

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 384 973**

51 Int. Cl.:  
**C12N 5/02** (2006.01)  
**C07H 21/04** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **04711259 .4**  
96 Fecha de presentación: **13.02.2004**  
97 Número de publicación de la solicitud: **1594955**  
97 Fecha de publicación de la solicitud: **16.11.2005**

54 Título: **Un casete y vector de expresión para la expresión transitoria o estable de moléculas exógenas**

30 Prioridad:  
**14.02.2003 US 448179 P**

45 Fecha de publicación de la mención BOPI:  
**16.07.2012**

45 Fecha de la publicación del folleto de la patente:  
**16.07.2012**

73 Titular/es:  
**BIAGEN IDEC MA INC.  
14 CAMBRIDGE CENTER  
CAMBRIDGE, MASSACHUSETTS 02142, US**

72 Inventor/es:  
**SISK, William P. y  
PRENTICE, Holly**

74 Agente/Representante:  
**Ponti Sales, Adelaida**

**ES 2 384 973 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

Un casete y vector de expresión para la expresión transitoria o estable de moléculas exógenas

### 5 CAMPO TÉCNICO

**[0001]** Esta invención se refiere a vectores de expresión y casetes de expresión, y más en particular a procedimientos, composición, y sistemas para la expresión de una molécula exógena en un organismo.

### 10 ANTECEDENTES

**[0002]** La introducción de polipéptidos moleculares, péptidos, y moléculas pequeñas de ácidos nucleicos en células y tejidos diana se utiliza tanto para sistemas de administración terapéuticos como en la producción de moléculas terapéuticas *in vitro*. La aplicabilidad de este enfoque se ha incrementado con una mejor comprensión de las células hospedadoras y la biología molecular de los mecanismos celulares de división, diferenciación y expresión. El documento de EE.UU. 6.194.176 describe procedimientos para la producción de polipéptidos heterólogos utilizando una variedad de líneas de células secretoras manipuladas genéticamente de manera recombinante.

### 20 RESUMEN

**[0003]** Se ha descubierto que la transcripción conducida por un promotor del CMV y terminada por un dominio de poliA a partir de una variante del gen de la hormona del crecimiento humana (hGHv) es más eficiente que otros vectores de expresión que carecen de uno de los dos o de ambos elementos. Por tanto, la descripción, que incluye la presente invención, proporciona un casete de expresión y un vector de expresión útiles en la expresión de polinucleótidos de interés. El casete de expresión de la descripción, que incluye el casete de expresión de la invención, comprende una combinación de elementos reguladores que proporcionan una transcripción eficiente, una terminación de la transcripción eficiente y una estabilidad mejorada del ARNm de los productos transcritos. En la terminación de la transcripción, y en un caso de la descripción, el casete de expresión incluye un promotor/potenciador del citomegalovirus humano, un sitio de clonación o polinucleótido de interés, y un dominio de la señal de poliadenilación del hGHv. Opcionalmente, puede estar presente una secuencia interviniente de longitud variable.

**[0004]** La invención proporciona un casete de expresión que comprende una región promotora/potenciadora inmediata temprana 1 del CMV humano (hCMV IE1); un polinucleótido de interés; un dominio de la señal de poliA de la hormona del crecimiento humana; y una secuencia interviniente de longitud variable (VLIVS) que comprende un sitio donador de corte y empalme y un sitio aceptor de corte y empalme, en donde el dominio de la señal de poliA tiene una longitud de al menos 100 nucleótidos, contiene la secuencia AATAAA, y es idéntica en al menos un 92% a la secuencia expuesta en la SEQ ID NO: 18; y en donde la VLIVS comprende un intrón A de un gen hCMV IE1.

**[0005]** La invención proporciona adicionalmente un vector de expresión que incluye un casete de expresión de la invención así como células hospedadoras que contienen un casete de expresión o un vector de expresión de la invención.

**[0006]** La descripción también incluye un procedimiento de administración de un agente de interés *in vivo*. El procedimiento incluye la administración a un sujeto de una composición que comprende un casete de expresión. El casete de expresión incluye una región promotora/potenciadora hCMV IE1, una secuencia interviniente de longitud variable que comprende un sitio donador de corte y empalme y un sitio aceptor de corte y empalme; un polinucleótido que codifica el agente de interés; y un dominio de la señal de poliA de la hormona del crecimiento humana (hGH) o una de sus variantes.

**[0007]** La descripción además incluye un sistema de expresión. El sistema de expresión incluye una célula hospedadora transfectada o transformada con un casete de expresión de la invención, en donde las células hospedadoras se cultivan en condiciones para la expresión el polinucleótido de interés; y la recuperación del agente de interés.

**[0008]** La presente invención y sus formas de realización se presentan en las reivindicaciones anexas. Los detalles de la invención se exponen en los dibujos acompañantes y la descripción siguiente. Otras características, objetos, y ventajas de la invención serán evidentes a partir de la descripción y los dibujos, y de las reivindicaciones.

### DESCRIPCIÓN DE LOS DIBUJOS

**[0009]** La Figura 1 es el mapa plasmídico del vector pV10. El promotor inmediato temprano 1 de citomegalovirus (CMV IE1)/intrón (IVS) se generó por PCR y se clonó en los sitios HindIII y BamHI. También se indica el gen de resistencia a ampicilina,  $\beta$  lactamasa (*bla*).

**[0010]** La Figura 2 es el mapa plasmídico del vector pV40. Se indica el promotor inmediato temprano 1 de citomegalovirus (CMV IE1), un intrón (IVS), un dominio de la señal de poliadenilación (poliA) de hGHv de una longitud de aproximadamente 600 pares de bases, y el gen de resistencia a ampicilina,  $\beta$  lactamasa (*bla*).

5 **[0011]** La Figura 3 es una representación esquemática del plásmido pV70. Se indican el promotor CMV IE1, la IVS que incluye la unión de supresión y los sitios donador de corte y empalme (SD) y aceptor de corte y empalme (SA), un dominio de la señal de poliA de la hGHv, y el gen *bla*. La unión de eliminación representa el resultado de la ligación de los extremos romos de BspB1'/HpaI.

10 **[0012]** La Figura 4 es una representación esquemática de la generación del vector pXLC.1. El vector pXLC.1 se construyó utilizando el vector de expresión pV70 y un producto de la PCR que contiene la secuencia codificante de la cadena ligera. pV70 se linealizó con BamHI, el producto de la PCR se digirió con BamHI y estos dos fragmentos se ligaron para generar el vector pXLC.1.

15 **[0013]** La Figura 5 es una representación esquemática de la generación del vector pXLC.2.

**[0014]** La Figura 6 es una representación esquemática de un vector pV60 y la generación del vector pXHC. El vector pXHC se construyó utilizando el vector de expresión pV60 y un producto de la PCR que contiene la mayor parte de la secuencia codificante de la cadena pesada (no se han incluido 52 aminoácidos a partir del extremo N-terminal). pV60 se linealizó con BamHI, el producto de la PCR se digirió con BamHI y estos dos fragmentos se ligaron.

20

**[0015]** La Figura 7 es una representación esquemática de un vector pXHC.1.

25 **[0016]** La Figura 8 es una representación esquemática de un vector pXHC.3.

**[0017]** La Figura 9 es una representación esquemática de un vector pXHC.5, resultado de la adición de un casete *dhfr* a pXHC.3. Los elementos de control del casete de expresión se obtuvieron en pSI (Promega, N° de acceso del Genbank #U47121) e incluyen el promotor/potenciador de SV40, un intrón artificial y la secuencia de poliadenilación tardía de SV40.

30

**[0018]** La Figura 10 es una representación esquemática del vector pV80. Se indican un fragmento del promotor inmediato temprano 1 de citomegalovirus (CMV IE1)/intrón (IVS) que incluye los sitios donador de corte y empalme (SD) y aceptor de corte y empalme (SA), un dominio de la señal de poliadenilación (poliA) de hGHv, el gen de resistencia a ampicilina,  $\beta$  lactamasa (*bla*), el intrón artificial del promotor/potenciador de SV40 y la secuencia de poliadenilación tardía de SV40.

35

**[0019]** Las Figuras 11A y 11B son representaciones esquemáticas del vector pV90 y las secuencias anotadas correspondientes. La Figura 11A es un mapa del vector. Se indican un fragmento del promotor inmediato temprano 1 de citomegalovirus (CMV IE1)/intrón (IVS) que incluye los sitios donador de corte y empalme (SD) y aceptor de corte y empalme (SA), un dominio de la señal de poliadenilación (poliA) de hGHv, el gen de resistencia a ampicilina,  $\beta$  lactamasa (*bla*), el intrón artificial y la secuencia de poliadenilación tardía de SV40. El vector carece del sitio NotI en el casete de expresión *dhfr*. La Figura 11B es la secuencia anotada del vector pV90. En la Fig. 11B, la secuencia a partir de los nucleótidos 1275 a 1866 de la SEQ ID NO: 19 representa una señal de poliA de la hGHv de una longitud de aproximadamente 600 nucleótidos.

40

45

**[0020]** La Figura 12 es un mapa del vector pSI-DBFR.2. El promotor de SV40 dirige el gen *dhfr*.

**[0021]** La Figura 13 es una gráfica que representa las productividades específicas relativas de las fracciones reunidas de muestras transfectadas. Se analizaron tres fracciones reunidas de muestras para pCMV-hGHvEA-SEAP y pSEAP2, y uno para pUC18.

50

**[0022]** La Figura 14 es un par de gráficas que muestran las productividades específicas relativas de las muestras aisladas. Las 20 primeras muestras aisladas procedentes de cada construcción se muestran en orden de importancia.

55

**[0023]** La Figura 15 muestra una secuencia con el N° de acceso del GeoBank K00470 (SEQ ID NO: 18).

#### DESCRIPCIÓN DETALLADA

60

**[0024]** La descripción, que incluye la presente invención, se refiere a un casete y vectores de expresión que contienen el casete de expresión. El casete de expresión incluye una región reguladora de la transcripción capaz de llevar a cabo la transcripción en un hospedador eucariota y una región de terminación de la transcripción. Los casetes y el vector de expresión de la invención proporcionan una iniciación y una detención fuertes de la transcripción, así como una estabilidad mejorada del ARNm de los productos transcritos.

65

**[0025]** La descripción, que incluye la presente invención, proporciona un promotor y opcionalmente elementos potenciadores procedentes de cualquier cepa de citomegalovirus, tales como las descritas en el presente documento o en referencias como las de la patente de EE.UU. N° 5.658.759. Por ejemplo, regiones promotoras inmediatas tempranas de CMV adecuadas útiles en casetes de expresión de la invención se pueden obtener a partir del vector de expresión de la  $\beta$ -galactosidasa promovido por CMV, CMV $\beta$  (MacGregor y col., Nucl. Acids Res. 17:236S (1989)).

**[0026]** Como se describe en profundidad en el presente documento, el dominio de la señal de poliadenilación de hGHv proporciona una señal fuerte de detención de la transcripción al mismo tiempo que incrementa la estabilidad del transcrito del ARNm. El elemento regulador/de expresión puede estar separado del dominio de la señal de poliadenilación de hGHv mediante, por ejemplo, un polinucleótido de interés o un sitio de clonación (por ejemplo, un sitio de clonación múltiple).

**[0027]** En un aspecto de la descripción se proporciona un polinucleótido que comprende un casete de expresión que incluye una región reguladora de la transcripción de citomegalovirus (CMV), una secuencia interviniente de longitud variable (por ejemplo, procedente del intrón A de CMV), un polinucleótido de interés, y un dominio de la señal de poliadenilación. La descripción, que incluye la presente invención, se refiere adicionalmente a procesos y vectores de expresión para la producción y recuperación de polipéptidos heterólogos a partir de células hospedadoras.

**[0028]** En otro aspecto de la descripción, un casete de expresión incluye, unido de manera operable, (i) una región promotora/potenciadora inmediata temprana 1 (IE1) principal de CMV y una secuencia interviniente de longitud variable (por ejemplo, derivada del intrón A), (ii) un polinucleótido de interés, y (iii) un dominio de la señal de poliadenilación de hGHv. El término "unido de manera operable" se refiere a una yuxtaposición en la que los componentes se encuentran en una relación que les permite funcionar de la manera prevista (por ejemplo, unidos funcionalmente). Así, por ejemplo, un promotor/potenciador unido de manera operable a un polinucleótido de interés está ligado a este último de tal forma que la expresión del polinucleótido de interés se consigue en condiciones que son compatibles con la activación de la expresión a partir del promotor/potenciador.

**[0029]** En una forma de realización específica de la invención, el casete de expresión incluye una secuencia como se expone en la SEQ ID NO: 1 desde el nucleótido 1 aproximadamente al nucleótido 1867 aproximadamente (por ejemplo, desde 1 aproximadamente a 1865, 1866, 1867, 1868, ó 1869). El casete de expresión expuesto desde el nucleótido 1 a 1867 de la SEQ ID NO: 1 incluye una serie de dominios distintos tales como una región promotora/potenciadora IE1 de CMV que presenta una secuencia como la expuesta desde  $x_1$  aproximadamente a  $x_2$  aproximadamente de la SEQ ID NO: 1, en donde  $x_1$  es un nucleótido entre 1-20 y  $x_2$  es un nucleótido entre 715-720 aproximadamente (por ejemplo, entre 1 y 719 aproximadamente de la SEQ ID NO: 1). Otro dominio del casete de expresión incluye una secuencia interviniente de longitud variable (VLIVS) que contiene un sitio donador de corte y empalme y un sitio aceptor de corte y empalme. La VLIVS puede tener una longitud de al menos 50 pb (por ejemplo, una longitud de al menos 100, 150, 200, ó 250 pb) y puede incluir donadores y aceptores de corte y empalme procedentes de cualquier fuente conocida en la materia. Véase, por ejemplo, Varani y col., Annu Rev Biophys Biomol Struct 27:407-45 (1998) y Koning, Eur J Biochem 219:25-42 (1994). Un dominio interviniente adecuado puede incluir todos los intrones A del genoma de CMV de cualquier cepa o puede incluir un fragmento más pequeño que comprenda una secuencia 5' que contenga un sitio donador de corte y empalme ligado a una secuencia 3' que contenga un sitio aceptor de corte y empalme. Por ejemplo, la VLIVS incluye nucleótidos desde  $x_3$  aproximadamente a  $x_4$  aproximadamente de la SEQ ID NO: 1, en donde  $x_3$  es un nucleótido entre 715-720 y  $x_4$  es un nucleótido entre 1236-1254 (por ejemplo, de 719 a 1236 de la SEQ ID NO: 1). La secuencia interviniente a continuación del promotor/potenciador IE1 de CMV puede variar de tamaño en hasta 317 nucleótidos con respecto a la presente en la SEQ ID NO: 1. Por ejemplo, se suprimieron 317 nucleótidos de la secuencia IVS como se representa en pV40 y pV70 (véanse, por ejemplo, las Figuras 2 y 3, respectivamente) para producir la VLIVS de la SEQ ID NO: 1. Así, en otro aspecto de la invención el casete de expresión incluye una secuencia entre el nucleótido 1 aproximadamente y el nucleótido 1254 aproximadamente de la SEQ ID NO: 1 (por ejemplo, un promotor/potenciador IE1 de CMV y una secuencia interviniente). Después de la región IVS (es decir en dirección 3') puede haber presente un sitio de clonación múltiple (por ejemplo, los nucleótidos 1255-1272 de la SEQ ID NO: 1 incluye los sitios BamHI y un sitio NotI). Mediante ingeniería genética se pueden añadir sitios de restricción adicionales o diferentes en el casete de expresión utilizando técnicas conocidas por aquellos expertos en la materia. El casete de expresión además incluye un dominio de poliA.

**[0030]** El dominio de la señal de poliA procede de un gen hGHv, que puede variar en su secuencia 3'UTR, por ejemplo, entre alelos. Un alelo del gen hGHv está descrito en el GenBank con el N° de acceso K00470 (SEQ ID NO: 18), mientras que otra secuencia está descrita en la Fig. 11B como SEQ ID NO: 19, que corresponde a los nucleótidos 2032 a 2625 de la SEQ ID NO: 18 (véase Figura 15). Se pueden preparar variantes no naturales del dominio de la señal de poliA mediante técnicas de mutagénesis, incluyendo las aplicadas a polinucleótidos, células u organismos. Una variante de poliA de un gen hGHv incluye un dominio de la señal de poliA que es diferente de un dominio de la señal de poliA de la hGHv de tipo silvestre pero que retiene la capacidad para señalar la detención de la transcripción y/o estabilizar el ARNm. Por ejemplo, el dominio de la señal de poliadenilación puede incluir una secuencia del dominio de la señal de poliadenilación de hGHv como se expone en las SEQ ID NOs: 18 ó 19. La persona experta en materia de biología molecular también comprenderá que no es necesario que las secuencias

alcancen una longitud de 600 nucleótidos aproximadamente. Más bien, se incluye cualquier dominio de la secuencia de poliA que comprenda una secuencia de nucleótidos contigua de al menos 100 nt (por ejemplo, de al menos 200, 300, 400, 500, ó 600 nt), incluyendo el sitio canónico AATAAA, de un gen hGHv. La invención engloba secuencias que difieren de las secuencias anteriores en hasta un 8% (por ejemplo, presentan una identidad del 92% a la SEQ ID NO: 18 ó 19 o un dominio distinto de las mismas). Por ejemplo, un polinucleótido con una longitud de 100 nucleótidos y una identidad del 95% a los nucleótidos 1-1867 de la SEQ ID NO: 1 y que incluya la secuencia AATAAA retendría la capacidad para terminar la transcripción.

**[0031]** En otro aspecto de la descripción, que incluye la presente invención, se proporciona un vector que comprende un casete de expresión. Como se usa en el presente documento, un "vector" es una molécula de ácidos nucleicos (ADN o ARN) capaz de replicarse autónomamente tras su introducción en una célula receptora (por ejemplo, una bacteria como *E. coli*). Plásmidos, virus y bacteriófagos son ejemplos de vectores. El proceso de expresión a partir de un vector de expresión es muy conocido, e incluye la utilización de enzimas y procesos celulares para producir un producto de expresión a partir de un polinucleótido de interés. Los vectores de expresión son vectores capaces de mediar en la expresión de un polinucleótido clonado en una célula hospedadora, que puede o puede no ser del mismo tipo de célula utilizada para la replicación o propagación del vector. Muchos vectores de expresión en mamíferos se pueden propagar en bacterias corrientes (célula receptora), pero pueden expresar el polinucleótido de interés en células de mamífero (célula hospedadora) y no en bacterias.

**[0032]** La descripción, que incluye la presente invención, se refiere al diseño y utilización de vectores que son capaces de permitir una transcripción e introducción eficientes de polinucleótidos en células eucariotas (por ejemplo, células de mamífero, y más en particular células humanas, murinas, de simio, bovinas, porcinas, de roedor, u ovinas). Los vectores de la descripción incluyen un casete de expresión como se ha expuesto anteriormente que comprende un dominio de la señal de poliadenilación que proporciona una terminación de la transcripción eficiente y estabilidad al ARNm.

**[0033]** Los vectores de la descripción incluyen: un sitio de clonación para acoger un polinucleótido de interés; elementos reguladores de la transcripción (por ejemplo, regiones promotoras/potenciadoras IE1 de CMV) suficientes para permitir la transcripción de un polinucleótido insertado en el sitio de clonación en una célula hospedadora; elementos de traducción suficientes para permitir la traducción de un transcrito de ARN de dicho polinucleótido en una célula hospedadora y (si se desea) elementos de replicación suficientes para permitir la replicación de dicho vector en una célula hospedadora u otra célula receptora utilizada para la propagación del vector. Los vectores de la descripción, que incluyen los vectores de la invención, son capaces de mediar en dicha expresión de manera transitoria o estable en las células hospedadoras.

**[0034]** En una forma de realización específica, un vector de la invención incluye (1) una secuencia como la expuesta en la SEQ ID NO: 1; (2) una secuencia que es complementaria a la secuencia expuesta en la SEQ ID NO: 1; (3) una secuencia que es al menos un 80% (preferentemente al menos un 90%; 95%; 98% ó 99%) idéntica a la SEQ ID NO: 1 o a su complementario; o (4) un vector que comprende la SEQ ID NO: 1 entre el nucleótido 1 aproximadamente y el nucleótido 1867 aproximadamente y que comprende un polinucleótido de interés y/o un marcador seleccionable.

