAAATCTGGACAACGATGAGGACAACCAGAAAAACAATGACAAGGGCTGG
AATTTTGAGGCCATCAAAAACTTCGATCAGAAGTTTATCGACATCACCAAGAAAAACAACAACCTGAGCCTGAAACAGAATAAG
CGCTACCTGGACGATCGATTCATCAACGATGCTATTCTGTCCCCTGGGGTGAAGCGAATCCTGCGGGAGGCAACCAAGGTCTTT
AATGCCATTCTGAAACAGTTCTCTGAAGAGTACGACGTGACAAAGGTGGTCATCGAACTGGCTCGCGAGCTGAGCGAAGAGAAG
GAACTGGAGAACACAAAGAACTACAAGAAACTGATCAAGAAAAACGGCGACAAGATTAGTGAGGGCCTGAAAGCACTGGGGATC
TCAGAAGATGAGATCAAAGACATTCTGAAGAGTCCCACTAAATCATACAAGTTTCTGCTGTGGCTGCAGCAGGACCACATCGAT
CCTTATAGCCTGAAGGAGATCGCCTTCGACGATATTTTTACCAAAACAGAAAAGTTCGAGATCGACCATATCATTCCCTACAGC
ATTTCCTTCGACGATTCTAGTTCAAACAAGCTGCTGGTGCTGGCTGAAAGTAATCAGGCAAAGTCAAACCAGACTCCTTATGAG
TTCATCAGCTCCGGAAACGCAGGCATTAAGTGGGAAGATTACGAGGCCTATTGCCGCAAGTTCAAGGATGGGGACTCTAGTCTG
CTGGACAGCACCCAGCGGTCCAAGAAATTCGCCAAAATGATGAAAACCGATACCTCAAGCAAGTACGACATCGGATTTCTGGCT
CGAAATCTGAACGATACTCGGTACGCAACCATTGTGTTCCGGGACGCCCTGGAGGACTATGCTAATAACCACCTGGTCGAGGAC
AAACCCATGTTTAAGGTGGTCTGTATCAATGGGTCCGTGACCTCTTTCCTGCGGAAGAACTTTGACGATTCCTCTTACGCCAAG
AAAGATAGAGACAAGAATATCCACCATGCTGTGGATGCAAGTATCATCTCAATTTTCAGCAACGAGACAAAGACTCTGTTCAAC
CAGCTGACTCAGTTTGCTGACTATAAACTGTTCAAGAACACCGATGGCAGCTGGAAGAAAATCGACCCTAAGACAGGGGTGGTC
ACTGAAGTGACCGACGAGAATTGGAAGCAGATTAGGGTGCGCAACCAGGTGAGCGAAATCGCCAAAGTCATTGAGAAGTACATC
CAGGATAGCAACATCGAAAGAAAGGCTAGGTATTCCCGCAAAATCGAGAATAAGACTAACATTTCCCTGTTTAATGACACCGTG
TACTCTGCCAAGAAAGTCGGCTATGAGGATCAGATCAAAAGAAAGAACCTGAAAACCCTGGACATTCACGAATCTGCTAAAGAG
AATAAGAACAGTAAAGTGAAGCGGCAGTTTGTCTACAGAAAGCTGGTGAATGTCAGCCTGCTGAATAACGATAAGCTGGCAGAC
CTGTTCGCCGAAAAAGAGGATATCCTGATGTATAGGGCCAATCCATGGGTCATCAACCTGGCTGAGCAGATTTTCAATGAATAC
ACTGAGAACAAGAAAATCAAGTCCCAGAACGTGTTTGAAAAATATATGCTGGACCTGACCAAAGAGTTCCCCGAGAAGTTCAGC
GAGTTTCTGGTGAAGTCCATGCTGAGAAACAAGACCGCCATCATCTACGACGATAAGAAAAACATTGTCCATCGAATCAAACGG
CTGAAGATGCTGAGTTCAGAACTGAAAAGAGAATAAGCTGTCTAACGTGATCATTAGGTCTAAGAATCAGAGTGGGACCAAACTG
TCATACCAGGATACAATCAACAGCCTGGCCCTGATGATTATGCGCAGCATCGACCCTACTGCTAAGAAACAGTATATTCGAGTG
CCACTGAATACCCTGAACCTGCACCTGGGAGATCATGACTTTGATCTGCACAATATGGATGCTTACCTGAAGAAACCAAAATTC
GTGAAGTATCTGAAAGCAAACGAAATCGGCGACGAGTACAAGCCCTGGAGGGTCCTGACATCTGGCACTCTGCTGATCCATAAG
AAGGATAAGAAACTGATGTACATCAGCTCCTTCCAGAATCTGAACGACGTGATCGAAATTAAGAATCTGATCGAAACCGAGTAT
AAAGAGAACGACGATTCTGATAGTAAGAAAAAGAAAAAGGCAAACCGCTTTCTGATGACCCTGAGCACAATCCTGAATGACTAC
ATTCTGCTGGACGCCAAGGATAACTTCGACATCCTGGGGCTGTCTAAAAATCGGATCGATGAGATTCTGAACAGTAAGCTGGGA
CTGGACAAGATTGTGAAATAAGAATTC

#24
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAGGATT
CTGGGGTTTGACATTGGCATTAACAGCATCGGGTGGGCTTTTGTGGAGAACGACGAACTGAAGGACTGCGGAGTGCGGATCTTC
ACAAAGGCCGAGAACCCAAAAAATAAGGAAAGCCTGGCACTGCCCCGGAGAAATGCACGCAGCTCCAGGCGCCGACTGAAACGG
AGAAAGGCCCGGCTGATCGCTATTAAGAGAATCCTGGCCAAAGAGCTGAAGCTGAACTACAAGGACTATGTCGCAGCTGATGGA
GAGCTGCCAAAGGCCTACGAAGGATCCCTGGCATCTGTGTACGAGCTGCGGTATAAGGCCCTGACACAGAACCTGGAAACTAAA
GATCTGGCCAGAGTGATCCTGCACATTGCTAAGCATAGGGGGTACATGAACAAGAACGAGAAGAAATCAAACGACGCTAAGAAA
GGAAAGATCCTGAGCGCTCTGAAAAACAATGCACTGAAGCTGGAGAACTACCAGAGCGTGGGCGAATACTTCTACAAGGAGTTC
TTTCAGAAATACAAGAAAAACACAAAGAACTTCATCAAGATCCGCAACACTAAGGATAATTACAACAATTGCGTGCTGTCTAGT
GACCTGGAAAAAGAGCTGAAGCTGATCCTGGAAAAACAGAAGGAGTTCGGCTACAACTACTCTGAAGATTTCATCAACGAGATT

FIG. 9O

CTGAAGGTCGCCTTCTTTCAGCGGCCCCTGAAGGACTTCAGTCACCTGGTGGGGGCCTGCACTTTCTTTGAGGAAGAGAAAAGG
GCCTGTAAGAACAGCTACTCTGCCTGGGAGTTTGTGGCTCTGACCAAGATCATTAACGAGATCAAGAGCCTGGAGAAGATCAGC
GGCGAAATTGTGCCAACCCAGACAATCAACGAGGTCCTGAATCTGATCCTGGACAAGGGGTCTATCACCTACAAGAAATTCAGA
AGTTGTATCAATCTGCATGAGAGTATCAGCTTCAAGAGCCTGAAGTATGATAAAGAAAACGCCGAGAATGCTAAACTGATCGAC
TTCCGCAAGCTGGTGGAGTTTAAGAAAGCCCTGGGAGTCCACAGCCTGTCCCGGCAGGAACTGGATCAGATCTCCACTCATATC
ACCCTGATTAAGGACAACGTGAAGCTGAAAACCGTCCTGGAGAAATACAACCTGAGTAATGAACAGATCAACAATCTGCTGGAA
ATTGAGTTCAACGATTATATCAACCTGAGCTTCAAGGCCCTGGGAATGATTCTGCCACTGATGCGCGAGGGCAAACGATACGAC
GAGGCCTGCGAGATCGCCAATCTGAAACCTAAGACCGTGGACGAGAAGAAAGATTTCCTGCCAGCATTTTGTGATTCCATTTTC
GCCCACGAGCTGTCTAACCCCGTGGTCAATAGGGCTATCAGCGAATACCGCAAGGTGCTGAACGCACTGCTGAAGAAATATGGA
AAGGTCCACAAAATTCATCTGGAGCTGGCTCGCGACGTGGGCCTGTCCAAGAAAGCACGAGAGAAGATCGAAAAAGAGCAGAAG
GAAAACCAGGCCGTGAATGCATGGGCCCTGAAGGAATGCGAGAATATTGGCCTGAAGGCCAGCGCAAAGAACATCCTGAAACTG
AAGCTGTGGAAAGAACAGAAGGAGATCTGTATCTACTCCGGAAATAAGATCTCTATTGAGCACCTGAAAGATGAAAAGGCCCTG
GAGGTGGACCATATCTACCCCTATTCTAGGAGTTTCGACGATTCTTTTATCAACAAAGTGCTGGTGTTCACCAAGGAAAATCAG
GAGAAACTGAACAAGACACCTTTCGAGGCCTTTGGCAAGAATATTGAAAAATGGAGCAAGATCCAGACCCTGGCTCAGAACCTG
CCATACAAGAAAAAGAATAAGATTCTGGACGAGAACTTCAAAGATAAGCAGCAGGAGGACTTTATCTCTCGAAATCTGAACGAC
ACCCGGTATATCGCTACACTGATTGCAAAATACACAAAGGAGTATCTGAACTTCCTGCTGCTGAGCGAAAATGAGAACGCCAAT
CTGAAGAGTGGCGAAAAAGGGTCAAAGATCCACGTGCAGACTATTAGCGGGATGCTGACCTCCGTCCTGAGGCACACATGGGGG
TTTGACAAAAAGGATCGCAACAATCATCTGCACCATGCACTGGATGCCATCATTGTGGCCTACAGTACAAATTCAATCATTAAG
GCTTTCAGCGATTTCCGGAAAAACCAGGAGCTGCTGAAGGCCAGATTCTACGCTAAAGAACTGACTTCCGATAACTATAAACAT
CAGGTCAAGTTCTTTGAGCCTTTCAAGAGTTTTAGAGAAAAAATCCTGTCAAAGATCGACGAGATTTTCGTGTCCAAACCACCT
CGAAAGCGAGCTAGGCGCGCACTGCACAAGGATACCTTTCATTCTGAGAACAAGATCATTGACAAGTGCAGCTACAACTCCAAG
GAAGGCCTGCAGATTGCCCTGAGCTGTGGAAGAGTGAGGAAAATCGGCACTAAGTATGTCGAGAATGATACCATCGTGAGGGTC
GACATTTTCAAAAAGCAGAACAAGTTTTACGCTATCCCAATCTACGCAATGGATTTTGCCCTGGGGATCCTGCCCAATAAGATC
GTGATTACTGGAAAAGATAAGAACAATAACCCCAAACAGTGGCAGACCATTGACGAATCATACGAGTTCTGCTTTAGCCTGTAT
AAGAATGACCTGATCCTGCTGCAGAAAAAGAACATGCAGGAACCTGAGTTCGCCTACTATAACGATTTTTCAATCAGCACATCA
AGCATTTGTGTGGAGAAACACGACAACAAGTTCGAAAATCTGACTAGCAACCAGAAGCTGCTGTTTTCCAATGCAAAAGAGGGC
TCTGTGAAGGTCGAAAGTCTGGGGATCCAGAACCTGAAAGTGTTCGAGAAGTACATCATTACCCCCCTGGGAGATAAAATTAAG
GCTGACTTTCAGCCTCGAGAAAACATCAGCCTGAAAACCAGTAAAAAGTATGGCCTGAGGTAAGAATTC

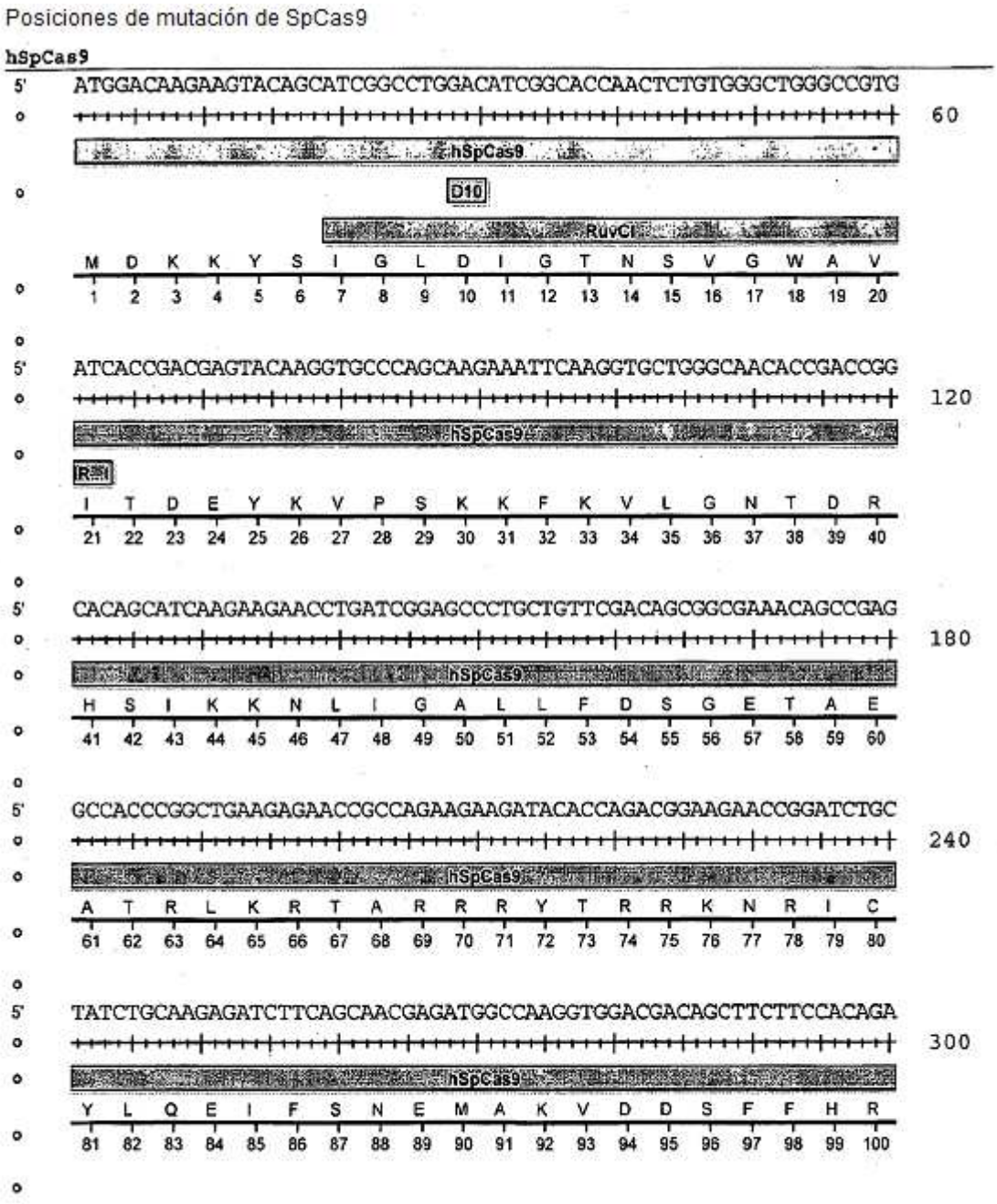

Posiciones de mutación de SpCas9
hSpCas9
5' ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
D10
RuvCI
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5' ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
RuvCII
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5' CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG
180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5' GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5' TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIG. 10A

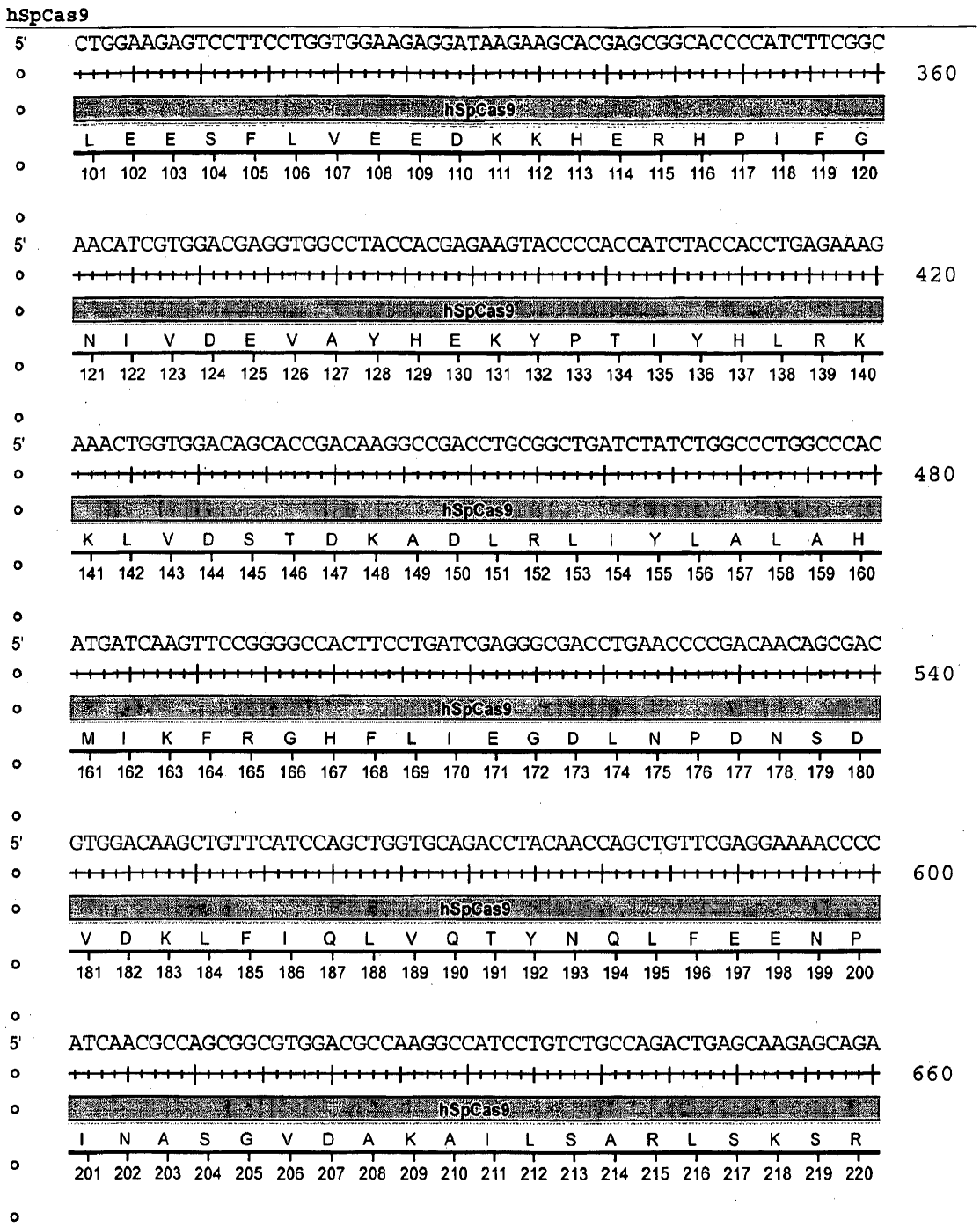

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

FIG. 10B

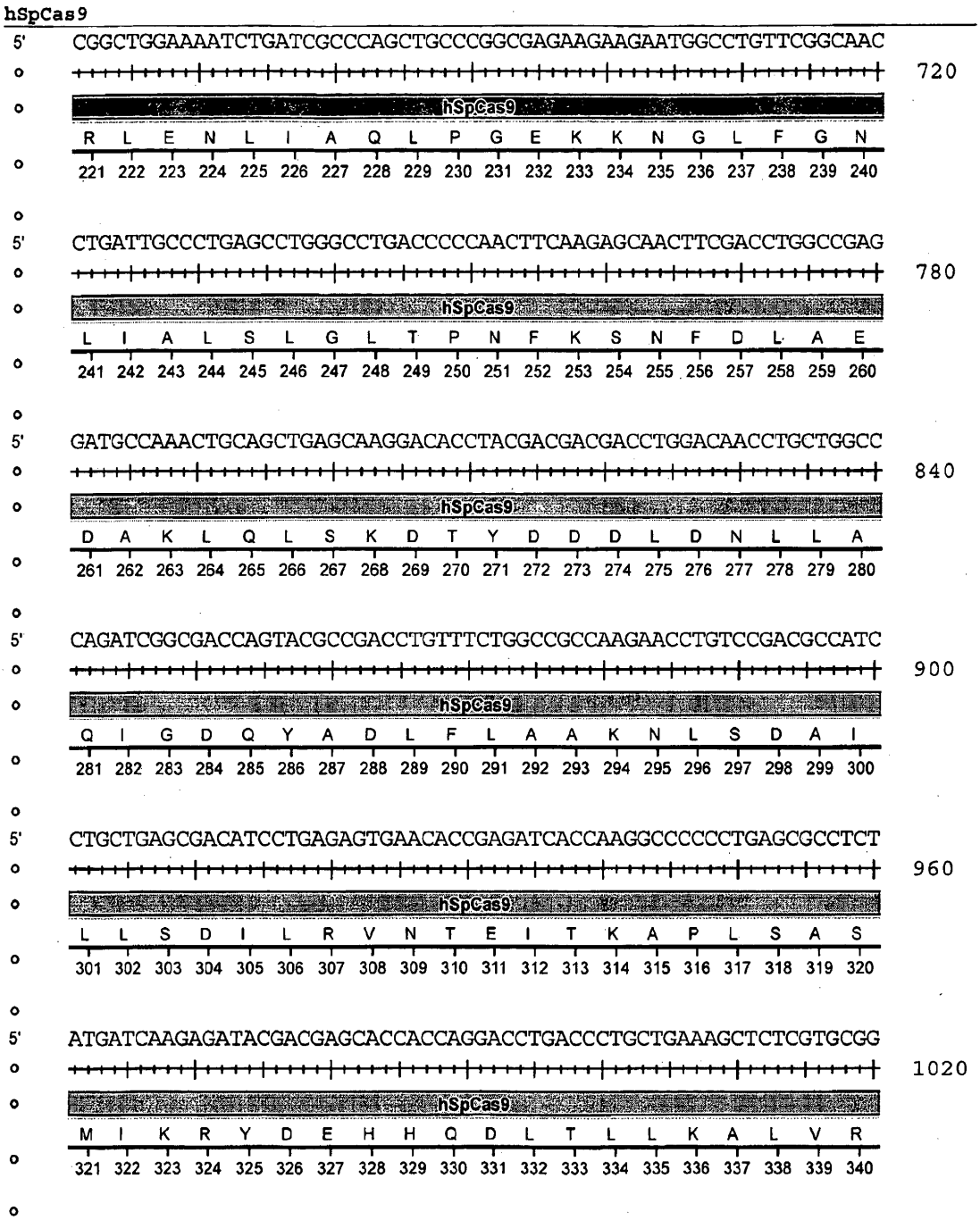

hSpCas9
5' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC 720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5' CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG 780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC 840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5' CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC 900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5' CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT 960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5' ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG 1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

FIG. 10C

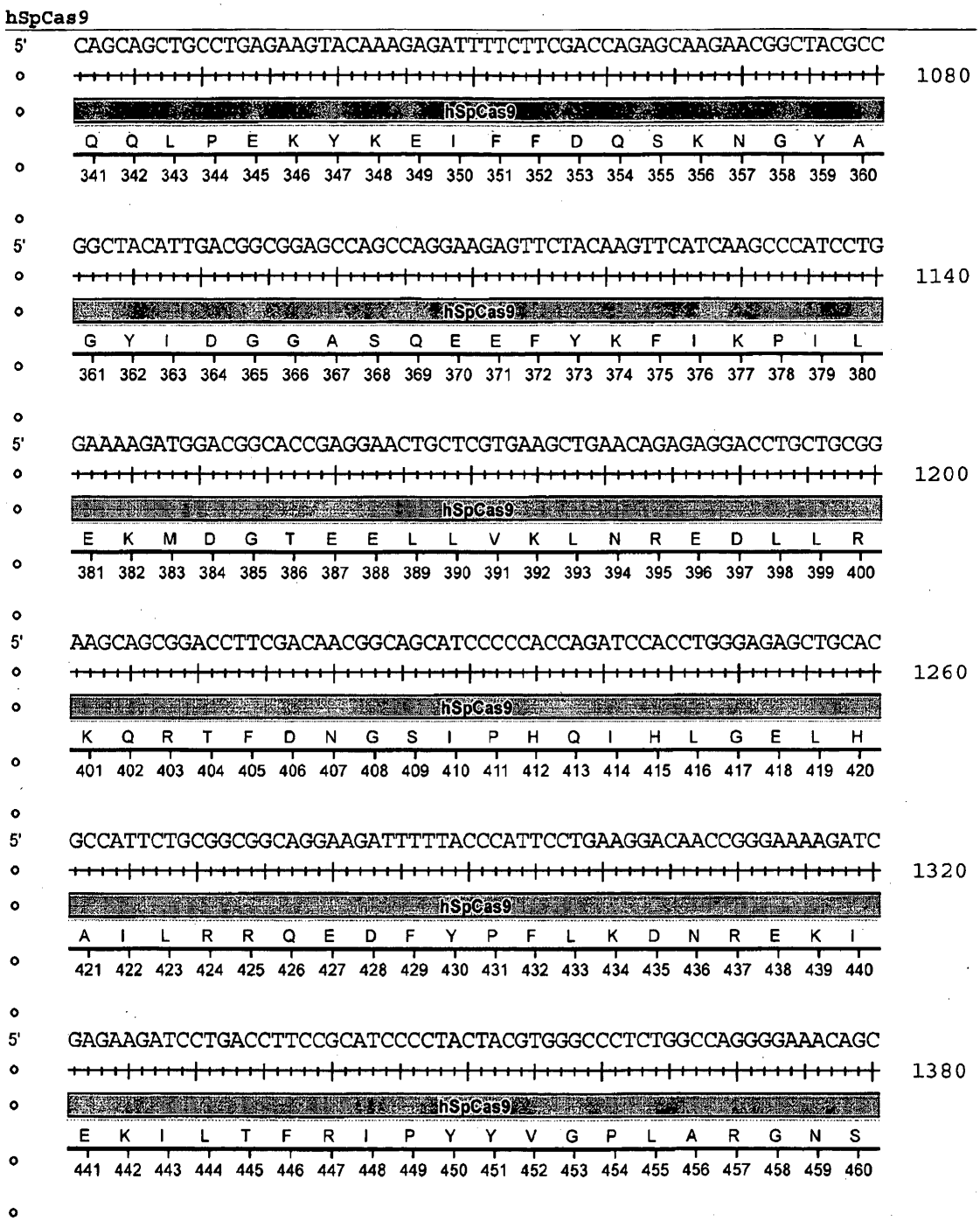

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG
1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC
1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

FIG. 10D

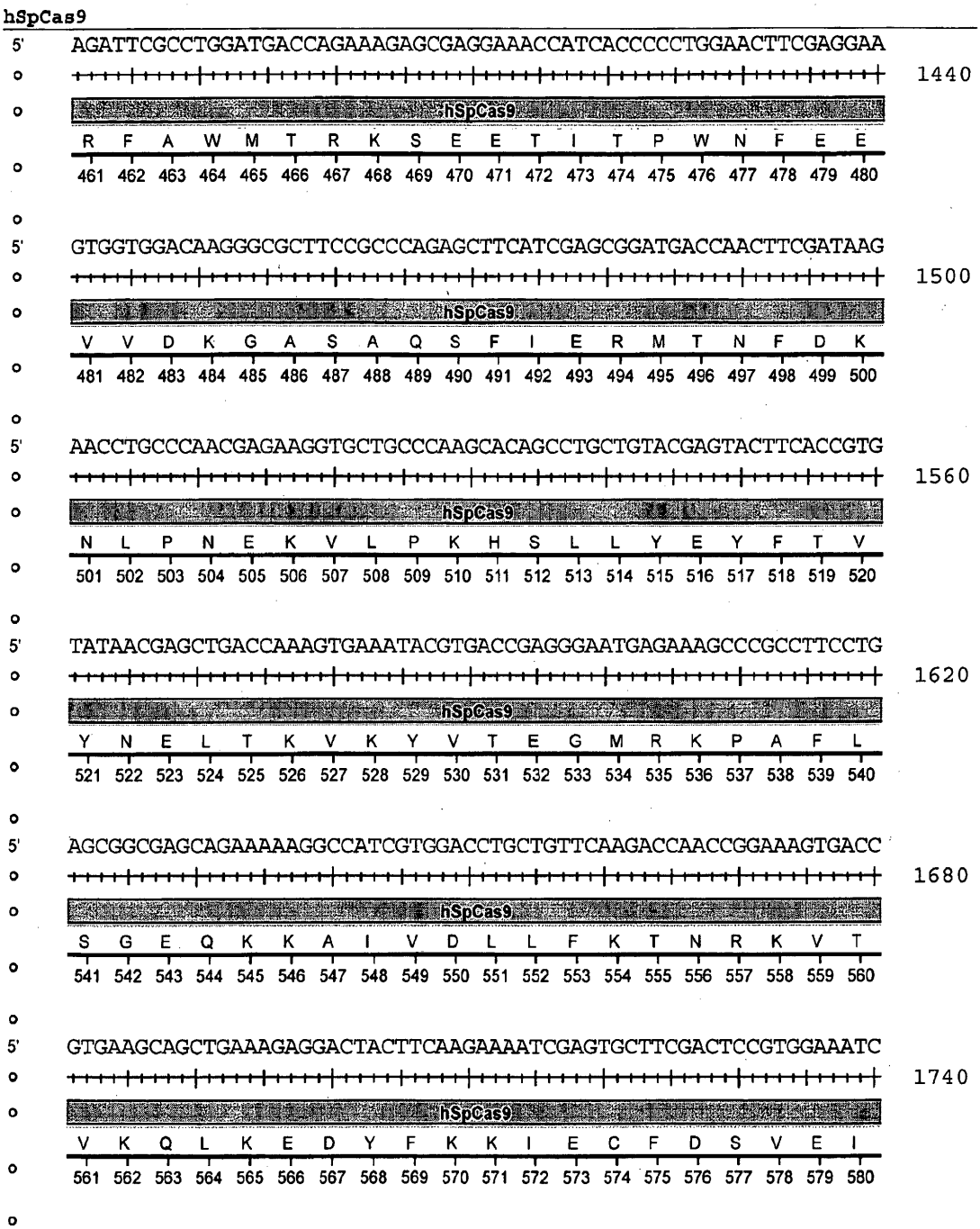

hSpCas9
5' AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

FIG. 10E

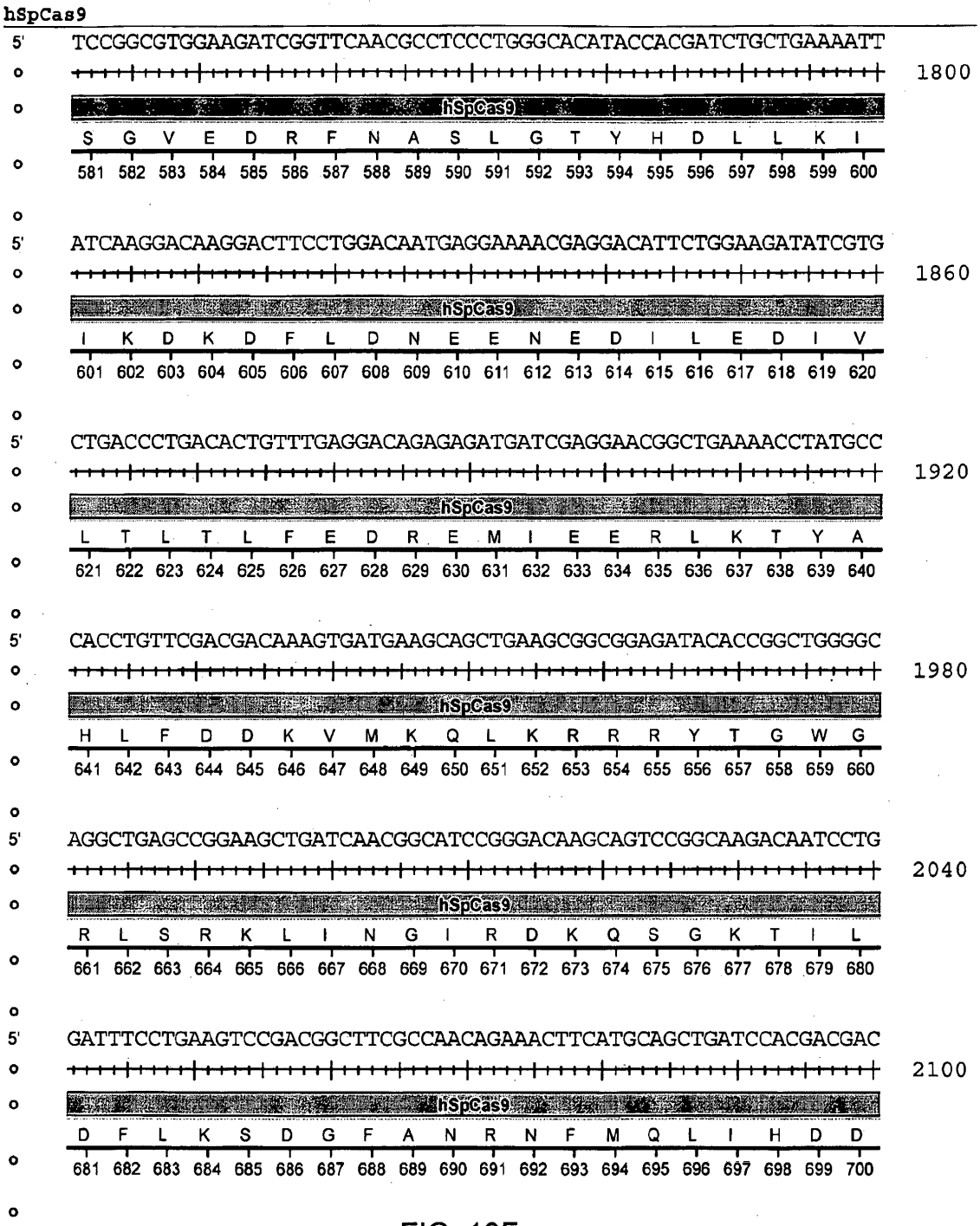

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG
1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

