

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 823 173**

51 Int. Cl.:

C12N 15/86 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.01.2017 PCT/US2017/015130**

87 Fecha y número de publicación internacional: **03.08.2017 WO17132376**

96 Fecha de presentación y número de la solicitud europea: **26.01.2017 E 17705220 (6)**

97 Fecha y número de publicación de la concesión europea: **15.07.2020 EP 3408397**

54 Título: **Promotor híbrido y usos del mismo**

30 Prioridad:

27.01.2016 US 201662388391 P

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

06.05.2021

73 Titular/es:

**JUST BIOTHERAPEUTICS, INC. (100.0%)
401 Terry Avenue North
Seattle, Washington 98109, US**

72 Inventor/es:

**MCGREW, JEFFREY T.;
SMIDT, PAULINE S. y
ONG, E-CHING**

74 Agente/Representante:

CONTRERAS PÉREZ, Yahel

ES 2 823 173 T3

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

DESCRIPCIÓN

Promotor híbrido y usos del mismo

- 5 La presente solicitud reivindica prioridad con respecto a la solicitud de patente provisional de los Estados Unidos con N.º de serie 62/388.391, presentada en la Oficina de Marcas y Patentes de los Estados Unidos el 27 de enero de 2016.

Listado de secuencias

- 10 La presente solicitud contiene un listado de secuencias que se ha presentado electrónicamente en formato ASCII y se incorpora al presente documento por referencia en su totalidad. Dicha copia en formato ASCII, creada el 25 de enero de 2017, se denomina JUST0051_SL.txt y tiene un tamaño de 25.408 bytes.

Antecedentes de la invención

15

1. Campo de la invención

Esta invención se refiere a la producción recombinante de proteínas en células de mamífero.

20 **2. Comentarios sobre la técnica relacionada**

La producción a escala industrial de proteínas terapéuticas mediante la expresión recombinante en cultivo de células de mamífero, es una iniciativa relativamente reciente que se esfuerza por lograr la mayor eficiencia de expresión recombinante posible y características de calidad del producto que cumplan con las directrices legislativas o las superen.

Para la producción de proteínas recombinantes, se han utilizado diferentes líneas celulares de mamífero, incluyendo varias líneas de células de ovario de hámster chino (CHO, *Chinese Hamster Ovary*). (Véase, por ejemplo, Hu et al., Overexpressing Cyclin D1 in a Eukaryotic Cell line, patente de Estados Unidos N.º 6.210.924; Goepfert et al., Protein Expression from Multiple Nucleic Acids, patente de Estados Unidos N.º 8.771.988; y Wurm, F.M., CHO quasispecies-Implications for Manufacturing Processes, Processes 1:296-311; doi:10.3390/pr1030296 (2013)).

Para potenciar la expresión recombinante en una variedad de células hospedadoras, se han utilizado promotores híbridos que combinan elementos potenciadores y promotores de diferentes fuentes genéticas. (Por ejemplo, Harvey, Hybrid promoters, WO 2008/020960 A1; US 2008/12492 A1).

Estudios previos han demostrado que el promotor del CMV humano y el promotor del CMV murino pueden impulsar una expresión de alto nivel de expresión heteróloga en las células de mamíferos. (Mizushima y Nagata, pEF-BOS, a powerful mammalian expression vector. Nucl. Acids Res., 18(17):5322 (1990); Masayuki y Tanaka, The CMV Enhancer Stimulates Expression of Foreign Genes from the Human EF-1α Promoter, Analytical Biochemistry 247:179-181 (1997); Chattellard, P. et al., The Lupac bifunctional marker and its use in protein production, WO 2006/058900 A1; Chattellard, P. et al., Expression vectors comprising the mCMV IE2 promoter, patente de Estados Unidos N.º 7.824.907; Hjelmstrom et al., Single IFN-beta fused to a mutated IgG Fc fragment, WO 2009/053368 A1; Gaucher et al., Cell line having a high transcription activity for the production of proteins, in particular therapeutic proteins, WO 2008/096070 A2; Flannery et al., Recombinant lubricin molecules and uses thereof, patente de Estados Unidos N.º 7.642.236; Mosyak et al., Method for identifying or designing a candidate agent that interacts with LINGO-1 polypeptide using a LINGO-1 three-dimensional structure, patente de Estados Unidos N.º 7.693.698; Mosyak et al., WO 2007/092370 A1).

La adición de intrones también puede aumentar la expresión de proteínas heterólogas (Lacy-Hulbert, A. et al., Interruption of coding sequences by heterologous introns can enhance the functional expression of recombinant genes, Gene Ther. 8(8):649-653 (2001); Brinster, R.L., et al., Introns increase transcriptional efficiency in transgenic mice, Proc. Natl. Acad. Sci. USA 85: 836-40 (1988)). Se ha demostrado que el promotor y los intrones del EF 1 alfa (factor 1 de elongación alfa) de ser humano y de hámster promueven de manera eficaz la expresión génica (Mizushima y Nagata, pEF-BOS, a powerful mammalian expression vector, Nucl. Acids Res. 18(17):5322 (1990); Running Deer, J., y Allison, D. S., High-level expression of proteins in mammalian cells using Transcription Regulatory sequences from Chinese Hamster Ovary EF-1α Gene, Biotechnology Progress 20:880-889 (2004); Allison, D.S., Recombinant method for making multimeric proteins, WO 2006/063292 A1; Orlova et al., Improved elongation factor-1α-based vectors for stable high-level expression of heterologous proteins in Chinese hamster ovary cells, BMC Biotechnology 14:56 (2014)). En el documento WO 2011/110864 A1 se desvelan vectores que tienen un promotor de CMV humano que contiene un intrón A. En el documento D1 también se desvela un intrón del EF 1 alfa de ser humano.

Las combinaciones de promotores de CMV humano o murino combinadas con intrones del EF 1 alfa (factor 1 de elongación alfa) humano, mostraron una mejora significativa sobre las construcciones de expresión sin intrones (Kim, S.-Y. et al., El primer intrón del factor 1 de elongación alfa (EF-1 alfa) de ser humano mejora sumamente la expresión de genes extraños del promotor del citomegalovirus murino, J. Biotechnol. 93(2):183-87 (2002)).

El represor Tet (TetR), que está codificado por el transposón bacteriano, Tn10, se ha utilizado para regular la expresión génica en células de mamífero. TetR se une a una secuencia de ADN (TetO) en ausencia de tetraciclina. Al unirse a la tetraciclina, TetR sufre un cambio conformacional que anula la unión de TetR con TetO. Yao et al., habían demostrado que el promotor potenciador de CMV humano podía regularse mediante tetraciclina incorporando una secuencia TetO entre la caja TATA y el sitio de inicio de la transcripción (TSS, *transcriptional start site*). (Véase, Yao et al., Tetracycline repressor, tetR, rather than the tetR mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells, Hum. Gene Ther. 9(13):1939-50 (1998); Yao et al., patente de Estados Unidos N.º 5.972.650; Yao et al., WO99/00510).

10

Es necesario mejorar la expresión recombinante de las proteínas por parte de las células de mamífero en grandes lotes o en cultivos continuos para mantener la fuerte producción de productos biológicos en una variedad de líneas celulares de mamífero, que proporciona la presente invención.

15 Resumen de la invención

La presente invención se refiere a un promotor híbrido para regular la expresión recombinante de una o más proteínas de interés. El promotor híbrido incluye: (i) una secuencia potenciadora de citomegalovirus murino (mCMV), que comprende un elemento potenciador de mCMV (mCMV-E) y una secuencia promotora de CMV (CMV-P) en su extremo 3', unida operativamente en 5' a una secuencia intrónica del factor 1 de elongación alfa (EF 1 alfa) de rata; (ii) una primera secuencia líder intermedia unida operativamente, en 3' a la secuencia promotora de CMV de la secuencia potenciadora de mCMV, y en 5' a la secuencia intrónica del EF 1 alfa de rata; y (iii) una segunda secuencia líder unida operativamente en 3' a la secuencia intrónica del EF 1 alfa. Esta ingeniosa combinación de una secuencia potenciadora de mCMV con una secuencia intrónica del EF1 alfa de rata, en el promotor híbrido, facilita la expresión recombinante de proteínas de interés, particularmente por células CHO, con altos títulos y alta productividad específicos adecuados para la producción industrial de moléculas biológicas, tales como, pero sin limitación, una proteína de unión a antígeno, una inmunoglobulina, un anticuerpo o fragmento de anticuerpo, (por ejemplo, como terapéutica humana para tratar enfermedades), y fue generalmente superior a la de otros promotores sometidos a ensayo en algunas líneas celulares.

Por consiguiente, la presente invención también se refiere a un casete de expresión que contiene el inventivo promotor híbrido, unido operativamente en 5' a un marco abierto de lectura que codifica una proteína exógena de interés; y un sitio de poliadenilación unido operativamente en 3' al marco abierto de lectura.

Otro aspecto de la invención es un vector de expresión recombinante que comprende el inventivo casete de expresión que contiene el promotor híbrido.

También se incluye en la invención una célula hospedadora de mamífero que contiene el inventivo vector de expresión recombinante que contiene el casete de expresión con el promotor híbrido, tal como una célula procedente de una célula de ovario de hámster chino (CHO, por las siglas del inglés *Chinese Hamster Ovary*), por ejemplo, una célula CHO-K1, una célula DXB11 o una célula DG44.

La presente invención también se refiere a un método para producir una proteína de interés, *in vitro*, que implica cultivar la célula hospedadora de mamífero que contiene el inventivo vector de expresión, en un medio acuoso en condiciones fisiológicas que permitan la expresión de la proteína de interés; y recuperar la proteína de interés del medio. La presente invención también incluye un método para producir una proteína de interés que implica la expresión inducible por tetraciclina de la proteína de interés.

El resumen anterior no pretende definir con detalle cada uno de los aspectos de la invención y en otros apartados, tal como en el de Descripción detallada de realizaciones, se describen realizaciones adicionales. Se pretende que todo el documento se relacione como una divulgación unificada.

Además de lo anterior, la invención desvela realizaciones de la invención de menor alcance en cualquier forma que las variaciones definidas por los párrafos específicos anteriores.

55 Breve descripción de los dibujos

La Figura 1 muestra mapas esquemáticos de algunas construcciones de expresión ejemplares.

La figura 2 muestra un mapa esquemático del Vector D ejemplar.

La Figura 3A-B muestra el título de la proteína Fc-A producida por células DXB11 (Figura 3A) y la productividad específica (Figura 3B) en medio de cultivo ProCHOTM4 que contiene metotrexato (MTX) 150 nM el día 8. Cada barra de error se construyó utilizando 1 error estándar de la media.

La Figura 4A-B muestra el título de la proteína Fc-A (Figura 4A) y la productividad específica producidos por células DXB11 (Figura 4B) en medio PowerCHOTM 2 con MTX 500 nM el Día 8. Cada barra de error se construyó utilizando 1 error estándar de la media. La Figura 5A-B muestra el título de la proteína Fc-A (Figura 5A) y la productividad específica (Figura 5B) de células DG44 en medio PowerCHOTM 2 con MTX y DXB11 1 µM en medio Excell302 con

MTX 500 nM, el día 10. Cada barra de error se construyó utilizando 1 error estándar de la media.

La Figura 6 muestra un mapa esquemático de pJV56 con el inventivo promotor híbrido potenciador de mCMV/intrón de rEF-1 α que dirige la expresión de LacZ.

5 La Figura 7 muestra un mapa esquemático de pJV57, que era igual que pJV56, excepto que parte del promotor de mCMV (mCMV-P) se reemplazó por el promotor de CMV humano (hCMV-P) y las secuencias TetO.

La Figura 8 muestra un mapa esquemático de pJV59, que era igual que pJV56, pero con un promotor de CMV humano (hCMV-P) optimizado; según Patwardhan et al., High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis, Nature Biotechnology 27(12):1173-75 (2009), sustituido por parte de la secuencia promotora de mCMV (mCMV-P).

10 La Figura 9 muestra un mapa esquemático de pJV60, que era igual que pJV57, excepto por cambios en las secuencias TetO (manteniendo la unión de TetR) para emparejarse con la secuencia promotora de CMV humano (hCMV-P) optimizada.

La Figura 10 a continuación muestra una representación esquemática del operador Tet-promotor de CMV humano (TetO; "Tet") insertado en la secuencia promotora de mCMV ("mCMVP"), en 3' con respecto a la secuencia del elemento potenciador de mCMV ("MCVE"). Esta inserción reemplazó parte del promotor de mCMV por una secuencia promotora de hCMV (hCMVP). La posición relativa de la primera secuencia líder también se representa como "TPL" (por las siglas del inglés *tripartite leader*, líder tripartita).

15 La Figura 11 muestra una comparación esquemática de las secuencias de ADN de un segmento de pJV56 (SEQ ID NO: 16) y de un segmento de pJV57 (SEQ ID NO: 15). El segmento de pJV57 mostrado en SEQ ID NO: 15, incluye un segmento (SEQ ID NO: 9) que incorpora una secuencia promotora parcial de hCMV (SEQ ID NO: 24) y, en 3' con respecto a SEQ ID NO: 24, una secuencia de TetO (SEQ ID NO: 23, que contiene una secuencia de TetO más pequeña SEQ ID NO: 29). Las flechas que se muestran en la secuencia de TetO (SEQ ID NO: 29) indican sitios de unión de TetR palindrómicos. La secuencia promotora de mCMV (mCMV-P) y la secuencia promotora de hCMV (hCMV-P) son las secuencias que van desde la caja TATA hasta el sitio de inicio de la transcripción. SEQ ID NO: 25 es el extremo 3' de una secuencia de mCMV-P que se encuentra tanto en pJV57 como en pJV56, y en pJV57 se encuentra en 3' con respecto a SEQ ID NO: 9. El inicio de la transcripción es el resto de guanina en la subsecuencia taccg en 3' de SEQ ID NO: 25. Observe que la secuencia de TetO insertada (SEQ ID NO: 23) sin duda tiene impacto sobre el sitio de inicio de la transcripción, ya que normalmente el sitio de inicio de la transcripción está ~ 30 bases en 3' con respecto a T en 5' de la caja TATA. La posición relativa de la primera secuencia líder también se representa como "TPL", y se muestran algunos restos de nucleótidos en el extremo 5' de la secuencia TPL.

20 La Figura 12 muestra una comparación esquemática de segmentos de las secuencias de ADN de pJV57 y pJV60. En el segmento mostrado de pJV60 (SEQ ID NO: 18), las secuencias de TetO en el segmento correspondiente de pJV57 (SEQ ID NO: 17) se cambiaron para emparejarse con las secuencias que aumentaban la expresión alrededor del sitio de inicio de la transcripción en el estudio de Patwardhan et al., High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis, Nature Biotechnology 27(12):1173-75 (2009).

25 La Figura 13 muestra una comparación esquemática de segmentos de las secuencias de ADN de pJV56 y pJV59. En este segmento de pJV59 (SEQ ID NO: 20), las secuencias promotoras de hCMV optimizadas en la secuencia correspondiente de pJV56 (SEQ ID NO: 19) reemplazaron parte de la secuencia promotora de mCMV (mCMV-P) en una variación del promotor híbrido potenciador de mCMV/intrón del EF la de rata de la invención.

30 La Figura 14 muestra la regulación de la expresión específica de β -galactosidasa para cada uno de los vectores indicados. Células CHO T-RExTM que expresan TetR, se transfectaron transitoriamente con pJV56, pJV57, pJV59, pJV60, o con el plásmido de control pcDNA5/TO/lacZ. Las células se incubaron durante 24 horas, y después, se trataron durante 24 horas con tetraciclina (Tet) o se dejaron sin tratar. Las células se sometieron a lisis y LacZ se analizó enzimáticamente. El factor de diferencia entre (+Tet) en comparación con (-Tet) se muestra encima de las barras, excepto para el control positivo (pcDNA5/TO/lacZ).

La Figura 15 muestra un mapa esquemático de pJV40, incluyendo un marco abierto de lectura para TetR.

35 La Figura 16 ilustra la expresión de LacZ (proteína beta-galactosidasa) por células CHO-K1/TetR (clon 3E7) que expresan TetR, que se transfectaron transitoriamente con pJV56, pJV57, pJV59, pJV60, o con el plásmido de control pcDNA5/TO/lacZ. Las células se incubaron durante 24 horas, y después, se trataron durante 24 horas con tetraciclina (Tet) o se dejaron sin tratar. Las células se sometieron a lisis y LacZ se analizó enzimáticamente. La diferencia de veces entre (+Tet) en comparación con (-Tet), se muestra encima de las barras.

40 La Figura 17 ilustra la expresión de LacZ (proteína beta-galactosidasa) por células CHO-K1/TetR (clon 3F9) que expresan TetR, que se transfectaron transitoriamente con pJV56, pJV57, pJV59, pJV60, o con el plásmido de control pcDNA5/TO/lacZ. Las células se incubaron durante 24 horas, y después, se trataron durante 24 horas con tetraciclina (Tet) o se dejaron sin tratar. Las células se sometieron a lisis y LacZ se analizó enzimáticamente. La diferencia de veces entre (+Tet) en comparación con (-Tet), se muestra encima de las barras.

45 La Figura 18 ilustra la expresión de LacZ (proteína beta-galactosidasa) por células CHO-K1/TetR (clon 4G2) que expresan TetR, que se transfectaron transitoriamente con pJV56, pJV57, pJV59, pJV60, o con el plásmido de control pcDNA5/TO/lacZ. Las células se incubaron durante 24 horas, y después, se trataron durante 24 horas con tetraciclina (Tet) o se dejaron sin tratar. Las células se sometieron a lisis y LacZ se analizó enzimáticamente. La diferencia de veces entre (+Tet) en comparación con (-Tet), se muestra encima de las barras.

Descripción detallada de realizaciones

65

Los encabezados de los apartados utilizados en el presente documento tienen fines organizativos solamente, y no deben considerarse como limitantes de la materia objeto descrita.

Definiciones

5

A menos que se defina de otro modo en el presente documento, los términos científicos y técnicos utilizados en relación con la presente solicitud, tendrán los significados que comúnmente entienden los expertos familiarizados con la materia. Además, a no ser que el contexto requiera otra cosa, los términos en singular incluirán las pluralidades y los términos en plural incluirán el singular. Por tanto, como se usa en la presente memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un/uno/una" y "el/la", incluyen referencias en plural a menos que el contexto indique claramente lo contrario. Por ejemplo, la referencia a "una proteína" incluye una pluralidad de proteínas; la referencia a "una célula" incluye poblaciones de una pluralidad de células.

10

La presente invención se refiere a un promotor híbrido para regular la expresión recombinante de una o más proteínas de interés. El inventivo promotor incluye una secuencia potenciadora de mCMV unida operativamente a una secuencia intrónica del EF 1 alfa de rata. El término "mCMV", utilizado indistintamente con "muCMV", se refiere a citomegalovirus murino, que es un herpesvirus de la subfamilia *Betaherpesviridae*. El mCMV es un virus de ADN bicatenario con envoltura, con especificidad de hospedador por ratones. De manera similar, el término "hCMV", o indistintamente "huCMV", se refiere a citomegalovirus humano.

15

La "secuencia potenciadora de mCMV", con respecto a la presente invención, incluye:

(i) un elemento potenciador de mCMV ("mCMV-E"); y, en el extremo 3' de la secuencia potenciadora de mCMV, (ii) una secuencia promotora de CMV ("CMV-P"; es decir, un segmento de secuencia de nucleótidos que comienza en, e incluye, la caja TATA a través del sitio de inicio de la transcripción); la secuencia promotora de CMV puede proceder de mCMV, hCMV, CMV de simio, CMV de rata, o de cualquier otra variedad de CMV, o puede ser una versión optimizada de una CMV-P, que se funcional para permitir la transcripción en células de mamífero, tales como células CHO.

25

La expresión "recombinante" indica que el material (por ejemplo, un ácido nucleico o un polipéptido) se ha alterado artificial o sintéticamente (es decir, de manera no naturalmente) por intervención humana. La alteración puede realizarse en el material dentro, o eliminado de, su entorno o estado natural. Por ejemplo, un "ácido nucleico recombinante" es aquel que se produce recombinando ácidos nucleicos, por ejemplo, durante la clonación, barajado de ADN u otros procedimientos biológicos moleculares bien conocidos. Se encuentran ejemplos de dichos procedimientos de biología molecular en Maniatis et al., *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982). Una "molécula de ADN recombinante" está compuesta por segmentos de ADN unidos entre sí mediante dichas técnicas de biología molecular. La expresión "proteína recombinante" o "polipéptido recombinante" como se usa en el presente documento, se refiere a una molécula de proteína que se expresa utilizando una molécula de ADN recombinante. Una "célula hospedadora recombinante" es una célula que contiene y/o expresa un ácido nucleico recombinante.

30

35

La expresión "de origen natural", como se usa a lo largo de la memoria descriptiva en relación con materiales biológicos, tales como polipéptidos, ácidos nucleicos, células hospedadoras y similares, se refiere a materiales que se encuentran en la naturaleza.

40

La expresión "secuencia de control" o "señal de control" se refiere a una secuencia de polinucleótidos que puede, en una célula hospedadora particular, influir en la expresión y el procesamiento de secuencias codificantes con las que está ligada. La naturaleza de dichas secuencias de control puede depender del organismo hospedador. En realizaciones particulares, las secuencias de control de procariotas pueden incluir un promotor, un sitio de unión al ribosoma y una secuencia de terminación de la transcripción. Las secuencias de control de eucariotas pueden incluir promotores que comprenden uno o una pluralidad de sitios de reconocimiento para factores de transcripción, secuencias o elementos potenciadores de la transcripción, sitios de poliadenilación y secuencias de terminación de la transcripción. Las "secuencias de control" pueden incluir secuencias líder y/o secuencias compañeras de fusión. Los promotores y potenciadores consisten en series cortas de ADN que interaccionan específicamente con proteínas celulares implicadas en la transcripción (Maniatis, et al., *Science* 236: 1237 (1987)). Se han aislado elementos promotores y potenciadores de una variedad de fuentes eucariotas, incluidos genes en levaduras, células y virus de insectos y mamíferos (elementos de control análogos, es decir, promotores, también se encuentran en procariotas). La selección de un promotor y potenciador particular depende del tipo de célula que se utilice para expresar la proteína de interés. Algunos promotores y potenciadores eucariotas tienen una amplia gama de hospedadores, mientras que otros son funcionales en un subconjunto limitado de tipos de células (para una revisión, véase Voss, et al., *Trends Biochem. Sci.*, 11:287 (1986) y Maniatis, et al., *Science* 236:1237 (1987)).

45

50

55

Un "promotor" es una región de ADN que incluye un sitio en el cual la ARN polimerasa se une para iniciar la transcripción de ARN mensajero por uno o más genes estructurales cadena abajo. Los promotores se encuentran cerca de los sitios de inicio de la transcripción de los genes, sobre la misma cadena y cadena arriba en el ADN (hacia

60

65

la región 5' de la cadena codificante). Normalmente los promotores tienen una longitud de aproximadamente 100-1 000 pb.

En general, un "potenciador" es una región corta (de 50 a 1 500 pb) de ADN que puede unirse con una o más proteínas activadoras (factores de transcripción) para activar la transcripción de un gen.

Una secuencia potenciadora de mCMV útil en el inventivo promotor híbrido, comprende una secuencia de nucleótidos que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a la secuencia de nucleótidos de la SEQ ID NO: 2, que contiene una secuencia promotora de CMV (CMV-P) en su extremo 3' (secuencia subrayada en negrita en la SEQ ID NO: 2) comenzando en la posición del nucleótido 588 (que incluye una caja TATA y un sitio de inicio de la transcripción, es decir, TATAAGAGG CGCGA C CAGCG TCGG TACCG//SEQ ID NO:28; subrayado en negrita en la SEQ ID NO: 2 a continuación; la SEQ ID NO: 28 también comprende la SEQ ID NO: 26):

GTCAACAGGA AAGTTCCATT GGAGCCAAGT ACATTGAGTC

AATAGGGACT TTCCAATGGG TTTTGCCCAG TACATAAGGT CAATGGGAGG

TAAGCCAATG GGTTTTTCCC ATTACTGGCA CGTATACTGA GTCATTAGGG

ACTTTCCAAT GGGTTTTGCC CAGTACATAA GGTC AATAGG GGTGAATCAA

CAGGAAAGTC CCATTGGAGC CAAGTACACT GAGTCAATAG GGACTTTCCA

TTGGGTTTTG CCCAGTACAA AAGGTCAATA GGGGGTGAGT CAATGGGTTT

TTCCATTAT TGGCACGTAC ATAAGGTCAA TAGGGGTGAG TCATTGGGTT

TTTCCAGCCA ATTTAATTAA AACGCCATGT ACTTTCCCAC CATTGACGTC

AATGGGCTAT TGAAACTAAT GCAACGTGAC CTTTAAACGG TACTTTCCCA

TAGCTGATTA ATGGGAAAGT ACCGTTCTCG AGCCAATACA CGTCAATGGG

AAGTGAAAGG GCAGCCAAA CGTAACACCG CCCC GGTTTT CCCCTGGAAA

TTCCATATTG GCACGCATTC TATTGGCTGA GCTGCGTTCT ACGTGGG **TAT**

AAGAGGCGCG ACCAGCGTCG GTACCG//SEQ ID NO:2.

15

Otra secuencia potenciadora de mCMV ejemplar, útil en el inventivo promotor híbrido, comprende una secuencia de nucleótidos que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a la secuencia de nucleótidos de la SEQ ID NO: 33, que contiene una secuencia promotora de CMV en su extremo 3' (secuencia subrayada en negrita en la SEQ ID NO: 33) comenzando en la posición 588, que incluye una caja TATA y un sitio de inicio de la transcripción, es decir, TATATAAGCAGAGCTCGTTTAGTGAACCGTCAGTTCGTCCTAGACGCCAACC G//SEQ ID NO:35; secuencia subrayada en negrita en la SEQ ID NO: 33 a continuación; la SEQ ID NO: 35 también comprende la SEQ ID NO: 24):

20

GTCAACAGGAAAGTTCCATTGGAGCCAAGTACATTGAGTCAATAG
 GGACTTTCCAATGGGTTTTGCCAGTACATAAGGTCAATGGGAGGTAAGCCA
 ATGGGTTTTTCCCATTACTGGCACGTATACTGAGTCATTAGGGACTTTCCAAT
 GGGTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCC
 ATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCATTGGGTTTTGCCAG
 TACAAAAGGTCAATAGGGGGTGAATCAACAGGAAAGTCCC
 ACATAAGGTCAATAGGGGTGAGTCATTGGGTTTTTCCAGCCAATTTAATTAAA
 ACGCCATGTACTTTCCACCATTGACGTCAATGGGCTATTGAACTAATGCAA
 CGTGACCTTTAAACGGTACTTTCCCATAGCTGATTAATGGGAAAGTACCGTTC
 TCGAGCCAATACAGTCAATGGGAAGTGAAAGGGCAGCCAAAACGTAAACAC
 CGCCCCGGTTTTCCCTGGAAATTCATATTGGCACGCATTCTATTGGCTGAG
 CTGCGTTCTACGTGGGTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGTT
CGTCTCTAGACGCCAACCG//SEQ ID NO:33.

En algunas realizaciones útiles, en las que se desea la expresión inducible por tetraciclina, el inventivo promotor híbrido comprende una o más secuencias de TetO unidas operativamente en 3' a la secuencia potenciadora de mCMV, insertadas dentro de la secuencia promotora de CMV (CMV-P) en el extremo 3' de la secuencia potenciadora de mCMV. Una secuencia de "TetO" significa una secuencia de nucleótidos, que conserva la capacidad de unirse a la proteína represora de tetraciclina (TetR). La secuencia de TetO se coloca de manera que, en presencia de la proteína TetR, hay unión de TetR con la secuencia de TetO, interrumpiendo así la transcripción a un nivel detectable, en comparación con un control que no tiene TetO en el promotor que dirige la transcripción de un gen de interés, o con un control que no tiene TetR. Los ejemplos de la secuencia de TetO incluyen una secuencia que tiene al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100% de identidad de secuencia con la SEQ ID NO:29 (TCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGA//SEQ ID NO: 29) o que tiene al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % de identidad de secuencia con la SEQ ID NO: 34 (CTCCCTATCAGTGATCAGTTCCCTATCAGTGATAGAGA//SEQ ID NO: 34). En algunas realizaciones útiles, puede haber restos de nucleótidos adicionales 5' o 3' a la(s) secuencia(s) de TetO, o puede haber secuencias enlazadoras de nucleótidos intermedias, siempre que no se elimine la capacidad de unirse a la proteína TetR.

En el contexto de la invención, la expresión "represión" o "reprimido/a, se refiere a la interferencia de la transcripción de un gen de interés (que codifica una proteína de interés), que ocurre cuando la proteína TetR se une a un sitio de unión a TetO en el promotor que dirige la expresión del gen de interés, dando como resultado una disminución de la expresión de la proteína de interés por parte de la(s) célula(s) (que es una célula o células que expresa(n) TetR). Se dice que la expresión de un gen de interés o de una proteína de interés está "desreprimida", cuando, en presencia de tetraciclina en el medio, la expresión de la proteína de interés es al menos 1,5 veces superior a los niveles basales de expresión de la(s) célula(s) en ausencia de tetraciclina en el medio.

"Tetraciclina" significa tetraciclina o un análogo de tetraciclina, tal como doxiciclina, anhidrotetraciclina, minociclina, oxitetraciclina, metaciclina, clortetraciclina o COL-3 (tetraciclina-3 químicamente modificada).

Una secuencia intrónica del EF 1 alfa de rata útil en el inventivo promotor híbrido, comprende una secuencia de nucleótidos que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a la secuencia de nucleótidos de SEQ ID NO: 4:

GTGAGTGGCGGGTGTGGCTTCCGCGGGCCCCGGAGCTGGAGCCCTG
 CTCTGAGCGGGCCGGGCTGATATGCGAGTGTCTCCGCAGGGTTTAGCTGTG
 AGCATTCCCACTTCGAGTGGCGGGCGGTGCGGGGGTGAGAGTGCAGGCCTA
 GCGGCAACCCCGTAGCCTCGCCTCGTGTCCGGCTTGAGGCCTAGCGTGGTGTC
 CGCCGCCGCGTGCCACTCCGGCCGCACTATGCGTTTTTTGTCCTTGCTGCCCTC
 GATTGCCTTCCAGCAGCATGGGCTAACAAAGGGAGGGTGTGGGGCTCACTCT
 TAAGGAGCCCATGAAGCTTACGTTGGATAGGAATGGAAGGGCAGGAGGGGC
 GACTGGGGCCCGCCCGCCTTCGGAGCACATGTCCGACGCCACCTGGATGGGG
 CGAGGCCTGTGGCTTCCGAAGCAATCGGGCGTGAGTTTAGCCTACCTGGGCC
 ATGTGGCCCTAGCACTGGGCACGGTCTGGCCTGGCGGTGCCGCGTTCCCTTGC
 CTCCAACAAGGGTGAGGCCGTCCCGCCCGGCACCAGTTGCTTGCGCGGAAA
 GATGGCCGCTCCCGGGGCCCTGTTGCAAGGAGCTCAAATGGAGGACGCGGC
 AGCCCGGTGGAGCGGGCGGGTGAGTCACCCACACAAAGGAAGAGGGCCTTG
 CCCCTCGCCGGCCGCTGCTTCTGTGACCCCGTGGTCTATCGGCCGCATAGTC
 ACCTCGGGCTTCTCTTGAGCACCGCTCGTCGCGGGCGGGGGAGGGGATCTAA
 TGGCGTTGGAGTTTGTTCACATTTGGTGGGTGGAGACTAGTCAGGCCAGCCTG
 GCGCTGGAAGTCATTCTTGAATTTGCCCTTTGAGTTTGGAGCGAGGCTAAT
 TCTCAAGCCTCTTAGCGGTTCAAAGGTATTTTCTAAACCCGTTTCCAG//SEQ ID
 NO:4.

