



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 367 530**

51 Int. Cl.:
C12N 15/85 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07712283 .6**

96 Fecha de presentación : **21.02.2007**

97 Número de publicación de la solicitud: **1987150**

97 Fecha de publicación de la solicitud: **05.11.2008**

54 Título: **Selección de células hospedadoras que expresan proteína a altos niveles.**

30 Prioridad: **21.02.2006 US 359953**
02.05.2006 US 416490
02.05.2006 EP 06113354

45 Fecha de publicación de la mención BOPI:
04.11.2011

45 Fecha de la publicación del folleto de la patente:
04.11.2011

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 2 367 530 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 hospedadoras que expresan proteína a altos niveles.

5 **Campo de la invención**

La invención se refiere al campo de la biología molecular y la biotecnología. Más específicamente, la invención se refiere a medios y procedimientos para mejorar la selección de células hospedadoras que expresan proteínas a altos niveles.

10

Antecedentes de la invención

Las proteínas pueden producirse en diversas células hospedadoras para un amplio intervalo de aplicaciones en biología y biotecnología, por ejemplo como productos biofarmacéuticos. Se prefieren con este fin células hospedadoras eucarióticas, y particularmente de mamífero, para la expresión de muchas proteínas, por ejemplo, cuando dichas proteínas tienen ciertas modificaciones postraduccionales tales como glucosilación. Los procedimientos para dicha producción están bien establecidos y suponen generalmente la expresión en una célula hospedadora de un ácido nucleico (también designado como "transgén") que codifica la proteína de interés. En general, se introduce el transgén junto con un marcador seleccionable en una célula precursora, se seleccionan las células para la expresión del gen marcador seleccionable, se identifican el uno o más clones que expresan la proteína de interés a altos niveles y se usan para la expresión de la proteína de interés.

Un problema asociado a la expresión de transgenes es que es impredecible, partiendo de la alta probabilidad de que el transgén se vuelva inactivo debido al silenciamiento génico (McBurney y col., 2002), y por lo tanto tiene que ensayarse en muchos clones de células hospedadoras una alta expresión del transgén.

Son conocidos procedimientos para seleccionar células hospedadoras recombinantes que expresan niveles relativamente altos de proteínas deseadas, y varios de dichos procedimientos se debaten en la introducción del documento WO 2006/048459.

30

En ciertos procedimientos ventajosos de la técnica anterior, se han descrito vectores de expresión bicistrónicos para una creación rápida y eficaz de líneas celulares de mamífero estables que expresan proteína recombinante. Estos vectores contienen un sitio de entrada interna del ribosoma (IRES) entre la secuencia de codificación en dirección 5' de la proteína de interés y la secuencia de codificación en dirección 3' del marcador de selección (Rees y col., 1996). Dichos vectores están comercialmente disponibles, por ejemplo, los vectores pIRES1 de Clontech (CLONTECHniques, octubre de 1996). Usando dichos vectores para introducción en células hospedadoras, la selección de una expresión suficiente de la proteína marcadora en dirección 3' selecciona entonces automáticamente altos niveles de transcripción del ARNm multicistrónico, y por tanto se prevé una probabilidad fuertemente aumentada de alta expresión de la proteína de interés usando dichos vectores. Preferiblemente en dichos procedimientos, el IRES usado es un IRES que da un nivel relativamente bajo de traducción del gen marcador de selección, para mejorar adicionalmente la oportunidad de seleccionar células hospedadoras con un alto nivel de expresión de la proteína de interés seleccionando la expresión de la proteína marcadora de selección (véanse, por ejemplo, los documentos WO 03/106684 y WO 2006/005718).

La presente invención se dirige a proporcionar medios y procedimientos mejorados para la selección de células hospedadoras que expresan altos niveles de proteínas de interés.

Sumario de la invención

El documento WO 2006/048459 se presentó antes, pero se publicó después, de la fecha de prioridad de la presente solicitud. El documento WO 2006/048459 da a conocer un concepto para la selección de células hospedadoras que expresan altos niveles de polipéptidos de interés, designándose el concepto en la misma como "traducción interdependiente recíproca". En ese concepto, se usa una unidad de transcripción multicistrónica en la que una secuencia que codifica un polipéptido marcador seleccionable está en dirección 5' de una secuencia que codifica un polipéptido de interés, y en la que la traducción del polipéptido marcador seleccionable está dañada por mutaciones en la misma, mientras que la traducción del polipéptido de interés es muy alta (véase, por ejemplo, la FIG. 13 de la misma para una visión esquemática). La presente invención proporciona medios y procedimientos alternativos para seleccionar células hospedadoras que expresan altos niveles de polipéptido.

En un aspecto, la invención proporciona una molécula de ADN que comprende una unidad de transcripción multicistrónica que codifica i) un polipéptido de interés y ii) un polipéptido marcador seleccionable funcional en una célula hospedadora eucariótica, en la que el polipéptido de interés tiene una secuencia de iniciación de la traducción separada de la del polipéptido marcador seleccionable, y en la que la secuencia de codificación del polipéptido de interés está en dirección 5' de la secuencia de codificación del polipéptido marcador seleccionable en dicha unidad

de transcripción multicistónica, y en la que está presente un sitio de entrada interna del ribosoma (IRES) en dirección 3' a la secuencia de codificación del polipéptido de interés y en dirección 5' de la secuencia de codificación del polipéptido marcador seleccionable, y en la que la secuencia de ácido nucleico que codifica el polipéptido marcador seleccionable en la hebra de codificación comprende una secuencia de inicio de la traducción elegida del grupo constituido por: a) un codón de inicio GTG; b) un codón de inicio TTG; c) un codón de inicio CTG; d) un codón de inicio ATT y e) un codón de inicio ACG.

La secuencia de inicio de la traducción en la hebra de codificación del polipéptido marcador seleccionable comprende un codón de inicio diferente del codón de inicio ATG, tal como una de las secuencias GTG, TTG, CTG, ATT o ACG, siendo los dos primeros de estas las más preferidas. Dichos codones de inicio no ATG están flanqueados preferiblemente por secuencias que proporcionan un reconocimiento relativamente bueno de las secuencias no ATG como codones de inicio, tales como al menos la traducción de inicio de algunos ribosomas a partir de estos codones de inicio, concretamente, la secuencia de inicio de la traducción comprende preferiblemente la secuencia ACC[codón de inicio no ATG]G o GCC[codón de inicio no ATG]G.

En realizaciones preferidas, la proteína marcadora seleccionable proporciona resistencia frente a los efectos letales y/o inhibidores del crecimiento de un agente de selección tal como un antibiótico.

La invención proporciona adicionalmente módulos de expresión que comprenden una molécula de ADN según la invención, comprendiendo dichos módulos de expresión adicionalmente un promotor en dirección 5' de la unidad de expresión multicistónica, y siendo funcionales en una célula hospedadora eucariótica para la iniciación de la transcripción de la unidad de expresión multicistónica, y comprendiendo adicionalmente dichos módulos de expresión una secuencia de terminación de la transcripción en dirección 3' de la unidad de expresión multicistónica.

En realizaciones preferidas de los mismos, dichos módulos de expresión comprenden adicionalmente al menos un elemento de control de cromatina elegido del grupo constituido por una región de unión a matriz o armazón (MAR/SAR), una secuencia aislante, un elemento de apertura de cromatina ubicuo (UCOE) y una secuencia antirrepresora. Se prefieren en este aspecto secuencias antirrepresoras, y en ciertas realizaciones se eligen dichas secuencias antirrepresoras del grupo constituido por: a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66; b) fragmentos de una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en las que dichos fragmentos tienen actividad antirrepresora; c) secuencias que son al menos un 70% idénticas en la secuencia nucleotídica a a) o a b), en las que dichas secuencias tienen actividad antirrepresora y d) el complemento de una cualquiera de a) a c).

La invención proporciona también células hospedadoras que comprenden moléculas de ADN según la invención.

La invención proporciona adicionalmente procedimientos para generar células hospedadoras que expresan un polipéptido de interés, comprendiendo el procedimiento las etapas de: introducir en una pluralidad de células hospedadoras precursoras una molécula de ADN o un módulo de expresión según la invención, cultivar las células en condiciones de selección de la expresión del polipéptido marcador seleccionable y seleccionar al menos una célula hospedadora productora del polipéptido de interés.

En un aspecto adicional, la invención proporciona procedimientos para la producción de un polipéptido de interés, comprendiendo los procedimientos cultivar una célula hospedadora, comprendiendo dicha célula hospedadora un módulo de expresión según la invención, y expresar el polipéptido de interés a partir del módulo de expresión. En realizaciones preferidas de la misma, el polipéptido de interés se aísla adicionalmente a partir de las células hospedadoras y/o del medio de cultivo de células hospedadoras.

Breve descripción de los dibujos

FIG. 1. Resultados con construcciones de expresión según la invención. La construcción de expresión contiene las secuencias que codifican el polipéptido de interés (ejemplificado aquí por d2EGFP) en dirección 5' de un IRES, que está en dirección 5' de la secuencia que codifica el marcador seleccionable según la invención (ejemplificado aquí por el gen de resistencia a zeocina, con un codón de inicio TTG (TTG Zeo) (o en controles con su codón de inicio ATG normal (ATG Zeo)). Véase el ejemplo 1 para detalles. Los puntos indican los puntos de datos individuales, las líneas indican los niveles de expresión media, las construcciones usadas se indican en el eje horizontal y se exhiben esquemáticamente por encima de la gráfica, el eje vertical designa la señal de d2EGFP.

FIG. 2. Resultados con módulos de expresión tricistónicos con dhfr como marcador de mantenimiento. La construcción de expresión contiene un gen marcador seleccionable de zeocina con un codón de inicio TTG y carece de secuencias ATG internas en dirección 5' de la secuencia que codifica el polipéptido de interés (ejemplificado aquí por d2EGFP), que adicionalmente está ligado operativamente mediante un IRES con un gen dhfr marcador de selección metabólica en dirección 3' (con un codón de inicio ATG). Los puntos indican los puntos de datos individuales (señal de fluorescencia de GFP en colonias Zeo^R en el eje vertical), las líneas indican los niveles de expresión media. Se muestra la construcción usada por encima de la gráfica, las condiciones se indican en el eje

horizontal (d: día). Véase el ejemplo 2 para detalles.

FIG. 3. Como la Fig. 2, pero con el gen dhfr teniendo el codón de inicio GTG.

5 FIG. 4. Como la Fig. 2, pero con el gen dhfr teniendo el codón de inicio TTG.

FIG. 5. Números de copias de clones con la enzima dhfr (codón de inicio ATG) en diferentes condiciones. Véase el ejemplo 3 para detalles.

10 FIG. 6. Como la Fig. 5, pero con el gen dhfr teniendo el codón de inicio GTG.

FIG. 7. Como la Fig. 5, pero con el gen dhfr teniendo el codón de inicio TTG.

15 Descripción detallada de la invención

En un aspecto, la invención proporciona una molécula de ADN según la reivindicación 1. Dicha molécula de ADN puede usarse según la invención para obtener células hospedadoras eucarióticas que expresan altos niveles del polipéptido de interés seleccionando la expresión del polipéptido marcador seleccionable. Posterior o simultáneamente, pueden identificarse una o más células hospedadoras que expresan el polipéptido de interés y usarse adicionalmente para la expresión de altos niveles del polipéptido de interés.

20 El término "gen monocistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica un polipéptido. Una "unidad de transcripción multicistrónica", también designada como un gen multicistrónico, se define como un gen capaz de proporcionar una molécula de ARN que codifica al menos dos polipéptidos. El término "gen bicistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica dos polipéptidos. Por lo tanto, un gen bicistrónico está comprendido dentro de la definición de un gen multicistrónico. Un "polipéptido" como se usa en la presente memoria comprende al menos cinco aminoácidos ligados por enlaces peptídicos, y puede ser, por ejemplo, una proteína o una parte, tal como una subunidad de la misma. La mayoría de veces, los términos polipéptido y proteína se usan intercambiabilmente en la presente memoria. Un "gen" o una "unidad transcripcional" como se usa en la presente invención puede comprender ADN cromosómico, ADNc, ADN artificial, combinaciones de los mismos y similares. Las unidades de transcripción que comprenden varios cistrones se transcriben como un ARNm individual.

35 Una unidad de transcripción multicistrónica según la invención es preferiblemente una unidad de transcripción bicistrónica que codifica de 5' a 3' un polipéptido de interés y un polipéptido marcador seleccionable. Por tanto, el polipéptido de interés está codificado en dirección 5' de la secuencia de codificación del polipéptido marcador seleccionable. El IRES está ligado operativamente con la secuencia que codifica el polipéptido marcador seleccionable, y por tanto el polipéptido marcador seleccionable depende del IRES para su traducción.

40 Se prefieren usar unidades de transcripción separadas para la expresión de diferentes polipéptidos de interés, también cuando estos forman parte de una proteína multimérica (véase, por ejemplo, el ejemplo 13 del documento WO 2006/048459: la cadena pesada y ligera de un anticuerpo, cada una codificada por una unidad de transcripción separada, siendo cada una de estas unidades de expresión una unidad de expresión bicistrónica).

45 Las moléculas de ADN de la invención pueden estar presentes en forma de ADN bicatenario, que tiene con respecto al polipéptido marcador seleccionable y el polipéptido de interés una hebra codificante y una hebra no codificante, siendo la hebra codificante la hebra con la misma secuencia que el ARN traducido, excepto por la presencia de T en lugar de U. Por tanto, un codón de inicio AUG está codificado en la hebra codificante por una secuencia ATG, y la hebra que contiene esta secuencia ATG correspondiente al codón de inicio AUG en el ARN se designa como la hebra codificante del ADN. Resultará evidente para el especialista en la materia que los codones de inicio o secuencias de iniciación de la traducción están de hecho presentes en una molécula de ARN, pero que estos pueden considerarse igualmente materializados en una molécula de ADN que codifica dicha molécula de ARN; por tanto, siempre que la presente invención designe un codón de inicio o secuencia de iniciación de la traducción, se pretende incluir la correspondiente molécula de ADN que tiene la misma secuencia que la molécula de ARN menos por la presencia de una T en lugar de una U en la hebra codificante de dicha molécula de ADN y viceversa, excepto cuando se especifica explícitamente otra cosa. En otras palabras, un codón de inicio es por ejemplo una secuencia AUG en ARN, pero la correspondiente secuencia ATG en la hebra codificante de ADN se designa como codón de inicio también en la presente invención. Se usa lo mismo para la referencia de secuencias codificantes "en fase", lo que significa tripletes (3 bases) en la molécula de ARN que se traducen a un aminoácido, pero también han de interpretarse como las correspondientes secuencias trinucleotídicas en la hebra codificante de la molécula de ADN.

60 El polipéptido marcador seleccionable y el polipéptido de interés codificados por el gen multicistrónico tienen cada uno su propia secuencia de iniciación de la traducción, y por lo tanto tienen cada uno su propio codón de inicio (así como codón de terminación), concretamente, están codificados por marcos de lectura abiertos separados.