FIG. 10F

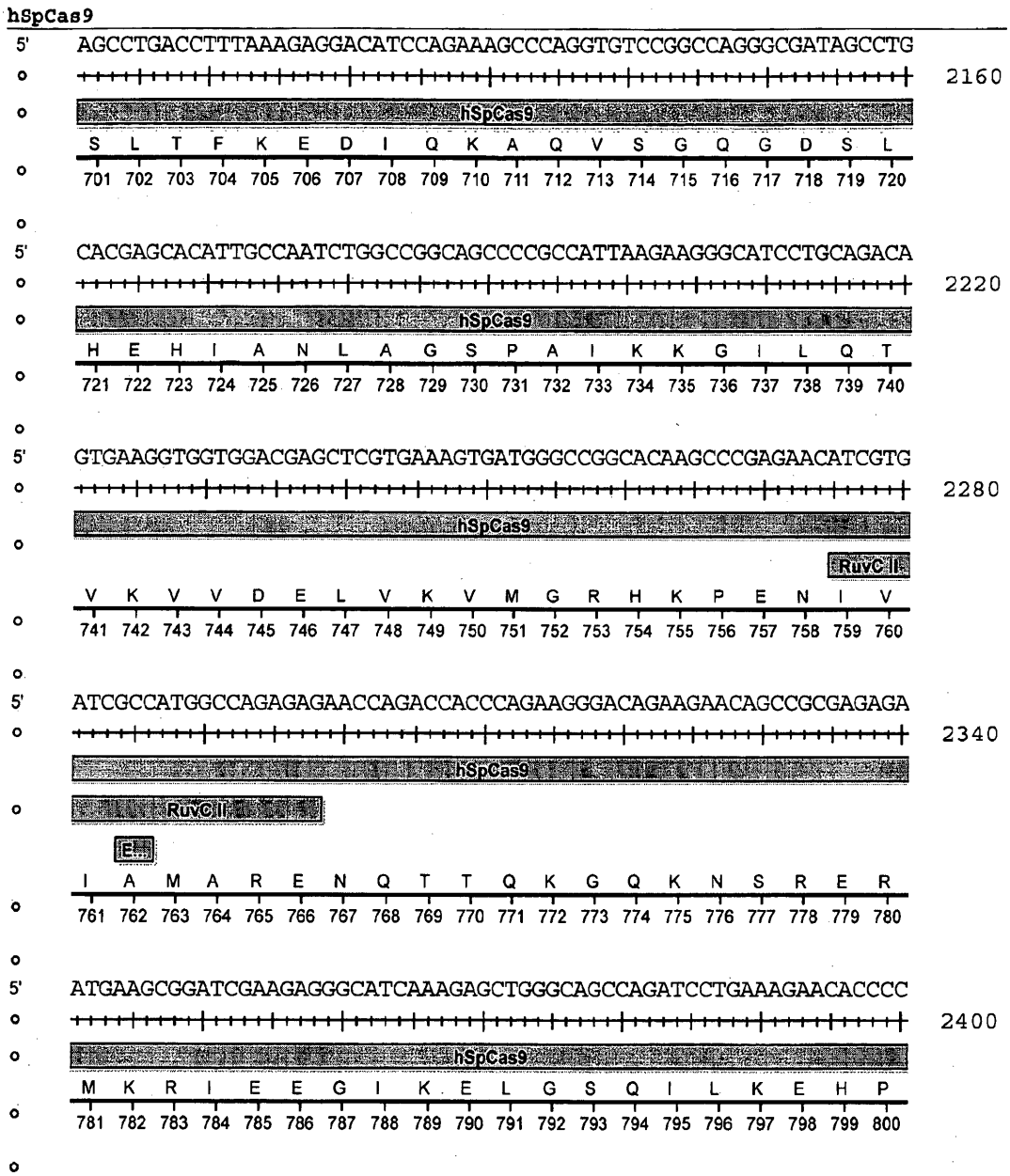

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

FIG. 10G

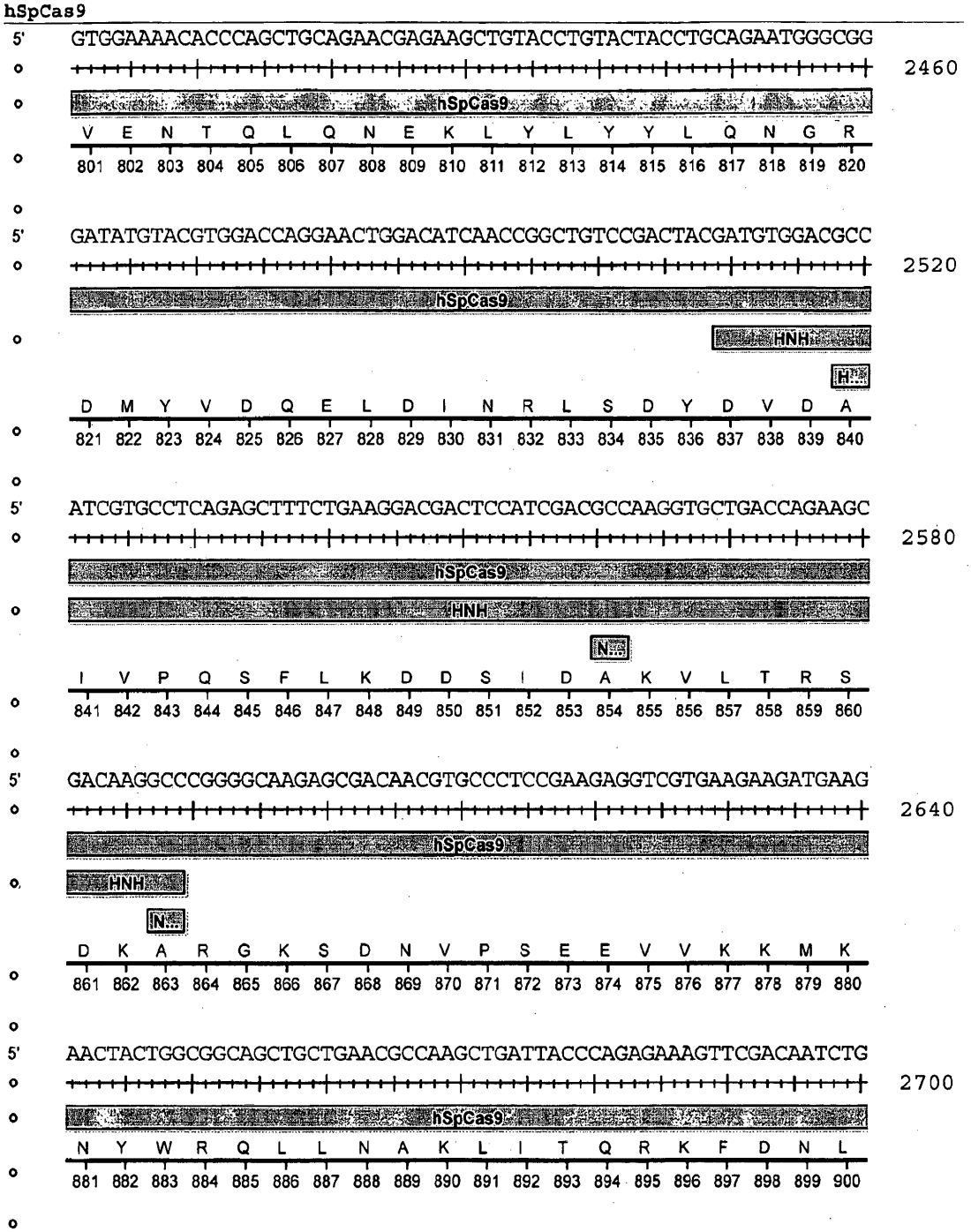

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

FIG. 10H

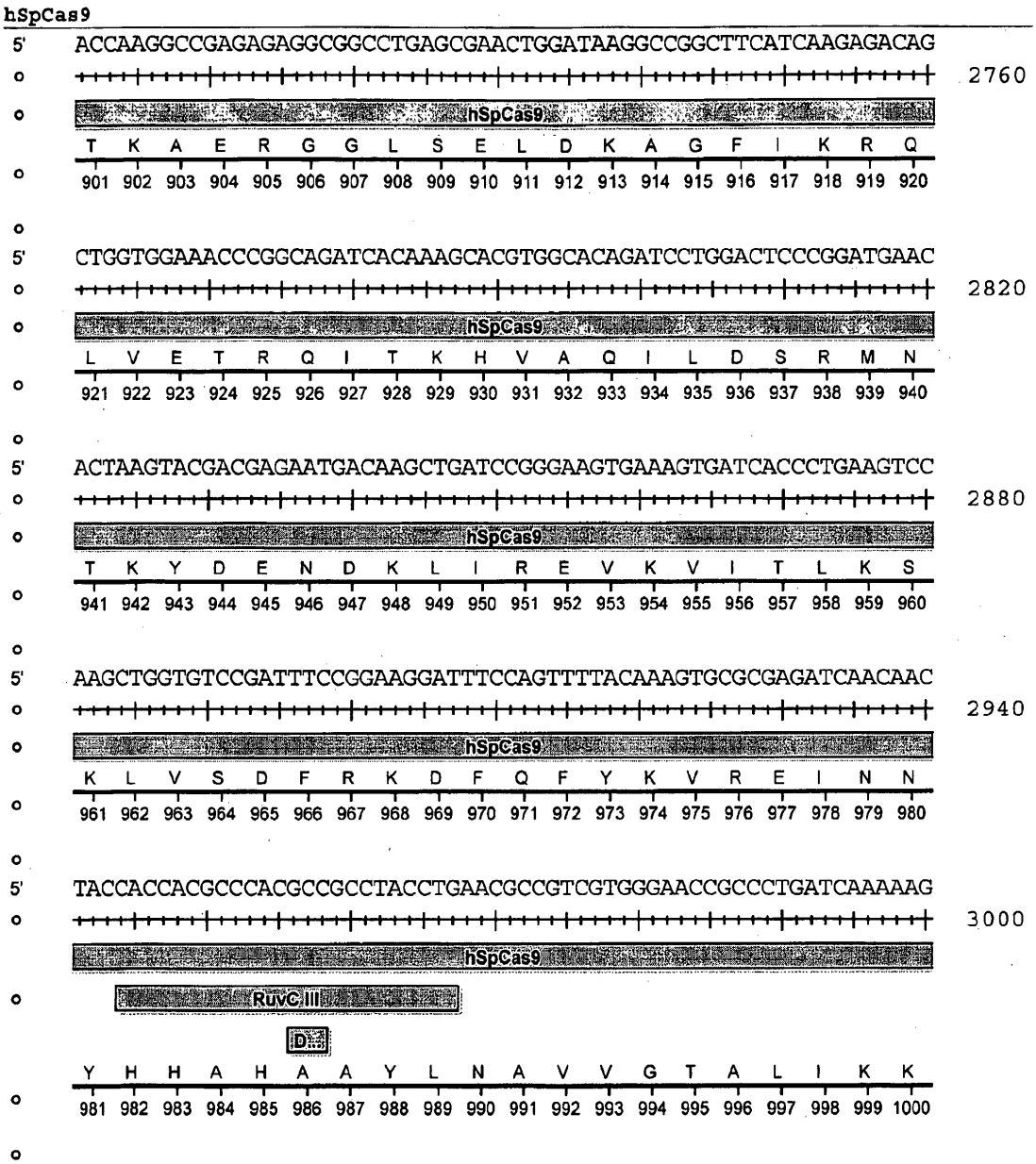

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5' ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5' AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5' TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

FIG. 10I

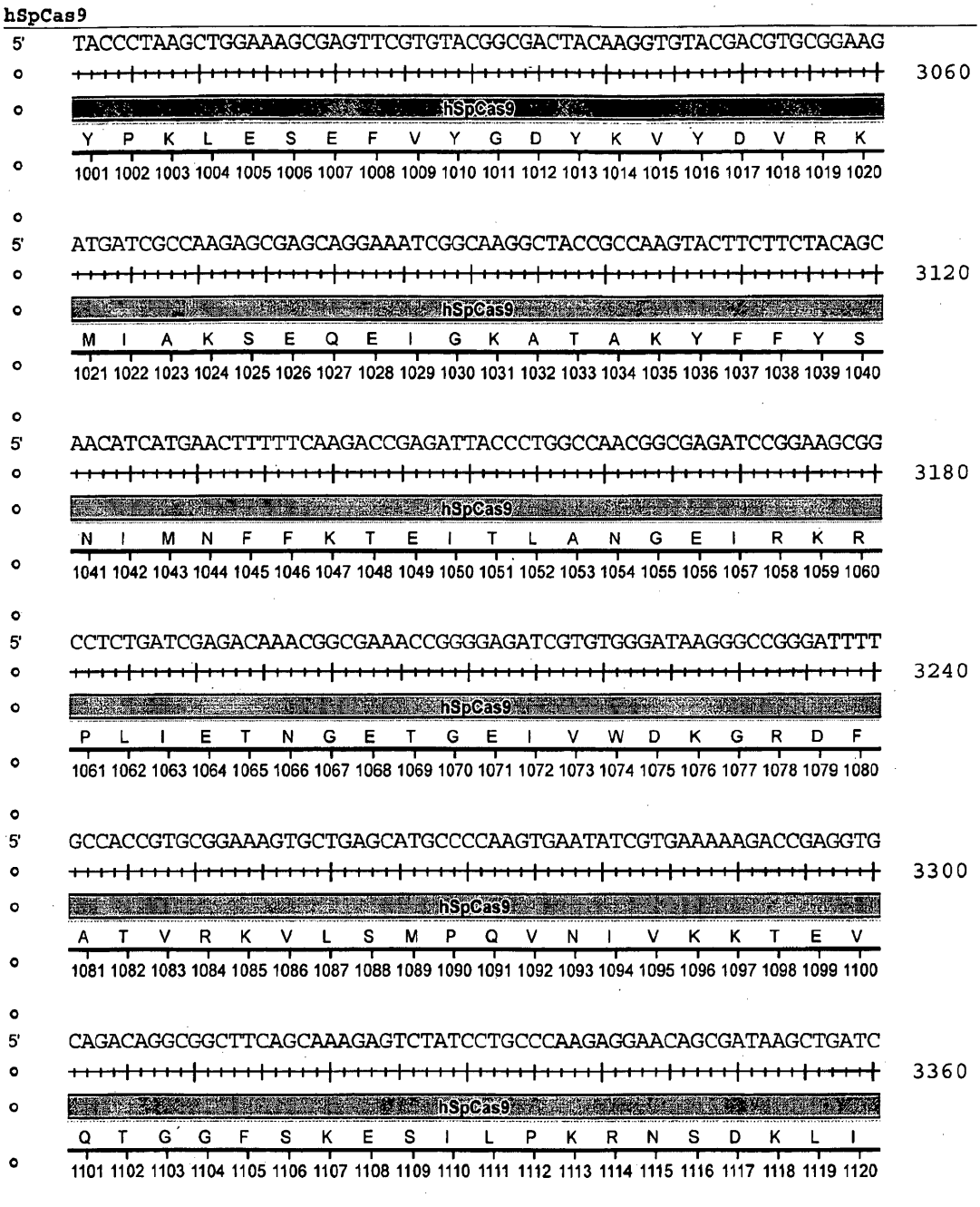

hSpCas9
5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
3060
hSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
3120
hSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
3180
hSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
3240
hSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
3300
hSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC
3360
hSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

FIG. 10J

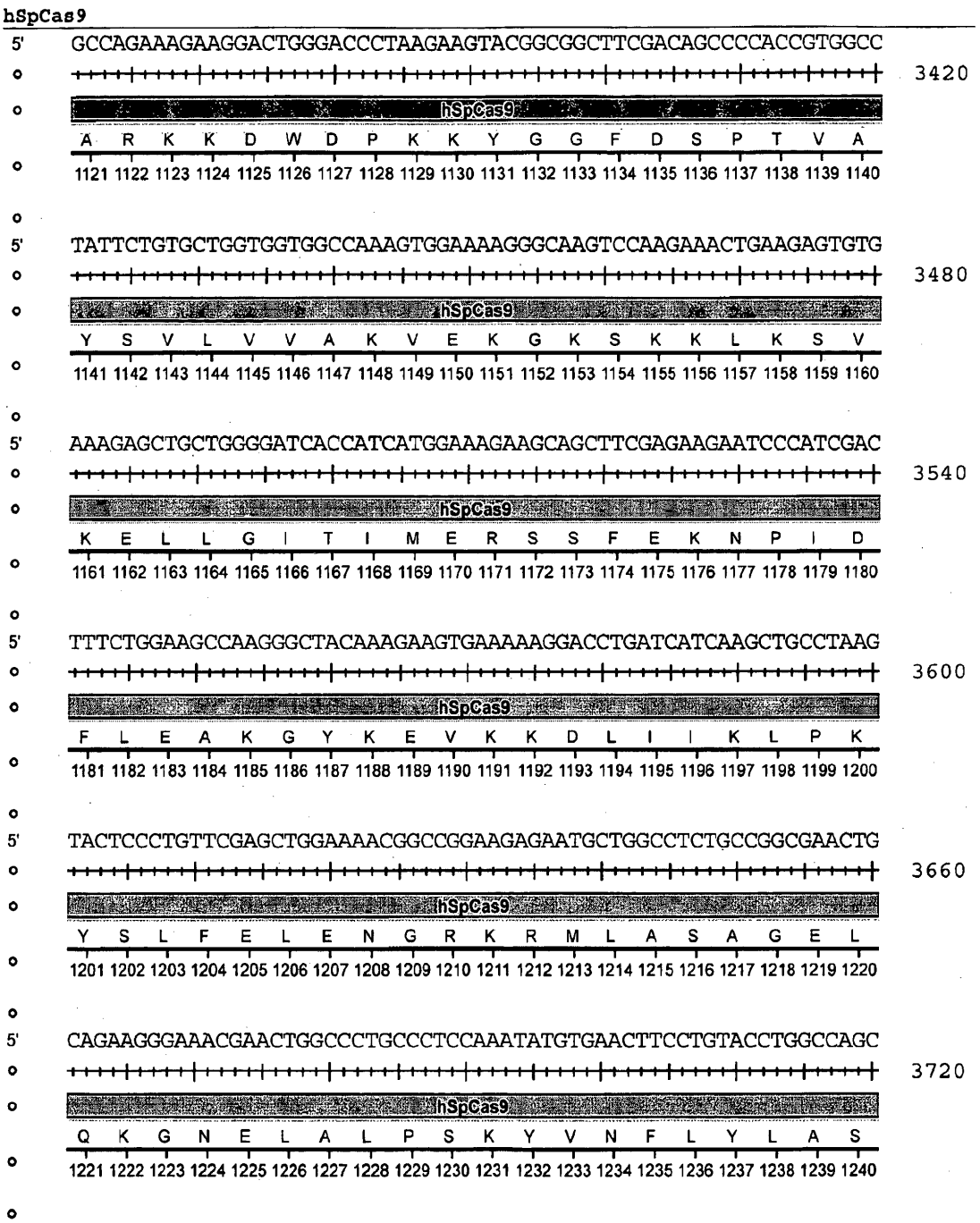

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

FIG. 10K

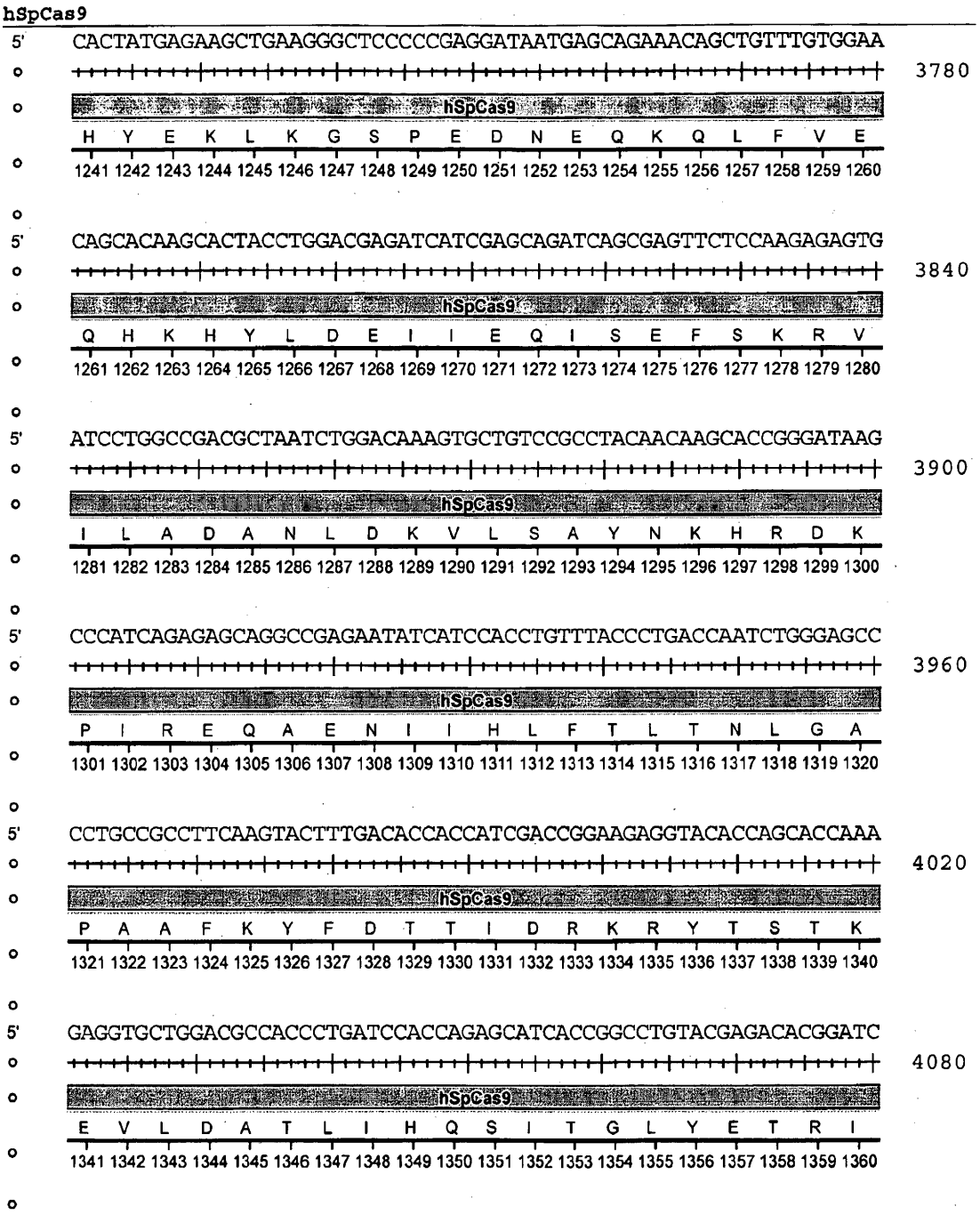

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA 3780
hSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG 3840
hSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG 3900
hSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC 3960
hSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA 4020
hSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC 4080
hSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

FIG. 10L

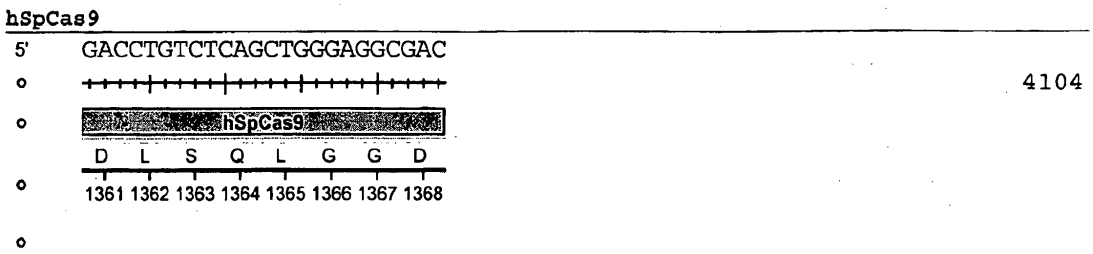

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

FIG. 10M

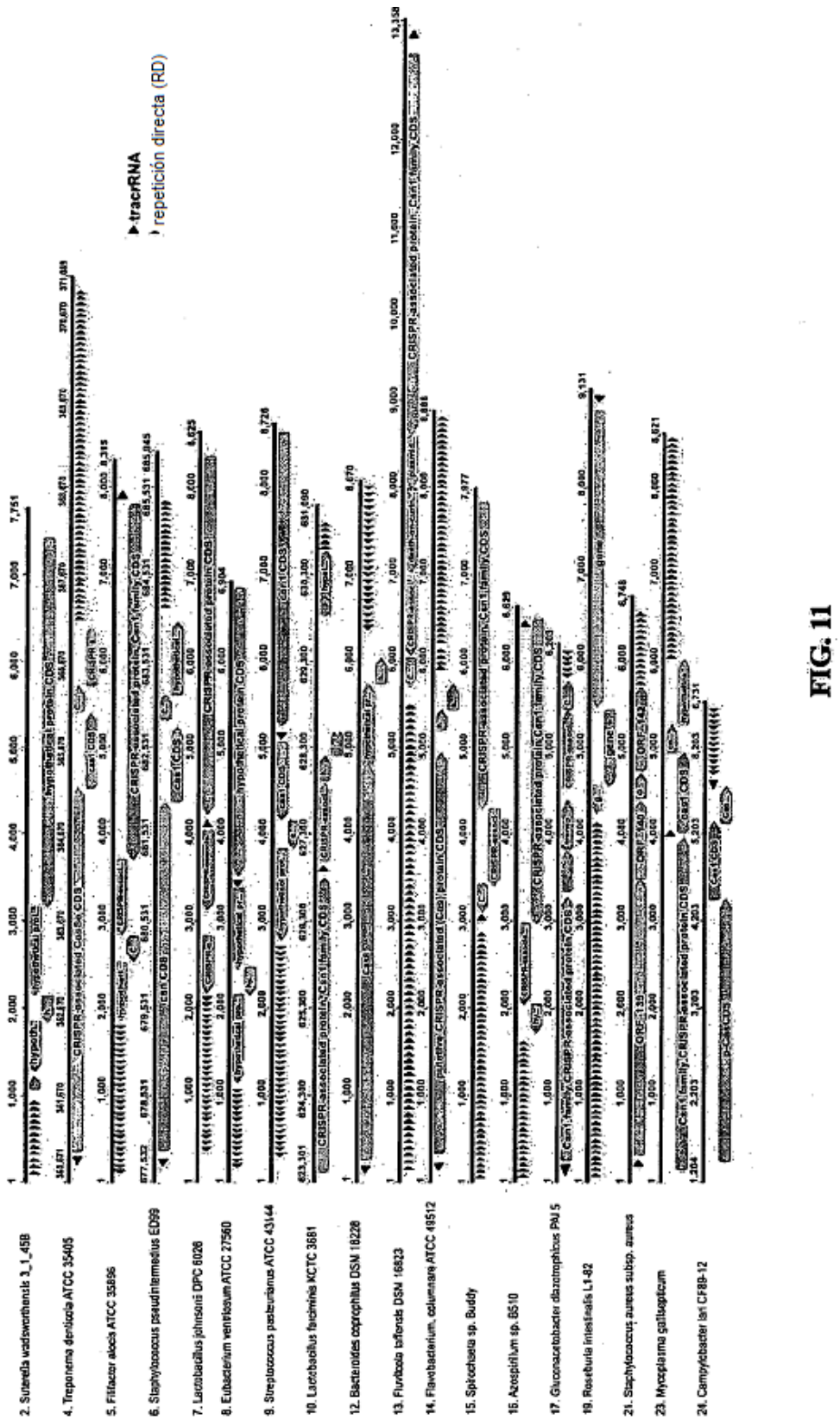

2. Sutterella wadsworthensis 3_1_45B
4. Treponema denticola ATCC 35405
5. Filifactor alocis ATCC 35896
6. Staphylococcus pseudintermedius ED99
7. Lactobacillus johnsonii DPC 6026
8. Eubacterium ventriosum ATCC 27560
9. Streptococcus pasteurianus ATCC 43144
10. Lactobacillus farciminis KCTC 3681
12. Bacteroides coprophilus DSM 18228
13. Fluviicola taffensis DSM 16823
14. Flavobacterium. columnare ATCC 49512
15. Spirochaeta sp. Buddy
16. Acetohalobium sp. 8510
17. Gluconacetobacter diazotrophicus PAl 5
19. Roseburia intestinalis L1-82
21. Staphylococcus aureus subsp. aureus
23. Mycoplasma gallisepticum
24. Campylobacter lari CF89-12
▶ tracrRNA
repetición directa (RD)

FIG. 11

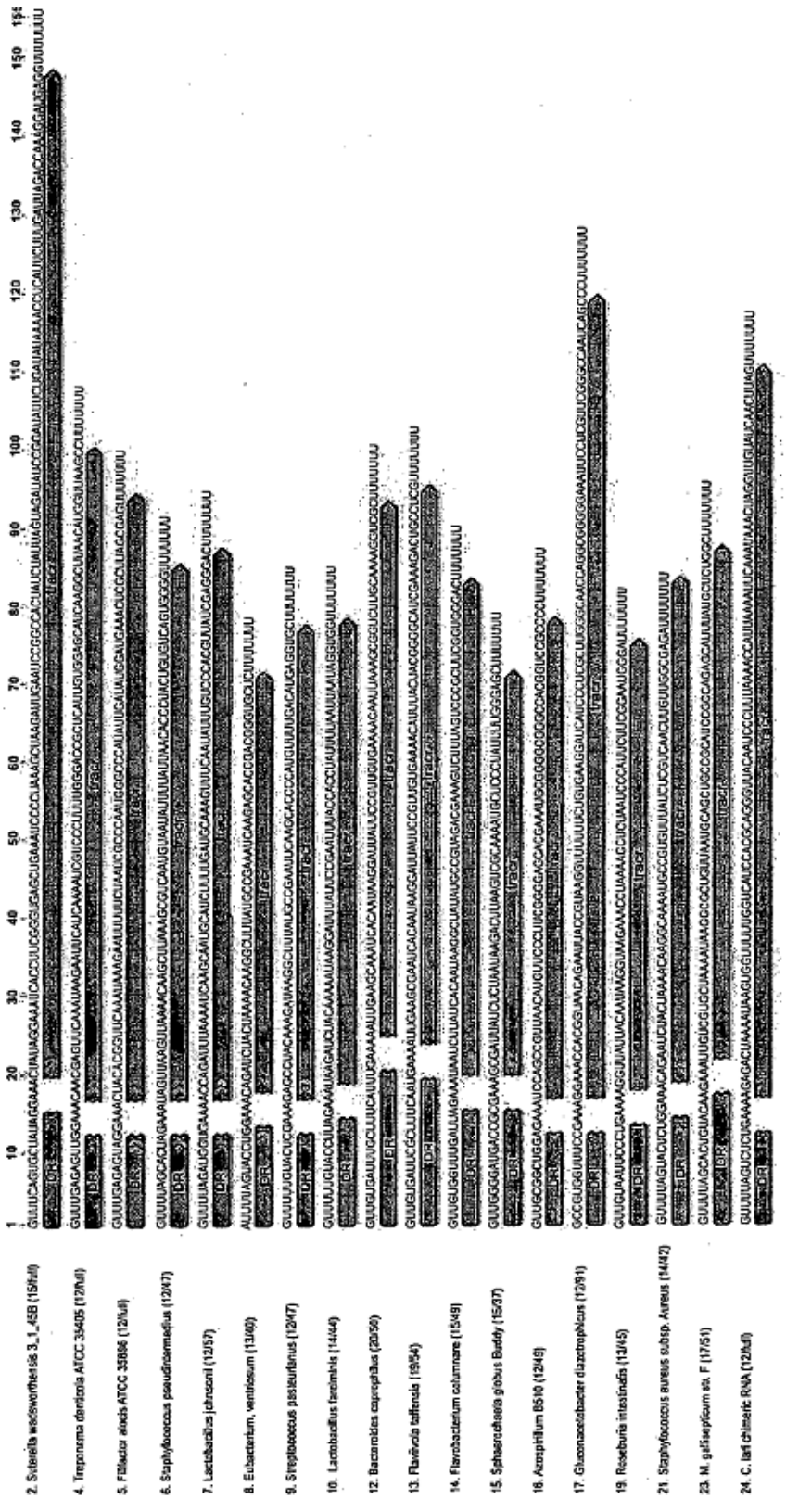

2. Sutterella wadsworthensis 3_1_45B (15/full)
4. Treponema denticola ATCC 35405 (12/full)
5. Filifactor alocis ATCC 35896 (12/full)
6. Staphylococcus pseudintermedius (12/47)
7. Lactobacillus johnsonii (12/57)
8. Eubacterium, ventriosum (13/60)
9. Streptococcus pasteurianus (12/47)
10. Lactobacillus farciminis (14/44)
12. Bacteroides coprophilus (20/56)
13. Flavivirga taiwania (19/54)
14. Flavobacterium columnare (15/49)
15. Sphaerochaeta globus Buddy (15/37)
18. Azospirillum B510 (12/49)
17. Gluconacetobacter diazotrophicus (17/91)
19. Roseburia intestinalis (13/45)
21. Staphylococcus aureus subsp. Aureus (14/42)
23. M. gilvisepticum str. F (17/51)
24. C. lari chinese RNA (12/full)

FIG. 12

FIG. 13A

Especie: Corynebacter diphtheriae	
Secuenciación código de barras	AGCAATTC
ID	1
ARN guía quimérico	NNNNNNNNNNNNNNNNNNNNACUGGGGUUCAGGAAACUGAACCUCAGUAAGCAUUGGCUCGUUUCCAAUG UUGAUUGCUCCCGCCGGUGCUCCCUUAUUUUUUAAGGGCGCCGGCUUUUUUUU
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRLYRRKRRRLQQLDK FIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKA IREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAREIQEICRMQEIGQELYRKIIDVVFAAESP KGSASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEI LGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAKV QAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPAVDR VLKTVSRWLESATKTWGAPERVIIEHVYREGFVTEKRARAREMDGDMRRRAARNAKLFQEMQEKLNVQGKPSRADLWR YQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVE RTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVYRGSLTA EARRASGISGKLKFFDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRA AWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTR EPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAFAMLRVYTID LLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRV DGFFSPSKLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAHLPVTWKVQ	
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGAAGTACCATGTCGGAATCGATGTCCGGAACCTTTTCTGTGGGGCTGGCTGCTATTGAAGTGGATGACGCTGG AATGCCTATTAAGACCCTGAGTCTGGTGTCACACATTCATGACTCAGGACTGGATCCTGACGAGATCAAGAGCGC TGTGACCAGGCTGGCAAGCTCCGGAATCGCCCGGAGAACAAGGCGCCTGTACCGACGGAAGAGAGGCGCCTGC AGCAGCTGGATAAGTTCATCCAGAGGCAGGGCTGGCCAGTGATCGAGCTGGAAGATTACAGCGACCCCCTGTAT CCTTGGAAGGTGCGCGCCGAACTGGCCGCTTCTTATATTGCTGACGAGAAGGAACGGGGGGAGAAACTGAGTGT GGCTCTGAGACACATCGCAAGGCATCGGGATGGAGGAACCCTTACGCCAAGGTGTCTAGTCTGTATCTGCCAGA TGGCCCCCTCAGACGCCTTCAAGGCTATTAGGGAGGAAATCAAACGCGCTAGCGGCCAGCCTGTGCCAGAGACTG CAACCGTCGGGCAGATGGTGACCCTGTGCGAACTGGGCACACTGAAGCTGCCGAGGAGAGGAGGAGTGCTGAGT GCACGGCTGCAGCAGTCAGATTACGCCCGCGAGATCCAGGAAATTTGTCGAATGCAGGAGATCGGCCAGGAACT GTATCGCAAGATCATTGACGTGGTGTTCGCAGCCGAGTCCCCAAAGGGCTCTGCCTCAAGCCGGGTGGGGAAAG ATCCTCTGCAGCCAGGAAAGAACAGAGCACTGAAAGCCAGCGACGCTTTTCAGCGATACCGGATTGCTGCACTG	

FIG. 13B

	ATCGGCAATCTGAGAGTCAGGGTGGATGGGGAGAAGAGGATTCTGAGCGTGGAGGAGAAGAACCTGGTGTTCGA CCACCTGGTGAATCTCTGACTCCAAAGAAAGAGCCCGAATGGGTGACCATCGCCGAAATTCTGGGCATCGATCGCG GGCAGCTGATCGGAACAGCTACTATGACCGACGATGGAGAGCGAGGAGCCCGACCCCCTACACACGATACT AACAGAAGTATTGTGAACAGCCCGATCGCACCACTGGTCGACTGGTGGAAAACAGCTAGCGCACTGGAGCAGCA CGCCATGGTGAAGGCACTGTCCAACGCCGAAGTCGACGATTTTGATTCTCCCGAGGGAGCAAAAGTGCAGGCAT TCTTTGCCGATCTGGACGATGACGTCCACGCCAAGCCTGGACAGCCTGCATCTGCCTGTGGGACGAGCAGCTTACT CCGAGGACACTCTGGTCAGACTGACCCGACGGATGCTGAGTGATGGGGTGGACCTGTATACCGCCCCGGCTGCAG GAGTTCGGAATTGAACCTAGCTGGACCCCACCCACACCAAGAATCGGAGAGCCTGTCGGCAATCCAGCCGTCGA CCGGGTGCTGAAAACAGTGAGCAGATGGCTGGAATCCGCAACAAAGACTTGGGGCGCCCCAGAGAGGGTCATCA TTGAGCACGTGCGCGAAGGCTTCGTCACTGAGAAACGCGCTCGAGAAATGGATGGGACATGAGAAGGCGCGCA GCCCGGAACGCCAAGCTGTTTCAGGAGATGCAGGAAAAGCTGAATGTGCAGGGCAAACCCAGTCGAGCCGATCT GTGGAGATACCAGTCAGTGCAGAGACAGAACTGCCAGTGTGCCTATTGCGGGTCCCCAATTACCTTTTCTAATAG TGAAATGGACCACATCGTGCCCAGAGCAGGGCAGGGATCCACCACACACAAGGAGAATCTGGTCGCCGTGTGCC ATCGCTGTAACCAGTCTAAGGGCAATACACCCTTCGCTATTTGGGCAAAAAACACTTCTATCGAAGGGGTCAGTG TGAAGGAGGCCGTGGAACGGACCAGACATTGGGTCACTGATACCGGCATGAGAAGCACTGACTTCAAGAAGTTC ACCAAGGCTGTGGTCGAGCGGTTTCAGAGAGCAACAATGGATGAGGAAATCGACGCCAGAAGCATGGAATCCGT CGCCTGGATGGCTAATGAGCTGAGGAGCCGCGTGGCTCAGCACTTCGCATCCCATGGAACCACAGTCAGGGTGT ACCGAGGCAGCCTGACAGCAGAGGCTCGACGGGCATCTGGGATCAGTGGAAAGCTGAAATTCTTTGATGGCGTG GGGAAGTCCAGGCTGGATAGAAGGCACCATGCTATTGACGCTGCAGTGATCGCATTCACCTCTGACTATGTGGCC GAAACACTGGCTGTCCGCTCAAACCTGAAACAGAGCCAGGCCCACCGACAGGAGGCTCCTCAGTGGAGAGAGTT CACCGGCAAGGATGCAGAGCATCGAGCAGCTTGGAGAGTGTGGTGCCAGAAGATGGAAAAACTGAGCGCCCTGC TGACCGAGGACCTGCGAGATGACCGGGTGGTCGTGATGTCTAACGTGCGACTGCGGCTGGGAAATGGCAGTGCC CACAAGGAAACCATTGGCAAACTGTCAAAGGTGAAACTGTCCTCTCAGCTGTCAGTCAGCGATATCGACAAAGC AAGTTCAGAGGCCCTGTGGTGTGCTCTGACCAGAGAGCCCGGATTCGATCCTAAGGAAGGCCTGCCCGCTAACCC TGAGAGACACATCAGGGTGAATGGGACACATGTCTACGCCGGGGACAATATTGGACTGTTTCCAGTGTCAGCAG GAAGCATCGCACTGAGGGGAGGATACGCAGCTGGGCAGCTCCTTCCACCATGCTCGCGTGTATAAAATTTACTT CCGGCAAGAAACCCGCATTTGCCATGCTGAGGGTGTACACCATCGATCTGCTGCCTTATCGCAACCAGGACCTGT TTAGCGTGGAACTGAAGCCACAGACAATGTCCATGAGGCAGGCTGAGAAGAAACTGCGCGACGCTCTGGCAACT GGGAATGCAGAATATCTGGGATGGCTGGTCGTGGATGACGAGCTGGTCGTGGATACATCTAAGATTGCCACTGA CCAGGTCAAAGCAGTGGAGGCCCGAACTGGGGACTATCCGCCGGATGGCGGGTGGATGGATTCTTTTCCCCCTCTAA ACTGAGACTGAGGCCTCTGCAGATGTCCAAGGAGGGGATCAAGAAAGAGTCCGCTCCCGAACTGTCTAAAATCA TTGACAGACCAGGATGGCTGCCCGCCGTGAACAAGCTGTTCTCTGATGGAAATGTCACCGTCGTGCGGAGAGACT CTCTGGGACGCGTGCGACTGGAGAGTACAGCCCACCTGCCTGTCACTTGGAAGGTGCAGTAAGAATTC
Longitud de proteína	1084

FIG. 13C

sin NLS y marcador HA añadido	
PAM	**NGG**
Longitud del espaciador	28
Especie	Sutterella_wadsworthensis
Secuenciación código de barras	AACTTGAC
ID	2
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNGUUUCAGUGCUAUAGGAAACUAUAGGAAAUCACCUUCGGGUGAGCUGAAA UCCCCUAAAGCUAAGAUUGAAUCCGGCCACUAUCUAUUAGUAGAUAUCCGGAUAUUCUGAUAUAAAACCU CAUUCUUUGAUUAGACCAAAGGAUGAGGUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MTQSERRFSCSIGIDMGAKYTGVFYALFDREELPTNLNSKAMTLVMPETGPRYVQAQRTAVRHRLRGQKRYTLARKL AFLVVDDMIKKQEKRLTDEEWKRGREALSGLLKRRGYSRPNADGEDLTPLENVRADVFAAHPAFSTYFSEVRSLAEQ WEEFTANISNVEKFLGDPNIPADKEFIEFAVAEGLIDKTEKKAYQSALSTLRANANVLTGLRQMGHKPRSEYFKAIEAD LKKDSRLAKINEAFGGAERLARLLGNLSNLQLRAERWYFNAPDIMKDRGWEPDRFKKTLVRAFKFFHPAKDQNKQH LELIKQIENSEDIIETLCTLDPNRTIPPYFDQNNRRPPLDQTLLLSPEKLTRQYGEIWKTWSARLTSAEPTLAPAAEILERS TDRKSRVAVNGHEPLPTLAYQLSYALQRAFDRSKALDPYALRALAAGSKSNKLTSARTALENCIGGQNVKTFLDCAR RYYREADDAKVGLWFDNADGLLERSDLHPPMKKKILPLLVANILQTDETTGQKFLDEIWRKQIKGRETVASRCARIET VRKSFGGGFNIAYNTAQYREVNKLPRNAQDKELLTIRDRVAETADFIAANLGLSDEQKRKFANPFSLAQFYTLIETEVS GFSATTLAVHLENAWRMTIKDAVINGETVRAAQCSRLPAETARPFDGLVRRLVDRQAWEIAKRVSTDIQSKVDFSNGI VDVSIFVEENKFEFSASVADLKKNKRVKDKMLSEAEKLETRWLIKNERIKKASRGTCPYTGDRLAEGGEIDHILPRSLI KDARGIVFNAEPNLIYASSRGNQLKKNQRYSLSDLKANYRNEIFKTSNIAAITAEIEDVVTKLQQTHRLKFFDLLNEHE QDCVRHALFLDDGSEARDAVLELLATQRRTRVNGTQIWMIKNLANKIREELQNWCKTTNNRLHFQAAATNVSDAKN LRLKLAQNQPDFEKPDIQPIASHSIDALCSFAVGSADAERDQNGFDYLDGKTVLGLYPQSCEVIHLQAKPQEEKSHFDS VAIFKEGIYAEQFLPIFTLNEKIWIGYETLNAKGERCGAIEVSGKQPKELLEMLAPFFNKPVGDLSAHATYRILKKPAYE FLAKAALQPLSAEEKRLAALLDALRYCTSRKSLMSLFMAANGKSLKKREDVLKPKLFQLKVELKGEKSFKLNGSLTLP VKQDWLRICDSPELADAFGKPCSADELTSKLARIWKRPVMRDLAHAPVRREFSLPAIDNPSGGFRIRRTNLFGNELYQ VHAINAKKYRGFASAGSNVDWSKGILFNELQHENLTECGGRFITSADVTPMSEWRKVVAEDNLSIWIAPGTEGRRYV RVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQKAVGTELSELLGQPRSEIFIENVGNAKHIRFWYIVVSSN KKMNESYNNVSKS

