

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 612**

51 Int. Cl.:

**C12N 15/63**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.08.2016** **E 22195354 (0)**

97 Fecha y número de publicación de la concesión europea: **27.12.2023** **EP 4155409**

54 Título: **Método de aumento de la replicación de una molécula de ADN circular**

30 Prioridad:

**10.08.2015 WO PCT/EP2015/068398**

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

**28.06.2024**

73 Titular/es:

**CUREVAC MANUFACTURING GMBH (100.0%)  
Friedrich-Miescher-Straße 15  
72076 Tübingen, DE**

72 Inventor/es:

**WILLIAMS, JIM**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 974 612 T3

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

## DESCRIPCIÓN

Método de aumento de la replicación de una molécula de ADN circular

5 **Campo de la invención**

La presente invención se refiere a un molde de ADN plasmídico lineal tal como se define en las reivindicaciones en el que la secuencia homopolimérica no está ubicada directamente junto al origen de replicación en la dirección de replicación, pero en el que hay una determinada distancia entre la secuencia homopolimérica y el origen de replicación. La presente invención se refiere además a un método de producción del molde de ADN plasmídico lineal a partir de una molécula de ADN recombinante circular cerrado covalentemente. La presente invención también se refiere a un método de producción de ARN mediante transcripción *in vitro* usando el molde de ADN lineal.

15 **Antecedentes de la invención**

En la terapia génica y en muchas otras aplicaciones bioquímicas y biotecnológicas de importancia terapéutica, el uso de ácidos nucleicos con fines terapéuticos y de diagnóstico es de gran importancia. Por ejemplo, en los últimos años se han producido rápidos avances en el campo de la terapia génica y se han logrado resultados prometedores. Por tanto, los ácidos nucleicos se consideran herramientas importantes para la terapia génica y la vacunación profiláctica y terapéutica contra enfermedades infecciosas y neoplasias malignas. Por ejemplo, el ARN mensajero amplificado a partir de células tumorales se transfectó en células dendríticas para inducir inmunidad tumoral (Boczkowski *et al.* (2000) Cancer Research 60, 1028-1034).

Las moléculas terapéuticas de ácido ribonucleico (ARN) representan una clase prometedora de fármacos. Los agentes terapéuticos basados en ARN incluyen moléculas de ARNm que codifican antígenos para su uso como vacunas (Fotin-Mleczek *et al.* (2012) J. Gene Med. 14(6): 428-439). Además, se prevé usar moléculas de ARN para terapias de reemplazo, por ejemplo, para proporcionar proteínas ausentes como factores de crecimiento o enzimas a los pacientes (Karikó *et al.* (2012) Mol. Ther. 20(5): 948-953; Kormann *et al.* (2012) Nat. Biotechnol. 29(2): 154-157). Además, se considera el uso terapéutico de moléculas de ARN inmunoestimulantes no codificantes (Heidenreich *et al.* Int J Cancer. 21 de diciembre de 2014. doi: 10.1002/ijc.29402.) y otros ARN no codificantes como microARN y ARN largos no codificantes (Esteller (2011) Nat. Rev. Genet. 15(12): 861-74).

Los agentes terapéuticos basados en ARN presentan algunas propiedades superiores a la transfección de ADN en células. Como se sabe generalmente, la transfección de moléculas de ADN puede conducir a problemas graves. Por ejemplo, la aplicación de moléculas de ADN conlleva el riesgo de que el ADN se integre en el genoma del huésped. La integración de ADN foráneo en el genoma del huésped puede influir en la expresión de los genes del huésped y posiblemente desencadena la expresión de un oncogén o la destrucción de un gen supresor de tumores. Además, un gen y, por tanto, el producto génico, que es esencial para el huésped, puede inactivarse mediante la integración del ADN foráneo en la región codificante de este gen. Puede existir un peligro particular si la integración del ADN tiene lugar en un gen que participa en la regulación del crecimiento celular. No obstante, el ADN sigue representando una herramienta importante, aunque algunos riesgos están asociados con la aplicación de ADN. Estos riesgos no se producen si se usa ARN, particularmente ARNm, en lugar de ADN. Una ventaja de usar ARN en lugar de ADN es que no es necesario administrar ningún elemento promotor derivado del virus *in vivo* y no puede producirse integración en el genoma. Además, el ARN no tiene que cruzar la barrera hacia el núcleo.

En los enfoques de terapia génica, se usa normalmente ADN aunque el ARN también se conoce en desarrollos recientes. Es importante destacar que en todos estos enfoques de terapia génica, el ARNm funciona como mensajero de la información de secuencia de la proteína codificada, independientemente de si se usa ADN, ARN viral o ARNm. En general, el ARN se considera una molécula inestable: las ARNasas son ubicuas y notoriamente difíciles de inactivar. Además, el ARN también es químicamente más lábil que el ADN. Por tanto, quizás sea sorprendente que el "estado por defecto" de un ARNm en una célula eucariota se caracteriza por una estabilidad relativa y se requieran señales específicas para acelerar la descomposición de los ARNm individuales. El motivo principal de este hallazgo parece ser que la descomposición del ARNm dentro de las células es catalizada casi exclusivamente por exonucleasas. Sin embargo, los extremos de los ARNm eucariotas están protegidos contra estas enzimas por estructuras terminales específicas y sus proteínas asociadas: una caperuza de m7GpppN en el extremo 5' y normalmente una secuencia de poli(A) en el extremo 3'. La eliminación de estas dos modificaciones terminales se considera, por tanto, un factor que limita la tasa de descomposición del ARNm. Aunque se ha caracterizado un elemento estabilizador en la 3' UTR del ARNm de alfa-globina, las secuencias de ARN que afectan el recambio de ARNm eucariotas actúan normalmente como un promotor de la descomposición, acelerando habitualmente la desadenilación (revisado en Meyer, S., C. Temme, *et al.* (2004), Crit Rev Biochem Mol Biol 39(4): 197-216.).

Para la fabricación de agentes terapéuticos basados en nucleótidos, se requiere la producción eficaz de vectores plasmídicos. Sin embargo, los presentes inventores han descubierto que las secuencias homopoliméricas están deteriorando el proceso de replicación conduciendo a una velocidad de replicación disminuida o incluso a la interrupción del proceso de replicación. Como consecuencia, el rendimiento de la producción fermentativa de plásmidos que comprenden secuencias homopoliméricas disminuye en comparación con los plásmidos sin tales

secuencias. Además, las secuencias homopoliméricas se eliminan a menudo al menos parcialmente durante la producción fermentativa de los plásmidos.

Por tanto, existe la necesidad de una replicación eficaz de vectores que comprenden secuencias homopoliméricas con alta fidelidad.

Revisión C de Agilent: "pCMV-Script Vector Instructions Manual" (descargado de <https://www.agilent.com/cs/library/usermanuals/public/212220.pdf>) y Novopro: "pCMV-Script vector map and sequence" descargado de <https://www.novoprolabs.com/vector/V11243>) describen el vector plasmídico pCMV-Script. El vector comprende una secuencia señal de poli(A) de TK de VHS y una secuencia señal de poli(A) de SV40.

Addgene pCRII-TOPO CMV-cGFP-bGH Poly(A) Sense - Sequence Analyzer (descargado de <https://www.addgene.org/browse/sequence/66680>) y Addgene: pCRII-TOPO CMV-cGFP-bGH Poly(A) Sense (descargado de <https://www.addgene.org/46835>) describen el vector plasmídico pCRII-TOPO CMV-cGFP-bGH Poly(A) Sense. El vector comprende una secuencia señal de poli(A) de bGH.

El documento WO 2014/152027 se refiere a métodos de fabricación para la producción de transcritos de ARN.

Mignon *et al.*, Pathogens 4 (2015), 157-181 describen la selección libre de antibióticos en productos bioterapéuticos.

## Sumario de la invención

Una solución a estos problemas es la provisión de un molde de ADN plasmídico lineal tal como se define en las reivindicaciones en el que la secuencia homopolimérica no está ubicada directamente junto al origen de replicación en la dirección de replicación, pero en el que hay una determinada distancia entre la secuencia homopolimérica y el origen de replicación. En un aspecto adicional, la presente invención se refiere a un método de producción del molde de ADN plasmídico lineal a partir de una molécula de ADN recombinante circular cerrado covalentemente.

En un aspecto aún adicional, la presente invención se refiere a un método de producción de ARN mediante transcripción *in vitro* usando el molde de ADN lineal.

Por consiguiente, la presente invención se refiere a un molde de ADN plasmídico lineal para transcripción *in vitro* que comprende

- un origen bacteriano de replicación,

- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,

- un promotor para una ARN polimerasa dependiente de ADN,

- un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.

En un aspecto adicional, la presente invención se refiere a un método de producción de un molde de ADN plasmídico lineal para transcripción *in vitro* que comprende la etapa de:

linealizar un ADN plasmídico, en el que el ADN plasmídico comprende

- un origen bacteriano de replicación,

- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,

- un promotor para una ARN polimerasa dependiente de ADN,

- un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.

La molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de

ADN plasmídico lineal de la invención mediante linealización es un plásmido.

Normalmente, el origen de replicación es de origen bacteriano. Preferiblemente, el origen de replicación es un origen de alto número de copias. El origen de replicación puede derivarse del plásmido pBR322, plásmido pUC, plásmido pMB1, plásmido ColE1, plásmido R6K, plásmido p15A, plásmido pSC101 o fagémido F1. Preferiblemente, el origen de replicación se deriva del plásmido pUC.

En algunas realizaciones, la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención comprende además un sitio de ensamblaje del primosoma en la cadena pesada (PAS-BH).

Normalmente, la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención comprende además un marcador de selección, por ejemplo, un gen de resistencia a antibióticos o un marcador seleccionable de sacarosa.

En realizaciones preferidas, la región homopolimérica comprendida en el inserto de la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención comprende una secuencia de poli(A). En una realización, la región homopolimérica comprende una secuencia de poli(C) y una de poli(A). La secuencia de poli(A) comprende una secuencia de aproximadamente 60 a aproximadamente 250 nucleótidos de adenosina. En algunas realizaciones, la región homopolimérica comprende una secuencia de poli(A) de aproximadamente 60 a aproximadamente 70 nucleótidos de adenosina y una secuencia de poli(C) de aproximadamente 20 a 30 nucleótidos de citidina.

Cuando se usa la molécula de ADN recombinante circular cerrado adecuada para la preparación del molde de ADN plasmídico lineal de la invención en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, el tiempo de fermentación puede acortarse en comparación con el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está ubicado a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación.

Cuando se usa la molécula de ADN recombinante circular cerrado adecuada para la preparación del molde de ADN plasmídico lineal de la invención en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación, el tiempo de fermentación se acorta en comparación con el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación.

Un aspecto adicional de la invención se refiere a un método para la producción fermentativa de una molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención que comprende las etapas de:

(a) proporcionar un microorganismo que comprende la molécula de ADN recombinante circular cerrado covalentemente según la invención;

(b) fermentar el microorganismo de la etapa (a).

En una realización específica, el método comprende además la siguiente etapa de:

(c) aislar la molécula de ADN recombinante circular cerrado covalentemente a partir del microorganismo de la etapa (b).

En realizaciones particularmente preferidas, el microorganismo es una bacteria, preferiblemente *E. coli*. En realizaciones preferidas, el microorganismo es una bacteria (por ejemplo, *E. coli*) que contiene un plásmido de ADN de alto número de copias inducible por temperatura recombinante circular cerrado covalentemente y la etapa (b) comprende las siguientes etapas de:

(i) hacer crecer las bacterias que contienen un plásmido de ADN de alto número de copias inducible por temperatura recombinante circular cerrado covalentemente a una temperatura en el intervalo de 25°C a 32°C durante la fase de crecimiento de la fase semicontinua en la que el sustrato se suministra a una tasa tal que la tasa de crecimiento es de  $\mu =$  de 0,05 a 0,3 h<sup>-1</sup> durante la fase semicontinua,

(ii) inducir la producción de dicho plásmido de ADN después de la fase de crecimiento aumentando la temperatura a de 36°C a 45°C; y

(iii) continuar la fermentación a la temperatura aplicada en (ii) para acumular el dicho plásmido de ADN.

En una realización más específica, la temperatura aplicada en la etapa (i) es de aproximadamente 30°C y la temperatura aplicada en la etapa (ii) es de aproximadamente 42°C.

En los métodos descritos anteriormente, el rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación se aumenta en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está ubicado a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación.

El rendimiento del ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación se aumenta en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación. Además, el tiempo de fermentación se acorta, en comparación con el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación.

Finalmente, la invención se refiere a uso de la molécula de ADN recombinante circular cerrado covalentemente según la invención para la transcripción *in vitro* de ARN.

La presente invención proporciona un método de producción de ARN que comprende las etapas de:

(a) obtener ARN mediante transcripción *in vitro* dependiente de ADN de un molde de ADN plasmídico linealizado usando una ARN polimerasa dependiente de ADN, en el que el ADN plasmídico comprende

- un origen bacteriano de replicación,

- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,

- un promotor para una ARN polimerasa dependiente de ADN,

- un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación; y

(b) purificar el ARN.

#### Breve descripción de las figuras

Figura 1: mapa de vector de P1140

La orientación de P1140 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Figura 2: mapa de vector de P1140-AF2

La orientación de P1140-AF2 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es la misma que la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Figura 3: mapa de vector de P1140-AF1

La orientación de P1140-AF1 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Figura 4: mapa de vector de P1140-K2

La orientación de P1140-K2 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de

vector como origen de pUC) es la misma que la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Figura 5: mapa de vector de P1140-K1

La orientación de P1140-K1 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Figura 6: ARNm optimizado en GC que codifica para H1N1(Países Bajos 2009)-HA (SEQ ID NO: 2)

## Definiciones

En aras de la claridad y la legibilidad, se proporcionan las siguientes definiciones. Cualquier característica técnica mencionada para estas definiciones puede leerse en todas y cada una de las formas de realización de la invención. Pueden proporcionarse específicamente definiciones y explicaciones adicionales en el contexto de estas realizaciones. A menos que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento generalmente tienen el mismo significado que el comúnmente entendido por un experto en la técnica a la que pertenece esta invención. Generalmente, la nomenclatura usada en el presente documento y los procedimientos de laboratorio en cultivo celular, genética molecular, química orgánica y química e hibridación de ácidos nucleicos son los conocidos bien y comúnmente empleados en la técnica. Se usan técnicas convencionales para ensayos de biotecnología molecular. Las técnicas y procedimientos se realizan generalmente según métodos convencionales en la técnica y diversas referencias generales (por ejemplo, Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, 2ª ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), que se proporcionan a lo largo de este documento.

ADN recombinante circular cerrado covalentemente: el ADN recombinante circular cerrado covalentemente se refiere a una molécula de ADN circular bicatenaria que está intacta, es decir en la que ambas cadenas están sin cortar. Preferiblemente, la molécula no se produce de forma natural, sino que se obtiene por ingeniería, por ejemplo, mediante técnicas de clonación molecular. En el contexto de la presente invención, se prefieren particularmente los plásmidos tal como se definen en el presente documento.

Proteína: una proteína comprende normalmente uno o más péptidos o polipéptidos. Normalmente, una proteína se pliega en una forma tridimensional, que puede ser necesaria para que la proteína ejerza su función biológica. La secuencia de una proteína o péptido se entiende normalmente como el orden, es decir, la sucesión de sus aminoácidos.

ARN, ARNm: ARN es la abreviatura habitual de ácido ribonucleico. Es una molécula de ácido nucleico, es decir, un polímero que consiste en monómeros de nucleótidos. Estos nucleótidos suelen ser monómeros de adenosina monofosfato, uridina monofosfato, guanosina monofosfato y citidina monofosfato, que están conectados entre sí a lo largo de una denominada estructura principal. La estructura principal está formada por enlaces fosfodiéster entre el azúcar, es decir, la ribosa, de un primero y un resto fosfato de un segundo monómero adyacente. El orden específico de los monómeros, es decir, el orden de las bases unidas a la estructura principal de azúcar/fosfato, se denomina secuencia de ARN. Habitualmente, el ARN puede obtenerse mediante la transcripción de una secuencia de ADN, por ejemplo, dentro de una célula. En las células eucariotas, la transcripción se realiza normalmente dentro del núcleo o las mitocondrias. *In vivo*, la transcripción de ADN da como resultado habitualmente el denominado ARN prematuro, que debe procesarse para dar el llamado ARN mensajero, habitualmente abreviado como ARNm. El procesamiento del ARN prematuro, por ejemplo, en organismos eucariotas, comprende una variedad de modificaciones postranscripcionales diferentes tales como corte y empalme, ocupación del extremo 5', poliadenilación, exportación desde el núcleo o las mitocondrias y similares. La suma de estos procesos también se denomina maduración de ARN. El ARN mensajero maduro proporciona habitualmente la secuencia de nucleótidos que puede traducirse en una secuencia de aminoácidos de un péptido o una proteína particular. Normalmente, un ARNm maduro comprende una caperuza 5', opcionalmente una 5' UTR, un marco de lectura abierto, opcionalmente una 3' UTR y una secuencia de poli(A). Aparte del ARN mensajero, existen varios tipos de ARN no codificantes que pueden estar implicados en la regulación de la transcripción y/o traducción y la inmunoestimulación. El término "ARN" abarca además otras moléculas de ARN codificante, tales como ARN viral, ARN retroviral y ARN replicón, ARN de interferencia pequeño (ARNip), ARN antisentido, ARN CRISPR, ribozimas, aptámeros, ribointerruptores, ARN inmunoestimulante, ARN de transferencia (ARNt), ARN ribosómico (ARNr), ARN nuclear pequeño (ARNnn), ARN nucleolar pequeño (ARNsno), microARN (miARN) y ARN que interactúa con Piwi (piARN). En el contexto de la presente invención, se prefieren particularmente los ARN codificantes tales como ARNm, ARN viral, ARN retroviral y ARN replicón.

La región codificante del ARNm para su uso según la presente invención puede producirse como un ARNm monocistrónico, dicistrónico o incluso multicistrónico, es decir, un ARNm, que porta las secuencias codificantes de una, dos o más proteínas o péptidos. Tales secuencias codificantes en ARNm dicistrónicos o incluso multicistrónicos pueden estar separadas por al menos una secuencia de sitio interno de entrada al ribosoma (IRES), por ejemplo, tal

como se describe en el presente documento o por péptidos señal que inducen la escisión del polipéptido resultante, que comprende varias proteínas o péptidos.

ARN bicistrónico/multicistrónico: el término ARN "bicistrónico" o "multicistrónico" tal como se usa en el presente documento es normalmente ARN, preferiblemente un ARNm, que normalmente tiene dos (bicistrónicos) o más (multicistrónicos) marcos de lectura abiertos (ORF). Un marco de lectura abierto en este contexto es una secuencia de codones que puede traducirse en un péptido o proteína.

ARN de interferencia pequeño (ARNip): "ARN de interferencia pequeño" (ARNip) se refiere a moléculas de ARN bicatenario que tienen una longitud de desde aproximadamente 10 hasta aproximadamente 30 nucleótidos y que se denominan por su capacidad para interferir específicamente con la expresión de proteínas. Preferiblemente, las moléculas de ARNip tienen una longitud de 12-28 nucleótidos, más preferiblemente una longitud de 15-25 nucleótidos, todavía más preferiblemente una longitud de 19-23 nucleótidos y lo más preferiblemente una longitud de 21-23 nucleótidos. Por tanto, las moléculas de ARNip preferidas tienen una longitud de 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 ó 29 nucleótidos.

ARN antisentido: el término "ARN antisentido" incluye cualquier molécula de ARN que es adecuada como ARN antisentido para regiones de ARN, preferiblemente ARNm, y más preferiblemente para elementos reguladores del mismo, y que puede provocar inhibición de la expresión génica mediante unión complementaria a estas regiones. El ARN antisentido también puede comprender secuencias de ADN.

ARN CRISPR: el término "ARN CRISPR" se deriva de CRISPR (grupos de repeticiones palindrómicas cortas en intervalos regulares) que se desarrollaron en bacterias como un sistema inmunitario adaptativo para defenderse del ataque viral. Los "grupos de repeticiones palindrómicas cortas en intervalos regulares" o "CRISPR", tal como se usa en el presente documento, se refiere a loci que contienen múltiples repeticiones directas cortas que se encuentran en los genomas de aproximadamente el 40% de las bacterias secuenciadas y el 90% de las arqueas secuenciadas. El sistema CRISPR es un sistema de nucleasa microbiano implicado en la defensa contra los fagos y plásmidos invasores que proporciona una forma de inmunidad adquirida. Tras la exposición a un virus, se integran segmentos cortos de ADN viral en el locus CRISPR. El "ARN CRISPR" se transcribe a partir de una porción del locus CRISPR que incluye la secuencia viral. Ese ARN, que contiene una secuencia complementaria al genoma viral, media el direccionamiento de una proteína Cas9 a la secuencia en el genoma viral. La proteína Cas9 escinde y, por tanto, silencia la diana viral. Los agrupamientos CRISPR se transcriben y se procesan para dar ARN CRISPR (grupos de repeticiones palindrómicas cortas en intervalos regulares) maduro (ARN CRISPR).

Ribozimas: las "ribozimas" son normalmente moléculas de ARN que tienen una actividad enzimática que puede escindir repetidamente otras moléculas diana de ARN separadas de una manera específica de la secuencia de bases de nucleótidos, bloqueando su expresión y afectando las funciones normales de otros genes. Las ribozimas pueden seleccionar como diana prácticamente cualquier transcrito de ARN y lograr una escisión eficaz *in vitro*.

Aptámeros: el término "aptámeros" se entiende en el presente documento que se refiere a moléculas de ácido nucleico que tienen afinidad de unión específica a moléculas a través de interacciones distintas del apareamiento de bases clásico de Watson-Crick. El término "aptámero de ARN" tal como se usa en el presente documento es un aptámero que comprende unidades de ribonucleósidos. Los aptámeros, preferiblemente aptámeros de ARN, pueden unirse específicamente a dianas seleccionadas y modular la actividad de la diana. Los aptámeros tienen varias características deseables para su uso como agentes terapéuticos y de diagnóstico que incluyen alta especificidad y afinidad, eficacia biológica y excelentes propiedades farmacocinéticas. Los ribointerruptores son elementos de ARN que residen en la región 5' no traducida (UTR) de los genes y modulan su expresión utilizando mecanismos transcripcionales o traduccionales (Roth y Breaker (2009) Annu Rev Biochem 78: 305-334).

ARN nuclear pequeño (ARNnn): la expresión "ARN nuclear pequeño" (ARNnn) se refiere a moléculas pequeñas de ARN que se sintetizan y/o funcionan en el nucleoplasma y/o el nucléolo de la célula.

ARN nucleolares pequeños (ARNsno): los "ARN nucleolares pequeños" (ARNsno) son ARN no codificantes implicados en el procesamiento del ARN. Hay dos subclases principales de ARNsno, denominadas ARNsno de caja C/D y caja H/ACA. Estas dos clases contienen secuencias guía que se sabe que se emparejan canónicamente con regiones complementarias en un pre-ARNr diana, formando un dúplex de ARN y facilitando la actividad enzimática de metilasas y uridilasas que modifican de forma específica del sitio las bases del pre-ARNr por 2'-O-metilación o pseudouridilación, respectivamente. Estas modificaciones del ARNr son críticas para el ensamblaje y la viabilidad de los ribosomas.

Micro ARN (miARN): los "micro ARN" (miARN) desempeñan un papel importante en la regulación de la actividad génica. Los microARN son ARN monocatenarios de normalmente 22 nucleótidos que se procesan a partir de precursores de ARN en horquilla de aproximadamente 70 nucleótidos por la ARNasa III nucleasa. Al igual que los ARNip, los miARN pueden silenciar la actividad génica mediante la destrucción del ARNm homólogo en plantas o el bloqueo de su traducción en plantas y animales. Estos ARN no codificantes de 20-22 nucleótidos tienen la capacidad de hibridar mediante apareamiento de bases con ARNm diana específicos y regular por disminución la

expresión de estos transcritos, al mediar la escisión del ARN o la represión de la traducción. Estudios recientes han indicado que los miARN tienen funciones importantes durante el desarrollo. Además, los miARN se transcriben por la ARN polimerasa 11 como mensajes poliadenilados y ocupados en los extremos conocidos como pri-miARN.

ARN que interactúa con Piwi (piARN): el "ARN que interactúa con Piwi" (piARN) es una clase de pequeñas moléculas de ARN no codificantes que se expresan o pueden introducirse en células animales (véase, por ejemplo, Seto *et al.* (2007) *Molecular Cell*, 26(5): 603-609; Siomi *et al.* (2011) *Nat. Rev. Mol. Cell. Biol.*, 12: 246-258). Los piARN forman complejos de ARN-proteína a través de interacciones con las proteínas Piwi. Estos complejos piARN se han relacionado con el silenciamiento génico tanto epigenético como postranscripcional de retrotransposones y otros elementos genéticos.

ARN inmunoestimulante: los términos "ARN inmunoestimulante" o "ARN inmunoestimulador" (isARN) en el contexto de la invención pueden ser normalmente un ARN que puede inducir una respuesta inmunitaria innata. Un isARN habitualmente no tiene un marco de lectura abierto y, por tanto, no proporciona un péptido-antígeno, pero provoca una respuesta inmunitaria innata, por ejemplo, uniéndose a receptores de patrones moleculares asociados a patógenos (PAMP) (por ejemplo, receptor de tipo Toll (TLR) u otros sensores de ARN intracelulares (por ejemplo, RIG-I, MDA-5 o PKR).

Secuencia homopolimérica: una secuencia homopolimérica comprende un tramo de nucleótidos idénticos, tal como una secuencia de nucleótidos de adenina, una secuencia de nucleótidos de citosina, una secuencia de nucleótidos de guanina, una secuencia de nucleótidos de timina, una secuencia de nucleótidos modificados idénticos o una secuencia de análogos de nucleótidos idénticos. Preferiblemente, la secuencia homopolimérica comprende una secuencia de poli(A) que comprende al menos 60 nucleótidos de adenina.

Región homopolimérica: una región homopolimérica comprende al menos una secuencia homopolimérica. Preferiblemente, la región homopolimérica comprende al menos dos secuencias homopoliméricas. Las secuencias homopoliméricas pueden estar conectadas por una secuencia ligadora que podría ser una secuencia heteropolimérica. Por tanto, la región homopolimérica puede comprender diferentes nucleótidos, es decir, ser una secuencia heteropolimérica. En general, la región homopolimérica es una secuencia de nucleótidos, por ejemplo, de hasta aproximadamente 1000 nucleótidos, por ejemplo, desde aproximadamente 20 hasta aproximadamente 800, preferiblemente desde aproximadamente 50 hasta aproximadamente 600, más preferiblemente desde aproximadamente 50 hasta aproximadamente 500, incluso más preferiblemente desde aproximadamente 50 hasta aproximadamente 300, lo más preferiblemente desde aproximadamente 60 hasta aproximadamente 250 nucleótidos, que se encuentra ubicado preferiblemente en 3' del marco de lectura abierto comprendido opcionalmente en el inserto de la molécula de ADN recombinante circular cerrado covalentemente según la invención.