**[0035]** El vector que comprende la SEQ ID NO: 1 presenta una serie de dominios y regiones codificantes diferentes. Por ejemplo, en el vector está presente una región promotora/potenciadora IE1 de CMV que presenta una secuencia como la expuesta desde  $x_1$  aproximadamente a  $x_2$  aproximadamente de la SEQ ID NO: 1, en donde  $x_1$  es un nucleótido entre 1-20 y  $x_2$  es un nucleótido entre 715-720 aproximadamente (por ejemplo, entre 1 y 719 aproximadamente de la SEQ ID NO: 1). Otro dominio de un casete de expresión de la invención incluye una secuencia interviniente de longitud variable (VLIVS) que contiene un sitio donador de corte y empalme y un sitio aceptor de corte y empalme. La VLIVS incluye nucleótidos de  $x_3$  aproximadamente a  $x_4$  aproximadamente de la SEQ ID NO: 1, en donde  $x_3$  es un nucleótido entre 715-720 y  $x_4$  incluye un nucleótido entre 1236-1254 (por ejemplo, nucleótidos de 719 a 1236 aproximadamente de la SEQ ID NO: 1). Un sitio de clonación múltiple del vector de expresión incluye los nucleótidos 1255-1272 de la SEQ ID NO: 1 (por ejemplo, los sitios BamHI y un sitio NotI). Mediante ingeniería genética se pueden añadir sitios de restricción adicionales o diferentes en el vector de expresión utilizando técnicas conocidas por aquellos expertos en la materia. El vector de expresión además incluye un dominio de poliA. El dominio de la señal de poliA es un dominio de la señal de poliA de la hGHv u otra variante del dominio de la señal de poliA de la hGHv. Por ejemplo, un dominio de la señal de poliA incluye una secuencia del dominio de la señal de poliA de la hGHv como se expone en la SEQ ID NO: 19. En un vector de la invención también hay presentes uno o más marcadores seleccionables. Por ejemplo, la SEQ ID NO: 1 incluye un gen de la dihidrofolato reductasa (*dhfr*) (por ejemplo, entre el nucleótido 2568 aproximadamente y el nucleótido 3132 aproximadamente de la SEQ ID NO: 1). Un vector de la invención puede incluir elementos promotores/potenciadores y regiones reguladoras adicionales (por ejemplo, dominios de poliadenilación) además de los proporcionados anteriormente. Dichos elementos reguladores y dominios de poliadenilación adicionales pueden flanquear (por ejemplo, estar inmediatamente adyacentes a, en 5' y 3' de) un marcador seleccionable o polinucleótido de interés. Por ejemplo, el vector que comprende la SEQ ID NO: 1 contiene un gen de la dihidrofolato reductasa (*dhfr*) entre el nucleótido 2568 aproximadamente y el nucleótido 3132 aproximadamente de la SEQ ID NO: 1. El gen *dhfr* está flanqueado por un elemento promotor/potenciador de SV40 y una región de poliadenilación de SV40 (por ejemplo, entre el nucleótido

1868 aproximadamente y el nucleótido 2210 aproximadamente y el nucleótido 3144 aproximadamente y el nucleótido 3440 aproximadamente de la SEQ ID NO: 1, respectivamente).

**[0036]** Ejemplos específicos de marcadores seleccionables son aquellos que codifican proteínas que confieren resistencia a fármacos citostáticos o citocidas, tales como la proteína DHFR, que confiere resistencia a metotrexato (Wigler y col., Proc. Natl. Acad. Sci. USA 77:3567 (1980); O'Hare y col., Proc. Natl. Acad. Sci. USA 78:1527 (1981)); la proteína GPF, que confiere resistencia al ácido micofenólico (Mulligan & Berg, Proc. Natl. Acad. Sci. USA 78:2072 (1981)), el marcador de resistencia a neomicina, que confiere resistencia al aminoglicósido G-418 (Colberre-Garapin y col., J. Mol. Biol. 150:1 (1981)); la proteína Hygro, que confiere resistencia a higromicina (Santerre y col., Gene 30:147 (1984)); y el marcador de resistencia Zeocin™ (disponible comercialmente en Invitrogen). Además, la timidina quinasa del herpesvirus simple (Wigler y col., Cell 11:223 (1977)), la hipoxantina-guanina fosforribosiltransferasa (Szybalska & Szybalski, Proc. Natl. Acad. Sci. USA 48:2026 (1962)), y la adenina fosforribosiltransferasa (Lowy y col., Cell 22:817 (1980)) se pueden emplear en células tk<sup>-</sup>, hgprt<sup>-</sup> o aprt<sup>-</sup>, respectivamente. Otros marcadores seleccionables codifican la puromicina N-acetiltransferasa o adenosina desaminasa.

**[0037]** Los términos "idéntico" o porcentaje "de identidad", en el contexto de dos o más moléculas de ácidos nucleicos, se refieren a dos o más secuencias o subsecuencias que son iguales o presentan un porcentaje especificado de nucleótidos que son iguales, cuando se comparan y se alinean para su máxima correspondencia a lo largo de una ventana de comparación, medido utilizando un algoritmo de comparación o alineamiento manual e inspección visual. Esta definición también se refiere al complementario de una secuencia (por ejemplo, el complementario de una secuencia como la expuesta en la SEQ ID NO: 1 o uno de sus fragmentos que comprende un casete de expresión). Por ejemplo, el casete de expresión y sus fragmentos incluyen aquellos con una identidad de secuencia de nucleótidos que es al menos el 80% aproximadamente, el 90% aproximadamente, y el 95% aproximadamente, 97% aproximadamente, 98% aproximadamente o 99% aproximadamente idéntica a una porción de la SEQ ID NO: 1 (por ejemplo, los nucleótidos 1-719, 1-1254, y similares, de la SEQ ID NO: 1). Así, si una secuencia presenta la identidad de secuencia requerida para la secuencia completa de la SEQ ID NO: 1 o uno de sus dominios, entonces también puede funcionar como casete de expresión o dominio de la invención, respectivamente.

**[0038]** Para la comparación de secuencias, normalmente una secuencia actúa como secuencia de referencia, con la que se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, se designan las coordenadas de las subsecuencias, si fuera necesario, y se designan los parámetros del algoritmo de secuencia del programa. Se pueden utilizar los parámetros por defecto del programa, o se pueden designar parámetros alternativos. A continuación el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la secuencia(s) de prueba en relación con la secuencia de referencia, basándose en los parámetros del programa designados o por defecto. Una "ventana de comparación", como se usa en el presente documento, incluye la referencia a un segmento de una cualquiera de una serie de posiciones contiguas seleccionadas del grupo constituido por entre 25 y 600, normalmente entre 50 aproximadamente y 200 aproximadamente, más habitualmente entre 100 aproximadamente y 150 aproximadamente en la que una secuencia se puede comparar con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias se hayan alineado de manera óptima. Los procedimientos de alineamiento de secuencias para su comparación son muy conocidos en la materia. El alineamiento óptimo de secuencias para su comparación se puede llevar a cabo, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), mediante el algoritmo de alineamiento por homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), mediante la búsqueda con el procedimiento de similitud de Pearson & Lipman, Proc. Natl. Acad. Sci. USA 85:2444 (1988), mediante implementaciones computerizadas de estos algoritmos (GAP, PILEUP, BESTFIT, FASTA, y TFASTA del paquete de *software* de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante alineamiento manual e inspección visual.

**[0039]** Un ejemplo de algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia (es decir, similitud o identidad sustancial) es el algoritmo BLAST, descrito en Altschul, J. Mol. Biol. 215:403-410, 1990. El *software* para llevar a cabo los análisis BLAST está disponible para el público en el Centro nacional para la información biotecnológica (en Internet en la dirección [ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/)). Este algoritmo supone identificar en primer lugar los pares de secuencia con la puntuación más elevada (HSPs) identificando palabras cortas de longitud W en la secuencia de búsqueda, que cumple o satisface unas puntuaciones umbral T con un valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. "T" se denomina umbral de puntuación de palabras vecinas. Estas coincidencias iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSPs más largos que las contengan. Estas coincidencias en la palabra a continuación se extienden en ambas direcciones a lo largo de cada secuencia tanto como se pueda incrementar la puntuación de alineamiento acumulativo. Las puntuaciones acumulativas se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos coincidentes; siempre > 0) y N (puntuación de penalización para restos no coincidentes, siempre < 0). La extensión de las coincidencias en la palabra en cada dirección se detienen cuando: la puntuación de alineamiento acumulativo decae en una cantidad X de su valor máximo conseguido; la puntuación acumulativa alcanza cero o inferior, debido a la acumulación de uno o

- más alineamientos de restos con puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST, W, T, y X, determinan la sensibilidad y velocidad del alineamiento. En una forma de realización, para determinar si una secuencia de ácidos nucleicos está dentro del alcance de la invención, se utiliza el programa BLASTN (para secuencias de nucleótidos) incorporando como parámetros por defecto una longitud de la palabra (W) de 11, una expectación (E) de 10, M=5, N=4, y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP utiliza como parámetros por defecto una longitud de la palabra (W) de 3, una expectación (E) de 10, y la matriz de puntuación BLOSUM62 (véase, por ejemplo, Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989).
- 10 **[0040]** El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin, Proc. Natl. Acad. Sci. USA 90:5873-5787, 1993). Una medida de la similitud proporcionada por el algoritmo BLAST es la suma de probabilidades más pequeña (P(N)), que proporciona una indicación de la probabilidad de que se pueda producir por azar una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la suma de
- 15 probabilidades más baja en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es inferior a 0,1 aproximadamente, más preferentemente inferior a 0,01 aproximadamente, y lo más preferentemente inferior a 0,001 aproximadamente.
- [0041]** En la invención también se incluyen polinucleótidos que se hibriden específicamente con una secuencia de polinucleótidos como la expuesta en la SEQ ID NO: 1 entre el nucleótido 1 y 1867 aproximadamente o uno de sus fragmentos. La frase "se híbrida selectiva (o específicamente)" se refiere a la unión, la formación de un dúplex, o hibridación de una molécula a un polinucleótido de referencia particular en condiciones de hibridación rigurosas. La frase "condiciones de hibridación rigurosas" se refiere a condiciones en las que una sonda se hibridará principalmente con su subsecuencia diana, normalmente en una mezcla compleja de ácidos nucleicos, pero no a
- 20 otras secuencias. Las condiciones rigurosas dependen de la secuencia y serán diferentes en diferentes circunstancias, por ejemplo, dependiendo de la longitud de la sonda. Secuencias más largas se hibridan específicamente a temperaturas superiores. Una guía exhaustiva para la hibridación de ácidos nucleicos se encuentra en Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generalmente, las
- 25 condiciones rigurosas se seleccionan para que estén aproximadamente 5-10°C por debajo del punto de fusión térmico (T<sub>m</sub>) para la secuencia específica a una fuerza iónica y un pH definidos. La T<sub>m</sub> es la temperatura (en condiciones definidas de fuerza iónica, pH, y concentración de ácidos nucleicos) a la que el 50% de las sondas complementarias a la diana se hibridan con la secuencia diana en equilibrio (puesto que las secuencias diana están presentes en exceso, a T<sub>m</sub>, el 50% de las sondas están ocupadas en equilibrio). Las condiciones rigurosas serán
- 30 aquellas en las que la concentración salina sea inferior a una concentración de iones sodio de 1,0 M aproximadamente, normalmente una concentración de iones sodio (u otras sales) de 0,01 a 1,0 M aproximadamente a un pH de pH 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30°C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos aproximadamente) y de al menos aproximadamente 60°C para sondas largas (por ejemplo, mayores a 50 nucleótidos aproximadamente). Las condiciones rigurosas también se pueden conseguir con la adición
- 35 de agentes desestabilizantes tales como la formamida. Para la hibridación selectiva o específica, una señal positiva (por ejemplo, la identificación de un ácido nucleico de la invención) es aproximadamente 2 veces la hibridación de fondo. Las condiciones de hibridación "rigurosas" que se utilizan para identificar ácidos nucleicos sustancialmente idénticos dentro del alcance de la invención incluyen la hibridación en un tampón que comprende el 50% de formamida, 5x SSC, y el 1% de SDS a una temperatura comprendida entre 42°C y 65°C, con un lavado de 0,2x SSC
- 40 y el 0,1% de SDS a 65°C, para sondas largas. No obstante, como será evidente para alguien con conocimientos ordinarios en la materia, las condiciones de hibridación se pueden modificar dependiendo de la composición de la secuencia. "Condiciones de hibridación moderadamente rigurosas" ejemplares incluyen una hibridación en un tampón del 40% de formamida, NaCl 1 M, y el 1% de SDS a 37°C, y un lavado en 1X SSC a 45°C. Una hibridación positiva es al menos dos veces la hibridación de fondo. Aquellos con conocimientos ordinarios reconocerán
- 45 fácilmente que se pueden utilizar condiciones alternativas de hibridación y de lavado para proporcionar condiciones de rigurosidad similar. Así, un casete de expresión de la invención puede incluir un dominio de la señal de poliA de la hGHv que se hibride en condiciones de alta rigurosidad a un ADN de cadena sencilla que contenga la secuencia de nucleótidos de las SEQ ID NO: 18 ó 19.
- 50 **[0042]** El casete de expresión se puede utilizar en forma de construcción de ácidos nucleicos desnudos. Alternativamente, el casete de expresión se puede introducir como parte de un vector de ácidos nucleicos (por ejemplo, un vector de expresión como los descritos anteriormente). Dichos vectores incluyen plásmidos y vectores víricos. Un vector puede incluir secuencias que flanquean el casete de expresión que incluyen secuencias homólogas a secuencias genómicas eucariotas, tales como secuencias genómicas de mamífero, o secuencias genómicas víricas. Esto permitirá la introducción del casete de expresión en el genoma de células eucariotas o de virus mediante recombinación homóloga. Por ejemplo, se puede utilizar un vector plasmídico que comprende el casete de expresión flanqueado por secuencias víricas para preparar un vector vírico adecuado para la introducción del casete de expresión en un vertebrado, incluyendo células de peces, aves o mamíferos. Las técnicas empleadas son muy conocidas por la persona experta.
- 55 **[0043]** Se pretende que el término "polinucleótido de interés" cubra moléculas de ácidos nucleicos que, al menos,
- 60

se puedan traducir. La molécula puede estar en orientación codificante o no codificante con respecto al promotor. Las construcciones en dirección no codificante se pueden utilizar para inhibir la expresión de un gen en una célula según técnicas muy conocidas. El polinucleótido de interés puede incluir un polinucleótido heterólogo. El término polinucleótido heterólogo engloba cualquier gen. Un polinucleótido heterólogo normalmente tiene su origen en una especie exógena, en comparación con el elemento regulador al cual está unido de manera operable en el casete o vector de expresión, o si tiene su origen en la misma fuente, es el gen modificado a partir de su forma original. Por tanto, un polinucleótido heterólogo unido de manera operable a un promotor procede de una fuente diferente de la que se obtiene el promotor o, si tiene su origen en la misma fuente, es el promotor modificado a partir de su forma original. La modificación de un polinucleótido heterólogo se puede producir, por ejemplo, tratando el ADN con una enzima de restricción para generar un fragmento de ADN que se pueda unir de manera operable al promotor. La mutagénesis dirigida de sitio también es útil para modificar un polinucleótido heterólogo. Los polinucleótidos heterólogos también pueden incluir genes marcadores (por ejemplo, que codifican para la  $\beta$ -galactosidasa o la proteína verde fluorescente) o genes cuyos productos regulan la expresión de otros genes. Así, los polinucleótidos que sirven como moldes para el ARNm, ARNt y ARNr están incluidos dentro de esta definición. El gen heterólogo también puede ser cualquier variante alélica de un gen de tipo silvestre, o puede ser un gen mutante. El ARNm opcionalmente incluirá parte o todas las regiones 5' y/o 3' flanqueantes transcritas pero sin traducir asociadas de manera natural, o de otra forma, a la secuencia codificante traducida.

**[0044]** El polinucleótido de interés además puede incluir opcionalmente los elementos de control de la transcripción asociados normalmente a las moléculas transcritas, por ejemplo señales de detención de la transcripción, dominios de poliadenilación y elementos potenciadores en dirección 3'. El polinucleótido de interés puede codificar o servir como molde para un producto terapéutico, que puede ser, por ejemplo, un péptido, polipéptido, proteína, o ácido ribonucleico. El polinucleótido de interés normalmente es una secuencia de ADN (como ADNc o ADN genómico) que codifica para un producto polipeptídico tal como enzimas (por ejemplo,  $\beta$ -galactosidasa); derivados sanguíneos; hormonas; citoquinas; interleuquinas; interferones; TNF; factores de crecimiento (por ejemplo, IGF-1); moléculas receptoras solubles (por ejemplo, moléculas receptoras del TNF solubles); neurotransmisores o sus precursores; factores tróficos tales como el BDNF, CNTF, NGF, IGF, GMF, aFGF, bFGF, NT3 y NT5; apolipoproteínas tales como ApoAI y ApoAIV; distrofina o una minidistrofina; proteínas supresoras de tumores tales como p53, Rb, Rap1A, DCC y k-rev; factores involucrados en la coagulación tales como los factores VII, VIII y IX; o alternativamente toda o una parte de una inmunoglobulina natural o artificial (por ejemplo, Fab y ScFv, o la cadena ligera o pesada de una IgG clonada).

**[0045]** Un polinucleótido de interés también puede incluir un molde para la generación de una molécula en dirección no codificante, cuya transcripción en una célula diana permite la expresión génica o la transcripción de ARNm celular a controlar. Dichas moléculas, por ejemplo, se pueden transcribir en una célula diana en ARN complementarios a los ARNm celulares y así pueden bloquear su traducción en proteínas, según técnicas conocidas en la materia. En particular, se pueden utilizar moléculas en dirección no codificante para bloquear la producción de citoquinas inflamatorias o catabólicas en el tratamiento de la artritis y pérdida tisular provocada por estas citoquinas.

**[0046]** La secuencia de polinucleótidos de interés normalmente codificarán un polipéptido para su utilización diagnóstica o terapéutica. El polipéptido se puede producir *in vitro* en biorreactores utilizando diversas células hospedadoras (por ejemplo, células COS o células CHO o sus derivados) que contienen el casete de expresión de la invención. Alternativamente, el casete y/o vector de expresión de la invención se puede utilizar para la introducción de genes, introducción de proteínas, y/o terapia génica.

**[0047]** La invención también se puede utilizar para la expresión de factores y polipéptidos tóxicos. Estos últimos, en particular, pueden ser venenos celulares (tales como la toxina diftérica, la toxina de pseudomonas y la ricina A), un producto que induzca sensibilidad a un agente externo (por ejemplo, timidina quinasa y citosina desaminasa) o alternativamente factores capaces de inducir la muerte celular (por ejemplo, Grb3-3 y ScFv anti-ras).

**[0048]** Por una utilización terapéutica se quiere decir una utilización que pueda proporcionar alivio de un trastorno o enfermedad, curación de un trastorno o enfermedad, y/o mejora de la gravedad de un trastorno o enfermedad. Una utilización diagnóstica incluye la utilización de moléculas capaces de determinar o proporcionar información con respecto a la causa o la relación de una molécula con un proceso patológico o la determinación de la presencia o ausencia de un trastorno o enfermedad. Un agente diagnóstico no contribuye directamente a la mejora del trastorno o enfermedad.

**[0049]** Un polinucleótido de interés también puede codificar un polipéptido antigénico para su utilización como vacuna. Los polipéptidos antigénicos o moléculas de ácidos nucleicos proceden de organismos patógenos tales como, por ejemplo, una bacteria o un virus. Por ejemplo, los polipéptidos antigénicos incluyen determinantes antigénicos presentes en los genomas o productos génicos de un organismo patógeno, por ejemplo, septicemia hemorrágica vírica, enfermedad renal bacteriana, vibriosis, y forunculosis. Los polipéptidos antigénicos se pueden seleccionar, por ejemplo, a partir de regiones del genoma y productos génicos del virus de la hepatitis C.

**[0050]** Como se usa en el presente documento, "aislado", cuando se refiere a una molécula o composición, tal como por ejemplo un vector o un casete de expresión de la invención, o un polinucleótido de interés, significa que la molécula o composición se encuentra separada de al menos otro compuesto, tal como una proteína, ADN, ARN, u



otros contaminantes con los que está asociada *in vivo* o en su estado natural. Así, un polinucleótido de interés se considera aislado cuando haya sido aislado de cualquier otro componente al cual está asociado de forma natural. No obstante, una composición aislada también puede ser sustancialmente pura. Una composición aislada puede estar en estado homogéneo. Puede estar en seco/liofilizada o en disolución acuosa. La pureza y homogeneidad se puede determinar, por ejemplo, utilizando técnicas de química analítica tales como, por ejemplo, electroforesis en gel de poliacrilamida (PAGE), electroforesis en gel de agarosa o cromatografía líquida de alta presión (HPLC).

**[0051]** Como se usan en el presente documento, los términos "molécula de ácidos nucleicos" y "polinucleótido" se utilizan de manera intercambiable, e incluyen oligonucleótidos (es decir, polinucleótidos cortos). También se refieren a moléculas de ácidos nucleicos sintéticas y/o que no son de origen natural (por ejemplo, que comprenden análogos de nucleótidos o restos o uniones modificados del esqueleto). Los términos también se refieren a oligonucleótidos desoxirribonucleótidos o ribonucleótidos en forma de cadena sencilla o de doble cadena. Los términos engloban ácidos nucleicos que contienen análogos de nucleótidos naturales. Los términos también engloban estructuras similares a los ácidos nucleicos con esqueletos sintéticos. Los análogos con esqueletos de ADN proporcionados por la invención incluyen fosfodiésteres, fosforotioato, fosforoditioato, metil-fosfonato, fosforamidato, alquilfosfotriéster, sulfamato, 3'-tioacetal, metileno (metilimino), 3'-N-carbamato, morfolino carbamato, y ácidos nucleicos peptídicos (PNA); véase *Oligonucleotides and Analogues, a Practical Approach*, editado por F. Eckstein, IRL Press en Oxford University Press (1991); *Antisense Strategies*, *Annals of the New York Academy of Sciences*, Volume 600, Eds. Baserga y Denhardt (NYAS 1992); Milligan (1993) *J. Med. Chem.* 36:1923-1937; *Antisense Research and Applications* (1993, CRC Press). Los PNA contienen esqueletos no iónicos, tales como unidades de N-(2-aminoetil) glicina. Las uniones de fosforotioato se describen en los documentos WO 97/03211; WO 96/39154; Mata (1997) *Toxicol. Appl. Pharmacol.* 144:189-197. Otros esqueletos sintéticos incluidos por el término comprenden uniones de metilfosfonato o uniones alternantes de metilfosfonato y fosfodiéster (Strauss-Soukup (1997) *Biochemistry* 36:8692-8698), y uniones de bencil-fosfonato (Samstag (1996) *Antisense Nucleic Acid Drug Dev* 6:153-156).

