

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 679 291**

51 Int. Cl.:

C12N 15/09 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.12.2012 PCT/IB2012/056977**

87 Fecha y número de publicación internacional: **13.06.2013 WO13084157**

96 Fecha de presentación y número de la solicitud europea: **05.12.2012 E 12816132 (0)**

97 Fecha y número de publicación de la concesión europea: **27.06.2018 EP 2788483**

54 Título: **Casete de expresión**

30 Prioridad:
07.12.2011 US 201161567675 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
23.08.2018

73 Titular/es:
**GLENMARK PHARMACEUTICALS S.A. (100.0%)
Chemin de la Combeta, 5
2300 La Chaux-de-Fonds, CH**

72 Inventor/es:
**LUESCHER, DANIEL;
AEBISCHER-GUMY, CHRISTEL;
MORETTI, PIERRE y
BERTSCHINGER, MARTIN**

74 Agente/Representante:
CAMPello ESTEBARANZ, Reyes

ES 2 679 291 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

Casete de expresión

5 Campo de la invención

La presente invención se refiere a un casete de expresión útil para la expresión de una secuencia de polinucleótidos que codifica un polipéptido. La presente invención también se dirige a vectores y células huésped que comprenden el casete de expresión y usos del casete de expresión para la producción de un polipéptido a partir de una célula huésped.

Antecedentes de la invención

Los sistemas de expresión para la producción de polipéptidos recombinantes se conocen bien en el estado de la técnica y se describen por, por ejemplo, Marino MH (1989) *Biopharm*, 2: 18-33; Goeddel DV et al., (1990) *Methods Enzymol* 185: 3-7; Wurm F & Bernard A (1999) *Curr Opin Biotechnol* 10: 156-159. Los polipéptidos para su uso en aplicaciones farmacéuticas se producen preferiblemente en células de mamífero tales como células CHO, células NSO, células SP2/0, células COS, células HEK, células BHK, o similares. Los elementos esenciales de un vector de expresión usado para este propósito normalmente se seleccionan de una unidad de propagación de plásmido procarionta, por ejemplo, *E. coli*, que comprende un origen de replicación procarionta y un marcador de selección procarionta, opcionalmente un marcador de selección eucariota, y uno o más casetes de expresión para la expresión de uno o más genes estructurales de interés, comprendiendo cada uno un promotor, una secuencia de polinucleótidos que codifica un polipéptido y, opcionalmente, un terminador de la transcripción que incluye una señal de poliadenilación. Para la expresión transitoria en células de mamífero puede incluirse un origen de replicación de mamífero, tal como SV40 Ori u OriP. Como promotor, se puede seleccionar un promotor constitutivo o inducible. Para la transcripción optimizada, se puede incluir una secuencia Kozak en la región 5' no traducida. Para el procesamiento de ARNm, en particular el corte y empalme de ARNm y la terminación de la transcripción, se pueden incluir señales de corte y empalme de ARNm, dependiendo de la organización del gen estructural (organización de exón/intrón), así como una señal de poliadenilación. La expresión de un gen se realiza de forma transitoria o usando una línea celular estable. El nivel de expresión estable y alta de un polipéptido en una línea celular de producción es crucial para el proceso global de la producción de polipéptidos recombinantes. La demanda de moléculas biológicas tales como proteínas y específicamente anticuerpos o fragmentos de anticuerpos ha aumentado significativamente en los últimos años. El alto coste y el bajo rendimiento han sido factores limitantes en la disponibilidad de moléculas biológicas y ha sido un gran desafío desarrollar procesos robustos que aumenten el rendimiento de moléculas biológicas deseables a escala industrial.

El documento EP1308510 enseña el uso de todo el promotor GAPDH para el cribado de agentes inductores de la apoptosis. En todas las realizaciones descritas en la solicitud citada, se usa el promotor completo. Incluso cuando las secuencias no traducidas se incluyen en las construcciones, se usan junto con el promotor completo.

Por lo tanto, todavía existe la necesidad de mejorar la eficacia de los vectores de expresión para obtener alta expresión en la producción de polipéptidos recombinantes.

Resumen de la invención

La presente invención se refiere en general a sistemas de expresión tales como casetes de expresión y vectores de expresión que pueden usarse para obtener una expresión aumentada en la producción de polipéptidos recombinantes. En un aspecto, la presente descripción proporciona un casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor de Gliceraldehído 3-fosfato deshidrogenasa (GAPDH) eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región que abarca desde la posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos.

En un aspecto adicional, la presente descripción proporciona un casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas

arriba de un promotor GAPDH eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de 200 a aproximadamente 2800 nucleótidos, con la condición de que el casete de expresión no comprenda un promotor GAPDH eucariota o fragmentos del mismo.

En un aspecto adicional, la presente descripción proporciona un vector de expresión que comprende un casete de expresión y una célula huésped que comprende un casete de expresión o un vector de expresión que comprende un casete de expresión.

En aún aspectos adicionales, la presente descripción proporciona un método *in vitro* para la expresión de un polipéptido, que comprende transfectar una célula huésped con un casete de expresión o un vector de expresión y recuperar el polipéptido y el uso de un casete de expresión o un vector de expresión para la expresión de un polipéptido heterólogo de una célula huésped de mamífero.

Breve descripción de las figuras

La **Figura 1** muestra la construcción de la expresión indicadora (REP) que consiste en el promotor de citomegalovirus de ratón (mCMV), fragmento aceptor de donante de Ig (IgDA) que contiene el primer intrón, cadena ligera de anticuerpo IgG1 (IgG1 LC), sitios de entrada ribosómica interna derivados del virus de la encefalomiocarditis (IRES), cadena pesada del anticuerpo IgG1 (IgG1 HC), proteína verde fluorescente (GFP) y señal de poliadenilación del virus simio 40 (poli (A)).

La **Figura 2** muestra la expresión transitoria del anticuerpo IgG1 en células CHO-S el día 5 después de la transfección (la media de títulos de IgG se representa gráficamente para dos transfecciones independientes). Las células se transfectaron usando los vectores GAPDH_A y GAPDH_B (GAPDH_A y GAPDH_B), los mismos vectores sin elementos GAPDH aguas arriba y aguas abajo (A y B) y el vector pGLEX41 como control (pGLEX41). La concentración del anticuerpo IgG1 acumulado en el sobrenadante se determinó usando el instrumento Octet (Fortebio, Menlo, CA, Estados Unidos).

La **Figura 3** muestra la expresión del anticuerpo IgG1 en células HEK293 EBNA. Las células se transfectaron usando los vectores GAPDH_A y GAPDH_B (GAPDH_A y GAPDH_B) y el vector pGLEX41 como control (pGLEX41). El sobrenadante se recolectó y se analizó el día 10 después de la transfección usando el instrumento Octet. Los datos representan N = 3 transfecciones independientes en tubospin por vector.

La **Figura 4** muestra un estudio de nivel de expresión en una producción por lotes que usa grupos celulares. Las células se transfectaron y se crearon grupos de células estables usando vectores GAPDH_A y GAPDH_B (GAPDH_A(1), GAPDH_A(2), GAPDH_B(1) y GAPDH_B(2)), los mismos vectores sin los elementos de GAPDH aguas arriba y aguas abajo (A(1) y A(2)) y el vector pGLEX41 como control (pGLEX41). Después de 7 días de cultivo, el sobrenadante se analizó usando el instrumento Octet para el anticuerpo acumulado en el sobrenadante. La media de los títulos de IgG se da ($\mu\text{g/ml}$) para cada grupo. Los datos representan N = 2 lotes por grupo.

La **Figura 5** muestra un estudio de nivel de expresión en poblaciones generadas por transfección estable y dilución limitante. Las células se transfectaron usando los vectores GAPDH_A y GAPDH_B (GAPDH_A y GAPDH_B), los mismos vectores sin los elementos GAPDH aguas arriba y aguas abajo (A y B) y el vector pGLEX41 como control (pGLEX41). El valor medio de la fluorescencia de GFP expresado por clones y minigrupos de transfecciones estables se leyó 14 días después de la transfección. Las células se cultivaron bajo presión de selección en placas de 96 pocillos. Los datos representan N = 48 clones o minigrupos por vector.

La **Figura 6** muestra el efecto de los aditivos de medio insulina y PMA (forbol 12-miristato 13-acetato, un éster de forbol) sobre la expresión del anticuerpo IgG1 en el sobrenadante. Después de la transfección con el vector GAPDH_A (GAPDH_A) y el vector pGLEX41 como control (pGLEX41), las células se diluyeron en medio PowerCHO2, Gln 4 mM, +/- insulina y PowerCHO2, Gln 4 mM, PMA +/- insulina. No se pudo observar diferencia en la expresión en comparación con el medio estándar para pGLEX41 (barras de color negro) o GAPDH_A (barras de color blanco).

La **Figura 7** muestra una visión general del locus GAPDH humano. El gen GAPDH está flanqueado por los genes NCAPD2 e IFFO1.

La **Figura 8** muestra detalles del gen GAPDH humano, los elementos GAPDH aguas arriba y aguas abajo y los fragmentos creados para el análisis del estudio de fragmentación aguas arriba de GAPDH. El sitio de restricción NruI se introdujo para facilitar las etapas de clonación y no es parte de la secuencia genómica 5'

de GAPDH aguas arriba (por lo tanto, se resalta usando un asterisco). Los tamaños de los fragmentos son: Fragmento 1 (SEQ ID NO: 9): 511 bps, Fragmento 2 (SEQ ID NO: 10): 2653 bps, Fragmento 3 (SEQ ID NO: 11): 1966 bps, Fragmento 4 (SEQ ID NO: 12): 1198 pb, Fragmento 8 (SEQ ID NO: 13): 259 pb, Fragmento 9 (SEQ ID NO: 14): 1947 pb, Fragmento 11 (SEQ ID NO: 15): 1436 pb, y Fragmento 17 (SEQ ID NO: 16): 1177 pb.

La **Figura 9** muestra los resultados de expresión de la fragmentación de los elementos de GAPDH aguas arriba y aguas abajo. Los resultados de expresión se obtuvieron en la transfección transitoria en células CHO el día 10 después de la transfección. La cuantificación se realizó utilizando el instrumento Octet. El vector pGLEX41 sirve como control negativo. pGLEX41-amp^r también es un control negativo que muestra la expresión inicial del vector sin los elementos flanqueantes de GAPDH. pGLEX41-ascendente/descendente contiene las regiones flanqueantes de longitud completa (aguas arriba y aguas abajo) y sirve como control positivo. pGLEX41-ascendente contiene solo la región flanqueante aguas arriba y pGLEX41-descendente contiene solo la región flanqueante aguas abajo. Todas las demás construcciones contienen los fragmentos descritos en la Figura 8. Los fragmentos 2 y 3 se clonaron en la misma dirección que IgG1 LC e IgG1 HC o en dirección opuesta en relación con IgG1 LC e IgG1 HC (AS).

La **Figura 10** muestra la expresión transitoria del anticuerpo IgG1 en células CHO-S el día 8 después de la transfección (la media de títulos de IgG se representa gráficamente para tres transfecciones independientes; barras de error: DE +/-). Las células se transfectaron usando vectores con el elemento aguas arriba de GAPDH de hámster chino en combinación con el CMV de ratón (A_GAPDH_UP) o el promotor GAPDH de hámster chino (A_GAPDH_UP_PR). Los plásmidos que tienen solo el CMV de ratón (A) o el promotor GAPDH de hámster chino (A_PR) se transfectaron como control. La concentración del anticuerpo IgG1 acumulado en el sobrenadante se determinó usando el instrumento Octet QK (Fortebio, Menlo, CA, Estados Unidos).

25 Descripción detallada de la invención

La presente descripción se refiere a casetes de expresión y vectores de expresión que comprenden un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor de Gliceraldehído 3-fosfato deshidrogenasa (GAPDH) eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región que abarca desde la posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos.

La presente descripción se refiere además a un casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de 200 a aproximadamente 2800 nucleótidos, con la condición de que el casete de expresión no comprenda un promotor GAPDH eucariota o fragmentos del mismo.

El término "casete de expresión" como se usa en el presente documento incluye una secuencia de polinucleótidos que codifica un polipéptido a expresar y secuencias que controlan su expresión tales como un promotor y opcionalmente una secuencia potenciadora, que incluye cualquier combinación de elementos de control transcripcional de actuación en cis. Las secuencias que controlan la expresión del gen, es decir, su transcripción y la traducción del producto de transcripción, se denominan comúnmente unidad reguladora. La mayoría de las partes de la unidad reguladora se localizan aguas arriba de la secuencia de codificación del gen y están operativamente unidas a la misma. El casete de expresión también puede contener una región 3' no traducida aguas abajo que comprende un sitio de poliadenilación. La unidad reguladora de la invención está unida operativamente al gen que se va a expresar, es decir, la unidad de transcripción, o está separada de la misma por un ADN intermedio tal como, por ejemplo, por la región 5' no traducida del gen heterólogo. Preferiblemente, el casete de expresión está flanqueado por uno o más sitios de restricción adecuados para permitir la inserción del casete de expresión en un vector y/o su escisión en un vector. Por lo tanto, el casete de expresión de acuerdo con la presente invención se puede usar para la construcción de un vector de expresión, en particular un vector de expresión de mamífero. El

casete de expresión de la presente invención puede comprender una o más, por ejemplo, dos, tres o incluso más secuencias de ADN genómico no traducidas aguas abajo de un promotor GAPDH eucariota o fragmentos del mismo, y/o una o más, por ejemplo, dos, tres o incluso más secuencias de ADN genómico no traducidas aguas arriba de un promotor GAPDH eucariota o fragmentos del mismo. Si el casete de expresión de la presente invención comprende más de una secuencia de ADN aguas abajo y/o aguas arriba de un promotor GAPDH eucariota o fragmentos del mismo, estas secuencias de ADN pueden estar unidas directamente, es decir, pueden comprender secuencias de unión, por ejemplo, secuencias enlazadoras que contienen sitios de restricción que están unidos a los extremos 5' y 3' y que permiten la clonación secuencial cómoda de las secuencias o fragmentos de las mismas. Como alternativa, las secuencias de ADN aguas abajo y/o aguas arriba de un promotor GAPDH eucariota o fragmentos de las mismas pueden no estar directamente unidas, es decir, pueden clonarse con secuencias de ADN intermedias.

El término "secuencia de polinucleótidos que codifica un polipéptido" como se usa en el presente documento incluye ADN que codifica un gen, preferiblemente un gen heterólogo que expresa el polipéptido.

Los términos "secuencia de codificación heteróloga", "secuencia génica heteróloga", "gen heterólogo", "gen recombinante" o "gen" se usan indistintamente. Estos términos se refieren a una secuencia de ADN que codifica un producto de proteína recombinante, en particular de proteína heteróloga recombinante, que se pretende expresar en una célula huésped, preferiblemente en una célula de mamífero y se cosecha. El producto del gen puede ser un polipéptido. La secuencia del gen heterólogo no está presente naturalmente en la célula huésped y se deriva de un organismo de la misma especie o de una diferente y puede modificarse genéticamente.

Los términos "proteína" y "polipéptido" se usan indistintamente para incluir una serie de residuos de aminoácidos conectados entre sí por enlaces peptídicos entre los grupos alfa-amino y carboxi de residuos adyacentes.

El término "secuencia de ADN genómico no traducida" como se usa en el presente documento incluye ADN que constituye información genética de un organismo. El genoma de casi todos los organismos es ADN, con la única excepción de algunos virus que tienen un genoma de ARN. Las moléculas de ADN genómico en la mayoría de los organismos están organizadas en complejos ADN-proteína llamados cromosomas. El tamaño, el número de cromosomas y la naturaleza del ADN genómico varían entre diferentes organismos. Los genomas de ADN virales pueden ser de cadena sencilla o doble, lineales o circulares. Todos los demás organismos tienen genomas de ADN bicatenario. Las bacterias tienen un solo cromosoma circular. En eucariotas, la mayoría del ADN genómico se encuentra dentro del núcleo (ADN nuclear) como cromosomas lineales múltiples de diferentes tamaños. Las células eucariotas contienen adicionalmente ADN genómico en las mitocondrias y, en plantas y eucariotas inferiores, los cloroplastos. Este ADN generalmente es una molécula circular y está presente como copias múltiples dentro de estos orgánulos. Una secuencia de ADN genómico no traducida normalmente no está unida operativamente a un promotor y, por lo tanto, no está traducida. Puede contener uno o más genes que no están traducidos, por lo tanto, uno o más genes que codifican, por ejemplo, una proteína que no se expresa.

El término "secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota" como se usa en el presente documento corresponde al ADN 3' genómico eucariota no traducido de un promotor GAPDH eucariota. La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota normalmente comienza en la posición de nucleótido de aproximadamente +1, preferiblemente en la posición de nucleótido +1, en la que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, es decir, relativa al origen del inicio de la transcripción del gen eucariota que codifica GAPDH. La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota es normalmente del mismo origen que el promotor GAPDH eucariota, por ejemplo, si el promotor GAPDH es de origen humano, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH humano también es de origen humano y corresponde a la secuencia de ADN genómico humano de origen natural aguas abajo del promotor GAPDH humano.

El término "secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota" como se usa en el presente documento corresponde al ADN 5' genómico eucariota no traducido de un promotor GAPDH eucariota. La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota normalmente comienza en una posición de nucleótido alrededor del extremo 5' del promotor GAPDH eucariota, preferiblemente en la posición de nucleótido inmediatamente después del extremo 5' del promotor GAPDH eucariota. La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota es normalmente del mismo origen que el promotor GAPDH eucariota, por ejemplo, si el promotor GAPDH es de origen humano la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH humano es también de origen humano y corresponde a la secuencia de ADN genómico humana de origen natural aguas arriba del promotor GAPDH humano.

Las posiciones del promotor GAPDH eucariota, la secuencia de ADN genómico no traducida aguas abajo o aguas arriba del promotor GAPDH eucariota, y otras secuencias de ADN como se indica en el presente documento son relativas al inicio de la transcripción del ARNm de GAPDH, por ejemplo, son relativas al origen del inicio de la transcripción de la GAPDH eucariota si no se indica específicamente de otra manera.

El término "secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota se extiende a" o "secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota se extiende a" se usa para definir la extensión de la longitud de la secuencia de ADN genómico no traducida aguas arriba y/o aguas abajo de un promotor GAPDH eucariota desde el inicio hasta un elemento genético particular, por ejemplo, extensión a un intrón. Esta extensión incluye la longitud completa de la secuencia de ADN que codifica el elemento genético, por ejemplo, el intrón o una parte del mismo.

El promotor GAPDH eucariota y el ADN genómico eucariota aguas arriba y/o aguas abajo del promotor GAPDH se pueden encontrar para seres humanos, ratas y ratones en el banco de datos público NCBI (las entradas para el gen GAPDH humano, de ratón, de rata y de hámster chino son genes ID 2597 (ARNm: NM_002046.3), 14433 (ARNm: NM_008084.2), 24383 (ARNm: NM_017008.3) y 100736557 (ARNm: NM_001244854.2), respectivamente; National Center for Biotechnology Information (NCBI): <http://www.ncbi.nlm.nih.gov/>) y se muestran de forma ilustrativa en la Figura 7 y 8 para el gen GAPDH humano.

El promotor GAPDH eucariota se considera habitualmente que se extiende desde aproximadamente los pb -500 a aproximadamente +50 con respecto al inicio de la transcripción del ARNm de GAPDH. El promotor GAPDH humano se encuentra en el cromosoma 12. El promotor GAPDH humano es considerado por Graven *et al.* (Graven et al., (1999) *Biochimica et Biophysica Acta*, 147: 203-218) para extenderse desde los pb -488 a +20 en relación con el comienzo de la transcripción del ARNm de GAPDH basándose en un estudio de fragmentación. De acuerdo con el banco de datos público NCBI, el promotor GAPDH humano se extiende desde los pb -462 a +46 en relación con el inicio de la transcripción del ARNm de GAPDH como se define en el banco de datos público NCBI. Si no se indica específicamente lo contrario, el promotor GAPDH humano como se menciona en el presente documento se extiende desde -462 hasta la posición +46 con respecto al inicio de la transcripción del ARNm de GAPDH que corresponde a la secuencia que se extiende desde los pb 4071 hasta 4578 de SEQ ID NO: 17.

La numeración utilizada para el ADN del gen GAPDH, el gen IFF01 y el gen NCAPD2 de origen humano, de ratón y de rata, como se refiere en el presente documento, corresponde a la numeración utilizada para estos genes en el banco de datos público NCBI (<http://www.ncbi.nlm.nih.gov/>).

El término "promotor" como se usa en el presente documento define una secuencia reguladora de ADN generalmente localizada aguas arriba de un gen que media el inicio de la transcripción al dirigir la ARN polimerasa para que se una al ADN e inicie la síntesis del ARN.

El término "potenciador" como se usa en el presente documento define una secuencia de nucleótidos que actúa para potenciar la transcripción génica independientemente de la identidad del gen, la posición de la secuencia en relación con el gen, o la orientación de la secuencia. Los vectores de la presente invención incluyen opcionalmente potenciadores.

Los términos "unidas funcionalmente" y "unidas operativamente" se usan indistintamente y se refieren a una relación funcional entre dos o más segmentos de ADN, en particular secuencias génicas a expresar y aquellas secuencias que controlan su expresión. Por ejemplo, una secuencia promotora y/o potenciadora, que incluye cualquier combinación de elementos de control transcripcional de actuación en cis, está unida operativamente a una secuencia codificante si estimula o modula la transcripción de la secuencia codificante en una célula huésped apropiada u otro sistema de expresión. Las secuencias reguladoras del promotor que están unidas operativamente a la secuencia del gen transcrito están físicamente contiguas a la secuencia transcrita.

"Orientación" se refiere al orden de los nucleótidos en una secuencia de ADN dada. Por ejemplo, una orientación de una secuencia de ADN en dirección opuesta en relación con otra secuencia de ADN es aquella en la que el orden de 5' a 3' de la secuencia en relación con otra secuencia se invierte cuando se compara con un punto de referencia en el ADN del que se obtuvo la secuencia. Dichos puntos de referencia pueden incluir la dirección de transcripción de otras secuencias de ADN especificadas en el ADN de origen y/o el origen de replicación de vectores replicables que contienen la secuencia.

El término "vector de expresión" como se usa en el presente documento incluye una molécula de ADN aislada y purificada que tras la transfección en una célula huésped apropiada proporciona una expresión de alto nivel de un producto génico recombinante dentro de la célula huésped. Además de la secuencia de ADN que codifica el producto genético o recombinante, el vector de expresión comprende secuencias de ADN reguladoras que se requieren para una transcripción eficaz de la secuencia de codificación de ADN en ARNm y para una traducción eficaz de los ARNm en proteínas en la línea celular huésped.

Los términos "célula huésped" o "línea celular huésped" como se usan en el presente documento, incluyen cualquier célula, en particular células de mamífero, que sea capaz de crecer en cultivo y expresar una proteína de producto recombinante deseada.

El término "fragmento" como se usa en el presente documento incluye una secuencia de nucleótidos respectiva, por ejemplo, una porción de la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota o una porción de la secuencia de nucleótidos que codifica un elemento genético particular tal como un promotor. Los fragmentos de una secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota pueden conservar la actividad biológica y, por lo tanto, alterar, por ejemplo, aumentar los patrones de expresión de las secuencias codificantes unidas operativamente a un promotor. Los fragmentos de una secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota pueden variar de nucleótidos de al menos aproximadamente 100 a aproximadamente 3000 pb, preferiblemente de aproximadamente 200 a aproximadamente 2800 pb, más preferiblemente de aproximadamente 300 a aproximadamente 2000 pb, en particular, nucleótidos de aproximadamente 500 a aproximadamente 1500 pb. Con el fin de clonar los fragmentos de la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota en el casete de expresión de la presente invención, habitualmente las secuencias enlazadoras que contienen sitios de restricción que permiten una clonación cómoda se unen a los extremos 5' y 3' de los fragmentos.

El término "identidad de secuencia de nucleótidos" o "secuencia de nucleótidos idéntica" como se usa en el presente documento incluye el porcentaje de nucleótidos en la secuencia candidata que son idénticos a la secuencia de nucleótidos de, por ejemplo, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota, después de alinear las secuencias e introducir huecos, si es necesario, para lograr el porcentaje máximo de identidad de secuencia. Por lo tanto, la identidad de secuencia puede determinarse mediante métodos estándar que se usan comúnmente para comparar la similitud en la posición de los nucleótidos de dos secuencias de nucleótidos. Habitualmente, la identidad de secuencia de nucleótidos de la secuencia candidata con respecto a la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota es de al menos el 80 %, preferiblemente al menos el 85 %, más preferiblemente al menos el 90 %, y mucho más preferiblemente al menos el 95 %, en particular 96 %, más particular el 97 %, incluso más particular el 98 %, mucho más particular el 99 %, incluyendo, por ejemplo, el 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, y el 100 %.

El término "sitio CpG" como se usa en el presente documento incluye regiones de ADN en las que se produce un nucleótido de citosina junto a un nucleótido de guanina en la secuencia lineal de bases a lo largo de su longitud. "CpG" es la abreviatura de "-C-fosfato-G-", es decir, citosina y guanina separadas por un solo fosfato; el fosfato une dos nucleósidos juntos en el ADN. La notación "CpG" se usa para distinguir esta secuencia lineal del emparejamiento de bases de CG de citosina y guanina.

El término "uso de codones alternativos" como se usa en el presente documento incluye el uso de codones alternativos que codifican el mismo aminoácido con el fin de evitar el motivo de la secuencia CpG. Esto incluye el uso preferiblemente de codones que no tengan un sitio CpG interno (por ejemplo, codificación GCG para alanina y que contenga un sitio CpG, podría reemplazarse por GCT, GCC o GCA), así como evitar la unión de dos codones que conduce a un nuevo sitio CpG.

El término "aproximadamente" como se usa en el presente documento en relación con la longitud de una secuencia de ADN y en relación con una posición de nucleótido que es relativa al inicio de la transcripción del ARNm de GAPDH, por ejemplo, es relativa al origen del inicio de la transcripción de la GAPDH eucariota incluye valores con desviaciones de un máximo de ± 50 %, usualmente de un máximo de ± 10 % de los valores establecidos, por ejemplo, "aproximadamente 3000 nucleótidos" incluye valores de 2700 a 3300 nucleótidos, preferiblemente de 2900 a 3100 nucleótidos, más preferiblemente de 2995 a 3005 nucleótidos, "aproximadamente 100 nucleótidos" incluye valores de 50 a 150 nucleótidos, preferiblemente de 90 a 110 nucleótidos, más preferiblemente de 95 a 105 nucleótidos, "aproximadamente 15000 nucleótidos" incluye valores de 13500 a 16500 nucleótidos, preferiblemente de 14500 a

15500 nucleótidos, más preferiblemente de 14990 a 15010 nucleótidos, mucho más preferiblemente de 14995 a 15005 nucleótidos, "aproximadamente 200 nucleótidos" incluye valores de 150 a 250 nucleótidos, preferiblemente de 190 a 210 nucleótidos, más preferiblemente de 195 a 205 nucleótidos, "aproximadamente 8000 nucleótidos" incluye valores de 7200 a 8800, preferiblemente de 7500 a 8500 nucleótidos, más preferiblemente de 7990 a 8010 nucleótidos, mucho más preferiblemente de 7995 a 8005 nucleótidos, "aproximadamente 500 nucleótidos" incluye valores de 450 a 550 nucleótidos, preferiblemente de 475 a 525, más preferiblemente de 490 a 510, mucho más preferiblemente de 495 a 505 nucleótidos, "aproximadamente 5000 nucleótidos" incluye valores de 4500 a 5500 nucleótidos, preferiblemente de 4750 a 5250, más preferiblemente de 4990 a 5010, mucho más preferiblemente de 4995 a 5005 nucleótidos, "aproximadamente 1000 nucleótidos" incluye valores de 900 a 1100 nucleótidos, 10 preferiblemente de 950 a 1050, más preferiblemente de 990 a 1010, mucho más preferiblemente de 995 a 1005 nucleótidos, "aproximadamente 4500 nucleótidos" incluye valores de 4050 a 4950 nucleótidos, preferiblemente de 4250 a 4750, más preferiblemente de 4490 a 4510, mucho más preferiblemente de 4495 a 4505 nucleótidos, "aproximadamente 1500 nucleótidos" incluye valores de 1350 a 1650 nucleótidos, preferiblemente de 1450 a 1550, más preferiblemente de 1490 a 1510, mucho más preferiblemente de 1495 a 1505 nucleótidos, "aproximadamente 15 4000 nucleótidos" incluye valores de 3600 a 4400 nucleótidos, preferiblemente de 3800 a 4200, más preferiblemente de 3990 a 4010, más preferiblemente de 3995 a 4005 nucleótidos, "aproximadamente 2000 nucleótidos" incluye valores de 1800 a 2200 nucleótidos, preferiblemente de 1900 a 2100, más preferiblemente de 1990 a 2010, mucho más preferiblemente de 1995 a 2005 nucleótidos, "aproximadamente 3500 nucleótidos" incluye valores de 3150 a 3850 nucleótidos, preferiblemente de 3300 a 3700, más preferiblemente de 3490 a 3510, mucho más preferiblemente de 3495 a 3505 nucleótidos, "aproximadamente 2700 nucleótidos" incluye valores de 2430 a 2970 nucleótidos, preferiblemente de 2600 a 2800, más preferiblemente de 2690 a 2710, mucho más preferiblemente de 2695 a 2705 nucleótidos, "aproximadamente 3300 nucleótidos" incluye valores de 2970 a 3630 nucleótidos, preferiblemente de 3100 a 3500, más preferiblemente de 3290 a 3310, mucho más preferiblemente de 3295 a 3305 nucleótidos, "aproximadamente 3200 nucleótidos" incluye valores de 2880 a 3520 nucleótidos, preferiblemente de 25 3000 a 3400, más preferiblemente de 3190 a 3210, mucho más preferiblemente de 3195 a 3205 nucleótidos, aproximadamente +7000 o aproximadamente la posición +7000 incluye las posiciones +6300 a +7700, preferiblemente las posiciones +6700 a +7300, más preferiblemente las posiciones +6990 a +7010, mucho más preferiblemente las posiciones +6995 a +7005, aproximadamente +1 o aproximadamente la posición +1 incluye las posiciones -10 a +10, preferiblemente las posiciones -5 a +5, más preferiblemente las posiciones -1 a +2, 30 aproximadamente -3500 o aproximadamente la posición -3500 incluye las posiciones -3150 a -3850, preferiblemente las posiciones -3300 a -3700, más preferiblemente las posiciones -3490 a -5010, mucho más preferiblemente las posiciones -3495 a -3505.

El término "aproximadamente" como se usa en el presente documento en relación con la numeración usada para el ADN del gen GAPDH, el gen IFF01 y el gen NCAPD2 de origen humano, de ratón y de rata como se refiere en el presente documento, o se usa en el presente documento en relación con una posición en una secuencia de un número de SEQ ID incluye valores con desviaciones de un máximo de ± 500 pb, preferiblemente ± 100 pb, más preferiblemente ± 10 pb, mucho más preferiblemente ± 5 pb.

En una realización, la presente descripción proporciona un casete de expresión que comprende un promotor, una 40 secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región que abarca de la posición de nucleótido aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de 45 GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos.

En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al segundo último intrón del gen IFF01 o a una parte del mismo. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de 50 un promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al último intrón del gen IFF01.

El gen IFF01 humano está localizado en el ADN humano aproximadamente de los pb 6665249 a 6648694 del cromosoma 12 (gen NCBI ID: 25900). En una realización, la longitud de la secuencia de ADN genómico no traducida 55 aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen IFF01 en seres humanos se extiende en su máximo a aproximadamente los pb 6650677 del cromosoma 12 que codifica el gen IFF01 en seres humanos (posición +7021). En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen IFF01 en seres humanos se extiende en su máximo a aproximadamente los pb 6657230 del cromosoma 12 que

codifica el gen IFF01 en seres humanos (posición +13574). Las secuencias de ADN genómico no traducidas aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen IFF01 en seres humanos y al segundo último intrón del gen IFF01 en seres humanos, respectivamente, se incluyen en la SEQ ID NO: 17 que muestra los pb 6657230 a 6639125 del cromosoma 12 (gen NCBI ID: 25900). La secuencia de ADN
 5 genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al último intrón se extiende a aproximadamente los pb 11553 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 17 y la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón se extiende a aproximadamente los pb 18106 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 17.

10

El gen IFF01 de ratón (gen NCBI ID: 320678) se sitúa en el ADN de ratón de aproximadamente los pb 125095259 a 125111800 del cromosoma 6. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen IFF01 en ratones se extiende en su máximo a aproximadamente los pb 125109211 del cromosoma 6 que codifica el gen IFF01 en
 15 ratones (posición +6391). En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen IFF01 en ratones se extiende en su máximo a aproximadamente los pb 125103521 del cromosoma 6 que codifica el gen IFF01 en ratones (posición +12081). Las secuencias de ADN genómico no traducidas aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón y al segundo último intrón del gen IFF01 en
 20 ratones, respectivamente, se incluyen en la SEQ ID NO: 18 que muestra los pb 125103521 a 125119832 del cromosoma 6 (gen NCBI ID: 320678). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al último intrón del gen IFF01 en ratones se extiende a aproximadamente los pb 10622 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 18 y la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón del gen IFF01 en
 25 ratones se extiende a aproximadamente los pb 16312 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 18.

El gen IFF01 de rata (gen NCBI ID: 362437) se encuentra en el ADN de rata de aproximadamente los pb 161264966 a 161282150 del cromosoma 4. En una realización, la longitud de la secuencia de ADN genómico no traducida
 30 aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen IFF01 en ratas se extiende en su máximo a aproximadamente los pb 161280937 del cromosoma 4 que codifica el gen IFF01 en ratas (posición + 5154). En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen IFF01 en ratas se extiende en su máximo a aproximadamente los pb 161279451 del cromosoma 4 que codifica el gen IFF01 en las
 35 ratas (posición +6640).

Las secuencias de ADN genómico no traducidas aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón y al segundo último intrón del gen IFF01 en ratas, respectivamente, se incluyen en la SEQ ID NO: 19 que muestra los pb 161279451 a 161290508 del cromosoma 4 (gen NCBI ID: 362437). La secuencia de
 40 ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al último intrón del gen IFF01 se extiende a aproximadamente los pb 9572 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 19 y la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón del gen IFF01 se extiende a aproximadamente los pb 11058 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 19.

45

El gen IFF01 de hámster chino (gen NCBI ID: 100753382) se encuentra en el ADN de hámster chino de aproximadamente los pb 3577293 a 3593683. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen IFF01 en hámster chino se extiende en su máximo a aproximadamente los pb 3579014 que codifica el gen IFF01 en
 50 hámster chino (posición +6883). En una realización, la longitud de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen IFF01 en hámster chino se extiende en su máximo a aproximadamente los pb 3585061 que codifican el gen IFF01 en hámster chino (posición +12930). La ubicación cromosómica aún no está anotada en el banco de datos NCBI y la información de secuencia actual contiene muchas bases desconocidas. Por lo tanto, la anotación precisa de los
 55 límites puede cambiar con la disponibilidad de información de secuencia más precisa.

Las secuencias de ADN genómico no traducidas aguas abajo de un promotor GAPDH eucariota que se extiende en su máximo al último intrón y al segundo último intrón del gen IFF01 en hámster chino, respectivamente, se incluyen en la SEQ ID NO: 29 que muestra los pb 3567932 a 3585061. La secuencia de ADN genómico no traducida aguas

abajo de un promotor GAPDH eucariota que se extiende al último intrón del gen IFFO1 se extiende a aproximadamente los pb 11083 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 29 y la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón del gen IFFO1 se extiende a aproximadamente los pb 17130 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 29.

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota se inicia en el sitio de poliadenilación de GAPDH eucariota, por ejemplo, se inicia en el primer nucleótido que codifica el sitio de poliadenilación de GAPDH eucariota. Preferiblemente, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota comienza aguas abajo del sitio de poliadenilación de GAPDH eucariota, por ejemplo, se inicia después del último nucleótido que codifica el sitio de poliadenilación de GAPDH eucariota. Incluso más preferida, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia aguas abajo del sitio de poliadenilación de GAPDH eucariota y la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al segundo último intrón del gen IFFO1.

En una realización, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región que abarca de la posición de nucleótido aproximadamente +3881 a la posición de nucleótido aproximadamente +5000, preferiblemente en una región que abarca de la posición de nucleótido de aproximadamente +3931 a la posición de nucleótido de aproximadamente +5000, más preferiblemente en una región que abarca de la posición de nucleótido de aproximadamente +4070 a la posición de nucleótido de aproximadamente +5000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH.

Una secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota que se inicia, por ejemplo, aguas abajo del sitio de poliadenilación GAPDH eucariota usado en la presente invención usualmente se inicia en una posición de nucleótido de aproximadamente la +3931, preferiblemente en una posición de nucleótido de aproximadamente +4070, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH.

En el ser humano, la secuencia de ADN genómico no traducida aguas abajo del sitio de poliadenilación de GAPDH humano se inicia en aproximadamente la posición de nucleótido +3931 (con respecto al inicio de la transcripción del ARNm de GAPDH que corresponde al pb 8463 como se muestra en la SEQ ID NO: 17). Preferiblemente, si la secuencia de ADN genómico no traducida aguas abajo del sitio de poliadenilación de GAPDH es de ser humano, la secuencia de ADN genómico no traducida aguas abajo del sitio de poliadenilación de GAPDH se inicia en aproximadamente +3931 (con respecto al inicio de la transcripción del ARNm de GAPDH; que corresponde al pb 8463 como se muestra en la SEQ ID NO: 17) y su longitud es de aproximadamente 3357 pb correspondientes a la secuencia de aproximadamente los pb 8463 a aproximadamente 11819 como se muestra en la SEQ ID NO: 17, más preferiblemente, se inicia en aproximadamente +4070 (con respecto al inicio de la transcripción del ARNm de GAPDH que corresponde al pb 8602 como se muestra en la SEQ ID NO: 17) y su longitud es de aproximadamente 3218 pb correspondientes a la secuencia de aproximadamente los pb 8602 a aproximadamente 11819 como se muestra en la SEQ ID NO: 17.

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota comprende la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 8 y 21 o fragmentos de las mismas.

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota comprende una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 8 y 21 o fragmentos de las mismas.

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota comprende una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 8 y 21 o fragmentos de las mismas.

En algunas realizaciones, la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 8 y 21 o fragmentos de las mismas, comprende cinco o menos, preferiblemente cuatro o menos, más preferiblemente tres o menos, mucho más preferiblemente dos o menos, en particular una modificación de ácido nucleico, en donde la modificación o modificaciones de ácidos nucleicos son preferiblemente una sustitución de ácido nucleico.

En una realización adicional, la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es preferiblemente de aproximadamente 200 a aproximadamente 8000 nucleótidos, más preferiblemente de aproximadamente 500 a aproximadamente 5000 nucleótidos, incluso más preferiblemente de aproximadamente 1000 a aproximadamente 4500 nucleótidos, mucho más preferiblemente de aproximadamente 1500 a aproximadamente 4000 nucleótidos, en particular de aproximadamente 2000 a aproximadamente 3500 nucleótidos, más en particular de aproximadamente 2700 a aproximadamente 3300, incluso más en particular aproximadamente 3200, mucho más en particular 3218 nucleótidos. La longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota como se define en el presente documento no incluye ninguna secuencia enlazadora añadida a la secuencia de ADN genómico no traducida.

10

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota está orientada en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido.

En una realización adicional, la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota está orientada en dirección opuesta en relación con la secuencia de polinucleótidos que codifica un polipéptido.

15

En algunas realizaciones, el casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota comprende además una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos.

20

En otra realización, el casete de expresión comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de 200 a aproximadamente 2800 nucleótidos, con la condición de que el casete de expresión no comprenda un promotor GAPDH eucariota o fragmentos del mismo.

30

En algunas realizaciones, el casete de expresión comprende además una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota, en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región una secuencia de polinucleótidos que +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos. En estas realizaciones, la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota usado es, por ejemplo, como se ha descrito anteriormente.

40

45

En una realización adicional, la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al codón de inicio del gen NCAPD2. En una realización adicional, la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al tercer último intrón del gen NCAPD2. En una realización adicional, la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al segundo último intrón del gen NCAPD2. En una realización adicional, la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de al menos 100 nucleótidos y se extiende en su máximo al último intrón del gen NCAPD2.

50

55

El gen NCAPD2 humano (gen NCBI ID: 9918) se encuentra en el ADN humano de aproximadamente los pb 6603298 a 6641132 del cromosoma 12. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen NCAPD2 en seres humanos se extiende en su máximo a aproximadamente 6640243 pb del cromosoma 12 que

codifica el gen NCAPD2 en seres humanos (posición -3414 con respecto al inicio de la transcripción del gen GAPDH que corresponde a los pb 1119 en la SEQ ID NO: 17).

5 En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen NCAPD2 in human se extiende en su máximo a aproximadamente 6639984 pb del cromosoma 12 que codifica el gen NCAPD2 en seres humanos (posición -3673 con respecto al inicio de la transcripción del gen GAPDH que corresponde a los pb 860 en la SEQ ID NO: 17).

10 En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al tercer último intrón del gen NCAPD2 en seres humanos se extiende en su máximo a aproximadamente 6639125 pb del cromosoma 12 que codifica el gen NCAPD2 en seres humanos (posición -4532 con respecto al inicio de la transcripción del gen GAPDH; que corresponde al pb 1 en la SEQ ID NO: 17).

15 La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón, al segundo último intrón y al tercer último intrón del gen NCAPD2 en seres humanos, respectivamente, se incluyen en la SEQ ID NO: 17, que muestra los pb 6657230 a 6639125 del cromosoma 12 (gen NCBI ID: 9918).

20 El gen NCAPD2 de ratón (Gen ID: 68298) se encuentra en el ADN de ratón de aproximadamente la posición 125118025 a 125141604 del cromosoma 6. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota (estimada con una longitud de 500 pb aguas arriba del inicio de la transcripción) que se extiende en su máximo al último intrón del gen NCAPD2 en ratones se extiende en su máximo a aproximadamente los pb 125118607 del cromosoma 6 que codifica el gen NCAPD2 en ratones.

En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen NCAPD2 en ratones se extiende en su máximo a aproximadamente 125118880 pb del cromosoma 6 que codifica el gen NCAPD2 en ratones.

30 En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al tercer último intrón del gen NCAPD2 en ratones se extiende en su máximo a aproximadamente 125119832 pb del cromosoma 6 que codifica el gen NCAPD2 en ratones.

35 La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón, al segundo último intrón y al tercer último intrón del gen NCAPD2 en ratones, respectivamente, se incluyen en la SEQ ID NO: 18, que muestra los pb 125103521 a 125119832 del cromosoma 6 (gen NCBI ID: 68298). La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende al último intrón se extiende a aproximadamente los pb 1226 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 18 (-3006 con respecto al inicio de la transcripción del ARNm de GAPDH de ratón). La

40 secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón se extiende a aproximadamente los pb 953 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 18 (-3279 con respecto al inicio de la transcripción del ARNm de GAPDH de ratón). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al tercer último intrón se extiende a aproximadamente 1 pb de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 18 (-4231

45 con respecto al inicio de la transcripción del ARNm de GAPDH de ratón).

El gen NCAPD2 de rata (Gen ID: 362438) se encuentra en el ADN eucariota de aproximadamente la posición 161288671 a 161310417 del cromosoma 4. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen

50 NCAPD2 en ratas se extiende en su máximo a aproximadamente 161289191 pb del cromosoma 4 que codifica el gen NCAPD2 en ratas. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen NCAPD2 en ratas se extiende en su máximo a aproximadamente 161289446 pb del cromosoma 4 que codifica el gen NCAPD2 en ratas. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor

55 GAPDH eucariota que se extiende en su máximo al tercer último intrón del gen NCAPD2 en ratas se extiende en su máximo a aproximadamente 161290508 pb del cromosoma 4 que codifica el gen NCAPD2 en ratas.

La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón, al segundo último intrón y al tercer último intrón del gen NCAPD2 en ratas, respectivamente,

se incluyen en la SEQ ID NO: 19, que muestra los pb 161279451 a 161290508 del cromosoma 4 (gen NCBI ID: 362438). La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende al último intrón se extiende a aproximadamente los pb 1318 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 19 (-3101 con respecto al inicio de la transcripción del ARNm de GAPDH de rata). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón se extiende a aproximadamente los pb 1063 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 19 (posición -3356 con respecto al inicio de la transcripción del ARNm de GAPDH de rata). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al tercer último intrón se extiende a aproximadamente 1 pb de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 19 (posición -4418 con respecto al inicio de la transcripción del ARNm de GAPDH de rata).

El gen NCAPD2 de hámster chino (Gen ID: 100753087) se encuentra en el ADN eucariota de aproximadamente la posición 3544184 a 3569879. La ubicación cromosómica no está disponible en la base de datos NCBI. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón del gen NCAPD2 en hámster chino se extiende en su máximo a aproximadamente 3569380 pb en hámster chino. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al segundo último intrón del gen NCAPD2 en hámster chino se extiende en su máximo a aproximadamente 3569131 pb en hámster chino. En una realización, la longitud de la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al tercer último intrón del gen NCAPD2 en hámster chino se extiende en su máximo a aproximadamente 3567932 pb en hámster chino.

