

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 599 386**

51 Int. Cl.:

**C12N 15/67** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.04.2013 PCT/EP2013/057808**

87 Fecha y número de publicación internacional: **24.10.2013 WO13156443**

96 Fecha de presentación y número de la solicitud europea: **15.04.2013 E 13718536 (9)**

97 Fecha y número de publicación de la concesión europea: **14.09.2016 EP 2839011**

54 Título: **Método para la expresión de polipéptidos usando ácidos nucleicos modificados**

30 Prioridad:

**17.04.2012 EP 12164430**

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

**01.02.2017**

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)  
Grenzacherstrasse 124  
4070 Basel, CH**

72 Inventor/es:

**KLOSTERMANN, STEFAN;  
KOPETZKI, ERHARD y  
SCHWARZ, URSULA**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 599 386 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Método para la expresión de polipéptidos usando ácidos nucleicos modificados

5 Los métodos que se describen en el presente documento se incluyen en el campo de la optimización de un ácido nucleico que codifica un polipéptido y de la expresión mejorada de un polipéptido codificado por un ácido nucleico optimizado con el método que se describe en el presente documento.

Antecedentes en la invención

10 Cannarozzi, G., et al. describen la función del orden de los codones en la dinámica de la traducción (Cell 141 (2010) 355-367). Plotkin, J.B. y Kudla, G. (Nat. Rev. Gen. 12 (2011) 32-42) describen la causa y consecuencia del sesgo codónico. Weygand-Durasevic, I. e Ibba, M., describen nuevas funciones para el uso de codones (Science 329 (2010) 1473-1474). Itzkovitz, S., et al. (Gen. Res. 20 (2010) 1582-1589) describen códigos solapantes dentro de las secuencias que codifican proteínas.

15 En el documento WO 97/11086 se describe la expresión de proteínas a alto nivel. En el documento WO 03/70957 se describe la producción de polipéptidos en plantas. En el documento WO 03/85114 se describe un método de diseño de secuencias de ácido nucleico sintéticas para la expresión de proteínas óptima en una célula hospedadora. En el documento US 5.082.767 se describe la optimización con pares de codones. En el documento WO 2008/000632 se describe un método para conseguir la expresión mejorada de polipéptidos. En los documentos WO 2007/142954 y US 8.128.938 se describe un método de optimización con codones.

20 Watkins, N.E., et al., describen la termodinámica del vecino más próximo de pares de deoxinosina en dúplex de ADN (Nucl. Acids Res. 33 (2005) 6258-6267).

25 Plotkin, J. B., et al. elucubran sobre el hecho de que las causas y las consecuencias del sesgo codónico son comparables y no iguales (Nat. Rev. Genet. 12 (2011) 32-42). Gustafsson, C., et al. (Trends. Biotechnol. 22 (2004) 346-353) describieron sobre el sesgo codónico y la expresión de proteínas heterólogas. G. Cannarozzi et al. (Cell 141 (2010) 355-367) describieron una función para el orden de codones en la dinámica de la traducción.

Sumario de la invención

35 Se ha descubierto que para la expresión de un polipéptido en una célula, el uso de un ácido nucleico que codifica un polipéptido con las características que se describen en el presente documento, resulta ser beneficioso. El ácido nucleico que codifica un polipéptido se caracteriza por que cada aminoácido está codificado por un grupo de codones, por lo cual cada codón en el grupo de codones está definido por una frecuencia de uso específica dentro del grupo que está relacionada con la frecuencia de uso total de ese codón en el genoma de la célula, y por lo cual, la frecuencia de uso (total) de los codones en el ácido nucleico que codifica un polipéptido es aproximadamente la misma que la frecuencia de uso dentro del grupo respectivo.

40 Un aspecto, como se describe en el presente documento, es un método para la producción recombinante de un polipéptido en una célula bacteriana, que comprende la etapa de cultivar una célula que comprende un ácido nucleico que codifica el polipéptido, y recuperar el polipéptido de la célula bacteriana o del medio de cultivo, en el que cada uno de los restos de aminoácido del polipéptido está codificado por uno o más (al menos uno) codones, por lo cual los (diferentes) codones que codifican el mismo resto de aminoácido se combinan en un grupo y cada uno de los codones en un grupo está definido por una frecuencia de uso específica dentro del grupo, que es la frecuencia con la cual un solo codón de un grupo de codones puede encontrarse en un ácido nucleico que codifica un polipéptido en relación con todos los codones de un grupo, por lo cual la suma de las frecuencias de uso específicas de todos los codones en un grupo es del 100 %, donde la frecuencia de uso total de cada codón en el ácido nucleico que codifica un polipéptido es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.

45 En una realización, cada uno de los restos de aminoácido G, A, V, L, I, P, F, S, T, N, Q, Y, C, K, R, H, D y E está codificado por un grupo de codones y los restos de aminoácido M y W están codificados por un solo codón.

50 En una realización, cada uno de los restos de aminoácido G, A, V, L, I, P, F, S, T, N, Q, Y, C, K, R, H, D y E está codificado por un grupo de codones que comprende al menos dos codones y los restos de aminoácido M y W están codificados por un solo codón.

55 En una realización la frecuencia de uso específica de un codón es del 100 % si el resto de aminoácido está codificado exactamente por un codón.

60 En una realización el resto de aminoácido G está codificado por un grupo de a lo sumo 4 codones. En una realización, el resto de aminoácido A está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido V está codificado por un grupo de a lo sumo 4 codones. En una realización, el resto de

aminoácido L está codificado por un grupo de a lo sumo 6 codones. En una realización, el resto de aminoácido I está codificado por un grupo de a lo sumo 3 codones. En una realización el resto de aminoácido M está codificado exactamente por 1 codón. En una realización, el resto de aminoácido P está codificado por un grupo de a lo sumo 4 codones. En una realización, el resto de aminoácido F está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido W está codificado exactamente por 1 codón. En una realización el resto de aminoácido S está codificado por un grupo de a lo sumo 6 codones. En una realización el resto de aminoácido T está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido N está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido Q está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido Y está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido C está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido K está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido R está codificado por un grupo de a lo sumo 6 codones. En una realización el resto de aminoácido H está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido D está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido E está codificado por un grupo de a lo sumo 2 codones.

En una realización el resto de aminoácido G está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido A está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido V está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido L está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido I está codificado por un grupo de 1 a 3 codones. En una realización el resto de aminoácido M está codificado por un grupo de 1 codón, es decir, exactamente por 1 codón. En una realización el resto de aminoácido P está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido F está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido W está codificado por un grupo de 1 codón, es decir, exactamente por 1 codón. En una realización el resto de aminoácido S está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido T está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido N está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido Q está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido Y está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido C está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido K está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido R está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido H está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido D está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido E está codificado por un grupo de 1 a 2 codones.

En una realización, cada uno de los grupos comprende únicamente codones con una frecuencia de uso total dentro del genoma de la célula de más del 5 %. En una realización, cada uno de los grupos comprende únicamente codones con una frecuencia de uso total dentro del genoma de célula del 8 % o más. En una realización cada uno de los grupos comprende únicamente codones con una frecuencia de uso total dentro del genoma de la célula del 10 % o más. En una realización, cada uno de los grupos comprende únicamente codones con una frecuencia de uso total dentro del genoma de la célula del 5 % o más.

En una realización la secuencia de codones en el ácido nucleico que codifica el polipéptido para un resto de aminoácido específico en la dirección 5' a 3' es, es decir corresponde a, la secuencia de codones en un motivo codónico de aminoácidos s respectivo.

En una realización para cada aparición secuencial de un aminoácido específico en el polipéptido, comenzado desde el extremo N del polipéptido, el ácido nucleico codificante comprende el codón que es el mismo que el de la posición secuencial correspondiente en el motivo codónico de aminoácidos del aminoácido específico respectivo, por lo cual después de la primera aparición del resto de aminoácido en la secuencia de aminoácidos del polipéptido, el primer codón del motivo codónico de aminoácidos se usa en el ácido nucleico codificante correspondiente, después de la segunda aparición del resto de aminoácido se usa el segundo codón del motivo codónico de aminoácidos y así sucesivamente.

En una realización la frecuencia de uso de un codón en el motivo codónico de aminoácidos es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.

En una realización después de que el codón final del motivo codónico de aminoácidos haya llegado a la siguiente aparición del aminoácido específico en el polipéptido, el ácido nucleico codificante comprende el codón que está en la primera posición del motivo codónico de aminoácidos.

En una realización los codones en el motivo codónico de aminoácidos se distribuyen al azar a lo largo del motivo codónico de aminoácidos.

En una realización el motivo codónico de aminoácidos se selecciona de un grupo de motivos codónicos de aminoácidos que comprenden todos los posibles motivos codónicos de aminoácidos obtenibles permutando codones en su interior donde todos los motivos tienen el mismo número de codones y los codones en cada motivo tienen la

misma frecuencia de uso específica.

5 En una realización los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente por lo cual todos los codones de una frecuencia de uso se suceden directamente entre sí. En una realización los codones de una frecuencia de uso de codones se agrupan conjuntamente.

En una realización los (diferentes) codones en el motivo codónico de aminoácidos se distribuyen uniformemente a lo largo del motivo codónico de aminoácidos.

10 En una realización los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente por lo cual después del codón con la frecuencia de uso específica más baja o del codón con la segunda frecuencia de uso específica más baja, está presente (se usa) el codón con la frecuencia de uso específica más alta.

15 En una realización los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente por lo cual después del codón con la frecuencia de uso específica más baja está presente (se usa) el codón con la frecuencia de uso específica más alta.

En una realización la célula bacteriana es una célula de *E.coli*.

20 En una realización el motivo codónico de aminoácidos para alanina se selecciona de SEQ ID NO: 01,02, 03, 04 y 05, y/o arginina se selecciona de SEQ ID NO: 06 y 07, y/o asparraguina se selecciona de SEQ ID NO: 08, 09, 10, 11, y 12, y/o  
 25 ácido aspártico se selecciona de SEQ ID NO: 13 y 14, y/o, cisteína se selecciona de SEQ ID NO: 15, 16 y 17, y/o glutamina se selecciona de SEQ ID NO: 18, 19, 20, y 21, y/o ácido glutámico se selecciona de SEQ ID NO: 22, 23 y 24, y/o glicina se selecciona de SEQ ID NO: 25 y 26, y/o  
 30 histidina se selecciona de SEQ ID NO: 27 y 28, y/o isoleucina se selecciona de SEQ ID NO: 29 y 30, y/o leucina se selecciona de SEQ ID NO: 31,32 y 33, y/o lisina se selecciona de SEQ ID NO: 34, 35, 36 y 37, y/o fenilalanina se selecciona de SEQ ID NO: 38, 39 y 40, y/o  
 35 prolina se selecciona de SEQ ID NO: 41,42, 43, 44, 45 y 46, y/o serina se selecciona de, SEQ ID NO: 47 y 48, y/o treonina se selecciona de SEQ ID NO: 49, 50 y 51, y/o tirosina se selecciona de SEQ ID NO: 52 y 53, y/o valina se selecciona de SEQ ID NO: 54, 55 y 56.

40 En una realización el motivo codónico de aminoácidos para alanina es SEQ ID NO: 03, arginina es SEQ ID NO: 07, asparraguina es SEQ ID NO: 10,  
 45 ácido aspártico es SEQ ID NO: 13, cisteína es SEQ ID NO: 17, glutamina es SEQ ID NO: 20, ácido glutámico es SEQ ID NO: 23, glicina es SEQ ID NO: 26, histidina es SEQ ID NO: 28, isoleucina es SEQ ID NO: 30, leucina es SEQ ID NO: 33;  
 50 lisina es SEQ ID NO: 36, fenilalanina es SEQ ID NO: 39, prolina es SEQ ID NO: 43, serina es SEQ ID NO: 48, treonina es SEQ ID NO: 51, tirosina es SEQ ID NO: 53, y valina es SEQ ID NO: 56.

60 En una realización el motivo codónico de aminoácidos para alanina se selecciona de SEQ ID NO: 64, 65, 66, 67 y 68, y/o arginina se selecciona de SEQ ID NO: 69 y 70, y/o asparraguina se selecciona de SEQ ID NO: 71 y 72, y/o  
 65 ácido aspártico se selecciona de SEQ ID NO: 73 y 74, y/o, cisteína se selecciona de SEQ ID NO: 75 y 76, y/o

- glutamina se selecciona de SEQ ID NO: 77, 78, 79, y 80, y/o  
 ácido glutámico se selecciona de SEQ ID NO: 81 y 82, y/o  
 glicina se selecciona de SEQ ID NO: 83 y 84, y/o  
 histidina se selecciona de SEQ ID NO: 85 y 86, y/o  
 5 isoleucina se selecciona de SEQ ID NO: 87 y 88, y/o  
 leucina se selecciona de SEQ ID NO: 89, 90 y 91, y/o  
 lisina se selecciona de SEQ ID NO: 92 y 93, y/o  
 fenilalanina se selecciona de SEQ ID NO: 94 y 95, y/o  
 prolina se selecciona de SEQ ID NO: 96 y 97, y/o  
 10 serina se selecciona, SEQ ID NO: 98, 99 y 100, y/o  
 treonina se selecciona de SEQ ID NO: 101, 102 y 103, y/o  
 tirosina se selecciona de SEQ ID NO: 104 y 105, y/o  
 valina se selecciona de SEQ ID NO: 106, 107 y 108.
- 15 En una realización el motivo codónico de aminoácidos para  
 alanina es SEQ ID NO: 68,  
 arginina es SEQ ID NO: 69,  
 asparraguina es SEQ ID NO: 72,  
 20 ácido aspártico es SEQ ID NO: 74,  
 cisteína es SEQ ID NO: 76,  
 glutamina es SEQ ID NO: 79,  
 ácido glutámico es SEQ ID NO: 82,  
 glicina es SEQ ID NO: 84,  
 histidina es SEQ ID NO: 86,  
 25 isoleucina es SEQ ID NO: 88,  
 leucina es SEQ ID NO: 90;  
 lisina es SEQ ID NO: 93,  
 fenilalanina es SEQ ID NO: 95,  
 prolina es SEQ ID NO: 97,  
 30 serina es SEQ ID NO: 99,  
 treonina es SEQ ID NO: 103,  
 tirosina es SEQ ID NO: 105, y  
 valina es SEQ ID NO: 108.
- 35 En una realización el polipéptido es un anticuerpo, o un fragmento de anticuerpo, o un polipéptido de fusión con anticuerpo.
- En el presente documento se describe un ácido nucleico que codifica un polipéptido, caracterizado por que cada uno de los restos de aminoácidos del polipéptido está codificado por uno o más (al menos uno) codones,  
 40 por lo cual los diferentes codones que codifican el mismo resto de aminoácido se combinan en un grupo y cada uno de los codones en un grupo se define por una frecuencia de uso específica dentro del grupo, que es la frecuencia con la que puede encontrarse un solo codón de un grupo de codones en un ácido nucleico que codifica un polipéptido en relación con todos los codones de un grupo, por lo cual la suma de las frecuencias de uso específicas de todos los codones en un grupo es del 100 %, en el que la frecuencia de uso de un codón en el ácido nucleico que  
 45 codifica el polipéptido es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.
- En una realización, cada uno de los restos de aminoácidos G, A, V, L, I, P, F, S, T, N, Q, Y, C, K, R, H, D y E están codificados por un grupo de codones y los restos de aminoácidos M y W están codificados por un solo codón.
- 50 En una realización cada uno de los restos de aminoácidos G, A, V, L, I, P, F, S, T, N, Q, Y, C, K, R, H, D y E están codificados por un grupo de codones que comprenden al menos dos codones y los restos de aminoácidos M y W están codificados por un solo codón.
- En una realización la frecuencia de uso específica de un codón es del 100 % si el resto de aminoácido está  
 55 codificado exactamente por un codón.
- En una realización el resto de aminoácido G está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido A está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido V está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido L está codificado por un grupo de a lo sumo 6 codones. En una realización el resto de aminoácido I está  
 60 codificado por un grupo de a lo sumo 3 codones. En una realización el resto de aminoácido M está codificado exactamente por 1 codón. En una realización el resto de aminoácido P está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido F está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido W está codificado exactamente por 1 codón. En una realización el resto de aminoácido S está codificado por un grupo de a lo sumo 6 codones. En una realización el resto de aminoácido T está codificado por un grupo de a lo sumo 4 codones. En una realización el resto de aminoácido N está codificado
- 65

- 5 por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido Q está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido Y está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido C está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido K está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido R está codificado por un grupo de a lo sumo 6 codones. En una realización el resto de aminoácido H está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido D está codificado por un grupo de a lo sumo 2 codones. En una realización el resto de aminoácido E está codificado por un grupo de a lo sumo 2 codones.
- 10 En una realización el resto de aminoácido G está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido A está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido V está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido L está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido I está codificado por un grupo de 1 a 3 codones. En una realización el resto de aminoácido M está codificado por un grupo de 1 codón. En una realización el resto de aminoácido P está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido F está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido W está codificado por un grupo de 1 codón. En una realización el resto de aminoácido S está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido T está codificado por un grupo de 1 a 4 codones. En una realización el resto de aminoácido N está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido Q está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido Y está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido C está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido K está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido R está codificado por un grupo de 1 a 6 codones. En una realización el resto de aminoácido H está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido D está codificado por un grupo de 1 a 2 codones. En una realización el resto de aminoácido E está codificado por un grupo de 1 a 2 codones.
- 15 En una realización cada uno de los grupos comprende solo codones con una frecuencia de uso total dentro del genoma de la célula de más del 5 %. En una realización cada uno de los grupos comprende solo codones con una frecuencia de uso total dentro del genoma de la célula del 8 % o más. En una realización, cada uno de los grupos comprende solo codones con una frecuencia de uso total dentro del genoma de la célula del 10 % o más. En una realización, cada uno de los grupos comprende solo codones con una frecuencia de uso total dentro del genoma de la célula del 15 % o más.
- 20 En una realización la secuencia de codones en el ácido nucleico que codifica el polipéptido para un resto de aminoácido específico en dirección 5' a 3' es, es decir, corresponde a, la secuencia de codones en un motivo codónico de aminoácidos respectivo.
- 25 En una realización para cada aparición secuencial de un aminoácido específico en el polipéptido, comenzado desde el extremo N del polipéptido, el ácido nucleico codificante comprende el codón que es el mismo que el de la posición secuencial correspondiente en el motivo codónico de aminoácidos del aminoácido específico respectivo, por lo cual después de la primera aparición del resto de aminoácido en la secuencia de aminoácidos del polipéptido, el primer codón del motivo codónico de aminoácidos se usa en el ácido nucleico codificante correspondiente, después de la segunda aparición del resto de aminoácido se usa el segundo codón del motivo codónico de aminoácidos y así sucesivamente.
- 30 En una realización la frecuencia de uso de un codón en el motivo codónico de aminoácidos es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.
- 35 En una realización después de que el codón final del motivo codónico de aminoácidos haya llegado a la siguiente aparición del aminoácido específico en el polipéptido el ácido nucleico codificante comprende el codón que está en la primera posición del motivo codónico de aminoácidos.
- 40 En una realización cada uno de los codones en el motivo codónico de aminoácidos se distribuye al azar a lo largo del motivo codónico de aminoácidos.
- 45 En una realización cada uno de los codones en el motivo codónico de aminoácidos se distribuye uniformemente a lo largo del motivo codónico de aminoácidos.
- 50 En una realización los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente por lo cual después del codón con la frecuencia de uso específica más baja o del codón con la segunda frecuencia de uso específica más baja, se usa el codón con la frecuencia de uso específica más alta.
- 55 En una realización los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente por lo cual después del codón con la frecuencia de uso específica más baja se usa el codón con la frecuencia de uso específica más alta.
- 60
- 65

En el presente documento se describe una célula que comprende un ácido nucleico como se indica en el presente documento.