FIG. 13D

Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGACTCAGAGCGAGCGACGAATTTTCTTGCAGCATTGGCATTGACATGGGGGGCTAAGTACACTGGGGGTGTTCTA CGCACTGTTCGACCGGGAGGAACTGCCCACAAACCTGAACAGCAAGGCCATGACCCTGGTCATGCCTGAGACAG GGCCAAGATACGTGCAGCACAGAGAACTGCCGTCAGACACAGGCTGCGCGGACAGAAGAGATATACCCTGGCT AGGAAACTGGCATTTCTGGTGGTCGACGATATGATCAAGAAACAGGAAAAGAGGCTGACTGATGAGGAATGGAA ACGAGGACGGGAGGCCCTGTCCGGCCTGCTGAAGCGGAGAGGGTACTCTCGGCCCAACGCTGACGGCGGAAGATC TGACCCCTCTGGAGAATGTGAGAGCAGACGTGTTCGCCGCTCATCCTGCCTTCAGCACATATTTTTCCGAAGTGC GCTCTCTGGCTGAGCAGTGGGAGGAGTTCACCGCAAACATCAGCAATGTCGAGAAGTTTCTGGGCGACCCAAAC ATCCCCGCCGATAAAGAGTTCATTGAATTTGCCGTGGCTGAAGGGCTGATTGACAAGACCGAGAAGAAGCCTA CCAGTCAGCTCTGAGCACCCTGAGGGCAAACGCCAATGTGCTGACAGGACTGCGGCAGATGGGCCACAAGCCTA GATCAGAATATTTTAAAGCAATCGAGGCCGACCTGAAGAAAGATAGCCGCCTGGCCAAGATTAACGAAGCATTC GGAGGAGCAGAGCGCCTGGCTCGACTGCTGGGAAACCTGTCCAATCTGCAGCTGCGGGCAGAAAGATGGTACTT CAATGCCCCCGACATCATGAAGGATAGGGGCTGGGAGCCTGATCGCTTCAAGAAAACACTGGTGCGGGCTTTTA AGTTCTTTCACCCAGCAAAGGACCAGAACAAACAGCATCTGGAACTGATCAAACAGATTGAGAACAGCGAAGAT ATCATTGAGACTCTGTGCACCCTGGACCCAAACAGAACCATCCCCCCTTACGAGGATCAGAACAATAGGCGCCCA CCCCTGGACCAGACTCTGCTGAGTCCCGAAAAAGCTGACCCGGCAGTATGGCGAGATCTGGAAAACATGGAG CGCCAGACTGACCTCCGCTGAACCCACACTGGCACCTGCAGCCGAGATTCTGGAAAGATCTCACCGACAGGAAGA GTCGCGTGGCAGTCAACGGACACGAGCCACTGCCTACACTGGCTTACCAGCTGAGTTATGCACTGCAGAGAGCCT TCGACAGGTCAAAAGCCCTGGATCCATATGCTCTGAGGGCACTGGCTGCAGGCTCAAAAAGCAATAAGCTGACA TCCGCCCGCACTGCTCTGGAGAACTGCATCGGAGGCCAGAATGTGAAAACCTTCCTGGACTGTGCCCGACGGTAC TATCGGGAAGCAGACGATGCCCAAAGTCGGGCTGTGGTTCGACAACGCCGATGGACTGCTGGAGAGATCTGACCT GCATCCTCCAATGAAGAAAAAGATCCTGCCCCTGCTGGTGGCCAATATTCTCTGCAGACAGATGAAACCACACAGGCC AGAAGTTTCTGGACGAGATCTGGCGAAAACAGATTAAGGGGCGGGAAACTGTGGCTAGCCGATGTGCACGGATC GAGACAGTGCGGAAATCCTTCGGGGGAGGCTTTAACATTGCCTACAATACCGCTCAGTATAGGGAGGTGAACAA GCTGCCCCGCAATGCCCAGGATAAAGAACTGCTGACAATCAGAGATAGGGTGGCTGAGACTGCAGACTTCATTG CCGCTAACCTGGGGCTGTCTGACGAGCAGAAAAGAAAGTTCGCCAATCCTTTTAGTCTGGCTCAGTTCTACACCC TGATCGAGACAGAAGTGTCCGGATTTTCTGCAACTACCCTGGCCGTCCACCTGGAGAACGCCTGGAGGATGACAA TCAAGGATGCTGTGATTAATGGGGAAACTGTCAGAGCAGCACAGTGCAGCAGGCTGCCTGCAGAGACAGCTCGC CCATTCGATGGACTGGTGAGAAGGCTGGTCGACAGACAGGCTTGGGAGATCGCAAAGAGGGTGTCAACTGACAT TCAGAGCAAAGTCGATTTCTCCAACGGCATCGTGGACGTCAGCATTTTTGTGGAGGAAAATAAGTTCGAGTTTTC CGCATCTGTGGCCGATCTGAAAAAGAACAAACGGGTCAAAGACAAGATGCTGTCCGAGGCCGAAAAGCTGGAA ACCAGATGGCTGATCAAAAATGAGCGGGATCAAGAAGGCCAGCCGGGGAACTTGTCCCTACACCGGCGATAGGCT GGCTGAGGGGGGAGAAATCGACCACATTCTGCCCCGAAGCCTGATCAAGGATGCCCGGGGAATTGTGTGTTTAACG CTGAGCCTAATCTGATCTATGCAAGCTCCCGCGGCAACCAGCTGAAAAAGAATCAGCGATACAGTCTGTCAGATC TGAAGGCCAACTATCGGAATGAGATCTTCAAAACTAGCAACATCGCTGCAATTACCGCCGAGATTGAGGACGTG

FIG. 13E

	GTCACTAAGCTGCAGCAGACCCATAGACTGAAATTCTTTGATCTGCTGAATGAGCACGAACAGGACTGCGTGCGG CACGCCCTGTTCCTGGACGATGGCAGCGAAGCTCGCGACGCAGTGCTGGAGCTGCTGGCAACACAGCGCCGAAC TCGCGTCAACGGGACACAGATCTGGATGATTAAGAACCTGGCCAACAAGATCCGAGAGGAACTGCAGAATTGGT GTAAGACAACTAACAATAGACTGCACTTTCAGGCCGCTGCAACTAACGTGTCCGATGCTAAAGAATCTGAGGCTG AAACTGGCCCAGAACCAGCCCGACTTCGAGAAGCCAGATATCCAGCCCATTGCCAGCCATTCCATCGACGCCCTG TGCTCTTTCGCTGTGGGGAGTGCTGACGCAGAACGCGATCAGAATGGATTTGACTACCTGGATGGCAAGACCGTG CTGGGACTGTATCCACAGAGCTGTGAGGTCATTCACCTGCAGGCCAAGCCCCAGGAGGAAAAAAGTCATTTCGA TTCAGTGGCTATCTTTAAGGAAGGCATCTACGCAGAGCAGTTCCTGCCTATCTTTACCCTGAACGAAAAGATCTG GATTGGATATGAGACACTGAATGCCAAAGGCGAAAGATGCCGGGGCTATTGAGGTGAGTGGCAAACAGCCAAAG GAGCTGCTGGAAATGCTGGCCCCCTTCTTTAACAAGCCTGTGGGCGACCTGTCAGCCCACGCTACTTACCGGATC CTGAAAAAGCCTGCATATGAGTTTCTGGCAAAGGCAGCTCTGCACAGCCACTGAGCGCAGAGGAAAAAAGACTGGC AGCCCTGCTGGATGCTCTGCGCTACTGTACCAGTCGAAAGTCACTGATGAGCCTGTTCATGGCTGCAAACGGGAA ATCCCTGAAAAAGCGGGAGGACGTGCTGAAACCCAAGCTGTTCCAGCTGAAGGTCGAAGGCGAAAAG AGCTTCAAGCTGAACGGGAGCCTGACCCTGCCTGTGAAGCAGGACTGGCTGAGAATCTGCGATAGCCCAGAACT GGCAGACGCCTTTGGCAAACCCTGTTCCGCCGATGAGCTGACATCTAAGCTGGCTCGCATTTGGAAACGACCTGT GATGCGGGATCTGGCTCATGCACCAGTCCGGAGAGAGTTCAGCCTGCCCGCAATCGACAACCCAAGTGGAGGGT TCAGGATTAGGCGCACCAACCTGTTTGGCAATGAGCTGTACCAGGTGCACGCCATCAACGCTAAAAAGTATCGCG GCTTCGCCTCCGCTGGGTCTAATGTCGACTGGTCCAAGGGGATCCTGTTTAACGAGCTGCAGCATGAAAATCTGA CCGAGTGCGGAGGCAGGTTCATTACAAGCGCCGATGTGACTCCTATGTCCGAATGGCGCAAGGTGGTCGCAGAG GACAACCTGTCTATCTGGATTGCTCCAGGGACAGAAGGACGGTACGTGAGGGTCGAGACAACATTCATCCA GGCCAGTCACTGGTTTGAACAGTCAGTGGAGAATTGGGCCATTACTAGTCCTCTGTCACTGCCAGCTTCCTTCAA GGTGGACAAACCAGCTGAGTTTCAGAAGGCAGTCGGAACCGAGCTGTCAGAACTGCTGGGCCAGCCCAGGAGCG AAATCTTCATTGAGAACGTGGGCAATGCCAAGCATATCCGCTTTTGGTACATTGTGGTGAGCAGCAACAAAAAGA TGAACGAGTCTTACAACAATGTGTCTAAGAGTTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1422
PAM	**NGG**
Longitud de espaciador	35-39
Especie:	**Legionella_pneumophila_Paris**

FIG. 13F

Secuenciación código de barras	TGTCGGAT
ID	3
ARN guía químérico	NNNNNNNNNNNNNNNNNNNNGUUUCAGUGGUGAAAACCCUGAAAUCAACAAAAUUAAAGAUUGAAUCGUU UUUCUAUGCUCGUCUUAAAUAGCGAGCAUAUAACGAUUUUUUU
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRATRHIRVRNKKRNQFVKRVALQ LFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKILVSKV EKKLDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSNLLQWKNLH RYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEARTNTGMEKDQS LLLNPEKLNNLYPNWRNLIPGIIDAHPFLEKDLEHTKLRDRKRIISPSKQDEKRDSYILQRYLDLNKKIDKFKIKKQLSFL GQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKERFDAAQGIWFDNAFSLCELSNINPPRKQKILPLLVGAILSE DFINNKDKWAKFKIFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNNKALIKIIQTIPDIIQAIQSHL GHNDSQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISYASRLPADSVRPFDGVLARMMQR LAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQWEEKPQRIINASMNICPYKG ASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEEHYLLEHLSPLYLKHQFGTDNVSDIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQIT AEEVHIDHRELLSKQEPKLVKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDNSWFINHLMPDEVHLNPVRSKEKYNKP NISSTPLFKDSLYAERFIPVWVKGETFAIGFSEKDLFEIKPSNKEKLFTLLKTYSTKNPGESLQELQAKSKAKWLYFPINK TLALEFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESITVKILKEPMPVLSVKFESSKKNVLGSFKHTIALPATKDWER LFNHPNFLALKANPAPNPKEFNEFIRKYFLSDNNPNSDIPNNGHNIKPQKHKAVRKVFSLPVIPGNAGTMMRIRRKDNK GQPLYQLQTIDDTPSMGIQINEDRLVKQEVLMDAYKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEIIKLEMKP HSKTRRYIRITQSLADFIKTIDEALMIKPSDSIDDPLNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQ WIQTLYVTQLKKQP
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGGTC CATGGAGTCCTCTCAGATCCTGTCACCTATCGGCATCGACCTGGGCGGAAAGTTTACTGGAGTCTGCCTGTCACA CCTGGAAGCATTTGCTGAGCTGCCTAACCACGCAAATACAAAGTACTCAGTGATCCTGATTGATCATAACAATTT CCAGCTGAGCCAGGCACAGCGGAGAGCCACTAGACACACAGGGTGCGCAACAAGAAAAAGAAATCAGTTCGTGAAG AGGGTCGCCCTGCAGCTGTTTCAGCACATCCTGTCCCGGGACCTGAACGCCAAGGAGGAAACCGCTCTGTGCCAT TACCTGAACAATAGAGGCTACACCTATGTGGACACAGATCTGGACGAGTATATCAAAGATGAAACCACAATTAA CCTGCTGAAGGAGCTGCTGCCAAGCGAGTCCGAACATAATTTCATTGACTGGTTTCTGCAGAAGATGCAGAGCTC CGAGTTCCGGAAGATCCTGGTGAGCAAAGTCGAGGAAAAGAAAGACGATAAAGAACTGAAGAACGCCGTGAAG AACATCAAGAACTTCATCACCGGATTCGAGAAAAATAGCGTGGAAGGCCACCGACATCGGAAGGTCTACTTTGA GAACATCAAGTCCGATATTACAAAAGACAATCAGCTGGATTCTATCAAGAAAAAGATTCCTAGCGTGTCTGTC

FIG. 13G

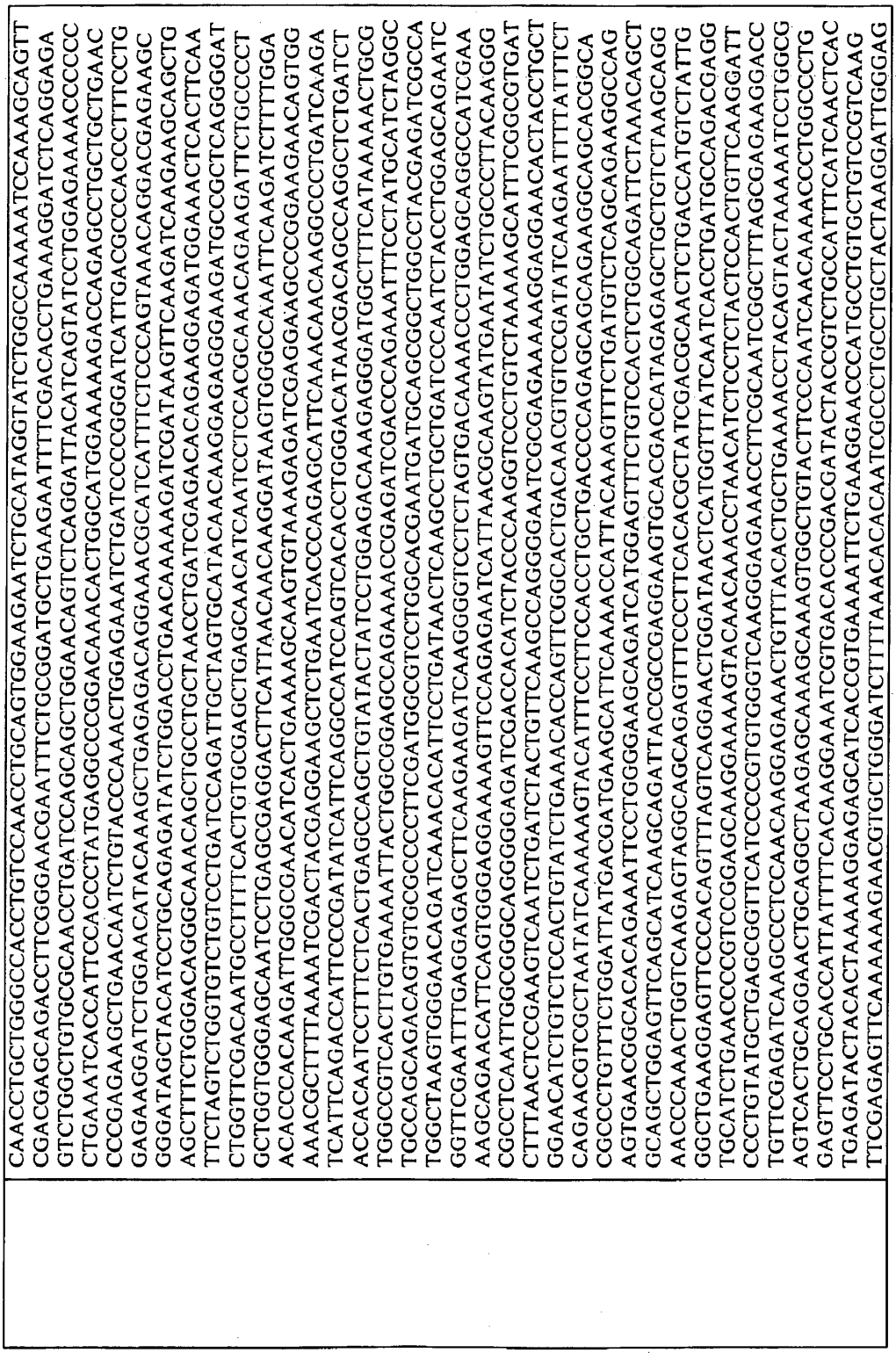

CAACCTGCTGGGCCACCTGTCCAACCTGCAGTGGAAGAATCTGCATAGGTATCTGGCCAAAAATCCAAAGCAGTT
CGACGAGCAGACCTTCGGGAACGAATTTTCTGCGGATGCTGAAGAATTTTCGACACCTGAAAGGATCTCAGGAGA
GTCTGGCTGTGCGCAACCTGATCCAGCAGCTGGAACAGTCTCAGGATTACATCATCAGTATCCTGGAGAAAACCCCCC
CTGAAATCACCATTCCACCCTATGAGGCCCGGACAAACACTGGCATGGAAAAAGACCAGAGCCTGCTGCTGAAC
CCCGAGAAGCTGAACAATCTGTACCCAAACTGGAGAAATCTGATCCCCGGGATCATTGACGCCCACCCTTTCCTG
GAGAAGGATCTGGAACATACAAAGCTGAGAGACAGGAAACGCATCATTTCTCCCAGTAAACAGGACGAGAAGC
GGGATAGCTACATCCTGCAGAGATATCTGGACCTGAACAAAAAGATCGATAAGTTCAAGATCAAGAAGCAGCTG
AGCTTTCTGGGACAGGGCAAACAGCTGCCTGCTAACCTGATCGAGACACAGAAGGAGATGGAAACTCACTTCAA
TTCTAGTCTGGTGTCTGTCCTGATCCAGATTGCTAGTGCATACAACAAGGAGAGGGAAGATGCCGCTCAGGGGAT
CTGGTTCGACAATGCCTTTTCACTGTGCGAGCTGAGCAACATCAATCCTCCACGCAAACAGAAGATTCTCTGCCCCT
GCTGGTGGGAGCAATCCTGAGCGAGGACTTCATTAACAACAAGGATAAGTGGGCCAAATTCAAGATCTTTTGGA
ACACCCACAAGATTGGGCGAACATCACTGAAAAGCAAGTGTAAAGAGATCGAGGAAGCCCGGAAGAACAGTGG
AAACGCTTTTAAAATCGACTACGAGGAAGCTCTGAATCACCCAGAGCATTCAAACAACAAGGCCCTGATCAAGA
TCATTCAGACCATTCCCGATATCATTCAGGCCATCCAGTCACACCTGGGACATAACGACAGCCAGGCTCTGATCT
ACCACAATCCTTTCTCACTGAGCCAGCTGTATACTATCCTGGAGACAAAGAGGGATGGCTTTCATAAAAACTGCG
TGGCCGTCACTTGTGAAAATTACTGGCGGAGCCAGAAAACCGAGATCGACCCAGAAATTTCCTATGCATCTAGGC
TGCCAGCAGACAGTGTGCGCCCCTTCGATGGCGTCCTGGCACGAATGATGCAGCGGCTGGCCTACGAGATCGCCA
TGGCTAAGTGGGAACAGATCAAACACATTCCTGATAACTCAAGCCTGCTGATCCCAATCTACCTGGAGGCAGAATC
GGTTCGAATTTGAGGAGAGCTTCAAGAAGATCAAGGGGTCCTCTAGTGACAAAACCCTGGAGCAGGCCATCGAA
AAGCAGAACATTCAGTGGGAGGAAAAGTTCCAGAGAATCATTAACGCAAGTATGAATATCTGCCCTTACAAGGG
CGCCTCAATTGGCGGCAGGGGGAGATCGACCACATCTACCCAAGGTCCCTGTCTAAAAAGCATTTCGGCGTGAT
CTTTAACTCCGAAGTCAATCTGATCTACTGTTCAAGCCAGGGGAAATCGCGAGAAAAAGGAGGAACACTACCTGCT
GGAACATCTGTCTCCACTGTATCTGAAACACCAGTTCGGCACTGACAACGTGTCCGATATCAAGAATTTTATTTCT
CAGAACGTCGCTAATATCAAAAAGTACATTTCCTTCCACCTGCTGACCCCAGAGCAGCAGAAGGCAGCACGGCA
CGCCCTGTTTCTGGATTATGACGATGAAGCATTCAAAAACCATTACAAAGTTTCTGATGTCTCAGCAGAAGGCCAG
AGTGAACGGCACACAGAAATTCCTGGGGAAGCAGATCATGGAGTTTCTGTCCACTCTGGCAGATTCTAAACAGCT
GCAGCTGGAGTTCAGCATCAAGCAGATTACCGCCGAGGAAGTGCACGACCATAGAGAGCTGCTGTCTAAGCAGG
AACCCAAACTGGTCAAGAGTAGGCAGCAGAGTTTCCCTTCACACGCTATCGACGCAACTCTGACCATGTCTATTG
GGCTGAAGGAGTTCCCACAGTTTAGTCAGGAACTGGATAACTCATGGTTTATCAATCACCTGATGCCAGACGAGG
TGCATCTGAACCCCGTCCGGAGCAAGGAAAAGTACAACAAACCTAACATCTCCTCTACTCCACTGTTCAAGGATT
CCCTGTATGCTGAGCGGTTCATCCCCGTGTGGGTCAAGGGAGAAACCTTCGCAATCGGCTTTAGCGAGAAGGACC
TGTTCGAGATCAAGCCCTCCAACAAGGAGAAACTGTTTACACTGCTGAAAACCTACAGTATCTAAAAAATCCTGGCG
AGTCACTGCAGGAACTGCAGGCTAAGAGCAAAGTGGCTGTACTTCCCAATCAACAAAACCCTGGCCCTG
GAGTTCCTGCACCATTATTTTCACAAGGAAATCGTGACACCCGACGATACTACCGTCTCGCCATTTCATCAACTCAC
TGAGATACTACACTAAAAAGGAGAGCATCACCGTGAAAATTTCTGAAGGAACCCATGCCTGTGTCCGTCAAG
TTCGAGAGTTCAAAAAAGAACGTGCTGGGATCTTTTAAACACACAATCGCCCTGCCTGCTACTAAGGATTTGGGAG

FIG. 13H

Longitud de proteína (sin NLS y marcador HA añadido)	AGGCTGTTCAACCATCCAAATTTTCTGGCACTGAAGGCCAACCCAGCTCCCAATCCTAAAGAGTTCAATGAGTTC ATCCGCAAGTACTTCCTGAGCGACAACAATCCCAACTCCGATATCCCTAACAATGGCCACAATATCAAGCCCCAG AAACATAAGGCCGTGCGAAAGGTCTTTAGCCTGCCAGTGATCCCCGGGAACGCTGGAACCATGATGCGCATTAG GCGCAAAGACAATAAGGGACAGCCACTGTATCAGCTGCAGACAATCGACGATACTCCCAGCATGGGCATCCAGA TTAACGAGGATCGCCTGGTGAAACAGGAAGTCCTGATGGACGCCTACAAGACACGAAATCTGAGCACTATCGAT GGGATTAACAATTCCGAGGGACAGGCATATGCCACATTCGACAACTGGCTGACCCTGCCCGTGAGCACCTTCAAG CCTGAGATCATCAAGCTGGAAATGAAGCCTCACTCTAAAACCCGACGGTACATCAGAATTACACAGAGTCTGGC CGACTTCATCAAAACTATTGATGAGGCTCTGATGATCAAGCCCAGTGACTCAATTGACGATCCTCTGAACATGCC AAATGAGATCGTGTGTAAAAACAAGCTGTTCGGGAATGAACTGAAGCCTAGGGATGGAAAAATGAAGATCGTGA GCACTGGCAAGATTGTCACCTACGAGTTTGAAAGCGACTCCACCCCCCAGTGGATCCAGACCCTGTATGTGACAC AGCTGAAAAAGCAGCCTTAAGAATTC
Longitud del espaciador	1372
PAM	**NGG**
	34-36
Especie: **Treponema_denticola**	
Secuenciación código de barras	TCGCCTTG
ID	4
Secuenciación código de barras	**NNNNNNNNNNNNNNNNNNNGUUUGAGAGUUGGAAACAACGAGUUCAAAUAAGAAUUCAUCAAAAUCGUC CCUUUUGGGACCGCUCAUUGUGGAGCAUCAAGGCUUAACAUGGUUAAGCCUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRCFETAETAEVRRLHRGARRRIERRKKRIKLLQEL FSQEIAKTDEGFFQRMKESPFYAEDKTILQENTLFNDKDFADKTYHKAYPTINHLIKAWIENKVKPDPRLLYLACHNIIK KRGHFLFEGDFDSENQFDTSIQALFEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQSRLNKILGLKPSDKQKKAITN LISGNKINFADLYDNPDLKDAEKNSISFSKDDFDALSDDLASILGDSFELLLKAKAVYNCSVLSKVIGDEQYLSFAKVKI YEKHKTDLTKLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVGVCKTKSKKLIINNSVNQEDFYKFLKTILSAKSEI KEVNDILTEIETGTFLPKQISKSNAEIPYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKIPYYIGPINDN

FIG. 13I

	HKKFFPDRCWVVKKEKSPSGKTTPWNFFDHIDKEKTAEAFITSRTNFCTYLVGESVLPKSSLLYSEYTVLNEINNLQIII
	DGKNICDIKLKQKIYEDLFKYKKITQKQISTFIKHEGICNKTDEVIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNML
	EEIIRWATIYDEGEGKTILKTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLETVTSEMPGFSEPVNIITAMRETQNN
	LMELLSSEFTFTENIKKINSGFEDAEKQFSYDGLVKPLFLSPSVKKMLWQTLKLVKEISHITQAPPKKIFIEMAKGAELEP
	ARTKTRLKILQDLYNNCKNDADAFSSEIKDLSGKIENEDNLRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNYDID
	HIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNNFISLEKLNRLTRATPISDDETAKFIARQL
	VETRQATKVAAKVLEKMFPETKIVYSKAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVYNTKFTNNPWNFI
	KEKRDNPKIADIYNYYKVFDYDVKRNNITAWEKGKTIITVKDMLKRNTPIYTRQAACKKGELFNQTIMKKGLGQHPL
	KKEGPFSNISKYGGYNKVSAAYYTLIEYEEKGNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSL
	LKINGFPCHITGKTNDSFLLRPAVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKTISPYEDLSFRSYIKENLWKKTKNDEIG
	EKEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPNSATIDILVKGKEKFKSLIIENQFEVILEILKLFSATRNVSDLQHIGGS
	KYSGVAKIGNKISSLDNCILIYQSITGIFEKRIDLLKV
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCCGCAAGGTCGAAGCGTC
	CATGAAAAAAGAAATCAAAGACTACTTCCTGGGGCTGGATGTGGGGACTGGGAGCGTGGGGTGGGCTGTGACCG
	ATACTGACTACAAACTGCTGAAGGCTAACCGAAAAGACCTGTGGGGCATGAGATGCTTCGAGACAGCCGAAACT
	GCTGAGGTGCGGAGACTGCACAGGGGAGCCAGGCGCCGAATCGAGCGGAGAAAGAAACGCATTAAGCTGCTGC
	AGGAGCTGTTCTCTCAGGAAATCGCCAAAACCGATGAGGGCTTCTTTCAGAGAATGAAGGAAAGCCCCTTTTACG
	CTGAGGACAAAACAATCCTGCAGGAAAACACTCTGTTCAATGACAAGGATTTTTGCTGATAAGACTTACCACAAA
	GCATATCCTACCATTAATCATCTGATCAAGGCTTGGATTGAGAACAAGGTGAAACCAGACCCCCGACTGCTGTAC
	CTGGCATGTCACAACATCATTAAGAAAAGGGGGACATTTCCTGTTTGAAGGCGACTTCGATTCAGAGAATCAGTTT
	GATACCAGCATCCAGGCACTGTTCGAGTATCTGCGCGAGGACATGGAGGTGGACATCGATGCCGACAGCCAGAA
	GGTCAAAGAGATTCTGAAGGATAGCTCCCTGAAGAACTCTGAAAAACAGAGTCGGCTGAATAAGATCCTGGGGC
	TGAAGCCTTCCGACAAACAGAAGAAAGCCATCACAAACCTGATTTCTGGAAACAAGATCAATTTCGCCGATCTGT
	ACGACAATCCAGATCTGAAGGACGCTGAGAAAAACTCAATCAGCTTCTCCAAGGACGATTTTTGATGCACTGAGT
	GACGATCTGGCCTCAATTCTGGGCGACAGCTTTGAACTGCTGCTGAAGGCCAAAGCTGTCTATAACTGCTCTGTG
	CTGAGTAAGGTCATCGGGGACGAGCAGTACCTGAGCTTCGCCAAGGTGAAAAATCTACGAAAAAGCACAAAACCGA
	TCTGACAAAGCTGAAAAACGTGATCAAGAAACATTTCCCCAAGGACTACAAGAAGGTCTTTGGATACAACAAGA
	ACGAGAAAAACACAACACATTACTCCGGCTATGTGGGAGTCTGTAAGACCAAGAGTAAGAAACTGATCATTAAC
	AACTCAGTCAACCAGGAAGATTTCTACAAGTTTCTGAAAACTATCCTGTCAGCCAAGAGCGAGATCAAGGAAGT
	GAATGACATCCTGACCGAGATTGAAACTGGCACCTTTCTGCCAAAGCAGATCTCTAAAAGTAACGCAGAGATTCC
	CTATCAGCTGAGGAAAATGGAGCTGGAAAAGATCCTGTCCAATGCCGAAAAGCACTTCTCTTTTCTGAAGCAGAA
	AGACGAAAAAGGACTGTCACATAGCGAGAAGATCATTATGCTGCTGACCTTCAAGATCCCTTACTATATTGGCCC
	AATCAACGATAATCACAAGAAATTCTTTCCCGACAGATGCTGGGTGGTCAAGAAAGAGAAAATCCCCCTTCTGGCA
	AGACCACACCATGGAACTTCTTTGATCATATCGACAAGGAAAAAACAGCAGAGGCCTTCATTACTTCTAGGACCA
	ATTTTTGCACTTACCTGGTGGGAGAGCGTCCTGCCTAAGTCTAGTCTGCTGTACTCCGAATATACCGTGCTGAA

FIG. 13J

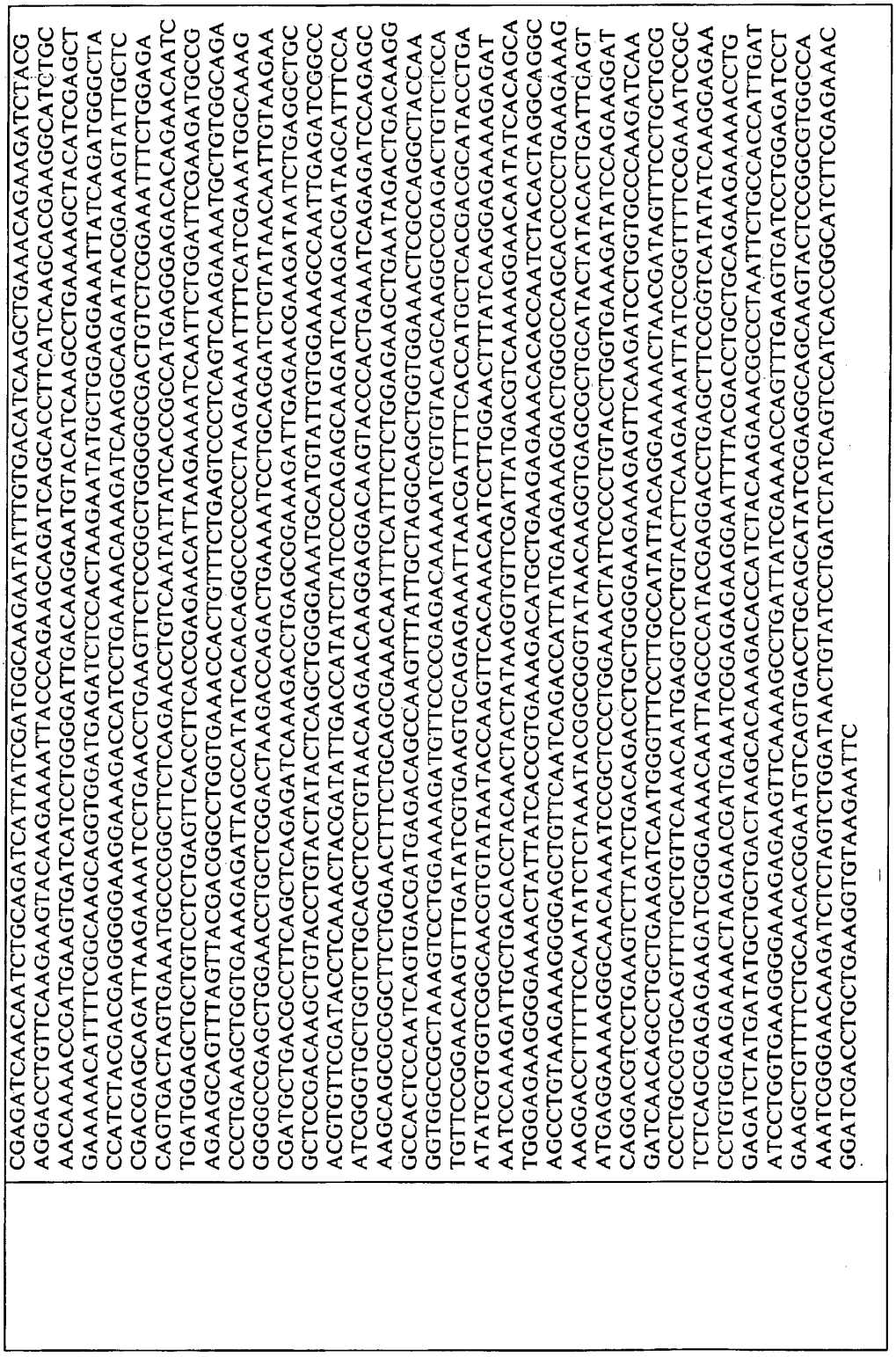

CGAGATCAACAATCTGCAGATCATTATCGATGGCAAGAATATTTGTGACATCAAGCTGAAACAGAAGATCTACG
AGGACCTGTTCAAGAAGTACAAGAAAATTACCCAGAAGCAGATCAGCACCTTCATCAAGGCACGAAGGCATCTGC
AACAAAACCGATGAAGTGATCATCCTGGGGATTGACAAGGAATGTACATCAAGCCTGAAAAGCTACATCGAGCT
GAAAAACATTTTCGGCAAGCAGGTGGATGAGATCTCCACTAAGAATATGCTGGAGGAAATTATCAGATGGGCTA
CCATCTACGACGAGGGGGAAGGAAAGACCATCCTGAAAACAAAGATCAAGGCAGAATACGGAAAGTATTGCTC
CGACGAGCAGATTAAGAAAATCCTGAACCTGAAGTTCTCCGGCTGGGGGCGACTGTCTCGGAAATTTCTGGAGA
CAGTGACTAGTGAAATGCCCCGGCTTCTCAGAACCTGTCAATATTATCACCGCCATGAGGGAGACACAGAACAATC
TGATGGAGCTGCTGTCCTCTGAGTTCACCTTCACCGAGAACATTAAGAAAATCAATTCTGGATTCGAAGATGCCG
AGAAGCAGTTTAGTTACGACGGCCTGGTGAAACCACTGTTTCTGAGTCCCTCAGTCAAGAAAATGCTGTGGCAGA
CCCTGAAGCTGGTGAAAGAGATTAGCCATATCACACAGGCCCCCCCCTAAGAAAATTTTCATCGAAATGGCAAAG
GGGGCCGAGCTGGAACCTGCTCGGACTAAGACCAGACTGAAAATCCTGCAGGATCTGTATAACAATTGTAAGAA
CGATGCTGACGCCTTCAGCTCAGAGATCAAAGACCTGAGCGGAAAGATTGAGAACGAAGATAATCTGAGGCTGC
GCTCCGACAAGCTGTACCTGTACTATACTCAGCTGGGGAAATGCATGTATTGTGGAAAGCCAATTGAGATCGGCC
ACGTGTTCGATACCTCAAACTACGATATTGACCATATCTATCCCCAGAGCAAGATCAAAGACGATAGCATTTCCA
ATCGGGTGCTGGTCTGCAGCTCCTGTAACAAGAACAAGGAGGACAAGTACCCACTGAAATCAGAGATCCAGAGC
AAGCAGCGCGGCTTCTGGAACTTTCTGCAGCGAAACAATTTCATTTCTCTGGAGAAGCTGAATAGACTGACAAGG
GCCACTCCAATCAGTGACGATGAGACAGCCAAGTTTATTGCTAGGCAGCTGGTGGAAACTCGCCAGGCTACCAA
GGTGGCGCTAAAGTCCTGGAAAAGATGTTTCCCCGAGACAAAAATCGTGTACAGCAAGGCCGAGACTGTCTCCA
TGTTCCGGAACAAGTTTGATATCGTGAAGGTGCAGAGAAATTAACGATTTTCACCATGCTCACGACCGCATACCTGA
ATATCGTGGTCGGCAACGTGTATAATACCAAGTTCACAAACAATCCTTGGAACTTTATCAAGGAGAAAAAGAGAT
AATCCAAAGATTGCTGACACCTACTATAAGGTGTTCGATTATGACGTCAAAAGGAACAATATCACAGCA
TGGGAGAAGGGGAAAACTATTATCACCGTGAAAGACATGCTGAAGAAACACACCAATCTACACTAGGCAGGC
AGCCTGTAAGAAAGGGGAGCTGTTCAATCAGACCATTATGAAGAAAGGACTGGGCCAGCACCCCCTGAAGAAAG
AAGGACCTTTTTCCAATATCTCTAAATACGGCGGGTATAACAAGGTGAGCGCTGCATACTATACACTGATTGAGT
ATGAGGAAAAAGGGCAACAAAATCCGCTCCCTGAAACTATTCCCCTGTACCTGGTGAAAGATATCCAGAAGGAT
CAGGACGTCCTGAAGTCTTATCTGACAGACCTGCTGGGGAAGAAAGAGTTCAAGATCCTGGTGCCCAAGATCAA
GATCAACAGCCTGCTGAAGATCAATGGGTTTCCTTGCCATATTACAGGAAAAACTAACGATAGTTTTCCTGCTGCG
CCCTGCCGTGCAGTTTTGCTGTTCAAACAATGAGGTCCTGTACTTCAAGAAAATTATCCGGTTTTTCCGAAATCCGC
TCTCAGCGAGAGAAGATCGGGAAAACAATTAGCCCATACGAGGACCTGAGCTTCCGGTCATATATCAAGGAGAA
CCTGTGGAAGAAAACTAAGAACGATGAAATCGGAGAAGGAATTTTACGACCTGCTGCAGAAGAAAAACCTG
GAGATCTATATGCTGCTGACTAAGCACACAAAGACACCATCTACAAGAAACGCCCTAATTCTGCCACCATTGAT
ATCCTGGTGAAGGGGAAAGAGAAGTTCAAAAGCCTGATTATCGAAAACCAGTTTGAAGTGATCCTGGAGATCCT
GAAGCTGTTTTCTGCAACACGGAATGTCAGTGACCTGCAGCATATCGGAGGCAGCAAGTACTCCGGCGTGGCCA
AAATCGGGAACAAGATCTCTAGTCTGGATAACTGTATCCTGATCTATCAGTCCATCACCGGCATCTTCGAGAAAC
GGATCGACCTGCTGAAGGTGTAAGAATTC