La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende en su máximo al último intrón, al segundo último intrón y al tercer último intrón del gen NCAPD2 en hámster chino, respectivamente, se incluyen en la SEQ ID NO: 29, que muestra los pb 3567932 a 3585061. La secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota que se extiende al último intrón se extiende a aproximadamente los pb 1449 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 29 (-2752 con respecto al inicio de la transcripción de ARNm de GAPDH de hámster chino). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al segundo último intrón se extiende a aproximadamente los pb 1200 de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 29 (posición -3001 con respecto al inicio de la transcripción de ARNm de GAPDH de hámster chino). La secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota que se extiende al tercer último intrón se extiende a aproximadamente 1 pb de la secuencia de nucleótidos como se muestra por la SEQ ID NO: 29 (posición -4200 con respecto al inicio de la transcripción de ARNm de GAPDH de hámster chino).

In human la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH humano se inicia en aproximadamente la posición de nucleótido -463 (con respecto al inicio de la transcripción del ARNm de GAPDH que corresponde a los pb 4533 como se muestra en la SEQ ID NO: 17). Preferiblemente, si la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH es de ser humano, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH se inicia en aproximadamente -500 (con respecto al inicio de la transcripción del ARNm de GAPDH; que corresponde a los pb 4533 como se muestra en la SEQ ID NO: 17). Más preferiblemente, si la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH es de ser humano, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH se inicia en aproximadamente -576 (con respecto al inicio de la transcripción del ARNm de GAPDH; que corresponde a los pb 4533 como se muestra en la SEQ ID NO: 17) y su longitud es de aproximadamente 3158 pb correspondientes a la secuencia de aproximadamente los pb 800 a aproximadamente 3957 como se muestra en la SEQ ID NO: 17.

En una realización adicional, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota comprende una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 o fragmentos de las mismas, preferiblemente una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 y 16 o fragmentos de las mismas, o una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 20, 22, 23, 24, 25, 26, 27, 28 y 16 o fragmentos de las mismas. Es más preferida una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 10, 12, 15 y 16 o fragmentos de las mismas, más preferiblemente, una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 10, 12, 15 y 16 o fragmentos de las mismas, en donde las secuencias de nucleótidos que comprenden las SEQ ID NOs: 10 y/o 16 se orientan en la dirección opuesta con respecto a la secuencia de polinucleótidos que codifica un polipéptido, y secuencias de nucleótidos que comprenden las SEQ ID NOs: 12 y/o 15 se orientan en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido. Es más preferida igualmente una secuencia de nucleótidos seleccionada

del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, más preferiblemente, una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, en donde las secuencias de nucleótidos que comprenden las SEQ ID NOs: 23 y/o 16 se orientan en la dirección opuesta con respecto a la secuencia de polinucleótidos que codifica un polipéptido, y secuencias de
5 nucleótidos que comprenden las SEQ ID NOs: 25 y/o 28 se orientan en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido.

En una realización adicional, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota comprende una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del
10 grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 o fragmentos de la misma, preferiblemente una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 y 16 o fragmentos de la misma, o una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs:
15 20, 22, 23, 24, 25, 26, 27, 28 y 16 o fragmentos de las mismas. Es más preferida una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 10, 12, 15 y 16 o fragmentos de las mismas. Es más preferida igualmente una secuencia de nucleótidos complementaria a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas.

En una realización adicional, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota comprende una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 o fragmentos de las mismas, preferiblemente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 y 16 o
25 fragmentos de las mismas, o una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 20, 22, 23, 24, 25, 26, 27, 28 y 16 o fragmentos de las mismas. Es más preferida una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 10, 12, 15 y 16 o fragmentos de las mismas, más preferiblemente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 10, 12, 15 y 16 o fragmentos de las mismas, en donde las secuencias de nucleótidos que comprenden las SEQ ID NOs: 10 y/o 16 se orientan en la dirección opuesta con respecto a la secuencia de polinucleótidos que codifica un polipéptido, y secuencias de nucleótidos que comprenden las SEQ ID NOs: 12 y/o 15 se orientan en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido. Es más preferida igualmente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, más preferiblemente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, en donde las secuencias de nucleótidos que comprenden las SEQ ID NOs: 23 y/o 16 se orientan en la dirección opuesta con respecto a la secuencia de polinucleótidos que codifica un polipéptido, y secuencias de nucleótidos que comprenden las SEQ ID
30 NOs: 12 y/o 15 se orientan en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido. Es más preferida igualmente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, más preferiblemente una secuencia de nucleótidos al menos un 80 % idéntica a la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 23, 25, 28 y 16 o fragmentos de las mismas, en donde las secuencias de nucleótidos que comprenden las SEQ ID NOs: 23 y/o 16 se orientan en la dirección opuesta con respecto a la secuencia de polinucleótidos que codifica un polipéptido, y secuencias de nucleótidos que comprenden las SEQ ID
35 NOs: 25 y/o 28 se orientan en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido.

En algunas realizaciones, la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 o fragmentos de las mismas, comprende cinco o menos, preferiblemente cuatro o menos, más preferiblemente tres o menos, mucho más preferiblemente dos o menos, en particular una modificación de ácido nucleico, en donde la modificación o modificaciones de ácidos nucleicos son preferiblemente una sustitución de ácido nucleico.
45

En algunas realizaciones, la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 11, 14, 20, 22, 24 y 27 o fragmentos de las mismas, comprende cinco o menos, preferiblemente cuatro o menos, más preferiblemente tres o menos, mucho más preferiblemente dos o menos, en particular una modificación de ácido nucleico, en donde la modificación o modificaciones de ácidos nucleicos son preferiblemente una sustitución de ácido nucleico.
50

En algunas realizaciones, la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 7, 9, 11, 14, o fragmentos de las mismas, comprende una sustitución de ácido nucleico en la posición 16 con respecto al inicio de la secuencia de nucleótidos de SEQ ID NOs: 7, 9, 11, 14. Preferiblemente, G en la posición 16 con respecto al inicio de la secuencia de nucleótidos se reemplaza con T.
55

En algunas realizaciones, la secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NOs: 20, 22,

24 y 27 o fragmentos de las mismas, comprende una sustitución de ácido nucleico en la posición 13 con respecto al inicio de la secuencia de nucleótidos de SEQ ID NOs: 20, 22, 24 y 27. Preferiblemente, G en la posición 13 con respecto al inicio de la secuencia de nucleótidos se reemplaza con T.

- 5 En una realización adicional, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se orienta en la misma dirección que la secuencia de polinucleótidos que codifica un polipéptido.

En una realización adicional, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota está orientada en dirección opuesta en relación con la secuencia de polinucleótidos que codifica un polipéptido.

En una realización preferida, el casete de expresión comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota y una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota como se ha descrito anteriormente. Preferiblemente, el origen de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota y la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota es el mismo, es decir, es de la misma especie. Más preferiblemente, el origen de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota, la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota y la célula huésped es el mismo, es decir, es de la misma especie, por ejemplo, el origen de la secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota, la secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota y la célula huésped es del mismo mamífero, por ejemplo, de un ser humano.

En algunas realizaciones, si la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota es una secuencia de ADN genómico no traducida de una especie, el promotor del casete de expresión no es un promotor GAPDH de la misma especie.

En algunas realizaciones, si la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota es una secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de origen humano, el promotor del casete de expresión no es un promotor GAPDH humano.

En algunas realizaciones, el promotor del casete de expresión no es un promotor GAPDH.

En una realización, si el casete de expresión comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y una secuencia de ADN genómico no traducida aguas abajo de un promotor de Gliceraldehído 3-fosfato deshidrogenasa (GAPDH) eucariota, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota se inicia en una región una secuencia de polinucleótidos que +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos y en el que el casete de expresión comprende además una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota es de aproximadamente 200 a aproximadamente 2800 nucleótidos, el promotor del casete de expresión puede ser un promotor GAPDH eucariota, preferiblemente un promotor GAPDH de mamífero, más preferiblemente un promotor GAPDH de roedor o humano. En esta realización, la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH eucariota que se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH eucariota a la posición de nucleótido de aproximadamente -3500 se encuentra preferiblemente directamente aguas abajo del promotor GAPDH eucariota, más preferiblemente en esta realización, el casete de expresión comprende la secuencia de ADN genómico de origen natural que comprende el promotor GAPDH eucariota y que se extiende a la posición de nucleótido de aproximadamente -3500, en donde la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH.

En algunas realizaciones, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota es de origen mamífero, por ejemplo, el promotor GAPDH eucariota es un promotor GAPDH de mamífero y secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH de mamífero se usa como se describe en el presente documento.

En algunas realizaciones, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota es de origen de roedor o humano, por ejemplo, el promotor GAPDH eucariota es un promotor GAPDH de roedor o humano y una secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH de roedor o humano se usa como se describe en el presente documento.

Preferiblemente, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota se selecciona de origen humano, de rata o ratón, más preferiblemente de origen humano o de ratón, mucho más preferiblemente de origen humano.

En algunas realizaciones, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH eucariota no está unida operativamente a la secuencia de polinucleótidos que codifica el polipéptido.

En algunas realizaciones, el casete de expresión comprende un sitio de poliadenilación. Preferiblemente, el sitio de poliadenilación se selecciona del grupo que consiste en SV40 poli(A) y BGH (hormona de crecimiento bovina) poli(A).

En algunas realizaciones, el promotor y la secuencia de polinucleótidos que codifica un polipéptido del casete de expresión están unidos operativamente.

En algunas realizaciones, el promotor del casete de expresión se selecciona del grupo que consiste en promotor SV40, promotor tk humano, promotor MPSV, CMV de ratón, CMV humano, CMV de rata, EF1 alfa humano, EF1 alfa de hámster chino, GAPDH humana, promotores híbridos que incluyen el promotor MYC, HYK y CX.

En algunas realizaciones, el polipéptido codificado por el casete de expresión puede ser un polipéptido no glucosilado y glucosilado. Los polipéptidos glucosilados se refieren a polipéptidos que tienen al menos una cadena de oligosacáridos.

Los ejemplos de proteínas no glucosiladas son, por ejemplo, hormonas no glucosiladas; enzimas no glucosiladas; factores de crecimiento no glucosilados de la familia del factor de crecimiento nervioso (NGF), del factor de crecimiento epitelial (EGF) y de la familia del factor de crecimiento fibroblástico (FGF) y receptores no glucosilados para hormonas y factores de crecimiento.

Los ejemplos de proteínas glucosiladas son hormonas y factores liberadoras de hormonas, factores de coagulación, factores anticoagulantes, receptores de hormonas o factores de crecimiento, factores neurotróficos, citocinas y sus receptores, receptores de linfocitos T, proteínas de membrana superficial, proteínas de transporte, receptores de localización, adreínas, proteínas reguladoras, anticuerpos, proteínas quiméricas, tales como inmunoadhesinas, y fragmentos de cualquiera de las proteínas glucosiladas. Preferiblemente, el polipéptido se selecciona del grupo que consiste en anticuerpos, fragmentos de anticuerpo o derivados de anticuerpo (por ejemplo, proteínas de fusión Fc y formatos de anticuerpo particulares como anticuerpo biespecíficos). El fragmento de anticuerpo como se usa en el presente documento incluyen, pero sin limitación, (i) un dominio, (ii) el fragmento Fab que consiste en los dominios VL, VH, CL o CK y CH1, incluyendo Fab' y Fab'-SH, (iii) el fragmento Fd que consiste en los dominios VH y CH1, (iv) el fragmento dAb (Ward ES et al., (1989) Nature, 341(6242): 544-6) que consiste en un único dominio variable, (v) fragmentos F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab unidos, (vi) moléculas Fv monocatenarias (scFv), en las que un dominio VH y un dominio VL están unidos por un enlazador peptídico que permite que los dos dominios se asocien para formar un sitio de unión a antígeno (Bird RE et al., (1988) Science 242(4877): 423-6; Huston JS et al., (1988) Proc Natl Acad Sci U S A, 85(16): 5879-83), (vii) "diacuerpos" o "triacuerpos", fragmentos multivalentes o multiespecíficos construidos por fusión génica (Holliger P et al., (1993) Proc Natl Acad Sci U S A, 90(14): 6444-8; Holliger P et al., (2000) Methods Enzymol, 326: 461-79), (viii) scFv, diacuerpo o anticuerpo de dominio fusionado a una región Fc e (ix) scFv fusionado al mismo o un anticuerpo diferente.

En algunas realizaciones, el casete de expresión comprende además un elemento genético seleccionado del grupo que consiste en un promotor adicional, un potenciador, elementos de control de la transcripción y un marcador seleccionable, preferiblemente un marcador seleccionable que se expresa en células animales. Los elementos de control transcripcional son, por ejemplo, secuencias de Kozak o elementos de terminación de la transcripción.

En una realización, el elemento genético es un marcador seleccionable en el que el contenido de sitios CpG contenidos en la secuencia de polinucleótidos que codifica el marcador seleccionable es 45 o menos,

preferiblemente 40 o menos, más preferiblemente 20 o menos, en particular 10 o menos, más en particular 5 o menos, más en particular 0 (los sitios CpG han sido completamente eliminados).

En un aspecto adicional, la presente descripción proporciona un vector de expresión, preferiblemente un vector de expresión de mamífero que comprende un casete de expresión como se ha descrito anteriormente.

En algunas realizaciones, el vector de expresión comprende al menos dos unidades de transcripción separadas. Un vector de expresión con dos unidades de transcripción separadas también se denomina vector de doble gen. Un ejemplo del mismo es un vector, en el que la primera unidad de transcripción codifica la cadena pesada de un anticuerpo o un fragmento del mismo y la segunda unidad de transcripción codifica la cadena ligera de un anticuerpo. Otro ejemplo es un vector de doble gen, en el que las dos unidades de transcripción codifican dos subunidades diferentes de una proteína tal como una enzima. Sin embargo, también es posible que el vector de expresión de la presente invención comprenda más de dos unidades de transcripción separadas, por ejemplo tres, cuatro o incluso más unidades de transcripción separadas cada una de las cuales comprende una secuencia de nucleótidos diferente que codifica una cadena polipeptídica diferente. Por lo tanto, un ejemplo es un vector con cuatro unidades de transcripción separadas, cada una de las cuales contiene una secuencia de nucleótidos diferente que codifica una subunidad de una enzima que consiste en cuatro subunidades diferentes.

En algunas realizaciones, el vector de expresión comprende además un elemento genético seleccionado del grupo que consiste en un promotor adicional, un potenciador, elementos de control de la transcripción, un origen de replicación y un marcador seleccionable.

En algunas realizaciones, el vector de expresión comprende además un origen de replicación y un marcador seleccionable en el que el contenido de los sitios CpG contenidos en la secuencia de polinucleótidos del vector de expresión que codifica el origen de replicación y el marcador seleccionable es 200 o menos, preferiblemente 150 o menos, en particular 100 o menos, más particularmente 50 o menos, mucho más particular 30 o menos.

Se puede usar cualquier marcador de selección comúnmente empleado tal como timidina cinasa (tk), dihidrofolato reductasa (DHFR), puromicina, neomicina o glutamina sintetasa (GS) para el casete de expresión o el vector de expresión de la presente invención. Preferiblemente, los vectores de expresión de la invención también comprenden un número limitado de sitios de restricción útiles para la inserción del casete de expresión para la secreción de una proteína heteróloga de la presente invención. Cuando se usan en particular solo para expresión transitoria/episómica, los vectores de expresión de la invención pueden comprender además un origen de replicación tal como el origen oriP del virus Epstein Barr (EBV) o virus SV40 para la replicación autónoma/mantenimiento episómico en células huésped eucariotas, pero pueden estar desprovistos de un marcador seleccionable. La expresión transitoria en células que carecen de factores relevantes para facilitar la replicación del vector también es posible. El vector de expresión que alberga el casete de expresión puede comprender además un casete de expresión que codifica un marcador fluorescente, un casete de expresión que codifica un ARNnc, un casete de expresión que codifica una proteína antiapoptótica o un casete de expresión que codifica una proteína que aumenta la capacidad de la ruta secretora.

En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

a) una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH de mamífero, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero se inicia en una región que abarca del extremo 5' del promotor GAPDH de mamífero a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, en el que la longitud de la secuencia de ADN genómico aguas arriba del promotor GAPDH de mamífero es de 200 a aproximadamente 2800 nucleótidos

b) un promotor

c) una secuencia de polinucleótidos que codifica un polipéptido

d) un sitio de poliadenilación

e) un potenciador

f) una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH de mamífero,

en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, en el que la secuencia de ADN genómico aguas abajo del promotor GAPDH de mamífero se inicia en una región que abarca desde la posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero es de aproximadamente 200 a

aproximadamente 2800 nucleótidos.

En algunas realizaciones, la presente descripción proporciona un vector de expresión, que comprende en orden:

- 5 a) una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota
b) un promotor
c) una secuencia de polinucleótidos que codifica un polipéptido
d) un sitio de poliadenilación
e) un potenciador
10 f) una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota,

en el que la inclusión del potenciador es opcional.

En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

- 15 a) una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota
b) un potenciador
c) un promotor
d) una secuencia de polinucleótidos que codifica un polipéptido
20 e) un sitio de poliadenilación
f) una secuencia de ADN genómico no traducida aguas abajo de una GAPDH eucariota,

en el que la inclusión del potenciador es opcional.

25 En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

- a) un potenciador
b) una secuencia de ADN genómico no traducida aguas arriba de una GAPDH eucariota
c) un promotor
30 d) una secuencia de polinucleótidos que codifica un polipéptido
e) un sitio de poliadenilación
f) secuencia de ADN genómico no traducida aguas abajo de una GAPDH eucariota,

en el que la inclusión del potenciador es opcional.

35 En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

- a) una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota
b) un promotor
40 c) una secuencia de polinucleótidos que codifica un polipéptido
d) un sitio de poliadenilación
e) un potenciador
f) una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH eucariota,

45 en el que la inclusión del potenciador es opcional.

En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

- a) una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota
50 b) un potenciador
c) un promotor
d) una secuencia de polinucleótidos que codifica un polipéptido
e) un sitio de poliadenilación
f) una secuencia de ADN genómico no traducida aguas arriba de una GAPDH eucariota,

55 en el que la inclusión del potenciador es opcional.

En un aspecto adicional, la presente descripción proporciona un vector de expresión, que comprende en orden:

- a) un potenciador
- b) una secuencia de ADN genómico no traducida aguas abajo de una GAPDH eucariota
- c) un promotor
- d) una secuencia de polinucleótidos que codifica un polipéptido
- e) un sitio de poliadenilación
- f) secuencia de ADN genómico no traducida aguas arriba de una GAPDH eucariota,

5

en el que la inclusión del potenciador es opcional.

- 10 La secuencia de ADN genómico no traducida aguas arriba de una GAPDH eucariota, un potenciador, un promotor, una secuencia de polinucleótido que codifica un polipéptido, un sitio de poliadenilación y una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH eucariota de los vectores de expresión son, por ejemplo, como se ha descrito anteriormente.
- 15 En un aspecto adicional, la presente descripción proporciona una célula huésped que comprende un casete de expresión o un vector de expresión como se ha descrito anteriormente. La célula huésped puede ser una célula humana o no humana. Las células huésped preferidas son células de mamífero. Los ejemplos preferidos de células huésped de mamífero incluyen, sin restricción, células de riñón embrionario humano (Graham FL et al., J. Gen. Virol. 36: 59-74), fibroblastos humanos MRC5, células de melanoma humano 983M, células de riñón canino MDCK, fibroblastos de pulmón de rata cultivados RF aislados de ratas Sprague-Dawley, células de melanoma murino B16BL6, células de mastocitoma murino P815, células de adenocarcinoma mamario murino MT1 A2, células PER:C6 (Leiden, Países Bajos) y células o líneas celulares de ovario de hámster chino (CHO) (Puck et al., 1958, J. Exp. Med. 108: 945-955).
- 20
- 25 En una realización preferida particular, la célula huésped es una célula o línea celular de ovario de hámster chino (CHO). Las líneas celulares CHO adecuadas incluyen, por ejemplo, CHO-S (Invitrogen, Carlsbad, CA, Estados Unidos), CHO K1 (ATCC CCL-61), CHO pro3-, CHO DG44, CHO P12 o la línea celular dhfr- CHO DUK-BII (Chasin et al., PNAS 77, 1980, 4216-4220), DUXBI 1 (Simonsen et al., PNAS 80, 1983, 2495-2499), o CHO-K1SV (Lonza, Basel, Suiza).
- 30
- En un aspecto adicional, la presente descripción proporciona un método *in vitro* para la expresión de un polipéptido, que comprende transfectar una célula huésped con el casete de expresión o un vector de expresión como se describe anteriormente y recuperar el polipéptido. El polipéptido es preferiblemente un polipéptido heterólogo, más preferiblemente un polipéptido humano.
- 35
- Para transfectar el casete de expresión o el vector de expresión en una célula huésped de acuerdo con la presente invención, cualquier técnica de transfección tales como las ya conocidas en la técnica, por ejemplo, electroporación, coprecipitación con fosfato de calcio, transfección de DEAE-dextrano, lipofección, pueden emplearse si es apropiado para un tipo de célula huésped dado. Debe observarse que la célula huésped transfectada con el casete de expresión o el vector de expresión de la presente invención debe interpretarse como una línea celular transfectada de forma transitoria o estable. Por lo tanto, de acuerdo con la presente invención, el presente casete de expresión o el vector de expresión se puede mantener episómicamente, es decir transfectarse transitoriamente, o puede integrarse de forma estable en el genoma de la célula huésped, es decir, transfectarse de forma estable.
- 40
- 45 Una transfección transitoria se caracteriza por la no aplicación de ninguna presión de selección para un marcador de selección portado por vector. En experimentos de expresión transitoria que comúnmente duran de 2 hasta 10 días después de la transfección, el casete de expresión transfectado o el vector de expresión se mantienen como elementos episómicos y aún no están integrados en el genoma. Es decir, el ADN transfectado normalmente no se integra en el genoma de la célula huésped. Las células huésped tienden a perder el ADN transfectado y crecer en exceso las células transfectadas en la población tras el cultivo del grupo de células transfectadas transitoriamente.
- 50
- Por lo tanto, la expresión es más fuerte en el periodo inmediatamente posterior a la transfección y disminuye con el tiempo. Preferiblemente, un transfectante transitorio de acuerdo con la presente invención se entiende como una célula que se mantiene en cultivo celular en ausencia de presión de selección hasta un tiempo de 2 a 10 días después de la transfección.
- 55
- En una realización preferida de la invención, la célula huésped, por ejemplo, la célula huésped CHO se transfecta de forma estable con el casete de expresión o el vector de expresión de la presente invención. La transfección estable significa que el ADN foráneo recientemente introducido, tal como el ADN del vector, se está incorporando al ADN genómico, habitualmente mediante eventos de recombinación aleatorios no homólogos. El número de copias del

ADN del vector y, concomitantemente, la cantidad del producto génico se puede aumentar seleccionando líneas celulares en las que las secuencias del vector se han amplificado después de la integración en el ADN de la célula huésped. Por lo tanto, es posible que dicha integración estable dé lugar, tras la exposición a aumentos adicionales en la presión de selección para la amplificación génica, a cromosomas de doble minuto en células CHO. Además, una transfección estable puede dar como resultado la pérdida de partes de la secuencia del vector no relacionadas directamente con la expresión del producto génico recombinante, tal como, por ejemplo, las regiones de control del número de copias bacterianas se vuelven superfluas tras la integración genómica. Por lo tanto, una célula huésped transfectada ha integrado al menos parte o diferentes partes del casete de expresión o el vector de expresión en el genoma.

10

En un aspecto adicional, la presente descripción proporciona el uso del casete de expresión o un vector de expresión como se ha descrito anteriormente para la expresión de un polipéptido heterólogo de una célula huésped de mamífero, en particular, el uso del casete de expresión o un vector de expresión como se ha descrito anteriormente para la expresión *in vitro* de un polipéptido heterólogo de una célula huésped de mamífero.

15

La expresión y la recuperación de la proteína se pueden realizar de acuerdo con métodos conocidos por el experto en la técnica.

Para la expresión de un polipéptido, la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota del casete de expresión o del vector de expresión como se ha descrito anteriormente y la célula huésped como se ha descrito anteriormente se usan y normalmente son del mismo origen. Sorprendentemente se ha encontrado que se obtiene un aumento de la expresión si la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba de un promotor GAPDH eucariota del casete de expresión o del vector de expresión y la célula huésped son de origen diferente, por ejemplo, si las secuencias de ADN humano aguas abajo y/o aguas arriba de un promotor GAPDH eucariota se usan en células CHO.

En un aspecto adicional, la presente descripción proporciona el uso del casete de expresión o el vector de expresión como se ha descrito anteriormente para la preparación de un medicamento para el tratamiento de un trastorno.

En un aspecto adicional, la presente descripción proporciona el casete de expresión o el vector de expresión como se ha descrito anterior para su uso como un medicamento para el tratamiento de un trastorno.

En un aspecto adicional, la presente descripción proporciona el casete de expresión o el vector de expresión como se ha descrito anteriormente para su uso en la terapia génica.

35

Ejemplos

Ejemplo 1: Clonación de vectores de expresión:

40 I. Materiales y métodos

I.1 Construcciones plasmídicas

I.1.1. Placas de cultivo LB

45

Se mezclaron 500 ml de agua y se hirvieron con 16 g de agar LB (Invitrogen, Carlsbad, CA, Estados Unidos) (1 litro de LB contiene 10 g de triptona, 5 g de extracto de levadura y 10 g de NaCl). Después de enfriar, se añadió el antibiótico respectivo a la solución que después se colocó en placas (placas de ampicilina a 100 µg/ml y placas de kanamicina a 50 µg/ml).

50

I.1.2. Reacción en cadena de la polimerasa (PCR)

Se realizaron todas las PCR utilizando 1 µl de dNTP (10 mM para cada dNTP; Invitrogen, Carlsbad, CA, Estados Unidos), 2 unidades de ADN polimerasa Phusion® (Finnzymes Oy, Espoo, Finlandia), 25 nmol de Cebador A (Mycrosynth, Balgach, Suiza), 25 nmol de Cebador B (Mycrosynth, Balgach, Suiza), 10 µl de tampón HF 5X (MgCl₂ 7,5 mM, Finnzymes, Espoo, Finlandia), 1,5 µl de dimetilsulfóxido (DMSO, Finnzymes, Espoo, Finlandia) y 1-3 µl de la plantilla (1-2 µg) en un volumen final de 50 µl. Todos los cebadores utilizados se enumeran en la Tabla 1.

La PCR se inició por una desnaturalización inicial a 98 °C durante 3 minutos, seguida de 35 ciclos de 30 s de

desnaturalización a 98 °C, 30 s de hibridación a una temperatura específica del cebador (según el contenido de CG) y 2 min de elongación a 72 °C. Se realizó una elongación final a 72 °C durante 10 minutos antes de enfriar y mantener a 4 °C.

5 **Tabla 1: Resumen de cebadores usados en PCR.** GAPDH: Secuencia de gliceraldehído 3-fosfato deshidrogenasa, 5': secuencia aguas arriba, 3: secuencia aguas abajo. La "T" (subrayada) en el cebador GlnPr1 172 se introdujo para evitar la formación de dímeros de cebadores.

| Cebador | Secuencia de cebador | Secuencia amplificada | Seq ID |
|------------|---|-----------------------|--------------|
| GlnPr 1171 | ATTATTCGCGATGGCTCCTGGCA TCTCTGGGACCGAGGC | 5' GAPDH | SEQ ID No: 1 |
| GlnPr 1172 | ATCGTCGCGAAGCTTGAGATTG <u>T</u> CCAAGCAGGTAGCCAG | | SEQ ID No: 2 |
| GlnPr 1173 | AGCAAGTACTTCTGAGCCTTCA GTAATGGCTGCCTG | 3'GAPDH | SEQ ID No: 3 |
| GlnPr 1174 | TGGCAGTACTAAGCTGGCACCA CTACTTCAGAGAACAAG | | SEQ ID No: 4 |

I.1.3. Digestión de restricción

10

Para todas las digestiones de restricción se mezcló aproximadamente 1 µg de ADN plasmídico (cuantificado con NanoDrop, espectrofotómetro ND-1000 (Thermo Scientific, Wilmington, DE, Estados Unidos)) con 10-20 unidades de cada enzima, 4 µl de 10X NEBuffer correspondiente (NEB, Ipswich, MA, Estados Unidos), y el volumen se completó a 40 µl con H₂O estéril. Sin más indicaciones, las digestiones se incubaron durante 1 hora a 37 °C.

15

Después de cada digestión preparativa de la estructura principal, se añadió 1 unidad de fosfatasa alcalina intestinal de ternera (CIP; NEB, Ipswich, MA, Estados Unidos) y la mezcla se incubó durante 30 min a 37 °C.

Si la digestión se realizó en NEBuffer 3 (NEB, Ipswich, MA, Estados Unidos), el tampón se cambió a tampón NEB 4 antes de añadir CIP porque esta enzima tiene una actividad fuerte en este tampón y también puede digerir algunos de los nucleótidos en los extremos externos.

20

I.1.4. Purificación por PCR y electroforesis en gel de agarosa

I.1.4.1. Limpieza por PCR

25

Para permitir la digestión, todos los fragmentos de PCR se limpiaron antes de las digestiones de restricción usando el kit Macherey Nagel Extract II (Macherey Nagel, Oensingen, Suiza) siguiendo el manual del fabricante usando 40 µl de tampón de elución. Este protocolo también se usó para cambiar los tampones de muestras de ADN.

I.1.4.2. Extracción de ADN

30

Para la electroforesis en gel, se prepararon geles al 1 % usando Agarosa UltraPure™ (Invitrogen, Carlsbad, CA, Estados Unidos) y tampón 50X de ácido tris acético EDTA (TAE, pH 8,3; Bio RAD, Munich, Alemania). Para la tinción de ADN, se añadió 1 µl de colorante rojo en gel (Biotum, Hayward, CA, Estados Unidos) a 100 ml de gel de agarosa. Como marcador de tamaño, se usaron 2 µg de la escalera de ADN de 1 kb (NEB, Ipswich, MA, Estados Unidos). La electroforesis se realizó durante aproximadamente 1 hora a 125 voltios. Las bandas de interés se cortaron del gel de agarosa y se purificaron usando el kit Extract II (Macherey-Nagel, Oensingen, Suiza), siguiendo el manual del fabricante usando 40 µl de tampón de elución.

35

I.1.5. Ligación

Para cada ligación, se mezclaron 4 µl de inserto con 1 µl de vector, 400 unidades de ligasa (ADN ligasa T4, NEB, Ipswich, MA, Estados Unidos), 1 µl de tampón ligasa 10X (tampón ADN ligasa T4; NEB, Ipswich, MA, Estados Unidos) en un volumen de 10 µl. La mezcla se incubó durante 1-2 h a TA.

I.1.6. Transformación de productos de ligación en bacterias competentes

10 Para la clonación de pGLEX41-[REP] y para construcciones hechas con el vector pCR-Blunt que contienen un origen de replicación estándar, se usó TOP 10 (*E. coli* competente TOP 10 de One Shot®; Invitrogen, Carlsbad, CA, Estados Unidos).

15 Para el inicio de la replicación del plásmido que contiene el origen de replicación R6K, se requiere la expresión de la proteína π, codificada por la secuencia pir. La proteína π se expresa por *E. coli* competente de One Shot® PIR1 (Invitrogen, Carlsbad, CA, Estados Unidos). Estas bacterias se usaron para todos los vectores que contienen la secuencia R6K.

20 Para transformar bacterias competentes con el producto de ligación, se descongelaron 25-50 µl de bacterias en hielo durante 5 minutos. Después, se añadieron 3-5 µl de producto de ligación a las bacterias competentes y se incubaron durante 20-30 minutos en hielo antes del choque térmico durante 1 minuto a 42 °C. Después, se añadieron 500 µl de medio S.O.C. (Invitrogen, Carlsbad, CA, Estados Unidos) por tubo y se incubaron durante 1 hora a 37 °C en agitación. Finalmente, las bacterias se colocan en una placa LB con ampicilina (Sigma-Aldrich, St. Louis, MO, Estados Unidos) y se incuban durante una noche a 37 °C. Para la clonación en vectores pCR-Blunt, se usaron 25 placas con kanamicina (Sigma-Aldrich, St. Louis, MO, Estados Unidos).

I.1.7. Preparación de plásmidos a pequeña (mini) y media (midi)**I.1.7.1. Minipreparación**

30 Para la minipreparación, se cultivaron colonias de bacterias transformadas durante 6-16 horas en 2,5 ml de LB y ampicilina o kanamicina a 37 °C, 200 rpm. El ADN se extrajo con un kit de purificación de plásmidos para *E. coli* (QuickPure, Macherey Nagel, Oensingen, Suiza), siguiendo el manual proporcionado.

35 El ADN plasmídico de minipreparaciones se cuantificó una vez con el espectrofotómetro NanoDrop ND-1000 (Thermo Scientific, Wilmington, DE, Estados Unidos) midiendo la absorbancia a 260 nm y evaluando la relación de OD260 nm/OD280 nm que tenía que estar entre 1,8 y 2. Se realizó una digestión de control antes de enviar la muestra a Fasteris SA (Ginebra, Suiza) para la confirmación de la secuencia.

40 Para la extracción de BAC, se utilizó el kit QuickPure (Macherey Nagel, Oensingen, Suiza) con la siguiente modificación del protocolo: Se sembraron 10 ml de LB y cloranfenicol (12,5 µg/ml) (Sigma-Aldrich, St. Louis, MO, Estados Unidos) con bacterias que contenían el vector pBACe3.6. Después de la incubación en una plataforma de agitación a 37 °C durante una noche, el cultivo se centrifugó durante 5 min a 13300 rpm antes de suspenderse de nuevo en 500 µl de tampón A1. Se añadieron 500 µl de Tampón de Lisis A2 y la solución se incubó durante 5 45 minutos a TA. Después, se neutralizó con 600 µl de tampón A3 y se centrifugó durante 10 minutos a 13300 rpm. El sobrenadante se cargó en una columna y a partir de esta etapa se usó el protocolo estándar del kit QuickPure miniprep.

I.1.7.2. Midipreparación

50 Para la midipreparación, las bacterias transformadas se cultivaron a 37 °C durante una noche en 200 a 400 ml de LB y ampicilina (o kanamicina). A continuación, el cultivo se centrifugó durante 20 minutos a 725 g y el plásmido se purificó usando un kit comercial (NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Suiza) siguiendo el protocolo de plásmido bajo proporcionado en el manual del fabricante.

55 El ADN plasmídico de la midipreparación se cuantificó tres veces con el espectrofotómetro NanoDrop ND-1000, se confirmó mediante digestión de restricción y finalmente se envió para su secuenciación (Fasteris SA, Ginebra, Suiza).

II. Resultados y análisis

II.1. Clonación de regiones de ADN aguas arriba y aguas abajo del casete de expresión de GAPDH (GAPDH 5' y 3')

5 El clon BAC RPCIB753F11841Q se pidió en Imagen (Berlín, Alemania). Este clon contiene la secuencia de GAPDH humana en una estructura principal del vector pBACe3.6, que contiene un gen de resistencia al cloranfenicol. Después de la extracción del ADN por minipreparación, la concentración del vector se determinó mediante Nanodrop a 27 ng/μl.

10 Las secuencias de ADN que rodean inmediatamente el casete de expresión de GAPDH aguas arriba del promotor y aguas abajo del sitio de poliadenilación se amplificaron por PCR usando 27 ng del clon RPCIB753F11841Q purificado como plantilla. El fragmento de 3 kb aguas arriba del promotor se amplificó con los cebadores GlnPrI 171 (SEQ ID NO: 1) y GlnPrI 172 (SEQ ID NO: 2) lo que condujo al amplicón con SEQ ID No. 5. Como cebador GlnPrI 172 (SEQ ID NO: 2) lleva un cambio de base (G a T) con respecto a la secuencia de plantilla, todas las secuencias derivadas de esta reacción de PCR también llevarán este cambio de base. El cambio se encuentra en la posición -3721 con respecto al inicio de la transcripción del gen GAPDH (pb 812 de SEQ ID NO: 17, posición 23 con respecto al inicio de la SEQ ID NO: 5). El fragmento de 3 kb aguas abajo del sitio de poliadenilación se amplificó con los cebadores GlnPrI 173 (SEQ ID NO: 3) y GlnPrI 174 (SEQ ID NO: 4) lo que condujo al amplicón con SEQ ID NO: 6 (Tabla 1). La temperatura de hibridación utilizada para estas PCR fue de 72 °C.

Los fragmentos de GAPDH 5' y 3' (SEQ ID NOS: 5 y 6) se clonaron en pCR-Blunt, un vector de clonación de producto de PCR disponible comercialmente (pCR-Blunt, kit de clonación PCR Zero Blunt, Invitrogen). Los productos de ligación se transformaron en bacterias competentes TOP10 y se pusieron en placas de kanamicina LB-agar. Las colonias se amplificaron y los plásmidos se aislaron mediante minipreparaciones. Se realizaron digestiones de control para identificar clones positivos que produjeron las construcciones pCR-Blunt-5'GAPDH y pCR-Blunt-3'GAPDH.

II.2. Preparación del fragmento de ADN que codifica las proteínas indicadoras GFP y un anticuerpo monoclonal recombinante IgG1 (LC-IRES-HC-IRES-GFP)

La construcción indicadora (REP) usada en el presente trabajo consistió en un gen policistrónico: cadena ligera (LC) del anticuerpo monoclonal IgG1-IRES-cadena pesada (HC) del anticuerpo monoclonal IgG1-IRES-proteína fluorescente verde (GFP). La presencia de sitios de entrada de ribosomas internos (IRES) derivados del virus de la encefalomiocarditis (Gurtu et al., Biochem Biophys Res Commun.; 229(1): 295-298, 1996) permite la traducción de la cadena ligera (LC) del anticuerpo monoclonal IgG1 de 3 péptidos, cadena pesada (HC) del anticuerpo monoclonal IgG1 y GFP (Figura 1). Las células transfectadas, por lo tanto, secretarán el anticuerpo monoclonal IgG1 y acumularán GFP intracelular de una manera dependiente. Sin embargo, los ARNm policistrónicos no son comunes en células eucariotas y su traducción no es muy eficiente, lo que conduce a títulos relativamente bajos de expresión de IgG1 y GFP.

Se digirió un vector que contenía la construcción de REP usando las enzimas de restricción NheI y BstBI (BstBI se usa a 65 °C). El fragmento REP que contenía la construcción de expresión se cortó, se purificó y se usó para etapas de clonación adicionales.

II.3. Clonación de vectores de expresión

El vector pGLEX41, un vector de expresión derivado de pcDNA3.1 (+) (Invitrogen, Carlsbad, CA) se usó para la producción de líneas celulares estables. Se usó como estructura principal inicial que se había modificado para generar las segundas generaciones de vectores A y B con y sin las secuencias de GAPDH. Para todos los vectores se usó la misma combinación promotor-intrón (mCMV y un fragmento donante-aceptor que codifica el primer intrón (IgDA)) (Gorman et al., (1990) Proc Natl Acad Sci USA, 87: 5459-5463).

Clonación del vector intermedio pGLEX41-HM-MCS-amp^r:

55 El desarrollo de la nueva generación de vectores se inició a partir de pGLEX41. Este vector se cortó usando las enzimas de restricción NruI y BspHI para liberar el casete de resistencia a la ampicilina. El fragmento de la estructura principal se sometió a CIP y se purificó mediante electroforesis en gel. El fragmento de ADN que codifica una versión optimizada de codón (para expresión en *E. coli*) del gen de resistencia a ampicilina (que incluye el promotor bla) se

ha pedido en GeneArt. El inserto se cortó del vector de clonación GeneArt n.º 1013237 usando las enzimas de restricción NruI y BspHI (las mismas enzimas que se usan para la estructura principal), se purificó y se clonó en la estructura principal. Las minipreparaciones se analizaron mediante digestión de restricción. El clon pGLEX41-HM-MCS-ampA#2 tenía el perfil de restricción esperado y la integración del fragmento correcto se confirmó mediante 5 secuenciación.

Clonación del vector intermedio pGLEX41-MCS-R6K-ampA

Para intercambiar el origen de replicación de pUC del vector pGLEX41-HM-MCS-ampA#2, el vector se digirió 10 usando PvuI y BspHI. El fragmento de la estructura principal se sometió a CIP y se purificó. El nuevo fragmento de inserción contiene el origen de replicación de R6K y una secuencia de poli(A) de SV40 modificada como parte del casete de expresión. Se han eliminado las secuencias innecesarias de la estructura principal bacteriana o viral alrededor del SV40 poli (A) (véase la Tabla 2 a continuación). El fragmento de inserción se ha pedido a GeneArt; se 15 cortó del vector de clonación GeneArt n.º 1013238 usando las enzimas PvuI y BspHI (las mismas que se usan para la estructura principal), se purificó y se clonó en el fragmento de la estructura principal. Las minipreparaciones se prepararon y se confirmaron mediante análisis de secuencia. El clon pGLEX41-MCS-R6K-ampA#1 tenía la secuencia correcta.

Tabla 2: Contenido de CpG en los diferentes vectores

| | Contenido de CpG en vectores de expresión | | |
|--------------------------|---|--|-----------------------------------|
| | pGLEX41 | Vectores optimizados por codón: "A" | Vectores reducidos por CpG "B" |
| Resistencia a ampicilina | 49 | 43 | 19 |
| Resistencia a puromicina | 93 | 36 | 1 |
| Resistencia a geneticina | 74 | 51 | 0 |
| Origen de replicación | 45 | 9 | 9 |
| Suma: | 261 | 139 | 29 |

20

Clonación del vector intermedio pGLEX41-MCS-R6K-ampB

El vector pGLEX41- MCS-R6K-ampA#1 se abrió utilizando la enzima de restricción BspHI y se sometió a CIP para liberar la resistencia a la ampicilina. El nuevo fragmento de inserción contiene el codón de resistencia a ampicilina 25 optimizado para la expresión en *E. coli*, pero todas las secuencias de CpG que podrían eliminarse mediante el uso de codones alternativos se habían reemplazado (véase la Tabla 2 anterior). Este fragmento se pidió en GeneArt. Para liberar el fragmento de inserto, el vector de clonación GeneArt n.º 1016138 se digirió usando BspHI. Después de la purificación de los fragmentos tanto de inserto como de estructura principal mediante electroforesis en gel, se ligaron y se transformaron en bacterias PIR1. Las minipreparaciones se enviaron directamente para su 30 secuenciación. pGLEX41-R6K-MCS-ampB#1 tiene la secuencia correcta y se usó para etapas de clonación adicionales.

Clonación de la construcción indicadora en vectores de expresión derivados de pGLEX41

35 Para clonar la construcción indicadora REP en los vectores de expresión pGLEX41-MCS-R6K-ampA y pGLEX41-MCS-R6K-ampB, los vectores se cortaron usando las enzimas de restricción NheI y ClaI. El vector de expresión pGLEX41-HM-MCS se abrió usando las enzimas de restricción NheI y BstBI (a 65 °C). Todas las estructuras principales de vectores se trataron con CIP después de la digestión y las estructuras principales se purificaron mediante electroforesis en gel. Las estructuras principales se ligaron con el fragmento NheI/BstBI (BstBI es 40 compatible con ClaI) que codifica la construcción indicadora REP. Los productos de ligación se transformaron en bacterias competentes PIR1 o TOP 10 y se sembraron en placas de ampicilina LB-agar. Las colonias se amplificaron y los plásmidos se aislaron mediante minipreparaciones. Los clones positivos se pudieron identificar por digestión de restricción de minipreparaciones y confirmación de secuencia posterior por FASTERIS SA.

45 Adición de secuencias flanqueantes de GAPDH en vectores de expresión derivados de pGLEX41

Todos los análisis de restricción de este párrafo se realizaron en un volumen final de 80 µl y se incubaron durante una noche a 37 °C.

50 La secuencia 5'GAPDH (SEQ ID NO: 7) se escindió de pCR-blunt-5'GAPDH usando la enzima de restricción NruI y

se ligó en los vectores de expresión pGLEX41-R6K-ampⁱA-[REP] y pGLEX41-R6K-ampⁱB-[REP] que se linealizaron usando NruI y se trataron con CIP para evitar la recirculación. Después de la amplificación de las colonias PIR1 (obtenidas por transformación de productos de ligación) se analizaron minipreparaciones por digestión de restricción. Los clones pGLEX41-R6K-ampⁱA-5'GAPDH-[REP] #2 y pGLEX41-R6K-ampⁱB-5'GAPDH-[REP] #1 mostraron bandas del tamaño esperado en el análisis de restricción, se confirmaron posteriormente mediante secuenciación y se usaron para más etapas de clonación. Estos nuevos vectores se abrieron entonces con Scal y se trataron con CIP. El fragmento 3'GAPDH (SEQ ID NO: 8) se escindió de pCR-Blunt-3'GAPDH usando la misma enzima y se ligó en las dos estructuras principales con el fin de generar los vectores de expresión pGLEX41-R6K-ampⁱA-GAPDH-[REP] y pGLEX41-R6K-ampⁱB-GAPDH-[REP].

10

La digestión de control de los clones pGLEX41-R6K-ampⁱA-GAPDH-[REP] # 2 y pGLEX41-R6K-ampⁱB-GAPDH-[REP] # 8 mostró bandas del tamaño esperado en el análisis de restricción. La inserción del fragmento 3'GAPDH en la orientación correcta se confirmó posteriormente por secuenciación (Fasteris).