Secuencia de poli(N): una secuencia de poli(N), también denominada cola de poli(N) o cola de 3'-poli(N), se entiende habitualmente que es una secuencia de nucleótidos, por ejemplo, de hasta aproximadamente 400 nucleótidos, por ejemplo, desde aproximadamente 20 hasta aproximadamente 400, preferiblemente desde aproximadamente 50 hasta aproximadamente 400, más preferiblemente desde aproximadamente 50 hasta aproximadamente 300, incluso más preferiblemente desde aproximadamente 50 hasta aproximadamente 250, lo más preferiblemente desde aproximadamente 60 hasta aproximadamente 250 nucleótidos, que se añade preferiblemente al extremo terminal 3' de un ARN, en particular ARNm. Una secuencia de poli(N) está ubicada normalmente en el extremo 3' de un ARN(m). En el contexto de la presente invención, una secuencia de poli(N) puede estar ubicada dentro de un ARN(m) o cualquier otra molécula de ácido nucleico, tal como, por ejemplo, en un vector, por ejemplo, en un vector que sirve como molde para la generación de un ARN, preferiblemente un ARNm, por ejemplo, mediante transcripción del vector. Los nucleótidos pueden ser adenosina, citidina, uridina, guanosina, nucleótidos modificados, análogos de nucleótidos y mezclas de los mismos.

Secuencia de poli(A): una secuencia de poli(A), también denominada cola de poli(A) o cola de 3'-poli(A), se entiende habitualmente que es una secuencia de desde aproximadamente 60 hasta aproximadamente 250 nucleótidos de adenina, que se añade preferiblemente al extremo terminal 3' de un constructo que codifica para un ARNm. Una secuencia de poli(A) está ubicada normalmente en el extremo 3' de un ARN, en particular ARNm. En el contexto de la presente invención, una secuencia de poli(A) puede estar ubicada dentro de un ARN(m) o cualquier otra molécula de ácido nucleico, tal como, por ejemplo, en un vector, por ejemplo, en un vector que sirve como molde para la generación de un ARN, preferiblemente un ARNm, por ejemplo, mediante transcripción del vector. En general, la secuencia de poli(A) consiste de monofosfatos de adenosina.

Modificaciones de azúcar: los nucleósidos y nucleótidos modificados, que pueden usarse en el contexto de la presente invención, pueden modificarse en el resto de azúcar. Por ejemplo, el grupo hidroxilo 2' (OH) puede modificarse o reemplazarse con varios sustituyentes "oxilo" o "desoxilo" diferentes. Los ejemplos de modificaciones del grupo hidroxilo "oxilo"-2' incluyen, pero no se limitan a, alcoxilo o ariloxilo (-OR, por ejemplo, R = H, alquilo, cicloalquilo, arilo, aralquilo, heteroarilo o azúcar); polietilenglicoles (PEG), -O(CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>OR; ácidos nucleicos "bloqueados" (LNA) en los que el hidroxilo 2' está conectado, por ejemplo, mediante un puente de metileno, al carbono 4' del mismo azúcar ribosa; y grupos amino (-O-amino, en los que el grupo amino, por ejemplo, NRR, puede ser alquilamino, dialquilamino, heterociclilo, arilamino, diarilamino, heteroarilamino o diheteroarilamino,

etilendiamina, poliamino) o aminoalcoxilo.

Las modificaciones de "desoxilo" incluyen hidrógeno, amino (por ejemplo, NH<sub>2</sub>; alquilamino, dialquilamino, heterociclilo, arilamino, diarilamino, heteroarilamino, diheteroarilamino o aminoácido); o el grupo amino puede estar unido al azúcar a través de un grupo de unión, en el que el grupo de unión comprende uno o más de los átomos C, N y O.

El grupo de azúcar también puede contener uno o más carbonos que poseen la configuración estereoquímica opuesta a la del carbono correspondiente en la ribosa. Por tanto, un nucleótido modificado puede incluir nucleótidos que contienen, por ejemplo, arabinosa como azúcar.

Modificaciones de la estructura principal: la estructura principal de fosfato puede modificarse además en los nucleósidos y nucleótidos modificados. Los grupos fosfato de la estructura principal pueden modificarse reemplazando uno o más de los átomos de oxígeno con un sustituyente diferente. Además, los nucleósidos y nucleótidos modificados pueden incluir el reemplazo completo de un resto de fosfato no modificado con un fosfato modificado tal como se describe en el presente documento. Los ejemplos de grupos fosfato modificados incluyen, pero no se limitan a, fosforotioato, fosforoselenatos, boranofosfatos, ésteres de boranofosfato, hidrogenofosfonatos, fosforamidatos, fosfonatos de alquilo o arilo y fosfotriésteres. Los fosforoditioatos tienen ambos oxígenos no de enlace reemplazados por azufre. El grupo de unión de fosfato también puede modificarse mediante el reemplazo de un oxígeno de enlace con nitrógeno (fosforotioatos en puente), azufre (fosforotioatos en puente) y carbono (fosfonatos de metileno en puente).

Modificaciones de base: los nucleósidos y nucleótidos modificados, que pueden usarse en la presente invención, pueden modificarse adicionalmente en el resto de nucleobase. Los ejemplos de nucleobases que se encuentran en el ARN incluyen, pero no se limitan a, adenina, guanina, citosina y uracilo. Por ejemplo, los nucleósidos y nucleótidos descritos en el presente documento pueden modificarse químicamente en la cara del surco mayor. En algunas realizaciones, las modificaciones químicas del surco mayor pueden incluir un grupo amino, un grupo tiol, un grupo alquilo o un grupo halógeno.

En realizaciones particularmente preferidas de la presente invención, los análogos/las modificaciones de nucleótidos se seleccionan de modificaciones de bases, que se seleccionan preferiblemente de 2-amino-6-cloropurinaribósido-5'-trifosfato, 2-aminopurina-ribósido-5'-trifosfato; 2-aminoadenosina-5'-trifosfato, 2'-amino-2'-deoxicitidina-trifosfato, 2-tiocitidina-5'-trifosfato, 2-tiouridina-5'-trifosfato, 2'-fluorotimidina-5'-trifosfato, 2'-O-metilinosina-5'-trifosfato, 4-tiouridina-5'-trifosfato, 5-aminoalilcitidina-5'-trifosfato, 5-aminoaliluridina-5'-trifosfato, 5-bromocitidina-5'-trifosfato, 5-bromouridina-5'-trifosfato, 5-bromo-2'-deoxicitidina-5'-trifosfato, 5-bromo-2'-deoxiuridina-5'-trifosfato, 5-yodocitidina-5'-trifosfato, 5-yodo-2'-deoxicitidina-5'-trifosfato, 5-yodouridina-5'-trifosfato, 5-yodo-2'-deoxiuridina-5'-trifosfato, 5-metilcitidina-5'-trifosfato, 5-metiluridina-5'-trifosfato, 5-propinil-2'-deoxicitidina-5'-trifosfato, 5-propinil-2'-deoxiuridina-5'-trifosfato, 6-azacitidina-5'-trifosfato, 6-azauridina-5'-trifosfato, 6-cloropurinaribósido-5'-trifosfato, 7-deazaadenosina-5'-trifosfato, 7-deazaguanosina-5'-trifosfato, 8-azaadenosina-5'-trifosfato, 8-azidoadenosina-5'-trifosfato, bencimidazol-ribósido-5'-trifosfato, N1-metiladenosina-5'-trifosfato, N1-metilguanosina-5'-trifosfato, N6-metiladenosina-5'-trifosfato, O6-metilguanosina-5'-trifosfato, pseudouridina-5'-trifosfato o puomicina-5'-trifosfato, xantosina-5'-trifosfato. Se da preferencia particular a nucleótidos para modificaciones de bases seleccionados del grupo de nucleótidos modificados con bases que consiste en 5-metilcitidina-5'-trifosfato, 7-deazaguanosina-5'-trifosfato, 5-bromocitidina-5'-trifosfato y pseudouridina-5'-trifosfato.

En algunas realizaciones, los nucleósidos modificados incluyen piridin-4-ona-ribonucleósido, 5-aza-uridina, 2-tio-5-aza-uridina, 2-tiouridina, 4-tio-pseudouridina, 2-tio-pseudouridina, 5-hidroxiuridina, 3-metiluridina, 5-carboximetiluridina, 1-carboximetil-pseudouridina, 5-propinil-uridina, 1-propinil-pseudouridina, 5-taurinometiluridina, 1-taurinometil-pseudouridina, 5-taurinometil-2-tio-uridina, 1-taurinometil-4-tio-uridina, 5-metil-uridina, 1-metil-pseudouridina, 4-tio-1-metil-pseudouridina, 2-tio-1-metil-pseudouridina, 1-metil-1-deaza-pseudouridina, 2-tio-1-metil-1-deaza-pseudouridina, dihidrouridina, dihidropseudouridina, 2-tio-dihidrouridina, 2-tio-dihidropseudouridina, 2-metoxiuridina, 2-metoxi-4-tio-uridina, 4-metoxi-pseudouridina y 4-metoxi-2-tio-pseudouridina.

En algunas realizaciones, los nucleósidos modificados incluyen 5-aza-citidina, pseudoisocitidina, 3-metil-citidina, N4-acetilcitidina, 5-formilcitidina, N4-metilcitidina, 5-hidroximetilcitidina, 1-metil-pseudoisocitidina, pirrolo-citidina, pirrolo-pseudoisocitidina, 2-tio-citidina, 2-tio-5-metil-citidina, 4-tio-pseudoisocitidina, 4-tio-1-metil-pseudoisocitidina, 4-tio-1-metil-1-deaza-pseudoisocitidina, 1-metil-1-deaza-pseudoisocitidina, zebularina, 5-aza-zebularina, 5-metil-zebularina, 5-aza-2-tio-zebularina, 2-tio-zebularina, 2-metoxi-citidina, 2-metoxi-5-metil-citidina, 4-metoxi-pseudoisocitidina y 4-metoxi-1-metil-pseudoisocitidina.

En otras realizaciones, los nucleósidos modificados incluyen 2-aminopurina, 2,6-diaminopurina, 7-deaza-adenina, 7-deaza-8-aza-adenina, 7-deaza-2-aminopurina, 7-deaza-8-aza-2-aminopurina, 7-deaza-2,6-diaminopurina, 7-deaza-8-aza-2,6-diaminopurina, 1-metiladenosina, N6-metiladenosina, N6-isopenteniladenosina, N6-(cis-hidroxiisopentenil)adenosina, 2-metil-N6-(cis-hidroxiisopentenil)adenosina, N6-glicinilcarbamoiladenosina, N6-treonilcarbamoiladenosina, 2-metil-N6-treonilcarbamoiladenosina, N6,N6-dimetiladenosina, 7-metiladenina, 2-

metiltio-adenina y 2-metoxi-adenina.

En otras realizaciones, los nucleósidos modificados incluyen inosina, 1-metil-inosina, wiosina, wibutosina, 7-deaza-guanosina, 7-deaza-8-aza-guanosina, 6-tio-guanosina, 6-tio-7-deaza-guanosina, 6-tio-7-deaza-8-aza-guanosina, 7-metil-guanosina, 6-tio-7-metil-guanosina, 7-metilinosina, 6-metoxi-guanosina, 1-metilguanosina, N2-metilguanosina, N2,N2-dimetilguanosina, 8-oxo-guanosina, 7-metil-8-oxo-guanosina, 1-metil-6-tio-guanosina, N2-metil-6-tio-guanosina y N2,N2-dimetil-6-tio-guanosina.

En algunas realizaciones, el nucleótido puede modificarse en la cara del surco mayor y puede incluir el reemplazo de hidrógeno en el C-5 de uracilo con un grupo metilo o un grupo halógeno. En realizaciones específicas, un nucleósido modificado es 5'-O-(1-tiofosfato)-adenosina, 5'-O-(1-tiofosfato)-citidina, 5'-O-(1-tiofosfato)-guanosina, 5'-O-(1-tiofosfato)-uridina o 5'-O-(1-tiofosfato)-pseudouridina.

En realizaciones específicas adicionales, los nucleótidos modificados incluyen modificaciones de nucleósidos seleccionadas de 6-aza-citidina, 2-tio-citidina,  $\alpha$ -tio-citidina, pseudo-iso-citidina, 5-aminoalil-uridina, 5-yodo-uridina, N1-metil-pseudouridina, 5,6-dihidrouridina,  $\alpha$ -tio-uridina, 4-tio-uridina, 6-aza-uridina, 5-hidroxi-uridina, desoxi-timidina, 5-metil-uridina, pirrolo-citidina, inosina,  $\alpha$ -tio-guanosina, 6-metil-guanosina, 5-metil-citidina, 8-oxo-guanosina, 7-deaza-guanosina, N1-metil-adenosina, 2-amino-6-cloro-purina, N6-metil-2-amino-purina, pseudo-iso-citidina, 6-cloro-purina, N6-metil-adenosina,  $\alpha$ -tio-adenosina, 8-azido-adenosina, 7-deaza-adenosina.

Además, previamente se han descrito nucleótidos modificados (véanse, por ejemplo, los documentos WO 2013/052523 y WO 2015/089511).

Caperuza 5': una caperuza 5' es una entidad, normalmente una entidad de nucleótido modificado, que generalmente "ocupa" el extremo 5' de un ARNm maduro y aumenta su estabilidad. Una caperuza 5' puede formarse normalmente mediante un nucleótido modificado (análogo de caperuza), particularmente mediante un derivado de un nucleótido de guanina. Preferiblemente, la caperuza 5' se une al extremo terminal 5' del ARN a través de un enlace 5'-5'-trifosfato. Una caperuza 5' puede metilarse, por ejemplo m7GpppN (por ejemplo m7G(5')ppp(5')G (m7G)), en el que N es el nucleótido 5' terminal del ácido nucleico que porta la caperuza 5', normalmente el extremo 5' de un ARN, m7 es un grupo metilo unido a la posición 7 de la guanina (G) y ppp es el trifosfato. Los ejemplos adicionales de estructuras de caperuza 5' incluyen glicerilo, residuo abásico de desoxilo invertido (resto), 4',5'-metilen-nucleótido, 1-(beta-D-eritrofuranosil)nucleótido, 4'-tio-nucleótido, nucleótido carbocíclico, 1,5-anhidrohexitol-nucleótido, L-nucleótidos, alfa-nucleótido, nucleótido de base modificado, treo-pentofuranosil-nucleótido, 3',4'-sec-nucleótido acíclico, 3,4-dihidroxibutil-nucleótido acíclico, 3,5-dihidroxipentil-nucleótido acíclico, resto de nucleótido invertido en 3'-3', resto abásico invertido en 3'-3', resto de nucleótido invertido en 3'-2', resto abásico invertido en 3'-2', fosfato de 1,4-butanediol, 3'-fosforamidato, hexilfosfato, fosfato, 3'-fosfato de aminohexilo, 3'-fosforotioato, fosforoditioato, o resto de metilfosfonato en puente o no en puente. Estructuras modificadas de la caperuza 5' adicionales que pueden usarse en el contexto de la presente invención con cap1 (metilación adicional de la ribosa en el nucleótido adyacente de m7GpppN), cap2 (metilación adicional de la ribosa del 2º nucleótido aguas abajo de m7GpppN), cap3 (metilación adicional de la ribosa del 3º nucleótido aguas abajo de m7GpppN), cap4 (metilación adicional de la ribosa del 4º nucleótido aguas abajo de m7GpppN), ARCA (análogo de caperuza inverso), ARCA modificado (por ejemplo, ARCA modificado con fosfotioato), inosina, N1-metil-guanosina, 2'-fluoro-guanosina, 7-deaza-guanosina, 8-oxo-guanosina, 2-amino-guanosina, LNA-guanosina y 2-azido-guanosina.

ARN, ARNm: ARN es la abreviatura habitual de ácido ribonucleico. Es una molécula de ácido nucleico, es decir, un polímero que consiste en nucleótidos. Estos nucleótidos suelen ser monómeros de adenosina monofosfato, uridina monofosfato, guanosina monofosfato y citidina monofosfato, que están conectados entre sí a lo largo de la denominada estructura principal. La estructura principal está formada por enlaces fosfodiéster entre el azúcar, es decir, la ribosa, de un primero y resto de fosfato de un segundo monómero adyacente. La sucesión específica de los monómeros se denomina secuencia de ARN. Normalmente, el ARN puede obtenerse mediante la transcripción de una secuencia de ADN, por ejemplo, dentro de una célula. En las células eucariotas, la transcripción se realiza normalmente dentro del núcleo o las mitocondrias. *In vivo*, la transcripción del ADN da como resultado habitualmente el denominado ARN prematuro, que debe procesarse en el denominado ARN mensajero, abreviado habitualmente como ARNm. El procesamiento del ARN prematuro, por ejemplo, en organismos eucariotas, comprende una variedad de modificaciones postranscripcionales diferentes tales como corte y empalme, ocupación del extremo 5', poliadenilación, exportación desde el núcleo o las mitocondrias y similares. La suma de estos procesos también se denomina maduración de ARN. El ARN mensajero maduro proporciona habitualmente la secuencia de nucleótidos que puede traducirse para dar una secuencia de aminoácidos de un péptido o proteína particular. Normalmente, un ARNm maduro comprende una caperuza 5', una 5'-UTR, un marco de lectura abierto, una 3'-UTR y una secuencia de poli(A). Aparte del ARN mensajero, existen varios tipos de ARN no codificantes, que pueden estar implicados en la regulación de la transcripción y/o traducción. Las moléculas de ARN corto pueden sintetizarse mediante métodos químicos, mientras que los ARN largos se producen normalmente mediante reacciones de transcripción *in vitro* que contienen un molde de ADN adecuado con un promotor derivado de bacteriófago, una ARN polimerasa, por ejemplo, ARN polimerasa de bacteriófago SP6, T3 o T7 y ribonucleósidos trifosfatos (NTP).

ADN: ADN es la abreviatura habitual de ácido desoxirribonucleico. Es una molécula de ácido nucleico, es decir, un

polímero que consiste en monómeros de nucleótidos. Estos nucleótidos son habitualmente monómeros de desoxi-adenosina monofosfato, desoxi-timidina monofosfato, desoxi-guanosina monofosfato y desoxi-citidina monofosfato que están, por sí mismos, compuestos por un resto de azúcar (desoxirribosa), un resto de base y un resto de fosfato, y polimerizados por una estructura de la estructura principal característica. La estructura de la estructura principal está formada, normalmente, por enlaces fosfodiéster entre el resto de azúcar del nucleótido, es decir, desoxirribosa, de un primero y un resto de fosfato de un segundo monómero adyacente. El orden específico de los monómeros, es decir, el orden de las bases unidas a la estructura principal de azúcar/fosfato, se denomina secuencia de ADN. El ADN puede ser monocatenario o bicatenario. En la forma bicatenaria, los nucleótidos de la primera cadena se hibridan normalmente con los nucleótidos de la segunda cadena, por ejemplo, por apareamiento de bases A/T y apareamiento de bases G/C.

Secuencia de una molécula de ácido nucleico/secuencia de ácido nucleico: la secuencia de una molécula de ácido nucleico se entiende normalmente como el orden particular e individual, es decir, la sucesión de sus nucleótidos.

Secuencia de moléculas de aminoácido/secuencia de aminoácidos: la secuencia de una proteína o un péptido se entiende normalmente como el orden, es decir, la sucesión de sus aminoácidos.

Identidad de secuencia: dos o más secuencias son idénticas si presentan la misma longitud y orden de nucleótidos o aminoácidos. El porcentaje de identidad describe normalmente el grado en el que dos secuencias son idénticas, es decir, describe normalmente el porcentaje de nucleótidos que corresponden en su posición de secuencia a nucleótidos idénticos de una secuencia de referencia. Para la determinación del grado de identidad, se considera que las secuencias que van a compararse presentan la misma longitud, es decir, la longitud de la secuencia más larga de las secuencias que van a compararse. Esto significa que una primera secuencia que consiste en 8 nucleótidos/aminoácidos es el 80% idéntica a una segunda secuencia que consiste en 10 nucleótidos/aminoácidos que comprenden la primera secuencia. Dicho de otro modo, en el contexto de la presente invención, la identidad de secuencias se refiere preferiblemente al porcentaje de nucleótidos/aminoácidos de una secuencia, que tienen la misma posición en dos o más secuencias que tienen la misma longitud. Los huecos se consideran habitualmente posiciones no idénticas, independientemente de su posición real en una alineación.

Origen de replicación: el término "origen de replicación" se refiere a una secuencia de ADN que puede promover el inicio de la replicación de la molécula de ADN (por ejemplo, ADN genómico, ADN recombinante circular cerrado covalentemente). En el contexto de la invención, el término se refiere especialmente a una secuencia de ADN en una molécula de ADN recombinante circular cerrado covalentemente que puede promover el inicio de la replicación de la molécula de ADN recombinante circular cerrado covalentemente. Habitualmente, el origen de replicación es unidireccional. La replicación de un origen de replicación unidireccional se lleva a cabo en una dirección. El origen de replicación puede tener la secuencia identificada en SEQ ID NO: 6 (derivada del plásmido pUC), la secuencia identificada en SEC ID NO: 7 (derivada del plásmido pBR322), la secuencia identificada en SEQ ID NO: 8 (derivada del plásmido pMB1), la secuencia identificada en SEQ ID NO: 9 (derivada del plásmido ColE1), las secuencias de R6K gamma (SEQ ID NO: 10), R6K beta (SEQ ID NO: 11), R6K alfa (SEQ ID NO: 12) derivadas del plásmido R6K, la secuencia identificada en SEQ ID NO: 13 (derivada del plásmido p15 A), la secuencia identificada en SEQ ID NO: 14 (derivada del plásmido pSC101) o la secuencia identificada en SEQ ID NO: 15 (derivada del fagémido F1). También se abarcan las secuencias que tienen una identidad del 70%, el 80%, el 90%, el 95%, el 98% o el 99% con las secuencias de SEQ ID NO: 6-14.

Vector: el término "vector" se refiere a una molécula de ácido nucleico, preferiblemente a una molécula de ácido nucleico artificial. Un vector en el contexto de la presente invención es adecuado para incorporar o albergar una secuencia de ácido nucleico deseada, tal como una secuencia de ácido nucleico que comprende un marco de lectura abierto. Tales vectores pueden ser vectores de almacenamiento, vectores de expresión, vectores de clonación, vectores de transferencia, etc. Un vector de almacenamiento es un vector que permite el almacenamiento conveniente de una molécula de ácido nucleico, por ejemplo, de una molécula de ARNm. Por tanto, el vector puede comprender una secuencia correspondiente, por ejemplo, a una secuencia de ARNm deseada o una parte de la misma, tal como una secuencia correspondiente al marco de lectura abierto y la 3'-UTR de un ARNm. Puede usarse un vector de expresión para la producción de productos de expresión tales como ARN, por ejemplo, ARNm, o péptidos, polipéptidos o proteínas. Por ejemplo, un vector de expresión puede comprender secuencias necesarias para la transcripción de un tramo de secuencia del vector, tal como una secuencia promotora, por ejemplo, una secuencia promotora de ARN polimerasa. Un vector de clonación es normalmente un vector que contiene un sitio de clonación, que puede usarse para incorporar secuencias de ácido nucleico en el vector. Un vector de clonación puede ser, por ejemplo, un vector plasmídico o un vector bacteriófago. Un vector de transferencia puede ser un vector que sea adecuado para transferir moléculas de ácido nucleico a células u organismos, por ejemplo, vectores virales. Un vector en el contexto de la presente invención puede ser, por ejemplo, un vector de ARN o un vector de ADN. Preferiblemente, un vector es una molécula de ADN. Preferiblemente, un vector en el sentido de la presente solicitud comprende un sitio de clonación, un marcador de selección, tal como un factor de resistencia a antibióticos, y una secuencia adecuada para la multiplicación del vector, tal como un origen de replicación. Preferiblemente, un vector en el contexto de la presente solicitud es un vector plasmídico.

Región no traducida 3' (3'-UTR): generalmente, el término "3'-UTR" se refiere a una parte de la molécula de ácido

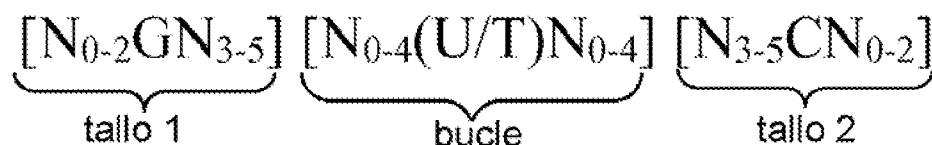
nucleico artificial que está ubicada en 3' (es decir, "aguas abajo") de un marco de lectura abierto y que no se traduce para dar una proteína. Normalmente, una 3'-UTR es la parte de un ARNm que está ubicada entre la región codificante de la proteína (marco de lectura abierto (ORF) o la secuencia codificante (CDS)) y la secuencia de poli(N/A) del ARN(m). En el contexto de la invención, una 3'-UTR de la molécula de ácido nucleico artificial puede comprender más de un elemento de 3'-UTR, que pueden ser de diferente origen, tales como elementos de secuencia derivados de la 3'-UTR de varios genes (no relacionados) que se producen de manera natural. Por consiguiente, el término 3'-UTR también puede comprender elementos, que no están codificados en el molde, a partir de los cuales se transcribe un ARN, pero que se añaden después de la transcripción durante la maduración, por ejemplo, una secuencia de poli(N/A). Una 3'-UTR del ARNm no se traduce en una secuencia de aminoácidos. La secuencia de 3'-UTR está codificada generalmente por el gen, que se transcribe en el ARNm respectivo durante el proceso de expresión génica. En primer lugar, se transcribe la secuencia genómica en ARNm prematuro, que comprende intrones opcionales. El ARNm prematuro se procesa luego adicionalmente para dar ARNm maduro en un proceso de maduración.

Este proceso de maduración comprende las etapas de ocupar el extremo 5', cortar y empalmar el ARNm prematuro para eliminar intrones opcionales y modificaciones del extremo 3', tal como polinucleotidilación/poliadenilación del extremo 3' del ARNm prematuro y escisiones con endonucleasas o exonucleasas opcionales, etc. En el contexto de la presente invención, una 3'-UTR corresponde a la secuencia de un ARNm maduro que está ubicada entre el codón de terminación de la región codificante de la proteína, preferiblemente inmediatamente 3' con respecto al codón de terminación de la región codificante de la proteína, y la secuencia de poli(N/A) del ARNm. El término "corresponde a" significa que la secuencia de 3'-UTR puede ser una secuencia de ARN, tal como en la secuencia de ARNm usada para definir la secuencia de 3'-UTR, o una secuencia de ADN, que corresponde a dicha secuencia de ARN. En el contexto de la presente invención, el término "una 3'-UTR de un gen", tal como "una 3'-UTR de un gen de proteína ribosómico", es la secuencia que corresponde a la 3'-UTR del ARNm maduro derivado de este gen, es decir, el ARNm obtenido por transcripción del gen y maduración del ARNm prematuro. El término "3'-UTR de un gen" abarca la secuencia de ADN y la secuencia de ARN (tanto cadena sentido como antisentido y tanto madura como inmadura) de la 3'-UTR. Tal como se usa en el presente documento, el término "elemento 3'-UTR" se refiere normalmente a un fragmento de una 3'-UTR tal como se define en el presente documento. En particular, el término comprende cualquier elemento de secuencia de ácido nucleico, que se encuentra 3' con respecto al ORF en la molécula de ácido nucleico artificial, preferiblemente el ARNm, según la invención. Por consiguiente, el término cubre, por ejemplo, elementos de secuencia derivados de 3'-UTR de un gen heterólogo, así como elementos tales como una secuencia de poli(A), una secuencia de poli(C) o un tallo-bucle de histona.

