

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 763 526**

51 Int. Cl.:

**C12P 21/00** (2006.01)

**C07K 16/00** (2006.01)

**C12N 15/63** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.07.2015 PCT/US2015/039086**

87 Fecha y número de publicación internacional: **07.01.2016 WO16004370**

96 Fecha de presentación y número de la solicitud europea: **02.07.2015 E 15739141 (8)**

97 Fecha y número de publicación de la concesión europea: **23.10.2019 EP 3164492**

54 Título: **Sistemas de expresión de polipéptidos**

30 Prioridad:

**03.07.2014 US 201462020838 P**

**17.11.2014 US 201462080698 P**

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

**29.05.2020**

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)  
Grenzacherstrasse 124  
4070 Basel, CH**

72 Inventor/es:

**HOTZEL, ISIDRO y  
SHANG, YONGLEI**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

ES 2 763 526 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**

Sistemas de expresión de polipéptidos

5 **CAMPO DE LA INVENCION**

La presente invención se refiere a sistemas de expresión de polipéptidos para la expresión modular y la producción de polipéptidos.

10 **ANTECEDENTES**

Los polipéptidos recombinantes se expresan en ocasiones como fusiones de dominios o marcas individuales para fines funcionales o de purificación. Los procedimientos de ADN recombinante se usan tradicionalmente para unir las secuencias que codifican cada módulo, lo que requiere una construcción diferente para cada combinación. Esto plantea un desafío para las tecnologías que implican la expresión de grandes colecciones de proteínas compuestas de módulos recurrentes unidos en diferentes combinaciones, ya que el número de construcciones se incrementa geoméricamente en función del número de módulos utilizados.

Si bien los sistemas de alto rendimiento para la subclonación pueden manejar una gran cantidad de inserciones en paralelo, normalmente requieren muchos recursos y generan una gran cantidad de construcciones que, en última instancia, no son necesarias después de las etapas iniciales de caracterización.

Schlessinger *et al.* (Journal of Immunological Methods, 2003, 282:175-186) se refieren al uso de empalme *trans* para la producción *in vivo* de anticuerpos. Los autores utilizaron el empalme *trans* de ARN dirigido para la generación de transcritos de scFv en células eucariotas, específicamente linfocitos B. En primer lugar se evaluaron los empalmes *trans* individuales y a continuación se combinaron en una reacción doble de empalme *trans*.

Pergolizzi *et al.* (Comptes rendus-biologies, 2004, 327(8): 695-709) divulga diferentes enfoques para su uso en el empalme *trans* individual en el tratamiento, en particular para reemplazar exones mutados o para administrar toxinas a las células.

Tesar *et al.* (Protein Engineering design and selection, 2013, 26(10): 655-662) divulga un sistema de vectores para la presentación en fagos de Fab que permite la expresión de un fragmento Fab fusionado a una proteína de la cubierta del fago M13 en *Escherichia coli* y la expresión de la IgG completa en células de mamífero sin subclonación. Para lograr esto, la proteína de la cubierta se incluyó en un intrón sintético optimizado dentro del gen de cadena pesada de inmunoglobulina. En las células de mamíferos, este intrón se elimina del ARN mediante empalme *cis*.

Por tanto, existe una necesidad insatisfecha en el campo del desarrollo de un sistema de expresión de polipéptidos que permita la expresión modular y la producción de polipéptidos recombinantes.

40 **SUMARIO**

La presente invención se refiere a sistemas de expresión de polipéptidos para la expresión modular y la producción de polipéptidos.

La materia objeto de la invención se define por las reivindicaciones adjuntas.

En un aspecto, la invención presenta un sistema de expresión de polipéptidos que comprende una primera molécula de ácido nucleico y una segunda molécula de ácido nucleico, en el que: (a) la primera molécula de ácido nucleico comprende un primer casete de expresión que comprende los siguientes componentes: (i) un primer promotor eucariota (P1<sub>Euk1</sub>), (ii) una primera secuencia que codifica polipéptidos (PES1<sub>1</sub>), (iii) un primer sitio de empalme en 5' (5'ss1<sub>1</sub>) y (iv) una secuencia de hibridación (HS1), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como P1<sub>Euk1</sub>-PES1<sub>1</sub>-5'ss1<sub>1</sub>-HS1; el primer casete de expresión comprende además un módulo promotor procariota escindible (ePPM<sub>1</sub>) que comprende los siguientes componentes: (i) un sitio de empalme en 5' (5'ss1<sub>2</sub>), (ii) un promotor procariota (P1<sub>Prok1</sub>) y (iii) un sitio de empalme en 3' (3'ss1<sub>1</sub>), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como 5'ss1<sub>2</sub>-P1<sub>Prok1</sub>-3'ss1<sub>1</sub>, y en el que el ePPM<sub>1</sub> se sitúa entre el P1<sub>Euk1</sub> y la PES1<sub>1</sub>; y (b) la segunda molécula de ácido nucleico comprende los siguientes componentes: (i) un promotor eucariota (P2<sub>Euk</sub>), (ii) una secuencia de hibridación capaz de hibridar con HS1 (HS2), (iii) un sitio de empalme en 3' (3'ss2), (iv) una secuencia que codifica polipéptidos (PES2), y (v) un sitio de poliadenilación (pA2), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como P2<sub>Euk</sub>-HS2-3'ss2-PES2-pA2. En algunos modos de realización, el P1<sub>Euk1</sub> es un promotor de citomegalovirus (CMV) o un promotor del virus del simio 40 (SV40). En algunos modos de realización, el P2<sub>Euk</sub> es un promotor CMV o un promotor SV40. En algunos modos de realización, el primer casete de expresión comprende además una primera secuencia de ácido nucleico que codifica una secuencia señal eucariota (ESS1<sub>1</sub>), en el que la ESS1<sub>1</sub> se sitúa entre el P1<sub>Euk1</sub> y la PES1<sub>1</sub>. En algunos modos de realización, la ESS1<sub>1</sub> se deriva del gen de la cadena pesada variable (VH).

En algunos modos de realización, el P1<sub>Prok1</sub> se selecciona del grupo que consiste en un promotor PhoA, un promotor Tac, un promotor Lac y un promotor Tphac. En algunos modos de realización, el ePPM<sub>1</sub> comprende además una primera secuencia de ácido nucleico que codifica una secuencia señal procariota (PSS1<sub>1</sub>). En algunos modos de realización, la PSS1<sub>1</sub> se deriva del gen de enterotoxina II (stII) termoestable. En algunos modos de realización, el sistema de expresión de polipéptidos comprende además un tramo de polipirimidina situado entre la PSS1<sub>1</sub> y el 3'ss1<sub>1</sub> (PPT1<sub>1</sub>). En un modo de realización, el PPT1<sub>1</sub> comprende la secuencia de aminoácidos de TTCCTTTTTCTCTTCC (SEQ ID NO: 1). En algunos modos de realización, la PES1<sub>1</sub> no comprende un sitio críptico de empalme en 5'. En algunos modos de realización, la HS1 es un gen que codifica toda o una parte de una proteína de la cubierta o una proteína adaptadora. En algunos modos de realización, la proteína de la cubierta se selecciona del grupo que consiste en pl, pll, plll, pIV, pV, pVI, pVII, pVIII, pIX y pX del bacteriófago M13, f1 o fd. En algunos modos de realización, la proteína de la cubierta es la proteína plll del bacteriófago M13. En algunos modos de realización, el fragmento plll comprende los residuos de aminoácidos 267-421 de la proteína plll o los residuos de aminoácidos 262-418 de la proteína plll. En algunos modos de realización, la proteína adaptadora es una cremallera de leucina. En algunos modos de realización, la cremallera de leucina comprende la secuencia de aminoácidos de SEQ ID NO: 4 o 5.

En algunos modos de realización, la primera molécula de ácido nucleico comprende además un segundo casete de expresión que comprende un segundo promotor eucariota (P1<sub>Euk2</sub>), (ii) una segunda secuencia que codifica polipéptidos (PES1<sub>2</sub>) y (iii) un sitio de poliadenilación (pA1), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como P1<sub>Euk2</sub>-PES1<sub>2</sub>-pA1. En algunos modos de realización, el P1<sub>Euk2</sub> es un promotor CMV o un promotor SV40. En algunos modos de realización, el segundo casete de expresión comprende además una segunda secuencia de ácido nucleico que codifica una secuencia señal eucariota (ESS1<sub>2</sub>). En algunos modos de realización, la ESS1<sub>2</sub> se deriva del gen de la proteína de unión a inmunoglobulina murina (mBiP). En algunos modos de realización, la ESS1<sub>2</sub> comprende la secuencia de ácido nucleico de ATG AAN TTN ACN GTN GTN GCN GCN GCN CTN CTN CTN GGN, en la que N es A, T, C o G (SEQ ID NO: 6).

En algunos modos de realización, el segundo casete de expresión comprende además un módulo promotor procariota escindible (ePPM<sub>2</sub>) que comprende los siguientes componentes: (i) un sitio de empalme en 5' (5'ss1<sub>3</sub>), (ii) un promotor procariota (P1<sub>Prok2</sub>) y (iii) un sitio de empalme en 3' (3'ss1<sub>2</sub>), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como 5'ss1<sub>3</sub>-P1<sub>Prok2</sub>-3'ss1<sub>2</sub>, y en el que el ePPM<sub>2</sub> se sitúa entre el P1<sub>Euk2</sub> y la PES1<sub>2</sub>. En algunos modos de realización, el P1<sub>Prok2</sub> se selecciona del grupo que consiste en un promotor PhoA, un promotor Tac y un promotor Lac. En algunos modos de realización, el ePPM<sub>2</sub> comprende además una secuencia de ácido nucleico que codifica una secuencia señal procariota (PSS1<sub>2</sub>). En algunos modos de realización, la PSS1<sub>2</sub> se deriva del gen de enterotoxina II (stII) termoestable. En algunos modos de realización, el sistema de expresión de polipéptidos comprende además un tramo de polipirimidina situado entre la PSS1<sub>2</sub> y el 3'ss1<sub>2</sub> (PPT1<sub>2</sub>). En un modo de realización, el PPT1<sub>2</sub> comprende la secuencia de aminoácidos de TTCCTTTTTCTCTTCC (SEQ ID NO: 1). En algunos modos de realización, el segundo casete de expresión se sitúa en dirección 5' al primer casete de expresión. En algunos modos de realización, el sistema de expresión de polipéptidos comprende además un potenciador intrónico de empalme (ISE) situado entre el 5'ss1<sub>1</sub> y la HS1 (ISE1). En algunos modos de realización, el ISE1 comprende una secuencia rica en G (G-run) que comprende tres o más residuos de guanina consecutivos. En algunos modos de realización, el ISE1 comprende una secuencia rica en G (G-run) que comprende nueve residuos de guanina consecutivos. En algunos modos de realización, el sistema de expresión de polipéptidos comprende además un tramo de polipirimidina situado entre la HS2 y el 3'ss2 (PPT2). En un modo de realización, el PPT2 comprende la secuencia de aminoácidos de TTCCTTTTCCCTTCTCTCC (SEQ ID NO: 7). En algunos modos de realización, el sistema de expresión de polipéptidos comprende además un ISE situado entre la HS2 y el 3'ss2 (ISE2). En algunos modos de realización, el ISE2 comprende una secuencia rica en G (G-run) que comprende tres o más residuos de guanina consecutivos. En algunos modos de realización, el ISE2 comprende una secuencia rica en G (G-run) que comprende nueve residuos de guanina consecutivos. En un modo de realización, el 5'ss1<sub>1</sub> comprende la secuencia de aminoácidos de GTAAGA (SEQ ID NO: 8).

En algunos modos de realización, la expresión por un promotor eucariota se produce en una célula de mamífero. En algunos modos de realización, la célula de mamífero es una célula Expi293F, una célula CHO, una célula 293T o una célula NSO. En algunos modos de realización, la célula de mamífero es una célula Expi293F. En algunos modos de realización, la expresión por un promotor procariota se produce en una célula bacteriana. En algunos modos de realización, la célula bacteriana es una célula de *E. coli*. En algunos modos de realización, la PES1<sub>1</sub> codifica todo o una parte de un anticuerpo. En algunos modos de realización, la PES1<sub>1</sub> codifica un polipéptido que comprende un dominio VH. En algunos modos de realización, el polipéptido comprende además un dominio CH1. En algunos modos de realización, la PES2 codifica todo o una parte de un anticuerpo. En algunos modos de realización, la PES2 codifica un polipéptido que comprende un dominio CH2 y un dominio CH3. En algunos modos de realización, la PES1<sub>2</sub> codifica todo o una parte de un anticuerpo. En algunos modos de realización, la PES1<sub>2</sub> codifica un polipéptido que comprende un dominio VL y un dominio CL.

En otro aspecto, la invención presenta un conjunto de vectores que comprende un primer vector y un segundo vector, en el que el primer y el segundo vector comprenden la primera y la segunda molécula de ácido nucleico, respectivamente, de cualquiera de los sistemas de expresión de polipéptidos divulgados en el presente documento.

En otro aspecto, la invención presenta células huésped que comprenden los conjuntos de vectores anteriores. En algunos modos de realización, la célula huésped es una célula procariota. En algunos modos de realización, la célula procariota es una célula bacteriana. En algunos modos de realización, la célula bacteriana es una célula de *E. coli*. En otros modos de realización, la célula huésped es una célula eucariota. En algunos modos de realización, la célula eucariota es una célula de mamífero. En algunos modos de realización, la célula de mamífero es una célula Expi293F, una célula CHO, una célula 293T o una célula NSO. En un modo de realización, la célula de mamífero es una célula Expi293F.

En otro aspecto, la invención presenta un procedimiento para producir un polipéptido que comprende cultivar una célula huésped que comprende uno o más de los ácidos nucleicos, vectores y/o conjuntos de vectores anteriores en un medio de cultivo. En algunos modos de realización, el procedimiento comprende además recuperar el polipéptido de la célula huésped o del medio de cultivo.

### Breve descripción de los dibujos

La FIG. 1 es un diagrama esquemático que muestra la organización relativa de las moléculas de ácido nucleico pDV2 y pRK-Fc de un sistema de expresión de polipéptidos para expresión modular de proteínas. El diagrama también muestra los productos pre-ARNm generales después de la transcripción de las moléculas de ácido nucleico en una célula eucariota, el evento de empalme *trans* esperado entre dos productos pre-ARNm generados y los productos resultantes después de la traducción de las moléculas de ARNm empalmadas.

La FIG. 2, que engloba las FIGS. 2A y 2B, es un diagrama de secuencia parcial del vector pDV2. Los sitios de empalme 5'ss y 3'ss y el tramo de polipirimidina (PPT) están en negrita y subrayados. La región que codifica la secuencia del gen III de 150 nt que hibrida con el transcrito derivado de pRK-Fc y pRK-Fc2 está en cursiva y subrayada en la Fig. 2B. Las mutaciones en el PPT natural están en negrita, cursiva y subrayadas. El sitio potencial de poliadenilación AATAAA en el gen III del vector pDV2 se muestra encima de la secuencia de las mutaciones silenciosas (en negrita, cursiva y subrayadas) introducidas en los vectores pDV2 variantes pDV2c y pDC2d. Las Figs. 2A y 2B divulgan las SEQ ID NO: 2, 3, 9, 10 y 19-22, respectivamente, en orden de aparición.

La FIG. 3 es un diagrama de secuencia parcial del vector pRK-Fc. La secuencia consenso del punto de ramificación (BP), el tramo de polipirimidina y el sitio de empalme 3'ss están en negrita o en negrita y subrayados y denotados en la Fig. 3. La secuencia del gen III antisentido de 150 pb está en cursiva y subrayada. El primer codón ATG sin cambio de pauta de lectura después del promotor CMV está en negrita, cursiva y subrayado. La Fig. 3 divulga las SEQ ID NO: 23 y 24, respectivamente, en orden de aparición.

La FIG. 4A es un gráfico que muestra el efecto de añadir secuencias de ISE o de eliminar el motivo potencial de poliadenilación del gen III en pDV2 y de complementar los vectores pRK-Fc y pRK-Fc2 a los niveles de expresión de IgG (en µg/ml) en células Expi293F.

La FIG. 4B es un gráfico que muestra el efecto de las proporciones de plásmidos de pDV2c y pRK-Fc2 a los niveles de expresión de IgG (en µg/ml) en células Expi293F. Los valores mostrados son valores promedio y error estándar de la media de un experimento representativo de dos experimentos independientes realizados por triplicado.

La FIG. 5, que engloba las FIGS. 5A y 5B, es un diagrama de secuencia parcial del vector pDV2b. Los sitios de empalme 5'ss y 3'ss se denotan y las secuencias están en negrita. El tramo de polipirimidina (PPT) y el ISE rico en G (G-run) de 9 nt se denotan y están en negrita y resaltados, respectivamente. La región que codifica la secuencia del gen III de 150 nt que hibrida con el transcrito derivado de pRK-Fc y pRK-Fc2 está en cursiva en la Fig. 5B. Las mutaciones en la secuencia señal de stII natural y el gen M13 III están en negrita y cursiva, con los residuos de nucleótidos naturales mostrados encima de la secuencia. El motivo del sitio de poliadenilación AATAAA potencial se muestra encima de la secuencia. Los aminoácidos entre paréntesis están codificados por *E. coli* y también por codones creados por empalme en células de mamíferos. Los sitios de restricción BsiWI y RsrII en el extremo 3' de la secuencia señal utilizada para la clonación de secuencia de región variable se muestran encima de las secuencias. Las Figs. 5A y 5B divulgan las SEQ ID NO: 2, 25, 9, 10, 19, 20, 26 y 27, respectivamente, en orden de aparición.

La FIG. 6 es un diagrama de secuencia parcial del vector pRK-Fc2. La secuencia consenso del punto de ramificación (BP), el tramo de polipirimidina y el sitio de empalme 3'ss están en negrita. El ISE rico en G (G-run) de 9 nt está resaltado. La secuencia del gen III antisentido de 150 pb está en cursiva. El primer triplete ATG y el codón de parada sin cambio de pauta de lectura están subrayados. El primer codón ATG sin cambio de pauta de lectura después del promotor CMV está en negrita y cursiva. La secuencia TATA del promotor TATA CMV y el sitio de inicio de la transcripción se indican encima de la secuencia. El residuo de ácido glutámico entre paréntesis está codificado por un codón creado por empalme *trans* en células de mamífero. La Fig. 6 divulga las SEQ ID NO: 28 y 29, respectivamente, en orden de aparición.

La FIG. 7A es un conjunto de gráficos que muestran la masa desconvolucionada del análisis de espectrometría de masas de las cadenas pesadas (panel izquierdo) y ligeras (panel derecho) de IgG purificada expresada en células Expi293F.

La FIG. 7B es una tabla que muestra las masas esperadas y observadas de las cadenas pesadas y ligeras de la Fig. 7A.

5 La FIG. 8A es un gráfico que muestra los rendimientos (en mg) de moléculas de IgG de 5 especificidades diferentes purificadas del sobrenadante (30 ml) de cultivos celulares de células Expi293F cotransfectados con los vectores pDV2d (que contiene ISE y sin el motivo AATAAA en el gen III) y pRK-Fc2, n=3. Las barras de error muestran el error estándar de la media.

10 La FIG. 8B es un gráfico que muestra los rendimientos (en mg) de moléculas de IgG de 5 especificidades diferentes purificadas del sobrenadante (30 ml) de cultivos celulares de células 293T y CHO cotransfectados con los vectores pDV2d (que contiene ISE y sin el motivo AATAAA en el gen III) y pRK-Fc2, n=4. Las barras de error muestran el error estándar de la media.

15 La FIG. 9 es un diagrama de secuencia parcial del vector pRK-Fab-Flag que muestra la región entre la secuencia TATA del promotor TATA CMV y la región de la bisagra superior de IgG1 humana fusionada con la secuencia de la marca Flag. Las secuencias la bisagra y de la marca Flag van seguidas de una señal de poliadenilación SV40 (no mostrada). Los sitios de empalme 3'ss, que incluyen los tramos de polipirimidina y el punto de ramificación (BP) consenso, están indicados y en negrita o están en negrita y subrayados. La secuencia del ISE se denota y está en negrita, subrayada y en cursiva. La secuencia del gen III antisentido que media en la hibridación con el transcrito donante está en cursiva y subrayada. La Fig. 9 divulga las SEQ ID NO: 30 y 31, respectivamente, en orden de aparición.

25 La FIG. 10 es un diagrama esquemático que muestra la organización relativa de las posibles primera y segunda moléculas de ácido nucleico para la expresión modular general de un producto polipeptídico. El diagrama también muestra los productos pre-ARNm generales después de la transcripción de las moléculas de ácido nucleico en una célula eucariota, el evento de empalme *trans* esperado entre dos productos pre-ARNm generados y los productos resultantes después de la traducción de las moléculas de ARNm empalmadas.

30 La FIG. 11 es un diagrama esquemático que muestra la organización relativa de las posibles primera y segunda moléculas de ácido nucleico para la expresión modular general de más de un producto polipeptídico. El diagrama también muestra los productos pre-ARNm generales después de la transcripción de las moléculas de ácido nucleico en una célula eucariota, el evento de empalme *trans* esperado entre dos productos pre-ARNm generados y los productos resultantes después de la traducción de las moléculas de ARNm empalmadas.

35 La FIG. 12 es un conjunto de inmunoelctrotransferencias que muestran la expresión de la proteína de fusión de cadena pesada Mab1 y Fd-cP3 en células Expi293F cotransfectadas con variantes de pDV2 y con pRK-Fc2. Los lisados de Expi293F transfectados se redujeron con ditioneitol (DTT) y se analizaron mediante inmunoelctrotransferencia con anticuerpos anti-Fc de IgG1 (panel superior) o anti-M13 p3 (panel inferior). Los vectores de control GFP y HC expresan la proteína fluorescente verde y la cadena pesada de IgG1 humana, respectivamente. HC indica la cadena pesada de IgG1 humana de longitud completa. Gen III AATAAA indica la presencia del sitio potencial de poliadenilación en el gen III. Fc\* indica un supuesto producto de expresión de fragmento Fc citoplasmático, truncado en el extremo N. NA indica no aplicable.

45 La FIG. 13 es un gel de electroforesis en gel de poliacrilamida con dodecilsulfato de sodio (SDS-PAGE) que muestra el análisis de fragmentos Fab e IgG purificados expresados en células Expi293F. Las IgG y Fab purificadas expresadas en sobrenadantes de células Expi293F cotransfectadas con pDV2d y pRK-Fc2 (IgG) o pRK-Fab-F (fragmento Fab) se purificaron, se resolvieron mediante SDS-PAGE con gradiente de 4 a 20 % en condiciones reductoras o no reductoras y se tiñeron con azul brillante de Coomassie. Las identidades de las bandas se muestran a la derecha. HC, cadena pesada. LC, cadena ligera. Fd, fragmento Fd de cadena pesada (VH + CH1 + bisagra superior). Las bandas de HC y LC (no reducidas) son cadenas pesadas y ligeras que no pueden formar enlaces disulfuro intercatenarios, pero que pueden tener enlaces disulfuro intracatenarios en las muestras de IgG. La banda de aproximadamente 25 kDa en la muestra de Fab no reducida tiene cadenas pesadas y ligeras que migran conjuntamente y que no formaron enlaces disulfuro intercatenarios, pero que pueden tener enlaces disulfuro intracatenarios.

50 La FIG. 14 es un gráfico que muestra la presentación en fagos de fragmentos Fab con el fagémido pDV2 detectado mediante ensayo de inmunoadsorción enzimática (ELISA) en fagos. El fago Fab-zip se produjo mediante infección de células de *E. coli* que albergan el fagémido pFab-zip con el fago auxiliar M13KO7. El fago pDV2 se produjo infectando células de *E. coli* que albergan el vector pDV2d con el fago Amber-2614 KO7.

## 60 DESCRIPCIÓN DETALLADA DE LOS MODOS DE REALIZACIÓN DE LA INVENCÓN

### I. DEFINICIONES

65 El término "anticuerpo" se usa en el presente documento en el sentido más amplio y engloba diversas estructuras de anticuerpo, incluyendo, pero sin limitarse a, anticuerpos monoclonales, anticuerpos policlonales, anticuerpos

multiespecíficos (por ejemplo, anticuerpos biespecíficos) y fragmentos de anticuerpo, siempre que presenten la actividad de unión a antígeno deseada.

