

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 373 549**

51 Int. Cl.:
C12Q 1/68 (2006.01)
C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **03779044 .1**
96 Fecha de presentación: **02.12.2003**
97 Número de publicación de la solicitud: **1573058**
97 Fecha de publicación de la solicitud: **14.09.2005**

54 Título: **MÉTODO PARA MEJORAR LA PRODUCCIÓN DE PROTEÍNAS.**

30 Prioridad:
18.12.2002 EP 02080347

45 Fecha de publicación de la mención BOPI:
06.02.2012

45 Fecha de la publicación del folleto de la patente:
06.02.2012

73 Titular/es:
CHROMAGENICS B.V.
ARCHIMEDES WEG 4
2333 CN LEIDEN, NL

72 Inventor/es:
OTTE, Arie, Pieter y
VAN BLOKLAND, Henricus, Johannes, Maria

74 Agente: **Arias Sanz, Juan**

ES 2 373 549 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

Método para mejorar la producción de proteínas

La invención se refiere a los campos de bioquímica, biología molecular y farmacología. Más específicamente la presente invención se refiere a la producción de proteínas en una célula (hospedadora).

5 Las proteínas se producen en sistemas para una amplia gama de aplicaciones en biología y biotecnología. Estas incluyen investigación en la función celular y molecular, producción de proteínas como reactivos de diagnóstico o productos biofarmacéuticos y modificación de los caracteres o fenotipos de animales de ganado y cultivos. Los productos biofarmacéuticos son habitualmente proteínas que tienen una función extracelular, tales como anticuerpos para inmunoterapia u hormonas o citocinas para provocar una respuesta celular. Las proteínas con funciones
10 extracelulares salen de la célula mediante la ruta secretora, y se someten a modificaciones tras la traducción durante la secreción (Chevet *et al.* 2001). Las modificaciones (principalmente glicosilación y formación de enlaces disulfuro) no se producen en las bacterias. Además, los oligosacáridos específicos unidos a las proteínas mediante enzimas glicosilantes son específicos para el tipo de célula y especie. Estas consideraciones a menudo limitan la elección de células hospedadoras para la producción de proteínas heterólogas a células eucariotas (Kaufman, 2000). Para la
15 expresión de proteínas terapéuticas humanas, las células hospedadoras tales como bacterias, levaduras o plantas pueden ser inapropiadas. Incluso las diferencias sutiles en la glicosilación de proteínas entre roedores y seres humanos, por ejemplo, puede ser suficiente para hacer que las proteínas producidas en células de roedor sean inaceptables para uso terapéutico (Sheeley *et al.*, 1997). Las consecuencias de la glicosilación inapropiada (es decir no humana) incluyen inmunogenicidad, semivida funcional reducida y pérdida de actividad. Esto limita
20 adicionalmente la elección de células hospedadoras, a líneas celulares humanas o a líneas celulares tales como células de ovario de hámster chino (CHO), que pueden producir glicoproteínas con estructuras de hidratos de carbono similares a la humana (Liu, 1992).

25 Algunas proteínas de interés biotecnológico son funcionales como multímeros, es decir consisten en dos o más cadenas de polipéptidos posiblemente diferentes en su forma biológica y/o biotecnológicamente activa, por ejemplo los anticuerpos (Wright y Morrison, 1997). La producción de tales proteínas multiméricas en sistemas heterólogos es difícil técnicamente debido a varias limitaciones de los sistemas de expresión actuales. Estas limitaciones incluyen (1) dificultades al aislar líneas celulares/células recombinantes que producen los polipéptidos monoméricos en niveles altos (previsibilidad y rendimiento), y (2) disminuciones en los niveles de expresión durante el ciclo de
producción industrial de las proteínas (estabilidad). Estos problemas se describen en más detalle a continuación.

30 (1) Las proteínas recombinantes tales como los anticuerpos que se usan como compuestos terapéuticos necesitan producirse en grandes cantidades. Las células hospedadoras usadas para la producción de proteínas recombinantes deben ser compatibles con la escala de los procesos industriales que se emplean. Específicamente, el sistema de expresión del transgén (o el gen que codifica para una proteína de interés, los dos términos se usan
35 intercambiabilmente en el presente documento) usado para la proteína heteróloga necesita retenerse por las células hospedadoras en una forma activa y estable durante las fases de crecimiento de ampliación a escala y de producción. Esto se logra mediante la integración del transgén dentro del genoma de la célula hospedadora. Sin embargo, la creación de líneas celulares recombinantes mediante medios convencionales es un proceso ineficaz y costoso debido a la imprevisibilidad de la expresión transgénica entre las células hospedadoras recombinantes. La imprevisibilidad es producto de la alta probabilidad de que el transgén se vuelva inactivo debido al silenciamiento
40 génico (McBurney *et al.*, 2002). Usando tecnologías convencionales, la proporción de células hospedadoras recombinantes que producen un polipéptido a niveles altos oscila entre el 1-2%. Con el fin de construir una línea celular que produce dos polipéptidos a niveles altos, los dos transgenes están integrados generalmente de manera independiente. Si los dos transgenes se transfectan simultáneamente en dos ácidos nucleicos separados, la proporción de células que producirán ambos polipéptidos en niveles altos será el producto aritmético de las
45 proporciones para transgenes individuales. Por tanto la proporción de tales líneas celulares recombinantes oscila desde uno de cada 2.500 hasta uno de cada 10.000. Para proteínas multiméricas con tres o más subunidades, las proporciones disminuyen adicionalmente. Estas líneas celulares de alta producción deben identificarse y aislarse posteriormente del resto de la población. Los métodos requeridos para seleccionar estas líneas celulares de alta expresión poco comunes llevan mucho tiempo y son caros.

50 Una alternativa a la transfección simultánea de dos ácidos nucleicos que portan transgenes es la transfección secuencial. En este caso la proporción de clones de alto rendimiento será la suma de las proporciones para transgenes individuales, es decir el 2-4%. Sin embargo la transfección secuencial tiene desventajas (importantes), incluyendo altos costes y mala estabilidad. Los altos costes resultan de diversos factores: en particular, el tiempo y los recursos requeridos para seleccionar líneas celulares de alta expresión se duplican, puesto que debe
55 seleccionarse por separado la alta expresión de cada subunidad. La mala estabilidad global de células hospedadoras que expresan dos polipéptidos es una consecuencia de la inestabilidad inherente de cada uno de los dos transgenes.

60 (2) El silenciamiento de la expresión transgénica durante el cultivo de células hospedadoras prolongado es un fenómeno comúnmente observado. En células de vertebrados puede provocarse mediante la formación de heterocromatina en el locus del transgén, que previene la transcripción del transgén. El silenciamiento transgénico

- es estocástico; puede producirse poco después de la integración del transgén en el genoma, o sólo tras varias divisiones celulares. Esto da como resultado poblaciones de células heterogéneas tras un cultivo prolongado, en las que algunas células continúan expresando niveles altos de proteína recombinante mientras que otras expresan niveles bajos o indetectables de la proteína (Martin y Whitelaw, 1996, McBurney *et al.*, 2002). Una línea celular que se usa para la producción de proteínas heterólogas se deriva de una única célula, sin embargo a menudo se aumenta la escala, y se mantiene durante periodos de tiempo largos, a densidades celulares de más de diez millones de células por mililitro en cultivadores de 1.000 litros o más. Estas poblaciones de células grandes (10^{14} - 10^{16} células) son propensas a graves disminuciones en la productividad debido al silenciamiento transgénico (Migliaccio *et al.*, 2000, Strutzenberger *et al.*, 1999).
- La inestabilidad de la expresión de células hospedadoras recombinantes es particularmente grave cuando se amplifican los números de copias de transgenes en un intento por aumentar los rendimientos. La amplificación transgénica se logra por ejemplo incluyendo un gen de marcador seleccionable tal como dihidrofolato reductasa (DHFR) con el transgén durante la integración (Kaufman, 2000). Concentraciones aumentadas del agente de selección (en el caso de DHFR, el fármaco metotrexato) seleccionado para células que han amplificado el número de genes DHFR en el cromosoma (Kaufman y Sharp 1982). Puesto que el transgén y DHFR están ubicados conjuntamente en el cromosoma, el número de copias del transgén también aumenta. Esto se correlaciona con un aumento en el rendimiento de la proteína heteróloga (Kaufman, 1990). Sin embargo, las repeticiones en tándem de transgenes que resultan de la amplificación son sumamente susceptibles al silenciamiento (Garrick *et al.*, 1998, Kaufman, 1990, McBurney *et al.*, 2002).
- Existe por tanto una necesidad clara de una tecnología de expresión de proteínas (heterólogas) alternativa y específicamente un método de expresión de proteínas que supera los problemas explicados anteriormente. Se necesita aún más un sistema de expresión que i) proporcione alta previsibilidad de la expresión, permitiendo la expresión balanceada de múltiples cadenas, ii) proporcione altos rendimientos y, iii) proporcione estabilidad durante un periodo extendido durante el cual se necesita producir la proteína en grandes cantidades. Esta estabilidad se necesita particularmente cuando están presentes altos números de copias en una célula y es probable que se produzca el silenciamiento.
- En una realización la presente invención usa una secuencia de pausa de transcripción (TRAP) para potenciar una expresión de proteína característica de una unidad de expresión de proteínas. Se cree que una TRAP previene al menos en parte la formación de ARN antisentido o previene al menos en parte que la transcripción entre a dicha unidad de expresión de proteínas. Sin limitarse por la teoría se cree que el presente bloqueo contraintuitivo de la transcripción conduce a la transcripción estable de un transgén. Debido al bloqueo no se forma ARN antisentido y por tanto se inhibe la formación de ARNbc (bicatenario). Esto puede conducir a una reducción o prevención completa del denominado ARNi, que implica la formación de ARNbc de 21 a 23 pares de bases. Se cree que ARNi está implicado en el silenciamiento génico. Una manera de función de la presente invención puede ser que tal silenciamiento inducido por ARNi se previene al menos en parte.
- ¿Cómo puede producirse ARNbc a partir de transgenes que no están diseñados para hacerlo? La situación es lo más fácil de imaginar cuando los transgenes se integran como copias múltiples en orientaciones invertidas en el genoma y cuando como resultado la transcripción de los transgenes es convergente (figura 1A). En este caso la transcripción que empieza en un transgén continua en el siguiente, dando como resultado un ARN que es autocomplementario (sentido y antisentido). La formación de ARNbc también puede producirse con copias múltiples de ácido nucleico que alberga dos unidades de expresión transgénica, gen 1 y gen 2, que están orientadas divergentes (figura 1B). En este ejemplo puede formarse tanto ARNm sentido como ARN antisentido del gen 2: el ARNm del gen 2 sentido mediante el promotor en el ácido nucleico izquierdo que impulsa el gen 2, el ARN del gen 2 antisentido mediante el promotor en el ácido nucleico derecho que impulsa el gen 1. También puede formarse tanto ARNm sentido como ARN antisentido del gen 1: el ARNm del gen 1 sentido mediante el promotor en el ácido nucleico derecho que impulsa el gen 1, el ARN del gen 1 antisentido mediante el promotor en el ácido nucleico izquierdo que impulsa el gen 2.
- Incluso cuando un transgén en un ácido nucleico se integra como única copia, puede formarse ARNbc cuando la transcripción empieza de un promotor endógeno que está ubicado fuera del transgén. Esto puede suceder fácilmente si por casualidad la única copia se integra en una ubicación genómica con un promotor endógeno presente en una orientación tal que se produce ARN antisentido (figura 1C) (Stam *et al.* 2000). Sin embargo, la formación de ARNbc es lo más probable cuando se integran copias múltiples del transgén como repeticiones invertidas, debido a que las cadenas complementarias siempre estarán conectadas. Esto es particularmente relevante puesto que en la mayoría de los casos un transgén se integrará con copias múltiples. Es práctica común que en el extremo 3' de un gen se coloque un terminador transcripcional de SV40. Sin embargo, incluso la presencia de una señal de poliadenilación de este tipo en el sentido de 3' desde la unidad de expresión en el sentido de 5' es insuficiente para prevenir la transcripción ultralectura en la segunda unidad de transcripción en el sentido de 3' (Eszterhas *et al.* 2002).
- Habitualmente se usan secuencias de ADN tales como la señal de poliadenilación de SV40 para terminar la transcripción colocando la señal de poliadenilación de SV40 inmediatamente en el sentido de 3' de un gen que está expresado (figura 2A-C). En otras palabras, debe prevenirse que la transcripción continúe en el sentido de 3' del

gen. En la presente invención se colocan preferiblemente bloqueadores de transcripción (TRAP) tanto en el sentido de 5' como en el sentido de 3' de las unidades de expresión enteras, de tal manera que previenen que entre transcripción a las unidades de expresión, procedente desde el sentido de 5' o desde el sentido de 3' de las unidades de expresión (figura 2A-C). La orientación de TRAP cuando se coloca en el sentido de 3' es opuesta a la orientación habitual de las señales de poliadenilación de SV40 que se colocan en el sentido de 3' de los genes (figura 2A-C).

En una realización, la invención proporciona una unidad de expresión de proteínas que comprende funcionalmente unidos desde 5' hacia 3': i) un promotor, ii) un marco de lectura abierto que codifica para una proteína de interés, y iii) una secuencia señal de terminación de la transcripción, caracterizada porque dicha unidad de expresión de proteínas comprende además al menos una segunda secuencia señal de terminación de la transcripción ubicada en el sentido de 3' de dicho marco de lectura abierto, en la que dicha segunda secuencia señal de terminación de la transcripción está en orientación opuesta de la primera secuencia señal de terminación de la transcripción, y en la que dicha segunda secuencia de terminación de la transcripción se escoge de las secuencias presentadas en la tabla 1. Preferiblemente, dicha segunda secuencia señal de terminación de la transcripción está en una orientación 3'-5' (con respecto a dicha región codificante).

Preferiblemente, dicha secuencia de TRAP reduce la formación de ARN antisentido a un nivel no detectable. Debido a la presencia de dicha TRAP la formación de ARN antisentido se previene al menos en parte y por tanto la cantidad de ARNbc disminuye. Como consecuencia, el nivel de ARNbc pequeños de 21 a 23 pares de bases (ARNi) también disminuye y el ARN (de longitud completa) correspondiente que codifica para una proteína de interés no se degradará. Por tanto, la traducción de dicho ARN correspondiente da como resultado la expresión (aumentada) de una proteína de interés.

Sorprendentemente, tal como se da a conocer en el presente documento con la parte experimental (ejemplo 5) el uso de secuencias de TRAP mejora la estabilidad de la expresión.

En la realización anteriormente explicada, la secuencia de TRAP puede ser por ejemplo una secuencia señal de poliadenilación y/o terminadora, pero en una orientación que difiere de una secuencia de terminador posiblemente usada detrás de un marco de lectura abierto en dicha unidad de expresión de proteínas (véase por ejemplo la figura 2A). Sin embargo, es completamente posible que existan secuencias de TRAP que son bidireccionales. Por tanto en la realización anterior sólo es necesario que dicha TRAP comprenda una función de TRAP en la orientación inversa. En otra realización, la invención proporciona una unidad de expresión de proteínas anterior, caracterizada porque dicha unidad de expresión de proteínas comprende además una secuencia de TRAP adicional ubicada en el sentido de 5' de dicho promotor y previene al menos en parte que entre transcripción a dicha unidad de expresión de proteínas. Preferiblemente, dicha secuencia de TRAP adicional está en una orientación 5'-3' (con respecto a dicha región codificante).

Nuevamente, una secuencia de TRAP usada en la última realización puede ser una secuencia señal de poliadenilación y/o terminadora, pero esta vez la secuencia de TRAP está en una posición inusual con respecto al marco de lectura abierto, debido a que dicha TRAP está ubicada en el sentido de 5' del promotor que impulsa la expresión de dicho marco de lectura abierto.

En esta realización, la presencia de una secuencia de TRAP previene al menos en parte la transcripción de una secuencia promotora ubicada fuera de una unidad de expresión de proteínas. Por tanto, el ARN de la unidad de expresión de proteínas no tiene que competir con otro ARN y por tanto se proporciona un sistema de producción de proteínas más eficaz.

El uso de una TRAP para prevenir al menos en parte la formación de ARN antisentido o para prevenir al menos en parte que la transcripción entre a dicha unidad de expresión de proteínas aísla dicha unidad de expresión de proteínas de efectos negativos, como la formación de ARNi, del exterior de dicha unidad.

La integración de ácido nucleico en el genoma de una célula se produce frecuentemente en los denominados concatémicos. Un concatémico de dos ácidos nucleicos integrados puede tener una de dos organizaciones, o bien los dos ácidos nucleicos forman una repetición directa o bien una repetición invertida. Los concatémicos que tienen tres o más copias integradas pueden tener cualquier combinación de las dos formas básicas, junto con alteraciones ocasionales tales como deleciones/mutaciones, etc. Normalmente las copias integradas tienen sus propias señales de terminación de la transcripción o señales análogas. La presencia de las mismas normalmente da como resultado una reducción de la cantidad de transcripción que procede en el ácido nucleico integrado flanqueante. En realizaciones de la invención que también se dieron a conocer anteriormente, se previene al menos en parte que cualquier transcripción residual entre a la unidad de expresión de proteínas presente en el ácido nucleico integrado conjuntamente mediante una secuencia de TRAP que también está presente en el ácido nucleico integrado conjuntamente. En otra realización, se previene al menos en parte que esta transcripción residual entre en el ácido nucleico integrado conjuntamente mediante la presencia de una secuencia de TRAP dentro del propio ácido nucleico integrado, es decir previniendo así que entre transcripción al ácido nucleico integrado conjuntamente. En esta realización particular, una unidad de expresión de proteínas comprende una secuencia de TRAP en el sentido de 3' de dicha unidad de expresión de proteínas y en una orientación 5'-3' con respecto a la unidad de expresión de

- 5 proteínas, además de la(s) señal(es) usual(es) para terminar la transcripción de dicha unidad de expresión de proteínas. En una versión preferida de esta realización particular, dicha unidad de expresión de proteínas comprende dos secuencias de TRAP consecutivas en la misma orientación, en la que ambas de dichas secuencias de TRAP se ubican en el sentido de 3' del gen de interés y previenen al menos en parte que entre transcripción a la unidad de transcripción de una unidad integrada conjuntamente en un concatémero.
- 10 El término "expresión" se usa normalmente para referirse a la producción de un producto o productos de ARN específicos, o una proteína o proteínas específicas, en una célula. En el caso de productos de ARN, se refiere al proceso de transcripción. En el caso de productos de proteína, se refiere a los procesos de transcripción, traducción y opcionalmente las modificaciones tras la traducción. En el caso de proteínas secretadas, se refiere a los procesos de transcripción, traducción y opcionalmente modificación tras la traducción (por ejemplo glicosilación, formación de enlaces disulfuro, etc.), seguido por la secreción. En el caso de proteínas multiméricas, incluye opcionalmente el ensamblaje de la estructura multimérica a partir de los monómeros polipeptídicos. Los verbos correspondientes del sustantivo "expresión" tienen un significado análogo a dicho sustantivo.
- 15 Una proteína de interés puede ser cualquier proteína y ejemplos no limitativos son enzimas, cadenas de inmunoglobulinas, proteínas terapéuticas tales como proteínas anticancerosas o proteínas de diagnóstico.
- 20 Una proteína se define en el presente documento como que es o bien (i) un producto obtenido mediante los procesos de transcripción y traducción y posiblemente pero no necesariamente dicho producto es parte de una proteína multimérica (por ejemplo una subunidad) y/o bien (ii) un producto obtenido mediante los procesos de transcripción, traducción y modificación tras la traducción. El término "multimérico" o "proteína multimérica" se define normalmente como una proteína que comprende dos o más cadenas de polipéptido posiblemente no idénticas ("monómeros"). Los diferentes monómeros en una proteína multimérica pueden presentarse en números estequiométricamente iguales o distintos. En cualquier caso, la proporción de los monómeros se fija habitualmente mediante la estructura funcional de la proteína multimérica.
- 25 Las expresiones "célula"/"célula hospedadora" y "línea celular"/"línea celular hospedadora" se definen de manera respectiva normalmente como una célula eucariota y poblaciones preferiblemente homogéneas de la misma que se mantienen en cultivo celular mediante métodos conocidos en la técnica, y que tienen la capacidad de expresar proteínas heterólogas. Sin embargo, queda claro que un método según la invención también puede usarse para la expresión de proteínas en procariontes.
- 30 Las expresiones "célula (hospedadora) recombinante" y "línea celular (hospedadora) recombinante" se definen de manera respectiva normalmente como una célula hospedadora y poblaciones preferiblemente homogéneas de la misma en la que se ha introducido un transgén para el fin de la expresión de un producto génico, preferiblemente una proteína o proteínas.
- 35 La expresión "unidad de expresión de proteínas" se define en el presente documento como una unidad que puede proporcionar expresión de proteínas y normalmente comprende un promotor funcional, un marco de lectura abierto que codifica para una proteína de interés y opcionalmente un terminador, todos en configuración operativa. Un promotor funcional es un promotor que puede iniciar la transcripción en una célula particular. Promotores adecuados para obtener expresión en células eucariotas son el promotor de CMV, un promotor EF1-alfa de mamífero, un promotor de ubiquitina de mamífero, o un promotor de SV40, o una parte, derivado y/o análogo funcional de los mismos que tienen la misma función en cuanto a la clase pero no necesariamente en cantidad. Un terminador funcional puede proporcionar la terminación de la transcripción, como en el caso del terminador de SV40. La expresión "un marco de lectura abierto que codifica para una proteína de interés (o un transgén)" se define normalmente como un fragmento de ADN que codifica para un producto o productos de ARN específicos o una proteína o proteínas específicas, y que puede integrarse dentro del genoma de una célula hospedadora. Incluye elementos de ADN requeridos para la transcripción apropiada y, cuando se desea la proteína, los elementos necesarios para la traducción de la(s) región/regiones codificante(s) del transgén. Dicho ADN que codifica para dicha proteína de interés/transgén puede ser o bien un ADN que codifica para un producto obtenido mediante los procesos de transcripción y traducción (y posiblemente pero no necesariamente dicho producto es parte de una proteína multimérica, por ejemplo una subunidad) o un producto obtenido mediante los procesos de transcripción, traducción y modificación tras la traducción.
- 40
- 45
- 50 Una secuencia de TRAP se define funcionalmente en el presente documento como una secuencia que puede prevenir al menos en parte la formación de ARN antisentido o prevenir al menos en parte que entre transcripción a dicha unidad de expresión de proteínas. En otras palabras una secuencia de TRAP, cuando se coloca dentro de una unidad de transcripción, da como resultado un nivel reducido de transcripción del ácido nucleico presente en el lado 3' de la TRAP cuando se compara al nivel de transcripción observado en el ácido nucleico en el lado 5' de la TRAP.
- 55 Cuando en esta solicitud no se hace referencia particular con respecto a la orientación de la TRAP en un constructo particular, es en la orientación que bloquea que entre transcripción a una unidad de transcripción (posible), es decir la unidad de transcripción del ácido nucleico de interés. Preferiblemente, la secuencia de TRAP está físicamente unida a la unidad de expresión de proteínas que pretende aislar de manera transcripcional de cualquiera de las unidades de transcripción flanqueantes, al menos antes de transfectar la unidad dentro del genoma de la célula.
- 60 Tras la integración de la unidad, la unidad y elementos unidos a la misma se unen a secuencias en el genoma y el

