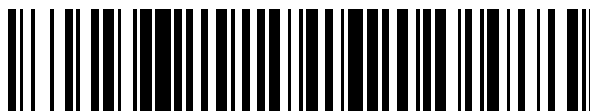


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 421 152**

51 Int. Cl.:

C12N 15/09 (2006.01)

C12N 15/11 (2006.01)

C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.07.2009 E 09800055 (7)**

97 Fecha y número de publicación de la concesión europea: **03.07.2013 EP 2310501**

54 Título: **Nuevos elementos reguladores**

30 Prioridad:

23.07.2008 EP 08161029

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

29.08.2013

73 Titular/es:

**BOEHRINGER INGELHEIM PHARMA GMBH & CO.
KG (100.0%)
Binger Strasse 173
55216 Ingelheim am Rhein, DE**

72 Inventor/es:

ENENKEL, BARBARA

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 421 152 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

Nuevos elementos reguladores

Antecedentes de la invención

5 Campo técnico

La invención se refiere al campo de la tecnología del cultivo celular. Se refiere a nuevos elementos reguladores así como a un método para mejorar la expresión de polipéptidos a partir de ácidos nucleicos tales como genes clonados y a la producción de varios polipéptidos en una célula hospedante eucariota utilizando dichos nuevos elementos reguladores.

10 Antecedentes

El mercado de los compuestos biofarmacéuticos para uso en terapia humana continúa creciendo a gran velocidad con más de 900 compuestos biofarmacéuticos evaluados en estudios clínicos y unas ventas estimadas de 50.000 millones en 2010. Actualmente, se produce un número creciente de compuestos biofarmacéuticos a partir de células de mamífero debido a la capacidad de éstas para producir y modificar correctamente las proteínas humanas. Por lo tanto las proteínas recombinantes son compatibles con los seres humanos tanto funcional como farmacocinéticamente. Una deficiencia en comparación con los sistemas de expresión procarióticos es a menudo el nivel de expresión de proteínas significativamente más bajo. La producción satisfactoria y con alto rendimiento de compuestos biofarmacéuticos a partir de células de mamífero es por lo tanto crucial y está determinada por diferentes factores incluyendo la línea de la célula hospedante, el sistema de expresión, el crecimiento y la productividad celular, los medios de cultivo y de alimentación, el procedimiento de producción y purificación, la estructura y secuencia de la proteína, la estabilidad y formulación de la proteína.

La expresión de la proteína recombinante requiere un vector de expresión que codifica el gen de interés deseado. Se han empleado varios métodos para optimizar los vectores de expresión para una producción eficiente de proteína. La expresión génica se regula a niveles de transcripción y de traducción. Por lo tanto muchos métodos se refieren a la identificación y optimización de promotores y potenciadores fuertes para mejorar la eficiencia con la que se transcriben los genes que codifican la proteína. Ejemplos de ellos son el promotor y potenciador temprano inmediato de CMV (citomegalovirus), el promotor y potenciador de SV40 (virus de simio 40), el promotor del factor de elongación (EF), el potenciador de polioma, y el promotor de [beta]-actina de pollo. Asimismo, se utilizan también las secuencias de señales de poliadenilación fuertes que estabilizan los ARNm y potencian la terminación de la transcripción, para aumentar la expresión de la proteína a partir de genes codificados por los vectores de expresión. Entre los métodos para mejorar la eficiencia con la que se traduce el ARNm resultante están el uso de los sitios de iniciación de la traducción (AUG), los sitios de unión óptima al ribosoma tales como la secuencia de Kozak (GCCGCCACCAUGG; AUG constituye el codón de inicio) o los sitios internos de entrada al ribosoma (IRES).

Uno de los métodos empleados para optimizar los vectores de expresión con el fin de obtener niveles más altos de expresión de genes recombinantes en células eucariotas se refiere al uso y selección de las señales de poliadenilación. Se utilizan una variedad de señales de poliadenilación en los vectores para la expresión de proteínas recombinantes. Las más comúnmente usadas incluyen por ejemplo las señales de poliadenilación procedentes de la hormona de crecimiento bovina (BGH) (documentos US 5.122.458; EP 0173552), región tardía y temprana del virus de simio 40, beta-globina de conejo, inmunoglobulinas humanas o de ratón, región tardía de virus polioma.

En el ARN mensajero (ARNm) de eucariotas la región 3' no traducida (3'UTR) es un elemento regulador importante. En muchos casos dicta la estabilidad del ARNm y puede regular también la eficiencia de la traducción. Las señales de poliadenilación son secuencias de nucleótidos dentro de la 3'UTR que dirigen la unión de un complejo de proteína de poliadenilación con una secuencia AAUAAA dentro de la secuencia señal. El complejo contiene una endonucleasa que corta el ARNm aproximadamente 14 a 30 nucleótidos corriente abajo de la secuencia AAUAAA y una polimerasa que incorpora post-transcripcionalmente una cadena de aproximadamente 100 a 200 nucleótidos de adenina (cola poliA) al extremo 3' escindido. Se cree que la cola poliA influye en muchos aspectos del metabolismo de ARNm, incluyendo la estabilidad, la eficiencia de la traducción, y el transporte desde el núcleo al citoplasma. Típicamente, la señal de poliadenilación consiste en dos elementos de reconocimiento que flanquean el sitio de escisión y poliadenilación: una secuencia AAUAAA altamente conservada de aproximadamente 14 a 30 nucleótidos corriente arriba del sitio de escisión y una región rica en G/U o en U pobremente conservada de aproximadamente 20 a 50 nucleótidos corriente abajo de la secuencia AAUAAA. La escisión entre estos dos elementos se efectúa usualmente en el lado 3' de un residuo A. La eficiencia con la que se procesan diferentes sitios de poliadenilación varía considerablemente *in vivo*. La velocidad de ensamblaje del complejo de la proteína de poliadenilación es un proceso en multietapas y guarda correlación con la resistencia de la secuencia de la señal de poliadenilación. Por

ejemplo, debido a la velocidad más rápida de ensamblaje, la escisión en la señal de poliadenilación tardía fuerte de SV40 tiene lugar más rápidamente que en la señal de poliadenilación temprana más débil de SV40 (Chao et al., Molecular and Cellular Biology, Vol. 19 (8), 5588-5600, 1999).

5 Existe la necesidad de identificar señales alternativas de poliadenilación fuertes o incluso muy fuertes para acelerar la generación de líneas celulares altamente productoras para la producción de proteínas recombinantes. El uso de señales de poliadenilación fuertes o incluso muy fuertes mejora la terminación de la transcripción que a su vez da como resultado un aumento de la producción, estabilidad, exportación nuclear y/o traducción del ARNm codificado por el vector. Esto debería llevar a niveles más altos de ARNm y por lo tanto ocasionar una productividad más alta de las células productoras.

10 **Compendio de la invención/Solución**

Se describe aquí una nueva señal de poliadenilación aislada de la hormona de crecimiento del hámster chino (*Cricetus griseus*). Sorprendentemente, se ha encontrado que esta señal de poliadenilación recién identificada, denominada HGH, aventaja a las señales fuertes de poliadenilación BGH y SV40 tardía. Cuando se utilizan vectores que comprenden HGH como secuencia de la señal de poliadenilación, los títulos de proteína en las transfecciones transitorias de las células CHO-DG44 aumentaron hasta un 35 % en comparación con las células que comprenden BGH. En células estables se obtuvieron altas productividades específicas de hasta 45 pg/célula/día y títulos en los procesos de lote alimentado de hasta 6,3 g/L.

Una realización de la presente invención es una secuencia de polinucleótidos que comprende al menos una señal de poliadenilación HGH y al menos una secuencia de nucleótidos heterólogos que codifica un producto de interés. La señal de poliadenilación HGH está corriente abajo y está ligada operativamente a la secuencia o secuencias de nucleótidos heterólogos. Otra realización de la presente invención es un nuevo vector que comprende al menos una secuencia de nucleótidos heterólogos que codifica un producto de interés y al menos una señal de poliadenilación HGH. La señal de poliadenilación HGH está corriente abajo y está ligada operativamente a la secuencia o secuencias de nucleótidos heterólogos. Otra realización de la presente invención es un nuevo vector o secuencia de polinucleótidos que comprende al menos una señal de poliadenilación HGH ligada operativamente a un sitio de clonación múltiple corriente arriba que permite la clonación del gen de interés mediante secuencias de reconocimiento de las endonucleasas de restricción. Otra realización más de la presente invención es una célula eucariota, preferiblemente una célula de mamífero, que comprende la señal de poliadenilación HGH. Todavía una realización más de la presente invención es un método para producir un producto de interés que comprende cultivar células eucariotas, preferiblemente células de mamífero transfectadas con vectores o secuencias de polinucleótidos que comprenden la señal de poliadenilación HGH. En una realización preferida el producto de interés es un polipéptido y el polipéptido deseado se recupera del medio de cultivo.

Los datos de la presente invención muestran el impacto de la secuencia de la señal de poliadenilación HGH sobre la expresión transitoria de sICAM (Figura 6). Sorprendentemente, la expresión más alta de sICAM se obtiene con la secuencia de la señal de poliadenilación derivada del gen de la hormona de crecimiento de hámster. El título aumenta hasta un 21 % (serie de transfección #1) en comparación con las células transfectadas con el vector pJR110 que contiene la señal de poliadenilación BGH y aumenta hasta un 40 % (serie de transfección #1 en comparación con las células transfectadas con el vector pJR106 que contiene la señal de poliadenilación SV40 tardía).

Los datos de la presente invención muestran además el impacto de la señal de poliadenilación HGH sobre la expresión transitoria de un anticuerpo IgG4. Sorprendentemente, los títulos obtenidos con la secuencia de la señal de poliadenilación HGH son como promedio un 35 % más altos que para la señal de poliadenilación BGH (Figura 7).

Los datos de la presente invención muestran adicionalmente una prueba de diferentes variantes HGH (Figura 8). La secuencia HGH más corta de 113 pares de bases (bp) contenida en el vector de expresión pJR135 lleva hasta un 78 % de reducción de la expresión de sICAM en comparación con la secuencia HGH secuencia de 324 bp contenida en el vector de expresión pJR131. Por otro lado la secuencia HGH de 189 bp contenida en el vector de expresión pJR134 produce una expresión de sICAM muy buena comparable al nivel de expresión alcanzado con BGH (Figura 7). El mejor resultado de expresión se alcanza con la secuencia HGH de 324 bp contenida en el vector de expresión pJR131 (Figura 8), que es mucho mejor (35 %) que la expresión alcanzada con la señal de poliadenilación BGH (Figura 7).

Esto demuestra que entre las regiones de HGH de 190 a 324 bp de la SEQ ID NO:8 hay localizadas secuencias que contribuyen a una expresión eficiente de un gen de interés.

Los datos de la presente invención muestran además la expresión estable de proteínas a altos niveles utilizando la señal de poliadenilación HGH. Se obtienen conjuntos de células y clones de células con productividades específicas en el intervalo de 10-45 pg/célula/día y títulos en los procesos de lote alimentado de hasta 6,3 g/L (Figura 9).

- En una realización específica la invención se refiere a una señal de poliadenilación que comprende un ácido nucleico que comprende una secuencia que es idéntica al menos en un 75 % a la SEQ ID NO:8. La invención se refiere específicamente a una señal de poliadenilación que comprende un ácido nucleico que consiste esencialmente en una secuencia que es idéntica al menos en un 75 % a la SEQ ID NO:8. La invención se refiere preferiblemente a una señal de poliadenilación que comprende un ácido nucleico que consiste en una secuencia idéntica al menos en un 75 % a la SEQ ID NO:8. La invención se refiere además a una señal de poliadenilación que comprende un ácido nucleico que comprende la SEQ ID NO:8.
- En otra realización de la presente invención la señal de poliadenilación comprende una secuencia idéntica al menos en un 80 %, 85 %, 90 %, 95 % o 98 % a la SEQ ID NO:8. En una realización específica la invención se refiere a una señal de poliadenilación que comprende un ácido nucleico que comprende una secuencia idéntica al menos en un 85 % a la SEQ ID NO:8. En otra realización específica la invención se refiere a una señal de poliadenilación que comprende un ácido nucleico que comprende una secuencia idéntica al menos en un 95 % a la SEQ ID NO:8.
- En una realización preferida se aísla dicha señal de poliadenilación. En otra realización preferida la invención se refiere a una señal de poliadenilación aislada que comprende un ácido nucleico que comprende una secuencia idéntica al menos en un 75 % a la SEQ ID NO:8. En otra realización preferida la invención se refiere a una señal de poliadenilación aislada que comprende un ácido nucleico que comprende una secuencia idéntica al menos en un 95 % a la SEQ ID NO:8. Todavía en otra realización preferida la invención se refiere a una señal de poliadenilación aislada que comprende un ácido nucleico que comprende la SEQ ID NO:8.
- Preferiblemente, la invención se refiere a un ácido nucleico aislado cuya secuencia comprende la SEQ ID NO:8.
- En una realización preferida dicha señal de poliadenilación está ligada operativamente a una secuencia de codificación heteróloga. En una realización específicamente preferida dicha señal de poliadenilación se caracteriza porque los títulos / niveles de expresión obtenidos con dicha señal de poliadenilación son al menos un 10 %, preferiblemente 20 % y lo más preferiblemente un 30 % más altos que los obtenidos para la señal de poliadenilación BGH. En la realización más preferida son al menos y/o como promedio un 35 % más altos que los obtenidos para la señal de poliadenilación BGH.
- La invención se refiere además a un ácido nucleico cuya secuencia comprende la SEQ ID NO:8 ligada operativamente a una secuencia de codificación heteróloga. Alternativamente, dicha secuencia consiste esencialmente en la SEQ ID NO:8 ligada operativamente a una secuencia de codificación heteróloga. Preferiblemente, dicha secuencia consiste en la SEQ ID NO:8 ligada operativamente a una secuencia de codificación heteróloga.
- Un ácido nucleico cuya secuencia comprende la SEQ ID NO: 8 y tiene función de terminación. Preferiblemente dicho ácido nucleico tiene función de terminador y está ligado operativamente a una secuencia de codificación heteróloga.
- La invención se refiere además a un vector que comprende una cualquiera de las señales de poliadenilación o secuencias de ácido nucleico descritas anteriormente. En una realización específica dichas señales de poliadenilación están ligadas operativamente a una unidad de expresión / casete de expresión. En otra realización de la invención el vector comprende el marcador de selección y/o de amplificación dihidrofolato reductasa (DHFR), glutamina sintetasa o neomicina fosfatasa (neo). En una realización preferida de la presente invención el vector comprende un gen heterólogo de interés que codifica un producto heterólogo de interés. Preferiblemente dicho producto es un polipéptido. Preferiblemente dicho polipéptido es un anticuerpo, fragmento de anticuerpo o proteína de fusión.
- La invención se refiere adicionalmente a una célula que comprende uno cualquiera de los vectores descritos antes. Preferiblemente, la célula comprende una cualquiera de las señales de poliadenilación descritas antes ligadas operativamente a una unidad de transcripción que codifica un producto de interés. Preferiblemente dicho producto de interés es un nucleótido / ácido nucleico de interés. En otra realización de la célula dicho producto de interés es un polipéptido de interés codificado por un gen de interés. Preferiblemente dicho polipéptido es un anticuerpo, fragmento de anticuerpo o proteína de fusión.
- En una realización específica dicha célula es una célula eucariota, una célula de mamífero, una célula de hámster o una célula murina. Preferiblemente dicha célula es una célula de hámster. Más preferiblemente, dicha célula es una célula de ovario de hámster chino (CHO). Lo más preferiblemente dicha célula es una célula CHO DG44, CHO-K1 o DUKX-B11. En otra realización preferida dicha célula es una célula NSO. En una realización preferida dichas células descritas son células cultivadas. Preferiblemente, dichas células se cultivan en medio exento de suero. Preferiblemente dichas células se hacen crecer en cultivo en suspensión. En otra realización preferida de la presente invención la célula se caracteriza porque los títulos / niveles de expresión obtenidos con dicha señal de poliadenilación son al menos un 10 %, preferiblemente un 20 % y lo más preferiblemente un 30 % más altos que los obtenidos con la señal de poliadenilación BGH. En la realización más preferida son al menos y/o como promedio un

35 % más altos que los obtenidos con la señal de poliadenilación BGH. Preferiblemente, dicha célula tiene niveles de expresión un 35 % más altos.