**[0052]** Como se usa en el presente documento, "recombinante" se refiere a un polinucleótido sintetizado o manipulado de otra forma *in vitro* (por ejemplo, "polinucleótido recombinante"), a procedimientos de utilización de polinucleótidos recombinantes para producir productos en células u otros sistemas biológicos, o a un polipéptido ("proteína recombinante") codificada por un polinucleótido recombinante. Los polinucleótidos recombinantes engloban moléculas de ácidos nucleicos procedentes de diferentes fuentes ligadas a un casete o vector de expresión para la expresión de, por ejemplo, una proteína de fusión; o aquellos producidos por la expresión inducible o constitutiva de un polipéptido (por ejemplo, un casete o vector de expresión de la invención unido de manera operable a un polinucleótido heterólogo, tal como una secuencia codificante de un polipéptido).

**[0053]** En un sistema de expresión típico, la producción de un polipéptido a partir de un polinucleótido heterólogo está no regulada o regulada mediante la modulación de la transcripción a partir de un promotor de la transcripción unido de manera operable en dirección 5' de un polinucleótido que codifica el polipéptido heterólogo. No obstante, la regulación también se debe producir adecuadamente en dirección 3' con el fin de proporcionar una terminación de la transcripción adecuada y estabilidad al ARNm. En un aspecto de la descripción, se proporciona un dominio de la señal de poliadenilación (poliA) de una variante de la hormona del crecimiento humana (hGHv) aguas abajo (3') de un polinucleótido de interés presente en un casete o vector de expresión de la descripción. El dominio de la señal de poliA de la hGHv incluye una secuencia derivada de la secuencia genética de la hormona del crecimiento humana. La secuencia del dominio de la señal de poliadenilación de la hGHv proporciona una terminación de la transcripción fuerte e incrementa la estabilidad del ARNm en células eucariotas. Este dominio de la señal de poliadenilación de la hGHv proporciona una ventaja característica sobre casetes y/o vectores de expresión previos que incluyen aquellos que puedan utilizar un promotor/potenciador de CMV.

**[0054]** También puede haber presentes elementos de traducción y está previsto que engloben las secuencias especializadas (tales como los sitios de unión a ribosomas y los codones de iniciación) que son necesarias para permitir la traducción de un transcrito de ARN en una proteína. Los elementos de traducción también pueden incluir secuencias consenso, secuencias líder, señales de corte y empalme, y similares, que sirven para facilitar o mejorar el alcance de la traducción, o incrementar la estabilidad del producto expresado. Por ejemplo, el dominio de la señal de poliadenilación de la hGHv incrementa la estabilidad del ARNm. Los vectores de la invención pueden poseer regiones anciliares de la transcripción, tales como intrones, señales de poliadenilación, señales de traducción de Shine/Dalgarno y secuencias consenso de Kozak (Shine y col., *Proc. Natl. Acad. Sci. (U.S.A.)* 71:1342-1346 (1974); Kozak, *Cell* 44:283-292 (1986)).

**[0055]** El término "elementos de replicación" está previsto que englobe las secuencias especializadas (tales como los orígenes de replicación) que son necesarias para permitir la replicación del vector en una célula receptora. En general, dichos vectores contendrán al menos un origen de replicación suficiente para permitir la replicación autónoma y estable del vector en una célula receptora.

**[0056]** Para facilitar la selección y el mantenimiento de un vector de la invención, en el vector se pueden incluir uno o más marcadores seleccionables (tales como polinucleótidos que confieran resistencia a antibióticos, o una capacidad celular para crecer en medio mínimo o en presencia de metabolitos tóxicos).

**[0057]** En una forma de realización adicional, la presente invención se refiere a células hospedadoras que contienen las construcciones descritas anteriormente (por ejemplo, el casete o vector de expresión de la invención). El casete de expresión de la invención se puede utilizar para modificar de manera recombinante una célula hospedadora, transfectando una célula hospedadora o transformando una célula hospedadora para que exprese un polinucleótido deseado de interés. Como se usa en el presente documento, el término "modificado de manera recombinante" significa la introducción de un casete o vector de expresión de la invención en una célula viva o en un sistema de expresión. Normalmente, el casete de expresión que comprende un polinucleótido de interés está presente en un vector (por ejemplo, un plásmido). Un sistema de expresión incluye una célula hospedadora viva en la que se haya introducido, como se describe en el presente documento, un polinucleótido de interés, cuyo producto se debe expresar.

**[0058]** Las células hospedadoras son células en las que se puede propagar un casete de expresión (incluyendo un vector que comprende un casete de expresión) y se pueden expresar polipéptidos que codifican productos. Una célula hospedadora también incluye cualquier progenie de la célula hospedadora objeto o de sus derivados. Se entiende que toda la progenie puede no ser idéntica a la célula parental puesto que puede haber mutaciones que se produzcan durante la replicación. No obstante, dicha progenie está incluida cuando se utiliza el término "célula hospedadora". Las células hospedadoras, que son útiles en la invención, incluyen células bacterianas, células fúngicas (por ejemplo, células de levadura), células de plantas y células animales. Por ejemplo, las células hospedadoras pueden ser de una célula eucariota superior, tal como una célula de mamífero, o de una célula eucariota inferior, tal como una célula de levadura, o la célula hospedadora puede ser de una célula procariota, tal como una célula bacteriana. La introducción de la construcción en la célula hospedadora se puede llevar a cabo mediante transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano o electroporación (Davis, L., Dibner, M., Battey, I., *Basic Methods in Molecular Biology* (1986)). Como ejemplos representativos de hospedadores adecuados, se pueden mencionar: células fúngicas, tales como levaduras; células de insectos tales como *Drosophila* *S2* y *Spodoptera Sf9*; células animales tales como CHO, COS o melanoma de Bowes; células de plantas, y similares. La selección de un hospedador adecuado se considera que está dentro de las competencias de aquellos expertos en la materia a partir de las enseñanzas del presente documento.

**[0059]** Las células hospedadoras para su utilización en la invención son células hospedadoras eucariotas (por ejemplo, células de mamífero). En un aspecto de la invención las células hospedadoras son células de producción de mamífero adaptadas por el crecimiento en cultivos celulares. Ejemplos de dichas células utilizadas habitualmente en la industria son las células CHO, VERO, BHK, HeLa, CV1 (incluyendo Cos; Cos-7), MDCK, 293, 3T3, C127, líneas celulares de mieloma (especialmente murinas), PC12 y W 138. Las células de ovario de hámster chino (CHO), que se han utilizado ampliamente para la producción de diversas proteínas recombinantes complejas, por ejemplo citoquinas, factores de coagulación, y anticuerpos (Brasel y col., *Blood* 88:2004-2012 (1996); Kaufman y col., *J. Biol Chem* 263: 6352-6362 (1988); McKinnon y col., *J Mol Endocrinol* 6:231-239 (1991); Wood y col., *J. Immunol* 145:3011-3016 (1990)). Las líneas celulares mutantes deficientes en dihidrofolato reductasa (DHFR) (Urlaub y col., *Proc Natl Acad Sci USA* 77:4216-4220 (1980)) son las líneas celulares hospedadoras CHO elegidas debido a que su eficiente sistema de expresión de genes seleccionables y amplificables de DHFR permite niveles elevados de expresión de proteínas recombinantes en estas células (Kaufman, *Meth Enzymol* 185:527-566 (1990)). Además, estas células son fáciles de manipular en forma de cultivos adherentes o en suspensión y presentan una estabilidad genética relativamente buena. Las células CHO y las proteínas recombinantes expresadas en ellas han sido caracterizadas exhaustivamente y aprobadas para su utilización en producción clínica por agencias reguladoras. Además, se contempla que se puedan utilizar células hospedadoras derivadas de cualquiera de las líneas celulares anteriores y con un fenotipo deseado. Por ejemplo, una célula hospedadora derivada incluye células CHO (por ejemplo, la línea celular DG44), que se ha cultivado selectivamente para un fenotipo deseado (por ejemplo, mediante procesos de selección positivos y/o negativos).

**[0060]** En un aspecto de la invención, se proporciona un sistema de expresión para la producción *in vitro* de un agente codificado por un polinucleótido de interés. Como se describe en el presente documento, el polinucleótido de interés puede codificar un polipéptido con un valor farmacéutico, médico, nutricional, y/o industrial. Por ejemplo, el polinucleótido de interés puede codificar un polipéptido con fundamento farmacológico. Normalmente dicho polipéptido se expresará en forma de producto extracelular. Por ejemplo, los polipéptidos que se pueden producir utilizando el casete y/o vector de expresión de la invención incluyen, pero no están limitados a, un ligando Flt3, un ligando CD40, eritropoyetina, trombopoyetina, calcitonina, ligando Fas, ligando para el activador del receptor de NF-kappa B (RANKL), ligando que induce la apoptosis relacionada con el TNF (TRAIL), ORK/Tek, linfopoyetina derivada de estroma tímico, factor estimulante de colonias de granulocitos, factor estimulante de colonias de granulocitos-macrófagos, factor de crecimiento de mastocitos, factor de crecimiento de blastocitos, factor de crecimiento epidérmico, RANTES, hormona del crecimiento, insulina, insulintropina, factores de crecimiento similares a la insulina, hormona paratiroidea, interferones (por ejemplo, interferón  $\beta$ ), factores de crecimiento nervioso, glucagón, interleuquinas 1 a 18, factores estimulantes de colonias, linfotóxina- $\beta$ , factor de necrosis tumoral, factor inhibidor de la leucemia, oncostatina-M, diversos ligandos para las moléculas de la superficie celular Elk y Hek (como los ligandos para las quinasas relacionadas con eph, o LERKS), y las cadenas ligera o pesada de los anticuerpos.

**[0061]** Los receptores para cualquiera de las proteínas anteriormente mencionadas también se pueden expresar utilizando los procedimientos y composiciones de la invención, incluyendo ambas formas del receptor del factor de

necrosis tumoral (denominadas p55 y p75), receptores de la interleuquina-1 (tipo 1 y 2), receptor de la interleuquina-4, receptor de la interleuquina-15, receptor de la interleuquina-17, receptor de la interleuquina-18, receptor del factor estimulante de colonias de granulocitos-macrófagos, receptor del factor estimulante de colonias de granulocitos, receptores para el factor inhibidor de la oncostatina-M y la leucemia, activador del receptor de NF-kappa B (RANK),  
 5 receptores para TRAIL, receptor del BAFF, receptor de la linfotóxina  $\beta$ , receptor del TGF $\beta$  tipos I y II, y receptores que incluyen los dominios de muerte, tales como Fas o al receptor que induce la apoptosis (AIR).

**[0062]** Otras proteínas que se pueden expresar utilizando el casete y/o los vectores de expresión de la invención incluyen antígenos de cúmulos de diferenciación (denominados proteínas CD), por ejemplo, los descritos en  
 10 Leukocyte Typing VI (Proceedings of the VIth International Workshop and Conference; Kishimoto, Kikutani y col., eds.; Kobe, Japón, 1996), o moléculas CD descritas en talleres posteriores. Ejemplos de dichas moléculas incluyen CD27, CD30, CD39, CD40, y sus ligandos (ligando de CD27, ligando de CD30 y ligando de CD40). Varios de estos son miembros de la familia de receptores TNF, que también incluye a 41BB y OX40; los ligandos a menudo son miembros de la familia TNF (como son el ligando 41BB y el ligando OX40); por consiguiente, los miembros de las  
 15 familias TNF y TNFR también se pueden expresar utilizando la invención.

**[0063]** Los polipéptidos que son enzimáticamente activos también se pueden expresar según la invención. Los ejemplos incluyen miembros de la familia de metaloproteinasas-desintegrinas, diversas quinasas, glucocerebrosidasa, superóxido dismutasa, activador del plasminógeno tisular, factor VIII, factor IX, apolipoproteína  
 20 E, apolipoproteína A-I, globinas, un antagonista de IL-2, anti-tripsina  $\alpha$ -1, enzima convertidora del TNF- $\alpha$  (TACE), y numerosas otras enzimas. Los ligandos para proteínas enzimáticamente activas también se pueden expresar utilizando el casete y vector de la invención.

**[0064]** Las composiciones y procedimientos de la invención también son útiles para la expresión de otros tipos de  
 25 proteínas y polipéptidos recombinantes, incluyendo moléculas de inmunoglobulina o fragmentos de las mismas y anticuerpos quiméricos (por ejemplo, un anticuerpo con una región constante humana acoplada a una región murina de unión al antígeno) o fragmentos de los mismos. Se conocen numerosas técnicas mediante las cuales ADN que codifican moléculas de inmunoglobulina se pueden manipular para dar ADN que codifican proteínas recombinantes tales como anticuerpos de cadena sencilla, anticuerpos con una afinidad mejorada, u otros polipéptidos basados en  
 30 anticuerpos (véase, por ejemplo, Larrick y col., *Biotechnology* 7:934-938 (1989); Reichmann y col., *Nature* 332:323-327 (1988); Roberts y col., *Nature* 328:731-734 (1987); Verhoeven y col., *Science* 239:1534-1536 (1988); Chaudhary y col., *Nature* 339:394-397 (1989)). Los anticuerpos humanizados clonados incluyen aquellos que se unen específicamente al receptor de la linfotóxina  $\beta$  e integrinas tales como VLA-1, VLA-4, y  $\alpha$ v $\beta$ 6. Dichos anticuerpos pueden ser agonistas o antagonistas.

**[0065]** Las diversas proteínas de fusión también se pueden expresar utilizando los procedimientos y composiciones de la invención. Ejemplos de dichas proteínas de fusión incluyen proteínas expresadas como fusión  
 40 con una porción de una molécula de inmunoglobulina, proteínas expresadas como proteínas de fusión con una resto de cremalleras, y nuevas proteínas polifuncionales como proteínas de fusión de una citoquina y un factor de crecimiento (por ejemplo, GM-CSF y IL-3, MGF y IL-3). Los documentos WO 93/08207 y WO 96/40918 describen la preparación de diversas formas oligoméricas solubles de una molécula denominada CD40L, incluyendo una proteína de fusión de inmunoglobulina y una proteína de fusión de cremallera, respectivamente; las técnicas descritas en dichos documentos son aplicables a otras proteínas.

**[0066]** Una vez expresado el polinucleótido de interés, el producto de expresión (por ejemplo, una proteína o polipéptido) se puede purificar utilizando técnicas habituales en la materia. Por ejemplo, cuando el polinucleótido de  
 45 interés codifica un polipéptido de fusión que comprende un marcador de purificación, el polipéptido se puede purificar utilizando anticuerpos que se unan específicamente al marcador. En un aspecto, un oligonucleótido que codifica una molécula marcador está ligado al extremo 5' o 3' de un polinucleótido de interés que codifica un polipéptido deseado; el oligonucleótido puede codificar una poliHis (tal como una hexaHis), u otro "marcador" tal como FLAG, HA (hemaglutinina del virus de la gripe) o myc para los que existen anticuerpos disponibles  
 50 comercialmente. Normalmente este marcador se fusiona al polipéptido tras la expresión del polipéptido, y puede servir como medio para la purificación por afinidad del polipéptido deseado a partir de la célula hospedadora. La purificación por afinidad se puede conseguir, por ejemplo, en cromatografía en columna utilizando anticuerpos contra el marcador como matriz de afinidad. Opcionalmente, el marcador se puede retirar del polipéptido purificado  
 55 posteriormente por diversos medios tales como escisión proteolítica.

**[0067]** El casete y los vectores de expresión de la invención se pueden utilizar para proporcionar una transferencia estable de un polinucleótido de interés en una célula hospedadora. Una transferencia estable significa que el  
 60 polinucleótido de interés se mantiene continuamente en el hospedador. El casete o vector de expresión de la invención también puede proporcionar una expresión transitoria de un polinucleótido de interés en una célula hospedadora. Las células hospedadoras transfectadas de manera transitoria pierden el ADN exógeno durante la replicación celular y el crecimiento.

**[0068]** Se puede utilizar un casete de expresión de la invención para introducir un agente terapéutico en un ser  
 65 humano o un animal que necesite tratamiento. Alternativamente, el casete de expresión de la invención se puede

utilizar para introducir un agente que codifica polipéptidos potencialmente inmunógenos *in vivo* con el fin de vacunar a un sujeto (por ejemplo, un ser humano), particularmente para la vacunación de animales domesticados incluyendo animales para el consumo como especies de peces, porcinas, equinas, bovinas, caninas, y felinas.

- 5 **[0069]** El casete de expresión de la invención se puede administrar directamente en forma de construcción de ácidos nucleicos desnudos, normalmente que comprende secuencias flanqueantes homólogas al genoma de la célula hospedadora. La captación de construcciones de ácidos nucleicos desnudos por células de vertebrados se potencia mediante diversas técnicas conocidas incluyendo la transformación biolística y la lipofección.
- 10 **[0070]** Alternativamente, el casete de expresión se puede administrar como parte de un vector, incluyendo un vector plasmídico o un vector vírico.

**[0071]** Normalmente el casete o vector de expresión se combina con un vehículo o diluyente farmacéuticamente aceptable para producir una composición farmacéutica. Los vehículos y diluyentes adecuados comprenden 15 soluciones salinas isotónicas incluyendo, por ejemplo, tampón fosfato salino. La composición que comprende el casete o vector de expresión se puede formular para diversos tipos de administración incluyendo, por ejemplo, administración por vía intramuscular.

**[0072]** Cuando la composición que comprende el casete o vector de expresión se utiliza en forma inyectable, 20 normalmente se mezcla con un vehículo que es farmacéuticamente aceptable para una formulación inyectable para su inyección directa en el sitio a tratar. El vehículo o diluyente farmacéuticamente aceptable puede estar, por ejemplo, en una disolución isotónica estéril. La composición que comprende el casete o vector de expresión también se puede formular en una forma oralmente activa.

25 **[0073]** La formulación definitiva utilizada puede ser determinada fácilmente por la persona experta y variará dependiendo de la naturaleza de la sustancia a administrar y la vía de administración.

**[0074]** La dosificación de la sustancia utilizada se puede ajustar según diversos parámetros, especialmente según la sustancia utilizada, la edad, el peso y condiciones del sujeto a tratar, el modo de administración utilizado y el 30 régimen clínico requerido. El facultativo será capaz de determinar la vía de administración y las dosificaciones necesarias para cualquier sujeto y dolencia particular.

## **EJEMPLOS**

### **35 Construcción de un vector pV10**

**[0075]** En la Figura 1 se indican detalles adicionales de la construcción de pV10. Se aisló ADN genómico a partir de fibroblastos diploides humanos infectados con la cepa AD 169 de citomegalovirus humano (ATCC No. VR-538) y se utilizó como molde para amplificar por PCR la región promotora/potenciadora inmediata temprana del gen 1 de 40 CMV (CMV IE1 P/E) (véase Figura 11 (B) (SEQ ID NO: 1) para los detalles de la 5'UTR del gen IE1 de CMV). El promotor se amplificó utilizando cebadores que contienen un sitio HindIII en el extremo 5'-terminal (ttt AAGCTT GACATTGATTATTGACTAG; SEQ ID NO: 2; sitio de restricción subrayado) y un sitio BamHI en el extremo 3'-terminal (tttt GGATCC CTGTCAAGGACGGTGACTGC; SEQ ID NO: 3; sitio de restricción subrayado). Los nucleótidos terminales "t" que preceden al sitio de restricción están incluidos en el diseño de los oligonucleótidos 45 para facilitar la digestión con enzimas de restricción y se eliminan en la etapa de clonación.

**[0076]** Todas las reacciones de PCR se realizaron en el termociclador DNA engine PTC-200 Pelier (MJ Research, Watertown, MA). El volumen de reacción total fue de 100 µl: tampón de la polimerasa Vent NEB 1X (KCl 10 mM, Tris 20 mM pH 8,8 a 25°C, (NH<sub>4</sub>)SO<sub>4</sub> 10 mM, MgSO<sub>4</sub> 2 mM, 0,1% de Triton X-100), dNTP 2,5 mM, 2 unidades de ADN 50 polimerasa Vent (New England Biolabs, Beverly, MA), 1 µg de cada cebador, 1 µg de ADN genómico aislado de las células infectadas por CMV como molde. Las condiciones de reacción fueron las siguientes: 99°C durante 1 minuto, 55°C durante 30 segundos, 75°C durante 1,5 minutos durante 15 ciclos. El fragmento resultante se digirió con las enzimas de restricción BamHI y HindIII (New England Biolabs) y se subclonó en el vector de clonación pUC19 digerido con BamHI y HindIII. Se determinó el análisis de la secuencia del inserto y era consistente con la secuencia 55 publicada de la región promotora/potenciadora IE1 de CMV clonada.