El sistema de numeración de Kabat se usa, en general, cuando se hace referencia a un residuo en el dominio variable (aproximadamente los residuos 1-107 de la cadena ligera y los residuos 1-113 de la cadena pesada) (por ejemplo, Kabat *et al.*, Sequences of Immunological Interest, 5.<sup>a</sup> ed., Public Health Service, National Institutes of Health, Bethesda, Md. (1991)). El "sistema de numeración EU" o "índice EU" se usa, en general, cuando se hace referencia a un residuo en una región constante de cadena pesada de inmunoglobulina (por ejemplo, el índice EU del que se informa en Kabat *et al.*, *supra*). El "índice EU según Kabat" se refiere a la numeración de residuos del anticuerpo humano IgG1 EU. A menos que se establezca de otro modo en el presente documento, las referencias a números de residuos en el dominio variable de anticuerpos se refieren a la numeración de residuos según el sistema de numeración de Kabat. A menos que se establezca de otro modo en el presente documento, las referencias a números de residuos en el dominio constante de cadena pesada de anticuerpos se refieren a la numeración de residuos según el sistema de numeración de Kabat.

Una unidad de anticuerpo tetracatenaria básica natural es una glucoproteína heterotetramérica compuesta de dos cadenas ligeras (LC) idénticas y dos cadenas pesadas (HC) idénticas (un anticuerpo IgM consiste en 5 de las unidades heterotetraméricas básicas junto con un polipéptido adicional denominado cadena J y, por lo tanto, contiene 10 sitios de unión a antígeno, mientras que los anticuerpos IgA secretados se pueden polimerizar para formar ensamblajes polivalentes que comprenden 2-5 de las unidades tetracatenarias básicas junto con la cadena J). En el caso de las IgG, la unidad tetracatenaria es, en general, de aproximadamente 150.000 dalton. Cada LC está enlazada a una HC mediante un enlace disulfuro covalente, mientras que las dos HC están enlazadas entre sí por uno o más enlaces disulfuro dependiendo del isotipo de la HC. Cada HC y LC también tienen puentes disulfuro intracatenarios regularmente espaciados. Cada HC tiene, en el extremo N, un dominio variable (VH) seguido de tres dominios constantes (CH1, CH2, CH3) para cada una de las cadenas  $\alpha$  y  $\gamma$  y cuatro dominios CH para los isotipos  $\mu$  y  $\epsilon$ . Cada LC tiene, en el extremo N, un dominio variable (VL) seguido de un dominio constante (CL) en su otro extremo. El VL se alinea con el VH y el CL se alinea con el primer dominio constante de la cadena pesada (CH1). CH1 se puede conectar al segundo dominio constante de la cadena pesada (CH2) mediante una región bisagra. Se cree que residuos de aminoácido particulares forman una interfase entre los dominios variables de la cadena ligera y de la cadena pesada. El emparejamiento de un VH y VL juntos forma un único sitio de unión a antígeno. Para la estructura y propiedades de las diferentes clases de anticuerpos, véase, por ejemplo, Basic and Clinical Immunology, 8.<sup>a</sup> edición, Daniel P. Stites, Abba I. Terr y Tristram G. Parslow (eds.), Appleton & Lange, Norwalk, CT, 1994, página 71 y capítulo 6.

El "dominio CH2" de una región Fc de IgG humana normalmente se extiende desde aproximadamente los residuos 231 a aproximadamente 340 de la IgG. El dominio CH2 es único porque no está estrechamente emparejado con otro dominio. Más bien, dos cadenas de carbohidratos ramificadas unidas a N se interponen entre los dos dominios CH2 de una molécula de IgG natural intacta. Se ha especulado que el carbohidrato puede proporcionar un sustituto para el emparejamiento dominio-dominio y ayudar a estabilizar el dominio CH2. Burton, Molec. Immunol. 22: 161-206 (1985).

El "dominio CH3" comprende el tramo de residuos desde el extremo C hasta un dominio CH2 en una región Fc (es decir, de aproximadamente el residuo aminoácido 341 a aproximadamente el residuo aminoácido 447 de una IgG).

La cadena ligera (LC) de cualquier especie de vertebrado se puede asignar a uno de dos tipos claramente distintos, llamados kappa y lambda, en base a las secuencias de aminoácidos de sus dominios constantes. Dependiendo de la secuencia de aminoácidos del dominio constante de sus cadenas pesadas (CH), las inmunoglobulinas se pueden asignar a diferentes clases o isotipos. Existen cinco clases de inmunoglobulinas: IgA, IgD, IgE, IgG e IgM, que tienen cadenas pesadas designadas como  $\alpha$ ,  $\delta$ ,  $\gamma$ ,  $\epsilon$  y  $\mu$ , respectivamente. Las clases  $\gamma$  y  $\alpha$  se dividen además en subclases en base a diferencias relativamente menores en la secuencia y función de CH, por ejemplo, los seres humanos expresan las siguientes subclases: IgG1, IgG2, IgG3, IgG4, IgA1 e IgA2.

El término "variable" se refiere al hecho de que determinados segmentos de los dominios variables difieren ampliamente en su secuencia de un anticuerpo a otro. El dominio V media en la unión a antígeno y define la especificidad de un anticuerpo particular por su antígeno particular. Sin embargo, la variabilidad no se distribuye uniformemente a lo largo de los 110 aminoácidos de los dominios variables. En cambio, las regiones V consisten en tramos relativamente invariables denominados regiones estructurales (FR) de 15-30 aminoácidos separados por regiones más cortas de variabilidad extrema denominadas "regiones hipervariables" que tienen cada una 9-12 aminoácidos de longitud. Los dominios variables de las cadenas pesada y ligera naturales comprenden cada uno cuatro FR, adoptando en gran medida una configuración de lámina beta, conectadas por tres regiones hipervariables, que forman bucles que conectan y, en algunos casos forman parte de, la estructura de lámina beta. Las regiones hipervariables en cada cadena se mantienen unidas en estrecha proximidad mediante las FR y, con las regiones hipervariables de la otra cadena, contribuyen a la formación del sitio de unión a antígeno de los anticuerpos (véase Kabat *et al.*, Sequences of Proteins of Immunological Interest, 5.<sup>a</sup> ed., Public Health Service, National Institutes of Health, Bethesda, MD. (1991)). Los dominios constantes no están implicados directamente en la unión de un anticuerpo a un antígeno, pero presentan diversas funciones efectoras, tales como la participación del anticuerpo en la citotoxicidad celular dependiente de anticuerpos (ADCC).

Un "fragmento de anticuerpo" se refiere a una molécula distinta de un anticuerpo intacto que comprende una parte de un anticuerpo intacto que se une al antígeno al que se une el anticuerpo intacto. Los ejemplos de fragmentos de anticuerpo incluyen, pero no se limitan a, Fv, Fab, Fab', Fab'-SH, F(ab')<sub>2</sub>; diacuerpos; anticuerpos lineales; moléculas de anticuerpo monocatenario (por ejemplo, scFv) y anticuerpos multiespecíficos formados a partir de fragmentos de anticuerpo.

El fragmento "Fab" es un fragmento de unión a antígeno generado por digestión con papaína de anticuerpos y consiste en una cadena L completa junto con el dominio de la región variable de la cadena H (VH) y el primer dominio constante de una cadena pesada (CH1). La digestión de los anticuerpos con papaína produce dos fragmentos Fab idénticos. El tratamiento con pepsina de un anticuerpo proporciona un único fragmento F(ab')<sub>2</sub> grande que corresponde aproximadamente a dos fragmentos Fab unidos por enlaces disulfuro que tiene actividad de unión a antígeno divalente y aún puede reticular el antígeno. Los fragmentos Fab' difieren de los fragmentos Fab porque tienen pocos unos residuos adicionales en el extremo carboxiterminal del dominio CH1, incluyendo una o más cisteínas de la región bisagra del anticuerpo. Fab'-SH es la designación en el presente documento para un Fab' en el que el/los residuo(s) de cisteína de los dominios constantes portan un grupo tiol libre. Los fragmentos de anticuerpo F(ab')<sub>2</sub> se produjeron originalmente como pares de fragmentos Fab' que tienen cisteínas bisagra entre ellos. También son conocidos otros acoplamientos químicos de fragmentos de anticuerpo.

Una "proteína adaptadora", como se usa en el presente documento, se refiere a una secuencia de una proteína que interactúa específicamente con otra secuencia de una proteína adaptadora en solución. En un modo de realización, la "proteína adaptadora" comprende un dominio de heteromultimerización. Dichas proteínas adaptadoras incluyen una proteína de cremallera de leucina o un polipéptido que comprende una secuencia de aminoácidos de SEQ ID NO: 4 (cJUN(R): **AS**IARLE**EE**KV KTL**KAQ**NYEL **ASTAN**MLREQ VAQLGGC) o SEQ ID NO: 5 (FosW(E): **AS**JDEL**QA**EV EQLE**ER**NYAL **R**KE**VED**L**Q**K**Q** **AE**KLGGC) o una variante de las mismas (los aminoácidos en SEQ ID NO: 4 y SEQ ID NO: 5 que se pueden modificar incluyen, pero no se limitan a, los que están subrayados y en negrita), en la que la variante tiene una modificación de aminoácidos en la que la modificación mantiene o incrementa la afinidad de la proteína adaptadora por otra proteína adaptadora, o un polipéptido que comprende la secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 11 (ASIALRERVKTLRARNYELRSRANMLRERVAQLGGC) o SEQ ID NO: 12 (ASLDELEAEIEQLEENYALEKEIEDLEKELEKLGCC), o un polipéptido que comprende una secuencia de aminoácidos de SEQ ID NO: 13 (GABA-R1: EEKSRLEKE NRELEKIAE KEERVSELRH QLQSVGGC) o SEQ ID NO: 14 (GABA-R2: TSRLEGLQSE NHRLRMKITE LDKDLEEVMT QLQDVGGC) o SEQ ID NO: 15 (Cys: AGSC) o SEQ ID NO: 16 (bisagra: CPPCPG). La molécula de ácido nucleico que codifica la proteína de la cubierta o la proteína adaptadora está comprendida dentro de un intrón sintético.

Como se usa en el presente documento, "dominio de heteromultimerización" se refiere a alteraciones en o adiciones a una molécula biológica para promover la formación de heteromultímeros e impedir la formación de homomultímeros. Cualquier dominio de heterodimerización que tenga una preferencia fuerte para formar heterodímeros sobre homodímeros está dentro del alcance de la invención. Los ejemplos ilustrativos incluyen pero no se limitan a, por ejemplo, la solicitud de patente de EE. UU. 20030078385 (Arathoon *et al.* - Genentech; que describe botón en ojales); el documento WO2007147901 (Kjærgaard *et al.* - Novo Nordisk; que describe interacciones iónicas); el documento WO 2009089004 (Kannan *et al.* - Amgen; que describe efectos de conducción electrostática); el documento WO 2011/034605 (Christensen *et al.* - Genentech; que describe superhélices). Véase también, por ejemplo, Pack, P. & Plueckthun, A., *Biochemistry* 31, 1579-1584 (1992) que describe la cremallera de leucinas o Pack *et al.*, *Bio/Technology* 11, 1271-1277 (1993) que describe el motivo hélice-giro-hélice. La frase "dominio de heteromultimerización" y "dominio de heterodimerización" se usan de manera intercambiable en el presente documento.

Como se usa en el presente documento, el término "sitio de clonación" se refiere a una secuencia de ácido nucleico que contiene un sitio de restricción para la clonación mediada por endonucleasa de restricción mediante la unión de una secuencia de ácido nucleico que contiene extremos cohesivos o romos compatibles, una región de secuencia de ácido nucleico que sirve como un sitio de cebado para la clonación mediada por PCR del ADN insertado por homología y "PCR de extensión por solapamiento", o un sitio de recombinación para la inserción mediada por recombinasa de secuencias de ácido nucleico diana por reacción de recombinación-intercambio, o extremos de mosaico para la inserción mediada por transposón de secuencias de ácido nucleico diana, así como otras técnicas comunes en la técnica.

Una "proteína de la cubierta", como se usa en el presente documento, se refiere a cualquiera de las cinco proteínas de la cápside que son componentes de las partículas de fago, incluyendo pIII, pVI, pVII, pVIII y pIX. En un modo de realización, la "proteína de la cubierta" se puede usar para presentar proteínas o péptidos (véase Phage Display, A Practical Approach, Oxford University Press, editado por Clackson y Lowman, 2004, pág. 1-26). En un modo de realización, una proteína de la cubierta puede ser la proteína pIII o alguna variante, parte y/o derivado de la misma. Por ejemplo, se puede usar una parte C terminal de la proteína de la cubierta pIII (cP3) del bacteriófago M13, tal como una secuencia que codifica los residuos C terminales 267-421 de la proteína III del fago M13. En un modo de realización, la secuencia pIII comprende la secuencia de aminoácidos de SEQ ID NO: 17

(AEDIEFASGGGSGAETVESCLAKPHTENSFTNVWKDDKTLDRYANYEGCLWNATGVVVCTGDETCQY  
 GTWVPIGLAIPENEGGGSEGGGSEGGGSEGGGSKPPEYGDTPIPGYTYINPLDGTYPGTEQNPANPNP  
 SLEESQPLNTFMFQNNRFRNRQGALTVYTGTVTQGTDPVKTYTYQYTPVSSKAMYDAYWNGKFRDCAFH  
 SGFNEDPFVCEYQQSSDLPQPPVNAGGGSGGGSGGGSEGGGSEGGGSEGGGSEGGGSGGGSGS  
 GDFDYEKMANANKGAMTENADENALQSDAKGKLDVATDYGAAIDGFIGDVSGLANGNGATGDFAGSN  
 SQMAVGDGDN SPLMNNFRQYLPSLPQSVECRPFVFSAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVF  
 STFANILRNKES). En un modo de realización, el fragmento pIII comprende la secuencia de aminoácidos de SEQ ID  
 NO: 18  
 (SGGGSGSGDFDYEKMANANKGAMTENADENALQSDAKGKLDVATDYGAAIDGFIGDVSGLANGNGA  
 TGDFAGSN SQMAVGDGDN SPLMNNFRQYLPSLPQSVECRPFVFGAGKPYEFSIDCDKINLFRGVFAFL  
 LYVATFMYVFSTFANILRNKES).

5 Un "casete de expresión", como se usa en el presente documento, significa un fragmento de ácido nucleico (por ejemplo, un fragmento de ADN) que comprende secuencias de ácido nucleico específicas con actividad biológica y/o bioquímica específicas. Las expresiones "casete", "casete de genes" o "casetes de ADN" se podrían usar de manera intercambiable y tener el mismo significado.

10 Los términos "célula huésped", "línea de células huésped" y "cultivo de células huésped" se usan de manera intercambiable y se refieren a células en las que se ha introducido ácido nucleico exógeno, incluyendo la descendencia de dichas células. Las células huésped incluyen "transformantes" y "células transformadas", que incluyen la célula transformada principal y la descendencia derivada de la misma, independientemente del número de pases. La descendencia puede no ser completamente idéntica en cuanto al contenido de ácido nucleico a una célula original, sino que puede contener mutaciones. La descendencia mutante que tiene la misma función o actividad biológica que la cribada o seleccionada en la célula transformada originalmente se incluye en el presente documento.

20 Los términos "enlazado" o "se enlaza" o "enlace", como se usan en el presente documento, tienen la intención de referirse a la unión covalente de dos secuencias de aminoácidos o dos secuencias de ácido nucleico entre sí a través de enlaces peptídicos o fosfodiéster, respectivamente, dicha unión puede incluir cualquier número de secuencias de aminoácidos o de ácido nucleico adicionales entre las dos secuencias de aminoácidos o secuencias de ácido nucleico que se están uniendo.

25 "Ácido nucleico" o "polinucleótido", como se usan de manera intercambiable en el presente documento, se refieren a polímeros de nucleótidos de cualquier longitud, e incluye ADN y ARN. Los nucleótidos pueden ser desoxirribonucleótidos, ribonucleótidos, bases o nucleótidos modificados y/o sus análogos, o cualquier sustrato que se pueda incorporar en un polímero mediante ADN o ARN polimerasa, o mediante una reacción sintética. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados y sus análogos. Si están presentes, se puede conferir una modificación a la estructura de los nucleótidos antes o después del ensamblaje del polímero. La secuencia de nucleótidos se puede interrumpir por componentes no nucleotídicos. Un polinucleótido se puede modificar adicionalmente después de la síntesis, tal como mediante conjugación con un marcador. Otros tipos de modificaciones incluyen, por ejemplo, "caperuzas", sustitución de uno o más de los nucleótidos naturales por un análogo, modificaciones internucleotídicas, tales como, por ejemplo, aquellas con enlaces no cargados (por ejemplo, metilfosfonatos, fosfotriésteres, fosfoamidatos, carbamatos, etc.) y con enlaces cargados (por ejemplo, fosforotioatos, fosforoditioatos, etc.), aquellas que contienen restos laterales, tales como, por ejemplo, proteínas (por ejemplo, nucleasas, toxinas, anticuerpos, péptidos señal, poli-L-lisina, etc.), aquellas con intercalantes (por ejemplo, acridina, psoraleno, etc.), aquellas que contienen quelantes (por ejemplo, metales, metales radiactivos, boro, metales oxidativos, etc.), aquellas que contienen alquilantes, aquellas con enlaces modificados (por ejemplo, ácidos nucleicos alfa anoméricos, etc.), así como formas no modificadas del/de los polinucleótido(s). Además, cualquiera de los grupos hidroxilo presentes de manera ordinaria en los azúcares se puede reemplazar, por ejemplo, por grupos fosfonato, grupos fosfato, proteger mediante grupos protectores estándar o activar para preparar enlaces adicionales a nucleótidos adicionales, o se puede conjugar a soportes semisólidos o sólidos. El OH de extremo 5' y 3' se puede fosforilar o sustituir por aminas o restos del grupo orgánico terminal de desde 1 hasta 20 átomos de carbono. También se pueden derivatizar otros hidroxilos a grupos protectores estándar. Los polinucleótidos también pueden contener formas análogas de los azúcares ribosa o desoxirribosa que se conocen en general en la técnica, incluyendo, por ejemplo, 2'-O-metil-, 2'-O-alil-, 2'-fluoro- o 2'-azido-ribosa, análogos de azúcares carbocíclicos, azúcares alfa anoméricos, azúcares epiméricos tales como arabinosa, xilosas o lixosas, azúcares de piranosa, azúcares de furanosa, sedoheptulosas, análogos acíclicos y análogos de nucleósidos básicos tales como metil-ribósido. Se pueden reemplazar uno o más enlaces fosfodiéster por grupos de enlace alternativos. Estos grupos de enlace alternativos incluyen, pero no se limitan a, modos de realización en los que el fosfato se reemplaza por P(O)S ("tioato"), P(S)S ("dilitioato"), (O)NR<sub>2</sub> ("amidato"), P(O)R, P(O)OR', CO o CH<sub>2</sub> ("formacetal"), en los que cada R o R' es independientemente H o alquilo (1-20 C) sustituido o no sustituido que contiene opcionalmente un enlace éter (-O-), arilo, alquenilo, cicloalquilo, cicloalquenilo o araldilo. No es necesario que todos los enlaces en un polinucleótido sean idénticos. La descripción precedente se aplica a todos los polinucleótidos a los que se hace referencia en el presente

documento, incluyendo ARN y ADN.

Un ácido nucleico se "enlaza de forma funcional" cuando se dispone en una relación estructural o funcional con otra secuencia de ácido nucleico. Por ejemplo, un segmento de ADN puede estar enlazado de forma funcional a otro segmento de ADN si se sitúan uno con respecto al otro en la misma molécula de ADN contigua y tienen una relación estructural o funcional, tal como un promotor o potenciador que se sitúa en relación con una secuencia de codificación para facilitar la transcripción de la secuencia de codificación; un sitio de unión a ribosomas que se sitúa en relación con una secuencia de codificación para facilitar la traducción; o una pre-secuencia o secuencia líder secretora que se sitúa en relación con una secuencia de codificación para facilitar la expresión de una pre-proteína (por ejemplo, una pre-proteína que participa en la secreción del polipéptido codificado). En otros ejemplos, las secuencias de ácido nucleico enlazadas de forma funcional existen no son contiguas, sino que se sitúan de manera que tienen una relación funcional entre sí como ácidos nucleicos o como proteínas que se expresan por ellos. Los potenciadores, por ejemplo, no tienen que ser contiguos. El enlace se puede lograr mediante unión en sitios de restricción convenientes o mediante el uso de adaptadores o conectores oligonucleotídicos sintéticos.

El término "señal de poliadenilación" o "sitio de poliadenilación" se usa en el presente documento para significar una secuencia suficiente para dirigir la adición de ácido ribonucleico de poliadenosina a una molécula de ARN expresada en una célula.

Un "promotor" es una secuencia de ácido nucleico que permite el inicio de la transcripción de una secuencia génica en un ARN mensajero, iniciando dicha transcripción con la unión de una ARN polimerasa en o cerca del promotor.

El término "sitio de empalme en 3'" pretende significar una secuencia de ácido nucleico, por ejemplo, una secuencia de pre-ARNm, en el límite intrón/exón 3' que puede ser reconocida y unida por la maquinaria de empalme.

El término "sitio de empalme en 5'" pretende significar una secuencia de ácido nucleico, por ejemplo, una secuencia de pre-ARNm, en el límite exón/intrón 5' que puede ser reconocida y unida por la maquinaria de empalme.

El término "sitio críptico de empalme" pretende significar un sitio de empalme normalmente inactivo en 5' o 3' que se activa por una mutación o de otro modo y puede servir como un elemento de empalme. Por ejemplo, una mutación puede activar un sitio de empalme en 5' que está situada corriente abajo del sitio de empalme en 5' natural o dominante. El uso de este sitio "críptico" de empalme da como resultado la producción de distintos productos de empalme de ARNm que no se producen mediante el uso del sitio de empalme natural o dominante.

El término "empalme *trans*", tal como se usa en el presente documento, significa la unión de exones contenidos en moléculas de ARN no contiguas separadas.

El término "región variable" o "dominio variable" se refiere al dominio de una cadena pesada o ligera de anticuerpo que está implicado en la unión del anticuerpo al antígeno. Los dominios variables de la cadena pesada y de la cadena ligera (VH y VL, respectivamente) de un anticuerpo natural tienen, en general, estructuras similares, comprendiendo cada dominio cuatro regiones estructurales (FR) conservadas y tres regiones hipervariables (HVR). (Véase, por ejemplo, Kindt *et al.*, *Kuby Immunology*, 6.<sup>a</sup> ed., W.H. Freeman and Co., página 91 (2007).) Un único dominio VH o VL puede ser suficiente para conferir especificidad de unión a antígeno. Además, los anticuerpos que se unen a un antígeno particular se pueden aislar usando un dominio VH o VL de un anticuerpo que se une al antígeno para cribar una colección de dominios VL o VH complementarios, respectivamente. Véase, por ejemplo, Portolano *et al.*, *J.Immunol.* 150:880-887 (1993); Clarkson *et al.* *Nature* 352:624-628 (1991).

El término "vector", como se usa en el presente documento, se refiere a una molécula de ácido nucleico que puede propagar otro ácido nucleico al que se enlaza. El término incluye el vector como una estructura de ácido nucleico autorreplicante, así como el vector incorporado en el genoma de una célula huésped en la que se ha introducido. Determinados vectores pueden dirigir la expresión de los ácidos nucleicos a los que se enlazan funcionalmente. Dichos vectores se denominan en el presente documento "vectores de expresión".

## II. SISTEMAS MODULARES DE EXPRESIÓN DE POLIPÉPTIDOS

La presente invención se basa, al menos en parte, en el descubrimiento de que el empalme *trans* de pre-ARNm se puede explotar en células de mamífero para permitir la expresión modular de proteínas recombinantes. El concepto de expresión modular y flexible de proteínas permite la unión precisa de dos secuencias de codificación de proteínas arbitrarias codificadas por dos construcciones diferentes en un solo ARNm que codifica una cadena de polipéptido, sin ninguno de los requisitos y limitaciones de otros procedimientos de empalme proteína-proteína. Este concepto se puede adaptar para simplificar y ampliar otras tecnologías que requieren la expresión en células de mamífero de grandes colecciones de proteínas con diferentes combinaciones de módulos recurrentes.