FIG. 13K

Longitud de proteína (sin NLS y marcador HA añadido)	1395
PAM	**NYAAAT**
Longitud del espaciador	30
Especie: **Filifactor_alocis**	
Secuenciación código de barras	AAGACACT
ID	5
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNNGUUUGAGAGUAGGAAACUACACGGUUCAAAUAAAGAAUUUUUCUAAUCGC CCAAUGGGGCCCCAUAUUGAUAUGGAUGAAACUCGCUUAGCCGAGUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MTKEYYLGLDVGTNSVGWAVTDSQYNLCKFKKKDMWGIRLFESANTAKDRRLQRGNRRRLERKKQRIDLLQEIFSP EICKIDPTFFIRLNESRLHLEDKSNDFKYPLFIEKDYSDIEYYKEFPTIFHLRKHLIESEEKQDIRLIYLALAHNIIKTRGHFLI DGDLQSAKQLRPILDTFLLSLQEEQNLSVSLSENQKDEYEEILKNRSIAKSEKVKKLKNLFEISDELEKEEKKAQSAVIE NFCKFIVGNKGDVCKFLRVSKEELEIDSFSFSEGKYEDDIVKNLEEKVPEKVYLFEQMKAMYDWNILVDILETEEYISF AKVKQYEKHKTNLRLLRDIILKYCTKDEYNRMFNDEKEAGSYTAYVGKLKKNNKKYWIEKKRNPEEFYKSLGKLLD KIEPLKEDLEVLTMMIEECKNHTLLPIQKNKDNGVIPHQVHEVELKKILENAKKYYSFLTETDKDGYSVVQKIESIFRFR IPYYVGPLSTRHQEKGSNVWMVRKPGREDRIYPWNMEEIIDFEKSNENFITRMTNKCTYLIGEDVLPKHSLLYSKYMV LNELNNVKVRGKKLPTSLKQKVFEDLFENKSKVTGKNLLEYLQIQDKDIQIDDLSGFDKDFKTSLKSYLDFKKQIFGEE IEKESIQNMIEDIIKWITIYGNDKEMLKRVIRANYSNQLTEEQMKKITGFQYSGWGNFSKMFLKGISGSDVSTGETFDIIT AMWETDNNLMQILSKKFTFMDNVEDFNSGKVGKIDKITYDSTVKEMFLSPENKRAVWQTIQVAEEIKKVMGCEPKKI FIEMARGGEKVKKRTKSRKAQLLELYAACEEDCRELIKEIEDRDERDFNSMKLFLYYTQFGKCMYSGDDIDINELIRGN SKWDRDHIYPQSKIKDDSIDNLVLVNKTYNAKKSNELLSEDIQKKMHSFWLSLLNKKLITKSKYDRLTRKGDFTDEEL SGFIARQLVETRQSTKAIADIFKQIYSSEVVYVKSSLVSDFRKKFPLNYLKSRRVNDYHHAKDAYLNIVVGNVYNKKFTS NPIQWMKKNRDTNYSLNKVFEHDVVINGEVIWEKCTYHEDTNTYDGGTLDRIRKIVERDNILYTEYAYCEKGELFNA TIQNKNGNSTVSLKKGLDVKKYGGYFSANTSYFSLIEFEDKKGDRARHIIGVPIYIANMLEHSPSAFLEYCEQKGYQNV RILVEKIKKNSLLIINGYPLRIRGENEVDTSFKRAIQLKLDQKNYELVRNIEKFLEKYVEKKGNYPIDENRDHITHEKMN QLYEVLLSKMKKFNKKGMADPSDRIEKSKPKFIKLEDLIDKINVINKMLNLLRCDNDTKADLSLIELPKNAGSFVVKK NTIGKSKIILVNQSVTGLYENRREL

FIG. 13L

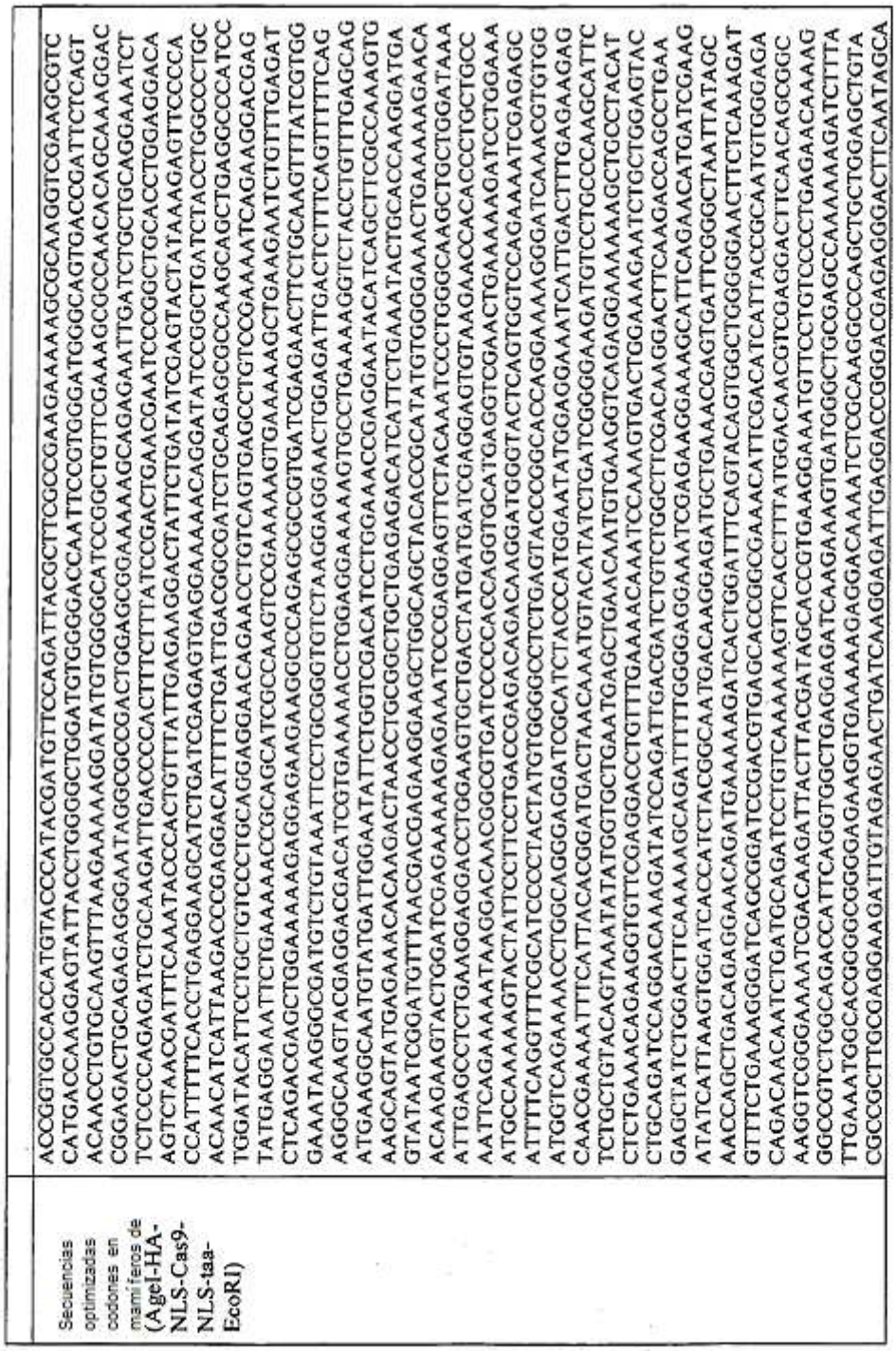

Secuencias optimizadas codones en mamíferos de (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC
CATGACCAAGGAGTATTACCTGGGGCTGGATGTGGGGACCAATTCCGTGGGATGGGCAGTGACCGATTCTCAGT
ACAACCTGTGCAAGTTTAAGAAAAAGGATATGTGGGGCATCCGGCTGTTCGAAAGCGCCAACACAGCAAAGGAC
CGGAGACTGCAGAGAGGGAATAGGCGCCGACTGGAGCGGAAAAAGCAGAGAATTGATCTGCTGCAGGAAATCT
TCTCCCCAGAGATCTGCAAGATTGACCCCACTTTCTTTATCCGACTGAACGAATCCCGGCTGCACCTGGAGGACA
AGTCTAACGATTTCAAATACCCACTGTTTATTGAGAAGGACTATTCTGATATCGAGTACTATAAAGAGTTCCCCA
CCATTTTCACCTGAGGAAGCATCTGATCGAGAGTGAGGAAAAACAGGATATCCGGCTGATCTACCTGGCCCTGC
ACAACATCATTAAGACCCGAGGACATTTTCTGATTGACGGCGATCTGCAGAGGCGCCAAGCAGCTGAGGCCCATCC
TGGATACATTCCTGCTGTCCCTGCAGGAGGAACAGAACCTGTCAGTGAGCCTGTCCGAAAATCAGAAGGACGAG
TATGAGGAAATTCTGAAAAACCGCAGCATCGCCAAGTCCGAAAAAGTGAAAAAGCTGAAGAATCTGTTTGAGAT
CTCAGACGAGCTGGAAAAAGAGGAGAAGGCCCAGAGCGCCGTGATCGAGAACTTCTGCAAGTTTATCGTGG
GAAATAAGGGCGATGTCTGTAAATTCCTGCGGGTGTCTAAGGAGGAACTGGAGATTGACTCTTTCAGTTTTTCAG
AGGGCAAGTACGAGGACGACATCGTGAAAAACCTGGAGGAAAAAGTGCCTGAAATAGGTCTACCTGTTTGAGCAG
ATGAAGGCAATGTATGATTGGAATATTCTGGTCGACATCCTGGAAACCGAGGAATACATCAGCTTCGCCAAAGTG
AAGCAGTATGAGAAACACAAGACTAACCTGCGGCTGCTGAGAGACATCATTCTGAAATACTGCACCAAGGATGA
GTATAATCGGATGTTTAACGACGAGAAGGAAGCTGGCAGCTACACCGCATATGTGGGGAAAGTCTGCTGGAGTAC
ACAAGAAGTACTGGATCGAGAAAAAGAGAAATCCCGAGGAGTTCTACAAATCCCTGGGCAAGCTGCTGGATAAA
ATTGAGCCTCTGAAGGAGGACCTGGAAGTGCTGACTATGATGATCGAGGAGTGTAAGAACCACACCCTGCTGCC
AATTCAGAAAAATAAGGACAACGGCGTGATCCCCCACCAGGTGCATGAGGTCGAACTGAAAAAGATCCCTGAAA
ATGCCAAAAAAGTACTATTCCTTCCTGACCGAGACAGACAAGGATGGGTACTCAGTGGTCCAGAAAATCGAGAGC
ATTTTCAGGTTTCGCATCCCCTACTATGTGGGGCCTCTGAGTACCCGGCACCAGGAAAAGGGATCAAACGTGTGG
ATGGTCAGAAAACCTGGCAGGGAGGATCGCATCTACCCATGGAATATGGAGGAAATCATTGACTTTGAGAAGAG
CAACGAAAATTTCATTACACGGATGACTAACAAATGTACATATCTGATCGGGGAAGATGTCCTGCCCAAGCATTC
TCTGCTGTACAGTAAATATATGGTGCTGAATGAGCTGAACAATGTGAAGGTCAGAGGAAAAAAGCTGCCTACAT
CTCTGAAACAGAAGGTGTTCGAGGACCTGTTTGAAAACACAAATCCAAAGTGACTGGAAAGAATCTGCTGAGTAC
CTGCAGATCCAGGACAAAGATATCCAGATTGACGATCTGTCTGGCTTCGACAAGGACTTCAAGGCCAGCCTGAA
GAGCTATCTGGACTTCAAAAAGCAGATTTTTGGGGAGGAAATCGAGAAGGAAGCATTCAGAACATGATCGAAG
ATATCATTAAGTGGATCACCATCTACGGCAATGACAAGGAGATGCTGAAACGAGTGATTCGGGCTAATTATAGC
AACCAGCTGACACAGAGGAACAGATGAAAAAGATCACTGGATTTCAGTACAGTGGCTGGGGGAACTTCTCAAAGAT
GTTTCTGAAAGGGATCAGCGGATCCGACGTGAGCACCGGCGAAACATTCGACATCATTACCGCAATGTGGGAGA
CAGACAACAATCTGATGCAGATCCTGTCAAAAAAGTTCACCTTTATGGACAACGTCGAGGACTTCAACAGCGGC
AAGGTCGGAAAATCGACAAGATTACTTACGATAGCACCGTGAAGGAAATGTTCCTGTCCCCTGAGAACAAAAG
GGCCGTCTGGCAGACCATTCAGGTGGCTGAGGAGATCAAGAAAGTGATGGGCTGCGAGCCAAAAAAGATCTTTA
TTGAAATGGCACGGGGGCGGGGAGAAGGTGAAAAAGAGGACAAAATCTCGCAAGGCCCAGCTGCTGGAGCTGTA
CGCCGCTTGCGAGGAAGATTGTAGAGAACTGATCAAGGAGATTGAGGACCGGGACGAGAGGGAACTTCAATAGCA

FIG. 13M

	TGAAGCTGTTTCTGTACTATACCCAGTTCGGGAAATGTATGTATTCCGGCGACATCGATATTAACGAGCTGA TTCGCGGCAATTCTAAGTGGGACCGAGATCACATCTACCCCCAGAGCAAAATTAAGGACGATTCCATCGATAACC TGGTGCTGGTCAATAAGACATATAATGCCAAAAAGTCCAATGAGCTGCTGTCTGAGGACATCCAGAAAAAGATG CATTCATTCTGGCTGAGCCTGCTGAACAAAAAGCTGATCACTAAAAGCAAGTACGACCGCCTGACTCGAAAGGG CGACTTTACCGATGAGGAACTGAGTGGGTTCATCGCTAGACAGCTGGTGGAAACAAGGCAGTCAACTAAGGCAA TCGCCGATATCTTCAAGCAGATCTACAGCTCCGAGGTGGTCTATGTGAAGAGCAGCCCTGGTGAGCGACTTCAGGA AAAAGCCACTGAACTACCTGAAGTCTCGGAGAGTCAATGATTACCACCATGCAAAAGACGCCCTATCTGAACATT GTGGTCGGGAACGTGTACAACAAAAAGTTTACCAGTAATCCCATCCAGTGGATGAAAAAGAATCGCGATACAAA CTATAGCCTGAACAAGGTGTTCGAACACGACGTGGTCATTAACGGAGAAGTGATCTGGGAAAAGTGCACATACC ATGAGGACACTAATACCTATGATGGAGGCACTCTGGACCGAATCCGGAAGATTGTGGAACGCGATAACATTCTG TACACCGAGTACGCTTATTGTGAGAAGGGCGAACTGTTTAATGCAACCATCCAGAACACAAAATGGAAACTCCAC AGTCTCTCTGAAAAAGGGCCTGGACGTGAAAAAGTACGGGGGGATACTTCAGCGCCAACACAAGTTACTTCTCACT GATCGAGTTTGAGGACAAGAAGGGGGGATAGAGCAAGGCACATCATTGGAGTGCCTATCTATTGCAAACATGC TGGAGCATTCTCCAAGTGCCTTCCTGGAGTACTGCGAACAGAAGGGGTATCAGAATGTGCGGATTCTGGTCGAGA AAATCAAAAAGAACAGCCTGCTGATCATTAATGGATACCCTCTGCGCATTCGAGGCGAGAACGAAGTGGATACT TCCTTTAAGAGGGCCATCCAGCTGAAGCTGGACCAGAAAAACTATGAGCTGGTCCGCAATATCGAGAAGTTCCTG GAAAAATACGTGGAGAAAAAGGGAAACTATCCAATTGACGAGAATAGAGATCACATCACACATGAAAAGATGA ACCAGCTGTACGAGGTGCTGCTGTCCAAAATGAAAAAAGTTCAACAAGAAGGGCATGGCCGACCCCTCTGATAGG ATCGAAAAGAGTAAGCCTAAATTCATCAAGCTGGAGGACCTGATCGATAAGATTAATGTGATCAACAAAATGCT GAACCTGCTGCGCTGTGACAATGATACTAAGGCCGACCTGTCTCTGATTGAGCTGCCCAAAAACGCTGGGAGTTT CGTGGTCAAAAAGAATACCATCGGAAAGTCAAAAAATCATCCTGGTGAATCAGAGCGTGACTGGACTGTACGAGA ATAGACGGGAACTGTAAGAATTC
Longitud de proteína (de tipo salvaje sin NLS añadido)	1365
PAM	**NNNAAGC**
Longitud del espaciador	30-31
Especie:	**Staphylococcus pseudintermedius**
Secuenciación	TCTCGGTC

FIG. 13N

código de barras	
ID	6
ARN guía químérico	NNNNNNNNNNNNNNNNNNGUUUUAGCACUAGAAAUAGUUAAGUUAAAACAAGCUUAAAGCGUCAAUGU AAUAUUUUAUUAACACCCUACUGUGUCAGUGGGGUUUUUUU
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MGRKPYILSLDIGTGSVGYACMDKGFNVLKYHDKDALGVYLFDGALTAQERRQFRTSRRRKNRRIKRLGLLQELLAP LVQNPNFYQFQRQFAWKNDNMDFKNKSLSEVLSFLGYESKKYPTIYHLQEALLLKDEKFDPELIYMALYHLVKYRGH FLFDHLKIENLTNNDNMHDFVELIETYENLNNIKLNLDYEKTKVIYEILKDNEMTKNDRAKRVKNMEKKLEQFSIMLL GLKFNEGKLFNHADNAEELKGANQSHTFADNYEENLTPFLTVEQSEFIERANKIYLSLTLQDILKGKKSMAMSKVAAY DKFRNELKQVKDIVYKADSTRTQFKKIFVSSKKSLKQYDATPNDQTFSSLCLFDQYLIRPKKQYSLLIKELKLKKIIPQDSE LYFEAENDTLLKVLNTTDNASIPMQINLYEAETILRNQQKYHAEITDEMIEKVLSLIQFRIPYYVGPLVNDHTASKFGW MERKSNESIKPWNFDEVVDRSKSATQFIRRMTNKCSYLINEDVLPKNSLLYQEMEVLNELNATQIRLQTDPKNRKYRM MPQIKLFAVEHIFKKYKTVSHSKFLEIMLNSNHRENFMNHGEKLSIFGTQDDKKFASKLSSYQDMTKIFGDIEGKRAQI EEIIQWITIFEDKKILVQKLKECYPELTSKQINQLKKLNYSGWGRLSEKLLTHAYQGHSIIELLRHSDENFMEILTNDVY GFQNFIKEENQVQSNKIQHQDIANLTTSPALKKGIWSTIKLVRELTSIFGEPEKIIMEFATEDQQKGKKQKSRKQLWDD NIKKNKLKSVDEYKYIIDVANKLNNEQLQQEKLWLYLSQNGKCMYSGQSIDLDALLSPNATKHYEVDHIFPRSFIKDD SIDNKVLVIKKMNQTKGDQVPLQFIQQPYERIAYWKSLNKAGLISDSKLHKLMKPEFTAMDKEGFIQRQLVETRQISV HVRDFLKEEYPNTKVIPMKAKMVSEFRKKFDIPKIRQMNDAHHAIDAYLNGVVYHGAQLAYPNVDLFDFNFKWEKV REKWKALGEFNTKQKSRELFFFKKLEKMEVSQGERLISKIKLDMNHFKINYSRKLANIPQQFYNQTAVSPKTAELKYE SNKSNEVVYKGLTPYQTYVVAIKSVNKKGKEKMEYQMIDHYVFDFYKFQNGNEKELALYLAQRENKDEVLDAQIVY SLNKGDLLYINNHPCYFVSRKEVINAKQFELTVEQQLSLYNVMNNNKETNVEKLLIEYDFIAEKVINEYHHYLNSKLKE KRVRTFFSESNQTHEDFIKALDELFKVVTASATRSDKIGSRKNSMTHRAFLGKGKDVKIAYTSISGLKTTKPKSLFKLA ESRNEL
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGGTC CATGGGGAGGAAACCTTACATCTGTCTCTGGATATTGGAACTGGGTCCGTCGGCTACGCTTGCATGGATAAAGG ATTCAACGTGCTGAAGTACCACGACAAAGATGCCCTGGGAGTGTATCTGTTCGACGGCGCTCTGACTGCACAGGA GCGGAGACAGTTTAGGACCTCCAGGCGCCGAAAGAACCGGAGAATCAAACGCCTGGGCCTGCTGCAGGAACTGC TGGCACCCCTGGTGCAGAACCCTAATTTCTACCAGTTTCAGCGGCAGTTCGCCTGGAAGAACGACAATATGGATT TTAAGAACAAGAGCCTGTCTGAGGTGCTGAGCTTCCTGGGATATGAATCCAAGAAATACCCTACCATCTACCACC TGCAGGAGGCTCTGCTGCTGAAAGACGAGAAGTTTTGATCCAGAACTGATCTACATGGCACTGTATCATCTGGTGA AATACAGAGGCCACTTTCTGTTCGAAGATCAGGAACCTGACTAACAATGACAATATGCACGATTTCG TGGAGCTGATTGAAACCTATGAGAACCTGAACAATATCAAGCTGGACTACGAGAAAAACCAAAGTGATC TATGAGATTCTGAAAGACAACGAAATGACTAAGAATGATAGAGCCCAAAAAGGGTCAAGAACATGGAGAAGAAAC TGGAACAGTTCTCTATCATGCTGCTGGGGCTGAAGTTCAATGAGGGAAAAACTGTTTAACCACCGCCGATAATGCTG AGGAACTGAAGGGGGCTAACCAGAGCCATACATTTGCAGACAACTACGAGGAAAATCTGACTCCCTTCCTGACC

FIG. 13O

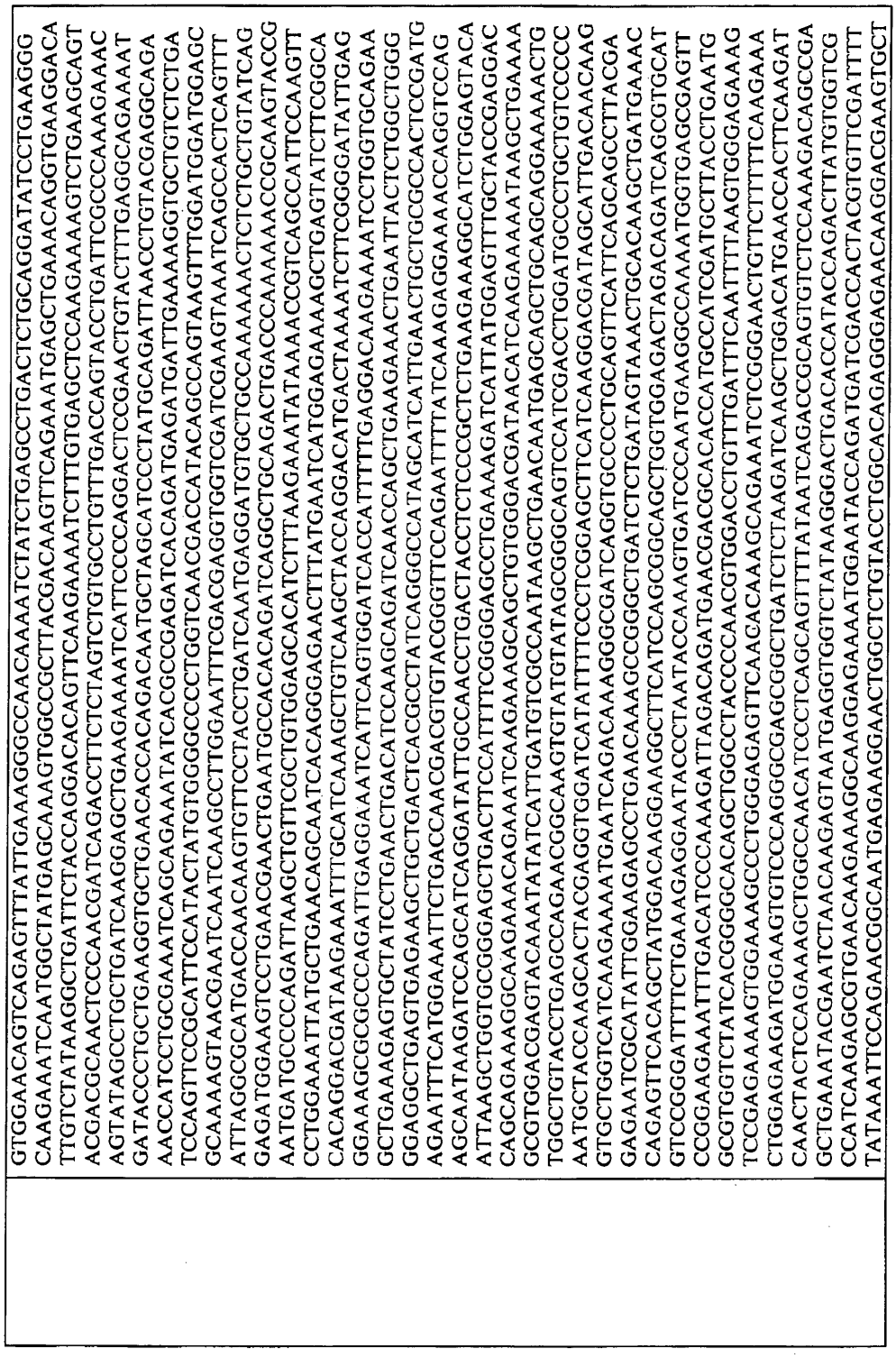

GTGGAACAGTCAGAGTTTATTGAAAGGGCCAACAAAATCTATCTGAGCCTGACTCTGCAGGATATCCTGAAGGG
CAAGAAATCAATGGCTATGAGCAAAGTGGCCGCTTACGACAAGTTCAGAAATGAGCTGAAACAGGTGAAGGACA
TTGTCTATAAGGCTGATTCTACCAGGACACAGTTCAAGAAAATCTTTGTGAGCTCCAAGAAAAGTCTGAAGCAGT
ACGACGCAACTCCCAACGATCAGACCTTCTCTAGTCTGTGCCTGTTTGACCAGTACCTGATTCGCCCAAAGAAAC
AGTATAGCCTGCTGATCAAGGAGCTGAAGAAATCATTCCCCAGGACTCCGAACTGTACTTTGAGGCAGAAAAT
GATACCCTGCTGAAGGTGCTGAACACCACAGACAATGCTAGCATCCCTATGCAGATTAACCTGTACGAGGCAGA
AACCATCCTGCGAAATCAGCAGAAATATCACGCCGAGATCACAGATGAGATTGATTGAAAAGGTGCTGTCTCTGA
TCCAGTTCCGCATTCCATACTATGTGGGGCCCCTGGTCAACGACCATACAGCCAGTAAGTTTGGATGGATGGAGC
GCAAAAGTAACGAATCAATCAAGCCTTGGAATTTCGACGAGGTGGTCGATCGAAGTAAATCAGCCACTCAGTTT
ATTAGGCGCATGACCAACAAGTGTTCCTACCTGATCAATGAGGATGTGCTGCCAAAAAACTCTCTGCTGTATCAG
GAGATGGAAGTCCTGAACGAACTGAATGCCACACAGATCAGGCTGCAGACTGACCCAAAAAACCGCAAGTACCG
AATGATGCCCCAGATTAAGCTGTTCGCTGTGGAGCACATCTTTAAGAAATATAAAACCGTCAGCCATTCCAAGTT
CCTGGAAATTATGCTGAACAGCAATCACAGGGAGAACTTTATGAATCATGGAGAAAAGCTGAGTATCTTCGGCA
CACAGGACGATAAGAAATTTGCATCAAAGCTGTCAAGCTACCAGGACATGACTAAAATCTTCGGGGATATTGAG
GGAAAGCGCGCCCAGATTGAGGAAATCATTCAGTGGATCACACCATTTTTGAGGACAAGAAAATCCTGGTGCAGAA
GCTGAAAGAGTGCTATCCTGAACTGACATCCAAGCAGATCAACCAGCTGAAGAAACTGAATTACTCTGGCTGGG
GGAGGCTGAGTGAGAAGCTGCTGACTCACGCCTATCAGGGCCATAGCATCATTGAACTGCTGCGCCACTCCGATG
AGAATTTCATGGAAATTCTGACCAACGACGTGTACGGGTTCCAGAATTTTTATCAAAGAGGAAAACCAGGTCCAG
AGCAATAAGATCCAGCATCAGGATATTGCCAACCTGACTACCTCTCCCGCTCTGAAGAAAGGCATCTGGAGTACA
ATTAAGCTGTGGCGGGAGCTGACTTCCATTTTCGGGGAGCCTGAAAAGATCATTATGGAGTTTGCTACCGAGGAC
CAGCAGAAAGGCAAGAAACAGAAAATCAAGAAAGCAGCTGTGGGACGATAACATCAAGAAAAATAAGCTGAAAA
GCGTGGACGAGTACAAATATATCATTGATGTCGCCAATAAGCTGAACAATGAGCAGCTGCAGCAGGAAAAACTG
TGGCTGTACCTGAGCCAGAACGGCAAGTGTATGTATAGCGGGCAGTCCATCGACCTGGATGCCCTGCTGTCCCCC
AATGCTACCAAGCACTACGAGGTGGATCATATTTTCCCTCGGAGCTTCATCAAGGACGATAGCATTGACAACAAG
GTGCTGGTCATCAGAAAATGAATCAGACAAAGGGCGATCAGGTGCCCCTGCAGTTCATTCAGCAGCCTTACGA
GAGAATCGCATATTGGAAGAGCCTGAACAAAGCCGGGCTGATCTCTGATAGTAAACTGCACAAGCTGATGAAAC
CAGAGTTCACAGCTATGGACAAGGAAGGCTTCATCCAGCGGCAGCTGGTGGAGACTAGACAGATCAGCGTGCAT
GTCCGGGATTTTCTGAAAGAGGAATACCCTAATACCAAAGTGATCCCAATGAAGGCCAAAATGGTGAGCGAGTT
CCGGAAGAAATTTGACATCCCAAAGATTAGACAGATGAACGACGCACACCATGCCATCGATGCTTACCTGAATG
GCGTGGTCTATCACGGGGCACAGCTGGCCTACCCCAACGTGGACCTGTTTGATTTCAATTTTAAGTGGGAGAAAG
TCCGAGAAAAGTGGAAAGCCCTGGGAGAGTTCAAGACAAAGCAGAAAATCTCGGGAACTGTTCTTTTTCAAGAAA
CTGGAGAAGATGGAAGTGTCCCAGGGCGAGCGGCTGATCTCTAAGATCAAGCTGGACATGAACCACTTCAAGAT
CAACTACTCCAGAAAGCTGGCCAACATCCCTCAGCAGTTTTATAATCAGACCGCAGTGTCTCCAAAGACAGCCGA
GCTGAAATACGAATCTAACAAGAGTAATGAGGTGGTCTATAAGGGACTGACACCATACCAGACTTATGTGGTCG
CCATCAAGAGCGTGAACAAGAAGGCAAGGAGAAAATGGAATACCAGATGATCGACCACTACGTGTTCGATTTT
TATAAATTCCAGAACGGCAATGAGAAGGAACTGGCTCTGTACCTGGCACAGAGGGAGAACAAGGACGAAGTGCT

FIG. 13P

	GGATGCTCAGATTGTCTATAGTCTGAATAAGGGGGATCTGCTGTACATCAACAATCATCCCTGCTATTTCGTGTCA CGCAAAGAGGTCATCAACGCAAAGCAGTTTGAGCTGACCGTGGAACAGCAGCTGTCTCTGTACAACGTGATGAA CAACAAGGAGACAAATGTCGAAAAAGCTGCTGATCGAGTATGACTTCATTGCCGAGAAAGTGATCAACGAATACC ACCATTATCTGAATAGCAAGCTGAAAGAAAAGCGAGTCCGGACCTTTTTCTCAGAGAGCAACCAGACACACGAG GACTTCATCAAGGCCCTGGACGAGCTGTTTAAGGTGGTCACCGCCATCCAAGGTCTGATAAAATCGGGAGT CGCAAGAACAGCATGACTCATCGAGCCTTCCTGGGAAAAGGCAAGGACGTGAAGATTGCTTACACCTCCATCTCT GGACTGAAAACACTAAACCTAAGAGTCTGTTTAAGCTGGCCGAGTCAAGAAACGAACTGTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1334
PAM	**NGG**
Longitud del espaciador	29
Especie: **Lactobacillus johnsonii**	
Secuenciación código de barras	AATGTTCT
ID	7
ARN guía químérico	**NNNNNNNNNNNNNNNNNNNNGUUUUAGAUGGUGAAAACCAGAUUUAAAAUCAAGCAAUGCAUCUUUUGAU GCAAAGUUUCAAUAUUUGUCCCACGUUAUCCGAGGGACUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MTKIKDDYIVGLDIGTDSCGWVAMNSNNDILKLQGKTAIGSRLFEGGKSAAERRLFRTTHRRIKRRRWRLKLLEEFFD PYMAEVDPYFFARLKESGLSPLDKRKTVSSIVFPTSAEDKKFYDDYPTIYHLRYKLMTEDEKFDLREVYLAIHHIIKYR GNFLYNTSVKDFKASKIDVKSSIEKLNELYENLGLDLNVEFNISNTAEIEKVLKDKQIFKRDKVKKIAELFAIKTDNKEQ SKRIKDISKQVANAVLGYKTRFDTIALKEISKDELSDWNFKLSDIDADSKFEALMGNLDENEQAILLTIKELFNEVTLNG IVEDGNTLSESMINKYNDHRDDLKLLKEVIENHIIDRKKAKELALAYDLYVNNRHGQLLQAKKKLGKIKPRSKEDFYK VVNKNLDDSRASKEIKKKIELDSFMPKQRTNANGVIPYQLQQLELDKIIENQSKYYPFLKEINPVSSHLKEAPYKLDELI RFRVPYYVGPLISPNESTKDIQTKKNQNFAWMIRKEEGRITPWNFDQKVDRIESANKFIKRMTTKDTYLFGEDVLPANS LLYQKFTVLNELNNIRINGKRISVDLKQEIYENLFKKHTTVTVKKLENYLKENHNLVKVEIKGLADEKKFNSGLTTYN RFKNLNIFDNQIDDLKYRNDFEKIIEWSTIFEDKSIYKEKLRSIDWLNEKQINALSNIRLQGWGRLSKKLLAQLHDHNGQ TIIEQLWDSQNNFMQIVTQADFKDAIAKANQNLLVATSVEDILNNAYTSPANKKAIRQVIKVVDDIVKAASGKVPKQI

FIG. 13Q

	AIEFTRDADENPKRSQTRGSKLQKVYKDLSTELASKTIAEELNEAIKDKKLVQDKYYLYFMQLGRDAYTGEPINIDEIQ KYDIDHILPQSFIKDDALDNRVLVSRAVNNGKSDNVPVKLEGNEMAANLGMTIRKMWEEWKNIGLISKTKYNNLLTD PDHINKYKSAGFIRRQLVETSQIIKLVSTILQSRYPNTEIITVKAKYNHYLREKFDLYKSREVNDYHHAIDAYLSAICGNL LYQNYPNLRPFFVYGQYKKFSSDPDKEKAIFNKTRKFSFISQLLKNKSENSKEIAKKLKRAYQFKYMLVSRETETRDQE MFKMTVYPRFSHDTVKAPRNLIPKKMGMSPDIYGGYTNNSDAYMVIVRIDKKKGTEYKILGIPTRELVNLKKAEKED HYKSYLKEILTPRILYNKNGKRDKKITSFEIVKSKIPYKQVIQDGDKKFMLGSSTYVYNAKQLTLSTESMKAITNNFDK DSDENDALIKAYDEILDKVDKYLPLFDINKFREKLHSGREKFIKLSLEDKKDTILKVLEGLHDNAVMTKIPTIGLSTPLG FMQFPNGVILSENAKLIYQSPTGLFKKSVKISDL
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGACAAAAATCAAAGACGACTACATCGTGGGACTGGACATCGGCACAGACTCCTGCGGGTGGGTGGCTATGA ACAGCAATAATGACATTCTGAAACTGCAGGGCAAGACCGCGACAATCGGGTCACCGCCTGTTCGAGGGAGGGAAGAGC GCAGCTGAACGGAGACTGTTTCGCACCACACACAGGCGCATCAAACGACGGAGATGGCGACTGAAGCTGCTGGA GGAGTTCTTCGACCCCTACATGGCAGAGGTGGATCCTTATTTCTTTGCCCGGCTGAAGGAATCTGGCCTGAGTCC ACTGGACAAAAGAAAGACCGTGAGCTCCATTGTGTTCCCCACATCCGCCGAGGATAAGAAGTTCTACGACGATT ACCCTACAATCTACCATCTGAGGTATAAACTGATGACTGAGGACGAAAAGTTCGATCTGCGCGAAGTGTACCTGG CTATCCACCATATCATTAAGTACCGAGGAAAACTTCCTGTGTATAATACCAGTGTGAAAGACTTCAAGGCATCAAAGA TCGATGTCAAATCTAGTATCGAGAAGCTGAACGAGGCTGTATGAAAATCTGGGCCTGGACCTGAACGTGGAGTTCA ACATTAGCAAATACTGCCGAGATCGAAAAGGTGCTGAAAGACAAGCAGATCTTCAAGCGGGATAAAGTCAAGAAA ATTGCCGAGCTGTTTGCTATCAAAACCGACAACAAGGAACAGAGCAAGAATCAAAGATATTTCCAAACAGGT GGCCAATGCTGTCCTGGGGTACAAGACCAGGTTCGACACAATCGCTCTGAAAGAGATTTCCAAGGACGAACTGT CTGATTGGAACTTCAAACTGTCAGACATCGATGCAGACAGCAAGTTTGAGGCCCTGATGGGAAACCTGGATGAG AATGAACAGGCCATCCTGCTGACTATTAAGGAGCTGTTTAACGAAGTGACCCTGAATGGAATTGTCGAGGACGG CAACACCCTGAGCGAATCCATGATCAACAAGTACAATGATCACCGGGACGATCTGAAGCTGCTGAAAGAAGTGA TCGAAAATCATATTGACAGAAAGAAAGCCAAGGAGCTGGCACTGGCCTACGATCTGTATGTCAACAATAGGCAC GGACAGCTGCTGCAGGCTAAGAAAAAAGCTGGGCAAAATCAAGCCCCGCTCTAAGGAGGACTTCTACAAAGTGGT CAACAAGAATCTGGACGATTCACGGGCAAGCAAGGAGATCAAAAAGAAAATTGAACTGGACAGCTTTATGCCTA AGCAGAAGCCAACGCCAATGGCGTGATCCCATACCAGCTGCAGCAGCTGGAGCTGGATAAGATCATCGAAAAC CAGTCTAAGTACTATCTCCATTCCTGAAGGAGATTAATCCCGTGTCAAGCCACCTGAAAGAGGCCCCCTATAAGCTG GACGAACTGATCCGATTTCGGGTGCCTTACTATGTCGGCCCCCTGATTTCTCCTAACGAGAGTACCAAGGATATC CAGACAAAGAAAAACCAGAATTTCGCCTGGATGATTCGCAAAGAGGAAGGGCGAATCACACCTTGGAACTTTGA CCAGAAGGTGGATCGAATTGAGAGCGCCAATAAGTTCATCAACGGATGACTACCAAGGACACTTACCTGTTTG GGGAGGATGTGCTGCCAGCTAACAGCCTGCTGTATCAGAAGTTCACCGTCCTGAACGAACTGAACATCAACCGG ATTAATGGAAAAAGAATCTCCGTGGACCTGAAGCAGGAGATCTACGAAAACCTGTTTAAGAAACACACAACTGT GACCGTCAAGAAACTGGAGAATTATCTGAAGGAAAACCATAATCTGGTGAAAGTCGAGATCAAGGGGCTGGCCG ATGAAAAGAAATTCAACAGCGGACTGACCACATACAATAGATTCAAGAACCTGAACATCTTTGACAACCAGATT

