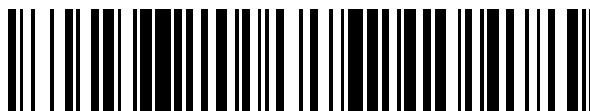


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 563 073**

51 Int. Cl.:

C12P 21/06 (2006.01)

C07H 21/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.08.2010 E 10807224 (0)**

97 Fecha y número de publicación de la concesión europea: **23.12.2015 EP 2462237**

54 Título: **Métodos para mejorar la expresión de proteínas recombinantes**

30 Prioridad:

06.08.2009 US 231906 P

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

10.03.2016

73 Titular/es:

**CMC ICOS BIOLOGICS, INC. (100.0%)
22021 20th Avenue SE
Bothell, WA 98021, US**

72 Inventor/es:

CLARKE, HOWARD ROBERT GRAHAME

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 563 073 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

Métodos para mejorar la expresión de proteínas recombinantes

Referencia cruzada a la solicitud relacionada

5 Esta solicitud reivindica el beneficio de prioridad de la solicitud provisional de EEUU n.º de serie 61/231.906, presentada el 6 de agosto, 2009, titulada "Métodos para mejorar la expresión de proteínas recombinantes", n.º de expediente del abogado 31351/44743.

Campo de la invención

10 Esta invención presenta una aplicación práctica en el campo de la expresión de proteínas recombinantes en células eucariotas por medio del aumento en la presión de selección sobre un vector, aumentando con ello la expresión de proteínas heterólogas asociadas con el vector.

Antecedentes

15 En el campo de la producción de proteínas recombinantes, el aumento en la expresión de un gen transfectado resulta una prioridad fundamental durante el desarrollo de una línea celular. El aumento en la transcripción, la traducción, el plegamiento de proteínas y la secreción son el objetivo de intensas investigaciones para aumentar las titulaciones de la proteína heteróloga.

Independientemente de los métodos utilizados en el pasado, en la técnica resulta necesario proporcionar mejores métodos para la producción de proteínas recombinantes que aumenten el rendimiento de la proteína deseada.

Sumario de la invención

20 En un aspecto, la invención proporciona un método para aumentar la expresión de proteínas heterólogas en una célula hospedante, que comprende las etapas de cultivar la célula hospedante que comprende una primera secuencia de polinucleótido heteróloga que codifica la proteína heteróloga bajo condiciones que permitan la expresión de proteínas, y el primer polinucleótido es codificado sobre un vector, y la célula hospedante comprende además una segunda secuencia de polinucleótido que presenta una secuencia codificadora de proteína para una proteína de un marcador seleccionable, y el segundo polinucleótido presenta una modificación en la secuencia, comparado con un polinucleótido de tipo salvaje, que codifica la proteína del marcador seleccionable, y la modificación de la secuencia reduce la eficacia de traducción de un ARNm codificado por el segundo polinucleótido, y el segundo polinucleótido presenta la modificación en la secuencia y el polinucleótido de tipo salvaje codifica secuencias de aminoácidos idénticas para la proteína del marcador seleccionable, y dicho primer polinucleótido y dicho segundo polinucleótido están codificados sobre un único vector de expresión del factor de alargamiento 1 del hámster chino (CHEF1), en el que el primer polinucleótido y el segundo polinucleótido están cada uno bajo el control transcripcional de promotores diferenciados.

35 En una realización del método, la modificación está en una región no traducida del segundo polinucleótido que codifica la proteína del marcador seleccionable y, en ciertos aspectos, la modificación está en una región no traducida 5' y/o la modificación está en una región no traducida 3'.

En otra realización del método, la modificación está en una región codificadora de proteína del gen que codifica la proteína del marcador seleccionable. En un aspecto, la modificación está a 25, 20, 15, 10 o 5 codones desde un codón de inicio de la región codificadora de proteína para la secuencia codificadora del gen del marcador seleccionable.

40 En otro aspecto del método, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del marcador seleccionable, y el codón modificado es un codón que no es un codón preferido para el aminoácido codificado para la célula hospedante. En un aspecto, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del marcador seleccionable, y el codón modificado es un codón que es un codón menos preferido para el aminoácido codificado para la célula hospedante.

45 En otro aspecto del método, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del marcador seleccionable, y la modificación introduce un cambio en la estructura secundaria del ARNm que reduce la eficacia de la traducción del ARNm. En una realización del método, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del marcador seleccionable, y la modificación aumenta el apareamiento de codones en el ARNm. En otra realización del método, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del

marcador seleccionable, y la modificación modifica el contenido en G+C del ARNm. En diversos aspectos, la modificación aumenta el contenido en G+C del ARNm, y en diversos aspectos, el contenido en G+C aumenta en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, o 100%. En otros aspectos, el contenido en G+C aumenta en más del 100%

En otro aspecto del método, la secuencia codificadora de proteína en la segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica la proteína del marcador seleccionable, y la modificación modifica el contenido en A+T del ARNm. En una realización, la modificación disminuye el contenido en A+T del ARNm en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, o 100%.

En otros aspectos del método, al menos 1%, al menos 2%, al menos 3%, al menos 4%, al menos 5%, al menos 6%, al menos 7%, al menos 8%, al menos 9%, al menos 10%, al menos 11%, al menos 12%, al menos 13%, al menos 14%, al menos 15%, al menos 16%, al menos 17%, al menos 18%, al menos 19%, al menos 20%, al menos 21%, al menos 22%, al menos 23%, al menos 24%, al menos 25%, al menos 26%, al menos 27%, al menos 28%, al menos 29%, al menos 30%, al menos 31%, al menos 32%, al menos 33%, al menos 34%, al menos 35%, al menos 36%, al menos 37%, al menos 38%, al menos 39%, al menos 40%, al menos 41%, al menos 42%, al menos 43%, al menos 44%, al menos 45%, al menos 46%, al menos 47%, al menos 48%, al menos 49%, al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% de los codones en la secuencia codificadora de proteína del segundo polinucleótido que codifica la proteína del marcador seleccionable son codones modificados.

En otros aspectos del método, la proteína del marcador seleccionable se selecciona del grupo que consiste en neomicina fosfotransferasa (npt II), higromicina fosfotransferasa (hpt), dihidrofolato reductasa (dhfr), zeocina, fleomicina, gen de resistencia a la bleomicina ble, gentamicina acetiltransferasa, estreptomycin fosfotransferasa, una forma mutante de la acetolactato sintasa (als), bromoxinil nitrilasa, fosfinotricina acetil transferasa (bar), enolpiruvilshikimato-3-fosfato (EPSP) sintasa (aro A), molécula del receptor de tirosina quinasa específica del músculo (MuSK-R), cobre-cinc superóxido dismutasa (sod 1), metalotioneínas (cup1, MT1), beta-lactamasa (BLA), puromicina N-acetil-transferasa (pac), blastidina acetil transferasa (bls), blastidina desaminasa (bsr), histidinol deshidrogenasa (HDH), N-succinil-5-aminoimidazol-4-carboxamida ribótido (SAICAR) sintetasa (ade1), argininosuccinato liasa (arg4), beta-isopropilmalato deshidrogenasa (leu2), invertasa (suc2), orotidina-5'-fosfato (OMP) descarboxilasa (ura3) y ortólogos de muchas de estas proteínas de marcadores.

En diversas realizaciones del método, la célula hospedante es una célula eucariota, la célula hospedante es una célula de mamífero, la célula hospedante es una célula humana, la célula hospedante es una célula de hámster chino, la célula hospedante es una célula de ovario de hámster chino, la célula hospedante es una célula de levadura, la célula hospedante es una célula de *Saccharomyces cerevisiae*, la célula hospedante es una célula de *Pichia pastoris*, la célula hospedante es una célula procariota, la célula hospedante es una célula de *Escherichia coli*, la célula hospedante es una célula de insecto, la célula hospedante es una célula de *Spodoptera frugiperda*, la célula hospedante es una célula vegetal, o la célula hospedante es una célula fúngica.

En el método, el vector de expresión es un vector de expresión del factor de alargamiento 1 del hámster chino (CHEF1). En otro aspecto, el método utiliza un segundo polinucleótido que comprende el polinucleótido indicado en la figura 2.

Descripción de los dibujos

La figura 1A es un polinucleótido que codifica DHFR, y la figura 1B es una secuencia del polipéptido de DHFR utilizada para la desoptimización de codones idéntica al ADNc BC005796m de *Mus musculus*.

La figura 2 muestra secuencias de ADN de las secuencias de DHFR con codones desoptimizados denominadas lisiada ("crippled", cr) y peor ("worst", wst).

La figura 3 muestra la DHFR desoptimizada (peor (wst) y lisiada (cr)) alineada con la secuencia de tipo salvaje (wt). Se indican los cambios en nucleótidos (*) que incluyen los codones menos preferidos de hámster (véase la tabla 4) y nuevas parejas de codones en tándem (en negrita; véase la tabla 5). Los símbolos degenerados son: B (C o G o T), D (A o G o T), H (A o C o T), V (A o C o G).

La figura 4 muestra el vector de expresión de CHEF1, pDEF38, con DHFR de tipo salvaje (WT). La DHFR con codones desoptimizados reemplaza a la DHFR WT para formar pDEF81 (DHFR lisiada) y pDEF82 (DHFR peor). El gen indicador de FIGI se clona en los sitios de clonación XhoI-XbaI para formar pDEF38:FIGI, pDEF81:FIGI y pDEF82:FIGI.

5 La figura 5 demuestra que la expresión de proteínas aumenta utilizando la DHFR con codones desoptimizados. Células CHO se transfectaron con DHFR de tipo salvaje (wt) y con codones desoptimizados (lisiada, pDEF81:FIGI, y peor, pDEF82:FIGI) que coexpresan una proteína de interés (FIGI). Los valores de las titulaciones determinadas mediante HPLC de proteína A e indicados en $\mu\text{g/ml}$ son la media de dos transfecciones independientes, cada una mediada por triplicado (seis ensayos de producción total). Los resultados indican una clara mejora en la titulación de la expresión para las agrupaciones de transfección seleccionadas para la DHFR con codones desoptimizados frente a las agrupaciones de DHFR de tipo salvaje.