En el presente documento se describe un método para aumentar la expresión de un polipéptido en una célula procariota o en una célula eucariota que comprende la etapa de,

- proporcionar un ácido nucleico que codifique el polipéptido,

en el que cada uno de los restos de aminoácido del polipéptido está codificado por al menos un codón, por lo cual los diferentes codones que codifican el mismo resto de aminoácido se combinan en un grupo y cada uno de los codones en un grupo se define por una frecuencia de uso específica dentro del grupo, por lo cual la suma de las frecuencias de uso específicas de todos los codones en un grupo es del 100 %,

en el que la frecuencia de uso de un codón en el ácido nucleico que codifica el polipéptido es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.

Descripción detallada de la invención

#### DEFINICIONES

El término "aminoácido", como se usa en la presente solicitud, significa el grupo de carboxi  $\alpha$  aminoácidos que directamente o en forma de un precursor, puede codificar un ácido nucleico. Cada uno de los aminoácidos individuales está codificado por ácidos nucleicos que consisten en tres nucleótidos, denominados codones o tripletes de bases. Cada aminoácido está codificado por al menos un codón. La codificación del mismo aminoácido por diferentes codones se conoce como "degeneración del código genético". El término "aminoácido", como se usa en la presente solicitud, significa los carboxi  $\alpha$  aminoácidos de origen natural y comprende alanina (código de tres letras: Ala, código de una letra: A), arginina (Arg, R), asparragina (Asn, N), ácido aspártico (Asp, D), cisteína (Cys, C), glutamina (Gln, Q), ácido glutámico (Glu, E), glicina (Gly, G), histidina (His, H), isoleucina (Ile, I), leucina (Ile, L), lisina (Lys, K), metionina (Met, M), fenilalanina (Phe, F), prolina (pro, P), serina (ser, S), treonina (Thr, T), triptófano (Trp, W), tirosina (Tyr, Y) y valina (val, V).

El término "anticuerpo" se usa en el presente documento en el sentido más general e incluye las diversas estructuras de anticuerpos, incluyendo, pero sin limitación, anticuerpos monoclonales, anticuerpos policlonales, anticuerpos multiespecíficos (por ejemplo, anticuerpos biespecíficos), y fragmentos de anticuerpo siempre que exhiban la actividad de unión antigénica deseada.

Un "fragmento de anticuerpo" se refiere a una molécula distinta de un anticuerpo intacto que comprende una porción de un anticuerpo intacto que se une al antígeno al cual se une el anticuerpo intacto. Los ejemplos de fragmentos de anticuerpo incluyen, pero sin limitación Fv, Fab, Fab', Fab'-SH, F(ab')<sub>2</sub>; diacuerpos; anticuerpos lineales; moléculas de anticuerpo de una sola cadena (por ejemplo, scFv) y anticuerpos multiespecíficos formados a partir de fragmentos de anticuerpo.

El término "codón" significa un oligonucleótido que consiste en tres nucleótidos, es decir, que codifica un aminoácido definido. Debido a la degeneración del código genético la mayoría de los aminoácidos están codificados por más de un codón. Estos diferentes codones que codifican el mismo aminoácido tienen diferentes frecuencias de uso relativas en células hospedadoras individuales. Por tanto, un aminoácido específico está codificado bien por un codón exactamente o por un grupo de codones diferentes. Asimismo, la secuencia de aminoácidos de un polipéptido puede estar codificada por ácidos nucleicos diferentes. Por lo tanto, un aminoácido (resto) específico en un polipéptido puede estar codificado por un grupo de codones diferentes, por lo cual cada uno de estos codones tiene una frecuencia de uso dentro de una célula hospedadora determinada.

Dado que se dispone de una gran cantidad de secuencias génicas para diversas células hospedadoras frecuentemente usadas, las frecuencias relativas del uso de codones pueden calcularse. Las tablas de las frecuencias calculadas del uso de codones están disponibles, por ejemplo, en la "Base de Datos de Uso de Codones" ([www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/)), Nakamura, Y., et al., Nucl. Acids Res. 28 (2000) 292.

Las tablas de uso de codones de levaduras, *E. coli*, *Homo sapiens* y hámster se han reproducido a partir de "EMBOSS: The European Molecular Biology Open Software Suite" (Rice, P., et al., Trends Gen. 16 (2000) 276-277, Release 6.0.1, 15.07.2009) y se muestran en las siguientes tablas. Las frecuencias del uso de codones diferentes para los 20 aminoácidos de origen natural de *E. coli*, levaduras, células humanas y células CHO, se han calculado para cada aminoácido, en lugar de para los 64 codones.

## ES 2 599 386 T3

Tabla: Frecuencia de uso total de codones de *Saccharomyces cerevisiae* (aminoácido codificado | codón | frecuencia de uso [%])

Ala	GCG	1	Gly	GGG	1	Pro	CCG	1
Ala	GCA	6	Gly	GGA	3	Pro	CCA	80
Ala	GCT	64	Gly	GGT	89	Pro	CCT	17
Ala	GCC	29	Gly	GGC	7	Pro	CCC	2
Arg	AGG	3	His	CAT	37	Ser	AGT	6
Arg	AGA	77	His	CAC	63	Ser	AGC	5
Arg	CGG	0				Ser	TCG	2
Arg	CGA	0	He	ATA	2	Ser	TCA	8
Arg	CGT	19	He	ATT	52	Ser	TCT	49
Arg	CGC	1	He	ATC	45	Ser	TCC	31
Asn	AAT	23	Leu	CTG	3	Thr	ACG	1
Asn	AAC	77	Leu	CTA	9	Thr	ACA	8
			Leu	CTT	4	Thr	ACT	52
Asp	GAT	49	Leu	CTC	1	Thr	ACC	40
Asp	GAC	51	Leu	TTG	64			
			Leu	TTA	20	Trp	TGG	100
Cys	TGT	87						
Cys	TGC	13	Lys	AAG	74	Tyr	TAT	22
			Lys	AAA	26	Tyr	TAC	78
Gln	CAG	6						
Gln	CAA	94	Met	ATG	100	Val	GTG	4
						Val	GTA	3
Glu	GAG	11	Phe	TTT	29	Val	GTT	54
Glu	GAA	89	Phe	TTC	71	Val	GTC	38

Tabla: Frecuencia de uso total de codones total de *Escherichia coli* (aminoácido codificado | codón | frecuencia de uso [%])

Ala	GCG	32	Gly	GGG	4	Pro	CCG	75
Ala	GCA	24	Gly	GGA	2	Pro	CCA	14
Ala	GCT	28	Gly	GGT	51	Pro	CCT	10
Ala	GCC	16	Gly	GGC	43	Pro	CCC	1
Arg	AGG	0	His	CAT	29	Ser	AGT	4
Arg	AGA	0	His	CAC	71	Ser	AGC	25
Arg	CGG	1				Ser	TCG	7
Arg	CGA	1	He	ATA	0	Ser	TCA	5
Arg	CGT	65	He	ATT	32	Ser	TCT	32
Arg	CGC	33	He	ATC	68	Ser	TCC	27
Asn	AAT	16	Leu	CTG	79	Thr	ACG	12
Asn	AAC	84	Leu	CTA	1	Thr	ACA	4
			Leu	CTT	5	Thr	ACT	28
Asp	GAT	46	Leu	CTC	8	Thr	ACC	56
Asp	GAC	54	Leu	TTG	5			
			Leu	TTA	3	Trp	TGG	100
Cys	TGT	36						
Cys	TGC	64	Lys	AAG	20	Tyr	TAT	36
			Lys	AAA	80	Tyr	TAC	64
Gln	CAG	82						
Gln	CAA	18	Met	ATG	100	Val	GTG	27
						Val	GTA	20
Glu	GAG	24	Phe	TTT	28	Val	GTT	40
Glu	GAA	76	Phe	TTC	72	Val	GTC	13

5

Tabla: Frecuencia de uso total de codones de *Homo sapiens* (aminoácido codificado | codón | frecuencia de uso [%])

Ala	GCG	10	Gly	GGG	24	Pro	CCG	11
Ala	GCA	22	Gly	GGA	25	Pro	CCA	28
Ala	GCT	27	Gly	GGT	17	Pro	CCT	28
Ala	GCC	41	Gly	GGC	34	Pro	CCC	33
Arg	AGG	20	His	CAT	40	Ser	AGT	15
Arg	AGA	20	His	CAC	60	Ser	AGC	25
Arg	CGG	20				Ser	TCG	6



ES 2 599 386 T3

Arg	CGA	11	He	ATA	15	Ser	TCA	14
Arg	CGT	9	He	ATT	35	Ser	TCT	18
Arg	CGC	20	He	ATC	50	Ser	TCC	23
Asn	AAT	45	Leu	CTG	42	Thr	ACG	12
Asn	AAC	55	Leu	CTA	7	Thr	ACA	27
			Leu	CTT	13	Thr	ACT	24
Asp	GAT	46	Leu	CTC	20	Thr	ACC	37
Asp	GAC	54	Leu	TTG	12			
			Leu	TTA	7	Trp	TGG	100
Cys	TGT	44						
Cys	TGC	56	Lys	AAG	59	Tyr	TAT	43
			Lys	AAA	41	Tyr	TAC	57
Gln	CAG	74						
Gln	CAA	26	Met	ATG	100	Val	GTG	47
						Val	GTA	11
Glu	GAG	58	Phe	TTT	45	Val	GTT	17
Glu	GAA	42	Phe	TTC	55	Val	GTC	25

Tabla: Frecuencia de uso total de codones de Hámster (aminoácido codificado | codón | frecuencia de uso [%])

Ala	GCG	9	Gly	GGG	24	Pro	CCG	7
Ala	GCA	23	Gly	GGA	25	Pro	CCA	29
Ala	GCT	30	Gly	GGT	19	Pro	CCT	29
Ala	GCC	38	Gly	GGC	33	Pro	CCC	34
Arg	AGG	22	His	CAT	42	Ser	AGT	14
Arg	AGA	20	His	CAC	58	Ser	AGC	24
Arg	CGG	19				Ser	TCG	5
Arg	CGA	9	He	ATA	15	Ser	TCA	15
Arg	CGT	10	He	ATT	35	Ser	TCT	18
Arg	CGC	19	He	ATC	51	Ser	TCC	24
Asn	AAT	39	Leu	CTG	44	Thr	ACG	10
Asn	AAC	61	Leu	CTA	6	Thr	ACA	29
			Leu	CTT	13	Thr	ACT	21
Asp	GAT	39	Leu	CTC	19	Thr	ACC	40
Asp	GAC	61	Leu	TTG	12			
			Leu	TTA	6	Trp	TGG	100
Cys	TGT	42						
Cys	TGC	58	Lys	AAG	67	Tyr	TAT	39
			Lys	AAA	33	Tyr	TAC	61
Gln	CAG	78						
Gln	CAA	22	Met	ATG	100	Val	GTG	48
						Val	GTA	11
Glu	GAG	64	Phe	TTT	44	Val	GTT	16
Glu	GAA	36	Phe	TTC	56	Val	GTC	25

5 El término “expresión”, como se usa en el presente documento, se refiere a los procesos de transcripción y/o de traducción que se producen en una célula. El nivel de transcripción de una secuencia de ácido nucleico de interés en una célula puede terminarse basándose en la cantidad de ARNm correspondiente que está presente en la célula. Por ejemplo, un ARNm transcrito a partir de una secuencia de interés puede cuantificarse por RT-PCR (cRT-PCR) o mediante hibridación Northern (véase, Sambrook, J., et al., 1989, citado anteriormente). Los polipéptidos codificados por un ácido nucleico de interés pueden cuantificarse mediante diversos métodos, por ejemplo, por Elisa, mediante ensayo de la actividad biológica del polipéptido o empleando ensayos que son independientes de dicha actividad, tales como transferencia de Western y radioinmunoensayo, usando inmunoglobulinas que reconocen el polipéptido y se unen al mismo (véase, Sambrook, J., et al., 1989, citado anteriormente).

15 Un “casete de expresión” se refiere a una construcción que contiene los elementos reguladores necesarios, tales como un promotor y un sitio de poliadenilación, para la expresión de al menos el ácido nucleico contenido en una célula.

20 La expresión de un gen se realiza bien como expresión transitoria o como expresión permanente. El polipéptido o polipéptidos de interés son, en general, polipéptidos secretados y por lo tanto contienen una extensión N terminal (también conocida como las secuencia señal) que es necesaria para el transporte/secreción del polipéptido a través de la pared celular hacia el interior del medio extracelular. En general, la secuencia señal puede proceder de cualquier gen que codifique un polipéptido secretado. Si se usa una secuencia señal heteróloga, esta es

preferentemente una que es reconocida y procesada (es decir, escindida por una peptidasa señal) por la célula hospedadora. Para la secreción en levaduras, por ejemplo, la secuencia señal nativa de un gen heterólogo a expresar puede sustituirse por una secuencia de señal de levadura homóloga procedente de un gen secretado, tal como la secuencia señal invertasa de levadura, el líder factor  $\alpha$  (incluyendo líderes de factor  $\alpha$  de *Saccharomyces*, *Kluyveromyces*, *Pichia* y *Hansenula*, el segundo se describe en el documento US 5.010.182), la secuencia señal de la fosfatasa ácida, o la secuencia señal de la glucoamilasa de *C. albicans* (documento EP 0 362 179). En la expresión de células de mamífero, la secuencia señal nativa de la proteína de interés es satisfactoria, aunque pueden ser adecuadas otras secuencias señal de mamífero, tales como las secuencias señal de polipéptidos secretados de la misma especie o de especies relacionadas, por ejemplo, para inmunoglobulinas de origen humano o murino, así como secuencias señal secretoras víricas, por ejemplo, la secuencia señal de la glucoproteína D del herpes simple. El fragmento de ADN que codifica dicho pre-segmento está ligado en fase, es decir, está unido operativamente, al fragmento de ADN que codifica un polipéptido de interés.

El término "célula", o la expresión "célula hospedadora", se refieren a una célula en la que un ácido nucleico, por ejemplo, que codifica un polipéptido heterólogo, puede estar transfectedo o se transfecta. El término "célula" incluye células tanto procariotas, que se usan para la expresión de un ácido nucleico y para la producción del polipéptido codificado incluyendo la propagación de plásmidos, como células eucariotas, que se usan para la expresión de un ácido nucleico y para la producción del polipéptido codificado. Como se usa en el presente documento, el término "célula" incluye la célula en cuestión y su descendencia. Por tanto, las palabras "transformante" y "célula transformada" incluyen la célula primaria en cuestión y los cultivos procedentes de la misma sin importar el número de transferencias o subcultivos. También se entiende que toda la descendencia no ha de ser exactamente idéntica en cuanto al contenido de ADN, debido a mutaciones intencionadas o accidentales. También se incluye la descendencia variante que tiene la misma función o actividad biológica que la detectada en la célula originalmente transformada.

En una realización, la célula bacteriana es una célula de *Escherichia*, o una célula de *Salmonella*, o una célula de *Bacillus*, o una célula de *Streptococcus*.

La expresión "optimización de codones" significa el intercambio de uno, de al menos uno, o de más de un codón, en un ácido nucleico que codifica un polipéptido para un codón diferente con una frecuencia de uso diferente en una célula respectiva.

La expresión "ácido nucleico optimizado por codones" significa que un ácido nucleico que codifica un polipéptido se ha adaptado para mejorar la expresión en una célula, por ejemplo, una célula de mamífero o una célula bacteriana, reemplazando uno, al menos uno, o más de un codón en un ácido nucleico que codifica un polipéptido parental, con un codón que codifica el mismo resto de aminoácido con una frecuencia de uso relativa diferente en la célula.

Un "gen" significa un ácido nucleico que es un segmento, por ejemplo, en un cromosoma o en un plásmido que puede efectuar la expresión de un péptido, polipéptido o proteína. Además de la región codificante, es decir, el gen estructural, un gen comprende otros elementos funcionales, por ejemplo, una secuencia señal, uno o más promotores, intrones y/o terminadores.

La expresión "grupos de codones" y sus equivalentes semánticos, significa un número definido de codones diferentes que codifican un (es decir, el mismo) resto de aminoácido. Los codones individuales de un grupo difieren en su frecuencia de uso total en el genoma de una célula. Cada codón en un grupo de codones tiene una frecuencia de uso específica dentro del grupo que depende del número de codones en el grupo. Esta frecuencia de uso específica dentro del grupo puede ser diferente de la frecuencia de uso total en el genoma de una célula pero esto depende (en relación con eso) de la frecuencia de uso total. Un grupo de codones puede comprender solo un codón, pero también puede comprender hasta seis codones.

La expresión "frecuencia de uso total en el genoma de una célula" significa la frecuencia de la aparición de un codón específico en todo el genoma de una célula.

La expresión "frecuencia de uso específica" de un codón en un grupo de codones significa la frecuencia con la que puede encontrarse un solo codón (es decir, específico) de un grupo de codones, en relación con todos los codones de un grupo, en un ácido nucleico que codifica un polipéptido obtenido con un método como se describe en el presente documento. El valor de la frecuencia de uso específica depende de la frecuencia de uso total del codón específico en el genoma de una célula y del número de codones en el grupo. Por lo tanto, como un grupo de codones no comprende necesariamente todos los posibles codones que codifican un resto de aminoácido específico, la frecuencia de uso de codones específica de un codón en un grupo de codones es al menos la misma que su frecuencia de uso total en el genoma de una célula y a lo sumo el 100 %, es decir, es al menos la misma, pero puede ser mayor que la frecuencia de uso total en el genoma de una célula. La suma de las frecuencias de uso de codones específicas de todos los miembros de un grupo de codones es siempre de aproximadamente 100 %.

La expresión "motivo codónico de aminoácidos" significa una secuencia de codones, que todos son miembros del mismo grupo de codones y, por tanto, codifican el mismo resto de aminoácido. El número de codones diferentes en

un motivo codónico de aminoácidos es el mismo que el número de codones diferentes en un grupo de codones, pero cada codón puede estar presente más de una vez en el motivo codónico de aminoácidos. Adicionalmente, cada codón está presente en el motivo codónico de aminoácidos en su frecuencia de uso específica. Por lo tanto, el motivo codónico de aminoácidos representa una secuencia de codones diferentes que codifican le mismo resto de aminoácido en el que cada uno de los diferentes codones está presente en su frecuencia de uso específica, en el que la secuencia comienza con el codón que tiene la frecuencia de uso específica más alta, y en el que los codones se disponen en una secuencia definida. Por ejemplo, el grupo de codones que codifica el resto de aminoácido alanina comprende los cuatro codones GCG, GCT, GCA y GCC con una frecuencia de uso específica del 32 %, 28 %, 24 % y 16 %, respectivamente (correspondiente a una proporción de 4:4:3:2). El motivo codónico de aminoácidos para el resto de aminoácido alanina se define por que comprende los cuatro codones GCG, GCT, GCA, y GCC a una proporción de 4:4:3:2, en el que el primer codón es GCG. Un motivo codónico de aminoácidos ejemplar para la alanina es gcg gcg gcg gcg gct gct gct gct gca gca gca gcc gcc (SEQ ID NO: 01). Este motivo consiste en trece codones secuenciales (4+4+3+2=13). Después de la primera aparición del resto de aminoácido alanina en la secuencia de aminoácidos de un polipéptido se usa el primer codón del motivo codónico de aminoácidos en el ácido nucleico codificante correspondiente. Después de la segunda aparición de la alanina, se usa el segundo codón del motivo codónico de aminoácidos y así sucesivamente. Después de la decimotercera aparición de la alanina en la secuencia de aminoácidos del polipéptido se usa el codón en la decimotercera posición, es decir, la última posición del motivo codónico de aminoácidos, en el ácido nucleico codificante correspondiente. Después de la decimocuarta aparición del aminoácido alanina en la secuencia de aminoácidos del polipéptido de nuevo se usa el primer codón del motivo codónico de aminoácidos y así sucesivamente.

