



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 356 485**

51 Int. Cl.:
C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07728440 .4**

96 Fecha de presentación : **24.04.2007**

97 Número de publicación de la solicitud: **2013349**

97 Fecha de publicación de la solicitud: **14.01.2009**

54 Título: **Selección de células huésped que expresan proteína en altos niveles.**

30 Prioridad: **02.05.2006 US 416490**
02.05.2006 EP 06113354

45 Fecha de publicación de la mención BOPI:
08.04.2011

45 Fecha de la publicación del folleto de la patente:
08.04.2011

73 Titular/es: **CHROMAGENICS B.V.**
Archimedesweg 4
2333 CN Leiden, NL

72 Inventor/es: **Otte, Arie, Pieter;**
Van Blokland, Henricus, Johannes, Maria;
Kwaks, Theodorus, Hendrikus, Jacobus y
Sewalt, Richard, George, Antonius, Bernardus

74 Agente: **Arias Sanz, Juan**

ES 2 356 485 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

SELECCIÓN DE CÉLULAS HUÉSPED QUE EXPRESAN PROTEÍNA EN ALTOS NIVELES

5

ANTECEDENTES DE LA INVENCION

Campo de la invención: La invención se refiere al campo de la biología molecular y la biotecnología. Más específicamente, la presente invención se refiere a medios y métodos para mejorar la selección de células huésped que expresan proteínas en altos niveles.

10

Las proteínas pueden producirse en diversas células huésped para una amplia gama de aplicaciones en biología y biotecnología, por ejemplo como agentes biofarmacéuticos. Para este fin se prefieren células huésped eucariotas y particularmente de mamífero para la expresión de muchas proteínas, por ejemplo cuando tales proteínas tienen determinadas modificaciones postraduccionales tales como glicosilación. Los métodos para tal producción están bien establecidos, y generalmente implican la expresión en una célula huésped de un ácido nucleico (también denominado "transgén") que codifica para la proteína de interés. En general, el transgén se introduce junto con un gen marcador seleccionable en una célula precursora, se seleccionan las células para determinar la expresión del gen marcador seleccionable, y se identifican uno o más clones que expresan la proteína de interés en altos niveles, y se usan para la expresión de la proteína de interés.

15

20

25

30

Se conocen métodos para seleccionar células huésped recombinantes que expresan niveles relativamente altos de proteínas deseadas (véanse, por ejemplo las introducciones en los documentos WO 2006/048459 y US 2006/0141577).

35

Un concepto novedoso para seleccionar células huésped que expresan altos niveles de polipéptidos de interés se dio a conocer en la solicitud internacional PCT/EP2005/055794 (publicada como el documento WO 2006/048459), que se presentó antes pero se publicó tras la fecha de prioridad de la presente solicitud. Se dio a conocer una alternativa en la solicitud de

patente estadounidense n.º 11/359.953 (publicada como el documento US 2006/0141577) y en la solicitud internacional PCT/EP2007/051696, presentada también antes pero publicada tras la fecha de prioridad de la presente solicitud. En resumen,
5 estas solicitudes enseñan el uso de una secuencia que codifica para un polipéptido marcador seleccionable con un codón de iniciación distinto de ATG, por ejemplo un GTG o TTG. Esto dio como resultado la posibilidad de seleccionar clones con alta rigurosidad y se usó para obtener clones de células huésped con
10 niveles de expresión muy altos.

La presente invención tiene como objetivo proporcionar medios más mejorados y métodos para la selección de células huésped que expresan altos niveles de proteínas de interés.

15 BREVE SUMARIO DE LA INVENCION

Los documentos PCT/EP2005/055794 (WO 2006/048459), US 11/359.953 (US 2006/0141577) y PCT/EP2007/051696 enseñan el uso de una secuencia que codifica para un polipéptido marcador seleccionable con un codón de iniciación distinto de ATG, por
20 ejemplo un GTG o TTG. Esto dio como resultado la posibilidad de seleccionar clones con alta rigurosidad y se usó para obtener clones de células huésped con niveles de expresión muy altos.

La presente invención da a conocer genes marcadores seleccionables mejorados con un codón de iniciación GTG o TTG.
25 Tales genes marcadores seleccionables mejorados pueden usarse por ejemplo en las unidades de transcripción y métodos de uso de las mismas descritos en los documentos WO 2006/048459 y US 2006/0141577. Esto conduce a (una selección de) células huésped más mejoradas con altos niveles de expresión.

30 En un aspecto, la invención proporciona una molécula de ADN que comprende una secuencia de marco de lectura abierto que codifica para un polipéptido marcador seleccionable, comprendiendo dicha molécula de ADN en la cadena codificante una secuencia de inicio de la traducción para el polipéptido
35 marcador seleccionable elegida del grupo que consiste en: a) un codón de iniciación GTG; y b) un codón de iniciación TTG; y en la que se ha mutado la secuencia de marco de lectura abierto que

codifica para la proteína marcadora seleccionable para sustituir al menos la mitad de sus dinucleótidos de CpG en comparación con la secuencia nativa de marco de lectura abierto que codifica para la proteína marcadora seleccionable.

5 Según la invención, la proteína marcadora seleccionable proporciona resistencia frente a los efectos mortales y/o inhibitorios del crecimiento de un agente de selección, tal como un antibiótico. En particular, el polipéptido marcador seleccionable proporciona resistencia frente a zeocina o frente
10 a neomicina.

La invención proporciona además una molécula de ADN según la invención, en la que la secuencia de marco de lectura abierto que codifica para un polipéptido marcador seleccionable es parte de una unidad de transcripción multicistrónica que comprende
15 además una secuencia de marco de lectura abierto que codifica para un polipéptido de interés.

La invención proporciona además un casete de expresión que comprende tales moléculas de ADN, comprendiendo dicho casete de expresión un promotor en el sentido de 5' de dicha unidad de
20 transcripción multicistrónica y preferiblemente una secuencia de terminación de la transcripción en el sentido de 3' de la unidad de transcripción multicistrónica.

La invención proporciona además células huésped que comprenden una molécula de ADN o un casete de expresión según la
25 invención.

La invención proporciona además un método de expresión de un polipéptido de interés, que comprende cultivar una célula huésped que comprende el casete de expresión de la invención, y expresar el polipéptido de interés a partir del casete de
30 expresión.

LEYENDAS DE LAS FIGURAS

Figura 1. Resultados con un marcador de resistencia a zeocina con contenido de CpG reducido en células CHO-K1. Los
35 puntos indican puntos de datos individuales; las líneas indican los niveles de expresión promedio; el eje vertical indica señal de d2EGFP. Véase el ejemplo 1 para detalles.

Figura 2. Tal como en la figura 1, pero en este caso en células CHO-DG44. Véase el ejemplo 1 para detalles.

Figura 3. Resultados con marcador de resistencia a neomicina "pobre en CpG" que tiene diferentes mutaciones. Los puntos indican puntos de datos individuales; las líneas indican los niveles de expresión promedio; el eje vertical indica señal de d2EGFP. Véase el ejemplo 2 para detalles.

DESCRIPCIÓN DETALLADA DE LA INVENCION

10 La expresión "gen monocistrónico" se define como un gen que puede proporcionar una molécula de ARN que codifica para un polipéptido. Una "unidad de transcripción multicistrónica", también denominada gen multicistrónico, se define como un gen que puede proporcionar una molécula de ARN que codifica para al menos dos polipéptidos. La expresión "gen bicistrónico" se define como un gen que puede proporcionar una molécula de ARN que codifica para dos polipéptidos. Por tanto, un gen bicistrónico se encuentra abarcado dentro de la definición de un gen multicistrónico. Un "polipéptido" tal como se usa en el presente documento comprende al menos cinco aminoácidos unidos por enlaces peptídicos y por ejemplo puede ser una proteína o una parte, tal como una subunidad, de la misma. Puede comprender modificaciones postraduccionales, por ejemplo glicosilación. En general, los términos polipéptido y proteína se usan de manera intercambiable en el presente documento. Un "gen" o una "unidad de transcripción" tal como se usa en la presente invención puede comprender ADN cromosómico, ADNc, ADN artificial, combinaciones de los mismos y similares. "Operativamente unido" se refiere a una situación en la que los componentes descritos están en una relación que les permite funcionar en su forma prevista. Por tanto, por ejemplo, un promotor "operativamente unido" a un cistrón está unido de tal manera que se consigue la expresión del cistrón en condiciones compatibles con el promotor. De manera similar, una secuencia de nucleótidos de un IRES operativamente unida a un cistrón está unida de tal manera que se consigue la traducción del cistrón en condiciones compatibles con el IRES.

Las moléculas de ADN de la invención pueden estar presentes en forma de ADN bicatenario, que tienen con respecto al polipéptido marcador seleccionable y el polipéptido de interés una cadena codificante y una cadena no codificante, siendo la

5 cadena codificante la cadena con la misma secuencia que el ARN traducido, excepto por la presencia de T en lugar de U. Por tanto, se codifica un codón de iniciación AUG en la cadena codificante mediante una secuencia ATG, y la cadena que contiene esta secuencia ATG correspondiente al codón de iniciación AUG en

10 el ARN se denomina como la cadena codificante del ADN. Será evidente para el experto que los codones de iniciación o las secuencias de iniciación de la traducción están presentes, de hecho, en una molécula de ARN, pero que estas pueden considerarse incluidas de igual manera en una molécula de ADN

15 que codifica para tal molécula de ARN; por tanto, siempre que la presente invención haga referencia a un codón de iniciación o una secuencia de iniciación de la traducción, se pretende que se incluya la molécula de ADN correspondiente que tiene la misma secuencia que la secuencia de ARN excepto por la presencia de

20 una T en lugar de un U en la cadena codificante de dicha molécula de ADN, y viceversa, excepto cuando se especifique explícitamente lo contrario. En otras palabras, un codón de iniciación es por ejemplo una secuencia AUG en el ARN, pero la secuencia ATG correspondiente en la cadena codificante del ADN

25 también se denomina codón de iniciación en la presente invención. Se usa lo mismo para la referencia de secuencias codificantes "en marco", que significan tripletes (3 bases) en la molécula de ARN que se traducen en un aminoácido, pero también ha de interpretarse como las secuencias de

30 trinucleótidos correspondientes en la cadena codificante de la molécula de ADN.