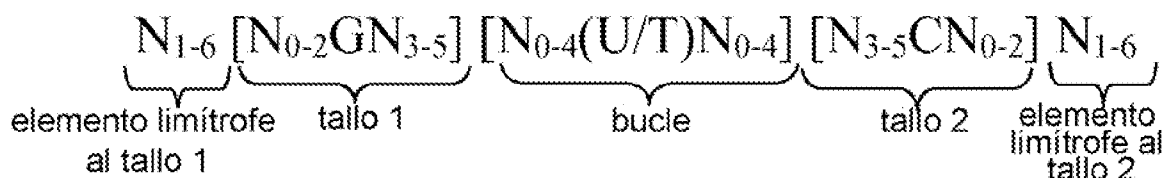
Tallo-bucle de histona: las secuencias de tallo-bucle de histona se seleccionan preferiblemente de secuencias de tallo-bucle de histona tal como se dan a conocer en el documento WO 2012/019780.

Una secuencia de tallo-bucle de histona adecuada para su uso dentro de la presente invención se selecciona preferiblemente de al menos una de las siguientes fórmulas (I) o (II):

**fórmula (I)** (secuencia de tallo-bucle sin elementos limítrofes a los tallos):



**fórmula (II)** (secuencia de tallo-bucle con elementos limítrofes a los tallos):



en las que:

elementos limítrofes  $N_{1-6}$  al tallo 1 o tallo 2

es una secuencia consecutiva de 1 a 6, preferiblemente de 2 a 6, más preferiblemente de 2 a 5, incluso más preferiblemente de 3 a 5, lo más preferiblemente de 4 a 5 ó 5 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G y C, o un análogo de nucleótido del mismo;

tallo 1 [N<sub>0-2</sub>GN<sub>3-5</sub>]

es inversamente complementario o inversamente complementario de manera parcial con el elemento tallo 2, y es una secuencia consecutiva entre de 5 a 7 nucleótidos;

5 en el que N<sub>0-2</sub> es una secuencia consecutiva de 0 a 2, preferiblemente de 0 a 1, más preferiblemente de 1 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo;

10 en el que N<sub>3-5</sub> es una secuencia consecutiva de 3 a 5, preferiblemente de 4 a 5, más preferiblemente de 4 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo, y

15 en el que G es guanosina o un análogo de la misma, y puede reemplazarse opcionalmente por una citidina o un análogo de la misma, siempre que su nucleótido complementario citidina en el tallo 2 se reemplace por guanosina;

secuencia de bucle [N<sub>0-4</sub>(U/T)N<sub>0-4</sub>]

20 está ubicada entre los elementos tallo 1 y tallo 2, y es una secuencia consecutiva de 3 a 5 nucleótidos, más preferiblemente de 4 nucleótidos;

25 en el que cada N<sub>0-4</sub> es independiente de otra secuencia consecutiva de 0 a 4, preferiblemente de 1 a 3, más preferiblemente de 1 a 2 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo; y

en el que U/T representa uridina, u opcionalmente timidina;

tallo 2 [N<sub>3-5</sub>CN<sub>0-2</sub>]

30 es inversamente complementario o inversamente complementario de manera parcial con el elemento tallo 1, y es una secuencia consecutiva entre de 5 y 7 nucleótidos;

35 en el que N<sub>3-5</sub> es una secuencia consecutiva de 3 a 5, preferiblemente de 4 a 5, más preferiblemente de 4 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo;

40 en el que N<sub>0-2</sub> es una secuencia consecutiva de 0 a 2, preferiblemente de 0 a 1, más preferiblemente de 1 N, en el que cada N se selecciona independientemente unos de otros de un nucleótido seleccionado de A, U, T, G o C o un análogo de nucleótido del mismo; y

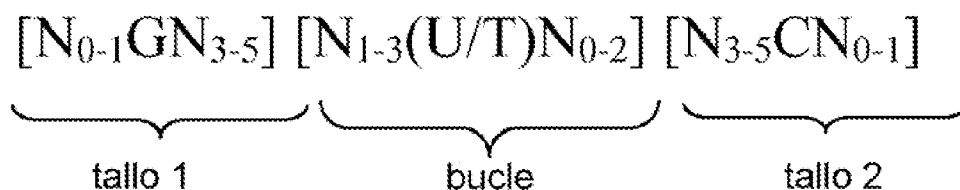
en el que C es citidina o un análogo de la misma, y puede reemplazarse opcionalmente por una guanosina o un análogo de la misma siempre que su nucleósido complementario guanosina en el tallo 1 se reemplace por citidina;

en las que

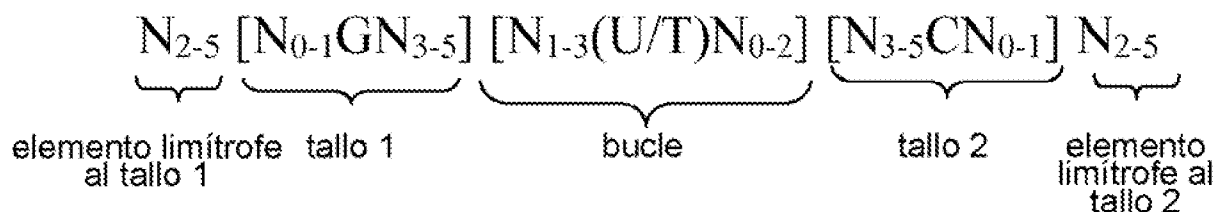
45 el tallo 1 y el tallo 2 pueden realizar apareamiento de bases entre sí formando una secuencia complementaria inversa, en la que el apareamiento de bases puede ocurrir entre el tallo 1 y el tallo 2, por ejemplo, por apareamiento de bases de Watson-Crick de nucleótidos A y U/T o G y C o por apareamiento de bases no de Watson-Crick, por ejemplo, apareamiento de bases titubeo (*wobble*), apareamiento de bases de Watson-Crick inverso, apareamiento de bases de Hoogsteen, apareamiento de bases de Hoogsteen inverso o pueden realizar apareamiento de bases entre sí formando una secuencia complementaria parcialmente inversa, en la que puede producirse un apareamiento de bases incompleto entre el tallo 1 y el tallo 2, sobre la base de que una o más bases en un tallo no tienen una base complementaria en la secuencia complementaria inversa del otro tallo.

55 Según una realización adicional preferida, la secuencia de tallo-bucle de histona puede seleccionarse según al menos una de las siguientes fórmulas específicas (Ia) o (IIa):

fórmula (Ia) (secuencia de tallo-bucle sin elementos limítrofes a los tallos):



fórmula (IIa) (secuencia de tallo-bucle con elementos limítrofes a los tallos):

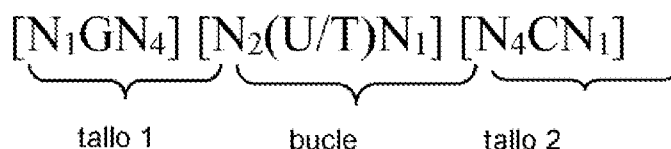


5 en las que:

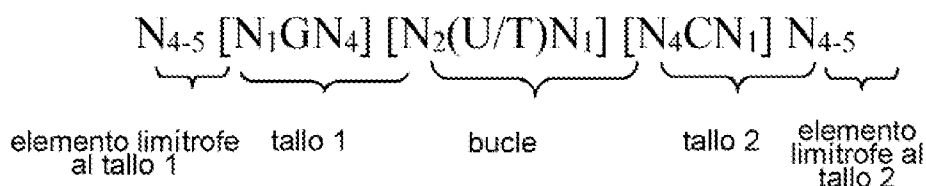
N, C, G, T y U son tal como se definieron anteriormente.

10 Según una realización particularmente más preferida adicional del primer aspecto, la secuencia de la molécula de ácido nucleico artificial puede comprender al menos una secuencia de tallo-bucle de histona según al menos una de las siguientes fórmulas específicas (Ib) o (IIb):

fórmula (Ib) (secuencia de tallo-bucle sin elementos limítrofes a los tallos):



fórmula (IIb) (secuencia de tallo-bucle con elementos limítrofes a los tallos):



15 en las que:

N, C, G, T y U son tal como se definieron anteriormente.

20 Una secuencia de tallo-bucle de histona preferida en particular es la secuencia según SEQ ID NO: 16: CAAAGGCTCTTTTCAGAGCCACCA o más preferiblemente la secuencia de ARN correspondiente según SEQ ID NO: 17: caaaggcucuuuucagagccacca

25 Región no traducida 5' (5'-UTR): una 5'-UTR se entiende normalmente como una sección particular de ARN mensajero (ARNm). Está ubicada en 5' del marco de lectura abierto del ARNm. Normalmente, la 5'-UTR comienza con el sitio de inicio transcripcional y termina un nucleótido antes del codón de inicio del marco de lectura abierto. La 5'-UTR puede comprender elementos para controlar la expresión génica, que también se denominan elementos reguladores. Tales elementos reguladores pueden ser, por ejemplo, sitios de unión ribosómica. La 5'-UTR puede modificarse de manera postranscripcional, por ejemplo, mediante la adición de una caperuza 5'. En el contexto de la

presente invención, una 5'-UTR corresponde a la secuencia de un ARNm maduro, que está ubicada entre la caperuza 5' y el codón de inicio.

Preferiblemente, la 5'-UTR corresponde a la secuencia, que se extiende desde un nucleótido ubicado en 3' con respecto a la caperuza 5', preferiblemente desde el nucleótido ubicado inmediatamente en 3' con respecto a la caperuza 5', hasta un nucleótido ubicado en 5' con respecto al codón de inicio de la región codificante de la proteína, preferiblemente hasta el nucleótido ubicado inmediatamente en 5' con respecto al codón de inicio de la región codificante de la proteína. El nucleótido ubicado inmediatamente en 3' con respecto a la caperuza 5' de un ARNm maduro corresponde normalmente al sitio de inicio transcripcional. El término "corresponde a" significa que la secuencia 5'-UTR puede ser una secuencia de ARN, tal como en la secuencia de ARNm usada para definir la secuencia 5'-UTR, o una secuencia de ADN, que corresponde a tal secuencia de ARN. En el contexto de la presente invención, el término "una 5'-UTR de un gen" es la secuencia, que corresponde a la 5'-UTR del ARNm maduro derivado de este gen, es decir, el mRNA obtenido por transcripción del gen y maduración del ARNm prematuro. El término "5'-UTR de un gen" abarca la secuencia de ADN y la secuencia de ARN de la 5'-UTR.

Sitio de clonación: un sitio de clonación se entiende normalmente como un segmento de una molécula de ácido nucleico que es adecuado para la inserción de una molécula de ácido nucleico, por ejemplo, una molécula de secuencia de ácido nucleico que comprende un marco de lectura abierto. La molécula de ácido nucleico puede insertarse mediante cualquier método de biología molecular conocido por el experto en la técnica, por ejemplo, por restricción y ligamiento. Un sitio de clonación comprende normalmente uno o más sitios de reconocimiento de enzimas de restricción (sitios de restricción). Estos uno o más sitios de restricción pueden ser reconocidos por enzimas de restricción que escinden el ADN en estos sitios. Un sitio de clonación que comprende más de un sitio de restricción también puede denominarse sitio de clonación múltiple (MCS) o poliligador.

Marco de lectura abierto: un marco de lectura abierto (ORF) en el contexto de la invención puede ser normalmente una secuencia de varios tripletes de nucleótidos que pueden traducirse para dar un péptido o una proteína. Un marco de lectura abierto contiene preferiblemente un codón de inicio, es decir, una combinación de tres nucleótidos posteriores que codifican habitualmente para el aminoácido metionina (ATG), en su extremo 5' y una región posterior que presenta habitualmente una longitud que es un múltiplo de 3 nucleótidos. Un ORF está preferiblemente terminado por un codón de terminación (por ejemplo, TAA, TAG, TGA). Normalmente, este es el único codón de terminación del marco de lectura abierto. Por tanto, un marco de lectura abierto en el contexto de la presente invención es preferiblemente una secuencia de nucleótidos que consiste en varios nucleótidos que pueden dividirse entre tres, que comienza con un codón de inicio (por ejemplo, ATG) y que preferiblemente termina con un codón de terminación (por ejemplo, TAA, TGA o TAG). El marco de lectura abierto puede aislarse o puede incorporarse en una secuencia de ácido nucleico más larga, por ejemplo, en un vector o un ARNm. Un marco de lectura abierto también puede denominarse "región codificante de proteínas".

Transcripción *in vitro*: el término "transcripción *in vitro*" se refiere a un proceso en el que el ARN se sintetiza en un sistema libre de células (*in vitro*). El ADN, particularmente el ADN plasmídico, se usa como molde para la generación de transcritos de ARN. El ARN puede obtenerse mediante transcripción *in vitro* dependiente de ADN de un molde de ADN apropiado, que según la presente invención es preferiblemente un molde de ADN plasmídico linealizado. El promotor para controlar la transcripción *in vitro* puede ser cualquier promotor de cualquier ARN polimerasa dependiente de ADN. Los ejemplos particulares de ARN polimerasas dependientes de ADN son las ARN polimerasas T7, T3 y SP6. Un molde de ADN para la transcripción *in vitro* de ARN puede obtenerse clonando un ácido nucleico, en particular el ADNc correspondiente al ARN respectivo que va a transcribirse *in vitro*, e introduciéndolo en un vector apropiado para la transcripción *in vitro*, por ejemplo en ADN plasmídico. En el sentido de la presente invención, el molde de ADN se linealiza mediante el método de la invención, antes de que se transcriba *in vitro*. El ADNc puede obtenerse mediante transcripción inversa de ARNm o síntesis química. Además, el molde de ADN para la síntesis de ARN *in vitro* también puede obtenerse mediante síntesis génica.

Los métodos para la transcripción *in vitro* son conocidos en la técnica (Geall *et al.* (2013) Semin. Immunol. 25(2): 152-159; Brunelle *et al.* (2013) Methods Enzymol. 530:101-14). Los reactivos usados en dicho método incluyen normalmente:

- 1) un molde de ADN linealizado (tal como se definió anteriormente) con una secuencia promotora que tiene una alta afinidad de unión por su respectiva ARN polimerasa tal como ARN polimerasas codificadas por bacteriófagos
- 2) ribonucleósidos trifosfatos (NTP) para las cuatro bases (adenina, citosina, guanina y uracilo);
- 3) opcionalmente, un análogo de caperuza tal como se definió anteriormente (por ejemplo m7G(5')ppp(5')G (m7G));
- 4) una ARN polimerasa dependiente de ADN que puede unirse a la secuencia promotora dentro del molde de ADN linealizado (por ejemplo, ARN polimerasa T7, T3 o SP6);
- 5) opcionalmente, un inhibidor de ribonucleasa (ARNasa) para inactivar cualquier ARNasa contaminante;

6) opcionalmente, una pirofosfatasa para degradar el pirofosfato, que puede inhibir la transcripción;

7)  $\text{MgCl}_2$ , que suministra iones  $\text{Mg}^{2+}$  como cofactor para la polimerasa;

- 5 8) un tampón para mantener un valor de pH adecuado, que también puede contener antioxidantes (por ejemplo, DTT) y poliaminas tales como espermidina en concentraciones óptimas.

Según una realización preferida, el tampón (de transcripción) se selecciona del grupo que consiste en ácido 4-(2-hidroxietil)-1-piperazinetanosulfónico (HEPES) y tris(hidroximetil)aminometano (Tris). Preferiblemente, el  
10 tampón se usa a una concentración de desde 10 hasta 100 mM, de 10 a 75 mM, de 10 a 50 mM, de 10 a 40 mM, de 10 a 30 mM o de 10 a 20 mM. El valor de pH del tampón puede ajustarse con, por ejemplo, NaOH, KOH o HCl. Preferiblemente, el tampón tiene un valor de pH desde 6 hasta 8,5, desde 6,5 hasta 8,0, desde 7,0 hasta 7,5, incluso más preferido 7,5. Lo más preferido es un tampón seleccionado del grupo que consiste en HEPES/KOH 80 mM, pH 7,5 y Tris/HCl 40 mM, pH 7,5.

Según una realización preferida de la invención, la ARN polimerasa se selecciona del grupo que consiste en ARN polimerasa T3, T7 y SP6. Preferiblemente, la concentración de la ARN polimerasa es de desde aproximadamente 1 hasta 100 nM, de 1 a 90 nM, de 1 a 80 nM, de 1 a 70 nM, de 1 a 60 nM, de 1 a 50 nM, de 1 a 40 nM, de 1 a 30 nM, de 1 a 20 nM, o aproximadamente de 1 a 10 nM. Incluso más preferido, la concentración de la ARN polimerasa es  
20 de desde aproximadamente 10 hasta 50 nM, de 20 a 50 nM, o de 30 a 50 nM. Lo más preferido es una concentración de ARN polimerasa de aproximadamente 40 nM. El experto en la técnica entenderá que la concentración del molde de ADN influye en la elección de la concentración de ARN polimerasa. Por tanto, en realizaciones específicas, la concentración de la ARN polimerasa es de entre 1 y 1000 U/ $\mu\text{g}$  de ADN de molde, preferiblemente entre 10 y 100 U/ $\mu\text{g}$  de ADN, particularmente si se usa ADN plasmídico como ADN de molde.

Según una realización preferida de la invención, la concentración del molde de ADN lineal está en un intervalo de desde aproximadamente 1 hasta 50 nM, de 1 a 40 nM, de 1 a 30 nM, de 1 a 20 nM, o aproximadamente de 1 a 10 nM. Incluso más preferido, la concentración del molde de ADN es de desde aproximadamente 10 hasta 30 nM. Lo más preferido, la concentración del molde de ADN es de aproximadamente 20 nM. En el caso de usarse ADN  
30 plasmídico como molde de ADN, la concentración del molde de ADN es preferiblemente entre 1 y 100  $\mu\text{g}/\text{ml}$ , particularmente en una concentración de aproximadamente 50  $\mu\text{g}/\text{ml}$ .

Según una realización preferida de la invención, la transcripción *in vitro* se realiza en presencia de pirofosfatasa. Preferiblemente, la concentración de la pirofosfatasa es de desde aproximadamente 1 hasta 20 unidades/ml, de 1 a 15 unidades/ml, de 1 a 10 unidades/ml, de 1 a 5 unidades/ml, o de 1 a 2,5 unidades/ml. Incluso más preferido, la  
35 concentración de la pirofosfatasa es de aproximadamente 5 unidades/ml.

Según una realización preferida de la invención, la mezcla de reacción de la transcripción *in vitro* comprende iones  $\text{Mg}^{2+}$ . Preferiblemente, los iones  $\text{Mg}^{2+}$  se proporcionan en forma de  $\text{MgCl}_2$  o  $\text{Mg}(\text{OAc})_2$ . Preferiblemente, la  
40 concentración inicial de  $\text{Mg}^{2+}$  libre es de desde aproximadamente 1 hasta 100 mM, de 1 a 75 mM, de 1 a 50 mM, de 1 a 25 mM, o de 1 a 10 mM. Incluso más preferido, la concentración inicial de  $\text{Mg}^{2+}$  libre es de desde aproximadamente 10 hasta 30 mM o aproximadamente de 15 a 25 mM. Lo más preferido es una concentración inicial de  $\text{Mg}^{2+}$  libre de aproximadamente 24 mM. El experto en la técnica entenderá que la concentración de NTP total inicial influye en la elección de la concentración de  $\text{Mg}^{2+}$

Según una realización preferida de la invención, la mezcla de reacción de transcripción *in vitro* comprende un agente reductor (antioxidante) para mantener la ARN polimerasa en su estado activo. Preferiblemente, el agente reductor se  
50 selecciona del grupo que consiste en ditioneitol (DTT), ditioneitol (DTE), tris(2-carboxietil)fosfina (TCEP) y  $\beta$ -mercaptoetanol. Preferiblemente, la concentración del reactivo reductor es de desde aproximadamente 1 hasta 50 mM, de 1 a 40 mM, de 1 a 30 mM, o de 1 a 20 mM, o de 1 a 10 mM. Incluso más preferido, la concentración del reactivo reductor es de desde 10 hasta 50 mM o de 20 a 40 mM. Lo más preferido es una concentración de DTT de 40 mM.

Según una realización preferida de la invención, la mezcla de reacción de transcripción *in vitro* comprende una poliamina. Preferiblemente, la poliamina se selecciona del grupo que consiste en espermina y espermidina. Preferiblemente, la concentración de la poliamina es de desde aproximadamente 1 hasta 25 mM, de 1 a 20 mM, de 1 a 15 mM, de 1 a 10 mM, de 1 a 5 mM o aproximadamente de 1 a 2,5 mM. Incluso más preferido, la concentración  
55 de poliamina es de aproximadamente 2 mM. Lo más preferido es una concentración de espermidina de 2 mM.

Según una realización preferida de la invención, la mezcla de reacción de transcripción *in vitro* comprende un inhibidor de ribonucleasa. Preferiblemente, la concentración del inhibidor de ribonucleasa es de desde  
60 aproximadamente 1 hasta 500 unidades/ml, de 1 a 400 unidades/ml, de 1 a 300 unidades/ml, de 1 a 200 unidades/ml o de 1 a 100 unidades/ml. Incluso más preferido, la concentración del inhibidor de ribonucleasa es de aproximadamente 200 unidades/ml.

Según una realización preferida de la invención, la concentración de NTP total en la mezcla de reacción de transcripción *in vitro* es de entre 1 y 100 mM, preferiblemente entre 10 y 50 mM, y lo más preferiblemente entre 10 y 20 mM.

- 5 Según la invención, el término concentración de nucleótidos total significa la concentración de NTP total, por ejemplo, la suma de las concentraciones de ATP, GTP, CTP, UTP y/o análogo de caperuza presentes inicialmente en la transcripción *in vitro* cuando los diversos componentes de la reacción se han ensamblado en el volumen final para llevar a cabo la reacción de transcripción *in vitro*. Naturalmente, a medida que avanza la reacción, los nucleótidos se incorporarán a la molécula de ARN y, en consecuencia, la concentración de nucleótidos total se
- 10 reducirá progresivamente desde su valor inicial.

En este contexto, se prefiere particularmente que los nucleótidos individuales se proporcionen en una concentración de entre 0,1 y 10 mM, preferiblemente entre 1 y 5 mM y lo más preferiblemente en una concentración de 4 mM.

- 15 En el caso de que deba generarse una caperuza 5' tal como se definió anteriormente en el extremo 5' del ARN, la mezcla de reacción de transcripción *in vitro* comprende además un análogo de caperuza. En este contexto, la concentración de GTP se reduce preferiblemente en comparación con los otros nucleótidos (ATP, CTP y UTP). Preferiblemente, el análogo de caperuza se añade con una concentración inicial en el intervalo de aproximadamente 1 a 20 mM, de 1 a 17,5 mM, de 1 a 15 mM, de 1 a 12,5 mM, de 1 a 10 mM, de 1 a 7,5 mM. Lo más preferiblemente,
- 20 el análogo de caperuza se añade en una concentración de 5,8 mM y la concentración de GTP se reduce a una concentración de 1,45 mM mientras que ATP, CTP y UTP están comprendidos en la reacción en una concentración de 4 mM cada uno.

- Los ribonucleósidos trifosfatos (NTP) GTP, ATP, CTP y UTP o análogos de los mismos pueden proporcionarse con un catión monovalente o divalente como contraión. Preferiblemente, el catión monovalente se selecciona del grupo que consiste en Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, NH<sub>4</sub><sup>+</sup> o tris(hidroximetil)-aminometano (Tris). Preferiblemente, el catión divalente se selecciona del grupo que consiste en Mg<sup>2+</sup>, Ba<sup>2+</sup> y Mn<sup>2+</sup>.
- 25

- Según una realización preferida de la invención, una parte o la totalidad de al menos un ribonucleósido trifosfato en la mezcla de reacción de transcripción *in vitro* se reemplaza con un nucleósido trifosfato modificado (tal como se define en el presente documento). En una realización preferida de la invención, dicho nucleósido trifosfato modificado se selecciona del grupo que consiste en pseudouridina-5'-trifosfato, 1-metilpseudouridina-5'-trifosfato, 2-tiouridina-5'-trifosfato, 4-tiouridina-5'-trifosfato y 5-metilcitidina-5'-trifosfato.
- 30

- Después de que se ha producido la transcripción *in vitro*, el producto de ARN se somete a métodos de purificación. En este contexto, puede usarse cualquier método de purificación (por ejemplo, digestión con molde de ADN, extracción con fenol-cloroformo, precipitación con LiCl, HPLC, etc.).
- 35

- ADN plasmídico: el término "ADN plasmídico" se refiere a una molécula de ácido nucleico circular, preferiblemente a una molécula de ácido nucleico artificial. Un ADN plasmídico en el contexto de la presente invención es adecuado para incorporar o albergar una secuencia de ácido nucleico deseada, tal como una secuencia de ácido nucleico que comprende una secuencia codificante de ARN y/o un marco de lectura abierto que codifica al menos para un péptido o polipéptido. Tales constructos de ADN plasmídico pueden ser vectores de almacenamiento, vectores de expresión, vectores de clonación, vectores de transferencia, etc. Un vector de almacenamiento es un vector que permite el almacenamiento conveniente de una molécula de ácido nucleico, por ejemplo, de una molécula de ARN. Por tanto, el ADN plasmídico puede comprender una secuencia que corresponde a (que codifica para), por ejemplo, una secuencia de ARN diana o una parte de la misma, tal como una secuencia que corresponde al marco de lectura abierto y la 5' y/o 3' UTR de un ARNm. Puede usarse un vector de expresión para la producción de productos de expresión tales como ARN, por ejemplo, ARNm en un proceso denominado transcripción *in vitro*. Por ejemplo, un vector de expresión puede comprender secuencias necesarias para la transcripción *in vitro* de un tramo de secuencia del vector, tal como una secuencia promotora, por ejemplo, una secuencia promotora de ARN, preferiblemente secuencias promotoras de ARN T3, T7 o SP6. Un vector de clonación es normalmente un vector que contiene un sitio de clonación, que puede usarse para incorporar secuencias de ácido nucleico (inserto) en el vector. Un vector de clonación puede ser, por ejemplo, un vector plasmídico o un vector bacteriófago. Un vector de transferencia puede ser un vector que sea adecuado para transferir moléculas de ácido nucleico a células u organismos, por ejemplo, vectores virales. Preferiblemente, un vector de ADN plasmídico en el sentido de la presente invención comprende un sitio de clonación múltiple, una secuencia promotora de ARN, opcionalmente un marcador de selección, tal como un factor de resistencia a antibióticos, y una secuencia adecuada para la multiplicación del vector, tal como un origen de replicación. El ADN, que forma el molde del ARN que va a transcribirse *in vitro*, puede prepararse mediante proliferación fermentativa y posterior aislamiento como parte de un plásmido que puede replicarse en bacterias. Particularmente preferidos en el contexto de la presente invención son los vectores de ADN plasmídico, o vectores de expresión, que comprenden promotores de ARN polimerasas dependientes de ADN tales como T3, T7 y Sp6. Como estructura principal del plásmido, se prefieren particularmente pUC19 y pBR322.
- 40
- 45
- 50
- 55
- 60
- 65

Antígeno: en el contexto de la presente invención, "antígeno" se refiere normalmente a una sustancia, que puede ser

reconocida por el sistema inmunitario, preferiblemente por el sistema inmunitario adaptativo, y puede desencadenar una respuesta inmunitaria específica de antígeno, por ejemplo, mediante la formación de anticuerpos y/o células T específicas de antígeno como parte de una respuesta inmunitaria adaptativa. Normalmente, un antígeno puede ser o puede comprender un péptido o una proteína que puede presentar el MHC a las células T y comprende al menos un epítipo. En particular, el antígeno puede ser un antígeno patógeno o un antígeno tumoral.