Un "ácido nucleico" o una "secuencia de ácido nucleico", cuyos términos se usan indistintamente en esta solicitud, se refieren a una molécula polimérica que consta de los nucleótidos individuales (también denominados bases) a, c, g, y t (o u en el ARN), por ejemplo, para ADN, ARN o modificaciones de los mismos. Esta molécula polinucleotídica puede ser una molécula polinucleotídica de origen natural o una molécula polinucleotídica sintética o una combinación de una o más moléculas polinucleotídicas de origen natural con una o más moléculas polinucleotídicas sintéticas. También se incluyen en esta definición las moléculas polinucleotídicas de origen natural en las que se uno o más nucleótidos se cambian (por ejemplo, por mutagénesis), se delecionan o se añaden. Un ácido nucleico puede estar bien aislado, o integrado en otro ácido nucleico, por ejemplo, en un casete de expresión, en un plásmido, o en el cromosoma de una célula hospedadora. Un ácido nucleico se caracteriza por su secuencia de ácido nucleico que consta de nucleótidos individuales.

Para un experto en la técnica, son muy conocidos los procedimientos y métodos para transformar una secuencia de aminoácidos, por ejemplo, de un polipéptido, en una secuencia de ácido nucleico correspondiente que codifica esta secuencia de aminoácidos. Por lo tanto, un ácido nucleico se caracteriza por su secuencia de ácido nucleico que consiste en nucleótidos individuales y del mismo modo por la secuencia de aminoácidos de un polipéptido codificada por la misma.

Un "gen estructural" significa la región de un gen sin una secuencia señal, es decir, la región codificante.

Un "vector de transfección" es un ácido nucleico (también denominado molécula de ácido nucleico) que proporciona todos los elementos necesarios para la expresión en el vector de transfección que comprende los ácidos nucleicos/gen o genes estructurales en una célula hospedadora. Un vector de transfección comprende una unidad de propagación de plásmido procariota, por ejemplo, para *E. coli*, que, a su vez, comprende un origen de replicación procariota, y un ácido nucleico que confiere resistencia contra un agente de selección procariota adicionalmente comprende el vector de transfección, uno o más ácidos nucleicos que confieren resistencia contra un agente de selección eucariota y uno o más ácidos nucleicos que codifican un polipéptido de interés. Preferentemente son los ácidos nucleicos los que confieren resistencia contra un agente de selección y el ácido nucleico o los ácidos nucleicos que codifican un polipéptido de interés colocados cada uno dentro de un casete de expresión, por lo cual cada casete de expresión comprende un promotor, un ácido nucleico codificante, y un terminador de la transcripción incluyendo una señal de poliadenilación. La expresión génica se coloca normalmente bajo el control de un promotor, y se dice que dicho gen estructural está "unido operativamente" al promotor. De manera similar, un elemento regulador y un promotor núcleo están unidos operativamente si el elemento regulador modula la actividad del promotor núcleo.

El término "vector", como se usa en el presente documento, se refiere a una molécula de ácido nucleico capaz de propagar otro ácido nucleico al cual está unido. El término incluye el vector como una estructura de ácido nucleico autorreplicante, así como el vector incorporado en el genoma de una célula hospedadora en la que se ha introducido. Determinados vectores tienen la capacidad de dirigir la expresión de ácidos nucleicos con los cuales están unidos operativamente. En el presente documento dichos vectores reciben el nombre de "vectores de expresión".

## MÉTODOS RECOMBINANTES

Los anticuerpos pueden producirse usando métodos recombinantes y composiciones, por ejemplo, como se describe en el documento US 4.816.567. En una realización, se proporciona ácido nucleico aislado que codifica un anticuerpo como se indica en el presente documento. Dicho ácido nucleico puede codificar una secuencia de aminoácidos que

comprende la región VL y/o una secuencia de aminoácidos que comprende la región VH del anticuerpo (por ejemplo, la cadena ligera y/o pesada del anticuerpo). En una realización, se proporciona uno o más vectores (por ejemplo, vectores de expresión) que comprenden dicho ácido nucleico. En una realización, se proporciona una célula que comprende dicho ácido nucleico. En una realización, una célula comprende (por ejemplo, se ha transformado con):

5 (1) un vector que comprende un ácido nucleico que codifica una secuencia de aminoácidos que comprende la región VL del anticuerpo y una secuencia de aminoácidos que comprende la región VH del anticuerpo, o (2) un primer vector que comprende un ácido nucleico que codifica una secuencia de aminoácidos que comprende la región VL del anticuerpo y un segundo vector que comprende un ácido nucleico que codifica una secuencia de aminoácidos que comprende la región VH del anticuerpo. En una realización, la célula es eucariota, por ejemplo, una célula de ovario de hámster chino (CHO) o una célula linfocítica (por ejemplo, una célula Y0, NS0, Sp2/0). En una realización, se proporciona un método para fabricar un anticuerpo, en el que el método comprende cultivar una célula que comprenda un ácido nucleico que codifique el anticuerpo, como se proporciona en el presente documento, en condiciones adecuadas para la expresión del anticuerpo, y opcionalmente recuperar el anticuerpo de la célula (o del medio de cultivo).

15 Para la producción recombinante de un anticuerpo, un ácido nucleico que codifica un anticuerpo, por ejemplo, como se indica en el presente documento, se aísla y se inserta en uno o más vectores para su clonación y/o expresión adicional en una célula. Dicho ácido nucleico puede aislarse y secuenciarse fácilmente usando procedimientos convencionales (por ejemplo, usando sondas oligonucleotídicas que tienen la capacidad de unirse específicamente a genes que codifican las cadenas pesada o ligera del anticuerpo).

20 Como células adecuadas para la clonación o expresión de vectores que codifican anticuerpos se incluyen células procariotas o eucariotas descritas en el presente documento. Por ejemplo, los anticuerpos pueden producirse en bacterias, en particular cuando no se requieren la glucosilación ni la función efectora de Fc. Para la expresión de fragmentos de anticuerpos y de polipéptidos en bacterias, véanse, por ejemplo, los documentos US 5.648.237, US 5.789.199, y US 5.840.523; Charlton, K.A., In: Methods in Molecular Biology, Vol. 248, Lo, B.K.C. (ed.), Humana Press, Totowa, NJ (2003) págs. 245-254, que describen la expresión de fragmentos de anticuerpos en *E. coli*. Después de la expresión, el anticuerpo puede aislarse del concentrado celular bacteriano en una fracción soluble y adicionalmente puede purificarse.

30 Además de los microbios procariotas, los microbios eucariotas, tales como hongos filamentosos o levaduras, son hospedadores adecuados para la clonación o expresión de hospedadores para vectores que codifican anticuerpos, incluyendo cepas de hongos y levaduras cuyas rutas de glucosilación se han "humanizado", dando como resultado la producción de un anticuerpo con un patrón de glucosilación parcial o completamente humano (véase, Gerngross, T.U., Nat. Biotech. 22 (2004) 1409-1414; y Li, H., et al., Nat. Biotech. 24 (2006) 210-215).

35 Las células hospedadoras adecuadas para la expresión de un anticuerpo glucosilado también proceden de organismos multicelulares (invertebrados y vertebrados). Los ejemplos de células de invertebrados incluyen células vegetales y de insectos. Se han identificado numerosas cepas de baculovirus que pueden usarse junto con células de insecto, particularmente para la transfección de células de *Spodoptera frugiperda*.

40 Como hospedadores también pueden utilizarse cultivos de células vegetales (véanse, por ejemplo, los documentos US 5,959,177, US 6,040,498, US 6,420,548, US 7,125,978, y US 6,417,429 (que describen la tecnología PLANTIBODIES™ para producir anticuerpos en plantas transgénicas).

45 Como hospedadores también pueden utilizarse células de vertebrados. Por ejemplo, pueden ser útiles líneas de células de mamífero que están adaptadas al crecimiento en suspensión. Otros ejemplos de líneas de células de mamífero útiles son la línea CV1 de riñón de mono transformada por SV40 (COS-7); la línea de riñón embrionario humano (células 293 como se describe, por ejemplo, en Graham, F.L., et al., J. Gen. Virol. 36 (1977) 59-74); las células de riñón de cría de hámster (BHK); las células de Sertoli de ratón (células TM4 como se describe, por ejemplo, en Mather, J.P., Biol. Reprod. 23 (1980) 243-252); las células de riñón de mono (CV1); las células de riñón de mono verde africano (VERO-76); las células de carcinoma de cuello uterino humano (HELA); las células de riñón canino (MDCK, riñón canino de Madin-Darby); las células de hígado de rata búfalo (BRL 3A); las células de pulmón humano (W138); las células de riñón humano (Hep G2); tumor mamario de ratón (MMT 060562), las células TRI, como se describe, por ejemplo, en Mather, J.P., et al., Annals N.Y. Acad. Sci. 383 (1982) 44-68; las células MRC 5 y las células FS4. Otras líneas de células de mamífero útiles incluyen células de ovario de hámster chino (CHO), incluyendo células CHO DHFR- (Urlaub, G., et al., Proc. Natl. Acad. Sci. USA 77 (1980) 4216-4220); y líneas de células de mieloma tales como Y0, NS0 y Sp2/0. Para una revisión de determinadas líneas de células de hospedadoras de mamífero adecuadas para la producción de anticuerpos, véase, por ejemplo, Yazaki, P. y Wu, A.M., Methods in Molecular Biology, Vol. 248, Lo, B.K.C. (ed.), Humana Press, Totowa, NJ (2004) págs. 255-268.

#### USO DE CODONES

50 Se puede acceder fácilmente a las tablas de uso de codones (para ejemplos, véanse las tablas anteriores), por ejemplo, a través de la "base de datos de uso de codones" disponible en la <http://www.kazusa.or.jp/codon/> y estas tablas pueden adaptarse de diversas maneras (Nakamura, Y., et al., Nucl. Acids Res. 28 (2000) 292).

Para la expresión de polipéptidos recombinantes a alto rendimiento, los ácidos nucleicos codificantes desempeñan una función importante. Generalmente, los ácidos nucleicos codificantes de origen natural y aislados de la naturaleza, no están optimizados para su expresión a alto rendimiento, especialmente si se expresan en una célula hospedadora heteróloga. Debido a la degeneración del código genético, un resto de aminoácido puede estar codificado por más de un triplete de nucleótidos (codones), a excepción de los aminoácidos triptófano y metionina. Por tanto, para una secuencia de aminoácidos son posibles diferentes codones codificantes (= secuencias de ácidos nucleicos codificantes correspondientes).

Los diferentes codones que codifican un resto de aminoácido son empleados por organismos diferentes con frecuencias relativas diferentes (uso de codones). Generalmente un codón específico se usa con mayor frecuencia que los otros codones posibles.

En el documento WO 2001/088141 se describe la optimización de una fase de lectura, de acuerdo con el uso de codones, encontrada en genes de mamífero expresados en altas cantidades. Para esa finalidad, se generó una matriz considerando casi exclusivamente aquellos codones que se usaban más frecuentemente y, menos preferentemente, aquellos que se usaban más frecuentemente como segundos codones en genes de mamífero expresados en grandes cantidades, como se representa en la siguiente tabla. Utilizando estos codones a partir de genes humanos expresados en grandes cantidades, se creó una fase de lectura completamente sintética que no era de origen natural, que, sin embargo codificaba el mismo producto que el de la construcción génica de tipo silvestre.

En el documento US 8.128.938 se describen diferentes métodos de optimización con codones usando la frecuencia de uso de codones individuales, tales como optimización uniforme, optimización completa y optimización mínima.

En la siguiente tabla se muestra el codón más frecuentemente usado (codón I) y el segundo codón más frecuentemente usado (codón II) encontrados en genes de mamífero expresados en grandes cantidades.

**Tabla.**

<b>aminoácido</b>	<b>codón 1</b>	<b>codón 2</b>
<i>Ala</i>	GCC	GCT
<i>Arg</i>	AGG	AGA
<i>Asn</i>	AAC	AAT
<i>Asp</i>	GAC	GAT
<i>Cys</i>	TGC	TGT
<i>End</i>	TGA	TAA
<i>Gln</i>	CAG	CAT
<i>Glu</i>	GAG	GAA
<i>Gly</i>	GGC	GGA
<i>His</i>	CAC	CAT
<i>Ile</i>	ATC	ATT
<i>Lcar</i>	CTG	CTC
<i>Lys</i>	AAG	AAT
<i>Met</i>	ATG	ATG
<i>Phe</i>	TTC	TTT
<i>Pro</i>	CCC	CCT
<i>Ser</i>	AGC	TCC
<i>Thr</i>	ACC	ACA
<i>Trp</i>	TGG	TGG
<i>Tyr</i>	TAC	TAT
<i>Val</i>	GTG	GTC

(Ausubel, F.M., et al., Current Protocols in Molecular Biology 2 (1994), A1.8-A1.9).

Pueden hacerse pocas desviaciones del estricto apego del uso de codones más frecuentemente encontrados (i) para dar cabida a la introducción de sitios de restricción únicos, (ii) para romper tramos G o C que se extienden más de 7 pares de bases para permitir la amplificación por PCR consecutiva y la secuenciación del producto génico sintético.

MÉTODO PARA LA PRODUCCIÓN DE UN POLIPÉPTIDO EXPRESANDO UN ÁCIDO NUCLEICO CODIFICANTE CON USO DE CODONES MODIFICADO

5 Se ha descubierto que para la expresión de un polipéptido en una célula, el uso de un ácido nucleico que codifica un polipéptido, en el que cada aminoácido está codificado por un grupo de codones, por lo cual cada codón en el grupo de codones se define por una frecuencia de uso específica en el grupo correspondiente a la frecuencia de uso total en el genoma de la célula, y por lo cual la frecuencia de uso de los codones en el ácido nucleico que codifica el polinucleótido (total) es aproximadamente la misma que la frecuencia de uso dentro del grupo, es beneficioso.

10 Un aspecto del presente documento es un método para la producción de un polipéptido que comprende la etapa de cultivar una célula, que comprende un ácido nucleico que codifica el polipéptido, y recuperar el polipéptido de la célula o del medio de cultivo, en el que cada uno de los restos de aminoácido del polipéptido está codificado por al menos un codón, por lo cual los diferentes codones que codifican el mismo resto de aminoácido están combinados en un grupo y cada uno de los codones en un grupo está definido por una frecuencia de uso específica dentro del grupo, por lo cual la suma de las frecuencias de uso específicas de todos los codones en un grupo es del 100 %, en el que la frecuencia de uso de un codón en el ácido nucleico que codifica el polinucleótido es aproximadamente la misma que su frecuencia de uso específico dentro de su grupo.

20 En el método como se indica en el presente documento, para expresar el polipéptido, se usa un ácido nucleico optimizado con codones que codifica un polipéptido, por lo cual el rendimiento de expresión obtenible está aumentado en comparación con otros ácidos nucleicos.

25 Se ha descubierto que usando un grupo de codones para codificar cada aminoácido en un polipéptido que está en la naturaleza también codificado por un grupo de codones y usando la misma proporción relativa entre los codones individuales de un grupo también dentro del ácido nucleico completo, puede proporcionarse un ácido nucleico que codifique un polipéptido que, después de la expresión, dé como resultado un rendimiento mejorado, por ejemplo, en comparación con un ácido nucleico en el que siempre está presente el codón con la frecuencia de uso más alta.

30 En una realización la célula es una célula procariota. En una realización la célula es una célula bacteriana.

En una realización preferida, la célula es una célula de E. coli. En la siguiente tabla se proporciona la frecuencia del uso de codones total, teniendo en cuenta todos los codones que codifican un resto de aminoácido específico para E. coli.

35

Tabla.									
Ala	GCG	32	Gly	GGG	4	Pro	CCG	75	
Ala	GCA	24	Gly	GGA	2	Pro	CCA	14	
Ala	GCT	28	Gly	GGT	51	Pro	CCT	10	
Ala	GCC	16	Gly	GGC	43	Pro	CCC	1	
Arg	AGG	0	His	CAT	29	Ser	AGT	4	
Arg	AGA	0	His	CAC	71	Ser	AGC	25	
Arg	CGG	1				Ser	TCG	7	
Arg	CGA	1	He	ATA	0	Ser	TCA	5	
Arg	CGT	65	He	ATT	32	Ser	TCT	32	
Arg	CGC	33	He	ATC	68	Ser	TCC	27	
Asn	AAT	16	Leu	CTG	79	Thr	ACG	12	
Asn	AAC	84	Leu	CTA	1	Thr	ACA	4	
			Leu	CTT	5	Thr	ACT	28	
Asp	GAT	46	Leu	CTC	8	Thr	ACC	56	
Asp	GAC	54	Leu	TTG	5				
			Leu	TTA	3	Trp	TGG	100	
Cys	TGT	36							
Cys	TGC	64	Lys	AAG	20	Tyr	TAT	36	
			Lys	AAA	80	Tyr	TAC	64	
Gln	CAG	82							
Gln	CAA	18	Met	ATG	100	Val	GTG	27	
						Val	GTA	20	
Glu	GAG	24	Phe	TTT	28	Val	GTT	40	
Glu	GAA	76	Phe	TTC	72	Val	GTC	13	

40 Para codificar el resto de aminoácido alanina se dispone de cuatro codones diferentes: GCG, GCA, GCT, y GCC. Por tanto, el grupo de codones que codifica el resto de aminoácido puede comprender a lo sumo cuatro codones. En el grupo que comprende cuatro codones, cada codón tiene una frecuencia de uso específica en el grupo que es la misma que la frecuencia de uso total en el genoma de la célula, es decir, el codón GCG tiene una frecuencia de uso

específica y total de 30 %, el codón GCA tiene una frecuencia de uso específica y total de 24 %, el codón GCT tiene una frecuencia de uso específica y total de 28 % y el codón GCG tiene una frecuencia de uso específica y total de 16 %. Si en el grupo se reduce el número de codones, por ejemplo, excluyendo del grupo los codones GCA y GCC con una frecuencia de uso específica de 24 % y de 16 %, respectivamente, la frecuencia de uso específica de los miembros restantes del grupo, es decir, GCG y GCT, cambia a 53 % ( $=32/(32+28) * 100$ ) y a 47 % ( $=28/(32+28) * 100$ ), respectivamente, ya que la suma de las frecuencias de uso específicas de todos los codones en un grupo es de 100 %. Por tanto, en el grupo que comprende los dos codones GCG y GCT, cada codón tiene una frecuencia de uso específica en el grupo que es más alta que su frecuencia de uso total en el genoma de la célula, es decir, el codón GCG tiene una frecuencia de uso específica de 53 % y una frecuencia de uso total de 32 %, y el codón GCT tiene una frecuencia de uso específica de 47 % y una frecuencia de uso total de 28 %.

Si el grupo de codones que codifica el resto de aminoácido alanina comprende los cuatro codones disponibles, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina comprende los codones GCG, GCT, GCA y GCC a una proporción de 32:28:24:16, que corresponde a 8:7:6:4. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 25 posiciones, esto se ajusta a 8:8:6:4, que corresponde a 4:4:3:2.

