

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 750 525**

51 Int. Cl.:

C07K 14/34 (2006.01)

C12N 15/73 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.12.2013 PCT/EP2013/077968**

87 Fecha y número de publicación internacional: **03.07.2014 WO14102265**

96 Fecha de presentación y número de la solicitud europea: **24.12.2013 E 13815768 (0)**

97 Fecha y número de publicación de la concesión europea: **21.08.2019 EP 2938363**

54 Título: **Procedimientos y composiciones relacionados con CRM197**

30 Prioridad:

27.12.2012 US 201261746366 P

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

26.03.2020

73 Titular/es:

**GLAXOSMITHKLINE BIOLOGICALS S.A. (50.0%)
Rue de l'Institut, 89
1330 Rixensart, BE y
EIDGENÖSSISCHE MATERIALPRÜFUNGS- UND
FORSCHUNGSANSTALT (50.0%)**

72 Inventor/es:

**IHSEN, JULIAN;
KOWARIK, MICHAEL y
THÖNY-MEYER, LINDA CHRISTIANE**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 750 525 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

Procedimientos y composiciones relacionados con CRM197

1. Introducción

5 La presente invención proporciona nuevos procedimientos de producción de toxina diftérica. En particular, la presente invención proporciona nuevos procedimientos de producción de formas no tóxicas de toxina diftérica, por ejemplo, CRM197. La presente invención también proporciona nuevas composiciones que comprenden toxina diftérica o formas no tóxicas de toxina diftérica, por ejemplo, CRM197.

2. Antecedentes

10 La proteína CRM197 es un portador dependiente de células T seguro y eficaz para sacáridos y actualmente se está usando en muchas formulaciones de vacuna diferentes llamadas vacunas conjugadas. La toxina de la difteria es una proteína exotoxina producida por la bacteria *Corynebacterium diphtheriae* tras la infección con el fago β 197. Tanto la toxina de la difteria ("DT") como CRM197 son componentes de muchas vacunas, como por ejemplo contra *Bordetella pertussis*, *Clostridium tetani*, *C. diphtheriae*, virus de la hepatitis B y *Haemophilus influenzae* tipo B (documentos WO 9324148, WO 9700697, WO 02055105). Además, ha habido un creciente interés en CRM197 debido a su potencial actividad antitumoral relacionada con su capacidad para unirse a la forma soluble de HB-EGF (documento US 2006/0270600A1).

20 El CRM197 es producido por *C. diphtheriae* infectado por el fago no toxigénico β 197tox. β 197tox fue creado por la mutagénesis por nitrosoguanidina del corinefago toxigénico β (Uchida, T. et al. 1971, Nature New Biology 233:8-11). La proteína CRM197 es una forma no tóxica de la toxina de la difteria, pero es inmunológicamente indistinguible de la toxina de la difteria. DT tiene una masa de 58,350 kDa (CRM197 = 58,415 kDa) y consta de los dominios N-terminal A y C-terminal B (21 y 37 kDa) que están unidos por un puente disulfuro que conecta Cys186 y Cys201. El fragmento A es tóxico después de ser liberado de su socio unido por disulfuro, el fragmento B. El corte de la holotoxina por proteólisis leve en el péptido de conexión en las posiciones 191-3 es un requisito previo para la activación del fragmento A. El fragmento B no tiene actividad enzimática aparente, pero se requiere para la toxicidad, probablemente debido a que la holotoxina se dirige a las membranas celulares diana (Broker M, Costantino P, De Tora L, McIntosh ED, Rappuoli R: Biochemical and biological characteristics of cross-reacting material 197 (CRM197), a non-toxic mutant of diphtheria toxin: use as a conjugation protein in vaccines and other potential clinical applications. Biologicals, 2011, 39(4):195-204).

30 Los cultivos de *C. diphtheriae* infectados secretan la proteína CRM197 a través de la membrana citoplasmática fuera de la célula al medio de cultivo. La proteína CRM197 tiene aproximadamente el mismo peso molecular que la toxina de la difteria, pero difiere de ella por un cambio de base única (guanina a adenina) en el gen estructural. Este cambio de base única provoca una sustitución de aminoácidos (ácido glutámico para glicina, G52E) en la proteína madura y elimina las propiedades tóxicas de la toxina diftérica (Giannini G, Rappuoli R, Ratti G: The amino-acid sequence of two non-toxic mutants of diphtheria toxin: CRM45 y CRM197. Nucleic Acids Res 1984, 12(10):4063-4069).

35 Los procedimientos de preparación de DT y CRM197 se describen en los documentos US 4709017, US 5843711, US 5601827, y US 5917017. Actualmente hay tres sistemas diferentes usados para la preparación industrial de CRM197. Dos sistemas se basan en el uso de células de *C. diphtheriae* infectadas con fagos. El desarrollo más reciente constituye un sistema de expresión recombinante en *Pseudomonas fluorescens*. El procedimiento emplea un enfoque de secreción en el periplasma en una cepa de *P. fluorescens* genéticamente optimizada usando un gen CRM197 equipado con un péptido señal para secreción en el periplasma (documento US20110287443).

40 Por ejemplo, la toxina de la difteria se aísla de cultivos de *C. diphtheriae* cepa C7 (B197) y/o *C. diphtheriae* cepa C7 (B197) pPx350 cultivados en un medio a base de extracto de levadura y casaminoácidos en condiciones aeróbicas. Se demostró que el ajuste de los componentes del medio mejora los rendimientos (documentos US 4925792, WO 2006 100108). CRM197 o DT se cosechan del sobrenadante del cultivo y se concentran por ultrafiltración. La precipitación con sulfato de amonio es la primera, y la cromatografía de intercambio aniónico es una segunda etapa de purificación.

45 Sin embargo, la producción de cantidades significativas de la proteína CRM197 para su uso en vacunas se ha visto obstaculizada debido a la baja abundancia de proteínas (documento WO 2006 100108).

50 Se han desarrollado técnicas para impulsar la producción de proteínas CRM usando lisógenos dobles (Isolation and characterization of *C. diphtheriae* nontandem double lysogens hyperproducing CRM197. R Rappuoli, Appl. Environ. Microbiol. September 1983 46:560-564; U.S. Pat. No. 4,925,792 issued to R. Rappuoli; e Integration of corynebacteriophages beta tox+, omega tox+, and gamma tox- into two attachment sites on the *C. diphtheriae* chromosome. R Rappuoli, J L Michel, and J R Murphy; J. Bacteriol. March 1983 153:1202-1210) del corinefago no toxigénico β 197. Rappuoli informa rendimientos de CRM197 de lisógenos dobles y triples hasta tres veces más altos que los lisógenos individuales. Los niveles de producción de CRM197 por lisógenos individuales son adecuados pero económicamente insatisfactorios para la producción de vacunas que usan la proteína CRM197. Es importante

tener en cuenta que la construcción de cepas lisogénicas dobles y triples para aumentar la eficiencia de expresión en *C. diphtheriae* es un procedimiento largo que requiere una fase de cribado laboriosa.

5 Se desarrollaron plásmidos para la expresión recombinante de CRM197 en *C. diphtheriae* (las Patentes de los Estados Unidos US patent 5614382, 1995/5614382 _1997). Esto hace posible aumentar el número de copias del gen (hasta 5-10 por célula) sin tener que seleccionar cepas bacterianas pluri-lisogénicas.

Como en el caso de las cepas de *Corynebacterium* infectadas por el fago β 197tox, CRM197 se expresa en medios de cultivo especiales con un bajo contenido ferroso. A pesar de una reducción en el tiempo requerido para el manejo genético de la cepa bacteriana, la producción de CRM197 no aumenta dramáticamente en comparación con el uso de dobles lisógenos.

10 Las células huésped de expresión alternativa para DT incluyeron una cepa de vacuna *Salmonella typhi* cvd 908-htra (Orr N, Galen JE, Levine MM: Expression and immunogenicity of a mutant diphtheria toxin molecule, CRM197, and its fragments in *S. typhi* vaccine strain CVD 908-htrA. Infect Immun 1999, 67(8):4290-4294). *Salmonella* es una bacteria Gram negativa y un huésped de expresión similar a *E. coli*. Los niveles de expresión de diversas construcciones (con, sin péptido señal) en cvd 908-htra fueron bajos y la solubilidad e inmunogenicidad fueron
15 pobres. El uso del sistema alternativo de translocación dependiente no-Sec del operón de hemolisina mejoró la expresión de DT soluble, pero los niveles aún eran bajos.