10 La figura 6 demuestra que un modelo de producción de alimentación discontinua de agrupaciones de transfección proporciona una mayor productividad en líneas celulares con codones desoptimizados. Este experimento se realizó durante 12 días en tubos de centrifuga de 50 ml con los transfectantes agrupados: DHFR de tipo salvaje (pDEF38:FIGI, azul) y con codones desoptimizados (pDEF81:FIGI, morado, y pDEF82:FIGI, rosa) que coexpresan la proteína de interés FIGI. Dos agrupaciones de transfección (Ay B) se realizaron por duplicado. Las agrupaciones con codones desoptimizados muestran mayor productividad que las muestras de tipo salvaje.

15 La figura 7 demuestra que las células seleccionadas para DHFR con codones desoptimizados tienen una menor expresión de DHFR y una mayor expresión de la proteína de interés. Células CHO se transfectaron con DHFR de tipo salvaje (T462) y con codones desoptimizados (pDEF81:FIGI, T463 y pDEF82:FIGI, T464) que coexpresan la proteína de interés FIGI. Las agrupaciones de transfección se tiñeron con metotrexato fluorescente (F-MTX) para detectar DHFR, y un anticuerpo marcado fluorescente que reconoce a FIGI (RPE). Las células teñidas se analizaron mediante citometría de flujo en un FACSCalibur. La figura 7A muestra los perfiles FACS de tinción dual de 10.000 células individuales procedentes de cada transfección que representa la expresión combinada de DHFR (F-MTX) y FIGI (RPE:FIGI). La figura 7B muestra el promedio de intensidad de fluorescencia de F-MTX (DHFR) y RPE procedente de dos poblaciones de 10.000 células promediada para cada transfección. Estos resultados indican que ambas agrupaciones de DHFR con codones desoptimizados presentan una menor producción de DHFR y una mayor producción de FIGI cuando se comparan con las células de tipo salvaje.

20 La figura 8 demuestra que los clones de DHFR con codones desoptimizados tienen una menor expresión de DHFR y una mayor expresión de la proteína de interés. La agrupaciones de transfección de células CHO (tipo salvaje T462, lisiada T463 y peor T464) se clonaron mediante una dilución limitante, y 23 líneas celulares monoclonales confirmadas se expandieron de cada transfección. Las células clonales se tiñeron con metotrexato fluorescente (DHFR RFU) para detectar DHFR y con un anticuerpo fluorescente anti-FIGI marcado con RPE (FIGI RFU). Se analizó un total de 10.000 células teñidas de cada población clonal mediante citometría de flujo en un FACSCalibur. La figura 8A muestra el promedio de fluorescencia de células teñidas con F-MTX. Cada punto de datos es un clon individual. Los clones se clasifican desde un promedio bajo de fluorescencia a un promedio alto. La figura 8B muestra el promedio de fluorescencia de las células teñidas con RPE. Cada punto de datos es un clon individual. Los clones se clasifican desde un promedio bajo de fluorescencia a un promedio alto.

25 La figura 9 demuestra que los clones con codones desoptimizados tienen una mayor productividad, comparado con los clones de tipo salvaje. Las titulaciones de los clones se determinaron mediante HPLC de proteína A en los sobrenadantes recolectados el día 8 procedentes de modelos de producción de 6 pocillos. Los clones se clasificaron según la titulación de alta a baja. Los clones con codones desoptimizados, pDEF81:FIGI y pDEF82:FIGI, muestran una mayor productividad de FIGI que los clones de DHFR de tipo salvaje (pDEF38:FIGI).

Descripción detallada de la invención

30 La presente invención proporciona una nueva generación de vectores de expresión y sus usos, que mejoran los rendimientos de proteínas recombinantes. Los vectores de la invención permiten una mayor expresión de un gen de interés (GOI) en una célula hospedante y reducen la eficacia de traducción de un marcador seleccionable cotransformado, aumentando con ello la rigurosidad de la selección. Los marcados seleccionables se emplean en experimentos de transfección para complementar las deficiencias de proteínas de la célula hospedante o para conferir resistencia a un agente que sería tóxico en otro caso, y así se puede seleccionar la presencia (expresión) de genes de interés cotransformados. Los vectores que proporcionan la menor eficacia de traducción de la proteína del marcador seleccionable se diseñan de modo que el polinucleótido que codifica la proteína del marcador de selección está "desoptimizado" con respecto a uno o más parámetros. El uso de los vectores proporcionados es contrario a la lógica de los materiales y métodos utilizados para potenciar la expresión de proteínas recombinantes. En efecto, una mayor expresión de proteínas generalmente se lleva a cabo "optimizando" un polinucleótido que codifica una proteína de interés, aumentando con ello la eficacia de la traducción y la expresión de proteínas. Por extensión, se optimiza la región codificadora de proteína para el gen del marcador seleccionable de la misma manera. En la presente, sin embargo, se demuestra de modo inesperado que la modificación de un polinucleótido que codifica una secuencia de un gen de un marcador seleccionable para que sea menos que óptimo para la traducción, independientemente de realizar cambios similares en el gen de interés, permite el aislamiento de células

hospedantes transformadas o transfectadas con un polinucleótido que codifica un GOI y un polinucleótido que codifica un marcador seleccionable, en las que la proteína codificada por el GOI se expresa a unos niveles inesperadamente altos.

5 Por consiguiente, el término “desoptimizado”, tal como se emplea en la presente referido a un polinucleótido, significa que el polinucleótido se ha modificado de tal forma que la traducción de una proteína codificada por el polinucleótido es menos que óptima para la célula hospedante en la que se ha introducido el polinucleótido. Un polinucleótido se desoptimiza en una multitud de formas, y la presente invención no se limita a los métodos ejemplificados en la presente.

10 Los métodos para la optimización de codones han sido descritos por otros (Itakura 1987, Kotula 1991, Holler 1993, Seed 1998). Sin embargo, existen ejemplos limitados de utilidades para la desoptimización de codones. Uno de estos ejemplos es la desoptimización de genes víricos para reducir la aptitud replicativa mediante la incorporación de codones menos preferidos o pares de codones no aleatorizados (Burns 2006, Mueller 2006, Coleman 2008, Kew 2008). En la presente se describen las consideraciones metodológicas para reducir la eficacia transcripcional de un gen *dhfr* para su uso en células hospedantes mediante la incorporación de codones menos preferidos específicos de especie y parejas de codones en tándem. Los métodos presentados en general son aplicables para desoptimizar codones en un polinucleótido que codifica cualquier marcador seleccionable para su hospedante de una especie específica.

20 Sin pretender quedar limitado por ningún mecanismo de acción concreto, la menor traducción del marcador seleccionable puede conducir a un aumento compensatorio en la producción de la misma proteína a través de una vía alternativa distinta de la traducción, tal como, por ejemplo y sin limitación, una mayor transcripción o secreción, para permitir la supervivencia de las células que portan el gen ineficaz. Así, las células hospedantes que son capaces de superar el debilitamiento del gen del marcador y, por tanto, sobrevivir, también pueden expresar el GOI a mayor tasa. Independientemente del mecanismo exacto, en la presente se demuestra, de modo inesperado, que, en contra del conocimiento convencional, la modificación de la secuencia del polinucleótido del gen del marcador
25 seleccionable de una manera que se reduzca la eficacia traduccional de alguna manera aumenta la expresión del gen cotransformado que codifica el GOI.

Los vectores y los métodos de la invención pueden utilizarse con cualquier gen de un marcador seleccionable que proporcione una selección positiva. Los ejemplos de marcadores seleccionables incluyen, sin limitación, genes de resistencia a antibióticos que codifican la neomicina fosfotransferasa (*npt II*), higromicina fosfotransferasa (*hpt*), dihidrofolato reductasa (*dhfr*), zeocina, fleomicina, gen de resistencia a la bleomicina *ble* (la enzima no se conoce), gentamicina acetiltransferasa, estreptomina fosfotransferasa, una forma mutante de la acetolactato sintasa (*als*), bromoxinil nitrilasa, fosfinotricina acetil transferasa (*bar*), enolpiruvilshikimato-3-fosfato (EPSP) sintasa (*aro A*), molécula del receptor de tirosina quinasa específica del músculo (*MuSK-R*), cobre-cinc superóxido dismutasa (*sod 1*), metalotioneínas (*cup1*, *MT1*), beta-lactamasa (*BLA*), puromicina N-acetil-transferasa (*pac*), blastidina acetil transferasa (*bls*), blastidina desaminasa (*bsr*), histidinol deshidrogenasa (*HDH*), N-succinil-5-aminoimidazol-4-carboxamida ribótido (SAICAR) sintetasa (*ade1*), argininosuccinato liasa (*arg4*), beta-isopropilmalato deshidrogenasa (*leu2*), invertasa (*suc2*), y orotidina-5'-fosfato (OMP) descarboxilasa (*ura3*).

40 Tal como se entiende en la técnica, el código genético presenta codones que dirigen la adición de aminoácidos específicos en un polipéptido traducido. Como también se conoce en la técnica, los veinte aminoácidos naturales son codificados por un número diferente de codones, que varía de uno a seis codones diferentes para cada aminoácido. Tal como se emplea en la presente, los diferentes codones que codifican el mismo aminoácido se denominan “codones sinónimos”. Estos codones sinónimos se indican en la siguiente tabla 1.