### **Construcción de un vector pV40**

**[0077]** La construcción de pV40 se indica en la Figura 2. El 3'UTR del gen hGHv que incluye la señal de poliA se 60 amplificó por PCR a partir de ADN genómico aislado de fibroblastos humanos. La hebra (+) del cebador 5' (TTTTGGATCCCTGCCCGGTGGCATCC; SEQ ID NO: 20) contenía un sitio de restricción BamHI terminal y la hebra (-) del cebador 3' contenía un sitio EcoRI terminal (TTTTGAATTCATGAGAGGACAGTGCCAAGC; SEQ ID NO: 21). Las condiciones de la PCR fueron las mismas que las descritas para la construcción de pV10. El fragmento resultante de la PCR se digirió con BamHI y EcoRI, se purificó en gel y se ligó en el vector pV10 digerido con BamHI 65 y EcoRI. El plásmido resultante, designado pV40, se verificó por análisis con enzimas de restricción. La secuenciación posterior indicaba que un pequeño número de diferencias en los nucleótidos entre la 3'UTR de este

gen hGHv (SEQ ID NO: 19) y la secuencia publicada del gen hGHv en el GenBank con el No. de Acceso K00470 (SEQ ID NO: 18). Al menos algunos de los cambios se deben a variaciones alélicas.

### Construcción de un vector pV70

5 **[0078]** El vector pV40 se digirió con BspE1 y HpaI para eliminar una sección de 317 nucleótidos de la región del intrón A (IVS) (véase, por ejemplo, las Figuras 2 y 3). El vector pV60 se generó por ligación de extremos romos en BspE1-HpaI de la región codificante de *dhfr* de pV40. El vector de expresión pV70 contiene la región promotora/potenciadora principal inmediata temprana del gen 1 de citomegalovirus humano (hCMV IE1) para regular la transcripción. También contiene el 5'UTR de hCMV IE1 y el intrón A, en donde el intrón contiene una supresión de 317 pares de bases. Para la terminación de la transcripción, el vector contiene la región de la variante de poliadenilación de la hormona del crecimiento humana (hGHv poli A), SEQ ID NO: 19. La construcción de pV40, pV60, y pV70 se detalla a continuación y se describe en profundidad en las figuras.

#### 15 A. Generación de pXLC.1

**[0079]** Un producto de PCR que contiene una secuencia codificante de la cadena ligera para un anticuerpo se digirió con BamHI y se clonó en un sitio BamHI único en el vector de expresión pV70 (Figura 3). La región codificante de la cadena ligera se insertó en un sitio BamHI único entre la secuencia 5'UTR en el extremo 3' del intrón hCMV IE1 y el extremo 5' de la región poliA de hGHv. Este plásmido fue designado pXLC.1. La Figura 4 muestra un esquema de la generación del vector pXLC.1.

#### B. La adición de un neo casete - pXLC.2

25 **[0080]** Un (neo) casete de expresión de neomicina transferasa se introdujo en pXLC.1 para actuar como marcador seleccionable (Figura 5). El neo casete se preparó como fragmento BamHI/EcoRI a partir de un plásmido disponible comercialmente denominado pGT-N28 (New England Biolabs, Catálogo # 307-28). En este plásmido, el neo gen está dirigido por el promotor de la fosfoglicerato quinasa (PGK) y termina en un sitio de poli-adenilación PGK. El extremo BamHI en el extremo 5' del neo casete se convirtió en un extremo NarI utilizando los oligos adaptadores siguientes:

EcoRI  
compatible con BamHI GATCGAT GAATTC GG (SEQ ID NO: 4)  
CTACTTAAGCCGC (SEQ ID NO: 5) compatible con NarI

35 **[0081]** Con estos enlazadores, también se añadió un nuevo sitio EcoRI. El adaptador se ligó en primer lugar al neo fragmento cortado con BamHI/EcoRI, y el fragmento NarI/EcoRI convertido se clonó en pXLC.1 digerido con NarI y EcoRI. De esta manera el neo casete de expresión se insertó en el extremo 3' de la secuencia de la cadena ligera del plásmido. Este plásmido fue designado como pXLC.2 (Figura 5).

#### 40 **Construcción del vector de expresión de la cadena pesada - pXHC.5**

##### A. RT-PCR de la cadena pesada

45 **[0082]** La cadena pesada se amplificó a partir de la reacción de RT-PCR utilizando el cebador 5' de la PCR *TTTTGGATCC ATG TACTGGGTGAAGCAG* (SEQ ID NO: 6), en donde la secuencia en cursiva es una región enlazadora añadida con un sitio BamHI, y las bases subrayadas corresponden a la segunda metionina en la secuencia codificante de la cadena pesada. El cebador 3' de la PCR utilizado, *GCCCGGATC CTC ATTTACCCGGAGACAG* (SEQ ID NO: 7), también contiene una región enlazadora añadida con un sitio BamHI (cursiva) y una secuencia que corresponde a la secuencia codificante de la cadena pesada que incluye el codón de terminación (subrayado). Se obtuvo el producto esperado de la PCR de 1268 pares de bases.

##### B. Construcción de pXHC

55 **[0083]** Debido a que el cebador 5' de la PCR utilizado se hibridó con el segundo codón ATG en la región codificante en lugar de con la región codificante de iniciación ATG contenida en el producto de la PCR, la región codificante de la cadena pesada estaba incompleta. El fragmento BamHI, que contiene la cadena pesada incompleta, se clonó en el plásmido pV60 (Figura 6). Este vector es idéntico a pV70 (descrito anteriormente), excepto por que contiene la región codificante *dhfr* en el sitio de la supresión en el intrón. La región codificante de la cadena pesada se insertó en un sitio BamHI único entre la secuencia 5' sin traducir en el extremo 3' del intrón hCMV IE1 y el extremo 5' de la región de poliA de la variante hGHv. Este plásmido, con la cadena pesada incompleta, se designó como pXHC.

##### C. Compleción de la secuencia codificante de la cadena pesada - pXHC.1

65 **[0084]** La región codificante que faltaba en la secuencia de la cadena pesada se insertó en pXHC para generar el

plásmido denominado pXHC.1 (Figura 7). Para ello, se generó un fragmento por PCR utilizando como molde un plásmido que contiene la secuencia codificante para el anticuerpo. El cebador 5' de la PCR utilizado fue:

PstI BamHI

5 ***TTTTCTGCAGTCACCGTCCTTGACACGGGATCCATGGACTGGACCTTGGAGGG*** (SEQ ID NO: 8). La secuencia en cursiva corresponde al extremo 5' de la secuencia pXHC del sitio BamHI y la secuencia en negrita corresponde a la secuencia de partida dos bases antes del codón de iniciación de la secuencia de cadena pesada. El cebador 3' utilizado fue CTGAGGAGACGGTGACCAGGGTCCCTTGGCCCC (SEQ ID NO: 9). Este cebador se hibrida con el extremo del primer exón, la región variable de la cadena pesada. Tal y como se esperaba se obtuvo un producto de PCR de 445 pares de bases, cortado con PstI y StuI. El fragmento PstI/StuI se clonó en pXHC cortado con PstI y StuI para dar pXHC.1.

#### D. Eliminación del *dhfr* intrónico - pXHC.3

15 **[0085]** pXHC.1 contenía un gen *dhfr* en el intrón hCMV IE 1. Debido a problemas potenciales con la amplificación encontrados con esta configuración, el gen *dhfr* se eliminó del intrón y se insertó un casete de expresión en 3' del casete de la cadena pesada (Figura 8). El primer paso fue la eliminación del gen *dhfr* del intrón. Esto se consiguió cortando la secuencia codificante de la cadena pesada de pXHC.1 como un fragmento PstI/EcoRI y su clonación en el plásmido pXLC.1 cortado con las mismas enzimas. Esta etapa de clonación simplemente intercambió la cadena ligera por la cadena pesada en el plásmido. El plásmido resultante era idéntico a pXHC.1 excepto por que el intrón que contiene el gen *dhfr* fue sustituido por un intrón con una supresión como se ha descrito anteriormente para pXLC.1. Este plásmido de la cadena pesada fue designado como pXHC.3 (Figura 8).

#### 25 E. Inserción del casete *dhfr* - pXHC.5

**[0086]** La segunda etapa en la alteración de la configuración de *dhfr* fue la inserción del casete de expresión de *dhfr* en 3' del casete de expresión de la cadena pesada (Figura 9). El casete de expresión de *dhfr* se obtuvo del plásmido pSI-DHFR sobre un fragmento BglII/BamHI e incluye un promotor temprano de SV40, el gen *dhfr* y una región poliA de SV40. Este fragmento se clonó en el sitio EcoRI localizado en el extremo 3' de la región poliA de hGHv de pXHC.3 utilizando los oligos adaptadores siguientes:

Sall

compatible con EcoRI AATTC **GTCCGAC** A (SEQ ID NO: 10)

35 GCAGCTGTCTAG (SEQ ID NO: 11) compatible con BamHI/BglII

Este adaptador se ligó en primer lugar al plásmido cortado con EcoRI y a continuación el casete de *dhfr* BglII/BamHI se ligó al plásmido adaptado. Este plásmido fue designado como pXHC.5 (Figura 9).

#### 40 Caracterización de pXLC.2 y pXHC.5

**[0087]** Ambos plásmidos se analizaron utilizando enzimas de restricción para confirmar la presencia y orientación de los fragmentos insertados. Además, la región de codificación de los plásmidos se secuenció para verificar que no se habían acumulado mutaciones durante la PCR o las etapas de clonación.

45 **[0088]** La presencia de un neo marcador de selección funcional se confirmó mediante la transfección de pXLC.2 en células CHO y que demuestran resistencia a G418. La capacidad para llevar a cabo una selección doble se demostró cuando los plásmidos pXLC.2 y pXHC.5 se co-transfectaron en un hospedador CHO DG44 adaptado libre de suero. Las colonias crecieron a partir de la co-transfección en un medio de selección doble (a - MEM 10% al dFBS con 400 mg/ml de G418). En las mismas condiciones, cualquiera de las dos selecciones solas fue capaz de matar a las células sin transfectar (a - MEM o 400 mg/ml de G418).

#### Construcción de los vectores: pV80 y pV90

55 **[0089]** pV80 se generó a partir del vector de expresión de la cadena pesada pXHC.5 (Figura 10). La secuencia codificante de la cadena pesada se eliminó de pXHC.5 utilizando BamHI. El esqueleto se ligó para recircularizarlo en el sitio BamHI.

60 **[0090]** Se introdujeron dos alteraciones en el vector pV80 para generar pV90 (Figura 11). Se destruyó el sitio NotI que se encuentra en la construcción pV80 en posición 3166 (al final de la región codificante *dhfr*, véase secuencia adjunta). Para conseguirlo, el plásmido se digirió con NotI y los flecos se "rellenaron" usando la polimerasa Klenow. Como resultado, el plásmido religado había perdido el sitio NotI en posición 3166. A continuación se creó un nuevo sitio NotI en el sitio de clonación por digestión del vector con BamHI y la introducción de un enlazador NotI con la hibridación del siguiente 14-mero consigo mismo: GATCCGCGGCCGCG (SEQ ID NO: 12). Cuando se hibridan  
65 juntos, la secuencia enlazadora es:

NotI

compatible con BamHI GATCC **GCGGCCGC** (SEQ ID NO: 13)  
CGCCGCGCCTAGG (SEQ ID NO: 14) compatible con BamHI

5 **[0091]** Esta etapa de clonación recrea los sitios BamHI a cada lado del sitio NotI. Estos sitios BamHI pueden ser útiles en el análisis genético de líneas celulares estables generadas con este vector. La secuencia del sitio de clonación de pV90 se confirmó por análisis de la secuencia.

**[0092]** En el vector pV80 se puede clonar un polinucleótido (por ejemplo, una secuencia codificante) en el sitio BamHI **GGATCC** CTGCCCGGGT (SEQ ID NO: 15). La secuencia en negrita representa el sitio BamHI. En el vector pV90 se puede clonar un polinucleótido (por ejemplo, una secuencia codificante) en el sitio BamHI o NotI **GGATCC GCGGCCGC GGATCC** CTGCCCGGGT (SEQ ID NO: 16). Aquí, la secuencia en negrita representa el sitio BamHI y la secuencia subrayada representa el sitio NotI. Para obtener resultados óptimos cuando se utiliza el sitio NotI, se debe añadir una "C" antes del codón de iniciación en los cebadores de la PCR para un mejor ajuste a la secuencia de Kozak (por ejemplo, GGATCC GCGGCCGC C ATG (SEQ ID NO: 17)).

**Sitios de restricción en pV80 y pV90**

**[0093]** pV80 y pV90 son idénticos con la excepción de los sitios de restricción NotI: pV80 tiene un único sitio NotI en posición 3166 (extremo de la región codificante *dhfr*) y pV90 tiene un único sitio NotI en posición 1260 (sitio de clonación).

**[0094]** Los sitios de restricción habituales, de los que se encontraron 2 o menos, incluyen aquellos enumerados en la Tabla 1.

25

Tabla 1

AatI(1) 2274	Apal(1) 1733	AspEI(2) 1382,4807	BamHI(1) 1254
BsgI(1) 1494	BsiXI(1) 3404	Clal(1) 3404	EgeI(1) 3585
HindIII (2) 2293, 6059	HpaI(1) 3308	KasI(1) 3585	Kpn2I(1) 1121
KpnI (2) 1927, 3144	NarI (1) 3585	NdeI (2) 254, 3637	NheI (1) 2557
NotI(1) véase más arriba	PstI (2) 1231, 2331	PvuI (2) 3544, 4440	SacI (2) 585, 2819
SacII(2) 673, 1258	Smal(2) 1271, 3161	SpeI(1) 20	StuI(1) 2274
XbaI (1) 3150	XhoI (1) 3127	XmaI (2) 1271, 3161	

No se encontraron los siguientes sitios de corte con enzimas habituales:

30

Tabla 2

AatII	AfaI	AfeI	AgeI	AluI	ApaLI	AspHI	AspI	AvaI	Avall	BsiWI	DpnI
DpnII	DraI	Drall	DrallI	EaeI	EagI	EarI	EcoRI	EcoRV	FspI	HaeII	HaeIII
HincII	HpaI	NaeI	NcoI	NdeII	NspI	Pacl	Sall	SphI	XhoII	XmaIII	

**Clonación de los genes informadores**

35

**[0095]** El casete/vector de expresión que incluye el IE1 CMV, un fragmento del intrón A, y la construcción de poliA de la hGHv se comparó con un vector de alta expresión basado en SV40 disponible comercialmente.

**[0096]** La fosfatasa alcalina secretada (SEAP) se utilizó como informador para la comparación de los plásmidos. El vector de expresión basado en SV40, control pSEAP2 (Cat. # 6052-1, Clontech), expresa la SEAP bajo el control del promotor temprano de SV40 y el potenciador de SV40. A la secuencia codificante de la SEAP le sigue la señal de poliadenilación tardía de SV40 para asegurar el procesamiento adecuado y eficaz del transcrito SEAP en células eucariotas. Un bloqueador sintético de la transcripción (TB), compuesto por sitios de pausa de la poliadenilación y la transcripción adyacentes, que se encuentra en dirección 5' del MCS reduce la transcripción de fondo. El vector incorpora una serie de características que mejoran la sensibilidad de la SEAP mediante el aumento de la eficacia de la expresión de SEAP o que mejoran la utilidad de los vectores. Estos incluyen: un sitio de iniciación de la traducción consenso de Kozak mejorado; la eliminación del pequeño intrón t de SV40, que puede causar un procesamiento alternativo críptico y la expresión reducida en algunos genes y/o tipos celulares; el intercambio entre la señal de poliadenilación temprana a tardía de SV40, que normalmente provoca un incremento de cinco veces en los niveles de ARNm; un sitio de clonación múltiple expandido (MCS); un tamaño compacto del plásmido, y la eliminación de secuencias extrañas de la región 3' sin traducir del ARNm de la SEAP. Número de acceso del GenBank U89938.

50

**[0097]** Con el fin de generar el plásmido pCMV-hGHvPA-SEAP, la secuencia codificante de SEAP se extrajo a

partir del plásmido control pSEAP2 por PCR y se clonó en un vector pV110. El plásmido pV110 es un derivado de pV90 (sin casete de expresión de *dhfr*, un polienlazador diferente, pero por lo demás idéntico). La construcción del plásmido pCMV-hGHvPA-SEAP se verificó por secuenciación.

- 5 **[0098]** El hospedador utilizado para las transfecciones fue la línea celular DG44 de ovario de hámster chino deficiente en dihidrofolato reductasa (DHFR) (Urlaub y col., Cell 33, 405-412 (1983)). Los plásmidos informadores CMVSEAP y control pSEAP2 se co-transfectaron con un plásmido que codifica la dihidrofolato reductasa (DHFR) de manera que se pudieron seleccionar transfectantes estables para (PSI-DHFR.2, Figura 12). Cada transfección contenía 50 µg de un plásmido informador y 5 µg de PSI-DHFR.2. Todo el ADN se preparó con el kit Megaprep Kit
- 10 (Qiagen). Antes de la transfección, el ADN se precipitó con EtOH, se lavó en EtOH al 70%, se secó, se resuspendió en HEBS (Hepes 20 mM, pH 7,05, NaCl 137 mM, KCl 5 mM, Na<sub>2</sub>HPO<sub>4</sub> 0,7 mM, dextrosa 6 mM), y se cuantificó antes de la transfección. Los controles negativos incluyen pUC18 (ATCC N° 37.253) como control informador y una transfección sin ADN como control de la transfección (Tabla 3).
- 15 **[0099]** Las células y el ADN se transfectaron por electroporación en 0,8 ml de HEBS utilizando una cubeta de 0,4 cm (BioRad) en 0,28 kV y 950 mF. Se utilizaron 5 x 10<sup>6</sup> células para cada transfección. Después del pulso de electroporación, las células se dejan incubar en la cubeta durante 5-10 minutos a temperatura ambiente. A continuación se transfirieron a un tubo de centrifuga que contiene 10 ml de α-MEM con nucleósidos y dFBS al 10% y se sedimentaron a 1000 rpm durante 5 min. Los precipitados resuspendidos se sembraron en placas de 6 pocillos en
- 20 α-MEM sin nucleósidos con dFBS al 10% y se incubaron a 36°C con CO<sub>2</sub> al 5% en un incubador humidificado hasta formar colonias.

Plásmido informador (50 µg de cada uno)	Plásmido DHFR (5 µg de cada uno)	Nº de transfecciones
pSEAP2-control	pSIDHFR.2	3
pCMVSEAP	pSIDHFR.2	1
pUC18	pSIDHFR.2	1
Sin ADN	Sin ADN	1

- 25 **[0100]** Aproximadamente 2 semanas después de la transfección, se habían formado colonias en las transfecciones que contienen el plásmido pSIDHFR.2 solamente. Se analizaron transfectantes estables, como fracciones reunidas o aislados. La productividad específica se evaluó en ensayos en los que el medio se intercambió por medio fresco y 24 horas después se tomaron muestras del medio y las células se sometieron a recuento. El título de producto se normalizó para el número de células al final del ensayo de 24 horas, y las productividades se expresaron como actividad SEAP por célula.

- 30 **[0101]** Ensayo de la SEAP. Se analizó medio acondicionado utilizando el sistema Great EscAPe™ SEAP Reporter System 3 (Clontech). Este ensayo utiliza un sustrato fluorescente para detectar la actividad SEAP en el medio acondicionado. El kit se utilizó en un formato de 96 pocillos de acuerdo con las instrucciones del fabricante, con las siguientes excepciones. El tampón de ensayo del kit se sustituyó por dietanolamina 1,5 M, MgCl<sub>2</sub> 0,75 mM, L-
- 35 homoarginina 15 mM, y 10% de Emerald II (Cat. # 9761, Applied Biosystems). Todos los patrones y las muestras se diluyeron en medio fresco en lugar del tampón de dilución previsto. En vez de hacer una lectura después de 60 minutos, se tomaron lecturas múltiples a intervalos de 10-20 min y se utilizaron para expresar la actividad SEAP como unidades relativas de fluorescencia por minuto (RFU/min). El filtro de emisión para el lector de placas (Cytofluor II, PerSeptive Biosystems) se encontraba a 460 nm en lugar de a los 449 nm recomendados.

- 40 **[0102]** Los valores de RFU/min se normalizaron a una curva patrón basada en un patrón suministrado con el kit. Debido a que el patrón proporcionado no estaba cuantificado, todos los valores son relativos. Estos valores relativos se normalizaron al número de células y al periodo de incubación para generar productividades relativas específicas (actividad SEAP por célula y día).

- 45 **[0103]** Fracciones reunidas. Después de la aparición de colonias, las células procedentes de cada transfección se recogieron y se reunieron. Las fracciones reunidas se sembraron a ~2x10<sup>5</sup> células por pocillo en placas de 6 pocillos. Al día siguiente, el medio se cambió por 2 ml de medio fresco. Después de 24 horas las células se contaron y se utilizó una muestra del medio para evaluar la actividad SEAP. Los resultados de los ensayos de las fracciones
- 50 reunidas se muestran en la Figura 13.

- [0104]** Aislados. Los aislados se obtuvieron "recogiendo" colonias de la transfección. La "recolección" se consiguió aspirando directamente sobre una colonia con una pipeta P200 Pipetman™ fijada a 50 µl. La colonia aspirada se transfirió en primer lugar a una placa de 48 pocillos y a continuación a una placa de 6 pocillos cuando un número
- 55 suficiente de células. Las productividades específicas se evaluaron en placas de 6 pocillos a densidades celulares de casi confluencia a confluencia utilizando el ensayo de 24 horas descrito anteriormente (Figura 14).