Los informes para la producción de CRM197 en *E. coli* muestran bajos rendimientos de CRM197 soluble y formación de producto insoluble en cuerpos de inclusión. Los enfoques de truncamiento se han usado en un intento de mejorar la expresión a niveles superiores. (Bishai WR, Miyahara A, Murphy JR: Cloning and expression in *E. coli* of three
20 fragments of diphtheria toxin truncated within fragment B. Journal of Bacteriology 1987, 169(4):1554-1563)

Se usó un plásmido de expresión de cadena sencilla para CRM197 que contiene el gen de la toxina diftérica mutada que codifica CRM197 para la expresión en *E. coli* (Bishai WR, Rappuoli R, Murphy JR: High-level expression of a proteolytically sensitive diphtheria toxin fragment in *Escherichia coli*. Journal of Bacteriology 1987, 169(11):5140-5151; Bishai 1987). En esta publicación, la transcripción de CRM197 fue controlada por el promotor P_{tox} endógeno y constitutivo. Además, el DT fusionado en C-terminalmente con la hormona estimulante de melanocitos alfa ("ABM508") fue expresado por el promotor P_{lambda} inducible por calor o el promotor P_{tac} para la expresión.
25

Bishai 1987 especuló que el atasco del aparato de secreción debido a la inducción de proteínas de alto nivel causó una detención del crecimiento después de la inducción de la expresión de las variantes periplásmicas de DT/CRM197. Esto puede ser un problema general en la expresión de proteínas periplásmicas que se ha observado
30 y dado como resultado previamente bajos rendimientos volumétricos de proteínas (Benson SA, Hall MN, Silhavy TJ: Genetic analysis of protein export in *E. coli* K12. Annual Review of Biochemistry 1985, 54:101-134). El bloqueo del aparato de secreción y la formación de proteína insoluble sugirió una incapacidad de las células de *E. coli* para proporcionar un entorno productivo de translocación y plegamiento para la biogénesis CRM197.

Como consecuencia, Bishai 1987 razonó que la expresión citoplasmática evitaría el bloqueo del translocón. Por lo tanto, Bishai 1987 eliminó el péptido señal para dirigir la expresión al citoplasma. Solo a bajas temperaturas y cuando se eliminaron las proteasas citoplasmáticas, las construcciones de expresión citoplasmática produjeron producto soluble. La producción fue ineficiente y condujo a agregados a temperaturas elevadas, y cuando las proteasas estaban presentes.
35

Bishai 1987 no pudo mostrar la producción de altos niveles de proteína de fusión de proteína CRM197 soluble, es decir, con péptido señal para el direccionamiento periplásmico. En un gel de poliacrilamida SDS teñido con Coomassie, los extractos que contienen la construcción de expresión ABM508 mostraron una banda de proteína intensa correspondiente a ABM508, mientras que las células que expresan CRM197 expresadas desde el promotor natural con el péptido señal de tipo salvaje no muestran una banda obvia para CRM197 en el tamaño esperado de 58 kDa.
40

De este modo, la expresión periplásmica no se consideró una estrategia de producción eficiente y hasta la fecha no existe un sistema de expresión periplásmico de *E. coli* establecido para la producción de CRM197 o DT soluble y correctamente plegado.
45

El primer sistema de expresión basado en *E. coli* para CRM197 (documento WO2010 150230) proporciona un sistema de producción que se basa en la expresión citoplasmática de CRM197 insoluble en cuerpos de inclusión seguido de solubilización, purificación y replegamiento de la proteína. El CRM197 de tipo salvaje sin aminoácidos adicionales solo se puede obtener con este sistema cuando se aplica una etapa de proteólisis adicional.
50

Los péptidos de señal inducen la secreción de proteínas al periplasma y tienen diversos efectos sobre la biogénesis de proteínas. (Powers T, Walter P: Co-translational protein targeting catalyzed by the *E. coli* signal recognition particle and its receptor. The EMBO Journal 1997, 16(16):4880-4886.) (Schierle CF, Berkmen M, Huber D, Kumamoto C, Boyd D, Beckwith J: The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *E. coli* periplasm via the signal recognition particle pathway. Journal of Bacteriology 2003,
55

185(19):5706-5713). La expresión y purificación de CRM197 se ha divulgado en los documentos WO 11/42516 y Alessandra Stefan et al Journal of Biotechnology 156; 245-252 (2011).

3. Sumario

5 En este documento se proporciona un procedimiento de producción de CRM197 en el que el procedimiento comprende cultivar una célula de *E. coli* que comprende un ácido nucleico que codifica CRM197 en el que CRM197 se fusiona con un péptido señal heterólogo que se dirige a CRM197 al periplasma de la célula de *E. coli*, en el que el ácido nucleico codifica un sitio de escisión entre el péptido señal que dirige CRM197 al periplasma y la proteína CRM197 en el que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en el que aa1 es Ala, Ser, Gly, Cys, Thr o Gln; aa2 se selecciona de cualquier aminoácido natural; 10 aa3 se selecciona de cualquier aminoácido natural excepto Phe, His, Tyr, Trp, Asp, Glu, Lys, Arg, Asn y Gln; aa4 a 8 se selecciona de ala-asp-asp-val y met-gly-ala-asp; o en la que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que aa4 a 8 se selecciona de ala-asp-asp-val y met-gly-ala-asp; y en la que los primeros 70 aa del primer marco de lectura abierto dan como resultado una puntuación Y cuando es analizada por el servidor SignalP4,0 de más de 0,72. El procedimiento consiste en producir 15 CRM197 soluble, plegado, de longitud completa con altos rendimientos (por ejemplo, al menos 0,5 mg/1) en cepas de expresión de *E. coli*. En particular, se usa un péptido señal para dirigir la secreción de la proteína al espacio periplásmico.

20 En este documento se proporcionan procedimientos de producción de CRM197 en el que el procedimiento comprende cultivar una célula de *E. coli* que comprende un ácido nucleico que codifica CRM197 en el que CRM197 se fusiona con un péptido señal heterólogo que se dirige a CRM197 al periplasma de la célula de *E. coli*. En realizaciones más específicas, el péptido señal de tipo salvaje de CRM197 se ha eliminado. En realizaciones aún más específicas, el péptido señal de tipo salvaje de CRM197 ha sido reemplazado por el péptido señal heterólogo. El péptido señal heterólogo se puede seleccionar del grupo que consiste en el péptido señal de la enterotoxina termolábil de *E. coli*, porina de membrana externa de *E. coli* (OmpA), proteína de unión a maltosa de *E. coli* (MalE), 25 pectato liasa de *E. carotovorus* (PelB) y *Bacillus sp.* endoxilanasas (XynA). CRM197 se puede producir a una concentración de al menos 5, 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 300, 400, 500, 600, 700, 800, 900, o al menos 1000 mg de proteína por litro de medio de cultivo. Al menos el 50% de la proteína producida se pliega adecuadamente según lo determinado por el diroísmo circular. Al menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99,5%, 99,8%, o al menos 99,9% de la proteína producida está plegada 30 adecuadamente. Al menos el 50% de la proteína producida no está presente en los agregados. Al menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99,5%, 99,8%, o al menos 99,9% de la proteína producida no está presente en los agregados. Al menos el 50% de la proteína producida es soluble. Al menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99,5%, 99,8%, o al menos 99,9% de la proteína producida es soluble.

35 En ciertas realizaciones, la secuencia de nucleótidos heteróloga codifica un sitio de escisión entre el péptido señal que dirige CRM197 al periplasma y la proteína CRM197 en la que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que

aa1 se selecciona entre Ala, Ser, Gly, Cys, Thr, y Gln;

aa2 se selecciona de cualquier aminoácido natural;

40 aa3 se selecciona de cualquier aminoácido natural excepto Phe, His, Tyr, Trp, Asp, Glu, Lys, Arg, Asn, y Gln;

aa4 a 8 se selecciona entre ala-asp-asp-val y met-gly-ala-asp;

45 o en la que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que aa4 a 8 se selecciona entre ala-asp-asp-val y met-gly-ala-asp; y en la que los primeros 70 aa del marco de lectura abierto dan como resultado una puntuación Y cuando es analizado por SignalP 4,0 Server de más de 0,72.

50 En ciertas realizaciones específicas, la secuencia de nucleótidos heterólogos codifica la proteína de la SEQ ID NO: 1 o 2. La secuencia de nucleótidos heterólogos se puede unir operativamente a un promotor seleccionado del grupo que consiste en el promotor araBAD inducible por 1-arabinosa (PBAD), el promotor lac, el promotor rhaP BAD inducible por 1-ramnosa, el promotor de ARN polimerasa T7, el promotor trc y tac, el promotor de fago lambda p L y el promotor/operador tetA inducible por anhidrotetraciclina.

