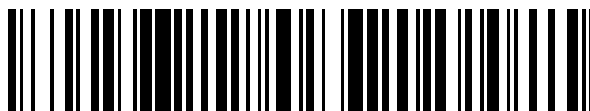


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 780 692**

51 Int. Cl.:

C07K 16/00 (2006.01)
C07K 16/24 (2006.01)
C12P 21/00 (2006.01)
C12N 15/85 (2006.01)
C12N 15/67 (2006.01)
A61K 39/395 (2006.01)
A61K 48/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **24.12.2014** **PCT/NL2014/050907**
87 Fecha y número de publicación internacional: **09.07.2015** **WO15102487**
96 Fecha de presentación y número de la solicitud europea: **24.12.2014** **E 14825469 (1)**
97 Fecha y número de publicación de la concesión europea: **15.01.2020** **EP 3089760**

54 Título: **Constructo y secuencia para la expresión genética mejorada**

30 Prioridad:

31.12.2013 EP 13199873
31.12.2013 EP 13199875

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.08.2020

73 Titular/es:

PROTEONIC BIOTECHNOLOGY IP B.V. (100.0%)
J.H. Oortweg 19-21
2333 CH Leiden, NL

72 Inventor/es:

VAN DER HEIJDEN, MAURICE WILHELMUS y
ENGELS, BART MARINUS

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 780 692 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

Constructo y secuencia para la expresión genética mejorada

Campo de la invención

[0001] La invención se refiere a un método de transcripción y expresión usando un constructo de ácidos nucleicos que se caracteriza por la presencia de un promotor seguido de un promotor intrónico. La invención se refiere además a dicho constructo de ácidos nucleicos, un vector de expresión y una célula que comprende dicho constructo y su uso.

[0002] La invención también se refiere a métodos para la transcripción y opcionalmente la expresión usando una secuencia de nucleótidos. La invención se refiere además a dicha secuencia de nucleótidos y un constructo, vector de expresión y célula que comprende dicha secuencia de nucleótidos y su uso.

Antecedentes de la invención

[0003] Christensen et al., 1996, transgenic res., 5 (3): p. 213-218 se refiere a vectores basados en promotores de ubiquitina para la expresión de alto nivel de genes marcadores seleccionables y/o cribables en plantas monocotiledóneas. Sivamani et al. (DOI: 10.1007/S11103-005-3853-Z) se refiere a la mejora de la expresión de un promotor del gen de la poliubiquitina de arroz. Bianchi et al. (DOI: 10.1371 /journal.pone.0065932) se refiere a la forma en que las secuencias de unión intrónica de Yin Yang 1 y el empalme provocan la mejora mediada por intrones de la expresión del gen de ubiquitina C. Neno et al. (DOI: 10.1016/S0378-1119 (96) 00380-0) se refiere a las regiones aguas arriba del gen de poliubiquitina evolutivamente equivalente de humanos y hámsteres chinos. Neno et al. discuten el mismo tema en Biochim Biophys Acta (1994) 1204(2): pág. 271-278.

[0004] Todavía existe la necesidad en la técnica de métodos alternativos y preferiblemente mejorados para regular la transcripción de un transcrito y opcionalmente regular la expresión de una proteína o un polipéptido de interés en células huésped.

Resumen de la invención

[0005] La presente invención se refiere a un método para la transcripción y opcionalmente a la purificación del transcrito producido que comprende los pasos de:

- a) proporcionar un constructo de ácidos nucleicos que comprende un primer promotor, un segundo promotor y una secuencia de nucleótidos de interés, en donde dichos primer promotor y segundo promotor están operativamente enlazados a dicha secuencia de nucleótidos de interés, y en donde dicho segundo promotor está flanqueado por una primera secuencia intrónica ubicada aguas arriba de dicho promotor y una segunda secuencia intrónica ubicada aguas abajo de dicho promotor; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada produzca un transcrito de la secuencia de nucleótidos de interés; y opcionalmente
- d) purificar dicho transcrito producido.

[0006] La presente invención se refiere además a un método para expresar y opcionalmente purificar una proteína o un polipéptido de interés que comprende el paso de:

- a) proporcionar un constructo de ácidos nucleicos que comprende un primer promotor, un segundo promotor y una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés, en donde dichos primer promotor y segundo están operativamente enlazados a dicha secuencia de nucleótidos que codifica una proteína o un polipéptido de interés, y en donde dicho segundo promotor está flanqueado por una primera secuencia intrónica ubicada aguas arriba de dicho promotor y una segunda secuencia intrónica ubicada aguas abajo de dicho promotor; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada exprese la proteína o polipéptido de interés; y opcionalmente

purificar dicha proteína o polipéptido de interés.

[0007] Preferiblemente, dicha primera secuencia intrónica comprende al menos un sitio de empalme donador y dicha segunda secuencia intrónica comprende al menos un sitio de empalme aceptor. Además, el constructo de ácidos nucleicos del paso a) del método de la invención comprende las siguientes secuencias de nucleótidos indicadas aquí en sus posiciones relativas en la dirección 5' a 3': (i) un primer promotor (ii) una primera secuencia intrónica que comprende al menos un sitio de empalme donador, (iii) un segundo promotor, (iv) una segunda secuencia intrónica que comprende al menos un sitio de empalme aceptor; y (v) una secuencia de nucleótidos

que codifica una proteína o un polipéptido de interés, en donde preferiblemente dicho primer promotor, dicha primera secuencia intrónica que comprende al menos un sitio de empalme donador, dicho segundo promotor y dicha segunda secuencia intrónica que comprende al menos un sitio de empalme aceptor están todos operativamente enlazados a dicha secuencia de nucleótidos que codifica una proteína o un polipéptido de interés.

[0008] Preferiblemente, dicho primer promotor tiene al menos un 50 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1 o los nucleótidos 1-614 de la SEC ID n.º: 2 en toda su longitud. En la tabla 1 se proporciona una descripción general de todos los n.º de SEC ID. Preferiblemente, una secuencia de nucleótidos que comprende dicho primer promotor y dicha primera secuencia intrónica que comprende al menos un sitio de empalme donador tiene al menos un 50 % de identidad con las SEC ID n.º: 1 o 2 en toda su longitud. Preferiblemente, dicho segundo promotor tiene al menos un 50 % de identidad de secuencia con la SEC ID n.º: 57 o la SEC ID n.º: 58 en toda su longitud.

[0009] La presente invención se refiere además a un constructo de ácidos nucleicos que comprende un primer promotor y un segundo promotor, en donde dichos primer promotor y segundo promotor están configurados para estar operativamente enlazados a una secuencia de nucleótidos de interés, y en donde dicho segundo promotor está flanqueado por una primera secuencia intrónica ubicada aguas arriba de dicho promotor y una segunda secuencia intrónica ubicada aguas abajo de dicho promotor. Preferiblemente, dicha primera secuencia intrónica comprende al menos un sitio de empalme donador y preferiblemente dicha segunda secuencia intrónica comprende al menos un sitio de empalme aceptor. Además, preferiblemente un constructo de ácidos nucleicos de la invención comprende las siguientes secuencias de nucleótidos indicadas aquí en sus posiciones relativas en la dirección 5' a 3': (i) un primer promotor (ii) una primera secuencia intrónica que comprende al menos un sitio de empalme donador, (iii) un segundo promotor, (iv) una segunda secuencia intrónica que comprende al menos un sitio de empalme aceptor; y opcionalmente (v) una secuencia de nucleótidos de interés, en donde preferiblemente dicho primer promotor, dicha primera secuencia intrónica que comprende al menos un sitio de empalme donador, dicho segundo promotor y dicha segunda secuencia intrónica que comprende al menos un sitio de empalme aceptor están todos configurados para estar operativamente enlazado a dicha secuencia de nucleótidos opcional de interés.

[0010] Preferiblemente, dicho primer promotor tiene al menos un 50 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1 o los nucleótidos 1-614 de la SEC ID n.º: 2 en toda su longitud. Preferiblemente, una secuencia de nucleótidos que comprende dicho primer promotor y dicha primera secuencia intrónica que comprende al menos un sitio de empalme donador tiene al menos un 50 % de identidad con las SEC ID n.º: 1 o 2 en toda su longitud. Preferiblemente, dicho segundo promotor tiene al menos un 50 % de identidad de secuencia con la SEC ID n.º: 57 o la SEC ID n.º: 58 en toda su longitud.

[0011] Preferiblemente, dicho constructo de ácidos nucleicos es un constructo aislado. Preferiblemente, dicho constructo de ácidos nucleicos es un constructo de ácidos nucleicos recombinante. Preferiblemente, dicha secuencia de nucleótidos opcional de interés es una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o polipéptido heterólogo.

[0012] La presente invención se refiere además a un vector de expresión que comprende un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante tal como se define en el presente documento.

[0013] La presente invención se refiere además a una célula que comprende un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante tal como se define en el presente documento y/o un vector de expresión tal como se define en el presente documento.

[0014] La presente invención también se refiere a un uso de un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante tal como se define en el presente documento y/o un vector de expresión tal como se define en el presente documento y/o una célula tal como se define en el presente documento para la transcripción de una secuencia de nucleótidos de interés.

[0015] La presente invención se refiere además a un uso de un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante tal como se define en el presente documento y/o un vector de expresión tal como se define en el presente documento y/o una célula tal como se define en el presente documento para la expresión de una proteína o un polipéptido de interés.

[0016] La presente divulgación se refiere además a un método para la transcripción y opcionalmente a la purificación del transcrito producido que comprende los pasos de:

- a) proporcionar un constructo de ácidos nucleicos que comprende un elemento potenciador de la expresión, un promotor heterólogo y una secuencia de nucleótidos de interés de la divulgación, en donde dicho

elemento potenciador de la expresión y dicho promotor heterólogo están operativamente enlazados a dicha secuencia de nucleótidos de interés; y,

b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,

c) permitir que dicha célula transformada produzca un transcrito de la secuencia de nucleótidos de interés; y opcionalmente

d) purificar dicho transcrito producido.

[0017] La presente divulgación se refiere además a un método para expresar y opcionalmente purificar una proteína o un polipéptido de interés que comprende el paso de:

a) proporcionar un constructo de ácidos nucleicos que comprende un elemento potenciador de la expresión, un promotor heterólogo y una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés, en el que dicho elemento potenciador de la expresión y dicho promotor heterólogo están operativamente enlazados a dicha secuencia de nucleótidos que codifica una proteína o un polipéptido de interés; y,

b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,

c) permitir que dicha célula transformada exprese la proteína o polipéptido de interés; y opcionalmente

d) purificar dicha proteína o polipéptido de interés.

[0018] Preferiblemente, dicho constructo de ácidos nucleicos de dicho método de transcripción y/o expresión y opcionalmente de purificación de un transcrito y/o proteína o polipéptido de interés comprende además un elemento regulador adicional de la expresión operativamente enlazado a dicha secuencia de nucleótidos de interés y/o dicha secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicho elemento regulador adicional de la expresión comprende una secuencia intrónica. Un elemento regulador de expresión adicional preferido comprende o es un elemento potenciador de expresión adicional. Más preferiblemente, dicho elemento regulador adicional de la expresión comprende además un elemento potenciador de la traducción.

[0019] La presente divulgación se refiere además a una molécula de ácidos nucleicos que está representada por una secuencia de nucleótidos que comprende un elemento potenciador de la expresión de la divulgación, es decir, una secuencia de nucleótidos que tiene al menos un 50 % de identidad con las SEC ID n.º: 1 o 2 en toda su longitud. En la tabla 1 se proporciona un resumen general de todos los n.º de SEC ID. Preferiblemente, dicha molécula de ácidos nucleicos es una molécula de ácidos nucleicos aislada. Preferiblemente, dicha molécula de ácidos nucleicos o molécula de ácidos nucleicos aislada está representada por una secuencia de nucleótidos que tiene al menos un 50 % de identidad de secuencia con las SEC ID n.º: 1 o 2 en toda su longitud. Preferiblemente, dicha molécula de ácidos nucleicos o molécula de ácidos nucleicos aislada está representada por una secuencia de nucleótidos que comprende una secuencia derivada del gen *Cricetulus griseus* para poliubiquitina de como máximo 8000 nucleótidos. La presente invención se refiere además a un constructo de ácidos nucleicos que comprende una molécula de ácidos nucleicos de la invención. Preferiblemente, dicho constructo de ácidos nucleicos está representado por una secuencia de nucleótidos que comprende además un promotor heterólogo, en el que preferiblemente dicho elemento potenciador de la expresión y dicho promotor heterólogo están configurados para estar los dos operativamente enlazados a una secuencia de nucleótidos opcional de interés. Preferiblemente, dicho constructo de ácidos nucleicos comprende además un elemento regulador adicional de la expresión, en donde preferiblemente dicho elemento potenciador de la expresión, dicho promotor heterólogo y dicho elemento regulador adicional de la expresión están configurados para estar todos operativamente enlazados a dicha secuencia de nucleótidos opcional de interés. Preferiblemente, dicho elemento regulador adicional de la expresión comprende además un elemento potenciador de la traducción. Preferiblemente, dicho elemento regulador adicional de la expresión comprende una secuencia intrónica. Preferiblemente, dicha secuencia de nucleótidos opcional de interés es una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o polipéptido heterólogo.

[0020] Preferiblemente, dicho constructo de ácidos nucleicos es un constructo de ácidos nucleicos recombinante y/o aislado.

[0021] La presente invención se refiere además a un vector de expresión que comprende una molécula de ácidos nucleicos o una molécula de ácidos nucleicos aislada tal como se define en el presente documento, y/o un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante y/o aislado tal como se define en el presente documento.

[0022] La presente invención se refiere además a una célula que comprende una molécula de ácidos nucleicos o una molécula de ácidos nucleicos aislada tal como se define en el presente documento, y/o un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante tal como se define en el presente documento, y/o un vector de expresión tal como se define en el presente documento.

[0023] La presente invención también se refiere al uso de una molécula de ácidos nucleicos o una molécula de ácidos nucleicos aislada tal como se define en el presente documento, y/o un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante y/o aislado tal como se define en el presente documento, y/o un vector de expresión tal como se define en el presente documento y/o una célula tal como se define en el presente documento para la transcripción de una secuencia de nucleótidos de interés.

[0024] La presente invención se refiere además a un uso de una molécula de ácidos nucleicos o una molécula de ácidos nucleicos aislada tal como se define en el presente documento, y/o un constructo de ácidos nucleicos o un constructo de ácidos nucleicos recombinante y/o aislado tal como se define en el presente documento, y/o un vector de expresión tal como se define en el presente documento y/o una célula tal como se define en el presente documento para la expresión de una proteína o un polipéptido de interés.
Descripción de la invención.

[0025] Los inventores identificaron un constructo de expresión para aumentar la expresión de una proteína o un polipéptido de interés. El constructo de expresión de la invención se caracteriza por dos promotores operativamente enlazados a una secuencia codificante de una proteína o un polipéptido de interés. Un constructo de expresión de la invención típicamente comprende un primer promotor seguido de un segundo promotor, una secuencia codificante de la proteína o polipéptido de interés y una secuencia de poliadenilación, en donde dicho segundo promotor está flanqueado por secuencias intrónicas. Dicho promotor flanqueado por secuencias intrónicas se denomina en este documento promotor intrónico. Se pueden insertar secuencias reguladoras de la expresión adicionales aguas arriba y aguas abajo de dichos primer y/o segundo promotor y/o aguas abajo de la secuencia de poliadenilación. Los inventores descubrieron sorprendentemente que un constructo de expresión de la invención que comprende un promotor seguido de un promotor intrónico operativamente enlazado a una secuencia codificante de una proteína o un polipéptido de interés, da como resultado un aumento significativo en la expresión de dicha proteína o polipéptido de interés en comparación con un constructo de expresión que comprende solo un promotor operativamente enlazado a dicha secuencia de codificación. Los inventores han descubierto que la expresión de proteínas en principio pobremente expresadas aumenta a niveles apreciables cuando se usa la combinación de un promotor y un promotor intrónico de la invención en lugar de un único promotor en un constructo de expresión que codifica estas proteínas, como se muestra en los ejemplos, más específicamente en el ejemplo 1. La combinación de un promotor y un promotor intrónico de la invención en un constructo de expresión para una proteína en principio pobremente expresada facilita la generación de líneas clonales y permite la generación de líneas clonales con niveles de expresión aumentados y relevantes, como se muestra en los ejemplos, más específicamente en el ejemplo 3. Además, la expresión de proteínas en principio altamente expresadas se incrementa aún más cuando se usa la combinación de un promotor y un promotor intrónico de la invención en lugar de un único promotor en un constructo de expresión que codifica estas proteínas, como se muestra en los ejemplos, más específicamente en el ejemplo 5. Además, se encontró un aumento en la cantidad total de ARNm y en la expresión medido en el nivel de proteína tal como se detalla a continuación en el presente documento y se muestra en los ejemplos adjuntos. Además, el porcentaje de líneas celulares de alto productor en un grupo transfectado de manera estable es significativamente mayor en comparación con los grupos con un único promotor operativamente enlazado a la secuencia codificante. Ya que la secuencia de nucleótidos de la invención que comprende tanto un promotor como un promotor intrónico operativamente enlazados a una secuencia de nucleótidos de interés da como resultado un aumento en la transcripción, la presente invención no se limita al uso de esta secuencia en la expresión de proteínas y/o polipéptidos y/o en la producción de proteínas y/o polipéptidos, sino que extiende al uso de esta combinación de un promotor y un promotor intrónico en métodos en los que se desean niveles más altos de transcripción, por ejemplo, en métodos para producir transcritos de ARN no codificantes tal como se especifica adicionalmente en este documento. Además, un beneficio adicional de la invención es que, aparte de un aumento en el nivel de transcripción y/o un incremento en el nivel de expresión de la proteína o polipéptido de interés, la invención permite que se formen diferentes transcritos tal como se muestra más detalladamente en el presente documento.

[0026] Los inventores identificaron un elemento potenciador de la expresión para aumentar la expresión de una proteína o un polipéptido de interés. La presente divulgación se refiere a dicho elemento potenciador de la expresión. La aplicación del elemento potenciador de la expresión de la divulgación en un constructo de expresión que comprende además un promotor heterólogo operativamente enlazado a una secuencia que codifica una proteína o un polipéptido de interés, da como resultado un marcado aumento en la expresión de dicha proteína o polipéptido de interés en comparación con dicha expresión usando un constructo de expresión similar que solo difiere del constructo de expresión anterior en que el elemento potenciador de la expresión de la divulgación está ausente. Los inventores han encontrado que la expresión de proteínas en principio pobremente expresadas aumenta a niveles apreciables después de la inserción del elemento en un constructo de expresión que codifica estas proteínas tal como se muestra en los ejemplos, más específicamente en el ejemplo 1. La inserción del elemento potenciador de la expresión en un constructo de expresión para una proteína en principio pobremente expresada facilita la generación de líneas clonales y permite la generación de líneas clonales con niveles de expresión relevantes, como se muestra en los ejemplos, más específicamente en el ejemplo 3. Además, la expresión de proteínas en principio altamente expresadas se incrementa aún más después de la

inserción del elemento en un constructo de expresión que codifica estas proteínas tal como se muestra en los ejemplos, más específicamente en el ejemplo 5. Además, se encontró un aumento en la cantidad total del nivel de ARNm y/o en la expresión medido en el nivel de proteína tal como se detalla a continuación en el presente documento y se muestra en los ejemplos adjuntos. Ya que el elemento potenciador de la expresión de la divulgación puede dar como resultado un aumento en la transcripción, la presente divulgación no se limita al uso de este elemento en la expresión de proteínas y/o polipéptidos y/o la producción de proteínas y/o polipéptidos, sino que se extiende al uso de este elemento en métodos donde se desean niveles más altos de transcripto, por ejemplo, en métodos para producir transcripciones de ARN no codificantes tal como se especifica más adelante en este documento.

Primer aspecto

[0027] En un primer aspecto, la presente invención proporciona un constructo de ácidos nucleicos para aumentar la transcripción y/o expresión, que comprende un primer promotor y un segundo promotor, que están configurados para estar ambos operativamente enlazados a una secuencia de nucleótidos opcional de interés dentro de un constructo de expresión. «Opcional» se entiende en el presente documento como no necesariamente presente en un constructo de expresión. Por ejemplo, dicha secuencia de nucleótidos de interés no necesita estar presente en un vector de expresión comercializado, sino que una persona experta en la técnica puede introducirla fácilmente antes de usarla en un método de la invención.

[0028] Preferiblemente, dentro de este primer aspecto, dicho constructo de nucleótidos de la invención que comprende un primer promotor y un segundo promotor es capaz de aumentar la transcripción de una secuencia de nucleótidos de interés que está bajo el control de dichos primer promotor y segundo promotor. Alternativamente o en combinación con la transcripción incrementada, dicho constructo de nucleótidos también es preferiblemente capaz de aumentar la expresión de una proteína o un polipéptido de interés codificado por dicha secuencia de nucleótidos de interés. Preferiblemente, los niveles de transcripción se evalúan en un sistema de expresión usando un constructo de expresión que comprende dichos primer promotor y segundo promotor operativamente enlazados a una secuencia de nucleótidos de interés usando un ensayo adecuado tal como RT-qPCR. Preferiblemente, dentro de este primer aspecto, el constructo de nucleótidos de la invención que comprende el primer promotor y el segundo promotor de la invención permite un aumento en la transcripción de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % de dicha secuencia de nucleótidos de interés en comparación con la transcripción que usa un constructo que solo difiere en que la secuencia de nucleótidos de interés está bajo el control de un único promotor, preferiblemente cuando se ensaya en un sistema como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, la transcripción de una secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide preferiblemente en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un primera secuencia promotora y una segunda secuencia promotora que se van a ensayar operativamente enlazadas a dicha secuencia de nucleótidos de interés. La transcripción se mide preferiblemente usando RT-qPCR y los niveles de transcripción se comparan con los niveles de transcripción de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que dichos primer promotor y segundo promotor se reemplazan por un único promotor, preferiblemente un promotor de CMV representado por la SEC ID n.º: 57, en el vector de expresión utilizado.

[0029] Preferiblemente, dentro de este primer aspecto, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión que comprende dichos primer promotor y segundo promotor operativamente enlazados a una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo de inmunoabsorción ligada a enzimas (ELISA), una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, el primer promotor y el segundo promotor de la invención permiten un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la expresión de dicha proteína o polipéptido que usa un constructo que solo difiere en que la secuencia de codificación de la proteína o polipéptido de interés está bajo el control de un único promotor, preferiblemente cuando se ensaya en un sistema como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un primera secuencia promotora y una segunda secuencia promotora que se van a ensayar operativamente enlazadas a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión solo difiere en que dichos primer promotor y segundo promotor son

reemplazados por un único promotor, preferiblemente un promotor de CMV representado por la SEC ID n.º: 57, en el vector de expresión utilizado.

[0030] Preferiblemente, dentro de dicho primer aspecto, el aumento de la expresión de proteínas o polipéptidos es independiente del número de copias según lo establecido por un ensayo adecuado para determinar la dependencia del número de copias realizado por una persona experta tal como, por ejemplo, un ensayo Taqman triplex tal como se detalla adicionalmente en el ejemplo 4 de la presente invención.

[0031] Preferiblemente, dentro de dicho primer aspecto, dicho primer promotor está ubicado aguas arriba o en el sitio 5' de dicho segundo promotor. Preferiblemente, dicho segundo promotor tal como se define en el presente documento debería estar desprovisto de elementos de secuencia que actuarán como terminadores de la transcripción. Los terminadores de la transcripción bien conocidos por los expertos en la técnica son secuencias que pueden dar como resultado la terminación prematura de la transcripción, como por ejemplo, estructuras de horquilla estables, secuencias repetidas tales como repeticiones terminales largas (LTR) o repeticiones Alu, motivos de poliadenilación y elementos transponibles

[0032] Dentro del contexto del primer aspecto de la invención, un promotor es un promotor capaz de iniciar la transcripción en la célula huésped elegida. Los promotores tal como se usan en el presente documento incluyen promotores específicos de tejido, preferidos de tejido, específicos del tipo de célula, inducibles y constitutivos, tal como se define en el presente documento en la sección Definiciones. Los promotores que pueden estar comprendidos dentro de dichos primer o segundo promotor tal como se define en el presente documento son promotores que pueden emplearse en la transcripción de secuencias de nucleótidos de interés y/o la expresión de proteínas o polipéptidos de interés, preferiblemente en células de mamífero, e incluyen, entre otros, el promotor de citomegalovirus (CMV) humano o murino, un promotor del virus del simio (SV40), un promotor de ubiquitina C (UBC) humana o de ratón, un promotor del factor de elongación alfa (EF1-a) humano, de ratón o de rata, un promotor de beta-actina de ratón o hámster o un promotor de hámster rpS21. Los elementos de respuesta de Tet-Off y Tet-On aguas arriba de un promotor mínimo tal como un promotor de CMV son un ejemplo de promotor inducible de mamífero. Ejemplos de promotores adecuados de levadura y hongos son el promotor Leu2, el promotor de galactosa (Gall o Ga17), el promotor de alcohol deshidrogenasa I (ADH1), el promotor de glucoamilasa (Gla), el promotor de triosa fosfato isomerasa (TPI), el factor de elongación de traducción EF-1 alfa (TEF2), el promotor de la gliceraldehído-3-fosfato deshidrogenasa (gpdA), promotor de la alcohol oxidasa (AOX1) o el promotor de la glutamato deshidrogenasa (gdhA). Un ejemplo de un promotor ubicuo fuerte para la expresión en plantas es el promotor 35S del virus del mosaico de la coliflor (CaMV).

[0033] En una realización dentro de dicho primer aspecto, dichos primer promotor y segundo promotor son promotores similares. Preferiblemente, dicho primer promotor tiene al menos un 20 %, un 25 %, un 30 %, un 35 %, un 40 %, un 45 %, un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con dicho segundo promotor.

[0034] En otra realización dentro de dicho primer aspecto, dichos primer promotor y segundo promotor son promotores distintos o diferentes. Preferiblemente, dicho primer promotor tiene menos de un 80 %, un 75 %, un 70 %, un 65 %, un 60 %, un 55 %, un 50 %, un 45 %, un 40 %, un 35 %, un 30 %, un 25 %, un 20 %, un 15 %, un 10 % o un 5 % de identidad de secuencia con dicho segundo promotor.

[0035] En una realización preferida dentro de dicho primer aspecto, dicha primera secuencia promotora comprende o consta de un promotor de UBC o un promotor de CCT8 y dicho segundo promotor comprende o consta de un promotor de CMV, o viceversa. Preferiblemente, dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1 o con los nucleótidos 1-614 de la SEC ID n.º: 2. Preferiblemente, dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 58, o preferiblemente con la SEC ID n.º: 57. Se prefiere dentro de dicho primer aspecto una secuencia de nucleótidos de la invención en la que dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1 y dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 57. También se prefiere dentro de dicho primer aspecto una secuencia de nucleótidos de la invención en la que dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-614 de la SEC ID n.º: 2 y dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65

%, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 57. También se prefiere dentro de dicho primer aspecto una secuencia de nucleótidos de la invención en la que dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 58, o preferiblemente con la SEC ID n.º: 57 y en donde dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º 1 o con los nucleótidos 1-614 de la SEC ID n.º 2. Preferiblemente dentro de dicho primer aspecto una secuencia de nucleótidos de la invención en la que dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 57 y dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1. También se prefiere una secuencia de nucleótidos de la invención en la que dicho primer promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: Preferiblemente y dicho segundo promotor comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-614 de la SEC ID n.º: 2.

[0036] Debe entenderse que dentro de dicho primer aspecto, dichos primer promotor y/o segundo promotor no constan solo de una secuencia potenciadora del promotor, tal como una secuencia seleccionada del grupo que consta de las SEC ID n.º: 52-54. Preferiblemente, dichos primer promotor y segundo promotor no constan solo de una secuencia potenciadora del promotor, tal como una secuencia seleccionada del grupo que consta de las SEC ID n.º: 52-54. Preferiblemente, una secuencia de nucleótidos de la invención no comprende o no consta de las SEC ID n.º: 55 o 56.

[0037] En una realización preferida dentro de dicho primer aspecto, dicho segundo promotor está flanqueado por una primera secuencia intrónica en el sitio 5' o aguas arriba de dicho segundo promotor y una segunda secuencia intrónica en el sitio 3' o aguas abajo de dicho segundo promotor. Estar «flanqueado» se entiende aquí como estar posicionado entre dichas secuencias indicadas opcionalmente separadas por 1-50, 1-60, 1-70, 1-80, 1-90, 1-100, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-800, 1-900, 1-1000, 1-5000 o 1-100 000 nucleótidos, entendiéndose que estos nucleótidos abarcan la 5'-UTR. Se entiende que una secuencia intrónica es al menos parte de la secuencia de nucleótidos de un intrón. Preferiblemente, dicha primera secuencia intrónica en el sitio 5' o aguas arriba de dicho segundo promotor comprende al menos un sitio de empalme donador o sitio de empalme GT. Un sitio de empalme donador se entiende en el presente documento como un sitio de empalme que, cuando se combina con un sitio de empalme aceptor tal como se define en el presente documento, da como resultado la formación de un intrón según se define en la sección Definición. Preferiblemente, una secuencia de nucleótidos es un intrón si al menos un 2 %, un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 % o un 100 % del ARN primario pierde esta secuencia mediante empalme de ARN usando un ensayo adecuado para detectar el empalme de intrones, como por ejemplo, entre otros, la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) seguido de un análisis de tamaño o secuencia de la RT-PCR. Los sitios de empalme donadores preferidos de la invención son M-W-G-[corte]-G-T-R-A-G-K o M-A-R-[corte]-G-T-R-A-G-K en caso de que la célula huésped sea una célula de mamífero, AG-[corte]-G-T-A-W-K en caso de que la célula huésped sea una célula vegetal, [corte]-G-T-A-W-G-T-T en caso de que la célula huésped sea una célula de levadura y RG-[corte]-G-T-R-A-G, en caso de que la célula huésped sea una célula de insecto. «[corte]» debe entenderse aquí como el sitio de corte específico donde se realizará el empalme. El empalme de intrones se puede evaluar funcionalmente usando un ensayo según se detalla en la sección Definición bajo «intrón». Más preferiblemente, el sitio de empalme donador comprendido dentro de la primera secuencia intrónica de la invención es C-T-G- [corte]-G-T-G-A-G-G o A-A-A- [corte]-G-T-G-A-G-G. Preferiblemente, dicha primera secuencia intrónica consta de dicho sitio de empalme donador o sitio de empalme GT. Preferiblemente, dicha primera secuencia intrónica comprende un único sitio de empalme donador tal como se define en el presente documento. Preferiblemente, dicha primera secuencia intrónica está libre de un sitio de empalme aceptor tal como se define más adelante en el presente documento.

[0038] Preferiblemente, dentro de dicho primer aspecto, dicha segunda secuencia intrónica en el sitio 3' o aguas abajo de dicho promotor comprende al menos un sitio de empalme aceptor que se entiende aquí como el sitio de empalme AG precedido preferiblemente por una secuencia rica en pirimidina o una secuencia de nucleótidos del tracto de polipirimidina, opcionalmente separado del sitio de empalme AG por 1-50 nucleótidos, y que opcionalmente comprende además un sitio de ramificación que comprende la secuencia Y-T-N-A-Y, en el sitio 5' de la secuencia de nucleótidos del tracto de polipirimidina, en donde el sitio de ramificación puede tener la secuencia de nucleótidos C-Y-G-A-C. Un sitio de empalme aceptor se entiende en el presente documento como

un sitio de empalme que, cuando se combina con un sitio de empalme donador incluido dentro de un constructo, da como resultado la formación de un intrón tal como se define en la sección Definición. Preferiblemente, el sitio de empalme aceptor o el sitio de empalme AG tienen la secuencia [Y-rico]-N-Y-A-G-[corte]. Preferiblemente, el sitio de empalme aceptor o sitio de empalme AG tienen la secuencia [Y-rico]-N-Y-A-G-[corte]-R en caso de que la célula huésped sea una célula de mamífero, [Y-rico]-D-Y-A-G-[corte]-R o [Y-rico]-DY-A-G-[corte]-R-W en caso de que la célula huésped sea una célula vegetal, [Y-rico]-A-Y-A-G-[corte] en caso de que la célula huésped sea una célula de levadura y [Y-rico]-N-Y-A-G-[corte] en caso de que la célula huésped sea una célula de insecto. «[Y-rico]» debe entenderse aquí como el tracto de polipirimidina que se define preferiblemente como una secuencia consecutiva de al menos 10 nucleótidos que comprende al menos 6, 7, 8, 9 o preferiblemente 10 nucleótidos de pirimidina. Preferiblemente, dicho sitio de empalme aceptor o sitio de empalme GT tiene la secuencia Y-A-G-[corte]-R. Preferiblemente, dicha segunda secuencia intrónica comprende un único sitio de empalme aceptor. En una realización, dicha segunda secuencia intrónica está libre de un sitio de empalme donador tal como se define en el presente documento. En una realización alternativa, dicha segunda secuencia intrónica comprende tanto un sitio de empalme donador como un sitio de empalme aceptor tal como se define en el presente documento. Más preferiblemente, dicha segunda secuencia intrónica es un intrón tal como se define en la sección Definición. Preferiblemente, el segundo promotor y las secuencias intrónicas que flanquean al segundo promotor están configurados para formar un promotor intrónico (se hace referencia a la figura 1). Un promotor intrónico es conocido por una persona experta en la técnica como un promotor ubicado dentro de una secuencia intrónica. Preferiblemente, dicho promotor intrónico es un intrón tal como se define en la sección Definición. Preferiblemente, los límites del promotor intrónico de la presente invención están formados por el sitio de empalme donador de la secuencia intrónica en el sitio 5' o aguas arriba del segundo promotor de la invención y el sitio de empalme aceptor de la secuencia intrónica en el sitio 3' o aguas abajo del segundo promotor de la invención. El promotor intrónico de la invención puede tener una longitud que sea comparable o similar a los intrones de origen natural, preferiblemente comparable o similar a los intrones de origen natural en la célula u organismo huésped tal como se define en el presente documento. Preferiblemente, dicho promotor intrónico tal como se define en el presente documento tiene como máximo 12 000 nucleótidos de longitud. Preferiblemente, dicha primera secuencia intrónica en el sitio 5' o aguas arriba de dicho segundo promotor está ubicada en el sitio 3' o aguas abajo de dicho primer promotor. Preferiblemente, el primer promotor y el segundo promotor, las secuencias intrónicas que flanquean al segundo promotor y una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés están configurados de tal manera que el primer promotor esté aguas arriba del segundo promotor, en donde el segundo promotor está flanqueado por dichas secuencias intrónicas para formar un promotor intrónico, y en donde dichos primer promotor y segundo promotor están configurados para estar tanto aguas arriba como operativamente enlazados a la secuencia de nucleótidos que codifica una proteína o un polipéptido de interés (figura 1). El promotor intrónico puede comprender más elementos potenciadores de la expresión, pero preferiblemente el promotor intrónico está libre de sitios de empalme adicionales además de los sitios de empalme donador y aceptor tal como se define en el presente documento dentro de la primera y segunda secuencias intrónicas tal como se define en el presente documento. Preferiblemente, una o más secuencias potenciadoras de la expresión están comprendidas dentro de dicho primer y/o dicho segundo promotor. Sin querer estar sujeto a ninguna teoría, la transcripción que comienza desde cualquiera de los dos promotores puede dar como resultado diferentes transcritos (pre-ARNm) que, al empalmar, dan como resultado diferentes ARNm tal como se ilustra en la figura 1. En apoyo de esta teoría, los inventores descubrieron que se forman diferentes transcritos usando un constructo de la invención (a este respecto se hace referencia a la figura 1, el ejemplo 8 y la figura 10). Además, se encuentra que la actividad aumentada se ve severamente disminuida por la mutación de 4 nucleótidos en el promotor intrónico que evita el empalme intrónico correcto (a este respecto se hace referencia a la figura 9 y al ejemplo 7), lo que apoya también la teoría de que ambos promotores están activos en el constructo. Por lo tanto, un beneficio adicional de la invención es que, aparte de un aumento en la transcripción de la secuencia de nucleótidos de interés y/o un aumento en el nivel de expresión de la proteína o el polipéptido de interés, la invención permite que se formen diferentes transcritos. Los «diferentes transcritos» se entienden aquí como transcritos que son estructuralmente diferentes o distintos, es decir, que tienen una secuencia de nucleótidos diferente o distinta. Por lo tanto, un beneficio adicional de la invención es dirigir o redirigir el empalme de una secuencia de nucleótidos de interés. Dependiendo de la ubicación de los sitios de empalme intrónico, los transcritos pueden tener una secuencia UTR diferente o distinta y/o una secuencia codificante diferente o distinta. También es posible que solo se forme un tipo de transcrito, p. ej., en caso de que las secuencias 5'-UTR de dichas primera y segunda secuencias intrónicas sean las mismas. La evaluación de si se forman diferentes transcritos se puede hacer usando cualquier método adecuado conocido por la persona experta en la técnica, tal como, entre otros, la amplificación rápida de extremos de ADNc con reacción en cadena de polimerasa (RACE-PCR).

[0039] Preferiblemente, dentro de dicho primer aspecto, dicha primera secuencia intrónica en el sitio 5' o aguas arriba de dicho segundo promotor, tiene al menos un 80 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con los nucleótidos 970-1449 de la SEC ID n.º: 1 o al menos un 80 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con los nucleótidos 667-1228 de SEC ID n.º: 2, que comprende preferiblemente al menos un sitio de empalme donador o sitio de empalme GT.

[0040] Preferiblemente, dentro de dicho primer aspecto, dicha secuencia intrónica aguas abajo o en el sitio 3 'de

dicho segundo promotor comprende una secuencia de nucleótidos que tiene al menos un 80 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con los nucleótidos 171-277 de la SEC ID n.º: 14, los nucleótidos 171-274 de la SEC ID n.º: 19, los nucleótidos 133-210 de la SEC ID n.º: 20, los nucleótidos 134-211 de la SEC ID n.º: 21, los nucleótidos 134-226 de la SEC ID n.º: 22, los nucleótidos 134-226 de la SEC ID n.º: 23, los nucleótidos 133-225 de la SEC ID n.º: 24, los nucleótidos 134-226 de la SEC ID n.º: 25, los nucleótidos 146-257 de la SEC ID n.º: 26 o los nucleótidos 147-223 de la SEC ID n.º: 27, que comprende preferiblemente al menos un sitio de empalme aceptor AG precedido por una secuencia de nucleótidos rica en TC, opcionalmente separada del sitio de empalme AG por 1-50 nucleótidos y un sitio de ramificación que comprende la secuencia Y-T-N-A-Y o C-Y-G-A-C, en el sitio 5' de la secuencia de nucleótidos rica en TC.

[0041] En una realización preferida dentro de dicho primer aspecto, dicho primer promotor está flanqueado en su sitio 3' por dicha primera secuencia intrónica. En una realización, dicho primer promotor y dicha primera secuencia intrónica no están alineadas en la naturaleza sino alineadas en un constructo de la invención mediante recombinación. En otra realización, dicha secuencia que comprende dicho primer promotor flanqueado en su sitio 3' por dicha primera secuencia intrónica se deriva de una secuencia de origen natural. En una realización preferida, dicha secuencia de nucleótidos que comprende tanto un primer promotor como una primera secuencia intrónica según la presente invención es una secuencia derivada del gen de ubiquitina UBC. Preferiblemente, dicha secuencia se deriva de un gen de ubiquitina UBC de mamífero. Más preferiblemente, dicha secuencia de nucleótidos que comprende tanto un primer promotor como una primera secuencia intrónica según la presente invención se deriva del gen *Cricetulus griseus* homólogo del gen de ubiquitina UBC humana, dicho gen se indica como gen *Cricetulus* sp. para poliubiquitina o CRUPUQ (GenBank D63782). En una realización preferida, dicha secuencia de nucleótidos derivada de CRUPUQ comprende tanto un primer promotor como una primera secuencia intrónica de la invención y es una secuencia contigua de al menos 500, 600, 700, 800, 900, 1000 o 1117 de longitud, preferiblemente al menos 1449 nucleótidos de longitud de la SEC ID n.º: 1. Preferiblemente, dicha secuencia de nucleótidos tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 1 tiene como máximo 8000 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 de longitud. Más preferiblemente, dicha secuencia tiene 1449 nucleótidos de longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 65 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 70% de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 75% de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 80% de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 85% de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 90% de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 95% de identidad con la SEC ID n.º: 1 en toda su longitud. También se prefiere una secuencia de, como máximo, 8000 nucleótidos que tengan al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 1.

[0042] En una realización preferida adicional dentro de dicho primer aspecto, dicha secuencia de nucleótidos que comprende tanto un primer promotor como una primera secuencia intrónica según la presente invención es una secuencia derivada del gen de ubiquitina UBC. Preferiblemente, dicha secuencia se deriva de un gen CCT8 de mamífero. Más preferiblemente, dicha secuencia de nucleótidos que comprende tanto un primer promotor como una primera secuencia intrónica según la presente invención se deriva del gen CCT8 humano u homo sapiens. En una realización preferida, dicha secuencia de nucleótidos derivada de dicho gen CCT8 que comprende tanto un primer promotor como una primera secuencia intrónica de la invención es una secuencia contigua de al menos 500, 600, 700, 791 o 1223 de longitud, preferiblemente al menos 1228 nucleótidos de longitud de la SEC ID n.º: 2. Preferiblemente, dicha secuencia de nucleótidos tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 2 tiene como máximo 8000 nucleótidos de longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 de longitud. Más preferiblemente, dicha secuencia tiene 1228 nucleótidos de longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 65 % de identidad con la SEC

ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 70% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 75% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 80% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 85% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 90% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 95% de identidad con la SEC ID n.º: 2 en toda su longitud. También se prefiere una secuencia de, como máximo, 8000 nucleótidos que tengan al menos un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 2 en toda su longitud.

[0043] Preferiblemente dentro de dicho primer aspecto, dicha secuencia de nucleótidos que comprende un primer promotor y una primera secuencia intrónica tal como se define en el presente documento tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con cualquiera de SEC ID n.º: 1, 2 y 59-61 en toda su longitud. Preferiblemente, dicha secuencia de nucleótidos que comprende un primer promotor y una primera secuencia intrónica tal como se define en el presente documento comprende o consta de cualquiera de las secuencias seleccionadas del grupo que consta de las SEC ID n.º: 1, 2 y 59-61. Más preferiblemente, dicha secuencia de nucleótidos que comprende un primer promotor y una primera secuencia intrónica tal como se define en el presente documento comprende o consta de cualquiera de las secuencias seleccionadas del grupo que consta de las SEC ID n.º: 1, 2 y 59.

[0044] Preferiblemente, el constructo de nucleótidos del primer aspecto comprende además una o más secuencias reguladoras de la expresión adicionales, en donde preferiblemente dicho primer promotor, dichas secuencias intrónicas tal como se definen en el presente documento y, opcionalmente, dicha secuencia reguladora adicional de la expresión están todos configurados para estar operativamente enlazados a una secuencia de nucleótidos opcional de interés. Una «secuencia reguladora adicional de la expresión» debe entenderse en el presente documento como una secuencia o elemento además del primer y/o segundo promotor y/o la primera y/o segunda secuencia intrónica tal como se ha definido anteriormente en el presente documento, y puede ser una secuencia de mejora adicional de la expresión y/o una secuencia que mejora la expresión distinta. Una secuencia reguladora adicional de la expresión que abarca la presente invención puede ser, por ejemplo, una regulación transcripcional y/o traduccional de un gen, que incluye, entre otras, una 5'-UTR, una 3'-UTR, un potenciador, un promotor, un intrón, una señal de poliadenilación y elementos de control de cromatina como S/MAR (región de asociación a andamiaje/matriz), un elemento de apertura de cromatina ubicuo, una isla de guanina fosfodiéster citosina y un STAR (elemento estabilizador y antirrepresor), y cualquier derivado de los mismos. Otras secuencias reguladoras opcionales que pueden estar presentes en el constructo de ácidos nucleicos de la invención incluyen, entre otras, secuencias codificantes de nucleótidos de secuencias de nucleótidos homólogas y/o heterólogas, que incluyen el elemento sensible al hierro (IRE), elemento traslacional cis-regulador (TLRE) o uORF en las 5'-UTR y tramos de poli(U) en las 3'-UTR. Dichos elemento o elementos reguladores de la expresión adicionales, preferiblemente potenciadores, pueden ubicarse en cualquier posición en el constructo, preferiblemente alineándose directamente o comprendidos dentro de dichos primer promotor y/o segundo promotor.