elemento presente en el mismo, en el caso de integración de concatémero la unidad integrada puede unirse a unidades integradas conjuntamente u otro ácido nucleico transfectado. En estas realizaciones una TRAP puede estar presente en el sentido de 5' o en el sentido de 3' de la unidad de transcripción que pretende aislar. Cuando está presente en el sentido de 5', la orientación de la TRAP es tal que puede reducir al menos en parte la transcripción que se origina en el sentido de 5' de la unidad de transcripción y la TRAP y que avanza hacia la unidad de transcripción. Cuando está presente en el sentido de 3' de la unidad de transcripción la TRAP está, en estas realizaciones, en una orientación que reduce al menos en parte el origen de la transcripción en el sentido de 3' de la unidad de transcripción a la que está unida a y avanza hacia la unidad de transcripción. La orientación en el sentido de 5' o en el sentido de 3' son normalmente imágenes especulares entre sí. Sin embargo, tal como se mencionó anteriormente, en la situación en la que se forman concatémeros tras la integración de una unidad de expresión de proteínas en el genoma, también es posible prevenir que entre transcripción a una unidad de transcripción integrada conjuntamente flanqueante mediante la colocación de una secuencia de TRAP en el sentido de 3' de la unidad de expresión de proteínas en la orientación que reduce la transcripción que se inicia dentro de la unidad de expresión de proteínas. En esta realización, la TRAP está, antes de la integración, físicamente unida a la unidad de transcripción cuya transcripción puede entrar a una unidad de transcripción flanqueante. Mediante la unión de la TRAP a la unidad antes de la integración, esta posibilidad se reduce al menos en parte. Esta secuencia de TRAP está además de señales de poliadenilación y/o de terminación de la transcripción normales presentes una unidad de expresión de proteínas. Con respecto a la colocación de una TRAP con respecto a la unidad de expresión de proteínas que pretende proteger frente a transcripción entrante se entiende que la TRAP se coloca preferiblemente cerca del casete de expresión que pretende aislar de manera transcripcional. En otras palabras se prefiere que no haya elementos promotores potencialmente activos que activen el dominio que codifica para la proteína entre la TRAP y el dominio que codifica para la proteína de la unidad de expresión que pretende aislar de manera transcripcional, distintos del promotor diseñado para dirigir la transcripción en la unidad de transcripción (es decir necesario para impulsar la proteína de interés).

Tal como se da a conocer en el presente documento dentro de la parte experimental, una secuencia de TRAP puede ser por ejemplo un sitio de poliadenilación y/o un sitio de pausa, en el que la ARN polimerasa II se detiene. Una TRAP puede derivarse de cualquier fuente, siempre que se logre una terminación eficaz de la transcripción. En una realización se identifica una TRAP basándose en su capacidad para prevenir al menos en parte la formación de ARN antisentido o para prevenir al menos en parte que entre transcripción a dicha unidad de expresión de proteínas. El ejemplo 1 proporciona un método para someter a prueba el efecto de supuestas TRAP sobre la transcripción. Se muestra que los elementos STAR 7, 17 y 40 bloquean mal la transcripción.

Por otro lado, determinadas regiones del fago λ así como una secuencia de poliA sintética satisfacen los criterios de una TRAP, puesto que son todas potentes bloqueantes de la transcripción.

En una realización preferida, dicha al menos una secuencia de TRAP está ubicada en el sentido de 5' de dicho promotor y en la que dicha secuencia de TRAP está en una orientación 5'-3'. Aún en otra realización preferida, dicha al menos una secuencia de TRAP está ubicada en el sentido de 3' de dicho marco de lectura abierto y en la que dicha secuencia de TRAP está en una orientación 3'-5' con respecto a la orientación del marco de lectura abierto. Queda claro a partir de los ejemplos dados a conocer en el presente documento que el potencial de secuencias de TRAP depende de la orientación. Por tanto, queda claro que la orientación en la que se aplica una TRAP al flanco de un transgén puede ser de importancia para su funcionamiento apropiado. Sin embargo, queda claro que también hay secuencias de TRAP que actúan independientemente de su orientación.

En una realización preferida, dicha unidad de expresión de proteínas comprende al menos dos secuencias de TRAP. Una versión particularmente preferida de la realización de al menos dos TRAP es la presencia de al menos una TRAP en el sentido de 5' y al menos una TRAP en el sentido de 3' de la unidad de transcripción de interés. Por tanto, preferiblemente, dichas al menos dos secuencias de TRAP están dispuestas de manera que dichas secuencias de TRAP flanquean la combinación formada por dicho promotor y dicho marco de lectura abierto. La figura 2A muestra aún otra disposición. Sin embargo, cuando están presentes múltiples unidades de expresión de proteínas en la misma parte de información genética también es posible inhibir o bloquear al menos parcialmente la transcripción a partir de una unidad de expresión de proteínas a otra unidad de expresión de proteínas. En este caso, se coloca una secuencia de TRAP entre unidades de expresión de proteínas (posiblemente diferentes), la orientación de esta secuencia de TRAP es por supuesto en la orientación 5'-3' con respecto a la transcripción para la que se pretende el bloqueo. En la situación esbozada en la figura 2A, se coloca una secuencia de TRAP entre el terminador del gen bicistrónico y el promotor de SV40. En otra realización, se une una tercera secuencia de TRAP al casete de expresión. Cuando se integran dos casetes de expresión de manera convergente (figura 1A), pueden crearse dominios transcripcionales inertes colocando secuencias de TRAP en una configuración tal que se previene que entre transcripción a las unidades de transcripción. Esta configuración se muestra en la figura 2A. Sin embargo, puede considerarse que la "inertidad" de las unidades de transcripción puede reforzarse cuando también se previene que la transcripción impulsada por CMV en las unidades de la figura 1A escape de la unidad de transcripción. Normalmente, se usa el terminador transcripcional de SV40 para este fin. Sin embargo, este terminador no detiene completamente la transcripción. Por tanto, se incorpora una tercera secuencia de TRAP en el sentido de 5' del elemento STAR en 3' en el casete de expresión (figura 2C). Esta secuencia de TRAP se coloca en una orientación 5'-3', con el fin de detener la transcripción que pudiera escapar a través del terminador

transcripcional de SV40. En esta configuración, todo el casete de expresión se ha vuelto esencialmente inerte para filtraciones de transcripción tanto de entrada como de salida.

5 Por tanto, en otro aspecto, se da a conocer el uso de una TRAP para aislar al menos en parte un elemento genético del avance de la transcripción. En una realización preferida, el elemento genético es un elemento STAR. Por tanto, la invención proporciona además un elemento STAR junto con una secuencia de TRAP de la invención. Preferiblemente, un elemento STAR flanqueado por al menos dos elementos STAR en ambos lados. La orientación del elemento de TRAP en estas realizaciones es tal que se previene al menos en parte la transcripción que avanza desde fuera del elemento STAR hasta el elemento STAR. Esta realización es en particular relevante si estaban presentes dos repeticiones invertidas en el elemento STAR. Estas repeticiones invertidas pueden iniciar la formación de ARNbc. A su vez, esto conduciría a silenciamiento génico de genes adyacentes. Por tanto, esta configuración específica de elementos TRAP-STAR-TRAP no sólo puede prevenir la formación de ARNbc en el elemento genético, es decir, el elemento STAR, sino que también proporciona protección adicional de toda la unidad de expresión.

10 Las TRAP tienen un efecto beneficioso sobre la expresión transgénica, en particular cuando están presentes altos números de copias en una célula. Sin querer limitarse por la teoría, se cree que altos números de copias significan grandes cantidades de secuencias de repeticiones (invertidas) así como numerosas posibilidades de que se produzca ultralectura transcripcional, esto podría conducir a ARNi y silenciamiento génico.

15 Aún en otra realización, la invención proporciona un método para la expresión de al menos una proteína de interés en una célula que comprende proporcionar a dicha célula al menos una unidad de expresión de proteínas tal como se reivindica. Se cree que dicha secuencia de TRAP previene al menos en parte la formación de ARN antisentido o previene al menos en parte que entre transcripción a dicha unidad de expresión de proteínas y en la que dicha unidad de expresión de proteínas comprende además al menos una secuencia anti-represora estabilizante (STAR). Se muestran ejemplos no limitativos de secuencias STAR en la tabla 2.

20 Una secuencia STAR (anti-represora estabilizante) (o elemento STAR; los términos se usarán de manera intercambiable en el presente documento) es un elemento de ADN que se produce de manera natural que se da a conocer en una solicitud de patente en tramitación junto con la presente PCT/NL02/00390 (que reivindica prioridad del documento EP 01202581.3; ambas solicitudes de patente se incorporan en el presente documento como referencia). Pueden identificarse por ejemplo secuencias STAR (tal como se da a conocer por ejemplo en el ejemplo 1 del documento EP 01202581.3) usando un método de detección, y opcionalmente de selección, de una secuencia de ADN con una cualidad de modulación de la transcripción génica. Sin embargo, queda claro a partir de dicha solicitud que pueden obtenerse secuencias STAR de diversos modos. Para tales métodos y los (nuevos) elementos STAR que resultan de los mismos se hace referencia al documento PCT/NL02/00390. Una secuencia STAR comprende la capacidad para influir en la transcripción de genes en cis y/o proporcionar un efecto de potenciación y/o estabilización. El nivel de expresión del transgén es estable a lo largo de muchas generaciones celulares y no manifiesta silenciamiento estocástico. Por tanto, las secuencias STAR confieren un grado de expresión independiente de la posición en transgenes que no es posible con sistemas transgénicos convencionales. La independencia de la posición significa que los transgenes que están integrados en ubicaciones genómicas que darían como resultado silenciamiento transgénico se mantienen, con la protección de los elementos STAR, en un estado transcripcionalmente activo.

25 En una realización preferida, dicha unidad de expresión de proteínas comprende además al menos dos secuencias STAR. Tal como se da a conocer en el presente documento dentro de la parte experimental, las secuencias de TRAP y los elementos STAR pueden proteger a transgenes individuales del silenciamiento. Las unidades de expresión que no están flanqueadas por TRAP y/o elementos STAR pueden experimentar silenciamiento significativo tras sólo 5-60 pases de cultivo, tiempo durante el cual el silenciamiento de las unidades protegidas con los elementos STAR y/o TRAP es insignificante. La presente invención usa preferiblemente secuencias de TRAP y STAR para la producción de una o más proteínas y de ese modo la invención proporciona (1) un aumento de la previsibilidad en la creación de líneas celulares recombinantes que producen eficazmente la proteína heteróloga de interés, (2) un aumento del rendimiento de la proteína heteróloga, (3) expresión estable de la proteína heteróloga, incluso durante cultivo prolongado en ausencia de agente de selección, (4) la invención también proporciona características de expresión transgénica favorable sin amplificación del transgén. El aumento del rendimiento de una proteína heteróloga proporcionado por la invención puede obtenerse a bajos números de copias del transgén, sin coamplificación selectiva usando, por ejemplo, el sistema de DHFR/metotrexato. Esto da como resultado una mayor estabilidad, puesto que el número de copias del transgén es bajo y no es susceptible de disminuir debido a recombinación (McBumey *et al.*, 2002) o silenciamiento génico inducido por repeticiones (Garrick *et al.*, 1998) y/o (5), la amplia aplicabilidad del método de la invención incluye su utilidad en una amplia gama de líneas celulares hospedadoras. Esto es por ejemplo útil/deseable cuando una proteína particular se expresa preferiblemente por una línea celular hospedadora particular (por ejemplo expresión de anticuerpos a partir de líneas celulares hospedadoras derivadas de linfocitos). Las ventajas mencionadas anteriormente no sólo son relevantes para la expresión de una única proteína en una célula. La combinación de una TRAP con un elemento STAR es particularmente favorable cuando va a expresarse una proteína multimérica en la célula. El aumento de previsibilidad de la expresión génica foránea que se obtiene usando esta combinación da como resultado un número de células significativamente superior que producen eficazmente la proteína multimérica, pueden obtenerse rendimientos superiores de proteína multimérica, las líneas celulares producen de manera más estable la proteína multimérica incluso en ausencia de

presión de selección, no es necesaria amplificación y/o puede usarse en una amplia variedad de líneas celulares.

El uso de TRAP y/o STAR para flanquear al menos una unidad de expresión de proteínas es uno de los aspectos de los niveles de expresión equilibrados y proporcionales de una o más proteínas y más específicamente para la expresión de los monómeros de proteínas multiméricas. Las TRAP y STAR crean dominios de cromatina de potencial transcripcional definido y estable. Como resultado, los promotores que impulsan la transcripción de ARNm monocistrónico o bicistrónico funcionan a niveles definidos, estables. Una línea celular hospedadora recombinante creada mediante el método de la invención se identifica fácilmente porque estos niveles dan como resultado que se expresen proporciones apropiadas de cada monómero de la proteína multimérica de interés a altos rendimientos.

Se cree que el uso de STAR y TRAP previene el silenciamiento de la expresión transgénica mediante una acción combinada de mantener fuera la represión asociada a la cromatina (elementos STAR) y simultáneamente crear dominios en los que no se permite la transcripción aberrante y perjudicial (TRAP). Un elemento de TRAP en la orientación de la presente invención proporciona más expresión estable del transgén heterólogo, particularmente en el contexto de un elemento STAR.

Preferiblemente, dichas al menos dos secuencias STAR están dispuestas de manera que dichas secuencias STAR flanquean la combinación formada por dicho promotor y dicho marco de lectura abierto (tal como se esboza en la figura 2A). Incluso más preferiblemente, dichas al menos dos secuencias de TRAP y dichos al menos dos elementos STAR están dispuestos de manera que una primera secuencia de TRAP en 5' está en el sentido de 5' de una primera secuencia STAR y que una segunda secuencia de TRAP en 3' está en el sentido de 3' de una segunda secuencia STAR.

La figura 2 proporciona una representación esquemática, no limitativa de una de las realizaciones de esta parte de la invención. Esta es la configuración de los elementos de ADN de las unidades de expresión en el ácido nucleico así como tras su integración en el genoma. La unidad de expresión uno se muestra en la figura 2A. Contiene un marco de lectura abierto para un transgén (un gen indicador, gen 1). Está en el sentido de 5' del IRES de EMCV atenuado (Martinez-Sals *et al.* 1999; Mizuguchi *et al.* 2000; Rees *et al.* 1996), y del marco de lectura abierto que codifica para la proteína de marcador seleccionable de resistencia a zeocina (zeo). El gen tiene el terminador transcripcional de SV40 en su extremo 3' (t). Este transgén bicistrónico se transcribe a altos niveles a partir del promotor de CMV. A continuación de éste está el marcador seleccionable de resistencia a puomicina (puro), transcrito como un ARN monocistrónico a partir del promotor de SV40. El gen tiene el terminador transcripcional de SV40 en su extremo 3' (t). Elementos STAR flanquean las unidades de expresión. Todo el casete con múltiples genes más STAR está flanqueado por TRAP en una orientación tal que puede prevenirse que entre transcripción a las unidades de expresión en el ácido nucleico o una TRAP está orientada de manera que no se forma o apenas se forma ARN antisentido. Para mejorar adicionalmente la expresión, se coloca una secuencia de TRAP y/o una secuencia STAR entre dicho gen bicistrónico y dicho gen monocistrónico.

En la figura 2B se representa otra configuración de las unidades de expresión. El constructo consiste en dos transgenes (dos genes indicadores o los marcos de lectura abiertos para dos subunidades de una proteína heterodimérica (gen 1 y gen 2) de los cuales el gen 1 está en el sentido de 5' del IRES de EMCV atenuado y la proteína de resistencia a puomicina (puro) y el gen 2 está en el sentido de 5' del IRES de EMCV y la proteína de resistencia a zeocina (zeo). Estos transgenes bicistrónicos se transcriben a altos niveles a partir del promotor de CMV, que están dirigidos en diferentes orientaciones para prevenir la interferencia transcripcional. Ambos genes bicistrónicos tienen el terminador transcripcional de SV40 en sus extremos 3' (t). Los elementos STAR flanquean las unidades de expresión. Todo el casete con múltiples genes más STAR está flanqueado por TRAP en una orientación tal que puede prevenirse que entre transcripción a las unidades de expresión en el ácido nucleico o están orientados de tal manera que no se forma o apenas se forma ARN antisentido.

Queda claro para un experto en la técnica que la secuencia en la que se colocan las TRAP y STAR para flanquear las unidades de expresión puede variar. En el ejemplo dado, los STAR se colocan entre la unidad de expresión y las TRAP, sin embargo, también es posible colocar las TRAP entre la unidad de expresión y el elemento STAR.

También queda claro para un experto en la técnica que son posibles otros marcadores de selección y otras combinaciones de marcadores de selección. Se proporcionan ejemplos de posibles combinaciones de antibióticos en el presente documento. El antibiótico que es particularmente ventajoso es zeocina, porque la proteína de resistencia a zeocina (zeocina-R) actúa uniéndose al fármaco y haciéndolo inocuo. Por tanto, es fácil valorar la cantidad de fármaco que destruye células con bajos niveles de expresión de zeocina-R, mientras que se permite que sobrevivan las que expresan altos niveles. Todas las otras proteínas de resistencia a antibióticos de uso común son enzimas, y por tanto actúan catalíticamente (no 1:1 con el fármaco). Cuando se realiza una selección de dos etapas, es por tanto ventajoso usar una proteína de resistencia a antibióticos con este modo de acción de unión 1:1. Por tanto, el antibiótico zeocina es un marcador de selección preferido. Por conveniencia, el antibiótico zeocina se combina en un método de selección de dos etapas con puomicina-R o blasticidina-R en el segundo gen bicistrónico, y puomicina-R o higromicina-R en el gen monocistrónico.

También queda claro para un experto en la técnica que pueden usarse diferentes promotores siempre que sean funcionales en la célula usada. El promotor de CMV se considera el más fuerte disponible, de modo que se elige

preferiblemente para el gen bicistrónico con el fin de obtener el rendimiento de producto más alto posible. Otros ejemplos de promotores adecuados son por ejemplo promotores de mamíferos para EF1-alfa o el promotor de ubiquitina C. La buena expresión y estabilidad del promotor de SV40 hace que sea muy adecuado para la expresión del gen monocistrónico; se produce suficiente proteína de marcador de selección (por ejemplo la proteína de resistencia a antibióticos puromicina-R en el ejemplo mencionado en el presente documento) para conferir alta expresión de dicho marcador de selección. Por tanto, se usa preferiblemente dicho promotor de SV40 como promotor para impulsar la expresión del marcador de selección.

En una realización preferida, la invención proporciona un método para la expresión de al menos una proteína de interés, que comprende además proporcionar a dicha célula una segunda unidad de expresión de proteínas. Esto es particularmente ventajoso cuando es necesario expresar dos proteínas (por ejemplo dos monómeros) según un método de la invención.

Preferiblemente, al menos una de dichas unidades de expresión de proteínas comprende un gen monocistrónico que comprende un marco de lectura abierto que codifica para dicha proteína de interés y en las que dicho gen monocistrónico está bajo el control de un promotor funcional. La expresión "gen monocistrónico" se define normalmente como un gen que puede proporcionar una molécula de ARN que codifica para una proteína/polipéptido. Aún en otra realización preferida, se usa un gen monocistrónico para la expresión de un marcador de selección. Se proporciona un ejemplo en la figura 2A, en la que se clona un gen de puromicina detrás de un promotor de SV40.

Aún en otra realización preferida, al menos una de dichas unidades de expresión de proteínas comprende un gen bicistrónico que comprende un marco de lectura abierto que codifica para dicha proteína de interés, un sitio de iniciación de la traducción de proteínas con eficacia de traducción reducida, un marcador de selección y en la que dicho gen bicistrónico está bajo el control de un promotor funcional. La expresión "gen bicistrónico" se define normalmente como un gen que puede proporcionar una molécula de ARN que codifica para dos proteínas/polipéptidos.