En ciertas realizaciones, el ácido nucleico que codifica CRM197 se inserta en un plásmido de expresión de copia alta. El plásmido de expresión de copia alta puede ser pEC415, pBR322, pBAD, serie pET, serie pUC, pACT3, pEXT22, pEXT20, serie pBLUESCRIPT, serie pGEM.

55 En ciertas realizaciones, la expresión de CRM197 se puede inducir a una densidad de cultivo de OD600 > 0,3. Específicamente, la expresión de CRM197 se puede inducir a una densidad de cultivo de OD600 >0,5, >1, o >1,5.

El CRM197 se puede expresar a una temperatura de 37 °C. CRM197 se puede expresar a una temperatura de 20, 25, 30, 32 o 35 °C.

- 5 En ciertas realizaciones, al menos el 50% de la proteína CRM197 que se ha producido de acuerdo con los procedimientos proporcionados en este documento tiene una terminación N de ADDV, GADDV o MGADDV. Más específicamente, al menos 60%, 70%, 80%, 90%, 95%, 98%, 99%, al menos 99,5% o 100% del CRM197 expresado tiene una terminación N de ADDV, GADDV, o MGADDV. En ciertas realizaciones, al menos el 50% del CRM197 expresado tienen un enlace disulfuro entre Cys186 y Cys201. Al menos 60%, 70%, 80%, 90%, 95%, 98%, 99%, al menos 99,5% o 100% del CRM197 expresado tienen un enlace disulfuro entre Cys186 y Cys201.

3.1 Terminología

- 10 La toxina DT-difteria

CRM197-Material reactivo cruzado 197, DT con una mutación de glicina 52 a glutamato, G52E

AMB508-Proteína de fusión que consiste en una secuencia de hormona estimulante de alfa melanocitos fusionada con la secuencia de CRM197 proteína-preproteína β 197tox-no-toxigénico corinefago 197

que incluye la proteína madura del péptido señal,

- 15 proteína procesada- proteína después de la escisión del péptido señal

IPTG-Isopropil- β -D-tiogalactopiranosid, inductor para lac, trc y promotores relacionados, que se agregarán a los medios de crecimiento en el momento de la inducción

Promotor de paraBAD del operón araBAD, inducible por la adición de L-arabinosa a los medios de crecimiento

4. Breve descripción de los dibujos

- 20 Figura 1 Expresión de CRM197 en *E. coli*. Se probaron diferentes plásmidos de expresión (1, 2, 3, 4, 5 que indican p932, p934, p722, una variante citoplasmática CRM197, y p150 (véase, la Tabla 2)) para la expresión de CRM197 en células *E. coli* BL21 (carriles indicados por A) o W3110 (B). Los extractos celulares totales normalizados a OD600 después de 2 horas de inducción se prepararon y analizaron. El panel superior muestra una transferencia Western usando antisuero anti DT, el panel inferior se detectó usando antisuero anti his tag.

- 25 Figura 2. Purificación de CRM197 soluble de *E. coli*. Se cultivaron dos cepas diferentes que contenían plásmidos de expresión (1, 2, que indican p932, p933) en BL21 como se describe en el texto. Las fracciones de elución se separaron mediante SDS PAGE y se tiñeron con azul Coomassie (panel izquierdo) y se inmunodetectaron después de la electrotransferencia a membranas de nitrocelulosa usando antisuero anti DT.

- 30 Figura 3. Niveles de expresión periplásmica de CRM197 fusionados a diferentes péptidos señal N-terminales en comparación con una proteína de referencia bien secretada EPA-6H, exotoxina A (EPA) recombinante, genéticamente desintoxicada de *Pseudomonas aeruginosa*, (Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thony-Meyer L: Production of glycoprotein vaccines in *E. coli*. Microbial cell factories 2010, 9:61). Gel SDS-PAGE teñido con Coomassie con extractos periplásmicos de cepas de *E. coli* que albergan los plásmidos 1 a 16 como se describe en la tabla 2. Las células se cultivaron en matraces de agitación a la temperatura de expresión óptima en medio TB y se indujeron a un OD₆₀₀ de 0,4-0,6 por la adición de 4 g/l de L-arabinosa. Se tomaron muestras equivalentes a OD para la extracción de proteínas periplásmicas solubles (procedimiento de sacarosa-lisozima) 4 h después de la inducción, con la excepción de 13*: muestra tomada antes de la inducción. M: mezcla de marcadores proteicos, escala de proteínas pre-teñidas Thermo-Scientific # 26616, 3 μ l (concentración de proteínas individuales 0,1-0,2 mg/ml).

40 5. Descripción detallada

En este documento se proporcionan procedimientos para la expresión de toxina diftérica y formas no tóxicas de toxina diftérica, por ejemplo, CRM197. Más específicamente, se proporcionan en este documento procedimientos para la expresión de CRM197 y la secreción de CRM197 en el espacio periplásmico de células de *E. coli* en una forma soluble y plegada correctamente.

- 45 El CRM197 se expresa usando un péptido señal heterólogo que se dirige al CRM197 en el espacio periplásmico de la célula huésped. En ciertas realizaciones más específicas, la célula huésped es *E. coli*. Se puede construir un casete de expresión que contiene el gen CRM197 y un péptido señal heterólogo usando técnicas estándar de biología molecular. Específicamente, el péptido señal de tipo salvaje de CRM197 se elimina y en su lugar se introduce un péptido señal heterólogo. A nivel de ácido nucleico, se debe tener cuidado de que la secuencia que codifica el péptido señal se clone en marco con el ácido nucleico que codifica el resto de CRM197. En ciertas realizaciones específicas, el péptido señal heterólogo reemplaza al péptido señal de tipo salvaje. En otras realizaciones, el péptido señal de tipo salvaje se elimina o se inactiva funcionalmente y el péptido señal heterólogo

se introduce en una ubicación diferente de la proteína. En ciertas realizaciones, se introduce un sitio de escisión proteolítica entre el péptido señal y el resto de la molécula. El sitio de escisión proteolítica se puede reconocer y escindir en el periplasma de la célula huésped. En ciertas realizaciones más específicas, la señal peptidasa se expresa de forma recombinante en la célula huésped.

- 5 En ciertas realizaciones, los siguientes parámetros pueden afectar la expresión de la proteína de interés. Se proporciona información más detallada sobre estos diversos aspectos en las siguientes secciones.

El ácido nucleico puede codificar el CRM197 maduro y secretado (SEQ ID NO: 6):

i) de manera optimizada para el uso de codones de *E. coli*.

- 10 ii) Se puede usar una secuencia de señal heteróloga para dirigir CRM197 al espacio periplásmico en *E. coli*. Mediante procedimientos de clonación estándar, las secuencias de ADN sintéticas que codifican un péptido señal heterólogo se pueden fusionar en la terminación N del gen de CRM197 maduro. Diferentes péptidos señal N-terminales, tal como a partir de la enterotoxina termolábil de *E. coli*, la porina A de membrana externa de *E. coli* (OmpA), la proteína de unión a maltosa de *E. coli* (MalE), *E. coli* DsbA, peptato liasa de *Erwinia carotovora*s (PelB) o endoxilanasas de *Bacillus sp.* (XynA) se puede usar con los procedimientos proporcionados en este documento. En ciertas realizaciones, se ha demostrado que un péptido señal heterólogo particular confiere secreción de proteínas recombinantes al espacio periplásmico de *E. coli*

- 20 iii) El sitio de escisión del péptido señal (esto es, la secuencia entre el péptido señal y la proteína secretada); Por ejemplo, los programas de predicción de escisión de péptidos señal, tal como por ejemplo el programa de servidor SignalP 4.0 (alojado en el sitio web del the Center for Biological Sequence Analysis of the Technical University of Denmark), se pueden usar para diseñar sitios de escisión de péptidos señal alternativos. Este programa predice i) la probabilidad del sitio de escisión, y ii) la ubicación del sitio de escisión, es decir, entre qué aminoácidos es más probable que se produzca la escisión. En realizaciones específicas, los sitios de escisión del péptido señal se diseñan de una manera que da como resultado una terminación N de CRM197 lo más similar posible a la terminación N natural.