Antígenos patógenos: el ARNm puede codificar para una proteína o un péptido, que comprende un antígeno patógeno o un fragmento, variante o derivado del mismo. Tales antígenos patógenos se derivan de organismos patógenos, en particular organismos patógenos bacterianos, virales o protozoológicos (multicelulares), que provocan una reacción inmunológica en un sujeto, en particular un sujeto mamífero, más particularmente un ser humano. Más específicamente, los antígenos patógenos son preferiblemente antígenos de superficie, por ejemplo, proteínas (o fragmentos de proteínas, por ejemplo, la porción exterior de un antígeno de superficie) ubicadas en la superficie del virus o del organismo bacteriano o protozoológico.

Los antígenos patógenos son antígenos peptídicos o proteicos derivados preferiblemente de un patógeno asociado con una enfermedad infecciosa, que se seleccionan preferiblemente de antígenos derivados de los patógenos *Acinetobacter baumannii*, *Anaplasma* genus, *Anaplasma phagocytophilum*, *Ancylostoma braziliense*, *Ancylostoma duodenale*, *Arcanobacterium haemolyticum*, *Ascaris lumbricoides*, género *Aspergillus*, *Astroviridae*, género *Babesia*, *Bacillus anthracis*, *Bacillus cereus*, *Bartonella henselae*, virus BK, *Blastocystis hominis*, *Blastomyces dermatitidis*, *Bordetella pertussis*, *Borrelia burgdorferi*, género *Borrelia*, *Borrelia* spp., género *Brucella*, *Brugia malayi*, familia *Bunyaviridae*, *Burkholderia cepacia* y otras especies de *Burkholderia*, *Burkholderia mallei*, *Burkholderia pseudomallei*, familia *Caliciviridae*, género *Campylobacter*, *Candida albicans*, *Candida* spp., *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, prión de CJD, *Clonorchis sinensis*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium* spp., *Clostridium tetani*, *Coccidioides* spp., coronavirus, *Corynebacterium diphtheriae*, *Coxiella burnetii*, virus de la fiebre hemorrágica de Crimea-Congo, *Cryptococcus neoformans*, género *Cryptosporidium*, citomegalovirus (CMV), virus del dengue (DEN-1, DEN-2, DEN-3 y DEN-4), *Dientamoeba fragilis*, virus del Ébola (EBOV), género *Echinococcus*, *Ehrlichia chaffeensis*, *Ehrlichia ewingii*, género *Ehrlichia*, *Entamoeba histolytica*, género *Enterococcus*, género *Enterovirus*, principalmente virus de Cocksackie A y enterovirus 71 (E71), *Epidermophyton* spp., virus de Epstein-Barr (VEB), *Escherichia coli* O157: H7, O111 y O104: H4, *Fasciola hepatica* y *Fasciola gigantica*, prión FFI, superfamilia *Filarioidea*, Flavivirus, *Francisella tularensis*, género *Fusobacterium*, *Geotrichum candidum*, *Giardia intestinalis*, *Gnathostoma* spp., prión GSS, virus Guanarito, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, henipavirus (virus Hendra, virus Nipah), virus de la hepatitis A, virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), virus de la hepatitis D, virus de la hepatitis E, virus del herpes simple 1 y 2 (VHS-1 y VHS-2), *Histoplasma capsulatum*, VIH (virus de la inmunodeficiencia humana), *Hortaea werneckii*, Bocavirus humano (BoVh), herpesvirus humano 6 (HHV-6) y herpesvirus humano 7 (HHV-7), metapneumovirus humano (MPVh), virus del papiloma humano (VPH), virus de la parainfluenza humana (VPIH), virus de la encefalitis japonesa, virus JC, virus Junin, *Kingella kingae*, *Klebsiella granulomatis*, prión Kuru, virus de Lassa, *Legionella pneumophila*, género *Leishmania*, género *Leptospira*, *Listeria monocytogenes*, virus de la coriomeningitis linfocítica (VCML), virus Machupo, *Malassezia* spp., virus de Marburgo, virus del sarampión, *Metagonimus yokagawai*, filo de microsporidios, virus del molusco contagioso (VMC), virus de las paperas, *Mycobacterium leprae* y *Mycobacterium lepromatosis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Mycoplasma pneumoniae*, *Naegleria fowleri*, *Necator americanus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Nocardia asteroides*, *Nocardia* spp, *Onchocerca volvulus*, *Orientia tsutsugamushi*, familia *Orthomyxoviridae* (gripe), *Paracoccidioides brasiliensis*, *Paragonimus* spp, *Paragonimus westermani*, parvovirus B19, género *Pasteurella*, género *Plasmodium*, *Pneumocystis jirovecii*, virus de la poliomielitis, virus de la rabia, virus respiratorio sincitial (VRS), rinovirus, *Rickettsia akari*, género *Rickettsia*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia typhi*, virus de la fiebre del valle del Rift, rotavirus, virus de la rubéola, virus de Sabiá, género *Salmonella*, *Sarcoptes scabiei*, coronavirus SARS, género *Schistosoma*, género *Shigella*, virus Sin Nombre, Hantavirus, *Sporothrix schenckii*, género *Staphylococcus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Strongyloides stercoralis*, género *Taenia*, *Taenia solium*, virus de la encefalitis transmitida por garrapatas (VETG), *Toxocara canis* o *Toxocara cati*, *Toxoplasma gondii*, *Treponema pallidum*, *Trichinella spiralis*, *Trichomonas vaginalis*, *Trichophyton* spp., *Trichuris trichiura*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Ureaplasma urealyticum*, virus de la varicela zoster (VVZ), viruela mayor o viruela menor, prion vCJD, virus de la encefalitis equina venezolana, *Vibrio cholerae*, virus del Nilo Occidental, virus de la encefalitis equina occidental, *Wuchereria bancrofti*, virus de la fiebre amarilla, *Yersinia enterocolitica*, *Yersinia pestis* y *Yersinia pseudotuberculosis*.

En este contexto, se prefieren particularmente los antígenos de los patógenos seleccionados de virus de la gripe, virus respiratorio sincitial (VRS), virus del herpes simple (VHS), virus del papiloma humano (VPH), virus de la inmunodeficiencia humana (VIH), *Plasmodium*, *Staphylococcus aureus*, virus del dengue, *Chlamydia trachomatis*, citomegalovirus (CMV), virus de la hepatitis B (VHB), *Mycobacterium tuberculosis*, virus de la rabia y virus de la fiebre amarilla.

En una realización preferida, el ARNm codifica para una proteína o un péptido del virus de la rabia o un fragmento antigénico del mismo. Preferiblemente, el ARNm codifica para una proteína o péptido antigénico seleccionado del grupo que consiste en glicoproteína G (RAV-G), nucleoproteína N (RAV-N), fosfoproteína P (RAV-P), proteína de matriz M (RAV-M) o ARN polimerasa L (RAV-L) del virus de la rabia, o un fragmento, variante o derivado del mismo.

En otra realización preferida, el ARNm según la invención codifica para una proteína o un péptido del virus respiratorio sincitial (VRS) o un fragmento antigénico del mismo. Preferiblemente, el ARNm según la invención codifica para una proteína o un péptido antigénico seleccionado del grupo que consiste en la proteína de fusión F, la glicoproteína G, la proteína hidrófoba corta SH, la proteína de matriz M, la nucleoproteína N, la polimerasa grande L, la proteína M2-1, la proteína M2-2, la fosfoproteína P, la proteína no estructural NS1 o la proteína no estructural NS2 del virus respiratorio sincitial (VRS), o un fragmento, variante o derivado del mismo.

Antígenos tumorales: en una realización adicional, el ARNm según la presente invención puede codificar para una proteína o un péptido, que comprende un péptido o proteína que comprende un antígeno tumoral, un fragmento, variante o derivado de dicho antígeno tumoral, preferiblemente, en el que el antígeno tumoral es un antígeno específico de melanocito, un antígeno de cáncer de testículo o un antígeno específico de tumor, preferiblemente un antígeno CT-X, un antígeno CT no en X, un componente de unión para un antígeno CT-X o un componente de unión para un antígeno CT no en X o un antígeno específico de tumor, más preferiblemente un antígeno CT-X, un componente de unión para un antígeno CT no en X o un antígeno específico de tumor o un fragmento, variante o derivado de dicho antígeno tumoral; y en el que cada una de las secuencias de ácido nucleico codifica para un péptido o una proteína diferente; y en el que al menos una de las secuencias de ácido nucleico codifica para 5T4, 707-AP, 9D7, AFP, AlbZIP, HPG1, alfa-5-beta-1-integrina, alfa-5-beta-6-integrina, alfa-actinina-4/m, alfa-metilacil-coenzima A racemasa, ART-4, ARTC1/m, B7H4, BAGE-1, BCL-2, bcr/abl, beta-catenina/m, BING-4, BRCA1/m, BRCA2/m, CA 15-3/CA 27-29, CA 19-9, CA72-4, CA125, calreticulina, CAMEL, CASP-8/m, cathepsina B, cathepsina L, CD19, CD20, CD22, CD25, CDE30, CD33, CD4, CD52, CD55, CD56, CD80, CDC27/m, CDK4/m, CDKN2A/m, CEA, CLCA2, CML28, CML66, COA-1/m, proteína similar a la coactosina, collage XXIII, COX-2, CT-9/BRD6, Cten, ciclina B1, ciclina D1, cyp-B, CYPB1, DAM-10, DAM-6, DEK-CAN, EFTUD2/m, EGFR, ELF2/m, EMMPRIN, EpCam, EphA2, EphA3, ErbB3, ETV6-AML1, EZH2, FGF-5, FN, Frau-1, G250, GAGE-1, GAGE-2, GAGE-3, GAGE-4, GAGE-5, GAGE-6, GAGE7b, GAGE-8, GDEP, GnT-V, gp100, GPC3, GPNMB/m, HAGE, HAST-2, hepsina, Her2/neu, HERV-K-MEL, HLA-A\*0201-R17I, HLA-A11/m, HLA-A2/m, HNE, homeobox NKX3.1, HOM-TES-14/SCP-1, HOM-TES-85, HPV-E6, HPV-E7, HSP70-2M, HST-2, hTERT, iCE, IGF-1R, IL-13Ra2, IL-2R, IL-5, receptor de laminina inmaduro, calicreína-2, calicreína-4, Ki67, KIAA0205, KIAA0205/m, KK-LC-1, K-Ras/m, LAGE-A1, LDLR-FUT, MAGE-A1, MAGE-A2, MAGE-A3, MAGE-A4, MAGE-A6, MAGE-A9, MAGE-A10, MAGE-A12, MAGE-B1, MAGE-B2, MAGE-B3, MAGE-B4, MAGE-B5, MAGE-B6, MAGE-B10, MAGE-B16, MAGE-B17, MAGE-C1, MAGE-C2, MAGE-C3, MAGE-D1, MAGE-D2, MAGE-D4, MAGE-E1, MAGE-E2, MAGE-F1, MAGE-HI, MAGEL2, mamaglobina A, MART-1/melan-A, MART-2, MART-2/m, proteína de la matriz 22, MC1R, M-CSF, ME1/m, mesotelina, MG50/PXDN, MMP11, antígeno MN/CA IX, MRP-3, MUC-1, MUC-2, MUM-1/m, MUM-2/m, MUM-3/m, miosina clase I/m, NA88-A, N-acetilglucosaminiltransferasa-V, Neo-PAP, Neo-PAP/m, NFYC/m, NGEP, NMP22, NPM/ALK, N-Ras/m, NSE, NY-ESO-1, NY-ESO-B, OA1, OFA-iLRP, OGT, OGT/m, OS-9, OS-9/m, osteocalcina, osteopontina, p15, bcr-abl menor de p190, p53, p53/m, PAGE-4, PAI-1, PAI-2, PAP, PART-1, PATE, PDEF, Pim-1-cinasa, Pin-1, Pml/PARalfa, POTE, PRAME, PRDX5/m, prosteína, proteinasa-3, PSA, PSCA, PSGR, PSM, PSMA, PTPRK/m, RAGE-1, RBAF600/m, RHAMM/CD168, RU1, RU2, S-100, SAGE, SART-1, SART-2, SART-3, SCC, SIRT2/m, Sp17, SSX-1, SSX-2/HOM-MEL-40, SSX-4, STAMP-1, STEAP-1, survivina, survivina 2B, SYT-SSX-1, SYT-SSX-2, TA-90, TAG-72, TARP, TEL-AML1, TGFbeta, TGFbetaRII, TGM-4, TPI/m, TRAG-3, TRG, TRP-1, TRP-2/6b, TRP/INT2, TRP-p8, tirosinasa, UPA, VEGFR1, VEGFR-2/FLK-1, WT1 y un idiotipo de inmunoglobulina de un glóbulo sanguíneo linfóide o un idiotipo de receptor de células T de un glóbulo sanguíneo linfóide, o un fragmento, variante o derivado de dicho antígeno tumoral; preferiblemente survivina o un homólogo de la misma, un antígeno de la familia MAGE o un componente de unión del mismo o un fragmento, variante o derivado de dicho antígeno tumoral.

Son particularmente preferidos en este contexto los antígenos tumorales NY-ESO-1, 5T4, MAGE-C1, MAGE-C2, survivina, Muc-1, PSA, PSMA, PSCA, STEAP y PAP.

Las siguientes combinaciones de antígenos son particularmente preferidas:

- Muc-1, PSA, PSMA, PSCA y STEAP
- Muc-1, PSA, PSMA, PSCA y PAP
- Muc-1, PSA, PSMA, STEAP y PAP
- Muc-1, PSA, PSCA, STEAP y PAP
- Muc-1, PSMA, PSCA, STEAP y PAP
- PSA, PSMA, PSCA, STEAP y PAP
- Muc-1, PSA, PSMA, PSCA, STEAP y PAP

Además, las siguientes combinaciones de antígenos son particularmente preferidas:

- NY-ESO-1, 5T4, MAGE-C1, MAGE-C2 y survivina
- 5 • NY-ESO-1, 5T4, MAGE-C1, MAGE-C2 y Muc-1
- NY-ESO-1, 5T4, MAGE-C1, survivina y Muc-1
- 10 • NY-ESO-1, 5T4, MAGE-C2, survivina y Muc-1
- NY-ESO-1, MAGE-C1, MAGE-C2, survivina y Muc-1
- 5T4, MAGE-C1, MAGE-C2, survivina y Muc-1
- 15 • NY-ESO-1, 5T4, MAGE-C1, MAGE-C2, survivina y Muc-1

Proteínas terapéuticas: las proteínas terapéuticas, tal como se definen en el presente documento, son péptidos o proteínas que son beneficiosos para el tratamiento de cualquier enfermedad heredada o adquirida o que mejoran el estado de un individuo. Particularmente, las proteínas terapéuticas desempeñan un gran papel en la creación de agentes terapéuticos que podrían modificar y reparar errores genéticos, destruir células cancerosas o células infectadas por patógenos, tratar trastornos del sistema inmunitario, tratar trastornos metabólicos o endocrinos, entre otras funciones. Por ejemplo, la eritropoyetina (EPO), una hormona proteica puede utilizarse en el tratamiento de pacientes con deficiencia de eritrocitos, que es una causa común de complicaciones renales. Además, las proteínas adyuvantes, los anticuerpos terapéuticos están abarcados en las proteínas terapéuticas y también en la terapia sustitutiva hormonal que se usa, por ejemplo, en la terapia de mujeres en la menopausia. En los enfoques más nuevos, las células somáticas de un paciente se usan para la reprogramación en células madre pluripotentes que reemplazan la terapia con células madre en disputa. Además, estas proteínas usadas para la reprogramación de células somáticas o usadas para la diferenciación de células madre se definen en el presente documento como proteínas terapéuticas. Además, las proteínas terapéuticas pueden usarse para otros fines, por ejemplo, cicatrización de heridas, regeneración de tejidos, angiogénesis, etc.

Por lo tanto, pueden usarse proteínas terapéuticas para diversos fines, incluyendo el tratamiento de diversas enfermedades como por ejemplo, enfermedades infecciosas, neoplasias (por ejemplo, cáncer o enfermedades tumorales), enfermedades de la sangre y de los órganos hematopoyéticos, enfermedades endocrinas, nutricionales y metabólicas, enfermedades del sistema nervioso, enfermedades del aparato circulatorio, enfermedades del aparato respiratorio, enfermedades del aparato digestivo, enfermedades de la piel y del tejido subcutáneo, enfermedades del aparato locomotor y del tejido conjuntivo, y enfermedades del aparato genitourinario, independientemente de si son hereditarias o adquiridas.

En este contexto, las proteínas terapéuticas particularmente preferidas que pueden usarse, entre otros, en el tratamiento de trastornos metabólicos o endocrinos se seleccionan de: esfingomielinasa ácida (enfermedad de Niemann-Pick), adipotida (obesidad), agalsidasa-beta (galactosidasa A humana) (enfermedad de Fabry; previene la acumulación de lípidos que podría conducir a complicaciones renales y cardiovasculares), alglucosidasa (enfermedad de Pompe (enfermedad por almacenamiento de glucógeno tipo II)), alfa-galactosidasa A (alfa-GAL A, agalsidasa alfa) (enfermedad de Fabry), alfa-glucosidasa (enfermedad por almacenamiento de glucógeno (EAG), enfermedad de Pompe), alfa-L-iduronidasa (mucopolisacaridosis (MPS), síndrome de Hurler, síndrome de Scheie), alfa-N-acetilglucosaminidasa (síndrome de Sanfilippo), anfirregulina (cáncer, trastorno metabólico), angiopoyetina ((Ang1, Ang2, Ang3, Ang4, ANGPTL2, ANGPTL3, ANGPTL4, ANGPTL5, ANGPTL6, ANGPTL7) (angiogénesis, estabilizar vasos), betacelulina (trastorno metabólico), beta-glucuronidasa (síndrome de Sly), proteína morfogenética ósea BMP (BMP1, BMP2, BMP3, BMP4, BMP5, BMP6, BMP7, BMP8a, BMP8b, BMP10, BMP15) (efecto regenerativo, estados relacionados con los huesos, enfermedad renal crónica (ERC)), proteína CLN6 (enfermedad CLN6 - atípica infantil tardía, variante de aparición tardía, juvenil temprana, lipofuscinosis cerioide neuronal (LCN)), factor de crecimiento epidérmico (EGF) (cicatrización de heridas, regulación del crecimiento, proliferación y diferenciación celular), Epigen (trastorno metabólico), epirregulina (trastorno metabólico), factor de crecimiento de fibroblastos (FGF) , FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7, FGF-8, FGF-9, FGF-10, FGF-11, FGF-12, FGF-13, FGF-14, FGF-16, FGF-17, FGF-18, FGF-19, FGF-20, FGF-21, FGF-22, FGF-23) (cicatrización de heridas, angiogénesis, trastornos endocrinos, regeneración tisular), galsulfasa (mucopolisacaridosis VI), grelina (síndrome del intestino irritable (SII), obesidad, síndrome de Prader-Willi, diabetes mellitus tipo II), glucocerebrosidasa (enfermedad de Gaucher), GM-CSF (efecto regenerador, producción de glóbulos blancos, cáncer), factor de crecimiento similar al EGF que se une a heparina (HB-EGF) (cicatrización de heridas, hipertrofia cardíaca y desarrollo y función cardíaca), factor de crecimiento de hepatocitos HGF (efecto regenerador, cicatrización de heridas), hepcidina (trastornos del metabolismo del hierro, beta-talasemia), albúmina humana (disminución de la producción de albúmina) (hipoproteinemia), aumento de la pérdida de albúmina (síndrome nefrótico), hipovolemia, hiperbilirrubinemia), idursulfasa (iduronato-2-sulfatasa) (mucopolisacaridosis II (síndrome de Hunter)), integrinas  $\alpha$ Vb3,  $\alpha$ Vb5 y  $\alpha$ 5b1 (unión de macromoléculas y proteinasas de la matriz, angiogénesis), iduronato sulfatasa (síndrome de Hunter),

laronidasa (formas de mucopolisacaridosis I de Hurler y Hurler-Scheie), N-acetilgalactosamina-4-sulfatasa (rhASB; galsulfasa, arilsulfatasa A (ARSA), arilsulfatasa B (ARSB)) (deficiencia de arilsulfatasa B, síndrome de Maroteaux-Lamy, mucopolisacaridosis VI), N-acetilglucosamina-6-sulfatasa (síndrome de Sanfilippo), factor de crecimiento nervioso (NGF, factor neurotrófico derivado del cerebro (BDNF), neurotrofina-3 (NT-3) y neurotrofina 4/5 (NT-4/5)) (efecto regenerador, enfermedades cardiovasculares, aterosclerosis coronaria, obesidad, diabetes tipo 2, síndrome metabólico, síndromes coronarios agudos, demencia, depresión, esquizofrenia, autismo, síndrome de Rett, anorexia nerviosa, bulimia nerviosa, cicatrización de heridas, úlceras cutáneas, úlceras corneales, enfermedad de Alzheimer), neuregulina (NRG1, NRG2, NRG3, NRG4) (trastorno metabólico, esquizofrenia), neuropilina (NRP-1, NRP-2) (angiogénesis, guía de axones, supervivencia, migración celular), obestatina (síndrome del intestino irritable (SII), obesidad, síndrome de Prader-Willi, diabetes mellitus tipo II), factor de crecimiento derivado de plaquetas (PDGF (PDGF-A, PDGF-B, PDGF-C, PDGF-D) (efecto regenerador, cicatrización de heridas, trastorno en la angiogénesis, arteriosclerosis, fibrosis, cáncer), receptores de TGF beta (endoglin, receptor de TGF-beta 1, receptor de TGF-beta 2, receptor de TGF-beta 3) (fibrosis renal, enfermedad renal, diabetes, en última instancia, insuficiencia renal en estadio terminal (IRT), angiogénesis), trombopoyetina (THPO) (factor de crecimiento y desarrollo de megacariocitos (MGDF)) (trastornos plaquetarios, plaquetas para donación, recuperación del recuento plaquetario después de la quimioterapia mielosupresora), factor de crecimiento transformante (TGF (TGF-a, TGF-beta (TGFbeta1, TGFbeta2 y TGFbeta3))) (efecto regenerativo, cicatrización de heridas, inmunidad, cáncer, enfermedad cardíaca, diabetes, síndrome de Marfan, síndrome de Loeys-Dietz), VEGF (VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGF-E, VEGF-F y PlGF) (efecto regenerador, angiogénesis, cicatrización de heridas, cáncer, permeabilidad), nesiritida (insuficiencia cardíaca congestiva aguda descompensada), tripsina (úlceras de decúbito, úlcera varicosa, desbridamiento de escara, herida dehiscente, quemaduras solares, íleo meconial), hormona adrenocorticotrófica (ACTH) ("enfermedad de Addison, carcinoma de células pequeñas, adrenoleucodistrofia, hiperplasia suprarrenal congénita, síndrome de Cushing, síndrome de Nelson, espasmos infantiles), péptido auricular-natriurético (ANP) (trastornos endocrinos), pancreocimina (diversa), gastrina (hipogastrinemia), leptina (diabetes, hipertrigliceridemia, obesidad), oxitocina (estimula la lactancia materna, no progresión del parto), somatostatina (tratamiento sintomático del síndrome carcinoide, hemorragia aguda por varices y acromegalia, enfermedades poliquísticas del hígado y riñón, acromegalia y síntomas provocados por tumores neuroendocrinos), vasopresina (hormona antidiurética) (diabetes insípida), calcitonina (osteoporosis posmenopáusica, hipercalcemia, enfermedad de Paget, metástasis óseas, dolor en el miembro fantasma, estenosis espinal), exenatida (diabetes tipo 2 resistente al tratamiento con metformina y una sulfonilurea), hormona del crecimiento (GH), somatotropina (retraso del crecimiento debido a deficiencia de GH o insuficiencia renal crónica, síndrome de Prader-Willi, síndrome de Turner, emaciación por SIDA o caquexia con terapia antiviral), insulina (diabetes mellitus, cetoacidosis diabética, hiperpotasiemia), factor de crecimiento similar a la insulina 1 IGF-1 (retraso del crecimiento en niños con delección del gen de GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedades cardiovasculares, insuficiencia cardíaca), mecasermina rinfabato, análogo de IGF-1 (retraso del crecimiento en niños con delección del gen de GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedades cardiovasculares, insuficiencia cardíaca), mecasermina, análogo de IGF-1 (retraso del crecimiento en niños con delección del gen de GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedades cardiovasculares, insuficiencia cardíaca), pegvisomant (acromegalia), pramlintida (diabetes mellitus, en combinación con insulina), teriparatida (residuos de hormona paratiroidea humana 1-34) (osteoporosis grave), becaplermina (adyuvante de desbridamiento para úlceras diabéticas), dibotermín-alfa (proteína morfogenética ósea 2) (cirugía de fusión espinal, reparación de lesiones óseas), acetato de histrelinina (hormona liberadora de gonadotropina; GnRH) (pubertad precoz), octreotida (acromegalia, alivio sintomático del adenoma secretor de VIP y tumores carcinoides metastásicos) y palifermin (factor de crecimiento de queratinocitos; KGF) (mucositis oral grave en pacientes sometidos a quimioterapia, cicatrización de heridas) (entre paréntesis está la enfermedad particular para la que se usa la proteína terapéutica en el tratamiento).

Se entiende que estas y otras proteínas son terapéuticas, ya que están destinadas a tratar al sujeto reemplazando su producción endógena defectuosa de una proteína funcional en cantidades suficientes. Por consiguiente, tales proteínas terapéuticas son normalmente proteínas de mamíferos, en particular proteínas humanas.