Una secuencia de inicio de la traducción a menudo se denomina en el campo "secuencia de Kozak", y una secuencia de Kozak óptima es RCCATGG, el codón de iniciación subrayado,

35 siendo R una purina, es decir A o G (véase Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Por tanto, además del propio codón de iniciación, el contexto del mismo, en particular los nucleótidos

de -3 a -1 y +4, son relevantes, y una secuencia de inicio de la traducción óptima comprende un codón de iniciación óptimo (es decir ATG) en un contexto óptimo (es decir, el ATG precedido directamente por RCC y seguido directamente por G). La traducción por los ribosomas es más eficaz cuando está presente una secuencia de Kozak óptima (véase Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Sin embargo, en un pequeño porcentaje de acontecimientos, el ribosoma reconoce y usa secuencias de iniciación de la traducción no óptimas para iniciar la traducción. La presente invención hace uso de este principio, y permite disminuir la cantidad de traducción y por tanto la expresión del polipéptido marcador seleccionable, que por tanto puede usarse para aumentar la rigurosidad del sistema de selección.

La expresión "marcador de selección" o "marcador seleccionable" se usa normalmente para hacer referencia a un gen y/o una proteína cuya presencia puede detectarse directa o indirectamente en una célula, por ejemplo un polipéptido que inactiva un agente de selección y protege a la célula huésped de los efectos mortales o inhibitorios del crecimiento del agente (por ejemplo, un gen y/o proteína de resistencia a antibióticos). Los polipéptidos marcadores seleccionables se conocen bien en la técnica y se usan de manera rutinaria cuando se han de obtener clones de células huésped eucariotas, y se proporcionan varios ejemplos de proteínas marcadoras seleccionables adecuadas en el documento WO 2006/048459. Se conocen secuencias de ADN que codifican para tales polipéptidos marcadores seleccionables, y en el documento WO 2006/048459 se proporcionan varios ejemplos de secuencias de ADN de tipo natural que codifican para proteínas marcadoras seleccionables (por ejemplo, las figuras 15-21 en el mismo). Será evidente que los mutantes o derivados de marcadores seleccionables también pueden usarse adecuadamente, y por tanto se incluyen dentro del alcance de la expresión "polipéptido marcador seleccionable", siempre que la proteína marcadora seleccionable sea aún funcional. Por ejemplo, también se abarca cualquier mutación silenciosa que no altera la proteína codificada debido a la

redundancia del código genético. También se abarcan mutaciones adicionales que conducen a mutaciones de aminoácidos conservativas o a otras mutaciones, siempre que la proteína codificada aún tenga actividad, que puede o no ser inferior a la de la proteína de tipo natural que se codifica en las secuencias indicadas. En particular, se prefiere que la proteína codificada sea al menos el 70%, preferiblemente al menos el 80%, más preferiblemente al menos el 90%, aún más preferiblemente al menos el 95% idéntica con respecto a las proteínas codificadas por las secuencias indicadas respectivas. Las pruebas para detectar la actividad de las proteínas marcadoras seleccionables pueden llevarse a cabo mediante métodos de rutina. Un polipéptido marcador seleccionable según la invención es una proteína que se codifica por ácido nucleico, polipéptido que puede usarse funcionalmente para la selección, por ejemplo porque proporciona resistencia frente a un agente de selección tal como un antibiótico. Por tanto, cuando se usa un antibiótico como un agente de selección, el ADN codifica para un polipéptido que confiere resistencia frente al agente de selección, siendo el polipéptido el polipéptido marcador seleccionable. El polipéptido marcador seleccionable se codifica por el ADN de la invención. El polipéptido marcador seleccionable según la invención debe ser funcional en una célula huésped eucariota, y por tanto puede seleccionarse en células huésped eucariotas. Ejemplos de genes marcadores seleccionables adecuados para la presente invención son zeocina y neomicina. Otros candidatos adecuados incluyen, por ejemplo, blasticidina, puromicina, bleomicina, higromicina, DHFR, GS, etc. (véase también el documento WO 2006/048459). Otros genes marcadores seleccionables que podrían usarse, y sus agentes de selección, se describen por ejemplo en la tabla 1 de la patente estadounidense 5.561.053: véase también Kaufman, *Methods in Enzymology*, 185:537-566 (1990), para una revisión de los mismos. El término "selección" se define normalmente como el procedimiento de usar un marcador de selección/marcador seleccionable y un agente de selección para identificar células huésped con propiedades genéticas específicas (por ejemplo, que la célula huésped contenga un

transgén integrado en su genoma). Por conveniencia y tal como aceptan generalmente los expertos, en muchas publicaciones así como en el presente documento, a menudo el gen y la proteína que codifica para la resistencia frente a un agente de selección se denomina "gen de (resistencia a) agente seleccionable" o "proteína de (resistencia frente a) agente de selección", respectivamente, aunque los nombres oficiales pueden ser diferentes, por ejemplo el gen que codifica para la proteína que confiere resistencia frente a neomicina (así como frente a G418 y kanamicina) a menudo se denomina gen de (resistencia a) neomicina (o neo^r), mientras que el nombre oficial es gen de aminoglicósido 3'-fosfotransferasa.

Las secuencias codificantes de la proteína marcadora seleccionable de los documentos WO 2006/048459 y US 2006/0141577 en realizaciones preferidas tienen un codón de iniciación GTG o más preferiblemente uno TTG. Esto da como resultado una selección muy rigurosa y expresión muy alta de la proteína de interés en los clones que se obtienen. En la presente invención, las secuencias codificantes de la proteína marcadora seleccionable se mejoran adicionalmente reduciendo el contenido de CpG en las mismas, lo que da como resultado una rigurosidad aún mayor y niveles de expresión mejorados adicionalmente.

Preferiblemente, la secuencia de inicio de la traducción en la cadena codificante para el polipéptido marcador seleccionable comprende un codón de iniciación TTG. Preferiblemente, el codón de iniciación GTG o TTG está flanqueado por secuencias que proporcionan reconocimiento relativamente bueno de las secuencias distintas de ATG como codones de iniciación, tal que al menos algunos ribosomas inician la traducción a partir de estos codones de iniciación, es decir la secuencia de inicio de la traducción comprende preferiblemente la secuencia ACC[codón de iniciación GTG o TTG]G o GCC[codón de iniciación GTG o TTG]G.

En un aspecto, la invención proporciona una molécula de ADN que comprende una secuencia de marco de lectura abierto que codifica para un polipéptido marcador seleccionable, comprendiendo dicha molécula de ADN en la cadena codificante una secuencia de inicio de la traducción para el polipéptido

marcador seleccionable elegida del grupo que consiste en: a) un codón de iniciación GTG; y b) un codón de iniciación TTG; y en la que se ha mutado la secuencia de marco de lectura abierto que codifica para la proteína marcadora seleccionable para sustituir al menos el 10% de sus dinucleótidos de CpG (cualquier "CG" en la secuencia) en comparación con la secuencia nativa de marco de lectura abierto que codifica para la proteína marcadora seleccionable. Una molécula de ADN de este tipo puede usarse según la invención para obtener células huésped eucariotas que expresan altos niveles del polipéptido de interés, seleccionando para detectar la expresión del polipéptido marcador seleccionable. Posterior o simultáneamente, puede(n) identificarse una o más célula(s) huésped que expresa(n) el polipéptido de interés, y usarse adicionalmente para la expresión de altos niveles del polipéptido de interés.

En el presente documento se muestra que la reducción del contenido de CpG del gen marcador seleccionable de la invención, es decir que tiene un codón de iniciación TTG o GTG, puede conducir a una expresión mejorada de un polipéptido de interés que se traduce a partir de una unidad de transcripción multicistrónica a partir de la cual también se traduce el polipéptido marcador seleccionable. Sin querer limitarse a la teoría, se cree que la reducción del contenido de CpG puede reducir la posibilidad de silenciar la transcripción, porque los dinucleótidos de CpG pueden metilarse y silenciarse en eucariotas. Los polipéptidos marcadores seleccionables que se codifican por genes con un contenido de CpG relativamente alto, a menudo derivados de secuencias bacterianas, por ejemplo zeocina y neomicina, pueden beneficiarse mucho más de la reducción del contenido de CpG, aunque ya puede encontrarse algún beneficio para la selección de genes con un contenido de CpG relativamente bajo. En determinadas realizaciones, se eliminan los dinucleótidos de CpG de una secuencia que codifica para un polipéptido marcador seleccionable sin cambiar la secuencia de aminoácidos codificada. Esto puede llevarse a cabo aprovechando la redundancia del código genético, tal como conoce

bien y resulta rutinario para el experto en la técnica de la biología molecular.

Se espera que un efecto positivo de la eliminación de los dinucleótidos de CpG será evidente cuando se ha sustituido al menos el 10% de los dinucleótidos de CpG en la secuencia codificante del gen marcador seleccionable. Se espera que la eliminación de más dinucleótidos de CpG aumentará el efecto, y por tanto en determinadas realizaciones, se muta al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70% o al menos el 80% de los dinucleótidos de CpG en comparación con la secuencia nativa de marco de lectura abierto que codifica para la proteína marcadora seleccionable. En determinadas realizaciones ventajosas, se ha sustituido al menos la mitad de los dinucleótidos de CpG de la secuencia de marco de lectura abierto que codifica para el polipéptido marcador seleccionable en comparación con la secuencia nativa de marco de lectura abierto que codifica para el polipéptido marcador seleccionable.

Una secuencia nativa de marco de lectura abierto que codifica para el polipéptido marcador seleccionable que proporciona resistencia frente a zeocina se facilita como la SEQ. ID. NO. 1 (que contiene secuencias ATG internas), y la mutación de A en la posición 280 por T en esta secuencia facilita una secuencia que carece de secuencias ATG internas, y en la que la metionina codificada internamente en la posición 94 se sustituye por leucina. Para las secuencias de ADN de la invención, el codón de iniciación (los primeros tres nucleótidos de las secuencias de ADN) se muta por un codón de iniciación GTG o por uno TTG.

En determinadas realizaciones ventajosas, el polipéptido marcador seleccionable proporciona resistencia frente a zeocina. En determinadas realizaciones del mismo, la molécula de ADN comprende la SEQ. ID. NO. 1, en la que se ha sustituido al menos la mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, con la condición de que se sustituye el codón de iniciación (los primeros tres nucleótidos en la secuencia) por un codón de iniciación elegido de GTG o

TTG. En una realización alternativa, la molécula de ADN comprende SEQ. ID. NO. 1 en la que se sustituye el nucleótido A en la posición 280 por T, tal que se sustituye el aminoácido codificado 94 (metionina) por leucina, en la que se ha
5 sustituido al menos la mitad de los dinucleótidos de CpG sin mutar adicionalmente la secuencia de aminoácidos que se codifica, con la condición de que se sustituye el codón de iniciación (los primeros tres nucleótidos en la secuencia) por un codón de iniciación elegido de GTG o TTG. Esta realización
10 carece de secuencias ATG en la secuencia codificante para el gen de resistencia a zeocina, y por tanto es adecuado en las unidades de transcripción multicistrónicas de la invención en las que la secuencia codificante para el polipéptido marcador seleccionable está en el sentido de 5' de la secuencia
15 codificante para el polipéptido de interés. En una realización preferida de la misma, la molécula de ADN comprende SEQ. ID. NO. 3.