- 25 iv) Se puede usar un plásmido de expresión de alto número de copias apropiado con los procedimientos proporcionados en este documento.

v) La expresión de CRM 197 se puede colocar bajo el control del promotor de arabinosa de inducción de alto nivel.

vi) El medio de crecimiento;

- 30 vii) El tiempo de expresión, es decir, el punto de inducción durante el crecimiento y el tiempo entre la inducción y la cosecha de las células;

viii) Cantidad de inductor;

ix) La temperatura de expresión;

- 35 En general, un procedimiento proporcionado en este documento se realiza como sigue. Primero, un plásmido de expresión como se describe en este documento se introduce en una célula huésped (por ejemplo, cepa de expresión de *E. coli* o *Salmonella sp.*). La mezcla de transformación se puede colocar en placas en medios enriquecidos complementados con el antibiótico para el cual el plásmido de expresión lleva un marcador de resistencia. Se puede usar una sola colonia para inocular un pequeño volumen de cultivo (por ejemplo, 5 ml) que consiste en, por ejemplo, medio TB o un medio rico similar que contiene glicerol como fuente de carbono y que carece o se complementa con el antibiótico apropiado. El cultivo se puede incubar luego entre 20-35 °C hasta la fase estacionaria y luego diluirse en medio fresco de composición idéntica o similar, precalentado a 20-35 °C, en una proporción de 1:50 a 1: 100. El cultivo fresco puede crecer hasta la fase de crecimiento exponencial (OD₆₀₀ de 0,6-1,2) y la expresión se induce mediante la adición del inductor apropiado, que depende del promotor usado en el plásmido de expresión. Los ejemplos de inductores incluyen arabinosa o un químico diferente de la condición física para la inducción de proteínas de alto nivel. Luego, la expresión continúa. En ciertas realizaciones, la expresión continúa antes de la cosecha durante al menos 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, o al menos 30 horas.

Se pueden tomar muestras en cualquier momento. En todos los puntos de tiempo, la cantidad de CRM197 soluble formado se evalúa mediante análisis a través de SDS-PAGE y tinción con Coomassie de extractos periplásmicos. Una comparación con el CRM197 comercial de concentración conocida permite estimar el rendimiento.

- 50 En ciertas realizaciones, el siguiente control se usa para determinar que la proteína CRM197 expresada es soluble. Sin estar limitado por la teoría, dicha proteína soluble está plegada correctamente. CRM197 detectado a alrededor de 58 kDa en extractos periplásmicos preparados por el procedimiento de sacarosa-lisozima (Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF et al: Definition of the

bacterial N-glycosylation site consensus sequence. The EMBO journal 2006, 25(9):1957-1966) se puede usar como un estándar para la proteína soluble.

5 En ciertas realizaciones, en este documento se proporcionan procedimientos que dan como resultado una concentración de CRM197 plegado adecuadamente de al menos 1 mg/l, 2 mg/l, 3 mg/l, 4 mg/l, 5 mg/l, 6 mg/l, 7 mg/l, 8 mg/l, 9 mg/l, 10 mg/l, 11 mg/l, 12 mg/l, 13 mg/l, 14 mg/l, 15 mg/l, 20 mg/l, 25 mg/l, 50 mg/l, 75 mg/l, o al menos 100 mg/l.

En ciertas realizaciones, dependiendo de la secuencia del sitio de escisión del péptido señal, la terminación N de CRM197 puede ser ADDV o MGADDV.

10 En ciertas realizaciones, al menos el 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, o 100% de las proteínas CRM197 resultantes tienen un disulfuro enlace entre Cys186 y Cys201 conectando los fragmentos A y B. La presencia de este enlace disulfuro se puede demostrar usando un ensayo de tior (Hansen RE, Ostergaard H, Norgaard P, Winther JR: Quantification of protein thiols and dithiols in the picomolar range using sodium borohydride and 4,4'-dithiodipyridine. Anal Biochem 2007, 363(1):77-82).

15 En ciertas realizaciones, al menos el 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, o 100% de las proteínas CRM197 resultantes son solubles y no están presentes en los agregados.

5.1 Proteínas de interés

La presente invención describe procedimientos de producción para CRM197. Las secuencias de proteínas ilustrativas para CRM197 se proporcionan como SEQ ID NOS: 3 y 4, y las secuencias de ADN del plásmido de expresión de longitud completa correspondientes son SEQ ID NOS: 1 y 2.

5.2 Orientación periplásmica

20 En *E. coli* existen diversas rutas secretoras desde el citoplasma hasta el periplasma. Estas vías incluyen la vía Sec, la vía dependiente de SRP y la vía de la arginina gemela para la secreción. Se cree que el poro TatABC es responsable de la secreción de proteínas plegadas. Sin estar limitado por la teoría, el péptido señal puede determinar qué vía secretora es elegida por la célula. (Driessen AJ, Nouwen N: Protein translocation across the bacterial cytoplasmic membrane. Annual review of biochemistry 2008, 77:643-667.)

25 En ciertas realizaciones, el péptido señal heterólogo para usar con los presentes procedimientos es una secuencia señal hidrófoba amino terminal que se escinde durante el procedimiento de translocación. En ciertas realizaciones, las condiciones para un procedimiento proporcionado en este documento se eligen de manera que la proteína no se pliegue en su estructura tridimensional estable en el citoplasma de la célula huésped. Sin estar limitado por la teoría, el plegado en el citoplasma puede evitar la exportación. En ciertas otras realizaciones, el péptido señal heterólogo codifica una firma de arginina que dirige la proteína a la ruta de la arginina gemela para la secreción.

30 En ciertas realizaciones, una proteína desplegada se puede mantener en un estado competente de exportación de varias maneras diferentes: (i) el péptido señal heterólogo se puede elegir de modo que la proteína se pueda translocar a través de una membrana simultáneamente con la traducción de la proteína, asegurando de este modo que ni siquiera sus estructuras secundarias se forman en el citoplasma debido a la ausencia de polímero de aminoácidos; (ii) chaperonas o factores antiplegamiento que evitan el plegamiento en el citoplasma (Randall LL, Topping TB, Smith VF, Diamond DL, Hardy SJ: SecB: a chaperone from *E. coli*. Methods Enzymol 1998, 290:444-459.) se pueden proporcionar; (iii) las secuencias de señal heterólogas se eligen y/o insertan de manera que actúen como chaperonas intrapolipéptidas para evitar el plegamiento rápido; y/o (iv) el DT o CRM197 se modifica de modo que contenga características en su estructura final (por ejemplo, enlaces disulfuro) que no se forman en el entorno del citoplasma de modo que las proteínas no puedan alcanzar sus conformaciones plegadas finales en el citoplasma.

5.2.1 Péptidos de señal

45 Los péptidos señal heterólogos ilustrativos que se pueden usar con los procedimientos proporcionados en este documento son: la secuencia señal de *E. coli* DsbA, los péptidos señal MalE, OmpA y PelB. Sin estar limitado por la teoría, la elección del péptido señal puede determinar la ruta de secreción, por ejemplo, la ruta dependiente de SRP frente a la dependiente de SecB hacia el translocón. Las condiciones de expresión óptimas pueden diferir para diferentes rutas de focalización. Hay informes que afirman que las tecnologías que permiten la identificación de la vía de secreción dirigida (Marrichi M, Camacho L, Russell DG, DeLisa MP: Genetic toggling of alkaline phosphatase folding reveals signal peptides for all major modes of transport across the inner membrane of bacteria. J Biol Chem 2008, 283(50):35223-35235).

55 Los péptidos señal preferidos se seleccionan de proteínas secretadas conocidas y predichas que se exportan eficientemente al periplasma de *E. coli* a través de rutas cotraduccionales. Entre otros, se pueden usar los péptidos señal de enterotoxina termolábil de *E. coli*, porina A de membrana externa de *E. coli* (OmpA), proteína de unión de maltosa de *E. coli* (MalE), peptato liasa de *E. carotovorus* (PelB), o endoxilanasas de *Bacillus sp.* (XynA).

5.2.2 Sitios de escisión

Sin estar limitados por la teoría, los péptidos señal se escinden de la preproteína mediante una peptidasa señal, y en *E. coli* hay SPasel y II. SPasel está escindiendo algunos péptidos señal de proteínas de membrana y más solubles, mientras que SPasel escinde péptidos señal de lipoproteínas. SPasel es la peptidasa señal responsable en la presente invención. Se pudo determinar el uso de SPasel (Paetzel M, Karla A, Strynadka NC, Dalbey RE: Signal peptidases. Chemical reviews 2002, 102(12):4549-4580).