[0045] Una secuencia reguladora preferida adicional dentro de dicho primer aspecto comprende o consta de un elemento potenciador de la traducción. Preferiblemente, un elemento potenciador de la traducción permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la expresión de dicha proteína o polipéptido que usa un constructo que solo difiere en que está libre de dicho elemento potenciador de la traducción, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, preferiblemente la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la traducción que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión está libre de dicho elemento potenciador de la traducción que se va a ensayar.

[0046] Preferiblemente, dentro de dicho primer aspecto, dicho elemento potenciador de la traducción comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con cualquiera de las SEC ID n.º: 3-51 en toda su longitud, o un elemento potenciador de la traducción que comprende o consta de una secuencia de nucleótidos que comprende:

i) una secuencia de nucleótidos repetida GAA, una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T;

ii) una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T, dicha primera secuencia de nucleótidos no comprende una secuencia de nucleótidos repetida GAA; o,

iii) una secuencia de nucleótidos repetida GAA, una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T, en donde dicha secuencia de nucleótidos repetida GAA se encuentra en 3' de cualquiera de dichas secuencias de nucleótidos ricas en TC, secuencias de nucleótidos ricas en A y/o secuencias de nucleótidos ricas en GT.

[0047] La secuencia de nucleótidos repetida GAA se define en el presente documento de manera que comprende al menos 3 repeticiones GAA. La secuencia de nucleótidos repetida GAA puede comprender una repetición GAA imperfecta. La secuencia de nucleótidos repetida GAA puede tener al menos un 50 % de identidad de secuencia o al menos un 60 % de identidad de secuencia o al menos un 70 % de identidad de secuencia o al menos un 80 % de identidad de secuencia o al menos un 90 % de identidad de secuencia o un 100% de identidad de secuencia con los nucleótidos 14-50 de la SEC ID n.º: 3. La repetición imperfecta de GAA puede comprender la secuencia de nucleótidos (GAA)₃ATAA (GAA)₈.

[0048] La secuencia de nucleótidos rica en TC se define en el presente documento de modo que tiene al menos un 70 %, un 80 %, un 90 % o un 100 % de identidad de secuencia con los nucleótidos 54-68 de la SEC ID n.º: 3.

[0049] La secuencia de nucleótidos rica en A se define en el presente documento de modo que tiene al menos una identidad de secuencia de un 70 %, un 80 %, un 90 % o un 100 % con cualquiera de los nucleótidos 77-87, los nucleótidos 93-105, los nucleótidos 111-121, los nucleótidos 126-132 o los nucleótidos 152-169 de la SEC ID n.º: 3, respectivamente.

[0050] La secuencia de nucleótidos rica en GT se define en el presente documento de modo que tiene al menos un 70 %, un 80 %, un 90 % o un 100 % de identidad de secuencia con los nucleótidos 133-148 de la SEC ID n.º: 3.

[0051] Preferiblemente, dentro de dicho primer aspecto, dicha secuencia potenciadora de la traducción comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con cualquiera de las SEC ID n.º: 19. Preferiblemente, dicha secuencia potenciadora de la traducción se encuentra aguas abajo o en el sitio 3' de la segunda secuencia promotora de la invención y aguas arriba o en el sitio 5' de una secuencia de nucleótidos opcional que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha secuencia potenciadora de la traducción está ubicada aguas abajo o en el sitio 3' de la segunda secuencia promotora de la invención y aguas arriba o en el sitio 5' de la segunda secuencia intrónica tal como se define en el presente documento.

[0052] Más preferiblemente dentro de dicho primer aspecto, dicho constructo de ácidos nucleicos del primer aspecto de la invención que comprende un primer promotor, una primera secuencia intrónica, un segundo promotor y una segunda secuencia intrónica, tiene al menos un 45%, un 50%, un 55%, un 60%, un 65%, un 70%, un 75%, un 80%, un 85%, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con las SEC ID n.º: 73, 74, 75 o 76.

[0053] Preferiblemente dentro de dicho primer aspecto, el constructo de ácidos nucleicos de la invención comprende además una secuencia de nucleótidos de interés operativamente enlazada y/o bajo el control de dichos primer promotor y segundo promotor y opcionalmente dicha secuencia reguladora adicional de la expresión tal como se define en el presente documento. Debe entenderse que dichos primer promotor y segundo promotor y opcionalmente dicha secuencia reguladora adicional de la expresión, están configurados para estar operativamente enlazados a la misma secuencia de nucleótidos sencilla de interés. En una realización preferida, dicha secuencia de nucleótidos de interés es una secuencia de nucleótidos que codifica una proteína o un

polipéptido de interés. La proteína o el polipéptido de interés pueden ser una proteína o un polipéptido homólogos, pero en una realización preferida de la invención, la proteína o el polipéptido de interés son una proteína o un polipéptido heterólogos. Una secuencia de nucleótidos que codifica una proteína o un polipéptido heterólogos puede derivarse total o parcialmente de cualquier fuente conocida en la técnica, incluyendo un genoma o episoma bacteriano o viral, ADN eucariótico nuclear o de plásmido, ADNc o ADN sintetizado químicamente. La secuencia de nucleótidos que codifica una proteína o un polipéptido de interés puede constituir una región codificante ininterrumpida o puede incluir uno o más intrones unidos por uniones de empalme apropiadas, puede además estar compuesta de segmentos derivados de diferentes fuentes naturales o sintéticas. La secuencia de nucleótidos que codifica la proteína o el polipéptido de interés según el método de la invención es preferiblemente una secuencia de nucleótidos de longitud completa, pero también puede ser una parte funcionalmente activa u otra parte de dicha secuencia de nucleótidos de longitud completa. La secuencia de nucleótidos que codifica la proteína o el polipéptido de interés también puede comprender secuencias de señal que dirigen la proteína o polipéptido de interés cuando se expresa en una ubicación específica de la célula o tejido. Además, la secuencia de nucleótidos que codifica la proteína o el polipéptido de interés también puede comprender secuencias que facilitan la purificación de proteínas y la detección de proteínas mediante, por ejemplo, transferencia de Western y ELISA (por ejemplo, secuencias de c-myc o polihistidina).

[0054] Dentro del contexto de la invención, la proteína o el polipéptido de interés pueden tener aplicaciones industriales o médicas (farmacéuticas). Los ejemplos de proteínas o polipéptidos con aplicaciones industriales incluyen enzimas tales como, p. ej., lipasas (p. ej., utilizadas en la industria de los detergentes), proteasas (utilizadas, entre otras cosas, en la industria de los detergentes, en la elaboración de cerveza y similares), enzimas que degradan la pared celular (tales como celulasas, pectinasas, beta-1,3/4-beta-1,6-glucanasas, ramnogalacturonasas, mananasas, xilanasas, pululanases, galactanasas, esterasas y similares, utilizadas en el procesamiento de frutas, en fabricación de vino y similares o en piensos), fitasas, fosfolipasas, glucosidasas (tales como amilasas, beta-glucosidasas, arabinofuranosidasas, ramnosidasas, apiosidasas y similares) y enzimas lácteas (por ejemplo, quimosina). Las proteínas o polipéptidos y/o enzimas de mamíferos, y preferiblemente humanos, con aplicaciones terapéuticas, cosméticas o de diagnóstico incluyen, entre otros, insulina, albúmina sérica (HSA), lactoferrina, hemoglobina α y β , activador del plasminógeno tisular (tPA), eritropoyetina (EPO), factores de necrosis tumoral (TNF), BMP (proteína morfogenética ósea), factores de crecimiento (G-CSF, GM-CSF, M-CSF, PDGF, EGF y similares), hormonas peptídicas (por ejemplo, calcitonina, somatomedina, somatotropina, hormonas de crecimiento, hormona foliculoestimulante (FSH), interleucinas (IL-x), interferones (IFN-y), fosfatasas, anticuerpos y proteínas similares a los anticuerpos, como, entre otros, anticuerpos multispecíficos como DART (reorientación de doble afinidad) y proteína tricuerpo y fragmentos de anticuerpos como Fc, Fab, Fab2, Fv y scFv. También se incluyen antígenos bacterianos y víricos, por ejemplo, para su uso como vacunas, que incluyen, por ejemplo, la subunidad β de la toxina termolábil, la subunidad β de la toxina del cólera, la proteína de la superficie de la envoltura del virus de la hepatitis B, la proteína de la cápside del virus de Norwalk, la glucoproteína B de citomegalovirus humano, la glucoproteína S, interferones y receptores de coronavirus de gastroenteritis transmisible y similares. Además se incluyen genes que codifican mutantes o análogos de dichas proteínas.

[0055] Dentro del contexto de la invención, en una realización alternativa, dicha secuencia de nucleótidos de interés no es una secuencia codificante de una proteína o un polipéptido, sino que puede ser una secuencia de nucleótidos funcional tal como, entre otras, una secuencia que codifica un ARN no codificante, en donde se entiende que un ARN no codificante es un ARN que no codifica una proteína o polipéptido. Preferiblemente, dicho ARN no codificante es una secuencia de referencia o molécula reguladora que puede regular la expresión de genes o regular la actividad o localización de proteínas o polipéptidos. Por ejemplo, un ARN no codificante puede ser una molécula de ARN o ARNm antisentido. Puesto que se cree que el primer promotor y el segundo promotor de la invención funcionan a nivel de transcripción, es decir, se cree que el aumento de la expresión mediante la secuencia de la invención que comprende dichos primer promotor y segundo promotor tal como se muestra en el presente documento es el resultado de un aumento en la transcripción, el constructo de la invención también puede usarse para producir niveles aumentados de transcritos, así como para producir transcritos con diferentes secuencias. Los niveles de transcripción pueden cuantificarse usando métodos de cuantificación de transcripción regulares conocidos por la persona experta en la técnica, tales como, entre otros, transferencia de Northern y RT-qPCR.

Segundo aspecto

[0056] En un segundo aspecto, la presente invención proporciona un vector de expresión que comprende un constructo de ácidos nucleicos según el primer aspecto de la invención. El constructo de ácidos nucleicos según la invención es preferiblemente un vector, en particular un plásmido, cósmido o fago o secuencia de nucleótidos, lineal o circular, de un ADN o ARN monocatenario o bicatenario, derivado de cualquier fuente, en el que varias secuencias de nucleótidos se han unido o recombinado en un constructo único capaz de introducir cualquiera de las secuencias de nucleótidos de la invención en una célula en las orientaciones sentido o antisentido. La elección del vector depende de los procedimientos recombinantes seguidos y de la célula huésped utilizada. El vector puede ser un vector de replicación autónoma o puede replicarse junto con el cromosoma en el que se ha integrado. Preferiblemente, el vector contiene un marcador de selección. Los marcadores útiles dependen de la

célula huésped elegida, son bien conocidos por los expertos en la técnica y se seleccionan de, entre otros, los marcadores de selección tal como se define en el tercer aspecto de la invención. Un vector de expresión preferido es el vector de expresión ADNpc3.1. Los marcadores de selección preferidos son el gen de resistencia a la neomicina, el gen de resistencia a la zeocina y el gen de resistencia a la blastidina.

Tercer aspecto

[0057] En un tercer aspecto, la presente invención proporciona una célula que comprende un constructo de ácidos nucleicos según el primer aspecto de la invención y/o un vector de expresión según el segundo aspecto de la invención tal como se define aquí.

[0058] Dentro del contexto de la invención, una célula puede ser una célula de mamífero, incluyendo una célula humana, una célula de planta, una célula de animal, una célula de insecto, una célula de hongo, una célula de levadura o una célula bacteriana. Una célula huésped recombinante, por ejemplo, una célula de mamífero, incluyendo humanos, de planta, de animal, de insecto, de hongo o bacteriana, que contiene una o más copias de un constructo de ácidos nucleicos según la invención es un tema adicional de la invención. Por célula huésped se entiende una célula que contiene un constructo de ácidos nucleicos, por ejemplo, un vector y respalda la replicación y/o expresión del constructo de ácidos nucleicos. Ejemplos de bacterias adecuadas son bacterias gram positivas, tal como varias especies de los géneros *Bacillus*, *Streptomyces* y *Staphylococcus* o bacterias Gram negativas, tales como varias especies de los géneros *Escherichia* y *Pseudomonas*. Las células fúngicas incluyen células de levadura. La expresión en la levadura se puede lograr mediante el uso de cepas de levadura tales como *Pichia pastoris*, *Saccharomyces cerevisiae* y *Hansenula polymorpha*. Otras células fúngicas de interés incluyen células de hongos filamentosos como *Aspergillus niger*, *Trichoderma reesei* y similares. Además, las células de insecto tales como las células o líneas celulares de *Drosophila melanogaster*, *Spodoptera frugiperda* y *Trichoplusia ni*, tales como, entre otras, células S2, Sf9, Sf21 y High Five, pueden usarse como células huésped. Alternativamente, un sistema de expresión adecuado puede ser un sistema de baculovirus o sistemas de expresión que usan células de mamífero tales como células CHO, COS, CPK (riñón porcino), MDCK, BHK, Sp2/0, NS0 y Vero. Una célula humana o línea celular humana adecuada es una célula o línea celular de astrocito, adipocito, condrocito, endotelial, epitelial, fibroblastos, cabello, queratinocitos, melanocitos, osteoblastos, músculo esquelético, músculo liso, célula madre o sinoviocito. Los ejemplos de líneas celulares humanas adecuadas también incluyen HEK 293 (riñón embrionario humano), HeLa, Per.C6, CAP (líneas celulares derivadas de amniocitos humanos primarios) y células de melanoma de Bowes. En una realización, una célula humana no es una célula madre embrionaria.

[0059] Por lo tanto, otro aspecto de la invención se refiere a una célula huésped que está modificada genéticamente, preferiblemente mediante un método de la invención, en el sentido de que una célula huésped comprende un constructo de ácidos nucleicos tal como se ha definido en el presente documento anteriormente. La célula huésped es una célula que ha sido modificada genéticamente. La célula huésped de expresión puede reemplazarse por una célula modificada o una célula transformada o una célula recombinante o una célula huésped modificada o una célula huésped transformada o una célula huésped recombinante. Para los procedimientos de transformación en plantas, las bacterias adecuadas incluyen *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes*.

[0060] Un constructo de ácidos nucleicos preferiblemente se mantiene de manera estable, ya sea como un elemento de replicación autónoma, o, más preferiblemente, el constructo de ácidos nucleicos se integra en el genoma de la célula huésped, en cuyo caso el constructo generalmente se integra en posiciones aleatorias en el genoma de la célula huésped, por ejemplo, mediante recombinación no homóloga. Las células huésped transformadas de forma estable se producen mediante métodos conocidos. El término «transformación estable» se refiere a exponer las células a métodos para transferir e incorporar ADN extraño en su genoma. Estos métodos incluyen, entre otros, la transferencia de ADN purificado a través de reactivos lipídicos catiónicos y polietilenimida (PEI), coprecipitación de fosfato de calcio, bombardeo de micropartículas, electroporación de protoplastos y microinyección o uso de fibras de silicio para facilitar la penetración y transferencia de ADN en la célula huésped.

[0061] Alternativamente, una proteína o un polipéptido de interés pueden expresarse en una célula huésped, por ejemplo, una célula de mamífero, apoyándose en la expresión transitoria de los vectores.

[0062] Un constructo de ácidos nucleicos según la invención también comprende preferiblemente un gen marcador que puede proporcionar capacidad de selección o cribado en una célula huésped tratada. Los marcadores seleccionables generalmente se prefieren para los eventos de transformación del huésped, pero no están disponibles para todas las células del huésped. Un constructo de ácidos nucleicos descrito en el presente documento también puede incluir una secuencia de nucleótidos que codifica un producto marcador. Se puede usar un producto marcador para determinar si el constructo o parte del mismo se ha suministrado a la célula y una vez suministrado se está expresando. Los ejemplos de genes marcadores incluyen, entre otros, el gen lacZ de *E. coli*, que codifica la β -galactosidasa, y un gen que codifica la proteína verde fluorescente.

[0063] Dentro del contexto de la invención, los ejemplos de marcadores seleccionables adecuados para células de mamífero incluyen, entre otros, dihidrofolato reductasa (DHFR), glutatión sintetasa (GS), timidina quinasa, neomicina, análogo de neomicina G418, higromicina, blastidicina, zeocina y puromicina.

[0064] Otros marcadores seleccionables adecuados incluyen, entre otros, genes resistentes a antibióticos, genes metabólicos, genes auxotróficos o genes resistentes a herbicidas que, cuando se insertan en una célula huésped en cultivo, conferirían a esas células la capacidad de resistir la exposición a un antibiótico. Los genes marcadores metabólicos o auxotróficos permiten que las células transformadas sinteticen un componente esencial, generalmente un aminoácido, que permite que las células crezcan en medios que carecen de este componente. Otro tipo de gen marcador es uno que puede cribarse mediante análisis histoquímico o bioquímico, incluso aunque el gen no pueda seleccionarse. Un gen marcador adecuado que es útil en dicha experiencia de transformación de células huésped es un gen de luciferasa. La luciferasa cataliza la oxidación de la luciferina, lo que da lugar a la producción de oxiluciferina y luz. Por lo tanto, el uso de un gen de luciferasa proporciona un ensayo conveniente para la detección de la expresión de ADN introducido en células huésped mediante análisis histoquímico de las células. En un ejemplo de un proceso de transformación, una secuencia de nucleótidos que intenta expresarse en una célula huésped puede acoplarse en tándem con el gen de la luciferasa. El constructo en tándem podría transformarse en células huésped, y las células huésped resultantes podrían analizarse para la expresión de la enzima luciferasa. Una ventaja de este marcador es el procedimiento no destructivo de aplicación del sustrato y la posterior detección.

[0065] Cuando dichos marcadores seleccionables se transfieren con éxito a una célula huésped, la célula huésped transformada puede sobrevivir si se coloca bajo presión selectiva. Hay dos categorías distintas de regímenes selectivos ampliamente utilizados. La primera categoría se basa en el metabolismo de una célula y el uso de una línea celular mutante que carece de la capacidad de crecer de manera independiente de un medio suplementado. Dos ejemplos no limitantes son las células CHO DHFR y las células LTK de ratón. Estas células carecen de la capacidad de crecer sin la adición de nutrientes tales como la timidina o la hipoxantina. Debido a que estas células carecen de ciertos genes necesarios para una ruta completa de síntesis de nucleótidos, no pueden sobrevivir a menos que los nucleótidos ausentes se proporcionen mediante un medio suplementado. Una alternativa para suplementar los medios es introducir un gen DHFR o TK intacto en las células que carecen de los genes respectivos, alterando así sus requisitos de crecimiento. Las células individuales que no se transformaron con el gen DHFR o TK no serán capaces de sobrevivir en medios no suplementados.

[0066] La segunda categoría es la selección dominante que se refiere a un esquema de selección utilizado en cualquier tipo de célula y no requiere el uso de una línea celular mutante. Estos esquemas suelen usar un medicamento para detener el crecimiento de una célula huésped. Aquellas células que tienen un gen nuevo expresarían una proteína que transmite resistencia frente a los medicamentos y sobrevivirían a la selección. Los ejemplos de tal selección dominante usan los medicamentos neomicina, (Southern P. y Berg, P., J. Molec. Appl. Genet. 1: 327 (1982)), ácido micofenólico, (Mulligan, R.C. y Berg, P. Science 209: 1422 (1980)) o higromicina, (Sugden, B. et al., Mol. Cell. Biol. 5: 410-413 (1985)). Los tres ejemplos emplean genes bacterianos bajo control eucariótico para transmitir resistencia frente al fármaco apropiado G418 o neomicina (geneticina), xgpt (ácido micofenólico) o higromicina, respectivamente. Otros incluyen el análogo de neomicina G418 y la puromicina. Otros marcadores útiles dependen de la célula huésped elegida y son bien conocidos por los expertos en la técnica.

[0067] Cuando se obtiene una célula huésped transformada con un método según la invención (véase más adelante), se puede regenerar un tejido huésped a partir de dicha célula transformada en un medio adecuado, que opcionalmente puede contener antibióticos o biocidas conocidos en la técnica para la selección de células transformadas.

[0068] Los tejidos huésped transformados resultantes se identifican preferiblemente mediante selección usando un gen marcador de selección presente en un constructo de ácidos nucleicos tal como se define en el presente documento.

Cuarto aspecto

[0069] En un cuarto aspecto, la presente invención proporciona un método para expresar y opcionalmente purificar una proteína o un polipéptido de interés que comprende el paso de:

- a) proporcionar un constructo de ácidos nucleicos según el primer aspecto de la invención que comprende una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada exprese la proteína o polipéptido de interés; y opcionalmente
- d) purificar dicha proteína o polipéptido de interés.

[0070] El método de la invención puede ser un método *in vitro* o *ex vivo*. El método de la invención se puede aplicar en un cultivo celular, cultivo de organismos o cultivo de tejidos. Alternativamente, junto a la expresión en células huésped, la proteína o el polipéptido de interés pueden producirse en sistemas de traducción sin células usando ARN derivados de los constructos de ácidos nucleicos de la presente invención. El método de la invención puede ejecutarse en células cultivadas.

[0071] El experto en la técnica es capaz de transformar células según el paso b). Los métodos de transformación que se usan en el paso b) incluyen, entre otros, la transferencia de ADN purificado a través de reactivos lipídicos catiónicos y polietilenimida (PEI), coprecipitación de fosfato de calcio, bombardeo de micropartículas, electroporación de protoplastos y microinyección o uso de fibras de silicio para facilitar la penetración y transferencia de ADN en la célula huésped.

[0072] En el paso c) se permite que la célula transformada exprese la proteína o el polipéptido de interés, y opcionalmente dicha proteína o polipéptido se recupera a continuación. Por ejemplo, la célula transformada puede someterse a condiciones que producen la expresión de la proteína o el polipéptido de interés. Los expertos en la técnica conocen bien las técnicas que se van a utilizar para expresar o sobreexpresar la proteína o el polipéptido de interés. Los métodos en los que la célula transformada no necesita someterse a condiciones específicas que producen la expresión de la proteína o el polipéptido de interés pero en los que la proteína o el polipéptido de interés se expresan automáticamente (por ejemplo, constitutivamente), también se incluyen en el método de la presente invención.

[0073] Dentro del contexto de la invención, los pasos de purificación dependen de la proteína o el polipéptido expresados y la célula huésped utilizada, pero pueden comprender el aislamiento de la proteína o el polipéptido. Cuando se aplica a una proteína/polipéptido, el término «aislamiento» indica que la proteína o el polipéptido se encuentran en una condición diferente a la de su entorno nativo. En una forma preferida, la proteína o el polipéptido aislado está sustancialmente libre de otras proteínas, particularmente otras proteínas homólogas. Se prefiere proporcionar la proteína o el polipéptido en una forma con una pureza superior al 40 %, más preferiblemente en una forma con una pureza superior al 60 %. Incluso más preferiblemente, se prefiere proporcionar la proteína o el polipéptido en una forma altamente purificada, es decir, una pureza superior al 80 %, más preferiblemente una pureza superior al 95% y aún más preferiblemente una pureza superior al 99%, según lo determinado mediante SDS-PAGE. Si se desea, la secuencia de nucleótidos que codifica una proteína o un polipéptido de interés se puede ligar a una secuencia de nucleótidos heteróloga para codificar una proteína o un polipéptido de fusión para facilitar la purificación de proteínas y la detección de proteínas, por ejemplo, en transferencia Western y ELISA. Las secuencias heterólogas adecuadas incluyen, entre otras, las secuencias de nucleótidos que codifican proteínas tales como, por ejemplo, glutatión-S-transferasa, proteína de unión a maltosa, polihistidina de unión a metal, proteína fluorescente verde, luciferasa y beta-galactosidasa. La proteína o el polipéptido también se pueden acoplar a vehículos, etiquetas o marcadores no peptídicos que facilitan el rastreo de la proteína o el polipéptido, tanto *in vivo* como *in vitro*, y permiten la identificación y la cuantificación de la unión de la proteína o el polipéptido a sustratos. Dichas etiquetas, marcadores o vehículos son bien conocidos en la técnica e incluyen, entre otros, biotina, marcadores radiactivos y marcadores fluorescentes.

[0074] Preferiblemente, el método de este cuarto aspecto de la invención permite un aumento en la expresión de una proteína o un polipéptido de interés. Preferiblemente, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión según el primer aspecto de la invención que comprende un primer y un segundo promotor según el primer aspecto de la invención, operativamente enlazados a una secuencia de nucleótidos que codifica una proteína o polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo ELISA, una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, el método de la invención permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50%, un 60 %, un 70%, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho primer promotor y dicho segundo promotor y está operativamente enlazado a un único promotor, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende una primera secuencia promotora y una segunda secuencia promotora que se van a ensayar operativamente enlazadas a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que dichos primer promotor y segundo promotor son reemplazados por un único promotor, preferiblemente un promotor de CMV representado por la SEC ID n.º: 57, en el vector de expresión utilizado.

Quinto aspecto

[0075] En un quinto aspecto, la presente invención proporciona un método para expresar una proteína o un polipéptido de interés en un organismo, que comprende los pasos de:

- a) proporcionar un constructo de ácidos nucleicos según el primer aspecto de la invención que comprende una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés; y,
- b) poner en contacto una célula diana y/o tejido diana de un organismo con dicho constructo de ácidos nucleicos para obtener una célula diana transformada y/o tejido diana transformado, permitiendo que dicha célula transformada exprese la proteína o el polipéptido de interés; y opcionalmente
- c) permitir que dicha célula diana transformada y/o tejido diana se desarrolle en un organismo transformado; y, opcionalmente,
- d) permitir que dicho organismo transformado exprese la proteína o el polipéptido de interés, por ejemplo, sometiendo dicho organismo transformado a condiciones que produzcan la expresión de la proteína o el polipéptido de interés, y opcionalmente recuperar dicha proteína o polipéptido.

[0076] Dentro del contexto de la invención, la célula diana puede ser una célula diana embrionaria, p. ej., una célula madre embrionaria, por ejemplo, derivada de un mamífero no humano, tal como de las especies bovina, porcina, etc. Preferiblemente, dicha célula diana no es una célula madre embrionaria humana. En el caso de un hongo multicelular, dicha célula diana puede ser una célula fúngica que puede proliferar en dicho hongo multicelular. Cuando se obtiene un tejido vegetal transformado o una célula vegetal (por ejemplo, trozos de hoja, segmentos de tallo, raíces, pero también protoplastos o células vegetales cultivadas por suspensión) con el método según la invención, se pueden regenerar plantas enteras a partir de dicho tejido o célula transformados en un medio adecuado, que opcionalmente puede contener antibióticos o biocidas conocidos en la técnica para la selección de células transformadas. El método de la invención se puede aplicar en vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano. Se incluye dentro de la presente invención un método de tratamiento que comprende el método del presente aspecto, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico. La invención también se refiere a un constructo del primer aspecto de la invención para el tratamiento, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico. Además, la invención también se refiere a un constructo del primer aspecto de la invención para la fabricación de un medicamento, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico.

[0077] Además, una parte de la invención es un organismo transformado no humano. Dicho organismo se transforma con una secuencia de nucleótidos, un constructo de ácidos nucleicos recombinante o un vector según la presente invención, y es capaz de producir el polipéptido de interés. Esto incluye un organismo transgénico no humano, tal como un mamífero transgénico no humano, una planta transgénica (incluyendo la propagación, la cosecha y el material de tejido de dicha planta transgénica, que incluye, entre otros, hojas, raíces, brotes y flores), un hongo multicelular y similares.

[0078] Preferiblemente, el método de este quinto aspecto de la invención permite un aumento en la expresión de una proteína o un polipéptido de interés en dicho organismo o al menos en un tejido u órgano u órgano de dicho organismo. Preferiblemente, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión según el primer aspecto de la invención que comprende un primer y un segundo promotor según el primer aspecto de la invención, operativamente enlazados a una secuencia de nucleótidos que codifica una proteína o polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo ELISA, una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, el método de la invención permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en dicho organismo o al menos en un tejido u órgano u órgano de dicho organismo en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho primer promotor y dicho segundo promotor y está operativamente enlazado a un único promotor, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un primera secuencia promotora y una segunda secuencia promotora que se van a ensayar operativamente enlazadas a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que dichos primer promotor y segundo promotor son reemplazados por un único promotor, preferiblemente un promotor de CMV representado por la

SEC ID n.º: 57, en el vector de expresión utilizado. Preferiblemente, dicho aumento de la expresión de proteínas o polipéptidos es independiente del número de copias según lo establecido por un ensayo adecuado para determinar la dependencia del número de copias realizado por una persona experta tal como, por ejemplo, un ensayo Taqman triplex tal como se detalla adicionalmente en el ejemplo 4 de la presente invención.

Sexto aspecto

[0079] En un sexto aspecto, la presente invención proporciona un método para la transcripción y opcionalmente la purificación del transcrito producido que comprende el paso de:

- a) proporcionar un constructo de ácidos nucleicos según el primer aspecto de la invención que comprende una secuencia de nucleótidos de interés; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada produzca un transcrito de la secuencia de nucleótidos de interés; y opcionalmente
- d) purificar dicho transcrito producido.

[0080] En una realización preferida del método según la invención, se usa un constructo de ácidos nucleicos tal como se ha definido anteriormente. El método de la invención puede ser un método *in vitro* o *ex vivo*. El método de la invención se puede aplicar en un cultivo celular, cultivo de organismos o cultivo de tejidos. El método de la invención se puede aplicar en vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, preferiblemente en un ser humano. Se incluye dentro de la presente invención un método para el tratamiento que comprende o consta del método del presente aspecto, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico. La invención también se refiere a un constructo del primer aspecto de la invención para su uso en el tratamiento, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico. Además, la invención también se refiere a un constructo del primer aspecto de la invención para la fabricación de un medicamento, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico.

[0081] El experto en la técnica es capaz de transformar células según el paso b). Los métodos de transformación que se usan en el paso b) incluyen, entre otros, la transferencia de ADN purificado a través de reactivos lipídicos catiónicos y polietilenimida (PEI), coprecipitación de fosfato de calcio, bombardeo de micropartículas, electroporación de protoplastos y microinyección o uso de fibras de silicio para facilitar la penetración y transferencia de ADN en la célula huésped.

[0082] En el paso c) se permite que la célula transformada produzca un transcrito de la secuencia de nucleótidos de interés y opcionalmente el transcrito producido se recupera posteriormente. Por ejemplo, la célula transformada puede someterse a condiciones que producen la transcripción de la secuencia de nucleótidos de interés. Los expertos en la técnica conocen bien las técnicas que se van a utilizar para la transcripción de la secuencia de nucleótidos de interés. Los métodos en los que la célula transformada no necesita someterse a condiciones específicas que producen la transcripción de la secuencia de nucleótidos de interés, pero en los que la secuencia de nucleótidos de interés se transcribe automáticamente (por ejemplo, constitutivamente), también se incluyen en el método de la presente invención.

[0083] Los pasos de purificación dependen del transcrito producido. El término «aislamiento» indica que el transcrito se encuentra en una condición distinta de su entorno nativo. En una forma preferida, el transcrito aislado está sustancialmente libre de otros componentes celulares, particularmente otros componentes celulares homólogos tales como proteínas homólogas. Se prefiere proporcionar el transcrito en una forma con una pureza superior al 40 %, más preferiblemente en una forma con una pureza superior al 60 %. Incluso más preferiblemente, se prefiere proporcionar el transcrito en una forma altamente purificada, es decir, con una pureza superior al 80 %, más preferiblemente con una pureza superior al 90 % e incluso más preferiblemente con una pureza superior al 99 %, según se determina mediante transferencia Northern.

[0084] Preferiblemente, el método de este aspecto de la invención permite un aumento en la transcripción de una secuencia de nucleótidos de interés. Preferiblemente, los niveles de transcripción se establecen en un sistema de expresión usando un constructo de expresión según el primer aspecto de la invención que comprende un primer y un segundo promotor según el primer aspecto de la invención, operativamente enlazados a una secuencia de nucleótidos que codifica una proteína o polipéptido de interés. Preferiblemente, la transcripción de dicha secuencia de nucleótidos de interés se detecta mediante un ensayo adecuado tal como RT-qPCR. Preferiblemente, el método de la invención permite un aumento en la transcripción de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho primer promotor y dicho segundo promotor y está operativamente enlazado a un único promotor, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, la transcripción de

una secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide preferiblemente en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un primera secuencia promotora y una segunda secuencia promotora que se van a ensayar operativamente enlazadas a dicha secuencia de nucleótidos de interés. La transcripción se mide preferiblemente usando RT-qPCR y los niveles de transcripción se comparan con los niveles de transcripción de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que dichos primer promotor y segundo promotor se reemplazan por un único promotor, preferiblemente un promotor de CMV representado por la SEC ID n.º: 57, en el vector de expresión utilizado.

10 Séptimo aspecto

[0085] En un séptimo aspecto, la presente divulgación proporciona un método para empalmar o redirigir el empalme de una secuencia de nucleótidos de interés, y opcionalmente purificar los transcritos producidos que comprende la etapa de:

- a) proporcionar un constructo de ácidos nucleicos según el primer aspecto de la invención que comprende una secuencia de nucleótidos de interés; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada produzca transcritos de la secuencia de nucleótidos de interés que da como resultado la producción de un transcrito; y opcionalmente
- d) purificar dichos transcritos producidos.

[0086] Preferiblemente dentro de este aspecto, dicho constructo de ácidos nucleicos usado en el paso a) comprende una secuencia de nucleótidos aguas arriba o en el sitio 5' de la segunda secuencia intrónica de la invención que es diferente o distinta de la secuencia de nucleótidos aguas arriba o en el sitio 5' de la primera secuencia intrónica de la invención. Preferiblemente, la secuencia de nucleótidos entre el primer promotor y el 5' de dicha primera secuencia intrónica y la secuencia de nucleótidos entre el segundo promotor y el 5' de dicha segunda secuencia intrónica difiere al menos en un 5 %, un 10 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 % o un 100 % en la secuencia de nucleótidos. Preferiblemente, un método de este aspecto de la divulgación en el que se usa dicho constructo de ácidos nucleicos da como resultado la producción de dos transcritos diferentes o distintos, que difieren en la secuencia de nucleótidos en el sitio 5' de los transcritos. En caso de que la secuencia de nucleótidos de interés esté ubicada aguas abajo o en el sitio 3' de la segunda secuencia intrónica, los transcritos resultantes diferirán en la secuencia aguas arriba de dicha secuencia de nucleótidos de interés tal como puede detectarse mediante cualquier ensayo adecuado conocido por la persona experta en la técnica, por ejemplo, entre otros, 5' RACE-PCR.

[0087] En una realización preferida dentro de este aspecto, la secuencia de nucleótidos de interés es una secuencia que codifica una proteína o un polipéptido de interés. El método de este aspecto se puede usar para producir dos proteínas o polipéptidos con N-terminales diferentes o distintos usando el constructo de la invención. Preferiblemente, una primera proteína o polipéptido que comprende un primer N-terminal y una segunda proteína o polipéptido que comprende un segundo N-terminal se producen usando el método de este aspecto, en el que preferiblemente, una primera secuencia de nucleótidos que codifica dicho primer N-terminal se ubica directamente aguas arriba o en el sitio 5' de dicha primera secuencia intrónica y una segunda secuencia de nucleótidos que codifica dicho segundo N-terminal está ubicada directamente aguas arriba o en el sitio 5' de dicha segunda secuencia intrónica. Preferiblemente, dicha primera secuencia de nucleótidos que codifica dicho primer N-terminal está ubicada aguas abajo o en el sitio 3' de dicho primer promotor y aguas arriba o en el sitio 5' de dicha primera secuencia intrónica. Preferiblemente, dicha segunda secuencia de nucleótidos que codifica dicho segundo N-terminal está ubicada aguas abajo o en el sitio 3' de dicho segundo promotor y aguas arriba o en el sitio 5' de dicha segunda secuencia intrónica. Preferiblemente, dicho constructo de ácidos nucleicos comprende además una secuencia de nucleótidos que codifica un C-terminal situado aguas abajo o en el sitio 3' de dicha segunda secuencia intrónica. En esta realización, se requiere que dicha segunda secuencia intrónica sea un intrón tal como se define en la sección Definición. La diferencia entre los N terminales puede estar limitada a una secuencia señal y dar como resultado proteínas o polipéptidos expresados idénticos, en donde la ubicación de las proteínas o polipéptidos puede diferir. Si se realiza en un sistema de expresión tal como se ha definido anteriormente en el presente documento, el método de esta realización preferiblemente da como resultado la producción de dos proteínas o polipéptidos de interés, en donde dicha primera proteína o polipéptido comprenderá dicho primer N-terminal enlazado a dicho C-terminal y dicha segunda proteína o polipéptido comprenderá dicho segundo N-terminal enlazado a dicho C-terminal, según puede detectarse mediante cualquier ensayo adecuado conocido por la persona experta en la técnica, tales como, entre otros, ensayo ELISA, transferencia Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por la persona experta en la técnica. Preferiblemente, dicho ensayo usado para detectar las dos proteínas o polipéptidos diferentes o distintos producidos está adaptado específicamente para distinguir entre las diferentes proteínas o polipéptidos producidos, por ejemplo, usando un anticuerpo de detección que se une específicamente al primer o al segundo N-terminal de proteínas o polipéptidos producidos.

Octavo aspecto

5 [0088] En un octavo aspecto, la presente invención proporciona un uso de un constructo de ácidos nucleicos según el primer aspecto de la invención y/o un uso de un vector de expresión según el segundo aspecto de la invención y/o un uso de una célula según el tercer aspecto de la invención para la expresión de una proteína o polipéptido de interés.

Noveno aspecto

10 [0089] En un noveno aspecto, la presente invención proporciona un constructo de ácidos nucleicos según el primer aspecto de la invención y/o un vector de expresión según el segundo aspecto de la invención y/o una célula según el tercer aspecto de la invención para usar como medicamento. La invención también se refiere a un método de tratamiento que comprende la administración de un constructo de ácidos nucleicos según el primer
15 aspecto de la invención y/o un vector de expresión según el segundo aspecto de la invención y/o una célula según el tercer aspecto de la invención, en donde preferiblemente dicha administración se realiza a un mamífero, más preferiblemente a un humano. Preferiblemente, dicho tratamiento es una vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano. Además, la invención se refiere al uso de un constructo de ácidos nucleicos según el tercer aspecto de la invención y/o el
20 uso de un vector de expresión según el segundo aspecto de la invención y/o el uso de una célula según el tercer aspecto de la invención para la preparación de un medicamento. Preferiblemente dicho medicamento es para vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano.

Décimo aspecto

[0090] En un décimo aspecto, la presente divulgación proporciona una molécula de ácidos nucleicos que está representada por una secuencia de nucleótidos que comprende o consta de un elemento potenciador de la expresión para aumentar la transcripción de una secuencia de nucleótidos de interés y/o la expresión de una
30 proteína o un polipéptido de interés. Preferiblemente, el elemento potenciador de la expresión de la divulgación es capaz de aumentar la transcripción de una secuencia de nucleótidos de interés y/o la expresión de una proteína o un polipéptido de interés. Preferiblemente en este aspecto, el elemento potenciador de la expresión de la divulgación capaz de aumentar la transcripción de una secuencia de nucleótidos de interés y/o la expresión de una proteína o un polipéptido de interés se encuentra aguas arriba o en el sitio 5' de un promotor que está
35 operativamente enlazado a dicha secuencia de nucleótidos de interés.

[0091] Preferiblemente dentro de este aspecto, los niveles de transcripción se establecen en un sistema de expresión usando un constructo de expresión que comprende dicho elemento potenciador de la expresión operativamente enlazado a una secuencia de nucleótidos de interés usando un ensayo adecuado tal como RT-qPCR. Preferiblemente, el elemento potenciador de la expresión de la divulgación permite un aumento en la
40 transcripción de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con los niveles de transcripción usando un constructo que solo difiere en que está libre de dicho elemento potenciador de la expresión, preferiblemente como se muestra en los ejemplos que están incluidos en el presente documento. Más específicamente, preferiblemente la transcripción de una secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento
45 potenciador de la expresión que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La transcripción se mide preferiblemente usando RT-qPCR y los niveles de transcripción se comparan con los niveles de transcripción de dicha secuencia de nucleótidos de interés medida en las mismas condiciones, excepto que el vector de expresión usado está libre de dicho elemento potenciador de la expresión que se va a ensayar.

[0092] Preferiblemente dentro de este aspecto, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión que comprende dicho elemento potenciador de la expresión operativamente enlazado a una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo de inmunoabsorción ligada a enzimas (ELISA), una transferencia de Western o, dependiendo de la identidad de la
60 proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, el elemento potenciador de la expresión de la divulgación permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la expresión de dicha proteína o polipéptido que usa un constructo que solo difiere en que está libre de dicho elemento potenciador de la expresión, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente
65

documento. Más específicamente, preferiblemente la expresión de una secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la expresión que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión está libre de dicho elemento potenciador de la expresión que se va a ensayar.

[0093] Preferiblemente dentro de este aspecto, dicha molécula de ácidos nucleicos es una molécula de ácidos nucleicos aislada como tal como se define en el presente documento. En una realización preferida, dicho elemento potenciador de la expresión es una secuencia que se deriva del gen de ubiquitina UBC. Preferiblemente, dicho elemento potenciador de la expresión se deriva de un gen de ubiquitina UBC de mamífero. Más preferiblemente, dicho elemento potenciador de la expresión se deriva del gen homólogo de *Cricetulus griseus* del gen de ubiquitina UBC humana, indicando dicho gen como el gen *Cricetulus* sp. para poliubiquitina o CRUPUQ (GenBank D63782).