Tabla 1 - El código genético

	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC Phe (F)	TCC Ser (S)	TAC Tyr (Y)	TGC Cys (C)
	TTA Leu (L)	TCA Ser (S)	TAA FIN	TGA FIN
	TTG Leu (L)	TCG Ser (S)	TAG FIN	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)
	CTA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)
	CTG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
	ATC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)
	ATA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)
	GTA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)
	GTG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)

Debido a que los codones sinónimos codifican el mismo aminoácido, la alteración de la secuencia codificadora de una proteína reemplazando un codón de tipo salvaje por un codón sinónimo no cambia la secuencia de aminoácidos de la secuencia polipeptídica codificada. Sin embargo, se altera la secuencia del ARNm subyacente que codifica la proteína, y el cambio en la secuencia de nucleótidos del ARNm puede alterar la expresión génica influyendo en la eficacia traduccional (Ikemura 1981a, Ikemura 1981b, Ikemura 1985).

Los factores específicos que gobiernan la eficacia de la traducción incluyen la incorporación de codones “preferidos”, codones en tándem o consecutivos (Rosenberg 1993), sesgos de parejas de codones (Gutman 1989, Boycheva 2003), estructura secundaria del ARN (Kozak 2005, Kudla 2009), contenido en GC y estructuras de repetición de nucleótidos (Hall 1982, Zhang 1991, Carlini 2003, Griswold 2003, Gustafsson 2004). Muchos de estos factores producen, por ejemplo y sin pretender quedar limitado por ningún mecanismo específico, sitios de pausa de la traducción que no solo detienen la traducción, sino que pueden afectar a la cinética de plegamiento de la proteína, y ambas situaciones, en último término, alteran la expresión de la proteína. Un ejemplo bien caracterizado de pausa en la traducción se produce durante la síntesis de genes biosintéticos de aminoácidos en bacterias, y se conoce ampliamente como atenuación (Watson 1988).

Preferencia de codones

En un aspecto, la invención proporciona vectores y métodos para aumentar la expresión de una proteína recombinante codificada por un GOI, utilizando un vector de expresión que comprende el GOI y que también codifica una proteína de un marcador seleccionable en un polinucleótido sintético diseñado con codones que no son preferidos en la célula hospedante. Se sabe en la técnica que, en diferentes especies, ciertos codones sinónimos se utilizan con más frecuencia que otros. Los codones que se utilizan con más frecuencia se denominan “codones preferidos” para esa especie. Otros han propuesto que la preferencia por ciertos codones es una función del número relativo de ARN de transferencia (ARNt) específicos codificados en el genoma de una especie, y se han desarrollado programas para determinar el número preciso de cada ARNt codificado en un genoma específico (Lowe y Eddy, 1997). Así, en un aspecto, la selección de los codones menos preferidos se basa en la frecuencia de utilización, previamente determinada, de codones sinónimos en una especie de célula hospedante concreta de origen.

En un aspecto, la invención proporciona un polinucleótido que codifica un marcador seleccionable, en el que la región codificadora de proteína del polinucleótido incluye al menos una modificación de un codón, y dicha modificación es la sustitución de un codón de tipo salvaje por un codón que es un codón no preferido para la célula hospedante. En otro aspecto, la modificación es la sustitución de un codón de tipo salvaje por un codón que es un codón menos preferido para la célula hospedante. Se contempla cualquier número de dichas sustituciones de codones, con la condición de que al menos una de dichas modificaciones se incorpore en la región codificadora de proteína. Por consiguiente, la invención contempla cualquier número de modificaciones desde una modificación de un codón de este tipo a la modificación de todos los codones en la región codificadora de proteína del gen del marcador seleccionable.

De forma más específica, en diversos aspectos, al menos 1%, al menos 2%, al menos 3%, al menos 4%, al menos 5%, al menos 6%, al menos 7%, al menos 8%, al menos 9%, al menos 10%, al menos 11%, al menos 12%, al menos 13%, al menos 14%, al menos 15%, al menos 16%, al menos 17%, al menos 18%, al menos 19%, al menos 20%, al menos 21%, al menos 22%, al menos 23%, al menos 24%, al menos 25%, al menos 26%, al menos 27%, al menos 28%, al menos 29%, al menos 30%, al menos 31%, al menos 32%, al menos 33%, al menos 34%, al menos 35%, al menos 36%, al menos 37%, al menos 38%, al menos 39%, al menos 40%, al menos 41%, al menos 42%, al menos 43%, al menos 44%, al menos 45%, al menos 46%, al menos 47%, al menos 48%, al menos 49%, al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% de los codones en la secuencia codificadora de proteína del polinucleótido que codifica el gen del marcador seleccionable son codones modificados.

Utilizando secuencias de nucleótidos de acceso público y tablas de utilización de codones conocidas en la técnica y ejemplificadas en las tablas 2 y 3 (Nakamura *et al.*, 2000) se puede crear una versión con codones desoptimizados de cualquier marcador seleccionable incorporando una selección aleatoria de los codones menos preferidos para la especie de origen de la célula hospedante seleccionada para la expresión de una proteína recombinante codificada por el GOI. Un ejemplo de los codones menos preferidos del hámster (*Cricetulus griseus*) se muestra en la tabla 4. Estos codones se emplean para reemplazar preferentemente codones sinónimos en la secuencia de un gen nativo que codifica un gen de un marcador, de modo que al menos uno a todos los codones sinónimos son reemplazados por cualquier codón que no sea el codón preferido para un resto aminoácido concreto.

Tabla 2 - Tabla de utilización de codones del hámster

Un ejemplo de tabla de utilización de codones del hámster (*Cricetulus griseus*) a partir de 331 regiones codificadoras de proteínas y 153527 codones. Para cada codón, el primer número es la frecuencia por mil y el segundo número es el número real de veces que se observó ese codón.

UUU 19,6 (3005)	UCU 16,0 (2450)	UAU 13,1 (2017)	UGU 9,1 (1397)
UUC 22,0 (3381)	UCC 16,5 (2529)	UAC 16,4 (2519)	UGC 10,3 (1589)
UUA 6,4 (978)	UCA 10,3 (1577)	UAA 0,6 (93)	UGA 1,2 (177)
UUG 14,1 (2169)	UCG 3,4 (529)	UAG 0,5 (84)	UGG 13,1 (2012)
CUU 13,2 (2023)	CCU 16,7 (2563)	CAU 10,2 (1563)	CGU 5,6 (863)
CUC 18,4 (2818)	CCC 17,0 (2608)	CAC 12,9 (1980)	CGC 9,3 (1429)
CUA 7,6 (1174)	CCA 15,6 (2388)	CAA 10,3 (1587)	CGA 7,2 (1102)
CUG 38,8 (5955)	CCG 4,3 (657)	CAG 33,4 (5122)	CGG 10,1 (1558)
AUU 17,4 (2673)	ACU 14,1 (2172)	AUU 17,4 (2671)	AGU 11,4 (1756)
AUC 24,8 (3808)	ACC 20,3 (3118)	AAC 21,2 (3248)	AGC 16,4 (2521)
AUA 6,9 (1053)	ACA 15,7 (2418)	AAA 24,6 (3782)	AGA 10,1 (1557)
AUG 23,0 (3538)	ACG 4,5 (685)	AAG 38,4 (5895)	AGG 10,2 (1570)
GUU 11,6 (1780)	GCU 22,4 (3434)	GAU 24,6 (3781)	GGU 12,8 (1968)
GUC 15,7 (2408)	GCC 25,9 (3973)	GAC 28,1 (4310)	GGC 21,3 (3568)
GUA 7,8 (1202)	GCA 16,3 (2497)	GAA 28,4 (4355)	GGA 15,8 (2425)
GUG 30,1 (4628)	GCG 5,0 (765)	GAG 41,1 (6311)	GGG 13,4 (2063)

Tabla 3 - Tabla de utilización de codones del ser humano

Se proporciona un ejemplo de una tabla de utilización de codones del ser humano (*Homo sapiens*) determinada a partir de 93487 regiones codificadoras de proteínas y el análisis de 40662582 codones. Para cada codón, el primer número es la frecuencia por mil y el segundo número es el número real de veces que se observó ese codón.

5

UUU 17,6 (714298)	UCU 15,2 (618711)	UAU 12,2 (495699)	UGU 10,6 (430311)
UUC 20,3 (824692)	UCC 17,7 (718892)	UAC 15,3 (622407)	UGC 12,6 (513028)
UUA 7,7 (311881)	UCA 12,2 (496448)	UAA 1,0 (40285)	UGA 1,6 (63237)
UUG 12,9 (525688)	UCG 4,4 (179419)	UAG 0,8 (32109)	UGG 13,2 (535595)
CUU 13,2 (536515)	CCU 17,5 (713233)	CAU 10,9 (441711)	CGU 4,5 (184609)
CUC 19,6 (796638)	CCC 19,8 (804620)	CAC 15,1 (613713)	CGC 10,4 (423516)
CUA 7,2 (290751)	CCA 16,9 (688038)	CAA 12,3 (501911)	CGA 6,2 (250760)
CUG 39,6 (1611801)	CCG 6,9 (281570)	CAG 34,2 (1391973)	CGG 11,4 (464485)
AUU 16,0 (650473)	ACU 13,1 (533609)	AUU 17,0 (689701)	AGU 12,1 (493429)
AUC 20,8 (846466)	ACC 19,9 (768147)	AAC 19,1 (776603)	AGC 19,5 (791383)
AUA 7,5 (304565)	ACA 15,1 (614523)	AAA 24,4 (993621)	AGA 12,2 (494682)
AUG 22,0 (896005)	ACG 6,1 (246105)	AAG 31,9 (1295568)	AGG 12,0 (486463)
GUU 11,0 (448607)	GCU 18,4 (750096)	GAU 21,8 (885429)	GGU 10,8 (437126)
GUC 14,5 (588138)	GCC 27,7 (1127679)	GAC 25,1 (1020595)	GGC 22,2 (903565)
GUA 7,1 (287712)	GCA 15,8 (643471)	GAA 29,0 (1177632)	GGA 16,5 (669873)
GUG 28,1 (1143534)	GCG 7,4 (299495)	GAG 39,6 (1609975)	GGG 16,5 (669768)