Aquí describimos la generación de múltiples sistemas de expresión de polipéptidos que permiten la expresión modular de diferentes formatos de anticuerpos en el contexto de un sistema de expresión de presentación en fagos. Los componentes de ácido nucleico requeridos, los vectores, las células huésped y los procedimientos de uso de los

sistemas de expresión de polipéptidos de la invención se describen en el presente documento.

#### A. Modos de llevar a cabo la invención

5 La práctica de la presente invención empleará, a menos que se indique de otro modo, técnicas convencionales de biología molecular (incluyendo técnicas recombinantes), microbiología, biología celular, bioquímica e inmunología, que están dentro de la experiencia de la técnica. Dichas técnicas se explican con detalle en la literatura, tal como, "Molecular Cloning: A Laboratory Manual", 2.<sup>a</sup> edición (Sambrook *et al.*, 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.);  
10 "Handbook of Experimental Immunology", 4.<sup>a</sup> edición (D.M. Weir & C.C. Blackwell, eds., Blackwell Science Inc., 1987); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F.M. Ausubel *et al.*, eds., 1987); "PCR: The Polymerase Chain Reaction", (Mullis *et al.*, eds., 1994); y "Current Protocols in Immunology" (J.E. Coligan *et al.*, eds., 1991).

#### 15 B. Sistema modular de expresión de proteínas

Los sistemas de expresión de polipéptidos de la invención pueden soportar la expresión de polipéptidos (por ejemplo, proteínas de fusión) en la misma o diferentes formas (por ejemplo, reformateadas). La presente invención proporciona un medio para generar dichos sistemas de expresión de polipéptidos para la expresión modular y la producción de  
20 diferentes formas (por ejemplo, diferentes formatos o diferentes formas de fusión) de la proteína de interés de una manera dependiente de la célula huésped usando el procedimiento de empalme *trans*.

##### 1. Componentes de ácido nucleico del sistema modular de expresión de proteínas

##### 25 a. Estructura de los componentes de ácido nucleico del sistema modular de expresión de proteínas

El sistema de expresión de proteínas usa al menos dos moléculas de ácido nucleico que juntas permiten la expresión flexible y modular de cualquier polipéptido deseado a través del procedimiento de empalme *trans* dirigido de pre-ARNm. La primera molécula de ácido nucleico incluye un primer casete de expresión que incluye un promotor eucariota (P1<sub>EUK1</sub>) (por ejemplo, un promotor de citomegalovirus (CMV), un promotor del virus simio 40 (SV40), una región U3 del virus de la leucemia murina de Moloney, una región U3 del virus de la artritis-encefalitis caprina, una región U3 del virus visna o una secuencia de la región U3 retroviral), que está enlazado de forma funcional a una secuencia que codifica polipéptidos (PES1<sub>1</sub>). En algunos casos, la secuencia que codifica polipéptidos codifica solo una parte del polipéptido deseado, y la parte restante es suministrada por una secuencia que codifica polipéptidos (PES2) contenida  
30 en una segunda molécula de ácido nucleico. La primera molécula de ácido nucleico incluye un sitio de empalme en 5' (5'ss1<sub>1</sub>) (por ejemplo, GTAAGA (SEQ ID NO: 8) situado corriente abajo de (en dirección 3' a) la PES1<sub>1</sub>, pero corriente arriba de (en dirección 5' a) una secuencia de hibridación (HS1).

La secuencia HS1 puede contener un gen que codifica toda o una parte de una marca de polipéptido, marcador, proteína de la cubierta y/o proteína adaptadora, que se puede situar sin cambio de pauta de lectura con PES1<sub>1</sub> de modo que la expresión dé como resultado la proteína codificada por PES1<sub>1</sub> fusionada a la proteína codificada por HS1. En un caso, la HS1 es un gen que codifica toda o una parte de una proteína de la cubierta seleccionada del grupo que consiste en pI, pII, pIII, pIV, pV, pVI, pVII, pVIII, pIX y pX del bacteriófago M13, f1 o fd. Por ejemplo, la PES1<sub>1</sub> puede codificar todo o una parte de un anticuerpo o fragmento Fab del mismo y la secuencia HS1 puede codificar una proteína de la cubierta (por ejemplo, toda o una parte de la proteína pIII del bacteriófago M13, por ejemplo, un fragmento de pIII comprende los residuos de aminoácidos 267-421 de la proteína pIII o los residuos de aminoácidos 262-418 de la proteína pIII), dando como resultado un producto de fusión del anticuerpo o fragmento Fab con la proteína pIII. En otro caso, la HS1 es un gen que codifica toda o una parte de una proteína adaptadora, tal como una cremallera de leucina, en la que la cremallera de leucina comprende la secuencia de aminoácidos de SEQ ID NO: 4 o 5.  
40

Además, la primera molécula de ácido nucleico puede codificar una secuencia señal eucariota (ESS1<sub>1</sub>) situada en dirección 3' a P1<sub>Euk1</sub> y en dirección 5' a PES1<sub>1</sub>. En consecuencia, la primera molécula de ácido nucleico puede incluir los componentes anteriores unidos (por ejemplo, enlazados de forma funcional) entre sí en dirección de 5' a 3' como P1<sub>Euk1</sub>-ESS1<sub>1</sub>-PES1<sub>1</sub>-5'ss1<sub>1</sub>-HS1.  
50

La segunda molécula de ácido nucleico del sistema de expresión de proteínas incluye un promotor eucariota (P2<sub>Euk</sub>) (por ejemplo, un promotor de citomegalovirus (CMV) o un promotor del virus simio 40 (SV40)), que está enlazado de forma funcional a una secuencia que codifica polipéptidos (PES2). En algunos casos, la secuencia que codifica polipéptidos codifica solo una parte del polipéptido deseado, y la parte restante es suministrada por la secuencia que codifica polipéptidos contenida en la primera molécula de ácido nucleico (PES1<sub>1</sub>). La segunda molécula de ácido nucleico incluye un sitio de empalme en 3' (3'ss2) situado en dirección 5' a PES2. La segunda molécula de ácido nucleico puede incluir una secuencia de hibridación capaz de hibridarse con HS1 (HS2), que se sitúa entre P2<sub>Euk</sub> y 3'ss2. Además, la segunda molécula de ácido nucleico incluye un sitio de poliadenilación (pA2), en el que los componentes de la segunda molécula de ácido nucleico están enlazados de forma funcional entre sí en una dirección 5' a 3' como P2<sub>Euk</sub>-HS2-3'ss2-PES2-pA2.  
60  
65

El empalme *trans* entre el primer y el segundo producto de pre-ARNm de ácido nucleico en una célula eucariota (por ejemplo, una célula de mamífero) sería por lo tanto inducido por la hibridación de secuencias complementarias (es decir, HS1 y HS2) localizadas en las moléculas de ARNm separadas, de modo que el único sitio de empalme en 5' de la primera molécula (5'ss<sub>1</sub>) y el único sitio de empalme en 3' de la segunda molécula (3'ss<sub>2</sub>) se acercan para que se produzca el empalme *trans* y apoyen la formación del transcrito de ARNm con empalme *trans* deseado. Además, para promover el empalme *trans*, la primera molécula de ácido nucleico puede incluir un potenciador intrónico de empalme (ISE) situado entre el 5'ss<sub>1</sub> y la HS1 (ISE1). El ISE1 puede, por ejemplo, incluir una secuencia rica en G que tiene tres o más residuos de guanina consecutivos, tal como una secuencia rica en G que tiene nueve residuos de guanina consecutivos. Además, el empalme *trans* entre el primer y el segundo producto de pre-ARNm de ácido nucleico se puede inducir tras su transcripción en células eucariotas (por ejemplo, células de mamífero, por ejemplo, células Expi293F, 293T o CHO) genomanipulando la primera molécula de ácido nucleico para que carezca de un sitio de poliadenilación estándar corriente abajo de su componente PES<sub>1</sub> y/o HS1. Esto minimizaría la formación de transcritos de ARNm maduros que se exportarían al citoplasma antes de que se pueda producir el empalme *trans* con el transcrito de ARNm de la segunda molécula de ácido nucleico.

En algunos casos, puede ser deseable expresar concomitantemente un producto polipeptídico separado. Por ejemplo, puede ser deseable expresar un segundo producto polipeptídico que se pueda autoensamblar con el primer producto polipeptídico codificado por la primera y la segunda molécula de ácido nucleico para formar un producto proteico heteromultimérico deseado (por ejemplo, un anticuerpo que está compuesto de cadenas pesadas y ligeras). Para este fin, la primera y/o segunda molécula de ácido nucleico puede incluir adicionalmente un segundo casete de expresión. Por ejemplo, en los casos en los que la primera molécula de ácido nucleico incluye un segundo casete de expresión, el segundo casete de expresión puede incluir un segundo promotor eucariota (P<sub>1Euk2</sub>), (ii) una segunda secuencia de ácido nucleico que codifica una secuencia señal eucariota (ESS<sub>12</sub>), (iii) una segunda secuencia que codifica polipéptidos (PES<sub>12</sub>), y (iv) un sitio de poliadenilación (pA1), en el que los componentes están enlazados de forma funcional entre sí en una dirección 5' a 3' como P<sub>1Euk2</sub>-ESS<sub>12</sub>-PES<sub>12</sub>-pA1. En algunos casos, el segundo casete de expresión puede no incluir un componente ESS<sub>12</sub> (por ejemplo, cuando la secreción del polipéptido expresado no es necesaria o deseable). En consecuencia, la primera molécula de ácido nucleico codificaría dos productos polipeptídicos en promotores separados, con lo que uno de los transcritos de ARNm que codifica uno de los productos polipeptídicos de la primera molécula de ácido nucleico se formó a través de un empalme *trans* directo con un transcrito de ARNm codificado por la segunda molécula de ácido nucleico. En algunos casos, el segundo casete de expresión se sitúa en dirección 5' al primer casete de expresión. En otros casos, el segundo casete de expresión se sitúa en dirección 3' respecto al primer casete de expresión.

#### *b. Expresión de polipéptidos en células procariontas y eucariotas*

El sistema de expresión de polipéptidos está diseñado para la expresión de polipéptidos en el contexto de células tanto procariontas como eucariotas. En consecuencia, la primera molécula de ácido nucleico incluye un módulo promotor procarionta escindible (ePPM<sub>1</sub>) que se sitúa entre el P<sub>1Euk1</sub> y la PES<sub>1</sub>. El ePPM<sub>1</sub> puede incluir un sitio de empalme en 5' (5'ss<sub>12</sub>), un promotor procarionta (P<sub>1Prok1</sub>), una secuencia de ácido nucleico que codifica una secuencia señal procarionta (PSS<sub>11</sub>) y un sitio de empalme en 3' (3'ss<sub>11</sub>) localizados uno con respecto al otro en dirección de 5' a 3' como 5'ss<sub>12</sub>-P<sub>1Prok1</sub>-PSS<sub>11</sub>-3'ss<sub>11</sub>, y enlazados de forma funcional para dar lugar a la transcripción del polipéptido codificado por PES<sub>1</sub> o PES<sub>1</sub> y HS1. En algunos casos, el ePPM<sub>1</sub> puede no incluir un componente PSS<sub>11</sub> (por ejemplo, cuando la secreción del polipéptido expresado no es necesaria o deseable). Por tanto, el ePPM<sub>1</sub> daría lugar a la transcripción del polipéptido codificado por PES<sub>1</sub> de la primera molécula de ácido nucleico en una célula procarionta. Por otra parte, en una célula eucariota (por ejemplo, célula de mamífero), el P<sub>1Euk1</sub> daría lugar a la expresión de la transcripción del polipéptido codificado por PES<sub>1</sub> de la primera molécula de ácido nucleico, y el ePPM<sub>1</sub> se eliminaría del transcrito de pre-ARNm mediante empalme *cis* en virtud de los resultados positivos que flanquean los componentes 5'ss<sub>12</sub> y 3'ss<sub>11</sub>.

En algunos casos, el ePPM<sub>1</sub> también incluye un tramo de polipirimidina situado entre la PSS<sub>11</sub> y el 3'ss<sub>11</sub> (PPT<sub>11</sub>). El PPT<sub>11</sub> puede incluir la secuencia de, por ejemplo, TTCCTTTTTCTCTTCC (SEQ ID NO: 1). La segunda molécula de ácido nucleico también puede incluir un tramo de polipirimidina (PPT<sub>2</sub>) que, por ejemplo, se puede situar entre la HS2 y el 3'ss<sub>2</sub>. El PPT<sub>2</sub> puede incluir la secuencia de, por ejemplo, TTCCTGTTCCCTTCTCTCC (SEQ ID NO: 7). Además, la segunda molécula de ácido nucleico puede incluir además un ISE situado entre la HS2 y el 3'ss<sub>2</sub> (ISE2). El ISE2 puede, por ejemplo, incluir una secuencia rica en G que tiene tres o más residuos de guanina consecutivos, tal como una secuencia rica en G que tiene nueve residuos de guanina consecutivos.

En algunos modos de realización en los que la primera molécula de ácido nucleico del sistema de expresión de polipéptidos incluye un segundo casete de expresión, el segundo casete de expresión puede incluir además un módulo promotor procarionta escindible (ePPM<sub>2</sub>) situado entre el P<sub>1Euk2</sub> y la PES<sub>12</sub> e incluye los siguientes componentes: (i) un sitio de empalme en 5' (5'ss<sub>13</sub>), (ii) un promotor procarionta (P<sub>1Prok2</sub>), (iii) una secuencia de ácido nucleico que codifica una secuencia señal procarionta (PSS<sub>12</sub>) y (iv) un sitio de empalme en 3' (3'ss<sub>12</sub>), con lo que los componentes se localizan uno con respecto al otro en una dirección de 5' a 3' como 5'ss<sub>13</sub>-P<sub>1Prok2</sub>-PSS<sub>12</sub>-3'ss<sub>12</sub>, y están enlazados de forma funcional para dar lugar a la transcripción del polipéptido codificado por PES<sub>12</sub>. En algunos casos, el ePPM<sub>2</sub> puede no incluir un componente PSS<sub>12</sub> (por ejemplo, cuando la secreción del polipéptido expresado no es necesaria o deseable). El segundo módulo promotor procarionta escindible funcionaría de manera similar a la del primer módulo

promotor procariota escindible descrito anteriormente.

El(los) promotor(es) procariotas del(de los) módulo(s) promotor(es) procariota(s) escindible(s) puede(n) ser un promotor *phoA*, *Tac*, *Lac* o *Tphac* (véase, por ejemplo, Kim *et al.*, *PLoS One*. 7(4): e35844), u otro promotor procariota conocido en la técnica.

Un desafío adicional en la construcción de un vector que pueda expresar proteínas de interés tanto en células procariotas (por ejemplo, células de *E. coli*) como células eucariotas (células de mamífero, por ejemplo, células Expi293F) surge de las diferencias en las secuencias señal encontradas en estos tipos de células. Si bien determinadas características de las secuencias señal se conservan en general tanto en células procariotas como en eucariotas (por ejemplo, un parche de residuos hidrófobos localizados en el medio de la secuencia y residuos polares/cargados adyacentes al sitio de escisión en el extremo N del polipéptido maduro), otras son más características de un tipo de célula que las otras. Además, es conocido en la técnica que diferentes secuencias señal pueden tener un impacto significativo en los niveles de expresión en células de mamífero, incluso si las secuencias son todas de origen mamífero (Hall *et al.*, *J of Biological Chemistry*, 265:19996-19999 (1990); Humphreys *et al.*, *Protein Expression and Purification*, 20: 252-264 (2000)). Por ejemplo, las secuencias señal bacterianas típicamente tienen residuos cargados positivamente (más comúnmente lisina) directamente después de la metionina de inicio, mientras que estos no siempre están presentes en las secuencias señal de mamífero.

Se puede usar cualquier secuencia señal (incluidas las secuencias señal consenso) que se dirija al polipéptido de interés para el periplasma en procariotas y para el retículo endoplásmico en eucariotas, si se necesita o desea la secreción de la proteína expresada. Por ejemplo, la secuencia señal eucariota (por ejemplo, ESS1<sub>1</sub> o ESS1<sub>2</sub>) se puede derivar de o incluir toda o una parte de la secuencia señal de la proteína de unión a inmunoglobulina murina (mBiP) (UniProtKB: n.º acceso P20029) o una secuencia señal de cadena pesada o ligera de anticuerpo (por ejemplo, una secuencia señal del gen VH murino). En algunos modos de realización, la secuencia señal procariota (por ejemplo, PSS1<sub>1</sub> o PSS1<sub>2</sub>) se puede derivar de o incluir todo o una parte del gen de enterotoxina II (stII) termoestable. Otras secuencias señal que se pueden utilizar incluyen secuencias señal de la hormona de crecimiento humana (hGH) (UniProtKB: n.º acceso BIA4G6), luciferasa de *Gaussia princeps* (UniProtKB: n.º acceso Q9BLZ2) y endo-1,3-glucanasa de levadura (yBGL2) (UniProtKB: n.º acceso P15703). La secuencia señal puede ser una secuencia señal natural o sintética. En algunos modos de realización, la secuencia señal sintética es una secuencia señal de secreción optimizada que conduce a niveles de presentación a un nivel optimizado en comparación con su secuencia señal natural no optimizada.

## 2. Vectores, células huésped y procedimientos de producción.

La invención también presenta un conjunto de vectores que incluye un primer vector y un segundo vector, en el que el primer y el segundo vector incluyen la primera y la segunda molécula de ácido nucleico, respectivamente, de un sistema de expresión de polipéptidos descrito anteriormente.

Además de los componentes de las moléculas de ácido nucleico descritas en detalle anteriormente, los vectores o conjuntos de vectores pueden incluir un origen de replicación bacteriano, un origen de replicación de mamífero y/o un ácido nucleico que codifica polipéptidos útiles como control (por ejemplo, proteína *gD*) o útiles para actividades (por ejemplo, purificación de proteínas, marcado de proteínas o etiquetado de proteínas).

También se proporcionan procedimientos para producir un polipéptido que comprende cultivar una célula huésped que comprende el(los) conjunto(s) de vectores anteriores en un medio de cultivo y, opcionalmente, recuperar el anticuerpo de la célula huésped (o medio de cultivo de la célula huésped).

## C. Sistema de vectores de presentación en fagos para expresión modular y reformateo de anticuerpos

En algunos modos de realización se pueden producir anticuerpos (por ejemplo, anticuerpos de longitud completa, por ejemplo, anticuerpos IgG de longitud completa, o fragmentos de los mismos, por ejemplo, fragmentos Fab) usando un sistema de expresión de polipéptidos de la invención. Demostramos la aplicación de sistemas de expresión modulares de proteínas mediante el diseño de un sistema de vectores de presentación en fagos que permite la expresión de diferentes formatos de anticuerpos en células humanas del mismo clon. La región de unión a antígeno de cadena pesada y parte de la región constante codificada por el vector de presentación en fagos se fusionaron de forma directa y precisa a las secuencias codificadas en una segunda construcción complementaria, uniando las secuencias que codifican diferentes partes del polipéptido mediante el empalme *trans* de pre-ARNm durante la expresión en células.

El uso del sistema de expresión de polipéptidos con el propósito de permitir la expresión directa de IgG en células de mamífero sin necesidad de subclonar las secuencias Fab de fagos se describe en los ejemplos 1 y 2 a continuación. En algunos casos, la primera molécula de ácido nucleico del sistema de expresión de polipéptidos se puede diseñar para codificar la totalidad de los componentes del fragmento Fab. En consecuencia, la primera molécula de ácido nucleico puede incluir un componente PES1<sub>1</sub> que codifica un polipéptido que tiene un dominio VH y un dominio CH1 del Fab. La primera molécula de ácido nucleico también puede incluir un componente PES1<sub>2</sub> que codifica un dominio VL y un dominio CL. La transcripción de la primera molécula de ácido nucleico daría como resultado dos productos de

pre-ARNm no contiguos, que juntos forman un fragmento Fab que se puede marcar adecuadamente (por ejemplo, fusionado a pIII de M13) para propósitos de presentación en fagos.

El procedimiento de reformatear el fragmento Fab en un anticuerpo IgG de longitud completa se puede lograr posteriormente mediante la expresión de la primera molécula de ácido nucleico en una célula eucariota (por ejemplo, una célula de mamífero, por ejemplo, una célula Expi293F), junto con una segunda molécula de ácido nucleico que proporciona la parte restante del anticuerpo (es decir, los dominios CH2 y CH3). Por ejemplo, la segunda molécula de ácido nucleico puede incluir un componente PES2 que codifica un polipéptido que tiene un dominio CH2 y un dominio CH3. La transcripción de la primera y la segunda molécula en la célula eucariota daría como resultado la generación de tres transcritos de pre-ARNm, induciendo a la cadena pesada que codifica los transcritos de pre-ARNm para que se sometan a empalme *trans* entre sí para generar la cadena pesada de longitud completa reformateada del anticuerpo IgG deseado. A continuación, los ARNm procesados se traducirían y darían como resultado la producción de cadenas ligeras y pesadas de la molécula de IgG, y dicha generación no requeriría la necesidad de una laboriosa subclonación.

La capacidad de expresar diferentes formatos de anticuerpos a partir del mismo clon es útil en el descubrimiento de anticuerpos cuando se requieren diferentes formatos de anticuerpos, tales como IgG natural, fragmentos Fab o IgG con modificaciones en Fc para formatos biespecíficos, para diferentes ensayos de cribado. El sistema de expresión de polipéptidos de la invención permite, en principio, cualquiera de estos formatos o formatos adicionales simplemente clonando una secuencia adecuada para añadirla después de la región CH1 en el plásmido complementario. Además, la organización modular del sistema permite la expresión de nuevos formatos de anticuerpos sin necesidad de volver a crear reservas de colecciones de presentación en fagos, ya que esto solo requiere la construcción de un plásmido complementario novedoso. Los ácidos nucleicos también se podrían adaptar para permitir el uso de cualquier región CH1 mediante desplazamiento del sitio de empalme 5'ss corriente abajo de la región que codifica CH1 hasta la región J (FR4) en el dominio VH o en el punto de unión J-CH1, separando por tanto el dominio VH y toda la región constante de la cadena pesada en dos ácidos nucleicos diferentes. Las moléculas de ácido nucleico son compatibles con los procedimientos tradicionales para la expresión de fragmentos Fab en *E. coli*, simplemente agregando un codón de parada después de la secuencia que codifica la bisagra superior. Sin embargo, los codones de parada ámbar en el punto de unión de las secuencias de cadena pesada y del gen III en las colecciones de presentación en fagos de Fab normalmente dan como resultado niveles significativamente menores de presentación, lo que requiere por tanto reformatear clones después de la selección, al menos en el caso de las colecciones de repertorios indiferenciados (Lee *et al.*, *Journal of immunological methods*. 284: 119-132, 2004). La expresión de fragmentos Fab en células de mamífero usando los mismos procedimientos usados para la expresión de IgG evita esta necesidad de reformateo, con rendimientos comparables a los obtenidos normalmente en *E. coli*.

Los anticuerpos producidos por este sistema de expresión de polipéptidos pueden incluir anticuerpos quiméricos, humanizados y/o humanos generados recombinantemente. En algunos casos, los anticuerpos son fragmentos de anticuerpo, por ejemplo, fragmentos Fab, Fv, Fab', scFv, diacuerpos o F(ab')<sub>2</sub>. En otros casos, los anticuerpos son anticuerpos de longitud completa, por ejemplo, anticuerpos IgG1, IgG2, IgG3 o IgG4 intactos u otros anticuerpos de otra clase o isotipo, como se define en el presente documento.

Los anticuerpos expresados pueden incorporar cualquiera de las características, individualmente o en combinación, como se describe en las secciones 1-7 a continuación:

#### 1. Afinidad de anticuerpos

El anticuerpo (por ejemplo, Fab o anticuerpo IgG de longitud completa) proporcionado por un sistema de expresión de polipéptidos descrito en el presente documento puede tener una constante de disociación (K<sub>d</sub>) de  $\leq 1 \mu\text{M}$ ,  $\leq 100 \text{ nM}$ ,  $\leq 10 \text{ nM}$ ,  $\leq 1 \text{ nM}$ ,  $\leq 0,1 \text{ nM}$ ,  $\leq 0,01 \text{ nM}$  o  $\leq 0,001 \text{ nM}$  (por ejemplo,  $10^{-8} \text{ M}$  o menos, por ejemplo, de  $10^{-8} \text{ M}$  a  $10^{-13} \text{ M}$ , por ejemplo, de  $10^{-9} \text{ M}$  a  $10^{-13} \text{ M}$ ).