El uso de los diferentes codones que codifican un aminoácido específico es alternativo dentro del genoma. Esta alternancia se refleja en el presente documento por la definición de un motivo codónico de aminoácidos. Dentro de un motivo codónico de aminoácidos, los codones individuales se distribuyen teniendo en cuenta la frecuencia de uso específica, por lo cual primero se seleccionan los codones con una frecuencia más alta. El motivo de la secuencia de aminoácidos comprende una secuencia específica de codones, en la que el número total de codones en un motivo codónico de aminoácidos es al menos el mismo o incluso más alto que el número de codones en un grupo de codones, para permitir un mapeo de la frecuencia de uso de un grupo de codones con respecto al motivo codónico de aminoácidos correspondiente.

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcg gcg gcg gct gct gct gca gca gca gcc gcc (SEQ ID NO: 01).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcg gct gca gcc gcg gct gca gcg gct gcc gcg gct gca (SEQ ID NO: 02).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gct (SEQ ID NO: 03).

Si el grupo de codones que codifica el resto de aminoácido arginina comprende los dos codones GCG y GCT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina comprende los codones GCG y GCT a una proporción de 53:47.

Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones, esto se ajusta a 50:50, que corresponde a 1:1.

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcg gct (SEQ ID NO: 04).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gct gcg (SEQ ID NO: 05).

Si el grupo de codones que codifica el resto de aminoácido arginina comprende los dos codones GCT y CGC, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina comprende los codones CGT y CGC a una proporción de 66:34, que corresponde a 33:17. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 51 posiciones, esto se ajusta a 66:33, que corresponde a 2:1.

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina es cgt cgt gcg (SEQ ID NO: 06).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina es cgt cgc cgt (SEQ ID NO: 07).

Si el grupo de codones que codifica el resto de aminoácido asparragina comprende los dos codones AAC y AAT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido asparraguina comprende los codones AAC y AAT a una proporción de 84:16, que corresponde a 21:4. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 25 posiciones esto se ajusta a 20:4, que corresponde a 5:1.

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido asparraguina es aac aac aac aac aac aat (SEQ ID NO: 08).





## ES 2 599 386 T3

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido glutámico es gaa gag gaa gaa (SEQ ID NO: 24).

5 Si el grupo de codones que codifica el resto de aminoácido de glicina comprende los dos codones GGT y GGC, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina comprende los codones GGT y GGC a una proporción de 54:46, que corresponde a 27:23. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones, esto se ajusta a 30:24, teniendo en cuenta que la glicina puede estar codificada por cuatro codones diferentes, que corresponde a 5:4.

10 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina es ggt ggt ggt ggt ggc ggc ggc ggc (SEQ ID NO: 25).

15 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina es ggt ggc ggt ggc ggt ggc ggt (SEQ ID NO: 26).

20 Si el grupo de codones que codifica el resto de aminoácido histidina comprende los dos codones CAC y CAT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido histidina comprende los codones CAC y CAT a una proporción de 71:29. Como esto daría como resultado un motivo codónico de aminoácido que comprende 100 posiciones, esto se ajusta a 66:33, que corresponde a 2:1.

25 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido histidina es cac cac cat (SEQ ID NO: 27).

30 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido histidina es cac cat cac (SEQ ID NO: 28).

35 Si el grupo de codones que codifica el resto de aminoácido isoleucina comprende los dos codones ATC y ATT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina comprende los codones ATC y ATT a una proporción de 68:32, que corresponde a 17:8. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 25 posiciones, esto se ajusta a 16:8, que corresponde a 2:1.

40 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina es atc atc att (SEQ ID NO: 29).

45 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina es atc att atc (SEQ ID NO: 30).

50 Si el grupo de codones que codifica el resto de aminoácido leucina comprende los dos codones CTG y CTC el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina comprende los codones CTG y CTC a una proporción de 91:9. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones, esto se ajusta a 90:10, que corresponde a 9:1.

55 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctg ctg ctg ctg ctg ctg ctg ctc (SEQ ID NO: 31).

60 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctg ctg ctg ctc ctg ctg ctg (SEQ ID NO: 32).

65 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctg ctg ctc ctg ctg ctg (SEQ ID NO: 33).

70 Si el grupo de codones que codifica el resto de aminoácido lisina comprende los dos codones AAA y AAG el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina comprende los codones AAA y AAG a una proporción de 80:20, que corresponde a 4:1.

75 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aaa aaa aaa aaa aag (SEQ ID NO: 34).

80 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aaa aaa aaa aag aaa (SEQ ID NO: 35).

85 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aaa aaa aag aaa aaa (SEQ ID NO: 36).

90 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aaa aag aaa aaa aaa (SEQ ID NO: 37).

## ES 2 599 386 T3

- Si el grupo de codones que codifica el resto de aminoácido fenilalanina comprende los dos codones TTC y TTT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina comprende los codones TTC y TTT a una proporción de 72:27, que corresponde a 18:7. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 25 posiciones, esto se ajusta a 18:6, que corresponde a 3:1.
- 5 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina es ttc ttc ttc ttt (SEQ ID NO: 38).
- 10 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina es ttc ttc ttt ttc (SEQ ID NO: 39).
- 15 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina es ttc ttt ttc ttc (SEQ ID NO: 40).
- 20 Si el grupo de codones que codifica el resto de aminoácido prolina comprende los tres codones CCG, CCA y CCT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina comprende los codones CCG, CCA y CCT a una proporción de 76:14:10, que corresponde a 38:7:5. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones, esto se ajusta a 35:7:7, que corresponde a 5:1:1.
- 25 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg ccg ccg ccg ccg cca cct (SEQ ID NO: 41).
- 30 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg ccg ccg cca ccg cct ccg (SEQ ID NO: 42).
- 35 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg ccg cca ccg ccg cct ccg (SEQ ID NO: 43).
- 40 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg ccg cca ccg ccg cct ccg ccg (SEQ ID NO: 44).
- 45 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg cca ccg ccg cct ccg ccg (SEQ ID NO: 45).
- 50 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccg cca ccg cct ccg ccg ccg (SEQ ID NO: 46).
- 55 Si el grupo de codones que codifica el resto de aminoácido prolina comprende los tres codones CTC, TTC y AGC, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina comprende los codones CTC, TTC y AGC a una proporción de 38:32:30, que corresponde a 19:16:15. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones, esto se ajusta a 18:15:15, que corresponde a 6:5:5.
- 60 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina es tct tct tct tct tcc tcc tcc tcc tcc agc agc agc agc (SEQ ID NO: 47).
- 65 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina es tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct (SEQ ID NO: 48).
- 70 Si el grupo de codones que codifica el resto de aminoácido treonina comprende los tres codones ACC, ACT y ACG, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina comprende los codones ACC, ACT y ACG a una proporción de 58:29:13. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones, esto se ajusta a 56:28:14, que corresponde a 4:2:1.
- 75 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc acc acc acc act act acg (SEQ ID NO: 49).
- 80 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc act acc act acc acg acc (SEQ ID NO: 50).
- 85 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc act acc acg acc act acc (SEQ ID NO: 51).
- 90 Si el grupo de codones que codifica el resto de aminoácido tirosina comprende los dos codones CAC y CAT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina comprende los codones CAC y CAT a una proporción de 64:34, que corresponde a 32:17. Como esto daría como resultado un

motivo codónico de aminoácidos que comprende 49 posiciones, esto se ajusta a 32:16, que corresponde a 2:1.

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina es tac tac tat (SEQ ID NO: 52).

5 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina es tac tat tac (SEQ ID NO: 53).

10 Si el grupo de codones que codifica el resto de aminoácido valina comprende los dos codones GTT, GTG, GTA y GTC, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina comprende los codones GTT, GTG, GTA y GTC a una proporción de 40:27:20:13. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones, esto se ajusta a 40:30:20:10, que corresponde a 4:3:2:1.

15 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtt gtt gtt gtt gtg gtg gta gta gtc (SEQ ID NO: 54).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtt gtg gta gtc gtt gtg gta gtt gtg gtt (SEQ ID NO: 55).

20 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtt gtg gta gtt gtg gtt gtc gtt gtg gta (SEQ ID NO: 56).

A continuación, el método como se indica en el presente documento se ilustra con un polipéptido que tiene la siguiente secuencia de aminoácidos:

DSAVDSQGTS	FSEYVGAFFVS	VDAGHCAAES
QASVSSVYNL	AVPAYRASVY	RSDTSDIDTA
AVSSPDVVDI	IERVKSYSRG	SVTAAYAIGV
RYDWSRSHSG	SETSTSNFAY	TYSLNSTQTF
VYASKARSAL	AAVVGVRES	ITGSSGQVFF
AATSTASDAH	ASTGADIDPT	AVVHTDVSVV
ISAFAVAAGH	VARVHHVIAS	IDYAVDAGAA
GAAGSSGGTR	IAGVVVSVTI	RGFSLTGLGA
GDVGPHTARY	AGESFSVDCS	RGASHVASSA
KPASVTDMTP	YRSVTDDASD	DGPASVSDGY

(SEQ ID NO: 57).

25  
30 Para facilitar la purificación, una etiqueta de purificación puede unirse al extremo N o C del polipéptido. La etiqueta de purificación puede fusionarse directamente con la secuencia de aminoácidos del polipéptido o puede estar separada de la secuencia de aminoácidos mediante un enlazador corto o mediante un sitio de escisión de proteasas. Un ejemplo de etiqueta de purificación es la etiqueta de seis histidinas con un enlazador GS N terminal para la fusión con el extremo C del polipéptido:

GSHHHHHH  
(SEQ ID NO: 58),

35 que está codificada por la siguiente secuencia de ácidos nucleicos:

ggttctcaccaccaccaccac  
(SEQ ID NO: 59).

40 El polipéptido de ensayo puede ir precedido por un péptido transportador corto. Un ejemplo de péptido transportador procede de la parte N terminal del interferón alfa humano maduro:

MCDLPQTHSL GS  
(SEQ ID NO: 60),

45 que está codificado por la siguiente secuencia de ácidos nucleicos:

atgtgcgacctgccgagaccactcccttgatcc  
(SEQ ID NO: 61).

50 La secuencia de ácidos nucleicos que codifica el polipéptido de ensayo de SEQ ID NO: 57, que se obtiene con un método de retrotraducción usando siempre el codón con la frecuencia de uso más alta en la célula respectiva, es

gactctgCGGTTgactctcagggTaccTcttctctgaatacgttggtgCGTTCg  
ttctctgTTgacgCGGgtcacaagCGGCGgaatctcagGCGTctgTTctctctgT  
ttacaacctggCGGttccGGCGTaccgTgCGTcttacgTTCgTtctgacacctct  
gacatCGacaccCGGCGGttctctctccGGacgTtGttgacatcatCGaacgTg  
ttaaactcttactctcgtGGttctgTtaccCGGCGTaccGgatCGgtgTtcgTta  
CGactGGTctcgtTctcactctGGTctgaaacctctacctetaacttcGCGTAC  
acctactctctgaaactctaccCagaccttcgTttacGCGTctaaagCGGttctg  
CGctGGCGGCGGttgTtGttGGTgTtcgTgaatctatcaccGGTctctctGGTca  
GGtttctctcGCGGCGacctctaccGCGTctgacGCGcaccGCGTctaccGGTgCG  
gacatCGaccCGaccCGGttGttcacaccGacgTtctctgTtGttatctctGCGT  
tcGCGGttGCGGCGcaccGGTgTtGCGCGTgTtcaccacgTtatcGCGTctatCGa  
ctacGCGGttgacGCGGgtGCGGCGGgtGCGGCGGgttctctctGGTGGTaccCGT  
atcGCGGgtGttGttGttctGttaccatccGtGGtttctctctgaccGGTctGG  
gtGCGGgtGacgTtGGTccGcaccCGCGGttacGCGGgtgaatcttctctctgT  
tgactgctctcgtGGTgCGTctcaccgTtGCGTcttctgCGaaaccGGCGTctgTt  
  
accgacatgacCCCGTaccgTtctgTtaccGacgacGCGTctgacgacGGTccGG  
CGTctgTtctgacGGttac (SEQ ID NO: 62).

5 La secuencia de ácidos nucleicos que codifica el polipéptido de ensayo de SEQ ID NO: 57, que se obtiene con un método como se indica en el presente documento, es

gactctgCGGTTgattcccagggTaccagcttctctgaatacgtgggCGcttctcG  
tatccgTtgacGcaggtcacaagccGCGGaaagccagGcttctgTgtccagCGT  
ttataacctggcagTcccGGcctaccgTgCGTcttacgTTCgTccgatactagc  
gacatCGataccGctgCagTgTcttccccGGacgtagTtgatattatcGagCGTg  
tgaaaagctattctcgtGGctctgTaaCGGCGGCGTaccGtatcGGTgTtcGctA  
CGactGGTcccGtagccattctGGctccGaaaccagcacttctaactttgcatat  
acctactccctgaacagcaccCAAacttctcgtgTaccGctctaaGGCGCGTtccG  
ctctGGcagCGTtGtCGTtGGTgTgCGCGaaagcattaccGGctcttccGGTca  
GGtattctctcGCGGctacgagcaccGcatctgatgCGcaccGCGTctactGGTgct  
gacatCGatccaaccGcagTtGtGcaccCGacgTatccgTtGtgatcagCGcct  
ttGCGGttGctGcaccatGGCGTgccccGtGttcaccacgTgattGCGTctatCGa  
ttatgctgtagacGcaggtGCGGCGGCGGctGcaggttccagCGGCGGtactCGT  
atcGccGGCGTtGtGGtatctgTtaccattcGCGGttctccctgacGGGtctcG  
gCGCGGgtgatGtGGGccccGcataccGctcgttacGcaggtgaaagcttctctgT  
tgactgctcccGtGGCGCCagCCacgTcGCGTcttccGctaaaccGGcAagCGTt  
actgatatgacCCcttaccGctctgTgaccGacgatGCGTctgacgatGGTccGG  
CGTccGtaagCGacGGctat (SEQ ID NO: 63).

10 Comparando la secuencia de ácidos nucleicos que se ha obtenido usando siempre el codón con la frecuencia de uso más alta con el ácido nucleico que se ha obtenido con un método como se indica en el presente documento puede observarse que los dos ácidos nucleicos difieren en 146 codones de los 300 codones, es decir, las secuencias optimizadas difieren en 48,7 % de todos los codones codificantes (los codones que difieren aparecen subrayados en la siguiente alineación).

ES 2 599 386 T3

gactctgcggttgactctcagggtaacctctttctctgaatacgttggtgcgttc  
gactctgcggttgattcccagggtaccagctttctctgaatacgtggcgctttc  
D S A V D S Q G T S F S E Y V G A F

gtttctgttgacgcgggtcacaaagcggcggaatctcaggcgtctgtttcttct  
gtatccgttgacgcaggtcacaaagccgcggaaagccaggcttctgtgtccagc  
V S V D A G H K A A E S Q A S V S S

gtttacaacctggcgggttccggcgtaccgtgcgtcttacgttcgttctgacacc  
gtttataacctggcagtcccggcctaccgtgcgtcttacgttcgctccgatact  
V Y N L A V P A Y R A S Y V R S D T

tctgacatcgacaccgcggcgggtttcttctccggacgttgttgacatcatcgaa  
agcgacatcgataccgctgcagtgtcttccccggacgtagttgatattatcgag  
S D I D T A A V S S P D V V D I I E

ES 2 599 386 T3

cggttaaatcttactctcgtggttctgttaccgcggggtacgcgatcggtgt  
R V K S Y S R G S V T A A Y A I G V

cgtacgactggtctcgttcactctgaaaccttaccttaacttc  
cgtacgactggtcccgtagccattctggctccgaaaccagcactttaacttt  
R Y D W S R S H S G S E T S T S N F

gcgtacactactctgaacttaccgaccttcgttacgcgtctaagcg  
gcatatacctaccctgaacagcaccaaactttcgtgtacgcctctaaggcg  
A Y T Y S L N S T Q T F V Y A S K A

cgttctgcgctggcggcgttgttgtgttcgtgaatcatcaccgttct  
cgtccgctctggcagcgttgtcgtgtgtgcgcgaagcattaccggctct  
R S A L A A V V V G V R E S I T G S

tctggtcaggtttcttcgcggcacctctaccgctctgacgcgcacgctct  
tccggtcaggtattcttcgcggtacgaccgcatctgatgcgcacgctct  
S G Q V F F A A T S T A S D A H A S

accggtgcggacatcgacccgaccgcggtgttcacaccgacgtttctgttgt  
actggtgctgacatcgatccaaccgcagttgtgcacaccgacgtatccgttgg  
T G A D I D P T A V V H T D V S V V

atctctgcgttcgcggtgcggcacggtgttgcgcgttcaccacgttatc  
atcagcgcctttgcggtgctgcacatggcgtcgcccgtgttcaccacgtgat  
I S A F A V A A H G V A R V H H V I

gcgtctatcgactacgcggtgacgcggtgcggcggtgcggcggttcttct  
gcgtctatcgattgctgtagacgcaggtgcggcggcgctgcaggtccagc  
A S I D Y A V D A G A A G A A G S S

ggtggtaccgtatcgcggtgttgttctgttaccatccgtggttctct  
ggcggtactcgtatcgccggcgttgtggtatctgttaccattcggttccc  
G G T R I A G V V V S V T I R G F S

ctgaccggtctgggtgcggtgacgttggtccgcacaccgcgcgttacgcggt  
ctgacggtctggcgcggtgatgtgggcccgcataccgctcgttacgcagt  
L T G L G A G D V G P H T A R Y A G

gaatctttctgttgactgctctcgtggtgcgtcacgttgcgtcttcgc  
gaagcttctgttgactgctccgtggcgccagccacgtcgcgtctccgct  
E S F S V D C S R G A S H V A S S A

aaaccggcgtctgttaccgacatgaccccgtaccgttctgttaccgacgacgcg  
aaaccggaagcgttactgatgacccctaccgctctgtgaccgacgatgcg  
K P A S V T D M T P Y R S V T D D A

tctgacgacggtccggctctgtttctgacggtac

tctgacgatgggtccggcgtccgtaagcgcgacggctat  
 S D D G P A S V S D G Y

(fila superior: SEQ ID NO: 62; fila central: SEQ ID NO: 63; fila inferior: SEQ ID NO: 57).

- 5 En la siguiente tabla se proporcionan los codones de los ácidos nucleicos codificantes respectivos en secuencias de su aparición comenzando desde el extremo 5' del ácido nucleico respectivo.