Tabla 4 - Codones menos preferidos del hámster

Un ejemplo de los codones menos preferidos del hámster (*Cricetulus griseus*)

Aminoácido	Codón menos preferido
Alanina	GCG, GCA
Arginina	CGT, CGA, CGC
Ácido aspártico	GAT
Asparagina	AAT
Cisteína	TGT
Ácido glutámico	GAA
Glutamina	CAA
Glicina	GGT, GGG

Aminoácido	Codón menos preferido
Isoleucina	ATA, ATT
Histidina	CAT
Leucina	TTA, CTA, CTT
Lisina	AAA
Fenilalanina	TTT
Prolina	CCG, CCA
Serina	AGT, TCG, TCA
Treonina	ACG, ACT
Tirosina	TAT
Valina	GTA, GTT

La desoptimización de codones puede realizarse mediante una diversidad de métodos, por ejemplo, seleccionando codones que son menos preferidos para su uso en genes altamente expresados en una célula hospedante concreta. Pueden emplearse algoritmos informáticos que incorporen tablas de frecuencia de codones, tales como “Ecohigh.cod” para la preferencia de codones de genes bacterianos altamente expresados, y que pueden obtenerse en the University of Wisconsin Package Version 9.0, Genetics Computer Group, Madison, WI. Otras tablas de frecuencia de codones útiles incluyen “Celegans_high.cod”, “Celegans_low.cod”, “Drosophila_high.cod”, “Human_high.cod”, “Maize_high.cod”, y “Yeast_high.cod”.

Sesgo de parejas de codones

En otro aspecto, la invención proporciona vectores y métodos para aumentar la expresión de una proteína recombinante codificada por un GOI transfectado, utilizando un vector de expresión que codifica una proteína de un marcador seleccionable en un polinucleótido sintético diseñado con parejas de codones que son menos preferidas en la especie de la célula hospedante de origen. Recientes resultados experimentales apoyan la idea de que las tasas de traducción están influidas por las compatibilidades de ARNt adyacentes en los sitios A y P sobre la superficie de los ribosomas de traducción (Smith y Yarus, 1989; Yarus y Curran, 1992). En la actualidad se entiende que algunas parejas de codones se emplean en secuencias codificadores de proteínas con mucha más frecuencia que la esperada partiendo de la utilización de los codones individuales de estas parejas (parejas de codones sobrerrepresentadas), y que algunas parejas de codones se observan con mucha menos frecuencia que la esperada (parejas de codones infrarrepresentadas). Coleman *et al.* (2008) han demostrado que una pareja de codones infrarrepresentada se traduce con más lentitud que una pareja de codones sobrerrepresentada, y que cuanto más infrarrepresentada esté una pareja de codones, con más lentitud se traduce.

Como ejemplo, en los seres humanos, los estudios han descubierto que el codón GCC de Ala se emplea cuatro veces más frecuentemente que el codón sinónimo GCG y que otras parejas de codones sinónimos se emplean con más o menos frecuencia que la esperada (Coleman *et al.*, 2008). Esta frecuencia de parejas de codones específicos se denomina “sesgo de parejas de codones”. Por ejemplo y de nuevo en seres humanos, basándose en la utilización de codones preferidos, se espera que la pareja de aminoácidos Ala-Glu sea codificada por GCCGAA y GCAGAG con una frecuencia aproximadamente igual. De hecho, la pareja de codones GCCGAA está muy infrarrepresentada, aunque contiene el codón Ala más frecuente, de modo que solo se utiliza una séptima parte de las veces que se utiliza GCAGAG.

Apareamiento de codones en tándem

En otro aspecto, la invención proporciona vectores y métodos para aumentar la expresión de una proteína recombinante codificada por un GOI, utilizando un vector de expresión que comprende el GOI y que también codifica una proteína de un marcador seleccionable en un polinucleótido sintético diseñado con parejas de codones en tándem. La frecuencia y la composición de las parejas de codones en una secuencia génica puede influir en la tasa de traducción, según se evidencia por la atenuación (Watson, 1988) y el desplazamiento de marco traduccional (Gurvich *et al.*, 2005). El mecanismo de atenuación implica la detención de los ribosomas en las parejas en tándem o repeticiones multiméricas del mismo codón, y se ve influido por la concentración de ARNt activados específicos de codones. Cuando se aparean codones raros, la detención de las moléculas de ARNt cognadas puede conducir no solo a la detención, sino al desplazamiento de marco, produciendo una reducción de la proteína traducida con precisión. Ambos mecanismos de acción de apareamiento de codones en tándem pueden utilizarse para

desoptimizar la expresión de un gen de un marcador seleccionable.

En la tabla 5 se muestran ejemplos de parejas de codones en tándem menos preferidos del hámster incorporados en los genes dhfr desoptimizados.

Tabla 5 - Parejas de codones en tándem

5 Los codones son los menos preferidos, excepto los que aparecen en negrita

Aminoácido	Parejas de codones en tándem
Ácido aspártico	GAC GAC
Ácido glutámico	GAA GAA
Glicina	GGG GGG
Leucina	CTA CTA
Lisina	AAA AAA
Fenilalanina	TTC TTC
Prolina	CCG CCG
Serina	TCG TCG
Serina	TCA TCA
Treonina	ACG ACG
Treonina	ACT ACT

Así, en una realización del método, los restos aminoácidos repetidos en tándem en la proteína del marcador seleccionable, en la que el mismo aminoácido está presente en más de una copia en la estructura primaria en tándem, son codificados por codones que no son un codón preferido para ese aminoácido. En otra realización, los restos aminoácidos repetidos en tándem en la proteína del marcador seleccionable, en la que el mismo aminoácido está presente en más de una copia en la estructura primaria en tándem, son codificados por codones que son los codones menos preferidos para ese aminoácido. En otra realización, los mismos aminoácidos presentes en más de una copia y en tándem en la estructura primaria son codificados por el mismo codón.

Estructura secundaria

15 En otro aspecto, la invención proporciona métodos para aumentar la expresión de una proteína recombinante codificada por un GOI, utilizando un vector de expresión que codifica una proteína de un marcador seleccionable en un polinucleótido diseñado con modificaciones en la secuencia que alteran la estructura secundaria del ARN.

En esta realización, se toma en cuenta la estructura del ARNm cuando se diseña un gen para la desoptimización de codones. El contexto de la secuencia, por ejemplo, de los codones rediseñados puede modular la estructura secundaria del ARN, que ha demostrado regular la estabilidad y la traducibilidad del mensaje del ARNm (Griswold 2003, Kozak 2005, Kudla 2009). Los factores a considerar cuando se diseña un marcador seleccionable con codones desoptimizados incluyen, pero no se limitan a la estabilidad de la estructura secundaria y la energía libre mínima (MFE) del ARN completo o del extremo 5' del ARN, según se puede determinar mediante un software de predicción de la estructura del ARN de libre acceso, tal como RNAfold (Gruber *et al.*, 2008). El contexto de la secuencia del gen desoptimizado en las regiones que lo rodean o en parte de un codón menos preferido también puede ser importante. Los factores que pueden reducir la eficacia traduccional incluyen el contenido en GC, G+C en la tercera posición del codón (Sueoka y Kawanishi, 2000), y las puntuaciones del índice de adaptación del codón (Sharp y Li, 1987). En efecto, las pruebas han demostrado que un contenido mayor en GC en el ARNm aumenta la probabilidad de formación de una estructura secundaria que obstaculice la eficacia de la traducción, y la reducción en el contenido de GC desestabiliza estas estructuras secundarias (Bulmer, 1989). Y a la inversa, para reducir la eficacia de traducción según se propone en los presentes métodos, el aumento del contenido en GC, mediante la sustitución de codones de tipo salvaje en la región codificadora de proteína por codones sinónimos con mayor contenido en GC o simplemente modificando las regiones no traducidas para incluir un mayor contenido en GC, proporciona un aumento en la estructura secundaria, reduciendo con ello la eficacia de la traducción.

35 En la técnica se sabe que la estructura primaria y secundaria de la región no codificadora 5' del ARNm modula la eficacia traduccional; se ha demostrado que la eficacia traduccional es inversamente proporcional al grado de

estructura secundaria en la región no codificadora 5' del ARNm (Pelletier y Sonenberg, 1987). En otro aspecto, se proporciona un método en el que el polinucleótido que codifica la proteína del marcador seleccionable se modifica fuera del contexto de la región codificadora de proteína, y las modificaciones en el gen se realizan de modo que las regiones no traducidas del ARNm codificado tienen mayor estructura secundaria, comparado con el ARNm de tipo salvaje. En un aspecto, se introducen una o más modificaciones en una región no traducida 5' y/o 3' que no es necesaria para la traducción. En otro aspecto, la modificación o las modificaciones se introducen en una región 5' y/o 3' que necesaria para la traducción.

Vectores y células hospedantes

El término "vector" se emplea según se reconoce en la técnica para indicar cualquier molécula (por ejemplo, ácido nucleico, plásmido o virus) empleada para transferir información codificadora a una célula hospedante. La expresión "célula hospedante" se emplea para indicar una célula que ha sido transformada o que es capaz de ser transformada por un vector que porta un gen de interés seleccionado, que después es expresado por la célula. La expresión incluye células de mamífero, levadura, fúngicas, de insecto, vegetales y protozoarias, y la progenie de la célula parental, independientemente de que la progenie sea idéntica en morfología o en constitución genética al progenitor original, con la condición de que el gen seleccionado esté presente. En general, la selección de un vector apropiado, en un aspecto, se basa en la célula hospedante seleccionada para la expresión del GOI.