La invención se refiere adicionalmente a un método para preparar un polipéptido de interés codificado por un gen de interés, comprendiendo el método:

- 5 (a) Proporcionar una célula hospedante que comprende un vector como se ha descrito antes o proporcionar una célula como se ha descrito antes,
- (b) Cultivar dichas células, en condiciones que permitan la proliferación de las células y la expresión del gen de interés,
- (c) Recoger el polipéptido de interés y
- 10 (d) Purificar el polipéptido de interés.

En una realización específica de dicho método la célula es una célula eucariota, una célula de mamífero, una célula de hámster o una célula murina. Preferiblemente dicha célula es una célula CHO, lo más preferiblemente una célula CHO DG44, CHO-K1 o DUKX-B11. La más preferida es una célula NSO.

- 15 En una realización preferida de dicho método el polipéptido de interés es una proteína recombinante, preferiblemente un polipéptido secretado, más preferiblemente una proteína terapéutica. Lo más preferiblemente el polipéptido de interés es un anticuerpo, tal como un anticuerpo monoclonal, policlonal, multiespecífico o de cadena sencilla, o un fragmento del mismo, p.ej. fragmentos Fab, Fab', F(ab')₂, Fc y Fc', cadenas de inmunoglobulinas pesadas y ligeras y sus regiones constantes, variables o hipervariables así como los fragmentos Fv y Fd. En otra realización preferida de dicho método el polipéptido de interés es una proteína de fusión o una proteína de almacén.

- 20 La invención se refiere además al uso de la célula como se ha descrito antes para la fabricación de proteínas.

La invención se refiere también al uso de una cualquiera de las señales de poliadenilación como se han descrito antes para uso como un aislante.

Adicionalmente, la invención se refiere al uso de una cualquiera de las señales de poliadenilación como se han descrito antes para la generación de líneas mejoradas de células hospedantes.

- 25 La invención se refiere específicamente al uso de una cualquiera de las señales de poliadenilación como se han descrito antes para uso en terapia génica.

La invención se refiere además a un kit que comprende una cualquiera de las señales de poliadenilación como se han descrito antes, un vector, una célula y un medio de cultivo celular para cultivo de dicha célula.

Descripción de las figuras

- 30 Figura 1: Vectores de expresión básicos

La Figura 1 presenta esquemáticamente diseños de vectores de expresión utilizados para la transfección de células CHO-DG44. "P/E" significa una unidad compuesta que contiene ambos el elemento potenciador y el elemento promotor de CMV, "P" significa un elemento promotor y "T" una señal de terminación de la transcripción, que se necesitan para la poliadenilación del ARN mensajero transcrito. Las señales de poliadenilación "BGH", "SV40L" y "HGH" son señales de terminación de la transcripción derivadas de la región 3' no traducida de la hormona de crecimiento bovina (SEQ ID NO:12), de la región génica SV40 tardía (SEQ ID NO:11) y de la región 3' no traducida de la hormona de crecimiento de hámster chino (SEQ ID NO:8), respectivamente. Estas señales de poliadenilación están flanqueadas por los sitios de la enzima de restricción para "Sfil" y "XbaI". La posición y dirección de la iniciación de la transcripción dentro de cada unidad de transcripción están indicadas por una flecha. Para la clonación del gen de interés se inserta una región de secuencia con múltiples sitios de corte para las endonucleasas de restricción (sitios múltiples de clonación - "mcs") después del elemento promotor/potenciador. El marcador seleccionable amplificable dihidrofolato reductasa se abrevia como "dhfr" y el marcador seleccionable neomicina fosfotransferasa se abrevia como "npt".

Figura 2: Región aislada del gen de la hormona de crecimiento de *Cricetus griseus*

- 45 La Figura 2 presenta la secuencia de nucleótidos de la región del gen de la hormona de crecimiento que se amplificó a partir de ADN genómico de CHO-DG44 (línea de células de ovario de hámster chino; *Cricetus griseus*) utilizando PCR anidada con un total de 362 bp (SEQ ID NO:7). La flecha indica la dirección, longitud y posición del cebador específico del gen GH for2 utilizado en la reacción de amplificación, la secuencia del propio cebador está remarcada

en letra cursiva (SEQ ID NO:2). El codón de terminación TAG de la secuencia del gen de la hormona de crecimiento está remarcado por letras en negrita subrayadas y va seguido por 324 bp de la región 3' no traducida.

Figura 3: Alineamiento de las regiones 3' no traducidas de los genes de la hormona de crecimiento

5 En este alineamiento la región 3' no traducida aislada de la hormona de crecimiento de *Cricetus griseus* (SEQ ID NO:8) se compara con la región 3' no traducida de la hormona de crecimiento del hámster sirio *Mesocricetus auratus* (Genbank S66299), *Mus musculus* (Genbank Z46663), *Rattus norvegicus* (Genbank V01239) y *Bos taurus* (Genbank J00008). El sombreado indica los nucleótidos que difieren de la secuencia de *C. griseus*.

Figura 4: Derivados por delección de HGH de la región 3' no traducida de la hormona de crecimiento de *Cricetus griseus*

10 En este alineamiento se muestran los derivados por delección de los 362 nucleótidos de la secuencia de la hormona de crecimiento (HGH) de *Cricetus griseus* (SEQ ID NO:7) que contienen sólo la región 3' no traducida. Todos los derivados tienen un extremo 5' idéntico y difieren en su extremo 3'. El derivado más largo con SEQ ID NO:8 consiste en 324 nucleótidos. La SEQ ID NO:9 consiste en 189 nucleótidos y la SEQ ID NO:10 en sólo 113 nucleótidos. El codón de terminación TAG de la secuencia del gen de la hormona de crecimiento está remarcado por letras en negrita subrayadas y el sitio de unión potencial del complejo de la proteína de poliadenilación AATAAA está remarcado en letra cursiva.

Figura 5: Vectores de expresión recombinantes para evaluación del comportamiento de la HGH

20 Todos los vectores de expresión recombinantes codifican el gen de interés "sICAM" bajo el control del elemento potenciador y promotor ("P/E") de CMV. La transcripción de sICAM se termina o por la región 3' no traducida de la hormona de crecimiento bovina "BGH" (SEQ ID NO:12), o por la región del gen SV40 tardía "SV40L" (SEQ ID NO:11) o por la región 3' no traducida de la hormona de crecimiento de hámster chino "HGH" (SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10). Se indica el tamaño de la última en pares de bases. Estas señales de poliadenilación están flanqueadas por los sitios de la enzima de restricción para "SfiI" y "XbaI". "P" indica un elemento promotor y "T" una señal de terminación de la transcripción. La posición y dirección del inicio de la transcripción dentro de cada unidad de transcripción están indicadas por una flecha. El marcador seleccionable amplificable dihidrofolato reductasa se abrevia como "dhfr".

Figura 6: Evaluación del comportamiento de la HGH en transfecciones transitorias

30 En dos series independientes, se transfectan células CHO-DG44 con vectores de expresión pJR106, pJR110 y pJR131 todos los cuales codifican sICAM bajo el potenciador/promotor de CMV. Para la terminación de la transcripción se utilizan o la señal de poliadenilación de SV40 tardía (SEQ ID NO:11), o la región 3' no traducida de la hormona de crecimiento bovina BGH (SEQ ID NO:12) o la región 3' no traducida de la hormona de crecimiento de hámster chino HGH (SEQ ID NO:8). Después de un período de 48 horas se determinan los títulos de sICAM en los sobrenadantes utilizando ELISA. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP y se mide la actividad de SEAP. Utilizando HGH como señal de poliadenilación se aumenta el título hasta un 21 % en comparación a la terminación con BGH y hasta un 40 % en comparación a la terminación con SV40 tardía.

Figura 7: Evaluación del comportamiento de la HGH en la expresión transitoria de un anticuerpo IgG4/Kappa

40 Se co-transfectan células CHO-DG44 con la combinación de vectores pBID/IgG4 y pBIN/kappa (n=6) en la cual la transcripción de la cadena pesada (IgG4) y de la cadena ligera (kappa) del anticuerpo está terminada por la región 3' no traducida de 324 bp de la hormona de crecimiento del hámster chino HGH (SEQ ID NO:8). Como control se co-transfectan células CHO-DG44 con la combinación de vectores pBID-B/IgG4 y pBIN-B/kappa (n=6) que contienen la señal de poliadenilación BGH (SEQ ID NO:12). Aparte de las diferentes secuencias de poliadenilación la organización genética de los diferentes vectores es idéntica. Después de un período de 48 horas se determinan los títulos de anticuerpo en los sobrenadantes utilizando ELISA. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP y se mide la actividad de SEAP. Utilizando HGH como señal de poliadenilación los títulos son como promedio un 35 % más altos que para las señales de poliadenilación BGH.

Figura 8: Ensayo de diferentes variantes de delección de HGH en transfecciones transitorias

50 En dos series independientes, se transfectan células CHO-DG44 con vectores de expresión pJR131, pJR134 y pJR135 todos los cuales codifican sICAM bajo el potenciador/promotor de CMV. Para terminación de la transcripción se utilizan o 324 bp (SEQ ID NO:8), o 189 bp (SEQ ID NO:9) o 113 bp (SEQ ID NO:10) de la región 3' no traducida de la hormona de crecimiento de hámster chino HGH. Todas las variantes tienen un extremo 5' idéntico pero difieren en su extremo 3'. Aparte de las diferentes secuencias de poliadenilación la organización genética de los diferentes

5 vectores es idéntica. 48 horas después de las transfecciones se determinan los títulos de sICAM en los sobrenadantes utilizando ELISA. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP y se mide la actividad de SEAP. En comparación con las células transfectadas con vectores que contienen la secuencia de HGH de 324 bp, las células transfectadas con vectores que contienen las variantes de delección de HGH de 189 bp y de 113 bp presentan una reducción en los niveles de expresión de sICAM de 23 % y 78 %, respectivamente.

Figura 9: Expresión de altos niveles de proteína en células transfectadas estables utilizando HGH

10 En la Figura 9 se resumen las productividades específicas y los títulos de clones de células o conjuntos de células CHO-DG44 transfectadas establemente que expresan anticuerpos IgG1, IgG2 y IgG4 o proteínas de fusión IgG1 y IgG2 Fc en procesos de lote alimentado realizados en biorreactores o en matraces de agitación. Las productividades específicas están en el intervalo de 10-45 pg/célula/día y los títulos están en el intervalo de 2,1-6,3 g/L. La organización genética de los vectores utilizados para la expresión de las diferentes proteínas es idéntica. Todos contienen la región 3' no traducida de 324 bp de la hormona de crecimiento de hámster chino (SEQ ID NO:8) como señal de poliadenilación para terminar la transcripción del gen de interés. 2 días después de la transfección se seleccionan conjuntos de células estables utilizando una selección basada en DHFR y NPT seguida por 2 etapas sucesivas de amplificación del gen mediada por DHFR mediante la adición de MTX 100 nM y 800 nM al medio de cultivo. Se obtienen clones de células individuales por clonación por dilución o por sedimentación basada en FACS de las células individuales en los pocillos de una placa de 96 pocillos.

Descripción detallada de la invención

20 Las realizaciones generales "que comprende" o "comprende" incluyen la realización más específica "que consiste en". Además, las formas singular y plural no se utilizan de modo limitante.

25 La presente invención proporciona nuevos elementos reguladores y métodos para preparar y seleccionar líneas de células de mamífero que permiten una alta expresión de productos génicos heterólogos, preferiblemente polipéptidos o proteínas biofarmacéuticamente relevantes. Los procedimientos según la invención se basan principalmente en el uso de nuevas señales de poliadenilación aisladas de la hormona de crecimiento del hámster chino (*Cricetus griseus*). Sorprendentemente, se ha encontrado que esta señal de poliadenilación recién identificada, denominada HGH (SEQ ID No:8), aventaja a las fuertes señales de poliadenilación BGH y SV40 tardía llevando a una productividad más alta de las células productoras.

Los términos usados a lo largo de esta presente invención tienen el siguiente significado.

30 Los términos "señal de poliadenilación", "sitio de poliadenilación", "señal poliA", "sitio poliA" o "señal de terminación" o "terminador" se refieren a secuencias de nucleótidos dentro de la 3'UTR que dirigen la unión de un complejo de proteínas de poliadenilación con una secuencia AAUAAA dentro de la secuencia señal. El complejo contiene una endonucleasa que corta el ARNm aproximadamente 14 a 30 nucleótidos corriente abajo de la secuencia AAUAAA y una polimerasa que incorpora post-transcripcionalmente una cadena de aproximadamente 100 a 200 nucleótidos de adenina (cola poliA) al extremo 3' escindido. Se cree que la cola poliA influye en muchos aspectos del metabolismo de ARNm, incluyendo la estabilidad, la eficiencia de traducción, y el transporte del núcleo al citoplasma. Típicamente, la señal de poliadenilación consiste en dos elementos de reconocimiento que flanquean el sitio de escisión y de poliadenilación: una secuencia AAUAAA altamente conservada de aproximadamente 14 a 30 nucleótidos corriente arriba del sitio de escisión y una región rica en G/U o en U pobremente conservada de aproximadamente 20 a 50 nucleótidos corriente abajo de la secuencia AAUAAA. La escisión entre estos dos elementos usualmente tiene lugar en el lado 3' de un residuo A. Se conocen varias señales de poliadenilación tales como tk poliA (Cole et al., Mol. Cell. Biol., 5, 2104-2113, 1985), SV40 tardía (Schek et al., Mol. Cell. Biol. 12, 5386-5393, 1992) y poliA temprana o poliA de BGH (descritas por ejemplo en la patente de Estados Unidos N° 5.122.458).

45 Aunque en la señal de poliadenilación se prefiere la secuencia AAUAAA descrita antes, ésta puede ser sustituida con otras secuencias de hexanucleótidos que tienen homología con la AAUAAA siempre que sean capaces de señalar la poliadenilación de los ARNm. Los ejemplos de secuencias de hexanucleótidos homólogas incluyen AAAAAA, AUUAAA, AAUAUA, AAUAUA, UAUAAA, AAUUAA, AAUAAG, AGUAAA, GAUAAA, AAUGAA, AAUAGA, AAGAAA, ACUAAA, CAUAAA, AAUCAA, AACAAA, AAUCAA, y AAUAAC. Por lo tanto, en una realización la señal de poliadenilación de HGH comprende una secuencia de hexanucleótidos seleccionada del grupo que consiste en AAUAAA, AUUAAA, AAUAUA, AAUAUA, UAUAAA, AAUUAA, AAUAAG, AGUAAA, GAUAAA, AAUGAA, AAUAGA, AAGAAA, ACUAAA, CAUAAA, AAUCAA, AACAAA, AAUCAA, y AAUAAC en lugar de la presente AAUAAA siempre que estos hexanucleótidos sean capaces de señalar la poliadenilación de los ARNm.