[0094] En una realización preferida, dicho elemento potenciador de la expresión derivado de CRUPUQ comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1. Preferiblemente, dicha secuencia tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-969 de la SEC ID n.º: 1 es un promotor tal como se define en la sección Definición. Preferiblemente, dicho promotor es capaz de iniciar la transcripción de una secuencia de nucleótidos de interés y/o la expresión de un polipéptido o una proteína de interés codificados por una secuencia de nucleótidos en una célula huésped tal como se define más adelante en el presente documento. En una realización preferida adicional, dicho elemento potenciador de la expresión derivado de CRUPUQ comprende o consta de una secuencia intrónica. Se entiende que una secuencia intrónica es al menos parte de la secuencia de nucleótidos de un intrón. Preferiblemente, dicha secuencia intrónica comprende al menos un sitio de empalme donador o un sitio de empalme GT. Un sitio de empalme donador se entiende en el presente documento como un sitio de empalme que, cuando se combina con un sitio de empalme aceptor tal como se define en el presente documento, da como resultado la formación de un intrón según se define en la sección Definición. Preferiblemente, una secuencia de nucleótidos es un intrón si al menos un 2 %, un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 % o un 100 % del ARN primario pierde esta secuencia mediante empalme de ARN usando un ensayo adecuado para detectar el empalme de intrones, como por ejemplo, entre otros, la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) seguido de un análisis de tamaño o secuencia de la RT-PCR. Los sitios de empalme donadores preferidos de la divulgación son M-W-G-[corte]-G-T-R-A-G-K en caso de que la célula huésped sea una célula de mamífero, AG-[corte]-G-T-A-W-K en caso de que la célula huésped sea una célula vegetal, [corte]-G-T-A-W-G-T-T en caso de que la célula huésped sea una célula de levadura y RG-[corte]-G-T-R-A-G, en caso de que la célula huésped sea una célula de insecto. «[corte]» debe entenderse aquí como el sitio de corte específico donde se realizará el empalme. El empalme de intrones se puede evaluar funcionalmente usando un ensayo según se detalla en la sección Definición bajo «intrón». Más preferiblemente, el sitio de empalme donador comprendido dentro del elemento potenciador de la expresión de la divulgación es C-T-G-[corte]-G-T-G-A-G-G. Preferiblemente, dicha secuencia intrónica incluida dentro de dicho elemento potenciador de la expresión consiste en dicho sitio de empalme donador o sitio de empalme GT. Preferiblemente, dicha secuencia intrónica incluida dentro de dicho elemento potenciador de la expresión está libre de un sitio de empalme aceptor tal como se define más adelante en el presente documento. Preferiblemente, dicho elemento potenciador de la expresión que comprende una secuencia intrónica tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 970-1449 de la SEC ID n.º: 1. En una realización preferida, dicho elemento potenciador de la expresión comprende o consta de un promotor y una secuencia intrónica tal como se define en el presente documento. Preferiblemente, el elemento potenciador de la expresión de la divulgación tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión de la divulgación que tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 1 en toda su longitud comprende tanto un promotor como una secuencia intrónica tal como se define en el presente documento. Preferiblemente, dicho elemento potenciador de la expresión para aumentar la expresión es una secuencia contigua de al menos 500, 600, 700, 800, 900, 1000, 1100 o 1117 nucleótidos de longitud, preferiblemente al menos 1449 nucleótidos de longitud de la SEC ID n.º: 1. Preferiblemente, dicho elemento potenciador de la expresión con al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un

- 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 1 tiene como máximo 8000 nucleótidos de longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud. Más preferiblemente, dicha secuencia tiene 1449 nucleótidos de longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 65 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 70 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 75 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 80 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 85 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 90 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1449 nucleótidos de longitud y tiene al menos un 95 % de identidad con la SEC ID n.º: 1 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión comprende o consta de una secuencia que está representada por la SEC ID n.º: 1.
- [0095] En una realización preferida adicional dentro de este aspecto, dicho elemento potenciador de la expresión de la divulgación es una secuencia derivada del gen CCT8. Preferiblemente, dicho elemento se deriva de un gen CCT8 de mamífero. Más preferiblemente, dicho elemento potenciador de la expresión se deriva del gen CCT8 humano u *Homo sapiens*.
- [0096] En una realización preferida dentro de dicho aspecto, dicho elemento potenciador de la expresión derivado del gen CCT8 de *Homo sapiens* comprende o consta de una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-614 de la SEC ID n.º: 2. Preferiblemente, dicha secuencia tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 1-614 de la SEC ID n.º: 2 es un promotor tal como se define en la sección Definición. En una realización preferida adicional, dicho elemento potenciador de la expresión derivado del gen CCT8 de *Homo sapiens* comprende o consta de una secuencia intrónica. Preferiblemente, dicha secuencia intrónica es una secuencia intrónica tal como se ha definido anteriormente en el presente documento que comprende al menos un sitio de empalme donador o sitio de empalme GT tal como se ha definido anteriormente en el presente documento. Preferiblemente, dicho sitio de empalme donador tiene la secuencia M-A-R-[corte]-G-T-R-A-G-K, más preferiblemente A-A-A-[corte]-G-T-G-A-G-G. Preferiblemente, dicha secuencia intrónica incluida dentro de dicho elemento potenciador de la expresión consiste en dicho sitio de empalme donador o sitio de empalme GT. Preferiblemente, dicho elemento potenciador de la expresión que comprende una secuencia intrónica tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con los nucleótidos 667-1228 de la SEC ID n.º: 2.
- [0097] En una realización preferida dentro de dicho aspecto, dicho elemento potenciador de la expresión comprende o consta de un promotor y una secuencia intrónica tal como se define en el presente documento. Preferiblemente, el elemento potenciador de la expresión de la divulgación tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión de la divulgación que tiene al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 2 en toda su longitud comprende tanto un promotor como una secuencia intrónica tal como se define en el presente documento. Preferiblemente, dicho elemento potenciador de la expresión para aumentar la expresión es una secuencia contigua de al menos 500, 600, 700, 791 o 1223 nucleótidos de longitud, preferiblemente al menos 1228 nucleótidos de la SEC ID n.º: 2. Preferiblemente, dicho elemento potenciador de la expresión con al menos un 50 %, un 51 %, un 52 %, un 53 %, un 54 %, un 55 %, un 56 %, un 57 %, un 58 %, un 59 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 2 tiene como máximo 8000 nucleótidos de longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud. Más preferiblemente,

dicha secuencia tiene 1228 nucleótidos de longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 65 % de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 70% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 75% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 80% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 85% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 90% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión tiene como máximo 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1900, 1800, 1700, 1600, 1500 o 1228 nucleótidos de longitud y tiene al menos un 95% de identidad con la SEC ID n.º: 2 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión comprende o consta de una secuencia que está representada por la SEC ID n.º: 2.

[0098] Más preferida es una secuencia de nucleótidos que comprende un elemento potenciador de la expresión tal como se define en el presente documento que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90%, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con cualquiera de SEC ID n.º: 59-61 en toda su longitud. Preferiblemente, dicho elemento potenciador de la expresión comprende o consta de cualquiera de las secuencias seleccionadas del grupo que consiste en la SEC ID n.º: 59-61. Más preferiblemente, dicho elemento potenciador de la expresión comprende o consta de una secuencia que está representada por la SEC ID n.º: 59.

Undécimo aspecto

[0099] En un undécimo aspecto, la presente divulgación proporciona un constructo de ácidos nucleicos que comprende una molécula de ácidos nucleicos según el décimo aspecto de la divulgación. Un constructo de ácidos nucleicos de la divulgación comprende un elemento potenciador de la expresión según el décimo aspecto de la presente divulgación. Preferiblemente, dicho constructo de ácidos nucleicos es un constructo recombinante y/o aislado tal como se define en el presente documento. Preferiblemente, dicho constructo de ácidos nucleicos comprende además un promotor heterólogo, en el que preferiblemente dicho elemento potenciador de la expresión y dicho promotor heterólogo están configurados para estar los dos operativamente enlazados a una secuencia de nucleótidos opcional de interés tal como se define más adelante en el presente documento. El «promotor heterólogo» debe entenderse en el presente documento como un promotor que no está operativamente enlazado de forma natural al elemento potenciador de la expresión de la divulgación, es decir, una secuencia contigua que comprende dicho elemento potenciador de la expresión y dicho promotor heterólogo no se produce en la naturaleza como secuencias vecinas, pero puede sintetizarse como una secuencia recombinante.

[0100] Preferiblemente dentro de este aspecto, dicho promotor heterólogo está ubicado dentro de un constructo de ácidos nucleicos de la divulgación aguas abajo o en el sitio 3' del elemento potenciador de la expresión de la divulgación. Preferiblemente, dicho promotor heterólogo se encuentra dentro de un constructo de la divulgación aguas arriba o en el sitio 5' de la secuencia de nucleótidos de la divulgación que codifica una proteína o un polipéptido de interés. En una realización de la divulgación, el promotor heterólogo es un promotor capaz de iniciar la transcripción en la célula huésped elegida. Los promotores tal como se usan en el presente documento incluyen promotores específicos de tejido, preferidos de tejido, específicos del tipo de célula, inducibles y constitutivos, tal como se define en el presente documento. Los promotores heterólogos y/o las secuencias reguladoras que pueden emplearse en la expresión de polipéptidos de según la presente divulgación, preferiblemente en células de mamífero, incluyen, entre otros, el promotor de citomegalovirus (CMV) humano o murino, un promotor del virus del simio (SV40), un promotor de ubiquitina C humana o de ratón, un promotor alfa (EF1-a) del factor de elongación humano o de ratón o rata, un promotor de beta-actina de ratón o hámster o un promotor de hámster rpS21. Los elementos Tet-Off y Tet-On aguas arriba de un promotor mínimo tal como un promotor de CMV forman un ejemplo de promotor inducible de mamífero. Ejemplos de promotores adecuados de levadura y hongos son el promotor Leu2, el promotor de galactosa (Gall o Ga17), el promotor de alcohol deshidrogenasa I (ADH1), el promotor de glucoamilasa (Gla), el promotor de triosa fosfato isomerasa (TPI), el factor de elongación de traducción EF-I alfa (TEF2), el promotor de la gliceraldehído-3-fosfato deshidrogenasa (gpdA), promotor de la alcohol oxidasa (AOX1) o el promotor de la glutamato deshidrogenasa (gdhA). Un ejemplo de un promotor ubicuo fuerte para la expresión en plantas es el promotor 35S del virus del mosaico de la coliflor (CaMV). Preferiblemente, el constructo de ácidos nucleicos de la divulgación comprende un promotor heterólogo que está representado por una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un

97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 58. Más preferiblemente, el constructo de ácidos nucleicos de este aspecto de la divulgación comprende un promotor heterólogo que está representado por una secuencia que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con la SEC ID n.º: 57.

[0101] En una realización preferida adicional dentro de este aspecto, el constructo de ácidos nucleicos de la divulgación comprende además uno o varios elementos reguladores de la expresión adicionales, en donde preferiblemente dicho elemento potenciador de la expresión, dicho promotor heterólogo y dicho elemento o elementos reguladores de la expresión adicionales están configurados para estar todos operativamente enlazados a una secuencia de nucleótidos opcional de interés tal como se define más adelante en el presente documento. Un «elemento regulador de expresión adicional» debe entenderse en el presente documento como un elemento adicional al elemento potenciador de la expresión y/o promotor tal como se ha definido anteriormente en el presente documento, que puede ser un elemento potenciador de la expresión adicional o un elemento potenciador de la expresión distinto o un elemento regulador de la expresión en su sentido más amplio. Un elemento regulador de la expresión adicional como el incluido en la presente divulgación puede estar implicado en la regulación transcripcional y/o traduccional de un gen, que incluye, entre otros, una 5'-UTR, una 3'-UTR, un potenciador, un promotor, una secuencia intrónica, una señal de poliadenilación y elementos de control de cromatina tales como regiones de unión de andamiaje/matriz, un elemento de apertura de cromatina ubicuo, pares de guanina fosfodiéster citosina y elementos estabilizadores y antirrepresores, y cualquier derivado de los mismos. Otros elementos reguladores opcionales que pueden estar presentes en el constructo de ácidos nucleicos de la divulgación incluyen, entre otros, secuencias codificantes de nucleótidos de secuencias de nucleótidos homólogas y/o heterólogas, que incluyen el elemento sensible al hierro (IRE), elemento traslacional cis-regulador (TLRE) o uORF en las 5'-UTR y tramos de poli(U) en las 3'-UTR.

[0102] Preferiblemente dentro de este aspecto, dicho elemento regulador de la expresión adicional comprende o consta de una secuencia intrónica tal como se define en el presente documento. Preferiblemente, la secuencia intrónica incluida dentro del elemento regulador de la expresión adicional comprende al menos un sitio de empalme aceptor que se entiende en el presente documento que comprende el sitio de empalme AG, precedido preferiblemente por una secuencia de nucleótidos del tracto de polipirimidina, opcionalmente separada del sitio de empalme AG por 1-50 nucleótidos, y que opcionalmente comprende además un sitio de ramificación que comprende la secuencia Y-T-N-A-Y, en el sitio 5' de la secuencia de nucleótidos del tracto de polipirimidina, en donde el sitio de ramificación puede tener la secuencia de nucleótidos C-Y-G-A-C. Un sitio de empalme aceptor se entiende en el presente documento como un sitio de empalme que, cuando se combina con un sitio de empalme donador incluido dentro de un constructo, da como resultado la formación de un intrón tal como se define en la sección Definición. Preferiblemente, el sitio de empalme aceptor o el sitio de empalme AG tienen la secuencia [Y-rico]-N-Y-A-G-[corte]. Preferiblemente, el sitio de empalme aceptor o sitio de empalme AG tienen la secuencia [Y-rico]-N-Y-A-G-[corte]-R en caso de que la célula huésped sea una célula de mamífero, [Y-rico]-D-Y-A-G-[corte]-R o [Y-rico]-DY-A-G-[corte]-R-W en caso de que la célula huésped sea una célula vegetal, [Y-rico]-A-Y-A-G-[corte] en caso de que la célula huésped sea una célula de levadura y [Y-rico]-N-Y-A-G-[corte] en caso de que la célula huésped sea una célula de insecto. «[Y-rico]» debe entenderse aquí como un tracto de polipirimidina que se define preferiblemente como una secuencia consecutiva de al menos 10 nucleótidos que comprende al menos 6, 7, 8, 9 o preferiblemente 10 nucleótidos de pirimidina. Preferiblemente, dicho sitio de empalme aceptor o sitio de empalme GT tiene la secuencia Y-A-G-[corte]-R. Preferiblemente, dicha secuencia intrónica incluida dentro de dicho elemento regulador de expresión adicional consta de dicho sitio de empalme aceptor o sitio de empalme AG. La secuencia intrónica preferiblemente comprende o consta de una secuencia de nucleótidos que tiene al menos un 80 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con los nucleótidos 171-277 de la SEC ID n.º: 14, los nucleótidos 171-274 de la SEC ID n.º: 19, los nucleótidos 133-210 de la SEC ID n.º: 20, los nucleótidos 134-211 de la SEC ID n.º: 21, los nucleótidos 134-226 de la SEC ID n.º: 22, los nucleótidos 134-226 de la SEC ID n.º: 23, los nucleótidos 133-225 de la SEC ID n.º: 24, los nucleótidos 134-226 de la SEC ID n.º: 25, los nucleótidos 146-257 de la SEC ID n.º: 26 o los nucleótidos 147-223 de la SEC ID n.º: 27 o los nucleótidos 970-1449 de la SEC ID n.º: 1 o los nucleótidos 667-1228 de la SEC ID n.º: 2. Preferiblemente, dicha secuencia intrónica comprendida dentro de dicho elemento regulador adicional de la expresión comprende además un sitio de empalme donador tal como se define en el presente documento. Aún más preferido, dicha secuencia intrónica incluida dentro del elemento regulador adicional de la expresión es un intrón tal como se define en la sección Definición. Más preferiblemente, la secuencia intrónica incluida dentro del elemento regulador de la expresión adicional comprende o consta de una secuencia de nucleótidos que tiene al menos un 80 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad de secuencia con los nucleótidos 171-277 de la SEC ID n.º: 14, los nucleótidos 171-274 de la SEC ID n.º: 19, los nucleótidos 133-210 de la SEC ID n.º: 20, los nucleótidos 134-211 de la SEC ID n.º: 21, los nucleótidos 134-226 de la SEC ID n.º: 22, los nucleótidos 134-226 de la SEC ID n.º: 23, los nucleótidos 133-225 de la SEC ID n.º: 24, los nucleótidos 134-226 de la SEC ID n.º: 25, los nucleótidos 146-257 de la SEC ID n.º: 26 o los nucleótidos 147-223 de la SEC ID n.º: 27.

[0103] También se prefiere dentro de este aspecto un elemento regulador de la expresión que es un elemento

potenciador de la traducción. Preferiblemente, un elemento potenciador de la traducción permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la expresión de dicha proteína o polipéptido que usa un constructo que solo difiere en que está libre de dicho elemento potenciador de la traducción, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, preferiblemente la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la traducción que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión está libre de dicho elemento potenciador de la traducción que se va a ensayar.

[0104] Preferiblemente, dentro de este aspecto, dicho elemento potenciador de la traducción comprende o consta de una secuencia de nucleótidos que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con cualquiera de las SEC ID n.º: 3-51 en toda su longitud. Preferiblemente, dicho elemento potenciador de la traducción comprende o consta de una secuencia de nucleótidos que tiene al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de identidad con cualquiera de las SEC ID n.º: 19 en toda su longitud. Más preferiblemente, dicho elemento potenciador de la traducción comprende o consta de una secuencia de nucleótidos que tiene al menos un 90 % de identidad con la SEC ID n.º: 3-51 en toda su longitud. También se prefiere dentro de este aspecto un elemento potenciador de la traducción que comprende o consta de una secuencia de nucleótidos que comprende:

- i) una secuencia de nucleótidos repetida GAA, una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T;
- ii) una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T, dicho elemento potenciador de la expresión no comprende una secuencia de nucleótidos repetida GAA; o,
- iii) una secuencia de nucleótidos repetida GAA, una secuencia de nucleótidos rica en TC que comprende al menos 8 nucleótidos C o T consecutivos, al menos 3 secuencias de nucleótidos ricas en A que comprenden al menos 5 nucleótidos A consecutivos, una secuencia de nucleótidos rica en GT que comprende al menos 10 nucleótidos, al menos el 80 % de los cuales son nucleótidos G o T, en donde dicha secuencia de nucleótidos repetida GAA se encuentra en 3' de cualquiera de dichas secuencias de nucleótidos ricas en TC, secuencias de nucleótidos ricas en A y/o secuencias de nucleótidos ricas en GT.

[0105] La secuencia de nucleótidos repetida GAA, la secuencia de nucleótidos rica en TC, la secuencia de nucleótidos rica en A, la secuencia de nucleótidos rica en GT ya se han definido en el presente documento en el primer aspecto de la invención. Estas definiciones también se aplican aquí.

[0106] Preferiblemente dentro de dicho aspecto, dicho elemento regulador adicional de la expresión comprende un elemento potenciador de la traducción tal como se define en el presente documento y una secuencia intrónica.

[0107] Preferiblemente dentro de dicho aspecto, dicho elemento regulador de la expresión adicional está ubicado dentro de un constructo de ácidos nucleicos de la descripción aguas abajo o en el sitio 3' de un promotor heterólogo. Preferiblemente, dicho elemento regulador adicional de la expresión está ubicado dentro de un constructo de ácidos nucleicos de la divulgación aguas arriba o en el sitio 5' de una secuencia de ácidos nucleicos que codifica una proteína o un polipéptido de interés. Además, preferiblemente un constructo de ácidos nucleicos de la divulgación comprende las siguientes secuencias de nucleótidos indicadas aquí en sus posiciones relativas en la dirección 5' a 3': (i) un elemento potenciador de la expresión, (ii) un promotor heterólogo, opcionalmente (iii) un elemento regulador de expresión adicional, y opcionalmente (iv) una secuencia de nucleótidos de interés, en donde preferiblemente dicho elemento potenciador de la expresión, dicho promotor heterólogo y dicho elemento regulador de expresión adicional están configurados para estar todos operativamente enlazados a dicha secuencia de nucleótidos opcional de interés tal como se define más adelante en el presente documento. Debe entenderse que dicho elemento potenciador de la expresión, dicho promotor heterólogo y, opcionalmente, dicho elemento regulador de la expresión adicional del constructo de ácidos nucleicos de la divulgación están todos configurados para estar operativamente enlazados a la misma secuencia de nucleótidos sencilla de interés.

[0108] Los inventores encontraron un efecto sinérgico inesperado cuando el elemento potenciador de la expresión de la divulgación se combina con un elemento regulador de la expresión adicional tal como se define en el presente documento en un constructo de expresión para expresar una proteína o un polipéptido de interés.

En un grupo transfectado de manera estable con un elemento potenciador de la expresión y un elemento regulador de la expresión adicional, el rendimiento de proteína fue significativamente mayor que el rendimiento esperado basado en la suma de los efectos separados de cualquiera de los elementos. Preferiblemente, dicho aumento de la expresión de proteínas o polipéptidos es independiente del número de copias según lo establecido por un ensayo adecuado para determinar la dependencia del número de copias realizado por una persona experta tal como, por ejemplo, un ensayo Taqman triplex tal como se detalla adicionalmente en el ejemplo 4. Preferiblemente, dicho constructo de ácidos nucleicos es un constructo recombinante y/o aislado tal como se define en el presente documento. Preferiblemente, dicho constructo de ácidos nucleicos comprende además una secuencia de nucleótidos de interés operativamente enlazada y/o bajo el control de dicho elemento potenciador de la expresión, dicho promotor heterólogo y opcionalmente dicho elemento regulador adicional de la expresión tal como se define en el presente documento. La presencia de una secuencia de nucleótidos de interés es opcional. «Opcional» debe entenderse en el presente documento como no necesariamente presente en un constructo de expresión. Por ejemplo, dicha secuencia de nucleótidos de interés no necesita estar presente en un vector de expresión comercializado, sino que una persona experta en la técnica puede introducirla fácilmente antes de usarla en un método de la divulgación. Debe entenderse que dicho elemento potenciador de la expresión, dicho promotor heterólogo y, opcionalmente, dicho elemento regulador de la expresión adicional están configurados para estar operativamente enlazados a la misma secuencia de nucleótidos sencilla de interés. Preferiblemente, dicho constructo de ácidos nucleicos del décimo aspecto de la divulgación tiene al menos un 20%, un 25%, un 30%, un 35%, un 40%, un 45%, un 50%, 55%, un 60%, un 65%, un 70%, un 75 %, un 80%, un 85%, un 90%, un 91%, un 92%, un 93%, un 94%, un 95%, un 96%, un 97%, un 98%, un 99% o un 100 % de identidad con las SEC ID n.º 73, 74, 75 o 76.

[0109] En una realización preferida dentro de este aspecto, dicha secuencia de nucleótidos de interés es una secuencia que codifica una proteína o un polipéptido de interés. La proteína o el polipéptido de interés pueden ser una proteína o un polipéptido homólogos, pero en una realización preferida de la divulgación, la proteína o el polipéptido de interés son una proteína o un polipéptido heterólogos. Una secuencia de nucleótidos que codifica una proteína o un polipéptido heterólogos puede derivarse total o parcialmente de cualquier fuente conocida en la técnica, incluyendo un genoma o episoma bacteriano o viral, ADN eucariótico nuclear o de plásmido, ADNc o ADN sintetizado químicamente. La secuencia de nucleótidos que codifica una proteína o un polipéptido de interés puede constituir una región codificante ininterrumpida o puede incluir uno o más intrones unidos por uniones de empalme apropiadas. Además puede estar compuesta por segmentos derivados de diferentes fuentes, de origen natural o sintético. La secuencia de nucleótidos que codifica la proteína o el polipéptido de interés según el método de la divulgación es preferiblemente una secuencia de nucleótidos de longitud completa, pero también puede ser una parte funcionalmente activa u otra parte de dicha secuencia de nucleótidos de longitud completa. La secuencia de nucleótidos que codifica la proteína o el polipéptido de interés también puede comprender secuencias de señal que dirigen la proteína o polipéptido de interés cuando se expresa en una ubicación específica de la célula o tejido. Además, la secuencia de nucleótidos que codifica la proteína o el polipéptido de interés también puede comprender secuencias que facilitan la purificación de proteínas y la detección de proteínas mediante, por ejemplo, transferencia de Western y ELISA (por ejemplo, secuencias de c-myc o polihistidina).

[0110] La proteína o el polipéptido de interés en este aspecto ya se ha definido anteriormente aquí en el primer aspecto de la invención.

[0111] En una realización alternativa, dicha secuencia de nucleótidos de interés no es una secuencia codificante para una proteína o un polipéptido, sino que puede ser una secuencia de nucleótidos funcional. Esta realización alternativa de este aspecto ya se ha definido anteriormente aquí en el primer aspecto de la invención.

Duodécimo aspecto

[0112] En un duodécimo aspecto, la presente divulgación proporciona un vector de expresión que comprende un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación. El vector de expresión de la divulgación es preferiblemente una secuencia de plásmido, cósmido o fago o nucleótido, lineal o circular, de un ADN o ARN monocatenario o bicatenario, derivado de cualquier fuente, en el que varias secuencias de nucleótidos se han unido o recombinado en un constructo único capaz de introducir cualquiera de las secuencias de nucleótidos de la divulgación en una célula en las orientaciones sentido o antisentido. La elección del vector depende de los procedimientos recombinantes seguidos y de la célula huésped utilizada. El vector puede ser un vector de replicación autónoma o puede replicarse junto con el cromosoma en el que se ha integrado. Preferiblemente, el vector contiene un marcador de selección. Los marcadores útiles dependen de la célula huésped elegida, son bien conocidos por los expertos en la técnica y se seleccionan de, entre otros, los marcadores de selección tal como se define en el tercer aspecto de la invención. Un vector de expresión

preferido es el vector de expresión ADNpc3.1. Los marcadores de selección preferidos son el gen de resistencia a la neomicina, el gen de resistencia a la zeocina y el gen de resistencia a la blasticidina.

Decimotercer aspecto

[0113] En un decimotercer aspecto, la presente divulgación proporciona una célula que comprende una molécula de ácidos nucleicos según el décimo aspecto de la divulgación y/o un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación y/o un vector de expresión según el duodécimo aspecto de la divulgación tal como se define en el presente documento. El tipo de célula dentro del contexto de este aspecto es el mismo que el definido en el contexto del tercer aspecto. Por lo tanto, otro aspecto de la divulgación se refiere a una célula huésped que está modificada genéticamente, preferiblemente mediante un método de la divulgación, en el sentido de que una célula huésped comprende un constructo de ácidos nucleicos tal como se ha definido anteriormente en el decimotercer aspecto. Para los procedimientos de transformación en plantas, las bacterias adecuadas incluyen *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes*.

[0114] Un constructo de ácidos nucleicos dentro del contexto de este decimotercer aspecto es como el del tercer aspecto: preferiblemente se mantiene de manera estable, ya sea como un elemento de replicación autónoma, o, más preferiblemente, el constructo de ácidos nucleicos se integra en el genoma de la célula huésped, en cuyo caso el constructo generalmente se integra en posiciones aleatorias en el genoma de la célula huésped, por ejemplo, mediante recombinación no homóloga. Las células huésped transformadas de forma estable se producen mediante métodos conocidos. La definición del término transformación estable y los métodos incluidos para la transformación estable ya se han proporcionado bajo el tercer aspecto.

[0115] Alternativamente, una proteína o un polipéptido de interés pueden expresarse en una célula huésped, por ejemplo, una célula de mamífero, apoyándose en la expresión transitoria de los vectores.

[0116] Un constructo de ácidos nucleicos según este aspecto preferiblemente también comprende un gen marcador que puede proporcionar capacidad de selección o cribado en una célula huésped tratada.

[0117] Todas las definiciones relacionadas con marcadores seleccionables y tipos de marcadores seleccionables, incluyendo el ejemplo del uso del gen de la luciferasa como marcador seleccionable, el ejemplo de una primera categoría de marcador basada en el metabolismo celular y el uso de una línea celular mutante que carece de la capacidad para crecer independientemente de un medio complementado y el ejemplo de selección dominante ya se han proporcionado en el tercer aspecto. También se aplican aquí en el decimotercer aspecto de la divulgación.

[0118] Cuando se obtiene una célula huésped transformada con un método según la divulgación (véase más adelante), se puede regenerar un tejido huésped a partir de dicha célula transformada en un medio adecuado, que opcionalmente puede contener antibióticos o biocidas conocidos en la técnica para la selección de células transformadas.

[0119] Los tejidos huésped transformados resultantes se identifican preferiblemente mediante selección usando un gen marcador de selección presente en un constructo de ácidos nucleicos tal como se define en el presente documento.

Decimocuarto aspecto

[0120] En un decimocuarto aspecto, la presente divulgación proporciona un método para expresar y opcionalmente purificar una proteína o un polipéptido de interés que comprende el paso de:

- a. proporcionar un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación que comprende una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés; y,
- b. poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c. permitir que dicha célula transformada exprese la proteína o polipéptido de interés; y opcionalmente
- d. purificar dicha proteína o polipéptido de interés.

[0121] En una realización preferida del método según la divulgación, se usa un constructo de ácidos nucleicos tal como se ha definido anteriormente en el undécimo aspecto de la divulgación. El método de la divulgación puede ser un método in vitro o ex vivo. El método de la divulgación se puede aplicar en un cultivo celular, cultivo de organismos o cultivo de tejidos. Alternativamente, junto a la expresión en células huésped, la proteína o el polipéptido de interés pueden producirse en sistemas de traducción sin células usando ARN derivados de los constructos de ácidos nucleicos de la presente divulgación. El método de la divulgación puede ejecutarse en células cultivadas.

[0122] El experto en la técnica es capaz de transformar células según el paso b). Los métodos de transformación

que se usan en el paso b) incluyen, entre otros, la transferencia de ADN purificado a través de reactivos lipídicos catiónicos y polietilenimida (PEI), coprecipitación de fosfato de calcio, bombardeo de micropartículas, electroporación de protoplastos y microinyección o uso de fibras de silicio para facilitar la penetración y transferencia de ADN en la célula huésped.

[0123] En el paso c) se permite que la célula transformada exprese la proteína o el polipéptido de interés, y opcionalmente dicha proteína o polipéptido se recupera a continuación. Por ejemplo, la célula transformada puede someterse a condiciones que producen la expresión de la proteína o el polipéptido de interés. Los expertos en la técnica conocen bien las técnicas que se van a utilizar para expresar o sobreexpresar la proteína o el polipéptido de interés. Los métodos en los que la célula transformada no necesita someterse a condiciones específicas que producen la expresión de la proteína o el polipéptido de interés pero en los que la proteína o el polipéptido de interés se expresan automáticamente (por ejemplo, constitutivamente), también se incluyen en el método de la presente divulgación.

[0124] Los pasos de purificación y las definiciones relacionadas con estos pasos, por ejemplo, la definición de una proteína o un polipéptido aislado, son los mismos que en el método del cuarto aspecto y se han definido anteriormente en el presente documento. Si se desea tal como se define en el método del cuarto aspecto, la secuencia de nucleótidos que codifica una proteína o un polipéptido de interés se puede ligar a una secuencia de nucleótidos heteróloga para codificar una proteína o un polipéptido de fusión para facilitar la purificación y la detección de proteínas, por ejemplo, en transferencia Western y en un ELISA. Las secuencias heterólogas adecuadas incluyen, entre otras, las secuencias de nucleótidos que codifican proteínas tales como, por ejemplo, glutatión-S-transferasa, proteína de unión a maltosa, polihistidina de unión a metal, proteína fluorescente verde, luciferasa y beta-galactosidasa. La proteína o el polipéptido también se pueden acoplar a vehículos, etiquetas o marcadores no peptídicos que facilitan el rastreo de la proteína o el polipéptido, tanto in vivo como in vitro, y permiten la identificación y la cuantificación de la unión de la proteína o el polipéptido a sustratos. Dichas etiquetas, marcadores o vehículos son bien conocidos en la técnica e incluyen, entre otros, biotina, marcadores radiactivos y marcadores fluorescentes.

[0125] Preferiblemente, el método de este decimocuarto aspecto de la divulgación permite un aumento en la expresión de una proteína o un polipéptido de interés. Preferiblemente, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión según el undécimo aspecto de la divulgación que comprende un elemento potenciador de la expresión y un promotor heterólogo operativamente enlazado a una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés según el undécimo aspecto de la divulgación. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo ELISA, una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, el método de la divulgación permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho elemento potenciador de la expresión, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, preferiblemente la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema de células de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la expresión que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión está libre de dicho elemento potenciador de la expresión que se va a ensayar. Preferiblemente, dicho aumento de la expresión de proteínas o polipéptidos es independiente del número de copias según lo establecido por un ensayo adecuado para determinar la dependencia del número de copias realizado por una persona experta tal como, por ejemplo, un ensayo Taqman triplex tal como se detalla adicionalmente en el ejemplo 4.

Decimoquinto aspecto

[0126] En un decimoquinto aspecto, la presente divulgación proporciona un método para expresar una proteína o un polipéptido de interés en un organismo, que comprende los pasos de:

- a) proporcionar un constructo de ácidos nucleicos según el undécimo aspecto que comprende una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés de la divulgación; y,
- b) poner en contacto una célula diana y/o tejido diana de un organismo con dicho constructo de ácidos nucleicos para obtener una célula diana transformada y/o tejido diana transformado, permitiendo que dicha célula transformada exprese la proteína o el polipéptido de interés; y opcionalmente

- c) permitir que dicha célula diana transformada se desarrolle en un organismo transformado; y, opcionalmente,
- d) permitir que dicho organismo transformado exprese la proteína o el polipéptido de interés, por ejemplo, sometiendo dicho organismo transformado a condiciones que produzcan la expresión de la proteína o el polipéptido de interés, y opcionalmente recuperar dicha proteína o polipéptido.

[0127] La célula diana puede ser una célula diana embrionaria, p. ej., una célula madre embrionaria, por ejemplo, derivada de un mamífero no humano, tal como de las especies bovina, porcina, etc. Preferiblemente, dicha célula diana no es una célula madre embrionaria humana. En el caso de un hongo multicelular, dicha célula diana puede ser una célula fúngica que puede proliferar en dicho hongo multicelular. Cuando se obtiene un tejido vegetal transformado o una célula vegetal (por ejemplo, trozos de hoja, segmentos de tallo, raíces, pero también protoplastos o células vegetales cultivadas por suspensión) con este método según la divulgación, se pueden regenerar plantas enteras a partir de dicho tejido o célula transformados en un medio adecuado, que opcionalmente puede contener antibióticos o biocidas conocidos en la técnica para la selección de células transformadas. Este método de la divulgación se puede aplicar en vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano. En la presente divulgación se incluye un método de tratamiento que comprende el método del presente aspecto, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico. La divulgación también se refiere a un constructo del undécimo aspecto de la divulgación para el tratamiento, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico. Además, la divulgación también se refiere a un constructo del undécimo aspecto de la divulgación para la fabricación de un medicamento, en el que la proteína o polipéptido de interés es una proteína o polipéptido terapéutico y/o inmunogénico.

[0128] Además, una realización de la divulgación es un organismo transformado no humano. Dicho organismo se transforma con una secuencia de nucleótidos, un constructo de ácidos nucleicos recombinante o un vector según la presente divulgación, y es capaz de producir el polipéptido de interés. Esto incluye un organismo transgénico no humano, tal como un mamífero transgénico no humano, una planta transgénica (incluyendo la propagación, la cosecha y el material de tejido de dicha planta transgénica, que incluye, entre otros, hojas, raíces, brotes y flores), un hongo multicelular y similares.

[0129] Preferiblemente, el método de este aspecto de la divulgación permite un aumento en la expresión de una proteína o un polipéptido de interés en dicho organismo o al menos en un tejido u orgánulo u órgano de dicho organismo. Preferiblemente, los niveles de expresión se establecen en un sistema de expresión usando un constructo de expresión según el undécimo aspecto de la divulgación que comprende un elemento potenciador de la expresión y un promotor heterólogo operativamente enlazado a una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés según el undécimo aspecto de la divulgación. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo ELISA, una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Preferiblemente, este método de la divulgación permite un aumento en la expresión de proteínas o polipéptidos de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en dicho organismo o al menos en un tejido u orgánulo u órgano de dicho organismo en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho elemento potenciador de la expresión, preferiblemente cuando se prueba en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, preferiblemente la expresión de la secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema de células de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la expresión que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La expresión se mide preferiblemente midiendo la conversión de cualquier sustrato de fosfatasa alcalina adecuado y los niveles de expresión se comparan con los niveles de expresión de dicha secuencia de nucleótidos de interés que se miden en las mismas condiciones, excepto que el vector de expresión está libre de dicho elemento potenciador de la expresión que se va a ensayar.

[0130] Preferiblemente, dicho aumento de la expresión de proteínas o polipéptidos es independiente del número de copias según lo establecido por un ensayo adecuado para determinar la dependencia del número de copias realizado por una persona experta tal como, por ejemplo, un ensayo Taqman triplex tal como se detalla adicionalmente en el ejemplo 4 de la presente divulgación.

Decimosexto aspecto

[0131] En un decimosexto aspecto, la presente divulgación proporciona un método para la transcripción y opcionalmente la purificación del transcrito producido que comprende el paso de:

- a) proporcionar un constructo de ácidos nucleicos según el undécimo aspecto que comprende una secuencia de nucleótidos de interés de la divulgación; y,
- b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
- c) permitir que dicha célula transformada produzca un transcrito de la secuencia de nucleótidos de interés; y opcionalmente
- d) purificar dicho transcrito producido.

[0132] En una realización preferida de este método según la divulgación, se usa un constructo de ácidos nucleicos tal como se ha definido anteriormente en el undécimo aspecto. El método de la divulgación puede ser un método in vitro o ex vivo. El método de la divulgación se puede aplicar en un cultivo celular, cultivo de organismos o cultivo de tejidos. El método de la divulgación se puede aplicar en vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, preferiblemente en un ser humano. En la presente divulgación se incluye un método para el tratamiento que comprende o consta del método del presente aspecto, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico. La divulgación también se refiere a un constructo del undécimo aspecto de la divulgación para su uso en el tratamiento, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico. Además, la divulgación también se refiere a un constructo del undécimo aspecto de la divulgación para la fabricación de un medicamento, en el que la secuencia de nucleótidos de interés codifica un transcrito terapéutico.

[0133] El experto en la técnica es capaz de transformar células según el paso b). Los métodos de transformación que se usan en el paso b) incluyen, entre otros, la transferencia de ADN purificado a través de reactivos lipídicos catiónicos y polietilenimida (PEI), coprecipitación de fosfato de calcio, bombardeo de micropartículas, electroporación de protoplastos y microinyección o uso de fibras de silicio para facilitar la penetración y transferencia de ADN en la célula huésped.

[0134] En el paso c) se permite que la célula transformada produzca un transcrito de la secuencia de nucleótidos de interés y opcionalmente el transcrito producido se recupera posteriormente. Por ejemplo, la célula transformada puede someterse a condiciones que producen la transcripción de la secuencia de nucleótidos de interés. Los expertos en la técnica conocen bien las técnicas que se van a utilizar para la transcripción de la secuencia de nucleótidos de interés. Los métodos en los que la célula transformada no necesita someterse a condiciones específicas que producen la transcripción de la secuencia de nucleótidos de interés, pero en los que la secuencia de nucleótidos de interés se transcribe automáticamente (por ejemplo, constitutivamente), también se incluyen en el método de la presente divulgación.

[0135] Los pasos de purificación dependen del transcrito producido. El término «aislamiento» indica que el transcrito se encuentra en una condición distinta de su entorno nativo. En una forma preferida, el transcrito aislado está sustancialmente libre de otros componentes celulares, particularmente otros componentes celulares homólogos tales como proteínas homólogas. Se prefiere proporcionar el transcrito en una forma con una pureza superior al 40 %, más preferiblemente en una forma con una pureza superior al 60 %. Incluso más preferiblemente, se prefiere proporcionar el transcrito en una forma altamente purificada, es decir, con una pureza superior al 80 %, más preferiblemente con una pureza superior al 90 % e incluso más preferiblemente con una pureza superior al 99 %, según se determina mediante transferencia Northern.

[0136] Preferiblemente, el método de este aspecto de la divulgación permite un aumento en la transcripción de una secuencia de nucleótidos de interés. Preferiblemente, los niveles de transcripción se establecen en un sistema de expresión usando un constructo de expresión según el segundo aspecto de la divulgación que comprende un elemento potenciador de la expresión y un promotor heterólogo operativamente enlazado a una secuencia de nucleótidos de interés según el segundo aspecto de la divulgación. Preferiblemente, la transcripción de dicha secuencia de nucleótidos de interés se detecta mediante un ensayo adecuado tal como RT-qPCR. Preferiblemente, el método de la divulgación permite un aumento en la transcripción de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con un método que solo difiere en que se utiliza un constructo en el paso a) que está libre de dicho elemento potenciador de la expresión, preferiblemente cuando se ensaya en un sistema tal como se muestra en los ejemplos que se incluyen en el presente documento. Más específicamente, preferiblemente la transcripción de una secuencia de nucleótidos de interés que codifica fosfatasa alcalina secretada (SeAP) se mide en un sistema celular de mamífero, más preferiblemente en células CHO, usando un vector de expresión ADNpc3.1 que comprende un elemento potenciador de la expresión que se va a ensayar y un promotor de CMV representado por la SEC ID n.º: 57, operativamente enlazado a dicha secuencia de nucleótidos de interés. La transcripción se mide preferiblemente usando RT-qPCR y los niveles de transcripción se comparan con los niveles de transcripción de dicha secuencia de nucleótidos de interés medida en las mismas condiciones, excepto que el vector de expresión usado está libre de dicho elemento potenciador de la expresión que se va a ensayar.

Decimoséptimo aspecto

[0137] En un decimoséptimo aspecto, la presente divulgación proporciona un uso de una molécula de ácidos nucleicos según el décimo aspecto de la divulgación y/o un uso de un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación y/o un uso de un vector de expresión según el duodécimo aspecto de la divulgación y/o un uso de una célula según el decimotercer aspecto de la divulgación, para la transcripción de una secuencia de nucleótidos de interés y/o la expresión de una proteína o un polipéptido de interés.

Decimotavo aspecto

[0138] En un decimotavo aspecto, la presente divulgación proporciona una molécula de ácidos nucleicos de acuerdo con el décimo aspecto de la divulgación y/o un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación y/o un vector de expresión según el duodécimo aspecto de la divulgación y/o una célula de acuerdo con el decimotercer aspecto de la divulgación para su uso como medicamento. La divulgación también se refiere a un método de tratamiento que comprende la administración de una molécula de ácidos nucleicos según el décimo aspecto de la divulgación, y/o de un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación y/o un vector de expresión según el duodécimo aspecto de la divulgación y/o una célula según el decimotercer aspecto de la divulgación, en donde preferiblemente dicha administración se realiza a un mamífero, más preferiblemente a un humano. Preferiblemente, dicho tratamiento es una vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano. Además, la divulgación se refiere al uso de una molécula de ácidos nucleicos según el décimo aspecto de la divulgación y/o el uso de un constructo de ácidos nucleicos según el undécimo aspecto de la divulgación y/o el uso de un vector de expresión según el duodécimo aspecto de la divulgación y/o el uso de una célula según el decimotercer aspecto de la divulgación para la preparación de un medicamento. Preferiblemente dicho medicamento es para vacunación basada en ácidos nucleicos y/o terapia génica preferiblemente en un mamífero, más preferiblemente en un ser humano.

Definiciones

[0139] La frase «ácido nucleico», tal como se usa en el presente documento, se refiere a un oligonucleótido o polinucleótido sintético o natural, ya sea ADN o ARN o híbrido de ADN-ARN, monocatenario o bicatenario, sentido o antisentido, que es capaz de hibridarse con un ácido nucleico complementario mediante apareamiento de bases Watson-Crick. Un ácido nucleico de la invención se modifica preferiblemente en comparación con su contrapartida natural cuando comprende al menos 1, 2, 3, 4, 5, 10, 20, 30 o 50 mutaciones de nucleótidos en comparación con su contrapartida natural. Preferiblemente, un ácido nucleico de la invención no se produce en la naturaleza. Los ácidos nucleicos de la invención también pueden incluir análogos de nucleótidos (por ejemplo, BrdU) y enlaces internucleosídicos no fosfodiéster (por ejemplo, enlaces de ácido peptidonucleico (APN) o tiodiéster). En particular, los ácidos nucleicos pueden incluir, entre otros, ADN, ARN, ADNc, ADNg, ADNmc, ADNbc, ARNmc, ARNbc, ARN no codificante, ARNnh, ARNpre-m, ARNm madurado o cualquier combinación de los mismos. Los términos «secuencia de ácidos nucleicos» y «secuencia de nucleótidos» tal como se usan en el presente documento son intercambiables y tienen su significado habitual en la técnica. El término se refiere a una molécula de ADN o ARN en forma monocatenaria o bicatenaria. Una «secuencia de ácidos nucleicos aislada» se refiere a una secuencia de ácidos nucleicos que ya no se encuentra en el entorno natural del que se aisló. Una molécula de ácidos nucleicos está representada por una secuencia de nucleótidos. Además, un elemento tal como, por ejemplo, un elemento potenciador de la expresión y un elemento regulador de la transcripción está representado por una secuencia de nucleótidos.

[0140] Un «constructo recombinante» (o constructo quimérico) se refiere a cualquier secuencia o molécula de ácidos nucleicos, que normalmente no se encuentra en la naturaleza en una especie, en particular una secuencia, molécula o gen de ácidos nucleicos en el que están presentes una o más partes de la secuencia de ácidos nucleicos que no están asociadas entre sí en la naturaleza. Por ejemplo, un constructo recombinante comprende un promotor que no está asociado en la naturaleza con parte o la totalidad de la región transcrita o con otra región reguladora comprendida dentro de dicho constructo recombinante. Se entiende que el término «constructo recombinante» incluye constructos de expresión en los que un promotor o una secuencia reguladora de la expresión están operativamente enlazados a una o varias secuencias sentido (por ejemplo, secuencias codificantes) o antisentido (complemento inverso de la cadena sentido) o secuencia de repetición invertida (sentido y antisentido, en donde el transcrito de ARN forma ARN bicatenario tras la transcripción) o cualquier otra secuencia que codifique una molécula de ARN funcional.