Tal como se esbozó anteriormente, un método según la invención puede comprender al menos dos secuencias de TRAP y al menos dos secuencias STAR. Preferiblemente, dichas al menos dos secuencias de TRAP son esencialmente idénticas. Incluso más preferiblemente, dichas al menos dos secuencias STAR son esencialmente idénticas.

Secuencias de TRAP y/o STAR esencialmente idénticas se definen en el presente documento como secuencias de TRAP y/o STAR que son idénticas en sus dominios importantes (los dominios de confieren la cualidad de potenciación o estabilización de la transcripción), pero que pueden variar dentro de sus dominios menos importantes, por ejemplo una mutación puntual, delección o inserción en una posición menos importante dentro de la secuencia de TRAP y/o STAR. Preferentemente, dichas secuencias de TRAP y/o STAR esencialmente idénticas proporcionan cantidades iguales de actividad de potenciación o estabilización de la transcripción. Se esbozan ejemplos de secuencias de TRAP y/o STAR adecuadas en la parte experimental en el presente documento.

Aún otra característica preferida de un método según la invención es la introducción de un sitio de unión interna al ribosoma (débil) (IRES) como ejemplo de un sitio de iniciación de la traducción de proteínas con una eficacia de traducción reducida, entre el marco de lectura abierto de la proteína de interés y el marco de lectura abierto del marcador de selección. En combinación con por ejemplo la secuencia de TRAP y/o STAR, este componente de la presente invención comprende una mejora marcada en sistemas transgénicos para la expresión de dos o más proteínas.

Se conocen elementos de sitio de unión interna al ribosoma (IRES) de genes virales y de mamífero (Martinez-Salas, 1999), y también se han identificado en selecciones de oligonucleótidos sintéticos pequeños (Venkatesan y Dasgupta, 2001). Se ha analizado en detalle el IRES del virus de la encefalomiocarditis (Mizuguchi *et al.*, 2000). Un IRES es un elemento codificado en el ADN que da como resultado una estructura en el ARN transcrito a la que pueden unirse ribosomas eucariotas e iniciar la traducción. Un IRES permite que se produzcan dos o más proteínas a partir de una única molécula de ARN (la primera proteína se traduce por ribosomas que se unen al ARN en la estructura de caperuza de su extremo 5' terminal (Martinez-Salas, 1999)). La traducción de proteínas a partir de elementos IRES es menos eficaz que la traducción dependiente de la caperuza: la cantidad de proteína a partir de marcos de lectura abiertos (ORF) dependientes de IRES oscila entre menos del 20% y el 50% de la cantidad del primer ORF (Mizuguchi *et al.*, 2000). Esto hace que los elementos IRES no sean deseables para la producción de todas las subunidades de una proteína multimérica a partir de un ARN mensajero (ARNm), puesto que no es posible lograr una expresión equilibrada y proporcional de dos o más monómeros de proteína a partir de un ARN bicistrónico o multicistrónico. Sin embargo, la eficacia reducida de la traducción dependiente de IRES proporciona una ventaja que se aprovecha por la presente invención. Además, la mutación de los elementos IRES puede atenuar su actividad, y disminuir la expresión a partir de los ORF dependientes de IRES hasta menos del 10% del primer ORF (Lopez de Quinto y Martinez-Salas, 1998, Rees *et al.*, 1996). La ventaja aprovechada por la invención es la siguiente: cuando el ORF dependiente de IRES codifica para una proteína de marcador seleccionable, su bajo nivel relativo de traducción significa que deben producirse altos niveles de transcripción absolutos con el fin de que se seleccione la célula hospedadora recombinante. Por tanto, los aislados de células hospedadoras recombinantes

expresarán por necesidad altas cantidades del ARNm transgénico. Puesto que la proteína recombinante se traduce a partir del ORF dependiente de la caperuza, puede producirse en abundancia dando como resultado altos rendimientos de producto.

5 Queda claro para un experto en la técnica que pueden hacerse cambios en el IRES sin alterar la esencia de la función del IRES (por tanto, proporcionando un sitio de iniciación de la traducción de proteínas con una eficacia de traducción reducida), dando como resultado un IRES modificado. Por tanto también se incluye en esta invención el uso de un IRES modificado que todavía puede proporcionar un pequeño porcentaje de traducción (en comparación con una traducción de caperuza en 5').

10 Cuando se usa un método según la invención para la expresión de dos o más proteínas codificadas por diferentes unidades de expresión de proteínas, preferiblemente cada una de dichas unidades de expresión de proteínas se encuentra en un ADN-portador separado. En cada unidad de transcripción, el ORF del monómero se produce mediante traducción dependiente de la caperuza eficaz. Esta característica de la invención contribuye a que se aislen células hospedadoras recombinantes que tienen altos rendimientos de cada monómero, a niveles que están equilibrados y son proporcionales a la estequiometría de la proteína multimérica. El aumento de previsibilidad a la
15 que se aíslan tales células hospedadoras recombinantes da como resultado una mejora en la eficacia de la selección de tales aislados en un factor de diez o más. En una realización preferida, dicho ADN-portador es un vector (o un ácido nucleico aislado o recombinante, un vector preferido es un plásmido; los términos se usan de manera intercambiable en el presente documento). En otra realización dicho vector es un vector viral y en una realización más preferida dicho vector viral es un vector adenoviral o un vector retroviral. Queda claro para el experto en la técnica que también pueden usarse otros vectores virales en un método según la invención.

Sistemas de expresión convencionales son moléculas de ADN en forma de un plásmido recombinante o un genoma viral recombinante, aunque existen otros medios que no se excluyen de ningún modo. El ácido nucleico, preferiblemente un plásmido o el genoma viral, se introduce en células (hospedadoras de mamífero) y se integra en sus genomas mediante métodos conocidos en la técnica. La presente invención también usa estos tipos de
25 moléculas de ADN para suministrar su sistema de expresión transgénica mejorada. Una realización preferida de la invención es el uso de ácido nucleico, preferiblemente ADN de plásmido para el suministro del sistema de expresión. Un plásmido contiene varios componentes: componentes convencionales, conocidos en la técnica, son un origen de replicación y un marcador seleccionable para la propagación del plásmido en células bacterianas; un marcador seleccionable que funciona en células eucariotas para identificar y aislar células hospedadoras que portan un sistema de expresión transgénica integrado; la proteína de interés, cuya transcripción a alto nivel se ocasiona preferiblemente mediante un promotor que es funcional en la célula hospedadora (por ejemplo el potenciador/promotor temprano inmediato principal de citomegalovirus humano, pCMV (Boshart *et al.*, 1985)); y terminadores transcripcionales virales (por ejemplo el sitio de poliadenilación de SV40 (Kaufman y Sharp, 1982)) para el transgén de interés y el marcador seleccionable. El vector usado puede ser cualquier vector adecuado para,
35 por ejemplo, clonar ADN y que puede usarse en un sistema de transcripción. Cuando se usan células hospedadoras se prefiere que el vector sea un vector de integración, sin embargo, también puede usarse un vector de replicación episomal. En un vector episomal, se evitan efectos debidos a diferentes sitios de integración del vector. Elementos de ADN que flanquean al vector en el sitio de integración pueden tener efectos sobre el nivel de la transcripción del promotor y de ese modo imitan efectos de fragmentos que comprenden secuencias de ADN con una cualidad de modulación de la transcripción génica. En una realización preferida, dicho vector comprende un origen de replicación del virus de Epstein-Barr (VEB), OriP y un antígeno nuclear (EBNA-1). Tales vectores pueden replicarse en muchos tipos de células eucariotas y se ensamblan para dar cromatina en condiciones apropiadas. Otro método para reducir al menos en parte los efectos del sitio de integración sobre la unidad de expresión es el uso de uno o más de los elementos STAR anteriormente mencionados. Por tanto, la invención proporciona además un vector que comprende un elemento de TRAP y un elemento STAR, preferiblemente en combinación con una unidad de expresión.

Preferiblemente, se usa un método según la invención para expresar/producir al menos una cadena de inmunoglobulina como dicha al menos una proteína de interés, por ejemplo una cadena pesada de inmunoglobulina o una cadena ligera de inmunoglobulina. Según esta realización se obtiene una proteína multimérica, un anticuerpo.
50 Queda claro para el experto en la técnica que es posible proporcionar una célula que expresa una cadena pesada de inmunoglobulina a partir de una unidad de expresión de proteína y una cadena ligera de inmunoglobulina a partir de otra unidad de expresión de proteína codificando una tercera unidad de expresión de proteína para un componente de secreción o una cadena de unión. De esta manera se proporciona la producción de por ejemplo sIgA e IgM pentamérica. Preferiblemente, dicha primera proteína de interés y dicha segunda proteína de interés comprenden al menos la parte variable de una cadena ligera de inmunoglobulina y una cadena pesada de inmunoglobulina. Preferiblemente, dicha primera proteína de interés comprende al menos la parte variable de una cadena pesada de inmunoglobulina, mientras que dicha segunda proteína de interés comprende una cadena ligera de inmunoglobulina o derivado y/o análogo de la misma. Esta realización garantiza que una proporción aumentada de las células presentarán una tendencia a sobreproducir ligeramente cadena pesada de inmunoglobulina permitiendo de ese modo una producción más eficaz de una proteína multimérica. La tecnología de
55 inmunoglobulinas está muy avanzada en la actualidad y es posible generar dominios que codifican para anticuerpos que no tienen un anticuerpo complementario en la naturaleza, es decir, un anticuerpo completamente artificial. Tales anticuerpos también están dentro del alcance de la presente invención. Para un resumen de la tecnología relevante para anticuerpos, su selección y producción se hace referencia a (Chad, HE, y Chamow, SM, 2001)

- Las mejoras proporcionadas por un método según la invención tienen tres aspectos integrados. (1) Con los sistemas existentes, sólo pueden crearse líneas celulares recombinantes que expresan simultáneamente cantidades aceptables de los monómeros de proteínas multiméricas a frecuencias muy bajas; la presente invención aumenta la previsibilidad de crear líneas de células hospedadoras recombinantes de alto rendimiento en un factor de diez o más. (2) Los sistemas existentes no proporcionan cantidades estequiométricamente equilibradas y proporcionales de las subunidades de proteínas multiméricas; la presente invención garantiza que los niveles de expresión de las subunidades serán equilibrados y proporcionales. (3) Los sistemas existentes no proporcionan un medio de protección de los transgenes que codifican para las subunidades de proteína frente al silenciamiento transgénico.
- 5 El ejemplo 1 proporciona un método para identificar una secuencia de TRAP y por tanto en otra realización la presente invención también proporciona un método para identificar una secuencia de pausa de transcripción (TRAP) que comprende
- 10 - proporcionar a una célula un ácido nucleico, preferiblemente un plásmido que comprende
- una secuencia promotora
 - una secuencia intermedia (IV) en el sentido de 3' de dicho promotor
- 15 - una supuesta secuencia de TRAP ubicada en dicha IV
- una secuencia cuyo producto es detectable y cuya secuencia se ubica en el sentido de 3' de dicha IV
- determinar la cantidad de dicho producto detectable y comparar dicha cantidad con la cantidad de producto obtenido en una célula que está provista de un ácido nucleico de control, preferiblemente un plásmido sin dicha supuesta secuencia de TRAP.
- 20 La clonación de la supuesta secuencia de TRAP se realiza en la secuencia intermedia para evitar la posibilidad de que la adición de una secuencia de cómo resultado un aumento de la inestabilidad del ARN. Esto también daría como resultado una señal de ARN inferior en una transferencia, pero esto no tendría nada que ver con el bloqueo de la transcripción. Colocar la secuencia que va a someterse a prueba en secuencias intermedias da como resultado la transcripción de esa secuencia para dar ARN, pero posteriormente se corta y empalma, de modo que se forma un ARNm funcional, en este caso particular codA. Esto sucede independientemente de si había una secuencia adicional dentro de la secuencia intermedia o no. Por tanto, cualquier disminución en la señal de ARNm no se debe a pérdida de estabilidad del ARN, sino a una consecuencia directa de la terminación de la transcripción debida a la secuencia de TRAP.
- 25 En una realización preferida, la invención proporciona un método para identificar una secuencia de pausa de transcripción (TRAP) que comprende
- 30 - proporcionar a una célula un ácido nucleico, preferiblemente un plásmido que comprende
- una secuencia promotora
 - una secuencia intermedia (IV) en el sentido de 3' de dicho promotor
 - una supuesta secuencia de TRAP ubicada en dicha IV
- 35 - una secuencia cuyo producto es detectable y cuya secuencia se ubica en el sentido de 3' de dicha IV
- dicho ácido nucleico, preferiblemente un plásmido comprende además un marcador de selección ubicado fuera de la combinación de dicho promotor, IV, supuesta TRAP y dicha secuencia cuyo producto es detectable
- seleccionar una célula por medio de dicho marcador de selección de dicho ácido nucleico, preferiblemente un plásmido, obteniendo de ese modo una célula que comprende dicho ácido nucleico, preferiblemente un plásmido
- 40 - determinar la cantidad de dicho producto detectable y comparar dicha cantidad con la cantidad de producto obtenido en una célula que está provista de un ácido nucleico de control, preferiblemente un plásmido sin dicha supuesta secuencia de TRAP.
- La expresión "marcador de selección o marcador seleccionable" se usa normalmente para hacer referencia a un gen y/o proteína cuya presencia puede detectarse directa o indirectamente en una célula, por ejemplo un gen y/o una proteína que inactiva un agente de selección y protege a la célula hospedadora frente a los efectos letales o inhibidores del crecimiento del agente (por ejemplo un gen y/o proteína de resistencia a antibióticos). Otra posibilidad es que dicho marcador de selección induzca fluorescencia o un depósito de color (por ejemplo proteína fluorescente verde y derivados, luciferasa o fosfatasa alcalina).
- 45 La expresión "agente de selección" se define normalmente como un compuesto químico que puede destruir o retrasar el crecimiento de células hospedadoras (por ejemplo un antibiótico).
- 50

El término “selección” se define normalmente como el proceso de usar un marcador de selección/marcador seleccionable y un agente de selección para identificar células hospedadoras con propiedades genéticas específicas (por ejemplo que la célula hospedadora contiene un transgén integrado en su genoma).

5 Preferiblemente, la invención proporciona un método en el que dicha secuencia cuyo producto es detectable es un gen suicida. Un gen suicida se define normalmente como un gen cuyo producto puede destruir, o bien directa o bien indirectamente, una célula. Más preferiblemente, dicho gen suicida es *codA* o *codA::upp*. Incluso más preferiblemente, dicho producto detectable es ARNm. Sin embargo, queda claro para un experto en la técnica que también puede usarse una proteína como producto detectable. En este caso se determinan cantidades/niveles de proteína por ejemplo mediante inmunotransferencia de tipo Western o detectando directamente la proteína, por ejemplo GFP o realizando una reacción enzimática (color) basándose en las propiedades de la proteína correspondiente.

15 El uso de un gen suicida, por ejemplo *codA* o *codA::upp* es particularmente ventajoso para la selección de una biblioteca de secuencias. Cuando se clonan secuencias de una biblioteca en la secuencia intermedia (IV), una secuencia de TRAP se identifica fácilmente porque el gen suicida no se transcribe ni se traduce y por tanto el producto letal de dicho gen suicida no se produce y la célula que comprende dicha secuencia de TRAP sobrevive. Cuando la secuencia clonada no es una secuencia de TRAP, la célula muere debido al producto letal formado. Queda claro que pueden usarse diferentes tipos de genes suicidas. El gen *codA* codifica para la enzima citosina desaminasa, enzima que convierte citosina en uracilo. Puede usarse *codA* como gen suicida metabólico en combinación con el profármaco 5-fluorocitosina. La enzima puede convertir el profármaco no tóxico en monofosfato de 5-fluorouracilo que destruye las células alterando la síntesis de ADN, desencadenando de ese modo la apoptosis. *CodA::upp* es una fusión entre un gen de citosina desaminasa y un gen de uracilo fosforibosil transferasa. Ambas enzimas actúan sinérgicamente para convertir 5-fluorocitosina en monofosfato de fluorouracilo, un compuesto tóxico. La fusión de los genes conduce a un sistema más eficaz. Otro ejemplo de un gen suicida y un profármaco no tóxico es timidina cinasa y ganciclovir. Sin embargo, queda claro que también es posible usar un gen suicida que no depende de la presencia de un profármaco.

Por tanto, la invención también proporciona un método para identificar una secuencia de TRAP que comprende

- proporcionar a una célula un ácido nucleico, preferiblemente un plásmido que comprende

- una secuencia promotora

- una secuencia intermedia (IV) en el sentido de 3' de dicho promotor

30 - una supuesta secuencia de TRAP ubicada en dicha IV

- una secuencia que codifica para un producto suicida y cuya secuencia se ubica en el sentido de 3' de dicha IV

- determinar si dicha célula sobrevive.

Preferiblemente, dicha supuesta secuencia de TRAP se deriva de una biblioteca. Aún en otra realización preferida, se usa un profármaco que se convierte en un compuesto tóxico por el producto de dicho gen suicida.

35 En el ejemplo 1, se usa el marco de lectura abierto de *codA::upp* como secuencia cuyo producto es detectable y se determina la cantidad de ARN. Se describen secuencias de TRAP según la invención en la parte experimental y en la tabla 1. Preferiblemente, dicha secuencia de TRAP comprende la secuencia de lambda 35711-38103 tal como se representa en la tabla 1. En otra realización preferida, dicha secuencia de TRAP comprende una secuencia de poliA preferiblemente una secuencia de poliA sintética (SPA), por ejemplo una secuencia de SPA tal como se representa en la tabla 1. Aún en otra realización preferida, dicha secuencia de TRAP comprende una combinación de una SPA y la señal de pausa del gen de globina $\alpha 2$ humana, por ejemplo una combinación de una SPA y la señal de pausa del gen de globina $\alpha 2$ humana tal como se representa en la tabla 1.

45 En el presente documento se define un equivalente funcional y/o un fragmento funcional de una secuencia representada en la tabla 1 ó 2 tal como sigue. Un equivalente funcional de una secuencia tal como se representa en la tabla 1 ó 2 es una secuencia derivada con la información facilitada en la tabla 1 ó 2. Por ejemplo, una secuencia que puede derivarse de una secuencia en la tabla 1 mediante delección, modificación y/o inserción de bases en o a partir de una secuencia indicada en la tabla 1 ó 2, en la que dicha secuencia derivada comprende la misma actividad en cuanto a la clase, no necesariamente en cantidad, de una secuencia tal como se representa en la tabla 1 ó 2. Un equivalente funcional es además una secuencia que comprende una parte de dos o más secuencias representadas en la tabla 1 ó 2. Un fragmento funcional de una secuencia en la tabla 1 ó 2 puede obtenerse por ejemplo por delecciones del extremo 5' o el extremo 3' o del interior de dichas secuencias o cualquier combinación de las mismas, en la que dicha secuencia derivada comprende la misma actividad en cuanto a la clase, no necesariamente en cantidad.

55 Aún en otra realización, se da a conocer el uso de una secuencia de TRAP para prevenir al menos en parte la formación de ARN antisentido. Dicha secuencia de TRAP se selecciona de las secuencias representadas en la tabla

1. Incluso más preferiblemente dicha secuencia de TRAP se combina con una secuencia STAR tal como se representa en la tabla 2. El uso según la invención es particularmente ventajoso cuando se aplica a la expresión de al menos una proteína de interés.

5 Aún en otra realización, la invención proporciona una unidad de expresión de proteína que comprende un promotor funcionalmente unido a un marco de lectura abierto que codifica para una proteína de interés, caracterizada porque dicha unidad de expresión de proteína comprende además al menos una secuencia de TRAP, en la que dicha secuencia de TRAP impide al menos en parte la formación de ARN antisentido. En una realización preferida, dicha al menos una secuencia de TRAP está ubicada en el sentido de 3' de dicho marco de lectura abierto y dicha secuencia de TRAP está en una orientación 3'-5'. Incluso más preferido, dicha unidad de expresión de proteína comprende al menos dos secuencias de TRAP. Preferiblemente, dichas al menos dos secuencias de TRAP están dispuestas de manera que dicha unidad de expresión de proteína está flanqueada en ambos lados por al menos una secuencia de TRAP.

15 En otra realización preferida, dicha unidad de expresión de proteína comprende además al menos una secuencia anti-represora estabilizante (STAR). En una realización incluso más preferida, dicha unidad de expresión de proteína comprende además al menos dos secuencias STAR. Preferiblemente, dichas al menos dos secuencias STAR están dispuestas de tal manera que dichas secuencias STAR flanquean la combinación formada por dicho promotor y dicha región codificante. Incluso más preferiblemente, dichas al menos dos secuencias de TRAP y dichos al menos dos STAR están dispuestos de manera que una primera secuencia de TRAP 5' está en el sentido de 5' de una primera secuencia STAR y que una segunda secuencia de TRAP 3' está en el sentido de 3' de una segunda secuencia STAR.

20 Aún en otra realización preferida la invención proporciona una unidad de expresión de proteína, en la que dichas al menos dos secuencias de TRAP son esencialmente idénticas y/o dichas al menos dos secuencias STAR son esencialmente idénticas.

25 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. Aún en 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 están presentes estas dos unidades de expresión de proteína en la misma célula (hospedadora), se ensambla una proteína multimérica y más específicamente un anticuerpo.

30 La invención también incluye una célula (hospedadora) que comprende al menos una unidad de expresión de proteína según la invención. Entonces se usa una célula (hospedadora) de este tipo por ejemplo para procesos de producción a gran escala.

La invención también incluye una célula que puede obtenerse según uno cualquiera de los métodos tal como se describen en el presente documento.

35 En una realización preferida, una célula según la invención es una célula vegetal. El ARNi desempeña un papel importante en la célula vegetal y por tanto aplicar un método según la invención a una célula vegetal es particularmente ventajoso. Ejemplos de una célula vegetal dicotiledónea son una célula de patata, una célula de tomate o una célula de *Arabidopsis*. Ejemplos de una célula monocotiledónea son una célula de arroz, una célula de trigo o una célula de cebada.