En un modo de realización, se mide la K<sub>d</sub> mediante un ensayo de unión a antígeno radiomarcado (RIA) realizado con la versión Fab de un anticuerpo de interés y su antígeno como se describe mediante el siguiente ensayo. La afinidad de unión en solución de los Fab por el antígeno se mide equilibrando los Fab con una concentración mínima de antígeno marcado con (<sup>125</sup>I) en presencia de una serie de valoraciones de antígeno no marcado, capturando a continuación el antígeno unido con una placa recubierta con anticuerpo anti-Fab (véase, por ejemplo, Chen *et al.*, *J. Mol. Biol.* 293:865-881(1999)). Para establecer las condiciones para el ensayo, se recubren placas de múltiples pocillos MICROTITER® (Thermo Scientific) durante la noche con 5 µg/ml de un anticuerpo anti-Fab de captura (Cappel Labs) en carbonato de sodio 50 mM (pH 9,6) y posteriormente se bloquea con seroalbúmina bovina al 2 % (p/v) en PBS durante de dos a cinco horas a temperatura ambiente (aproximadamente 23 °C). En una placa no adsorbente (Nunc n.º 269620), se mezcla [<sup>125</sup>I]-antígeno 100 pM o 26 pM con diluciones en serie de un Fab de interés (por ejemplo, consecuente con la evaluación del anticuerpo anti-VEGF, Fab-12, en Presta *et al.*, *Cancer Res.* 57:4593-4599 (1997)). A continuación, el Fab de interés se incuba durante la noche; sin embargo, la incubación puede continuar durante un período más largo (por ejemplo, de aproximadamente 65 horas) para garantizar que se alcanza el equilibrio. Después de esto, se transfieren las mezclas a la placa de captura para su incubación a temperatura ambiente (por ejemplo, durante una hora). A continuación, se retira la solución y la placa se lava ocho veces con polisorbato 20 (TWEEN-20®) al 0,1 % en PBS. Cuando las placas se hayan secado, se añaden 150 µl/pocillo de centelleador (MICROSCINT-20™,

Packard) y las placas se cuentan en un contador gamma TOPCOUNT™ (Packard) durante diez minutos. Se eligen concentraciones de cada Fab que dan menos de o igual a un 20 % de unión máxima para su uso en ensayos de unión competitiva.

5 De acuerdo con otro modo de realización, la  $K_d$  se mide usando ensayos de resonancia de plasmón superficial usando un dispositivo BIACORE®-2000 o uno BIACORE®-3000 (BIAcore, Inc., Piscataway, NJ) a 25 °C con matrices CM5 de antígeno inmovilizado a ~10 unidades de respuesta (UR). En resumen, se activan matrices de biosensor de dextrano carboximetilado (CM5, BIACORE, Inc.) con clorhidrato de *N*-etil-*N'*-(3-dimetilaminopropil)carbodiimida (EDC) y *N*-hidroxisuccinimida (NHS) de acuerdo con las instrucciones del proveedor. El antígeno se diluye con acetato de sodio  
10 10 mM, pH 4,8, hasta 5 µg/ml (~0,2 µM) antes de la inyección a un caudal de 5 µl/minuto para lograr aproximadamente 10 unidades de respuesta (UR) de proteína acoplada. Tras la inyección del antígeno, se inyecta etanolamina 1 M para bloquear los grupos sin reaccionar. Para las mediciones cinéticas, se inyectan diluciones en serie al 50 % de Fab (de 0,78 nM a 500 nM) en PBS con tensioactivo polisorbato 20 (TWEEN-20™) (PBST) al 0,05 % a 25 °C a un caudal de aproximadamente 25 µl/min. Se calculan las velocidades de asociación ( $k_{aso}$ ) y velocidades de disociación ( $k_{dis}$ ) usando un modelo de unión de Langmuir uno a uno sencillo (programa informático de evaluación de BIACORE®, versión 3.2) ajustando simultáneamente los sensogramas de asociación y disociación. Se calcula la constante de disociación en equilibrio ( $K_d$ ) como la proporción  $k_{dis}/k_{aso}$ . Véase, por ejemplo, Chen *et al.*, *J. Mol. Biol.* 293:865-881 (1999). Si la velocidad de asociación supera  $10^6 M^{-1} s^{-1}$  mediante el ensayo de resonancia de plasmón superficial anterior, entonces la velocidad de asociación se puede determinar usando una técnica de extinción fluorescente que mide el incremento o disminución de la intensidad de emisión de fluorescencia (excitación = 295 nm; emisión = 340 nm, paso de banda 16 nm) a 25 °C de un anticuerpo anti-antígeno (forma Fab) 20 nM en PBS, pH 7,2, en presencia de concentraciones crecientes de antígeno como se mide en un espectrómetro, tal como un espectrofotómetro equipado con interrupción de flujo (Aviv Instruments) o un espectrofotómetro SLM-AMINCO™ serie 8000 (ThermoSpectronic) con una cubeta agitada.

## 25 2. Fragmentos de anticuerpo

En determinados modos de realización, el anticuerpo producido por un sistema de expresión de polipéptidos descrito en el presente documento es un fragmento de anticuerpo. Los fragmentos de anticuerpo incluyen, pero no se limitan a: fragmentos Fab, Fab', Fab'-SH, F(ab')<sub>2</sub>, Fv y scFv, y otros fragmentos que se describen a continuación. Para una revisión de determinados fragmentos de anticuerpo, véase Hudson *et al.*, *Nat. Med.* 9:129-134 (2003). Para una revisión de fragmentos scFv, véase, por ejemplo, Pluckthün, en *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenberg y Moore eds., (Springer-Verlag, Nueva York), pp. 269-315 (1994); véase también el documento WO 93/16185; y las patentes de EE. UU. n.º 5.571.894 y 5.587.458. Para un análisis de los fragmentos Fab y F(ab')<sub>2</sub> que comprenden residuos de epítipo de unión al receptor de rescate y que tienen una semivida *in vivo* incrementada, véase la patente de EE. UU. n.º 5.869.046.

Los diacuerpos son fragmentos de anticuerpo con dos sitios de unión a antígeno que pueden ser bivalentes o biespecíficos. Véase, por ejemplo, el documento EP 404.097; el documento WO 1993/01161; Hudson *et al.*, *Nat. Med.* 9:129-134 (2003); y Hollinger *et al.*, *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993). También se describen triacuerpos y tetracuerpos en Hudson *et al.*, *Nat. Med.* 9:129-134 (2003).

Los anticuerpos de dominio único son fragmentos de anticuerpo que comprenden todo o una parte del dominio variable de la cadena pesada o todo o una parte del dominio variable de la cadena ligera de un anticuerpo. En determinados modos de realización, un anticuerpo de un único dominio es un anticuerpo de un único dominio humano (Domantis, Inc., Waltham, MA; véase, por ejemplo, la patente de EE. UU. n.º 6.248.516 B1).

## 45 3. Anticuerpos quiméricos y humanizados

50 En determinados modos de realización, el anticuerpo (por ejemplo, Fab o anticuerpo IgG de longitud completa) producido por un sistema de expresión de polipéptidos descrito en el presente documento es un anticuerpo quimérico. Determinados anticuerpos quiméricos se describen, por ejemplo, en la patente de EE. UU. n.º 4.816.567; y Morrison *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:6851-6855 (1984). En un ejemplo, un anticuerpo quimérico comprende una región variable no humana (por ejemplo, una región variable derivada de un ratón, rata, hámster, conejo o primate no humano, tal como un mono) y una región constante humana. En otro ejemplo, un anticuerpo quimérico es un anticuerpo de "clase cambiada" en el que se ha cambiado la clase o subclase respecto a la del anticuerpo original. Los anticuerpos quiméricos incluyen fragmentos de unión a antígeno de los mismos.

60 En determinados modos de realización, un anticuerpo quimérico es un anticuerpo humanizado. Típicamente, se humaniza un anticuerpo no humano para reducir la inmunogenicidad en los seres humanos, mientras que se retienen la especificidad y afinidad del anticuerpo no humano original. En general, un anticuerpo humanizado comprende uno o más dominios variables en los que las HVR, por ejemplo, las CDR (o partes de las mismas), se derivan de un anticuerpo no humano y las FR (o partes de las mismas) se derivan de secuencias de anticuerpos humanos. Un anticuerpo humanizado también comprenderá opcionalmente al menos una parte de una región constante humana.  
65 En algunos modos de realización se sustituyen algunos residuos de FR de un anticuerpo humanizado por los residuos correspondientes de un anticuerpo no humano (por ejemplo, el anticuerpo del que se derivan los residuos de HVR),

por ejemplo, para restablecer o mejorar la especificidad o afinidad del anticuerpo.

Los anticuerpos humanizados y los procedimientos de preparación de los mismos se revisan, por ejemplo, en Almagro y Fransson, *Front. Biosci.* 13:1619-1633 (2008), y se describen adicionalmente, por ejemplo, en Riechmann *et al.*, *Nature* 332:323-329 (1988); Queen *et al.*, *Proc. Nat'l Acad. Sci. USA* 86:10029-10033 (1989); las patentes de EE. UU. n.º 5.821.337, 7.527.791, 6.982.321 y 7.087.409; Kashmiri *et al.*, *Methods* 36:25-34 (2005) (que describe el injerto en la SDR (a-CDR)); Padlan, *Mol. Immunol.* 28:489-498 (1991) (que describe el "rebarnizado"); Dall'Acqua *et al.*, *Methods* 36:43-60 (2005) (que describen la "reordenamiento de FR"); y Osbourn *et al.*, *Methods* 36:61-68 (2005) y Klimka *et al.*, *Br. J. Cancer*, 83:252-260 (2000) (que describe el enfoque de "selección guiada" para la reordenamiento de FR).

Las regiones estructurales humanas que se pueden usar para la humanización incluyen, pero no se limitan a: regiones estructurales seleccionadas usando el procedimiento de "mejor ajuste" (véase, por ejemplo, Sims *et al.*, *J. Immunol.*, 151:2296 (1993)); regiones estructurales derivadas de la secuencia consenso de anticuerpos humanos de un subgrupo particular de regiones variables de la cadena ligera o pesada (véase, por ejemplo, Carter *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:4285 (1992); y Presta *et al.*, *J. Immunol.*, 151:2623 (1993)); regiones estructurales maduras (mutadas somáticamente) humanas o regiones estructurales de la línea germinal humana (véase, por ejemplo, Almagro y Fransson, *Front. Biosci.*, 13:1619-1633 (2008)); y regiones estructurales derivadas del cribado de colecciones de FR (véase, por ejemplo, Baca *et al.*, *J. Biol. Chem.* 272:10678-10684 (1997) y Rosok *et al.*, *J. Biol. Chem.* 271:22611-22618 (1996)).

#### 4. Anticuerpos humanos

En determinados modos de realización, el anticuerpo (por ejemplo, Fab o anticuerpo IgG de longitud completa) producido por un sistema de expresión de polipéptidos descrito en el presente documento es un anticuerpo humano. El anticuerpo humano puede ser un anticuerpo humano recombinante que se preparó originalmente y cuya secuencia se identificó posteriormente usando diversas técnicas conocidas en la técnica. Se describen, en general, anticuerpos humanos en van Dijk y van de Winkel, *Curr. Opin. Pharmacol.* 5: 368-74 (2001) y Lonberg, *Curr. Opin. Immunol.* 20:450-459 (2008).

#### 5. Anticuerpos derivados de colecciones

En virtud de la utilidad del sistema de expresión de polipéptidos descrito en el presente documento que es útil en sistemas de presentación en fagos, los anticuerpos (por ejemplo, Fab o anticuerpos IgG de longitud completa) producidos por un sistema de expresión de polipéptidos de la invención se pueden haber aislado cribando colecciones combinatorias para anticuerpos con la actividad o actividades deseadas. Véase, por ejemplo, Hoogenboom *et al.*, en *Methods in Molecular Biology* 178:1-37 (O'Brien *et al.*, ed., Human Press, Totowa, NJ, 2001) y, adicionalmente, por ejemplo, en McCafferty *et al.*, *Nature* 348:552-554; Clackson *et al.*, *Nature* 352:624-628 (1991); Marks *et al.*, *J. Mol. Biol.* 222:581-597 (1992); Marks y Bradbury, en *Methods in Molecular Biology* 248:161-175 (Lo, ed., Human Press, Totowa, NJ, 2003); Sidhu *et al.*, *J. Mol. Biol.* 338(2): 299-310 (2004); Lee *et al.*, *J. Mol. Biol.* 340(5): 1073-1093 (2004); Fellouse, *Proc. Natl. Acad. Sci. USA* 101(34): 12467-12472 (2004); y Lee *et al.*, *J. Immunol. Methods* 284(1-2): 119-132(2004).

#### 6. Anticuerpos multiespecíficos

En determinados modos de realización, el anticuerpo (por ejemplo, Fab o anticuerpo IgG de longitud completa) producido por un sistema de expresión de polipéptidos descrito en el presente documento es un anticuerpo multiespecífico, por ejemplo, un anticuerpo biespecífico. Los anticuerpos multiespecíficos son anticuerpos monoclonales que tienen especificidades de unión por al menos dos sitios diferentes. En determinados modos de realización, una de las especificidades de unión es para un primer antígeno y la otra es para cualquier otro antígeno. En determinados modos de realización se pueden unir anticuerpos biespecíficos a dos epítomos diferentes del primer antígeno. También se pueden usar anticuerpos biespecíficos para localizar agentes citotóxicos para las células que expresan el primer antígeno. Se pueden preparar anticuerpos biespecíficos como anticuerpos de longitud completa o fragmentos de anticuerpo.

En el presente documento también se incluyen anticuerpos genomanipulados con tres o más sitios de unión a antígeno funcionales, incluyendo los "anticuerpos pulpo" (véase, por ejemplo, el documento US 2006/0025576A1).

El anticuerpo o fragmento en el presente documento también incluye un "Fab de doble acción" o "DAF" que comprende un sitio de unión a antígeno que se une a un primer antígeno, así como a otro antígeno diferente (véase, el documento US 2008/0069820, por ejemplo).

#### 7. Variantes de anticuerpo

En determinados modos de realización se contemplan variantes de secuencia de aminoácidos de los anticuerpos proporcionados en el presente documento. Por ejemplo, puede ser deseable mejorar la afinidad de unión y/u otras propiedades biológicas del anticuerpo. Las variantes de secuencia de aminoácidos de un anticuerpo se pueden

preparar mediante la introducción de modificaciones apropiadas en una o más de las moléculas de ácido nucleico que codifican todo o una parte del anticuerpo. Dichas modificaciones incluyen, por ejemplo, deleciones de y/o inserciones en y/o sustituciones de residuos dentro de las secuencias de aminoácidos del anticuerpo. Se puede realizar cualquier combinación de deleción, inserción y sustitución para llegar a la construcción final, siempre que la construcción final posea las características deseadas, por ejemplo, unión a antígeno.

En determinados modos de realización, los sistemas de expresión y procedimientos de la invención pueden producir una colección de variantes de anticuerpos que tienen una o más sustituciones aminoacídicas entre sí. Los sitios de interés para la mutagénesis de sustitución incluyen las HVR y las FR. Las sustituciones conservadoras se muestran en la tabla 1 bajo el encabezamiento "sustituciones conservadoras". Se proporcionan cambios más sustanciales en la tabla 1 bajo el encabezamiento "sustituciones ejemplares" y como se describe además a continuación en referencia a las clases de cadenas laterales de aminoácidos. Se pueden introducir sustituciones de aminoácidos en un anticuerpo de interés y cribar los productos para determinar una actividad deseada, por ejemplo, unión a antígeno retenida/mejorada, inmunogenicidad disminuida, o ADCC o CDC mejorada.

**Tabla 1. Sustituciones aminoacídicas ejemplares y preferentes**

Residuo original	Sustituciones ejemplares	Sustituciones preferentes
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Asp, Lys; Arg	Gln
Asp (D)	Glu; Asn	Glu
Cys (C)	Ser; Ala	Ser
Gln (Q)	Asn; Glu	Asn
Glu (E)	Asp; Gln	Asp
Gly (G)	Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe; norleucina	Leu
Leu (L)	Norleucina; Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Trp; Leu; Val; Ile; Ala; Tyr	Tyr
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Val; Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala; norleucina	Leu

Los aminoácidos se pueden agrupar de acuerdo con propiedades de cadena lateral comunes:

- (1) hidrófobos: norleucina, Met, Ala, Val, Leu, Ile;
- (2) hidrófilos neutros: Cys, Ser, Thr, Asn, Gln;
- (3) ácidos: Asp, Glu;
- (4) básicos: His, Lys, Arg;
- (5) residuos que influyen en la orientación de la cadena: Gly, Pro;
- (6) aromáticos: Trp, Tyr, Phe.

Las sustituciones no conservadoras supondrán intercambiar un miembro de una de estas clases por otra clase.

Un tipo de variante de sustitución implica sustituir uno o más residuos de la región hipervariable de un anticuerpo original (por ejemplo, un anticuerpo humanizado o humano). En general, la(s) variante(s) resultante(s) seleccionada(s) para su estudio adicional tendrá(n) modificaciones (por ejemplo, mejoras) en determinadas propiedades biológicas (por ejemplo, afinidad incrementada, inmunogenicidad reducida) en relación con el anticuerpo original y/o habrá(n)

retenido sustancialmente determinadas propiedades biológicas del anticuerpo original. Una variante de sustitución ejemplar es un anticuerpo madurado en afinidad, que se puede generar convenientemente, por ejemplo, usando técnicas de maduración de la afinidad basadas en presentación en fagos, tales como las descritas en el presente documento. En resumen, se mutan uno o más residuos de HVR y los anticuerpos variantes se presentan en fagos y se criban para una actividad biológica particular (por ejemplo, afinidad de unión).

Se pueden realizar alteraciones (por ejemplo, sustituciones) en las HVR, por ejemplo, para mejorar la afinidad del anticuerpo. Se pueden realizar dichas alteraciones en "puntos calientes" de la HVR, es decir, residuos codificados por codones que se someten a una mutación con alta frecuencia durante el procedimiento de maduración somática (véase, por ejemplo, Chowdhury, *Methods Mol. Biol.* 207:179-196 (2008)), y/o SDR (a-CDR), sometiendo a prueba el VH o VL variante resultante para determinar su afinidad de unión. La maduración de la afinidad mediante construcción y reelección de colecciones secundarias se ha descrito, por ejemplo, en Hoogenboom *et al.* en *Methods in Molecular Biology* 178:1-37 (O'Brien *et al.*, ed., Human Press, Totowa, NJ, (2001)). En algunos modos de realización de la maduración de la afinidad se introduce diversidad en los genes variables elegidos para la maduración mediante cualquiera de una variedad de procedimientos (por ejemplo, PCR propensa a error, reordenamiento de cadenas o mutagénesis dirigida por oligonucleótidos). A continuación, se crea una colección secundaria. A continuación, se criba la colección para identificar cualquier variante de anticuerpo con la afinidad deseada. Otro procedimiento para introducir diversidad implica enfoques dirigidos a HVR, en los que se aleatorizan varios residuos de HVR (por ejemplo, 4-6 residuos a la vez). Se pueden identificar específicamente los residuos de HVR implicados en la unión a antígeno, por ejemplo, usando modelado o mutagénesis por barrido de alanina. A menudo se seleccionan en particular CDR-H3 y CDR-L3.

En determinados modos de realización se pueden producir sustituciones, inserciones o deleciones dentro de una o más HVR, siempre que dichas alteraciones no reduzcan sustancialmente la capacidad del anticuerpo de unirse al antígeno. Por ejemplo, en las HVR se pueden realizar alteraciones conservadoras (por ejemplo, sustituciones conservadoras como se proporciona en el presente documento) que no reduzcan sustancialmente la afinidad de unión. Dichas alteraciones pueden estar fuera de los "puntos calientes" de la HVR o los SDR. En determinados modos de realización de las secuencias de VH y VL variantes proporcionadas anteriormente, cada HVR está inalterada o bien no contiene más de una, dos o tres sustituciones de aminoácidos.

Un procedimiento útil para la identificación de residuos o regiones de un anticuerpo que se pueden seleccionar para la mutagénesis se llama "mutagénesis por barrido de alanina" como se describe por Cunningham y Wells (1989) *Science*, 244:1081-1085. En este procedimiento se identifican un residuo o grupo de residuos diana (por ejemplo, residuos cargados, tales como arg, asp, his, lys y glu) y se reemplazan por un aminoácido neutro o cargado negativamente (por ejemplo, alanina o polialanina) para determinar si la interacción del anticuerpo con el antígeno se ve afectada. Se pueden introducir otras sustituciones en las localizaciones de aminoácidos que demuestren sensibilidad funcional a las sustituciones iniciales. De forma alternativa, o adicionalmente, una estructura cristalina de un complejo antígeno-anticuerpo permite identificar los puntos de contacto entre el anticuerpo y el antígeno. Dichos residuos de contacto y residuos adyacentes se pueden seleccionar o eliminar como candidatos para la sustitución. Las variantes se pueden cribar para determinar si contienen las propiedades deseadas.

Las inserciones en la secuencia de aminoácidos incluyen fusiones amino y/o carboxiterminales que varían en longitud desde un residuo hasta polipéptidos que contienen cien o más residuos, así como inserciones intrasecuenciales de residuos de aminoácidos únicos o múltiples. Los ejemplos de inserciones terminales incluyen un anticuerpo con un residuo de metionilo N terminal. Otras variantes de inserción de la molécula de anticuerpo incluyen la fusión del extremo N o C del anticuerpo a una enzima (por ejemplo, para ADEPT) o un polipéptido que incrementa la semivida en suero del anticuerpo.

Aunque el concepto de expresión modular de proteínas a través del empalme *trans* de pre-ARNm en el contexto de un sistema de vectores de presentación de anticuerpos en fagos se describe en detalle en el presente documento, la aplicación del concepto, como se ejemplifica mediante el uso de las moléculas de ácido nucleico, vectores, conjuntos de vectores, células huésped y procedimientos descritos en el presente documento se pueden adaptar y ampliar, por ejemplo, a otras tecnologías que requieren la expresión en células de mamíferos de grandes colecciones de proteínas con diferentes combinaciones de módulos recurrentes.

### III. EJEMPLOS

Los siguientes son ejemplos de la invención. Se entiende que se pueden poner en práctica otros modos de realización diversos, dada la descripción general proporcionada anteriormente.

#### **Ejemplo 1. Generación de un sistema modular de expresión de proteínas para el reformato de anticuerpos en el contexto de vectores de presentación en fagos**

Describimos la generación de sistemas de expresión de polipéptidos para la expresión modular y la producción de polipéptidos. La invención se basa, al menos en parte, en hallazgos experimentales que demuestran que el empalme *trans* de pre-ARNm se puede explotar en células de mamífero para permitir la expresión modular de proteínas

recombinantes. El concepto de expresión modular de proteínas permite la unión precisa de dos secuencias de codificación de proteínas arbitrarias codificadas por dos construcciones diferentes en un solo ARNm que codifica una cadena de polipéptido, sin ninguno de los requisitos y limitaciones de otros procedimientos de empalme proteína-proteína. El concepto de expresión modular de proteínas a través del empalme *trans* de pre-ARNm se puede adaptar para simplificar y ampliar otras tecnologías que requieren la expresión en células de mamífero de grandes colecciones de proteínas con diferentes combinaciones de módulos recurrentes. Por ejemplo, este concepto encontrará aplicación en otros entornos que requieran la expresión de combinaciones de parejas de proteínas de fusión o mutaciones en polipéptidos individuales. Esta tecnología es simple y poderosa, permitiendo la aplicación a cualquier escala y tiene una gran importancia para el campo de la expresión de proteínas recombinantes en células de mamíferos, la base de gran parte de la biotecnología moderna.