[0141] Un «constructo de ácidos nucleicos» se define como un polinucleótido que se aísla de un gen natural o que se ha modificado para contener segmentos de polinucleótidos que se combinan o yuxtaponen de una manera que de otra manera no existiría en la naturaleza. Opcionalmente, un polinucleótido presente en un constructo de ácidos nucleicos está operativamente enlazado a una o más secuencias de control, que dirigen la producción o transcripción de una secuencia de nucleótidos de interés y/o la expresión de un péptido o polipéptido de interés en una célula o en un sujeto.

[0142] Se entiende que un «vector» o «plásmido» significa una molécula de ácidos nucleicos artificial

(generalmente circular) que resulta del uso de tecnología de ADN recombinante y que se usa para administrar ADN exógeno a una célula huésped. Los vectores generalmente comprenden elementos genéticos adicionales para facilitar su uso en la clonación molecular, como p. ej., marcadores seleccionables, sitios de clonación múltiples y similares (ver más adelante). Un constructo de ácidos nucleicos también puede ser parte de un vector vírico recombinante para la expresión de una proteína en una planta o célula vegetal (por ejemplo, un vector derivado del virus del mosaico de la coliflor, CaMV o el virus del mosaico del tabaco, TMV) o en un organismo mamífero o un sistema celular mamífero (por ejemplo, un vector derivado del virus de la leucemia murina de Moloney (MMLV; un retrovirus), un lentivirus, un virus adenoasociado (AAV) o un adenovirus (AdV)).

[0143] Una «célula transformada» es un término que se refiere a una nueva célula (u organismo) individual, que surge como resultado de la introducción en dicha célula de al menos una molécula de ácidos nucleicos, que comprende especialmente un constructo quimérico o recombinante que codifica una proteína o una secuencia de ácidos nucleicos deseadas que tras la transcripción produce un ARN antisentido para silenciar un gen diana/familia de genes. La célula huésped puede ser una célula vegetal, una célula bacteriana (por ejemplo, una cepa de *Agrobacterium*), una célula fúngica (que incluye una célula de levadura), una célula animal (que incluye insectos, mamíferos), etc. La célula transformada puede contener el constructo de ácidos nucleicos como una molécula replicante extracromosómica (episomal), como una molécula no replicante o comprende el constructo recombinante integrado en el ADN nuclear o de orgánulo de la célula huésped. El término «organismo», tal como se usa en el presente documento, abarca todos los organismos que consisten en más de una célula, *es decir* organismos multicelulares, e incluye hongos multicelulares. «Transformación» y «transformado» se refieren a la transferencia de una secuencia de ácidos nucleicos, generalmente una secuencia de ácidos nucleicos que comprende un constructo recombinante o un gen de interés (GOI), en el genoma nuclear de una célula para crear una célula u organismo «transgénico» que comprende un transgén. La secuencia de ácidos nucleicos introducida generalmente está integrada en el genoma del huésped, pero no siempre. Cuando la secuencia de ácidos nucleicos introducida no está integrada en el genoma del huésped, se puede hablar de «transfección», «transfectada transitoriamente» y «transfectada». Para los fines de la presente memoria descriptiva de la patente, los términos «transformación», «transfectado transitoriamente» y «transfección» se usan indistintamente y se refieren a la presencia estable o transitoria de una secuencia de ácidos nucleicos en una célula u organismo. Cuando la célula es una célula bacteriana, el término generalmente se refiere a un vector extracromosómico, autorreplicante, que alberga una resistencia a los antibióticos seleccionable.

[0144] La «identidad de secuencia» o «identidad» en el contexto de la secuencia de aminoácidos o de ácidos nucleicos se define en el presente documento como una relación entre dos o más secuencias de aminoácidos (péptidos, polipéptidos o proteínas) o dos o más ácidos nucleicos (nucleótidos, secuencias de polinucleótidos), según lo determinado mediante la comparación de las secuencias. En la técnica, «identidad» también significa el grado de relación de secuencia entre las secuencias de aminoácidos o nucleótidos, según sea el caso, de acuerdo con lo determinado mediante la coincidencia entre cadenas de tales secuencias. Dentro de la presente invención, la identidad de secuencia con una secuencia particular indicada con una SEC ID n.º particular significa preferiblemente identidad de secuencia en toda la longitud de dicha secuencia de polipéptidos o polinucleótidos particular indicada con dicha SEC ID n.º particular. Sin embargo, la identidad de secuencia con una secuencia particular indicada con una SEC ID n.º particular también puede significar que la identidad de secuencia se evalúa sobre una parte de dicha SEC ID n.º. Una parte puede significar al menos 50 %, un 60 %, un 70 %, un 80 %, un 90 % o un 95 % de la longitud de dicha SEC ID n.º. La información de secuencia que se proporciona en este documento no debe interpretarse tan estrictamente como para requerir la inclusión de bases identificadas erróneamente. La persona experta en la técnica es capaz de identificar tales bases identificadas erróneamente y sabe cómo corregir dichos errores.

[0145] Cualquier secuencia de nucleótidos capaz de hibridarse con las secuencias de nucleótidos de la invención se define como parte de los elementos de acción en *cis* de la invención. Las condiciones de hibridación rigurosas se definen en el presente documento como condiciones que permiten que una secuencia de ácidos nucleicos de al menos 25, preferiblemente 50, 75 o 100 y más preferiblemente 150 o más nucleótidos, se hibride a una temperatura de aproximadamente 65 °C o de 65 °C en una solución que comprende aproximadamente 1 M de sal o 1 M de sal, preferiblemente 6 x SSC o cualquier otra solución que tenga una fuerza iónica comparable, y se lave a 65 °C en una solución que comprende aproximadamente 0.1 M de sal, o 0.1 M de sal o menos, preferiblemente 0.2 x SSC o cualquier otra solución que tenga una fuerza iónica comparable. Preferiblemente, la hibridación se realiza durante la noche, *es decir* al menos durante 10 horas y preferiblemente el lavado se realiza durante al menos una hora con al menos dos cambios de la solución de lavado. Estas condiciones generalmente permitirán la hibridación específica de secuencias que tienen aproximadamente un 90 % o más de identidad de secuencia o al menos un 90 % de identidad de secuencia. Las condiciones de hibridación moderadas se definen en el presente documento como condiciones que permiten que una secuencia de ácidos nucleicos de al menos 50, preferiblemente 150 o más nucleótidos, se hibride a una temperatura de aproximadamente 45 °C o de 45 °C en una solución que comprende aproximadamente 1 M de sal o 1 M de sal, preferiblemente 6 x SSC o cualquier otra solución que tenga una fuerza iónica comparable, y se lave a temperatura ambiente en una solución que comprende aproximadamente 1 M de sal, o 1 M de sal preferiblemente 6 x SSC o cualquier otra solución que tenga una fuerza iónica comparable. Preferiblemente, la hibridación se realiza durante la noche, *es decir* al menos durante 10 horas y preferiblemente el lavado se realiza durante al menos una hora con al menos dos

cambios de la solución de lavado. Estas condiciones generalmente permitirán la hibridación específica de secuencias que tienen una identidad de secuencia de hasta un 50 %. La persona experta en la materia podrá modificar estas condiciones de hibridación para identificar específicamente secuencias que varían en identidad entre un 50 % y un 90 %.

[0146] La «identidad» se puede calcular fácilmente mediante métodos conocidos, incluidos, entre otros, los descritos en Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of Sequence Data, Parte I, Griffin, A. M. y Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991; y Carillo, H. y Lipman, D., SIAM J. Applied Math., 48: 1073 (1988).

[0147] Los métodos preferidos para determinar la identidad están diseñados para proporcionar la mayor coincidencia entre las secuencias ensayadas. Los métodos para determinar la identidad y la similitud se codifican en programas informáticos disponibles para el público. Los métodos informáticos preferidos para determinar la identidad y similitud entre dos secuencias incluyen, p. ej., el paquete del programa GCG (Devereux, J., et al., Nucleic Acids Research 12 (1): 387 (1984)), BestFit, BLASTP, BLASTN y FASTA (Altschul, S. F. et al., J. Mol. Biol. 215: 403-410 (1990) El programa BLAST X está disponible públicamente en NCBI y otras fuentes (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215: 403-410 (1990) El conocido algoritmo Smith Waterman también se puede usar para determinar la identidad.

[0148] Los parámetros preferidos para la comparación de secuencias de polipéptidos incluyen los siguientes: Algoritmo: Needleman y Wunsch, J. Mol. Biol. 48: 443-453 (1970); matriz de comparación: BLOSSUM62 de Hentikoff y Hentikoff, Proc. Natl. Acad. Sci. EE. UU. 89: 10915-10919 (1992); penalización de gap: 12; y penalización de longitud de gap: 4. Un programa útil con estos parámetros está disponible para el público como el programa «Ogap» de Genetics Computer Group, ubicado en Madison, WI. Los parámetros antes mencionados son los parámetros predeterminados para las comparaciones de aminoácidos (junto con ausencia de penalización de gaps finales).

[0149] Los parámetros preferidos para la comparación de ácidos nucleicos incluyen los siguientes: Algoritmo: Needleman y Wunsch, J. Mol. Biol. 48: 443-453 (1970); matriz de comparación: coincidencias = + 10, falta de coincidencia = 0; penalización de gap: 50; penalización de longitud de gap: 3. Disponible como el programa Gap de Genetics Computer Group, ubicado en Madison, Wisconsin. Los parámetros predeterminados para las comparaciones de ácido nucleico son los anteriores.

[0150] El programa y el parámetro preferidos para evaluar la identidad de la comparación de ácidos nucleicos se calculan utilizando el algoritmo de alineación de nucleótidos EMBOSS Needle con los siguientes parámetros: matriz DNAfull con las siguientes penalizaciones de gap: abierto = 10; extensión = 0.5 tal como se lleva a cabo en el ejemplo 9.

[0151] El término «derivado de» en el contexto de la derivación a partir de un gen o secuencia de origen natural particular se define en el presente documento de modo que está sintetizado químicamente de acuerdo con un gen o secuencia natural y/o aislado y/o purificado a partir de un gen o secuencia de origen natural. Las técnicas para síntesis química, aislamiento y/o purificación de moléculas de ácidos nucleicos son bien conocidas en la técnica. En general, una secuencia derivada es una secuencia parcial del gen o secuencia de origen natural o una fracción del gen o secuencia de origen natural. Opcionalmente, la secuencia derivada comprende sustituciones o mutaciones de ácidos nucleicos, preferiblemente dando como resultado una secuencia de al menos un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 75 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 % o un 99 % idénticos en toda su longitud al gen o gen parcial o secuencia o secuencia parcial del gen de origen natural.

[0152] «Polipéptido», tal como se usa en el presente documento, se refiere a cualquier péptido, oligopéptido, polipéptido, producto génico, producto de expresión o proteína. Un polipéptido está compuesto de aminoácidos consecutivos. El término «polipéptido» abarca moléculas naturales o sintéticas. Un polipéptido está representado por una secuencia de aminoácidos. Un polinucleótido está representado por una secuencia de nucleótidos. Un polipéptido está representado por una secuencia de aminoácidos.

[0153] El término «homólogo» cuando se usa para indicar la relación entre una molécula de ácidos nucleicos o de polipéptido dada y un organismo huésped o célula huésped dada, significa que en la naturaleza la molécula de ácidos nucleicos o de polipéptido es producida por una célula u organismo huésped de la misma especie, preferiblemente de la misma variedad o cepa. Si es homóloga a una célula huésped, una secuencia de ácidos nucleicos de interés, que preferiblemente codifica un polipéptido, estará operativamente enlazada a otra secuencia promotora o, si corresponde, a otra secuencia señal secretora y/o secuencia de terminación que en su entorno natural.

[0154] Cuando se usa para indicar la relación de dos secuencias de ácidos nucleicos, el término «homólogo» significa que una secuencia de ácido nucleico monocatenario puede hibridarse con una secuencia complementaria de ácido nucleico monocatenario. El grado de hibridación puede depender de una serie de factores, incluido el grado de identidad entre las secuencias y las condiciones de hibridación, tales como la temperatura y la concentración de sal, según se describe más adelante. Preferiblemente, la región de identidad es mayor que 5 pb, más preferiblemente la región de identidad es mayor que 10 pb.

[0155] El término «heterólogo» cuando se usa para indicar la relación entre una molécula de ácidos nucleicos o de polipéptido dada (recombinante) y un organismo o célula huésped dados, se entiende que significa una molécula de ácido nucleico o de polipéptido de una célula extraña que no ocurre de manera natural como parte del organismo, célula, genoma o secuencia de ADN o ARN en la que está presente, o que se encuentra en una célula o ubicación o ubicaciones en el genoma o las secuencias de ADN o ARN que difiere de aquellas en las que se encuentra en la naturaleza. Los ácidos nucleicos o proteínas heterólogos no son endógenos a la célula en la que se introducen, sino que se han obtenido de otra célula o se han producido de forma sintética o recombinante.

[0156] Cuando se usa para indicar la relación de dos secuencias de ácidos nucleicos, el término «secuencia heteróloga» o «ácido nucleico heterólogo» es aquel que no se encuentra operativamente enlazado de manera natural como secuencia vecina de la otra secuencia. Tal como se usa en el presente documento, el término «heterólogo» puede significar «recombinante». «Recombinante» se refiere a una entidad genética distinta de la que generalmente se encuentra en la naturaleza. Según se aplica a una secuencia de nucleótidos o molécula de ácidos nucleicos, esto significa que dicha secuencia de nucleótidos o molécula de ácidos nucleicos es el producto de varias combinaciones de pasos de clonación, restricción y/o de ligadura y otros procedimientos que dan como resultado la producción de un constructo que es distinto de una secuencia o molécula encontrada en la naturaleza.

[0157] «Operativamente enlazado» se define en el presente documento como una configuración en la que una secuencia de control o secuencia reguladora se coloca apropiadamente en una posición relativa a la secuencia de nucleótidos de interés, preferiblemente codificando el polipéptido de interés de tal manera que la secuencia de control o regulación dirige o afecta a la transcripción y/o producción o expresión de la secuencia de nucleótidos de interés, que codifica preferiblemente un péptido o polipéptido de la invención en una célula y/o un sujeto. Por ejemplo, un promotor está operativamente enlazado a una secuencia codificante si el promotor puede iniciar o regular la transcripción o la expresión de una secuencia codificante, en cuyo caso la secuencia codificante debe entenderse que está «bajo el control» del promotor. Cuando una o más secuencias de nucleótidos y/o elementos comprendidos dentro de un constructo se definen en el presente documento como «configurados para estar operativamente enlazados a una secuencia de nucleótidos opcional de interés», se entiende que dichas secuencias de nucleótidos y/o elementos están configurados dentro de dicho constructo de tal manera que estas secuencias de nucleótidos y/o elementos están operativamente enlazados a dicha secuencia de nucleótidos de interés una vez que dicha secuencia de nucleótidos de interés está presente en dicho constructo.

[0158] «Promotor» se refiere a una secuencia de ácidos nucleicos ubicada aguas arriba o en 5' de un codón de inicio de traducción de un marco de lectura abierto (o región codificante de proteínas) de un gen y que está involucrado en el reconocimiento y la unión de la ARN polimerasa II y otras proteínas (factores de transcripción de acción en trans) para iniciar la transcripción. El término promotor se refiere a un fragmento de ácidos nucleicos que funciona para controlar la transcripción de uno o más genes, ubicados aguas arriba con respecto a la dirección de transcripción del sitio de inicio de la transcripción del gen y se identifica estructuralmente por la presencia de un sitio de unión para ARN polimerasa dependiente de ADN, sitios de iniciación de la transcripción y cualquier otra secuencia de ADN, incluyendo, entre otras, sitios de unión a factores de transcripción, sitios de unión a proteínas represoras y activadoras y cualquier otra secuencia de nucleótidos conocida por una persona experta en la técnica para actuar directamente o indirectamente para regular la cantidad de transcripción del promotor. El promotor no incluye el sitio de inicio de la transcripción (TSS), sino que termina en el nucleótido -1 del sitio de transcripción y no incluye secuencias de nucleótidos que se convierten en regiones no traducidas en el ARNm transcrito tales como la 5'-UTR. Los promotores de la invención pueden ser promotores específicos de tejido, preferidos de tejido, específicos de tipo celular, inducibles y constitutivos. Los promotores específicos de tejido son promotores que inician la transcripción solo en ciertos tejidos y se refieren a una secuencia de ADN que proporciona señales de reconocimiento para la ARN polimerasa y/u otros factores necesarios para que comience la transcripción, y/o para controlar la expresión de la secuencia codificante precisamente dentro de ciertos tejidos o dentro de ciertas células de ese tejido. La expresión de una manera específica de tejido puede ser solo en tejidos individuales o en combinaciones de tejidos. Los promotores preferidos de tejido son promotores que inician preferentemente la transcripción en ciertos tejidos. Los promotores específicos del tipo celular son promotores que impulsan principalmente la expresión en ciertos tipos de células. Los promotores inducibles son promotores que son capaces de activar la transcripción de una o más secuencias de ADN o genes en respuesta a un inductor. Las secuencias o genes de ADN no se transcribirán cuando el inductor esté ausente. La activación de un promotor inducible se establece mediante la aplicación del inductor. Los promotores

constitutivos son promotores activos bajo muchas condiciones ambientales y en muchos tipos de tejidos diferentes. Preferiblemente, la capacidad para iniciar la transcripción se establece en un sistema de expresión usando un constructo de expresión que comprende dicho promotor operativamente enlazado a una secuencia de nucleótidos de interés usando un ensayo adecuado tal como RT-qPCR o transferencia Northern. Se dice que un promotor es capaz de comenzar la transcripción si se puede detectar una transcripción o si se encuentra un aumento en el nivel de transcripción de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la transcripción usando un constructo que solo difiere en que está libre de dicho promotor. En una realización preferida adicional, la capacidad para iniciar la expresión se establece en un sistema de expresión usando un constructo de expresión que comprende dicho promotor operativamente enlazado a una secuencia de nucleótidos que codifica una proteína o un polipéptido de interés. Preferiblemente, dicha proteína o polipéptido de interés es una proteína o un polipéptido secretado y la expresión de dicha proteína o polipéptido de interés se detecta mediante un ensayo adecuado tal como un ensayo ELISA, una transferencia de Western o, dependiendo de la identidad de la proteína o polipéptido de interés, cualquier ensayo de identificación y/o cuantificación de proteínas adecuado conocido por el experto en la técnica. Se dice que un promotor es capaz de iniciar la expresión si se puede detectar la proteína o el polipéptido de interés o si se encuentra un aumento en el nivel de expresión de al menos un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 100 %, un 200 %, un 300 %, un 500 %, un 1000 %, un 1500 % o un 2000 % en comparación con la expresión usando un constructo que solo difiere en que está libre de dicho promotor. Se puede usar como primer y segundo promotor de la invención un promotor inducido o constitutivo o una combinación de los mismos en la presente invención.

[0159] Un «intrón» es una secuencia de nucleótidos dentro de un transcrito de ARN primario que se elimina mediante empalme de ARN o empalme de intrón mientras se genera el producto de ARN maduro final. Se puede evaluar si se produce el empalme de intrones utilizando cualquier método adecuado conocido por la persona experta en la materia, como por ejemplo, entre otros, la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) seguida de análisis de tamaño o secuencia del producto RT-PCR. Preferiblemente, una secuencia de nucleótidos es un intrón si al menos un 2 %, un 5 %, un 10 %, un 15 %, un 20 %, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 % o un 100 % del ARN primario pierde esta secuencia mediante empalme de ARN usando un ensayo adecuado para detectar el empalme de intrones tal como se ha indicado anteriormente. Preferiblemente, un intrón comprende un sitio de empalme GT en el extremo 5' de la secuencia de nucleótidos y un sitio de empalme AG en el extremo 3' de la secuencia de nucleótidos, cuyo sitio de empalme AG está precedido por una secuencia de nucleótidos rica en pirimidina o un tracto de polipirimidina, opcionalmente separado del sitio de empalme AG por 1-50 nucleótidos. Un intrón puede comprender además un sitio de ramificación que comprende la secuencia Y-T-N-A-Y en el lado 5' del tracto de polipirimidina. El sitio de ramificación puede tener la secuencia de nucleótidos C-Y-G-A-C. Se entiende que una «secuencia intrónica» es al menos parte de la secuencia de nucleótidos de un intrón.

[0160] Se entenderá que «expresión» incluye cualquier paso implicado en la producción del péptido o polipéptido, que incluye, entre otros, transcripción, modificación postranscripcional, traducción, modificación postraducciona y secreción.

[0161] Opcionalmente, un promotor representado por una secuencia de nucleótidos presente en un constructo de ácidos nucleicos está operativamente enlazado a otra secuencia de nucleótidos que codifica un péptido o un polipéptido tal como se identifica en el presente documento.

[0162] Un vector de expresión puede ser cualquier vector que pueda someterse convenientemente a procedimientos de ADN recombinante y puede provocar la expresión de una secuencia de nucleótidos que codifica un polipéptido de la invención en una célula y/o un sujeto.

[0163] Tal como se usa en el presente documento, el «5'-UTR» es la secuencia que comienza con el nucleótido 1 del ARNm y termina con el nucleótido -1 del codón de inicio. Es posible que una parte reguladora del promotor esté comprendida dentro de la secuencia de nucleótidos que se convierte en una 5'-UTR; sin embargo, en tal caso, la 5'-UTR todavía no es parte del promotor tal como se define en el presente documento.

[0164] El término "secuencias de control" se define en el presente documento para incluir todos los componentes, que son necesarios o beneficiosos para la expresión de un polinucleótido o un polipéptido. Cada secuencia de control puede ser nativa o extranjera para la secuencia de ácidos nucleicos que alberga o codifica el polinucleótido o el polipéptido. Dichas secuencias de control incluyen, entre otras, un líder, secuencias de iniciación de traducción óptimas (tal como se describe en Kozak, 1991, J. Biol. Chem 266: 19867-19870), una secuencia de poliadenilación, una secuencia de propéptido, una secuencia de prepropéptido, un promotor, una secuencia señal y un terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor, y señales de parada transcripcionales y traduccionales.

[0165] Las secuencias de control pueden proporcionarse con enlazadores con el fin de introducir sitios de

restricción específicos que faciliten la ligadura de las secuencias de control con la región codificante de la secuencia de nucleótidos que codifica un polipéptido.

[0166] La secuencia de control puede ser una secuencia promotora apropiada, una secuencia de ácidos nucleicos, que es reconocida por una célula huésped para la expresión de la secuencia de ácidos nucleicos. La secuencia promotora contiene secuencias de control transcripcionales que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de ácidos nucleicos que muestre actividad transcripcional en la célula, incluyendo promotores mutantes, truncados e híbridos, y puede obtenerse a partir de genes que codifican polipéptidos extracelulares o intracelulares, ya sean homólogos o heterólogos a la célula huésped.

[0167] La secuencia de control también puede ser una secuencia terminadora de transcripción adecuada, una secuencia reconocida por una célula huésped para terminar la transcripción. La secuencia terminadora está operativamente enlazada al terminal 3' de la secuencia de ácidos nucleicos de interés, que codifica preferiblemente un polipéptido de interés. Cualquier terminador que sea funcional en la célula puede usarse en la presente invención.

[0168] La secuencia de control también puede ser una secuencia líder adecuada, una región no traducida de un ARNm que es importante para la traducción por parte de la célula huésped. La secuencia líder está operativamente enlazada al terminal 5' de la secuencia de ácidos nucleicos de interés, que codifica preferiblemente un polipéptido de interés. Cualquier secuencia líder que sea funcional en la célula puede usarse en la presente invención.

[0169] La secuencia de control también puede ser una secuencia de poliadenilación, una secuencia operativamente enlazada al terminal 3' de la secuencia de ácidos nucleicos y que, cuando se transcribe, es reconocida por la célula huésped como una señal para añadir residuos de adenina al ARNm transcrito. Cualquier secuencia de poliadenilación que sea funcional en la célula puede usarse en la presente invención.

[0170] En este documento y en sus reivindicaciones, el verbo «comprender» y sus conjugaciones se utilizan en su sentido no limitante para indicar que se incluyen los elementos que siguen a la palabra, pero no se excluyen los elementos que no se mencionan específicamente. Además, el verbo «constar» puede reemplazarse por «constar esencialmente de» que significa que un producto o una composición o una molécula de ácidos nucleicos o un péptido o polipéptido de un constructo de ácidos nucleicos o un vector o una célula tal como se definen en el presente documento pueden comprender componente(s) adicional(es) a los identificados específicamente; dicho(s) componente(s) adicional(es) no altera(n) la característica única de la invención. Además, la referencia a un elemento por los artículos indefinidos «un» o «una» no excluye la posibilidad de que más de uno de los elementos esté presente, a menos que el contexto requiera claramente que haya uno y solo uno de los elementos. De este modo, los artículos indefinidos «un» o «una» significan generalmente «al menos uno».

Tabla 1: Identificación de secuencia

SEC ID n.º:	Descripción
1	Elemento potenciador de la expresión 1
2	Elemento potenciador de la expresión 2
3	UN1
4-13	secuencias derivadas de UN1
14	UN2
15	UN1dGAA
16	UN2dGAA
17	R3
18	fUN1
19	UN2-2
20	UN2-3
21	UN2-4
22	UN2-5
23	UN2-6

SEC ID n.º:	Descripción
24	UN2-7
25	UN2-8
26	UN2-9
27	UN2-10
28	UN1dGAA-2
29	UN1dGAA-3
30	UN1dGAA-4
31	UN1dGAA-5
32	UN1dGAA-6
33	UN2dGAA-2
34	UN2dGAA-3
35	UN2dGAA-4
36	UN1 barajado
37	UN1 barajado-2
38	UN1 barajado-3
39	UN1 barajado-4
40	UN1 barajado-5
41	UN1 barajado-6
42	UN2 barajado-1
43	UN2 barajado-2
44	CAA1
45	CAA2
46	CAA3
47	CAA4
48	CAA5
49	CAA6
50	TATA1
51	TATA2
52	Secuencia potenciadora del promotor de CMV
53	Región potenciadora de UBC
54	Secuencia potenciadora del promotor de CMV
55	constructo
56	constructo
57	Secuencia promotora de CMV
58	Secuencia promotora mínima de CMV
59	EEE1-Xt
60	EEE1-80

SEC ID n.º:	Descripción
61	EEE1-60
62	EEE1-50
63	EEE1-SL
64	cebador HC RACE
65	Secuencia de vector de cadena ligera
66	Secuencia de vector de cadena pesada
67	Cadena ligera de proteínas HuMab1
68	Proteína HuMab1 pesada
69	Cadena ligera de proteínas HuMab2
70	Proteína HuMab2 y cadena pesada
71	ADNpc3.1 (+)
72	Proteína SeAP
73	EEE1 + CMV + TEE
74	EEE1-Xt + CMV + TEE
75	EEE1-80 + CMV + TEE
76	EEE1-60 + CMV + TEE
77	pPNic384
78	inserto pPNic602
79	Promotor EF1a
80	EEE1-A1
81	EEE1-A2
82	EEE1-A3
83	EEE1-B1
84	EEE1-B2
85	EEE1-B3
86	EEE1-B4
87	EEE1-B5
88	Secuencia reguladora de la transcripción

Figuras

[0171]

5

10

15

Fig. 1. Mapa esquemático del constructo del promotor intrónico y diferentes transcritos. El constructo comprende 2 promotores. La transcripción por el promotor 1 da como resultado un transcrito primario que incluye el intrón que contiene la secuencia completa del promotor 2 y está bordeado por los sitios de empalme 5' y 3'. Después del empalme de intrones, dicho transcrito primario da como resultado un ARNm sin que dicho intrón (transcrito 1) codifique un «gen». La transcripción a partir del promotor 2 también da como resultado un ARNm (transcrito 2) que codifica el mismo «gen».

Fig. 2. Mapa esquemático de los elementos EEE1 (A) y EEE2 (B) que muestra algunas características de los genes de UBC y CCT8 relevantes para su actividad promotora en un contexto genómico. Las características incluyen el sitio de inicio de transcripción pronosticado (TSS), las 5'-UTR y la información de exón e intrón.

Fig. 3. Mapa esquemático de un vector de expresión para una cadena ligera de Ig (IgLC) con la secuencia EEE1 integrada aguas arriba del promotor de CMV.

Fig. 4. Comparación de la producción de HuMab1 mediante grupos de CHO-S transfectados de manera estable con constructos de referencia o EEE1. Vector de expresión sin (A) y con (B) el elemento regulador de expresión adicional. Las barras representan los títulos promedio de cuatro grupos derivados de dos transfecciones independientes. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 5. Comparación de la producción de HuMab1 mediante conjuntos de CHO-S transfectadas de manera estable con constructos de referencia o EEE2. Las barras representan los títulos promedio de cuatro grupos derivados de dos transfecciones independientes. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 6. Análisis de la producción de HuMab1 mediante las 10 líneas celulares clonales CHO-S principales transfectadas de manera estable con constructos EEE1-TEE que albergan EEE1. Las células se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 7. Comparación de la producción de HuMab2 mediante los grupos CHO-S transfectados de manera estable con EEE1 en el vector de referencia (A, panel izquierdo) y en el vector con un elemento regulador adicional (A, panel derecho). Las barras representan los títulos promedio de cuatro grupos derivados de dos transfecciones independientes. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO. (B) Análisis de la producción de HuMab2 mediante las 12 líneas celulares clonales principales de referencia y EEE1-TEE CHO-S. Las células se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 8. Comparación de la actividad de SeAP en los medios de los grupos CHO-S transfectados de manera estable con constructos de referencia o EEE1. Vector de expresión sin (panel izquierdo) y con (panel derecho) elemento regulador de expresión adicional. Las barras representan las actividades promedio de cuatro grupos derivados de dos transfecciones independientes, medidas utilizando el kit de ensayo de genes SEAP Reporter, Abcam. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 9. Comparación de la actividad de SeAP en los medios de los grupos CHO-S transfectados de manera estable con constructos que contienen diferentes versiones del elemento EEE1. Las barras representan las actividades promedio de cuatro grupos derivados de dos transfecciones independientes, medidas utilizando el kit de ensayo de genes SEAP Reporter, Abcam. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO.

Fig. 10. Amplificación 5'-RACE de los extremos 5' de transcritos de cadena pesada de clones de CHO-S transfectados de manera estable con constructos EEE1-TEE que expresan HuMab1 (A). Se detectan dos bandas en gel de agarosa, que corresponden a las transcripciones generadas por el promotor de CMV (transcrito 1) y el promotor de UBC (transcrito 2). La diferencia de tamaño entre el transcrito 1 y el transcrito 2 se explica en un mapa esquemático del constructo del promotor intrónico y diferentes transcritos. (B). El constructo comprende 2 promotores, UBC y CMV. El promotor de CMV está vinculado a una secuencia TEE que también comprende un intrón corto. La transcripción mediante el promotor de CMV da como resultado, después del empalme de intrones, ARNm con TEE como 5'-UTR (transcrito 1). El promotor de UBC está vinculado a una región parcial 5'-UTR de UBC que comprende un sitio de empalme donador 5' y precede al promotor de CMV. La transcripción por el promotor de UBC da como resultado ARNm con la secuencia 5'-UTR de UBC (transcrito 2). El intrón grande que se empalma a partir del transcrito primario se extiende desde la secuencia de empalme donadora 5' en la secuencia de UBC hasta el sitio de empalme aceptor 3' en TEE y contiene la secuencia completa de CMV.

Fig. 11. Efecto de EEE en *Pichia pastoris* que expresa interleucina humana recombinante 8. Cada barra representa la expresión promedio de 10 clones independientes.

[0172] La invención se explicará con más detalle en la siguiente sección de ejemplos, haciendo referencia a las figuras adjuntas. Los ejemplos sirven solo para fines ilustrativos y no pretenden limitar la presente invención de ninguna manera.

Ejemplos

[0173] El elemento potenciador de la expresión representado por la SEC ID n.º: 1 se basa en el gen de ubiquitina C (UBC) de hámster chino (*Cricetulus griseus*). Comprende la secuencia promotora pronosticada y parte de la región 5' no traducida (Fig. 2A). El elemento potenciador de la expresión representado por la SEC ID n.º: 2 se basa en el gen CCT8 humano (chaperonina que contiene TCP1, subunidad 8). Comprende la secuencia promotora pronosticada y parte de la región 5' no traducida, así como una secuencia corta que codifica 27 aminoácidos y parte del primer intrón (Fig. 2B).

Ejemplo 1

[0174] Los plásmidos de expresión se construyeron basándose en el vector de expresión ADNpc3.1 (SEC ID n.º: 71). El vector se modificó mediante la eliminación del f1-ori. Se insertaron secuencias codificantes para genes de cadena pesada (representada por una secuencia que tiene al menos un 96 % de identidad con la SEC ID n.º: 68) y cadena ligera (representada por una secuencia que tiene al menos un 99 % de identidad con la SEC ID n.º: 67) de IgG1 (HuMab1) en este vector (la secuencia codificante de cadena ligera se insertó en el vector representado por la SEC ID n.º: 65 y la secuencia codificante de cadena pesada se insertó en el vector representado por la

SEC ID n.º: 66), lo que dio como resultado los constructos de referencia. Para generar los vectores EEE1 (elemento potenciador de la expresión 1), la secuencia EEE1 (SEC ID n.º: 1) se insertó aguas arriba del promotor de CMV (SEC ID n.º: 57) (Fig. 3). EEE2 (SEC ID n.º: 2) se introdujo de manera similar, lo que dio como resultado los vectores EEE2. Se generó un vector con un elemento regulador de la expresión adicional reemplazando la 5'-UTR de ADNpc3.1 para la SEC ID n.º: 19 (elemento potenciador de la transcripción; TEE).

[0175] Las células CHO-S (Life Technologies) se mantuvieron según las instrucciones del fabricante. Se realizaron transfecciones duplicadas usando células 3E7, 50 µg de ADN linealizado y reactivo FreeStyle MAX (Life Technologies). Los grupos postransfección se dividieron en dos y se seleccionaron en medio CD FortiCHO suplementado con glutamina 8 mM y 800 µg/ml de G418. Se sembraron grupos seleccionados en 30 ml del mismo medio a una densidad de 3E5 células/ml en matraces de agitación de 125 ml. Los títulos de HuMab1 se determinaron mediante ELISA (Fig. 4). Los títulos de los grupos de referencia eran demasiado bajos para una determinación precisa. Esto indica que el anticuerpo está mal expresado. Los grupos EEE1 produjeron aproximadamente 1 µg/ml (Fig. 4A). Se observó un efecto similar en el vector con un elemento regulador adicional de la expresión. En este vector, la introducción de EEE1 aumentó la producción desde aproximadamente 0.2 µg/ml hasta 9-12 µg/ml en grupos estables de tres experimentos de transfección independientes (Fig. 4B). Estos datos muestran que un anticuerpo mal expresado puede expresarse a niveles significativamente más altos mediante la introducción del elemento EEE1.

Ejemplo 2

[0176] El efecto de introducir el elemento EEE2 se estudió en un vector que alberga un elemento regulador de la expresión adicional (véase el ejemplo 1, Fig. 4B). Las células CHO-S se transfectaron con la referencia o con los constructos de EEE2 tal como se ha descrito anteriormente. Los títulos de anticuerpos de los grupos de EEE2 transfectados de forma estable eran más de 20 veces más altos que los grupos de referencia (Fig. 5).

Ejemplo 3

[0177] El EEE1-TEE y los grupos de referencia generados previamente se sembraron en seis placas de 96 pocillos a una densidad de 0.5 células/pocillo en medio de selección CD FortiCHO. Las células de referencia mostraron un crecimiento deteriorado en comparación con los clones EEE1-TEE y, por lo tanto, no se produjo HuMab1. Los clones de EEE1-TEE mostraron un crecimiento normal y producción de HuMab1 (ver más adelante). Se evaluaron 100 líneas clonales EEE1-TEE para la producción de HuMab1 en placas de microtitulación. Los 10 clones con la productividad específica más alta se expandieron a matraces de agitación de 125 ml. Los clones se sembraron en 30 ml de medio de selección CD FortiCHO a una densidad de 3E5 células/ml y los títulos de HuMab1 se determinaron mediante ELISA (Fig. 6). Los clones produjeron hasta 0.25 µg/ml de HuMab1. Estos datos indican que el EEE1 puede facilitar la generación de líneas clonales y permite la generación de líneas clonales con niveles de expresión relevantes.

Ejemplo 4

[0178] Se determinó el número de copias de clones que comprenden EEE que expresan anticuerpos. El programa PrimerExpress (Life Technologies) se utilizó para diseñar cebadores y sondas Taqman específicos para las cadenas pesadas y ligeras de la microglobulina HuMab1 y β-2. Los cebadores se combinaron en un ensayo triplex Taqman para medir copias de genes en muestras de ADNg de clones y grupos de HuMab1 de EEE1-TEE. Los números de copias de genes se compararon con los títulos de HuMab1 (Tabla 2). Las líneas celulares clonales que producían títulos similares de HuMab1 tenían diferentes números de copias de genes de cadena ligera y pesada (clon 1 y 2). Además, los clones que producían títulos de HuMab1 muy diferentes tenían números de copias de genes similares (clones 3 y 4). En los grupos, se emparejaron números relativamente altos de genes de cadena ligera y pesada con niveles de expresión relativamente bajos. Estos datos (Tabla 2) indican que no hay correlación entre el EEE que comprende el número de copia del gen y los niveles de expresión de HuMab1.

Tabla 2: Títulos de IgG1 y números de copias de genes

	Título de IgG	CL	CP
Clon1	123.7	36.8	25.1
Clon2	118.2	1.7	1.7
Clon3	143.6	3.1	1.2
Clon4	7.2	4.6	0.7
Grupo	9.0	17.5	21.1

Ejemplo 5

[0179] Los genes de cadena pesada y ligera HuMab1 de los ejemplos anteriores fueron reemplazados por genes de cadena pesada y ligera (SEC ID n.º: 69 y 70) que codifican un anticuerpo biosimilar (HuMab2 derivado del número de acceso de DrugBank DB00072). Los constructos se usaron para generar grupos de CHO-S tal como se ha descrito anteriormente. Los títulos se determinaron usando ELISA. Los datos (6.3 µg/ml sin elemento potenciador) indican que este anticuerpo se produce a un nivel más alto que el anticuerpo de los ejemplos anteriores. Sin la introducción de ningún elemento regulador de la expresión adicional del EEE, se produjo un aumento de 3.7 veces (Fig. 7A, panel izquierdo), en el vector modificado el aumento es de 7 veces (Fig. 7A, panel derecho). Dado que el elemento regulador de expresión adicional da como resultado un aumento del 40 %, los datos también indican un efecto sinérgico entre el EEE y el elemento regulador de expresión adicional. Las líneas clonales se aislaron de los grupos de referencia y EEE1-TEE tal como se ha descrito anteriormente. Los mejores clones de EEE1-TEE produjeron títulos de HuMab2 tres veces más altos en comparación con los mejores clones de referencia (Fig. 7B). Estos datos indican que el elemento EEE1 se puede aplicar con éxito para mejorar la expresión de proteínas recombinantes a partir de líneas celulares estables.

Ejemplo 6

[0180] El gen de la cadena ligera HuMab1 de los constructos del ejemplo 1 se reemplazó por el gen que codifica la fosfatasa alcalina secretada (SeAP; SEC ID n.º: 72). Los constructos se usaron para generar grupos de CHO-S tal como se ha descrito anteriormente. La actividad de SeAP se midió en el medio usando el kit de ensayo de genes SEAP Reporter, Abcam. Los grupos de EEE1 mostraron una actividad dos veces mayor en comparación con los grupos de referencia (Fig. 8). En los grupos de EEE1-TEE, el aumento fue casi cuatro veces mayor en comparación con el grupo de referencia. Estos datos muestran que el EEE1 mejora la expresión de una proteína de subunidad única no anticuerpo en una línea celular transfectada.

Ejemplo 7

[0181] Todos los constructos de SeAP usados en el ejemplo 6 comprendían el promotor de CMV. Se hicieron dos variantes del vector TEE que contenían el promotor EF-1α humano en lugar de CMV (SEC ID n.º: 79). Los constructos se usaron para generar grupos de CHO-S tal como se ha descrito anteriormente. La actividad de SeAP se midió en el medio usando el kit de ensayo de genes SEAP Reporter, Abcam. El grupo de EEE1-TEE con EF-1α como promotor intrónico produjo una actividad de SeAP 2,8 veces mayor en comparación con el grupo de promotores EF-1α de referencia sin EEE1. Estos datos muestran que EEE1 mejora la expresión de una proteína en un constructo de promotor intrónico cuando el promotor intrónico no es el promotor de CMV, por ejemplo, el promotor EF-1α.

Ejemplo 8

[0182] El elemento EEE1 del vector de expresión de SeAP EEE1 fue reemplazado por las siguientes variantes: 1. EEE1-80 representado por la SEC ID n.º: 60 tiene un truncamiento de 290 pb desde el extremo 5'; 2. EEE1-60 representado por la SEC ID n.º: 61 con un truncamiento de 580 pb desde el extremo 5'; 3. EEE1-50 representado por la SEC ID n.º: 62 con un truncamiento de 725 pb desde el extremo 5'; 4. EEE1-Xt representado por la SEC ID n.º: 59 con una extensión de 800 pb desde la secuencia de UBC de *C. griseus* genómica en el extremo 5'; 5. El EEE1-SL (SEC ID n.º: 63) tiene los principales sitios de empalme pronosticados donadores y aceptores mutados. La actividad de SeAP en el sobrenadante de células transfectadas con el elemento EEE1 se estableció en el 100 % y disminuyó hasta un 39 % de actividad sin el elemento EEE1 (Fig. 9). Los truncamientos 5' de los constructos EEE1-80 y EEE1-60 disminuyeron gradualmente la actividad, pero seguían mostrando una actividad mejorada en comparación con el constructo sin EEE. El elemento EEE1-50 disminuyó la actividad de SeAP hasta el 40 %, que es similar a la del constructo sin EEE. El constructo EEE1-Xt mostró casi un 40 % más de actividad en comparación con el constructo EEE1. Los datos sugieren que las secuencias con más del 50% de identidad con EEE1 pueden funcionar como elementos potenciadores de la expresión. El constructo EEE1-Xt produjo casi un 40 % más de actividad en comparación con el constructo EEE1, que muestra que las secuencias potenciadoras adicionales residen en la región aguas arriba de la secuencia genómica de la que se tomó el EEE1. La actividad del elemento EEE1 se ve gravemente perjudicada por mutaciones de 4 nt del constructo EEE1-SL que evitan el empalme correcto de intrones, lo que da como resultado una reducción significativa en la expresión de SeAP en comparación con el constructo EEE1.

Ejemplo 9

[0183] El elemento EEE1 del vector de expresión de SeAP de EEE1 fue reemplazado por nueve variantes del elemento EEE1, que se pueden agrupar basándose en dos tipos diferentes de mutaciones. El primer tipo de variantes de EEE1 (EEE1-A) experimentó cambios dentro de los elementos EEE1 o EEE1-Xt con más del 30 por ciento de nucleótidos mutados, cada uno en otra de las tres regiones que constaban de al menos 244 pb. El segundo tipo de variantes de EEE1 (EEE1-B) también tenía el mismo tamaño que el elemento EEE1 (1449 pb)

con al menos un 96 por ciento de identidad de secuencia, con mutaciones que apuntaban a secuencias funcionales de la secuencia EEE1. Las diferentes mutaciones se enumeran en la tabla 3.