15 II.4. Clonación de vectores de resistencia

El punto de partida para la clonación de los vectores de resistencia fue el vector pGLEX-MCS-R6K-ampⁱA#1. En cuanto a la expresión de genes de resistencia, un promotor débil es suficiente, el promotor mCMV fue reemplazado por el promotor SV40. Los genes que codifican los genes de resistencia se pidieron a GeneArt SA (Regensburg, Alemania) y se optimizaron para la expresión en hámster chino (puromicina: puroA y neomicina: neoA) o se redujeron en contenido de CpG mediante el uso de codones selectivos (puromicina: puroB y neomicina: neoB).

20

Clonación de pGLEX-R6K-AmpⁱA-PuroA/PuroB:

Para clonar la resistencia a la puromicina en el casete de expresión, se abrió el vector pGLEX41-MCS-R6K-ampⁱA#1 usando las enzimas de restricción NruI y XbaI, seguido de tratamiento con CIP. El fragmento de inserción se pidió a GeneArt y se proporcionó como inserto en el vector de clonación GeneArt n.º 1013239. Contiene el promotor SV40 y el gen optimizado por codón para la resistencia a la puromicina (para el uso de codones de células CHO). El inserto se cortó del vector de clonación GeneArt usando las enzimas NruI y XbaI (las mismas que se usan para la estructura principal), se purificó y se clonó en el fragmento de la estructura principal. Las minipreparaciones se prepararon y se analizaron mediante digestión de restricción. El clon pGLEX-MCS-R6K-ampⁱA-puroA#1 mostró el perfil correcto y pudo confirmarse mediante secuenciación.

30

Este vector se usó para la clonación del vector pGLEX-MCS-R6K-ampⁱA-puroB mediante el intercambio de la región codificante por el gen de resistencia a la puromicina, mientras se abandona el promotor SV40. El nuevo fragmento de inserción contiene una versión optimizada por codones del gen puromicina, donde se han reemplazado todas las secuencias CpG que podrían eliminarse debido al uso de codones alternativos. El fragmento se ha pedido por GeneArt y se administró en el vector de clonación n.º 1016139. Para liberar el fragmento de inserción, el vector GeneArt se digirió usando las enzimas de restricción XbaI y NotI. El fragmento de inserción se purificó mediante electroforesis en gel y se clonó en la estructura principal de pGLEX-MCS-R6K-ampⁱA-puroA, después de la liberación del marco de lectura abierto de puromicina mediante digestión de restricción usando XbaI y NotI, seguido de tratamiento con CIP. El vector resultante pGLEX-MCS-R6K-ampⁱA-puroB#1 se confirmó directamente mediante análisis de secuencia.

35

40

45 **Clonación de los vectores pGLEX-R6K-ampⁱA-NeoA y pGLEX-R6K-ampⁱA-NeoB**

Para clonar la resistencia a la neomicina en el casete de expresión, el vector pGLEX-R6K-puroA#1 se abrió usando las enzimas de restricción XbaI y NotI, seguido de tratamiento con CIP. Los fragmentos de inserción se pidieron a partir de GeneArt y se proporcionaron como insertos en los vectores de clonación GeneArt n.º 1013242 (neoA) y n.º 1026894 (neoB). Contienen el gen optimizado por codón para la resistencia a la neomicina para el uso de codones de las células CHO y la versión reducida de CpG de la resistencia a la neomicina, respectivamente. Los insertos se cortaron de los vectores de clonación GeneArt usando las enzimas XbaI y NotI (las mismas que se usan para la estructura principal), se purificaron y se clonaron en el fragmento de la estructura principal. Las minipreparaciones se prepararon y los clones se confirmaron por secuenciación.

50

55

Clonación de vectores pGLEX-R6K-ampⁱB-NeoB y pGLEX41-R6K-ampⁱB-puroB:

El vector pGLEX41-R6K-puroB#1 se abrió usando la enzima de restricción BspHI y posteriormente se sometió a CIP. El fragmento de inserción contiene el gen de resistencia a la ampicilina que estaba optimizado por codón para la

expresión en *E. coli*, mientras que todas las secuencias CpG que podrían eliminarse debido al uso de codones alternativos se han reemplazado. Este fragmento se ha pedido a GeneArt y ha llegado en el vector de clonación n. 1016138. Para liberar el fragmento de inserto, el vector de clonación GeneArt se digirió usando BspHI. Después de la purificación del fragmento tanto de inserción como de estructura principal mediante electroforesis en gel, se ligaron y se transformaron en bacterias PIR1. Las minipreparaciones se enviaron directamente para su secuenciación y pudieron confirmarse (pGLEX41-ampB-R6K-puroB#1).

La clonación que conduce al vector pGLEX-R6K-neoB-ampB se realizó abriendo pGLEX-R6K-neoB-ampA usando las enzimas de restricción BspHI para crear el fragmento de la estructura principal. La digestión de pGLEX-R6K-ampB-hygroB usando la misma combinación de enzima de restricción produjo el fragmento de inserción que codifica ampB. El inserto ampB se clonó en la estructura principal pGLEX-R6K-neoB-ampA.

II.5 Adición de secuencias aguas arriba y aguas abajo del gen GAPDH humano en vectores de resistencia

El vector pCR-blunt-5'GAPDH se digirió con NruI para obtener el inserto 5'GAPDH (3164 pb). Los vectores que codifican los genes de resistencia se digirieron con NruI, posteriormente se trataron con CIP (fosfatasa intestinal de ternero, NEB, Ipswich, MA) con el fin de preparar los fragmentos de la estructura principal. Los 4 fragmentos de estructura principal diferentes (pGLEX-R6K-neoA-ampA, pGLEX-R6K-neoB-ampB, pGLEX-R6K-puroA-ampA u pGLEX-puroB-ampB) se ligaron con el inserto GAPDH 5' de 364 pb y se transformaron en bacterias competentes PIR1. La digestión de restricción de minipreparaciones usando ApalI permitió la identificación de los clones pGLEX-R6K-neoB-ampB-5'GAPDH#5, pGLEX-R6K-neoA-ampA-5'GAPDH #6, pGLEX-R6K-puroA-ampA-5'GAPDH #16 y pGLEX-puroB-ampB-5'GAPDH #5.

Estos vectores intermedios se cortaron entonces con la enzima de restricción Scal y se trataron con CIP para preparar las estructuras principales para la ligación. El vector que porta el segundo fragmento de inserción, pCR-Blunt-3'GAPDH, se cortó usando Scal para liberar el fragmento de inserción (3224 pb) de la región flanqueante aguas abajo de GAPDH. Las cuatro moléculas de estructura principal diferentes se ligaron con el fragmento de inserción de 3224 pb purificado y se transformaron en células competentes PIR1. Las minipreparaciones se analizaron mediante digestión de restricción. Los clones que muestran fragmentos de restricción del tamaño esperado eran pGLEX-R6K-neoB-ampB-GAPDH #8, pGLEX-R6K-neoA-ampA-GAPDH #1, pGLEX-R6K-puroA-ampA-GAPDH #1 u pGLEX-puroB-ampB-GAPDH #4. Los clones se confirmaron posteriormente por análisis de secuenciación (Fasteris, Ginebra, Suiza).

II.1.5. Midipreparaciones de plásmidos clonados para transfección

Para tener cantidades suficientes de plásmidos, se prepararon midipreps usando el kit Macherey Nagel (NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Suiza). Después de la confirmación por digestión de restricción y secuenciación, los plásmidos se linealizaron y se usaron para la transfección en células CHO-S. La Tabla 3 resume las concentraciones de lotes de ADN plasmídico obtenidos en midipreparaciones, preparaciones de ADN linealizadas que se habían preparado para la transfección, las enzimas utilizadas para la linealización y los archivos de secuencias de Fasteris SA confirmando la identidad y la información de secuencia del plásmido respectivo. Todas las midipreparaciones se confirmaron por secuenciación antes de ser utilizados para transfecciones.

Tabla 3: Resumen de plásmidos clonados. Concentración de midipreparación de ADN y midipreparación linealizada (con la enzima correspondiente). El número GSC codifica el plásmido respectivo y permite identificar los archivos de secuenciación relevantes.

| Plásmidos | Conc. de midipreparación (µg/ml) | Enzima para linealización | Conc. de plásmidos linealizados (µg/ml) | Código de plásmido Glenmark |
|------------------------------|----------------------------------|---------------------------|---|-----------------------------|
| pGLEX41-R6K-AmpA-[REP]-GAPDH | 1538 | EcoRV | 1019 | GSC 2774 |
| pGLEX41-R6K-AmpB-[REP]-GAPDH | 1243 | EcoRV | 1233 | GSC 2775 |
| pGLEX-R6K-AmpA-neoA-GAPDH | 890 | AseI | 766 | GSC 2776 |
| pGLEX-R6K-AmpB-neoB-GAPDH | 594 | AseI | 979 | GSC 2777 |

| | | | | |
|-----------------------------|------|-------|------|----------|
| pGLEX-R6K-AmpiA-puroA-GAPDH | 917 | AseI | 859 | GSC 2778 |
| pGLEX-AmpiB-puroB-GAPDH | 869 | AseI | 1049 | GSC 2779 |
| pGLEX41-[REP] | 2119 | BspHI | 868 | GSC 2239 |
| pGLEX41-R6K-AmpiA-[REP] | 865 | BspHI | 779 | GSC 2240 |
| pGLEX41-R6K-AmpiB-[REP] | 1751 | BspHI | 806 | GSC 2249 |
| pGLEX-R6K-AmpiA-neoA | 890 | BspHI | 764 | GSC 2214 |
| pGLEX-R6K-AmpiB-neoB | 767 | BspHI | 654 | GSC 2244 |
| pGLEX-R6K-AmpiA-puroA | 708 | BspHI | 659 | GSC 2220 |
| pGLEX-R6K-AmpiB-puroB | 574 | BspHI | 746 | GSC 2213 |

Ejemplo 2: Transfección de células con vectores de expresión:

1. Materiales y métodos

5

Células CHO-S y células HEK293

Las células de mamífero son el huésped preferido para expresar proteínas porque son capaces de plegar, ensamblar y modificar post-transcripcionalmente de forma correcta proteínas recombinantes. La línea celular CHO se usó porque están bien caracterizadas y no sirven como huésped para la mayoría de los virus patógenos humanos, lo que los convierte en un huésped relativamente seguro para la producción estable de proteínas terapéuticas. Las células de ovario de hámster chino (CHO-S, Invitrogen, Carlsbad, CA, Estados Unidos) se cultivaron en suspensión en medio PowerCHO-2 CD (Lonza, Verviers, Bélgica), se complementaron con L-glutamina 4 mM (Applichem, Alemania) y se incubaron en una incubadora de sacudidas (200 rpm con una carrera circular de 2,5 cm) a 37 °C, CO₂ al 5 % y 80 % de humedad. Las células HEK293 se utilizan porque son fáciles de transfectar y permiten la producción rápida de proteínas recombinantes hasta cantidades de menos gramos. Las células utilizadas son células HEK293-EBNA (ATCC, Manassas, VA) y se cultivan rutinariamente en suspensión en medio Ex-cell 293 (Sigma-Aldrich, St. Louis, MI).

Se realizaron rutinariamente subcultivos de células CHO-S y HEK293 EBNA cada 3-4 días usando una densidad de siembra de $0,5 \times 10^6$ células viables/ml en medio fresco. Las células se cultivaron utilizando 10 ml de medio en tubos de biorreactor de 50 ml (Tubespin Bioreactor 50; TPP, Trasadingen, Suiza) que contenían un filtro permeable que permitía el intercambio de gases. La viabilidad celular y la concentración se determinaron con el contador celular automatizado Countess (Invitrogen, Carlsbad, CA, Estados Unidos) utilizando el método de exclusión de células azules tripánicas. La concentración celular se confirmó mediante la determinación del método del volumen celular empaquetado (PCV) usando tubos de PCV (TPP, Trasadingen, Suiza) para células CHO-S.

Volumen celular empaquetado (PCV)

El método de PCV se basa en la centrifugación de un volumen específico de líquido de cultivo en un tubo de mini-PCV (PCV Tubo de volumen celular empaquetado; TPP, Trasadingen, Suiza) durante 1 minuto a 5000 rpm. Durante la centrifugación, las células se sedimentan en el capilar graduado en la base del tubo. El porcentaje de volumen celular empaquetado se determina luego evaluando el volumen del sedimento en relación con la cantidad de fluido de cultivo celular centrifugado. Por ejemplo, el PCV al 1 % indicó que 10 µl de sedimento celular estaban presentes en 1 ml de fluido de cultivo.

Para el recuento celular de rutina de las células, se añadieron por pipeteo 200 µl de cada muestra en un tubo de PCV y se leyó el volumen del sedimento correspondiente (en µl) con una regla (dispositivo de medición de "lectura fácil"; TPP, Trasadingen, Suiza). Este volumen se multiplicó por 5 para tener el valor de 1 ml y luego se multiplicó usando un factor de correlación específico de células para obtener una estimación de la concentración de células

viables (en millones de células/ml).

Recuento celular "automatizado"

5 La concentración celular y la viabilidad se determinaron con el contador celular automatizado Countess® (Invitrogen, Carlsbad, CA, Estados Unidos) en la mezcla de la muestra con la misma cantidad de azul de tripano. Después, la solución se añade por pipeteo en el portaobjetos de la cámara Countess® antes de leerse por el instrumento. Este instrumento permite una lectura automática de la cámara de Neubauer que, después de la calibración, determina la viabilidad celular y la concentración de células muertas y vivas.

10

Análisis de citometría de flujo

La citometría de flujo es una técnica para el análisis de múltiples parámetros de células individuales. Esta técnica permite el análisis cuantitativo y cualitativo de células que son fenotípicamente diferentes entre sí, por ejemplo, 15 muertas de células viables (de acuerdo con el tamaño y la granularidad de las células). También permite la cuantificación de células que expresan una proteína de interés, tal como GFP. Las células se recogieron del cultivo mediante pipeteo estéril de 300 µl de muestras y se analizaron con un citómetro de flujo Calibur de clasificación celular asociada a fluorescencia (FACS) (Becton, Dickinson and Company, Franklin Lakes, NJ, Estados Unidos) equipado con un láser de argón refrigerado por aire que emite a 488 nm. Los análisis se realizaron con el software 20 CellQuest. La emisión de GFP se detectó con FL-1, utilizando un filtro de paso de banda de 530/30 nm.

En la primera puerta, los restos celulares, así como las células muertas se excluyeron del análisis en un gráfico de puntos de SSC/FSC en escala lineal. Después, la fluorescencia GFP de las células vivas se mostró en un histograma en escala logarítmica. El valor medio de la distribución de fluorescencia se utilizó para evaluar el nivel de 25 expresión de GFP de las poblaciones celulares analizadas.

Método de cuantificación de IgG: OCTET QK

El sistema Octet QK (FortéBio, Menlo Park, CA, Estados Unidos) realiza la cuantificación sin etiqueta de anticuerpos, 30 proteínas, péptidos, ADN y otras biomoléculas y proporciona la caracterización cinética de las interacciones de unión biomolecular. Una correlación entre la velocidad de unión (nm) y la concentración de IgG1 acumulada (µg/ml) de la muestra permite la cuantificación del título de IgG con una curva de calibración.

Las muestras celulares se centrifugaron durante 5 minutos a 300 g. El sobrenadante luego se diluyó (1/5 para el 35 anticuerpo IgG1) con el tampón octeto en una placa de 96 pocillos antes de analizarse con el octeto utilizando biosensores de proteína A (DIP de proteína A y Biosensor READ™, Forté Bio, Estados Unidos) para obtener la concentración de anticuerpo por pocillos.

Transfección transitoria usando JetPEI

40

Se realizó la transfección transitoria y estable de células CHO-S y HEK293 EBNA usando polietilenimina (PEI, JetPEI, Polyplus-transfection, Illkirch, Francia). PEI es un polímero catiónico que puede formar complejos con moléculas cargadas negativamente tal como ADN. El complejo de ADN-PEI cargado positivamente se une a la 45 superficie celular cargada negativamente y se internaliza mediante endocitosis. Alcanza el compartimiento lisosómico desde donde se libera por lisis al núcleo. La alta eficacia de transfección con complejos de ADN-PEI se debe a la capacidad de PEI para proteger el ADN de la degradación lisosómica. Las células se transfectaron de acuerdo con el manual proporcionado por el fabricante.

Todos los plásmidos se linealizaron antes de la transfección estable (100 µg de ADN resuspendido en 100 µl de Tris- 50 EDTA, pH 7,5). Para la transfección transitoria, se usaron directamente plásmidos circulares a partir de ADN de midipreparación. En este estudio, las transfecciones transitorias se mantuvieron en tubos de biorreactor de 50 ml y no se añadieron antibióticos.

Se obtuvieron clones de CHO-S estables que expresaban IgG1 y GFP por cotransfección de un vector de expresión 55 y dos vectores de resistencia (que codifican resistencia a puromicina o neomicina, respectivamente).

Selección de grupos y minigrupos estables

La eficacia de transfección se determinó 24 h después de la transfección mediante citometría de flujo (citómetro BD

FACS Calibur, n.º 1293) mediante el análisis de la expresión de GFP intracelular. Si el porcentaje de células positivas a GFP era superior al 20 %, las células transfectadas se diluyeron con medio selectivo y se distribuyeron en placas de 96 pocillos (para limitar la dilución para generar minigrupos estables aislados) o en T-Flasks (para generar grupos estables). El medio selectivo utilizado fue PowerCHO-2, glutamina 4 mM, complementado con diferentes concentraciones de geneticina y puromicina.

Siete días después de la transfección, la rigurosidad de la selección se renovó añadiendo medio de selección a las células. Tan pronto como las colonias en placas de 96 pocillos fueron confluentes, las placas se leyeron usando un lector de fluorescencia.

Los grupos en T-Flasks se expandieron a escala de tubospin usando PowerCHO-2 libre de antibióticos, L-glutamina 4 mM. Su viabilidad y concentración se evaluaron con el contador celular automatizado Countess (Invitrogen, Carlsbad, CA, Estados Unidos). Tan pronto como la densidad celular lo permitió, se inició un tren de siembra para cada grupo sembrando células a una densidad de $0,5 \times 10^6$ células/ml en 10 ml de medio en tubos de biorreactor de 50 ml (incubados en un agitador (200 rpm) en CO₂ al 5 %, 37 °C y 80 % de humedad). Cada tren de siembra se hizo pasar dos veces por semana sembrando las células a $0,5 \times 10^6$ células/ml en medio de crecimiento (la concentración de células se determinó por análisis de PCV). El tren de siembra se utilizó para el inóculo de todas las realizaciones de producción (lotes).

Durante las siguientes 4-5 semanas, las realizaciones de producción se sembraron una vez por semana en duplicados. La estabilidad del grupo se evaluó mediante expresión de FACS e IgG como se ha descrito anteriormente para poblaciones clonales.

Realizaciones de producción (fermentación por lotes)

Se sembraron las realizaciones en lote de grupos de células a una concentración de $0,5 \times 10^6$ células/ml usando el tren de siembra para la inoculación y luego se cultivaron las células durante 7 días en medio de alimentación. Los días 4 y 8, se centrifugaron 200 µl de células durante 5 minutos a 300 g y se analizó el sobrenadante en busca de IgG acumulada usando el Octeto. Además, la expresión de GFP de cada lote fue analizada por FACS.

2. Resultados

2.1 Expresión en transitorio en células CHO:

Los vectores comparados en este estudio difieren principalmente en su estructura principal. El casete de expresión completo (promotor, primer intrón, construcción de expresión, poli (A)) es exactamente el mismo para todos los vectores. Los vectores se derivan del vector pGLEX41 como se describe en el Ejemplo 1. En un vector, el gen de resistencia a ampicilina se optimizó por codón para la expresión en *E. coli* y la estructura principal bacteriana se redujo a un mínimo: pGLEX41-R6K-AmpiA-[REP] (en resumen A). En un segundo vector, el gen de resistencia a la ampicilina se optimizó por codón para la expresión en *E. coli*, pero se evitaron todas las secuencias CpG, mediante el uso de codones alternativos (cuando fue posible): Este vector se denomina pGLEX41-R6K-AmpiB-[REP] (en resumen B). La tercera modificación incluyó el uso de las secuencias flanqueantes de GAPDH que se clonaron aguas arriba y aguas abajo del casete de expresión de los vectores A y B, dando los vectores pGLEX41-R6K-AmpiA-[REP]-GAPDH (en resumen GAPDH A) y pGLEX41-R6K-AmpiB-[REP]-GAPDH (en resumen GAPDH_B).

Se realizaron transfecciones transitorias de células CHO-S (Invitrogen) con el fin de comparar el nivel de expresión de las proteínas indicadoras expresadas en el contexto de las diferentes estructuras principales del plásmido. Las transfecciones (por duplicado) se realizaron en tubos de biorreactor de 50 ml (TPP, Trasadingen, Suiza) usando 10 ml de volumen de medio final y se analizaron el día 5 después de la transfección con Octet (Figura 2).

Todos los vectores (A y B) con la estructura principal corregida muestran un nivel de expresión ligeramente más alto que los vectores de control pGLEX41. Solo hay una pequeña diferencia entre los vectores A y B. Esto es esperado, porque la única diferencia en la estructura principal es la resistencia a la ampicilina que no debería tener un impacto en la expresión transitoria.

La observación más llamativa es el efecto positivo de las secuencias de GAPDH en la expresión. Se obtiene un nivel de expresión 2 veces mayor con el plásmido que alberga las secuencias flanqueantes de GAPDH en comparación con los que no tienen las secuencias de GAPDH. Esto es cierto para las construcciones A y B. En comparación con el vector pGLEX41, se puede observar una expresión 3 veces mayor. Esto es incluso más sorprendente si se tiene

en cuenta el tamaño de los plásmidos. El vector A (7048 pb) es casi la mitad del tamaño en comparación con el vector GAPDH-A (13436 pb). Por lo tanto, se supone que la cantidad de ADN entregado durante el proceso de transfección transitoria es la misma para todos los plásmidos, solo la mitad de la cantidad molar de GAPDH-A se administra al núcleo.

5

2.2 Expresión en transitorio en células HEK293

Se realizaron transfecciones transitorias de células HEK293 EBNA con el fin de comparar el nivel de expresión de las proteínas indicadoras expresadas en el contexto de las diferentes estructuras principales del plásmido. Las transfecciones (por duplicado) se realizaron en tubos de biorreactor de 50 ml (TPP, Trasadingen, Suiza) usando 10 ml de volumen de medio final y se analizaron el día 10 después de la transfección con Octet (Figura 3).

10

Los resultados mostrados en la Figura 3 muestran un aumento significativo en la expresión que se puede obtener usando las regiones flanqueantes de GAPDH en células HEK293 EBNA. El vector GAPDH-B muestra un aumento de tres veces en la expresión, mientras que el vector GAPDH-A muestra un aumento aún mayor en la expresión de 5 veces. Estos vectores no contienen el elemento oriP y, por lo tanto, pueden tener un potencial para títulos incluso mayores.

15

2.3 Expresión en líneas celulares CHO estables

20

Establecimiento de células transfectadas estables

Se generaron poblaciones estables co-transfectando un vector de expresión y vectores que codifican genes de resistencia, seguido de presión de selección mediada por antibióticos. La presión de selección se eliminó 14 días después de la transfección. Estas etapas permitieron la generación de minigrupos estables y grupos estables que se cultivaron en intervalos regulares en las realizaciones de producción con el fin de comparar los niveles de expresión de las proteínas indicadoras (anticuerpo IgG1 y GFP) de las diferentes construcciones y la estabilidad de la expresión.

25

30 Estudio de expresión de proteína indicadora en realizaciones de producción realizadas con grupos de células

Los grupos se generaron por transfección estable. Durante el procedimiento de selección (los primeros 14 días después de la transfección) los grupos se analizaron mediante FACS. Se pudo observar un aumento de la fracción de células positivas para GFP junto con la viabilidad del cultivo a lo largo del tiempo. La presión de selección mediada por los antibióticos se eliminó de los grupos después de 14 días. Usando este enfoque, no se pueden obtener grupos celulares transfectadas con los plásmidos "B". El nivel de expresión de los grupos generados se ensayó tan pronto como las células pudieron cultivarse en tubos de biorreactor de 50 ml. Los lotes se hicieron por duplicado. Las células se analizaron mediante FACS para la expresión de GFP y la acumulación de IgG en el sobrenadante se ensayó mediante Octet después de 8 días de expresión.

35

40

Se pudo observar una relación proporcional entre los títulos de IgG y la expresión de GFP de los grupos. Por lo tanto, solo los datos de IgG se muestran en la Figura 4. Todos los grupos transfectados con vectores que contienen la secuencia de GAPDH muestran una expresión más alta en comparación con el vector pGLEX41 o con el mismo vector sin la secuencia de GAPDH (factor de 2,8 entre A y A-GAPDH. No puede sacar ninguna conclusión entre B y B-GAPDH ya que no ha sobrevivido ningún grupo B). Las transfecciones realizadas con A-GAPDH y B-GADPH indujeron una mayor expresión de IgG (2,7 y 3,5 veces más, respectivamente) que la transfección de pGLEX41 (para el lote 2). Por lo tanto, en los grupos, las secuencias flanqueantes de GAPDH parecen ser favorables para la producción de proteínas. Finalmente, las transfecciones realizadas con vectores B-GADPH indujeron una mayor expresión de IgG que la transfección realizada con A-GAPDH (factor de 1,25). Por lo tanto, la reducción de CpG en los genes de resistencia también parece ser favorable para la producción estable de proteínas.

45

50

Estudio de nivel de expresión en poblaciones clónicas

Las células se transfectaron y se distribuyeron en placas de 96 pocillos en medio selectivo con el fin de obtener poblaciones clonales u oligoclonales. Después de 7 días, la presión de selección se refrescó mediante la adición de medio selectivo a las células. La expresión de GFP se evaluó 14 días después de la transfección usando un lector de placas ELISA. Los resultados se muestran en la Figura 5.

55

Confirmando los resultados obtenidos en grupos celulares, las células transfectadas con vectores que contenían secuencias flanqueantes de GAPDH expresaron significativamente más GFP que la misma estructura principal sin secuencias aguas arriba y aguas abajo de GAPDH (factores de 1,7 a 2 veces) o los demás vectores usados como control (pGLEX41: 2,5 veces) (Figura 5). Además, las poblaciones con vectores que contenían secuencias de resistencia que habían sido reducidas por CpG (B) indujeron una expresión más alta que los vectores correspondientes que solo habían sido optimizados por codón (A) (1,5 veces entre A y B; 1,2 veces entre B y B-GAPDH).

Del estudio de expresión se sacaron varias conclusiones. En primer lugar, la secuencia de GAPDH aguas arriba y aguas abajo permite una expresión más alta que el vector estándar que se usó como punto de referencia (pGLEX41). También se obtiene un nivel de expresión más bajo cuando las células se transfectan con la misma estructura principal del vector sin que las secuencias de GAPDH confirmen que el efecto beneficioso sobre la expresión está relacionado con las secuencias flanqueantes de GAPDH insertadas. Además, la reducción del número de CpG en los plásmidos de expresión y selección parece ser ligeramente favorable para la expresión, también.

Ejemplo 3: Nivel de expresión transitoria de células CHO-S GMP transfectadas con nuevos vectores diseñados

Se ha descrito en la bibliografía que la región 5' del promotor GAPDH alberga una insulina potencial, así como un elemento de respuesta de éster de forbol (Alexander-Bridges et al., (1992) *Advan Enzyme Regul*, 32: 149-159). El elemento de respuesta de éster de forbol (-1040 -1010 pb) está situado aguas arriba de lo que habitualmente se conoce como el promotor GAPDH (-488 - +20). En un estudio de delección realizado en líneas celulares de Hepatoma H35 estables, los autores no pudieron demostrar un efecto significativo de la eliminación de pares de bases -1200 a -488 (en relación con el punto de inicio de la transcripción). Por lo tanto, el elemento de respuesta de éster de forbol podría no estar funcionalmente unido a la expresión impulsada por el promotor GAPDH. Sin embargo, se realizó un experimento de transfección transitoria para evaluar la contribución de insulina y PMA (forbol-12-miristato-13-acetato, el éster de forbol más común) en el aumento de la expresión transitoria y estable que se observó utilizando los plásmidos que contenían los elementos flanqueantes de GAPDH.

Para obtener un medio de crecimiento libre de insulina, se preparó PowerCHO2 a partir de un medio en polvo y no se añadió insulina. PMA se adquirió en Sigma (St. Louis, MO), y se dosificó a una concentración final de 1,6 μ M (correspondiente a la concentración utilizada por Alexander-Bridges en líneas celulares de Hepatoma H35) en PowerCHO2 (+/- Insulina).

Las transfecciones se realizaron en tubos de biorreactor de 50 ml (Tubespins, TPP, Trasadingen, Suiza) como se ha descrito previamente. Con el fin de evitar la presencia de insulina proporcionada por OptiMEM (Life technologies, Carlsbad, CA), el medio de transfección se cambió a RPMI1640 (PAA, Pasching, Austria) complementado con 4 Gln mM y HEPES 25 mM. Después de la transfección, las células se distribuyeron en placas de 12 pocillos y se añadió 1 ml de los cuatro medios diferentes (PowerCHO2, Gln 4 mM, +/- insulina; PowerCHO2, Gln 4 mM, PMA 1,6 μ M, +/- insulina). Nuevamente, se usó la construcción indicadora que expresa IgG1 y GFP utilizando dos IRES (descritos en el Ejemplo 2). Este vector permitió la verificación de la eficacia de la transfección. El porcentaje y la viabilidad de las células transfectadas se encontraron similares en las cuatro preparaciones de medios diferentes.

Como se muestra en la Figura 6, no se pudo observar ningún efecto significativo del agotamiento de la insulina y/o la adición de PMA durante este experimento. Se obtuvieron títulos similares en todos los medios utilizados para la expresión. Esto sugiere que los potenciales ésteres de forbol y los elementos de respuesta a la insulina presentes en la secuencia flanqueante aguas arriba del gen GAPDH no afectan a la expresión transgénica transitoria.

Ejemplo 4: Análisis de fragmentación del ADN que flanquea el casete de expresión de GAPDH aguas arriba del promotor y aguas abajo del sitio poli A con el fin de estudiar el efecto sobre la expresión del gen indicador

El locus GAPDH humano está localizado en el cromosoma 12 del genoma humano. GAPDH se describe como constitutivamente activo en todas las células de origen mamífero, ya que la enzima es un elemento clave en el metabolismo de la glucosa. Aguas arriba del promotor, el gen GAPDH está flanqueado por NCAPD2, un gen que se extiende en más de 30000 pb. Aguas abajo del sitio de poliadenilación, el gen GAPDH está flanqueado por IFFO1 (véase la Figura 7 para más detalles).

No solo GAPDH y el promotor, sino también las regiones flanqueantes están bien conservadas entre diferentes especies (véase la Tabla 4).

Tabla 4: Tramos de homologías altas entre las regiones flanqueantes de GAPDH humanas, de rata y de ratón. El análisis se realizó usando el administrador de clones 9 (ScieED, Cary, NC, Estados Unidos). La numeración es relativa a la primera base del elemento flanqueante aguas arriba o aguas abajo, respectivamente (Secuencia ID NO: 7 y Secuencia ID NO: 8, respectivamente). Las secuencias usadas para la alineación fueron para las bases de ratón 532-3731 (aguas arriba) y 8164-11364 (aguas abajo) de la Secuencia ID No 18 y para las bases de rata 719-3918 (aguas arriba) y 8495-11058 (aguas abajo) de la Secuencia ID No 19.

| Región aguas arriba | | | | Aguas abajo | | | |
|--------------------------------|---------|---------------------------------|---------|--------------------------------|-----------|---------------------------------|-----------|
| Secuencias de homología [rata] | | Secuencias de homología [ratón] | | Secuencias de homología [rata] | | Secuencias de homología [ratón] | |
| >80 % | >90 % | >80 % | >90 % | >80 % | >90 % | >80 % | >90 % |
| 161-249 | 279-331 | 15-69 | 278-329 | 1608-1764 | 1706-1764 | 1614-1671 | 1904-2061 |
| 256-338 | 554-623 | 159-249 | 546-626 | 1894-2067 | 1912-2061 | 1888-2072 | 2927-3071 |
| 515-659 | | 273-342 | | | | 2918-3082 | |
| 2296-2349 | | 515-647 | | | | | |
| 2381-2513 | | 1143-1223 | | | | | |
| 2736-2818 | | 1957-2009 | | | | | |
| | | 2029-2080 | | | | | |
| | | 2375-2485 | | | | | |
| | | 2730-2821 | | | | | |

Una comparación de la homología de ADN entre roedores y seres humanos muestra un mínimo de conservación de ADN del 38 %. La presencia de un tramo conservado de ADN fuera de una región promotora o una región que codifica un gen indica que puede haber una presión de selección sobre la célula para mantener la secuencia de ADN o para permitir solo ciertos cambios/cambios menores. En este caso específico, las regiones flanqueantes de GAPDH podrían ser importantes para las células porque mantienen un alto nivel de expresión de los genes GAPDH. Los cambios en la secuencia de ADN que conducen a la disminución de la expresión se seleccionarán en contra.

Para evaluar la contribución del elemento GAPDH aguas arriba y aguas abajo al aumento observado en la expresión, se prepararon construcciones que contenían solo la región flanqueante de GAPDH aguas arriba (SEQ ID NO: 7), fragmentos de la región flanqueante de GAPDH aguas arriba o la región flanqueante de GAPDH aguas abajo (SEQ ID NO: 8). El anticuerpo indicador de tipo IgG1 se expresó mediante una construcción IRES (cadena ligera-IRES-cadena pesada), evitando por lo tanto la cotransfección de plásmidos múltiples. Los detalles sobre la fragmentación del fragmento aguas arriba de GAPDH se muestran en la Figura 8. Se usaron los siguientes fragmentos de la región flanqueante de GAPDH aguas arriba: Fragmento 1 (SEQ ID NO: 9), fragmento 2 (SEQ ID NO: 10), fragmento 3 (SEQ ID NO: 11), fragmento 4 (SEQ ID NO: 12), fragmento 8 (SEQ ID NO: 13), fragmento 9 (SEQ ID NO: 14), fragmento 11 (SEQ ID NO: 15), fragmento 17 (SEQ ID NO: 16).

La región flanqueante de GAPDH aguas arriba (SEQ ID NO: 7) usada contiene 2 veces 3 (en total 6) nucleótidos del sitio de restricción NruI de los cuales tres están unidos al ADN genómico en su 5' y tres están unidos al ADN genómico en su extremo 3'. La región flanqueante GAPDH aguas abajo (SEQ ID NO: 8) usada contiene dos veces 3 (en total 6) nucleótidos del sitio de restricción Scal de los cuales tres están unidos al ADN genómico en su 5' y tres están unidos al ADN genómico en su extremo 3'. La región flanqueante de GAPDH aguas arriba y la región flanqueante de GAPDH aguas abajo sin los nucleótidos del sitio de restricción respectivo se muestran en la SEQ ID NO: 20 (región flanqueante de GAPDH aguas arriba sin sitios de restricción) y SEQ ID NO: 21 (región flanqueante de GAPDH aguas abajo sin sitios de restricción). Los fragmentos de la región flanqueante de GAPDH aguas arriba utilizada contienen cada uno 3 nucleótidos del sitio de restricción respectivo en su extremo 5' y/o 3' unido al ADN genómico (el Fragmento 1 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 5'; el Fragmento 2 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 3'; el Fragmento 3 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 5'; el Fragmento 4 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 3'; el Fragmento 8 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 3'; el Fragmento 9 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 5' y 3 nucleótidos del sitio de restricción NruI en su extremo 3'; el Fragmento 11 contiene 3 nucleótidos del sitio de restricción NruI en su extremo 3'). El Fragmento 17 no contiene nucleótidos de un sitio de restricción. Los fragmentos de la región flanqueante de GAPDH aguas arriba sin los nucleótidos del sitio de restricción respectivo se muestran en la SEQ ID NO: 22 (fragmento 1 sin sitio de restricción), SEQ ID NO: 23 (fragmento 2 sin sitio de restricción) SEQ ID NO: 24 (fragmento 3 sin sitio de restricción), SEQ ID

NO: 25 (fragmento 4 sin sitio de restricción), SEQ ID NO: 26 (fragmento 8 sin sitio de restricción), SEQ ID NO: 27 (fragmento 9 sin sitios de restricción), SEQ ID NO: 28 (fragmento 11 sin sitio de restricción).

El efecto de los elementos GAPDH aguas arriba y aguas abajo sobre la expresión se evaluó el día 10 después de la transfección usando el octeto (Fortebio, Menlo, CA, Estados Unidos) para cuantificar la cantidad de IgG1 secretada en el sobrenadante (véase la Figura 9) pGLEX41, el vector original está dando resultados de expresión más bajos (80 %) en comparación con el diseño de vector nuevo mejorado utilizado en las estructuras principales pGLEX41-amp^r. En comparación con la estructura original pGLEX41, el nuevo diseño incluye la optimización por codón del gen amp^r necesario para la resistencia a la ampicilina en *E. coli*, un origen de replicación diferente (R6K en lugar del origen de replicación de pUC) y la eliminación de secuencias enlazadoras innecesarias (o espaciadoras) de origen bacteriano. Ambos vectores tienen aproximadamente el mismo tamaño.

Sorprendentemente, pGLEX41-amp^r que incluye el elemento aguas arriba (SEQ ID NO: 7) y aguas abajo (SEQ ID NO: 8), (denominado pGLEX41-ascendente/descendente en la Figura 9 que muestra los resultados de la expresión) está dando una mayor expresión (factor 1,5) en comparación con el mismo vector sin las secuencias aguas arriba y aguas abajo. Si se considera la diferencia de tamaño (los fragmentos ascendentes/descendentes aumentan el tamaño del plásmido en aproximadamente 6000 pb) y, por lo tanto, las diferencias en las copias de plásmido entregadas durante la transfección, el efecto podría ser aún más importante por plásmido.

El vector que contiene solo el fragmento aguas arriba (ascendente) muestra un nivel de expresión similar a la construcción de expresión original pGLEX41-amp^r. El vector que contiene solo el fragmento aguas abajo (descendente) muestra un aumento significativo (factor 1,2) en la expresión en comparación con la construcción de expresión original pGLEX41-amp^r. Se puede observar un aumento adicional en la expresión si ambos, el fragmento ascendente y el descendente están presentes. Esto se confirma por la fragmentación de los fragmentos aguas arriba. El Fragmento 9 y el fragmento proximal promotor 8 no muestran ninguna diferencia en la expresión en comparación con pGLEX41-amp^r. Los fragmentos 1, 11 y 17 muestran un aumento en la expresión. El mayor aumento se observó para el fragmento 4. Debe destacarse que el fragmento proximal promotor 8 no muestra ningún efecto. Por lo tanto, el aumento en la expresión no puede explicarse por secuencias publicadas previamente (Alexander-Bridges et al., (1992) *Advan Enzyme Regul*, 32: 149-159), Graven et al., (1999) *Biochimica et Biophysica Acta* 147: 203-218).

De forma interesante, los fragmentos 2 y 3 conducen a una disminución significativa de la expresión. Esto es inesperado, especialmente en vista del hecho de que estos fragmentos clonados en la dirección opuesta (antisentido (AS) en la Figura 9) no causan este efecto. Para los fragmentos 1, 8, 9, 11 y 17 no se observó diferencia en la expresión para fragmentos que se integraron en orientación sentido o antisentido (datos no mostrados). El fragmento 11, aunque es una parte del fragmento 2, no muestra este efecto. Por lo tanto, el elemento de secuencia que parece ser perjudicial para la expresión debería estar al menos parcialmente en el fragmento BstBI-BstBI que se eliminó en el fragmento 2 para obtener el fragmento 11.

Además, la hipótesis de que un elemento negativo se localiza (al menos parcialmente) en el fragmento BstBI-BstBI está respaldada por el aumento en la expresión observada entre el fragmento 3 (que incluye el fragmento BstBI-BstBI) y el fragmento 1.

Si bien parece fácil localizar el fragmento que tiene un efecto negativo (BstBI-BstBI), a partir de este estudio es menos obvio cómo este efecto negativo observado para el fragmento 2 y 3 se compensa mediante elementos de secuencia presentes en el fragmento aguas arriba completo. Podría ser que este efecto negativo se equilibre con el pequeño efecto positivo que se observó en el fragmento 1 y el fragmento 4 (pero el aumento en la expresión del fragmento 1 es menor que en el fragmento 4). Sin embargo, el efecto positivo para el fragmento 4 (factor 1,25) observado parece menos importante en comparación con el efecto negativo (factor 0,4). Además, el fragmento 9, que es toda la región aguas arriba sin el fragmento BstBI-BstBI, no muestra expresión aumentada en comparación con toda la región flanqueante aguas arriba de GAPDH (sin embargo, el fragmento 9 incluye el fragmento EcoRV-BstBI que es parte del fragmento 2 y 3 y podría tener un efecto negativo sobre la expresión).

Solo se puede especular sobre el mecanismo detrás de los efectos observados. La dependencia de la orientación del efecto negativo sobre la expresión observada con los fragmentos 2 y 3 excluye la expresión de marcos de lectura abiertos no identificados (por ejemplo, expresión de un ARNnc), porque no hay promotores circundantes que puedan desencadenar la expresión de una sola orientación. El hecho de que la expresión se reduzca por debajo del nivel inicial muestra, no solo la ausencia de un efecto positivo (por ejemplo, una actividad potenciadora), sino más bien la presencia de un efecto negativo dependiente de la orientación.

En resumen, se observa un aumento sorprendente de la expresión transitoria en células CHO si ambas regiones flanqueantes, la región aguas arriba y aguas abajo, están presentes en el plásmido de expresión. Aunque el fragmento 4 parece tener un efecto positivo significativo sobre la expresión, no se pudo identificar ningún fragmento
 5 único que sea responsable del aumento total de la expresión que se observó. El aumento de la expresión del vector de expresión pGLEX41-amp^rA (ascendente/descendente) parece ser el efecto resumen de las regiones flanqueantes tanto aguas arriba y aguas abajo.

Ejemplo 5: Clonación de la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino y el promotor de hámster chino

1.1 Clonación de la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino en un vector de expresión

15 La secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino se amplificó en ADN genómico de células CHO-S (Life Technologies) por PCR. El ADN genómico se extrajo como se describe en el Ejemplo 1. Las construcciones se prepararon usando el promotor CMV de ratón o el promotor GAPDH de hámster chino para la expresión de la construcción del gen indicador [REP] descrito en el Ejemplo 1.

20 Para la clonación de la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino en combinación con el promotor CMV de ratón, se usaron los cebadores GlnPr1896 y GlnPr1897 para la amplificación del fragmento de 3 kbs (pb 672 a 3671 de SEQ ID No 29) usando el protocolo de PCR descrito en el Ejemplo 1 y que conduce al amplicón con la SEQ ID No 30. El amplicón contiene la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino y los sitios de restricción 5' y 3' que se introdujeron por los cebadores.

25 Para la clonación de la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino en combinación con el promotor GAPDH de hámster chino, se usaron los cebadores GlnPr1902 y GlnPr1905 para amplificar el fragmento de 3508 pb que contenía la secuencia de ADN genómico que incluía la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino y el promotor GAPDH (pb 672 a 4179 de SEQ ID No 29) que
 30 conduce al amplicón con la SEQ ID No 31. En una segunda PCR, se usaron GlnPr1901 y GlnPr1902 para la amplificación del fragmento de 508 pb que contenía solo la región promotora (pb 3672 a 4179 de SEQ ID No 29), que conduce a SEQ ID No 32. El intrón utilizado en el vector "A" (descrito en el Ejemplo 1) se amplificó usando los cebadores GlnPr1903 y GlnPr1904.

35 Se realizó una primera PCR de fusión con los cebadores GlnPr1904 y GlnPr1901 usando el amplicón con SEQ ID NO: 32 y el amplicón con la secuencia del intrón como plantillas. El amplicón contiene el promotor GAPDH de hámster chino, un intrón y sitios de restricción 5' y 3' que fueron introducidos por los cebadores. Todos los cebadores se muestran en la Tabla 5.

40 Se realizó una segunda PCR de fusión con los cebadores GlnPr1905 y GlnPr1904 usando el amplicón con SEQ ID No. 31 y el amplicón con la secuencia del intrón como plantillas. El amplicón contiene la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino, el promotor GAPDH de hámster chino, un intrón y los sitios de restricción 5' y 3' que se introdujeron por los cebadores.

45 Después de la purificación en un gel de agarosa al 1 %, las bandas de interés se cortaron y se purificaron usando el kit "NucleoSpin Gel and PCR Clean-up" (Macherey Nagel, Oensingen, Suiza). Los fragmentos purificados se clonaron en el plásmido pCR Blunt usando el kit de clonación por PCR Zero Blunt (Invitrogen, Carlsbad, CA, Estados Unidos). Los productos de ligación se transformaron en *E. coli* competente TOP10 (*E. coli* competente de One Shot® TOP 10; Invitrogen, Carlsbad, CA, Estados Unidos) y se analizaron mediante análisis de restricción de
 50 minipreparaciones. Esto condujo a los plásmidos pCR_blunt[CHO-upstreamGAPDH], que contenían la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino, pCR_Blunt[CHO-upstreamGAPDH_GAPDHpromoter] que contenía la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino y el promotor GAPDH e intrón del vector "A" y pCR_Blunt[CHO-GAPDHpromoter] que contenía el promotor GAPDH y el intrón del vector "A".