Los ejemplos incluyen, pero no se limitan a células de mamífero, tales como células de ovario de hámster chino (CHO) (ATCC n.º CCL61); células CHO DHFR, células 293 o 293T de riñón embrionario humano (HEK) (ATCC n.º CRL1573); o células 3T3 (ATCC n.º CCL92). Otras líneas celulares de mamífero adecuadas son las líneas celulares COS-1 (ATCC n.º CRL1650) y COS-7 (ATCC n.º CRL1651) de mono, y la línea celular CV-1 (ATCC n.º CCL70). Otras líneas celulares de mamífero adecuadas incluyen, pero no se limitan a Sp2/0, células de hibridoma de ratón NS1 y NS0, células N2A de neuroblastoma de ratón, HeLa, células L-929 de ratón, líneas 3T3 derivadas de ratones Swiss, Balb-c o NIH, líneas de células de hámster BHK o HaK, que también están disponibles en el ATCC.

Otros ejemplos de células hospedantes de mamífero incluyen líneas celulares de primates y líneas celulares de roedores, que incluyen las líneas celulares transformadas. También son adecuadas las células diploides normales, cepas de células derivadas de un cultivo *in vitro* de tejido primario, así como explantes primarios.

Otras líneas celulares que también son útiles incluyen, por ejemplo, diversas cepas de *E. coli* (por ejemplo, HB101, (ATCC n.º 33694) DH5 α , DH10, y MC1061 (ATCC n.º 53338)), diversas cepas de *B. subtilis*, *Pseudomonas* spp., *Streptomyces* spp., *Salmonella typhimurium* y similares.

Muchas cepas de células de levadura conocidas por los expertos en la técnica también están disponibles como células hospedantes para la expresión de un GOI e incluyen, por ejemplo, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, cepas de *Kluyveromyces*, *Candida*, *Pichia ciferrii* y *Pichia pastoris*.

Además, cuando se desee, pueden utilizarse sistemas de células de insecto en los métodos de la presente invención. Estos sistemas incluyen, por ejemplo y sin limitación, Sf-9 y Hi5 (Invitrogen, Carlsbad, CA).

Los ejemplos de células fúngicas incluyen, sin limitación, *Thermoascus aurantiacus*, *Aspergillus* (hongo filamentoso), que incluye, sin limitación, *Aspergillus oryzae*, *Aspergillus nidulans*, *Aspergillus terreus*, y *Aspergillus niger*, *Fusarium* (hongo filamentoso), que incluye, sin limitación, *Fusarium venenatum*, *Penicillium chrysogenum*, *Penicillium citrinum*, *Acremonium chrysogenum*, *Trichoderma reesei*, *Mortierella alpina*, y *Chrysosporium lucknowense*.

Los ejemplos de células protozoarias incluyen, sin limitación, cepas de *Tetrahymena* y cepas de *Trypanosoma*.

Ejemplos

En una realización, la presente invención se ejemplifica empleando codones de hámster menos preferidos para generar un gen que codifica DHFR con codones desoptimizados (CDD) adecuado para la selección en células de ovario de hámster chino (CHO).

El gen de inicio es idéntico a un ADNc que codifica DHFR de *Mus musculus*, n.º de registro BC005796, y codifica el polipéptido de DHFR de tipo salvaje (véase la figura 1). Se sintetizaron dos versiones de un polinucleótido que codifica DHFR con codones desoptimizados, denominados en la presente lisiado y peor, que representan las secuencias codificadoras desoptimizadas en un grado intermedio y máximo, respectivamente. Estos polinucleótidos se diseñaron utilizando un algoritmo de utilización de codones GENEART AG CHO. Las secuencias de polinucleótidos que codifican DHFR con codones desoptimizados se muestran en la figura 2. Los genes de DHFR con codones desoptimizados se alinean con la secuencia del gen de DHFR de tipo salvaje en la figura 3, y se resaltan las diferencias en los nucleótidos que resultan de la introducción de los codones menos preferidos de hámster y parejas de codones en tándem. Los productos de la traducción de los tres genes, el tipo salvaje, el lisiado y el peor, son idénticos.

Las secuencias de polinucleótidos que codifican DHFR con codones desoptimizados se introdujeron en el vector de

expresión pDEF38, un vector de expresión de CHEF1 (patente de EEUU n.º 5.888.809), para sustituir la secuencia codificadora de DHFR de tipo salvaje (figura 4). Los plásmidos resultantes se denominaron (DHFR lisiada) y pDEF82 (DHFR peor). El gen indicador de interés, FIGI, que codifica una proteína de fusión IgG1 Fc, se clonó en el sitio de clonación múltiple (XhoI a XbaI) de pDEF38, pDEF81 y pDEF82 para crear los vectores de expresión pDEF38:FIGI, pDEF81:FIGI y pDEF82:FIGI, respectivamente.

Estos vectores de expresión de FIGI se transfectaron en células CHO DG44, se cultivaron durante dos días en medio sin selección que contenía hipoxantina y timidina (HT), y después se seleccionaron en medio que carecía de HT (-HT). Las poblaciones celulares seleccionadas, o agrupaciones, se expandieron y se dividieron en cultivos de modelo de producción para evaluar la productividad.

Las agrupaciones de transfección se diluyeron hasta células individuales de siembra en pocillos individuales de placas de 96 pocillos. Se formaron imágenes de las placas con el Clone Select Imager (Genetix), y los pocillos que contenían células que expresan FIGI derivadas de una única célula se expandieron. Se seleccionaron aleatoriamente 23 clones a partir de placas de dilución limitante para cada transfección (DHFR de tipo salvaje, lisiada o peor) a partir de los conjuntos monoclonales confirmados.

Los modelos de producción de 6 pocillos se inocularon con un total de un millón de células en 3 ml de medio de cultivo celular con FBS al 10% y se cultivaron durante 4 días a 37 °C, después durante 4 días a 34 °C. Los sobrenadantes de la recolección se filtraron a través de filtros de 0,2 micrómetros y se ensayaron para la producción de FIGI con una HPLC de proteína A. Los modelos de producción de alimentación discontinua se sembraron a 0,5 millones de células/ml en medio de cultivo suplementado con FBS al 10% en tubos de centrifugación. Los tubos de centrifugación de 50 ml se ensayaron con un volumen de trabajo de 15 ml. Después de sembrar, las muestras se cultivaron a 37 °C y con CO₂ al 6% durante 3 días, con alimentación y desplazamiento de la temperatura hasta 34 °C que comienza en el día 4. Las muestras para la titulación y la densidad celular se recogieron en los días 3, 5, 7, 10 y 12. El estudio concluyó en el día 12.

Se realizaron análisis FACS con células del cultivo normales del día 2 que fueron recolectadas y teñidas con metotrexato marcado con isotiocianato-fluoresceína (F-MTX) para detectar la proteína de DHFR y un Fc anti-IgG1 marcado con R-ficoeritrina (RPE) para detectar el FIGI.

Se prepararon líneas celulares estables que expresan la proteína indicadora FIGI utilizando genes de tipo salvaje y con codones desoptimizados que codifican el marcador seleccionable DHFR. Se realizaron transfecciones por duplicado (T462-T464, A y B) con los plásmidos de DHFR de tipo salvaje, lisiada y peor que expresan la proteína indicadora FIGI. Las colonias individuales contadas para cada transfección se indican como "n.º de transfectantes". Tal como puede observarse en la tabla 6, los resultados de la transfección indican que la presión de selección aumenta cuando se emplea DHFR con codones desoptimizados (CDD), comparado con la DHFR de tipo salvaje. Este resultado se considera como una reducción en el número de transfectantes CDD seleccionados en un medio que carece de HT.

Tabla 6 - Número de transfectantes por transfección

Transfección	Plásmido	Marcador de DHFR	n.º de transfectantes
T462A	pDEF38:FIGI	tipo salvaje	33834
T462B	pDEF38:FIGI	tipo salvaje	22663
T463A	pDEF81:FIGI	lisiado	1915
T463B	pDEF81:FIGI	lisiado	4309
T464A	pDEF82:FIGI	peor	7342
T464B	pDEF82:FIGI	peor	6863

La cantidad de proteína de FIGI producida a partir de los transfectantes agrupados en los modelos de producción de la placa de 6 pocillos, día 8 (figura 5) y del tubo de centrifugación, alimentación discontinua durante 12 días (figura 6) mostraron un aumento inesperado en la productividad del GOI con el gen del marcador seleccionable de DHFR con codones desoptimizados, frente al gen de DHFR de tipo salvaje. El gen de DHFR lisiado produjo la titulación mayor. Este resultado resulta coherente con la observación de que la selección de DHFR lisiada fue la más rigurosa (véase la tabla 6) y sugiere que la diversidad en la población puede reducirse, pero la célula promedio expresa más POI. Esta conclusión es evidente en la distribución FACS de la DHFR lisiada (T463) en la figura 7A, que muestra un apretado grupo de células que se tiñen muy intensamente con RPE:FIGI, con una tinción reducida con F-MTX concomitante. Las células de DHFR peor muestran un patrón de tinción de RPE:FIGI similar pero más amplio, comparado con la DHFR lisiada, que resulta coherente con una titulación ligeramente menor en el modelo de

5 producción. Comparado con el patrón de tinción de tipo salvaje, ambas agrupaciones con codones desoptimizados presentan un desplazamiento notable en la tinción, con una reducción en la DHFR y más FIGI. Esta diferencia se observa más claramente en el mayor promedio de fluorescencia de las agrupaciones de CDD frente a la agrupaciones de tipo salvaje (figura 7B), y corrobora la conclusión de que la selección con DHFR con codones desoptimizados produce una mayor producción de POI.