FIG. 13R

	GACGATCTGAAGTACAGGAACGATTTCGAGAAGATCATCGAATGGTCTACAATTTTTGAGGACAAGAGTATCTAC AAAGAAAAGCTGAGGAGCATCGATTGGCTGAACGAGAAGCAGATTAACGCTCTGTCTAATATCAGACTGCAGGG GTGGGGAAGGCTGAGTAAGAAACTGCTGGCACACGCTGCACGACCATAATGGCCAGACCATCATTGAGCAGCTGT GGGATTCCCCAGAACAATTTCATGCAGATTGTGACACAGGCCGACTTTAAAGATGCTATCGCAAAGGCCAACCAG AATCTGCTGGTGGCTACCTCAGTCGAGGACATTCTGAACAATGCATACACAAGCCCCGCAAACACAAGAAAGCCAT CAGACAGGTCATCAAGGTGGTCGACGATATCGTGAAGGCAGCCTCCGGAAAGGTCCCCAAAACAGATCGCCATTG AGTTCACTAGGGATGCTGACGAAAAATCCCAAGAGAAGTCAGACCAGGGGCTCAAAGCTGCAGAAAGTGTACAAG GACCTGAGCACTGAGCTGGCCTCCAAGACCATTGCTGAGGAACTGAACGAAGCAATCAAAGACAAGAAACTGGT GCAGGATAAGTACTATCTGTACTTTATGCAGCTGGGGCGGGACGCCTATACAGGAGAGCCTATCAATATCGATGA AATCCAGAAGTACGATATCGACCACATTCTGCCACAGTCTTTCATCAAGGACGATGCCCCTGGACAACAGGGTGCT GGTGAGCCGGCTGTGAACAATCTGATAATGTGCCTGTCAAGCTGTTTGGCAACGAGATGGCTGCAA ATCTGGGGATGACTATCAGGAAAATGTGGGAGGAATGGAAGAACATCGGCCTGATTAGCAAAACAAAGTACAAC AATCTGCTGACTGATCCCGACCACATTAACAAGTATAAGAGTGCCGGTTCATCAGGCGCCAGCTGGTGGAGAC ATCACAGATCATCAAGCTGGTGAGCACTATCCTGCAGAGTCGCTACCCTAACACTGAAATCATTACCGTGAAGGC TAAGTACAATCATTATCTGCGGGAGAAATTTGACCTGTATAAGAGCAGAGAAGTCAACGACTACCACCATGCTAT TGATGCATATCTGTGCCGCCATCTGCGGAAATCTGCTGTACCAGAACTATCCAAATCTGCGGCCCTTCTTTGTGTAC GGCCAGTATAAGAAATTCTCCTCTGATCCTGACACAAAGAGAAGGCCATTTTTAACAAAACCCGCAAGTTCTCCTTT ATCTCTCAGCTGCTGAAAAACAAGAGTGAGAACAGCAAGGAAATCGCTAAGAAACTGAAACGGGCATACCAGTT CAAGTATATGCTGGTGTCTCGAGAGACTGAAACCCGGGACCAGGAGATGTTCAAAATGACCGTGTACCCCCGGTT CAGCCACGATACAGTCAAGGCTCCTAGGAACCTGATTCCAAAGAAAATGGGCATGTCCCCTGACATCTACGGAG GCTATACAAACAATTCTGACGCATACATGGTCATCGTCCGCATTGATAAGAAAAAGGGAACTGAGTATAAGATC CTGGGCATTCCAACCCGGGAACTGGTGAAATCTGAAAAAGGCCGAGAAGGAGGACCATTACAAAAGCTATCTGAA GGAGATCCTGACACCAAGGATTCTGTACAACAAAAATGGGAAGCGCGATAAAAAGATCACTTCCTTCGAAATTG TGAAATCTAAGATCCCCTATAAGCAGGTCATCCAGGATGGGGACAAAAAGTTTATGCTGGGAAGTTCAACATAC GTGTATAACGCAAAGCAGCTGACACTGAGCACTGAGTCCATGAAAGCCATCACTAACAATTTCGATAAGGACAG CGATGAGAACGACGCTCTGATTAAGGCATACGATGAAATCCTGGACAAAAGTGGATAAGTATCTGCCACTGTTCG ACATCAACAAGTTCCGGGAGAAGCTGCACAGTGGGCGAGAAAAGTTCATCAAGCCTGAGCCTGGAGGACAAAAA GGATACCATCCTGAAAGTGCTGGAAGGACTGCATGATAACGCTGTCATGACAAAGATCCCTACTATTGGCCTGTC CACACCACTGGGGTTCATGCAGTTTCCCAACGGCGTGATTCTGAGCGAGAATGCCCAAACTGATCTACCAGTCCCC CACCGGGCTGTTCAAAAAGTCAGTGAAGATCAGCGACCTGTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1375
PAM	**NTAAA**

FIG. 13S

Longitud del espaciador	30
Especie:	*Neisseria cinerea*
Secuenciación código de barras	ATTGTCTG
ID	18
ARN guía químérico	NNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCCAUUCUCGAAAGAGAACCGUUGCUACAAUAAGGCCGUCU GAAAAGAUGUGCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGGGCAUCGUUUAUUUCGGUUAA AAUGCCGUCUCUGAAAACCGGUUUUUAGGUUUCAGACGGCA
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MAAFKPNPMNYILGLDIGIASVGWAIVEIDEENPIRLIDLGVRVFERAEVPKTGDSLAAARRLARSVRRLTRRRAHRL LRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELG ALLKGVADNTHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFNRKDLQAELNLLFEKQKEFGNPHVSDGLKE GIETLLMTQRPALSGDAVQKMLGHCTFEPTEPKAAKNTYTAERFVWLTKLNNLRILEQGSERPLTDTERATLMDEPYR KSKLTYAQARKLLDLDDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLF KTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGNRYDEACTEIYGDHYGKKNTEEKIYLPPIPA DEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKSAAKFREYFPNFVGEP KSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNG KDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYINRFLCQFVADHMLLTGKGKRRVFASNGQIT NLLRGFWGLRKVRAENDRHHALDAVVVACSTIAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKAHFPQPWEFF AQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHKYVTPLFISRAPNRKMSGQGHMETVKSAKRLDEGI SVLRVPLTQLKLKDLEKMVNRERPKLYEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTG VWVHNHNGIADNATIVRVDVFEKGGKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWTVMDDSFEFKFVLYANDLIK LTAKKNEFLGYFVSLNRATGAIDIRTHDTDSTKGKNGIFQSVGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGGCTGCCTTCAAACCTAATCCTATGAACTACATCCTGGGCCTGGACATTGGAATCGCTTCTGTCGGGTGGGCT ATCGTGGAAATCGAGGAAGAGAACCCTATCAGACTGATTGATCTGGGAGTCAGAGTGTTTGAAAGGGCAGA GGTGCCAAAGACCGGCGCACTCCCTGGCCGCTGCACGGAGACTGGCTCTGTCAGGCCGCCTGACACGACGGA GAGCACACAGGCTGCTGCGAGCTAGGCGCCTGCTGAAGAGAGGGCGTGCTGCAGGCCGCTGACTTCGATGAA AACGGCCTGATCAAGAGCCTGCCCAATACTCCTTGGCAGCTGAGAGCAGCCGCTCTGGACAGGAAGCTGACCCC ACTGGAGTGGTCTGCCGTGCTGCTGCACCTGATCAAGCATCGCGGCTACCTGAGTCAGCGAAAAAATGAAGGGG

FIG. 13T

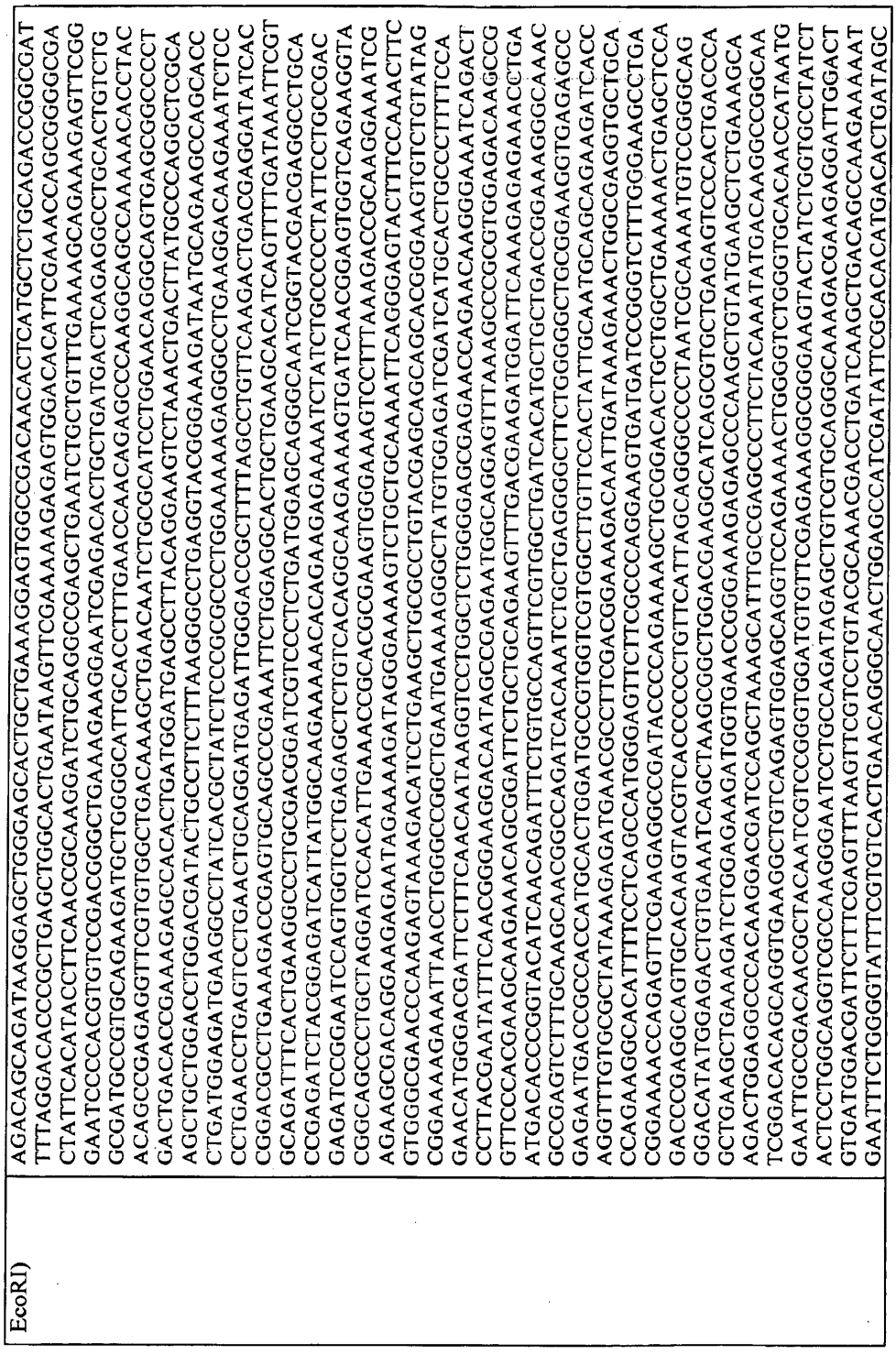

EcoRI)
AGACAGCAGATAAGGAGCTGGGAGCACTGCTGAAAGGAGTGGCCGACAACACTCATGCTCTGCAGACCGGCGAT
TTTAGGACACCCGCTGAGCTGGCACTGAATAAGTTCGAAAAAAGAGAGTGGACACATTCGAAACCAGCGGGGCGA
CTATTCACATACCTTCAACCGCAAGGATCTGCAGGCCGAGCTGAATCTGCTGTTTGAAAAGCAGAAAGAGTTCGG
GAATCCCCACGTGTCCGACGGGCTGAAAGGAATCGAGACACTGCTGATGACTCAGAGGCCTGCACTGTCTG
GCGATGCCGTGCAGAAGATGCTGGGGCATTGCACCTTTGAACCCAACAGAGCCCAAGGCAGCCAAAAACACCTAC
ACAGCCGAGAGGTTCGTGTGGCTGACAAAGCTGAACAATCTGCGCATCCTGGAACAGGGCAGTGAGCGGCCCCT
GACTGACACCGAAAGAGCCACACTGATGGATGAGCCTTACAGGAAGTCTAAACTGACTTATGCCCAGGCTCGCA
AGCTGCTGGACCTGGACGATACTGCCTTCTTTAAGGGCCTGAGGTACGGGAAAGATAATGCAGAAGCCAGCACC
CTGATGGAGATGAAGGCCTATCACGCTATCTCCCGCGCCCTGGAAAAAGAGGGCCTGAAGGACAAGAAATCTCC
CCTGAACCTGAGTCCTGAACTGCAGGATGAGATTGGGACCGCTTTTAGCCTGTTCAAGACTGACGAGGATATCAC
CGGACGCCTGAAAGACCGAGTGCAGCCCGAAATTCTGGAGGCACTGCTGAAGCACATCAGTTTTGATAAATTCGT
GCAGATTTCACTGAAGGCCCTGCGACGGATCGTCCCTCTGATGGAGCAGGGCAATCGGTACGACGAGGCCTGCA
CCGAGATCTACGGAGATCATTATGGCAAGAAAAACACAGAAGAGAAAATCTATCTGCCCCCTATTCCTGCCGAC
GAGATCCGGAATCCAGTGGTCCTGAGAGCTCTGTCACAGGCAAGAAAAAGTGATCAACGGAGTGGTCAGAAGGTA
CGGCAGCCCTGCTAGGATCCACATTGAAACCGCACGCGAAGTGGGAAAGTCCTTTAAAGACCGCAAGGAAATCG
AGAAGCGACAGGAAGAATAGAAAAAGATCTGCTGCAAAATTCAGGGAGTACTTTCCAAACTTC
GTGGGCGAACCCAAGAGTAAAGACATCCTGAAGCTGCGCCTGTACGAGCAGCACGGGAAGTGTCTGTATAG
CGGAAAAGAAATTAACCTGGGCCGGCTGAATGAAAAGGGCTATGTGGAGATCGATCATGCACTGCCCTTTTCCA
GAACATGGGACGATTCTTTCAACAATAAGGTCCTCTGGCTCTGGGGAGCGAGAACCAGAACAAGGGAAATCAGACT
CCTTACGAATATTTCAACGGGAAGGACAATAGCCGAGAATGGCAGGAGTTTAAAGCCCCGCGTGGAGACAAGCCG
GTTCCCACGAAGCAAGAAACAGCGGATTCTGCTGCAGAAGTTTGACGAAGATGGATTCAAAGAGAAACCTGA
ATGACACCCGGTACATCAACAGATTTCTGTGCCAGTTCGTGGCTGATCACATGCTGCTGACCGGAAAGGGCAAAC
GCCGAGTCTTTGCAAGCAACGGCCAGATCACAAATCTGCTGAGGGGCTTCTGGGGGCTGCGGAAGGTGAGAGCC
GAGAATGACCGCCACCATGCACTGGATGCCGTGGTCGTGGCTTGTTCCACTATTGCAATGCAGCAGAAGATCACC
AGGTTTGTGCGCTATAAAGAGATGAACGCCTTCGACGGAAAGACAATTGATAAAGAAACTGGCGAGGTGCTGCA
CCAGAAGGCACATTTTCCTCAGCCATGGGAGTTCTTCGCCCAGGAAGTGATGATCCGGGTCTTTGGGAAGCCTGA
CGGAAAACCAGAGTTCGAAGAGGCCGATACCCCAGAAAAGCTGCCGGACACTGCTGGCTGAAAAACTGAGCTCCA
GACCCGAGGCAGTGCACAAGTACGTCACCCCCCTGTTCATTAGCAGGGCCCCCTAATCGCAAAATGTCCGGGCAG
GGACATATGGAGACTGTGAAATCAGCTAAGCGGCTGGACGAAGGCATCAGCGTGCTGAGAGTCCCACTGACCCA
GCTGAAGCTGAAAGATCTGGAGAAGATGGTGAACCGGGAAAGAGAGCCCAAGCTGTATGAAGCTCTGAAAGCA
AGACTGGAGGCCCACAAGGACGATCCAGCTAAAGCATTTGCCGAGCCCTTCTACAAATATGACAAGGCCGGCAA
TCGGACACAGCAGGTGAAGGCTGTCAGAGTGGAGCAGGTCCAGAAAACTGGGGTCTGGGTGCACAACCATAATG
GAATTGCCGACAACGCTACAATCGTCCGGGTGGATGTGTTCGAGAAAGGCGGGAAGTACTATCTGGTGCCTATCT
ACTCCTGGCAGGTCGCCAAGGGAATCCCTGCCAGATAGAGCTGTGTCGTGCAGGGCAAAGACGAAGAGGATTGGACT
GTGATGGACGATTCTTTCGAGTTTAAGTTCGTCCTGTACGCAAACGACCTGATCAAGCTGACAGCCAAGAAAAAT
GAATTTCTGGGGTATTTCGTGTCACTGAACAGGGCAACTGGAGCCATCGATATTCGCACACATGACACTGATAGC

FIG. 13U

	ACCAAGGGAAAAAACGGCATCTTTCAGTCTGTGGGCGTCAAGACCGCCCTGAGTTTCCAGAAATATCAGATTGAC GAACTGGGGAAGGAGATCCGACCCTGTCGGCTGAAGAAACGACCACCCGTGCGGTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1082
PAM	**NNNNGTAA**
Longitud del espaciador	30
Especie: **Parvibaculum lavamentivorans**	
Secuenciación código de barras	CGCCTTCC
ID	20
ARN guía químérico	**NNNNNNNNNNNNNNNNNNNNGCUGCGGAUUGCGGCGAAAUCGCUUUUCGCAAGCAAAUUGACCCCCUUGUGC GGGCUCGGCAUCCCCAAGGUCAGCUGCCGGUUAUUAUCGAAAAGGCCCACCGCAAGCAGCGCGUGGGCCU UUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQQRRQKRMMRRQLRRRRIRRKALNETLHEA GFLPAYGSADWPVVMADEPYELRRRGLEEGLSAYEFGRAIYHLAQHRHFKGRELEESDTPDPDVDDEKEAANERAAT LKALKNEQTTLGAWLARRPPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEMRARISDTIFAQRPVFWRKNT LGECRFMPGEPLCPKGSWLSQQRRMLEKLNNLAIAGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQR GEPGAEKSLKFNLELGGESKLLGNALEAKLADMFGPDWPAHPRKQEIRHAVHERLWAADYGETPDKKRVIILSEKDR KAHREAAANSFVADFGITGEQAAQLQALKLPTGWEPYSIPALNLFLAELEKGERFGALVNGPDWEGWRRTNFPFHRNQ PTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIRIEVGRDVGKSKRERAEEIQSGIRRNEKQ RKKATEDLIKNGIANPSRDDVEKWILWKEGQERCPYTGDQIGFNALFREGRYEVEHIWPRSRSFDNSPRNKTLCRKDV NIEKGNRMPFEAFGHDEDRWSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQILAQLK RLWPDMGPEAPVKVEAVTGQVTAQLRKLWTLNNILADDGEKTRADHRHHAIDALTVACTHPGMTNKLSRYWQLRD DPRAEKPALTPPWDTIRADAEKAVSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGELD EIRDPRIKEIVAAHVAGRGGDPKKAFPPYPCVSPGGPEIRKVRLTSKQQLNLMAQTGNGYADLGSNHHIAIYRLPDGKA DFEIVSLFDASRRLAQRNPIVQRTRADGASFVMSLAAGEAIMIPEGSKKGIWIVQGVVASGQVVLERDTDADHSTTTR PMPNPILKDDAKKVSIDPIGRVRPSND

FIG. 13V

Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC
	CATGGAGAGGATTTTCGGCTTTGACATCGGCACAACAAGTATCGGATTCAGCGTGATTACAGTAGCACCCA
	GTCCGCAGGCAACATCCAGAGGCTGGCGCATTTTCCCTGAGGCAAGGGACCCAGATGGGACCCCCCTGA
	ACCAGCAGCGGAGACAGAAACGCATGATGAGGCGCCAGCTGCGACGGAGAAGGATTCGCCGAAAGGCACTGAA
	TGAGACACTGCACGAAGCCGGCTTTCTCGCCAGCTTACGGGTCTGCAGATTGGCCCGTGGTCATGGCCGACGAGCC
	TTATGAACTGCGGAGAAGGGGACTGGAGGAAGGCCTGAGTGCTTACGAGTTCGGACGGGCAATCTATCATCTGG
	CCCAGCACCGGCATTTTAAAGGCAGAGAACTGGAGGAATCCGATACACCCGACCCTGATGTGGACGATGAGAAG
	GAAGCCGCTAACGAGAGAGCAGCCACTCTGAAGGCCCTGAAAAATGAACACAGACCACACTGGGAGCATGGCTGG
	CCCGCCGACCCCCTTCTGACCGCAAGGCGAGGAATCCACGCCCATAGGAACGTGGTCGCTGAGGAGTTCGAGCGC
	CTGTGGGAAGTGCAGTCCAAGTTTCACCCCGCTCTGAAATCTGAGGAAATGCGGGCAAGAATCAGTGATACAATT
	TTCGCCCAGAGGCCTGTGTTTTGGCGCAAGAACACTCTGGGAGAGTGCAGATTCATGCCTGGCGAACCACTGTGT
	CCCAAGGGGTCCTGGCTGTCTCAGCAGCGGAGAATGCTGGAGAAACTGAACAATCTGGCTATCGCAGGCGGGAA
	TGCTAGGCCACTGGATGCAGAGGAACGCGACGCCATTCTGAGTAAGCTGCAGCAGGCCAGCATGTCCTGGC
	CAGGCGTGCGGTCAGCTCTGAAGGCACTGTACAAACAGAGAGGCGAGCCCGGGGCTGAAAAGAGCCTGAAATTC
	AACCTGGAGCTGGGAGGCGAATCCAAGCTGCTGGGAAATGCCCTGGAGGCTAAACTGGCAGATATGTTTGGCCC
	TGACTGGCCAGCTCACCCCCGAAAGCAGGAGATCCGGCACGCAGTGCATGAACGGCTGTGGGCTGCAGATTACG
	GCGAGACACCCGACAAGAAAAGAGTCATCATTCTGTCCGAGAAGGATCGAAAAAGCTCATCGGAAGCCGCTGCA
	AACTCTTTCGTGGCAGACTTTGGAATTACTGGCGAGCAGGCAGCTCAGCTGCAGGCCCTGAAGCTGCCAACCGGC
	TGGGAACCTTATAGCATCCCAGCACTGAACCTGTTCCTGGCCGAGCTGGAAAAGGGGGAGAGGTTTGGAGCCCT
	GGTGAATGGACCTGATTGGGAAGGCTGGAGGCGCACAAACTTCCCCCACCGCAATCAGCCTACTGGGGAGATCC
	TGGACAAGCTGCCCAAGTCCCGCCTCAAAAGAGGGAAAGGAACGCATTAGCCAGCTGCGCAACCCAACCGTGGTC
	CGAACACAGAATGAGCTGAAGAAAGGTGGTCAACAATCTGATCGGGCTGTATGGAAAACCCGATCGAATCCGGAT
	TGAAGTGGGCCGGGACGTCGGGAAGTCCAAAAGAGAAAGGGAGGAAATCCAGTCTGGCATTCGACGGAACGAG
	AAGCAGAGAAAGAAAGCCACTGAAGATCTGATCAAAAACGGAATTGCTAATCCTAGCCGGGACGATGTGGAGA
	AGTGGATCCTGTGGAAAGAGGGCCAGGAAAGATGCCCATACACCGGCGACCAGATTTGGCTTCAATGCCCTGTTT
	AGAGAAGGCAGATATGAGGTGGAACACATCTGGCCTCGCTCTCGAAGTTTTGATAACAGCCCAAGGGAATAAGAC
	ACTGTGTCGCAAAGACGTGAACATCGAGAAGGGAAATAGGATGCCTTTCGAGGCATTTGGCCATGACGAAGATC
	GGTGGAGCGCCATCCAGATTAGACTGCAGGGCATGTGTGTCAGCCAAAGGGGGAACTGGGATGAGCCCCGGAAA
	GGTCAAACGCTTCCTGGCTAAGACCATGCCTGAGGATTTTGCAGCCCGGCAGCTGAACGACACAAGATACGCTGC
	AAAGCAGATCCTGGCCCAGCTGAAAAGGCTGTGGCCAGACATGGGACTCTGAAGGCTCCAGTGAAGGTCGAAGCAG
	TGACTGGACAGGTCACCGCCCAGCTGCGCAAACTGTGGACTCTGAACAATATTCTGGCTGACGATGGGGAGAAA
	ACCAGAGCAGATCACAGGCACCATGCCATCGACGCTCTGACACGTGGCCTGCACTCATCCTGGAATGACCAACAA
	GCTGAGCAGGTATTGGCAGCTGCGCGACGATCCACGAGCAGAAGCCAGCTCTGACTCCACCCTGGGATACCA
	TCCGCGCCGACGCTGAGAAAGCCGTGTCTGAAATTGTGGTCAGTCACCGGGTGAGAAAGAAAGTCAGCGGCCCA
	CTGCATAAGGAGACTACCTACGGCGATACAGGGACTGACATTAAGACCAAATCCGGCACATATAGACAGTTCGT

FIG. 13W

	GACCAGGAAGAAAATCGAGTCACTGAGCAAGGGGGAGCTGGATGAAATTCGCGACCCCCGAATCAAAGAAATT GTGGCAGCTCACGTCGCAGGACGAGGAGGCGACCCCAAGAAGGCCTTCCCTCCATACCCCTGTGTGTCTCCCGGA GGCCCTGAGATCCGGAAGGTCAGACTGACCAGTAAACAGCAGCTGAACCTGATGGCCCAGACAGGGAATGGATA CGCTGACCTGGGCTCCAACCACCATATCGCAATCTACCGGCTGCCCGATGGGAAGGCCGACTTCGAGATTGTGTC ACTGTTTGATGCTAGCAGAAGGCTGGCACAGAGAAATCCAATCGTGCAGAGGACACGAGCAGACGGAGCCAGCT TCGTCATGTCCCTGGCAGCCGGAGAGGCCATCATGATTCCCGAAGAAAGGGATCTGGATTGTGCAG GGAGTCTGGGCAAGCGGACAGGTGGTCCTGGAGAGGGACACCGATGCTGACCACTCTACAACTACCCGCCCTAT GCCAAACCCCATCCTGAAGGACGATGCCAAGAAAGTGAGTATCGATCCTATTGGCCGAGTCCGGCCATCAAATG ACTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1037
PAM	**NNNCAT**
Longitud del espaciador	30
Especie: **Staphylococcus aureus subsp. Aureus – double check**	
Secuenciación código de barras	CTATGCGT
ID	21
ARN guía químérico	**NNNNNNNNNNNNNNNNNNNNGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGCAAAAUGCCGUG UUUAUCUCGUCAACUUGUUGGCGAGAUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTD HSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERL KKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMG HCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYR VTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAR EKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIP RSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQK

FIG. 13X

	DFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKK DNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYE EAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKK YSTDILGNLYEVKSKKHPQIIKKG
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGGTC CATGAAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAA CAAGGACCGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAG CAAGAGGGGAGCCAGGCGCCTGAAACGACGGAAGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGAT TACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAAGGCCTGAGTCAG AAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGTGCATAACGTCAATGA GGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTCACGCAATAGCAAAGCTCTGGAAGAG AAGTATGTCGCAGAGCTGCAGCTGGAAGAAGATGGCGAGGTGAGAGGGTCAATTATAATAGGTTCAA GACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGGCTTACCACCAGCTGGATCAGAGCT TCATCGATACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTC GGATGGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAGAAGAGCTGAGAAG CGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGA AAACGAGAAACTGGAATACTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACAC TGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAAA ACCAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAAATCATTGAGAACG CCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAGCTGACTA ACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAA CCTGTCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAA CCGGCTGAAGCTGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACACACTGGTGGACGATT TCATTCTGTCACCCGTGGTCAAGCCGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACG GCCTGCCCAATGATATCATTATCGAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAG ATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCCGAACTACCGGGAAAGAGAACGCAA AGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCC TGGAGGACCTGCTGAACAATCCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAATT CCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTTCCAGTACCTG TCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGC ATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTAT TAACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCTATTTCCGGGTGAA CAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCACATCTTTTCTGAGGCGCAAATGGAAGTTTTAAAAA

FIG. 13Y

	GGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGG AGTGGAAAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCGAATCTAT GCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTTCATCACTCCTCACCAGATCAAGCATATCAAGGATT TCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGT ACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACA AGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAAA CTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTA CCTGACCAAGTATAGCAAAAAGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATG CCCATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGAT TCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAAAAGGAGAACT ACTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAAGCTGAAAAAGATTAGCAACCAGGCAGAGTTCATC GCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGGTCATCGGGGTGAACAATGATCTG CTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAGCGCCCC CCTCGAATTATCAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTG TATGAGGTGAAGAGCAAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1053
PAM	**NNGRRT**
Longitud del espaciador	30
Especie: **Mycoplasma gallisepticum str. F**	
Secuenciación código de barras	CGCTATGT
ID	23
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNNGUUUUAGCACUGUACAAGAAAUUGUCGUGCUACAAAAUAAGGCGCCUGUUAAU GCAGCUGCCCGCAUCCGCCAGAGCAUUUAUGCUCUGGCUUUUUUUU**
Secuencia de proteína Cas9	MNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRLIRRRKYKLKRFVNLIWKYN SYFGFKNKEDILNNYQEQQKLHNTVLNLKSEALNAKIDPKALSWILHDYLKNRGHFYEDNRDFNVYPTKELAKYFDK

FIG. 13Z

(de tipo salvaje sin NLS añadida)	YGVYKGIIDSKEDNDNKLEEELTKYKFSNKHWLEEVKKVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGI YHLDEKEGKVVQKYNNIWDKTIGKCNIFPDEYRAPKNSPIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDL LIKTNGEKPIDARQFKKLREETIAESIGKETLKDVENEEKLEKEDHKWKLKGLKLNTNGKIQYNDLSSLAKFVHKLKQ HLKLDFLLEDQYATLDKINFLQSLFVYLGKHLRYSNRVDSANLKEFSDSNKLFERILQKQKDGLFKLFEQTDKDDEKIL AQTHSLSTKAMLLAITRMTNLDNDEDNQKNNDKGWNFEAIKNFDQKFIDITKKNNNLSLKQNKRYLDDRFINDAILSP GVKRILREATKVFNAILKQFSEEYDVTKVVIELARELSEEKELENTKNYKKLIKKNGDKISEGLKALGISEDEIKDILKSP TKSYKFLLWLQQDHIDPYSLKEIAFDDIFTKTEKFEIDHIIPYSISFDDSSSNKLLVLAESNQAKSNQTPYEFISSGNAGIK WEDYEAYCRKFKDGDSSLLDSTQRSKKFAKMMKTDTSSKYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKP MFKVVCINGSVTSFLRKNFDDSSYAKKDRDKNIHHAVDASIISIFSNETKTLFNQLTQFADYKLFKNTDGSWKKIDPKT GVVTEVTDENWKQIRVRNQVSEIAKVIEKYIQDSNIERKARYSRKIENKTNISLFNDTVYSAKKVGYEDQIKRKNLKTL DIHESAKENKNSKVKRQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINLAEQIFNEYTENKKIKSQNVFEK YMLDLTKEFPEKFSEFLVKSMLRNKTAIIYDDKKNIVHRIKRLKMLSSELKENKLSNVIIRSKNQSGTKLSYQDTINSLA LMIMRSIDPTAKKQYIRVPLNTLNLHLGDHDFDLHNMDAYLKKPKFVKYLKANEIGDEYKPWRVLTSGTLLIHKKDK KLMYISSFQNLNDVIEIKNLIETEYKENDDSDSKKKKKANRFLMTLSTILNDYILLDAKDNFDILGLSKNRIDEILNSKLG LDKIVK
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGAACAATAGCATCAAATCTAAACCTGAAGTGACCATCGGGCTGGACCTGGGAGTGGGAAGCGTGGGGTGGG CAATCGTGGATAACGAAACAAACATCATTCACCATCTGGGCTCCAGGCTGTTTTCTCAGGCCAAGACTGCTGAGG ATCGGAGATCTTTCCGCGGGGTGAGGCGCCTGATCCGACGGAGAAAATACAAGCTGAAACGATTCGTCAATCTG ATTTGGAAGTACAACACAGCTATTTCGGCTTCAAGAACACAAAGAGGACATCCTGAACAATTTATCAGGAGCAGCAGAA GCTGCACAATACCGTGCTGAACCTGAAATCAGAGGCACTGAATGCCAAGATCGATCCTAAAGCACTGAGCTGGA TTCTGCACGACTACCTGAAGAACAGAGGCCATTTTTATAGAGGACAATAGGGATTTCAACCGTGTACCCAACAAAGG AGCTGGCCAAGTACTTCGATAAGTACGGGTACTACAAGGGAATCATTGACAGCAAGGAGGACAATGATAACAAA CTGGAGGAAGAGCTGACAAAGTACAAATTCTCCAATAAGCACTGGCTGGAAGAGGTGAAGAAATTATTCAGAGG CCAGACTGGCCTGCCAGAAAAGTTTAAAGAAGAGTATGAGTCACTGTTCAGCTACGTGAGAAATTATTCAGAGG GCCCAGGGAGCATCAACTCTGTCAGTCCCTACGGGATCTCACCATCTGGACGAAAAAGAGGGAAAGGTGGTCCAG AAGTACAACAACATCTGGGATAAGACAATCGGAAAGTGCAACATCTTCCCTGACGAGTATAGAGCTCCCAAGAA CAGTCCTATCGCAATGATTTTCAATGAAATCAACGAGCTGTCCACAATCAGGTCATACAGCATCTACCTGACTGG CTGGTTCATTAATCAGGAGTTCAAGAAAGCCTACCTGAACAAGCTGCTGGATCTGCTGATCAAAACCACAACGGAG AGAAGCCAATTGACGCAAGGCAGTTCAAGAAACTGCGCGAAGAGACAATCGCCGAAAGCATTGGCAAAGAGAC ACTGAAGGATGTGGAGAATGAAGAGAAACTGGAAAAGGAGGACCACAAGTGGAAACTGAAGGGACTGAAGCTG AATACCAACGGCAAAATCCAGTACAACGATCTGAGCTCCCTGGCTAAGTTTGTGCACAAACTGAAGCAGCATCTG AAACTGGATTTCCTGCTGGAGGACCAGTATGCAACACTGGACAAGATCAATTTCCTGCAGTCCCTGTTTGTGTAC CTGGGCAAGCACCTGAGATATTCCAATAGGGTCGATTCTGCCAACCTGAAGGAATTTTCCGACTCTAACAAACTG TTCGAGCGCATCCTGCAGAAACAGAAGGATGGGCTGTTCAAGCTGTTTGAACAGACTGACAAAGACGATGAGAA

FIG. 13AA

	GATCCTGGCCCAGACACATAGTCTGTCAACTAAGGCCATGCTGCTGGCTATTACCCGGATGACAAATCTGGACAA CGATGAGGACAACCAGAAAAACAATGACAAGGGCTGGAATTTTGAGGCCATCAAAAACTTCGATCAGAAGTTTA TCGACATCACCAAGAAAAACAACAACCTGAGCCTGAAACAGAATAAGCGCTACCTGGACGATCGATTCATCAAC GATGCTATTCTGTCCCCTGGGGTGAAGCGAATCCTGCGGGAGGCAACCAAGGTCTTTAATGCCATTCTGAAACAG TTCTCTGAAGAGTACGACGTGACAAAGGTGGTCATCGAACTGGCTCGCGAGCTGAGCAAGGAAGAGAAGGAACTGGA GAACACAAAGAACTACAAGAAACTGATCAAGAAAAACGGCGACACCCAGCGGTCCAAGAAATTCGCCAA ATCTCAGAAGATGAGATCAAAGACATTCTGAAGGAGATCGCCTTCGACGATATTTTTACCAAAACAGAAAAGTTCGAGATC GACCACATCGATCCTTATAGCCTGAAGGAGATCGCCTTCGACGATATTTTTACCAAAACAGAAAAGTTCGAGATC GACCATATCATTCCCTACAGCATTTCCTTCGACGATTCTAGTTCAAACAAGCTGCTGGTGCTGGCTGAAAGTAATC AGGCAAAGTCAAACCAGACTCCTTATGAGTTCATCAGCTCCGGAAAACGCAGGCATTAAGTGGGAAGATTACGAG GCCTATTGCCGCAAGTTCAAGGATGGGGACTCTAGTCTGCTGGACAGCACCCAGCGGTCCAAGAAATTCGCCAA AATGATGAAAACCGATACCTCAAGCAAGTACGACATCGGATTTCTGGCTCGAAATCTGAACGATACTCGGTACGC AACCATTGTGTTCCGGGACGCCCTGGAGGACTATGCTAATAACCACCTGGTCGAGGACAAACCCATGTTTAAGGT GGTCTGTATCAATGGGTCCGTGACCTCTTTCCTGCGGAAGAACTTTGACGATTCCTCTTACGCCAAGAAAGATAG AGACAAGAATATCCACCATGCTGTGGATGCAAGTATCATCTCAATTTTCAGCAACGAGACAAAGACTCTGTTCAA CCAGCTGACTCAGTTTGCTGACTATAAACTGTTCAAGAACACCGATGGCAGCTGGAAGAAATCGACCCTAAGA CAGGGGTGGTCACTGAAGTGACCGACGAGAATTGGAAGCAGATTAGGGTGCGCAACCAGGTGAGCGAAATCGCC AAAGTCATTGAGAAGTACATCCAGGATAGCAACATCGAAAGAAAGGCTAGGTATTCCCGCAAATCGAGAATAA GACTAACATTTCCCTGTTTAATGACACCGTGTACTCTGCCAAGAAAGTCGGCTATGAGGATCAGATCAAAAGAAA GAACCTGAAAACCCTGGACATTCACGAATCTGCTAAAGAGAATAAGAACAGTAAAGTGAAGCGGCAGTTTGTCT ACAGAAAGCTGGTGAATGTCAGCCTGCTGAATAACGATAAGCTGGCAGACCTGTTCGCCGAAAAAGAGGATATC CTGATGTATAGGGCCAATCCATGGGTCATCAACCTGGCTGAGCAGATTTTCAATGAATACACTGAGAACACAAGAA AATCAAGTCCCAGAACGTGTTTGAAAAATATATGCTGGACCTGACCAAAGAGTTCCCCGAGAAGTTCAGCGAGTT TCTGGTGAAGTCCATGCTGAGAAACAAGACCGCCATCATCTACGACGATAAGAAAAACATTGTCCATCGAATCA AACGGCTGAAGATGCTGAGTTCAGAACTGAAAGAGAATAAGCTGTCTAACGTGATCATTAGGTCTAAGAATCAG AGTGGGACCAAACTGTCATACCAGGATACAATCAACAGCCTGGCCCTGATGATTATGCGCAGCATCGACCCTACT GCTAAGAAACAGTATATTCGAGTGCCACTGAATACCCTGAACCTGCACCTGGGAGATCATGACTTTGATCTGCAC AATATGGATGCTTACCTGAAGAAACCAAAATTCGTGAAGTATCTGAAAGCAAACGAAATCGGCGACGAGTACAA GCCCTGGAGGGTCCTGACATCTGGCACTCTGCTGATCCATAAGAAGGATAAGAACTGATGTACATCAGCTCCTT CCAGAATCTGAACGACGTGATCGAAATTAAGAATCTGATCGAAACCGAGTATAAAGAGAACGACGATTCTGATA GTAAGAAAAAGAAAAAGGCAAACCGCTTTCTGATGACCCTGAGCACAATCCTGAATGACTACATTCTGCTGGAC GCCAAGGATAACTTCGACATCCTGGGGCTGTCTAAAAATCGGATCGATGAGATTCTGAACAGTAAGCTGGGACT GGACAAGATTTGTGAAATAAGAATTC
Longitud de proteína	1269