- [0105]** Como se resume en las Figuras 13 y 14, un vector de expresión basado en la combinación de un promotor de CMV IE y un dominio de la señal de poliA de la hGHv fue muy superior a un vector comercialmente disponible



que presume de altas capacidades de expresión.

**LISTADO DE SECUENCIAS**

**5 [0106]**

<110> Biogen Idec Inc.

<120> CASETE Y VECTOR DE EXPRESIÓN PARA LA EXPRESIÓN TRANSITORIA O ESTABLE DE MOLÉCULAS EXÓGENAS

10 <130> 13751-013WO1

<150> US 60/448.179

<151> 2003-02-14

<160> 21

<170> FastSEQ para Windows Versión 4.0

15 <210> 1

<211> 6069

<212> ADN

<213> Homo sapiens

<400> 1

20

```

agcttgacat  tgattattga  ctagttatta  atagtaatca  attacggggt  cattagttca      60
tagcccatat  atggagttcc  gcgttacata  acttacggta  aatggccgcg  ctggctgacc     120
gcccacagac  ccccgcccat  tgacgtcaat  aatgacgtat  gttcccatag  taacgccaat     180
agggactttc  cattgacgtc  aatgggtgga  gtatttacgg  taaactgccc  acttggcagt     240
acatcaagtg  tatcatatgc  caagtacgcc  ccctattgac  gtcaatgacg  gtaaatggcc     300
cgcctggcat  tatgcccagt  acatgacctt  atgggacttt  cctacttggc  agtacatcta     360
cgtattagtc  atcgctatta  ccatggtgat  gcggttttgg  cagtacatca  atgggcgtgg     420
atagcggttt  gactcacggg  gatttccaag  tctccacccc  attgacgtca  atgggagttt     480
gttttggaac  caaaatcaac  gggactttcc  aaaatgtcgt  aacaactccg  ccccattgac     540
gcaaatgggc  ggtaggcgtg  tacggtggga  ggtctatata  agcagagctc  gtttagtgaa     600
ccgtccagtc  gcctggagac  gccatccaag  ctgttttgac  ctccatagaa  gacaccggga     660
ccgatccagc  ctccgcggcc  gggaacggtg  cattggaacg  cggattcccc  gtgccaagag     720
tgacgtaagt  accgcctata  gagtctatag  gccaccccc  ttggcttctt  atgcatgcta     780
tactgttttt  ggcttggggt  ctatacacc  ccgcttcttc  atgttatagg  tgatggtata     840
gcttagccta  taggtgtggg  ttattgacca  ttattgacca  ctcccctatt  ggtgacgata     900
ctttccatta  ctaatccata  acatggctct  ttgccacaac  tctctttatt  ggctatatgc     960
caatacactg  tcttccagag  actgacacgg  actctgtatt  ttacaggat  ggggtctcat    1020
ttattattta  caaattcaca  tatacaaac  caccgtcccc  agtgcccgca  gtttttatta    1080
aacataacgt  gggatctcca  cgcgaatct  c  gggtacgtgt  tcoggaacgg  tggagggcag    1140
tgtagtctga  gcagtaactg  ttgctgccc  gcgcgccacc  agacataata  gctgacagac    1200
taacagactg  ttoccttcca  tgggtctttt  ctgcagtcac  cgtccttgac  acgggatccg    1260
cggcccgga  tccctgcccg  ggtggcatcc  ctgtgacccc  tcccagtg  ctctcctggt    1320
cgtggaaggt  gctactccag  tgcccaccag  ccttgtccta  ataaaattaa  gttgcatcat    1380
tttgtttgac  taggtgtcct  tgtataatat  tatggggtgg  aggcgggtgg  tatggagcaa    1440
ggggcaggtt  ggaagacaa  cctgtagggc  cttcagggtc  tattgggaac  caggctggag    1500
tgacgtggca  cgatcttggc  tcgctgcaat  ctccgectcc  tgggttcaag  cgattctcct    1560
gcctcagctc  ccggaatagt  tgggattcca  ggcattgcag  accaggctca  gctaattttt    1620
gtattttttg  tagagacggg  gtttcaccat  attggccagt  ctggtctoca  tctcctgacc    1680
tcaggtaatc  cgcctcctc  ggctcccaa  attgctggga  ttacaggat  gagccactgg    1740
gccttccct  gtctgtgat  tttaaaata  ttataccagc  agaaggacgt  ccagacacag    1800
catgggctac  ctggccatgc  ccagccagtt  ggacatttga  gttgtttgc  tggcactgtc    1860
ctctcatgaa  ttctgtcgaca  gatctgcgca  gcacatggc  ctgaaataac  ctctgaaaga    1920
ggaacttgg  taggtacctt  ctgaggcgga  aagaaccagc  tgtggaatgt  gtgtcagtta    1980
gggtgtgaa  agtccccagg  ctcccagca  ggcagaagta  tgcaaagcat  gcatctcaat    2040
tagtcagcaa  ccaggtgtgg  aaagtcccca  ggctccccag  caggcagaag  tatgcaaage    2100
atgcatctca  attagtcagc  aaccatagtc  ccgcccctaa  ctccgcccct  cccgccccta    2160
actccgcca  gttccgcca  ttctccgccc  catggtgac  taattttttt  tatttatgca    2220
gagggcagg  ccgctcggc  ctctgagcta  ttccagaagt  agtgaggagg  cttttttgga    2280

```

ES 2 384 973 T3

ggocctaggct tttgcaaaaa gcttgattct tctgacacaa cagtctcgaa cttaaactgc 2340  
 agaagttggc cgtgaggcac tgggcaggta agtatcaagg ttacaagaca ggtttaagga 2400  
 gaccaataga aactgggctt gtcgagacag agaagactct tgcgtttctg ataggcacct 2460  
 attggctctta ctgacatcca ctttgccttt ctctccacag gtgtccactc ccagttcaat 2520  
 tacagctctt aaggctagag tacttaatac gactcactat aggctagcat ggttcgacca 2580  
 ttgaactgca tgcgccccgt gtoccaaaat atggggattg gcaagaacgg agacctacc 2640  
 tggcctccgc tcaggaacga gttcaagtac ttccaaagaa tgaccacaac ctcttcagt 2700  
 gaaggtaaac agaactctgt gattatgggt aggaaaaact ggttctccat tccctgagaag 2760  
 aatcgacctt taaaggacag aattaatata gttctcagta gagaactcaa agaaccacca 2820  
 cgaggagctc attttcttgc caaaagtttg gatgatgctt taagaactat tgaacaaccg 2880  
 gaattggcaa gtaaagtaga catggtttgg atagtcggag gcagttctgt ttaccaggaa 2940  
 gccatgaatc aaccaggcca cctcagactc tttgtgacaa ggatcatgca ggaatttgaa 3000  
 agtgacacgt ttttccaga aattgatttg gggaaatata aacttctccc agaataccca 3060  
 ggcgctctct ctgaggcca ggaggaaaaa ggcatacaagt ataagtttga agtctacgag 3120  
 aagaaagact aactcgagaa ttcacgcgtg gtacctctag agtcgacccg ggcggccggc 3180  
 cgcttcgagc agacatgata agatacattg atgagtttgg acaaacccca actagaatgc 3240  
 agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat tgccttattt gtaaccatta 3300  
 taagctgcaa taaacaagt aacaacaaca attgcattca ttttatgttt caggttcagg 3360  
 gggaggtgtg ggaggtttt taaagcaagt aaaacctcta caaatgtggt aaaatcgata 3420  
 aggatctgtc gacgaattca ctggccgtcg ttttacaacg tcgtgactgg gaaaacctg 3480  
 gcgttaccga acttaatcgc cttgcagcac atccccctt cgcagctgg cgtaatagcg 3540  
 aagagggccg caccgatcgc ccttcccac agttgcgcag cctgaatggc gaatggcgcc 3600  
 tgcagcggta ttttctcctt acgcctctgt tgcgtatttc acaccgcata tgggtgactc 3660  
 tcagtacaat ctgctctgat gccgcacagt cagccagcc cgcacaccg ccaacaccg 3720  
 ctgacgcgcc ctgacggcct tgtctgtctc cggeatccgc ttacagacaa gctgtgaccg 3780  
 tctccgggag ctgcatgtgt cagaggtttt caccgtcctc accgaaacgc gcgagacgaa 3840  
 agggcctcgt gatacgccta ttttatagg ttaatgtcat gataataatg gtttcttaga 3900  
 cgtcaggctg cacttttcgg ggaatgtgc gggaaacccc tatttgttta ttttctaaa 3960  
 tacattcaaa tatgtatccg ctcatgagac aataacctg ataaatgctt caataatatt 4020  
 gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc cottattccc tttttgctg 4080  
 cattttgcct tctgttttt gctcaccag aaacgctgg gaaagtaaaa gatgctgaa 4140  
 atcagttggg tgcacgagtg ggttacatcg aactggatct caacagcgg aagatcctg 4200  
 agagttttgc ccccgagaa cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg 4260  
 gcgcggatatt atcccgatt gacgcggggc aagagcaact cggtcgccgc atacactatt 4320  
 ctcagaatga cttggttgag tactcaccag tcacagaaaa gcatottacg gatggcatga 4380  
 cagtaagaga attatgcagt gctgccataa ccatgagtga taacactgcg gccaacctac 4440  
 ttctgacaac gatcggagga ccgaaggagc taaccgctt tttgcacaac atgggggatc 4500  
 atgtaactcg ccttgatcgt tgggaaccgg agctgaatga agccatacca aacgcagagc 4560  
 gtgacaccac gatgcctgta gcaatggcaa caactgtgc caaactatta actggcgaac 4620  
 tacttactct agcttcccga caacaattaa tagactggat ggaggcggat aaagttgcag 4680  
 gaccacttct gcgctcggcc cttccggtg gctggtttat tgcgtataaa tctggagccg 4740  
 gtgagcgtgg gtctcgcggg atcattgcag cactggggcc agatggtaag cctcccgt 4800  
 tgcgtagttat ctacacgacg gggagtcagg caactatgga tgaaacgaaat agacagatcg 4860  
 ctgagatagg tgcctcactg attaagcatt ggttaactgtc agaccaagtt tactcatata 4920  
 tactttagat tgaattaaa attcatttt aattttaaag gatctaggta agatccttt 4980  
 ttgataatct catgacaaa atocctaac gtgagtttc gttccactga gcgtcagacc 5040  
 ccgtagaaaa gatcaaagga tcttcttgag atccttttt tctgcgcgta atctgctgct 5100  
 tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgtt gcccgatcaa gagctaccaa 5160  
 ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag 5220  
 tgtagccgta gttaggccac cacttcaaga actctgtagc accgcctaca tacctcctc 5280  
 tgctaatoct gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttg 5340  
 actcaagacg atagttaccg gataaggcgc agcggctggg ctgaaocggg ggttcgtgca 5400  
 cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat 5460  
 gagaaagcgc cacgcttccc gaaggagaa aggggacag gtatccggt agcggcagg 5520  
 tgggaacagg agagcgcagc agggagctc cagggggaaa ccgctggtat ctttatagtc 5580  
 ctgtcgggtt tgcaccctc tgacttgagc gtcgattttt gtgatgctc tcaggggggc 5640  
 ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc 5700  
 cttttgctca catgttcttt cctgcgttat cccctgattc tgtggataac cgtattaccg 5760  
 cctttgagtg agctgatacc gctcgcgcc aaccgacgac cagcgcagc gagtcagtga 5820  
 gcgaggaagc ggaagagcgc ccaatacgc aaccgcctct ccccgccgt tggccgatc 5880  
 attaatgcag ctggcacgac aggtttccc actggaagc ggcagtgag gcgaacgcaa 5940

ttaatgtgag ttagctcact cattagggcacc cccaggcttt acactttatg ctccggctc 6000  
 gtatggtgtg tggaaattgtg agcgggataac aatttcacac aggaaacagc tatgaccatg 6060  
 attacgcca 6069

<211> 28  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 5 <223> cebador  
 <400> 2  
 ttaagcttg acattgatta ttgactag 28  
 <210> 3  
 <211> 30  
 10 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> cebador  
 <400> 3  
 15 ttttgatcc ctgtcaagga cggtgactgc 30  
 <210> 4  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 20 <220>  
 <223> oligonucleótido adaptador  
 <400> 4  
 gatc gatgaa ttcg 15  
 <210> 5  
 25 <211> 13  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido adaptador  
 30 <400> 5  
 ctacttaagc cgc 13  
 <210> 6  
 <211> 28  
 <212> ADN  
 35 <213> Secuencia artificial  
 <220>  
 <223> cebador  
 <400> 6  
 ttttgatcc atgtactggg tgaagcag 28  
 40 <210> 7  
 <211> 28  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 45 <223> cebador  
 <400> 7  
 gcccgatcc tcattacc ggagacaq 28  
 <210> 8  
 <211> 53  
 50 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> cebador  
 <400> 8  
 55 tttctgcag tcaccgtcct tgacacggga tccatggact ggacctgga ggg 53  
 <210> 9  
 <211> 33  
 <212> ADN  
 <213> Secuencia artificial  
 60 <220>  
 <223> cebador  
 <400> 9  
 ctgaggagac ggtgaccagg gtccttggc ccc 33  
 <210> 10  
 65 <211> 12  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> oligonucleótido adaptador  
 <400> 10  
 5 aattcgtcga ca 12  
 <210> 11  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> oligonucleótido adaptador  
 <400> 11  
 gcagctgtct ag 12  
 <210> 12  
 15 <211> 14  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido generado sintéticamente  
 20 <400> 12  
 gatccgcggc cgcg 14  
 <210> 13  
 <211> 13  
 <212> ADN  
 25 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido generado sintéticamente  
 <400> 13  
 gatccgcggc cgc 13  
 30 <210> 14  
 <211> 14  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 35 <223> oligonucleótido generado sintéticamente  
 <400> 14  
 cgccggcgcc tagg 14  
 <210> 15  
 <211> 16  
 40 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido generado sintéticamente  
 <400> 15  
 45 ggatccctgc ccgggt 16  
 <210> 16  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 50 <220>  
 <223> oligonucleótido generado sintéticamente  
 <400> 16  
 ggatccgagg ccgcatcc ctgcccgggt 30  
 <210> 17  
 55 <211> 18  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido generado sintéticamente  
 60 <400> 17  
 ggatccgagg ccgcatg 18  
 <210> 18  
 <211> 2660  
 <212> ADN  
 65 <213> Homo sapiens  
 <400> 18

ES 2 384 973 T3

gaattcagca	ctgaatcatg	cccagaaccc	ccgcaatcta	ttggetgtgc	tttggcccct	60
tttcccaaca	cacacattct	gtctgggtggg	tggaggggaa	acatgcgggg	aggaggaag	120
gaataggata	gagagtggga	tggggtcggg	aggggtctca	aggactggcc	tatcctgaca	180
tccttctccg	cgttcagggt	ggccaccatg	gcctgctgcc	agagggcacc	cacgtgaccc	240
ttaaagagag	gacaagttag	gtggtatctc	tggetgacat	tctgtgcaca	acctcacia	300
cgctggtgat	ggtgggaagg	gaaagatgac	aagtcagggg	gcatgatccc	agcatgtgtg	360
ggaggagott	ctaaattatc	cattagcaca	agcccgtcag	tggccccagg	cctaaacatg	420
cagagaaaca	ggtgaggaga	agcagcgaga	gagaaggggc	caggtataaa	aagggccac	480
aagagaccag	ctcaaggatc	ccaaggccca	actccccgaa	ccactcaggg	tcctgtggac	540
agctcaactag	cggcaatggc	tgcaggtaa	cgccccataa	atccctttgg	cacaatgtgt	600
cctgagggga	gaggcggcgt	cctgtagatg	ggacgggggc	actaaccttc	aggtttgggg	660
cttatgaatg	ttagctatcg	ccatctaagc	ccagtatttg	gccaatctct	gaatgttctc	720
ggtccctgga	ggaggcagag	agagagagag	agaaaaaaaa	aaccagctc	ctggaacagg	780
gagagcgctg	gocctctgct	ctccagctcc	ctctggtgcc	tcgggtttct	ccccaggtc	840
ccggacgtcc	ctgctcctgg	cttttggcct	gctctgocctg	tcctggcttc	aagagggcag	900
tgccttccca	accattccct	tatccaggct	ttttgacaac	gctatgctcc	gctgcccctg	960
ccgttaccag	ctggcaatag	acacctatca	ggagtttcta	agctcttggg	taatgggtgc	1020
gcttcagagg	tggcaggaag	gggtgaattt	cccccgctgg	gaagtaatgg	gaggagacta	1080
aggagctcag	ggttgttttc	tgaagtga	atgcaggcag	atgagcatac	gctgagttag	1140
gttcccagaa	aagtaacaat	gggagcaggt	ctccagcata	gaccttgggtg	ggcggctcctt	1200
ctcctaggaa	gaagcctata	tctgaagga	gcagaagtat	tcattcctgc	agaacccccca	1260
gacctccctc	tgtctctcag	agtctatctc	aacaccttcc	aacaggtgta	aaacgcagca	1320
gaaatctgtg	agtggatgcc	ttctccccag	gtgggatggg	gtagacctgt	ggtcagagcc	1380
cccgggcagc	acagccactg	ccggctcctc	ccctgcagaa	cctagagctg	ctccgcatct	1440
ccctgctgct	catccagtca	tggctggagc	ccgtgcagct	cctcaggagc	gtcttcgcca	1500
acagcctggt	gtatggcgcc	tcggacagca	acgtctatcg	ccacctgaag	gacctagagg	1560
aaggcatcca	aacgctgatg	tgggtgaggg	tggcaccagg	atccaatcct	ggggccccac	1620
tggcttccag	ggactgggga	gagaaacact	gctgcocctc	ttttagcagt	caggcgctga	1680
cccaagagaa	ctcaccgtat	tcttcatttc	ccctcgtgaa	tcctccagge	ctttctctac	1740
aacctggagg	ggagggagga	aatggatga	atgagagagg	gaggaacag	tgcccagcgt	1800
cttggcctct	ccttctcttc	cttcactttg	cagaggctgg	aagatggcag	cccccgact	1860
gggcagatct	tcaatcagtc	ctacagcaag	tttgacacaa	aatcgcaaa	cgatgacgca	1920
ctgctcaaga	actacgggct	gctctactgc	ttcaggaagg	acatggacaa	ggtcagagaca	1980
ttcctgcgca	tcgtgcagtg	ccgctctgtg	gagggcagct	gtggctteta	gctgcccggg	2040
tggcatecct	gtgacccttc	cccagtgctc	ctcctggctg	tggaagggtg	tactccagtg	2100
cccaccagcc	ttgtcctaat	aaaattaagt	tgcateattt	tgtttgacta	ggtgtccttg	2160
tataatatta	tggggtggag	gcgggtggta	tggagcaagg	ggccagggtg	ggaagacaac	2220
ctgtagggcc	ttcagggctc	attcgggaac	caggctggag	tgcagtgcca	gtcttggtct	2280
getgcaatct	cgcctcctg	ggttcaagcg	attctcctgc	ctcagctctc	ogaatagttg	2340
cgattccagg	catgcaagac	caggctcagc	taatttttgt	atttttggta	gagacggggt	2400
ttcaccaat	tggccagctc	ggtctccatc	tcctgacctc	aggtaatccg	ccgcctcgg	2460
cotcccaaat	tgtctgggatt	acaggataga	gccaactggc	ccttccctgt	cctgtgattt	2520
taaaataatt	ataccagcag	aaggacgtcc	agacacagca	tgggctacct	ggccatgccc	2580
agccagttgg	acatttgagt	tgtttgcttg	gcaactgtcct	ctcatgcatt	gggtccactc	2640
agtagatgct	tgttgaattc					2660

- <210> 19
- 5 <211> 594
- <212> ADN
- <213> Homo sapiens
- <400> 19

ES 2 384 973 T3

ctgcccggt	ggcatccctg	tgaccctcc	ccagtgcctc	tcctggctcg	ggaaggtgct	60
actccagtgc	ccaccagcct	tgtcctaata	aaattaagtt	gcatactttt	gtttgactag	120
gtgtccttgt	ataatattat	ggggtggagg	cgggtgggat	ggagcaaggg	gccaggttgg	180
gaagacaacc	tgtagggcct	tcagggctca	ttcgggaacc	aggctggagt	gcagtggcag	240

tcttggctcg	ctgcaatctc	cgctcctgg	gttcaagcga	ttctcctgcc	tcagtctccc	300
gaatagttgc	gattccaggc	atgcaagacc	aggctcagct	aatttttgta	tttttggtag	360
agacggggtt	tcaccatatt	ggccagtctg	gtctccatct	cctgacctca	ggtaatccgc	420
ccgcctcggc	ctcccaaatt	gctgggaita	caggtatgag	ccaactgggc	cttccctgtc	480
ctgtgatttt	aaaataatta	taccagcaga	aggacgtcca	gacacagcat	gggctaacctg	540
gccatgccca	gccagttgga	catttgagtt	gtttgcttgg	cactgtcctc	tcata	594