10 El aumento observado en la productividad con las agrupaciones de CDD también se confirma en los clones individuales. Clones seleccionados aleatoriamente se expandieron, después se analizaron mediante citometría de flujo y se introdujeron en el modelo de producción de 6 pocillos. Los perfiles FACS de los clones individuales demostraron que las células seleccionadas con codones desoptimizados se tiñen con más intensidad para el POI (figura 8B), pero tienen unos niveles inferiores de DHFR (figura 8A), comparado con los clones seleccionados de DHFR de tipo salvaje. Estos datos son coherentes con los datos de las agrupaciones de transfección. La productividad de los clones en el ensayo de la proteína A se muestra en la figura 9 y demuestra un aumento en la titulación para clones aleatorios procedentes de las agrupaciones seleccionadas de CDD. Las diferencias en la titulación para los clones de CDD son entre 2 y 3 veces mayores que el tipo salvaje.

15

REIVINDICACIONES

- 1.- Un método para aumentar la expresión de proteínas heterólogas en una célula hospedante, que comprende las etapas de cultivar la célula hospedante que comprende una primera secuencia de polinucleótido heteróloga que codifica dicha proteína heteróloga bajo condiciones que permiten la expresión de proteínas, y dicha célula hospedante comprende además una segunda secuencia de polinucleótido que presenta una secuencia codificadora de proteína para una proteína de un marcador seleccionable, y dicho segundo polinucleótido presenta una modificación en la secuencia, comparado con un polinucleótido de tipo salvaje, que codifica dicha proteína del marcador seleccionable, y dicha modificación de la secuencia reduce la eficacia de traducción de un ARNm codificado por dicho segundo polinucleótido, y dicho segundo polinucleótido presenta dicha modificación en la secuencia y dicho polinucleótido de tipo salvaje codifica secuencias de aminoácidos idénticas para dicha proteína del marcador seleccionable; y dicho primer polinucleótido y dicho segundo polinucleótido están codificados sobre un único vector de expresión del factor de alargamiento 1 del hámster chino (CHEF1), en el que el primer polinucleótido y el segundo polinucleótido están cada uno bajo el control transcripcional de promotores diferenciados.
- 2.- El método de la reivindicación 1, en el que la modificación se produce en una región no traducida de dicho segundo polinucleótido que codifica dicha proteína de un marcador seleccionable.
- 3.- El método de la reivindicación 2, en el que la modificación se produce en una región no traducida 5' y/o en una región no traducida 3'.
- 4.- El método de la reivindicación 1, en el que la modificación se produce en una región codificadora de proteína del gen que codifica la proteína del marcador seleccionable.
- 5.- El método de la reivindicación 4, en el que la modificación está a 25, 20, 15, 10 o 5 codones desde un codón de inicio de la región codificadora de proteína para el gen del marcador seleccionable.
- 6.- El método de la reivindicación 1, en el que dicha secuencia codificadora de proteína en dicha segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica dicha proteína del marcador seleccionable, y dicho codón modificado es un codón que no es un codón preferido o es un codón menos preferido para el aminoácido codificado para la célula hospedante.
- 7.- El método de la reivindicación 1, en el que dicha secuencia codificadora de proteína en dicha segunda secuencia de polinucleótido comprende al menos un codón modificado que no es un codón de tipo salvaje en un polinucleótido de tipo salvaje que codifica dicha proteína del marcador seleccionable, y dicha modificación introduce un cambio en la estructura secundaria de dicho ARNm que reduce la eficacia de traducción de dicho ARNm, aumenta el apareamiento de codones en dicho ARNm, aumenta el contenido en G+C de dicho ARNm, o modifica el contenido en A+T de dicho ARNm.
- 8.- El método de la reivindicación 8, en el que al menos 1%, al menos 2%, al menos 3%, al menos 4%, al menos 5%, al menos 6%, al menos 7%, al menos 8%, al menos 9%, al menos 10%, al menos 11%, al menos 12%, al menos 13%, al menos 14%, al menos 15%, al menos 16%, al menos 17%, al menos 18%, al menos 19%, al menos 20%, al menos 21%, al menos 22%, al menos 23%, al menos 24%, al menos 25%, al menos 26%, al menos 27%, al menos 28%, al menos 29%, al menos 30%, al menos 31%, al menos 32%, al menos 33%, al menos 34%, al menos 35%, al menos 36%, al menos 37%, al menos 38%, al menos 39%, al menos 40%, al menos 41%, al menos 42%, al menos 43%, al menos 44%, al menos 45%, al menos 46%, al menos 47%, al menos 48%, al menos 49%, al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% de los codones en dicha secuencia codificadora de proteína del segundo polinucleótido son codones modificados.
- 9.- El método de la reivindicación 1, en el que la célula hospedante se selecciona de una célula eucariota, una célula procariota, una célula de *Escherichia coli*, una célula de levadura, una célula de *Saccharomyces cerevisiae*, una célula de *Pichia pastoris*, una célula de insecto, una célula de *Spodoptera frugiperda*, una célula vegetal, una célula protozoaria, una célula de mamífero, una célula humana, una célula de hámster chino, y una célula de ovario de hámster chino.
- 10.- El método de la reivindicación 1, en el que el segundo polinucleótido comprende el polinucleótido indicado en la figura 2.
- 11.- El método de la reivindicación 1, en el que la proteína del marcador seleccionable es la dihidrofolato reductasa.
- 12.- El método de la reivindicación 11, en el que el segundo polinucleótido comprende la secuencia de la dihidrofolato reductasa lisiada (cr) indicada en la figura 2.

13.- El método de la reivindicación 11, en el que el segundo polinucleótido comprende la secuencia de la dihidrofolato reductasa peor (wst) indicada en la figura 2.

Figura 1

A) Polinucleótido de DHFR de *Mus musculus*

ATG GTT CGA CCA TTG AAC TGC ATC GTC GCC GTG TCC CAA AAT ATG GGG ATT GGC AAG AAC GGA
 GAC CTA CCC TGG CCT CCG CTC AGG AAC GAG TTC AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT
 TCA GTG GAA GGT AAA CAG AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG
 AAG AAT CGA CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA CCA CCA
 CGA GGA GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT GAA CAA CCG GAA
 TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC AGT TCT GTT TAC CAG GAA GCC ATG
 AAT CAA CCA GGC CAC CTC AGA CTC TTT GTG ACA AGG ATC ATG CAG GAA TTT GAA AGT GAC ACG
 TTT TTC CCA GAA ATT GAT TTG GGG AAA TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT
 GAG GTC CAG GAG GAA AAA GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA GAC TAA

B) Polipéptido de DHFR de *Mus musculus*

MVRPLNCIVAVSQNMGIGKNGDLPWPLRNEFKYFQRMTTTTSSVEGKQNLVIMGRKTWFS
 IPEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSS
 VYQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLKGYKLLPEYPGVLSEVQEEKGIKYKF
 EVYEKKD-

Figura 2

>cr dhfr
 ATG GTT CGA CCG CTT AAC TGC ATA GTA GCA GTA TCA CAA AAC ATG GGG ATA GGG AAA AAT GGG
 GAT CTT CCG TGG CCG CCG TTG CGT AAC GAA TTC AAA TAC TTC CAA CGT ATG ACT ACT ACT TCA
 TCA GTA GAA GGG AAA CAA AAC CTT GTA ATA ATG GGG CGT AAA ACA TGG TTC TCA ATA CCG GAA
 AAA AAC CGT CCG CTT AAA GAC CGT ATA AAC ATA GTA CTT TCA CGT GAA CTT AAA GAA CCG CCG
 CGT GGG GCA CAT TTT CTT GCA AAA TCA CTT GAC GAC GCA CTT CGT CTT ATA GAA CAA CCG GAA
 CTT GCA TCA AAA GTA GAC ATG GTT TGG ATA GTA GGG GGG TCA TCA GTA TAC CAA GAA GCA ATG
 AAC CAA CCG GGG CAC CTT CGT CTT TTC GTA ACT CGT ATA ATG CAA GAA TTC GAA TCA GAC ACT
 TTC TTC CCG GAA ATA GAC CTT GGG AAA TAC AAA CTT CTT CCG GAA TAC CCG GGG GTA TTG TCA
 GAA GTA CAA GAA GAA AAA GGG ATA AAA TAC AAA TTC GAA GTA TAC GAA AAA AAA GAC TAG
 >wst dhfr
 ATG GTT CGA CCG CTA AAC TGC ATA GTA GCG GTA TCG CAA AAC ATG GGG ATA GGG AAA AAT GGG
 GAC TTA CCG TGG CCG CCG TTA CGA AAC GAA TTC AAA TAC TTC CAA CGT ATG ACG ACG ACG TCG
 TCG GTA GAA GGG AAA CAA AAC CTA GTA ATA ATG GGG CGT AAA ACG TGG TTT TCG ATA CCG GAA
 AAA AAC CGT CCG CTA AAA GAC CGT ATA AAC ATA GTA CTA TCG CGT GAA CTA AAA GAA CCG CCG
 CGT GGG GCG CAT TTT TTA GCG AAA TCG CTA GAC GAC GCG CTA CGT CTA ATA GAA CAA CCG GAA
 CTA GCG TCG AAA GTA GAC ATG GTT TGG ATA GTA GGG GGG TCG TCG GTA TAT CAA GAA GCG ATG
 AAC CAA CCG GGG CAC TTA CGT TTA TTC GTA ACG CGA ATA ATG CAA GAA TTC GAA TCG GAC ACG
 TTC TTC CCG GAA ATA GAC CTA GGG AAA TAC AAA CTA CTA CCG GAA TAC CCG GGG GTA CTA TCG
 GAA GTA CAA GAA GAA AAA GGG ATA AAA TAC AAA TTC GAA GTA TAC GAA AAA AAA GAC TAG