El inventivo promotor híbrido también incluye una primera secuencia líder intermedia, unida operativamente en 3' al potenciador de mCMV y en 5' a la secuencia intrónica del EF 1 alfa en una orientación operativa. La primera secuencia líder intermedia tiene una longitud de aproximadamente 10 a 200 restos de nucleótidos, más preferiblemente una longitud de aproximadamente 10 a 60 restos de nucleótidos o incluso más preferiblemente una longitud de aproximadamente 20 a 50 restos de nucleótidos, y carece de cualquier estructura secundaria con una estabilidad superior a 20 kcal y carece de sitios de inicio de traducción ATG. Un ejemplo de una primera secuencia líder útil, es una secuencia líder no traducida (5'UTR, *untranslated*), procedente de la secuencia líder tripartita (TPL) de adenovirus, es decir, la secuencia de nucleótidos de SEQ ID NO: 3: TACCTCTTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGG//SEQ ID NO:3.

El inventivo promotor híbrido también incluye una segunda secuencia líder unida operativamente en 3' a la secuencia intrónica del EF 1 alfa en una orientación operable. La segunda secuencia líder tiene una longitud de aproximadamente 5 a 200 restos de nucleótidos, más preferiblemente una longitud de aproximadamente 10 a 200 restos de nucleótidos o incluso más preferiblemente una longitud de aproximadamente 10 a 150 restos de nucleótidos y carece de cualquier estructura secundaria con una estabilidad superior a 20 kcal y carece de sitios de inicio de traducción ATG. Un ejemplo de una segunda secuencia líder útil, es otra región de secuencia líder no traducida (5'UTR), procedente de la secuencia líder tripartita (TPL) de adenovirus, es decir, la secuencia de nucleótidos de SEQ ID NO: 5:

CTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGG
 ATCGGAAACCCGTCGGCCTCCGAACGGTACTCCGCCACCGAGGGACCTGAGC
 GAGTCCGCATCGACCGGATCGGAAAACCTC//SEQ ID NO:5.

Para otros ejemplos de secuencias líder útiles y principios de diseño, véase, por ejemplo, Mignone, F. et al., Untranslated regions of mRNAs. Genome Biology, 3(3): reviews0004.1-0004.10 (2002)), incorporado en el presente documento por referencia en su totalidad. Cualquier secuencia líder adecuada, con las características anteriormente mencionadas, puede usarse en la práctica de la presente invención.

5

Una realización útil del inventivo promotor híbrido es un promotor que comprende una secuencia de nucleótidos que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a la secuencia de nucleótidos de la SEQ ID NO: 1, a continuación, que contiene los nucleótidos 4067-4682 de mCMV del GenBank: L06816.1; los nucleótidos 22137-21728 del intrón del EF 1 alfa de rata del Genbank: AC158987.3; y una primera

10

secuencia líder intermedia subrayada en letras minúsculas (SEQ ID NO: 3); y una segunda secuencia líder subrayada en letras minúsculas en cursiva (SEQ ID NO: 5):

GTCAACAGGAAAGTTCCATTGGAGCCAAGTACATTGAGTCAATAG
GGACTTTCCAATGGGTTTTGCCAGTACATAAGGTCAATGGGAGGTAAGCCA
ATGGGTTTTTCCCATTACTGGCACGTATACTGAGTCATTAGGGACTTTCCAAT
GGGTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCC
ATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCATTGGGTTTTGCCAG
TACAAAAGGTCAATAGGGGGTGAGTCAATGGGTTTTTCCCATTATTGGCACGT

ACATAAGGTCAATAGGGGTGAGTCATTGGGTTTTTCCAGCCAATTTAATTA
 ACGCCATGTA CTTTCCCACCATTGACGTCAATGGGCTATTGAACTAATGCAA
 CGTGACCTTTAAACGGTACTTTCCCATAGCTGATTAATGGGAAAGTACCGTTC
 TCGAGCCAATACACGTCAATGGGAAAGTGAAAGGGCAGCCAAAACGTAACAC
 CGCCCCGGTTTTCCCCTGGAAATTCCATATTGGCACGCATTCTATTGGCTGAG
 CTGCGTTCTACGTGGGTATAAGAGGGCGCGACCAGCGTCGGTACCGtacctcttcgca
tcgctgtctcgagggccagctgttgggGTGAGTGGCGGGTGTGGCTTCCGCGGGCCCCGGAG
 CTGGAGCCCTGCTCTGAGCGGGCCGGGCTGATATGCGAGTGTCTGCCAGG
 GTTTAGCTGTGAGCATTCCACTTCGAGTGGCGGGCGGTGCGGGGGTGAGAG
 TGCGAGGCCTAGCGGCAACCCCGTAGCCTCGCCTCGTGTCCGGCTTGAGGCCT
 AGCGTGGTGTCCGCCGCCGCGTGCCACTCCGGCCGCACTATGCGTTTTTTGTC
 CTTGCTGCCCTCGATTGCCTTCCAGCAGCATGGGCTAACAAAGGGAGGGTGT
 GGGGCTCACTCTTAAGGAGCCCATGAAGCTTACGTTGGATAGGAATGGAAGG
 GCAGGAGGGGCGACTGGGGCCCCGCCGCTTCGGAGCACATGTCCGACGCCA
 CCTGGATGGGGCGAGGCCTGTGGCTTTCCGAAGCAATCGGGCGTGAGTTTAG
 CCTACCTGGGCCATGTGGCCCTAGCACTGGGCACGGTCTGGCCTGGCGGTGC
 CGCGTTCCCTTGCCCTCCAACAAGGGTGAGGCCGTCCCGCCCGGCACCAGTTG
 CTTGCGCGGAAAGATGGCCGCTCCCGGGGCCCTGTTGCAAGGAGCTCAAAT
 GGAGGACGCGGCAGCCCGGTGGAGCGGGCGGGTGAGTACCCACACAAAGG
 AAGAGGGCCTTGCCCCTCGCCGGCCGCTGCTTCTGTGACCCCGTGGTCTATC
 GGCCGCATAGTCACCTCGGGCTTCTCTTGAGCACCGCTCGTCGCGGCGGGGG
 GAGGGGATCTAATGGCGTTGGAGTTTGTTCACATTTGGTGGGTGGAGACTAGT
 CAGGCCAGCCTGGCGCTGGAAGTCATTCTTGGAATTTGCCCTTTGAGTTTGG
 AGCGAGGCTAATTCTCAAGCCTCTTAGCGGTTCAAAGGTATTTTCTAAACCCG
 TTTCCAGGtcgagggtgaggacaaacttctcggggtttccagtaacttggatcggaacccgleggcctccgaacgg
tactcgccaccgaggacactgagcgagtccgcategaccggateggaaaacctc//SEQ ID NO:1.

5 Como ejemplos de realizaciones útiles adicionales del inventivo promotor híbrido se incluyen (i) la SEQ ID NO: 30 (secuencia promotora híbrida que incluye TetO, como en pJV57), (ii) la SEQ ID NO: 31 (promotor híbrido que incluye un elemento potenciador de mCMV con secuencia promotora de hCMV optimizada, como en pJV59) y (iii) la SEQ ID NO: 32 (promotor híbrido que incluye un elemento potenciador de mCMV con secuencia promotora de hCMV optimizada y secuencia de TetO optimizada, como en pJV60):

GTCAACAGGAAAGTTCCATTGGAGCCAAGTACATTGAGTCAATAG
 GGACTTTCCAATGGGTTTTGCCAGTACATAAGGTCAATGGGAGGTAAGCCA

ATGGGTTTTTCCCATTACTGGCACGTATACTGAGTCATTAGGGACTTTCCAAT
GGGTTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCC
ATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCATTGGGTTTTGCCAG
TACAAAAGGTCAATAGGGGGTGAGTCAATGGGTTTTTCCCATTATTGGCACGT
ACATAAGGTCAATAGGGGTGAGTCATTGGGTTTTTCCAGCCAATTTAATTA
ACGCCATGTACTTTCCCACCATTGACGTCAATGGGCTATTGAACTAATGCAA
CGTGACCTTTAAACGGTACTTTCCCATAGCTGATTAATGGGAAAGTACCGTTC
TCGAGCCAATACAGTCAATGGGAAGTGAAAGGGCAGCCAAAACGTAACAC
CGCCCCGGTTTTCCCCTGGAAATTCCATATTGGCACGCATTCTATTGGCTGAG
CTGCGTTCTACGTGGGTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATC
TCCCTATCAGTGATAGAGATCGTCGACGAGCTCAGCGTCGGTACCGTACCTCT
TCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTGGCGGGTGTGGC
TTCCGCGGGCCCCGGAGCTGGAGCCCTGCTCTGAGCGGGCCGGGCTGATATG
CGAGTGTTCGTCAGGGTTTAGCTGTGAGCATTCCCCTCGAGTGGCGGGC
GGTGCAGGGGTGAGAGTGCAGGGCCTAGCGGCAACCCCGTAGCCTCGCCTCG
TGTCCGGCTTGAGGCCTAGCGTGGTGTCCGCCGCCGCGTGCCACTCCGGCCGC
ACTATGCGTTTTTTGTCTTGCTGCCCTCGATTGCCTTCCAGCAGCATGGGCTA
ACAAAGGGAGGGTGTGGGGCTACTCTTAAGGAGCCCATGAAGCTTACGTTG
GATAGGAATGGAAGGGCAGGAGGGGCGACTGGGGCCCCCGCCCTTCGGAG
CACATGTCCGACGCCACCTGGATGGGGCGAGGCCTGTGGCTTTCCGAAGCAA
TCGGGCGTGAGTTTAGCCTACCTGGGCCATGTGGCCCTAGCACTGGGCACGGT
CTGGCCTGGCGGTGCCGCGTTCCCTTGCCCTCCCAACAAGGGTGAGGCCGTCCC
GCCCCGCACCAGTTGCTTGCGCGGAAAGATGGCCGCTCCCGGGGCCCTGTTG
CAAGGAGCTCAAAATGGAGGACGCGGCAGCCCGGTGGAGCGGGCGGGTGAG
TCACCCACACAAAGGAAGAGGGCCTTGCCCCTCGCCGGCCGCTGCTTCCTGT
GACCCCGTGGTCTATCGGCCGCATAGTCACCTCGGGCTTCTCTTGAGCACCGC
TCGTCGCGGGCGGGGGGAGGGGATCTAATGGCGTTGGAGTTTGTTCACATTG
GTGGGTGGAGACTAGTCAGGCCAGCCTGGCGCTGGAAGTCATTCTTGGAATT
TGCCCCTTTGAGTTTGGAGCGAGGCTAATTCTCAAGCCTCTTAGCGGTTCAA
GGTATTTTCTAAACCCGTTTCCAGCTCGCGGTTGAGGACAACTCTTCGCGGT
CTTTCCAGTACTCTTGGATCGGAAACCCGTCGGCCTCCGAACGGTACTCCGCC
ACCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTC//SEQ ID
NO:30;

GTCAACAGGAAAGTTCCATTGGAGCCAAGTACATTGAGTCAATAG
 GGACTTTCCAATGGGTTTTGCCAGTACATAAGGTCAATGGGAGGTAAGCCA
 ATGGGTTTTTCCCATTACTGGCACGTATACTGAGTCATTAGGGACTTTCCAAT
 GGGTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCC
 ATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCATTGGGTTTTGCCAG
 TACAAAAGGTCAATAGGGGGTGAGTCAATGGGTTTTTCCCATTATTGGCACGT
 ACATAAGGTCAATAGGGGTGAGTCATTGGGTTTTTCCAGCCAATTTAATTTAA
 ACGCCATGTACTTTCCACCATTGACGTCAATGGGCTATTGAACTAATGCAA
 CGTGACCTTTAAACGGTACTTTCCCATAGCTGATTAATGGGAAAGTACCGTTC
 TCGAGCCAATACACGTCAATGGGAAGTGAAAGGGCAGCCAAAACGTAACAC
 CGCCCCGGTTTTCCCCTGGAAATTCCATATTGGCACGCATTCTATTGGCTGAG
 CTGCGTTCTACGTGGGTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGTTCG
 TCTCTAGACGCCAACCGCCTCTTCCGCATCGCTGTCTGCGAGGGCCAGCTGTT
 GGGGTGAGTGGCGGGTGTGGCTTCCGCGGGCCCCGGAGCTGGAGCCCTGCTC
 TGAGCGGGCCGGGCTGATATGCGAGTGTCTCCGCAGGGTTTAGCTGTGAGC
 ATTCCCACTTCGAGTGGCGGGCGGTGCGGGGGTGAGAGTGCAGGCCTAGCG
 GCAACCCCGTAGCCTCGCCTCGTGTCCGGCTTGAGGCCTAGCGTGGTGTCCGC
 CGCCGCGTGCCACTCCGGCCGCACTATGCGTTTTTTGTCTTGCTGCCCTCGAT
 TGCCTTCCAGCAGCATGGGCTAACAAAGGGAGGGTGTGGGGCTCACTCTTAA
 GGAGCCCATGAAGCTTACGTTGGATAGGAATGGAAGGGCAGGAGGGGGCGAC
 TGGGGCCCCGCCCTTCGGAGCACATGTCCGACGCCACCTGGATGGGGCGA
 GGCTGTGGCTTTCCGAAGCAATCGGGCGTGAGTTTAGCCTACCTGGGCCATG
 TGGCCCTAGCACTGGGCACGGTCTGGCCTGGCGGTGCCGCTTCCCTTGCCTC
 CCAACAAGGGTGAGGCCGTCCCGCCCGGCACCAGTTGCTTGCGCGGAAAGAT
 GGCCGCTCCCGGGGCCCTGTTGCAAGGAGCTCAAATGGAGGACGCGGCAGC
 CCGGTGGAGCGGGCGGGTGAGTACCCACACAAAGGAAGAGGGCCTTGCCC
 CTCGCCGGCCGCTGCTTCTGTGACCCCGTGGTCTATCGGCCGCATAGTCACC
 TCGGGCTTCTCTTGAGCACCGCTCGTCGCGGGGGGAGGGGATCTAATGG
 CGTTGGAGTTTGTTACATTTGGTGGGTGGAGACTAGTCAGGCCAGCCTGGCG
 CTGGAAGTCATTCTTGGAATTTGCCCTTTGAGTTTGGAGCGAGGCTAATTCT
 CAAGCCTCTTAGCGGTTCAAAGGTATTTCTAAACCCGTTTCCAGCTCGCGGT
 TGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGGATCGGAAACCCGTCG
 GCCTCCGAACGGTACTCCGCCACCGAGGGACCTGAGCGAGTCCGCATCGACC
 GGATCGGAAAACCTC//SEQ ID NO:31;Y

GTCAACAGGAAAGTTCCATTGGAGCCAAGTACATTGAGTCAATAG
GGACTTTCCAATGGGTTTTGCCAGTACATAAGGTCAATGGGAGGTAAGCCA
ATGGGTTTTTCCCATTACTGGCACGTATACTGAGTCATTAGGGACTTTCCAAT
GGGTTTTGCCAGTACATAAGGTCAATAGGGGTGAATCAACAGGAAAGTCCC
ATTGGAGCCAAGTACACTGAGTCAATAGGGACTTTCCATTGGGTTTTGCCAG
TACAAAAGGTCAATAGGGGGTGAAGTCAATGGGTTTTTCCCATTATTGGCACGT
ACATAAGGTCAATAGGGGTGAGTCATTGGGTTTTTCCAGCCAATTTAATTTAA
ACGCCATGTACTTTCCACCATTGACGTCAATGGGCTATTGAACTAATGCAA
CGTGACCTTTAAACGGTACTTTCCCATAGCTGATTAATGGGAAAGTACCGTTC
TCGAGCCAATACAGTCAATGGGAAGTGAAAGGGCAGCCAAAACGTAACAC
CGCCCCGGTTTTCCCTGGAAATTCATATTGGCACGCATTCTATTGGCTGAG
CTGCGTTCTACGTGGGTATATAAGCAGAGCTCTCCCTATCAGTGATCAGTTCC
TCCCTATCAGTGATAGAGATCGTCGACGAGCTCAGCGTCGGTACCGTACCTCT
TCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTGGCGGGTGTGGC
TTCCGCGGGCCCCGGAGCTGGAGCCCTGCTCTGAGCGGGCCGGGCTGATATG
CGAGTGTCGTCCGCAGGGTTTAGCTGTGAGCATTCCCACTTCGAGTGGCGGGC
GGTGCGGGGGTGAGAGTGCGAGGCCTAGCGGCAACCCCGTAGCCTCGCCTCG
TGTCCGGCTTGAGGCCTAGCGTGGTGTCCGCCGCCGCGTGCCACTCCGGCCGC
ACTATGCGTTTTTTGTCTTGCTGCCCTCGATTGCCTTCCAGCAGCATGGGCTA
ACAAAGGGAGGGTGTGGGGCTCACTCTTAAGGAGCCCATGAAGCTTACGTTG
GATAGGAATGGAAGGGCAGGAGGGGCGACTGGGGCCCGCCCGCCTTCGGAG
CACATGTCCGACGCCACCTGGATGGGGCGAGGCCTGTGGCTTTCCGAAGCAA
TCGGGCGTGAGTTTAGCCTACCTGGGCCATGTGGCCCTAGCACTGGGCACGGT
CTGGCCTGGCGGTGCCGCTTCCCTTGCCTCCCAACAAGGGTGAGGCCGTCCC
GCCCCGGCACCAAGTTGCTTGCGCGGAAAGATGGCCGCTCCCGGGGCCCTGTTG
CAAGGAGCTCAAATGGAGGACGCGGCAGCCCGGTGGAGCGGGCGGGTGAG
TCACCCACACAAAGGAAGAGGGCCTTGCCCCCTCGCCGGCCGCTGCTTCCCTGT
GACCCCGTGGTCTATCGGCCGCATAGTCACCTCGGGCTTCTCTTGAGCACCGC
TCGTGCGGGCGGGGGGAGGGGATCTAATGGCGTTGGAGTTTGTTACATTTG
GTGGGTGGAGACTAGTCAGGCCAGCCTGGCGCTGGAAGTCATTCTTGAAT
TGCCCCTTGAGTTTGGAGCGAGGCTAATTCTCAAGCCTCTTAGCGGTTCAA
GGTATTTTCTAAACCCGTTTCCAGCTCGCGGTTGAGGACAACTCTTCGCGGT
CTTCCAGTACTCTTGATCGGAAACCCGTCGGCCTCCGAACGGTACTCCGCC

ACCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTC// SEQ ID

NO:32.

Las expresiones "en combinación operativa", "en orden operativo" y "unido(a) operativamente", como se usan en el presente documento, se refieren al enlace de secuencias de ácido nucleico de tal manera que se produce una molécula de ácido nucleico capaz de dirigir la transcripción de un gen dado y/o la síntesis de una molécula de proteína deseada. Las expresiones también se refieren al enlace de secuencias de aminoácidos de tal manera que se produce una proteína funcional. Por ejemplo, en un vector, una secuencia de control que está "unida operativamente" a una secuencia codificante de proteína, está ligada a la misma de manera que se consigue la expresión de la secuencia codificante de la proteína en condiciones compatibles con la actividad transcripcional de las secuencias de control.

En el presente documento, los términos "polipéptido" y "proteína" se utilizan indistintamente e incluyen una cadena molecular de dos o más aminoácidos unidos por enlace covalente a través de enlaces peptídicos. Los términos no se refieren a una longitud específica del producto. Por tanto, en la definición de polipéptido se incluyen "péptidos" y oligopéptidos". Los términos incluyen modificaciones postraduccionales del polipéptido, por ejemplo, glucosilaciones, acetilaciones, fosforilaciones y modificaciones similares. Además, en el significado de polipéptido se incluyen fragmentos proteicos, proteínas análogas, mutadas o variantes, proteínas de fusión y fragmentos similares. Los términos también incluyen moléculas en las que se incluyen uno o más análogos de aminoácidos o aminoácidos no canónicos o no naturales que pueden expresarse de forma recombinante usando técnicas conocidas de ingeniería de proteínas. Además, las proteínas de fusión pueden derivatizarse como se describe en el presente documento mediante técnicas de química orgánica bien conocidas.

Una "variante" de un polipéptido (por ejemplo, una inmunoglobulina o un anticuerpo) comprende una secuencia de aminoácidos en donde uno o más restos de aminoácidos se insertan, se eliminan y/o se sustituyen en la secuencia de aminoácidos en relación con otra secuencia polipeptídica. Las variantes incluyen proteínas de fusión.

La expresión "proteína de fusión" indica que la proteína incluye componentes polipeptídicos procedentes de más de una proteína o polipéptido precursor. Normalmente, una proteína de fusión se expresa a partir de un "gen de fusión" en el que una secuencia de nucleótidos que codifica una secuencia de polipéptidos de una proteína se adjunta en marco con, y opcionalmente se separa mediante un enlazador de, una secuencia de nucleótidos que codifica una secuencia de polipéptidos de una proteína diferente. Después, una célula hospedadora recombinante puede expresar el gen de fusión como una sola proteína.

Una proteína "secretada" se refiere a aquellas proteínas capaces de dirigirse al RE, a vesículas secretoras o al espacio extracelular como resultado de una secuencia peptídica señal secretora, así como aquellas proteínas liberadas en el espacio extracelular sin contener necesariamente una secuencia señal. Si la proteína secretada se libera en el espacio extracelular, la proteína secretada puede sufrir un procesamiento extracelular para producir una proteína "madura". La liberación al espacio extracelular puede producirse mediante muchos mecanismos, incluyendo la exocitosis y la escisión proteolítica. En algunas otras realizaciones del inventivo método, la célula hospedadora puede sintetizar la proteína de interés como una proteína secretada, que después también puede purificarse del espacio y/o medio extracelular.

Como se usa en el presente documento, cuando se hace referencia a una proteína producida por tecnología de ADN recombinante en una célula hospedadora, una proteína "soluble", es una proteína que existe en solución acuosa; si la proteína contiene una secuencia de aminoácidos señal de arginina gemela, la proteína soluble se exporta al espacio periplasmático en hospedadores bacterianos gram negativos, o se secreta al medio de cultivo por parte de células hospedadoras eucariotas capaces de secreción, o por hospedadores bacterianos que poseen los genes apropiados (por ejemplo, el gen *kil*). Por tanto, una proteína soluble es una proteína que no se encuentra en un cuerpo de inclusión dentro de la célula hospedadora. Como alternativa, dependiendo del contexto, una proteína soluble es una proteína que no se encuentra integrada en las membranas celulares, o, *in vitro*, se disuelve, o es capaz de disolverse en un tampón acuoso en condiciones fisiológicas sin formar cantidades significativas de agregados insolubles (es decir, forma agregados de menos del 10 %, y normalmente de menos de aproximadamente 5 %, de proteína total) cuando se suspende sin otras proteínas en un tampón acuoso de interés en condiciones fisiológicas, no conteniendo dicho tampón ningún detergente ni agente caotrópico, tal como urea, clorhidrato de guanidinio o perclorato de litio. En cambio, una proteína insoluble es aquella que existe en forma desnaturalizada dentro de los gránulos citoplasmáticos (denominados cuerpos de inclusión) en la célula hospedadora, o nuevamente dependiendo del contexto, una proteína insoluble es aquella que está presente en las membranas celulares, incluyendo, pero sin limitación, membranas citoplasmáticas, membranas mitocondriales, membranas de cloroplasto, membranas del retículo endoplasmático, etc., o en un tampón acuoso *in vitro* en condiciones fisiológicas que forma cantidades significativas de agregados insolubles (es decir, forma agregados iguales o superiores al 10 % de la proteína total) cuando se suspende sin otras proteínas (a una temperatura fisiológicamente compatible) en un tampón acuoso de interés en condiciones fisiológicas, no conteniendo dicho tampón ningún detergente ni agente caotrópico, tal como urea, clorhidrato de guanidinio o perclorato de litio.

El término "polinucleótido" o "ácido nucleico" incluye polímeros de nucleótidos tanto monocatenarios como bicatenarios que contienen dos o más restos de nucleótidos. Los restos de nucleótidos que comprenden el polinucleótido pueden ser ribonucleótidos o desoxirribonucleótidos o una forma modificada de cualquiera de los dos tipos de nucleótidos.

5 Dichas modificaciones incluyen modificaciones de bases, tales como derivados de bromouridina e inosina, modificaciones de ribosa, tales como 2',3'-didesoxirribosa y modificaciones de enlaces internucleotídicos, tales como fosforotioato, fosforoditioato, fosforoselenoato, fosforodiselenoato, fosforoanilitioato, fosforaniladato y fosforoamidato.

El término "oligonucleótido" significa un polinucleótido que comprende 200 o menos nucleótidos. En algunas realizaciones, los oligonucleótidos tienen una longitud de 10 a 60 bases. En otras realizaciones, los oligonucleótidos tienen una longitud de 12, 13, 14, 15, 16, 17, 18, 19 o 20 a 40 nucleótidos. Los oligonucleótidos pueden ser monocatenarios o bicatenarios, por ejemplo, para su uso en la construcción de un gen mutante. Los oligonucleótidos pueden ser oligonucleótidos de sentido o antisentido. Un oligonucleótido puede incluir un marcador, incluyendo un radiomarcador, un marcador fluorescente, un hapteno o un marcador antigénico, para ensayos de detección. Pueden usarse oligonucleótidos, por ejemplo, como cebadores para la PCR, cebadores de clonación o sondas de hibridación.

Una "secuencia de polinucleótidos" o "secuencia de nucleótidos" o "secuencia de ácido nucleico", tal como se usa indistintamente en el presente documento, es la secuencia primaria de restos de nucleótidos en un polinucleótido, incluyendo de un oligonucleótido, un ADN y ARN, un ácido nucleico, o una cadena de caracteres que representa la secuencia primaria de restos de nucleótidos, dependiendo del contexto. A partir de cualquier secuencia de polinucleótidos especificada, se puede determinar el ácido nucleico dado o la secuencia de polinucleótidos complementaria. Se incluyen ADN o ARN de origen genómico o sintético que pueden ser monocatenarios o bicatenarios y representan la cadena de sentido o antisentido. A menos que se especifique lo contrario, el extremo izquierdo de cualquier secuencia de polinucleótido monocatenaria analizada en el presente documento, es el extremo 5'; la dirección hacia la izquierda de las secuencias de polinucleótidos bicatenarias se cita como la dirección 5'. La dirección de adición de 5' a 3' de los transcritos de ARN nacientes se cita como la dirección de transcripción; las regiones de secuencia en la cadena de ADN que tienen la misma secuencia que la del transcrito de ARN que se encuentran en 5' respecto del extremo 5' del transcrito de ARN, se denominan "secuencias cadena arriba"; las regiones de secuencia en la cadena de ADN que tienen la misma secuencia que la del transcrito de ARN que se encuentran en 3' respecto del extremo 3' del transcrito de ARN, se denominan "secuencias cadena abajo".

Como se usa en el presente documento, una "molécula de ácido nucleico aislada" o "secuencia de ácido nucleico aislada" es una molécula de ácido nucleico que (1) se identifica y se separa de al menos una molécula de ácido nucleico contaminante con la que normalmente está asociada en la fuente natural del ácido nucleico o (2) se clona, amplifica, etiqueta, o se distingue de otro modo de los ácidos nucleicos de fondo, de manera que pueda determinarse la secuencia del ácido nucleico de interés. Una molécula de ácido nucleico aislada está en una forma o configuración distinta de en la que se encuentra en la naturaleza. Sin embargo, una molécula de ácido nucleico aislada incluye una molécula de ácido nucleico contenida en células que habitualmente expresan la inmunoglobulina (por ejemplo, el anticuerpo) donde, por ejemplo, la molécula de ácido nucleico está en una ubicación cromosómica diferente de la de las células naturales.

Como se usa en el presente documento, las expresiones "molécula de ácido nucleico codificante", "secuencia de ADN codificante", y "ADN codificante", se refieren al orden o a la secuencia de desoxirribonucleótidos a lo largo de una cadena de ácido desoxirribonucleico. El orden de estos desoxirribonucleótidos determina el orden de los ribonucleótidos a lo largo de la cadena de ARNm y también determina el orden de los aminoácidos a lo largo de la cadena del polipéptido (proteína). Por tanto, la secuencia de ADN codifica la secuencia de ARN y la secuencia de aminoácidos.

El término "gen" se usa ampliamente para referirse a cualquier ácido nucleico asociado a una función biológica. Los genes normalmente incluyen secuencias codificantes y/o secuencias reguladoras necesarias para la expresión de dichas secuencias codificantes. El término "gen" se aplica a una secuencia genómica o recombinante específica, así como a un ADNc o ARNm codificado por esa secuencia. Los genes también incluyen segmentos de ácido nucleico no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Las secuencias reguladoras no expresadas incluyen elementos de control transcripcional a los que se unen proteínas reguladoras, tales como factores de transcripción, dando como resultado la transcripción de secuencias adyacentes o cercanas.

"Expresión de un gen" o "expresión de un ácido nucleico" significa la transcripción de ADN en ARN (incluyendo opcionalmente modificación del ARN, por ejemplo, corte y empalme), traducción de ARN en un polipéptido (posiblemente incluida la posterior modificación postraduccional del polipéptido), o tanto la transcripción como la traducción, según lo indique el contexto.

La presente invención se refiere a un casete de expresión que incluye el inventivo promotor híbrido. Un "casete de expresión" eucariota se refiere a la parte de un vector de expresión que permite la producción de proteínas en una célula eucariota, tal como una célula de mamífero. Este incluye un promotor, operativo en una célula eucariota, para la transcripción de ARNm, uno o más genes que codifican proteínas de interés y una señal de procesamiento y

terminación de ARNm. Un casete de expresión puede incluir provechosamente entre las secuencias codificantes, un gen útil como marcador selectivo. En el casete de expresión de la presente invención, el promotor híbrido está unido operativamente en 5' a un marco abierto de lectura que codifica una proteína exógena de interés; y un sitio de poliadenilación está unido operativamente en 3' al marco abierto de lectura. Una realización de una secuencia de sitio de poliadenilación útil es un sitio de poliadenilación de SV40 tardío de SEQ ID NO: 14:

CACACATCATAAGATACATTGATGAGTTTGGACAAACCACAACACTAG
 AATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTAT
 TTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCAT
 TTTATGTTTCAGGTTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAA
 CCTCTACAAATGTGGTA//SEQ ID NO:14.