FIG. 13BB

(sin NLS y marcador HA añadido)	
PAM	**NNGAT**
Longitud del espaciador	30
Especie: **Campylobacter lari CF89-12**	
Secuenciación código de barras	CTGTGGCG
ID	24
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNGUUUUAGUCUCUGAAAAGAGACUAAAAUAAGUGGUUUUUGGUCAUCCACG CAGGGUUACAAUCCCUUUAAAACCAUUAAAAUUCAAAUAAACUAGGUUGUAUCAACUUAGUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MRILGFDIGINSIGWAFVENDELKDCGVRIFTKAENPKNKESLALPRRNARSSRRRLKRRKARLIAIKRILAKELKLNYK DYVAADGELPKAYEGSLASVYELRYKALTQNLETKDLARVILHIAKHRGYMNKNEKKSNDAKKGKILSALKNNALK LENYQSVGEYFYKEFFQKYKKNTKNFIKIRNTKDNYNNCVLSSDLEKELKLILEKQKEFGYNYSEDFINEILKVAFFQR PLKDFSHLVGACTFFEEEKRACKNSYSAWEFVALTKIINEIKSLEKISGEIVPTQTINEVLNLILDKGSITYKKFRSCINLH ESISFKSLKYDKENAENAKLIDFRKLVEFKKALGVHSLSRQELDQISTHITLIKDNVKLKTVLEKYNLSNEQINNLLEIEF NDYINLSFKALGMILPLMREGKRYDEACEIANLKPKTVDEKKDFLPAFCDSIFAHELSNPVVNRAISEYRKVLNALLKK YGKVHKIHLELARDVGLSKKAREKIEKEQKENQAVNAWALKECENIGLKASAKNILKLKLWKEQKEICIYSGNKISIE HLKDEKALEVDHIYPYSRSFDDSFINKVLVFTKENQEKLNKTPFEAFGKNIEKWSKIQTLAQNLPYKKKNKILDENFKD KQQEDFISRNLNDTRYIATLIAKYTKEYLNFLLLSENENANLKSGEKGSKIHVQTISGMLTSVLRHTWGFDKKDRNNH LHHALDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFEPFKSFREKILSKIDEIFVSKPPRKRARR ALHKDTFHSENKIIDKCSYNSKEGLQIALSCGRVRKIGTKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILPNKIVIT GKDKNNNPKQWQTIDESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSISTSSICVEKHDNKFENLTSNQKLLFSNAK EGSVKVESLGIQNLKVFEKYIITPLGDKIKADFQPRENISLKTSKKYGLR
Secuencias optimizadas en codones de mamíferos (AgeI-HA-	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CATGAGGATTCTGGGGTTTGACATTGGCATTAACAGCATCGGGTGGGCTTTTGTGGAGAACGACGAACTGAAGG ACTGCGGAGTGCGGATCTTCACAAAGGCCGAGAACCCAAAAAATAAGGAAAGCCTGGCACTGCCCCGGAGAAAT GCACGCAGCTCCAGGCGCCGACTGAAACGGAGAAAGGCCCGGCTGATCGCTATTAAGAGAATCCTGGCCAAAGA GCTGAAGCTGAACTACAAGGACTATGTCGCAGCTGATGGAGAGCTGCCAAAGGCCTACGAAGGATCCCTGGCAT CTGTGTACGAGCTGCGGTATAAGGCCCTGACACAGAACCTGGAAACTAAAGATCTGGCCAGAGTGATCCTGCAC

FIG. 13CC

NLS-Cas9-NLS-taa-EcoR1)	ATTGCTAAGCATAGGGGGTACATGAACAAGAACGAGAAGAAATCAAACGACGCTAAGAAAGGAAAGATCCTGA GCGCTCTGAAAAAACAATGCACTGAAGCTGGAGAACTACCAGAGCGTGGGCGAATACTTCTACAAGGAGTTCTTT CAGAAATACAAGAAAAACACAAAGAACTTCATCAAGATCCGCAACACTAAGGATAATTACAACAATTGCGTGCT GTCTAGTGACCTGGAAAAAGAGCTGAAGCTGATCCTGGAAAAACAGAAGGAGTTCGGCTACAACTACTCTGAAG ATTTCATCAACGAGATTCTGAAGGTCGCCTTCTTTCAGCGGCCCCTGAAGGACTTCAGTCACCTGGTGGGGCCT 'GCACTTTCTTTGAGGAAGAGAAAAGGGCCTGTAAGAACAGCTACTCTGCCTGGGAGTTTGTGGCTCTGACCAAGA TCATTAACGAGATCAAGAGCCTGGAGAAGATCAGCGGCGAAATTGTGCCAACCCAGACAATCAACGAGGTCCTG AATCTGATCCTGGACAAGGGGTCTATCACCTACAAGAAATTCAGAAGTTGTATCAATCTGCATGAGAGTATCAGC TTCAAGAGCCTGAAGTATGATAAAGAAAACGCCGAGAATGCTAAACTGATCGACTTCCGCAAGCTGGTGGAGTT TAAGAAAGCCCTGGGAGTCCACAGCCTGTCCCGGCAGGAACTGGATCAGATCTCCACTCATATCACCCTGATTAA GGACAACGTGAAGCTGAAAACCGTCCTGGAGAAATACAACCTGAGTAATGAACAGATCAACAATCTGCTGGAAA TTGAGTTCAACGATTATATCAACCTGAGCTTCAAGGCCCTGGGAATGATTCTGCCACTGATGCGCGAGGGCAAAC GATACGACGAGGCCTGCGAGATCGCCAATCTGAAACCTAAGACCGTGGACGAAGAAAGATTTCCTGCCAGCA TTTTGTGATTCCATTTTCGCCCACGAGCTGTCTAACCCCGTGGTCAATAGGGCTATCAGCGAATACCGCAAGGTGC TGAACGCACTGCTGAAGAAATATGGAAAGGTCCACAAAATTCATCTGGAGCTGGCTCGCGACGTGGGCCTGTCC AAGAAAGCACGAGAGAAGATCGAAAAAGAGCAGAAGGAAAACCAGGCCGTGAATGCATGGGCCCTGAAGGAAT GCGAGAATATTGGCCTGAAGGCCAGCGCAAAGAACATCCTGAAACTGAAGCTGTGGAAAGAACAGAAGGAGAT CTGTATCTACTCCGGAAATAAGATCTCTATTGAGCACCTGAAAGATGAAAAGGCCCTGGAGGTGGACCATATCTA CCCCTATTCTAGGAGTTTCGACGATTCTTTTATCAACAAAGTGCTGGTGTTCACCAAGGAAAATCAGGAGAAACT GAACAAGACACCTTTCGAGGCCTTTGGCAAGAATATTGAAAAATGGAGCAAGATCCAGACCCTGGCTCAGAACC TGCCATACAAGAAAAAGAATAAGATTCTGGACGAAGAACTTCAAAGATAAGCAGCAGGAGGACTTTATCTCTCGA AATCTGAACGACACCCGGTATATCGCTACACTGATTGCAAAATACACAAAGGAGTATCTCGAACTTCCTGCTGCTG AGCGAAAATGAGAACGCCAATCTGAAGAGTGGCGAAAAAGGGTCAAAGATCCACGTGCAGACTATTAGCGGGA TGCTGACCTCCGTCCTGAGGCACACATGGGGGTTTGACAAAAAGGATCGCAACAATCATCTGCACCATGCACTGG ATGCCATCATTGTGGCCTACAGTACAAATTCAATCATTAAGGCTTTCAGCGATTTCCGGAAAAACCAGGAGCTGC TGAAGGCCAGATTCTACGCTAAAGAACTGACTTCCGATAACTATAAACATCAGGTCAAGTTCTTTGAGCCTTTCA AGAGTTTTAGAGAAAAAATCCTGTCAAAGATCGACGAGATTTTCGTGTCCAAACCACCTCGAAAGCGAGCTAGG CGCGCACTGCACAAGGATACCTTTCATTCTGAGAACAAGATCATTGACAAGTGCAGCTACAACTCCAAGGAAGG CCTGCAGATTGCCCTGAGCTGTGGAAGAGTGAGGAAAATCGGCACTAAGTATGTCGAGAATGATACCATCGTGA GGGTCGACATTTTCAAAAAGCAGAACAAGTTTTACGCTATCCCAATCTACGCAATGGATTTTGCCCTGGGGATCC TGCCCAATAAGATCGTGATTACTGGAAAAGATAAGAACAATAACCCCAAACAGTGGCAGACCATTGACGAATCA TACGAGTTCTGCTTTAGCCTGTATAAGAATGACCTGATCCTGCTGCAGAAAAAGAACATGCAGGAACCTGAGTTC GCCTACTATAACGATTTTTCAATCAGCACATCAAGCATTTGTGTGGAGAAACACGACAACAAGTTCGAAAATCTG ACTAGCAACCAGAAGCTGCTGTTTCCAATGCAAAAGAGGGCTCTGTGAAGGTCGAAAGTCTGGGGATCCAGAA CCTGAAAGTGTTCGAGAAGTACATCATTACCCCCCTGGGAGATAAAATTAAGGCTGACTTTCAGCCTCGAGAAAA CATCAGCCTGAAAACCAGTAAAAAGTATGGCCTGAGGTAAGAATTC

FIG. 13DD

Longitud de proteína (sin NLS y marcador HA añadido)	1003
PAM	**NNGGGYA**
Longitud del espaciador	29
Especie: **Streptococcus pyogenes SF370**	
Secuenciación código de barras	CCAGTTAG
ID	25
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU**
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadida)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR ICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST	439

FIG. 13EE

	KEVLDATLIHQSITGLYETRIDLSQLGGD
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACA AGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCC CTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAATACACCAGAC GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCAC AGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGA CGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGG CCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACC TGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGA AAATCTGATCGCCCAGCTGCCCGGCGAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTT GACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACG ACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCG ACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCA AGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTAC AAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTT CTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGG ACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCC ATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTC CGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGA AACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGA CCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGT ATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAA AAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCA AGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATC GTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGA CGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTT CATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCG ATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAG GTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAG ACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACCTG

FIG. 13FF

	TGCAGAATGGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCAT ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGG CAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCA AGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCC GGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGAT GAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGT CCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACGCCCACGACGCCTACC TGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTAC AAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGAT CGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGA GCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCC AAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCC CCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAA GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGG CTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGA AGAGGATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTC CTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGA ACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACG CTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGGCAGGCCGAGAATATC ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAG AGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACG GATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAAGAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1368
PAM	**NGG**
Longitud del espaciador	30
Especie: **Streptococcus thermophilus LMD9 (CRISPR-1)**	

FIG. 13GG

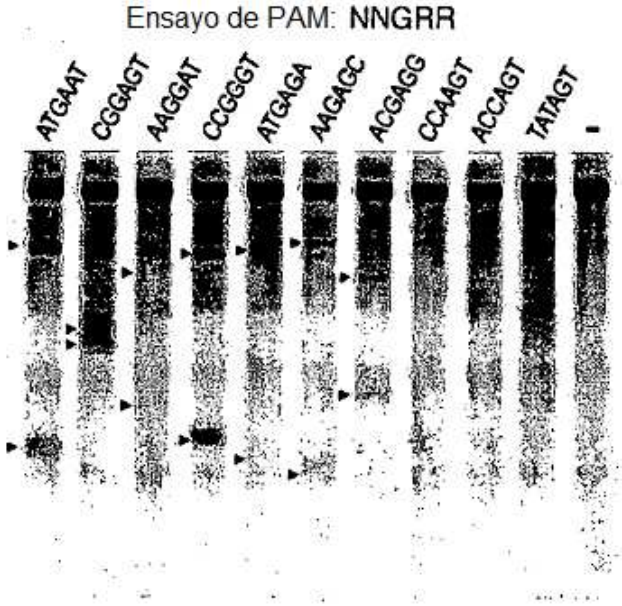
Secuenciación código de barras	TCGGAATG
ID	26
ARN guía quimérico	**NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCUGUUUUGAAUGGUCCCAAAACUCGUACUCUCAAGAUUUAGAAAUAAAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCACCCUGUCAUUUUAUGGCAGGGUGUUUU**
Secuencia de proteína Cas9 (de tipo salvaje en NLS añadida)	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFT KISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERY QTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDY GRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPA KLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEF ADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTE EIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELP HSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPY QALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDT KVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESV FKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIY KKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGN HIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFT LYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKV RTDVLGNQHIIKNEGDKPKLM
Secuencias optimizadas en codones de mamíferos (AgeI-HA-NLS-Cas9-NLS-taa-EcoRI)	ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGGCCGAAGAAAAAGCGCAAGGTCGAAGCGTC CTCCGACCTGGTACTTGGACTGGATATTGGTATCGGTTCGGGTGGGAGTCGGAAATCCTCAACAAGGTCACCGGGGGA GATCATTCACAAGAACTCCGCGGATCTTCCCCGCAGCTCAGGCTGAGAACACAACTTGGTGCGGAGAACGAATAGGC AGGGCAGGCGACTGGCCGAGGAGGAAGAAACACAGGAGAGTCCGATTGAACCGGCTGTTCGAGGAGTCCGGTTT GATCACCGACTTTACGAAAATCTCGATTAACCTTAATCCCTATCAGCTTCGGGTGAAAGGCCTGACAGACGAACT TTCGAATGAGGAACTTTTCATCGCGCTGAAAAACATGGTCAAGCACACAGAGGGATTTCCTACCTCGATGACGCCTC GGATGACGGAAATTCCTCAGTAGGAGATTATGCACAGATCGTGAAAGAGAACTCAAAGCAACTGGAAACAAAG ACACCGGGGCAGATCCAACTTGAAAGATACCAGAGACATACGGACAGCTCAGAGGAGATTTTACGGTGGAGAAGGA CGGTAAAAAGCACAGACTCATTAACGTATTTCCCACGTCGGCGTACAGATCCGAAGCCGCTCCGCATCCTTCAGAC TCAACAGGAGTTCAACCCGCAAATTACTGATGAGTTCATCAACCGCTATTTGGAAATCTTGACCGGAAAAGCGCAA GTATTATCATGGGCCGGGTAATGAGAAATCCAGAACAGATTACGGCCGATACAGAACTTCGGGGGAAACCTTGG ATAACATCTTTGGTATTTTGATTGGAAAGTGCACCTTTTACCCGGACGAGTTTCGAGCGGCCAAGGCGTCATACA CAGCACAAGAGTTTAATCTCTTGAATGATTTGAACAACTTGACGGTCCCCACGGAGACAAAGAAGCTCTCCAAAG AGCAAAAGAACCAAATCATCAACTACGTCAAGAACGAGAAGGCTATGGGGCCAGCGAAGCTGTTCAAGTATATC

FIG. 13HH

	GCTAAACTTCTCAGCTGTGATGTGGCCGGACATCAAAGGGTACCGAATCGACACAAGTCGGGAAAAGCGGAAATTCA
	CACGTTTGAAGCATATCGAAAGATGAAAACGTTGGAAACACTGGACATTGAGCAGATGGACCGGGAAACGCTCG
	ACAAACTGGCATACGTGCTCACGTTGAATACTGAACGAGAGGGAATCCAAGAGGCCCTTGAACATGAGTTCGCC
	GATGGATCGTTCAGCCAGAAGCAGGTCGACGAACTTGTGCAATTCCGCAAGGCGAATAGCTCCATCTTCGGGAA
	GGGATGGCACAACTTTTCGGTCAAACTCATGATGGAGTTGATCCCAGAACTTTATGAGACTTCGGAGGAGCAAAT
	GACGATCTTGACGCGCTTGGGGAAACAGAAAACGACAAGCTCATCGAACAAAACTAAGTACATTGATGAGAAAT
	TGCTGACGGAAGAAATCTATAATCCGGTAGTAGCGAAAATCGGTAAGACAAGCGATCAAAATCGTGAACGCGGCG
	ATCAAGGAATATGGTGACTTTGATAACATCGTAATTGAAATGGCTAGAGAGACGAACGAAGATGACGAGAAAAA
	GGCAATCCAGAAGATCCAGAAGGCCAACAAGGATGAAAAAGATGCAGCGGCTTAAAGCGGCCAACCAATAC
	AATGGAAAGGCGGAGCTGCCCCCATTCAGTGTTTCACGGTCATAAACAGTTGGCGACCAAGATCCGACTCTGGCAT
	CAGCAGGGTGAGCGGTGTCTCTACACCGGAAAGACTATCTCCATCATGACTTGATTAACAATTCGAACCAGTTT
	GAAGTGGATCATATTCTGCCCCTGTCAATCACCTTTGACGACTCGCTTGCGAACAAGGTGCTCGTGTACGCAACG
	GCAAATCAGGAGAAAGGCCAGCGGACTCCGTATCAGGCGCTCGACTCAATGGACGATGCGTGGTCATTCCGGGA
	GCTGAAGGCGTTCGTACGCGAGAGCAAGACACTGAGCAACACAAAAGAAAGAGTATCTGCTGACAGAGGAGGAC
	ATCTCGAAATTCGATGTCAGGAAGAAGTTCATCGAGCGGAATCTTGTCGACACTCGCTACGCTTCCAGAGTAGTA
	CTGAACGCGCTCCAGGAACACTTTAGAGCGCACAAAATTGACACGAAGGTGTCAGTGGTGAGAGGGCAGTTCAC
	ATCCCAACTCCGCCGACATTGGGGCATCGAAAAGACGCGGGACACATATCACCATCATGCGGTGGACGCGCTGA
	TTATTGCCGCTTCGTCCCAGTTGAATCTCTGGAAAAAGCAGAAGAACACGCTGGTGTCGTATTCGGAGGATCAGC
	TTTTGGACATCGAAACCGGGGAGCTGATTTCCGACGATGAATACAAAGAATCGGTGTTTAAGGCACCATATCAGC
	ATTTCGTGGACACGCTGAAGAGCAAAGAGTTTGAGGACAGCATCCTCTTTTCGTACCAAGTGGACTCGAAGTTTA
	ATCGCAAGATTTCAGACGCCACAATCTACGCGACGAGGCAGGCGAAGGTGGGCAAAGATAAAGCAGATGAAAC
	CTACGTCCTTGGTAAAATCAAGGACATCTACACTCAGGACGGTACGATGCGTTCATGAAAATCTACAAGAAGG
	ATAAGTCGAAGTTTCTCATGTACCGCCACGATCCACAGACTTTCGAAAAAGTCATTGAGCCTATTTTGGAGAACT
	ACCCTAACAAGCAAATCAACGAGAAAGGGAAAGAAGTCCCGTGCAACCCCTTTCTGAAGTACAAGGAAGAGCAC
	GGTTATATCCGCAAATACTCGAAGAAAGGAAATGGGCCTGAGATTAAGTCGCTTAAGTATTACGACTCAAAGTTG
	GGTAACCACATCGACATTACCCCGAAAGACTCCAACACAAAGTCGTGTTGCAGTCCGTCTCGCCCCTGGCGAGCA
	GATGTGTATTTTAATAAGACGACCGGCAAATATGAGATCCTTGGACTCAAATACGCAGACCTTCAATTCGAAAAG
	GGGACGGGCACTTATAAGATTTTCACAAGAGAAGTACAACGACATCAAGAAAAAGGAAGGGGTCGATTCAGATTC
	GGAGTTCAAATTCACCCTCTACAAAAACGACCTCCTGCTTGTGAAGGACACAGAAACGAAGGAGCAGCAGCTCT
	TTCGGTTCCTCTCACGCACGATGCCCAAACAAAAACATTACGTCGAACTTAAACCTTACGATAAGCAAAAGTTTG
	AAGGGGGAGAGGCACTGATCAAAGTATTGGGTAACGTAGCCAATAGCGGACAGTGTAAGAAAGGGCTGGGAAA
	GTCCAATATCTCGATCTATAAAGTACGAACAGATGTATTGGGAAACCAGCATATCATCAAAAATGAGGGGGATA
	AACCCAAACTCGATTTCAGCCCCAAGAAGAAGAGGTGGAGGCCAGCTAAGAATTC
Longitud de proteína	1120

FIG. 13II

(sin NLS y marcador HA añadido)	
PAM	**NNAGAA**
Longitud del espaciador	30



Secuencias de espaciador para PAMs ensayadas

Espaciador	PAM
GCCCGGGTGGAACTGGTAGCC	ATGAAAT
GTTGAAGATGAAGCCCAGAG	CGGAGT
GCTTCCGACGAGGTGGCCATC	AAGGAT
GCACCATCTCTCCGTGGTACC	CCGGGT
GGTGGAACTGGTAGCCATGA	ATGAGA
GCCATGAATGAGACCGACCCA	AAGAGC
GCATCCTCGTGGGCACTTCCG	ACGAGG
GCAGAGCGGAGTGCTGTTCTC	CCAAGT
GGTCGGTCTCATTCATGGCT	ACCAGT
GCAATAAAAGGTGCTATTGC	TATAGT

FIG. 14

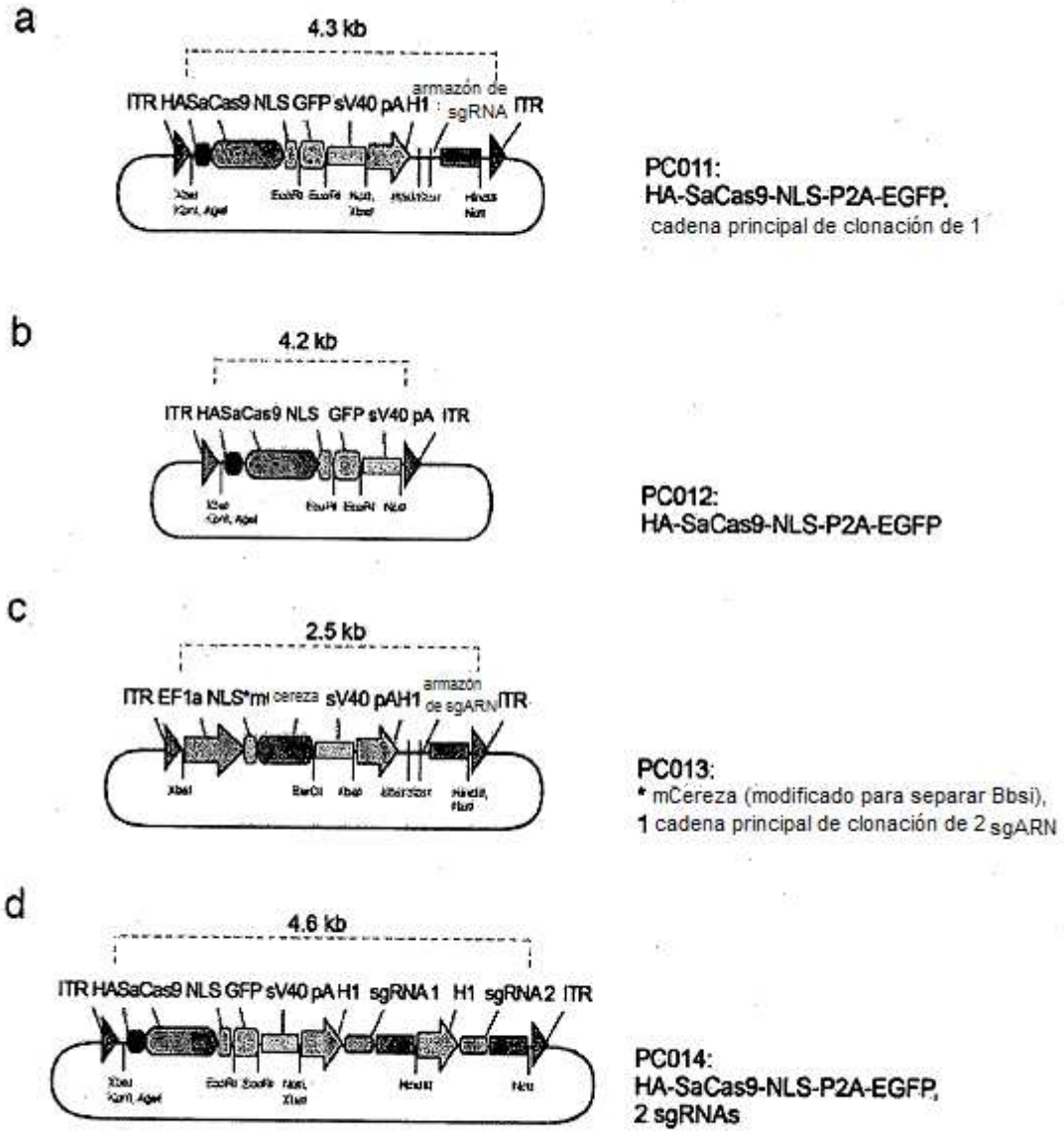
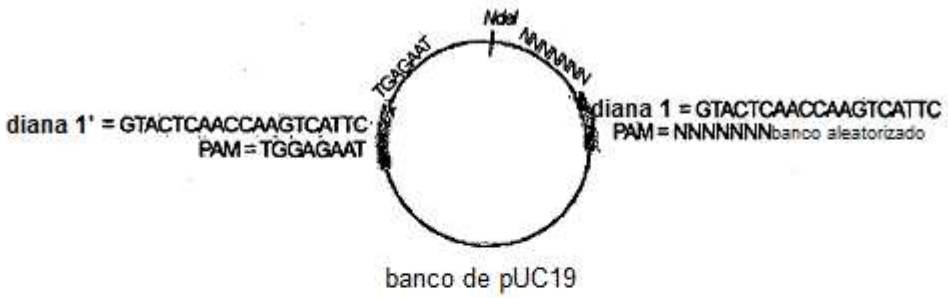

a
4.3 kb
ITR HASaCas9 NLS GFP sV40 pA H1
armazón de sgRNA
ITR
PC011:
HA-SaCas9-NLS-P2A-EGFP,
cadena principal de clonación de 1
b
4.2 kb
ITR HASaCas9 NLS GFP sV40 pA ITR
PC012:
HA-SaCas9-NLS-P2A-EGFP
c
2.5 kb
ITR EF1a NLS*mCereza sV40 pAH1
armazón de sgARN
ITR
PC013:
* mCereza (modificado para separar Bbsi),
1 cadena principal de clonación de 2 sgARN
d
4.6 kb
ITR HASaCas9 NLS GFP sV40 pA H1 sgRNA 1 H1 sgRNA 2 ITR
PC014:
HA-SaCas9-NLS-P2A-EGFP,
2 sgRNAs

FIG. 15


NdeI
NNNNNNN
TGAGAAT
diana 1' = GTACTCAACCAAGTCATTC
PAM = TGGAGAAT
diana 1 = GTACTCAACCAAGTCATTC
PAM = NNNNNNNbanco aleatorizado
banco de pUC19

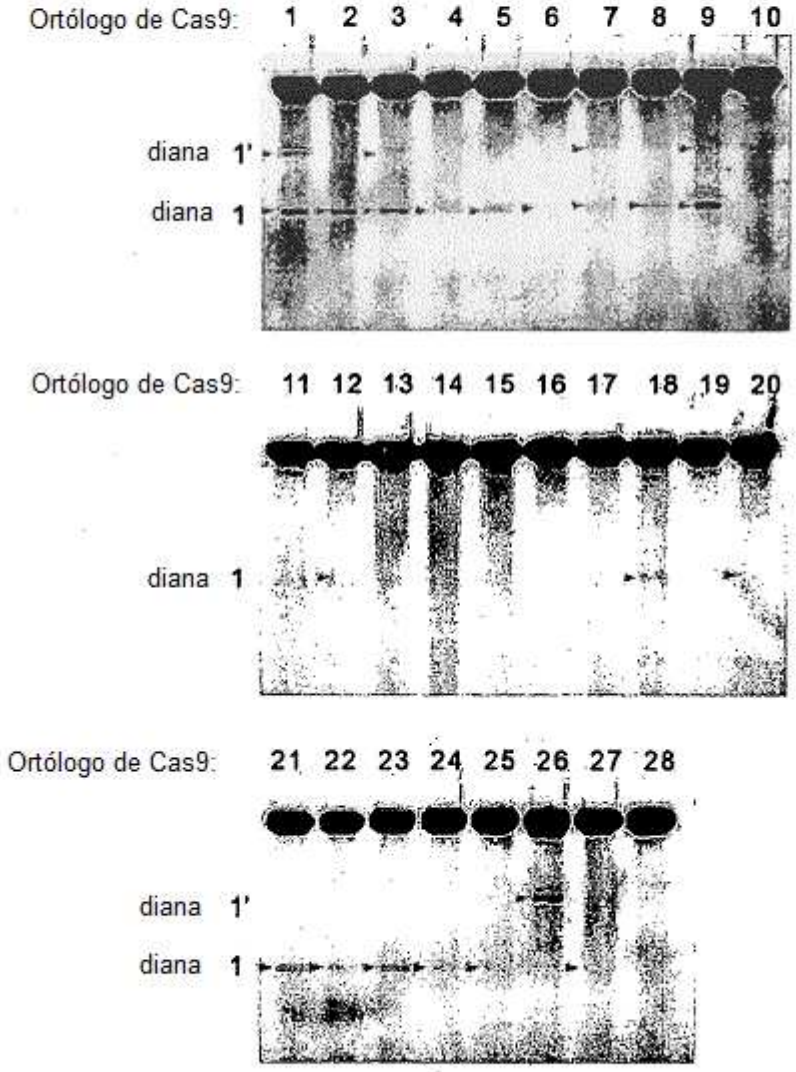

Ortólogo de Cas9: 1 2 3 4 5 6 7 8 9 10
diana 1'
diana 1
Ortólogo de Cas9: 11 12 13 14 15 16 17 18 19 20
diana 1
Ortólogo de Cas9: 21 22 23 24 25 26 27 28
diana 1'
diana 1

FIG. 16

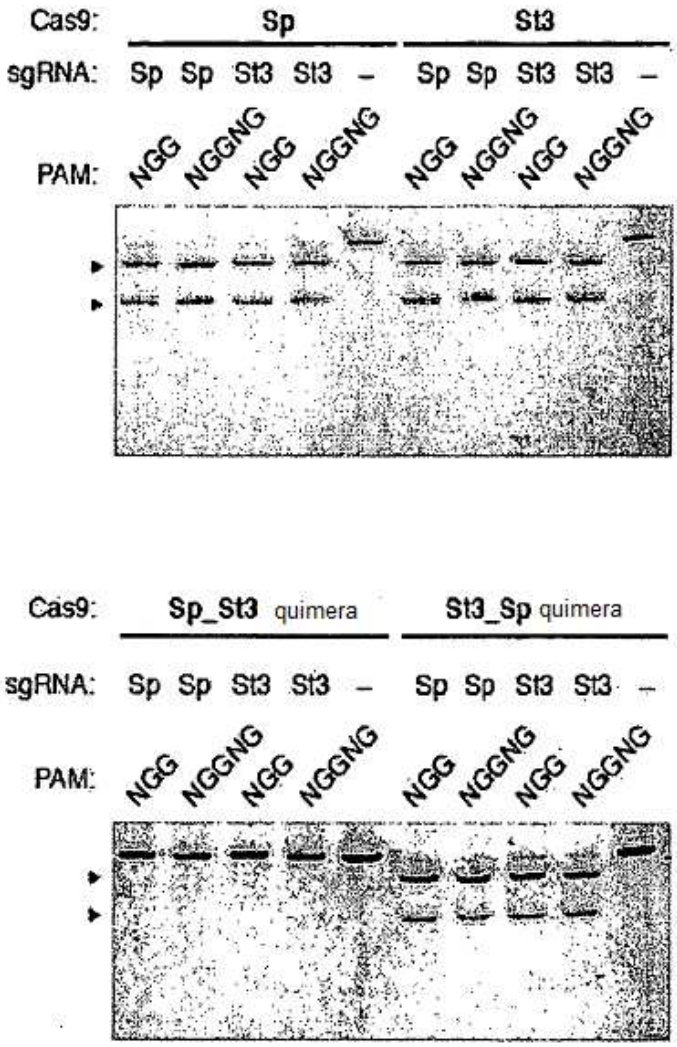

Cas9:
Sp
St3
sgRNA:
Sp Sp St3 St3 – Sp Sp St3 St3 –
PAM:
NGG NGGNG NGG NGGNG NGG NGGNG NGG NGGNG
Cas9:
Sp_St3 quimera
St3_Sp quimera
sgRNA:
Sp Sp St3 St3 – Sp Sp St3 St3 –
PAM:
NGG NGGNG NGG NGGNG NGG NGGNG NGG NGGNG

FIG. 17

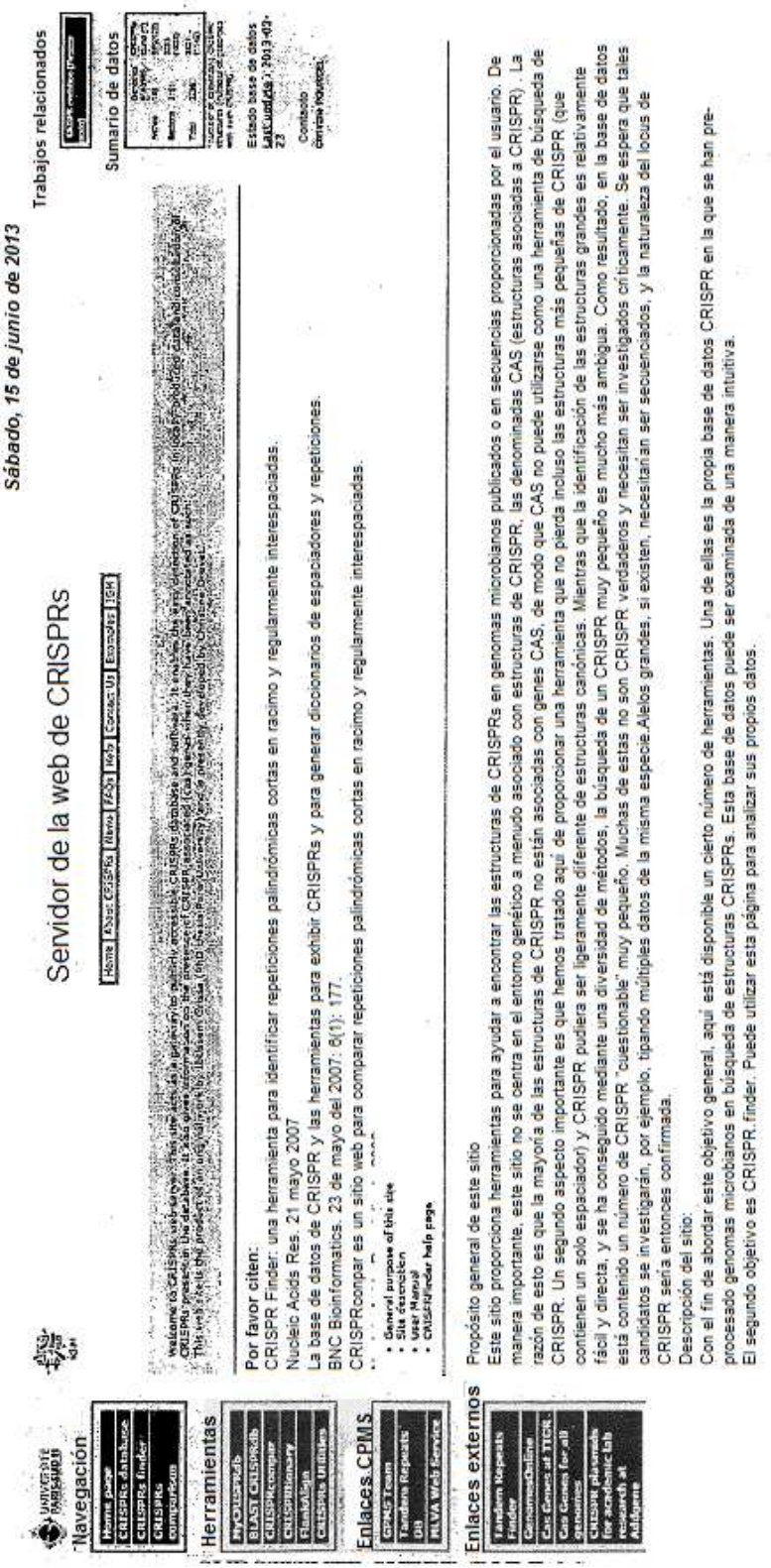

Sábado, 15 de junio de 2013
Trabajos relacionados
Servidor de la web de CRISPRs
Home | About CRISPRs | News | FAQs | Help | Contact Us | Examples | TBM
Sumario de datos
Navegación
Herramientas
Enlaces CPMS
Enlaces externos
Por favor citen:
CRISPR Finder: una herramienta para identificar repeticiones palindrómicas cortas en racimo y regularmente interespaciadas.
Nucleic Acids Res. 21 mayo 2007
La base de datos de CRISPR y las herramientas para exhibir CRISPRs y para generar diccionarios de espaciadores y repeticiones.
BNC Bioinformatics. 23 de mayo del 2007; 8(1): 177.
CRISPRcompar es un sitio web para comparar repeticiones palindrómicas cortas en racimo y regularmente interespaciadas.
Propósito general de este sitio
Este sitio proporciona herramientas para ayudar a encontrar las estructuras de CRISPRs en genomas microbianos publicados o en secuencias proporcionadas por el usuario. De manera importante, este sitio no se centra en el entorno genético a menudo asociado con estructuras de CRISPR, las denominadas CAS (estructuras asociadas a CRISPR). La razón de esto es que la mayoría de las estructuras de CRISPR no están asociadas con genes CAS, de modo que CAS no puede utilizarse como una herramienta de búsqueda de CRISPR. Un segundo aspecto importante es que hemos tratado aquí de proporcionar una herramienta que no pierda incluso las estructuras más pequeñas de CRISPR (que contienen un solo espaciador) y CRISPR pudiera ser ligeramente diferente de estructuras canónicas. Mientras que la identificación de las estructuras grandes es relativamente fácil y directa, y se ha conseguido mediante una diversidad de métodos, la búsqueda de un CRISPR muy pequeño es mucho más ambigua. Como resultado, en la base de datos está contenido un número de CRISPR 'cuestionable' muy pequeño. Muchas de estas no son CRISPR verdaderos y necesitan ser investigados críticamente. Se espera que tales candidatos se investigarán, por ejemplo, tipando múltiples datos de la misma especie.Alelos grandes, si existen, necesitarían ser secuenciados, y la naturaleza del locus de CRISPR sería entonces confirmada.
Descripción del sitio:
Con el fin de abordar este objetivo general, aquí está disponible un cierto número de herramientas. Una de ellas es la propia base de datos CRISPR en la que se han pre-procesado genomas microbianos en búsqueda de estructuras CRISPRs. Esta base de datos puede ser examinada de una manera intuitiva.
El segundo objetivo es CRISPR finder. Puede utilizar esta página para analizar sus propios datos.

