

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 620 325**

51 Int. Cl.:

C12N 15/85 (2006.01)

C12N 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.11.2013 PCT/EP2013/074114**

87 Fecha y número de publicación internacional: **30.05.2014 WO2014079819**

96 Fecha de presentación y número de la solicitud europea: **18.11.2013 E 13792030 (2)**

97 Fecha y número de publicación de la concesión europea: **21.12.2016 EP 2922962**

54 Título: **Casete de expresión optimizado para expresar un polipéptido con alto rendimiento**

30 Prioridad:
20.11.2012 US 201261728459 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
28.06.2017

73 Titular/es:
**NOVARTIS AG (100.0%)
Lichtstrasse 35
4056 Basel, CH**

72 Inventor/es:
**DRAGIC, ZORICA;
SCHMITZ, RITA;
WILMS, BURKHARD y
GEISSE, SABINE**

74 Agente/Representante:
CARVAJAL Y URQUIJO, Isabel

ES 2 620 325 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

Casete de expresión optimizado para expresar un polipéptido con alto rendimiento

Campo de la invención

5 La presente invención se refiere, entre otras cosas, a un casete de expresión adecuado para expresar un polipéptido de interés donde dicho casete de expresión comprende una combinación de una 5'UTR específica y la secuencia líder secretora de hCD33. Se descubrió que esta combinación específica de elementos genéticos tiene como resultado, sorprendentemente, un nivel de expresión significativamente mejor del polipéptido de interés en comparación con las combinaciones de otras 5'UTR con otras secuencias líder secretoras. Por lo tanto, utilizar este casete de expresión es ventajoso para producir un polipéptido de interés con alto rendimiento.

Antecedentes de la invención

15 La capacidad para clonar y expresar un polipéptido de interés en grandes cantidades se ha vuelto cada vez más importante. La capacidad de producir y purificar altos niveles de proteínas es en particular importante en el campo biotecnológico y farmacéutico humano, por ejemplo para producir fármacos de proteína así como en el ámbito de investigación básica, por ejemplo para cristalizar proteínas para permitir la determinación de su estructura tridimensional. Las proteínas que son, de otro modo, difíciles de obtener en cantidad se pueden sobreexpresar en una célula huésped y posteriormente aislar y purificar.

20 La elección de un sistema de expresión para la producción de proteínas recombinantes depende de muchos factores, que incluyen características de crecimiento celular, niveles de expresión, expresión intracelular y extracelular, modificaciones postraduccionales y actividad biológica de la proteína de interés, así como asuntos reguladores y consideraciones económicas en la producción de proteínas terapéuticas. Las ventajas clave de las células de mamífero sobre otros sistemas de expresión tal como bacteria o levadura son la capacidad de llevar a cabo un plegamiento de proteína apropiado, glicosilación unida a *N* complejo y glicosilación unida a *O* auténtica, así como un amplio espectro de otras modificaciones postraduccionales. Debido a las ventajas descritas, las células eucariotas y, en particular, de mamífero son actualmente el sistema de expresión de elección para producir proteínas terapéuticas complejas tales como anticuerpos monoclonales.

30 El enfoque más común para obtener células huésped de expresión alta (también llamadas de producción alta) es para generar un vector de expresión apropiado para expresar el producto de interés como un primer paso. El vector de expresión impulsa la expresión del polinucleótido que codifica el producto de interés en la célula huésped y usualmente comprende al menos un marcador seleccionable para generar la línea celular recombinante. Los vectores de expresión utilizados para expresar un polipéptido en una célula huésped usualmente comprende además del polinucleótido que codifica la proteína de interés elementos de control transcripcional adecuados para impulsar la transcripción tal como, por ejemplo, promotores, potenciadores, señales de poliadenilación, interrupción de la transcripción o señales de terminación como elemento de un casete de expresión. Además, los elementos de control traduccional adecuados están usualmente incluidos y unidos de forma operativa a los polinucleótidos para ser expresados tales como, por ejemplo, 5'UTR y 3'UTR apropiadas.

40 Para aumentar la eficacia de tal sistema de expresión, se optimizan diferentes elementos, especialmente las secuencias de ADN que contribuyen a la eficiencia de la transcripción y traducción, síntesis de la proteína, plegamiento correcto en ER y secreción de la proteína. Los sistemas de expresión de alto rendimiento sin traducción optimizada y componentes de secreción podrían llevar de forma potencial, a la inestabilidad del ARNm, a una secreción de proteína insuficiente y acumulación de proteína mal plegada (inactiva) en la membrana o citosol de la célula. Por lo tanto, los sistemas de expresión que permiten la traducción estable y consistente y la secreción de proteínas correctamente plegadas en el medio de cultivo celular son de particular interés. Tales sistemas secretadores ofrecen las ventajas de una traducción de ARNm estable y eficaz; plegamiento de proteína correcto y secreción eficaz, procedimientos de purificación del producto simples y rápidos, así como un rendimiento aumentado en comparación con los sistemas citosólicos. Sin embargo, los rendimientos del producto de la mayoría de los sistemas secretadores disponibles aun no se encuentran completamente optimizados. Para mejorar la eficiencia de la productividad y secreción, un objetivo es optimizar las señales de secreción (también denominadas en la presente como péptido señal o secuencia líder secretora), así como su combinación con diferentes secuencias 5'UTR para obtener una combinación de elementos genéticos que tiene como resultado el alto nivel de expresión deseado.

55 La mayoría de las proteínas secretadas y unidas a la membrana de los organismos procariontes o eucariotas poseen un péptido líder de extremo amino (también referido como secuencia líder secretora o péptido señal) que se escinde del polipéptido precursor naciente durante la biosíntesis. Los péptidos líder secretadores tienen usualmente 5 a 60 aminoácidos de largo. Esta secuencia es necesaria y suficiente para la secreción. Los análisis de un gran número de estos péptidos líder secretadores han revelado un motivo estructural común que

ocurre en la ausencia de homología de secuencia de aminoácidos significativa [Von Heijne, 1981; Perlman et al, 1983]. En general, una secuencia líder secretora consiste en extremos amino cargados positivamente (n), un núcleo hidrofóbico (h) y más extremos carboxi polares (c) que define el sitio de escisión de la señal peptidasa. La región "n" del péptido líder secretor tiene alrededor de 5 a 8 aminoácidos de largo y se caracteriza por la presencia de residuos básicos. La región "h" contiene 8 a 12 aminoácidos no polares que están compuestos en promedio de 37 % de leucina, 15 % de alanina, 10 % de valina, 10 % de fenilalanina, 7 % de isoleucina y 21 % de aminoácidos hidrofóbicos tales como glicina, metionina, prolina o triptófano. Esta región tiene una alta propensión a la formación alfa-hélice, una conformación que puede facilitar la interacción con el interior de la bicapa lípida. Los estudios sobre las características estructurales de los péptidos líder secretores, basados principalmente en proteínas bacterianas, han sugerido que la región "h" es fundamental para la función de la secuencia señal [Gierasch, 1990]. La interrupción de la región h por eliminación o por reemplazo de los residuos hidrofóbicos con aminoácidos hidrofílicos o de carga lleva a la pérdida de la función de la señal, mientras que las alteraciones a la región "n" tienen poco efecto [Bird et al, 1990]. El extremo carboxi, o la región de escisión, tiene típicamente alrededor de 6 aminoácidos de largo. Esta región está implicada en el reconocimiento de la señal peptidasa y la escisión, que es usualmente requerida para alcanzar el plegamiento y la secreción final de la proteína.

La 5'UTR (región no traducida 5') es una parte de la secuencia de ADN que se transcribe en ARNm, pero no en proteína. Usualmente comienza en el inicio de la transcripción, y termina un nucleótido antes del codón de comienzo. Una 5'UTR puede contener secuencias que regulan la eficiencia de la traducción o la estabilidad del ARNm, sitios de unión para proteínas, elementos reguladores, y secuencias que promueven el inicio de la traducción. Las secuencias de 5'UTR pueden variar en su longitud y pueden comprender un par de decenas de nucleótidos hasta un par de cientos o incluso miles de nucleótidos. En las eucariotas, la longitud media de las 5'UTR es de aproximadamente 100 a 200 nt. El papel específico de las 5'UTR y sus elementos aun no se han aclarado completamente, parcialmente también porque la secuencia no está traducida en la proteína funcional. Sin embargo, se sabe que la combinación de una 5'UTR específica y una secuencia líder secretora específica puede mejorar intensamente la eficiencia de la traducción y secreción del sistema de producción y por lo tanto puede aumentar el rendimiento de la expresión. Sin embargo, como una plétora de 5'UTR y secuencias líder secretoras se encuentran disponibles, es un desafío obtener una combinación eficiente que realmente mejore la expresión. Por lo tanto, a pesar de la plétora de casetes de expresión y vectores de expresión disponibles, obtener una fuerte producción de polipéptidos/proteínas con un alto rendimiento en células eucariotas es aun desafiante.

Por lo tanto, es el propósito de la presente invención proporcionar un casete de expresión que permite la secreción de un polipéptido de interés con alto rendimiento cuando dicho casete de expresión se introduce en una célula huésped. Además, es un propósito de la presente invención proporcionar un vector de expresión que permite la expresión de un polipéptido de interés con alto rendimiento. Además, es un propósito de la presente invención proporcionar un método adecuado para expresar un polipéptido de interés con alto rendimiento.

Resumen de la invención

La presente invención se basa en el descubrimiento de que la combinación de una secuencia de polinucleótidos de 5'UTR específicos (véase SEQ ID NO 1, que también está contenida en la SEQ ID NO 2, 3 y 4 y también 5, 6 y 7) y la secuencia líder secretora de CD33 humana en un casete de expresión tiene como resultado una expresión sorprendentemente alta del polipéptido de interés que está codificada por dicho casete de expresión. En varios ejemplos, se mostró que el casete de expresión, de acuerdo con la presente invención, es superior sobre los casetes de expresión conocidos en la técnica previa dado que el rendimiento de expresión se podría aumentar hasta 4 veces. El rendimiento superior del casete de expresión de acuerdo con la presente invención se confirmó en numerosos experimentos donde los diferentes polipéptidos de interés se expresaron a partir de dicho casete de expresión. Por lo tanto, el casete de expresión de acuerdo con la presente invención es particularmente ventajoso para producir un polipéptido de interés con alto rendimiento. De manera adicional, se encontró que el procesamiento correcto de la secuencia líder se puede mejorar. Por lo tanto, la presente invención proporciona una contribución valiosa a la técnica, ya que este diseño novedoso del casete de expresión mejora de forma considerable la expresión del polipéptido de interés.

De acuerdo con un primer aspecto, la presente invención proporciona un casete de expresión para expresar un polipéptido de interés que comprende:

- a) un promotor
- b) una secuencia de polinucleótidos de 5'UTR, donde dicha secuencia de polinucleótidos de 5'UTR se selecciona del grupo que consiste en
 - i) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacgggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccccccccttggttcgttagaacgcggct
acaattaatacataaccttatgtatcatacacatacagatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgca
(SEQ ID NO 1);

ii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacgggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccccccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacagatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cg (SEQ ID NO 2);

iii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacgggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccccccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacagatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgaaaacgcgtccacc (SEQ ID NO 3);

5

iv) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacgggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccccccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacagatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgcgattgaattccccgggatcctctagggtgaccgTTTggtgccgccacc
(SEQ ID NO 4);

v) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgcg

taagtaccgcctatagagtctataggcccacccccttggttcggttagaacgcggctaca
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgaaaacgcgcctctagagccgccacc (SEQ ID NO 5);

vi) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccacccccttggttcggttagaacgcggct
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggt
tctatcgaaaaacgcgtgcccgccacc (SEQ ID NO 6);

vii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccacccccttggttcggttagaacgcggct
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggt
tctatcgaaaaacgcgtgcccgccacc (SEQ ID NO 7);

5

viii) una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a una secuencia mostrada en la SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3 o SEQ ID NO 4 o SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7;

c) un polinucleótido que codifica una secuencia líder secretora de hCD33; y

10 d) un polinucleótido que codifica un polipéptido de interés o un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés.

De acuerdo a un segundo aspecto, la presente invención proporciona un vector de expresión para expresar un polipéptido de interés, que comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención.

15 De acuerdo a un tercer aspecto, la presente invención proporciona una célula huésped eucariota que comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención y/o al menos un vector de expresión de acuerdo con el segundo aspecto de la presente invención.

20 De acuerdo a un cuarto aspecto, la presente invención proporciona un método para producir la célula huésped de acuerdo con el tercer aspecto de la presente invención, donde un vector de expresión de acuerdo con el segundo aspecto de la presente invención se introduce en una célula huésped eucariota.

De acuerdo a un quinto aspecto, la presente invención proporciona un método para producir un polipéptido de interés, dicho método comprende el cultivo de células huésped de acuerdo con el tercer aspecto de la presente invención en un cultivo celular en condiciones que permiten la expresión de dicho polipéptido de interés.