Tabla 3: Modificaciones de EEE1

Tipo A: más del 30 % de mutaciones en 3 regiones de EEE1				
Variante	SEC ID n.º:	Identidad con EEE1 ¹⁾	Región de EEE1 modificada	Región de tamaño modificado (pb)
EEE1-A1	80	71.6 % ²⁾	región promotora 5'	1526
EEE1-A2	81	95.0 %	región promotora 3'	244
EEE1-A3	82	81.8 %	región intrón	480
Tipo B: mutaciones que apuntan a dominios específicos de EEE1				
Variante	SEC ID n.º:	Identidad con EEE1 ¹⁾	Modificación de la secuencia de EEE1	
EEE1-B1	83	97.9 %	1: 7 nt cambiados en nt 144-152 2: 4 nt cambiados en nt 612-615 3: 4 nt cambiados en nt 667-670 4: 6 nt cambiados en nt 816-823 5: 5 nt cambiados en nt 1.106-1.112 6: 5 nt cambiados en nt 1.432-1.438	
EEE1-B2	84	96.5 %	50 mutaciones de pb sencillas = 50 % de los CG mutados en nt 227-1409; se mantienen los sitios de unión del factor de transcripción pronosticados.	
EEE1-B3	85	99.7 %	5 mutaciones de pb sencillas = 50 % de los CG mutados en nt 549-603	
EEE1-B4	86	96.5 %	50 mutaciones de pb sencillas = 50 % de los CG mutados en nt 227-1409; 8 sitios pronosticados para los factores de transcripción SP1, HSF y NFKB afectados.	
EEE1-B5	87	96.0 %	se realizaron 51 mutaciones de pb en 12 regiones con actividad de unión del factor de transcripción pronosticada que abarcan nt 105-1449.	
EEE1-B6 ³⁾	63	99.7 %	4 mutaciones de pb sencillas que eliminan los sitios de empalme donadores o aceptores pronosticados y conocidos, incluyendo el sitio donador conocido (nt 970), nt 545 y 552 en la región promotora, nt 1267 en la región intrón.	
¹⁾ Identidad calculada usando el algoritmo de alineación de nucleótidos EMBOSS con los siguientes parámetros: matriz DNFull con las siguientes penalizaciones de gap: abierto = 10; extensión = 0.5				
²⁾ % de identidad EEE1-A1 calculado en relación con EEE1-Xt				
³⁾ Esto se conoce como EEE1-SL en el ejemplo 8				

5

[0184] Se midió la actividad de SeAP en el sobrenadante de células transfectadas con las diferentes variantes. La actividad de las células con el elemento EEE1 se utilizó como referencia (100 %). Sin el elemento EEE1, la actividad fue del 24 % en este experimento. La actividad de SeAP de las células transfectadas con los constructos EEE1-A2 y EEE1-A3 se redujo a hasta el 75 % y el 48 % en relación con EEE1, respectivamente (Tabla 4). Esto es superior al 24 % de actividad observada con el constructo sin EEE en este experimento. El constructo EEE1-A1 disminuyó la actividad de SeAP hasta el 30 % en relación con el constructo EEE1-Xt en el que se basa, que sigue siendo superior al constructo sin EEE que produce solo el 18 % de la actividad de SeAP en relación con el constructo EEE1-Xt. Los datos muestran que las variantes de EEE1 con tan solo un 72 % de identidad global y un 50% de identidad local con la secuencia genómica de UBC pueden funcionar como elementos potenciadores de la expresión.

10

15

[0185] La actividad de SeAP de las células transfectadas con los constructos EEE1-B1 hasta B6 se redujo hasta en un 42% con respecto al constructo EEE1 (Tabla 4). Los datos muestran que las mutaciones en regiones con

una funcionalidad pronosticada en la actividad promotora intrónica del elemento EEE1 pueden limitar significativamente la capacidad de mejora de la expresión del elemento EEE1. Por ejemplo, la mutación de 4 nt implicados en el empalme de intrones dio como resultado una disminución del 38 % en los títulos de SeAP (EEE1-B6). La mutación de diferentes conjuntos de CpG también dio como resultado una disminución de los

5

Tabla 4: Actividad de SeAP de variantes de EEE1

Constructo	Actividad relativa a EEE1 (%) ¹⁾
Sin EEE1	24
EEE1	100
EEE1-A2	75
EEE1-A3	48
EEE1-B1	58
EEE1-B2	60
EEE1-B3	73
EEE1-B4	58
EEE1-B5	84
EEE1-B6	62
Constructo	Actividad relativa a EEE1-Xt (%) ¹⁾
Sin EEE1	18
EEE1-Xt	100
EEE1-A1	30

¹⁾ Los valores representan las actividades promedio de cuatro grupos derivados de dos transfecciones independientes, medidas con el kit de ensayo de genes SEAP Reporter, Abcam. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO

Ejemplo 10

10

[0186] Se cultivó un clon de EEE1-TEE de CHO-S del ejemplo 3 y las células se cosecharon en la fase logarítmica. El ARN total se aisló de las células usando el mini kit AllPrep DNA/RNA (Qiagen). El ADNc se sintetizó usando el kit 5' y 3' RACE de mRNA eucariótico Epicentre Exact Start. El ADNc de la primera cadena se amplificó usando el cebador RACE 5' del kit combinándolo con un cebador específico de cadena pesada (SEC ID n.º: 64) y la ADN polimerasa ZymoTaq. El producto de PCR se analizó en geles de agarosa al 1.2 %, mostrando dos bandas discretas (Fig. 10A) que se aislaron por separado y se insertaron en un vector PCR4-TOPO (Life Technologies). El análisis de secuenciación reveló que la banda superior vista en el gel de agarosa corresponde al transcrito iniciado desde el promotor de CMV. La banda inferior corresponde al transcrito iniciado desde el promotor UBC. Ambos productos tienen la secuencia intrónica pronosticada correctamente dividida. Las diferencias de tamaño corresponden a las diferentes longitudes de las 5'-UTR, tal como se muestra en la figura 10B. Los datos muestran que ambos promotores contribuyen a la transcripción.

15

20

Ejemplo de referencia 11

25

[0187] Los grupos de CHO-S transfectados de forma estable con constructos de tres promotores individuales diferentes se compararon mediante la actividad de SeAP en el sobrenadante. Los grupos se cultivaron en matraces de agitación de 125 ml en 30 ml de medio de selección CD FortiCHO y la actividad de SeAP se midió usando el kit de ensayo de genes SEAP Reporter (Abcam) en 4 grupos por constructo derivados de 2 transfecciones independientes. Los constructos contenían el promotor de CMV (ejemplo 6), el promotor EF-1 α (ejemplo 7) o el promotor de UBC (Ejemplo 11). El promotor de UBC produjo una actividad de SeAP 2.7 veces mayor en comparación con el constructo del promotor de CMV. El promotor de UBC produjo una actividad de SeAP seis veces mayor en comparación con el promotor EF-1 α . Los datos muestran que la expresión con el promotor de UBC en solitario es mayor en comparación con el promotor de CMV o el promotor EF-1 α en solitario.

30

35

Ejemplo 12 Secreción inducida por metanol de IL8 en transformantes integradores GS115 de Pichia pastoris

[0188] Los plásmidos para la transformación estable de Pichia pastoris con constructos de expresión de

interleucina 8 humana (hIL-8) se generaron en el plásmido pPIC9K (Life Technologies). La inserción del gen hIL-8 en pPIC9K dio como resultado el plásmido pPNic384 (SEC ID n.º: 77), que contiene el gen hIL-8 bajo el control del promotor AOX1. La secuencia EEE1 se insertó aguas arriba del promotor AOX1 como un fragmento de AatII-Alel (SEC ID n.º: 78) en pPNic384, lo que dio como resultado el plásmido pPNic602.

[0189] Los vectores de expresión se linealizaron mediante digestión con Sa1I y se transformaron en la cepa GS115 de *P. pastoris* usando la electroporación según se recomienda (Invitrogen, 2008). Los transformantes se colocaron en placas de agar en RDB (medio de dextrosa de regeneración, un medio que carece de histidina). Después de la incubación a 30 °C durante 48 h, se observaron grandes colonias. Se realizó una transformación de control sin ADN que dio como resultado una ausencia de colonias. Se recogieron al azar de la placa de transformación 10 clones por constructo y se cultivaron hasta la saturación en 800 µl de BMG (medio mínimo tamponado con glicerol al 1 %), en placas de pocillos profundos de 2 ml. La placa se mantuvo en una incubadora con agitación (Infors-HT Microton) a 30 °C, 1000 rpm durante 18 horas. La densidad óptica del cultivo estaba entre 5-10 unidades de absorbancia a 600 nm. Las células se cosecharon y el medio se reemplazó por 800 µl de BMM (medio mínimo tamponado con metanol al 0.5 %) en placas de pocillos profundos de 2 ml. Las células se cultivaron en la incubadora con agitación y cada 24 horas se añadió metanol al 0.5 % (concentración final) al cultivo para mantener la inducción. Después de 72 horas de inducción de metanol, se recogió el sobrenadante de cultivo y se ensayó el rendimiento secretado de hIL8 usando el kit de hIL8 AlphaLISA (Perkin Elmer). Los datos muestran (Fig. 11) que existe una diferencia significativa entre el rendimiento de IL8 de la referencia y los transformantes EEE1, lo que sugiere que la secuencia EEE1 aguas arriba del promotor mejora el rendimiento de hIL8 en comparación con el plásmido de expresión sin la secuencia EEE1.

LISTADO DE SECUENCIAS

[0190]

<110> R1 B3 Holding B.V.
 <120> Constructo y secuencia para la expresión genética mejorada
 <130> P6045820PCT
 <150> EP 13199873
 <151> 2013-12-31
 <150> EP 13199875
 <151> 2013-12-31
 <160> 88
 <170> versión de PatentIn 3.3
 <210> 1
 <211> 1449
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1

ES 2 780 692 T3

<400> 1
 tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
 gcttcaaagg gttcaatctc ttaggttagct aaatatgaac aaaatttggg aaatgtgacc 120
 ttttccttag tgacagtcag atagaacctt ctcgagtgcaggacaccaa gtgcaaacag 180
 gctcaagaac agcctggaaa ggtctagtgc tatggggcct caggtcgaat gccaaactgtt 240
 ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
 cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
 gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480
 tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540
 ggggtaagcg ggtgggtgtc tggggcgcgga agccgactg cgcatgcgcc gaggtccgct 600
 ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660
 cacgctgata cctccgcggg gccgcacagg gtcacagcct tccccctccc caaaaagccc 720
 cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca 780
 ggcgcctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttccgc 840
 ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcgggtc 900
 ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960
 ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcg 1020
 ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc gtggcccgcg aggaaggcga 1080
 ctgtcctgag gcgaggacc cagcggcaag atggcgcca agtggaagcc tgaggggata 1140
 ggcgagcggc cctgagggcgc tcgacggggt tgggggggaa gcaggcccgc gaggcagctg 1200
 cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc cgtaggtgcc 1260
 tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt 1320
 tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttg tggcgggtga 1380
 acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgagga 1440
 catttccct 1449
 <210> 2
 <211> 1228
 <212> ADN
 <213> artificial
 <220>
 <223> EEE2

5

10

ES 2 780 692 T3

	<400> 2	
	gtaaagcaga tcacacagaa tatggcacac ttgagcactt gatgtgtact acattactct	60
	tagtgacgac ttttaattatc gtgcgcattc ccagcgcttc ctatggtgcc caacacagag	120
	cggacgccta gagacaattt tgggggatgg ggcagatgct ctgcctcggg aaaaaaaaaag	180
	cacacctgcc ctgacgttgg tggctgggtc tggaagatac gtggaaatta agctaaggat	240
	gtgtggcttc cagatcaaaa accgcaaaaa tctaacgccg tgactactga ctacggtcag	300
	agagcacaga ctggagcaac ctctcacggc ctgggctgtc tgcgcgtgcg tgagccagaa	360
	acccgagggg ctccctgggc ccgccctatc gatcgacctg atcggggatc gtcagcttgg	420
	ttctggccac agaggttgct cttctcgcga tgcttcagac ctggcggcag ggaaagggtg	480
	ggctaattgg agagccagga agagcgtgag gcggcccccac gctgctttcc cagaaggctg	540
	tgcgtgctcc tcgcttcctc cgcggtcttc cgagcggtcg cgtgaactgc ttccagcagg	600
	ctggccatgg cgcttcacgt tcccaaggct ccgggctttg cccagatgct caaggaggga	660
	gcgaaagtaa gggctgaagg aaaggaatga ggtgggagcg tcagcatagg gctgcggcgg	720
	cggcggcgaa gtaggagggc ctactaacgg gctgagcgtg ctgccctggc tcagcggccg	780
	ggggaagaga agattccaga aaggaggtg attttggaag ggctcggcca ccggagcctg	840
	cgggcacttc tcttcttcg cgaccgggag aaggccgagg gatcggcggc acgatcgaca	900
	ttgtacacct tgaaggtgga cggatgtgaa gccgcgcgtg cgttttgcct ccatccgtaa	960
	atgggggctaa ggcccgtcac ccttaaagga ggttgtgagg gtgaaattga ataacgtaga	1020
	tgaaattgtc ttgagaactg cgacgtcgat tatcacatag ctcgcgagtt gtaggatggg	1080
	gaagaacgag aactagccga tccagagaag agagtgggaa aaagggccgg gtcttggttg	1140
	cttgcttccc agtgagaaac atacggcttt cagcttagtt gacagaagcc atgcgttgta	1200
5	gccaaatgag ttccggtccc aacttatg	1228
	<210> 3	
	<211> 173	
	<212> ADN	
10	<213> artificial	
	<220>	
	<223> TEE	
	<400> 3	
	caagctctag caggaagaag aaataagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aataaaactc ccaaaaaaaaa gaaatcatc aaaaaaacia	120
	atttcacaaaa gagtttttgt gtttggggat taaagaataa aaaaaacaac gtc	173
15	<210> 4	
	<211> 175	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> TEE	

ES 2 780 692 T3

5	<400> 4	
	caagctctag caggaagaag aaataagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aataaaactc ccaaaaaaaaa gaaaatcatc aaaaaaacia	120
	atttcaaaaa gagtttttgt gtttgggat taaagaataa aaaaaacaac gtccc	175
	<210> 5 <211> 189 <212> ADN <213> artificial <220> <223> TEE <400> 5	
10	agatcactag aagcttcaag ctctagcagg aagaagaaat aagaagaaga agaagaagaa	60
	gaagaagcgt ctctcttct tcttgtgaga gtaaaaaata aaactcccaa aaaaaagaaa	120
	atcatcaaaa aaacaaatct caaaaagagt ttttgtgttt ggggattaaa gaataaaaaa	180
	aacaacgcc	189
	<210> 6 <211> 189 <212> ADN <213> artificial <220> <223> TEE <400> 6	
15	agatcactag aagcttcaag ctctagcagg aagaagaaag aagaagaaga agaagaagaa	60
	gaagaagcgt ctctcttct tcttgtgaga gtaaaaaaga aaactcccaa aaaaaagaaa	120
	atcatcaaaa aaacaaatct caaaaagagt ttttgtgttt ggggattaaa gaagaaaaaa	180
	aacaacgcc	189
	<210> 7 <211> 191 <212> ADN <213> artificial <220> <223> TEE <400> 7	
20	agatcactag aagcttcaag ctctagcagg aagaagaaat aagaagaaga agaagaagaa	60
	gaagaagcgt ctctcttct tcttgtgaga gtaaaaaata aaactcccaa aaaaaagaaa	120
	atcatcaaaa aaacaaatct caaaaagagt ttttgtgttt ggggattaaa gaataaaaaa	180
	aacaacagggc c	191
	<210> 8 <211> 284 <212> ADN <213> artificial <220> <223> TEE <400> 8	
25	ctttttcgca acgggtttgc cgccagaaca caggtgtcgt gaggaattag cttggtacta	60
	atacgactca ctataggag acccaagctg gctaggttaag cttggtaccc aagctctagc	120
	aggaagaaga aataagaaga agaagaagaa gaagaagaag cgtctcctct tcttcttctg	180
	agagtaaaaa ataaaactcc caaaaaaaag aaaatcatca aaaaaacaaa tttcaaaaag	240
	agtttttctg tttggggatt aaagaataaa aaaaacaacg tccc	284
30		

5	<210> 9		
	<211> 230		
	<212> ADN		
	<213> artificial		
	<220>		
10	<223> TEE		
	<400> 9		
	aacccactgc ttactggctt atcgaaatta atacgactca ctatagggag acccaagctc	60	
	tagcaggaag aagaaataag aagaagaaga agaagaagaa gaagcgtctc ctcttcttct	120	
	tgtgagagta aaaaataaaa ctcccaaaaa aaagaaaatc atcaaaaaaa caaatattcaa	180	
15	aaagagtttt tgtgttttggg gattaaagaa taataaaaaac aacctccacc	230	
	<210> 10		
	<211> 177		
	<212> ADN		
	<213> artificial		
20	<220>		
	<223> TEE		
	<400> 10		
	caagctctag cagcaacaac aaataacaac aacaacaaca acaacaacaa gcgtctcctc	60	
	ttcttcttgt gagagtataa aataaaactc ccaaaaaaaaa gaaatcatc aaaaaaaca	120	
25	atttcaaaaa gagtttttgt gtttggggat taaagaataa aaaaaacaac ctccacc	177	
	<210> 11		
	<211> 191		
	<212> ADN		
	<213> artificial		
30	<220>		
	<223> TEE		
	<400> 11		
	agatcactag aagcttcaag ctctagcagg aagaagaaat aagaagaaga agaagaataa	60	
	gaagaagcgt ctcttcttct tcttgtgaga gtaaaaaata aaactcccaa aaaaaataaa	120	
35	atcatcaaaa aaagaaattt caaaaagagt ttttgtgttt ggggattaaa gaataaaaaa	180	
	aacaacaggc c	191	
	<210> 12		
	<211> 189		
	<212> ADN		
40	<213> artificial		
	<220>		
	<223> TEE		
	<400> 12		
	agatcactag aagcttcaag ctctagcagg aagaagaaat aataagaaga agaagaataa	60	
45	gaagaagcgt ctcttcttct tcttgtgaga gtaaaaaata aaactcccaa aaaaaataaa	120	
	atcatcaaaa aaataaattt caaaaagagt ttttgtgttt ggggattaaa gaataaaaaa	180	
	aacaacgcc	189	
	<210> 13		
	<211> 173		
50	<212> ADN		
	<213> artificial		
	<220>		
	<223> TEE		

ES 2 780 692 T3

	<400> 13	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaacia	120
	atttcaaaaa gagtttttgt gtttgggat taaagaakaa aaaaaacaac gtc	173
5	<210> 14	
	<211> 274	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> UN2	
	<400> 14	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaacia	120
	atttcaaaaa gagtttttgt gtttgggat taaagaakaa aaaaaacaac aggtgagtaa	180
	gcgcagttgt cgtctcttgc ggtgccgttg ctggttctca caccttttag gtctgttctc	240
10	gtcttcogtt ctgactctct ctttttcggt gcag	274
	<210> 15	
	<211> 123	
	<212> ADN	
	<213> artificial	
15	<220>	
	<223> UNldGAA	
	<400> 15	
	ggcgtctcct cttcttcttg tgagagtaaa aaataaaact ccaaaaaaaaa aaaaaatcat	60
	caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka aaaaaacaac	120
	gtc	123
20	<210> 16	
	<211> 257	
	<212> ADN	
	<213> artificial	
	<220>	
25	<223> UN2dGAA	
	<400> 16	
	aaagtatcaa caaaaaagct tcgtctcctc ttcttcttgt gagagtaaaa aakaaaactc	60
	caaaaaaaaa kaaatcatc aaaaaaacia atttcaaaaa gagtttttgt gtttgtaagt	120
	caggactcta gctttctact gtagtatact ctaaaggact gctgttctgt gcacccctt	180
	cctttgttta tcatagcgca cgacaagagt actaactaat taacttaggg ggattaaaga	240
	akaaaaaaaa caacaaa	257
30	<210> 17	
	<211> 173	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> R3	
	<400> 17	
	caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact ccaaaaaaaaa	60
	akaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka	120
35	aaaaaaacia ggaagaagaa akaagaagaa gaagaagaag aagaagaagc ctc	173
	<210> 18	
	<211> 384	

5	<212> ADN		
	<213> artificial		
	<220>		
	<223> fUN1		
	<400> 18		
	caagctctag caggaagaag aaataagaag aagaagaaga agaagaagaa gcgtctcctc	60	
	ttcttcttgt gagagtataa aataaaactc ccaaaaaaaaa gaaatcatc aaaaaaaca	120	
	atttcaaaaa gagtttttgt gtttggggat taaagaataa aaaaaacaac gtctggacaa	180	
	accacaacta gaatgcagtg aaaaaaatgc tttatttgtg aaatttgtga tgctattgct	240	
	ttatttgtaa ccattataag ctgcaataaa caagttaaca acaacaattg cattcatttt	300	
10	atgtttcagg ttcaggggga ggtgtgggag gtttttttaa gcaagtaaaa cctctacaaa	360	
	tgtggtaaaa tcgataagga tccg	384	
	<210> 19		
	<211> 277		
	<212> ADN		
	<213> artificial		
	<220>		
	<223> UN2-2		
	<400> 19		
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60	
15	ttcttcttgt gagagtataa aakaaaactc ccaaaaaaaaa gaaatcatc aaaaaaaca	120	
	atttcaaaaa gagtttttgt gtttggggat taaagaaga aaaaaacaac aggtgagtaa	180	
	gcgcagttgt cgtctcttgc ggtgccgttg ctggttctca caccttttag gtctgttctc	240	
	gtcttccggt ctgactctct ctttttcggt gcaggcc	277	
	<210> 20		
	<211> 252		
	<212> ADN		
	<213> artificial		
	<220>		
	<223> UN2-3		
20	<400> 20		
	aagctctagc aggaagaaga aakaagaaga agaagaagaa gaagaagaag cgtctcctct	60	
	tcttcttgtg agagtataaa akaaaactcc caaaaaaak aaaatcatca aaaaaacaaa	120	
	tttcaaaaag agtaggtaag attatctctt cccaaaattg attacttttt tattgaacaa	180	
	ttattaacca atcatggctt aacgaaaaac aggttttgtg tttggggatt aaagaakaaa	240	
	aaaaacaaaa ca	252	
	<210> 21		
	<211> 254		
	<212> ADN		
	<213> artificial		
25	<220>		
	<223> UN2-4		

ES 2 780 692 T3

	<400> 21	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaaca	120
	atttcaaaaa gagtaggtaa gattatctct tcccaaaatt gattactttt attattgaac	180
	aattactaac atttcatggc ttaacgaaaa acagggttttg tgtttgggga ttaaagaaka	240
	aaaaaaaca aaca	254
5	<210> 22	
	<211> 266	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> UN2-5	
	<400> 22	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaaca	120
	atttcaaaaa gagtgaggta agattatcga tatttaaatt atttatttct tcttttccat	180
	ttttttggct aacattttcc atgggttttat gatatcatgc aggtacgttt tgtgtttggg	240
10	gattaaagaa kaaaaaaaaac aaaaca	266
	<210> 23	
	<211> 266	
	<212> ADN	
	<213> artificial	
15	<220>	
	<223> UN2-6	
	<400> 23	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaaca	120
	atttcaaaaa gagtgaggta agattatcga tatttaaatt atttatttct tcttttccat	180
	ttttttggct aacattttcc taggttttat tatatctagc aggtacgttt tgtgtttggg	240
	gattaaagaa kaaaaaaaaac aaaaca	266
20	<210> 24	
	<211> 265	
	<212> ADN	
	<213> artificial	
	<220>	
25	<223> UN2-7	
	<400> 24	
	aagctctagc aggaagaaga aakaagaaga agaagaagaa gaagaagaag cgtctcctct	60
	tcttcttgtg agagtaaaaa aaaaaactcc caaaaaaaak aaaatcatca aaaaaacaaa	120
	tttcaaaaag agtgaggtaa gattatcgat atttaaatta tttatttctt cttttccatt	180
	tttttggcta acattttcct aggttttatt atatctagca ggtacgtttt gtgtttgggg	240
	attaaagaak aaaaaaaca aaaca	265
30	<210> 25	
	<211> 266	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> UN2-8	

ES 2 780 692 T3

	<400> 25	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaacia	120
	atttcaaaaa gagtgaggta agattatcga tatttaaatt atttatttct tcttttccat	180
	ttttttggct aacattttcc taggttttat tataatctagc aggtacgttt tgtgtttggg	240
	gattaaagaa kaaaaaaaaac aaaacc	266
5	<210> 26 <211> 287 <212> ADN <213> artificial <220> <223> UN2-9 <400> 26	
	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaacia	120
	atttcaaaaa gagtttttgt gtttgtaagt caggactcta gctttctact gtagtatcct	180
	ctaaaggact gctgttctgt gcacccctt cctttgttta tcatagcgca cgacaagagt	240
10	actaactaat taacttaggg ggattaaaga akaaaaaaaaa caacaaa	287
	<210> 27 <211> 251 <212> ADN <213> artificial <220> <223> UN2-10 <400> 27	
15	caagctctag caggaagaag aaakaagaag aagaagaaga agaagaagaa gcgtctcctc	60
	ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc aaaaaaacia	120
	atttcaaaaa gagtttttgt gtttgggtaa gtaattgcct tactcggaaa ataataatc	180
	atcatactaa cgcaagaggc gctgatattg cggttatata gggattaaag aakaaaaaaaa	240
	acaacgtcac c	251
20	<210> 28 <211> 143 <212> ADN <213> artificial <220> <223> UN1dGAA-2 <400> 28	
25	aaagtatcaa caaaaaagct tcgtctcctc ttcttcttgt gagagtaaaa aakaaaactc	60
	caaaaaaaaa kaaatcatc aaaaaaacia atttcaaaaa gagtttttgt gtttggggat	120
	taaagaakaa aaaaaacaac aaa	143
30	<210> 29 <211> 123 <212> ADN <213> artificial <220> <223> UN1dGAA-3	

ES 2 780 692 T3

	<400> 29	
	tcgtctcctc ttcttcttgt gagagtaaaa aakaaaactc ccaaaaaaaaa kaaatcatc	60
	aaaaaaaaa atttcaaaaa gagtttttgt gtttggggat taaagaakaa aaaaaacaac	120
	gtc	123
5	<210> 30	
	<211> 134	
	<212> ADN	
	<213> artificial	
	<220>	
	<223> UN1dGAA-4	
	<400> 30	
	caagctctag cacgtctcct ctcttcttgt tgagagtaaa aaakaaaact ccaaaaaaaaa	60
	akaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka	120
10	aaaaaaaaa cgcc	134
	<210> 31	
	<211> 140	
	<212> ADN	
15	<213> artificial	
	<220>	
	<223> UN1dGAA-5	
	<400> 31	
	ggcaagctct agcacgtctc ctcttcttct tgtgagagta aaaaakaaaa ctcccaaaaa	60
	aaakaaatc atcaaaaaa caaatttcaa aaagagtttt tgtgtttggg gattaaagaa	120
	kaaaaaaaaac aacctccacc	140
20	<210> 32	
	<211> 130	
	<212> ADN	
	<213> artificial	
	<220>	
25	<223> UN1dGAA-6	
	<400> 32	
	caagcgtctc ctcttcttct tgtgagagta aaaaakaaaa ctcccaaaaa aaakaaatc	60
	atcaaaaaa caaatttcaa aaagagtttt tgtgtttggg gattaaagaa kaaaaaaaaac	120
	aacctccacc	130
	<210> 33	
	<211> 178	
	<212> ADN	
30	<213> artificial	
	<220>	
	<223> UN2dGAA-2	
	<400> 33	
	cgtctcctct tcttcttgtg agagtaaaaa akaaaactcc caaaaaaak aaaatcatca	60
	aaaaaaca aa tttcaaaaag agtttttgtg tttggggatt aaagaakaaa aaaaacaacc	120
	tcgtgcggtg tgccgattcg cgtacgaata cgccttgtgc tgacacttct gtagcacc	178
35	<210> 34	
	<211> 238	
	<212> ADN	
	<213> artificial	
	<220>	
40	<223> UN2dGAA-3	

ES 2 780 692 T3

5	<400> 34 caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa 60
	akaaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka 120
	aaaaaaacaa caggtgagta agcgcagttg tcgtctcttg cggtgccgtt gctggttctc 180
	acacctttta ggtctgttct cgtcttccgt tctgactctc tctttttcgt tgcaggcc 238
	<210> 35 <211> 229 <212> ADN <213> artificial <220> <223> UN2dGAA-4 <400> 35
10	atcaagctct agcacgtctc ctcttcttct tgtgagagta aaaaakaaaa ctcccaaaaa 60
	aaakaaaatc atcaaaaaaa caaatttcaa aaagagttag gtaagattat cgatatttaa 120
	attattttatt tcttcttttc catttttttg gctaacattt tcctagggtt tattatatct 180
	agcaggtagc ttttgtgttt ggggattaaa gaakaaaaaa aacaaaaca 229
	<210> 36 <211> 188 <212> ADN <213> artificial <220> <223> UN1shuffle <400> 36
15	caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa 60
	akaaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka 120
	aaaaaaacaa ggaagaagaa akaagaagaa gaagaagaag aagaagaagc gccgcccgcc 180
	ccttcacc 188
	<210> 37 <211> 173 <212> ADN <213> artificial <220> <223> UN1shuffle-2 <400> 37
20	caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa 60
	akaaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka 120
	aaaaaaacaa ggaagaagaa akaagaagaa gaagaagaag aagaagaagc gcc 173
	<210> 38 <211> 154 <212> ADN <213> artificial <220> <223> UN1shuffle-3 <400> 38
	gaagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa 60
30	akaaaatcat caaaaaaaca aatttcaaaa agagtttttg tgtttgggga ttaaagaaka 120
	aaaaaaacaa ggaagaagaa gaagaagaag cgcc 154
	<210> 39 <211> 179 <212> ADN <213> artificial
35	

ES 2 780 692 T3

	<220>		
	<223> UN1shuffle-4		
	<400> 39		
	ggcaagctct agcacgtctc ctcttcttct tgtgagagta aaaaakaaaa ctccccaaaa	60	
	aaakaaaatc atcaaaaaaa caaatattcaa aaagagtttt tgtgtttggg gattaaagaa	120	
5	kaaaaaaac aaggaagaag aaakaagaag aagaagaaga agaagaagaa gcctccacc	179	
	<210> 40		
	<211> 160		
	<212> ADN		
	<213> artificial		
10	<220>		
	<223> UN1shuffle-5		
	<400> 40		
	ggcaagctct agcacgtctc ctcttcttct tgtgagagta aaaaataaaa ctccccaaaa	60	
	aaagaaaatc atcaaaaaaa caaatattcaa aaagagtttt tgtgtttggg gattaaagaa	120	
	taaaaaaaac aaggaagaag aagaagaaga agcctccacc	160	
	<210> 41		
	<211> 177		
15	<212> ADN		
	<213> artificial		
	<220>		
	<223> UN1shuffle-6		
	<400> 41		
	caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa	60	
	akaaaatcat caaaaaaaca aattttcaaaa agagtttttg tgtttgggga ttaaagaaka	120	
20	aaaaaaacaa ggaagaagaa akaagaagaa gaagaagaag aagaagaagc ctccacc	177	
	<210> 42		
	<211> 277		
	<212> ADN		
	<213> artificial		
25	<220>		
	<223> UN2shuffle-1		
	<400> 42		
	caagctctag cacgtctcct cttcttcttg tgagagtaaa aaakaaaact cccaaaaaaa	60	
	akaaaatcat caaaaaaaca aattttcaaaa agagtttttg tgtttgggga ttaaagaaka	120	
	aaaaaaacaa ggaagaagaa akaagaagaa gaagaagaag aagaagaagc aggtgagtaa	180	
	gcgagtttgt cgtctcttgc ggtgccgttg ctggttctca caccttttag gtctgttctc	240	
	gtcttccggt ctgactctct ctttttcggt gcaggcc	277	
30	<210> 43		
	<211> 267		
	<212> ADN		
	<213> artificial		
	<220>		
35	<223> UN2shuffle-2		
	<400> 43		
	atcaagctct agcacgtctc ctcttcttct tgtgagagta aaaaakaaaa ctccccaaaa	60	
	aaakaaaatc atcaaaaaaa caaatattcaa aaagagttag gtaagattat cgatatttaa	120	
	attatttatt tcttcttttc cttttttttg gctaacattt tcctaggttt tattatatct	180	
	agcaggtacg ttttgtgttt ggggattaaa gaakaaaaaa aacaaggaag aagaakaag	240	
	aagaagaaga agaagaagaa gaaaaca	267	

ES 2 780 692 T3

[illegible]

	<400> 48		
	caacctctac caccaacaac aaacaacaac aacaacaaca acaacaacaa ccctctccac	60	
	atctccctct cagagtaaaa aacaaaattg acaaaaaaaaa gattttataa taaaaacaaa	120	
	tttcaaaaag aattcaactc attcaatatt acaacaagaa caaaggaggt cacat	175	
5	<210> 49 <211> 177 <212> ADN <213> artificial <220> <223> CAA6 <400> 49		
	caagctctac caccaagaac aaacaacaac aacatatata aaacaacaac caccatctcc	60	
	tcttcttctt gtcaactcca aatcaaaact ccaaaaaaaaa agcaaatcat caaaaccaca	120	
10	aatttcaaac aacaacaaca aacaacaaac aacattaaca tcatatcaac ctccacc	177	
	<210> 50 <211> 175 <212> ADN <213> artificial <220>		
15	<223> TATA1 <400> 50		
	caagctctag caggaagaag aaataagaag aagaagaaga agaagaagaa gcgtctcctc	60	
	ttcttcttga cagagtaaaa aataactttt ataataaaga aaatcatcaa aaaaacaaat	120	
	ttcaaaaaga gtttttgtgt ttggggatta aagaataaaa aaaaggaggt cacat	175	
20	<210> 51 <211> 177 <212> ADN <213> artificial <220>		
25	<223> TATA2 <400> 51		
	caagctctag caggaagaag aaataagaag aagtatataa aagaagaaga agcgtctcct	60	
	cttcttcttg tgaagtaaaa aataaaactc ccaaaaaaaaa gaaaatcatc aaaaaacaa	120	
	atttcaaaaa gagtttttgt gtttggggat taaagaataa aaaaaacaac ctccacc	177	
30	<210> 52 <211> 315 <212> ADN <213> Artificial <220> <223> fragmento <400> 52		
	ggagttccgc gttacataac ttacggtaaa tggcccgctt ggctgaccgc ccaacgaccc	60	
	ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca	120	
	ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtac atcaagtgt	180	
	tcatatgcc agtacgcccc ctattgacgt caatgacggt aaatggcccg cctggcatta	240	
	tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg tattagtc	300	
35	cgctattacc atggt	315	
	<210> 53 <211> 303 <212> ADN <213> Artificial		

ES 2 780 692 T3

	<220>	
	<223> fragmento	
	<400> 53	
	ggcctccgcg ccgggttttg gcgcctcccg cgggcgcccc cctcctcacg gcgagcgctg	60
	ccacgtcaga cgaagggcgc agcgagcgtc ctgatccttc cgcccggacg ctcaggacag	120
	cggcccgcgtg ctcataagac tcggccttag aaccccagta tcagcagaag gacattttag	180
	gacgggactt gggtgactct agggcactgg ttttctttcc agagagcgga acaggcgagg	240
	aaaagtagtc ccttctcggc gattctgcgg agggatctcc gtggggcggt gaacgccgat	300
5	gat	303
	<210> 54	
	<211> 305	
	<212> ADN	
	<213> Artificial	
10	<220>	
	<223> fragmento	
	<400> 54	
	cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt	60
	gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca	120
	atgggtggag tatttacggg aaactgccc cttggcagta catcaagtgt atcatatgcc	180
	aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	240
	catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	300
	catga	305
	<210> 55	
	<211> 1428	
	<212> ADN	
	<213> Artificial	
15	<220>	
	<223> fragmento	

ES 2 780 692 T3

<400> 55
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120
atgggtggag tatttacggt aaactgccc a ttggcagta catcaagtgt atcatatgcc 180
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240
catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac 300
catgaattgg tttgatctga ttataaccta ggtcgaggaa ggtttcttca actcaaattc 360
atccgcctga taattttctt atattttcct aaagaaggaa gagaagcgca tagaggagaa 420
gggaaataat tttttaggag cttttcttac ggctatgagg aatttggggc tcagttgaaa 480
agcctaaact gcctctcggg aggttgggag cggcgaacta ctttcagcgg cgcacggaga 540
cggcgtctac gtgaggggtg ataagtacg caacactcgt tgcataaatt tgcctccgcc 600
agcccgagc atttaggggc ggttggcttt gttgggtgag cttgtttgtg tccctgtggg 660
tgacgtggt tggtgattgg caggatcctg gtatccgcta acaggtactg gcccgagcc 720
gtaacgacct tgggggggtg tgagaggggg gaatgggtga ggtcaagggt gaggccttctt 780
ggggttgggt gggccgctga ggggagggcg tgggggaggg gagggcgagg tgacgcggcg 840
ctgggccttt ccgggacagt gggccttggt gacctgaggg gggcgagggc gggtggcgcg 900
cgcggggtga cggaaactaa cggacgccta accgatcggc gattctgtcg agtttacttc 960
gcggggaagg cggaaaagag gtagtttggt tggtttctgg aagcctttac tttggaatct 1020
cagtgtgaga aaggtgcccc ttcttggtgt tcaatgggat ttttatttcg cgagtcttgt 1080
gggtttggtt ttgttttcag tttgcctaac accgtgctta ggtttgaggc agattggagt 1140
tcggtcgggg gagtttgaat atccggaaca gttagtgggg aaagctgtgg acgattggta 1200
agagagcgct ctggattttc cgctgttgac gttgaaacct tgaatgacga atttcgtatt 1260
aagtgactta gccttgtaaa attgagggga ggcttgcgga atattaacgt atttaaggca 1320
ttttgaagga atagttgcta attttgaaga atattaggtg taaaagcaag aaatacaatg 1380
atcctgaggt gacacgctta tgttttactt ttaactaggt tcagcatg 1428
<210> 56
<211> 1055
<212> ADN
<213> fragmento

ES 2 780 692 T3

<400> 56
ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc ccaacgaccc 60
ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca 120
ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagtac atcaagtgt 180
tcatatgcc agtacgcccc ctattgacgt caatgacggt aaatggcccc cctggcatta 240
tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg tattagtcac 300
cgctattacc atggctcgagg tgagccccac gttctgcttc actctcccca tctccccccc 360
ctccccaccc ccaattttgt atttatttat tttttaatta ttttgtgcag cgatgggggc 420
gggggggggg gggggcgcg cccaggcggg gcggggcggg gcgaggggcg gggcggggcg 480
aggcggagag gtgcggcggc agccaatcag agcggcgcg tccgaaagt tccctttatg 540
gcgagggcgg ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc gggagtcgct 600
gcgcgctgcc ttgcgcccgt gcccgcctcc gccgcccgcct cgcgcgccc gcccgcgctc 660
tgactgaccg cgttactaaa acaggtaatg ccggcctccg cgcggggttt tggcgccctcc 720
cgcgggcgcc cccctcctca cggcgagcgc tgccacgtca gacgaagggc gcagcgagcg 780
tcctgatcct tccgcccga cgctcaggac agcggcccg tgctcataag actcggcctt 840
agaaccccag tatcagcaga aggacatttt aggacgggac ttgggtgact ctagggcact 900
ggttttcttt ccagagagcg gaacaggcga ggaaaagtag tcccttctcg gcgattctgc 960
ggaggggatct ccgtggggcg gtgaacgccg atgatgcctc tactaaccat gttcatgttt 1020
tctttttttt tctacaggtc ctgggtgacg aacag 1055

<210> 57

<211> 753

<212> ADN

<213> Artificial

<220>

<223> CMV promoter

<400> 57

tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccg ccatgttggc attgattatt gactagttat taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgccaacg acccccgcgc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcg cccctattg acgtcaatga 420
cggtaaattg cccgcctggc attatgccca gtacatgacc ttacgggact ttcctacttg 480

ES 2 780 692 T3

	gcagtacatc tacgtattag tcatcgctat taccatagtg atgcggtttt ggcagtacac	540
	caatgggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt	600
	caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaataaccc	660
	cgccccgttg acgcaaattgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc	720
	tcgttttagtg aaccgtcaga tcactagaag ctt	753
5	<210> 58 <211> 69 <212> ADN <213> Artificial <220> <223> promotor mínimo de CMV <400> 58	
	taggcgtgta cgggtgggagg tctatataag cagagctcgt ttagtgaacc gtcagatcac	60
10	tagaagctt	69
	<210> 59 <211> 2248 <212> ADN <213> Artificial	
15	<220> <223> EEE1-Xt <400> 59	
	tgggtgaccct gtctcaaaaa accctcaaaa agtggtggga ttagtggcat gcaccaccat	60
	tcccacaaaa ggtttatattt taataatatg tgtgtgagtg tgtatcacta tgagtatatg	120
	tcaatatgtg tcaatgtccc caggagacatt taaagagccc ctgaagctgg agtcataggc	180
	cattatgaac tgcctgacat ggctaattggg aattgaactc agattttctg gaagttatac	240
	ctgctcttac tgctgagcca tgtctctgaa gaccccaggg atttttttttt ttttttgaga	300
	caggatatttt ctgtatagcc ctggctgtcc tgaaagcact ctctatatgt agaccaggct	360
	tgcctggagc ttggatatgc acctgcttct gcctcaggaa tgggtgggatt gaaggtgtgc	420
	accaccacat ccgctaacat gcacaattct taatgggttt atatcttatt taatgaatga	480
	aagggtttggg ggatggatgt agcttaatgg aaaatgactg aagatttcaa ttaaaaaatct	540
	ggggcttagc tgcgcggtgg gtggtgcctg ctttagtcc cagtactggg gaggcagagg	600
	aaggaggatc tctgtgagtt cgaggccagc tgggtctataa cgtgagttcc aggacagcca	660
	gagatacaca gacaaaccct gtctcaccaa aacaaaacaa caacaacaac aacaaatctg	720
	ggacgtaggc ttggtgtggt ggcacacatt ttgattccag cacttgggaag gaagaggcct	780
	gcatggtcta catagcttgt ttcaggcaac cagagctaca tagtgagatc ctgtctcaac	840

ES 2 780 692 T3

	aaaaataaaa taatctaagg cttcaaaggg ttcaatctct taggtagcta aatatgaaca	900
	aaatttgga aatgtgacct ttcccttagt gacagtcaga tagaaccttc tcgagtgcaa	960
	ggacaccaag tgcaaacagg ctcaagaaca gcctggaaag gtctagtgt atggggcttc	1020
	aggtcgaatg ccaactgttt tcaagaactg tgtggatttt tctgcctgta acgaattcag	1080
	attcattttt caaaactcgg ggagagtttt cccctttat aatttttttt ttaaatttat	1140
	taaactttgt ttcggtcccc ttgttttgag aattgcagag tcatccacc tgtcacagt	1200
	ccagggagct cagggatggg cccagggggc tggcggggct gaaggggctg gggaagcgag	1260
	ggctccaaag ggaccccagt gtggcaggag ccaaagccct aggtccctag aacgcagagg	1320
	ccaccgggac ccccagacg gggtaagcgg gtgggtgtct ggggcgcgaa gccgcactgc	1380
	gcatgcgcg aggtccgctc cggccgcgct gatccaagcc gggttctcgc gccgacctg	1440
	tcgtgattga caagtacac acgctgatcc ctccgcgggg ccgcacaggg tcacagcctt	1500
	tcccctcccc acaaagcccc ctactctctg ggcaccacac acgaacattc cttgagcgtg	1560
	accttggttg ctctagtcag ggcctccgg tgcagagact ggaacggcct tgggaagtag	1620
	tccctaaccg catttccgcg gagggatcgt cgggagggcg tggcttctga ggattatata	1680
	aggcgactcc gggcgggtct tagctagttc cgtcggagac ccgagttcag tcgccgcttc	1740
	tctgtgagga ctgctgccgc cgcgctgggt gaggagaagc cgcgcgctt ggcgtagctg	1800
	agagacgggg agggggcgcg gacacgaggg gcagcccgcg gcctggacgt tctgtttccg	1860
	tggcccgcga ggaaggcgac tgtcctgagg cggaggacc agcggcaaga tggcggccaa	1920
	gtggaagcct gaggggatag gcgagcggcc ctgaggcgct cgacgggggt gggggggaag	1980
	caggcccgcg aggcagctgc agccgggaac gtgcggccaa ccccttattt tttttgacgg	2040
	gttgccgggcc gtaggtgcct ccgaagtgag agccgtgggc gtttgactgt cgggagaggt	2100
	cggtcggatt ttcatccgtt gctaaagacg gaagtgcgac tgagacggga agggggggga	2160
	gtcggttggt ggcggttgaa cctggactaa ggcgcacatg acgtcgcggt ttctatgggc	2220
	tcataatggg tggtaggac atttcctt	2248
5	<210> 60	
	<211> 1159	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> EEE1-80	
	<400> 60	
	tcaaaactcg gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg	60
	tttcggtccc cttgttttga gaattgcaga gtcacccacc ctgtcacagt gccagggagc	120
10	tcagggatgg gccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa	180