<210> 20

<211> 27

5 <212> ADN

<213> Secuencia artificial

<220>

<223> cebador

<400> 20

10 ttttgatcc ctgcccggt ggcatcc 27

<210> 21

<211> 30

<212> ADN

<213> Secuencia artificial

15 <220>

<223> cebador

<400> 21

tttgaattc atgagaggac agtgccaagc 30

## REIVINDICACIONES

1. Un casete de expresión que comprende una región promotora/potenciadora inmediata temprana 1 del CMV humano (hCMV IE1); un polinucleótido de interés; un dominio de la señal de poliA de la hormona del crecimiento humana; y una secuencia interviniente de longitud variable (VLIVS) que comprende un sitio donador de corte y empalme y un sitio aceptor de corte y empalme, en donde el dominio de la señal de poliA tiene una longitud de al menos 100 nucleótidos, contiene la secuencia AATAAA, y es idéntica en al menos un 92% a la secuencia expuesta en la SEQ ID NO: 18; y en donde la VLIVS comprende un intrón A de un gen hCMV IE1.
2. El casete de expresión de la reivindicación 1, en el que la región promotora/potenciadora hCMV IE1 comprende una secuencia entre el nucleótido  $x_1$  y el nucleótido  $x_2$  de la SEQ ID NO: 1, en la que  $x_1$  denota un nucleótido desde el nucleótido 1 hasta el nucleótido 20 de la SEQ ID NO: 1 y  $x_2$  denota un nucleótido desde el nucleótido 715 hasta el nucleótido 720 de la SEQ ID NO: 1.
3. El casete de expresión de la reivindicación 1, en el que el casete de expresión comprende la secuencia expuesta en la SEQ ID NO: 1.
4. El casete de expresión de la reivindicación 1, en el que el casete de expresión comprende una secuencia complementaria a la secuencia expuesta en la SEQ ID NO: 1.
5. El casete de expresión de la reivindicación 1, en el que el casete de expresión comprende una secuencia al menos un 80% idéntica a la SEQ ID NO: 1 o su complementario.
6. El casete de expresión de la reivindicación 5, en el que el casete de expresión comprende una secuencia al menos un 90% idéntica a la SEQ ID NO: 1 o su complementario.
7. El casete de expresión de la reivindicación 6, en el que el casete de expresión comprende una secuencia al menos un 95% idéntica a la SEQ ID NO: 1 o su complementario.
8. Los casetes de expresión de la reivindicación 7, en los que el casete de expresión comprende una secuencia al menos un 97% idéntica a la SEQ ID NO: 1 o su complementario.
9. El casete de expresión de la reivindicación 8, en el que el casete de expresión comprende una secuencia al menos un 98% idéntica a la SEQ ID NO: 1 o su complementario.
10. El casete de expresión de la reivindicación 9 en el que el casete de expresión comprende una secuencia al menos un 99% idéntica a la SEQ ID NO: 1 o su complementario.
11. El casete de expresión de la reivindicación 1, en el que el casete de expresión comprende la secuencia entre el nucleótido 1 y el nucleótido 1867 de la SEQ ID NO: 1.
12. El casete de expresión de la reivindicación 1, en el que la VLIVS comprende un intrón A de un gen hCMV IE1 que tiene una supresión entre el aceptor de corte y empalme y el donador de corte y empalme del intrón A.
13. El casete de expresión de la reivindicación 1, en el que la VLIVS comprende una secuencia entre el nucleótido  $x_3$  y el nucleótido  $x_4$  de la SEQ ID NO: 1, en la que  $x_3$  denota un nucleótido desde el nucleótido 715 hasta el nucleótido 720 de la SEQ ID NO: 1 y  $x_4$  denota un nucleótido desde el nucleótido 1236 hasta el nucleótido 1254 de la SEQ ID NO: 1.
14. El casete de expresión de la reivindicación 1, que comprende además un elemento promotor/potenciador o región reguladora adicional.
15. El casete de expresión de la reivindicación 14, en el que el elemento promotor/potenciador es un elemento promotor/potenciador de SV40.
16. El casete de expresión de la reivindicación 14, en el que la región reguladora es una región de poliadenilación de SV40.
17. El casete de expresión de la reivindicación 1, que comprende además un marcador seleccionable.
18. El casete de expresión de la reivindicación 17, en el que el marcador seleccionable se selecciona del grupo constituido por la dihidrofolato reductasa, GPF, neomicina, Hígro, Zeocin™, timidina quinasa del herpesvirus simplex, hipoxantina-guanina fosforribosiltransferasa, adenina fosforribosiltransferasa, puromicina N-acetiltransferasa o adenosina desaminasa.

19. El casete de expresión de la reivindicación 17, en el que el marcador seleccionable es la dihidrofolato reductasa.
20. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés codifica un agente terapéutico.
21. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés está en la orientación no codificante con respecto al promotor.
- 10 22. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés es un polinucleótido heterólogo.
23. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés incluye, además, elementos de control de la transcripción.
- 15 24. El casete de expresión de la reivindicación 23, en el que los elementos de control de la transcripción se seleccionan del grupo constituido por señales de detención de la transcripción, dominios de poliadenilación y elementos potenciadores en dirección 3'.
- 20 25. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés codifica un polipéptido de uso diagnóstico o terapéutico.
26. El casete de expresión de la reivindicación 1, en el que el polinucleótido de interés codifica un polipéptido antigénico para su uso como una vacuna.
- 25 27. El casete de expresión de la reivindicación 1, en el que el dominio de la señal de poliA comprende al menos 100 nucleótidos contiguos de la SEQ ID NO: 19.
28. El casete de expresión de la reivindicación 27, en el que el dominio de la señal de poliA comprende la SEQ ID NO: 19.
- 30 29. El casete de expresión de la reivindicación 27, en el que el dominio de la señal de poliA es la SEQ ID NO: 19.
- 35 30. El casete de expresión de la reivindicación 1 en la forma de construcción de ácido nucleico desnudo.
31. Un polinucleótido que se hibrida específicamente con una secuencia de polinucleótidos como se expone en la SEQ ID NO: 1 desde el nucleótido 1 hasta el nucleótido 1867 o un fragmento de la misma.
- 40 32. Un vector de expresión que comprende un casete de expresión de una cualquiera de las reivindicaciones 1 a 30.
33. Una célula hospedadora que comprende un vector de expresión de la reivindicación 32.
- 45 34. Una célula hospedadora que comprende un casete de expresión de una cualquiera de las reivindicaciones 1 a 30.
35. Un procedimiento *in vitro* para producir un polipéptido que comprende la propagación de una célula hospedadora de la reivindicación 33 o la reivindicación 34 y la expresión del polinucleótido de interés.
- 50 36. El procedimiento de la reivindicación 35, en el que el polinucleótido de interés codifica un agente terapéutico y el procedimiento comprende la recuperación del agente terapéutico.
37. El uso de un casete de expresión de cualquiera de las reivindicaciones 1 a 30 o un vector de expresión de la reivindicación 32 para la fabricación de un medicamento para su uso en terapia génica.
- 55 38. El vector genético pV40, que comprende, en orden:
- a. un promotor inmediato temprano 1 de citomegalovirus (CMV IE1),
- 60 b. un intrón;
- c. un dominio de la señal de poliadenilación (poliA) de hGHv como se establece en la SEQ ID NO: 19; y
- d. un gen de la  $\beta$  lactamasa (*bla*);

en donde dicho vector genético se puede obtener mediante el uso de los cebadores de ácidos nucleicos de la SEQ ID NO: 2 y SEQ ID NO: 3 para amplificar un fragmento que contiene el promotor IE1 de CMV y el intrón adyacente a partir del ADN genómico de fibroblastos diploides humanos que están infectados con la cepa AD169 de



- citomegalovirus humano depositada en la ATCC con el número de acceso ATCC VR-538; la digestión del fragmento resultante con BamHI y HindIII; la inserción de dicho fragmento digerido en un vector pUC19 que se digiere con BamHI y HindIII, para así formar el vector pV10; utilizando los cebadores de ácidos nucleicos de la SEQ ID NO: 20 y SEQ ID NO: 21 para amplificar un fragmento que contiene el dominio de la señal de poliA de la hGHv a partir del ADN genómico de fibroblastos humanos; digiriendo el fragmento resultante con BamHI y EcoRI; e insertando dicho fragmento digerido en dicho vector pV10 que se digiere con BamHI y EcoRI.
39. El vector genético pV60, que se puede obtener a partir del vector genético pV40 según la reivindicación 38 mediante la sustitución de la región de 317 pares de bases entre los sitios BspEI y HpaI dentro del intrón de pV40 con una región codificante *dhfr*.
40. El vector genético pV70, que se puede obtener a partir del vector genético pV40 según la reivindicación 38 mediante la supresión de la región de 317 pares de bases entre los sitios BspEI y HpaI dentro del intrón de pV40.
41. El vector genético pXLC.1, que se puede obtener a partir del vector genético pV70 tal como se define en la reivindicación 40 mediante la inserción de una secuencia de ADNc que codifica para la cadena ligera de un anticuerpo en el único sitio BamHI en el extremo 3' de dicho intrón.
42. El vector genético pXLC.2, que se puede obtener a partir del vector genético pXLC.1 tal como se define en la reivindicación 41 por escisión del fragmento BamHI/EcoRI que contiene el neo casete del plásmido pGT-N28; la conversión del extremo BamHI de dicho fragmento en un extremo NarI por ligación de dicho fragmento a los oligonucleótidos adaptadores de la SEQ ID NO: 4 y SEQ ID NO: 5; y la inserción del fragmento resultante en dicho vector XLC.1 que se digiere con NarI y EcoRI.
43. El vector genético pXHC.1, que se puede obtener a partir del vector genético de pV60 tal como se define en la reivindicación 39 mediante la inserción de una secuencia codificante de ADNc para la cadena pesada de un anticuerpo en el sitio BamHI único en el extremo 3' de dicho intrón.
44. El vector genético pXHC.3, que es idéntico al vector genético pXHC.1 tal como se define en la reivindicación 43, excepto porque carece de la región codificante *dhfr* dentro de dicho intrón.
45. El vector genético pXHC.5, que se puede obtener a partir del vector genético de pXHC.3 tal como se define en la reivindicación 44 por escisión del fragmento BglII/BamHI que contiene el promotor temprano de SV40, el gen *dhfr*, y la región poliA de SV40 del plásmido pSI-DHFR, la digestión del plásmido pXHC.3 con EcoRI para producir extremos EcoRI; la ligación de los extremos EcoRI a los oligonucleótidos adaptadores de la SEQ ID NO: 10 y SEQ ID NO: 11 para producir extremos compatibles con BglII/BamHI; y la ligación de dicho fragmento BglII/BamHI a partir del plásmido pSI-DHFR a dichos extremos compatibles, para formar así dicho vector genético pXHC.5.
46. El vector genético pV80, que se puede obtener a partir del vector genético pXHC.5 tal como se define en la reivindicación 45 por escisión de la secuencia codificante de la cadena pesada a partir de pXHC.5 utilizando BamHI.
47. El vector genético pV90 como se expone en la SEQ ID NO: 1.

