(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 791 152**

(51) Int. CI.:

C12N 15/67 (2006.01)
C12N 15/85 (2006.01)
A61K 39/12 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **11.12.2015 PCT/EP2015/002501**

(87) Fecha y número de publicación internacional: **16.06.2016 WO16091391**

(96) Fecha de presentación y número de la solicitud europea: **11.12.2015 E 15817068 (8)**

(97) Fecha y número de publicación de la concesión europea: **19.02.2020 EP 3230458**

(54) Título: **Moléculas de ácido nucleico artificiales para una expresión proteica mejorada**

(30) Prioridad:

12.12.2014 WO PCT/EP2014/003334

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.11.2020

(73) Titular/es:

CUREVAC AG (100.0%)
Paul-Ehrlich-Str. 15
72076 Tübingen, DE

(72) Inventor/es:

THESS, ANDREAS;
SCHLAKE, THOMAS y
GRUND, STEFANIE

(74) Agente/Representante:

BUENO FERRÁN , Ana María

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

Moléculas de ácido nucleico artificiales para una expresión proteica mejorada

Campo de la invención

La invención se define en las reivindicaciones adjuntas y se refiere a una molécula de ácido nucleico artificial que comprende un marco de lectura abierto y una 3'-UTR que comprende al menos dos secuencias poli(A) separadas como se define en las reivindicaciones para su uso como un medicamento, para su uso como una vacuna o para su uso en la terapia génica, estando la molécula de ácido nucleico artificial asociada con o complejada con un compuesto catiónico o policatiónico. La invención se refiere además a un vector que comprende la molécula de ácido nucleico artificial, a una célula que comprende la molécula de ácido nucleico artificial o el vector, a una composición farmacéutica que comprende la molécula de ácido nucleico artificial o el vector y a un kit que comprende la molécula de ácido nucleico artificial, el vector y/o la composición farmacéutica. La invención también se refiere a un método *in vitro* para aumentar la expresión de proteínas a partir de una molécula de ácido nucleico artificial y al uso de una 3'-UTR para un método para aumentar la producción de proteínas a partir de una secuencia de ácido nucleico artificial.

Antecedentes de la invención

La terapia génica y la vacunación genética son parte de los métodos más prometedores y de rápido desarrollo de la medicina moderna. Pueden proporcionar opciones altamente específicas e individuales para la terapia de una gran variedad de enfermedades. Particularmente, enfermedades genéticas heredadas, pero también enfermedades autoinmunes, enfermedades cancerosas o relacionadas con tumores, así como enfermedades inflamatorias pueden ser el objetivo de tales procedimientos de tratamiento. También, se prevé evitar el inicio (precoz) de tales enfermedades mediante estas soluciones.

El concepto fundamental de la terapia génica es la apropiada modulación de una expresión génica deteriorada asociada a condiciones patológicas de enfermedad específicas. La expresión génica patológicamente alterada puede dar como resultado una carencia o sobreproducción de productos génicos esenciales, por ejemplo factores de señalización como hormonas, factores de mantenimiento, enzimas metabólicas, proteínas estructurales o similares. La expresión génica alterada puede deberse no solo a la regulación o transcripción y/o traducción defectuosa, sino también a mutaciones dentro de un ORF que codifica una proteína particular. Las mutaciones patológicas pueden estar provocadas, por ejemplo, por una aberración cromosómica o por más mutaciones específicas, como mutaciones puntuales o de desplazamiento de marco, todas ellas resultando en una funcionalidad limitada y, potencialmente, pérdida total de función del producto génico. Sin embargo, también aparecer una regulación defectuosa de la transcripción o traducción cuando las mutaciones afectan a genes que codifican las proteínas implicadas en la maquinaria de transcripción o de traducción celular. Tales mutaciones pueden conducir a una sub- o sobre-regulación patológica de los genes que, como tales, son funcionales. Genes que codifican los productos génicos que ejercen tales funciones reguladoras, pueden ser, por ejemplo, factores de transcripción, receptores de señal, proteínas mensajeras y similares. Sin embargo, la pérdida de función de estos genes que codifican proteínas reguladoras bajo ciertas circunstancias puede revertirse por la introducción artificial de otros factores que actúan adicionalmente aguas abajo del producto génico deteriorado. Tales defectos génicos también se pueden compensar con la terapia génica mediante la sustitución del gen afectado en sí mismo.

La vacunación genética permite provocar una respuesta inmunitaria deseada a determinados antígenos, tales como componentes característicos de las superficies bacterianas, partículas virales, antígenos tumorales o similares. En general, la vacunación es uno de los logros fundamentales de la medicina moderna. Sin embargo, las vacunas efectivas sólo están actualmente disponibles para una pequeña variedad de enfermedades. Por consiguiente, las infecciones que no se pueden prevenir por la vacunación aún afectan a millones de personas cada año.

Normalmente, las vacunas se pueden subdividir en vacunas de "primera", "segunda" y "tercera" generación. Las vacunas de "Primera generación" son típicamente vacunas de organismos enteros. Se basan tanto en patógenos vivos y atenuados como exterminados, por ejemplo virus, bacterias y similares. La desventaja principal de las vacunas vivas y atenuadas es el riesgo de reversión a las variantes que ponen en peligro la vida. Así, aunque están atenuados, tales patógenos aún pueden conllevar intrínsecamente riesgos no predecibles. Los patógenos exterminados pueden no ser tan eficaces como se desea para generar una respuesta inmunitaria específica. A fin de minimizar estos riesgos, se desarrollaron vacunas de "segunda generación". Éstas son típicamente vacunas de subunidad, que consisten en antígenos definidos o componentes de proteína recombinante derivados de patógenos.

Las vacunas genéticas, es decir, vacunas para la vacunación genética, generalmente se consideran vacunas de “tercera generación”. Se componen típicamente de moléculas de ácido nucleico genéticamente diseñadas que permiten la expresión de fragmentos de péptido o proteína (antígeno) característicos de un patógeno o de un antígeno tumoral in vivo. Las vacunas genéticas se expresan con la administración a un paciente y captación por las células competentes. La expresión de los ácidos nucleicos administrados da como resultado la producción de proteínas codificadas. En caso de que estas proteínas sean reconocidas como extrañas por el sistema inmunitario del paciente, se desencadena una respuesta inmunitaria.

Como se puede observar de lo anterior, ambos métodos, la terapia génica y la vacunación genética, se basan esencialmente en la administración de moléculas de ácido nucleico a un paciente y en la transcripción y/o traducción subsecuente de la información genética codificada. Alternativamente, la vacunación genética o la terapia génica también pueden comprender métodos que incluyen el aislamiento de células corporales específicas de un paciente a tratar, la transfección *in vitro* subsecuente de tales células y la re-administración de las células tratadas al paciente.

Como moléculas de ácido nucleico para la administración en el contexto de la terapia génica o la vacunación genética se puede emplear ADN y ARN. Es conocido que el ADN es relativamente estable y fácil de manejar. Sin embargo, el uso de ADN conlleva el riesgo de inserción indeseada de fragmentos de ADN administrados en el genoma del paciente, dando como resultado una potencial pérdida de función de los genes deteriorados. Como riesgo adicional, surge una generación indeseada de anticuerpos anti-ADN. Otra desventaja es el nivel de expresión limitado del péptido o proteína codificada que se logra con la administración de ADN y su transcripción/traducción. Entre otras razones, el nivel de expresión del ADN administrado será dependiente de la presencia de factores de transcripción específicos que regulan la transcripción de ADN. En ausencia de tales factores, la transcripción de ADN no producirá cantidades satisfactorias de ARN. Como resultado, se limita el nivel de péptido o proteína traducida obtenida.

Al utilizar ARN en lugar de ADN para la terapia o la vacunación genética, el riesgo de integración y generación genómica indeseada de anticuerpos anti-ADN se minimiza o evita. Sin embargo, el ARN se considera una especie molecular más bien inestable que se puede degradar fácilmente por ARNasas ubicuas.

In vivo, la degradación de ARN contribuye a la regulación del tiempo de vida media del ARN. Ese efecto se consideró y probó para afinar la regulación de la expresión génica eucariótica (Friedel y col., Conserved principles of mammalian transcriptional regulation revealed by RNA half-life, Nucleic Acid Research, 2009, 1-12). Por consiguiente, cada ARNm de origen natural tiene una vida media individual dependiendo del gen del cual se derive el ARNm. Contribuye a la regulación del nivel de expresión de este gen. Los ARN inestables son importantes para llevar a cabo la expresión génica transitoria en distintos puntos en el tiempo. Sin embargo, los ARN de larga vida se pueden asociar con la acumulación de distintas proteínas o expresión continua de genes. In vivo, la vida media de los ARNm también es dependiente de factores ambientales, como tratamiento hormonal, como se ha observado, por ejemplo, para el factor de crecimiento similar a insulina I, actina y ARNm de albúmina (Johnson y col., Newly synthesized RNA: Simultaneous measurement in intact cells of transcription rates and RNA stability of insulin-like growth factor I, actin, and albumin in growth hormone-stimulated hepatocytes, Proc. Natl. Acad. Sci., Vol. 88, páginas 5287-5291, 1991).

Para la terapia génica y la vacunación genética, normalmente se desea un ARN estable. Esto es, por una parte, debido al hecho de que el producto codificado por la secuencia de ARN se acumulará *in vivo*. Por otra parte, el ARN tiene que mantener su integridad estructural y funcional cuando se prepara para una forma de dosificación adecuada, durante su almacenamiento y cuando se administra. Así, se ha dedicado una atención considerable a proporcionar moléculas de ARN estables para la terapia génica o la vacunación genética a fin de evitar que se vean afectadas por la degradación o la descomposición temprana.

Se ha informado que el contenido de G/C de las moléculas de ácido nucleico puede influir en su estabilidad. Así, los ácidos nucleicos que comprenden una mayor cantidad de residuos de guanina (G) y/o citosina (C) pueden ser funcionalmente más estables que los ácidos nucleicos que contienen una gran cantidad de adenina (A) y timina (T) o nucleótidos uracilo (U). En este contexto, la WO02/098443 proporciona una composición farmacéutica que contiene un ARNm que se estabiliza por modificaciones de secuencia en la región traducida. Tal modificación de secuencia tiene ventajas en la degeneración del código genético. Por consiguiente, los codones que contienen una combinación menos favorable de nucleótidos (menos favorable en términos de estabilidad de ARN) se puede sustituir por codones alternativos sin alterar la secuencia de aminoácidos codificada. Este método de estabilización de ARN está limitado por las provisiones de la secuencia de nucleótidos específica de cada molécula de ARN individual, no permitiéndose dejar el espacio deseado de la secuencia de aminoácidos. Igualmente, este procedimiento está restringido a las regiones de codificación del ARN.

Como una opción alternativa para la estabilización del ARNm, se ha descubierto que las moléculas de ARNm eucarióticas de origen natural contienen elementos de estabilización característicos. Por ejemplo, pueden comprender las llamadas regiones no traducidas (UTR) en su extremo 5' (5'UTR) y/o en su extremo 3' (3'UTR), así como otras características estructurales, como una estructura cap 5' o una cola 3'-poli(A). Ambas, 5'UTR y 3'UTR se transcriben típicamente del ADN genómico y, así, son un elemento del ARNm prematuro. Los aspectos estructurales característicos del ARNm maduro, tal como la cap 5' y la cola 3'-poli(A) (también llamada cola poli(A) o secuencia poli(A)) normalmente se agregan al ARNm transcrito (prematuro) durante el procesamiento del ARNm.

Una cola 3'-poli(A) es típicamente un tramo de secuencia monótono de nucleótidos de adenina que está enzimáticamente unido al extremo 3' del ARNm naciente. Típicamente, la cola poli(A) de un mamífero contiene hasta aproximadamente 250 nucleótidos de adenina. Se descubrió que la longitud de esta cola 3'-poli(A) es un elemento potencialmente crítico para la estabilidad del ARNm individual. En este contexto, Holtkamp et al. informaron que una cola poli(A) consistente en 120 nucleótidos resultaba en una molécula de ARNm más estable, que se expresaba de manera más eficiente, que una cola poli(A) más corta (Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells, Blood, Vol. 108, pp. 4009-4017, 2006). Sin embargo, según Holtkamp et al., una extensión adicional de la cola poli(A) no conduce a un aumento adicional de la estabilidad o de la expresión de ARNm. Se informó además que la adenilación enzimática de un ARNm que comprende una cola poli(A) mejora aún más la expresión del ARNm después de electroporación en células T ((Zhao et al., Multiple Injections of Electroporated Autologous T Cells Expressing a Chimeric Antigen Receptor Mediate Regression of Human Disseminated Tumor, Caner Res., Vol. 70(22), pp. 9053-9061, 2010). Las solicitudes de patente internacionales WO 2016/005324 y WO 2016/005004, respectivamente, que se han publicado posteriormente y representan el estado de la técnica anterior según el Art. 54(3) EPC, describen moléculas de ácido nucleico que contienen regiones poli(dA:dT) que contienen una interrupción por una secuencia que no codifica una secuencia compuesta únicamente por residuos A. Secuencias de ácido nucleico similares también se describen en el documento WO 2015/101415, que también se publicó posteriormente.

Casi todos los ARNm eucariotas terminan en una secuencia poli(A) que se añade en su extremo 3' mediante la maquinaria ubicua de escisión/poliadenilación. La presencia de una secuencia poli (A) en el extremo 3' es una de las características más reconocibles de los ARNm eucariotas. Después de la escisión, la mayoría de los pre-ARNm, con la excepción de los transcriptos de histonas dependientes de la replicación, adquieren una cola poliadenilada. En este contexto, el procesamiento del extremo 3' es un proceso de co-transcripción nuclear que promueve el transporte de ARNm desde el núcleo al citoplasma y afecta a la estabilidad y traducción de los ARNm. La formación de este extremo 3' se produce en una reacción de dos pasos dirigida por la maquinaria de escisión/poliadenilación y depende de la presencia de dos elementos de secuencia en los precursores del ARNm (pre-ARNm): un hexanucleótido altamente conservado AAUAAA (señal de poliadenilación) y una secuencia rica en G/U aguas abajo. En un primer paso, los pre-ARNm se escinden entre estos dos elementos. En un segundo paso estrechamente unido al primer paso, el extremo 3' recién formado se extiende por adición de una secuencia poli(A) consistente en 200-250 adenilatos, que afecta posteriormente a todos los aspectos del metabolismo del ARNm, incluida la exportación, estabilidad y traducción de ARNm (Dominski, Z. y WF Marzluff (2007), Gene 396 (2): 373-90.).

La única excepción conocida a esta regla son los ARNm de histona dependientes de la replicación, que terminan en un tallo-bucle de histona en lugar de una secuencia poli(A). Ejemplos de secuencias tallo-bucle de histona se describen en Lopez et al. (Dávila López, M. y Samuelsson, T. (2008), ARN (Nueva York, N.Y.), 14 (1), 1-10. Doi: 10.1261 / rna.782308.).

Los tallos-bucle de histona en los pre-ARNm suelen ir seguidos de una secuencia rica en purinas conocida como elemento aguas abajo de histona (HDE). Estos pre-ARNm se procesan en el núcleo mediante una única división endonucleolítica aproximadamente 5 nucleótidos aguas abajo del tallo-bucle, por catálisis por el snRNP U7 mediante el emparejamiento de bases del ARNsn U7 y el HDE.

Debido al requerimiento de empaquetar el ADN recién sintetizado en la cromatina, la síntesis de histonas se regula en concierto con el ciclo celular. El aumento de la síntesis de proteínas histona durante la fase S se logra mediante la activación transcripcional de los genes de histona, así como por la regulación postranscripcional de los niveles de ARNm de histona. Se podría demostrar que el tallo-bucle de histona es esencial para todos los pasos postranscripcionales de la regulación de la expresión de histona. Para un procesamiento eficiente, es necesaria la exportación del ARNm al citoplasma, la carga en los polirribosomas y la regulación de la estabilidad del ARNm.

En el contexto anterior, se identificó una proteína de 32 kDa asociada al tallo-bucle de histona en el extremo 3' de los mensajes de histona tanto en el núcleo como en el citoplasma. El nivel de expresión de esta proteína de

unión a tallo-bucle (SLBP) es un ciclo celular regulado y es más alto durante la fase S, cuando los niveles de ARNm de histona se incrementan. La SLBP es necesaria para el procesamiento eficiente del extremo 3' del pre-ARNm de histona por U7 snRNP. Después de la terminación del procesamiento, la SLBP permanece asociada al tallo-bucle en el extremo de los ARNm de histona maduros y estimula su traducción en las proteínas de histona en el citoplasma. (Dominski, Z. and W. F. Marzluff (2007), Gene 396(2): 373-90). De forma interesante, el dominio de unión a ARN de SLBP se conserva en metazoarios y protozoarios (Dávila López, M., & Samuelsson, T. (2008), RNA (Nueva York, N.Y.), 14(1), 1-10. doi:10.1261/rna.782308) y se podría demostrar que su unión a la secuencia tallo-bucle de histona es dependiente de la estructura tallo-bucle y que el sitio de unión mínima contiene al menos 3 nucleótidos 5' y 2 nucleótidos 3' del tallo-bucle (Pandey, N. B., y col. (1994), Molecular and Cellular Biology, 14(3), 1709-1720 y Williams, A. S., & Marzluff, W. F., (1995), Nucleic Acids Research, *23*(4), 654-662).

Incluso aunque los genes de histona se clasifican generalmente como "dependientes de replicación", dando origen al ARNm que termina en un tallo-bucle de histona, o como "tipo de reemplazo", dando origen a un ARNm que lleva una cola poli(A), en cambio, los ARNm de origen natural que contienen tanto un tallo-bucle de histona como una poli(A) u oligo(A) 3' del mismo se identifican solo en algunos casos muy raros. Sanchez y col. examinaron el efecto de las colas oligo(A) de origen natural adjuntas a 3' del tallo-bucle de histona del ARNm de histona durante la ovogénesis de Xenopus utilizando luciferasa como proteína reporter y descubriendo que la cola oligo(A) es una parte activa del mecanismo de represión de la traducción que silencia el ARNm de histona durante la ovogénesis y su eliminación es parte del mecanismo que activa la traducción de los ARNm de histona (Sanchez, R. y W. F. Marzluff (2004), Mol Cell Biol 24(6): 2513-25).

Además, se han investigado los requisitos para la regulación de las histonas dependientes de replicación en el nivel del procesamiento de pre-ARNm y la estabilidad del ARNm utilizando constructos artificiales que codifican la proteína marcadora alfa-globina, en base al hecho de que el gen de globina contiene intrones en oposición a los genes de histona sin intrones. Para este propósito se generaron constructos donde una secuencia de codificación de alfa-globina se siguió de una señal tallo-bucle de histona (tallo-bucle de histona seguido por el elemento aguas abajo de histona) y una señal de poliadenilación (Whitelaw, E., y col. (1986). Nucleic Acids Research, 14(17), 7059-7070; Pandey, N. B., & Marzluff, W. F. (1987). Molecular and Cellular Biology, 7(12), 4557-4559; Pandey, N. B., y col. (1990). Nucleic Acids Research, 18(11), 3161-3170).

También, se demostró que la 3'UTR del ARNm de α-globina puede ser un factor importante para la estabilidad bien conocida de un ARNm de α-globina (Rodgers y col., Regulated α-globin RNAm decay is a cytoplasmic event proceeding through 3'-to-5' exosome-dependent decapping, RNA, 8, pp. 1526-1537, 2002). La 3'UTR de un ARNm de α-globina está obviamente implicado en la formación de un complejo de ribonucleoproteína específico, un α-complejo, cuya presencia se correlaciona con la estabilidad del ARNm *in vitro* (Wang y col., An ARNm stability complex functions with poli(A)-binding protein to stabilize ARNm *in vitro*, Molecular and Cellular biology, Vol 19, No. 7, Julio de 1999, p. 4552-4560).

Independientemente de los factores que afectan a la estabilidad del ARNm, la traducción eficaz de las moléculas de ácido nucleico administradas por las células o tejido objetivo es crucial para cualquier procedimiento que utiliza moléculas de ácido nucleico para la terapia génica o la vacunación genética. Junto con la regulación de la estabilidad, también la traducción de la mayoría de los ARNm se regula por las características estructurales similares a UTR, cap 5' y cola 3'-poli(A). En este contexto, se ha informado que la longitud de la cola poli(A) puede desempeñar una función importante también en la eficiencia de la traducción. La estabilización de elementos 3', sin embargo, también puede tener un efecto de atenuación de la traducción.

Otros elementos reguladores que pueden tener influencia en los niveles de expresión se pueden encontrar en la 5'UTR. Por ejemplo, se ha informado que la síntesis de proteínas particulares, por ejemplo proteínas que pertenecen al aparato traduccional, se pueden regular no solo a nivel transcripcional sino también a nivel traducción. Por ejemplo, la traducción de las proteínas codificadas por los llamados "genes TOP" se pueden subregular por la represión traduccional. Aquí, el término "gen TOP" se refiere a un gen que corresponde a un ARNm caracterizado por la presencia de una secuencia TOP en el extremo 5' y, en la mayoría de casos, por una regulación de traducción asociada al crecimiento (Iadevaia y col., All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5'-terminal oligopyrimidine (TOP) ARNms; RNA, 2008, 14:1730-1736). En este contexto, una secuencia TOP – también llamada "tracto de oligopirimidina 5'-terminal" consiste típicamente en residuos C en el sitio cap, seguido por una secuencia no interrumpida de hasta 13 o aún más pirimidinas (Avni y col., Vertebrate mRNAs with a 5'-terminal pyrimidine tract are candidates for translational repression in quiescent cells: characterization of the translational cis-regulatory element, Molecular and Cellular Biology, 1994, p. 3822-3833). Se indica que estas secuencias TOP están presentes en muchos ARNm que codifican componentes de la maquinaria traduccional y son responsables de la represión selectiva de la traducción de estos ARNm que contienen TOP debido a la detección del crecimiento (Meyuhas y col., Translational Control of Ribosómica Protein mRNAs in Eukaryotes, Translational Control. Cold Spring

Harbor Monograph Archive. Cold Spring Harbor Laboratory Press, 1996, p. 363–388).

Es el objeto de la invención proporcionar moléculas de ácido nucleico que pueden ser adecuadas para su uso como un medicamento o vacuna, preferentemente para la aplicación en la terapia génica y/o la vacunación genética. Particularmente, es objeto de la invención proporcionar moléculas de ácido nucleico artificiales, tales como una especie de ARNm, que proporcionan una producción de proteínas mejorada a partir de dichas moléculas de ácido nucleico artificiales. Otro objeto de la presente invención es proporcionar moléculas de ácido nucleico que codifican tal especie de ARNm superior que puede ser apta para su uso como medicamento o vacuna, preferentemente en la terapia génica y/o la vacunación genética. Es otro objeto adicional de la presente invención proporcionar una composición farmacéutica, preferentemente para su uso como un medicamento o vacuna, preferentemente en la terapia génica y/o la vacunación genética. En resumen, es objeto de la presente invención proporcionar especies de ácido nucleico mejoradas que superen las desventajas planteadas en el estado anterior de la técnica mediante un procedimiento rentable y sencillo.

El objetivo que subyace aquí se resuelve por el contenido de las reivindicaciones.

Por razones de claridad y legibilidad, se proporcionan las siguientes definiciones. Cualquier característica técnica mencionada en estas definiciones se puede leer en todas y cada una de las realizaciones de la invención. Las definiciones y explicaciones adicionales se pueden proporcionar específicamente en el contexto de estas realizaciones.

Respuesta inmunitaria adaptiva: La respuesta inmunitaria adaptiva se entiende típicamente como una respuesta específica de antígeno del sistema inmunitario. La especificidad de antígeno permite generar respuestas que se adaptan a patógenos específicos o células infectadas por patógenos. Normalmente, la capacidad de desarrollar estas respuestas adaptadas es mantenida en el cuerpo por “células de memoria”. Si un patógeno infecta el cuerpo más de una vez, se usan estas células de memoria específicas para eliminarlo rápidamente. En este contexto, la primera etapa de una respuesta inmunitaria adaptiva es la activación de células T específicas de antígeno naive o diferentes células inmunitarias capaces de inducir una respuesta inmunitaria específica de antígeno por células presentadoras de antígeno. Esto ocurre en los tejidos linfoides y órganos a través de los cuales las células T naive pasan constantemente. Los tres tipos de células que pueden servir como células presentadoras de antígeno son células dendríticas, macrófagos y células B. Cada una de estas células tiene una función distinta provocando respuestas inmunitarias. Las células dendríticas pueden tomar antígenos por fagocitosis y macropinocitosis y pueden volverse a estimular por contacto con, por ejemplo, un antígeno exterior para migrar al tejido linfoide local, donde se diferencian en células dendríticas maduras. Los macrófagos ingieren antígenos de partículas como bacterias y se inducen por agentes infecciosos u otros estímulos apropiados para expresar las moléculas MHC. La capacidad única de las células B de enlazar e interiorizar antígenos de proteína solubles por medio de sus receptores también puede ser importante para inducir las células T. Las moléculas MHC son típicamente responsables de la presentación de un antígeno a células T. A este respecto, presentar el antígeno en moléculas MHC lleva a la activación de las células T que inducen su proliferación y diferenciación en células T efectoras armadas. La función más importante de las células T efectoras es la eliminación de células infectadas por células T citotóxicas CD8+ y la activación de macrófagos por células Th1 que juntos conforman la inmunidad mediada por célula, y la activación de células B por ambas células Th2 y Th1 para producir diferentes clases de anticuerpo, conduciendo así a la respuesta inmunitaria humoral. Las células T reconocen un antígeno por sus receptores de células T que no reconocen y enlazan el antígeno directamente, pero en su lugar reconocen fragmentos peptídicos cortos, por ejemplo de antígenos de proteína derivados de patógeno, por ejemplo los llamados epítopos, que se enlazan a moléculas MHC en las superficies de otras células.

Sistema inmunitario adaptativo: El sistema inmunitario adaptativo se dedica esencialmente a eliminar o impedir el crecimiento patogénico. Típicamente regula la respuesta inmunitaria adaptiva dotando al sistema inmunitario vertebrado de la capacidad para reconocer y recordar patógenos específicos (para generar inmunidad) y para desarrollar ataques más fuertes cada vez que se encuentra el patógeno. El sistema es altamente adaptable debido a la hipermutación somática (un proceso de mutaciones somáticas aceleradas) y a la recombinación V(D)J (una recombinación genética irreversible de segmentos de gen de receptor de antígeno). Este mecanismo permite a un número pequeño de genes generar un gran número de diferentes receptores de antígeno, que luego se expresan únicamente en cada linfocito individual. Debido a que el reordenamiento del gen lleva a un cambio irreversible del ADN de cada célula, toda la progenie (descendencia) de tal célula heredará entonces genes que codifican la misma especificidad de receptor, que incluye células B de Memoria y células T de Memoria, claves para la inmunidad específica longeva.

Adyuvante/componente adyuvante: Un adyuvante o un componente adyuvante en el sentido más amplio es típicamente un agente farmacológico y/o inmunológico que puede modificar, por ejemplo mejorar, el efecto de otros agentes, tal como un fármaco o vacuna. Esto debe interpretarse en un sentido amplio y se refiere a un

espectro amplio de sustancias. Típicamente, estas sustancias son capaces de incrementar la inmunogenicidad de antígenos. Por ejemplo, los adyuvantes se pueden ser reconocidos por los sistemas inmunitarios innatos y, por ejemplo, pueden provocar una respuesta inmunitaria innata. Los "adyuvantes" típicamente no provocan una respuesta inmunitaria adaptiva. A este respecto, los "adyuvantes" no califican como antígenos. Su modo de acción es distinto de los efectos desencadenados por antígenos que resultan en una respuesta inmunitaria adaptiva.

Antígeno: En el contexto de la presente descripción, "antígeno" se refiere típicamente a una sustancia que puede ser reconocida por el sistema inmunitario, preferiblemente por el sistema inmunitario adaptativo, y es capaz de desencadenar una respuesta inmunitaria específica de antígeno, por ejemplo por la formación de anticuerpos y/o células T específicas de antígeno como parte de una respuesta inmunitaria adaptiva. Típicamente, un antígeno puede ser o puede comprender un péptido o proteína que comprende al menos un epítopo y que se puede presentar por la MHC a células T. En el sentido de la presente descripción, un antígeno puede ser el producto de la traducción de una molécula de ácido nucleico proporcionada, preferiblemente un ARNm, como se define aquí. En este contexto, también los fragmentos, las variantes y los derivados de péptidos y proteínas que comprenden al menos un epítopo se entienden como antígenos. En el contexto de la presente descripción, son particularmente preferentes los antígenos tumorales y los antígenos patogénicos como se definen aquí.

Molécula artificial de ácido nucleico: Una molécula artificial de ácido nucleico se puede entender típicamente como una molécula de ácido nucleico, por ejemplo un ADN o un ARN, que no es de origen natural. En otras palabras, una molécula artificial de ácido nucleico se puede entender como una molécula de ácido nucleico no natural. Tal molécula de ácido nucleico puede ser no natural debido a su secuencia individual (que no se presenta naturalmente) y/o debido a otras modificaciones, por ejemplo modificaciones estructurales de nucleótidos que no se presentan naturalmente. Una molécula artificial de ácido nucleico puede ser una molécula de ADN, una molécula de ARN o una molécula híbrida que comprende porciones de ADN y ARN. Típicamente, las moléculas artificiales de ácido nucleico se pueden diseñar y/o generar por métodos de ingeniería genética en correspondencia con una secuencia artificial deseada de nucleótidos (secuencia heteróloga). En este contexto una secuencia artificial normalmente es una secuencia que no puede estar presente naturalmente, es decir difiere de la secuencia de tipo natural en al menos un nucleótido. El término "tipo natural" se puede entender como una secuencia presente en la naturaleza. Además, el término "moléculas de ácido nucleico artificiales" no se restringe a "una molécula individual" sino, típicamente, se entiende que comprende un conjunto de moléculas idénticas. Por consiguiente, se puede referir a una pluralidad de moléculas idénticas contenidas en una alícuota.

ARN bicistrónico, ARN multicistrónico: Un ARN bicistrónico o multicistrónico es típicamente un ARN, preferentetemente un ARNm, que puede tener típicamente dos (bicistrónico) o más (multicistrónico) marcos de lectura abiertos (ORF). Un marco de lectura abierto en este contexto es una secuencia de codones que es traducible en un péptido o proteína.

Portador/portador polimérico: Un portador en el contexto de la presente descripción puede ser típicamente un compuesto que facilita el transporte y/o la formación de un complejo con otro compuesto (carga). Un portador polimérico es típicamente un portador que se forma en un polímero. Un portador se puede asociar a su carga por interacción covalente o no covalente. Un portador puede transportar ácidos nucleicos, por ejemplo ARN o ADN, a las células diana. El portador puede - para algunas realizaciones – ser un componente catiónico.

Componente catiónico: El término "componente catiónico" se refiere típicamente a una molécula cargada que está cargada positivamente (catión) a un pH típico de 1 a 9, de manera preferente un pH de o inferior a 9 (por ejemplo de 5 a 9), de o por debajo de 8 (por ejemplo de 5 a 8), de o por debajo de 7 (por ejemplo de 5 a 7), de manera preferente a pH fisiológico, por ejemplo de 7,3 a 7,4. Por consiguiente, un componente catiónico puede ser cualquier compuesto o polímero cargado positivamente, de manera preferente un péptido o proteína catiónico cargado positivamente bajo condiciones fisiológicas, en particular bajo condiciones fisiológicas *in vivo*. Un "péptido o proteína catiónica" puede contener al menos un aminoácido cargado positivamente o más de un aminoácido cargado positivamente, por ejemplo seleccionado de Arg, His, Lys u Orn. Así, los componentes "policatiónicos", que tienen más de una carga positiva bajo las condiciones dadas, también están dentro del alcance.

Cap 5': Cap 5' es una entidad, típicamente una entidad de nucleótidos modificada, que generalmente "tapa" el extremo 5' de un ARNm maduro. Un cap 5' se puede formar típicamente por un nucleótido modificado, en particular por un derivado de un nucleótido de guanina. De manera preferente, cap 5' se une al extremo 5' mediante un enlace 5'-5'-trifosfato. Un cap 5' puede estar metilado, por ejemplo m7GpppN, donde N es el nucleótido 5' terminal del ácido nucleico que lleva el cap 5', típicamente el extremo 5' de un ARN. Ejemplos adicionales de estructuras cap 5' incluyen glicerilo, residuo no básico de desoxi invertido (porción), nucleótido

4',5'- metileno, nucleótido de 1-(beta-D-eritrofuranosilo), nucleótido 4'-tio, nucleótido carbocíclico, nucleótido de 1,5-anidrohexitol, L-nucleótidos, alfa-nucleótido, nucleótido base modificado, nucleótido treopentofuranosilo, nucleótido 3',4'-seco acíclico, nucleótido 3,4-dihidroxibutil-acíclico, nucleótido 3,5-dihidroxipentil-acíclico, porción de nucleótido 3'-3'-invertido, porción abásica 3'-3'-invertida, porción de nucleótido 3',2'-invertido, porción abásica 3',2'-invertida, fosfato de 1,4-butanodiol, 3'-fosforamidato, hexilfoshato, fosfato de aminohexilo, 3'-fosfato, 3'-fosforotioato, fosforoditioato, o porción de metilfosfonato puente o no puente.

Inmunidad celular/respuesta inmunitaria celular: La inmunidad celular se relaciona típicamente con la activación de macrófagos, células asesinas naturales (NK), linfocitos T citotóxicos específicos de antígeno y la liberación de diversas citoquinas en respuesta a un antígeno. En términos más generales, la inmunidad celular no se basa en anticuerpos, sino en la activación de células del sistema inmunitario. Típicamente, una respuesta inmunitaria celular se puede caracterizar por ejemplo activando linfocitos T citotóxicos específicos de antígeno que son capaces de inducir la apoptosis celular, por ejemplo en células inmunitarias específicas como células dendríticas u otras células, que detectan epítopos de antígenos extraños en su superficie. Tales células se pueden infectar por virus o con bacterias intracelulares o con células de cáncer que detectan antígenos tumorales. Las características adicionales pueden ser la activación de macrófagos y células asesinas naturales, permitiéndoles destruir patógenos y la estimulación de células para secretar una variedad de citoquinas que influyen en la función de otras células implicadas en respuestas inmunitarias adaptativas e innatas.

ADN: ADN es la abreviatura usual para ácido desoxirribonucleico. Es una molécula de ácido nucleico, es decir un polímero consistente en nucleótidos. Estos nucleótidos son normalmente monómeros de desoxiadenosin-monofosfato, desoxitimidin-monofosfato, desoxiguanosin-monofosfato y desoxicitidin-monofosfato que son a su vez están compuestos por una porción de azúcar (desoxirribosa), una porción base y una porción fosfato, y se polimerizan en un esqueleto principal característico. El esqueleto de principal está formado típicamente por uniones fosfodiéster entre la porción de azúcar del nucleótido, es decir la desoxirribosa, de una primera porción fosfato de un segundo monómero adyacente. El orden específico de los monómeros, es decir el orden de las bases ligadas a la cadena principal azúcar/fosfato, se denomina secuencia de ADN. El ADN puede ser mono- o bi-catenario. En la forma de doble hebra, los nucleótidos de la primera hebra se hibridan típicamente con los nucleótidos de la segunda hebra, por ejemplo por un apareamiento de bases A/T y G/C.

Epítopo: Los epítopos (también llamados 'determinantes de antígeno') pueden distinguirse en epítopos de células T y epítopos de células B. Los epítopos de células T o partes de las proteínas en el contexto de la presente invención pueden comprender fragmentos que, preferiblemente, tienen una longitud de alrededor de 6 a alrededor de 20 o aún más aminoácidos, por ejemplo fragmentos tal como se procesan y presentan por moléculas de clase I MHC, preferiblemente con una longitud de alrededor de 8 a alrededor de 10 aminoácidos, por ejemplo 8, 9, o 10, (o aún 11, o 12 aminoácidos), o fragmentos tales como se procesan y presentan por moléculas de clase II MHC, preferiblemente de una longitud de alrededor de 13 o más aminoácidos, por ejemplo 13, 14, 15, 16, 17, 18, 19, 20 o aún más aminoácidos, donde estos fragmentos se pueden seleccionar de cualquier parte de la secuencia de aminoácidos. Estos fragmentos son típicamente reconocidos por células T en forma de un complejo que consiste en el fragmento de péptido y una molécula MHC, así, los fragmentos típicamente no son reconocidos en su forma nativa. Los epítopos de células B son típicamente fragmentos ubicados en la superficie externa de proteínas (nativas) o antígenos de péptido como se define aquí, preferiblemente de 5 a 15 aminoácidos, más preferiblemente de 5 a 12 aminoácidos, aún más preferiblemente de 6 a 9 aminoácidos, que pueden ser reconocidos por anticuerpos, esto es en su forma nativa. Tales epítopos de proteínas o péptidos además se pueden seleccionar de cualquiera de las variantes aquí mencionadas de tales proteínas o péptidos. En este contexto, los determinantes antigénicos pueden ser epítopos conformacionales y discontinuos que se componen de segmentos de las proteínas o péptidos tal como se definen aquí que son discontinuos en la secuencia de aminoácidos de las proteínas o péptidos como se definen aquí, pero se reúnen en la estructura tridimensional, o epítopos continuos o lineales que se componen de una única cadena de polipéptidos.

Fragmento de una secuencia: Un fragmento de una secuencia es típicamente una parte más corta de una secuencia de longitud completa de, por ejemplo, una secuencia de ácido nucleico o de aminoácidos. En consecuencia, un fragmento de una secuencia típicamente consiste en una secuencia que es idéntica al tramo o tramos correspondientes dentro de la secuencia de longitud completa. Un fragmento preferente de una secuencia en el contexto de la presente invención consiste en un tramo continuo de entidades, como nucleótidos o aminoácidos, que corresponden a un tramo continuo de entidades en la molécula de la que se deriva el fragmento, lo que representa al menos el 5%, 10%, 20%, preferiblemente al menos el 30%, más preferiblemente al menos el 40%, más preferiblemente al menos el 50%, incluso más preferiblemente al menos el 60%, incluso más preferiblemente al menos el 70% y con total preferencia al menos el 80% de la molécula total (esto es longitud completa) de la cual se deriva el fragmento.

Modificado en G/C: Un ácido nucleico modificado en G/C puede ser típicamente un ácido nucleico, preferentemente una molécula artificial de ácido nucleico como se define aquí, basada en una secuencia de tipo natural modificada, que comprende un número preferentemente incrementado de nucleótidos guanosina y/o citosina en comparación con la secuencia de tipo natural. Tal número incrementado se puede generar por sustitución de codones que contienen nucleótidos de adenosina o timidina por codones que contienen nucleótidos de guanosina o citosina. Si el contenido de G/C enriquecido aparece en una región de codificación de ADN o ARN, hace uso de la degeneración del código genético. Por consiguiente, las sustituciones del codón preferentemente no alteran los residuos de aminoácidos codificados, sino incrementan exclusivamente el contenido de G/C de la molécula de ácido nucleico.

Terapia génica: La terapia génica se puede entender como un tratamiento de elementos corporales o aislados del cuerpo de un paciente, por ejemplo tejidos/células aisladas, con ácidos nucleicos que codifican un péptido o proteína. Típicamente puede comprender al menos una de las ecaps de a) administrar un ácido nucleico, de manera preferente una molécula artificial de ácido nucleico como se define aquí, directamente al paciente – por cualquier vía de administración – o *in vitro* a células/tejidos aislados del paciente, lo cual tiene como resultado la transfección de las células del paciente ya sea *in vivo/ex vivo o in vitro*; b) la transcripción y/o traducción de la molécula de ácido nucleico introducida; y opcionalmente c) la re-administración de las células transfectadas aisladas al paciente si el ácido nucleico no se ha administrado directamente al mismo.

Vacunación genética: La vacunación genética típicamente se puede entender como la vacunación por administración de una molécula de ácido nucleico que codifica un antígeno o un inmunógeno o fragmentos de los mismos. La molécula de ácido nucleico se puede administrar a un cuerpo del sujeto o a células aisladas de un sujeto. Tras la transfección de ciertas células del cuerpo o tras la transfección de las células aisladas, el antígeno o inmunógeno puede ser expresado por aquellas células y posteriormente ser presentado al sistema inmunitario, provocando una respuesta inmunitaria adaptiva, esto es específica de antígeno. En consecuencia, la vacunación genética típicamente comprende al menos una de las etapas de a) administrar un ácido nucleico, preferiblemente un ARN (aislado) como se define aquí, a un sujeto, preferiblemente un paciente, o a células aisladas de un sujeto, preferiblemente de un paciente, que usualmente resulta en la transfección de las células del sujeto ya sea *in vivo* o *in vitro*; b) transcripción y/o traducción de la molécula de ácido nucleico introducida; y opcionalmente c) re-administración de células aisladas transfectadas al sujeto, preferiblemente al paciente, si el ácido nucleico no se ha administrado directamente al paciente.

Secuencia heteróloga: Típicamente se entiende que dos secuencias son “heterólogas” cuando no se derivan del mismo gen, es decir, aunque las secuencias heterólogas se pueden derivar del mismo organismo, no se presentan naturalmente (en la naturaleza) en la misma molécula de ácido nucleico, tal como en el mismo ARNm.

Inmunidad humoral/respuesta inmunitaria humoral: La inmunidad humoral se refiere típicamente a la producción de anticuerpos y opcionalmente a procesos accesorios que acompañan q la producción de anticuerpos. Una respuesta inmunitaria humoral se puede caracterizar típicamente, por ejemplo, por la activación de Th2, la producción de citoquinas, formación central germinal y cambio de isotipo, maduración por afinidad y generación de células de memoria. La inmunidad humoral también se puede referir típicamente a funciones efectoras de los anticuerpos, incluyendo la naturalización de patógenos y toxinas, activación de complemento clásica y la promoción de opsonina de fagocitosis y eliminación de patógenos.

Inmunógeno: En el contexto de la presente descripción, un inmunógeno se puede entender típicamente como un compuesto capaz de estimular una respuesta inmunitaria. Preferentemente, un inmunógeno es un péptido, polipéptido o proteína. En una realización particularmente preferente, un inmunógeno en el sentido de la presente descripción es el producto de la traducción de una molécula de ácido nucleico proporcionada, preferentemente una molécula artificial de ácido nucleico como se define aquí. Típicamente un inmunógeno induce al menos una respuesta inmunitaria adaptativa.

Composición inmunoestimuladora: En el contexto de la presente descripción, una composición inmunoestimuladora se puede entender típicamente como una composición que contiene al menos un componente capaz de inducir una respuesta inmunitaria o del cual se deriva un componente que es capaz de inducir una respuesta inmunitaria. Tal respuesta inmunitaria puede ser preferentemente una respuesta inmunitaria innata o una combinación de una respuesta inmunitaria adaptativa e innata. De manera preferente, una composición inmunoestimuladora en el contexto de la presente descripción contiene al menos una molécula artificial de ácido nucleico, en especial un ARN, por ejemplo una molécula de ARNm. El componente inmunoestimulador, tal como el ARNm, se puede complejar con un portador adecuado. Así, la composición inmunoestimuladora puede comprender un complejo ARNm/portador. Adicionalmente, la composición inmunoestimuladora puede comprender un adyuvante y/o un vehículo adecuado para el componente inmunoestimulador, tal como el ARNm.

Respuesta inmunitaria: Una respuesta inmunitaria puede ser típicamente una reacción específica del sistema inmunitario adaptativo a un antígeno particular (llamada así respuesta inmunitaria específica o adaptativa) o una reacción no específica del sistema inmunitario innato (llamada entonces respuesta inmunitaria no específica o innata), o una combinación de las mismas.