Para el tratamiento de trastornos sanguíneos, enfermedades del aparato circulatorio, enfermedades del aparato respiratorio, cáncer o enfermedades tumorales, enfermedades infecciosas o inmunodeficiencias pueden usarse las siguientes proteínas terapéuticas: alteplasa (activador del plasminógeno tisular; tPA) (embolia pulmonar, infarto de miocardio, accidente cerebrovascular isquémico agudo, oclusión de dispositivos de acceso venoso central), anistreplasa (trombólisis), antitrombina III (AT-III) (deficiencia hereditaria de AT-III, tromboembolismo), bivalirudina (reduce el riesgo de coagulación de la sangre en la angioplastia coronaria y la trombocitopenia inducida por heparina), darbepoetina alfa (tratamiento de la anemia en pacientes con insuficiencia renal crónica e insuficiencia renal crónica (+/- diálisis)), drotrecogina alfa (proteína C activada) (septicemia grave con alto riesgo de muerte), eritropoyetina, epoetina alfa (anemia de enfermedad crónica, mielodisplasia, anemia por fallo renal o quimioterapia, preparación preoperatoria), factor IX (hemofilia B), factor VIIa (hemorragia en pacientes con hemofilia A o B e inhibidores del factor VIII o factor IX), factor VIII (hemofilia A), lepirudina (trombocitopenia inducida por heparina), concentrado de proteína C (trombosis venosa, púrpura fulminante), reteplasa (delección de la muteína del tPA) (manejo del infarto agudo de miocardio, mejora de la función ventricular), estreptocinasas (infarto agudo de miocardio transmurales evolutivos, embolia pulmonar, trombosis venosa profunda, trombosis arterial o embolia, oclusión de la cánula arteriovenosa), tenecteplasa (infarto agudo de miocardio), urocinasas (embolia pulmonar), angiostatina

- (cáncer), inmunotoxina anti-CD22 (leucemia mieloide aguda CD33+ recidivante), denileucina diftotox (linfoma cutáneo de células T (LCCT)), inmunocianina (cáncer de vejiga y próstata), MPS (metalopanestimulina) (cáncer), aflibercept (cáncer de pulmón de células no pequeñas (CPCNP), cáncer colorrectal metastásico (CCRM), cáncer de próstata metastásico resistente a hormonas, degeneración macular húmeda), endostatina (cáncer, enfermedades inflamatorias tales como artritis reumatoide, así como enfermedad de Crohn, retinopatía diabética, psoriasis y endometriosis), collagenasa (desbridamiento de úlceras dérmicas crónicas y áreas con quemaduras graves, contractura de Dupuytren, enfermedad de Peyronie), desoxirribonucleasa I humana, dornasa (fibrosis quística; disminuye las infecciones de las vías respiratorias en pacientes seleccionados con CVF superior al 40% de lo previsto), hialuronidasa (usada como adyuvante para aumentar la absorción y dispersión de fármacos inyectados, en particular anestésicos en cirugía oftálmica y ciertos agentes de obtención de imágenes), papaína (desbridamiento del tejido necrótico o licuación de esfacelos en lesiones agudas y crónicas, tal como úlceras de decúbito, úlceras varicosas y diabéticas, quemaduras, heridas posoperatorias, heridas de quistes pilonidales, ántrax y otras heridas), L-asparaginasa (leucemia linfocítica aguda, que requiere asparagina exógena para la proliferación), Peg-asparaginasa (leucemia linfocítica aguda, que requiere asparagina exógena para la proliferación), rasburicasa (pacientes pediátricos con leucemia, linfoma y tumores sólidos que están recibiendo terapia contra el cáncer que puede provocar síndrome de lisis tumoral), gonadotropina coriónica humana (HCG) (reproducción asistida), hormona foliculoestimulante humana (FSH) (reproducción asistida), lutropina alfa (infertilidad con deficiencia de hormona luteinizante), prolactina (hipoprolactinemia, deficiencia de prolactina sérica, disfunción ovárica en mujeres, ansiedad, disfunción eréctil arteriogénica, eyaculación precoz, oligozoospermia, astenospermia, hipofunción de vesículas seminales, hipoandrogenismo en hombres), inhibidor de alfa-1-proteinasa (deficiencia congénita de antitripsina), lactasa (gases, distensión abdominal, calambres y diarrea debido a la incapacidad para digerir la lactosa), enzimas pancreáticas (lipasa, amilasa, proteasa) (fibrosis quística, pancreatitis crónica, insuficiencia pancreática, cirugía de derivación gástrica posterior a Billroth II, obstrucción del conducto pancreático, esteatorrea, mala digestión, gases, distensión abdominal), adenosina desaminasa (pegademasa bovina, PEG-ADA) (enfermedad de inmunodeficiencia combinada grave debido a deficiencia de adenosina desaminasa), abatacept (artritis reumatoide (especialmente cuando es resistente a la inhibición del TNF $\alpha$ )), alefacept (psoriasis en placas), anakinra (artritis reumatoide), etanercept (artritis reumatoide, artritis reumatoide juvenil de curso poliarticular, artritis psoriásica, espondilitis anquilosante, psoriasis en placas, espondilitis anquilosante), antagonista del receptor de interleucina-1 (IL-1), anakinra (inflamación y degradación del cartilago asociada con la artritis reumatoide), timulina (enfermedades neurodegenerativas, reumatismo, anorexia nerviosa), antagonista del TNF- $\alpha$  (trastornos autoinmunitarios tales como artritis reumatoide, espondilitis anquilosante, enfermedad de Crohn, psoriasis, hidrosadenitis supurativa, asma resistente al tratamiento), enfuvirtida (infección por VIH-1) y timosina  $\alpha$ 1 (hepatitis B y C) (entre paréntesis está la enfermedad particular para la cual se usa la proteína terapéutica en el tratamiento).
- Además, las proteínas adyuvantes o inmunoestimulantes también están abarcados en el término proteínas terapéuticas. Pueden usarse proteínas adyuvantes o inmunoestimulantes en este contexto para inducir, alterar o mejorar una respuesta inmunitaria en un individuo para tratar una enfermedad particular o para mejorar el estado del individuo.
- En este contexto, las proteínas adyuvantes pueden seleccionarse de proteínas adyuvantes de mamíferos, en particular proteínas adyuvantes humanas, que normalmente comprenden cualquier proteína o péptido humano, que pueda provocar una respuesta inmunitaria innata (en un mamífero), por ejemplo, como una reacción de la unión de un ligando de TLR exógeno a un TLR. Más preferiblemente, las proteínas adyuvantes humanas se seleccionan del grupo que consiste en proteínas, que son componentes y ligandos de las redes de señalización de los receptores de reconocimiento de patrones, incluidos TLR, NLR y RLH, incluidos TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, TLR11; NOD1, NOD2, NOD3, NOD4, NOD5, NALP1, NALP2, NALP3, NALP4, NALP5, NALP6, NALP7, NALP8, NALP9, NALP10, NALP11, NALP12, NALP13, NALP14, 1 IPAF, NAIP, CIITA, RIG-I, MDA5 y LGP2, los transductores de señal de la señalización de TLR que incluyen las proteínas adaptadoras, incluyendo, por ejemplo, Trif y Cardif; componentes de la señalización de pequeñas GTPasas (RhoA, Ras, Rac1, Cdc42, Rab, etc.), componentes de la señalización de PIP (PI3K, Src-cinasas, etc.), componentes de la señalización dependiente de MyD88 (MyD88, IRAK1, IRAK2, IRAK4, TIRAP, TRAF6, etc.), componentes de la señalización independiente de MyD88 (TICAM1, TICAM2, TRAF6, TBK1, IRF3, TAK1, IRAK1, etc.); las cinasas activadas incluyendo, por ejemplo, Akt, MEKK1, MKK1, MKK3, MKK4, MKK6, MKK7, ERK1, ERK2, GSK3, cinasas PKC, cinasas PKD, cinasas GSK3, JNK, p38MAPK, TAK1, IKK y TAK1; los factores de transcripción activados incluyendo, por ejemplo, NF- $\kappa$ B, c-Fos, c-Jun, c-Myc, CREB, AP-1, Elk-1, ATF2, IRF-3, IRF-7.

Las proteínas adyuvantes de mamíferos, en particular proteínas adyuvantes humanas, pueden seleccionarse además del grupo que consiste en proteínas de choque térmico, tales como HSP10, HSP60, HSP65, HSP70, HSP75 y HSP90, gp96, fibrinógeno, dominio adicional A de repetición TyplII de fibronectina; o componentes del sistema del complemento, incluyendo C1q, MBL, C1r, C1s, C2b, Bb, D, MASP-1, MASP-2, C4b, C3b, C5a, C3a, C4a, C5b, C6, C7, C8, C9, CR1, CR2, CR3, CR4, C1qR, C1INH, C4bp, MCP, DAF, H, I, P y CD59, o genes diana inducidos incluyendo, por ejemplo, beta-defensina, proteínas de la superficie celular; o proteínas adyuvantes humanas incluyendo trif, ligando de flt-3, Gp96 o fibronectina, etc., o cualquier homólogo de especie de cualquiera de las proteínas adyuvantes humanas anteriores.

Las proteínas adyuvantes de mamíferos, en particular proteínas adyuvantes humanas, pueden comprender además

citocinas que inducen o mejoran una respuesta inmunitaria innata, incluyendo IL-1 alfa, IL1 beta, IL-2, IL-6, IL-7, IL-8, IL-9, IL-12, IL-13, IL-15, IL-16, IL-17, IL-18, IL-21, IL-23, TNFalfa, IFNalfa, IFNbeta, IFNgamma, GM-CSF, G-CSF, M-CSF, quimiocinas incluyendo IL-8, IP-10, MCP-1, MIP-1 alfa, RANTES, eotaxina, CCL21; citocinas que se liberan de macrófagos, incluyendo IL-1, IL-6, IL-8, IL-12 y TNF-alfa; así como IL-1R1 e IL-1 alfa.

Las proteínas terapéuticas para el tratamiento de trastornos sanguíneos, enfermedades del aparato circulatorio, enfermedades del aparato respiratorio, cáncer o enfermedades tumorales, enfermedades infecciosas o inmunodeficiencias o proteínas adyuvantes son normalmente proteínas de origen mamífero, preferiblemente de origen humano, según el animal que será tratado. Un sujeto humano, por ejemplo, se trata preferiblemente con una proteína terapéutica de origen humano.

Las proteínas adyuvantes patógenas comprenden normalmente una proteína adyuvante patógena que puede provocar una respuesta inmunitaria innata (en un mamífero), más preferiblemente seleccionadas de entre proteínas adyuvantes patógenas derivadas de bacterias, protozoos, virus u hongos, etc., por ejemplo, proteínas bacterianas (adyuvantes), proteínas protozoarias (adyuvantes) (por ejemplo, proteína similar a la profilina de *Toxoplasma gondii*), proteínas virales (adyuvantes) o proteínas fúngicas (adyuvantes), etc.

Particularmente, las proteínas bacterianas (adyuvantes) pueden seleccionarse del grupo que consiste en proteínas bacterianas de choque térmico o chaperonas, incluyendo Hsp60, Hsp70, Hsp90, Hsp100; OmpA (proteína de la membrana externa) de bacterias gramnegativas; porinas bacterianas, incluyendo OmpF; toxinas bacterianas, incluyendo la toxina pertúsica (PT) de *Bordetella pertussis*, toxina pertúsica adenilato ciclasa CyaA y CyaC de *Bordetella pertussis*, mutante PT-9K/129G de la toxina pertúsica, toxina pertúsica adenilato ciclasa CyaA y CyaC de *Bordetella pertussis*, toxina tetánica, toxina del cólera (CT), subunidad B de la toxina del cólera, mutante CTK63 de la toxina del cólera, mutante CTE112K de CT, enterotoxina termolábil de *Escherichia coli* (LT), subunidad B de mutantes de enterotoxina termolábiles (LTB) de *Escherichia coli* con toxicidad reducida, incluyendo LTK63, LTR72; modulina soluble en fenol; proteína activadora de neutrófilos (HP-NAP) de *Helicobacter pylori*; proteína tensioactiva D; lipoproteína de la proteína A de la superficie externa de *Borrelia burgdorferi*, Ag38 (antígeno de 38 kDa) de *Mycobacterium tuberculosis*; proteínas de fimbrias bacterianas; enterotoxina CT de *Vibrio cholerae*, pilina de pilosidades de bacterias gramnegativas y proteína tensioactiva A; etc., o cualquier homólogo de especie de cualquiera de las proteínas bacterianas (adyuvantes) anteriores. Las proteínas bacterianas (adyuvantes) también pueden comprender flagelinas bacterianas. En el contexto de la presente invención, las flagelinas bacterianas pueden seleccionarse de flagelinas de organismos incluyendo, sin limitarse a lo mismo, *Agrobacterium*, *Aquifex*, *Azospirillum*, *Bacillus*, *Bartonella*, *Bordetella*, *Borrelia*, *Burkholderia*, *Campylobacter*, *Caulobacter*, *Clostridium*, *Escherichia*, *Helicobacter*, *Lachnospiraceae*, *Legionella*, *Listeria*, *Proteus*, *Pseudomonas*, *Rhizobium*, *Rhodobacter*, *Roseburia*, *Salmonella*, *Serpulina*, *Serratia*, *Shigella*, *Treponema*, *Vibrio*, *Wolinella*, *Yersinia*, más preferiblemente de las flagelinas de las especies que incluyen, sin limitarse a lo mismo, *Agrobacterium tumefaciens*, *Aquifex pyrophilus*, *Azospirillum brasilense*, *Bacillus subtilis*, *Bacillus thuringiensis*, *Bartonella bacilliformis*, *Bordetella bronchiseptica*, *Borrelia burgdorferi*, *Burkholderia cepacia*, *Campylobacter jejuni*, *Caulobacter crescentus*, clon 1 de cepa Bennett de *Clostridium botulinum*, *Escherichia coli*, *Helicobacter pylori*, *Lachnospiraceae bacterium*, *Legionella pneumophila*, *Listeria monocytogenes*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Pseudomonas syringae*, *Rhizobium meliloti*, *Rhodobacter sphaeroides*, *Roseburia cecicola*, *Roseburia hominis*, *Salmonella typhimurium*, *Salmonella bongori*, *Salmonella typhi*, *Salmonella enteritidis*, *Serpulina hyodysenteriae*, *Serratia marcescens*, *Shigella boydii*, *Treponema phagedenis*, *Vibrio alginolyticus*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Wolinella succinogenes* y *Yersinia enterocolitica*.

Las proteínas protozoarias (adyuvantes) son un ejemplo adicional de proteínas adyuvantes patógenas. Las proteínas protozoarias (adyuvantes) pueden seleccionarse en este contexto de cualquier proteína protozoaria que muestre propiedades adyuvantes, más preferiblemente, del grupo que consiste en, sin limitarse a lo mismo, Tc52 de *Trypanosoma cruzi*, PFTG de *Trypanosoma gondii*, proteínas de choque térmico de protozoos, LelF de *Leishmania* spp., proteína de tipo perfilado de *Toxoplasma gondii*, etc.

Las proteínas virales (adyuvantes) son otro ejemplo de proteínas adyuvantes patógenas. En este contexto, las proteínas virales (adyuvantes) pueden seleccionarse de cualquier proteína viral que muestre propiedades adyuvantes, más preferiblemente, del grupo que consiste en, sin limitarse a lo mismo, glucoproteína de fusión (proteína F) del virus respiratorio sincitial, proteína de la envuelta del virus MMT, proteína del virus de la leucemia de ratón, proteína hemaglutinina del virus del sarampión de tipo natural, etc.

Las proteínas fúngicas (adyuvantes) son incluso un ejemplo adicional de proteínas adyuvantes patógenas. En el contexto de la presente invención, las proteínas fúngicas (adyuvantes) pueden seleccionarse de cualquier proteína fúngica que muestre propiedades adyuvantes, más preferiblemente, del grupo que consiste en proteína inmunomoduladora fúngica (FIP; LZ-8), etc.

Finalmente, las proteínas adyuvantes pueden seleccionarse además del grupo que consiste en hemocianina de lapa californiana (KLH), OspA, etc.

En una realización adicional, pueden usarse proteínas terapéuticas para la terapia sustitutiva hormonal,

particularmente para la terapia de mujeres en la menopausia. Estas proteínas terapéuticas se seleccionan preferiblemente entre estrógenos, progesterona o progestinas y, a veces, testosterona.

Además, pueden usarse proteínas terapéuticas para reprogramar células somáticas en células madre pluripotentes u omnipotentes. Para este fin, se describen varios factores, en particular Oct-3/4, familia de genes Sox (Sox1, Sox2, Sox3 y Sox15), familia Klf (Klf1, Klf2, Klf4 y Klf5), familia Myc (c-myc, L-myc y N-myc), Nanog y LIN28.

Tal como se mencionó anteriormente, los anticuerpos terapéuticos se definen también en el presente documento como proteínas terapéuticas. Estos anticuerpos terapéuticos se seleccionan preferiblemente de anticuerpos que se usan, entre otros, para el tratamiento de cáncer o enfermedades tumorales, por ejemplo, 131I-tositumomab (linfoma folicular, linfomas de células B, leucemias), 3F8 (neuroblastoma), 8H9, abagovomab (cáncer de ovario), adecatumumab (cáncer de próstata y mama), afutuzumab (linfoma), alacizumab pegol, alemtuzumab (leucemia linfocítica crónica de células B, linfoma de células T), amatuximab, AME-133v (linfoma folicular, cáncer), AMG 102 (carcinoma de células renales avanzado), anatumomab mafenatox (carcinoma de pulmón de células no pequeñas), apolizumab (tumores sólidos, leucemia, linfoma no Hodgkin, linfoma), bavituximab (cáncer, infecciones virales), bectumomab (linfoma no Hodgkin), belimumab (linfoma no Hodgkin), bevacizumab (cáncer de colon, cáncer de mama, tumores del cerebro y del sistema nervioso central, cáncer de pulmón, carcinoma hepatocelular, cáncer de riñón, cáncer de mama, cáncer de páncreas, cáncer de vejiga, sarcoma, melanoma, cáncer de esófago; cáncer de estómago, carcinoma metastásico de células renales; cáncer de riñón, glioblastoma, cáncer de hígado, retinopatía diabética proliferativa, degeneración macular), bivatuzumab mertansina (carcinoma de células escamosas), blinatumomab, brentuximab vedotina (cánceres hematológicos), cantuzumab (cáncer de colon, cáncer de estómago, cáncer de páncreas, CPCNP), cantuzumab mertansina (cáncer colorrectal), cantuzumab ravtansina (cánceres), capromab pendetido (cáncer de próstata), carlumab, catumaxomab (cáncer de ovario, neoplasias de las trompas de Falopio, neoplasias peritoneales), cetuximab (cáncer colorrectal metastásico y cáncer de cabeza y cuello), citatuzumab bogatox (cáncer de ovario y otros tumores sólidos), cixutumumab (tumores sólidos), clivatuzumab tetraxetano (cáncer de páncreas), CNTO 328 (linfoma no Hodgkin de células B, mieloma múltiple, enfermedad de Castleman, cáncer de ovario), CNTO 95 (melanoma), conatumumab, dacetuzumab (cánceres hematológicos), dalotuzumab, denosumab (mieloma, tumor de células gigantes de hueso, cáncer de mama, cáncer de próstata, osteoporosis), detumomab (linfoma), drozitumab, ecomeximab (melanoma maligno), edrecolomab (carcinoma colorrectal), elotuzumab (mieloma múltiple), elsilimomab, enavatuzumab, ensituximab, epratuzumab (enfermedades autoinmunitarias, lupus eritematoso sistémico, linfoma no Hodgkin, leucemia), ertumaxomab (cáncer de mama), ertumaxomab (cáncer de mama), etaracizumab (melanoma, cáncer de próstata, cáncer de ovario), farletuzumab (cáncer de ovario), FBTA05 (leucemia linfocítica crónica), ficlatuzumab (cáncer), figitumumab (carcinoma adrenocortical, carcinoma de pulmón de células no pequeñas), flanvotumab (melanoma), galiximab (linfoma de células B), galiximab (linfoma no Hodgkin), ganitumab, GC1008 (carcinoma avanzado de células renales; melanoma maligno, fibrosis pulmonar), gemtuzumab (leucemia), gemtuzumab ozogamicina (leucemia mielógena aguda), girentuximab (carcinoma de células renales de células claras), glembatumumab vedotina (melanoma, cáncer de mama), GS6624 (fibrosis pulmonar idiopática y tumores sólidos), HuC242-DM4 (cáncer de colon, cáncer de estómago, cáncer de páncreas), HuHMF1 (cáncer de mama), HuN901-DM1 (mieloma), ibritumomab (linfoma no Hodgkin (NHL) de células B con recaída o resistente al tratamiento en grado bajo, folicular o transformado), icrucumab, ID09C3 (linfoma no Hodgkin), indatuximab ravtansina, inotuzumab ozogamicina, intetumumab (tumores sólidos (cáncer de próstata, melanoma)), ipilimumab (sarcoma, melanoma, cáncer de pulmón, cáncer de ovario, leucemia, linfoma, tumores de cerebro y del sistema nervioso central, cáncer de testículo, cáncer de próstata, cáncer de páncreas, cáncer de mama), iratumumab (linfoma de Hodgkin), labetuzumab (cáncer colorrectal), lexatumumab, lintuzumab, lorvotuzumab mertansina, lucatumumab (mieloma múltiple, linfoma no Hodgkin, linfoma de Hodgkin), lumiliximab (leucemia linfocítica crónica), mapatumumab (cáncer de colon, mieloma), matuzumab (cáncer de pulmón, cáncer de cuello uterino, cáncer de esófago), MDX-060 (linfoma de Hodgkin, linfoma), MEDI 522 (tumores sólidos, leucemia, linfoma, cáncer de intestino delgado, melanoma), mitumomab (carcinoma de pulmón de células pequeñas), mogamulizumab, MORab-003 (cáncer de ovario, cáncer de trompas de Falopio, cáncer peritoneal), MORab-009 (cáncer de páncreas, mesotelioma, cáncer de ovario, cáncer de pulmón de células no pequeñas, cáncer de trompas de Falopio, cáncer de cavidad peritoneal), moxetumomab pasudotox, MT103 (linfoma no Hodgkin), nacolomab tafenatox (cáncer colorrectal), naptumomab estafenatox (carcinoma de pulmón de células no pequeñas, carcinoma de células renales), narnatumumab, necitumumab (carcinoma de pulmón de células no pequeñas), nimotuzumab (carcinoma de células escamosas, cáncer de cabeza y cuello, cáncer de nasofaringe, glioma), nimotuzumab (carcinomas de células escamosas, glioma, tumores sólidos, cáncer de pulmón), olaratumab, onartuzumab (cáncer), oportuzumab monatox, oregovomab (cáncer de ovario), oregovomab (cáncer de ovario, cáncer de trompas de Falopio, cáncer de cavidad peritoneal), PAM4 (cáncer de páncreas), panitumumab (cáncer de colon, cáncer de pulmón, cáncer de mama; cáncer de vejiga; cáncer de ovario), patritumab, pentumomab, pertuzumab (cáncer de mama, cáncer de ovario, cáncer de pulmón, cáncer de próstata), primumab (cáncer de cerebro), racotumomab, radretumab, ramucirumab (tumores sólidos), rilatumumab (tumores sólidos), rituximab (urticaria, artritis reumatoide, colitis ulcerosa, encefalitis focal crónica, linfoma no Hodgkin, linfoma, leucemia linfocítica crónica), robatumumab, samalizumab, SGN-30 (linfoma de Hodgkin, linfoma), SGN-40 (linfoma no Hodgkin, mieloma, leucemia, leucemia linfocítica crónica), sibrotuzumab, siltuximab, tabalumab (cánceres de células B), tacatuzumab tetraxetano, taplitumomab paptox, tenatumomab, teprotumumab (tumores hematológicos), TGN1412 (leucemia linfocítica crónica, artritis reumatoide), ticilimumab (= tremelimumab), tigatuzumab, TNX-650 (linfoma no Hodgkin), tositumomab (linfoma folicular, linfomas de células B, leucemias, mieloma), trastuzumab

(cáncer de mama, cáncer de endometrio, tumores sólidos), TRBS07 (melanoma), tremelimumab, TRU-016 (leucemia linfocítica crónica), TRU-016 (linfoma no Hodgkin), Tucotuzumab celmoleucina, ublituximab, urelumab, veltuzumab (linfoma no Hodgkin), veltuzumab (IMMU-106) (linfoma no Hodgkin), volociximab (carcinoma de células renales, cáncer de páncreas, melanoma), votumumab (tumores colorrectales), WX-G250 (carcinoma de células renales),  
5 zalutumumab (cáncer de cabeza y cuello, cáncer de células escamosas) y zanolimumab (linfoma de células T); anticuerpos, que se usan entre otros, para el tratamiento de trastornos inmunitarios, por ejemplo, efalizumab (psoriasis), epratuzumab (enfermedades autoinmunitarias, lupus eritematoso sistémico, linfoma no Hodgkin, leucemia), etrolizumab (enfermedad inflamatoria del intestino), fontolizumab (enfermedad de Crohn), ixekizumab (enfermedades autoinmunitarias), mepolizumab (síndrome de hipereosinofilia, asma, gastroenteritis eosinofílica,  
10 síndrome de Churg-Strauss, esofagitis eosinofílica), milatuzumab (mieloma múltiple y otras neoplasias hematológicas malignas), inmunoglobulinas agrupadas (inmunodeficiencias primarias), priliximab (enfermedad de Crohn, esclerosis múltiple), rituximab (urticaria, artritis reumatoide, colitis ulcerosa, encefalitis focal crónica, linfoma no Hodgkin, linfoma, leucemia linfocítica crónica), rontalizumab (lupus eritematoso sistémico), ruplizumab (enfermedades reumáticas), sarilumab (artritis reumatoide, espondilitis anquilosante), vedolizumab (enfermedad de Crohn, colitis ulcerosa), visilizumab (enfermedad de Crohn, colitis ulcerosa), reslizumab (inflamaciones de las vías respiratorias, la piel y el tracto gastrointestinal), adalimumab (artritis reumatoide, enfermedad de Crohn, espondilitis anquilosante, artritis psoriásica), aselizumab (pacientes con lesiones graves), atinumab (tratamiento de sistemas neurológicos), atlizumab (artritis reumatoide, artritis idiopática juvenil sistémica), bertilimumab (trastornos alérgicos graves), besilesomab (lesiones inflamatorias y metástasis), BMS-945429, ALD518 (cáncer y artritis reumatoide),  
20 briakinumab (psoriasis, artritis reumatoide, enfermedades inflamatorias del intestino, esclerosis múltiple), brodalumab (enfermedades inflamatorias), canakinumab (artritis reumatoide), canakinumab (síndromes periódicos asociados a criopirina (CAPS), artritis reumatoide, enfermedad pulmonar obstructiva crónica), certolizumab pegol (enfermedad de Crohn), erlizumab (infarto de miocardio, accidente cerebrovascular, choque traumático), fezakinumab (artritis reumatoide, psoriasis), golimumab (artritis reumatoide, artritis psoriásica, espondilitis anquilosante), gomiliximab (asma alérgica), infliximab (artritis reumatoide, enfermedad de Crohn, espondilitis anquilosante, artritis psoriásica, psoriasis en placas, enfermedad de Bechterew, colitis ulcerosa), mavrilimumab (artritis reumatoide), natalizumab (esclerosis múltiple), ocrelizumab (esclerosis múltiple, artritis reumatoide, lupus eritematoso, cáncer hematológico), odulimomab (prevención del rechazo de trasplante de órganos, enfermedades inmunológicas), ofatumumab (leucemia linfocítica crónica, linfoma folicular no Hodgkin, linfoma de células B, artritis reumatoide, esclerosis múltiple remitente recurrente, linfoma, leucemia linfocítica crónica de células B), ozoralizumab (inflamación), pexelizumab (reducción de los efectos secundarios de cirugía cardíaca), rovelizumab (choque hemorrágico), SBI-087 (artritis reumatoide), SBI-087 (lupus eritematoso sistémico), secukinumab (uveítis, artritis reumatoide, psoriasis), sirukumab (artritis reumatoide), talizumab (reacción alérgica), tocilizumab (artritis reumatoide, artritis idiopática juvenil sistémica, enfermedad de Castleman), toralizumab (artritis reumatoide, nefritis lúpica),  
35 TRU-015 (artritis reumatoide), TRU-016 (enfermedad autoinmunitaria e inflamación), ustekinumab (esclerosis múltiple, psoriasis, artritis psoriásica), ustekinumab (bloqueador de IL-12/IL-23) (psoriasis en placas, artritis psoriásica, esclerosis múltiple, sarcoidosis, este último frente a), vepalimomab (inflamación), zolimomab aritox (lupus eritematoso sistémico, enfermedad de injerto contra huésped), sifalimumab (LES, dermatomiositis, polimiositis), lumiliximab (alergias) e inmunoglobulina Rho(D) (enfermedad de Rhesus); o se seleccionan de anticuerpos usados para el tratamiento de enfermedades infecciosas, por ejemplo, afelimomab (septicemia), CR6261 (enfermedad infecciosa/gripe A), edobacomab (septicemia provocada causada por bacterias gramnegativas), efungumab (infección invasiva por *Candida*), exbivirumab (hepatitis B), felvizumab (infección por virus respiratorio sincitial), foravirumab (rabia (profilaxis)), ibalizumab (infección por VIH), libivirumab (hepatitis B), motavizumab (virus respiratorio sincitial (prevención)), nebacumab (septicemia), tuvirumab (hepatitis B crónica), urtoxazumab (diarrea causada por *E. coli*), bavituximab (diversas infecciones virales), pagibaximab (septicemia (por ejemplo, *Staphylococcus*)), palivizumab (prevención de la infección por virus respiratorio sincitial en pacientes pediátricos de alto riesgo), panobacumab (infección por *Pseudomonas aeruginosa*), PRO 140 (infección por VIH), rafivirumab (rabia (profilaxis)), raxibacumab (carbúnculo (profilaxis y tratamiento)), regavirumab (infección por citomegalovirus), sevirumab (infección por citomegalovirus), suvizumab (infecciones virales) y tefibazumab (infección por *Staphylococcus aureus*); anticuerpos, que se usan, entre otros, para el tratamiento de trastornos sanguíneos, por ejemplo, abcximab (intervención coronaria percutánea), atorolimumab (enfermedad hemolítica del recién nacido), eculizumab (hemoglobulinuria paroxística nocturna), mepolizumab (síndrome de hipereosinofilia, asma, gastroenteritis eosinofílica, síndrome de Churg-Strauss, esofagitis eosinofílica) y milatuzumab (mieloma múltiple y otras neoplasias hematológicas malignas); anticuerpos, que se usan, entre otros, para inmunorregulación, por ejemplo, globulina antitimocítica (rechazo agudo de trasplante de riñón, anemia aplásica), basiliximab (profilaxis contra el rechazo de aloinjerto en pacientes con trasplante renal que reciben una pauta posológica inmunosupresora que incluye ciclosporina y corticosteroides), cedelizumab (prevención del rechazo de trasplante de órganos, tratamiento de enfermedades autoinmunitarias), daclizumab (profilaxis contra rechazo agudo de aloinjerto en pacientes que reciben trasplantes renales, esclerosis múltiple), gavilimomab (enfermedad de injerto contra huésped), inolimomab (enfermedad de injerto contra huésped), muromonab-CD3 (prevención de rechazos de trasplante de órganos), muromonab-CD3 (rechazo agudo de aloinjerto renal o rechazo de aloinjerto cardíaco o hepático resistente a esteroides), odulimomab (prevención de rechazos de trasplante de órganos, enfermedades inmunológicas) y siplizumab (psoriasis, enfermedad de injerto contra huésped (prevención)); anticuerpos que se usan para el tratamiento de la diabetes, por ejemplo, gevokizumab (diabetes), otelixizumab (diabetes mellitus tipo 1) y teplizumab  
65 (diabetes mellitus tipo 1); anticuerpos que se usan para el tratamiento de la enfermedad de Alzheimer, por ejemplo, bapineuzumab, crenezumab, gantenerumab, ponezumab, R1450 y solanezumab; anticuerpos que se usan para el

tratamiento del asma, por ejemplo, benralizumab, enokizumab, keliximab, lebrikizumab, omalizumab, oxelumab, pascolizumab y tralokinumab; y anticuerpos que se usan para el tratamiento de diversos trastornos, por ejemplo, blosozumab (osteoporosis), CaroRx (caries dental), fresolimumab (fibrosis pulmonar idiopática, glomeruloesclerosis focal y segmentaria, cáncer), fulranumab (dolor), romosozumab (osteoporosis), stamulumab (distrofia muscular), tanezumab (dolor) y ranibizumab (degeneración macular asociada con la edad neovascular).