40 Aún en otra realización, la invención proporciona una línea celular que comprende una célula tal como se describió anteriormente. Preferiblemente, dicha línea celular comprende una línea celular de osteosarcoma U-2 OS, CHO, 293, mieloma HuNS-1, retinoblastoma WERI-Rb-1, BHK, Vero, mieloma de ratón no secretante Sp2/0-Ag 14, mieloma de ratón no secretante NSO, o de carcinoma de glándula suprarrenal NCI-H295R.

Una célula o una línea celular según la invención es particularmente ventajosa cuando se usa para la producción de una molécula proteica.

45 Además, la invención también proporciona un método para producir una proteína de interés que comprende cultivar una célula o una línea celular según la invención y recoger dicha proteína de interés a partir del cultivo correspondiente.

50 La evaluación anterior y los siguientes ejemplos se proporcionan para fines ilustrativos, y no se pretende que limiten el alcance de la invención tal como se reivindica en el presente documento. Simplemente proporcionan algunas de las realizaciones preferidas de la invención. Las modificaciones y variaciones, que se le pueden ocurrir a un experto en la técnica, están dentro del alcance pretendido de esta invención. Varias otras realizaciones se aplican a la presente invención, incluyendo: otros genes marcadores seleccionables; otros elementos de IRES o medios de atenuación de la actividad de IRES; otros elementos que afectan a la transcripción incluyendo promotores, potenciadores, intrones, terminadores y sitios de poliadenilación; otros órdenes y/o orientaciones de los genes monocistrónicos y bicistrónicos; otros elementos anti-represores o partes, derivaciones y/o análogos de los mismos; otros sistemas de vector para el suministro de las moléculas de ADN de la invención en células hospedadoras eucariotas; y aplicaciones del método de la invención a otros sistemas transgénicos.

Parte experimental y resultados

Ejemplo 1: Identificación de una secuencia de pausa de transcripción (TRAP)

Un objeto de esta invención es identificar elementos de ADN que actúan como secuencia de pausa de transcripción (TRAP). En este ejemplo se proporciona un método para identificar secuencias de TRAP.

5 Materiales y métodos

Plásmidos

Para construir pIRES-6, se toma pIRES (Clontech) como plásmido de partida. Se corta pIRES con *Bgl*II y *Dra*I, y se ligan el promotor de CMV, la secuencia intermedia (IV), el IRES y la señal de poliadenilación de SV40 en pBSKS (Stratagene), se corta con *Bam*HI y *Eco*RV para crear pIRES-1. Se ligan los oligos STOP 1 (CTAGCTAAGTAAGTAAGCTTGG) y STOP 2 (AATTCCAAGCTTACTTA CTTAG) en pIRES1 cortado con *Nhe*I-*Xho*I, creando pIRES-2. Esto da como resultado tres codones de terminación, en tres marcos de lectura diferentes, delante de IRES. Se aparearon los oligos *Bam*HI-*Bgl*II-*Asc*I (TTAAGGATCCAGATCTGGCGCGCC) y *Asc*I-*Bgl*II-*Bam*HI (TTAAGGCGCGCCAGATCTGGATCC) y se ligaron en pIRES-2 cortado con *Bsa*I, creando pIRES-3. De esta manera, se crean sitios *Asc*I, *Bam*HI y *Bgl*II en las secuencias intermedias. Se corta pORFCODA::UPP (InvitroGen) con *Nco*I-*Nhe*I, se rellena con el fragmento Klenow y se liga en 3' del IRES, en el *Sma*I de pIRES-3, creando pIRES-4. Se realiza la clonación en la secuencia intermedia para evitar la posibilidad de que la adición de una secuencia dé como resultado aumento de la inestabilidad del ARN. Esto también daría como resultado una señal de ARN inferior en una transferencia, pero esto no tendría nada que ver con el bloqueo de la transcripción. Colocar la secuencia que va a someterse a prueba en secuencias intermedias da como resultado la transcripción de esta secuencia para dar ARN, pero se corta y empalma posteriormente, de modo que se forma un ARNm de *codA* funcional. Esto sucede independientemente de si había una secuencia adicional dentro de la secuencia intermedia o no. Por tanto, cualquier disminución en la señal de ARNm no se debe a pérdida de estabilidad del ARN, sino a una consecuencia directa de la terminación de la transcripción debida a la secuencia de TRAP.

A continuación, se corta pIRES-4 con *Xho*I-*Xba*I, se rellena con el fragmento Klenow y se liga en pBSKS cortado con *Sma*I, creando pIRES-5. Se corta pIRES 5 con *Sal*I, y se liga en pPURO parcialmente digerido con *Sal*I, creando pIRES-6. Por tanto el casete completo consiste en el promotor de CMV, la IV, IRES, *codA::upp* y la señal de poliadenilación de SV40 en una estructura principal de pREP4 (Invitrogen). En esta estructura principal está presente un gen de resistencia a higromicina, para permitir la selección de transformantes sobre higromicina.

Para crear pIRES-31, se sustituyen el IRES y *codA::upp* en pIRES-6 por sólo *codA::upp*. Se aparean los oligos *Not*I-*Bcl*I-EV (GGCCGCTGATCAGATATCGCGG) y *Nhe*I-*Eco*RV-*Bcl*I (CTAGCCGCGATATCTGATCAGC) y se ligan a pIRES 6 digerido con *Not*I-*Nhe*I, que libera el IRES y *codA::upp*. Esto crea pIRES-30. Entonces se liga el ORF de *CodA::upp* como un fragmento de *Bam*HI en pIRES-30 cortado con *Bcl*I, creando pIRES-31 (fig. 3).

Transfección y cultivo de células U2-OS

Transfección y cultivo de células U-2 OS con los plásmidos pIRES-6 y pIRES-31: Se cultiva la línea celular U-2 OS de osteosarcoma humano (n.º de la ATCC HTB-96) en medio de Eagle modificado por Dulbecco + suero de ternero fetal al 10% que contiene glutamina, penicilina y estreptomycin (citado anteriormente) a 37°C/CO₂ al 5%. Se transfectan las células con el vector pIRES-6 y -31 que contiene supuestas TRAP en MCS1 usando SuperFect. Se completa la selección con higromicina en 2 semanas, tiempo tras el cual se aísla un conjunto de clones U-2 OS resistentes a higromicina y se aísla el ARN usando protocolos convencionales (Sambrook *et al* 1989).

40 Resultados

Se transfectan varios constructos en células U-2 OS

- 1) El vector de control vacío tal como se muestra en la figura 3.
- 2) Un ADN de 2400 pb de longitud de fago λ (pb 35711-38103) en orientación 5'-3'
- 3) Un ADN de 2400 pb de longitud de fago λ (pb 35711-38103) en orientación 3'-5'
- 45 4) La secuencia de ADN de MAZ de 60 pb de longitud
- 5) STAR7
- 6) STAR 40
- 7) El vector de control vacío tal como se muestra en la figura 3
- 8) Una secuencia de políA sintética (SPA) de 50 pb de longitud en orientación 5'-3'

- 9) Una combinación de la secuencia de SPA de 50 pb de longitud y un señal de pausa del gen de globina $\alpha 2$ de 92 pb de longitud en orientación 5'-3'
- 10) Una secuencia de poli sintética de 50 pb de longitud en orientación 3'-5'
- 11) Una combinación de la secuencia de SPA de 50 pb de longitud y una señal de pausa del gen de globina $\alpha 2$ de 92 pb de longitud en orientación 3'-5'

Tras la transfección, se realiza la selección mediante higromicina. Tras tres semanas, se recoge todo el conjunto de células y se aísla el ARNm. Se usa todo el conjunto de células y no colonias individuales puesto que el vector de la figura 3 se replica de manera episomal, lo que previene los efectos de posición que podrían producirse cuando los vectores se integran de manera estable.

- 10 Tal como se muestra en la figura 4, el fragmento de lambda (pb 35711-37230) (carril 2) bloquea eficazmente la transcripción del gen *codA* impulsada por el promotor de CMV, en comparación con el control vacío (carriles 1 y 7). También una secuencia de poliA sintética (SPA) de 50 pb de longitud (Levitt *et al* 1989) (AATAAAAGATCCTTATTTTAC TAGTTCTGTGTGTTGGTTTTTGTGTG) o bien sola (carril 8) o bien en combinación con una señal de pausa de 92 pb del gen de globina $\alpha 2$ humana (ACATACGCTCTCCATCAAACAAAACGAAACAAAACAACTAGCAAATAGGCTGTCCCCAGTGCAAGTGCAG GTGCCAGAACATTTCTCT) (Enriquez-Harris *et al* 1991) (carril 9) bloquea de manera potente la transcripción del gen *codA* impulsada por el promotor de CMV. Se usa el gen de resistencia a higromicina como control interno, indicando el número de copias de los plásmidos. Se ha notificado que la secuencia de MAZ de 60 pb de longitud es un poderoso bloqueante de la transcripción (Ashfield *et al* 1994). Sin embargo, en el sistema de prueba, la secuencia de MAZ (carril 4) no previene la transcripción. Además, STAR 7 (carril 5) y STAR 40 (carril 6) no previenen la transcripción del gen *codA* impulsada por el promotor de CMV. La cuantificación de las señales usando un instrumento de obtención de imágenes de fósforo mostró que el fago lambda (pb 35711-38103), la secuencia de SPA y el 95% del bloque de combinación SPA/Pausa del promotor de CMV impulsaban la transcripción. Se concluye que el fragmento de fago lambda (pb 35711-38103) y las secuencias de SPA, SPA/Pausa (tabla 1) sirven como TRAP.

Ejemplo 2: Las secuencias de TRAP bloquean la transcripción de una manera direccional

Materiales y métodos

Se hace referencia a los experimentos del ejemplo 1.

Resultados

- 30 Tal como se muestra en la figura 4, la orientación de la TRAP es un parámetro esencial en la acción de las secuencias de TRAP. El fago lambda (pb 35711-38103) sirve sólo como TRAP en una orientación 5'-3' (carril 2). Cuando se somete a prueba en la contraria, la orientación 3'-5' (carril 3), no se encuentra ningún bloqueo de la transcripción impulsada por CMV. De manera similar, la SPA y la secuencia combinada SPA/Pausa bloquean la transcripción sólo en la orientación 5'-3' (carriles 8 y 9) y no en la orientación 3'-5' (carriles 10 y 11). La dependencia de la orientación de las secuencias de TRAP es de importancia para la orientación en la que pueden usarse cuando flaquean transgenes.

Ejemplo 3: Las TRAP mejoran los efectos de elementos STAR sobre el nivel de expresión de transgenes

Un objeto de esta invención es mejorar la expresión transgénica para la producción de proteínas heterólogas, aumentando así el rendimiento de la proteína heteróloga.

Materiales y métodos

Plásmidos

- Se describe a continuación la construcción del vector pPlug&Play-d2EGFP-ires-Zeo (PP). Se modifica el plásmido pd2EGFP (Clontech 6010-1) mediante inserción de un conector en el sitio *Bs*WI para producir pd2EGFP-link. El conector (preparado apareando los oligonucleótidos GTACGGATATCAGATCTTTAATTAAG y GTACCTTAATT AAAGATCTGATAT) introduce sitios para las endonucleasas de restricción *PacI*, *BglII* y *EcoRV*. Esto crea el sitio de clonación múltiple MCSII para la inserción de elementos STAR. Entonces se usan cebadores (GATCAGATCTGGCGGCCATTTAAATCGTC TCGCGGTTTCGGTGATGACGG) y (AGGCGGATCCGAATGTATTTAGA AAAATAACAATAGGGG) para amplificar una región de 0,37 kb de pd2EGFP, que se inserta en el sitio *BglII* de pIRES (Clontech 6028-1) para producir pIRES-stuf. Esto introduce sitios para las endonucleasas de restricción *Ascl* y *Swal* en MCSI, y actúa como "fragmento de relleno" para evitar la posible interferencia entre elementos STAR y promotores adyacentes. Se digiere pIRES-stuf con *BglII* y *FspI* para liberar un fragmento de ADN compuesto por el fragmento de relleno, el promotor de CMV, el elemento IRES (flanqueado por sitios de clonación múltiple MCS A y MCS B) y la señal de poliadenilación de SV40. Se liga este fragmento con la estructura principal de vector de pd2EGFP-link producido mediante digestión con *Bam*HI y *Stu*I,

para producir pd2IRES-link.

Se insertan los marcos de lectura abiertos de los genes de resistencia a zeocina en los sitios *Bam*HI/*Not*I de MCS B en pd2IRES-link tal como sigue: se amplifica el ORF de resistencia a zeocina mediante PCR con cebadores (GATCGGATCCTTC GAAATGGCCAAGTTGACCAGTGC) y (AGGCGCGGCCGCAATTCTCAG TCCTGCTCCTC) del plásmido pEM7/zeo, se digiere con *Bam*HI y *Not*I y se liga con pd2IRES-link digerido con *Bam*HI/*Not*I para dar pd2IRES-link-zeo.

Se introduce el ORF indicador de SEAP en pd2IRES-link-zeo mediante amplificación por PCR de pSEAP2-basic con cebadores (GATCGAATTCTCGCGACTTCG CCCACCATGC) y (AGGCGAATTCACCGGTGTTAAACTCATGTCTGCTC GAAGCGGCCGG), e inserción del casete de SEAP digerido con *Eco*RI en los sitios *Eco*RI en MCS A de los plásmidos pd2IRES-link-zeo (para producir el plásmido PP2). Se corta PP2 con *Eco*RI y *Mlu*I para eliminar el gen SEAP y se introduce p2EGFP con cebadores (GATCGAATTCATGGTGTAGCAAGGGCGAGGAG) y (AGGCACGCGTGTTAACCTACACATTGATCCTAGCAGAAGC).

Se clonan fragmentos de STAR de *Ascl* en el sitio *Ascl* de MCS I de ppp2EGFP. Se amplifica un fragmento de ADN de lambda de 2,4 kb (TRAP) usando cebadores (GATCATTTAAATGT CGACCTGAATTGCTATGTTTAGTGAGTTG) y (GATCGTTCGACGTTTGG CTGATCGGC), y se clona como un fragmento de *Sal*I en MCS I, en 5' con respecto al STAR. Entonces se amplifican STAR y TRAP usando cebadores (GATCTTAATTAACCAAGCTTGCATGCCTGCAG) y (AGGCGATATCGCG CGAGACGATTTAAATGG), se cortan con *Eco*RV y *Pac*I, y se ligan en el mismo vector, cortado con *Eco*RV y *Pac*I, a partir del cual se amplificaron.

Transfección y cultivo de células CHO

Se cultiva la línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) en medio HAMS-F12 + suero de ternero fetal al 10% que contiene glutamina 2 mM, penicilina 100 U/ml y estreptomocina 100 microgramos/ml a 37°C/CO₂ al 5%. Se transfectan las células con los plásmidos indicados usando SuperFect (QIAGEN) tal como se describe por el fabricante. En resumen, se siembran las células en recipientes de cultivo y se cultivan durante la noche hasta una confluencia del 70-90%. Se combina el reactivo SuperFect con el ADN de plásmido a una razón de 6 microlitros por microgramo (por ejemplo para una placa de Petri de 10 cm, 20 microgramos de ADN y 120 microlitros de SuperFect) y se añade a las células. Tras otra incubación durante la noche se añade zeocina a una concentración de 50 µg/ml y se cultivan las células adicionalmente. Tras otros tres días, se sustituye el medio por medio nuevo que contiene zeocina (100 µg/ml) y se cultiva adicionalmente. Cuando se hacen visibles colonias individuales (aproximadamente diez días tras la transfección), se elimina el medio y se sustituye por medio nuevo sin zeocina. Se aíslan clones individuales y se transfieren a placas de 24 pocillos en medio sin zeocina. Un día tras el aislamiento de las colonias se añade zeocina al medio. Se evalúa la expresión del gen indicador GFP aproximadamente 3 semanas tras la transfección.

Los constructos sometidos a prueba consisten básicamente en un gen bicistrónico con el gen GFP, un IRES y el gen de resistencia a zeocina bajo el control del promotor de CMV y un gen monocistrónico que codifica para el gen de resistencia a puromicina bajo el control del promotor de SV40 (figura 2A). Se crea diversidad en los constructos mediante la adición de la TRAP de lambda de 2400 pb (pb 35711-38103) (tabla 1) a los extremos 5' y 3', STAR 40 (tabla 2) a los extremos 5' y 3' o la combinación de STAR 40 y la TRAP de lambda (pb 35711-38103) a los extremos 5' y 3'. Se transfectan los constructos en células CHO-K1. Se expanden colonias estables antes de que se determine la señal de GFP en un citómetro de flujo XL-MCL Beckman Coulter. Se toma la media de la señal de GFP como medida para el nivel de la expresión de GFP y esto se representa en la figura 5.

Resultados

La fig 5 muestra que flanquear todo el constructo GFP-IRES-Zeo (figura 2A) con la TRAP de lambda (pb 35711-38103) en la orientación 5'-3' (véase la figura 4) no da como resultado colonias de CHO estables que expresan niveles significativamente superiores de la proteína GFP, en comparación con el control "vacío" sin las secuencias de TRAP (control). Sin embargo, flanquear todo el casete con la TRAP de lambda (pb 35711-38103) (orientación 5'-3') y STAR 40 combinados da como resultado señales de GFP significativamente superiores (aproximadamente del 200%) en comparación con las señales de GFP más altas que se obtienen con un constructo que está flanqueado por elementos STAR 40 solos. Por tanto, se concluye que la TRAP de lambda (pb 35711-38103) potencia la capacidad de los elementos STAR para transmitir niveles de expresión superiores a un transgén.

Ejemplo 4: La influencia de una TRAP sobre el rendimiento de proteína depende de la orientación.

Materiales y métodos

Se hace referencia a los experimentos del ejemplo 3.

Resultados

Tal como se muestra en la figura 6, la orientación de la TRAP es un parámetro esencial en la acción de las

5 secuencias de TRAP. La TRAP de fago lambda (pb 35711-37230) sirve sólo como TRAP en la orientación 5'-3' (figura 3, carril 2). Cuando se somete a prueba en la orientación 3'-5', que no transmite bloqueo de la transcripción (figura 3, carril 3), tampoco se observa efecto sobre los niveles de expresión de la proteína GFP. No se observa ningún efecto de la secuencia de lambda por sí misma y ningún efecto cuando se combina con STAR 40 (figura 6). La dependencia de la orientación de las secuencias de TRAP es de importancia para la orientación en la que van a usarse cuando flanquean transgenes.

Ejemplo 5: La estabilidad de la expresión transgénica se mejora mediante las TRAP

10 Durante el cultivo de células hospedadoras recombinantes, es una práctica común mantener la selección con antibióticos. Esto pretende prevenir el silenciamiento transcripcional del transgén, o la pérdida del transgén del genoma mediante procesos tales como recombinación. Sin embargo, no se desea para la producción de proteínas, por una serie de motivos. En primer lugar, los antibióticos que se usan son bastante caros, y contribuyen significativamente al coste unitario del producto. En segundo lugar, para uso biofarmacéutico, la proteína debe ser pura de manera demostrable, sin trazas del antibiótico en el producto. Una ventaja de STAR y TRAP para la producción de proteínas heterólogas es que confieren expresión estable en los transgenes durante el cultivo prolongado, incluso en ausencia de selección con antibióticos; esta propiedad se demuestra en este ejemplo.

Materiales y métodos

Se miden los niveles de expresión de GFP en las colonias que se describen en el ejemplo 3 tras periodos de una semana. Tras las tres semanas iniciales tras la transfección cuando se realizan las primeras mediciones de GFP, se cultivan las colonias en medio sin zeocina u otros antibióticos. Esto continuó durante el resto del experimento.

20 Resultados

La figura 7 muestra los datos sobre la expresión de GFP de colonias que se transfectaron de manera estable con el constructo de GFP que está flanqueado por la TRAP de lambda (pb 35711-38103) en la orientación 5'-3' y STAR 40 combinados. Se escogen las colonias con los niveles de expresión de GFP más altos en la figura 5 para análisis de la estabilidad de la expresión a lo largo del tiempo en ausencia de presión de selección mediante antibióticos. La expresión de la proteína GFP indicadora permanece estable en las células CHO en los tres puntos de tiempo. El primer punto de tiempo representa el inicio del experimento cuando la presión de selección se elimina. Se realizan mediciones tras una, dos y tres semanas, lo que significa aproximadamente 10, 20 y 30 ciclos celulares respectivamente. Las colonias que contienen la TRAP de lambda y STAR 40 combinados son estables en ausencia de antibióticos. Esto demuestra que la aplicación de la capacidad de una combinación de elementos de TRAP y STAR protege a los transgenes frente al silenciamiento durante el cultivo prolongado. También demuestra que esta propiedad es independiente de la selección con antibióticos.

Ejemplo 6: Las secuencias de TRAP bloquean la transcripción en el contexto de elementos STAR

Materiales y métodos

El cultivo y la determinación de los niveles de expresión de ARN son como en el ejemplo 1.

35 Resultados

Se colocan los siguientes insertos en los constructos de codA:

- 1) El vector de control vacío tal como se muestra en la figura 3.
- 2) STAR 7
- 3) La secuencia de SPA/pausa en la orientación 5'-3'
- 40 4) La secuencia de SPA/pausa en la orientación 3'-5'
- 5) La secuencia de SPA/pausa en la orientación 5'-3' y colocada en 5' de STAR7
- 6) La secuencia de SPA/pausa en la orientación 3'-5' y colocada en 5' de STAR7
- 7) La secuencia de SPA/pausa en la orientación 5'-3' y colocada en 3' de STAR7
- 8) La secuencia de SPA/pausa en la orientación 3'-5' y colocada en 3' de STAR7

45 Los ocho constructos se transfectan posteriormente en células CHO, se recogen conjuntos de colonias transfectadas de manera estable y se miden los niveles de ARNm de codA mediante análisis de transferencia de tipo Northern. Se usa el gen de resistencia a higromicina como control interno, indicando el número de copias de los plásmidos. Se compara la señal de codA con la señal del vector vacío (carril 1).