25 De acuerdo a un sexto aspecto, la presente invención pertenece a un uso de una secuencia de 5'UTR en

combinación con una secuencia líder secretora de hCD33 en un casete de expresión para expresar un polipéptido de interés con alto rendimiento de dicho casete de expresión, donde dicha secuencia de polinucleótidos de 5'UTR se selecciona del grupo que consiste en

i) una secuencia de polinucleótidos de 5'UTR que comprende la secuencia

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga  
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga  
cgtaagtaccgcctatagagtctataggccccaccccccttggttcgttagaacgcggct  
acaattaatacataaccttatgtatcacatacagatttaggtgacactatagaata  
acatccactttgcctttctctccacaggtgtccactcccagggtccaactgca
```

(SEQ ID NO 1);

5

ii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggccccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cg (SEQ ID NO 2);
```

iii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggccccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cgaaaacgcgtccacc (SEQ ID NO 3);
```

10 iv) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg
```

```
taagtaccgcctatagagtctataggccccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cgcgattgaattccccggggatcctctagggtgaccgtttggtgcccacc
```

(SEQ ID NO 4);

v) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccccccccttggttcggttagaacgcggctaca
attaatacataaccttatgtatcatacacatacagatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat
cgaaaacgcgcctctagagccgccacc (SEQ ID NO 5);

vi) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccccccccttggttcggttagaacgcggct
acaattaatacataaccttatgtatcatacacatacagatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggt
tctatcgaaaaacgcgtgccgccacc
(SEQ ID NO 6);

vii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccccccccttggttcggttagaacgcggct
acaattaatacataaccttatgtatcatacacatacagatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggt
tctatcgaaaaacgcgtgccgccacc (SEQ ID NO 7);

5
viii) una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a una secuencia mostrada en la SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3 o SEQ ID NO 4 o SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7.

10 Otros objetos, características, ventajas y aspectos de la presente solicitud serán evidentes para los expertos en la técnica a partir de la siguiente descripción y de las reivindicaciones adjuntas. Se debería comprender, sin embargo, que la siguiente descripción, reivindicaciones adjuntas y ejemplos específicos, en tanto indican realizaciones preferidas de la solicitud, se dan únicamente a modo de ilustración. Diversos cambios y modificaciones dentro del espíritu y alcance de la invención divulgada se volverán fácilmente evidentes para los expertos en la técnica al leer lo siguiente.

15 Descripción detallada de la invención

La presente invención proporciona un casete de expresión para expresar un polipéptido de interés que comprende una combinación nueva de elementos genéticos, a saber, una secuencia de polinucleótidos de 5'UTR específica y la secuencia líder secretora de hCD33. Esta combinación específica de elementos genéticos tiene como resultado un aumento significativo en la eficiencia de la expresión del polipéptido codificado de interés. Además, se encontró que el procesamiento de la secuencia líder se puede mejorar para que se puedan reducir las extensiones o truncamientos de secuencias no deseadas (también denominadas "pinzamientos") debido al procesamiento erróneo de la secuencia líder.

20 De acuerdo con un primer aspecto, se proporciona un casete de expresión para expresar un polipéptido de interés que comprende:

25 a) un promotor

b) una secuencia de polinucleótidos de 5'UTR, donde dicha secuencia de polinucleótidos de 5'UTR se selecciona del grupo que consiste en

i) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggct
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata
acatccaactttgcctttctctccacaggtgtccactcccaggtccaactgca

(SEQ ID NO 1);

5 ii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cg (SEQ ID NO 2);

iii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg

taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgaaaacgcgtccacc (SEQ ID NO 3);

iv) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgcgattgaattccccggggatcctctagggtgaccgtttggtgcccacc

(SEQ ID NO 4);

10

v) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcgccgggaacggtgcattggaacgcggtattccccgtgccaagagtgcg
taagtaccgctatagagtctataggcccccccccttggttcgttagaacgcggtaca
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cgaaaacgcgctctagagccgccacc (SEQ ID NO 5);

vi) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccggg
accgatccagcctccgcgccgggaacggtgcattggaacgcggtattccccgtg
ccaagagtgcgtaagtaccgctatagagtctataggcccccccccttggttc
cgttagaacgcggtacaattaatacataaccttatgtatcatacacatacgat
ttaggtgacactatagaataacatccactttgcctttctctccacaggtgtcca
ctcccaggtccaactgcacctcggttctatcgaaaaacgcggtgccgccacc
(SEQ ID NO 6);

vii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcgccgggaacggtgcattggaacgcggtattccccgtgccaagagtga
cgtaagtaccgctatagagtctataggcccccccccttggttcgttagaacgcggt
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggt
tctatcgaaaacgcggtgccgccacc (SEQ ID NO 7);

5

viii) una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a una secuencia mostrada en la SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3 o SEQ ID NO 4 o SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7;

c) un polinucleótido que codifica una secuencia líder secretora de hCD33; y

10 d) un polinucleótido que codifica un polipéptido de interés o un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés.

15 El término "casete de expresión" tal como se usa en la presente en particular se refiere a un segmento de ADN que es capaz, en un entorno apropiado, de conducir la expresión de un polinucleótido que codifica un polipéptido de interés que está incorporado en dicho casete de expresión. Cuando se introduce en una célula huésped, un casete de expresión, entre otros, es capaz de dirigir el mecanismo de la célula para transcribir un polinucleótido incorporado que codifica un polipéptido de interés en ARN, el cual luego se procesa adicionalmente y finalmente se traduce en el polipéptido de interés. El casete de expresión se puede comprimir en un vector de expresión tal como se describirá en mayor detalle más adelante. Los elementos individuales del casete de expresión de acuerdo con la presente invención se explican en detalle posteriormente.

20

25 El casete de expresión de acuerdo con la presente invención comprende como elemento a) un promotor. El término "promotor" tal como se usa en la presente en particular se refiere a un elemento de ADN que facilita la transcripción de un polinucleótido al que el promotor está unido de forma operativa. El promotor también puede formar parte de un elemento promotor/potenciador. Aunque los límites físicos entre los elementos "promotor" y "potenciador" no son siempre claros, el término "promotor" generalmente se refiere a un sitio en la molécula de ácido nucleico al cual se une una polimerasa de ARN y/o cualquier factor asociado y en el cual

5 se inicia la transcripción. Los potenciadores potencian la actividad del promotor, de forma temporal y espacial. Se conocen muchos promotores en la técnica previa que tienen actividad transcripcional en un amplio rango de tipos de células. Los promotores se pueden dividir en dos clases, aquellos que funcionan de forma constitutiva y aquellos que se regulan por inducción o desrepresión. Ambas clases son adecuadas para la expresión de la proteína. Los promotores que se utilizan para la producción de polipéptidos de alto nivel en células eucariotas y en particular de mamífero deberían ser fuertes y preferentemente deberían estar activas en un amplio rango de tipos de células. Los promotores constitutivos fuertes que son capaces de llevar la expresión en muchos tipos de células son conocidos en la técnica previa y por lo tanto, no necesitan ser descritos en detalles en la presente. Preferentemente, se utiliza un promotor de citomegalovirus (CMV) de acuerdo con las enseñanzas de la presente invención. Un promotor o promotor/potenciador derivado de la región inmediata temprana (IE, por sus siglas en inglés) del citomegalovirus humano (hCMV, por sus siglas en inglés) es particularmente adecuado como promotor en el casete de expresión de acuerdo con la presente invención. La región inmediata temprana (IE) del citomegalovirus humano (hCMV) y fragmentos que promueven la expresión funcional y/o fragmentos que aumentan la expresión funcional derivada de este están por ejemplo, descritos en EP 0 173 177 y EP 0 323 997 y también son conocidos en la técnica previa. Por lo tanto, varios fragmentos de la región inmediata temprana (IE) de hCMV se pueden utilizar como promotor y/o promotor/potenciador. De acuerdo con una realización, se utiliza un promotor de CMV humano en el casete de expresión de acuerdo con la presente invención. El término "un promotor de CMV humano" tal como se usa en la presente en particular se refiere a un promotor que deriva de la región inmediata temprana (IE) del hCMV.

De acuerdo con una realización, se utiliza un promotor de CMV humano que se selecciona del grupo que consiste en:

a) un promotor que comprende la siguiente secuencia:

```
tcaatattgg    ccattagcca    tattattcat    tggttatata    gcataaatca
atattggcta    ttggccattg    catacgttgt    atctatatca    taatatgtac
atztatattg    gctcatgtcc    aatatgaccg    ccatgttggc    attgattatt
gactagttat    taatagtaat    caattacggg    gtcattagtt    catagcccat
atatggagtt    ccgcgttaca    taacttacgg    taaatggccc    gcctggctga
ccgccaacg    acccccgccc    attgacgtca    ataatgacgt    atgttcccat
agtaacgcca    atagggactt    tccattgacg    tcaatgggtg    gagtatttac
ggtaaactgc    ccaacttgga    gtacatcaag    tgtatcatat    gccaagtacg
ccccctattg    acgtcaatga    cggtaaattg    cccgcctggc    attatgcca
gtacatgacc    ttatgggact    ttccacttg    gcagtacatc    tacgtattag
tcatcgctat    taccatgggtg    atgcggtttt    ggcagtacat    caatgggcgt
ggatagcggg    ttgactcaag    gggatttcca    agtctccacc    ccattgacgt
caatgggagt    ttgttttggc    accaaaatca    acgggacttt    ccaaaatgtc
gtaacaactc    cgccccattg    acgcaaattg    gcggtaggcg    tgtacggtgg
gaggctctata    taagcagagc    tcgttttagtg    aaccgtc (SEQ ID NO 8);
```

25 o un fragmento funcional de este que funciona como promotor;

b) un promotor que comprende una secuencia que es al menos 80 %, preferentemente al menos 85 %, más preferentemente al menos 90 %, más preferentemente al menos 95 %, más preferentemente al menos 97 %, más preferentemente al menos 99 % idéntica a la secuencia mostrada en la SEQ ID NO 8 o un fragmento funcional de este que funciona como promotor.

30 El % de identidad se puede calcular sobre la longitud total de la secuencia de referencia.

Se encontró que un promotor respectivo funciona particularmente bien en combinación con la secuencia de polinucleótidos de 5'UTR específica seleccionada del grupo anterior y de la secuencia señal hCD33 tal como se enseña en la presente invención.

35 El casete de expresión de acuerdo con la presente invención comprende además como elemento b) una secuencia de polinucleótidos de 5'UTR (región cinco prima no traducida) específica. El término una

5 "secuencia de polinucleótidos de 5'UTR" tal como se usa en la presente, en particular se refiere a una secuencia de ADN que se transcribe en ARNm pero luego no se traduce en un polipéptido. Una secuencia de polinucleótidos de 5'UTR generalmente comienza en el sitio de inicio de transcripción y finaliza antes del codón de inicio de la región de codificación actual. En el casete de expresión de acuerdo con la presente invención, la secuencia de polinucleótidos de 5'UTR finaliza antes de que comience el polinucleótido que codifica la secuencia líder secretora de hCD33.

La secuencia de polinucleótidos de 5'UTR utilizada en el casete de expresión de acuerdo con la presente invención preferentemente comprende la secuencia mostrada en la SEQ ID NO 1. Se describe una secuencia de polinucleótidos respectiva de 5'UTR en WO 2009/080720.

10 De acuerdo con una realización, la secuencia de polinucleótidos de 5'UTR comprende una secuencia seleccionada de las secuencias SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3 y SEQ ID NO 4 o comprende una secuencia que se selecciona de las secuencias SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7. La secuencia de 5' UTR que se muestra en la SEQ ID NO 1 es una secuencia consenso que también está comprendida en las secuencias de 5' UTR mostradas como SEQ ID NO 2, 3 y 4 así como en las secuencias de 5'UTR mostradas como SEQ ID NO 5, 6 y 7. Preferentemente, la secuencia de 5'UTR que comprende o consiste en la SEQ ID NO 3 se utiliza para expresar la cadena ligera de un anticuerpo o un fragmento funcional de esta y la secuencia de 5'UTR comprende o consiste en la SEQ ID NO 1 o la SEQ ID NO 4 se utiliza para expresar la cadena pesada de un anticuerpo o un fragmento funcional de esta. La secuencia de 5'UTR, que comprende o consiste en la SEQ ID NO: 5 de acuerdo con una realización, es utilizada para expresar la cadena ligera de un anticuerpo o un fragmento funcional de este. La SEQ ID NO 5 también se puede utilizar por ejemplo para la expresión de un nanocuerpo como una realización de ejemplo. La secuencia de 5'UTR, que comprende o consiste en la SEQ ID NO: 6 de acuerdo con una realización, es utilizada para expresar la cadena pesada de un anticuerpo o un fragmento funcional de esta. La secuencia de 5'UTR, que comprende o consiste en la SEQ ID NO: 7 de acuerdo con una realización, es utilizada para expresar la cadena pesada de un anticuerpo o un fragmento funcional de esta. También se probó en los ejemplos.