Una secuencia nativa de marco de lectura abierto que codifica para el polipéptido marcador seleccionable que
20 proporciona resistencia frente a neomicina se facilita como SEQ. ID. NO. 5 (que contiene secuencias ATG internas) y como SEQ. ID. NO. 7 (que carece de secuencias ATG internas). En realizaciones ventajosas, estas secuencias pueden contener una o más mutaciones adicionales de modo que el polipéptido codificado
25 tiene una mutación de valina en la posición 201 a glicina (201V>G), de ácido glutámico en la posición 185 a ácido aspártico (185E>D), o ambas (185E>D, 201V>G).

En otras realizaciones ventajosas, el polipéptido marcador seleccionable proporciona resistencia frente a neomicina. En
30 determinadas realizaciones del mismo, la molécula de ADN comprende una secuencia elegida del grupo que consiste en una cualquiera de: a) SEQ. ID. NO. 5, con la condición de que se ha sustituido al menos la mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, y con la
35 condición adicional de que se sustituye el codón de iniciación (la primera secuencia de ATG) por o bien GTG o bien TTG; b) SEQ. ID. NO. 7, con la condición de que se ha sustituido al menos la

mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, y con la condición adicional de que se sustituye el codón de iniciación (la primera secuencia de ATG) por o bien GTG o bien TTG; y c) SEQ. ID. NO. 5 o SEQ. ID. NO. 7, que contiene una mutación para codificar para una variante de proteína de resistencia frente a neomicina en comparación con las secuencias codificadas por las secuencias indicadas, teniendo dicha variante glicina en la posición 201 en la proteína codificada (variante 201G), o ácido aspártico en la posición 185 (variante 185D), o tanto glicina en la posición 201 como ácido aspártico en la posición 185 (variante 185D, 201G), con la condición de que se ha sustituido al menos la mitad de los dinucleótidos de CpG en la secuencia de ADN facilitada sin mutar adicionalmente la secuencia de aminoácidos que se codifica, y con la condición adicional de que se sustituye el codón de iniciación (la primera secuencia de ATG) por o bien GTG o bien TTG. La variante 185D se obtiene por ejemplo sustituyendo el codón de la posición 553-555 en las secuencias de ácido nucleico proporcionadas por la secuencia GAC, y la variante 201G se obtiene por ejemplo sustituyendo el codón de la posición 601-603 en la secuencia de ácido nucleico proporcionada por GGT. En una realización preferida, la molécula de ADN comprende SEQ. ID. NO. 9, con la condición de que se sustituye el nucleótido A en la posición 555 por C (para codificar para la variante 185E>D), y que se sustituye el nucleótido T en la posición 602 por G y que se sustituye el nucleótido G en la posición 603 por T (para codificar para la variante 201 V>G), y con la condición adicional de que se sustituye el codón de iniciación (ATG en las posiciones 1-3) por o bien GTG o bien TTG. Será evidente para el experto que el experto puede preparar variaciones adicionales sin apartarse de la enseñanza de la presente invención, y tales variaciones adicionales se abarcan con la presente invención siempre que el codón de iniciación no sea ATG y que la proteína codificada proporcione resistencia frente a neomicina (o G418). Según la presente invención, las variantes 185D y 201G mejoran adicionalmente la rigurosidad de selección.

En determinadas realizaciones, el polipéptido marcador seleccionable comprende además una mutación que reduce la actividad del polipéptido marcador seleccionable en comparación con su equivalente de tipo natural. Esto puede usarse para
5 aumentar aún más la rigurosidad de la selección. Como ejemplos no limitativos, puede mutarse la prolina en la posición 9 en el polipéptido de resistencia a zeocina, por ejemplo a Thr o Phe, y para el polipéptido de resistencia a neomicina, puede mutarse además el residuo de aminoácido 182 ó 261 o ambos (véase por
10 ejemplo el documento WO 01/32901).

En principio, las moléculas de ADN de la invención, que codifican para el polipéptido marcador seleccionable, pueden usarse en cualquier vector de expresión, por ejemplo como un gen monocistrónico. Proporcionan criterios de selección rigurosos.
15 Sin embargo, en realizaciones preferidas el ORF que codifica para un polipéptido marcador seleccionable es parte de una unidad de transcripción multicistrónica que comprende además una secuencia de ORF que codifica para un polipéptido de interés.

Una unidad de transcripción multicistrónica según la invención puede ser por ejemplo una unidad de transcripción multicistrónica que comprende secuencias que codifican de 5' a 3' para un polipéptido marcador seleccionable y para un polipéptido de interés, o por ejemplo una unidad de transcripción multicistrónica que comprende secuencias que
20 codifican de 5' a 3' para un polipéptido de interés y para un polipéptido marcador seleccionable. En el primer caso, la secuencia codificante para el polipéptido marcador seleccionable preferiblemente carece de secuencias ATG en la cadena codificante (véase el documento WO 2006/048459). En el último
25 caso, el polipéptido de interés se codifica en el sentido de 5' de la secuencia codificante para el polipéptido marcador seleccionable y se une operativamente un sitio interno de entrada al ribosoma (IRES) a la secuencia que codifica para el polipéptido marcador seleccionable, y por tanto el polipéptido
30 marcador seleccionable es dependiente del IRES para su traducción (véase el documento US 2006/0141577). Por tanto en una realización, una unidad de transcripción multicistrónica de

la invención comprende en el siguiente orden: a) un promotor; b) la secuencia que codifica para la proteína marcadora seleccionable; y c) una secuencia que codifica para una proteína de interés. En otra realización, una unidad de transcripción multicistrónica de la invención comprende en el siguiente orden:
5 a) un promotor; b) una secuencia que codifica para una proteína de interés; y c) un sitio interno de entrada al ribosoma (IRES), operativamente unido a d) la secuencia que codifica para la proteína marcadora seleccionable.

10 En determinadas realizaciones, las unidades de transcripción multicistrónicas comprenden un tercer cistrón en el sentido de 3' del segundo cistrón, dicho tercer cistrón preferiblemente está operativamente unido a un IRES, y por ejemplo codifica para un segundo polipéptido marcador
15 seleccionable. En determinadas realizaciones, este segundo polipéptido marcador seleccionable es DHFR, preferiblemente con un codón de iniciación GTG o TTG para permitir la selección continua de células deficientes en dhfr (véase, por ejemplo el documento PCT/EP2007/051696).

20 En determinadas realizaciones, la invención proporciona un casete de expresión que comprende una molécula de ADN de la invención, comprendiendo dicho casete de expresión un promotor en el sentido de 5' de una unidad de transcripción multicistrónica de la invención y una secuencia de terminación
25 de la transcripción en el sentido de 3' de la misma. Dicho casete de expresión es funcional en una célula huésped eucariota para accionar la transcripción de la unidad de transcripción multicistrónica

Un "casete de expresión" tal como se usa en el presente
30 documento es una secuencia de ácido nucleico que comprende al menos un promotor unido funcionalmente a una secuencia cuya expresión se desea. Preferiblemente, un casete de expresión contiene además secuencias de terminación de la transcripción y de poliadenilación. Se conocen bien ejemplos de promotores
35 adecuados y de secuencias de terminación de la transcripción/poliadenilación y están fácilmente disponibles

para el experto, y se tratan por ejemplo en el documento WO 2006/048459, págs. 28-29.

También pueden incluirse otras secuencias reguladores tales como potenciadores. El promotor debe poder funcionar en una
5 célula huésped eucariota, es decir debe poder accionar la transcripción de la unidad de transcripción. Por tanto, el promotor está operativamente unido a la unidad de transcripción. El casete de expresión puede contener además opcionalmente otros elementos conocidos en la técnica, por ejemplo sitios de corte y
10 empalme, para comprender intrones, y similares. En las realizaciones en las que el polipéptido marcador seleccionable se codifica en el sentido de 3' del polipéptido de interés, se une operativamente un IRES al cistrón que contiene la secuencia que codifica para el polipéptido marcador seleccionable. En las
15 realizaciones en las que el polipéptido marcador seleccionable se codifica en el sentido de 5' del polipéptido de interés, la secuencia que codifica para el polipéptido marcador seleccionable carece de secuencias ATG en la cadena codificante.

Tal como se usa en el presente documento, un "sitio interno de entrada al ribosoma" o "IRES" se refiere a un elemento que
20 promueve la entrada directa del ribosoma interno al codón de iniciación, tal como normalmente un ATG, pero en esta invención preferiblemente GTG o TTG, de un cistrón (una región que codifica para la proteína), conduciendo de ese modo a la traducción independiente de caperuza del gen. El experto en la
25 técnica conoce bien las secuencias IRES y el uso de las mismas para la expresión, tal como se enseña en los documentos US 2006/0141577 y PCT/EP2007/051696. Véase también, por ejemplo, Jackson R J, Howell M T, Kaminski A (1990) Trends Biochem Sci 15
30 (12): 477-83), Jackson R J y Kaminski, A. (1995) RNA 1 (10): 985-1000, Martinez-Salas, 1999, Venkatesan & Dasgupta, 2001, Rees *et al*, 1996, y Mizuguchi *et al.*, 2000. Un ejemplo de una secuencia IRES adecuada se facilita en el ejemplo 19 del documento US 2006/0141577 (SEQ ID NO. 127 en el mismo).

35 Las moléculas de ADN según la invención pueden generarse mediante métodos de biología molecular convencionales disponibles para el experto. Por ejemplo, pueden mutarse

secuencias nativas, por ejemplo a partir de plásmidos comercialmente disponibles, mediante métodos de rutina. Además, en la actualidad también es posible sintetizar a voluntad (si se requiere usando etapas de subclonación) secuencias de ADN que

5 tienen una longitud suficiente para un ORF de un polipéptido marcador seleccionable, y tales secuencias de ADN sintéticas en la actualidad pueden pedirse comercialmente de diversas compañías. Por tanto, usando las enseñanzas de la presente invención, el experto en la técnica puede diseñar secuencias

10 apropiadas según la invención que codifican para un polipéptido marcador seleccionable (con un codón de iniciación GTG o TTG, y con contenido de CpG reducido, y en determinadas realizaciones que no tienen secuencias ATG internas), hacer que se sintetice esta secuencia y someter a prueba la molécula de ADN para

15 determinar la funcionalidad del marcador seleccionable codificado introduciendo la molécula de ADN en células huésped eucariotas y someter a prueba para detectar la expresión del polipéptido marcador seleccionable funcional. La disponibilidad comercial de tales secuencias también hace factible proporcionar

20 sin demasiada carga secuencias que codifican para marcador de selección que carecen de secuencias ATG internas, en las que la secuencia codificante de tipo natural del polipéptido marcador de selección comprende varias secuencias ATG internas de este tipo (véase el documento WO 2006/048459).