Sistema inmunitario: El sistema inmunitario puede proteger a los organismos de infecciones. Si un patógeno tiene éxito traspasando una barrera física de un organismo y entra en éste, el sistema inmunitario innato proporciona una respuesta inmediata, pero no específica. Si los patógenos evaden esta respuesta innata, los vertebrados tienen una segunda capa de protección, el sistema inmunitario adaptativo. Aquí, el sistema inmunitario adapta su respuesta durante una infección para mejorar su reconocimiento del patógeno. Esta respuesta mejorada se mantiene así después de que el patógeno se ha eliminado en forma de una memoria inmunológica y permite al sistema inmunitario adaptativo desarrollar ataques más rápidos y fuertes cada vez que se encuentra este patógeno. De acuerdo con esto, el sistema inmunitario comprende el sistema inmunitario innato y el adaptativo. Cada una de estas dos partes típicamente contiene los denominados componentes humorales y celulares.

ARN inmunoestimulador: Un ARN inmunoestimulador (isARN) en el contexto de la presente descripción típicamente puede ser un ARN que es capaz de inducir una respuesta inmunitaria innata. Éste usualmente no tiene un marco de lectura abierto y, por tanto, no proporciona un péptido-antígeno o inmunógeno, pero provoca una respuesta inmunitaria innata, por ejemplo por enlace a una clase específica de receptor tipo Toll (TLR) u otros receptores adecuados. Sin embargo, por supuesto también los ARNm que tienen un marco de lectura abierto y que codifican para un péptido/proteína puede inducir una respuesta inmunitaria innata y, por tanto, pueden ser ARN inmunoestimuladores.

Sistema inmunitario innato: El sistema inmunitario innato, también conocido como sistema inmunitario no específico (o inespecífico), típicamente comprende células y mecanismos que defienden al huésped de la infección por otros organismos de manera no específica. Esto significa que las células del sistema innato pueden reconocer y responder a patógenos de forma genérica, pero, a diferencia del sistema inmunitario adaptativo, no confiere inmunidad de larga duración o de protección al huésped. El sistema inmunitario innato puede ser activado, por ejemplo, por ligandos de receptores tipo Toll (TLR) u otras sustancias auxiliares, como lipopolisacáridos, TNF-alfa, ligando CD40, o citoquinas, monoquinas, linfoquinas, interleuquinas o quimiocinas, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IFN-alfa, IFN-beta, IFN-gama, GM-CSF, G-CSF, M-CSF, LT-beta, TNF-alfa, factores de crecimiento y hGH, un ligando de receptor tipo Toll humano TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, un ligando de receptor tipo Toll de murino TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, TLR11, TLR12 o TLR13, un ligando de un receptor tipo NOD, un ligando de un receptor tipo RIG-I, un ácido nucleico inmunoestimulador, un ARN inmunoestimulador (isARN), un CpG-ADN, un agente antibacteriano o un agente anti-viral. La composición farmacéutica aquí descrita puede comprender una o más de tales sustancias. Típicamente, una respuesta del sistema inmunitario innato incluye células inmunitarias de reclutamiento a sitios de infección mediante la producción de factores químicos, incluyendo mediadores químicos especializados denominados citoquinas; activación de la cascada del complemento; identificación y eliminación de sustancias extrañas presentes en órganos, tejidos, sangre y linfa por glóbulos blancos especializados; activación del sistema inmunitario adaptativo; y/o actuando como barrera física y química a agentes infecciosos.

Sitio de clonación: Un sitio de clonación se entiende típicamente como un segmento de una molécula de ácido nucleico adecuada para la inserción de una secuencia de ácidos nucleicos, por ejemplo una secuencia de ácidos nucleicos que comprende un marco de lectura abierto. La inserción se puede llevar a cabo por cualquier método biológico molecular conocido por el experto en la técnica, por ejemplo por restricción y ligación. Un sitio de clonación comprende típicamente uno o más sitios de reconocimiento de enzimas de restricción (sitios de restricción). Estos sitios de una o más restricciones pueden ser reconocidos por las 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.

Molécula de ácido nucleico: Una molécula de ácido nucleico es una molécula que comprende, preferentemente que consiste en, componentes de ácido nucleico. El término molécula de ácido nucleico se refiere preferentemente a moléculas de ADN o ARN. Se utiliza preferentemente como sinónimo del término “polinucleótido”. De manera preferente, una molécula de ácido nucleico es un polímero que comprende o que consiste en monómeros de nucleótidos que se unen covalentemente entre sí por enlaces fosfodiéster a un esqueleto azúcar/fosfato. El término “molécula de ácido nucleico” también abarca moléculas de ácido nucleico modificadas, tales como moléculas de ADN o ARN modificadas con base o modificadas con azúcar o modificadas en su esqueleto.

Marco de lectura abierto: Un marco de lectura abierto (ORF) en el contexto de la presente descripción típicamente puede ser una secuencia de diversos tripletes de nucleótidos que se pueden traducir en un péptido o proteína. Un marco de lectura abierto preferiblemente contiene un codón de inicio, esto es una combinación de tres nucleótidos subsiguientes que codifican usualmente para el aminoácido metionina (ATG o AUG), en su extremo 5'- y una región posterior que usualmente tiene una longitud múltiplo de 3 nucleótidos. Un ORF preferiblemente está terminado por un codón de parada (por ejemplo, TAA, TAG, TGA). Típicamente, éste es únicamente codón de parada del marco de lectura abierto. Por tanto, un marco de lectura abierto en el contexto de la presente descripción es preferiblemente una secuencia de nucleótidos que consiste en un número de nucleótidos que se pueden dividir entre tres, que inician con un codón de inicio (por ejemplo ATG o AUG) y que preferiblemente terminan con un codón de parada (por ejemplo, TAA, TGA, o TAG o UAA, UAG, UGA, respectivamente). El marco de lectura abierto se puede aislar o se puede incorporar en una secuencia de ácido nucleico más larga, por ejemplo en un vector o un ARNm. Un marco de lectura abierto también se puede denominar "región de codificación de proteína" o "región de codificación".

Péptido: Un péptido o polipéptido es típicamente un polímero de monómeros de aminoácidos unidos por enlaces peptídicos. Contiene típicamente menos de 50 unidades monoméricas. No obstante, el término péptido no excluye moléculas que tienen más de 50 unidades monoméricas. Los péptidos largos también se denominan polipéptidos, que tienen típicamente entre 50 y 600 unidades monoméricas.

Cantidad farmacéuticamente efectiva: Una cantidad farmacéuticamente efectiva en el contexto de la presente descripción se entiende típicamente como una cantidad que es suficiente para inducir un efecto farmacéutico, tal como una respuesta inmunitaria, que altera un nivel patológico de un péptido o proteína expresado, o sustituye un producto carente de gen, por ejemplo, en el caso de una situación patológica.

Proteína: Una proteína comprende típicamente uno o más péptidos o polipéptidos. Una proteína típicamente está plegada en una forma tridimensional, que puede ser necesaria para que la proteína ejerza su función biológica.

Secuencia Poli(A): Una secuencia poli(A), también llamada cola poli(A) o cola 3'-poli(A), se entiende típicamente como una secuencia de nucleótidos adenina, por ejemplo de hasta 400 nucleótidos de adenina, por ejemplo de aproximadamente 20 a aproximadamente 400, de manera preferente de aproximadamente 50 a aproximadamente 400, de manera más preferente de aproximadamente 50 a aproximadamente 300, aun de manera más preferente de aproximadamente 50 a aproximadamente 250, de manera mucho más preferente de aproximadamente 60 a aproximadamente 250 nucleótidos de adenina, que preferentemente se añade al terminal 3' de un ARNm. La secuencia poli(A) se sitúa típicamente en el extremo 3' de un ARNm. En el contexto de la presente descripción, una secuencia poli(A) se puede situar dentro de un ARNm o de cualquier otra molécula de ácido nucleico, por ejemplo un vector, por ejemplo un vector que sirve como plantilla para la generación de un ARN, preferentemente un ARNm, por ejemplo por transcripción del vector. En el contexto de la presente descripción, el ´termino "secuencia poli(A)" también comprende además elementos de secuencia, preferentemente elementos de secuencia artificiales, que son parte de la 3'-UTR o que se localizan en el extremo 3' de la molécula de ácido nucleico artificial, y que preferentemente comprenden hasta 1.100 nucleótidos adenina, en especial al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900, o al menos 1.000 nucleótidos adenina.

Poliadenilación: La poliadenilación se entiende típicamente como la adición de una secuencia poli(A) a una secuencia de ácido nucleico, tal como una molécula de ARN, por ejemplo un ARNm prematuro. La poliadenilación se puede inducir por una llamada señal de poliadenilación. Esta señal se sitúa preferentemente dentro de un tramo de nucleótidos cerca o en el extremo 3' de una molécula de ácido nucleico, tal como una molécula de ARN, que se poliadenila. En el contexto de la presente descripción, una señal de poliadenilación también puede estar comprendida en la 3'-UTR de la molécula de ácido nucleico artificial. Típicamente una señal de poliadenilación comprende un hexámero que consiste en adenina y nucleótidos uracilo/timina, preferentemente la secuencia de hexámero AAUAAA. Otras secuencias, de manera preferente secuencias de hexámero, también son concebibles. La poliadenilación aparece típicamente durante el procesamiento de un pre-ARNm (también llamado ARNm prematuro). Típicamente, la maduración del ARN (de pre-ARNm a ARNm maduro) comprende la etapa de poliadenilación. Tal como se usa en el contexto de la presente descripción, el término puede relacionarse con la poliadenilación del ARN como un proceso celular, así como a una poliadenilación llevada a cabo por reacción enzimática *in vitro* o mediante síntesis química.

Sitio de restricción: Un sitio de restricción, también llamado "sitio de reconocimiento de enzimas de restricción", es una secuencia de nucleótidos reconocida por la enzima de restricción. Un sitio de restricción es típicamente una secuencia de nucleótidos, preferentemente palindrómicos corta, por ejemplo una secuencia que comprende de 4 a 8 nucleótidos. Preferentemente, un sitio de restricción es reconocido por una enzima de restricción. La enzima de restricción escinde típicamente una secuencia de nucleótidos que comprende un sitio

de restricción en este sitio. En una secuencia de nucleótidos de doble hebra, tal como una secuencia de ADN de doble hebra, la enzima de restricción corta típicamente ambas hebras de la secuencia de nucleótidos.

ARN, ARNm: ARN es la abreviatura usual para el ácido ribonucleico. Es una molécula de ácido nucleico, es decir un polímero que consiste en nucleótidos. Estos nucleótidos son usualmente monómeros de adenosina-monofosfato, uridina-monofosfato, guanosina-monofosfato y citidina-monofosfato que se unen entre sí a lo largo de un esqueleto principal. El esqueleto principal está formado por la unión fosfodiéster entre el azúcar, es decir ribosa, de una primera porción de fosfato y de un segundo monómero adyacente. La sucesión específica de monómeros se denomina secuencia de ARN. Usualmente el ARN puede ser obtenible por la transcripción de una secuencia de ADN, por ejemplo dentro de una célula. En las células eucarióticas, la transcripción se lleva a cabo típicamente dentro del núcleo de la mitocondria. *In vivo*, la transcripción del ADN da por resultado normalmente el llamado ARN prematuro, que tiene que ser procesado en el llamado ARN mensajero, usualmente abreviado como ARNm. El procesamiento del ARN prematuro, por ejemplo en organismos eucarióticos, comprende una variedad de diferentes modificaciones pos-transcripcionales, como empalme, cap 5', poliadenilación, exportación del núcleo o la mitocondria y similares. La suma de estos procesos también se denomina maduración del ARN. El ARN mensajero maduro proporciona usualmente la secuencia de nucleótidos que se puede traducir en una secuencia de aminoácidos de un péptido o proteína particular. Típicamente, un ARNm maduro comprende una cap 5', 5'UTR, un marco de lectura abierto, una 3'UTR y una secuencia poli(A). Además del ARN mensajero, existen varios tipos de ARN no codificantes que pueden estar implicados en la regulación de la transcripción y/o traducción.

Secuencia de una molécula de ácido nucleico: La secuencia de una molécula de ácido nucleico se entiende típicamente como el orden particular e individual, es decir la sucesión de sus nucleótidos. La secuencia de una proteína o péptido se entiende típicamente como el orden, es decir la sucesión de sus aminoácidos.

Identidad de secuencia: Dos o más secuencias son idénticas si tienen la misma longitud y orden de nucleótidos o aminoácidos. El porcentaje de identidad típicamente describe el grado en que dos secuencias son idénticas, esto es típicamente describe el porcentaje de nucleótidos que corresponden en su posición de secuencia con nucleótidos idénticos de una secuencia de referencia. Para determinar el grado de identidad, las secuencias a comparar se consideran de la misma longitud, esto es la longitud de la secuencia más larga de las secuencias a comparar. Esto significa que una primera secuencia que consiste en 8 nucleótidos es un 80% idéntica a una segunda secuencia que consiste en 10 nucleótidos que comprende la primera secuencia. En otras palabras, en el contexto de la presente invención, la identidad de secuencia preferiblemente se refiere al porcentaje de nucleótidos de una secuencia que tiene la misma posición en dos o más secuencias de la misma longitud. Los espacios son usualmente considerados como posiciones no idénticas, independientemente de su posición actual en una alineación.

Molécula de ácido nucleico estabilizada: Una molécula de ácido nucleico estabilizada es una molécula de ácido nucleico, preferentemente una molécula de ADN o ARN, que se modifica de forma que es más estable a la desintegración o degradación, por ejemplo por factores ambientales o digestión enzimática, tal como por degradación con exo o endonucleasas, que la molécula de ácido nucleico sin la modificación. De manera preferente, una molécula de ácido nucleico estabilizado en el contexto de la presente descripción se estabiliza en una célula, tal como una célula procariótica o eucariótica, de manera preferente en una célula de mamífero, tal como una célula humana. El efecto de estabilización también se puede ejercer fuera de las células, por ejemplo con una solución tampón, etc., por ejemplo en un proceso de fabricación para una composición farmacéutica que comprende la molécula de ácido nucleico estabilizada.

Transfección: El término 'transfección' se refiere a la introducción de moléculas de ácido nucleico, tal como moléculas de ADN o ARN (por ejemplo ARNm), en células, preferiblemente en células eucariotas. En el contexto de la presente descripción, el término 'transfección' abarca cualquier método conocido por el experto para introducir moléculas de ácido nucleico en las células, preferiblemente en células eucariotas, tal como en células de mamífero. Tales métodos abarcan, por ejemplo, electroporación, lipofección, por ejemplo basado en lípidos catiónicos y/o liposomas, precipitación de fosfato de calcio, transfección basada en nanopartículas, transfección basada en virus o transfección basada en polímeros catiónicos, tal como DEAE-dextrano o polietilenimina, etc. De manera preferente, la introducción no es viral.

Vacuna: Una vacuna se entiende típicamente como un material profiláctico o terapéutico que proporciona al menos un antígeno, preferiblemente un inmunógeno. El antígeno o inmunógeno se puede derivar de cualquier material adecuado para la vacunación. Por ejemplo, el antígeno o inmunógeno se puede derivar de un patógeno, tal como de bacterias o partículas virales, etc., o de un tumor o tejido canceroso. El antígeno o inmunógeno estimula el sistema inmunitario adaptativo del cuerpo para proporcionar una respuesta inmunitaria adaptiva.

Vector: El término “vector” se refiere a una molécula de ácido nucleico, de manera preferente a una molécula de ácido nucleico artificial. Un vector en el contexto de la presente descripción es adecuado para incorporar o alojar 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. Así, el vector puede comprender una secuencia que corresponde, por ejemplo, a una secuencia de ARNm deseada o a una parte de la misma, tal como una secuencia que corresponde al marco de lectura abierto y la 3’UTR de un ARNm. Se puede emplear 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 las secuencias necesarias para la transcripción de un tramo de secuencia del vector, tal como una secuencia promotora, por ejemplo una secuencia promotora de ARN. Un vector de clonación es típicamente un vector que contiene un sitio de clonación, que se puede utilizar para incorporar secuencias de ácidos nucleicos en el vector. Un vector de clonación puede ser, por ejemplo, un vector plásmido o bacteriófago. Un vector de transferencia puede ser un vector adecuado para transferir moléculas de ácido nucleico en células u organismos, por ejemplo vectores virales. Un vector en el contexto de la presente descripción puede ser, por ejemplo, un vector de ARN o un vector de ADN. Preferentemente, un vector es una molécula de ADN. De manera preferente 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 multiplicación del vector, tal como un origen de replicación. De manera preferente, un vector en el contexto de la presente solicitud es un vector de plásmido.

Vehículo: Un vehículo se entiende típicamente como un material adecuado para almacenar, transportar y/o administrar un compuesto, tal como un compuesto farmacéuticamente activo. Por ejemplo, puede ser un líquido fisiológicamente aceptable adecuado para almacenar, transportar y/o administrar un compuesto farmacéuticamente activo.

Región 3’ no traducida (3’UTR): En general, el término “3’-UTR” se refiere a una parte de una molécula de ácido nucleico artificial que se localiza 3? (esto es “aguas abajo”) de marco de lectura abierto y que no se traduce en una proteína. Típicamente una 3’-UTR es la parte de un ARNm entre la región de codificación de la proteína (es decir el marco de lectura abierto (ORF)) y la secuencia poli(A) del ARNm. En el contexto de la presente descripción, una 3’-UTR de la molécula de ácido nucleico artificial puede comprender más de un elemento 3’-UTR, pudiendo ser de origen diferente, tal comoelementos de secuencia derivados de la 3’-UTR de diversos genes naturales (no relacionados). Así, el término 3’-UTR puede comprender también elementos que no son codificados en el patrón a partir del cual de transcribe el ARN, sino que son añadidos después de la transcripción, durante la maduración, por ejemplo una secuencia poli(A). Una 3’-UTR del ARNm no se traduce en una secuencia de aminoácidos. La secuencia 3’-UTR es codificada generalmente por el gen que se transcribe en el ARNm respectivo durante el proceso de expresión génica. La secuencia genómica primero se transcribe en el ARNm pre-maduro, que comprende intrones opcionales. El ARNm pre-maduro luego se procesa adicionalmente en ARNm maduro en un proceso de maduración. Este proceso de maduración comprende las etapas de cap 5’, empalme del ARNm pre-maduro para escindir intrones opcionales y modificaciones del extremo 3’, tal como poliadenilación del extremo 3’ del ARNm pre-maduro y escisiones de endo o exonucleasa opcionales, etc. En el contexto de la presente descripción, una 3’-UTR corresponde a la secuencia de un ARNm maduro que se sitúa 3’ al codón de parada de la región que codifica la proteína, preferentemente inmediatamente 3’ al codón de parada de la región que codifica a la proteína, y la secuencia poli(A) del ARNm. El término “corresponde a” significa que la secuencia 3’-UTR puede ser una secuencia de ARN , tal como en la secuencia de ARNm utilizada para definir la secuencia 3’-UTR, o una secuencia de ADN que se corresponde con tal secuencia de ARN. En el contexto de la presente descripción, el término “3’-UTR de un gen”, tal como “3’-UTR de un gen de proteína ribosómica”, 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 la maduración de la ARNm pre-maduro. El término “3’-UTR de un gen” abarca la secuencia de ADN y la secuencia de ARN (ambas hebras sentido y anti-sentido y ambos, maduro e inmaduro) de la 3’-UTR. Tal como se usa aquí, el término “elemento 3’-UTR” típicamente se refiere a un fragmento de una 3’-UTR como se define aquí. En particular, el término comprende cualquier elemento de secuencia de ácido nucleico que se sitúa 3’ al ORF en la molécula de ácido nucleico artificial, preferente el ARNm, aquí descrita. Así, el término abarca, por ejemplo, elementos de secuencia derivados de la 3’-UTR de un gen heterólogo, así como elementos tales como una secuencia poli(C) o un tallo-bucle de histona.

Región no traducida 5' (5'-UTR): Una 5'-UTR se entiende típicamente como una sección particular de ARN mensajero (ARNm). Se sitúa 5' del marco de lectura abierto del ARNm. Típicamente, la 5'-UTR comienza con el sitio de inicio de la transcripción 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, también denominados elementos reguladores. Dichos elementos reguladores pueden ser, por ejemplo, sitios de unión a ribosomas. La 5'-UTR

puede modificarse postranscripcionalmente, por ejemplo por adición de una 5'-cap. En el contexto de la presente descripción, una 5'-UTR corresponde a la secuencia de un ARNm maduro entre la 5'-cap y el codón de inicio. Preferiblemente, la 5'-UTR corresponde a la secuencia que se extiende desde un nucleótido situado 3' a la 5'-cap, preferiblemente desde el nucleótido ubicado inmediatamente 3' a la 5'-cap, hasta un nucleótido situado 5' del codón de inicio de la región de codificación de proteínas, preferiblemente al nucleótido ubicado inmediatamente 5' al codón de inicio de la región de codificación de proteínas. El nucleótido localizado inmediatamente 3' a la 5'-cap de un ARNm maduro típicamente corresponde al sitio de inicio de la transcripción. El término "corresponde a" significa que la secuencia 5'-UTR puede ser una secuencia de ARN, como en la secuencia de ARNm utilizada para definir la secuencia 5'-UTR, o una secuencia de ADN que corresponde a dicha secuencia de ARN. En el contexto de la presente descripción, el término "una 5'-UTR de un gen" es la secuencia que corresponde a la 5'-UTR del ARNm maduro derivado de ese gen, es decir, el ARNm obtenido por transcripción del gen y maduración del ARNm premaduro. El término "5'-UTR de un gen" abarca la secuencia de ADN y la secuencia de ARN de la 5'-UTR.

Tracto de oligopirimidina terminal 5' (TOP): El tracto de oligopirimidina terminal 5' (TOP) es típicamente un tramo de nucleótidos pirimidina en la región terminal 5' de una molécula de ácido nucleico, tal como la región terminal 5' de ciertas moléculas de ARNm o la región terminal 5' de una entidad funcional, por ejemplo la región transcrita, de ciertos genes. La secuencia comienza con una citidina, que normalmente corresponde al sitio de inicio de la transcripción, y es seguida por un tramo de aproximadamente 3 a 30 nucleótidos pirimidina. Por ejemplo, la TOP puede comprender 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 , 25, 26, 27, 28, 29, 30 o incluso más nucleótidos. El tramo de pirimidina y, por tanto la TOP 5', termina en un nucleótido 5' al primer nucleótido purina ubicado aguas abajo de la TOP. El ARN mensajero que contiene un tracto de oligopirimidina terminal 5' a menudo se denomina ARNm TOP. En consecuencia, los genes que proporcionan tales ARN mensajeros se denominan genes TOP. Se han encontrado secuencias TOP, por ejemplo, en genes y ARNm que codifican factores de elongación de péptidos y proteínas ribosómicas.

Motivo TOP: En el contexto de la presente descripción, un motivo TOP es una secuencia de ácido nucleico que corresponde a una 5' TOP como se definió anteriormente. Así, un motivo TOP en el contexto de la presente descripción es preferiblemente un tramo de nucleótidos pirimidina de una longitud de 3-30 nucleótidos. Preferiblemente, el motivo TOP consiste en al menos 3 nucleótidos pirimidina, preferiblemente al menos 4 nucleótidos pirimidina, preferiblemente al menos 5 nucleótidos pirimidina, más preferiblemente al menos 6 nucleótidos, más preferiblemente al menos 7 nucleótidos, con mayor preferencia al menos 8 nucleótidos pirimidina, donde el tramo de nucleótidos de pirimidina comienza preferiblemente en su extremo 5' con un nucleótido citosina. En los genes TOP y los ARNm TOP, el motivo TOP comienza preferiblemente en su extremo 5' con el sitio de inicio de la transcripción y termina en un nucleótido 5' al primer residuo de purina en dicho gen o ARNm. Un motivo TOP en el sentido de la presente descripción preferiblemente se dispone en el extremo 5' de una secuencia que representa una 5'-UTR o en el extremo 5' de una secuencia que codifica una 5'-UTR. Por tanto, preferiblemente, un tramo de 3 o más nucleótidos de pirimidina se denomina "motivo TOP" en el sentido de la presente descripción si este tramo está situado en el extremo 5' de una secuencia respectiva, tal como la molécula de ácido nucleico artificial, el elemento 5'-UTR de la molécula de ácido nucleico artificial o la secuencia de ácido nucleico que se deriva de la 5'-UTR de un gen TOP como se describe aquí. En otras palabras, un tramo de 3 o más nucleótidos pirimidina que no se encuentra en el extremo 5' de un elemento 5'-UTR o de una 5'-UTR, sino en cualquier lugar dentro de un elemento 5'-UTR o de una 5'-UTR, preferiblemente no se denomina "motivo TOP".

Gen TOP: Los genes TOP se caracterizan típicamente por la presencia de un tracto de oligopirimidina 5' terminal. Además, la mayoría de los genes TOP se caracterizan por una regulación de traducción asociada al crecimiento. Sin embargo, también se conocen genes TOP con una regulación de traducción específica de tejido. Como se definió anteriormente, la 5'-UTR de un gen TOP corresponde a la secuencia de una 5'-UTR de un ARNm maduro derivado de un gen TOP, que preferiblemente se extiende desde el nucleótido 3' al 5'-CAP hasta el nucleótido 5' del codón de inicio. Una 5'-UTR de un gen TOP típicamente no comprende ningún codón de inicio, preferiblemente sin AUG (uAUG) aguas arriba o marcos de lectura abiertos (uORF) aguas arriba. Allí, los AUG aguas arriba y los marcos de lectura abiertos aguas arriba se entienden típicamente como AUG y marcos de lectura abiertos que ocurren 5' del codón de inicio (AUG) del marco de lectura abierto a traducirse. Las 5'-UTR de los genes TOP generalmente son bastante cortas. Las longitudes de las 5'-UTR de los genes TOP pueden variar entre 20 nucleótidos y hasta 500 nucleótidos y típicamente son inferiores a aproximadamente 200 nucleótidos, preferiblemente inferiores a aproximadamente 150 nucleótidos, con mayor preferencia inferiores a aproximadamente 100 nucleótidos. Ejemplos de 5'-UTR de genes TOP en el sentido de la presente descripción son aquellas secuencias de ácido nucleico que se extienden desde el nucleótido en la posición 5 hasta el nucleótido ubicado inmediatamente 5' al codón de inicio (por ejemplo ATG) en las secuencias de acuerdo con las SEQ ID NO. 1-1363 de la solicitud de patente WO2013/143700. En este contexto, un fragmento particularmente preferente de una 5'-UTR de un gen TOP es una 5'-UTR de un gen

TOP que carece del motivo 5' TOP. Los términos “5'-UTR de un gen TOP” o “5’-UTR TOP” se refieren preferiblemente a la 5'-UTR de un gen TOP natural.

En un primer aspecto, la presente invención se refiere a una molécula de ácido nucleico artificial que comprende

a) al menos un marco de lectura abierto (ORF); y
b) al menos una región 3’-no traducida (3’-UTR) que comprende
al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, y
donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende un tallo-bucle de histona según la SEQ ID NO: 11,

para su uso como un medicamento, para su uso como una vacuna o para su uso en la terapia génica,

donde la molécula de ácido nucleico artificial está asociada a o complejada con un compuesto catiónico policatiónico.

También se describe aquí una molécula de ácido nucleico artificial que comprende

a) al menos un marco de lectura abierto (ORF); y
b) i) al menos una región 3’-no traducida (3’-UTR) que comprende
al menos una secuencia poli(A), donde la al menos una secuencia poli(A) comprende al menos 70 nucleótidos adenina o
b) ii) una señal de poliadenilación.

La molécula de ácido nucleico artificial aquí descrita puede ser un ARN, tal como un ARNm, un ADN, tal como un vector de ADN, o puede ser una molécula de ARN o ADN modificada. Se puede proporcionar como una molécula bicatenaria con una cadena sentido y una cadena antisentido, por ejemplo como una molécula de ADN con una cadena sentido y una cadena antisentido.

Preferentemente la molécula de ácido nucleico artificial aquí descrita se caracteriza preferiblemente por una expresión de proteína aumentada con respecto a una molécula de ácido nucleico de referencia.

En el contexto de la presente descripción, una "molécula de ácido nucleico de referencia" es una molécula de ácido nucleico, que típicamente comprende el mismo ORF que la molécula de ácido nucleico artificial, y que carece de una 3'-UTR o que comprende una 3'- UTR de referencia. Preferiblemente, la molécula de ácido nucleico de referencia comprende una 3'-UTR (es decir, una "3'-UTR de referencia") que no comprende una secuencia poli(A) o que comprende una secuencia poli(A) donde el número total de nucleótidos adenina comprendidos en una o más secuencias poli(A) de la 3'-UTR es inferior al número total de nucleótidos adenina comprendidos en la al menos una secuencia poli(A) de la 3'-UTR de la molécula de ácido nucleico artificial aquí descrita, preferentemente inferior a 70. Preferiblemente, la molécula de ácido nucleico de referencia tiene, además del contenido diferente de nucleótidos adenina en una secuencia poli(A), la misma estructura general, es decir, comprende las mismas características estructurales, tales como, por ejemplo, una estructura 5'-cap, una 5'-UTR o una 3'-UTR. Con especial preferencia, la molécula de ácido nucleico de referencia comprende o consiste en, además del contenido diferente de nucleótidos adenina en una secuencia poli(A), la misma secuencia de ácido nucleico que la molécula de ácido nucleico artificial. En el contexto de la presente descripción, la secuencia de ácido nucleico natural (por ejemplo un ARNm) que comprende el ORF de la molécula de ácido nucleico artificial también puede ser la molécula de ácido nucleico de referencia.

En ciertas realizaciones, una molécula de ARN comprende una secuencia 5'-UTR, un ORF y una secuencia 3'-UTR, donde la secuencia 5'-UTR o la secuencia 3'-UTR es heteróloga en relación al ORF del ARNm (por ejemplo donde el ARN no comprende la secuencia 5'-UTR y/o la secuencia 3'-UTR de un ARN de tipo salvaje que codifica el ORF).

Tal como se usa aquí, el término "número total de nucleótidos de adenina" se refiere típicamente a la suma de todos los nucleótidos adenina que están comprendidos en una o más secuencias poli(A) en la 3'-UTR de la molécula de ácido nucleico artificial aquí descrita. En particular, en caso de que la 3'-UTR comprenda más de una secuencia poli(A), el término se refiere a la suma de los nucleótidos adenina comprendidos en todas las secuencias poli(A) comprendidas en la 3'-UTR de la molécula de ácido nucleico artificial.

En una realización aquí descrita, la molécula de ácido nucleico artificial comprende una 3'-UTR donde el

número total de nucleótidos adenina comprendidos en una o más secuencias poli(A) de la 3'-UTR es al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1.000 o 1.100 nucleótidos adenina. De acuerdo con esa realización, la 3'-UTR puede comprender, por ejemplo, dos secuencias poli(A), preferiblemente separadas, donde la suma de nucleótidos adenina comprendidos en dichas dos secuencias poli(A) es al menos 60, 70, 80 , 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1.000 o 1.100.

En una realización adicional, se proporciona una composición que comprende una pluralidad de moléculas de ARN de las realizaciones en un vehículo farmacéuticamente aceptable, donde al menos aproximadamente el 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% , 98%, 99% o más del ARN de la composición comprende una secuencia poli(A) que difiere en longitud en no más de 10 nucleótidos. En una realización preferente, al menos aproximadamente el 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más del ARN de la composición comprende una secuencia poli(A) de idéntica longitud. En ciertas realizaciones, la secuencia poli(A) se dispone en el extremo 3' del ARN, sin otros nucleótidos posicionados en 3' con respecto a la secuencia poli(A). En otra realización adicional, se proporciona una composición que comprende una pluralidad de moléculas de ARN de las realizaciones en un vehículo farmacéuticamente aceptable, donde dicha pluralidad de moléculas de ARN comprende tanto ARN terminados en cap como no terminados en cap. Por ejemplo, en algunos aspectos, una composición comprende una pluralidad de moléculas de ARN donde no más del 95%, 90%, 80%, 70% o 60% de los ARN comprenden una cap y las moléculas de ARN restantes están no terminadas en cap.

En una realización preferente, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos una secuencia poli(A), donde la al menos una secuencia poli(A) comprende al menos 60, 70, 80, 90, 100, 1 10, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos adenina .

El número total de nucleótidos adenina comprendidos en una o más secuencias poli(A) puede ser de hasta aproximadamente 1.200 nucleótidos adenina, por ejemplo de aproximadamente 70 a aproximadamente 1.100, preferiblemente de aproximadamente 80 a aproximadamente 800, más preferiblemente de aproximadamente 90 a aproximadamente 700, incluso más preferiblemente de aproximadamente 100 a aproximadamente 500, con especial preferencia de aproximadamente 120 a aproximadamente 450 nucleótidos adenosina. Preferiblemente, la 3'-UTR de la molécula de ácido nucleico artificial comprende al menos 75, 80, 85, 95, 100, 105, 1 10, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160 , 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, o al menos 225, más preferiblemente al menos 220, nucleótidos adenina comprendidos en las una o más secuencias poli(A). Preferiblemente, la 3'-UTR de la molécula de ácido nucleico artificial comprende de 180 a 1.100 nucleótidos adenina comprendidos en la una o más secuencias poli(A), en especial de 200 a 1.100 nucleótidos adenina, incluso más preferiblemente de 210 a 250 nucleótidos adenina, más preferiblemente de 215 a 240 nucleótidos adenina. En una realización particularmente preferente, la 3'-UTR comprende aproximadamente 220 nucleótidos, preferiblemente aproximadamente 224 nucleótidos. Alternativamente, la 3'-UTR comprende al menos 300 nucleótidos adenina comprendidos en la una o más secuencias poli(A), preferiblemente al menos 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430 o al menos 440 nucleótidos adenina comprendidos en la una o más secuencias poli(A). En esa realización, la 3'-UTR de la molécula de ácido nucleico artificial preferentemente comprende de 300 a 650 nucleótidos adenina comprendidos en la una o más secuencias poli(A), más preferiblemente de 400 a 500 nucleótidos adenina, incluso más preferiblemente de 420 a 470 nucleótidos adenina, más preferiblemente de 430 a 460 nucleótidos adenina. En una realización particularmente preferente, la 3'-UTR comprende un número total de aproximadamente 440 nucleótidos adenina, preferiblemente aproximadamente 444 nucleótidos adenina, comprendidos en la una o más secuencias poli(A). Alternativamente, la 3'-UTR comprende al menos 900 nucleótidos adenina comprendidos en la una o más secuencias poli(A), preferiblemente al menos 900, 950 o al menos 1.000 nucleótidos adenina comprendidos en la una o más secuencias poli(A). En esa realización, la 3'-UTR de la molécula de ácido nucleico artificial preferentemente comprende de 900 a 1.100 nucleótidos adenina comprendidos en la una o más secuencias poli(A). En una realización particularmente preferente, la 3'-UTR comprende un número total de aproximadamente 1.064 nucleótidos adenina comprendidos en la una o más secuencias poli(A).

En una realización preferente, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos una secuencia poli(A) que comprende al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240 o 250 nucleótidos adenina, más preferiblemente al menos 150 nucleótidos adenina, incluso más preferiblemente al menos 160 nucleótidos adenina. Preferentemente, el número de nucleótidos adenina comprendidos en la al menos una secuencia poli(A) es de 110 a 200, de 120 a 200, de 130 a 190, de 140 a 180 o de 150 a 170. Alternativamente, la 3'-UTR comprende al menos una secuencia poli(A) que preferiblemente comprende al menos 300 nucleótidos adenina, más preferiblemente al menos 350 nucleótidos adenina, incluso más preferiblemente al menos 380 nucleótidos adenina. Preferentemente, la al menos una secuencia poli(A) comprende de 320 a 430, de 330 a 420, de 340 a 410, de

350 a 400, de 360 a 400, o de 370 a 390. Alternativamente, la 3'-UTR comprende al menos una secuencia poli(A) que preferiblemente comprende al menos 900 nucleótidos adenina, más preferiblemente al menos 900, 950 o al menos 1.000 nucleótidos adenina. En esa realización, la 3'-UTR de la molécula de ácido nucleico artificial preferentemente comprende al menos una secuencia poli(A) que preferiblemente comprende de 900 a 1.100 o de 1.000 a 1.100 nucleótidos adenina.

La al menos una secuencia poli(A) puede ubicarse en cualquier posición dentro de la 3'-UTR. Así, la al menos una secuencia poli(A) puede ubicarse en el extremo 3' de la 3'-UTR, es decir, la 3'-UTR preferiblemente no contiene más de 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótidos dispuestos en 3' de dicha secuencia poli(A); más preferiblemente, la 3'-UTR no contiene otros elementos ubicados 3' a dicha secuencia poli(A). En una realización preferente, la al menos una secuencia poli(A) está ubicada en el extremo 3' de la molécula de ácido nucleico artificial, es decir, la molécula de ácido nucleico artificial preferiblemente no contiene más de 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótidos ubicados en 3' de dicha secuencia poli(A). Alternativamente, la al menos una secuencia poli(A) puede ubicarse en el extremo 5' de la 3'-UTR, es decir, inmediatamente 3' del ORF de la molécula de ácido nucleico artificial, o ubicarse dentro de la 3'-UTR, es decir flanqueada en el lado 5' y en el lado 3' por otros elementos 3'-UTR. Preferentemente, la al menos una secuencia poli(A) está flanqueada en el lado 3' por una secuencia poli(C) y/o por una secuencia tallo-bucle de histona. Además o alternativamente, la al menos una secuencia poli(A) está flanqueada en el lado 5' por un elemento 3'-UTR derivado de un gen de albúmina o globina, preferiblemente humano, preferiblemente albúmina7 como se define aquí.

En una realización preferente, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos una secuencia poli(A) donde al menos una secuencia poli(A) se localiza en el extremo 3' de la 3'-UTR y donde la secuencia poli(A) en el extremo 3' de la 3'-UTR comprende al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos adenina, más preferiblemente al menos 150 nucleótidos adenina, incluso más preferiblemente al menos 160 nucleótidos de adenina. Preferentemente, el número de nucleótidos adenina comprendidos en la secuencia poli(A) ubicada en el extremo 3' de la 3'-UTR es de 110 a 200, de 120 a 200, de 130 a 190, de 140 a 180 o de 150 a 170. Alternativamente, la secuencia poli(A) en el extremo 3' de la 3'-UTR comprende preferiblemente al menos 300 nucleótidos adenina, más preferiblemente al menos 350 nucleótidos adenina, incluso más preferiblemente al menos 380 nucleótidos adenina. Preferentemente, el número de nucleótidos adenina comprendido en la secuencia poli(A) ubicada en el extremo 3' de la 3'-UTR es de 320 a 430, de 330 a 450, de 340 a 410, de 350 a 400, de 360 a 400 o de 370 a 390. En una realización preferente, la secuencia poli(A) ubicada en el extremo 3' de la 3'-UTR comprende aproximadamente 160, aproximadamente 380 o aproximadamente 430 nucleótidos de adenina. Además, alternativamente, la secuencia poli(A) en el extremo 3' de la 3'-UTR comprende preferiblemente al menos 900 nucleótidos adenina, más preferiblemente al menos 900, 950 o al menos 1.000 nucleótidos adenina. Preferiblemente, el número de nucleótidos adenina comprendidos en la secuencia poli(A) ubicada en el extremo 3' de la 3'-UTR es de 900 a 1.100 o de 1.000 a 1.100 nucleótidos adenina. En una realización particularmente preferente, la secuencia poli(A) en el extremo 3' de la 3'-UTR comprende preferiblemente aproximadamente 1.000 nucleótidos adenina.

En una realización, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende solo una secuencia poli(A) y, opcionalmente, otros elementos 3'-UTR como se definen aquí. Así, la 3'-UTR preferiblemente comprende además un elemento 3'-UTR opcional, preferiblemente como se define aquí, una secuencia poli(C) y/o una secuencia tallo-bucle de histona. En una realización particularmente preferente, la 3'-UTR comprende, preferiblemente en la dirección de 5' a 3', un elemento 3'-UTR opcional, preferiblemente como se define aquí, una secuencia poli(A), una secuencia poli(C) y una secuencia tallo-bucle de histona.

En otra realización, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos dos secuencias poli(A). Preferentemente, una primera secuencia poli(A) está ubicada en el terminal 5' de la 3'-UTR o ubicada en una posición dentro de la 3'-UTR, es decir, flanqueada en el lado 5' y en el lado 3' por otros elementos UTR, mientras que una segunda secuencia poli(A) se encuentra en una posición dentro de la 3'-UTR o en el extremo 3' de la 3'-UTR. Preferiblemente, la primera secuencia poli(A) está flanqueada en el lado 3' por una secuencia poli(C) y/o una secuencia tallo-bucle de histona. Además o alternativamente, la primera secuencia poli(A) está flanqueada en el lado 5’ por un elemento 3’-UTR derivado de un gen de albúmina o globina, preferentemente humano, preferentemente albúmina7 como se define aquí.

En una realización preferente, la molécula de ácido nucleico artificial comprende al menos dos secuencias poli(A) que están separadas una de la otra por una secuencia de nucleótidos que comprende o consiste en al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145 o 150 nucleótidos, donde la secuencia de nucleótidos preferiblemente no comprende más de 10, 9, 8, 7, 6, 5, 4, 3 o 2 nucleótidos adenina consecutivos. Preferiblemente, la secuencia de nucleótidos que separa la primera y la segunda secuencia poli(A) comprende de 1 a aproximadamente 200 nucleótidos, preferiblemente de 10 a 90, de 20 a 85, de 30 a 80, de 40 a 80, de 50 a 75 o de 55 a 85 nucleótidos,

más preferiblemente de 55 a 80 nucleótidos, donde la secuencia de nucleótidos preferiblemente no comprende más de 10, 9, 8, 7, 6, 5, 4, 3 o 2 nucleótidos de adenina consecutivos

Preferentemente, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos dos secuencias poli(A), donde la primera y/o la segunda secuencia poli(A) comprende preferiblemente al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos adenina, más preferiblemente al menos 150 nucleótidos adenina, incluso más preferiblemente al menos 160 nucleótidos adenina. En una realización preferente, la primera secuencia poli(A) comprende al menos 20, 30, 40, 50, 60, 70, 80 o 90 nucleótidos adenina. La primera secuencia poli(A) puede comprender además de 20 a 90, de 25 a 85, de 35 a 80 o de 45 a 75, preferiblemente de 60 a 70, nucleótidos adenina. En otra realización preferente, la primera secuencia poli(A) comprende aproximadamente 60 nucleótidos adenina. En una realización particularmente preferente, la primera secuencia poli(A) comprende o consiste en aproximadamente 64 nucleótidos adenina. La segunda secuencia poli(A) comprende preferiblemente al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos adenina, más preferiblemente al menos 150 nucleótidos adenina, incluso más preferiblemente al menos 160 nucleótidos adenina. En una realización particularmente preferente, la segunda secuencia poli(A) comprende aproximadamente 160 nucleótidos adenina. Alternativamente, la segunda secuencia poli(A) comprende preferiblemente al menos 300 nucleótidos adenina, más preferiblemente al menos 350 nucleótidos adenina, incluso más preferiblemente al menos 380 nucleótidos adenina o al menos 430 nucleótidos adenina. En una realización preferente, la segunda secuencia poli(A) comprende aproximadamente 380 o aproximadamente 430 residuos adenina. Preferiblemente, el número de nucleótidos adenina comprendidos en la segunda secuencia poli(A) es preferiblemente de 110 a 200, de 120 a 200, de 130 a 190, de 140 a 180 o de 150 a 170. Alternativamente, el número de nucleótidos adenina comprendidos en la segunda secuencia poli(A) es preferiblemente de 320 a 450, de 330 a 420, de 340 a 410, de 350 a 400, de 360 a 400 o de 370 a 390. Además, alternativamente, la segunda secuencia poli(A) comprende preferiblemente al menos 900 nucleótidos adenina, más preferiblemente al menos 900, 950 o al menos 1.000 nucleótidos de adenina. Preferiblemente, el número de nucleótidos adenina comprendidos en la segunda secuencia poli(A) es de 900 a 1.100 o de 1.000 a 1.100 nucleótidos adenina. En una realización particularmente preferente, la segunda secuencia poli(A) comprende preferiblemente aproximadamente 1.000 nucleótidos de adenina.