También pueden incluirse otras secuencias de control adecuadas siempre que el casete de expresión permanezca operativo. El marco abierto de lectura puede incluir opcionalmente una secuencia codificante de más de una proteína de interés.

Como se usa en este documento, la expresión "región codificante" o "secuencia codificante" cuando se usa en referencia a un gen estructural, se refiere a las secuencias de nucleótidos que codifican los aminoácidos que se encuentran en el polipéptido naciente como resultado de la traducción de una molécula de ARNm. La región codificante está limitada, en eucariotas, en el lado 5' por el triplete de nucleótidos "ATG" que codifica la metionina iniciadora y en el lado 3' por uno de los tres tripletes que especifican los codones de parada (es decir, TAA, TAG, TGA).

La presente invención también incluye un vector de expresión recombinante que comprende el inventivo casete de expresión.

El término "vector" significa cualquier molécula o entidad (por ejemplo, ácido nucleico, plásmido, bacteriófago o virus) utilizada para transferir información codificante de proteínas a una célula hospedadora.

La expresión "vector de expresión" o "construcción de expresión", como se usa en este documento, se refiere a una molécula de ADN recombinante que contiene una secuencia codificante deseada y secuencias de control de ácido nucleico apropiadas necesarias para la expresión de la secuencia codificante unida operativamente en una célula hospedadora particular. Un vector de expresión puede incluir, pero sin limitación, secuencias que afectan a, o controlan, la transcripción, la traducción y, si hay intrones, afectan al corte y empalme de ARN de una región codificante unida operativamente a los mismos. Las secuencias de ácido nucleico necesarias para la expresión en procariontes incluyen un promotor, opcionalmente una secuencia de operador, un sitio de unión al ribosoma y posiblemente otras secuencias. Se sabe que las células eucariotas utilizan promotores, potenciadores y señales de terminación y poliadenilación. Una secuencia de péptido señal secretora también puede, opcionalmente, estar codificada por el vector de expresión, ligado operativamente a la secuencia codificante de interés, de modo que el polipéptido expresado pueda ser secretado por la célula hospedadora recombinante, para un aislamiento más fácil del polipéptido de interés de la célula, si se desea. Dichas técnicas son muy conocidas en la materia. (Por ejemplo, Goodey, Andrew R.; et al., Peptide and DNA sequences, patente de Estados Unidos N. °5.302.697; Weiner et al., Compositions and methods for protein secretion, patente de Estados Unidos N. °6.022.952 y patente de Estados Unidos N. °6.335.178; Uemura et al., Protein expression vector and utilization thereof, patente de Estados Unidos N. °7.029.909; Ruben et al., 27 human secreted proteins, US 2003/0104400 A1). Para la expresión de proteínas de interés de múltiples subunidades, para transformar una célula hospedadora, pueden utilizarse vectores de expresión distintos en número y proporciones adecuados, conteniendo cada uno una secuencia codificante para cada uno de los diferentes monómeros. En otras realizaciones, para expresar las diferentes subunidades de la proteína de interés puede utilizarse un solo vector de expresión.

La presente invención también se refiere a una célula hospedadora de mamífero que comprende el inventivo vector de expresión recombinante.

La expresión "célula hospedadora" significa una célula que se ha transformado o tiene la capacidad de transformarse, con una secuencia de ácido nucleico y de este modo expresa un gen de interés. La expresión incluye la descendencia de la célula precursora, ya sea idéntica o no la descendencia en cuanto a su morfología o composición genética a la de la célula precursora original, en tanto que esté presente el gen de interés. Puede usarse cualquiera de un gran número de células hospedadoras disponibles y bien conocidas en la práctica de la presente invención. La selección de un hospedador determinado depende de diversos factores reconocidos en la técnica. Estos incluyen, por ejemplo, la compatibilidad con el vector de expresión elegido, la toxicidad de los péptidos codificados por la molécula de ADN,

la tasa de transformación, la facilidad de recuperación de los péptidos, las características de expresión, la bioseguridad y los costes. Se debe lograr un equilibrio de estos factores sabiendo que no todos los hospedadores pueden ser igualmente eficaces para la expresión de una secuencia de ADN particular. Dentro de estas pautas generales, las células hospedadoras microbianas útiles en cultivo incluyen bacterias (tales como *Escherichia coli* sp.), levaduras (tales como *Saccharomyces* sp.) y otras células fúngicas, células de insecto, células vegetales, células de mamífero (incluidos seres humanos), por ejemplo, células CHO y células HEK-293. También puede realizarse modificaciones a nivel de ADN. La secuencia de ADN que codifica el péptido puede cambiarse por codones más compatibles con la célula hospedadora elegida. En la técnica se conocen codones optimizados para la especie *E. coli*. Los codones pueden sustituirse para eliminar sitios de restricción o para incluir sitios de restricción silenciosos, que puede ayudar en el procesamiento del ADN en la célula hospedadora seleccionada. A continuación, el hospedador transformado se cultiva y purifica. Las células hospedadoras pueden cultivarse en condiciones de fermentación o cultivo convencionales para que se expresen los compuestos deseados. Dichas condiciones de fermentación y cultivo son muy conocidas en la técnica.

15 Son ejemplos de líneas de células hospedadoras de mamífero útiles, las células de ovario de hámster chino, incluidas las células CHO-K1 (p. ej., ATCC CCL61), las células DXB-11, DG-44 y de ovario de hámster chino/-DHFR (CHO, Urlaub et al, Proc. Natl. Acad. Sci. USA 77: 4216 (1980)); la línea CVI de riñón de mono transformada por SV40 (COS-7, ATCC CRL 1651); la línea de riñón embrionario humano (293 o células 293 subclonadas para su crecimiento en cultivo en suspensión, [Graham et al, J. Gen Virol. 36: 59 (1977)]; las células de riñón de cría de hámster (BHK, ATCC CCL 10); las células de Sertoli de ratón (TM4, Mather, Biol. Reprod. 23: 243-251 (1980)); las células de riñón de mono (CVI ATCC CCL 70); las células de riñón de mono verde africano (VERO-76, ATCC CRL-1587); las células de carcinoma cervicouterino humano (HELA, ATCC, CCL 2); las células de riñón canino (MDCK, ATCC CCL 34); las células de hígado de rata búfalo (BRL 3A, ATCC CRL 1442); las células de pulmón humano (W138, ATCC CCL 75); las células de hepatoma humano (Hep G2, HB 8065); las células de tumor mamario de ratón (MMT 060562, ATCC CCL51); las células TRI (Mather et al, Annals N.Y Acad. Sci. 383: 44-68 (1982)); las células MRC 5 o células FS4; o las células de mieloma de mamífero.

Las expresiones "célula", "línea celular", y "cultivo celular" a menudo se usan indistintamente y en el presente documento todas estas designaciones incluyen la descendencia celular. Por ejemplo, una célula "procedente de" una célula CHO es una descendencia celular de una célula de ovario de hámster chino, que se puede eliminar de la célula precursora primaria original mediante cualquier número de generaciones, y que también puede incluir una célula descendiente transformante. Los transformantes y las células transformadas incluyen la célula objeto primaria y los cultivos procedentes de la misma sin tener en cuenta el número de transferencias. Se entiende también que toda la descendencia puede no ser exactamente idéntica en cuanto al contenido de ADN, debido a mutaciones deliberadas o involuntarias. Se incluyen descendientes mutantes que tienen la misma función o actividad biológica que la usada para cribar la célula transformada originalmente.

Las células hospedadoras se transforman o transfectan con los ácidos nucleicos o vectores descritos anteriormente para la producción de polipéptidos (incluidas las proteínas de unión a antígenos, como los anticuerpos) y se cultivan en un medio nutritivo convencional modificado según sea apropiado para inducir promotores, seleccionar transformantes o amplificar los genes que codifican las secuencias deseadas. Además, los nuevos vectores y líneas celulares transfectadas con múltiples copias de unidades de transcripción separadas por un marcador selectivo, son particularmente útiles para la expresión de polipéptidos, tales como anticuerpos.

El término "transfección" significa la captación de ADN extraño o exógeno por una célula y una célula se ha "transfectado" cuando el ADN exógeno se ha introducido dentro de la membrana celular. Se conocen bien en el campo técnico, diversas técnicas de transfección y se divulgan en el presente documento. Véase, por ejemplo, Graham et al., 1973, *Virology* 52:456; Sambrook et al., 2001, *Molecular Cloning: A Laboratory Manual*, citado anteriormente; Davis et al., 1986, *Basic Methods in Molecular Biology*, Elsevier; Chu et al., 1981, *Gene* 13:197. Dichas técnicas pueden utilizarse para introducir uno o más residuos de ADN exógenos en células hospedadoras adecuadas.

El término "transformación" se refiere a un cambio en las características genéticas de una célula y una célula se ha transformado cuando se ha modificado para que contenga ADN o ARN nuevo. Por ejemplo, una célula está transformada en caso de que esté genéticamente modificada con respecto a su estado nativo mediante la introducción de nuevo material genético por transfección, transducción u otras técnicas. Después de la transfección o transducción, el ADN transformante puede recombinarse con el de la célula integrándose físicamente en un cromosoma de la célula o puede mantenerse transitoriamente como un elemento episómico sin replicarse o puede replicarse de manera independiente como un plásmido. Se considera que una célula se ha "transformado de manera estable" cuando el ADN transformante se replica con la división de la célula.

La presente invención también se refiere a un método para producir una proteína de interés que implica el cultivo de la célula hospedadora de mamífero, en medio acuoso en condiciones fisiológicas que permitan la expresión de la proteína de interés; y recuperar la proteína de interés del medio.

Las células hospedadoras utilizadas para producir los polipéptidos útiles en la invención, pueden cultivarse en una

- variedad de medios. Los medios comercialmente disponibles tales como F10 de Ham (Sigma), medio mínimo esencial ((MEM), (Sigma), RPMI-1640 (Sigma), medio de Eagle modificado por Dulbecco ((DMEM), Sigma), son adecuados para cultivar células hospedadoras. Además, cualquiera de los medios descritos en Ham et al., Meth. Enz. 58: 44 (1979), Barnes et al., Anal. Biochem. 102: 255 (1980), en las patentes de Estados Unidos N. °4.767.704; 4.657.866; 5 4.927.762; 4.560.655; o 5.122.469; en los documentos WO90103430 y WO 87/00195; o en la patente de Estados Unidos Re N. °30.985, pueden usarse como medio de cultivo para las células hospedadoras. Cualquiera de estos medios puede complementarse según sea necesario con hormonas y/u otros factores de crecimiento (tales como insulina, transferrina o factor de crecimiento epidérmico), sales (tales como cloruro de sodio, calcio, magnesio y fosfato), tampones (tal como HEPES), nucleótidos (tales como adenosina y timidina), antibióticos (tal como el fármaco
- 10 GENTAMYCIN™), oligoelementos (definidos como compuestos inorgánicos habitualmente presentes en concentraciones finales en el intervalo micromolar) y glucosa o una fuente de energía equivalente, de tal manera que las condiciones fisiológicas de la célula en, o sobre, el medio promueven la expresión de la proteína de interés por la célula hospedadora; también se pueden incluir otros suplementos a concentraciones apropiadas que conocerían los expertos en la materia. Las condiciones de cultivo, tales como la temperatura (normalmente, pero no necesariamente,
- 15 una temperatura de aproximadamente 37 °C, el pH (normalmente, pero no necesariamente, un pH de aproximadamente 6,5 a 7,5), la oxigenación y similares, son las utilizadas previamente con la célula hospedadora seleccionada para la expresión de la proteína de interés, y serán obvias para el experto en la materia. El medio de cultivo puede incluir una cantidad adecuada de suero tal como suero bovino fetal
- 20 (FBS, *fetal bovine serum*), o preferiblemente, las células hospedadoras se pueden adaptar para cultivo en medio sin suero. En algunas realizaciones, el medio acuoso es líquido, de manera que las células hospedadoras se cultivan en una suspensión celular dentro del medio líquido. Las células hospedadoras pueden cultivarse de manera útil en cultivo discontinuo o en sistemas de cultivo continuo.
- 25 En otras realizaciones, las células hospedadoras de mamífero se pueden cultivar en medio acuoso sólido o semisólido, por ejemplo, que contiene agar o agarosa, para formar un medio o superficie de sustrato a la que se adhieren las células y forman una capa de adhesión.

Al cultivar las células hospedadoras, el polipéptido recombinante puede producirse intracelularmente, en el espacio

30 periplasmático, o secretarse directamente al medio. Si el polipéptido, tal como una proteína de unión a antígeno (por ejemplo, un anticuerpo), se produce intracelularmente, como una primera etapa, los restos de partículas, ya sean células hospedadoras o fragmentos resultantes de la lisis, se eliminan, por ejemplo, mediante centrifugación o ultrafiltración.

- 35 Una proteína de interés, tal como un anticuerpo o fragmento de anticuerpo puede purificarse usando, por ejemplo, cromatografía de hidroxil apatita, cromatografía de intercambio catiónico o aniónico, o preferiblemente cromatografía de afinidad, utilizando el antígeno de interés o proteína A o proteína G como ligando de afinidad. La proteína A puede usarse para purificar proteínas que incluyen polipéptidos basados en cadenas pesadas $\gamma 1$, $\gamma 2$ o $\gamma 4$ humanas (Lindmark et al., J. Immunol. Meth. 62: 1-13 (1983)). La proteína G se recomienda para todos los isotipos de ratón y para $\gamma 3$
- 40 humano (Guss et al., EMBO J. 5:15671575 (1986)). La matriz a la que se une el ligando de afinidad suele ser agarosa, pero otras matrices están disponibles. Las matrices mecánicamente estables, tales como vidrio o poli(estirendivinil)benceno de poro controlado, permiten caudales más rápidos y tiempos de procesamiento más cortos que los que se pueden obtener con la agarosa. Cuando la proteína comprende un dominio CH 3, la resina Bakerbond ABX™ (JT Baker, Phillipsburg, NJ) es útil para la purificación. Otras técnicas de purificación de proteínas como la
- 45 precipitación con etanol, HPLC de fase inversa, cromatografía de intercambio iónico, SDS-PAGE y precipitación con sulfato de amonio, también están disponibles dependiendo del anticuerpo que se vaya a recuperar.

"En condiciones fisiológicas", con respecto a tampones de incubación e inmunoglobulinas, u otros reactivos de ensayos de unión, significa incubación en condiciones de temperatura, pH y fuerza iónica, que permitan que se

50 produzca una reacción bioquímica, tal como una reacción de unión no covalente. Normalmente, la temperatura es una temperatura de lugar o ambiente hasta aproximadamente 37 °C y a un pH de 6,5 a 7,5.

"Sal fisiológicamente aceptable" de una composición de materia, por ejemplo, una sal de una inmunoglobulina, tal como de un anticuerpo o de otra proteína de interés, significa cualquier sal o sales, que se sabe, o que se descubre

55 más tarde, que es (son) farmacéuticamente aceptable(s). Algunos ejemplos no limitantes de sales farmacéuticamente aceptables son: acetato; trifluoroacetato; hidroháluros, tales como hidrocloreto e hidrobromuro; sulfato; citrato; maleato; tartrato; glicolato; gluconato; succinato; mesilato; besilato; sales de ésteres de ácido gálico (el ácido gálico también se conoce como ácido 3,4, 5 trihidroxibenzoico) tales como pentagalactoglucosa (PGG) y galato de epigallocatequina (EGCG), sales de colesteril sulfato, sales de pamoato, tanato y oxalato.

60 Una "mezcla de reacción" es una mezcla acuosa que contiene todos los reactivos y factores necesarios, que en condiciones fisiológicas de incubación, permiten que se produzca una reacción bioquímica *in vitro* de interés, tal como una reacción de unión covalente o no covalente.

- 65 Un "dominio" o una "región" (utilizados indistintamente en el presente documento) de un polinucleótido, es cualquier

parte de todo el polinucleótido, hasta e incluyendo el polinucleótido completo, pero normalmente comprende menos que el polinucleótido completo. Un dominio puede, pero no necesariamente, plegarse independientemente (p. ej., plegamiento en horquilla de ADN) del resto de la cadena de polinucleótidos y/o correlacionarse con una función o ubicación biológica, bioquímica o estructural particular, tal como una región codificante o una región reguladora.

5 Un "dominio" o una "región" (utilizados indistintamente en el presente documento) de una proteína, es cualquier parte de toda la proteína, hasta e incluyendo la proteína completa, pero normalmente comprende menos que la proteína completa. Un dominio puede, pero no necesariamente, plegarse independientemente del resto de la cadena proteica y/o correlacionarse con una función o ubicación biológica, bioquímica o estructural particular (por ejemplo, un dominio
10 de unión a ligando, o un dominio citosólico, transmembrana o extracelular).

El término "anticuerpo", o indistintamente "Ab", se utiliza en el sentido más amplio e incluye anticuerpos completamente ensamblados, anticuerpos monoclonales (incluidos anticuerpos humanos, humanizados o quiméricos), anticuerpos policlonales, anticuerpos multiespecíficos (por ejemplo, anticuerpos biespecíficos) y fragmentos de anticuerpo que
15 pueden unirse al antígeno (por ejemplo, Fab, Fab', F(ab')₂, Fv, anticuerpos monocatenarios, diacuerpos), que comprenden las regiones determinantes de la complementariedad (CDR, siglas del inglés *complementarity determining regions*) de los anteriores, siempre que muestren la actividad biológica deseada. Se contemplan multímeros o agregados de moléculas y/o fragmentos intactos, incluidos los anticuerpos químicamente derivatizados. Se contemplan anticuerpos de cualquier clase o subclase de isotipo, incluyendo IgG, IgM, IgD, IgA e IgE, IgG1, IgG2,
20 IgG3, IgG4, IgA1 e IgA2, o de cualquier alotipo. Los diferentes isotipos tienen diferentes funciones efectoras; por ejemplo, los isotipos IgG1 e IgG3 tienen actividad de citotoxicidad celular dependiente de anticuerpos (ADCC, siglas del inglés *antibody-dependent cellular cytotoxicity*).

Una proteína "aislada", por ejemplo, una inmunoglobulina, tal como un anticuerpo o un fragmento de anticuerpo, es
25 una que se ha identificado y separado de uno o más componentes de su ambiente natural o de un medio de cultivo en el que una célula productora lo ha secretado. En algunas realizaciones, la proteína aislada está sustancialmente exenta de proteínas o polipéptidos u otros contaminantes que se encuentran en su ambiente natural o medio de cultivo que podrían interferir con su uso desde un punto de vista terapéutico, diagnóstico, profiláctico, de investigación u otro uso. Los componentes "contaminantes" de su ambiente o medio natural son materiales que podrían interferir con los
30 usos diagnósticos o terapéuticos de la proteína, por ejemplo, un anticuerpo, y puede incluir enzimas, hormonas y otros solutos proteicos o no proteicos (por ejemplo, polinucleótidos, lípidos, hidratos de carbono). Normalmente, una "proteína aislada" constituye al menos aproximadamente un 5 %, al menos aproximadamente un 10 %, al menos aproximadamente un 25 % o al menos aproximadamente un 50 % de una muestra dada. En algunas realizaciones, la proteína de interés, por ejemplo, un anticuerpo, se purificará (1) hasta un nivel mayor del 95 % en peso de proteína, y
35 lo más preferentemente, mayor del 99 % en peso, o (2) hasta la homogeneidad por SDS-PAGE en condiciones reductoras o no reductoras, opcionalmente usando tinción, por ejemplo, tinción con azul de Coomassie o con plata. El anticuerpo de origen natural aislado incluye el anticuerpo *in situ* en el interior de células recombinantes, ya que al menos un componente del entorno natural de la proteína no estará presente. Sin embargo, normalmente, la proteína aislada de interés (por ejemplo, un anticuerpo) se preparará mediante al menos una etapa de purificación.

40 La expresión "anticuerpo monoclonal", como se usa en el presente documento, se refiere a un anticuerpo obtenido de una población de anticuerpos sustancialmente homogénea, es decir, los anticuerpos individuales que comprenden la población, son idénticos salvo por las posibles mutaciones de origen natural que puede haber en cantidades minoritarias. Los anticuerpos monoclonales, que son proteínas de unión a antígeno, son proteínas de unión
45 sumamente específicas, que se dirigen contra un sitio o epítipo antigénico individual, a diferencia de las preparaciones de anticuerpos policlonales que normalmente incluyen anticuerpos diferentes dirigidos contra epítipos diferentes. Como ejemplos no limitantes de anticuerpos monoclonales se incluyen anticuerpos murinos, de conejo, rata, pollo, quiméricos, humanizados o humanos, anticuerpos completamente ensamblados, anticuerpos multiespecíficos (incluidos anticuerpos biespecíficos), fragmentos de anticuerpo que pueden unirse a un antígeno (incluidos, Fab, Fab',
50 F(ab)₂, Fv, anticuerpos monocatenarios, diacuerpos), maxicuerpos, nanocuerpos y péptidos recombinantes que comprenden las CDR de los anteriores siempre que muestren la actividad biológica deseada, o variantes o derivados de los mismos.

El modificador "monoclonal" indica la naturaleza del anticuerpo obtenido de una población de anticuerpos
55 sustancialmente homogénea y no debe interpretarse como que requiere la producción del anticuerpo por ningún método concreto. Por ejemplo, los anticuerpos monoclonales para su uso de acuerdo con la presente invención, pueden prepararse mediante el método de hibridoma descrito por primera vez por Kohler et al., *Nature*, 256:495 [1975], o pueden prepararse mediante métodos de ADN recombinante (véase, por ejemplo, la patente de Estados Unidos N.º 4.816.567). Los "anticuerpos monoclonales" también pueden aislarse de fagotecas de anticuerpos usando las
60 técnicas descritas en Clackson et al., *Nature*, 352:624-628[1991] y en Marks et al., *J. Mol. Biol.*, 222:581-597 (1991), por ejemplo.

Un agente de unión o proteína o anticuerpo de unión a antígeno "multiespecífico", es uno que se dirige a más de un antígeno o epítipo.

65

Un agente de unión o proteína o anticuerpo de unión a antígeno "bienespecifico", "dual-especifico" o "bifuncional", es un híbrido que tiene dos sitios de unión a antígeno diferentes. Las proteínas de unión a antígeno, las proteínas y los anticuerpos de unión a antígeno son una especie de proteína de unión a múltiples antígenos, proteína de unión a antígeno o anticuerpo multiespecifico y puede producirse mediante una variedad de métodos que incluyen, pero sin limitación, fusión de hibridomas o enlace de fragmentos Fab'. (Véase, por ejemplo, Songsivilai y Lachmann, 1990, Clin. Exp. Immunol. 79:315-321; Kostelny et al., 1992, J. Immunol. 148:1547-1553; Spiess et al., Alternative molecular formats and therapeutic applications for bispecific antibodies, Mol. Immunol. 67:95-106 (2015)). Los dos sitios de unión de una proteína o anticuerpo de unión a antígeno bienespecifico se unirán a dos epítomos diferentes, que pueden encontrarse en las mismas dianas proteicas o en diferentes dianas proteicas.

10

El término "inmunoglobulina" abarca anticuerpos completos que comprenden dos cadenas pesadas (HC, siglas del inglés *heavy chains*) dimerizadas, cada una de ellas unida por enlace covalente a una cadena ligera (LC, siglas del inglés *light chain*); una sola cadena pesada de inmunoglobulina no dimerizada y una cadena ligera unida por enlace covalente (HC + LC), o una inmunoglobulina quimérica (heterodímero de cadena ligera + cadena pesada -Fc (denominado "hemiacuerpo"). Una "inmunoglobulina" es una proteína, pero no es necesariamente una proteína de unión a antígeno.

15

En un "anticuerpo", cada tetrámero está compuesto por dos pares de cadenas polipeptídicas idénticas, teniendo cada par una cadena "ligera" de aproximadamente 220 aminoácidos (aproximadamente 25 kDa) y una cadena "pesada" de aproximadamente 440 aminoácidos (aproximadamente 50-70 kDa). La parte amino-terminal de cada cadena incluye una región "variable" ("V") de aproximadamente 100 a 110 o más aminoácidos, principalmente responsable del reconocimiento antigénico. La parte carboxi-terminal de cada cadena define una región constante, principalmente responsable de la función efectora. La región variable difiere entre diferentes anticuerpos. La región constante es la misma entre diferentes anticuerpos. Dentro de la región variable de cada cadena pesada o ligera, hay tres subregiones hipervariables que ayudan a determinar la especificidad del anticuerpo por el antígeno en el caso de un anticuerpo que sea una proteína de unión a antígeno. Sin embargo, dentro del alcance de la presente invención, una realización de la inmunoglobulina, por ejemplo, un anticuerpo, no necesita ser una proteína de unión a antígeno, o no es necesario saber que se une específicamente a un antígeno. Los restos del dominio variable entre las regiones hipervariables se denominan restos estructurales y generalmente son algo homólogos entre diferentes anticuerpos. Las inmunoglobulinas se pueden asignar a diferentes clases dependiendo de la secuencia de aminoácidos del dominio constante de sus cadenas pesadas. Las cadenas ligeras humanas se clasifican como cadenas ligeras kappa (κ) y lambda (λ). En las cadenas ligeras y pesadas, las regiones variable y constante están unidas por una región "J" de aproximadamente 12 o más aminoácidos, incluyendo también la cadena pesada una región "D" de aproximadamente 10 aminoácidos más. Véase, en general, Fundamental Immunology, cap. 7 (Paul, W., ed., 2.^a ed. Raven Press, N.Y. (1989)). En el alcance de la invención, un "anticuerpo" también incluye un anticuerpo fabricado de manera recombinante, y anticuerpos que están glicosilados o que carecen de glicosilación.

20

25

30

35

La expresión "cadena ligera" o "cadena ligera de inmunoglobulina" incluye una cadena ligera de longitud completa y fragmentos de la misma que tienen suficiente secuencia de región variable para conferir especificidad de unión. Una cadena ligera de longitud completa incluye un dominio de región variable, V_L y un dominio de región constante, C_L . El dominio de región variable de la cadena ligera se encuentra en el extremo amino del polipéptido. Las cadenas ligeras incluyen cadenas kappa y cadenas lambda.

40

La expresión "cadena pesada" o "cadena pesada de inmunoglobulina" incluye una cadena pesada de longitud completa y fragmentos de la misma que tienen suficiente secuencia de región variable para conferir especificidad de unión. Una cadena pesada de longitud completa incluye un dominio de región variable, V_H y tres dominios de región constante, C_{H1} , C_{H2} y C_{H3} . El dominio V_H se encuentra en el extremo amino del polipéptido y los dominios C_H se encuentran en el extremo carboxilo, siendo el C_{H3} el más cercano al extremo carboxilo del polipéptido. Las cadenas pesadas se clasifican como mu (μ), delta (δ), gamma (γ), alfa (α) y épsilon (ϵ), y definen el isotipo del anticuerpo como IgM, IgD, IgG, IgA e IgE, respectivamente. En formas de realización distintas de la invención, las cadenas pesadas pueden ser de cualquier isotipo, incluyendo IgG (incluyendo los subtipos IgG1, IgG2, IgG3 e IgG4), IgA (incluyendo los subtipos IgA1 e IgA2), IgM e IgE. Varios de estos pueden dividirse además en subclases o isotipos, por ejemplo, IgG1, IgG2, IgG3, IgG4, IgA1 e IgA2. Diferentes isotipos de IgG pueden tener diferentes funciones efectoras (mediadas por la región Fc), tales como citotoxicidad celular dependiente de anticuerpos (ADCC) y citotoxicidad dependiente del complemento (CDC). En la ADCC, la región Fc de un anticuerpo se une a receptores de Fc (Fc γ R) en la superficie de células efectoras inmunitarias tales como linfocitos citolíticos naturales (*natural killers*) y macrófagos, conduciendo a la fagocitosis o a la lisis de las células diana. En la CDC, los anticuerpos destruyen a las células diana activando la cascada del complemento en la superficie celular.

45

50

55

60

Una "región Fc", o como se usa indistintamente en el presente documento, un "dominio Fc" o "dominio Fc de inmunoglobulina", contiene dos fragmentos de cadena pesada, que en un anticuerpo completo comprenden los dominios C_{H1} y C_{H2} del anticuerpo. Los dos fragmentos de cadena pesada se mantienen juntos mediante dos o más enlaces disulfuro y mediante interacciones hidrófobas de los dominios C_{H3} .

65

La expresión "epítomo de unión al receptor de rescate" se refiere a un epítomo de la región Fc de una molécula de IgG

(por ejemplo, IgG₁, IgG₂, IgG₃ o IgG₄), que es responsable de aumentar la semivida en suero de la molécula de IgG *in vivo*.

En muchas realizaciones de la invención, la proteína de interés es una proteína de unión a antígeno, tal como, pero sin limitación, un anticuerpo, una subunidad de anticuerpo o un fragmento de anticuerpo. Para una descripción detallada sobre la estructura y generación de anticuerpos, véase Roth, D. B. y Craig, N. L., *Cell*, 94:411-414 (1998), incorporado por referencia en su totalidad en el presente documento. Resumiendo, el proceso de generación de ADN codificante de las secuencias de inmunoglobulina de cadena pesada y ligera, se produce principalmente en linfocitos B en desarrollo. Antes de la reorganización y de la unión de varios segmentos génicos de inmunoglobulina, los segmentos génicos V, D, J y constante (C) se encuentran generalmente en una proximidad relativamente estrecha en un solo cromosoma. Durante la diferenciación de los linfocitos B, uno de cada uno de los miembros apropiados de la familia de los segmentos génicos V, D, J (o solo V y J en el caso de genes de cadena ligera), se recombinan para formar regiones variables funcionalmente reorganizadas de los genes de inmunoglobulina pesada y ligera. Este proceso de reorganización de segmentos génicos parece ser secuencial. En primer lugar, se hacen uniones D-a-J de cadena pesada, seguido de uniones V-a-DJ de cadena pesada y uniones V-a-J de cadena ligera. Además de la reorganización de los segmentos V, D y J, se genera mayor diversidad en el repertorio primario de las cadenas ligera y pesada de inmunoglobulina mediante recombinación variable en las ubicaciones donde se unen los segmentos V y J de la cadena ligera y donde se unen los segmentos D y J de la cadena pesada. Dicha variación en la cadena ligera se produce normalmente dentro del último codón del segmento génico V y del primer codón del segmento J. En el cromosoma de la cadena pesada entre los segmentos D y J_H se produce una imprecisión similar en la unión y puede extenderse hasta 10 nucleótidos. Por otro lado, entre los segmentos génicos D y J_H y entre V_H y D, pueden insertarse varios nucleótidos que no están codificados por ADN genómico. La adición de estos nucleótidos se conoce como diversidad de la región N. El efecto neto de dichas reorganizaciones en los segmentos génicos de la región variable y la recombinación variable que puede producirse durante dicha unión, es la producción de un repertorio de anticuerpos primarios.