25 En determinadas realizaciones, una molécula de ADN según la invención es parte de un vector, por ejemplo un plásmido. Tales vectores pueden manipularse fácilmente mediante métodos bien conocidos por el experto en la técnica, y pueden diseñarse por ejemplo para poder replicarse en células procariotas y/o

30 eucariotas. Además, muchos vectores pueden usarse directamente o en forma de fragmento deseado aislado de los mismos para la transformación de células eucariotas y se integrarán en su totalidad o en parte en el genoma de tales células, lo que da como resultado células huésped estables que comprenden el ácido

35 nucleico deseado en su genoma.

El vector usado puede ser cualquier vector que es adecuado para clonar ADN y que puede usarse para la transcripción del

ácido nucleico de interés. Cuando se usan las células huésped, se prefiere que el vector sea un vector de integración. Alternativamente, el vector puede ser un vector de replicación episomal.

5 Se aprecia en gran medida que la estructura de la cromatina y otros mecanismos de control epigenético pueden influir sobre la expresión de transgenes en células eucariotas (por ejemplo Whitelaw *et al.* 2001). Las unidades de expresión multicistrónicas según la invención forman parte de un sistema
10 de selección con un régimen de selección bastante riguroso. Generalmente esto requiere altos niveles de transcripción en las células huésped de elección. Para aumentar la posibilidad de encontrar clones de células huésped que sobreviven el riguroso régimen de selección, y posiblemente para aumentar la
15 estabilidad de la expresión en los clones obtenidos, generalmente será preferible aumentar la predictibilidad de la transcripción. Por tanto, en realizaciones preferidas, un casete de expresión según la invención comprende además al menos un elemento de control de cromatina. Un "elemento de control de
20 cromatina" tal como se usa en el presente documento es una expresión colectiva para secuencias de ADN que pueden de algún modo tener un efecto sobre la estructura de la cromatina y con ello sobre el nivel de expresión y/o estabilidad de la expresión de transgenes en su proximidad (funcionan "en cis", y por tanto se colocan preferiblemente dentro de 5 kb, más preferiblemente dentro de 2 kb, aún más preferiblemente dentro de 1 kb del transgén) dentro de células eucariotas. Preferiblemente un elemento de control de cromatina de este tipo se elige del grupo que consiste en una secuencia aislante, un elemento de apertura
25 de cromatina ubicuo (UCOE), regiones de unión a matriz o estructura (MAR/SAR) y secuencias antirrepresoras (STAR). Ejemplos de elementos de control de cromatina, así como métodos para obtenerlos y usarlos y someterlos funcionalmente a prueba, se facilitan en el documento WO 2006/048459, páginas 32-37. En
30 determinadas realizaciones, dicho al menos un elemento de control de cromatina es un elemento antirrepresor elegido del grupo que consiste en una cualquiera de SEQ. ID. NO. 1 a SEQ.

ID. NO. 66 del documento WO 2006/048459, y fragmentos de las mismas. En determinadas realizaciones del mismo, dicho casete de expresión comprende la SEQ. ID. NO. 66 del documento WO 2006/048459, o un fragmento del mismo ubicado en el sentido de 5' del promotor que acciona la transcripción de la unidad de transcripción multicistrónica. En otras realizaciones, la unidad de transcripción multicistrónica está flanqueada en ambos lados por al menos una secuencia antirrepresora elegida del grupo que consiste en una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 65 del documento WO 2006/048459, o fragmentos de las mismas. Preferiblemente, el elemento de control de cromatina se elige del grupo que consiste en STAR67, STAR7, STAR9, STAR17, STAR27, STAR29, STAR43, STAR44, STAR45, STAR47, STAR61, o un fragmento funcional o derivado de dichas secuencias STAR (véase por ejemplo el documento WO 2006/048459 para las secuencias y usos preferidos de estos elementos STAR).

Un polipéptido de interés según la invención puede ser cualquier proteína, y puede ser una proteína monomérica o una (parte de una) proteína multimérica. Una proteína multimérica comprende al menos dos cadenas de polipéptido. Ejemplos no limitativos de una proteína de interés según la invención son enzimas, hormonas, cadenas de inmunoglobulinas, proteínas terapéuticas tales como proteínas anticancerígenas, proteínas de la coagulación sanguínea tales como factor VIII, proteínas multifuncionales, tales como eritropoyetina, proteínas de diagnóstico, o proteínas o fragmentos de las mismas útiles para fines de vacunación, todas conocidas por el experto en la técnica.

El polipéptido de interés puede ser de cualquier fuente, y en determinadas realizaciones es una proteína de mamífero, una proteína artificial (por ejemplo una proteína de fusión o una proteína mutada), y preferiblemente es una proteína humana.

Las moléculas de ADN que comprenden unidades de transcripción multicistrónicas y/o casetes de expresión según la invención pueden usarse para mejorar la expresión del ácido nucleico, preferiblemente en células huésped. Las expresiones "célula"/"célula huésped" y "línea celular"/"línea celular

huésped" normalmente se definen respectivamente como una célula y poblaciones homogéneas de la misma que pueden mantenerse en cultivo celular mediante métodos conocidos en la técnica, y que tienen la capacidad de expresar proteínas homólogas o heterólogas. La invención proporciona además células huésped que comprenden una molécula de ADN o un casete de expresión según la presente invención.

Las células huésped procariotas pueden usarse para propagar y/o realizar ingeniería genética con las moléculas de ADN de la invención, especialmente cuando están presentes en plásmidos que pueden replicarse en células huésped procariotas tales como bacterias.

Una célula huésped según la presente invención es preferiblemente una células eucariota, más preferiblemente una célula de mamífero, tal como una célula de roedor (por ejemplo ratón, hámster) o una célula humana o fusión entre diferentes células. En determinadas realizaciones no limitativas, dicha célula huésped es una célula de osteosarcoma U-2 OS, célula HEK 293, célula de mieloma HuNS-1, célula de retinoblastoma WERI-Rb-1, célula BHK, célula COS, célula Vero, célula de mieloma de ratón no secretor Sp2/0-Ag 14, célula de mieloma de ratón no secretor NS0, célula de carcinoma de glándula suprarrenal NCI-H295R o una célula PER.C6[®]. Las células PER.C6 para el fin de la presente invención significan células de un pasaje aguas arriba o aguas abajo o un descendiente de un pasaje aguas arriba o aguas abajo de células tal como se depositaron con el n.º ECACC 96022940 (véase por ejemplo la patente estadounidense 5.994.128), es decir que tienen las características de estas células. Anteriormente se ha mostrado que tales células pueden expresar proteínas en altos niveles (por ejemplo el documento WO 00/63403, y Jones *et al*, 2003). En determinadas realizaciones preferidas, las células huésped son células CHO (ovario de hámster chino), por ejemplo CHO-K1, CHO-S, CHO-DG44, CHO-DUKXB11 y similares. En determinadas realizaciones, dichas células CHO tienen un fenotipo dhfr⁻.

Tales células huésped eucariotas pueden expresar polipéptidos deseados, y a menudo se usan para este fin. Pueden

obtenerse mediante la introducción de una molécula de ADN de la invención, preferiblemente en forma de un casete de expresión, en las células. Preferiblemente, el casete de expresión se integra en el genoma de las células huésped, que puede estar en diferentes posiciones en diversas células huésped, y la selección proporcionará un clon en el que el transgén está integrado en una posición adecuada, lo que conduce a un clon de célula huésped con las propiedades deseadas en cuanto a niveles de expresión, estabilidad, características de crecimiento y similares. Alternativamente, la unidad de transcripción puede seleccionarse como diana o seleccionarse al azar para su integración en una región cromosómica que se activa de manera transcripcional, por ejemplo detrás de un promotor presente en el genoma.

Preferiblemente, las células huésped provienen de un clon estable que puede seleccionarse y propagarse según procedimientos convencionales conocidos por el experto en la técnica. Un cultivo de un clon de este tipo puede producir el polipéptido de interés, si las células comprenden la unidad de transcripción que codifica para el mismo.

La invención también proporciona un método de generación de una célula huésped que puede expresar un polipéptido de interés, comprendiendo dicho método las etapas de: a) introducir en una pluralidad de células precursoras una molécula de ADN o un casete de expresión según la invención, b) cultivar la pluralidad de células precursoras en condiciones adecuadas para la expresión del polipéptido marcador seleccionable, y c) seleccionar al menos una célula huésped que expresa el polipéptido marcador seleccionable. La selección para determinar la expresión del polipéptido marcador seleccionable se lleva a cabo por ejemplo aplicando presión de selección (por ejemplo, cultivar en presencia del agente de selección) y asegurará la expresión del polipéptido de interés en las unidades de transcripción multicistrónicas y casetes de expresión de la invención. Este método novedoso proporciona un muy buen resultado en cuanto a la razón de clones obtenidos frente a clones con alta expresión del polipéptido deseado: se obtienen

muchas menos colonias usando la misma concentración de agente de selección que con sistemas de selección conocidos, y un porcentaje relativamente alto de los clones obtenidos producen el polipéptido de interés en altos niveles.

5 La invención proporciona además un método para producir un polipéptido de interés, que comprende cultivar una célula huésped que comprende un casete de expresión según la invención, para expresar el ácido nucleico que codifica para la proteína de interés en dicha célula. En realizaciones preferidas, la
10 proteína de interés se recoge de dicha célula o del medio de cultivo o de ambos. En una realización preferida, dicha célula es una célula de mamífero, por ejemplo, una célula CHO.

 La introducción del ácido nucleico que ha de expresarse en una célula, puede llevarse a cabo mediante uno de varios
15 métodos, que como tal el experto en la técnica conoce, también dependen del formato del ácido nucleico que va introducirse. Dichos métodos incluyen, pero no se limitan a, transfección, infección, inyección, transformación y similares.

 En determinadas realizaciones, el agente de selección está
20 presente en el medio de cultivo al menos parte del tiempo durante el cultivo, o bien en concentraciones suficientes para seleccionar células que presentan el polipéptido marcador seleccionable o bien en concentraciones menores. En otras realizaciones, el agente de selección ya no está presente en el
25 medio de cultivo durante la fase de producción cuando se expresa el polipéptido.