En una realización preferente, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos dos secuencias poli(A), donde la primera secuencia poli(A) está ubicada en el extremo 5' de la 3'-UTR o en una posición dentro de la 3'-UTR, es decir, flanqueada en el lado 5' y en el lado 3' por otros elementos UTR, y comprende al menos 20, 30, 40, 50, 60, 70, 80 o 90 nucleótidos adenina, preferiblemente de 20 a 90, de 25 a 85, de 35 a 80 o de 45 a 75, preferiblemente de 60 a 70, más preferiblemente aproximadamente 64, nucleótidos adenina. En esa realización, la segunda secuencia poli(A) comprende preferiblemente al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos adenina, más preferiblemente al menos 150 nucleótidos adenina, incluso más preferiblemente al menos 160 nucleótidos adenina, preferiblemente de 110 a 200, de 120 a 200, de 130 a 190, de 140 a 180 o de 150 a 170 nucleótidos de adenina. Allí, la segunda secuencia poli(A) se encuentra preferiblemente 3' de la primera secuencia poli(A), más preferiblemente en el extremo 3' de la 3'-UTR como se define aquí, o incluso más preferiblemente en el extremo 3' de la molécula de ácido nucleico artificial como se define aquí, donde la primera y la segunda secuencia poli (A) preferentemente están separadas, en especial separadas como se define aquí.

En otra realización preferida adicional, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos dos secuencias poli(A), donde la primera secuencia poli(A) está ubicada en el extremo 5' de la 3'-UTR o en una posición dentro de la 3'-UTR, es decir, flanqueada en el lado 5' y en el lado 3' por otros elementos UTR, y comprende al menos 20, 30, 40, 50, 60, 70, 80 o 90 nucleótidos adenina, preferiblemente de 20 a 90, de 25 a 85, de 35 a 80 o de 45 a 75, preferiblemente de 60 a 70, más preferiblemente aproximadamente 64 nucleótidos adenina. En esa realización, la segunda secuencia poli(A) comprende preferiblemente al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000, más preferiblemente 300 nucleótidos adenina, más preferiblemente al menos 350 nucleótidos adenina, incluso más preferiblemente al menos 380 nucleótidos adenina, preferiblemente de 320 a 430, de 330 a 420, de 340 a 410, de 350 a 400, de 360 a 400, o de 370 a 390 nucleótidos adenina. Allí, la segunda secuencia poli(A) se encuentra preferiblemente a 3' de la primera secuencia poli(A), más preferiblemente en el extremo 3' de la 3'-UTR de la molécula de ácido nucleico artificial, donde la primera y la segunda secuencias poli(A) preferiblemente están separadas, más preferiblemente separadas como se define aquí.

En una realización preferida adicional, la molécula de ácido nucleico artificial comprende una 3'-UTR que comprende al menos dos secuencias poli(A), donde la primera secuencia poli(A) está ubicada en el extremo 5'

de la 3'-UTR o en una posición dentro de la 3'-UTR, es decir, flanqueada en el lado 5' y en el lado 3' por otros elementos UTR, y comprende al menos 20, 30, 40, 50, 60, 70, 80 o 90 nucleótidos adenina, preferiblemente de 20 a 90, de 25 a 85, de 35 a 80 o de 45 a 75, preferiblemente de 60 a 70, más preferiblemente aproximadamente 64, nucleótidos adenina. En esa realización, la segunda secuencia poli(A) comprende preferiblemente al menos 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 350, 400, 500, 600, 700, 800, 900 o 1.000, más preferiblemente al menos 900 nucleótidos adenina, más preferiblemente al menos 900, al menos 950 o al menos 1.000 nucleótidos adenina. Preferiblemente, el número de nucleótidos de adenina comprendido en la segunda secuencia poli(A) es de 900 a 1.100 o de 1.000 a 1.100 nucleótidos adenina. En una realización particularmente preferente, la segunda secuencia poli(A) comprende preferiblemente aproximadamente 1.000 nucleótidos de adenina. Allí, la segunda secuencia poli(A) se ubica preferiblemente 3' a la primera secuencia poli(A), más preferiblemente en el extremo 3' de la 3'-UTR de la molécula de ácido nucleico artificial, donde la primera y la segunda secuencias poli(A) preferiblemente están separadas, más preferiblemente separadas como se define aquí.

Una molécula de ácido nucleico artificial como se describe aquí, tal como una molécula de ADN que comprende un ORF seguido de una 3'-UTR, puede contener al menos un tramo de nucleótidos timidina que corresponde a la al menos una secuencia poli(A) como se define aquí y que puede transcribirse en una secuencia poli(A) como se define aquí en el ARNm resultante.

Por ejemplo, la molécula de ácido nucleico artificial aquí descrita puede comprender una secuencia de ácido nucleico correspondiente a la secuencia de ADN

GTCCACCTGTCCCTCCTGGGCTGCTGGATTGTCTCGTTTTCCTGCCAAATAAACAGGATCAGCGCT
TTACAGATCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAA (SEQ ID No. 1).

La transcripción de tales secuencias puede dar como resultado moléculas de ácido nucleico artificiales que comprenden la secuencia

GUCCACCUGUCCCUCCUGGGCUGCUGGAUUGUCUCGUUUUCCUGCCAAAUAAACAGGAUCAGCGCU
UUACAGAUCUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAA (SEQ ID No. 2).

Dichas moléculas de ARN artificiales, es decir, moléculas de ácido nucleico artificiales que comprenden una secuencia de acuerdo con la SEQ ID NO 2 también pueden obtenerse *in vitro* mediante métodos comunes de síntesis química sin transcribirse necesariamente desde un progenitor de ADN.

Alternativa o adicionalmente, se pueden añadir múltiples nucleótidos adenina a una molécula de ácido nucleico artificial aquí descrita mediante cualquier otro método conocido en la técnica, por ejemplo mediante síntesis química o mediante una reacción de adenilación. Preferiblemente, los nucleótidos adenina se añaden a una molécula de ácido nucleico artificial como la aquí descrita mediante una reacción de adenilación enzimática. Por ejemplo, una molécula de ARN como se describe aquí puede poliadenilarse enzimáticamente mediante incubación con una enzima adecuada, tal como poli(A) polimerasa de *E. coli*.

En una realización, la molécula de ácido nucleico artificial aquí descrita comprende un ORF y una 3'-UTR, donde la 3'-UTR comprende una señal de poliadenilación. En el contexto de la presente descripción, la señal de poliadenilación se encuentra dentro de la 3'-UTR, en el extremo 3' de la 3'-UTR o aguas abajo del extremo 3' del elemento 3'-UTR. Preferiblemente, la señal de poliadenilación como se usa aquí está comprendida en la 3'-UTR de la molécula de ácido nucleico artificial. Aún más preferiblemente, la señal de poliadenilación se encuentra en el extremo 3' de la 3'-UTR, es decir, la 3'-UTR preferiblemente no contiene más de 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótidos localizados 3' a dicha señal de poliadenilación; más preferiblemente, la 3'-UTR no contiene otros elementos ubicados 3' a dicha señal de poliadenilación. En una realización preferente, la señal de poliadenilación se encuentra en el extremo 3' de la molécula de ácido nucleico artificial, es decir, la molécula de ácido nucleico artificial preferiblemente no contiene más de 12, 11, 10, 9, 8, 7, 6, 5 , 4, 3, 2 o 1 nucleótidos ubicados 3' de dicha secuencia poli(A).

En una realización preferente, la molécula de ácido nucleico artificial aquí descrita comprende una 3'-UTR que comprende al menos una secuencia poli(A), preferiblemente como se define aquí, y una señal de poliadenilación. Allí, la señal de poliadenilación se encuentra preferiblemente 3' de la al menos una secuencia poli(A) comprendida en la 3'-UTR, más preferiblemente ubicada 3' de la mayoría de la secuencia 3'-poli(A) en la 3'-UTR. Más preferiblemente, la señal de poliadenilación se localiza 3' de la al menos una señal poli (A), más preferiblemente 3' de la mayoría de la secuencia 3'-poli(A), y está separada de dicha señal poli(A) por una

secuencia de nucleótidos que comprende o consiste en al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 1 15, 120, 125, 130, 135, 140, 145 o 150 nucleótidos, donde la secuencia de nucleótidos preferiblemente no comprende más de 10, 9, 8, 7, 6, 5, 4, 3 o 2 nucleótidos adenina consecutivos. Preferentemente, la secuencia de nucleótidos que separa al menos una secuencia poli(A) y la señal de poliadenilación comprende de 1 a aproximadamente 200 nucleótidos, preferiblemente de 10 a 90, de 20 a 85, de 30 a 80, de 40 a 80, de 50 a 75 o de 55 a 85 nucleótidos, más preferiblemente de 55 a 80 nucleótidos, donde la secuencia de nucleótidos preferiblemente no comprende más de 10, 9, 8, 7, 6, 5, 4, 3 o 2 nucleótidos adenina consecutivos.

En una realización preferente, la molécula de ácido nucleico artificial aquí descrita comprende una 3'-UTR que comprende al menos dos secuencias poli(A), preferiblemente como se define aquí, y una señal de poliadenilación, donde la poliadenilación se encuentra 3' de las al menos dos secuencias poli(A) y donde preferiblemente cada una de las al menos dos secuencias poli(A) y la señal de poliadenilación están separadas de otra secuencia poli(A) o de la señal de poliadenilación, respectivamente, siendo las secuencias de nucleótidos de separación como se definieron anteriormente. Allí, la señal de poliadenilación se localiza preferiblemente en el extremo 3' de la 3'-UTR como se define aquí, más preferiblemente en el extremo 3' de la molécula de ácido nucleico artificial como se define aquí.

En una realización, la molécula de ácido nucleico artificial es una molécula de ADN que comprende una 3'-UTR que comprende una señal de poliadenilación, preferiblemente como se define aquí, donde la molécula de ADN además comprende opcionalmente al menos una secuencia poli(A), preferiblemente como se define aquí. Alternativamente, la molécula de ácido nucleico artificial es una molécula de ARN que comprende una 3'-UTR que comprende una señal de poliadenilación, preferiblemente como se define aquí, donde la molécula de ARN opcionalmente comprende además al menos una secuencia poli(A), preferiblemente como se define aquí.

Preferentemente, la señal de poliadenilación comprende la secuencia consenso NN(U/T)ANA, con N = A o U, preferiblemente AA(U/T)AAA o A(U/T)(U/T)AAA. Dicha secuencia consenso puede ser reconocida por la mayoría de los sistemas celulares animales y bacterianos, por ejemplo por los factores de poliadenilación, como el factor de especificidad de escisión/poliadenilación (CPSF) que coopera con CstF, PAP, PAB2, CFI y/o CFII. Preferiblemente, la señal de poliadenilación, en especial la secuencia consenso NNUANA, se localiza a menos de aproximadamente 50 nucleótidos, más preferiblemente a menos de aproximadamente 30 bases, lo más preferiblemente a menos de aproximadamente 25 bases, por ejemplo 21 bases, aguas abajo del extremo 3' del elemento 3'-UTR opcional como se define aquí. Más preferiblemente, la señal de poliadenilación, preferiblemente la secuencia consenso definida anteriormente, se localiza como se describe anteriormente.

Se describe aquí la transcripción de una molécula de ácido nucleico artificial, por ejemplo de una molécula de ADN artificial que comprende una señal de poliadenilación aguas abajo (es decir, en la dirección 3') de la 3'-UTR resultará en un ARN prematuro que contiene la señal de poliadenilación aguas abajo de su 3'-UTR.

El uso de un sistema de transcripción apropiado conducirá a la unión de una secuencia poli(A) al ARN prematuro. Por ejemplo, la molécula de ácido nucleico artificial aquí descrita puede ser una molécula de ADN que comprende una 3'-UTR como se describe aquí que comprende una señal de poliadenilación, que puede dar como resultado la poliadenilación de un ARN tras la transcripción de esta molécula de ADN. Por consiguiente, el ARN resultante puede comprender una 3'-UTR que comprende al menos una secuencia poli (A) y donde la 3'-UTR es seguida por una secuencia poli(A) adicional.

Sistemas de transcripción potenciales son sistemas de transcripción *in vitro* o sistemas de transcripción celular, etc. Así, la transcripción de una molécula de ácido nucleico artificial aquí descrita, por ejemplo la transcripción de una molécula de ácido nucleico artificial que comprende un marco de lectura abierto, una 3'-UTR como se define aquí que comprende una señal de poliadenilación, puede dar como resultado una molécula de ARNm que comprende un marco de lectura abierto, una 3'-UTR como se define aquí que comprende una secuencia poli(A) adicional.

Los inventores han descubierto sorprendentemente que la 3'-UTR que comprende al menos una secuencia poli(A) como se define aquí o una señal de poliadenilación como se define aquí da como resultado una mayor expresión de la proteína codificada por el ORF de la molécula de ácido nucleico artificial.

"Expresión de proteína aumentada" o "expresión de proteína mejorada" en el contexto de la presente descripción preferiblemente significa una expresión de proteína aumentada/mejorada en un punto temporal después del inicio de la expresión o una cantidad total aumentada/mejorada de proteína expresada en comparación con la expresión inducida por una molécula de ácido nucleico de referencia. Así, el nivel de proteína observado en un cierto punto temporal después del inicio de la expresión, por ejemplo después de la transfección, de la molécula de ácido nucleico artificial aquí descrita o después de la administración, por ejemplo

por inyección, de la molécula de ácido nucleico artificial a un tejido, por ejemplo después de la transfección, o la administración de un ARNm aquí descrito, por ejemplo 6, 12, 24, 48 o 72 horas después de la transfección o administración, respectivamente, preferiblemente es mayor que el nivel de proteína observado en el mismo punto temporal después del inicio de la expresión , por ejemplo después de la transfección o la administración, de una molécula de ácido nucleico de referencia, tal como un ARNm de referencia que comprende una 3'-UTR de referencia o que carece de una 3'-UTR. En una realización preferente, la cantidad máxima de proteína (según se determina, por ejemplo, por la actividad o masa de la proteína) expresada a partir de la molécula de ácido nucleico artificial está aumentada con respecto a la cantidad de proteína expresada a partir de un ácido nucleico de referencia que comprende una 3'-UTR de referencia o que carece de una 3'-UTR. Los niveles de expresión máximos se alcanzan preferiblemente dentro de las 48 horas, más preferiblemente dentro de las 24 horas e incluso más preferiblemente dentro de las 12 horas después de, por ejemplo, la transfección o la administración a un tejido.

Preferentemente, la producción total de proteínas a partir de una molécula de ácido nucleico artificial aquí descrita aumenta o mejora con respecto a un ácido nucleico de referencia. En particular, la producción de proteínas aumenta o mejora preferiblemente durante el período de tiempo en que la proteína es producida a partir de una molécula de ácido nucleico artificial, preferiblemente en un tejido diana o en un sistema de expresión mamífero, como en células de mamífero, por ejemplo células HeLa o HDF, en comparación con una molécula de ácido nucleico de referencia que carece de una 3'-UTR o que comprende una 3'-UTR de referencia. Según una realización preferente, la cantidad acumulativa de proteína expresada a lo largo del tiempo aumenta cuando se usa la molécula de ácido nucleico artificial aquí descrita.

La molécula de ácido nucleico artificial aquí descrita se caracteriza preferiblemente por una expresión aumentada de la proteína codificada, en comparación con una molécula de ácido nucleico respectiva que carece de al menos un elemento 3'-UTR o que comprende una 3'-UTR de referencia ("ácido nucleico de referencia"), que comprende una secuencia de ácido nucleico que se deriva de la 3'-UTR de un gen de proteína ribosómica o de una variante de la 3'-UTR de un gen de proteína ribosómica. Con el fin de evaluar la producción de proteína *in vivo* por la molécula de ácido nucleico artificial aquí descrita, la expresión de la proteína codificada se determina después de la inyección/transfección de la molécula de ácido nucleico artificial aquí descrita en células/tejidos diana y se compara con la expresión de proteína inducida por el ácido nucleico de referencia. Los métodos cuantitativos para determinar la expresión de proteínas son conocidos en la técnica (por ejemplo Western-Blot, FACS, ELISA, espectometría de masas). Particularmente útil en este contexto es la determinación de la expresión de proteínas reporter como luciferasa, proteína verde fluorescente (GFP) o fosfatasa alcalina secretada (SEAP). Por tanto, un ácido nucleico artificial aquí descrito o un ácido nucleico de referencia se introduce en el tejido o célula diana, por ejemplo por transfección o inyección. Varias horas o varios días (por ejemplo 6, 12, 24, 48 o 72 horas) después del inicio de la expresión o después de la introducción de la molécula de ácido nucleico, se recoge una muestra de células diana y se mide mediante FACS y/o se lisa. Posteriormente, los lisados se pueden usar para detectar la proteína expresada (y así determinar la eficiencia de la expresión de la proteína) usando varios métodos, por ejemplo Western-Blot, FACS, ELISA, espectrometría de masas o por medición de fluorescencia o luminiscencia.

Así, cuando se compara la expresión de proteína a partir de una molécula de ácido nucleico artificial aquí descrita con la expresión de proteína de una molécula de ácido nucleico de referencia en un punto de tiempo específico (por ejemplo 6, 12, 24, 48 o 72 horas después del inicio de la expresión o después de la introducción de la molécula de ácido nucleico), ambas moléculas de ácido nucleico se introducen por separado en el tejido/células diana, se recoge una muestra del tejido/células después de un momento determinado, los lisados de proteínas se preparan de acuerdo con el protocolo particular ajustado al método de detección particular (por ejemplo Western Blot, ELISA, etc., como se conoce en la técnica) y la proteína se detecta mediante el método de detección elegido. Como alternativa a la medida de la cantidad de proteína expresada en los lisados celulares o, además de la medida de la cantidad de proteína en los lisados celulares antes de la lisis de las células recolectadas o usando una alícuota en paralelo, la cantidad de proteína también se puede determinar usando análisis FACS.

Si se va a medir la cantidad total de proteína durante un período de tiempo específico, se puede recolectar tejido o células después de varios puntos de tiempo después de la introducción de la molécula de ácido nucleico artificial (por ejemplo 6, 12, 24, 48 y 72 horas después del inicio de la expresión o después de la introducción de la molécula de ácido nucleico; generalmente de diferentes animales de ensayo), y la cantidad de proteína por punto de tiempo se puede determinar tal como se explicó anteriormente. Para calcular la cantidad de proteína acumulada, se puede usar un método matemático para determinar la cantidad total de proteína, por ejemplo el área bajo la curva (AUC) se puede determinar de acuerdo con la siguiente fórmula:

$$\text{AUC} = \int_a^b f(x)\, d(x)$$

Para calcular el área bajo la curva para la cantidad total de proteína, se calcula la integral de la ecuación de la curva de expresión de cada punto final (a y b).

Así, "producción de proteína total" preferentemente se refiere al área bajo la curva (AUC) que representa la producción de proteína a lo largo del tiempo.

Preferiblemente, la molécula de ácido nucleico aquí descrita comprende una 3'-UTR que comprende al menos una secuencia poli(A) como se define aquí, que opcionalmente comprende además al menos un elemento 3'-UTR adicional que es distinto de una secuencia poli(A). Allí, el al menos otro elemento 3'-UTR preferentemente aumenta la expresión de proteínas a partir de dicha molécula de ácido nucleico artificial.

En este contexto, el término "elemento 3'-UTR" se refiere típicamente a un elemento opcional de la 3'-UTR de la molécula de ácido nucleico artificial como se describe aquí donde el elemento 3'-UTR es una secuencia de ácido nucleico distinta a una secuencia poli(A), es decir, no es una secuencia poli(A). En particular, el término se refiere a una secuencia de ácido nucleico que comprende o consiste en una secuencia de ácido nucleico que se deriva de una 3'-UTR o de una variante de una 3'-UTR. Un "elemento 3'-UTR" se refiere preferiblemente a una secuencia de ácido nucleico que representa una 3'-UTR, o parte de la misma, de una secuencia de ácido nucleico artificial, tal como un ARNm artificial, o que codifica una 3'-UTR de una molécula de ácido nucleico artificial. Por consiguiente, en el sentido de la presente descripción, preferiblemente un elemento 3'-UTR puede ser la 3'-UTR de un ARNm, preferiblemente de un ARNm artificial, o puede ser la plantilla de transcripción para una 3'-UTR de un ARNm. Así, un elemento 3'-UTR preferiblemente es una secuencia de ácido nucleico que corresponde a la 3'-UTR de un ARNm, preferiblemente a la 3'-UTR de un ARNm artificial, tal como un ARNm obtenido por transcripción de un vector constructo genéticamente modificado. Preferiblemente, un elemento 3'-UTR en el sentido de la presente descripción funciona como una 3'-UTR o codifica una secuencia de nucleótidos que cumple la función de una 3'-UTR.

Preferiblemente, el al menos un marco de lectura abierto y la 3'-UTR son heterólogos. El término "heterólogo" en este contexto significa que el marco de lectura abierto y el elemento 3'-UTR o la 3'-UTR no ocurren naturalmente (en la naturaleza) en esta combinación. Preferiblemente, el elemento 3'-UTR o la 3'-UTR se deriva de un gen diferente al del marco de lectura abierto. Por ejemplo, el ORF puede derivarse de un gen diferente al de la 3'-UTR, por ejemplo que codifica una proteína diferente o la misma proteína pero de una especie diferente, etc. Preferiblemente, el marco de lectura abierto no codifica para una proteína ribosómica. En realizaciones específicas, se prefiere que el marco de lectura abierto no codifique para una proteína reporter, por ejemplo seleccionada del grupo consistente en proteínas globina (particularmente beta-globina), proteína luciferasa, proteínas GFP o variantes de las mismas, por ejemplo variantes que tienen al menos un 70% de identidad de secuencia con una proteína globina, una proteína luciferasa o una proteína GFP.

Preferiblemente, la al menos una 3'-UTR está funcionalmente unida al ORF. Preferentemente, esto significa que la 3'-UTR está asociada al ORF de modo que puede ejercer una función, tal como una función de aumento, mejora o estabilización de la expresión del péptido o proteína codificado o una función estabilizadora de la molécula de ácido nucleico artificial. Preferiblemente, el ORF y la 3'-UTR están asociados en la dirección 5' → 3'. Por tanto, preferiblemente, la molécula de ácido nucleico artificial comprende la estructura 5'-ORF-enlazador (opcional)-3'-UTR-3', donde el enlazador puede estar presente o ausente. Por ejemplo, el enlazador puede ser uno o más nucleótidos, como un tramo de 1 a 50 o 1 a 20 nucleótidos, por ejemplo comprendiendo o consistiendo en uno o más sitios de reconocimiento de enzimas de restricción (sitios de restricción).

En una realización, la 3'-UTR de la molécula de ácido nucleico artificial aquí descrita comprende opcionalmente un elemento 3'-UTR que puede derivarse de un gen que está relacionado con un ARNm con una vida media mejorada (que proporciona un ARNm estable), por ejemplo un elemento 3'-UTR como se define y describe a continuación.

En una realización particularmente preferente, el elemento 3'-UTR opcional comprende o consiste en una secuencia de ácido nucleico que se deriva de una 3'-UTR de un gen seleccionado del grupo consistente en un gen de albúmina, un gen de α-globina, un gen de β-globina, un gen de proteína ribosómica, un gen de tirosina-hidroxilasa, un gen de lipoxigenasa y un gen de colágeno alfa, tal como un gen de colágeno alfa 1 (I), o de una variante de una 3'-UTR de un gen seleccionado del grupo consistente en un gen de albúmina, un gen de α-globina, un gen de β-globina, un gen de proteína ribosómica, un gen de tirosina-hidroxilasa, un gen de lipoxigenasa y un gen de colágeno alfa, tal como un gen de colágeno alfa 1 (I) de acuerdo con la SEQ ID NO

1369-1390 de la solicitud de patente WO2013/143700. En una realización particularmente preferente, el elemento 3'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 3'-UTR de un gen de albúmina, preferiblemente de un gen de albúmina de vertebrado, más preferiblemente de un gen de albúmina de mamífero, con total preferencia de un gen de albúmina humana de acuerdo con la SEQ ID No. 3.

3'-UTR de albúmina humana SEQ ID NO 3:

CATCACATTT AAAAGCATCT CAGCCTACCA TGAGAATAAG AGAAAGAAAA TGAAGATCAA
AAGCTTATTC ATCTGTTTTT CTTTTTCGTT GGTGTAAAGC CAACACCCTG TCTAAAAAAC
ATAAATTTCT TTAATCATTT TGCCTCTTTT CTCTGTGCTT CAATTAATAA AAAATGGAAA
GAATCT

(correspondiente a la SEQ ID NO 1369 de la solicitud de patente WO2013/143700).

En este contexto, es particularmente preferente que la 3'-UTR de la molécula de ácido nucleico artificial aquñi descrita comprenda opcionalmente un elemento 3'-UTR que comprende una secuencia de ARN correspondiente derivada de los ácidos nucleicos de acuerdo con las SEQ ID NO 1369-1390 de la solicitud de patente WO2013/143700 o un fragmento, homólogo o variante del mismo.

Con mayor preferencia, el elemento 3'-UTR opcional comprende la secuencia de ácido nucleico derivada de un fragmento del gen de albúmina humana de acuerdo con la SEQ ID NO 4:

3'-UTR de albúmina7

CATCACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAATAGCTT
ATTCATCTCTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTT
AATCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAACCT

(SEQ ID NO 4 correspondiente a la SEQ ID NO 1376 de la solicitud de patente WO2013/143700)

En este contexto, se prefiere particularmente que el elemento 3'-UTR opcional de la molécula de ácido nucleico artificial aquí descrita comprenda o consista en una secuencia de ARN correspondiente a la secuencia de ácido nucleico de acuerdo con la SEQ ID NO 4.

En otra realización particularmente preferente, el elemento 3'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 3'-UTR de un gen de α-globina, preferiblemente un gen de α-globina o β-globina de vertebrado, más preferiblemente de un gen de α-globina o β-globina de mamífero, en especial de un gen de α-globina o β-globina humana de acuerdo con las SEQ ID NO 5-7:

3'-UTR de hemoglobina alfa 1 (HBA1) de Homo sapiens

GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTG
CACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGC

(SEQ ID No: 5 correspondiente a SEQ ID NO 1370 de la solicitud de patente WO2013/143700)

3'-UTR de hemoglobina alfa 2 (HBA2) de Homo sapiens

GCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTCCCAACGGGCCCTCCTCCCCTCCTTG
CACCGGCCCTTCCTGGTCTTTGAATAAAGTCTGAGTGGGCAG

(SEQ ID NO 6 correspondiente a SEQ ID NO 1371 de la solicitud de patente WO2013/143700)

3'-UTR de beta hemoglobina (HBB) de Homo sapiens

GCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACT
GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGC

(SEQ ID NO 7 correspondiente a la SEQ ID NO 1372 de la solicitud de patente WO2013/143700).

Por ejemplo, el elemento opcional 3'UTR puede comprender o consistir en la parte central que se une al complejo ct de la 3'UTR de un gen de α-globina, tal como un gen de α -globina humana, preferiblemente de acuerdo con la SEQ ID NO 8:

Centro, parte de unión al complejo α de la 3'-UTR de un gen de α-globina (también denominado aquí "muag")

GCCCGATGGGCCTCCCAACGGGCCCTCCTCCCCTCCTTGCACCG
(SEQ ID NO 8 correspondiente a la SEQ ID NO 1393 de la solicitud de patente WO2013/143700).

En este contexto, es particularmente preferente que el elemento 3'-UTR del ARNm aquí descrito comprenda o consista en una secuencia de ARN correspondiente de la secuencia de ácido nucleico de acuerdo con la SEQ ID NO 8 o un homólogo, fragmento o variante de la misma.

En una realización preferente, el al menos un elemento 3'-UTR (opcional) comprende una secuencia de ácido nucleico que se deriva de la 3'-UTR de un gen de proteína ribosómica eucariota, preferiblemente de la 3'-UTR de un gen de proteína ribosómica de vertebrado, más preferiblemente de la 3'-UTR de un gen de proteína ribosómica de mamífero, incluso más preferiblemente de la 3'-UTR de un gen de proteína ribosómica de primate, en particular de un gen de proteína ribosómica humana.

En una realización preferente, el elemento 3'-UTR opcional comprende o corresponde a una secuencia de ácido nucleico que se deriva de la secuencia 3'-UTR de un transcripto seleccionado del grupo consistente en NM_000661.4, NM_001024921.2, NM_000967.3, NM_001033853.1, NM_000968.3, NM_000969.3, NM_001024662.1, NM_000970.3, NM_000971.3, NM_000972.2, NM_000975.3, NM_001199802.1, NM_000976.3, NM_000977.3, NM_033251.2, NM_001243130.1, NM_001243131, NM_000978.3, NM_000979.3, NM_001270490.1, NM_000980.3, NM_000981.3, NM_000982.3, NM_000983.3, NM_000984.5, NM_000985.4, NM_001035006.2, NM_001199340.1, NM_001199341.1, NM_001199342.1, NM_001199343.1, NM_001199344.1, NM_001199345.1, NM_000986.3, NM_000987.3, NM_000988.3, NM_000989.3, NM_000990.4, NM_001136134.1, NM_000991.4, NM_001136135.1, NM_001136136.1, NM_001136137.1, NM_000992.2, NM_000993.4, NM_001098577.2, NM_001099693.1, NM_000994.3, NM_001007073.1, NM_001007074.1, NM_000996.2, NM_000997.4, NM_000998.4, NM_000999.3, NM_001035258.1, NM_001000.3, NM_001002.3, NM_053275.3, NM_001003.2, NM_213725.1, NM_001004.3 , NM_001005.4, NM_001256802.1, NM_001260506.1, NM_001260507.1 , NM_001006.4, NM_001267699.1, NM_001007.4, NM_001008.3, NM_001009.3, NM_001010.2, NM_001011.3, NM_001012.1, NM_001013.3, NM_001203245.2, NM_001014.4, NM_001204091.1, NM_001015.4, NM_001016.3, NM_001017.2, NM_001018.3, NM_001030009.1, NM_001019.4, NM_001020.4, NM_001022.3, NM_001146227.1, NM_001023.3, NM_001024.3, NM_001025.4, NM_001028.2, NM_001029.3, NM_001030.4, NM_002954, NM_001135592.2, NM_001177413.1, NM_001031.4, NM_001032.4, NM_001030001.2, NM_002948.3, NM_001253379.1, NM_001253380.1, NM_001253382.1, NM_001253383.1, NM_001253384.1, NM_002952.3, NM_001034996.2, NM_001025071.1, NM_001025070.1, NM_005617.3, NM_006013.3, NM_001256577.1, NM_001256580.1, NM_007104.4, NM_007209.3, NM_012423.3, NM_001270491.1, NM_033643.2, NM_015414.3, NM_021029.5, NM_001199972.1, NM_021104.1, NM_022551.2, NM_033022.3, NM_001142284.1, NM_001026.4, NM_001142285.1, NM_001142283.1, NM_001142282.1, NM_000973.3, NM_033301.1, NM_000995.3, NM_033625.2, NM_001021.3, NM_002295.4, NM_001012321.1, NM_001033930.1, NM_003333.3, NM_001997.4, NM_001099645.1, NM_001021.3, NM_052969.1, NM_080746.2, NM_001001.4 , NM_005061.2 , NM_015920.3 , NM_016093.2 , NM_198486.2 , NG_011172.1, NG_011253.1, NG_000952.4, NR_002309.1, NG_010827.2, NG_009952.2, NG_009517.1

En una realización preferente, el elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita se deriva de la región 3'-UTR de un gen que codifica una proteína ribosómica, preferiblemente de la región 3'-UTR de la proteína ribosómica L9 (RPL9), proteína ribosómica L3 (RPL3), proteína ribosómica L4 (RPL4), proteína ribosómica L5 (RPL5), proteína ribosómica L6 (RPL6), proteína ribosómica L7 (RPL7), proteína ribosómica L7a (RPL7A), proteína ribosómica L1 1 (RPL1 1), proteína ribosómica L12 (RPL12), proteína ribosómica L13 (RPL13), proteína ribosómica L23 (RPL23), proteína ribosómica L18 (RPL18), proteína ribosómica L18a (RPL18A), proteína ribosómica L19 (RPL19), proteína ribosómica L21 (RPL21), proteína ribosómica L22 (RPL22), proteína ribosómica L23a (RPL23A), proteína ribosómica L17 (RPL17), proteína ribosómica L24 (RPL24), proteína ribosómica L26 (RPL26), proteína ribosómica L27 (RPL27), proteína ribosómica L30 (RPL30), proteína ribosómica L27a (RPL27A), proteína ribosómica L28 (RPL28), proteína ribosómica L29 (RPL 29), proteína ribosómica L31 (RPL31), proteína ribosómica L32 (RPL32), proteína ribosómica L35a (RPL35A), proteína ribosómica L37 (RPL37), proteína ribosómica L37a (RPL37A), proteína ribosómica L38 (RPL38), proteína ribosómica L39 (RPL39), proteína ribosómica grande P0 (RPLP0), proteína ribosómica grande P1 (RPLP1), proteína ribosómica grande P2 (RPLP2), proteína ribosómica S3 (RPS3), proteína ribosómica S3A (RPS3A), proteína ribosómica S4, X-ligado (RPS4X), proteína ribosómica S4, Y-ligado 1 (RPS4Y1), proteína ribosómica S5 (RPS5), proteína ribosómica S6 (RPS6), proteína ribosómica S7 (RPS7), proteína ribosómica S8 (RPS8), proteína ribosómica S9 (RPS9), proteína ribosómica S10 (RPS10), proteína ribosómica S1 1 (RPS1 1), proteína ribosómica S12 (RPS12), proteína ribosómica S13 (RPS13), proteína ribosómica S15 (RPS15), proteína ribosómica S15a (RPS15A), proteína ribosómica S16 (RPS16), proteína ribosómica S19 (RPS19), proteína ribosómica S20 (RPS20), proteína ribosómica S21 (RPS21), proteína ribosómica S23 (RPS23), proteína ribosómica S25 (RPS25), proteína ribosómica S26 (RPS26), proteína ribosómica S27 (RPS27), proteína ribosómica S27a (RPS27a), proteína ribosómica S28 (RPS28), proteína ribosómica S29 (RPS29),

proteína ribosómica L15 (RPL15), proteína ribosómica S2 (RPS2), proteína ribosómica L14 (RPL14), proteína ribosómica S14 (RPS14), proteína ribosómica L10 (RPL10), proteína ribosómica L10a (RPL10A), proteína ribosómica L35 (RPL35), proteína ribosómica L13a (RPL13A), proteína ribosómica L36 (RPL36), proteína ribosómica L36a (RPL36A), proteína ribosómica L41 (RPL41), proteína ribosómica S18 (RPS18), proteína ribosómica S24 (RPS24), proteína ribosómica L8 (RPL8), proteína ribosómica L34 (RPL34), proteína ribosómica S17 (RPS17), proteína ribosómica SA (RPSA) o proteína ribosómica S17 (RPS17). En una realización alternativa, el elemento 3'-UTR puede derivarse de un gen que codifica una proteína ribosómica o de un gen seleccionado del producto 1 de fusión de proteína ribosómica residuo de ubiquitina A-52 (UBA52), virus del sarcoma murino de Finkel-Biskis-Reilly (FBR-MuSV) expresado de forma ubicua (FAU), proteína ribosómica L22 tipo 1 (RPL22L1), proteína ribosómica tipo L39 (RPL39L), proteína ribosómica tipo L10 (RPL10L), proteína ribosómica tipo L36a (RPL36AL), proteína ribosómica tipo L3 (RPL3L), proteína ribosómica tipo S27 (RPS27L), proteína ribosómica L26 tipo 1 (RPL26L1), proteína ribosómica L7 tipo 1 (RPL7L1), pseudogen de proteína ribosómica L13a (RPL13AP), proteína ribosómica L37a pseudogen 8 (RPL37AP8), pseudogen 5 de la proteína ribosómica S10 (RPS10P5), pseudogen 11 de proteína ribosómica S26 (RPS26P11), pseudogen 5 de proteína ribosómica L39 (RPL39P5), pseudogen P0 de proteína ribosómica grande 6 (RPLP0P6) y pseudogen 14 de proteína ribosómica L36 (RPL36P14).

Preferentemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita comprende o consiste en una secuencia de ácido nucleico que tiene una identidad de al menos aproximadamente un 1, 2, 3, 4, 5, 10, 15, 20, 30 o 40%, preferiblemente de al menos aproximadamente un 50%, preferiblemente de al menos aproximadamente un 60%, preferiblemente de al menos aproximadamente un 70%, más preferiblemente de al menos aproximadamente un 80%, más preferiblemente de al menos aproximadamente un 90%, incluso más preferiblemente de al menos aproximadamente un 95%, incluso más preferiblemente de al menos aproximadamente un 99%, con total preferencia de un 100% con la secuencia de ácido nucleico de una 3'-UTR de un gen de proteína ribosómica, tal como las secuencias de ácido nucleico de acuerdo con las SEQ ID NO: 10 a 115 como se definen en la solicitud de patente internacional PCT/EP2013/003946, o la secuencia de ARN correspondiente.

En una realización preferente, el al menos un elemento 3'-UTR (opcional) comprende o consiste en una secuencia de ácido nucleico que tiene una identidad de al menos aproximadamente un 40%, preferiblemente de al menos aproximadamente un 50%, preferiblemente de al menos aproximadamente un 60%, preferiblemente de al menos aproximadamente un 70%, más preferiblemente de al menos aproximadamente un 80%, más preferiblemente de al menos aproximadamente un 90%, incluso más preferiblemente de al menos aproximadamente un 95%, incluso más preferiblemente de al menos aproximadamente un 99%, con total preferencia del 100% con la secuencia 3'-UTR de la proteína ribosómica Small 9 (RPS9). Con particular preferencia, el al menos un elemento 3'-UTR (opcional) comprende o consiste en una secuencia de ácido nucleico que tiene una identidad de al menos aproximadamente un 40%, preferiblemente de al menos aproximadamente un 50%, preferiblemente de al menos aproximadamente un 60%, preferiblemente de al menos aproximadamente un 70%, más preferiblemente de al menos aproximadamente un 80%, más preferiblemente de al menos aproximadamente un 90%, incluso más preferiblemente de al menos aproximadamente un 95%, incluso más preferiblemente de al menos aproximadamente un 99%, con total preferencia preferiblemente de un 100% con la SEQ ID NO: 9 o la SEQ ID NO: 10

SEQ ID NO: 9

GTCCACCTGTCCCTCCTGGGCTGCTGGATTGTCTCGTTTTCCTGCCAAATAAACAGGATCAGCGCT
TTAC

SEQ ID NO: 10

GUCCACCUGUCCCUCCUGGGCUGCUGGAUUGUCUCGUUUUCCUGCCAAAUAAACAGGAUCAGCGCU
UUAC

El término "una secuencia de ácido nucleico que se deriva de la 3'-UTR de un gen (…)" se refiere preferiblemente a una secuencia de ácido nucleico que se basa en la secuencia 3'-UTR de un gen o de una parte del mismo, tal como en la 3'-UTR de un gen de albúmina, un gen de α-globina, un gen de β-globina, un gen de proteína ribosómica, un gen de tirosina-hidroxilasa, un gen de lipoxigenasa o un gen de colágeno alfa, tal como un gen de colágeno alfa 1 (I), preferiblemente de un gen de albúmina o en parte del mismo. Este término incluye secuencias correspondientes a la secuencia 3'-UTR completa, es decir, la secuencia 3'-UTR completa de un gen, y secuencias correspondientes a un fragmento de la secuencia 3'-UTR de un gen, como un gen de albúmina, de α-globina, de β-globina, de proteína ribosómica, de tirosina-hidroxilasa, de lipoxigenasa o de colágeno alfa, tal como un gen de colágeno alfa 1 (I), preferiblemente de un gen de albúmina.

La frase "secuencia de ácido nucleico, que se deriva de la 3'-UTR de un gen (…)" preferiblemente se refiere a

una secuencia de ácido nucleico que se basa en la secuencia 3'-UTR de un gen (…), preferentemente un gen descrito anteriormente, o en un fragmento o parte del mismo. Esta frase incluye secuencias correspondientes a la secuencia 3'-UTR completa, es decir, la secuencia 3'-UTR de longitud completa de dicho gen, y secuencias correspondientes a un fragmento de la secuencia 3'-UTR de dicho gen. Preferiblemente, un fragmento de un 3-UTR de un gen (...) consiste en un tramo continuo de nucleótidos que corresponde a un tramo continuo de nucleótidos de la 3'-UTR de longitud completa de un gen que representa al menos el 5%, 10%, 20%, preferiblemente al menos el 30%, más preferiblemente al menos el 40%, más preferiblemente al menos el 50%, incluso más preferiblemente al menos el 60%, incluso más preferiblemente al menos el 70%, incluso más preferiblemente al menos el 80% y con total preferencia al menos el 90% de la 3'-UTR de longitud completa de un gen. Tal fragmento, en el sentido de la presente descripción, es preferiblemente un fragmento funcional como se describe aquí. Preferentemente, el fragmento conserva una función reguladora para la traducción del ORF unido a la 3'-UTR o a un fragmento de la misma. El término "3'-UTR de un gen (…)" preferiblemente se refiere a la 3'-UTR de un gen natural, preferiblemente como se describe aquí.

Los términos "variante de la 3'-UTR de un gen (...)" y "variante del mismo, en ese contexto, se refieren a una variante de la 3'-UTR de un gen natural. Dicha variante puede ser una 3'-UTR modificada de dicho gen. Por ejemplo, una variante 3'-UTR puede presentar una o más deleciones, inserciones, adiciones y/o sustituciones de nucleótidos en comparación con la 3'-UTR natural de la cual se deriva la variante. Preferiblemente, una variante de una 3'-UTR de un gen como se usa aquí es al menos un 40%, preferiblemente al menos un 50%, más preferiblemente al menos un 60%, más preferiblemente al menos un 70%, incluso más preferiblemente al menos un 80%, incluso más preferiblemente al menos un 90%, con total preferencia al menos un 95% idéntica a la 3'-UTR de origen natural de la que se deriva la variante. Preferiblemente, la variante es una variante funcional como se describe aquí.