55 Para la evaluación de los amplicones en su efecto sobre la expresión de un gen secretado, se usó el vector "A" (descrito en el Ejemplo 1). Como se ha descrito previamente, el casete de expresión utilizado en este vector contiene un gen policistrónico que codifica una IgG1 y GFP secretadas (véase el Ejemplo 1). Las células transfectadas, por lo tanto, secretarán el anticuerpo monoclonal IgG1 y acumularán GFP intracelular de una manera dependiente.

Para liberar el fragmento de inserción de 3 kb que contenía la secuencia de ADN genómico aguas arriba del gen GAPDH de hámster chino, el plásmido pCR_Blunt[CHO-upstreamGAPDH] se digirió usando la enzima de restricción NaeI. Este inserto se clonó en la estructura principal de "A", se digirió usando la enzima de restricción NruI y se sometió a CIP (CIP, NEB, Ipswich, MA, Estados Unidos). La estructura principal y el inserto se ligaron entre sí usando la ADN ligasa T4 (ADN ligasa T4, NEB, Ipswich, MA, Estados Unidos) y posteriormente se transformaron en *E. coli* competente PIR1. Los clones se seleccionaron para la preparación de minipreparaciones y el posterior análisis de restricción. El plásmido resultante se denominó "A_GAPDH_UP", se confirmó mediante análisis de secuenciación y se produjo a escala de midipreparación utilizando el kit NucleoBond Xtra Midi (Macherey Nagel, Oensingen, Suiza).

Para la clonación de construcciones de expresión usando el promotor GAPDH de hámster chino, los fragmentos de inserción se liberaron de los plásmidos pCR_Blunt[promotor CHO-upstreamGAPDH_GAPDH] y pCR_Blunt[CHO-GAPDHpromoter] por digestión usando las enzimas de restricción NheI y NruI. Los fragmentos resultantes se clonaron en la estructura principal del vector "A", se abrieron utilizando las mismas enzimas y se sometieron a CIP. Después de la ligación con ADN ligasa T4 y la transformación en *E. coli* competente PIR1, se seleccionaron clones para el análisis de restricción de minipreparación. Los plásmidos resultantes se denominaron "A_GAPDH_UP_Prom" (plásmido con secuencia de ADN genómico no traducida aguas arriba de GAPDH de hámster chino y el promotor) y "A_PR" (plásmido con solo el promotor), se confirmaron mediante análisis de secuenciación y se produjeron a escala de midipreparación usando el kit NucleoBond Xtra Midi (Macherey Nagel, Oensingen, Suiza).

2. Evaluación del efecto de la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino sobre la expresión de la construcción del gen indicador

Se transfectaron células CHO-S en biorreactores tubespín usando 10 ml de volumen de medio (como se describe en el Ejemplo 2). Las células transfectadas se incubaron en una incubadora de agitación con agitación de 200 rpm a 37 °C, CO₂ al 5 % y 80 % de humedad. Los sobrenadantes de las células se analizaron para determinar la expresión de IgG1 usando el sistema Octet QK con biosensores de Proteína A (FortéBio, Menlo Park, CA, Estados Unidos). Los resultados se muestran en la Figura 10.

El nivel de expresión del plásmido que contiene el promotor GAPDH ("A PR") en comparación con el promotor CMV de ratón (A) se reduce en un 50 %, lo que indica que el promotor GAPDH de hámster chino no es tan fuerte como el promotor viral. El plásmido que contiene la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino en combinación con el promotor GAPDH de hámster chino ("A GAPDH UPProm") muestra un aumento de expresión doble en comparación con la construcción que tiene solo el promotor GAPDH ("A PR"). El plásmido que contiene la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino y el promotor CMV de ratón ("A GAPDH UP") muestra la expresión más alta y un aumento de más del 40 % sobre el plásmido que contiene solo el promotor CMV de ratón ("A"). Esto confirma que la secuencia de ADN genómico no traducida aguas arriba del gen GAPDH de hámster chino tiene un efecto potenciador sobre la expresión de la proteína indicadora.

Tabla 5: Cebadores usados para la clonación en el Ejemplo 5

| Cebador | SEQ ID No | Secuencia | Orientación | Sitio de restric. |
|------------|--------------|--|-------------|-------------------|
| GlnPr 1896 | SEQ ID No 33 | TACGGCCGGCTTCACTGTACAGTGGCACAT | directo | NaeI |
| GlnPr 1897 | SEQ ID No 34 | TCAGGCCGGCCGTGGTTCTTCGGTAGTGAC | inverso | NaeI |
| GlnPr 1901 | SEQ ID No 35 | TACTCGCGAAGAAGATCCTCAACTTTTCCACAGCC | directo | NruI |
| GlnPr 1902 | SEQ ID No 36 | GTTCCTAAACGAGCTCTGCTATTTATAGGAACTGGGGTG | inverso | / |
| GlnPr 1903 | SEQ ID No 37 | CACCCCAGTTCCTATAAATAGCAGAGCTCGTTTAGTGAAC | directo | / |
| GlnPr 1904 | SEQ ID No 38 | CGCTAGCACCGGTTCGATCGA | inverso | NheI |
| GlnPr 1905 | SEQ ID No 39 | TACTCGCGATTCACTGTACAGTGGCACATAC | directo | NruI |

LISTA DE SECUENCIAS

<110> Glenmark Pharmaceuticals SA

5 <120> Casete de expresión

<130> GBR2001/PCT

10 <150> US 61/567.675

<151> 07-12-2011

<160> 39

15 <170> PatentIn versión 3.3

<210> 1

<211> 39

<212> ADN

<213> Artificial

20 <220>

<223> Cebador 1

25 <400> 1

attattcgcg atggctcctg gcatctctgg gaccgagcc 39

<210> 2

<211> 39

<212> ADN

30 <213> Artificial

<220>

<223> Cebador 2

35 <400> 2

atcgtcgcga agcttgagat tgtccaagca ggtagccag 39

<210> 3

<211> 36

<212> ADN

40 <213> Artificial

<220>

<223> Cebador 3

45 <400> 3

agcaagtact tctgagcctt cagtaatggc tgcctg 36

<210> 4

<211> 39

<212> ADN

<213> Artificial

50 <220>

55 <223> Cebador 4

<400> 4

tggcagtact aagctggcac cactacttca gagaacaag 39

ES 2 679 291 T3

<210> 5
 <211> 3179
 <212> ADN
 <213> Artificial

5

<220>
 <223> Fragmento 5' GAPDH

<400> 5
 atcgtcgcga agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa 60
 ccatccactg gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct 120
 gcagtgattt gtttcttctt cttttttaaa tcacgttttc ctgccttttc taggttctag 180
 gtaccagcct ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg 240
 tactaccctg cggcatccaa acaccagca gcgagcttcc aaaaagaaac ccaaagtgtg 300
 cttctcaagt gatgagtcca gtgaggaagg tatgatgctc ccgctgttc ccggccgaga 360
 aggcacacag ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt 420
 gtagatgctc tgccccactg ccgcagaagg gcctctcctg tacagcttgg attttatttc 480
 ttctgtgcgg tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg 540
 tgactcagct ttttcttatt cctttaattc tttgcataga tctttcagca gagatgacag 600
 aagacgagac acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat 660
 cctaggaagt ctgttcctgt cctccctgtg cagggtatcc tgtagggtga cctggaattc 720
 gaattctggt tcccttgtaa aatatttgtc tgtctctttt ttttaaaaaa aaaaaaggcc 780
 gggcactgtg gctcacgcct gtaatcccag cactttgcga taccaaggcg ggtggataac 840
 ctgaggtagg gagttcgaga ccagcctgac caacatggag aaaccccatc tctactaaaa 900
 ataaaaaatt agccgggctg attggcgtgc gcctgtaatc ccagctactc aagaggctga 960
 ggcaggagaa tcgcctgaac ccagaggcgg aggttgtagt gagccgaaat cacaccattg 1020
 cactccagct tgggcaacaa tagcgaacct ccatctcaa ttaaaaaaa aatgcctaca 1080
 cgctctttaa aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct 1140
 tcaactttgg aatatatggt tgccaatctc cttgttttct aatgaataaa tgtttttata 1200
 tactttttaga cattttttcc taagcttgtc tttgtttcat ctttcacatt agcccagttt 1260
 catgcagcag agagaggggt atcagtgcag agagagatga gtgagcccag agtcctaggg 1320
 cctgtcccgg gatggcagat gagcttctctg ccccgctcact gccacctttc ccctctcaac 1380
 ctctggacc cgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggg 1440
 tccagatcag tgcttctgaa ccgggggcaa ttttgtctgc cagaggacat ctgacaacac 1500
 ctggggcctg ttttgttgtc atagcctata ggggaagaat gctaccagca tttgtgggaa 1560
 gaggccaggg atgtggctca acatcctgca gtgcacagga tggcccctca acaaagaatc 1620

10

ES 2 679 291 T3

acacggccca caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg 1680
ctttctgocg tgtgcctcac occaccccca ctcgtgttcc ctaatcccat ctccaaaggt 1740
tggcagcaga ccggcccagc ctcgtggaag ttcagatcat gatcccctcc agctctgcag 1800
gagacaagac ctgtctccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct 1860
atgctgcagg cgggctgccc cctgaagcct gcgcaccctc ctccagctcc tcaagtcttc 1920
tctgctgagt caccttcgaa ccggaggctg tgagctggct gtcgtgacca cactgggtgcc 1980
tctgctgtca tgacaacagc aactacgtc agtagtgctc cctgggact gagctccctc 2040
tttgccggga gaagacagta atgaaaaatg acaagcatga ggcagagggg aagatcacgc 2100
ttgggtggtg caggagcatg gaggtgctct taatgctctc aatgagaaag ggtaaacggt 2160
cctggttgca ggaatagctg agtcagaggt ggggcttctc cactcccc accccacccc 2220
tttcaccatt agggacctc ttgccttgct ctgtctactc tgctctgggt ggtcattgtg 2280
aaaagccgc accaacctg ccagtgagcag ccagacgagg acacagcctg gctctgggtc 2340
ccagcaggaa aggcaatccc agaaaggcag ggtcagggac tggagtcctg tgggtgcttt 2400
ttaagcaaag attatcacca ggcaggttaa acttagcaac cggcttttag ctagaagggc 2460
aggggctggt tgctcaggta tgctgggcca gcaagaggc ccgggatccc cctccatgc 2520
acctgctgat gggccaaggc cccccaccc caccocctc cttacaagtg ttcagcacc 2580
tcccatacca cactcacaaa cctggccctc tgcctccta ccagaagaat ggatcccctg 2640
tgggaggggg caggggacct gttcccaccg tgtgcccaag acctctttc ccacttttc 2700
cctcttcttg actcaccctg ccctcaatat ccccggcgc agccagtgaaggaggctccc 2760
tggctcctgg ctgcctgca cgtcccaggc cggggaggga cttccgcctc cacgtcccgc 2820
tcttcgcccc aggctggatg gaatgaaagg cactctgtc ctctccctag gcagcacagc 2880
ccacaggttt ccaggagtgc ctttgtggga ggcctctggg ccccaccag ccacctgtc 2940
ctccgcctgg ggccccagcc cggagagagc cgctggtgca cacagggccg ggattgtctg 3000
ccctaattat caggtccagc ctacagggtc gcaggacatc gtgaccttc gtgcagaaac 3060
ctccccctcc ccctcaagcc gcctcccagc cctccttctc ctccaggccc ccagtgccca 3120
gtgccagtg cccagcccag gcctcggctc cagagatgcc aggagccatc gcgaataat 3179

<210> 6
<211> 3238
<212> ADN
<213> Artificial

<220>
<223> Fragmento 3'-GAPDH

<400> 6
tggcagtact aagctggcac cactacttca gagaacaagg ccttttcctc tcctcgtccc 60

5

10

ES 2 679 291 T3

agtcctaggc tatctgctgt tggccaaaca tggagaagc tattctgtgg gcagccccag 120
 ggagctgac aggtggagga agtcagggct cgcactgggc tctgacgctg actggttagt 180
 ggagctcagc ctggagctga gctgcagcgg gcaattccag cttggcctcc gcagctgtga 240
 ggtcttgagc acgtgctcta ttgctttctg tgcctctgtg tcttatctga ggacatcgtg 300
 gccagcccct aaggtcttca agcaggattc atctaggtaa accaagtacc taaaacctag 360
 cccaaggcgg taaggactat ataatgttta aaaatcggta aaaatgccca cctcgcatag 420
 ttttgaggaa gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg 480
 gaacttgggt aggggcaagg cgtgtagctg ggacctaggt ccagaccctt ggctctgcca 540
 ctgaacggct cagttgcttt gggcagttac tcccggcctt cactttgcac gtgtgcttac 600
 ctagtggaga caaaagtaca tacctcggta gagcgcgcac gcctgtaacc ccagcacttt 660
 gggaggccaa ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat 720
 ggtgaaactc cgtctctact aaaattacaa aatcagcca ggcttcatgg cacatgccta 780
 tagtcccagc tacaggcatg ctgaagcagg agaatcgctt gcaccccgga ggcagaggct 840
 gcagtgagct gagaccacac cactgcactc cagcctaggc aacagagtat gagactccat 900
 ctcaaaaaaa aaaaaagtac ctacctcaga gttcaacta gtgaatatta ggaagtgctt 960
 gagacagtga caccaaagtg cacaataaat actcgccagt ttcatatta ttaaagaatc 1020
 catttgaatg tcagctcaac acagcctcct ataccgaggc attgtgaacc gcactcctcc 1080
 agcttctcca ggcttttcca agaatcaggg aactgttagc ctgttggtct cagtgtatga 1140
 cagacacgga ggaagcaca ctttagctga tacttaaaaca gagaccctga gcgcacatac 1200
 acccgcgcac acatgcatgg agcttcacct tctctgtcat tctgcagtga ccaggagagc 1260
 aagagctccc acctcccttc aaaacactgt gccatcccg ggcactaagg cctctttaa 1320
 gcacggcacc tccacgaggg agggccacag ccacatacac tccacctggc aggtggacag 1380
 cgtgagcacg tggacatag cagggacaag gtgccccggc cagcccacac gccctctgcc 1440
 gctgacaggg acagaagccc tctccagctg cgtgtgctgc agaggccatg cgtagcctcc 1500
 agctgcattc tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga 1560
 gctggctccg gacaacaggg atggaggaaa ggtcccacat tcacattcct gatagctgga 1620
 caagtgaggg ggccgcaatc gctctggcag cattttaaag atggggaagt agcagacacc 1680
 cacgcgtgaa ggcaggagag cccaactgt ggtggaaatg gcccagaat ggtaggcca 1740
 agcctagctc cagacacccc agagccctgg agaagccaag actgagggag aaagcctgag 1800
 ggaggagcgc cccagctccc agggaccggc ctggtgcaga gctgcagctg atgttcccct 1860
 ctgtgcagcc ccacctctg cctcgtgag ctccctgctg cgagggcctc gggtgcaagg 1920

ES 2 679 291 T3

gggaggcagg tctctatctc atggagctgt cagatgagac atcgcgatcg gagtccctcag 1980
cctcgcttgg cggcggcggc gggtcgctaa gcgggaccgc agtgaaagca ggagactttc 2040
tagaaaaaaa caccagttgt caacctggg gcaggcagga atcctgaaga cggacggcac 2100
tcctcctcct gctgcctcac cctctggcag cccgtgagaa gtaccggaag cgagggcggg 2160
gccgcgggat ggcgagggag cggcagggac tgaactctct ccaaaccac cctgacaggg 2220
aaatgggccc cgcctgtgtc ttgggaactc agaggctgag gtcaggcatg atggctcacg 2280
cctgtaattc cagcactttg ggaggcagag gcgggtggat cacgacgtca ggagtccaag 2340
gcaagcctgg ccaacatggt gaaaccccat ctctactaaa aatgcaaaa ttagccaggt 2400
gtggtggcgg gcgccttga ggctgaggca agataatcac ttgaacctgg gaggcggagg 2460
ttgtagttag ccaaaaaaaa aaaaaaaga aatagctgaa gtcacagtag gagagaagct 2520
gctgagcctc cagcaccctg actctagggc cttggcttta tgtctatctg cagtattttt 2580
gtgattttta aaaattcact ttcttgttgc ggtgtaactt acacagggtc aaatgcacaa 2640
atcatggccc tgactttaga taaaaatctg cccccacaac ccttctgttc cttgccagtt 2700
tttaactgc ctctaaccag ggaaccacc agagctgggt gccttgggag gtttcagccc 2760
tcccgtcatg aatggacata gctcatccaa ctgccaaggg agagagctgt gggctctgggc 2820
cagccccacc aggtaactcc caaagggcag cccacagca agatgtgacc cagtcattgc 2880
ctgagggctc ctggggctgt gttccaacct ttctccccgc tgtgtcccc tggaggcccc 2940
catgcccagg ggaggcgcct accggtctcc agactgggtg atgagccggc ggcaggtctc 3000
catctgcacg tccaggccgc gcttcatgct gcacatctcc atgtactcgt gcaggtgccg 3060
gttcatgtcg ttcttggccg tggccaactc cagctgtggt ggggcaggca gggccatcgg 3120
ggttaacagg tggcgttcac agcgcctctg ttgccccgc caggaggcca acacgccaag 3180
agcagtggct gggccggggg ccaggcagc cattactgaa ggctcagaag tacttgct 3238

<210> 7
<211> 3164
<212> ADN
<213> Artificial

5

<220>
<223> Región flanqueante de GAPDH aguas arriba

10

<400> 7
cgaagcttga gattgtccaa gcaggtagcc agagagcgc atcagccaag aaaccatcca 60
ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120
tttgtttctt cttctttttt aaatcacgtt ttctctgctt ttctaggttc taggtaccag 180
cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240
cgtcggcatc caaacaccca gcagcgagct tccaaaaaga aaccctaaagt tgtcttctca 300

ES 2 679 291 T3

agtgatgagt ccagtgagga aggtatgatg ctccccgctg ttccccggccg agaaggcaca 360
 cagctagggg gcagagggct ggttccata ggacctgctg cgggggcctg agtgtagatg 420
 ctctgcccc ctgccgaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480
 cgggtgtggga ttgtctcact tgttctctga tatctatfff ttcaccatct ttgtgactca 540
 gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600
 gacacccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660
 agtctgttcc tgtcctccct gtgcagggtg tcctgtaggg tgacctgga ttcgaattct 720
 gtttcccttg taaaatattt gtctgtctct tttttttaa aaaaaaaag gccgggcaact 780
 gtggctcacg cctgtaatcc cagcactttg cgataccaag gcgggtggat aacctgaggt 840
 agggagtcc agaccagcct gaccaacatg gagaaacccc atctctacta aaaataaaaa 900
 attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagaggc tgaggcagga 960
 gaatcgccctg aaccagaggg cggagggtgt agtgagccga aatcacacca ttgactcca 1020
 gcttgggcaa caatagcgaa cctccatctc aaattaaaa aaaaatgcct acacgctctt 1080
 taaaatgcaa ggctttctct taaattagcc taactgaact gcgttgagct gcttcaactt 1140
 tggaatatat gtttgccaat ctcttggtt tctaatgaat aaatgtttt atatactttt 1200
 agacatffff tctaagcct gtctttggtt catctttcac attagcccag tttcatgcag 1260
 cagagagagg gttatcagt cagagagaga tgagtgagcc cagagtccta gggcctgtcc 1320
 cgggatggca gatgagcttc ctgccccgtc actgccacct tccccctctc aacctctgga 1380
 ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtgccta ggctccagat 1440
 cagtgccttc gaaccggggg caatfffgtc tgccagagga catctgacaa cacctggggc 1500
 ctgttttgtt gtcatagcct ataggggaag aatgctacca gcatttggg gaagaggcca 1560
 gggatgtggc tcaacatcct gcagtgcaca ggatggcccc tcaacaaaga atcacacggc 1620
 ccacaatgtc aatagcgtca cagttgagaa aacctgctct agaccaaggg ttgctttctg 1680
 ccgtgtgcct caccccaccc cactcgtgt tocctaatec catctccaaa ggttggcagc 1740
 agaccggccc aggctcgtgg aagttcagat catgatcccc tccagctctg caggagacaa 1800
 gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860
 aggcgggctg ccccctgaag cctgcgcacc cctctccagc tcctcaagtc ttctctgctg 1920
 agtcaccttc gaaccggagg ctgtgagctg gctgtcgtga ccacactggt gcctctgctg 1980
 tcatgacaac agcacactac gtcagtagtg ctccctgggc actgagctcc ctctttgagg 2040
 ggagaagaca gtaatgaaaa atgacaagca tgaggcagag ggaagatca cgcttgggtg 2100
 gtgcaggagc atggaggtgc tcttaatgct ctcaatgaga aagggttaac ggtcctgggt 2160

ES 2 679 291 T3

gcaggaatag ctgagtcaga ggtggggctt cctccactcc cccaccccac ccctttcacc 2220
 attagggacc ttcttgccct gctcttgcta ctctgctctg ggtggtcatt gtgaaaagcc 2280
 cgcaccaacc atgccagtgg cagccagacg aggacacagc ctggctctgg gtcccagcag 2340
 gaaaggcaat cccagaaag cagggtcagg gactggagtc ctgtgggtgc tttttaagca 2400
 aagattatca ccaggcaggc taaacttagc aaccggcttt tagctagaag ggcagggggc 2460
 tgggtgcagg ttatgctggg ccagcaaaga ggccogggat ccccctcca tgcacctgct 2520
 gatgggcaa ggccaccca cccaccccc ttcttataa gtgttcagca ccctccatc 2580
 ccacactcac aaacctggcc ctctgccctc ctaccagaag aatggatccc ctgtgggagg 2640
 gggcagggga cctgttccca ccgtgtgcc aagacctctt ttcccacttt ttccctcttc 2700
 ttgactcacc ctgccctcaa tatccccgg cgcagccagt gaaagggagt ccctggctcc 2760
 tggctcgcct gcacgtccca gggcggggag ggacttccgc cctcacgtcc cgctcttcgc 2820
 cccaggctgg atggaatgaa aggcacactg tctctctccc taggcagcac agcccacagg 2880
 tttccaggag tgcctttgtg ggaggcctct gggcccccac cagccatcct gtcctcgcgc 2940
 tggggcccca gcccgagag agccgctggt gcacacaggg cgggattgt ctgccctaat 3000
 tatcaggtcc aggctacag gctgcaggac atcgtgacct tccgtgcaga aacctcccc 3060
 tccccctcaa gccgcctccc gagcctcctt cctctccagg ccccagtgcc ccagtgccca 3120
 gtgccagcc caggcctcgg tcccagagat gccaggagcc atcg 3164

<210> 8
 <211> 3224
 <212> ADN
 <213> Artificial

5

<220>
 <223> Región flanqueante de GAPDH aguas abajo

10

<400> 8
 actaagctgg caccactact tcagagaaca aggcctttc ctctcctcgc tccagtccta 60
 ggctatctgc tgttggccaa acatggaaga agctattctg tgggcagccc cagggaggct 120
 gacaggtgga ggaagtcagg gctcgcactg ggctctgacg ctgactggtt agtggagctc 180
 agcctggagc tgagctgcag cgggcaattc cagcttggcc tccgcagctg tgaggtcttg 240
 agcacgtgct ctattgcttt ctgtgccctc gtgtcttacc tgaggacatc gtggccagcc 300
 cctaaggtct tcaagcagga ttcatctagg taaaccaagt acctaaaacc atgccaagg 360
 cggtaaggac tatataatgt ttaaaaatcg gtaaaaatgc ccacctcgca tagttttgag 420
 gaagatgaac tgagatgtgt cagggtgact tatttccatc atcgtcctta ggggaacttg 480
 ggtaggggca aggcgtgtag ctgggaccta ggtccagacc cctggctctg cactgaacg 540
 gctcagttgc tttgggcagt tactccccgg cctcactttg cacgtgtgct tacctagtgg 600

ES 2 679 291 T3

agacaaaagt acatacctcg gtagagcgcg cacgcctgta accccagcac tttgggaggc 660
 caagtggggt gtatcacctg aggtcaggag tttgagacca gcctggccea catggtgaaa 720
 ctccgtctct actaaaatta caaaaatcag ccaggcttca tggcacatgc ctatagtcct 780
 agctacagggc atgctgaagc aggagaatcg cttgcacccc ggaggcagag gctgcagtga 840
 gctgagacca caccactgca ctccagccta ggcaacagag tatgagactc catctcaaaa 900
 aaaaaaaaaag tacctacctc agagttcaaa ctagtgaata ttaggaagtg cttgagacag 960
 tgacacccaa gtgcacaata aatactcgcc agtttcatta ttattaaaga atccatttga 1020
 atgtcagctc aacacagcct cctataccga ggcattgtga accgcatctc cccagcttct 1080
 ccaggctttt ccaagaatca gggacactgt agcctgttgg tctcagtgtg tgacagacac 1140
 ggaggaagca catctttagc tgatacttaa acagagaccc tgagcgcaca tacaccgcg 1200
 cacacatgca tggagcttca ccttctctgt cattctgcag tgaccaggag agcaagagct 1260
 cccacctccc ttcaaaacac tgtgccatc cggggcacta aggctcttt aaagcacggc 1320
 acctccacga gggagggcca cagccacata cactccacct ggcagggtga cagcgtgagc 1380
 acgtggacca tagcaggagc aagtgcccc gccagcccc aacgcctct gccgctgaca 1440
 gggacagaag ccctctccag ctgctgtgtc tgcagaggcc atgcgtagcc tccagctgca 1500
 ttctattcca ctccagtgcc tgggccagtt agcaccagtg tggaagacag tgagctggct 1560
 ccggacaaca gggatggagg aaaggtccca cattcacatt cctgatacgt ggacaagggt 1620
 aggggcccga atcgctctgg cagcatttta aagatgggga agtagcagac acccacgct 1680
 gaaggcagga gagccccaac tgtggtgaa atggcccag aatggtaggg ccaagcctag 1740
 ctccagacac cccagagccc tggagaagcc aagactgagg gagaaagcct gagggaggag 1800
 cgccccagtc cccagggacc ggctggtgc agagctgcag ctgatgttcc cctctgtgca 1860
 gccccaccct ctgcctcgct gagctccctg ctgcgagggc ctcggtgca agggggaggg 1920
 aggtctctat ctcatggagc tgtcagatga gacatcgcga tcggagtcct cagcctcgct 1980
 tggcggggc ggcggtcgc taagcgggac cgcagtgaaa gcaggagact ttctagaaaa 2040
 aaacaccagt tgtcaacctt ggggcaggca ggaatcctga agacggacgg cactcctcct 2100
 cctgctgcct caccctctgg cagcccgtga gaagtaccgg aagcgagggc ggggccgcg 2160
 gatggcgagg gagcggcagg gactgaactc totccaaacc caccctgaca gggaaatggg 2220
 cccgcctgt gtcttgggaa ctcagaggct gaggtcaggc atgatggctc acgcctgtaa 2280
 ttccagcact ttgggaggca gaggcgggtg gatcacgacg tcaggagttc aaggcaagcc 2340
 tggccaacat ggtgaaaccc catctctact aaaaatgcaa aaattagcca ggtgtggtgg 2400
 cgggcgcctt ggaggctgag gcaagataat cacttgaacc tgggagggcg aggtttagt 2460

ES 2 679 291 T3

gagccaaaaa aaaaaaaaaa agaaatagct gaagtcacag taggagagaa gctgctgagc 2520
ctccagcacc ctgactctag ggccttggt ttatgtctat ctgcagtatt tttgtgattt 2580
ttaaaaattc actttcttgt tgcggtgtaa cttacacagg gtcaaagca caaatcatgg 2640
ccctgacttt agataaaaat ctgccccac aacccttctg ttccttgcca gtttttaaac 2700
tgcctctaac caggggaacc accagagctg gtggccttg gaggtttcag ccctcccgtc 2760
atgaatggac atagctcatc caactgcaa gggagagagc tgtgggtctg ggccagcccc 2820
accagtaac tcccaaagg cagccccaca gcaagatgtg acccagtcac tgcctgaggg 2880
tctctggggc tgtgttccaa cttttctccc cgctgtgtcc ccctggaagg ccccatgccc 2940
aggggagggc cctaccggtc tccagactgg gtgatgagcc ggcggcaggc ctccatctgc 3000
acgtccaggc cgcgcttcat gctgcacatc tccatgtact cgtgcaggctg ccggttcatg 3060
tcgttcttgg ccgtggccaa ctccagctgt ggtggggcag gcagggccat cggggttaac 3120
aggtggcggt cacagcgct ctggtgcccc cgccaggagg ccaacacgcc aagagcagtg 3180
gctgggccgg gggcccaggc agccattact gaaggtcag aagt 3224

<210> 9
<211> 511
<212> ADN
<213> Artificial

5

<220>
<223> Fragmento 1

10

<400> 9
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60
ctggtacgta aggcagcctg tgcgggagag accagactgg ggcctcccct cctgcagtga 120
tttgtttctt cttctttttt aaatcacggt ttctgcctt ttctaggttc taggtaccag 180
cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240
cgtcggcatc caaacaccca gcagcgagct tccaaaaaga aaccxaaagt tgtcttctca 300
agtgatgagt ccagtgagga aggtatgatg ctcccgctg ttcccggccg agaaggcaca 360
cagctagggg gcagagggct ggtttccata ggacctgctg cgggggcctg agttagatg 420
ctctgcccc a ctgcccaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480
cgggtgtggga ttgtctcact tgttctctga t 511

<210> 10
<211> 2653
<212> ADN
<213> Artificial

15

<220>
<223> Fragmento 2

20

<400> 10

ES 2 679 291 T3

atctatTTTT tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
 gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
 gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat 180
 cctgtagggg gacctggaat tcgaattctg tttcccttgt aaaatatttg tctgtctctt 240
 ttttttaaaa aaaaaaaagg ccgggcactg tggctcacgc ctgtaatccc agcactttgc 300
 gataccaagg cgggtggata acctgaggta gggagttcga gaccagcctg accaacaatgg 360
 agaaacccca tctctactaa aaataaaaaa ttagccgggc gtattggcgt gcgcctgtaa 420
 tcccagctac tcaagaggct gaggcaggag aatcgctga acccagaggc ggaggttga 480
 gtgagccgaa atcacacat tgcactccag cttgggcaac aatagcgaac ctccatctca 540
 aattaaanaa aaaatgccta cacgctcttt aaaatgcaag gctttctctt aaattagcct 600
 aactgaactg cgttgagctg cttcaacttt ggaatatatg tttgccaatc tccttgTTTT 660
 ctaatgaata aatgttttta tatactttta gacatttttt cctaagcttg tctttgtttc 720
 atctttcaca ttagcccagt ttcattgcagc agagagaggg ttatcagtcg agagagagat 780
 gagtgagccc agagtcctag ggcctgtccc gggatggcag atgagcttcc tgccccgtca 840
 ctgccacctt tcccctctca acctctggac cctgcacagt gaccagacag cctctctggg 900
 gagaattatg cagtgcctag gctccagatc agtgcttctg aaccgggggc aattttgtct 960
 gccagaggac atctgacaac acctggggcc tgttttgttg tcatagccta taggggaaga 1020
 atgctaccag catttggtgg aagaggccag ggatgtggct caacatcctg cagtgcacag 1080
 gatggcccct caacaaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa 1140
 acctgctcta gaccaagggg tgctttctgc cgtgtgcctc accccacccc cactcgtggt 1200
 ccctaatccc atctccaaag gttggcagca gaccggccca ggctcgtgga agttcagatc 1260
 atgatcccct ccagctctgc aggagacaag acctgtctcc cagcattcct cattgttccc 1320
 gggctctcag agggcgtgag ctatgctgca ggcgggctgc cccctgaagc ctgctcacc 1380
 ctctccagct cctcaagtct tctctgctga gtcacctog aaccggaggc tgtgagctgg 1440
 ctgtcgtgac cacactggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc 1500
 tccctgggca ctgagctccc tctttgcggg gagaagacag taatgaaaa tgacaagcat 1560
 gaggcagagg ggaagatcac gcttgggtgg tgcaggagca tggaggtgct cttaatgctc 1620
 tcaatgagaa agggttaacg gtccctgggtg caggaatagc tgagtcagag gtggggcttc 1680
 ctccactccc ccaccccacc cctttcacca ttagggacct tcttgccctg ctcttgctac 1740
 tctgctctgg gtggctcattg tgaaaagccc gcaccaacca tgccagtgcc agccagacga 1800
 ggacacagcc tggctctggg tcccagcagg aaaggcaatc ccagaaaggc agggtcaggg 1860

ES 2 679 291 T3

actggagtcc tgtgggtgct ttttaagcaa agattatcac caggcaggct aaacttagca 1920
accggccttt agctagaagg gcagggggct ggtgtcaggt tatgctgggc cagcaaagag 1980
gccccggatc cccctcccat gcacctgctg atgggccaaag gccaccccac cccaccccct 2040
tccttacaag tgttcagcac cctcccatcc cacactcaca aacctggccc tctgccctcc 2100
taccagaaga atggatcccc tgtgggaggg ggcaggggac ctggtccac cgtgtgccca 2160
agacctcttt tcccactttt tccctcttct tgactcaccc tgcctcaat atccccggc 2220
gcagccagtg aaagggagt cctggctcct ggctgcctg cacgtcccag ggcggggagg 2280
gacttccgcc ctcacgtccc gctcttcgcc ccaggctgga tggaatgaaa ggcacactgt 2340
ctctctccct aggcagcaca gccacaggt ttccaggagt gcctttgtgg gaggcctctg 2400
ggccccacc agccatcctg tcctccgct ggggcccag cccggagaga gccgctggtg 2460
cacacagggc cgggattgtc tgccctaatt atcaggcca ggctacaggg ctgcaggaca 2520
tcgtgacctt ccgtgcagaa acctcccct cccctcaag ccgcctccg agcctcctc 2580
ctctccaggc cccagtgcc cagtgccag tgcccagccc aggcctcggc cccagagatg 2640
ccaggagcca tcg 2653

<210> 11
<211> 1966
<212> ADN
<213> Artificial

5

<220>
<223> Fragmento 3

10

<400> 11
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60
ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120
tttgtttctt cttctttttt aaatcacggt ttctgcctt ttctaggttc taggtaccag 180
cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240
cgtcggcatc caaacaccca gcagcgagct tccaaaaaga aaccxaaagt tgtcttctca 300
agtgatgagt ccagtgagga aggtatgatg ctccgcctg ttcccggccg agaaggcaca 360
cagctagggg gcagagggct ggtttccata ggacctgctg cgggggcctg agtgtagatg 420
ctctgcccc a tgcgcgaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480
cgggtgtggga ttgtctcact tgttctctga tatctatfff ttcacatct ttgtgactca 540
gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600
gacacccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660
agtctgttcc tgtcctccct gtgcagggta tcctgtaggg tgacctgaa ttcgaattct 720
gtttcccttg taaaatattt gtctgtctct tttttttaa aaaaaaaaa gcccggcact 780

ES 2 679 291 T3

gtggctcacg cctgtaatcc cagcactttg cgataccaag gcggtggat aacctgaggt 840
 agggagttcg agaccagcct gaccaacatg gagaaacccc atctctacta aaaataaaaa 900
 attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagaggc tgaggcagga 960
 gaatcgctg aaccagagc cggaggtgt agtgagccga aatcacacca ttgactcca 1020
 gcttgggcaa caatagcgaa cctccatctc aaattaaaaa aaaaatgcct acacgctctt 1080
 taaaatgcaa ggctttctct taaattagcc taactgaact gcgttgagct gcttcaactt 1140
 tggaatatat gtttgccaat ctcttgttt tctaatgaat aatgttttt atatactttt 1200
 agacattttt tcctaagcct gtctttgttt catctttcac attagcccag tttcatgcag 1260
 cagagagagg gttatcagtg cagagagaga tgagtgagcc cagagtccta gggcctgtcc 1320
 cgggatggca gatgagcttc ctgccccgtc actgccacct ttcccctctc aacctctgga 1380
 ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtccta ggctccagat 1440
 cagtgtctct gaaccggggg caattttgtc tgccagagga catctgaca cacctggggc 1500
 ctgttttgtt gtcatagcct ataggggaag aatgctacca gcatttgtgg gaagaggcca 1560
 gggatgtggc tcaacatcct gcagtcaca ggatggccc tcaacaaaga atcacacggc 1620
 ccacaatgtc aatagcgtca cagttgagaa aacctgctct agaccaaggg ttgctttctg 1680
 ccgtgtgctt caccaccacc ccaactcgtg tccctaatec catctccaaa ggttggcagc 1740
 agaccggccc aggctcgtg aagttcagat catgatcccc tccagctctg caggagacaa 1800
 gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860
 aggcgggctg cccoctgaag cctgcgcacc cctctccagc tcctcaagtc ttctctgctg 1920
 agtcaccttc gaaccggagc ctgtgagctg gctgtcgtga ccacac 1966

<210> 12
 <211> 1198
 <212> ADN
 <213> Artificial

5

<220>
 <223> Fragmento 4

10

<400> 12
 tgggtcctct gctgtcatga caacagcaca ctacgtcagt agtgctccct gggcactgag 60
 ctccctcttt gcggggagaa gacagtaatg aaaaatgaca agcatgaggc agaggggaag 120
 atcacgcttg ggtggtgcag gagcatggag gtgctcttaa tgctctcaat gagaaagggc 180
 taacggctct ggttgagcga atagctgagt cagaggtggg gcttctcca ctccccacc 240
 ccacccttt caccattagg gaccttcttg ccttgctctt gctactctgc tctgggtggt 300
 cattgtgaaa agcccgacc aaccatgccg gtggcagcca gacgaggaca cagcctggct 360

ES 2 679 291 T3

ctgggtccca gcaggaaag caatcccaga aaggcagggt cagggactgg agtcctgtgg 420
 gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg cttttagcta 480
 gaagggcagg gggctgggtg caggttatgc tgggccagca aagaggcccg ggatccccct 540
 cccatgcacc tgctgatggg ccaaggccac cccaccccac ccccttcctt acaagtgttc 600
 agcaccctcc catcccacac tcacaaacct ggcctctgc cctcctacca gaagaatgga 660
 tcccctgtgg gagggggcag gggacctgtt cccaccgtgt gcccaagacc tcttttccca 720
 ctttttccct cttcttgact caccctgccc tcaatatccc cgggcgcagc cagtgaaagg 780
 gagtcctctg ctctctggctc gcctgcacgt cccaggggcg ggagggactt ccgacctcac 840
 gtcccgtctc tcgcccagc ctggatggaa tgaaggcac actgtctctc tccctaggca 900
 gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca 960
 tcctgtcctc cgcctggggc cccagcccgg agagagccgc tgggtgcacac agggccggga 1020
 ttgtctgccc taattatcag gtccaggcta cagggtgca ggacatcgtg accttccgtg 1080
 cagaaacctc cccctcccc tcaagccgcc tcccagacct ccttcctctc caggccccc 1140
 gtgcccagtg cccagtgcc agcccaggcc tcggtcccag agatgccagg agccatcg 1198

5 <210> 13
 <211> 259
 <212> ADN
 <213> Artificial

10 <220>
 <223> Fragmento 8

<400> 13
 cctctggggc cccaccagcc atcctgtcct cgcctgggg cccagcccg gagagagccg 60
 ctggtgcaca cagggccggg attgtctgcc ctaattatca ggtccaggct acagggtgc 120
 aggacatcgt gaccttccgt gcagaaacct cccctcccc ctcaagccgc ctcccagacc 180
 tccttcctct ccaggcccc agtgcccagt gccagtgcc cagcccaggc ctcggtccca 240
 gagatgccag gagccatcg 259

15 <210> 14
 <211> 1947
 <212> ADN
 <213> Artificial

20 <220>
 <223> Fragmento 9

<400> 14
 cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60
 ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120
 tttgtttcct cttctttttt aaatcacggt ttctgcctt ttctaggttc taggtaccag 180

ES 2 679 291 T3

cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240
 cgtcggcatc caaacacca gcagcgagct tccaaaaaga aacccaaagt tgtcttctca 300
 agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccggccg agaaggcaca 360
 cagctagggg gcagagggct ggtttccata ggacctgctg cgggggcctg agtgtagatg 420
 ctctgcccc ctgccgcaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480
 cgggtgggga ttgtctcact tgttctctga tatctatfff ttcaccatct ttgtgactca 540
 gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600
 gacacccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660
 agtctgttcc tgtcctccct gtgcagggta tcctgtaggg tgacctggaa ttcgaaccgg 720
 aggctgtgag ctggctgtcg tgaccacact ggtgcctctg ctgtcatgac aacagcacac 780
 tacgtcagta gtgctccctg ggcactgagc tcctctttg cggggagaag acagtaatga 840
 aaaatgacaa gcatgaggca gaggggaaga tcacgcttgg gtggtgcagg agcatggagg 900
 tgctcttaat gctctcaatg agaaagggtt aacggctcctg gttgcaggaa tagctgagtc 960
 agaggggggg ctctctccac tccccaccc caccctttc accattaggg accttcttgc 1020
 cttgctcttg ctactctgct ctgggtggtc attgtgaaaa gcccgacca accatgccag 1080
 tggcagccag acgaggacac agcctggctc tgggtcccag caggaaaggc aatcccagaa 1140
 aggcagggtc agggactgga gtccctgtggg tgccttttaa gcaaagatta tcaccaggca 1200
 ggctaaactt agcaaccggc ttttagctag aagggcaggg ggctggtgct aggttatgct 1260
 gggccagcaa agaggcccg gatccccctc ccatgcacct gctgatgggc caaggccacc 1320
 ccacccacc cccttctta caagtgtca gcacctccc atcccacact cacaaacctg 1380
 gccctctgcc ctctaccag aagaatgat cccctgtggg agggggcagg ggacctgttc 1440
 ccaccgtgtg cccaagacct cttttccac tttttcctc ttcttgactc acctgcct 1500
 caatatcccc cggcgcagcc agtgaaagg agtccctggc tcctggctcg cctgcacgtc 1560
 ccagggcggg gagggacttc cgccctcacg tcccgcctt cgcccaggc tggatggaat 1620
 gaaaggcaca ctgtctctct ccctaggcag cacagcccac aggtttccag gagtgccttt 1680
 gtgggagggc tctgggcccc caccagccat cctgtcctcc gcctggggcc ccagcccgga 1740
 gagagccgct ggtgcacaca gggccgggat tgtctgccct aattatcagg tccaggctac 1800
 agggctgcag gacatcgtga ccttccgtgc agaaacctcc ccctccccct caagccgct 1860
 cccgagcctc ctctctctcc agggccccag tgcccagtgc ccagtgccca gccaggcct 1920
 cgggtcccaga gatgccagga gccatcg 1947

<210> 15
 <211> 1436
 <212> **ADN**
 <213> Artificial

ES 2 679 291 T3

<220>
<223> Fragmento 11

5 <400> 15
atctatTTTT tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acaccoaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggat 180
cctgtagggg gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gcaactgagct 300
ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
cacgcttggg tgggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
acggctcctg ttgcaggaat agctgagtcg gaggtggggc ttcctccact cccccacccc 480
acccttttca ccattagga ccttcttgcc ttgctcttgc tactctgctc tgggtgggtca 540
ttgtgaaaag cccgcaccaa ccatgccagt ggagccaga cgaggacaca gcctggctct 600
gggtcccagc aggaaaggca atcccagaaa ggaggggtca gggactggag tctctgggt 660
gctttttaag caaagattat caccaggcag gctaaactta gcaaccggct tttagctaga 720
agggcagggg gctgggtgtca ggttatgctg ggccagcaaa gaggcccggg atccccctcc 780
catgcacctg ctgatgggcc aaggccaccc cccccaccc ccttccttac aagtgttcag 840
caccctccca tcccacactc acaaacctgg ccctctgccc tcctaccaga agaatggatc 900
ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc tttcccact 960
ttttccctct tcttgactca ccctgccctc aatatcccc ggcgagcca gtgaaagga 1020
gtccctggct cctggctcgc ctgcacgtcc caggggggg agggacttcc gccctcacgt 1080
cccgtcttc gcccaggt ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
acagcccaca ggtttccag agtgccttg tggaggcct ctgggcccc accagccatc 1200
ctgtcctccg cctggggccc cagcccggag agagccgctg gtgcacacag ggccgggatt 1260
gtctgcccta attatcaggt ccaggctaca gggctgcagg acatcgtgac cttccgtgca 1320
gaaacctccc cctccccctc aagccgctc ccgagcctc ttcctctcca ggccccagt 1380
gccagtgcc cagtgccag cccaggcctc ggtcccagag atgccaggag ccatcg 1436