Tal como se muestra en la figura 8, la orientación de la TRAP de SPA/pausa es un parámetro esencial en la acción

de secuencias de TRAP. La SPA/pausa sólo sirve como TRAP en la orientación 5'-3' (carril 3), y no en la orientación 3'-5' (carril 4). STAR7 no sirve como TRAP (carril 2). Cuando se coloca en 5' de STAR7, la secuencia de SPA/pausa todavía sirve como TRAP en la orientación 5'-3' (carril 5), pero no en la orientación 3'-5' (carril 6). Cuando se coloca en 3' de STAR7, la secuencia de SPA/pausa también sirve como TRAP en la orientación 5'-3' (carril 7), pero no en la orientación 3'-5' (carril 8). Se concluye que la secuencia de SPA/pausa todavía funciona como TRAP en el contexto de un elemento STAR. Esto es importante puesto que la secuencia de TRAP se usa en el contexto de un elemento STAR cuando flanquea un transgén. También se concluye que la dependencia de la orientación de la TRAP sigue siendo igual en el contexto del elemento STAR: sólo la orientación 5'-3' proporciona función de TRAP. La dependencia de la orientación de las secuencias de TRAP es de importancia para la orientación en la que pueden usarse cuando flanquean transgenes. Finalmente se concluye que colocar la secuencia de SPA/pausa en 5' o en 3' del elemento STAR no influye en su función de TRAP. Sólo importa la orientación de la propia secuencia de TRAP.

Tabla 1. Secuencias de TRAP que se usan para las pruebas en los ejemplos descritos.

>fragmento de lambda 35711-38103

5'—AGATCTGAATTGCTATGTTTAGTGAGTTGTATCTATTTATTTTCAATAAAATACAATTGGTTAT
 GTGTTTTGGGGCGATCGTGAGGCAAAGAAAACCCGGCGCTGAGGCCGGTTATTCTTGTTCTCTGGTCAA
 ATTATATAGTTGGAAAACAAGGATGCATATATGAATGAACGATGCAGAGGCAATGCCGATGGCGATAGTG
 GGTATCATGTAGCCGCTTATGCTGGAAAAGAAGCAATAACCCGCAGAAAAACAAAGCTCCAAGCTCAACAA
 AACTAAGGGCATAGACAATAACTACCGATGTCATATACCCATACTCTCTAATCTTGGCCAGTCGGCGCGTT
 CTGCTTCCGATTAGAAACGTCAAGGCAGCAATCAGGATTGCAATCATGGTTCCTGCATATGATGACAATGT
 CGCCCCAAGACCATCTCTATGAGCTGAAAAAGAAACACCAGGAATGTAGTGCGGAAAAGGAGATAGCAA
 ATGCTTACGATAACGTAAGGAATTACTATGTAAACACCAGGCATGATTCTGTTCCGCATAAATTACTCCT
 GATAATTAATCCTTAACTTTGCCACCTGCCTTTTAAAAACATTCCAGTATATCACTTTTTCATTCTTGGGTAGC
 AATATGCCATCTCTCAGTATCTCAGCATTGGTGACCTTGTTCAGAGGCGCTGAGAGATGGCCTTTTCTG
 ATAGATAATGTTCTGTTAAAATATCTCCGGCCTCATCTTTTGGCCGAGGCTAATGTCTGAAAATTGAGGTG
 ACGGGTAAAAATAATATCCTTGGCAACCTTTTTTATAATCCCTTTTAAATTTTGGCTTAATGACTATATCCAA
 TGAGTCAAAAAGCTCCCTTCAATATCTGTTGCCCTAAGACCTTTAATATATCGCCAAATACAGGTAGCTT
 GGCTTCTACCTTACCGTTGTTCCGGCCGATGAAATGCATATGCATAACATCGTCTTTGGTGGTTCCCTCAT
 CAGTGGCTCTATCTGAACGCGCTCTCCACTGCTTAATGACATTCCTTTCCCGATTAAAAAATCTGTCAAGTC
 GGATGTGGTCCGCCCCGAAAACAGTTCTGGCAAAACCAATGGTGTCCGCTTCAACAAACAAAAAGATGGG
 AATCCCAATGATTTCGTCATCTGCGAGGCTGTTCTTAATATCTTCAACTGAAGCTTTAGAGCGATTTATCTTC
 TGAACCAAGACTCTTGTCATTTGTTTTGGTAAAGAGAAAAGTTTTCCATCGATTTTATGAATATACAAATAA
 TTGGAGCCAACCTGCAGGTGATGATTATCAGCCAGCAGAGAATTAAGGAAAACAGACAGGTTTATTGAGC
 GCTTATCTTTCCCTTTATTTTTGCTGCGGTAAGTCGCATAAAAACCATTTCTTCAATTCATTTACTA
 TGTATGTTCTGAGGGGAGTAAAAATCCCTAATTTCGATGAAGATTCTTGCTCAATTGTTATCAGCTATGC
 GCCGACCAGAACCTTGGCGATCAGCCAAACGTCCTTTCAGGCCACTGACTAGCGATAACTTTCCCCACA
 ACGGAACAACCTCTCATTGCATGGGATCATTGGGTACTGTGGGTTTGTGTTGTA AAAACACCTGACCGCT
 ATCCCTGATCAGTTTCTTGAAGGTA AACTCATCCCCCAAGTCTGGCTATGCAGAAATCACCTGGTCAAC
 AGCCTGCTCAGGTCAACGAGAATTAACATTCCTCAGGAAAGCTTGGCTTGGAGCCTGTTGGTGGGTC
 TGGAATTACCTTCAACCTCAAGCCAGAATGCAGAACTACTGGCTTTTTTGGTTGTGCTTACCCATCTCTCCG
 CATCACCTTTGGTAAAGGTTCTAAGCTCAGGTGAGAACATCCCTGCCTGAACATGAGAAAAAACAGGGTAC
 TCATACTCACTTCAAGTGACGGCTGCATACTAACCCTTTCATACATCTCGTAGATTTCTCTGGCGATTGAA
 GGGCTAAATTTCAACGCTAACTTTGAGAATTTTTGCAAGCAATGCGGCGTTATAAGCATTTAATGCATTG
 ATGCCATTAATAAAGCACCAACGCCTGACTGCCCCATCCCCATCTTGTCTGCGACAGATTTCTGGGATAA
 GCCAAGTTCAATTTCTTTTTTTTATAAATTTGCTTTAAGGCGACGTGCGTCTCAAGCTGCTCTTGTGTTAAT
 GGTTTCTTTTTGTGCTACAGTTAAATCTATCACCCGAAGGATAAATATCTAACCCGTGCGTGTGAC
 TATTTTACCTCTGGCGGTGATAATGGTTGCATGTAATAAGGAGGTTGTATGGAACAACGCATAACCCTGAA
 AGATTATGCAATGCGCTTTGGGCAACCAAGACAGCTAAAGATCT—3'

>fragmento de lambda 22425-27972

5'-

CTGCAGATCTGGAAATTGCAACGAAGGAAGAAACCTCGTTGCTGGAAGCCTGGAAGAAGTATCGGGTGTT
GCTGAACCGTGTTGATACATCAACTGCACCTGATATTGAGTGGCCTGCTGTCCCTGTTATGGAGTAATCGTT
TTGTGATATGCCGCAGAAACGTTGTATGAAATAACGTTCTGCGGTTAGTTAGTATATTGTAAAGCTGAGTAT
TGGTTTATTTGGCGATTATTATCTTCAGGAGAATAATGGAAGTTCTATGACTCAATTGTTTCATAGTGTTTAC
ATCACCGCCAATTGCTTTTAAGACTGAACGCATGAAATATGGTTTTTTCGTCATGTTTTGAGTCTGCTGTTGA
TATTTCTAAAGTCGGTTTTTTTTCTTCGTTTTCTCTAACTATTTTCCATGAAATACATTTTTGATTATTATTTG
AATCAATTCCAATTACCTGAAGTCTTTCATCTATAATTGGCATTGTATGTATTGGTTTTATTGGAGTAGATGC
TTGCTTTTCTGAGCCATAGCTCTGATATCCAAATGAAGCCATAGGCATTTGTTATTTTGGCTCTGTCAGCTG
CATAACGCCAAAAAATATATTTATCTGCTTGATCTTCAAATGTTGTATTGATTAAATCAATTGGATGGAATT
GTTTATCATAAAAAATTAATGTTTGAATGTGATAACCGTCCTTTAAAAAAGTCGTTTCTGCAAGCTTGGCTG
TATAGTCAACTAACTCTTCTGTGCGAAGTGATATTTTTAGGCTTATCTACCAGTTTTAGACGCTCTTTAATATC
TTCAGGAATTATTTTATTGTCATATTGTATCATGCTAAATGACAATTTGCTTATGGAGTAATCTTTTAATTTT
AAATAAGTTATTCTCCTGGCTTCATCAAATAAAGAGTCGAATGATGTTGGCGAAATCACATCGTCACCCAT
TGGATTGTTTATTTGTATGCCAAGAGAGTTACAGCAGTTATACATTCTGCCATAGATTATAGCTAAGGCATG
TAATAATTCGTAATCTTTTAGCGTATTAGCGACCCATCGTCTTTCTGATTTAATAATAGATGATTCAGTTAA

ATATGAAGGTAATTTCTTTTGTGCAAGTCTGACTAACTTTTTTATACCAATGTTTAAACATACTTTTCATTTGTA
 ATAAACTCAATGTCATTTTCTTCAATGTAAGATGAAATAAGAGTAGCCTTTGCCTCGCTATACATTTCTAAA
 TCGCCTTGTTTTTCTATCGTATTGCGAGAATTTTLAGCCCAAGCCATTAATGGATCATTTTTCCATTTTTCAA
 TAACATTATTGTTATACCAAATGTCATATCCTATAATCTGGTTTTTGTTTTTTTGAATAATAAATGTTACTGT
 TCTTGCGTTTTGGAGGAATTGATTCAAATTCAAGCGAAATAATTCAGGGTCAAAATATGTATCAATGCAGC
 ATTTGAGCAAGTGCATAAAATCTTAAAGTCTTCTTTCCCATGGTTTTTLAGTCATAAAACTCTCCATTTTGAT
 AGGTTGCATGCTAGATGCTGATATATTTTAGAGGTGATAAAATTAAGTCTTAACTGTCAATGTAATACAA
 GTTGTGTTGATCTTTGCAATGATTCTTATCAGAAACCATATAGTAAATTAGTTACACAGGAAATTTTTAATAT
 TATTATTATCATTATTATGTATTAATAATTAGAGTTGTGGCTTGGCTCTGCTAACACGTTGCTCATAGGAGA
 TATGGTAGAGCCGCAGACACGTCGTATGCAGGAACGTGCTGCGGCTGGCTGGTGAACCTCCGATAGTGCGG
 GTGTTGAATGATTTCCAGTTGCTACCGATTTTACATATTTTTTGTCATGAGAGAATTTGTACCACCTCCCACC
 GACCATCTATGACTGTACGCCACTGTCCCTAGGACTGCTATGTGCCGGAGCGGACATTACAAACGTCCTTCT
 CGGTGCATGCCACTGTTGCCAATGACCTGCCTAGGAATTGGTTAGCAAGTTACTACCGGATTTTGTAAAAA
 CAGCCCTCTCATATAAAAAAGTATTCGTTCACTTCCGATAAGCGTCGTAATTTTCTATCTTTTCATCATATTCT
 AGATCCCTCTGAAAAAATCTTCCGAGTTTGTAGGCACTGATACATAACTCTTTTCCAATAATTGGGGAAGT
 CATTCAAATCTATAATAGGTTTCAGATTTGCTTCAATAAATTCTGACTGTAGCTGCTGAAACGTTGCGGTTG
 AACTATATTTCTTATAACTTTTACGAAAGAGTTTCTTTGAGTAATCACTTCACTCAAGTGCCTCCCTGCCTC
 CAAACGATACCTGTTAGCAATATTTAATAGCTTGAAATGATGAAGAGCTCTGTGTTTGTCTTCTGCCTCCA
 GTTCGCCGGGCATTCAACATAAAAACTGATAGCACCCGGAGTTCCGGAAACGAAATTTGCATATACCCATT
 GCTCACGAAAAAAAATGTCCTTGTTCGATATAGGGATGAATCGCTTGGTGTACCTCATCTACTGCGAAAACT
 TGACCTTTCTCTCCCATATTCAGTCGCGGCACGATGGAACATAATTAATAGGCATCACCGAAAAATTCAGG
 ATAATGTGCAATAGGAAGAAAATGATCTATATTTTTTGTCTGTCTATATCACCCAAAAATGGACATTTTTTC
 ACCTGATGAAACAAGCATGTCATCGTAATATGTTCTAGCGGGTTTGTTTTTATCTCGGAGATTATTTTCATA
 AAGCTTTTCTAATTTAACCTTTGTTCAGGTTACCAACTACTAAGGTTGTAGGCTCAAGAGGGTGTGTCTGTCT
 GTAGGTAAATAACTGACCTGTGAGCTTAATATTCTATATTGTTGTTCTTTCTGCAAAAAAGTGGGGAAGTG
 AGTAATGAAATTATTTCTAACATTTATCTGCATCATACTTCCGAGCATTATTAAGCATTTCGCTATAAGTT
 CTCGCTGGAAGAGGTAGTTTTTTTCATTGTACTTTACCTTCATCTCTGTTCAATTATCATCGCTTTTAAACGGT
 TCGACCTTCTAATCCTATCTGACCATTATAATTTTTTAGAATGGTTTCATAAGAAAGCTCTGAATCAACGGA
 CTGCGATAATAAGTGGTGGTATCCAGAATTTGTCACTTCAAGTAAAAACACCTCACGAGTTAAAAACACCTA
 AGTTCTCACCGAATGTCTCAATATCCGGACGGATAATTTTATTGCTTCTCTTGACCGTAGGACTTTCCACA
 TGCAGGATTTTGGAACTCTTGCAGTACTACTGGGGAATGAGTTGCAATTTATTGCTACACCATTGCGTGCAT
 CGAGTAAGTCGCTTAATGTTTCGTAAAAAAGCAGAGAGCAAAGGTGGATGCAGATGAACCTCTGGTTTCATC
 GAATAAACTAATGACTTTTCCCAACGACATCTACTAATCTTGTGATAGTAAATAAAACAATTGCATGTC
 CAGAGCTCATTGCAAGCAGATATTTCTGGATATTGTCATAAAACAATTTAGTGAATTTATCATCGTCCACTT
 GAATCTGTGGTTCATTACGTCTTAACTCTTCATATTTAGAAATGAGGCTGATGAGTTCATATTTGAAAAGT
 TTTTCATCACTACTTAGTTTTTTGATAGCTTCAAGCCAGAGTTGTCTTTTTTCTATCTACTCTCATAACAACCAAT
 AAATGCTGAAATGAATTTCTAAGCGGAGATCGCCTAGTGATTTTAAACTATTGCTGGCAGCATTCTTGAGTC
 CAATATAAAAGTATTGTGTACCTTTTGTCTGGGTGAGGTTGTCTTTAGGAGGAGTAAAAGGATCAAATGCA
 CTAACGAAACTGAAACAAGCGATCGAAAATATCCCTTTGGGATTCTTGACTCGATAAGTCTATTATTTTCA
 GAGAAAAAATATTCATTGTTTTCTGGGTTGGTGATTGCACCAATCATTCCATTCAAAATTGTTGTTTTACCA
 CACCCATTCCGCCCGATAAAAAGCATGAATGTTTCGTGCTGGGCATAGAATTAACCGTCACCTCAAAAGGTAT
 AGTTAAATCACTGAATCCGGGAGCACTTTTTCTATTAATGAAAAGTGGAATCTGACAATTCTGGCAAAC
 CATTAAACACACGTGCGAACTGTCCATGAATTTCTGAAAGAGTTACCCCTCTAAGTAATGAGGTGTTAAGG

ACGCTTTCATTTTCAATGTGCGGCTAATCGATTTGGCCATACTACTAAATCCTGAATAGCTTTAAGAAGGTTA
 TGTTTAAAACCATCGCTTAATTTGCTGAGATTAACATAGTAGTCAATGCTTTCACCTAAGGAAAAAACATT
 TCAGGGAGTTGACTGAATTTTTTATCTATTAATGAATAAGTGCTTACTTCTCTTTTTGACCTACAAAACCA
 ATTTTAAACATTTCCGATATCGCATTTTTTACCATGCTCATCAAAGACAGTAAGATAAAAACATTGTAACAAAG
 GAATAGTCATTCCAACCATCTGCTCGTAGGAATGCCTTATTTTTTCTACTGCAGGAATATACCCGCCTCTTT
 CAATAACACTAAACTCCAACATATAGTAACCTTAATTTTATTAATAAACCAGCAATTTATTTGGCGGCAAC
 ACAGGATCTCTTTTTAAGTTACTCTCTATTACATACGTTTTCCATCTAAAAATTAGTAGTATTGAACTTAAC
 GGGCATCGTATTGTAGTTTTCCATATTTAGCTTTCTGCTTCTTTTTGGATAACCCACTGTTATTCATGTTGC
 ATGGTGCCTGTTTATACCAACGATATAGTCTATTAATGCATATATAGTATCGCCGAACGATTAGCTCTTCA
 GGCTTCTGAAGAAGCGTTTCAAGTACTAATAAGCCGATAGATAGCCACGGACTTCGTAGCCATTTTTTATA
 AGTGTAACTTCCGCTCCTCGCTCATAACAGACATTCACTACAGTTATGGCGGAAAGGTATGCATGCTGGG
 TGTGGGGAAGTCGTGAAAAGAAAAGAACGCTGCGTCTTTGACATCACTGCTATCTTCTTACTGGTTAT
 GCAGGTCGTAGTGGGTGGCACACAAAAGCTTTCAGTGGATTGCGAGGCTTTGTGCTTCTCTGGAGTGGAC
 AGTTTTGATGACAAAAAATTAGCGCAAGAAGACAAAAATCACCTTGCCTAATGCTCTGTTACAGGTCACT
 AATACCATCTAAGTAGTTGATTCATAGTACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTA
 TGCAAATCTAATTTAATATATTGATATTTATATCATTTTTACGTTTCTCGTTCAGCTTTTTTATACTAAGTTG
 GCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAAATAAAATCATT
 ATTTGATTTCAATTTTGTCCACTCCCTGCCTCTGTCATCACGATACTGTGATGCCATGGTGTCCGACTTATG
 CCCGAGAAGATGTTGAGCAAACCTTATCGCTTATCTGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACT
 CGTGAAAGGTAGGCGGATCC—3'

> Secuencia de poliA sintética (SPA) combinada y señal de pausa del gen de globina $\alpha 2$ humana

5'--ATAAAAGATCCTTAAATTTCACTAGTTCTGTGTGTTGGTTTTTTGTGTGAACATACTGCTCTCCATCA
 AAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAA
 CATTCTCT—3'

5 > Inter-histona H3FA-H4F (<http://genome.ucsc.edu/cgi-bin/> gTracks? hgside=13148179&
 position=chr6%3A26063) (cromosoma 6, pb 26063887-26064766)

5'—TATTTGAGGACACTAACCTGTGCGCCATCCACGCCAAAGCGCGTCACTATCAATGCCCAAGGACAT
 CCAGCTCGCCCGCCGATCCGCGGAGAGAGGGCGTGATTACTGTGGTCTCTCTGACGGTCCAAGCA
 AAGGCTCTTTTCAGAGCCACCACCTTTTCAAGTAAAGTAGCTGTAAGAAACCAATTTAAGACAAAAGGGAA
 TGCATTGGGAGCACTTTTCGTTTTAATGCTACTGAAGGCTTCAAACCAATCGATTTCCGGCCGGTTCGCGGTG
 ACTCACGCCTGTAATTCAAGCACTTTGAGAGGCTGAGGCGGGCGGATTACCAGAAATCAGGAGTTCGGGAT
 CAGCCTGGCCAACATGGCCGAATCCCGTCTCTACGAAAAATACAAAAACACGCCGGGCGCGACGGCGAGC
 GCTTGTAATCCAGCTACACTCTGAAGGCTGAGGCAGGAGAAACACTTGAACCTGAGAGGCAGAGGTTTC
 AGTGAATCGAGATGGCTCTAATGTACTCCAGTCTGGGCGACAGAGAGATTCCGGTAAAAAAAAGTTTCGA
 CTTAAAAATAATTCTGGAGTCAGAATGGGTTTACATTTAATTTTAAACCCAGTTCCTCAAAGCCTGTAGCTCT
 GTTAAGAAAATAAAGGCCATTGGTCAAGCCTGTGGTCCCACCTCATCTCCCCACCTCCCCAATCGCT
 GCTCCCGCCATTTCTGGGGCTTGGAGGAGGGGTTAAAGGAGCGGACTGTAGGCGTCACATTTCCCGCCTG
 CGCGCTTTTCAGTCTCAGTGTCCGCTGGAGGTGGGGGCAGGGGTAACGTAGATATATAAAGATCGGTTTCC
 TATTCTCTCACTTGTCTTGGTTCACTTCT—3'

> Inter-histona H1F4-H2BFB (cromosoma 6: 26214737-26215909)

5'—AAGGCGCCCAAGAGCCAGCGAAGGCCAAAGCAGTTAAACCCAAGGCGGCTAAACCAAAGAC
 CGCCAAGCCCAAGGCAGCCAAGCCAAAGAAGGCGGCAGCCAAGAAAAAGTAGAAAAGTTCCTTTGGCCAAC
 TGCTTAGAAGCCCAACACAACCCAAAGGCTCTTTTCAGAGCCACCCACCGCTCTCAGTAAAAGAGCTGTTG
 CACTATTAGGGGGCGTGGCTCGGGAAAACGCTGCTAAGCAGGGGGCGGGTCTCCCGGGAACAAAGTCGGGG
 AGAGGAGTGGGATTTTGTGTCTCCGAGCTATTTTACTAAGGCGTCCGCTCGCCCAAGCCGAGGATGC
 AGTGGCGTCATCTCGATTTTGCCTTCTCGAGTGTCCGAGTTGAACCCATTTGGGCTCCCTTGTGCTTTGCA
 CTTTTAGCAGGCCCTGGCCTCCAGATAGCATGGGAAAAAAAATGTTGGGATTTTCCCGGGTTTCTAAGCTG
 GGTTTTTCCGAGTTCCAAACACGGCACAGTGTATCAGTTTCTGTGCTGGTTACAAGCCTACTGGTTATCCCT
 ATCGAGTATGGCAGGCAGTGAAGGACTTCAGAGGAGTACGTCTTAGGACAAGTGGCATAGTACTGACATT
 ATTTCCGAAGGGCTACATTTCAAGTGCTTGGGGAGACTACTGCCACATAACTGAAAATTAGAAACCGACAC

ES 2 373 549 T3

TGCAGAAAATACTTGGTCCTTAAATGTGGCATTGGATGGATTAAGGACTTGCCGAAACGTAAAACCTGAC
AGACTTGGGGGGGGGGGATGTCCCAATTAGCACGGCTTCTGTATGCAACGAGTCCCATACTTTGTTAAAGG
AAGAAAGGAATGTGAGTTCTCCTAATCTGTTAAGTATCTTCGGTGTAAAGTTCTGACACCACAATGTTAAA
AAAGTCGGATCTCAAAAACCAACTGCTCCAAGCGAAGTGCACAGCTGTCTTGCCTAAAGAGGCCTATTTAT
AGTAGCCTCGGGTAGTCTGGTCTGGGCTTCTCATTGGGTACAAGTAAAGGAACGAAATAGCCAATGAAAA
GGTAGACTTTTAAAGTGTGTTTACATTGGCATTGTGACGACACTCTAAAATTAATCCAATCATAAACGAAA
TCTGATTAACCTCATTGAATACCGCATCTATAAATGAACAGGGCC-- 3'

Tabla 2. Los elementos STAR que se usan para las pruebas en los ejemplos descritos.