55 Las señales de poliadenilación se pueden usar también como "aislantes" o "secuencias aisladoras". Las secuencias aisladoras son segmentos de ADN que bloquean las interacciones o interferencia de las secuencias génicas vecinas. Por ejemplo, las secuencias aisladoras pueden reducir la lectura transcripcional a través de un promotor de un gen vecino o promotores espurios en secuencias de nucleótidos adyacentes. Es decir, bloquean la interacción de

- un potenciador en un lado de la secuencia aisladora con un promotor de un gen vecino en el otro lado de la secuencia aisladora. La característica que define una secuencia aisladora dentro del significado de la presente invención es su capacidad para aislar o proteger una unidad de transcripción definida que está ligada operativamente a un elemento regulador a partir de la influencia de un elemento genético que interfiere corriente arriba o corriente abajo. Para este fin la secuencia aisladora está colocada entre la secuencia genética que interfiere (potencial) y la secuencia reguladora de la unidad de transcripción a ser aislada. La secuencia aisladora puede estar colocada en cualquiera o en ambos lados de la unidad de transcripción en una o más copias. En una realización preferida de la presente invención la secuencia aisladora es una señal de poliadenilación. En una realización preferida de esta invención la secuencia de poliadenilación es la secuencia de poliadenilación HGH.
- 5 Es posible también usar derivados funcionales de la secuencia de poliadenilación HGH tales como subfragmentos o subsecuencias así como también mutantes/variantes funcionales de la secuencia completa o subfragmentos de la misma que han sido modificados, por ejemplo, por sustitución, inserción, adición y/o delección. Los subfragmentos o subsecuencias, mutantes o variantes correspondientes se denominan también de aquí en adelante “terminadores modificados” o “derivados”.
- 10 Un “terminador modificado” o “derivado” es un derivado funcional de la SEQ ID NO:8, que incluye subfragmentos o subsecuencias y mutantes/variantes funcionales, y preferiblemente lleva a niveles de expresión de un producto de interés comparables con los niveles de expresión obtenidos con la secuencia de nucleótidos dada en la SEQ ID NO:8. Un terminador modificado demuestra ser útil para los fines de la invención si el nivel de expresión de un gen informador ligado operativamente es al menos el 60 %, preferiblemente al menos el 75 %, más preferiblemente al menos el 90 % y lo más preferiblemente al menos el 100 % del nivel de expresión obtenido con la SEQ ID NO:8 en un ensayo comparativo del gen informador. Son particularmente preferidos los terminadores modificados que tienen una homología mínima de secuencia con la secuencia de tipo natural SEQ ID NO:8 de la señal de poliadenilación de la hormona de crecimiento de hámster o su secuencia complementaria de al menos 75 %, preferiblemente al menos 80 %, preferiblemente al menos 85 %, más preferiblemente al menos 95 % y lo más preferiblemente al menos 97 % y llevan a niveles de expresión correspondientes en un ensayo comparativo del gen informador.
- 15 20 25 En un “ensayo del gen informador” comparativo correspondiente los fragmentos de terminación a ser analizados incluyendo la secuencia de referencia SEQ ID NO:8 están clonados corriente abajo de un gen informador. Este gen informador codifica, por ejemplo la luciferasa, la fosfatasa alcalina secretada o la proteína fluorescente verde (GFP). Alternativamente, se pueden usar como genes informadores otros polipéptidos o proteínas, por ejemplo un anticuerpo o sICAM. Estas construcciones se introducen a continuación en las células de ensayo, p.ej. CHO-DG44, mediante transfección y se determina la influencia del terminador modificado en cuestión sobre el nivel de expresión del gen informador por ejemplo midiendo el contenido en proteína del gen informador. Un ensayo correspondiente se describe en los ejemplos 2, 3 y 4 de la presente invención.
- 30 Una señal preferida de poliadenilación de HGH es la secuencia de nucleótidos que comprende la secuencia de SEQ ID NO:8 o una subsecuencia de la misma que comprende la secuencia de SEQ ID NO: 9. En otras realizaciones, la señal de poliadenilación es una secuencia de nucleótidos que comprende o consiste en una secuencia de nucleótidos con homología o identidad de secuencia con las SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 o SEQ ID NO:10. Como se usa aquí, dos secuencias tienen identidad u homología de secuencia cuando las secuencias de nucleótidos son homólogas o idénticas en al menos un 75 %, preferiblemente un 80 %, preferiblemente un 85 %, más preferiblemente un 90 %, y aún más preferiblemente un 95 % o más. La identidad sustancial existe también cuando la secuencia de ácido nucleico se hibrida en condiciones restrictivas con el complemento de la cadena.
- 35 40 Como se usa aquí, el término “se hibrida en condiciones restrictivas” describe las condiciones de hibridación y lavado que son conocidas por los expertos en la técnica. Generalmente, se seleccionan condiciones restrictivas que sean de aproximadamente 5-10 °C más bajas que el punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos. La T_m es la temperatura (a una fuerza iónica, pH y concentración de ácido nucleico definidos) a la cual el 50 % de las sondas complementarias a la diana se hibridan con la secuencia diana en equilibrio. Las condiciones restrictivas serán aquellas en las que la concentración salina es inferior a aproximadamente 1,0 M de ion sodio, típicamente aproximadamente 0,01 a 1,0 M de concentración de ion sodio (u otras sales) a pH 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30 °C para sondas cortas (p.ej. 10 a aproximadamente 50 nucleótidos) y de al menos aproximadamente 60 °C para sondas largas (p.ej. mayores que aproximadamente 50 nucleótidos). Los ejemplos de condiciones restrictivas incluyen la hibridación a 60 a 65 °C en un tampón de hibridación con 5×SSC y el lavado a 42 °C con 0,2×SSC/0,1% de SDS. Una señal de hibridación positiva está al menos 2 veces por encima de la hibridación básica.
- 45 50 55 La secuencia de poliadenilación de la hormona de crecimiento del hámster y los terminadores modificados, que pueden incluir también, por ejemplo las secuencias de nucleótidos naturales que están corriente arriba o corriente abajo de la secuencia de HGH aislada de la SEQ ID NO:7 o fragmentos seleccionados de la misma, puede ser obtenida por un experto en la técnica con un conocimiento de la secuencia o secuencias homólogas utilizando diversos métodos estándar conocidos en la técnica y en la presente invención en el ejemplo 1 se describe también

un método adecuado. Partiendo de la secuencia descrita en la SEQ ID NO:7 se puede seleccionar un fragmento adecuado, por ejemplo, y se puede sintetizar químicamente una sonda de oligonucleótidos que contiene la secuencia de esta fracción. Se puede usar una sonda de este tipo, por ejemplo para clonar el gen de la hormona de crecimiento de hámster o la región 3' no traducida u otros fragmentos de la misma, por ejemplo mediante hibridación a partir de una genoteca del genoma de hámster. Utilizando el ensayo del gen informador descrito antes el experto para la técnica está en situación de identificar los fragmentos de terminación funcionales sin gran esfuerzo y utilizarlos para los fines de la presente invención. La región 3' no traducida o fragmentos especiales de la misma se pueden obtener fácilmente por amplificación por PCR con los correspondientes cebadores a partir del ADN genómico o de una genoteca. Se pueden obtener también fragmentos de la región 3' no traducida mediante digestión limitada con exonucleasa III a partir de fragmentos de ADN más grandes. Tales moléculas de ADN se pueden sintetizar también químicamente o producir a partir de fragmentos sintetizados químicamente mediante ligamiento. Se pueden producir también mutantes de delección, inserción, adición y sustitución por mutagénesis específica del sitio, técnicas de mutagénesis basadas en la PCR y/o métodos de síntesis química conocidos por los expertos en la técnica. Preferiblemente, un mutante está alterado en posiciones de hasta 3, 6, 10, 20 o 50 bp. Preferiblemente un mutante está alterado en las posiciones de 6 bp.

Un método similar como se describe en la presente invención en el ejemplo 1 se puede usar para aislar por ejemplo las señales de poliadenilación de la hormona de crecimiento de ratón, rata o hámster sirio o de hormonas de crecimiento de otras especies. Se puede analizar su comportamiento en ensayos del gen informador como se describe en los ejemplos 2, 3 o 4 de la presente invención. Mediante hibridación cruzada con sondas derivadas de la secuencia de la hormona de crecimiento de hámster, preferiblemente de la región 3' no traducida, es posible también identificar y aislar secuencias terminadoras aisladas a partir de los correspondientes genes homólogos de otras especies, preferiblemente mamíferos. Los métodos adecuados son conocidos por los expertos en la técnica.

Los términos "homología", "homólogo", "identidad", "idéntico", "identidad de secuencia" o "secuencia homóloga" se utilizan de modo intercambiable. Los métodos para calcular la "homología" o "identidad" son bien conocidos en la técnica. Para comparación de secuencias típicamente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de ensayo. Las secuencias se alinean con una correspondencia máxima. Se pueden introducir huecos en cualquiera de las secuencias de ácido nucleico en la comparación para un alineamiento óptimo. El porcentaje de identidad entre dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco que es necesario introducir para el alineamiento óptimo de las dos secuencias. La comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias se puede realizar utilizando algoritmos matemáticos. Se pueden usar parámetros por defecto de un programa o se pueden designar parámetros alternativos. El algoritmo de comparación de secuencias calcula entonces el porcentaje de identidad para la secuencia o secuencias de ensayo con respecto a la secuencia de referencia, basado en los parámetros designados o en los parámetros por defecto de un programa. Un ejemplo de un algoritmo que es adecuado para determinar la identidad es el algoritmo BLAST (Altschul et al., J. Mol. Biol. 215, 403-410, 1990; Gish et al., Nature Genetics 3, 266-272, 1993; Madden et al., Meth. Enzymol. 266, 131-141, 1996; Zhang et al., Genome Res. 7, 649-656, 1997; Altschul et al., Nucleic Acids Res. 25, 3389-3402, 1997). Otras aplicaciones de ordenador de algoritmos de alineamiento son GAP, PILEUP, BESTFIT, FASTA y TFASTA en el paquete de software de Wisconsin Genetics. Sin embargo, el porcentaje de identidad se puede determinar también por alineamiento manual e inspección visual y cálculo.

El término "vector" como se usa aquí, se refiere a construcciones presentes en la naturaleza o generadas sintéticamente para absorción, proliferación, expresión o transmisión de ácidos nucleicos en una célula, p.ej. plásmidos, minicírculos, fagémidos, cósmidos, cromosomas artificiales/mini-cromosomas, bacteriófagos, virus tales como baculovirus, retrovirus, adenovirus, virus adeno-asociados, virus del herpes simple, bacteriófagos. Los métodos usados para construir vectores son bien conocidos por los expertos en la técnica y están descritos en varias publicaciones. En particular las técnicas para construir vectores adecuados, incluyendo una descripción de los componentes funcionales y reguladores tales como promotores, potenciadores, señales de terminación y de poliadenilación, marcadores de selección, orígenes de replicación, y señales de corte y empalme, son conocidas por los expertos en la técnica. Los vectores de expresión eucariotas contendrán típicamente también secuencias procariotas que facilitan la propagación del vector en bacterias tal como genes de origen de la replicación y de resistencia a antibióticos para selección en bacterias. Una variedad de vectores de expresión eucariotas, que contienen un sitio de clonación al cual puede estar ligado operativamente un polinucleótido, son bien conocidos en la técnica y algunos están comercialmente disponibles de compañías tales como Stratagene, La Jolla, CA; Invitrogen, Carlsbad, CA; Promega, Madison, WI o BD Biosciences Clontech, Palo Alto, CA.

Una realización preferida de la invención son los vectores o secuencias de polinucleótidos que contienen una o más unidades de transcripción que codifican genes de interés que comprenden al menos una señal de poliadenilación HGH para la terminación y estabilización del transcrito y/o como secuencia aisladora. También se prefieren según la invención vectores o secuencias de polinucleótidos que comprenden señales de poliadenilación HGH para la terminación y estabilización del transcrito y/o como secuencia aisladora que en lugar de los genes de interés tienen

solamente un sitio de clonación múltiple que permite la clonación del gen de interés mediante secuencias de reconocimiento para las endonucleasas de restricción.

El término “promotor” indica una secuencia de polinucleótidos que permite y controla la transcripción de los genes o de las secuencias conectadas operativamente con ellos. Un promotor contiene secuencias de reconocimiento para la unión de la ARN polimerasa y el sitio de iniciación de la transcripción (sitio de inicio de la transcripción). Con el fin de expresar una secuencia deseada en un cierto tipo de célula o en una célula hospedante se debe elegir un promotor funcional adecuado. Un gran número de promotores, incluyendo los promotores constitutivos, inducibles y reprimibles de una variedad de diferentes fuentes, son bien conocidos en la técnica (y están identificados en bases de datos tales como GenBank) y se encuentran disponibles como elementos separados o elementos clonados dentro de secuencias de polinucleótidos de fuentes comerciales (p.ej. depósitos tales como ATCC así como otras fuentes comerciales) o fuentes individuales. En los promotores inducibles la actividad del promotor se puede aumentar o reducir en respuesta a una señal. Por ejemplo, el promotor de tetraciclina (tet) que contiene la secuencia del operador de la tetraciclina (tetO) puede ser inducido por una proteína transactivadora regulada por tetraciclina (tTA). La unión de la tTA al tetO se inhibe en la presencia de tet. Son ejemplos de otros promotores inducibles jun, fos, metalotioneina y los promotores de shock térmico. De los promotores que son particularmente adecuados para una alta expresión en eucariotas, están por ejemplo el promotor de ubiquitina/S27a del hámster (documento WO 97/15664), el promotor temprano de SV40, el promotor tardío principal de adenovirus, el promotor de metalotioneina I de ratón, la región de repetición terminal larga del virus de sarcoma de Rous, el promotor temprano de Citomegalovirus humano (CMV). Son ejemplos de otros promotores heterólogos de mamífero la actina, la inmunoglobulina o los promotores de shock térmico.

Los promotores mencionados antes son bien conocidos en la técnica. Un promotor heterólogo correspondiente se puede conectar funcionalmente a otras secuencias reguladoras con el fin de aumentar/regular la actividad de transcripción en una casete de expresión. Por ejemplo, el promotor se puede ligar funcionalmente a secuencias potenciadoras con el fin de aumentar la actividad transcripcional. Para esto, se pueden usar uno o más potenciadores y/o varias copias de una secuencia potenciadora, p.ej. un potenciador CMV o SV40. Por consiguiente, un vector de expresión según la invención, en otra realización, contiene una o más secuencias potenciadoras/mejoradoras, preferiblemente un potenciador CMV o SV40.

El término “potenciador” indica una secuencia de polinucleótidos que en la posición *cis* actúa sobre la actividad de un promotor y estimula de este modo la transcripción de un gen o secuencia de codificación conectada funcionalmente a este promotor. A diferencia de los promotores el efecto de los potenciadores es independiente de la posición y orientación y por lo tanto pueden estar posicionados delante o detrás de una unidad de transcripción, dentro de un intrón o incluso dentro de la región codificadora. El potenciador puede estar localizado tanto en la vecindad inmediata de la unidad de transcripción como a una distancia considerable del promotor. Es posible también que tenga un solapamiento físico y funcional con el promotor. Los expertos en la técnica están al tanto de un número de potenciadores de diferentes fuentes (y depositados en genotecas tales como GenBank, p.ej. potenciadores de SV40, potenciadores de CMV, potenciadores de polioma, potenciadores de adenovirus) que están disponibles como elementos independientes o elementos clonados dentro de secuencias de polinucleótidos (p.ej. depositados en la ATCC o de fuentes comerciales e individuales). Una serie de secuencias promotoras contienen también secuencias potenciadoras tales como el promotor de CMV utilizado frecuentemente. El potenciador de CMV humano es uno de los potenciadores más fuertes identificados hasta ahora. Un ejemplo de un potenciador inducible es el potenciador de metalotioneina, que puede ser estimulado por glucocorticoides o metales pesados.