**Tabla**

Resto de aminoácido	SEQ ID NO: 62	SEQ ID NO: 63
G	ggg ggt ggt ggt ggt ggt ggg ggt ggt ggt ggt ggt ggg ggt ggt ggt ggt ggt ggg ggt ggt ggt ggt ggt ggg ggt	ggg ggc ggt ggc ggt ggc ggg ggc ggt ggt ggc ggt ggc ggt ggc ggt ggc ggt ggg ggc ggt ggc ggt ggc ggg ggc
A	gcg gcg	gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg
V	ggt gtt gtt gtt gtt gtt gtt ggt gtt gtt gtt gtt gtt gtt ggt gtt gtt gtt gtt gtt gtt ggt gtt gtt gtt gtt gtt gtt ggt gtt gtt gtt gtt gtt gtt ggt gtt gtt gtt gtt gtt gtt	ggt gtt gta gtt gtt gtt gtc ggt gtt gta gtt gtt gta gtt gtg gtt gtc gtt gtt gta gtt gtg gta gtt gtt gtt gtc gtt gtg gta gtt gtt gta gtt gtt ggt gtc gtt gtt gta
L	ctg ctg ctg ctg ctg	ctg ctg ctg ctg ctg
I	atc atc atc atc atc atc atc atc atc atc	atc att atc atc att atc atc att atc atc att
M	atg	atg
P	ccg ccg ccg ccg ccg ccg	ccg ccg cca ccg ccg cct ccg
F	ttc ttc ttc ttc ttc ttc ttc	ttc ttc ttt ttc ttc ttt ttc
W	tgg	tgg
S	tct tct	tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tct tcc agc tct tcc agc tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc tct tcc agc tct tcc agc
T	acc acc	acc act acc acg acc act acc acc act acc acg acc act acc acc act acc acg acc act acc acc
N	aac aac aac	aac aac aac
Q	cag cag cag cag	cag cag caa cag
Y	tac tac tac tac tac tac tac tac tac tac tac tac	tac tat tac tac tat tac tac tac tac tat tac tac tat
C	tgc	tgc
K	aaa aaa aaa aaa	aaa aaa aag aaa
R	cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt	cgt cgc cgt cgt cgc cgt cgc cgt cgt cgc cgt cgc
H	cac cac cac cac cac cac cac cac cac	cac cat cac cac cat cac cac cat cac

ES 2 599 386 T3

Resto de aminoácido	SEQ ID NO: 62	SEQ ID NO: 63
D	gac gac	gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat
E	gaa gaa gaa gaa gaa gaa	gaa gaa gag gaa gaa gaa

En la siguiente tabla se proporciona la frecuencia de uso de codones en los motivos codónicos de aminoácidos con respecto a la frecuencia de uso total en las células.

**Tabla**

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo	Frecuencia relativa de codones usados en el grupo
G	ggt (único codón usado)	1/1 = 100 %	100 %	ggt ggc ggt ggc ggt ggc ggt ggc ggt	ggt: 5/9 = 56 % ggc: 4/9 = 44 %	2	ggt: 54 % ggc: 46 %
A	gcg (único codón usado)	1/1 = 100 %	100 %	gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg	gcg: 4/12 = 33 % gct: 3/12 = 25 % gca: 3/12 = 25 % gcc: 2/12 = 17 %	4	gcg: 32 % gct: 28 % gca: 24 % gcc: 16 %
V	gtt (único codón usado)	1/1 = 100 %	100 %	gtt gtg gta gtt gtg gtt gtc gtt gtg gta	gtt: 4/10 = 40 % gtg: 3/10 = 30 % gta: 2/10 = 20 % gtc: 1/10 = 10 %	4	gtt: 40 % gtg: 27 % gta: 20 % gtc: 12 %
L	ctg (único codón usado)	1/1 = 100 %	100 %	ctg ctg ctg ctc	ctg: 4/5 = 80 % ctc: 1/5 = 20 %	2	ctg: 91 % ctc: 9 %
I	atc (único codón usado)	1/1 = 100 %	100 %	atc att atc	atc: 2/3 = 67 % att: 1/3 = 33 %	2	atc: 68 % att: 32 %
M	atg (único codón usado)	1/1 = 100 %	100 %	atg	1/1 = 100 %	1	100 %



ES 2 599 386 T3

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo	Frecuencia relativa de codones usados en el grupo
P	ccg (único codón usado)	1/1 = 100 %	100 %	ccg ccg cca ccg ccg cct	ccg: 4/6 = 67 % cca: 1/6 = 16,5 % cct: 1/6 = 16,5 %	3	ccg: 76 % cca: 14 % cct: 10 %
F	ttc (únicamente codón usado)	1/1 = 100 %	100 %	ttc ttc ttt ttc	ttc: 3/4 = 75 % ttt: 1/4 = 25 %	2	ttc: 72 % ttt: 28 %
W	tgg (único codón usado)	1/1 = 100 %	100 %	tgg	1/1 = 100 %	1	100 %
S	tct (único codón usado)	1/1 = 100 %	100 %	tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct	tct: 6/16 = 38 % tcc: 5/16 = 31 % agc: 5/16 = 31 %	3	tct: 38 % tcc: 32 % agc: 30 %
T	acc (único codón usado)	1/1 = 100 %	100 %	acc act acc acg acc act acc	acc: 4/7 = 57 % act: 2/7 = 29 % acg: 1/7 = 14 %	3	acc: 58 % act: 29 % acg: 13 %
N	aac (único codón usado)	1/1 = 100 %	100 %	aac aac aac	1/1 = 100 %	2	aac: 84 % aat: 16 %
Q	cag (único codón usado)	1/1 = 100 %	100 %	cag cag caa	cag: 2/3 = 67 % caa: 1/3 = 33 %	2	cag: 82 % caa: 18 %
Y	tac (único codón usado)	1/1 = 100 %	100 %	tac tat tac	tac: 2/3 = 67 % tat: 1/3 = 33 %	2	tac: 64 % tat: 36 %
C	tgc (único codón usado)	1/1 = 100 %	100 %	tgc	1/1 = 100 %	2	tgc: 64 % tgt: 36 %

ES 2 599 386 T3

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo	Frecuencia relativa de codones usados en el grupo
K	aaa (único codón usado)	1/1 = 100 %	100 %	aaa cLcLcL aag	aaa 2/3 = 67 % aag: 1/3 = 33 %	2	80 % aag: 20 %
R	cgt (único codón usado)	1/1 = 100 %	100 %	cgt cgc cgt	cgt: 2/3 = 67 % cgc: 1/3 = 33 %	2	cgt: 66 % cgc: 34 %
H	cac (único codón usado)	1/1 = 100 %	100 %	cac cat cac	cac: 2/3 = 67 % cat: 1/3 = 33 %	2	cac: 71 % cat: 29 %
D	gac (único codón usado)	1/1 = 100 %	100 %	gac gat	gac: 1/2 = 50 % gat: 1/2 = 50 %	2	gac: 54 % gat: 46 %
E	gaa (único codón usado)	1/1 = 100 %	100 %	gaa gaa gag gaa	gaa: 3/4 = 75 % gag: 1/4 = 25 %	2	gaa: 76 % gag: 24 %

En la siguiente tabla se da la frecuencia de los codones en todo el ácido nucleico codificante.

**Tabla.**

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo
G	ggt ggt	26/26 = 100 %	100 %	ggt ggc ggt ggc ggt ggc ggt ggc ggt ggt ggc ggt ggc ggt ggc ggt ggc ggt ggt ggc ggt ggc ggt ggc ggt ggc	ggt: 14/26 = 54 % ggc: 12/26 = 46 %	ggt: 54 % ggc: 46 %
A	gcg gcg	49/49 = 100 %	100 %	gcg gct gca gcc gcg gct gca gcc gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcg gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gcg gct gca gcc gfg	gcg: 17/49 35 % gct: 12/49 = 24,5 % gca: 12/49 = 24,5 % gcc: 8/49 = 16 %	gcg: 32 % gct: 28 % gca: 24 % gcc: 16 %

ES 2 599 386 T3

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo
V	gtt gtt	32/32 = 100 %	100 %	gtt gtg gta gtt gtg gtt gtc gtt gtg gta gtt gtt gta gtt gtt gtt gtc gtt gtt gta gtt gtt gta gtt gtg gtt gtc gtt gtg gta gtt gtt gta gtt gtt gtt gtc gtt gtt gta	gtt: 16/40 = 40 %  gtg: 12/40 = 30 %  gta: 8/40 = 20 %  gtc: 4/40 = 10 %	gtt: 40 %  gtg: 27 %  gta: 20 %  gtc: 12 %
L	ctg ctg ctg ctg ctg	5/5 = 100 %	100 %	ctg ctg ctg ctg ctc	ctg: 4/5 = 80 % ctc: 1/5 = 20 %	ctg: 91 % ctc: 9 %
I	atc atc atc atc atc atc atc atc atc atc atc	11/11 = 100 %	100 %	atc att atc atc att atc atc att atc atc att	atc: 7/11 = 64 % att: 4/11 = 36 %	atc: 68 % att: 32 %
M	atg	1/1 = 100 %	100 %	atg	1/1 = 100 %	100 %
P	ccg ccg ccg ccg ccg ccg ccg	7/7 = 100 %	100 %	ccg ccg cca ccg ccg cct ccg	ccg: 5/7 = 72 % cca: 1/7 = 14 % cct: 1/7 = 14 %	ccg: 76 % cca: 14 % cct: 10 %
F	ttc ttc ttc ttc ttc ttc ttc ttc ttc	9/9 = 100 %	100 %	ttc ttc ttt ttc ttc ttc ttt ttc ttc	ttc: 7/9 = 78 % ttt: 2/9 = 22 %	ttc: 72 % ttt: 28 %
W	tgg	1/1 = 100 %	100 %	tgg	1/1 = 100 %	100 %

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo
S	tct tct	51/51 = 100 %	100 %	tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc agc tct tcc tct tcc agc tct tct tcc agc tct tct tcc agc tct	tct: 19/51 = 38 % tcc: 16/51 = 31 % agc: 16/51 = 31 %	tct: 38 % tcc: 32 % agc: 30 %
T	acc acc	22/22 = 100 %	100 %	acc act acc acg acc act acc acc act acc acg acc act acc acc act acc acg acc act acc acc	acc: 13/22 = 59 % act: 6/22 = 27 % acg: 3/22 = 14 %	acc: 58 % act: 29 % acg: 13 %
N	aac aac aac	3/3 = 100 %	100 %	aac aac aac	3/3 = 100 %	100 %
Q	cag cag cag cag	4/4 = 100 %	100 %	cag cag caa cag	cag: 3/4 = 75 % caa: 1/4 = 25 %	cag: 82 % caa: 18 %
Y	tac tac tac tac tac tac tac tac tac tac tac tac tac tac	14/14 = 100 %	100 %	tac tat tac tac tat tac tac tat tac tac tat tac tac tat	tac: 9/14 = 64 % tat: 5/14 = 36 %	tac: 64 % tat: 36 %
C	tgc	1/1 = 100 %	100 %	tgc	1/1 = 100 %	100 %
K	aaa aaa aaa aaa	4/4 = 100 %	100 %	aaa aaa aag aaa	aaa 3/4 = 75 % aag: 1/4 = 25 %	aaa 80 % aag: 20 %
R	cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt	14/14 = 100 %	100 %	cgt cgc cgt cgt cgc cgt cgt cgc cgt cgt cgc cgt cgtcgc	cgt: 9/14 = 64 % cgc: 5/14 = 36 %	cgt: 66 % cgc: 34 %
H	cac cac cac cac cac cac cac cac cac	9/9 = 100 %	100 %	cac cat cac cac cat cac cac cat cac	cac: 6/9 = 67 % cat: 3/9 = 33 %	cac: 71 % cat: 29 %

Resto de aminoácido	Secuencia de codones (motivo) en SEQ ID NO: 62	Frecuencia relativa de codones usados en motivos	Frecuencia relativa de codones usados en el grupo para restos de aminoácidos específicos	Secuencia de codones (motivo) en SEQ ID NO: 63	Frecuencia relativa de codones usados en motivos	N.º de codones en el grupo
D	gac gac	23/23 = 100 %	100 %	gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat gac gat	gac: 12/23 = 52 % gat: 11/23 = 48 %	gac: 54 % gat: 46 %
E	gaa gaa gaa gaa gaa gaa	6/6 = 100 %	100 %	gaa gaa gag gaa gaa gaa	gaa: 5/6 = 83 % gag: 1/6 = 17 %	gaa: 76 % gag: 24 %

5 Ambos ácidos nucleicos se han expresado en *E. coli* que comprende un péptido transportador de interferón alfa en el extremo N terminal de SEQ ID NO: 60 codificado por un ácido nucleico de SEQ ID NO: 61 y una etiqueta de purificación C terminal de SEQ ID NO: 58 codificada por un ácido nucleico de SEQ ID NO: 59. El vector de expresión de *E. coli*, incluyendo el casete de expresión, era idéntico para el polipéptido de ensayo codificado de manera diferente. El rendimiento de la expresión del polipéptido de ensayo codificado de manera diferente se ha determinado por análisis de transferencia de Western cuantitativo. Por lo tanto, se prepararon lisados de células enteras de *E. coli* y se fraccionaron en un sobrenadante soluble y en una fracción de sedimento celular insoluble por centrifugación. Después de esto, las proteínas se separaron electroforéticamente mediante SDS PAGE, se transfirieron electroforéticamente a una membrana de nitrocelulosa y después se tiñeron con un anticuerpo conjugado con POD que reconoce la etiqueta de purificación de poli-His. La poli-His teñida contenía el polipéptido de ensayo codificado de forma diferente se cuantificó por comparación con un patrón de referencia de proteína pura que contenía la misma etiqueta de purificación poli-His (fragmento de anticuerpo scFv-poli-His) de concentración proteica scFv-poli-His conocida usando el analizador Lumi-Imager F1 (Roche Molecular Biochemicals) y el programa informático Lumi Analyst, versión 3.1.

20 En las Figuras 1 y 2 se muestran las transferencias de Western del polipéptido de ensayo marcado con poli-His codificado de manera diferente. Los lisados de células enteras de *E. coli* se fraccionaron en un sobrenadante soluble y en una fracción de sedimento celular insoluble antes de realizar SDS PAGE e inmunotransferencia. Como referencia se ha usado un patrón de proteína de peso molecular y un fragmento de anticuerpo scFv que comprendía una etiqueta de poli-His de concentración proteica conocida.

25 Los carriles 2, 4, 6, y 8 son muestras que muestran la cantidad de polipéptido de ensayo expresado obtenido con el ácido nucleico generado usando siempre el codón con la frecuencia de uso más alta después de 0 horas, 4 horas, 6 horas y 21 horas de cultivo.

30 Los carriles 3, 5, 7 y 9 son muestras que muestran la cantidad de polipéptido de ensayo expresado obtenido con el ácido nucleico generado con el nuevo método de retrotraducción de proteínas de la invención como se indica en el presente documento, después de 0 horas, 4 horas, 6 horas y 21 horas de cultivo.

Los carriles 11 a 14 corresponden al patrón de referencia de la proteína scFv-poli-His purificada de concentración conocida (5 ng, 10 ng, 20 ng, 30 ng y 40 ng).

35 La cantidad de polipéptido de ensayo expresada se determinó en la fracción de residuo celular insoluble (sedimento) después de la solubilización/extracción de agregados proteicos insolubles con tampón de muestra SDS ya que la fracción principal del polipéptido de ensayo expresado se encontró en la fracción de sedimento celular insoluble después de la lisis celular y del fraccionamiento celular (agregados proteicos insolubles precipitados, también conocido como cuerpos de inclusión).

40 En la siguiente tabla se muestran las cantidades determinadas del polipéptido de ensayo solubilizadas de la fracción de sedimento celular insoluble.

Tabla.

Carril	Muestra	Cantidad de muestra [μl]	Señal Lumi Imager [ULB]	Cantidad determinada de polipéptido de ensayo por carril [ng]	Cantidad total de polipéptido de ensayo por muestra [ng]
2	referencia - 0 horas	5	Sin señal	0	0
4	referencia - 4 horas	0,02*	9703	6,3	315,7
6	referencia - 6 horas	0,02*	13461	9,0	448,9
8	referencia - 21 horas	0,02*	1448	0,5	22,9
3	Indicado en el presente documento 0 horas	5	Sin señal	0	0
5	Indicado en el presente documento 4 horas	0,02*	19440	13,2	660,9
7	Indicado en el presente documento 6 horas	0,02*	24014	16,5	823,1
9	Indicado en el presente documento 21 horas	0,02*	7738	4,9	246,0
10	patrón 1 (5 ng)	1	4803		
11	patrón 2 (10 ng)	2	16847		
12	patrón 3 (20 ng)	4	32733		
13	patrón 4 (30 ng)	6	40476		
14	patrón 5 (40 ng)	8	43858		

\*: muestra diluida a 1:50 con tampón de muestra buffer; volumen analizado 5μl

- 5 Como puede observarse en la tabla anterior, el rendimiento de la expresión que se obtiene usando un ácido nucleico que codifica un polipéptido de ensayo, en el que los codones codificantes se seleccionan de acuerdo con el método indicado en el presente documento, es al menos aproximadamente 1,8 veces el rendimiento que se obtiene usando un método clásico de optimización por codones.

A continuación, el método que se indica en el presente documento se ilustra usando la célula eucariota.

- 10 La célula puede ser una célula CHO. La frecuencia de uso de codones total teniendo en cuenta todos los codones que codifican un resto de aminoácido específico para especies de *Cricetulus* (células CHO; especies de *Mesocricetus*; hámster) se da en la siguiente tabla.

Tabla.

Ala	GCG	9	Gly	GGG	24	Pro	CCG	7
Ala	GCA	23	Gly	GGA	25	Pro	CCA	29
Ala	GCT	30	Gly	GGT	19	Pro	CCT	29
Ala	GCC	38	Gly	GGC	33	Pro	CCC	34
Arg	AGG	22	His	CAT	42	Ser	AGT	14
Arg	AGA	20	His	CAC	58	Ser	AGC	24
Arg	CGG	19				Ser	TCG	5
Arg	CGA	9	He	ATA	15	Ser	TCA	15
Arg	CGT	10	He	ATT	35	Ser	TCT	18
Arg	CGC	19	He	ATC	51	Ser	TCC	24
Asn	AAT	39	Leu	CTG	44	Thr	ACG	10
Asn	AAC	61	Leu	CTA	6	Thr	ACA	29
			Leu	CTT	13	Thr	ACT	21
Asp	GAT	39	Leu	CTC	19	Thr	ACC	40
Asp	GAC	61	Leu	TTG	12			
			Leu	TTA	6	Trp	TGG	100
Cys	TGT	42						
Cys	TGC	58	Lys	AAG	67	Tyr	TAT	39
			Lys	AAA	33	Tyr	TAC	61
Gln	CAG	78						
Gln	CAA	22	Met	ATG	100	Val	GTG	48
						Val	GTA	11
Glu	GAG	64	Phe	TTT	44	Val	GTT	16
Glu	GAA	36	Phe	TTC	56	Val	GTC	25

## ES 2 599 386 T3

- Para codificar al resto de aminoácido alanina se dispone de cuatro codones diferentes: GCG, GCA, GCT, y GCC. Por tanto, el grupo de codones que codifica el resto de aminoácido puede comprender a lo sumo cuatro codones. En el grupo que comprende cuatro codones, cada codón tiene una frecuencia de uso específica en el grupo que es el mismo que el de la frecuencia de uso total en el genoma de la célula, es decir, el codón GCG tiene una frecuencia de uso específica y total de 9 %, el codón GCA tiene una frecuencia de uso específica y total de 23 %, el codón GCT tiene una frecuencia de uso específica y total del 30 % y el codón GCC tiene una frecuencia de uso específica y total de 38 %. Si el número de codones en el grupo se reduce, por ejemplo, excluyendo del grupo los codones GCG y GCA con una frecuencia de uso específica de 9 % y 23 %, respectivamente, la frecuencia de uso específica de los miembros restantes del grupo, es decir, GCT y GCC, cambia a 44 % ( $=30/(30+38) * 100$ ) y 56 % ( $=38/(30+38) * 100$ ), respectivamente, ya que la suma de las frecuencias de uso específicas de todos los codones en un grupo es 100 %. Por tanto, en el grupo que comprende los dos codones GCT y GCC, cada codón tiene una frecuencia de uso específica en el grupo que es más alta que su frecuencia de uso total en el genoma de la célula, es decir, el codón GCT tiene una frecuencia de uso específica de 44 % y una frecuencia de uso total de 30 %, y el codón GCC tiene una frecuencia de uso específica de 56 % y una frecuencia de uso total de 38 %.
- Si el grupo de codones que codifica el resto de aminoácido alanina comprende los cuatro codones disponibles, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina comprende los codones GCC, GCT, GCA y GCG a una proporción de 38:30:23:9. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 40:30:20:10, que corresponde a 4:3:2:1.
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcc gcc gcc gct gct gct gca gca gcg (SEQ ID NO: 64).
- Como los codones se distribuyen dentro del genoma también se usa una distribución dentro del motivo codónico de aminoácidos teniendo en cuenta la proporción y la frecuencia de uso anteriores, por lo que los codones con una frecuencia más alta son los primeros que se seleccionan.
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcc gct gca gcc gct gca gcg gcc gct gcc (SEQ ID NO: 65).
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcc gct gcc gct gca gcg gcc gct gca gcc (SEQ ID NO: 66).
- Si el grupo de codones que codifica el resto de aminoácido alanina comprende los tres codones GCC, GCT y GCA el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina comprende los codones GCC, GCT y GCA a una proporción de 42:33:25. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 40:30:30, que corresponde a 4:3:3.
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcc gcc gcc gct gct gct gca gca gca (SEQ ID NO: 67).
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido alanina es gcc gct gca gcc gct gca gcc (SEQ ID NO: 68).
- Si el grupo de codones que codifica el resto de aminoácido arginina comprende los cuatro codones AGG, AGA, CGG y CGC el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina comprende los codones AGG, AGA, CGG y CGC a una proporción de 27:25:24:24. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 25:25:25:25, que corresponde a 1:1:1:1.
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina es agg aga cgg cgc (SEQ ID NO: 69).
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido arginina es agg cgg aga cgc (SEQ ID NO: 70).
- Si el grupo de codones que codifica el resto de aminoácido asparraguina comprende los dos codones AAC y AAT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido asparraguina comprende los codones AAC y AAT a una proporción de 61:39. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 60:40, que corresponde a 3:2.
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido asparraguina es aac aac aac aat aat (SEQ ID NO: 71).
- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido asparraguina es aac aat aac aat aac (SEQ ID NO: 72).