El término “marcador de selección” o “marcador seleccionable” se usa típicamente para designar un gen y/o 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 hospedadora de los efectos letales o inhibidores del crecimiento del agente (por ejemplo, un gen y/o proteína de resistencia a antibiótico). Es otra posibilidad que dicho marcador de selección induzca fluorescencia o un depósito coloreado (por ejemplo, proteína fluorescente verde (GFP) y derivados (por ejemplo, d2EGFP), luciferasa, lacZ, fosfatasa alcalina, etc.), que puede usarse para seleccionar células que expresan el polipéptido inductor del depósito coloreado, por ejemplo, usando un clasificador celular activado por fluorescencia (FACS) para seleccionar células que expresan GFP. Preferiblemente, el polipéptido marcador seleccionable según la invención proporciona resistencia frente a los efectos letales y/o inhibidores del crecimiento de un agente de selección. El polipéptido marcador seleccionable está codificado por el ADN de la invención. El polipéptido marcador seleccionable según la invención debe ser funcional en una célula hospedadora eucariótica, y por tanto poder seleccionarse en células hospedadoras eucarióticas. Puede usarse en principio según la presente invención cualquier polipéptido marcador seleccionable que cumpla este criterio. Dichos polipéptidos marcadores seleccionables son bien conocidos en la técnica y se usan rutinariamente cuando han de obtenerse clones de células hospedadoras eucarióticas, y se proporcionan varios ejemplos en la presente memoria. En ciertas realizaciones, es un marcador de selección usado para la invención la zeocina. En otras realizaciones, se usa blasticidina. El especialista en la materia sabrá que están disponibles y pueden usarse otros marcadores de selección, por ejemplo, neomicina, puromicina, bleomicina, higromicina, etc. En otras realizaciones, se usa kanamicina. En aún otras realizaciones, se usa el gen *dhfr* como marcador seleccionable, que puede seleccionarse con metotrexato, especialmente aumentando la concentración de metotrexato, pueden seleccionarse células con números aumentados de copias del gen *dhfr*. El gen *dhfr* puede usarse también para complementar la deficiencia de *dhfr*, por ejemplo, en células CHO que tienen un fenotipo *dhfr*⁻, en medio de cultivo con folato y que carece de glicina, hipoxantina y timidina. De forma similar, puede usarse el gen de glutamina sintetasa (GS), para el que es posible la selección en células que tienen insuficiente GS (por ejemplo, células NS-0) cultivando en medio sin glutamina o, como alternativa, en células que tienen suficiente GS (por ejemplo, células CHO) añadiendo un inhibidor de GS, la metionina sulfoximina (MSX). Se describen otros genes marcadores seleccionables que podrían usarse, y sus agentes de selección, por ejemplo, en la tabla 1 de la patente de EE.UU. 5.561.053; véase también Kaufman, *Methods in Enzymology*, 185: 537-566 (1990), para una revisión de estos. Si el polipéptido marcador seleccionable es *dhfr*, se cultiva la célula hospedadora en realizaciones ventajosas en un medio de cultivo que contiene folato y cuyo medio de cultivo está esencialmente desprovisto de hipoxantina y timidina, y preferiblemente también de glicina.

Cuando han de seleccionarse dos unidades de transcripción multicistrónicas según la invención en una sola célula hospedadora, cada una contiene preferiblemente la secuencia de codificación de un marcador seleccionable diferente, para permitir la selección de ambas unidades de transcripción multicistrónicas. Por supuesto, ambas unidades de transcripción multicistrónicas pueden estar presentes en una sola molécula de ácido nucleico o, como alternativa, cada una puede estar presente en una molécula de ácido nucleico separada.

El término “selección” se define típicamente como el proceso de uso de un marcador de selección/marcador selectivo y un agente de selección para identificar células hospedadoras con propiedades genéticas específicas (por ejemplo, que la célula hospedadora contenga un transgén integrado en su genoma). Resulta evidente para un especialista en la materia que son posibles numerosas combinaciones de marcadores de selección. Es un antibiótico que es particularmente ventajoso la zeocina, porque la proteína de resistencia a zeocina (zeocina-R) actúa uniéndose al fármaco y volviéndolo inofensivo. Por lo tanto, es fácil valorar la cantidad de fármaco que mata las células con bajos niveles de expresión de zeocina-R, mientras que permite sobrevivir a las de alta expresión. Todas las demás proteínas de resistencia a antibiótico de uso común son enzimas, y por tanto actúan catalíticamente (no 1:1 con el fármaco). Por tanto, el antibiótico zeocina es un marcador de selección preferido. Es otro marcador de selección preferido una enzima sintetizadora de 5,6,7,8-tetrahidrofolato (*dhfr*). Sin embargo, la invención funciona también con otros marcadores de selección.

Es un polipéptido marcador seleccionable según la invención la proteína codificada por el ácido nucleico de la invención, cuyo polipéptido puede usarse funcionalmente para la selección, por ejemplo, debido a que proporciona resistencia ante un agente de selección tal como un antibiótico. Por tanto, cuando se usa un antibiótico como agente de selección, el ADN codifica un polipéptido que confiere resistencia al agente de selección, siendo dicho polipéptido el polipéptido marcador seleccionable. Son conocidas las secuencias de ADN que codifican dichos polipéptidos marcadores seleccionables, y se proporcionan en la presente memoria varios ejemplos de secuencias de tipo natural de ADN que codifica proteínas marcadores seleccionables (por ejemplo, Fig. 26-32 del documento WO 2006/048459). Resultará evidente que los mutantes o derivados de los marcadores seleccionables pueden usarse también adecuadamente según la invención, y están por lo tanto incluidos dentro del alcance del término “polipéptido marcador seleccionable”, a condición de que la proteína marcadora seleccionable siga funcional.

Por conveniencia, y como se acepta generalmente por los especialistas en la materia, en muchas publicaciones así como en la presente memoria, a menudo el gen y la proteína que codifica resistencia ante un agente de selección se designan como el “gen (de resistencia) del agente seleccionable” o la “proteína (de resistencia) del agente de

selección”, respectivamente, aunque los nombres oficiales pueden ser diferentes, por ejemplo, el gen que codifica la proteína que confiere resistencia a neomicina (así como a G418 y kanamicina) se designa a menudo como gen (de resistencia) de neomicina (o neo’), mientras que el nombre oficial es gen de aminoglucósido 3'-fosfotransferasa.

5 Para la presente invención, es beneficioso tener bajos niveles de expresión del polipéptido marcador seleccionable, de modo que sea posible una selección rigurosa. En la presente invención, se produce esto usando una secuencia de codificación de marcador seleccionable con un codón de inicio no ATG. Tras la selección, se seleccionarán solo las células que tengan no obstante suficientes niveles de polipéptido marcador seleccionable, lo que significa que
10 dichas células deben tener suficiente transcripción de la unidad de transcripción multicistónica y suficiente traducción del polipéptido marcador seleccionable, lo que proporciona una selección de células en que la unidad de transcripción multicistónica se ha integrado o está presente de otro modo en las células hospedadoras en un lugar en que los niveles de expresión de esta unidad de transcripción son altos.

15 Las moléculas de ADN según la invención tienen la secuencia de codificación del polipéptido marcador seleccionable en dirección 3' de la secuencia de codificación del polipéptido de interés. Por tanto, la unidad de transcripción multicistónica comprende, en la dirección 5' a 3' (tanto en la hebra transcrita del ADN como en el ARN transcrito resultante) la secuencia que codifica el polipéptido de interés y la secuencia que codifica el polipéptido marcador seleccionable. El IRES está en dirección 5' de la secuencia de codificación del polipéptido marcador
20 seleccionable.

Según la invención, la región de codificación del gen de interés se traduce preferiblemente a partir del ORF dependiente de caperuza, y el polipéptido de interés se produce en abundancia. El polipéptido marcador
25 seleccionable se traduce a partir de un IRES. Para reducir la traducción del cistron marcador seleccionable, según la invención la secuencia de ácido nucleico que codifica el polipéptido marcador seleccionable comprende una mutación en el codón de inicio que reduce la eficacia de la iniciación de la traducción del polipéptido marcador seleccionable en una célula hospedadora eucariótica. Preferiblemente, se introduce por ingeniería genética un codón de inicio GTG, o más preferiblemente un codón de inicio TTG, en el polipéptido marcador seleccionable. La eficacia de traducción es menor que la de la correspondiente secuencia de tipo natural en la misma célula, concretamente, la mutación da como resultado menos polipéptido por célula por unidad de tiempo, y por tanto
30 menos polipéptido marcador seleccionable.

Una secuencia de inicio de la traducción se designa a menudo en el campo como una “secuencia Kozak”, y es una secuencia Kozak óptima RCCATGG, con el codón de inicio subrayado, siendo R una purina, concretamente A o G (véanse Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Por tanto, además del codón de inicio mismo, es relevante
35 el contexto del mismo, en particular los nucleótidos -3 a -1 y +4, y una secuencia de inicio de la traducción óptima comprende un codón de inicio óptimo (concretamente, ATG) en un contexto óptimo (concretamente, el ATG precedido directamente por RCC y seguido directamente por G). La traducción por los ribosomas es más eficaz cuando está presente una secuencia Kozak óptima (véanse Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Sin embargo, en un pequeño porcentaje de eventos, se reconocen secuencias de iniciación de la traducción no óptimas y se usan por el ribosoma para iniciar la traducción. La presente invención hace uso de este principio, y permite reducir e incluso un ajuste fino de la cantidad de traducción, y por tanto de la expresión del polipéptido marcador
40 seleccionable, lo que puede usarse por lo tanto para aumentar el rigor del sistema de selección.

El codón de inicio ATG del polipéptido marcador seleccionable de la invención se muta a otro codón, lo que se ha
45 notificado que proporciona cierta iniciación de la traducción, por ejemplo, a GTG, TTG, CTG, ATT o ACG (designados colectivamente en la presente memoria como “codones de inicio no ATG”). En realizaciones preferidas, el codón de inicio ATG se muta a un codón de inicio GTG. Esto proporciona niveles de expresión aún menores (menor traducción) que con el codón de inicio ATG intacto, pero en un contexto no óptimo. Más preferiblemente, el codón de inicio ATG se muta a un codón de inicio TTG, que proporciona niveles de expresión aún menores del polipéptido marcador seleccionable que con el codón de inicio GTG (Kozak M, 1986, 1987, 1989, 1990, 1997, 2002; véanse también los ejemplos 9-13 en el documento WO 2006/048459). No se ha dado a conocer ni sugerido en la técnica anterior el uso de codones de inicio no ATG en la secuencia de codificación de un polipéptido marcador
50 seleccionable en una unidad de transcripción multicistónica según la presente invención y, preferiblemente en combinación con elementos de control de cromatina, conduce a niveles muy altos de expresión del polipéptido de interés, como se muestra también en el documento WO 2006/048459.

Para el uso de un codón de inicio no ATG según la invención, es muy preferido proporcionar un contexto óptimo para dicho codón de inicio, concretamente, los codones de inicio no ATG preferiblemente están precedidos
60 directamente por los nucleótidos RCC en posiciones -3 a -1 y seguidos directamente por un nucleótido G (posición +4). Sin embargo, se ha notificado que usando la secuencia TTTGTGG (codón de inicio subrayado), se observa cierta iniciación al menos *in vitro*, así que aunque muy preferido, puede no ser absolutamente necesario proporcionar un contexto óptimo para los codones de inicio no ATG.

Las secuencias de ATG en la secuencia de codificación de un polipéptido, pero excluyendo el codón de inicio ATG,

se designan como "ATG internos", y si estos están en fase con el ORF y codifican por lo tanto metionina, la metionina resultante en el polipéptido se designa como "metionina interna". En la invención del documento WO 2006/048459, la región de codificación (después de codón de inicio, no incluyendo necesariamente el codón de inicio) que codifica el polipéptido marcador seleccionable está desprovista de cualquier secuencia ATG en la hebra de codificación del ADN hasta (pero sin incluir) el codón de inicio del polipéptido de interés. El documento WO 2006/048459 da a conocer cómo producir esto y cómo ensayar la funcionalidad de los polipéptidos marcadores resultantes. Para la presente invención, cuando la secuencia de codificación del polipéptido marcador seleccionable está en dirección 3' de un IRES y en dirección 3' de la secuencia de codificación del polipéptido de interés, los ATG internos en la secuencia que codifica el polipéptido marcador seleccionable pueden permanecer intactos.

Evidentemente, es muy preferido según la presente invención que la secuencia de inicio de la traducción del polipéptido de interés comprenda una secuencia de inicio de la traducción óptima, concretamente, que tenga la secuencia consenso RCCATGG (codón de inicio subrayado). Esto dará como resultado una traducción muy eficaz del polipéptido de interés.

Al proporcionar a la secuencia de codificación del marcador diferentes mutaciones que conducen a varios niveles de eficacia de traducción reducida, puede aumentarse el rigor de la selección. Es por tanto posible el ajuste fino del sistema de selección usando las unidades de transcripción multicistrónicas según la invención: por ejemplo, usando un codón GTG de inicio para el polipéptido marcador de selección, se traducirán solo unos pocos ribosomas a partir de este codón de inicio, dando como resultado bajos niveles de proteína marcadora seleccionable y por tanto un alto rigor de selección; usar un codón de inicio TTG aumenta adicionalmente el rigor de la selección debido a que aún menos ribosomas traducirán el polipéptido marcador seleccionable a partir de este codón de inicio.

Se demuestra en el documento WO 2006/048459 que las unidades de expresión multicistrónicas dadas a conocer en la presente memoria pueden usarse en un sistema de selección muy robusto, conduciendo a un porcentaje muy alto de clones que expresan el polipéptido de interés a altos niveles, según se desee. Además, los niveles de expresión obtenidos para el polipéptido de interés parecen ser significativamente mayores que los obtenidos cuando se criban un número aún mayor de colonias usando los sistemas de selección conocidos hasta ahora.

Además de una eficacia de iniciación de la traducción reducida, podría ser beneficioso proporcionar también una eficacia de alargamiento de la traducción reducida del polipéptido marcador seleccionable, por ejemplo, mutando la secuencia de codificación del mismo de modo que comprenda varios codones no preferidos de la célula hospedadora, para reducir adicionalmente los niveles de traducción del polipéptido marcador y permitir condiciones de selección aún más rigurosas, si se desea. En ciertas realizaciones, además de la mutación o mutaciones que reducen la eficacia de la traducción según la invención, el polipéptido marcador seleccionable comprende adicionalmente una mutación que reduce la actividad del polipéptido marcador seleccionable en comparación con su contrapartida de tipo natural. Eso puede usarse para aumentar el rigor de la selección más aún. Como ejemplos no limitantes, puede mutarse la prolina en posición 9 del polipéptido de resistencia a zeocina, por ejemplo, a Thr o Phe (véase, por ejemplo, en ejemplo 14 del documento WO 2006/048459), y para el polipéptido de resistencia a neomicina, pueden mutarse adicionalmente los restos aminoacídicos 182 o 261 o ambos (véase, por ejemplo, el documento WO 01/32901).