La expresión región "hipervariable" se refiere a los restos de aminoácidos de un anticuerpo que son responsables de la unión con el antígeno. La región hipervariable comprende restos de aminoácidos de una región determinante de la complementariedad o CDR [es decir, los restos 24-34 (L1), 50-56 (L2) y 89-97 (L3) en el dominio variable de la cadena ligera y 31-35 (H1), 50-65 (H2) y 95-102 (H3) en el dominio variable de la cadena pesada, como describen Kabat et al., *Sequences of Proteins of Immunological Interest*, ^a Ed Public Health Service, National Institutes of Health, Bethesda, Md. (1991)]. Incluso una sola CDR puede reconocer a un antígeno y unirse al mismo, aunque con una afinidad menor que todo el sitio de unión al antígeno que contiene todas las CDR.

Chothia et al., describen una definición alternativa de restos de un "bucle" hipervariable, en *J. Mol. Biol.* 196: 901-917 (1987) como restos 26-32 (L1), 50-52 (L2) y 91-96 (L3) en el dominio variable de la cadena ligera y 26-32 (H1), 53-55 (H2) y 96-101 (H3) en el dominio variable de la cadena pesada.

Los restos "estructurales" o "FR" son los restos de la región variable distintos de los de la región hipervariable.

Los "fragmentos de anticuerpos" comprenden una parte de un anticuerpo intacto de longitud completa, preferentemente, la región de unión al antígeno o variable del anticuerpo intacto. Como ejemplos de fragmentos de anticuerpo se incluyen fragmentos Fab, Fab', F(ab')₂ y Fv; diacuerpos; anticuerpos lineales (Zapata et al., *Protein Eng.*, 8(10):1057-1062 (1995)); moléculas de anticuerpo monocatenario; y anticuerpos multiespecíficos formados a partir de fragmentos de anticuerpos.

La digestión con papaína de los anticuerpos produce dos fragmentos de unión a antígeno idénticos, denominados fragmentos "Fab", cada uno con un solo sitio de unión al antígeno y un fragmento "Fc" residual que contiene la región constante. El fragmento Fab contiene todo el dominio variable, así como el dominio constante de la cadena ligera y el primer dominio constante (CH1) de la cadena pesada. El fragmento Fc muestra hidratos de carbono y es responsable de muchas funciones efectoras de anticuerpos (tales como la unión del complemento y de receptores celulares), que distinguen una clase de anticuerpo de otra.

El tratamiento con pepsina produce un fragmento F(ab')₂ que tiene dos fragmentos de anticuerpo "Fv monocatenario" o "scFv" que comprenden los dominios V_H y V_L de anticuerpo, en donde estos dominios están presentes en una sola cadena polipeptídica. Los fragmentos Fab difieren de los fragmentos Fab' por la inclusión de algunos restos adicionales en el extremo carboxilo del dominio CH1 de cadena pesada incluyendo una o más cisteínas de la región bisagra del anticuerpo. Preferentemente, el polipéptido Fv comprende además un enlazador polipeptídico entre los dominios V_H y V_L que permite al Fv formar la estructura deseada para la unión del antígeno. Para una revisión de scFv véase Plückthun en *The Pharmacology of Monoclonal Antibodies*, vol. 1 13, Rosenberg y Moore eds., Springer-Verlag, Nueva York, págs. 269-315 (1994).

Un "fragmento Fab" comprende una cadena ligera y las regiones C_{H1} y variable de una cadena pesada. La cadena pesada de una molécula de Fab no puede formar un enlace disulfuro con otra molécula de cadena pesada.

65

Un "fragmento Fab'" contiene una cadena ligera y una parte de una cadena pesada que contiene el dominio V_H y el dominio C_{H1} y también la región entre los dominios C_{H1} y C_{H2} , de tal manera que entre las dos cadenas pesadas de dos fragmentos Fab' puede formarse un enlace disulfuro intercadena para formar una molécula de $F(ab')_2$.

- 5 Un "fragmento $F(ab')_2$ " contiene dos cadenas ligeras y dos cadenas pesadas que contienen una parte de la región constante entre los dominios C_{H1} y C_{H2} , de tal forma que se forma un enlace disulfuro intercadena entre las dos cadenas pesadas. Por tanto, un fragmento $F(ab')_2$ está compuesto por dos fragmentos Fab' que se mantienen juntos mediante un enlace disulfuro entre las dos cadenas pesadas.
- 10 El fragmento "Fv" es el fragmento mínimo de anticuerpo que contiene un sitio completo de reconocimiento del antígeno y de unión con el mismo. Esta región consiste en un dímero de un dominio variable de cadena pesada y uno de cadena ligera en asociación estrecha, no covalente. Es en esta configuración, en la que las tres CDR de cada dominio variable interaccionan para definir un sitio de unión con el antígeno en la superficie del dímero $VH-VL$. Un solo dominio variable (o la mitad de un Fv que comprende solo tres CDR específicas para un antígeno) tiene la capacidad de reconocer al
- 15 antígeno y unirse al mismo, aunque con una afinidad menor en comparación con el sitio de unión completo.

Los "anticuerpos monocatenarios" son moléculas de Fv en las que las regiones variables de cadena pesada y ligera están conectadas a través de un enlazador flexible para formar una sola cadena polipeptídica, que forma una región de unión a antígeno. Los anticuerpos monocatenarios se analizan con detalle en la publicación de solicitud

20 internacional de patente n.º WO 88/01649 y en las patentes de Estados Unidos n.º 4.946.778 y 5.260.203, cuyas divulgaciones se han incorporado por referencia en su totalidad.

Los fragmentos de anticuerpos "Fv monocatenarios" o "scFv" comprenden los dominios V_H y V_L del anticuerpo, en donde estos dominios están presentes en una sola cadena polipeptídica y, opcionalmente, comprenden un enlazador

25 polipeptídico entre los dominios V_H y V_L que permiten que el Fv forme la estructura deseada para la unión con el antígeno (Bird et al., Science 242:423-426, 1988 y Huston et al., Proc. Natl. Acad. Sci. USA 85:5879-5883, 1988). Un fragmento "Fd" consiste en los dominios V_H y C_{H1} .

El término "diacuerpos" se refiere a pequeños fragmentos de anticuerpos con dos sitios de unión al antígeno, cuyos

30 fragmentos comprenden un dominio variable de cadena pesada (V_H) conectado a un dominio variable de cadena ligera (V_L) en la misma cadena polipeptídica (V_H-V_L). Utilizando un enlazador que sea demasiado corto como para permitir el emparejamiento entre los dos dominios de la misma cadena, se obliga a los dominios a emparejarse con los dominios complementarios de otra cadena y a crear dos sitios de unión al antígeno. Se describen diacuerpos de manera más completa, por ejemplo, en los documentos EP 404.097; WO 93/11161; y en Hollinger et al., Proc. Natl.

35 Acad. Sci. USA, 90:6444-6448 (1993).

Un "anticuerpo de dominio" es un fragmento de inmunoglobulina inmunológicamente funcional que contiene únicamente la región variable de una cadena pesada o la región variable de una cadena ligera. En algunos casos, para crear un anticuerpo de dominio bivalente, dos o más regiones V_H se unen mediante enlace covalente con un

40 enlazador peptídico. Las dos regiones V_H de un anticuerpo de dominio bivalente pueden dirigirse al mismo antígeno a antígenos diferentes.

La expresión "proteína de unión a antígeno" (ABP, *antigen binding protein*) incluye anticuerpos o fragmentos de anticuerpos, como se definió anteriormente, y péptidos u otros compuestos recombinantes que contienen secuencias

45 procedentes de las CDR que tienen las propiedades de unión al antígeno deseadas, de modo que se unen específicamente a un antígeno diana de interés.

En general, una proteína de unión a antígeno, por ejemplo, un anticuerpo o fragmento de anticuerpo, "se une específicamente" a un antígeno de interés cuando tiene una afinidad de unión significativamente más alta y, en consecuencia, es capaz de distinguir, ese antígeno, en comparación con su afinidad por otras proteínas no

50 relacionadas, en condiciones de ensayo de unión similares. Normalmente, se dice que una proteína de unión a antígeno se "une específicamente" a su antígeno diana cuando la constante de disociación (K_D) tiene un valor de 10^{-8} M o inferior. La proteína de unión al antígeno se une específicamente al antígeno con "alta afinidad" cuando la K_D tiene un valor de 10^{-9} M o inferior, y con "muy alta afinidad" cuando la K_D tiene un valor de 10^{-10} M o inferior.

55 Las expresiones "región de unión al antígeno" o "sitio de unión al antígeno", significan una parte de una proteína que se une específicamente a un antígeno especificado. Por ejemplo, la parte de una proteína de unión al antígeno que contiene los restos de aminoácidos que interactúan con un antígeno y que confieren a la proteína de unión al antígeno su especificidad y afinidad por el antígeno, se denomina "región de unión al antígeno". Una región de unión al antígeno

60 incluye normalmente una o más "regiones determinantes de la complementariedad" ("CDR"). Ciertas regiones de unión al antígeno también incluyen una o más regiones "estructurales" ("FR", siglas del inglés *framework*). Una "CDR" es una secuencia de aminoácidos que contribuye a la especificidad y afinidad de unión al antígeno. Las regiones "estructurales" pueden ayudar a mantener la conformación adecuada de las CDR para promover la unión entre la región de unión al antígeno y un antígeno. En un anticuerpo tradicional, las CDR están incluidas en una región

65 estructural en la región variable de la cadena pesada y ligera donde constituyen las regiones responsables de la unión

al antígeno y el reconocimiento del mismo. Una región variable de una proteína de inmunoglobulina de unión al antígeno, comprende al menos tres CDR de cadena pesada o ligera, véase, citado anteriormente (Kabat et al., 1991, Sequences of Proteins of Immunological Interest, Public Health Service N.I.H., Bethesda, Md.); véase también Chothia y Lesk, 1987, J. Mol. Biol. 196:901-917; Chothia et al., 1989, Nature 342: 877-883), dentro de una región estructural (denominadas regiones estructurales 1-4, FR1, FR2, FR3 y FR4, por Kabat et al., 1991, citado anteriormente; véase también Chothia y Lesk, 1987, citado anteriormente).

El término "antígeno" se refiere a una molécula o a una parte de una molécula que puede unirse a un agente de unión selectivo, tal como una proteína de unión a antígeno (incluyendo, por ejemplo, un anticuerpo o fragmento inmunológico funcional del mismo) y que además puede utilizarse en un animal para producir anticuerpos que pueden unirse a ese antígeno. Un antígeno puede poseer uno o más epítomos que pueden interactuar con diferentes proteínas de unión a antígeno, por ejemplo, anticuerpos.

Un "epítomo" es la parte de una molécula que está unida por una proteína de unión a antígeno (por ejemplo, por un anticuerpo). El término incluye cualquier determinante que pueda unirse específicamente a una proteína de unión a antígeno, tal como un anticuerpo o un receptor de linfocitos T. Un epítomo puede ser contiguo o no contiguo (por ejemplo, en un polipéptido monocatenario, los restos de aminoácidos que no son contiguos entre sí en la secuencia polipeptídica pero que dentro del contexto de la molécula están unidos por la proteína de unión al antígeno). En determinadas realizaciones, los epítomos pueden ser miméticos ya que comprenden una estructura tridimensional que es similar a la de un epítomo utilizado para generar la proteína de unión al antígeno, pero no comprenden ninguno o solo algunos de los restos de aminoácidos que se encuentran en ese epítomo utilizado para generar la proteína de unión al antígeno. Con mucha frecuencia, los epítomos se encuentran en las proteínas, pero en algunos casos, pueden encontrarse en otros tipos de moléculas, tales como ácidos nucleicos. Los determinantes epitópicos pueden incluir agrupaciones superficiales químicamente activas de moléculas tales como aminoácidos, cadenas laterales de azúcar, grupos fosforilo o sulfonilo, y pueden tener características estructurales tridimensionales específicas y/o características de carga específicas. En general, en una mezcla compleja de proteínas y/o macromoléculas, los anticuerpos específicos para un antígeno diana particular, reconocerán preferentemente un epítomo en el antígeno diana.

El término "identidad" se refiere a una relación entre las secuencias de dos o más moléculas polipeptídicas o dos o más moléculas de ácido nucleico, determinada mediante alineamiento y comparación de las secuencias. El "porcentaje de identidad" significa el porcentaje de restos idénticos entre los aminoácidos o nucleótidos en las moléculas comparadas y se calcula basándose en el tamaño de la más pequeña de las moléculas que se estén comparando. Para estos cálculos, los huecos en los alineamientos (en caso de haberlos) se abordan preferentemente mediante un modelo matemático o programa informático particular (es decir, un "algoritmo"). Los métodos que pueden usarse para calcular la identidad de los ácidos nucleicos o los polipéptidos alineados incluyen los descritos en Computational Molecular Biology, (Lesk, A. M., ed.), 1988, Nueva York: Oxford University Press; Biocomputing Informatics and Genome Projects, (Smith, D. W., ed.), 1993, Nueva York: Academic Press; Computer Analysis of Sequence Data, Parte I, (Griffin, A. M., y Griffin, H. G., eds.), 1994, Nueva Jersey: Humana Press; von Heinje, G., 1987, Sequence Analysis in Molecular Biology, Nueva York: Academic Press; Sequence Analysis Primer, (Gribskov, M. y Devereux, J., eds.), 1991, Nueva York: M. Stockton Press; y Carillo et al., 1988, SIAM J. Applied Math. 48:1073. Por ejemplo, la identidad de secuencia puede determinarse mediante métodos estándar que se utilizan comúnmente para comparar la similitud en la posición de los aminoácidos de dos polipéptidos. Utilizando un programa informático, tal como BLAST o FASTA, dos polipéptidos o dos secuencias de polinucleótidos se alinean para un emparejamiento óptimo de sus respectivos restos (ya sea a lo largo de la longitud completa de una o ambas secuencias, o a lo largo de una parte predeterminada de una o ambas secuencias). Los programas proporcionan una penalización de apertura por defecto y una penalización de hueco por defecto, y junto con el programa informático puede utilizarse una matriz de puntuación, tal como PAM 250 [una matriz de puntuación estándar; véase Dayhoff et al., en Atlas of Protein Sequence and Structure, vol. 5, supp. 3 (1978)]. Por ejemplo, el porcentaje de identidad puede calcularse como: el número total de emparejamientos idénticos multiplicado por 100 y después dividido entre la suma de la longitud de la secuencia más larga dentro del tramo emparejado y el número de huecos introducidos en las secuencias más largas para alinear las dos secuencias. Al calcular el porcentaje de identidad, las secuencias que se están comparando se alinean de manera que se proporciona el máximo grado de emparejamiento entre las secuencias.

El paquete del programa GCG es un programa informático que puede utilizarse para determinar el porcentaje de identidad, cuyo paquete incluye GAP (Devereux et al., 1984, Nucl. Acid Res. 12:387; Genetics Computer Group, University of Wisconsin, Madison, Wis.). El algoritmo informático GAP se utiliza para alinear los dos polipéptidos o polinucleótidos para los que se va a determinar el porcentaje de identidad de secuencia. Las secuencias se alinean para un emparejamiento óptimo de sus aminoácidos o nucleótidos respectivos (el "tramo emparejado", determinado por el algoritmo). Junto con el algoritmo, se utiliza una penalización por apertura de hueco (que se calcula como 3 veces la diagonal promedio, en donde la "diagonal promedio" es el promedio de la diagonal de la matriz de comparación que se está utilizando; la "diagonal" es la puntuación o el número asignado a cada emparejamiento perfecto de aminoácidos por la matriz de comparación particular) y una penalización por extensión de hueco (que generalmente es 1/10 veces la penalización por apertura de hueco), así como una matriz de comparación, tal como PAM 250 o BLOSUM 62. En determinadas realizaciones, para el algoritmo también se utiliza una matriz de comparación estándar (véase, Dayhoff et al., 1978, Atlas of Protein Sequence and Structure 5:345-352 para la matriz

de comparación PAM 250; Henikoff et al., 1992, Proc. Natl. Acad. Sci. U.S.A. 89: 10915-10919 para la matriz de comparación BLOSUM 62).

Los parámetros recomendados para determinar el porcentaje de identidad para secuencias de polipéptidos o de nucleótidos utilizando el programa GAP, incluyen los siguientes:

- Algoritmo: Needleman et al., 1970, J. Mol. Biol. 48:443-453;
- Matriz de comparación: BLOSUM 62 de Henikoff et al., 1992, citado anteriormente;
- Penalización por hueco: 12 (pero sin penalización por huecos terminales)
- 10 Penalización por longitud de hueco: 4
- Umbral de similitud: 0

Ciertos esquemas de alineamiento para alinear dos secuencias de aminoácidos pueden dar como resultado el emparejamiento de tan solo una región corta de las dos secuencias y esta región pequeña alineada puede tener una identidad de secuencia muy elevada, incluso a pesar de que no haya una relación significativa entre las dos secuencias de longitud completa. Por consiguiente, el método de alineamiento seleccionado (programa GAP) puede ajustarse, si así se desea, para dar como resultado un alineamiento que abarque al menos 50 aminoácidos contiguos del polipéptido diana.

20 El término "modificación", cuando se usa en relación con proteínas de interés (por ejemplo, inmunoglobulinas, incluyendo anticuerpos y fragmentos de anticuerpo), incluye, aunque sin limitación, uno o más cambios de aminoácidos (incluidas las sustituciones, inserciones o deleciones); modificaciones químicas; modificación covalente por conjugación con agentes terapéuticos o de diagnóstico; etiquetado (por ejemplo, con radionúclidos o varias enzimas); unión de polímero covalente tal como PEGilación (derivatización con polietilenglicol) e inserción o sustitución
25 por síntesis química de aminoácidos no naturales. Las inmunoglobulinas modificadas de la invención conservarán las propiedades de unión (o no unión) de las moléculas no modificadas de la invención. Mediante métodos conocidos por el experto en la materia, las inmunoglobulinas, tales como los anticuerpos u otras proteínas, pueden "genomanipularse" o modificarse para mejorar la afinidad, selectividad, estabilidad y/o capacidad de fabricación de la diana, antes de que la secuencia codificante de la proteína "genomanipulada" se incluya en el inventivo casete de
30 expresión.

El término "derivado", cuando se usa en relación con proteínas de interés, tales como inmunoglobulinas (incluidos anticuerpos y fragmentos de anticuerpos), se refiere a proteínas que se modifican por enlace covalente mediante conjugación con agentes terapéuticos o de diagnóstico, etiquetado (por ejemplo, con radionúclidos o varias enzimas),
35 unión de polímero covalente tal como PEGilación (derivatización con polietilenglicol) e inserción o sustitución por síntesis química de aminoácidos no naturales. Los derivados de la invención conservarán las propiedades de unión de las moléculas no derivatizadas.

En el alcance de la invención, las proteínas de interés pueden ser proteínas terapéuticas o "productos biológicos",
40 para el tratamiento de enfermedades, incluyendo, pero sin limitación, enfermedades humanas o trastornos humanos. "Tratamiento" o "tratar" es una intervención realizada con la intención de prevenir el desarrollo o alterar la patología de un trastorno. Por consiguiente, "tratamiento" se refiere tanto a tratamiento terapéutico como a medidas profilácticas o preventivas. Los que necesitan tratamiento incluyen los que ya tienen el trastorno así como en los que va a prevenirse la enfermedad. "Tratamiento" incluye cualquier indicación de éxito en la mejora de una lesión, patología o afección,
45 incluyendo cualquier parámetro objetivo o subjetivo tal como alivio; remisión; disminución de los síntomas o hacer que la lesión, la patología o la afección sea más tolerable para el paciente; ralentización de la velocidad de degeneración o empeoramiento; haciendo que el punto final de la degeneración sea menos debilitante; mejorando el bienestar físico o mental del paciente. El tratamiento o mejora de los síntomas puede basarse en parámetros objetivos o subjetivos; incluidos los resultados de una exploración física realizada por un médico u otro profesional sanitario, el autoinforme
50 de un paciente, los exámenes neuropsiquiátricos y/o la evaluación psiquiátrica.

Una "cantidad eficaz" de un agente terapéutico es generalmente una cantidad suficiente para reducir la gravedad y/o frecuencia de los síntomas, eliminar los síntomas y/o la causa subyacente, prevenir la aparición de los síntomas y/o su causa subyacente, y/o mejorar o remediar el daño resultante de una jaqueca o asociado a ella. En algunas
55 realizaciones, la cantidad eficaz es una cantidad terapéuticamente eficaz o una cantidad profilácticamente eficaz. Una "cantidad terapéuticamente eficaz" es una cantidad suficiente para remediar una patología (por ejemplo, rechazo de trasplante o EICH, inflamación, esclerosis múltiple, cáncer, enfermedad cardiovascular, diabetes, neuropatía, dolor) o uno o más síntomas, particularmente un estado o uno o más síntomas asociados con la patología, o de otra manera, para prevenir, impedir, retardar o revertir la progresión de la patología o de cualquier otro síntoma no deseable
60 asociado a la enfermedad de cualquier forma, sea la que sea (es decir, que proporcione "eficacia terapéutica"). Una "cantidad profilácticamente eficaz" es una cantidad de una composición farmacéutica que, cuando se administra a un sujeto, tendrá el efecto profiláctico esperado, por ejemplo, prevenir o retrasar la aparición (o reincidencia) de los síntomas de la jaqueca o de la esclerosis múltiple, o reducir la probabilidad de la aparición (o reincidencia) de la jaqueca, los síntomas de la jaqueca o los síntomas de la esclerosis múltiple. El efecto terapéutico o profiláctico
65 completo no tiene que producirse necesariamente después de la administración de una dosis, y puede producirse

después de la administración de una serie de dosis. Por tanto, se puede administrar una cantidad terapéutica o profilácticamente eficaz en una o más administraciones.

La proteína de interés puede ser cualquier proteína, pero en muchas realizaciones la proteína es una proteína o un péptido farmacológicamente activo.

En algunas realizaciones de la invención, la proteína de interés puede ser un péptido mimético o agonista. Los términos "péptido mimético", "peptidomimético", y "péptido agonista", se refieren a un péptido o a una proteína que tiene una actividad biológica comparable a la de una proteína de interés de origen natural. Estos términos incluyen además péptidos que imitan indirectamente la actividad de una molécula peptídica de origen natural, tal como potenciando los efectos de la molécula de origen natural.

En algunas realizaciones de la invención, la proteína de interés puede ser un péptido antagonista o un péptido inhibidor. El término "péptido antagonista", "antagonista peptídico", y "péptido inhibidor", se refiere a un péptido que bloquea o de alguna manera interfiere con la actividad biológica de un receptor de interés, o tiene una actividad biológica comparable a la de un antagonista o inhibidor conocido de un receptor de interés (tal como, pero sin limitación, un canal iónico o un receptor acoplado a proteína G (GPCR, *G-Protein Coupled Receptor*)).

Los ejemplos de proteínas farmacológicamente activas que pueden usarse en la presente invención incluyen, aunque sin limitación, un péptido de toxina, un péptido de unión a IL-6, una proteína de unión a CD3, una proteína de unión a CD19, una proteína de unión a CD20, una proteína de unión a CD22, una proteína de unión a HER2, una proteína de unión a HER3, una proteína de unión a VEGF-A, una proteína de unión a TNF- α , una proteína de unión a EGFR, una proteína de unión al ligando RANK, una proteína de unión a IL-1 α , una proteína de unión a IL-1 β , una proteína de unión a IL-17A, una proteína de unión a EPCAM (CD326), un antagonista del péptido CGRP, un antagonista del péptido del receptor de bradicinina B1, un péptido agonista de la hormona paratiroidea (PTH), un péptido antagonista de la hormona paratiroidea (PTH), un péptido de unión a ang-1, un péptido de unión a ang-2, un péptido de unión a miostatina, un péptido mimético de eritropoyetina (mimético de EPO), un péptido FGF21, un péptido mimético de trombopoyetina (mimético de TPO) (por ejemplo, AMP2 o AMPS), un péptido de unión al factor de crecimiento nervioso (NGF), un péptido de unión al factor de activación de linfocitos B (BAFF) y un péptido similar al glucagón (GLP)-1 o un peptidomimético del mismo o GLP-2 o un peptidomimético del mismo.

El término "análogo" de péptido o proteína se refiere a un polipéptido que tiene una secuencia que difiere de una secuencia peptídica existente en la naturaleza en al menos una sustitución de un resto de aminoácido, adición interna o deleción interna de al menos un aminoácido y/o truncamientos o adiciones de los extremos amino o carboxilo). Una "deleción interna" se refiere a la ausencia de un aminoácido de una secuencia existente en la naturaleza en una posición distinta al extremo N o C terminal. Asimismo, una "adición interna" se refiere a la presencia de un aminoácido en una secuencia existente en la naturaleza en una posición distinta al extremo N o C terminal.

Clonación de ADN

La clonación de ADN se lleva a cabo utilizando técnicas estándar (véase, por ejemplo, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Guide*, volúmenes 1-3, Cold Spring Harbor Press, que se incorpora en el presente documento como referencia). Por ejemplo, se puede construir una biblioteca de ADNc mediante transcripción inversa de ARNm de poliA⁺, preferiblemente ARNm asociado a la membrana, y la biblioteca puede explorarse usando sondas específicas de secuencias de genes de polipéptidos de inmunoglobulina humana. Sin embargo, en una realización, la reacción en cadena de la polimerasa (PCR, *polymerase chain reaction*) se usa para amplificar los ADNc (o partes de los ADNc completos) que codifican un segmento de interés del gen de inmunoglobulina (por ejemplo, un segmento variable de cadena ligera o pesada). Las secuencias amplificadas pueden clonarse fácilmente en cualquier vector adecuado, por ejemplo, en vectores de expresión, vectores minigénicos o vectores de presentación de fagos. Se apreciará que el método de clonación particular utilizado no es crítico, siempre que sea posible determinar la secuencia de alguna parte del polipéptido de inmunoglobulina de interés.

Una fuente de ácidos nucleicos que codifican anticuerpos es un hibridoma producido obteniendo un linfocito B de un animal inmunizado con el antígeno de interés y fusionándolo con una célula inmortal. Como alternativa, el ácido nucleico se puede aislar de los linfocitos B (o del bazo completo) del animal inmunizado. Otra fuente más de ácidos nucleicos que codifican anticuerpos es una biblioteca de tales ácidos nucleicos generados, por ejemplo, a través de la tecnología de presentación de fagos. Los polinucleótidos que codifican péptidos de interés, por ejemplo, péptidos de región variable con características de unión deseadas, pueden identificarse mediante técnicas estándar tales como barrido (o paneo).

La secuencia que codifica toda una región variable del polipéptido de inmunoglobulina puede determinarse; sin embargo, a veces será adecuado secuenciar solo una parte de una región variable, por ejemplo, la parte que codifica la CDR. La secuenciación se lleva a cabo utilizando técnicas estándar (véase, por ejemplo, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Guide*, volúmenes 1-3, Cold Spring Harbor Press y Sanger, F. et al. (1977) *Proc. Natl. Acad. Sci. USA* 74: 5463-5467, que se incorporan en el presente documento como referencia). Al comparar la

secuencia del ácido nucleico clonado con las secuencias publicadas de genes de inmunoglobulina humana y ADNc, un experto en la materia podrá determinar fácilmente, dependiendo de la región secuenciada, (i) el uso del segmento de la línea germinal del polipéptido de inmunoglobulina del hibridoma (incluido el isotipo de la cadena pesada) y (ii) la secuencia de las regiones variables de la cadena pesada y ligera, incluyendo las secuencias resultantes de la adición de la región N y del proceso de mutación somática. Una fuente de información sobre la secuencia de genes de inmunoglobulina es el National Center for Biotechnology Information, la National Library of Medicine y los National Institutes of Health, Bethesda, Md.

El ADN aislado puede unirse operativamente a secuencias de control o colocarse en vectores de expresión, que después se transfectan en células hospedadoras que de otro modo no producen la proteína de inmunoglobulina, para dirigir la síntesis de anticuerpos monoclonales en las células hospedadoras recombinantes. La producción recombinante de anticuerpos es bien conocida en la técnica.

El ácido nucleico está unido operativamente cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN de una presecuencia o secuencia líder secretora está unido operativamente a ADN de un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o potenciador está unido operativamente a una secuencia codificante si afecta a la transcripción de la secuencia; o un sitio de unión al ribosoma está unido operativamente a una secuencia codificante si se coloca de manera que facilite la traducción. En general, unido operativamente significa que las secuencias de ADN que se unen son contiguas y, en el caso de una secuencia líder secretora, contiguas y en fase de lectura. Sin embargo, no es necesario que los potenciadores sean contiguos. La unión se realiza mediante ligamiento en sitios de restricción convenientes. Si no existen dichos sitios, se usan adaptadores o enlazadores oligonucleotídicos sintéticos de acuerdo con la práctica habitual.

En la técnica se conocen muchos vectores. Los componentes del vector pueden incluir uno o más de lo siguiente: una secuencia señal (que puede, por ejemplo, dirigir la secreción del anticuerpo; por ejemplo, ATGGACATGAGGGTGCCCGCTCAGCTCCTGGGGCTCCTGCTGCTGTGGCTGAG AGGT GCGCGCTGT//SEQ ID NO:7, que codifica la secuencia del péptido señal VK-1 MDMRVPAQLLGLLLLWLRGARC//SEQ ID NO:8), un origen de replicación, uno o más genes marcadores selectivos (que pueden, por ejemplo, conferir resistencia a antibióticos o a otros fármacos, complementar las deficiencias auxotróficas o suministrar nutrientes fundamentales no disponibles en los medios), un elemento potenciador, un promotor y una secuencia de terminación de la transcripción, todos ellos bien conocidos en la técnica.

Ejemplos

Ejemplo 1. Materiales y métodos.

Construcción de vectores. Todos los vectores C, D, E, F y G, incluían una secuencia de ADN que codifica una proteína exógena de interés "Fc-A" (o "FC-A"; una fusión de Fc ejemplar con un anticuerpo antiinflamatorio terapéutico), un IRES (siglas del inglés *Internal Ribosome Entry Site*, Sitio Interno de Entrada al Ribosoma) procedente del virus EMCV (virus de la encefalomiocarditis), el gen murino de la dihidrofolato reductasa y una poliadenilación tardía de 221 pb procedente del virus SV40 como se describió anteriormente (Kaufman, R.J. et al., Improved vectors for stable expression of foreign genes in mammalian cells by use of the untranslated leader sequence from EMC virus, Nucl. Acids Res. 19(16): 4485-4490 (1991); Schek, et al., Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using *in vitro* analyses, Mol. Cell. Biol. 12(12):5386-93 (1992)). La unión entre el IRES y la DHFR fue *aaaacacgatTGCTCGAGAGTTGCCACCCATCafg* //SEQ ID NO:6, donde las letras minúsculas representan la secuencia de IRES endógena y el sitio de inicio de la transcripción ATG de la DHFR está en minúsculas en cursiva. A partir de estas diversas construcciones, se inicia una sola transcripción, que después incluye el gen de Fc-A, el IRES del EMCV, el gen murino de la DHFR y termina en el sitio de poliadenilación de SV40.