Sin estar limitados por la teoría, las ubicaciones del sitio de escisión se definen por i) la estructura de la organización del péptido señal característico con un núcleo hidrófobo, una terminación N cargada y una terminación C hidrófila, y ii) por la secuencia primaria alrededor de la posición de escisión (a menudo A-X-A) (Heijne G: The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. The EMBO journal 1986, 5(11):3021-3027.). Ambos parámetros son bien entendidos y los programas de predicción tienen una alta precisión (Petersen TN, Brunak S, von Heijne G, Nielsen H: SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011, 8(10):785-786). El programa SignalP 4.0 server proporciona una probabilidad de escisión basada en la secuencia de los primeros 70 aminoácidos de la preproteína. En ciertas realizaciones, los sitios de escisión diseñados para su uso con los procedimientos proporcionados en este documento tienen una puntuación Y de al menos 0,4, 0,5, 0,6, 0,7, 0,72, 0,75, 0,8, 0,85, 0,9, o al menos 0,95.

En ciertas realizaciones, el sitio de escisión del péptido señal está diseñado de tal manera que la terminación N prevista es la terminación N de la proteína que existe naturalmente. En otras realizaciones, los péptidos señal y la terminación N del CRM197 están diseñados de tal manera que la terminación N esté lo más cerca posible de la terminación N encontrado de forma nativa.

La terminación N natural después de la escisión del péptido señal de la proteína CRM197 es GADDV... (Bell CE, Eisenberg D: Crystal structure of nucleotide-free diphtheria toxin. Biochemistry 1997, 36(3):481-488). En ciertos aspectos de la divulgación, la terminación N de CRM197 expresado en *E. coli* usando el péptido señal DsbA puede ser:

MKKIWLALAGLVLAFA**ASA**-(escisión)-ADDVVDSSK... y usando el péptido señal PeIB

MKKIWLALAGLVLAFA**AMA**-(escisión)-GADDVVDSSKS...

Tenga en cuenta el motivo AXA en el sitio de escisión, donde la escisión tiene lugar después del segundo A.

Otras secuencias de escisión y combinaciones de sitios de escisión de péptido señal se exponen en la Tabla 2 a continuación.

5.3 Plásmidos de expresión

Se conoce una amplia variedad de vectores de expresión para la expresión recombinante en células de *E. coli*. En principio, se puede usar cualquier esqueleto de vectores. Los vectores ilustrativos son: pEC415 (Schulz H, Hennecke H, Thony-Meyer L: Prototype of a heme chaperone essential for cytochrome c maturation. Science 1998, 281(5380): 1197-1200), pBR322 (Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, Crosa JH, Falkow S: Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 1977, 2(2):95-113), pBAD (Invitrogen corporation, Carlsbad, CA), serie pET (Invitrogen), serie pUC (Lin-Chao S, Chen WT, Wong TT: High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. Mol Microbiol 1992, 6(22):3385-3393), pACT3, pEXT22, pEXT20 (Dykxhoorn DM, St Pierre R, Linn T: A set of compatible tac promoter expression vectors. Gene 1996, 177(1-2):133-136.), serie pBLUESCRIPT (Stratagene, Agilent Technologies, Santa Clara, CA), serie pGEM (Promega Corp., Madison, WI). Todos estos vectores podrían usarse para clonar el casete de expresión de la preproteína bajo el control de un promotor inducible.

Se proporcionan plásmidos ilustrativos como SEQ ID NOs: 1 y 2. El esqueleto del vector se basa en pBR322 que contiene un origen de replicación pMB1 de copia media a alta, un casete de resistencia a ampicilina que se puede intercambiar por un casete de kanamicina, el regulón del operón araBAD que codifica el represor AraC y el promotor araBAD para la inducción de expresión de proteínas de alto nivel.

En ciertas realizaciones, CRM197 se expresa a partir de construcciones integradas cromosómicamente. Esta estrategia requiere tecnologías adicionales que son bien conocidas para los expertos en el arte y darían como resultado una construcción de expresión integrada al genoma que consta de los mismos elementos que un plásmido de expresión pero que no requiere el casete de selección (solo para la selección tras la integración genómica) y origen de replicación.

5.4 Promotores

Entre los promotores inducibles de alta expresión bien conocidos, se puede usar cualquiera que sea funcional a la temperatura para la expresión de la proteína de interés. En ciertas realizaciones, un promotor que se usará con los procedimientos proporcionados en este documento es activo por debajo de la temperatura de 37 °C, por debajo de

36 °C, 35 °C, 34 °C, 33 °C, 32 °C, 31 °C, o por debajo de 30 °C. La siguiente lista contiene promotores ilustrativos de expresión bacteriana que se pueden usar con los procedimientos proporcionados en este documento (Tabla 1):

Tabla 1: Promotores inducibles usados en la expresión bacteriana (Fuente: sitio web del: The Wolfson Centre for Applied Structural Biology of the Hebrew University of Jerusalem)

Promotor	Fuente	Regulación	Inducción	Nivel de expresión	Información adicional
lac	<i>E. coli</i>	lacI, lacI ^{q*}	IPTG	bajo	
lacUV5	<i>E. coli</i>	lacI, lacI ^{q*}	IPTG	bajo	Teóricamente no está sujeto a la regulación dependiente de AMPc
tac (híbrido)	<i>E. coli</i>	lacI, lacI ^{q*}	IPTG	Permite la acumulación de proteínas en aproximadamente el 15-30% de la proteína celular total	Consiste en la región -35 del promotor <i>trp</i> y la región -10 del promotor <i>lac</i> (difiere del promotor <i>trc</i> en 1 pb)
trc (híbrido)	<i>E. coli</i>	lacI, lacI ^{q*}	IPTG	Permite la acumulación de proteínas en aproximadamente el 15-30% de la proteína celular total	Consiste en la región -35 del promotor <i>trp</i> y la región -10 del promotor <i>lac</i> (difiere del promotor <i>tac</i> en 1 pb)
trp	<i>E. coli</i>	La adición de fructosa al medio de crecimiento aumenta la regulación descendente en condiciones no inducidas.	La inanición por triptófano o la adición de ácido B-indolecrílico		
araBAD	<i>E. coli</i>	araC	1-arabinosa	Más débil que el promotor <i>tac</i>	Existe una gran heterogeneidad en las poblaciones celulares tratadas con concentraciones subsaturadas de 1-arabinosa (algunas bacterias son completamente inducidas y otras no).
phoA	<i>E. coli</i>	phoB (positivo) phoR (negativo)	Inanición de fosfato		Estrechamente controlada. La inducción requiere inanición de fosfato, por lo que puede limitar la duración de la síntesis de proteínas.
recA	<i>E. coli</i>	lexA	Ácido nalidíxico		
proU	<i>E. coli</i>		Omolaridad		
cst-1	<i>E. coli</i>		Inanición de glucosa		

(continuación)

Promotor	Fuente	Regulación	Inducción	Nivel de expresión	Información adicional
tetA	<i>E. coli</i>		tetraciclina		
cadA	<i>E. coli</i>	cadR	pH		
nar	<i>E. coli</i>	fhr	Condiciones anaeróbicas		
cspA	<i>E. coli</i>		Choque térmico en frío (cambio a menos de 20 °C)		El promotor central de cspA solo es débilmente inducido por el cambio descendente de temperatura. Una región no traducida de 159 nucleótidos de largo en la terminación 5' de las transcripciones impulsadas por cspA los hace altamente inestables a 37 °C y aumenta significativamente su estabilidad a bajas temperaturas. Esta región también favorece su participación mediante maquinaria de traducción modificada en frío. El sistema cspA se reprime 1-2 horas después del cambio descendente de temperatura.
SP6	Fago de <i>Salmonella</i>				
T7	Fago T7	clts857	térmica		
Operador T7-lac	Fago T7	lacI ^q *	IPTG	Permite la acumulación de proteínas en aproximadamente el 40-50% de la proteína celular total	
Operador T3-lac	Fago T3	lacI ^q *	IPTG		
Operador T5-lac	Fago T5	lacI, lacI ^q *	IPTG		Este promotor es reconocido por la ARN polimerasa de <i>E. coli</i>
Gen 32 de T4	Fago T4		Infección T4		
Operador nprM-lac	<i>Bacillus</i>	lacI ^q *	IPTG		
VHb	<i>Vitreoscilla</i>		oxígeno		

5.5 Medio de cultivo

5 El medio de cultivo para la producción de proteínas puede ser cualquier medio definido, semidefinido o complejo apropiado para la sobreexpresión de proteínas recombinantes en *E. coli*. Se prefiere un medio complejo rico como el caldo Terrific (TB), pero también se pueden usar medios de sales minerales definidos. El caldo Terrific está compuesto de 24 g/l de extracto de levadura, 12 g/l de triptona o peptona (es decir, caseína digerida proteolíticamente, proteína de soja u otra proteína) y 4% (v/v) de glicerol. Además, el medio está tamponado.