>STAR4

GATCTGAGTCATGTTTTAAGGGGAGGATTCCTTTGGCTGCTGAGTTGAGATTAGGTTGAGGGTAGTGAAGG
TAAAGGCAGTGAGACCACGTAGGGGTCATTGCAGTAATCCAGGCTGGAGATGATGGTGGTTTCAGTTGGAA
TAGCAGTGCATGTGCTGTAACAACCTCAGCTGGGAAGCAGTATATGTGGCGTTATGACCTCAGCTGGAACA
GCAATGCATGTGGTGGTGAATGACCCACAGTGGGTAGGGTGCATGTGATGGAACAACCTCAGCTGGGTAG
CAGTGTACTTGGATAAAAATGTTGGCATACTCTAGATTTGTTATGAGGGTAGTGCCATTAATTTCTCCACAAA
TTGGTTGTCACGTATGAGTGAAGAGAGGAAGTATGGAAGACTTCAGTGCTTTTGGCCTGAATAAATAGAA
GACGTCATTTCCAGTTAATGGAGACAGGGAAGACTAAAGGTAGGGTGGGATTCAGTAGAGCAGGTGTTCA
GTTTTGAATATGATGAACTCTGAGAGAGGAAAACTTTTCTACCTCTTAGTTTTTGTGACTGGACTTAAGA
ATTAAGTGCATAAAGACAGAGTAACAAGACAAAAATATGCGAGGTTATTTAATATTTTTACTTGCAGAGG
GGAATCTTCAAAAAGAAAAATGAAGACCCAAAGAAGCCATTAGGGTCAAAAGCTCATATGCCTTTTTAAGT
AGAAAATGATAAATTTTAAACAATGTGAGAAGACAAAGGTGTTTGAGCTGAGGCAATAAATTTGTGGACA
GTGATTAAGAAATATATGGGGGAAATGAAATGATAAGTTATTTTTAGTAGATTTATCTTCATACTATTTTTG
GCTTCAACTTCCAGTCTCTAGTGATAAGAATGTTCTTCTCTCCTGGTACAGAGAGAGCACCTTTCTCATGG
GAAATTTTATGACCTTGCTGTAAGTAGAAAGGGGAAGATCTCCTGTTTCCAGCATCAGGATGCAAACATT
TCCCTCCAATCCAGTTCTCAACCCCATGGCTGGGCCTCATGGCATTCCAGCATCGCTATGAGTGCACCTTC
CTGCAGGCTCGCTCGGTAGCTGGTGCAGTCTAGGTCAGTCTATGTGACCAGGAGCTGGGCCTCTGGGCA
ATGCCAGTTGGCAGCCCCCATCCCTCCACTGCTGGGGCCCTCTATCCAGAAGGGCTTGGTGTGCAGAACG
ATGGTGCACCATCATCAATCCCACTTGGCATTTTTAGGGGACAGCCAGCTGCTTTGGCGCGGCAAAAA
ACACCCAACCTCACTCCTCTCAGGGGCTCTGGTCTGATGCCACCACAGGACATCCTTGAGTGTGCTGGGCAG
TCTGAGGACAGGGAAGGAGTATGACCACAAAACAGGAATGGCAGCAGCAGTGCAGGAGGAAGTCAAA
GGCTTGTGTGCTGCCCCGCTGAGGGCTGGCGAGGGCCCTGGGATGGCGCTCAGTGCCTGGTGGCTGC
AAGAGGCCAGCCCTCTGCCATGAGGGGAGCTGGCAGTGACCAAGCTGCACTGCCCTGGTGGTGCATTTCC
TGCCCCACTCTTCTCTTAAGATC

>STAR6

GATCTGACCCACCACAGACATCCCCTCTGGCCTCCTGAGTGGTTTTCTTCAGCACAGCTTCCAGAGCCAAATT
AAACGTTCACTCTATGTCTATAGACAAAAAGGGTTTTGACTAAACTCTGTGTTTTAGAGAGGGAGTTAAAT
GCTGTTAACTTTTTAGGGGTGGGCGAGAGGAATGACAAATAACAACCTGTCTGAATGTTTTACATTTCTCCC
CACTGCCTCAAGAAGGTTCAACACGAGGTCAATCATGATAAGGAGTAAGACCTCCCAGCCGACTGTCCCT
CGGCCCCAGAGGACACTCCACAGAGATATGCTAACTGGACTTGGAGACTGGCTCACACTCCAGAGAAAA
GCATGGAGCACGAGCGCACAGAGCAGGGCCAAGGTCCCAGGGACAGAATGTCTAGGAGGGAGATTGGGG
TGAGGGTAATCTGATGCAATTAAGTGGCAGCTCAACATTAAGGGAGGGGGAAGAAAGAAACAGTCCCT
GTCAAGTAAGTTGTGCAGCAGAGATGGTAAGCTCCAAAATTTGAACTTTGGCTGCTGGAAAGTTTTAGGG
GGCAGAGATAAGAAGACATAAGAGACTTTGAGGGTTACTACACACTAGACGCTCTATGCATTTATTTATT
TATTATCTCTTATTTACTTTGTATAACTCTTATAATAATCTTATGAAAACGGAAACCCTCATATACCCAT
TTTACAGATGAGAAAAGTGACAATTTGAGAGCATAGCTAAGAATAGCTAGTAAGTAAAGGAGCTGGGAC
CTAAACCAACCCCTATCTCACAGAGATACACTCTTTTTTTTCCAGTGTAATTTTTTTTTAATTTTTATT
ACTTTAAGTTCTGGGATACATGTGCAGAAGGTATGGTTTTGTACATAGGTATATGTGTGCCATAGTGGATTG
CTGCACCTATCAACCCGTCATCTAGGTTTAAAGCCCCACATGCATTAGCTATTTGTCTGATGCTCTCCCTCCC
CTCCCCACACCAGACAGGCCTTGGTGTGTGATGTTCCCTCCCTGTGTCCATGTGTTCTCACTGTTCACTCC
CACTTATGAGTGAGAACATGTGGTATTTGGTTTTCTGTTCTGTGTTAGTTTGCTGAGGATGATGGCTTCCA
GTTTATCCATGTCCCTGCAAAGGACACGATC

>STAR7

GATCACCCGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCTCGTCTCTACTAAAAAATA
CGAAAAATTAGCTGGTTGTGGTGGTGCCTGTTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC
ACTTGAATCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATAGTCCATTGCACTCCAGCTGGGCAACAGA
CGGAGACTCTGTCTCCAAAAAATAAATAAATCTTAGAGGACAAGAAATGGCTCTCTCAAACTTTTGAA
GAAAGAATAAATAAATATGCAAGTTCTAGAAGAAGTAATGGGGATATAGGTGCAGCTCATGATGAGGAAG
ACTTAGCTTAACTTTCAATAATGCATCTGTCTGGCCTAAGCCTGGTGGTGGTATGTTTATGTTCTGAAAACATTC
AATATAGAATGATAAATAAATCACTTCTGACCCCTTTTTTTTCTCTCCCTAGACTGTGAAGCAGAAAC
CCCATATTTTTCTTAGGGAAGTGGCTACGCATTTGTATTTATTAACAACCTACCTTATCAGGAAATTCAT
ATTGTTGCCCTTTTATGGATGGGGAACTGGACAAGTGACAGAGCAAAATCCAAACACAGCTGGGGATTTCC
CCTCTTTTAGATGATGATTTTAAAGAATGCTGCCAGAGAGATTCTTGCAAGTGGGAGGACATATATGAC
CTTTAAGATATTTTCCAGCTCAGAGATGCTATGAATGTATCCTGAGTGCATGGATGGACCTCAGTTTTGCG
ATTCTGTAGCTTATAACAATTTGGTGGTTTTCTTTAGAAGAAAAAACAACATTTAATAATTAATAATGAGCC
CAAGACCTTACAAGGCAATCATACAATAAGAGGCTGCTGAAAGTTGAGTTTGTCTCACTTTCTAGTTAATTA
TCTCCTGCTGTTTGTCAATAATGCGTTTGTAGAGGAGCTGCTAATGACAGGTTCTCCAACAGAGTGTGGA
AGAAGGAGATGACGGCTGGCTTCCCCTCTGGGACAGCCTCAGAGCTAGTGGGGAACTATGTTAGCAGAG
TGATGCAGTGACCAAGAAAATAGCACTAGGAGAAAGCTGGTCCATGAGCAGCTGGTGGAGAAAAGGGGTGG
TAATCATGATGCCCTTCTGTTTTATTTTTATTTGGTTTTCTTTTGCCTCTCAATTCCTTCTGACAATACA

AAATGTTGGTTGGAACATGGAGCACCTGGAAGTCTGGTTCATTTTCTCTCAGTCTCTTGATGTTCTCTCGGG
 TTCACTGCCTATTGTTCTCAGTCTACACTTGAGCAATCTCCTCAATAGCTAAAGCTTCCACAATGCAGATT
 TTGTGATGACAAATTCAGCATCACCCAGCAGAACTTAGGTTTTTTCTGTCCTCCGTTTCTGACCTTTTTCT
 TCTGAGTGCTTTATGTCACCTCGTGAACCATCCTTTCCTTAGTCATCTACCTAGCAGTCCCTGATTCTTTTGAC
 TTGTCTCCCTACACCACAATAAATCACTAATTAATGGAATTCAATCCCTAAAATTTGCACAAAACCTTGCAAA
 TAGATTACGGGTTGAAACTTAGAGATTTCAAACCTTGAGAAAAAAGTTTAAATCAAGAAAAATGACCTTTAC
 CTTGAGAGTAGAGGCAATGTCATTTCCAGGAATAATTATAATAATATTGTGTTTAAATTTTGTATGTAACAT
 TTGAATACCTTCAATGTTCTTATTTGTGTTATTTAATCTCTTGATGTTACTAACTCATTGGTAGGGAAGAA
 AACATGCTAAAAATAGGCATGAGTGTCTTATTAATGTTGACAAGTGAATAGATGGCAGAAGGTGGATTTCATA
 TTCAGTTTTCCATCACCTGGAAATCATGCGGAGATGATTTCTGCTTGCAAATAAACTAACCCAATGAGG
 GAAACAGCTGTTCTTAGGTGAAAACAAAACACGCCAAAAACCTTTATTCTCTTTATTATGAATCAAA
 TTTTTCTCTCAGATAATTGTTTTATTTATTTATTTTATTATTATTGTTATTATGTCCAGTCTCACTCTGTGC
 CTAAGCTGGCATGATC

>STAR12

ATCCTGCTTCTGGGAAGAGAGTGGCCTCCCTTGTGCAAGTGACTTTGGCAGGACCAGCAGAAAACCCAGGTT
 TCCTGTCAGGAGGAAGTGCTCAGCTTATCTCTGTGAAGGGTCGTGATAAGGCACGAGGAGGCAGGGGCTTG
 CCAGGATGTTGCCCTTCTGTGCCATATGGGACATCTCAGCTTACGTTGTTAAGAAATATTTGGCAAGAAGAT
 GCACACAGAAATTTCTGTAACGAATAGGATGGAGTTTTAAGGGTACTACGAAAAAAGAAAACACTACTGGA
 GAAGAGGGAAGCCAAACACCACCAAGTTTGAATCGATTTTATTGGACGAATGTCTCACTTTAAATTTAAA
 TGGAGTCCAACCTTCTTTCTCACCCAGACGCGAAGGTGGCATTCAAAATGTTTACACTTGTTCATCT
 GCCTTTTTGCTAAGTCTGGTCCCCTACCTCCTTTCCCTCACTTACATTTGTCGTTTCATCGCACACATATG
 CTCATCTTTATATTTACATATATAAATTTTATATATGGCTTGTGAAATATGCCAGACGAGGGATGAAATA
 GTCCTGAAAACAGCTGGAAAATTATGCAACAGTGGGGAGATTGGGCACATGTACATTCTGTACTGCAAAGT
 TGCACAACAGACCAAGTTTGTATAAGTGAGGCTGGGTGGTTTTTATTTTTCTCTAGGACAACAGCTTGCC
 TGGTGGAGTAGGCCTCCTGCAGAAGGCATTTTCTTAGGAGCCTCAACTCCCCAAGAAGAGGAGAGGGCG
 AGACTGGAGTTGTGCTGGCAGCACAGAGACAAGGGGGCACGGCAGGACTGCAGCCTGCAGAGGGGCTGG
 AGAAGCGGAGGCTGGCACCAGTGGCCAGCGAGGCCAGGTCCAAGTCCAGCGAGGTGAGGCTAGAGT
 ACAGCAAGGCCAAGGTCCAAGGTCAAGTCAAGTCCAAGTCCATGGTCAGTGAGGCTGAGACCCAGGGTCCA
 ATGAGGCCAAGGTCCAGAGTCCAGTAAGGCCGAGATC

>STAR18

CTAAAGGCATTTTATATAGAGCTGTGGTTTTTGTGGTTTACCTGTGGCCGTGGCCAGAGGTTCTCTGGGAGGC
 TAACAGGTGTTTTTGGAGGTTGGGGCTTGGGTGGGGTGGGGTGAATTTCTGTITCTAGGATGTGCTTGG
 TGTGTAATAGCTAGGCTTATGACTGATGCTGGTTAATTTCTAGGGTTGATGGTTTATTGGGCTTGTGTTGT
 ATGAGATGGAATTTTAAATATTTTTAAATGTTTCTCTAGTCTTAGAGAAATTTTAAAGCAACTCAAGATAG
 GCTCTTCCCGCATATGATAATCCGTCAGGTGAATTTGGATTCTTTTATATCACAAAATGAATCCATGTTTTG
 GGAGGTAATGGTATCAGAATATATGGTGCAGGTCTTGGTAAAAACCAATAGATCTTTGAGAAATACAAG
 ACATCTCTGTGTTGAAACATCGTGTGTTTCTTATTTGCCAGAGTAGGAAAAGAGTAGATCTTTTTGCTCTCT
 AAATGATTGATGGGTTGTGTTTTTTTTCCACCTGCTAATAAATATTACATTGCAACATTCTTCCCTCAACT
 TCAAACTGCTGAACTGAAACAATATGCATAAAAAGAAAATCCTTTGCAGAAGAAAAAAGCTATTTTCTCC
 CACTGATTTTGAATGGCACTTGGGATGCAGTTCGAAAATCCTATTGCCATTCCCTCATGAACATTGTGAA
 ATGAAAACCTTTGGACAGTCTGCCGATGCGCATGAGACTGCCTGCGCAAGGCAAGGGTATGGTTCCCAAA
 GCACCCAGTGGTAAATCCTAACTTATTATCCCTTAAAATTCCAATGTAACAACGTGGGCCATAAAAAGAGT
 TTCTGAACAAAACATGTCATTTGTGGAAAGGTGTTTTTCGTAATTAATGATGGAATCATGCTCATTTCAA
 ATGGAGGTCCACGATTTGTGGCCAGCTGATGCCTGCAAATTAATCCTGGATCACTAACTCTGA

>STAR35

CGACTTGGTGATGCGGGCTCTTTTTTGGTTCCATATGAACTTTAAAGTAGTCTTTTCCAATTCTGTGAAGAA
 AGTCATTGGTAGGTTGATGGGGATGGCATTGAATCTGTAATTAACCTTGGGCAGTATGGCCATTTTCACAAT
 GTTGATTCCTCCTATCCATGATGATGGAATGTTCTTCCATTAGTTTGTATCCTCTTTTATTTCTTGGAGCAGT
 GGTTTGTAGTCTCCTTGAAGAGGTCCTTACATCCCTTGTAAAGTTGGATTCTAGGATTTTTATTCTCTTTG
 AAGCAATTGTGAATGGGAGTTCACCTCACGATTTGGCTCTCTGTTTGTCTGCTGGTGTATAAGAATGTTTGTG
 ATTTTTGTACATTGATTTTGTATCCTGAGACTTTGCTGAAGTTGCTTATCAGCTTAAGGAGCTTTTGGGCTGA
 GACAATGGGATTTTCTAGATATACAATCATGTCTGTGCAAAACAGGGACAATTTGACTTCTCTTTTCTAA
 TTGAATACACTTTATCTCTTCTCTGCTAATTGCCCTGGGCAGAACTTCCAACACTATGTTGAATAGGAG
 TGGTGAGAGAGGGCATCCCTGTCTTGTGCCAGTTTTCAAAGGGAATGCTTCCAGTTTTTGGCCATTCAGTAT
 GATATTGGCTGTGGGTTTGTATAGATAGCTCTTATTTATTTTGAATGTGTCCCATCAATACCTAATTTATTG
 AGAGTTTTTAGCATGAAGCATTGTTGAATTTTGTCAAAGGCTTTTTCTGCATCTATTGAGATAATCATGTGG
 TTTTTGTCTTTGGCTCTGTTTATATGCTGGATTACATTTATTGATTTGTGTATATTGAACCAGCCTTGCATCC
 CAGGGATGAAGCCCACTTGATC

>STAR40

GATCAAGAAAGCACTCCGGGCTCCAGAAGGAGCCTTCCAGGCCAGCTTTGAGCATAAGCTGCTGATGAGC
 AGTGAGTGTCTTGGTAGTGTTCAGGGCAGCATGTTACCATTTCATGCTTGACTTCTAGCCAGTGTGACGAG

```

AGGCTGGAGTCAGGTCTCTAGAGAGTTGAGCAGCTCCAGCCTTAGATCTCCCAGTCTTATGCGGTGTGCC
ATTGCTTTGTGTCTGCAGTCCCCTGGCCACACCCAGTAACAGTTCTGGGATCTATGGGAGTAGCTTCCTTA
GTGAGCTTTCCCTCAAATACTTTGCAACCAGGTAGAGAAGTTTGGAGTGAAGTTTTGTTCTTCGTTTCTT
CACAATATGGATATGCATCTTCTTTTAAAAATGTTAAAGTAAATTACCTCTCTTTTCAGATACTGTCTTCATG
CGAACTTGGTATCCTGTTTCCATCCCAGCCTTCTATAACCCAGTAACATCTTTTTTGAACCAGTGGGTGAG
AAAGACACCTGGTCAGGAACGCGGACCACAGGACAACCTCAGGCTCACCCACGGCATCAGACTAAAGGCAA
ACAAGGACTCTGTATAAAGTACCGGTGGCATGTGTATTAGTGGAGATGCAGCCTGTGCTCTGCAGACAGGG
AGTCACACAGACACTTTTCTATAATTTCTTAAGTGCTTTGAATGTTCAAGTAGAAAAGTCTAACATTTAAATTT
GATTGAACAATTGTATATTCATGGAATATTTTGGAAACGGAATACCAAAAAATGGCAATAGTGGTTCTTTCT
GGATGGAAGACAAACTTTTCTTTTAAAAATAAATTTTATTTTATATTTGAGGTTGACCACATGACCTTA
AGGATACATATAGACAGTAAACTGGTTACTACAGTGAAGCAAATTAACATATCTACCATCGTACATAGTTA
CATTTTTTGTGTGACAGGAACAGCTAAAATCTACGTATTTAACAAAACTCCTAAAGACAATACATTTTTAT
TAACTATAGCCCTCATGATGTACATTAGATC

```

DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Diagrama esquemático de la invención.