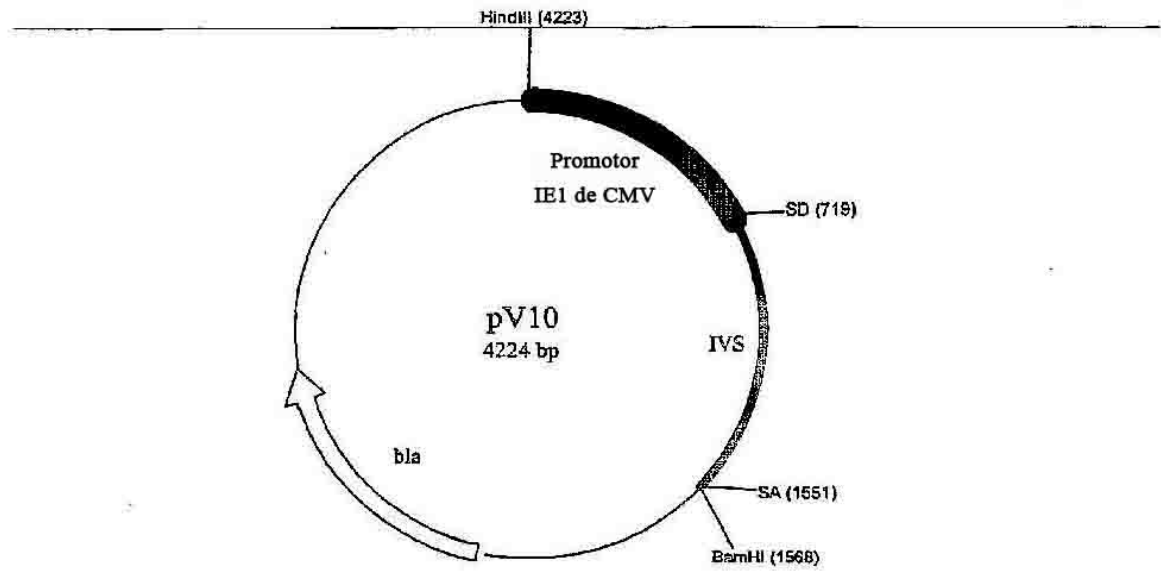


FIG. 1

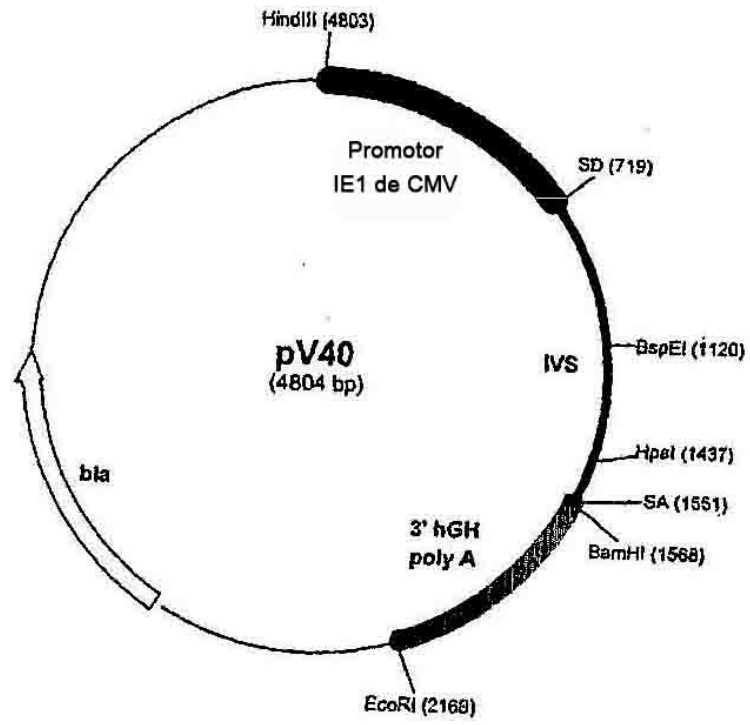


FIG. 2

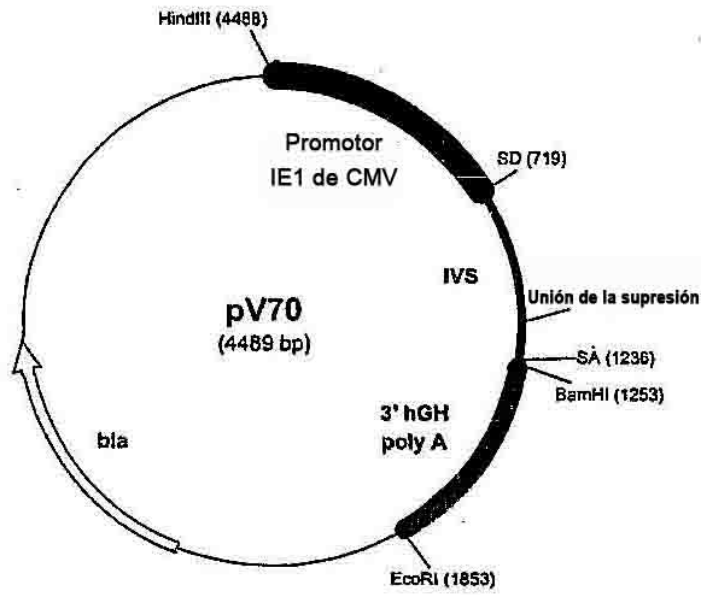


FIG. 3

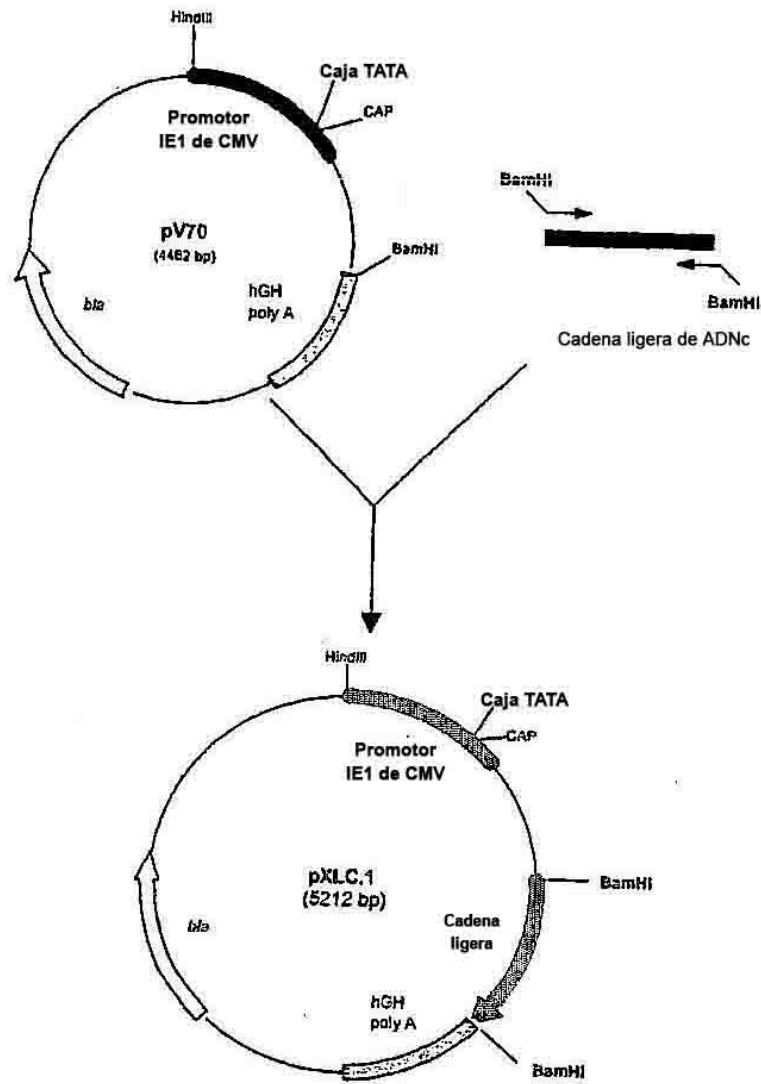


Fig. 4

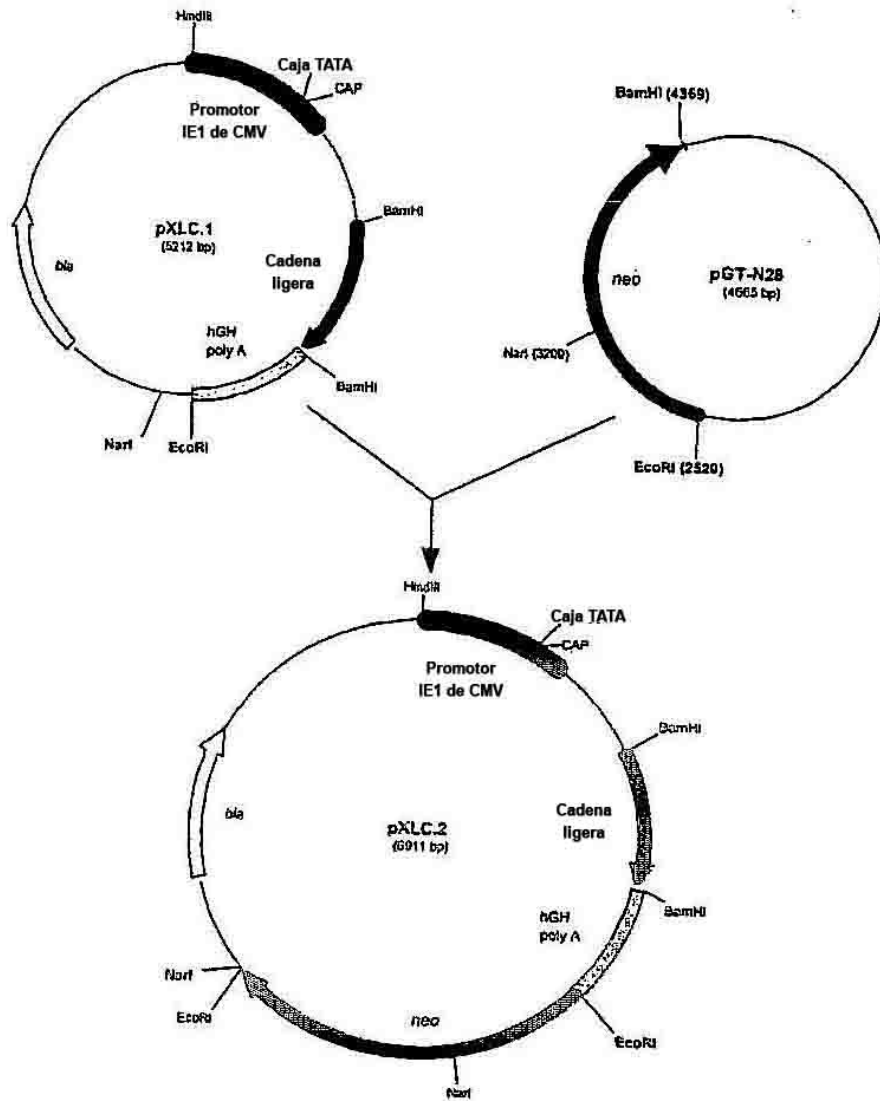


FIG. 5

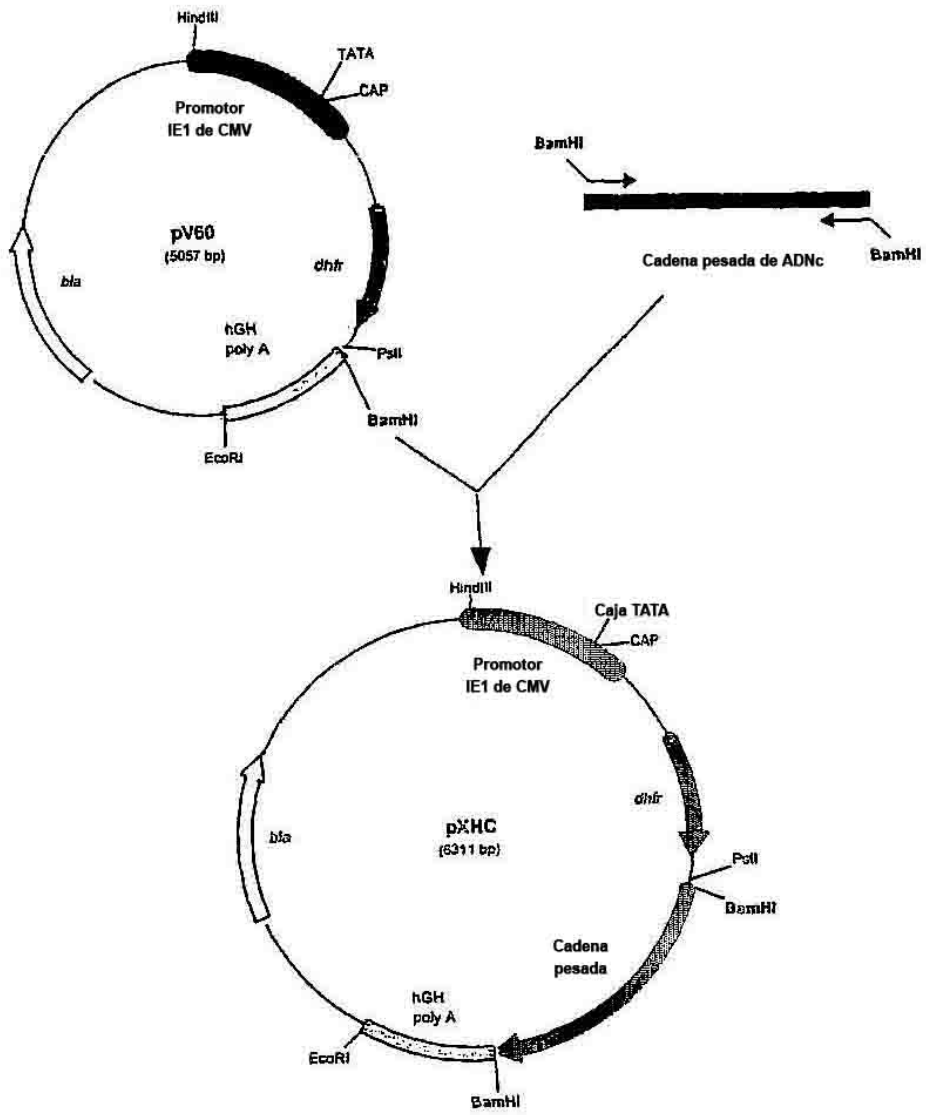


FIG. 6

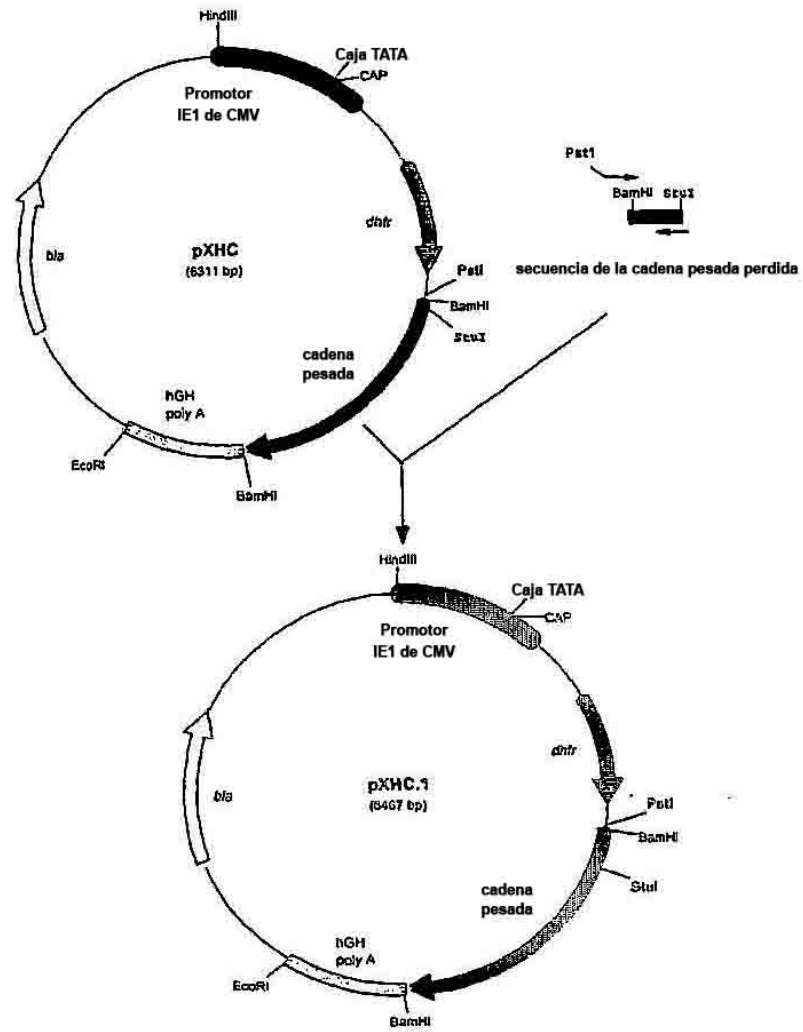


FIG. 7



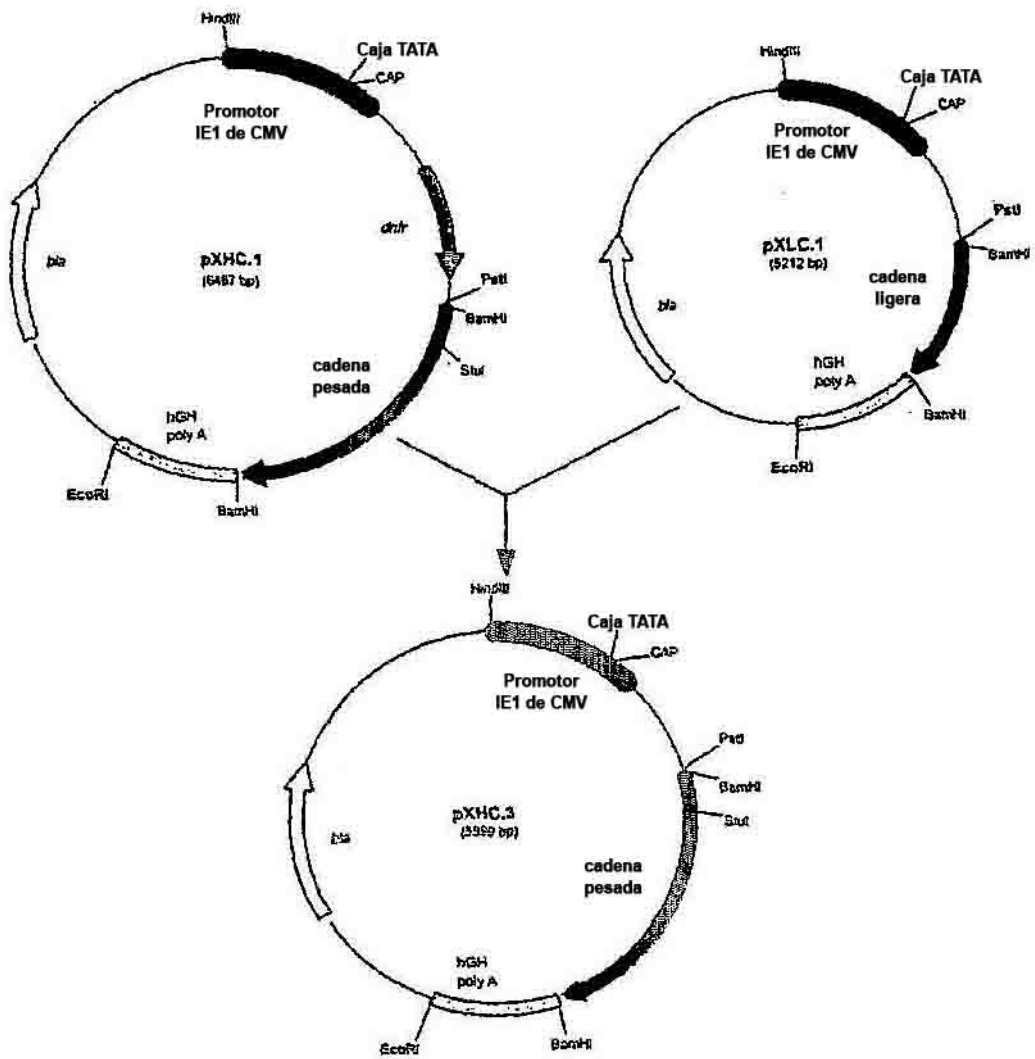


FIG. 8

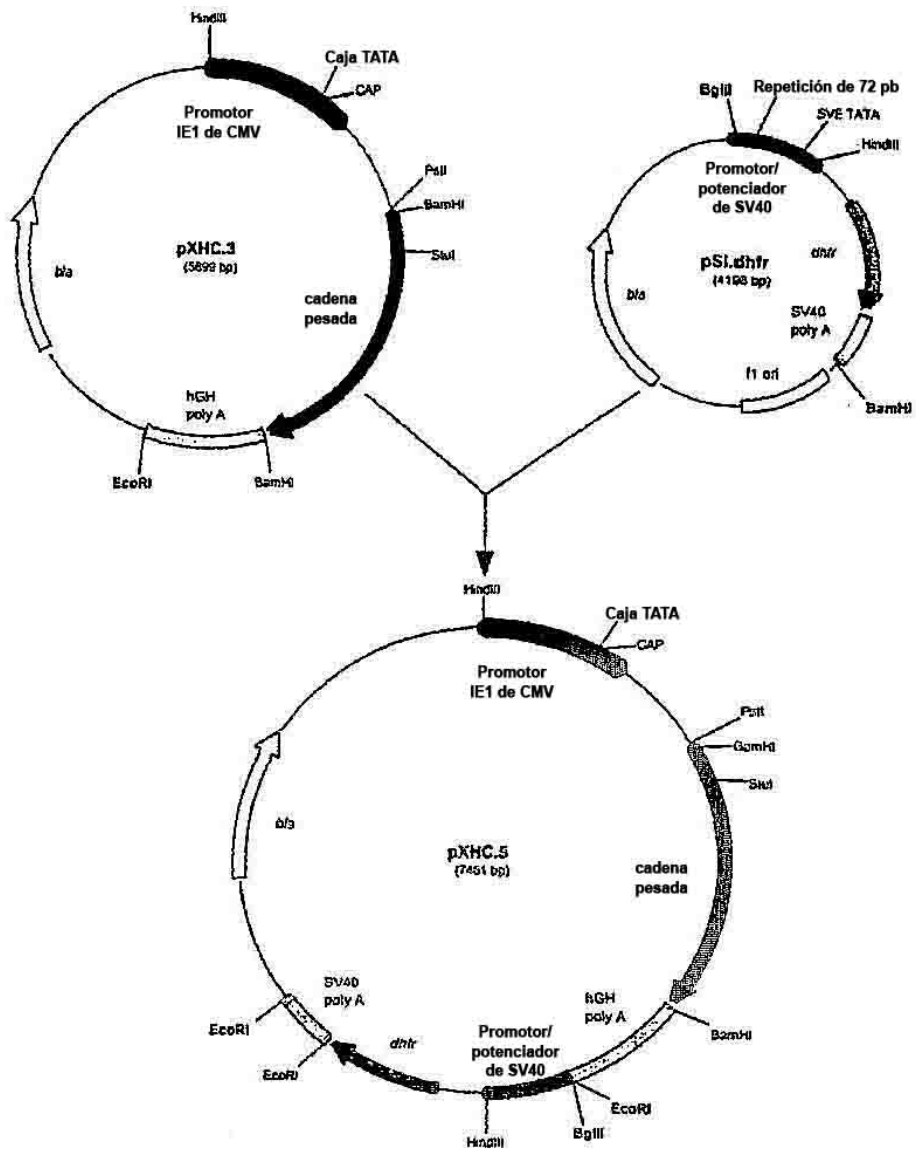


FIG. 9

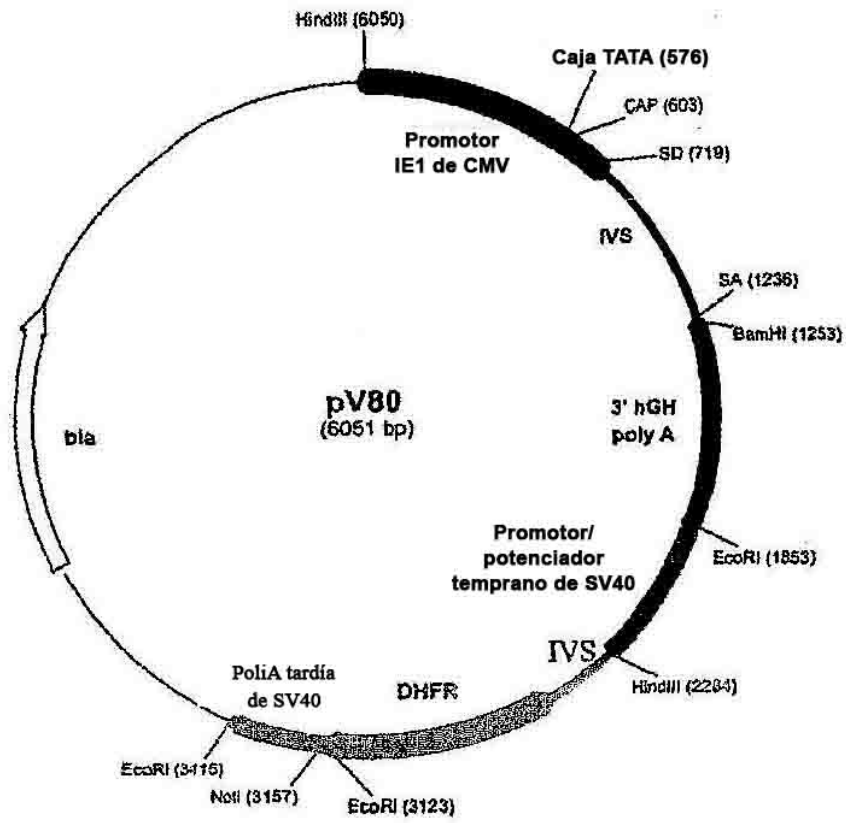


FIG. 10

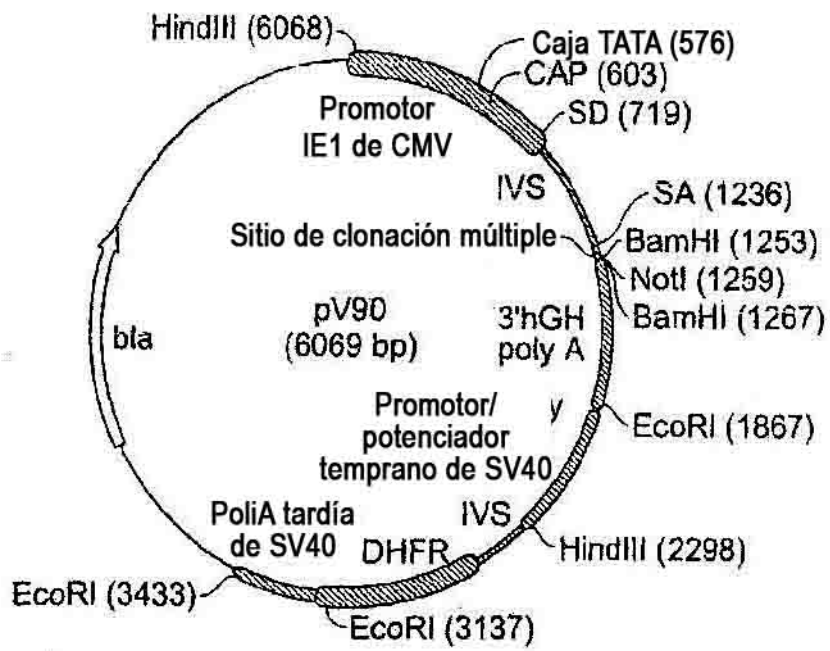


FIG. 11A

Secuencia de pV90 anotada

1 AGCTTGCATTCATTATGACTAGTTATTARTAGTARTCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGGCTTA  
87 CATRACTTACGGTAAATGGCCCGCTGCTGACCGCCACACGACCCCGCCCATTTGACGTCAATATGACGTATGTTCCCATAGTA  
173 ACGCCAATAGGGACTTCCATTGACGTCAATGGTGGAGTATTTACGGTAACTGCCACTTGGCAGTACATCAAGTGTATCATAT  
259 GCCAGTACGCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGBCATTATGCCAGTACATGACCTTATGGACTTTCCTA  
345 CTTGGCAGTACATCTACGTATYAGTCATCGCTATTACCATGGTGTATGCCGGTTTGGCAGTACATCAATGGCCGTGGATAGCGGTT  
431 GACTCACGGGGATTTCCAAAGTCTCCACCCCATTTGACGTCAATGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAATG  
Caja TATA (576)

517 TCGTACACACTCCGCCCATTTGACGCAAAATGGCCGGTAGGGGTGACGGTGGGAGGTTCTA TATAAGACAGGCTCGTTTATGTAA  
CAP (603)

681 CCGTCAGATCGCTTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAGACACCGGGACCGATCCAGCCTCCCGGGCCGGGAAAC  
687 GGTGCATTTGGAGCGGGATTCGCCGTGCCAAG  
donador de corte y empalme (719)

728 GTGAGTAAAGTACCGCTATAGAGTCTATAGCCCGACCCCTTGGCTTCTATGCTGCTATACGTTGTTTTGGCTTGGGGTCTATA  
806 GACCCCGCTTCCCTCATGTTATAGGTGATGGTATAGCTTAGCCATAGGTTGGTTATTTGACCATATTTGACCACTCCCATATG  
892 GTGACATACCTTCCATTAATCCATACATGCTCTTTCGACCACTCTCTTATTGCTATATGCCATACACTGTCCTTCA  
978 GAGACTGACACCGACTCTGATTTTACAGCATGGGCTCATTTATTTTACAAATTCACATATACACACCCCGTCCCGGT  
1064 GCCCCGATTTTTATTAACATACCTGGGATCTCCACCGCACTCTCGGATACGTGTTCCGAACTGTGAGGCAATGTAGTCTG  
1158 AGGCTACTGGTTGCTGGCCCGCCCGACAGCAATATAGCTGACAGCTACAGACTGTTCCCTTTCATGGTCTTTTCGCA  
BamHI (1253) BamHI (1267)  
Sitio de clonación múltiple

aceptor de corte y empalme (1236) NotI (1259)  
1236 @TCACCGTCTT TCACCG ATCCGGGCGGGGATCCTGCCCCGGTGGCATCCCTGTGACCCCTCCCCAGTGGC  
BamHI NotI BamHI  
sitio de poliadenilación (1359)

1312 TCTCCTGGTCTBTGG AASGTGCT ACTCCAGTGGCC ACCAGCCTTGCTCT AAT AAAATT AABTTGC ATCATTTGT  
1386 TTGACTAGBTGCTCTTGT AT AAT ATTTATGGGGTGGAGGCGGTGGT ATGGAGC AAGGGGCAGBTGGG AABACA  
1468 ACCTBTAGGGCCTTCAGGGTCT ATTGGGAACCAGGCTGGAGTGCAGTGGCAGCATCTTGGCTCGCTGC AATCTC  
1534 CGCTCTCCTGGGTGC AAGCGATTCTCCTGCCTCAGTCTCCCGAATAGTTGGGATTCAGGCACTGCACGACAGG  
1608 TCAGCTAATTTTTGTATTTTTGGT AGAGACGGGGTTTACC ATATTGGCCAGTCTGGTCTCCATCTCCTGACCT  
1682 CAGGTAAATCCCGCCCGCTCGCCCTCCAAATGCTGGGATTAACAGGTATGAGCCACTGGGCGCTTCCCTGCTCT  
1756 GTGATTTTAAAATAATTATACCAGCAGAAAGACGTCCAGACACAGCATGGGCTACCTGGCCATGCCCCAGCCAT  
SalI (1873)