En la secuencia descrita anteriormente, se incluyeron varias secuencias promotoras de la transcripción en 5' o en 5' y 3'. Esto permitió seleccionar todos los plásmidos utilizando el gen de DHFR y Fc-A como marcador de la expresión relativa. En todos los plásmidos, la cadena principal del plásmido era pUC57-simple Genscript (Piscataway, NJ).

En la Figura 1 se representan esquemáticamente varios plásmidos ejemplares. El vector C contenía las secuencias intrónicas de CMV humano y del EF 1 alfa humano descritas anteriormente (Kim, Lee, Shin, Kang, y Kim, 2002). El vector F (Figura 1) contenía secuencias que flanqueaban el gen del EF 1 alfa endógeno de células de ovario de hámster chino (CHO) como se describe (Running Deer, J., y Allison, D. S., High-level expression of proteins in mammalian cells using Transcription Regulatory sequences from Chinese Hamster Ovary EF-1alpha Gene, Biotechnology Progress 20: 880-889 (2004)) con secuencias en 5' de 4 083 pb que flanquean el codón ATG y secuencias en 3' de 4 174 pb que flanquean después del sitio de poliadenilación. El Vector G (Figura 1) se construyó eliminando una secuencia de 240 pb del sitio de inicio de la transcripción del gen del EF 1 alfa de CHO en el sitio FSEI y reemplazándola por un fragmento de ADN que abarcaba el fragmento de CMV murino desde la posición -615 hasta el sitio de inicio de la transcripción. El vector D (Figura 1) se construyó sintetizando el fragmento mostrado en la SEQ ID NO: 1 (que contenía el CMV murino GenBank: L06816.1, nucleótidos 4067-4682, el intrón del EF 1 alfa de rata

GenBank: AC158987.3, nucleótidos 22137-21728) y colocando estas secuencias en 5' de la secuencia codificante de Fc-A. El vector C y el vector D también contenían la secuencia líder tripartita de adenovirus que reemplaza la secuencia líder no traducida 3' endógena (Kaufman et al., 1991). El vector pUC57-simple no se muestra en la Figura 1. El mapa completo del Vector D se muestra en la Figura 2.

5

Electroporación y cultivo de células. La electroporación de larga duración se realizó como se describió anteriormente (Bodwell, et al., Long Duration electroporation for achieving high level expression of glucocorticoid receptors in mammalian cell lines. *J. Steroid Biochem. y Mol. Biol.*, 68(e 8), 77-82 (1999)). Resumiendo, 2×10^7 células se resuspendieron en 0,3 ml de tampón HBS (HEPES 0,01 M pH 7,4, NaCl 0,15 M, EDTA 3 mM, Polisorbato20 al 0,005 % v/v), se añadieron 25 mg de ADN y se sometieron a electroporación en una cubeta de 4 mm de espacio usando un BTX EM 600 con una capacitancia de $3175 \mu\text{F}$, 720 ohmios y 200 voltios.

Las células DXB11 y DG44 se cultivaron en PowerCHO2 (Lonza), ProCHO4 (Lonza) o Ex-cell 302 (Sigma), siguiendo las instrucciones del fabricante. Las células se cultivaron cada 3-4 días y se sembraron a 4×10^5 células/ml durante 3 días en cultivo o a 3×10^5 células/ml durante 4 días en cultivo.

El título de proteína se determinó mediante cromatografía líquida de alta resolución (HPLC, *Performance Liquid Chromatography*) de afinidad, utilizando una columna de proteína A POROS A/20.

20 Ejemplo 2. Transfección de células CHO

Con diversas construcciones que expresaban la proteína de fusión Fc-A, se transfectaron células CHO. En las Figuras 1 y 2 se muestran estas construcciones. Se compararon grupos de células transfectadas para determinar su expresión con respecto a la proteína de fusión Fc-A.

25

Transfección de células DXB11. Utilizando electroporación, los diversos plásmidos de expresión de Fc-A se utilizaron para transfectar células DXB11, una línea celular de CHO mutante con respecto a DHFR. Estas células se cultivaron en ProCHO4 y primero se seleccionaron para el crecimiento en medios que carecían de hipoxantina y timidina y después se seleccionaron con metotrexato (MTX) 150 nM. Grupos de células seleccionadas con MTX 150 nM se sembraron a 1×10^6 células/ml y los días 2, 4 y 7 se alimentaron con Ex-cell Advanced CHO Feed 1 (SAFC) y glucosa. El fluido sobrenadante se recogió el día 8 y se analizó utilizando una columna de proteína A poros como se describió anteriormente.

En las Figuras 3A y 3B se muestra el título de proteína Fc-A y la productividad específica (picogramos por célula por día), respectivamente. Los resultados demostraron que el promotor híbrido que contenía la secuencia potenciadora de CMV murino combinada con una secuencia intrónica del EF 1 alfa de rata y secuencias líder adecuadas (por ejemplo, en este caso, la secuencia líder tripartita de adenovirus; vector D), aumentaba la expresión de la proteína recombinante Fc-A mediante grupos de células transfectadas. Los títulos de proteína obtenidos fueron superiores a los de otras construcciones y la productividad específica fue tan buena o mejor que la obtenida utilizando otras construcciones.

Transfección de DXB-11 con los Vectores D, F y G. Los Vectores D, F y G (Figura 1) se utilizaron para transfectar células DXB11 que crecían en PowerCHO2. Las células se seleccionaron primero para el crecimiento en medios que carecían de hipoxantina y timidina, se amplificaron en MTX a 150 nM y después se re-amplificaron en MTX 500 nM para mejorar aún más la expresión. Estos grupos se sembraron a 1×10^6 células/ml y los días 2, 3 y 6 se alimentaron con Ex-cell Advanced CHO Feed 1 y glucosa. El fluido sobrenadante se recogió el día 8. Como se muestra en las Figuras 4A y 4B, respectivamente, el título de los grupos transfectados con el Vector D fue significativamente mayor en comparación con el de los transfectados con los Vectores F y G. Las productividades específicas del Vector D fueron generalmente más altas que las de los Vectores F y G.

50

Transfección de DG44 y DXB-11 con los Vectores D y F. Células CHO se transfectaron con los Vectores D y F que expresaban la proteína de fusión Fc-A. Las células DXB11 que crecían en Ex-cell 302 se transfectaron como se describió anteriormente. DG44, otra línea celular mutante con respecto a DHFR, se cultivó en PowerCHO2 y se transfectó de la misma manera que las células DXB11. Se compararon grupos de células transfectadas para determinar su expresión con respecto a Fc-A. Las células se transfectaron utilizando electroporación, se seleccionaron para el crecimiento en medios que carecían de hipoxantina y timidina, y después se amplificaron en MTX 150 nM. Adicionalmente, las células DXB11 se amplificaron en MTX 500 nM y las células DG44 se amplificaron adicionalmente en MTX 1 μM . A continuación, grupos de células se sembraron mediante una dilución 1:5 en medio de crecimiento complementado con Ex-cell Advanced CHO Feed 1, los días 3, 6 y 8 se alimentaron con Ex-cell Advanced CHO Feed 1 y glucosa. El fluido sobrenadante se recogió el día 10 y se analizó utilizando una columna de proteína A poros como se describió anteriormente. En las Figuras 5A y Figura 5B se muestra el título y la productividad específica, respectivamente.

Los grupos de células DXB11 transfectadas con el Vector D expresaron casi el doble de proteína de fusión Fc-A que los grupos de células transfectadas con el Vector F. Los grupos de células DG44 transfectadas con los Vectores D y

65

F expresaron a niveles similares. Sin embargo, los grupos de células transfectadas con el Vector D tenían una productividad específica ligeramente superior. La menor productividad específica de las células DG44 en comparación con las células DXB11, implica que las células DG44 son menos sensibles al MTX que las células DXB 11.

- 5 El vector D también incluyó una secuencia líder no traducida en 5' (5'UTR), procedente de la secuencia líder tripartita (TPL) de adenovirus. Se ha demostrado que la secuencia TPL mejora la expresión de las proteínas en condiciones de estrés (Logan & Shenk, Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. Proc. Natl. Acad. Sci. USA 81(12): 3655-59 (1984)). En estos estudios no se aclara si la secuencia TPL mejora la expresión, pero otros estudios han demostrado que las secuencias exactas de nucleótidos de la 5'UTR que pueden usarse en esta posición, varían siempre que no contengan grandes estructuras secundarias o codones iniciadores antes del auténtico codón iniciador (Mignone, F. et al., Untranslated regions of mRNAs. Genome Biology, 3(3):reviews0004.1-0004.10 (2002)).

Ejemplo 3. Expresión inducible por tetraciclina del inventivo promotor híbrido

- 15 En las células que expresan el represor Tet (TetR), la tetraciclina puede utilizarse para regular la expresión de los promotores que contienen la secuencia del operador Tet (Yao et al.). La introducción de la secuencia del operador Tet (TetO) justo en la posición 3' de la caja TATA, impide la transcripción de este promotor en presencia del TetR. Presumiblemente, TetR se une a TetO e impide que los factores de transcripción interactúen con el sitio de inicio de la transcripción o interfieran con la formación del complejo de inicio de la transcripción (Yao et al). En particular, la colocación de las secuencias de TetO es fundamental para determinar si TetR puede modular eficazmente la transcripción. (Véase, Yao et al., Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells, Hum. Gene Ther. 9(13):1939-50 (1998)).
- 20
- 25 Yao et al. plantearon la hipótesis de que la colocación de la secuencia de TetO 10 pb cadena abajo de la secuencia TATATAA, permitía la unión de TetR a la misma superficie que la de la proteína de unión a TATA. De acuerdo con esta hipótesis, Kim et al., descubrieron que la inserción de las secuencias de TetO en la posición 0 o 15 cadena abajo de la secuencia TATAAG de la proteína US11 de hCMV, no logró preservar la represión en presencia de TetR. (Véase, Kim et al., Tetracycline repressor-regulated gene repression in recombinant human cytomegalovirus, J. Virol. 69: 2565-30 257 (1995)).

- Al añadir tetraciclina (Tet) al cultivo, ésta se unirá a TetR e inhibirá la unión con la secuencia de TetO, permitiendo así la transcripción. Para crear un poderoso promotor regulado por Tet basado en la ingeniosa secuencia potenciadora de mCMV de la invención/secuencias intrónicas del promotor híbrido del EF 1a de rata, primero se generó un plásmido que dirigía la expresión de LacZ (proteína beta-galactosidasa), JV56_pJV39_pJV10_pUC57S (abreviado en el presente documento como "pJV56"; Figura 6). El marco abierto de lectura de LacZ es SEQ ID NO: 12:
- 35

ATGGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGA
CTGGTGGACAGCAAATGGGTCTGGGATCTGTACGACGATGACGATAAGGTACC
TAAGGATCAGCTTGGAGTTGATCCCGTCGTTTTACAACGTCGTGACTGGGAAA
ACCCTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGC
TGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCA
GCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAAGCGGTGCC
GGAAAGCTGGCTGGAGTGCATCTTCTGAGGCCGATACTGTCTGTCGTCCTCCCT
CAAACCTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTAACCTA
TCCATTACGGTCAATCCGCCGTTTTGTTCCACGGAGAATCCGACGGGTTGTT
ACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCG
AATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCT
GGGTCTGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCA
TTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCGTTGGAGTG
ACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCG
TGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTG
CCTACTCGCTTTAATGATGATTTACAGCCGCGCTGTACTGGAGGCTGAAGTTCAG
ATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGG
TGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGAT
GAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACC

CGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACT
GCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTC
CGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGC
TGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATG
GATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACCTTA
ACGCCGTGCGCTGTTTCGATTATCCGAACCATCCGCTGTGGTACACGCTGTGC
GACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCA
TGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAG
CGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATC
ATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGT
ATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGC
GGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGG
ATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAATGGCT
TTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCACGCG
ATGGGTAACAGTCTTGGCGTTTTCGCTAAATACTGGCAGGCGTTTTCGTCAGTA
TCCCGTTTTACAGGGCGGCTTCGCTCTGGGACTGGGTGGATCAGTCGCTGATTA
AATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGA
TACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCA
CGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTT
CCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATA
GCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGC
AAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAA
CTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTAC
GCGTAGTGCAACCGAACGCGACCGCATGGTTCAGAAGCCGGGCACATCAGCGC
CTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCG
TCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCT
GGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGT
GGATTGGCGATAAAAAACAACCTGCTGACGCCGCTGCGCGATCAGTTCACCCG
TGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGCATTGACCCT
AACGCCTGGGTGCAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCA
GCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGAC
CGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACC
TACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGC
GAGCGATAACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAG

GTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACATATCCCG
 ACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATG
 TATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCG
 AATTGAATTATGGCCCACACCAGTGGCGCGGGCGACTTCCAGTTCAACATCAG
 CCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCAC
 GCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTG
 GCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAGTTCCAGCTGAGCGCCGG
 TCGCTACCATTACCAGTTGGTCTGGTGTCAAAAATAA//SEQ ID NO:12.

La idea básica era intercambiar parte de la secuencia del elemento promotor de mCMV (mCMV-P) con el elemento promotor de CMV humano (hCMV-P) que incluye los sitios de unión de TetO responsables de la regulación por tetraciclina (Tet). Esta construcción, JV57_JV56_pJV39_pJV10 (abreviada en el presente documento como "pJV57"; Figura 7), incluye el elemento regulador de TetO sensible a tetraciclina para la expresión de LacZ. La Figura 10 muestra esquemáticamente el hCMV-P/TetO insertado en la secuencia promotora de mCMV. Esta inserción reemplaza parte de la secuencia promotora de mCMV-P con una parte de la secuencia promotora de hCMV-P. Se creó otra construcción para incorporar cambios que aumentaban la expresión en un promotor de hCMV descrito por Patwardhan et al. en el contexto de un promotor no regulado (potenciador híbrido de mCMV/promotor de hCMV (parcial)/intrón del EF 1 alfa de rata). (Patwardhan et al., High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis, *Nature Biotechnology* 27(12):1173-75 (2009)). Esta construcción fue JV59_JV56_pJV39_pJV10 (abreviada en el presente documento como "pJV59"; Figura 8). En otra construcción, JV60_JV56_pJV39_pJV10 (abreviada en el presente documento como "pJV60"; Figura 9), las secuencias de TetO de pJV57 se cambiaron para emparejarse con las secuencias que aumentaban la expresión alrededor del sitio de inicio de la transcripción en el estudio de Patwardhan et al. (véanse las Figuras 9 y 12).

La Figura 11 muestra la secuencia de ADN de hCMV-P/TetO que se utilizó para reemplazar las secuencias promotoras de mCMV (mCMVP). Las flechas indican sitios de unión de TetR palindrómicos. Las secuencias promotoras de mCMV-P y hCMV-P mostradas, son las secuencias desde la caja TATA hasta el sitio de inicio de la transcripción. El inicio de la transcripción es el resto g en la secuencia taccg en la posición 3'. Hay que tener en cuenta que la secuencia de TetO ciertamente impacta en el sitio de inicio de la transcripción, ya que normalmente el sitio de inicio de la transcripción está ~ 30 bases en 3' con respecto a T en 5' de la caja TATA.

Para comprobar el nivel de expresión proteico se expresó la beta-galactosidasa a partir de varias construcciones. Esta secuencia sustituta insertada, que incluía elemento potenciador de muCMV (mCMV-E)/promotor de huCMV (parcial)/TetO/intrón del EF 1 alfa de rata, estaba en efecto regulada por tetraciclina en el medio (Véase, por ejemplo, la Figura 14, pJV57 +/- Tet). Por otro lado, la construcción pJV57 produjo aproximadamente 10 veces más proteína recombinante que un control de CMV humano TetO (pcDNA5/TO/lacZ; Thermo Fisher Scientific), cuando había Tet en el medio, pero aproximadamente 3,5 veces menos beta-galactosidasa en ausencia de tetraciclina, como se muestra en la Figura 14. Células CHO T-Rex, que llevaban el promotor híbrido potenciador de muCMV/intrón del EF 1 alfa de rata en pJV56, en ausencia de una secuencia de TetO dentro del segmento CMV-P de la secuencia potenciadora de mCMV, también expresaron aproximadamente 10 veces más proteína que el control de CMV humano/TetO (pcDNA5/TO/lacZ), independientemente de si había tetraciclina en el medio.

35 Materiales y métodos.

Células CHO T-REx™. La línea celular CHO T-REX (Thermo Fisher Scientific, número de producto R71807) se cultivó en medio F12 de Ham + glutamina + Blast 10 µg/ml, siguiendo el protocolo del fabricante. Las células CHO T-REx se sometieron a electroporación utilizando el método de electroporación de larga duración (LDE, siglas del inglés *long duration electroporation*) descrito por Bodwell et al. (Bodwell et al., Long Duration Electroporation for Achieving High Level Expression of Glucocorticoid Receptors in Mammalian Cell Lines, *J. Steroid Biochem. Molec. Biol.* 68:77-82 (1999)), con varias construcciones pJV56, pJV57, pJV59 y pJV60, que expresan la proteína beta-galactosidasa, como se muestra en la Figuras 6, 7, 8 y 9, respectivamente. Se permitió que las células transfectadas se recuperasen durante 24 horas y después se trataron durante 24 horas con o sin tetraciclina 1 µg/ml. Las células se sometieron a lisis y la actividad beta-galactosidasa se analizó usando un kit de ensayo de β-Gal siguiendo el protocolo del fabricante (Thermo Fisher Scientific, número de producto K1455-01). Resumiendo, los lisados se diluyeron 10 veces y para determinar la concentración de proteínas del lisado, se analizaron alícuotas de 10 µl utilizando un kit de ensayo de proteínas BCA (Thermo Fisher Scientific, número de producto 23227). Los lisados se diluyeron 100 veces y para determinar la

actividad beta-galactosidasa se analizaron alícuotas de 10 μ . Todas las muestras se transfectaron al menos por duplicado. La actividad beta-galactosidasa específica se normalizó a la cantidad de proteína analizada y la actividad de fondo de las células transfectadas de forma simulada se restó de las muestras transfectadas. Los resultados se representaron gráficamente mostrando la desviación estándar (Figura 14).

5 Células CHO-K1/TetR. En experimentos distintos, el represor de tetraciclina (TetR) se subclonó a través de Genscript® (Piscataway, NJ) en el vector de expresión pcDNA3.1 + para producir el vector pJV40_pcDNA3.1+tetR (abreviado en el presente documento como "pJV40"; Figura 15). Las células CHO-K1 se cultivaron de la manera habitual en medio CHO sin suero, PowerCHO™ 2 químicamente definido (Lonza, número de producto BE12-771Q) y las células se
10 sometieron a electroporación utilizando el método de LDE, como se ha descrito anteriormente para pJV40. Los grupos de células transfectadas se seleccionaron durante 10 días en PowerCHO™ 2 + G418 600 μ g/ml (Gibco, número de producto 10131027). A continuación, las células se diluyeron en serie y se sembraron en microplacas de 96 pocillos (Corning® CellBIND® 96 Well, número de producto 3340) a 0,75 células por pocillo en medio de clonación (medio sin suero EX-CELL® 302 para células CHO; Sigma-Aldrich, número de producto 14326C SIGMA) + medio acondicionado
15 al 15 % en presencia de G418). Las poblaciones clonales se identificaron después de 11-12 días y se transfirieron a placas de 12 pocillos en medios de clonación 50/50/medio PowerCHO™ 2 durante 3-4 días más. Finalmente, los clones se cambiaron y se cultivaron en una placa profunda de 24 pocillos (VWR, número de producto P-DW-10ML-24-CS) con agitación constante (220 rpm) en medio PowerCHO™ 2 + G418 600 μ g/ml.

20 Para explorar clones que expresaran altos niveles de TetR, los clones CHO-K1/TetR se transfectaron con pJV57 utilizando el kit de transfección ExpiFectamine™ CHO (Thermo Fisher Scientific, número de producto A29130) siguiendo el protocolo del fabricante. De los clones de TetR+ con alta expresión, se eligieron algunos clones posteriormente para ser transfectados con el vector pJV56, pJV57, pJV59 o pJV60, utilizando el kit de transfección ExpiFectamine™ CHO (Thermo Fisher Scientific, número de producto A29130), nuevamente transfectados siguiendo
25 el protocolo del fabricante. Se permitió que los clones transfectados se recuperasen durante 24 horas y después se trataron con tetraciclina 1 μ g/ml durante 24 horas. Para medir la expresión del producto proteico LacZ, las células se sometieron a lisis y la actividad beta-galactosidasa se analizó utilizando un kit de ensayo de β -Gal siguiendo el protocolo del fabricante (Thermo Fisher Scientific, número de producto K1455-01).

30 Resumiendo, los lisados se diluyeron 10 veces y para determinar la concentración de proteínas del lisado, se analizaron alícuotas de 10 μ utilizando un kit de ensayo de proteínas BCA (Thermo Fisher Scientific, número de producto 23227). Para determinar la actividad beta-galactosidasa, se analizaron diez (10) μ l de lisado no diluido. Todas las muestras se transfectaron al menos por duplicado. La actividad beta-galactosidasa específica se normalizó a la cantidad de proteína analizada y la actividad de fondo de las células transfectadas de forma simulada se restó de las
35 muestras transfectadas. Los resultados se representaron gráficamente mostrando la desviación estándar (véanse, por ejemplo, las Figuras 16, 17 y 18). Las Figuras 16 (clon 3E7), 17 (clon 3F9) y 18 (clon 4G2) muestran resultados representativos de los clones CHO-K1/TetR transfectados con diversas construcciones (pJV56,

pJV57, pJV59 y pJV60) que expresan la proteína beta-galactosidasa (o transfectados con control positivo: pcDNA5/TO/lacZ; Thermo Fisher Scientific, número de producto V1033-20). Como se ha indicado anteriormente, las
40 células se transfectaron con pJV57 utilizando el kit de transfección ExpiFectamine™ CHO (Thermo Fisher Scientific, número de producto A29130) siguiendo el protocolo del fabricante. Se permitió que los clones transfectados se recuperasen durante 24 horas y después se trataron con tetraciclina 1 μ g/ml durante 24 horas. Las células se sometieron a lisis y la actividad beta-galactosidasa se analizó utilizando un kit de ensayo de β -Gal siguiendo el
45 protocolo del fabricante (Thermo Fisher Scientific, número de producto K1455-01). Resumiendo, los lisados se diluyeron 10 veces y para determinar la concentración de proteínas del lisado, se analizaron alícuotas de 10 μ utilizando un kit de ensayo de proteínas BCA (Thermo Fisher Scientific, número de producto 23227). Para determinar la actividad beta-galactosidasa, se analizaron diez (10) μ l de lisado no diluido. Todas las muestras se transfectaron al menos por duplicado. La actividad beta-galactosidasa específica se normalizó a la cantidad de proteína analizada y la
50 actividad de fondo de las células transfectadas de forma simulada se restó de las muestras transfectadas. La actividad beta-galactosidasa específica se normalizó a la cantidad de proteína analizada y la actividad de fondo de las células transfectadas de forma simulada se restó de las muestras transfectadas. Los resultados se representaron gráficamente mostrando la desviación estándar.

55 Resultados.

Células CHO T-REx™. En experimentos con células CHO T-REx (Thermo Fisher Scientific), una línea celular comercial que expresa de manera estable TetR, se midió la actividad beta-galactosidasa (beta-gal) como una lectura para probar la capacidad de la tetraciclina (Tet) para regular la construcción pJV57. Como se muestra en la Figura 14,
60 la construcción pJV57 tenía una expresión equivalente a la construcción pJV56 cuando se inducía debido a la presencia de tetraciclina (+Tet). Además, la construcción pJV57 tenía una expresión aproximadamente 10 veces mayor de LacZ en presencia de tetraciclina (+Tet) en comparación con pcDNA5/TO/lacZ, un vector que muestra una expresión contundente regulada por tetraciclina. La construcción pJV57 también mostró una expresión regulada 3,5 veces por tetraciclina, una observación clave que demuestra que es un promotor poderosamente regulado. El factor
65 de diferencia en la regulación de Tet para el control positivo pcDNA5/TO/lacZ no se incluyó en la Figura 14, debido a

la baja relación de la señal con respecto al ruido entre pcDNA5/TO/lacZ no inducido y el fondo, dando como resultado una actividad beta-gal específica negativa en el control.

En un intento por mejorar aún más la construcción pJV57, se planteó la hipótesis de que las secuencias de TetO podrían reducir la expresión de pJV57, ya que estudios anteriores demostraron que los cambios alrededor del sitio de inicio de la transcripción reducían la expresión del promotor de hCMV (Patwardhan et al., High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis, *Nat Biotechnol.* 27(12): 1173-1175 (2009)). Las secuencias de TetO en pJV57 se cambiaron para emparejarse con las secuencias que aumentaban la expresión alrededor del sitio de inicio de la transcripción en el estudio de Patwardhan et al. (pJV60, Figuras 9 y 12). Por otro lado, estos cambios se diseñaron para mantener la unión de TetR de alta afinidad (la afinidad de diferentes mutaciones de TetO se describe en Sizemore et al. y se revisa en Hillen y Berens). (Véase, Sizemore et al., Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants, *Nucleic Acids Research* 18 (10) 2875-2880 (1990); Hillen y Berens, Mechanisms underlying expression of Tn10 encoded tetracycline resistance, *Ann. Rev. Micro.* 48:345-369 (1994)). Sin embargo, esta construcción optimizada, pJV60, no pareció aumentar la expresión en comparación con pJV57 en células CHO T-REx (Figura 14). Se mantuvo la expresión regulada por tetraciclina, aunque posiblemente la regulación era ligeramente inferior (2,68 veces en comparación con 3,50 veces (Figura 14).

Las secuencias optimizadas del promotor de hCMV (hCMV-P) sustituyeron a las secuencias del promotor de mCMV (mCMV-P) en el promotor híbrido secuencia potenciadora de mCMV/intrón del EF 1 a de rata de la invención (véase la comparación de secuencias en la Figura 13). Los cambios incorporados en pJV59 no parecieron aumentar la expresión en relación con pJV56 (Figura 14). Por tanto, se llegó a la conclusión de que en las células CHO T-REx, los cambios alrededor del sitio de inicio de la transcripción tienen poco o ningún impacto sobre la expresión, en el contexto del poderoso promotor de secuencia potenciadora de mCMV/intrón del EF 1 alfa de rata.

Células CHO-K1/TetR. En experimentos distintos, se generó una línea celular CHO-K1 que expresaba TetR de manera estable mediante la transfección de pJV40 (véase la Figura 15) que tiene el gen TetR subclonado en el vector pcDNA3.1+. Los grupos de CHO-K1 resistentes a genética se diluyeron en serie para obtener colonias unicelulares y se exploraron para que tuvieran una alta expresión de TetR, es decir, represión de más del doble de LacZ en ausencia de Tet (datos no mostrados). Se seleccionaron tres clones para un análisis adicional, como se muestra respectivamente en las Figuras 16, 17 y 18, denominados en el presente documento clones CHO-K1/TetR. Los clones se transfectaron transitoriamente con pcDNA5/TO/lacZ como control, y los vectores de expresión de la presente invención, es decir, los vectores pJV56, pJV57, pJV59 y pJV60, como se ha descrito anteriormente, para determinar la eficacia de un elemento promotor regulado por Tet, basándose en las secuencias del promotor híbrido secuencia potenciadora de mCMV/secuencias intrónicas del EF 1 alfa de rata, utilizando células generadas internamente que expresaban TetR.

Como se muestra en las Figuras 16, 17 y 18, en dos de los 3 clones CHO-K1/TetR, la construcción pJV56 mostró una expresión de LacZ de 2 a 3 veces mayor en comparación con pcDNA5/TO/lacZ en presencia de tetraciclina (+Tet). La adición de secuencias de hCMV-P/TetO en pJV57 permitió a la tetraciclina regular la expresión de la proteína hasta un nivel de potenciación de 3,24-, 9,30- y 3,15 veces la expresión de LacZ en presencia de tetraciclina (+Tet), en comparación con (-Tet) (véase, pJV57, en las Figuras 16, 17 y 18 respectivamente), aunque el nivel de expresión fue variable dependiendo del clon estudiado. Esta es una comparación con pJV56, que no incluía las secuencias de hCMV-P/TetO (1,27-, 0,95-, 1,19 veces de diferencia entre los tratamientos con tetraciclina (+Tet) y sin tetraciclina (-Tet), como se muestra en las Figuras 16, 17 y 18, respectivamente).

De manera interesante, la expresión total de LacZ aumentó en los tres clones CHO-K1/TetR transfectados con pJV59, aunque, como era de esperar, no hubo esencialmente ninguna regulación de la expresión sensible a la tetraciclina (comparación con \pm Tet: diferencia de 0,99-, 1,06-, 1,00 veces), como se muestra en las Figuras 16, 17 y 18, en consonancia con nuestra hipótesis de que la expresión proteica puede aumentar optimizando las secuencias que rodean el sitio de inicio de la transcripción (véase, Patwardhan et al.). Sin embargo, la adición de sitios TetO que se mutaron para mantener la unión de TetR (pJV60) - en el contexto del promotor optimizado - mostró una expresión de LacZ disminuida (Figuras 16 y 17) o sin cambios (Figura 18) en comparación con pJV59. No obstante, la expresión de LacZ aumentó en los tres clones portadores de pJV60 en comparación con pJV57, en presencia de tetraciclina (2,31-, 2,82 y 1,56 veces mayor expresión en comparación con su control no tratado con Tet; Figuras 16, 17 y 18, respectivamente). Esto sugiere que la mutación de las secuencias de TetO afectaba adversamente a la unión de TetR, como se observa por una mayor expresión de LacZ basal en todos los clones en ausencia de Tet y un factor de inducción más bajo después de la adición de +Tet.

En general, todas las construcciones pJV56, pJV57, pJV59 y pJV60 en células CHO/TetR, demostraron un aumento de la expresión de LacZ en comparación con células CHO que se expresaban a partir de pcDNA5/TO/lacZ (es decir, controles positivos) en presencia de tetraciclina, y las células CHO que llevaban el inventivo promotor híbrido con secuencias de hCMV-P/TetO, tenían expresión de beta-galactosidasa regulada positivamente por tetraciclina (diferencias de 1,56 a 9,30 veces), en comparación con (-Tet), como se muestra en las Figuras 16, 17 y 18. Cabe señalar que el método de transfección es diferente entre las células CHO T-REx-(adherentes) y las células CHO-K1/TetR (suspensión celular en medio acuoso líquido), y que la eficiencia de la transfección puede explicar las

diferencias observadas en los niveles generales de actividad beta-galactosidasa específica.