FIG. 18

FIG. 19A

1 10 20 30 40 50 60
21 Staphylococcus aureus subsp. Aureus Cas9 translación ------MKRNYILGLDIGITSVGYGIIDYETRDVIDA----------GVRLFK
26 Streptococcus thermophilus CRISPR1 Cas9 translación --------MSDLVLGLDIGIGSVGVGILNKVTGEIIHK----------NSRIFP
4 Treponema denticola Cas9 (modificado) translación ---MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANR----------KDLMGMRCFE
5 Filifactor alocis Cas9 (modificado) translación ------MTKEYYIGLDVGTNSVGWAVTDSQYNLCKFKK----------KDMWGIRLFE
18 Neisseria cinerea Cas9 (modificado) translación MAAFKPNPMNYILGLDIGIASVGWAIVEIDEEENPIR----------LIDLGVRVFE
24 Campylobacter lari Cas9 (modificado) translación ----------MRILGFDIGINSIGWAFVENDE----------LKDCGVRIFT
Streptococcus pyogenes Cas9 translación ------MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
7 Lactobacillus johnsonii Cas9 (modificado) translación ---MTKIKDDYIVGLDIGTDSCGWVMMNSNNDILKLQG----------KTAIGSRLFE
23 Mycoplasma gallisepticum Cas9 (modificado) translación MKNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLG----------SRLFS
20 Parvibaculum lavamentivorans Cas9 (modificado) translación ---------MERIFGFDIGTTSIGFSVIDYSSTQSAGN----------IQRLGVRIFP
1 Corynebacter diptheriae Cas9 (modificado) translación ---------MKYHVGIDVGTFSVGLAAIEVDDAGMPIKT----------LSLVSHIHD
2 Sutterella wadsworthensis Cas9 (modificado) translación -MTQSERRFSCSIGIDMGAKYTGVFYALFDREELPTNLN----------SKAMTLVM
21 Staphylococcus aureus subsp. Aureus Cas9 translación EANVENNE--------GRRSKRGARRLKRRRRHRIQRVKKLLF----------
26 Streptococcus thermophilus CRISPR1 Cas9 translación AAQAENNL--------VRRTNRQGRRLARRKKHRRVRLNRLFE----------
4 Treponema denticola Cas9 (modificado) translación TAETAEV----------RRLHRGARRRIERKKKRIKLLQELFSQEIAKTDEGFFQRMKES
5 Filifactor alocis Cas9 (modificado) translación SANTAKD----------RRLQRGNRRRLERKKQRIDLLQEIFSPEICKIDPTFFIRLNES
18 Neisseria cinerea Cas9 (modificado) translación RAEVPKTGDSLAA---ARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFD------
24 Campylobacter lari Cas9 (modificado) translación KAENPKNKESLAL---PRRNARSSRRRLKRRKARLIAIKRILAKELKLNYKDYV------
Streptococcus pyogenes Cas9 translación SGETAEAT----------RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
7 Lactobacillus johnsonii Cas9 (modificado) translación GGKSAAER----------RLFRTTHRRIKRRRWRLKLLEEFFDPYMAEVDPYFFARLKES
23 Mycoplasma gallisepticum Cas9 (modificado) translación QAKTAEDR----------RSFRGVRRLIRRRKYKLKRFVNLIWKYNSYFGFRNK------
20 Parvibaculum lavamentivorans Cas9 (modificado) translación EARDPDGTP-LNQ---QRRQKRWWRQLRRRRIRRKALNETLHEAG----------
1 Corynebacter diptheriae Cas9 (modificado) translación SGLDPDEIKSAVTRLASSGIARTRRLVFRKKRRRLQQLDKFIQR----------
2 Sutterella wadsworthensis Cas9 (modificado) translación PETGPRYVQAQRT--AVRHRLRGQKRYTLARKLAFLVVDDMIKKQEKRLITDEEWKRGREA
21 Staphylococcus aureus subsp. Aureus Cas9 translación ----------DYNLLITDHSELSG-INPYEARVKGLSQKLSEEE-----
26 Streptococcus thermophilus CRISPR1 Cas9 translación ----------ESGLITDFTKISINLNPYQLRVKGLTDELSNEE-----
4 Treponema denticola Cas9 (modificado) translación PFYAEDKTIL----------QENTLFNDKDFADKTYHKAYPTINHLIKAWIENKVKPDP

FIG. 19B

```
 5 Filifactor alocis Cas9 (modificado) translación RLHLEDKSND----------FKYPLFIEKDYSDIEYYKEFPTIPHLRKHLISSEKQDI
10 Neisseria cinerea Cas9 (modificado) translación ----------ENG-----LIKSLPNTPWQLRAAALDRKLTPLE-----
24 Campylobacter lari Cas9 (modificado) translación ----------AADGELPKAYEGSLASVYSLRYKALTQNLETKD-----
          Streptococcus pyogenes Cas9 translación FLVEEDKK----------HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
 7 Lactobacillus johnsonii Cas9 (modificado) translación GLSPLDKRK----------TVSSIVFPTSAEDKKFYDDYPTIYHLRYKLMTEDEKFDL
23 Mycoplasma gallisepticum Cas9 (modificado) translación ----------EDILNNYQEQQKLHNTVLMLKSEALNAKIDPKA-----
20 Parvibaculum lavamentivorans Cas9 (modificado) translación ----------FLPAYGSADWPVVMADEPYSLRRRGLEEGLSAYE-----
 1 Corynebacter diptheriae Cas9 (modificado) translación ----------QGNFPIELEDYSDPLYPWKVRASLAASYIADEKERGE-
 2 Sutterella wadsworthensis Cas9 (modificado) translación LSGLLKRRGYSRPNADGEDLTPLENVRADVFAAMPAFSTYFSEVRSLAEQWEEPTANISN

21 Staphylococcus aureus subsp. Aureus Cas9 translación --FSAALLHLAKRRGVHNVN----------EVEEDTGNELSTKEQIS
26 Streptococcus thermophilus CRISPR1 Cas9 translación --LFIALKMMVKHRGISYLD----------DASDDGNSSVGDYAQIV
 4 Treponema denticola Cas9 (modificado) translación RLLYLACHNIIKKRGHFLFEGDFDSEN----QFDVSIQAL-FEYLSRDMEVDIDADSQK-
 5 Filifactor alocis Cas9 (modificado) translación RLIYLALHNIIKTRGHFLIDGDLQSAK----QLSPILDTF-LISLQEEQNLSVSLSENQK
18 Neisseria cinerea Cas9 (modificado) translación --WSAVLLHLIKHRGYLSQR----------KNEGETADKELGALLKGV
24 Campylobacter lari Cas9 (modificado) translación --LARVILHIAKHRGYMN----------KNEKKSNDAKKGKILSAL
          Streptococcus pyogenes Cas9 translación RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT-YNQLFEENPINASGVDAKA
 7 Lactobacillus johnsonii Cas9 (modificado) translación RSVYLAIHHIIKYRGNFLYNTSVKGDFKASKIDVKSSITEKLNELYENLGLDLNVEPNISNT
23 Mycoplasma gallisepticum Cas9 (modificado) translación --LSNILHDYLKNRGHFYEDN----------RDFNVYFTKELAKYFDKY
20 Parvibaculum lavamentivorans Cas9 (modificado) translación --FGRAIVHLAQHRHFKGRE----------LEESDTFDPDVDDEKHAA
 1 Corynebacter diptheriae Cas9 (modificado) translación -KLSVALRHIARHRGWRNPYAKVSS----------LYLPDGPSDAFKAIRESI
 2 Sutterella wadsworthensis Cas9 (modificado) translación VEKPLGDPNIPADKEFIEFAVASG----------LIDKTEKKAYQSALSTL

21 Staphylococcus aureus subsp. Aureus Cas9 translación R-NSKALEEKYVA----------
26 Streptococcus thermophilus CRISPR1 Cas9 translación XENSKQLETKTPG----------
 4 Treponema denticola Cas9 (modificado) translación --VKEILKDSSLKKNSEKQSRLNKILGLK--PSDKQKKAITNLIS------GNKINFADL
 5 Filifactor alocis Cas9 (modificado) translación DEYEEILKNRSIAKSEKVKKLKNLPEISDELEKEEKKAQSAVIENFCKFTVGNKGDVCKF
18 Neisseria cinerea Cas9 (modificado) translación ADYTHALQTGD----------
24 Campylobacter lari Cas9 (modificado) translación KNNALKLENYQSVG----------
          Streptococcus pyogenes Cas9 translación ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN----------
 7 Lactobacillus johnsonii Cas9 (modificado) translación AEIEKVLKDKQIFKRDKVKKTAELFAIKTDNKSQSKRIKDISKQVANAVLGYKTKFDTIA
23 Mycoplasma gallisepticum Cas9 (modificado) translación GYYKGIDSKEDMD----------
```

FIG. 19C

FIG. 19D

```
 4 Treponema denticola Cas9 (modificado) translación  CKTKSKKLIINNSVNQEDFYKFLKTILSAKSRIKEVNDILT----EIETGTFLPKQISKS
 5 Filifactor alocis Cas9 (modificado) translación  LKKNNKKYWIEKKRNPEEFYKSIGKLLDKIEPLKEDLEVLTMMIESCKNHTLLPIQKNKD
18 Neisseria cinerea Cas9 (modificado) translación  ----------------------------------------------------NQRGDY
24 Campylobacter lari Cas9 (modificado) translación  ----------------------------------------------------NTKDNY
                Streptococcus pyogenes Cas9 translación  QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN-------REDLLRKQRTFD
 7 Lactobacillus johnsonii Cas9 (modificado) translación  LLQAKKKLGKIKPRSKEDFYKVVNKMLDGSAASKEIKKKIELDS------FMPKQRTNA
23 Mycoplasma gallisepticum Cas9 (modificado) translación  ----------------------------------------------LPEKPKEEY
20 Parvibaculum lavamentivorans Cas9 (modificado) translación  ----------------------------------------------------
 1 Corynebacter diptheriae Cas9 (modificado) translación  ----------------------------------------------------
 2 Sutterella wadsworthensis Cas9 (modificado) translación  ----------------------------------------------NLSNLQLRAERWYFNAP

21 Staphylococcus aureus subsp. Aureus Cas9 translación  INRFKTSDYVKEAKQLLKVQKAYH-QLDQS---------FIDTYIDLLETRRTVYEGPGE
26 Streptococcus thermophilus CRISPR1 Cas9 translación  INVFPTSAYRSEALRIIQTQQEFNPQITDE---------FINRYLEILTGKRKYYNGPGN
 4 Treponema denticola Cas9 (modificado) translación  NAEIPYQLRKMELEKILSNAEKHFSFLKQKD---EKGLSHSEKIIMLLTFKIPYYIGPIN
 5 Filifactor alocis Cas9 (modificado) translación  NGVIPHQVHEVELKKILENAKKYYSFLTEID---KDGYSVVQKIESIFRFRIPYYVGPLS
18 Neisseria cinerea Cas9 (modificado) translación  SHTPWNKDLQAELNLLFEKQKEFGNPHVSD---------GLKEGIETLLMTQRPAL----
24 Campylobacter lari Cas9 (modificado) translación  NNCVLSSDLEKELKLILEKQKEFGYNYSED---------FINEILKVAPFQR--------
                Streptococcus pyogenes Cas9 translación  NGSIPHQIHIGELHAILRRQEDFYPFLKDN---------REKIEKILTFRIPYYVGPLA
 7 Lactobacillus johnsonii Cas9 (modificado) translación  NGVIPYQLQQLELDKIIENQSKYYPFLKEINPVSSHLKEAPYKLDELIRFRVPYYVGPLI
23 Mycoplasma gallisepticum Cas9 (modificado) translación  ESLFSYVRNYSEGPGSINSVSPYGIYHLDEK------EGKVVQKVNNTWDKTIGKCNIFPD
20 Parvibaculum lavamentivorans Cas9 (modificado) translación  GIHARRNVVAEEFRRLMEVQSKFHPALKSEE---------MRARISQTIPAQRPVFNR---
 1 Corynebacter diptheriae Cas9 (modificado) translación  GGVLSARLQQSDYAREIQEICRMQEIGQELY----------RKIIDVVFAAES-------
 2 Sutterella wadsworthensis Cas9 (modificado) translación  DIMKDRGWEPDRPKKTLVRAFKFFHPAKDQNKQHLELIKQIENSEDIIETLCTLDPNRTI

21 Staphylococcus aureus subsp. Aureus Cas9 translación  ---GSPFG----------------------------------------WKDIKEWYEMLMGHCTYFPEE
26 Streptococcus thermophilus CRISPR1 Cas9 translación  EKSRTDYGR----------------------------------------YRTSGETLDNIFGILIGKCTFYPDE
 4 Treponema denticola Cas9 (modificado) translación  DNHKKFFPDRCW----VVKKEKSPSGKTTPWNFPDHIDKEKTAEAFTTSKTNFCTYLVGE
 5 Filifactor alocis Cas9 (modificado) translación  TRHQEXGSN-VW----MVRKP-GREDRIYFWNMEEIIDFEKSNENFITRMTNKCTYLIGE
18 Neisseria cinerea Cas9 (modificado) translación  ----------------------------------------SGDAVQKMLGHCTFEPTE
24 Campylobacter lari Cas9 (modificado) translación  ----------------------------------------PLKDFSHLVGACTFPEEE
                Streptococcus pyogenes Cas9 translación  RGNSRFAMMTR----------KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
 7 Lactobacillus johnsonii Cas9 (modificado) translación  SPNESTKDIQTKKQNFAWMIRKEEGRITPWNFDQKVDRIESANKFIKRMTTKDTYLPGE
```

FIG. 19E

23 Mycoplasma gallisepticum Cas9	(modificado)	translación	EYRAPKNSP----------i-IANIFNEINELSTIRSYSIYLTGW
20 Parvibaculum lavamentivorans Cas9	(modificado)	translación	----------KNTLGRCRFMPGS
1 Corynebacter diptheriae Cas9	(modificado)	translación	----------PKGSASSRVGKDPLQPGK
2 Sutterella wadsworthensis Cas9	(modificado)	translación	PPYEDQNNRRPP-----LDQTLLLSPEKLTRQYGEIWKTWSARLTSAEPTLAPAAEILER
21 Staphylococcus aureus subsp. Aureus Cas9		translación	LRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKK---PTLKQIA
26 Streptococcus thermophilus CRISPR1 Cas9		translación	FRAAKASYTAQEFNLLNDLNNLTVP-TETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIA
4 Treponema denticola Cas9	(modificado)	translación	SVLPKSSLLYSEYTVLNEINNLQIIIDGKNICDIKLKQKIYEDLFKKYKK---TTQKQIS
5 Filifactor alocis Cas9	(modificado)	translación	DVLPKHSLLYSKYMVLNELN--NVKVRGKKLP-TSLKQKVFEDLFENKSK---VTGKNLL
18 Neisseria cinerea Cas9	(modificado)	translación	PKAAKNTYTAERFVWLTKLM--NLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARK
24 Campylobacter lari Cas9	(modificado)	translación	KRACKNSYSAWEFVALTKIN-EIKSLEKISGEIVPTQTTNEVLNLILDKGSITYKKFRS
Streptococcus pyogenes Cas9		translación	KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
7 Lactobacillus johnsonii Cas9	(modificado)	translación	DVLPANSLLYQKFTVLNELNN---IRINGKRISVDLKQEIYENLFKKHTT----VTVKKL
23 Mycoplasma gallisepticum Cas9	(modificado)	translación	FINQEFKKAYLNKLLDLIKTNGEKPIDARQFNKLREETIAESIGKETLKDVENEEKLEK
20 Parvibaculum lavamentivorans Cas9	(modificado)	translación	PLCPKGSWLSQQRRMLEKLNNLAIAGGNARPLDAEERDAILSKLQQQASNSWPGVRSALK
1 Corynebacter diptheriae Cas9	(modificado)	translación	MRALKASDAPQRYRIAALIGNLERVRVDGEKRILSVEEKNLVFDHLVNLTPKEPEWVTIA
2 Sutterella wadsworthensis Cas9	(modificado)	translación	STDRKSRVAVNGHEPLFTLAYQLSYALQRAFDRSKALDPYALRALAAGSKSNKLTSARTA
21 Staphylococcus aureus subsp. Aureus Cas9		translación	KEILVNEEDIKG---YRVT--STGKPEFTNLKVYHDIKD--ITA--RKEIIENAELLDQI
26 Streptococcus thermophilus CRISPR1 Cas9		translación	KLLSCDVADIKG---YRID--KSGKAEIHTFEAYRKMKT--LET--LDIEQMDRETLDKL
4 Treponema denticola Cas9	(modificado)	translación	TFIKHEGICNKT---DEVIILGIDKECTSSLKSYIELKN-IFGK--QVDEISTKNMLEEI
5 Filifactor alocis Cas9	(modificado)	translación	EYLQIQDKDIQI---DDLS--GFDKDPKTSLKSYLDFKKQIFGE--EIEKESIQNMIEDI
18 Neisseria cinerea Cas9	(modificado)	translación	LLDLDDTAFFKG---LRYGKDNAEASTLMEMKAYHAISR-ALEK--EGLKDKKSPLNLSP
24 Campylobacter lari Cas9	(modificado)	translación	CINLNESISFKS---LKYDKENAENAKLIDFKKLVEFKX-ALG----------VHSLSR
Streptococcus pyogenes Cas9		translación	EDYFKKIECFDS---VEIS--GVEDRFNASLGTYHDLLKIIKDKD-FLDNEENEDILEDI
7 Lactobacillus johnsonii Cas9	(modificado)	translación	ENYLKENNNLVK---VEIKGLADEKKFNSGLGTYNRFKNLMIFDN-QIDDLKYRNDFEKI
23 Mycoplasma gallisepticum Cas9	(modificado)	translación	EDNEWKLKGLKLNTWGKIQYNDLSSLAKFVIKLKQHLKLDFLLED-QYATLDKTNFLQSL
20 Parvibaculum lavamentivorans Cas9	(modificado)	translación	ALYKQRGEPGAEKS-LKFNLELGGSESKLLGNALEAKLADMPGPDWPAHPRKQEIRHAVHE
1 Corynebacter diptheriae Cas9	(modificado)	translación	EILGIDRGQLIG------TATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASAL
2 Sutterella wadsworthensis Cas9	(modificado)	translación	LENCIGGQNVKT--FLDCARYYREADDAKVGLWFDNADGLLERSDLHPPMKKKILPLLV
21 Staphylococcus aureus subsp. Aureus Cas9		translación	AKILTTYQ--SSEDIQEELTN--LNSELTQEEIEQISNLKGYT----------

FIG. 19F

```
25 Streptococcus thermophilus CRISPR1 Cas9   translación   AYVLTLNT--ERBGIQEALEHEFADGSPSQKQVDELVQFRKANS----------
   4 Treponema denticola Cas9   (modificado)   translación   IRWATIYDEGEGKTILKTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRK----------
   5 Filifactor alocis Cas9   (modificado)   translación   IKWITIY--GNDKEMLKRVIRANYSNQLTEEQMKKITGFQYSGWGNFSKM----------
  18 Neisseria cinerea Cas9   (modificado)   translación   ELQQQIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKP----------
  24 Campylobacter lari Cas9   (modificado)   translación   QELDQISTHTILIKDNVKLKTVLEKIWLSNEQINNLLEIEFNDY----------
Streptococcus pyogenes Cas9   translación   VLTLTLFE----DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK----------
   7 Lactobacillus johnsonii Cas9   (modificado)   translación   IENSTIFE-----DKSIYKEKLRSIDWLNEKQINALSNTRLQGWG----------
  23 Mycoplasma gallisepticum Cas9   (modificado)   translación   FVYLGKHLAYSNRVDSANLKEFSDSNKLFERILQKQKDGLFKLFEQTDKDD----------
  20 Parvibaculum lavamentivorans Cas9   (modificado)   translación   RLWAADYGETPDKKRVIILSEKDRKAHREAAANSFVADFGITVGEQAAQLQA----------
   1 Corynebacter diptheriae Cas9   (modificado)   translación   EQHAMVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLP----------
   2 Sutterella wadsworthensis Cas9   (modificado)   translación   ANILQTDETTGQKFLDEIWRKQIKGRETVASRCARIETVRKSPGGGFNIAYNTAQYREVN

  21 Staphylococcus aureus subsp. Aureus Cas9   translación   ----------GTHNLSLKAINLILDELWHTND----------NQIAIFN--------
  26 Streptococcus thermophilus CRISPR1 Cas9   translación   --------SIFGKGWHNFSVKLNMELIPELYETSE----------EQMTILT--------
   4 Treponema denticola Cas9   (modificado)   translación   -------FLETVTSEMPGFSEPVNIITAMRETQN----------NLMELLSSEFTFTS
   5 Filifactor alocis Cas9   (modificado)   translación   -------PLKGISGSDVSIGETFDIITAMWETDN----------NLMQILSKKFTFMD
  18 Neisseria cinerea Cas9   (modificado)   translación   ----------VQISLKALRRIVPLMEQGNRYDE----------ACTEIYG--------
  24 Campylobacter lari Cas9   (modificado)   translación   ----------INLSFKALGMILPLMREGKKYDE----------ACEIANLK------
Streptococcus pyogenes Cas9   translación   -------LINGIRDKQSGKTILDFLKSDGFANRN----------FMQLIHDDS----L
   7 Lactobacillus johnsonii Cas9   (modificado)   translación   --------RLSKKLLAQLHDEMGQTIRQLMDSQN----------NPMQIVTQAD----
  23 Mycoplasma gallisepticum Cas9   (modificado)   translación   --------EKILAQTHSLSTKAMLLAITRMTNLDNDEDMQKWNNDKGWNFEAIKNFDQKFID
  20 Parvibaculum lavamentivorans Cas9   (modificado)   translación   -------LKLPTGWEPYSIPALNLFLAELEKGERFG-----ALVNGPDWEGWRKTNFFHR
   1 Corynebacter diptheriae Cas9   (modificado)   translación   ----------VGRAAVSEDTLVRLTRRMLSDGVD----------
   2 Sutterella wadsworthensis Cas9   (modificado)   translación   KLPRNAQDKELLTIRDRVAETADFIAANLGLSDEQKRKFANPFSLAQFYTLIETEVSGFS

  21 Staphylococcus aureus subsp. Aureus Cas9   translación   RLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG---LPNDII
  26 Streptococcus thermophilus CRISPR1 Cas9   translación   RLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYG---DFDNIVI
   4 Treponema denticola Cas9   (modificado)   translación   NIKKINSGFEDAEKQFSYDGLVKPLFLSPSVKKMLMQTLKLVKEISHITQA--PPKKIFI
   5 Filifactor alocis Cas9   (modificado)   translación   NVEDFNSGKVGKIDKITYDSTVKEMFLSPEMKRAVWQTIQVAEEIKKVMGC--EPKKIFI
  18 Neisseria cinerea Cas9   (modificado)   translación   ----DHYGKKNTEEKIYLFPIPADEIRNPVVLRALSQAKKVINGVVRRYG---SPARIHI
  24 Campylobacter lari Cas9   (modificado)   translación   ---PKTVDEKKDEFLPAFCDSIFAHELSNFVVNRAISEYRKVLNALLKKYG---KVHKIHL
Streptococcus pyogenes Cas9   translación   TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH-KPENIVI
```

FIG. 19G

7 Lactobacillus johnsonii Cas9 (modificado) translación FKDAIAKANQNLLVATSVEDILNNAYTSPANKKAIRQVIKVVDDIVKAASGK-VPKQIAI
23 Mycoplasma gallisepticum Cas9 (modificado) translación ITKKNNNLSLKQNKRYLDDRFINDAILSPGVKRILREATKVFNAILKQPSEEYDVTKVVI
20 Parvibaculum lavamentivorans Cas9 (modificado) translación NQPTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVVNNLIGLYG---KPDRIRI
1 Corynebacter diptheriae Cas9 (modificado) translación LYTARLQEPGIEPSWTPPTPRIGEPVGNPAVDRVLKTVSRNLESATKTWG---APERVII
2 Sutterella wadsworthensis Cas9 (modificado) translación ATTLAVHLENAWRMTIKDAVINGETVRAAQCSRLPAETARPFDGLVRRLVDR-QANEIAK
21 Staphylococcus aureus subsp. Aureus Cas9 translación ELAR----EKNSKD-AQKMINEMQKRNRQTNERIEEIIRTTGKEN----------AKY
26 Streptococcus thermophilus CRISPR1 Cas9 translación EMAR----ETNEDD-EKKAIQKIQKANK--DEKDAAMLKAANQYNGKAELPHSVFHGHKQ
4 Treponema denticola Cas9 (modificado) translación EMAKGA--ELEPAR-TKTRLKILQDLYNNCKNDADAFSSEIKDLSGKIENE----DNLRL
5 Filifactor alocis Cas9 (modificado) translación EMARGG--EKVKKR-TKSRKAQLLELYAACEEDCRELIKEIEDR----------DERDF
18 Neisseria cinerea Cas9 (modificado) translación ETAREV--GKSFKD-RKEIEKRQEENRKDREKSAAXFREYFPNFVG---------EPKSK
24 Campylobacter lari Cas9 (modificado) translación ELARDV--GLSKKA-REKIEKEQKEN----QAVNAWALKECENIGL---------KASAK
Streptococcus pyogenes Cas9 translación EMAREN--QTTQKG-QKNSRERMKRIEEGIKELGSQILKEHPVEN----------TQL
7 Lactobacillus johnsonii Cas9 (modificado) translación EFTRDA--DENPKR-SQTRGSKLQKVYKDLSTELASKTIAEELNEAIK--------DKKL
23 Mycoplasma gallisepticum Cas9 (modificado) translación ELARELSEEKELEN-TKNYKKLIKKNGDKISEGLKALGISEDEIKD--------ILKSPT
20 Parvibaculum lavamentivorans Cas9 (modificado) translación EVGRDVG--KSKRE-RBEIQSGIRRNEKQRKKATEDLIKNGIANPS----------RD
1 Corynebacter diptheriae Cas9 (modificado) translación EHVREG--FVTEKR-AREMDGDMRRRAARNAKLFQENQEKLNVQGK---------PSRA
2 Sutterella wadsworthensis Cas9 (modificado) translación RVSTDIQSKVDFSNGIVDVSIPVEENKFEPSASVADLKKNKRVKDKMLS--EAEKLETRW
21 Staphylococcus aureus subsp. Aureus Cas9 translación LIEKIXLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNS----FNNKVLV
26 Streptococcus thermophilus CRISPR1 Cas9 translación LATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDS----LANKVLV
4 Treponema denticola Cas9 (modificado) translación RSDKLYLYYTQLGKCMYCGKPIEIGHVFD-TSNYDIDHIYPQSKIKDDS----ISNRVLV
5 Filifactor alocis Cas9 (modificado) translación NSMKLFLYYTQFGKCMYSGDDIDINELIRGNSKWDRDHIYPQSKIKDDS----IDNLVLV
18 Neisseria cinerea Cas9 (modificado) translación DILKLRLYEQQHGKCLYSGKEINLGRLNE-KGYVEIDHALPFSRTWDDS----FNNKVLA
24 Campylobacter lari Cas9 (modificado) translación NILKLKLWKEQKEICIYSGNKISIEHLKD-EKALEVDHIYPYSRSFDDS----FINKVLV
Streptococcus pyogenes Cas9 translación QNEKLYLYYLQNGRDMYVDQELDINRLSD----YDVDHIVPQSFLKDDS----IDNKVLT
7 Lactobacillus johnsonii Cas9 (modificado) translación VQDKYYLYFMQLGRDAYTGEPINIDEIQX----YDIDHILPQSFIKDDA----LDNRVLV
23 Mycoplasma gallisepticum Cas9 (modificado) translación KSYKFLLWLQQDHIDPYSLKEIAFDDIFTKTEKFEIDHIIPYSISFDDS----SSNKLLV
20 Parvibaculum lavamentivorans Cas9 (modificado) translación DVEKWILWKEGGERCPYTGDQIGFNALFR-EGRYEVEHIWPRSRSFDNS----PRNKTLC
1 Corynebacter diptheriae Cas9 (modificado) translación DLWRYQSVQRQNCQCAYCGSPITFS-------NSEMDHIVPRAGQGSTN----TRENLVA
2 Sutterella wadsworthensis Cas9 (modificado) translación LIKNERIKKASRGTCPYTGDRLAEGG--------EIDHILPRSLIKDARGIVFNAEPNLI

FIG. 19H

```
21 Staphylococcus aureus subsp. Aureus Cas9   translación   KQEEMSKKGNRTPFQYLSSSDGKISYETFKKHIINLA-KGKGRISKTKKEYLLEEMDINR
26 Streptococcus thermophilus CRISPR1 Cas9    translación   YATANQEKQQRTPYQALDSMDDAWSFRELKAFVR-----ESKTLSNKKKEYLLTEEDISK
     4 Treponema denticola Cas9 (modificado)  translación   CSSCNKNKSDKYPLKSEIQSKQRQPWNFLQ---------HNNFISLEKIANRLTRATPISD
        5 Filifactor alocis Cas9 (modificado) translación   NKTYNAKKSNE-LLSEDIQKKNHSFNLSLL---------NKKLITKSKYDRLTRKGDFTD
      18 Neisseria cinerea Cas9 (modificado)  translación   LGSENQNKGNQTPYEYFNGKDNSHENQSEFKAR------VETSRFPRSKKQRILLQK--PD
      24 Campylobacter lari Cas9 (modificado) translación   PTKENQEKLNKTPPEAPG--KNIEKMSKIQTL------AQM--LPYKKKNKILDEN--FK
                Streptococcus pyogenes Cas9   translación   RSDKNRGKSDNVPSEEVVKKMKNYWRQLLN---------AKLITQRKFDNLTKAERGGL
     7 Lactobacillus johnsonii Cas9 (modificado) translación   SRAVNNGKSDNVPVKLPGNEMAANLGMTIRKMNEEN--KNIGLISKTKYNNLLTDPDNIN
   23 Mycoplasma gallisepticum Cas9 (modificado) translación   LAESNQAKSNQTPYEFISSGNAGIKMEDYEAYCRKFNDGDSSLLDSTQRSKKFAKMMKTD
20 Parvibaculum lavamentivorans Cas9 (modificado) translación   RKDVNIEKGNRMPFRAFGHDEDRWSAIQIR---------LQGMVSAKGGTGMSPGKVKRF
     1 Corynebacter diptheriae Cas9 (modificado) translación   VCHRCNQSKGNTPFAIMKKVTSIEGVSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVERF
     2 Sutterella wadsworthensis Cas9 (modificado) translación   YASSRGMQLKKNQRYSLSDLKANYRMEIFKTSNIAAITAEIEDVVTKLQQTNRLKFFDLL

21 Staphylococcus aureus subsp. Aureus Cas9   translación   FSVQKD--FINRNLVDTRYATRGLNNLLRSYFRVNNLD----------------------
26 Streptococcus thermophilus CRISPR1 Cas9    translación   FDVRKK--FIERNLVDTRYASRVVLNALQEHFRANKID----------------------
     4 Treponema denticola Cas9 (modificado)  translación   DETAK---FIARQLVETRQATKYAAKVLEKMFPETKIV----------------------
        5 Filifactor alocis Cas9 (modificado) translación   EELSG---FIARQLVETRQSTKAIADIFKQIY-SSEVV----------------------
      18 Neisseria cinerea Cas9 (modificado)  translación   E---DG--FKERNLNDTRYINRFLCQFVADHMLLTGKG----------------------
     24 Campylobacter lari Cas9 (modificado)  translación   DKQQED--FISRNLNDTRYIATLIAKYTKEYLNFLLLSENENANL----------------
                Streptococcus pyogenes Cas9   translación   SELDKAG-FIKRQLVETRQITKHVAQILDSRMNTKYDENDK----------------------
     7 Lactobacillus johnsonii Cas9 (modificado) translación   KYKSAG--FIRRQLVETSQIIKLVSTILQSRYPNTEII----------------------
   23 Mycoplasma gallisepticum Cas9 (modificado) translación   TSSKYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKP----------------------
20 Parvibaculum lavamentivorans Cas9 (modificado) translación   LAKTWPEDFAARQLNDTRYAAKQILAQLKRLWPDMGFEAP----------------------
     1 Corynebacter diptheriae Cas9 (modificado) translación   QRATWDEEIDARSMKSVAWMANELRSRVAQHFASKGTT----------------------
     2 Sutterella wadsworthensis Cas9 (modificado) translación   NEHEQDCVSHALFLDDGSEARDAVLKLLATQRRTRVNGTQIWNIKNLANKIPEELQMNCK

21 Staphylococcus aureus subsp. Aureus Cas9   translación   ----------VKVKSINGGFTSFLRR---KWKFKK------SRNKGYKHHAEDALIIANAD
26 Streptococcus thermophilus CRISPR1 Cas9    translación   ----------TKVSVVRGQFTSQLRR---HWGIEK------TRDI-YHHHAVDALIIAASS
     4 Treponema denticola Cas9 (modificado)  translación   ----------YSKATVSMPRNK--------PDIVK------CRRINDFHHAHDAYLNIVVG
        5 Filifactor alocis Cas9 (modificado) translación   ----------YVKSSLVSDPRKKP------LNYLK------SRRVNDYHHAKDAYLNIVVG
      18 Neisseria cinerea Cas9 (modificado)  translación   ---------KRRVFASNGQITNLLRG---FWGLRK------VRAENDRHHALDAVVVACST
     24 Campylobacter lari Cas9 (modificado)  translación   -KSGEKGSKIHVQTISGMLTSVLRH---TWGFDK------KDRNNHLHHALDAIVVAYST
```

FIG. 19I

```
                Streptococcus pyogenes Cas9  translación  -----LIREVKVITLKSKLVSDFRK---DFQFYK------VREINNYHHAHDAYLNAVVG
                 7 Lactobacillus johnsonii Cas9 (modificado) translación  ----------TVKAKYNHHYLRE---KFDLYK------SREVNDYHHAIDAYLSAICG
               23 Mycoplasma gallisepticum Cas9 (modificado) translación  ----------MFKVVCINGSVTSFLRKNFDDSSYAK------KDRDKNIHHAVDASIISIFS
            20 Parvibaculum lavamentivorans Cas9 (modificado) translación  ----------VKVEAVTGQVTAQLRKLWTLNNILADDG--EKTRACHRHHAIDALTVACTH
                 1 Corynebacter diptheriae Cas9 (modificado) translación  ----------VRVVSGSLTAEARRASGISGKLKFFDGVGKSRLDRRHHAIDAAVIAFTS
                 2 Sutterella wadsworthensis Cas9 (modificado) translación  TYNNRLHTQAAATNVSDAKNLRLKLAQNPDDFEK------PDIQPIASHSIDALCSPAVG

                21 Staphylococcus aureus subsp. Aureus Cas9 translación  FIFKENKKIDKAKKVN----------------------------ENQMFEEKQAESMPEIET-EQEYKE
                26 Streptococcus thermophilus CRISPR1 Cas9 translación  QLNLWKKQKNTLVSYS----------------------------EDQLLDIETGSLISDDEYKESVFKA
                   4 Treponema denticola Cas9 (modificado) translación  NVYNTKFTNNPWNFIK----------------------------EKRDNPKIADTYNYYKVFDYDVKRN
                   5 Filifactor alocis Cas9 (modificado) translación  NVYNKKFTSNPIQWMK----------------------------KNRD-----TNYSLNKVFEHDVVIN
                  18 Neisseria cinerea Cas9 (modificado) translación  IANQQKIT-REVRYKE----------------------------MNAFDGKTIDKETGRVLHQKAHFPQ
                  24 Campylobacter lari Cas9 (modificado) translación  NSIIKAFS-DFRKNQE----------------------------LLKAR----------------
                        Streptococcus pyogenes Cas9  translación  7ALIKKYPKLESEFVYGDYK-----VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
                 7 Lactobacillus johnsonii Cas9 (modificado) translación  NLLYQNYPNLRPFFVYG----------------------------QYKKFSSDPDKEKAIFNKTRKFSFI
               23 Mycoplasma gallisepticum Cas9 (modificado) translación  METKTLFNQLTQFADYKLFK-----NTDGSWKKIDPKTGVVTEVTDENMKQIRVRNQVSE
            20 Parvibaculum lavamentivorans Cas9 (modificado) translación  PGNTNKLSRYHQLRDD----------------------------------------------
                 1 Corynebacter diptheriae Cas9 (modificado) translación  DYVASTLAVRSNLKQSQ----------------------------AHRQEAFQWREFTGKDAEHRAAWRV
                 2 Sutterella wadsworthensis Cas9 (modificado) translación  SADAERDQGWGFDYILDGKTVLGLYPQSCSVIHLQRKPQEENSHFDSVAIFKEGIYAEQFLP

                21 Staphylococcus aureus subsp. Aureus Cas9 translación  IFITPHQIKHIKDFKD-YKYSHRVDKKPNRELINDTLYSTR-----KDDKGNTLIVNNLN
                26 Streptococcus thermophilus CRISPR1 Cas9 translación  PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIK
                   4 Treponema denticola Cas9 (modificado) translación  NITAWKR----------GKTIITVKDMLKRNTPIYTRQAACKGELFNQTIMKKGL
                   5 Filifactor alocis Cas9 (modificado) translación  GEVIWEKCTYHEDTNT---YDGGTLGRIRKIVSRDNILYTEVAYGRKGELFNATIQNKNG
                  18 Neisseria cinerea Cas9 (modificado) translación  PWEFFAQSVMIRVFGK--PDGKPEFEEADTFEKLRTLLAEKLSSRPEAVHKYVTPLFISR
                  24 Campylobacter lari Cas9 (modificado) translación  ---FYAKELTSDNY----------------------------KHQVKFFEPKSFREKILSKIDEIFVSK
                        Streptococcus pyogenes Cas9  translación  ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES
                 7 Lactobacillus johnsonii Cas9 (modificado) translación  SQLLKNKSENSKEIAK----------KLKRAYQFKYMVLVSRETETRDQMFKHTVIPRFS
               23 Mycoplasma gallisepticum Cas9 (modificado) translación  IAKVIEKYIQDSNIERKKAEYSRKIENKTNISLFNDTVYSAKKVGYEDQIKRKNLKTLDIH
            20 Parvibaculum lavamentivorans Cas9 (modificado) translación  ----------------------------PRAEKPALTEPMDTIRADAEKAVSEIVVSH
                 1 Corynebacter diptheriae Cas9 (modificado) translación  MCQKMEKLSALLTEDLR----------DDRVVMSMPVRLRGNGSAHKETTGKLSKVKLSS
                 2 Sutterella wadsworthensis Cas9 (modificado) translación  IFTLNEKIWIGYETLNAKGERCGAIEVSGKQPKELLEMLARFFNKPVGDLSAHATYRILK
```