Figura 3

```

>wt dhfr      #1  ATG GTT CGA CCA TTG AAC TGC ATC GTC GCC GTG TCC CAA AAT ATG GGG ATT GGC AAG AAC GGA
>cr dhfr      #1  ATG GTT CGA CCG CTT AAC TGC ATA GTA GCA GTA TCA CAA AAC ATG GGG ATA GGG AAA AAT GCG
>wst dhfr     #1  ATG GTT CGA CCG CTA AAC TGC ATA GTA GCG GTA TCG CAA AAC ATG GGG ATA GGG AAA AAT GGG
Contig[0001]  #1  ATG GTT CGA CCG CTD AAC TGC ATA GTA GCV GTA TCV CAA AAC ATG GGG ATA GGG AAA AAT GGG
* * * * *
>wt dhfr      #64  GAC CTA CCC TGG CCT CCG CTC AGG AAC GAG TTC AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT
>cr dhfr      #64  GAT CTT CCG TGG CCG CCG TTG CGT AAC GAA TTC AAA TAC TTC CAA CGT ATG ACT ACT ACT TCA
>wst dhfr     #64  GAC TTA CCG TGG CCG CCG TTA CGA AAC GAA TTC AAA TAC TTC CAA CGT ATG ACG ACG ACG TCG
Contig[0001]  #64  GAC CTA CCG TGG CCG CCG TTV CGD AAC GAA TTC AAA TAC TTC CAA CGT ATG ACB ACD ACB TCD
* * * * *
>wt dhfr      #127  TCA GTG GAA GGT AAA CAG AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG
>cr dhfr      #127  TCA GTA GAA GGG AAA CAA AAC CTT GTA ATA ATG GGG CGT AAA ACA TGG TTC TCA ATA CCG GAA
>wst dhfr     #127  TCG GTA GAA GGG AAA CAA AAC CTA GTA ATA ATG GGG CGT AAA ACG TGG TTT TCG ATA CCG GAA
Contig[0001]  #127  TCA GTA GAA GGG AAA CAA AAC CTD GTA ATA ATG GGG CGT AAA ACV TGG TTC TCV ATA CCG GAA
* * * * *
>wt dhfr      #190  AAG AAT CGA CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA CCA CCA
>cr dhfr      #190  AAA AAC CGT CCG CTT AAA GAC CGT ATA AAC ATA GTA CTT TCA CGT GAA CTT AAA GAA CCG CCG
>wst dhfr     #190  AAA AAC CGT CCG CTA AAA GAC CGT ATA AAC ATA GTA CTA TCG CGT GAA CTA AAA GAA CCG CCG
Contig[0001]  #190  AAA AAC CGT CCG CTA AAA GAC CGT ATA AAC ATA GTA CTH TCD CGT GAA CTH AAA GAA CCG CCG
* * * * *
>wt dhfr      #253  CGA GGA GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT GAA CAA CCG GAA
>cr dhfr      #253  CGT GGG GCA CAT TTT CTT GCA AAA TCA CTT GAC GAC GCA CTT CGT CTT ATA GAA CAA CCG GAA
>wst dhfr     #253  CGT GGG GCG CAT TTT TTA GCG AAA TCG CTA GAC GAC GCG CTA CGT CTA ATA GAA CAA CCG GAA
Contig[0001]  #253  CGT GGG GCD CAT TTT CTT GCV AAA TCD CTD GAC GAC GCV CTA CGT CTT ATA GAA CAA CCG GAA
* * * * *
>wt dhfr      #316  TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC AGT TCT GTT TAC CAG GAA GCC ATG
>cr dhfr      #316  CTT GCA TCA AAA GTA GAC ATG GTT TGG ATA GTA GGG GGG TCA TCA GTA TAC CAA GAA GCA ATG
>wst dhfr     #316  CTA GCG TCG AAA GTA GAC ATG GTT TGG ATA GTA GGG GGG TCG TCG GTA TAT CAA GAA GCG ATG
Contig[0001]  #316  CTD GCA TCD AAA GTA GAC ATG GTT TGG ATA GTA GGG GGG TCD TCD GTA TAC CAA GAA GCV ATG
* * * * *
>wt dhfr      #379  AAT CAA CCA GGC CAC CTC AGA CTC TTT GTG ACA AGG ATC ATG CAG GAA TTT GAA AGT GAC ACG
>cr dhfr      #379  AAC CAA CCG GGG CAC CTT CGT CTT TTC GTA ACT CGT ATA ATG CAA GAA TTC GAA TCA GAC ACT
>wst dhfr     #379  AAC CAA CCG GGG CAC TTA CGT TTA TTC GTA ACG CGA ATA ATG CAA GAA TTC GAA TCG GAC ACG
Contig[0001]  #379  AAC CAA CCG GGG CAC CTH CGT CTH TTC GTA ACD CGD ATA ATG CAA GAA TTC GAA TCD GAC ACG
* * * * *
>wt dhfr      #442  TTT TTC CCA GAA ATT GAT TTG GGG AAA TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT
>cr dhfr      #442  TTC TTC CCG GAA ATA GAC CTT GGG AAA TAC AAA CTT CTT CCG GAA TAC CCG GGG GTA TTG TCA
>wst dhfr     #442  TTC TTC CCG GAA ATA GAC CTA GGG AAA TAC AAA CTA CTA CCG GAA TAC CCG GGG GTA CTA TCG
Contig[0001]  #442  TTC TTC CCG GAA ATA GAC CTD GGG AAA TAC AAA CTT CTH CCG GAA TAC CCG GGG GTA CTV TCD
* * * * *
>wt dhfr      #505  GAG GTC CAG GAG GAA AAA GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA GAC TAA
>cr dhfr      #505  GAA GTA CAA GAA GAA AAA GGG ATA AAA TAC AAA TTC GAA GTA TAC GAA AAA AAA GAC TAG
>wst dhfr     #505  GAA GTA CAA GAA GAA AAA GGG ATA AAA TAC AAA TTC GAA GTA TAC GAA AAA AAA GAC TAG
Contig[0001]  #505  GAA GTA CAA GAA GAA AAA GGG ATA AAA TAC AAA TTC GAA GTA TAC GAA AAA AAA GAC TAG
* * * * *