LISTADO DE SECUENCIAS

5 <110> JUST BIOTHERAPEUTICS, INC.
<120> PROMOTOR HÍBRIDO Y USOS DEL MISMO
<130> JUST0051
10 <140>
<141>
<150> 62/388.391
15 <151> 27/01/2016
<160> 35
<170> PatentIn versión 3.5
20 <210> 1
<211> 1717
<212> ADN
<213> Secuencia artificial
25 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"
30 <400> 1

ES 2 823 173 T3

gt caacagga aagt t ccat t ggagccaagt acat t gaggc aat agggact t t ccaat ggg 60
 t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca 120
 cgt at act ga gt cat t aggg act t t ccaat ggg t t t t gcc cagt acat aa ggt caat agg 180
 ggt gaat caa caggaaagt c ccat t ggagc caagt aact gagt caat ag ggact t t cca 240
 t t ggg t t t g cccagt acaa aaggt caat a ggggg t gagg caat ggg t t t t cccat t at 300
 t ggcacgt ac at aaggt caa t aggggt gag t cat t ggg t t t t ccagcca at t t aat t aa 360
 aacgccat gt act t t cccac cat t gacgt c aat gggct at t gaaact aat gcaacgt gac 420
 ct t t aaacgg t act t t ccca t agct gat t a at gggaaagt accgt t ct cg agccaat aca 480
 cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccggt t t t cccct ggaaa 540
 t t ccat at t g gcacgcat t c t at t ggct ga gct gcgt t ct acgt ggg t at aagaggcgcg 600
 accagcgt cg gt accgt acc t ct t ccgcat cgct gt ct gc gagggccagc t gt t ggggt g 660
 agt ggcgggt gt ggct t ccg cgggccccgg agct ggagcc ct gct ct gag cgggccccggc 720
 t gat at gcga gt gt cgt ccg cagggt t t ag ct gt gagcat t cccact t cg agt ggcgggc 780
 ggt gcggggg t gagagt gcg aggcct agcg gcaaccccg t agcct cgcct cgt gt ccggc 840
 t t gaggcct a gcgt ggt gt c cgccgcccgg t gccact ccg gccgact at gcgt t t t t g 900
 t cct t gct gc cct cgat t gc ct t ccagcag cat gggct aa caaaggagg gt gt ggggt 960
 cact ct t aag gagcccat ga agct t acgt t ggat agaat ggaagggcag gaggggcgac 1020
 t ggggccccgc ccgct t cgg agcacat gt c cgacgccacc t ggat ggggc gaggcct gt g 1080
 gct t t ccgaa gcaat cgggc gt gagg t t ag cct acct ggg ccat gt ggcc ct agcact gg 1140
 gcacggt ct g gcct ggcgg t gccgct t cc ct t gcct ccc aacaagggt g aggccgt ccc 1200

 gcccggcacc agt t gct t gc gcggaaagat ggccgct ccc gggggcct gt t gcaaggagc 1260
 t caaat gga ggacgcggca gcccgg t gga gcgggcgggt gagg caccca cacaaaggaa 1320
 gagggcct t g cccct cgccg gccgct gct t cct gt gacc cgt ggt ct at cggccgcat a 1380
 gt cacct cgg gct t ct ct t g agcaccgct c gt cgcgccgg ggggagggga t ct aat ggcg 1440
 t t ggagt t t g t t cacat t t g gt ggg t ggag act agt cagg ccagcct ggc gct ggaagt c 1500
 at t ct t ggaa t t t gccct t t gagg t t gga gcgaggct aa t t ct caagcc t ct t agcgg t 1560
 t caaagg t at t t t ct aaacc cgt t t ccagc t cgcggt t ga ggacaaact c t t cgcggt ct 1620
 t t ccagt act ct t ggat cgg aaacccgt cg gcct ccgaac ggt act ccgc caccgagggga 1680
 cct gagcgag t ccgcat cga ccggat cggaa aaacct c 1717

- 5 <210> 2
- <211> 616
- <212> ADN
- <213> Secuencia artificial

ES 2 823 173 T3

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

5 <400> 2

gt caacagga aagt t ccat t ggagccaagt acat t gaggc aat agggact t t ccaat ggg	60
t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca	120
cgt at act ga gt cat t aggg act t t ccaat ggg t t t gcc cagt acat aa ggt caat agg	180
ggg gaat caa caggaaagt c ccat t ggagc caagt aact gagg caat ag ggact t t cca	240
t t ggg t t t g cccagt acaa aaggt caat a gggggg gagg caat ggg t t t t cccat t at	300
t ggcacgt ac at aaggt caa t aggggt gag t cat t ggg t t t t ccagcca at t t aat t aa	360
aacgccat gt act t t cccac cat t gaggc aat ggggt at t gaaact aat gcaacgt gac	420
ct t t aaacgg t act t t ccca t agct gat t a at gggaaagt accgt t ct cg agccaat aca	480
cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccggt t t t cccct ggaaa	540
t t ccat at t g gcacgcat t c t at t ggct ga gct gcgt t ct acgt ggg at aagaggcgcg	600
accagcgt cg gt accg	616

10 <210> 3
 <211> 41
 <212> ADN
 <213> Secuencia artificial

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

20 <400> 3
 tacctctccgcatcgctgtctgagggccagctgtggg 41

25 <210> 4
 <211> 932
 <212> ADN
 <213> *Rattus norvegicus*

<400> 4

ES 2 823 173 T3

gt gagt ggcg ggt gt ggct t ccgcgggccc cggagct gga gccct gct ct gagcgggccc 60
ggct gat at g cgagt gt cgt ccgcagggt t t agct gt gag cat t cccact t cgagt ggcg 120
ggcggg ggcg gggg gagagt gcgaggcct a gcggcaaccc cgt agcct cg cct cgt gt cc 180
ggct t gaggc ct agcgt ggt gt ccgccgcc gcgt gccact ccggccgcac t at gcgt t t t 240
t t gt cct t gc t gccct cgat t gcct t ccag cagcat gggc t aacaaaggg aggggt gt ggg 300
gct cact ct t aaggagccca t gaagct t ac gt t ggat agg aat ggaaggg caggaggggc 360
gact ggggcc cgcccgcct t cggagcacat gt ccgacgcc acct ggat gg ggcgaggcct 420
gt ggct t t cc gaagcaat cg ggcgt gagt t t agcct acct gggccat gt g gccct agcac 480
t gggcacggt ct ggcct ggc ggt gccgcgt t ccct t gcct cccaacaagg gt gaggccgt 540
ccgcccggc accagt t gct t gcgcggaaa gat ggccgct ccgggggccc t gt t gcaagg 600
agct caaaat ggaggacgcg gcagcccggg ggagcgggcg ggt gagt cac ccacacaaag 660
gaagagggcc t t gccct cg ccggccgct g ct t cct gt ga ccccgt ggt c t at cggccgc 720
at agt cacct cgggct t ct c t t gagcaccg ct cgt cgcgg cggggggagg ggat ct aat g 780
gcgt t ggagt t t gt t cacat t t ggt gggg g gagact agt c aggccagcct ggcgct ggaa 840
gt cat t ct t g gaat t t gcc ct t t gagt t t ggagcgagggc t aat t ct caa gcct ct t agc 900
ggt t caaagg t at t t t ct aa acccgt t t cc ag 932

5 <210> 5
<211> 128
<212> ADN
<213> Secuencia artificial

10 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

<400> 5

ct cgcggt t g aggacaaact ct t cgcggt c t t t ccagt ac t ct t ggat cg gaaaccgct c 60
ggcct ccgaa cggt act ccg ccaccgaggg acct gagcga gt ccgcat cg accggat cgg 120
aaaacct c 128

15 <210> 6
<211> 36
<212> ADN
<213> Secuencia artificial

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

25 <400> 6
aaaaacacgattgctcgagagtgccaccatcatg 36

<210> 7
<211> 66

ES 2 823 173 T3

<212> ADN
 <213> Secuencia artificial

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

 <400> 7

 at ggacat ga gggg gcccgc t cagct cct g gggct cct gc t gct gt ggct gagaggt gcg 60
 10 cgct gt 66

 <210> 8
 <211> 22
 <212> PRT
 15 <213> Secuencia artificial

 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Péptido sintético"
 20
 <400> 8

 Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
 1 5 10 15

 Leu Arg Gly Ala Arg Cys
 20

 25 <210> 9
 <211> 69
 <212> ADN
 <213> Secuencia artificial

 30 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

 <400> 9
 35
 t at at aagca gagct ct ccc t at cagt gat agagat ct cc ct at cagt ga t agagat cgt 60
 cgacgagct 69

 <210> 10
 <211> 9
 40 <212> ADN
 <213> *Mus* sp.

 <400> 10
 45 cagcgtcgg 9

 <210> 11
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 50
 <220>
 <221> fuente

ES 2 823 173 T3

<223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

<400> 11

ggcgccgccacc

12

5

<210> 12

<211> 3171

<212> ADN

<213> Secuencia artificial

10

<220>

<221> fuente

<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

15

<400> 12

at ggggggt t ct cat cat ca t cat cat cat ggt at ggct a gcat gact gg t ggacagcaa 60
at ggggt cggg at ct gt acga cgat gacgat aagggt acct a aggat cagct t ggagt t gat 120
cccgt cgt t t t acaacgt cg t gact gggaa aacctt ggcg t t acccaact t aat cgcct t 180
gcagcacat c cccct t t cgc cagct ggcgt aat agcgaag aggcccgcac cgat cgcct 240
t cccaacagt t ggcgagcct gaat ggcgaa t ggcgct t t g cct ggt t t cc ggcaccagaa 300
gcggt gccgg aaagct ggct ggagt gcgat ct t cct gagg ccgat act gt cgt cgt cccc 360
t caaact ggc agat gcacgg t t acgat gcg cccat ct aca ccaacgt aac ct at cccat t 420
acggt caat c cgccgt t t gt t cccacggag aat ccgacgg gt t gt t act c gct cacat t t 480
aat gt t gat g aaagct ggct acaggaaggc cagacgcgaa t t at t t t t ga t ggcgt t aac 540
t cggcgt t t c at ct gt ggt g caacgggcgc t ggggt cgggt t acggccagga cagt cgt t t g 600
ccgt ct gaat t t gacct gag cgcatt t t t a cgcgccggag aaaaccgcct cgcggt gat g 660
gt gct gcgt t ggagt gacgg cagt t at ct g gaagat cagg at at gt ggcg gat gagcggc 720
at t t t ccgt g acgt ct cgt t gct gcat aaa ccgact acac aaat cagcga t t t ccat gt t 780
gccact cgct t t aat gat ga t t t cagccgc gct gt act gg aggct gaagt t cagat gt gc 840
ggcgagt t gc gt gact acct acgggt aaca gt t t ct t t at ggcagggt ga aacgcaggt c 900
gccagcggca ccgcgcct t t cggcgggt gaa at t at cgat g agcgt ggt gg t t at gccgat 960
cgcgt cacac t acgt ct gaa cgt cgaaaac ccgaaact gt ggagcggcga aat cccgaat 1020
ct ct at cgt g cgggt ggt t ga act gcacacc gccgacggca cgct gat t ga agcagaagcc 1080
t gcgat gt cg gt t t ccgcga ggt gcggat t gaaaat ggt c t gct gct gct gaacggcaag 1140
ccgt t gct ga t t cgaggcgt t aaccgt cac gagcat cat c ct ct gcat gg t caggt cat g 1200
gat gagcaga cgat ggt gca ggat at cct g ct gat gaagc agaacaact t t aacgccgt g 1260
cgct gt t cgc at t at ccgaa ccat ccgct g t ggt acacgc t gt gcgaccg ct acggcct g 1320
t at gt ggt gg at gaagccaa t at t gaaacc cacggcat gg t gccaat gaa t cgt ct gacc 1380
gat gat ccgc gct ggct acc ggcgat gaggc gaacgcgt aa cgcgaat ggt gcagcgcgat 1440

ES 2 823 173 T3

cgt aat cacc cgagt gt gat cat ct ggt cg ct ggggaat̄ g aat caggcca cggcgt aat 1500
cacgacgcgc t gt at cgct g gat caaat ct gt cgat cct t cccgcccggg gcagt at gaa 1560
ggcggcgggag ccgacaccac ggccaccgat at t at t t gcc cgat gt acgc gcgcgt ggat 1620
gaagaccagc cct t cccggc t gt gccgaaa t ggt ccat ca aaaaat ggct t t cgct acct 1680
ggagagacgc gcccgct gat cct t t gcgaa t acgcccacg cgat gggg aa cagt ct t ggc 1740
ggg t t cgct a aat act ggca ggcgt t t cgt cagt at cccc gt t t acaggg cggct t cgt c 1800
t gggact ggg t ggat cagt c gct gat t aaa t at gat gaaa acggcaaccc gt ggt cggct 1860
t acggcggg g at t t t ggcga t acgcccgaac gat cgccagt t ct gt at gaa cggg ct ggt c 1920
t t t gccgacc gcacgccgca t ccagcgt g acggaagcaa aacaccagca gcagt t t t t c 1980
cagt t ccgt t t at ccgggca aacct cgaa gt gaccagcg aat acct gt t ccgt cat agc 2040
gat aacgagc t cct gact g gat ggt ggcg ct ggat ggt a agccgct ggc aagcggg gaa 2100
gt gcct ct gg at gt cgct cc acaaggt aaa cagt t gat t g aact gcct ga act accgcag 2160
ccggagagcg ccgggcaact ct ggct caca gt acgcgt ag t gcaaccgaa cgcgaccgca 2220
t ggt cagaag ccgggcacat cagcgcct gg cagcagt ggc gt ct ggcgga aaacct cagt 2280
gt gacgct cc ccgccgct c ccacgccat c ccgcat ct ga ccaccagcga aat ggat t t t 2340
t gcat cgagc t gggg aat aa gcgt t ggcaa t t t aaccgcc agt caggct t t ct t t cacag 2400
at gt ggat t g gcgat aaaaa acaact gct g acgcccgt gc gcgat cagt t caccct gca 2460
ccgct ggat a acgacat t gg cgt aagt gaa gcgacccgca t t gaccct aa cgcct gggg c 2520
gaacgct gga aggcggcggg ccat t accag gccgaagcag cgt t gt t gca gt gcacggca 2580
gat aact t g ct gat gcggg gct gat t acg accgct cacg cgt ggcagca t caggggaaa 2640
acct t at t t a t cagccggaa aacct accgg at t gat ggt a gt ggt caaat ggcgat t acc 2700
gt t gat gt t g aagt ggcgag cgat acaccg cat ccggcgc ggat t ggcct gaact gccag 2760
ct ggcgcagg t agcagagcg ggt aaact gg ct cggat t ag ggccgcaaga aaact at ccc 2820
gaccgcct t a ct gccgcct g t t t t gaccgc t gggat ct gc cat t gt caga cat gt at acc 2880
ccgt acgt ct t cccgagcga aaacggg ct g cgct gcggga cgcgcgaat t gaat t at ggc 2940
ccacaccagt ggcgcggcga ct t ccagt t c aacat cagcc gct acagt ca acagcaact g 3000
at ggaacca gccat cgcca t ct gct gcac gcggaagaag gcacat ggct gaat at cgac 3060
ggg t t ccat a t ggggat t gg t ggcgacgac t cct ggagcc cgt cagt at c ggcggagt t c 3120
cagct gagcg ccggg cgct a ccat t accag t t ggt ct ggt gt caaaaat a a 3171

<210> 13

<211> 14

5 <212> ADN

<213> Secuencia artificial

<220>

<221> fuente

10 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

ES 2 823 173 T3

<400> 13
gcggccgcgctagc 14

5 <210> 14
<211> 222
<212> ADN
<213> Secuencia artificial

10 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

15 <400> 14
cacacat cat aagat acat t gat gagg t t g gacaaaccac aact agaat g cagt gaaaaa 60
aat gct t t at t t gt gaaat t t gt gat gct a t t gct t t at t t gt aaccat t at aagct gca 120
at aaacaagt t aacaacaac aat t gcat t c at t t t at gt t t cagg t t cag ggggagat gt 180
gggaggt t t t t t aaagcaag t aaaacct ct acaaat gt gg t a 222

20 <210> 15
<211> 102
<212> ADN
<213> Secuencia artificial

25 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

<400> 15
gt t ct acgt g ggt at at aag cagagct ct c cct at cagt g at agagat ct ccct at cagt 60
gat agagat c gt cgacgagc t cagcgt cgg t accgt acct ct 102

30 <210> 16
<211> 48
<212> ADN
<213> Secuencia artificial

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

40 <400> 16
gttctacgtgggtataagaggcgaccagcgtcggtagcctct 48

45 <210> 17
<211> 56
<212> ADN
<213> Secuencia artificial

50 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético "

<400> 17
tatataagcagagctctccctatcagtgatagagatctccctatcagtgatagaga 56

ES 2 823 173 T3

<210> 18
 <211> 56
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"
 10 <400> 18
 tataaagcagagctctccctatcagtgatcagttcctccctatcagtgatagaga 56
 <210> 19
 <211> 36
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 <221> fuente
 20 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"
 <400> 19
 tataagaggcgcgaccagcgtcggtagcctct 36
 25 <210> 20
 <211> 59
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"
 35 <400> 20
 tataaagcagagctcgtttagtgaaccgctcagttcgtctctagacccaaccgcctct 59
 <210> 21
 <211> 29
 <212> ADN
 40 <213> Secuencia artificial
 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"
 45 <400> 21
 tataagaggcgcgaccagcgtcggtagc 29
 <210> 22
 <211> 27
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"
 55 <400> 22
 tataaagcagagcagcgtcggtagc 27
 60 <210> 23
 <211> 56
 <212> ADN
 <213> Secuencia artificial
 65

ES 2 823 173 T3

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

5 <400> 23
 ctctccctatcagtgatagagatctccctatcagtgatagagatcgtagcagagct 56

<210> 24
 <211> 15
 10 <212> ADN
 <213> Secuencia artificial

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

<400> 24
 tatataagcagagct 15

20 <210> 25
 <211> 14
 <212> ADN
 <213> Secuencia artificial

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

30 <400> 25
 cagcgtcgtaccg 14

<210> 26
 <211> 15
 <212> ADN
 35 <213> Secuencia artificial

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

40 <400> 26
 tataagaggcgcgac 15

45 <210> 27
 <211> 12
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

55 <400> 27
 gttctacgtggg 12

<210> 28
 <211> 29
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <221> fuente
 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

65 <400> 28

ES 2 823 173 T3

tataagaggcgcgaccagcgtcgtaccg 29

<210> 29
<211> 40
5 <212> ADN
<213> Secuencia artificial

<220>
<221> fuente
10 <223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

<400> 29
tccctatcagtgatagagatctccctatcagtgatagaga 40

15 <210> 30
<211> 1771
<212> ADN
<213> Secuencia artificial

20 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

25 <400> 30

ES 2 823 173 T3

gt caacagga aagt t ccat t ggagccaagt acat t gaggc aat agggact t t ccaat ggg 60
t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca 120
cgt at act ga gt cat t aggg act t t ccaat ggg t t t t gcc cagt acat aa ggt caat agg 180
ggg gaat caa caggaaagt c ccat t ggagc caagt aact gagg caat ag ggact t t cca 240
t t ggg t t t g cccagt acaa aaggt caat a gggggg gagg caat ggg t t t t cccat t at 300
t ggcacgt ac at aaggt caa t aggggg gagg t cat t ggg t t t t ccagcca at t t aat t aa 360
aacgccat gt act t t cccac cat t gacgt c aat gggct at t gaaact aat gcaacgt gac 420
ct t t aaacgg t act t t ccca t agct gat t a at gggaaagt accgt t ct cg agccaat aca 480
cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccggt t t t cccct ggaaa 540
t t ccat at t g gcacgcat t c t at t ggct ga gct gcgt t ct acgt ggg t at at aagcagag 600
ct ct cct at cagt gat aga gat ct cct a t cagt gat ag agat cgt cga cgagct cagc 660
gt cgg t accg t acct ct t cc gcat cgct gt ct gcgagggc cagct gt t gg ggt gagg ggc 720

gggt gt ggct t ccgcgggcc ccggagct gg agccct gct c t gaggggcc gggct gat at 780
gcgagt gt cg t ccgcaggg t t agct gt ga gcat t cccac t t cgagt ggc gggcggt gcg 840
ggggg gagag t gcgaggcct agcggcaacc ccgt agcct c gcct cgt gt c cggct t gagg 900
cct agcgt gg t gt ccgccc cgcgt gccac t ccggccgca ct at gcgt t t t t t gt cct t g 960
ct gccct cga t t gcct t cca gcagcat ggg ct aacaaagg gagggt gt gg ggct cact ct 1020
t aaggagccc at gaagct t a cgt t ggat ag gaat ggaagg gcaggagggg cgact ggggc 1080
ccgcccgcct t cggagcaca t gt ccgacgc cacct ggat g gggcgaggcc t gt ggct t t c 1140
cgaagcaat c gggcgt gagg t t agcct acc t gggccat gt ggccct agca ct gggcacgg 1200
t ct ggct gg cggg gccggt t t cct t gcc t cccaacaag ggt gaggccg t cccgcccgg 1260
caccagt t gc t t gcgaggaa agat ggccgc t cccggggcc ct gt t gcaag gagct caaaa 1320
t ggaggacgc ggcagcccgg t ggagcgggc gggg gagg ca cccacacaaa ggaagagggc 1380
ct t gccct c gccggccgt gct t cct gt g acccct ggt ct at cggccg cat agt cacc 1440
t cgggct t ct ct t gaggacc gct cgt cgcg gcgggggggag gggat ct aat ggcgt t ggag 1500
t t t gt t caca t t t ggt gggg ggagact agt caggccagcc t ggcgct gga agt cat t ct t 1560
ggaat t t gcc cct t t gagg t t ggagcggg ct aat t ct ca agcct ct t ag cgg t t caaag 1620
gt at t t t ct a aaccct t t c cagct cgcg t t gaggacaa act ct t cgcg gt ct t t ccag 1680
t act ct t gga t cggaaaccc gt cggcct cc gaacggg act ccgccaccga gggacct gag 1740
cgagt ccgca t cgaccggat cggaaaacct c 1771

<210> 31
<211> 1740
<212> ADN
<213> Secuencia artificial

ES 2 823 173 T3

<220>

<221> fuente

<223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"

5

<400> 31

gt caacagga aagt t ccat t ggagccaagt acat t gaggc aat agggact t t ccaat ggg	60
t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca	120
cgt at act ga gt cat t aggg act t t ccaat ggg t t t gcc cagt acat aa ggt caat agg	180
ggg gaat caa caggaaagt c ccat t ggagc caagt aact gagg caat ag ggact t t cca	240
t t ggg t t t g cccagt acaa aaggt caat a gggggg gagg caat ggg t t t t cccat t at	300
t ggcaggt ac at aaggt caa t aggggg gagg t cat t ggg t t t t ccagcca at t t aat t aa	360
aagccat gt act t t cccac cat t gaggc aat gggct at t gaaact aat gcaacgt gac	420
ct t t aaggg t act t t ccca t aggt gat t a at gggaaagt accgt t ct cg agccaat aca	480
cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccggt t t t cccct ggaaa	540
t t ccat at t g gcagcat t c t at t ggct ga gct gcgt t ct acgt ggg t at at aagcagag	600
ct cgt t t agt gaaccgt cag t t cgt ct ct a gacgccaacc gcct ct t ccg cat cgct gt c	660
t gcgagggcc agct gt t ggg gt gagg ggcg ggt gt ggct t ccgccccccc cggagct gga	720
gccct gct ct gaggggccg ggct gat at g cagggt gt cgt ccgaggggt t t agct gt gag	780
cat t cccact t cagggt ggcg ggcgggt ggcg gggg gagggt gcgagggct a gcggcaacc	840
cgt agcct cg cct cgt gt cc ggct t gaggc ct agcgt ggt gt ccgcccgc gcgt gccact	900
ccggccgcac t at gcgt t t t t t gt cct t gc t gccct cgat t gcct t ccag cagcat gggc	960
t acaaaagg aggggt gt ggg gct cact ct t aaggagccca t gaagct t ac gt t ggat agg	1020
aat ggaagg caggaggggc gact ggggcc cggccgct t cggagcacaat gt ccgacgcc	1080
acct ggat gg ggcgaggcct gt ggct t t cc gaagcaat cg ggcgt gagg t t agcct acct	1140
gggccaat gt g gccct agcac t gggcagggt ct ggcct ggc ggt gccgcgt t cct t gcct	1200
ccaacaagg gt gaggcct cccgcccggc accagt t gct t gcgcccgaat gat ggccgct	1260
cccggggccc t gt t gcaagg agct caaaat ggaggacgcg gcagcccggg ggagcggggc	1320
ggg gagg cac ccacacaaag gaagagggcc t t gccct cg ccggccgct g ct t cct gt ga	1380
ccccgt ggt c t at cggccc at agt cacct cgggct t ct c t t gaggaccg ct cgt cgcgg	1440
cggggggagg ggat ct aat g gcgt t gagg t t gt t cacat t t ggt ggg t g gaggact agt c	1500
aggccagcct ggcgt ggaa gt cat t ct t g gaat t t gcc ct t t gagg t t ggagcggaggc	1560
t aat t ct caa gccct ct t agc ggt t caaagg t at t t t ct aa acccgt t t cc agct cgcgg	1620
t gaggacaaa ct ct t cgcgg t ct t t ccagt act ct t ggat cggaaaccgc t cggcct ccg	1680
aagcgt act c cggccagg ggacct gagg gagg ccgcac cgaccggat c ggaaaacct c	1740

ES 2 823 173 T3

<210> 32
 <211> 1771
 <212> ADN
 <213> Secuencia artificial

5

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

10

<400> 32

gt caacagga aagt t ccat t ggagccaagt acat t gagt c aat agggact t t ccaat ggg	60
t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca	120
cgt at act ga gt cat t aggg act t t ccaat ggg t t t gcc cagt acat aa ggt caat agg	180
ggg gaat caa caggaaagt c ccat t ggagc caagt aact gagt caat ag ggact t t cca	240
t t ggg t t t g cccagt acaa aaggt caat a gggggg gagt caat ggg t t t t cccat t at	300
t ggcacgt ac at aaggt caa t aggggt gag t cat t ggg t t t t ccagcca at t t aat t aa	360
aacgccat gt act t t cccac cat t gacgt c aat gggct at t gaaact aat gcaacgt gac	420
ct t t aaacgg t act t t ccca t agct gat t a at gggaaagt accgt t ct cg agccaat aca	480
cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccgg t t t cccct ggaaa	540

ES 2 823 173 T3

t t ccat at t g gcacgcat t c t at t ggct ga gct gcgt t ct acgt gggat at at aagcagag 600
ct ct ccct at cagt gat cag t t cct ccct a t cagt gat ag agat cgt cga cgagct cagc 660
gt cggat accg t acct ct t cc gcat cgct gt ct gcgagggc cagct gt t gg ggt gagt ggc 720
gggt gt ggct t ccgcggggc ccggagct gg agccct gct c t gagcggggc gggct gat at 780
gcgagt gt cg t ccgcagggt t t agct gt ga gcat t cccac t t cgagt ggc gggcggg gcg 840
ggggt gagag t gcgaggcct agcggcaacc ccgt agcct c gcct cgt gt c cggct t gagg 900
cct agcgt gg t gt ccgccgc cgcgt gccac t ccggccgca ct at gcgt t t t t t gt cct t g 960
ct gccct cga t t gcct t cca gcagcat ggg ct aacaaagg gaggggt gt gg ggct cact ct 1020
t aaggagccc at gaagct t a cgt t ggat ag gaat ggaagg gcaggagggg cgact ggggc 1080
ccgcccgcct t cggagcaca t gt ccgacgc cacct ggat g gggcgaggcc t gt ggct t t c 1140
cgaagcaat c gggcgt gagt t t agcct acc t gggccat gt ggccct agca ct gggcacgg 1200
t ct ggcct gg cggg gccgcg t t ccct t gcc t cccaacaag ggt gaggccg t cccgcccgg 1260
caccagt t gc t t gcgcggaa agat ggccgc t cccggggcc ct gt t gcaag gagct caaaa 1320
t ggaggacgc ggcagcccgg t ggagcgggc ggggt gagt ca cccacacaaa ggaagagggc 1380
ct t gccct c gccggccgct gct t cct gt g accccgt ggt ct at cggccg cat agt cacc 1440
t cgggct t ct ct t gagcacc gct cgt cgcg gcggggggag gggat ct aat ggcgt t ggag 1500
t t t gt t caca t t t ggt gggg ggagact agt caggccagcc t ggcgct gga agt cat t ct t 1560
ggaat t t gcc cct t t gagt t t ggagcgagg ct aat t ct ca agcct ct t ag cggg t caaag 1620
gt at t t t ct a aaccct t t c cagct cgcg t t gaggacaa act ct t cgcg gt ct t t ccag 1680
t act ct t gga t cggaaaccc gt cggcct cc gaacggg act ccgccaccga gggacct gag 1740
cgagt ccgca t cgaccggat cggaaaacct c 1771

- <210> 33
- <211> 641
- 5 <212> ADN
- <213> Secuencia artificial
- <220>
- <221> fuente
- 10 <223> / observación= "Descripción de la secuencia artificial: Polinucleótido sintético"
- <400> 33

ES 2 823 173 T3

gt caacagga aagt t ccat t ggagccaagt acat t gaggc aat agggact t t ccaat ggg 60
t t t t gccag t acat aaggt caat gggagg t aagccaat g ggt t t t t ccc at t act ggca 120
cgt at act ga gt cat t aggg act t t ccaat gggg t t t gcc cagt acat aa ggt caat agg 180
ggg gaat caa caggaaagt c ccat t ggagc caagt aact gagg caat ag ggact t t cca 240
t t gggg t t t g cccagt acaa aaggt caat a gggggg gagg caat gggg t t t t cccat t at 300
t ggcaggt ac at aaggt caa t aggggg gagg t cat t gggg t t t t ccagcca at t t aat t aa 360
aacgcat gt act t t cccac cat t gaggc aat ggggt at t gaaact aat gcaacgt gac 420

ct t t aaacgg t act t t ccca t aggt gat t a at gggaaagt accgt t ct cg agccaat aca 480
cgt caat ggg aagt gaaagg gcagccaaaa cgt aacaccg ccccggg t t t cccct ggaaa 540
t t ccat at t g gcacgcat t c t at t gggt ga gct gcgt t ct acgt gggg at at aagcagag 600
ct cgt t t agt gaaccgt cag t t cgt ct ct a gacgccaacc g 641

5 <210> 34
<211> 41
<212> ADN
<213> Secuencia artificial

10 <220>
<221> fuente
<223> / observación= "Descripción de la secuencia artificial: Oligonucleótido sintético"