En ciertas realizaciones específicas, la concentración de iones de magnesio es como máximo 10nM, 50nM, 100nM, 250nM, 500nM, 750nM, o como máximo 1mM. En ciertas realizaciones específicas, no se agrega magnesio. En ciertas realizaciones específicas, no se agrega MgCl₂ al medio de cultivo.

10 En ciertas realizaciones específicas, el pH del medio de cultivo está entre 6 y 9. En ciertas condiciones específicas, el extracto de levadura puede estar presente en el medio de cultivo a una concentración entre 10-30 g/l. En ciertas realizaciones específicas, el medio de cultivo comprende glicerol del 2,5% al 10%. En ciertas otras realizaciones, el medio de cultivo comprende glicerol al menos 5%, 10%, 15% o al menos 20%.

5.6 Inducción y expresión

15 Los cultivos de expresión antes de la inducción se pueden cultivar a diferentes temperaturas, por ejemplo, temperaturas que oscilan entre 4-35 °C o 18-37 °C. En ciertas realizaciones, los cultivos de expresión antes de la inducción se cultivan a una temperatura dentro del intervalo de 18-20 °C, 20-22 °C, 22-24 °C, 24-26 °C, 26-28 °C, 28-30 °C, 30-32 °C, 32-34 °C, o 34-36 °C. En ciertas realizaciones, los cultivos de expresión antes de la inducción se cultivan a una temperatura de aproximadamente 18 °C, 19 °C, 20 °C, 21 °C, 22 °C, 23 °C, 24 °C, 25 °C, 26 °C, 27 °C, 28 °C, 29 °C, 30 °C, 31 °C, 32 °C, 33 °C, 34 °C, 35 °C, 36 °C, o 37 °C.

20 Las temperaturas de cultivo después de la inducción pueden caer en ciertos intervalos, por ejemplo, temperaturas que oscilan entre 4-35 °C o 18-37 °C, y pueden ser diferentes de las condiciones de inducción anteriores. Por ejemplo, un cultivo de preinducción se puede cultivar a temperaturas más altas, por ejemplo, una temperatura descrita anteriormente, y luego cambiarse a una temperatura más baja, por ejemplo, una temperatura en el intervalo de 15-30 °C, para la producción. En ciertas realizaciones, los cultivos después de la inducción se cultivan a una temperatura dentro del intervalo de 18-20 °C, 20-22 °C, 22-24 °C, 24-26 °C, 26-28 °C, 28-30 °C, 30-32 °C, 32-34 °C, o 34-36 °C. En una realización específica, dicha temperatura cae dentro de un intervalo que es inferior al intervalo en el que se cultiva el cultivo de preinducción. En ciertas realizaciones, los cultivos después de la inducción se cultivan a una temperatura de aproximadamente 18 °C, 19 °C, 20 °C, 21 °C, 22 °C, 23 °C, 24 °C, 25 °C, 26 °C, 27 °C, 28 °C, 29 °C, 30 °C, 31 °C, 32 °C, 33 °C, 34 °C, 35 °C, 36 °C, o 37 °C. En una realización específica, dicha temperatura cae dentro de un intervalo que es inferior al intervalo en el que se cultiva el cultivo de preinducción.

35 Dependiendo de la construcción, el tiempo de expresión puede ser de 2-20 h. Las concentraciones de inductor son, dependiendo del promotor, de 0,01 a 1% (p/v) de arabinosa (ParaBAD), o de 10 a 1000 μM de IPTG. La inducción se puede realizar a valores de OD₆₀₀ obtenidos durante la fermentación entre 0,3 a 1,5 en cultivos en matraces de agitación, y a OD₆₀₀ entre 5 a 200 en fermentaciones del biorreactor. En ciertas realizaciones específicas, la inducción se realiza a una OD₆₀₀ entre 5 y 50, 25 y 75, 50 y 100, 75 y 125, 100 y 150, 125 y 175, 150 y 200, o 175 y 200. En ciertas realizaciones, la inducción se realiza al comienzo de la fase log en el matraz de agitación. Las fermentaciones del biorreactor se pueden realizar a valores constantes de pO₂ que oscilan desde 0% a 40%. La regulación de pO₂ se puede realizar regulando la velocidad del agitador o la velocidad de aireación.

40 En ciertas realizaciones, el promotor es inducible con arabinosa; las concentraciones de arabinosa pueden ser al menos 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, o al menos 1 % (p/v) arabinosa. En ciertas realizaciones, la concentración del inductor arabinosa es como máximo 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, o como máximo 1 % (p/v) arabinosa

45 En ciertas realizaciones, el promotor es inducible con IPTG; las concentraciones de IPTG pueden ser al menos 10, 25, 50, 75, 100, 200, 250, 300, 400, 500, 600, 700, 750, 800, 900 o al menos 1000 μM de IPTG. En ciertas realizaciones, la concentración del inductor IPTG es como máximo 10, 25, 50, 75, 100, 200, 250, 300, 400, 500, 600, 700, 750, 800, 900 o como máximo 1000 μM de IPTG.

50 En ciertas realizaciones, la expresión se realiza en cultivos en matraces de agitación. Los valores de OD₆₀₀ en el momento de la inducción son al menos 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,1, 1,2, 1,3, 1,4, o al menos 1,5 en cultivos en matraces de agitación. En ciertas realizaciones, los valores de OD₆₀₀ en el momento de inducción son como máximo 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,1, 1,2, 1,3, 1,4, o como máximo 1,5 en cultivos en matraces de agitación.

55 En ciertas realizaciones, la expresión se realiza en fermentaciones del biorreactor. Los valores de OD₆₀₀ en el momento de la inducción son al menos 5, 10, 15, 20, 25, 50, 75 o al menos 100 en fermentaciones del biorreactor. En ciertas realizaciones, los valores de OD₆₀₀ en el momento de la inducción son como máximo 5, 10, 15, 20, 25, 50, 75 o como máximo 100 en fermentaciones del biorreactor.

Las fermentaciones del biorreactor se pueden realizar a valores constantes de pO₂ de al menos 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, o al menos 40%. En ciertas realizaciones, las fermentaciones del biorreactor se pueden realizar a valores constantes de pO₂ de como máximo 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, o como máximo 40%.

5 5.7 Células huésped

Las cepas de expresión para la producción recombinante de la proteína diana pueden ser, pero no se limitan a, cepas de *E. coli* K12 y B, como W3110, DB1, DH5a, BL21, BL21(DE3), C43, JM109, JM101, JM110, y derivados de los mismos (Huang CJ, Lin H, Yang X: Industrial production of recombinant therapeutics in *E. coli* and its recent advancements. Journal of Industrial Microbiology & Biotechnology 2012, 39(3):383-399). Las células huésped se pueden modificar cromosómicamente para acomodar la expresión óptima de la proteína CRM197. Por ejemplo, las proteasas periplásmicas como DepP, Prc, Spr y/o proteasa III pueden eliminarse en cepas de producción. Las deleciones pueden ser útiles solas o en combinación con otras proteasas. Además, mutaciones supresoras como, por ejemplo, sprW148R (Chen C, Snedecor B, Nishihara JC, Joly JC, McFarland N, Andersen DC, Battersby JE, Champion KM: High-level accumulation of a recombinant antibody fragment in the periplasm of *E. coli* requires a triple-mutant (degP prc spr) host strain. Biotechnology and bioengineering 2004, 85(5):463-474.) pueden aumentar el rendimiento de la proteína CRM197.

5.8 Ensayos

Los procedimientos para caracterizar el rendimiento, la pureza, la estabilidad, el grado de corte, la toxicidad y el contenido de endotoxinas están bien establecidos y definen la calidad para el uso de CRM197 en una vacuna. El análisis de CRM197 se realiza mediante, por ejemplo, cromatografía de exclusión por tamaño de alto rendimiento, enfoque isoeléctrico, SDS-PAGE y Western Blot, determinación de peso molecular por MS, secuenciación de terminal N, análisis de aminoácidos, cromatografía líquida de fase inversa, espectroscopía de masas por electroaspersión y mapeo de péptidos por espectroscopía de masas después de la digestión triptica.