```

Figura 4

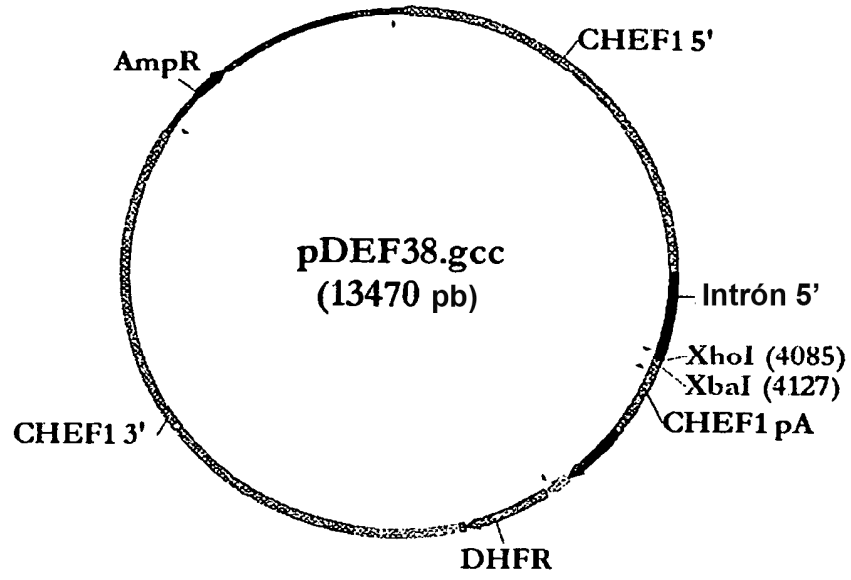


Figura 5

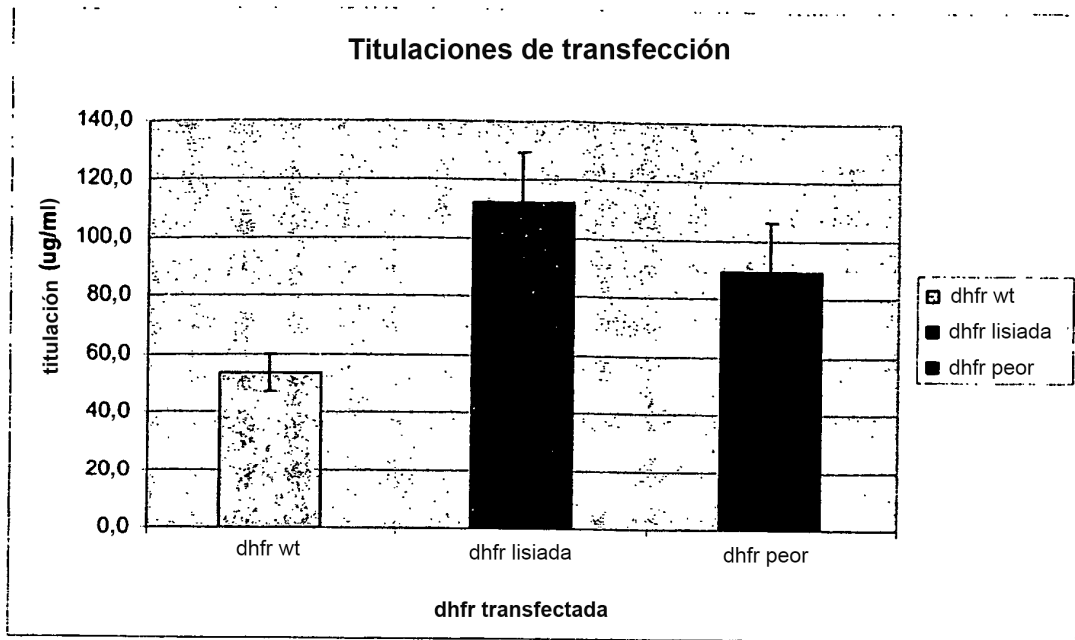


Figura 6

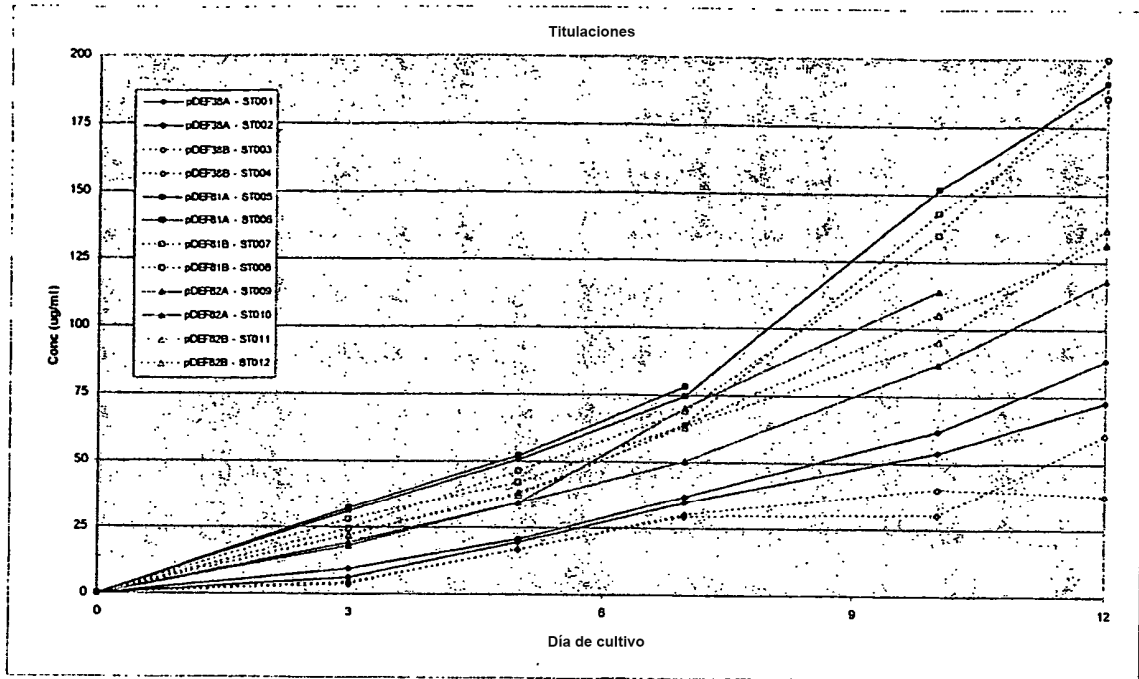


Figura 7A

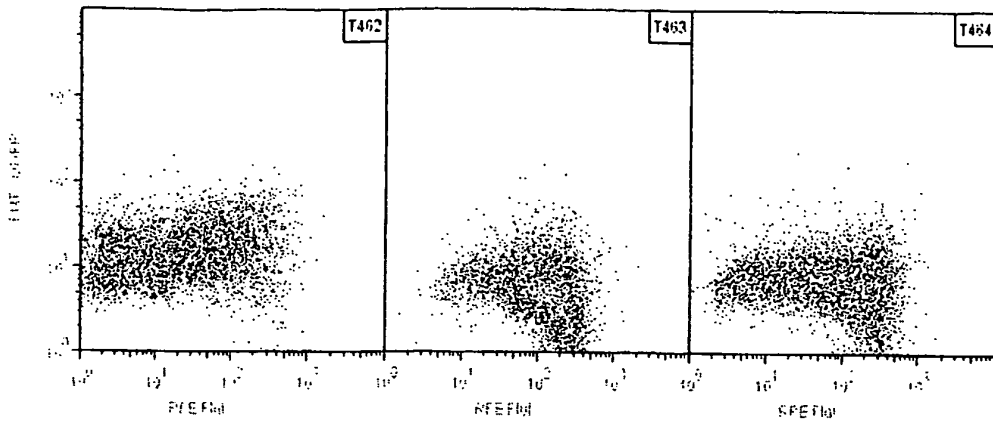


Figura 7B

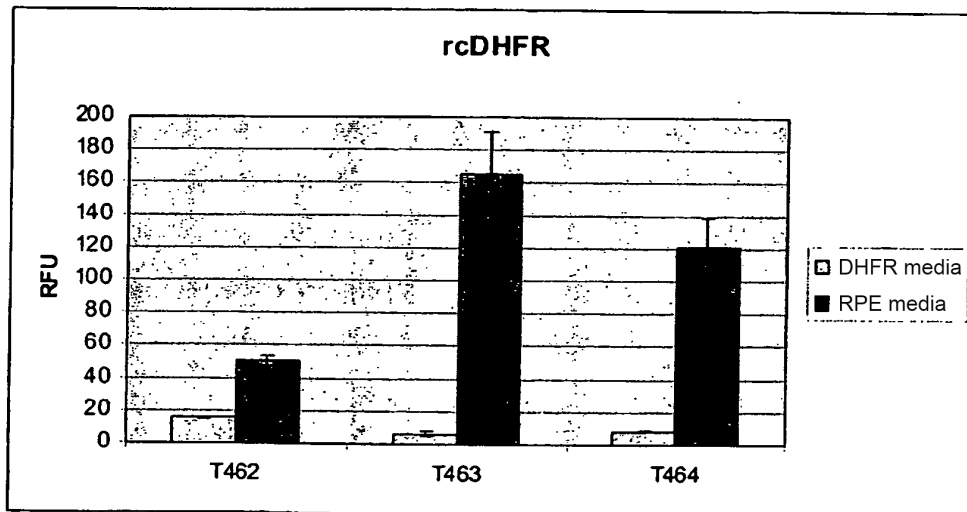


Figura 7:

Figura 8

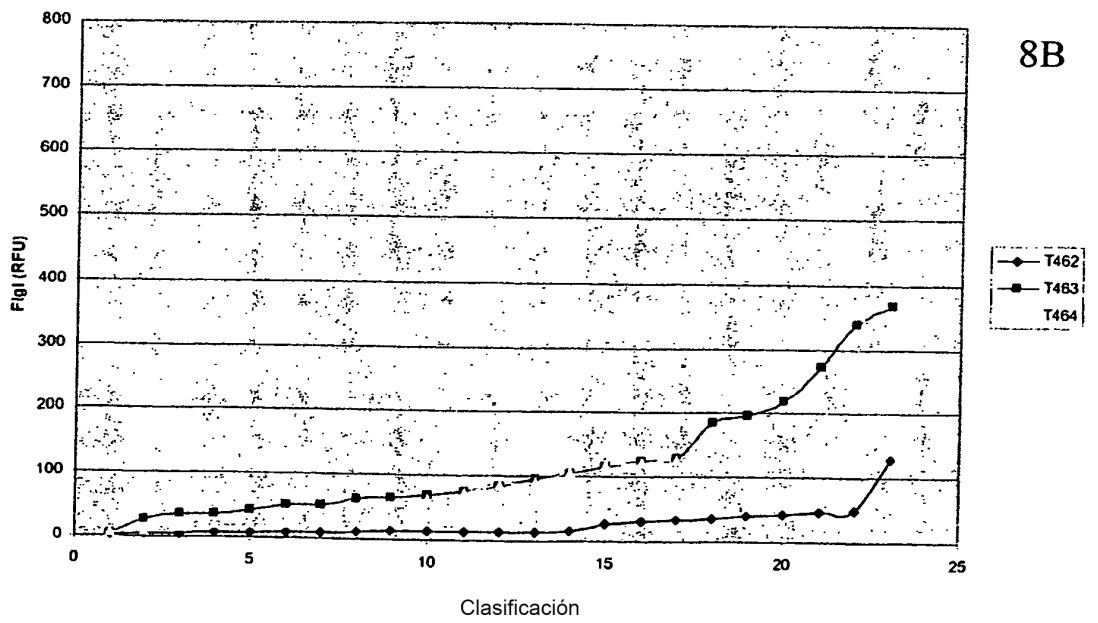
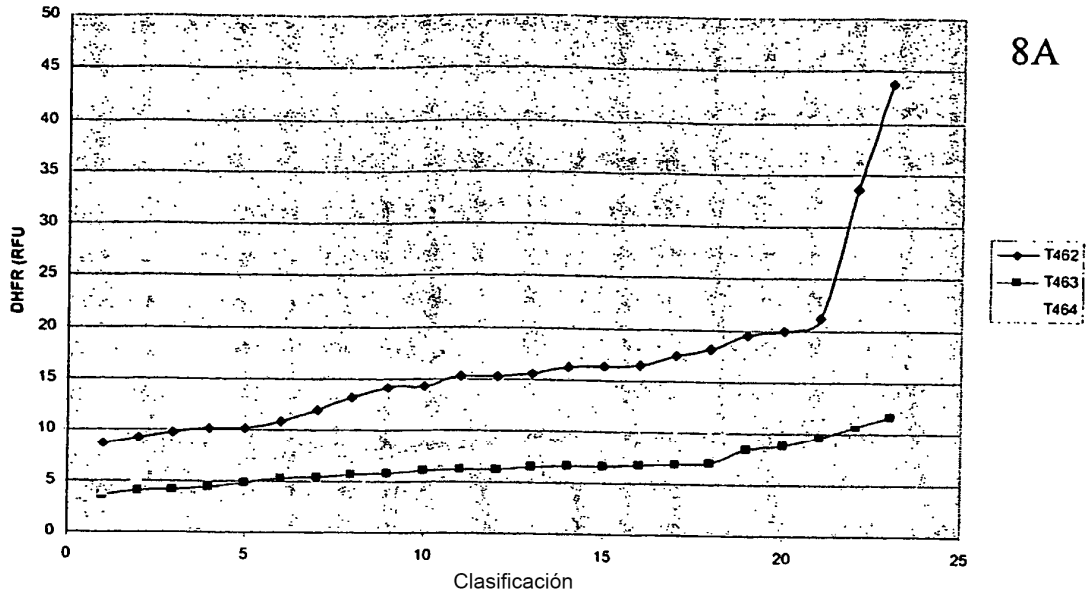


Figura 9