FIG. 19J

21	Staphylococcus aureus subsp. Aureus Cas9	translación	GLYDKD-NDKLKKLINKSPEKLLMYHHDP----------QTYQK-LKLI
26	Streptococcus thermophilus CRISPR1 Cas9	translación	DIYTQDGYDAFMKIYKKDKSKFLMYRHDP----------QTFEKVIEPI
4	Treponema denticola Cas9	(modificado)	translación	GQNPLKKEGPPSNISKYGG-----YNKVS----------AAYYTLIEYE
5	Filifactor alocis Cas9	(modificado)	translación	NSTVSLKKG--LDVKKYGG-----YFSAN----------TSYFSLIEFE
18	Neisseria cinerea Cas9	(modificado)	translación	APNRKMSGQGHMETVKSAKRLDEGISVLR----------VPLTQLKLKD
24	Campylobacter lari Cas9	(modificado)	translación	PPRKARRALHKDTPHSENKIIDKCSYN----------VAYSVLVVAK
	Streptococcus pyogenes Cas9	translación	ILPKRNSDKLIARKKDWDPKKYGGFDSPT----------DAYMVIVRID
7	Lactobacillus johnsonii Cas9	(modificado)	translación	HDTVKAPRNLIPKKMGMSPDIYGGYTNNS----------ESAKENKNSKVKRQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANFWVINLAEQIFNE
23	Mycoplasma gallisepticum Cas9	(modificado)	translación	RVRKKVSGPLHKETTYGVDTGTDIKTKSGT----------
20	Parvibaculum lavamentivorans Cas9	(modificado)	translación	QLSVSDIDKASSEALNCALTREPGFDFKE----------
1	Corynebacter diptheriae Cas9	(modificado)	translación	KPAYEFLAKAALQPLSAEEKRLAALLDALRYCTSRKSLMSLFVAANGRSLKKREDVLKPK
2	Sutterella wadsworthensis Cas9	(modificado)	translación	
21	Staphylococcus aureus subsp. Aureus Cas9	translación	MEQYGD----EK------NPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDI
26	Streptococcus thermophilus CRISPR1 Cas9	translación	LENYPNKQINEKGKEVPCNPFLKYKEEHG-YIKKYSKKGNGPEIKSLKYYDSKLGNHIDI
4	Treponema denticola Cas9	(modificado)	translación	EK-GNKIRSLETIP----LYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKI
5	Filifactor alocis Cas9	(modificado)	translación	DKKGDRARHIIGVP----IYIANMLEHSPSAFLEYCE-QKGYQNVRILVEKIKKNSLLII
18	Neisseria cinerea Cas9	(modificado)	translación	LEKMVNRERLPKIYEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKT
24	Campylobacter lari Cas9	(modificado)	translación	----------SKEGLQIALSCGRVRKI
	Streptococcus pyogenes Cas9	translación	VEKGKSKK----------LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL
7	Lactobacillus johnsonii Cas9	(modificado)	translación	KKKGTEYKILG-------IPTRELVNLKKAEKEDHYKSYLKEILTPRILYNKNGKRDFKTI
23	Mycoplasma gallisepticum Cas9	(modificado)	translación	YTENKKIKSQNVPEKYNLDLTKEFPEKPSEFLVKSMLRNKTALIYDDKKNIVHRIKRLKM
20	Parvibaculum lavamentivorans Cas9	(modificado)	translación	----------YRQFVTRKKIESLSKGELDEIRDPRIKEI
1	Corynebacter diptheriae Cas9	(modificado)	translación	----------GLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAE
2	Sutterella wadsworthensis Cas9	(modificado)	translación	LFQLKVELKGEKSFKLNGSLTLPVKQDWLRICDSPELADAFGKPCSADELTSKLARIWKR
21	Staphylococcus aureus subsp. Aureus Cas9	translación	TDDYFNSRNKVVKLSLKPYRFDVYLDN--GVYKFVTVKNLDVIKKE----------
26	Streptococcus thermophilus CRISPR1 Cas9	translación	TPK--DSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGT----------
4	Treponema denticola Cas9	(modificado)	translación	NGFPCHITGKTNDSFLLRPAVQFCCSNNEVLYFKKIIRFSERSQRKIGKTISPYEDLS
5	Filifactor alocis Cas9	(modificado)	translación	NGYPLRIRGENEVDTSFKRAIQLKLDQKNYELVRNIEKFLEKYVEK----------
18	Neisseria cinerea Cas9	(modificado)	translación	GVWVHHNGIADNATIVRVDVFEKGGKYYLVPIYSWQVAKGILPDR----------

FIG. 19K

```
24 Campylobacter lari Cas9 (modificado) translación G------TKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILPNK----------
Streptococcus pyogenes Cas9 translación FKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK----------
7 Lactobacillus johnsonii Cas9 (modificado) translación TSFSIVKSNIPYKQVIQDGDKKFMLGSSTYVYNAKQLTLSTESMKA----------
23 Mycoplasma gallisepticum Cas9 (modificado) translación LSSELKENKLSNVIIRSKNQSGTKLSYQDTINSELAMIMRSIDPTAKKQYIRVPLN----
20 Parvibaculum lavamentivorans Cas9 (modificado) translación VAAHVAGRGGDPKKAFPPYPCVSPGGPEIRKVRLTSKQQLNLMAQTG----------
1 Corynebacter diptheriae Cas9 (modificado) translación LGSSFHHARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKP----------
2 Sutterella wadsworthensis Cas9 (modificado) translación PVMRDLAHAPVRREFSLPAIDNPSGGFRIRRTNLPGMELYQVHATNAKKYRGPASMG---

21 Staphylococcus aureus subsp. Aureus Cas9 translación -----NYYEVNSKCYEDAKKLKKISNQAEFIASFYNNDLIKING------ELYRVIG-VN
26 Streptococcus thermophilus CRISPR1 Cas9 translación -----GTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTN
4 Treponema denticola Cas9 (modificado) translación FRSYIKENLWKKTKNDRIGEKEFYDLLQKKNLEIYDMLLTKEKDTIYKKRPNSATIDILV
5 Filifactor alocis Cas9 (modificado) translación -----KGNYPIDENRDMIT--------HEKMNQLYEVLLSKMKK--FNKKGNADPSDRIE
18 Neisseria cinerea Cas9 (modificado) translación ------AVVQGKDEED----WTVMDDSFEFKFVLYANDLIKLTAK--KNEFLGYFVSLNR
24 Campylobacter lari Cas9 (modificado) translación -------IVITGKDKNNNNFKQWQTIDESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSI
Streptococcus pyogenes Cas9 translación ----------LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
7 Lactobacillus johnsonii Cas9 (modificado) translación ----------ITNNFDKDSDENDALIKAYDEILDKVDKYLPLFDIN--------KFREKLH
23 Mycoplasma gallisepticum Cas9 (modificado) translación -----TLNELHLGDHDEFDLHNMDAYLKKPKFVKYLKANEIGDEYKPWRVLTSGTLLTHKKD
20 Parvibaculum lavamentivorans Cas9 (modificado) translación -------NGYADLGSNHHIAIYRLPDGKADFEIVSLFDASRRLAQR---NPIVQRTRADGA
1 Corynebacter diptheriae Cas9 (modificado) translación --------QTMSNRQAEKKLRDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELG
2 Sutterella wadsworthensis Cas9 (modificado) translación ----SNVDMSKGILPNELQHENLTECGGRFITSADVTPMSEWRKVVAEENLSINIAPGTE

21 Staphylococcus aureus subsp. Aureus Cas9 translación NDLLNRIEVNMIDITYREYLENM----------NDKR-----PPRIIKTIASKTQSIKKYS
26 Streptococcus thermophilus CRISPR1 Cas9 translación PKQKHYVSLKPYDKQKFEGCEAL----------IKVLGNVANSGQCKKGLGKSNISIYKVR
4 Treponema denticola Cas9 (modificado) translación KGKEKFKSLI-IENQFEVILEIL----------KLFSATRN-VSDLQHIGGSKYSGVAKIG
5 Filifactor alocis Cas9 (modificado) translación KSKPKFIKLEDLIDKINVINKML----------NLLRCDNDTKADLSLIELPKNAGSFVVK
18 Neisseria cinerea Cas9 (modificado) translación ATGAIDIRTHD----------TDSTKGKNGIFQSVGVKTALSFQKYQ
24 Campylobacter lari Cas9 (modificado) translación STSSICVEKHDNKFENLTSNQKL----------LFSNNAKEGSVKVESLGIQNLKVFEKYI
Streptococcus pyogenes Cas9 translación KHRDKPIREQAENIIHLFTLTNL----------GAPAAFKYFDTTIDRK---RYTSTKEVL
7 Lactobacillus johnsonii Cas9 (modificado) translación SGREKFIKLSLEDKKDTILKVLE----------GLMDN--AVMTKIPTIGLSTPLGFMQPP
23 Mycoplasma gallisepticum Cas9 (modificado) translación KKLMYISSFQNLNDVIEIKNLIE----------TSYKENDDSDSKKKKANRFLMTLSTIL
20 Parvibaculum lavamentivorans Cas9 (modificado) translación SFVMSLAAGEAIMIFEQSKKGIW----------IVQGVMASGQVVLERDTDADHSTTTRPM
1 Corynebacter diptheriae Cas9 (modificado) translación TIRRMRVDGFFSPSKLRLRPLQM----------SKRGIKKESAPELSKIIDRGWLPAVNK
```

FIG. 19L

2	Sutterella wadsworthensis Cas9	(modificado)	translación	GRRYVRVETTFIQASHNPEQSVRNWAITSPLSILPASFKVDKPAEFQKAVGTELSELLGQP
21	Staphylococcus aureus subsp. Aureus Cas9		translación	TDILGNLYEVKSKKHPQIIKKG----------
26	Streptococcus thermophilus CRISPR1 Cas9		translación	TDVLGNQNIIKKNEGDKPKLDF----------
4	Treponema denticola Cas9	(modificado)	translación	NKISSLDNCIIIIYQSITGIFEKRIDLLKV----------
5	Filifactor alocis Cas9	(modificado)	translación	KNTIGKSKIILVNQSVTGLYENRREL----------
18	Neisseria cinerea Cas9	(modificado)	translación	IDELGKEIRP----------CRLKKRPPVR----------
24	Campylobacter lari Cas9	(modificado)	translación	ITPLGDKIKADFQPRENISLKTSKKVGLR----------
	Streptococcus pyogenes Cas9		translación	DATLIHQSITGLYETRIDLSQLGGD----------
7	Lactobacillus johnsonii Cas9	(modificado)	translación	NGVILSENAKLIYQSPTGLFKKSVKISDL----------
23	Mycoplasma gallisepticum Cas9	(modificado)	translación	NDYILLDAKDNFDILGLSKNRIDEILNKSKLGLDKIVK-
20	Parvibaculum lavamentivorans Cas9	(modificado)	translación	PNPILKDDAKKVSIDPIGRVRPSND----------
1	Corynebacter diptheriae Cas9	(modificado)	translación	LFSDGNVTVVRRDSLGRVKLESTAHLPVTWKVQ----
2	Sutterella wadsworthensis Cas9	(modificado)	translación	RSEIPIENVGNAKHIRFMYTVVSSNKKMNESYNWVSKS

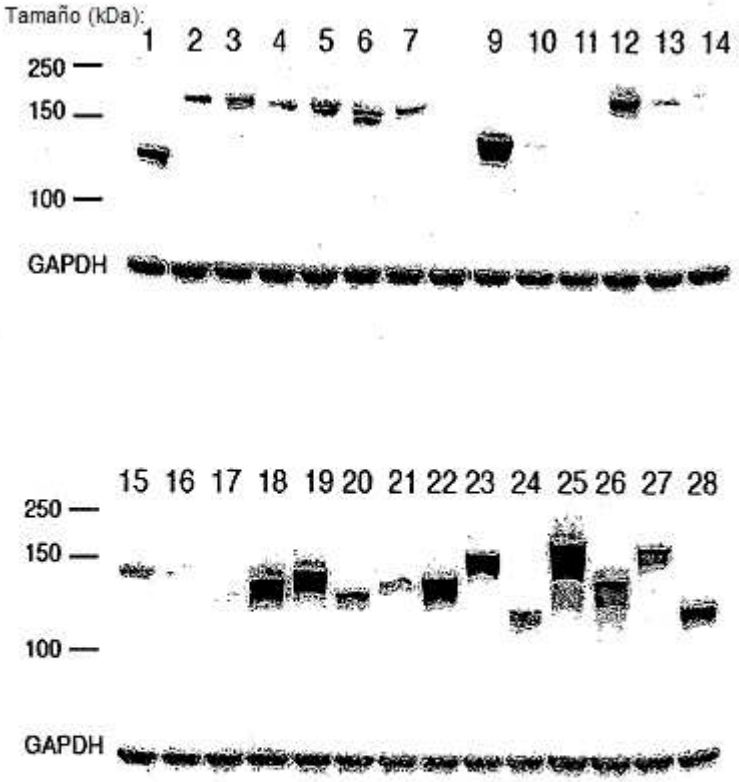
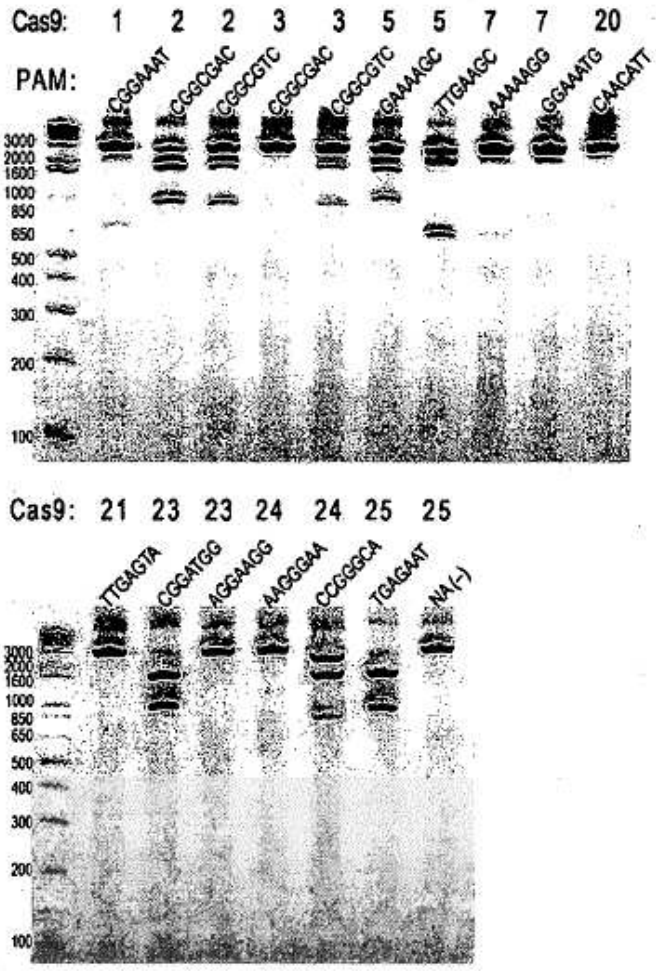


FIG. 20

1 Corynebacter_diphtheriae	119.24
2 Sutterella_wadsworthensis	156.42
3 Legionella_pneumophila_Paris	150.92
4 Treponema_denticola	153.45
5 Filifactor_alocis	150.15
6 Staphylococcus_pseudintermedius	146.74
7 Lactobacillus_johnsonii	151.25
*8 Eubacterium_ventriosum	121.77
9 Streptococcus_pasteurianus	124.3
10 Lactobacillus_farciminis	123.86
11 Mycoplasma_mobile	135.96
12 Bacteroides_coprophilus	165.99
13 Flavilvola_taffensis	160.38
14 Flavobacterium_columnare	168.85
15 Sphaerochaeta_globus_Buddy	129.69
16 Azospirillum_B510	128.48
17 Gluconacetobacter_diazotrophicus	115.5
18 Neisseria_cinerea	119.02
19 Roseburia_intestinalis	125.4
20 Parvibaculum_lavamentivorans	114.07
21 Staphylococcus aureus subsp. Aureus –	115.83
22 Nitratifractor salsuginis DSM 16511	124.52
23 Mycoplasma gallisepticum str. F	139.59
24 Campylobacter lari CF89-12	110.33
25 Streptococcus pyogenes	158
26 Streptococcus thermophilus CRISPR 1	127
27 Streptococcus thermophilus CRISPR 3	153
28 Campylobacter jejuni	110

* sin ADN añadido para nº 8



Cas9 nº	especie	PAM putativo	diana pUC19
1	Corynebacter_diphtheriae		AAAAGGGAATAAGGGCGACA
2	Sutterella_wadsworthensis	NGGNNNN	CATTCTGAGAATAGTGTATG
2	Sutterella_wadsworthensis		GCGACCGAGTTGCTCTTGCC
3	Legionella_pneumophila_Paris	NGGNNNN	CATTCTGAGAATAGTGTATG
3	Legionella_pneumophila_Paris		GCGACCGAGTTGCTCTTGCC
5	Filifactor_alocis	NNNAAGC	TTGAGTACTCACCAGTCACA
5	Filifactor_alocis		CTCTTCCTTTTTCAATATTA
7	Lactobacillus_johnsonii	NNAAANN	TAAATGCTTCAATAATATTG
7	Lactobacillus_johnsonii		AAAGGGAATAAGGGCGACAC
20	Parvibaculum_lavamentivorans	NNNCATN	AAAGGAAGAGTATGAGTATT
21	Staphylococcus aureus subsp. Aureus		CTATTCTCAGAATGACTTGG
23	Mycoplasma gallisepticum str. F	NNGANgg	AGTCACAGAAAAGCATCTTA
23	Mycoplasma gallisepticum str. F		GTTTCTGGGTGAGCAAAAAC
24	Campylobacter lari CF89-12	NNGGGNA	GAAGGCAAAATGCCGCAAAA
24	Campylobacter lari CF89-12		GGTATTATCCCGTATTGACG
25	SpCas9		GGGACTCAACGAAGTCATTC
25	SpCas9		

FIG. 21

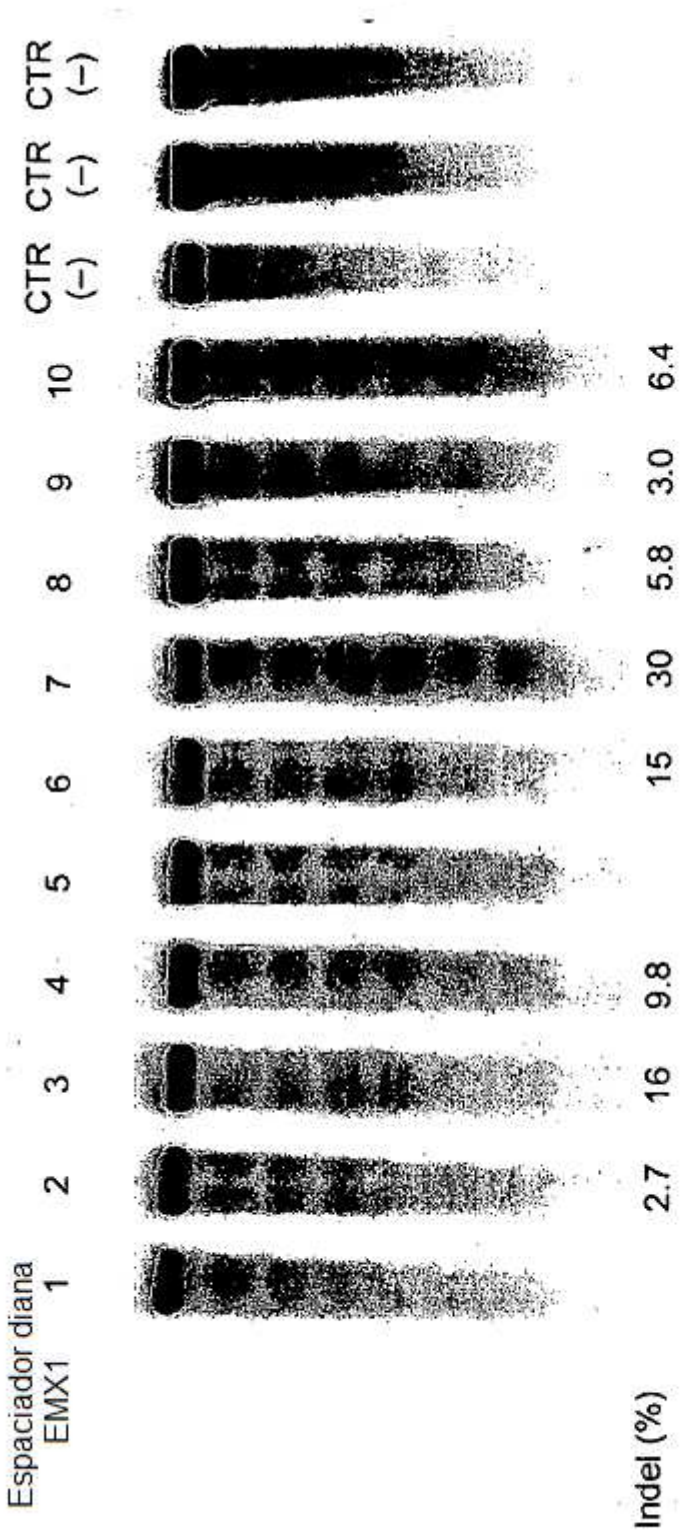

Espaciador diana EMX1
1
2
3
4
5
6
7
8
9
10
CTR (–)
CTR (–)
CTR (–)
Indel (%)
2.7
16
9.8
15
30
5.8
3.0
6.4

FIG. 22

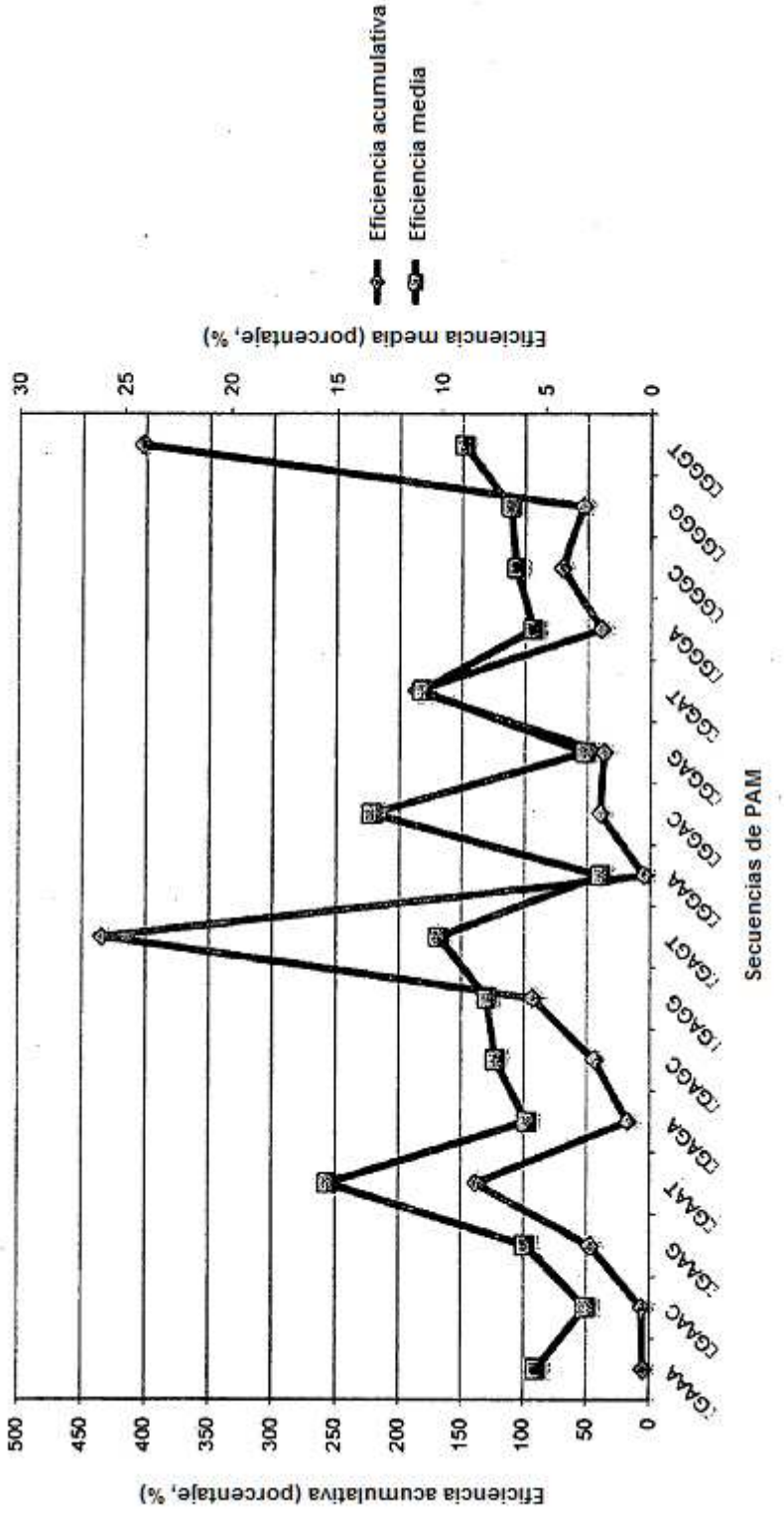

Eficiencia acumulativa
Eficiencia media
Eficiencia media (porcentaje, %)
Eficiencia acumulativa (porcentaje, %)
Secuencias de PAM
0
5
10
15
20
25
30
0
50
100
150
200
250
300
350
400
450
500
TGAA
TGAC
TGAG
TGAT
TGCA
TGCG
TGCC
TGCT
TGGA
TGGG
TGGC
TGGT

FIG. 23

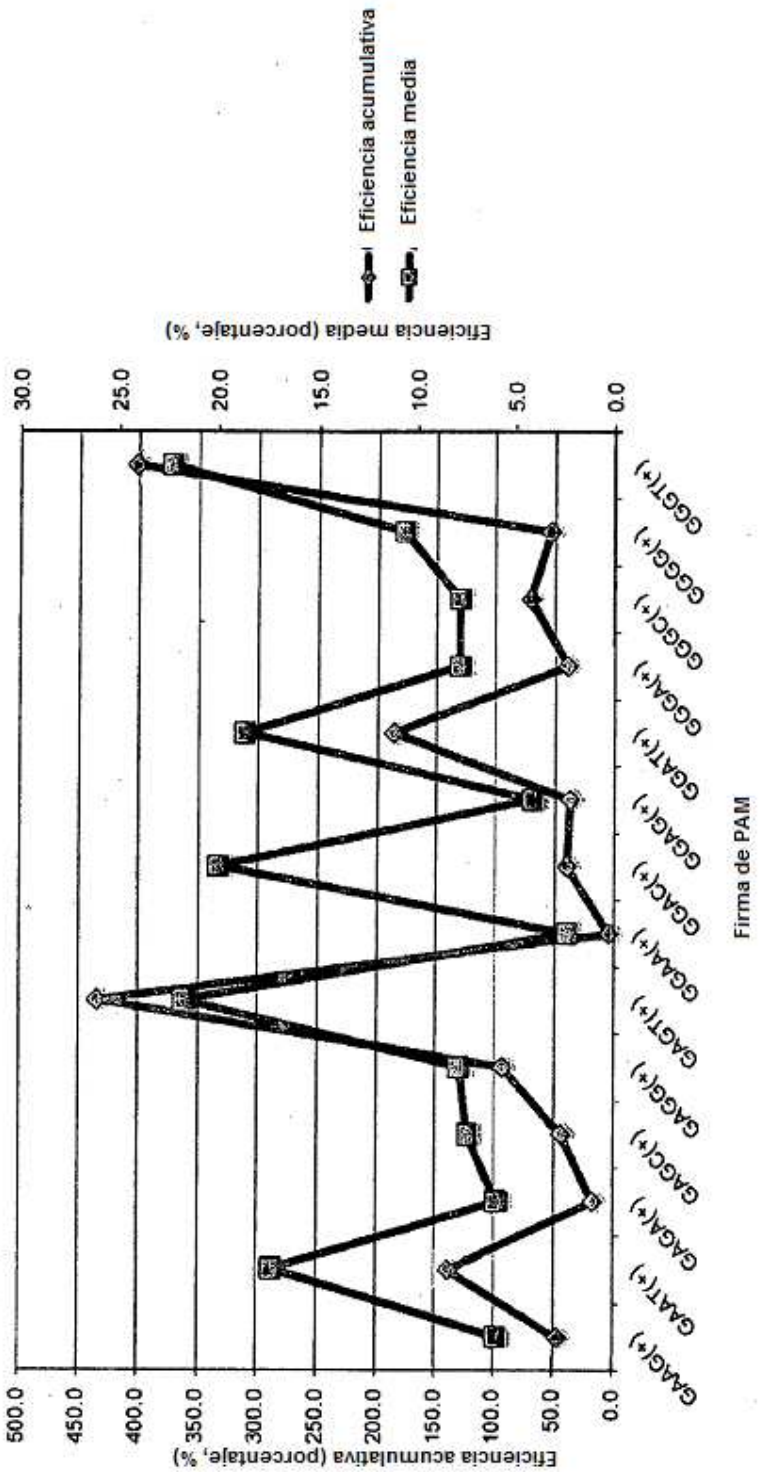

Eficiencia acumulativa
Eficiencia media
Eficiencia media (porcentaje, %)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
500.0
450.0
400.0
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
Eficiencia acumulativa (porcentaje, %)
GAAG(+)
GAAT(+)
GAGA(+)
GAGC(+)
GAGG(+)
GAGT(+)
GGAA(+)
GGAC(+)
GGAG(+)
GGAT(+)
GGGA(+)
GGGC(+)
GGGG(+)
GGGT(+)
Firma de PAM

FIG. 24

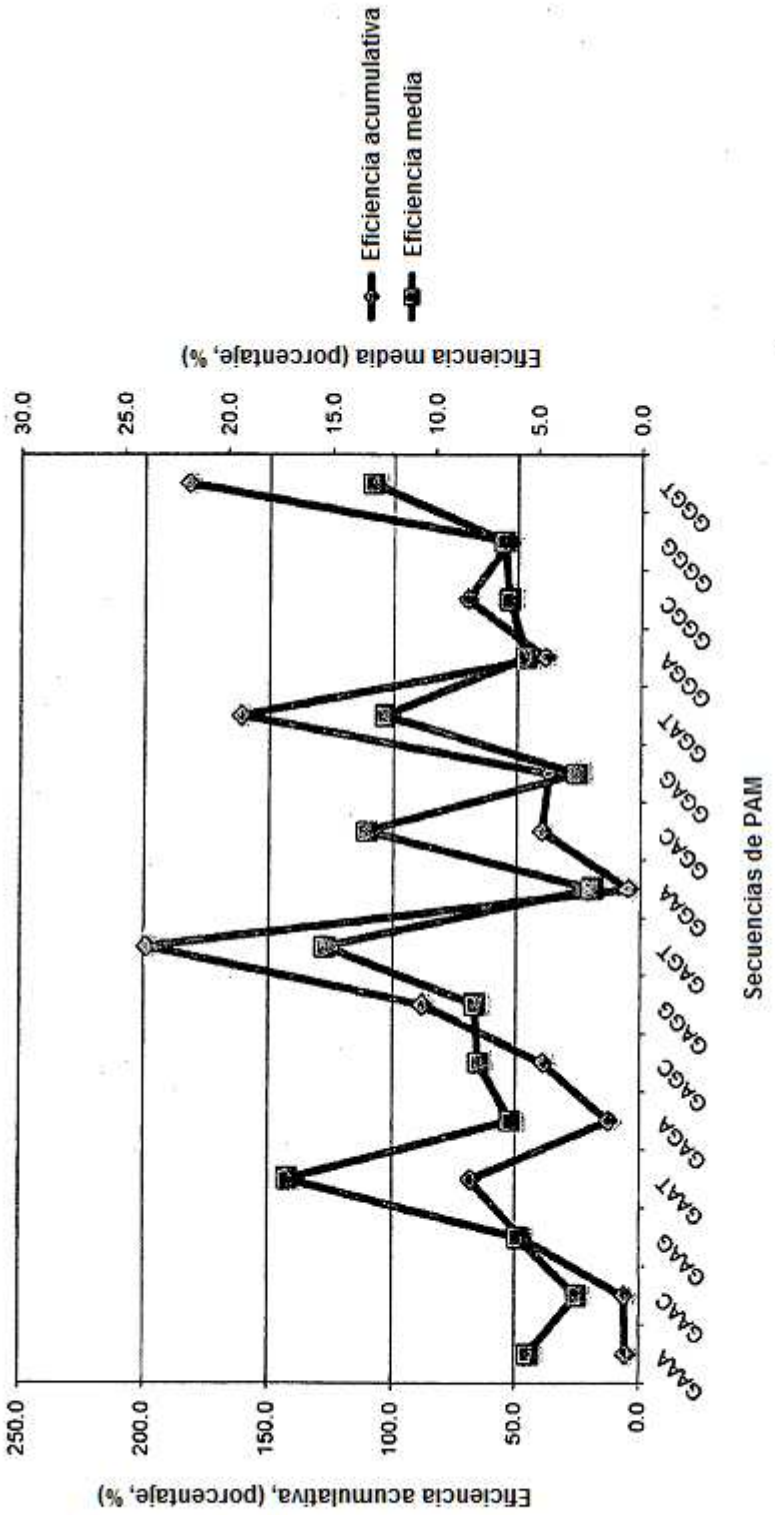

Eficiencia acumulativa
Eficiencia media
Eficiencia media (porcentaje, %)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
250.0
200.0
150.0
100.0
50.0
0.0
Eficiencia acumulativa, (porcentaje, %)
GAAA
GAAC
GAAG
GAAT
GACA
GACC
GACG
GACT
GAGA
GAGC
GAGG
GAGT
GATA
GATC
GATG
GATT
GCAA
GCAC
GCAG
GCAT
GCCA
GCCC
GCCG
GCCT
GGAA
GGAC
GGAG
GGAT
GGCA
GGCC
GGCG
GGCT
GGGA
GGGC
GGGG
GGGT
Secuencias de PAM

FIG. 25

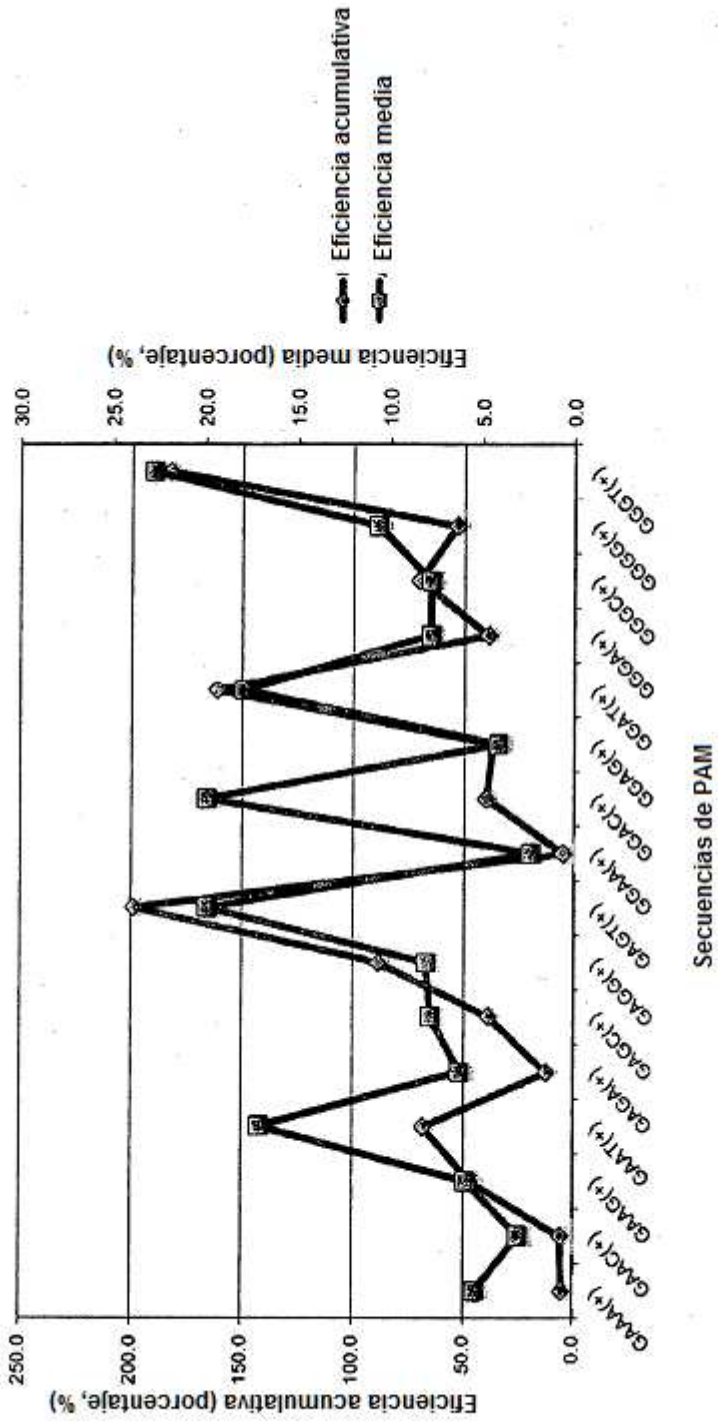

Eficiencia acumulativa
Eficiencia media
Eficiencia media (porcentaje, %)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Eficiencia acumulativa (porcentaje, %)
250.0
200.0
150.0
100.0
50.0
0.0
GAAA(+)
GACA(+)
GAAG(+)
GAAT(+)
GACA(+)
GACC(+)
GAGA(+)
GAGC(+)
GAGG(+)
GAGT(+)
GAAA(+)
GACA(+)
GAAG(+)
GAAT(+)
GGCA(+)
GGGC(+)
GGGG(+)
GGGT(+)
Secuencias de PAM

FIG. 26

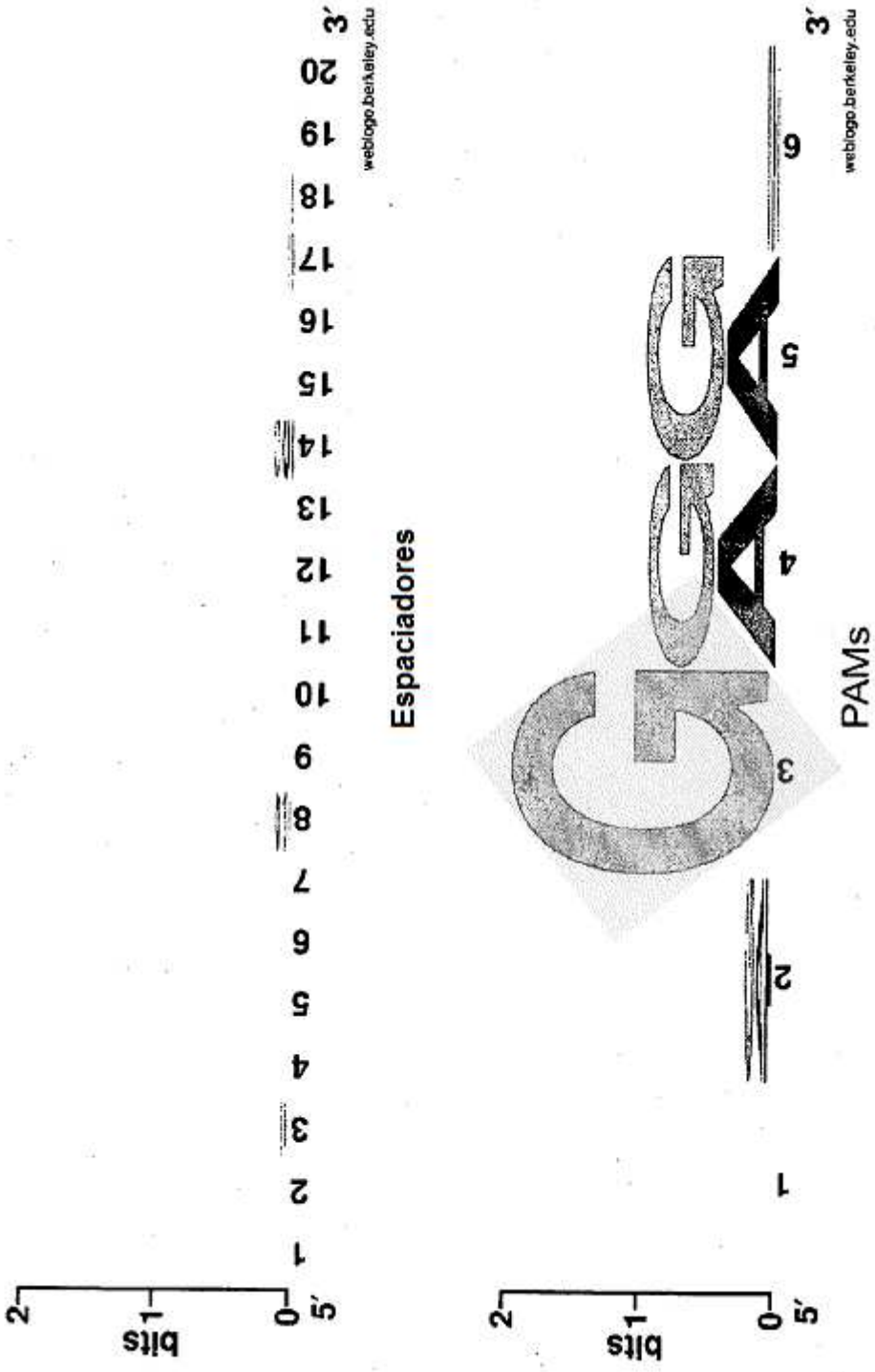

bits
2
1
0
5'
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
3'
weblogo.berkeley.edu
Espaciadores
bits
2
1
0
5'
1
2
3
4
5
6
3'
weblogo.berkeley.edu
PAMs

FIG. 27