5 La figura 1A muestra una única unidad de expresión, gen 1, bajo el control del promotor de CMV en un plásmido. Este plásmido se ha integrado como múltiples copias en el genoma, en una orientación tal que la transcripción es convergente. Por consiguiente, habrá una transcripción de ultralectura a partir de la copia uno en la copia dos y viceversa. Esto dará como resultado la formación de ARNbc. Este plásmido presenta silenciamiento de los genes.

10 La figura 1B muestra dos unidades de expresión, gen 1 y gen 2, ambas bajo el control del promotor de CMV y ubicadas en una orientación divergente en un plásmido. Este plásmido se ha integrado como múltiples copias en el genoma. Independientemente de la orientación, habrá siempre transcripción de ultralectura a partir de un gen en la copia uno a otro gen en la copia dos. Esto da como resultado la formación de ARNbc. Este plásmido presenta silenciamiento de los genes.

15 La figura 1C muestra una única unidad de expresión, gen 1, bajo el control del promotor de CMV en un plásmido. Este plásmido se ha integrado como una única copia en el genoma. Cuando la integración es adyacente a un promotor que está orientado de una manera convergente en relación con el plásmido, habrá transcripción de ultralectura a partir de ese promotor en el gen 1 del plásmido. Esto da como resultado la formación de ARNbc. Este plásmido presenta silenciamiento de los genes.

Figura 2. Diagrama esquemático de la invención.

20 La figura 2A muestra la primera unidad de expresión. Está flanqueada por elementos TRAP y STAR, y comprende un gen bicistrónico que contiene (de 5' a 3') un transgén (que codifica por ejemplo para un gen indicador o una subunidad de una proteína multimérica; gen), un IRES y un marcador seleccionable (zeo, que confiere resistencia a zeocina) bajo el control del promotor de CMV. Se incluye un marcador seleccionable monocistrónico (puro) bajo el control del promotor de SV40. Ambos genes tienen el terminador de la transcripción de SV40 en sus extremos 3' (t). La TRAP se dibuja con una flecha que indica que en esa orientación particular no entra en la unidad de expresión transcripción impulsada por ningún promotor fuera de la unidad de expresión.

30 La figura 2B muestra dos unidades de expresión en un plásmido. Tanto el gen 1 como el gen 2 son ambos parte de un gen bicistrónico que contiene (de 5' a 3') un transgén (gen 1), un IRES y un marcador seleccionable (zeo con el gen 1 y puro con el gen 2) bajo el control del promotor de CMV y el terminador de la transcripción de SV40 (t). Todo el casete está rodeado por elementos STAR y TRAP, estas últimas orientadas de tal manera que se mantiene la transcripción fuera del casete y los elementos STAR.

35 La figura 2C muestra una unidad de expresión. Está flanqueada por elementos TRAP y STAR, y comprende un gen bicistrónico que contiene (de 5' a 3') un transgén (que codifica por ejemplo para un gen indicador o una subunidad de una proteína multimérica; gen), un IRES y un marcador seleccionable (zeo, que confiere resistencia a zeocina) bajo el control del promotor de CMV. El gen tiene el terminador de la transcripción de SV40 en sus extremos 3' (t). Se coloca una secuencia de TRAP en el sentido de 5' del STAR que está adyacente al promotor de CMV. Se colocan dos TRAP para flanquear el STAR en 3'. La secuencia de TRAP en el sentido de 3' (3'-5') está orientada para prevenir que la transcripción impulsada por cualquier promotor fuera de la unidad de expresión entre en la unidad de expresión.

Figura 3. El plásmido pcodA para identificar y someter a prueba supuestas TRAP.

40 El plásmido pIRES-31 contiene el promotor de CMV en el sentido de 5' de una secuencia intermedia IV (Clontech) que contiene un sitio de clonación múltiple en el que se clonan supuestas TRAP. En el sentido de 3' está el gen suicida codA::upp. El plásmido comprende además el gen de resistencia a higromicina que está bajo el control del promotor de SV40. El plásmido también tiene un origen de replicación (ori) y gen de resistencia a ampicilina (amp^R) para la propagación en *Escherichia coli* y el antígeno nuclear EBNA-1 para una replicación episomal de alto número de copias. La TRAP se dibuja como una flecha que indica que la TRAP bloquea la transcripción impulsada por el

promotor de CMV en esta orientación particular. Esto es de importancia para la orientación de TRAP en la figura 2, que también se dibujan como flechas para indicar la orientación específica de las TRAP para prevenir que la transcripción impulsada por cualquier promotor fuera de la unidad de expresión entre en esta unidad de expresión.

Figura 4. Las TRAP bloquean eficazmente la transcripción impulsada por el promotor de CMV

- 5 Los constructos indicados con posibles TRAP que están ubicadas entre el promotor de CMV y el gen *codA* se transfectan en células U-2 OS.
- 1) El vector de control vacío sin secuencias entre el promotor de CMV y el gen *codA* (figura 3)
 - 2) Un ADN de 2400 pb de longitud de fago λ (pb 35711-38103) en orientación 5'-3'
 - 3) Un ADN de 2400 pb de longitud de fago λ (pb 35711-38103) en orientación 3'-5'
- 10 4) La secuencia de ADN de MAZ de 60 pb de longitud (Ashfield *et al* 1994)
- 5) STAR7
 - 6) STAR 40
 - 7) El vector de control vacío tal como se muestra en la figura 3
 - 8) Una secuencia de poliA sintética (SPA) de 50 pb de longitud (Levitt *et al* 1989) en orientación 5'-3'
- 15 9) Una combinación de la secuencia de SPA de 50 pb de longitud y una señal de pausa del gen de globina $\alpha 2$ de 92 pb de longitud en orientación 5'-3'
- 10) Una secuencia de poliA sintética de 50 pb de longitud en orientación 3'-5'
 - 11) Una combinación de la secuencia de SPA de 50 pb de longitud y una señal de pausa del gen de globina $\alpha 2$ de 92 pb de longitud en orientación 3'-5'
- 20 Tras la selección de transfección mediante higromicina se aísla y se transfiere el ARNm. La transferencia se incuba con una sonda marcada radiactivamente que abarca el gen *codA*. Como control de carga también se incuba la transferencia con una sonda radiactiva que abarca el gen de resistencia a higromicina. El fragmento de lambda (pb 35711-38103) bloquea eficazmente la transcripción del gen *codA* impulsada por el promotor de CMV en la orientación 5'-3' (carril 2), pero no en la orientación 3'-5' (carril 3). Además, una secuencia de poliA sintética (SPA) o bien sola (carril 8) o bien en combinación con una señal de pausa de globina $\alpha 2$ de 92 pb de longitud (carril 9) bloquea eficazmente la transcripción en la orientación 5'-3', pero no en la orientación 3'-5' (carriles 10 y 11). Ni la secuencia de MAZ (carril 4), STAR 7 (carril 5) ni STAR 40 (carril 6) impiden la transcripción del gen *codA* impulsada por el promotor de CMV. Todas las señales se comparan con el vector de control que no contiene ninguna supuesta secuencia de TRAP (carriles 1 y 7).

30 **Figura 5. TRAP mejoran los efectos de elementos STAR sobre la expresión transgénica**

Se transfectan constructos que están flanqueados con la TRAP de lambda (pb 35711-38103) en la orientación A, STAR 40 o TRAP de lambda (pb 35711-38103)/STAR40 combinados en células CHO-K1. La orientación 5'-3' de TRAP de lambda (pb 35711-38103) da como resultado el bloqueo de la transcripción (figura 3) y las TRAP se colocan para flanquear todo el constructo de manera que la transcripción no puede entrar en las unidades de expresión. Se expanden colonias estables (14 de cada constructo) y se determina la señal de GFP en un citómetro de flujo XL-MCL Beckman Coulter. Para cada colonia independiente se representa gráficamente la media de la señal de GFP. Esto se toma como medida para el nivel de expresión de GFP. Se comparan los resultados con colonias que se transfectan con un constructo que no contiene ni TRAP de lambda (pb 35711-38103) ni elemento STAR (control).

40 **Figura 6. TRAP actúan de una manera dependiente de la orientación**

Se transfectan constructos que están flanqueados con la TRAP de lambda (pb 35711-38103) en la orientación 3'-5', STAR 40 o TRAP de lambda (pb 35711-38103)/STAR40 combinados en células CHO-K1. La orientación 3'-5' de TRAP de lambda (pb 35711-38103) no da como resultado el bloqueo de la transcripción (figura 3). El análisis de colonias estables es como en la figura 5.

45 **Figura 7. TRAP y STAR mejoran la estabilidad de la expresión transgénica**

Se expanden colonias transfectadas de manera estable que contienen o bien un constructo de GFP sin TRAP de lambda (pb 35711-38103)/STAR (control) o bien el constructo de GFP que está flanqueado por TRAP de lambda (pb 35711-38103)/STAR40 combinados. De las dos categorías se escogen cuatro colonias con los niveles de GFP más altos (véase la figura 5). Estas colonias se cultivan adicionalmente sin el antibiótico (zeocina) y se determina la señal

de GFP con intervalos de una semana, que representan aproximadamente 10 ciclos celulares. Se representa gráficamente la media de la señal de GFP como en la figura 3. La primera barra de cada colonia representa la señal de GFP en el momento en el que se elimina la presión de selección con antibióticos. Las tres barras adyacentes representan la señal de GFP que se mide después de una, dos y tres semanas.

5 **Figura 8. Función de TRAP en el contexto de un elemento STAR.**

La secuencia de SPA/pausa se coloca en 5' o 3' de STAR7 y posteriormente se somete a prueba en el vector de codA, tras la transfección en células CHO. De esta manera se someten a prueba siete insertos diferentes para determinar su capacidad para bloquear la transcripción de codA impulsada por CMV, esto en comparación con la señal del vector vacío (carril 1). Los siguientes carriles indican los insertos en los constructos de codA:

- 10 1) El vector de control vacío tal como se muestra en la figura 3.
 2) STAR 7
 3) La secuencia de SPA/pausa en la orientación 5'-3'
 4) La secuencia de SPA/pausa en la orientación 3'-5'
 5) La secuencia de SPA/pausa en la orientación 5'-3' y colocada en 5' de STAR7
 15 6) La secuencia de SPA/pausa en la orientación 3'-5' y colocada en 5' de STAR7
 7) La secuencia de SPA/pausa en la orientación 5'-3' y colocada en 3' de STAR7
 8) La secuencia de SPA/pausa en la orientación 3'-5' y colocada en 3' de STAR7

20 Se usa el gen de resistencia a higomicina como control interno, indicando el número de copias de los plásmidos. La secuencia de SPA/pausa funciona como TRAP sólo cuando se usa en la orientación 5'-3' (carriles 3, 5 y 7), o bien cuando se usa sola (carril 3), colocada en 5' de STAR7 (carril 5) o bien colocada en 3' de STAR 7 (carril 7). Cuando se usa en la orientación 3'-5' (carriles 4, 6 y 8), la secuencia de SPA/pausa no funciona como TRAP, independientemente de usarse sola o en combinación con STAR 7.

Bibliografía

- 25 Ashfield, R, Patel, AJ, Bossone, SA, Brown, H, Campbell, RD, Marcu, KB, y Proudfoot, NJ. MAZ-dependent termination between closely spaced human complement genes. *EMBO J* 13, 5656-5667.
- Berger, J, Hauber, J, Hauber, R, Geiger, R, and Cullen, BR. (1988) Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells *Gene* 66, 1-10.
- Boshart, M, Weber, F, Jahn, G, Dorsch-Hasler, K, Fleckenstein, B, y Schaffner, W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus *Cell* 41, 521-30.
- 30 Chad, HE, y Chamow, SM, 2001, Therapeutic antibody expression technology *Curr Opin Biotechnol* 12, 188-194. Das, RC, 2001. Proteins and antibodies make advances as therapeutic products. *Am Clin Lab* 20, 8-14
- Chevet, E, Cameron, PH, Pelletier, MF, Thomas, DY, y Bergeron, JJ. (2001) The endoplasmic reticulum: integration of protein folding, quality control, signaling and degradation *Curr Opin Struct Biol* 11, 120-4.
- 35 Das, GC, Niyogi, SK, y Salzman, NP. (1985) SV40 promoters and their regulation *Prog Nucleic Acid Res Mol Biol* 32, 217-36.
- Eszterhas, SK, Bouhassira, EE, Martin, DI, y Fiering, S. (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position *Mol Cell Biol* 22, 469-79.
- Solicitud de patente europea 01202581.3
- 40 Garrick, D, Fiering, S, Martin, DI, y Whitelaw, E. (1998) Repeat-induced gene silencing in mammals *Nat Genet* 18, 56-9.
- Kain, SR. (1997) Use of secreted alkaline phosphatase as a reporter of gene expression in mammalian cells *Methods Mol Biol* 63, 49-60.
- Kaufman, RJ. (2000) Overview of vector design for mammalian gene expression *Mol Biotechnol* 16, 151-60.
- 45 Kaufman, RJ. (1990) Selection and coamplification of heterologous genes in mammalian cells *Methods in Enzymology* 185, 536-566.

- Kaufman, RJ, y Sharp, PA. (1982) Construction of a modular dihydrofolate reductase cDNA gene: analysis of signals utilized for efficient expression *Mol Cell Biol* 2, 1304-19.
- Levitt N, Briggs D, Gil A, y Proudfoot NJ. (1989) Definition of an efficient synthetic poly(A) site. *Genes & Dev* 3, 1019-1025.
- 5 Liu, DT. (1992) Glycoprotein pharmaceuticals: scientific and regulatory considerations, and the US Orphan Drug Act *Trends Biotechnol* 10, 114-20.
- Lopez de Quinto, S, y Martinez-Salas, E. (1998) Parameters influencing translational efficiency in aphthovirus IRES-based bicistronic expression vectors *Gene* 217, 51-6.
- Martin, DI, y Whitelaw, E. (1996) The vagaries of variegating transgenes *Bioessays* 18, 919-23.
- 10 Martinez-Salas, E. (1999) Internal ribosome entry site biology and its use in expression vectors *Curr Opin Biotechnol* 10, 458-64.
- 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.
- Meyer, P. (2000) Transcriptional transgene silencing and chromatin components *Plant Mol Biol* 43, 221-34.
- 15 Migliaccio, AR, Bengra, C, Ling, J, Pi, W, Li, C, Zeng, S, Keskinetepe, M, Whitey, B, Sanchez, M, Migliaccio, G, y Tuan, D. (2000) Stable and unstable transgene integration sites in the human genome: extinction of the Green Fluorescent Protein transgene in K562 cells *Gene* 256, 197-214.
- Mizuguchi, H, Xu, Z, Ishii-Watabe, A, Uchida, E, y Hayakawa, T. (2000) IRES-dependent second gene expression is significantly lower than cap- dependent first gene expression in a bicistronic vector *Mol Ther* 1, 376-82.
- 20 Solicitud de patente PCT, PCT/NL02/00390
- 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, 102-4, 106, 108-10.
- 25 Sambrook, J, Fritsch, EF, y Maniatis, T (1989) *Molecular Cloning: A Laboratory Manual*, segunda ed., Cold Spring Harbor Laboratory Press, Plainview NY.
- Sanger, F, Nicklen, S, y Coulson, AR. (1977) DNA sequencing with chain-terminating inhibitors *Proc Natl Acad Sci U S A* 74, 5463-7.
- Schorpp, M, Jager, R, Schellander, K, Schenkel, J, Wagner, EF, Weiher, H, y Angel, P. (1996) The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice *Nucleic Acids Res* 24, 1787-8.
- 30 Sheeley, DM, Merrill, BM, y Taylor, LC. (1997) Characterization of monoclonal antibody glycosylation: comparison of expression systems and identification of terminal alpha-linked galactose *Anal Biochem* 247, 102-10.
- Stam, M, Viterbo, A, Mol, JN, y Kooter, JM. (1998) Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants *Mol Cell Biol* 18, 6165-77.
- 35 Stam, M, De Bruin, R, Van Blokland, R, Ten Hoom, RAL, Mol, JN, y Kooter, JM. (2000) Distinct features of posttranscriptional gene silencing by antisense transgenes in single copy and inverted T-DNA repeat loci. *Plant J* 21, 27-42.
- Strutzenberger, K, Borth, N, Kunert, R, Steinfellner, W, y Katinger, H. (1999) Changes during subclone development and ageing of human antibody- producing recombinant CHO cells *J Biotechnol* 69, 215-26.
- 40 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.
- Wright, A, y Morrison, SL. (1997) Effect of glycosylation on antibody function: implications for genetic engineering *Trends Biotechnol* 15, 26-32.
- 45 Yang, TT, Sinai, P, Kitts, PA, y Kain, SR. (1997) Quantification of gene expression with a secreted alkaline phosphatase reporter system *Biotechniques* 23, 1110-4.

REIVINDICACIONES

1. Unidad de expresión de proteína que comprende funcionalmente unidos de 5' a 3': i) un promotor, ii) un marco de lectura abierto que codifica para una proteína de interés y iii) una secuencia señal de terminación de la transcripción, caracterizada por que dicha unidad de expresión de proteína comprende además al menos una segunda secuencia señal de terminación de la transcripción ubicada en el sentido de 3' de dicho marco de lectura abierto, en la que dicha segunda secuencia señal de terminación de la transcripción está en orientación opuesta de la primera secuencia señal de terminación de la transcripción, y en la que dicha segunda secuencia señal de terminación de la transcripción se escoge de las secuencias presentadas en la tabla 1.
2. Unidad de expresión de proteína según la reivindicación 1, comprendiendo además dicha unidad de expresión de proteína una tercera secuencia señal de terminación de la transcripción en el sentido de 5' de dicho promotor, en la que dicha segunda secuencia señal de terminación de la transcripción se escoge de las secuencias presentadas en la tabla 1.
3. Unidad de expresión de proteína según la reivindicación 1 ó 2, comprendiendo además dicha unidad de expresión de proteína al menos una secuencia anti-represora estabilizante (STAR) presentada en la tabla 2, escogida de STAR4, STAR6, STAR7, STAR12, STAR18, STAR35 y STAR40.
4. Unidad de expresión de proteína según una cualquiera de las reivindicaciones anteriores, comprendiendo además dicha unidad de expresión de proteína al menos dos secuencias anti-represoras estabilizantes (STAR) presentadas en la tabla 2, escogidas de STAR4, STAR6, STAR7, STAR12, STAR18, STAR35 y STAR40.
5. Unidad de expresión de proteína según la reivindicación 4, en la que dichas al menos dos secuencias STAR están dispuestas de manera que dichas secuencias STAR flanquean la combinación formada por dicho promotor, marco de lectura abierto y señal de terminación de la transcripción.
6. Unidad de expresión de proteína según la reivindicación 5, que comprende en el siguiente orden:
 - a) una secuencia señal de terminación de la transcripción, escogida de las secuencias presentadas en la tabla 1,
 - b) una secuencia STAR,
 - c) la combinación formada por dicho promotor, marco de lectura abierto y señal de terminación de la transcripción,
 - d) una secuencia STAR, y
 - e) una secuencia señal de terminación de la transcripción en orientación opuesta, escogida de las secuencias presentadas en la tabla 1.
7. Unidad de expresión de proteína según una cualquiera de las reivindicaciones anteriores, en la que dicha proteína de interés es una cadena de inmunoglobulina o una parte funcional, derivado y/o análogo de la misma.
8. Unidad de expresión de proteína según una cualquiera de las reivindicaciones anteriores, comprendiendo dicha unidad de expresión de proteína un gen monocistrónico que comprende funcionalmente unidos de 5' a 3': i) un promotor, ii) el marco de lectura abierto que codifica para dicha proteína de interés y iii) una señal de terminación de la transcripción.
9. Unidad de expresión de proteína según una cualquiera de las reivindicaciones 1-8, comprendiendo dicha unidad de expresión de proteína un gen bicistrónico que comprende funcionalmente unidos de 5' a 3': i) un promotor, ii) el marco de lectura abierto que codifica para dicha proteína de interés, iii) un marcador de selección y iv) una señal de terminación de la transcripción, y preferiblemente dicho gen bicistrónico comprende un sitio de iniciación de la traducción de proteínas con eficacia de traducción reducida ubicado en el sentido de 3' de dicho marco de lectura abierto.
10. Unidad de expresión de proteína según la reivindicación 9, en la que dicho sitio de iniciación de la traducción de proteínas con eficacia de traducción reducida comprende un sitio interno de entrada al ribosoma.
11. Método para la expresión de al menos una proteína de interés en una célula que comprende proporcionar a dicha célula al menos una unidad de expresión de proteína según una cualquiera de las reivindicaciones anteriores.
12. Método según la reivindicación 11, que comprende además proporcionar a dicha célula una segunda

unidad de expresión de proteína.

13. Célula que comprende al menos una unidad de expresión de proteína según una cualquiera de las reivindicaciones 1-10.
14. Célula según la reivindicación 13, que es una célula vegetal o una célula de mamífero.
- 5 15. Método para producir una proteína de interés que comprende cultivar una célula según la reivindicación 13 ó 14 y recoger dicha proteína de interés del cultivo correspondiente.
16. Método para identificar una secuencia señal de terminación de la transcripción que comprende
- proporcionar a una célula un ácido nucleico que comprende
 - una secuencia promotora
- 10 una secuencia intermedia (IV) en el sentido de 3' de dicho promotor, que comprende una supuesta secuencia señal de terminación de la transcripción
- una secuencia cuyo producto es detectable y cuya secuencia se ubica en el sentido de 3' de dicha IV
- determinar la cantidad de dicho producto detectable y comparar dicha cantidad con la cantidad de producto obtenido en una célula que está provista de un ácido nucleico de control sin dicha supuesta secuencia señal de terminación de la transcripción.
- 15 17. Método según la reivindicación 16, en el que dicho ácido nucleico comprende además un marcador de selección ubicado fuera de la combinación de dicho promotor, IV, secuencia señal de terminación de la transcripción supuesta y dicha secuencia cuyo producto es detectable y que comprende además
- 20 seleccionar una célula mediante dicho marcador de selección de dicho ácido nucleico, obteniendo de ese modo una célula que comprende dicho ácido nucleico.
18. Método según la reivindicación 16 ó 17, en el que dicha secuencia cuyo producto es detectable es un gen suicida, preferiblemente *codA* o *codA::upp*.