 Se lleva a cabo el cultivo de una célula para permitirle metabolizar, y/o crecer y/o dividirse y/o producir proteínas recombinantes de interés. Esto puede lograrse mediante métodos
30 bien conocidos por los expertos en la técnica, e incluye, pero no se limita a, proporcionar nutrientes para la célula. Los métodos comprenden crecimiento adhiriéndose a superficies, crecimiento en suspensión, o combinaciones de los mismos. El cultivo puede llevarse a cabo por ejemplo en placas, botellas de
35 cultivo rotatorias o en biorreactores, usando sistemas discontinuos, semicontinuos, continuos tales como sistemas de perfusión y similares. Con el fin de conseguir una producción

(continua) a gran escala de proteínas recombinantes mediante cultivo celular, en la técnica se prefiere tener células que puedan crecer en suspensión, y se prefiere tener células que puedan cultivarse en medio de cultivo libre de suero, o incluso libre de proteínas.

El experto en la técnica conoce las condiciones para hacer crecer y multiplicar células (véase por ejemplo *Tissue Culture*, Academic Press, Kruse and Paterson, editores (1973)) y las condiciones para la expresión del producto recombinante. En general, los principios, protocolos y las técnicas prácticas para maximizar la productividad de cultivos de células de mamíferos pueden encontrarse en *Mammalian Cell Biotechnology: a Practical Approach* (M. Butler, ed., IRL Press, 1991).

En una realización preferida, la proteína expresada se recoge (aisla), o bien de las células o bien del medio de cultivo o bien de ambos. Entonces puede purificarse adicionalmente usando métodos conocidos, por ejemplo filtración, cromatografía en columna, etc., mediante métodos generalmente conocidos por el experto en la técnica.

Obviamente, también pueden usarse las configuraciones de los casetes de expresión cuando el objetivo final no es la producción de un polipéptido de interés, sino el propio ARN, por ejemplo para producir cantidades aumentadas de ARN a partir de un casete de expresión, que puede usarse para fines de regulación de otros genes (por ejemplo iARN, ARN antisentido), terapia génica, producción de proteínas *in vitro*, etc.

La práctica de esta invención empleará, a menos que se indique lo contrario, técnicas convencionales de inmunología, biología molecular, microbiología, biología celular y ADN recombinante, que están dentro de la experiencia de la técnica. Véase por ejemplo Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual*, 2ª edición, 1989; *Current Protocols in Molecular Biology*, Ausubel FM, *et al*, eds, 1987; la serie *Methods in Enzymology* (Academic Press, Inc.); PCR2: *A Practical Approach*, MacPherson MJ, Hams BD, Taylor GR, eds, 1995; *Antibodies: A Laboratory Manual*, Harlow and Lane, eds, 1988.

La invención se explica adicionalmente en los siguientes ejemplos. Los ejemplos no limitan la invención en ninguna manera. Simplemente sirven para aclarar la invención.

5

EJEMPLOS

Ejemplo 1: La eliminación de dinucleótidos de CpG de la secuencia que codifica para marcador seleccionable mejora la expresión usando un método de selección de la invención

10

15

20

25

30

35

Los métodos de selección que usan diferentes codones de iniciación de la traducción para el marcador seleccionable, tal como GTG o TTG, pueden dar como resultado una selección muy rigurosa, y niveles muy altos de producción para el polipéptido de interés (véanse los documentos WO 2006/048459 y US 2006/0141577, por ejemplo los ejemplos 1-19 en el último). En este ejemplo, se modificó la región codificante del propio gen de polipéptido marcador seleccionable eliminando dinucleótidos de CpG. El fundamento es que el nucleótido C en el nucleótido CpG puede ser propenso a metilación, lo que puede dar como resultado el silenciamiento génico del marcador seleccionable, y por tanto la eliminación de los dinucleótidos de CpG puede mejorar los resultados. Se tomó el gen de resistencia a zeocina con un codón de iniciación TTG como el marcador, y se eliminaron tantos dinucleótidos de CpG como fue posible, sin cambiar la secuencia de aminoácidos de la proteína de resistencia frente a zeocina, y además sin introducir secuencias ATG en la cadena codificante, para evitar la iniciación de la traducción no deseada dentro de la región codificantes de la proteína de resistencia frente a zeocina (tal como se explica por ejemplo en el documento WO 2006/048459). Por tanto, algunos CpG no se eliminaron. El contenido de CpG de la secuencia nativa (en este caso: que contiene un codón de iniciación TTG, y una mutación para eliminar la secuencia ATG interna) es del 13,3%, mientras que tras mutar los CpG, el contenido de CpG se redujo hasta el 1,8% [denominado "TTG Zeo (pobre en CpG)"]. Se clonó el gen de resistencia a zeocina con contenido de CpG disminuido en el sentido de 5' de la secuencia que codifica para d2EGFP para dar

como resultado una construcción de expresión multicistrónico. Se midieron los niveles de expresión de d2EGFP.

Se prepararon construcciones que contenían STAR 7 y 67 en el sentido de 5' del promotor de CMV, seguido del marcador de selección TTG Zeo (pobre en CpG) (sintetizado por GeneArt GmbH, Regensburg, Alemania; véase SEQ. ID. NO 3; véase SEQ. ID. NO. 1 para la secuencia codificante de resistencia a zeocina con su contenido de CpG natural), el gen d2EGFP y STAR 7 (figura 1). Se transfectaron las construcciones a células CHO-K1. Se transfectó ADN usando Lipofectamine 2000 (Invitrogen) y se hicieron crecer las células en presencia de zeocina 150 µg/ml en medio HAM-F12 (Invitrogen) + FBS al 10% (Invitrogen).

Tras la transfección aparecieron ocho colonias con la construcción TTG Zeo "rico en CpG" de control (A en la figura 1) y ninguna con la construcción que contenía TTG Zeo "pobre en CpG" (C en la figura 1). Por el contrario, aparecieron más de 24 colonias con los marcadores de selección tanto TTG Zeo "rico en CpG" (B en la figura 1) como TTG Zeo "pobre en CpG" (D en la figura 1) cuando se incluyó STAR 7/67-7 en la construcción. Con el marcador de selección TTG Zeocina "rico en CpG" (A en la figura 1), la expresión promedio de d2EGFP con la construcción de control sin STAR fue de 140, y con la construcción que contenía STAR 1332 (B en la figura 1). Esto es un aumento debido a la presencia de los elementos STAR. La expresión promedio de d2EGFP con la construcción que contenía STAR y el Zeo "pobre en CpG" fue de 2453 (D en la figura 1), un aumento de casi el doble en comparación con el TTG Zeo "rico en CpG" (B en la figura 1). Además, el valor de d2EGFP más alto conseguido con la construcción TTG Zeo "rico en CpG" (B) fue de 2481 y con el TTG Zeo "pobre en CpG" (D) de 4308.

Se concluye que la disminución del contenido de CpG del gen marcador de zeocina aumenta la rigurosidad del sistema de selección. Esto da como resultado mayores valores de expresión de d2EGFP cuando se incluyen elementos STAR en la construcción y ninguna colonia con la construcción de control.

También se transfectaron las mismas construcciones a células CHO-DG44. Esto se llevó a cabo con Lipofectamine 2000 (Invitrogen) y se realizó la selección con zeocina 150 µg/ml en el medio de cultivo. El medio de cultivo consistió en HAMF12:DMEM = 1:1, + suero bovino fetal al 10%. Con el marcador de selección TTG Zeocina "rico en CpG", la expresión promedio de d2EGFP con la construcción de control sin STAR fue de 43 (A en la figura 2), y la expresión promedio de d2EGFP con las construcciones que contenían STAR fue de 586 (B en la figura 2). Esto es un aumento debido a la presencia de los elementos STAR. La expresión promedio de d2EGFP con las construcciones con STAR y el Zeo "pobre en CpG" fue de 1152 (D en la figura 2), un aumento de casi el doble en comparación con el TTG Zeo "rico en CpG" (B en la figura 2). Además, el valor de d2EGFP más alto conseguido con la construcción TTG Zeo "rico en CpG" fue de 1296 (B en la figura 2) y con el TTG Zeo "pobre en CpG" de 2416 (D en la figura 2). A diferencia de CHO-K1, en las que no apareció ninguna colonia de control con la construcción TTG Zeo "pobre en CpG" (C en la figura 1), aparecieron colonias de control con CHO-DG44, pero el valor promedio de d2EGFP fue de 52 y el valor más alto en una colonia fue de 115 (C en la figura 2).

Se concluye que también en CHO-DG44 la adición del marcador de selección TTG Zeo "pobre en CpG" a la construcción da como resultado una mayor expresión de proteína cuando se emplean elementos STAR.

Ejemplo 2: Modificaciones en la secuencia codificante de resistencia a neomicina en el sistema de selección de la invención

En este ejemplo, además del codón de iniciación, también se modificó la región codificante del gen de resistencia a neomicina, eliminando tantos dinucleótidos de CpG del gen de resistencia a neomicina (sin ATG, de modo que ya carece de secuencias ATG en la cadena codificante) como fue posible, mientras que no se cambió la secuencia de aminoácidos de la proteína de resistencia frente a neomicina (excepto por las

mutaciones Met>Leu en las que las secuencias ATG internas estaban en marco y se sustituyeron por CTG en comparación con la secuencia de tipo natural: obviamente esto se realizó por razones de eliminación de las secuencias ATG de la cadena codificante e independiente del esfuerzo de reducción del contenido de CpG, véase el ejemplo 17 del documento WO 2006/048459), y sin introducir nuevas secuencias ATG en la cadena codificante, de manera análoga a lo que se realizó en el ejemplo 1 para el gen de resistencia a zeocina. El contenido de CpG del gen marcador de selección de neomicina de "tipo natural" es del 10,4% (SEQ. ID. NO. 5), mientras que tras los cambios se redujo el contenido de CpG hasta el 2,3% (SEQ. ID. NO. 9). Las construcciones que contienen las secuencias para el gen de resistencia a neomicina en este ejemplo se pidieron a GeneArt GmbH, Regensburg, Alemania. Como codón de iniciación, se usó TTG en este ejemplo. Por tanto, las secuencias usadas consistieron en SEQ. ID. NO. 9, con la condición de que se sustituyó el codón de iniciación (los primeros tres nucleótidos, ATG) por un codón de iniciación TTG, y además en determinados casos contenían una de las mutaciones indicadas anteriormente.