Los “elementos reguladores de la transcripción” normalmente comprenden un promotor corriente arriba de la secuencia del gen a expresar, sitios de iniciación y terminación de la transcripción y una señal de poliadenilación.

El término “sitio de iniciación de la transcripción” se refiere a un ácido nucleico de la construcción que corresponde al primer ácido nucleico incorporado en el transcrito primario, esto es, el precursor de ARNm. El sitio de inicio de la transcripción puede solaparse con las secuencias del promotor.

El término “sitio de terminación de la transcripción” se refiere a una secuencia de nucleótidos normalmente representada en el extremo 3' del gen de interés o del tramo de secuencias a transcribir, que provoca que la ARN polimerasa termine la transcripción.

Una “unidad de transcripción”, “unidad de expresión” o “casete de expresión” define una región dentro de un vector, construcción o secuencia de polinucleótidos que contiene uno o más genes a transcribir, en donde los genes contenidos dentro del segmento están ligados operativamente uno con otro. Ellos se transcriben a partir de un único promotor y la transcripción se termina por al menos una señal de poliadenilación. Como resultado, los diferentes genes se encuentran ligados al menos transcripcionalmente. Más de una proteína o producto se pueden transcribir y expresar a partir de cada unidad de transcripción (unidad de transcripción multicistriónica). Cada unidad de transcripción comprenderá los elementos reguladores necesarios para la transcripción y traducción de cualquiera de las secuencias seleccionadas que están contenidas dentro de la unidad. Y cada unidad de transcripción puede

contener los mismos elementos reguladores o diferentes. Por ejemplo, cada unidad de transcripción puede contener el mismo terminador. Se pueden usar el elemento IRES o los intrones para la unión funcional de los genes con una unidad de transcripción. Un vector o secuencia de polinucleótidos puede contener más de una unidad de transcripción.

5 Los “elementos reguladores de la traducción” comprenden un sitio de iniciación de la traducción (AUG), un codón de terminación y una señal poliA para cada polipéptido individual a ser expresado. En algunas construcciones se puede incluir un sitio interno de entrada al ribosoma (IRES). El IRES se define más adelante. Con el fin de optimizar la expresión puede ser aconsejable separar, añadir o alterar las regiones 5' y/o 3' no traducidas de la secuencia de ácido nucleico a expresar para eliminar todos los codones de inicio de la traducción alternativos inapropiados potencialmente extra u otras secuencias que puedan interferir o reducir la expresión, ya sea a nivel de transcripción o de traducción. Se pueden insertar sitios de consenso de unión al ribosoma (secuencia de Kozak: GCCGCCACCAUGG (SEQ ID NO:13); AUG constituye el codón de inicio) inmediatamente corriente arriba del codón de inicio para mejorar la traducción y de este modo la expresión. El aumento de los contenidos de A/U alrededor de este sitio de unión al ribosoma favorece una unión más eficiente al ribosoma. Para producir un polipéptido secretado el gen de interés usualmente incluye una secuencia señal que codifica un péptido líder o péptido señal que dirige el polipéptido recién sintetizado a la membrana ER y a través de ella, donde el polipéptido puede ser enviado para secreción. El péptido líder o señal está a menudo pero no siempre en el terminal amino de una proteína secretada y se separa por escisión mediante peptidasas señal una vez que la proteína cruza la membrana ER. La secuencia génica contiene generalmente, pero no necesariamente, su propia secuencia de péptido señal. Cuando está ausente la secuencia de péptido señal nativo, se puede fusionar una secuencia de péptido señal heterólogo con la secuencia seleccionada. Por tanto la secuencia del péptido señal nativo puede ser reemplazada por una heteróloga. Numerosas secuencias de péptido señal son conocidas por los expertos en la técnica y están depositadas en genotecas de secuencias tales como GenBank y EMBL.

25 Un “sitio interno de entrada al ribosoma” o “IRES” describe una secuencia que promueve funcionalmente el inicio de la traducción con independencia del gen 5' del IRES y permite que dos cistrones (marcos de lectura abiertos) sean traducidos a partir de un único transcrito en una célula animal. El IRES proporciona un sitio de entrada independiente al ribosoma para la traducción del marco de lectura abierto situado inmediatamente corriente abajo del mismo. A diferencia del ARNm bacteriano que puede ser policistrónico, esto es, que codifica varios polipéptidos diferentes que se traducen secuencialmente a partir de los ARNm, la mayor parte de los ARNm de células animales son monocistrónicos y codifican la síntesis de un solo polipéptido. Con un transcrito policistrónico en una célula eucariota, la traducción se iniciará a partir del sitio de inicio 5' de la mayor parte de la traducción, terminará en el primer codón de terminación, y el transcrito será liberado del ribosoma, dando como resultado la traducción de solamente el primer polipéptido codificado en el ARNm. En una célula eucariota, un transcrito policistrónico que tiene un IRES ligado operativamente al segundo o posterior marco de lectura abierto en el transcrito permite la traducción secuencial de dicho marco de lectura abierto corriente abajo para producir los dos o más polipéptidos codificados por el mismo transcrito. El IRES puede ser de longitud variable y proceder de diferentes fuentes, p.ej. virus de la encefalomiocarditis (EMCV), picornavirus (p.ej. FMDV), polio virus (PV), o virus de la hepatitis C (HCV). Se han descrito varias secuencias de IRES y su uso en la construcción de vectores y son bien conocidas en la técnica. La secuencia de codificación corriente abajo está ligada operativamente al extremo 3' del IRES a cualquier distancia que no afecte negativamente la expresión del gen corriente abajo. La distancia óptima o permisible entre el IRES y el inicio del gen corriente abajo se puede determinar fácilmente variando la distancia y midiendo la expresión como una función de la distancia.

45 El término “intrón” como se usa aquí, se refiere a una secuencia de ácido nucleico no codificadora de longitud variable, normalmente presente dentro de muchos genes eucariotas, que se separa de un precursor de ARNm recién transcrito mediante el proceso de corte y empalme para el que son necesarias secuencias altamente conservadas en el extremo o cerca del extremo del intrón. En general, el proceso de corte y empalme requiere que los extremos 5' y 3' del intrón estén correctamente escindidos y que los extremos resultantes del ARNm estén exactamente unidos, de tal modo que se produzca un ARNm maduro que tiene el marco de lectura apropiado para la síntesis de proteínas. Han sido caracterizados y descritos y son conocidos por los expertos muchos sitios donadores de empalme y aceptores de empalme, que significan las secuencias que rodean inmediatamente las uniones exón-intrón e intrón-exón.

55 Los términos “gen de interés”, “secuencia deseada”, “polinucleótido de interés” o “gen deseado” como se usan aquí tienen el mismo significado y se refieren a una secuencia de polinucleótidos de cualquier longitud que codifica un producto de interés. La secuencia seleccionada puede ser un gen de longitud completa o un gen truncado, un gen de fusión o un gen marcado, y puede ser un ADNc, un ADN genómico, o un fragmento de ADN. Puede ser la secuencia nativa, esto es en la forma natural, o puede estar mutada o modificada de otra forma según se desee. Estas modificaciones incluyen optimizaciones de codones para optimizar la utilización del codón en la célula hospedante seleccionada, la humanización o el marcado. Además pueden incluir separaciones o adiciones de sitios que actúan en *cis* tales como sitios donadores de empalme (crípticos), sitios aceptores y puntos de ramificación, señales de poliadenilación, cajas TATA, sitios chi, sitios de entrada al ribosoma, secuencias repetidas, estructuras

60

secundarias (p.ej. horquillas (stem loops)), sitios de unión para los factores de transcripción u otros factores reguladores, sitios de la enzima de restricción etc. para dar sólo algunos ejemplos, pero no limitantes. La secuencia seleccionada puede codificar un polipéptido secretado, citoplásmico, nuclear, unido a la membrana o de la superficie celular.

5 Dentro del alcance de la presente descripción los términos “unión funcional”, “ligado funcionalmente” o “ligado operativamente” significan que dos o más secuencias o elementos de secuencia de ácido nucleico están colocados de tal modo que les permite funcionar de la manera que se pretende. Por ejemplo, un promotor/potenciador o terminador está ligado funcionalmente a una secuencia génica codificadora si es capaz de controlar o modular la transcripción de la secuencia génica ligada en la posición *cis*. Generalmente, pero no necesariamente, las
10 secuencias de ADN que están ligadas funcionalmente son contiguas y, cuando es necesario unir dos regiones que codifican el polipéptido o en el caso de un péptido señal de secreción, son contiguas y en marco de lectura. Sin embargo, aunque un promotor ligado operativamente generalmente está localizado corriente arriba o un terminador ligado operativamente generalmente está localizado corriente abajo de la secuencia codificadora, no es necesariamente contiguo a ella. Los potenciadores no tienen que ser contiguos siempre que aumenten la
15 transcripción de la secuencia codificadora. Para esto pueden estar localizados corriente arriba o corriente abajo de la secuencia codificadora e incluso a alguna distancia. Un sitio de poliadenilación está ligado operativamente a una secuencia codificadora si está localizado en el extremo 3' de la secuencia codificadora de modo que la transcripción tiene lugar a través de la secuencia codificadora hacia la señal de poliadenilación. El ligamiento se consigue por métodos recombinantes conocidos en la técnica, p.ej. utilizando la metodología de la PCR, mediante ligación a sitios
20 de restricción adecuados o por anillamiento. Se pueden usar enlaces o adaptadores de oligonucleótidos sintéticos de acuerdo con la práctica convencional si no hay presentes sitios de restricción adecuados.

El término “ácido nucleico”, “secuencia de ácido nucleico”, “secuencia de nucleótidos”, “polinucleótido”, “secuencia de polinucleótidos” o “secuencia de ADN” como se usa aquí se refiere a un oligonucleótido, nucleótido o polinucleótido y fragmentos y porciones de los mismos y a ADN o ARN de origen genómico o sintético, que pueden ser de cadena sencilla o doble y que representan la cadena sentido o la cadena antisentido. La secuencia puede ser una secuencia no codificadora, una secuencia codificadora o una mezcla de ambas. Las secuencias de ácido nucleico de la presente invención se pueden preparar utilizando métodos estándar bien conocidos por los expertos en la técnica.

El término “que codifica” o “codificador/a” se refiere a la propiedad inherente de secuencias específicas de nucleótidos en un ácido nucleico, tal como un gen en los cromosomas o un ARNm, para servir como modelos para la síntesis de otros polímeros y macromoléculas en procesos biológicos que tienen una secuencia definida de nucleótidos (esto es ARNr, ARNt, otras moléculas de ARN) o aminoácidos y las propiedades biológicas que resultan de ellos. Por consiguiente, un gen codifica una proteína si la proteína deseada se produce en una célula o en otro sistema biológico por transcripción y subsiguiente traducción del ARNm. Tanto la cadena codificadora, cuya
30 secuencia de nucleótidos es idéntica a la secuencia de ARNm y usualmente está disponible en los listados de secuencias de las genotecas, p.ej. EMBL o GenBank, como la cadena no codificadora, usada como el modelo para la transcripción, de un gen o ADNc se pueden denominar codificadoras de la proteína o de otro producto de dicho gen o ADNc. Un ácido nucleico que codifica una proteína incluye todos los ácidos nucleicos que tienen diferentes secuencias de nucleótidos pero codifican la misma secuencia de aminoácidos de la proteína debido a la degeneración del código genético. Las secuencias de ácidos nucleicos y nucleótidos que codifican proteínas pueden incluir intrones. En el Listado de secuencias, se presentan las secuencias como secuencias de ADN en lugar de secuencias de ARN. Por ejemplo, cuando se presentan como ADN el codón de inicio se presenta como ATG en lugar de AUG.

El término “ADNc” en el contexto de esta invención se refiere a los ácidos desoxirribonucleicos producidos por transcripción inversa y típicamente síntesis de la segunda cadena de ARNm o de otro ARN producido por un gen. Si tiene doble cadena, una molécula de ADNc tiene tanto una cadena codificadora o cadena sentido como una cadena no codificadora o cadena antisentido.

El término “expresión” como se usa aquí se refiere a la transcripción y/o traducción de una secuencia de ácido nucleico heterólogo dentro de una célula hospedante. El nivel de expresión de un producto deseado en una célula hospedante se puede determinar sobre la base de la cantidad del correspondiente ARN o ARNm que está presente en la célula, o de la cantidad del polipéptido deseado codificado por la secuencia seleccionada. Por ejemplo, el ARNm transcrito desde una secuencia seleccionada se puede cuantificar por hibridación de transferencia Northern, protección de la ARN ribonucleasa, hibridación *in situ* con el ARN celular o por PCR. Las proteínas codificadas por una secuencia seleccionada se pueden cuantificar por diferentes métodos, p.ej. por ELISA, por transferencia
50 Western, por radioinmunoensayos, por inmunoprecipitación, por ensayo de la actividad biológica de la proteína, o por inmunotinción de la proteína seguida por PCR con análisis FACS.

El término “polipéptido” se usa de forma intercambiable con “secuencia de restos aminoácidos” o con el término “proteína” y se refiere a polímeros de aminoácidos de cualquier longitud. Estos términos incluyen también proteínas

que son modificadas después de la traducción por reacciones que incluyen, pero no se limitan a glucosilación, glucación, acetilación, fosforilación, oxidación, amidación o procesado de proteínas. Se pueden hacer modificaciones y cambios, por ejemplo fusiones con otras proteínas, sustituciones, deleciones o inserciones en la secuencia de aminoácidos, en la estructura de un polipéptido siempre que la molécula mantenga su actividad biológica funcional. Por ejemplo se pueden hacer ciertas sustituciones en la secuencia de aminoácidos de un polipéptido o en su secuencia subyacente codificadora de ácido nucleico y se puede obtener una proteína con propiedades similares. Se pueden preparar modificaciones de aminoácidos por ejemplo realizando mutagénesis específica del sitio o mutagénesis mediada por la reacción en cadena de la polimerasa o su secuencia subyacente de ácido nucleico. El término “polipéptido” por lo tanto incluye también, por ejemplo, proteínas de fusión que consisten en un componente de inmunoglobulina, p.ej. el componente Fc, y un factor de crecimiento, p.ej. una interleucina.

Como se usa aquí, el término “anticuerpo” incluye un anticuerpo policlonal, monoclonal, bi-específico, multi-específico, humano, humanizado, o quimérico, un anticuerpo de cadena sencilla, un fragmento de unión al antígeno de un anticuerpo (p.ej., un fragmento Fab o F(ab')₂), un Fv ligado a disulfuro, etc. Dichos anticuerpos se pueden producir por síntesis química, por medios recombinantes o transgénicos, por cultivo de células (p.ej., hibridoma), o por otros medios.

Los fragmentos Fab (Fragmento de unión al antiígeno = Fab) consisten en las regiones variables de ambas cadenas que se mantienen juntas por la región constante adyacente. Estos fragmentos se pueden formar por digestión con proteasas, p.ej. con papaina, a partir de anticuerpos convencionales, pero también se pueden producir fragmentos Fab similares por ingeniería genética. Otros fragmentos de anticuerpos incluyen fragmentos F(ab')₂, que se pueden preparar por escisión proteolítica con pepsina.