<210> 16
<211> 1177
<212> ADN
<213> Artificial

10

<220>
<223> Fragmento 17

15

<400> 16

ES 2 679 291 T3

atctatTTTT tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acaccoaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctagga gtctgttctt gtctccctg tgcagggtat 180
cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gactgagct 300
ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
cacgcttggg tggcgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggta 420
acggctcctg ttgcaggaat agctgagtca gaggtggggc ttcctccact cccccaccc 480
accctttca ccattagga ccttcttgc ttgctcttgc tactctgctc tgggtggctca 540
ttgtgaaaag cccgcaccaa ccatgccagt ggagccaga cgaggacaca gcctggctct 600
gggtcccagc aggaaaggca atcccagaaa ggaggggtca gggactggag tcctgtgggt 660
gctttttaag caaagattat caccagcag gctaaactta gcaaccggct tttagctaga 720
aggcagggg gctgggtgca gggtatgctg ggccagcaaa gaggcccggg atccccctcc 780
catgcacctg ctgatgggcc aaggccacc caccocacc ccttccttac aagtgttcag 840
caccctcca tcccacactc acaaacctgg ccctctgcc toctaccaga agaatggatc 900
ccctgtggga gggggcagg gacctgttcc caccgtgtgc ccaagacctc ttttcccact 960
ttttccctct tcttgactca ccctgccctc aatatcccc ggcgagcca gtgaaagga 1020
gtccctggct cctggctcgc ctgcacgtcc cagggcgggg agggacttcc gccctcacgt 1080
cccgtcttc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
acagcccaca ggtttccag agtgcctttg tgggagg 1177

<210> 17

<211> 18106

<212> **ADN**

<213> Homo sapiens

5

<400> 17

gtgagcagca caggacactt caatgcctgt tgggttctgg gctggctaag acatctgccg 60
gccctgggca gcatacggct cttgcagtca ccttccctc ctcttatcc ccagctgggt 120
tgcaacaaaa ttgccagagt gacctaaagc cagatctttg tctccagttc ttttttatt 180
actcaaaaa cacaacaaaa gcagcatctc atccaattct tgtttgtttg tttttaatag 240
tttttatttt tcagagcagt tttaggttca aagcaaaatt gagcagaaag tacagggagt 300
tcccttctac cccttgcccc tacacatcac agccttcccc accttcaaca tcctgcacca 360
gggtggcaca tttgttacag ctgaacctac acttacacat catctcctaa agtcatgggt 420

10

ES 2 679 291 T3

taccttggag ttcactgcac gtaatgacat gtaccacca ttgcagtatc atacagaaga 480
gtttcaactgc cttacaaaac ccctgcactc cacctattta tccctctctc cccacaacc 540
ctgatctttt tactgttgcc atcactttgt cttttccaga atgtatcatt ggaatgatcc 600
ggtatggagc cttctcacct tggcttctta gtaatgtgog ttttaaggcct ccatgtcttc 660
catggccttg tttcttttta atcagaagta actgttttca ggctgtctct gaatctcctt 720
ttctccctcc aggctataat agatgaattt gagcagaagc ttcgggcctg tcataccaga 780
ggtttggatg gaatcaagga gcttgagatt ggccaagcag gtagccagag agcgcctca 840
gccaagaaac catccactgg tacgtaaggc agcctgtgog ggcgagacca gactgggccc 900
tcccctcctg cagtgatttg tttcttctc ttttttaaat cacgttttcc tgccttttct 960
aggttctagg taccagctc tggcttctac agcctcagac aatgactttg tcacaccaga 1020
gccccgccgt actaccctc ggcatccaaa caccagcag cgagcttcca aaaagaacc 1080
caaagtgtgc ttctcaagtg atgagtccag tgaggaaggt atgatgctcc cgcctgttcc 1140
cggccgagaa ggcacacagc tagggtgcag agggctggt tccataggac ctgctgcggg 1200
ggcctgagtg tagatgctc gcccactgc cgcagaaggg cctctcctgt acagcttga 1260
ttttatttct tctgtgcggt gtgggattgt ctcaactgtt ctctgatatc tattttttca 1320
ccatctttgt gactcagctt tttcttattc ctttaattct ttgcatagat ctttcagcag 1380
agatgacaga agacgagaca cccaagaaaa caactcccat tctcagagca tcggctcgca 1440
ggcacagatc ctaggaagtc tgttctgtc ctccctgtc agggatcct gtaggtgac 1500
ctggaattcg aattctgtt ccctgtaaa atattgtct gtctctttt tttaaaaaa 1560
aaaaaggccg ggcactgtgg ctcaogcctg taatccagc actttgcat accaaggcgg 1620
gtggataacc tgaggtaggg agttogagac cagcctgacc aacatggaga aacccatct 1680
ctactaaaaa taaaaaatta gccggcgta ttggcgtgog cctgtaatcc cagctactca 1740
agaggctgag gcaggagaat cgctgaacc cagaggcggg ggtgtagtg agccgaaatc 1800
acaccattgc actccagctt gggcaacaat agcgaacctc catctcaaat taaaaaaaa 1860
atgctacac gctctttaa atgcaaggct ttctcttaa ttagcctaac tgaactgcgt 1920
tgagctgctt caactttgga atatatggtt gccaatctcc ttgttttcta atgaataaat 1980
gtttttatat acttttagac attttttct aagcttgtct ttgtttctc tttcacatta 2040
gcccagtttc atgcagcaga gagagggtta tcagtgcaga gagagatgag tgagcccaga 2100
gtcctagggc ctgtcccggg atggcagatg agcttctgc cccgtcactg ccacctttcc 2160
cctctcaacc tctggaccct gcacagtgac cagacagcct ctctggggag aattatgcag 2220
tgcctaggct ccagatcagt gcttctgaac cgggggcaat tttgtctgcc agaggacatc 2280

ES 2 679 291 T3

tgacaacacc tggggcctgt tttgtgtca tagcctatag ggaagaatg ctaccagcat 2340
 ttgtgggaag aggccaggga tgtggctcaa catcctgcag tgcacaggat ggcccctcaa 2400
 caaagaatca cacggcccac aatgtcaata gcgtcacagt tgagaaaacc tgctctagac 2460
 caagggttgc tttctgccgt gtgcctcacc ccacccccac tcgtgttccc taatcccac 2520
 tccaaagggtt ggcagcagac cggcccaggc tcgtggaagt tcagatcatg atcccctcca 2580
 gctctgcagg agacaagacc tgtctcccag cattcctcat tgttcccggg tctgcagagg 2640
 gcgtgageta tgetgcaggc gggctgcccc ctgaagcctg cgcacccctc tccagctcct 2700
 caagtcttct ctgctgagtc acctcgaac cggaggctgt gagctggctg tcgtgaccac 2760
 actggtgcct ctgctgtcat gacaacagca cactacgtca gtagtgctcc ctgggcactg 2820
 agctccctct ttgcggggag aagacagtaa tgaaaaatga caagcatgag gcagagggga 2880
 agatcacgct tgggtggtgc aggagcatgg aggtgctctt aatgctctca atgagaaagg 2940
 gttaacggtc ctggttgagc gaatagctga gtcagaggtg gggcttctc cactccccca 3000
 cccccccct ttcaccatta gggaccttct tgccttgctc ttgctactct gctctgggtg 3060
 gtcattgtga aaagcccgca ccaacctgac cagtggcagc cagacgagga cacagcctgg 3120
 ctctgggtcc cagcaggaaa ggcaatccca gaaaggcagg gtcagggact ggagtccctgt 3180
 ggggtgctttt taagcaaaga ttatcaccag gcaggctaaa cttagcaacc ggcttttagc 3240
 tagaagggca gggggtctgt gtcaggttat gctgggccag caaagaggcc cgggatcccc 3300
 ctcccatgca cctgctgatg ggccaaggcc accccacccc acccccttc ttacaagtgt 3360
 tcagcaccct cccatcccac actcacaaac ctggccctct gcctcctac cagaagaatg 3420
 gatcccctgt gggagggggc aggggacctg tcccaccogt gtgcccaaga cctcttttcc 3480
 cactttttcc ctcttcttga ctcacctgc cctcaatata ccccggcgca gccagtgaaa 3540
 gggagtccct ggctcctggc tcgcctgcac gtcccagggc ggggagggac ttccgccctc 3600
 acgtcccgtct ctccgcccc aacctggatgg aatgaaaggc aactgtctc tctccctagg 3660
 cagcacagcc cacaggtttc caggagtgc tttgtgggag gcctctgggc ccccaccagc 3720
 catcctgtcc tccgctggg gccccagccc ggagagagcc gctggtgcac acagggccgg 3780
 gattgtctgc cctaattatc aggtccaggc tacaggctg caggacatg tgaccttccg 3840
 tgcagaaaacc tccccctccc cctcaagccc cctcccagc ctcttctc tccaggcccc 3900
 cagtgccag tgcccagtgc ccagcccagg cctcggtccc agagatgcca ggagccagga 3960
 gatggggagg ggaagtggg ggctgggaag gaaccacggg ccccgcctcg aggccatgg 4020
 gccctccta ggcctttgcc tgagcagtc ggtgtcacta ccgcagagcc tcgaggagaa 4080
 gttccccaac tttccgcct ctccagcttt gaaagaaaga aaggggaggg ggcaggccgc 4140
 gtgcagccgc gagcgggtgt gggctccggc tccaattccc catctcagtc gttcccaaag 4200

ES 2 679 291 T3

tcctcctggt tcatccaagc gtgtaagggt ccccgtcctt gactccctag tgtcctgctg 4260
 cccacagtcc agtcctggga acccagcaccg atcacctccc atcggggcaa tctcagtccc 4320
 ttccccocta cgtcggggcc cacacgctcg gtgctgccc agttgaacca ggcggctgcg 4380
 gaaaaaaaa agcggggaga aagtagggcc cggctactag cggttttacg ggcgcacgta 4440
 gctcaggcct caagacctg ggctgggact ggctgagcct ggcgggaggc ggggtccgag 4500
 tcaccgctg ccgccgcgcc cccggtttct ataaattgag cccgcagcct cccgcttcgc 4560
 tctctgctcc tctgttcga cagtacgccg catcttcttt tgcgtcgcca ggtgaagacg 4620
 ggcggagaga aaccgggag gctagggacg gcctgaaggc ggcaggggcg ggcgcaggcc 4680
 ggatgtgttc gcgccgctgc ggggtgggccc cgggcggcct ccgcattgca ggggcgggcg 4740
 gagcagctga tgcggcgcgg gctgggcatg gaggcctggt gggggagggg aggggagggc 4800
 tgtgtgtcgg ccggggccac taggcgctca ctgttctctc cctccgcgca gccgagccac 4860
 atcgtcaga caccatgggg aagtggaag tggagtcga cgggtgagtt cgcgggtggc 4920
 tggggggccc tgggctgca ccgccccga accgcgtcta cgagccttc gggctccggg 4980
 tctttgcagt cgtatggggg cagggtagct gttccccga aggagagctc aaggtcagcg 5040
 ctcggaactg gcggagcccc gcaccaggc tgtggcggcc tgtgcagctc cgccttgcg 5100
 gcgccatctg ccgggagcct ccttccccta gtccccagaa acaggaggtc cctactcccg 5160
 cccgagatcc cgaccggac ccctaggtgg gggacgcttt ctttccttc gcgctctgcg 5220
 gggtcacgtg tcgcagagga gccctcccc cacggcctcc ggcaccgag gccccgggat 5280
 gctagtgcgc agcgggtgca tcctgtccg gatgctgcgc ctgcggtaga gcggcccca 5340
 tgttgcaacc gggaagaaa tgaatgggca gccgttagga aagcctgccg gtgactaacc 5400
 ctgcctcct gcctcagatg gtggagtcgc gtgtggcggg gaagtcaagt ggagcagggc 5460
 tagctggccc gatctctcct ccgggtgatg cttttcctag attattctct ggtaaatcaa 5520
 agaagtgggt ttatggaggt cctcttgtgt cccctcccg cagaggtgtg gtggctgtgg 5580
 catggtgcca agccgggaga agctgagtca tgggtagttg gaaaaggaca tttccaccgc 5640
 aaaatggccc ctctgggtgt gggcccttc tgcagcggc gctcacctca cggccccgcc 5700
 cttcccctgc cagcctagcg ttgaccgac cccaaaggcc aggctgtaa tgtcaccggg 5760
 aggattgggt gtctgggccc ctcggggaac ctgccttct cccattccg tcttccggaa 5820
 accagatctc ccaccgacc ctggtctgag gttaaatata gctgctgacc tttctgtagc 5880
 tgggggectg ggctggggct ctctccatc ccttctccc acacacatgc acttacctgt 5940
 gctcccactc ctgatttctg gaaaagagct aggaaggaca ggcaacttg caaatcaaag 6000
 ccctgggact aggggggtaa aatacagctt cccctcttc caccggccc agtctctgtc 6060

ES 2 679 291 T3

cctttttagtag gagggactta gagaaggggt gggcttgccc tgtccagtta atttctgacc 6120
 ttactcctg ccctttgagt ttgatgatgc tgagtgtaca agcgtttct ccctaaagg 6180
 tgcagctgag ctaggcagca gcaagcattc ctgggggtggc atagtgggggt ggtgaatacc 6240
 atgtacaaag cttgtgccc gactgtgggt ggcagtgcc cacatggccg cttctcctg 6300
 aagggtctcg tatgactggg ggtgttgggc agccctggag ccttcagttg cagccatgcc 6360
 ttaagccagg ccagcctggc agggaagctc aaggagata aaattcaacc tcttgggccc 6420
 tcctgggggt aaggagatgc tgcattgcc ctcttaatgg ggaggtggcc tagggctgct 6480
 cacatattct ggaggagcct cccctcctca tgccttcttg cctcttgtct cttagatttg 6540
 gtcgtattgg gcgctgtg accagggctg cttttaactc tggtaaagt gatattgtg 6600
 ccatcaatga ccccttcatt gacctcaact acatggtgag tgctacatgg tgagcccaa 6660
 agctggtgtg ggaggagcca cctggctgat gggcagccc tcataccct cacgtattcc 6720
 cccaggttta catgttccaa tatgattcca cccatggcaa attccatggc accgtcaagg 6780
 ctgagaacgg gaagcttgc atcaatgaa atcccacac catcttcag gaggagtgg 6840
 aagacagaat ggaagaaatg tgctttgggg aggcaactag gatggtgtgg ctcccttggg 6900
 tatatggtaa ccttgtgtcc ctcaatatgg tctgtcccc atctccccc cccccata 6960
 ggcgagatcc ctccaaaatc aagtggggcg atgctggcgc tgagtacgtc gtggagtcca 7020
 ctggcgtcct caccacatg gagaaggctg gggtgagtgc aggagggccc gcgggagggg 7080
 aagctgactc agccctgcaa aggcaggacc cgggttcata actgtctgct tctctgctgt 7140
 aggctcattt gcaggggga gccaaaagg tcatcatctc tgccccctct gctgatgccc 7200
 ccatgttcgt catgggtgtg aacctgaga agtatgaaa cagcctcaag atcatcaggt 7260
 gaggaaggca gggcccgtg agaagcggcc agcctggcac cctatggaca cgctcccctg 7320
 acttgcgccc cgctccctct ttctttgcag caatgcctcc tgcaccacca actgcttagc 7380
 acccctggcc aaggtcatcc atgacaactt tggtatcgtg gaaggactca tggtatgaga 7440
 gctggggaat gggactgagg ctcccacctt tctcatccaa gactggctcc tccctgccgg 7500
 ggctgcgtgc aaccctgggg ttgggggttc tggggactgg ctttccata atttccttc 7560
 aagggtggga gggaggtaga ggggtgatgt ggggagtacg ctgcagggcc tcaactcctt 7620
 tgcagaccac agtccatgcc atcaactgcca cccagaagac tgtggatggc ccctccggga 7680
 aactgtggcg tgatggccgc ggggctctcc agaacatcat cctgcctct actggcgtg 7740
 ccaaggctgt gggcaaggtc atccctgagc tgaacgggaa gctcactggc atggccttc 7800
 gtgtcccac tgccaacgtg tcagtgtgg acctgacctg ccgtctagaa aaacctgcca 7860
 aatgatgatga catcaagaag gtggtgaagc aggcgtcgga gggccccctc aaggcatcc 7920
 tgggtacac tgagcaccag gtggtctcct ctgactcaa cagcgacacc cactcctcca 7980

ES 2 679 291 T3

cctttgacgc tggggctggc attgccctca acgaccactt tgtcaagctc atttcctggt 8040
 atgtggctgg ggccagagac tggctcttaa aaagtgcagg gtctggcgcc ctctggtggc 8100
 tggctcagaa aaagggccct gacaactctt ttcattctct aggtatgaca acgaatttgg 8160
 ctacagcaac aggggtggtg acctcatggc ccacatggcc tccaaggagt aagaccctg 8220
 gaccaccagc cccagcaaga gcacaagagg aagagagaga ccctcactgc tggggagtcc 8280
 ctgccacact cagtcccca ccacactgaa tctcccctcc tcacagttgc catgtagacc 8340
 ccttgaagag gggaggggcc tagggagccg caccttgtca tgtaccatca ataaagtacc 8400
 ctgtgctcaa ccagttactt gtcctgtctt attctagggg ctggggcaga ggggagggaa 8460
 gctgggcttg tgtcaaggtg agacattctt gctggggagg gacctggtat gttctcctca 8520
 gactgagggg agggcctcca aacagccttg cttgcttoga gaaccatttg cttcccgtcc 8580
 agacgtcttg agtgctacag gaagctggca ccactacttc agagaacaag gccttttct 8640
 ctcctcgctc cagtccctagg ctatctgctg ttggcctaac atggaagaag ctattctgtg 8700
 ggcagcccca gggaggctga caggtggagg aagtcagggc tcgcaactgg ctctgacgct 8760
 gactggttag tggagctcag cctggagctg agctgcagcg ggcaattcca gcttgccctc 8820
 cgcagctgtg aggtcttgag cacgtgctct attgctttct gtgccctcgt gtcttatctg 8880
 aggacatcgt ggcagcccc taaggtcttc aagcaggatt catctaggta aaccaagtac 8940
 ctaaaacat gcccaaggcg gtaaggacta tataatgttt aaaaatcggg aaaaatgccc 9000
 acctcgcata gttttgagga agatgaactg agatgtgtca ggtgactta tttccatcat 9060
 cgtccttagg ggaacttggg taggggcaag gcggtgtagct gggacctagg tccagacccc 9120
 tggctctgcc actgaacggc tcagttgctt tgggcagtta ctcccgggcc tcactttgca 9180
 cgtgtgctta cctagtggag acaaaagtac atacctcggg agagcgcgca cgcctgtaac 9240
 cccagcactt tgggaggcca aggtgggtgt atcacctgag gtcaggagt tggagaccagc 9300
 ctggccaaca tggtgaaact ccgtctctac taaaattaca aaaatcagcc aggcctcatg 9360
 gcacatgcct atagtcccag ctacaggcat gctgaagcag gagaatcgct tgcaccccgg 9420
 aggcagaggg tgcagtgagc tgagaccaca ccactgcaact ccagcctagg caacagagta 9480
 tgagactcca tctcaaaaa aaaaaaagta cctacctcag agttcaaact agtgaatatt 9540
 aggaagtgc tggagacagt acaccaagt gcacaataaa tactcgccag tttcattatt 9600
 attaaagaat ccatttgaat gtcagctcaa cacagcctcc tataccgagg cattgtgaac 9660
 cgcactctcc cagcttctcc aggttttcc aagaatcagg gacactgtag cctggtggtc 9720
 tcagtgtatg acagacacgg aggaagcaca tctttagctg atacttaaac agagaccctg 9780
 agcgcacata caccgcgca cacatgcatg gagcttcacc ttctctgtca ttctgagtg 9840

ES 2 679 291 T3

accaggagag caagagctcc cacctccctt caaaacactg tgcccatccc gggcactaag 9900
gcctctttaa agcacggcac ctccacgagg gagggccaca gccacatata ctccacctgg 9960
caggtggaca gcgtgagcac gtggaccata gcagggacaa ggtgccccgg ccagcccaaa 10020
cgccctctgc cgctgacagg gacagaagcc ctctccagct gcgtgtgctg cagaggccat 10080
gcgtagcctc cagctgcatt ctattccact ccagtgctg ggcagttag caccagtggt 10140
gaagacagtg agctggctcc ggacaacagg gatggaggaa aggtcccaca ttcacattcc 10200
tgatacgtgg acaagtgtag gggccgcaat cgctctggca gcattttaa gatggggaag 10260
tagcagacac ccacgcgtga aggcaggaga gcccactg tggtggaat ggcccagaa 10320
tggtagggcc aagcctagct ccagacacc cagagccctg gagaagcaa gactgagga 10380
gaaagcctga gggaggagcg cccagtccc cagggaccgg cctggtgag agctgcagct 10440
gatgttccc tctgtgcagc cccaccctct gcctcgctga gctccctgct gcgagggcct 10500
cgggtgcaag ggggagcag gtctctatct catggagctg tcagatgaga catgcgatc 10560
ggagtctca gcctcgctt gggcgcgcg cgggtcgcta agcgggaccg cagtgaaagc 10620
aggagacttt ctagaaaaa acaccagttg tcaaccttgg ggcaggcagg aatcctgaag 10680
acggacggca ctctctctcc tgctgcctca ccctctggca gccctgaga agtaccgaa 10740
gcgagggcgg ggcccgggga tggcgagga gggcagga ctgaactctc tccaaacca 10800
ccctgacagg gaaatgggccc ccgcctgtgt cttgggaact cagaggctga ggtcaggcat 10860
gatggctcac gcctgtaatt ccagcacttt gggagcaga ggcgggtgga tcacgacgc 10920
aggagttcaa ggcaagcctg gccaacatgg tgaaaccca tctctactaa aaatgcaaaa 10980
attagccagg tgtggtggcg ggcgccttgg aggctgaggc aagataatca cttgaacctg 11040
ggagcgggag gttgtagtga gccaaaaaa aaaaaaaag aaatagctga agtcacagta 11100
ggagagaagc tgctgagcct ccagcaccct gactctaggg ccttggcttt atgtctatct 11160
gcagtathtt tgtgattttt aaaaattcac tttcttgtt cggtgtaact tacacagggt 11220
caaatgcaca aatcatggcc ctgactttag ataaaaatct gccccacaa cccttctgtt 11280
ccttgccagt ttttaactg cctctaacca ggggaaccac cagagctggt gccttggga 11340
ggtttcagcc ctcccgtcat gaatggacat agctcatcca actgccaagg gagagagctg 11400
tgggtctggg ccagccccac caggtaaact ccaaagggca gccccacagc aagatgtgac 11460
ccagtcattg cctgagggtc tctgggctg tgttccaacc tttctcccog ctgtgtcccc 11520
ctggaaggcc ccatgccag gggagggccc taccggtctc cagactgggt gatgagccgg 11580
cggcaggtct ccatctgcac gtccaggccg cgcttcatgc tgcacatctc catgtactcg 11640
tgcaggtgcc ggttcatgtc gttctggcc gtggccaact ccagctgtgg tggggcagcc 11700
agggccatcg gggtaacag gtggcgttca cagcgctct gttgccccog ccaggaggcc 11760

ES 2 679 291 T3

aacacgccaa gagcagtggc tgggccgggg gccaggcag ccattactga aggctcagat 11820
 tttaaaataa accagaccaa ggcagaaggc agcgaataca ccattgtttc ctctgctttt 11880
 cctggaaatc tttgcctatg gaggaatgca aaggtactgc tgcaggctgt gttcctttta 11940
 gactgataag agtggtgagg aaagtcaagt tgggcgggat ggagcagatc cacaggctcc 12000
 cagggccaca caggttccta tggggcccgg ctcaccctc atcccctgta ctgcgctggg 12060
 cctgggctgt cccctcactg cactgtgagc tcccgtaagg gttagcgaca ggttttcctc 12120
 atcctggcag acccagacct cgaagggaat taactctctc tgtgaccctg ggcatacaag 12180
 cctggctaac tgctgaaagg ggccaatcc tcttaccatt ctctccttat cactttat 12240
 tctcacctaa aagccatttt cagacactaa aaggaagact cccatattgg aggtgccatt 12300
 gattctgaga cctgtcctga tttcagaaat gtcacacgtg gaaaacacgg ggctcagaat 12360
 caaagaaatc cggctctcct ccccgggcc attctgcagc ccacactctg aggcaggcgg 12420
 agaggcaggg cccaggcagc acaaactcag tgccaggac agtcgactgc ctctctttgt 12480
 tttgggtaca aagggcaggg gccgcaaggg gaagcaaaat gaaacttctt ggagatgacg 12540
 ttcgtgactg cctgagtga gagaacgggg aggataccag ctgccataat gtggattgga 12600
 cagagagggg gggaaacttag agaaagaagg aaacaggtta cagggccaga cagcagagat 12660
 ggtcaggaca gtgacatttg gggtgagatg ggggctggct gggtgagaag ccctagtggg 12720
 gacagcttcc ctggcaggct tggatccaca gctggggccc cacagtagtg ggaggagcag 12780
 gacacggtga ctgttcaacc caggggaaa atacagtgcc accagcctt gttcagatca 12840
 gctgcctaca cccaatttc tccaccttg aagaaccgga gtgcaggag ctgggttctg 12900
 gaaccatcac tcaggagac tgggaagaag gaaatgaacc aaatatattc tactgaaata 12960
 caaggctagt actgccacca gtagtctctg catatgtcta tgtcccact taattcacia 13020
 atttacagca ggaacaaggc caggaggtga acgtgcatgt caggagatgt ccttgacaag 13080
 ccaccttga aactgcaggc tcgctgtgtg aacagacttg ggggtgggac gggaggagac 13140
 agagagaggt gaggagcggg ggggaagaac tgcctcccgc ccaccactct gggccttggc 13200
 tttaccttgc tccatcttag tgacctgag agatcaaagg tcacaaagat aaaggcacct 13260
 tgggtactcc tgaggtcagc aagtcctcct aacggttccc ctaatgcaga agtccccaaa 13320
 ccccaggcca cggaccagta cctgtctgtg gcctgttagg aaccaggcca aatagcagga 13380
 ggtgagcagt gggcaagtga gtgaagcttc atctgtat 13440
 gcattacagc ctgagctcca cctcctgtca gatccgaggc ggcattagat tctcacagga 13500
 tcctgaacct taccttacgt gaactgtgca tgcgagggat ctaggctgtg tgctccttat 13560
 gagaatctaa tgctgatga tctgtcacta tctccatca ccccagatg ggaccgtcta 13620

ES 2 679 291 T3

gttgcggaaa acaagttcat ggctcccact gattctacat tatggtgaga tgtataatta 13680
 tttcattata tattataatg taataataat agaaataaag tgcacaggcc gggtagcggg 13740
 gctcatgcct gtaatcccag cactttggga ggccgagggtg gtggatcacc tgaggtcggg 13800
 agttcgagac cagcttgacc aacatggaga aaccccatct ctactaaaaa taaaaaatta 13860
 gccaggcgtg gtggcacatg cccgtaatcc cagctacttg ggaggctgag gcaggagaat 13920
 tgcttgagcc cgggaggcgg aggctgcagt gagccaagat tgcgccattg cactccagcc 13980
 tgggcaacaa gagcaaaact ctgtctctaa aaataataat aataattttt aaaaatagta 14040
 ataaagtaca caataaatgt aatgtttgaa tcatctcaa atcatttccc ccgggtcagt 14100
 ggaaaaactg tcttccacaa accagtccct ggtacccaaa aggttgggga ctgctgccta 14160
 aagtcattgc actgggagag aggaggagtc agatgatttg aaaatacttt tttttttttt 14220
 tttctgaga tggaatcttg ctctgcacc caggctgtgg tgcagtggca caatctcagc 14280
 tcaactgcagc ctccacctcc cagattccag caattctcct gcctcagcct cccaagtagc 14340
 tgggattaca ggtgccacc acaatgcctg gctacttttt gtatttttag tagagacggg 14400
 gtttcacatg ttggtggcca ggctgtctc acactcctga cctcaagtga tctgcccgcg 14460
 ttggcctccc aaagtgctgg gattacaggc gagagccacc acgccagacc aaaaatagct 14520
 taaaaggcat gattccattt aagatattat tttcgttctt tcagatcatg ttttactctg 14580
 ggcacttgag gtccagggaa gactaggag agatggcagc ccgagatctc ttctcacgaa 14640
 gatggccgca caaaggcccc tagttggttg aaggcacagt gaaaacacaa accttaaact 14700
 caccatccaa aagagggaga caagaatcag gtcaggcagg aggaggaac agtgagccgc 14760
 ggtagacggg aggggtggga gtactggttt cccttgccat gtcttttagct agcagcacia 14820
 atccagggaa cctcatctac ctggccaaga aaaccagggt tgatgcagga gcaggacagt 14880
 ttgatgtgtg ttgggagaat gagtatatag gaaaataatt tctcgacttc cttctcaaac 14940
 acctccaatg tgacagcaat aagaataaag tcaaggccgg gcgcggtggc tcacgcctgt 15000
 aatcccagca ctttgaagc ccgaggcggg tggatcacga ggtcaggaga tcgagacat 15060
 cctggctaac atggtgaaac cccgtctcta ctaaaaaata caaaaaatta gccgggcgta 15120
 gtggcaggcg cctgtagtcc cagctactct ggaggctgag gcaggagaat ggcgtgaacc 15180
 tgggagggtg agcttgagc gagctgagat cgcgccactg cactccagcc tgggcgacac 15240
 agcgagacta cgtctcaaaa aaataaata ataaataaat aatgaaaag aataaagtca 15300
 ctctctcctt atccactggt ttcactttcc acggcttcag gtaccaagg tgaaccttga 15360
 tctgaaaata ttaaatgaaa tattccagaa ataacaatt cctaagtttt aaattgcagc 15420
 ctgttctgag tggcctgata aaatctcact ctctctgccg ggccggtggc ctcacgccta 15480
 taatcccagc actttgggag gccgaggtgg gcggatcatg aggtcaggag atcgagatca 15540

ES 2 679 291 T3

tcctggctaa cacggtgaaa ccccatctct actaaaaaat aaaaagaaa ttagccgggc 15600
 gtgatggcgg gcgcctgtag tcccagctac tcaggaggct gaggcaggag aatggcgtga 15660
 acccaggagg tggagcttgc agtgagccga gttcaagcca ctgcactcca gcctgggtga 15720
 cacagcgaga ctccgtctcg gaaaaaaaa aaaaaacctc actctatcca gaccaggaca 15780
 tgaacctctc ctctgtccca aggatccaca ctgtctatac tgcttgccta ttagttactt 15840
 aggagccatc tcagtgtatga gattgactgt cgtggtatgg cagtgccctg gttcaaggaa 15900
 ccctatcttt tacttcataa tggcccaaaa gcacaagagt agtgatgctg gcaatttggg 15960
 tatgtgaaag agaagccaaa atatgctccc ttttaagtga aaggtaaaag ttttcaactt 16020
 aataagaaaa agaatccgat gctgagggtg gtaagctcta cagtaagaat gaatcttcta 16080
 tccgtgaaat tgtaaagaag gaaaaagaaa tttgtgccag ttttgctgtc acacctcgaa 16140
 ctgcaaaagg taccgccaca gcgcacgcta agtgcttagt taagatagaa aaggcattga 16200
 atgtgtgggt gggagacatg aacagaaaca cgttccagtt gacagcaact ggaccaagc 16260
 tatcaggagg tcagtgtctat ctttggtttc aggcattccac tgtgggtctt ggaacacatg 16320
 ccctgaagat aaggggggac tacgtatgaa tctcccatgt gggcctgaag agaggctgag 16380
 acgcagagaa ggctccatgc ctctccaaa gtcagagcgt cgggaaagga ccacctcacc 16440
 aggccaggag agctcacggg ggtcccgctc atccctcctc ccacagcaac agcaacctcc 16500
 attctgtggt tccacaaatc cttcttctt tgtttgatt tttgagacgg agtcttgctc 16560
 tgtcggccag gctggagtgc agtggtgcca tctcggctca ctgcaagctc cgcctcccg 16620
 gttcacgcca ttctcctgcc tcagcctccc gagtagctgg gactacaggt gcctgccacc 16680
 acggccagct aatctttttt ttttgatatt ttcgtagaga cggggtttca ccgtgttagc 16740
 caggatggtc tcgatctcct gacctogtga tcttcccgcc tcggcctccc aaagtctggt 16800
 gattacaggc gtgagccacc gcgcccggcc tccacaaatc cttcttgtag ctctaaagc 16860
 ctagagtttg agctaaagca ccagctccac ttgcctctag gtacctctag aaggtaactc 16920
 aaggcttggg aaaaagcatg gagatttctc tgttatctca aaggctgagt ggaagaaaca 16980
 ggaaagtccc agtccatctc tgagtgccat ggcactgata gtgttaataa taataaaaca 17040
 gtccccacta tgctacactc gctatgagcc agacactgcc ctaagctcta tatatgcatt 17100
 atttcatctg atcttctaaa caacctata ggtgaggtat tattaaattc ctcattttac 17160
 aaataaaaag gcccggcaca gtggctcatg cctgtaatcc caacactttg tggggccgag 17220
 ttaggaggac tgcttgaatc cagaagtttg agggcagcct gaaaaacata gtgacacct 17280
 gttctctgta aaaaaaatca aaaaattagc cgggcgtggt agcacacgcc tgtggtccca 17340
 gctacttggg aggctgaggt aggaggattg cttggatcca gaagatcacg gctgcagtgg 17400

ES 2 679 291 T3

gcctggagtg acagagcaag accctgtctc aaaaaaagaa agaaagaaag aaagaaagaa 17460
aaagaaacag gtgtagggct ggacacatgg ctcatgtctc taattctagt actttgggag 17520
gccgaggcag gaggatcact taagcccaag agtttaagac ctgcctggga aacacagggg 17580
gacaccatct ctacaaaaaa atacaaataa aaattagctg ggtatggtgg cacacacctg 17640
tgggcccagc tactcagaag gctggggtgg gaggatcact ccagcctagg caacagagtg 17700
aaacagagtg agaccttgtc tcaaaaaaga aaggacagaa agacaaggga gagaggggag 17760
gagggataga gggggagggg aagagagaga gagagagaaa ggaaggaagg aaggaaggaa 17820
agaaaggaag gaaagggaaa ggaagggagg ggctcagagg gaacaagtaa cttattacaa 17880
gttcagggtg gttgtaagta acagagctga gatttgaacc gatatctgtc tgtttccaga 17940
gcctgaggtc ttcagagccc accaggagga ctatagctga gagatcactc ctccctccaa 18000
gaccttggtc tctagggctc tgcagcaggg atgaggccac tgcgcctgca gccccactca 18060
aaacctctg ggacaccagc ccctggctt ccgctcctgc ccttac 18106

<210> 18
<211> 16312
<212> ADN
<213> Mus sp.

5

<400> 18
gtgagtggtc ctgtaggtt ccggtgcagt aaacttcctc tgcctcccta agtgctctga 60
gactcctgtc tgaactttct gccttgctag cctcactaca gctatggaac cagagtgact 120
aagaccagat cttaattcct ggtccctttt ttaaaggcca ttcatttctt tttttgtaa 180
catgcattgg tattttgcct gtgtatatgt ttgcatgaag gtgccaagtc cccttggaat 240
tggagttaga ctggtgtggg ccctgagaac tgaacacagg tcttttgtaa gagcagccag 300
tgctcctaac cactatgcca ccttcccagc cccagtgaag gccttgagat agggctctgc 360
tatgtagacc cagctgcctc agactaacag agattcccct gcctctgtct cctggtagta 420
gaggatgccc gcacatgttt gcaactggact gcttctttag agtagtttca tgtttaaagc 480
aaaagtccag aagtagagct cccatagct ctgtcccaca cccgcatggc ctccacccca 540
tgaccttctt gtccctggat acgtggctct ggatgaagga acgtgattga cagatcagca 600
tccaaagtcc actcactgga cacagtggca catacttacc acttgtgtca tactgggcat 660
tttcaggact ttgaggtcac cttaccccc acccctgtgg tggcatcagt tagttggaat 720
cacagtttca gattggtttc tttagttttc tctcttttca tggctatttc tttttaagca 780
gggagataat gtttccagcc ctgtctttaa ctttctgtcc tccaggctat aatagatgaa 840
tttgagcaga agcttcgtgc ctgtcacacc agaggcatgg atggaataga ggagtttgaa 900
actggccagg gaggcagcca gcgagccctg tctgccaaga aacctccgc tggatgaga 960

10

ES 2 679 291 T3

gggatgcaag cccacatggg gctcctatga gtggccccca cccttttttt gttcctttgt 1020
 ggtttgtttt tggttttgaa atggccccct tgcctttcct ttctagtctc aaggctacag 1080
 cctctgactt ctgtagactc agacaatgac tttgtcacac ctaagccgcg acgtacccaa 1140
 cctggtcgtc cacaaactca gcagcgcaag aagtcccaga ggaaagccaa agttgtcttc 1200
 ttgagtgacg agtccagtga ggacggtatg atacaccctc ccaacagatc agctgcctgt 1260
 gtggggctct ccagtggggc ttgcctagct gtggattcat tcttcctggc catgtcttcc 1320
 tcactatgca gctcggctga ctttcacagt gtgccgtgtg cgtgtgtgtt tgtgtgtgtg 1380
 tttgtgtgtg tgtgtgcaca cgtgtgcgcg tgcgcacgcg cgcagtgtgt tactttgaga 1440
 tccattttgc cccatcttta tgactcagct tttgttccta taattctttg aatagaactt 1500
 tcagcagaga tgacagaaga agagacacc c aagagaacca ccccatccg cagagcatct 1560
 gggcgaagac acaggtccta ggagctgcct ttgcctactc ccgtccctgc cgtgcagggt 1620
 atcttactgg atgatcttgg attcagtgcc cccccccct tgtagaagtg cttggtgtct 1680
 gtgggtttta catggaaatg gcctttgcat tgaggtcttt ctcttagagc cctgctgcc 1740
 cttcccttt ctccgagcct gtgccaattg ctgttttcta atgaataaat gtttttatag 1800
 acttttaagc atcctgtcct ggctggtctg tcccatcctt agccctgagc tgtgtctgga 1860
 tcttcatggc ccctgtggtg cctcatcctt gcacaactga aactctgctg tgcagaatta 1920
 tgcggtttct aggttcacgt gaacagttct cagcagtctg gggactttg caccocagaa 1980
 aacattcggg aatacctaga gacttatttc tgtttgcctt gtggggcagg gggctgcagt 2040
 gaaagcctta agatgaaggc cacctcatg cacattgtgc ccataaacac agcatgtcag 2100
 gagtacacaa cttgagaaat gctgctctag acctaagggt tgtttcctct gtgocccca 2160
 ccccatcccc cgaagactgg caaatgactc agcctggcta atggcagaga ttatgaaccc 2220
 cactcagttc tataggtgac aaggccatc gcccagagc cagtccttca ctacactaag 2280
 gggctctgag gaggtccagc tgtaagccat gctgtgtgcc caccatcccc aagcccttaa 2340
 tctgctgagt cacttggagc aggaggcaca actgctgttt ctgctgccat gacaacagca 2400
 cagtgctgct gaagtgtcct ctacctctg gtgaggggac agtaatgata gatgacaggc 2460
 gtgaggcaaa agggaaagggt cacctctgaa acactagggt agggctgggc tatggaaggg 2520
 ctctttgggt aggaaaaggg ttaacgctgt gcatagagcc tcggtaggaa ggcagaaaag 2580
 ggactctgaa tctgccatgc ctctcctccc caccactgcg ctggggccac tcccaccctt 2640
 cccaccctgt tcatctggcc tgctccccgc tcatgggaag cctgggcacg ggcacacag 2700
 ctgtcagtct aacacagcct tggctctggg gcccacagga ggtagggcaa tctggaaggg 2760
 caaggagcca agactagatt tgggggtcag cccagcgtgc tcctgcctt ttaagcaaag 2820
 gttatcacca ggcagctaa acttagcaat aggctcttga gctagatgag cagggggctg 2880

ES 2 679 291 T3

gtttcatggt gcaactggcct agcaaagagg cgccagagtc ccctgcct gcacctgcta 2940
 cagtgtccc aagccctcc acccttcca gcctccctac ccttggagt gaagaatccc 3000
 ggtctcgaaa agacaaaata aaaccaaaga atgcctttc tccctccac ttgtggcaag 3060
 aggctagggg cttccctgtc ctggctcagg ggtataatat ttcctctcct gtgttctccc 3120
 ctcaactgac taccctgtg tccacgaggg cagccaaaaa aggagagtcc ctggctccag 3180
 ggccagttag cacgtccctg gatggggagg gacttccgcc ctcaactccc aactctccac 3240
 cctggggcta cagtgggtga aagggcagt gtctcctagc ctggggggg cagctctcag 3300
 gttccgagga gggatactat aggaggcctc tgagcctcct ccaattcaac ccttaagagg 3360
 gatgtgccc ttaccocggg gtcccagctt aggttcatca ggtaaactca ggagagtgtt 3420
 tgtaagtctc aattatcagg tttccaggct gcagggcata ctgacctatg gcgtagcaat 3480
 ctcttttca agcgtctcc tggccacccc cttgccctca cagcgcggtc cacaagccag 3540
 gcctctggcc aaagacagaa gccaggaggg gggggggggc agataggaa atggggctcc 3600
 tgagggtccc cgcgggagga aaagtcccc caccctggca tttcttcca ctctttttt 3660
 tttcggggg gggggttct gtgtcaactc cgaagaaca cgaggagaag atcctcaact 3720
 tttccgcagc cttttcaata atggggagag gttcgtgat gcagtggcag ggagaccac 3780
 acttctocat tcccctgtt ctcccattt actcgggaag cagcattcag gtctctgggt 3840
 cctggatgtc cttggtgac actccaagga ctctctgtcc ttaagttcat agtctgtatt 3900
 ccctgagtcc taccctggga accatcaccc ggtcacctcc tgagcggggc aatctcagct 3960
 cccctcccc tatcagttcg gagccacac gcttgggtcg tgcacattc aaaaatgagg 4020
 cgggtccaaa gagaggagg aggggaaatg agagaggccc agctactcgc ggctttacgg 4080
 gtgcacgtag ctcaaggctc tgcgccttg agctaggact ggataagcag ggcgggaggc 4140
 ggggcgcgcg tcatcagctc cccccacca tccgggttcc tataaatacg gactgcagcc 4200
 ctccctggtg ctctctgtc ctccctgtc cagagacggc cgcactctt tgtgcagtgc 4260
 caggtgaaaa tcgaggagtg ggccgcagga ggccggggac agtcggaaac tgggaagggg 4320
 agtgggcaact gtacgggtct agggatgctg gtgcgaagtg tgcaagccg acccaggctc 4380
 cgcattgcag ggggggatg atggaggacg tgatggggcg cacggcggga atggaggcgg 4440
 ggtgggggag gggactgcct ggtgtcctc gggccacgct aatctcattt tcttctctg 4500
 cagcctcgtc ccgtagacaa aatggtgaag gtcgggtgta acgggtgagt tccaggcgg 4560
 ggccctgctc cgttgcctac gcaggctctg ctgacccggg gctctgcag tactgtgggg 4620
 aggtggatga ggtggccgaa gcgccaagg agacctcaag gtcagcgtc ggacctggcg 4680
 atggctcgca cttgcggccc ccagctgctg cacctctggt aactccgct ttgcggggat 4740

ES 2 679 291 T3

gagcggcccc gagtcttaag tattaggaac aacccccacgc gcccggtcag acccatcccc 4800
taatccccag tccggggcctt ctttctttac tttcgcgccc tgaggagtca cgtgccagga 4860
gggaagcccc tccccatct ccccttcct cggggtcggg ggcatggcgt tgtgaggtgc 4920
atacctttgc gcatcatctc ccagccttg gcttccttta ggtaactgg ccgccccat 4980
gttgcaaacg ggaagaaaat gaatgaaccg ccgttatgaa atcttgctta gccttcctt 5040
cttcctagct tgtgactaac ctcatcctc tcggctgggt ggagtgcct ttatcctgta 5100
ggccagggtga tgcaaggcct ccggtcctc gagagagttc tacctcaaaa tctgtctcac 5160
cttattagcc ttaaaagccc ttgagcctta ttgtcctcgg gcataatgog tattctagat 5220
tattctctga aaatcaaagc ggacttacag aggtccgctt gacctccaa cccagaggt 5280
agttatggcg tagtgacagag ccgtgggatg gggagctgag tcatgggtgt tctgaaaaga 5340
aattttccac cacaaaatgg ctccctgtagt agcagcccct tccatcccct gcacttccca 5400
tcacagcctc gcactgaccc aggcctata ggcaggatg taaaggtcat taagaggatt 5460
gggtgtccct gcgcctcaga atcctgccct tctccccgtt ccatcctcca gaaaccagat 5520
ctctccactc cgccctgatc tgaggttaaa tttagccgtg tgacctttct ggatctgggg 5580
tctgagcggg ctctccaccc tgctccccct acacacatct gttgctccgg ctctcatttt 5640
tgcccagaaa gaacagggtgt ttcgcgaacg agccctggga ttagggttg aaacccccca 5700
catgttttct cagtctttcc ccttagttcg agggacttgg aggacacagg tgggcccgcc 5760
ctgtgctgct cacgctgacc tttagccttg ccctttgagc ttgctgatga atgagttcac 5820
aggtctgccc tgtccagggg gtgtagcctg aagtccagcc atgctggaac aaacttccca 5880
gggcatgagt gatgggggtg atgtgccaag cttgtaccca gactggggca aactgccact 5940
tcttaagaga cttagaatga cttggaggag gtttgctggg caagcaatca cctcttgac 6000
aggaaagaaa cctccacttt ataaccgtgc tataaaagcc ctgccaggcc tcggctgctc 6060
aaagaatata aaattagatc tctttggact ttctaggggt ggaacagtct atatttgggt 6120
tgtacatcca agcattcaac tagctttatt aaagggaatt ctgaacaaaa catgaacttc 6180
ctgatgcata catttctaata gtactgtgtc tgcataaagg cttgtacctt tgttgggta 6240
cgtgcatagc tgatggctgc aggttctcca cacctatggt gcaacagtat tccactctga 6300
agaacatgag atagcctggg gctcactaca gacctatgag gatttctgat ctgagctccc 6360
ctgtttcttg tctttcagat ttggccgtat tgggcgcctg gtcaccaggg ctgccatttg 6420
cagtggcaaa gtggagattg ttgccatcaa cgacccttc attgacctca actacatggt 6480
ctacatgttc cagtatgact ccactcacgg caaattcaac ggcacagtca aggccgagaa 6540
tggaagcctt gtcatcaacg ggaagcccat caccatcttc caggagcgag accccactaa 6600
catcaaatgg ggtgaggccg gtgctgagta tgtcgtggag tctactggtg tcttcaccac 6660