En el gen de resistencia a neomicina "pobre en CpG", se realizaron algunas mutaciones para cambiar aminoácidos de la proteína de resistencia frente a neomicina, para someter a prueba si estas tienen influencia sobre los niveles de expresión del polipéptido de interés cuando se usan en las unidades transcripción multicistrónicas de la invención. Las mutaciones (Sautter *et al*, 2005; se indica que la secuencia neo usada en la presente solicitud codifica para tres aminoácidos adicionales inmediatamente después del codón de iniciación en comparación con las secuencia usada por (Sautter *et al*, 2005), y por tanto la numeración de aminoácidos en la presente solicitud es mayor en tres en comparación con la numeración en (Sautter *et al*, 2005)) consistieron en un cambio del aminoácido valina 201 (198 en Sautter *et al*, 2005) a glicina 201 (TTG Neo 201V>G), ácido glutámico 185 (182 en Sautter *et al*, 2005) a ácido aspártico 185 (TTG Neo 185E>D) y una doble mutación en la que se cambiaron tanto el aminoácido valina 201 como ácido glutámico 185 a

glicina 201 y ácido aspártico 185, respectivamente (TTG Neo 185E>D/201V>G) (figura 3). Estas modificaciones se compararon con la neomicina control (TTG Neo pobre en CpG 185E/201V). En todos los casos se prepararon construcciones con y sin elementos STAR (figura 3).

Se incorporó el marcador de selección TTG Neo modificado en una construcción que contenía STAR 7 y 67 en el sentido de 5' del promotor de CMV, seguido por el marcador de selección TTG Neo, el gen d2EGFP y STAR 7 (figura 3). Se transfectaron las construcciones a células CHO-K1. Se transfectó ADN usando Lipofectamine 2000 (Invitrogen) y se hicieron crecer las células en presencia de geneticina G418 500 µg/ml en medio HAM-F12 (Invitrogen) + FBS al 10% (Invitrogen).

Con la construcción Neo de control (185E/201V) sólo se observó un efecto muy limitado de los elementos STAR. Esto puede deberse al menos en parte a las numerosas colonias que se generaron con geneticina G418 500 µg/ml, lo que indicó que la rigurosidad de la modificación de TTG neomicina es baja. Sin embargo, la neomicina con modificaciones de la invención es operativa: en la construcción TTG Neo 185E 201V se eliminaron todas las secuencias ATG de la cadena codificante del gen de resistencia a neomicina, y aunque los valores de d2EGFP fueron bajos, es evidente que la eliminación de las secuencias ATG aún permitió una selección apropiada bajo la presión de selección con geneticina. Cuando se modificó adicionalmente el gen de resistencia a neomicina, se observó un efecto distintivo de la adición de los elementos STAR. La media de 21 colonias de control TTG Neo 201V>G fue de 65 (A2 en la figura 3), mientras que la señal media de d2EGFP de las 24 colonias TTG Neo 201V>G con elementos STAR fue de 150 (B2 en la figura 3). Se aumentó adicionalmente la rigurosidad de selección con la mutación TTG Neo 185E>D, puesto que ninguna colonia de control sobrevivió sin elementos STAR (A3 en la figura 3), mientras que la señal media de d2EGFP de 17 colonias TTG Neo 185E>D con STAR que sobrevivieron fue de 204 (B3 en la figura 3). Esta fluorescencia media de GFP es mayor que con las colonias TTG Neo 201V>G (B2 en

la figura 3). Además el valor de d2EGFP más alto en colonias TTG Neo 185E>D fue de 715, en comparación con 443 en las colonias TTG Neo 201V>G (comparar B3 y B2 en la figura 3). Se observó la rigurosidad más alta en el mutante Neo doble, TTG Neo 185E>D
5 201V>G. No sobrevivió ninguna colonia de control (A4 en la figura 3) y el valor medio de d2EGFP de 7 colonias TTG Neo 185E>D 201V>G con STAR que sobrevivieron fue de 513, con el valor de d2EGFP más alto de 923 (B4 en la figura 3).

Se concluye que la introducción de mutaciones específicas
10 aumenta la rigurosidad de la selección del gen de resistencia a neomicina cuando se usa según la invención. Algunas de estas modificaciones transmiten tal rigurosidad de selección al gen de resistencia a neomicina que sólo pueden sobrevivir las colonias tras la incorporación de elementos STAR, debido a mayores
15 valores de expresión. Esto da como resultado al mismo tiempo mayores valores de expresión de d2EGFP. Evidentemente, las realizaciones ventajosas descritas en el presente documento del gen de resistencia a neomicina mejoran adicionalmente la idoneidad de este gen para su uso según la presente invención.

Será evidente que la configuración en la que un gen de resistencia a neomicina con un contenido de CpG disminuido y con un codón de iniciación GTG o TTG, y con las mutaciones indicadas (185E>D y/o 201V>G) también podría colocarse en el sentido de 3'
20 de la secuencia codificante para el polipéptido de interés (en este caso d2EGFP como modelo) cuando las secuencias que codifican para la proteína de resistencia frente a neomicina se colocan bajo el control de un IRES (véase por ejemplo, el ejemplo 19 en el documento US 2006/0141577). Lo mismo se considera para el gen de resistencia a zeocina (ejemplo 1). En
25 tal caso, no necesita tenerse ningún cuidado de que la mutación de dinucleótidos de CpG introduzca secuencias ATG. Se espera que también en tales realizaciones pueden obtenerse buenos resultados, es decir que la reducción del contenido de CpG y la mutación específica en las posiciones indicadas de la secuencia
30 que codifica para la proteína marcadora seleccionable mejorarán los niveles de expresión.

BIBLIOGRAFÍA

Jones D, Kroos N, Anema R, Van Montfort B, Vooy's A, Van Der Kraats S, Van Der Helm E, Smits S, Schouten J, Brouwer K, Lagerwerf F, Van Berkel P, Opstelten D-J, Logtenberg T, Bout A
5 (2003) High-level expression of recombinant IgG in the human cell line PER:C6. *Biotechnol. Prog.* 19: 163-168.

Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44: 283-292.

10 Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15: 8125-8148.

Kozak M. (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol.* 9: 5073-5080.

15 Kozak M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA* 87:8301-8305.

Kozak M. (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not
20 generally affected by the nucleotides in positions +5 and +6. *EMBO J.* 16: 2482-2492.

Kozak M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299: 1-34.

25 Martinez-Salas, E. (1999) Internal ribosome entry site biology and its use in expression vectors. *Curr Opin Biotechnol* 10, 458-64.

Mizuguchi, H, Xu, Z, Ishii-Watabe, A, Uchida, E, and Hayakawa, T. (2000) IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in
30 a bicistronic vector. *Mol Ther* 1, 376-82.

Rees, S, Coote, J, Stables, J, Goodson, S, Harris, S, and Lee, MG. (1996) Bicistronic vector for the creation of stable mammalian cell lines that predisposes all antibiotic-resistant cells to express recombinant protein. *Biotechniques* 20, 102-104,
35 106, 108-110.

Sautter, K, Enenkel, B. 2005. Selection of high-producing CHO cells using NPT selection marker with reduced enzyme activity. *Biotechnol Bioeng.* 89, 530-538.

5 Venkatesan, A, and Dasgupta, A. (2001) Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements. *Mol Cell Biol* 21, 2826-37.

Whitelaw, E, Sutherland, H, Kearns, M, Morgan, H, Weaving, L, and Garrick, D. (2001) Epigenetic effects on transgene expression. *Methods Mol Biol* 158, 351-68.

10

LISTADO DE SECUENCIAS

<110> ChromaGenics B.V.

Otte, Arie P.

Kwaks, Theodorus H.J.

15

Sewalt, Richard G.A.B.

van Blokland, Rik

<120> Selección de células huésped que expresan proteína en altos niveles

<130> 0117 B WO 00 ORD

20

<150> Documento EP 06113354.2

<151> 02-05-2006

<150> Documento US 11/416.490

25

<151> 02-05-2006

<160> 10

<170> PatentIn versión 3.3

30

<210> 1

<211> 375

<212> ADN

<213> Artificial

35

<220>

<223> Gen de resistencia a zeocina de tipo natural (wt)

<220>

<221> CDS

<222> (1)..(375)

<400> 1

atg gcc aag ttg acc agt gcc gtt ccg gtg ctc acc gcg cgc gac gtc	48
Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val	
1 5 10 15	
gcc gga gcg gtc gag ttc tgg acc gac cgg ctc ggg ttc tcc cgg gac	96
Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp	
20 25 30	
ttc gtg gag gac gac ttc gcc ggt gtg gtc cgg gac gac gtg acc ctg	144
Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu	
35 40 45	
ttc atc agc gcg gtc cag gac cag gtg gtg ccg gac aac acc ctg gcc	192
Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala	
50 55 60	
tgg gtg tgg gtg cgc gcc ctg gac gag ctg tac gcc gag tgg tcc gag	240
Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu	
65 70 75 80	
gtc gtg tcc acg aac ttc cgg gac gcc tcc ggg ccg gcc atg acc gag	288
Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu	
85 90 95	
atc gcc gag cag ccg tgg ggg cgg gag ttc gcc ctg cgc gac ccg gcc	336
Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala	
100 105 110	
ggc aac tgc gtg cac ttc gtg gcc gag gag cag gac tga	375
Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp	
115 120	

5

<210> 2

<211> 124

<212> PRT

<213> Artificial

<220>

10

<223> Construcción sintética

<400> 2

Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val
 1 5 10 15
 Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp
 20 25 30
 Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu
 35 40 45
 Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala
 50 55 60
 Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu
 65 70 75 80
 Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu
 85 90 95
 Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala
 100 105 110
 Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp
 115 120

<210> 3

<211> 375

<212> ADN

5 <213> Artificial

<220>

<223> Secuencia de resistencia a zeocina (zeo) pobre en CpG y sin ATG

<220>

10 <221> CDS

<222> (1)..(375)

<400> 3

ttg gcc aag ttg acc agt gct gtc cca gtg ctc aca gcc agg gac gtg	48
Leu Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val	
1 5 10 15	
gct gga gct gtt gag ttc tgg act gac agg ttg ggg ttc tcc aga gat	96
Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp	
20 25 30	
ttt gtg gag gac gac ttt gca ggt gtg gtc aga gac gac gtc acc ctg	144
Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu	
35 40 45	

ttc atc tca gca gtc cag gac cag gtg gtg cct gac aac acc ctg gct	192
Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala	
50 55 60	
tggtgggtggtgaga gga ctg gac gag ctg tac gct gag tgg agt gag	240
Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu	
65 70 75 80	
gtg gtc tcc acc aac ttc agg gac gcc agt ggc cct gcc ttg aca gag	288
Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Leu Thr Glu	
85 90 95	
att gga gag cag ccc tgg ggg aga gag ttt gcc ctg aga gac cca gca	336
Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala	
100 105 110	
ggc aac tgt gtg cac ttt gtg gca gag gag cag gac tga	375
Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp	
115 120	

<210> 4
 <211> 124
 <212> PRT
 <213> Artificial
 <220>
 <223> Construcción sintética
 <400> 4

5

Leu Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val	
1 5 10 15	
Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp	
20 25 30	
Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu	
35 40 45	
Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala	
50 55 60	
Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu	
65 70 75 80	
Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Leu Thr Glu	
85 90 95	
Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala	
100 105 110	
Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp	
115 120	