En algunas realizaciones de la invención, se dispone una denominada secuencia espaciadora en dirección 3' de la secuencia que codifica el codón de inicio del polipéptido marcador seleccionable, siendo preferiblemente dicha secuencia espaciadora una secuencia en fase con el codón de inicio, que codifica unos pocos aminoácidos y que no contiene una estructura secundaria (Kozak, 1990). Puede usarse dicha secuencia espaciadora para reducir adicionalmente la frecuencia de iniciación de la traducción si está presente una estructura secundaria en el ARN (Kozak, 1990) del polipéptido marcador seleccionable (por ejemplo, de zeocina, posiblemente de blasticidina) y por tanto aumentar el rigor del sistema de selección según la invención (véase, por ejemplo, el ejemplo 14 del documento WO 2006/048459).

Resultará evidente que puede usarse también cualquier molécula de ADN como se ha descrito, pero que tenga mutaciones en la secuencia en dirección 3' del primer ATG (codón de inicio) que codifica la proteína marcadora seleccionable, y por tanto están también comprendidas en la invención, a condición de que la proteína marcadora seleccionable codificada respectiva siga teniendo actividad. Por ejemplo, está también comprendida cualquier mutación silenciosa que no altere la proteína codificada debido a la redundancia del código genético. Están también comprendidas mutaciones adicionales que conduzcan a mutaciones aminoacídicas conservativas o a otras mutaciones, a condición de que la proteína codificada siga teniendo actividad, que puede ser menor o no que la de la proteína de tipo natural codificada por las secuencias indicadas. En particular, se prefiere que la proteína codificada sea al menos un 70%, preferiblemente al menos un 80%, más preferiblemente al menos un 90%, aún más preferiblemente al menos un 95% idéntica a las proteínas codificadas por las secuencias indicadas respectivas (por ejemplo, como se proporciona en las SEQ ID NO. 68-80 de la lista de secuencias de la presente solicitud). El ensayo de la actividad de las proteínas marcadoras seleccionables puede realizarse mediante procedimientos rutinarios.

Es un aspecto preferido de la invención proporcionar un módulo de expresión que comprende la molécula de ADN según la invención, que tiene la unidad de transcripción multicistrónica. Dicho módulo de expresión es útil para expresar secuencias de interés, por ejemplo, en células hospedadoras. Un "módulo de expresión" como se usa en la presente memoria es una secuencia de ácido nucleico que comprende al menos un promotor ligado funcionalmente con una secuencia de la que se desea la expresión. Preferiblemente, un módulo de expresión contiene adicionalmente secuencias de terminación de la transcripción y de poliadenilación. Pueden incluirse también otras secuencias reguladoras tales como potenciadores. Por tanto, la invención proporciona un módulo de expresión que comprende en el siguiente orden: 5'– promotor – unidad de transcripción multicistrónica según la invención que codifica un polipéptido de interés y en dirección 3' de la misma un polipéptido marcador seleccionable – secuencia de terminación de la transcripción – 3'. El promotor debe ser capaz de funcionar en una célula hospedadora eucariótica, concretamente, debe ser capaz de dirigir la transcripción de la unidad de transcripción multicistrónica. El promotor está así ligado operativamente con la unidad de transcripción multicistrónica. Adicionalmente, el módulo de expresión puede contener opcionalmente otros elementos conocidos en la materia, por ejemplo, sitios de corte y empalme que comprenden intrones y similares. En algunas realizaciones, está presente un intrón detrás del promotor y antes de la secuencia que codifica el polipéptido de interés. Un IRES está ligado operativamente con el cistron que contiene la secuencia que codifica el polipéptido marcador seleccionable. En realizaciones adicionales, está presente una secuencia que codifica un segundo marcador seleccionable en la unidad de transcripción multicistrónica – (concretamente, esta es al menos una unidad de transcripción tricistrónica en estas realizaciones). En realizaciones preferidas del mismo, dicha secuencia que codifica un segundo polipéptido marcador seleccionable: a) tiene una secuencia de iniciación de la traducción separada de la del polipéptido de interés, b) está dispuesta en dirección 5' de dicha secuencia que codifica un polipéptido de interés, c) no tiene una secuencia ATG en la hebra de codificación después del codón de inicio de dicho segundo polipéptido marcador seleccionable hasta el codón de inicio del polipéptido de interés y d) tiene una secuencia de inicio de la traducción no óptima, por ejemplo, un codón de inicio GTG o un codón de inicio TTG. Para dichas realizaciones, es un polipéptido marcador seleccionable preferido la enzima sintetizadora de 5,6,7,8-tetrahidrofolato (dhfr). Esto permite una selección continua de altos niveles de expresión del polipéptido de interés, como se ejemplifica en el ejemplo 2.

Para obtener la expresión de secuencias de ácido nucleico que codifican proteína, es bien conocido por los especialistas en la materia que las secuencias capaces de dirigir dicha expresión pueden estar ligadas funcionalmente con las secuencias de ácido nucleico que codifican proteína, dando como resultado moléculas de ácido nucleico recombinantes que codifican una proteína en formato expresable. En la presente invención, el módulo de expresión comprende una unidad de transcripción multicistrónica. En general, la secuencia promotora se dispone en dirección 5' de las secuencias que deberían expresarse. Están disponibles en la técnica vectores de expresión muy usados, por ejemplo, la serie de vectores pcDNA y pEF de Invitrogen, pMSCV y pTK-Hyg de BD Sciences, pCMV-Script de Stratagene, etc., que pueden usarse para obtener secuencias promotoras y/o terminadoras de la transcripción adecuadas, secuencias de poliA y similares.

Cuando la secuencia que codifica el polipéptido de interés se inserta apropiadamente con respecto a las secuencias que rigen la transcripción y traducción del polipéptido codificado, el módulo de expresión resultante es útil para producir el polipéptido de interés, designado como expresión. Las secuencias que dirigen la expresión pueden incluir promotores, potenciadores y similares, y combinaciones de los mismos. Estos deberían ser capaces de funcionar en la célula hospedadora, dirigiendo así la expresión de las secuencias de ácido nucleico que están ligadas funcionalmente con ellos. El especialista en la materia es consciente de que pueden usarse diversos promotores para obtener la expresión de un gen en células hospedadoras. Los promotores pueden ser constitutivos o regulados, y pueden obtenerse de diversas fuentes incluyendo virus, fuentes procarióticas o eucarióticas o diseñarse artificialmente. La expresión de los ácidos nucleicos de interés puede ser a partir del promotor natural o un derivado del mismo o a partir de un promotor enteramente heterólogo (Kaufman, 2000). Según la presente invención, se prefieren promotores fuertes que dan altos niveles de transcripción en las células eucarióticas de elección. Los promotores adecuados son bien conocidos y están disponibles para el especialista en la materia, y se describen varios en el documento WO 2006/048459 (por ejemplo, páginas 28-29), incluyendo el promotor inmediato temprano (IE) de CMV (designado en la presente memoria como el promotor de CMV) (obtenible, por ejemplo, a partir de pcDNA, Invitrogen), y muchos otros.

En ciertas realizaciones, una molécula de ADN según la invención es parte de un vector, por ejemplo, un plásmido. Dichos vectores pueden manipularse fácilmente mediante procedimientos bien conocidos por el especialista en la materia, y pueden diseñarse, por ejemplo, para ser capaces de replicación en células procarióticas y/o eucarióticas. Además, pueden usarse muchos vectores directamente o en forma de un fragmento deseado aislado a partir del mismo para la transformación de células eucarióticas, y se integrará en todo o en parte en el genoma de dichas células, dando como resultado células hospedadoras estables que comprenden el ácido nucleico deseado en su genoma.

Los sistemas de expresión convencionales son moléculas de ADN en forma de un plásmido recombinante o un genoma vírico recombinante. El plásmido o genoma vírico se introduce en células (hospedadoras eucarióticas) y

preferiblemente se integra en sus genomas mediante procedimientos conocidos en la materia, y se han descrito varios aspectos del mismo en el documento WO 2006/048459 (pág. 30-31).

Es ampliamente apreciado que la estructura de cromatina y otros mecanismos de control epigenético pueden influir en la expresión de transgenes en células eucarióticas (por ejemplo, Whitelaw y col., 2001). Las unidades de expresión multicistricas según la invención forman parte de un sistema de selección con un régimen de selección bastante riguroso. Este requiere generalmente altos niveles de transcripción en las células hospedadoras de elección. Para aumentar la oportunidad de encontrar clones de células hospedadoras que sobrevivan al régimen de selección riguroso, y posiblemente aumentar la estabilidad de la expresión en los clones obtenidos, será preferiblemente generalmente aumentar la predecibilidad de la transcripción. Por lo tanto, en realizaciones preferidas, un módulo de expresión según la invención comprende adicionalmente al menos un elemento de control de cromatina. Un "elemento de control de cromatina" como se usa en la presente memoria es un término colectivo para secuencias de ADN que pueden tener de algún modo un efecto sobre la estructura de cromatina y con ello sobre el nivel de expresión y/o estabilidad de la expresión de los transgenes de su vecindad (funcionan "en cis" y por tanto se disponen preferiblemente a 5 kb, más a 2 kb, aún más preferiblemente a 1 kb del transgén) en células eucarióticas. Dichos elementos se han usado a veces para aumentar el número de clones que tienen los niveles deseados de expresión transgénica. Se han descrito varios tipos de dichos elementos que pueden usarse según la presente invención en el documento WO 2006/048459 (por ejemplo, página 32-34) y, con el fin de la presente invención, se eligen elementos de control de cromatina del grupo constituido por regiones de unión a matriz o armazón (MAR/SAR), aislantes tales como el elemento aislante de beta-globina (5' HS4 del locus de beta-globina de pollo), scs, scs' y similares, un elemento de apertura de cromatina ubicuo (UCOE) y secuencias antirrepresoras (también designadas como secuencias "STAR").

Preferiblemente, dicho elemento de control de cromatina es una secuencia antirrepresora, preferiblemente elegida del grupo constituido por: a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66; b) fragmentos de una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en los que dichos fragmentos tienen actividad antirrepresora ("fragmentos funcionales"); c) secuencias que son al menos un 70% idénticas en la secuencia nucleotídica a a) o a b), en los que dichas secuencias tienen actividad antirrepresora ("derivados funcionales"); y d) el complemento de una cualquiera de a) a c). Preferiblemente, dicho elemento de control de cromatina se elige del grupo constituido por STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61) o un fragmento o derivado funcional de dichas secuencias STAR. En una realización preferida, dicha secuencia STAR es STAR67 (SEQ. ID. NO. 66) o un fragmento o derivado funcional de la misma. En ciertas realizaciones preferidas, STAR67 o un fragmento o derivado funcional de la misma se dispone en dirección 5' de un promotor que dirige la expresión de la unidad de transcripción multicistricas. En otras realizaciones preferidas, los módulos de expresión según la invención están flanqueados por ambos lados por al menos una secuencia antirrepresora, por ejemplo, por una de SEQ. ID. NO. 1 a SEQ. ID. NO. 65 en ambos lados, preferiblemente cada una con el extremo 3' de estas secuencias enfrentado a la unidad de transcripción. En ciertas realizaciones, se proporcionan módulos de expresión según la invención que comprenden en orden 5' a 3': secuencia antirrepresora A – secuencia antirrepresora B - [promotor – unidad de transcripción multicistricas según la invención (que codifica el polipéptido de interés y en dirección 3' del mismo la proteína marcadora seleccionable funcional) – secuencia de terminación de la transcripción] – secuencia antirrepresora C, en los que A, B y C pueden ser iguales o diferentes.

Se han descrito secuencias que tienen actividad antirrepresora (secuencias antirrepresoras) y las características de las mismas, así como fragmentos o derivados funcionales de las mismas, y las definiciones estructurales y funcionales de las mismas, y procedimientos para obtenerlas y usarlas, siendo dichas secuencias útiles para la presente invención, en el documento WO 2006/048459 (por ejemplo, páginas 34-38).

Para la producción de proteínas multiméricas, pueden usarse dos o más módulos de expresión. Preferiblemente, ambos módulos de expresión son módulos de expresión multicistricos según la invención, codificando cada uno una proteína marcadora seleccionable diferente, de modo que sea posible la selección de ambos módulos de expresión. Esta realización ha probado dar buenos resultados, por ejemplo, para la expresión de la cadena pesada y ligera de anticuerpos. Resultará evidente que ambos módulos de expresión pueden disponerse en una molécula de ácido nucleico o ambos pueden estar presentes en una molécula de ácido nucleico separada antes de introducirse en las células hospedadoras. Una ventaja de disponerlas en una molécula de ácido nucleico es que los dos módulos de expresión están presentes a una relación predeterminada única (por ejemplo, 1:1) cuando se introducen en células hospedadoras. Por otro lado, cuando están presentes en dos moléculas de ácido nucleico diferentes, esto permite la posibilidad de variar la relación molar de los dos módulos de expresión cuando se introducen en células hospedadoras, lo que puede ser una ventaja si la relación molar preferida es diferente de 1:1 o cuando es desconocido previamente cuál es la relación molar preferida, de modo que pueda efectuarse fácilmente por un especialista en la materia la variación de la misma y encontrar empíricamente la óptima. Según la invención, preferiblemente al menos uno de los módulos de expresión, pero más preferiblemente cada uno de ellos, comprende un elemento de control de cromatina, más preferiblemente una secuencia antirrepresora.

En otra realización, las diferentes subunidades o partes de una proteína multimérica están presentes en un solo módulo de expresión.

5 Se han descrito configuraciones útiles de antirrepresores combinados con módulos de expresión en el documento WO 2006/048459 (por ejemplo, página 40).

10 En ciertas realizaciones, se proporcionan unidades de transcripción o módulos de expresión según la invención que comprenden adicionalmente una secuencia de pausa de la transcripción (TRAP), esencialmente como se describe en las páginas 40-41 del documento WO 2006/048459. Se da un ejemplo no limitante de secuencia TRAP en la SEQ. ID. NO. 81. Se han descrito ejemplos de otras secuencias TRAP, procedimientos para encontrar estas y usos de las mismas en el documento WO 2004/055215.

15 Las moléculas de ADN que comprenden unidades de transcripción multicistricas y/o módulos de expresión según la presente invención pueden usarse para mejorar la expresión de ácidos nucleicos, preferiblemente en células hospedadoras. Los términos "célula"/"célula hospedadora" y "línea celular"/"línea celular hospedadora" típicamente se definen respectivamente como una célula y poblaciones homogéneas de la misma que pueden mantenerse en cultivo celular mediante procedimientos conocidos en la técnica, y que tienen la capacidad de expresar proteínas heterólogas u homólogas.

20 Se han descrito varias células hospedadoras ejemplares que pueden usarse en el documento WO 2006/048459 (por ejemplo, páginas 41-42), y dichas células incluyen por ejemplo células de mamífero incluyendo, pero sin limitación, células CHO, por ejemplo, CHO-K1, CHO-S, CHO-DG44, CHO-DUKXB11, incluyendo células CHO que tienen un fenotipo *dhfr*⁻, así como células de mieloma (por ejemplo, Sp2/0, NS0), células HEK 293 y células PER.C6.