ES 2 679 291 T3

catggagaag gccggggtaa gtggccggaa agctgaaggt gacgggcacc cttgatatgg 6720
 tgcaacctga aaaccaagaa ctgagtctga aatcaacttc tttcccttaa acaggcccac 6780
 ttgaagggtg gagccaaaag ggtcatcatc tccgccctt ctgccgatgc ccccatgttt 6840
 gtgatgggtg tgaaccacga gaaatatgac aactcactca agattgtcag caatgcatcc 6900
 tgcaccacca actgcttagc ccccctggcc aaggatcatcc atgacaactt tggcattgtg 6960
 gaagggctca tggatgtag gcagtgggga gacagctcat gcatttctta tcttaccctg 7020
 ccatgagtgg acccttcttt gtaggtgtcc cttttgggta gaggggtgcc gtgcaggacc 7080
 tcactcattg ccccctgttt ttctagacca cagtccatgc catcactgcc acccagaaga 7140
 ctgtggatgg cccctctgga aagctgtggc gtgatggccg tggggctgcc cagaacatca 7200
 tccctgcatc cactggtgct gccaaaggctg tgggcaaggc catcccagag ctgaacggga 7260
 agctcactgg catggccttc cgtgttccta cccccaatgt gtccgtcgtg gatctgacgt 7320
 gccgcctgga gaaacctgta tgtatgggga gagctgggct tgtctctggt gtgacagtga 7380
 cttgggacaa ggatagtcac tttggggttt gttcttatca ggccaagtat gatgacatca 7440
 agaaggtggt gaagcaggca tctgagggcc cactgaaggc catcttgggc taccactgagg 7500
 accaggttgt ctctgcgac ttcaacagca actcccactc ttccaccttc gatgccgggg 7560
 ctggcattgc tctcaatgac aactttgtca agctcatttc ctggtatgtg gggatgggaa 7620
 acctgacttt taagagcaac tgggggtttg gtgccctctg gtggctagct cagaaaagaa 7680
 acccaaaacta acagttgtcc caatttgttc taggtatgac aatgaatacg gctacagcaa 7740
 cagggtggtg gacctcatgg cctacatggc ctccaaggag taagaaaccc tggaccaccc 7800
 accccagcaa ggacactgag caagagaggc cctatcccaa ctccggcccc aacactgagc 7860
 atctccctca caatttccat cccagacccc cataataaca ggaggggcct agggagccct 7920
 ccctactctc ttgaatacca tcaataaagt togctgcacc cacctctcct tggttttaat 7980
 gcggtgttgg gggaggagg gtgtgtatgg ggaggtgaag caggctcaat caacttttgt 8040
 gcatgtatct ttattggctc tgacatccag ctgggtgccg gaagtccagg gctacattct 8100
 atccaagtcc cagcccaggc cgctgtcct gaacagctga tatgtctggc aatgccggagg 8160
 gtggggcttg acaatgggac aggaaatgca gtctgagact cacttgttgg acagcactga 8220
 cttccagagc tcaactgctc tgagaaagat cgggtgcctt ttatctgcaa acaaaggtta 8280
 attgtacca gccctcctgc agacagacgt aaaattagat aaaattggaa gggcgcccag 8340
 tacagtaaaa cctagtggta tccatgaaga ggggctgtag cactggtcga tgctttggtt 8400
 cgtgctcac acagtgactc agtgcttggg aagtattcca caccgtcatt gcttttaaag 8460
 gttccattcc aatagttcaa catagcgtct cctgggaggt cgtgactgac tccctaggtc 8520

ES 2 679 291 T3

aggacatct taaccttttg gttttgtcag tgacagccat ggcggcatac cctcagctgc 8580
 tcaactggaga aacaggcccc gagcaccagt aacctgcagg ccacacccat gtgcggggcag 8640
 cagcttcacc tcctcagccg cgacctgggc ccagcacagg acttcccaaa ggcgagcctg 8700
 cacggccatg ttcaactgcaa gaggcaagt aacagctggg agcatgtgga cagagcacc 8760
 cacaccctcc aaagccttac tagggagcaa agccctatcc caaggtacat gatgaatctg 8820
 cttcacacga gtgcctaggc aagttagcac cagtacaaaa tccaagatac ccaaggaata 8880
 caggcaggtc ccgtgatcac attcettaca aagtggaaaa gggaggctca caagtgtcct 8940
 ggcagcattc cgaagatgga gaagtagcag ataccacat gtgaggaagg caggaaggag 9000
 tgtcccaatg tcatgggctg tgtacagga ggggtgggct gacccatagct ccagacacc 9060
 cagaacctgg ggaggacag cccagggccca ggcttcatgc agtggcacag aagaccaca 9120
 cccctgcaga cccacacccc tgcccagtg agctctgggg tgccagacac tggtttctat 9180
 ctcatggagc taccagatga gacgtctcga tcggagtct cagtctcact cggcgggtggc 9240
 ggcgggtcgc taagcgggac cgcagtgaaa gcaggagact ttctagaaaa aacaccagc 9300
 tgtcaccttg gaccaagcag gaagctgca gacttctaag cttgctttct catggccttc 9360
 tgaggtacag ggggtgccag gcaaggctg gggaggaag aagacacttc cttttcaac 9420
 ttgatgggga gcagtgcctc tcttatcttg ggggctctg agaggctcag agatcagaaa 9480
 agtgcggaagt cacagcacgg agatgcccc tgagccacgc tggggcctta gcttctgtc 9540
 cagcagtaat ctggtatggg ggtgggggct gctggacccc aaaactcctg ggtgcaagca 9600
 atcctcttgc ctccaaagtg gctgggggtg agccacagat ctcttctaag catgcacact 9660
 atactcagct tgatgtaatt tgggagattg ggttgggttt tttgttctt ttgagacagg 9720
 gcctgtacgc attcctgtct ggcctggaac tctatacaga ggtgaggctg gcctggtact 9780
 tggagtacct gcctgttct tctatgcatt tgaatgacag atatgcgcta gcacacctga 9840
 cttactgtag tttcttggat tttgtttgtt gttgctatct ggtttgtttg ggacaaatct 9900
 ttgtctgtgt cgtccaggac agcctaaaac tcatgacaat ccttctgcct ctgccttcca 9960
 agactgaaac tgcagacatg taccactctg cctggtttgc agtaaatttt aaaaaaatt 10020
 tttaatgtgt gtgtgtctgt gtatgtgtat ctgtctgtgt gtctgtgtgt gtatgtgtgt 10080
 gtctgtgtgt gtatctgtct gtgtgtgttt atctgtgtgc ctgtatctgt ctgtgtgtat 10140
 gtctgtgtgt atctgtgtgt ctgtgtctgt ctgtgtgtgt atctgtgtgt gtgtgtttgt 10200
 atctgtgtgt gtgtctgtgt gtctgactgt ctgcctgtct gtctctgtgt gtgtgtgtgt 10260
 gtgtacacag ggccccacac aggctaggca agtcctctac cactaagcta tactccgaaa 10320
 cccagatctt actcttacac aaggtcaaac acataaacc gtgatttggg taagatcctg 10380
 tcccogtatc tttggtcctt gcctggttta ggttctaaga cactcagagt tgggacactg 10440

ES 2 679 291 T3

gtcagtttca tctgatggcc aaggcagggga gaggaagctt cattctgggc cacacctaca 10500
 ggcgactccc aaggacagct ggacataaag atgtgacctg gtcacctgcc acagaccatc 10560
 ggggcctatt ccagccttcc tccctgggag agagaccag cccgctcggg gcaggccctt 10620
 accggtcccc ggactgtgtg atgagccggc ggcaggtctc catctgcacg tccaggcccc 10680
 gcttcatgct gcacatctcc atgtactcat gcaggtgtcg gttcatgtcg ttcttggctg 10740
 tggccagctc cagctgcagt aaggtggggt cagggaggac cacggttagc aaaccaggc 10800
 aaggccctgg gggccaagca gctcaggaga caagtaagct acaactgcct gtttccccca 10860
 ggatatggag gaggggtggg gcaactggcg atctgacttc aggcctcac tgtggctagg 10920
 caagggtca ggacagtgct ccccgctcgg ccctgtcctg ccaatccact cctactgcca 10980
 cccaagaatc aatcagctga tgtctgagtt ctctgtaaga acagcaagag cacacaacac 11040
 agtaaggatg gtgcggtgcg gtgcggtgcg gggagcacgc atccctccag cgcttactcg 11100
 gctcaggcag cagggctccc aaggcaaacc taaatataaa acccggtctc aaaaaggggt 11160
 aaggaggcca caccaggcgg ctcagtggat aaggctcctg cgcactctggc aatgtgaatt 11220
 tgttctctaa cctccacaca tgcaccgtgg catgtggaga gagagagaca cacacacact 11280
 aataacaata ataataaaat ttagtatttt gggggctggt gagatggctc agtgggtaag 11340
 agcaccggac tgctcttcca aaggtccaga gttcaaatcc cagcaaccac atgggtggctc 11400
 acaaccatcc gtaacgagat ctgacgccct cttctggtgt gtctgaagac agctacagtg 11460
 tacttacata taataaataa ataaatctta aaaaaaaaaa aaaaaattta gtattttggt 11520
 tgtttgtttt togaaacagg gtctctccat gtagccctgg ctgtcctgga actcactcta 11580
 tagaccaggc tggcctggaa ctcagaaatg cgcctgcctc tgcctcccga gtgttgggat 11640
 taaagggtgtg caccaccact cccggcaagg cagcagttct taacctgtga gtctcaatcc 11700
 ccttgggggt tgactatcag aactctgca tgtctgtatt tatgttataa ttcgtaacag 11760
 tagcaaaatt agttatgaag cagcaatgaa atagttatg cctggggcca ccacagcgtg 11820
 aagcactgtg ttaagggtca cagcgtctgg aaggctgagc accactgctt taaggacaga 11880
 cagggccact catctgtcca tctttgccac tgagcctccc ctcaactgaa gctaattcac 11940
 tgtgaaatat tccaaagcct cgctgcaatg gagggaaacct ggctgcttct gtttatagca 12000
 tttcattcca catcatcttc tccctatat cctgctcgcc cacataagag agagaccttc 12060
 ctctgagtg attacagggt aaatcccaa caaaagaaat gggggggggg agtctctgta 12120
 tcaagaaaag tctcccttcc agagctgact ctgcagccag gtttgaaggc agcaaaaaaa 12180
 gtaggctgag gcggaaaagt atctaggcca gagacaaact gcggcctctc tttgtttggt 12240
 atacaaaggc caggggccag ataggagaaa tcagactgcg gtggggagat ttgcttggct 12300

ES 2 679 291 T3

gcctgagggc aaagacaagg aggccgagtg gccttgaagc atgtggatca gctgtgggaa 12360
gtcaccggaa acaaatgtga actggaccgg gtgggagtgg gcccaggag cagacaaagg 12420
agacagccag gacactggca ttttgggtca tctgctgaac agctgggtgg taagccacag 12480
aggagatggc ttgaaaggca tgcttgttcc cacagcagac tgcaccattt gtgagaaaaa 12540
catgacactg aggatcagag ctcagctgcc acccccaca gacaaatgca cacagcattc 12600
caggtcagag atcagctgcc acccatgtac atcaacacgc acacacacac aactattac 12660
aggtcagaga gcagctgtca cccatacata ccaacacaca cacacactat tacaggtcag 12720
agagcagctg tcaccatac ataccaacac acacacacac acacacacac acacactatt 12780
acaggtcaga gagcagctgt caccattca taccaacaca cacacacaca cacacacact 12840
attacaggtc gagagcagct gtcaccata aatacaagca taccctcac tccctcccct 12900
accccacctc cattatagtc tctcaaagca taaggcagta ctaccagcca tgtctctaag 12960
tgtcatatgt caccctgatt tacacactga aggggaagat gcacattcta gaagatgcct 13020
tcaacaaatc accccaaatc ttcagactca gtgtgggaga ggcaggggta gaggtagata 13080
ggagcagatg agggcggggg ccatggaaga aaactccacc caccactctg catcaccatt 13140
caataatggg ggctggtcta tctctatgcc aataaccctg agatcaaggg tcatggagat 13200
aaaggtccta cagatgctcc agggagaccg aggagctact ctgctaccaa cgctaccaag 13260
tgagcatcct ttggctttct cctcaactca ttctcccagg agagaacaga agtcatacaa 13320
cttaggaaca gcttcaaaca tgttcccaat tttttggga gggcaggggg agagtagagg 13380
tttgaagcag ggcttctctg tgcagccctg gccatcctgg aagtagttct gttcccagg 13440
ctgtcctcga actcacagag atcctcctgc ctgttttccc actaccagga ctaaagggtg 13500
caccaccaca ctccatttta taggtactgg tatcaaacct agggcttcac gcatgaaggt 13560
aggcagcat tctaccatga gagcacagtc ccagcccagg gtgctcttct ctgtcctacc 13620
acatcatggt tacacttagt tatttggggc ctaagggatg aaccacagaa gggcagcagc 13680
tcctgatcac tgtgctaaga ctctacctg gctgacggta gggaaaaaag gaccaagcct 13740
taaactctacc gtctgtaaga ggtggacaca acccaggag ggggacagga aggaggaagg 13800
agcagtaaga ggagcggaca ggagagatg ggagatcatt tgtccccagg caaagcttaa 13860
ctggtggcag aacactggct ggggtggaat tctgggttaa aaaccaaag cactggtttt 13920
cctcagagta gcacggccat ctgacatgtg ctggccactg cagaagagtt acttccaaag 13980
ttccctttca aaatgcaaac agagtgaccc ccatgggcac gggaatgcac cacagggctc 14040
gcagggagcc caatggcatt agtctcctgc ctcatccctg catcagtgtc taagacatag 14100
agccctgttc actgtccaat ggcgaccccc ttctgcagct cctgcctgc agtcacagag 14160
gacaaactaa accccagccc caccaccatt cataccgat gccccaaggc tcaaaaaaca 14220

ES 2 679 291 T3

aaagagggag acacctccag gagctccaag gctgagccaa ggagacagga gtccctctgg 14280
gtgctacggc acttaggggtg ttcccagtaa taaatgcgca cactaccgca tgtcaacctg 14340
ccagatocctg tcctaagagc cacgcgcgcc tcaactctatc tgatcatctt aataacctg 14400
cctgcctacc agcctgtcgg gagagattat gaacatctct attggacaaa taaagagaca 14460
ggccccacct tagcagggtg gaaaccatct ctaaccctaa caggctgagg caggaggatt 14520
gcctcaagtt tgactggcta cacagaaatc tggctctgggt gggggcttag gatggctcag 14580
tagttaagag catgccctgt tcttgacagag gacccaagtt tattttccag cgcccacatc 14640
aggttcacaa acacctgtga gcgccagctc ctggaaacct gatgccacc tctgacctcc 14700
aaggagagag ggggagggag ggaggaaggg aaggagggag ggagggaggg agggagggag 14760
ggagggaggg agagagagag aaagagagag agagagaaag agagaggaga gatgggggag 14820
tggatctgag aggaaggcag gcaggaaga tgaatgtgtg tgagagccta caacaaaaga 14880
ataaaaagga atgaaggagc tagctgagga acacctcatc agtgggaggc ttactgcaca 14940
gcatgaagac ctgagtttga tcttagaacc catctagaaa ataagctagg gctgaagaga 15000
tggctcagca gttaagagca ctgtgttgc tcttgcaatt cccagcaacc acatgtagc 15060
tcacaacctg ggacctgatg ccctcttctg gcatgcaggt gtacatacag atagaacct 15120
catatattaa ataataaat aaataaataa atcttaaaaa aataaataa actagccata 15180
gtagaacaca cttgcaatcc caaagctaag aatgtacaga cagacagaca gacagacaga 15240
cagacaaaag atgcctaggg ctccctggcc agctcggcct actcagcaag tttcagacca 15300
gtgaaagacc ctgtctcaat taaaaaaaaa aaacaaaaaa cccttaacca cactcaaaga 15360
acactaccga ggtttgtctt ctggcctcca cacttgtatg tgtacacatg tgccgctgca 15420
cgcacacaca cacacacaca cacacacaca cacacacaca cacagaggaa aggaggaaga 15480
cccaaggaag gataaaaagc tggtcagtca aagctggcct aaagctcctg ccgtaatcct 15540
atcgcacagg aggcgtgggc agaagcacca ggagcttgag gccaacctag actacaagac 15600
aagatcttga aaagaaacag agaggcaggg gcaggaagga cactcacaag tgcgagagag 15660
cttgacagttg tgatgcacac ttttaatccc aggagaggca cagtaacttc tccagagtga 15720
gttccaggac agccagggct acagagaagc cctgtctgtc agggggaaaa gaaaaaaaaa 15780
agaaaggggg ggacagaaaa gaaaaagata gtctggtggg aatgattca ccgggtaaaag 15840
gcatgagcca ccaagtcca agacctaac ttgattgctg tcaccccat ggtaaaagag 15900
aacttctcc cacacgttgt tctctgacct ctacacatgt gcatggactg cgcatgctct 15960
ctgacactcc atccatccat ccatccatcc atccatccat ccatccagtg tgtgtgtgtg 16020
tgtgtatgtg tgtgtgtgtg tgtgtgtaaa ataaataaaa aggaaggagg gagggacagc 16080

ES 2 679 291 T3

aaaagatttg ttccatgtgg tgacagagcc gggatttgaa cccacatcta tgtgcctcca 16140
gatcctgaag tctcaaggac ctgctacagc taccacaact gagaggatca ctactgacaa 16200
cagatggaag ctccgggccct ggggttcagc gaggaggagg ccagcaacac acctacagcg 16260
tggctcctga gccctggacc tacagcgtgg ctctgaggc cggcaccctt ac 16312

<210> 19
<211> 11058
<212> ADN
<213> Rattus sp.

5

<400> 19
gtgagtggac ccggatcact ggcctgttg ggtccgggtg taggaaacgg aatccgctct 60
ccctaagtgc tcagagatcc ctgtctggac ttgctagcct cactacagcc gtgcaaccga 120
agtgacttaa ggccagacct tttttttttt ttttttctga gctggggacc gaaccaggg 180
ccttgcgctt gctaggcaag tgctctacca ctgagctaaa tcccacccc caggccagac 240
cttttttaaa ggcattcgtt ttcctctttt atttattttt tagtgtgcac tggatatttg 300
cctatgttta tgtctgtgaa ggagcccagt gccctggagt tggagccaca gacagttgtg 360
ggtgctgaga actaaacaca ggtcttaacc actatgccgc ccccccccc cccagcctca 420
gtaaaggcca ttttctactt gcgatagggt ctgctatgt agaccagct ggcctcaaac 480
taacagagat tcccctgcct ctgtctccag tgtgctggta gtaaagggtg accccatgt 540
tcggcactag actgcttctt tagagtggtc tcatgtttaa agtaaacctg agtcccctc 600
tgcctgtccc acaccctcat ggcctctacc ccatcacctt atcctcggtg cgtgatccta 660
caatctacat tgactgggat gaaggaatga gattgacaga tcagcatcca aagtccactc 720
actgaatcca gtggcacatt cttgccacta gtgtcactt gggcattttc aggactgtga 780
ggtcacctca ctccccacc tgtggtggta tcagtcagtt ggaatcacag tttcagagtt 840
tgtttcttta acttgctgta tctttcatgg ctgcttttta aatagggaga taatgcttcc 900
gaccccgctt ttaccttttt ctccaggctg taatagatga atttgagcag aagcttcgtg 960
cctgtcacac cagaggcatg gatggaatag aggagcttga aactggccaa ggaggcagcc 1020
agcgggccct gtctgccaag aaaccctccg ccggtgtgag aggcaagccc acgctgggcc 1080
ccacagcgag tcccctcttt tctttttttt cctttgtggt ttgtttttgg ttttgaatc 1140
atgtccctct gcatattcca gtctcaaggc aacagcctct gacttctgta gactcagaca 1200
atgactttgt cacacctaag cctcgacgta ccaaaccctg tcgtccacaa actcagcaac 1260
gcaagtccca gaggaaagcc aaagttgtct tctcaagtga cgagtccagt gaagatggta 1320
agatacacc tcccacagc acaactgctt gtgtggggtc tccagccggg cttgcctggc 1380
tgtggattca ttctccctgg cagtgtcttc cccactgtgc agctgctggg ctgactttca 1440

10

ES 2 679 291 T3

ctctgtgtgt gtgtttgtgg tgtgtgtcca gatccgtttt gccccatctt tatgactcag 1500
 cgtttgttcc tataattctt tgaatagaac tttcagcggg gatgaccgaa gaagagacac 1560
 ccaagaaaac ccccccatc cgcagagcat ctgcgcaaag acacaggcc taggagctgc 1620
 ctttgccat tcccatccct gctgtgcagg gtatcttact gaatgatcga gaattcaacg 1680
 ccccccccc ccttgaaaa tacttggtgc ctcttggtt ttaacatgta aatggctctt 1740
 tcgagcaagg cttttctctt agagcccctg actgctactt tccaagatcc aatcaccatg 1800
 ttttctagt aataaatggt tttatatact ttttaagcatt ctggccggct ggtctgttcc 1860
 attattaggc agcagatagc agtccagaga gagaacccca ggcatcctga gccatctctg 1920
 gaccttcag gaccctgtgg tgcttcatct ttgcacaact gctgagattt gctgtagaga 1980
 gttctgtggt ttctaggttc acataagcac ttctcagcct gggggtactt tgcaccccag 2040
 aacacgtttg gcaatatcta gagaccagg gtgggagaga tggctcagcg attaggagca 2100
 cttactgccc ttccagaggt cctgagttca attcccagca accacatggt ggctcacagc 2160
 catctgtaat gagatctggg gccctcttcc ggtgtgtcta aagacagcta tataataaat 2220
 aaataaatct ttaaaaaaaaa aatctagaga cctatttctt ctgcccacct tgtggggcag 2280
 gggctgctgt aggcacccag tcagtcagag cctaaggagg aagactcctt catccatagc 2340
 actcccaaca cagcatgtcg ggactgtgca actagagggt ctgctctaga cctgcggttt 2400
 gtttctgtg tgccacaccc catcccagaa gactggcaaa tgagtcagcc tggctggtgg 2460
 cagagatcat gccattagct ctggttggtga caaggccac ctccaagag tcagtcctcc 2520
 actatgctaa ggggtctctc gctcagctgc aggccaggc tgtgctatgt gctgccacc 2580
 ttcccacag ccttaatctg ctgagtcact tggagcagga ggcaggactg ttgtttccgc 2640
 tgtcatgaca acagcacagt gcagctgagc ctctccctac ctcttggtat ggaagagta 2700
 atgacaggcg tgaggcagaa gaagatgcac ctcagaaaca ctaggtagg ggtggactgt 2760
 ggaagggctc ttgagacctc tgggtaggaa aagggttaac tcggtggata gacctcagt 2820
 aggaagacag aaaagggaca gctgagctct agatgcctgt ctcccagca ctgcaactgg 2880
 gccactccta ccctttccac cctgtttttg gcctgctccc tgctcatgaa aagcctaggc 2940
 atggaccaca cagctgtcag tctaaacaca gctttggctc tggggcccca aggaggtagg 3000
 gcaatcccag aagggcaagg agccaggact agatttgggg tgcagcccag cgtgctccct 3060
 gccttttaag caaaggttat caccaggcca gctaaactta gcaataggct cttcagctag 3120
 aagtgcaggg ggctggtctc aagttgcact gggctagcaa agaagcccca gactccccct 3180
 ggccgcgacc tgctgatggc ccgcacctcc caatcaaccc cttccacctt ttctagcctc 3240
 cctaccctgc aaagcaaaaa agtccttttc tccctccact gtgcaaagag gctaggggtt 3300
 gcctgttctg gctcgggata cttcccttct gctttcttcc ctttctgatc tcaccctgct 3360

ES 2 679 291 T3

tccacgaggg cagccaagga aagagagtcc ttggctcctg ggcagtcag caggtcgcct 3420
ggatggggag ggacttccgc cctcacgtcc cagctctcca ccctggggct gcggtgggtg 3480
aaaggacag tgtctccaat cctgggaggc acaactcagg ttccggggag ggaaggctc 3540
tgagcctcct ccaacccaac cctcaacagg gatgcttacc ccggggcccc ggtataggtt 3600
cagcaggtaa actcagggaa gtgtgtctc aatcaggtt ccagactgca ggggatcccc 3660
gacctatggc gtaggaatct ccttttcaag cctctagtca cttccttgct ttcacatcgc 3720
ggtagacaag ccaggactct ggcctaagag atagaagcca ggcggtgggg agagagggaa 3780
atggggctgc gaagggcccc caccggagga aaagtgcccc ccccgctcct cccccagcct 3840
tttcttctgc tcttttttcc gggcgggggg gggcgtgctg tgtcactacc gaagaacaac 3900
gagaagatcc tcaacttctc ctaagccttt tcactaatag ggagaagttc gatggggcag 3960
ccttgggcag acccacactt ctgctcatt tocctggttc ctgcagctct cagattctcc 4020
cattttattc gggaagcagc tttctggttt ctgggtcctg gatgtccttg gtgcacactc 4080
caaggactcc tcgtccttaa tccatagtct gtattccctg agtccctatcc tgggaaccct 4140
catccggtca cttcctcggc gggacaatct cagctcccct cccctctca ggtcggagcc 4200
cacacgcttg gtgcgtgcac atttcaaaaa cgaggcgggt ccaaaaagag ggaggggggg 4260
aatgagagag gccagctac tcgaggcttt acgggtgcac gtagctcagg cctctgcgcc 4320
cttgagctgg gactggatga gccgagcggg aggcggggcg cgcgtcatca gctccccca 4380
ccatccagtt cctataaata cggactgcag ccctccctgg tgctctctgc tctccctgt 4440
tctagagaca gccgcatctt cttgtgcagt gccagtgaa aaagaaaaga aaaaaagga 4500
ctgggcccga ggaggccgga gacgaatgga aattaggaat ggggggaag acgctgtacg 4560
ggtttagggg cgctgggtcg aggtccggaa gccgagccca ggctccgcat tgcagaggat 4620
ggtagaggac gtgatggggc atgcggcggg aatggaggcg ggtgggggga ggggactggc 4680
cacgctaact tgactttctt ctcccgcagc ctcgtctcat agacaagatg gtgaaggtcg 4740
gtgtgaacgg gtgagttcca gggcggggcc ccgctccgtt gcgcatgccg gtctggctaa 4800
attcggggct gtgctgtagc gtgcggggag gtggataggg tggccgaagt acccaaggag 4860
acctcaaggt cagcgtcggc acctggggaa ggctcgcact taacgggacc ccagctgttg 4920
ctccccttg aactccgcct ttgctgggga ggggcccgga gtctaaagta gcaggagcac 4980
ccccgacact ggcgcgcacc cgctcaggat ccgaccgta atccccaggc tggggcttct 5040
ttctttactt tcgccccagc aggagtcacg tgccagaagg ggagcccctc ccccatcgtc 5100
cccttctca ggtcgggtg cacgggggtg tgaggtgcat acctttgcgc atcttctaag 5160
ggcttactt atggtaactg gccgcccga tgttgcaaac ggaagggaaa tgaatgaacc 5220

ES 2 679 291 T3

accgtaaga aatcttcctt cggccttccct cattcttagc ttgtgactaa ctttctcatt 5280
 cctctcagct ggggtgagtg tcctttattc tgtaggccag gtgatggatg gcagggtttc 5340
 cgtgctctca ttagagctct actatacaat ccgtctctcg tcagccttta taggcctcca 5400
 gccttatttc cctcgggcat aacttatatt ctagattggt ctctggaagt cgaagtgcac 5460
 ttacagaggt ctacttggcc tcccagcccc ggaggtgagg tagcaatggc gtagtgccga 5520
 gcccgggggt ggggtgggga gctgagtcac gatggctctg aaaagaaatt ttctgccaca 5580
 aaatggctcc ggtgtagca gccccctccc tccagggcct ccaactccat cccagctcag 5640
 cactgaccca aaccctatag gccaggatgt aaaggctact aagaggattg ggtgtctctg 5700
 agcctcagga tcctaccctt tcccccaacc catcctccag aaaccagatc tccccattcc 5760
 gccctgatct ggggttaaata ttagctgtct gacctttctg tatctgggggt ctgagctggg 5820
 ctctctgccc tgttcatccc tccacacatc tgttgctcct gctccgattt tttttttttt 5880
 ttttttttgg ccgggaaaga cagggtgtttt gcaaatgagt cctgggatta gggctggaaa 5940
 atcactgggt ttcttgatct tccccccaga aagggacttg gaggaagcag gtgggctggc 6000
 cctgtcctgc tcaactctgac ctttagcctt gccctttgag ctgctgatga atgagttctc 6060
 tcctgtccgg gagtgtagcc tgaagtccag ccatgctgga accggcttcc caggcctgtg 6120
 tgggtgggggt gatgaatacc atgtgccaag cttgtacca gaccggggca aaccgccact 6180
 tcttaagaga cttaaaatga ctttgggggt ccgggcaagg ctgtgggcct aatcacctct 6240
 tggacaggaa aggaacctcc actttatagc tgctataaaa gtctgccag ccctgggtgg 6300
 ctcaaggaat ataaaattag atctctttgg actttctagg gagggaacat tctatatttg 6360
 ggtgtacat ccaagcattc aaccggcttt attaaaggaa attctggaca aaacatgaac 6420
 ttctgtctgc ttacgtttct gatgctgtac tgaatcagcg taagaacttg cacctttttg 6480
 tgatgcgtgt gtagcgggct gctgtaggat ctccacgccc atggtgcagc gatgctttac 6540
 tttctgaaga acatgcgttg gcccagggct gactacaaac ccaggagggg ttactgatc 6600
 ccaactaact cgcctatttc ttgcctcaga tttggccgta tcggacgcct ggttaccagg 6660
 gctgccttct cttgtgacaa agtgacatt gttgccatca acgaccctt cattgacctc 6720
 aactacatgg tctacatggt ccagtatgac tctaccacg gcaagttcaa cggcacagtc 6780
 aaggctgaga atgggaagct ggtcatcaac gggaaacca tcaccatctt ccaggagtac 6840
 gtatggaaac atgcacaggg tacttcgagg agatggtgtaagg gggcggcatt aactctgtcc 6900
 tgttccccgc ctgtaggcga gatcccgcta acatcaaatg ggggtgatgct ggtgctgagt 6960
 atgtcgtgga gtctactggc gtcttcacca ccatggagaa ggctggggta agtggccagg 7020
 aagctggagg taaggggaac cttgatatg ggtgcaacct ccaaactgaa gagctgagtc 7080
 tgaatcaac ttctcctgca caggctcacc tgaagggtgg ggccaaaagg gtcacatct 7140

ES 2 679 291 T3

ccgccccttc cgctgatgcc cccatgtttg tgatgggtgt gaaccacgag aaatatgaca 7200
 actccctcaa gattgtcagc aatgcatcct gcaccaccaa ctgcttagcc ccctggcca 7260
 aggtcatcca tgacaacttt ggcatogtgg aagggctcat ggtatgtagg caatggagac 7320
 agctcatgca tgagtggacc tttctttgaa gatgtccctt tgggtaggag ggctgccctg 7380
 caagacctca cccattgcct ctatgctttc tagaccacag tccatgccat cactgccact 7440
 cagaagactg tggatggccc ctctggaaag ctgtggcgtg atggccgtgg ggcagcccag 7500
 aacatcatcc ctgcatccac tgggtctgcc aaggctgtgg gcaaggctcat ccagagctg 7560
 aacgggaagc tcaactggcat ggccttccgt gttcctacc ccaatgtatc cgttgtggat 7620
 ctgacatgcc gcctggagaa acctgtacgt agggggagag ctgggtttgt tttgtctatg 7680
 gtatgggtgt gacttggggc aagagcagtc atctttggtt ttgcccttac aaggccaagt 7740
 atgatgacat caagaaggtg gtgaagcagg cggccgaggg cccactaaag ggcacccctg 7800
 gctacactga ggaccaggtt gtctcctgtg acttcaacag caactcccat tcttccacct 7860
 ttgatgctgg ggctggcatt gctctcaatg acaactttgt gaagctcatt tcctggtatg 7920
 tggggaccgc aagcctggct cttgagagta actgaaggtt tggtgccctc tgggtggccag 7980
 cttagaaaag aagcccaaac taaccgttgt cccaatctgt tctaggtatg acaatgaata 8040
 tggctacagc aacagggttg tggacctcat ggcctacatg gcctccaagg agtaagaaac 8100
 cctggaccac ccagcccagc aaggatactg agagcaagag agaggccctc agttgctgag 8160
 gagtccccat cccaactcag cccccaacac tgagcatctc cctcacaatt ccatcccaga 8220
 ccccataaca acaggagggg cctggggagc cctcccttct ctggaatacc atcaataaag 8280
 ttcgctgcac cctcttcttg tgtgtgaatg cagtgttgtg ggggttatgg gcaggtgaag 8340
 gaggtccat tcggttttgt gcacatatct ttattggctc tgccatccag ctgggtgcta 8400
 gaggtccagg gctacattct aaccgagtc cagcccaggc cgctgtcca gacacctgat 8460
 gtgtctggca gtgtggaggg tggggcttga cactggaaca gaaggtgcag tctttgaaca 8520
 gcactgactt tcagagcttg ctgctcttga gaaagatctg tgtcctttat ctgcaaaaa 8580
 aggttaattg taccagctc tcaaagtctc cagatgttaa attaggtaaa gttggaaggg 8640
 tgcccagtac agtaggacta atttgtcca tgaagaggga aggtccctgc tgagctgttc 8700
 acctgtcttc aggtgtgggc acttgaaggt gctttggttc tggggcagtg actagccagc 8760
 gctcactggt ggttgggaat gaaatgaagg tgcattacac ggtgactcga ctcaacgcct 8820
 tcattgcttt caaaggctcc attcaaatgt cagctctgca tagcctctcc tgggagatcc 8880
 tagctcccta ggaccagga cattgtaacc ttttggtttt gtcagtgaca gccatggagg 8940
 cattccctca gctggctact gaggagacac cctgagcacc cctaaccctg tgcaggccac 9000

ES 2 679 291 T3

acccaagtgt ggcagcagct tcatctcagt ggcgacctgg cccaggcact ggacttccca 9060
 aaggcacagc catgttcact gcagtgaggca ggtgggcagc tagctgtgag catgtgcagt 9120
 gtccacgccc tcaaggcctt cctagggagc aaggctctat ccaaaggtac atgcggaatc 9180
 tgtttcacat gagtgcctag gccagttagc accagtacaa aaccaagga aaacaggcag 9240
 gtcccataac cacattcctt acaaatggaa aagggaggct ccaagtgtgc tggcagcatt 9300
 ccaaagatgg agaagtagca gatacccaca tgtgaggaag gcaggagtgc cccaatgtgt 9360
 tgggagtgtg acaggggggtg gggccgagcc tagctccaga cccccagaa ccctggagag 9420
 gacaagccag ggccaggcctt catgcggtgg cacagaagat ggtcctctgc agaccacac 9480
 ccctgcccc a gagactctg gggcaccagc ccctggttcc tatctcatgg agctgtcaga 9540
 tgagacatct cgatcggagt cctcagtctc gcttgccggc ggccggcgggt cgctaagcgg 9600
 gaccgcagtg aaagcaggag actttctaga aagaaatacc agttgtcacc ttggggcaaa 9660
 gcaggaaaagc tcctgagact tccaaacttg ctttctctgg ccttatcttt tgaggtagag 9720
 ggggatgaca ggcaaggacg ggggagggaa ggaaaacttg ttttcaactt gatgggagga 9780
 acagtttcgc tctttgggac tcagagatca gaaaagtctg aagtcacagc atggagatgt 9840
 tccttgagcc aactaaggc cctagcttct tatccagcaa taattcgggtg tgcggggggc 9900
 tgtggggact gtaggagggtg agactggggg agtatggggg tttagggggc tgtttcactc 9960
 tgttgtcact gctgaccca aattcctggg tgcaagcaat cctcttgctt tcagttgatt 10020
 ttcttttttt ttttaattta ttttatttat atgagtacac tgtaactggt ttcagacata 10080
 ccagaagagg gcatcagatc ccattacaga tggttgtgag ccacatgtg gttgctggga 10140
 attgaactca ggacctctgg aagagcagtc agtgctctta accgctgagc catctctcca 10200
 gccctctctg ccttcaaagt ggctgggggtg gccacaaacc cagttagcgt gctcaccata 10260
 caatgcaatt tggggggctg ggttgggttg ggtttttgt tctctttgag atgagtcagt 10320
 ttgcattcct gcctggcctt gaactcctca cagaggcgag gctggctag tacttgtggc 10380
 aattcacctg cccgttgctt ctgtgcactc gaatgatagg tatgcgctca taagcctgac 10440
 ttaactgcca ttccttgat tttgtttgt tggtgtgtgct gtttggtttg tgtgggacaa 10500
 ggttttactg ctgtccagaa cagcctaaaa ctcatgacaa tccttctgcc tccgccttcc 10560
 aagactgaga ctgcagacat gtaccgctag gcctggttta gagatttcta aaaaaatctt 10620
 agtgtgtgtg tgtgtgtctg tgtgtgtgtg tgtgtgtgtc tgtgtgtgtg tgtgtgtgtg 10680
 tctgtgtgtg tctgtgtgtg tctgtgtgtg tgctagacag gctagacaag gactctacca 10740
 ctgagctata ttccaaaacc ccagatttac tcttaacaa ggtcaaacag ctaaactgtg 10800
 atttggattt ggataatcct gtcctctat ttctttgctc cttgccaggt ttgggctctc 10860
 agaccctcag agttgggcac tagtcggttt catctgacgg ccaaggcagg gagaggaagc 10920

ES 2 679 291 T3

ttcattctag gccacaccta cagggactcc caaggacagc tggacataaa gatgtgacct 10980
 ggtcactccc acaggccact ggggcctgtc cagccttctt ccctgggaga cagccacgcc 11040
 actcagggcg gcccttac 11058

<210> 20
 <211> 3158
 <212> ADN
 <213> Homo sapiens

5

<400> 20
 agcttgagat tgtccaagca ggtagccaga gagcgcctac agccaagaaa ccatccactg 60
 gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
 gtttcttctt cttttttaa tcaagttttc ctgccttttc taggttctag gtaccagcct 180
 ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgcg tactaccctg 240
 cggcatccaa acaccagca gcgagcttc aaaaagaaac ccaaagttgt cttctcaagt 300
 gatgagtcca gtgaggaagg tatgatgctc ccgctgttc ccggccgaga aggcacacag 360
 ctagggtgca gagggctggt ttccatagga cctgctgagg gggcctgagt gtagatgctc 420
 tgccccactg ccgcagaagg gcctctcctg tacagcttg attttatttc ttctgtgagg 480
 tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg tgactcagct 540
 ttttcttatt ctttaattc tttgcataga tctttcagca gagatgacag aagacgagac 600
 acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660
 ctgttctgt cctccctgtg cagggtatcc tgtagggatga cctggaattc gaattctgtt 720
 tccctgttaa aatatttgc tgtctctttt ttttaaaaa aaaaaggcc gggcactgtg 780
 gctcacgcct gtaatccag cactttgcga taccaaggcg ggtggataac ctgaggtagg 840
 gagttcagaga ccagcctgac caacatggag aaacccatc tctactaaa ataaaaaatt 900
 agccgggctg attggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa 960
 tcgcctgaac ccagaggcgg aggtttagt gagccgaaat cacaccattg cactccagct 1020
 tgggcaaaa tagcgaacct ccatctcaa ttaaaaaaa aatgcctaca cgctcttaa 1080
 aatgcaaggc tttctcttaa attagcctaa ctgaactgag ttgagctgct tcaactttgg 1140
 aatatatggt tgccaatctc cttgtttct aatgaataaa tgtttttata tactttttaga 1200
 cttttttcc taagcttgtc tttgtttcat ctttcacatt agcccagttt catgcagcag 1260
 agagagggtt atcagtgcag agagagatga gtgagcccag agtcctaggg cctgtcccgg 1320
 gatggcagat gagcttcctg ccccgctcact gccaccttc ccctctcaac ctctggacct 1380
 tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc tccagatcag 1440
 tgcttctgaa ccgggggcaa tttgtctgc cagaggacat ctgacaacac ctggggcctg 1500

10

ES 2 679 291 T3

ttttgtgtgc atagcctata ggggaagaat gctaccagca tttgtgggaa gaggccaggg 1560
 atgtggctca acatcctgca gtgcacagga tggccoctca acaaagaatc acacggoccca 1620
 caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg ctttctgccg 1680
 tgtgcctcac cccaccccca ctcgtgttcc ctaatcccat ctccaaaggt tggcagcaga 1740
 ccggcccagg ctcgtggaag ttcagatcat gatcccctcc agctctgcag gagacaagac 1800
 ctgtctccca gcatcctca ttgttcccgg gtctgcagag ggcgtgagct atgctgcagg 1860
 cgggctgcc cctgaagcct gcgcaccctc ctccagctcc tcaagtcttc tctgtgagt 1920
 caccttoga cgggaggctg tgagctggct gtctgacca cactggtgcc tctgtgtca 1980
 tgacaacagc aactacgtc agtagtgctc cctgggcact gagctccctc tttgcgggga 2040
 gaagacagta atgaaaaatg acaagcatga ggcagagggg aagatcacgc ttgggtgggtg 2100
 caggagcatg gaggtgctct taatgctctc aatgagaaag ggtaacggc cctggttgca 2160
 ggaatagctg agtcagaggt ggggcttctc cactccccc accccacccc tttcaccatt 2220
 agggaccttc ttgccttgct cttgctactc tgctctgggt ggtcattgtg aaaagccgc 2280
 accaaccatg ccagtggcag ccagacgagg acacagcctg gctctgggtc ccagcaggaa 2340
 aggcaatccc agaaggcag ggtcagggac tggagtctg tgggtgcttt ttaagcaaag 2400
 attatcacca ggcaggctaa acttagcaac cggcttttag ctagaagggc agggggctgg 2460
 tgtcaggtta tgtggggcca gcaaagaggc ccgggatccc cctcccatgc acctgctgat 2520
 gggccaaggc caccaccacc cacccttc cttacaagtg ttcagcacc tcccatocca 2580
 cactacaaa cctggccctc tgccctccta ccagaagaat ggatcccctg tgggaggggg 2640
 caggggacct gtcccaccg tgtgcccaag acctcttttc ccacttttcc cctcttcttg 2700
 actcaccctg ccctcaatat ccccggcgc agccagtga agggagtccc tggctcctgg 2760
 ctgcctgca cgtcccaggc ggggagggga ctccgccct cacgtcccgc tcttcgcccc 2820
 aggtggatg gaatgaaag cacactgtct ctctccctag gcagcacagc ccacaggttt 2880
 ccaggagtgc ctttgtggga ggcctctggg ccccaccag ccacctctgc ctccgcctgg 2940
 ggcccagcc cggagagagc cgctggtgca cacagggccg ggattgtctg ccctaattat 3000
 caggtccagg ctacagggt gcaggacatc gtgaccttcc gtgcagaaac ctccccctcc 3060
 ccctcaagcc gcctcccag cctcctcct ctccaggccc ccagtgccca gtgccagtg 3120
 cccagcccag gcctcgggtcc cagagatgcc aggagcca 3158