10

<210> 5
 <211> 804
 <212> ADN
 <213> Artificial

<220>

<223> Secuencia de resistencia a neomicina (Neo) de tipo natural

<220>

5

<221> CDS

<222> (1)..(804)

<400> 5

atg gga tcg gcc att gaa caa gat gga ttg cac gca ggt tct ccg gcc Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala 1 5 10 15	48
gct tgg gtg gag agg cta ttc ggc tat gac tgg gca caa cag aca atc Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Thr Ile 20 25 30	96
ggc tgc tct gat gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro 35 40 45	144
ggt ctt ttt gtc aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln 50 55 60	192
gac gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys 65 70 75 80	240
gca gct gtg ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu 85 90 95	288
ttg ggc gaa gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro 100 105 110	336
gcc gag aaa gta tcc atc atg gct gat gca atg cgg cgg ctg cat acg Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr 115 120 125	384
ctt gat ccg gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile 130 135 140	432
gag cga gca cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp 145 150 155 160	480
ctg gac gaa gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg 165 170 175	528
ctc aag gcg cgc atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly 180 185 190	576
gat gcc tgc ttg ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly 195 200 205	624
ttc atc gac tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile 210 215 220	672
gcg ttg gct acc cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala 225 230 235 240	720
gac cgc ttc ctc gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg 245 250 255	768
atc gcc ttc tat cgc ctt ctt gac gag ttc ttc tga Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe 260 265	804

<210> 6

<211> 267

<212> PRT

<213> Artificial

<220>

<223> Construcción sintética

5

<400> 6

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
1 5 10 15

Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
20 25 30

Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro
35 40 45

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

Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
65 70 75 80

Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
85 90 95

Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
100 105 110

Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr
115 120 125

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
130 135 140

Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp
145 150 155 160

Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
165 170 175

Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly
180 185 190

Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly
195 200 205

Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
210 215 220

Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
225 230 235 240

Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
245 250 255

Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
260 265

<210> 7

<211> 804

10

<212> ADN

<213> Artificial

<220>

<223> Gen de resistencia a neomicina modificado que carece
secuencias ATG internas

<220>

<221> CDS

5

<222> (1)..(804)

<400> 7

atg gga tcg gcc att gaa caa gac gga ttg cac gca ggt tct ccg gcc	48
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala	
1 5 10 15	
gct tgg gtg gag agg cta ttc ggc tac gac tgg gca caa cag aca atc	96
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Thr Ile	
20 25 30	
ggc tgc tct gac gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg	144
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro	
35 40 45	
gtt ctt ttt gtc aag acc gac ctg tcc ggt gcc ctg aac gaa ctg cag	192
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln	
50 55 60	
gac gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc	240
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys	
65 70 75 80	
gca gct gtg ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta	288
Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu	
85 90 95	
ttg ggc gaa gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct	336
Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro	
100 105 110	
gcc gag aaa gta tcc atc ctg gct gac gca ctg cgg cgg ctg cat acg	384
Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr	
115 120 125	
ctt gat ccg gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc	432
Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile	
130 135 140	
gag cga gca cgt act cgg ctg gaa gcc ggt ctt gtc gat cag gac gat	480
Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Leu Val Asp Gln Asp Asp	
145 150 155 160	
ctg gac gaa gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg	528
Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg	
165 170 175	
ctc aag gcg cgc ctg ccc gac ggc gac gat ctc gtc gtg acc cac ggc	576
Leu Lys Ala Arg Leu Pro Asp Gly Asp Asp Leu Val Val Thr His Gly	
180 185 190	
gac gcc tgc ttg ccg aat atc ctg gtg gaa aac ggc cgc ttt tct gga	624
Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly	
195 200 205	
ttc atc gac tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata	672
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile	
210 215 220	
gcg ttg gct acc cgt gat att gct gaa gag ctt ggc ggc gag tgg gct	720
Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala	
225 230 235 240	
gac cgc ttc ctc gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc	768
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg	
245 250 255	
atc gcc ttc tat cgc ctt ctt gac gag ttc ttc tga	804
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
260 265	

5

<210> 8
 <211> 267
 <212> PRT
 <213> Artificial
 <220>
 <223> Construcción sintética
 <400> 8

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
 1 5 10 15

Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
 20 25 30

Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro
 35 40 45

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

Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
 65 70 75 80

Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
 85 90 95

Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
 100 105 110

Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
 115 120 125

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
 130 135 140

Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Leu Val Asp Gln Asp Asp
 145 150 155 160

Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
 165 170 175

Leu Lys Ala Arg Leu Pro Asp Gly Asp Asp Leu Val Val Thr His Gly
 180 185 190

Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly
 195 200 205

Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
 210 215 220

Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
 225 230 235 240

Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
 245 250 255

Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
 260 265

10

<210> 9
 <211> 804

<212> ADN

<213> Artificial

<220>

<223> Secuencia de resistencia a Neo pobre en CpG

5

<220>

<221> CDS

<222> (1) .. (804)

<400> 9

```

atg gga agt gcc att gaa caa gac gga ttg cac gca ggt tct cct gca      48
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
1           5           10           15

gct tgg gtg gag agg cta ttt ggc tac gac tgg gca caa cag aca ata      96
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
           20           25           30

ggc tgc tct gac gca gca gtg ttc aga ctg tca gca cag ggg aga cca      144
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro
           35           40           45

gtt ctt ttt gtc aag act gac ctg tca ggt gcc ctg aac gaa ctg cag      192
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln
           50           55           60

gac gag gca gca aga cta agt tgg ctg gcc act act ggt gtt cct tgt      240
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
           65           70           75           80

gca gct gtg ttg gac gtt gtc act gaa gca gga agg gac tgg ctg cta      288
Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
           85           90           95

ttg ggt gaa gtg cct ggg cag gat ctc ctg tca tct cac ctt gct cct      336
Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
           100           105           110

gca gag aaa gta tcc atc ctg gct gac gca ctg aga aga ctg cat act      384
Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
           115           120           125

ctt gat cca gct acc tgc cca ttt gac cac caa gca aaa cat aga att      432
Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
           130           135           140

gag aga gca cga act aga ctg gaa gca ggt ctt gta gat cag gac gat      480
Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Leu Val Asp Gln Asp Asp
           145           150           155           160

ctg gac gaa gag cat cag ggg ttg gca cca gca gaa ctg ttt gcc agg      528
Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
           165           170           175

ctc aag gca aga ctg cct gac ggt gaa gat ttg gtt gtg acc cac ggt      576
Leu Lys Ala Arg Leu Pro Asp Gly Glu Asp Leu Val Val Thr His Gly
           180           185           190

gac gcc tgc ttg cct aat atc ctg gtg gaa aac ggc aga ttt tct gga      624
Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly
           195           200           205

ttc att gac tgt ggc aga ctg ggt gtg gca gac aga tat cag gac ata      672
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
           210           215           220

gca ttg gct acc aga gat att gct gaa gag ctt ggt ggt gag tgg gct      720
Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
           225           230           235           240

gac aga ttc ttg gtg ctt tac ggt ata gcc gct cct gat tca cag aga      768
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
           245           250           255

ata gcc ttc tat aga ctt ctt gac gag ttc ttc tga      804
Ile Ala Phe Tyr Arg Leu Asp Glu Phe Phe
           260           265
    
```

10

<210> 10

<211> 267
 <212> PRT
 <213> Artificial
 <220>
 5 <223> Construcción sintética
 <400> 10

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
 1 5 10 15

Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
 20 25 30

Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro
 35 40 45

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

Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
 65 70 75 80

Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
 85 90 95

Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
 100 105 110

Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
 115 120 125

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
 130 135 140

Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Leu Val Asp Gln Asp Asp
 145 150 155 160

Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
 165 170 175

Leu Lys Ala Arg Leu Pro Asp Gly Glu Asp Leu Val Val Thr His Gly
 180 185 190

Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly
 195 200 205

Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
 210 215 220

Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
 225 230 235 240

Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
 245 250 255

Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe

REIVINDICACIONES

1. Molécula de ADN que comprende una secuencia de marco de lectura abierto que codifica para un polipéptido marcador seleccionable que proporciona resistencia frente a zeocina o frente a neomicina, caracterizada porque dicha molécula de ADN en la cadena codificante para el polipéptido marcador seleccionable tiene un codón de iniciación GTG o un codón de iniciación TTG, y porque se ha mutado la secuencia de marco de lectura abierto que codifica para la proteína marcadora seleccionable para sustituir al menos la mitad de sus dinucleótidos de CpG en comparación con la secuencia nativa de marco de lectura abierto que codifica para la proteína marcadora seleccionable.
2. Molécula de ADN según la reivindicación 1, en la que dicho codón de iniciación es TTG.
3. Molécula de ADN según la reivindicación 1 ó 2, que comprende una secuencia de marco de lectura abierto que codifica para un polipéptido que proporciona resistencia frente a zeocina, comprendiendo la molécula de ADN una secuencia elegida del grupo que consiste en:
 - a) SEQ. ID. NO. 1, con la condición de que se ha sustituido al menos la mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG; y
 - b) SEQ. ID. NO. 1, en la que el nucleótido A en la posición 280 se sustituye por T, y con la condición que se ha sustituido al menos la mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG.
4. Molécula de ADN según la reivindicación 3, que comprende SEQ. ID. NO. 3.
5. Molécula de ADN según la reivindicación 1 ó 2, que comprende una secuencia de marco de lectura abierto que

codifica para un polipéptido que proporciona resistencia frente a neomicina, comprendiendo la molécula de ADN una secuencia elegida del grupo que consiste en:

- 5 a) SEQ. ID. NO. 5, con la condición de que se ha sustituido al menos la mitad de los dinucleótidos de CpG sin mutar la secuencia de aminoácidos que se codifica, y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG; y
- 10 b) SEQ. ID. NO. 7, con la condición de que se ha sustituido al menos la mitad de los dinucleótidos de CpG de la cadena codificante sin mutar la secuencia de aminoácidos que se codifica, y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG; y
- 15 c) SEQ. ID. NO. 5 o SEQ. ID. NO. 7, con la condición de que contiene una mutación que codifica para cualquiera de las siguientes variantes de polipéptido en comparación con el polipéptido codificado por las secuencias nativas:
- 20 (i) sustitución de valina en la posición 201 por glicina (201V>G), o
- (ii) sustitución de ácido glutámico en la posición 185 por ácido aspártico (185E>D), o
- 25 (iii) una combinación de ambas mutaciones (i) y (ii) (185E>D y 201V>G),
- con la condición adicional de que se ha sustituido al menos la mitad de los dinucleótidos de CpG de la cadena codificante sin mutación adicional de la
- 30 secuencia de aminoácidos que se codifica más allá de la mutación indicada en (i)-(iii), y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG.
6. Molécula de ADN según la reivindicación 5, que comprende
- 35 SEQ. ID. NO. 9, con la condición de que el nucleótido A en la posición 555 se sustituye por C, y que el nucleótido T en la posición 602 se sustituye por G y que el nucleótido G

en la posición 603 se sustituye por T, y con la condición adicional de que el codón de iniciación es o bien GTG o bien TTG.