Se describen procedimientos analíticos y los parámetros que definen una calidad aceptable están bien establecidos para su uso en medicamentos. Se proporcionan información detallada y parámetros marco, por ejemplo, en las directrices publicadas por la agencia médica europea, EMEA, y se pueden encontrar en el sitio web de EMEA, por ejemplo, para la vacuna Prevenar que contiene CRM197.

5.8.1 Concentración del producto de expresión

Se podrían usar tecnologías de concentración de proteínas estándar como el ensayo Lowry, el ensayo BCA y los ensayos Bradford, así como la determinación de la absorción UV a 280 nm y la cuantificación de geles SDS-PAGE teñidos con Coomassie por densitometría o electroforesis en gel capilar por mediciones de intensidad de tinte fluorescente.

5.8.2 Plegado del producto de expresión

El plegamiento del producto se puede analizar directamente mediante espectroscopía de dicroísmo circular, espectroscopía de RMN de proteínas y HPSEC. Los procedimientos indirectos incluyen la medición de la solubilidad, la resistencia a la proteasa y los ensayos de actividad para determinar la toxicidad en el caso del fragmento DT A, y los ensayos de unión para los fragmentos DT B y CRM197.

5.8.3 Cuerpos de inclusión del producto de expresión

La formación del cuerpo de inclusión se cuantifica fácilmente mediante la primera homogeneización de las células cosechadas después de la fermentación, centrifugación de baja rotación para la sedimentación de la materia insoluble, y comparando las pellas y el sobrenadante lado a lado de una manera de densidad óptica equivalente. La intensidad de la banda de proteínas permite la estimación de la proporción en el sobrenadante (proteína soluble) y las pellas (agregados insolubles y cuerpos de inclusión).

5.8.4 Solubilidad del producto de expresión

La solución sobrenadante que contiene la proteína se puede centrifugar y filtrar estéril. Si la proteína permanece en solución y no se agota del filtrado y los sobrenadantes, la proteína es soluble. Un procedimiento más sofisticado es la dispersión dinámica de la luz. Permite la determinación del tamaño de partícula, que es indicativo del estado oligomérico o micro agregado de la proteína purificada.

La solubilidad es inversamente proporcional a la formación de agregados de tal manera que un hallazgo de alta solubilidad demuestra que no hay o tiene un bajo nivel de formación de agregados.

5.8.5 Localización periplásmica del producto de expresión

La localización periplásmica se mide mediante el fraccionamiento de las células y comparando los rendimientos proteicos específicos observados en las fracciones de periplasma y esferoplasto. El fraccionamiento se realiza

usando el procedimiento de sacarosa-lisozima (Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF et al: Definition of the bacterial N-glycosylation site consensus sequence. The EMBO journal 2006, 25(9):1957-1966), choque osmótico (Johansson HJ, Jagersten C, Shiloach J: Large scale recovery and purification of periplasmic recombinant protein from *E. coli* using expanded bed adsorption chromatography followed by new ion exchange media. J Biotechnol 1996, 48(1-2):9-14.), o polimixina (Schulz H, Hennecke H, Thony-Meyer L: Prototype of a heme chaperone essential for cytochrome c maturation. Science 1998, 281(5380): 1197-1200). Las alícuotas de las fracciones se normalizan en función de los volúmenes de muestra y el cultivo OD₆₀₀ y se analizan mediante SDS PAGE y transferencia Western.

5.8.6 Escisión de secuencia de señal

- 10 La escisión de los péptidos señal se analiza mediante i) análisis de desplazamiento de gel de células fraccionadas como se describe en 5.8.5. En este análisis, la preproteína no procesada se puede acumular en agregados citoplasmáticos o membranas, y la proteína procesada estará presente en fracciones periplásmicas, solubles. La movilidad electroforética diferente constituirá un desplazamiento entre preproteínas procesadas y no procesadas por SDS-PAGE (y transferencia Western si es necesario). Finalmente, la secuenciación de aminoácidos N-terminal se puede usar para determinar la terminación N procesado y, por lo tanto, definir el sitio de escisión experimentalmente.

5.8.7 Toxicidad potencial de CRM197

Se puede analizar la presencia de toxina activa en CRM197 midiendo la actividad ADP-ribosil transferasa. Además, se pueden usar otras pruebas (citotoxicidad en células HeLa o células Vero in vitro, letalidad en cobayas in vivo, prueba de toxicidad anormal) para demostrar la no toxicidad de CRM197

20 5.9 Composiciones

- El CRM197, producido de acuerdo con los procedimientos proporcionados en este documento, puede procesarse adicionalmente para obtener composiciones inmunogénicas o vacunas. Por ejemplo, la proteína se puede conjugar con un oligosacárido o un polisacárido para producir una composición o vacuna inmunogénica. En ciertas realizaciones específicas, dicha composición inmunogénica o vacuna tiene propiedades inmunogénicas mejoradas sobre las composiciones de la técnica anterior. Sin estar limitados por la teoría, los procedimientos proporcionados en este documento proporcionan una población más homogénea de proteína CRM197 soluble. Como tal, cualquier composición inmunogénica o vacuna es más efectiva que las composiciones de la técnica anterior.

- En ciertas realizaciones, se proporciona en este documento una composición que comprende CRM197, que se ha producido de acuerdo con los procedimientos proporcionados en este documento. En ciertas realizaciones más específicas, dicha composición es una composición farmacéutica. Incluso más específicamente, dicha composición farmacéutica comprende además y un portador farmacéuticamente aceptable.

6. Ejemplos

6.1 Ejemplo 1

- 35 Se probaron diferentes configuraciones experimentales y se determinó el rendimiento de CRM197 mediante transferencia Western usando antisuero toxina anti diftérica para la detección de CRM197.

- Un proveedor comercial (Genescript, Piscataway, NJ) sintetizó un marco de lectura abierto de ADN para la expresión de CRM197 en una forma optimizada de codón que contiene el péptido señal N-terminal de la proteína DsbA de *E. coli* en lugar del péptido señal natural, y una etiqueta de hexahistidina C terminal. La secuencia de proteína resultante es la SEQ ID 5. El marco de lectura abierto para ssDsbA-CRM197-his6 se insertó en los sitios NdeI y XbaI de pEC415 (Schulz H, Hennecke H, Thony-Meyer L: Prototype of a heme chaperone essential for cytochrome c maturation. Science 1998, 281(5380):1197-1200).

- 45 A partir de este plásmido, se hicieron diversos mutantes para analizar las diferencias de los rendimientos de CRM197. Las mutaciones se introdujeron en el sitio esperado de escisión del péptido señal mediante mutagénesis de cambio rápido según lo descrito por el fabricante (Stratagene, Agilent Technologies, Santa Clara, CA). Las construcciones resultantes se resumen en la Tabla 1.

- 50 Los plásmidos mencionados se transformaron en células BL21 y W3110 para realizar experimentos de expresión de proteínas. Las colonias transformadas se recogieron de una placa LB y se usaron para inocular el cultivo líquido medio LB, que se cultivaron durante la noche a 37 °C. Los cultivos de alta densidad se diluyeron a una OD₆₀₀ de 0,05 en medio LB nuevo y se cultivaron más hasta que la OD alcanzó un valor de OD₆₀₀ = 0,5. Luego se agregó arabinosa para la inducción de la expresión de proteínas recombinantes. Los experimentos iniciales usando algunas de las construcciones mencionadas se realizaron bajo diversas condiciones.

Sin embargo, no se detectó proteína CRM197 en extractos celulares en comparación con las células de control que no expresan proteínas o que expresan EPA (Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thony-Meyer L:

Production of glycoprotein vaccines in *E. coli*. Microbial cell factories 2010, 9:61). Ni a 30 ni a 37 °C, usando tiempos de inducción durante la noche y medio LB suplementado con ampicilina para el mantenimiento del plásmido.

Posteriormente, la expresión se realizó como sigue. Para la expresión, los cultivos de alta densidad de incubaciones durante la noche se diluyeron en caldo Terrific para una mejor viabilidad celular. Los cultivos se cultivaron hasta la fase exponencial y se indujeron durante 2 horas y durante la noche, y luego las células se cosecharon y los extractos celulares se prepararon disolviendo cantidades equivalentes de OD de biomasa en tampón de muestra Lämmli. Los extractos se separaron mediante SDS PAGE y se transfirieron por electrotransferencia a membranas de nitrocelulosa para la inmunodetección posterior usando antisueros anti DT y anti his tag. Sorprendentemente, se detectó una señal de proteína en la movilidad electroforética esperada de CRM197 a aproximadamente 60 kDa después de 2 horas de inducción. Las construcciones de expresión p932, p934 y p722 condujeron a señales detectables en inmunotransferencias antisuero anti DT y anti his tag. p932 parecía producir la mayoría, p934 menos, y p722 aún menos señales CRM197. Un extracto de control de células que contiene un plásmido de expresión que carece de una secuencia de péptido señal mostró CRM197 en el intervalo de peso molecular correcto y confirmó la identidad del material en los otros carriles.