25 Dichas células hospedadoras eucarióticas pueden expresar los polipéptidos deseados y a menudo se usan con ese fin. Pueden obtenerse mediante la introducción de una molécula de ADN de la invención, preferiblemente en forma de un módulo de expresión, en las células. Preferiblemente, el módulo de expresión se integra en el genoma de las células hospedadoras, pudiendo estar en diferentes posiciones en diversas células hospedadoras, y la selección proporcionará un clon en que el transgén se integra en una posición adecuada, conduciendo a un clon de célula hospedadora con las propiedades deseadas en términos de niveles de expresión, estabilidad, características de crecimiento y similares. Como alternativa, la unidad de transcripción multicistrica puede orientarse a o seleccionarse aleatoriamente para integración en una región cromosómica que sea transcripcionalmente activa, por ejemplo, detrás de un promotor presente en el genoma. La selección de células que contienen el ADN de la invención puede efectuarse seleccionando el polipéptido marcador seleccionable, usando procedimientos rutinarios conocidos por el especialista en la materia. Cuando dicha unidad de transcripción multicistrica se integra detrás de un promotor en el genoma, puede generarse *in situ* un módulo de expresión según la invención, concretamente, en el genoma de las células hospedadoras.

40 Preferiblemente, las células hospedadoras son de un clon estable que puede seleccionarse y propagarse según procedimientos estándares conocidos por el especialista en la materia. El cultivo de dicho clon puede producir el polipéptido de interés si las células comprenden la unidad de transcripción multicistrica de la invención.

45 La introducción de ácido nucleico que se va a expresar en una célula puede hacerse mediante uno de varios procedimientos, que son conocidos como tales por el especialista en la materia, dependiendo también del formato del ácido nucleico a introducir. Dichos procedimientos incluyen, pero sin limitación, transfección, infección, inyección, transformación y similares. Las células hospedadoras adecuadas que expresan el polipéptido de interés pueden obtenerse mediante selección.

50 En realizaciones preferidas, la molécula de ADN que comprende la unidad de transcripción multicistrica de la invención, preferiblemente en forma de un módulo de expresión, se integra en el genoma de la célula hospedadora eucariótica según la invención. Esto proporcionará una herencia estable de la unidad de transcripción multicistrica.

55 La selección de la presencia del polipéptido marcador seleccionable, y por tanto de la expresión, puede efectuarse durante la obtención inicial de las células. En ciertas realizaciones, el agente de selección está presente en el medio de cultivo al menos parte del tiempo durante el cultivo a concentraciones suficientes para seleccionar las células que expresan el polipéptido marcador seleccionable o a concentraciones menores. En realizaciones preferidas, el agente de selección ya no está presente en el medio de cultivo durante la fase de producción cuando se expresa el polipéptido.

60 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 polipeptídicas. Los ejemplos no limitantes de proteína de interés según la invención son enzimas, hormonas, cadenas de inmunoglobulina, proteínas terapéuticas como proteínas anticancerosas, proteínas de coagulación sanguínea tales como factor VIII, proteínas multifuncionales tales como eritropoyetina, proteínas de diagnóstico o proteínas o

fragmentos de las mismas útiles con fines de vacunación, todas conocidas por el especialista en la materia.

En ciertas realizaciones, un módulo de expresión de la invención codifica una cadena pesada o ligera de inmunoglobulina o una parte de unión a antígeno, derivado y/o análogo de la misma. En una realización preferida, se proporciona una unidad de expresión de proteína según la invención en la que dicha proteína de interés es una cadena pesada de inmunoglobulina. En aún otra realización preferida, se proporciona una unidad de expresión de proteína según la invención en la que dicha proteína de interés es una cadena ligera de inmunoglobulina. Cuando estas dos unidades de expresión de proteína están presentes en la misma célula (hospedadora), se ensambla una proteína multimérica y, más específicamente, una inmunoglobulina. Por tanto, en ciertas realizaciones, la proteína de interés es una inmunoglobulina tal como un anticuerpo, que es una proteína multimérica. Preferiblemente, dicho anticuerpo es un anticuerpo humano o humanizado. En ciertas realizaciones del mismo, es un anticuerpo de IgG, IgA o IgM. En una inmunoglobulina, pueden estar codificadas las cadenas pesada y ligera en diferentes módulos de expresión o en un solo módulo de expresión. Preferiblemente, la cadena pesada y ligera están presentes cada una en un módulo de expresión separado, que tiene cada uno su propio promotor (que puede ser el mismo o diferente para los dos módulos de expresión), comprendiendo cada uno una unidad de transcripción multicistriónica según la invención, siendo la cadena pesada y ligera el polipéptido de interés, y codificando preferiblemente cada uno una proteína marcadora seleccionable diferente, de modo que puede efectuarse la selección de módulo de expresión de ambas cadenas pesada y ligera cuando los módulos de expresión se introducen y/o están presentes en una célula hospedadora eucariótica.

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

Obviamente, las configuraciones de los módulos de expresión de la presente invención pueden usarse también cuando el objetivo último no se la producción de un polipéptido de interés, sino el ARN mismo, por ejemplo, para la producción de cantidades aumentadas de ARN a partir de un módulo de expresión, que puede usarse con fines de regular otros genes (por ejemplo, ARNi, ARN anticodificante), terapia génica, producción de proteína *in vivo*, etc.

En un aspecto, la invención proporciona un procedimiento para generar una célula hospedadora que expresa un polipéptido de interés, comprendiendo el procedimiento introducir en una pluralidad de células precursoras una molécula de ADN o un módulo de expresión según la invención, cultivar las células generadas en condiciones de selección y seleccionar al menos una célula hospedadora productora del polipéptido de interés. Las ventajas de este procedimiento novedoso son similares a las descritas para el procedimiento alternativo dado a conocer en el documento WO 2006/048459 (por ejemplo, páginas 46-47).

Aunque puede obtenerse clones que tienen números relativamente bajos de las unidades de transcripción multicistriónicas y altos niveles de expresión, el sistema de selección de la invención puede combinarse no obstante con procedimientos de amplificación para mejorar aún más los niveles de expresión. Esto puede lograrse, por ejemplo, mediante la amplificación de un gen *dhfr* cointegrado usando metotrexato, por ejemplo, disponiendo a *dhfr* en la misma molécula de ácido nucleico que la unidad de transcripción multicistriónica de la invención, o mediante cotransfección cuando el *dhfr* está en una molécula de ADN separada. El gen *dhfr* puede ser también parte de una unidad de expresión multicistriónica de la invención.

La invención proporciona también procedimientos para producir uno o más polipéptidos de interés, comprendiendo el procedimiento cultivar células hospedadoras de la invención.

El cultivo de una célula se realiza para posibilitar que metabolice y/o crezca y/o se divida y/o produzca proteínas recombinantes de interés. Esto puede lograrse mediante procedimientos bien conocidos por los especialistas en la materia e incluye, pero sin limitación, proporcionar nutrientes a la célula. Los procedimientos comprenden crecimiento adherente de superficies, crecimiento en suspensión o combinaciones de los mismos. El cultivo puede realizarse por ejemplo en platos, matraces giratorios o en biorreactores, usando sistemas discontinuos, semicontinuos, continuos tales como sistemas de perfusión y similares. Para conseguir la producción a gran escala (continua) de proteínas recombinante mediante el cultivo celular, se prefiere en la técnica tener células capaces de crecer en suspensión, y se prefiere tener células aptas para cultivarse en ausencia del suero derivado del animal o ser humano o de componentes séricos derivados del animal o ser humano.

Las condiciones de crecimiento o multiplicación de las células (véase, por ejemplo, "Tissue Culture", Academic Press, Kruse and Paterson, editores (1973)) y las condiciones de expresión del producto recombinante son conocidas para el especialista en la materia. En general, pueden encontrarse los principios, protocolos y técnicas prácticas para maximizar la productividad de cultivos de células de mamífero en "Mammalian Cell Biotechnology: a Practical Approach" (M. Butler, ed., IRL Press, 1991).

En una realización preferida, se recoge (aisla) la proteína expresada a partir de células o del medio de cultivo o de ambos. Puede purificarse adicionalmente entonces usando procedimientos conocidos, por ejemplo, filtración,

cromatografía en columna, etc., mediante procedimientos generalmente conocidos por el especialista en la materia.

El procedimiento de selección según la presente invención funciona en ausencia de elementos de control de cromatina, pero se obtienen resultados mejorados cuando se proporcionan dichos elementos a las unidades de expresión multicistrónicas. El procedimiento de selección según la presente invención funciona particularmente bien cuando se usa un módulo de expresión según la invención que comprende al menos una secuencia antirrepresora. Dependiendo del agente y de las condiciones de selección, la selección puede hacerse en ciertos casos tan rigurosa que solo unas pocas o incluso ninguna célula hospedadora sobrevive a la selección, a menos que estén presentes secuencias antirrepresoras. Por tanto, la combinación del procedimiento de selección novedoso y las secuencias antirrepresoras proporciona un procedimiento muy atractivo para obtener solo números limitados de colonias con una oportunidad mejorada en gran medida de alta expresión del polipéptido de interés en las mismas mientras que, al mismo tiempo, los clones obtenidos que comprenden los módulos de expresión con secuencias antirrepresoras proporcionan la expresión estable del polipéptido de interés, concretamente, tienden menos al silenciamiento u otros mecanismos de reducción de la expresión que los módulos de expresión convencionales.

En un aspecto, la invención proporciona una unidad de transcripción multicistrónica que tiene una configuración alternativa en comparación con la configuración dada a conocer en el documento WO 2006/048459: en la configuración alternativa de la presente invención, la secuencia que codifica el polipéptido de interés está en dirección 5' de la secuencia que codifica el polipéptido marcador seleccionable, y el polipéptido marcador seleccionable está ligado operativamente con una secuencia de iniciación de la traducción independiente de caperuza, preferiblemente un sitio de entrada interna del ribosoma (IRES). Dichas unidades de transcripción multicistrónica eran conocidas como tales (por ejemplo, Rees y col., 1996, WO 03/106684), pero no se habían combinado con un codón de inicio no ATG. Según la alternativa de la presente invención, el codón de inicio del polipéptido marcador seleccionable se cambia por un codón de inicio no ATG para reducir adicionalmente la tasa de iniciación de la traducción del marcador seleccionable. Esto conduce por lo tanto a un nivel de expresión reducido deseado del polipéptido marcador seleccionable, y puede dar como resultado células hospedadoras de selección altamente eficaces que expresan altos niveles del polipéptido de interés, como con las realizaciones dadas a conocer en el documento WO 2006/048459. Es una ventaja potencial de este aspecto alternativo de la presente invención, en comparación con las realizaciones expuestas en el documento WO 2006/048459, que la secuencia de codificación del polipéptido marcador seleccionable no necesita la modificación adicional de secuencias ATG internas, porque cualquier secuencia ATG interna en el mismo puede permanecer intacta puesto que no son ya relevantes para la traducción de polipéptidos en dirección 3' adicionales. Esto puede ser especialmente ventajoso si la secuencia de codificación del polipéptido marcador seleccionable contiene varias secuencias ATG internas, porque la tarea de cambiar estas y ensayar la funcionalidad de la construcción resultante no tiene que efectuarse para la presente invención: basta solo con la mutación del codón de inicio ATG en este caso. Se muestra a continuación en la presente memoria (ejemplo 1) que esta alternativa proporcionada por la presente invención conduce también a muy buenos resultados.

La secuencia de codificación del polipéptido marcador seleccionable en las moléculas de ADN de la presente invención está bajo el control traduccional del IRES, mientras que la secuencia de codificación de la proteína de interés está preferiblemente traducida de manera dependiente de caperuza. La secuencia de codificación del polipéptido de interés comprende un codón de terminación, de modo que la traducción del primer cistron termine en dirección 5' del IRES, estando dicho IRES ligado operativamente con el segundo cistron.

Como resultará fácilmente evidente para el especialista en la materia después de la lectura de la presente divulgación, la mayoría de las partes de estas unidades de expresión multicistrónica pueden variarse ventajosamente de forma similar a las unidades de expresión multicistrónica que tienen el orden opuesto a las secuencias de codificación del polipéptido de interés y el polipéptido marcador seleccionable (concretamente, las unidades de transcripción multicistrónicas del documento WO 2006/048459). Por ejemplo, pueden variarse los codones de inicio preferidos del polipéptido marcador seleccionable, la incorporación a módulos de expresión, las células hospedadoras, los promotores, la presencia de elementos de control de cromatina, etc., y usarse en realizaciones preferidas como se describe anteriormente. También el uso de estas unidades de expresión multicistrónica y módulos de expresión es como se describe anteriormente. Por lo tanto, este aspecto es verdaderamente una alternativa a los medios y procedimientos descritos en el documento WO 2006/048459, siendo la principal diferencia que el orden de los polipéptidos en las unidades de expresión multicistrónica está invertido, y que es ahora necesario un IRES para la traducción del polipéptido marcador seleccionable.

Como se usa en la presente memoria, un "sitio de entrada interna del ribosoma" o "IRES" designa un elemento que promueve la entrada directa interna del ribosoma al codón de iniciación, tal como normalmente un ATG, pero en esta invención preferiblemente GTG o TTG, de un cistron (una región que codifica proteína), conduciendo así a la traducción independiente de caperuza del gen. Véanse, por ejemplo, Jackson R J, Howell M T, Kaminski A (1990) *Trends Biochem. Sci.* 15 (12): 477-83) y Jackson R J y Kaminski, A. (1995) *RNA* 1 (10): 985-1000. La presente invención comprende el uso de cualquier secuencia de iniciación de la traducción independiente de caperuza, en particular, cualquier elemento IRES que sea capaz de promover la entrada directa interna del ribosoma al codón de

5 iniciación de un cistrón. “Bajo el control traduccional de un IRES” como se usa en la presente memoria significa que la traducción está asociada al IRES y procede de manera independiente de caperuza. Como se usa en la presente memoria, el término “IRES” comprende variaciones funcionales de secuencias de IRES a condición de que la variación sea capaz de promover la entrada directa interna del ribosoma al codón de iniciación de un cistrón. Como se usa en la presente memoria, “cistrón” designa una secuencia polinucleotídica o gen de una proteína, polipéptido o péptido de interés. “Ligado operativamente” designa una situación en que los componentes descritos están en una relación que les permite funcionar de la manera pretendida. Por tanto, por ejemplo, un promotor “ligado operativamente” con un cistrón está ligado de tal manera que se consigue la expresión del cistrón en condiciones compatibles con el promotor. De forma similar, una secuencia nucleotídica de un IRES ligada operativamente con un cistrón está ligada de tal manera que se consigue la traducción del cistrón en condiciones compatibles con el IRES.