Aquí describimos la generación de dicho sistema de expresión de polipéptidos que permite la expresión modular de diferentes formatos de anticuerpos en el contexto de un sistema de expresión de presentación en fagos. La presentación en fagos se usa ampliamente en el descubrimiento y la genomanipulación de fragmentos de anticuerpos para el desarrollo de anticuerpos terapéuticos y reactivos (McCafferty *et al.*, *Nature* 348: 552-554, 1990; Sidhu, *Current opinion in biotechnology* 11: 610-616, 2000; Smith, *Science* 228:1315-1317, 1985). La presentación en fagos permite tradicionalmente la selección rápida de aglutinantes específicos de antígeno, pero limita el cribado de los fragmentos de anticuerpo seleccionados. A menudo, la caracterización detallada de los fragmentos de anticuerpo requiere la expresión de inmunoglobulina G (IgG) de longitud completa, normalmente expresada en células de mamífero. Sin embargo, un paso limitante en este procedimiento es el reformateo de los clones de fagos en vectores de expresión de mamíferos para la expresión de IgG. Aunque los procedimientos de subclonación de alto rendimiento se pueden usar para reformatear una gran cantidad de clones, estos procedimientos normalmente son relativamente laboriosos y producen muchos clones que no se utilizarán más allá de la fase de cribado.

Para evitar la necesidad de subclonar y permitir la expresión modular de proteínas, generamos una primera molécula de ácido nucleico: el vector de doble huésped, pDV2 (Fig. 1). A diferencia del vector doble descrito previamente, pDV, que contiene un casete de expresión de IgG con una secuencia señal diseñada para la expresión de la cadena pesada en células bacterianas o de mamífero y que requiere la cotransfección de células de mamífero con un vector de expresión de mamífero que expresa la cadena ligera para la expresión de IgG completa (Tesar *et al.*, *Protein engineering, design & selection: PEDS* 26: 655-662, 2013), pDV2 contiene un promotor bacteriano y la mayor parte de una secuencia señal de stII incluida en un intrón que se elimina mediante empalme en células de mamífero.

La secuencia señal de stII en pDV2 se modificó para incluir un sitio de empalme en 3' (3'ss) y un tramo optimizado de polipirimidina (PPT) antes del 3'ss. Esto requirió la introducción de tres sustituciones aminoácidas relativamente conservadoras en la secuencia señal de stII, que no afectaron a la presentación en fagos de los fragmentos Fab (Fig. 2). Para permitir la expresión modular y flexible de formatos de anticuerpos del mismo clon, no agregamos un intrón y un exón completos que codifican regiones constantes corriente abajo de la región que codifica el dominio CH1. En cambio, buscamos añadir estas secuencias de cadena pesada en *trans* respecto a una segunda molécula de ácido nucleico. Para lograr esto, explotamos el procedimiento de empalme *trans* de pre-ARNm, la unión de dos pre-ARNm diferentes para formar un solo ARNm maduro. El empalme *trans* se puede inducir en células de mamífero trayendo pre-ARNm con sitios de empalme 5'ss y 3'ss solos mediante la hibridación de secuencias complementarias corriente abajo del sitio de empalme 5'ss y corriente arriba del sitio de empalme 3'ss para formar un solo pre-ARNm unido de forma no covalente que a continuación se empalma como un pre-ARNm normal (Konarska *et al.*, *Cell*. 42: 165-171, 1985; Puttaraju *et al.*, *Nature biotechnology*. 17: 246-252, 1999; Solnick, *Cell*. 42: 157-164, 1985). En este sistema de expresión de polipéptidos en particular utilizamos un fragmento de 150 pb del gen III (gIII) de M13 como secuencia de hibridación (Fig. 1). Esta secuencia del gen III sigue a un sitio de empalme 5'ss GTAAGA optimizado descrito previamente en el límite en 3' de la secuencia que codifica CH1 (Tesar *et al.*, *Protein engineering, design & selection: PEDS* 26: 655-662, 2013).

Para completar el sistema de expresión de polipéptidos, generamos una segunda molécula de ácido nucleico, pRK-Fc, que es un plásmido complementario que expresa un pre-ARNm que contiene una secuencia antisentido del gen III de 150 nt seguida de una secuencia conectora, un punto de ramificación consensuado, y un PPT, así como un 3'ss seguido por la región bisagra, las regiones CH2 y CH3 en un exón, y una señal de poliadenilación SV40 (Figs. 1 y 3). Este transcrito no codifica ninguna secuencia señal, y los dos primeros codones de iniciación potenciales se localizan en la secuencia antisentido del gen III y en la región bisagra, con cambio de pauta de lectura. Por tanto, a excepción del sitio 5'ss, el resto de las secuencias necesarias para el empalme están codificadas por pRK-Fc, en lugar de pDV2. La cotransfección de células Expi293F (Invitrogen) con pDV2 y pRK-Fc dio como resultado un nivel de expresión de IgG basal pero detectable (Fig. 4A).

## **Ejemplo 2. Generación de un sistema modular de expresión de proteínas optimizado para el reformateo de anticuerpos en el contexto de vectores de presentación en fagos**

Los rendimientos basales de IgG logrados por pDV2 y pRK-Fc se podrían deber a la ausencia de secuencias requeridas para un empalme *trans* eficaz o de secuencias en el vector que inhiban el empalme *trans*. Los motivos de nucleótidos en ambos exones e intrones pueden actuar como potenciadores o supresores de empalme o ambos, dependiendo de su localización. Para los propósitos del diseño de vectores, los potenciadores intrónicos de empalme

(ISE) se pueden añadir más fácilmente, ya que estos probablemente no afectarían a las secuencias de codificación en la expresión de células de mamífero. Un ISE bien descrito está compuesto por una secuencia de 3 o más residuos de guanina consecutivos, o una secuencia rica en G (G-run), localizada cerca de los límites del intrón, que están unidas por ribonucleoproteínas nucleares heterogéneas H o F para potenciar el empalme (Wang *et al.*, *Nature structural & molecular biology* 19: 1044-1052, 2012; Xiao *et al.*, *Nature structural & molecular biology* 16: 1094-1100, 2009). Además, también se ha demostrado que las secuencias de intrón ricas en purina cercanas al sitio 5'ss no limitadas por secuencias ricas en G (G-run) potencian el empalme (Hastings *et al.*, *RNA*. 7: 859-874, 2001).

Por tanto, creamos una variante de pDV2, pDV2b, que incluye una región rica en purina de 23 pares de bases (pb) situada 26 pb corriente abajo del sitio 5'ss, que tiene una secuencia rica en G (G-run) de 9 nt en la región que codifica el conector entre la bisagra superior y la parte C terminal de la proteína de cubierta pIII del bacteriófago M13 (cP3), así como una segunda secuencia rica en G (G-run) de 4 nt situada 10 nt corriente abajo (Fig. 5). Esta variante cambia el conector Gly-Arg-Pro entre la bisagra superior y la proteína cP3 por tres residuos de Gly. El vector no incluía un sitio de poliadenilación estándar para el casete de cadena pesada. El motivo de esto fue intentar minimizar la formación de ARNm maduros de cadena pesada a partir del vector, que luego se exportaran al citoplasma antes de que se pudiera producir el empalme *trans* y potencialmente dieran lugar a la expresión de la proteína de fusión Fab-cP3. La molécula pRK-Fc también se optimizó. También se ha demostrado que las secuencias ricas en G (G-run) intrónicas cercanas al sitio 3'ss estimulan el empalme *in vitro* (Martinez-Contreras, *PLoS biology*, 4: e21, 2006). Por lo tanto, también se agregó un ISE de 9 nt corriente arriba del sitio de ramificación para generar el plásmido complementario optimizado pRK-Fc2 (Fig. 6). La cotransfección de células Expi293F humanas con pDV2 y pRK-Fc (ISE-) o pRK-Fc2 (ISE+) dio como resultado un nivel basal de expresión de IgG (Fig. 4A). La cotransfección de células Expi293F con el plásmido ISE+ pDV2b y pRK-Fc o pRK-Fc2 dio como resultado niveles mayores de expresión de IgG, con los mayores niveles de expresión, hasta 25 µg/ml, proporcionados al cotransfectar los plásmidos ISE+ pDV2b y pRK-Fc2, lo que indica que las secuencias de ISE en ambos transcritos potencian la eficacia del empalme *trans*.

Los niveles de expresión basales de IgG en células Expi293F transfectadas se asociaron con lisis celular aparente 7 días después de la transfección, que también se observó cuando pDV2 o pDV2b pero no pRK-Fc o pRK-Fc2 se transfectaron solos. El análisis de los lisados celulares transfectados mediante inmunoelectrotransferencia con un anticuerpo anti-p3 de M13 reveló un polipéptido con un peso molecular aparente de aproximadamente 41 kDa (Fig. 12, panel inferior, carriles 3 a 6), consecuente con la expresión de un fragmento Fd de IgG1 (VH-CH1-bisagra superior) fusionado al péptido cP3 de M13. La expresión de este polipéptido fue mayor en las células transfectadas con pDV2 o pDV2b sin un plásmido complementario. Los resultados indicaron que los plásmidos pDV2 y pDV2b eran capaces de expresar un ARNm maduro que codifica un producto potencialmente tóxico a pesar del hecho de que ambos carecen de un sitio de poliadenilación de mamífero en el vector situado corriente abajo del casete de la cadena pesada.

La inspección visual de la secuencia del gen III que codifica cP3 reveló un motivo AATAAA que posiblemente podría actuar como un sitio de poliadenilación (Fig. 2). Introdujimos dos mutaciones silenciosas en este sitio para generar los plásmidos pDV2c (ISE-) y pDV2d (ISE+) para someter a prueba si esto reduciría la toxicidad y mejoraría la expresión de proteínas en células de mamíferos. La cotransfección de células Expi293F con pRK-Fc y pDV2c o pDV2d dio como resultado niveles de expresión de IgG aproximadamente 6 veces mayores en relación con los vectores pDV2 y pDV2b con el sitio potencial de poliadenilación del gen III (Fig. 4A). Este incremento en los niveles de expresión de IgG se asoció con una alta viabilidad de las células transfectadas y con una expresión significativamente reducida o indetectable de la proteína de fusión Fd-cP3 en las células transfectadas (Fig. 12, panel inferior, carriles 7 a 10). Esto indica que la presencia de un sitio potencial de poliadenilación en el vector donante dentro del gen III da lugar a la expresión de proteínas no deseadas del plásmido donante solo que tiene un impacto negativo significativo sobre la expresión de proteínas. La cotransfección de células Expi293F con pDV2c o pDV2d y el vector complementario ISE+ pRK-Fc2 dio como resultado un incremento adicional de 2 veces en la expresión de IgG en comparación con las cotransfecciones con el vector ISE- pRK-Fc (Fig. 4A). Estos resultados indican que el factor principal determinante de la expresión basal de la proteína en el vector pDV2 fue la presencia del sitio potencial de poliadenilación en el gen III, mientras que la adición de un ISE tiene un efecto menor en la expresión de proteínas cuando el motivo potencial de poliadenilación del gen III está ausente. Por el contrario, la adición de un ISE en el plásmido pRK-Fc2 complementario da como resultado un rendimiento de IgG aproximadamente 2 veces mayor cuando se cotransfecta con una variante de pDV2 sin un sitio potencial de poliadenilación en el gen III (Fig. 4A).

Se logró una optimización adicional de la expresión de proteínas determinando las proporciones óptimas de ADN para la transfección. El uso de un exceso 2:1 del plásmido complementario pRK-Fc2 en relación con pDV2d dio como resultado los mayores rendimientos de expresión de IgG en este sistema (Fig. 4B). Usando pDV2d y pRK-Fc2 con las proporciones de ADN optimizadas, el rendimiento de IgG purificada a partir de 30 ml de sobrenadantes de células Expi293F transfectadas fue de  $3,2 \pm 1,2$  mg ( $n = 3$ ). La IgG purificada de las células Expi293F cotransfectadas con estos plásmidos era indistinguible por espectrometría de masas y SDS-PAGE de la misma IgG expresada por vectores de expresión convencionales (Figs. 7A -7B y Fig. 13). La cotransfección de células Expi293F con pDV2d que codifica regiones variables de diferentes especificidades y pRK-Fc2 con las proporciones de ADN optimizadas dio como resultado una alta expresión de IgG de entre 2,5 y 5,5 mg de IgG purificada a partir de 30 ml de sobrenadantes de células Expi293F transfectadas (Fig. 8A). El sistema de expresión de polipéptidos no se limita al uso de células Expi293F para lograr altos niveles de expresión. Otras líneas celulares de mamíferos ampliamente utilizadas para la expresión de IgG, tales como las células 293T y CHO, también fueron eficaces. La cotransfección de células 293T o

CHO con pDV2d que expresa regiones variables de diferentes especificidades y pRK-Fc2 dio como resultado una alta expresión de IgG (Fig. 8B).

El vector pRK-Fc2 se modificó para la expresión de fragmentos Fab cuando se realizó la cotransfección con los plásmidos pDV2. Las secuencias que codifican la región de bisagra inferior y las regiones Fc en pRK-Fc2 se eliminaron y se reemplazaron por una marca Flag para proporcionar el vector pRK-Fab-Flag (Fig. 9). El rendimiento de los fragmentos Fab purificados a partir de 30 ml de sobrenadantes de células Expi293F cotransfectadas con pDV2d y pRK-Fab-Flag fue de  $0,8 \pm 0,06$  mg (media  $\pm$  desviación estándar,  $n = 3$ ). La correcta estructura del fragmento Fab marcado con Flag purificado se confirmó por espectrometría de masas y SDS-PAGE (Fig. 13). La masa de la cadena pesada observada fue de 25.169 Da, cercana a la masa esperada de 25.172 Da al excluir la lisina C terminal recortada.

La expresión de proteínas truncadas en el extremo N del transcrito complementario se ha observado en sistemas de empalme *trans* para terapia génica (Monjaret *et al.*, *Molecular therapy* 22: 1176-1187, 2014). Esto se debe a que el transcrito complementario que codifica el exón 3' tiene todos los elementos necesarios para la formación de un ARNm maduro, que podría dar lugar a la traducción a partir de codones de iniciación internos. Mediante inmunoelectrotransferencia de lisados de células transfectadas con pRK-Fc2 observamos la expresión de un polipéptido consecuente con un fragmento Fc traducido a partir del primer codón ATG sin cambio de pauta de lectura (Fig. 12, panel superior, carril 11). Este polipéptido presumiblemente carece de una secuencia señal de secreción y se debe expresar solo en el citoplasma. Aunque este producto se podría liberar en los medios de cultivo mediante lisis celular, no lo observamos en muestras de IgG purificadas por SDS-PAGE (Fig. 13, carril 2) y espectrometría de masas. La expresión de este producto truncado se redujo pero no se eliminó cuando se cotransfectaron pDV2c o pDV2d en células (Fig. 12, panel superior, carriles 8 y 10). La inserción de un marco de lectura abierto con cambio de pauta de lectura con un sitio óptimo de inicio de la traducción en la región intrónica situada corriente arriba de los codones de inicio de Fc potenciales no redujo significativamente la expresión del producto Fc truncado.

Una propiedad importante de los vectores de presentación en fagos que determina la eficacia de selección es el nivel de presentación de fragmentos de anticuerpos en las partículas de fagos que se logra. Usando el fago auxiliar Amber-2614 KO7 descrito previamente con expresión reducida de p3 en cepas supresoras de *E. coli* SupE, los niveles de presentación de fragmentos Fab logrados con el vector pDV2d fueron comparables a los niveles de presentación de Fab logrados con un vector de presentación de Fab especializado, Fab-zip-fago, utilizando el fago auxiliar M13KO7 estándar (Fig. 14).

La capacidad de expresar diferentes formatos de anticuerpos a partir del mismo clon es útil en el descubrimiento de anticuerpos cuando se requieren diferentes formatos de anticuerpos, tales como IgG natural, fragmentos Fab o IgG con modificaciones en Fc para formatos biospecíficos, para diferentes ensayos de cribado. El conjunto de vectores permite, en principio, cualquiera de estos formatos o formatos adicionales simplemente clonando una secuencia adecuada para añadirla después de la región CH1 en el plásmido complementario. Además, la organización modular del sistema permite la expresión de nuevos formatos de anticuerpos sin necesidad de volver a crear reservas de colecciones de presentación en fagos, ya que esto solo requiere la construcción de un plásmido complementario novedoso. El vector doble también se podría adaptar para permitir el uso de cualquier región CH1 mediante desplazamiento del sitio de empalme 5'ss de la región que codifica CH1 hasta la región J (FR4) en el dominio VH o en el punto de unión J-CH1, separando por tanto el dominio VH y toda la región constante de la cadena pesada en dos plásmidos diferentes. Los vectores pDV2 son compatibles con los procedimientos tradicionales para la expresión de fragmentos Fab en *E. coli*, simplemente añadiendo un codón de parada después de la secuencia que codifica la bisagra superior, con el conocimiento de que los codones de parada ámbar en el punto de unión de las secuencias de cadena pesada y del gen III en las colecciones de presentación en fagos de Fab normalmente dan como resultado niveles significativamente menores de presentación, lo que requiere por tanto reformatear clones después de la selección, al menos en el caso de las colecciones de repertorios indiferenciados (Lee *et al.*, *Journal of immunological methods* 284:119-132, 2004). La expresión de fragmentos Fab en células de mamífero usando los mismos procedimientos usados para la expresión de IgG evita esta necesidad de reformateo, con rendimientos comparables a los obtenidos normalmente en *E. coli*.

### Ejemplo 3. Sistemas modulares de expresión de proteínas

Los sistemas de expresión de polipéptidos generados y caracterizados en los ejemplos 1 y 2 demuestran que la expresión modular y flexible de polipéptidos de cualquier proteína deseada se puede lograr directamente mediante el uso de un sistema de expresión de polipéptidos, tal como los sistemas de expresión optimizados descritos anteriormente para el reformateo de proteínas en el contexto de presentación en fagos. En consecuencia, el sistema de expresión incluirá dos componentes de moléculas de ácido nucleico (secuencias que codifican polipéptidos PES1<sub>1</sub> y PES2) que codifican cada uno una parte de un único producto polipeptídico deseado, en el que estas regiones de codificación divididas de la proteína se unen entre sí con precisión *in vivo* a través de empalme *trans* de pre-ARNm sin la necesidad de subclonar el ácido nucleico que codifica la proteína. Como se muestra en la Fig. 10, la primera molécula de ácido nucleico incluye un casete de expresión que tiene la PES1<sub>1</sub> y también incluye un promotor eucariota (P1<sub>Euk1</sub>) y una secuencia señal eucariota (ESS1<sub>1</sub>) situada corriente arriba del componente PES1<sub>1</sub>, así como un sitio de empalme en 5' (5'ss<sub>1</sub>) y una secuencia de hibridación (HS1) localizada corriente abajo de la PES1<sub>1</sub>. La segunda molécula de ácido nucleico complementaria incluiría también un promotor eucariota (P2<sub>Euk</sub>), así como una secuencia

de hibridación capaz de hibridarse con HS1 (HS2) y un sitio de empalme en 3' (3'ss2) situado corriente arriba del componente PES2. Además, la segunda molécula de ácido nucleico incluiría un sitio de poliadenilación (pA2) situado corriente abajo del componente PES2. Por tanto, cuando se transcriben en una célula de mamífero, las dos moléculas de pre-ARNm generadas, una con un sitio de empalme 5'ss solo y la otra con un sitio de empalme 3'ss solo, serían dirigidas juntas por sus secuencias de hibridación complementarias (HS1 y HS2) y sufrirían empalme *trans* para formar un único ARNm contiguo capaz de traducción posterior y que codifica el producto proteico deseado.

La primera molécula de ácido nucleico incluye además un módulo promotor procariota escindible (ePPM<sub>1</sub>) que se sitúa entre el P1<sub>Euk1</sub> y la PES1<sub>1</sub>. El ePPM<sub>1</sub> puede incluir un sitio de empalme en 5' (5'ss1<sub>2</sub>), un promotor procariota (P1<sub>Prok1</sub>), una primera secuencia de ácido nucleico que codifica una secuencia señal procariota (PSS1<sub>1</sub>) y un sitio de empalme en 3' (3'ss1<sub>1</sub>) enlazado de forma funcional uno con respecto al otro en dirección de 5' a 3' como 5'ss1<sub>2</sub>-P1<sub>Prok1</sub>-PSS1<sub>1</sub>-3'ss1<sub>1</sub>. El ePPM<sub>1</sub> daría lugar a la transcripción del polipéptido codificado de la primera molécula de ácido nucleico en una célula procariota. Por otra parte, en una célula eucariota (por ejemplo, célula de mamífero), el P1<sub>Euk1</sub> daría lugar a la expresión de la transcripción del polipéptido codificado de la primera molécula de ácido nucleico, y el ePPM<sub>1</sub> se eliminaría del transcrito de pre-ARNm mediante empalme *cis* en virtud de los resultados positivos que flanquean los componentes 5'ss1<sub>2</sub> y 3'ss1<sub>1</sub>.

En algunos casos, puede ser deseable expresar también un segundo polipéptido. La primera molécula de ácido nucleico del sistema modular de expresión de proteínas se puede diseñar en consecuencia para incluir un segundo casete de expresión. Como se muestra en la Fig. 11, el segundo casete de expresión que codifica el segundo producto proteico (PES1<sub>2</sub>) se diseñaría de manera similar al primer casete de expresión, pero contendría un sitio de poliadenilación (pA1) situado corriente abajo de la secuencia PES1<sub>2</sub> para garantizar la generación de una molécula de pre-ARNm distinta después de la transcripción. En otros casos, el segundo casete de expresión se podría diseñar en la segunda molécula de ácido nucleico del sistema de expresión de polipéptidos.