ES 2 780 692 T3

gggaccccag tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga 240
 ccccccagac ggggtaagcg ggtgggtgtc tggggcgcgga agccgcactg cgcattgcgcc 300
 gaggtccgct ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg 360
 acaagtcaca cacgctgac cctccgcggg gccgcacagg gtcacagcct tccccctccc 420
 cacaaagccc cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg 480
 gctctagtca ggcgcctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc 540
 gcatttccgc ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc 600
 cgggcgggtc ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg 660
 actgctgccg ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg 720
 gagggggcgc ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc gtggcccgcg 780
 aggaaggcga ctgtcctgag gcggaggacc cagcggcaag atggcggcca agtgggaagcc 840
 tgaggggata ggcgagcggc cctgaggcgc tcgacggggt tgggggggaa gcaggcccgc 900
 gaggcagctg cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgccggc 960
 cgtaggtgcc tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat 1020
 tttcatccgt tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttgg 1080
 tggcggttga acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg 1140
 gtggtgagga catttccct 1159
 <210> 61
 <211> 869
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1-60
 <400> 61
 cgcattgcgcc gaggtccgct ccggccgcgc tgatccaagc cgggttctcg cgccgacctg 60
 gtcgtgattg acaagtcaca cacgctgac cctccgcggg gccgcacagg gtcacagcct 120
 tccccctccc cacaaagccc cctactctct gggcaccaca cacgaacatt ccttgagcgt 180
 gaccttggtg gctctagtca ggcgcctccg gtgcagagac tggaacggcc ttgggaagta 240
 gtccctaacc gcatttccgc ggagggatcg tcgggagggc gtggcttctg aggattatat 300
 aaggcgactc cgggcgggtc ttagctagtt ccgtcggaga cccgagttca gtcgccgctt 360
 ctctgtgagg actgctgccg ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct 420
 gagagacggg gagggggcgc ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc 480
 gtggcccgcg aggaaggcga ctgtcctgag gcggaggacc cagcggcaag atggcggcca 540

5

10

ES 2 780 692 T3

	agtggaagcc tgaggggata ggcgagcggc cctgaggcgc tcgacggggt tgggggggaa	600
	gcaggcccgc gaggcagctg cagccgggaa cgtgcggcca accccttatt ttttttgacg	660
	ggttgcgggc cgtaggtgcc tccgaagtga gagccgtggg cgtttgactg tcgggagagg	720
	tcggtcggat tttcatccgt tgctaaagac ggaagtgcga ctgagacggg aagggggggg	780
	agtcggttgg tggcggttga acctggacta aggcgcacat gacgtcgcgg tttctatggg	840
	ctcataatgg gtggtgagga catttcctt	869
5	<210> 62 <211> 724 <212> ADN <213> Artificial <220> <223> EEE1-50 <400> 62	
	tctctgggca ccacacacga acattccttg agcgtgacct tgttggctct agtcaggcgc	60
	ctccggtgca gagactggaa cggccttggg aagtagtccc taaccgcatt tccgcggagg	120
	gatcgtcggg agggcgtggc ttctgaggat tatataaggc gactccgggc gggctcttagc	180
	tagttccgtc ggagaccga gttcagtcgc cgcttctctg tgaggactgc tgccgccgcc	240
	gctggtgagg agaagccgcc gcgcttggcg tagctgagag acggggaggg ggcgcggaca	300
	cgaggggcag cccgcggcct ggacgttctg tttccgtggc ccgcgaggaa ggcgactgtc	360
	ctgaggcggg ggaccagcg gcaagatggc ggccaagtgg aagcctgagg ggataggcga	420
	gcggccctga ggcgctcgac ggggttgggg gggaagcagg cccgcgaggc agctgcagcc	480
	gggaacgtgc ggccaacccc ttatTTTTTT tgacgggttg cgggccgtag gtgcctccga	540
	agtgagagcc gtgggcgttt gactgtcggg agaggtcggg cggattttca tccgttgcta	600
	aagacggaag tgcgactgag acgggaaggg gggggagtcg gttggtggcg gttgaacctg	660
	gactaaggcg cacatgacgt cgcggtttct atgggctcat aatgggtggt gaggacattt	720
10	ccct	724
	<210> 63 <211> 1449 <212> ADN <213> Artificial <220> <223> EEE1-SL <400> 63	
15	tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag	60
	gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc	120
	ttttccttag tgacagtcag atagaacctt ctcgagtgcaggacaccaa gtgcaaacag	180

ES 2 780 692 T3

gctcaagaac agcctggaaa ggtctagtgc tatggggcctt caggtcgaat gccaaactggt 240
 ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
 cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
 gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480
 tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540
 ggggaaagcg gttgggtgtc tggggcgcgga agccgcactg cgcatgcgcc gaggtccgct 600
 ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660
 cacgctgatc cctccgcggg gccgcacagg gtcacagcct tccccctccc cacaaagccc 720
 cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca 780
 ggcgcctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttccgc 840
 ggaggggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcgggctc 900
 ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960
 ccgccgctgc tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcgc 1020
 ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc gtggcccgcg aggaaggcga 1080
 ctgtcctgag gcggaggacc cagcggcaag atggcgcca agtggaagcc tgaggggata 1140
 ggcgagcggc cctgaggcgc tcgacggggt tgggggggaa gcaggcccgc gaggcagctg 1200
 cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc cgtaggtgcc 1260
 tccgaattga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt 1320
 tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttg tggcggttga 1380
 acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgagga 1440
 catttccct 1449

5

<210> 64
 <211> 33
 <212> ADN
 <213> Artificial
 <220>
 <223> HC RACE PRIMER
 <400> 64

10

gctggtgccc aggtccttag cgcaatagta cac 33
 <210> 65

15

<211> 4128
 <212> ADN
 <213> Artificial
 <220>
 <223> Secuencia de vector de cadena ligera sin la secuencia codificante

ES 2 780 692 T3

<400> 65

tgccaggcggc cgcttttcagg caaccagagc tacatagtga gatcctgtct caacaaaaaat	60
aaaataatct aaggcttcaa aggggttcaat ctcttaggta gctaaatatg aacaaaattt	120
gggaaatgtg accttttcct tagtgacagt cagatagaac cttctcgagt gcaaggacac	180
caagtgcaaa caggctcaag aacagcctgg aaaggtctag tgctatgggg cttcagggtcg	240
aatgccaaact gttttcaaga actgtgtgga tttttctgcc tgtaacgaat tcagattcat	300
ttttcaaaac tcgggggagag ttttccccct ttataatttt ttttttaaata ttattaaact	360
ttgtttcggtt ccccttggtt tgagaattgc agagtcaccc accctgtcac agtgccagggg	420
agctcaggga tggggccagg ggcctggcgg ggctgaagg gctggggaag cgagggtcc	480
aaagggaccc cagtgtggca ggagccaaag ccctagggtcc ctagaacgca gaggccaccg	540
ggacccccca gacggggtaa gcggtgtggg gtctggggcg cgaagccgca ctgcgcacgc	600
gccgagggtcc gctccggccg cgctgatcca agccgggttc tcgcgccgac ctggtcgtga	660
ttgacaagtc acacacgctg atccctccgc ggggccgcac agggtcacag ctttccccct	720
ccccacaaag cccctactc tctgggcacc acacacgaac attccttgag cgtgaccttg	780
ttggctctag tcaggcgctt ccggtgcaga gactggaacg gccttgggaa gtagtcccta	840
accgcatttc cgcggaggga tcgtcgggag ggcgtggctt ctgaggatta tataaggcga	900
ctccggggcgg gtcttagcta gttccgtcgg agaccgagt tcagtcgccg cttctctgtg	960
aggactgctg ccgccgccgc tgggtaggag aagccgccgc gcttggcgta gctgagagac	1020
ggggaggggg cgccgacacg aggggcagcc cgccggcctgg acgttctgtt tccgtggccc	1080
gcgaggaagg cgaactgtct gaggcggagg acccagcggc aagatggcgg ccaagtggaa	1140
gcctgagggg ataggcgagc ggccctgagg cgctcgacgg ggttgggggg gaagcaggcc	1200
cgcgaggcag ctgcagccgg gaacgtgcgg ccaaccctt attttttttg acgggttgcg	1260
ggccgtaggt gcctccgaag tgagagccgt gggcgtttga ctgtcgggag aggtcggtcg	1320
gattttcatc cgttgctaaa gacggaagtg cgactgagac gggaaggggg gggagtcggt	1380
tgggtggcgg tgaacctgga ctaaggcgca catgacgtcg cggtttctat gggctcataa	1440
tgggtgggtga ggacatttcc ctgttttaac ttaaacaagt ttgtacaaaa aagcaggcta	1500
gatcttcaat attggccatt agccatatta ttcatgtgtt atatagcata aatcaatatt	1560
ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca	1620
tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt	1680
acgggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat	1740
ggcccgccctg gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatggt	1800

cccatagtaa	cgccaatagg	gactttccat	tgacgtcaat	gggtggagta	tttacggtaa	1860	
actgcccact	tggcagtaca	tcaagtgtat	catatgccaa	gtccgcccc	tattgacgtc	1920	
aatgacggta	aatggcccg	ctggcattat	gcccagtaca	tgaccttacg	ggactttcct	1980	
acttggcagt	acatctacgt	attagtcac	gctattacca	tagtgatg	gttttggcag	2040	
tacaccaatg	ggcgtggata	gcggtttgac	tcacggggat	ttccaagtct	ccacccatt	2100	
gacgtcaatg	ggagtttgtt	ttggcaccaa	aatcaacggg	actttccaaa	atgtcgtaat	2160	
aaccccgccc	cgttgacgca	aatgggcgg	aggcgtgtac	ggtgggaggt	ctatataagc	2220	
agagctcg	tagtgaaccg	tcagatcact	agaagcttaa	tacgactcac	tatagggaga	2280	
cccaagctgg	ctagcg	tttaaacggg	ccgccagtgt	gctggaattc		2340	
ggcttaactc	tagaccatgg	ggcgcgcgg	ttcagcctcg	actgtgcctt	ctagttgcc	2400	
gccatctgtt	gtttgcccct	cccccg	ttccttgacc	ctggaagg	ccactccac	2460	
tgtcctttcc	taataaaatg	aggaaattgc	atcgcatgt	ctgagtaggt	gtcattctat	2520	
tctgggggg	gggggtggg	aggacagcaa	gggggaggat	tgggaagaca	atagcaggca	2580	
tgctggggat	gcggtgggct	ctatggcttc	tgaggcgga	agaaccagct	ggatccatcc	2640	
gttagatatc	tgtggaatgt	gtgtcagtta	gggtgtggaa	agtccccagg	ctccccagca	2700	
ggcagaagta	tgcaaagcat	gcac	ctcaat	tagtcagcaa	ccagggtgtg	aaagtcccca	2760
ggctccccag	caggcagaag	tatgcaaagc	atgcac	ctca	attagtcagc	aaccatagtc	2820
ccgcccctaa	ctccgccc	at	cccccta	actccgcca	gttccgcca	ttctccggcc	2880
catgcctgac	taattttttt	tatttatgca	gaggccgagg	ccgcctctgc	ctctgagcta		2940
ttccagaagt	agtgaggagg	cttttttgg	ggcctaggct	tttgcaaaaa	gctcccgga		3000
gcttgatat	ccattttcgg	atctgatcaa	gagacaggat	gaggatcg	ttcacatgatt		3060
gaacaagatg	gattgcacgc	aggttctccg	gccgcttggg	tggagaggct	attcggtat		3120
gactgggcac	aacagacaat	cggctgctct	gatgccg	ccg	tgttccggct	gtcagcgcag	3180
gggcgcccgg	ttctttttgt	caagaccgac	ctgtccgg	tg	ccctgaatga	actgcaggac	3240
gaggcagcgc	ggctatcgtg	gctggccacg	acgggcgttc	cttg	gcgacg	tgtgctcgac	3300
gttgtcactg	aagcgggaag	ggactggctg	ctattggg	cg	aagtgccggg	gcaggatctc	3360
ctgtcatctc	accttgctcc	tgccgagaaa	gtatccatca	tggctgatgc	aatgcggcgg		3420
ctgcatacgc	ttgatccggc	tacctgcca	ttcgaccacc	aagcgaaaca	tcgcatcgag		3480
cgagcacgta	ctcgatgga	agccggtctt	gtcgatcagg	atgatctgga	cgaagagcat		3540
caggggctcg	cgccagccga	actgttcgcc	aggctcaagg	cgcgcatgcc	cgacggcgag		3600
gatctcgctg	tgacacatgg	cgatgcctgc	ttgccgaata	tcatggtgga	aaatggccgc		3660

5

```

ttttctggat tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg 3720
ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg 3780
ctttacggta tcgccgctcc cgattcgcag cgcacgcct tctatgcct tcttgacgag 3840
ttcttctagg taccacgaga tttcgattcc accgccgcct tctatgaaag gttgggcttc 3900
ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag 3960
ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata aagcaatagc 4020
atcacaaatt tcacaaataa agcatttttt tcaactgcatt ctagttgtgg tttgtccaaa 4080
ctcatcaatg tatcttatca tgtctcaggt tgatgagcat attttacc 4128

```

<210> 66

<211> 2647

<212> ADN

<213> Artificial

<220>

<223> Secuencia de vector de cadena pesada sin la secuencia codificante

<400> 66

```

tgcaggcggc cgctttcagg caaccagagc tacatagtga gatcctgtct caacaaaaat 60
aaaataatct aaggcttcaa agggttcaat ctcttaggta gctaaatatg aacaaaattt 120
gggaaatgtg accttttctt tagtgacagt cagatagaac cttctcgagt gcaaggacac 180
caagtgcaaa caggctcaag aacagcctgg aaaggcttag tgctatgggg cttcagggtcg 240
aatgccaaact gttttcaaga actgtgtgga tttttctgcc tgtaacgaat tcagattcat 300
ttttcaaaac tcggggagag ttttccccct ttataatttt ttttttaaata ttattaaact 360
ttgtttcgtt ccccttgttt tgagaattgc agagtcaccc accctgtcac agtgccaggg 420
agctcaggga tggggccagg ggcctggcgg ggctgaagg gctggggaag cgagggtcc 480
aaagggaccc cagtgtggca ggagccaaag ccctaggtcc ctagaacgca gaggccaccg 540
ggacccccca gacggggtaa gcgggtgggt gtctggggcg cgaagccgca ctgcgcatgc 600
gccgaggtcc gctccggccg cgctgatcca agccgggttc tcgcgccgac ctggtcgtga 660
ttgacaagtc acacacgctg atccctccgc ggggccgcac agggtcacag cctttcccct 720
ccccacaaag cccctactc tctgggcacc acacacgaac attccttgag cgtgacctg 780
ttggctctag tcaggcgctt ccggtgcaga gactggaacg gccttgggaa gtagtcccta 840
accgcatttc cgcggaggga tcgtcgggag ggcgtggctt ctgaggatta tataaggcga 900
ctccggggcg gtcttagcta gttccgtcgg agaccgagt tcagtcgccg cttctctgtg 960
aggactgctg ccgccgccgc tgggtaggag aagccgccgc gcttggcgta gctgagagac 1020
ggggaggggg cgcggacacg aggggcagcc cgcggcctgg acgttctgtt tccgtggccc 1080
gcgaggaagg cgactgtcct gaggcggagg acccagcggc aagatggcgg ccaagtggaa 1140

```

10

gcctgagggg atagggcgagc ggccctgagg cgctcgacgg ggttgggggg gaagcaggcc	1200
cgcgaggcag ctgcagccgg gaacgtgcgg ccaaccctt attttttttg acgggttgcg	1260
ggccgtaggt gcctccgaag tgagagccgt gggcgtttga ctgtcgggag aggtcggtcg	1320
gattttcatc cgttgctaaa gacggaagtg cgactgagac ggggaagggg gggagtcggt	1380
tggtggcggt tgaacctgga ctaaggcgca catgacgtcg cggtttctat gggctcataa	1440
tgggtggtga ggacatttcc ctgtttaaac ttaaacaagt ttgtacaaaa aagcaggcta	1500
gatcttcaat attggccatt agccatatta ttcattggtt atatagcata aatcaatatt	1560
ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca	1620
tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt	1680
acggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaa	1740
ggcccgccctg gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatggt	1800
cccatagtaa cgccaatagg gactttccat tgacgtcaat gggtgaggta ttacggtaa	1860
actgcccact tggcagtaca tcaagtgtat catatgccaa gtccgcccc tattgacgtc	1920
aatgacggta aatggccgc ctggcattat gcccagtaca tgacctacg ggactttcct	1980
acttggcagt acatctacgt attagtcac cgtattacca tagtgatgcg gttttggcag	2040
tacaccaatg ggcgtggata gcggtttgac tcacggggat ttccaagtct ccacccatt	2100
gacgtcaatg ggagtttggt ttggcaccaa aatcaacggg actttccaaa atgtcgtaat	2160
aaccccgccc cgttgacgca aatgggcggg aggcgtgtac ggtgggaggt ctatataagc	2220
agagctcggt tagtgaaccg tcagatcact agaagcttaa tacgactcac tatagggaga	2280
cccaagctgg ctagcgttta aacgggccct ctagtaacgg ccgccagtgt gctggaattc	2340
ggcttaactc tagaccatgg ggcgcgcggg ttcagcctcg actgtgcctt ctagttgcca	2400
gccatctggt gtttgcccct ccccggtgcc ttccttgacc ctggaagggt ccaactccac	2460
tgtcctttcc taataaaatg aggaaattgc atcgcatgt ctgagtaggt gtcattctat	2520
tctggggggg ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca	2580
tgctggggat gcggtgggct ctatggcttc tgaggcgga agaaccagct ggatccatcc	2640
gttagat	2647
<210> 67	
<211> 235	
<212> PRT	
<213> Artificial	
<220>	
<223> Cadena ligera de la proteína HuMab1	
<400> 67	

ES 2 780 692 T3

Met	Gly	Trp	Ser	Cys	Ile	Ile	Leu	Phe	Leu	Val	Ala	Thr	Ala	Thr	Gly	1	5	10	15
Val	His	Ser	Ala	Gln	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	Val	20	25	30	
Ser	Ala	Ser	Val	Gly	Asp	Arg	Val	Thr	Ile	Thr	Cys	Arg	Ala	Ser	Gln	35	40	45	
Gly	Ile	Ser	Ser	Trp	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Lys	Ala	50	55	60	
Pro	Lys	Leu	Leu	Ile	Tyr	Ala	Ala	Ser	Ser	Leu	Gln	Ser	Gly	Val	Pro	65	70	75	80
Ser	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile	85	90	95	
Ser	Ser	Leu	Gln	Pro	Glu	Asp	Phe	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Ala	100	105	110	
Asn	Asn	Phe	Pro	Leu	Thr	Phe	Gly	Gly	Gly	Thr	Lys	Val	Glu	Ile	Lys	115	120	125	
Arg	Thr	Val	Ala	Ala	Pro	Ser	Val	Phe	Ile	Phe	Pro	Pro	Ser	Asp	Glu	130	135	140	
Gln	Leu	Lys	Ser	Gly	Thr	Ala	Ser	Val	Val	Cys	Leu	Leu	Asn	Asn	Phe	145	150	155	160
Tyr	Pro	Arg	Glu	Ala	Lys	Val	Gln	Trp	Lys	Val	Asp	Asn	Ala	Leu	Gln	165	170	175	
Ser	Gly	Asn	Ser	Gln	Glu	Ser	Val	Thr	Glu	Gln	Asp	Ser	Lys	Asp	Ser	180	185	190	
Thr	Tyr	Ser	Leu	Ser	Ser	Thr	Leu	Thr	Leu	Ser	Lys	Ala	Asp	Tyr	Glu	195	200	205	
Lys	His	Lys	Val	Tyr	Ala	Cys	Glu	Val	Thr	His	Gln	Gly	Leu	Ser	Ser	210	215	220	
Pro	Val	Thr	Lys	Ser	Phe	Asn	Arg	Gly	Glu	Cys	225	230	235						

<210> 68
<211> 474

5

<212> PRT
<213> Artificial

10

<220>
<223> Cadena pesada de la proteína HuMab1
<400> 68

ES 2 780 692 T3

Met	Gly	Trp	Ser	Cys	Ile	Ile	Leu	Phe	Leu	Val	Ala	Thr	Ala	Thr	Gly	1	5	10	15
Val	His	Ser	Glu	Val	Gln	Leu	Leu	Glu	Ser	Gly	Gly	Gly	Leu	Val	Gln	20	25	30	
Pro	Gly	Gly	Ser	Leu	Arg	Leu	Ser	Cys	Ala	Ala	Ser	Gly	Phe	Thr	Phe	35	40	45	
Ser	Asn	Tyr	Ala	Met	Ser	Trp	Val	Arg	Gln	Ala	Pro	Gly	Lys	Gly	Leu	50	55	60	
Glu	Trp	Val	Ser	Ala	Ile	Ser	Ala	Ser	Gly	His	Ser	Thr	Tyr	Leu	Ala	65	70	75	80
Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ser	Lys	Asn	85	90	95	
Thr	Leu	Tyr	Leu	Gln	Met	Asn	Ser	Leu	Arg	Ala	Glu	Asp	Thr	Ala	Val	100	105	110	
Tyr	Tyr	Cys	Ala	Lys	Asp	Arg	Glu	Val	Thr	Met	Ile	Val	Val	Leu	Asn	115	120	125	
Gly	Gly	Phe	Asp	Tyr	Trp	Gly	Gln	Gly	Thr	Arg	Val	Thr	Val	Ser	Ser	130	135	140	
Ala	Ser	Thr	Lys	Gly	Pro	Ser	Val	Phe	Pro	Leu	Ala	Pro	Ser	Ser	Lys	145	150	155	160
Ser	Thr	Ser	Gly	Gly	Thr	Ala	Ala	Leu	Gly	Cys	Leu	Val	Lys	Asp	Tyr	165	170	175	
Phe	Pro	Glu	Pro	Val	Thr	Val	Ser	Trp	Asn	Ser	Gly	Ala	Leu	Thr	Ser	180	185	190	
Gly	Val	His	Thr	Phe	Pro	Ala	Val	Leu	Gln	Ser	Ser	Gly	Leu	Tyr	Ser	195	200	205	
Leu	Ser	Ser	Val	Val	Thr	Val	Pro	Ser	Ser	Ser	Leu	Gly	Thr	Gln	Thr	210	215	220	

ES 2 780 692 T3

Tyr Ile Cys Asn Val Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys
 225 230 235 240
 Arg Val Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys
 245 250 255
 Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro
 260 265 270
 Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys
 275 280 285
 Val Val Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp
 290 295 300
 Tyr Val Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu
 305 310 315 320
 Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu
 325 330 335
 His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn
 340 345 350
 Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly
 355 360 365
 Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu
 370 375 380
 Met Thr Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr
 385 390 395 400
 Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn
 405 410 415
 Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe
 420 425 430
 Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn
 435 440 445
 Val Phe Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr
 450 455 460
 Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
 465 470

5

<210> 69

<211> 233

<212> PRT

<213> Artificial

<220>

<223> Cadena ligera de la proteína HuMab2

<400> 69

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly
 1 5 10 15

Val His Ser Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala
 20 25 30

Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val
 35 40 45

Asn Thr Ala Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys
 50 55 60

Leu Leu Ile Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg
 65 70 75 80

Phe Ser Gly Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser
 85 90 95

Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr
 100 105 110

Thr Pro Pro Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Thr
 115 120 125

Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu
 130 135 140

Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro
 145 150 155 160

Arg Glu Ala Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly
 165 170 175

Asn Ser Gln Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr
 180 185 190

Ser Leu Ser Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His
 195 200 205

Lys Val Tyr Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val
 210 215 220

Thr Lys Ser Phe Asn Arg Gly Glu Cys
 225 230

<210> 70

<211> 470

<212> PRT

<213> Artificial

<220>

<223> Cadena pesada de la proteína HuMab2

<400> 70

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly
 1 5 10 15

Val His Ser Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln
 20 25 30

Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn Ile
 35 40 45

Lys Asp Thr Tyr Ile His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu
 50 55 60

Glu Trp Val Ala Arg Ile Tyr Pro Thr Asn Gly Tyr Thr Arg Tyr Ala
 65 70 75 80

Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Ala Asp Thr Ser Lys Asn
 85 90 95

Thr Ala Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val
 100 105 110

Tyr Tyr Cys Ser Arg Trp Gly Gly Asp Gly Phe Tyr Ala Met Asp Tyr
 115 120 125

Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly
 130 135 140

Pro Ser Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly Gly
 145 150 155 160

Thr Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val
 165 170 175

ES 2 780 692 T3

Thr	Val	Ser	Trp	Asn	Ser	Gly	Ala	Leu	Thr	Ser	Gly	Val	His	Thr	Phe	180	185	190	
Pro	Ala	Val	Leu	Gln	Ser	Ser	Gly	Leu	Tyr	Ser	Leu	Ser	Ser	Val	Val	195	200	205	
Thr	Val	Pro	Ser	Ser	Ser	Leu	Gly	Thr	Gln	Thr	Tyr	Ile	Cys	Asn	Val	210	215	220	
Asn	His	Lys	Pro	Ser	Asn	Thr	Lys	Val	Asp	Lys	Lys	Val	Glu	Pro	Pro	225	230	235	240
Lys	Ser	Cys	Asp	Lys	Thr	His	Thr	Cys	Pro	Pro	Cys	Pro	Ala	Pro	Glu	245	250	255	
Leu	Leu	Gly	Gly	Pro	Ser	Val	Phe	Leu	Phe	Pro	Pro	Lys	Pro	Lys	Asp	260	265	270	
Thr	Leu	Met	Ile	Ser	Arg	Thr	Pro	Glu	Val	Thr	Cys	Val	Val	Val	Asp	275	280	285	
Val	Ser	His	Glu	Asp	Pro	Glu	Val	Lys	Phe	Asn	Trp	Tyr	Val	Asp	Gly	290	295	300	
Val	Glu	Val	His	Asn	Ala	Lys	Thr	Lys	Pro	Arg	Glu	Glu	Gln	Tyr	Asn	305	310	315	320
Ser	Thr	Tyr	Arg	Val	Val	Ser	Val	Leu	Thr	Val	Leu	His	Gln	Asp	Trp	325	330	335	
Leu	Asn	Gly	Lys	Glu	Tyr	Lys	Cys	Lys	Val	Ser	Asn	Lys	Ala	Leu	Pro	340	345	350	
Ala	Pro	Ile	Glu	Lys	Thr	Ile	Ser	Lys	Ala	Lys	Gly	Gln	Pro	Arg	Glu	355	360	365	
Pro	Gln	Val	Tyr	Thr	Leu	Pro	Pro	Ser	Arg	Asp	Glu	Leu	Thr	Lys	Asn	370	375	380	
Gln	Val	Ser	Leu	Thr	Cys	Leu	Val	Lys	Gly	Phe	Tyr	Pro	Ser	Asp	Ile	385	390	395	400
Ala	Val	Glu	Trp	Glu	Ser	Asn	Gly	Gln	Pro	Glu	Asn	Asn	Tyr	Lys	Thr	405	410	415	
Thr	Pro	Pro	Val	Leu	Asp	Ser	Asp	Gly	Ser	Phe	Phe	Leu	Tyr	Ser	Lys	420	425	430	

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
435 440 445

Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu
450 455 460

Ser Leu Ser Pro Gly Lys
465 470

<210> 71

<211> 5428

<212> ADN

<213> Artificial

<220>

<223> vector de clonación pcADN3.1(+)

<400> 71

gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg	60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg	120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc	180
ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt	240
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata	300
tggagttccg cgttacataa cttacggtaa atggcccgc tggctgaccg cccaacgacc	360
cccgccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc	420
attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt	480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	540
atgcccagta catgacctta tgggactttc ctacttgga gtacatctac gtattagtca	600
tcgtattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg	660
actcacgggg atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc	720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg	780
gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaacca	840
ctgcttactg gcttatcgaa attaatacga ctactatag ggagacccaa gctggctagc	900
gtttaaactt aagcttggtta ccgagctcgg atccactagt ccagtgtggt ggaattctgc	960
agatatccag cacagtggcg gccgctcgag tctagagggc ccgtttaaac ccgctgatca	1020
gcctcgactg tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc	1080
ttgaccctgg aaggtgccac tcccactgtc ctttcctaataaaaatgagga aattgcatcg	1140
cattgtctga gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg	1200
gaggattggg aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag	1260

ES 2 780 692 T3

gcggaaagaa ccagctgggg ctctaggggg tatccccacg cgccctgtag cggcgcatta	1320
agcgcgggcg gtgtgggtgg tactgcgcagc gtgaccgcta cacttgccag cgccctagcg	1380
cccgcctcctt tcgcctttctt cccttccttt ctcgccacgt tcgcccggctt tccccgtcaa	1440
gctctaaatc gggggctccc tttaggggttc cgatttagtg ctttacggca cctcgacccc	1500
aaaaaacttg attaggggtga tgggtcacgt agtgggcat cgccctgata gacgggttttt	1560
cgccctttga cggtggagtc cacgttcttt aatagtggac tcttggtcca aactggaaca	1620
acactcaacc ctatctcggc ctattctttt gatttataag ggattttgcc gatttcggcc	1680
tattggttaa aaaatgagct gatttaacaa aaatttaacg cgaattaatt ctgtggaatg	1740
tgtgtcagtt aggggtgtgga aagtccccag gctccccagc aggcagaagt atgcaaagca	1800
tgcatctcaa ttagtcagca accaggtgtg gaaagtcccc aggcctccca gcaggcagaa	1860
gtatgcaaag catgcatctc aatagttagc caaccatagt cccgccccta actccgccc	1920
tcccgcccct aactccgccc agttccgccc attctccgcc ccatggctga ctaatttttt	1980
ttatttatgc agaggccgag gccgcctctg cctctgagct attccagaag tagtgaggag	2040
gcttttttgg aggcctaggc ttttgcaaaa agctcccggg agcttgata tccattttcg	2100
gatctgatca agagacagga tgaggatcgt ttcgcatgat tgaacaagat ggattgcacg	2160
caggttctcc ggccgcttgg gtggagaggc tattcggcta tgactgggca caacagacaa	2220
tcggctgctc tgatgccgcc gtgttccggc tgtcagcgca gggcgccccg gttctttttg	2280
tcaagaccga cctgtccggc gccctgaatg aactgcagga cgaggcagcg cggctatcgt	2340
ggctggccac gacgggctt ccttgccgag ctgtgctcga cgttgctact gaagcgggaa	2400
gggactggct gctattgggc gaagtgccgg ggcaggatct cctgtcatct caccttgctc	2460
ctgccgagaa agtatccatc atggctgatg caatgcggcg gctgcatacg cttgatccgg	2520
ctacctgcc attcgaccac caagcgaaac atcgcatcga gcgagcacgt actcggatgg	2580
aagccggtct tgatgatcag gatgatctgg acgaagagca tcaggggctc gcgccagccg	2640
aactgttcgc caggctcaag gcgcgcagtc ccgacggcga ggatctcgtc gtgacccatg	2700
gcgatgcctg cttgccgaat atcatgggtg aaaatggccg cttttctgga ttcatacgact	2760
gtggccggct ggggtgtggc gaccgctatc aggacatagc gttggctacc cgtgatattg	2820
ctgaagagct tggcggcgaa tgggctgacc gcttcctcgt gctttacggc atcgccgctc	2880
ccgattcgca gcgcacgcc ttctatcgcc ttcttgacga gttcttctga gcgggactct	2940
ggggttcgaa atgaccgacc aagcgacgcc caacctgcc tcacgagatt tcgattccac	3000
cgccgccttc tatgaaaggc tgggcttcgg aatcgttttc cgggacgccg gctggatgat	3060
cctccagcgc ggggatctca tgctggagtt cttcgccac cccaacttgt ttattgcagc	3120

ttataatggt	tacaaataaa	gcaatagcat	cacaaatttc	acaaataaag	catttttttc	3180
actgcattct	agttgtggtt	tgtccaaact	catcaatgta	tcttatcatg	tctgtatacc	3240
gtcgacctct	agctagagct	tggcgtaatc	atggtcatag	ctgtttcctg	tgtgaaattg	3300
ttatccgctc	acaattccac	acaacatacg	agccggaagc	ataaagtgta	aagcctgggg	3360
tgcctaataga	gtgagctaac	tcacattaat	tgcgttgccg	tcactgcccg	ctttccagtc	3420
gggaaacctg	tcgtgccagc	tgcattaatg	aatcggccaa	cgcgcgggga	gaggcggttt	3480
gcgtattggg	cgctcttccg	cttcctcgct	cactgactcg	ctgcgctcgg	tcgttcggct	3540
gcggcgagcg	gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	3600
taacgcagga	aagaacatgt	gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	3660
cgcgttgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	3720
ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	ttccccctgg	3780
aagctccctc	gtgcgctctc	ctgttccgac	cctgccgctt	accggatacc	tgtccgcctt	3840
tctcccttcg	ggaagcgtgg	cgcttttctca	tagctcacgc	tgtaggtatc	tcagttcggt	3900
gtaggtcggt	cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	3960
cgcttatatc	ggtaactatc	gtcttgagtc	caaccgggta	agacacgact	tatcgccact	4020
ggcagcagcc	actggtaaca	ggattagcag	agcgaggtat	gtaggcggtg	ctacagagtt	4080
cttgaagtgg	tggcctaact	acggctacac	tagaagaaca	gtatttggtg	tctgcgctct	4140
gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	4200
cgctggtagc	ggtttttttg	tttgcaagca	gcagattacg	cgcagaaaaa	aaggatctca	4260
agaagatcct	ttgatctttt	ctacggggtc	tgacgctcag	tggaacgaaa	actcacgtta	4320
agggattttg	gtcatgagat	tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	4380
atgaagtttt	aaatcaatct	aaagtatata	tgagtaaact	tggctcgaca	gttaccaatg	4440
cttaatcagt	gaggcaccta	tctcagcgat	ctgtctattt	cgttcatcca	tagttgcctg	4500
actccccgtc	gtgtagataa	ctacgatacg	ggagggctta	ccatctggcc	ccagtgcctg	4560
aatgataccg	cgagaccac	gctcaccggc	tccagattta	tcagcaataa	accagccagc	4620
cggaagggcc	gagcgcagaa	gtggtcctgc	aactttatcc	gcctccatcc	agtctattaa	4680
ttgttgccgg	gaagctagag	taagtagttc	gccagttaat	agtttgcgca	acgttgttgc	4740
cattgctaca	ggcatcgtag	tgtcacgctc	gtcgtttggt	atggcttcac	tcagctccgg	4800
ttcccaacga	tcaaggcgag	ttacatgatc	ccccatgttg	tgcaaaaaag	cggttagctc	4860
cttcggctct	ccgatcggtg	tcagaagtaa	gttgcccgca	gtgttatcac	tcattggttat	4920
ggcagcactg	cataattctc	ttactgtcat	gccatccgta	agatgctttt	ctgtgactgg	4980
tgagtactca	accaagtcac	tctgagaata	gtgtatgcgg	cgaccgagtt	gctcttgccc	5040

ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 5100
 aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 5160
 gtaaccact cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttcttg 5220
 gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg 5280
 ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 5340
 catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 5400
 atttccccga aaagtgccac ctgacgtc 5428

<210> 72

<211> 519

<212> PRT

<213> Artificial

<220>

<223> proteína SeAP

<400> 72

Met Leu Leu Leu Leu Leu Leu Leu Gly Leu Arg Leu Gln Leu Ser Leu
 1 5 10 15

Gly Ile Ile Pro Val Glu Glu Glu Asn Pro Asp Phe Trp Asn Arg Glu
 20 25 30

Ala Ala Glu Ala Leu Gly Ala Ala Lys Lys Leu Gln Pro Ala Gln Thr
 35 40 45

Ala Ala Lys Asn Leu Ile Ile Phe Leu Gly Asp Gly Met Gly Val Ser
 50 55 60

Thr Val Thr Ala Ala Arg Ile Leu Lys Gly Gln Lys Lys Asp Lys Leu
 65 70 75 80

Gly Pro Glu Ile Pro Leu Ala Met Asp Arg Phe Pro Tyr Val Ala Leu
 85 90 95

Ser Lys Thr Tyr Asn Val Asp Lys His Val Pro Asp Ser Gly Ala Thr
 100 105 110

Ala Thr Ala Tyr Leu Cys Gly Val Lys Gly Asn Phe Gln Thr Ile Gly
 115 120 125

Leu Ser Ala Ala Ala Arg Phe Asn Gln Cys Asn Thr Thr Arg Gly Asn
 130 135 140

Glu Val Ile Ser Val Met Asn Arg Ala Lys Lys Ala Gly Lys Ser Val

5

10

ES 2 780 692 T3

145		150		155		160										
Gly	Val	Val	Thr	Thr	Thr	Thr	Arg	Val	Gln	His	Ala	Ser	Pro	Ala	Gly	Thr
				165						170					175	
Tyr	Ala	His	Thr	Val	Asn	Arg	Asn	Trp	Tyr	Ser	Asp	Ala	Asp	Val	Pro	
			180					185					190			
Ala	Ser	Ala	Arg	Gln	Glu	Gly	Cys	Gln	Asp	Ile	Ala	Thr	Gln	Leu	Ile	
		195					200					205				
Ser	Asn	Met	Asp	Ile	Asp	Val	Ile	Leu	Gly	Gly	Gly	Arg	Lys	Tyr	Met	
	210					215					220					
Phe	Arg	Met	Gly	Thr	Pro	Asp	Pro	Glu	Tyr	Pro	Asp	Asp	Tyr	Ser	Gln	
225					230					235					240	
Gly	Gly	Thr	Arg	Leu	Asp	Gly	Lys	Asn	Leu	Val	Gln	Glu	Trp	Leu	Ala	
				245					250					255		
Lys	Arg	Gln	Gly	Ala	Arg	Tyr	Val	Trp	Asn	Arg	Thr	Glu	Leu	Met	Gln	
			260					265					270			
Ala	Ser	Leu	Asp	Pro	Ser	Val	Thr	His	Leu	Met	Gly	Leu	Phe	Glu	Pro	
		275					280					285				
Gly	Asp	Met	Lys	Tyr	Glu	Ile	His	Arg	Asp	Ser	Thr	Leu	Asp	Pro	Ser	
	290					295					300					
Leu	Met	Glu	Met	Thr	Glu	Ala	Ala	Leu	Arg	Leu	Leu	Ser	Arg	Asn	Pro	
305					310					315					320	
Arg	Gly	Phe	Phe	Leu	Phe	Val	Glu	Gly	Gly	Arg	Ile	Asp	His	Gly	His	
				325					330					335		
His	Glu	Ser	Arg	Ala	Tyr	Arg	Ala	Leu	Thr	Glu	Thr	Ile	Met	Phe	Asp	
			340					345					350			
Asp	Ala	Ile	Glu	Arg	Ala	Gly	Gln	Leu	Thr	Ser	Glu	Glu	Asp	Thr	Leu	
		355					360					365				
Ser	Leu	Val	Thr	Ala	Asp	His	Ser	His	Val	Phe	Ser	Phe	Gly	Gly	Tyr	
	370					375					380					
Pro	Leu	Arg	Gly	Ser	Ser	Ile	Phe	Gly	Leu	Ala	Pro	Gly	Lys	Ala	Arg	
385					390					395					400	

Asp Arg Lys Ala Tyr Thr Val Leu Leu Tyr Gly Asn Gly Pro Gly Tyr
405 410 415

Val Leu Lys Asp Gly Ala Arg Pro Asp Val Thr Glu Ser Glu Ser Gly
420 425 430

Ser Pro Glu Tyr Arg Gln Gln Ser Ala Val Pro Leu Asp Glu Glu Thr
435 440 445

His Ala Gly Glu Asp Val Ala Val Phe Ala Arg Gly Pro Gln Ala His
450 455 460

Leu Val His Gly Val Gln Glu Gln Thr Phe Ile Ala His Val Met Ala
465 470 475 480

Phe Ala Ala Cys Leu Glu Pro Tyr Thr Ala Cys Asp Leu Ala Pro Pro
485 490 495

Ala Gly Thr Thr Asp Ala Ala His Pro Gly Tyr Ser Arg Val Gly Ala
500 505 510

Ala Gly Arg Phe Glu Gln Thr
515

<210> 73

<211> 2522

<212> ADN

<213> Artificial

<220>

<223> EEE1+CMV+TEE

<400> 73

tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60

gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120

ttttccttag tgacagtcag atagaacctt ctcgagtgcaggacaccaa gtgcaaacag 180

gctcaagaac agcctggaaa ggtctagtgc tatggggcct caggtcgaat gccaaactggt 240

ttcaagaact gtgtggattht ttctgcctgt aacgaattca gattcattht tcaaaactcg 300

gggagagtht tcccccttha taathththt tttaaattta ttaaacttht tttcgttccc 360

cttgththtga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420

gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggacccccag 480

tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540

ggggtaagcg ggtgggtgtc tggggcgcca agccgcactg cgcattgcgc gaggtccgct 600

ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660

cacgctgata	cctccgcggg	gccgcacagg	gtcacagcct	ttcccctccc	cacaaagccc	720
cctactctct	gggcaccaca	cacgaacatt	ccttgagcgt	gaccttggtg	gctctagtca	780
ggcgccctcc	gtgcagagac	tggaacggcc	ttgggaagta	gtccctaacc	gcatttccgc	840
ggagggatcg	tcgggagggc	gtggcttctg	aggattatat	aaggcgactc	cgggcggggtc	900
ttagctagtt	ccgtcggaga	cccagagttca	gtcgccgctt	ctctgtgagg	actgctgccg	960
ccgccgctgg	tgaggagaag	ccgccgcgct	tggcgtagct	gagagacggg	gagggggcg	1020
ggacacgagg	ggcagcccg	ggcctggacg	ttctgtttcc	gtggcccgcg	aggaaggcga	1080
ctgtcctgag	gcggaggacc	cagcggcaag	atggcggcca	agtggaagcc	tgaggggata	1140
ggcgagcggc	cctgaggcgc	tcgacggggt	tggggggaa	gcaggcccg	gaggcagctg	1200
cagccgggaa	cgtgcggcca	accccttatt	tttttgacg	ggttgcgggc	cgtaggtgcc	1260
tccgaagtga	gagccgtggg	cgtttgactg	tcgggagagg	tcggtcggt	tttcatccgt	1320
tgctaaagac	ggaagtgcga	ctgagacggg	aagggggggg	agtcggttg	tggcggttga	1380
acctggacta	aggcgacat	gacgtcgcg	tttctatggg	ctcataatgg	gtggtgagga	1440
catttccctg	tttaaactta	aacaagtttg	tacaaaaaag	caggctagat	cttcaatatt	1500
ggccattagc	catattattc	attgggtata	tagcataaat	caatattggc	tattggccat	1560
tgcatacgtt	gtatctatat	cataatatgt	acatttatat	tggctcatgt	ccaatatgac	1620
cgccatgttg	gcattgatta	ttgactagtt	attaatagta	atcaattacg	gggtcattag	1680
ttcatagccc	atatatggag	ttccgcgtta	cataacttac	ggtaaattgg	ccgcctggct	1740
gaccgcccac	cgacccccgc	ccattgacgt	caataatgac	gtatgttccc	atagtaacgc	1800
caatagggac	tttccattga	cgtcaatggg	tgaggtat	acggtaaaact	gcccacttgg	1860
cagtacatca	agtgtatcat	atgccaagtc	cgccccctat	tgacgtcaat	gacggtaaat	1920
ggcccgccctg	gcattatgcc	cagtacatga	ccttacggga	ctttcctact	tggcagtaca	1980
tctacgtatt	agtcacgcgt	attaccatag	tgatgcggtt	ttggcagtac	accaatgggc	2040
gtggatagcg	gtttgactca	cggggatttc	caagtctcca	ccccattgac	gtcaatggga	2100
gtttgttttg	gcacaaaaat	caacgggact	ttccaaaatg	tcgtaataac	cccgccccgt	2160
tgacgcaa	gggcggtagg	cgtgtacgg	gggaggtcta	tataagcaga	gctcgttttag	2220
tgaaccgtca	gatcactaga	agcttcaagc	tctagcagga	agaagaaaga	agaagaagaa	2280
gaagaagaag	aagaagcgtc	tcctcttctt	cttgtgagag	taaaaaagaa	aactcccaaa	2340
aaaaagaaaa	tcataaaaaa	aacaaatttc	aaaaagagtt	tttgtgtttg	gggattaaag	2400
aagaaaaaaa	acaacaggtg	agtaagcgca	gttgcgtct	cttgcggtgc	cgttgctggt	2460
tctcacacct	tttaggtctg	ttctcgtctt	ccgttctgac	tctctctttt	tcgttgacag	2520
cc						2522