Promotor/potenciador temprano de SV40 (1868)  
EcoRI (1867)

1838 TGGACATTTGAGTTGTTGCTTGGC ACTGTCTCTCATGARTCGTCAACAGATCTGGCAGCACCATGBCCTGAAATRA  
1918 CTCTGAAAGAGGACTTGGTTAGGTACTTCTGAGGCGGAAAGAGACCGCTGTGGATGTGTGTCAAGTATGGGTGCGAAGTCC  
1996 CCAGGCTCCCGACAGGCAAGGATGCAAGCATGCATCTCAATTAATCAGCAACAGGTGTGGAAATCCCGAGGCTCCCGACG  
2082 AGGCAAGGATGCAAGCATGCATCTCAATTAATCAGCAACCATAGTCCCGCCCTACTCCGCCCCTCCCGCCCTACTCCG  
2168 CCAGTTCCGCCCTTCTCCGCCCTGGCTGACTAATTTTTTTTATTTATGCAAGGCGCGAGGCGCCCTCGCCCTGAGCTATT  
HindIII (2298)

2254 CAGAAATAGTGAAGGAGCTTTTTTGGAGGCTTAGGCTTTTGCARAAAGCTTGATTTCTGACACACAGCTCTCGAATTAGCTG  
2348 CAAGATTTGGTCTGAGGCACTGGGCAAGTATGATTCAGGTTACAGGACAGGTTTACGGAGCCCAATAGAACTGGCTTGTCA  
2426 GACAGAGAGACTCTTGCSTTCTGATAGGCACTATTTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTC  
DHFR coding sequence (2568)

2512 CAGTTCATTAACAGCTCTTAGGCTAGAGTACTTARTACGACTCACTATAGGctagcATGGTTCGACCATTTGAAGTGCATCG  
2594 TCGCCGTGTCCTCAAAATATGGGGATTGGCAGAGACGGGACCTACCCGCGCTCCGCTCAGGACGAGTTCAGG  
2658 TACTTCCAAAGAAATGACCACACCTCTTCAAGTGGAGGTAACAGAACTCTGGTATTTGGGTAGGAAACCTG  
2742 GTCTCCATTCCTGAGAAAGAAATCGACCTTTAAGGAGCAAAATTAATATAGTTCTCAGTAGAGAACTCAAGAAC  
2816 CACCAAGAGGAGCTCATTTCTTCCAAAGATTTGGATGATGCCTTAGACTTATTGACACACCGAAATTTGGCA  
2898 AGTAAGTAGACATGGTTTGGATAGTCCGAGGCAATTTCTGTTTACAGGAAAGCCATGAATCAACAGGCGCACT  
2964 CAGACTCTTTGTGACAGGATCATGCAAGAAATTTGAAAGTGAACAGCTTTTCCAGAAATGATTTGGGAAAT  
3038 ATAAATCTTCCAGAAATCCAGGCGTCTCTCTGAGGTCAGGAGGAAAGGCAATCAAGTATAGTATTGAA  
EcoRI (3137) SalI (3161)

3112 GTCTACGAGAAAGAAAGACTAACTCAGAAATTCACGCGTGGTACTCTAGAGTCAACCCGGCCGGCCGCTTCGAGCAGAC

FIG. 11B - 1

3195 ATGATAGATACATTGATGAGTTTGGACAAACCCACAACTAGGATGCGATGAAAAAARTGCTTTATTTGTGAAATTTGTGATGCTAT  
sitio de poliadenilación (3308)  
3281 TGCTTTATTTGTACCATTTATAGCTGCAATRAACRAGTTTRACRACRACRARTTGCAATTCATTTTATGTTTCAGGTTCAAGGGGAGG  
EcoRI (3433)  
Sall (3427)  
3367 TGTGGGAGGTTTTTAAAGCAGSTAAACCTCTACAAATGTGGTAAATCGATAGGATCTGTGACGAAATTCAGTGGCCGTGCTT  
3453 TTACRACGTCGTGACTGGGAAACCCCTGGCGTTACCCACTTAACTCGCCTTGCAGCACATCCCCCTTTCCGCRGCTGGCGTAAATAG  
3539 CGARGAGGCCCGCACCCGATCGCCCTTCCRACAGTTGCGCAGCCTGAATGGCGAATGGCGCTGATGCGGATTTTCTCCTTACGC  
3625 ATCTGTGCGGTATTTACACCCGCATATGGTGCACCTCTCAGTACARTCTGCTCTGATGCCGCATAGTTARGCCAGCCCGACACCCG  
3711 CCAACACCCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGRACAGCTGTGACCGTCTCCGGGAGCTCCAT  
3797 GTGTCAAGGTTTTACCGCTCATCCCGAAGCGCCGAGACGAAAGGGCCCTGATGATACGCTATTTTATAGGTTAATGTCAATGA  
3883 TAATAATGGTTTTCTTAGACGTCAAGTGGCACTTTTCGGGAAATGTGCCGGAAACCCCTATTTGTTTATTTTCTAARTACATTCA  
3969 AATATGTATCCGCTCATGAGACAAATACCCGTATAAATGCTTCAATATATTTGAAARAGGAGAGATATGATATTCACATTTCCG  
4855 TGTCGCCCTTATTCCTTTTTTGGCGCATTTTGCCTTCTGTTTTTGTCTACCCCGAARACGCTGGTGAARGTAAAGATGCTGARG  
4141 ATCRGTTGGGTGCACGAGTGGTTACATCGAATCGATCTCAACAGCGGTARATCCTTGAGAGTTTTCCGCCCGAAGACGTTTT  
4227 CCARTGATGAGCACTTTTAAAGTTCTGCTRTGTGGCGGATTTATCCCGTATGACGCCGGCCAGAGCACTCGGTCGCCGAT  
4313 ACACATTTCTCAGATGACTTGGTTGAGTACTCACAGTACAGAAAGCATCTTACGGATGGCATGACAGTARAGGATTTATGCA  
4399 GTGCTGCCATACCAATGAGTGTATACACTGCGGCCAATTTACTTCTGACRACGATCGSAGGACCGAAGGAGCTACCCGCTTTTTG  
4485 CACRACATGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAAGCCATACCAACCGACGAGCGTGCACCCAC  
4571 GATGCCCTGATGCAATGGCAACACGTTGGCGAACTATTACTGCGCACTACTTACTCTAGCTTCCCGCAACAAATTAATAGACT  
4657 GGATGGAGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTTATGCTGATAAATCTGGAGCCGGT  
4743 GAGCGTGGGTCTCGCGTATCATTTGCAGCACTGGGGCAGATGGTARGCCCTCCGATCGTAGTTATCTACRACGCGGGAGTCA  
4829 GGCACTATGATGAAAGCAATAGACAGATCGCTGAGTARGGTGCCCTCACTGATTAAGCATTGGTACTGTGAGCCAGGTTACT  
4915 CATATATACTTTAGATTTGATTTAAACTTCAATTTTTAATTTAAAGGATCTAGGTGAGATCCTTTTTGATATCTCATGACCAAA  
5001 ATCCCTTACGCTGAGTTTTGCTCCACTGAGCCTCAGACCCCGTAGAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCG  
5087 CGTAACTGCTGCTTGCRAACRAAAARACCCACCGCTACCAAGCGGTGGTTTTGTTTCCCGATCAAGAGCTACCACTCTTTTTCCGA  
5173 AGGTAACTGGCTTCAAGCAGGCGCAGATACCAATATCTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA  
5259 GCACCGCTTACATACCTCGCTCTGCTARTCTGTACRATGGCTGCTGCCAGTGGCGATARGTCTGTCTTACCGGTTGGACTC  
5345 AAGACGATATTTACCGGATAGGCGCAGCGGTGCGGCTGAAACGGGGGTTCTGTGCACACAGCCAGCTTGGAGCGACACCCCTACA  
5431 CCGAATGAGATACCTACAGCGTGAAGCTATGAGAAAGCGCCACGCTTCCGAAAGGAGAAAGGCGGACAGGATCCGATAGCGGC  
5517 AGGGTGGAGACGAGAGCGCACGAGGGAGCTTCCAGGGGAAACCGCTGGTATCTTTATAGTCTGTGCGGTTTCCGACCTCTG  
5603 ACTTGAAGCTCGATTTTGTGATGCTCTGTCAGGGGGGGAGCCATGAAARACCGCAGCAGCGGCCCTTTTACGGTTCTCTG  
5689 CCTTTGCTGGCCTTTGCTCAATGTTCTTCTGCGTTATCCCTGATTCGTGTGATACCGATTTACCGCTTTTGAAGTGAECT  
5775 GATACCGCTCGCCCGACCGACCGACCGACCGAGTCAAGTGTGSCGAGGAGCGGAGAGCGCCCAATACGCAACCGCTCT  
5861 CCCCAGCGGTTGGCCGATTCATTAATGCACTGGCACGACAGGTTTTCCGACTGGAAAGCGGGCAGTGAAGCGCAACCGCAATTAATG  
5947 TGAATAGCTCACTCATTAGGCACCCAGGCTTTRACCTTATGCTTCCGGCTCGATGTTGTGTGAAATTTGAGCCGAAACCA  
6033 TTTCAACAGGAAACAGCTATGACCATGATACGCCA

FIG. 11B-2

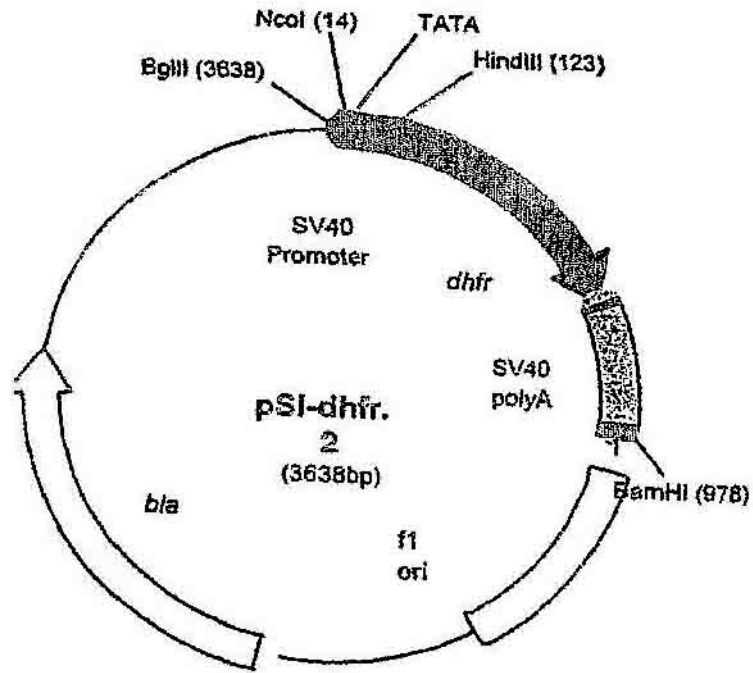


FIG. 12

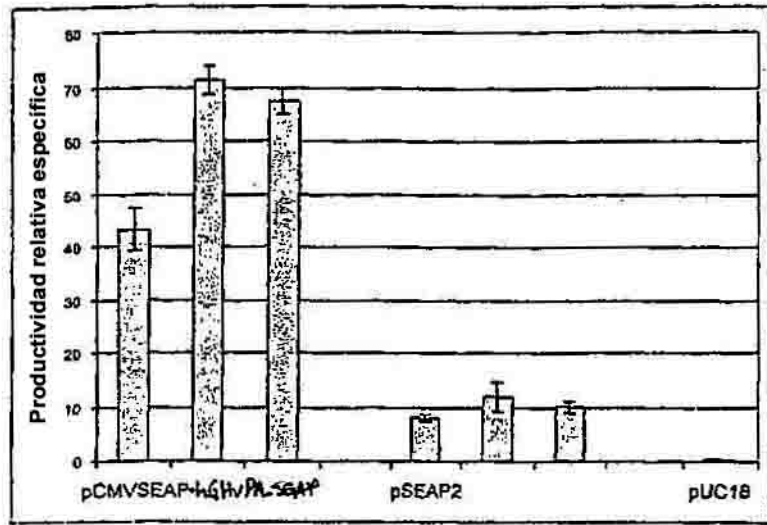
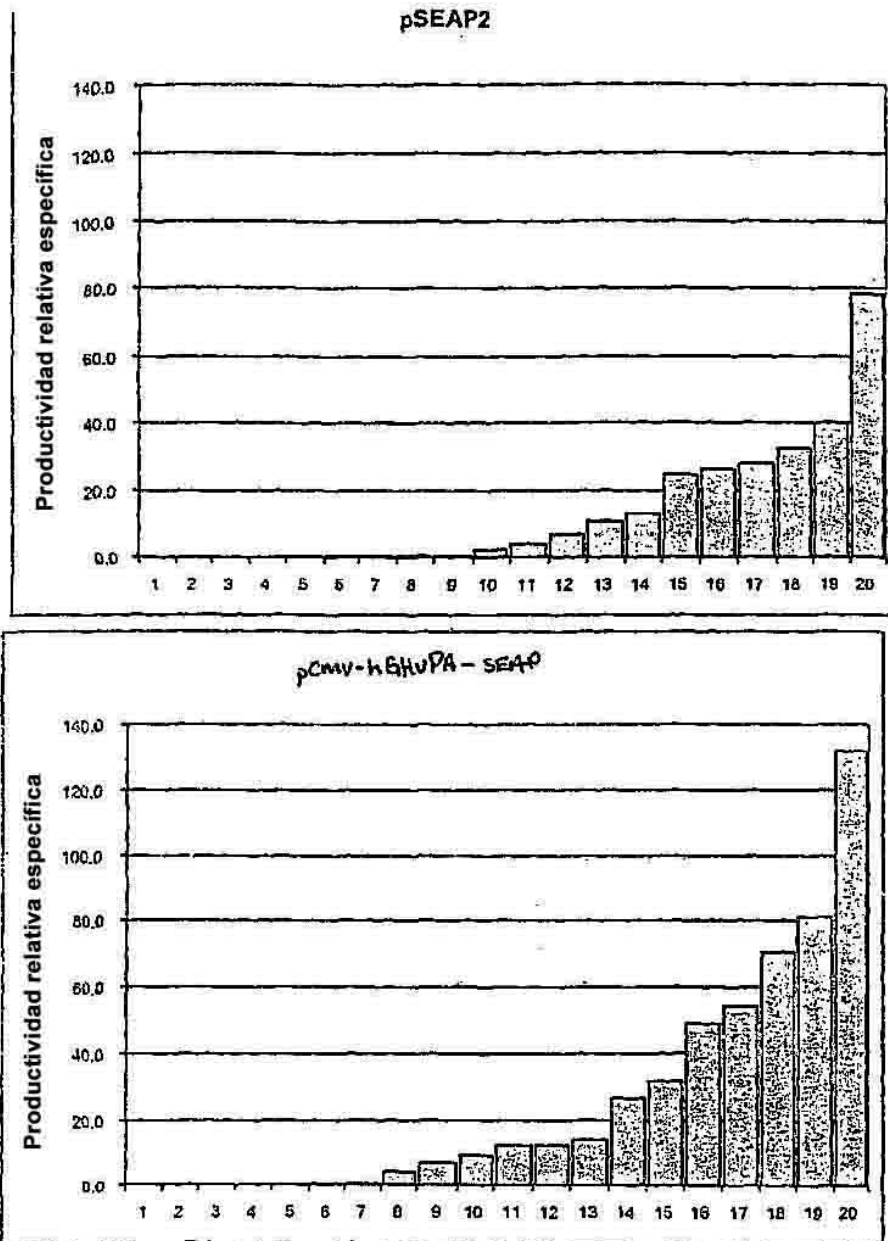


FIG. 13





Productividad relativa especifica de los aislados. Las 20 primeras muestras aisladas procedentes de cada construcción se muestran en orden de importancia.

FIG. 14

FIG. 15

2660 bp DNA linear (SEQ ID NO:18)  
 DEFINITION Homo sapiens growth hormone variant precursor (GH-V) gene, complete cda.  
 ACCESSION K00470 VERSION K00470.1 GI:183174  
sig peptide join(556..565,837..904) /gene="GH-V"  
mat peptide join(905..997,1208..1327,1419..1583,1834..2028) /gene="GH-V"  
 /product="growth hormone variant"  
 exon 837..997 /gene="GH-V" /number=2  
 exon 1208..1327 /gene="GH-V" /number=3  
 exon 1419..1583 /gene="GH-V" /number=4  
 exon 1834..2136 /gene="GH-V" /number=5  
 polyA signal 2118..2123 /gene="GH-V"  
 repeat region 2234..2505 /rpt\_family="Alu" BASE COUNT 593 a 719 c 753 g 595 t

```

gaattcagca ctgaatcatg cccagaaccc ccgcaatcta ttggctgtgc tttggcccct 61
ttcccacaac cacacattct gtctggtggg tggaggggaa acatgctggg aggaggaag 121
gaataggata gagagtggga tggggctcgt aggggtctca aggactggcc tatcctgaca 181
tccttctccg cgttcagggt ggccaccatg gcctgtgccc agagggcacc cacgtgacc 241
ttaagagag gagcaagtgg gtggtatctc tggctgacat tctgtgcaca accctcaca 301
cgctggtgat ggtgggaagg gaaagatgac aagtcagggg gcatgatccc agcatgtgtg 361
ggaggagctt ctaaattatc cattagcaca agcccgtcag tggccccagg cctaaacatg 421
cagagaaaca ggtgaggaga agcagcgaga gagaaggggc caggtataaa aagggcccac 481
aagagaccag ctcaaggatc ccaaggccca actccccgaa ccactcaggg tctgtgggac 541
agctcactag cggcaatggc tgcaggtaag cgcccctaaa atcccotttg cacaatgtgt 601
cctgagggga gaggcggcgt cctgtatagtg ggacgggggc actaacccctc aggtttgggg 661
cttatgaatg ttagctatcg ccatetaagc ccagtatttg gccaatctct gaatgttctt 721
ggtcccaggg ggaggcagag agagegagag agaaaaaaaa aaccagctc ctggaacagg 781
gagagcgtg gccctctgct ctccagctcc ctctgttggc tccggttctt ccccaggctc 841
ccggacgtcc ctgtccctgg cttttggcct gctctgcttg tcttggcttc aagagggcag 901
tgccctccca accattccct tatccaggct ttttgacaac gctatgctcc gcgccectcg 961
cctgtaccag ctggcatatg acacctatca ggagtttcta agctcttggg taatgggtgc 1021
gcttcagagg tggcaggaag ggggtgaattt cccccgctgg gaagtaatgg gaggagacta 1081
aggagctcag ggtgtgtttc tgaagtgaat atgcaggcag atgagcatac gctgagttag 1141
gttcccagaa aagtaacaat gggagcaggt ctccagcata gaccttgggt ggcggtctct 1201
tcctctaggaa gaagcctata tcctgaagga gcagaagtat tcatctctgc agaaccctca 1261
gacctccctc tgcttctcag agtctattcc aacaccttcc aacaggggtg aaacgcagca 1321
gaaatctgtg agtggatgce ttctcccag gtgggatggg gtagacctgt ggtcagagcc 1381
cccgggcagc acagccactg ccggctcctc cctgacagaa cctagagctg ctccgcactc 1441
ccctgctgct catccagtca tggctggagc ccgtgcagct cctcaggagc gtcttgcgca 1501
acagcctggg gtatggcgcc tcggacagca acgtctatcg ccacctgaag gacctagagg 1561
aaggcatcca aacgtgatg tgggtgaggg tggcaccagg atccaatcct ggggcccacc 1621
tggcttccag ggactgggga gagaaacact gctgcectct ttttagcagt caggcgctga 1681
cccaagagaa ctaccggtat tcttcaattc ccctcgtgaa tcctccaggc ctttctctac 1741
aacctggagg ggagggagga aaatggatga atgagagagg gagggaacag tgcccaagcg 1801
cttggcctct ccttctcttc cttcactttg cagaggttgg aagatggcag cccccggact 1861
gggcagatct tcaatcagtc ctacagcaag ttgacacaa aatcgcacaa cgatgacgca 1921
ctgctcaaga actacgggct gctctactgc ttcaggaaag acatggacaa ggtcagagca 1981
ttcctgcgca tctgacagtg ccgctctgtg gagggcagct gtggcttcta gctgcccggg 2041
tggcatcctt gtgacccctc cccagtgccct ctccctggctg tggaaagggtc tactccagtg 2101
cccaccagcc ttgtcctaataaaaattaaagt tgcatacttt tgtttgacta ggtgtccttg 2161
tataatatta tggggtggag gcgggtggta tggagcaagg ggccagggtg ggaagacaac 2221
ctgtagggcc ttcagggtct attcgggaac caggctggag tgcagtggca gtcttggctc 2281
gctgcaatct cgcctcctg ggttcaagcg atttctctgc ctccagtctcc cgaatagtctg 2341
cgattccagg catgcaagac caggctcagc taatttttgt atttttggta gagacggggt 2401
ttcaccatct tggccagtct ggtctccatc tcttgacctc agttaatccg cccgcctcgg 2461
cctcccacat tcttgggatt acaggtatga gccactgggc ccttccctgt cctgtgattt 2521
taaaataat ataccagcag aaggacgtcc agacacagca tgggctacct ggccatgccc 2581
agccagttgg acatttgagt tgtttgcttg gcactgtcct ctcatgcatt gggctccactc 2641
agtagatgct tgttgaattc
    
```