Se descubrió que la 5'UTR utilizada de acuerdo con la presente invención se compone de varios elementos y comprende una secuencia 5' de la región flanqueante del intrón, un intrón y una secuencia 3' de la región flanqueante del intrón. El intrón y las regiones flanqueantes del intrón comprendidas en dicha 5'UTR se originan de la cadena pesada de IgG de ratón (véase, por ejemplo, Eaton et al., 1986, Biochemistry 25, 8343-8347, Neuberger et al., 1983, EMBO J. 2(8), 1373-1378; puede obtenerse del vector pRK-5 (BD PharMingen)). La siguiente tabla ilustra los límites putativos de dichos elementos que están comprendidos en la 5'UTR que puede usarse de acuerdo con la presente invención:

Secuencia 5' de región flanqueante del intrón	agatcgcctggagacgccatccacgctgttttgacctccatagaagacacc gggaccgatccagcctccgcgccgggaacgggtgcattggaacgcggattc cccqtaccaaaqaatqac (SEQ ID NO 9)
Intrón	gtaagtaccgcctatagagtctataggccccaccccttgcttcgtagaa cgcggtacaattaatacataaccttatgtatcatacacatacgatttagg tgacactatagaataacatccactttgcctttctctccacag (SEQ ID NO 10)
Secuencia 3' de región flanqueante del intrón	gtgtccactcccagggtccaactgca (SEQ ID NO 11)

35 Además, puede usarse una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 %, preferentemente al menos 95 %, más preferido al menos 98 %, más preferido al menos 99 % idéntica a la secuencia mostrada en la SEQ ID NO: 1, 2, 3 o 4 o la secuencia mostrada en la SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7. La identidad puede calcularse sobre la longitud total de la secuencia de 5'UTR de referencia.

40 En numerosos experimentos los inventores encontraron que dichas secuencias de polinucleótidos de 5'UTR particulares tienen, en combinación con la secuencia señal hCD33, los efectos ventajosos descritos anteriormente en el rendimiento de expresión que se aumenta sustancialmente comparado con los casetes de expresión de la técnica previa.

45 El casete de expresión de acuerdo con la presente invención comprende además como elemento c) un polinucleótido que codifica una secuencia líder secretora de hCD33. El CD33 humano es un miembro de los receptores inhibidores de lectina similares a la inmunoglobulina que unen ácido siálico. El CD33 es una glicoproteína de transmembrana 67-kDa y se expresa específicamente en el linaje mielóide. El término "un polinucleótido que codifica una secuencia líder secretora de hCD33" tal como se usa en la presente hace

referencia a un polinucleótido que codifica una secuencia líder secretora que comprende la secuencia líder secretora hCD33. Por lo tanto, luego de la transcripción y posterior traducción se obtiene una secuencia líder secretora que comprende y preferentemente consiste en una secuencia líder secretora de hCD33. Dicha secuencia líder secretora produce el efecto de secretar, de forma eficaz, un polipéptido de interés fusionado a esta desde una célula huésped. La secuencia líder secretora se escinde durante la biosíntesis. Como se trató anteriormente, los inventores descubrieron sorprendentemente que la combinación específica de la secuencia de polinucleótidos de 5'UTR descrita anteriormente con una secuencia líder secretora de hCD33 da como resultado un aumento significativo de la expresión de polipéptidos. Por lo tanto, la combinación de acuerdo con la presente invención es superior a otras combinaciones que usan la misma 5'UTR en conjunto con otras secuencias líderes secretoras que no comprenden una secuencia líder secretora de hCD33 pero, por ejemplo, una secuencia líder secretora de inmunoglobulina como se describe en WO2009/080720. De acuerdo con una realización, la secuencia líder secretora de hCD33 utilizada en el casete de expresión de acuerdo con la presente invención comprende la siguiente secuencia de aminoácidos:

MPLLLLLLPLLWAGALA (SEQ ID NO 12)

Se describen diferentes secuencias de aminoácidos en la bibliografía para la secuencia líder secretora de hCD33. La mayoría de los documentos describen que la secuencia líder secretora de hCD33 consiste en la siguiente secuencia de aminoácidos

MPLLLLLLPLLWAGALAMD (SEQ ID NO 13).

Como es evidente, la SEQ ID NO 13 comprende dos aminoácidos adicionales en el extremo 3' en comparación con SEQ ID NO 12. De acuerdo con una realización, el polinucleótido que codifica una secuencia líder secretora de hCD33 codifica una secuencia líder secretora que consiste en la secuencia de aminoácidos mostrada en SEQ ID NO 12 o SEQ ID NO 13. Sin embargo, se prefiere utilizar la secuencia líder secretora de hCD33 más corta mostrada en SEQ ID NO 12.

Se asume que la secuencia de señal apenas más corta mejora la probabilidad de una escisión correcta de peptidasa, mejora el procesamiento correcto y, de este modo, asegura la calidad del producto. Utilizando la secuencia líder secretora de CD33 más larga mostrada en SEQ ID NO 13 podría suponer el riesgo de que el polipéptido de interés tenga aminoácidos adicionales en su extremo N luego de que la secuencia líder secretora se escindió, lo que es inaceptable en particular cuando se expresan polipéptidos farmacéuticos. El riesgo se evita cuando se utiliza la secuencia líder secretora de hCD33 mostrada en SEQ ID NO 12. Además, la secuencia más corta es aparentemente más eficaz para la secreción debido a una mejor distribución de la carga, en particular cuando se utiliza en combinación con la secuencia de polinucleótidos de 5'UTR de acuerdo con la presente invención. Por lo tanto, preferentemente, el polinucleótido codifica una secuencia líder secretora de hCD33 que consiste de la secuencia de aminoácidos mostrada en SEQ ID NO 12.

En el casete de expresión, el promotor (elemento a) se dispone 5' de la secuencia de polinucleótidos de 5'UTR (elemento b) que, a su vez, se dispone 5' del polinucleótido que codifica la secuencia líder secretora de hCD33 (elemento c). Dichos elementos están operativamente unidos en el casete de expresión con el fin de permitir una expresión eficaz de un polipéptido de interés si un polinucleótido que codifica un polipéptido de interés se inserta dentro de dicho casete de expresión.

El casete de expresión de acuerdo con la presente invención es "adecuado para expresar un polipéptido de interés". Este término en particular describe que un polipéptido es expresado de dicho casete de expresión si un polinucleótido que codifica dicho polipéptido se inserta dentro de dicho casete de expresión. Por lo tanto, de acuerdo con una realización, el casete de expresión de acuerdo con la presente invención comprende como elemento d) un polinucleótido que codifica un polipéptido de interés. Por lo tanto, dicho polinucleótido comprende la región codificadora del polipéptido de interés. Preferentemente, dicho polinucleótido es o se deriva de un ADNc. Preferentemente, dicho polinucleótido no comprende una secuencia líder secretora adicional. Dicho polinucleótido que codifica el polipéptido de interés se ubica 3' del polinucleótido que codifica la secuencia líder secretora de hCD33, por lo que, luego de la expresión, se obtiene un polipéptido de fusión que comprende la secuencia líder secretora de hCD33 y el polipéptido de interés. La 5'UTR y la secuencia líder secretora que se usan en el casete de expresión, de acuerdo con la presente invención, son heterólogas al polinucleótido que codifica el polipéptido de interés, es decir, no están naturalmente asociadas con dicho polinucleótido. La expresión del polinucleótido que codifica el polipéptido de interés se encuentra bajo el control postranscripcional de dicha 5'UTR y dicha secuencia líder secretora que comprende y preferentemente consiste en la secuencia líder hCD33.

De acuerdo con una realización alternativa, el casete de expresión comprende un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés, sin embargo, no comprende todavía un polinucleótido que codifica un polipéptido de interés. Dicho sitio de inserción se ubica 3' del polinucleótido que codifica la secuencia líder secretora de hCD33. Para este propósito, el casete de expresión puede comprender, por ejemplo, un sitio de clonación múltiple (MCS, por sus siglas en inglés) que puede, por ejemplo, usarse en todos los marcos de lectura. Un casete de expresión respectivo "vacío" que comprende un

sitio de inserción puede usarse para insertar un polinucleótido que codifica un polipéptido de interés deseado.

Esto tiene la ventaja que un cliente pueda insertar un polinucleótido que codifica un polipéptido deseado en dicho sitio de inserción completando, de esta manera, el casete de expresión. Luego de la expresión, el polipéptido deseado fusionado a la secuencia líder secretora de acuerdo con la presente invención se expresa de dicho casete de expresión completo y se secreta. Por lo tanto, un casete de expresión respectivo "vacío" proporciona una herramienta útil para expresar diferentes polipéptidos de interés dado que el casete de expresión puede adaptarse fácilmente al uso que se le pretende dar simplemente mediante la inserción del polinucleótido que codifica el polipéptido de interés deseado en el sitio de inserción.

Además, el casete de expresión puede comprender un sitio de terminación de transcripción adecuado. Los sitios de terminación de transcripción son ampliamente caracterizados en la técnica previa y su incorporación en los casetes de expresión ha demostrado tener efectos beneficiosos múltiples en la expresión génica. Como se describirá en mayor detalle más adelante, el casete de expresión, de acuerdo con la presente invención, pretende utilizarse en particular para expresar un polipéptido de interés en una célula huésped eucariota. La mayoría de ARNm nacientes eucariotas tienen una cola poli A en su extremo 3' que se agrega durante un proceso complejo que involucra la escisión de la transcripción primaria y una reacción de poliadenilación acoplada. La cola de poli A es, entre otros, ventajosa para la estabilidad del ARNm. Por lo tanto, el casete de expresión comprende preferentemente/proporciona respectivamente, un sitio de poliadenilación adecuado para la terminación de la transcripción y la poliadenilación. Existen varias señales de poliA eficaces que pueden usarse en casetes de expresión que se pretende utilizar para la expresión en células eucariotas, que incluyen aquellas derivadas de la hormona de crecimiento bovina (bgh, por sus siglas en inglés), beta globina de ratón, la unidad de transcripción temprana de SV40 y el gen de timidina cinasa del virus del herpes simple. Sin embargo, también se conocen sitios de poliadenilación sintéticos (véase, por ejemplo, el vector de expresión de pCI-neo de Promega que se basa en Levitt et al, 1989, Genes Dev. 3, (7): 1019-1025). Por lo tanto, el sitio de poliadenilación que se utiliza en el casete de expresión de acuerdo con la presente invención puede seleccionarse del grupo que consiste en el sitio poliASV40, tal como el sitio poli-A temprano y tardío de SV40 (véase, por ejemplo, el plásmido pSV2-DHFR como se describe en Subramani et al, 1981, Mol.Cell. Biol. 854-864), un sitio poliA sintético (véase, por ejemplo, el vector de expresión pCI-neo de Promega basado en Levitt et al, 1989, Genes Dev. 3, (7): 1019-1025) y un sitio de poliA de bgh (hormona de crecimiento bovino). El sitio de terminación de transcripción puede proporcionarse junto con el polinucleótido que codifica el polipéptido de interés o puede proporcionarse como un elemento separado en el casete de expresión.

Por lo tanto, de acuerdo con una realización, el casete de expresión comprende adicionalmente una secuencia de 3'UTR como elemento e). Una región no traducida tres prima (3' UTR) sigue la región codificadora y comprende los elementos reguladores. Los elementos que pueden estar comprendidos en dicha 3'UTR incluyen pero no se limitan a los sitios de unión para proteínas y/o moléculas que inducen el ARNi, tal como miARN. De acuerdo con una realización, se utiliza una 3'UTR que comprende la siguiente secuencia:

ggcgccgcttcccttagtgagggttaatgcttcgag (SEQ ID NO 14).

Además, el casete de expresión puede comprender una señal de poliadenilación como elemento f). Los sitios de poliadenilación adecuados se describen anteriormente. De acuerdo con una realización, se utiliza un sitio poli A, que comprende la siguiente secuencia:

cagacatgataagatacattgatgagtttgacaaaccacaactagaatgcagtgaaaaaatgcttt
atgtgtgaaattgtgtagctattgctttatgttaaccattataagctgcaataaacaagttaacaa
caacaattgcattcattttatgtttcaggttcagggggagatgtgggaggttttttaaagcaagtaaa
acctctacaaatgtggta (SEQ ID NO 15).

De acuerdo con una realización preferida, el casete de expresión comprende los siguientes elementos:

a) un promotor, preferentemente un promotor de CMV humano o un promotor derivado del promotor de CMV humano;

b) una secuencia de polinucleótidos de 5'UTR seleccionada del grupo que consiste en una secuencia de polinucleótidos de 5'UTR que comprende la SEQ ID NO 1, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 2, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 3, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 4, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 5, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 6, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 7 y una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a la secuencia mostrada en SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO

5, SEQ ID NO 6 o SEQ ID NO 7;

a) un polinucleótido que codifica una secuencia líder secretora de hCD33;

b) un polinucleótido que codifica un polipéptido de interés o un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés;

5 c) una secuencia de polinucleótidos de 3'UTR; y

d) un sitio de poli A.