- 5 Si el grupo de codones que codifica el resto de aminoácido ácido aspártico comprende los dos codones GAC y GAT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido aspártico comprende los codones GAC y GAT a una proporción de 61:39. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 60:40, que corresponde a 3:2.
- 5 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido aspártico es gac gac gac gat gat (SEQ ID NO: 73).
- 10 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido aspártico es gac gat gac gat gac (SEQ ID NO: 74).
- 15 Si el grupo de codones que codifica el resto de aminoácido cisteína comprende los dos codones TGC y TGT, el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido cisteína comprende los codones TGC y TGT a una proporción de 58:42, que corresponde a 29:21. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones esto se ajusta a 30:20, que corresponde a 3:2.
- 20 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido cisteína es tgc tgc tgc tgt tgt (SEQ ID NO: 75).
- 20 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido cisteína es tgc tgt tgc tgt tgc (SEQ ID NO: 76).
- 25 Si el grupo de codones que codifica el resto de aminoácido glutamina comprende los dos codones CAG y CAA el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glutamina comprende los codones CAG y CAA a una proporción de 78:22, que corresponde a 39:11. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones esto se ajusta a 40:10, que corresponde a 4:1.
- 30 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glutamina es cag cag cag caa (SEQ ID NO: 77).
- 30 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glutamina es cag cag cag caa cag (SEQ ID NO: 78).
- 35 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glutamina es cag cag caa cag cag (SEQ ID NO: 79).
- 35 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glutamina es cag caa cag cag cag (SEQ ID NO: 80).
- 40 Si el grupo de codones que codifica el resto de aminoácido ácido glutámico comprende los dos codones CAG y CAA el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido ácido glutámico comprende los codones CAG y CAA a una proporción de 64:36, que corresponde a 32:18. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones esto se ajusta a 32:16, que corresponde a 2:1.
- 45 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido de ácido glutámico es gag gag gaa (SEQ ID NO: 81).
- 50 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido de ácido glutámico es gag gaa gag (SEQ ID NO: 82).
- 55 Si el grupo de codones que codifica el resto de aminoácido glicina comprende todos los codones disponibles el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina comprende los codones GGC, GGA, GGG y GGT a una proporción de 33:25:24:19. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 35:25:25:20, que corresponde a 7:5:5:4.
- 60 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina es ggc ggc ggc ggc ggc ggc gga gga gga gga gga gga ggg ggg ggg ggg ggg ggt ggt ggt (SEQ ID NO: 83).
- 60 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido glicina es ggc gga ggg ggt ggc gga ggc ggg ggt ggc gga ggg ggt ggc gga ggc ggg ggc gga ggg ggt (SEQ ID NO: 84).
- 65 Si el grupo de codones que codifica el resto de aminoácido histidina comprende los dos codones CAC y CAT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido histidina comprende los codones CAC y CAT a una proporción de 58:42 que corresponde a 29:21. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 50 posiciones esto se ajusta a 30:20, que corresponde a 3:2.



- Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido histidina es cac cac cac cat cat (SEQ ID NO: 85).
- 5 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido histidina es cac cat cac cat cac (SEQ ID NO: 86).
- Si el grupo de codones que codifica el resto de aminoácido isoleucina comprende todos los codones disponibles el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina comprende los codones ATC, ATT y ATA a una proporción de 51:35:15. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 51:35:15, que corresponde a 10:7:3.
- 10 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina es atc atc atc atc atc atc atc atc att att att att att att att ata ata ata (SEQ ID NO: 87).
- 15 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido isoleucina es atc att atc atc att ata atc att atc atc att ata atc att atc atc att ata atc att (SEQ ID NO: 88).
- Si el grupo de codones que codifica el resto de aminoácido leucina comprende los cuatro codones CTG, CTC, CTT y TTG el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina comprende los codones CTG, CTC, CTT y TTG a una proporción de 44:19:13:12. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 40:20:10:10, que corresponde a 4:2:1: 1.
- 20 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctg ctg ctc ctc ctt ttg (SEQ ID NO: 89).
- 25 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctc ctg ctt ctg ctc ctg ttg (SEQ ID NO: 90).
- 30 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido leucina es ctg ctc ctt ttg ctg ctc ctg ctg (SEQ ID NO: 91).
- Si el grupo de codones que codifica el resto de aminoácido lisina comprende los dos codones AAG y AAA el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina comprende los codones AAG y AAA a una proporción de 67:33. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 66:33, que corresponde a 2:1.
- 35 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aag aag aaa (SEQ ID NO: 92).
- 40 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido lisina es aag aaa aag (SEQ ID NO: 93).
- Si el grupo de codones que codifica el resto de aminoácido fenilalanina comprende los dos codones TTC y TTT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina comprende los codones TTC y TTT a una proporción de 56:44, que corresponde a 14:11. Como esto daría como resultado un motivo codónico de aminoácidos que comprende las 25 posiciones esto se ajusta a 15:10, que corresponde a 3:2.
- 45 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina es ttc ttc ttc ttt ttt (SEQ ID NO: 94).
- 50 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido fenilalanina es ttc ttt ttc ttt ttc (SEQ ID NO: 95).
- Si el grupo de codones que codifica el resto de aminoácido prolina comprende los tres codones CCC, CCA y CCT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina comprende los codones CCC, CCA y CCT a una proporción de 34:29:29. Como esto daría como resultado un motivo codónico de aminoácidos que comprende las 100 posiciones esto se ajusta a 35:30:30, que corresponde a 7:6:6.
- 55 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccc ccc ccc ccc ccc ccc cca cca cca cca cca cca cct cct cct cct cct cct (SEQ ID NO: 96).
- 60 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido prolina es ccc cca cct ccc cca cct ccc cca cct ccc cca cct ccc cca cct ccc (SEQ ID NO: 97).
- 65 Si el grupo de codones que codifica el resto de aminoácido serina comprende los cuatro codones TCC, AGC, TCT y TCA el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina comprende los

codones TCC, AGC, TCT y TCA a una proporción de 24:24:18:15, que corresponde a 8:8:6:3. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 25 posiciones esto se ajusta a 9:9:6:3, que corresponde a 3:3:2:1.

5 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina es tcc tcc tcc agc agc agc tct tct tca (SEQ ID NO: 98).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina es tcc agc tct tca tcc agc tct tcc agc (SEQ ID NO: 99).

10 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido serina es tcc agc tct tcc agc tca tct tcc agc (SEQ ID NO: 100).

15 Si el grupo de codones que codifica el resto de aminoácido prolina comprende los tres codones ACC, ACA y ACT el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina comprende los codones ACC, ACA y ACT a una proporción de 45:32:23. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 50:30:20, que corresponde a 5:3:2.

20 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc acc acc acc acc aca aca aca act act (SEQ ID NO: 101).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc aca act acc aca acc aca act acc acc (SEQ ID NO: 102).

25 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido treonina es acc aca acc aca act acc acc aca act acc (SEQ ID NO: 103).

30 Si el grupo de codones que codifica el resto de aminoácido tirosina comprende los dos codones TAT y TAC el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina comprende los codones TAT y TAC a una proporción de 61:39. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 60:40, que corresponde a 3:2.

35 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina es tat tat tat tac tac (SEQ ID NO: 104).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido tirosina es tat tac tat tac tat (SEQ ID NO: 105).

40 Si el grupo de codones que codifica el resto de aminoácido valina comprende los cuatro codones disponibles el motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina comprende los codones GTG, GTC, GTT y GTA a una proporción de 48:25:16:11. Como esto daría como resultado un motivo codónico de aminoácidos que comprende 100 posiciones esto se ajusta a 48:24:18:12, que corresponde a 8:4:3:2.

45 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtg gtg gtg gtg gtg gtg gtc gtc gtc gtc gtt gtt gtt gta gta (SEQ ID NO: 106).

Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtg gtg gtc gtt gta gtg gtg gtc gtt gtg gtg gtc gtt gta gtg gtg gtc (SEQ ID NO: 107).

50 Por tanto, un motivo codónico de aminoácidos de los codones que codifican el resto de aminoácido valina es gtg gtc gtt gta gtg gtc gtt gta gtg gtc gtt gtg gtc gtg gtg gtg (SEQ ID NO: 108).

Los siguientes ejemplos, secuencias y figuras se proporcionan para ayudar a entender la presente invención, cuyo alcance verdadero se expone en las reivindicaciones adjuntas.

55 **Ejemplos**

Determinación de proteínas:

60 La concentración proteica se determinó determinando la densidad óptica (DO) a 280 nm, usando el coeficiente de extinción molar calculado basándose en la secuencia de aminoácidos.

Técnica de ADN recombinante:

65 Para manipular el ADN se usaron métodos convencionales como se describe en Sambrook, J., et al., Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York (1989). Se

usaron reactivos de biología molecular de acuerdo con las instrucciones del fabricante.

### Ejemplo 1

5 Fabricación y descripción de los plásmidos de expresión del polipéptido de ensayo.

Se preparó un polipéptido de ensayo de fusión por medios recombinantes. La secuencia de aminoácidos del polipéptido de ensayo de fusión expresado se codificó con un ácido nucleico que comprendía, en dirección de 5' a 3', un ácido nucleico de SEQ ID NO: 61 que codificaba el péptido transportador, un ácido nucleico de SEQ ID NO: 62 o  
10 SEQ ID NO: 63 que codificaba el polipéptido de ensayo, y un ácido nucleico de SEQ ID NO: 59 que codificaba una etiqueta de purificación de seis histidinas (etiqueta de poli-His).

El gen codificante de fusión se ensambló con métodos y técnicas recombinantes conocidos por conexión de segmentos de ácido nucleico apropiados. Las secuencias de ácido nucleico fabricadas por síntesis química se verificaron mediante secuenciación de ADN. Para la producción del polipéptido de fusión el plásmido de expresión se preparó como se indica más adelante.  
15

### Fabricación del plásmido de expresión en *E. coli*.

20 El plásmido 4980 (4980-pBRori-URA3-LACI-SAC) es un plásmido de expresión para la expresión de estreptavidina núcleo en *E. coli*. Esto se generó por ligamiento del fragmento del vector EcoRI/CeIII de 3124 pb de longitud procedente del plásmido 1966 (1966-pBRori-URA3-LACI-T-repetición; descrito en el documento EP-B 1 422 237) con una estreptavidina núcleo de 435 pb de longitud que codifica el fragmento EcoRI/CeIII.

25 El plásmido de expresión en *E. coli* para la estreptavidina núcleo comprende los siguientes elementos:

- el origen de replicación del vector pBR322 para la replicación en *E. coli* (correspondiente a los pb de las posiciones 2517-3160 según Sutcliffe, G., et al., Quant. Biol. 43 (1979) 77-90),
- el gen URA3 de *Saccharomyces cerevisiae* que codifica la orotidina 5'-fosfato descarboxilasa (Rose, M., et al., Gene 29 (1984) 113-124) que permite la selección del plásmido por complementación de cepas mutantes pyrF de *E. coli* (por auxotrofia a uracilo),
- el casete de expresión del núcleo de estreptavidina que comprende
  - el promotor híbrido de T5 (promotor híbrido T5-PN25/03/04 según Bujard, H., et al., Methods. Enzymol. 155 (1987) 416-433 y Stueber, D., et al., Immunol. Methods IV (1990) 121-152) incluyendo un sitio de unión al ribosoma sintético según Stueber, D., et al. (véase anteriormente),
  - el gen de la estreptavidina núcleo,
  - dos terminadores de la transcripción procedentes de dos bacteriófagos, el terminador  $\lambda$ -T0 (Schwarz, E., et al., Nature 272 (1978) 410-414) y el terminador fd (Beck E. y Zink, B. Gene 1-3 (1981) 35-58),
- el gen represor lacl de *E. coli* (Farabaugh, P.J., Nature 274 (1978) 765-769).

El plásmido de expresión final para la expresión del polipéptido de ensayo de fusión se preparó escindiendo el gen estructural de la estreptavidina núcleo del vector 4980 usando el sitio de escisión de EcoRI flanqueante singular y la endonucleasa de restricción CeIII e insertando el ácido nucleico flanqueado por el sitio de restricción EcoRII/CeIII que codifica el polipéptido de ensayo de fusión, en el fragmento del vector EcoRI/CeIII-4980 de 3142 pb de longitud.  
45

El plásmido de expresión que contenía el gen del polipéptido de ensayo generado con el uso de codones clásico se denominó 11020 mientras que el plásmido de expresión que contenía el gen del polipéptido de ensayo generado con el uso de codones novedoso se denominó 11021.  
50

### Ejemplo 2

Expresión del polipéptido de ensayo en *E. coli*

55 Para la expresión del polipéptido de ensayo de fusión se empleó un sistema de hospedador/vector *E. coli* que permite una selección de plásmido sin antibiótico por complementación de una auxotrofia en *E. coli* (PyrF) (documentos EP 0 972 838 y US 6.291.245).

60 Transformación, cultivo celular e inducción de células de *E. coli* transformadas

La cepa K12 de *E. coli* CSPZ-2 (leuB, proC, trpE, thi-1,  $\Delta$ pyrF) se transformó con el plásmido de expresión (11020 y 11021, respectivamente) obtenido en la etapa anterior. Las células CSPZ-2 transformadas se cultivaron primero a 37°C en placas de agar y posteriormente en un cultivo con agitación en medio mínimo M9 que contenía casaminoácidos al 0,5 % (Difco) hasta una densidad óptica de 550 nm (DO550) de 0,6–0,9 y posteriormente se indujo con IPTG (concentración final 1-5mmol/l).  
65

Ejemplo 3Expresión y análisis del polipéptido de ensayo

5 El polipéptido de ensayo de fusión expresado se visualizó después de SDS PAGE por análisis de transferencia de Western cuantitativo. A continuación, el lisado de *E. coli* se procesó mediante electroforesis en gel de poliacrilamida con dodecil sulfato sódico (SDS) (SDS-PAGE), y los polipéptidos separados se transfirieron a una membrana desde el gel y posteriormente se detectaron y cuantificaron mediante un método inmunológico.

10 Muestreo de células de *E. coli* y preparación de muestras para SDS-PAGE

Para los análisis de expresión, se extrajeron muestras de cultivo de células de *E. coli* desde el cultivo con agitación durante una evolución de 24 horas aproximadamente. Una muestra se extrajo antes de la inducción de la expresión de la proteína recombinante. Se tomaron muestras adicionales a momentos de tiempo dedicados, por ejemplo, 4, 6 y 15 21 horas después de la inducción.

Los sedimentos de células de *E. coli* de 3 unidades DO550 nm (1 DO550 nm = 1 ml de suspensión celular con una DO a 550 nm de 1) de medio de cultivo centrifugado se resuspendieron en 0,25 ml de tampón de fosfato potásico 10 mmol/l, pH 6,5 y las células se lisaron con tratamiento ultrasonido (dos pulsos de 30 segundos a una intensidad de 50 %). Los componentes celulares insolubles se sedimentaron (centrifugación a 14.000 rpm, 5 minutos) y una parte alícuota del sobrenadante clarificado se mezcló con un volumen 1/4 (v/v) de tampón de muestra 4xLDS y un volumen 1/10 (v/v) de 1,4-ditiotreitol (DTT) 0,5M. La fracción residual celular insoluble (sedimento) se resuspendió/extrajo en tampón de muestra 1xLDS 0,3 ml que contenía 1,4-ditiotreitol (DTT) 50 mM con agitación durante 15 minutos y de nuevo se centrifugó.

25 SDS-PAGE

Tampón de muestra LDS, concentrado cuatro veces (4x): glicerol 4g, TRIS-Base 0,682g, clorhidrato de TRIS 0,666 g, LDS (dodecil sulfato de litio) 0,8 g, EDTA (ácido etilendiaminatetracético) 0,006g, 0,75 ml de una solución al 1 % en peso (p/p) de Serva Blue G250 en agua, 0,75 ml de una solución al 1 % en peso (p/p) de rojo fenol, se añade agua para constituir un volumen total de 10 ml.

Se usó el sistema de gel Pre-Cast NuPAGE® (Invitrogen) de acuerdo con las instrucciones del fabricante (geles Pre-Cast Bis-TRIS Novex® NuPAGE® al 10 %, pH 6,4; N.º de catálogo: NP0301). Las muestras se incubaron durante 10 minutos a 70 °C y después de enfriar a temperatura ambiente se cargaron 5-40 µl en los geles. Además, en el gel, se cargaron 5 µl de patrón de proteína Western MagicMark™ XP (20-220 kDa) (Invitrogen, N.º de catálogo: LC5602), 5 µl de patrón de proteína previamente teñida Precision Plus Protein™ (Bio-Rad, N.º de catálogo.: 161-0373) y 1, 2, 4, 6 y 8 µl de patrón de cuantificación scFv-poli-His purificado (concentración de proteína: 5 ng/µl). La separación de las proteínas tuvo lugar en tampón de desarrollo reductor SDS MOPS de NuPAGE® (Invitrogen, N.º de catálogo: NP0001) durante 60 minutos a 180 V.