<400> 34
ctccctatcagtgatcagttcctccctatcagtgatagaga 41

15 <210> 35
<211> 54
<212> ADN
<213> Secuencia artificial

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

25 <400> 35
tatataagcagagctcgttagtgaaccgtcagttcgtctctagacgccaaccg 54

REIVINDICACIONES

1. Un promotor híbrido, que comprende:
 - 5 (i) una secuencia potenciadora de citomegalovirus murino (mCMV), que comprende un elemento potenciador de mCMV (mCMV-E) y una secuencia promotora de CMV (CMV-P) en su extremo 3', unida operativamente en 5' a una secuencia intrónica del factor 1 de elongación alfa (EF 1 alfa) de rata;
 - (ii) una primera secuencia líder intermedia unida operativamente, en 3' a la secuencia promotora de CMV de la secuencia potenciadora de mCMV, y en 5' a la secuencia intrónica del EF 1 alfa de rata; y
 - 10 (iii) una segunda secuencia líder unida operativamente en 3' a la secuencia intrónica del EF 1 alfa de rata.
2. El promotor híbrido de la reivindicación 1, en donde la secuencia promotora de CMV en el extremo 3' de la secuencia potenciadora de mCMV comprende un segmento que tiene la secuencia de nucleótidos de SEQ ID NO: 24 o SEQ ID NO: 26.
- 15 3. El promotor híbrido de cualquiera de las reivindicaciones 1-2, en donde la secuencia intrónica del EF 1 alfa de rata comprende la secuencia de nucleótidos de SEQ ID NO: 4.
4. El promotor híbrido de cualquiera de las reivindicaciones 1-3, en donde la secuencia potenciadora de mCMV comprende la secuencia de nucleótidos de SEQ ID NO: 2 o SEQ ID NO: 33; y la secuencia intrónica del EF 1 alfa de rata comprende la secuencia de nucleótidos de SEQ ID NO: 4.
- 20 5. El promotor híbrido de cualquiera de las reivindicaciones 1-4, en donde la primera secuencia líder comprende la secuencia de nucleótidos de SEQ ID NO: 3.
- 25 6. El promotor híbrido de cualquiera de las reivindicaciones 1-5, en donde la segunda secuencia líder comprende la secuencia de nucleótidos de SEQ ID NO: 5.
7. El promotor híbrido de cualquiera de las reivindicaciones 1-6, que comprende además una o más secuencias de TetO insertadas dentro de la secuencia promotora de CMV.
- 30 8. El promotor híbrido de cualquiera de las reivindicaciones 1-7, que comprende una secuencia de nucleótidos que tiene una identidad de secuencia del al menos 95 % con la secuencia de nucleótidos de SEQ ID NO:1, SEQ ID NO:30, SEQ ID NO: 31 o SEQ ID NO: 32.
- 35 9. Un casete de expresión, que comprende:
 - (a) el promotor híbrido de cualquiera de las reivindicaciones 1-8, unido operativamente en 5' a un marco abierto de lectura que codifica una proteína exógena de interés; y
 - 40 (b) un sitio de poliadenilación unido operativamente en 3' al marco abierto de lectura.
10. Un vector de expresión recombinante que comprende el casete de expresión de la reivindicación 9.
11. Una célula hospedadora de mamífero que comprende el vector de expresión recombinante de la reivindicación 10.
- 45 12. La célula hospedadora de mamífero de la reivindicación 11, que además puede expresar TetR.
13. La célula hospedadora de mamífero de la reivindicación 11 o la reivindicación 12, que procede de una célula de ovario de hámster chino (CHO).
- 50 14. Un método para producir una proteína de interés, que comprende:
 - (a) cultivar la célula hospedadora de mamífero de cualquiera de las reivindicaciones 11-13, en un medio acuoso en condiciones fisiológicas que permitan la expresión de la proteína de interés; y
 - 55 (b) recuperar la proteína de interés del medio.
15. Un método para producir una proteína de interés, que comprende:
 - (a) cultivar una célula hospedadora de mamífero que comprende un vector de expresión que comprende el promotor híbrido de cualquiera de las reivindicaciones 7-8, en un medio acuoso en condiciones fisiológicas, en donde la célula hospedadora de mamífero puede expresar TetR, de modo que, en ausencia de tetraciclina en el medio, se reprime la expresión de la proteína de interés;
 - (b) añadir tetraciclina al medio acuoso en una cantidad suficiente para unir TetR en la célula hospedadora, de modo que la célula hospedadora desreprime la expresión de la proteína de interés; y
 - 60 (c) recuperar la proteína de interés del medio.
 - 65