- 5 7. Molécula de ADN según una cualquiera de las reivindicaciones 1-6, en la que la secuencia de marco de lectura abierto que codifica para un polipéptido marcador seleccionable es parte de una unidad de transcripción multicistrónica que comprende además una secuencia de marco de lectura abierto que codifica para un polipéptido de
- 10 interés.
8. Molécula de ADN según la reivindicación 7, en la que el marco de lectura abierto que codifica para el polipéptido marcador seleccionable está en el sentido de 5' del marco de lectura abierto que codifica para el polipéptido de
- 15 interés, y en la que el marco de lectura abierto que codifica para el polipéptido marcador seleccionable no tiene ninguna secuencia de ATG en la cadena codificante.
9. Molécula de ADN según la reivindicación 7, en la que el marco de lectura abierto que codifica para el polipéptido
- 20 de interés está en el sentido de 5' del marco de lectura abierto que codifica para el polipéptido marcador seleccionable, y en la que el marco de lectura abierto que codifica para el polipéptido marcador seleccionable está operativamente unido a un sitio interno de entrada al
- 25 ribosoma (IRES).
10. Casete de expresión que comprende la molécula de ADN según una cualquiera de las reivindicaciones 7-9, comprendiendo dicho casete de expresión un promotor en el sentido de 5' de dicha unidad de expresión multicistrónica y una
- 30 secuencia de terminación de la transcripción en el sentido de 3' de la unidad de expresión multicistrónica.
11. Casete de expresión según la reivindicación 10, que comprende además al menos un elemento de control de cromatina.
- 35 12. Célula huésped que comprende la molécula de ADN según una cualquiera de las reivindicaciones 1-9 o un casete de

expresión según una cualquiera de las reivindicaciones 10-11.

13. Método de generación de una célula huésped que puede expresar un polipéptido de interés, comprendiendo dicho método las etapas de:
- 5 a) introducir en una pluralidad de células precursoras una molécula de ADN según una cualquiera de las reivindicaciones 7-9 o un casete de expresión según una cualquiera de las reivindicaciones 10-11,
- 10 b) cultivar la pluralidad de células precursoras en condiciones adecuadas para la expresión del polipéptido marcador seleccionable, y
- c) seleccionar al menos una célula huésped que expresa el polipéptido de interés.
- 15 14. Método de expresión de un polipéptido de interés, que comprende cultivar una célula huésped que comprende el casete de expresión según una cualquiera de las reivindicaciones 10-11, y expresar el polipéptido de interés a partir del casete de expresión.
- 20 15. Método según la reivindicación 14, que comprende además recoger el polipéptido de interés.

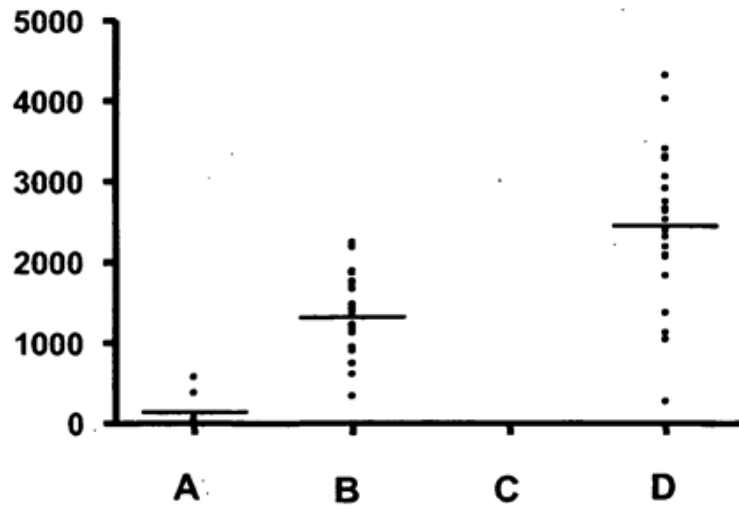
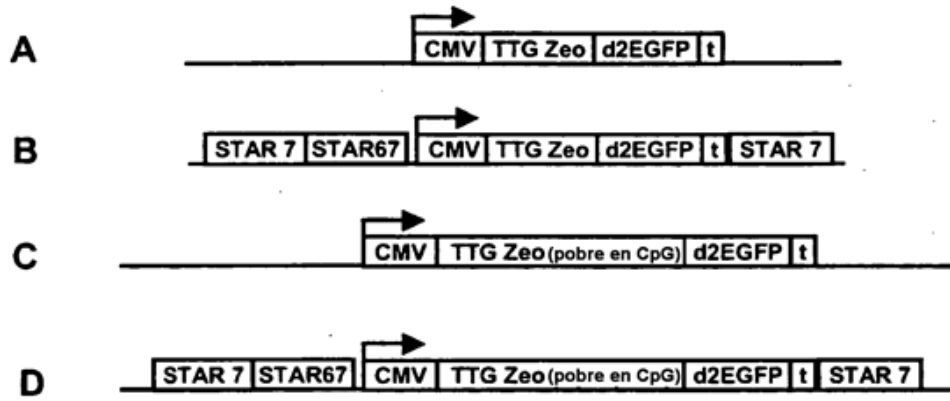


FIG. 1

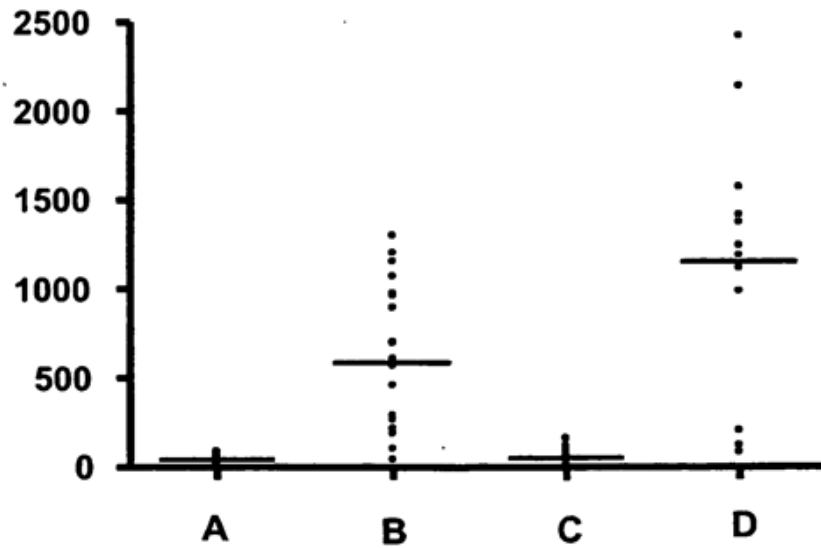
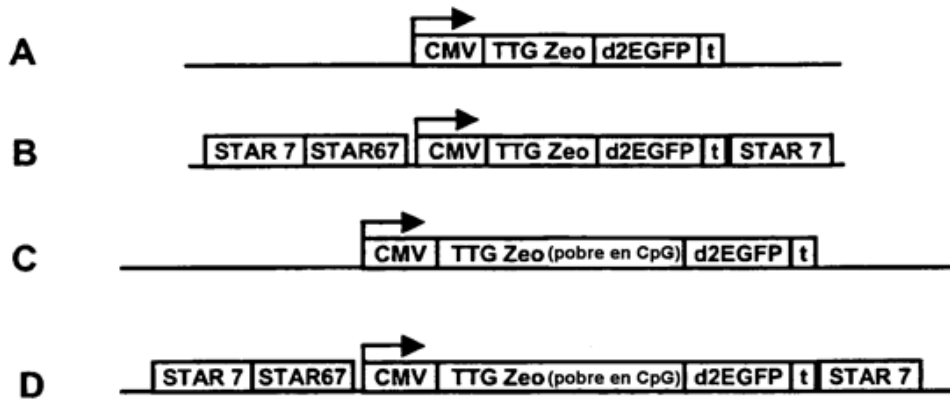


FIG. 2

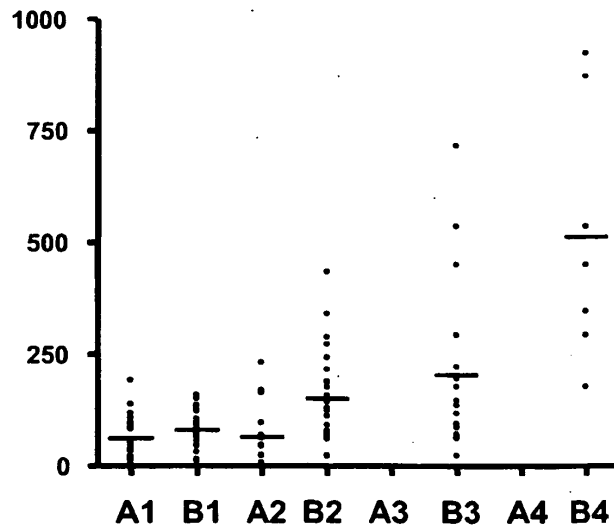
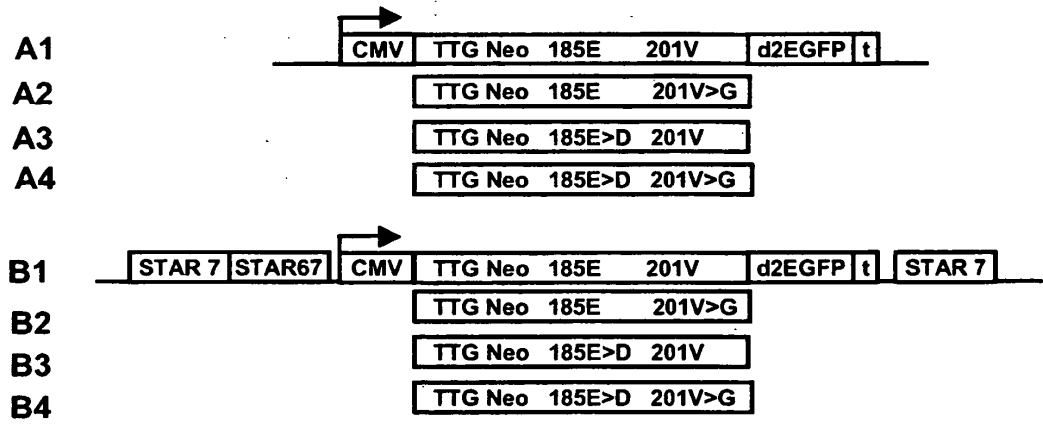


FIG. 3