Utilizando métodos de ingeniería genética es posible producir fragmentos acortados de anticuerpos que consisten solamente en las regiones variables de la cadena pesada (VH) y de la cadena ligera (VL). Estos fragmentos se conocen como fragmentos Fv (Fragmento variable = fragmento de la parte variable). Puesto que estos fragmentos Fv carecen de la unión covalente de las dos cadenas por las cisteínas de las cadenas constantes, los fragmentos Fv están a menudo estabilizados. Es ventajoso unir las regiones variables de la cadena pesada y de la cadena ligera por un fragmento peptídico corto, p.ej. de 10 a 30 aminoácidos, preferiblemente de 15 aminoácidos. De este modo se obtiene una cadena de un único péptido que consiste en VH y VL, ligadas por un péptido enlace. Una proteína de anticuerpo de este tipo es conocida como un Fv de cadena sencilla (scFv). Los ejemplos de proteínas de anticuerpo scFv de este tipo son conocidos de la técnica anterior.

En los últimos años, se han desarrollado varias estrategias para preparar scFv como un derivado multimérico. Se pretende llegar, en particular, a anticuerpos recombinantes con mejores propiedades farmacocinéticas y de biodistribución así como con mayor avidéz de unión. Con el fin de alcanzar la multimerización del scFv, se prepararon scFv como proteínas de fusión con dominios de multimerización. Los dominios de multimerización pueden ser, p.ej. la región CH3 de una IgG o estructura *coiled coil* (estructuras de hélice) tal como los dominios *Leucin-zipper*. Sin embargo, también hay estrategias en las que la interacción entre las regiones VH/VL del scFv se utiliza para la multimerización (p.ej. dia-, tri- y pentacuerpos). Por diacuerpo un experto quiere decir un derivado scFv bivalente homodimérico. El acortamiento del enlace (*Linker*) en una molécula de scFv a 5-10 aminoácidos lleva a la formación de homodímeros en los que tiene lugar una superposición entre cadenas VH/VL. Los diacuerpos se pueden estabilizar adicionalmente por la incorporación de puentes disulfuros. Los ejemplos de proteínas de anticuerpo diacuerpo son conocidos de la técnica anterior.

Por minicuerpo los expertos entienden un derivado scFv bivalente homodimérico. Consiste en una proteína de fusión que contiene la región CH3 de una inmunoglobulina, preferiblemente IgG, lo más preferiblemente IgG1 como la región de dimerización que está conectada al scFv por una región bisagra (*Hinge*) (p.ej. también de IgG1) y una región enlace (*Linker*). Los ejemplos de proteínas de anticuerpo minicuerpo son conocidos de la técnica anterior.

Por triacuerpo los expertos entienden un derivado scFv trivalente homotrimérico. Los derivados de ScFv en los que VH-VL se fusionan directamente sin una secuencia de enlace llevan a la formación de trímeros.

Los expertos estarán también familiarizados con los llamados minianticuerpos que tienen una estructura bi-, tri- o tetravalente y que se derivan de scFv. La multimerización se lleva a cabo por estructuras *coiled coil* di-, tri- o tetraméricas. En una realización preferida de la presente invención, el gen de interés codifica cualquiera de estos polipéptidos deseados mencionados antes, preferiblemente un anticuerpo monoclonal, un derivado o fragmento del mismo.

El “polipéptido de interés”, “proteína de interés” o “producto de interés” incluye proteínas, polipéptidos, fragmentos de los mismos, péptidos, proteínas de fusión, todos los cuales pueden ser expresados en la célula hospedante seleccionada. Las proteínas deseadas pueden ser por ejemplo anticuerpos, enzimas, citocinas, linfocinas, moléculas

de adhesión, receptores y derivados o fragmentos de los mismos, y cualquier otro polipéptido que pueda servir como agonista o antagonista y/o que pueda tener uso terapéutico o de diagnóstico.

5 Especialmente, las proteínas/polipéptidos deseados o proteínas de interés son por ejemplo, pero sin limitarse a ellos, insulina, factor de crecimiento tipo insulina, hGH, tPA, citocinas, tales como las interleucinas (IL), p.ej. IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, interferón (IFN) alfa, IFN beta, IFN gamma, IFN omega o IFN tau, factor de necrosis tumoral (TNF), tal como TNF alfa y TNF beta, TNF gamma, TRAIL; G-CSF, GM-CSF, M-CSF, MCP-1, VEGF y nanocuerpos. También se incluye la producción de eritropoyetina o cualquier otra hormona o factor de crecimiento y cualquier otro polipéptido que puedan servir como
10 agonistas o antagonistas y/o que tengan uso terapéutico o de diagnóstico. El método según la invención se puede usar también ventajosamente para la producción de anticuerpos, tales como anticuerpos monoclonales, policlonales, multiespecíficos y de cadena sencilla, o fragmentos de los mismos, p.ej. fragmentos Fab, Fab', F(ab')₂, Fc y Fc', cadenas pesadas y ligeras de inmunoglobulina y sus regiones constantes, variables o hipervariables así como los fragmentos Fv y Fd.

15 El "producto de interés" puede ser también un ARN antisentido, ARNt, ARNrs, otros ARN que son parte de riboproteínas u otros ARN reguladores.

El método de la presente invención se puede realizar en todas las células eucariotas. Las células y líneas celulares se pueden presentar p.ej. en un cultivo celular e incluyen pero no se limitan a células eucariotas, tales como células de levaduras, plantas, insectos o mamífero. Por ejemplo, las células pueden ser oocitos, excepto los oocitos humanos, células madre embrionarias, células madre hematopoyéticas o cualquier tipo de células diferenciadas. Se
20 prefiere un método en el que la célula eucariota es una célula de mamífero. Más preferido es un método en el que la célula de mamífero es una célula humana, de simio, murina, de rata, de conejo, de hámster, de cabra, bovina, de oveja o de cerdo. Las líneas celulares o "células hospedantes" preferidas para la producción de compuestos biofarmacéuticos son líneas celulares humanas, de ratón, de rata, de mono, o de roedor. Las células de hámster son más preferidas, preferiblemente las células BHK21, BHK TK⁻, CHO, CHO-K1, CHO-DUKX, CHO-DUKX B1, CHO-S y CHO-DG44 o los derivados/progenies de cualquiera de tales líneas celulares. Son particularmente preferidas las
25 células CHO-DG44, CHO-DUKX, CHO-K1, CHO-S y BHK21, y aún más preferidas las células CHO-DG44 y CHO-DUKX. Además, las células de mieloma murino, preferiblemente las células NS0 y Sp2/0 o los derivados/progenies de cualquiera de tales líneas de células son conocidas también como líneas celulares de producción de proteínas biofarmacéuticas. Los ejemplos de células murinas y de hámster que se pueden usar en la acepción de esta invención se resumen en la Tabla 1.

Tabla 1. Líneas celulares eucariotas de producción

Línea celular	Número de orden
NS0	ECACC No. 85110503
Sp2/0-Ag14	ATCC CRL-1581
BHK21	ATCC CCL-10
BHK TK	ECACC No. 85011423
HaK	ATCC CCL-15
2254-62.2 (derivado BHK-21)	ATCC CRL-8544
CHO	ECACC No. 8505302
CHO tipo natural	ECACC 00102307
CHO-K1	ATCC CCL-61
CHO-DUKX (=CHO duk ⁻ , CHO/dhfr ⁻)	ATCC CRL-9096
CHO-DUKX B11	ATCC CRL-9010
CHO-DG44	Urlaub et al., Cell 33 (2), 405-412, 1983
CHO Pro-5	ATCC CRL-1781
CHO-S	Invitrogen Cat No. 10743-029
Lec13	Stanley P. et al, Ann. Rev. Genetics 18, 525-552, 1984
V79	ATCC CCC-93
B14AF28-G3	ATCC CCL-14
HEK 293	ATCC CRL-1573
COS-7	ATCC CRL-1651
U266	ATCC TIB-196
HuNS1	ATCC CRL-8644
Per.C6	Fallaux, F.J. et al, <i>Human Gen Therapy</i> 9 (13), 1909-1917, 1998
CHL	ECACC No. 87111906

Las células hospedantes son las más preferidas, cuando son establecidas, adaptadas, y completamente cultivadas en condiciones exentas de suero, y opcionalmente en medios que están exentos de cualquier proteína/péptido de origen animal. Son ejemplos de soluciones nutritivas apropiadas los medios comercialmente disponibles tales como medio de Ham F12 (Sigma, Deisenhofen, Alemania), RPMI-1640 (Sigma), medio de Eagle modificado por Dulbecco (DMEM; Sigma), medio mínimo esencial (MEM; Sigma), medio de Dulbecco modificado por Iscove (IMDM; Sigma), CD-CHO (Invitrogen, Carlsbad, CA), CHO-S-SFMII (Invitrogen), medio de CHO exento de suero (Sigma), medio de CHO exento de proteína (Sigma), Medio EX-CELL (SAFC), CDM4CHO y SFM4CHO (HyClone). Cualquiera de los medios puede ser suplementado si es necesario con una variedad de compuestos, ejemplos de los cuales son hormonas y/o otros factores de crecimiento (tales como insulina, transferrina, factor de crecimiento epidérmico, factor de crecimiento tipo insulina), sales (tales como cloruro de sodio, fosfato de calcio, magnesio), tampones (tales como HEPES), nucleósidos (tales como adenosina, timidina), glutamina, glucosa u otras fuentes de energía equivalentes, antibióticos, elementos trazas. También se puede incluir cualquier otro suplemento necesario a concentraciones apropiadas que serán conocidas por los expertos en la técnica. En la presente invención se prefiere el uso de medio exento de suero, pero se puede usar también medio suplementado con una cantidad adecuada de suero para el cultivo de células hospedantes. Para el crecimiento y selección de células modificadas genéticamente que expresan un gen seleccionable se añade al medio de cultivo un agente de selección adecuado.

La “transfección” de células hospedantes eucariotas con secuencias de polinucleótidos o vectores de expresión, que da como resultado células genéticamente modificadas, células recombinantes o transgénicas, se puede realizar por cualquier método bien conocido por los expertos en la técnica. Los métodos de transfección incluyen pero no se limitan a la transfección mediada por liposomas, co-precipitación con fosfato de calcio, electroporación, transfección mediada por policationes (p.ej. transfección mediada por DEAE dextrano), fusión de protoplastos, microinyección e infecciones virales. Preferiblemente, la transfección es una transfección estable. El método de transfección que proporciona una frecuencia óptima de transfección y la expresión de los genes o polinucleótidos heterólogos en la línea y tipo de la célula hospedante particular es el favorable. Se pueden determinar métodos adecuados por procedimientos de rutina. Para los transfectantes estables las construcciones o bien se integran en el genoma de la célula hospedante o en un cromosoma/mini-cromosoma artificial o bien se localizan episómicamente de manera que se mantengan de modo estable dentro de la célula hospedante. Para la generación de células genéticamente modificadas que expresan el producto o productos de interés todos los genes heterólogos requeridos se pueden localizar en un único vector o secuencia de polinucleótidos en unidades de transcripción mono- o multicistónicas. En este caso la célula hospedante es transfectada con vectores individuales o secuencias de polinucleótidos. Los genes heterólogos se pueden localizar también en diferentes vectores o secuencias de polinucleótidos. En este caso las células hospedantes o bien son co-transfectadas con todos los vectores o secuencias de polinucleótidos y/o bien son transfectadas en series sucesivas con los vectores o secuencias de polinucleótidos que codifican los genes de interés.

Por definición, cada secuencia de polinucleótidos o cada gen insertado en una célula hospedante y la respectiva proteína o ARN codificados de ese modo se denomina “heterólogo”, “secuencia heteróloga”, “gen heterólogo”, “secuencia codificadora heteróloga”, “transgén” o “proteína heteróloga” con respecto a la célula hospedante. Esto se aplica incluso si la secuencia a ser introducida o el gen a ser introducido es idéntico a una secuencia endógena o a un gen endógeno de la célula hospedante. Por ejemplo, un gen de actina de hámster introducido en una célula hospedante de hámster es por definición un gen heterólogo.

El término “endógeno” significa que está contenido naturalmente en la célula u organismo. Un gen endógeno es por lo tanto un gen que se encuentra en el genoma de la célula natural no manipulada.

El término “gen marcador de selección” se refiere a un gen que solamente permite que las células portadoras del gen sean seleccionadas o no específicamente en presencia del agente de selección correspondiente. A título ilustrativo, se puede utilizar un gen de resistencia a un antibiótico como un gen marcador seleccionable positivo que permite que la célula hospedante transformada con el gen sea seleccionada positivamente en presencia del correspondiente antibiótico; una célula hospedante no transformada no sería capaz de crecer ni de sobrevivir en las condiciones de cultivo de selección. Los marcadores seleccionables pueden ser positivos, negativos o bifuncionales. Los marcadores seleccionables positivos permiten la selección de células que portan el marcador por conferir resistencia a un fármaco o compensar un defecto metabólico o catabólico en la célula hospedante. Por contraste, los marcadores de selección negativos permiten que las células que portan el marcador sean eliminadas selectivamente. Por ejemplo, la utilización del gen HSV-tk como un marcador hará que las células sean sensibles a agentes tales como aciclovir y ganciclovir. Los genes marcadores seleccionables utilizados aquí, incluyendo los genes seleccionables amplificables, incluirán los mutantes y variantes, fragmentos, equivalentes funcionales, derivados, homólogos y fusiones del gen marcador seleccionable nativo, obtenidos por ingeniería genética, siempre que el producto codificado retenga la propiedad seleccionable. Los derivados útiles generalmente tienen una similitud sustancial de secuencia (a nivel de aminoácidos) en regiones o dominios del marcador seleccionable asociada con la propiedad seleccionable. Se ha descrito una variedad de genes marcadores, bien conocidos por los expertos, incluyendo los marcadores bifuncionales (esto es positivos/negativos) (véase p.ej. los documentos WO 92/08796 y WO 94/28143), incorporados aquí como referencia. Por ejemplo, los genes seleccionables utilizados

comúnmente con células eucariotas incluyen los genes de aminoglucósido fosfotransferasa (APH), higromicina fosfotransferasa (HYG), dihidrofolato reductasa (DHFR), timidina cinasa (TK), glutamina sintetasa, asparagina sintetasa, y los genes que codifican la resistencia a la neomicina (G418), puromicina, histidinol D, bleomicina y fleomicina.

- 5 El "gen marcador seleccionable amplificable" usualmente codifica una enzima que es necesaria para el crecimiento de las células eucariotas en estas condiciones. Por ejemplo, el gen marcador seleccionable amplificable puede codificar DHFR cuyo gen se amplifica cuando una célula hospedante transfectada con el mismo se cultiva en presencia del agente selectivo, metotrexato (MTX). Por consiguiente, las células hospedantes modificadas genéticamente según cualquier método descrito aquí están incluidas en esta invención, en donde el gen marcador seleccionable amplificable codifica por ejemplo un polipéptido que tiene la función de dihidrofolato reductasa (DHFR), glutamina sintetasa, CAD, adenosina desaminasa, adenilato desaminasa, UMP sintetasa, IMP 5'-deshidrogenasa, xantina guanina fosforibosil transferasa, HGPRTasa, timidina cinasa, timidilato sintetasa, P glucoproteína 170, ribonucleótido reductasa, asparagina sintetasa, arginosuccinato sintetasa, ornitina descarboxilasa, HMG CoA reductasa, acetilglucosaminil transferasa, treonil-ARNt sintetasa o Na⁺K⁺-ATPasa. Para una revisión de los ejemplos de genes marcadores seleccionables amplificables listados en la Tabla 2 véase Kaufman, Methods in Enzymology, 185, 537-566, 1990.

20 Un gen marcador seleccionable amplificable particular es el gen que codifica la dihidrofolato reductasa (DHFR) que es necesaria para la biosíntesis de las purinas. Las células que carecen del gen DHFR no crecerán en un medio que carece de purinas. El gen DHFR es por lo tanto útil como un marcador seleccionable dominante para seleccionar y amplificar genes en aquellas células que crecen en un medio que carece de purinas. El agente de selección usado conjuntamente con un gen DHFR es metotrexato (MTX).