Como se describió, el % de identidad puede ser calculado sobre la longitud total de la secuencia de referencia. Como se discutió anteriormente y como se muestra en los ejemplos, el casete de expresión de acuerdo con la presente invención da como resultado una expresión aumentada del polipéptido de interés. Un "polipéptido" hace referencia a una molécula que comprende un polímero de aminoácidos unidos juntos mediante enlaces peptídicos. Los polipéptidos incluyen polipéptidos de cualquier longitud, que incluyen proteínas (por ejemplo, que tienen más de 50 aminoácidos) y péptidos (por ejemplo, que tienen 2-49 aminoácidos). Los polipéptidos incluyen proteínas y/o péptidos de cualquier actividad o bioactividad. El polipéptido que se producirá puede ser un compuesto farmacéuticamente o terapéuticamente activo, o una herramienta de búsqueda para utilizar en ensayos y similares. Un polipéptido, por consiguiente, no se limita a ningún polipéptido en particular o grupo de polipéptidos, pero puede, por el contrario, ser cualquier polipéptido, de cualquier tamaño, función u origen, que se desee seleccionar y/o expresar por los métodos descritos en la presente. Por consiguiente, pueden expresarse varios polipéptidos de interés diferentes del casete de expresión de acuerdo con la presente invención y/o las células huésped de acuerdo con la presente invención. Como se estableció anteriormente, el término polipéptido incluye proteínas y/o péptidos de cualquier actividad o bioactividad, que incluye, por ejemplo, polipéptidos bioactivos tales como proteínas o péptidos enzimáticos (por ejemplo, proteasas, cinasas, fosfatasa), proteínas o péptidos receptores, proteínas o péptidos transportadores, proteínas de unión a endotoxinas y/o bactericidas, proteínas o péptidos estructurales, polipéptidos inmunitarios, inmunoglobulinas, toxinas, antibióticos, hormonas, factores de crecimiento, vacunas o similares. Dicho polipéptido puede seleccionarse del grupo que consiste en hormonas peptídicas, interleucinas, activadores del plasminógeno del tejido, citocinas, inmunoglobulinas, en particular anticuerpos o fragmentos funcionales o derivados de estos y anticuerpos de dominio simple, denominados también nanocuerpos. De acuerdo con una realización, el polipéptido de interés es glicosilado. El polipéptido de interés que se expresa del casete de expresión de acuerdo con la presente invención también puede ser una subunidad o dominio de uno de los polipéptidos anteriores, tales como, por ejemplo, una cadena pesada o una cadena ligera de un anticuerpo o un fragmento o derivado funcional de este. En una realización preferida, el polipéptido de interés es una molécula de inmunoglobulina, más preferentemente un anticuerpo, o una subunidad o dominio de este tal como, por ejemplo, la cadena pesada o ligera de un anticuerpo o un anticuerpo de dominio simple. También se incluyen fragmentos o derivados funcionales de los anteriores o una subunidad o dominio de los anteriores tales como, por ejemplo, una cadena pesada o ligera de un anticuerpo o un fragmento o derivado funcional de este. De acuerdo con una realización, el polipéptido de interés no es hCD33.

El término "anticuerpo" como se usa en la presente hace referencia, particularmente, a una proteína que comprende al menos dos cadenas pesadas y dos cadenas ligeras conectadas mediante enlaces disulfuro. El término "anticuerpo" incluye anticuerpos de origen natural así como también todas las formas recombinantes de anticuerpos, por ejemplo, anticuerpos humanizados, anticuerpos totalmente humanos y anticuerpos quiméricos. Cada cadena pesada generalmente comprende una región variable de cadena pesada (VH) y una región constante de cadena pesada (CH). Cada cadena ligera generalmente comprende una región variable de cadena ligera (VL) y una región constante de cadena ligera (CL). La región constante de cadena pesada comprende tres o, en el caso de los anticuerpos del tipo IgM o IgE, cuatro dominios constantes de cadena pesada (CH1, CH2, CH3 y CH4) donde el primer dominio constante CH1 es adyacente a la región variable y puede estar conectado al segundo dominio constante CH2 mediante una región bisagra. La región constante de cadena ligera consiste solo en un dominio constante. Las regiones variables también se pueden subdividir en regiones de hipervariabilidad, denominadas regiones determinantes de complementariedad (CDR), intercaladas con regiones que están más conservadas, denominadas regiones marco (FR), donde cada región variable comprende tres CDR y cuatro FR. Las regiones variables de las cadenas pesada y ligera contienen un dominio de unión que interactúa con un antígeno. Las regiones constantes de los anticuerpos pueden mediar la unión de la inmunoglobulina a los factores o tejidos hospedadores, incluyendo diversas células del sistema inmunitario (por ejemplo, células efectoras) y el primer componente (C1q) del sistema de complemento clásico. El término "anticuerpo" de acuerdo con la invención, sin embargo, también incluye otros tipos y variantes de anticuerpos tales como anticuerpos de cadena pesada, es decir, anticuerpos compuestos solamente de una o más, en particular, dos cadenas pesadas y nanocuerpos, es decir, anticuerpos compuestos solamente de un dominio variable monomérico simple. Dichos nanocuerpos también pueden estar unidos para formar estructuras multivalentes. Preferentemente, el anticuerpo se encuentra, luego de la expresión, en la célula huésped glicosilada adecuada. Como se describe anteriormente, el polinucleótido que

codifica el polipéptido de interés puede también codificar uno o más subunidades o dominios de un anticuerpo, por ejemplo, una cadena pesada o ligera o un fragmento o derivado funcional, como polipéptido de interés.

5 Un anticuerpo de dominio simple, al que también se hace referencia como nanocuerpo es un fragmento de anticuerpo que consiste en un dominio de anticuerpo variable monomérico simple. Al igual que un anticuerpo completo, puede unirse selectivamente a un antígeno específico. Con un peso molecular de solo 12-15 kDa, los anticuerpos de dominio simple son mucho más pequeños que los anticuerpos comunes (150-160 kDa) que se componen de dos cadenas pesadas de proteínas y dos cadenas ligeras, e incluso más pequeños que los fragmentos Fab y fragmentos variables de cadena simple. Los primeros anticuerpos de dominio simple se diseñaron a partir de anticuerpos de cadena pesada encontrados en camélidos; estos se denominan fragmentos V_{HH}. Los anticuerpos de cadena pesada también se encuentran en otras especies. Un enfoque alternativo para separar los dominios variables dimericos de la inmunoglobulina G común (IgG) de humanos o ratones en monómeros. A pesar de que la mayor parte de la investigación de los anticuerpos de dominio simple se basa actualmente en los dominios variables de cadena pesada, se ha demostrado que los nanocuerpos derivados de cadenas ligeras se unen específicamente a epítopos objetivo. Las proteínas de anticuerpos obtenidas a partir de miembros de la familia del camello y el dromedario (*Camelus bactrianus* y *Camelus dromaderius*) que incluye miembros del nuevo mundo tales como las especies de llamas (*Lama paccos*, *Lama glama* y *Lama vicugna*) se han caracterizado con respecto al tamaño, complejidad estructural y antigenicidad para sujetos humanos. Ciertos anticuerpos IgG de esta familia de mamíferos tal como se encuentran en la naturaleza no tienen cadenas ligeras y son, por lo tanto, estructuralmente distintos de la estructura cuaternaria de cuatro cadenas típica que tiene dos cadenas pesadas y dos ligeras, para los anticuerpos de otros animales (véase WO94/04678). Una región del anticuerpo de camélido que es el dominio variable simple pequeño identificado como V_{HH} puede obtenerse mediante modificación genética para proporcionar una proteína pequeña que tiene una alta afinidad por un objetivo, lo que resulta en una proteína derivada de un anticuerpo de peso molecular bajo conocido como un "nanocuerpo de camélido" (véase US5,759,808; Stijlemans, B. et al., 2004 J Biol Chem 279: 1256-1261; Dumoulin, M. et al., 2003 Nature 424: 783-788; Pleschberger, M. et al. 2003 Bioconjugate Chem 14: 440-448; Cortez-Retamozo, V. et al. 2002 Int J Cancer 89: 456-62; y Lauwereys, M. et al. 1998 EMBO J 17: 3512-3520). Las bibliotecas modificadas de anticuerpos de camélidos y fragmentos de anticuerpos se encuentran comercialmente disponibles. Como sucede con otros anticuerpos de origen no humano, una secuencia de aminoácidos de un anticuerpo de camélido puede alterarse de forma recombinante para obtener una secuencia que se asemeja en gran medida a una secuencia humana, es decir, el nanocuerpo puede ser "humanizado". También pueden expresarse anticuerpos de dominio simple respectivos mediante las enseñanzas de la presente invención. Además, como se describió, los anticuerpos de dominio simple también pueden estar unidos para formar estructuras multivalentes. Los nanocuerpos multiméricos respectivos también pueden ser expresados mediante las enseñanzas de la presente invención. De acuerdo con una realización, se expresa un nanocuerpo multimérico de un casete de expresión simple.

Un "fragmento o derivado funcional" de un anticuerpo en particular hace referencia a una proteína o glicoproteína que se deriva de un anticuerpo y es capaz de unirse al mismo antígeno, en particular al mismo epítopo que el anticuerpo. Sucede lo mismo *mutatis mutandi* para un fragmento o derivado de una molécula de inmunoglobulina, una cadena pesada o la cadena ligera. Se ha demostrado que los fragmentos de un anticuerpo de longitud completa o derivados de este pueden llevar a cabo la función de un anticuerpo de unirse a un antígeno. Los ejemplos de fragmentos o derivados de un anticuerpo incluyen (i) fragmentos Fab, fragmentos monovalentes que consisten en la región variable y el primer dominio constante de cada una de las cadenas pesada y ligera; (ii) fragmentos F(ab)₂, fragmentos bivalentes que comprenden dos fragmentos Fab unidos mediante un puente de disulfuro en la región bisagra; (iii) fragmentos Fd que consisten en la región variable y el primer dominio constante CH1 de la cadena pesada; (iv) fragmentos Fv que consisten en la región variable de la cadena pesada y de la cadena ligera de un único brazo de un anticuerpo; (v) fragmentos scFv, fragmentos Fv que consisten en una cadena de polipéptido simple; (vi) fragmentos (Fv)₂ que consisten en dos fragmentos Fv unidos de forma covalente entre sí; (vii) un dominio variable de cadena pesada; y (viii) multicuerpos que consisten en una región variable de cadena pesada y una región variable de cadena ligera unidas de forma covalente entre sí de forma tal que la asociación de las regiones variables de cadena pesada y cadena ligera pueden ocurrir solamente de forma intermolecular, pero no intramolecular. Estos fragmentos y derivados de anticuerpos pueden obtenerse utilizando técnicas convencionales conocidas por aquellos expertos en la técnica.

Como es evidente a partir de los ejemplos de polipéptidos descritos anteriormente que pueden expresarse de acuerdo con las enseñanzas de la presente invención, el polipéptido final que la célula huésped producirá y secretará también puede ser una proteína dimérica o multimérica. Un ejemplo preferido de una proteína respectiva es una molécula de inmunoglobulina, en particular un anticuerpo que comprende, por ejemplo, cadenas pesadas y ligeras. Existen varias opciones para producir una proteína dimérica o multimérica respectiva.

De acuerdo con una realización, dos o más subunidades o dominios de dicha proteína dimérica o multimérica

5 se expresan a partir de un casete de expresión de acuerdo con la presente invención. En la presente
realización, se obtiene una transcripción larga a partir del casete de expresión respectivo que comprende las
regiones codificadoras de las subunidades o dominios individuales de la proteína dimérica o multimérica. De
acuerdo con una realización, al menos un elemento IRES (sitio de entrada de ribosoma interno) se encuentra
funcionalmente ubicado entre las regiones codificadoras de las subunidades o dominios individuales y cada
región codificadora está precedida por una secuencia líder secretora de hCD33 como se describe
anteriormente. De esta manera, se asegura la obtención de productos de traducción separados a partir de
dicha transcripción y que la proteína dimérica o multimérica final pueda ser ensamblada y secretada
correctamente. Además, también pueden expresarse nanocuerpos multiméricos a partir de un casete de
expresión único.

10 Sin embargo, también se encuentra dentro del alcance de la presente invención y para algunas realizaciones
incluso se prefiere expresar las subunidades o dominios individuales de una proteína dimérica o multimérica
de diferentes casetes de expresión. De acuerdo con una realización, el casete de expresión de acuerdo con la
presente invención es un casete de expresión monocistrónico. En esta realización, cada casete de expresión
15 comprende un polinucleótido que codifica una subunidad o dominio de la proteína dimérica o multimérica
como polipéptido de interés. De acuerdo con una realización, todos estos casetes de expresión están
diseñados de acuerdo con las enseñanzas de la presente invención. Sin embargo, también se encuentra
dentro del alcance de la presente invención el utilizar diferentes diseños para diferentes casetes de expresión.
Luego de la expresión de las subunidades/dominios individuales de los casetes de expresión individuales, la
20 proteína dimérica o multimérica final se ensambla y se secreta de la célula huésped. Esta realización se
explicará en mayor detalle junto con el vector de expresión de acuerdo con la presente invención. De acuerdo
con la realización preferida, el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de
interés la cadena pesada o la ligera de una molécula de anticuerpo o un fragmento o derivado funcional de
este.

25 De acuerdo con una realización, el casete de expresión de acuerdo con la presente invención ya comprende
un polinucleótido que codifica al menos una parte de una región constante de una molécula de
inmunoglobulina. El polinucleótido que codifica una parte variable correspondiente de la molécula de
inmunoglobulina puede ser insertado luego por el usuario/cliente en el casete de expresión al utilizar
estrategias de clonación apropiadas para completar el casete de expresión.