<210> 21
 <211> 3218
 <212> ADN
 <213> Homo sapiens

 <400> 21

5

ES 2 679 291 T3

aagctggcac cactacttca gagaacaagg ccttttcctc tcctcgctcc agtcctaggc 60
tatctgctgt tggccaaaca tggaagaagc tattctgtgg gcagccccag ggaggctgac 120
aggtggagga agtcagggct cgcactgggc totgacgctg actggttagt ggagctcagc 180
ctggagctga gctgcagcgg gcaattccag cttggcctcc gcagctgtga ggtcttgagc 240
acgtgctcta ttgctttctg tgcctctgtg tottatctga ggacatcgtg gccagcccct 300
aaggtcttca agcaggattc atctaggtaa accaagtacc taaaaccatg cccaaggcgg 360
taaggactat ataatgttta aaaatcggta aaaatgcca cctcgcatag ttttgaggaa 420
gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg gaacttgggt 480
aggggcaagg cgtgtagctg ggacctaggt ccagaccctt ggctctgccca ctgaacggct 540
cagttgcttt gggcagttac tcccgggcct cactttgcac gtgtgcttac ctagtggaga 600
caaaagtaca tacctcggta gagcgcgcac gcctgtaacc ccagcacttt gggaggccaa 660
ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat ggtgaaactc 720
cgtctctact aaaattaca aatcagcca ggcttcatgg cacatgccta tagtcccagc 780
tacaggcatg ctgaagcagg agaatcgctt gcaccccgga ggcagaggct gcagtgagct 840
gagaccacac cactgcactc cagcctaggc aacagagtat gagactccat ctcaaaaaaa 900
aaaaaagtac ctacctcaga gttcaaaacta gtgaatatta ggaagtgctt gagacagtga 960
caccaaagtg cacaataaat actcgcaggt ttattatta ttaaagaatc catttgaatg 1020
tcagctcaac acagcctcct ataccgaggc attgtgaacc gcatctcccc agcttctcca 1080
ggcttttcca agaatcaggg acaactgtagc ctgttggctt cagtgtatga cagacacgga 1140
ggaagcacat ctttagctga tacttaaaca gagaccctga gcgcacatac acccgcgcac 1200
acatgcatgg agcttcacct tctctgtcat totgcagtga ccaggagagc aagagctccc 1260
acctcccttc aaaacactgt gcccatcccg ggcactaagg cctctttaa gcacggcacc 1320
tccacgaggg agggccacag ccacatacac tccacctggc aggtggacag cgtgagcacg 1380
tggaccatag cagggacaag gtgcccggc cagcccaac gccctctgcc gctgacaggg 1440
acagaagccc tctccagctg cgtgtgctgc agaggccatg cgtagcctcc agctgcattc 1500
tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga gctggctccg 1560
gacaacaggg atggaggaaa ggtcccacat tcacattcct gatacgtgga caaggtgagg 1620
ggccgcaatc gctctggcag cattttaaag atggggaagt agcagacacc cacgcgtgaa 1680
ggcaggagag ccccaactgt ggtggaaatg gcccagaat ggtagggcca agcctagctc 1740
cagacacccc agagccctgg agaagccaag actgaggag aaagcctgag ggaggagcgc 1800
cccagtcccc agggaccggc ctggtgcaga gctgcagctg atgttcccct ctgtgcagcc 1860

ES 2 679 291 T3

ccaccctctg cctcgctgag ctccctgctg cgagggcctc gggtgcaagg gggaggcagg 1920
tctctatctc atggagctgt cagatgagac atcgcgatcg gagtcctcag cctcgcttgg 1980
cggcgggcggc gggtcgctaa gggggaccgc agtgaaagca ggagactttc tagaaaaaaa 2040
caccagttgt caaccttggg gcaggcagga atcctgaaga cggacggcac tcctcctcct 2100
gctgcctcac cctctggcag cccgtgagaa gtaccggaag cgagggcggg gccgcgggat 2160
ggcgagggag cggcagggac tgaactctct ccaaaccac cctgacaggg aaatgggccc 2220
cgcctgtgtc ttgggaactc agaggctgag gtcaggcatg atggctcacg cctgtaattc 2280
cagcactttg ggaggcagag gcgggtggat cacgacgtca ggagttcaag gcaagcctgg 2340
ccaacatggt gaaaccccat ctctactaaa aatgcaaaaa ttagccaggt gtggtggcgg 2400
gcgccttggg ggctgaggca agataatcac ttgaacctgg gaggcggagg ttgtagttag 2460
ccaaaaaaaa aaaaaaaaaa aatagctgaa gtcacagtag gagagaagct gctgagcctc 2520
cagcacccctg actctagggc cttggcttta tgtctatctg cagtattttt gtgattttta 2580
aaaattcact ttcttggtgc ggtgtaactt acacagggtc aaatgcacaa atcatggccc 2640
tgactttaga taaaaatctg cccccacaac ccttctgttc cttgccagtt tttaaactgc 2700
ctctaaccag ggaaccacc agagctggtg gccttgggag gtttcagccc tcccgtcatg 2760
aatggacata gctcatccaa ctgcccaagg agagagctgt gggctctgggc cagccccacc 2820
aggtaactcc caaagggcag cccacagca agatgtgacc cagtcattgc ctgagggctc 2880
ctggggctgt gttccaacct ttctcccgc tgtgtcccc tggaaggccc catgccagg 2940
ggaggcgcct accggtctcc agactgggtg atgagccggc gccaggtctc catctgcacg 3000
tccagggcgc gttcatgct gcacatctcc atgtactcgt gcaggtgccg gttcatgtcg 3060
ttcttgccg tggccaactc cagctgtggt ggggcaggca gggccatcgg ggttaacagg 3120
tggcgttcac agcgcctctg ttgccccgc caggaggcca acacgccaag agcagtggct 3180
gggcccgggg cccaggcagc cattactgaa ggctcaga 3218

<210> 22
<211> 508
<212> ADN
<213> Homo sapiens

<400> 22
agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60
gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
gtttcttctt cttttttaa tcacgttttc ctgccttttc taggttctag gtaccagcct 180
ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccctg 240
cggcatccaa acaccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300

5

10

ES 2 679 291 T3

gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc cgggccgaga aggcacacag 360
ctaggggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420
tgccccactg ccgcagaagg gcctctcctg tacagcttgg attttatttc ttctgtgcgg 480
tgtgggattg tctcacttgt tctctgat 508

<210> 23
<211> 2650
<212> ADN
<213> Homo sapiens

5

<400> 23
atctatTTTT tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acaccaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat 180
cctgtagggg gacctggaat tcgaattctg tttcccttgt aaaatatttg tctgtctctt 240
ttttttaaaa aaaaaaagg ccgggactg tggtcacgc ctgtaatccc agcactttgc 300
gataccaagg cgggtggata acctgaggta gggagttcga gaccagcctg accaacaatgg 360
agaaacccca tctctactaa aaataaaaaa ttagccgggc gtattggcgt gcgcctgtaa 420
tcccagctac tcaagaggct gaggcaggag aatcgcctga acccagaggc ggaggttgta 480
gtgagccgaa atcacaccat tgcactccag cttgggcaac aatagcgaac ctccatctca 540
aattaaaaaa aaaatgccta cacgctcttt aaaatgcaag gctttctctt aaattagcct 600
aactgaactg cgttgagctg cttcaacttt ggaatatatg tttgccaatc tccttgTTTT 660
ctaatgaata aatgttttta tatactttta gacatTTTT cctaagcttg tctttgtttc 720
atctttcaca ttagcccagt ttcatgcagc agagagaggg ttatcagtgc agagagagat 780
gagtgagccc agagtcctag ggcctgtccc gggatggcag atgagcttcc tgccccgtca 840
ctgccacctt tcccctctca acctctggac cctgcacagt gaccagacag cctctctggg 900
gagaattatg cagtgcctag gctccagatc agtgcttctg aaccgggggc aattttgtct 960
gccagaggac atctgacaac acctggggcc tgttttggtg tcatagccta taggggaaga 1020
atgctaccag catttggtgg aagaggccag ggatgtggct caacatcctg cagtgcacag 1080
gatggccctt caacaaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa 1140
acctgctcta gaccaagggt tgctttctgc cgtgtgcctc accccacccc cactcgtggt 1200
ccctaatacc atctccaaag gttggcagca gaccggcca ggctcgtgga agttcagatc 1260
atgatccctt ccagctctgc aggagacaag acctgtctcc cagcattcct cattgttccc 1320
gggtctgcag agggcgtgag ctatgctgca ggcgggctgc cccctgaagc ctgcgcaccc 1380
ctctccagct cctcaagtct tctctgctga gtcaccttcg aaccggaggc tgtgagctgg 1440

10

ES 2 679 291 T3

ctgtcgtgac cacactgggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc 1500
 tccctgggca ctgagctccc tctttgcggg gagaagacag taatgaaaaa tgacaagcat 1560
 gaggcagagg ggaagatcac gcttgggtgg tgcaggagca tggaggtgct cttaatgctc 1620
 tcaatgagaa agggttaacg gtcctggttg caggaatagc tgagtcagag gtggggcttc 1680
 ctccactccc ccaccccacc cctttcacca ttagggacct tcttgccttg ctcttgctac 1740
 tctgctctgg gtggctcattg tgaaaagccc gcaccaacca tgccagtggc agccagacga 1800
 ggacacagcc tggctctggg tcccagcagg aaaggaatc ccagaaaggc agggtcaggg 1860
 actggagtcc tgtgggtgct ttttaagcaa agattatcac caggcaggct aaacttagca 1920
 accggctttt agctagaagg gcagggggct ggtgtcaggt tatgctgggc cagcaaagag 1980
 gccgggatc cccctcccat gcacctgctg atgggccaag gccaccccac cccaccccct 2040
 tccttacaag tgttcagcac cctcccatcc cacactcaca aacctggccc tctgccctcc 2100
 taccagaaga atggatcccc tgtgggaggg ggagggggac ctggtccac cgtgtgccca 2160
 agacctcttt tcccactttt tccctcttct tgactcacc cgcctcaat atccccggc 2220
 gcagccagtg aaagggagt cctggctcct ggctgcctg cacgtcccag ggcggggagg 2280
 gacttccgcc ctcacgtccc gctcttcgcc ccaggctgga tggaatgaaa ggcacactgt 2340
 ctctctccct aggcagcaca gccacaggt ttccaggagt gcctttgtgg gaggcctctg 2400
 ggccccacc agccatcctg tcctccgctt ggggccccag cccggagaga gccgctgggtg 2460
 cacacagggc cgggattgtc tgccctaatt atcaggcca ggctacaggg ctgcaggaca 2520
 tcgtgacctt ccgtgcagaa acctccccct cccctcaag ccgcctcccg agcctccttc 2580
 ctctccaggc cccagtgcc cagtgccag tgcccagccc aggcctcggc cccagagatg 2640
 ccaggagcca 2650

<210> 24
 <211> 1963
 <212> ADN
 <213> Homo sapiens

5

<400> 24
 agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60
 gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
 gtttcttctt cttttttaa tcacgttttc ctgccttttc taggttctag gtaccagcct 180
 ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccctg 240
 cggcatccaa acaccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300
 gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc cgggccgaga aggcacacag 360
 ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420

10

ES 2 679 291 T3

tgccccactg cgcgagaagg gcctctcctg tacagcttgg attttatttc ttctgtgagg 480
 tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg tgactcagct 540
 ttttcttatt cctttaattc tttgcataga tctttcagca gagatgacag aagacgagac 600
 acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660
 ctgttctgtt cctccctgtg cagggtatcc tgtagggtga cctggaattc gaattctggt 720
 tcccttgtaa aatatttgtc tgtctctttt ttttaaaaaa aaaaaaggcc gggcactgtg 780
 gctcacgcct gtaatccag cactttgcca taccaaggcg ggtggataac ctgaggtagg 840
 gagttcgaga ccagcctgac caacatggag aaacccatc tctactaaa ataaaaaatt 900
 agccgggagt attggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa 960
 tcgcctgaac ccagagggcg aggtttagt gagccgaaat cacaccattg cactccagct 1020
 tgggcaacaa tagcgaacct ccatctcaa ttaaaaaaaaa aatgcctaca cgctctttaa 1080
 aatgcaaggc tttctcttaa attagcctaa ctgaactgag ttgagctgct tcaactttgg 1140
 aatatatggt tgccaatctc cttgttttct aatgaataaa tgtttttata tactttttaga 1200
 cttttttcc taagcttgtc tttgtttcat ctttcacatt agcccagttt catgcagcag 1260
 agagaggggt atcagtgag agagagatga gtgagcccag agtcctaggg cctgtcccgg 1320
 gatggcagat gagcttctg ccccgctact gccacctttc cctctcaac ctctggacct 1380
 tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc tccagatcag 1440
 tgcttctgaa ccgggggcaa ttttgtctgc cagaggacat ctgacaacac ctggggcctg 1500
 ttttgtgtc atagcctata ggggaagaat gctaccagca tttgtgggaa gaggccaggg 1560
 atgtggctca acatcctgca gtgcacagga tggcccctca acaaagaatc acacggccca 1620
 caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggtt ctttctgccg 1680
 tgtgcctcac cccaccccca ctctgttcc ctaatccat ctccaaagg tggcagcaga 1740
 ccggcccagg ctctggaag ttcagatcat gatcccctcc agctctgcag gagacaagac 1800
 ctgtctccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct atgtgcagg 1860
 cgggctgcc cctgaagcct gcgcaccct ctccagctcc tcaagtctc tctgtgagt 1920
 caccttcgaa ccggaggctg tgagctggct gtcgtgacca cac 1963

<210> 25
 <211> 1195
 <212> ADN
 <213> Homo sapiens

<400> 25
 tgggtgcctct gctgtcatga caacagcaca ctacgtcagt agtgctccct gggcactgag 60
 ctccctcttt gcggggagaa gacagtaatg aaaaatgaca agcatgaggc agaggggaag 120

5

10

ES 2 679 291 T3

atcacgcttg ggtggtgcag gagcatggag gtgctcttaa tgctctcaat gagaaaggg 180
 taacggctct ggttgcagga atagctgagt cagaggtggg gcttcctcca ctccccacc 240
 ccaccctttt caccattagc gaccttcttg ccttgctctt gctactctgc tctgggtgg 300
 cattgtgaaa agcccgccacc aaccatgccca gtggcagcca gacgaggaca cagcctggct 360
 ctgggtccca gcaggaaag caatcccaga aaggcagggc cagggactgg agtctgtgg 420
 gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg cttttagcta 480
 gaagggcagg gggctgggtg caggttatgc tgggccagca aagaggcccg ggatccccct 540
 cccatgcacc tgctgatggg ccaaggccac cccaccacc ccccttcctt acaagtgttc 600
 agcaccctcc catcccacac tcacaaacct ggcctctgc cctcctacca gaagaatgga 660
 tccccgtggg gagggggcag gggacctgtt cccaccgtgt gcccaagacc tcttttccca 720
 ctttttccct cttcttgact caccctgcc tcaatatccc ccggcgccagc cagtgaaggg 780
 gaggccctgg ctctggctc gcctgcacgt cccaggcgg ggagggactt ccgccctcac 840
 gtcccctctc tcgccccag ctggatggaa tgaaggcac actgtctctc tccctaggca 900
 gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca 960
 tcctgtcctc cgctggggc cccagcccg agagagccgc tgggtgcacac agggccggga 1020
 ttgtctgccc taattatcag gtccaggcta cagggtgca ggacatcgtg acctccgtg 1080
 cagaaacctc cccctcccc tcaagccgcc tcccgagcct ccttcctctc cagccccca 1140
 gtgccagtg cccagtgcc agcccaggcc tcggtcccag agatgccagg agcca 1195

<210> 26
 <211> 256
 <212> ADN
 <213> Homo sapiens

5

<400> 26
 cctctgggcc cccaccagcc atcctgtcct ccgcctgggg cccagcccg gagagagccg 60
 ctggtgcaca cagggccggg attgtctgcc ctaattatca ggtccaggct acagggctgc 120
 aggacatcgt gaccttccgt gcagaaacct cccctcccc ctcaagccgc ctcccagcc 180
 tccttctctc ccaggcccc agtgcccagt gccagtgcc cagcccaggc ctcggtccca 240
 gagatgccag gagcca 256

10

<210> 27
 <211> 1941
 <212> ADN
 <213> Homo sapiens

15

<400> 27
 agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60
 gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120

ES 2 679 291 T3

gtttcttctt cttttttaa tcacgttttc ctgccttttc taggttctag gtaccagcct 180
 ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccogt 240
 cggcatocaa acaccagca gcgagcttcc aaaaagaac ccaaagttgt cttctcaagt 300
 gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggccgaga aggcacacag 360
 ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420
 tgccccactg ccgcagaagg gcctctcctg tacagcttgg attttatttc ttctgtcggg 480
 tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg tgactcagct 540
 ttttcttatt cctttaattc tttgcataga tctttcagca gagatgacag aagacgagac 600
 acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660
 ctgttctctg cctccctgtg cagggtatcc tgtagggtga cctggaattc gaaccggagg 720
 ctgtgagctg gctgtcgtga ccacactggt gcctctgctg tcatgacaac agcactactac 780
 gtcagtagtg tccctcggc actgagctcc ctctttgcgg ggagaagaca gtaatgaaaa 840
 atgacaagca tgaggcagag ggaagatca cgcttgggtg gtgcaggagc atggaggtgc 900
 tcttaatgct ctcaatgaga aagggttaac ggtcctggtt gcaggaatag ctgagtcaga 960
 ggtggggctt cctccactcc cccaccccac ccctttcacc attagggacc ttcttgctt 1020
 gctcttgcta ctctgctctg ggtggtcatt gtgaaaagcc cgcaccaacc atgccagtgg 1080
 cagccagacg aggacacagc ctggctctgg gtcccagcag gaaaggcaat ccagaaaagg 1140
 cagggtcagg gactggagtc ctgtgggtgc tttttaagca aagattatca ccaggcaggc 1200
 taaacttagc aaccggcttt tagctagaag ggcagggggc tgggtgcagg ttatgctggg 1260
 ccagcaaaga ggcccgggat ccccctccca tgcacctgct gatgggcaa ggccacccca 1320
 ccccaccccc ttocttaca gtgttcagca ccctcccatc ccactcac aaacctggcc 1380
 ctctgccctc ctaccagaag aatggatccc ctgtgggagg ggcagggga cctgttcca 1440
 ccgtgtgcc aagacctctt ttcccacttt ttccctcttc ttgactcacc ctgccctcaa 1500
 tatccccgg cgcagccagt gaaaggagc coctggctcc tggctcgcct gcacgtcca 1560
 gggcggggag ggacttccgc cctcaogtcc cgctctcgc ccaggctgg atggaatgaa 1620
 aggcacactg tctctctccc taggcagcac agcccacagg tttccaggag tgcctttgtg 1680
 ggaggcctct gggccccac cagccatcct gtccctccgc tggggccca gcccgagag 1740
 agccgctggt gcacacaggg ccgggattgt ctgccctaat tatcaggctc aggctacagg 1800
 gctgcaggac atcgtgacct tccgtgcaga aacctcccc tccccctcaa gccgcctccc 1860
 gagctcctt cctctccagg ccccagtgcc ccagtgccca gtgccagcc caggcctcgg 1920
 tcccagagat gccaggagcc a 1941

<210> 28
 <211> 1433
 <212> ADN

ES 2 679 291 T3

<213> Homo sapiens

<400> 28

```

atctatTTTT tcaccatcct tgtgactcag ctttttctta ttcctttaat tctttgcata      60
gatctttcag cagagatgac agaagacgag acacccaaga aaacaactcc cattctcaga      120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat      180
cctgtagggg gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg      240
gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gactgagct      300
ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat      360
cacgcttggg tgggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggta      420
acggctcctg ttgcaggaat agctgagtca gaggtggggc ttcctccact cccccaccc      480
acccttttca ccattagggg ctttcttgcc ttgtcttgc tactctgctc tgggtggtca      540
ttgtgaaaag cccgcaccaa ccatgccagt ggagccaga cgaggacaca gcctggctct      600
gggtcccagc aggaaaggca atcccagaaa ggaggggtca gggactggag tcctgtgggt      660
gctttttaag caaagattat caccaggcag gctaaactta gcaaccggct tttagctaga      720
agggcagggg gctggtgtca ggttatgctg ggccagcaaa gagggccggg atccccctcc      780
catgcacctg ctgatgggcc aaggccaccc caccocaccc ccttccttac aagtgttcag      840
caccctccca tcccacactc acaaacctgg ccctctgccc tctaccaga agaatggatc      900
ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc ttttccact      960
ttttccctct tcttgactca ccctgccctc aatatcccc ggcgagcca gtgaaagga      1020
gtccctggct cctggctcgc ctgcacgtcc cagggcgggg agggacttcc gccctcacgt      1080
cccgtcttcc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc      1140
acagcccaca ggtttccagc agtgcctttg tgggaggcct ctgggcccc accagccatc      1200
ctgtcctccg cctggggccc cagcccggag agagccgctg gtgcacacag ggccgggatt      1260
gtctgcccta attatcaggt ccaggctaca gggctgcagg acatcgtgac cttccgtgca      1320
gaaacctccc cctccccctc aagccgctc ccgagcctcc ttcctctcca ggccccagt      1380
gcccagtgcc cagtgccag cccaggcctc ggtcccagag atgccaggag cca      1433

```

5

<210> 29

<211> 17130

<212> ADN

<213> Cricetulus griseus

10

<220>

<221> misc_feature

<222> (4674)..(4912)

<223> n es a, c, g, o t

15

<220>

<221> misc_feature

<222> (11347)..(13938)
 <223> n e s a, c, g, o t

5

<220>
 <221> misc_feature
 <222> (14757)..(14870)
 <223> n e s a, c, g, o t

<400> 29
 gtgagtgggtg caggggtgact gggcctattg ggttcctggt cagtgaacac agtctgttct 60
 tcgagtgctg tgggactcct aacttactgc cgtgctggcc tcaactacagc catttaacta 120
 gagtgacttg agatcacttt ctatgtccac ttcccgggtcc taaaaaaaaa atttattttg 180
 tatacattag tattttgcct gtgtgatgt ccatgtgaag gtgccagatc tggaacttga 240
 gttacaggca gttgtgagct gccacatggg tgctgggaac tgaaccagc tcttctgcag 300
 gagcagcaag tgccatctct ctagcccccga gtaaaagccg atttctagtt gagacatggt 360
 ctttctgtgt agctctagct ggcttagaat ttgctaacta gaccaggctg gccccaaatt 420
 cagattcccc tgcctctgcc tctcatgtgc tggtaggaaa ggcgtgagtc acatgcttgg 480
 cactagacta gctttgtgtt tagagcaaaa ctaagtaagt agagttcccg tttgtccctg 540
 tcctctgccc catcatcttc ataccccttg gtactgtgat cttacaatct acagtacact 600
 ggatgaatga gtgtgagtga caatgaatga cagagcagta cccagagtcc acagttcaca 660
 ctgatactgg gttcactgta cagtggcaca tacttgtcac cacagtgtca tactggacat 720
 tttcatgacc ctaaggccac atcacccttg tgggtgcact cagttgtaat ttcaccctat 780
 atagtgattt caggttggct ttttacttag aatataacat ttaaattatg tttctttttt 840
 ggttttgagt tttgtttttc gagacagggg ttctctgtat tgctttggag gttgtgctgg 900
 aacttgctct ggagaccagg ctggcttcga actcacagag atcctcccac ctctgcctcc 960
 cgagtgctgg gattaaaggc gtgcgccacc aatgccccag catctttctg tttctctggt 1020
 tttaaagagg gagataatgt ttccagtgtc gtttttaact tttttcctgt aggctacat 1080
 agatgaattt gagcagaagc ttcgggcatg tcataccaga ggcatggatg gaatagagga 1140
 gcttgaact ggccaagaag gcagccagca agccctgtct gccaaagaaac cctctgctgg 1200
 tatgagagac aagccacac ggactcccca caaatcatcc ctgtcattta ttggtttttg 1260
 tttttaaaat cacatctccc tgccatttct agtctcaagg cgacagcctc tgacttctat 1320
 agattcagac aatgactttg tcacacctaa gcctcgacgc accaaccgtc atcgtccaaa 1380
 cactcagcaa cgaaagtcca agaagaaagc caaagttgtc ttttcaagtg atgagtcag 1440

10

ES 2 679 291 T3

tgaggatggt atgactcgcc ctccccgcag ctgcttgctt gccttggggt ctccctgggg 1500
 cttgcctggt tgtagagtca ttcttcctgg taatgtgttc cccactgtgc acgttggctc 1560
 acctttgcag ctctatgttt taagatccac tttacaccat ctttgtgact cagcttttgt 1620
 tcctataatt ctttgaacag aactttcagc ggaaatgacg gaagaagaga ctcccaagag 1680
 aaccaccccc atccgcagat catctgcccg aagacacagg tcttaggagt tgccctcacct 1740
 attcccatcc ctgctgtgca gggatatctta attgattttt aattcatttc ctcccttgta 1800
 aaaatgttgc ctccctgttt ttaatatgta aatggtcttt gcattcaagg cctttctcct 1860
 agagcccatc tgactgctgc tttcactttc caacaaacac ttaaccaatc accatgtttt 1920
 ctaagaataa atgtttttat atacttttat tctttgctgg tctgtttcat ctccacctt 1980
 agcccagatt caggcagcaa tagagggtta tcaaaccaga gagaagatac accagcccag 2040
 gcatcctgag gctgtctctg ggtcttcatg tttcctcatt gattcatctt gcacaactga 2100
 tgagactcca ctgtagagaa ttctcacttc tcagcctggg ggtactttct acccagagac 2160
 ctatttctgt catagccttt tgggacacat aaagcttctt tcttgcatag aacaccccac 2220
 aacagtgtat caggagtgtg caagttgaca aacaatgctc tagacctacg gttttcttcc 2280
 tctgtgtgcc tcatcccaga agagatcatg actcccagga gtcagccttt actatggggt 2340
 ctgcaggggc gtccagcccc tcagcggcaa gccatgcca ccctcccaa gtccttaatc 2400
 tgctgagtca cttggaacag gagacactga ttctgctgtc atgacaacag cacattgcca 2460
 tagaaatgct ccctacctct tacgtgtgtg gtgggggaac agtaatgaca aaccataggc 2520
 aggaggcaaa aggggaagacg gcacctcaga aacatgtgtt aggttagggc agaactatgg 2580
 aggggctcct gagactcttt gatgggaaa ggtaatgct gtcctgaaa cctctgttgg 2640
 aaggcagaaa agggacaggg ctgagtcccc gcactgggac catttccatc ctctgcatcc 2700
 tgccccggc tcatggaaag cctgggcatg ggccacacag ctgtcagtct tggctctggg 2760
 gccccaaagga ggtagggcaa tcccagaatg gcaaggagcc aggactggat ttgggggtgca 2820
 gcccagcctg ctccctgcct ttaagcaaa ggttatcacc aggccagcta aacttagcaa 2880
 ttaggctcct cagctaaaag agcagggggc tggctcaag ttgcaactgac ctagcaaaga 2940
 ggccccagga tccccctgcc cagcacctgt ggctgagctc ccaagccctt cccgagagct 3000
 caggatccac cctttccacc ctccctactc ttcagaggag gaacccccct tctccttccc 3060
 acttgttgga ggggctggg gccaggctgt tctggcttgg ggtataatac ccctacccc 3120
 ttctactttc ccctcctctc agacctcacc ctgcctccac gagggcagcc aagaaaggag 3180
 agtccctggc tgcagggcca gtaggcacgt cccagagcgg ggaggactt ccgccctcac 3240
 gtcagctct ccgccctggg gctgcagtgg gtgaaagggg cagtgtctcc tagcctgggc 3300
 ggtgcaacc tcaggttccg aggaggaacg ctctgggagg cttctttgcc tcctccaacc 3360

ES 2 679 291 T3

caaccacaaa ccaggacatt gtctcacc cggggcccca acctagacct taactgagga 3420
acacagagggc cagtttghtaa gtctcaatta tgcagggcat cccgacctgt ggcgtaggga 3480
gccccctcc agggcgcttc cctagcctcc tcttgccct caccagcccag gcctctggcc 3540
caagaaatgg aagtgggggt gggggatgga actgcgaatg cgaagggccc ccgagggagg 3600
caaagtgacc cctcccgggc cttttctgct ccgagacttg tttttgctg tgtcactacc 3660
gaagaaccac gagaagatcc tcaacttttc cacagccttt gcataaaggg gagagggctg 3720
gcggtgcagc tgtggcacac acgcaacttc gctcaaccgg ccccccccg cccccgttc 3780
tgctcctcc caggttctcc ccattttatc ggggcggcaa cttttaggtc cctgggtcct 3840
ggaagtcctt agtacacact cttcgtcctt aagtccatag tctgtattcc ctcggtccta 3900
tcctgtcccc catcaccggg tcacctcccc agcgaagcaa tctcagttcc cctccccctc 3960
tcagccccga gcccacacgt ttggtgcgtg cacatttcaa aaacgaggcg ggtccaaaga 4020
gagggggtgg ggaggtgccg agtgcccag ctactcgcgg ctttacgggt gcacgtagct 4080
caggcctcag cgcccttgag ctgtgactgg atggatgagc gggcggggag gcggggcgag 4140
cgtcctggc gctccccacc accccagttc ctataaatac ggactgcagc cctccccggt 4200
gctctctgct cctccctggt ctagagacag ccgcatcttt ccgtgcagtg ccaggtgaaa 4260
accgcagagt gggccgcagg tggccgggga cggtcggaaa cggggaaggg gggcgctcag 4320
cccgggactg cgggcgctgg ggcgagctcc actgcccag cccgggctcc gcattgcaga 4380
ggctggaggg ggaagggatg gggggcgcgg cggggggcac gctgacctct ctttctctc 4440
tctccctccg cagcctcgct ccggagacgc aatggtgaa gtcggcgtga acgggtgagt 4500
tcgtggctgg gctaggggtg ggctccgggt cccgctccgt cgcgtatgca ggtctacccc 4560
accccggggc tctgcgggag cgtgggggtg ccggtgggtg gccgcagcac ccaaggagac 4620
ctcaaggtca gcgagccacc tcctcccttg cggggatgag cagcgcggag tccnnnnnnn 4680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnaggtcggt 4920
ggcaccgggc cgtgcggtgc ccgctttagc gcatccatca tctccaagg gcttccttta 4980
gggtggctgg ccgcccgat gttgcaaacg ggaaggaaat gaatgaacca ccgttaggaa 5040
acctcccttc ggccttcctc cttcctagcc cgtgactaac ctccccactc cctccccggg 5100
tggagtgcgc tctgtactgt aagccaggtg atgcaaggct tccgtgctct cgagagagct 5160
ctacctgcc agctgtctca tattattagc ctcaaagcag ccctcaagcc tcatttacct 5220

ES 2 679 291 T3

tgagcatatg atatatdddg tagattctct gagaatcgaa gcggacttgg agaggtctgc 5280
 ttgtccttct cccagcccaa aggtggtagc tatggcgtag cgccggaggg gggagtgggg 5340
 ggggagctga gtcattggtg ttctgaaaag aaaatttcca ccacaaaatg gctccggtgc 5400
 tagcatcccc ttcccccat aacctctgct tcccatcaca cctgaccca aacctgtag 5460
 gccagactgt aaaggtcact aagaggattg agtgtctgag cctcggaacc ctgcccttct 5520
 ccccatccca tcctctggaa accagatctc ccccgctcca ccctaactctg aggttatatt 5580
 tagcggctg acctttcagt atttggggtc tggggcccta cacacatctg ttgctcctgc 5640
 tcctgatttt tagctagcaa attcaagtgc tttgcaaatt agagcccagg gattaggggt 5700
 tggaaagctc agtggttttc tcagtctttc cctttagggg gagggacttg gaggaagcag 5760
 gtgggccgac ccctgtccta ctcatctga cctttaacct tggcctttga gcttgatgat 5820
 gctgagtgca cgagttcttc ctgtccaggg ggtgtagcct gaagccaggc caggctagaa 5880
 caaacttccc agggggtggg ggtagtgaat gccttgtgcc cacacagggg cacactgcca 5940
 cctctggag acttgaaatg actggtgggg gggttggaca aggctttgag cccaatcacc 6000
 tcttgacag gaaagtaacc cccactttat ggccctgctg taaaagccca gtcaaacctc 6060
 atttgtccaa ggaagataga cctctgggg ctccctaagg ataggggtgt tctatatttg 6120
 ggccctgctt ctaagcattc agccagcttt attaaaggaa attcataaca aaacttgaat 6180
 ttcctgcttc ttaaatacta atagtgtgct ggatctccat taaaaatgct gtcttgacaa 6240
 gtaggctatg gtttctgtgg gctctctaca gctatgggac aactggattc tgttttctga 6300
 agggcatgtg tcagctcagt actgactata gacctatgag ttctctgacc ccctaactca 6360
 cctttttttt tcttgctca gatttgccg tattggacgc ctggttacca gggctgcctt 6420
 cacttctggc aaagtggaag ttgttgccat caatgacccc tcattgacc tcaactacat 6480
 ggtctacatg ttccagtatg actctaccca tggcaagtcc aaaggcacag tcaaggctga 6540
 gaatggaaag cttgtcatca acgggaaggc catcaccatc ttocaggagc gagatcccgc 6600
 caacatcaaa tggggtgatg ctggcgccga gtatgtgtg gaatctactg gcgtcttcac 6660
 caccatggag aaggtgggg cccacttgaa gggcggggcc aagagggtca tcatctccgc 6720
 ccctctgct gatgccccca tgtttgtgat ggggtggaac caagacaagt atgacaactc 6780
 cctcaagatt gtcaggtgag gatgagcag ggtgtggca aagtgggcaa gcaggggcaa 6840
 ggttacaggt gggcgagcct cctaacctgt ctcttctctt cagcaatgag tcctgcacca 6900
 ccaactgctt agccccctg gccaaagtca tccatgacaa ctttggcatt gtggaaggac 6960
 tcatggtatg tagttcatct gtttcatcct gccagcagtg ggcgctgtgg tggggccct 7020
 gcaagacctc actccctgcc tctgtgtctt tcagaccacg gtccatgcca tcaactgccc 7080
 ccagaagact gtggatggcc cctccgggaa gctgtggcgt gatggccgtg gggctgcccc 7140

ES 2 679 291 T3

gaacatcacc cctgcatcca ctggcgctgc caaggctgtg ggcaaagtca gccagagct 7200
 gaacgggaag ctgactggca tggccttccg tgttcctacc cccaacgtgt ccgttggtga 7260
 tctgacatgt cgcttgaga aacctgtatg tctgggggtg gctgagggtt gtctctagtg 7320
 gtgaggttg ggcttgagta gtcacottga tttttgccct taataggcca agtatgagga 7380
 catcaagaag gtggtgaagc aggcattga gggcccactg aagggcatcc tgggctacac 7440
 cgaggaccag gttgtctcct gcgacttcaa cagtgactcc cactcttcca cctttgatgc 7500
 tggggctggc attgctctca atgacaactt tgtaaagctc atttctggt atgtgggatc 7560
 agaaacagtc ttttaaaagt aacttgggggt ttgtcaccoc ttggtgtcta actcaaaaag 7620
 taaacccttc ctaactgccc tcccaatgtg ttctaggtat gacaatgaat ttggctacag 7680
 caacagagtg gtggacctca tggcctacat ggcctccaag gagtaagaag cccaccctgg 7740
 accatccacc ccagcaagga ctgagcaag agggaggccc tggctgctga gcagtccctg 7800
 tccaataacc cccacaccga tcatctccct cacagtgtcc atcccagacc cccagaataa 7860
 ggaggggctt agggagccct actctcttga ataccatcaa taaagttcac tgcacccatc 7920
 ttctctggcc tttcaatgta aggttggggg agggaggctg tgacctagca aggttgggaa 7980
 tcctctgtgt cacttttcaa acagggcact agccacatgc cagcccaggt ttctgtcct 8040
 gaacagatga aaattcacct aaggtgtctt ggtgctggga ggagtggggg ttgacaatgg 8100
 gaccagtagt atggtcttag aggcttgggc tggactgcat caagttccag ggctgtgtgt 8160
 gtgttatctg caaaciaaag tcaattgtgt ctggaggcct taggtaaaat tgggaagatg 8220
 cccaacatag taaaaatgta tcagccaggg gaagtgacta cactgtatct aacctgaaac 8280
 agctgagctg taagccagca gctgtcacta tgttcagggtg tggccgctg gttctgggggt 8340
 ggtcacttgt atccagttg ttaggaagtg ttgtcattgc ttgttaggaa gacaacacat 8400
 ctgagctgg gcagtgttag tgctgcctc taatgccagc attcccagca cggaaagtggc 8460
 agaggcagga ggacagcctg gaataacaac ccagggcaga gccccatct cggaaaagac 8520
 aaaaaccaga aagtgcctaa aacattgaca caaagtgct caaatattcc ttctattgctt 8580
 ttagagattc cactgtcagc ttggcatggc ctctagttag acatcctgac ttggctccct 8640
 gctttccaag gtcaggagaa tgatagccac agaacgttcc ctcagctgat ggctggagaa 8700
 ccggggtccc tgagccccca cctcacacc catgtgcagg agggagcttc acctttcct 8760
 ccgagcagtg tctgccttca ggacaccgct ctcagcccag aactgagtc ttctttgtgt 8820
 gcatttcccc ggggaaaagc tacacattcc ctgtacagca ggcaggaggg cagctgtgag 8880
 ctcatggaca cgggacagaa ggccaggtac cctgcctccc tgcggtacct cgcctccctt 8940
 cgaggcctta ctaggaagca aagccttctc caggtgtacg tgatgcggaa agtctgcaga 9000

ES 2 679 291 T3

aagtcgggga ctcccagatg agtctgctgc acagtgcctg ggccagttag caccagaaca 9060
aaagtccaag atagccaagg agaacaggca ggtcccacaa tcacattcct taaaaacgga 9120
gaagggagga tcacaagtgt tctggcagcg tttcaaagat agagaagtag caaacagtca 9180
gagtgaagaa ggcaggagtg tcccaatgtc atggggatat ccaggaaggg tggggccaag 9240
cctagctcca gacaccagg gagcccgggc gaggagacag cccagggcc aggcttcatg 9300
cagtggcaca aaagaccccc accccggcct cggtgagctc cagggagcca ggcactggtc 9360
cctatctcat ggagctatca gatgagacat ggcaatcgga gtctcagtt tcgcttggcg 9420
ctggcggcgg gtcgctaagc gggaccgcag tgaaagcagg agactttcta gaaaaaaca 9480
ccagatgtca ccttggggca agcaggaagc ccaagcaggc ctgcatgctc gctctccttt 9540
cctatggcctt atcttctggg gtgcagctgg atgccaggca gggaaagtggg tgggaagact 9600
tgtccccatt cacctctatg aggaacagtt cccctcttgg aggctctgag aggctcagag 9660
atcaaaaatg tctcaagtca cagcatggag atgcttctctg agcccacagt gtgctgacac 9720
cagggccagc ttcttgtcca gctgtaaagg gggtggggct gtctcactat gttcccaagg 9780
ctgaccccca acacttggtg caaagcaatc ctcttgctcc agagtagctg ggatgacaag 9840
caaagccac taaacctagt gaacatgctg cactctgtgc ctacttaat gtaattgggg 9900
gagtgggttg gggttttggt ttgagacagg gtctcagtat gcattcctgg ctggcctgga 9960
attcacagag gagagaccgg cctggtactt gtggcaatct gtagggagac accgtaacca 10020
cgcctcctaa gagctggctc atcagggtgt ggtcaagggt atgtcagggg gacatgggag 10080
agggtgaaag ggataaggaa cgcacatggt agctcttttc cctgctgctt tggcttgctc 10140
tgctttcctg ccggttggct acctaataaa tatacctga ccatcacggg ctatgtatta 10200
gtaattaatc ctaatagccg gtgtctccct acaacaatcc acctgcctgt gacacctatg 10260
tgcttggatg acagatgtac accaccacac ctggcttact gtggttcctt agattttacc 10320
tgttgttctg ggttgtttta ttttaagaca agtttttctt attgtagtcc aggatagcct 10380
caaaactcac gcagctctttt gcttcagcct cccaaggctg ggcttgcaga catgtacat 10440
catgcctggg ttacacattt ttaatatata atttattttt attctgtgtt aattggtatt 10500
tgcctgcata tatgtctgtg taaaggtatc agaagccctg gaactggagt tacagacact 10560
cgtgtgggtg ctgggaattg aaccaggcc ctctggaaga gcatccagtg ttcttaacta 10620
atgagccatc tctcagcccc caggtttttt tttttttttt ttactctgt gtgtatgtgt 10680
gtgtgtgcac acgcacgcgc atgctcgaac atacaccagc acacagggtc tcacacaggc 10740
tagacaagta ctctaccact gagctatatt ccccaaacc aaatttactc ttacacaagg 10800
tcaataactt aaatcatgac ttagactttg gataaaaacc tgcccctgta tctcttcatt 10860
cctagccagg tttcggctct cagacactgg ggcactggta ggagtcgtct gaccaccagg 10920

ES 2 679 291 T3

gcagggagag gaagctatca atctgggcca cacctatagg tgactccaaa ggacagctcc 10980
acttgcagat gtgacctggt caccggccac tgggctggga cgagggccaa tcagagcagg 11040
tccttaccgg tcccccgact gcgtgatcag ggcaggtcct taccggctcc ccgactgtgt 11100
gatgaggcgg cggcaggtct ccatctgcac gtccaagccc cgtttcatgc tgcacatctc 11160
catgtactcg tgcaggtggc ggttcatgtc attcttggct gtggccagct ccagctgcag 11220
aaaagtaggg cagggcagga tgaggagggt agcaagccca aggcagacag gggattagag 11280
cccagctggg cccccggagc caggcagcca tggttaaaag ctcggccttc aaaagaaagc 11340
aagcagnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 11940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12780

ES 2 679 291 T3

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12840
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12900
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12960
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13020
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13080
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13140
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13200
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13260
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13320
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13380
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13440
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13500
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13560
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13620
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13680
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13740
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13800
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13860
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13920
 nnnnnnnnnn nnnnnnnnnnac acacacacac acacacacac acacacttac ttctgagga 13980
 agactgcagg gaacaaggct agatagaacc cattattcca ggagacaaaa gtgaagcaaa 14040
 aatggtctct taaagcacia agcagcacta ccaacagtgg tctcttcag cctatatatc 14100
 accctagttt acacattgaa gggaaagatg tacattctag aacattttca acaagtgcgc 14160
 ccccaaattc cagactcagc atgggagatc agagcagtag aggtagatgg aggagataga 14220
 ggggccaggg ccatggagga aaactcccac ctgtcatcag ctttaccatc caataatggg 14280
 cagctctacc tcagtaacac tgagagatta agggccatac agacaaaagg aacttttgcc 14340
 acaaagctac caagtgaact ttctttacct ttctcctcaa gtcattcttc caggaaggac 14400
 agaagtcaca caacttagaa acaactttcc atccatgtgc agtatacata tttttatgtg 14460
 tgtggataca gatgtgcatg tgtgtggtta catggggaag cctgatgtcg atgcaggaa 14520
 tcttcctagc tggttcttat actctgagtc tgtgtaacgg cattggtgtg cctggaactt 14580
 gatctgtaga ccagactggc ctcaaactca cagagatcta cctgtctctg cctcccgaat 14640
 gctgggatta aaagcgtgtg cctcagccgg acgttgggtg cgcattgcctt taatcccagc 14700

ES 2 679 291 T3

actcgggagg cagaggcagg aggatctctg tgagttctag accagcctgg tctacannnn 14760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 14820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ctagttccag 14880
ggcagcctcc aaagccacag agaaacctg tctcaaaaaa caaaacaaaa ataaataaat 14940
aaataaataa aatcctgcgc ctccactgcc cggctcagaa tattttttaa gacaaggatt 15000
atgtagccca ggctggcctc aaacttgta tgtagctaag atgatggggtg cctctgatcc 15060
tcctgcctct acctcccaag tgtaggact ataggtggc actaccacac ccggtttcat 15120
gtgggctagg atcacactca ggcagggtt catgcacatt ggcaggcat tctaccaact 15180
gagttacaca gtcccagccc aggatgtctc accagatcac ggttacagct aggcacctgt 15240
ggctctgaggt ttgcaccaca tggcagcccc agtactgag ctaagggtgaa actcatactt 15300
gactaaaggc acggaatctg aaaggggtgg gtacaaccca ggaggaaaat cggcggaagc 15360
aataagaggc gtcaacaggc aatggtcact tttcccaga caagtctcag ctggaagcag 15420
cacacgggct gggatctga aactcagaag tttcctgggt taaaacaaa agccctgggt 15480
tttcttagga tagtatgttc atttgatgtg tgctgggaca ctgagaacac aggaaaataa 15540
cttccaaagt ttgttctcaa aatgcagaca ggggtgaacct ttacagggt tgcagggagg 15600
ctgatacaag aggtctctg cctcattcac ccatcagagt ccaagacaca gtcctcacac 15660
agccccacc aaccaactgt cctccgacgg tgacctctc cctacaaacc tgcctgcagc 15720
ccagaaggac agtcagagcc ccagctccat agccattgt aaaggactcc ccaggctcaa 15780
aaacaaaagg gagactcctc caggagctca caggctgagc caaggaggca ggagtgtctc 15840
ggcacttagg gtgtccgggt gataatgcac aatatgtatg atgtgtcaac ctgccagatc 15900
ctgtcctaag cgccacgtgt ctatctgatc atcctaataa cccagcctgc ggggagagac 15960
atgcacatct cttctgaca aataaagagg caggccaag ctagcatgg cagtacacat 16020
ctgtaaccca gcaactcaaca ggcagagcag gaggattgac tgcctcctat tccaggccag 16080
tttggtaca cagaaagacc ctgtactgat gctaagagag atggctcaga gttaagagca 16140
cttctgctc ctggagagga cccaggttt ctttccagca cccatgtcag gctcacagtc 16200
gcctgtgaac accagctcca ggggatctga tggcctctc tggcctccag gggcacgcgc 16260
gtgtgcgctc acacacacac acacacacac acacacacac acacacacac acacacacac 16320
acagagtgca aagctgaagt gcagctcaca ccataatcc tagcaaatg agacaaaagg 16380
accaggagtt cgaggccaac ctgactaca agacaagtac ttgtaaataa atacatatat 16440
acgaaaataa acaaaccaaa aataatgaaa caggcactca gatcaaggga gccagtattg 16500
gtggcacacg cctttaatcc cagcactcag gaggcagagg cagggtgatc tctgagttt 16560