25 Otro marcador de selección y/o amplificación es el gen de la glutamina sintetasa (GS). El gen GS codifica la enzima glutamina sintetasa que es necesaria para la síntesis del aminoácido glutamina. Las células que carecen del gen GS o que expresan bajos niveles de GS endógena no crecerán en un medio exento de glutamina. El gen GS es por lo tanto útil como un marcador seleccionable dominante para seleccionar y amplificar genes en aquellas células que crecen en un medio que carece de glutamina. El agente de selección utilizado conjuntamente con el gen GS es metionina sulfoximina (MSX).

Tabla 2. Genes marcadores seleccionables amplificables

Gen marcador seleccionable amplificable	Número de acceso	Agente de selección
Dihidrofolato reductasa	M19869 (hámster) E00236 (ratón)	Metotrexato (MTX)
Metalotioneina	D10551 (hámster) M13003 (humano) M11794 (rata)	Cadmio
CAD (Carbamoil-fosfato sintetasa: Aspartato transcarbamilasa: Dihidrorotasa	M23652 (hámster) D78586 (humano)	N-Fosfoacetil-L-aspartato
Adenosina desaminasa	K02567 (humano) M10319 (ratón)	Xyl-A- o adenosina, 2'desoxicoformicina
AMP (adenilato) desaminasa	D12775 (humano) J02811 (rata)	Adenina, azaserina, coformicina
UMP sintasa	J03626 (humano)	6-Azauridina, pirazofurano
IMP 5' deshidrogenasa	J04209 (hámster) J04208 (humano) M33934 (ratón)	Ácido micofenólico
Xantina-guanina fosforibosiltransferasa	X00221 (<i>E. coli</i>)	Ácido micofenólico con xantina limitante
HGPRTasa mutante o timidina cinasa mutante	J00060 (hámster) M13542, K02581 (humano) J00423, M68489 (ratón) M63983 (rata) M36160 (herpesvirus)	Hipoxantina, aminopterina, y timidina (HAT)
Timidilato sintetasa	D00596 (humano) M13019 (ratón) L12138 (rata)	5-Fluorodesoxiuridina
P-glucoproteína 170 (MDR1)	AF016535 (humano) J03398 (ratón)	Múltiples fármacos, p.ej. adriamicina, vincristina, colchicina

Ribonucleótido reductasa	M124223, K02927 (ratón)	Afidicolina
Glutamina sintetasa	AF150961 (hámster) U09114, M60803 (ratón) M29579 (rata)	Metionina sulfoximina (MSX)
Asparagina sintetasa	M27838 (hámster) M27396 (humano) U38940 (ratón) U07202 (rata)	β -Aspartil hidroxamato, Albizziina, 5'Azacitidina
Argininosuccinato sintetasa	X01630 (humano) M31690 (ratón) M26198 (bovino)	Canavanina
Ornitina decarboxilasa	M34158 (humano) J03733 (ratón) M16982 (rata)	α -Difluorometilornitina
HMG-CoA reductasa	L00183, M12705 (hámster) M11058 (humano)	Compactina
N-Acetilglucosaminil transferasa	M55621 (humano)	Tunicamicina
Treonil-ARNt sintetasa	M63180 (humano)	Borrelidina
Na ⁺ K ⁺ -ATPasa	J05096 (humano) M14511 (rata)	Ouabaina

5 La selección se puede hacer también clasificando las células activadas por fluorescencia (FACS) utilizando por ejemplo un marcador de la superficie celular, β -galactosidasa bacteriana o proteínas fluorescentes (p.ej. proteínas fluorescentes verdes (GFP) y sus variantes de *Aequorea victoria* y *Renilla reniformis* u otras especies; proteínas fluorescentes rojas, proteínas fluorescentes y sus variantes de especies no bioluminiscentes (p.ej. *Discosoma* sp., *Anemonia* sp., *Clavularia* sp., *Zoanthus* sp.) para seleccionar las células recombinantes.

10 El término “agente de selección” se refiere a una sustancia que interfiere con el crecimiento o la supervivencia de una célula hospedante que es deficiente en un gen seleccionable particular. Por ejemplo, para seleccionar la presencia de un gen de resistencia a un antibiótico tal como APH (aminoglucósido fosfotransferasa) en una célula transfectada se utiliza el antibiótico genticina (G418). El agente de selección puede comprender también un “agente de amplificación” que se define para los fines de esta invención como un agente para amplificar copias del gen amplificable si el gen marcador seleccionable relacionado es un marcador seleccionable amplificable. Por ejemplo, MTX es un agente de selección útil para la amplificación del gen DHFR

15 Otra realización de los métodos mencionados antes se refiere a un método, en el que los polipéptidos/productos que son codificados por el gen o genes de interés y que se expresan en dicha célula hospedante, se aíslan de las células o del sobrenadante del cultivo celular, si se secretan en el medio de cultivo.

20 Dichas células de producción se cultivan preferiblemente en medio exento de suero y en cultivo en suspensión en condiciones que son favorables para la expresión del gen o genes deseados y aislando la proteína de interés de las células y/o del sobrenadante del cultivo celular. Preferiblemente la proteína de interés se recupera del medio de cultivo como un polipéptido secretado, o se puede recuperar de los lisados de células hospedantes si se expresan sin ninguna señal secretora. Es necesario purificar la proteína de interés de otras proteínas recombinantes, proteínas de la célula hospedante y contaminantes de modo que se obtengan preparaciones de la proteína de interés sustancialmente homogéneas. Como una primera etapa, a menudo se separan las células y/o restos de células en forma de partículas a partir del medio de cultivo o del lisado. Después se purifica el producto de interés de las proteínas solubles, polipéptidos y ácidos nucleicos contaminantes, por ejemplo, por fraccionamiento en columnas de inmunoadfinidad o de intercambio iónico, precipitación con etanol, HPLC de fase inversa, cromatografía en Sephadex, cromatografía en sílice o en una resina de intercambio catiónico tal como DEAE. En general, los métodos que enseñan a los expertos cómo purificar una proteína heteróloga expresada por células hospedantes, son bien conocidos en la técnica.

30 La práctica de la presente invención empleará, a menos que se indique otra cosa, métodos convencionales de biología celular, biología molecular, cultivo celular, inmunología y similares que son conocidos por los expertos en la técnica. Estos métodos están completamente descritos en las publicaciones científicas actuales.

Los siguientes ejemplos no son limitantes. Ellos presentan simplemente posibles realizaciones de la invención. Un experto en la técnica podrá ajustar fácilmente las condiciones a aplicar a otras realizaciones.

35 Ejemplos

Abreviaturas

	AP:	Fosfatasa alcalina
	BGH:	Hormona de crecimiento bovina
	bp:	Pares de bases
	CHO:	Ovario de hámster chino
5	DHFR:	Dihidrofolato reductasa
	ELISA:	Ensayo de inmunoabsorbente ligado a enzimas
	FACS:	Separación de células activadas por fluorescencia
	HGH:	Hormona de crecimiento de hámster
	HT:	Hipoxantina/timidina
10	HRPO:	Peroxidasa de rábano
	IgG:	Inmunoglobulina
	IRES:	Sitio interno de entrada al ribosoma
	kb:	Kilobase
	mAb:	Anticuerpo monoclonal
15	MTX:	Metotrexato
	NPT:	Neomicina fosfotransferasa
	nt:	Nucleótidos
	PBS:	Solución salina tamponada con fosfato
	PCR:	Reacción en cadena de la polimerasa
20	SEAP:	Fosfatasa alcalina secretada
	sICAM:	Molécula de adhesión intracelular soluble
	UTR:	Región no traducida

Materiales y métodos

Cultivo celular

- 25 Se cultivan células CHO-DG44/dhfr^{-/-} permanentemente en suspensión en el medio exento de suero CHO-S-SFMII (Invitrogen) suplementado con hipoxantina y timidina (HT). Se cultivan las células en frascos de cultivo celular a 37 °C en una atmósfera humidificada que contiene 5 % de CO₂. El número de células así como la viabilidad de las células se determinan con un contador de células CASY1 (Schaefer System, Alemania), un Cedex (Innovatis AG, Alemania) o mediante exclusión con colorante azul tripán. Se siembran las células a una concentración de 1-3×10⁵ células/mL en medio fresco cada dos a tres días.

Transfecciones

- 35 Las transfecciones de células CHO-DG44 se realizan utilizando reactivo Lipofectamina Plus (Invitrogen). En un pocillo de una cámara de 6 pocillos, se siembran por transfección 6×10⁵ células con crecimiento exponencial en 0,8 mL de medio CHO-S-SFMII suplementado con hipoxantina/timidina (HT). Se genera para cada transfección una mezcla de plásmido ADN, 4 µL de Lipofectamina y 6 µL de reactivo Plus en un volumen de 200 µL y se añade a las células, siguiendo el protocolo del fabricante. Después de incubación durante 3 horas, se añaden 2 mL de medio CHO-S-SFMII suplementado con HT.

- 40 Se realizan transfecciones transitorias por triplicado y se recogen el sobrenadante y las células 2 días después de la transfección. Para una selección basada en DHFR de células CHO-DG44 transfectadas estables se reemplaza el medio con medio CHO-S-SFMII exento de HT 48 horas después de la transfección. La amplificación del gen basada en DHFR se alcanza añadiendo al medio MTX en el intervalo de 5-2000 nM (Sigma) como agente de selección de amplificación. En el caso de co-transfecciones se realiza una selección basada en DHFR y NPT de células CHO-DG44 transfectadas estables transfiriendo las células 48 horas después de la transfección a medio CHO-S-SFMII exento de HT suplementado con G418 (Invitrogen) en una concentración de 200-400 µg/mL.

45 Vectores de expresión

- 50 Los vectores de expresión eucariotas son derivados del vector pAD-CMV1 (documento WO 9201055) y median en la expresión constitutiva de los genes heterólogos dirigida por el promotor/potenciador de CMV. Para la terminación y poliadenilación del transcrito del gen de interés, los vectores contienen la señal de poliadenilación de SV40 tardía (SEQ ID NO: 11) o la señal de poliadenilación de BGH (SEQ ID NO: 12). Los vectores pBID codifican un mini gen DHFR como marcador de selección amplificable (véase por ejemplo el documento EP 0 393 438) mientras que los vectores pBIN codifican un gen NPT como marcador de selección bajo el control del promotor de SV40 temprano y una señal de poliadenilación de timidina cinasa (Figura 1).

- 55 Los genes de interés que codifican la sICAM humana, los anticuerpos monoclonales de cadena pesada y ligera (isotipo IgG1, IgG2 o IgG4) o las proteínas de fusión Fc se clonan en los vectores utilizando los múltiples sitios de clonación localizados entre el promotor y la señal de poliadenilación.

ELISA

5 Los títulos de sICAM se cuantifican por ELISA con protocolos estándar utilizando dos anticuerpos monoclonales desarrollados internamente específicos de sICAM (como se describen por ejemplo en las patentes de Estados Unidos números 5.284.931 y 5.475.091), donde uno de los anticuerpos es un anticuerpo conjugado a HRPO. Como un estándar se utiliza una proteína sICAM.

Los títulos de mAb se cuantifican por ELISA con protocolos estándar utilizando un fragmento Fc de cabra anti-IgG humana (Dianova) y un anticuerpo de cabra conjugado a AP anti cadena ligera kappa humana (Sigma). El anticuerpo mAb purificado del mismo tipo del mAb expresado se utiliza como estándar.

Se analizan las muestras utilizando un lector Spectra Fluor Plus (TECAN, Crailsheim, Alemania).

10 La productividad celular (pg/célula/día) se calcula con la fórmula $pg/((Ct-Co)/\ln(Ct-Co))$ donde "Co" es el número de células en el momento de la siembra, "Ct" el número de células en el momento de la recogida y "t" el tiempo de cultivo.

Ensayo de SEAP

15 La actividad de la SEAP se determina con el ensayo del gen informador SEAP según el protocolo del fabricante (Roche Diagnostics).

Ejemplo 1. Aislamiento y clonación de la región de terminación de la transcripción de la hormona de crecimiento del hámster (HGH)

20 Para el aislamiento de la región completa de la señal de poliadenilación del gen de la hormona de crecimiento a partir del genoma de CHO-DG44 (hámster chino, *Cricetus griseus*), el ADN de CHO-DG44 genómico ligado a un adaptador sirve como plantilla en una PCR anidada. La PCR primaria se realiza con una combinación de cebadores con complementariedad con la secuencia del adaptador y con una secuencia del gen de la hormona de crecimiento, respectivamente. El cebador específico del gen, GH for1 (5'-GAGACCTACCTGCGGGTCA TGA-3'; SEQ ID NO: 1) se diseña sobre la base de una secuencia de ADNc de la hormona de crecimiento del hámster sirio (*Mesocricetus auratus*; Genbank S66299) y se localiza 35 bp corriente arriba del codón de terminación. Se realiza una PCR 25 secundaria sobre los productos de la PCR primaria con una combinación de un cebador adaptador interior y un segundo cebador específico del gen anidado, GH for2; 5'-AGTGCCGTCGCTTTGTGGAAA G-3'; SEQ ID NO: 2), colocado directamente corriente abajo de la posición del cebador GH for1. Los fragmentos de ADN resultantes se subclonan en un vector de clonación TA (Invitrogen) y se analizan después por análisis de secuencia. El fragmento más largo de ADN contiene aparte de la secuencia del cebador GH for2 otras 13 bp del extremo 3' de la región de 30 codificación seguidas por un codón de terminación y 324 bp de la región 3' no traducida de la hormona de crecimiento de *Cricetus griseus* (Figura 2; SEQ ID NO: 7).

35 Para obtener sólo la región 3' no traducida con 324 bp en total (SEQ ID NO: 8) se realiza otra PCR utilizando los cebadores GH Sfi for1 (SEQ ID NO: 3) y GH Xba rev1 (SEQ ID NO: 4). De ese modo el fragmento de ADN de 362 bp mencionado antes (SEQ ID NO: 7), subclonado en el vector TA, sirve como plantilla en la PCR. La secuencia 40 ampliada (SEQ ID NO:8) tiene las siguientes homologías con las regiones 3' no traducidas de la hormona de crecimiento de diferentes especies: 72,1 % con la secuencia del hámster sirio *Mesocricetus auratus* (Genbank S66299), 71,6 % con la secuencia de *Mus musculus* (Genbank Z46663), 61 % con la secuencia de *Rattus norvegicus* (Genbank V01239) y 50,4% con la secuencia BGH de *Bos taurus* (Genbank J00008) (Figura 3).

45 El método basado en la PCR se utiliza también para la generación de subclones con diferentes deleciones del extremo 3' de la región 3' no traducida aislada. Utilizando la combinación de cebadores GH Sfi for1 (SEQ ID NO: 3) y GH Xba rev2 (SEQ ID NO: 5) se genera un fragmento de 189 bp de la región 3' no traducida (SEQ ID NO: 9) y con la combinación de cebadores GH Sfi for1 (SEQ ID NO: 3) y GH Xba rev3 (SEQ ID NO: 6) se genera un subfragmento de 113 bp (SEQ ID NO: 10). De este modo, todos los fragmentos amplificados de la región 3' no traducida tienen un extremo 5' idéntico que corresponde al primer nucleótido después del codón de terminación y un extremo 3' variable (Figura 4).

50 Los productos de la PCR se digieren con SfiI y XbaI y los fragmentos de restricción resultantes se utilizan para reemplazar la secuencia de la señal de poliadenilación de SV40 tardía en el vector pJR106, que codifica la sICAM humana (Figura 5). Los vectores pJR131, pJR134 y pJR135 resultantes contienen ahora una secuencia de la señal de poliadenilación derivada de la hormona de crecimiento de *Cricetus griseus*, HGH corta requerida, con un tamaño de 324 bp (SEQ ID NO: 8), 189 bp (SEQ ID NO: 9) y 113 bp (SEQ ID NO: 10), respectivamente (Figura 5).