30 El casete de expresión puede comprender elementos adicionales que se pueden utilizar para aliviar y/o
mejorar la selección de células huésped de alta expresión. Un método de selección establecido conocido en la
técnica previa para seleccionar células huésped que expresen el polipéptido de interés con un alto
rendimiento se basa en el uso de la citometría de flujo, en particular, clasificación de células activadas por
fluorescencia (FACS, por sus siglas en inglés). Los métodos de selección que emplean la citometría de flujo
35 tienen la ventaja de que se pueden analizar grandes cantidades de células rápidamente. En un método de
selección que es particularmente útil para identificar clones de células de alta producción, una parte del
producto de interés, por ejemplo, un anticuerpo, se expresa como polipéptido de fusión unido a la membrana.
De este modo, una parte del producto se muestra como un polipéptido de fusión en la superficie celular. Como
la cantidad de polipéptido de fusión producido se correlaciona con la tasa de expresión general, las células
40 huésped se pueden seleccionar mediante citometría de flujo en función de la cantidad de polipéptido de fusión
expuesto en la superficie celular. Esto permite la selección rápida de células huésped de alta producción. El
casete de expresión de acuerdo con la presente invención se puede adaptar de forma ventajosa para que se
pueda utilizar en un método de selección respectivo que se basa en el uso de citometría de flujo. Para permitir
la selección eficiente utilizando la citometría de flujo, preferentemente FACS, se pueden utilizar un casete de
45 expresión especial para expresar el polipéptido de interés. Por lo tanto, de acuerdo con una realización, el
casete de expresión para expresar el polinucleótido que codifica el polipéptido de interés está diseñado de
modo que una parte del polipéptido expresado de interés, preferentemente menor que 10 %, más
preferentemente menor que 5 % o incluso menor que 2,5 %, comprende un ancla transmembrana. Existen
varias opciones para alcanzar ese resultado.

50 De acuerdo con una realización, dicho casete de expresión comprende adicionalmente al menos un codón de
terminación posterior al polinucleótido que codifica el polipéptido de interés, y un polinucleótido adicional
posterior al codón de terminación que codifica un ancla membrana y/o una señal para un ancla membrana.
Los elementos respectivos se encuentran unidos de forma operativa. Este diseño del casete de expresión
tiene el efecto de que a través de procesos de translectura traduccional (el codón de terminación tiene
55 "fugas") una parte del polipéptido de interés se produce como un polipéptido de fusión que comprende un
ancla membrana. Como resultado, este polipéptido de fusión está expuesto sobre la superficie celular y las
células que muestran altos niveles de polipéptido de fusión anclado a la membrana se pueden seleccionar
mediante citometría de flujo, preferentemente mediante FACS. De este modo, se seleccionan las células
huésped que tienen una tasa de expresión alta. Se describen detalles y realizaciones preferidas de esta
60 tecnología basada en el codón de terminación en WO2005/073375 y WO2010/022961. Se hace referencia a
esta divulgación.

De acuerdo con una realización alternativa dicho casete de expresión comprende, del lado posterior al polinucleótido que codifica el polipéptido de interés, al menos un intrón que comprende un sitio donante de corte y empalme 5' y un sitio aceptor de corte y empalme 3' y que comprende un codón de terminación traduccional en marco y una señal de poliadenilación y un polinucleótido posterior a dicho intrón que codifica un ancla membrana y/o una señal para un ancla membrana. Los elementos respectivos se encuentran unidos de forma operativa. Este diseño del casete de expresión tiene el efecto que a través de la transcripción y el procesamiento de la transcripción se obtienen al menos dos ARNm maduros diferentes (ARNm-POI) y (ARNm-POI-ANCLA) a partir del casete de expresión. Traducción de los resultados de ARNm-POI en el producto de interés. Traducción de los resultados ARNm-POI-ANCLA en un polipéptido de fusión que comprende el producto de interés y un ancla membrana. Como resultado, este polipéptido de fusión está expuesto una vez más sobre la superficie celular y las células que muestran altos niveles de polipéptido de fusión anclado a la membrana se pueden seleccionar mediante citometría de flujo, preferentemente FACS. De este modo, se seleccionan las células huésped que tienen una tasa de expresión alta. Se describen detalles y realizaciones preferidas de esta tecnología basada en el intrón en WO2007/131774. Se hace referencia a esta divulgación.

De acuerdo con una realización preferida que es en particular útil para la expresión de anticuerpos como producto de interés, el ancla membrana es un ancla transmembrana de inmunoglobulina. Otras anclas membrana adecuadas y realizaciones preferidas de un ancla transmembrana de inmunoglobulina se describen en WO2007/131774, WO2005/073375 y WO2010/022961. Se hace referencia a la divulgación respectiva.

De acuerdo a un segundo aspecto, la presente invención proporciona un vector de expresión para expresar un polipéptido de interés, que comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención. Dicho casete de expresión se describe en detalle anteriormente, por lo tanto se hace referencia a la divulgación anterior.

Un "vector de expresión" de acuerdo con la presente invención se refiere en particular a un polinucleótido capaz de portar al menos un fragmento de ácido nucleico extraño. Un vector funciona como un portador molecular, que suministra fragmentos de ácidos nucleicos, respectivamente polinucleótidos, a una célula huésped. Comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención que comprende las secuencias reguladoras necesarias para expresar de forma apropiada un polinucleótido que codifica un polipéptido de interés incorporado en él.

De acuerdo con una realización, el vector de expresión comprende de manera adicional, al menos un casete de expresión que comprende un polinucleótido que codifica a un marcador seleccionable. Dicho casete de expresión comprende las secuencias reguladoras necesarias para expresar de forma apropiada el polinucleótido que codifica el marcador seleccionable incorporado en él. Los marcadores seleccionables incluyen marcadores seleccionables que proporcionan células eucariotas huésped resistentes a agentes o fármacos tóxicos, tales como antibióticos, en particular antibióticos de aminoglicósidos, por ejemplo, un marcador seleccionable de neomicina. Los marcadores seleccionables también incluyen, de modo no taxativo, marcadores seleccionables eucariotas tal como reductasa de dihidrofolato (DHFR, por sus siglas en inglés) y glutamina sintetasa (GS, por sus siglas en inglés). Otros marcadores de selección adecuados tales como el receptor de ácido fólico se describen en WO 2009/080759 y WO 2010/097240. Preferentemente, el vector de expresión comprende al menos un casete de expresión que comprende un polinucleótido que codifica un marcador seleccionable amplificable. Un marcador seleccionable y amplificable permite que la selección de células huésped eucariotas que contienen un vector así como la amplificación génica. Un ejemplo no taxativo para un gen marcador de mamífero seleccionable y amplificable es el gen reductasa dihidrofolato (DHFR). Actualmente otros sistemas en uso son, entre otros, el sistema de la glutamina sintetasa (gs, por sus siglas en inglés) (Bebbington et al., 1992) y el sistema de selección conducido por histidinol (Hartmann and Mulligan, 1988). Estos marcadores amplificables también son marcadores seleccionables y por lo tanto pueden ser usados para seleccionar aquellas células que obtuvieron el vector. La DHFR y la glutamina sintetasa proporcionan buenos resultados. En ambos casos la selección generalmente ocurre en la ausencia del metabolito apropiado (hipoxantina y timidina en caso de DHFR, glutamina en el caso de GS), previniendo, de este modo, el crecimiento de células no transformadas. Con sistemas amplificables tales como el sistema DHFR, se puede aumentar la expresión de una proteína recombinante al exponer las células a determinados agentes que promueven la amplificación génica tales como antifolatos (por ejemplo, metotrexato (MTX)) en el caso del sistema DHFR. Un inhibidor adecuado para la amplificación del gen promotor GS es la sulfoximina de metionina (MSX). La exposición a MSX también tiene como resultado la amplificación génica. De acuerdo con una realización preferida, el vector de expresión comprende un casete de expresión que comprende un polinucleótido que codifica una enzima de reductasa de dihidrofolato (DHFR) como un marcador seleccionable.

Además, el vector de expresión también puede comprender un casete de expresión que comprende un polinucleótido que codifica un marcador seleccionable procariota. Un "marcador seleccionable procariota" es un marcador seleccionable que permite la selección células huésped procariotas en condiciones de selección

apropiadas. Ejemplos de marcadores seleccionables procariotas respectivos son marcadores que son resistentes a los antibióticos tales como, por ejemplo, la ampicilina, kanamicina, tetraciclina y/o cloranfenicol. Incluir un marcador seleccionable procariota en el vector de expresión tiene la ventaja de que el vector de expresión se puede proliferar rápidamente en las células huésped procariotas.

- 5 Preferentemente, el vector de expresión comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención, un casete de expresión que comprende un polinucleótido que codifica un marcador seleccionable eucariota que proporciona resistencia contra los antibióticos aminoglucósidos, preferentemente un marcador neoseleccionable, y un casete de expresión que comprende un polinucleótido que codifica un marcador seleccionable amplificable, preferentemente DHFR.
- 10 El vector de expresión de acuerdo con la presente invención puede comprender más de un casete de expresión para expresar un polipéptido de interés. Por lo tanto, de acuerdo con una realización, se colocan varios casetes de expresión para expresar polipéptidos iguales o diferentes de interés sobre el vector de expresión de acuerdo con la presente invención. Por lo tanto, la presente invención también proporciona un vector de expresión que comprende más de un casete de expresión donde cada casete de expresión codifica, por ejemplo, una subunidad o dominio de una proteína multimérica de alto orden o dimérica. Los casetes de expresión que codifican diferentes subunidades de una proteína multimérica, cada una incorporada en un casete de expresión diferente, se pueden colocar, por ejemplo, adyacentes entre sí. Para las proteínas multiméricas codificadas por al menos dos genes diferentes (por ejemplo, las cadenas ligeras y pesadas de un anticuerpo o fragmentos funcionales o derivados de estos), los polinucleótidos que codifican las subunidades o dominios deseados se insertan como polipéptidos de interés en los diferentes casetes de expresión. Una realización respectiva que utiliza al menos dos casetes de expresión distintos para expresar subunidades individuales o dominios de una proteína dimérica o multimérica como polipéptido de interés es particularmente ventajosa para expresar moléculas de inmunoglobulina tales como, por ejemplo, anticuerpos. En la célula huésped, la proteína dimérica o multimérica, por ejemplo, el anticuerpo se ensambla y se secreta.
- 15 De acuerdo con una realización, se utiliza el diseño de casete de expresión de acuerdo con la presente divulgación para expresar la cadena pesada de un anticuerpo. El casete de expresión de la cadena ligera puede tener en realizaciones un diseño diferente, por ejemplo, puede comprender una secuencia líder secretora diferente, tal como por ejemplo, una secuencia líder Ig. De acuerdo con una realización, ambos casetes de expresión están diseñados de acuerdo con las enseñanzas de la presente divulgación.
- 20 Por lo tanto, de acuerdo con una realización, el vector de expresión comprende al menos dos casetes de expresión de acuerdo con el primer aspecto de la presente invención. El polinucleótido comprendido en uno de dichos casetes de expresión codifica la cadena pesada de una molécula de inmunoglobulina o un fragmento o derivado funcional de este de interés, y el polinucleótido comprendido en el otro casete de expresión codifica la cadena ligera de una molécula de inmunoglobulina o un fragmento o derivado funcional de este como polipéptido de interés. Luego de la expresión de la cadena ligera y la cadena pesada de dichos casetes de expresión en la célula huésped, la molécula de inmunoglobulina funcional, que preferentemente es un anticuerpo, se ensambla y se secreta de la célula huésped. De acuerdo con una realización, la 5' UTR para la cadena ligera o un fragmento o derivado funcional de este comprende o consiste en la SEQ ID NO 3. De acuerdo con una realización, la 5' UTR para la cadena pesada o un fragmento o derivado funcional de este comprende o consiste en la SEQ ID NO 1 o SEQ ID NO 4. De acuerdo con una realización, la 5' UTR para la cadena ligera o un fragmento o derivado funcional de este comprende o consiste en la SEQ ID NO 5. De acuerdo con una realización, la 5' UTR para la cadena pesada o un fragmento o derivado funcional de este comprende o consiste en la SEQ ID NO 6 o SEQ ID NO 7.
- 25 Los vectores de expresión adecuados que se pueden utilizar junto con la presente invención se describen en WO 2009/080720, donde, sin embargo, en las enseñanzas de la presente invención, la secuencia líder secretora de inmunoglobulina descrita en WO 2009/080720 es reemplazada por la secuencia líder secretora de hCD33 en el o los casetes de expresión que comprende el polinucleótido que codifica el polipéptido de interés. Tal como se describe, en caso de que más de un casete de expresión se encuentre presente en el vector de expresión, es suficiente que un casete de expresión esté diseñado tal como se describe en la presente.
- 30 De acuerdo a un tercer aspecto, se proporciona una célula huésped eucariota que comprende al menos un casete de expresión de acuerdo con el primer aspecto de la presente invención y/o que comprende al menos un vector de expresión de acuerdo con el segundo aspecto de la presente invención. El casete de expresión y el vector de expresión de acuerdo con la presente invención fueron descritos en detalle anteriormente; hace referencia a la divulgación anterior que también aplica en la presente. Preferentemente, el casete de expresión comprende un polinucleótido que codifica un polipéptido de interés. Se puede introducir un casete de expresión respectivo, preferentemente comprendido en un vector de expresión, en una célula huésped mediante transfección.
- 35 De acuerdo con una realización, la célula huésped eucariota comprende al menos dos casetes de expresión de acuerdo con la presente invención. De acuerdo con una realización, cada uno de dichos casetes de