5

<210> 74
<211> 3321
<212> ADN
<213> Artificial
<220>

<223> EEE1-Xt+CMV+TEE

<400> 74

tggtgacct	gtctcaaaaa	accctcaaaa	agtgttgga	ttagtgcat	gcaccacct	60
tcccacaaa	ggtttatttt	taataatatg	tgtgtgagt	tgtatcacta	tgagtatatg	120
tcaatatgtg	tcaatgtccc	caggacatt	taaagagccc	ctgaagctgg	agtcataaggc	180
cattatgaac	tgctgacat	ggctaattgg	aattgaactc	agattttctg	gaagttatac	240
ctgctcttac	tgctgagcca	tgtctctgaa	gacccaggg	attttttttt	ttttttgaga	300
caggatattt	ctgtatagcc	ctggctgtcc	tgaagcact	ctctatatgt	agaccaggct	360
tgcttgagc	ttggatatgc	acctgcttct	gcctcaggaa	tggtgggatt	gaaggtgtgc	420
accaccacat	ccgctaacat	gcacaattct	taatgggttt	atatcttatt	taatgaatga	480
aaggtttggg	ggatggatgt	agcttaatgg	aaaatgactg	aagatttcaa	ttaaaaatct	540
ggggcttagc	tgcgcggtgg	gtgggtgcctg	cctttagtcc	cagtactggg	gaggcagagg	600
aaggaggatc	tctgtgagtt	cgaggccagc	tggtctataa	cgtgagttcc	aggacagcca	660
gagatacaca	gacaaacct	gtctcaccaa	aacaaaacaa	caacaacaac	aacaaatctg	720
ggacgtaggc	ttggtgtggt	ggcacacatt	ttgattccag	cacttggaag	gaagaggcct	780
gcatggtcta	catagcttgt	ttcaggcaac	cagagctaca	tagtgagatc	ctgtctcaac	840
aaaaataaaa	taatctaagg	cttcaaagg	ttcaatctct	taggtagcta	aatatgaaca	900
aaatttgga	aatgtgacct	tttccttagt	gacagtcaga	tagaaccttc	tcgagtgcaa	960
ggacaccaag	tgcaaacagg	ctcaagaaca	gcctggaaag	gtctagtgtc	atggggcttc	1020
aggtcgaatg	ccaactgttt	tcaagaactg	tgtggatttt	tctgcctgta	acgaattcag	1080
attcattttt	caaaactcgg	ggagagtgtt	ccccctttat	aatttttttt	ttaaatttat	1140
taaactttgt	ttcgttcccc	ttgttttgag	aattgcagag	tcatccaccc	tgtcacagt	1200
ccaggagct	cagggatggg	cccagggggc	tgccggggct	gaaggggctg	gggaagcgag	1260
ggctccaaag	ggacccagc	gtggcaggag	ccaaagccct	aggtccctag	aacgcagagg	1320
ccaccgggac	ccccagacg	gggtaagcgg	gtgggtgtct	ggggcgcgaa	gccgcactgc	1380
gcatgcgcg	aggtccgctc	cggccgcgct	gatccaagcc	gggttctcgc	gccgacctgg	1440
tcgtgattga	caagtacac	acgctgatcc	ctccgcgggg	ccgcacaggg	tcacagcctt	1500
tccccctccc	acaaagcccc	ctactctctg	ggcaccacac	acgaacattc	cttgagcgtg	1560

ES 2 780 692 T3

```

accttgttgg ctctagtcag ggcctccgg tgcagagact ggaacggcct tgggaagtag 1620
tccctaaccg ctttccgcg gagggatcgt cgggagggcg tggcttctga ggattatata 1680
aggcgactcc gggcggtct tagctagttc cgtcggagac ccgagttcag tcccgcttc 1740
tctgtgagga ctgctgccg cgcgctggg gaggagaagc cgcgcgctt ggcgtagctg 1800
agagacgggg agggggcgcg gacacgaggg gcagcccgcg gcctggacgt tctgtttccg 1860
tggcccgcg ggaaggcgac tgtcctgagg cggaggaccc agcggcaaga tggcgggcaa 1920
gtggaagcct gaggggatag gcgagcggcc ctgaggcgct cgacggggtt gggggggaag 1980
caggcccgcg aggcagctgc agccgggaac gtgcggccaa ccccttattt tttttgacgg 2040
gttgcgggcc gtaggtgcct ccgaagtgag agccgtgggc gtttgactgt cgggagaggt 2100
cggtcggatt ttcattcgtt gctaaagacg gaagtgcgac tgagacggga agggggggga 2160
gtcggttggg ggcggttgaa cctggactaa ggcgcacatg acgtcgcggt ttctatgggc 2220
tcataatggg tggtagggac atttccctgt ttaacttaa acaagtttgt aaaaaaagc 2280
aggctagatc ttcaatattg gccattagcc atattattca ttggttatat agcataaatc 2340
aatattggct attggccatt gcatacgttg tatctatatc ataatatgta catttatatt 2400
ggctcatgtc caatatgacc gccatgttgg cattgattat tgactagtta ttaatagtaa 2460
tcaattacgg ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg 2520
gtaaattggc cgcctggctg accgccaac gaccccgcc cattgacgtc aataatgacg 2580
tatgttccca tagtaacgcc aatagggact ttccattgac gtcaatgggt ggagtattta 2640
cggtaaactg cccacttggc agtacatcaa gtgtatcata tgccaagtcc gccccctatt 2700
gacgtcaatg acggtaaatg gccgcctgg cattatgccc agtacatgac cttacgggac 2760
tttcctactt ggcagtacat ctacgtatta gtcacgcta ttaccatagt gatgcggttt 2820
tggcagtaca ccaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccac 2880
cccattgacg tcaatgggag tttgttttgg caccaaaatc aacgggactt tccaaaatgt 2940
cgtaataacc ccgcccgtt gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat 3000
ataagcagag ctcgtttagt gaaccgtcag atcactagaa gcttcaagct ctagcaggaa 3060
gaagaaagaa gaagaagaag aagaagaaga agaagcgtct cctcttcttc ttgtgagagt 3120
aaaaaagaaa actcccaaaa aaaagaaaat catcaaaaaa acaaatttca aaaagagttt 3180
ttgtgttttg ggattaaaga agaaaaaaaa caacagggtga gtaagcgcag ttgtcgtctc 3240
ttgcggtgcc gttgctggtt ctcacacctt ttaggtctgt tctcgtcttc cgttctgact 3300
ctctcttttt cgttgcaggg c 3321
<210> 75
<211> 2232
<212> ADN
<213> Artificial

```

ES 2 780 692 T3

<220>

<223> EEE1-80+CMV+TEE

<400> 75

tcaaaaactcg gggagagttt tcccccttta taatTTTTTT tttaaattta ttaaactttg	60
tttcggttccc cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc	120
tcagggatgg gccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa	180
gggaccccag tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga	240
ccccccagac ggggtaagcg ggtgggtgtc tggggcgcg agccgcactg cgcattgcgcc	300
gagggtccgct ccggccgcgc tgatccaagc cgggttctcg cgcgcacctg gtcgtgattg	360
acaagtcaca cacgctgatc cctccgcggg gccgcacagg gtcacagcct tttccctccc	420
cacaaagccc cctactctct gggcaccaca cacgaacatt ccttgagcgt gacctgttg	480
gctctagtca ggcgcctccg gtgcagagac tggaaacggc ttgggaagta gtccctaacc	540
gcatttccgc ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc	600
cgggcgggtc ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg	660
actgctgccg ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg	720
gagggggcgc ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc gtggcccgcg	780
aggaaggcga ctgtcctgag gcgaggacc cagcggcaag atggcgcca agtgggaagcc	840
tgaggggata ggcgagcggc cctgaggcgc tcgacggggt tgggggggaa gcaggcccgc	900
gaggcagctg cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc	960
cgtaggtgcc tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat	1020
tttcatccgt tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttgg	1080
tggcggttga acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg	1140
gtggtgagga catttccctg tttaaactta aacaagtttg taaaaaaaag caggctagat	1200
cttcaatatt ggccattagc catattattc attggttata tagcataaat caatattggc	1260
tattggccat tgcatacgtt gtatctatat cataatatgt acatttatat tggctcatgt	1320
ccaatatgac cgccatgttg gcattgatta ttgactagtt attaatagta atcaattacg	1380
gggtcattag ttcatagccc atatatggag ttccgcgtta cataacttac ggtaaatggc	1440
ccgcctggct gaccgcccga cgacccccgc ccattgacgt caataatgac gtatgttccc	1500
atagtaacgc caatagggac tttccattga cgtcaatggg tggagtattt acggtaaact	1560
gcccacttgg cagtacatca agtgtatcat atgccaagtc cggcccctat tgacgtcaat	1620
gacggtaaat ggccgcgctg gcattatgcc cagtacatga ccttacggga ctttcctact	1680

	tggcagtaca tctacgtatt agtcatcgct attaccatag tgatgcgggtt ttggcagtac	1740
	accaatgggc gtggatagcg gtttgactca cggggatttc caagtctcca cccattgac	1800
	gtcaatggga gtttgttttg gcacaaaaat caacgggact ttccaaaatg tcgtaataac	1860
	cccgccccgt tgacgcaaat gggcggtagg cgtgtacggt gggaggtcta tataagcaga	1920
	gctcgttttag tgaaccgtca gatcactaga agcttcaagc tctagcagga agaagaaaga	1980
	agaagaagaa gaagaagaag aagaagcgtc tcctcttctt cttgtgagag taaaaaagaa	2040
	aactcccaaa aaaaagaaaa tcatcaaaaa aacaaatttc aaaaagagtt tttgtgtttg	2100
	gggattaaag aagaaaaaaa acaacaggtg agtaagcgca gttgtcgtct cttgcggtgc	2160
	cgttgctggt tctcacacct tttaggtctg ttctcgtctt ccgttctgac tctctctttt	2220
	tcgttgccagg cc	2232
	<210> 76	
	<211> 1942	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> EEE1-60+CMV+TEE	
	<400> 76	
	cgcacgcgcc gaggtccgct ccggccgcgc tgatccaagc cgggttctcg cgccgacctg	60
	gtcgtgattg acaagtcaca cacgctgac cctccgcggg gccgcacagg gtcacagcct	120
	ttccccctccc caaaaagccc cctactctct gggcaccaca cacgaacatt ccttgagcgt	180
	gaccttggtg gctctagtca ggcgcctccg gtgcagagac tggaacggcc ttgggaagta	240
	gtccctaacc gcatttccgc ggagggatcg tcgggagggc gtggcttctg aggattatat	300
	aaggcgactc cggggcgggtc ttagctagtt ccgtcggaga cccgagttca gtcgccgctt	360
	ctctgtgagg actgctgccg ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct	420
	gagagacggg gagggggcgc ggacacgagg ggcagccgc gccctggacg ttctgtttcc	480
	gtggcccgcg aggaaggcga ctgtcctgag gcggaggacc cagcggcaag atggcggcca	540
	agtggaagcc tgaggggata ggcgagcggc cctgagggcgc tcgacggggt tgggggggaa	600
	gcaggcccgcc gaggcagctg cagccgggaa cgtgcggcca accccttatt ttttttgacg	660
	ggttgcgggc cgtaggtgcc tccgaagtga gagccgtggg cgtttgactg tcgggagagg	720
	tcggtcggat tttcatccgt tgctaaagac ggaagtgcga ctgagacggg aagggggggg	780
	agtcggttgg tggcggttga acctggacta aggcgcacat gacgtcgcgg tttctatggg	840
	ctcataatgg gtggtgagga catttccctg tttaaactta aacaagtttg tacaaaaaag	900
	caggctagat cttcaatatt ggccattagc catattattc attggttata tagcataaat	960
	caatattggc tattggccat tgcatacgtt gtatctatat cataatatgt acatttatat	1020

tggtcatgt ccaatatgac cgccatgttg gcattgatta ttgactagtt attaatagta	1080
atcaattacg gggtcattag ttcataagccc atatatggag ttccgcgtta cataacttac	1140
ggtaaattggc ccgcctggct gaccgcccac cgacccccgc ccattgacgt caataatgac	1200
gtatgttccc atagtaacgc caatagggac tttccattga cgtcaatggg tggagtattt	1260
acggtaaaact gcccaacttg cagtacatca agtgtatcat atgccaagtc cgccccctat	1320
tgacgtcaat gacggtaaat ggccgcctg gcattatgcc cagtacatga ccttacggga	1380
ctttcctact tggcagtaca tctacgtatt agtcatcgct attaccatag tgatgcgggt	1440
ttggcagtac accaatgggc gtggatagcg gtttgactca cggggatttc caagtctcca	1500
ccccattgac gtcaatggga gtttggtttg gcacaaaaat caacgggact ttccaaaatg	1560
tcgtaataac cccgccccgt tgacgcaaat gggcggtagg cgtgtacggg gggagggtcta	1620
tataagcaga gctcgttttag tgaaccgtca gatcactaga agcttcaagc tctagcagga	1680
agaagaaaga agaagaaga gaagaagaag aagaagcgtc tcctcttctt cttgtgagag	1740
taaaaaaga aactcccaaa aaaaagaaaa tcatcaaaaa aacaaatttc aaaaagagtt	1800
tttgtgtttg gggattaaag aagaaaaaaa acaacagggtg agtaagcgca gttgtcgtct	1860
cttgcggtgc cgttgctggt tctcacacct tttaggtctg ttctcgtctt ccgttctgac	1920
tctctctttt tcgttgcagg cc	1942
<210> 77	
<211> 9513	
<212> ADN	
<213> Artificial	
<220>	
<223> pPNic384	
<400> 77	
agatctaaca tccaaagacg aaagggtgaa tgaaaccttt ttgccatccg acatccacag	60
gtccattctc acacataagt gccaaacgca acaggagggg atacactagc agcagaccgt	120
tgcaaacgca ggacctccac tcctcttctc ctcaacaccc acttttgcca tcgaaaaacc	180
agccagttta ttgggcttga ttggagctcg ctcatccaa ttccttctat taggctacta	240
acaccatgac tttattagcc tgtctatcct ggccccctg gcgagggttca tgtttgttta	300
tttccgaatg caacaagctc cgcattacac ccgaacatca ctccagatga gggctttctg	360
agtgtggggg caaatagttt catgttcccc aaatggccca aaactgacag tttaaacgct	420
gtcttggaac ctaatatgac aaaagcgtga tctcatccaa gatgaactaa gtttggttcg	480
ttgaaatgct aacggccagt tgggtcaaaaa gaaacttcca aaagtcgcca taccgtttgt	540
cttggttggt attgattgac gaatgctcaa aaataatctc attaatgctt agcgcagtct	600

5

10

ctctatcgct tctgaacccc ggtgcacctg tgccgaaacg caaatgggga aacacccgct	660
ttttggatga ttatgcattg tctccacatt gtatgcttcc aagattctgg tgggaatact	720
gctgatagcc taacgttcat gatcaaaatt taactgttct aaccctact tgacagcaat	780
atataaacag aaggaagctg ccctgtctta aacctttttt tttatcatca ttattagctt	840
actttcataa ttgcgactgg ttccaattga caagcttttg attttaacga cttttaacga	900
caacttgaga agatcaaaaa acaactaatt attcgaagga tccaaacgat gagatttcct	960
tcaattttta ctgcagtttt attcgcggcc tcctcggcct tagctgctcc agtcaacact	1020
acaacagaag atgaaacggc acaaattccg gctgaagctg tcatcgggta ctcagattta	1080
gaaggggatt tcgatgttgc tgttttgcca tttccaaca gcacaaataa cgggttattg	1140
tttataaata ctactattgc cagcattgct gctaaagaag aaggggtatc tctcgagaaa	1200
agagaggctg aagcttacgt agaattcgag ggtgctgtct tgcctagatc cgctaaagaa	1260
ttgagatgtc agtgtatcaa gacttactcc aagccattcc acccaaagt catcaaagag	1320
ttgagagtta tcgagtcggg tccacactgt gctaactctg agatcatcgt taagttgtcc	1380
gacggtagag agttgtgttt ggacccaaaa gagaactggg ttcagagagt tgttgagaag	1440
ttcttgaaga gagctgagaa ctctagtaa gcggccgcga attaatcgc cttagacatg	1500
actgttcctc agttcaagtt gggcacttac gagaagaccg gtcttgctag attctaata	1560
agaggatgtc agaatgccat ttgcctgaga gatgcaggct tcatttttga tactttttta	1620
tttgtaacct atatagtata ggattttttt tgtcattttg tttcttctcg tacgagcttg	1680
ctcctgatca gcctatctcg cagctgatga atatcttgtg gtaggggttt gggaaaatca	1740
ttcgagtttg atgtttttct tggatatttc cactcctctt cagagtacag aagattaagt	1800
gagaagtctg tttgtgcaag cttatcgata agctttaatg cggtagttta tcacagttaa	1860
attgctaacg cagtcaggca ccgtgtatga aatctaaca tgcgctcatc gtcctcctcg	1920
gcaccgtcac cctggatgct gtaggcatag gcttggttat gccggtactg ccgggcctct	1980
tgcgggatat cgtccattcc gacagcatcg ccagtcacta tggcgtgctg ctacgctat	2040
atgcgttgat gcaatttcta tgcgcacccg ttctcggagc actgtccgac cgctttggcc	2100
gccgcccagt cctgctcgct tcgctacttg gagccactat cgactacgag atcatggcga	2160
ccacacccgt cctgtggatc tatcgaatct aaatgtaagt taaaatctct aaataattaa	2220
ataagtccca gtttctccat acgaacctta acagcattgc ggtgagcatc tagaccttca	2280
acagcagcca gatccatcac tgcttgcca atatgtttca gtccctcagg agttacgtct	2340
tgtgaagtga tgaacttctg gaagggtgca gtgttaactc cgctgtattg acgggcatat	2400
ccgtacgttg gcaaagtgtg gttggtaccg gaggagtaat ctccacaact ctctggagag	2460
taggcaccaa caaacacaga tccagcgtgt tgtacttgat caacataaga agaagcattc	2520

tcgatttgca	ggatcaagtg	ttcaggagcg	tactgattgg	acattttccaa	agcctgctcg	2580
taggttgcaa	ccgatagggg	tgtagagtgt	gcaatacact	tgcgtacaat	ttcaaccctt	2640
ggcaactgca	cagcttggtt	gtgaacagca	tcttcaattc	tggcaagctc	cttgtctgtc	2700
atatcgacag	ccaacagaat	cacctgggaa	tcaataccat	gttcagcttg	agacagaagg	2760
tctgaggcaa	cgaaatctgg	atcagcgtat	ttatcagcaa	taactagaac	ttcagaaggc	2820
ccagcaggca	tgtcaatact	acacagggct	gatgtgtcat	tttgaaccat	catcttggca	2880
gcagtaacga	actggtttcc	tggaccaa	atcttgtcac	acttaggaac	agtttctggt	2940
ccgtaagcca	tagcagctac	tgcctggg	cctcctgcta	gcacgatata	cttagcacca	3000
accttgtggg	caacgtagat	gacttctggg	gtaagggtac	catccttctt	aggtggagat	3060
gcaaaaacaa	tttctttgca	accagcaact	ttggcaggaa	cacccagcat	cagggaagtg	3120
gaaggcagaa	ttgcggttcc	accaggaata	tagaggccaa	ctttctcaat	aggtcttgca	3180
aaacgagagc	agactacacc	agggcaagtc	tcaacttgca	acgtctccgt	tagttgagct	3240
tcatggaatt	tcctgacgtt	atctatagag	agatcaatgg	ctctcttaac	gttatctggc	3300
aattgcataa	gttctcttgg	gaaaggagct	tctaacacag	gtgtcttcaa	agcgactcca	3360
tcaaacttgg	cagttagtct	taaaagggtt	ttgtcaccat	tttgacgaac	attgtcgaca	3420
attggtttga	ctaattccat	aatctgttcc	gttttctgga	taggacgacg	aagggcatct	3480
tcaatttctt	gtgaggaggc	cttagaaacg	tcaattttgc	acaattcaat	acgaccttca	3540
gaagggactt	cttttaggtt	ggattcttct	ttaggttggt	ccttggtgta	tcctggcttg	3600
gcattctcct	tccttctagt	gacctttagg	gacttcatat	ccaggtttct	ctccacctcg	3660
tccaacgtca	caccgtactt	ggcacatcta	actaatgcaa	aataaaataa	gtcagcacat	3720
tcccaggcta	tatcttcctt	ggatttagct	tctgcaagtt	catcagcttc	ctccctaatt	3780
ttagcggtca	acaaaacttc	gtcgtcaa	aaccgtttgg	tataagaacc	ttctggagca	3840
ttgtctttac	gatccacaaa	ggtggcttcc	atggctctaa	gacctttgta	ttggccaaaa	3900
caggaagtgc	gttccaagtg	acagaaacca	acacctgttt	gttcaaccac	aaatttcaag	3960
cagtctccat	cacaatccaa	ttcgataccc	agcaactttt	gagttgctcc	agatgtagca	4020
cctttatacc	acaaaccgtg	acgacgagat	tggtagactc	cagtttgtgt	ccttatagcc	4080
tccggaatag	actttttgga	cgagtacacc	aggcccaacg	agtaattaga	agagtcagcc	4140
accaaagtag	tgaatagacc	atcgggg	tcagtagtca	aagacgccaa	caaaatttca	4200
ctgacaggga	actttttgac	atcttcagaa	agttcgtatt	cagtagtcaa	ttgccgagca	4260
tcaataatgg	ggattatacc	agaagcaaca	gtggaagtca	catctaccaa	ctttgcggtc	4320
tcagaaaaag	cataaacagt	tctactaccg	ccattagtga	aacttttcaa	atcgcccagt	4380

ES 2 780 692 T3

ggagaagaaa aaggcacagc gatactagca ttagcgggca aggatgcaac tttatcaacc	4440
agggtcctat agataaccct agcgcctggg atcatccttt ggacaactct ttctgccaaa	4500
tctaggtcca aaatcacttc attgatacca ttattgtaca acttgagcaa gttgtcgatc	4560
agctcctcaa attggtcctc tgtaacggat gactcaactt gcacattaac ttgaagctca	4620
gtcgattgag tgaacttgat cagggttggtc agctggtcag cagcataggg aaacacggct	4680
tttcctacca aactcaagga attatcaaac tctgcaacac ttgcgatgac aggtagcaag	4740
ggaaatgtca tacttgaagt cggacagtga gtgtagtctt gagaaattct gaagccgtat	4800
ttttattatc agtgagtcag tcatcaggag atcctctacg ccggacgcat cgtggccgac	4860
ctgcaggggg ggggggggct ctgaggtctg cctcgtgaag aaggtgttgc tgactcatac	4920
caggcctgaa tcgccccatc atccagccag aaagtgaggg agccacggtt gatgagagct	4980
ttgttgtagg tggaccagtt ggtgattttg aacttttgct ttgccacgga acggtctgctg	5040
ttgtcgggaa gatgcgtgat ctgaccttc aactcagcaa aagttcgatt tattcaacaa	5100
agccgccgtc ccgtcaagtc agcgtaatgc tctgccagtg ttacaaccaa ttaaccaatt	5160
ctgattagaa aaactcatcg agcatcaaat gaaactgcaa tttattcata tcaggattat	5220
caataccata tttttgaaaa agccgtttct gtaatgaagg agaaaactca ccgaggcagt	5280
tccataggat ggcaagatcc tgggtatcggc ctgcgattcc gactcgtcca acatcaatac	5340
aacctattaa tttccctcgt tcaaaaataa ggttatcaag tgagaaatca ccatgagtga	5400
cgactgaatc cggtgagaat ggcaaaagct tatgcatttc tttccagact tgttcaacag	5460
gccagccatt acgctcgtca tcaaaatcac tcgcatcaac caaacggtta ttcattcgtg	5520
attgcgcctg agcgagacga aatacgcgat cgctgttaaa aggacaatta caaacaggaa	5580
tcgaatgcaa ccggcgagcagg aacactgccg gcgcatcaac aatattttca cctgaatcag	5640
gatattcttc taatacctgg aatgctgttt tcccggggat cgcagtggtg agtaaccatg	5700
catcatcagg agtacggata aaatgcttga tggtcggaag aggcataaat tccgtcagcc	5760
agtttagtct gaccatctca tctgtaacat cattggcaac gctacctttg ccatgtttca	5820
gaaacaactc tggcgcatcg ggcttcccat acaatcgata gattgtcgca cctgattgcc	5880
cgacattatc gcgagcccat ttatacccat ataaatcagc atccatgttg gaatttaatc	5940
gcggcctcga gcaagacggt tcccgttgaa tatggctcat aacaccctt gtattactgt	6000
ttatgtaagc agacagtttt attgttcatg atgatatatt tttatcttgt gcaatgtaac	6060
atcagagatt ttgagacaca acgtggcttt ccccccccc cctgcaggtc ggcatcaccg	6120
gcgccacagg tgcggttgct ggcgctata tcgccgacat caccgatggg gaagatcggg	6180
ctcgccactt cgggctcatg agcgcttgtt tcggcgtggg tatgggtggca ggccccgtgg	6240
ccgggggact gttgggcgcc atctccttgc atgcaccatt ccttgcgggc gcggtgctca	6300

acggcctcaa cctactactg ggctgcttcc taatgcagga gtcgcataag ggagagcgtc	6360
gagtatctat gattggaagt atgggaatgg tgatacccg c attcttcagt gtcttgaggt	6420
ctcctatcag attatgcccc actaaagcaa ccggaggagg agatttcatg gtaaatttct	6480
ctgacttttg gtcacacagta gactcgaact gtgagactat ctcggttatg acagcagaaa	6540
tgtccttctt ggagacagta aatgaagtcc caccaataaa gaaatccttg ttatcaggaa	6600
caaacttctt gtttcgaact ttttcggtgc cttgaactat aaaatgtaga gtggatatgt	6660
cgggtaggaa tggagcgggc aaatgcttac cttctggacc ttcaagaggt atgtagggtt	6720
tgtagatact gatgccaaact tcagtgacaa cggttgctatt tcggtcaaac cattccgaat	6780
ccagagaaat caaagttggt tgtctactat tgatccaagc cagtgcggtc ttgaaactga	6840
caatagtgtg ctcggtgttt gaggtcatct ttgtatgaat aaatctagtc tttgatctaa	6900
ataatcttga cgagccaagg cgataaatac ccaaactctaa aactctttta aaacgttaaa	6960
aggacaagta tgtctgcctg tattaacccc caaatcagct cgtagtctga tcctcatcaa	7020
cttgaggggc actatcttgt tttagagaaa tttgcggaga tgcgatatcg agaaaaaggt	7080
acgtgattt taaacgtgaa atttatctca agatctctgc ctgcgcggtt tcggtgatga	7140
cgggtgaaaac ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcggg	7200
tgccggggagc agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggcgc	7260
agccatgacc cagtcacgta gcgatagcgg agtgtatact ggcttaacta tgcggcatca	7320
gagcagattg tactgagagt gcaccatatg cgggtgtgaa taccgcacag atgcgtaagg	7380
agaaaatacc gcatcaggcg ctcttcgct tcctcgctca ctgactcgct gcgctcggtc	7440
gttcgggtgc ggcgagcgggt atcagctcac tcaaaggcgg taatacgggt atccacagaa	7500
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt	7560
aaaaaggccg cggttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa	7620
aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcggtt	7680
ccccctggaa gctccctcgt gcgctctcct gttccgacct tgccgcttac cggatacctg	7740
tccgcctttc tcccttcggg aagcgtggcg ctttctcaat gctcacgctg taggtatctc	7800
agttcggtgt aggtcggtcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc	7860
gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggttaag acacgactta	7920
tcgccactgg cagcagccac tggtaacagg attagcagag cgaggatatgt aggcggtgct	7980
acagagttct tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggtatc	8040
tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa	8100
caaaccaccg ctggtagcgg tggttttttt gtttgcaagc agcagattac gcgcagaaaa	8160

aaaggatctc aagaagatcc ttgatcttt tctacggggt ctgacgctca gtggaacgaa 8220
 aactcacggt aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt 8280
 ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac 8340
 agttaccaat gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcatcc 8400
 atagttgcct gactccccgt cgtgtagata actacgatac gggaggggctt accatctggc 8460
 cccagtgcgt caatgatacc gcgagaccca cgctcaccgg ctccagattt atcagcaata 8520
 aaccagccag ccggaagggc cgagcgcaga agtggtcctg caactttatc cgcctccatc 8580
 cagtctatta attggtgccg ggaagctaga gtaagtagtt cgccagttaa tagtttgccg 8640
 aacgttggtg ccattgctgc aggcacgtg gtgtcacgct cgtcgtttgg tatggcttca 8700
 ttcagctccg gttcccaacg atcaaggcga gttacatgat ccccatggt gtgcaaaaaa 8760
 gcggttagct ccttcgggtc tccgatcgtt gtcagaagta agtggccgc agtggtatca 8820
 ctcatgggta tggcagcact gcataattct cttactgtca tgccatccgt aagatgcttt 8880
 tctgtgactg gtgagtactc aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt 8940
 tgctcttgcc cggcgtcaac acgggataat accgcgccac atagcagaac tttaaaagtg 9000
 ctcatcattg gaaaacgttc ttcggggcga aaactctcaa ggatcttacc gctgttgaga 9060
 tccagttoga tgtaaccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc 9120
 agcgtttctg ggtgagcaaa aacaggaagg caaatgccg caaaaaaggg aataagggcg 9180
 acacggaaat gttgaatact catactcttc ctttttcaat attattgaag catttatcag 9240
 gggtattgtc tcatgagcgg atacatattt gaatgtattt agaaaaataa acaaataggg 9300
 gttccgcgca catttccccg aaaagtgcc cctgacgtct aagaaaccat tattatcatg 9360
 acattaacct ataaaaatag gcgtatcacg aggccctttc gtcttcaaga attaattctc 9420
 atgtttgaca gcttatcatc gataagctga ctcatgttgg tattgtgaaa tagacgcaga 9480
 tcgggaacac tgaaaaataa cagttattat tcg 9513
 <210> 78
 <211> 1720
 <212> ADN
 <213> Artificial
 <220>
 <223> inserto pPNic602
 <400> 78
 cctgacgtct aagaaacat tattatcatg acattaacct ataaaaatag gcgtatcacg 60
 aggccctttc gtcttcaaga attaattctc atgtttgaca gcttatcatc gataagctga 120
 ctcatgttgg tattgtgaaa tagacgcaga tcgggaacac tgaaaaataa cagttattat 180
 tcgtttcagg caaccagagc tacatagtga gatcctgtct caacaaaaat aaaataatct 240

5

10

aaggcttcaa agggttcaat ctcttaggta gctaaatatg aacaaaattht gggaaatgtg 300
acctttttcct tagtgacagt cagatagaac cttctcagat gcaaggacac caagtgcaaa 360
caggctcaag aacagcctgg aaaggctctag tgctatgggg cttcaggtcg aatgccaaact 420
gtttttcaaga actgtgtgga tttttctgcc tgtaacgaat tcagattcat ttttcaaaac 480
tcggggagag ttttccccct ttataattht ttttttaaht ttattaaact ttgtttcgth 540
ccccctgtth tgagaattgc agagtcatcc accctgtcac agtgccaggg agctcaggga 600
tggggcccagg ggcctggcgg ggctgaaggg gctggggaag cgagggtcc aaagggaccc 660
cagtgtggca ggagccaaag ccctagggtcc ctagaacgca gaggccaccg ggacccccca 720
gacggggtaa gcgggtgggt gtctggggcg cgaagccgca ctgcgcatgc gccgagggtcc 780
gctccggccg cgctgatcca agccgggttc tcgcgccgac ctggtcgtga ttgacaagtc 840
acacacgctg atccctccgc ggggccgcac agggtcacag cctttccccct cccacaaaag 900
ccccctactc tctgggcacc acacacgaac attccttgag cgtgaccttg ttggctctag 960
tcaggcgcct ccggtgcaga gactggaacg gccttgggaa gtagtcccta accgcatttc 1020
cgcgagggga tcgtcgggag ggcgtggctt ctgaggatta tataaggcga ctccgggcgg 1080
gtcttagcta gttccgtcgg agaccgagat tcagtcgccg cttctctgtg aggactgctg 1140
ccgccgccgc tggtagaggag aagccgccgc gcttggcgta gctgagagac ggggaggggg 1200
cgcggacacg aggggcagcc cgcggcctgg acgttctgtt tccgtggccc gcgaggaagg 1260
cgactgtcct gagggcggagg acccagcggc aagatggcgg ccaagtggaa gcctgagggg 1320
ataggcgagc ggccctgagg cgctcgacgg ggttgggggg gaagcaggcc cgcgaggcag 1380
ctgcagccgg gaacgtgcgg ccaaccctt atththththt acgggttgcg ggccgtaggt 1440
gcctccgaag tgagagccgt gggcgthtga ctgtcgggag aggtcggtcg gattthtcatc 1500
cgthtctaaa gacggaagtg cgactgagac gggaaggggg gggagtcggt tgggtggcggt 1560
tgaacctgga ctaaggcgca catgacgacg cggthtctat gggctcataa tgggtggtga 1620
ggacattthc ctagatctaa catccaaaga cgaaaggtht aatgaaacct tthtgccatc 1680
cgacatccac aggtccattc tcacacataa gtgccaaaacg 1720
<210> 79
<211> 455
<212> ADN
<213> Artificial
<220>
<223> protomor de EF1a
<400> 79
ggatccttgg agctaagcca gcaatggtag agggaagatt ctgcacgtcc cttccaggcg 60

5

10

ES 2 780 692 T3

	gctccccgt caccaccccc cccaacccgc cccgaccgga gctgagagta attcatacaa	120
	aaggactcgc ccctgccttg gggaatccca gggaccgtcg ttaaactccc actaacgtag	180
	aaccagaga tcgctgcgtt cccgccccct caccgccccg ctctcgtcat cactgaggtg	240
	gagaagagca tgcgtgaggc tccggtgccc gtcagtgggc agagcgcaca tcgcccacag	300
	tccccgagaa gttgggggga ggggtcggca attgaaccgg tgcctagaga aggtggcgcg	360
	gggtaaactg ggaaagtgat gtcgtgtact ggctccgcct ttttcccgag ggtgggggag	420
	aaccgtatat aagtgcagta gtcgccgtga acgtt	455
5	<210> 80 <211> 2250 <212> ADN <213> Artificial <220> <223> EEE1-A1 <400> 80	
	tgcagtgcc a gacactacat agcctgatat actctagtgt taactgagaa ggagcaccat	60
	tcccaccata ccttaaatat aatttaaaag agagagtgtg agaaactcaa agtgaacaag	120
	actaaaagag actaagacac aaggcagaat aatacactcc ctcatggtgc tgactttgcc	180
	caatttcatac aggcagtctt ggggttaagcc taatcatcac tgttatacag catgattttg	240
	gtccacattc aggtcaccca agacacagta cagggctggg ttttatatat tgatatcaca	300
	gaggaaatat gtcttttggc ctgggtctgc agtatggagt gtgtttttct tgtgcagggt	360
	aggcagcacc atgcttttcc tcctccttct gcctcagcat tggaggggta gtaggagagg	420
	agctctagaa ccggttagaa ggagattaca taatcctttt ttatcctatt atatgaatga	480
	aagatttggc ctatgttctt ggcttttggg ataatgactt acaatactac ttatggttgt	540
	ggcccttacc agcgcggtgg gtggtgctgc gcttttagtc ctgatattgg ggaggcagag	600
	gaaggaggat cactctcact acgtgggcac ctggtgtttt acgagactac cacctctggc	660
	tgtgttacag tgaatagcac agaccctacc aatgcatata acaacaacta gatctattgt	720
	gccacgaagg gtagctctcc agcctctctt atagttagga cctcatggaa ggaagagggc	780
	aggaagctgt tctttggtag atacaccta ctagagttca ttagtgagtg catgcaccta	840
	atataaatat ttgatctaag gagcataggc tactaatctc ttaggtgggtt attttcactt	900
	ctatatattg gaaagagtcc tataccctag ctaaagttac aaagaacctt ctcgactcct	960
	aggtctggaa gtgcaaaagg gtgaaagaac agctgctatg ctgttgaggt ttggcccctt	1020
	cccctcgtaa ggctagtcta tactacatca gagagcaata tacaggcaga atcgtaatga	1080
	catacttata cctatacacg ggcacactat accccataaa attatacttt agagcaataa	1140
10	attatcatag atacgatggg catctatagg taattctatg accatccacc ctgcatctgt	1200

ggctgcctgg tgaggcaagc cccagggggg cagccggggc agtagcccct ggggtaccgt 1260
 ggggtgctat gcctccccac tctgggaggt ggctatgggg ttgctgcttt gtacggacag 1320
 ggctccggga cccccctgtc ggccttaccg gcagggagac aggcccgcg agccggagtc 1380
 cggaagcgcc gtggtgcggt gcgcccgcgc tgatcctacc cgctatgtcg cgccgtcctg 1440
 gacgagttag tctactgaga gacgctgac cctccgcggg gcccgagac actgagaccc 1500
 tatgcccacc ctagattacc cctcaatctc tgggcaccac acacgaacat tccttgagcg 1560
 tgaccttggt ggctctagtc aggcgcctcc ggtgcagaga ctggaacggc cttgggaagt 1620
 agtccctaac cgcatttccg cggaggggac gtcgggaggg cgtggcttct gaggattata 1680
 taaggcgact ccgggcgggt cttagctagt tccgtcggag acccgagttc agtcgccgct 1740
 tctctgtgag gactgctgcc gccgccgctg gtgaggagaa gccgccgcgc ttggcgtagc 1800
 tgagagacgg ggagggggcg cggacacgag gggcagcccg cggcctggac gttctgtttc 1860
 cgtggcccg gcaggaaggcg actgtcctga ggcggaggac ccagcggcaa gatggcggcc 1920
 aagtggaagc ctgaggggat aggcgagcgg ccctgaggcg ctcgacgggg ttggggggga 1980
 agcaggcccg cgaggcagct gcagccggga acgtgcggcc aacccttat tttttttgac 2040
 gggttgcggg ccgtaggtgc ctccgaagtg agagccgtgg gcgtttgact gtcgggagag 2100
 gtcggtcgga tttcatccg ttgctaaaga cggaagtgcg actgagacgg gaaggggggg 2160
 gagtcggttg gtggcggttg aacctggact aaggcgcaca tgacgtcgcg gtttctatgg 2220
 gctcataatg ggtggtgagg acatttcctt 2250
 <210> 81
 <211> 1449
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1-A2
 <400> 81
 tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
 gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120
 ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag 180
 gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt 240
 ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
 cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
 gccagggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480

5

10

tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac	540
ggggtaagcg ggtgggtgtc tggggcgcgga agccgcactg cgcattgcgcc gaggtccgct	600
ccggccgcgc tgatccaagc cgggttctcg cggcgacctg gtcgtgattg acaagtcaca	660
cacgctgac cctccgcggg gccgcacagg gtcacagcct tccccctccc cacaaagccc	720
cctactgtct ggccatctct ctcttagaat gcatcaccga gtgctagatg ccacaactga	780
ccgcctccg ctctgtgtc agcataggcc ttgggaagta gtccttttagc ggaataccgc	840
gcaccgttcg acgtgagggc gtggctattg aggattatat aaggcgtcac cggccggctg	900
taacctagtt ccgtcgcaca gccgagagta cccgccgctt ctgagagtgc agtccaggcg	960
ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcg	1020
ggacacgagg ggcagcccgc ggcctggacg ttctgtttcc gtggcccgcg aggaaggcga	1080
ctgtcctgag gcgaggacc cagcggcaag atggcggcca agtggaagcc tgaggggata	1140
ggcgagcggc cctgaggcgc tcgacggggg tgggggggaa gcaggcccgc gaggcagctg	1200
cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgccggc cgtaggtgcc	1260
tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt	1320
tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttgg tggcggttga	1380
acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgagga	1440
catttccct	1449
<210> 82	
<211> 1449	
<212> ADN	
<213> Artificial	
<220>	
<223> EEE1-A3	
<400> 82	
tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataactaag	60
gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc	120
ttttccttag tgacagtcag atagaacctt ctgagtgca aggacaccaa gtgcaaacag	180
gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt	240
ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaaactcg	300
gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc	360
cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg	420
gccagggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag	480
tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac	540
ggggtaagcg ggtgggtgtc tggggcgcgga agccgcactg cgcattgcgcc gaggtccgct	600

5

10

ES 2 780 692 T3

ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660

cacgctgata cctccgcggg gccgcacagg gtcacagcct tccccctccc cacaaagccc 720

cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca 780

ggcgccctcc gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttccgc 840

ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcggggtc 900

ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960

ccgccgctgg tgaggtcaat ccgccgcggg agccgattgt ctctctcggg gagggcccg 1020

gctgtcgtcc cctggggcgc ggggacctcg aagacaaagc gacctgcgcg aggaaggcga 1080

ctgtcctgag gcggaggtgc ctccggcaag atggcgctgt tcacctttgg actccccatt 1140

cccgtccggg ggactcccg acgtcgccca accccccctt cgtccggcgc gtccctccac 1200

tgctcgctt cgaccgccct tggggaattt ttttttctcg ggttgccggc cgatccaccc 1260

agcgttcact ctcccgctgg cgtttgtgac acgcctctcc acggtcgcta aaagtagcgt 1320

tgctaaagac gcttcaccgt gactgacggg aagggggggg agtcgcaacc acccggttct 1380

tggacctgaa aggcggtgta ctcgtcgcgc aaagataccc gactgggggtt ttggtgagga 1440

catttccct 1449

<210> 83

<211> 1449

<212> ADN

<213> Artificial

<220>

<223> EEE1-B1

<400> 83

tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataactaag 60

gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120

ttttccttag tgacagtcag ataactgcat tccgagtga aggacaccaa gtgcaaacag 180

gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt 240

ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaaactcg 300

gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360

cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420

gccagggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480

tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540

ggggtaagcg ggtgggtgtc tggggcgcg agccgcactg cgcatgcgcc gaggtccgct 600

ccggccgcgc ttgcacaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660

5

10

cacgcttgca cctccgcggg gccgcacagg gtcacagcct ttcccctccc caciaagccc	720
cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca	780
ggcgccctccg gtgcagagac tggaacggcc ttgggctgag atacctaacc gcatttccgc	840
ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcgggctc	900
ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg	960
ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcgc	1020
ggacacgagg ggcagcccg cgcctggacg ttctgtttcc gtggcccgcg aggaaggcga	1080
ctgtcctgag gcggaggacc cagcgtagag caggcggcca agtggagacc tgaggggata	1140
ggcgagcggc cctgagggcg tcgacggggg tgggggggaa gcaggcccg caggcagctg	1200
cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc cgtaggtgcc	1260
tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt	1320
tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttgg tggcggttga	1380
acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gaggggtgtga	1440
catttccct	1449
<210> 84	
<211> 1449	
<212> ADN	
<213> Artificial	
<220>	
<223> EEE1-B2	
<400> 84	
tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataataag	60
gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc	120
ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag	180
gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtccaat gccaaactgtt	240
ttcaagaact gtgtggatth ttctgcctgt aaccaattca gattcatttt tcaaaactgg	300
gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttccttccc	360
cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg	420
gcccaggggc ctgggggggc tgaaggggct ggggaagcca gggctccaaa gggaccccag	480
tgtggcagga gccaaagccc taggtcccta gaaccagag gccaccggga cccccagac	540
ggggtaaggg ggtgggtgtc tggggcccca agccccactg cccatgggcc cagggtcccct	600
ccggccgcgc tgatccaagc cgggttcttg cccccacctg gtggtgattg acaagtcaca	660
cacgctgac cctccgcggg gccgcacagg gtcacagcct ttcccctccc caciaagccc	720
cctactctct gggcaccaca caggaaacatt ccttgagcct gaccttggtg gctctagtca	780