Los términos "variante funcional", "fragmento funcional" y "fragmento funcional de una variante" (también denominado "fragmento variante funcional") en el contexto de la presente descripción significan que el fragmento de la 3'-UTR de un gen, la variante de la 3'-UTR de un gen o el fragmento de una variante de la 3'-UTR de un gen cumple al menos una función, preferiblemente más de una, de la 3'-UTR natural del respectivo gen del cual se deriva la variante, el fragmento o el fragmento de la variante. Dicha función puede ser, por ejemplo, estabilizar ARNm y/o potenciar, estabilizar y/o prolongar la producción de proteínas a partir de un ARNm y/o aumentar la expresión de proteínas o la producción total de proteínas a partir de un ARNm, preferiblemente en una célula de mamífero, tal como en una célula humana. Preferiblemente, la función de la 3'-UTR de un gen como se describe aquí se refiere a la traducción de la proteína codificada por el ORF. Más preferiblemente, la función comprende mejorar la eficiencia de traducción del ORF unido a la 3'-UTR o a un fragmento o variante de la misma. Es particularmente preferente que la variante, el fragmento y el fragmento variante en el contexto de la presente descripción cumplan la función de estabilizar un ARNm, preferiblemente en una célula de mamífero, tal como una célula humana, en comparación con un ARNm que comprende una 3'-UTR de referencia o que carece de una 3'-UTR, y/o la función de potenciar, estabilizar y/o prolongar la producción de proteínas a partir de un ARNm, preferiblemente en una célula de mamífero, tal como en una célula humana, en comparación con un ARNm que comprende una 3'-UTR de referencia o que carece de una 3'-UTR, y/o la función de aumentar la producción de proteínas a partir de un ARNm, preferiblemente en una célula de mamífero, tal como en una célula humana, en comparación con un ARNm que comprende una 3'-UTR de referencia o sin una 3'-UTR. Una 3'-UTR de referencia puede ser, por ejemplo, una 3'-UTR que se produce naturalmente en combinación con el ORF. Además, una variante funcional, un fragmento funcional o un fragmento de variante funcional de una 3'-UTR de un gen como se describe aquí preferiblemente no tiene un efecto esencialmente decreciente sobre la eficiencia de la traducción del ARNm que comprende dicha variante, fragmento o fragmento variante de una 3'-UTR en comparación con la 3'-UTR de tipo salvaje del cual se deriva la variante, el fragmento o el fragmento variante. Una función particularmente preferente de un "fragmento funcional", una "variante funcional" o un "fragmento funcional de una variante" de la 3'-UTR de un gen en el contexto de la presente descripción es el aumento, mejora, estabilización y/o prolongación de la producción de proteínas mediante la expresión de un ARNm que porta el fragmento funcional, la variante funcional o el fragmento funcional de una variante como se describió anteriormente.

Preferentemente, la eficiencia de la una o más funciones ejercidas por la variante funcional, el fragmento funcional o el fragmento de variante funcional, tal como el ARNm, y/o la eficiencia estabilizadora de la producción de proteínas y/o la eficiencia creciente de la producción de proteínas aumenta al menos en al menos un 5%, más preferiblemente al menos un 10%, más preferiblemente al menos un 20%, más preferiblemente al menos un 30%, más preferiblemente al menos un 40%, más preferiblemente al menos un 50%, más preferiblemente al menos un 60%, incluso más preferiblemente al menos un 70%, incluso más preferiblemente al menos un 80%, con total preferencia al menos un 90% con respecto a la eficiencia estabilizadora de la producción de ARNm y/o proteína y/o la eficiencia creciente de producción de proteína exhibida por la 3'-UTR del gen de origen natural respectivo del cual se deriva la variante, el fragmento o el fragmento variante.

En el contexto de la presente descripción, un fragmento de la 3'-UTR de un gen como se describe aquí o de una variante de la 3'-UTR de dicho gen preferentemente tiene una longitud de al menos aproximadamente 3 nucleótidos, preferiblemente de al menos aproximadamente 5 nucleótidos, más preferiblemente de al menos aproximadamente 10, 15, 20, 25 o 30 nucleótidos, incluso más preferiblemente de al menos aproximadamente 50 nucleótidos, con total preferencia de al menos aproximadamente 70 nucleótidos. Preferiblemente, dicho fragmento de la 3'-UTR de un gen o de una variante de la 3'-UTR de un gen es un fragmento funcional como se describió anteriormente. En una realización preferente, la 3'-UTR de un gen o un fragmento o variante del mismo tiene una longitud de entre 3 y aproximadamente 500 nucleótidos, preferiblemente de entre 5 y aproximadamente 150 nucleótidos, más preferiblemente de entre 10 y 100 nucleótidos, incluso más preferiblemente entre 15 y 90, con total preferencia entre 20 y 70.

Preferentemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita comprende o consiste en un "fragmento funcional", una "variante funcional" o un "fragmento funcional de una variante" de la 3'-UTR de un gen como se describe aquí.

En una realización preferente, el al menos otro elemento 3'-UTR que opcionalmente puede estar comprendido en la 3'-UTR de la molécula de ácido nucleico artificial y que es distinto de una secuencia poli(A) aumenta la estabilidad de la molécula de ácido nucleico artificial, por ejemplo aumenta la estabilidad de un ARNm tal como se describe aquí, en comparación con un ácido nucleico respectivo (ácido nucleico de referencia) que carece de una 3'-UTR o que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR opcional, tal como una 3'-UTR que ocurre naturalmente en combinación con el ORF. Preferentemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita aumenta la estabilidad de la producción de proteínas a partir de la molécula de ácido nucleico artificial aquí descrita, por ejemplo de un ARNm aquí descrito, en comparación con un ácido nucleico respectivo que carece de una 3'-UTR o que comprende una 3'-UTR de referencia, tal como una 3'-UTR que se produce naturalmente en combinación con el ORF. Preferentemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita prolonga la producción de proteínas a partir de la molécula de ácido nucleico artificial aquí descrita, por ejemplo de un ARNm aquí descrito, en comparación con un ácido nucleico respectivo que carece de una 3'-UTR o que comprende una 3'-UTR de referencia, tal como una 3'-UTR que se produce naturalmente en combinación con el ORF. Preferiblemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita aumenta la expresión de proteínas y/o la producción de proteína total a partir de la molécula de ácido nucleico artificial aquí descrita, por ejemplo de un ARNm aquí descrito, en comparación con un ácido nucleico respectivo que carece de una 3'-UTR o que comprende una 3'-UTR de referencia, tal como una 3'-UTR que se produce naturalmente en combinación con el ORF. Preferiblemente, el al menos un elemento 3'-UTR de la molécula de ácido nucleico artificial aquí descrita no influye negativamente en la eficiencia de traducción de un ácido nucleico en comparación con la eficiencia de traducción de un ácido nucleico respectivo que carece de una 3'-UTR o que comprende una 3'-UTR de referencia, tal como una 3'-UTR que ocurre naturalmente en combinación con el ORF. Incluso más preferiblemente, la eficiencia de traducción es mejorada gracias a la 3'-UTR en comparación con la eficiencia de traducción de la proteína codificada por el ORF respectivo en su contexto natural.

En este contexto, el término "molécula de ácido nucleico respectiva" o "molécula de ácido nucleico de referencia" significa que, aparte de las diferentes 3'-UTR, la molécula de ácido nucleico de referencia es comparable, preferiblemente idéntica, a la molécula de ácido nucleico artificial aquí descrita que comprende el elemento 3'-UTR. En particular, una molécula de ácido nucleico de referencia puede comprender una secuencia de nucleótidos y elementos, tales como ORF y 3'-UTR, que difiere de la molécula de ácido nucleico artificial aquí descrita solo en al menos un elemento 3'-UTR opcional que es distinto de una secuencia poli(A).

El término "estabilizar y/o prolongar la producción de proteínas" a partir de una molécula de ácido nucleico artificial, tal como un ARNm artificial, significa preferiblemente que la producción de proteínas a partir de la molécula de ácido nucleico artificial, tal como el ARNm artificial, se estabiliza y/o prolonga en comparación con la producción de proteínas a partir de una molécula de ácido nucleico de referencia, tal como un ARNm de referencia, por ejemplo que comprende una 3'-UTR de referencia o que carece del elemento 3'-UTR o que carece por completo de una 3'-UTR, preferiblemente en un sistema de expresión de mamífero, tal como en células HeLa o HDF. Por tanto, la proteína producida a partir de la molécula de ácido nucleico artificial, como el ARNm artificial, es observable durante un período de tiempo más largo que lo que se puede observar para una proteína producida a partir de una molécula de ácido nucleico de referencia. En otras palabras, la cantidad de proteína producida a partir de la molécula de ácido nucleico artificial, como el ARNm artificial, medida en el tiempo, sobrepasa un valor umbral en un punto de tiempo posterior que la cantidad de proteína producida a partir de una molécula de ácido nucleico de referencia, como un ARNm de referencia, medido con el tiempo. Tal valor umbral puede ser, por ejemplo, la cantidad de proteína medida en la fase inicial de la expresión, como 1, 2, 3, 4, 5 o 6 horas después del inicio de la expresión, tal como la transfección posterior de la molécula de ácido nucleico.

Por ejemplo, la producción de proteínas a partir de la molécula de ácido nucleico artificial, tal como un ARNm artificial, en una cantidad que es al menos la cantidad observada en la fase inicial de expresión, como 1, 2, 3, 4, 5 o 6 horas después del inicio de la expresión, como la transfección posterior de la molécula de ácido nucleico, se prolonga al menos aproximadamente 5 horas, preferiblemente al menos aproximadamente 10 horas, más preferiblemente al menos aproximadamente 24 horas en comparación con la producción de proteínas a partir de una molécula de ácido nucleico de referencia, tal como un ARNm de referencia, en un sistema de expresión de mamífero, como en células de mamífero, por ejemplo en células HeLa o HDF. Así, la molécula de ácido nucleico artificial aquí descrita preferentemente permite la producción prolongada de proteínas en una cantidad que es al menos la cantidad observada en la fase inicial de expresión, tal como 1, 2, 3, 4, 5 o 6 horas después del inicio de la expresión, como la transfección posterior, al menos aproximadamente 5 horas, preferiblemente al menos aproximadamente 10 horas, más preferiblemente al menos aproximadamente 24 horas en comparación con una molécula de ácido nucleico de referencia que carece de 3'-UTR o que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR.

En realizaciones preferentes, el período de producción de proteínas a partir de la molécula de ácido nucleico artificial aquí descrita se extiende al menos 1,5 veces, preferiblemente al menos 2 veces, más preferiblemente al menos 2,5 veces en comparación con la producción de proteínas a partir de una molécula de ácido nucleico de referencia que carece de una 3'-UTR o que comprende una 3'-UTR de referencia sin el elemento 3'-UTR.

Este efecto de prolongar la producción de proteínas puede determinarse (i) midiendo la cantidad de proteínas, por ejemplo obtenidas mediante la expresión de una proteína reporter codificada, tal como luciferasa, preferiblemente en un sistema de expresión de mamífero, tal como células HeLa o HDF, a lo largo del tiempo, (ii) determinando el punto de tiempo en el cual la cantidad de proteína sobrepasa la cantidad de proteína observada, por ejemplo a 1, 2, 3, 4, 5 o 6 horas después del inicio de la expresión, por ejemplo 1, 2, 3, 4, 5 o 6 horas después de la transfección de la molécula de ácido nucleico artificial, y (iii) comparar el punto de tiempo en el que la cantidad de proteína sobrepasa la cantidad de proteína observada en 1, 2, 3, 4, 5, o 6 horas después del inicio de la expresión en dicho punto de tiempo determinado para una molécula de ácido nucleico que carece de 3'-UTR o que comprende una 3'-UTR de referencia sin el elemento 3'-UTR.

Preferiblemente, se logra este efecto estabilizador y/o prolongador de la producción de proteínas a la vez que la cantidad total de proteína producida a partir de la molécula de ácido nucleico artificial aquí descrita, por ejemplo en un lapso de tiempo de 48 o 72 horas, es al menos la cantidad de proteína producida a partir de una molécula de ácido nucleico de referencia que carece de una 3'-UTR o que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR, tal como una 3'-UTR que ocurre naturalmente con el ORF de la molécula de ácido nucleico artificial. Por tanto, la presente invención proporciona una molécula de ácido nucleico artificial que permite la producción de proteínas prolongada y/o estabilizada en un sistema de expresión de mamífero, tal como células de mamífero, por ejemplo células HeLa o HDF, como se especifica anteriormente, donde la cantidad total de proteína producida a partir de dicha molécula de ácido nucleico artificial, por ejemplo en un lapso de tiempo de 48 o 72 horas, es al menos la cantidad total de proteína producida, por ejemplo en dicho intervalo de tiempo, a partir de una molécula de ácido nucleico de referencia que carece de 3'-UTR o que comprende una 3'-UTR de referencia, tal como una 3'-UTR que se produce naturalmente con el ORF de la molécula de ácido nucleico artificial.

Así, "expresión de proteína estabilizada" significa preferiblemente que hay una producción de proteína más uniforme a partir de la molécula de ácido nucleico artificial aquí descrita durante un período de tiempo predeterminado, tal como más de 24 horas, preferiblemente más de 48 horas, con mayor preferencia más de 72 horas, en comparación con una molécula de ácido nucleico de referencia, por ejemplo un ARNm que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR o que carece de una 3'-UTR por completo. En consecuencia, el nivel de producción de proteínas, por ejemplo en un sistema de mamífero, a partir de la molécula de ácido nucleico artificial que comprende un elemento 3'-UTR como se describe aquí, por ejemplo a partir de un ARNm aquí descrito, preferiblemente no cae en la medida observada para una molécula de ácido nucleico de referencia, tal como un ARNm de referencia como se describió anteriormente. Por ejemplo, la cantidad de proteína (codificada por el ORF) observada 6 horas después del inicio de la expresión, por ejemplo 6 horas después de la transfección de la molécula de ácido nucleico artificial aquí descrita en una célula, tal como una célula de mamífero, puede ser comparable a la cantidad de proteína observada 48 horas después del inicio de la expresión, por ejemplo 48 horas después de la transfección. Así, la proporción entre la cantidad de proteína codificada por el ORF, como una proteína reporter, por ejemplo luciferasa, observada 48 horas después del inicio de la expresión, 48 horas después de la transfección, y la cantidad de proteína observada 6 horas después del inicio de la expresión, por ejemplo 6 horas después de la transfección, preferiblemente es al menos aproximadamente 0,4, más preferiblemente al menos aproximadamente 0,5, más preferiblemente al menos aproximadamente 0,6, incluso más preferiblemente al menos aproximadamente 0,7. Preferiblemente, la relación está entre aproximadamente 0,4 y aproximadamente 4, preferiblemente entre aproximadamente 0,65 y aproximadamente 3, más preferiblemente entre

aproximadamente 0,7 y 2 para una molécula de ácido nucleico tal como la aquí descrita. Para una molécula de ácido nucleico de referencia respectiva, por ejemplo un ARNm que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR o que carece por completo de una 3'-UTR, dicha relación puede estar, por ejemplo, entre aproximadamente 0,05 y aproximadamente 0,3.

Por tanto, en una realización preferente, la presente descripción proporciona una molécula de ácido nucleico artificial que comprende un ORF y una 3'-UTR que comprende un elemento 3'-UTR opcional como se describió anteriormente donde la proporción entre la cantidad de proteína (reporter), por ejemplo la cantidad de luciferasa, observada 48 horas después del inicio de la expresión y la cantidad de proteína (reporter) observada 6 horas después del inicio de la expresión, preferiblemente en un sistema de expresión de mamífero, como en células de mamífero, por ejemplo en células HeLa, preferiblemente es superior a aproximadamente 0,4, más preferiblemente superior a aproximadamente 0,5, más preferiblemente superior a aproximadamente 0,6, incluso más preferiblemente superior a aproximadamente 0,7, por ejemplo entre aproximadamente 0,4 y aproximadamente 4, preferiblemente entre aproximadamente 0,65 y aproximadamente 3, más preferiblemente entre aproximadamente 0,7 y 2, donde preferentemente la cantidad total de proteína producida a partir de dicha molécula de ácido nucleico artificial, por ejemplo en un lapso de 48 horas, es al menos la cantidad total de proteína producida, por ejemplo dentro de dicho lapso de tiempo, a partir de una molécula de ácido nucleico de referencia que carece de 3'-UTR o que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR opcional, tal como una 3'-UTR que ocurre naturalmente con el ORF de la molécula de ácido nucleico artificial. En una realización preferente, la presente descripción proporciona una molécula de ácido nucleico artificial que comprende un ORF y una 3'-UTR que comprende un elemento 3'-UTR opcional como se describió anteriormente, donde la proporción entre la cantidad de proteína (reporter), por ejemplo la cantidad de luciferasa, observada 72 horas después del inicio de la expresión y la cantidad de proteína (reporter) observada 6 horas después del inicio de la expresión, preferiblemente en un sistema de expresión de mamífero, como en células de mamífero, por ejemplo células HeLa, es preferiblemente superior a aproximadamente 0,4, más preferiblemente superior a aproximadamente 0,5, más preferiblemente superior a aproximadamente 0,6, incluso más preferiblemente superior a aproximadamente 0,7, p. entre aproximadamente 0,4 y 1,5, preferiblemente entre aproximadamente 0,65 y aproximadamente 1,15, más preferiblemente entre aproximadamente 0,7 y 1,0, donde preferentemente la cantidad total de proteína producida a partir de dicha molécula de ácido nucleico artificial, por ejemplo en un lapso de tiempo de 72 horas, es al menos la cantidad total de proteína producida, por ejemplo dentro de dicho lapso de tiempo, a partir de una molécula de ácido nucleico de referencia que carece de 3'-UTR o que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR opcional, tal como una 3'-UTR que ocurre naturalmente con el ORF de la molécula de ácido nucleico artificial.

En una realización preferente, el al menos otro elemento 3'-UTR que opcionalmente puede estar comprendido en la 3'-UTR de la molécula de ácido nucleico artificial y que es distinto de una secuencia poli(A) aumenta o mejora la expresión de proteínas a partir de la molécula de ácido nucleico artificial como se define aquí.

Dicho aumento de la estabilidad de la molécula de ácido nucleico artificial, dicho aumento de la estabilidad de la producción de proteínas, dicha prolongación de la producción de proteínas y/o dicho aumento/mejora de la expresión de proteínas y/o la producción de proteínas totales se determina preferiblemente por comparación con una molécula de ácido nucleico de referencia respectiva que carece de una 3'- UTR, por ejemplo un ARNm que carece de una 3'-UTR, o con una molécula de ácido nucleico de referencia que comprende una 3'-UTR de referencia que carece del elemento 3'-UTR opcional, tal como una 3'-UTR que ocurre naturalmente con el ORF como se describió anteriormente.

El efecto y la eficacia de estabilización de la producción de ARNm y/o proteína y/o el efecto de aumento de la producción de proteína y la eficiencia del al menos un elemento 3'-UTR (opcional) de la molécula de ácido nucleico artificial aquí descrita puede determinarse por cualquier método adecuado para ello conocido por el experto. Por ejemplo, se pueden generar moléculas de ARNm artificiales que comprenden una secuencia de codificación/marco de lectura abierto (ORF) para una proteína reporter, como la luciferasa, y sin 3'-UTR, con una 3'-UTR que carece del elemento 3'-UTR opcional, tal como una 3'-UTR derivada de un gen natural o una 3'-UTR derivada de un gen de referencia (es decir, una 3'-UTR de referencia, como una 3'-UTR que ocurre naturalmente con el ORF). Tales ARNm pueden generarse, por ejemplo, por transcripción *in vitro* en vectores respectivos, tales como vectores plásmidos, por ejemplo comprendiendo un promotor T7 y una secuencia que codifica las secuencias de ARNm respectivas. Las moléculas de ARNm generadas pueden transfectarse en células mediante cualquier método de transfección adecuado para transfectar ARNm; por ejemplo, pueden electroporarse en células de mamífero, como células HELA, y las muestras pueden analizarse en ciertos puntos de tiempo después de la transfección, por ejemplo 6 horas, 24 horas, 48 horas y 72 horas después de la transfección. Dichas muestras pueden analizarse para cantidades de ARNm y/o cantidades de proteínas por métodos bien conocidos por el experto. Por ejemplo, la cantidad de ARNm reporter presente en las células en los puntos de tiempo de la muestra pueden determinarse mediante métodos PCR cuantitativos. La cantidad de

proteína reporter codificada por los respectivos ARNm puede determinarse, por ejemplo, mediante Western Blot, ensayos ELISA, análisis FACS o ensayos reporter, tales como ensayos de luciferasa, dependiendo de la proteína reporter utilizada. El efecto de estabilizar la expresión de proteínas y/o prolongar la expresión de proteínas puede analizarse, por ejemplo, determinando la proporción entre el nivel de proteína observado 48 horas después de la transfección y el nivel de proteína observado 6 horas después de la transfección. Cuanto más cercano sea dicho valor a 1, más estable será la expresión de la proteína en este período de tiempo. Por supuesto, tales medidas también se pueden realizar a las 72 horas o más y se puede determinar la proporción entre el nivel de proteína observado 72 horas después de la transfección y el nivel de proteína observado 6 horas después de la transfección para establecer la estabilidad de expresión de la proteína.

En algunas realizaciones, la 3'-UTR de la molécula de ácido nucleico artificial puede comprender un tallo-bucle de histona además de la al menos una secuencia poli (A) y el elemento 3'-UTR opcional como se describe aquí. Tal 3'-UTR de la molécula de ácido nucleico artificial aquí descrita puede comprender, por ejemplo, en la dirección 5'-a-3', un elemento 3'-UTR opcional como se describe aquí, al menos una poli(A), una secuencia poli(C) opcional, una secuencia tallo-bucle de histona opcional y opcionalmente otra secuencia poli(A) o una señal de poliadenilación.

En una realización preferente, la molécula de ácido nucleico artificial aquí descrita comprende al menos una secuencia tallo-bucle de histona.

Preferentemente, dichas secuencias tallo-bucle de histona se seleccionan de las secuencias tallo-bucle de histona descritas en la WO 2012/019780, como referencia.

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

fórmula (I) (secuencia de tallo-bucle sin elementos frontera de tallo):

$$\underbrace{[N_{0-2}GN_{3-5}]}_{\text{tallo 1}}\ \underbrace{[N_{0-4}(U/T)N_{0-4}]}_{\text{bucle}}\ \underbrace{[N_{3-5}CN_{0-2}]}_{\text{tallo 2}}$$

fórmula (II) (secuencia tallo-bucle con elementos frontera del tallo):

$$\underbrace{N_{1-6}}_{\text{elemento frontera de tallo 1}}\ \underbrace{[N_{0-2}GN_{3-5}]}_{\text{tallo1}}\ \underbrace{[N_{0-4}(U/T)N_{0-4}]}_{\text{bucle}}\ \underbrace{[N_{3-5}CN_{0-2}]}_{\text{tallo2}}\ \underbrace{N_{1-6}}_{\text{elemento frontera de tallo 2}}$$

donde:

elementos frontera del tallo1 o tallo2 N_{1-6}: es una secuencia consecutiva de 1 a 6, preferiblemente de 2 a 6, más preferiblemente de 2 a 5, aún más preferiblemente de 3 a 5, más preferiblemente de 4 a 5 o de 5 N, donde cada N se selecciona independientemente de un nucleótido seleccionado de entre A, U, T, G y C, o un análogo de nucleótido del mismo;

tallo1 $[N_{0-2}GN_{3-5}]$: es complementaria inversa o complementaria parcialmente inversa con el elemento tallo2 y es una secuencia consecutiva de entre 5 a 7 nucleótidos; donde N_{0-2} es una secuencia consecutiva de 0 a 2, preferiblemente de 0 a 1, más preferiblemente de 1 N, donde cada N se selecciona independientemente de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo; donde N_{3-5} es una secuencia consecutiva de 3 a 5, preferiblemente de 4 a 5, más preferiblemente de 4 N, donde cada N se selecciona independientemente de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo, y donde G es guanosina o un análogo del mismo, y se puede reemplazar opcionalmente por una citidina o un análogo de la misma, siempre que su citidina de nucleótido complementario en tallo2 se reemplace por guanosina;

secuencia bucle $[N_{0-4}(U/T)N_{0-4}]$: se ubica entre los elementos tallo1 y tallo2 y es una secuencia consecutiva de 3 a 5 nucleótidos, más preferiblemente de 4 nucleótidos; donde cada N_{0-4} es independiente de otra secuencia consecutiva de 0 a 4, preferiblemente de 1 a 3, más preferiblemente de 1 a 2 N, donde cada N se selecciona

de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo; donde U/T representa uridina, u opcionalmente timidina;

tallo2 [$N_{3-5}CN_{0-2}$]: es complementaria inversa o complementaria parcialmente inversa al elemento tallo1 y es una secuencia consecutiva de 5 a 7 nucleótidos; donde N_{3-5} es una secuencia consecutiva de 3 a 5, preferiblemente de 4 a 5, más preferiblemente de 4 N, donde cada N se selecciona independientemente de un nucleótido seleccionado de A, U, T, G y C o un análogo de nucleótido del mismo; siendo N_{0-2} una secuencia consecutiva de 0 a 2, preferiblemente de 0 a 1, más preferiblemente de 1 N, donde cada N se selecciona independientemente de un nucleótido seleccionado de A, U, T, G o C o un análogo de nucleótido del mismo; y donde C es citidina o un análogo de la misma, y se puede reemplazar opcionalmente por una guanosina o un análogo de la misma siempre que su guanosina de nucleósido complementario en tallo1 se reemplace por citidina;

donde tallo1 y tallo2 son capaces de apareamiento de bases entre sí formando una secuencia inversa complementaria, donde el apareamiento de bases puede ocurrir entre tallo1 y tallo2, por ejemplo por apareamiento de bases Watson-Crick de nucleótidos A y U/T o G y C o por apareamiento de bases no Watson-Crick, por ejemplo apareamiento de bases de tambaleo, apareamiento de bases Watson-Crick inverso, apareamiento de bases Hoogsteen, apareamiento de bases Hoogsteen inverso o son capaces de apareamiento de bases entre sí formando una secuencia complementaria parcialmente inversa, donde un apareamiento de bases incompleto puede ocurrir entre tallo1 y tallo2, en base a que una o más bases en un tallo no tienen una base complementaria en la secuencia complementaria inversa del otro tallo.

De acuerdo con una realización preferente adicional, la secuencia de tallo-bucle de histona se puede seleccionar de acuerdo con al menos una de las siguientes fórmulas específicas (Ia) o (IIa):

fórmula (Ia) (secuencia de tallo-bucle sin elementos frontera de tallo):

$$\underbrace{[N_{0-1}GN_{3-5}]}_{\text{tallo 1}}\ \underbrace{[N_{1-3}(U/T)N_{0-2}]}_{\text{bucle}}\ \underbrace{[N_{3-5}CN_{0-1}]}_{\text{tallo 2}}$$

fórmula (IIa) (secuencia de tallo-bucle con elementos frontera de tallo):

$$\underbrace{N_{2-5}}_{\text{elemento frontera de tallo 1}}\ \underbrace{[N_{0-1}GN_{3-5}]}_{\text{tallo 1}}\ \underbrace{[N_{1-3}(U/T)N_{0-2}]}_{\text{bucle}}\ \underbrace{[N_{3-5}CN_{0-1}]}_{\text{tallo 2}}\ \underbrace{N_{2-5}}_{\text{elemento frontera de tallo 2}}$$

donde N, C, G, T y U son como se definen anteriormente.

De acuerdo con otra realización adicional particularmente preferente del primer aspecto, la secuencia de la molécula de ácido nucleico artificial puede comprender al menos una secuencia tallo-bucle de histona acuerdo con al menos una de las siguientes fórmulas específicas (Ib) o (IIb):

fórmula (Ib) (secuencia de tallo-bucle sin elementos frontera de tallo):

$$\underbrace{[N_1GN_4]}_{\text{tallo 1}}\ \underbrace{[N_2(U/T)N_1]}_{\text{bucle}}\ \underbrace{[N_4CN_1]}_{\text{tallo 2}}$$

fórmula (IIb) (secuencia de tallo-bucle con elementos frontera de tallo):

$$\underbrace{N_{4-5}}_{\text{elemento frontera de tallo 1}}\ \underbrace{[N_1GN_4]}_{\text{tallo 1}}\ \underbrace{[N_2(U/T)N_1]}_{\text{bucle}}\ \underbrace{[N_4CN_1]}_{\text{tallo 2}}\ \underbrace{N_{4-5}}_{\text{elemento frontera de tallo 2}}$$

donde N, C, G, T y U son como se definen anteriormente.

Una secuencia tallo-bucle de histona particularmente preferente es la secuencia de acuerdo con la SEQ ID NO: 11: CAAAGGCTCTTTTCAGAGCCACCA o, más preferiblemente, la secuencia de ARN correspondiente de la secuencia de ácido nucleico de acuerdo con la SEQ ID NO: 11.

En algunas realizaciones, la molécula de ácido nucleico artificial comprende elementos adicionales, tales como una 5'-cap, una secuencia poli(C) y/o un motivo IRES. Se puede añadir una 5'-cap al extremo 5' de un ARN durante la transcripción o después de la transcripción. Además, la molécula de ácido nucleico artificial aquí descrita, en particular cuando el ácido nucleico está en forma de un ARNm o codifica un ARNm, puede modificarse mediante una secuencia de al menos 10 citidinas, preferiblemente al menos 20 citidinas, más preferiblemente al menos 30 citidinas (llamada "secuencia poli(C)"). En particular, la molécula de ácido nucleico artificial aquí descrita puede contener, especialmente si el ácido nucleico está en forma de un ARN(m) o codifica un ARNm, una secuencia poli(C) típicamente de aproximadamente 10 a 200 nucleótidos citidina, preferiblemente aproximadamente 10 a 100 nucleótidos citidina, más preferiblemente de aproximadamente 10 a 70 nucleótidos citidina o incluso más preferiblemente de aproximadamente 20 a 50 o incluso de 20 a 30 nucleótidos citidina. Con total preferencia, el ácido nucleico aquí descrito comprende una secuencia poli(C) de 30 residuos citidina. Así, preferiblemente, la molécula de ácido nucleico artificial aquí descrita comprende, preferiblemente en la dirección 5' a 3', un ORF, al menos un elemento 3'-UTR como se describió anteriormente, al menos una secuencia poli(A), una secuencia poli(C), una secuencia tallo-bucle de histona y, opcionalmente, otra secuencia poli(A).

Una secuencia de sitio de entrada de ribosoma interno (IRES) o un motivo IRES puede separar varios marcos de lectura abiertos, por ejemplo si la molécula de ácido nucleico artificial codifica dos o más péptidos o proteínas. Una secuencia IRES puede ser particularmente útil cuando la molécula de ácido nucleico artificial es una molécula de ácido nucleico bi o multicistrónico.

Además, la molécula de ácido nucleico artificial puede comprender elementos 5' adicionales, preferiblemente una 5'-UTR, un promotor o una 5'-UTR y una secuencia que contiene el promotor. El promotor puede conducir y/o regular la transcripción de la molécula de ácido nucleico artificial aquí descrita, por ejemplo de una molécula de ADN artificial aquí descrita. Además, la 5'-UTR puede consistir o puede comprender la 5'-UTR de un gen como se define aquí. Además, la 5'-UTR puede interactuar con la 3'-UTR de la molécula de ácido nucleico artificial aquí descrita y, por tanto, puede apoyar el efecto de la 3'-UTR de la molécula de ácido nucleico. Dichos elementos pueden apoyar aún más la estabilidad y la eficiencia de la traducción.

En una realización particularmente preferente aquí descrita, la molécula de ácido nucleico artificial comprende al menos un elemento de región no traducida 5' (elemento 5'-UTR) que comprende o consiste en una secuencia de ácido nucleico que se deriva de la 5'-UTR de un gen TOP o que se deriva de un fragmento, homólogo o variante de la 5'-UTR de un gen TOP.

Es particularmente preferente que el elemento 5'-UTR no comprenda un motivo TOP o un 5'-TOP como se definió anteriormente.

La secuencia de ácido nucleico que se deriva de la 5'-UTR de un gen TOP, preferiblemente se deriva de un gen TOP eucariota, preferiblemente un gen TOP vegetal o animal, en especial de un gen TOP de cordado, incluso con mayor preferencia de un gen TOP de vertebrado, más preferiblemente un gen TOP de mamífero, tal como un gen TOP humano.

Por ejemplo, el elemento 5'-UTR se selecciona preferiblemente de elementos 5'-UTR que comprenden o consisten en una secuencia de ácido nucleico que se deriva de una secuencia de ácido nucleico seleccionada del grupo consistente en las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de homólogos de las SEQ ID NO. 1- 1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de variantes de las mismas, o preferiblemente de una secuencia de ARN correspondiente. El término "homólogos de SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEQ ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700" se refiere a secuencias de otras especies que no son *homo sapiens* que son homólogas a las secuencias según las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700.

En una realización preferente, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una secuencia de ácido nucleico que se extiende desde la posición de nucleótido 5 (es decir, el nucleótido que se encuentra en la posición 5 de la secuencia) hasta la posición de nucleótido inmediatamente 5' al codón de inicio (ubicado en el extremo 3' de las secuencias), por ejemplo la posición de nucleótido

inmediatamente 5' respecto a la secuencia ATG de una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de homólogos de las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de variantes de las mismas, o de una secuencia de ARN correspondiente. Es particularmente preferente que el elemento 5'-UTR se derive de una secuencia de ácido nucleico que se extiende desde la posición del nucleótido inmediatamente 3' al 5'-TOP a la posición del nucleótido inmediatamente 5' al codón de inicio (ubicado en el extremo 3' de las secuencias), por ejemplo la posición de nucleótidos inmediatamente 5' respecto a la secuencia ATG de una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de homólogos de las SEQ ID NO. 1-1363, SEQ ID NO. 1395, SEC ID NO. 1421 y SEQ ID NO. 1422 de la solicitud de patente WO2013/143700, de variantes de las mismas, o de una secuencia de ARN correspondiente.

En una realización particularmente preferente, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 5'-UTR de un gen TOP que codifica una proteína ribosómica o de una variante de una 5'-UTR de un gen TOP que codifica una proteína ribosómica. Por ejemplo, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 5'-UTR de una secuencia de ácido nucleico de acuerdo con cualquiera de las SEQ ID NO: 170, 232, 244, 259, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, o 1360 de la solicitud de patente WO2013/143700, de una secuencia de ARN correspondiente, de un homólogo de la misma o de una variante de la misma como se describe aquí, preferiblemente que carece del motivo 5'-TOP. Como se describió anteriormente, la secuencia que se extiende desde la posición 5 al nucleótido inmediatamente 5’ al ATG (que se encuentra en el extremo 3' de las secuencias) corresponde a la 5'-UTR de dichas secuencias.

Preferentemente, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 5'-UTR de un gen TOP que codifica una proteína ribosómica Large (RPL) o de un homólogo o variante de una 5'-UTR de un gen TOP que codifica una proteína ribosómica Large (RPL). Por ejemplo, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 5'-UTR de una secuencia de ácido nucleico de acuerdo con cualquiera de las SEQ ID NO: 67, 259, 1284-1318, 1344, 1346, 1348-1354, 1357, 1358, 1421 y 1422 de la solicitud de patente WO2013/143700, de una secuencia de ARN correspondiente, de un homólogo de la misma o de una variante de la misma como se describe aquí, preferiblemente sin el motivo 5'-TOP.

En una realización particularmente preferente, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de la 5'-UTR de un gen de proteína ribosómica Large 32, preferiblemente de un gen de proteína ribosómica Large 32 de vertebrado (L32), más preferiblemente de un gen de proteína ribosómica Large 32 de mamífero (L32), con total preferencia de un gen de proteína ribosómica Large 32 humana (L32), o de una variante de la 5'-UTR de un gen de proteína ribosómica Large 32 (L32), preferiblemente de un gen de proteína ribosómica Large 32 (L32) de vertebrado, más preferiblemente de un gen de proteína ribosómica Large 32 (L32) de mamífero, más preferiblemente de un gen de proteína ribosómica Large 32 (L32) humana, donde preferiblemente el elemento 5'-UTR no comprende el 5’-TOP de dicho gen.

Así, en una realización particularmente preferente, el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que tiene una identidad de al menos aproximadamente un 40%, preferiblemente de al menos aproximadamente un 50%, preferiblemente de al menos aproximadamente un 60%, preferiblemente de al menos aproximadamente un 70%, más preferiblemente de al menos aproximadamente un 80%, más preferiblemente de al menos aproximadamente un 90%, incluso más preferiblemente de al menos aproximadamente un 95%, incluso más preferiblemente de al menos aproximadamente un 99% con la secuencia de ácido nucleico de acuerdo con la SEQ ID NO: 12 (5'-UTR de la proteína ribosómica humana Large 32 sin el tracto de oligopirimidina 5’-'terminal: GGCGCTGCCTACGGAGGTGGCAGCCATCTCCTTCTCGGCATC; correspondiente a la SEQ ID NO:. 1368 de la solicitud de patente WO2013/143700) o preferiblemente con una secuencia de ARN correspondiente, o donde el al menos un elemento 5'-UTR comprende o consiste en un fragmento de una secuencia de ácido nucleico que tiene una identidad de al menos aproximadamente un 40%, preferiblemente de al menos aproximadamente un 50%, preferiblemente de al menos aproximadamente un 60%, preferiblemente de al menos aproximadamente un 70%, más preferiblemente de al menos aproximadamente un 80%, más preferiblemente de al menos aproximadamente un 90%, incluso más preferiblemente de al menos aproximadamente un 95%, incluso más preferiblemente de al menos aproximadamente un 99% con la secuencia de ácido nucleico de acuerdo con la SEQ ID NO: 12 o más preferiblemente con una secuencia de

ARN correspondiente, donde, preferentemente, el fragmento es como se describió anteriormente, es decir, es un tramo continuo de nucleótidos que representa al menos el 20%, preferiblemente al menos el 30%, más preferiblemente al menos el 40% de la longitud total de la 5'-UTR. Preferiblemente, el fragmento tiene una longitud de al menos aproximadamente 20 nucleótidos o más, preferiblemente de al menos aproximadamente 30 nucleótidos o más, más preferiblemente de al menos aproximadamente 40 nucleótidos o más. Preferentemente, el fragmento es un fragmento funcional como se describe aquí.

En algunas realizaciones, la molécula de ácido nucleico artificial comprende un elemento 5'-UTR que comprende o consiste en una secuencia de ácido nucleico que se deriva de la 5'-UTR de un gen TOP de vertebrado, tal como mamífero, por ejemplo un gen TOP humano, seleccionado de RPSA, RPS2, RPS3, RPS3A, RPS4, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27A, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL27, RPL27A, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37A, RPL38, RPL39, RPL40, RPL41, RPLP0, RPLP1, RPLP2, RPLP3, RPLP0, RPLP1, RPLP2, EEF1A1, EEF1B2, EEF1D, EEF1G, EEF2, EIF3E, EIF3F, EIF3H, EIF2S3, EIF3C, EIF3K, EIF3EIP, EIF4A2, PABPC1, HNRNPA1, TPT1, TUBB1, UBA52, NPM1, ATP5G2, GNB2L1, NME2, UQCRB o de un homólogo o variante de los mismos, donde preferentemente el elemento 5'-UTR no comprende un motivo TOP o el 5'-TOP de dicho genes, y donde opcionalmente el elemento 5'-UTR comienza en su extremo 5' con un nucleótido en la posición 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aguas abajo del tracto de oligopirimidina 5' terminal (TOP) y donde además opcionalmente el elemento 5'-UTR que se deriva de una 5'-UTR de un gen TOP termina en su extremo 3' con un nucleótido en la posición 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aguas arriba del codón de inicio (A(U/T)G) del gen del que deriva.

Así, en algunas realizaciones, la descripción proporciona moléculas de ácido nucleico artificiales, preferiblemente moléculas de ARNm, que comprenden, en la dirección 5' a 3', al menos una de las siguientes estructuras

ORF - secuencia poli(A);
ORF - secuencia poli(A) - secuencia poli(A);
ORF - IRES - ORF - secuencia poli(A);
ORF - elemento 3'-UTR - secuencia poli(A);
ORF - secuencia poli(A) - elemento 3'-UTR;
ORF - elemento 3'-UTR - secuencia poli(A) - secuencia poli(C) – tallo-bucle de histona;
ORF - elemento 3'-UTR - secuencia poli(A) - secuencia poli(C) - secuencia poli(A);
ORF - elemento 3'-UTR - secuencia poli(A) – tallo-bucle de histona - secuencia poli(A);
ORF - elemento 3'-UTR - secuencia poli(A) - secuencia poli(C) – tallo-bucle de histona - secuencia poli(A);
5'-UTR - ORF - elemento 3'-UTR - secuencia poli(A) - secuencia poli(C) – tallo-bucle de histona - secuencia poli(A); o
5'-cap - 5'-UTR - ORF - elemento 3'-UTR - secuencia poli(A) - secuencia poli(C) – tallo-bucle de histona - secuencia poli(A).

Preferentemente, la molécula de ácido nucleico artificial aquí descrita, preferiblemente su marco de lectura abierto, está al menos parcialmente modificado en G/C. Por tanto, la molécula de ácido nucleico artificial aquí descrita puede estar estabilizada termodinámicamente modificando el contenido de G (guanosina) y/o C (citidina) (contenido G/C) de la molécula. El contenido G/C del marco de lectura abierto de una molécula de ácido nucleico artificial aquí descrita puede aumentar en comparación con el contenido G/C del marco de lectura abierto de una secuencia de tipo salvaje correspondiente, preferiblemente usando la degeneración del código genético. Por tanto, la secuencia de aminoácidos codificada de la molécula de ácido nucleico artificial preferentemente no se modifica por la modificación G/C en comparación con la secuencia de aminoácidos codificada de la secuencia particular de tipo salvaje. Los codones de la secuencia de codificación o la molécula de ácido nucleico artificial completa, por ejemplo ARNm, por tanto, pueden variar en comparación con la secuencia de codificación de tipo salvaje de modo que incluyan una mayor cantidad de nucleótidos G/C a la vez que se mantiene la secuencia de aminoácidos traducida. Debido al hecho de que varios codones codifican uno y el mismo aminoácido (la llamada degeneración del código genético), es factible alterar codones sin que se altere la secuencia de péptidos/proteína codificada (el llamado uso de codones alternativos). Por tanto, es posible introducir específicamente ciertos codones (a cambio de los respectivos codones de tipo salvaje que codifican el mismo aminoácido) que son más favorables con respecto a la estabilidad del ARN y/o con respecto al uso de codones en un sujeto (la llamada optimización de codones).

Dependiendo del aminoácido a codificar por la región de codificación de la molécula de ácido nucleico artificial como se define aquí, existen varias posibilidades de modificación de la secuencia de ácido nucleico, por

ejemplo del marco de lectura abierto, en comparación con su región de codificación de tipo salvaje. En el caso de los aminoácidos que están codificados por codones que contienen exclusivamente nucleótidos G o C, no es necesaria ninguna modificación del codón. Así, los codones para Pro (CCC o CCG), Arg (CGC o CGG), Ala (GCC o GCG) y Gly (GGC o GGG) no requieren modificación, ya que no hay A o U/T.

Por el contrario, los codones, que contienen los nucleótidos A y/o U/T pueden modificarse por la sustitución de otros codones que codifican los mismos aminoácidos pero no contienen A y/o U/T. Por ejemplo

los codones para Pro se pueden modificar de CC (U/T) o CCA a CCC o CCG;
los codones para Arg se pueden modificar de CG (U/T) o CGA o AGA o AGG a CGC o CGG;
los codones para Ala se pueden modificar de GC (U/T) o GCA a GCC o GCG;
los codones para Gly se pueden modificar de GG (U/T) o GGA a GGC o GGG.

En otros casos, aunque los nucleótidos A o (U/T) no pueden eliminarse de los codones, es posible disminuir el contenido de A y (U/T) mediante el uso de codones que contienen un menor contenido en nucleótidos A y/o (U/T). Ejemplos de éstos son:

los codones para Phe se pueden modificar de (U/T)(U/T)(U/T) a (U/T) (U/T)C;
los codones para Leu se pueden modificar de (U/T) (U/T)A, (U/T) (U/T)G, C(U/T) (U/T) o C(U/T)A C(U/T)C o C(U/T)G;
los codones para Ser se pueden modificar de (U/T)C(U/T) o (U/T)CA o AG(U/T) a (U/T)CC, (U/T)CG o AGC;
el codón para Tyr se puede modificar de (U/T)A(U/T) a (U/T)AC;
el codón para Cys se pueden modificar de (U/T)G(U/T) a (U/T)GC;
el codón para His se pueden modificar de CA(U/T) a CAC;
el codón para Gln se pueden modificar de CAA a CAG;
los codones para Ile se pueden modificar de A(U/T)(U/T) o A(U/T)A A(U/T)C;
los codones para Thr se pueden modificar de AC(U/T) o ACA a ACC o ACG;
el codón para Asn se pueden modificar de AA(U/T) ato AAC;
el codón para Lys se puede modificar de AAA a AAG;
los codones para Val se pueden modificar de G(U/T)(U/T) o G(U/T)A G(U/T)C o G(U/T)G;
el codón para Asp se pueden modificar de GA(U/T) a GAC;
el codón para Glu se pueden modificar de GAA a GAG;
el codón de parada (U/T)AA se puede modificar a (U/T)AG o (U/T)GA.