expresión comprende un polinucleótido que codifica al menos una subunidad o dominio de una proteína dimérica o multimérica como polipéptido de interés. Esta realización es particularmente adecuada para expresar moléculas de inmunoglobulina. De acuerdo con una realización preferida, la célula huésped comprende una primer casete de expresión de acuerdo con el primer aspecto de la presente invención, que comprende un polinucleótido que codifica la cadena pesada de una molécula de inmunoglobulina o un fragmento o derivado funcional de este como polipéptido de interés y un segundo casete de expresión de acuerdo con el primer aspecto de la presente invención, que comprende un polinucleótido que codifica la cadena ligera de una molécula de inmunoglobulina o un fragmento o derivado funcional de este como polipéptido de interés. Tal como se describe anteriormente, preferentemente, la 5'UTR que se utiliza en el casete de expresión para expresar la cadena ligera comprende o consiste en la SEQ ID NO 3 y la 5'UTR que se utiliza en el casete de expresión para expresar la cadena pesada comprende o consiste en la SEQ ID NO 1 o SEQ ID NO 4. De acuerdo con una realización, la 5'UTR que se utiliza en el casete de expresión para expresar la cadena ligera comprende o consiste en la SEQ ID NO 5 y la 5'UTR que se utiliza en el casete de expresión para expresar la cadena pesada comprende o consiste en la SEQ ID NO 6 o SEQ ID NO 7. Dichos casetes de expresión se pueden introducir en las células huésped al utilizar uno o más vectores de expresión apropiados tal como se describe anteriormente. De acuerdo con una realización, el primer casete de expresión fue introducido por un vector de expresión y el segundo casete de expresión fue introducido por un segundo vector de expresión. Sin embargo, se prefiere que ambos casetes de expresión sean introducidos utilizando un vector de expresión que lleva ambos casetes de expresión.

De acuerdo con una realización, la célula huésped eucariota comprende al menos un casete de expresión de acuerdo con el primer aspecto para expresar la cadena pesada de un anticuerpo. Tal como se describe, por ejemplo, una 5'UTR que comprende o consiste en la SEQ ID NO 1, 4, 6 o 7 se puede utilizar para este propósito.

Básicamente cualquier célula huésped eucariota se puede utilizar junto con la presente invención siempre y cuando permitan la expresión eficiente de un polipéptido a partir del casete de expresión de acuerdo con la presente invención. Preferentemente, la célula huésped eucariota es una célula de mamífero. Dicha célula de mamífero preferentemente se selecciona del grupo que consiste en células de roedores, células humanas y células de monos. Particularmente se prefiere el uso de las células de roedores, preferentemente seleccionadas del grupo que consiste en células CHO, células BHK, células NS0, células de fibroblastos 3T3 de ratón y células SP2/0. Se prefiere particularmente el uso de célula CHO como células huésped. Las células humanas se pueden seleccionar, por ejemplo, del grupo que consiste en células HEK293, células MCF-7, células PerC6 y células HeLa. Las células de mono se pueden seleccionar, por ejemplo, a partir de células COS y las células Vero. El vector de expresión de acuerdo con la presente invención es particularmente adecuado para producir polipéptidos en células de roedores tales como las células CHO, que incluyen las células CHO DHFR⁻ o las células CHO DHFR⁺. Preferentemente, la célula huésped es una célula CHO.

De acuerdo a un cuarto aspecto, se proporciona un método para producir la célula huésped de acuerdo con el tercer aspecto de la presente invención, donde el vector de expresión de acuerdo con el segundo aspecto de la presente invención se introduce en la célula huésped eucariota, que preferentemente es una célula huésped de mamífero. De este modo, el casete de expresión de acuerdo con un primer aspecto que preferentemente comprende un polinucleótido para expresar un polipéptido de interés se introduce en la célula huésped.

La introducción se puede lograr, por ejemplo, al transfectar el vector de expresión de acuerdo con el segundo aspecto de la presente invención. De acuerdo con una realización, el vector de expresión se integra en el genoma de la célula huésped (transfección estable). Los vectores de expresión adecuados que permiten la introducción del casete de expresión de acuerdo con el primer aspecto de la presente invención en la célula huésped se describen en detalle anteriormente junto con el segundo aspecto de acuerdo con la presente invención. Si el casete de expresión introducido no se inserta en el genoma (transfección temporal) se puede perder en la etapa posterior, por ejemplo, cuando las células son sometidas a mitosis. Los vectores de expresión adecuados también pueden mantenerse en la célula huésped sin integrarse en el genoma, por ejemplo, mediante replicación episomal. Hay varios métodos apropiados conocidos en la técnica previa para introducir un vector de expresión en una célula huésped eucariota, que incluyen células huésped de mamífero, en particular mediante transfección. Los métodos respectivos incluyen, de modo no taxativo, transfección de fosfato de calcio, electroporación, lipofección, transferencia génica mediada por polímeros y biolística. Además de los métodos a base de integración aleatoria tradicional, también se pueden utilizar enfoques mediados por recombinación. Tales métodos de recombinación pueden incluir el uso de recombinasas específicas de sitio tales como Cre, Flp o ΦC31 (véase, por ejemplo, Oumard et al, Cytotechnology (2006) 50: 93 - 108) que pueden mediar la inserción dirigida de transgenes. De manera alternativa, el mecanismo de recombinación homóloga se puede utilizar para insertar el casete de expresión de acuerdo con la presente invención (discutida en Sorrell et al, Biotechnology Advances 23 (2005) 431 - 469). La inserción de genes basada en la recombinación permite minimizar la cantidad de elementos a ser incluidos en el vector de expresión que se introduce a la célula huésped. Las realizaciones de un vector de expresión adecuado o combinaciones de vectores de expresión de acuerdo con la presente invención así

- como células huésped adecuadas se describen en detalle anteriormente; se refiere en la divulgación anterior. Tal como se describe anteriormente en conjunción con la realización, los casetes de expresión que comprenden los polinucleótidos que codifican la cadena pesada y la cadena ligera de una molécula de inmunoglobulina o un fragmento o derivado funcional de este se puede ubicar en vectores de expresión iguales o diferentes en caso de que una combinación de al menos dos vectores de expresión se utilice para transfectar las células huésped.
- De acuerdo con un quinto aspecto, se proporciona un método para producir un polipéptido de interés, dicho método comprende cultivar células huésped de acuerdo con el tercer aspecto de la presente invención en un cultivo celular bajo condiciones que permiten la expresión de dicho polipéptido de interés.
- Existen dos formatos principales de cultivos de células huésped, a saber, cultivos de células adherentes y cultivos de suspensión. Se prefiere el uso de cultivos de suspensión. De acuerdo con una realización, dichas células huésped se cultivan bajo condiciones libres de suero.
- El polipéptido de interés se expresa a partir del casete de expresión de acuerdo con la presente invención y se secreta de la célula huésped, por ejemplo, en el medio de cultivo, y el polipéptido secretado se puede obtener a partir de este. Si más de un casete de expresión se encuentra presente en la célula huésped, por ejemplo, cuando la proteína dimérica o multimérica se expresa (véase anteriormente), la proteína dimérica o multimérica se ensambla en la célula y luego se secreta de la célula huésped. Debido a la expresión extraordinaria que se logra al utilizar la combinación de la secuencia líder 5'UTR/secretora novedosa de acuerdo con la presente invención, los polipéptidos se pueden expresar y secretar con alto rendimiento. El polipéptido secretado también puede estar sujeto a etapas de procesamiento adicionales tales como, por ejemplo, etapas de purificación y/o modificación. Por consiguiente, el método para producir el polipéptido de interés puede comprender al menos una de las siguientes etapas:
- aislar el polipéptido de interés de dicho medio de cultivo celular; y/o
 - procesar el polipéptido de interés aislado.
- Por lo tanto, el polipéptido producido de acuerdo con la invención se puede recuperar y opcionalmente se puede procesar adicionalmente, por ejemplo, purificar adicionalmente, aislar y/o modificar mediante métodos conocidos en la técnica. Por ejemplo, el polipéptido se puede recuperar a partir del medio nutriente mediante procedimientos convencionales que incluyen, de modo no taxativo, centrifugación, filtración, ultrafiltración, extracción o precipitación. La purificación se puede realizar mediante varios procedimientos conocidos en la técnica que incluyen, de modo no taxativo, cromatografía (por ejemplo, intercambio iónico, afinidad, hidrofóbico, cromatografía de exclusión por tamaño), procedimientos electroforéticos (por ejemplo, enfoque isoeléctrico preparativo), solubilidad diferencial (por ejemplo, precipitación de sulfato de amonio) o extracción.
- Tal como se describe anteriormente, el polipéptido de interés es preferentemente una molécula de inmunoglobulina o un fragmento o derivado funcional de este, más preferentemente un anticuerpo o un fragmento o derivado funcional de este.
- De acuerdo con un sexto aspecto, la presente invención se relaciona con el uso de una secuencia de 5'UTR en combinación con una secuencia líder secretora de hCD33 en un casete de expresión para expresar un polipéptido de interés con alto rendimiento a partir de dicho casete de expresión, donde dicha secuencia de polinucleótidos de 5'UTR se selecciona del grupo que consiste en una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 1, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 2, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 3, una secuencia de polinucleótidos de 5'UTR que comprende SEQ ID NO 4 y una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica, a la secuencia mostrada en SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3 o SEQ ID NO 4 o donde dicha secuencia de polinucleótidos de 5'UTR se selecciona del grupo que consiste en una secuencia de polinucleótidos de 5'UTR que consiste en SEQ ID NO 5, una secuencia de polinucleótidos de 5'UTR que consiste en SEQ ID NO 6, una secuencia de polinucleótidos de 5'UTR que consiste en SEQ ID NO 7 y una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es la menos 85 %, preferentemente al menos 90 %, idéntica a la secuencia que se muestra en SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7. La identidad puede calcularse sobre la longitud total de la secuencia de referencia.
- Las ventajas de una combinación respectiva, realizaciones adecuadas y preferidas de la secuencia de polinucleótidos de 5'UTR y la secuencia líder secretor de hCD33 así como también realizaciones adecuadas y preferidas del casete de expresión y el polinucleótido que codifica el polipéptido de interés se describen en detalle anteriormente. Se hace referencia a la divulgación anterior que también se aplica a la presente.
- Preferentemente, el casete de expresión tiene el diseño del casete de expresión que se describe anteriormente en conjunción con el primer aspecto de la presente invención. Se hace referencia a la divulgación anterior. De acuerdo con una realización, el casete de expresión tiene una o más de las siguientes

características:

a) comprende un promotor;

b) comprende una secuencia de polinucleótidos de 5'UTR como se describe anteriormente;

5 c) comprende una secuencia líder secretora de hCD33 que comprende y preferentemente consiste en la secuencia MPLLLLLPLLWAGALA (SEQ ID NO 12);

d) comprende un polinucleótido que codifica un polipéptido de interés o un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés;

e) comprende una secuencia de 3'UTR y/o

f) comprende un sitio de poli A

10 Se describen anteriormente los elementos individuales y realizaciones y combinaciones preferidas en combinación con el primer aspecto de la presente invención. Se hace referencia a la divulgación anterior.

15 De acuerdo con una realización, el polinucleótido que codifica el polipéptido de interés codifica, como polipéptido de interés, dos o más subunidades o dominios de una proteína dimérica o multimérica, donde al menos un elemento IRES se ubica entre las regiones codificadoras de las subunidades o dominios individuales y cada región codificadora está precedida por una secuencia líder secretora de hCD33. Sin embargo, se prefiere el uso de los casetes de expresión monocistrónicos en el contexto de la presente invención.

20 De acuerdo con una realización alternativa, el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés una subunidad o dominio de una proteína dimérica o multimérica. Preferentemente, el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena pesada o la ligera de una molécula de anticuerpo o un fragmento o derivado funcional de este. De acuerdo con una realización, se utiliza un diseño de casete de expresión de acuerdo con la presente invención para expresar la cadena pesada de un anticuerpo. Aquí, el casete de expresión para expresar la cadena ligera puede tener un diseño diferente y puede, por ejemplo, comprender una secuencia líder diferente, por ejemplo, una secuencia líder de Ig, o puede también estar diseñada de acuerdo con las enseñanzas de la presente invención. De acuerdo con una realización, se utiliza un diseño de casete de expresión de acuerdo con la presente invención para expresar la cadena ligera de un anticuerpo. Aquí, el casete de expresión para expresar la cadena pesada puede tener un diseño diferente y puede, por ejemplo, comprender una secuencia líder diferente, por ejemplo, una secuencia líder de Ig, o puede también estar diseñada de acuerdo con las enseñanzas de la presente invención. De acuerdo con una realización, la materia descrita en la presente como que comprende ciertos elementos también hace referencia a la materia que consiste en los elementos respectivos. En particular, las 5'UTR descritas en la presente como que comprenden ciertas secuencias también pueden consistir en las secuencias respectivas. Se prefiere seleccionar y combinar realizaciones preferidas descritas en la presente y la materia específica que surge de una combinación respectiva de realizaciones preferidas también pertenece a la presente divulgación.