ES 2 780 692 T3

ggggcctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttcccc 840
 ggagggatcg tggggagggc gtggcttctg aggattatat aaggccactc ctgggggggtc 900
 ttagctagtt ccctgggaga cccagttca gtccccctt ctctgtgagg actgctgccc 960
 ccgccgctgg tgaggagaag cccctcgct tggcgtagct gagagacggg gagggggcgc 1020
 ggacacgagg ggcagcccc tgcctggacc ttctgtttcc gtggcccgcg aggaaggcga 1080
 ctgtcctgag gcggaggacc cagcggcaag atggcggcca agtggaagcc tgaggggata 1140
 ggccaggggc cctgaggccc tccagggggt tgggggggaa gcaggccgc caggcagctg 1200
 cagctgggaa cctgcggcca accccttatt ttttttgacg ggttgggggc cgtaggtgcc 1260
 tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt 1320
 tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttg tggggggtga 1380
 acctggacta agggccacat gacgtccctg tttctatggg ctcataatgg gtggtgagga 1440
 catttccct 1449
 <210> 85
 <211> 1449
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1-B3
 <400> 85
 tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
 gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120
 ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag 180
 gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt 240
 ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
 cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
 gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggacccccag 480
 tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540
 ggggtaagcg ggtgggtgtc tggggcgcca agccccactg cccatgcgcc gaggtcccct 600
 cctgccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660
 cacgctgac cctccgcggg gccgcacagg gtcacagcct tccccctccc cacaaagccc 720
 cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttgttg gctctagtca 780
 ggcgccctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttccgc 840

5

10

ggaggggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcggggtc 900
 ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960
 ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcgc 1020
 ggacacgagg ggcagcccg cgcctggacg ttctgtttcc gtggcccgcg aggaaggcga 1080
 ctgtcctgag gcggaggacc cagcggcaag atggcggcca agtggagcc tgaggggata 1140
 ggcgagcggc cctgagggcg tcgacggggg tgggggggaa gcaggcccg gaggcagctg 1200
 cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgccggc cgtaggtgcc 1260
 tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt 1320
 tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttg tgccggttga 1380
 acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgagga 1440
 catttcctt 1449
 <210> 86
 <211> 1449
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1-B4
 <400> 86
 tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
 gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120
 ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag 180
 gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtccaat gccaaactgtt 240
 ttcaagaact gtgtggattt ttctgcctgt aaggaattca gattcatttt tcaaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttccttccc 360
 cttgttttga gaattgcaga gtcacccacc ctgtcacagt gccagggagc tcagggatgg 420
 gcccaggggc ctgggggggc tgaaggggct ggggaaggga gggctccaaa gggaccccag 480
 tgtggcagga gccaaagccc taggtcccta gaaccagag gccaccggga cccccagag 540
 ggggtaagcg ggtgggtgtc tggggcccga agccccactg cgcattgcccc gaggtcccct 600
 ccggccccgc tgatccaagc ctggttctcg ccccgacctg gtcgtgattg acaagtcaca 660
 cacgctgac cctccgcggg gccccacagg gtcacagcct tttccctccc cacaaagccc 720
 cctactctct gggcaccaca caccaacatt ccttgagcgt gaccttgttg gctctagtca 780
 ggggcctccg gtgcagagac tggaagggcc ttgggaagta gtcctaacc gcatttcccc 840
 ggaggggatcc tcgggagggg gtggcttctg aggattatat aaggccactc cgggcggggtc 900
 ttagctagtt ccctcggaga cccagttca gtcgccctt ctctgtgagg actgctgccg 960

5

10

cccccgctgg tgaggagaag cccccgccct tggggtagct gagagacggg gagggggccc 1020
 ggacaggagg ggcagcccgg ggcctggacg ttctgtttcc ctggcccgcg aggaaggcca 1080
 ctgtcctgag ggggaggacc cagcggcaag atggcggcca agtgggaagcc tgaggggata 1140
 ggcgagggggc cctgaggcgc tccacggggg tgggggggaa gcaggccccc gaggcagctg 1200
 cagcctggaa cgtggggcca accccttatt ttttttgacg ggttgggggc cgtaggtgcc 1260
 tccccagtga gagccctggg cctttgactg tcgggagagag tcggtcggat tttcatccct 1320
 tgctaaagag ggaagtggga ctgagacggg aagggggggg agtgggttgg tggcggttga 1380
 acctggacta aggggcacat gacgtcccgg tttctatggg ctcataatgg gtggtgagga 1440
 catttcctt 1449
 <210> 87
 <211> 1449
 <212> ADN
 <213> Artificial
 <220>
 <223> EEE1-B5
 <400> 87
 tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
 gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaagtatta ggatgtgacc 120
 ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag 180
 gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt 240
 ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaactcg 300
 gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
 cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
 gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480
 tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccgccc agggccagac 540
 ggggtaagcg ggtgggtgtc tggggcgcgga agccgcactg cgcgtgcgcc gaggtccgct 600
 ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660
 cacgctgac cctccgcggg gccgcacagg gtcacagcct tccccctccc caciaagccc 720
 cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca 780
 ggcgcctccg gtgcagagac tggaacggcc tggatatgaa gtccctaacc gcatttccgc 840
 ggagggatcg tcggggagcc ggggtttctg aggattatat aaggcgactc cgggcggggtc 900
 ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960
 ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacgga gggggggcgc 1020

5

10

ggacacgagg ggcagcccg cgcctggacg ttctgtttcc gtggcccgcg ggaggcacgt 1080
gaggataggc acggtgcacc cagcggcaag atggcggcca agtgggaagcc tgaggggata 1140
ggcgagcggc cctgaggcgc tcgacggggt tgggggggaa gcaggcccg caggcagctg 1200
cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc cgtaggtgcc 1260
tccgaagtga gagccgtgtg cgtgtgactg tcgggagagg tcggtcggat tttcatccga 1320
tagttacaga cggaagtgcg actgaggcgg ggaggaggag agtcggttg tggcggttga 1380
acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgaggt 1440
cattcacct 1449
<210> 88
<211> 2617
<212> ADN
<213> Artificial
<220>
<223> Secuencia reguladora de la transcripción
<400> 88

tttcaggcaa ccagagctac atagtgagat cctgtctcaa caaaaataaa ataatctaag 60
gcttcaaagg gttcaatctc ttaggtagct aaatatgaac aaaatttggg aaatgtgacc 120
ttttccttag tgacagtcag atagaacctt ctcgagtga aggacaccaa gtgcaaacag 180
gctcaagaac agcctggaaa ggtctagtgc tatggggctt caggtcgaat gccaaactgtt 240
ttcaagaact gtgtggattt ttctgcctgt aacgaattca gattcatttt tcaaaaactcg 300
gggagagttt tcccccttta taattttttt tttaaattta ttaaactttg tttcgttccc 360
cttgttttga gaattgcaga gtcattccacc ctgtcacagt gccagggagc tcagggatgg 420
gcccaggggc ctggcggggc tgaaggggct ggggaagcga gggctccaaa gggaccccag 480
tgtggcagga gccaaagccc taggtcccta gaacgcagag gccaccggga cccccagac 540
ggggtaagcg ggtgggtgtc tggggcgcg agccgcactg cgcattgcgc gaggtccgct 600
ccggccgcgc tgatccaagc cgggttctcg cgccgacctg gtcgtgattg acaagtcaca 660
cacgctgac cctccgcggg gccgcacagg gtcacagcct tccccctccc cacaaagccc 720
cctactctct gggcaccaca cacgaacatt ccttgagcgt gaccttggtg gctctagtca 780
ggcgctccg gtgcagagac tggaacggcc ttgggaagta gtccctaacc gcatttccgc 840
ggagggatcg tcgggagggc gtggcttctg aggattatat aaggcgactc cgggcgggtc 900
ttagctagtt ccgtcggaga cccgagttca gtcgccgctt ctctgtgagg actgctgccg 960
ccgccgctgg tgaggagaag ccgccgcgct tggcgtagct gagagacggg gagggggcg 1020
ggacacgagg ggcagcccg cgcctggacg ttctgtttcc gtggcccgcg aggaaggcga 1080
ctgtcctgag gcggaggacc cagcggcaag atggcggcca agtgggaagcc tgaggggata 1140

ES 2 780 692 T3

```

ggcgagcggc cctgagggcg tcgacggggt tgggggggaa gcaggcccg c gaggcagctg 1200
cagccgggaa cgtgcggcca accccttatt ttttttgacg ggttgcgggc cgtaggtgcc 1260
tccgaagtga gagccgtggg cgtttgactg tcgggagagg tcggtcggat tttcatccgt 1320
tgctaaagac ggaagtgcga ctgagacggg aagggggggg agtcggttgg tggcggttga 1380
acctggacta aggcgcacat gacgtcgcgg tttctatggg ctcataatgg gtggtgagga 1440
catttccttg actatagctt tccctcagtt gtaggacagg gtttgggcct cggcctcggg 1500
ttaggtcttc cagagtgggc aggaaccgga aatccagagg ggggaaaagt gaggcctaaat 1560
tgagttttgt ttcttgctct atatggttta gagagagact cgctgcaaaa ccgtggctgg 1620
cctggaactc tagaccagaa ccctggcctt tgccgaccca catgattaga ttcaaggcct 1680
gtgccaccag cccaggcttt attattatgg tctgggattt ctgcgatttc atccctggtg 1740
ttttgggatg atgacttggt ggtcttccct cctccccctt actgtttctg tccatggcgt 1800
gtgttctaac ccaagtttgt tcttttgggg ggggtgggagg gttgcgataa aatgggatct 1860
atctctgcc tcccaacttg agatctgcct gtcagaagtc tcagtgtga gaataaaggt 1920
gtgcattggc tcagacctcg attttttttt tttttattat tttgtaggaa gtctgtagtc 1980
cttacttgat acataagacc agacaggatc tgatttcctg cctatgaatg gtagatcctc 2040
tcagtgactg cagtgtgaat ggggaccacg cttttctcca aactatgcag atagccatga 2100
aagccatgaa atgactttca gccactggta ctgcaatatc cactcaccat ttattatatg 2160
gaccaggttc accatgccta ggtggctttg cttttgagac acggtttctc tgtgtagcct 2220
tggttatggt tttttgtttg tttttttaat ttttttgggt ttttcgagac agggtttctc 2280
tgtgtagcct tggagcctat cctggcactt gctccggaga ccaggctggc ctccaactca 2340
gatctgcctg cctctgcctc ccgactgctg ggattaaaagt aaagccattc tgcaaccctg 2400
aataccactc aataggtttc ttatttgaaa tgtggtttta tgatttttat ttctggattt 2460
agaaaagaaa tcttcagaca gaagtcttca gacagaaact agctgtagtt tggctgtgtg 2520
aactaaattg gcatccattt cacagcaatc caactgtagg taccatacca cgaatatttg 2580
tcattcctga cctgtttttt gtttgtgtgt gtgacag 2617

```

REIVINDICACIONES

1. Método para la transcripción y/o expresión y, opcionalmente, la purificación del transcrito y/o proteína o polipéptido de interés producidos que comprende los pasos de:
 - a) proporcionar un constructo de ácidos nucleicos que comprende un primer promotor, un segundo promotor y una secuencia de nucleótidos de interés, en donde dichos primer promotor y segundo promotor están operativamente enlazados a dicha secuencia de nucleótidos de interés, y en donde dicho segundo promotor está flanqueado por una primera secuencia intrónica ubicada aguas arriba de dicho promotor y una segunda secuencia intrónica ubicada aguas abajo de dicho promotor; y,
 - b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada; y,
 - c) permitir que dicha célula transformada produzca un transcrito de la secuencia de nucleótidos de interés y/o exprese la proteína o el polipéptido de interés; y opcionalmente
 - d) purificar dicho transcrito y/o proteína o polipéptido de interés producidos.
2. Método según la reivindicación 1, en donde
 - dicha primera secuencia intrónica comprende al menos un sitio de empalme donador y en donde dicha segunda secuencia intrónica comprende al menos un sitio de empalme aceptor, preferiblemente en donde dicha secuencia que comprende dicho primer promotor y/o dicha primera secuencia intrónica tiene al menos un 50 % de identidad con la SEC ID n.º: 1 o 2 en toda su longitud y/o
 - en el que dicho segundo promotor tiene al menos un 50 % de identidad con la SEC ID n.º: 57 o 58 en toda su longitud.
3. Método según la reivindicación 1 o 2, en el que dicho constructo de ácidos nucleicos comprende además una secuencia reguladora de la expresión adicional, en donde preferiblemente dicha secuencia reguladora de la expresión adicional comprende o consta de un intrón.
4. Constructo de ácidos nucleicos tal como se define en el paso a) de la reivindicación 1 o como se define en las reivindicaciones 2 o 3, en donde la secuencia de nucleótidos de interés es opcional y preferiblemente donde dicha secuencia de nucleótidos opcional codifica una proteína o un polipéptido de interés, y más preferiblemente dicha proteína o polipéptido de interés son una proteína o un polipéptido heterólogos.
5. Vector de expresión que comprende un constructo de ácidos nucleicos según la reivindicación 4.
6. Célula que comprende un constructo de ácidos nucleicos según la reivindicación 4 y/o un vector de expresión según la reivindicación 5.
7. Constructo de ácidos nucleicos según la reivindicación 4 y/o un vector de expresión según la reivindicación 5 y/o una célula según la reivindicación 6 para usar como medicamento.
8. Uso in vitro de un constructo de ácidos nucleicos según la reivindicación 4 y/o un vector de expresión según la reivindicación 5 y/o una célula según la reivindicación 6 para la transcripción y/o expresión de una secuencia de nucleótidos de interés.

Fig. 1

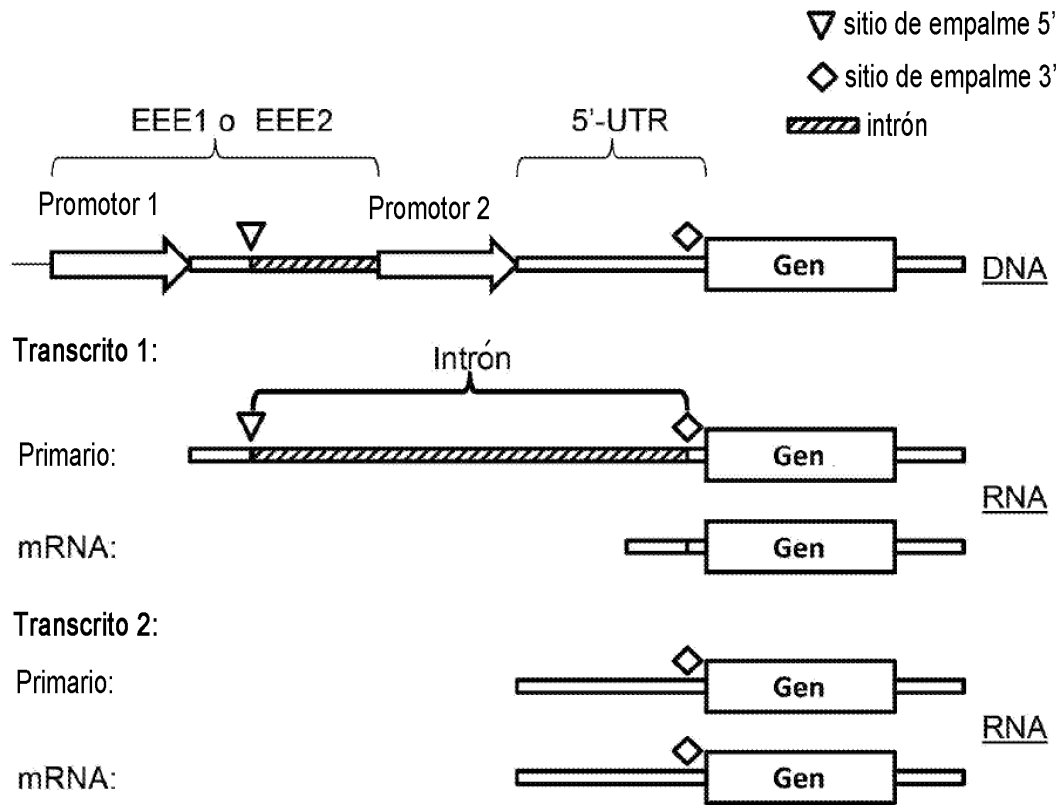


Fig. 2a

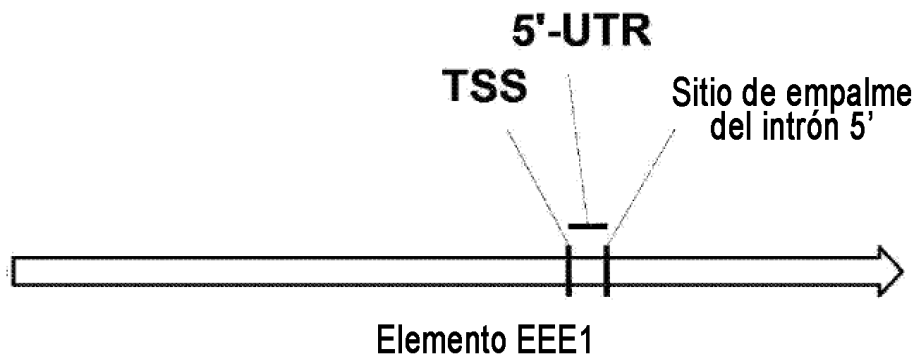


Fig. 2b

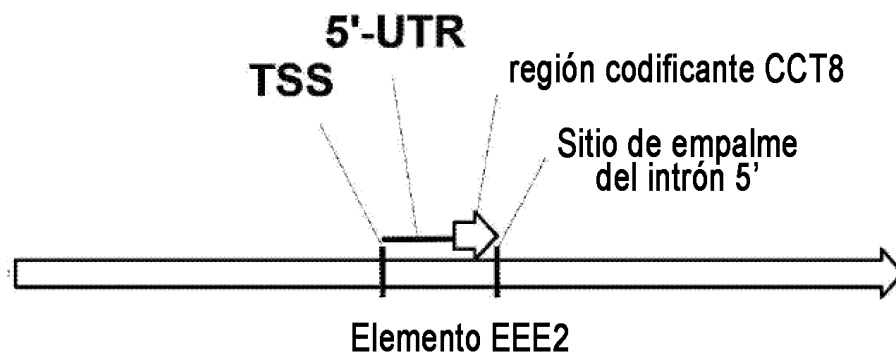


Fig. 3

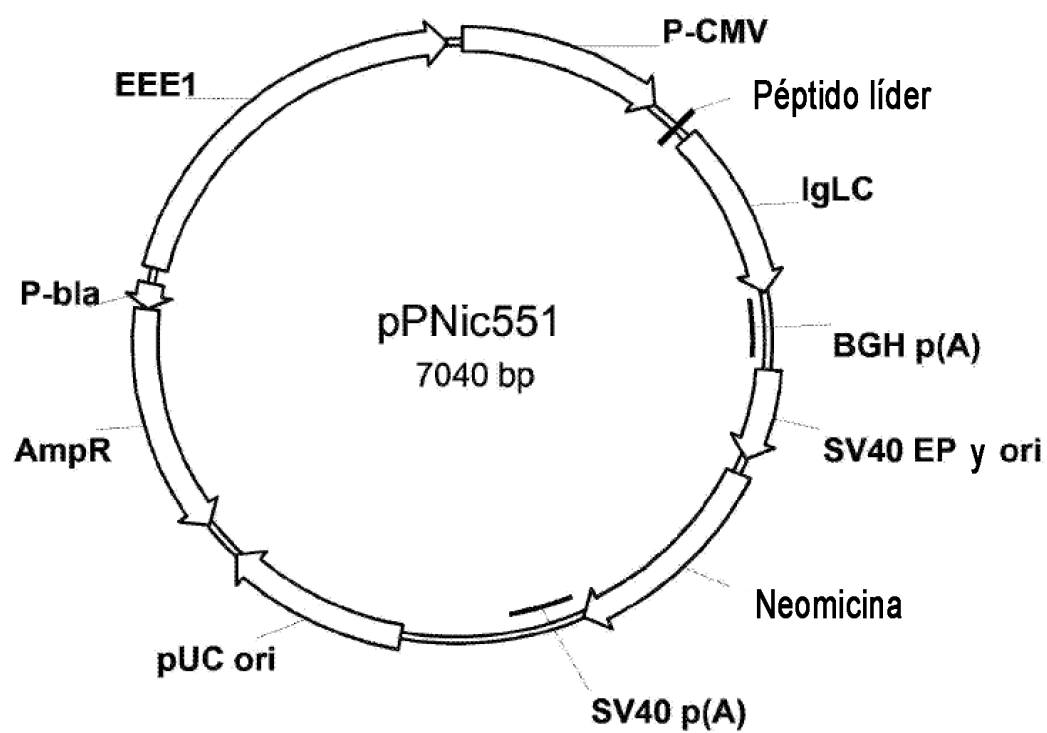


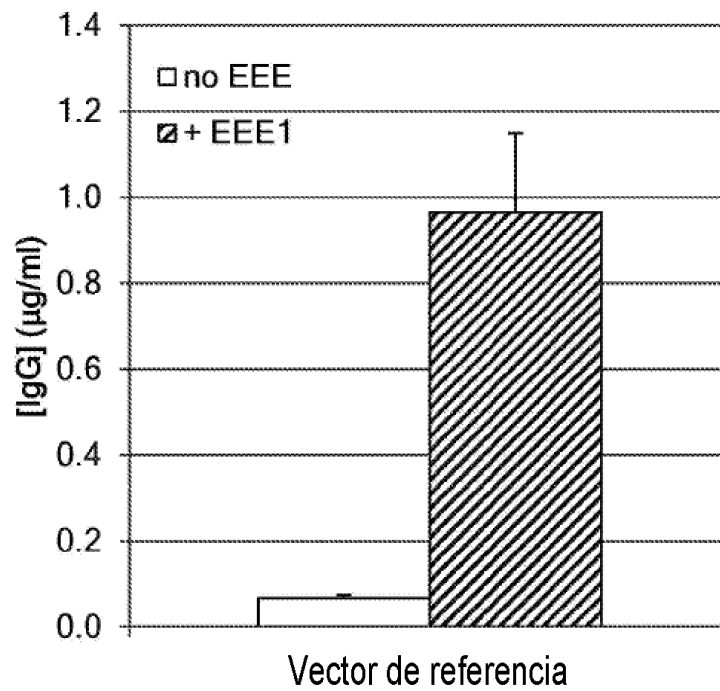
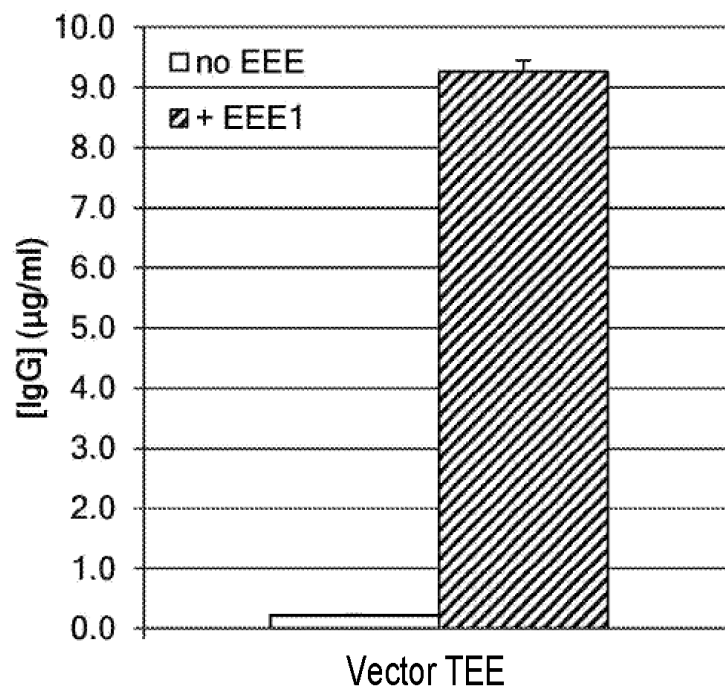
Fig. 4a*Fig. 4b*

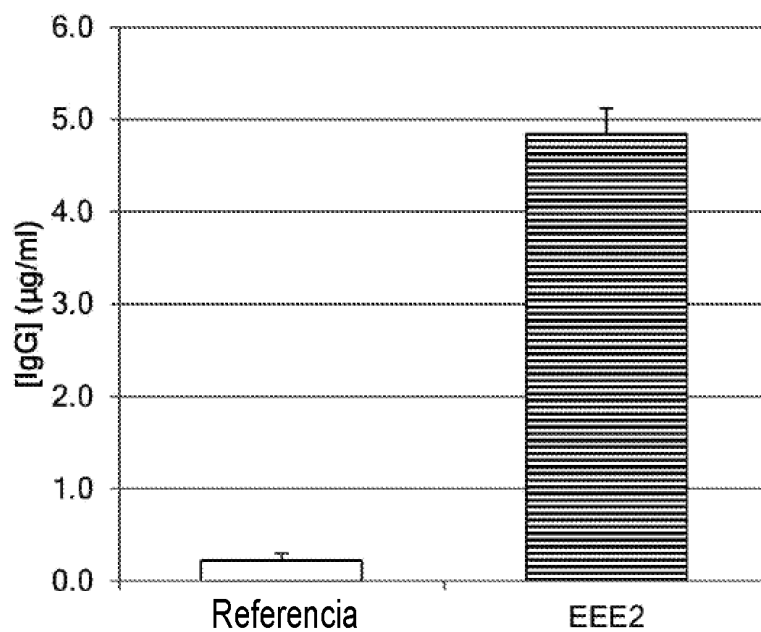
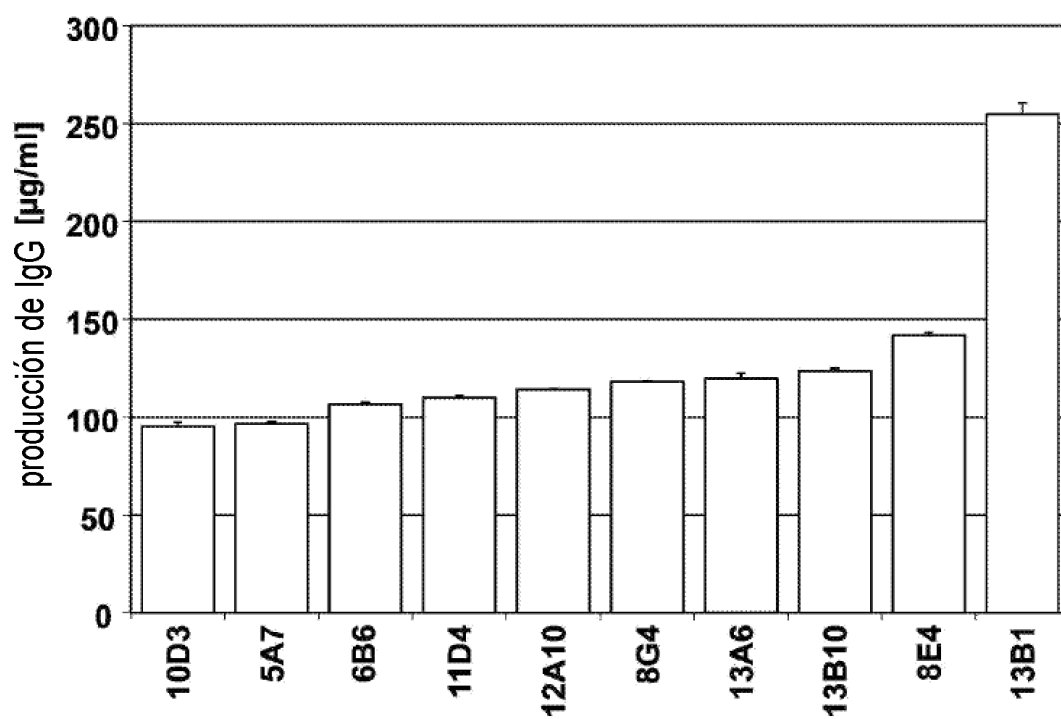
Fig. 5*Fig. 6*

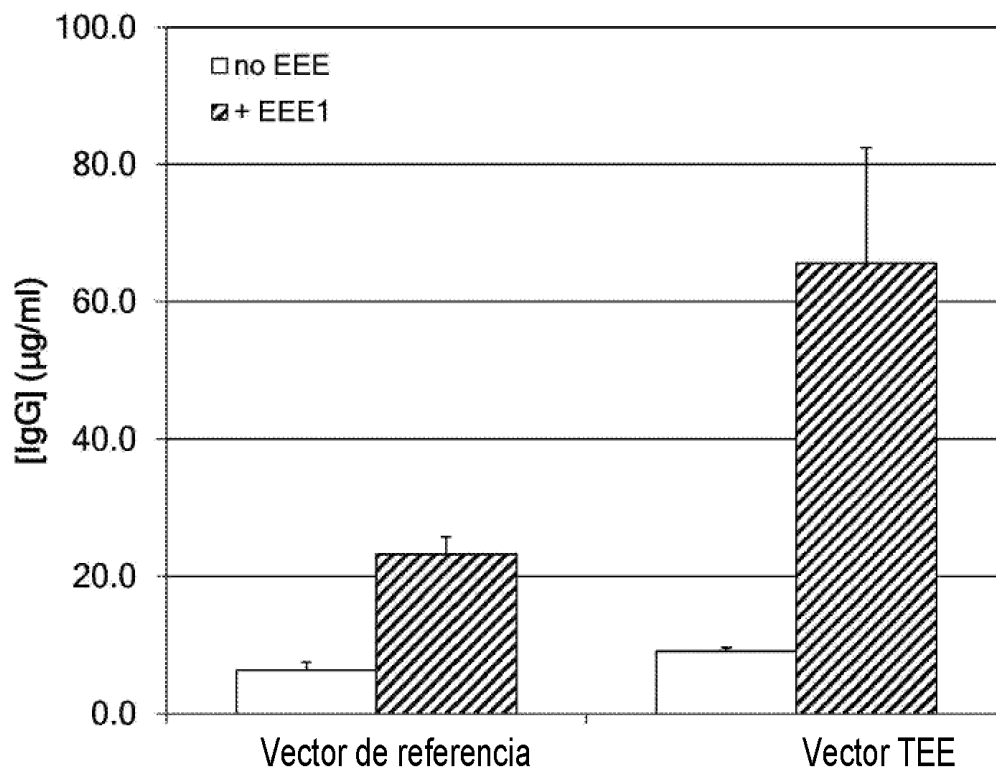
Fig. 7a

Fig. 7b

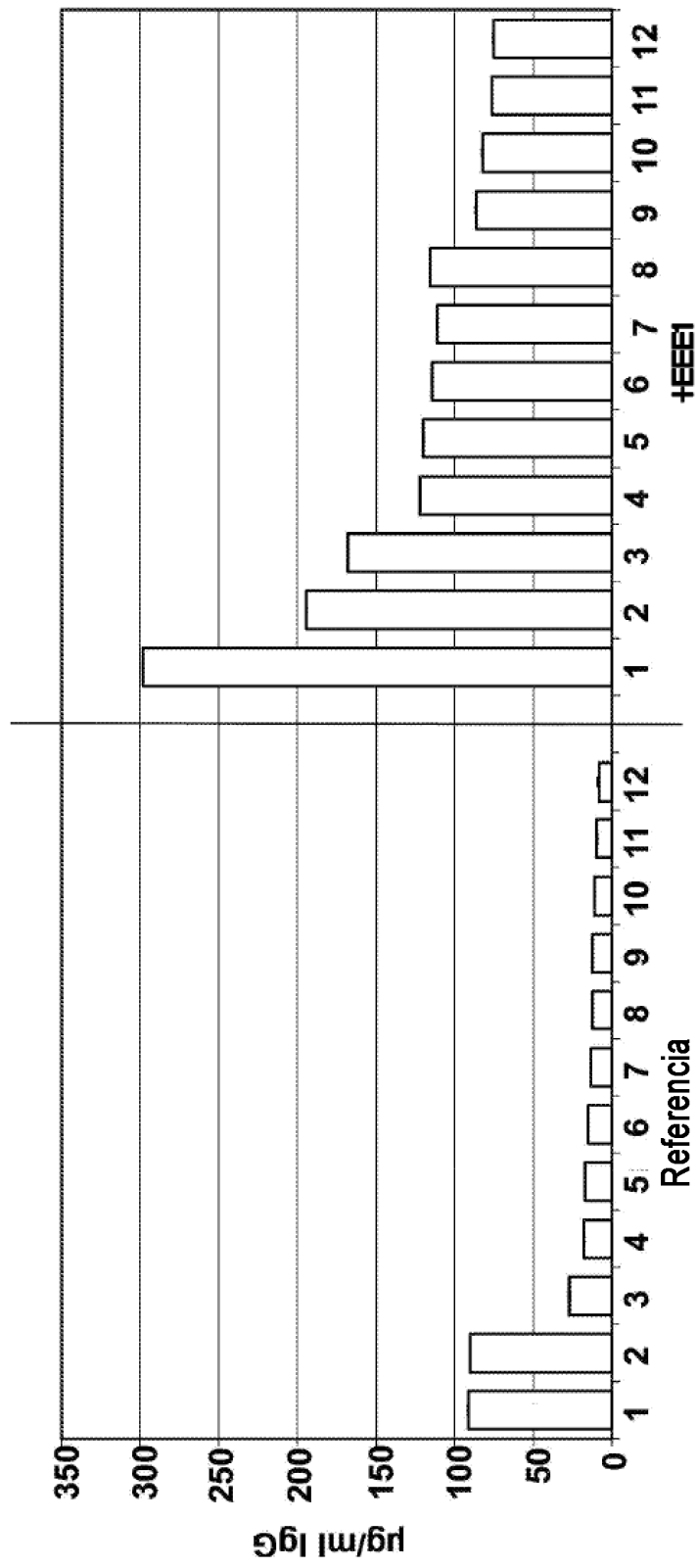


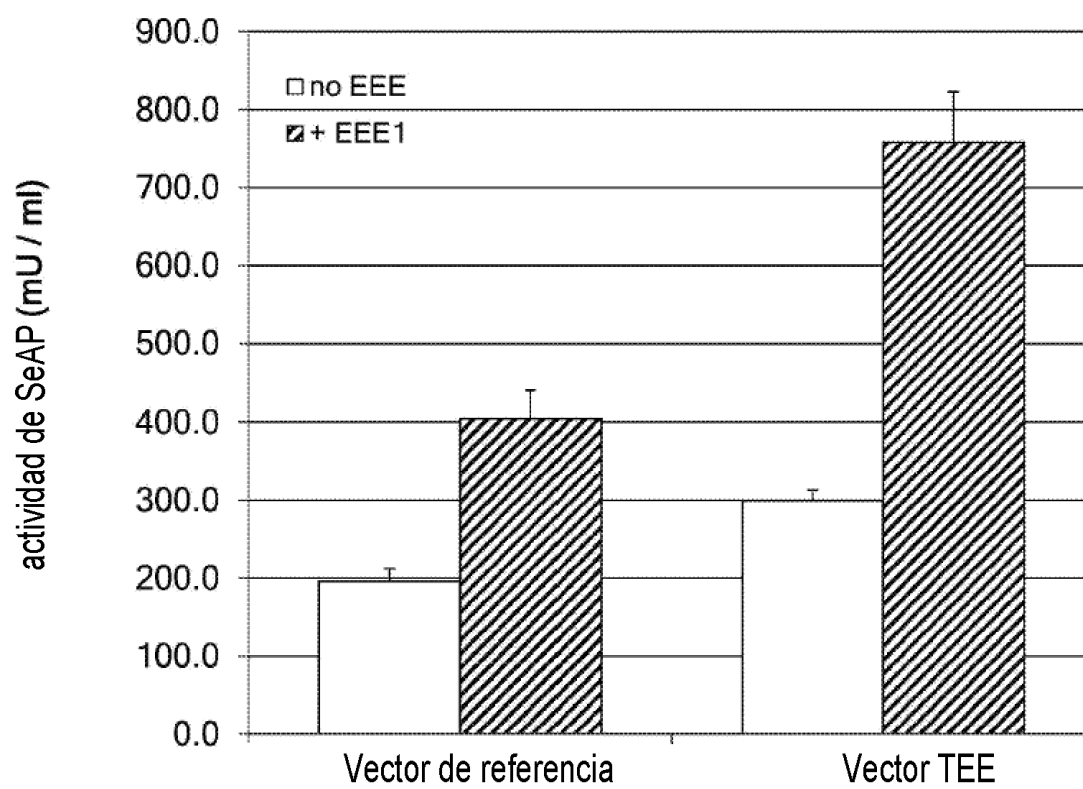
Fig. 8

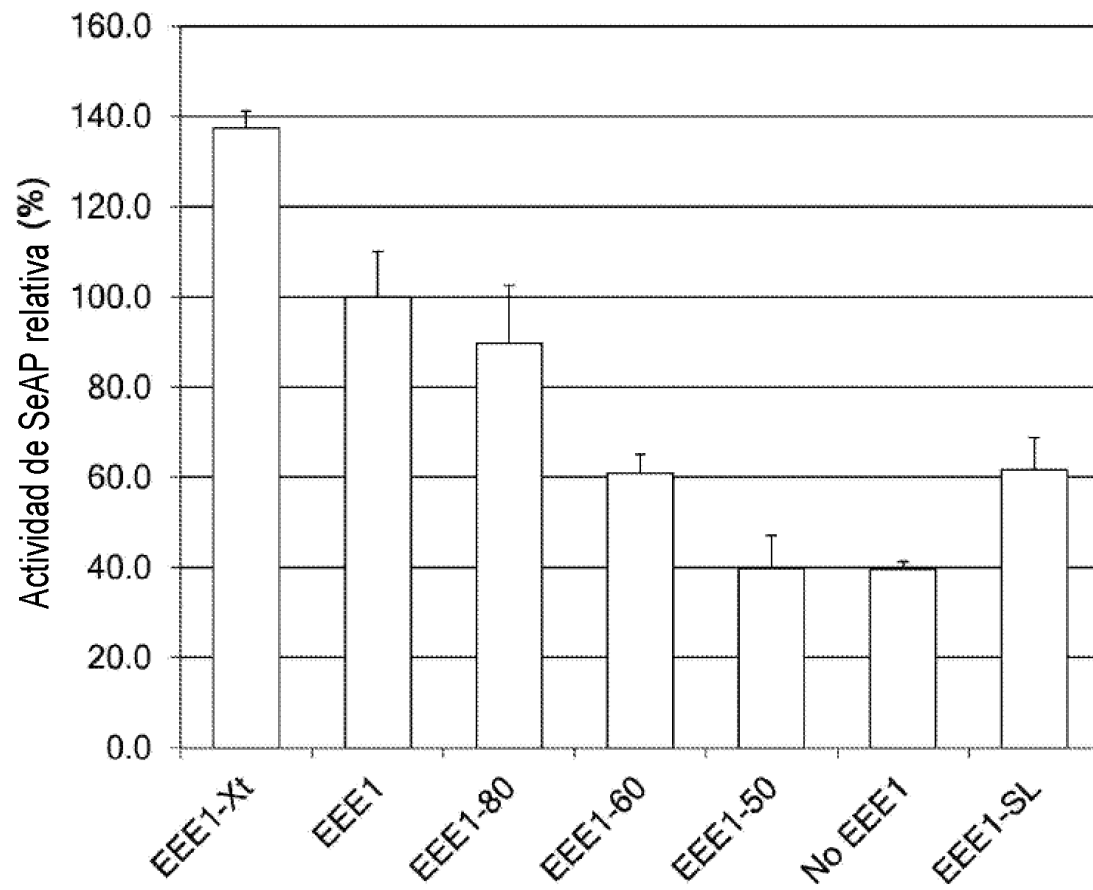
Fig. 9

Fig. 10a

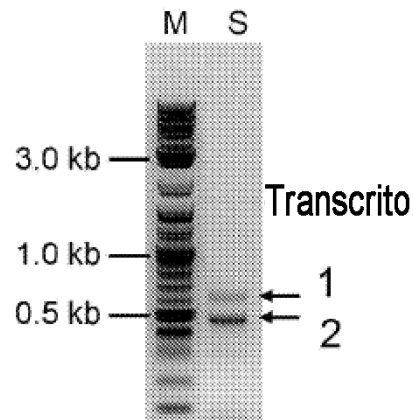


Fig. 10b

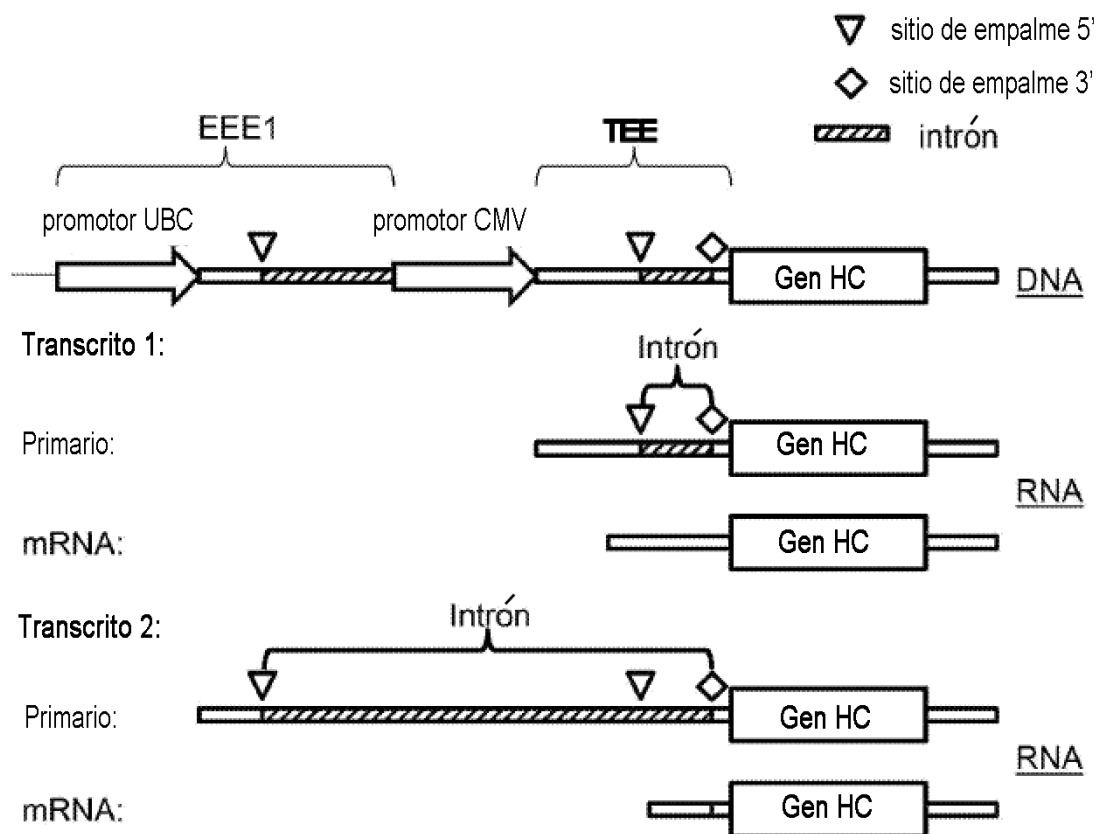


Fig. 11