25

LISTADO DE SECUENCIAS

<110> GENENTECH, INC.  
 F. HOFFMANN-LA ROCHE AG  
 5 <120> SISTEMAS DE EXPRESIÓN DE POLIPÉPTIDOS  
 <130> 50474-073WO3  
 10 <150> 62/080, 698  
 <151> 17/11/2014  
 <150> 62/020, 838  
 <151> 03/07/2014  
 15 <160> 31  
 <170> PatentIn versión 3.5  
 20 <210> 1  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial  
 25 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: oligonucleótido sintético"  
 <400> 1  
 30 **ttcctttttt ctctttcc** 18  
 <210> 2  
 <211> 400  
 <212> ADN  
 35 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"  
 40 <400> 2  
**gccaccatga tgaaatctac cgtggtggcg gcgcgctgc tgctgctggg cggtaagtac** 60  
**cgctataga gtctataggc ccaccccctt ggcttacgta aagaagttat tgaagcatcc** 120  
**tcgtcagtaa aaagttaatc ttttcaacag ctgtcataaa gttgtcacgg ccgagactta** 180  
**tagtcgcttt gtttttattt ttttatgtat ttgtaactag tacgcaagtt cacgtaaaaa** 240  
**gggtatgtag aggttgaggt gattttatga aaaagaatat cgcatttctt cttgcatcta** 300  
**tgttcctttt ttctctttcc acagccgtac gcgagatat ccagatgacc cagtccccga** 360  
**gctccctgtc cgctctgtg ggcgataggg tcaccatcac** 400  
 <210> 3  
 45 <211> 14  
 <212> PRT  
 <213> Secuencia artificial  
 <220>  
 50 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: péptido sintético"

<400> 3  
**Met Met Lys Phe Thr Val Val Ala Ala Ala Leu Leu Leu Leu**  
 1 5 10

5 <210> 4  
 <211> 37  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 4  
**Ala Ser Ile Ala Arg Leu Glu Glu Lys Val Lys Thr Leu Lys Ala Gln**  
 1 5 10 15

**Asn Tyr Glu Leu Ala Ser Thr Ala Asn Met Leu Arg Glu Gln Val Ala**  
 20 25 30

**Gln Leu Gly Gly Cys**  
 35

15 <210> 5  
 <211> 37  
 <212> PRT  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

25 <400> 5  
**Ala Ser Ile Asp Glu Leu Gln Ala Glu Val Glu Gln Leu Glu Glu Arg**  
 1 5 10 15

**Asn Tyr Ala Leu Arg Lys Glu Val Glu Asp Leu Gln Lys Gln Ala Glu**  
 20 25 30

**Lys Leu Gly Gly Cys**  
 35

30 <210> 6  
 <211> 42  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: oligonucleótido sintético"

40 <220>  
 <221> base\_modificada  
 <222> (6)..(6)  
 <223> a, c, t, g, desconocida u otras

45 <220>  
 <221> base\_modificada  
 <222> (9)..(9)  
 <223> a, c, t, g, desconocida u otras

5 <220>  
 <221> base\_modificada  
 <222> (12)..(12)  
 <223> a, c, t, g, desconocida u otras

10 <220>  
 <221> base\_modificada  
 <222> (15)..(15)  
 <223> a, c, t, g, desconocida u otras

15 <220>  
 <221> base\_modificada  
 <222> (18)..(18)  
 <223> a, c, t, g, desconocida u otras

20 <220>  
 <221> base\_modificada  
 <222> (21)..(21)  
 <223> a, c, t, g, desconocida u otras

25 <220>  
 <221> base\_modificada  
 <222> (24)..(24)  
 <223> a, c, t, g, desconocida u otras

30 <220>  
 <221> base\_modificada  
 <222> (27)..(27)  
 <223> a, c, t, g, desconocida u otras

35 <220>  
 <221> base\_modificada  
 <222> (30)..(30)  
 <223> a, c, t, g, desconocida u otras

40 <220>  
 <221> base\_modificada  
 <222> (33)..(33)  
 <223> a, c, t, g, desconocida u otras

45 <220>  
 <221> base\_modificada  
 <222> (36)..(36)  
 <223> a, c, t, g, desconocida u otras

50 <220>  
 <221> base\_modificada  
 <222> (39)..(39)  
 <223> a, c, t, g, desconocida u otras

55 <220>  
 <221> base\_modificada  
 <222> (42)..(42)  
 <223> a, c, t, g, desconocida u otras

<400> 6  
**atgaanttna cngtngtngc ngcngcncn ctncnctng gn**

42

60 <210> 7  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

65 <220>  
 <221> fuente

ES 2 763 526 T3

<223> /nota="Descripción de secuencia artificial: oligonucleótido sintético"

<400> 7  
**ttcctctttc cctttctctc c** 21

5  
 <210> 8  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial

10  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: oligonucleótido sintético"

15  
 <400> 8  
**gtaaga** 6

<210> 9  
 <211> 45  
 <212> PRT  
 <213> Secuencia artificial

20  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

25  
 <400> 9  
**Met Lys Lys Asn Ile Ala Phe Leu Leu Ala Ser Met Phe Leu Phe Ser**  
 1 5 10 15

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

**Ser Leu Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr**  
 35 40 45

30  
 <210> 10  
 <211> 400  
 <212> ADN  
 <213> Secuencia artificial

35  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"

<400> 10  
**gccaccatgg gatggtcacg tctctcctt tttctagtag caactgcaac aggtaagtac** 60

**cgcctataga gtctataggg ccacccctt ggcttacgta aagaagttat tgaagcatcc** 120

**tcgtcagtaa aaagttaatc ttttcaacag ctgtcataaa gttgtcacgg ccgagactta** 180

**tagtcgcttt gtttttattt ttttatcatc gtggccctag tacgcaaatt cacgtaaaaa** 240

**gggtaactag aggttgaggt gattttatga aaaagaatat cgcatttctt cttgcatcta** 300

**tgttcctttt ttctctttcc acagcgggtcc gcgcagaagt tcagctggtg gaggctggcg** 360

40  
**gtggcctggt gcagccaggg ggctcactcc gtttgcctg** 400

<210> 11  
 <211> 37

<212> PRT  
 <213> Secuencia artificial

<220>  
 5 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 11  
 Ala Ser Ile Ala Arg Leu Arg Glu Arg Val Lys Thr Leu Arg Ala Arg  
 1 5 10 15

Asn Tyr Glu Leu Arg Ser Arg Ala Asn Met Leu Arg Glu Arg Val Ala  
 20 25 30

Gln Leu Gly Gly Cys  
 35

10 <210> 12  
 <211> 37  
 <212> PRT  
 <213> Secuencia artificial

15 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

20 <400> 12  
 Ala Ser Leu Asp Glu Leu Glu Ala Glu Ile Glu Gln Leu Glu Glu Glu  
 1 5 10 15

Asn Tyr Ala Leu Glu Lys Glu Ile Glu Asp Leu Glu Lys Glu Leu Glu  
 20 25 30

Lys Leu Gly Gly Cys  
 35

25 <210> 13  
 <211> 38  
 <212> PRT  
 <213> Secuencia artificial

30 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 13  
 Glu Glu Lys Ser Arg Leu Leu Glu Lys Glu Asn Arg Glu Leu Glu Lys  
 1 5 10 15

Ile Ile Ala Glu Lys Glu Glu Arg Val Ser Glu Leu Arg His Gln Leu  
 20 25 30

Gln Ser Val Gly Gly Cys  
 35

35 <210> 14  
 <211> 38  
 <212> PRT  
 <213> Secuencia artificial

40 <220>  
 <221> fuente

ES 2 763 526 T3

<223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 14  
 Thr Ser Arg Leu Glu Gly Leu Gln Ser Glu Asn His Arg Leu Arg Met  
 1 5 10 15

Lys Ile Thr Glu Leu Asp Lys Asp Leu Glu Glu Val Thr Met Gln Leu  
 20 25 30

Gln Asp Val Gly Gly Cys  
 35

5

<210> 15  
 <211> 4  
 <212> PRT  
 <213> Secuencia artificial

10

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: péptido sintético"

15

<400> 15  
 Ala Gly Ser Cys  
 1

20

<210> 16  
 <211> 6  
 <212> PRT  
 <213> Secuencia artificial

25

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: péptido sintético"

<400> 16  
 Cys Pro Pro Cys Pro Gly  
 1 5

30

<210> 17  
 <211> 418  
 <212> PRT  
 <213> Secuencia artificial

35

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 17  
 Ala Glu Asp Ile Glu Phe Ala Ser Gly Gly Gly Ser Gly Ala Glu Thr  
 1 5 10 15

Val Glu Ser Cys Leu Ala Lys Pro His Thr Glu Asn Ser Phe Thr Asn  
 20 25 30

Val Trp Lys Asp Asp Lys Thr Leu Asp Arg Tyr Ala Asn Tyr Glu Gly  
 35 40 45

Cys Leu Trp Asn Ala Thr Gly Val Val Val Cys Thr Gly Asp Glu Thr  
 50 55 60

Gln Cys Tyr Gly Thr Trp Val Pro Ile Gly Leu Ala Ile Pro Glu Asn  
 65 70 75 80

40

ES 2 763 526 T3

Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser Glu  
 85 90 95

Gly Gly Gly Thr Lys Pro Pro Glu Tyr Gly Asp Thr Pro Ile Pro Gly  
 100 105 110

Tyr Thr Tyr Ile Asn Pro Leu Asp Gly Thr Tyr Pro Pro Gly Thr Glu  
 115 120 125

Gln Asn Pro Ala Asn Pro Asn Pro Ser Leu Glu Glu Ser Gln Pro Leu  
 130 135 140

Asn Thr Phe Met Phe Gln Asn Asn Arg Phe Arg Asn Arg Gln Gly Ala  
 145 150 155 160

Leu Thr Val Tyr Thr Gly Thr Val Thr Gln Gly Thr Asp Pro Val Lys  
 165 170 175

Thr Tyr Tyr Gln Tyr Thr Pro Val Ser Ser Lys Ala Met Tyr Asp Ala  
 180 185 190

Tyr Trp Asn Gly Lys Phe Arg Asp Cys Ala Phe His Ser Gly Phe Asn  
 195 200 205

Glu Asp Pro Phe Val Cys Glu Tyr Gln Gly Gln Ser Ser Asp Leu Pro  
 210 215 220

Gln Pro Pro Val Asn Ala Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly  
 225 230 235 240

Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly Ser Glu Gly Gly Gly  
 245 250 255

Ser Glu Gly Gly Gly Ser Gly Gly Gly Ser Gly Ser Gly Asp Phe Asp  
 260 265 270

Tyr Glu Lys Met Ala Asn Ala Asn Lys Gly Ala Met Thr Glu Asn Ala  
 275 280 285

Asp Glu Asn Ala Leu Gln Ser Asp Ala Lys Gly Lys Leu Asp Ser Val  
 290 295 300

Ala Thr Asp Tyr Gly Ala Ala Ile Asp Gly Phe Ile Gly Asp Val Ser  
 305 310 315 320

Gly Leu Ala Asn Gly Asn Gly Ala Thr Gly Asp Phe Ala Gly Ser Asn  
 325 330 335

Ser Gln Met Ala Val Gly Asp Gly Asp Asn Ser Pro Leu Met Asn Asn  
 340 345 350

Phe Arg Gln Tyr Leu Pro Ser Leu Pro Gln Ser Val Glu Cys Arg Pro  
 355 360 365

Phe Val Phe Ser Ala Gly Lys Pro Tyr Glu Phe Ser Ile Asp Cys Asp  
 370 375 380

ES 2 763 526 T3

Lys Ile Asn Leu Phe Arg Gly Val Phe Ala Phe Leu Leu Tyr Val Ala  
 385 390 395 400

Thr Phe Met Tyr Val Phe Ser Thr Phe Ala Asn Ile Leu Arg Asn Lys  
 405 410 415

Glu Ser

- <210> 18
- <211> 158
- 5 <212> PRT
- <213> Secuencia artificial
- <220>
- <221> fuente
- 10 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 18  
 Ser Gly Gly Gly Ser Gly Ser Gly Asp Phe Asp Tyr Glu Lys Met Ala  
 1 5 10 15

Asn Ala Asn Lys Gly Ala Met Thr Glu Asn Ala Asp Glu Asn Ala Leu  
 20 25 30

Gln Ser Asp Ala Lys Gly Lys Leu Asp Ser Val Ala Thr Asp Tyr Gly  
 35 40 45

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

Asn Gly Ala Thr Gly Asp Phe Ala Gly Ser Asn Ser Gln Met Ala Gln  
 65 70 75 80

Val Gly Asp Gly Asp Asn Ser Pro Leu Met Asn Asn Phe Arg Gln Tyr  
 85 90 95

Leu Pro Ser Leu Pro Gln Ser Val Glu Cys Arg Pro Phe Val Phe Gly  
 100 105 110

Ala Gly Lys Pro Tyr Glu Phe Ser Ile Asp Cys Asp Lys Ile Asn Leu  
 115 120 125

Phe Arg Gly Val Phe Ala Phe Leu Leu Tyr Val Ala Thr Phe Met Tyr  
 130 135 140

Val Phe Ser Thr Phe Ala Asn Ile Leu Arg Asn Lys Glu Ser  
 145 150 155

- 15 <210> 19
- <211> 15
- <212> PRT
- <213> Secuencia artificial
- <220>
- <221> fuente
- 20 <223> /nota="Descripción de secuencia artificial: péptido sintético"
- <400> 19

ES 2 763 526 T3

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

5 <210> 20  
 <211> 45  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 10 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 20  
 Met Lys Lys Asn Ile Ala Phe Leu Leu Ala Ser Met Phe Leu Phe Ser  
 1 5 10 15

Leu Ser Thr Ala Val Arg Ala Glu Val Gln Leu Val Glu Ser Gly Gly  
 20 25 30

Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys  
 35 40 45

15 <210> 21  
 <211> 600  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"

<400> 21  
 cacaagccca gcaacaccaa ggtcgacaag aaagtaggta agaaatcttg tgacaaaact 60  
 cacacatgcg gccggccctc tggttccggt gattttgatt atgaaaagat ggcaaacgct 120  
 aataaggggg ctatgaccga aaatgccgat gaaaacgcgc tacagtctga cgctaaaggc 180  
 aaacttgatt ctgtcgctac tgattacggt gctgctatcg atggtttcat tggtgacgtt 240  
 tccggccttg ctaatggtaa tgggtgctact ggtgattttg ctggctctaa ttcccaaatg 300  
 25 gctcaagtcg gtgacggtga taattcacct ttaatgaata atttccgtca atatttacct 360  
 tccctccctc aatcggttga atgtcgccct tttgtcttta gcgctggtaa accatatgaa 420  
 ttttctattg attgtgacaa gattaactta ttccgtggtg tctttgctgt tcttttatat 480  
 gttgccacct ttatgtatgt attttctacg tttgctaaca tactgcgtaa taaggagtct 540  
 taaccgtggg gtgaacgctc ggttgccgcc gggcgttttt tatgttcttt tggaggatcc 600

30 <210> 22  
 <211> 180  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 35 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 22

ES 2 763 526 T3

His Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Gly Lys Lys Ser  
 1 5 10 15

Cys Asp Lys Thr His Thr Cys Gly Arg Pro Ser Gly Ser Gly Asp Phe  
 20 25 30

Asp Tyr Glu Lys Met Ala Asn Ala Asn Lys Gly Ala Met Thr Glu Asn  
 35 40 45

Ala Asp Glu Asn Ala Leu Gln Ser Asp Ala Lys Gly Lys Leu Asp Ser  
 50 55 60

Val Ala Thr Asp Tyr Gly Ala Ala Ile Asp Gly Phe Ile Gly Asp Val  
 65 70 75 80

Ser Gly Leu Ala Asn Gly Asn Gly Ala Thr Gly Asp Phe Ala Gly Ser  
 85 90 95

Asn Ser Gln Met Ala Gln Val Gly Asp Gly Asp Asn Ser Pro Leu Met  
 100 105 110

Asn Asn Phe Arg Gln Tyr Leu Pro Ser Leu Pro Gln Ser Val Glu Cys  
 115 120 125

Arg Pro Phe Val Phe Ser Ala Gly Lys Pro Tyr Glu Phe Ser Ile Asp  
 130 135 140

Cys Asp Lys Ile Asn Leu Phe Arg Gly Val Phe Ala Phe Leu Leu Tyr  
 145 150 155 160

Val Ala Thr Phe Met Tyr Val Phe Ser Thr Phe Ala Asn Ile Leu Arg  
 165 170 175

Asn Lys Glu Ser  
 180

- 5 <210> 23  
 <211> 600  
 <212> ADN  
 <213> Secuencia artificial
- 10 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"
- <400> 23

ES 2 763 526 T3

```

gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg ccatccacgc      60
tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg ggaacggtgc      120
attggaacgc ggattccccg tgccaagagt gacgaacgtc accaatgaaa ccatcgatag      180
cagcaccgta atcagtagcg acagaatcaa gtttgccttt agcgtcagac tgtagcgcgt      240
tttcatcggc attttcggtc atagccccct tattagcgtt tgccatcttt tcataatcaa      300
aatcccaccc ctttggttcc gttagtagac tagatctaact ctactctact aaccgcctat      360
agagtctata ggcccacccc cttggcttta ctgacacacg aattcctctt tccctttctc      420
tccacagagc ccaaactctg tgacaaaact cacacatgcc caccgtgccc agcacctgaa      480
ctcctgggcg gaccgtcagt cttcctcttc cccccaaaac ccaaggacac cctcatgatc      540
tcccggaccc ctgaggtcac atgcgtggtg gtggacgtga gccacgaaga ccctgaggtc      600

```

5 <210> 24  
 <211> 58  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

```

<400> 24
Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala
1          5          10          15

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
          20          25          30

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
          35          40          45

Val Asp Val Ser His Glu Asp Pro Glu Val
          50          55

```

15 <210> 25  
 <211> 15  
 <212> PRT  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: péptido sintético"

```

<400> 25
Met Met Lys Phe Thr Val Val Ala Ala Ala Leu Leu Leu Leu Gly
1          5          10          15

```

30 <210> 26  
 <211> 600  
 <212> ADN  
 <213> Secuencia artificial

<220>

ES 2 763 526 T3

<221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"

<400> 26  
 gtggtgaccg tgcctccag cagcttgggc acccagacct acatctgcaa cgtgaatcac 60  
 aagcccagca acaccaaggt cgacaagaaa gtaggtaaga aatcttgtga caaaactcac 120  
 acatgcgggg gggggagtgg gagcggggat tttgattatg aaaagatggc aaacgctaata 180  
 aagggggcta tgaccgaaaa tgccgatgaa aacgcgctac agtctgacgc taaaggcaaaa 240  
 cttgattctg tcgctactga ttacggtgct gctatcgatg gtttcattgg tgacgtttcc 300  
 ggccttgcta atggtaatgg tgctactggt gattttgctg gctctaattc ccaaattggct 360  
 caagtcggtg acggtgataa ttcaccttta atgaataatt tccgtcaata tttaccttcc 420  
 ctccctcaat cggttgaatg tcgccctttt gtctttagcg ctggtaaacc atatgaattt 480  
 tctattgatt gtgacaagat caacttattc cgtggtgtct ttgcgtttct tttatatggt 540  
 5 gccaccttta tgtatgtatt ttctacgttt gctaacatac tgcgtaataa ggagtcttaa 600

<210> 27  
 <211> 199  
 <212> PRT  
 10 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 27  
 Val Val Thr Val Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys  
 1 5 10 15

ES 2 763 526 T3

Asn Val Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Gly  
 20 25 30

Lys Lys Ser Cys Asp Lys Thr His Thr Cys Gly Gly Gly Ser Gly Ser  
 35 40 45

Gly Asp Phe Asp Tyr Glu Lys Met Ala Asn Ala Asn Lys Gly Ala Met  
 50 55 60

Thr Glu Asn Ala Asp Glu Asn Ala Leu Gln Ser Asp Ala Lys Gly Lys  
 65 70 75 80

Leu Asp Ser Val Ala Thr Asp Tyr Gly Ala Ala Ile Asp Gly Phe Ile  
 85 90 95

Gly Asp Val Ser Gly Leu Ala Asn Gly Asn Gly Ala Thr Gly Asp Phe  
 100 105 110

Ala Gly Ser Asn Ser Gln Met Ala Gln Val Gly Asp Gly Asp Asn Ser  
 115 120 125

Pro Leu Met Asn Asn Phe Arg Gln Tyr Leu Pro Ser Leu Pro Gln Ser  
 130 135 140

Val Glu Cys Arg Pro Phe Val Phe Ser Ala Gly Lys Pro Tyr Glu Phe  
 145 150 155 160

Ser Ile Asp Cys Asp Lys Ile Asn Leu Phe Arg Gly Val Phe Ala Phe  
 165 170 175

Leu Leu Tyr Val Ala Thr Phe Met Tyr Val Phe Ser Thr Phe Ala Asn  
 180 185 190

Ile Leu Arg Asn Lys Glu Ser  
 195

<210> 28

<211> 600

5 <212> ADN

<213> Secuencia artificial

<220>

<221> fuente

10 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"

<400> 28

gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg ccatccacgc 60

tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg ggaacggtgc 120

ES 2 763 526 T3

attggaacgc ggattccccg tgccaagagt gacgaacgtc accaatgaaa ccatcgatag 180  
 cagcaccgta atcagtagcg acagaatcaa gtttgccttt agcgtcagac tgtagcgcgt 240  
 tttcatcggc attttcggtc atagccccct tattagcggt tgccatcttt tcataatcaa 300  
 aatcccaccc ccttggttc gttagtagac tagatctaact ctactctact agggggggggg 360  
 accgcctata gagtctatag gccaccccc ttggctttac tgacacacga attcctcttt 420  
 ccctttctct ccacagagcc caaatcttgt gacaaaactc acacatgccc accgtgcccc 480  
 gcacctgaac tcctggggcg accgtcagtc ttctctctcc ccccaaaacc caaggacacc 540  
 ctcatgatct cccggacccc tgaggtcaca tgcgtggtgg tggacgtgag ccacgaagac 600

5 <210> 29  
 <211> 55  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polipéptido sintético"

<400> 29  
 Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala  
 1 5 10 15

Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro  
 20 25 30

Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val  
 35 40 45

Val Asp Val Ser His Glu Asp  
 50 55

15 <210> 30  
 <211> 500  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: polinucleótido sintético"

<400> 30  
 gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg ccatccacgc 60

tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg ggaacggtgc 120

attggaacgc ggattccccg tgccaagagt gacgaacgtc accaatgaaa ccatcgatag 180

25 cagcaccgta atcagtagcg acagaatcaa gtttgccttt agcgtcagac tgtagcgcgt 240

ES 2 763 526 T3

```

tttcatcggc attttcggtc atagcccct tattagcgtt tgccatcttt tcataatcaa      300
aatcccaccc ccttggcttc gttagtagac tagatctaata ctactctact aggggggggg      360
accgcctata gagtctatag gccaccccc ttggetttac tgacacacga attcctcttt      420
ccctttctct ccacagagcc caaatcttgt gacaaaactc acacagatta taaggacgat      480
gacgataaat gagtgcgacg                                                    500

```

<210> 31

<211> 18

5 <212> PRT

<213> Secuencia artificial

<220>

<221> fuente

10 <223> /nota="Descripción de secuencia artificial: péptido sintético"

<400> 31

```

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Asp Tyr Lys Asp Asp Asp
1           5                10                15

```

**Asp Lys**

## REIVINDICACIONES

1. Un sistema de expresión de polipéptidos que comprende una primera molécula de ácido nucleico y una segunda molécula de ácido nucleico, en el que:
- 5 (a) la primera molécula de ácido nucleico comprende un primer casete de expresión que comprende los siguientes componentes: (i) un primer promotor eucariota ( $P1_{Euk1}$ ), (ii) una primera secuencia que codifica polipéptidos ( $PES1_1$ ), (iii) un primer sitio de empalme en 5' ( $5'ss1_1$ ) y (iv) una secuencia de hibridación ( $HS1$ ), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como  $P1_{Euk1}$ - $PES1_1$ - $5'ss1_1$ - $HS1$ ; en el que el primer casete de expresión comprende además un módulo promotor procariota escindible ( $ePPM_1$ ) que comprende los siguientes componentes: (i) un sitio de empalme en 5' ( $5'ss1_2$ ), (ii) un promotor procariota ( $P1_{Prok1}$ ) y (iii) un sitio de empalme en 3' ( $3'ss1_1$ ), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 2' como  $5'ss1_3$ - $P1_{Prok1}$ - $3'ss1_1$ , y en el que el  $ePPM_1$  se sitúa entre el  $P1_{Euk1}$  y la  $PES1_1$ , y
- 15 (b) la segunda molécula de ácido nucleico comprende los siguientes componentes: (i) un promotor eucariota ( $P2_{Euk}$ ), (ii) una secuencia de hibridación capaz de hibridar con  $HS1$  ( $HS2$ ), (iii) un sitio de empalme en 3' ( $3'ss2$ ), (iv) una secuencia que codifica polipéptidos ( $PES2$ ), y (v) un sitio de poliadenilación ( $pA2$ ), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como  $P2_{Euk}$ - $HS2$ - $3'ss2$ - $PES2$ - $pA2$ .
- 20 2. El sistema de expresión de polipéptidos de la reivindicación 1, en el que el  $P1_{Euk1}$  y/o  $P2_{Euk}$  es un promotor de citomegalovirus (CMV) o un promotor del virus del simio 40 (SV40).
3. El sistema de expresión de polipéptidos de la reivindicación 1 o 2, en el que el primer casete de expresión comprende además una primera secuencia de ácido nucleico que codifica una secuencia señal eucariota ( $ESS1_1$ ), en el que la  $ESS1_1$  se sitúa entre el  $P1_{Euk1}$  y la  $PES1_1$ , preferentemente en el que la  $ESS1_1$  se deriva del gen de la cadena pesada variable (VH).
- 25 4. El sistema de expresión de polipéptidos de la reivindicación 1, en el que el  $P1_{Prok1}$  se selecciona del grupo que consiste en un promotor PhoA, un promotor Tac, un promotor Lac y un promotor Tphac.
- 30 5. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1 a 4, en el que el  $ePPM_1$  comprende además una primera secuencia de ácido nucleico que codifica una secuencia señal procariota ( $PSS1_1$ ), preferentemente en el que la  $PSS1_1$  se deriva del gen de enterotoxina II (stII) termoestable.
- 35 6. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-5, que comprende además un tramo de polipirimidina situado
- a) entre la  $PSS1_1$  y el  $3'ss1_1$  ( $PPT1_1$ ); o
- 40 b) entre la  $PSS1_1$  y el  $3'ss1_1$  ( $PPT1_1$ ), en el que el  $PPT1_1$  comprende la secuencia de ácido nucleico de TTCCTTTTTTCTCTTCC (SEQ ID NO: 1).
7. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-6, en el que la  $PES1_1$  no comprende un sitio crítico de empalme en 5'.
- 45 8. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-7, en el que la  $HS1$  es un gen que codifica toda o una parte de
- a) una proteína de cubierta, preferentemente en el que la proteína de cubierta
- 50 1) se selecciona del grupo que consiste en pI, pII, pIII, pIV, pV, pVI, pVII, pVIII, pIX y pX del bacteriófago M13, f1 o fd; o
- 2) es la proteína pIII del bacteriófago M13; o
- 55 3) es la proteína pIII del bacteriófago M13, en el que la proteína pIII es un fragmento pIII que comprende los residuos de aminoácidos 267-421 de la proteína pIII o los residuos de aminoácidos 262-418 de la proteína pIII; o
- b) una proteína adaptadora, preferentemente en el que la proteína adaptadora es 1) una cremallera de leucina, o 2) una cremallera de leucina que comprende la secuencia de aminoácidos de SEQ ID NO: 4 o 5.
- 60 9. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-8, en el que la primera molécula de ácido nucleico comprende además (i) un segundo casete de expresión que comprende un segundo promotor eucariota ( $P1_{Euk2}$ ), (ii) una segunda secuencia que codifica polipéptidos ( $PES1_2$ ), y (iii) un sitio de poliadenilación ( $pA1$ ), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como  $P1_{Euk2}$ - $PES1_2$ - $pA1$ , preferentemente en el que el  $P1_{Euk2}$  es un promotor de CMV o un promotor de SV40.
- 65