35 Los siguientes ejemplos sirven para ilustrar la presente invención sin limitar de ninguna forma el alcance de esta. En particular, los ejemplos se relacionan con las realizaciones preferidas de la presente invención.

EJEMPLOS

Los ejemplos se llevaron a cabo de acuerdo con el siguiente protocolo:

40 EJEMPLO A

I. Material y métodos

Células huésped

Se utilizaron como células hospedadores, células CHO (ovario de hámster chino) derivadas de células CHO-K1.

45 Vectores de expresión

50 Como control estándar, se utiliza un vector de expresión que tiene un diseño como el que se describe en WO 2009/080720 (véase, en particular, el Ejemplo 6), que utiliza una secuencia líder secretora de Ig en los casetes de expresión para expresar las cadenas ligeras y pesadas de un anticuerpo. El promotor de CMV comprende la SEQ ID NO 8. Se utiliza la SEQ ID NO 3 como 5'UTR para expresar la cadena ligera y se utilizó la SEQ ID NO 4 como 5'UTR para expresar la cadena pesada. Se obtiene un vector de expresión de acuerdo con las descripciones de la invención a partir de dicho vector de control mediante la sustitución de la

secuencia líder secretora de inmunoglobulina de cadena ligera y cadena pesada por la siguiente secuencia líder secretora de hCD33:

MPLLLLLLPLLWAGALA (SEQ ID NO 12)

5 Dichos vectores de expresión se utilizan con el fin de expresar moléculas de anticuerpo. Un diseño de vector correspondiente también se utiliza con el fin de expresar un nanocuerpo como un polipéptido de interés, donde, sin embargo, se utiliza la 5'UTR de acuerdo con la SEQ ID NO 5. Aquí, sin embargo, se necesita solamente un casete de expresión. El vector de control tiene el mismo diseño, sin embargo, se utilizó nuevamente una secuencia líder de Ig.

Cultivo, transfección y selección celular

10 Las células se cultivaron utilizando medio de cultivo celular patentado producido en el laboratorio y se pasaron regularmente 2-3 veces a la semana para mantener la fase de crecimiento logarítmica a lo largo del estudio. Las células se transfectaron mediante métodos de nucleofección estándar. Se centrifugaron células 5E6/impulso de nucleofección y se resuspendieron en un amortiguador de transfección. Se agregaron 3 ug de ADN vector que codifica el polipéptido de interés y se lleva a cabo la nucleofección. Se transfirieron las células 15 transfectadas en un matraz de agitación de 125 ml y se cultivaron en condiciones de agitación durante 24-48 horas. Se llevó a cabo una primera etapa de selección con G-418 como se describe previamente en WO 2009/080720. La selección adicional se lleva a cabo con 2 etapas de selección de metotrexato (MTX) adicionales, principalmente MTX 500 nM y MTX 1000 nM. Los grupos seleccionados, que comprenden células 20 productoras resistentes, aceptables en su mayoría, se clonan a una densidad de 0,5-1 células/pocillos, ya sea mediante dilución limitante o mediante el sistema FACS. Se analizó la productividad y el crecimiento clonal en diferentes formatos de análisis para los clones obtenidos.

II. Resultados

25 La información experimental obtenida con el vector de expresión de acuerdo con la técnica previa (vector de expresión de control) que comprende la secuencia líder secretora de Ig y el vector de expresión de acuerdo con la presente invención que comprende la secuencia líder secretora hCD33 demuestra que la titulación del anticuerpo producido es aumentada notablemente cuando se utiliza la combinación nueva de la 5'UTR específica y la secuencia líder secretora de hCD33 de acuerdo con la presente invención. Como proteína de interés, se expresa un anticuerpo de IgG. El aumento de la productividad observado en el nivel de grupo fue en promedio de 2,88 veces cuando se utilizó el vector de expresión de acuerdo con la presente invención. Los resultados pueden mejorarse adicionalmente en el nivel de clones, donde podría obtenerse un aumento de la productividad de 4,1 veces cuando se utiliza el vector de expresión de acuerdo con la presente invención (cuando se compara el mejor clon de control obtenido con el vector de expresión de control y el mejor clon 30 obtenido con el vector de expresión de acuerdo con la presente invención).

35 Cuando se expresa un nanocuerpo como proteína de interés, la comparación de los mejores clones muestra un aumento en la productividad de 1,1 veces cuando se utiliza el vector de expresión de acuerdo con la presente invención y cuando se comparan los 6 mejores clones se obtiene un aumento de la productividad de 1,6 veces en promedio. Esto se ilustra en las siguientes tablas:

Tabla 2: Evaluación de la combinación nueva con el anticuerpo IgG: Se logró un aumento en la productividad de 2,88 veces en el nivel de grupo con el vector de expresión de acuerdo con la presente invención.

Titulación (g/L)	G-418	1000 nM MTX
péptido señal de hCD33	0,006	0,094
Control	0,005	0,038

40 Tabla 3: Evaluación de la combinación nueva con el anticuerpo IgG: 4,1 veces de aumento de productividad al comparar los mejores clones del vector de expresión conocido en la técnica previa y el vector de expresión de acuerdo con la presente invención y en promedio de los 6 mejores clones: la productividad aumentó 4 veces.

	6 mejores clones de hCD33	6 mejores clones de control
Titulación (g/L)	2,99	0,727
	2,69	0,656
	2,65	0,642

	2,21	0,622
	2,12	0,575
	2,11	0,503

Tabla 4: Evaluación de la combinación nueva con nanocuerpos: 1,1 veces de aumento de productividad al comparar los mejores clones del vector de expresión conocido en la técnica previa y el vector de expresión de acuerdo con la presente invención y en promedio de los 6 mejores clones: la productividad aumentó 1,6 veces.

5

	6 mejores clones de hCD33 SP	6 mejores clones de control
Titulación (g/L)	0,92	0,75
	0,92	0,58
	0,90	0,57
	0,88	0,57
	0,88	0,57
	0,87	0,57

EJEMPLO B

Vectores de expresión

10 Como control estándar, se utilizaron los vectores de expresión que tienen el diseño básico como el que se describe en WO 2009/080720 (véase, en particular, el Ejemplo 6), que utilizan secuencias líder secretoras de Ig en los casetes de expresión para expresar las cadenas ligeras y pesadas de un anticuerpo. El casete de expresión para la cadena pesada se modificó utilizando la tecnología de codón de detención de fugas descrita en WO2010/022961 para facilitar la selección de FACS. El promotor de CMV comprende la SEQ ID NO 8. Se utiliza la SEQ ID NO 5 como 5'UTR para expresar la cadena ligera y se utiliza la SEQ ID NO 7 como 5'UTR para expresar la cadena pesada. Se obtiene un vector de expresión de acuerdo con las descripciones de la invención a partir de dichos vectores de control mediante la sustitución de la secuencia líder secretora de inmunoglobulina de cadena pesada existente por la siguiente secuencia líder secretora de hCD33:

15

MPLLLLLPLLWAGALA (SEQ ID NO 12)

Dichos vectores de expresión se utilizan con el fin de expresar moléculas de anticuerpo diferentes.

20

La evaluación se llevó a cabo para abordar el procesamiento correcto de la expresión de IgG y del péptido de señal (secuencia líder) en el casete de cadena pesada para 3 anticuerpos diferentes. El procesamiento correcto de la secuencia líder, es decir, la escisión de la secuencia líder mediante la peptidasa de señal, es esencial para obtener la secuencia esperada del polipéptido de interés en su extremo N, y por lo tanto una molécula funcional con la calidad esperada. Se observó previamente que en algunos casos utilizando los diseños de la técnica previa el procesamiento del péptido señal fue incorrecto, y generó, por ejemplo, la cadena pesada o ligera con uno o más aminoácidos adicionales en el extremo N. Dependiendo del uno o más aminoácidos que permanecieron incorrectamente en el extremo N, esta extensión de la secuencia de aminoácidos por el uno o más aminoácidos restantes de la secuencia líder puede, por ejemplo, aumentar la tendencia de agregación de la molécula o inducir la formación de enlaces de disulfuro. Por lo tanto, puede tener un impacto grave en la calidad del producto. Además, se descubrió que el procesamiento incorrecto del péptido señal supone el riesgo de que el polipéptido de interés, tal como un producto de anticuerpo, se escinda mediante proteasa posterior al sitio de escisión predicho en el extremo N (también denominado pinzamiento). Esto genera un producto truncado, que también puede tener un impacto grave en la calidad de la molécula.

25

30

35

La tabla 5 ilustra los resultados del procesamiento del péptido señal (secuencia líder) obtenidos con la tecnología de la presente divulgación comparados con los controles. Como puede observarse, utilizando la tecnología de la invención redujo significativamente el procesamiento del péptido señal no deseado y, por lo tanto, mejoró la calidad.

Tabla 5: Evaluación de diferentes péptidos señal en los casetes de cadena pesada muestra un procesamiento de línea celular correcto y mejorado cuando se expresan tres modelos de anticuerpo diferentes.

Polipéptido de interés	Procesamiento/extensions del péptido señal y pinzamiento con hCD33	Procesamiento/extensions del péptido señal y pinzamiento con otros péptidos señal
Anticuerpo 1	<1 %	~3,6%
Anticuerpo 2	<1 %	~3,2%
Anticuerpo 3	Ninguno	~1,5% ~1,2%

Listado de secuencias

- 5 <110> Novartis AG
 <120> Casete de expresión optimizado para expresar un polipéptido con alto rendimiento
 <130> PAT055302-WO-PCT
 <150> US 61/728,459
 <151> 2012-11-20
- 10 <160> 15
 <170> PatentIn version 3.3
 <210> 1
 <211> 288
 <212> ADN
- 15 <213> artificial
 <220>
 <223> 5' UTR
 <400> 1
- ```

 agatcgctg gagacgcat ccacgtggt ttgacctca tagaagacac cgggaccgat 60
 ccagcctccg cggccgggaa cgggtgattg gaacgcggat tccccgtgcc aagagtgacg 120
 taagtaccgc ctatagagtc tataggccca ccccttggc ttcgtagaa cgcggtaca 180
 attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240
 ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgca 288

```
- 20 <210> 2  
 <211> 302  
 <212> ADN  
 <213> artificial  
 <220>
- 25 <223> 5' UTR



ES 2 620 325 T3

<400> 2

```

agatcgctg gagacgcat ccacgtgtt ttgacctca tagaagacac cgggaccgat 60
ccagcctccg cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120
taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180
attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240
ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300
cg 302

```

<210> 3

<211> 316

5 <212> ADN

<213> artificial

<220>

<223> 5' UTR

<400> 3

```

agatcgctg gagacgcat ccacgtgtt ttgacctca tagaagacac cgggaccgat 60
ccagcctccg cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120
taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180
attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240
ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300
cgaaaacgcg tccacc 316

```

10

<210> 4

<211> 352

<212> ADN

<213> artificial

15 <220>

<223> 5' UTR

<400> 4

ES 2 620 325 T3

agatcgctg gagacgcat ccacgctgtt ttgacctcca tagaagacac cgggaccgat 60  
 ccagcctcog cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120  
 taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180  
 attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240  
 ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300  
 cgcgattgaa ttccccgggg atcctctagg gtgaccgttt ggtgccgcca cc 352

<210> 5

<211> 327

<212> ADN

5 <213> Artificial

<220>

<223> 5'UTR

<400> 5

agatcgctg gagacgcat ccacgctgtt ttgacctcca tagaagacac cgggaccgat 60  
 ccagcctcog cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120  
 taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180  
 attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240  
 ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300  
 cgaaaacgcg cctctagagc cgccacc 327

10 <210> 6

<211> 321

<212> ADN

<213> Artificial

<220>

15 <223> 5'UTR

<400> 6

ES 2 620 325 T3

```

agatcgctg gagacgccat ccacgctggt ttgacctcca tagaagacac cgggaccgat 60
ccagcctccg cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120
taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180
attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240
ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300
cgaaaaacgc gtgccgccac c 321

```

<210> 7

<211> 320

<212> ADN

5 <213> Artificial

<220>

<223> 5'UTR

<400> 7

```

agatcgctg gagacgccat ccacgctggt ttgacctcca tagaagacac cgggaccgat 60
ccagcctccg cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgacg 120
taagtaccgc ctatagagtc tataggccca cccccttggc ttcgttagaa cgcggctaca 180
attaatacat aaccttatgt atcatacaca tacgatttag gtgacactat agaataacat 240
ccactttgcc tttctctcca caggtgtcca ctcccaggtc caactgcacc tcggttctat 300
cgaaaaacgcg tgccgccacc 320