Epítipo: los epítipos (también denominados “determinantes de antígeno”) pueden distinguirse en epítipos de células T y epítipos de células B. Los epítipos de células T o partes de las proteínas en el contexto de la presente invención pueden comprender fragmentos que tienen preferiblemente una longitud de aproximadamente 6 a aproximadamente 20 o incluso más aminoácidos, por ejemplo, fragmentos procesados y presentados por moléculas de MHC de clase I, que tienen preferiblemente una longitud de aproximadamente 8 a aproximadamente 10 aminoácidos, por ejemplo, 8, 9 ó 10 (o incluso 11 ó 12 aminoácidos), o fragmentos procesados y presentados por moléculas de MHC de clase II, que tienen preferiblemente una longitud de aproximadamente 13 o más aminoácidos, por ejemplo 13, 14, 15, 16, 17, 18, 19, 20 o incluso más aminoácidos, en los que estos fragmentos pueden seleccionarse de cualquier parte de la secuencia de aminoácidos. Estos fragmentos son reconocidos normalmente por las células T en forma de un complejo que consiste en el fragmento de péptido y una molécula de MHC, es decir, los fragmentos no son reconocidos normalmente en su forma nativa. Los epítipos de células B son fragmentos ubicados normalmente en la superficie externa de antígenos proteicos o peptídicos (nativos) tal como se define en el presente documento, que tienen preferiblemente de 5 a 15 aminoácidos, más preferiblemente de 5 a 12 aminoácidos, incluso más preferiblemente de 6 a 9 aminoácidos que pueden ser reconocidos por anticuerpos, es decir, en su forma nativa.

Además, tales epítipos de proteínas o péptidos pueden seleccionarse de cualquiera de las variantes mencionadas en el presente documento de tales proteínas o péptidos. En este contexto, los determinantes antigénicos pueden ser epítipos conformacionales o discontinuos que se componen de segmentos de las proteínas o los péptidos tal como se definen en el presente documento que son discontinuos en la secuencia de aminoácidos de las proteínas o los péptidos tal como se define en el presente documento, pero se juntan en los epítipos continuos o lineales o en la estructura tridimensional que se componen de una única cadena polipeptídica.

Proteína: una proteína comprende normalmente uno o más péptidos o polipéptidos. Normalmente, una proteína se pliega en una forma tridimensional que puede ser necesaria para que la proteína ejerza su función biológica.

Péptido: un péptido o polipéptido es normalmente un polímero de monómeros de aminoácidos unidos por enlaces peptídicos. Normalmente contiene menos de 50 unidades de monómero. Sin embargo, el término péptido no es una exclusión de moléculas que tienen más de 50 unidades de monómero. Los péptidos largos también se denominan polipéptidos y tienen normalmente entre 50 y 600 unidades monoméricas.

Fragmento o parte de una proteína: los fragmentos o partes de una proteína en el contexto de la presente invención se entienden normalmente como péptidos que corresponden a una parte continua de la secuencia de aminoácidos de una proteína, preferiblemente teniendo una longitud de aproximadamente 6 a aproximadamente 20 o incluso más aminoácidos, por ejemplo, partes procesadas y presentadas por moléculas de MHC de clase I, que tienen preferiblemente una longitud de aproximadamente 8 a aproximadamente 10 aminoácidos, por ejemplo, 8, 9 ó 10 (o incluso 11 ó 12 aminoácidos), o fragmentos procesados y presentados por moléculas de MHC de clase II, que tienen preferiblemente una longitud de aproximadamente 13 o más aminoácidos, por ejemplo, 13, 14, 15, 16, 17, 18, 19, 20 o incluso más aminoácidos, en los que estos fragmentos pueden seleccionarse de cualquier parte de la secuencia de aminoácidos. Estos fragmentos son reconocidos normalmente por las células T en forma de un complejo que consiste en el fragmento de péptido y una molécula de MHC, es decir, los fragmentos no son reconocidos normalmente en su forma nativa. Los fragmentos o partes de las proteínas tal como se definen en el presente documento también pueden comprender epítipos o sitios funcionales de esas proteínas. Preferiblemente, los fragmentos o partes de una proteína en el contexto de la invención son antígenos, particularmente inmunógenos, por ejemplo, determinantes de antígenos (también denominados “epítipos”), o tienen características antigénicas, provocando una respuesta inmunitaria adaptativa. Por tanto, los fragmentos de proteínas o péptidos pueden comprender al menos un epítipo de esas proteínas o péptidos. Además, también puede entenderse que los dominios de una proteína, como el dominio extracelular, el dominio intracelular o el dominio transmembrana, y las versiones acortadas o truncadas de una proteína, comprenden un fragmento de una proteína.

Microorganismo: un microorganismo es un organismo vivo microscópico que puede ser unicelular o multicelular. Los microorganismos son muy diversos e incluyen todas las bacterias y arqueas y casi todos los protozoos. También incluyen algunos hongos, algas y determinados animales, tales como los rotíferos. En el contexto de la presente invención, se prefieren particularmente los microorganismos que pueden replicar una molécula de ADN recombinante circular cerrado covalentemente que comprende un origen de replicación, por ejemplo, un plásmido. Particularmente preferidas son las bacterias, por ejemplo, *Escherichia coli* (*E. coli*). Particularmente preferidas en este contexto son las cepas de *E. coli* usadas para la producción de plásmidos, por ejemplo, DH1, DH5 $\alpha$ , DH10B, BL21 (por ejemplo, BL21  $\Delta$ recA $\Delta$ endA) o JM101. En algunas realizaciones, se prefieren particularmente las cepas DH5 $\alpha$ alfa, por ejemplo, DH5 $\alpha$  dcm- y DH5 $\alpha$  dcm-att  $\lambda$ :: P5/6 6/6-RNA-IN-SacB.

Inserto: un inserto es una secuencia de ADN comprendida opcionalmente en la molécula de ADN circular cerrado covalentemente según la invención, que codifica para una secuencia de ARN diana o una secuencia de péptido o proteína diana. En el caso de que el inserto codifique para un péptido o proteína, el inserto comprende un marco de lectura abierto.

Fermentación: los términos “fermentar”, “fermentación” o “producción fermentativa” se refieren al crecimiento a granel de microorganismos en un medio de crecimiento en condiciones aerobias o anaerobias. Preferiblemente, la fermentación presentada en el presente documento se refiere a bacterias en condiciones anaerobias. En el contexto de la presente invención, los términos se refieren particularmente a procedimientos de fermentación en los que el microorganismo replica una molécula de ADN circular cerrado covalentemente que comprende un origen de replicación, por ejemplo, un plásmido.

Fermentación semicontinua: la fermentación semicontinua es un procedimiento que comprende al menos una etapa de adición predeterminada o controlada de nutrientes en el cultivo de fermentación. Esta etapa de adición predeterminada o controlada de nutrientes también se denomina fase semicontinua y permite el control de la tasa de crecimiento a tasas  $< \mu_{\max}$ .

Sitio de ensamblaje del primosoma: el sitio de ensamblaje del primosoma es una secuencia de nucleótidos que soporta el ensamblaje del primosoma. El primosoma comprende las proteínas primasa DnaG, helicasa DnaB, asistente de helicasa DnaC, DnaT, PriA, Pri B y PriC. En cada horquilla de replicación, el primosoma se utiliza una vez en la hebra adelantada de ADN y de manera repetida, iniciando cada fragmento de Okazaki, en la hebra retrasada de ADN.

#### Descripción detallada de la invención

Para resolver el problema mencionado anteriormente, la presente invención emplea un molde de ADN plasmídico lineal tal como se define en las reivindicaciones en el que la secuencia homopolimérica no está ubicada directamente junto al origen de replicación en la dirección de replicación.

Por tanto, en un primer aspecto, la invención se refiere a un molde de ADN plasmídico lineal para transcripción *in vitro* que comprende

- un origen bacteriano de replicación,
- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,
- un promotor para una ARN polimerasa dependiente de ADN,

un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.

La molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención mediante linealización es un plásmido. El plásmido puede ser cualquier plásmido que sea útil para la expresión de ARN, tal como pUC, pBluescript, pGEM, pTZ, pBR322, pACYC, pSC101, pET, pGEX, PCole1, PR6K, PSC101.

La región homopolimérica está ubicada a una distancia de al menos 2200 pb, al menos 2300 pb, al menos 2400 pb desde el origen de replicación en la dirección de replicación. Preferiblemente, la región homopolimérica está ubicada a al menos 2200 pb desde el origen de replicación en la dirección de replicación.

Además, la molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación mejora el rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la producción por fermentación en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación.

La distancia desde el origen de replicación se refiere a la distancia entre el último nucleótido (extremo 3') del origen de replicación y el primer nucleótido (extremo 5') de la región homopolimérica. El experto en la técnica conoce las secuencias de orígenes de replicación usadas comúnmente, por lo que el último nucleótido (extremo 3') del origen

de replicación puede identificarse fácilmente.

Los plásmidos usados en los ejemplos de la presente solicitud contienen un origen derivado de un plásmido pUC. En algunos ejemplos, el origen de replicación tiene la secuencia identificada en SEQ ID NO: 6. Por tanto, el último nucleótido de la secuencia identificada en SEQ ID NO: 6 es el último nucleótido (extremo 3') del origen de replicación derivado de un plásmido pUC tal como se identifica en SEQ ID NO: 6.

La molécula de ADN recombinante circular cerrado covalentemente puede derivarse, por ejemplo, del vector P1140-AF2 o del vector P1140-K2.

Por tanto, la molécula de ADN recombinante circular cerrado covalentemente puede comprender una secuencia que es al menos un 90% idéntica a SEQ ID NO: 4 o SEQ ID NO: 5. Preferiblemente, la molécula de ADN recombinante circular cerrado covalentemente puede comprender una secuencia que se expone en SEQ ID NO: 4 o SEQ ID NO: 5.

La "dirección de replicación" de un origen de replicación indica la dirección en la que se realiza la replicación. Mientras que la hebra adelantada se ensambla continuamente en la dirección de replicación, la hebra retrasada se ensambla de manera discontinua, es decir, los polímeros de nucleótidos que se ensamblan en la dirección de 5' a 3' se unen para permitir un crecimiento general en la dirección de 3' a 5' de la hebra retrasada. La dirección de replicación se indica habitualmente en el mapa de vector de los plásmidos.

El inserto es una secuencia de ADN bicatenario que codifica para un ARN diana que comprende una región homopolimérica. Por consiguiente, el inserto comprende una región homopolimérica. Preferiblemente, el inserto comprende una secuencia de ácido nucleico que codifica para un ARN diana y una región homopolimérica. La región homopolimérica está ubicada en el extremo 3' del inserto. Preferiblemente, una secuencia de poli(A) y opcionalmente además una secuencia de poli(C) está ubicada en el extremo 3' del inserto. Más preferiblemente, una secuencia de poli(A) y opcionalmente además una secuencia de poli(C) está ubicada en el extremo 3' del inserto.

La secuencia del ARN diana puede comprender al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb, al menos 900 pb, al menos 1000 pb, al menos 1100 pb, al menos 1200 pb, al menos 1300 pb, al menos 1400 pb, al menos 1500 pb, al menos 1600 pb, al menos 1700 pb, al menos 1800 pb, al menos 1900 pb, al menos 2000 pb, al menos 2100 pb, al menos 2200 pb, al menos 2300 pb, al menos 2400 pb.

El ARN diana puede ser cualquier tipo de ARN descrito en el presente documento o puede codificar para proteínas o péptidos terapéuticamente activos, proteínas adyuvantes, antígenos (antígenos tumorales, antígenos patógenos (por ejemplo, seleccionados, de antígenos animales, de antígenos virales, de antígenos protozoarios, de antígenos bacterianos), antígenos alergénicos, antígenos autoinmunitarios u otros antígenos), alérgenos, anticuerpos, proteínas o péptidos inmunoestimuladores, receptores de células T específicos de antígeno, péptidos penetrantes de células, proteínas secretadas, proteínas de la membrana plasmática, proteínas citoplasmáticas o citoesqueléticas, proteínas unidas a la membrana intracelular, proteínas nucleares, proteínas asociadas con enfermedades humanas, restos de direccionamiento o aquellas proteínas codificadas por el genoma humano, para las que no se ha identificado ninguna indicación terapéutica pero que, no obstante, tienen utilidad en áreas de investigación y descubrimiento. Preferiblemente, el ARN diana es un ARN inmunoestimulante o un ARN codificante. Los ARN codificantes pueden ser, por ejemplo, ARNm, ARN vírico o ARN replicón.

El inserto puede comprender elementos adicionales tal como se describe en el presente documento, tales como una secuencia de tallo-bucle de histona, secuencias de estabilización, UTR, secuencias de IRES, etc.

La región homopolimérica comprendida en el inserto de la molécula de ADN recombinante circular cerrado covalentemente comprende al menos una secuencia de poli(A) tal como se define en las reivindicaciones y opcionalmente además al menos una secuencia de poli(C). En realizaciones incluso más preferidas, la región homopolimérica comprende una secuencia de poli(C) y una de poli(A). Preferiblemente, la secuencia de poli(A) comprende una secuencia de aproximadamente 60 a aproximadamente 250 nucleótidos de adenosina. En algunas realizaciones, la región homopolimérica comprende una secuencia de poli(A) de aproximadamente 60 a aproximadamente 70 nucleótidos de adenosina y una secuencia de poli(C) de aproximadamente 20 a 30 nucleótidos de citidina.

En una realización específica, la molécula de ADN recombinante circular cerrado covalentemente comprende una secuencia de poli(A) y una secuencia de poli(C).

La secuencia de poli(A) y la secuencia de poli(C) pueden estar conectadas mediante una secuencia ligadora. La secuencia ligadora puede ser una secuencia heteropolimérica. La secuencia ligadora puede comprender de 1 a 400, preferiblemente de 1 a 100, más preferiblemente de 2 a 50, incluso más preferiblemente de 3 a 50, lo más preferiblemente de 3 a 100 nucleótidos. La secuencia ligadora puede comprender una secuencia tal como 'tgcat'. Por consiguiente, la región homopolimérica que comprende una secuencia de poli(A) y una secuencia de poli(C)

puede tener, por ejemplo, una secuencia tal como se expone en SEQ ID NO: 1.

El término “la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación” significa que la secuencia se replica en la misma dirección que se transcribe, es decir que el extremo 5' del inserto es más cercano al origen de replicación en la dirección de la replicación que el extremo 3' del inserto. Por tanto entre el origen de replicación y la región homopolimérica existe la secuencia que codifica para un ARN diana.

Por tanto, en una molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de la transcripción del inserto es la misma que la dirección de replicación del origen de replicación, la distancia entre el origen de replicación y la región homopolimérica es mayor en comparación con una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de la transcripción del inserto es opuesta a la dirección de replicación del origen de replicación.

El rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación se mejora en la producción fermentativa en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación. Se entiende que las condiciones de fermentación de ambas moléculas de ADN recombinante circular cerrado covalentemente son las mismas.

Además, la molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación mejora el rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la producción por fermentación en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación. Se entiende que las condiciones de fermentación de ambas moléculas de ADN recombinante circular cerrado covalentemente son las mismas.

Cuando el rendimiento y/o tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación se compara con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación, se entiende que las condiciones de fermentación de ambas moléculas de ADN recombinante circular cerrado covalentemente son las mismas.

De manera correspondiente, cuando el rendimiento y/o tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación se compara con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación, se entiende que el inserto de ambas moléculas de ADN recombinante circular cerrado covalentemente es el mismo.

Tal como puede observarse a partir del mapa de vector representado en la figura 2, la orientación de P1140-AF2 es tal que la dirección de replicación del origen de aplicación es la misma que la dirección de transcripción del inserto. El mapa de vector de P1140-AF1 en la figura 3 muestra que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación. Tal como puede observarse a partir de la tabla 1, el rendimiento de plásmido para el constructo P1140-AF1 es de 3,0 mg//DO<sub>600</sub>, mientras que el rendimiento de plásmido aumenta hasta 7,5 mg//DO<sub>600</sub> cuando se usa el plásmido P1140-AF2. Por tanto, el rendimiento de plásmido se aumenta 2,5 veces para el plásmido P1140-AF2 en comparación con P1140-AF1.

Tal como puede observarse a partir del mapa de vector representado en la figura 4, la orientación de P1140-K2 es tal que la dirección de replicación del origen de replicación es la misma que la dirección de transcripción del inserto. El mapa de vector de P1140-K1 en la figura 5 muestra que la dirección de la transcripción del inserto es opuesta a la dirección de replicación del origen de replicación. Tal como puede observarse a partir de la tabla 1, el rendimiento de plásmido para el constructo P1140-K1 es de 1,9 mg//DO<sub>600</sub>, mientras que el rendimiento de plásmido aumenta hasta 9,3 mg//DO<sub>600</sub> cuando se usa el plásmido P1140-K2. Por tanto, el rendimiento de plásmido se aumenta 4,9 veces para el plásmido P1140-K2 en comparación con P1140-K1.

El rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación puede aumentarse en la producción fermentativa en al menos 1,1 veces, al menos 1,2 veces, al menos 1,3 veces, al menos 1,4 veces, al menos 1,5 veces, al menos

1,6 veces, al menos 1,7 veces, al menos 1,8 veces, al menos 1,9 veces, al menos 2,0 veces, al menos 2,1 veces, al menos 2,2 veces, al menos 2,3 veces, al menos 2,4 veces, al menos 2,5 veces, al menos 2,6 veces, al menos 2,7 veces, al menos 2,8 veces, al menos 2,9 veces, al menos 3,0 veces, al menos 3,1 veces, al menos 3,2 veces, al menos 3,3 veces, al menos 3,4 veces, al menos 3,5 veces, al menos 3,6 veces, al menos 3,7 veces, al menos 3,8 veces, al menos 3,9 veces, al menos 4,0 veces, al menos 4,1 veces, al menos 4,2 veces, al menos 4,3 veces, al menos 4,4 veces, al menos 4,5 veces, al menos 4,6 veces, al menos 4,7 veces, al menos 4,8 veces, al menos 4,9 veces, al menos 5,0 veces, al menos 5,1 veces, al menos 5,2 veces, al menos 5,3 veces, al menos 5,4 veces, al menos 5,5 veces o más en la producción fermentativa en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación. Se entiende que las condiciones de fermentación de ambas moléculas de ADN recombinante circular cerrado covalentemente son las mismas.

El rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación puede aumentar el rendimiento de la molécula de ADN recombinante circular cerrado covalentemente en la producción fermentativa en al menos 1,1 veces, al menos al menos 1,2 veces, al menos 1,3 veces, al menos 1,4 veces, al menos 1,5 veces, al menos 1,6 veces, al menos 1,7 veces, al menos 1,8 veces, al menos 1,9 veces, al menos 2,0 veces, al menos 2,1 veces, al menos 2,2 veces, al menos 2,3 veces, al menos 2,4 veces, al menos 2,5 veces, al menos 2,6 veces, al menos 2,7 veces, al menos 2,8 veces, al menos 2,9 veces, al menos 3,0 veces, al menos 3,1 veces, al menos 3,2 veces, al menos 3,3 veces, al menos 3,4 veces, al menos 3,5 veces, al menos 3,6 veces, al menos 3,7 veces, al menos 3,8 veces, al menos 3,9 veces, al menos 4,0 veces, al menos 4,1 veces, al menos 4,2 veces, al menos 4,3 veces, al menos 4,4 veces, al menos 4,5 veces, al menos 4,6 veces, al menos 4,7 veces, al menos 4,8 veces, al menos 4,9 veces, al menos 5,0 veces, al menos 5,1 veces, al menos 5,2 veces, al menos 5,3 veces, al menos 5,4 veces, al menos 5,5 veces o más en la producción por fermentación en comparación con el rendimiento de una molécula de ADN recombinante circular cerrado covalentemente en la que la región homopolimérica está ubicada a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación. Se entiende que las condiciones de fermentación de ambas moléculas de ADN recombinante circular cerrado covalentemente son las mismas.

Normalmente, el origen de replicación es de origen bacteriano. Habitualmente, el origen de replicación es unidireccional. La replicación de un origen de replicación unidireccional se lleva a cabo en una dirección. Preferiblemente, el origen de replicación es un origen de alto número de copias. El origen de replicación puede derivarse del plásmido pBR322, plásmido pUC, plásmido pMB1, plásmido ColE1, plásmido R6K, plásmido p15 A, plásmido pSC101 o fagémido F1. Los expertos en la técnica conocen los diferentes orígenes de replicación. Algunos de los orígenes de replicación mencionados anteriormente están relacionados. Por ejemplo, el origen de replicación derivado del plásmido pUC es una variante del origen de replicación derivado del plásmido pMB1 que difiere del origen de pMB1 en solo dos mutaciones que conducen a un mayor número de copias en comparación con el plásmido pMB1. Preferiblemente, el origen de replicación se deriva del plásmido pUC.

En algunas realizaciones, la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención comprende además un sitio de ensamblaje del primosoma, por ejemplo un sitio de ensamblaje del primosoma en la hebra adelantada (PAS-BH). Por ejemplo, el sitio de ensamblaje del primosoma puede comprender la secuencia SEQ ID NO: 3.

Normalmente, la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención comprende además una secuencia que codifica para un marcador de selección, en la que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa.

El término "secuencia que codifica para un marcador de selección" se refiere a una secuencia que codifica para un ARN o polipéptido que proporciona un fenotipo a la célula que contiene la secuencia que codifica para un marcador de selección que permite una selección o bien positiva o bien negativa de células que contienen la secuencia que codifica para un marcador de selección. La secuencia que codifica para un marcador de selección puede usarse para distinguir entre células transformadas y no transformadas o puede usarse para identificar células que han experimentado recombinación u otras clases de modificaciones genéticas.

El experto en la técnica conoce diferentes marcadores de selección tales como genes de resistencia a antibióticos, por ejemplo, gen de resistencia a kanamicina. Alternativamente, puede usarse un marcador seleccionable de sacarosa como marcador de selección. Un marcador seleccionable de sacarosa es, por ejemplo, el sistema ARN-OUT de Nature Technology Corporation (descrito Luke J *et al.*; Vaccine. 30 de octubre de 2009;27(46):6454-9). Los vectores contienen y expresan un ARN antisentido ARN-OUT de 150 pb. ARN-OUT reprime la expresión de un marcador de contraselección expresado constitutivamente integrado cromosómicamente (sacB), lo que permite la selección del plásmido en sacarosa. SacB codifica para una levansucrasa que es tóxica en presencia de sacarosa.

La molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de

ADN plasmídico lineal de la invención puede comprender al menos un promotor para la transcripción *in vitro*, tal como T3, Sp6 o T7, así como un terminador de transcripción, por ejemplo *trpA*. Además, la molécula de ADN recombinante circular cerrado covalentemente puede comprender una pluralidad de sitios de inserto y los sitios de inserto pueden agruparse como parte de un sitio de clonación múltiple. La molécula de ADN recombinante circular cerrado covalentemente también puede comprender más de un sitio de clonación múltiple, que pueden ser idénticos.

En un aspecto adicional, la presente invención se refiere a un método de producción de un molde de ADN plasmídico lineal para transcripción *in vitro* que comprende la etapa de:

- 10 linealizar un ADN plasmídico, en el que el ADN plasmídico comprende
  - un origen bacteriano de replicación,
  - una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,
  - un promotor para una ARN polimerasa dependiente de ADN,
  - un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.
- 25 En realizaciones específicas, se reduce preferiblemente el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico de la invención en la que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación en comparación con el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está ubicado a una distancia de menos de 500 pb desde el origen de replicación en la dirección de replicación.

El tiempo de fermentación reducido significa que el mismo rendimiento se logra dentro de un tiempo de fermentación más corto.

- 35 Se reduce preferiblemente el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende una región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación en comparación con el tiempo de fermentación de una molécula de ADN recombinante circular cerrado covalentemente en la que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es opuesta a la dirección de replicación del origen de replicación.

Esto significa, por ejemplo, que el tiempo de fermentación se reduce preferiblemente en más del 5%, en más del 10%, en más del 15%, en más del 16%, en más del 17%, en más del 18%, en más del 19%, en más del 20%, en más del 21%, en más del 22%, en más del 23%, en más del 24%, en más del 25%, en más del 26%, en más del 27%, en más del 28%, en más del 29%, en más del 30%, en más del 31%, en más del 32%. Preferiblemente, el tiempo de fermentación se reduce entre el 5% y el 70%, más preferiblemente entre el 10% y el 50%, incluso más preferiblemente entre el 15% y el 35%.