10. El sistema de expresión de polipéptidos de la reivindicación 9, en el que el segundo casete de expresión comprende además
- 5 a) una secuencia de ácido nucleico que codifica una secuencia señal eucariota (ESS1<sub>2</sub>), preferentemente en el que la ESS1<sub>2</sub> se deriva del gen de la proteína de unión a inmunoglobulina murina (mBiP); y/o
- b) una secuencia de ácido nucleico que codifica una secuencia señal eucariota (ESS1<sub>2</sub>), en la que la ESS1<sub>2</sub> comprende la secuencia de ácido nucleico de ATG AAN TTN ACN GTN GTN GCN GCN GCN CTN CTN CTN  
 10 CTN GGN, en la que N es A, T, C o G (SEQ ID NO: 6).
11. El sistema de expresión de polipéptidos de la reivindicación 9 o 10, en el que el segundo casete de expresión comprende además un módulo promotor procariota escindible (ePPM<sub>2</sub>) que comprende los siguientes componentes:
- 15 (i) un sitio de empalme en 5' (5'ss1<sub>3</sub>), (ii) un promotor procariota (P1<sub>Prok2</sub>) y (iii) un sitio de empalme en 3' (3'ss1<sub>2</sub>), en el que los componentes están enlazados de forma funcional entre sí en dirección de 5' a 3' como 5'ss1<sub>3</sub>-P1<sub>Prok2</sub>-3'ss1<sub>2</sub>, y en el que el ePPM<sub>2</sub> se sitúa entre el P1<sub>Euk2</sub> y la PES1<sub>2</sub>, preferentemente en el que
- (a) el P1<sub>Prok2</sub> se selecciona del grupo que consiste en un promotor PhoA, un promotor Tac y un promotor Lac; y/o
- 20 (b) el ePPM<sub>2</sub> comprende además una secuencia de ácido nucleico que codifica una secuencia señal procariota (PSS1<sub>2</sub>); y/o
- (c) la PSS1<sub>2</sub> se deriva del gen de enterotoxina II (stII) termoestable.
12. El sistema de expresión de polipéptidos de la reivindicación 11, que comprende además un tramo de polipirimidina situado
- 25 a) entre la PSS1<sub>2</sub> y el 3'ss1<sub>2</sub> (PPT1<sub>2</sub>); o
- b) entre la PSS1<sub>2</sub> y el 3'ss1<sub>2</sub> (PPT1<sub>2</sub>), en el que el PPT1<sub>2</sub> comprende la secuencia de ácido nucleico de TTCCTTTTTTCTCTTCC (SEQ ID NO: 1).
13. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 9-12, en el que el segundo casete de expresión se sitúa en dirección 5' al primer casete de expresión.
- 35 14. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-13, que comprende además un potenciador intrónico de empalme (ISE) situado
- a) entre el 5'ss1<sub>1</sub> y la HS1 (ISE1); o
- 40 b) entre el 5'ss1<sub>1</sub> y la HS1 (ISE1), en el que el ISE1 comprende una secuencia rica en G (G-run) que comprende 1) tres o más residuos de guanina consecutivos, o 2) nueve residuos de guanina consecutivos.
15. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-14, que comprende además un tramo de polipirimidina situado
- 45 a) entre la HS2 y el 3'ss2 (PPT2); o
- b) entre la HS2 y el 3'ss2 (PPT2), en el que el PPT2 comprende la secuencia de ácido nucleico de TTCCTCTTCCCTTCTCTCC (SEQ ID NO: 7).
- 50 16. El sistema de expresión de polipéptidos de la reivindicación 15, que comprende además un ISE situado
- a) entre la HS2 y el 3'ss2 (ISE2); o
- 55 b) entre la HS2 y el 3'ss2 (ISE2), en el que el ISE2 comprende una secuencia rica en G (G-run) que comprende 1) tres o más residuos de guanina consecutivos, o 2) nueve residuos de guanina consecutivos.
17. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-16, en el que el 5'ss1<sub>1</sub> comprende la secuencia de ácido nucleico de GTAAGA (SEQ ID NO: 8).
- 60 18. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-17, en el que la expresión por un promotor eucariota se produce en
- 65 a) una célula de mamífero; o

b) una célula de mamífero, en el que la célula de mamífero es una célula Expi293F, una célula CHO, una célula 293T o una célula NSO, preferentemente en el que la célula de mamífero es una célula Expi293F.

5 19. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-18, en el que la expresión por un promotor procariota se produce en

a) una célula bacteriana; o

10 b) una célula de *E. coli*.

20. El sistema de expresión de polipéptidos de las reivindicaciones 1-19, en el que la PES1<sub>1</sub> codifica

a) todo o una parte de un anticuerpo; o

15 b) un polipéptido que comprende un dominio VH; o

c) un polipéptido que comprende un dominio VH, en el que el polipéptido comprende además un dominio CH1.

21. El sistema de expresión de polipéptidos de la reivindicación 20, en el que la PES2 codifica

20 a) todo o una parte de un anticuerpo; o

b) todo o una parte de un anticuerpo, en el que la PES2 codifica un polipéptido que comprende un dominio CH2 y un dominio CH3.

25 22. El sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 9-21, en el que la PES1<sub>2</sub> codifica

a) todo o una parte de un anticuerpo; o

30 b) todo o una parte de un anticuerpo, en el que PES1<sub>2</sub> codifica un polipéptido que comprende un dominio VL y un dominio CL.

35 23. Un conjunto de vectores que comprende un primer vector y un segundo vector, en el que el primer y el segundo vector comprenden la primera y la segunda molécula de ácido nucleico, respectivamente, del sistema de expresión de polipéptidos de una cualquiera de las reivindicaciones 1-22.

24. Una célula huésped que comprende el conjunto de vectores de la reivindicación 23.

40 25. La célula huésped de la reivindicación 24, en la que la célula huésped es

(a) una célula procariota; o

(b) una célula procariota, en la que la célula procariota es una célula bacteriana, preferentemente en la que la célula bacteriana es una célula de *E. coli*; o

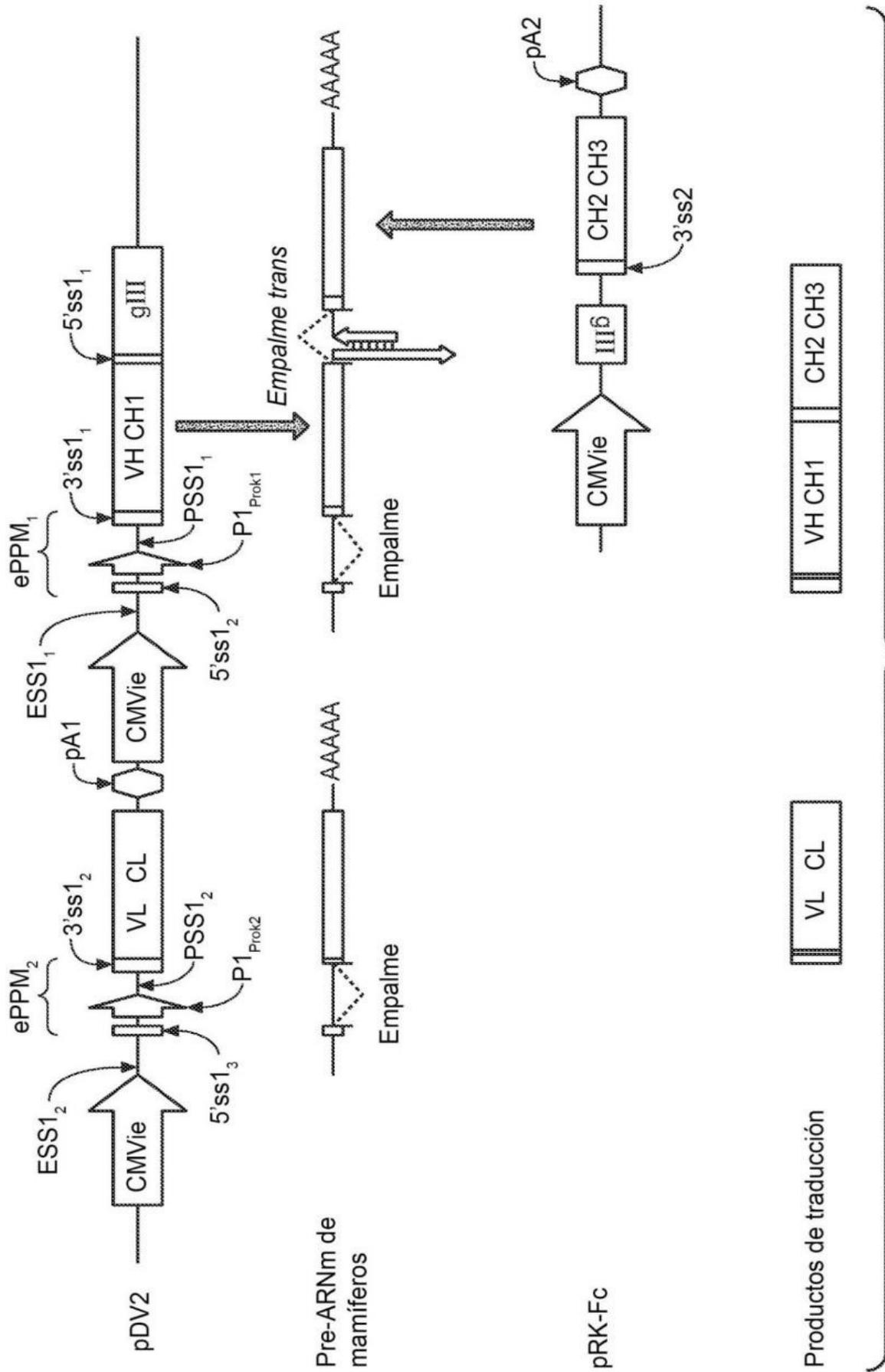
45 (c) una célula eucariota; o

(d) una célula eucariota, en la que la célula eucariota es una célula de mamífero, preferentemente en la que la célula de mamífero es una célula Expi293F, una célula CHO, una célula 293T o una célula NSO; o

50 (e) una célula eucariota, en la que la célula eucariota es una célula Expi293F.

26. Un procedimiento para producir un polipéptido que comprende cultivar una célula huésped que comprende el conjunto de vectores de la reivindicación 23 en un medio de cultivo, preferentemente en el que el procedimiento comprende además recuperar el polipéptido de la célula huésped o del medio de cultivo.

55



**FIG. 1**



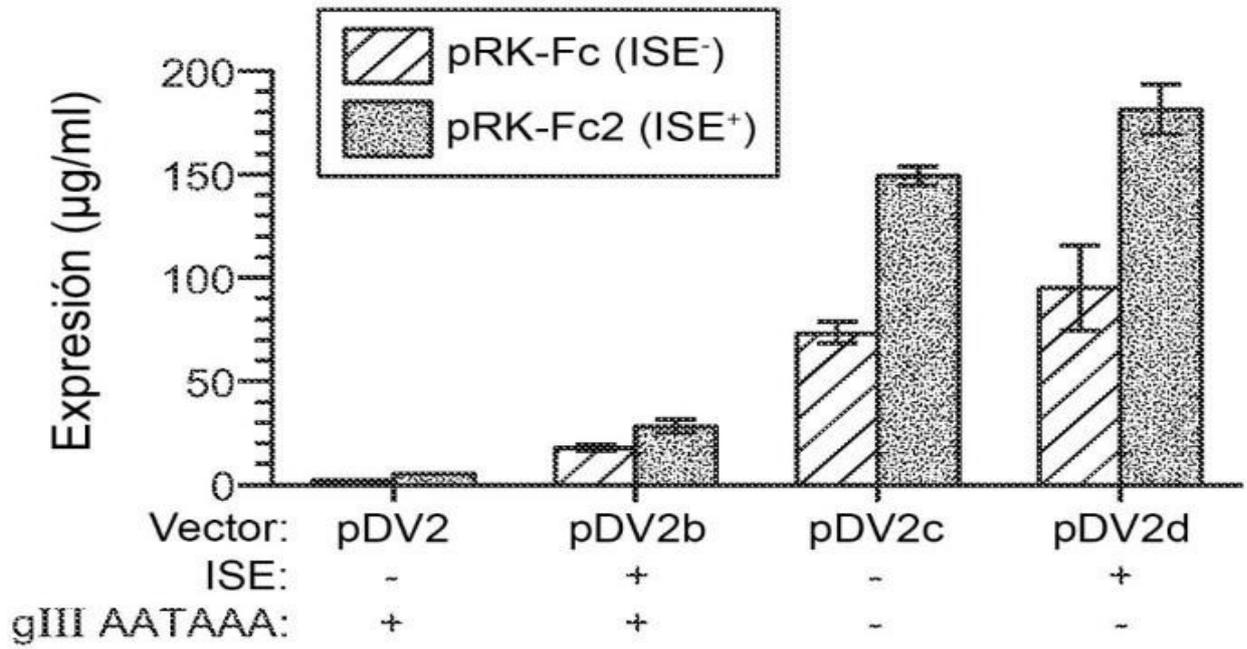
```

Tramo de polipirimidina 3'ss
TGTTCCCTTT TCTCTTCC ACAGCGGTCC GCCCAGAAGT TCAGCTGGTG GAGTCTGGCG GTGGCCTGGT GCAGCCAGGG GGCTCACTCC GTTTGTCCTG
(SEQ ID NO: 10)
F L F S L S T A V R A E V Q L V E S G G L V Q P G G S L R L S C
< Cadena pesada
                    5'ss
CACAAAGCCA GCAACACCAA GGTGACAAG AAAGTAGGTA AGAAATCTTG TGACAAAAC TACACATGG GCCGGCCCTC TGGTTCCGGT GATTTTGATT
H K P S N T K V D K K V G K K S C D K T H T C G R P S G S G D F D Y
CHI
                    < Conector > < P3 C terminal (cP3)
ATGAAAAGAT GGCAAACGCT AATAAGGGG CTATGACCGA AAATGCCGAT GAAAACGGC TACAGTCTGA CGCTAAAGG AACTTGATT CTGTCGCTAC
E K M A N A N K G A M T E N A D E N A L Q S D A K G K L D S V A T
TGATTACGGT GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT GGTGATTTG CTGGCTCTAA TTCCCAAATG
D Y G A A I D G F I G D V S G L A N G N G A T G D F A G S N S Q M
GCTCAAGTCG GTGACGGTGA TAAATCACCT TTAATGAATA ATTTCCGTCA ATATTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT TTTGTCTTTA
A Q V G D G D N S P L M N N F R Q Y L P S L P Q S V E C R P F V F S
                    AATAAA
GGCGTGGTAA ACCATATGAA TTTTCTATTG ATTTGACAAA GATTAACTTA TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTAATG
A G K P Y E F S I D C D K I N L F R G V F A F L L Y V A T F M Y V
ATTTTCTACG TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAACCGTGGG GTGAACGCTC GGTGCGGTTTT TATGTTCTTT TGGAGGATCC
(SEQ ID NO: 21)
F S T F A N I L R N K E S * (SEQ ID NO: 22)

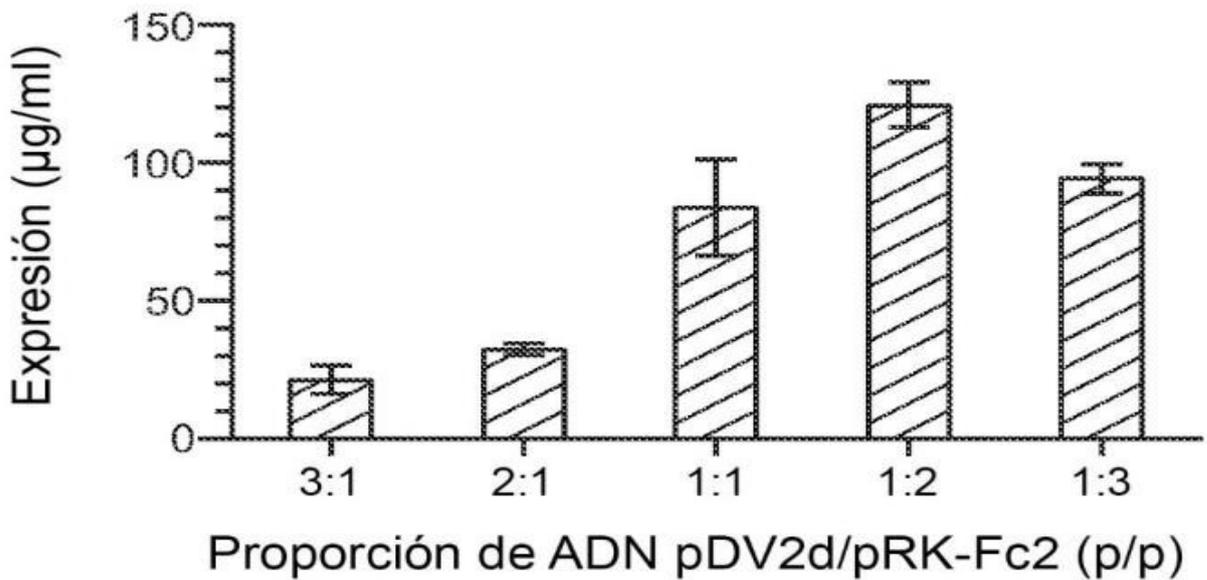
```

FIG. 2B





**FIG. 4A**



**FIG. 4B**

FIG. 5

FIG. 5A

FIG. 5B

GCCACCATGA TGA A A T T T A C C G T G G T G G G G C G G G C G C T G C T G C T G C T G G G C G G T A A G T A C C G C C T A T A G A G T C T A T A G G C C A C C C C C C T T G G C T T A C G T A <PhoA  
 M M K F T V V A A A L L L L L G (SEQ ID NO: 25)  
 ^ Secuencia señal de la cadena ligera (del gen BiP murino)

AAGAAGTTAT TGAAGCATCC TCGTCAGTAA AAAGTTAATC TTTTCAACAG CTGTCATAAA GTTGTACGG CCGAGACTTA TAGTCGCTTT GTTTTATTT >

TTTTATGTAT TTGTAAGTAG TACGCAAGTT CACGTAAAAA GGGTATGTAG AGGTTGAGGT GATTTTATGA AAAAGAATAT CGCATTTCCTT CTTGCATCTA  
 M K K N I A F L L A S M  
 < Secuencia señal de stII modificada

G PPT A G T 3'ss BsiWI  
 TGTTCCTTTT TCTCTTCC ACAGCCGTAC GCGCAGATAT CCAGATGACC CAGTCCCGA GCTCCCTGTC CGCCTCTGTG GCGATAGGG TCACCATCAC (SEQ ID NO: 2)  
 F L F S L S T (A) V R A D I Q M T Q S P S S L S A S V G D R V T I T  
 V I A < Región variable de la cadena ligera (SEQ ID NO: 9)

. [Cadena ligera, SV40 poli(A), promotor CMV]

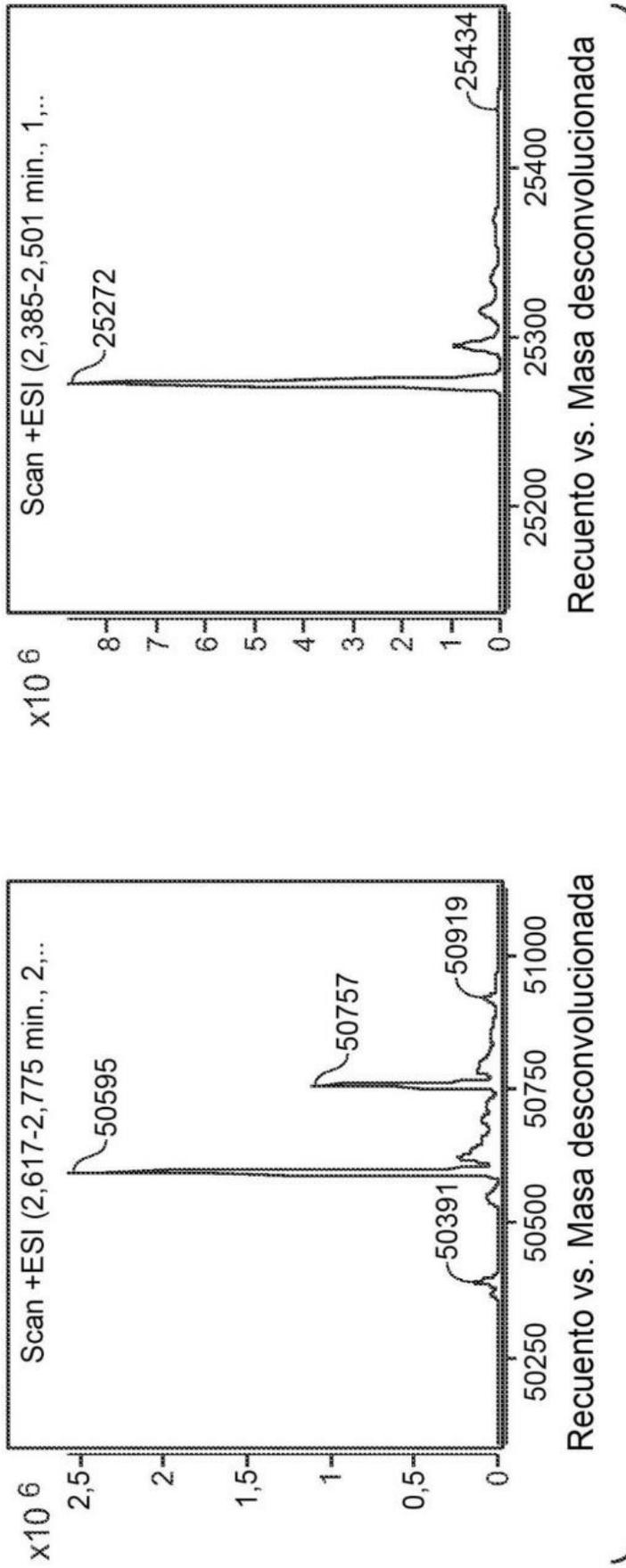
GCCACCATGG GATGGTCATG TATCATCCTT TTTCTAGTAG CAACTCGAAC AGGTAAGTAC CGCCTATAGA GTCTATAGGC CCACCCCTT GGCTTACGTA <PhoA  
 M G W S C I I L F L V A T A T (SEQ ID NO: 19)  
 ^ Secuencia señal de la cadena pesada (del gen VH murino)

AAGAAGTTAT TGAAGCATCC TCGTCAGTAA AAAGTTAATC TTTTCAACAG CTGTCATAAA GTTGTACGG CCGAGACTTA TAGTCGCTTT GTTTTATTT >

TTTTATCATC GTGGCCCTAG TACGCAAAAT CACGTAAAAA GGGTAACTAG AGGTTGAGGT GATTTTATGA AAAAGAATAT CGCATTTCCTT CTTGCATCTA  
 M K K N I A F L L A S M  
 < Secuencia señal de stII modificada