```

10 <210> 8

<211> 737

<212> ADN

<213> hCMV

<400> 8

ES 2 620 325 T3

tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60  
 ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120  
 aatatgaccg ccatggtggc attgattatt gactagttat taatagtaat caattacggg 180  
 gtcattagtt catagcccat atatggagtt ccgcggtaca taacttacgg taaatggccc 240  
 gcctggctga ccgcccaacg acccccgcc attgacgtca ataatgacgt atgttcccat 300  
 agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360  
 ccacttggca gtacatcaag tgtatcatat gccaaagtacg ccccctattg acgtcaatga 420  
 cggtaaatgg cccgcctggc attatgcccga gtacatgacc ttatgggact ttcctacttg 480

gcagtacatc tacgtattag tcatcgctat taccatgggtg atgcgggtttt ggcagtacat 540  
 caatgggctg ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600  
 caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc 660  
 cgccccattg acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc 720  
 tcgttttagtg aaccgtc 737

<210> 9

<211> 119

5 <212> ADN

<213> ratón

<400> 9

agatcgcttg gagacgcat ccacgctgtt ttgacctcca tagaagacac cgggaccgat 60  
 ccagcctccg cggccgggaa cgggtgcattg gaacgcggat tccccgtgcc aagagtgac 119

<210> 10

10 <211> 144

<212> ADN

<213> ratón

<400> 10

gtaagtaccg cctatagagt ctataggccc acccccttgg cttcggtaga acgcggctac 60  
 aattaatata taaccttatg tatcatcac atacgattta ggtgacacta tagaataaca 120  
 tccactttgc ctttctctcc acag 144

15 <210> 11

<211> 25

<212> ADN

<213> ratón

<400> 11

gtgtccactc ccagggtccaa ctgca 25

5 <210> 12

<211> 16

<212> PRT

<213> human

<400> 12

10 **Met Pro Leu Leu Leu Leu Pro Leu Leu Trp Ala Gly Ala Leu Ala**  
**1 5 10 15**

<210> 13

<211> 18

<212> PRT

<213> humano

15 <400> 13

**Met Pro Leu Leu Leu Leu Pro Leu Leu Trp Ala Gly Ala Leu Ala**  
**1 5 10 15**

**Met Asp**

<210> 14

<211> 39

<212> ADN

20 <213> artificial

<220>

<223> 3' UTR

<400> 14

gggcggccgc ttccttag tgagggttaa tgcttcgag 39

25 <210> 15

<211> 222

<212> ADN

<213> artificial

<220>

30 <223> sitio poli A

<400> 15

ES 2 620 325 T3

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa | 60  |
| aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca | 120 |
| ataaacaagt taacaacaac aattgcattc attttatggt tcaggttcag ggggagatgt | 180 |
| gggaggtttt ttaaagcaag taaaacctct acaaatgtgg ta                    | 222 |

**REIVINDICACIONES**

1. Un casete de expresión adecuado para expresar un polipéptido de interés, el cual comprende:

a) un promotor;

5 b) una secuencia de polinucleótidos de 5'UTR, en donde dicha secuencia de polinucleótidos de 5'UTR se selecciona a partir del grupo que consiste en:

i) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga  
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaaagagtga  
cgtaagtaccgcctatagagtctataggcccaccccccttggttcggttagaacgcggct  
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata  
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgca  
(SEQ ID NO 1);

ii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaaagagtgcg  
taagtaccgcctatagagtctataggcccaccccccttggttcggttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cg (SEQ ID NO 2);

10 iii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaaagagtgcg  
taagtaccgcctatagagtctataggcccaccccccttggttcggttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cgaaaacgcgtccacc (SEQ ID NO 3);

iv) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaaagagtgcg  
taagtaccgcctatagagtctataggcccaccccccttggttcggttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cgcgattgaattccccggggatcctctagggtgaccgtttggtgccgccacc  
(SEQ ID NO 4);

v) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cgaaaacgcgcctctagagccgccacc (SEQ ID NO 5);

vi) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cgaaaacgcgctgccgccacc (SEQ ID NO 6);

5 vii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccgat  
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccaccccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcacatacagatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggttctat  
cgaaaacgcgctgccgccacc (SEQ ID NO 7);

viii) una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a una secuencia mostrada en la SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7;

10 c) un polinucleótido que codifica una secuencia líder secretora de hCD33; y

d) un polinucleótido que codifica un polipéptido de interés o un sitio de inserción para insertar un polinucleótido que codifica un polipéptido de interés.

2. El casete de expresión de acuerdo con la reivindicación 1, en donde la secuencia líder secretora de hCD33 consiste en la secuencia MPLLLLLPLLWAGALA (SEQ ID NO 12).

15 3. El casete de expresión de acuerdo con la reivindicación 1 o 2, en donde dicho casete de expresión comprende además:

e) una secuencia de 3'UTR y/o

f) una señal de poli A.

20 4. El casete de expresión de acuerdo con una o más de las reivindicaciones 1 a 3, el cual tiene una o más de las siguientes características:

i) el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés dos o más subunidades o dominios de una proteína dimérica o multimérica, en donde al menos un elemento IRES se ubica entre las regiones codificadoras de las subunidades o dominios individuales y cada región codificadora está precedida por una secuencia líder secretora de hCD33;

25 ii) el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés una subunidad o



dominio de una proteína dimérica o multimérica;

iii) el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena pesada o la ligera de una molécula de anticuerpo o de un fragmento o derivado funcional del mismo;

5 iv) la secuencia de polinucleótidos de 5'UTR comprende o consiste en la SEQ ID NO 3 y el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena ligera de una molécula de anticuerpo o de un fragmento o derivado funcional del mismo;

v) la secuencia de polinucleótidos de 5'UTR comprende o consiste en la SEQ ID NO 1 o SEQ ID NO 4 y el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena pesada de una molécula de anticuerpo o de un fragmento o derivado funcional del mismo;

10 vi) la secuencia de polinucleótidos de 5'UTR comprende o consiste en la SEQ ID NO 5 y el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena ligera de una molécula de anticuerpo o de un fragmento o derivado funcional del mismo;

15 vii) la secuencia de polinucleótidos de 5'UTR comprende o consiste en la SEQ ID NO 6 o SEQ ID NO 7 y el polinucleótido que codifica el polipéptido de interés codifica como polipéptido de interés la cadena pesada de una molécula de anticuerpo o de un fragmento o derivado funcional del mismo;

viii) el casete de expresión es un casete de expresión monocistrónico;

ix) el promotor comprendido en el casete de expresión es un promotor de CMV humano que se selecciona a partir del grupo que consiste en:

aa) un promotor que comprende la siguiente secuencia:

```
tcaatattgg ccattagcca tattattcat tggttatata gcataaatca
atattggcta ttggccattg catacgttgt atctatatca taatatgtac
atztatattg gctcatgtcc aatatgaccg ccatgttggc attgattatt
gactagttat taatagtaat caattacggg gtcattagtt catagcccat
atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggetga
ccgcccacg acccccgccc attgacgtca ataatgacgt atgttcccat
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac
ggtaaactgc ccacttgcca gtacatcaag tgtatcatat gccaaagtacg
ccccctattg acgtcaatga cggtaaattg cccgcctggc attatgceca
gtacatgacc ttatgggact ttctacttg gcagtacatc tacgtattag
tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggctg
ggatagcgg ttgactcacg gggatttcca agtctccacc ccattgacgt
caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc
gtaacaactc cgccccattg acgcaaattg gcggtaggcg tgtacgggtg
gaggtctata taagcagagc tcgtttagtg aaccgtc (SEQ ID NO 8);
```

20

o un fragmento funcional de la misma que funciona como promotor;

bb) un promotor que comprende una secuencia que es al menos 80 % idéntica a la secuencia mostrada en la SEQ ID NO 8 o un fragmento funcional de la misma que funciona como promotor.

25 5. Un vector de expresión para expresar un polipéptido de interés, el cual comprende al menos un casete de expresión de acuerdo con una o más de las reivindicaciones 1 a 4.

6. El vector de expresión de acuerdo con la reivindicación 5, el cual comprende además uno más de los siguientes elementos:

a) al menos un casete de expresión que comprende un polinucleótido que codifica un marcador seleccionable;

b) al menos un segundo casete de expresión para expresar un polipéptido de interés.

7. El vector de expresión de acuerdo con la reivindicación 6, el cual comprende al menos dos casetes de expresión, en donde cada casete de expresión comprende un polinucleótido que codifica un polipéptido de interés, en donde el polinucleótido comprendido en un casete de expresión codifica la cadena pesada de una molécula de inmunoglobulina o un fragmento o derivado funcional de la misma, y el polinucleótido comprendido en el otro casete de expresión codifica la cadena ligera de una molécula de inmunoglobulina o un fragmento o derivado funcional de la misma.
8. El vector de expresión de acuerdo con la reivindicación 7, en donde la secuencia de polinucleótidos de 5'UTR del casete de expresión que codifica la cadena pesada de una molécula de inmunoglobulina o un fragmento o derivado funcional de la misma comprende o consiste en la SEQ ID NO 1 o SEQ ID NO 4, y en donde la secuencia de polinucleótidos de 5'UTR del casete de expresión que codifica la cadena ligera de una molécula de inmunoglobulina o un fragmento o derivado funcional de la misma comprende o consiste en la SEQ ID NO 3.
9. Una célula huésped eucariota que comprende al menos un casete de expresión de acuerdo con una o más de las reivindicaciones 1 a 4 y/o al menos un vector de expresión de acuerdo con una o más de las reivindicaciones 5 a 8.
10. La célula huésped de acuerdo con la reivindicación 9, en donde dicha célula huésped se selecciona a partir del grupo que consiste en células de roedores, células de primates y células humanas, y en donde dicha célula huésped es preferentemente una célula CHO.
11. Un método para producir la célula huésped de acuerdo con la reivindicación 9 o 10, en donde se introduce un vector de expresión de acuerdo con una o más de las reivindicaciones 5 a 8 en una célula huésped eucariota.
12. Un método para producir un polipéptido de interés, comprendiendo dicho método cultivar células huésped de acuerdo con la reivindicación 9 o 10 en un cultivo celular bajo condiciones que permitan la expresión de dicho polipéptido de interés.
13. El método de acuerdo con la reivindicación 12, en donde dicho polipéptido de interés se secreta dentro del medio de cultivo celular y es aislado del medio de cultivo celular, y el polipéptido aislado se procesa adicionalmente de manera opcional.
14. El método de acuerdo con la reivindicación 12 o 13, en donde dicho polipéptido es una molécula de inmunoglobulina o un fragmento de la misma.
15. Uso de una secuencia de 5'UTR en combinación con una secuencia líder secretora de hCD33 en un casete de expresión para expresar un polipéptido de interés con un alto rendimiento a partir de dicho casete de expresión, en donde dicha secuencia de polinucleótidos de 5'UTR se selecciona a partir del grupo que consiste en:

i) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga
tccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtga
cgtaagtaccgcctatagagtctataggcccccccccttggttcgtagaacgcggct
acaattaatacataaccttatgtatcatacacatacagatttaggtgacactatagaata
acatccactttgctttctctccacaggtgtccactcccaggtccaactgca
(SEQ ID NO 1);
```

ii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

```
agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat
ccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg
taagtaccgcctatagagtctataggcccccccccttggttcgtagaacgcggctaca
attaatacataaccttatgtatcatacacatacagatttaggtgacactatagaataacat
ccactttgctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat
cg (SEQ ID NO 2);
```

iii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcgccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cgaaaacgcgtccacc (SEQ ID NO 3);

iv) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcgccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cgcgattgaattccccggggatcctctaggggtgaccgtttggtgcccacc  
(SEQ ID NO 4);

5 v) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgat  
ccagcctccgcgccgggaacggtgcattggaacgcggattccccgtgccaagagtgacg  
taagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggctaca  
attaatacataaccttatgtatcatacacatacgatttaggtgacactatagaataacat  
ccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggttctat  
cgaaaacgcgcctctagagccgccacc (SEQ ID NO 5);

vi) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

agatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccga  
tccagcctccgcgccgggaacggtgcattggaacgcggattccccgtgccaagagtga  
cgtaagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggct  
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata  
acatccactttgcctttctctccacaggtgtccactcccaggtccaactgcacctcggt  
tctatcgaaaacgcgtgcccacc (SEQ ID NO 6);

vii) una secuencia de polinucleótidos de 5'UTR que comprende la siguiente secuencia:

## ES 2 620 325 T3

agatcgctggagacgccatccacgctgTTTTgacctccatagaagacaccgggaccga  
tccagcctccgcgggccggaacggtgcattggaacgcggattccccgtgccaaagagtga  
cgtaagtaccgcctatagagtctataggcccacccccttggttcgttagaacgcggct  
acaattaatacataaccttatgtatcatacacatacgatttaggtgacactatagaata  
acatccactttgcctttctctccacaggtgtccactcccagggtccaactgcacctcggt  
tctatcgaaaacgcgtgccgccacc (SEQ ID NO 7);

viii) una secuencia de polinucleótidos de 5'UTR que comprende una secuencia que es al menos 85 %, preferentemente al menos 90 % idéntica a una secuencia mostrada en la SEQ ID NO 1, SEQ ID NO 2, SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO 5, SEQ ID NO 6 o SEQ ID NO 7.