- 50 Un aspecto adicional de la invención se refiere a un método para la producción fermentativa de una molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención que comprende las etapas de:

(a) proporcionar un microorganismo que comprende la molécula de ADN recombinante circular cerrado covalentemente según la invención;

(b) fermentar el microorganismo de la etapa (a).

En una realización específica, el método comprende además la siguiente etapa de:

- 60 (c) aislar la molécula de ADN recombinante circular cerrado covalentemente a partir del microorganismo de la etapa (b).

El experto en la técnica conoce las técnicas convencionales usadas para el aislamiento de moléculas de ADN recombinante circular cerrado covalentemente.

En realizaciones preferidas, el microorganismo es una bacteria (por ejemplo, *E. coli*) que contiene un plásmido de

ADN de alto número de copias inducible por temperatura recombinante circular cerrado covalentemente y la etapa (b) comprende las siguientes etapas de:

(i) hacer crecer las bacterias que contienen un plásmido de ADN de alto número de copias inducible por temperatura recombinante circular cerrado covalentemente a una temperatura en el intervalo de 25°C a 32°C durante la fase de crecimiento de la fase semicontinua en la que el sustrato se suministra a una tasa tal que la tasa de crecimiento es de  $\mu =$  de 0,05 a 0,3 h<sup>-1</sup> durante la fase semicontinua,

(ii) inducir la producción de dicho plásmido de ADN después de la fase de crecimiento aumentando la temperatura hasta de 36°C a 45°C; y

(iii) continuar la fermentación a la temperatura aplicada en (ii) para acumular el dicho plásmido de ADN.

En realizaciones más específicas, la etapa (b) comprende las siguientes etapas de:

(i) hacer crecer las bacterias que contienen un plásmido de ADN de alto número de copias inducible por temperatura recombinante circular cerrado covalentemente a una temperatura de aproximadamente 30°C durante la fase de crecimiento de la fase semicontinua en la que el sustrato se suministra a una tasa tal que la tasa de crecimiento es de  $\mu =$  de 0,05 a 0,3 h<sup>-1</sup> durante la fase semicontinua,

(ii) inducir la producción de dicho plásmido de ADN después de la fase de crecimiento aumentando la temperatura hasta aproximadamente 42°C; y

(iii) continuar la fermentación a la temperatura aplicada en (ii) para acumular el dicho plásmido de ADN.

Este método de fermentación semicontinua se describe con detalle en el documento EP 1 781 800 B1 que se denomina en el presente documento como fermentación HyperGRO. Habitualmente el método se lleva a cabo con moléculas de ADN recombinante circular cerrado covalentemente inducible por temperatura (por ejemplo, plásmidos que contienen origen de pUC o pMM1) a tasa de crecimiento celular restringida y temperatura reducida durante la fase de crecimiento; luego se induce la producción de plásmidos aumentando la temperatura. De ese modo aumenta el rendimiento de la fermentación del ADN plasmídico, mientras que se mantiene o se mejora la estabilidad de la molécula de ADN recombinante circular cerrado covalentemente.

La fase de crecimiento en la etapa (b) (i) puede realizarse a temperaturas desde 25 hasta 37°C, preferiblemente a de 30 a 37°C. Para plásmidos de alto número de copias, la fase de crecimiento se realiza preferiblemente a de 30 a 32°C, más preferiblemente a 30°C. La fase de inducción en la etapa (b) (ii) puede llevarse a cabo a temperaturas desde 33 hasta 45°C, preferiblemente desde 37 hasta 42°C, lo más preferiblemente a 42°C.

El experto en la técnica conoce las diferentes estrategias de alimentación limitativa en la etapa (b) (i). Por ejemplo, puede emplearse una estrategia de alimentación exponencial limitativa del carbono. En resumen, se consume una cantidad inicial de sustrato de carbono durante la fase discontinua a una tasa de crecimiento específica  $\mu_{\max}$ . Tras agotarse el sustrato de carbono, comienza la fase semicontinua y se añade el nutriente de alimentación automáticamente a una tasa determinada por la siguiente ecuación

$$F(t) = \frac{\mu X_B V_B}{S_r Y_{X/S}} e^{\mu t}$$

en la que  $\mu$  = tasa de crecimiento específica deseada durante fase semicontinua,

$X_B$  = concentración de biomasa al final de la fase discontinua, g de DCW (peso seco de la célula)/l

$V_B$  = volumen de líquido inicial de cultivo, l;

$S_r$  = concentración de sustrato limitativa en el medio de alimentación de nutrientes, g/l;

$Y_{X/S}$  = coeficiente de rendimiento de biomasa del sustrato, g/g,

$t$  = tiempo desde el comienzo de la fase semicontinua

Es una ventaja de la molécula de ADN recombinante circular cerrado covalentemente adecuada para la preparación del molde de ADN plasmídico lineal de la invención que la secuencia de la molécula de ADN recombinante circular cerrado covalentemente permanezca estable durante la producción por fermentación.

Por ejemplo, la región homopolimérica permanece estable durante la producción por fermentación. En particular, la secuencia de poli(A) y/o secuencia de poli(C) permanece estable durante la producción por fermentación. El término

"permanece estable" tal como se usa en el presente documento significa que no existen mutaciones, es decir, tales como sustituciones de nucleótidos, deleciones de nucleótidos o adiciones de nucleótidos introducidas durante la producción por fermentación.

- 5 Finalmente la invención se refiere al uso de la molécula de ADN recombinante circular cerrado covalentemente según la invención para la transcripción *in vitro* de ARN.

La presente invención proporciona un método de producción de ARN que comprende las etapas de:

- 10 (a) obtener ARN mediante transcripción *in vitro* dependiente de ADN de un molde de ADN plasmídico linealizado usando una ARN polimerasa dependiente de ADN, en el que el ADN plasmídico comprende
- un origen bacteriano de replicación,
  - 15 - una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,
  - un promotor para una ARN polimerasa dependiente de ADN,
  - 20 - un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación; y
  - 25 (b) purificar el ARN.

Esta invención se realizó con el apoyo del gobierno según el acuerdo n.º HR0011-11-3-0001 concedido por DARPA. El gobierno tiene ciertos derechos con respecto a la invención.

### Ejemplos

Los ejemplos que se muestran a continuación son meramente ilustrativos y describirán la presente invención de una manera adicional. Estos ejemplos no se interpretarán como que limitan la presente invención a los mismos.

Plásmidos:

Se generaron diferentes vectores que codifican para ARNm optimizado en GC que codifica para H1N1(Paises Bajos 2009)-HA (tal como se muestra en la figura 6, SEQ ID NO: 2).

P1140 (mostrado en la figura 1):

La orientación de P1140 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Como marcador de selección, el vector codifica para kanamicina (indicada en el mapa de vector como KanR).

P1140-AF1 (mostrado en la figura 3):

La orientación de P1140-AF1 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Como marcador de selección, el vector codifica para el ARN-OUT, un ARN antisentido que reprime la expresión de un marcador de contraselección expresado constitutivamente integrado cromosómicamente (sacB), lo que permite la selección del plásmido en sacarosa (indicado en el mapa de vector como ARN-OUT).

P1140-AF2 (mostrado en la figura 2):

La orientación de P1140-AF2 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es la misma que la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Como marcador de selección, el vector codifica para el ARN-OUT, un ARN antisentido que reprime la expresión de un marcador de contraselección expresado constitutivamente integrado cromosómicamente (sacB), lo que permite la

selección del plásmido en sacarosa (indicado en el mapa de vector como ARN-OUT).

P1140-K1 (tal como se muestra en la figura 5):

- 5 La orientación de P1140-K1 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es opuesta a la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

Como marcador de selección, el vector codifica para kanamicina (indicada en el mapa de vector como KanR).

10

P1140-K2 (mostrado en la figura 4):

La orientación de P1140-K2 es tal que la dirección de replicación del origen de replicación (indicado en el mapa de vector como origen de pUC) es la misma que la dirección de transcripción del inserto (indicado en el mapa de vector como ARN).

15

Como marcador de selección, el vector codifica para kanamicina (indicada en el mapa de vector como KanR).

Fermentación HyperGRO:

20

Se transformaron los plásmidos P1140 (figura 1), P1140-AF1 (figura 3), P1140-AF2 (figura 2), P1140-K1 (figura 5) y P1140-K2 (figura 4) al interior de cepas de *E. coli* derivadas de DH5 $\alpha$ , DH5 $\alpha$ , DH5 $\alpha$  dcm- o DH5 $\alpha$  dcm-att $\lambda$ ::P5/6 6/6-RNA-IN-SacB (tal como se indica en la tabla 1) y se propagaron mediante fermentación HyperGRO. Este método de fermentación semicontinua se describe con detalle en el documento EP 1 781 800. En la fermentación HyperGRO, el vector P1140 fue de crecimiento lento y de mal rendimiento. En la fermentación HyperGRO, las líneas celulares P1140-AF1 y P1140-K1 también fueron de crecimiento lento y de bajo rendimiento (véase la tabla 1). En contraste, las líneas celulares P1140-AF2 y P1140-K2 tenían un crecimiento normal y altos rendimientos de producción (véase la tabla 1). Mediante la secuenciación de los plásmidos pudo confirmarse que la secuencia de poli(A) era estable durante la fermentación en todos los plásmidos sometidos a prueba.

25

30

Tabla 1: resumen de la fermentación HyperGRO

| Plásmido  | Línea celular                                         | Rendimiento de plásmido a 42°C (mg/l) | DO600 de la recogida | Duración de la fermentación (horas) |
|-----------|-------------------------------------------------------|---------------------------------------|----------------------|-------------------------------------|
| P1140     | DH5 $\alpha$ dcm-                                     | 75                                    | 68                   | 53                                  |
| P1140-AF1 | DH5 $\alpha$ dcm-att $\lambda$ ::P5/6 6/6-RNA-IN-SacB | 223                                   | 75                   | 45                                  |
| P1140-AF2 | DH5 $\alpha$ dcm-att $\lambda$ ::P5/6 6/6-RNA-IN-SacB | 684                                   | 91                   | 38                                  |
| P1140-K1  | DH5 $\alpha$                                          | 130                                   | 68                   | 38                                  |
| P1140-K2  | DH5 $\alpha$                                          | 798                                   | 86                   | 36                                  |

Fermentación discontinua:

35

Estos resultados se han confirmado usando fermentación discontinua convencional a 37°C, lo que demuestra que el aumento del rendimiento de plásmido es independiente del protocolo de fermentación específico.

# REIVINDICACIONES

1. Método de producción de ARN que comprende las etapas de:
  - 5 (a) obtener ARN mediante transcripción *in vitro* dependiente de ADN de un molde de ADN plasmídico linealizado usando una ARN polimerasa dependiente de ADN, en el que el ADN plasmídico comprende
    - un origen bacteriano de replicación,
    - 10 - una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,
    - un promotor para una ARN polimerasa dependiente de ADN,
    - 15 - un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación; y
    - 20 (b) purificar el ARN.
  2. Método según la reivindicación 1, en el que la secuencia de poli(A) del ADN plasmídico comprende un tramo de 60 a 250 nucleótidos de adenosina.
  3. Método según una cualquiera de las reivindicaciones 1 a 2, en el que la ARN polimerasa dependiente de ADN usada en la etapa (a) se selecciona del grupo que consiste en ARN polimerasa T3, ARN polimerasa T7 y ARN polimerasa SP6.
  - 30 4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que reactivos adicionales usados en la etapa (a) son ribonucleósidos trifosfatos (NTP) para las cuatro bases adenina, citosina, guanina y uracilo.
  5. Método según la reivindicación 4, en el que se reemplaza una parte o la totalidad de al menos un ribonucleósido trifosfato con un nucleósido trifosfato modificado, en el que el nucleósido trifosfato modificado se selecciona preferiblemente del grupo que consiste en pseudouridina-5'-trifosfato, 1-metilpseudouridina-5'-trifosfato, 2-tiouridina-5'-trifosfato, 4-tiouridina-5'-trifosfato y 5-metilcitidina-5'-trifosfato.
  - 35 6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que un reactivo adicional usado en la etapa (a) es (i) un análogo de caperuza y/o (ii) un inhibidor de ribonucleasa (ARNasa) y/o (iii) una pirofosfatasa y/o (iv) MgCl<sub>2</sub> y/o (v) un tampón que contiene opcionalmente antioxidantes y/o poliaminas.
  - 40 7. Método según una cualquiera de las reivindicaciones 1 a 6, en el que la purificación comprende una digestión con molde de ADN y/o extracción con fenol-cloroformo y/o precipitación con LiCl y/o HPLC.
  - 45 8. Método según una cualquiera de las reivindicaciones 1 a 7, en el que el ARN es ARNm.
  9. Método según la reivindicación 8, en el que el ARNm comprende una secuencia de marco de lectura abierto y una 5'-UTR y/o 3'-UTR.
  - 50 10. Método según la reivindicación 9, en el que la secuencia de marco de lectura abierto codifica para un antígeno patógeno o un fragmento del mismo, en el que el antígeno patógeno es preferiblemente un antígeno de superficie, más preferiblemente un antígeno de superficie ubicado en la superficie de un virus.
  - 55 11. Método según la reivindicación 10, en el que el antígeno patógeno se deriva de un patógeno asociado con una enfermedad infecciosa, en el que el patógeno se selecciona del grupo que consiste en *Acinetobacter baumannii*, *Anaplasma* genus, *Anaplasma phagocytophilum*, *Ancylostoma braziliense*, *Ancylostoma duodenale*, *Arcanobacterium haemolyticum*, *Ascaris lumbricoides*, género *Aspergillus*, *Astroviridae*, género *Babesia*, *Bacillus anthracis*, *Bacillus cereus*, *Bartonella henselae*, virus BK, *Blastocystis hominis*, *Blastomyces dermatitidis*, *Bordetella pertussis*, *Borrelia burgdorferi*, género *Borrelia*, *Borrelia* spp., género *Brucella*, *Brugia malayi*, familia *Bunyaviridae*, *Burkholderia cepacia* y otras especies de *Burkholderia*, *Burkholderia mallei*, *Burkholderia pseudomallei*, familia *Caliciviridae*, género *Campylobacter*, *Candida albicans*, *Candida* spp., *Chlamydia trachomatis*, *Chlamydomydia pneumoniae*, *Chlamydomydia psittaci*, prión de CJD, *Clonorchis sinensis*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium* spp., *Clostridium tetani*, *Coccidioides* spp., coronavirus, *Corynebacterium diphtheriae*, *Coxiella burnetii*, virus de la fiebre hemorrágica de Crimea-Congo, *Cryptococcus neoformans*, género *Cryptosporidium*,
  - 60
  - 65

citomegalovirus (CMV), virus del dengue (DEN-1, DEN-2, DEN-3 y DEN-4), *Dientamoeba fragilis*, virus del Ébola (EBOV), género *Echinococcus*, *Ehrlichia chaffeensis*, *Ehrlichia ewingii*, género *Ehrlichia*, *Entamoeba histolytica*, género *Enterococcus*, género *Enterovirus*, enterovirus, principalmente virus de Coxsackie A y enterovirus 71 (E71), *Epidermophyton* spp., virus de Epstein-Barr (VEB), *Escherichia coli* O157: H7, O111 y O104: H4, *Fasciola hepatica* y *Fasciola gigantica*, prión FFI, superfamilia *Filarioidea*, Flavivirus, *Francisella tularensis*, género *Fusobacterium*, *Geotrichum candidum*, *Giardia intestinalis*, *Gnathostoma* spp., prión GSS, virus Guanarito, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, henipavirus (virus Hendra, virus Nipah), virus de la hepatitis A, virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), virus de la hepatitis D, virus de la hepatitis E, virus del herpes simple 1 y 2 (VHS-1 y VHS-2), *Histoplasma capsulatum*, VIH (virus de la inmunodeficiencia humana), *Hortaea werneckii*, Bocavirus humano (BoVh), herpesvirus humano 6 (HVH-6) y herpesvirus humano 7 (HVH-7), metapneumovirus humano (MPVh), virus del papiloma humano (VPH), virus de la parainfluenza humana (VPIH), virus de la encefalitis japonesa, virus JC, virus Junin, *Kingella kingae*, *Klebsiella granulomatis*, prión Kuru, virus de Lassa, *Legionella pneumophila*, género *Leishmania*, género *Leptospira*, *Listeria monocytogenes*, virus de la coriomeningitis linfocítica (VCML), virus Machupo, *Malassezia* spp., virus de Marburgo, virus del sarampión, *Metagonimus yokagawai*, filo de microsporidios, virus del molusco contagioso (VMC), virus de las paperas, *Mycobacterium leprae* y *Mycobacterium lepromatosis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Mycoplasma pneumoniae*, *Naegleria fowleri*, *Necator americanus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Nocardia asteroides*, *Nocardia* spp., *Onchocerca volvulus*, *Orientia tsutsugamushi*, familia *Orthomyxoviridae* (gripe), *Paracoccidioides brasiliensis*, *Paragonimus* spp., *Paragonimus westermani*, parvovirus B19, género *Pasteurella*, género *Plasmodium*, *Pneumocystis jirovecii*, virus de la poliomyelitis, virus de la rabia, virus respiratorio sincitial (VRS), rinovirus, *Rickettsia akari*, género *Rickettsia*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia typhi*, virus de la fiebre del valle del Rift, rotavirus, virus de la rubéola, virus de Sabiá, género *Salmonella*, *Sarcoptes scabiei*, coronavirus SARS, género *Schistosoma*, género *Shigella*, virus Sin Nombre, Hantavirus, *Sporothrix schenckii*, género *Staphylococcus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Strongyloides stercoralis*, género *Taenia*, *Taenia solium*, virus de la encefalitis transmitida por garrapatas (VETG), *Toxocara canis* o *Toxocara cati*, *Toxoplasma gondii*, *Treponema pallidum*, *Trichinella spiralis*, *Trichomonas vaginalis*, *Trichophyton* spp., *Trichuris trichiura*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Ureaplasma urealyticum*, virus de la varicela zoster (VVZ), viruela mayor o viruela menor, prion vCJD, virus de la encefalitis equina venezolana, *Vibrio cholerae*, virus del Nilo Occidental, virus de la encefalitis equina occidental, *Wuchereria bancrofti*, virus de la fiebre amarilla, *Yersinia enterocolitica*, *Yersinia pestis* y *Yersinia pseudotuberculosis*.

12. Molde de ADN plasmídico lineal para transcripción *in vitro* que comprende

- un origen bacteriano de replicación,

- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,

- un promotor para una ARN polimerasa dependiente de ADN,

- un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.

13. Molde de ADN plasmídico lineal para transcripción *in vitro* según la reivindicación 12, en el que la secuencia de poli(A) comprende un tramo de 60 a 250 nucleótidos de adenosina.

14. Molde de ADN plasmídico lineal para transcripción *in vitro* según la reivindicación 12 ó 13, en el que el ADN plasmídico comprende además un sitio de ensamblaje del primosoma.

15. Molde de ADN plasmídico lineal para transcripción *in vitro* según una cualquiera de las reivindicaciones 11 a 14, en el que el ADN plasmídico comprende una secuencia de marco de lectura abierto y una 5'-UTR y/o 3'-UTR de un ARNm.

16. Molde de ADN plasmídico lineal para transcripción *in vitro* según la reivindicación 15, en el que la secuencia de marco de lectura abierto codifica para un antígeno patógeno o un fragmento del mismo, en el que el antígeno patógeno es preferiblemente un antígeno de superficie, más preferiblemente un antígeno de superficie ubicado en la superficie de un virus.

17. Molde de ADN plasmídico lineal para transcripción *in vitro* según la reivindicación 16, en el que el antígeno patógeno se deriva de un patógeno asociado con una enfermedad infecciosa, en el que el patógeno se

selecciona del grupo que consiste en *Acinetobacter baumannii*, *Anaplasma* genus, *Anaplasma phagocytophilum*, *Ancylostoma braziliense*, *Ancylostoma duodenale*, *Arcanobacterium haemolyticum*, *Ascaris lumbricoides*, género *Aspergillus*, *Astroviridae*, género *Babesia*, *Bacillus anthracis*, *Bacillus cereus*, *Bartonella henselae*, virus BK, *Blastocystis hominis*, *Blastomyces dermatitidis*, *Bordetella pertussis*, *Borrelia burgdorferi*, género *Borrelia*, *Borrelia* spp., género *Brucella*, *Brugia malayi*, familia *Bunyaviridae*, *Burkholderia cepacia* y otras especies de *Burkholderia*, *Burkholderia mallei*, *Burkholderia pseudomallei*, familia *Caliciviridae*, género *Campylobacter*, *Candida albicans*, *Candida* spp., *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, príon de CJD, *Clonorchis sinensis*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium* spp., *Clostridium tetani*, *Coccidioides* spp., coronavirus, *Corynebacterium diphtheriae*, *Coxiella burnetii*, virus de la fiebre hemorrágica de Crimea-Congo, *Cryptococcus neoformans*, género *Cryptosporidium*, citomegalovirus (CMV), virus del dengue (DEN-1, DEN-2, DEN-3 y DEN-4), *Dientamoeba fragilis*, virus del Ébola (EBOV), género *Echinococcus*, *Ehrlichia chaffeensis*, *Ehrlichia ewingii*, género *Ehrlichia*, *Entamoeba histolytica*, género *Enterococcus*, género *Enterovirus*, enterovirus, principalmente virus de Coxsackie A y enterovirus 71 (E71), *Epidermophyton* spp., virus de Epstein-Barr (VEB), *Escherichia coli* O157: H7, O111 y O104: H4, *Fasciola hepatica* y *Fasciola gigantica*, príon FFI, superfamilia *Filarioidea*, Flavivirus, *Francisella tularensis*, género *Fusobacterium*, *Geotrichum candidum*, *Giardia intestinalis*, *Gnathostoma* spp., príon GSS, virus Guanarito, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, henipavirus (virus Hendra, virus Nipah), virus de la hepatitis A, virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), virus de la hepatitis D, virus de la hepatitis E, virus del herpes simple 1 y 2 (VHS-1 y VHS-2), *Histoplasma capsulatum*, VIH (virus de la inmunodeficiencia humana), *Hortaea werneckii*, Bocavirus humano (BoVh), herpesvirus humano 6 (HVH-6) y herpesvirus humano 7 (HVH-7), metapneumovirus humano (MPVh), virus del papiloma humano (VPH), virus de la parainfluenza humana (VPIH), virus de la encefalitis japonesa, virus JC, virus Junin, *Kingella kingae*, *Klebsiella granulomatis*, príon Kuru, virus de Lassa, *Legionella pneumophila*, género *Leishmania*, género *Leptospira*, *Listeria monocytogenes*, virus de la coriomeningitis linfocítica (VCML), virus Machupo, *Malassezia* spp., virus de Marburgo, virus del sarampión, *Metagonimus yokagawai*, filo de microsporidios, virus del molusco contagioso (VMC), virus de las paperas, *Mycobacterium leprae* y *Mycobacterium lepromatosis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Mycoplasma pneumoniae*, *Naegleria fowleri*, *Necator americanus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Nocardia asteroides*, *Nocardia* spp, *Onchocerca volvulus*, *Orientia tsutsugamushi*, familia *Orthomyxoviridae* (gripe), *Paracoccidioides brasiliensis*, *Paragonimus* spp, *Paragonimus westermani*, parvovirus B19, género *Pasteurella*, género *Plasmodium*, *Pneumocystis jirovecii*, virus de la poliomieltis, virus de la rabia, virus respiratorio sincitial (VRS), rinovirus, *Rickettsia akari*, género *Rickettsia*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia typhi*, virus de la fiebre del valle del Rift, rotavirus, virus de la rubéola, virus de Sabiá, género *Salmonella*, *Sarcoptes scabiei*, coronavirus SARS, género *Schistosoma*, género *Shigella*, virus Sin Nombre, Hantavirus, *Sporothrix schenckii*, género *Staphylococcus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Strongyloides stercoralis*, género *Taenia*, *Taenia solium*, virus de la encefalitis transmitida por garrapatas (VETG), *Toxocara canis* o *Toxocara cati*, *Toxoplasma gondii*, *Treponema pallidum*, *Trichinella spiralis*, *Trichomonas vaginalis*, *Trichophyton* spp., *Trichuris trichiura*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Ureaplasma urealyticum*, virus de la varicela zoster (VZV), viruela mayor o viruela menor, príon vCJD, virus de la encefalitis equina venezolana, *Vibrio cholerae*, virus del Nilo Occidental, virus de la encefalitis equina occidental, *Wuchereria bancrofti*, virus de la fiebre amarilla, *Yersinia enterocolitica*, *Yersinia pestis* y *Yersinia pseudotuberculosis*.

18. Método de producción de un molde de ADN plasmídico lineal para transcripción *in vitro* que comprende la etapa de:

linealizar un ADN plasmídico, en el que el ADN plasmídico comprende

- un origen bacteriano de replicación,
- una secuencia que codifica para un marcador de selección, en el que la secuencia es un gen de resistencia a kanamicina o el marcador es un marcador seleccionable de sacarosa,
- un promotor para una ARN polimerasa dependiente de ADN,
- un inserto que comprende una región homopolimérica, en el que la región homopolimérica comprende al menos una secuencia de poli(A) que comprende un tramo de al menos 60 nucleótidos de adenosina, en el que la región homopolimérica está ubicada a una distancia de al menos 2200 pb desde el origen de replicación en la dirección de replicación, y en el que el inserto que comprende la región homopolimérica está orientado de manera que la dirección de transcripción del inserto es la misma que la dirección de replicación del origen de replicación.

19. Método según la reivindicación 18, en el que la secuencia de poli(A) del ADN plasmídico comprende un tramo de 60 a 250 nucleótidos de adenosina.