15 Los elementos de sitio de unión interna del ribosoma (IRES) son conocidos de los genes víricos y de mamíferos (Martínez-Salas, 1999), y se han identificado también en cribados de oligonucleótidos sintéticos pequeños (Venkatesan y Dasgupta, 2001). El IRES del virus de la encefalomiocarditis se ha analizado con detalle (Mizuguchi y col., 2000). Un IRES es un elemento codificado en el ADN que da como resultado una estructura en el ARN transcrito a la que los ribosomas eucarióticos pueden unirse e iniciar la traducción. Un IRES permite producir dos o más proteínas a partir de una sola molécula de ARN (la primera proteína se traduce por ribosomas que se unen al ARN en la estructura de caperuza de su extremo 5', (Martínez-Salas, 1999)). La traducción de proteínas a partir de elementos IRES es menos eficaz que la traducción dependiente de caperuza: la cantidad de proteína de marcos abiertos de lectura (ORF) dependientes de IRES está en el intervalo de menos de 20 a 50% de la cantidad del primer ORF (Mizuguchi y col., 2000). La eficacia reducida de la traducción dependiente de IRES proporciona una ventaja que es explotada por esta realización de la presente invención. Además, la mutación de elementos IRES puede atenuar su actividad y reducir la expresión a partir de los ORG dependientes de IRES a menos de un 10% del primer ORF (López de Quinto y Martínez-Salas, 1998, Rees y col., 1996). Resulta por lo tanto evidente para un especialista en la materia que pueden hacerse cambios al IRES sin alterar la esencia de la función del IRES (por tanto, proporcionando un sitio de iniciación de la traducción de proteína con una eficacia de traducción reducida), dando como resultado un IRES modificado. El uso de un IRES modificado que sigue siendo capaz de proporcionar un pequeño porcentaje de traducción (en comparación con una traducción de la caperuza 5'), se incluye por lo tanto también en esta invención. La presente invención usa codones de inicio no ATG para reducir adicionalmente de forma significativa la iniciación de la traducción del ORF del marcador seleccionable, mejorando adicionalmente con ello las oportunidades de obtener una célula hospedadora preferida, concretamente, una célula hospedadora que exprese altos niveles de la proteína recombinante de interés.

35 Las patentes de EE.UU. 5.648.267 y 5.733.779 describen el uso de una secuencia marcadora seleccionable dominante con una secuencia Kozak de consenso dañada ([Py]xxATG[Py], en la que [Py] es un nucleótido de pirimidina (concretamente, C o T), x es un nucleótido (concretamente, G, A, T o C) y el codón de inicio ATG está subrayado). La patente de EE.UU. 6.107.477 describe el uso de una secuencia Kozak no óptima (AGATCTTTATGGACC, en la que el codón de inicio ATG no está subrayado) de un gen marcador seleccionable. Ninguna de estas patentes describe el uso de un codón de inicio no ATG ni proporciona ninguna sugerencia de hacerlo. Además, no mencionan combinaciones con un IRES. Además, puesto que un IRES tiene por sí mismo una iniciación de la traducción reducida en comparación con la traducción dependiente de caperuza, no podía preverse antes de la presente invención que la combinación de un IRES con un codón de inicio no ATG del marcador seleccionable pudiera proporcionar suficiente traducción del polipéptido marcador seleccionable para dar algún nivel seleccionable del mismo. La presente invención muestra que ese es el caso y proporciona sistemas de selección sorprendentemente eficaces.

50 La invención proporciona también una molécula de ADN que comprende una secuencia que codifica un polipéptido marcador seleccionable ligado operativamente con una secuencia IRES, en la que la secuencia de codificación que codifica el polipéptido marcador seleccionable comprende una secuencia de inicio de la traducción seleccionada del grupo constituido por: a) un codón de inicio GTG, b) un codón de inicio TTG, c) un codón de inicio CTG, d) un codón de inicio ATT y e) un codón de inicio ACG.

55 El especialista en la materia entenderá que son posibles modificaciones adicionales de la presente invención, por ejemplo, las dadas en el documento US 2006/0195935, particularmente los ejemplos 20-27 del mismo.

60 En ciertas realizaciones, puede usarse la enzima dihidrofolato reductasa (dhfr) sintetizadora de 5,6,7,8-tetrahidrofolato de mamífero como marcador de selección en células que tienen un fenotipo *dhfr* (por ejemplo, células CHO-DG44), omitiendo hipoxantina y timidina (y preferiblemente también glicina) del medio de cultivo e incluyendo folato (o ácido (dihidro)fólico) en el medio de cultivo (Simonsen y col., 1988). El gen *dhfr* puede derivar, por ejemplo, del genoma de ratón o del ADNc de ratón y puede usarse según la invención, preferiblemente proporcionando un codón de inicio GTG o TTG (véase la SEQ. ID. NO. 73 para la secuencia del gen *dhfr*). En todas estas realizaciones, se entiende por “omitir del medio de cultivo” que el medio de cultivo tiene que estar esencialmente desprovisto del componente o componentes indicados, lo que significa que hay insuficiente componente indicado presente para mantener el crecimiento de las células en el medio de cultivo, de modo que es

posible una buena selección cuando la información genética de la enzima indicada se expresa en las células y el componente precursor indicado está presente en el medio de cultivo. Por ejemplo, el componente indicado está presente a una concentración de menos de un 0,1% de la concentración de aquel componente que se usa normalmente en el medio de cultivo para un cierto tipo celular. Preferiblemente, el componente indicado está ausente del medio de cultivo. Puede prepararse un medio de cultivo que carece del componente indicado según procedimientos estándares por el especialista en la materia, o puede obtenerse a partir de suministradores de medios comerciales. Una ventaja potencial del uso de estos tipos de enzimas metabólicas como polipéptidos marcadores seleccionables es que pueden usarse para mantener las unidades de transcripción multicistrónicas bajo selección continua, lo que puede dar como resultado una mayor expresión del polipéptido de interés.

En otro aspecto, la invención usa el marcador de selección metabólica dhfr como marcador de selección adicional en una unidad de transcripción multicistrónica según la invención. En dichas realizaciones, la selección de clones de células hospedadoras con alta expresión se establece en primer lugar mediante el uso de, por ejemplo, un marcador de selección antibiótico, por ejemplo, zeocina, neomicina, etc., cuyas secuencias de codificación tendrán un codón de inicio GTG o TTG según la invención. Después de la selección de los clones adecuados, se interrumpe la selección de antibiótico, y puede efectuarse entonces la selección continua o intermitente usando el marcador de selección enzimático cultivando las células en medio que carece de los componentes identificados apropiados descritos anteriormente y que contiene los componentes precursores apropiados descritos anteriormente. En este aspecto, el marcador de selección metabólica está ligado operativamente con un IRES, y puede tener su contenido normal de ATG, y el codón de inicio puede elegirse adecuadamente de GTG o TTG. Las unidades de transcripción multicistrónicas en este aspecto son al menos tricistrónicas.

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

La invención se explica adicionalmente en los siguientes ejemplos. Los ejemplos no limitan la invención en modo alguno. Sirven simplemente para clarificar la invención.

Ejemplos

El ejemplo 1 describe el sistema de selección con la unidad de transcripción multicistrónica de la presente invención, y resultará evidente que las variaciones descritas en los ejemplos 8-26 del documento WO 2006/048459 pueden aplicarse también y ensayarse para las unidades de transcripción multicistrónicas de la presente solicitud. Lo mismo es válido para las de los ejemplos 20-27 del documento US 2006/0195935.

Ejemplo 1: Selección rigurosa disponiendo un gen de resistencia a zeocina modificado detrás de una secuencia IRES

Los ejemplos 8-26 del documento WO 2006/048459 han mostrado un sistema de selección en que una secuencia que codifica una proteína marcadora seleccionable está en dirección 5' de una secuencia que codifica una proteína de interés en una unidad de transcripción multicistrónica, y en el que la secuencia de iniciación de la traducción del marcador seleccionable no es óptima, y en el que se han retirado los ATG internos adicionales de la secuencia de codificación del marcador seleccionable. Este sistema da como resultado un sistema de selección de alto rigor. Por ejemplo, se ha mostrado que el marcador de selección Zeo en el que se cambia el codón de iniciación de la traducción por TTG da un rigor de selección muy alto y muy altos niveles de expresión de la proteína de interés codificada en dirección 3'.

En otro posible sistema de selección (concretamente, el sistema de la presente invención), se dispone el marcador de selección, por ejemplo Zeo, en dirección 3' de una secuencia IRES. Esto crea un ARNm multicistrónico del que se traduce el producto del gen Zeo mediante iniciación dependiente de IRES. En la construcción d2EGFP-IRES-Zeo habitual (concretamente, una construcción de la técnica anterior, por ejemplo, el documento WO 2006/005718), el codón de inicio de Zeo es el ATG óptimo. Se ensayó si cambiar el codón de inicio de Zeo ATG, por ejemplo, a TTG (designado como IRES-TTG Zeo) da como resultado rigores de selección aumentados en comparación con el IRES-ATG Zeo habitual.

Resultados

Se muestran esquemáticamente en la FIG. 1 las construcciones usadas. La construcción de control consistía en un promotor de CMV, el gen d2EGFP, una secuencia IRES (la secuencia del IRES usado (Rees y col., 1996) en este ejemplo era:

GCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGAATAAGGCCGGTGTGCGTTTGTCTAT
 ATGTGATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGC
 ATTCCTAGGGGTCTTCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCT
 5 GGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCCTTTCAGGGCAGCGGAACCCCCACCTGGCGACAGG
 TGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAAGTGCCACGTTGTGA
 GTACCCCATTTGATGGGATCTGATCTGGGGCCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAA
 CGTCTAGGCCCCCGAACCCACGGGGACGTGGTTTTCTTTGAAAAACACGATGATAAGCTTGCCACAACCCCGG
 10 GATA; SEQ. ID. NO. 82), y un marcador de selección TTG Zeo, concretamente, el gen de resistencia a zeocina con
 un codón de inicio TTG ("d2EGFP-IRES-TTG Zeo"). La otra construcción era la misma, pero con una combinación
 de STAR7 y STAR67 dispuestos en dirección 5' del módulo de expresión y STAR7 en dirección 3' del módulo
 ("STAR7/67 d2EGFP-IRES-TTG Zeo STAR7"). Se transfectaron ambas construcciones a células CHO-K1 y se
 15 efectuó la selección con zeocina 100 µg/ml en el medio de cultivo. Surgieron cuatro colonias después de la
 transfección con la construcción de control y seis con la construcción que contiene STAR. Se aislaron y propagaron
 estas colonias independientes antes del análisis de los niveles de expresión de d2EGFP. Como se muestra en la
 FIG. 1, la incorporación de elementos STAR a la construcción dio como resultado la formación de colonias con altos
 niveles de expresión de d2EGFP. De las colonias de control sin elementos STAR ("d2EGFP-IRES-TTG Zeo"), solo
 una colonia exhibía cierta expresión de d2EGFP. Los niveles de expresión son también mucho mayores que los
 20 obtenidos con otras construcciones de control que contienen el IRES con un Zeo normal con codón de inicio ATG
 estándar, con o sin elementos STAR ("d2EGFP-IRES-ATG Zeo" y "STAR 7/67 d2EGFP-IRES-ATG Zeo STAR7";
 también en estas construcciones de ATG Zeo había un efecto potenciador de los elementos STAR, pero estos son
 modestos en comparación con la variante TTG Zeo novedosa).

Estos resultados muestran que disponer un marcador de selección Zeo con un codón de inicio TTG en dirección 3'
 25 de una secuencia URES, en combinación con elementos STAR, funciona bien y establece un sistema de selección
 riguroso.

A partir de estos datos y los ejemplos 8-26 del documento WO 2006/048459 y 20-27 del documento US
 2006/0195935, resultará evidente que el marcador puede variar de forma similar a los ejemplos 8-26 del documento
 30 WO 2006/048459 y 20-27 del documento US 2006/0195935. Por ejemplo, en lugar de un codón de inicio TTG,
 puede usarse un codón de inicio GTG, y el marcador puede cambiarse de Zeo a un marcador diferente, por ejemplo,
 Neo, Blas, dhfr, puro, etc., todos con GTG o TTG como codón de inicio. Los elementos STAR pueden variarse
 usando diferentes secuencias STAR o diferentes disposiciones de los mismos, o sustituyéndolos por otros
 35 elementos de control de cromatina, por ejemplo, secuencias MAR. Esto conduce a mejoras frente a los sistemas de
 selección de la técnica anterior que tienen un IRES con un marcador con un codón de inicio ATG normal.

Como ejemplo no limitante, en lugar del gen de resistencia Zeo modificado (TTG Zeo), se dispone un gen de
 resistencia a neomicina modificado en dirección 3' de una secuencia IRES. La modificación consiste en el reemplazo
 40 del codón de iniciación de la traducción ATG de la secuencia de codificación de Neo por un codón de iniciación de la
 traducción TTG, creando TTG Neo. La construcción CMV-d2EGF-IRES-TTG Neo, rodeada por elementos STAR o
 no, se transfecta en células CHO-K1. Se recogen las colonias, se propagan las células y se miden los valores de
 d2EGFP. Este ("IRES-TTG Neo") conduce a mejoras frente al sistema de selección conocido que tiene Neo con un
 codón de inicio ATG en dirección 3' de una IRES ("IRES-ATG Neo"). La mejora es especialmente evidente cuando la
 45 construcción TTG Neo comprende elementos STAR.

Ejemplo 2: Estabilidad de la expresión disponiendo un gen dhfr modificado detrás de una secuencia IRES

La modificación del codón de iniciación de la traducción del marcador de selección de zeocina por un codón de
 50 iniciación de la traducción que se usa mucho menos frecuentemente que el codón ATG habitual da como resultado
 un sistema de selección de alto rigor. En el sistema de selección descrito en el documento WO 2006/048459, se
 dispone el TTG Zeo en dirección 5' del gen de interés. En otro posible sistema de selección, se dispuso el marcador
 de selección Zeo en dirección 3' de una secuencia IRES (presente solicitud, véase el ejemplo 1). Esto crea un
 ARNm bicistrónico del que se traduce el producto del gen Zeo a partir de codones de iniciación de la traducción de la
 55 secuencia IRES.

En este experimento, se combinaron realizaciones de estos dos sistemas. Se dispuso un marcador de selección
 TTG en dirección 5' del gen informador y se acopló un marcador metabólico modificado con GTG o TTG con un
 IRES con el gen informador. Pueden usarse diferentes genes marcadores de selección, tales como los genes de
 resistencia a zeocina y neomicina, así como el gen dhfr. Se dispuso aquí un gen de resistencia a zeocina modificado
 60 TTG Zeo (véase el documento WO 2006/048459), en dirección 5' de un gen de interés y el gen de selección dhfr en
 dirección 3' del gen de interés, acoplado por un IRES (Fig. 2). El objetivo de este módulo de expresión era
 seleccionar un clon de célula de mamífero productor de un alto nivel de proteína, en primer lugar mediante selección
 por zeocina. La configuración TTG Zeo-gen de interés consigue muy eficazmente este objetivo. Después de esta
 fase de selección inicial, se emplean las características de esta proteína dhfr para conseguir el mantenimiento de

altos niveles de expresión en ausencia del antibiótico zeocina.

La presión de selección activa parece beneficiosa para mantener los niveles de expresión de proteína en una colonia seleccionada por TTG Zeo al mismo alto nivel durante un periodo prolongado de tiempo. Esto puede lograrse, por ejemplo, manteniendo una cantidad mínima de zeocina en el medio de cultivo, pero esto no está favorecido en instalaciones industriales por razones económicas y potencialmente regulatorias (la zeocina es tanto tóxica como costosa).