ES 2 679 291 T3

agccagcct agtctacaga atgagatcca ggactacaca gagaaaccct gtctcacct 16620
 cctctcccct aaaaaaatt aagaaagaaa agaaagagcc aggggctggt aagcaatgga 16680
 tcagcaggta aaggcacaag ccaccaagcc caacaacct aatttgatcc ctgggaccta 16740
 catggtagaa ggagagaact cactgcaagc tgtcctctga cctccacaca cacaccatgg 16800
 caggcacatg ctctccccgc cacatacata catacataca tacatacata catacataca 16860
 tacatacatg taaacagaaa agaaagtagg aagacaaaag gttcattcta agtggtgaca 16920
 gagccaagat ttgaaccac atctatctgc gtctagatcc tgaggctca agggcccatt 16980
 agagctacca aaactggggt ggggtgcaact cctcctgtca acaggtgaa ggtcctggtc 17040
 ttgggctctg gctgggagga ggccagccag cagcacacct ccagcctggt ttgtgagccc 17100
 tgggaacaga gccctgcc aacccttac 17130

5

<210> 30
 <211> 3020
 <212> ADN
 <213> Artificial

10

<220>
 <223> amplicón

<400> 30
 tacggccggc ttcactgtac agtggcacat acttgtcacc acagtgtcat actggacatt 60
 ttcattgacc taaggccaca taccctctgt ggtgtcactc agttgtaatt tcaccctata 120
 tagtgatttc aggttggtt tttacttaga atataacatt taaattatgt ttcttttttg 180
 gttttgagtt ttgtttttcg agacagggtt tctctgtatt gctttggagg ttgtgctgga 240
 acttgctctg gagaccaggc tggcttcgaa ctcacagaga tcctcccacc tctgcctccc 300
 gagtgctggg attaaaggcg tgcgccacca atgcccagc atctttctgt ttctctgttt 360
 ttaaagaggg agataatggt tccagtctg tttttaactt ttttcctgta ggctaccata 420
 gatgaatttg agcagaagct tcgggcatgt cataccagag gcatggatgg aatagaggag 480
 cttgaaactg gccaagaagg cagccagcaa gccctgtctg ccaagaaacc ctctgctggt 540
 atgagagaca agcccacacg gactccccac aaatcatccc tgtcatttat tggtttttgt 600
 ttttaaaatc acatctccct gccatttcta gtctcaaggc gacagcctct gacttctata 660
 gattcagaca atgactttgt cacacctaag cctcgaagca ccaaccgtca tcgtccaaac 720
 actcagcaac gaaagtccaa gaagaaagcc aaagtgtct tttcaagtga tgagtccagt 780
 gaggatggta tgactcgccc tcccggcagc tgcttgcttg ccttggggtc tccttggggc 840
 ttgcctggtt gtagagtcatt tcttcctggt aatgtgttcc cactgtgca cgttggtcca 900
 cctttgcagc tctatgtttt aagatccact ttacaccatc tttgtgactc agcttttgtt 960
 cctataattc tttgaacaga actttcagcg gaaatgacgg aagaagagac tcccaagaga 1020

ES 2 679 291 T3

accaccccca tccgcagatc atctgcccga agacacaggt ctaggagtt gcctcaccta 1080
 ttcccatccc tgctgtgcag ggtatcttaa ttgattttta attcatttcc tcccttgtaa 1140
 aatggtgccc tccctgtttt taatatgtaa atggctcttg cattcaaggc ctttctccta 1200
 gagcccatct gactgctgct ttcactttcc aacaaacact taaccaatca ccatgttttc 1260
 taagaataaa tgtttttata tacttttatt ctttctggtt ctgtttcatc ttccacctta 1320
 gccagattc aggcagcaat agagggttat caaacagag agaagataca ccagcccagg 1380
 catcctgagg ctgtctctgg gtcttcatgt ttcctcattg attcatcttg cacaactgat 1440
 gagactccac tgtagagaat tctcacttct cagcctgggg gtactttcta cccagagacc 1500
 tatttctgtc atagcctttt gggacacata aagcttcctt cctgcataga acaccccaca 1560
 acagtgtatc aggagtgtac aagttgacaa acaatgctct agacctacgg ttttcttct 1620
 ctgtgtgcct catcccagaa gagatcatga ctcccaggag tcagccttta ctatggggtc 1680
 tgcaggggag tccagcccct cagcggcaag ccatgccac cctcccgaag tccttaatct 1740
 gctgagtcac ttggaacagg agacactgat tctgctgtca tgacaacagc acattgcat 1800
 agaaatgctc cctacctctt acgtgtgtgg tgggggaaca gtaatgacaa acctataggca 1860
 ggaggcaaaa ggaagacgg cacctcagaa acatgtgtta ggttagggca gaactatgga 1920
 ggggctcctg agactctttg atgggaaagg gttaatgctg ctctgaaac ctctgttgg 1980
 aggcagaaaa gggacagggc tgagtccccg cactgggacc atttccatcc tctgcatcct 2040
 gccccggct catggaagc ctggcatgg gccacacagc tgcagtcct ggctctgggg 2100
 cccaaggag gtagggcaat cccagaatgg caaggagcca ggactggatt tggggtgcag 2160
 cccagcctgc tccctgcctt ttaagcaaag gttatcacca ggccagctaa acttagcaat 2220
 taggtctctc agctaaaaga gcagggggct ggtctcaagt tgcaactgacc tagcaaagag 2280
 gcccaggat cccctgccc agcacctgtg gctgagctcc caagccctc ccgagagctc 2340
 aggatccacc ctttccacc tccctactct tcagaggagg aaccccttt ctcttcca 2400
 cttgttgag gggctgggg ccaggctgtt ctggcttggg gtataatacc ccctaccct 2460
 tctactttcc cctcctctca gacctaccc tgcctccag agggcagcca agaaaggaga 2520
 gtcctggct gcagggccag taggcacgtc ccaggacggg gagggacttc cgccctcacg 2580
 tccagctctc cgccctgggg ctgcagtggg tgaaagggg agtgtctcct agcctggggc 2640
 gtgcaaccct caggttccga ggaggaacgc tctgggaggc ttctttgcct cctccaacc 2700
 aaccacaac caggacattg tctcaccctc ggggcccaca cctagacctt aactgaggaa 2760
 cacagaggcc agtttgtaag tctcaattat gcagggcatc ccgacctgtg gcgtagggag 2820
 cgccctcca ggccgctcc cttagcctcct cctggcctc acagcccagg cctctggccc 2880

ES 2 679 291 T3

aagaaatgga agtgggggtg ggggatggaa ctgcgaatgc gaagggcccc cgcaggaggc 2940
 aaagtgacct ctccccggcc ttttctgctc cgagacttgt ttttgcctgt gtcactaccg 3000
 aagaaccacg gccggcctga 3020

<210> 31
 <211> 3537
 <212> ADN
 <213> Artificial

<220>
 <223> amplicón

<400> 31
 tactcgcgat tcaactgtaca gtggcacata cttgtcacca cagtgtcata ctggacattt 60
 tcatgacctt aagccacat caccctgtg gtgtcactca gttgtaattt caccctatat 120
 agtgatttca ggttggcttt ttacttagaa tataacattt aaattatgtt tcttttttgg 180
 ttttgagttt tgtttttcga gacagggttt ctctgtattg ctttggaggt tgtgctggaa 240
 cttgctctgg agaccaggct ggcttcgaac tcacagagat cctcccacct ctgcctcccg 300
 agtgctggga ttaaaggcgt gcgccaccaa tgcccagca tctttctgtt tctctgtttt 360
 taaagagga gataatgttt ccagtgtgtt ttttaacttt tttcctgtag gctaccatag 420
 atgaatttga gcagaagctt cgggcatgtc ataccagagg catggatgga atagaggagc 480
 ttgaaactgg ccaagaaggc agccagcaag ccctgtctgc caagaaacct tctgtctgga 540
 tgagagacaa gccacacgg actccccaca aatcatccct gtcatttatt ggtttttgtt 600
 tttaaaatca catctccctg ccatttctag tctcaaggcg acagcctctg acttctatag 660
 attcagacaa tgactttgtc acacctaaag ctgcagcac caaccgtcat cgtccaaaca 720
 ctcagcaacg aaagtccaag aagaaagcca aagtgtctt ttcaagtgat gagtccagtg 780
 aggatggtat gactogccct cccggcagct gcttgcctgc cttggggctc ccttggggct 840
 tgcctggttg tagagtcatt cttcctggta atgtgttccc cactgtgcac gttggctcac 900
 ctttgcagct ctatgtttta agatccactt tacaccatct ttgtgactca gcttttgttc 960
 ctataattct ttgaacagaa ctttcagcgg aatgacgga agaagagact cccaagagaa 1020
 ccacccccat ccgcagatca tctgcccga gacacaggtc ttaggagttg cctcacctat 1080
 tccatccctt gctgtgcagg gtatcttaat tgatttttaa ttcatttcct cccttgtaaa 1140
 aatgttgctt ccttgttttt aatatgtaa tggcttttgc attcaaggcc tttctcctag 1200
 agcccatctg actgctgctt tcactttcca acaaacactt aaccaatcac catgttttct 1260
 aagaataaat gtttttatat acttttattc tttgctggtc tgtttcatct tccaccttag 1320
 cccagattca gccagcaata gagggttatc aaaccagaga gaagatacac cagcccaggc 1380
 atcctgaggc tgtctctggg tottcatggt tcctcattga ttcattctgc acaactgatg 1440

ES 2 679 291 T3

agactccact gtagagaatt ctcaacttctc agcctggggg tactttctac ccagagacct 1500
 atttctgtca tagccttttg ggacacataa agcttccttc ctgcatagaa caccaccaca 1560
 cagtgtatca ggagtgtaca agttgacaaa caatgctcta gacctacggt tttcttcctc 1620
 tgtgtgcctc atcccagaag agatcatgac tcccaggagt cagcctttac tatggggctc 1680
 gcaggggctc ccagcccctc agcggcaagc catgcccacc ctcccccaagt ccttaatctg 1740
 ctgagtcact tggaacagga gacactgatt ctgctgtcat gacaacagca cattgccata 1800
 gaaatgctcc ctacctctta cgtgtgtggt gggggaacag taatgacaaa ccataggcag 1860
 gaggcaaaag ggaagacggc acctcagaaa catgtgttag gttagggcag aactatggag 1920
 gggctcctga gactctttga tgggaaaggg ttaatgctgc tcctgaaacc tctgttgaa 1980
 ggagaaaag ggacagggct gagtccccgc actgggacca ttccatcct ctgcatcctg 2040
 cccccggctc atggaaagcc tgggcatggg ccacacagct gtcagtcttg gctctggggc 2100
 cccaaggagg tagggcaatc ccagaatggc aaggagccag gactggattt ggggtgcagc 2160
 ccagcctgct ccctgccttt taagcaaagg ttatcaccag gccagctaaa cttagcaatt 2220
 aggctcttca gctaaaagag cagggggctg gtctcaagtt gactgacct agcaaaggg 2280
 ccccaggatc ccctgcccc gacactgtgg ctgagctccc aagcccttcc cgagagctca 2340
 ggatccacc tttccaccct ccctactctt cagaggagga acccccttcc tccttcccac 2400
 ttgttgagg gggctggggc caggctgttc tggcttgggg tataataccc cctaccctt 2460
 ctactttccc ctctctcag acctcaccct gcctccacga gggcagccaa gaaaggagag 2520
 tccttggtg cagggccagt aggcacgtcc caggacgggg agggacttcc gccctcacgt 2580
 ccagctctcc gccctggggc tgcagtgggt gaaaggggca gtgtctccta gcctggggc 2640
 tgcaaccctc aggttccgag gaggaacgct ctgggaggt tctttgcctc ctccaacca 2700
 acccacaacc aggacatgt cctcaccctg gggcccaac ctagacctta actgaggaac 2760
 acagaggcca gtttgtaagt ctcaattatg cagggatcc cgacctgtgg cgtagggagc 2820
 gccctccag gcogcttccc tagcctctc ctggccctca cagcccagc ctctggcca 2880
 agaaatggaa gtgggggtgg gggatggaac tgcgaatgg aagggcccc gcaggaggca 2940
 aagtgacccc tcccgggcct tttctgctcc gagactgtt tttgcctgtg tcaactaccga 3000
 agaaccacga gaagatcctc aactttcca cagccttgc ataaagggga gaggtcggc 3060
 ggtgcagctg tggcacacac gcacttctgc tcaaccgcc cccccgcc cccgttctg 3120
 ttccttccca ggttctccc atttatcgg ggcggcaact ttaggtccc tgggtcctgg 3180
 aagtccttag tacacactct tcgtccttaa gtccatagtc tgtattccct cggctctatc 3240
 ctgtcccca tcaccgggtc acctcccag cgaagcaatc tcagttccc tcccctctc 3300

ES 2 679 291 T3

| | | |
|----|---|------|
| | agccccgagc ccacacgttt ggtgctgca catttcaaaa acgagggcggg tccaaagaga | 3360 |
| | gggggtgggg aggtgccgag tggcccagct actcggggct ttacgggtgc acgtagctca | 3420 |
| | ggcctcagcg cccttgagct gtgactggat ggatgagcgg gccggggaggc ggggagagcg | 3480 |
| | tcctcggcgc tccccaccac cccagttcct ataaatagca gagctcgttt agtgaac | 3537 |
| 5 | <210> 32 <211> 537 <212> ADN <213> Artificial | |
| 10 | <220> <223> amplicón | |
| | <400> 32 tactcgcgaa gaagatcctc aacttttcca cagcctttgc ataaagggga gagggtcggc | 60 |
| | ggtgcagctg tggcacacac gcacttctgc tcaaccggcc ccccccgcc cccgttctctg | 120 |
| | ttccttccca ggttctcccc attttatcgg ggcggcaact ttaggtccc tgggtcctgg | 180 |
| | aagtccttag tacacactct tcgtccttaa gtccatagtc tgtattccct cggtcctatc | 240 |
| | ctgtcccca tcaccgggtc acctccccag cgaagcaatc tcagttcccc tccccctctc | 300 |
| | agccccgagc ccacacgttt ggtgctgca catttcaaaa acgagggcggg tccaaagaga | 360 |
| | gggggtgggg aggtgccgag tggcccagct actcggggct ttacgggtgc acgtagctca | 420 |
| | ggcctcagcg cccttgagct gtgactggat ggatgagcgg gccggggaggc ggggagagcg | 480 |
| | tcctcggcgc tccccaccac cccagttcct ataaatagca gagctcgttt agtgaac | 537 |
| 15 | <210> 33 <211> 30 <212> ADN <213> Artificial | |
| 20 | <220> <223> Cebador GlnPr1896 | |
| | <400> 33 tacggccggc ttactgtac agtggcacat 30 | |
| 25 | <210> 34 <211> 30 <212> ADN <213> Artificial | |
| 30 | <220> <223> Cebador GlnPr1897 | |
| | <400> 34 tcaggccggc cgtggtctt cggtagtgac 30 | |
| 35 | <210> 35 <211> 35 <212> ADN <213> Artificial | |

<220>
 <223> Cebador GlnPr1901
 5 <400> 35
 tactcgcgaa gaagatcctc aactttcca cagcc 35
 <210> 36
 <211> 40
 10 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador GlnPr1902
 15 <400> 36
 gtcactaaa cgagctctgc tattatagg aactggggtg 40
 <210> 37
 <211> 40
 20 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador GlnPr1903
 25 <400> 37
 caccacagtt cctataaata gcagagctcg ttagtgaac 40
 30 <210> 38
 <211> 20
 <212> ADN
 <213> Artificial
 35 <220>
 <223> Cebador GlnPr1904
 <400> 38
 40 cgctagcacc ggtcgatcga 20
 <210> 39
 <211> 31
 <212> ADN
 <213> Artificial
 45 <220>
 <223> Cebador GlnPr1905
 <400> 39
 50 tactcgcgat tcactgtaca gtggcacata c 31

REIVINDICACIONES

1. Un casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica un polipéptido, y un elemento potenciador de la expresión en el que el elemento potenciador de la expresión
5 comprende una secuencia de ADN genómico no traducida aguas abajo de un promotor de gliceraldehído 3-fosfato deshidrogenasa (GAPDH) de mamífero, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico aguas abajo del promotor GAPDH de mamífero se inicia en una
10 región que abarca desde posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero es de aproximadamente 200 a aproximadamente 2800 nucleótidos, y en el que el casete de expresión comprende además una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH de mamífero, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero se
15 inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH de mamífero a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero es de aproximadamente 200 a aproximadamente 2800 nucleótidos.
2. Un casete de expresión que comprende un promotor, una secuencia de polinucleótidos que codifica
20 un polipéptido, y una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH de mamífero, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, y en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero se inicia en una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH de mamífero hasta la posición de nucleótido de
25 aproximadamente -3500, en el que la posición del nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, donde la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero es de 200 a aproximadamente 2800 nucleótidos, con la condición de que el casete de expresión no comprenda un promotor GAPDH de mamífero o fragmentos del mismo.
3. El casete de expresión de la reivindicación 2, en el que el casete de expresión comprende además
30 una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH de mamífero, en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero se inicia en una región que abarca desde la posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH
35 de mamífero es de aproximadamente 200 a aproximadamente 2800 nucleótidos.
4. El casete de expresión de una cualquiera de las reivindicaciones 1 a 3, en el que la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH de mamífero no está unida operativamente a la secuencia de polinucleótidos que codifica el polipéptido.
40
5. El casete de expresión de la reivindicación 1 o 3, en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero se inicia aguas abajo del sitio de poliadenilación de GAPDH de mamífero, y en el que la longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero es de al menos 100 nucleótidos y se extiende en su máximo al segundo último intrón
45 del gen IFF01.
6. El casete de expresión de la reivindicación 1 o 2, en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero es de al menos 100 nucleótidos y se extiende en su máximo al tercer último intrón del gen NCAPD2.
50
7. El casete de expresión de una cualquiera de las reivindicaciones 1 a 3, en el que la secuencia de ADN genómico no traducida aguas abajo y/o aguas arriba del promotor GAPDH de mamífero es de origen de roedor o humano.
- 55 8. El casete de expresión de la reivindicación 1 o 3, en el que la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero comprende la secuencia de nucleótidos seleccionada del grupo que consiste en: SEQ ID NOs: 8 y 21; una secuencia de nucleótidos al menos un 80 % idéntica a la SEQ ID NOs: 8 o 21; fragmentos de las mismas; o una secuencia de nucleótidos complementaria de las mismas.

9. El casete de expresión de la reivindicación 1 o 2, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero comprende una secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NO: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 que comprende cinco o menos modificaciones de ácidos nucleicos; una secuencia de nucleótidos al menos un 80 % idéntica a la
5 secuencia de nucleótidos seleccionada del grupo que consiste en SEQ ID NO: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 y 28 o fragmentos de las mismas; fragmentos de las mismas; o una secuencia de nucleótidos complementaria de las mismas.
10. El casete de expresión de una cualquiera de las reivindicaciones 1 a 3, en el que el promotor se
10 selecciona del grupo que consiste en el promotor SV40, promotor MPSV, CMV de ratón, tk humano, CMV humano, CMV de rata, EF1alfa humano, EF1alfa de hámster chino, GAPDH humana, promotores híbridos que incluyen promotor MYC, HYK y CX.
11. El casete de expresión de una cualquiera de las reivindicaciones 1 a 3, en el que el polipéptido se
15 selecciona del grupo que consiste en anticuerpos, fragmentos de anticuerpos o derivados de anticuerpos.
12. Un vector de expresión, que comprende en orden:
- 20 a) una secuencia de ADN genómico no traducida aguas arriba de un promotor GAPDH de mamífero, en el que la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero se inicia dentro de una región que abarca desde aproximadamente el extremo 5' del promotor GAPDH de mamífero a la posición de nucleótido de aproximadamente -3500, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la longitud de la secuencia de ADN genómico no traducida aguas arriba del promotor GAPDH de mamífero es de aproximadamente 200 a aproximadamente
25 2800 nucleótidos
- b) un promotor
- c) una secuencia de polinucleótidos que codifica un polipéptido
- d) un sitio de poliadenilación
- e) un potenciador
- 30 f) una secuencia de ADN genómico no traducida aguas abajo de un promotor GAPDH de mamífero, en el que el polipéptido codificado por la secuencia de polinucleótidos no es GAPDH, en el que la secuencia de ADN genómico aguas abajo del promotor GAPDH de mamífero se inicia en una región que abarca desde la posición de nucleótido de aproximadamente +1 a la posición de nucleótido de aproximadamente +7000, en el que la posición de nucleótido es relativa al inicio de la transcripción del ARNm de GAPDH, y en el que la
35 longitud de la secuencia de ADN genómico no traducida aguas abajo del promotor GAPDH de mamífero es de aproximadamente 200 a aproximadamente 2800 nucleótidos.
13. Una célula huésped que comprende un casete de expresión de una cualquiera de las reivindicaciones 1 a 11 o un vector de expresión de la reivindicación 12.
40
14. El casete de expresión de una cualquiera de las reivindicaciones 1 a 11 o el vector de expresión de la reivindicación 12 para su uso como un medicamento para el tratamiento de un trastorno o para su uso en terapia génica.
- 45 15. Un método *in vitro* para la expresión de un polipéptido, que comprende transfectar una célula huésped con el casete de expresión de una cualquiera de las reivindicaciones 1 a 11 o el vector de expresión de la reivindicación 12 y recuperar el polipéptido.
16. Uso de un casete de expresión de una cualquiera de las reivindicaciones 1 a 11 o el vector de
50 expresión de la reivindicación 12 para la expresión *in vitro* de un polipéptido heterólogo de una célula huésped de mamífero.

Figura 1

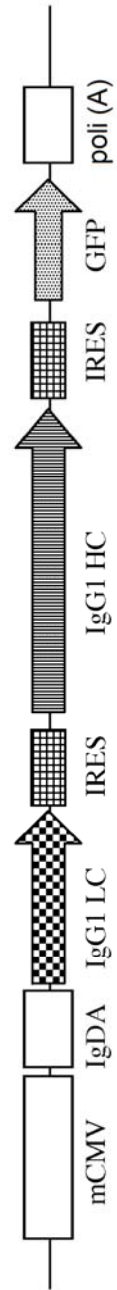


Figura 2

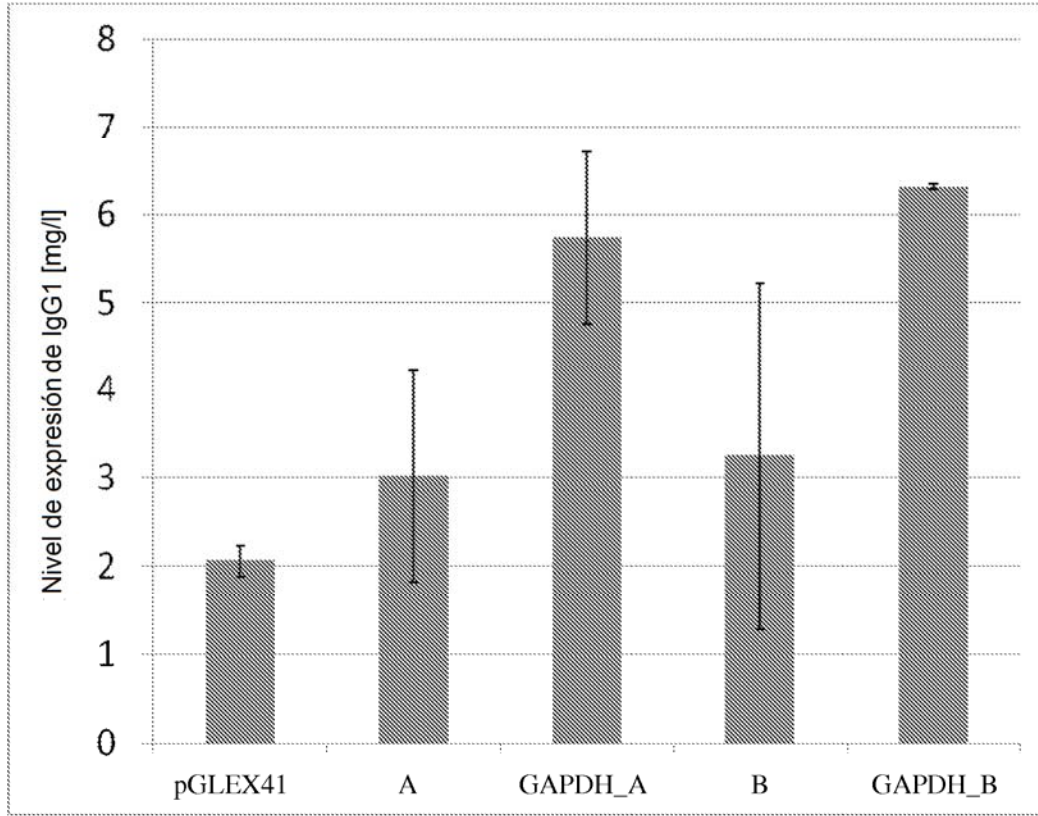


Figura 3

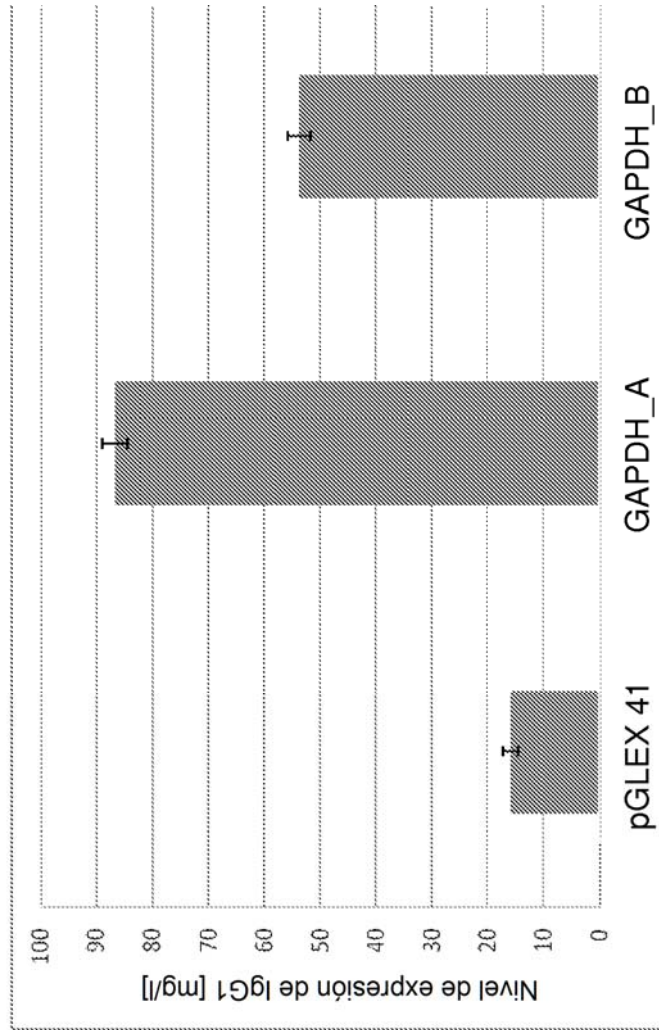


Figura 4

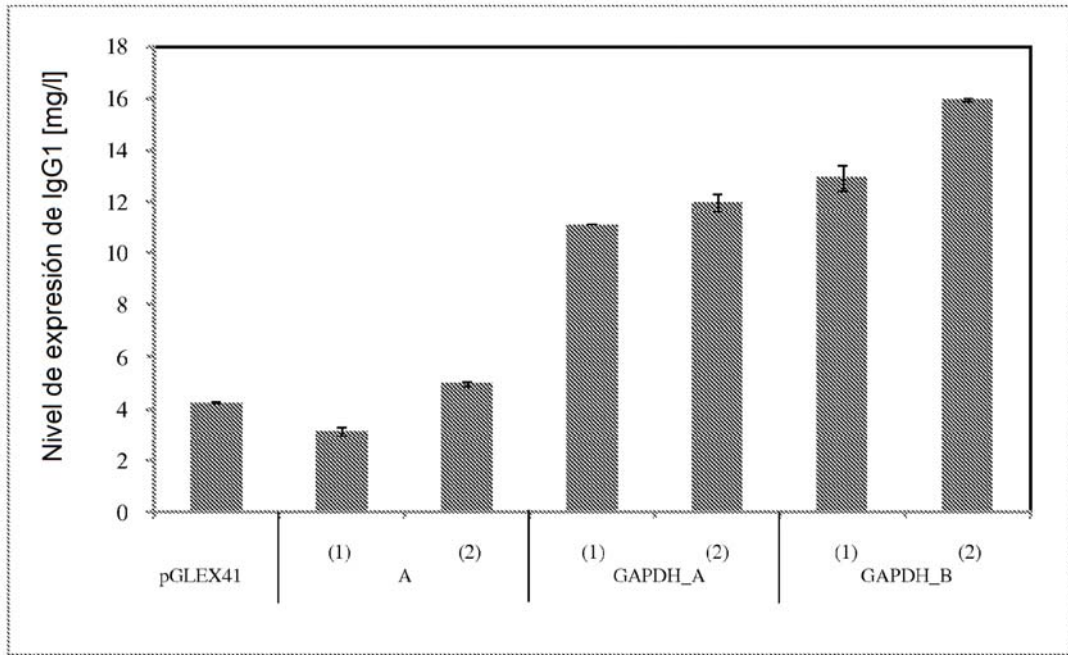


Figura 5

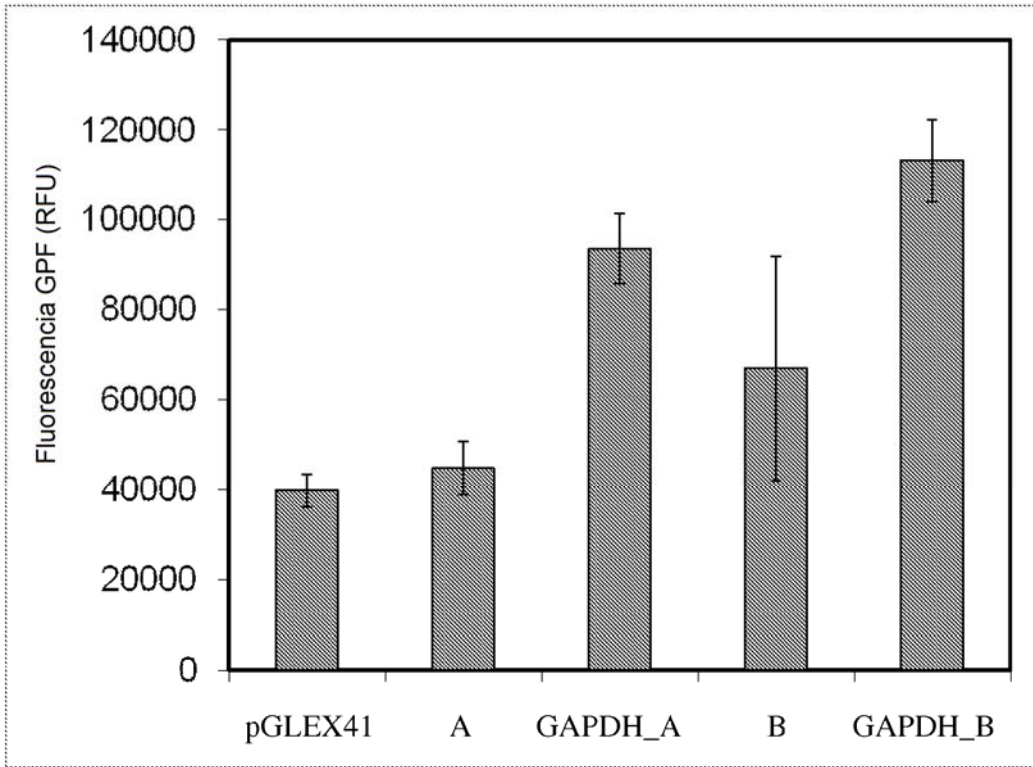


Figura 6

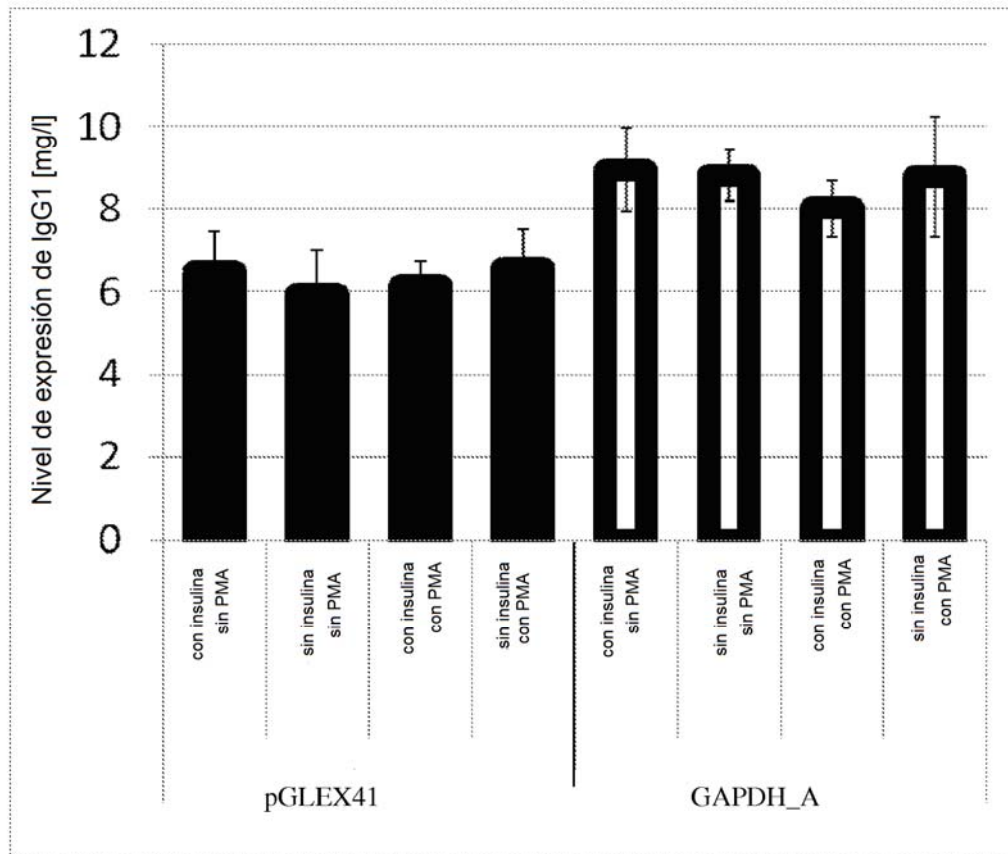
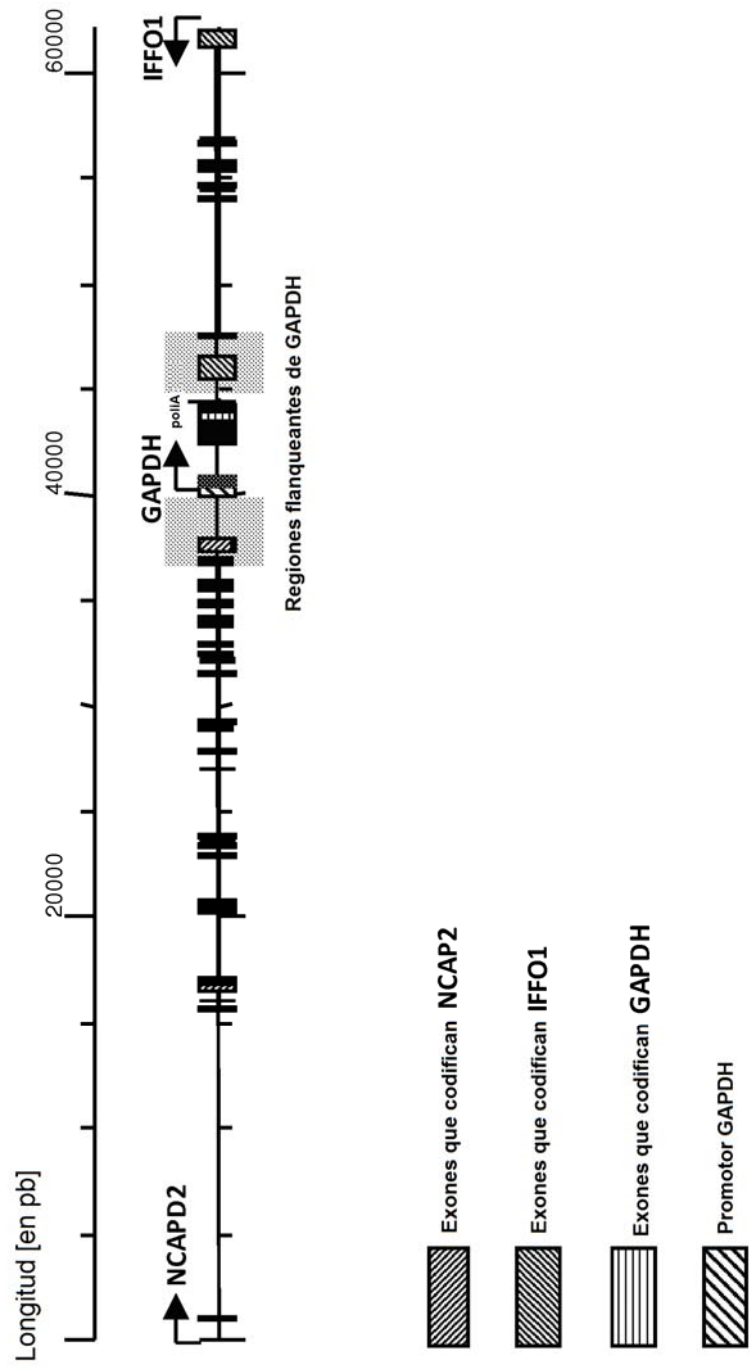
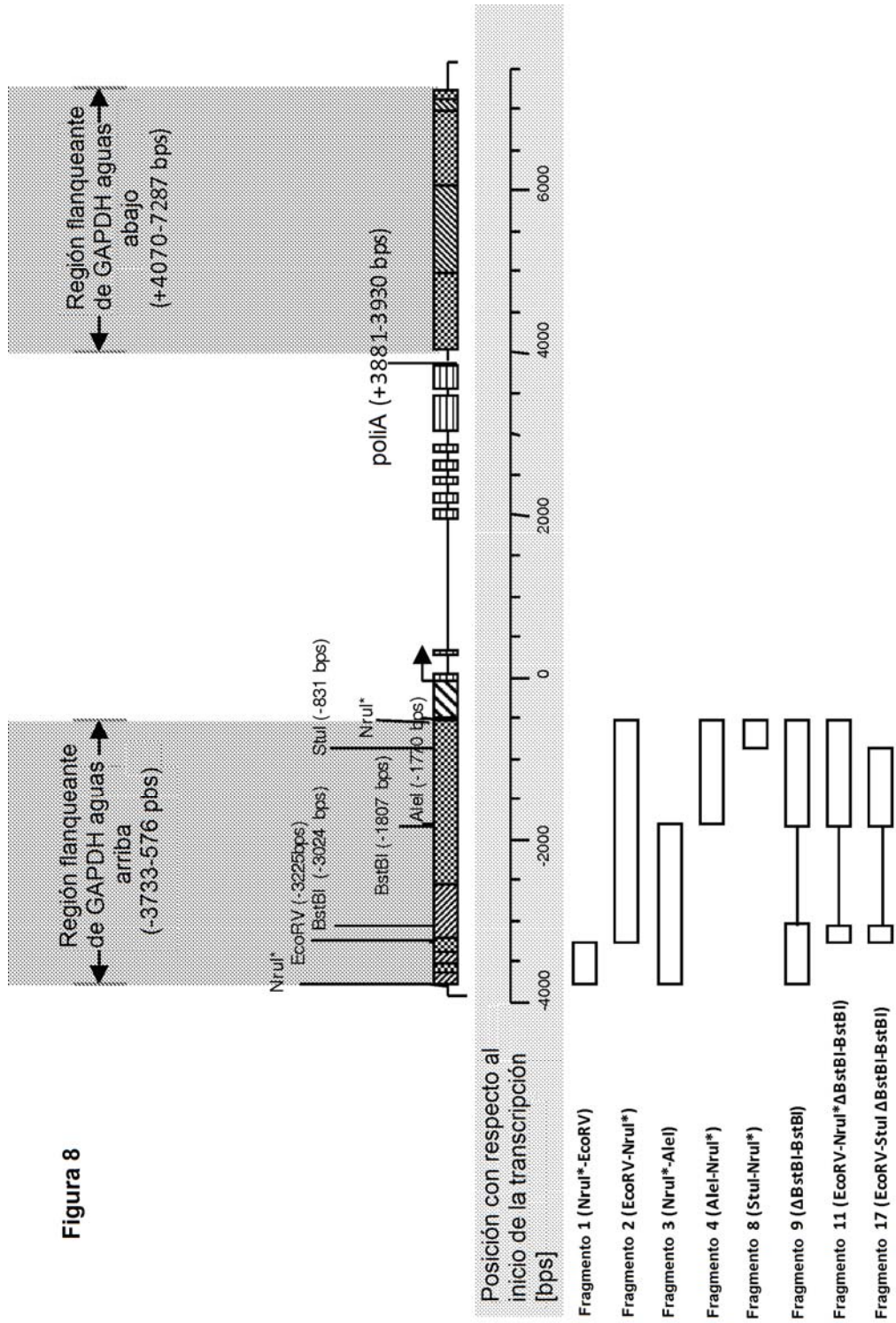


Figura 7





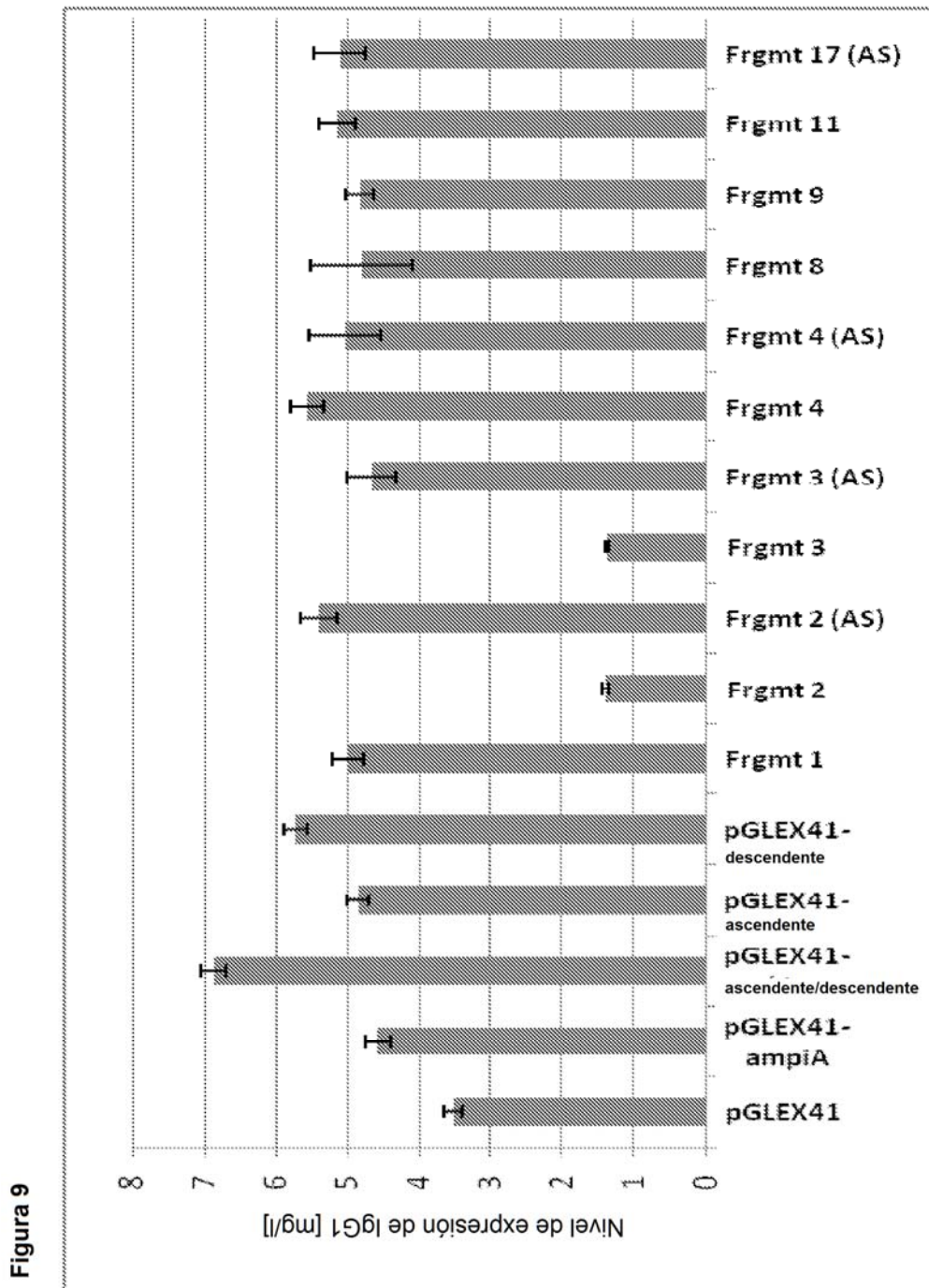


Figura 10