Ejemplo 2. Impacto de la secuencia de la señal de poliadenilación HGH sobre la expresión transitoria de sICAM

Para evaluar el impacto de la secuencia de la señal de poliadenilación derivada de la hormona de crecimiento de *Cricetus griseus* (HGH) sobre la expresión de un gen de interés, sICAM, se realizan transfecciones transitorias independientes de los sitios de integración cromosómica. Se transfectan células CHO-DG44 con el plásmido pJR131 que contiene 324 bp de la 3'UTR de la hormona de crecimiento del hámster (=HGH, SEQ ID NO: 8) (Figura 5). Los vectores que contienen las secuencias de la señal de poliadenilación SV40 tardía (pJR106) o BGH (pJR110) se usan como control (Figura 5). Aparte de las diferentes secuencias de terminación la organización genética de los diferentes vectores para la expresión de sICAM es idéntica.

Se recogen los sobrenadantes 2 días después de la transfección y se determinan los títulos de sICAM utilizando ELISA. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP (100 ng de ADN/reacción de transfección), que codifica la fosfatasa alcalina secretada, y se mide la actividad de SEAP.

La Figura 6 presenta los datos de 2 series de transfección transitoria independientes realizadas por duplicado. Sorprendentemente, la expresión más alta de sICAM se obtiene con la secuencia de la señal de poliadenilación derivada del gen de la hormona de crecimiento de hámster. El título se aumenta hasta un 21 % (serie de transfección #1) en comparación con las células transfectadas con el vector pJR110 que contiene la señal de poliadenilación BGH y se aumenta hasta un 40 % (serie de transfección #1) en comparación con las células transfectadas con el vector pJR106 que contiene la señal de poliadenilación SV40 tardía.

Ejemplo 3. Impacto de la señal de poliadenilación HGH sobre la expresión transitoria de un anticuerpo IgG4

Para evaluar el impacto de la secuencia de la señal de poliadenilación derivada de la hormona de crecimiento de *Cricetus griseus* (HGH) sobre la expresión de un gen de interés, mAb IgG4/kappa humanizado, se realizan transfecciones transitorias independientes de los sitios de integración cromosómica. Se co-transfectan las células CHO-DG44 con la combinación de vectores pBID/IgG4 y pBIN/kappa. Ambos vectores contienen 324 bp de la UTR 3' de la hormona de crecimiento de hámster (=HGH; SEQ ID NO: 8) como una secuencia de la señal de poliadenilación. Como control las células CHO-DG44 se co-transfectan con la combinación de vectores pBID-B/IgG4 y pBIN-B/kappa que contienen la señal de poliadenilación BGH (véase la Figura 1 para vectores básicos). Aparte de las diferentes secuencias de terminación la organización genética de los diferentes vectores para la expresión de IgG4/kappa mAb es idéntica.

Se recogen los sobrenadantes 2 días después de la transfección y se determinan los títulos de IgG4 utilizando ELISA. Se transfectan 6 conjuntos de células por combinación de vectores. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP (100 ng de ADN/reacción de transfección), que codifica la fosfatasa alcalina secretada, y se mide la actividad de SEAP.

Sorprendentemente, los títulos obtenidos con la secuencia de la señal de poliadenilación HGH son como promedio un 35 % más altos que para la señal de poliadenilación BGH (Figura 7).

Ejemplo 4. Ensayo de diferentes variantes HGH

Se generan por la PCR dos clones, con delección en 3', de la secuencia HGH de 324 bp derivada de la hormona de crecimiento de *Cricetus griseus* (SEQ ID NO: 8) y se colocan como una secuencia de la señal de poliadenilación corriente abajo del gen sICAM. Los vectores pJR134 y pJR135 resultantes (Figura 5) contienen un tramo más corto de la secuencia HGH de 189 bp (SEQ ID NO: 9) y 113 bp (SEQ ID NO: 10), respectivamente, que tienen una posición común en el extremo 5' (Figura 4).

Para evaluar el impacto de las variantes de delección de HGH sobre la expresión de un gen de interés, sICAM, se realizan transfecciones transitorias independientes de los sitios de integración cromosómica. Se transfectan células CHO-DG44 con los vectores pJR134 y pJR135. El vector pJR131 que contiene la secuencia de 324 bp de HGH se usa como control (Figura 5). Aparte de las diferentes secuencias de terminación la organización genética de los diferentes vectores para la expresión de sICAM es idéntica.

Se recogen los sobrenadantes 2 días después de la transfección y se determinan los títulos de sICAM utilizando ELISA. Para corregir las variaciones en la eficiencia de transfección se co-transfectan las células con el plásmido pCMV-SEAP (100 ng de ADN/reacción de transfección), que codifica la fosfatasa alcalina secretada, y se mide la actividad de SEAP.

La Figura 8 presenta los datos de 2 series de transfección transitoria independientes realizadas por duplicado. Ambas variantes de delección de HGH llevan a niveles reducidos de expresión de sICAM. La secuencia de HGH de 189 bp contenida en el vector de expresión pJR134 da como resultado una reducción más moderada de hasta 23 %. Por lo tanto el fragmento de 189 bp presenta un comportamiento comparable a la señal de poliadenilación BGH y SV40 tardía (véase el ejemplo 2 y 3). Sin embargo, la secuencia de HGH más corta de 113 bp contenida en el vector de expresión pJR135 lleva hasta una reducción del 78 % de la expresión de sICAM. Esto demuestra que entre las

regiones de HGH de 190 a 324 bp de la SEQ ID NO:8 están localizadas secuencias que contribuyen a una eficiente expresión de un gen de interés.

Ejemplo 5. Expresión estable de proteínas a altos niveles utilizando la señal de poliadenilación de HGH

5 Se co-transfectan células CHO-DG44 con las combinaciones de vectores que codifican la cadena pesada y la cadena ligera de mAbs de diferentes isotipos (IgG1, IgG2, IgG4) o las proteínas de fusión Fc donde la parte Fc se deriva de IgG1 o IgG2. Los vectores básicos pBID y pBIN (Figura 1) usados para la expresión contienen la secuencia de 324 bp de HGH (SEQ ID NO:8) como secuencia de la señal de poliadenilación en posición corriente abajo del gen de interés. Se seleccionan conjuntos de células estables utilizando una selección basada en DHFR y NPT, 2 días después de la transfección. La primera selección de transfectantes estables es seguida por dos etapas sucesivas de amplificación del gen mediada por DHFR añadiendo al medio de cultivo MTX 100 nM en la primera serie y posteriormente MTX 800 nM. Se obtienen clones de células individuales o por clonación por dilución o por un depósito basado en FACS de células individuales en pocillos de una placa de 96 pocillos.

15 Los datos experimentales demuestran que se puede alcanzar una alta expresión de una proteína de interés en transfectantes estables utilizando la señal de poliadenilación de HGH de *Cricetus griseus*. Se obtienen conjuntos de células y clones de células con productividades específicas en el intervalo de 10-45 pg/célula/día y títulos en procesos de lote alimentado de hasta 6,3 g/L (Figura 9).

Tabla de secuencias:

SEQ ID NO: 1	Cebador GH for1
SEQ ID NO: 2	Cebador GH for2
SEQ ID NO: 3	Cebador GH Sfi for1
SEQ ID NO: 4	Cebador GH Xba rev1
SEQ ID NO: 5	Cebador GH Xba rev2
SEQ ID NO: 6	Cebador GH Xba rev3
SEQ ID NO: 7	<i>Cricetus griseus</i> , secuencia de la hormona de crecimiento, parte de la región de codificación 3' y región no traducida 3' (362 nucleótidos)
SEQ ID NO: 8	<i>Cricetus griseus</i> , región no traducida 3' de la hormona de crecimiento (324 nucleótidos)
SEQ ID NO: 9	<i>Cricetus griseus</i> , región no traducida 3' de la hormona de crecimiento (189 nucleótidos)
SEQ ID NO: 10	<i>Cricetus griseus</i> , región no traducida 3' de la hormona de crecimiento (113 nucleótidos)
SEQ ID NO: 11	SV40, terminación tardía y secuencia de poliadenilación (222 nucleótidos)
SEQ ID NO: 12	<i>Bos taurus</i> , secuencia de terminación y de poliadenilación de la hormona de crecimiento (208 nucleótidos)
SEQ ID NO: 13	Secuencia de Kozak, sitio de consenso de unión a ribosoma (13 nucleótidos)

REIVINDICACIONES

1. Una señal de poliadenilación que comprende un ácido nucleico que comprende una secuencia que es idéntica al menos en un 75 % a la SEQ ID NO:8.
- 5 2. La señal de poliadenilación según la reivindicación 1, que comprende una secuencia idéntica al menos en un 80 %, 85 %, 90 %, 95 % o 98 % a la SEQ ID NO:8.
3. La señal de poliadenilación según las reivindicaciones 1 o 2, que comprende la SEQ ID NO:8.
4. La señal de poliadenilación según las reivindicaciones 1 a 3, en donde dicha señal de poliadenilación está ligada operativamente a una secuencia de codificación heteróloga.
5. Un vector que comprende una cualquiera de dichas señales de poliadenilación de las reivindicaciones 1 a 4.
- 10 6. El vector de la reivindicación 5, que comprende un gen heterólogo de interés que codifica un producto heterólogo de interés.
7. El vector de la reivindicación 6, en donde el producto de interés es un polipéptido de interés y dicho polipéptido de interés es un anticuerpo, fragmento de anticuerpo o proteína de fusión.
8. Una célula que comprende el vector de una cualquiera de las reivindicaciones 5 a 7.
- 15 9. La célula según la reivindicación 8, en donde la señal de poliadenilación de las reivindicaciones 1 a 4, está ligada operativamente a una unidad de transcripción que codifica un producto de interés y en donde el producto de interés es un polipéptido de interés codificado por un gen de interés.
10. La célula según las reivindicaciones 8 o 9, en donde dicha célula es una célula de hámster.
- 20 11. La célula según la reivindicación 10, en donde la célula de hámster es una célula de ovario de hámster chino (CHO), preferiblemente una célula CHO DG44.
12. Un método para preparar un polipéptido de interés codificado por un gen de interés, comprendiendo el método:
 - (a) Proporcionar una célula hospedante que comprende un vector según una cualquiera de las reivindicaciones 5 a 7, o proporcionar una célula según una cualquiera de las reivindicaciones 8 a 10
 - (b) Cultivar dicha célula, en condiciones que permitan la proliferación de la célula y la expresión del gen de interés,
 - 25 (c) Recoger el polipéptido de interés y
 - (d) Purificar el polipéptido de interés.
13. El uso de una cualquiera de las señales de poliadenilación de las reivindicaciones 1 a 4, como un aislante.
14. Un kit que comprende una cualquiera de las señales de poliadenilación de las reivindicaciones 1 a 4, un vector, una célula y un medio de cultivo celular para cultivo de dicha célula.

30

FIGURA 1

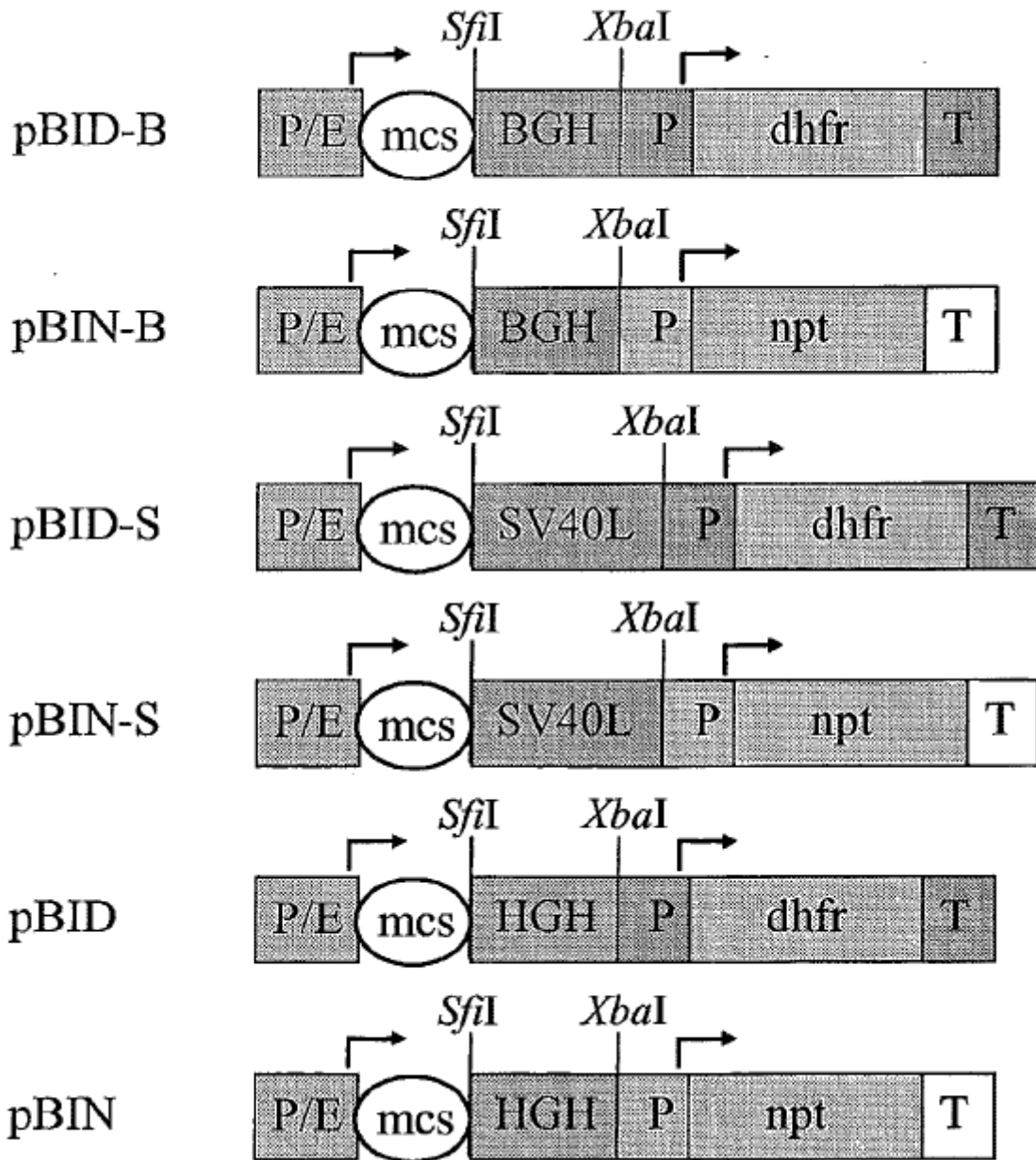


FIGURA 2

	GH for2 →			Stop		
1	AGTGCCGTCG	CTTGTGGAA	AGCAGCTGTG	CCTTTAGCA	GCGGCGTCTC	TGCTGGACTC
	TCACGGCAGC	GAAACACCTT	TCGTGACAC	GGAAAATCCT	GCCGCAGAG	ACGACCTGAG
61	CCCAGCGCCC	CCCTTTACCC	TGGCAACTGC	CCACCCCTAT	GCTTTGCCCT	AATAAAATGA
	GGGTCGCGGG	GGGAAATGGG	ACCGTTGACC	GGTGGGATA	GAAACGGGA	TTATTTTACT
121	AGATGCATTG	TATTGCTTGG	CTAGATGTAT	TTCTGTTGTG	GGATGGAGGG	TGGTGTCAAA
	TCTACGTAAC	ATAACGAACC	GATCTACATA	AAGACAACAC	CCTACCTCCC	ACCACAGTTT
181	GAGTCCTAGA	GCCCGACATG	CCTGTGGGCT	GCTGGAAGAA	CAGCCCTGAC	TTGCCTGGA
	CTCAGGATCT	CCGGCTGTAC	GGACACCCGA	CGACCTTCTT	GTCGGGACTG	AAACGGACCT
241	CCAAGTAGAG	TCAACACATC	ACTTCCCTG	TCTCGTGATG	AGECTGCTCC	CACTCCAGAG
	GGTTCATCTC	AGTTGTGTAG	TGAAGGGGAC	AGAGCACTAC	TCGGACGAGG	GTGAGGTCTC
301	TCAGAATCCC	AGCTCTCTGG	ACAGTCACAA	GGCGGCAAGG	TCCTATGTCA	CCCCATAAA
	AGTCTTAGGG	TCGAGAGACC	TGTCAGTGTT	CCGCCGTCC	AGGATACAGT	GGGGGTATTT
361	GT					
	CA					