Otro enfoque es acoplar el gen de interés con un marcador de selección que sea una enzima que metabolice una o más etapas esenciales en una ruta metabólica. Con esencial, se entiende que la célula no es capaz de sintetizar bloques de construcción metabólicos esenciales específicos por sí misma, implicando que estos bloques de construcción tienen que estar presentes en el medio de cultivo para permitir sobrevivir a la célula. Son ejemplos bien conocidos los aminoácidos esenciales que no pueden sintetizarse por una célula de mamífero y que tienen que estar presentes en el medio de cultivo para permitir sobrevivir a la célula. Otro ejemplo está relacionado con el gen dhfr sintetizador de 5,6,7,8-tetrahidrofolato. La correspondiente enzima dhfr es una enzima de la ruta del folato. La proteína dhfr convierte específicamente folato en 5,6,7,8-tetrahidrofolato, una lanzadera de grupos metilo necesaria para la síntesis *de novo* de purinas (hipoxantina), ácido timidílico (timidina) y el aminoácido glicina. Para funcionar, tiene que estar presente en el medio de cultivo la sustancia no tóxica folato (Urlaub y col., 1980). Además, el medio tiene que carecer de hipoxantina y timidina, puesto que cuando estas están disponibles para la célula, se evita la necesidad de enzima dhfr. Las células CHO-DG44 carecen del gen dhfr y estas células necesitan por lo tanto glicina, hipoxantina y timidina en el medio de cultivo para sobrevivir. Sin embargo, si los productos finales glicina, hipoxantina y timidina están ausentes del medio de cultivo, está presente folato y se proporciona el gen dhfr porque está presente en un módulo de expresión de la célula, la célula puede convertir folato en 5,6,7,8-tetrahidrofolato y sobrevivir por tanto en este medio de cultivo. Este principio se ha usado durante muchos años como metodología de selección para crear líneas celulares de mamífero transfectadas establemente.

Aquí se usa este principio no para seleccionar los clones estables inicialmente (esto se hace con zeocina), sino para mantener las células bajo presión de selección metabólica. La ventaja es que puede conseguirse una expresión de proteína muy alta mediante el sistema de selección TTG Zeo, y que estos altos niveles de expresión pueden mantenerse sin necesidad de mantener la zeocina en el medio de cultivo. En lugar de ello, la zeocina puede omitirse del medio y la ausencia de glicina, hipoxantina y timidina (GHT) o solo hipoxantina y timidina (HT) del medio de cultivo es suficiente para mantener la presión de selección suficientemente alta para garantizar altos niveles de expresión de proteína. Dicha configuración requiere la presencia de dos marcadores de selección, tanto el gen de resistencia a zeocina como el gen dhfr tienen que estar presentes en el módulo de expresión. Como se indica anteriormente, esto se consigue eficazmente cuando ambos genes están presentes con el gen de interés en una configuración tal que se transcribe un ARNm tricistrónico a partir de un solo promotor. Cuando se emplea el gen de resistencia a zeocina modificado (TTG Zeo) en dirección 5' del gen d2EGFP, el gen dhfr tiene que estar en dirección 3' acoplado con el gen d2EGFP, por ejemplo, mediante una secuencia IRES (Fig. 1).

Resultados

Se prepararon construcciones en que se dispuso el marcador de selección TTG Zeo en dirección 5' del gen informador d2EGFP y el marcador de selección dhfr en dirección 3' del gen d2EGFP, acoplado mediante una secuencia IRES (Fig. 2). Estas construcciones estaban flanqueadas por STAR 7/67/7. Se prepararon tres versiones de estas construcciones: ATG dhfr, GTG dhfr o TTG dhfr, indicando cada nombre el codón de inicio usado para el gen dhfr. Se transfectaron las construcciones en células CHO-DG44. Se transfectó el ADN usando Lipofectamine 2000 (Invitrogen) y se cultivaron las células en presencia de zeocina 400 µg/ml en medio IMDM (Gibco) + 10% de FBS (Gibco) + suplemento de HT.

El valor medio de d2EGFP en 14 clones de TTG Zeo IRES **ATG** dhfr era de 341 (día 1), cuando se medía en presencia de zeocina 400 µg/ml (Fig. 2). Después de estas medidas, se dividieron las células y se cultivaron adicionalmente en tres condiciones:

- (1) con zeocina 400 µg/ml y con hipoxantina y timidina (suplemento HT) en el medio,
- (2) sin zeocina, pero con suplemento de HT en el medio,
- (3) sin zeocina y sin suplemento de HT.

En resumen, en la condición 1, las células están solo bajo presión de selección de zeocina, en la condición 2 las células no están bajo presión de selección y en la condición 3 las células permanecen bajo presión de selección por dhfr. La última condición 3 requiere la expresión continua del gen dhfr para permitir como resultado la expresión de la proteína dhfr y la supervivencia celular.

Después de 65 días, se midieron de nuevo los valores de d2EGFP. El valor medio de d2EGFP en los clones TTG Zeo IRES **ATG** dhfr bajo selección con zeocina era ahora de 159 (Fig. 2). El valor medio de d2EGFP en los clones TTG Zeo IRES ATG dhfr sin zeocina y con suplemento de HT era de 20 (Fig. 2). El valor medio de d2EGFP en los

clones TTG Zeo IRES ATG dhfr sin selección por zeocina y sin suplemento de HT era de 37 (Fig. 2). Globalmente, se observó una caída de los valores de d2EGFP, pero mayor en ausencia de zeocina, independientemente de si estaba presente suplemento de HT o no.

5 Se siguió el mismo protocolo con la construcción TTG Zeo IRES **GTG** dhfr. El valor medio de d2EGFP en 15 clones TTG Zeo IRES GTG dhfr era de 455 (día 1), cuando se medía en presencia de zeocina 400 µg/ml (Fig. 3). Después de estas medidas, se dividieron las células y se cultivaron adicionalmente en las tres condiciones anteriormente descritas. Después de 65 días, se midieron de nuevo los valores de d2EGFP. El valor medio de d2EGFP en los clones TTG Zeo IRES GTG dhfr con selección por zeocina era ahora de 356 (Fig. 3). El valor medio de d2EGFP en los clones TTG Zeo IRES GTG dhfr sin selección por zeocina y con suplemento de HT era de 39 (Fig. 3). El valor medio de d2EGFP en los clones TTG Zeo IRES GTG dhfr sin selección por zeocina y sin suplemento de HT era de 705 (Fig. 3).

15 En este caso, se observó por tanto una caída de los valores de d2EGFP solo en ausencia de zeocina y en presencia del suplemento de HT (condición 2). En ausencia de zeocina, pero en ausencia también de suplemento de HT, los valores de d2EGFP se volvieron en realidad significativamente mayores (condición 3). Esto puede indicar que los niveles de expresión de la proteína dhfr, debido a la frecuencia de traducción dañada del ARNm de GTG dhfr, son suficientemente bajos para proporcionar un rigor de selección muy alto. Esta presión de selección, en ausencia de cualquier agente tóxico, es suficientemente alta para mantener altos niveles de expresión con el tiempo, y aparentemente mejora incluso estos niveles de expresión con el tiempo.

25 Se hizo lo mismo para la construcción TTG Zeo IRES **TTG** dhfr. El valor medio de d2EGFP en 18 clones TTG Zeo IRES TTG dhfr era de 531 (día 1), cuando se medía en presencia de zeocina 400 µg/ml (Fig. 4). Después de estas medidas, se dividieron las células y se cultivaron adicionalmente en las tres condiciones anteriormente descritas. Después de 65 días, se midieron de nuevo los valores de d2EGFP. El valor medio de d2EGFP en los clones TTG Zeo IRES TTG dhfr con selección por zeocina era ahora de 324 (Fig. 4). El valor medio de d2EGFP en los clones TTG Zeo IRES TTG dhfr sin selección por zeocina y en presencia de suplemento de HT era de 33 (Fig. 4). El valor medio de d2EGFP en los clones TTG Zeo IRES TTG dhfr sin selección por zeocina y sin suplemento con HT era de 1124 (Fig. 4).

30 Se observó de nuevo una caída de los valores de d2EGFP solo en ausencia de zeocina y en presencia de suplemento de HT (condición 2). En ausencia de zeocina, pero en ausencia de suplemento de HT, los valores de d2EGFP se volvieron aún mayores que con la construcción TTG Zeo IRES GTG dhfr (condición 3). Puesto que la variante TTG es más rigurosa que la variante GTG, se espera que se traduzcan aún menos proteína dhfr con TTG dhfr que con la variante GTG dhfr. La presión de selección aumentada en ausencia de cualquier agente tóxico con la variante TTG dhfr es suficientemente alta para mantener altos niveles de expresión de proteína con el tiempo, y aparentemente también mejora adicionalmente los niveles de expresión de proteína con el tiempo.

40 Los datos muestran que el acoplamiento de una variante de codón de inicio no ATG del gen dhfr mediante un IRES al gen d2EGFP permite un alto grado de estabilidad de una alta expresión de d2EGFP en células CHO-DG44. Esto ocurre en medio de cultivo sin zeocina y sin productos finales metabólicos esenciales. La selección previa con zeocina mediante el marcador de selección TTG Zeo modificado permite el establecimiento eficaz de colonias con altos niveles de expresión de d2EGFP. Ahora se requiere solo un sencillo cambio del medio de cultivo (retirando zeocina y HT) para mantener los altos niveles de expresión de d2EGFP, e incluso mejorar estos niveles de expresión.

Ejemplo 3: La expresión aumentada disponiendo un gen dhfr modificado detrás de una secuencia IRES debilitada no es el resultado de la amplificación génica

50 El uso del gen dhfr como marcador de selección en la técnica anterior se basaba a menudo en la amplificación del gen dhfr. Se usó un agente tóxico, metotrexato, en dichos sistemas para amplificar el gen dhfr y al mismo tiempo con él el transgén deseado, del que podrían encontrarse integradas hasta muchos miles de copias en el genoma de células CHO después de dicha amplificación. Aunque estos altos números de copias conducen a altos niveles de expresión, se consideran también una desventaja porque tantas copias pueden conducir a una inestabilidad genómica aumentada y la retirada adicional de metotrexato del medio de cultivo conduce a la rápida retirada de muchos de los loci amplificados.

60 En el ejemplo 2, no se usó metotrexato para inhibir la actividad de la enzima dhfr. Solo se retiraron hipoxantina y precursor de timidina del medio de cultivo, y esto fue suficiente para conseguir tanto la estabilidad de la expresión de proteína como niveles de expresión incluso aumentados. Por lo tanto, se determinó si el empleo de la enzima dhfr en este entorno daba como resultado la amplificación génica.

Resultados

Se aisló ADN de los clones que se describieron en el ejemplo 2, el mismo día (65) que se midieron los valores de d2EGFP. Con este ADN, se determinaron los números de copias de d2EGFP.

El número medio de copias de d2EGFP en los clones TTG Zeo IRES **ATG** dhfr con selección por zeocina era de 86 (condición 1) (Fig. 5). El número medio de copias de d2EGFP en los clones TTG Zeo IRES ATG dhfr sin selección por zeocina y en presencia del suplemento de HT era de 53 (condición 2) (Fig. 5). El número medio de copias de d2EGFP en los clones TTG Zeo IRES ATG dhfr sin selección por zeocina y sin suplemento de HT era de 59 (condición 3) (Fig. 5).

El número medio de copias de d2EGFP en los clones TTG Zeo IRES **GTG** dhfr con selección por zeocina era de 23 (condición 1) (Fig. 6). El número medio de copias de d2EGFP en los clones TTG Zeo IRES GTG dhfr sin selección por zeocina y en presencia del suplemento de HT era de 14 (condición 2) (Fig. 6). El número medio de copias de d2EGFP en los clones TTG Zeo IRES GTG dhfr sin selección por zeocina y sin suplemento de HT era de 37 (condición 3) (Fig. 6).

El número medio de copias de d2EGFP en los clones TTG Zeo IRES **TTG** dhfr con selección por zeocina era de 33 (condición 1) (Fig. 7). El número medio de copias de d2EGFP en los clones TTG Zeo IRES TTG dhfr sin selección por zeocina y en presencia del suplemento de HT era de 26 (condición 2) (Fig. 7). El número medio de copias de d2EGFP en los clones TTG Zeo IRES TTG dhfr sin selección por zeocina y sin suplemento de HT era de 32 (condición 3) (Fig. 7).

En ningún caso se observó un fuerte aumento de los números de copias de d2EGFP después de la retirada del suplemento de HT, lo que daba como resultado valores de d2EGFP aumentados en el caso de la variante GTG dhfr y TTG dhfr. El hecho de que con ambas construcciones los valores de d2EGFP permanecieran estables con el tiempo e incluso aumentaran significativamente debe ser debido a la acción de la proteína dhfr. Es más, no se observaron números de copias de d2EGFP aumentados en los clones TTG Zeo TTG dhfr en absoluto, y solo un modesto aumento en los clones TTG Zeo GTG dhfr. De forma interesante, los números de copias de d2EGFP globales en los menores productores, los clones TTG Zeo **ATG** dhfr, eran mayores que en las otras dos variantes, mientras que estos clones no mantenían los valores de fluorescencia de d2EGFP iniciales altos (véase el ejemplo 2). Se concluye a partir de estos datos que la amplificación génica conocida comúnmente, observada cuando se usa la proteína dhfr en combinación con la adición de metotrexato, no es la responsable de mantener los niveles de expresión de d2EGFP estables con el tiempo y del aumento observado de estos niveles de expresión. En lugar de ello, parece que se expresa más proteína d2EGFP por copia del gen d2EGFP con las variantes GTG y TTG dhfr.

Se han analizado adicionalmente los niveles de ARNm de d2EGFP de los diferentes clones y en diferentes condiciones como anteriormente, y se ha encontrado que estos niveles de ARNm seguían ampliamente la tendencia de los valores de fluorescencia de d2EGFP. Por lo tanto, se concluyó que los aumentos de los valores de fluorescencia de d2EGFP son debidos a niveles de ARNm aumentados, y no a eficacias de traducción alteradas.

Referencias

Kaufman, R.J. (2000) "Overview of vector design for mammalian gene expresión" Mol. Biotechnol. 16, 151-160.

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

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.

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 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.

López de Quinto, S y Martínez-Salas, E. (1998) "Parameters influencing translational efficiency in aphthovirus IRES-based bicistronic expresión vectors" Gene 217, 51-6.

- Martínez-Salas, E. (1999) "Internal ribosome entry site biology and its use in expresión vectors" Curr. Opin. Biotechnol. 10, 458-64.
- 5 McBurney, MW, Mai, T, Yang, X y Jardine, K. (2002) "Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes" Exp. Cell. Res. 274, 1-8.
- Mizuguchi, H, Xu, Z, Ishii-Watabe, A, Uchida, E y Hayakawa, T. (2000) "IRES-dependent second gene expresión is significantly lower than cap-dependent first gene expresión in a bicistronic vector" Mol. Ther. 1, 376-82.
- 10 Rees, S, Coote, J, Stables, J, Goodson, S, Harris, S y 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, 106, 108-110.
- 15 Urlaub, G. y Chasin, L.A. "Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity". Proc. Natl. Acad. Sci. USA 77, 4216-20 (1980)
- Venkatesan, A, y Dasgupta, A. (2001) "Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements" Mol. Cell. Biol. 21, 2826-37.
- 20 Whitelaw, E, Sutherland, H, Kearns, M, Morgan, H, Weaving, L y Garrick, D. (2001) "Epigenetic effects on transgene expresión" Methods Mol. Biol. 158, 351-68.