En el caso de los codones para Met (A(U/T)G) y Trp ((U/T)GG), por otra parte, no existe posibilidad de modificación de secuencia sin alterar la secuencia de aminoácidos codificada.

Las sustituciones citadas anteriormente se pueden utilizar ya sea individualmente o en todas las combinaciones posibles para incrementar el contenido de G/C del marco de lectura abierto de la secuencia de ácido nucleico artificial como se define aquí, comparado con su marco de lectura abierto de tipo natural particular (es decir la secuencia original). Así, por ejemplo, todos los codones para Thr que se presentan en la secuencia de tipo natural se pueden modificar a ACC (o ACG).

De manera preferente, el contenido de G/C del marco de lectura abierto de la molécula de ácido nucleico artificial como se define aquí se incrementa en al menos 7%, de manera más preferente al menos 15%, de manera preferente al menos 20%, en comparación con el contenido G/C de la región de codificación de tipo natural sin alterar la secuencia de aminoácidos codificada, esto es empleando la degeneración de código genético. De acuerdo con una realización específica al menos el 5%, 10%, 20%, 30%, 40%, 50%, 60%, de manera más preferente al menos 70%, de manera aún más preferente al menos 80% y de manera mucho más preferente al menos 90%, 95% o aún el 100% de los codones sustituibles en el marco de lectura abierto de la molécula de ácido nucleico artificial aquí descrita o de un fragmento, variante o derivada de la misma se sustituye, incrementando así el contenido de G/C del marco de lectura abierto.

En este contexto, es particularmente preferible incrementar el contenido de G/C del marco de lectura abierto de la secuencia de ácido nucleico artificial como se define aquí al máximo (es decir 100% de los codones sustituibles), en comparación con el marco de lectura abierto de tipo natural, sin alterar la secuencia de aminoácidos codificada.

Además, el marco de lectura abierto está preferiblemente optimizado al menos parcialmente en codón. La optimización de codones se basa en el hallazgo de que la eficiencia de la traducción puede estar determinada por una frecuencia diferente en la aparición de ARN de transferencia (ARNt) en las células. Así, si los denominados "codones raros" están presentes en la región de codificación de la molécula de ácido nucleico artificial como se define aquí en mayor medida, la traducción de la secuencia de ácido nucleico modificada

correspondiente es menos eficiente que en el caso en que están presentes codones que codifican ARNt relativamente "frecuentes".

Por tanto, el marco de lectura abierto de la molécula de ácido nucleico artificial aquí descrita se modifica preferiblemente en comparación con la región de codificación de tipo salvaje correspondiente de modo que al menos un codón de la secuencia de tipo salvaje que es reconocido por un ARNt que es relativamente raro en la célula, se intercambia por un codón que es reconocido por un ARNt que es relativamente frecuente en la célula y lleva el mismo aminoácido que el ARNt relativamente raro. Mediante esta modificación, el marco de lectura abierto de la molécula de ácido nucleico artificial como se define aquí se modifica de manera que los codones para los que están disponibles frecuentemente los ARNt pueden reemplazar los codones que corresponden a los ARNt raros. En otras palabras, según la presente descripción, con dicha modificación todos los codones del marco de lectura abierto de tipo salvaje que son reconocidos por un ARNt raro pueden intercambiarse por un codón que es reconocido por un ARNt que es más frecuente en la célula y que lleva el mismo aminoácido que el ARNt raro. Qué ARNt ocurren con relativa frecuencia en la célula y cuáles, en contraste, son relativamente raros son conocidos por el experto en la técnica; véase por ejemplo Akashi, Curr. Opin. Gineta. Dev. 2001 , 11 (6): 660-666. Por consiguiente, preferiblemente el marco de lectura abierto está optimizado en codón, preferentemente con respecto al sistema en el que la molécula de ácido nucleico artificial aquí descrita se va a expresar, preferiblemente con respecto al sistema en el que la molécula de ácido nucleico artificial aquí descrita se va a traducir. Preferentemente, el uso de codones del marco de lectura abierto se optimiza con codones de acuerdo con el uso de codones de mamíferos, más preferiblemente de acuerdo con el uso de codones humanos. Preferiblemente, el marco de lectura abierto está optimizado en codón y modificado su contenido G/C.

Para mejorar adicionalmente la resistencia a la degradación, por ejemplo a la degradación *in vivo* por una exo- o endonucleasa, y/o para mejorar adicionalmente la estabilidad de la producción de proteínas de la molécula de ácido nucleico artificial aquí descrita, la molécula de ácido nucleico artificial puede comprender además modificaciones tales como modificaciones de esqueleto, modificaciones de azúcar y/o modificaciones base, por ejemplo modificaciones de lípidos o similares. De manera preferente, la transcripción y/o la traducción de la molécula de ácido nucleico artificial aquí descrita no se deteriora significativamente por las modificaciones.

En general, la molécula de ácido nucleico artificial aquí descrita puede comprender cualquier nucleótido nativo (= natural), por ejemplo guanosina, uracilo, adenosina y/o citosina o un análogo de los mismos. A este respecto, los análogos de nucleótidos se definen como variantes de origen natural y no natural de los nucleótidos naturales adenosina, citosina, timidina, guanosina y uridina. Así, análogos son por ejemplo nucleótidos derivados químicamente con grupos funcionales de origen no nativo que preferentemente se añaden o eliminan del nucleótido natural o que sustituyen a los grupos funcionales naturales de un nucleótido. Por consiguiente, cada componente del nucleótido natural puede modificarse en, a saber, el componente base, el componente de azúcar (ribosa) y/o el componente fosfato que forma el esqueleto (véase más arriba) de la secuencia de ARN. Análogos de guanosina, uridina, adenosina, timidina y citosina incluyen, sin limitación, cualquier guanosina, uridina, adenosina, timidina o citosina de origen natural o no natural que haya sido alterada, por ejemplo químicamente, por ejemplo por acetilación, metilación, hidroxilación, etc., incluyendo 1-metiladenosina, 1-metilguanosina, 1 -metilinosina, 2,2-dimetilguanosina, 2,6-diaminopurina, 2'-amino-2'-desoxiadenosina, 2'-amino-2'-desoxicitidina, 2'-amino-2'-desoxiguanosina, 2'-amino-2'-desoxiuridina, 2-amino-6-cloropurinribósido, 2-aminopurinaribósido, 2'-araadenosina, 2'-aracitidina, 2'-arauridina, 2'-azido-2'-desoxiadenosina, 2’-azido-2'-desoxicitidina, 2'-azido-2'-desoxiguanosina, 2'-azido-2'-desoxiuridina, 2-cloroadenosina, 2'-fluor-2'-desoxiadenosina, 2'-fluor-2'-desoxicitidina, 2'-fluor-2'- desoxiguanosina, 2'-fluor-2'-desoxiuridina, 2'-fluorotimidina, 2-metiladenosina, 2- metilguanosina, 2-metiltio-N6-isopenteniladenosina, 2'-O-Metil-2-aminoadenosina, 2'-O-metil-2'-desoxiadenosina, 2'-O-metil-2'-desoxicitidina, 2'-O-metil-2'- desoxiguanosina, 2'-O-metil-2'-desoxiuridina, 2'-O-metil-5-metiluridina, 2'-O-metilinosina, 2'-O-metilpseudouridina, 2-tiocitidina, 2-tiocitosina, 3-metilcitosina, 4-acetilcitosina, 4-tiouridina, 5-(carboxihidroximetil)uracilo, 5,6-dihidrouridina, 5-aminoalilcitidina, 5-aminoalildesoxiuridina, 5-bromouridina, 5- carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluridina, 5-cloroaracitosina, 5-fluorouridina, 5-yodouridina, 5-metoxicarbonilmetil-uridina, 5-metoxiuridina, 5-metil-2-tiouridina, 6-azacitidina, 6-azauridina, 6-cloro-7-deaza-guanosina, 6-cloropurinribósido, 6-mercaptoguanosina, 6-metil-mercaptopurin-ribósido, 7-deaza-2'-desoxiguanosina, 7-deazaadenosina, 7-metilguanosina, 8-azaadenosina, 8-bromo-adenosina, 8-bromoguanosina, 8-mercaptoguanosina, 8-oxoguanosina, benzimidazol-ribósido, beta-D-manosilqueosina, dihidrouridina, inosina, N1-metiladenosina, N6-([6-aminohexil]carbamoilmetil)adenosina, N6-isopenteniladenosina, N6-metiladenosina, N7-metilxantosina, metil éster de ácido N-uracil-5-oxiacético, puromicina, queosina, ácido uracil-5-oxiacético, metil éster de ácido uracil-5-oxiacético, wibutoxosina, xantosina y xiloadenosina. La preparación de dichos análogos es conocida por los expertos en la técnica, por ejemplo de las Patentes US 4.373.071, US 4.401, 796, US 4.415.732, US 4.458.066, US 4.500.707, US 4.668.777, US 4.973.679, US 5.047.524, US 5.132.418, US 5.153.319 , US 5.262.530 y 5.700.642. En el caso de un análogo como se describe anteriormente, se puede dar particular preferencia en ciertas realizaciones aquí descritas a aquellos análogos que aumentan la expresión de la proteína o del péptido o proteína codificados o que aumentan la

inmunogenicidad de la molécula de ácido nucleico artificial aquí descrita y/o no interfieren con otra modificación de la molécula de ácido nucleico artificial que se ha introducido.

Según una realización particular, la molécula de ácido nucleico artificial aquí descrita puede contener una modificación lipídica.

En una realización preferente, la molécula de ácido nucleico artificial aquí descrita comprende, preferiblemente de 5' a 3', los siguientes elementos:

un ORF;
un elemento 3'-UTR opcional como se describe aquí;
una secuencia poli(A), que preferiblemente comprende o consiste en 64 residuos de adenosina;
una secuencia poli(C), que preferiblemente comprende o consiste en 30 residuos de citosina;
una secuencia tallo-bucle de histona, que comprende o consiste preferiblemente en la secuencia de ácido nucleico de acuerdo con la SEQ ID NO: 11; y
una secuencia poli(A), que comprende preferiblemente al menos 160 residuos de adenosina.

Preferentemente, la molécula de ácido nucleico artificial, que preferiblemente es un ARNm, opcionalmente comprende además una 5'-UTR, que comprende preferiblemente un elemento 5'-UTR que comprende o consiste en un ácido nucleico que se deriva de una 5'-UTR de un gen TOP que codifica una proteína ribosómica Large (RPL) o de un homólogo, fragmento o variante de la misma, preferiblemente sin el motivo 5'-TOP. Incluso más preferiblemente, la molécula de ácido nucleico artificial aquí descrita comprende además una estructura 5'-cap.

En una realización preferente, el al menos un marco de lectura abierto de la molécula de ácido nucleico artificial codifica una proteína o péptido terapéutico. En otra realización, un antígeno está codificado por al menos un marco de lectura abierto, tal como un antígeno patógeno, un antígeno tumoral, un antígeno alergénico o un antígeno autoinmune. Allí, la administración de la molécula de ácido nucleico artificial que codifica el antígeno se usa en un enfoque de vacunación genética contra una enfermedad que implica a dicho antígeno.

En una realización alternativa, un anticuerpo está codificado por el al menos un marco de lectura abierto de la molécula de ácido nucleico artificial aquí descrita.

Antígenos:

Antígenos patógenos:

La molécula de ácido nucleico artificial aquí descrita puede codificar una proteína o un péptido que comprende un antígeno patógeno o un fragmento, variante o derivado del mismo. Dichos 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 patogénicos son antígenos de péptido o proteína derivados de un patógeno asociado a una enfermedad infecciosa, preferiblemente seleccionados de antígenos derivados de los patógenos Acinetobacter baumannii, Anaplasma genus, Anaplasma phagocytophilum, Ancylostoma braziliense, Ancylostoma duodenale, Arcanobacterium haemolyticum, Ascaris lumbricoides, Aspergillus genus, Astroviridae, Babesia genus, Bacillus anthracis, Bacillus cereus, Bartonella henselae, virus BK, Blastocystis hominis, Blastomyces dermatitidis, Bordetella pertussis, Borrelia burgdorferi, Borrelia genus, Borrelia spp, Brucella genus, Brugia malayi, familia Bunyaviridae, Burkholderia cepacia y otras especies de Burkholderia, Burkholderia mallei, Burkholderia pseudomallei, familia de Caliciviridae, Campylobacter genus, Candida albicans, Candida spp, Chlamydia trachomatis, Chlamydophila pneumoniae, Chlamydophila psittaci, CJD prion, Clonorchis sinensis, Clostridium botulinum, Clostridium difficile, Clostridium perfringens, Clostridium perfringens, Clostridium spp, Clostridium tetani, Coccidioides spp, coronavirus, Corynebacterium diphtheriae, Coxiella burnetii, virus de fiebre hemorrágica de Crimean-Congo, Cryptococcus neoformans, Cryptosporidium genus, Citomegalovirus (CMV), virus del Dengue (DEN-1, DEN-2, DEN-3 and DEN-4), Dientamoeba fragilis, Ebolavirus (EBOV), Echinococcus genus, Ehrlichia chaffeensis, Ehrlichia ewingii, Ehrlichia genus, Entamoeba histolytica, Enterococcus genus, Enterovirus genus, Enteroviruses, principalmente virus Coxsackie A y Enterovirus 71 (EV71), Epidermophyton spp, Virus Epstein-Barr (EBV), Escherichia coli O157:H7, O111 y O104:H4, Fasciola hepatica y Fasciola gigantica, FFI prion, superfamilia Filarioidea, Flavivirus, Francisella tularensis, Fusobacterium genus,

Geotrichum candidum, Giardia intestinalis, Gnathostoma spp, GSS prion, virus Guanarito, Haemophilus ducreyi, Haemophilus influenzae, Helicobacter pylori, Henipavirus (virus Hendra virus Nipah), Virus de Hepatitis A, Virus de Hepatitis B (HBV), Virus de Hepatitis C (HCV), Virus de Hepatitis D, Virus de Hepatitis E, virus Herpes simple 1 y 2 (HSV-1 y HSV-2), Histoplasma capsulatum, VIH (Virus de inmunideficiencia humana), Hortaea werneckii, bocavirus humano (HBoV), herpesvirus 6 humano (HHV-6) y herpesvirus 7 humano (HHV-7), metapneumovirus humano (hMPV), virus del papiloma humano (HPV), virus de parainfluenza (HPIV), virus de encefalitis Japonesa, virus JC, virus Junín, Kingella kingae, Klebsiella granulomatis, Kuru prion, virus Lassa, Legionella pneumophila, Leishmania genus, Leptospira genus, Listeria monocytogenes, Virus de coriomeningitis linfocítica (LCMV), virus Machupo, Malassezia spp, virus Marburg, virus Measles, Metagonimus yokagawai, Microsporidia phylum, virus Molluscum contagiosum (MCV), virus Mumps, 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 de Orthomyxoviridae (Influenza), Paracoccidioides brasiliensis, Paragonimus spp, Paragonimus westermani, Parvovirus B19, Pasteurella genus, Plasmodium genus, Pneumocystis jirovecii, Poliovirus, Virus de la rabia, Virus sincitial respiratorio (RSV), Rinovirus, rinovirus, Rickettsia akari, Rickettsia genus, Rickettsia prowazekii, Rickettsia rickettsii, Rickettsia typhi, virus de la fiebre Rift Valley, Rotavirus, Virus de la rubéola, virus Sabia, género Salmonella, Sarcoptes scabiei, SARS coronavirus, género Schistosoma, género Shigella, virus Sin Nombre, Hantavirus, Sporothrix schenckii, Staphylococcus genus, Staphylococcus genus, Streptococcus agalactiae, Streptococcus pneumoniae, Streptococcus pyogenes, Strongyloides stercoralis, Taenia genus, Taenia solium, Virus de encefalitis transmitido por garrapatas (TBEV), Toxocara canis o Toxocara cati, Toxoplasma gondii, Treponema pallidum, Trichinella spiralis, Trichomonas vaginalis, Trichophyton spp, Trichuris trichiura, Trypanosoma brucei, Trypanosoma cruzi, Ureaplasma urealyticum, virus zóster de varicela (VZV), Virus zóster de varicela (VZV), Viruela mayor o Viruela menor, vCJD prion, Virus de encefalitis equina venezolana, Vibrio cholerae, virus del Nilo Occidental, Virus de encefalitis equina Occidental, Wuchereria bancrofti, Virus de fiebre amarilla, Yersinia enterocolitica, Yersinia pestis, y Yersinia pseudotuberculosis.

En este contexto, son particularmente preferentes antígenos de los patógenos seleccionados de virus de la gripe, virus sincitial respiratorio (RSV), virus de herpes simple (HSV), virus de papiloma humano (HPV), virus de inmunodeficiencia humana (VIH), Plasmodium, Staphylococcus aureus, virus del Dengue, Chlamydia trachomatis, Citomegalovirus (CMV), virus de hepatitis B (HBV), Mycobacterium tuberculosis, virus de la rabia y virus de fiebre amarilla.

Antígenos tumorales:

En una realización adicional, la molécula de ácido nucleico artificial aquí descrita puede codificar 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 donde el antígeno tumoral es un antígeno específico de melanocitos, un antígeno de cáncer de testículo o un antígeno específico de tumor, preferiblemente un antígeno CT-X, un antígeno CT no X, un compañero de unión para un antígeno de CT-X o un socio de unión para un antígeno CT no X o un antígeno específico de tumor, más preferiblemente un antígeno CT-X, un compañero de unión para un antígeno CT no X o un antígeno específico de tumor o un fragmento, variante o derivado de dicho antígeno tumoral; y donde cada una de las secuencias de ácido nucleico codifica un péptido o proteína diferente; y donde 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 1 5-3 / CA 27-29, CA 19-9, CA72-4, CA125, calreticulina, CAMEL, CASP-8/m, catepsina B, catepsina 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 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-R171, HLA-A11/m, HLA-A2/m, HNE, homeobox NKX3.1, HOM-TES-14 / SCP-1, HOM-TE S-85, HPV-E6, HPV-E7, HSP70-2M, HST-2, hTERT, iCE, IGF-1 R, IL-13Ra2, IL-2R, IL-5, receptor de laminina inmaduro, kalikreía-2, kalikreína-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-B1 6, MAGE-B1 7 , MAGE-C1, MAGE-C2, MAGE-C3, MAGE-D1, MAGE-D2, MAGE-D4, MAGE-E1, MAGE-E2, MAGE-F1, MAGE-H1, MAGEL2, mamaglobina A, MART-1/melan-A, MART-2, MART-2/m, matriz de proteína 22, MC1R, M-CSF, ME1/m, mesotelina, MG50/PXDN, MMP11, MN/CA IX-antígeno, 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, p190 minor bcr-abl, p53, p53/m, PAGE-4, PAI-1, PAI-2, PAP, PART-1, PATE, PDEF, Pim-1-Quinasa, 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 una célula sanguínea linfoide o un idiotipo de receptor de células T de una célula sanguínea linfoide, 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 una pareja de unión de la misma o un fragmento, variante o derivado de dicho antígeno tumoral. Particularmente preferidos en este contexto son los antígenos tumorales NY-ESO-1, 5T4, MAGE-C1, MAGE-C2, survivina, Muc-1, PSA, PSMA, PSCA, STEAP y PAP.

En una realización preferente, la molécula de ácido nucleico artificial codifica una proteína o un péptido que comprende una proteína terapéutica o un fragmento, variante o derivado de la misma.

Las proteínas terapéuticas como se definen aquí son péptidos o proteínas beneficiosos para el tratamiento de cualquier enfermedad hereditaria o adquirida, o que mejoran la condición de un individuo. En particular, las proteínas terapéuticas desempeñan un papel importante en la creación de agentes terapéuticos que podrían modificar y reparar errores genéticos, destruir células cancerígenas o células infectadas con patógenos, tratar trastornos del sistema inmunológico, tratar trastornos metabólicos o endocrinos, entre otras funciones. Por ejemplo, la eritropoyetina (EPO), una hormona proteica, se puede utilizar para tratar a pacientes con deficiencia de eritrocitos, que es una causa común de complicaciones renales. Además, las proteínas adyuvantes, los anticuerpos terapéuticos están incluidos en las proteínas terapéuticas y también la terapia de reemplazo hormonal, que por ejemplo se emplea en la terapia femenina en la menopausia. En enfoques más recientes, las células somáticas de un paciente se utilizan para reprogramarlas en células madre pluripotentes, que reemplazan la terapia de células madre en disputa. También estas proteínas utilizadas para la reprogramación de células somáticas o usadas para diferenciar las células madre se definen aquí como proteínas terapéuticas. Además, las proteínas terapéuticas se pueden usar para otros fines, por ejemplo para la curación de heridas, la regeneración de tejidos, la angiogénesis, etc. Además, los receptores y fragmentos de células B específicos de antígeno y sus variantes se definen aquí como proteínas terapéuticas.

Por tanto, las proteínas terapéuticas se pueden usar para diversos fines, incluido el tratamiento de diversas enfermedades tales como 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 sistema circulatorio, enfermedades del sistema respiratorio, enfermedades del sistema nervioso, del sistema digestivo, enfermedades de la piel y tejido subcutáneo, enfermedades del sistema musculoesquelético y tejido conectivo y enfermedades del sistema genitourinario, independientemente de si son heredadas o adquiridas.

En este contexto, las proteínas terapéuticas particularmente preferidas que se pueden usar, entre otras, en el tratamiento de trastornos metabólicos o endocrinos se seleccionan entre (entre paréntesis la enfermedad particular para la cual se utiliza la proteína terapéutica en el tratamiento): esfingomielinasa ácida (enfermedad de Niemann-Pick), adipotida (obesidad), agalsidasa-beta (galactosidasa humana A) (Enfermedad de Fabry previene la acumulación de lípidos que podrían conducir a complicaciones renales y cardiovasculares), alglucosidasa (enfermedad de Pompe (enfermedad de almacenamiento de glucógeno tipo II)), alfa-galactosidasa A (alfa-GAL A, agalsidasa alfa) (enfermedad de Fabry), alfa-glucosidasa (Enfermedad de almacenamiento de glucógeno (GSD), Morbus Pompe), alfa-L-iduronidasa (mucopolisacaridosis (MPS), síndrome de Hurler, síndrome de Scheie), alfa-N-acetilglucosaminidasa (síndrome de Sanfilippo), anfiregulina (cáncer, trastorno metabólico), angiopoyetina ((Ang1, Ang2, Ang3, Ang4, ANGPTL2, ANGPTL3, ANGPTL4, ANGPTL5, ANGPTL6, ANGPTL7) (angiogénesis, estabilización de vasos), Betacellulina (trastorno metabólico), Beta-glucuronidasa (síndrome de Sly), proteína morfogenética ósea BMP (BMP1 , BMP2, BMP3, BMP4, BMP5, BMP6, BMP7, BMP8a, BMP8b, BMP10, BMP1 5) (efecto regenerativo, afecciones relacionadas con los huesos, enfermedad renal crónica (ERC)), proteína CLN6 (enfermedad CLN6 - Infancia tardía atípica, tardía Variante de inicio, Juvenil Temprano, Lipofuscinosis Ceroides Neuronales (NCL), Factor de crecimiento epidérmico (EGF) (cicatrización de heridas, regulación del crecimiento celular, proliferación y diferenciación), Epigen (trastorno metabólico), Epiregulina (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-1 1, FGF- 12, FGF-13, FGF-14, FGF-1 6, FGF-1 7, FGF-1 7, FGF-1 8, FGF-1 9, FGF-20, FGF-21, FGF-22, FGF-23) (cicatrización de heridas, angiogénesis, trastornos endocrinos, regeneración de tejidos), Galsulfasae (Mucopolisacaridosis VI), Grelina (síndrome del intestino irritable (SII), obesidad, síndrome de Prader-Willi, diabetes mellitus tipo II), Glucocerebrosidasa (enfermedad de Gaucher), GM- CSF (efecto regenerativo, producción de glóbulos blancos, cáncer), factor de crecimiento de unión a heparina de tipo EGF (HB-EGF) (cicatrización de heridas, hipertrofia cardíaca y desarrollo y función del corazón), factor de crecimiento de hepatocitos HGF (efecto regenerativo, 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 αVβ3, αVβ5 y α5β1 (macromoléculas de matriz de enlace y proteinasas, angiogénesis), luduronato 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, Derivado del Cerebro) Factor (BDNF), Neurotrofina-3 (NT-3) y Neurotrofina 4/5 (NT-4/5) (efecto regenerativo, 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 axón, supervivencia celular, migración), Obestatina (síndrome del intestino irritable (SII), obesidad, síndrome de Prader-Willi, diabetes mellitus tipo II), factor de crecimiento derivado de las plaquetas (PDGF (PDFF-A) , PDGF-B, PDGF-C, PDGF-D) (efecto regenerativo, cicatrización de heridas, trastorno en angiogénesis, arteriosclerosis, fibrosis, cáncer, receptores de TGF beta (endoglina, receptor de TGF-beta 1, receptor de TGF-beta 2, receptor de TGF-beta 3) (fibrosis renal, enfermedad renal, diabetes, enfermedad renal en etapa terminal (ERT), angiogénesis), trombopoyetina (THPO) (factor de crecimiento y desarrollo de megacariocitos (MGDF)) (trastornos plaquetarios, plaquetas para donación, recuperación de recuentos de plaquetas después de la quimioterapia mielosupresora), factor de crecimiento transformante (TGF (TGF-a, TGF-beta) (TGFbetal, TGFbeta2 y TGFbeta3)) (efecto regenerativo, curación de heridas, inmunidad, cáncer, cardiopatía, diabetes, síndrome de Marfan, síndrome de Loeys-Dietz), VEGF (VEGF-A, VEGF-B, VEGF-C, VEGF) -D, VEGF-E, VEGF-F y PIGF) (efecto regenerativo, angiogénesis, cicatrización de heridas, cáncer, permeabilidad), Nesiritida (insuficiencia cardíaca congestiva descompensada aguda), tripsina (úlcera de decúbito, úlcera varicosa, desbridamiento de la escara, herida dehiscente) , quemadura solar, 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 atrio-natriurético (ANP) (trastornos endocrinos), Colecistocinina (diversa), Gastrina (hipogastrinemia), Leptina (Diabetes,hipertrigliceridemia, obesidad), oxitocina (estimulación de la lactancia materna, no progresión del parto), somatostatina (tratamiento sintomático del síndrome carcinoide, hemorragia aguda de las várices y acromegalia, poliquistosis hepática y renal, acromegalia y síntomas causados por tumores neuroendocrinos), vasopresina (hormona antidiurética) (diabetes insípida), calcitonina (osteoporosis posmenopáusica, hipercalcemia, enfermedad de Paget, metástasis óseas, dolor del miembro fantasma, estenosis espinal), exenatida (diabetes tipo 2 resistente al tratamiento con metformina y una sulfonilurea), hormona de crecimiento (HG), somatotropina (fallo de crecimiento debido a deficiencia de GH o insuficiencia renal crónica, síndrome de Prader-Willi, síndrome de Turner, pérdida del SIDA o caquexia con terapia antiviral), insulina (Diabetes mellitus, cetoacidosis diabética, hiperpotasemia), factor de crecimiento similar a la insulina 1 IGF-1 (Fallo de crecimiento en niños con deleción del gen GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedad cardiovascular, insuficiencia cardíaca), Mecasermina rionfabato, análogo de IGF-1 (Fallo de crecimiento en niños con deleción del gen GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedades cardiovasculares, insuficiencia cardíaca), Mecasermina, análogo de IGF-1 (falla de crecimiento en niños con Supresión del gen GH o deficiencia primaria grave de IGF1, enfermedad neurodegenerativa, enfermedades cardiovasculares, insuficiencia cardíaca), Pegvisomant (Acromegalia), Pramlintida (Diabetes mellitus, en combinación con insulina), Teriparatida (hormona paratiroidea humana residuos 1-34) (osteoporosis severa), Becaplermina (complemento de desbridamiento para las úlceras diabéticas), dibotermina alfa (proteína morfogenética ósea 2) (cirugía de fusión espinal, reparación de lesión ósea), acetato de histrelina (hormona liberadora de gonadotropina); GnRH) (pubertad precoz), octreotida (acromegalia, alivio sintomático de los tumores de adenoma secretor de VI P y metástasis de los carcinoides) y palifermina (factor de crecimiento de queratinocitos; KGF) (mucosidad oral grave en pacientes sometidos a quimioterapia, cicatrización de heridas).

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 típicamente mamíferas, en particular proteínas humanas.

Para el tratamiento de trastornos sanguíneos, enfermedades del sistema circulatorio, enfermedades del sistema respiratorio, cáncer o enfermedades tumorales, enfermedades infecciosas o inmunodeficiencias después de las proteínas terapéuticas, se pueden usar: Alteplasa (activador tisular del plasminógeno; tPA) (embolia pulmonar, infarto de miocardio, ictus isquémico agudo, oclusión de los dispositivos de acceso venoso central), anistreplasa (trombolisis), antitrombina III (AT-III) (deficiencia hereditaria de AT-III, tromboembolismo), bivalirudina (reduce el riesgo de coagulación sanguínea en la angioplastia coronaria y 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) (sepsis grave con alto riesgo de muerte), eritropoyetina, epoetina-alfa, eritropoyetina, ertropoyetina (anemia por enfermedad crónica,

mileodisplasia, anemia por insuficiencia renal o quimioterapia, preparación preoperatoria), Factor IX (Hemofilia B), Factor Vila (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, Purpura fulminans), Reteplasa (eliminación de muteína de tPA) (Manejo del infarto agudo de miocardio, mejoría de la función ventricular), Estreptoquinasa (Infarto de miocardio transmural evolutivo agudo, embolia pulmonar, trombosis venosa profunda, trombosis o embolia arterial, oclusión de la cánula arteriovenosa, cánula de Tenecte, infarto de miocardio), urocinasa (embolia pulmonar), angiostatina (cáncer), inmunotoxina anti-CD22 (recaída de leucemia mieloide aguda CD33+), denileucina diftitox (linfoma cutáneo de células T (CCD)), inmunocianina (cáncer de vejiga y próstata), MPS (Metalopanstimulina) (Cáncer), Aflibercept (Cáncer de pulmón no microcítico (CPNM), cáncer colorrectal metastásico (mCRC, cáncer metastásico refractario a las hormonas, degeneración macular húmeda), Endostatina (cáncer, enfermedades inflamatorias como la artritis reumatoide y la enfermedad de Crohn, retinopatía diabética, psoriasis y endometriosis), colagenasa (desbridamiento de úlceras dérmicas crónicas y áreas severamente quemadas, contractura de Dupuytren, enfermedad de Peyronie), desoxi-ribonucleasa I humana, Dornasa (Fibrosis quística; disminuye las infecciones del tracto respiratorio en pacientes seleccionados con FVC superior al 40% de lo previsto), hialuronidasa (utilizada como adyuvante para aumentar la absorción y dispersión de los fármacos inyectados, particularmente anestésicos en cirugía oftálmica y ciertos agentes de imagen), papaína (desecho de tejido necrótico o licuefacción de la muda en lesiones agudas y crónicas, como úlceras por presión, úlceras varicosas y diabéticas, quemaduras, heridas postoperatorias, heridas de quistes pilonidales, carbuncles y otras heridas), L-Asparaginasa (leucemia linfocítica aguda, que requiere una 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 en tratamiento contra el cáncer que puede causar el síndrome de lisis tumoral), gonadotropina coriónica humana (HCG) (Reproducción asistida), Hormona estimulante del folículo humano (FSH) (Reproducción asistida), Lutropina -alfa (Infertilidad con deficiencia de la hormona luteinizante), Prolactina (Hipoprolactinemia, deficiencia de prolactina en suero, disfunción ovárica en mujeres, ansiedad, disfunción eréctil arteriogénica, eyaculación precoz, oligozoospermia, astenospermia, hipofunción de vesículas seminales, hipoandrogenia), Inhibidor de proteinasa (deficiencia de antitripsina congénita), lactasa (gases, distensión abdominal, calambres y diarrea debido a la incapacidad de digerir la lactosa), enzimas pancreáticas (lipasa, amilasa, proteasa) (fibrosis quística, pancreatitis crónica, insuficiencia pancreática, post-Billroth II, cirugía de bypass geriátrica, obstrucción del conducto pancreático, esteatorrea, mala digestión, gases, distensión abdominal), adenosina desaminasa (pegademase bovine, PEG-ADA) (enfermedad de inmunodeficiencia combinada grave debida a deficiencia de adenosina desaminasa), abatacept (artritis reumatoide (especialmente cuando es refractaria a la inhibición de TNF)), Alefacept (psoriasis en placas), Anakinra (artritis reumatoide), Etanercept (artritis reumatoide, Artritis reumatoide de curso poliarticular juvenil, artritis psoriásica, espondilitis anquilosante, psoriasis en placas, espondilitis anquilosante), antagonista del receptor de la interleucina 1 (IL-1), Anakinra (inflamación y degfradación de cartílago asociada a artritis reumatoide), Thimilina (enfermedades neurodegenerativas, reumatismo, anorexia nerviosa), antagonista de TNF-alfa (trastornos autoinmunes como artritis reumatoide, espondilitis anquilosante, enfermedad de Crohn, psoriasis, hidradenitis supurativa, asma refractaria), enfuvirtida (infección por VIH-1) y timosina α1 (hepatitis B y C). (Entre paréntesis la enfermedad particular para la cual se utiliza la proteína terapéutica en el tratamiento).

En otro aspecto, la presente invención proporciona un vector como se define en las reivindicaciones, que comprende una molécula de ácido nucleico artificial que comprende

a) al menos un marco de lectura abierto ORF y
b) una región no traducida 3’ (3'-UTR) que comprende
al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, y donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende una secuencia tallo-bucle de histona según la SEQ ID NO: 11.

También se describe aquí una molécula de ácido nucleico artificial que comprende un ORF y una 3’-UTR, comprendiendo la 3’-UTR al menos una secuencia poli(A) como se define aquí, donde la al menos una secuencia poli(A) comprende al menos 70 nucleótidos adenina. En una realización preferente, la molécula de ácido nucleico artificial comprende una 3’-UTR que comprende al menos dos secuencias poli(A) como se define aquí.

La 3’UTR y el ORF son como se describen aquí anteriormente para la molécula de ácido nucleico artificial aquí descrita. El sitio de clonación puede ser cualquier secuencia adecuada para introducir un marco de lectura abierto o una secuencia que comprende un marco de lectura abierto, tal como uno o más sitios de restricción. Así, el vector que comprende un sitio de clonación preferentemente es adecuado para insertar un marco de lectura abierto en el vector, de manera preferente para insertar un marco de lectura abierto 5’ a la 3’-UTR.

Preferentemente, el sitio de clonación o el ORF se sitúa 5' a la 3'-UTR, de manera preferente en proximidad cercana al extremo 5' de la 3'-UTR. Por ejemplo, el sitio de clonación o el ORF se pueden unir directamente al extremo 5' de la 3'-UTR o se pueden unir a través de un tramo de nucleótidos, tal como un tramo de 2, 4, 6, 8, 10, 20, etc., nucleótidos como se describe anteriormente para la molécula de ácido nucleico artificial aquí descrita.

Preferentemente, el vector aquí descrito es adecuado para producir la molécula de ácido nucleico artificial aquí descrita, preferentemente para producir un ARNm artificial aquí descrito, por ejemplo insertando opcionalmente el marco de lectura abierto o una secuencia que comprende un marco de lectura abierto en el vector y transcribiendo el vector. Así, preferentemente el vector comprende elementos necesarios para la transcripción, como un promotor, por ejemplo un promotor de ARN polimerasa. De manera preferente, el vector es adecuado para la transcripción utilizando sistemas de transcripción eucarióticos, procarióticos, virales o en fago, tales como células eucarióticas, células procarióticas o sistemas de transcripción *in vitro* eucarióticos, procarióticos, virales o en fago. Así, por ejemplo, el vector puede comprender una secuencia promotora que es reconocida por una polimerasa, tal como por una ARN-polimerasa, por ejemplo una ARN-polimerasa eucariótica, procariótica, viral o en fago. En una realización preferente, el vector comprende un promotor de ARN-polimerasa en un fago tal como SP6, T3 o T7, preferentemente un promotor T7. De manera preferente, el vector es adecuado para la transcripción *in vitro* utilizando un sistema de transcripción *in vitro* basado en fago, tal como una ARN polimerasa T7 basada en el sistema de transcripción *in vitro*.

En otra realización preferente, el vector puede usarse directamente para la expresión del péptido o proteína codificado en células o tejidos. Para ello, el vector comprende elementos particulares que son necesarios para la expresión en esas células/tejido, por ejemplo secuencias promotoras particulares, tales como un promotor de CMV.

El vector puede comprender además una señal de poliadenilación como se describe anteriormente para la molécula de ácido nucleico artificial.

El vector puede ser un vector de ARN o un vector de ADN. De manera preferente, el vector es un vector de ADN. El vector puede ser cualquier vector conocido por el experto, tal como un vector viral o plásmido. De manera preferente, el vector es un vector plásmido, de manera preferente un vector plásmido de ADN.

En una realización preferente aquí descrita, la presente descripción proporciona un vector que comprende la molécula de ácido nucleico artificial aquí descrita.

De manera preferente, el vector es una molécula circular. Preferentemente, el vector es una molécula de doble hebra, tal como una molécula de ADN de doble hebra. Tal molécula de ADN preferente de doble hebra circular se puede utilizar convenientemente como una forma de almacenamiento para la molécula de ácido nucleico artificial aquí descrita. Además, se puede utilizar para la transfección de células, por ejemplo células cultivadas. También se puede utilizar para la transcripción *in vitro* con el fin de obtener una molécula de ARN artificial aquí descrita.

Preferentemente, el vector, en especial el vector circular, es linealizable, por ejemplo por digestión con enzimas de restricción. En una realización preferida, el vector comprende un sitio de escisión, tal como un sitio de restricción, de manera preferente un sitio de escisión único, situado inmediatamente 3' a la 3'-UTR o situado 3' respecto a la secuencia poli(A) o – si está presente – a la señal de poliadenilación, o – si está presente – situado 3' a la secuencia poli(C), o – si está presente – situado 3' al tallo-bucle de histona. Así, preferentemente, el producto obtenido al linealizar el vector termina en el extremo 3' con el extremo 3' de la 3'-UTR o con el extremo 3' de la secuencia poli(A) o – si está presente – de la señal de poliadenilación, o - si está presente – con el extremo 3' de la secuencia poli(C) o – si está presente – con el extremo 3' del tallo-bucle de histona. En la realización donde el vector aquí descrito comprende la molécula de ácido nucleico artificial aquí descrita, un sitio de restricción, preferentemente un único sitio de restricción, se sitúa de manera preferente inmediatamente 3' al extremo 3' de la molécula de ácido nucleico artificial.

Preferiblemente, un sitio de restricción para la linealización de la molécula del vector circular se encuentra en el extremo 3' de la 3'-UTR de la hebra de codificación.

En un aspecto adicional, la presente invención se refiere a una célula tal como se define en las reivindicaciones, que comprende la molécula de ácido nucleico artificial o el vector como se definen en las reivindicaciones. Además, se describe aquí una célula que comprende una molécula de ácido nucleico artificial o un vector aquí descritos. La célula puede ser cualquier célula, como una célula bacteriana, de insecto, una célula vegetal, una célula de vertebrada, por ejemplo de mamífero. Dicha célula puede usarse, por ejemplo, para la replicación del vector aquí descrito, por ejemplo en una célula bacteriana. Además, la célula puede usarse para transcribir la

molécula de ácido nucleico artificial o el vector aquí descritos y/o para traducir el marco de lectura abierto de la molécula de ácido nucleico artificial o del vector aquí descritos. Por ejemplo, la célula puede usarse para la producción de proteínas recombinantes.

Las células aquí descritas se pueden obtener, por ejemplo, mediante métodos estándar de transferencia de ácido nucleico, tales como métodos de transfección, transducción o transformación estándar. Por ejemplo, la molécula de ácido nucleico artificial o el vector aquí descritos pueden transferirse a la célula mediante electroporación, lipofección, por ejemplo basada en lípidos catiónicos y/o liposomas, precipitación de fosfato de calcio, transfección basada en nanopartículas, transfección basada en virus o en polímeros catiónicos, tales como DEAE-dextrano o polietilenimina, etc.

Preferiblemente, la célula es una célula de mamífero, tal como una célula de un sujeto humano, un animal doméstico, un animal de laboratorio, tal como una célula de ratón o rata. Preferiblemente, la célula es una célula humana. La célula puede ser una célula de una línea celular establecida, como CHO, BHK, 293T, COS-7, HELA, HEK, etc. o la célula puede ser una célula primaria, como una célula de fibroblastos dérmicos humanos (HDF) etc., preferiblemente una célula aislada de un organismo. En una realización preferente, la célula es una célula aislada de un sujeto mamífero, preferiblemente de un sujeto humano. Por ejemplo, la célula puede ser una célula inmune, tal como una célula dendrítica, una célula cancerosa o tumoral, o cualquier célula somática, etc., preferiblemente de un sujeto mamífero, en particular de un sujeto humano.

En otro aspecto, la presente invención proporciona una composición farmacéutica como se define en las reivindicaciones. Además se describe aquí una composición farmacéutica que comprende la molécula de ácido nucleico artificial como se describe aquí, el vector como se describe aquí o la célula como se describe aquí. Se puede usar la composición farmacéutica aquí descrita, por ejemplo, como una vacuna, por ejemplo para la vacunación genética. Así, el ORF puede, por ejemplo, codificar un antígeno para ser administrado a un paciente para la vacunación. Por tanto, en una realización preferente, la composición farmacéutica aquí descrita es una vacuna. Además, la composición farmacéutica aquí descrita puede usarse, por ejemplo, para la terapia génica.

Preferentemente, la composición farmacéutica comprende además uno o más vehículos, diluyentes y/o excipientes farmacéuticamente aceptables y/o uno o más adyuvantes. En el contexto de la presente descripción, un vehículo farmacéuticamente aceptable incluye típicamente una base líquida o no líquida para la composición farmacéutica aquí descrita. En una realización, la composición farmacéutica se proporciona en forma líquida. En este contexto, preferentemente el vehículo se basa en agua, tal como agua libre de pirógenos, solución salina isotónica o soluciones tampón (acuosas), por ejemplo soluciones tampón fosfato, citrato, etc. La solución tampón puede ser hipertónica, isotónica o hipotónica en relación al medio de referencia específico, es decir la solución tampón puede tener un contenido de sales más alto, idéntico o más bajo en relación al medio de referencia específico, donde preferentemente se emplean concentraciones de las sales mencionadas que no conducen al daño de las células del mamífero por ósmosis u otros efectos de la concentración. Medios de referencia son, por ejemplo, líquidos que se emplean en métodos "*in vivo*", tales como líquidos sanguíneos, linfoides, citosólicos u otros líquidos corporales, por ejemplo, líquidos que se pueden utilizar como medios de referencia en métodos "*in vitro*", tales como soluciones tampón habituales o líquidos. Tales soluciones tampón o líquidos comunes son conocidos del experto. La solución de lactato de Ringer es particularmente preferente como base líquida.