FIG 1

Pérdida de constructos de expresión génica

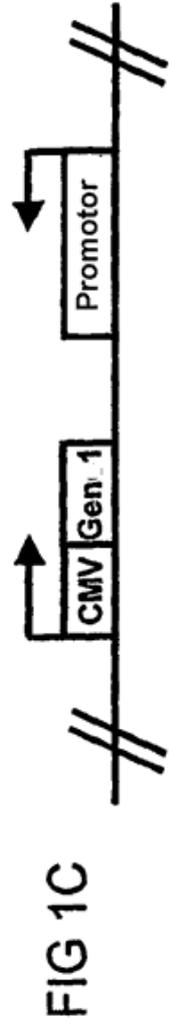


Fig 2

Diagrama esquemático de la invención

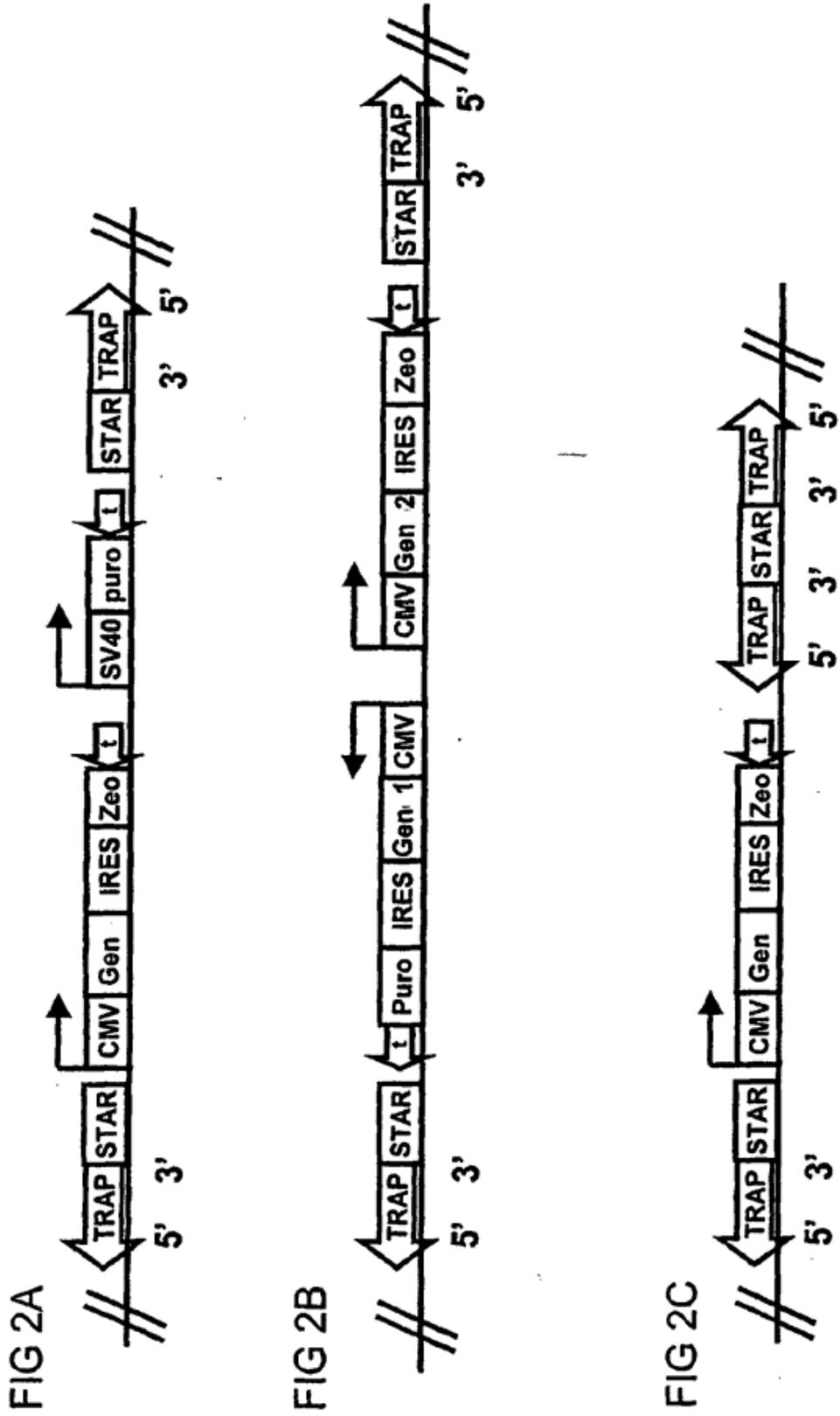


FIG 3

Plásmido para identificar y someter a prueba supuestas secuencias de TRAP

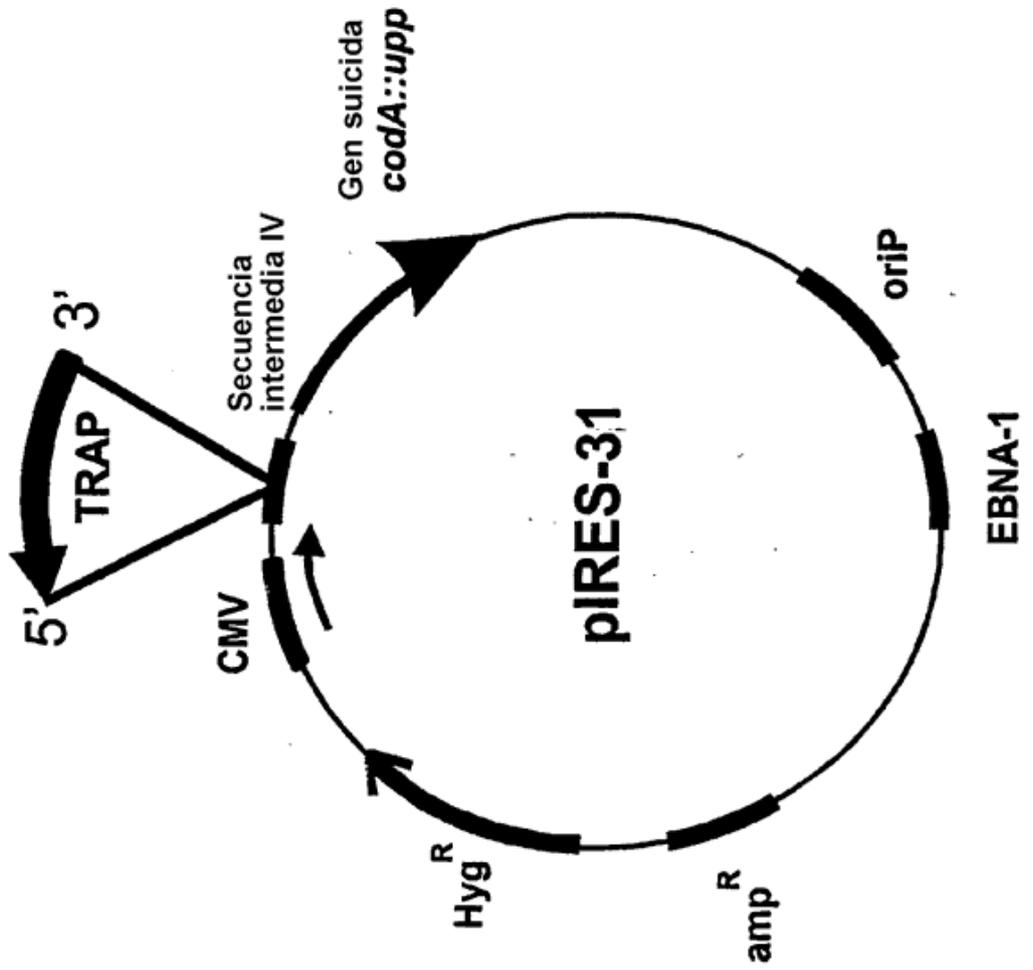


FIG 4

Secuencias de TRAP bloquean la transcripción impulsada por CMV

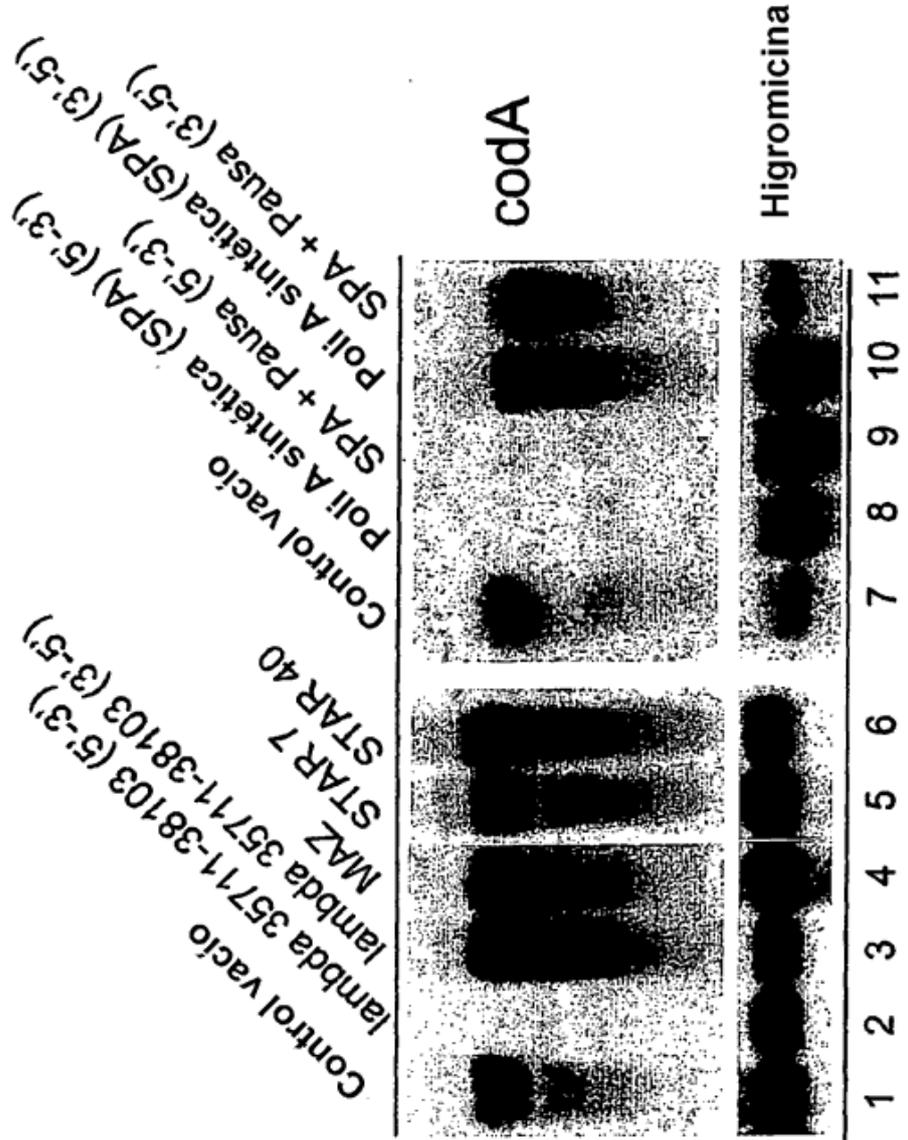


FIG 5

Una TRAP influye positivamente en la potenciación mediada por STAR de la expresión de GFP impulsada por el promotor de CMV en células CHO

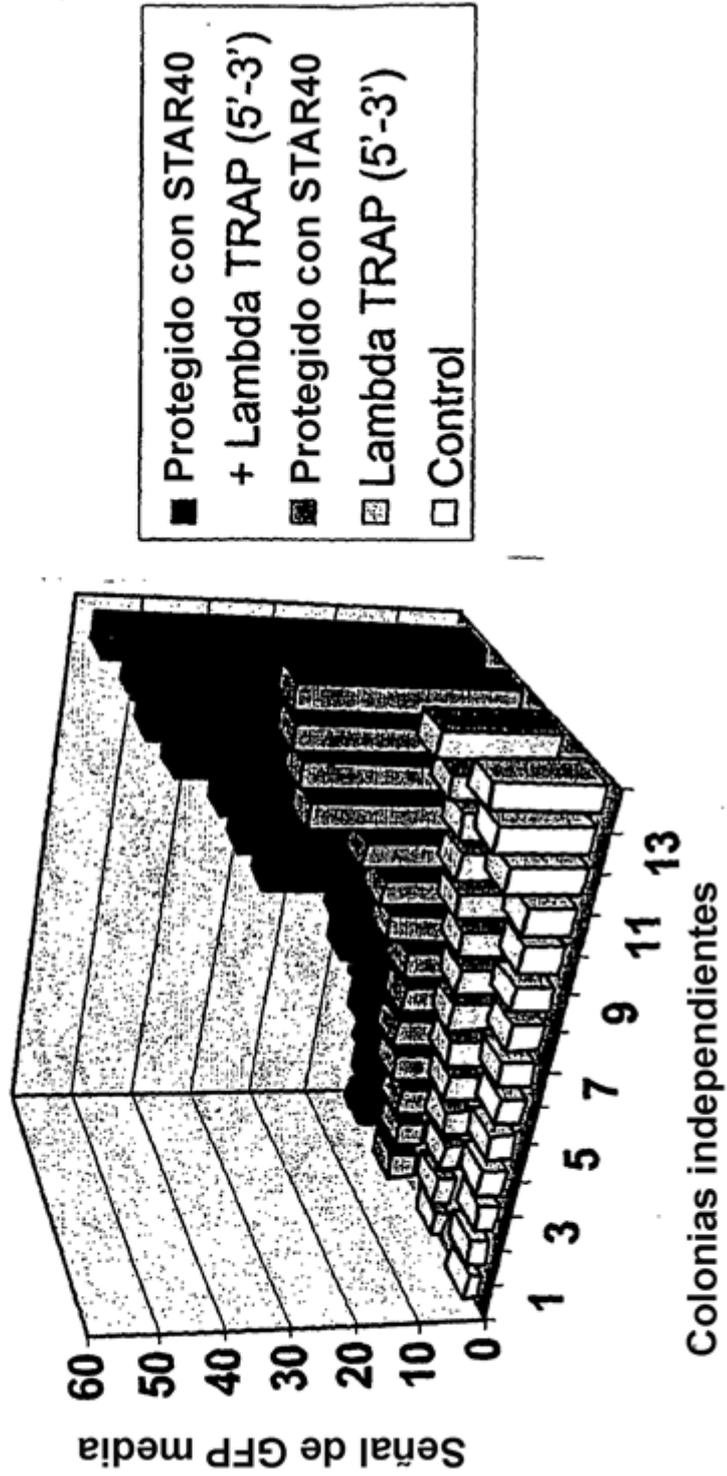


FIG 6

Dependencia de la orientación de la influencia de una TRAP sobre la expresión de GFP impulsada por el promotor de CMV en células CHO

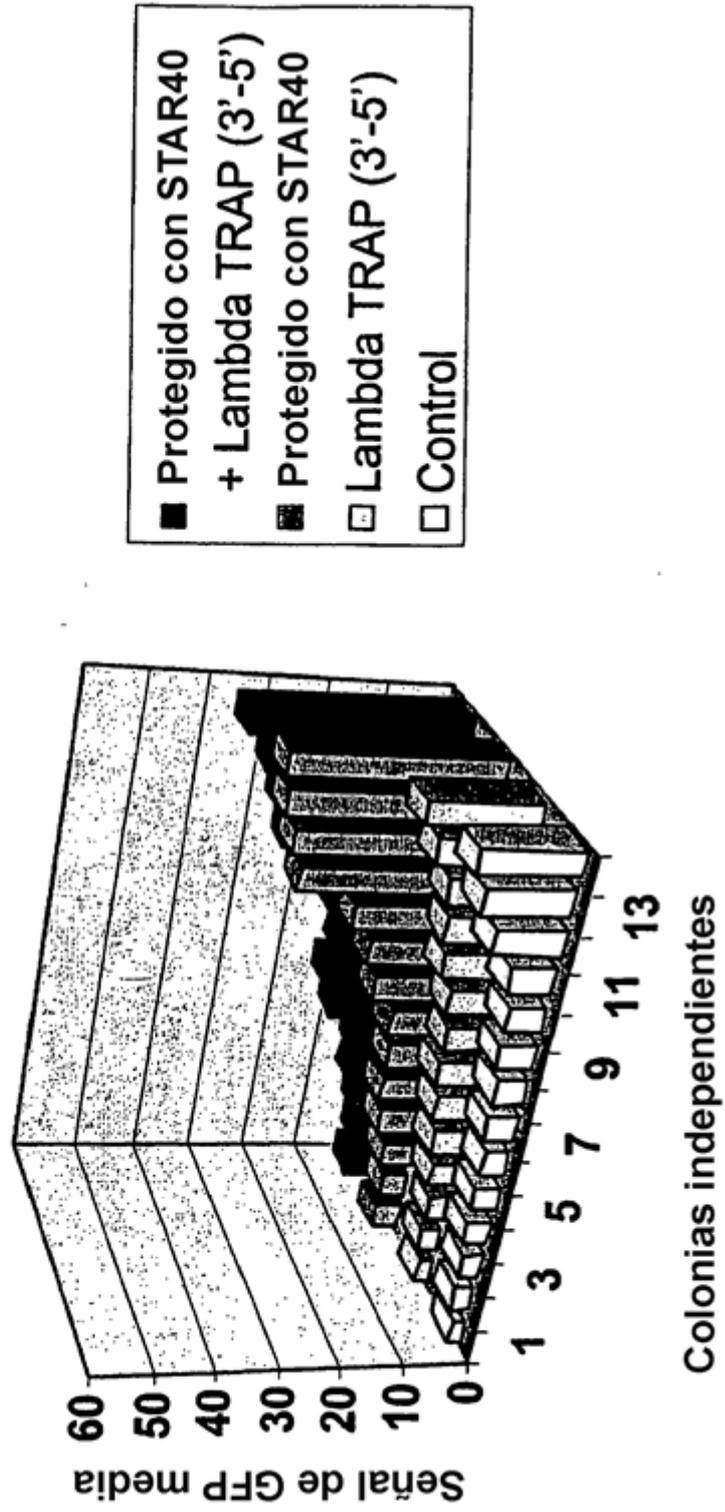
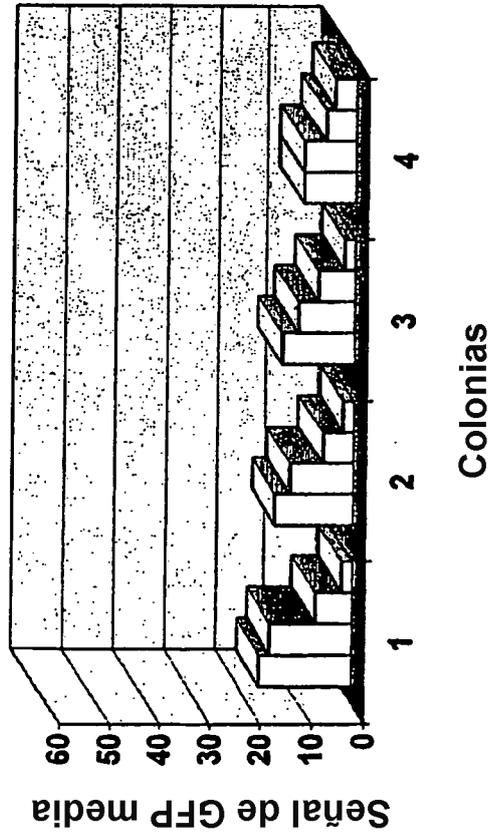


FIG 7

La aplicación de TRAP y STAR potencia la estabilidad de la expresión de GFP impulsada por el promotor de CMV en células CHO

Estabilidad en colonias sin STAR/TRAP



Efectos sobre la estabilidad de STAR + TRAP

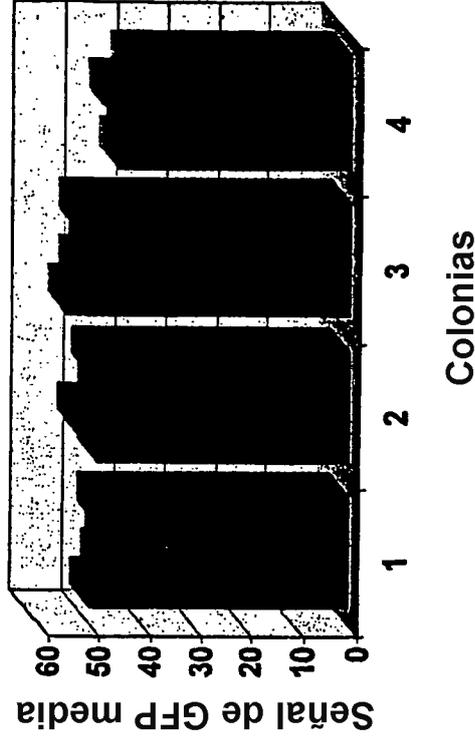


FIG 8
Secuencias de TRAP bloquean la transcripción impulsada por CMV
en el contexto de un elemento STAR