20. Método según una cualquiera de las reivindicaciones 18 a 19, en el que el ADN plasmídico comprende una secuencia de marco de lectura abierto y una 5'-UTR y/o 3'-UTR de un ARNm.
- 5 21. Método según la reivindicación 20, en el que la secuencia de marco de lectura abierto codifica para un antígeno patógeno o un fragmento del mismo, en el que el antígeno patógeno es preferiblemente un antígeno de superficie, más preferiblemente un antígeno de superficie ubicado en la superficie de un virus.
- 10 22. Método según la reivindicación 21, en el que el antígeno patógeno se deriva de un patógeno asociado con una enfermedad infecciosa, en el que el patógeno se selecciona del grupo que consiste en *Acinetobacter baumannii*, *Anaplasma* genus, *Anaplasma phagocytophilum*, *Ancylostoma braziliense*, *Ancylostoma duodenale*, *Arcanobacterium haemolyticum*, *Ascaris lumbricoides*, género *Aspergillus*, *Astroviridae*, género *Babesia*, *Bacillus anthracis*, *Bacillus cereus*, *Bartonella henselae*, virus BK, *Blastocystis hominis*, *Blastomyces dermatitidis*, *Bordetella pertussis*, *Borrelia burgdorferi*, género *Borrelia*, *Borrelia* spp., género *Brucella*, *Brugia malayi*, familia *Bunyaviridae*, *Burkholderia cepacia* y otras especies de *Burkholderia*, *Burkholderia mallei*, *Burkholderia pseudomallei*, familia *Caliciviridae*, género *Campylobacter*, *Candida albicans*, *Candida* spp., *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, prión de CJD, *Clonorchis sinensis*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium* spp., *Clostridium tetani*, *Coccidioides* spp., coronavirus, *Corynebacterium diphtheriae*, *Coxiella burnetii*, virus de la fiebre hemorrágica de Crimea-Congo, *Cryptococcus neoformans*, género *Cryptosporidium*, citomegalovirus (CMV), virus del dengue (DEN-1, DEN-2, DEN-3 y DEN-4), *Dientamoeba fragilis*, virus del Ébola (EBOV), género *Echinococcus*, *Ehrlichia chaffeensis*, *Ehrlichia ewingii*, género *Ehrlichia*, *Entamoeba histolytica*, género *Enterococcus*, género *Enterovirus*, enterovirus, principalmente virus de Cocksackie A y enterovirus 71 (E71), *Epidermophyton* spp., virus de Epstein-Barr (VEB), *Escherichia coli* O157: H7, O111 y O104: H4, *Fasciola hepatica* y *Fasciola gigantica*, prión FFI, superfamilia *Filarioidea*, Flavivirus, *Francisella tularensis*, género *Fusobacterium*, *Geotrichum candidum*, *Giardia intestinalis*, *Gnathostoma* spp., prión GSS, virus Guanarito, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, henipavirus (virus Hendra, virus Nipah), virus de la hepatitis A, virus de la hepatitis B (VHB), virus de la hepatitis C (VHC), virus de la hepatitis D, virus de la hepatitis E, virus del herpes simple 1 y 2 (VHS-1 y VHS-2), *Histoplasma capsulatum*, VIH (virus de la inmunodeficiencia humana), *Hortaea werneckii*, Bocavirus humano (BoVh), herpesvirus humano 6 (HVH-6) y herpesvirus humano 7 (HVH-7), metapneumovirus humano (MPVh), virus del papiloma humano (VPH), virus de la parainfluenza humana (VPIH), virus de la encefalitis japonesa, virus JC, virus Junin, *Kingella kingae*, *Klebsiella granulomatis*, prión Kuru, virus de Lassa, *Legionella pneumophila*, género *Leishmania*, género *Leptospira*, *Listeria monocytogenes*, virus de la coriomeningitis linfocítica (VCML), virus Machupo, *Malassezia* spp., virus de Marburgo, virus del sarampión, *Metagonimus yokagawai*, filo de microsporidios, virus del molusco contagioso (VMC), virus de las paperas, *Mycobacterium leprae* y *Mycobacterium lepromatosis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Mycoplasma pneumoniae*, *Naegleria fowleri*, *Necator americanus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Nocardia asteroides*, *Nocardia* spp, *Onchocerca volvulus*, *Orientia tsutsugamushi*, familia *Orthomyxoviridae* (gripe), *Paracoccidioides brasiliensis*, *Paragonimus* spp, *Paragonimus westermani*, parvovirus B19, género *Pasteurella*, género *Plasmodium*, *Pneumocystis jirovecii*, virus de la poliomieltis, virus de la rabia, virus respiratorio sincicial (VRS), rinovirus, *Rickettsia akari*, género *Rickettsia*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia typhi*, virus de la fiebre del valle del Rift, rotavirus, virus de la rubéola, virus de Sabiá, género *Salmonella*, *Sarcoptes scabiei*, coronavirus SARS, género *Schistosoma*, género *Shigella*, virus Sin Nombre, Hantavirus, *Sporothrix schenckii*, género *Staphylococcus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Strongyloides stercoralis*, género *Taenia*, *Taenia solium*, virus de la encefalitis transmitida por garrapatas (VETG), *Toxocara canis* o *Toxocara cati*, *Toxoplasma gondii*, *Treponema pallidum*, *Trichinella spiralis*, *Trichomonas vaginalis*, *Trichophyton* spp., *Trichuris trichiura*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Ureaplasma urealyticum*, virus de la varicela zoster (VVZ), viruela mayor o viruela menor, prion vCJD, virus de la encefalitis equina venezolana, *Vibrio cholerae*, virus del Nilo Occidental, virus de la encefalitis equina occidental, *Wuchereria bancrofti*, virus de la fiebre amarilla, *Yersinia enterocolitica*, *Yersinia pestis* y *Yersinia pseudotuberculosis*.
- 45
- 50

Figura 1

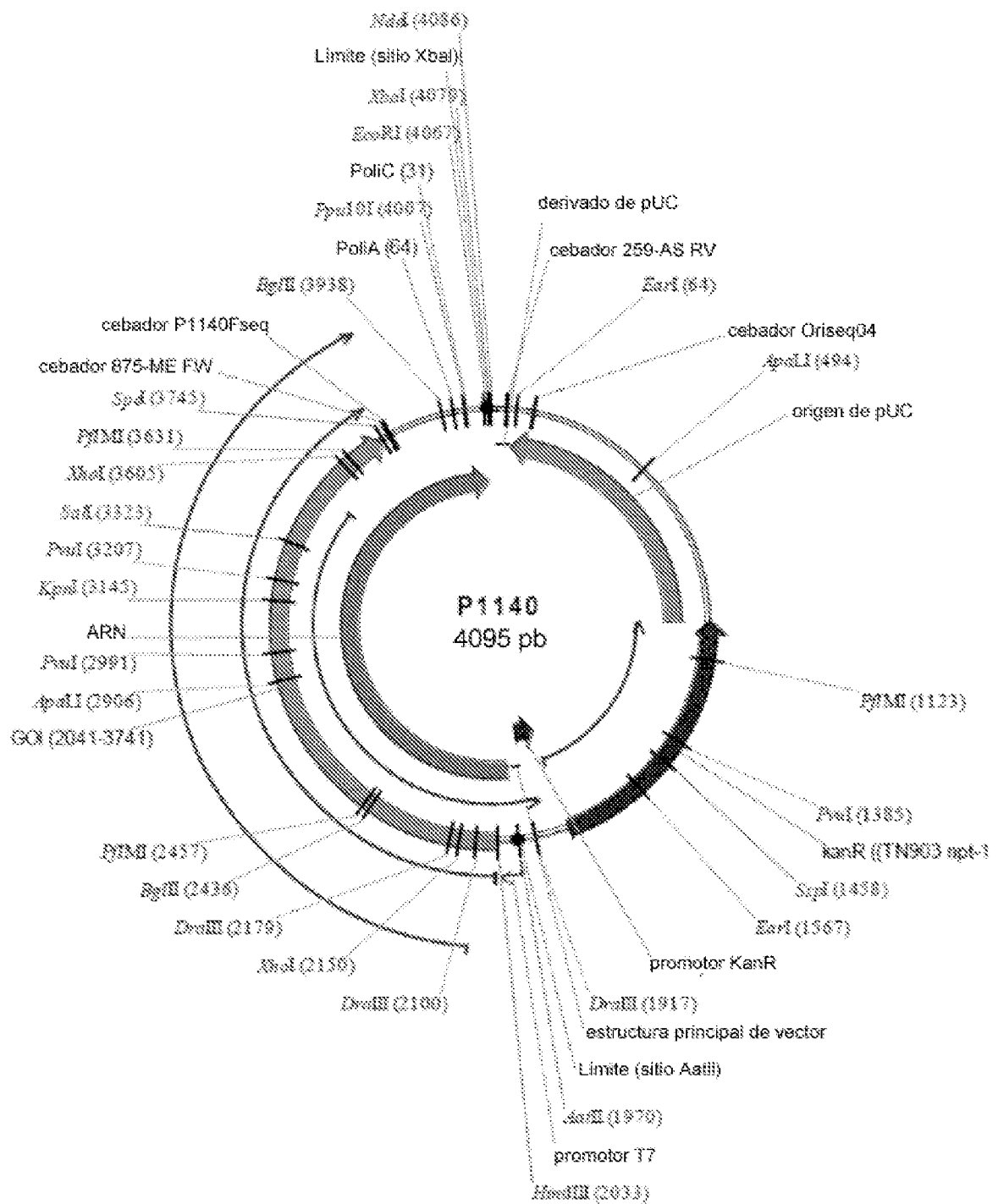


Figura 2

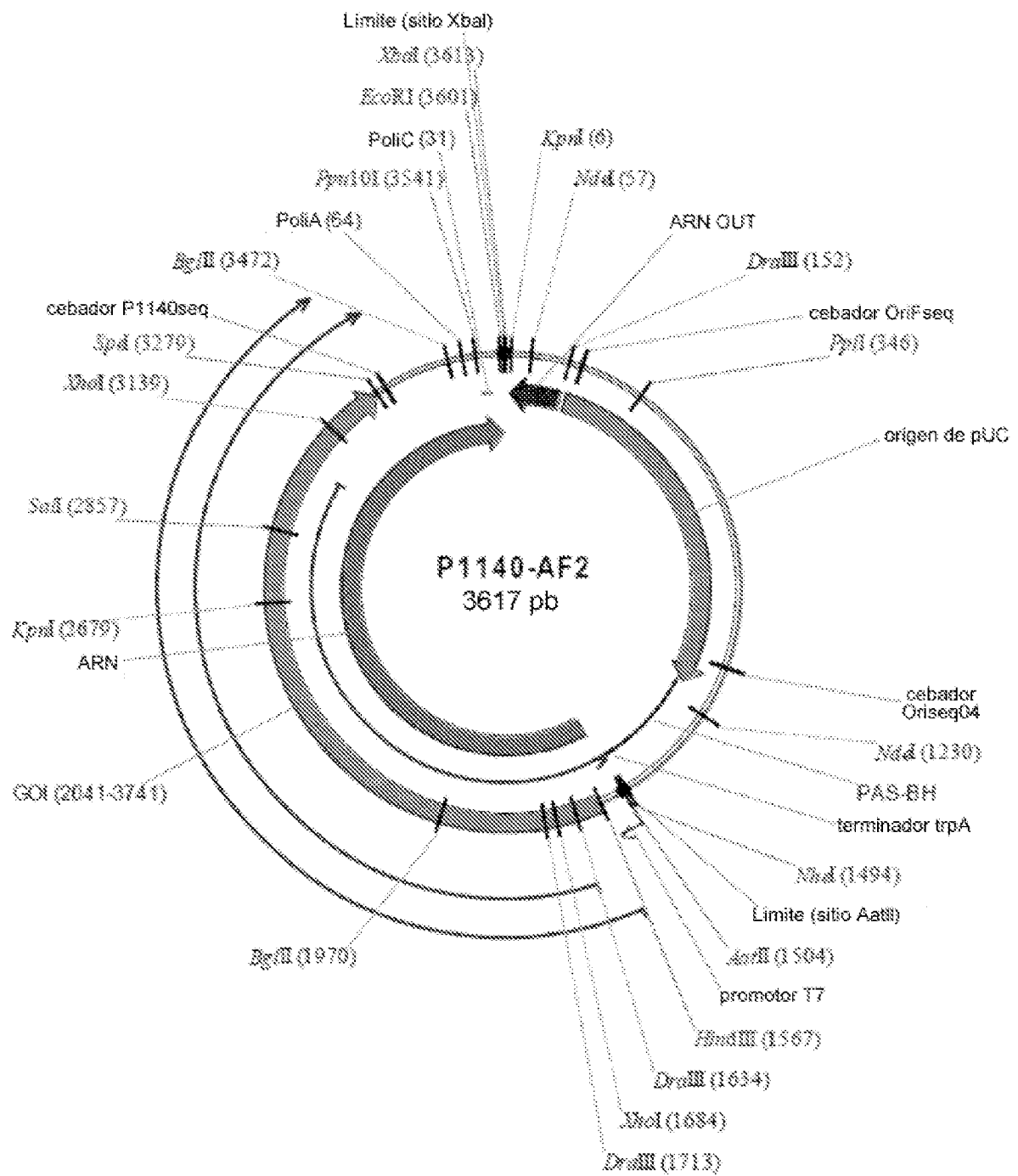


Figura 3

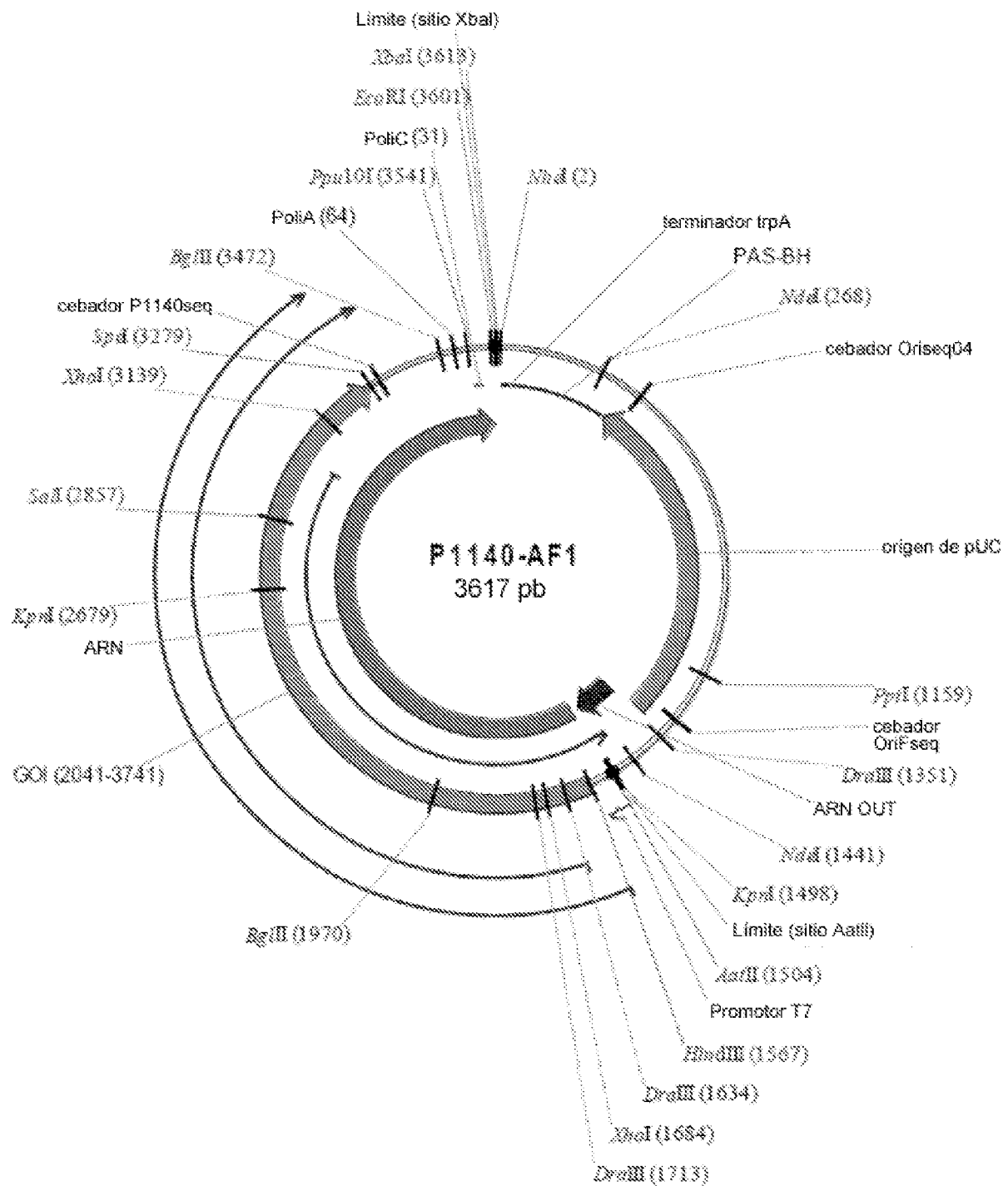


Figura 4

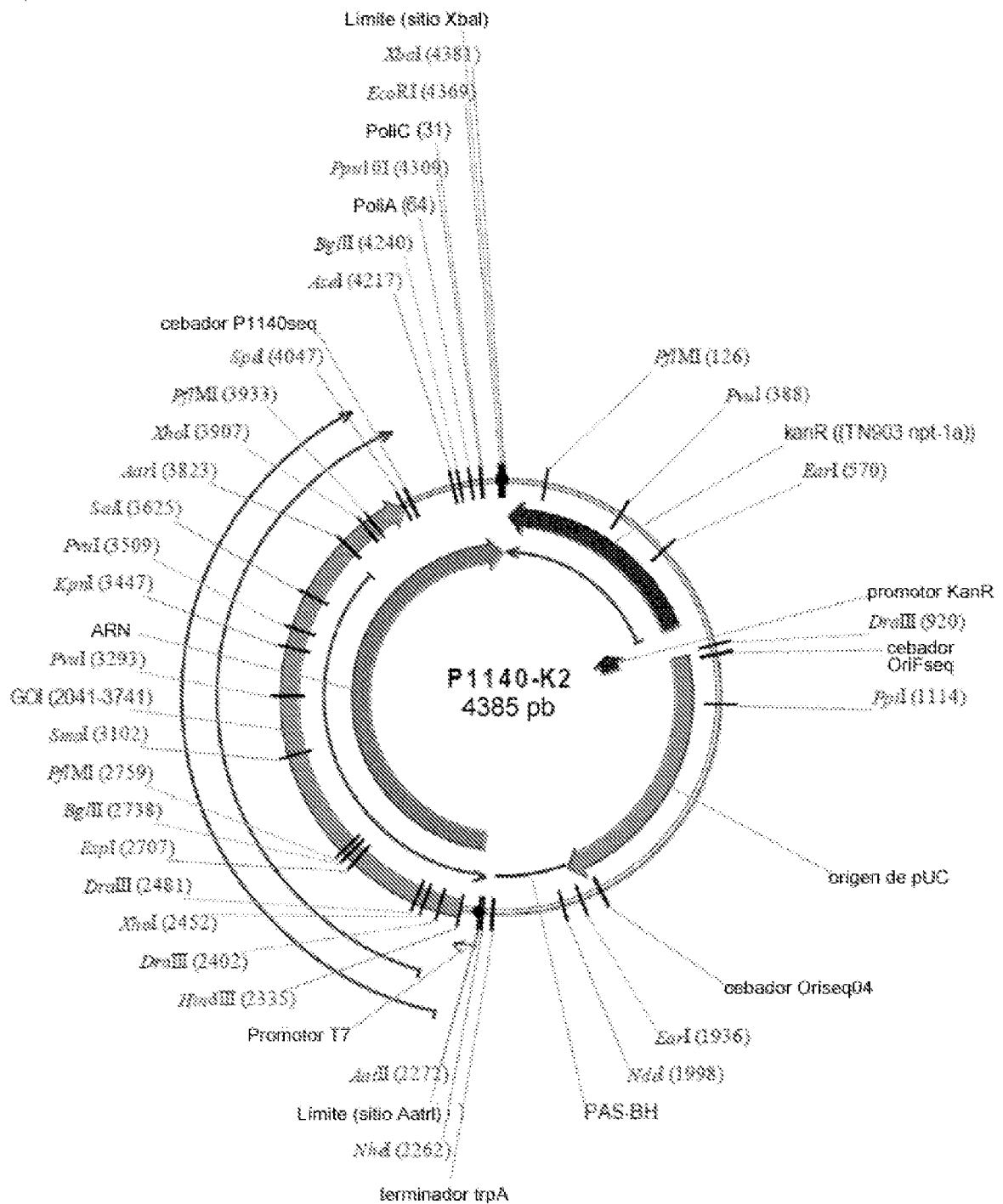
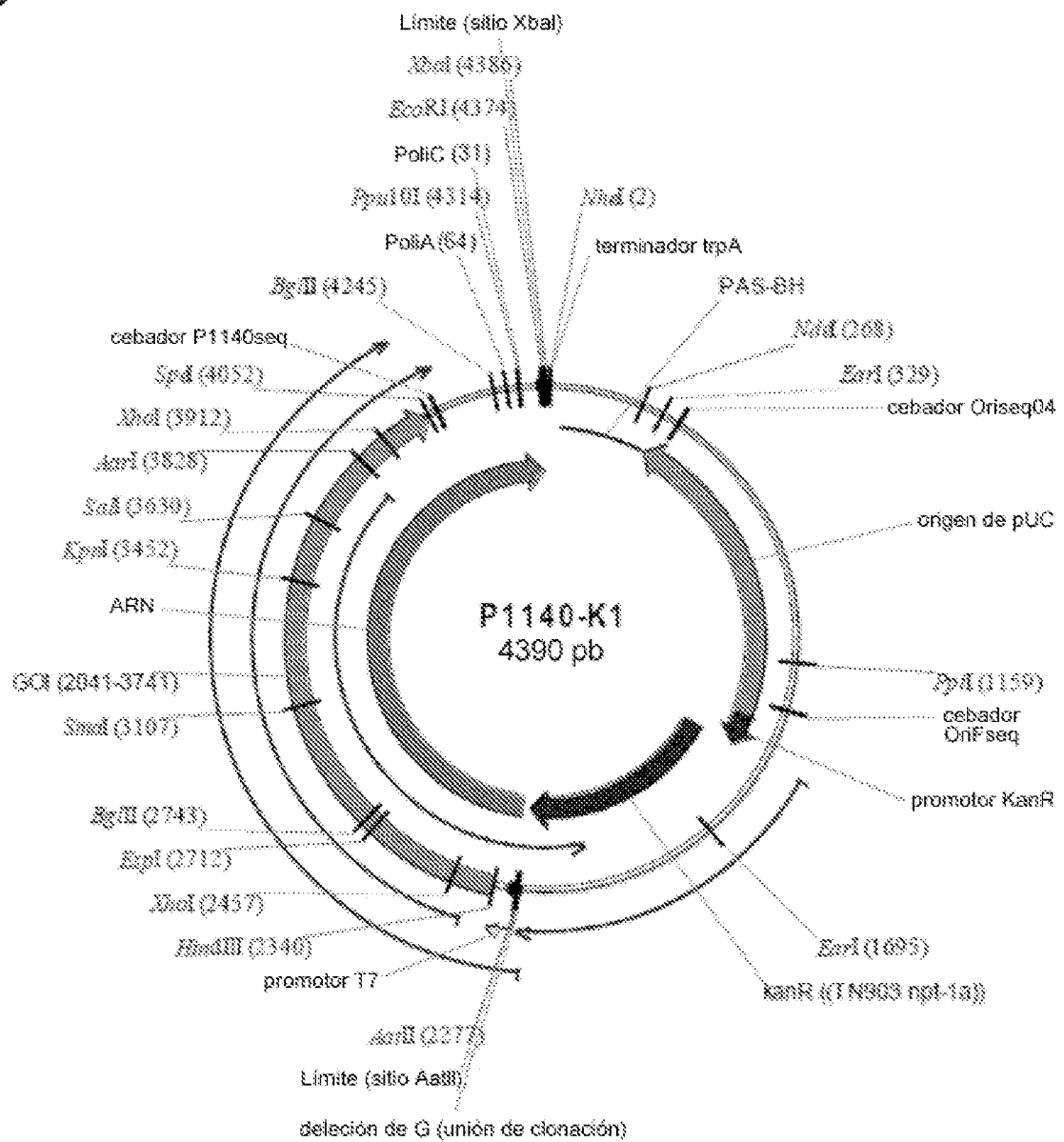


Figura 5



## Figura 6

ARNm optimizado en GC que codifica para H1N1 (Países Bajos 2009)-HA

```

GGGGCGCUGCCUACGGAGGUGGCAGCCAUCCUUCUCGGCAUCAAGCUUACCAUGAAGGCC
AUCCUGGUGGUGCCUCCUGUACACCUUCGCCACCGCGAACCGCCGACACGCUGUGCAUCGGCUA
CCACGCCAACAAACAGCACCGACACCGUGGACACCGUGCUCGAGAAGAACGUCACGGUGACCC
ACUCCGUGAACCUGCUGGAGGACAAGCACAAACGGGAAGCUCUGCAAGCUGCGGGGCGUCGCC
CCGCUGCACCUCGGGAAGUGCAACAUCGCCGGCUGGAUCCUGGGGAACCCGGAGUGCGAGAG
CCUGUCCACCGCCGAGCUCUCCUGGAGCUACAUCGUGGAGACCUCAGCUCGACAAACGGGCACGU
GCUACCCCGGGCGACUUCACUGGACUACGAGGAGCUCGCGGAGCAGCUGAGCUCGUGAGCUCU
UUCGAGCGGUUCGAGAUCUUCUCCCCAAGACCAGCUCUCCUGGCCCAACCACGACAGCAACAAGGG
GGUCACCGCCGCCUGCCCGCACGCCGGCGCGAAGUCCUUCUACAAGAACCUGAUCUGGCUCG
UGAAGAAGGGGAACAGCUACCCCAAGCUGUCCAAGAGCUACAUCAACGACAAGGGCAAGGAG
GUGCUGGUCCUCUGGGGGAUCCACCACCCAGCACCUCGCGGACCAGCAGAGCCUGUACCA
GAACCGCCGACGCCUACUGUUCGUGGGCUCAGCCGCUACUCCAAGAAGUUCAAGCCCGAGA
UCGCCAUCCGGCCGAAGGUCGCGGACAGGAGGGCCGGAUGAACUACUACUGGACGCUGGUG
GAGCCCGGGGACAAGAUCACCUUCGAGGCGACCGGCAACCUCGUGGUCGCGGCUACGCCUU
CGCCAUUGGAGCGGAACGCCGGGAGCGGCAUCAUCAUCUCCGACACCCCGUGGCACGACUGCA
ACACGACCUGCCAGACCCCGAAGGGCGCCAUCAACACCAGCCUGCCCUUCCAGAACAUCCAC
CCCAUCACGAUCGGGAAGUGCCCCAAGUACGUGAAGUCCACCAAGCUGCGCCUUCGCGACCGG
CCUGCGGAACGUCGCGGAGCAUCCAGUCCCGCGGGCUGUUCGGCGCCAUCGCGGGGUUCAUCG
AGGGCGGCGUGGACCGGGGAUGGUGGACGGCUGGUAACGGGUACCAACCAGAACGAGCAGGGC
AGCGGGUACGCGCGCGACCUCAAGUCCACGCGAAGCGGAUCGACGAGAUACCAACAAGGU
GAACAGCGUCAUCGAGAAAGAUGAACACCCAGUUCACCGCGUGGGCAAGGAGUUCAACCACC
UGGAGAAGCGGAUCGAGAACCUGAACAAGAAGGUCGACGACGGCUUCCUCGACAUCUGGACG
UACAACGCCGAGCUGCUGGUGCUCUCCUGGAGAACGAGCGCACCCUGGACUACCACGACUCCAA
CGUGAAGAACCUCUACGAGAAGGUCGGAGCCAGCUGAAGAACAACGCCAAGGAGAUCCGGGA
ACGGCUGCUUCGAGUUCUACCACAAGUGCGACAACACCUGCAUGGAGUCCUGAAGAACGGG
ACCUACGACUACCCCAAGUACAGCGAGGAGGCCAAGCUGAACCGCGAGGAGAUCCGACGGCGU
GAAGCUCGAGUCCACGCGGAUCUACCAGAUCCUGGCGAUCUACAGCACCUGCGCCAGCUCUCC
UGGUGCUCGUGGUCAGCCUGGGGGCCAUCUCCUUCUGGAUGUGCAGCAACGGCUCUCCUGCAG
UGCCGCAUCUGCAUCUGACCACUAGUGCAUCACAUUUAAAAGCAUCUCAGCCUACCAUGAGA
AUAAGAGAAAGAAAAUGAAGAUCAAUAGCUUAUUCUUCUUUUUCUUUUUCGUUGGUGUAA
AGCCAACACCCUGUCUAAAAACAUAUUUUUUUUAUCAUUUUUGCCUCUUUUUCUCUGUGCU
UCAAUUAAUAAAAAAUGGAAAGAACCUAGAUCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAUGCAUCCCCCCCCCCCCCCCCCCC
CCCCCCCCCAAGGCUCUUUUCAGAGCCACCAGAAU

```