Se pueden utilizar una o más cargas o sólidas o líquidas o diluyentes o compuestos encapsulantes adecuados para la administración a un paciente así como para la composición farmacéutica aquí descrita. El término "compatible" como se utiliza aquí significa preferentemente que estos componentes de la composición farmacéutica son capaces de mezclarse con el ácido nucleico artificial, con el vector o con las células como se definen aquí de manera que no aparece una interacción que reduce la efectividad farmacéutica de la composición farmacéutica bajo las condiciones de uso típicas.

La composición farmacéutica aquí descrita puede comprender opcionalmente además uno o más componentes farmacéuticamente activos adicionales. Un componente farmacéuticamente activo en este contexto es un compuesto que tiene un efecto terapéutico para sanar, mejorar o prevenir una indicación o enfermedad particular. Tales compuestos incluyen, sin limitación, péptidos o proteínas, ácidos nucleicos, compuestos orgánicos o inorgánicos de bajo peso molecular (terapéuticamente activos) (peso molecular inferior a 5.000, de manera preferente inferior a 1.000), azucares, antígenos o anticuerpos, agentes terapéuticos ya conocidos en la técnica anterior, células antigénicas, fragmentos celulares antigénicos, fracciones celulares, componentes de pared celular (por ejemplo polisacáridos), patógenos modificados, atenuados o desactivados (por ejemplo químicamente o por irradiación) (virus, bacterias, etc.).

Además, la composición farmacéutica aquí descrita puede comprender un portador para la molécula de ácido nucleico artificial o el vector. Tal portador puede ser adecuado para facilitar la disolución en líquidos fisiológicos

aceptables, para el transporte y la captación celular activa de la molécula de ácido nucleico artificial farmacéutica o del vector. Por consiguiente, tal portador puede ser un componente que puede ser adecuado para el almacenamiento y la administración de una molécula de ácido nucleico artificial o un vector aquí descritos. Tales componentes pueden ser, por ejemplo, portadores o compuestos catiónicos o policatiónicos que pueden servir como agentes de transfección o de complejación.

Agentes de transfección o complejación particularmente preferentes en este contexto son compuestos catiónicos o policatiónicos, incluyendo protamina, nucleolina, espermina o espermidina, u otros péptidos o proteínas catiónicas, como poli-L-lisina (PLL), poli-arginina, poli-péptidos básicos, péptidos de penetración de células (CPPs), incluyendo péptidos de unión a VIH, VIH-1 Tat (VIH), péptidos derivados de Tat, Penetratina, péptidos derivados o análogos de VP22, HSV VP22 (Herpes simple), MAP, KALA o dominios de transducción de proteínas (PTDs), PpT620, péptidos ricos en prolina, péptidos ricos en arginina, péptidos ricos en lisina, MPG-péptido(s), Pep-1, L-oligómeros, péptido(s) de calcitonina, péptidos derivados de Antennapedia (particularmente de Drosophila antennapedia), pAntp, pIsl, FGF, Lactoferrina, Transportano, Buforina-2, Bac715-24, SynB, SynB(1), pVEC, péptidos derivados de hCT, SAP, o histonas.

Adicionalmente, tales compuestos catiónicos o policatiónicos o portadores pueden ser péptidos o proteínas catiónicas o policatiónicas, que preferentemente comprenden o están modificados adicionalmente para comprender al menos una parte –SH. De manera preferente, un portador catiónico o policatiónico se selecciona de péptidos catiónicos que tienen la siguiente fórmula suma (I):

$$\{(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x\};\ \text{fórmula (I)}$$

donde l + m + n + o + x = 3-100, e l, m, n u o independientemente entre sí es cualquier número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90 y 91-100 con la condición de que el contenido total de Arg (Arginina), Lis (Lisina), His (Histidina) y Orn (Ornitina) representa al menos el 10% de todos los aminoácidos del oligopéptido; y Xaa es cualquier aminoácido seleccionado de aminoácidos nativos (= de origen natural) o no nativos excepto Arg, Lis, His u Orn; y x es cualquier número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, con la condición de que el contenido total de Xaa no exceda el 90% de todos los aminoácidos del oligopéptido. Cualquiera de los aminoácidos Arg, Lis, His, Orn y Xaa se puede situar en cualquier lugar del péptido. En este contexto los péptidos o proteínas catiónicas en un intervalo de 7-30 aminoácidos son particularmente preferidos.

Adicionalmente, el péptido o proteína catiónica o policatiónica, cuando se define de acuerdo con la formula $\{(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x\}$ (formula (I)) como se muestra anteriormente, y que comprende o está modificado adicionalmente para comprender al menos una parte –SH, puede seleccionarse, sin restringirse a, de la fórmula (Ia):

$$\{(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa')_x\,(Cys)_y\}\quad \text{subfórmula (Ia)}$$

donde $(Arg)_l;(Lys)_m;(His)_n;(Orn)_o$; y x son como se definen aquí, Xaa' es cualquier aminoácido seleccionado de aminoácidos nativos (= de origen natural) o no nativos excepto Arg, Lis, His, Orn o Cys e y es cualquier número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80 y 81-90, con la condición de que el contenido total de Arg (Arginina), Lis (Lisina), His (Histidina) y Orn (Ornitina) representa al menos un 10% de todos los aminoácidos del oligopéptido. Adicionalmente, el péptido catiónico o policatiónico se puede seleccionar de la fórmula (Ib):

$$Cys_1\{(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x\}Cys_2\ \ \text{subfórmula (Ib)}$$

donde la fórmula empírica $\{(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x\}$ (fórmula (III)) es como se define aquí y forma un núcleo de una secuencia de aminoácidos de acuerdo con la fórmula (semi empírica) (III) y donde Cys_1 y Cys_2 son Cisteínas próximas a, o terminales a $(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x$.

Compuestos catiónicos o policatiónicos preferidos adicionales que se pueden utilizar como agentes de transfección o formación de complejos pueden incluir polisacáridos catiónicos, por ejemplo quitosano, polibreno, polímeros catiónicos, por ejemplo polietilenimina (PEI), lípidos catiónicos, por ejemplo DOTMA: cloruro de [1-(2,3-sioleiloxi)propil)]-N,N,N-trimetilamonio, DMRIE, di-C14-amidina, DOTIM, SAINT, DC-Chol, BGTC, CTAP, DOPC, DODAP, DOPE: dioleil fosfatidiletanol-amina, DOSPA, DODAB, DOIC, DMEPC, DOGS: dioctadecilamidoglicilespermina, DIMRI: bromuro de dimiristo-oxipropildimetilhidroxietil-amonio, DOTAP: dioleoiloxi-3-(trimetilamonio)propano, DC-6-14: cloruro de O,O-ditetradecanoil-N-(α trimetilamonioacetil)dietanolamina, CLIP1: cloruro de rac-[(2,3-dioctadecil-oxipropil)(2-

hidroxietil)]dimetilamonio, CLIP6: rac-[2(2,3-dihexadeciloxipropil-oximetiloxi)etil]trimetilamonio, CLIP9: rac-[2(2,3-dihexadeciloxipropil-oxisuccinil-oxi)etil]-trimetilamonio, oligofectamina, o polímeros catiónicos o policatiónicos, por ejemplo poliaminoácidos modificados, tales como polímeros β-aminoácido o poliamidas inversas, etc., polietilenos modificados, tales como PVP, bromuro de (poli(N-etil-4-vinilpiridinio)), etc., acrilatos modificados, tales como pDMAEMA (poli(dimetilaminoetil metilacrilato)), etc., amidoaminas modificadas tales como pAMAM (poli(amidoamina)), etc., polibetaaminoéster modificado (PBAE), tal como polímeros de 1,4-butanodiol, diacrilato-co-5-amino-1-pentanol, modificados en el extremo diamina, etc., dendrímeros, tales como dendrímeros de polipropilamina o dendrímeros basados en pAMAM, etc., poliimina(s), como PEI: poli(etilenimina), poli(propilenimina), etc., polialilamina, polímeros basados en cadena principal de azúcar, tales como polímeros basados en ciclodextrina, polímeros basados en dextrano, quitosano, etc., polímeros basados en cadena principal de silano, tales como copolímeros PMOXA-PDMS, etc., polímeros en bloque que consisten en una combinación de uno o más bloques catiónicos (por ejemplo seleccionados de un polímero catiónico como se menciona anteriormente) y uno o más bloques hidrofílicos o hidrofóbicos (por ejemplo polietilenglicol); etc.

De acuerdo con una realización, la composición farmacéutica aquí descrita puede comprender un adyuvante para mejorar las propiedades inmunoestimuladoras de la composición farmacéutica. En este contexto, un adyuvante puede entenderse como cualquier compuesto adecuado para soportar la administración y el suministro de los componentes tales como la molécula de ácido nucleico artificial o el vector comprendido en la composición farmacéutica de acuerdo con la invención. Además, dicho adyuvante puede, sin estar unido a él, iniciar o aumentar una respuesta inmune del sistema inmune innato, es decir, una respuesta inmune no específica. En otras palabras, cuando se administra, la composición farmacéutica aquí descrita típicamente inicia una respuesta inmune adaptativa dirigida al antígeno codificado por la molécula de ácido nucleico artificial. Además, la composición farmacéutica aquí descrita puede generar una respuesta inmune innata (de apoyo) debido a la adición de un adyuvante como se define aquí a la composición farmacéutica aquí descrita.

Dicho adyuvante puede seleccionarse de cualquier adyuvante conocido por el experto y adecuado para el presente caso, es decir, que respalde la inducción de una respuesta inmune en un mamífero. Preferiblemente, el adyuvante se selecciona del grupo que consiste en, sin limitarse a, TDM, MDP, dipéptido muramilo, plurónicos, solución alum, hidróxido de aluminio, ADJUMER™ (polifosfaceno); gel de fosfato de aluminio; glucanos de algas; algamulina; gel de hidróxido de aluminio (alum); gel de hidróxido de aluminio altamente adsorbente de proteína; gel de hidróxido de aluminio de baja viscosidad; AF o SPT (emulsión de esqualeno (5%), Tween 80 (0,2%), Pluronics L121 (1,25%), solución salina regulada con fosfato, pH 7,4); AVRIDINE™ (propanodiamina); BAY R1005™ ((hidroacetato de N-(2-desoxi-2-L-leucilamino-b-D-glucopiranosil)-N-octadecildodecanoil-amida); CALCITRIOL™ (1-alfa-25-dihidroxivitamina D3); gel de fosfato de calcio; CAPTM (nanopartículas de fosfato de calcio); holotoxina de cólera, proteína de fusión de cólera-toxina-A1-proteína-A-D-fragmento, subunidad B de toxina de cólera; CRL 1005 (copolímero de bloqueo P1205); liposomas que contienen citocina; DDA (bromuro de dimetildioctadecilamonio); DHEA (dehidroespiroesterona); DMPC (dimiristoilfosfatidilcolina); DMPG (dimiristoilfosfatidilglicerol); complejo DOC/alum (sal de sodio de ácido desoxicólico); adyuvante completo de Freund; adyuvante incompleto de Freund; gamma-inulina; adyuvante de Gerbu (mezcla de: i) N-acetilglucosaminil-(P1-4)-N- acetilmuramil-L-alanil-D-glutamina (GMDP), ii) cloruro de dimetildioctadecilamonio (DDA), iii) complejo de sal de zinc-L-prolina (ZnPro-8); GM-CSF); GMDP (N-acetilglucosaminil-(b1-4)-N-acetilmuramil-L-alanil-D-isoglutamina); imiquimod (1-2-metipropil)-1H-imidazo[4,5-c]quinolin-4-amina); ImmTher™ (dipalmitato de N-acetilglucosaminil-N-acetilmuramil-L-Ala-D-isoGlu-L-Ala-glicerol); DRV (inmunoliposomas preparados por deshidratación-rehidratación de vesículas); interferon-gamma; interleucina-1-beta; interleucina-2; interleucina-7; interleucina-12; ISCOMS™; ISCOPREP 7.0.3.™; liposomas; LOXORIBINE™ (7-alil-8-oxoguanosina); adyuvante oral LT (Ecoli lábil enterotoxina-protoxina); microesferas y micropartículas de cualquier composición; MF59™; (emulsión escualeno-agua); MONTANIDE ISA 51™ (adyuvante de Freund incompleto purificado); MONTANIDE ISA 720™ (adyuvante de aceite metabolizable); MPL™ (3-Q-desacil-4'-monofosforil lípido A); liposomas MTP-PE y MTP-PE ((N-acetil-L-alanil-D-isoglutaminil-L-alanin-2-(sal monosódica de 1,2-dipalmitoil-sn-glicero-3-(hidroxifosforiloxi))etilamida); MURAMETIDE™ (Nac-Mur-L-Ala-D-GIn-OCH3); MURAPALMITINE™ y D-MURAPALMITINE™ (Nac-Mur-L-Thr-D-isoGIn-sn-gliceroldipalmitoilo); NAGO (neuraminidasa-galactosa oxidasa); nanoesferas o nanopartículas de cualquier composición; NISV (vesículas surfactantes no iónicas); PLEURAN™ (β-glucano); PLGA, PGA y PLA (homo- y co-polímero de ácido láctico y ácido glicólico; microesferas/nanoesferas); PLURONIC L121™; PMMA (metacrilato de polimetilo); PORDDS™ (microesferas proteinoides); derivados de carbamato de polietileno; poli-rA: poli-rU (complejo de ácido poliadenílico-ácido poliuridílico); polisorbato 80 (Tween 80); cocleatos de proteína (Avanti Porlar Lipids, Inc., Alabaster, AL); STIMULON™ (QS-21); Quil-A (Quil-A saporin); S-28463 (4-amino-otec-dimetil-2-etoximetil-1H-imidazo[4,5-c]quinolin-1-etanol); SAF-1™ ("Formulación de adyuvante Syntex"); proteoliposomas Sendai y matrices de lípidos que contienen Sendai; Span-85 (trioleato de sorbitano); Specol (emulsión de Marcol 52, Span 85 y Tween 85); escualeno o Robane® (2,6,10,15,19,23-hexametiltetracosan y 2,6,10,15,19,23-hexametil-2,6,10,14,18,22-tetracosahexano); estearíltirosina (clorhidrato de octadeciltirosina); Theramid® (N-acetilglucosaminil-N-acetilmuramil-L-Ala-D-isoGlu-L-Ala-dipalmitoxipropilamida); Teronil-MDP (Termurtide™ o [thr l]-MDP; N-acetilmuramil-L-treonil-D-isoglutamina);

partículas Ty (Ty-VLPs o partículas de tipo virus); liposomas Walter-Reed (liposomas que contienen lípidos A adsorbidos en hidróxido de aluminio) y lipopéptidos, incluyendo Pam3Cys, en particular sales de aluminio, tales como Adju-phos, Alhydrogel, Rehydragel; emulsiones, incluyendo CFA, SAF, IFA, MF59, Provax, TiterMax, Montanide, Vaxfectin; copolímeros que incluyen Optivax (CRL1005), LI 21, Poloaxmer4010), etc.; liposomas, incluyendo Stealth, cocleatos, incluyendo BIORAL; adyuvantes derivados de plantas, incluyendo QS21, Quil A, Iscomatrix, ISCOM; adyuvantes adecuados para la coestimulación que incluyen Tomatino, biopolímeros, incluyendo PLG, PMM, Inulina; adyuvantes derivados de microbios, incluyendo Romurtide, DETOX, MPL, CWS, Manosa, secuencias de ácido nucleico CpG, CpG7909, ligandos de TLR 1 -10 humano, ligandos de murina TLR 1 -13, ISS-1018, IC31, imidazoquinolinas, Ampligen, Ribi529, IMOxine, IRIV, VLP, toxina de cólera, toxina termo-lábil, Pam3Cys, Flagelina, ancla de GPI, LNFPIII/Lewis X, péptidos antimicrobianos, UC-1V150, proteína de fusión RSV, cdiGMP; y adyuvantes adecuados como antagonistas que incluyen neuropéptidos CGRP.

También pueden seleccionarse adyuvantes adecuados de compuestos catiónicos o policatiónicos donde el adyuvante se prepara preferiblemente después de la complejación de la molécula de ácido nucleico artificial o el vector de la composición farmacéutica con el compuesto catiónico o policatiónico. La asociación o complejación de la molécula de ácido nucleico artificial o del vector de la composición farmacéutica con compuestos catiónicos o policatiónicos como se definen aquí preferiblemente provee de propiedades adyuvantes y confiere un efecto estabilizador a la molécula de ácido nucleico artificial o al vector de la composición farmacéutica. Son particularmente preferentes los compuestos catiónicos o policatiónicos que se seleccionan de péptidos o proteínas catiónicos o policatiónicos que incluyen protamina, nucleolina, espermina o espermidina u otros péptidos o proteínas catiónicos, tales como poli-L-lisina (PLL), poli-arginina, polipéptidos básicos, péptidos penetradores de células (CPP), incluyendo péptidos de unión a VIH, Tat, VIH-1 Tat (VIH), péptidos derivados de Tat, Penetratina, péptidos VP22 derivados o análogos, HSV VP22 (Herpes simple), MAP, KALA o dominios de transducción de proteína (PTD, PpT620, péptidos ricos en prolina, péptidos ricos en arginina, péptidos ricos en lisina, MPG-péptidos, Pep-1, L-oligómeros, péptido(s) de calcitonina, péptidos derivados de Antenapedia (particularmente de Drosophila antennapedia), pAntp, plsl, FGF, Lactoferrina, Transportano, Buforin-2, Bac715-24, SynB, SynB(1), pVEC, péptidos derivados de hCT, SAP, protamina, espermina, espermidina o histonas. Otros compuestos catiónicos o policatiónicos preferidos pueden incluir polisacáridos catiónicos, por ejemplo quitosano, polibreno, polímeros catiónicos, por ejemplo polietilenimina (PEI), lípidos catiónicos, por ejemplo DOTMA: cloruro de [1 -(2,3-sioleiloxi)propil)]-N,N,N-trimetilamonio, DMRIE, di-C14-amidina, DOTIM, SAINT, DC-Choi, BGTC, CTAP, DOPC, DODAP, DOPE: Dioleil fosfatidiletanolamina, DOSPA, DODAB, DOIC, DMEPC, DOGS: Dioctadecilamidoglicilspermina, DIMRI: Bromuro de dimiristooxipropil-dimetil-hidroxietil-amonio, DOTAP: dioleoiloxi-3- (trimetilamonio)-propano, DC-6-14: cloruro de O,O-ditetradecanoil-N-(α- trimetilamonioacetil)dietanolamina, CLIP1: cloruro de rac-[(2,3-dioctadeciloxipropil)-(2-hidroxietil)]-dimetilamonio, CLIP6: rac-[2(2,3-dihexadeciloxipropil-oximetiloxi)etil]trimetilamonio, CLIP9: rac-[2(2,3-dihexadeciloxipropil-oxisucciniloxijetil-trimetilamonio, oligofectamina, o polímeros catiónicos o policatiónicos, por ejemplo poliaminoácidos modificados, tal como polímeros de β-aminoácidos o poliamidas inversas, polietilenos modificados, como PVP (bromuro de poli(N-etil-4-vinilpiridinio)), acrilatos modificados como pDMAEMA (metilacrilato de polidimetilaminoetilo)), amidoaminas modificadas como pAMAM (poli(amidoamina)), polibetaaminoéster (PBAE) modificado, como polímeros de diamina con extremo modificado con 1,4-butanodiol diacrilato-co-5-amino-1-pentanol, dendrímeros como dendrímeros de polipropilamina o dendrímeros a base de pAMAM, poliimina(s), como PEI: poli(etilenimina), poli(propilenimina), polialilamina, polímeros a base de estructuras de azúcar, tales como polímeros a base de ciclodextrina, polímeros a base de dextrano, quitosano, polímeros a base de estructuras silano, como copolímeros de PMOXA-PDMS, polímeros en bloque que consisten en una combinación de uno o más bloques catiónicos (por ejemplo seleccionados de un polímero catiónico como los que se mencionan arriba) y uno o más bloques hidrófilos o hidrófobos (por ejemplo polietilenglicol); etc.

De forma adicional, las proteínas o péptidos catiónicos o policatiónicos preferidos que pueden usarse como adyuvante por complejación con la molécula de ácido nucleico artificial o el vector, preferentemente un ARN, de la composición puede seleccionarse de las siguientes proteínas o péptidos que tienen la siguiente fórmula general (I): $(Arg)_l;(Lys)_m;(His)_n;(Orn)_o;(Xaa)_x$, en la que l + m + n +o + x = 8-15, y l, m, n u o, independientemente entre sí, pueden ser un número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15, siempre y cuando el contenido total de Arg, Lys, His y Orn represente al menos un 50% de todos los aminoácidos del oligopéptido; y Xaa puede ser cualquier aminoácido seleccionado de aminoácidos nativos (= que ocurren naturalmente) o no nativos excepto Arg, Lys, His u Orn; y x puede ser cualquier número seleccionado de 0, 1, 2, 3 o 4, siempre y cuando el contenido total de Xaa no exceda el 50% de todos los aminoácidos del oligopéptido. Oligoargininas particularmente preferidas en este contexto son por ejemplo Arg7, Arg8, Arg9, Arg7, H3R9, R9H3, H3R9H3, YSSR9SSY, $(RKH)_4$, $Y(RKH)_2R$, etc.

En una realización preferente, la molécula de ácido nucleico artificial, preferiblemente una molécula de ARN, está asociada o complejada con un compuesto catiónico o policatiónico o un vehículo polimérico, opcionalmente en una relación en peso seleccionada de un rango de aproximadamente 6:1 (p/p) a

aproximadamente 0,25:1 (p/p), más preferiblemente de aproximadamente 5:1 (p/p) a aproximadamente 0,5:1 (p/p), incluso más preferiblemente de aproximadamente 4:1 (p/p) a aproximadamente 1:1 (p/p) o de aproximadamente 3:1 (p/p) a aproximadamente 1:1 (p/p), y con total preferencia una relación de aproximadamente 3:1 (p/p) a aproximadamente 2:1 (p / p) ARN:compuesto catiónico o policatiónico y/o vehículo polimérico; u opcionalmente en una relación nitrógeno/fosfato de ARN:compuesto catiónico o policatiónico y/o vehículo polimérico en el intervalo de aproximadamente 0,1-10, preferiblemente en un intervalo de aproximadamente 0,3-4 o 0,3-1 y con total preferencia en un intervalo de aproximadamente 0,5-1 o 0,7-1, e incluso con mayor preferencia en un rango de aproximadamente 0,3-0,9 o 0,5-0,9.

La relación entre el ácido nucleico artificial o el vector y el compuesto catiónico o policatiónico se puede calcular en base a la relación nitrógeno/fosfato (relación N/P) de todo el complejo de ácido nucleico. Por ejemplo, 1 µg de ARN típicamente contiene aproximadamente 3 nmol de residuos fosfato, siempre que el ARN muestre una distribución estadística de bases. Además, 1 µg de péptido típicamente contiene aproximadamente x nmol de residuos nitrógeno, dependiendo del peso molecular y la cantidad de aminoácidos básicos. Cuando se calcula a modo de ejemplo para (Arg)9 (peso molecular 1424 g/mol, 9 átomos de nitrógeno), 1 µg de (Arg)9 contiene aproximadamente 700 pmol de (Arg)9 y, por tanto, 700 x 9 = 6300 pmol de aminoácidos básicos = 6,3 nmol de átomos de nitrógeno Para una relación en masa de aproximadamente 1:1 ARN/(Arg)9, se puede calcular una relación N/P de aproximadamente 2. Cuando se calcula a modo de ejemplo para protamina (peso molecular de aproximadamente 4250 g/mol, 21 átomos de nitrógeno, cuando se usa protamina de salmón) con una relación en masa de aproximadamente 2:1 con 2 µg de ARN, se deben calcular 6 nmol de fosfato para el ARN; 1 µg de protamina contiene aproximadamente 235 pmol de moléculas de protamina y, por tanto, 235 x 21 = 4935 pmol de átomos de nitrógeno básicos = 4,9 nmol de átomos de nitrógeno. Para una relación en masa de aproximadamente 2:1 ARN/protamina, se puede calcular una relación N/P de aproximadamente 0,81. Para una relación en masa de aproximadamente 8:1 ARN/protamina, se puede calcular una relación N/P de aproximadamente 0,2. En el contexto de la presente invención, una relación N/P está preferiblemente en el intervalo de aproximadamente 0,1-10, preferiblemente en un intervalo de aproximadamente 0,3-4 y con mayor preferencia en un intervalo de aproximadamente 0,5-2 o 0,7-2 con respecto a la relación ácido nucleico:péptido en el complejo, y con total preferencia en el rango de aproximadamente 0,7-1,5

La solicitud de patente WO2010/037539 describe una composición inmunoestimuladora y métodos para la preparación de una composición inmunoestimuladora. Así, en una realización preferida de la invención, la composición se obtiene en dos etapas separadas para obtener tanto un efecto inmunoestimulador eficaz como una traducción eficiente de la molécula de ácido nucleico artificial de acuerdo con la invención. En este documento, un llamado “componente adyuvante" se prepara complejando, en una primera etapa, la molécula o el vector de ácido nucleico artificial, preferiblemente un ARN, del componente adyuvante con un compuesto catiónico o policatiónico en una relación específica para formar un complejo estable. En este contexto, es importante que no quede compuesto catiónico o policatiónico libre o solo una pequeña cantidad despreciable en el componente adyuvante después de complejar el ácido nucleico. Por consiguiente, la proporción entre el ácido nucleico y el compuesto catiónico o policatiónico en el componente adyuvante se selecciona típicamente en un intervalo en el que el ácido nucleico está completamente complejado y no hay compuesto catiónico o policatiónico libre o solo queda una pequeña cantidad necesaria en la composición. Preferiblemente, la relación del componente adyuvante, es decir, la relación entre el ácido nucleico y el compuesto catiónico o policatiónico se selecciona de un intervalo de aproximadamente 6:1 (p/p) a aproximadamente 0,25:1 (p/p), más preferiblemente de aproximadamente 5:1 (p/p) a aproximadamente 0,5:1 (p/p), incluso más preferiblemente de aproximadamente 4:1 (p/p) a aproximadamente 1:1 (p/p) o de aproximadamente 3:1 (p/p) a aproximadamente 1:1 (p/p), y con total preferencia una relación de aproximadamente 3:1 (p/p) a aproximadamente 2:1 (p/p).

De acuerdo con una realización preferente, la molécula de ácido nucleico artificial o el vector, preferiblemente una molécula de ARN, como se describe aquí, se agrega en un segundo paso a la molécula de ácido nucleico complejado, preferiblemente un ARN, del componente adyuvante para formar la composición (inmunoestimuladora) aquí descrita. En ella, la molécula o el vector de ácido artificial, preferiblemente un ARN, como se describen aquí se agrega como ácido nucleico libre, es decir, ácido nucleico no complejado con otros compuestos. Antes de la adición, la molécula de ácido nucleico artificial libre o el vector no forman complejos y preferiblemente no experimentarán ninguna reacción de complejación detectable o significativa tras la adición del componente adyuvante.

Adyuvantes adecuados pueden seleccionarse además de ácidos nucleicos de fórmula (II): $G_lX_mG_n$, donde: G es guanosina, uracilo o un análogo de guanosina o uracilo; X es guanosina, uracilo, adenosina, timidina, citosina o un análogo de los nucleótidos mencionados; l es un número entero de 1 a 40, donde cuando l = 1 G es guanosina o un análogo de la misma, cuando l > 1 al menos el 50% de los nucleótidos son guanosina o análogos de la misma; m es un número entero y es al menos 3; donde cuando m = 3 X es uracilo o un análogo del mismo, cuando m > 3 existen al menos 3 uracilos o análogos de uracilo sucesivos; n es un número entero de 1 a 40, donde cuando n = 1 G es guanosina o un análogo de la misma, cuando n > 1 al menos el 50% de

los nucleótidos son guanosina o análogos de la misma.

Otros adyuvantes adecuados pueden seleccionarse además de ácidos nucleicos de fórmula (III): $C_lX_mC_n$, donde C es citosina, uracilo o un análogo de citosina o uracilo; X es guanosina, uracilo, adenosina, timidina, citosina o un análogo de los nucleótidos mencionados; l es un número entero de 1 a 40, donde cuando l = 1 C es citosina o un análogo de la misma, cuando l > 1 al menos el 50% de los nucleótidos son citosina o análogos de la misma; m es un número entero y es al menos 3; donde cuando m = 3 X es uracilo o un análogo del mismo, cuando m > 3 existen al menos 3 uracilos o análogos de uracilo sucesivos; n es un número entero de 1 a 40, donde cuando n = 1 C es citosina o un análogo de la misma, cuando n > 1 al menos el 50% de los nucleótidos son citosina o análogos de la misma.

Preferentemente, la composición farmacéutica aquí descrita comprende una “cantidad segura y efectiva” de los componentes de la composición farmacéutica, particularmente de la secuencia de ácido nucleico artificial aquí descrita, del vector y/o de las células como se definen aquí. Como se utiliza aquí, una “cantidad segura y efectiva” significa una cantidad suficiente para inducir significativamente una modificación positiva de una enfermedad o trastorno como se define aquí. Al mismo tiempo, sin embargo, una “cantidad segura y efectiva” preferentemente evita efectos secundarios serios y permite una relación sensible entre ventaja y riesgo. La determinación de estos límites está típicamente dentro del alcance del juicio médico sensible.

En otro aspecto, la presente descripción proporciona la molécula de ácido nucleico artificial como la aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita para su uso como un medicamento, por ejemplo como una vacuna (en la vacunación genética) o en la terapia génica.

La molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita son particularmente adecuados para cualquier aplicación medica que hace uso de la acción terapéutica o efecto de los péptidos, polipéptidos o proteínas, o donde es necesaria la suplementación de un péptido o proteína particular. Así, la presente descripción proporciona la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita para su uso en el tratamiento o la prevención de enfermedades o trastornos sensibles al tratamiento por la acción terapéutica o efecto de los péptidos, polipéptidos o proteínas sensibles al tratamiento mediante suplementación de un péptido, polipéptido o proteína particular. Por ejemplo, la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita se pueden utilizar para el tratamiento o la prevención de enfermedades genéticas, enfermedades autoinmunes, enfermedades cancerosas o relacionadas con tumores, enfermedades infecciosas, enfermedades crónicas o similares, por ejemplo, por vacunación genética o terapia génica.

En particular, tales tratamientos terapéuticos que se benefician de una presencia estable, prolongada y/o incrementada de péptidos terapéuticos, polipéptidos o proteínas en un sujeto a tratar son especialmente adecuados como aplicación medica en el contexto de la presente descripción, ya que la molécula de ácido nucleico artificial aquí descrita proporciona una expresión estable, aumentada y/o prolongada del péptido o de la proteína codificada por la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita en la vacunación. Así, la presente descripción proporciona la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita es la vacunación. De esta manera, la presente invención proporciona la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita para la vacunación de un sujeto, preferentemente un sujeto mamífero, en especial un sujeto humano. Los tratamientos de vacunación preferentes son la vacunación contra enfermedades infecciosas, como infecciones bacterianas, protozoarias o virales, y la vacunación anti-tumoral. Tales tratamientos de vacunación pueden ser profilácticos o terapéuticos.

El ORF se puede seleccionar dependiendo de la enfermedad a tratar o prevenir. Por ejemplo, el marco de lectura abierto puede codificar una proteína que tiene que ser suministrada a un paciente que sufre de carencia total o al menos pérdida parcial de la función de una proteína, tal como un paciente que sufre de una enfermedad genética. Adicionalmente, el marco de lectura abierto se puede elegir de un ORF que codifica un péptido o proteína que influye beneficiosamente en una enfermedad o en la condición de un sujeto. Adicionalmente, el marco de lectura abierto puede codificar un péptido o proteína que afecta la sub-regulación de una sobreproducción patológica de un péptido o proteína natural o por la eliminación de células que expresan patológicamente una proteína o péptido. Tal carencia, pérdida de función o sobreproducción pueden aparecer, por ejemplo, en el contexto de tumores y neoplasias, enfermedades autoinmunes, alergias, infecciones, enfermedades crónicas o similares. Adicionalmente, el marco de lectura abierto puede codificar para un antígeno o inmunógeno. Así, en realizaciones preferentes, la molécula de ácido nucleico artificial o el vector aquí descritos comprenden un ORF que codifica una secuencia de aminoácidos que comprende o consiste en un antígeno o inmunógeno, por ejemplo un epítopo de un patógeno o un antígeno asociado a tumor,

y una 3’-UTR como se describe anteriormente.

En el contexto de la aplicación médica, en particular en el contexto de la vacunación, es preferente que la molécula de ácido nucleico artificial aquí descrita sea ARN, en especial ARNm, ya que el ADN conlleva el riesgo de inducir una respuesta inmunitaria anti-ADN y tiende a insertarse en el ADN genómico. Sin embargo, en algunas realizaciones, por ejemplo si se utiliza un vehículo de suministro viral, tal como un vehículo adenoviral, para la administración de la molécula de ácido nucleico artificial o del vector aquí descritos, por ejemplo en el contexto de los tratamientos terapéuticos génicos, puede ser deseable que la molécula de ácido nucleico artificial o el vector sea una molécula de ADN.

La molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita se administran vía oral, parenteral, por rociado en inhalación, vía tópica, rectal, nasal, bucal, vaginal o mediante un depósito implantado o inyección con pistola. El término parenteral como se utiliza aquí incluye subcutáneo, intravenoso, intramuscular, intra-articular, intra-sinovial, intraesternal, intratecal, intrahepática, intralesional, intracraneal, transdérmica, intradérmica, intrapulmonar, intraperitoneal, intracardiaca, intraarterial, e inyección sublingual o técnicas de infusión. En una realización preferente, la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita se administran mediante inyección sin aguja (por ejemplo, inyección con pistola).

Preferentemente, la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita se administran parenteralmente, por ejemplo, por inyección parenteral, en especial por inyección subcutánea, intravenosa, intramuscular, intra-articular, intra-sinovial, intraesternal, intratecal, intrahepática, intralesional, intracraneal, transdérmica, intradérmica, intrapulmonar, intraperitoneal, intracardiaca, intraarterial, inyección sublingual o por técnicas de infusión. Es particularmente preferente la inyección intradérmica e intramuscular. Las formas inyectables estériles de la composición farmacéutica aquí descrita puede ser una suspensión acuosa u oleosa. Estas suspensiones se pueden formular según técnicas conocidas en la técnica utilizando agentes de dispersión o humectantes y agentes de suspensión adecuados. Preferiblemente, las soluciones o suspensiones se administran mediante inyección sin aguja (por ejemplo por inyección con pistola).

La molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita también se pueden administrar oralmente en cualquier forma de dosificación oralmente aceptable incluyendo, pero sin limitarse a, cápsulas, comprimidos, suspensiones o soluciones acuosas.

La molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita también se pueden administrar tópicamente, especialmente cuando el objetivo del tratamiento incluye áreas u órganos fácilmente accesibles por la aplicación tópica, por ejemplo incluyendo enfermedades de la piel o de cualquier otro tejido epitelial accesible. Las formulaciones tópicas adecuadas se preparan fácilmente para cada una de estas áreas u órganos. Para aplicaciones tópicas, la molécula de ácido nucleico artificial aquí descrita, el vector aquí descrito, la célula aquí descrita o la composición farmacéutica aquí descrita se pueden formular en una pomada adecuada suspendidos o disueltos en uno o más vehículos.

En una realización, el uso como un medicamento comprende la etapa de transfectar células de mamífero, preferentemente transfección *in vitro o ex vivo* de células de mamífero, en especial transfectar *in vitro* células aisladas de un sujeto a tratar por el medicamento. Si el uso comprende la transfección *in vitro* de células aisladas, el uso como medicamento puede comprender además la (re)administración de las células transfectadas al paciente. El uso de las moléculas de ácido nucleico artificiales o del vector aquí descritos como un medicamento puede comprender además la etapa de seleccionar las células aisladas exitosamente transfectadas. Así, puede ser beneficioso si el vector comprende además un marcador de selección. También, el uso como medicamento puede comprender la transfección *in vitro* de células aisladas y la purificación de un producto de expresión, es decir la proteína o péptido codificado a partir de estas células. Este péptido o proteína purificada se puede administrar subsecuentemente a un sujeto que lo necesite.

Según una realización preferente, la molécula de ácido nucleico aquí descrita, el vector aquí descrito, la célula aquí descrita o la vacuna o la composición farmacéutica aquí descrita se administran al sujeto vía intramuscular. Ventajosamente, la administración intramuscular de la molécula, vector, vacuna o composición de ácido nucleico aquí descritos da como resultado una mayor expresión del péptido o proteína codificada por el al menos un marco de lectura abierto de la molécula de ácido nucleico artificial como se define aquí. En una realización particularmente preferente, la expresión de la molécula de ácido nucleico artificial como se define aquí, preferiblemente de la molécula de ácido nucleico artificial que comprende una 3'-UTR que comprende al

menos una, más preferiblemente al menos dos secuencias poli(A) como se define aquí mejora cuando se administra vía intramuscular. Más preferiblemente, la administración intramuscular de la molécula de ácido nucleico artificial aquí descrita, preferiblemente de la molécula de ácido nucleico artificial que comprende al menos dos secuencias poli(A), resulta en una expresión aumentada y una respuesta inmune mejorada frente a un antígeno, preferiblemente un antígeno patógeno como se define aquí, más preferiblemente un antígeno asociado a una enfermedad infecciosa.

Como se describió anteriormente, la 3'-UTR de la invención es capaz de aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial. Por tanto, en un aspecto adicional, la presente invención se refiere a un método *in vitro* para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial, comprendiendo el método la etapa de asociar la molécula de ácido nucleico, preferiblemente la molécula de ARNm o el vector, con una región 3'-no traducida (3'-UTR), donde la 3'-UTR comprende al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, donde las al menos dos secuencias están separadas por una secuencia de ácido nucleico que comprende una secuencia tallo-bucle de histona según la SEQ ID NO: 11, y donde la molécula de ácido nucleico artificial está asociada o complejada con un compuesto catiónico o policatiónico. Además, se describe aquí un método *in vitro* para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial, comprendiendo el método la etapa de asociar la molécula de ácido nucleico, preferiblemente la molécula de ARNm o el vector, con una región 3'-no traducida (3'-UTR), donde la 3'-UTR comprende al menos una secuencia poli(A), donde la al menos una secuencia poli(A) comprende al menos 70 residuos adenosina, o una señal de poliadenilación. En una realización preferente, el método comprende asociar la molécula de ácido nucleico, preferiblemente el ARNm o el vector, con una 3'-UTR que comprende al menos dos secuencias poli(A) como se describe aquí.

El término "asociar la molécula de ácido nucleico artificial o el vector con una 3'-UTR" en el contexto de la presente descripción significa preferiblemente asociar funcionalmente o combinar funcionalmente la molécula de ácido nucleico artificial o el vector con la 3'-UTR. Esto significa que la molécula de ácido nucleico artificial o el vector y la 3'-UTR, preferiblemente la 3'-UTR como se describe anteriormente, están asociados o acoplados de manera que la función de la 3'-UTR, por ejemplo la función de aumentar la producción de proteínas, es ejercida. Típicamente, esto significa que la 3'-UTR está integrada en la molécula de ácido nucleico artificial o el vector, preferiblemente la molécula de ARNm, 3' a un marco de lectura abierto, preferiblemente inmediatamente 3' a un marco de lectura abierto, preferiblemente entre e marco de lectura abierto y una señal de poliadenilación. Preferentemente, la 3'-UTR está integrada en la molécula de ácido nucleico artificial o el vector, preferiblemente el ARNm, como 3'-UTR, es decir, de forma que la 3'-UTR es la 3'-UTR de la molécula de ácido nucleico artificial o del vector, preferiblemente del ARNm, es decir, de modo que se extiende desde el lado 3' del marco de lectura abierto hasta el extremo 5' de la molécula o hacia el lado 5' de una secuencia poli(A) o de una señal de poliadenilación, opcionalmente a través de un enlazador corto, tal como una secuencia que comprende o consiste en uno o más sitios de restricción. Así, preferentemente, el término "asociar la molécula de ácido nucleico artificial o el vector con una 3'-UTR" significa asociar funcionalmente la 3'-UTR con un marco de lectura abierto ubicado dentro de la molécula de ácido nucleico artificial o del vector, preferiblemente dentro de la molécula de ARNm. La 3'-UTR y el ORF son como se describieron anteriormente para la molécula de ácido nucleico artificial aquí descrita, por ejemplo preferiblemente el ORF y la 3'-UTR son heterólogos, por ejemplo derivados de diferentes genes, como se describió anteriormente.

En una realización preferida del método para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial, preferiblemente a partir de una molécula de ARNm o de un vector, la al menos una secuencia poli(A) se produce mediante una reacción de poliadenilación química o enzimática. Preferentemente, se emplea una enzima bacteriana, tal como poli(A) polimerasa de E. coli. En una realización preferente, la longitud de la al menos una secuencia poli(A) está regulada, entre otros parámetros, por la duración de la reacción de poliadenilación, es decir, una incubación más larga de un ácido nucleico con una enzima adecuada conduce típicamente a una secuencia poli(A) más larga. Preferiblemente, la molécula de ácido nucleico se incuba con una enzima adecuada durante al menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85 o 90 minutos, preferiblemente a una temperatura adecuada para la enzima respectiva, por ejemplo 37ºC. con mayor preferencia, la reacción de poliadenilación dura de aproximadamente 10 a aproximadamente 120 minutos, preferiblemente de aproximadamente 15, 20, 25 o 30 a aproximadamente 90 minutos, en especial de aproximadamente 30 a aproximadamente 60 minutos. Para obtener una población de moléculas de ácido nucleico que comparten aproximadamente el mismo grado de poliadenilación, es decir, que tienen aproximadamente el mismo número de adenilatos unidos a sus extremos 3', el experto puede elegir entre técnicas de separación estándar (por ejemplo, basadas en peso molecular o carga), tales como métodos cromatográficos, bien conocidos en la técnica y que se emplean típicamente después de la poliadenilación para separar o purificar los productos de reacción. Preferiblemente se usa una población de moléculas de ácido nucleico artificiales que es más o menos homogénea con respecto a la longitud de la secuencia poli(A) 3'-terminal. Preferentemente se usa una población de moléculas de ácido nucleico artificial donde al menos el 80%, más preferiblemente al menos el 85%, 90%, 95% o 98% de las moléculas se caracterizan por la misma

longitud de la secuencia 3' terminal poli(A). En este contexto, “la misma longitud” se refiere a una situación en la que el número de adenilatos en la secuencia poli(A) 3'-terminal varía de un valor dado (tal como 160 adenilatos, 380 adenilatos, 430 adenilatos, 1.000 adenilatos, etc.) en no más del 10%, preferiblemente en no más del 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% o no más del 1%.

En otro aspecto, la presente invención proporciona el uso de una 3'-UTR para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial donde la 3’-UTR comprende al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende una secuencia tallo-bucle de histona según la SEQ ID NO: 11 y donde la molécula de ácido nucleico artificial está asociada o complejada con un compuesto catiónico o policatiónico. Además, se describe aquí el uso de una 3’-UTR para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial como se describe aquí, preferiblemente a partir de una molécula de ARNm o de un vector, donde la 3'-UTR comprende al menos una secuencia poli(A), donde la al menos una secuencia poli(A) comprende al menos 70 residuos adenosina, o una señal de poliadenilación. En una realización preferente del uso aquí descrito, una 3'-UTR que comprende al menos dos secuencias poli(A), preferiblemente como se describe aquí, se usa para aumentar la producción de proteínas a partir de una molécula de ácido nucleico, preferiblemente a partir de una molécula de ácido nucleico artificial como se describe aquí.