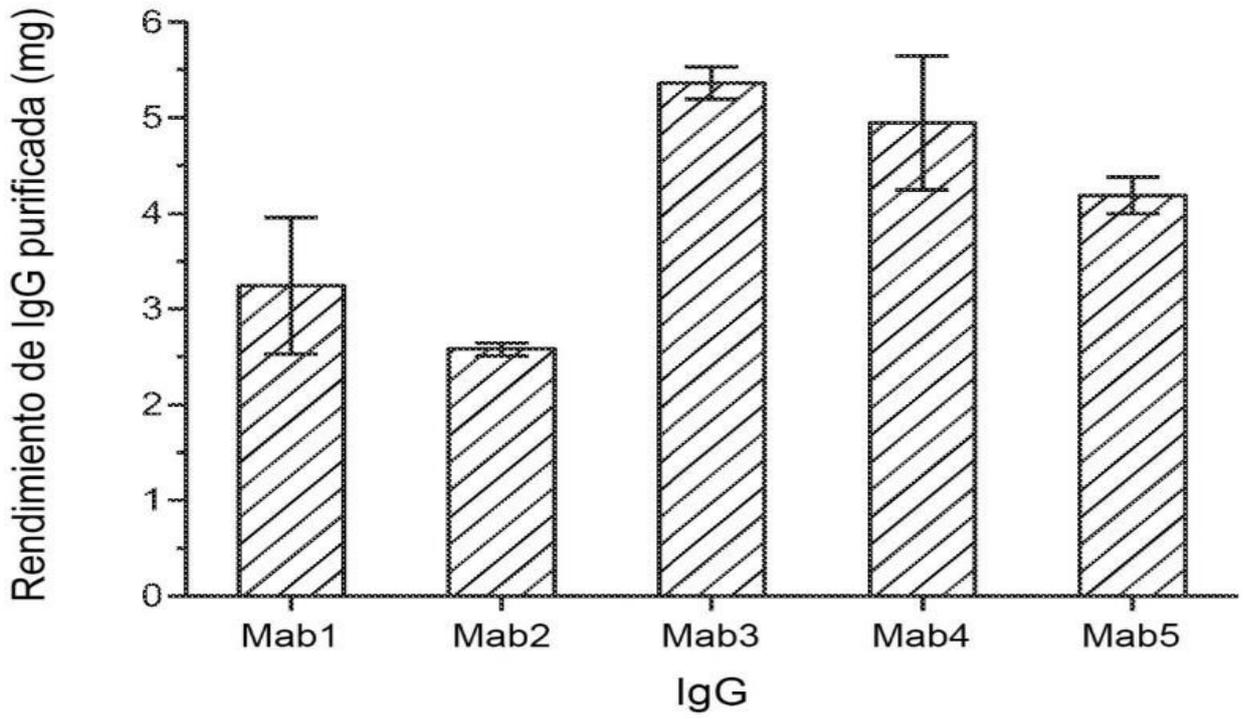




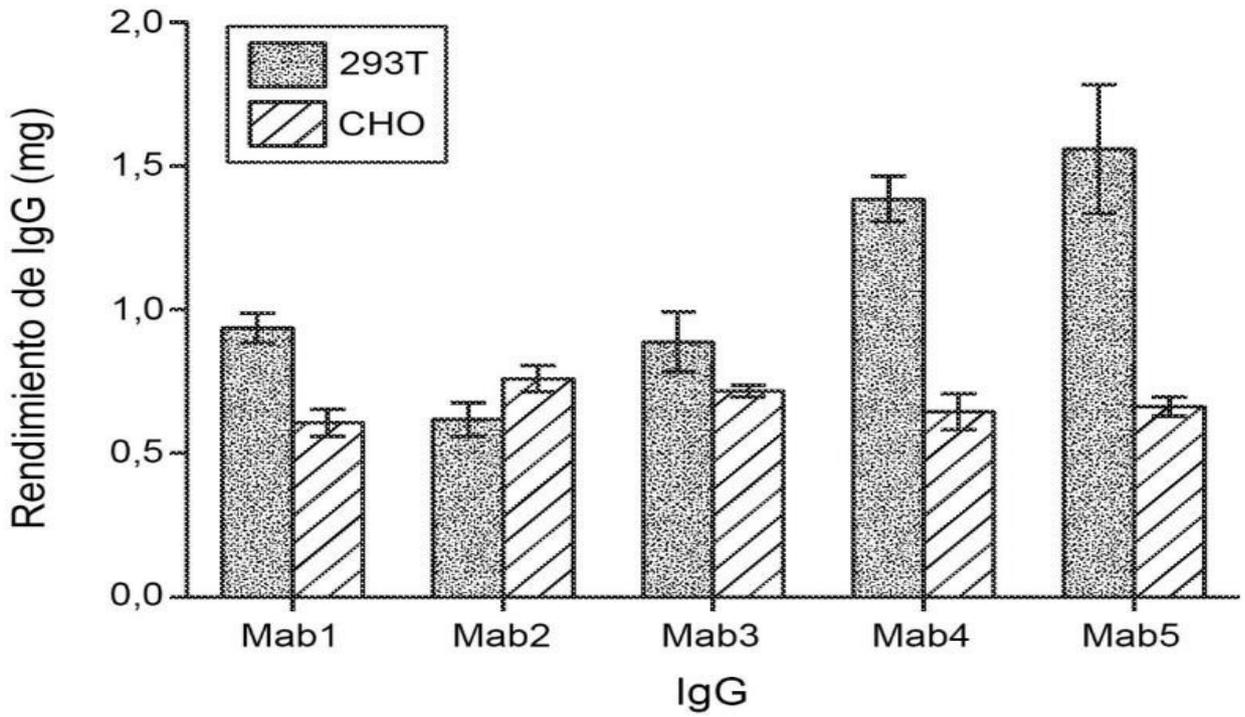


	Masa esperada	Masa observada
Cadena pesada, principales formas glucosiladas	50595 Da	50595 Da
	50757 Da	50757 Da
Cadena ligera	25275 Da	25272 Da

FIG. 7B

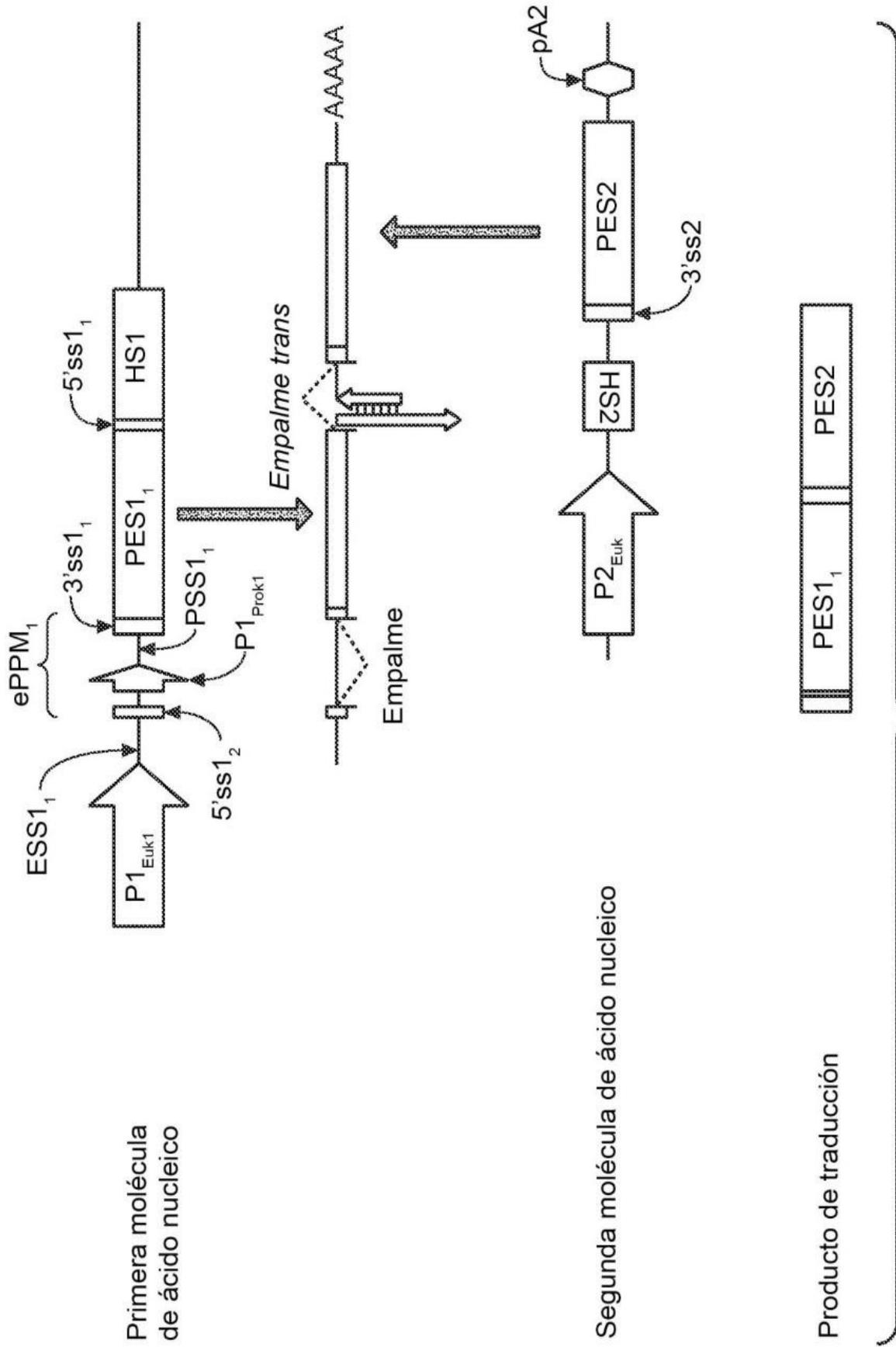


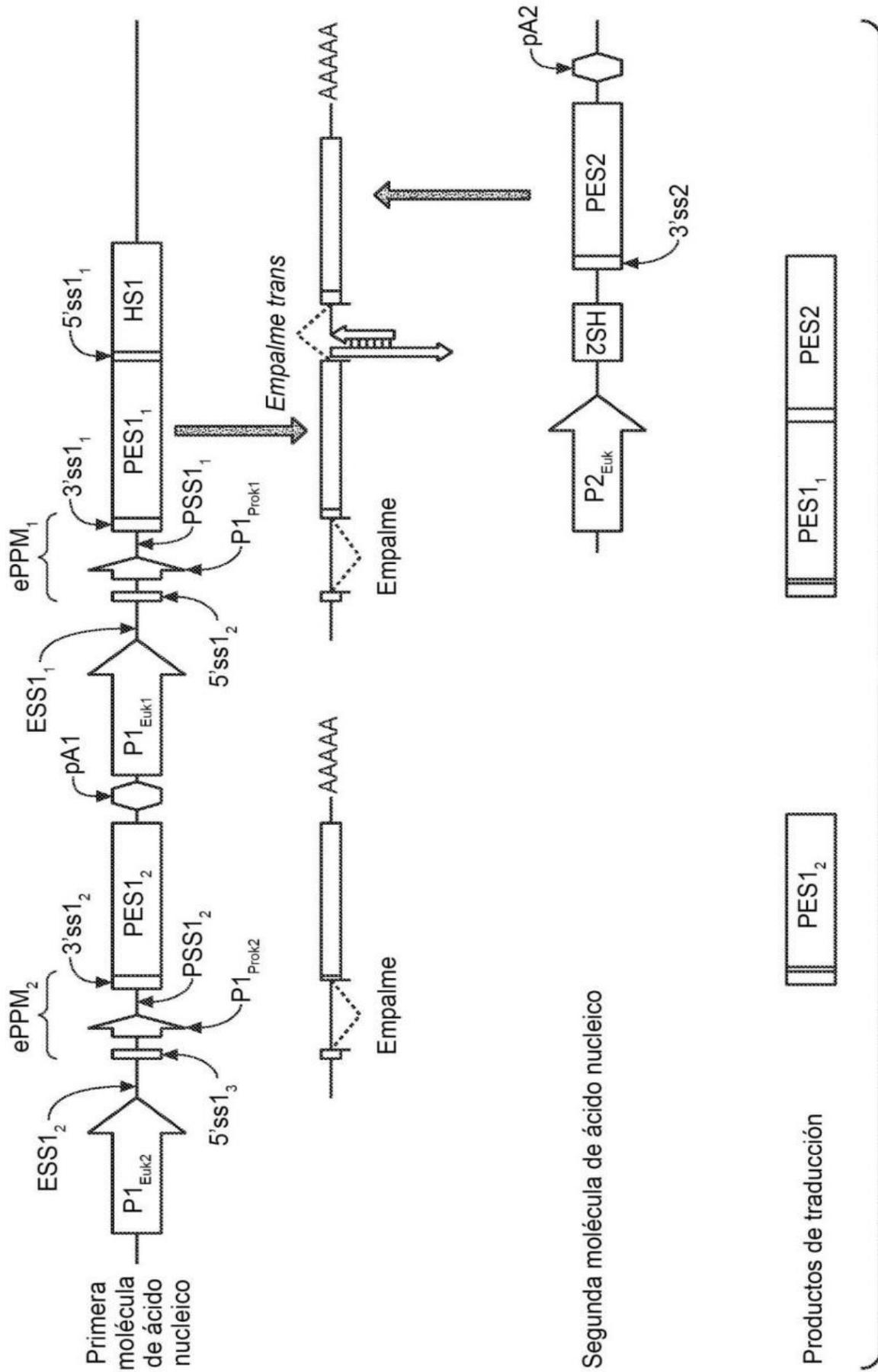
**FIG. 8A**



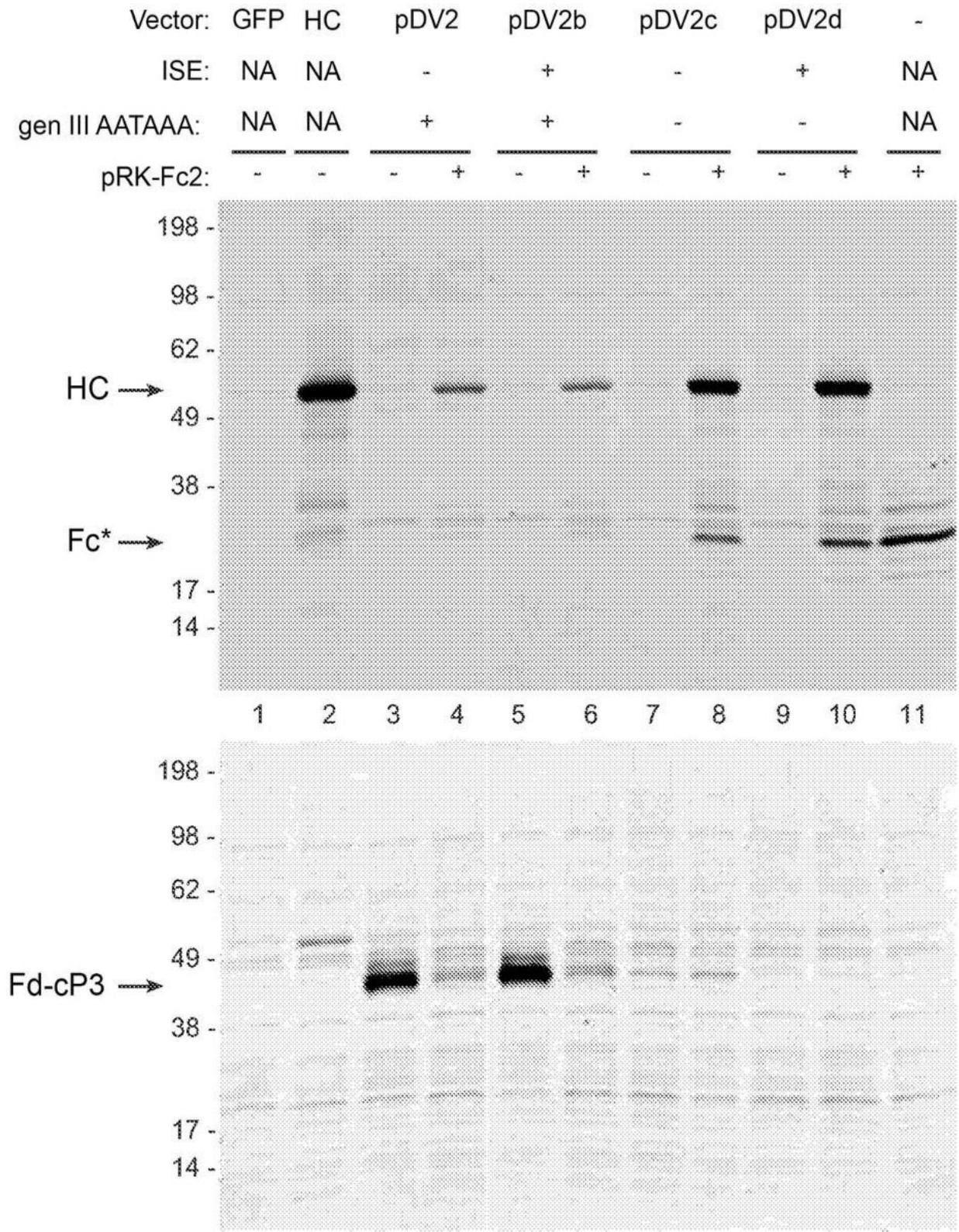
**FIG. 8B**



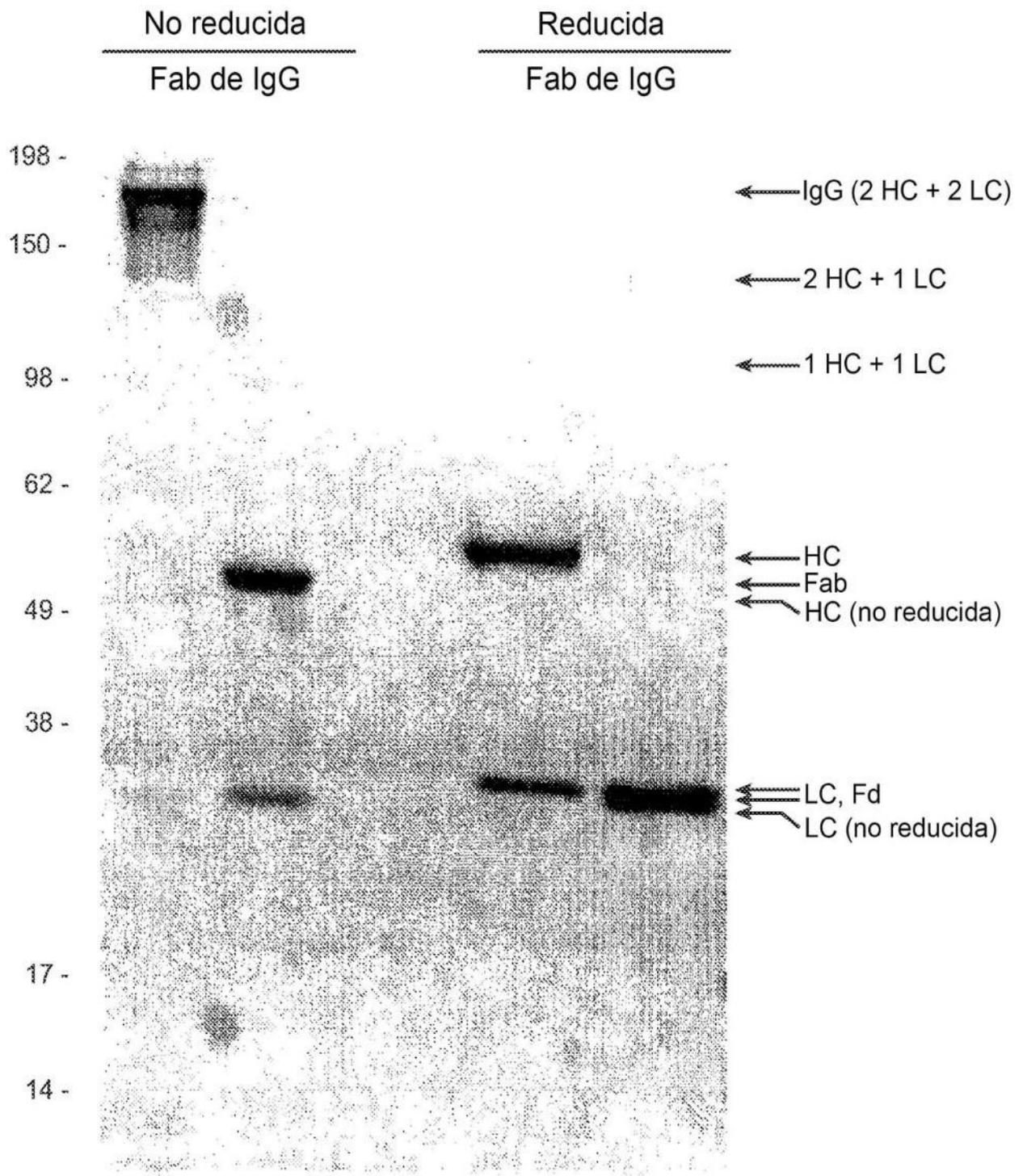




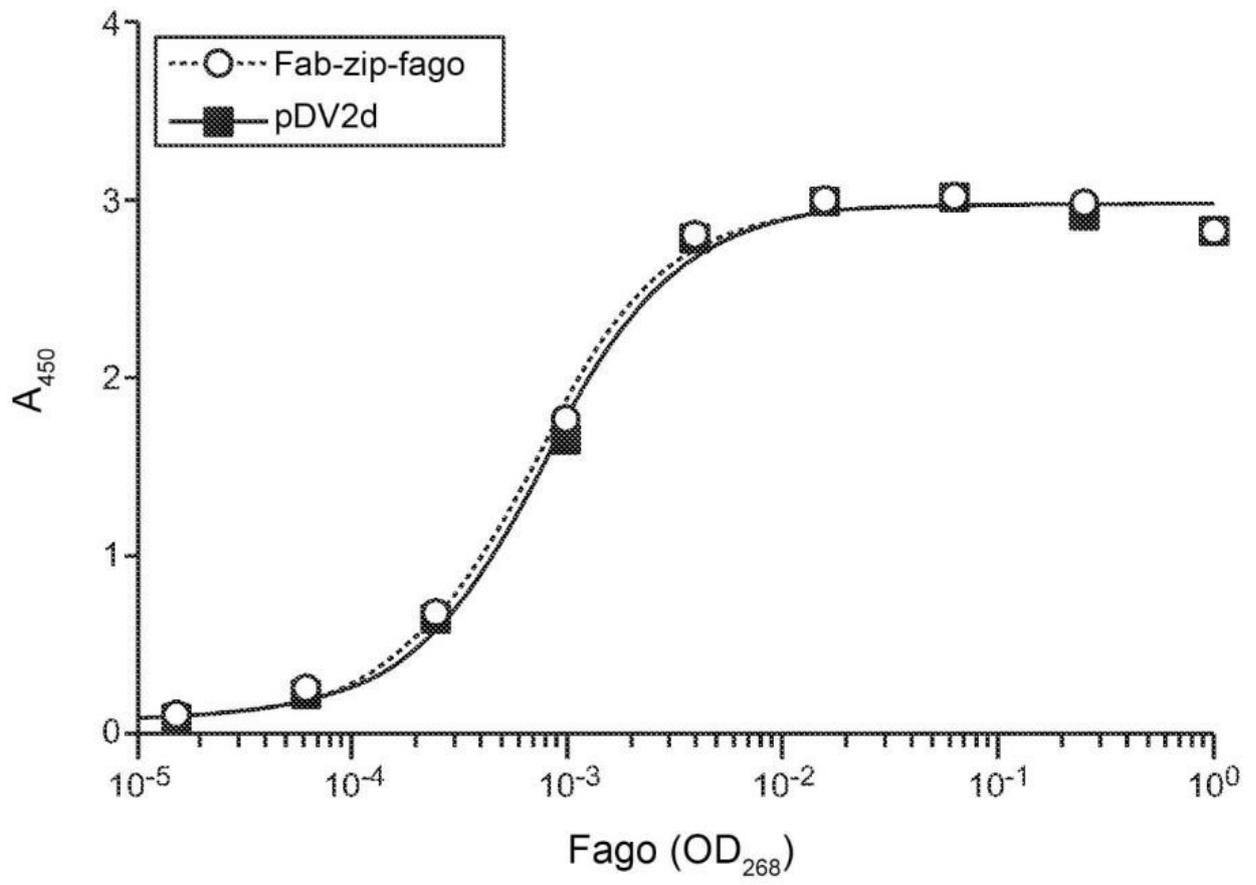
**FIG. 11**



**FIG. 12**



**FIG. 13**

**FIG. 14**