REIVINDICACIONES

1. Una molécula de ADN que comprende: una unidad de transcripción multicistrónica que comprende al menos una secuencia de codificación que codifica tanto
- 5 i) un polipéptido de interés, como
- ii) un polipéptido marcador seleccionable funcional en una célula hospedadora eucariótica, en la que el polipéptido de interés tiene una secuencia de iniciación de la traducción separada de la del polipéptido marcador seleccionable,
- 10 en la que al menos una secuencia de codificación del polipéptido de interés está en dirección 5' de al menos una secuencia de codificación del polipéptido marcador seleccionable en dicha unidad de transcripción multicistrónica,
- 15 en la que está presente un sitio de entrada interna del ribosoma (IRES) en dirección 3' a al menos una secuencia de codificación del polipéptido de interés y en dirección 5' a al menos una secuencia de codificación del polipéptido marcador seleccionable, y
- 20 caracterizada por que la secuencia de codificación que codifica el polipéptido marcador seleccionable comprende una secuencia de inicio de la traducción seleccionada del grupo constituido por:
- a) un codón de inicio GTG;
- b) un codón de inicio TTG;
- c) un codón de inicio CTG;
- 25 d) un codón de inicio ATT y
- e) un codón de inicio ACG.
2. Una molécula de ADN según la reivindicación 1, en la que la secuencia de inicio de la traducción para el polipéptido marcador seleccionable comprende un codón de inicio GTG o un codón de inicio TTG.
- 30 3. Una molécula de ADN según una cualquiera de las reivindicaciones precedentes, en la que el polipéptido marcador seleccionable proporciona resistencia frente a los efectos letales o inhibidores del crecimiento de un agente de selección.
4. Una molécula de ADN según la reivindicación 3, en la que dicho agente de selección se selecciona del grupo constituido por zeocina, puromicina, blasticidina, higromicina, neomicina, metotrexato, metionina sulfoximina y kanamicina.
- 35 5. Una molécula de ADN según la reivindicación 3, en la que el agente de selección es zeocina.
6. Una molécula de ADN según la reivindicación 1 o 2, en la que el polipéptido marcador seleccionable es una enzima sintetizadora de 5,6,7,8-tetrahidrofolato (dhfrdhfr).
- 40 7. Una molécula de ADN según una cualquiera de las reivindicaciones precedentes, en la que la unidad de transcripción multicistrónica comprende adicionalmente una secuencia que codifica un segundo polipéptido marcador seleccionable funcional en la célula eucariótica, en la que dicha secuencia codifica un segundo polipéptido marcador seleccionable:
- 45 a) tiene una secuencia de iniciación de la traducción separada de la del polipéptido de interés,
- b) está dispuesta en dirección 5' de dicha secuencia que codifica un polipéptido de interés,
- 50 c) no tiene ninguna secuencia ATG en la hebra de codificación después del codón de inicio de dicho polipéptido marcador seleccionable hasta el codón de inicio del polipéptido de interés y
- d) tiene un codón de inicio GTG o un codón de inicio TTG.
8. Un módulo de expresión que comprende una molécula de ADN según una cualquiera de las reivindicaciones precedentes, comprendiendo dicho módulo de expresión un promotor en dirección 5' de dicha unidad de transcripción multicistrónica y una secuencia de terminación de la transcripción en dirección 3' de la unidad de transcripción multicistrónica, en el que dicho módulo de expresión es funcional en una célula hospedadora eucariótica para iniciar la transcripción de la unidad de transcripción multicistrónica.
- 55 9. Un módulo de expresión según la reivindicación 8, que comprende adicionalmente al menos un elemento de control de cromatina seleccionado del grupo constituido por una región de unión a matriz o armazón (MAR/SAR), una secuencia aislante, un elemento de apertura de cromatina universal (UCOE) y una secuencia antirrepresora (STAR).
- 60

10. Un módulo de expresión según la reivindicación 9, en el que dicho al menos un elemento de control de cromatina es una secuencia antirrepresora seleccionada del grupo constituido por:
- a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66;
 - b) fragmentos de una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en los que dichos fragmentos tienen actividad antirrepresora;
 - c) secuencias que son al menos un 70% idénticas en secuencia nucleotídica a a) o a b), en las que dichas secuencias tienen actividad antirrepresora; y
 - d) el complemento de uno cualquiera de a) a c).
11. Una célula hospedadora que comprende una molécula de ADN según una cualquiera de las reivindicaciones 1-7 o un módulo de expresión según una cualquiera de las reivindicaciones 8-10, siendo preferiblemente la célula hospedadora una célula de mamífero, preferiblemente una célula de ovario de hámster chino (CHO).
12. Un procedimiento de generación de una célula hospedadora capaz de expresar un polipéptido de interés, comprendiendo dicho procedimiento las etapas de:
- a) introducir en una pluralidad de células precursoras una molécula de ADN según una cualquiera de las reivindicaciones 1-7 o un módulo de expresión según una cualquiera de las reivindicaciones 8-10 y
 - b) cultivar la pluralidad de células precursoras en condiciones adecuadas para la expresión del polipéptido marcador seleccionable, y
 - c) seleccionar al menos una célula hospedadora que expresa el polipéptido de interés.
13. Un procedimiento de expresión de un polipéptido de interés, que comprende cultivar una célula hospedadora que comprende un módulo de expresión según una cualquiera de las reivindicaciones 8-10 y que expresa el polipéptido de interés a partir del módulo de expresión.
14. Un procedimiento según la reivindicación 13, que comprende adicionalmente recoger el polipéptido de interés.
15. Un procedimiento según la reivindicación 13 ó 14, en el que dichas células hospedadoras son células CHO que tienen fenotipo *dhfr* y en el que el módulo de expresión comprende una secuencia de codificación de un polipéptido marcador seleccionable que es una enzima sintetizadora de 5,6,7,8-tetrahidrofolato (*dhfrdhfr*), en el que dichas células se cultivan en un medio de cultivo que contiene folato y estando dicho medio de cultivo esencialmente desprovisto de hipoxantina y timidina.

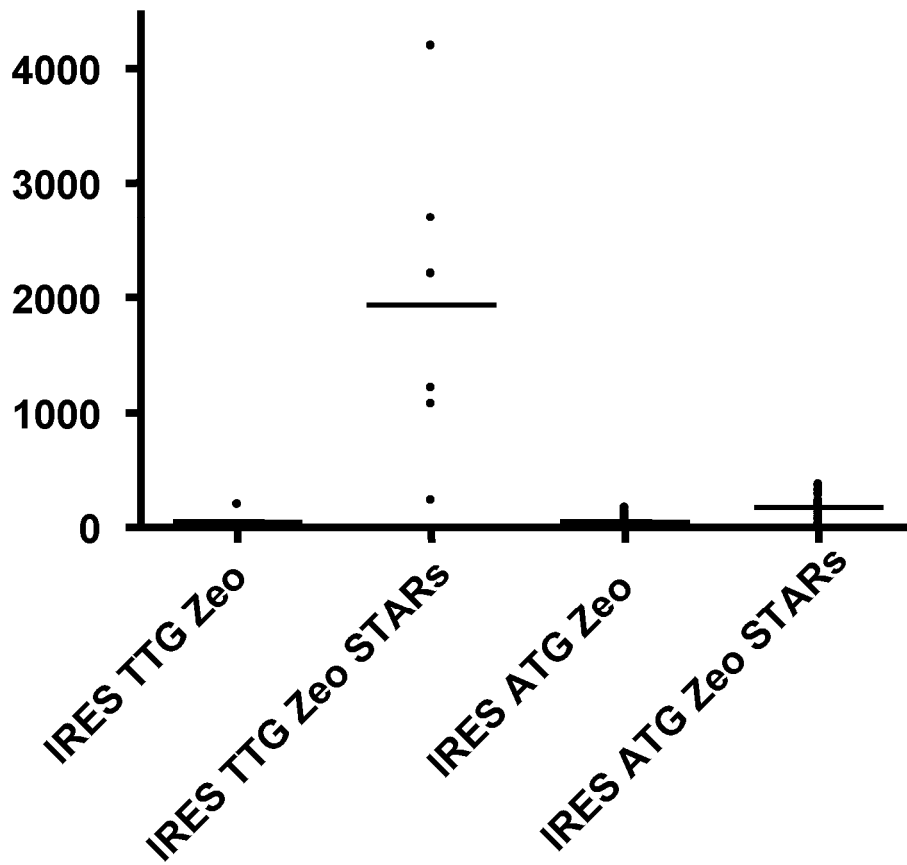
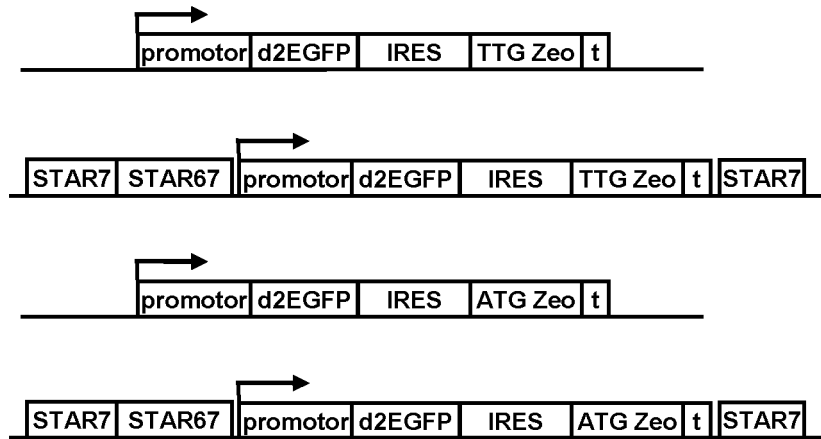