Los compuestos e ingredientes de la composición farmacéutica aquí descrita también se pueden fabricar y comercializar por separado entre sí. Así, la invención se refiere también a un kit o kit de partes tal como se define en las reivindicaciones. La descripción proporciona además un kit o kit de partes que comprende una molécula de ácido nucleico artificial aquí descrita, un vector aquí descrito, una célula aquí descrita y/o una composición farmacéutica aquí descrita. Preferiblemente, dicho kit o kits de partes puede además comprender instrucciones de uso, células para la transfección, un adyuvante, un medio para la administración de la composición farmacéutica, un vehículo farmacéuticamente aceptable y/o una solución farmacéuticamente aceptable para la disolución o dilución de la molécula de ácido nucleico artificial, del vector, de las células o de la composición farmacéutica.

Breve descripción de las figuras

Las figuras mostradas a continuación son meramente ilustrativas y describen la presente invención de forma adicional. Estas figuras no se interpretarán para limitar la presente invención a las mismas.

Figura 1: secuencia de ADN (SEQ ID NO: 13) que codifica la secuencia de ARNm que se ha utilizado en los experimentos y que comprende las secuencias que codifican los siguientes elementos: rpl32 - PpLuc (GC) - albúmina7 - A64 - C30 – tallo-bucle de histona. Dentro de la secuencia de ADN se resaltan los elementos de secuencia correspondientes a los siguientes elementos en el ARNm: PpLuc (GC) (ORF) en cursiva, rpl32 (5'-UTR) subrayado y albúmina7 (3'-UTR) subrayado.

Figura 2: Poliadenilación de la secuencia de ARNm correspondiente a la SEQ ID NO: 13: A. Se agregaron aprox. 160 adenilatos a un lote de ARNm (Lote 1). Se agregaron aprox. 380 adenilatos a un lote diferente de ARNm (Lote 2). El ARNm correspondiente a la SEQ ID NO: 13 se cargó en la línea izquierda, el ARNm adenilado respectivo se cargó en la línea derecha. Se cargó un marcador de tamaño molecular en las líneas más externas para la comparación de tamaño (los números en la Fig. 2A indican el número de nucleótidos comprendidos en las moléculas marcadoras; se usó el mismo marcador en la Fig. 2B y Fig. 2C). B. Se añadieron aprox. 430 adenilatos al ARNm correspondiente a la SEQ ID NO: 13. C. Se agregaron aprox. 1.000 adenilatos al ARNm correspondiente a la SEQ ID NO: 13.

Figura 3: Expresión de proteínas a partir del ARNm poliadenilado en células cultivadas: se transfectó ARNm correspondiente a la SEQ ID NO: 13, al que se habían añadido aprox. 160 adenilatos (ARNm lote 1) o aprox. 380 adenilatos (ARNm lote 2) por poliadenilación, en fibroblastos dérmicos humanos (HDF) y se midieron los niveles de luciferasa en los puntos de tiempo indicados. La Figura 3 muestra los resultados como RLU media (unidades de luz relativas) ± SEM (error estándar) para transfecciones por triplicado.

Figura 4: Expresión de proteínas a partir del ARNm poliadenilado después de la inyección intramuscular en ratones: 2 µg de ARNm correspondiente a la SEQ ID NO: 13 a la que se habían añadido aprox. 380 adenilatos por poliadenilación se inyectaron intramuscularmente en ratones. La Figura 4 muestra los resultados como la mediana de hasta 10 replicados.

Figura 5: Expresión de proteínas a partir del ARNm poliadenilado después de la inyección intramuscular en ratones: 10 µg de ARNm correspondiente a la SEQ ID NO: 13 a la que se habían añadido aprox. 430 adenilatos por poliadenilación se inyectaron intramuscularmente en ratones. La Figura 5 muestra los resultados como la mediana de hasta 10 replicados.

Figura 6: Expresión de proteínas a partir del ARNm poliadenilado después de la inyección intramuscular en ratones: 1 µg de ARNm correspondiente a la SEQ ID NO: 13 a la que se habían añadido aprox. 1.000 adenilatos por poliadenilación se inyectaron intramuscularmente en ratones. La Figura 6 muestra los resultados como la mediana de hasta 10 replicados.

Figura 7: Inducción de anticuerpos neutralizantes de HA por ARNm poliadenilado después de la inyección intramuscular en ratones.

Figura 8: Secuencia de ADN (SEQ ID NO: 14) que codifica la secuencia de ARNm que se ha utilizado en los experimentos y que comprende las secuencias que codifican los siguientes elementos: 32L4-H1N1 (Países Bajos, 2009) -HA(GC) -albúmina7- A64-N5-C30-histonaSL

Figura 9: Inducción de títulos neutralizantes de virus por ARNm poliadenilado después de inyección intramuscular en ratones.

Figura 10: Secuencia de ADN (SEQ ID NO: 15) que codifica la secuencia de ARNm que se ha utilizado en los experimentos y que comprende las secuencias que codifican los siguientes elementos: 32L4-RAVG(GC)-albumin7-A64-N5-C30-histonaSL.

Ejemplos

Los ejemplos que se muestran a continuación son meramente ilustrativos y describen la presente invención de otra manera. Estos ejemplos no se interpretarán para limitar la presente invención a los mismos.

1 . Preparación de plantillas de ADN

Se construyó un vector para la transcripción *in vitro* que contenía un promotor T7 seguido por una secuencia enriquecida en GC que codifica la luciferasa de *Photinus pyralis* (PpLuc(GC)).. La región no traducida 5' (5'-UTR) de la proteína ribosómica Large 32 se insertó 5' de PpLuc(GC). Se insertó una 3'-UTR derivada de albúmina humana (albumin7) 3' de PpLuc (GC), Además, se insertó una secuencia poli(A) A64, seguida de C30 y una secuencia tallo-bucle de histona 3' de albumin7 La secuencia tallo-bucle de histona fue seguida por un sitio de restricción utilizado para la linealización del vector antes de la transcripción *in vitro*. El ARNm obtenido a partir de este vector por transcripción *in vitro* se designa "rpl32 - PpLuc (GC) – albumin7 - A64 - C30 - histonaSL".

En resumen, se generó un vector que comprende la secuencia que codifica el ARNm que se usó en experimentos adicionales. La secuencia de ADN (SEQ ID NO: 13) que codifica dicho ARNm se muestra en la Fig. 1. El ARNm correspondiente a dicha secuencia de ADN se caracteriza por los siguientes elementos: rpl32 - PpLuc (GC) - albumin7 - A64 - C30 - histonaSL

Aquí, se utilizan las siguientes abreviaturas:

- PpLuc (GC): secuencia de ARNm enriquecida en GC que codifica la luciferasa *Photinus pyralis*
- rpl32: 5'-UTR de proteína ribosómica humana Large 32 que carece del tracto de oligopirimidina 5' terminal
- albumin7: 3'-UTR de albúmina humana con tres mutaciones de punto único introducidas para eliminar una señal de terminación T7, así como un sitio de restricción HindIII y XbaI
- A64: secuencia poli(A) de 64 adenilatos
- C30: secuencia poli(C) de 30 citidilatos
- histonaSL: secuencia tallo-bucle de histona de acuerdo con la SEQ ID NO: 11.

Otros constructos utilizados en los experimentos:

32L4-H1N1 (Países Bajos 2009) -HA (GC) -albumin7-A64-N5-C30-histonaSL (SEQ ID NO: 14)

32L4-RAV-G (GC) -albumin7-A64-N5-C30-histonaSL (SEQ ID NO: 15)

Las plantillas se prepararon como se describe para rpl32 - PpLuc (GC) - albumin7 - A64 - C30 - histonaSL.

2. Transcripción in vitro

La plantilla de ADN preparada en el Ejemplo 1 se linealizó y se transcribió *in vitro* usando polimerasa T7. La plantilla de ADN se digirió luego mediante tratamiento con ADNasa. Los transcriptos de ARNm contenían una estructura 5'-cap obtenida mediante la adición de un exceso de N7-metilguanosin-5'-trifosfato-5'-guanosina a la reacción de transcripción. El ARNm así obtenido se purificó y se resuspendió en agua.

3. Adenilación enzimática

El ARN se hizo reaccionar con la poli(A) polimerasa de *E. coli* (Cellscript) utilizando ATP 1 mM a 37ºC durante 30 o 60 min. Inmediatamente después, el ARN se purificó utilizando una columna giratoria (mini columna RNeasy, Quiagen). El ARN se hizo correr en un gel para evaluar la extensión del ARN.

Para los experimentos de vacunación, el ARNm se complejó opcionalmente con protamina. La complejación de ARNm consistió en una mezcla de ARNm desnudo al 50% y ARNm al 50% complejado con protamina en una relación en peso 2:1. Primero, el ARNm se complejó con protamina por adición de una solución de lactato de protamina-Ringer al ARNm. Después de incubación durante 10 minutos, cuando los complejos se generaron de manera estable, se añadió ARNm desnudo y la concentración final de la vacuna se ajustó con solución lactato de Ringer. La vacuna de ARNm formulada obtenida se usó para experimentos *in vivo*.

4. Expresión de proteínas por lipofección de ARNm

Se sembraron fibroblastos dérmicos humanos (HDF) en placas de 96 pocillos tres días antes de la transfección a una densidad de 10^4 células por pocillo. Inmediatamente antes de la lipofección, las células se lavaron en Opti-MEM. Las células se lipofectaron con 25 ng de ARNm que codifica PpLuc por pocillo complejado con Lipofectamine2000. El ARNm que codifica la luciferasa de *Renilla reniformis* (RrLuc) se transfectó junto con ARNm de PpLuc para controlar la eficacia de la transfección (2,5 ng de ARNm de RrLuc por pocillo). 90 minutos después del inicio de la transfección, Opti-MEM se cambió por medio. 6, 24, 48 y 72 horas después de la transfección, se aspiró el medio y las células se lisaron en 100 µl de tampón de lisis (Passive Lysis Buffer, Promega). Los lisados se almacenaron a -80ºC hasta que se midió la actividad luciferasa.

5. Medida de la luminiscencia en el lisado celular

La actividad de la luciferasa del lisado celular se midió como unidades de luz relativas (RLU) en un lector de placas Hidex Chameleon. La actividad de PpLuc se midió a los 2 segundos de tiempo de medida usando 20 µl de lisado y 50 µl de tampón luciferina (Beetle-Juice, PJK GmbH). La actividad de RrLuc se midió a los 2 segundos de tiempo de medida usando 20 µl de lisado y 50 µl de tampón coelenterazina (Renilla-Juice, PJK GmbH).

6. Expresión de proteínas por inyección intramuscular de ARNm

Los ratones se anestesiaron por inyección intraperitoneal de una mezcla de Ketavet y Rompun. Después de afeitar la parte inferior de la pata del animal, se inyectaron 2 µg de ARNm que codifica PpLuc en 20 µl de lactato de Ringer (80%) vía intramuscular (*M. tibialis* o *M. gastrocnemius*).

7. Imágenes de luminiscencia in vivo

Los ratones se anestesiaron por inyección intraperitoneal de una mezcla de Ketavet y Rompun. Se inyectaron 150 µl de solución de Luciferina (20 g/l) vía intraperitoneal. 10 minutos después de la inyección de Luciferina, se registró la luminiscencia en un sistema de imágenes MS Lumina II.

Resultados

8.1 La poliadenilación adicional del ARNm artificial aumenta la expresión de proteínas a partir del ARNm artificial *in vitro*

Para investigar el efecto de la poliadenilación adicional del ARNm artificial sobre la expresión de proteínas a partir del ARNm, el ARNm artificial se sintetizó mediante transcripción *in vitro* (rpl32 - PpLuc (GC) - albumin7 - A64 - C30 - histonaSL). Parte de un lote de ARNm se adenila enzimáticamente para agregar una cola poli(A) de aprox. 160 adenilatos (Lote 1). Parte de un lote diferente de ARNm se adenila enzimáticamente para agregar una cola poli(A) de aprox. 380 adenilatos (Lote 2) (ver Figura 2).

Los ARNm que codifican la luciferasa se transfectaron en fibroblastos dérmicos humanos (HDF) por triplicado. Los niveles de luciferasa se midieron 6, 24, 48 y 72 horas después de la transfección. A partir de estos datos, la proteína total expresada en 0 a 72 horas se calculó como el área bajo la curva (AUC) (ver la siguiente Tabla 1 y Figura 3).

Tabla 1: Actividad de luciferasa medida en fibroblastos dérmicos humanos (HDF)

Cola Poli(A)	RLU 6 horas	RLU 24 horas	RLU 48 horas	RLU 72 horas	AUC
Lot 1 incl. A64	37252	59085	29825	14612	2579000
Lot 1 incl. A64 + aprox. A160 (A224)	81043	246102	89506	41308	8784000
Lot 2 incl. A64	47959	61053	23001	14053	2578000
Lot 2 incl. A64 + aprox. A380 (A444)	69780	188560	69269	44478	6993000

La expresión de luciferasa total fue idéntica en ambos lotes de ARNm que contenían una secuencia A64 transcrita *in vitro*. La adición de aprox. 160 adenilatos en el extremo 3' del ARNm (que da como resultado una 3'-UTR que comprende aproximadamente 224 adenilatos comprendidos en una secuencia poli(A)) aumentó la expresión de luciferasa en un factor 3,4. La adición de aprox. 380 adenilatos en el extremo 3' del ARNm (que da como resultado una 3'-UTR que comprende aproximadamente 444 adenilatos comprendidos en una secuencia poli(A)) aumentó la expresión de luciferasa solo en un grado similar, en un factor 2,7.

Por tanto, la adición de adenilatos (adicionales) al extremo 3' del ARNm aumenta notablemente la expresión *in vitro* de la proteína codificada por el ARNm. En particular, una 3'-UTR que comprende más de 64 adenilatos comprendidos en una secuencia poli(A) aumenta notablemente la expresión de proteínas *in vitro*.

8.2 La poliadenilación adicional del ARNm artificial aumenta fuertemente la expresión de proteínas a partir del ARNm artificial después de la inyección intramuscular

Para investigar el efecto de la poliadenilación adicional del ARNm artificial sobre la expresión de proteínas a partir del ARNm inyectado vía intramuscular, el ARNm artificial se sintetizó mediante transcripción *in vitro* (rpl32 - PpLuc (GC) - albumin7 - A64 - C30 - histonaSL). Parte de este ARNm se adenila enzimáticamente para agregar una cola poli(A) de aprox. 380 adenilatos.

Se inyectaron 2 µg de ARNm codificadores para la luciferasa vía intramuscular (*M. tibialis* o *M. gastrocnemius*) en ratones BALB/c (10 replicados por grupo). La luminiscencia *in vivo* se registró los días siguientes (ver Figura 4). A partir de estos datos, la proteína total expresada en 0 a 15 días se calculó como el área bajo la curva. La luciferasa se expresó claramente a partir de ARNm inyectado intramuscularmente que contenía una secuencia A64. Sorprendentemente, sin embargo, la poliadenilación adicional del ARNm artificial con 380 adenilatos aumentó la expresión de luciferasa muy fuertemente, elevando la luciferasa total expresada ocho veces. La magnitud del aumento en la expresión debido a la cola (adicional) poli(A) no se anticipó considerando el efecto mucho menor observado en las células cultivadas. Por tanto, la adición de adenilatos (adicionales) comprendidos en una secuencia poli(A) aumenta la expresión de proteínas a partir de ARNm inyectado vía intramuscular de forma muy inesperada. Paralelamente, se probó el efecto de la poliadenilación adicional del ARNm artificial con aproximadamente 430 adenilatos (ver Figura 5) y aproximadamente 1.000 adenilatos (ver Figura 6), respectivamente. Si bien los dos ARNm que se poliadenilaron con aproximadamente 430 adenilatos y aproximadamente 1.000 adenilatos, respectivamente, condujeron a una mayor expresión de proteínas en comparación con el ARNm no poliadenilado, no se observó ningún aumento adicional con respecto al ARNm artificial poliadenilado con 380 adenilatos.

9. Vacunación con ARNm que codifica HA:

Se vacunaron ratones Balb/c 2 veces (d0 y d21) en ambos *M. tibialis*. Se vacunaron 8 ratones con 40 µg de R2564 (ARNm de HA desnudo), 8 animales se vacunaron con 40 µg de R2564 poliadenilado (SEQ ID NO: 14; HA desnudo, poliadenilado), se vacunaron 8 animales con 40 µg de R2630 (ARNm de HA formulado) y 8 animales fueron vacunados con 40 µg primero poliadenilados y luego formulados con R2564. 8 ratones inyectados con RiLa sirvieron como control. Se recogió sangre el día 35.

9.1. Ensayo de inhibición de la hemaglutinación (HI)

En una placa de 96 pocillos, los sueros obtenidos se mezclaron con el antígeno HA H1N1 (A/California/07/2009 (H1N1); NIBSC) y glóbulos rojos (4% de eritrocitos; Lohmann Tierzucht). En presencia de anticuerpos neutralizantes de HA, se puede observar una inhibición de la hemaglutinación de los eritrocitos. El nivel más bajo de suero titulado que resultó en una inhibición visible de la hemaglutinación fue el resultado del ensayo.

Resultados:

Los resultados muestran que se podrían alcanzar títulos de HI más altos mediante la poliadenilación del ARNm. Todos los ratones tratados con ARNm poliadenilado alcanzaron un nivel superior a los títulos neutralizantes de virus potencialmente protectores (> 40).

10. Vacunación con ARNm que codifica RAV G:

Se vacunaron ratones Balb/c 2 veces (d0 y d21) con 20 µg de ARNm de RAV-G en ambos M. tibialis. 8 animales fueron vacunados i.m. con R2506 (SEQ ID NO: 15; ARNm de RAV-G desnudo), se vacunaron 8 animales i.m. con R2506 poliadenilado (ARN desnudo) y 8 ratones fueron inyectados con RiLa como control. Se recogió sangre 28 días después del cebado. Se analizó el suero en busca de VNT.

10.1. Prueba de neutralización viral

La detección de la respuesta de anticuerpos neutralizantes de virus (respuesta inmune específica de células B) se llevó a cabo mediante un ensayo de neutralización de virus. El resultado de este ensayo se conoce como título de neutralización viral (VNT). Según los estándares de la OMS, un título de anticuerpos se considera protector si el VNT respectivo es de al menos 0,5 IU/ml. Por tanto, se tomaron muestras de sangre de ratones vacunados el día 28 y se prepararon sueros. Estos sueros se usaron en la prueba de neutralización viral con anticuerpos fluorescentes (FAVN) usando la cepa de virus de desafío adaptada al cultivo celular (CVS) del virus de la rabia según lo recomendado por la OIE (Organización Mundial para la Salud Animal) y se describió por primera vez en Cliquet F., Aubert M. & Sagne L. (1998); J. Immunol. Methods, 212, 79-87. En resumen, los sueros inactivados térmicamente se analizaron por cuadruplicado en series de diluciones de dos veces con respecto a su potencial para neutralizar 100 TCID50 (dosis infecciosas de cultivo de tejidos al 50%) de CVS en 50 µl de volumen. Así, las diluciones de suero se incubaron con virus durante 1 hora a 37ºC (en incubadora húmeda con 5% de CO_2) y posteriormente se añadieron células BHK-21 tripsinizadas ($4x10^5$ células/ml; 50 µl por pocillo). Los cultivos celulares infectados se incubaron durante 48 horas en una incubadora húmeda a 37ºC y 5% de CO_2. La infección de las células se analizó después de fijar las células usando acetona al 80% a temperatura ambiente con el conjugado antirrábico FITC. Las placas se lavaron dos veces con PBS y se eliminó el exceso de PBS. Los cultivos celulares se puntuaron como positivos o negativos con respecto a la presencia del virus de la rabia. Las células con puntuación negativa en los pozos tratados con suero representan la neutralización del virus de la rabia. Cada prueba FAVN incluyó suero estándar WHO u OIE (suero de referencia positivo) que sirvió como referencia para la estandarización del ensayo. La actividad de neutralización de los sueros de prueba se calculó con referencia al suero estándar proporcionado por la OMS y se mostró como Unidades Internacionales/ml (IU/ml).

Resultados

Como se puede ver en la Figura 9, el ARNm de RAV-G poliadenilado induce títulos de anticuerpos neutralizantes más altos, muy por encima del estándar de la OMS de 0,5 IU/ml.

Listado de Secuencias:

<110> CureVac AG

<120> Moléculas de ácido nucleico artificiales para la expresión mejorada de proteínas

<130> CU01P142WO1

<160> 15

<170> PatentIn version 3.5

<210> 1
<211> 140
<212> ADN
<213> Secuencia artificial
<220>
<223> secuencia de AND ilustrativa que comprende una secuencia poli(A)

<400> 1
gtccacctgt ccctcctggg ctgctggatt gtctcgtttt cctgccaaat aaacaggatc 60

agcgctttac agatctaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 120

aaaaaaaaaa aaaaaaaaaa 140

<210> 2
<211> 140
<212> ARN
<213> Secuencia artificial

<220>
<223> Secuencia de ARN ilustrativa que comprende una secuencia poli(A)

<400> 2

guccaccugu cccuccuggg cugcuggauu gucucguuuu ccugccaaau aaacaggauc 60

agcgcuuuac agaucuaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 120

aaaaaaaaaa aaaaaaaaaa 140

<210> 3
<211> 186
<212> ADN
<213> Homo sapiens

<400> 3

catcacattt aaaagcatct cagcctacca tgagaataag agaaagaaaa tgaagatcaa 60

aagcttattc atctgttttt ctttttcgtt ggtgtaaagc caacaccctg tctaaaaaac 120

ataaatttct ttaatcattt tgcctctttt ctctgtgctt caattaataa aaaatggaaa 180

gaatct 186

<210> 4
<211> 186
<212> ADN

<213> Homo sapiens

<400> 4

catcacattt aaaagcatct cagcctacca tgagaataag agaaagaaaa tgaagatcaa 60

tagcttattc atctcttttt ctttttcgtt ggtgtaaagc caacaccctg tctaaaaaac 120

ataaatttct ttaatcattt tgcctctttt ctctgtgctt caattaataa aaaatggaaa 180

gaacct 186

<210> 5
<211> 110
<212> ADN
<213> Homo sapiens

<400> 5

gctggagcct cggtggccat gcttcttgcc ccttgggcct ccccccagcc cctcctcccc 60

ttcctgcacc cgtacccccg tggtctttga ataaagtctg agtgggcggc 110

<210> 6
<211> 108
<212> ADN
<213> Homo sapiens

<400> 6

gctggagcct cggtagccgt tcctcctgcc cgctgggcct cccaacgggc cctcctcccc 60

tccttgcacc ggcccttcct ggtctttgaa taaagtctga gtgggcag 108

<210> 7
<211> 132
<212> ADN
<213> Homo sapiens

<400> 7

gctcgctttc ttgctgtcca atttctatta aaggttcctt tgttccctaa gtccaactac 60

taaactgggg gatattatga agggccttga gcatctggat tctgcctaat aaaaaacatt 120

tattttcatt gc 132

<210> 8
<211> 44
<212> ADN
<213> Homo sapiens

<400> 8

gcccgatggg cctcccaacg ggccctcctc ccctccttgc accg 44

<210> 9
<211> 70
<212> ADN

<213> Secuencia artificial

<220>
<223> secuencia 3'-UTR of proteína ribosómica Small 9 (RPS9)

<400> 9

gtccacctgt ccctcctggg ctgctggatt gtctcgtttt cctgccaaat aaacaggatc 60

agcgctttac 70

<210> 10
<211> 70
<212> ADN
<213> Secuencia artificial

<220>
<223> secuencia 3'-UTR of proteína ribosómica Small 9 (RPS9)

<400> 10

guccaccugu cccuccuggg cugcuggauu gucucguuuu ccugccaaau aaacaggauc 60

agcgcuuuac 70

<210> 11
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> secuencia tallo-bucle de histona

<400> 11

caaaggctct tttcagagcc acca 24

<210> 12
<211> 42
<212> ADN
<213> Secuencia artificial

<220>
<223> 5'-UTR de proteína ribosómica Large 32 humana sin tracto de oligopirimidina 5' terminal

<400> 12

ggcgctgcct acggaggtgg cagccatctc cttctcggca tc 42

<210> 13
<211> 2035
<212> ADN
<213> Secuencia artificial

<220>
<223> rpl32 - PpLuc(GC) - albumin7 - A64 - C30 – tallo-bucle de histona

<400> 13

ggggcgctgc ctacggaggt ggcagccatc tccttctcgg catcaagctt gaggatggag 60

gacgccaaga acatcaagaa gggcccggcg cccttctacc cgctggagga cgggaccgcc 120

ggcgagcagc tccacaaggc catgaagcgg tacgccctgg tgccgggcac gatcgccttc 180

accgacgccc acatcgaggt cgacatcacc tacgcggagt acttcgagat gagcgtgcgc 240

ctggccgagg ccatgaagcg gtacggcctg aacaccaacc accggatcgt ggtgtgctcg 300

gagaacagcc tgcagttctt catgccggtg ctgggcgccc tcttcatcgg cgtggccgtc 360

gccccggcga acgacatcta caacgagcgg gagctgctga acagcatggg gatcagccag 420

ccgaccgtgg tgttcgtgag caagaagggc ctgcagaaga tcctgaacgt gcagaagaag 480

ctgcccatca tccagaagat catcatcatg gacagcaaga ccgactacca gggcttccag 540

tcgatgtaca cgttcgtgac cagccacctc ccgccgggct tcaacgagta cgacttcgtc 600

ccggagagct tcgaccggga caagaccatc gccctgatca tgaacagcag cggcagcacc 660

ggcctgccga agggggtggc cctgccgcac cggaccgcct gcgtgcgctt ctcgcacgcc 720

cgggacccca tcttcggcaa ccagatcatc ccggacaccg ccatcctgag cgtggtgccg 780

ttccaccacg gcttcggcat gttcacgacc ctgggctacc tcatctgcgg cttccgggtg 840
gtcctgatgt accggttcga ggaggagctg ttcctgcgga gcctgcagga ctacaagatc 900

cagagcgcgc tgctcgtgcc gaccctgttc agcttcttcg ccaagagcac cctgatcgac 960

aagtacgacc tgtcgaacct gcacgagatc gccagcgggg gcgccccgct gagcaaggag 1020

gtgggcgagg ccgtggccaa gcggttccac ctcccgggca tccgccaggg ctacggcctg 1080

accgagacca cgagcgcgat cctgatcacc cccgaggggg acgacaagcc gggcgccgtg 1140

ggcaaggtgg tcccgttctt cgaggccaag gtggtggacc tggacaccgg caagaccctg 1200

ggcgtgaacc agcggggcga gctgtgcgtg cgggggccga tgatcatgag cggctacgtg 1260

aacaacccgg aggccaccaa cgccctcatc gacaaggacg gctggctgca cagcggcgac 1320

atcgcctact gggacgagga cgagcacttc ttcatcgtcg accggctgaa gtcgctgatc 1380

aagtacaagg gctaccaggt ggcgccggcc gagctggaga gcatcctgct ccagcacccc 1440

aacatcttcg acgccggcgt ggccgggctg ccggacgacg acgccggcga gctgccggcc 1500

gcggtggtgg tgctggagca cggcaagacc atgacggaga aggagatcgt cgactacgtg 1560

gccagccagg tgaccaccgc caagaagctg cggggcggcg tggtgttcgt ggacgaggtc 1620

ccgaagggcc tgaccgggaa gctcgacgcc cggaagatcc gcgagatcct gatcaaggcc 1680

aagaagggcg gcaagatcgc cgtgtaagac tagtgcatca catttaaaag catctcagcc 1740

taccatgaga ataagagaaa gaaaatgaag atcaatagct tattcatctc ttttttcttt 1800

tcgttggtgt aaagccaaca ccctgtctaa aaaacataaa tttctttaat cattttgcct 1860

cttttctctg tgcttcaatt aataaaaaat ggaaagaacc tagatctaaa aaaaaaaaaa 1920

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa atgcatcccc 1980

ccccccccc ccccccccc cccccccaaa ggctcttttc agagccacca gaatt 2035

<210> 14
<211> 2083
<212> ADN
<213> Secuencia artificial

<220>
<223> 32L4-H1N1(Netherlands2009)-HA(GC)-albumin7- A64-N5-C30-histonaSL

<400> 14

ggggcgctgc ctacggaggt ggcagccatc tccttctcgg catcaagctt accatgaagg 60
ccatcctggt ggtcctcctg tacaccttcg ccaccgcgaa cgccgacacg ctgtgcatcg 120
gctaccacgc caacaacagc accgacaccg tggacaccgt gctcgagaag aacgtcacgg 180
tgacccactc cgtgaacctg ctggaggaca agcacaacgg gaagctctgc aagctgcggg 240
gcgtcgcccc gctgcacctc gggaagtgca acatcgccgg ctggatcctg gggaacccgg 300
agtgcgagag cctgtccacc gcgagctcct ggagctacat cgtggagacc tccagctccg 360
acaacggcac gtgctacccc ggcgacttca tcgactacga ggagctccgc gagcagctga 420
gctccgtgag ctccttcgag cggttcgaga tcttccccaa gaccagctcc tggcccaacc 480
acgacagcaa caaggggtc accgccgcct gcccgcacgc cggcgcgaag tccttctaca 540
agaacctgat ctggctcgtg aagaagggga acagctaccc caagctgtcc aagagctaca 600
tcaacgacaa gggcaaggag gtgctggtcc tctgggggat ccaccacccc agcacctccg 660
ccgaccagca gagcctgtac cagaacgccg acgcctacgt gttcgtgggc tccagccgct 720
actccaagaa gttcaagccc gagatcgcca tccggccgaa ggtccgcgac caggagggcc 780
ggatgaacta ctactggacg ctggtggagc ccggggacaa gatcaccttc gaggcgaccg 840
gcaacctcgt ggtcccccgc tacgccttcg ccatggagcg gaacgccggg agcggcatca 900
tcatctccga cacccccgtg cacgactgca acacgacctg ccagaccccg aagggcgcca 960
tcaacaccag cctgcccttc cagaacatcc accccatcac gatcgggaag tgccccaagt 1020
acgtgaagtc caccaagctg cgcctcgcga ccggcctgcg gaacgtcccg agcatccagt 1080
cccgcgggct gttcggcgcc atcgccgggt tcatcgaggg cggctggacc gggatggtgg 1140
acggctggta cgggtaccac caccagaacg agcagggcag cgggtacgcc gccgacctca 1200
agtccacgca gaacgcgatc gacgagatca ccaacaaggt gaacagcgtc atcgagaaga 1260
tgaacaccca gttcaccgcc gtgggcaagg agttcaacca cctggagaag cggatcgaga 1320
acctgaacaa gaaggtcgac gacggcttcc tcgacatctg gacgtacaac gccgagctgc 1380
tggtgctcct ggagaacgag cgcaccctgg actaccacga ctccaacgtg aagaacctct 1440
acgagaaggt ccggagccag ctgaagaaca acgccaagga gatcgggaac ggctgcttcg 1500
agttctacca caagtgcgac aacacctgca tggagtccgt gaagaacggg acctacgact 1560
accccaagta cagcgaggag gccaagctga accgcgagga gatcgacggc gtgaagctcg 1620

agtccacgcg gatctaccag atcctggcga tctacagcac cgtcgccagc tccctggtgc 1680

tcgtggtcag cctgggggcc atctccttct ggatgtgcag caacggctcc ctgcagtgcc 1740

gcatctgcat ctgaccacta gtgcatcaca tttaaaagca tctcagccta ccatgagaat 1800

aagagaaaga aaatgaagat caatagctta ttcatctctt tttctttttc gttggtgtaa 1860

agccaacacc ctgtctaaaa aacataaatt tctttaatca ttttgcctct tttctctgtg 1920

cttcaattaa taaaaaatgg aaagaaccta gatctaaaaa aaaaaaaaaa aaaaaaaaaa 1980

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaat gcatcccccc ccccccccc 2040

ccccccccc ccccaaagg ctcttttcag agccaccaga att 2083

<210> 15
<211> 1957
<212> ADN
<213> Secuencia artificial

<220>
<223> 32L4-RAV-G(GC)-albumin7-A64-N5-C30-histonaSL

<400> 15
ggggcgctgc ctacggaggt ggcagccatc tccttctcgg catcaagctt accatggtgc 60

cccaggccct gctcttcgtc ccgctgctgg tgttcccct ctgcttcggc aagttcccca 120

tctacaccat ccccgacaag ctggggccgt ggagccccat cgacatccac cacctgtcct 180

gccccaacaa cctcgtggtc gaggacgagg gctgcaccaa cctgagcggg ttctcctaca 240

tggagctgaa ggtgggctac atcagcgcca tcaagatgaa cgggttcacg tgcaccggcg 300

tggtcaccga ggcggagacc tacacgaact tcgtgggcta cgtgaccacc accttcaagc 360

ggaagcactt ccgccccacg ccggacgcct gccgggccgc ctacaactgg aagatggccg 420

gggacccccg ctacgaggag tccctccaca acccctaccc cgactaccac tggctgcgga 480

ccgtcaagac caccaaggag agcctggtga tcatctcccc gagcgtggcg gacctcgacc 540

cctacgaccg ctccctgcac agccgggtct tccccggcgg gaactgctcc ggcgtggccg 600

tgagctccac gtactgcagc accaaccacg actacaccat ctggatgccc gagaacccgc 660

gcctggggat gtcctgcgac atcttcacca acagccgggg caagcgcgcc tccaagggca 720

gcgagacgtg cgggttcgtc gacgagcggg gcctctacaa gtccctgaag ggggcctgca 780

agctgaagct ctgcggcgtg ctgggcctgc gcctcatgga cgggacctgg gtggcgatgc 840

agaccagcaa cgagaccaag tggtgccccc ccggccagct ggtcaacctg cacgacttcc 900

ggagcgacga gatcgagcac ctcgtggtgg aggagctggt caagaagcgc gaggagtgcc 960

tggacgccct cgagtccatc atgacgacca agagcgtgtc cttccggcgc ctgagccacc 1020

tgcggaagct cgtgcccggg ttcggcaagg cctacaccat cttcaacaag accctgatgg 1080

aggccgacgc ccactacaag tccgtccgca cgtggaacga gatcatcccg agcaaggggt 1140

gcctgcgggt gggcggccgc tgccaccccc acgtcaacgg ggtgttcttc aacggcatca 1200

tcctcgggcc cgacggcaac gtgctgatcc ccgagatgca gtccagcctg ctccagcagc 1260

acatggagct gctggtctcc agcgtgatcc cgctcatgca ccccctggcg gacccctcca 1320

ccgtgttcaa gaacggggac gaggccgagg acttcgtcga ggtgcacctg cccgacgtgc 1380

acgagcggat cagcggcgtc gacctcggcc tgccgaactg ggggaagtac gtgctgctct 1440

ccgccggcgc cctgaccgcc ctgatgctga tcatcttcct catgacctgc tggcgccggg 1500

tgaaccggag cgagcccacg cagcacaacc tgcgcgggac cggccgggag gtctccgtga 1560

ccccgcagag cgggaagatc atctccagct gggagtccta caagagcggc ggcgagaccg 1620

ggctgtgagg actagtgcat cacatttaaa agcatctcag cctaccatga gaataagaga 1680

aagaaaatga agatcaatag cttattcatc tctttttctt tttcgttggt gtaaagccaa 1740

caccctgtct aaaaaacata aatttcttta atcattttgc ctcttttctc tgtgcttcaa 1800

ttaataaaaa atggaaagaa cctagatcta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaatgcatcc cccccccccc cccccccccc 1920

ccccccccca aaggctcttt tcagagccac cagaatt 1957

REIVINDICACIONES

1. Molécula de ácido nucleico artificial que comprende

a) al menos un marco de lectura abierto (ORF); y
b) al menos una región 3'-no traducida (3'-UTR) que comprende
al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, y

donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende un tallo-bucle de histona según la SEQ ID NO: 11
para su uso como un medicamento, para su uso como una vacuna o para su uso en la terapia génica,
donde la molécula de ácido nucleico artificial está asociada a o complejada con un compuesto catiónico policatiónico.

2. Molécula de ácido nucleico artificial para su uso según la reivindicación 1, donde la molécula de ácido nucleico artificial se administra intramuscularmente.

3. Molécula de ácido nucleico artificial para su uso según la reivindicación 1 o 2, donde el al menos un marco de lectura abierto codifica un antígeno.

4. Molécula de ácido nucleico artificial para su uso según la reivindicación 3, donde el antígeno se selecciona del grupo consistente en un antígeno patogénico, un antígeno tumoral, un antígeno alergénico y un antígeno autoinmune.

5. Molécula de ácido nucleico artificial para su uso según la reivindicación 4, donde el antígeno patogénico se deriva de un patógeno asociado a una enfermedad infecciosa.

6. Molécula de ácido nucleico artificial para su uso según la reivindicación 5, donde el patógeno se selecciona del grupo consistente en una patógeno bacteriano, viral, fúngico o protozoario.

7. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 7, donde el ácido nucleico artificial es una molécula de ADN o una molécula de ARN, preferentemente una molécula de ARNm.

8. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 7, donde el ácido nucleico artificial comprende de 180 a 1.100 nucleótidos adenina comprendidos en las al menos dos secuencias poli(A) separadas.

9. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 8, donde al menos una secuencia poli(A) se localiza en el extremo 3' de la molécula de ácido nucleico artificial.

10. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 9, donde la 3'-UTR comprende al menos otro elemento 3'-UTR, que es una secuencia de ácido nucleico derivada de la 3'-UTR de un gen seleccionado del grupo consistente en un gen de α-globina, un gen de albúmina, un gen de β-globina, un gen de proteína ribosómica, un gen de tirosina-hidroxilasa, un gen de lipoxigenasa y un gen de colágeno alfa, tal como un colágeno alfa 1 (I), o de una variante de una 3'-UTR de un gen seleccionado del grupo consistente en un gen de albúmina, un gen de α -globina, un gen de β-globina, un gen de proteína ribosómica, un gen de tirosina-hidroxilasa, un gen de lipoxigenasa y un gen de colágeno alfa, tal como un gen de colágeno alfa 1 (I).

11. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 10, donde la 3'-UTR comprende al menos otro elemento 3'-UTR que se deriva de la región 3'-UTR de un gen que codifica una proteína ribosómica.

12. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 10, donde la 3'-UTR comprende al menos otro elemento 3'-UTR que se deriva de la región 3'-UTR de proteína ribosómica L9 (RPL9), la proteína ribosómica L3 (RPL3), la proteína ribosoma L4 (RPL4), la proteína ribosómica L5 (RPL5), la proteína ribosómica L6 (RPL6), la proteína ribosómica L7 (RPL7), proteína ribosómica L7a (RPL7A), proteína ribosómica L1 1 (RPL1), proteína ribosómica L12 (RPL12), proteína ribosómica L13 (RPL13), proteína ribosómica L23 (RPL23), proteína ribosómica L18 (RPL18), proteína ribosómica L18a (RPL18A), proteína ribosómica L19 (RPL19), proteína ribosómica L21 (RPL21), proteína

ribosómica L22 (RPL22), proteína ribosómica L23a (RPL23A), proteína ribosómica L17 (RPL17), proteína ribosómica L24 (RPL24), proteína ribosómica L26 (RPL26) , proteína ribosómica L27 (RPL27), proteína ribosómica L30 (RPL30), proteína ribosómica L27a (RPL27A), proteína ribosómica L28 (RP L28), proteína ribosómica L29 (RPL29), proteína ribosómica L31 (RPL31), proteína ribosómica L32 (RPL32), proteína ribosómica L35a (RPL35A), proteína ribosómica L37 (RPL37), proteína ribosómica L37a (RPL37A), proteína ribosómica L38 (RPL38), proteína ribosómica L39 (RPL39), proteína ribosómica, grande, P0 (RPLP0), proteína ribosómica, grande, P1 (RPLP1), proteína ribosómica, grande, P2 (RPLP2), proteína ribosómica S3 (RPS3), proteína ribosómica S3A (RPS3A), proteína ribosómica S4, ligada a X (RPS4X), proteína ribosómica S4, ligada a Y1 (RPS4Y1), proteína ribosómica S5 (RPS5), proteína ribosómica S6 (RPS6), proteína ribosómica S7 (RPS7), proteína ribosómica S8 (RPS8), proteína ribosómica S9 (RPS9), proteína ribosómica S10 (RPS10), proteína ribosómica S11 (RPS1 1), proteína ribosómica S12 (RPS12), proteína ribosómica S13 (RPS13), proteína ribosómica S15 (RPS15), proteína ribosómica S15a (RPS15A), proteína ribosómica S16 (RPS16), proteína ribosómica S19 (RPS19), proteína ribosómica S20 (RPS20), proteína ribosómica S21 (RPS21), proteína ribosómica S23 (RPS23), proteína ribosómica S25 (RPS25), proteína ribosómica S26 (RPS26), proteína ribosómica S27 (RPS27), proteína ribosómica S27a (RPS27a), proteína ribosómica S28 (RPS28), proteína ribosómica S29 (RPS29), proteína ribosómica L15 (RPL15), proteína ribosómica S2 (RPS2), proteína ribosómica L14 (RPL14), proteína ribosómica S14 (RPS14), proteína ribosómica L10 (RPL10), proteína ribosómica L10a (RPL10A), proteína ribosómica L35 (RPL35), proteína ribosómica L 13a (RPL13A)), proteína ribosómica L36 (RPL36), proteína ribosómica L36a (RPL36A), proteína ribosómica L41 (RPL41), proteína ribosómica S18 (RPS18), proteína ribosómica S24 (RPS24), proteína ribosómica L8 (RPL8), proteína ribosómica L34 (RPL34) , proteína ribosómica S17 (RPS17), proteína ribosómica SA (RPSA) o proteína ribosómica S17 (RPS17), ubiquitina A-52 (UBA52), virus del sarcoma murino de Finkel-Biskis-Reilly (FBR-MuSV) expresado de forma ubicua (FAU), proteína ribosómica L22 tipo 1 (RPL22L1), proteína ribosómica tipo L39 (RPL39L), proteína ribosómica tipo L10 (RPL10L), proteína ribosómica tipo L36a (RPL36AL), proteína ribosómica tipo L3 (RPL3L), proteína ribosómica tipo S27 (RPS27L), proteína ribosómica L26 tipo 1 (RPL26L1), proteína ribosómica L7 tipo 1 (RPL7L1), pseudogen de proteína ribosómica L13a (RPL13AP), pseudogen 8 de proteína ribosómica L37a (RPL37AP8), pseudogen 5 de proteína ribosómica S10 (RPS10P5), pseudogen 11 de proteína ribosómica S26 (RPS26P11), pseudogen 5 de proteína ribosómica L39 (RPL39P5), pseudogen 6 de proteína ribosómica, grande, P0 (RPLP0P6) o pseudogen 14 de proteína ribosómica L36 (RPLP14)

13. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 10 a 12, donde el ORF no codifica un gen mencionado en cualquiera de las reivindicaciones 11 a 12, respectivamente.

14. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 13, que comprende además una estructura de 5'-cap, una secuencia poli(C) y/o un motivo IRES.

15. Molécula de ácido nucleico artificial para su uso según la reivindicación 15, donde el ácido nucleico comprende un elemento 5' adicional que se selecciona de una 5'-UTR, un promotor o una 5'-UTR y un promotor que contiene la secuencia.

16. Molécula de ácido nucleico artificial para su uso según la reivindicación 15, donde la 5'-UTR comprende o consiste en un elemento 5'-UTR que comprende una secuencia de ácido nucleico que se deriva de una 5'-UTR de un gen TOP, una secuencia de ARN correspondiente, un homólogo, un fragmento o una variante de la misma.