FIG. 1

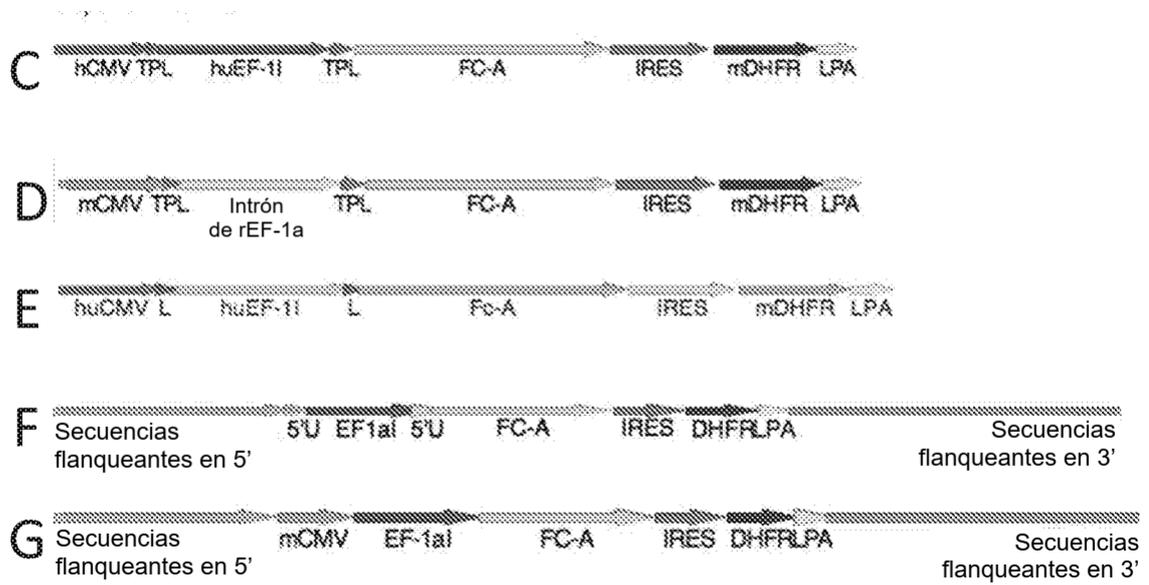


FIG.2

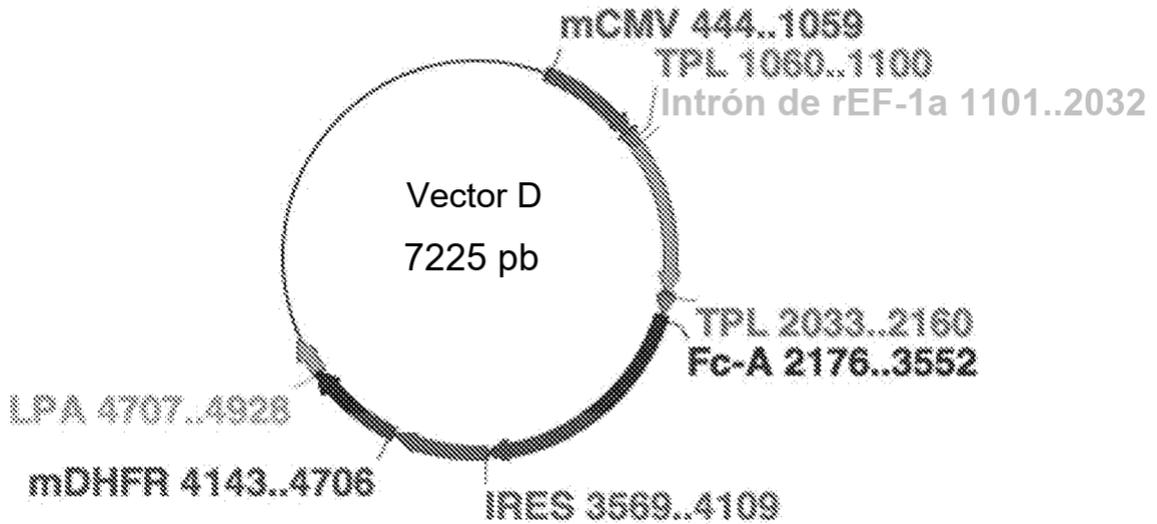


FIG. 3A

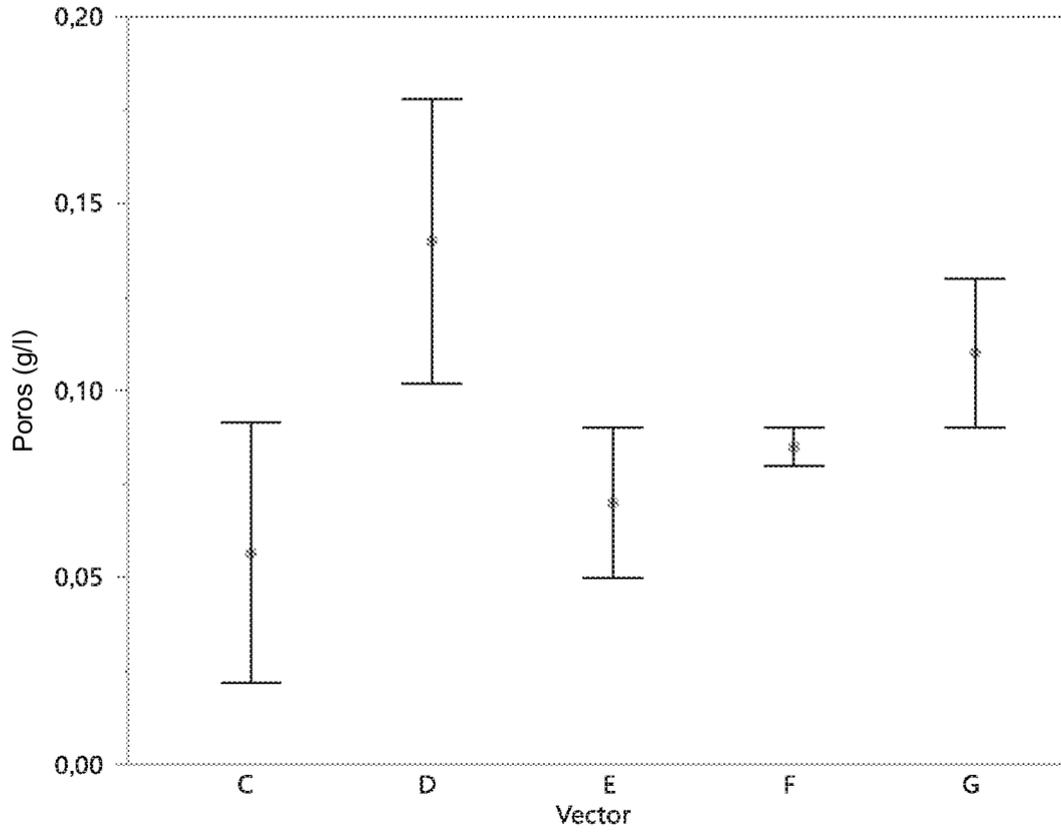


FIG. 3B

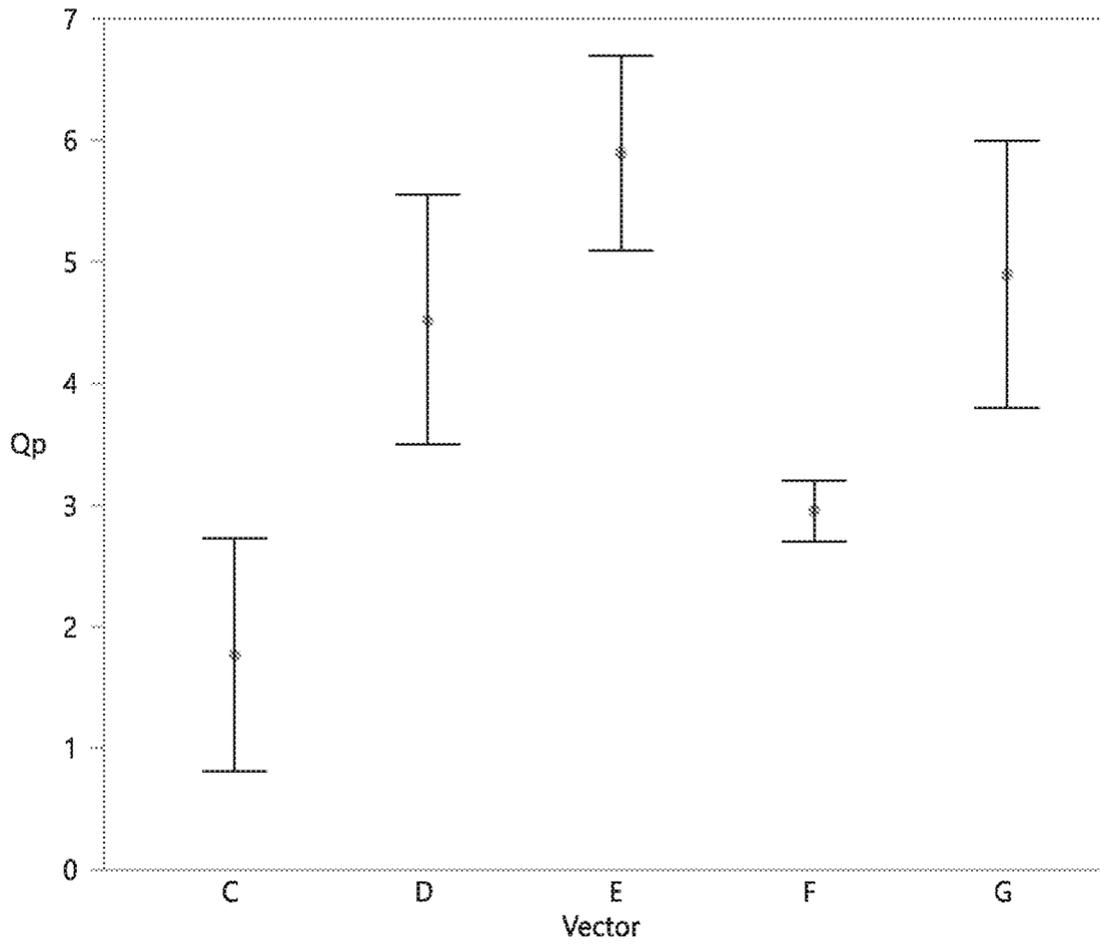


FIG. 4A

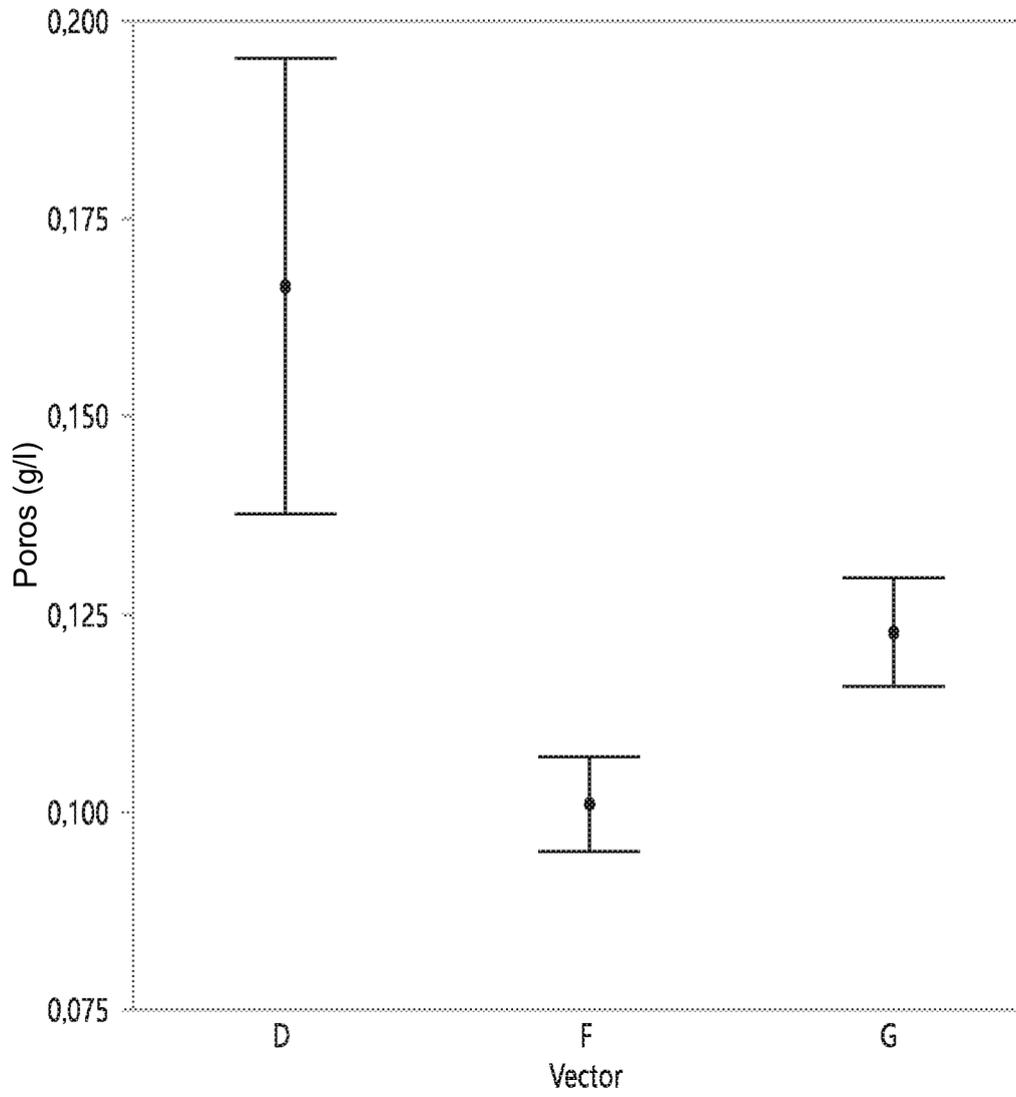


FIG. 4B

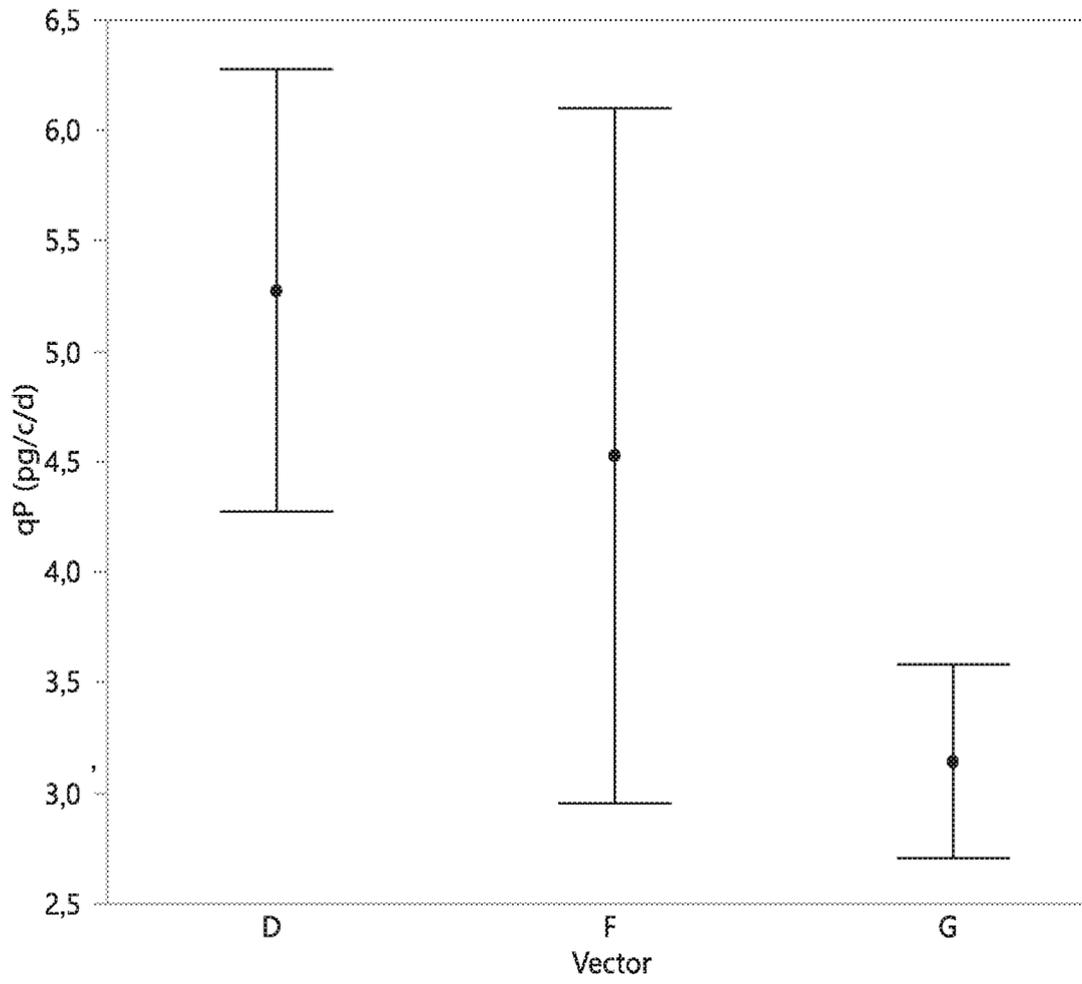


FIG. 5A

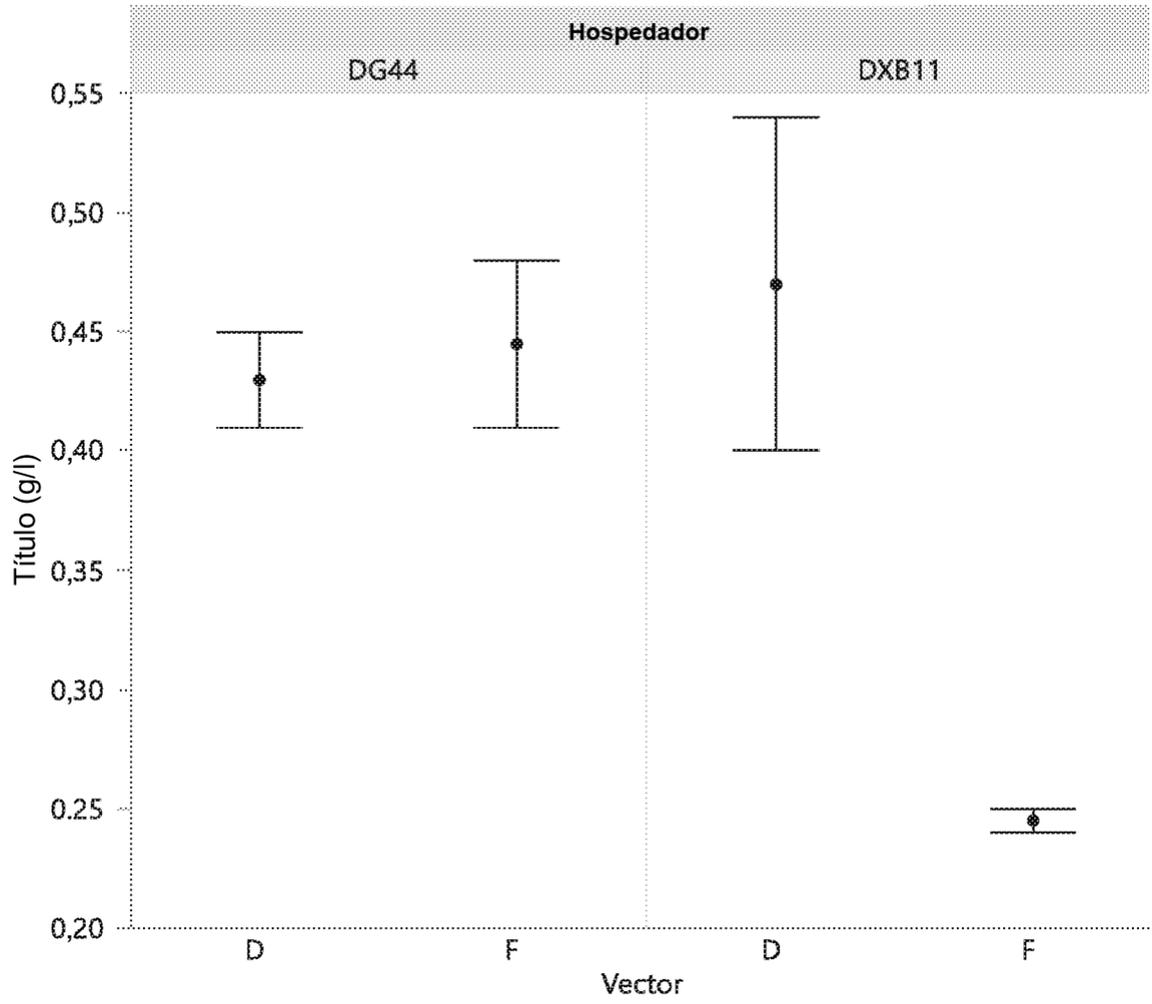


FIG. 5B

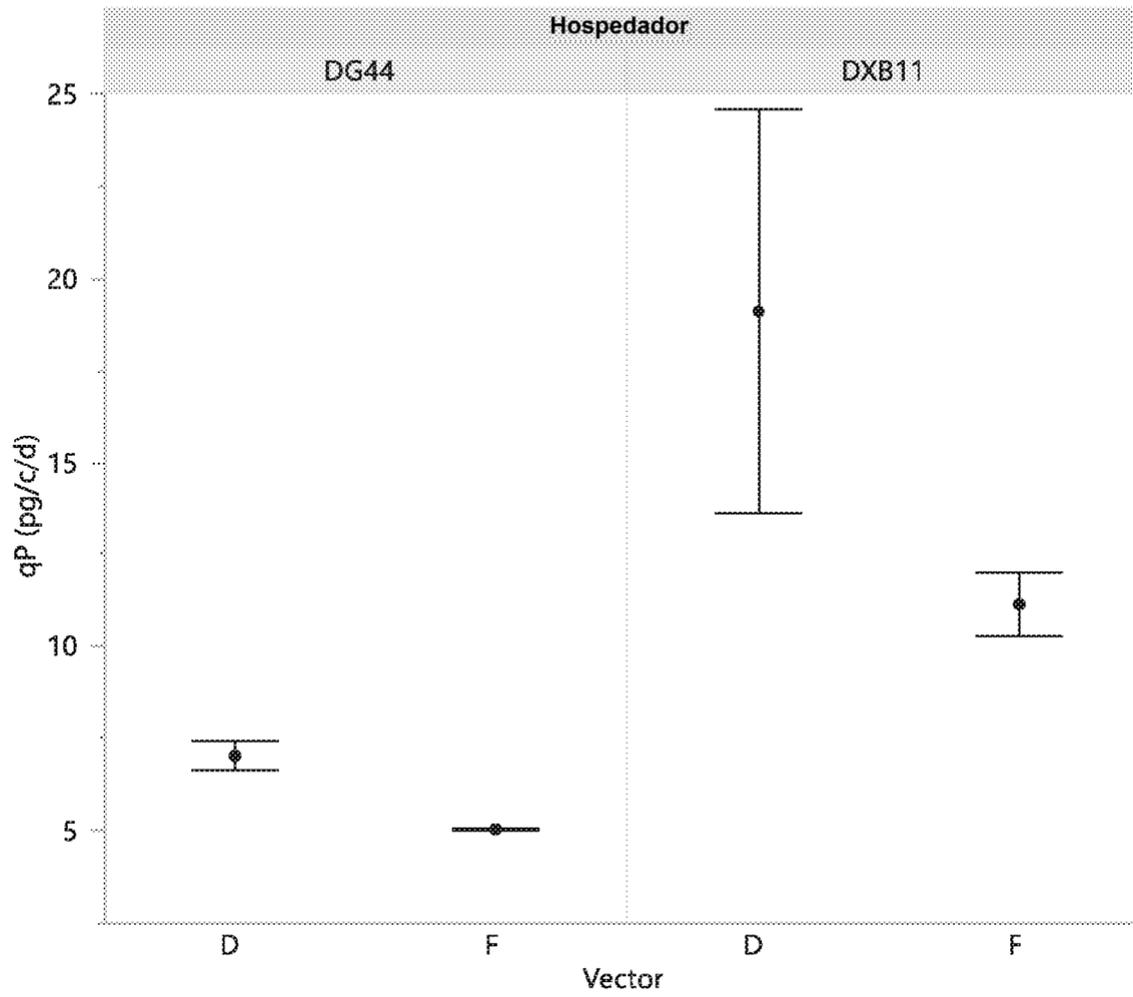


FIG. 6

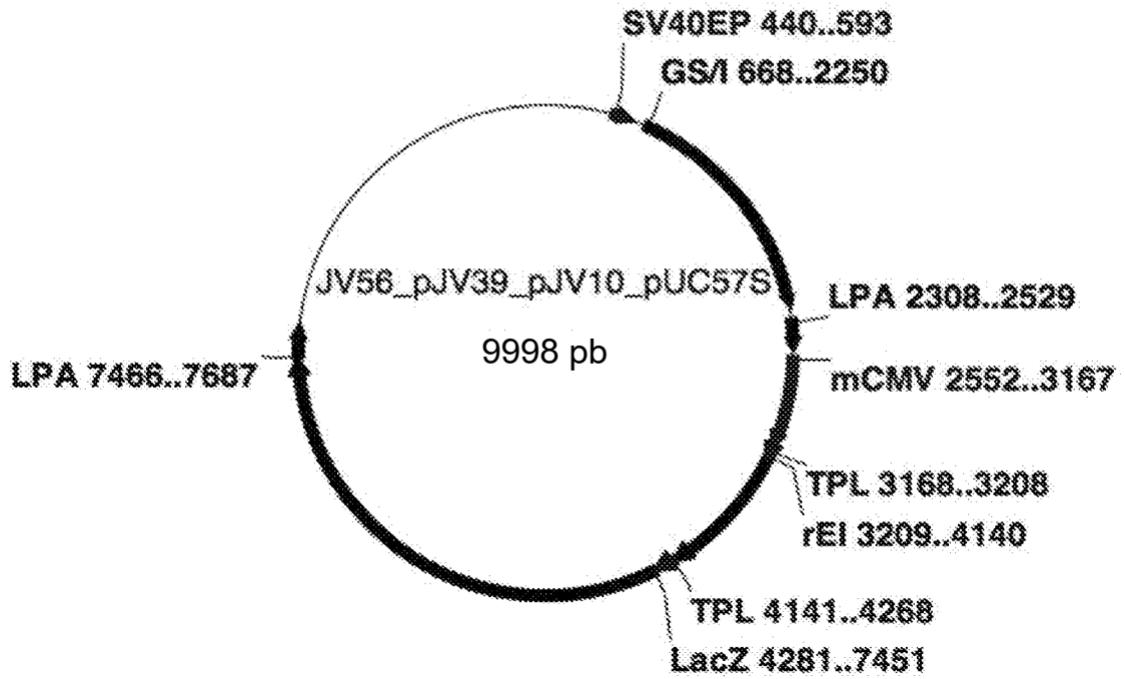


FIG. 7

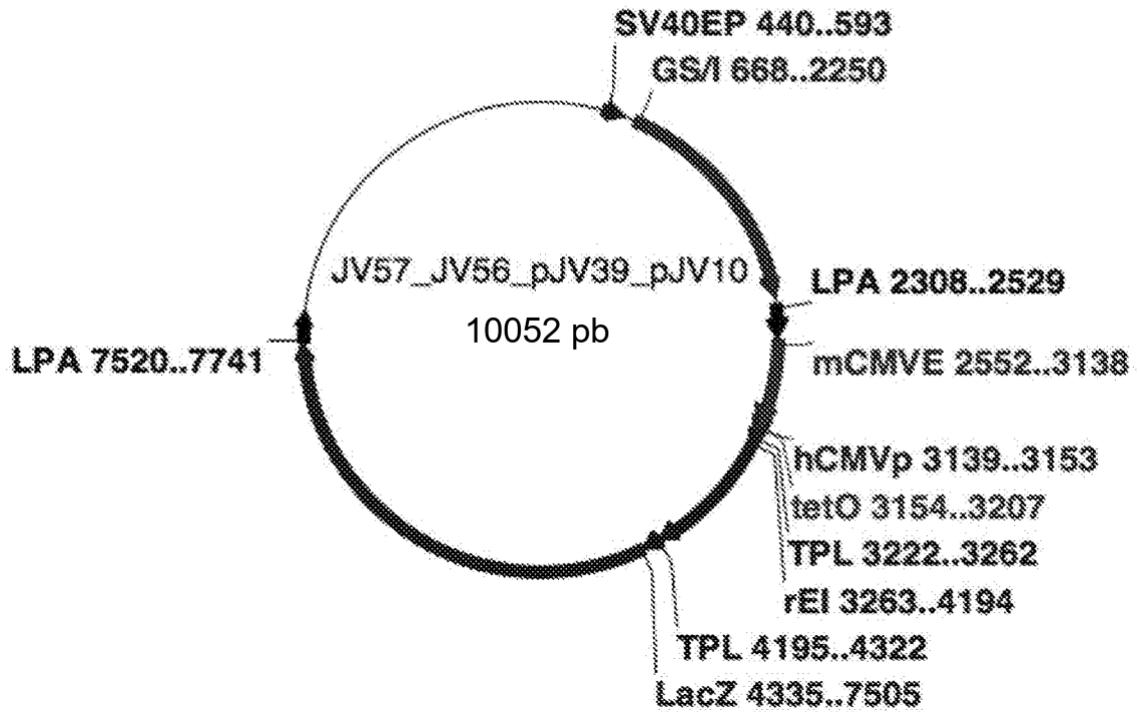


FIG. 8

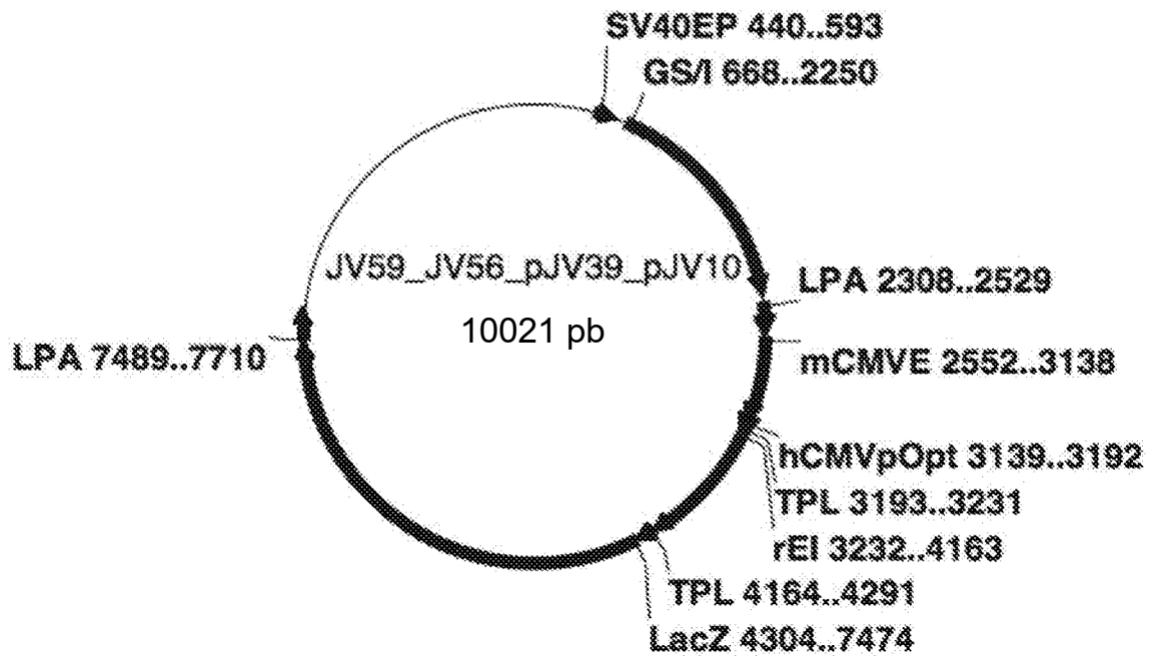


FIG. 9

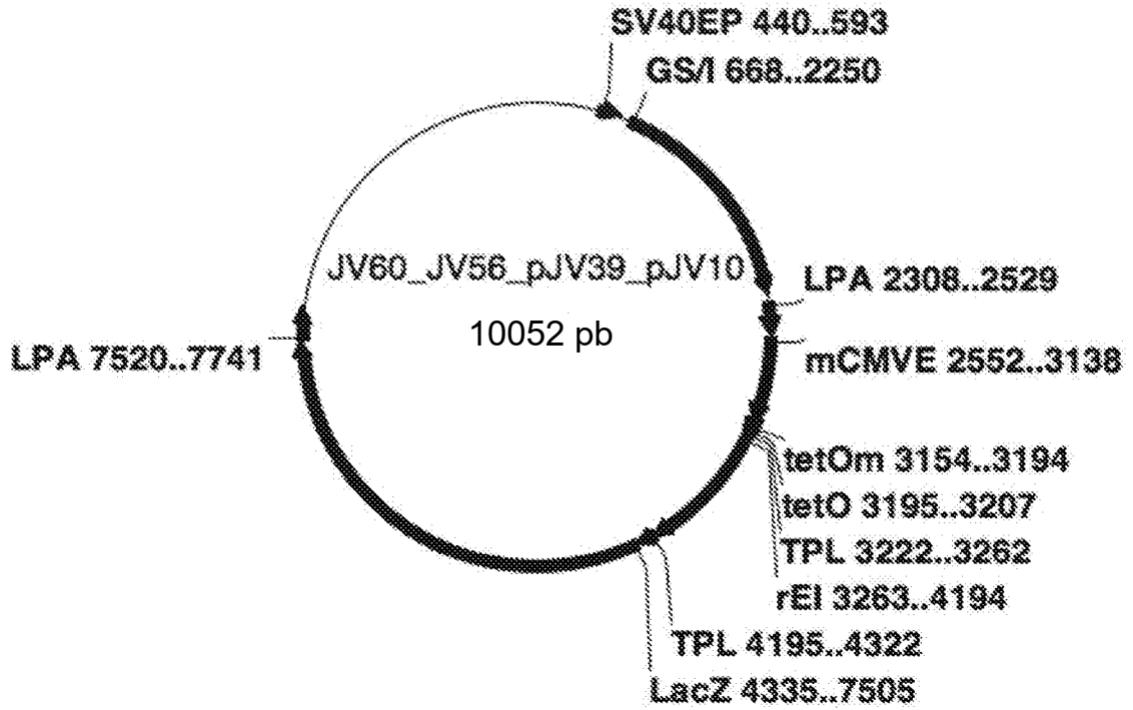


FIG. 11

	mCMV-E	huCMV-P	TetO	mCMV-P	TPL
pJV57	...gttctactgtggg	TATAtaAGcaGaGct	CTCCCTATCAGTGTAGAGAGATCTCCCTATCAGTGTAGAGATCGTTCGACGAGCT	cagcgtcggtaaccg	Taccctct...
			←		
			→		
pJV56	...gttctactgtggg	tataagaggcgcgac		cagcgtcggtaaccg	Taccctct...
	mCMV-E	mCMV-P		mCMV-P	TPL

FIG. 13

pJV56	TATAAGAGGGCGGACCAGCGTCGGT-ACCGTA	-----CCCTCT
pJV59	TATAT-AAGCAGAGCTCGTTTAGTGAACCGTCAGTTCGTTCTTAGACGCCAACCGCCTCT	

FIG. 14

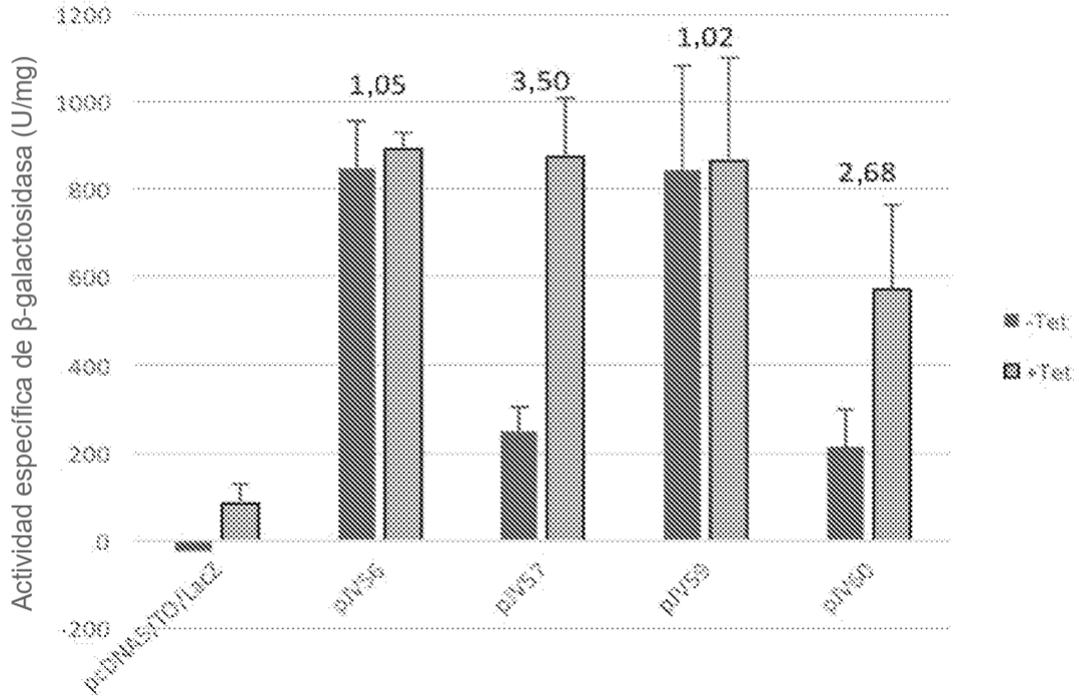


FIG. 15

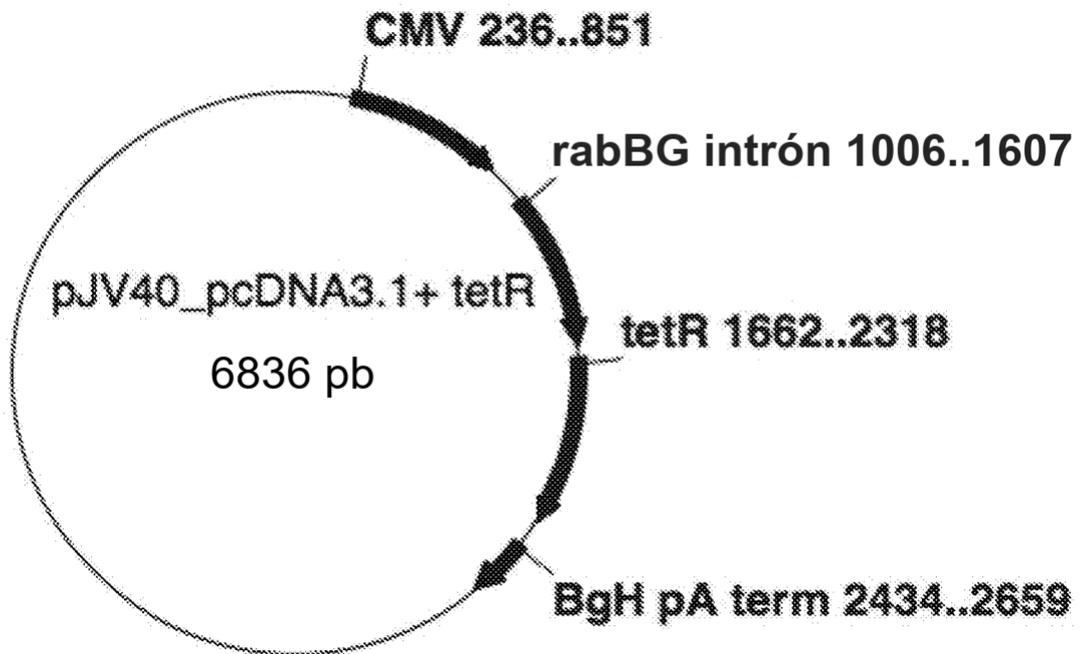


FIG. 16

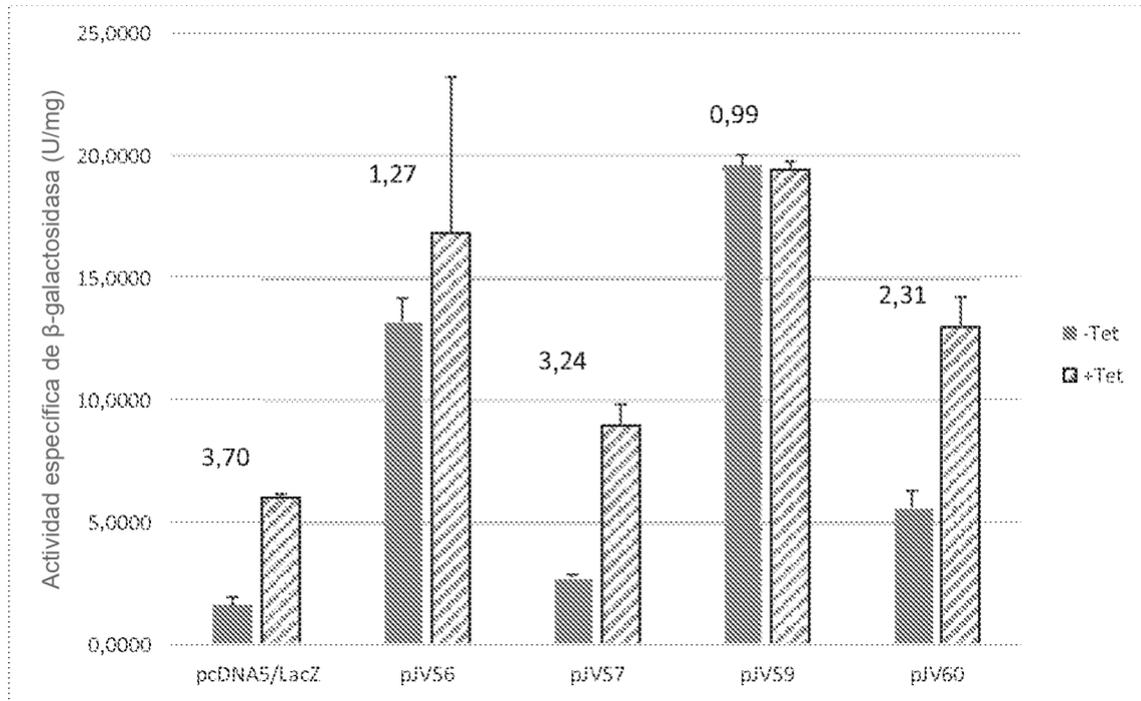


FIG. 17

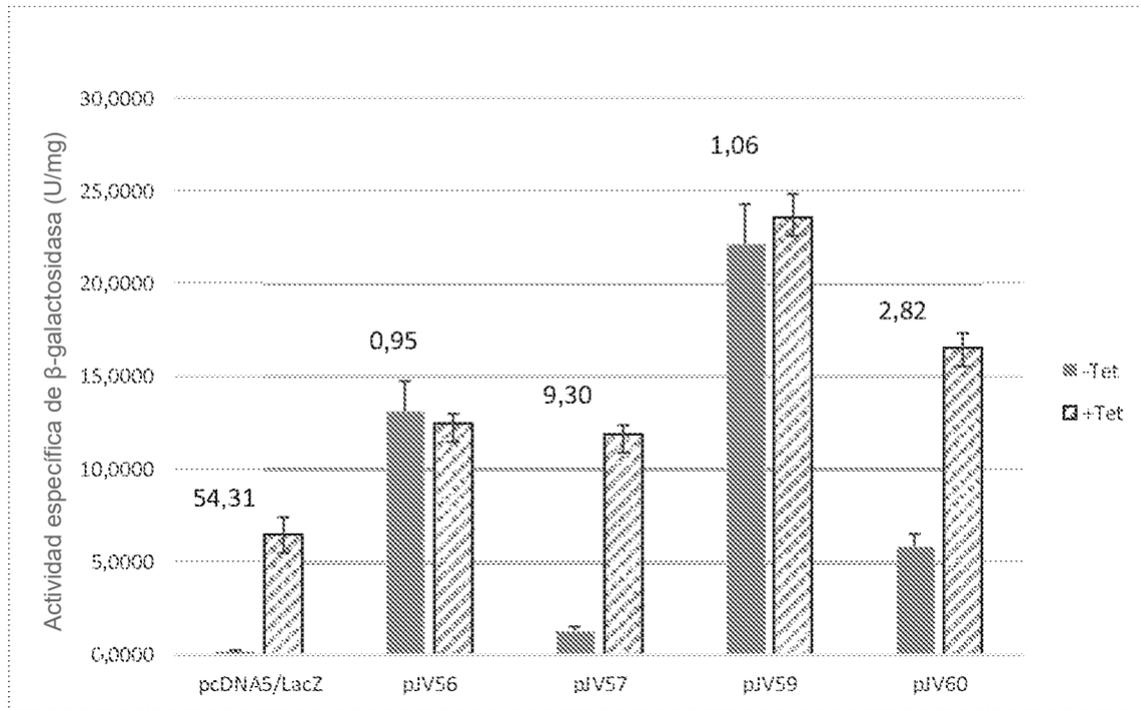
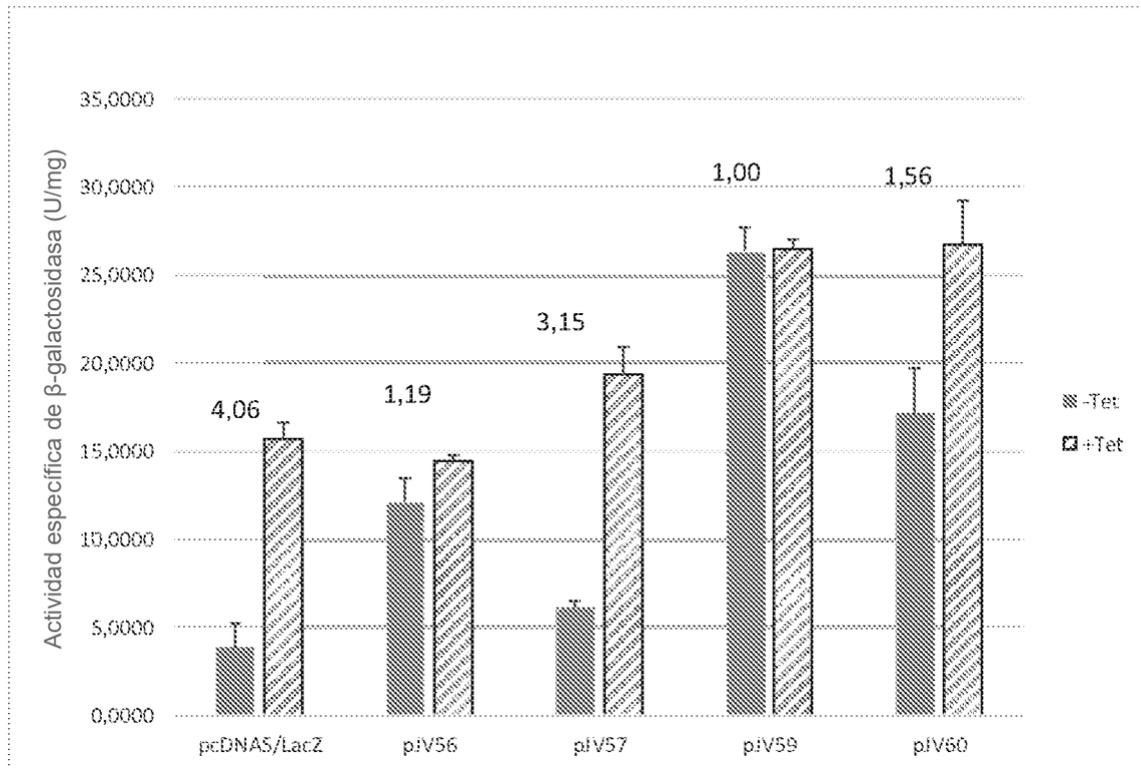


FIG. 18



REFERENCIAS CITADAS EN LA DESCRIPCIÓN

Esta lista de referencias citadas por el solicitante es únicamente para la comodidad del lector. No forma parte del documento de la patente europea. A pesar del cuidado tenido en la recopilación de las referencias, no se pueden excluir errores u omisiones y la EPO niega toda responsabilidad en este sentido.

Documentos de patentes citados en la descripción

- US 62388391 B [0001]
- US 6210924 B [0005]
- US 8771988 B [0005]
- WO 2008020960 A1 [0006]
- US 200812492 A1 [0006]
- WO 2006058900 A1 [0007]
- US 7824907 B [0007]
- WO 2009053368 A1 [0007]
- WO 2008096070 A2 [0007]
- US 7642236 B [0007]
- US 7693698 B [0007]
- WO 2007092370 A1, Mosyak [0007]
- WO 2006063292 A1 [0008]
- WO 2011110864 A1 [0008]
- US 5972650 A, Yao [0010]
- WO 9900510 A, Yao [0010]
- US 5302697 A [0058]
- US 6022952 A [0058]
- US 6335178 B [0058]
- US 7029909 B [0058]
- US 20030104400 A1 [0058]
- US 4767704 A [0067]
- US 4657866 A [0067]
- US 4927762 A [0067]
- US 4560655 A [0067]
- US 5122469 A [0067]
- WO 90103430 A [0067]
- WO 8700195 A [0067]
- US 30985 A [0067]
- US 4816567 A [0080]
- WO 8801649 A [0100]
- US 4946778 A [0100]
- US 5260203 A [0100]
- EP 404097 A [0102]
- WO 9311161 A [0102]
- WO 62388391 A

Literatura diferente de patentes citada en la descripción

- **WURM, F.M.** *CHO quasispecies-Implications for Manufacturing Processes, Processes*, 2013, vol. 1, 296-311 [0005]
- **MIZUSHIMA; NAGATA.** pEF-BOS, a powerful mammalian expression vector. *Nucl. Acids Res.*, 1990, vol. 18 (17), 5322 [0007] [0008]
- **MASAYUKI ; TANAKA.** The CMV Enhancer Stimulates Expression of Foreign Genes from the Human EF-1a Promoter. *Analytical Biochemistry*, 1997, vol. 247, 179-181 [0007]
- **LACY-HULBERT, A. et al.** Interruption of coding sequences by heterologous introns can enhance the functional expression of recombinant genes. *Gene Ther.*, 2001, vol. 8 (8), 649-653 [0008]
- **BRINSTER, R.L. et al.** Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, 836-40 [0008]
- **RUNNING DEER, J.; ALLISON, D. S.** High-level expression of proteins in mammalian cells using Transcription Regulatory sequences from Chinese Hamster Ovary EF-1alpha Gene. *Biotechnology Progress*, 2004, vol. 20, 880-889 [0008] [0130]
- **ORLOVA et al.** Improved elongation factor-1alpha-based vectors for stable high-level expression of heterologous proteins in Chinese hamster ovary cells. *BMC Biotechnology*, 2014, vol. 14, 56 [0008]
- **KIM, S.-Y. et al.** The human elongation factor 1 alpha (EF-1 alpha) first intron highly enhances expression of foreign genes from the murine cytomegalovirus promoter. *J. Biotechnol.*, 2002, vol. 93 (2), 183-87 [0009]
- **YAO et al.** Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives,

- regulates inducible gene expression in mammalian cells. *Hum. Gene Ther.*, 1998, vol. 9 (13), 1939-50 [0010] [0141]
- **PATWARDHAN et al.** High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nature Biotechnology*, 2009, vol. 27 (12), 1173-75 [0019] [0144]
 - **MANIATIS et al.** Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, 1982 [0024]
 - **MANIATIS et al.** *Science*, 1987, vol. 236, 1237 [0026]
 - **VOSS et al.** *Trends Biochem. Sci.*, 1986, vol. 11, 287 [0026]
 - **MIGNONE, F. et al.** Untranslated regions of mRNAs. *Genome Biology*, 2002, vol. 3 (3) [0037]
 - **URLAUB et al.** *Proc. Natl. Acad. Sci. USA*, 1980, vol. 77, 4216 [0061]
 - **GRAHAM et al.** *J. Gen Virol.*, 1977, vol. 36 (59) [0061]
 - **MATHER.** *Biol. Reprod.*, 1980, vol. 23, 243-251 [0061]
 - **MATHER et al.** *Annals N.Y Acad. Sci.*, 1982, vol. 383, 44-68 [0061]
 - **GRAHAM et al.** *Virology*, 1973, vol. 52, 456 [0064]
 - **SAMBROOK et al.** *Molecular Cloning: A Laboratory Manual*, 2001 [0064]
 - **DAVIS et al.** Basic Methods in Molecular Biology. Elsevier, 1986 [0064]
 - **CHU et al.** *Gene*, 1981, vol. 13, 197 [0064]
 - **HAM et al.** *Meth. Enz.*, 1979, vol. 58, 44 [0067]
 - **BARNES et al.** *Anal. Biochem.*, 1980, vol. 102, 255 [0067]
 - **LINDMARK et al.** *J. Immunol. Meth.*, 1983, vol. 62, 1-13 [0071]
 - **GUSS et al.** *EMBO J.*, 1986, vol. 5, 1567-1575 [0071]
 - **KOHLER et al.** *Nature*, 1975, vol. 256, 495 [0080]
 - **CLACKSON et al.** *Nature*, 1991, vol. 352, 624-628 [0080]
 - **MARKS et al.** *J. Mol. Biol.*, 1991, vol. 222, 581-597 [0080]
 - **SONGSIVILAI ; LACHMANN.** *Clin. Exp. Immunol.*, 1990, vol. 79, 315-321 [0082]
 - **KOSTELNY et al.** *J. Immunol.*, 1992, vol. 148, 1547-1553 [0082]
 - **SPIESS et al.** Alternative molecular formats and therapeutic applications for bispecific antibodies. *Mol. Immunol.*, 2015, vol. 67, 95-106 [0082]
 - Fundamental Immunology. Raven Press, 1989 [0084]
 - **ROTH, D. B. ; CRAIG, N. L.** *Cell*, 1998, vol. 94, 411-414 [0089]
 - **CHOTHIA et al.** *J. Mol. Biol.*, 1987, vol. 196, 901-917 [0091]
 - **ZAPATA et al.** *Protein Eng.*, 1995, vol. 8 (10), 1057-1062 [0093]
 - Pluckthun in The Pharmacology of Monoclonal Antibodies. Springer-Verlag, 1994, vol. 1 13, 269-315 [0095]
 - **BIRD et al.** *Science*, 1988, vol. 242, 423-426 [0101]
 - **HUSTON et al.** *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, 5879-5883 [0101]
 - **HOLLINGER et al.** *Proc. Natl. Acad. Sci. USA*, 1993, vol. 90, 6444-6448 [0102]
 - **KABAT et al.** Sequences of Proteins of Immunological Interest. Public Health Service N.I.H, 1991 [0106]
 - **CHOTHIA ; LESK.** *J. Mol. Biol.*, 1987, vol. 196, 901-917 [0106]
 - **CHOTHIA et al.** *Nature*, 1989, vol. 342, 877-883 [0106]
 - Computational Molecular Biology. Oxford University Press, 1988 [0109]
 - Biocomputing Informatics and Genome Projects. Academic Press, 1993 [0109]
 - Computer Analysis of Sequence Data. Humana Press, 1994 [0109]
 - **HEINJE, G.** Sequence Analysis in Molecular Biology. Academic Press, 1987 [0109]
 - Sequence Analysis Primer. M. Stockton Press, 1991 [0109]
 - **CARILLO et al.** *SIAM J. Applied Math.*, 1988, vol. 48, 1073 [0109]
 - **DAYHOFF et al.** Atlas of Protein Sequence and Structure. 1978, vol. 5 [0109]
 - **DEVEREUX et al.** *Nucl. Acid Res.*, 1984, vol. 12, 387 [0110]
 - Genetics Computer Group. University of Wisconsin [0110]
 - **DAYHOFF et al.** Atlas of Protein Sequence and Structure. 1978, vol. 5, 345-352 [0110]
 - **HENIKOFF et al.** *Proc. Natl. Acad. Sci. U.S.A.*, 1992, vol. 89, 10915-10919 [0110]
 - **NEEDLEMAN et al.** *J. Mol. Biol.*, 1970, vol. 48, 443-453 [0111]
 - **SAMBROOK et al.** Molecular Cloning: A Laboratory Guide. Cold Spring Harbor Press, 1989, vol. 1-3 [0122] [0124]
 - **SANGER, F. et al.** *Proc. Natl. Acad. Sci. USA*, 1977, vol. 74, 5463-5467 [0124]
 - **KAUFMAN, R.J. et al.** Improved vectors for stable expression of foreign genes in mammalian cells by use of the untranslated leader sequence from EMC virus. *Nucl. Acids Res.*, 1991, vol. 19 (16), 4485-4490 [0128]
 - **SCHEK et al.** Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Mol. Cell. Biol.*, 1992, vol. 12 (12), 5386-93 [0128]
 - **BODWELL et al.** Long Duration electroporation for achieving high level expression of glucocorticoid receptors in mammalian cell lines. *J. Steroid Biochem. y Mol. Biol.*, 1999, vol. 68 (e 8), 77-82 [0131]
 - **LOGAN ; SHENK.** Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proc. Natl. Acad. Sci. USA*, 1984, vol. 81 (12), 3655-59 [0140]
 - **MIGNONE, F. et al.** Untranslated regions of mRNAs. *Genome Biology*, 2002, vol. 3 (3) [0140]
 - **KIM et al.** Tetracycline repressor-regulated gene repression in recombinant human cytomegalovirus. *J. Virol.*, 1995, vol. 69, 2565-257 [0142]
 - **BODWELL et al.** Long Duration Electroporation for Achieving High Level Expression of Glucocorticoid Receptors in Mammalian Cell Lines. *J. Steroid Biochem. Molec. Biol.*, 1999, vol. 68, 77-82 [0147]
 - **PATWARDHAN et al.** High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.

- Nat Biotechnol.*, 2009, vol. 27 (12), 1173-1175 [0153]
- **SIZEMORE et al.** Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants. *Nucleic Acids Research*, 1990, vol. 18 (10), 2875-2880 [0153]
 - **HILLEN ; BERENS.** Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Ann. Rev. Micro.*, 1994, vol. 48, 345-369