Para la transferencia de Western/SDS-PAGE las muestras se distribuyeron como se muestra en las Figuras 1 y 2.

carril	muestra	fuentes	cantidad
1	Magic Mark™		5 µl
2	referencia, 0 h	sobrenadante	5 µl
3	Indicado en el presente documento, 0 h	sobrenadante	5 µl
4	referencia, 4 h	sobrenadante	2 µl
5	Indicado en el presente documento, 4 h	sobrenadante	2 µl
6	referencia, 6 h	sobrenadante	2 µl
7	Indicado en el presente documento, 6 h	sobrenadante	2 µl
8	referencia, 21 h	sobrenadante	5 µl
9	Indicado en el presente documento, 21 h	sobrenadante	5 µl
10	patrón		5 ng
11	patrón		10 ng
12	patrón		20 ng
13	patrón		30 ng
14	patrón		40 ng
15	Precision Plus Protein™		5 µl

carril	muestra	fuentes	cantidad
1	Magic Mark™		5 µl
2	referencia, 0h	Sedimento	5 µl
3	Indicado en el presente documento, 0h	Sedimento	5 µl
4	referencia, 4h	Sedimento	40 µl
5	Indicado en el presente documento, 4h	Sedimento	40 µl
6	referencia, 6h	Sedimento	40 µl
7	Indicado en el presente documento, 6h	Sedimento	40 µl
8	referencia, 21h	Sedimento	40 µl
9	Indicado en el presente documento, 21h	Sedimento	40 µl
10	patrón		5 ng
11	patrón		10 ng
12	patrón		20 ng
13	patrón		30 ng
14	patrón		40 ng
15	Precision Plus Protein™		5 µl

#### Transferencia de Western

5 Tampón de transferencia: glicina 39 mM, clorhidrato de TRIS 48 mM, 0,04 % en peso (p/p) de SDS y 20 % en volumen (v/v) de metanol.

Después de SDS-PAGE, los polipéptidos separados se transfirieron electroforéticamente a una membrana de filtro de nitrocelulosa (tamaño de poro: 0,45 µM, Invitrogen, N.º de Catálogo LC2001) según el "Método de Transferencia en Semiseco" de Burnette (Burnette, W.N., Anal. Biochem. 112 (1981) 195-203).

10 Detección inmunológica y cuantificación del polipéptido de ensayo con etiqueta de poli-His.

Después de la electrotransferencia las membranas se lavaron en Tris-HCl 50 mM, pH7,5, NaCl 150 mM (TBS, solución salina tamponada con tris) y se bloquearon sitios de unión inespecífica durante una noche a 4 °C en TBS, reactivo de bloqueo de Western al 1 % (p/v) (Roche, N.º de catálogo: 11921673001).

20 El anticuerpo anti-Penta-His monoclonal de ratón (Qiagen, N.º de catálogo 34660) se usó como anticuerpo primario a una dilución de 1:1.000 en TBS, solución de bloqueo de Western al 0,5 % (p/v). Después de dos lavados en TBS (Bio-Rad, Cat. No.: 170-6435) y de dos lavados en TBS complementado con Tween-20 (TBST) al 0,05 % (v/v) los polipéptidos que contenían la poli-His se visualizaron usando un como anticuerpo secundario un anticuerpo de conejo anti IgG de antirratón purificado conjugado con una peroxidasa (Roche Molecular Biochemicals, Cat. No.: 11693 506) a una dilución de 1:400 en TBS con leche en polvo descremada al 3 % (p/v).

25 Después de lavar las membranas tres veces con tampón TBTS y una vez con tampón TBS durante 10 minutos a temperatura ambiente, las membranas de la transferencia de Western se revelaron con una solución de Luminol/peróxido que genera luminiscencia (sustrato de transferencia de Western Lumi-LightPLUS, Roche Molecular Biochemicals, N.º de catálogo 12015196001). Por lo tanto las membranas se incubaron en solución de Luminol/peróxido 10 ml durante 10 segundos a 5 minutos y la luz emitida se detectó después con una analizador LUMI-Imager F1 (Roche Molecular Biochemicals) y se obtuvo una curva patrón de referencia de proteína representando gráficamente la concentración proteica conocida de las proteínas scFv-poli-His frente a sus señales homólogas medidas con el LUMI-Imager (la intensidad de las manchas se expresó en unidades ULB) que se usó para el cálculo de las concentraciones de la proteína diana en las muestras originales.

35 La intensidad de las manchas se cuantificó con el programa informático LumiAnalyst (Versión 3.1).

Tabla.

Carril	Muestra	Cantidad de muestra [μl]	Señal Lumi Imager [ULB]	Cantidad determinada de polipéptido de ensayo por carril [ng]	Cantidad total de polipéptido de ensayo por muestra [ng]
2	referencia - 0 horas	5	Sin señal	0	0
4	referencia - 4 horas	0,02*	9703	6,3	315,7
6	referencia - 6 horas	0,02*	13461	9,0	448,9
8	referencia - 21 horas	0,02*	1448	0,5	22,9
3	Indicado en el presente documento 0 horas	5	Sin señal	0	0
5	Indicado en el presente documento 4 horas	0,02*	19440	13,2	660,9
7	Indicado en el presente documento 6 horas	0,02*	24014	16,5	823,1
9	Indicado en el presente documento 21 horas	0,02*	7738	4,9	246,0
10	Patrón 1 (5 ng)	1	4803		
11	Patrón 2 (10 ng)	2	16847		
12	Patrón 3 (20 ng)	4	32733		
13	Patrón 4 (30 ng)	6	40476		
14	Patrón 5 (40ng)	8	43858		

\*: Muestra diluida a 1:50 con tampón de muestra; volumen analizado 5 μl

5 En la Figura 3 se muestra la curva patrón de referencia de las proteínas, obtenida a partir de cinco concentraciones de scFv-poli-His conocidas.

En la figura 1 se muestra la transferencia de Western del polipéptido que contenía los sobrenadantes. En la figura 2 se muestra la transferencia de Western de la fracción de sedimento celular extraída con SDS.

10 LISTADO DE SECUENCIAS

<110> F. Hoffmann-La Roche AG

<120> Método para le expresión de polipéptidos usando ácidos nucleicos modificados

15

<130> 30934 WO

<150> EP12164430.6

20

<151> 17-4-2012

<160> 108

<170> PatentIn versión 3.5

25

<210> 1

<211> 39

<212> ADN

<213> Secuencia artificial

30

<220>

<223> motivo codónico del aminoácido alanina

<400> 1

gcggcggcgg cggctgctgc tgctgcagca gcagccgc 39

35

<210> 2

<211> 39

<212> ADN

<213> Secuencia artificial

40

<220>

<223> motivo codónico del aminoácido alanina

<400> 2  
 gcggctgcag ccgcggtgc agcggctgcc gcggctgca 39

5 <210> 3  
 <211> 39  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido alanina

<400> 3  
 gcggctgcag ccgcggtgc agccggtgc gcagcggct 39

15 <210> 4  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido alanina

25 <400> 4  
 gcggct 6

<210> 5  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido alanina

35 <400> 5  
 gctgcg 6

<210> 6  
 <211> 9  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido arginina

45 <400> 6  
 cgtcgtcgc 9

50 <210> 7  
 <211> 9  
 <212> ADN  
 <213> Secuencia artificial

55 <220>  
 <223> motivo codónico del aminoácido arginina

<400> 7  
 cgtcggcgt 9

60 <210> 8  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

65 <220>  
 <223> motivo codónico del aminoácido asparragina

<400> 8  
 aacaacaaca acaacaat 18

5 <210> 9  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido asparragina

<400> 9  
 aacaacaaca acaataac 18

15 <210> 10  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido asparragina

<400> 10  
 aacaacaaca ataacaac 18

25 <210> 11  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido asparragina

<400> 11  
 aacaacaata acaacaac 18

35 <210> 12  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido asparragina

45 <400> 12  
 aacaataaca acaacaac 18

<210> 13  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <223> motivo codónico del aminoácido ácido aspártico

55 <400> 13  
 gacgat 6

60 <210> 14  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial

65 <220>  
 <223> motivo codónico del aminoácido ácido aspártico



<400> 14  
gatgac 6

5 <210> 15  
<211> 24  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> motivo codónico del aminoácido ácido cisteína

<400> 15  
tgctgctgct gctgctgtg ttgt 24

15 <210> 16  
<211> 24  
<212> ADN  
<213> Secuencia artificial

20 <220>  
<223> motivo codónico del aminoácido ácido cisteína

<400> 16  
tgctgttgct gctgttgctg ctgt 24

25 <210> 17  
<211> 24  
<212> ADN  
<213> Secuencia artificial

30 <220>  
<223> motivo codónico del aminoácido ácido cisteína

<400> 17  
tgctgctggt gctgttgctg ttgc 24

35 <210> 18  
<211> 15  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> motivo codónico del aminoácido glutamina

45 <400> 18  
cagcagcagc agcaa 15

<210> 19  
<211> 15  
<212> ADN  
<213> Secuencia artificial

50 <220>  
<223> motivo codónico del aminoácido glutamina

55 <400> 19  
cagcagcagc aacag 15

60 <210> 20  
<211> 15  
<212> ADN  
<213> Secuencia artificial

65 <220>  
<223> motivo codónico del aminoácido glutamina

<400> 20  
cagcagcaac agcag 15

5 <210> 21  
<211> 15  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> motivo codónico del aminoácido glutamina

<400> 21  
cagcaacagc agcag 15

15 <210> 22  
<211> 12  
<212> ADN  
<213> Secuencia artificial

20 <220>  
<223> motivo codónico del aminoácido ácido glutámico

<400> 22  
gaagaagaag ag 12

25 <210> 23  
<211> 12  
<212> ADN  
<213> Secuencia artificial

30 <220>  
<223> motivo codónico del aminoácido ácido glutámico

<400> 23  
gaagaagagg aa 12

35 <210> 24  
<211> 12  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> motivo codónico del aminoácido ácido glutámico

45 <400> 24  
gaagaggaag aa 12

<210> 25  
<211> 27  
50 <212> ADN  
<213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido glicina

55 <400> 25  
ggtggtggtg gtggtggcgg cggcggc 27

<210> 26  
<211> 27  
60 <212> ADN  
<213> Secuencia artificial

<220>  
65 <223> motivo codónico del aminoácido glicina

<400> 26  
ggtggcgggtg gcggtggcgg tggcggt 27

5 <210> 27  
<211> 9  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> motivo codónico del aminoácido histidina

<400> 27  
caccacat 9

15 <210> 28  
<211> 9  
<212> ADN  
<213> Secuencia artificial

20 <220>  
<223> motivo codónico del aminoácido histidina

<400> 28  
caccatcac 9

25 <210> 29  
<211> 9  
<212> ADN  
<213> Secuencia artificial

30 <220>  
<223> motivo codónico del aminoácido isoleucina

<400> 29  
atcatcatt 9

35 <210> 30  
<211> 9  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> motivo codónico del aminoácido isoleucina

45 <400> 30  
atcattatc 9

<210> 31  
<211> 30  
<212> ADN  
<213> Secuencia artificial

50 <220>  
<223> motivo codónico del aminoácido leucina

55 <400> 31  
ctgctgctgc tgctgctgct gctgctgctc 30

60 <210> 32  
<211> 30  
<212> ADN  
<213> Secuencia artificial

65 <220>  
<223> motivo codónico del aminoácido leucina

<400> 32  
 ctgctgctgc tgctgctcct gctgctgctg 30

5 <210> 33  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido leucina

<400> 33  
 ctgctgctgc tgctcctgct gctgctgctg 30

15 <210> 34  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido lisina

<400> 34  
 aaaaaaaaaa aaaag 15

25 <210> 35  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido lisina

<400> 35  
 aaaaaaaaaa agaaa 15

35 <210> 36  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido lisina

45 <400> 36  
 aaaaaaaaga aaaaa 15

<210> 37  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <223> motivo codónico del aminoácido lisina

55 <400> 37  
 aaaaagaaaa aaaaa 15

<210> 38  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

60 <220>  
 <223> motivo codónico del aminoácido fenilalanina

65 <220>  
 <223> motivo codónico del aminoácido fenilalanina

<400> 38  
 ttcttcttct tt 12

5 <210> 39  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido fenilalanina

15 <400> 39  
 ttcttctttt tc 12

20 <210> 40  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

25 <220>  
 <223> motivo codónico del aminoácido fenilalanina

30 <400> 40  
 ttcttttct tc 12

35 <210> 41  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido prolina

45 <400> 41  
 ccgccgccgc cgccgccacc t 21

50 <210> 42  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

55 <220>  
 <223> motivo codónico del aminoácido prolina

60 <400> 42  
 ccgccgccgc caccgcctcc g 21

65 <210> 43  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

70 <220>  
 <223> motivo codónico del aminoácido prolina

75 <400> 43  
 ccgccgccac cgccgcctcc g 21

80 <210> 44  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

85 <220>  
 <223> motivo codónico del aminoácido prolina

<400> 44  
 ccgccgccac cgcctccgcc g 21

5 <210> 45  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido prolina

<400> 45  
 ccgccaccgc cgcctccgcc g 21

15 <210> 46  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido prolina

<400> 46  
 ccgccaccgc ctccgccgcc g 21

25 <210> 47  
 <211> 48  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido serina

<400> 47  
 tcttcttctt ctctcttctt ctctctctcc tccagcagca gcagcagc 48

35 <210> 48  
 <211> 48  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido serina

45 <400> 48  
 tctccagct ctccagctc ttccagctct tccagctcct ctagctct 48

<210> 49  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <223> motivo codónico del aminoácido treonina

55 <400> 49  
 accaccacca ccactactac g 21

60 <210> 50  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

65 <220>  
 <223> motivo codónico del aminoácido treonina

<400> 50  
 accactacca ctaccacgac c 21

5 <210> 51  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido treonina

<400> 51  
 accactacca cgaccactac c 21

15 <210> 52  
 <211> 9  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido tirosina

<400> 52  
 tactactat 9

25 <210> 53  
 <211> 9  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido tirosina

<400> 53  
 tactattac 9

35 <210> 54  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> motivo codónico del aminoácido valina

45 <400> 54  
 gttgtgttg ttgtgtggt gtagtagtc 30

<210> 55  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <223> motivo codónico del aminoácido valina

55 <400> 55  
 gttgtgtag tcgtgtggt agttgtggtt 30

60 <210> 56  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

65 <220>  
 <223> motivo codónico del aminoácido valina

<400> 56  
gttgtggtag ttgtggtgt cgttggta 30

5 <210> 57  
<211> 300  
<212> PRT  
<213> Secuencia artificial

10 <220>  
<223> secuencia de aminoácidos del polipéptido de ejemplo

<400> 57



ES 2 599 386 T3

Asp Ser Ala Val Asp Ser Gln Gly Thr Ser Phe Ser Glu Tyr Val Gly  
 1 5 10 15  
 Ala Phe Val Ser Val Asp Ala Gly His Lys Ala Ala Glu Ser Gln Ala  
 20 25 30  
 Ser Val Ser Ser Val Tyr Asn Leu Ala Val Pro Ala Tyr Arg Ala Ser  
 35 40 45  
 Tyr Val Arg Ser Asp Thr Ser Asp Ile Asp Thr Ala Ala Val Ser Ser  
 50 55 60  
 Pro Asp Val Val Asp Ile Ile Glu Arg Val Lys Ser Tyr Ser Arg Gly  
 65 70 75 80  
 Ser Val Thr Ala Ala Tyr Ala Ile Gly Val Arg Tyr Asp Trp Ser Arg  
 85 90 95  
 Ser His Ser Gly Ser Glu Thr Ser Thr Ser Asn Phe Ala Tyr Thr Tyr  
 100 105 110  
 Ser Leu Asn Ser Thr Gln Thr Phe Val Tyr Ala Ser Lys Ala Arg Ser  
 115 120 125  
 Ala Leu Ala Ala Val Val Val Gly Val Arg Glu Ser Ile Thr Gly Ser  
 130 135 140  
 Ser Gly Gln Val Phe Phe Ala Ala Thr Ser Thr Ala Ser Asp Ala His  
 145 150 155 160  
 Ala Ser Thr Gly Ala Asp Ile Asp Pro Thr Ala Val Val His Thr Asp  
 165 170 175  
 Val Ser Val Val Ile Ser Ala Phe Ala Val Ala Ala His Gly Val Ala  
 180 185 190  
 Arg Val His His Val Ile Ala Ser Ile Asp Tyr Ala Val Asp Ala Gly  
 195 200 205  
 Ala Ala Gly Ala Ala Gly Ser Ser Gly Gly Thr Arg Ile Ala Gly Val  
 210 215 220  
 Val Val Ser Val Thr Ile Arg Gly Phe Ser Leu Thr Gly Leu Gly Ala  
 225 230 235 240

ES 2 599 386 T3

Gly Asp Val Gly Pro His Thr Ala Arg Tyr Ala Gly Glu Ser Phe Ser  
 245 250 255

Val Asp Cys Ser Arg Gly Ala Ser His Val Ala Ser Ser Ala Lys Pro  
 260 265 270

Ala Ser Val Thr Asp Met Thr Pro Tyr Arg Ser Val Thr Asp Asp Ala  
 275 280 285

Ser Asp Asp Gly Pro Ala Ser Val Ser Asp Gly Tyr  
 290 295 300

5 <210> 58  
 <211> 8  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <223> etiqueta de seis histidinas  
 <400> 58

Gly Ser His His His His His His  
 1 5

15 <210> 59  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> ácido nucleico que codifica una etiqueta de seis histidinas  
 <400> 59  
 ggttctcacc accaccacca ccac 24

25 <210> 60  
 <211> 12  
 <212> PRT  
 <213> Secuencia artificial

30 <220>  
 <223> secuencia de aminoácidos del péptido transportador  
 <400> 60

Met Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser  
 1 5 10

40 <210> 61  
 <211> 36  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> ácido nucleico que codifica el péptido transportador  
 <400> 61  
 atgtgcgacc tgccgcagac ccaactccctt ggatcc 36

ES 2 599 386 T3

<210> 62  
 <211> 900  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> ácido nucleico que codifica el polipéptido ejemplar obtenido usando solo el codón más abundante para cada resto de aminoácidos

10

<400> 62

```

gactctgcgg ttgactctca gggtagctct ttctctgaat acgttggtgc gttcgtttct      60
gttgacgcgg gtcacaaagc ggcggaatct caggcgtctg tttcttctgt ttacaacctg      120
gcggttcgag cgtaccgtgc gtcttacggt cgttctgaca cctctgacat cgacaccgag      180
gcggtttctt ctccggacgt tgttgacatc atcgaacgtg ttaaacttta ctctcgtggt      240
tctgttacag cggcgtacgc gatcgggtgt cgttacgact ggtctcgttc tcaactctggt      300
tctgaaacct ctacctataa cttcgcgtac acctactctc tgaactctac ccagaccttc      360
gtttacgcgt ctaaagcgag ttctgcgctg gggcgggttg ttggtggtgt tcgtgaatct      420
atcaccggtt cttctggta cggttttcttc gcggcgacct ctaccgcgtc tgacgcgcac      480
gcgtctaccg gtgcggacat cgaccgcacc gcggttggtc acaccgacgt ttctgttggt      540
atctctgcgt tcgcggttgc ggcgcacggt gttgcgcgtg ttcaccacgt tctcgcgtct      600
atcgactaag cggttgaagc gggtagcggc ggtgcggcgg gttcttctgg tggtagccgt      660
atcgcgggtg ttggtgtttc tgttaccatc cgtggtttct ctctgaccgg tctgggtgag      720
ggtgacgttg gtcgcacac cgcgcggttac ggggtgaat cttctctctg tgactgctct      780
cgtggtgagt ctcaagttgc gtcttctgag aaaccggcgt ctggtaccga catgaccccg      840
taccgttctg ttaccgacga cgcgtctgac gacggtccgg cgtctgtttc tgacggttac      900
    
```

15

<210> 63  
 <211> 900  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> ácido nucleico que codifica el polipéptido ejemplar obtenido con un método como se describe en el presente documento