17. Molécula de ácido nucleico artificial para su uso según la reivindicación 16, donde el gen TOP codifica una proteína ribosómica.

18. Molécula de ácido nucleico artificial para su uso según la reivindicación 16 o 17, donde el elemento 5'-UTR carece del motivo 5'-TOP.

19. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 16 a 18, donde el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que se deriva de una 5'-UTR de un gen TOP que codifica una proteína ribosómica Large (RPL) o de un homólogo, variante o derivado del mismo.

20. Molécula de ácido nucleico artificial para su uso según la reivindicación 19, donde el elemento 5'-UTR comprende o consiste en una secuencia de ácido nucleico que comprende o consiste en una secuencia de ARN correspondiente de la secuencia de ácido nucleico de acuerdo con la SEC ID NO. 12.

21. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 20, donde el marco de lectura abierto de la molécula de ácido nucleico artificial está modificado al menos

parcialmente en G/C en comparación con el marco de lectura abierto de tipo natural y/o donde el marco de lectura abierto comprende una región optimizada en codones.

22. Molécula de ácido nucleico artificial para su uso según la reivindicación 21, donde el contenido en G/C del marco de lectura abierto está aumentado en comparación con el marco de lectura abierto de tipo natural.

23. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 22, donde los al menos dos elementos de secuencia poli(A) están separados por una secuencia de ácido nucleico que comprende un elemento poli(C).

24. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 23, donde la molécula de ácido nucleico artificial comprende, en la dirección 5' a 3', al menos los siguientes elementos:

a) un ORF;
b) un elemento 3'-UTR como se define en la reivindicación 1;
c) una secuencia poli(A) que comprende o consiste en 64 residuos de adenosina;
d) una secuencia poli(C);
e) una secuencia tallo-bucle de histona; y
f) una secuencia poli(A) que comprende al menos 160 residuos de adenosina.

25. Molécula de ácido nucleico artificial para su uso según la reivindicación 24, que comprende además una estructura 5'-cap.

26. Molécula de ácido nucleico artificial para su uso según cualquiera de las reivindicaciones 1 a 25, donde la molécula de ácido nucleico artificial es una molécula de ARN asociada o complejada con un compuesto catiónico o policatiónico en una relación en peso seleccionada de un intervalo de aproximadamente 6:1 (p/p) a aproximadamente 0,25:1 (p/p).

27. Molécula de ácido nucleico artificial para su uso según la reivindicación 26, donde el ARN está asociado o complejado con una proteína o péptido catiónico, o con un lípido catiónico.

28. Molécula de ácido nucleico artificial para su uso según la reivindicación 26 o 27, donde el ARN está asociado o complejado con una proteína o péptido catiónico en una relación nitrógeno/fosfato ARN:compuesto catiónico o policatiónico en un rango de aproximadamente 0,1-10.

29. Vector que comprende la molécula de ácido nucleico artificial como se define en cualquiera de las reivindicaciones 1 a 28.

30. Vector según la reivindicación 29, que es un vector de ADN.

31. Vector según la reivindicación 29 o 30, que es un vector circular donde un sitio de restricción para la linealización de la molécula del vector circular está ubicado en el lado 3' de la 3'-UTR de la hebra de codificación.

32. Célula que comprende la molécula de ácido nucleico artificial como se define en cualquiera de las reivindicaciones 1 a 28 o el vector según cualquiera de las reivindicaciones 29 a 31.

33. Célula según la reivindicación 32, que es una célula de mamífero.

34. Composición farmacéutica que comprende la molécula de ácido nucleico artificial según cualquiera de las reivindicaciones 1 a 28, el vector según cualquiera de las reivindicaciones 29 a 21 o la célula según la reivindicación 32 o 33.

35. Composición farmacéutica según la reivindicación 34, que comprende además uno o más vehículos, diluyentes y/o excipientes farmacéuticamente aceptables y/o uno o más adyuvantes.

36. Vector según cualquiera de las reivindicaciones 29 a 31, célula según la reivindicación 32 o 33 o composición farmacéutica según la reivindicación 34 o 35 para uso como un medicamento, para su uso como una vacuna o para su uso en terapia génica.

37. Molécula de ácido nucleico artificial como se define según cualquiera de las reivindicaciones 1 a 28 o vector según cualquiera de las reivindicaciones 29 a 31 para su uso en un método de tratamiento o para prevenir un trastorno que comprende la transfección de una célula, donde la transfección de una célula se realiza *in vitro*/*ex vivo* y la célula transfectada se administra a un sujeto que lo necesita.

38. Método *in vitro* para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial, comprendiendo el método la etapa de asociar la molécula de ácido nucleico, preferiblemente la molécula de ARNm o el vector, con una región 3' no traducida (3'-UTR), donde la 3'-UTR comprende al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende una secuencia tallo-bucle de histona según la SEQ ID NO: 11 y donde la molécula de ácido nucleico artificial está asociada o complejada con un compuesto catiónico o policatiónico.

39. Método *in vitro* según la reivindicación 38, donde la al menos una secuencia poli(A) se produce mediante una reacción de poliadenilación enzimática.

40. Uso de una 3'-UTR para aumentar la producción de proteínas a partir de una molécula de ácido nucleico artificial, donde la 3'-UTR comprende al menos dos secuencias poli(A) separadas, donde la primera secuencia poli(A) comprende de 20 a 90 nucleótidos adenina y la segunda secuencia poli(A) comprende al menos 150 nucleótidos adenina, donde las al menos dos secuencias poli(A) están separadas por una secuencia de ácido nucleico que comprende una secuencia tallo-bucle de histona según la SEQ ID NO: 11 y donde la molécula de ácido nucleico artificial está asociada o complejada con un compuesto catiónico o policatiónico.

41. Método según la reivindicación 38 o 39, o uso según la reivindicación 40, donde la molécula de ácido nucleico artificial es una molécula de ácido nucleico artificial como se define en una cualquiera de las reivindicaciones 1 a 28.

42. Kit o kit de partes que comprende la molécula de ácido nucleico artificial como se define según cualquiera de las reivindicaciones 1 a 28, el vector según cualquiera de las reivindicaciones 29 a 31, la célula según la reivindicación 32 o 33 o la composición farmacéutica según la reivindicación 34 o 35.

GGGGCGCTGCCTACGGAGGTGGCAGCCATCTCCTTCTCGGCATCAAGCTTGAGGATGGAG
GACGCCAAGAACATCAAGAAGGGCCCGGCGCCCTTCTACCCGCTGGAGGACGGGACCGCC
GGCGAGCAGCTCCACAAGGCCATGAAGCGGTACGCCCTGGTGCCGGGCACGATCGCCTTC
ACCGACGCCCACATCGAGGTCGACATCACCTACGCGGAGTACTTCGAGATGAGCGTGCGC
CTGGCCGAGGCCATGAAGCGGTACGGCCTGAACACCAACCACCGGATCGTGGTGTGCTCG
GAGAACAGCCTGCAGTTCTTCATGCCGGTGCTGGGCGCCCTCTTCATCGGCGTGGCCGTC
GCCCCGGCGAACGACATCTACAACGAGCGGGAGCTGCTGAACAGCATGGGGATCAGCCAG
CCGACCGTGGTGTTCGTGAGCAAGAAGGGCCTGCAGAAGATCCTGAACGTGCAGAAGAAG
CTGCCCATCATCCAGAAGATCATCATCATGGACAGCAAGACCGACTACCAGGGCTTCCAG
TCGATGTACACGTTCGTGACCAGCCACCTCCCGCCGGGCTTCAACGAGTACGACTTCGTC
CCGGAGAGCTTCGACCGGGACAAGACCATCGCCCTGATCATGAACAGCAGCGGCAGCACC
GGCCTGCCGAAGGGGGTGGCCCTGCCGCACCGGACCGCCTGCGTGCGCTTCTCGCACGCC
CGGGACCCCATCTTCGGCAACCAGATCATCCCGGACACCGCCATCCTGAGCGTGGTGCCG
TTCCACCACGGCTTCGGCATGTTCACGACCCTGGGCTACCTCATCTGCGGCTTCCGGGTG
GTCCTGATGTACCGGTTCGAGGAGGAGCTGTTCCTGCGGAGCCTGCAGGACTACAAGATC
CAGAGCGCGCTGCTCGTGCCGACCCTGTTCAGCTTCTTCGCCAAGAGCACCCTGATCGAC
AAGTACGACCTGTCGAACCTGCACGAGATCGCCAGCGGGGGCGCCCCGCTGAGCAAGGAG
GTGGGCGAGGCCGTGGCCAAGCGGTTCCACCTCCCGGGCATCCGCCAGGGCTACGGCCTG
ACCGAGACCACGAGCGCGATCCTGATCACCCCCGAGGGGGACGACAAGCCGGGCGCCGTG
GGCAAGGTGGTCCCGTTCTTCGAGGCCAAGGTGGTGGACCTGGACACCGGCAAGACCCTG
GGCGTGAACCAGCGGGGCGAGCTGTGCGTGCGGGGGCCGATGATCATGAGCGGCTACGTG
AACAACCCGGAGGCCACCAACGCCCTCATCGACAAGGACGGCTGGCTGCACAGCGGCGAC
ATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTCGACCGGCTGAAGTCGCTGATC
AAGTACAAGGGCTACCAGGTGGCGCCGGCCGAGCTGGAGAGCATCCTGCTCCAGCACCCC
AACATCTTCGACGCCGGCGTGGCCGGGCTGCCGGACGACGACGCCGGCGAGCTGCCGGCC
GCGGTGGTGGTGCTGGAGCACGGCAAGACCATGACGGAGAAGGAGATCGTCGACTACGTG
GCCAGCCAGGTGACCACCGCCAAGAAGCTGCGGGGCGGCGTGGTGTTCGTGGACGAGGTC
CCGAAGGGCCTGACCGGGAAGCTCGACGCCCGGAAGATCCGCGAGATCCTGATCAAGGCC
AAGAAGGGCGGCAAGATCGCCGTGTAAGACTAGTGCATCACATTTAAAAGCATCTCAGCC
TACCATGAGAATAAGAGAAAGAAAATGAAGATCAATAGCTTATTCATCTCTTTTTCTTTT
TCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTTAATCATTTTGCCT
CTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAACCTAGATCTAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATCCCC
CCCCCCCCCCCCCCCCCCCCCCCCCCCCAAAGGCTCTTTTCAGAGCCACCAGAATT

Figura 1

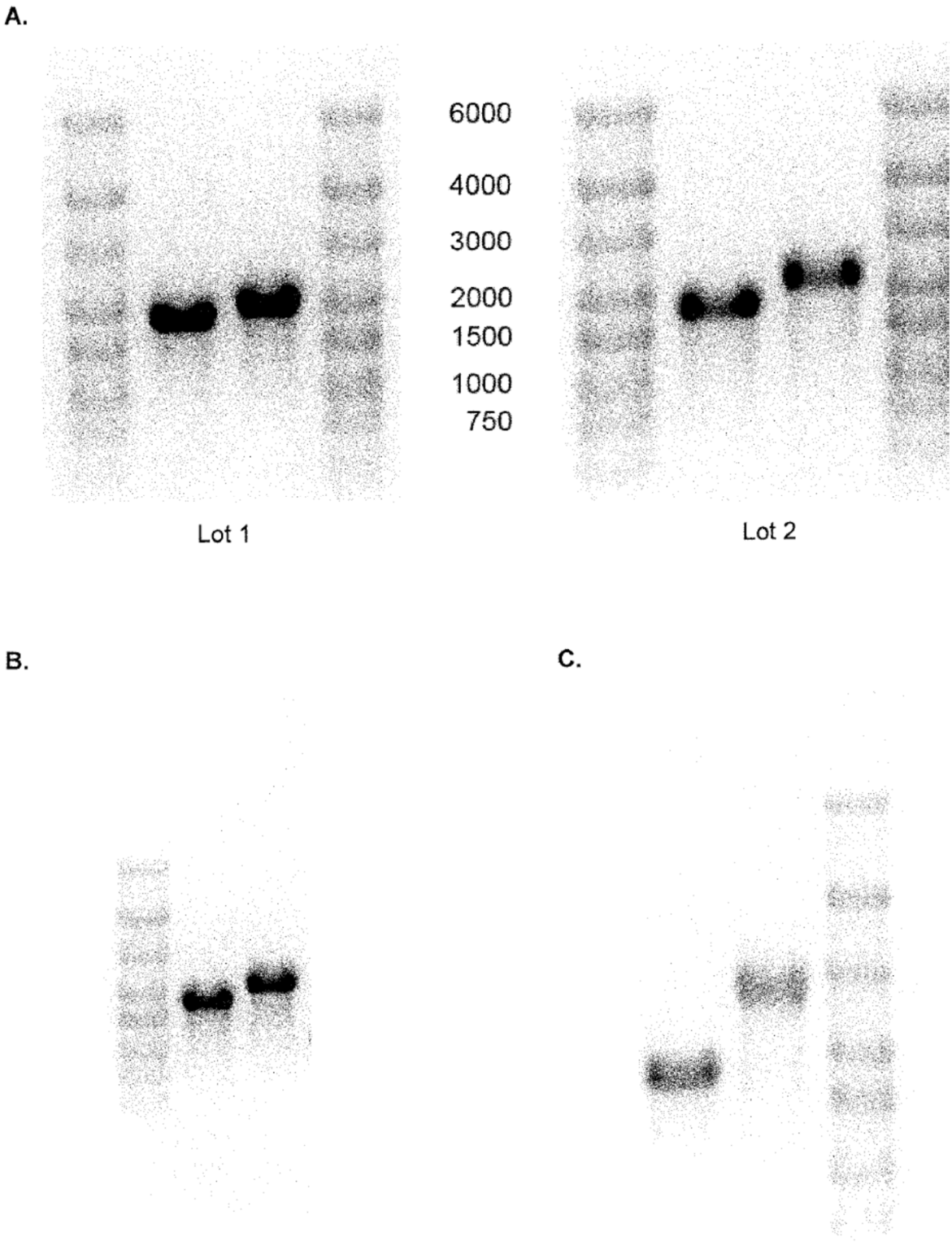

Figura 2

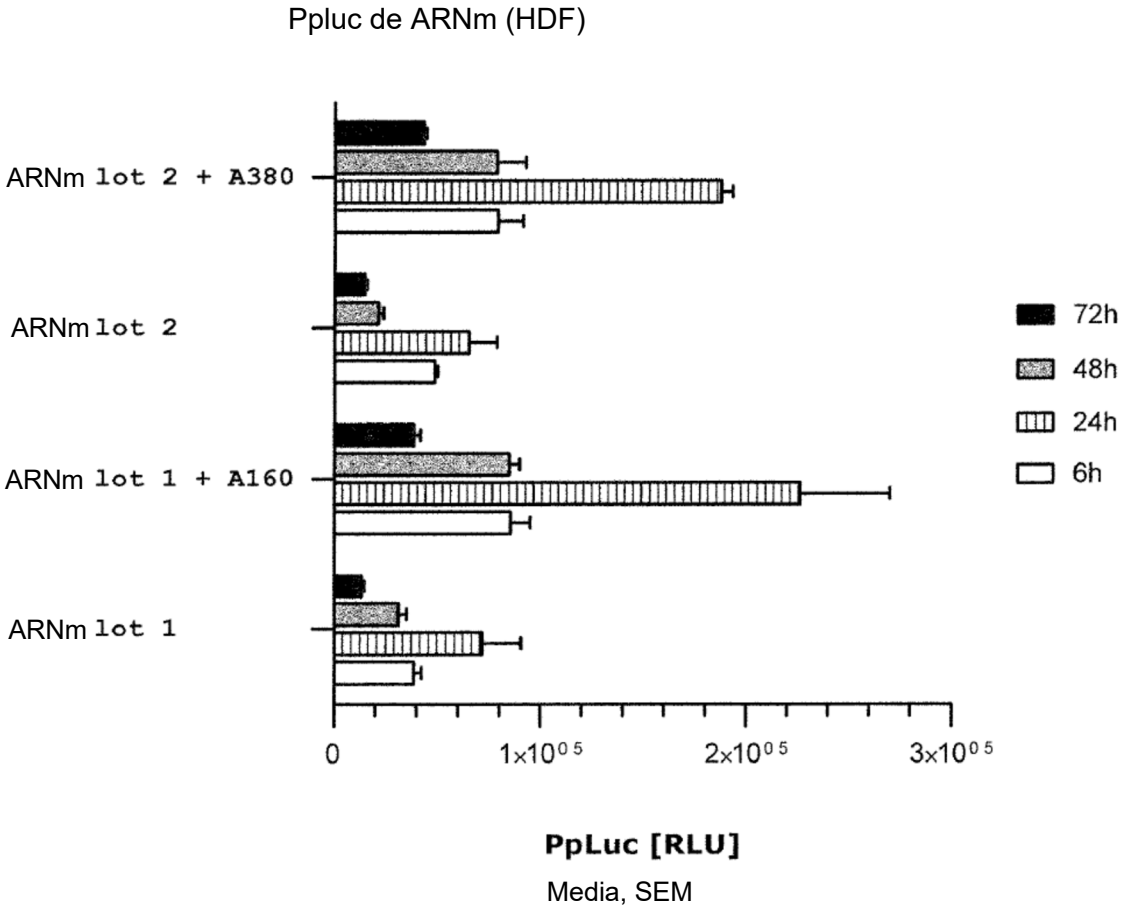
Ppluc de ARNm (HDF)
ARNm lot 2 + A380
ARNm lot 2
ARNm lot 1 + A160
ARNm lot 1
72h
48h
24h
6h
0
1x10^05
2x10^05
3x10^05
PpLuc [RLU]
Media, SEM

Figura 3

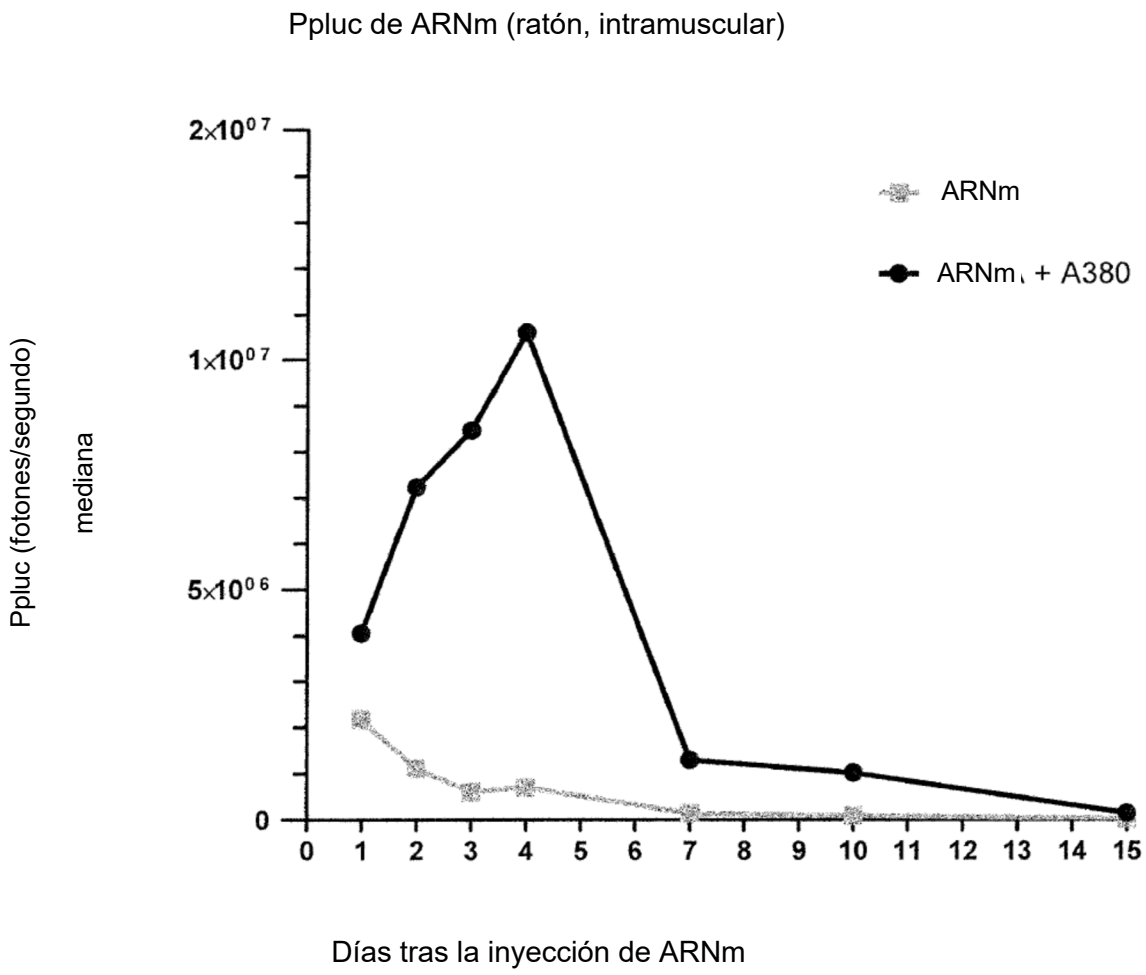
Ppluc de ARNm (ratón, intramuscular)
Ppluc (fotones/segundo)
mediana
2x10^07
1x10^07
5x10^06
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
ARNm
ARNm + A380
Días tras la inyección de ARNm

Figura 4

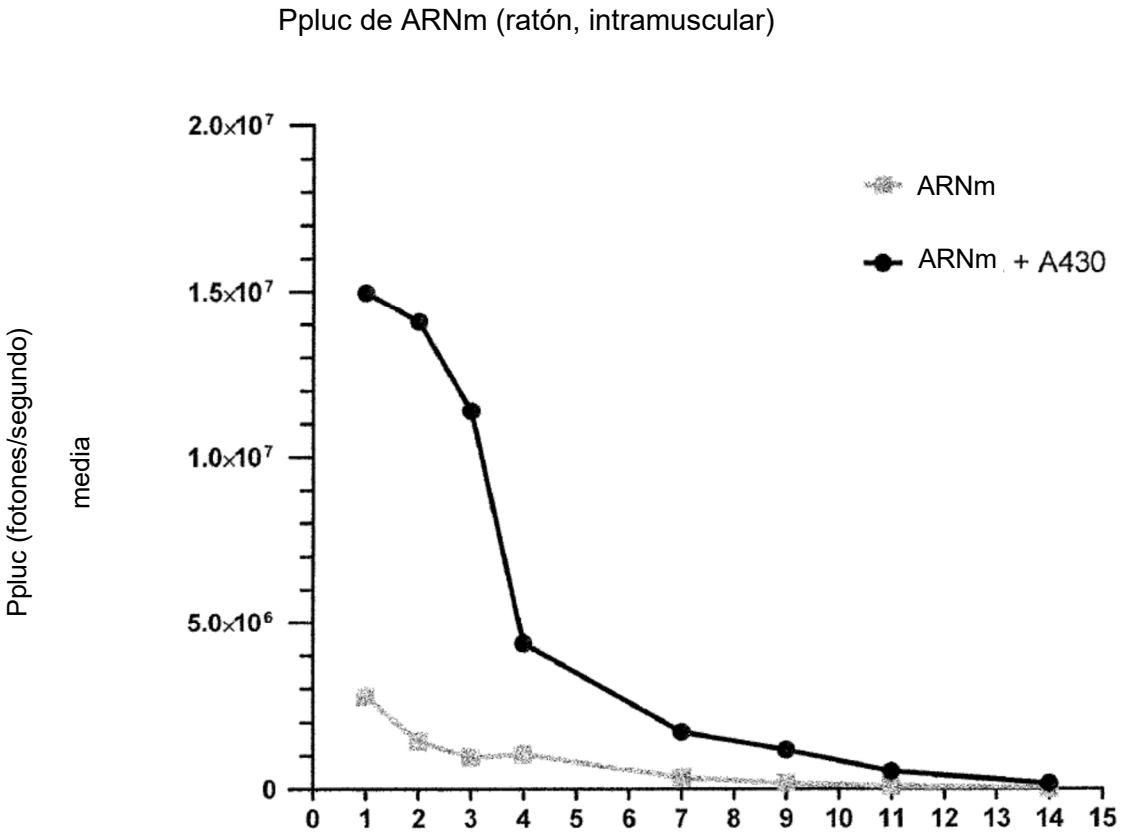
Ppluc de ARNm (ratón, intramuscular)
Ppluc (fotones/segundo)
media
2.0x10^7
1.5x10^7
1.0x10^7
5.0x10^6
0
ARNm
ARNm + A430
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Días tras la inyección de ARNm

Figura 5

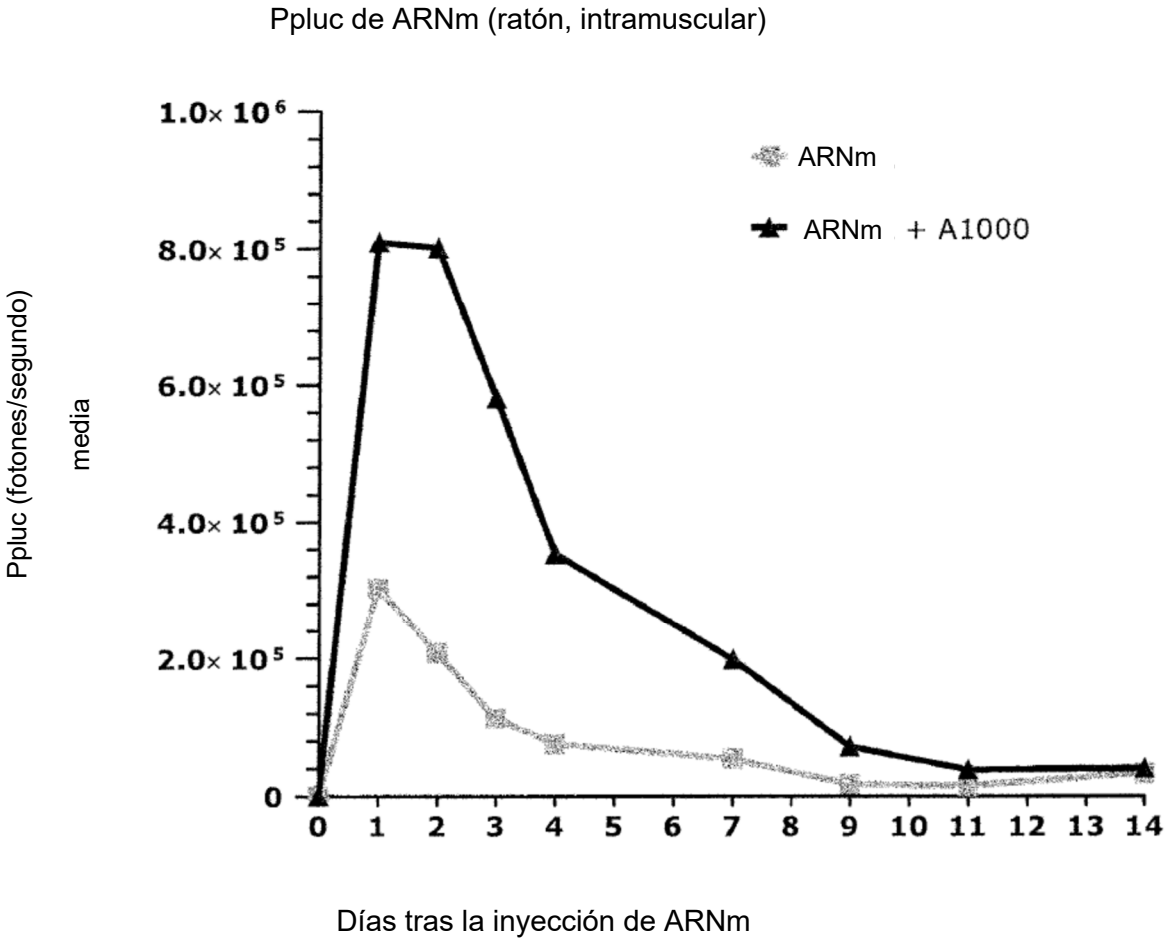
Ppluc de ARNm (ratón, intramuscular)
Ppluc (fotones/segundo)
media
1.0× 10⁶
8.0× 10⁵
6.0× 10⁵
4.0× 10⁵
2.0× 10⁵
0
ARNm
ARNm + A1000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Días tras la inyección de ARNm

Figura 6

Título HI – ARNm HA

D35 (2 semanas después reto)

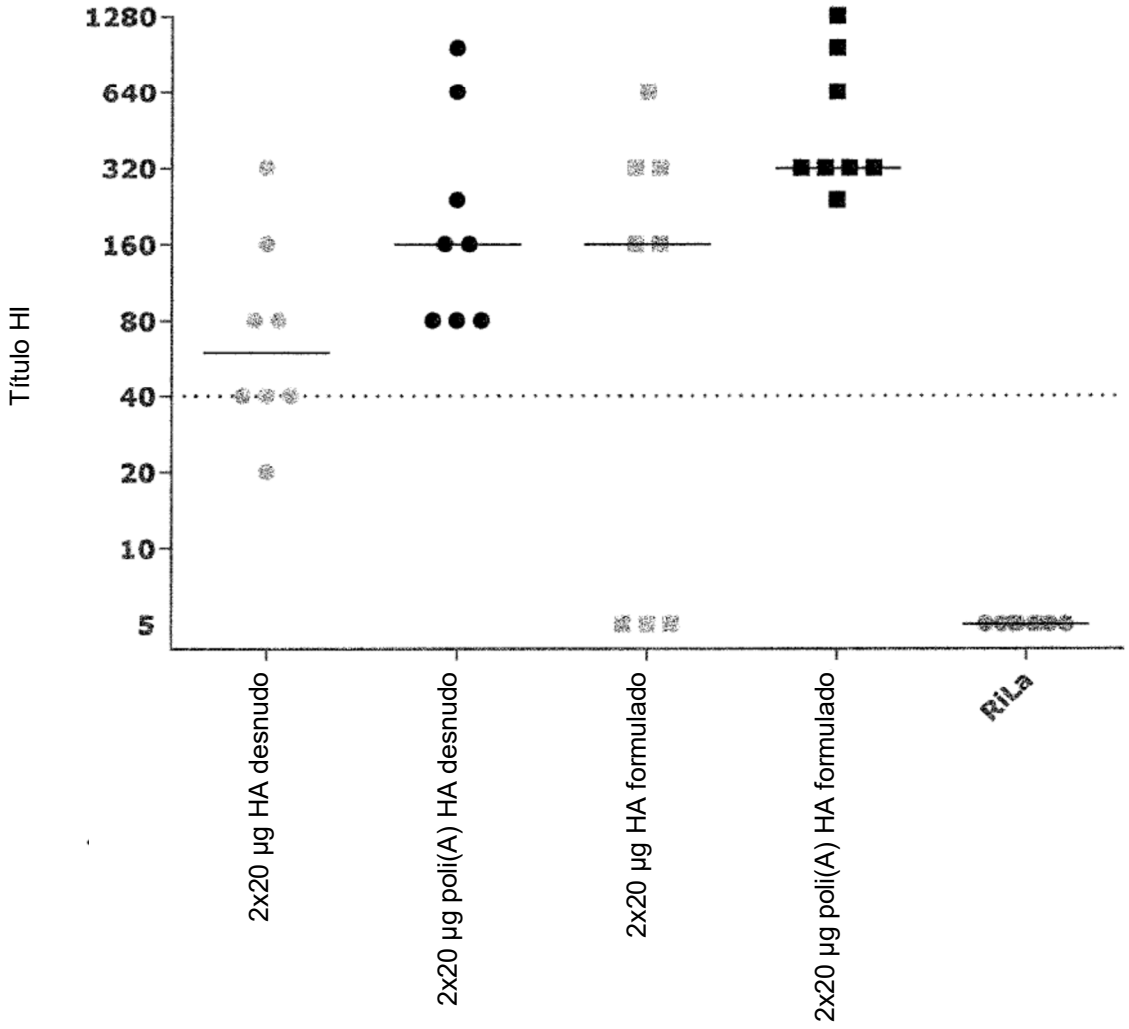

Figura 7

32L4-H1N1 (Netherlands 2009)-HA(GC)-albumin7-A64-N5-C30-histonaSL

GGGGCGCTGCCTACGGAGGTGGCAGCCATCTCCTTCTCGGCATCAAGCTTACCATGAAGGCCATCCTG
GTGGTCCTCCTGTACACCTTCGCCACCGCGAACGCCGACACGCTGTGCATCGGCTACCACGCCAACAA
CAGCACCGACACCGTGGACACCGTGCTCGAGAAGAACGTCACGGTGACCCACTCCGTGAACCTGCTGG
AGGACAAGCACAACGGGAAGCTCTGCAAGCTGCGGGGCGTCGCCCCGCTGCACCTCGGGAAGTGCAAC
ATCGCCGGCTGGATCCTGGGGAACCCGGAGTGCGAGAGCCTGTCCACCGCGAGCTCCTGGAGCTACAT
CGTGGAGACCTCCAGCTCCGACAACGGCACGTGCTACCCCGGCGACTTCATCGACTACGAGGAGCTCC
GCGAGCAGCTGAGCTCCGTGAGCTCCTTCGAGCGGTTCGAGATCTTCCCCAAGACCAGCTCCTGGCCC
AACCACGACAGCAACAAGGGGGTCACCGCCGCCTGCCCGCACGCCGGCGCGAAGTCCTTCTACAAGAA
CCTGATCTGGCTCGTGAAGAAGGGGAACAGCTACCCCAAGCTGTCCAAGAGCTACATCAACGACAAGG
GCAAGGAGGTGCTGGTCCTCTGGGGGATCCACCACCCCAGCACCTCCGCCGACCAGCAGAGCCTGTAC
CAGAACGCCGACGCCTACGTGTTCGTGGGCTCCAGCCGCTACTCCAAGAAGTTCAAGCCCGAGATCGC
CATCCGGCCGAAGGTCCGCGACCAGGAGGGCCGGATGAACTACTACTGGACGCTGGTGGAGCCCGGGG
ACAAGATCACCTTCGAGGCGACCGGCAACCTCGTGGTCCCCCGCTACGCCTTCGCCATGGAGCGGAAC
GCCGGGAGCGGCATCATCATCTCCGACACCCCCGTGCACGACTGCAACACGACCTGCCAGACCCCGAA
GGGCGCCATCAACACCAGCCTGCCCTTCCAGAACATCCACCCCATCACGATCGGGAAGTGCCCCAAGT
ACGTGAAGTCCACCAAGCTGCGCCTCGCGACCGGCCTGCGGAACGTCCCGAGCATCCAGTCCCGCGGG
CTGTTCGGCGCCATCGCCGGGTTCATCGAGGGCGGCTGGACCGGGATGGTGGACGGCTGGTACGGGTA
CCACCACCAGAACGAGCAGGGCAGCGGGTACGCCGCCGACCTCAAGTCCACGCAGAACGCGATCGACG
AGATCACCAACAAGGTGAACAGCGTCATCGAGAAGATGAACACCCAGTTCACCGCCGTGGGCAAGGAG
TTCAACCACCTGGAGAAGCGGATCGAGAACCTGAACAAGAAGGTCGACGACGGCTTCCTCGACATCTG
GACGTACAACGCCGAGCTGCTGGTGCTCCTGGAGAACGAGCGCACCCTGGACTACCACGACTCCAACG
TGAAGAACCTCTACGAGAAGGTCCGGAGCCAGCTGAAGAACAACGCCAAGGAGATCGGGAACGGCTGC
TTCGAGTTCTACCACAAGTGCGACAACACCTGCATGGAGTCCGTGAAGAACGGGACCTACGACTACCC
CAAGTACAGCGAGGAGGCCAAGCTGAACCGCGAGGAGATCGACGGCGTGAAGCTCGAGTCCACGCGGA
TCTACCAGATCCTGGCGATCTACAGCACCGTCGCCAGCTCCCTGGTGCTCGTGGTCAGCCTGGGGGCC
ATCTCCTTCTGGATGTGCAGCAACGGCTCCCTGCAGTGCCGCATCTGCATCTGACCACTAGTGCATCA
CATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAATAGCTTATTCATC
TCTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTTAATCATTTT
GCCTCTTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAACCTAGATCTAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCAAAGGCTCTTTTCAGAGCCACCAGAATT

Figura 8

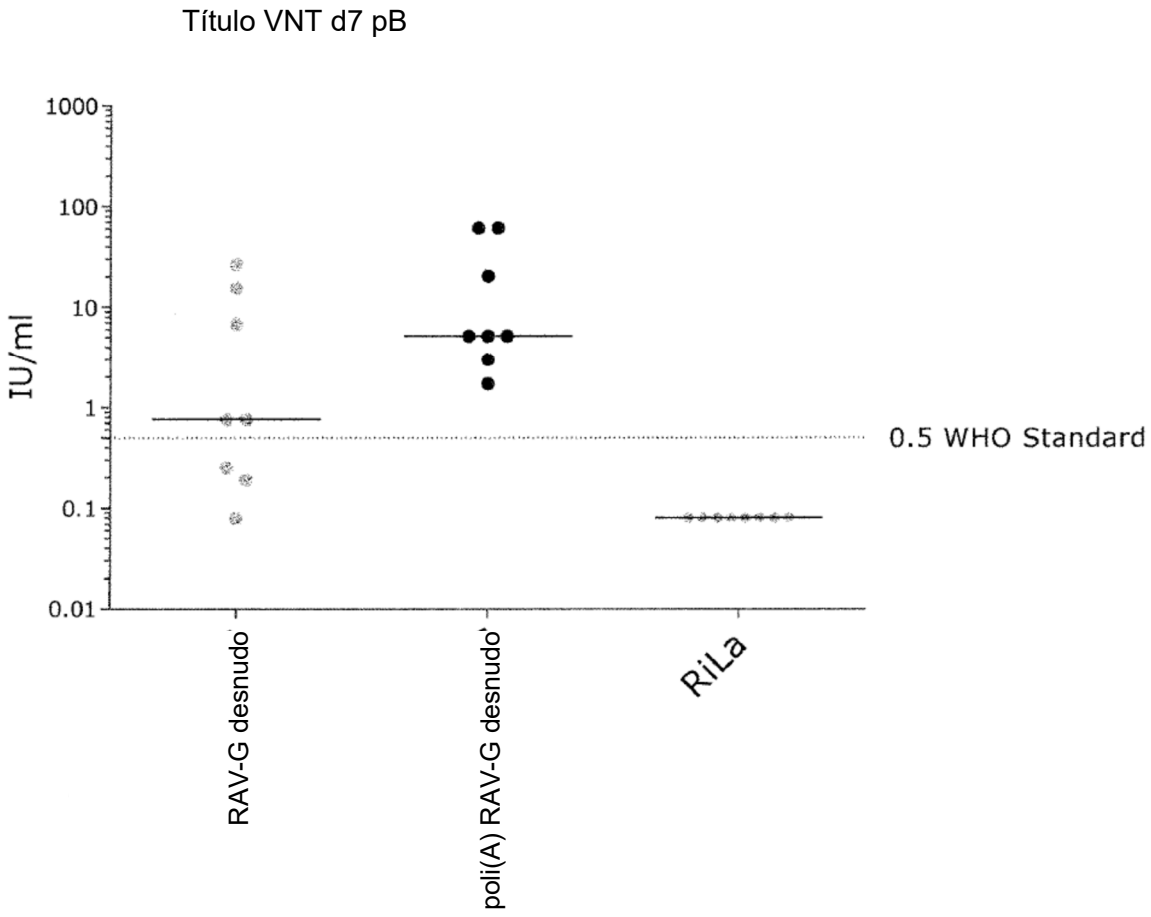
Título VNT d7 pB
1000
100
10
1
0.1
0.01
IU/ml
0.5 WHO Standard
RAV-G desnudo
poli(A) RAV-G desnudo
RiLa

Figura 9

32L4-RAV-HA(GC)-albumin7-A64-N5-C30-histonaSL

GGGGCGCTGCCTACGGAGGTGGCAGCCATCTCCTTCTCGGCATCAAGCTTACCATGGTGCCCCAGGCC
CTGCTCTTCGTCCCGCTGCTGGTGTTCCCCCTCTGCTTCGGCAAGTTCCCCATCTACACCATCCCCGA
CAAGCTGGGGCCGTGGAGCCCCATCGACATCCACCACCTGTCCTGCCCCAACAACCTCGTGGTCGAGG
ACGAGGGCTGCACCAACCTGAGCGGGTTCTCCTACATGGAGCTGAAGGTGGGCTACATCAGCGCCATC
AAGATGAACGGGTTCACGTGCACCGGCGTGGTCACCGAGGCGGAGACCTACACGAACTTCGTGGGCTA
CGTGACCACCACCTTCAAGCGGAAGCACTTCCGCCCCACGCCGGACGCCTGCCGGGCCGCCTACAACT
GGAAGATGGCCGGGGACCCCCGCTACGAGGAGTCCCTCCACAACCCCTACCCCGACTACCACTGGCTG
CGGACCGTCAAGACCACCAAGGAGAGCCTGGTGATCATCTCCCCGAGCGTGGCGGACCTCGACCCCTA
CGACCGCTCCCTGCACAGCCGGGTCTTCCCCGGCGGGAACTGCTCCGGCGTGGCCGTGAGCTCCACGT
ACTGCAGCACCAACCACGACTACACCATCTGGATGCCCGAGAACCCGCGCCTGGGGATGTCCTGCGAC
ATCTTCACCAACAGCCGGGGCAAGCGCGCCTCCAAGGGCAGCGAGACGTGCGGGTTCGTCGACGAGCG
GGGCCTCTACAAGTCCCTGAAGGGGGCCTGCAAGCTGAAGCTCTGCGGCGTGCTGGGCCTGCGCCTCA
TGGACGGGACCTGGGTGGCGATGCAGACCAGCAACGAGACCAAGTGGTGCCCCCCCGGCCAGCTGGTC
AACCTGCACGACTTCCGGAGCGACGAGATCGAGCACCTCGTGGTGGAGGAGCTGGTCAAGAAGCGCGA
GGAGTGCCTGGACGCCCTCGAGTCCATCATGACGACCAAGAGCGTGTCCTTCCGGCGCCTGAGCCACC
TGCGGAAGCTCGTGCCCGGGTTCGGCAAGGCCTACACCATCTTCAACAAGACCCTGATGGAGGCCGAC
GCCCACTACAAGTCCGTCCGCACGTGGAACGAGATCATCCCGAGCAAGGGGTGCCTGCGGGTGGGCGG
CCGCTGCCACCCCCACGTCAACGGGGTGTTCTTCAACGGCATCATCCTCGGGCCCGACGGCAACGTGC
TGATCCCCGAGATGCAGTCCAGCCTGCTCCAGCAGCACATGGAGCTGCTGGTCTCCAGCGTGATCCCG
CTCATGCACCCCCTGGCGGACCCCTCCACCGTGTTCAAGAACGGGGACGAGGCCGAGGACTTCGTCGA
GGTGCACCTGCCCGACGTGCACGAGCGGATCAGCGGCGTCGACCTCGGCCTGCCGAACTGGGGGAAGT
ACGTGCTGCTCTCCGCCGGCGCCCTGACCGCCCTGATGCTGATCATCTTCCTCATGACCTGCTGGCGC
CGGGTGAACCGGAGCGAGCCCACGCAGCACAACCTGCGCGGGACCGGCCGGGAGGTCTCCGTGACCCC
GCAGAGCGGGAAGATCATCTCCAGCTGGGAGTCCTACAAGAGCGGCGGCGAGACCGGGCTGTGAGGAC
TAGTGCATCACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAATAG
CTTATTCATCTCTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTT
TAATCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAACCTAGATCTAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATCCCCCC
CCCCCCCCCCCCCCCCCCCCCCCCCCAAAGGCTCTTTTCAGAGCCACCAGAATT

Figura 10