FIG. 1

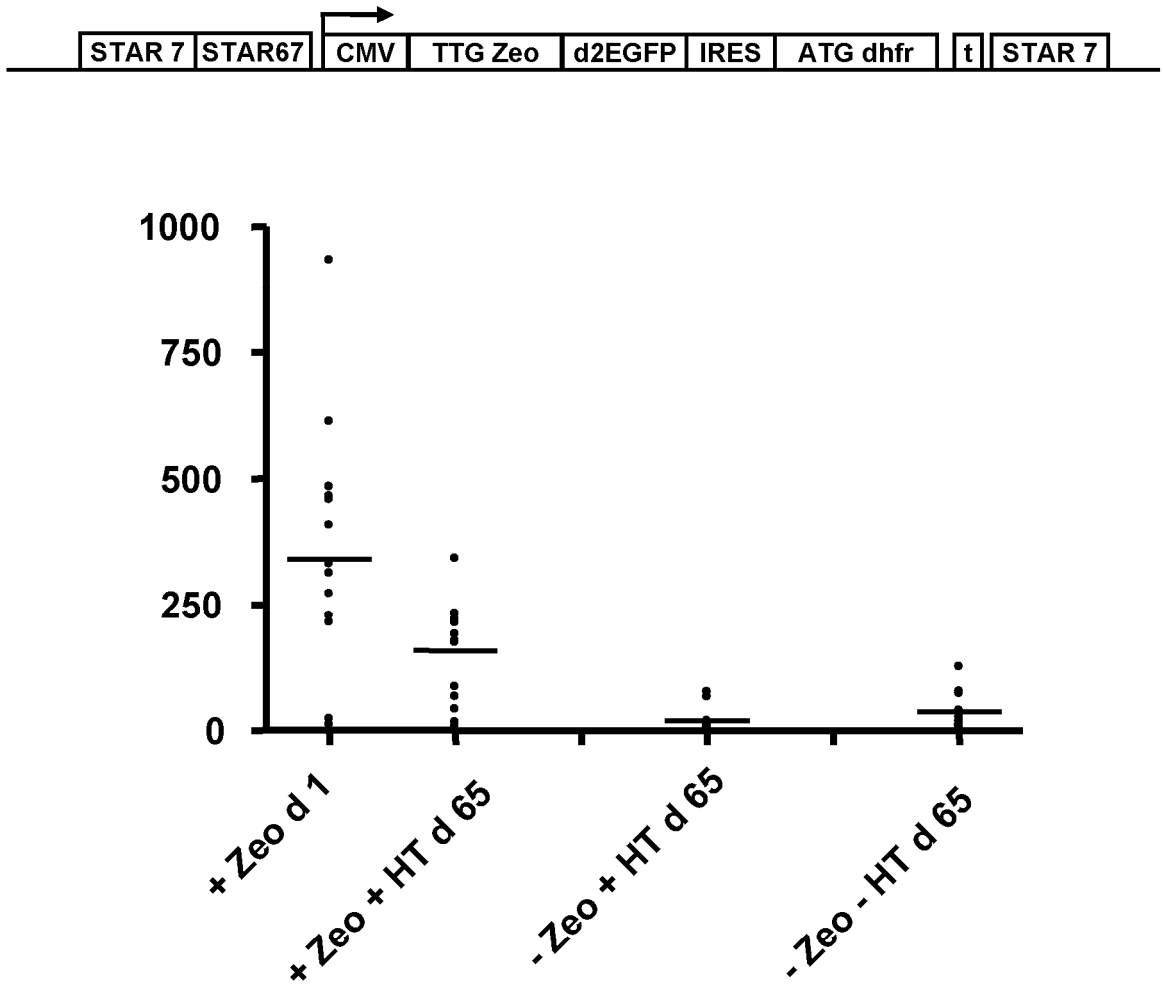


FIG. 2

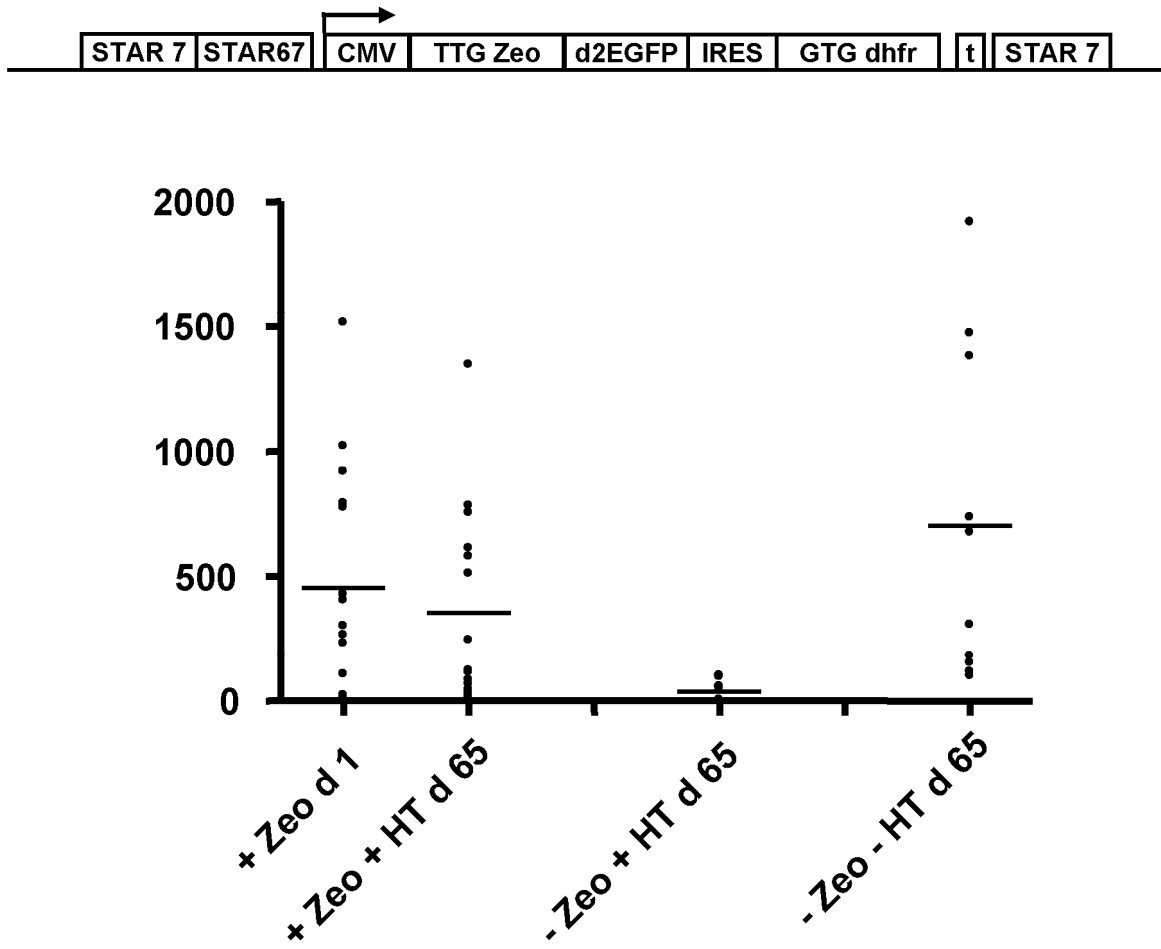


FIG. 3

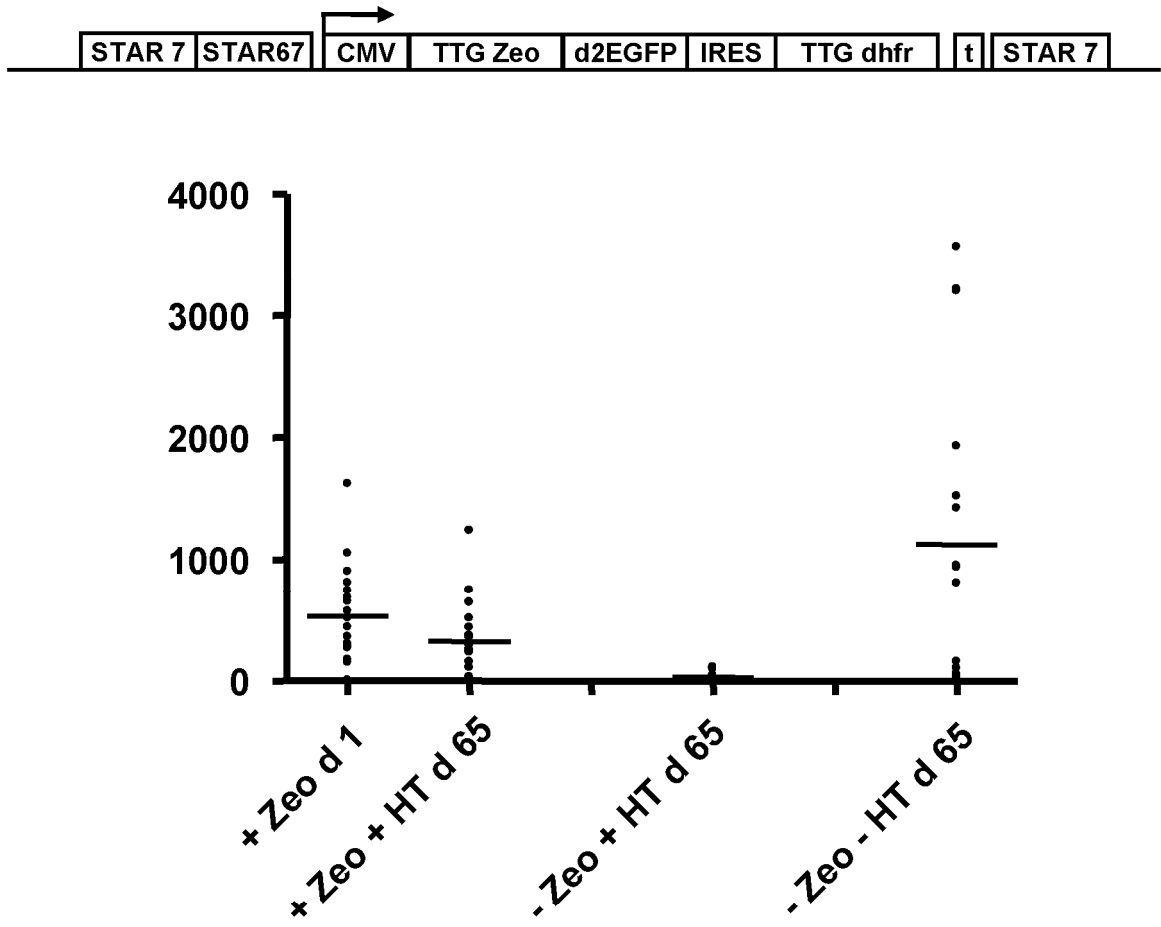


FIG. 4

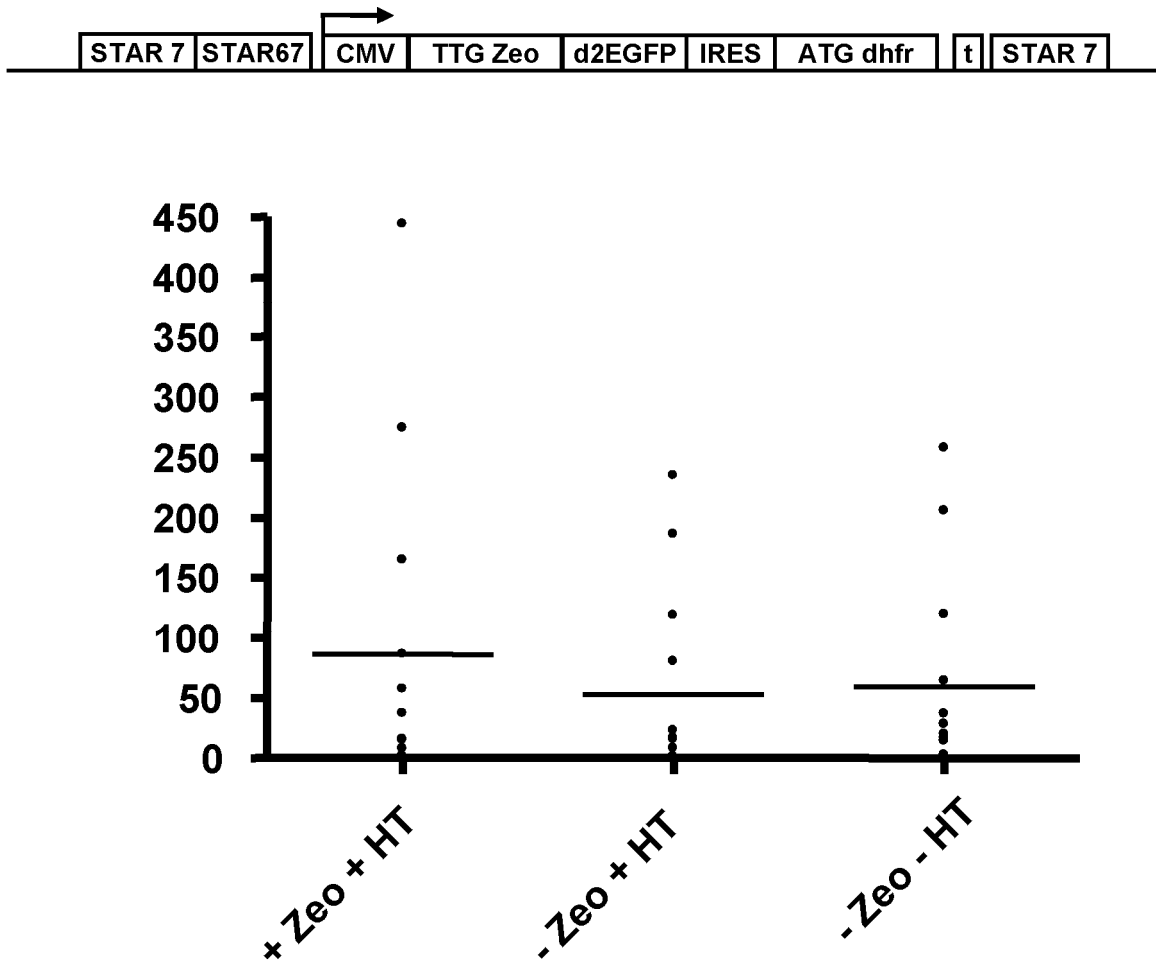


FIG. 5

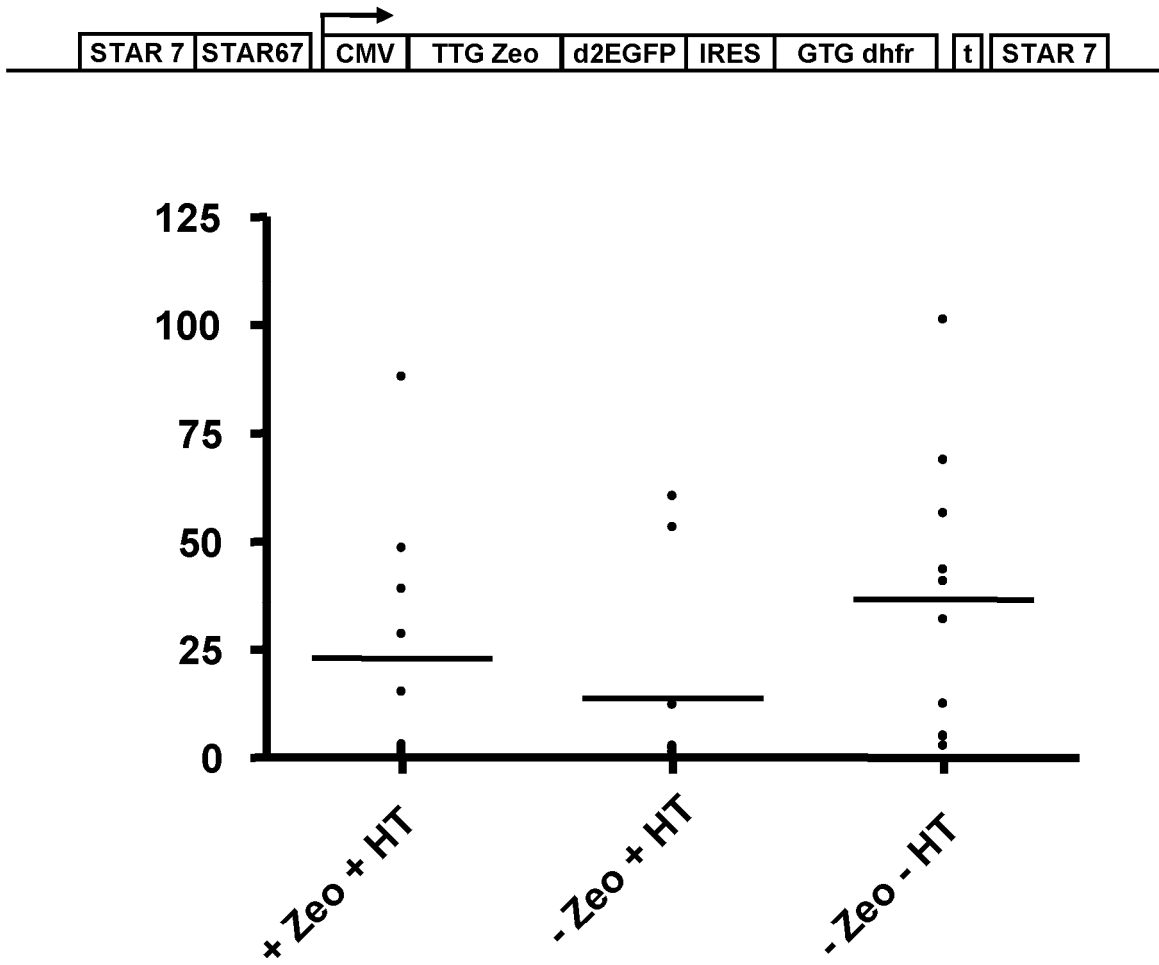


FIG. 6

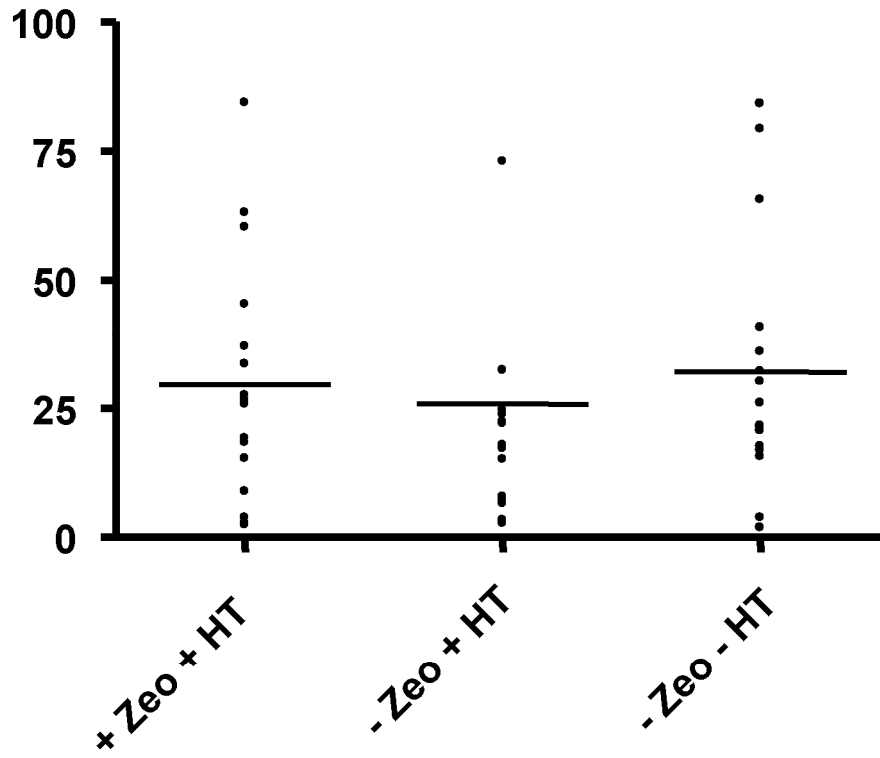
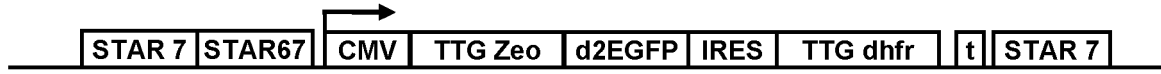


FIG. 7