<400> 63

ES 2 599 386 T3

gactctgcgg ttgattccca gggtagaccgac ttctctgaat acgtggggcgc tttcgtatcc 60  
 gttgacgcag gtcacaaaagc cgcggaaagc caggcttctg tgtccagcgt ttataaacctg 120  
 gcagtcccgg cctaccgtgc gtcttacggt cgctccgata ctaccgacat cgataaccgct 180  
 gcagtgtcct ccccggaagc agttgatatt atcagagcgtg tgaaaagcta ttctcgtggc 240  
 tctgtaacgg cggcgtagcgc tatcgggtgt cgctaccgact ggtcccgtag ccattctggc 300  
 tccgaaacca gcaattotaa ctttgcatat acctactccc tgaacagcac ccaaactttc 360  
 gtgtaccgct ctaaggecgc ttccgctctg gcagccgttg tcgttggtgt ggcgcgaaagc 420  
 attaccggct cttccgggtca ggtattcttc gcggctacga gcaccgcac tgatgogcac 480  
 gcgtctactg gtgctgacat cgatccaacc gcagttgtgc acaccgacgt atccggtgtg 540  
 atcagcgctt ttgcggttgc tgcacatggc gtccgctgag ttcaccacgt gattgogctc 600  
 atcgattatg ctgtagacgc aggtgcccgc ggccgctgag gttccagcgg cggtagctctg 660  
 atcgccggcg ttgtggtatc tgttaccatt cgcggttctt cctgacggg tctcggcgcg 720  
 ggtgatgtgg gcccgcatat cgctcgttac gcaggtgaaa gcttctctgt tgactgctcc 780  
 cgtggcgcca gccacgtcgc gtcttccgct aaaccggcaa gcgttactga tatgaccct 840  
 taccgctctg tgaccgacga tgcgtctgac gatggtccgc cgtccgtaag cgacggctat 900

5 <210> 64  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

10 <220>  
 <223> motivo codónico del aminoácido alanina

<400> 64  
 gccgcccgcg ccgctgctgc tgcagcagcg 30

15 <210> 65  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> motivo codónico del aminoácido alanina

<400> 65  
 gccgctgcag ccgctgcagc ggccgctgcc 30

25 <210> 66  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> motivo codónico del aminoácido alanina

<400> 66  
 gccgctgccg ctgcagcggc cgctgcagcc 30

35

<210> 67  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <223> motivo codónico del aminoácido alanina  
 <400> 67  
 gccgccgccg ccgctgctgc tgcagcagca 30  
 10  
 <210> 68  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 15  
 <220>  
 <223> motivo codónico del aminoácido alanina  
 20  
 <400> 68  
 gccgctgcag ccgctgcagc cgctgcagcc 30  
 25  
 <210> 69  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial  
 30  
 <220>  
 <223> motivo codónico del aminoácido arginina  
 <400> 69  
 aggagacggc gc 12  
 35  
 <210> 70  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial  
 40  
 <220>  
 <223> motivo codónico del aminoácido arginina  
 <400> 70  
 aggcggagac gc 12  
 45  
 <210> 71  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 50  
 <220>  
 <223> motivo codónico del aminoácido asparragina  
 <400> 71  
 aacaacaaca ataata 15  
 55  
 <210> 72  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 60  
 <220>  
 <223> motivo codónico del aminoácido asparragina  
 <400> 72  
 aacaataaca ataac 15  
 65

<210> 73  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <223> motivo codónico del aminoácido ácido aspártico  
 <400> 73  
 10 gacgacgacg atgat 15  
 <210> 74  
 <211> 15  
 <212> ADN  
 15 <213> Secuencia artificial  
 <220>  
 <223> motivo codónico del aminoácido ácido aspártico  
 20 <400> 74  
 gacgatgacg atgac 15  
 <210> 75  
 <211> 15  
 25 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> motivo codónico del aminoácido ácido cisteína  
 30 <400> 75  
 tgctgctgct gttgt 15  
 <210> 76  
 <211> 15  
 35 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> motivo codónico del aminoácido ácido cisteína  
 40 <400> 76  
 tgctgttgc gttgc 15  
 <210> 77  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 50 <220>  
 <223> motivo codónico del aminoácido glutamina  
 <400> 77  
 55 cagcagcagc agcaa 15  
 <210> 78  
 <211> 15  
 <212> ADN  
 <213> Secuencia artificial  
 60 <220>  
 <223> motivo codónico del aminoácido glutamina  
 <400> 78  
 65 cagcagcagc aacag 15

ES 2 599 386 T3

<210> 79  
<211> 15  
<212> ADN  
<213> Secuencia artificial  
5  
<220>  
<223> motivo codónico del aminoácido glutamina  
<400> 79  
cagcagcaac agcag 15  
10  
<210> 80  
<211> 15  
<212> ADN  
<213> Secuencia artificial  
15  
<220>  
<223> motivo codónico del aminoácido glutamina  
20  
<400> 80  
cagcaacagc agcag 15  
<210> 81  
<211> 9  
<212> ADN  
<213> Secuencia artificial  
25  
<220>  
<223> motivo codónico del aminoácido ácido glutámico  
30  
<400> 81  
gaggaggaa 9  
<210> 82  
<211> 9  
<212> ADN  
<213> Secuencia artificial  
35  
<220>  
<223> motivo codónico del aminoácido ácido glutámico  
40  
<400> 82  
gaggaagag 9  
45  
<210> 83  
<211> 63  
<212> ADN  
<213> Secuencia artificial  
50  
<220>  
<223> motivo codónico del aminoácido glicina  
<400> 83  
**ggcggcggcg gcggcggcgg cggaggagga ggaggagggg gggggggggg ggggtggtggt** 60  
**ggt** 63  
55  
<210> 84  
<211> 63  
<212> ADN  
<213> Secuencia artificial  
60  
<220>  
<223> motivo codónico del aminoácido glicina

ES 2 599 386 T3

	<400> 84	
	<b>ggcggagggg gtggcggagg cgggggtggc ggaggggtg gcgaggcgg gggcggaggg</b>	<b>60</b>
	<b>ggt</b>	<b>63</b>
5	<210> 85 <211> 15 <212> ADN <213> Secuencia artificial	
10	<220> <223> motivo codónico del aminoácido histidina	
	<400> 85 caccaccacc atcat 15	
15	<210> 86 <211> 15 <212> ADN <213> Secuencia artificial	
20	<220> <223> motivo codónico del aminoácido histidina	
	<400> 86 caccatcacc atcac 15	
25	<210> 87 <211> 60 <212> ADN <213> Secuencia artificial	
30	<220> <223> motivo codónico del aminoácido isoleucina	
	<400> 87 atcatcatca tcatcatcat catcatcatc attattatta ttattattat tataataata 60	
35	<210> 88 <211> 60 <212> ADN <213> Secuencia artificial	
40	<220> <223> motivo codónico del aminoácido isoleucina	
	<400> 88 atcattatca tcattataat cattatcatc attataatca ttatcatcat tataatcatt 60	
45	<210> 89 <211> 24 <212> ADN <213> Secuencia artificial	
50	<220> <223> motivo codónico del aminoácido leucina	
	<400> 89 ctgctgctgc tgctcctcct ttg 24	
55	<210> 90 <211> 24 <212> ADN <213> Secuencia artificial	
60	<210> 90 <211> 24 <212> ADN <213> Secuencia artificial	



<220>  
 <223> motivo codónico del aminoácido leucina  
  
 <400> 90  
 5 ctgctctgc ttctgctct gttg 24  
  
 <210> 91  
 <211> 24  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido leucina  
  
 <400> 91  
 15 ctgctccttt tgctgctct gctg 24  
  
 <210> 92  
 <211> 9  
 <212> ADN  
 20 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido lisina  
  
 <400> 92  
 25 aagaagaaa 9  
  
 <210> 93  
 <211> 9  
 <212> ADN  
 30 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido lisina  
  
 <400> 93  
 35 aagaaaaag 9  
  
 <210> 94  
 <211> 15  
 <212> ADN  
 40 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido fenilalanina  
  
 <400> 94  
 45 ttcttctct ttttt 15  
  
 <210> 95  
 <211> 15  
 <212> ADN  
 50 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido fenilalanina  
  
 <400> 95  
 55 ttcttttct ttttc 15  
  
 <210> 96  
 <211> 57  
 <212> ADN  
 60 <213> Secuencia artificial  
  
 <210> 96  
 <211> 57  
 <212> ADN  
 65 <213> Secuencia artificial

## ES 2 599 386 T3

<220>  
<223> motivo codónico del aminoácido prolina

<400> 96  
5 cccccccccc cccccccccc cccaccacca ccaccaccac ctctctctcc tcctctct 57

<210> 97  
<211> 57  
<212> ADN  
10 <213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido prolina

<400> 97  
15 cccccacctc cccccacctc cccacctccc ccacctcccc cacctcccc acctccc 57

<210> 98  
<211> 27  
20 <212> ADN  
<213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido serina

<400> 98  
25 tcctctcca gcagcagctc ttctca 27

<210> 99  
<211> 27  
30 <212> ADN  
<213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido serina

<400> 99  
35 tccagctctt catccagctc ttccagc 27

<210> 100  
<211> 27  
40 <212> ADN  
<213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido serina

<400> 100  
45 tccagctctt ccagctcctc ttccagc 27

<210> 101  
<211> 30  
50 <212> ADN  
<213> Secuencia artificial

<220>  
<223> motivo codónico del aminoácido treonina

<400> 101  
55 accaccacca ccaccacaac aacaactact 30

<210> 102  
<211> 30  
60 <212> ADN  
<213> Secuencia artificial

65

<220>  
 <223> motivo codónico del aminoácido treonina  
  
 <400> 102  
 5 accacaacta ccacaaccac aactaccacc 30  
  
 <210> 103  
 <211> 30  
 <212> ADN  
 10 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido treonina  
  
 15 <400> 103  
 accacaacca caactaccac cacaactacc 30  
  
 <210> 104  
 <211> 15  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido tirosina  
  
 25 <400> 104  
 tattattatt actac 15  
  
 <210> 105  
 <211> 15  
 30 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido tirosina  
  
 35 <400> 105  
 tattactatt actat 15  
  
 <210> 106  
 <211> 51  
 40 <212> ADN  
 <213> Secuencia artificial  
  
 45 <220>  
 <223> motivo codónico del aminoácido valina  
  
 <400> 106  
 50 gtggtggtgg tgggtggtg ggtggtcgtc gtcgtcgttg ttgtgtagt a 51  
  
 <210> 107  
 <211> 51  
 <212> ADN  
 55 <213> Secuencia artificial  
  
 <220>  
 <223> motivo codónico del aminoácido valina  
  
 <400> 107  
 60 gtggtggtcg ttgtagtgg ggtcgttggt gttggtcgtg tagtggtggt c 51  
  
 <210> 108  
 <211> 51  
 <212> ADN  
 65 <213> Secuencia artificial

<220>

<223> motivo codónico del aminoácido valina

<400> 108

5 gtggtcgttg tagtggcgt ttagtggtc gttgtggtcg tgggtggtg g 51

## REIVINDICACIONES

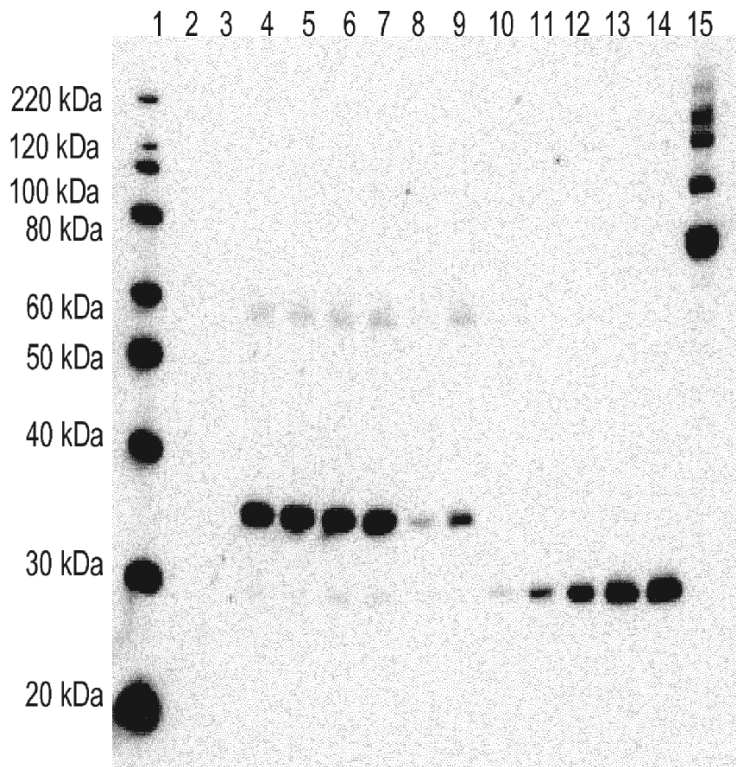
1. Un método para producir un polipéptido de manera recombinante en una célula bacteriana, que comprende la etapa de cultivar una célula bacteriana que comprende un ácido nucleico que codifica el polipéptido, y recuperar el polipéptido de la célula bacteriana o del medio de cultivo, en el que cada uno de los restos de aminoácido del polipéptido está codificado por al menos un codón, en el que el codón o los codones que codifican el mismo resto de aminoácido está combinado en un grupo y cada uno de los codones en un grupo está definido por una frecuencia de uso específica dentro del grupo, que es la frecuencia con la cual un solo codón de un grupo de codones puede encontrarse en un ácido nucleico que codifica un polipéptido en relación con todos los codones de un grupo, por lo cual la suma de las frecuencias de uso específicas de todos los codones en un grupo es de 100 %, en el que la frecuencia de uso total de cada codón en el genoma de la célula es aproximadamente la misma que su frecuencia de uso específica dentro de su grupo.
2. El método de acuerdo con la reivindicación 1, caracterizado por que los grupos comprenden solo codones con una frecuencia de uso total dentro del genoma de la célula de más de 5 %.
3. El método de acuerdo con la reivindicación 1, caracterizado por que los grupos comprenden solo codones con una frecuencia de uso total dentro del genoma de la célula de 8 % o más.
4. El método de acuerdo con la reivindicación 1, caracterizado por que los grupos comprenden solo codones con una frecuencia de uso total dentro del genoma de la célula de 10 % o más.
5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, caracterizado por que para cada aparición secuencial de un aminoácido específico en el polipéptido comenzando desde el extremo N del polipéptido en la posición correspondiente del ácido nucleico codificante el codón utilizado es idéntico al que está presente en la posición secuencial correspondiente en el motivo codónico de aminoácidos respectivo para el aminoácido específico, por lo cual después de la primera aparición del resto de aminoácido en la secuencia de aminoácidos del polipéptido, se usa el primer codón del motivo codónico de aminoácidos en el ácido nucleico codificante correspondiente, después de la segunda aparición del resto de aminoácido se usa el segundo codón del motivo codónico de aminoácidos y así sucesivamente.
6. El método de acuerdo con la reivindicación 5, caracterizado por que
- i) después del uso del codón final del motivo codónico de aminoácidos en la siguiente aparición del aminoácido específico en el polipéptido, el codón que está en la primera aposición del motivo codónico de aminoácidos respectivo se usa de nuevo en el ácido nucleico codificante correspondiente,
  - ii) para cada aparición secuencial adicional de este aminoácido específico en el polipéptido en la posición correspondiente del ácido nucleico codificante, el codón que se utiliza es el que está presente en la posición correspondiente en el motivo codónico de aminoácidos respectivo para el aminoácido específico.
7. El método de acuerdo con una cualquiera de las reivindicaciones 5 a 6, caracterizado por que los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente, por lo que después del codón con la frecuencia de uso específica más baja o del codón con la segunda frecuencia de uso específica más baja, se utiliza el codón con la frecuencia de uso específica más alta.
8. El método de acuerdo con la reivindicación 7, caracterizado por que los codones en el motivo codónico de aminoácidos se disponen con una frecuencia de uso específica decreciente, por lo que después del codón con la frecuencia de uso específica más baja, se utiliza el codón con la frecuencia de uso específica más alta.
9. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, caracterizado por que la célula es *E. coli*.
10. El método de acuerdo con la reivindicación 9, caracterizado por que el motivo codónico de aminoácidos para alanina se selecciona de SEQ ID NO: 01, 02, 03, 04 y 05, y/o arginina se selecciona de SEQ ID NO: 06 y 07, y/o asparraguina se selecciona de SEQ ID NO: 08, 09, 10, 11 y 12, y/o ácido aspártico se selecciona de SEQ ID NO: 13 y 14, y/o cisteína se selecciona de SEQ ID NO: 15, 16 y 17, y/o glutamina se selecciona de SEQ ID NO: 18, 19, 20 y 21, y/o ácido glutámico se selecciona de SEQ ID NO: 22, 23 y 24, y/o glicina se selecciona de SEQ ID NO: 25 y 26, y/o histidina se selecciona de SEQ ID NO: 27 y 28, y/o isoleucina se selecciona de SEQ ID NO: 29 y 30, y/o leucina se selecciona de SEQ ID NO: 31, 32 y 33, y/o lisina se selecciona de SEQ ID NO: 34, 35, 36 y 37, y/o fenilalanina se selecciona de SEQ ID NO: 38, 39 y 40, y/o prolina se selecciona de SEQ ID NO: 41, 42, 43, 44, 45 y 46, y/o

serina se selecciona de SEQ ID NO: 47 y 48, y/o  
treonina se selecciona de SEQ ID NO: 49, 50 y 51, y/o  
tirosina se selecciona de SEQ ID NO: 52 y 53, y/o  
valina se selecciona de SEQ ID NO: 54, 55 y 56.

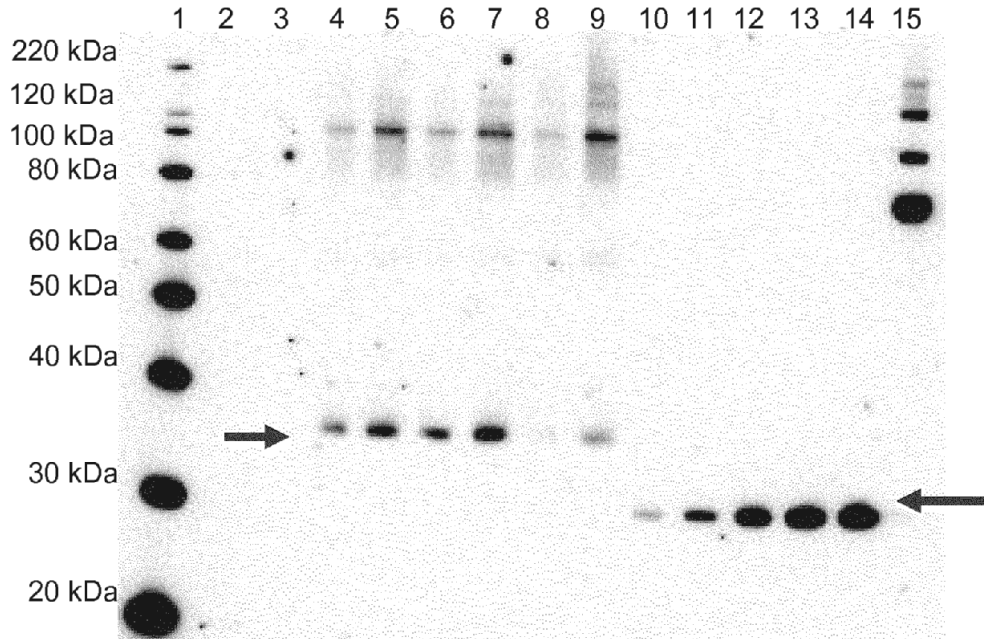
5

11. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 10, caracterizado por que el polipéptido es un anticuerpo, o un fragmento de anticuerpo, o un polipéptido de fusión de anticuerpo.

**Fig. 1**



**Fig. 2**





**Fig. 3**