FIGURA 3

			50
C.griseus (SEQ ID NO:8)	(1)	-----CAGCGGGCTCTCTGCTGGACTCCCCAGCGCCCCCTTTACCC	
M.auratus (S66299) 3'UTR	(1)	CGGCTCTCTACGGGTCTCTGCTGACTCCCCAGCGCCCTCCCTTTGCC	100
M. musculus (Z46663) 3'UTR	(1)	CGAGTCACCAAGTGTCTCTGCTGCACTCTCCCTGTGCCCTCCCTCCGCC	
R. norvegicus (V01239) 3'UTR	(1)	CGAGCACCACTGGTGTCTCTGCTGGACTCCCCCTTACCCCCCTGTACTC	
Bos taurus (J00008) 3'UTR	(1)	CTTGGCAGCACTGTCTCTGCTGGCCCCCCCCGGTGCCTTCCCTTACCC	51
C.griseus (SEQ ID NO:8)	(43)	TGGCAACTGCCACCCCTAIGCTTTGCCCTAATAAAAATGAAGA---TGCA	100
M.auratus (S66299) 3'UTR	(51)	TGGCAACTGCCATCCCTATCTTTGCCCTAATAAAAATGAAGA---TGCA	
M. musculus (Z46663) 3'UTR	(48)	TGGCAACTGCCACCCCTGGCTTTGTCCCTAATAAAAATGAAGA---TGCA	
R. norvegicus (V01239) 3'UTR	(49)	TGGCAACTGCCACCCCTAGACTTTGTCCCTAATAAAAATGAAGA---TGCA	
Bos taurus (J00008) 3'UTR	(48)	TGGCAACTGCCACCCCACTGCTTTGCCCTAATAAAAATGAAGAARTTGCA	101
C.griseus (SEQ ID NO:8)	(90)	TTGTATTGCTTGGCTAGATGTATTTCTGTTGTGGGATGGAGG---GTGGTGT	150
M.auratus (S66299) 3'UTR	(98)	TCNTAATAAAAATTTT-----	
M. musculus (Z46663) 3'UTR	(94)	TCATATGACCCGGCTAGAGGTCTTTCTCTTATGGGATGGAGGAGTCTGT	
R. norvegicus (V01239) 3'UTR	(95)	TCATATGACTCCGCTAGACATCTTTTCTTTTATGGGCGTCCGGTCTTTT	
Bos taurus (J00008) 3'UTR	(98)	TGGATTGCTTGACTAGGTCTATTCTATTCTGGGG---GCTGG---GTGGGG	151
C.griseus (SEQ ID NO:8)	(139)	CAAAGAG---TCCTAGAGGCCGACATGCCCTGTGGGCTGCTGGAAGAACAGC	200
M.auratus (S66299) 3'UTR	(112)	-----	
M. musculus (Z46663) 3'UTR	(144)	CAATCTTGTTCCTTCGAGGCCGCGCAAGATCCCT---	
R. norvegicus (V01239) 3'UTR	(145)	TTTTTAGATTATTTAATTTATTTAATAG---	
Bos taurus (J00008) 3'UTR	(147)	AGGATAG---CAAGCCCGAGGATTCGGTAGCAATAGCAGGCATCCCTGGC	201
C.griseus (SEQ ID NO:8)	(187)	CCTGACTTTGCCTGGACCAAGTAGAGTCAACACATCACCTCCCTGTCTC	250
M.auratus (S66299) 3'UTR	(112)	-----	
M. musculus (Z46663) 3'UTR	(176)	-----	
R. norvegicus (V01239) 3'UTR	(172)	-----	
Bos taurus (J00008) 3'UTR	(194)	GATGGGTTGGGCTCTATGGGTACCAGGGTGCCTGAAGAAATTGACCCGGTCTC	251
C.griseus (SEQ ID NO:8)	(237)	GTGATCAGCCCTGCTCCCACTCCAGAGTCAGAAATCCCACTCTCTGGACAG	300
M.auratus (S66299) 3'UTR	(112)	-----	
M. musculus (Z46663) 3'UTR	(176)	-----	
R. norvegicus (V01239) 3'UTR	(172)	-----	
Bos taurus (J00008) 3'UTR	(244)	CTGCTGGG-----	301
C.griseus (SEQ ID NO:8)	(287)	TCACAAGGCGGCAAGGTCTCTATGTCACCCCCATAAAGT	338
M.auratus (S66299) 3'UTR	(112)	-----	
M. musculus (Z46663) 3'UTR	(176)	-----	
R. norvegicus (V01239) 3'UTR	(172)	-----	
Bos taurus (J00008) 3'UTR	(252)	-----	

FIGURA 4

			1	Stop	50
HGH 362 nt (SEQ ID NO:7)	(1)	AGTGCCGTCGCTTTGTGGAAAGCAGCTGTGCCCTTTAGCAGCGGCGTCTC			
HGH 324 nt (SEQ ID NO:8)	(1)	-----CAGCGGCGTCTC			
HGH 189 nt (SEQ ID NO:9)	(1)	-----CAGCGGCGTCTC			
HGH 113 nt (SEQ ID NO:10)	(1)	-----CAGCGGCGTCTC			
			51		100
HGH 362 nt (SEQ ID NO:7)	(51)	TGCTGGACTCCCCAGCGCCCCCTTTACCCTGGCAACTGCCACCCCTAT			
HGH 324 nt (SEQ ID NO:8)	(13)	TGCTGGACTCCCCAGCGCCCCCTTTACCCTGGCAACTGCCACCCCTAT			
HGH 189 nt (SEQ ID NO:9)	(13)	TGCTGGACTCCCCAGCGCCCCCTTTACCCTGGCAACTGCCACCCCTAT			
HGH 113 nt (SEQ ID NO:10)	(13)	TGCTGGACTCCCCAGCGCCCCCTTTACCCTGGCAACTGCCACCCCTAT			
			101		150
HGH 362 nt (SEQ ID NO:7)	(101)	GCTTTGCCCTAATAAAAATGAAGATGCATTGTATTGCTTGGCTAGATGTAT			
HGH 324 nt (SEQ ID NO:8)	(63)	GCTTTGCCCTAATAAAAATGAAGATGCATTGTATTGCTTGGCTAGATGTAT			
HGH 189 nt (SEQ ID NO:9)	(63)	GCTTTGCCCTAATAAAAATGAAGATGCATTGTATTGCTTGGCTAGATGTAT			
HGH 113 nt (SEQ ID NO:10)	(63)	GCTTTGCCCTAATAAAAATGAAGATGCATTGTATTGCTTGGCTAGATGTAT			
			151		200
HGH 362 nt (SEQ ID NO:7)	(151)	TTCTGTTGTGGGATGGAGGGTGGTGTCAAAGAGTCCCTAGAGGCCGACATG			
HGH 324 nt (SEQ ID NO:8)	(113)	TTCTGTTGTGGGATGGAGGGTGGTGTCAAAGAGTCCCTAGAGGCCGACATG			
HGH 189 nt (SEQ ID NO:9)	(113)	TTCTGTTGTGGGATGGAGGGTGGTGTCAAAGAGTCCCTAGAGGCCGACATG			
HGH 113 nt (SEQ ID NO:10)	(113)	T-----			
			201		250
HGH 362 nt (SEQ ID NO:7)	(201)	CCTGTGGGCTGCTGGAAGAACAGCCCTGACTTTGCCTGGACCAAGTAGAG			
HGH 324 nt (SEQ ID NO:8)	(163)	CCTGTGGGCTGCTGGAAGAACAGCCCTGACTTTGCCTGGACCAAGTAGAG			
HGH 189 nt (SEQ ID NO:9)	(163)	CCTGTGGGCTGCTGGAAGAACAGCCCT-----			
HGH 113 nt (SEQ ID NO:10)	(114)	-----			
			251		300
HGH 362 nt (SEQ ID NO:7)	(251)	TCAACACATCACPTCCCTGTCTCGTGATGAGCCTGCTCCCACTCCAGAG			
HGH 324 nt (SEQ ID NO:8)	(213)	TCAACACATCACPTCCCTGTCTCGTGATGAGCCTGCTCCCACTCCAGAG			
HGH 189 nt (SEQ ID NO:9)	(190)	-----			
HGH 113 nt (SEQ ID NO:10)	(114)	-----			
			301		350
HGH 362 nt (SEQ ID NO:7)	(301)	TCAGAATCCCAGCTCTCTGGACAGTCACAAGGCGGCAAGGTCTATGTCA			
HGH 324 nt (SEQ ID NO:8)	(263)	TCAGAATCCCAGCTCTCTGGACAGTCACAAGGCGGCAAGGTCTATGTCA			
HGH 189 nt (SEQ ID NO:9)	(190)	-----			
HGH 113 nt (SEQ ID NO:10)	(114)	-----			
			351	362	
HGH 362 nt (SEQ ID NO:7)	(351)	CCCCATAAAGT			
HGH 324 nt (SEQ ID NO:8)	(313)	CCCCATAAAGT			
HGH 189 nt (SEQ ID NO:9)	(190)	-----			
HGH 113 nt (SEQ ID NO:10)	(114)	-----			

FIGURA 5

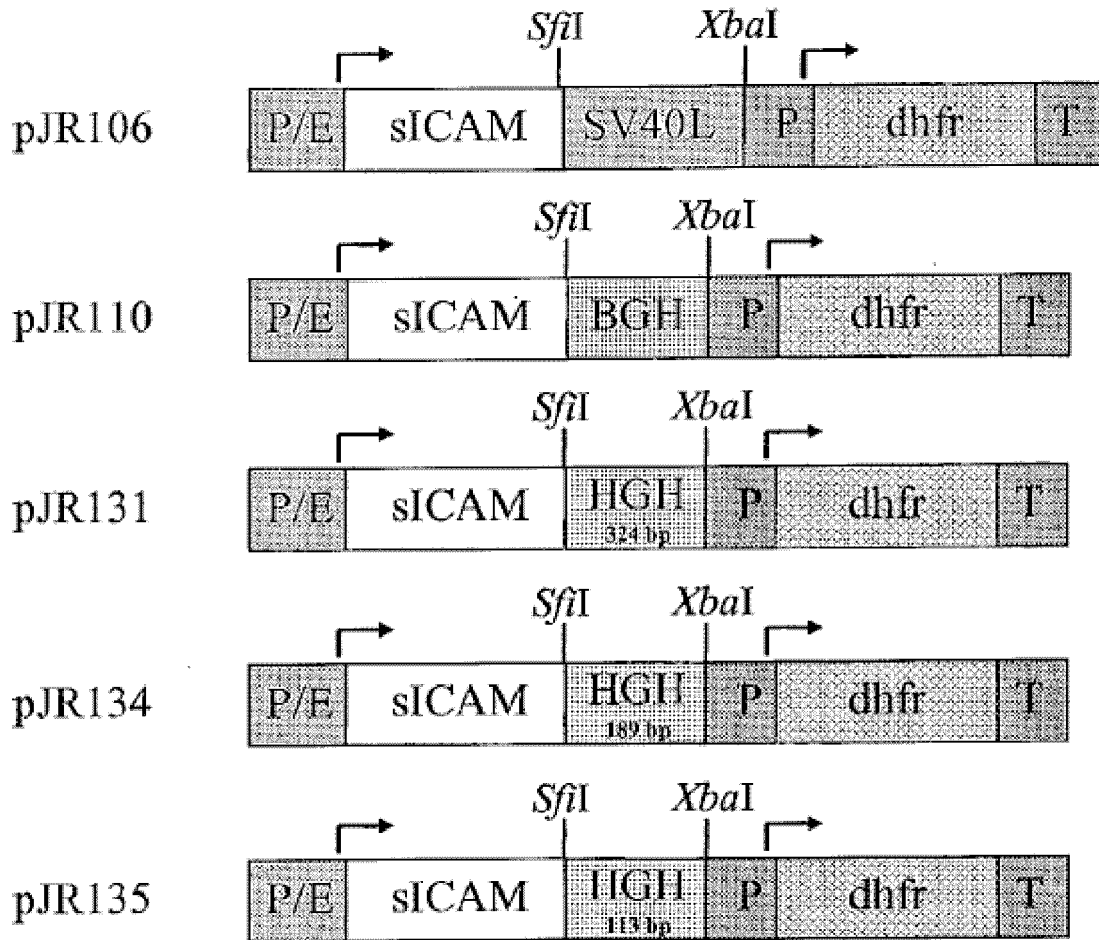


FIGURA 6

Vector	Secuencia de terminación	Transfección transitoria # 1 (n= 2)	Transfección transitoria # 2 (n= 2)
pJR106	SV40 tardía (SEQ ID NO: 11)	193 ng/mL sICAM	207 ng/mL sICAM
pJR110	BGH (SEQ ID NO: 12)	222 ng/mL sICAM	234 ng/mL sICAM
pJR131	HGH (SEQ ID NO: 8)	269 ng/mL sICAM	257 ng/mL sICAM

FIGURA 7

Combinación de vectores	Secuencia de terminación	Transfección transitoria (n= 6)
pBID-B/IgG4 + pBIN-B/kappa	BGH (SEQ ID NO: 12)	40 ± 6 ng/mL IgG4
pBID-B/IgG4 + pBIN-B/kappa	HGH (SEQ ID NO: 8)	54 ± 8 ng/mL IgG4

FIGURA 8

Vector	Secuencia de HGH (3'UTR)	Transfección transitoria # 1 (n= 2)	Transfección transitoria # 2 (n= 2)
pJR131	bp 1 - 324 (SEQ ID NO: 8)	111 ng/mL sICAM	108 ng/mL sICAM
pJR134	bp 1 - 189 (SEQ ID NO: 9)	98 ng/mL sICAM	83 ng/mL sICAM
pJR135	bp 1 - 113 (SEQ ID NO: 10)	25 ng/mL sICAM	25 ng/mL sICAM

FIGURA 9

Producto	Clones de células/ Conjunto de células	Productividad específica (pg/célula/día)	Título en proceso de lote alimentado
IgG1 (1)	clones de células	23	6,3 g/L (biorreactor de 40 L)
IgG1 (2)	conjunto de células (MTX 800 mM)	32	2,5 g/L (frasco de agitación)
IgG2	clones de células	10	2,1 g/L (biorreactor de 2 L)
IgG4	conjunto de células (MTX 800 mM)	44	2,1 g/L (frasco de agitación)
Proteína de fusión Fc (IgG1)	conjunto de células (MTX 800 mM)	45	3,9 g/L (frasco de agitación)
Proteína de fusión Fc (IgG2)	conjunto de células (MTX 800 mM)	34	3,4 g/L (frasco de agitación)