Estos experimentos mostraron que CRM197 se podía expresar, pero no si era soluble o plegado. Como se indica en la Figura 1, se detectó CRM197 sin un péptido señal y se esperaba en cuerpos de inclusión citoplasmáticos. Los rendimientos esperados son desconocidos y solo pueden estimarse en comparación con la expresión de EPA. En esta comparación, CRM197 alcanza rendimientos similares a los de EPA en función de las intensidades de señal observadas usando la transferencia Western antisuero anti his tag como se ilustra en la Figura 1 (compare los carriles 4A y 4B con 5A y 5B). El EPA en fermentaciones controladas de biorreactores conduce a hasta 0,5 g/l de proteína.

El orden de eficiencia para la producción de CRM197 fue p932> p934> p722. El residuo de metionina codificado en el sitio de escisión del CRM197 expresado a partir de p722 puede interferir con la productividad, y también el residuo de glicina tiene cierta influencia. Parece, sin embargo, que la formación de un N terminal de CRM197 con un aminoácido menos (ADDV...; p932) que el N terminal natural en combinación con el residuo de serina en la posición -2 en relación con el sitio de escisión conduce al contexto de expresión óptimo cuando se usa el péptido señal DsbA.

Sin embargo, fue posible detectar señales CRM197 en experimentos de expresión usando diferentes construcciones de expresión y medio TB.

6.2 Ejemplo 2

Para analizar la solubilidad y el rendimiento global en matraces de agitación, CRM197 se purificó a partir de cultivos celulares usando dos construcciones de expresión diferentes. La cepa de expresión fue BL21, los plásmidos de expresión p932 o p933. Los matraces de agitación de 5 litros que contenían 1 l TB suplementado con ampicilina se inocularon con un cultivo previo al cultivo en LB suplementado con ampicilina y se cultivaron a 30 °C. A una OD₆₀₀ de 0,5, se agregó arabinosa al 0,2% (p/v) y se permitió la expresión durante 2 (p932) o 4 horas (p933). Las células se cosecharon luego por centrifugación, se resuspendieron en tampón para la extracción del periplasma (sacarosa al 20% p/v, Tris·HCl 30 mM pH 8,0, EDTA 1 mM, 1 mg/ml de lisozima, 1 comprimido/80 mL de mezcla inhibidora de proteasa completa (Roche, Basilea, Suiza)) en una proporción de 20 OD por ml, se incubó en hielo durante 30 minutos y se centrifugó durante 15 minutos a 8000 rpm y 4 °C. El sobrenadante se trató adicionalmente con DNasa (Fluka, Balgach, Suiza), se centrifugó a 4 °C, y el sobrenadante se esterilizó por filtración. El filtrado se preparó para la purificación usando cromatografía de afinidad Ni²⁺. La carga, el lavado y la elución se realizaron a concentraciones específicas de imidazol (10, 20, 500 mM). Las fracciones de elución se analizaron mediante SDS PAGE y tinción azul brillante de Coomassie (Figura 2).

Se detectó una banda principal correspondiente a CRM197 en fracciones de elución de la purificación. La determinación de proteínas dio como resultado valores de aproximadamente 2 mg de proteína de la construcción p932, y aproximadamente 4 mg de la construcción p933 por litro de caldo de fermentación. La secuenciación N terminal y MALDI MSMS de las bandas de proteínas extirpadas de este gel SDS PAGE confirmaron la terminación N de CRM197 en ambos casos (véase la Tabla 1) y que la proteína es realmente CRM197.

La diferencia entre la proteína expresada de p932 y p933 es la secuencia del péptido señal y la terminación N de CRM197 maduro resultante. p933 produjo la terminación N de tipo salvaje correcto; aunque la puntuación Y para la eficiencia de escisión es menor que para p932. En la figura 1, la expresión transmitida por p934 parece ser incluso menos eficiente y, de acuerdo con lo anterior, la puntuación Y es menor. De este modo, una combinación de un alto valor de puntuación Y y una posición de escisión del péptido señal que da como resultado la terminación N nativo GADDV parece ser la configuración óptima para la expresión de CRM197 de alto rendimiento en *E. coli*. El tiempo de expresión, la temperatura, el medio y la concentración del inductor pueden influir en el rendimiento, la velocidad y la eficacia de la escisión del péptido señal y, de acuerdo con lo anterior, en los rendimientos de CRM197.

6.3 Ejemplo 3

Para analizar la productividad de diferentes construcciones en paralelo, se realizaron experimentos de expresión en matraces de agitación a pequeña escala, los extractos periplásmicos se prepararon y analizaron mediante SDS

PAGE para la intensidad de la banda CRM197 mediante tinción de Coomassie (Figura 3) y se cuantificaron (Tabla 2). Las condiciones de expresión detalladas se dan en la leyenda de la Figura 3 y en la Tabla 2.

5 Los péptidos señal DsbA, MalE y PelB dieron como resultado los mejores rendimientos en combinación con condiciones de expresión optimizadas. Las condiciones de expresión tuvieron una influencia más fuerte en los rendimientos que las configuraciones del sitio de escisión del péptido señal. Sin embargo, la importancia de la secuencia del sitio de escisión del péptido señal se muestra, por ejemplo, por los bajos rendimientos obtenidos con el plásmido de expresión p722 (a 25 °C). Aunque p722 codifica la señal DsbA, el rendimiento es bajo en comparación con otras secuencias (codificadas, por ejemplo, en p932, p933, p934 o p936). Las configuraciones del sitio de escisión del péptido de señal se pueden clasificar de acuerdo con sus eficiencias de rendimiento: ASA-ADD y AMA-GADD parecen mejores que ASA-GADD, y AMG-ADD es el sitio menos eficiente. Las puntuaciones Y no se correlacionan con los niveles de expresión.

10 Todas las construcciones probadas que contienen la señal de PelB dieron como resultado altos rendimientos a una temperatura de expresión de 30 °C. Las diferencias en la secuencia del sitio de escisión del péptido señal no influyeron drásticamente en los rendimientos. Sin embargo, las diferencias en la secuencia del sitio de escisión del péptido señal fueron pequeñas en este conjunto de construcciones.

15

Tabla 2 Plásmidos y péptidos señal usados para la expresión periplásmica de CRM197 y una proteína de referencia bien secretada (EPA-6H). Los plásmidos se transformaron en *E. coli* (cepa derivada de W3110) y se cultivaron en medio TB a temperaturas que producen los niveles más altos de proteínas recombinantes. Se tomaron muestras para la preparación de extractos periplásmicos 4 h después de la inducción con 4 g de L⁻¹ L-arabinosa, la concentración de células resuspendidas en tampón de extracción de sacarosa-isozima se normalizó a OD₆₀₀ = 20. Las concentraciones de proteínas recombinantes sobreexpresadas en extractos periplásmicos se estimaron mediante análisis de imagen de un gel SDS-PAGE teñido con Coomassie usando bandas marcadoras a 55 kDa como referencia. Los rendimientos de proteínas en matraces de agitación se volvieron a calcular mediante OD₆₀₀ en el momento del muestreo. n.a.: no analizado.

Plásmido No.	Nombre GVXN	Nombre Proteína	SEQ ID	Péptido señal	Sitio de escisión previsto	Probabilidad de escisión (puntuación Y Señal P 4,0)	Terminal N determinado después de la exportación	Terminal C 6xHis tag	Temperatura de expresión óptima	Conc. de proteína (CRM197/EPA) en extractos periplásmicos (µg mL ⁻¹)	OD ₆₀₀ en el muestreo	Rendimiento proteico en matraz de agitación (mg L ⁻¹)
Plásmidos para la expresión periplásmica de CRM197, esqueleto pEC415												
1	p722	5	DsbA	AFS <u>AMG</u> -ADDV	0,763	n.a.	Sí	25°C	24	1,28	1,6	
2	p932	2	DsbA	AFS <u>ASA</u> -ADDV	0,878	ADDV...	Sí	25°C	66	2,66	8,8	
3	p933	4	DsbA	AFS <u>AMA</u> -GADDV	0,783	GADDV ...	Sí	25°C	51	1,30	3,3	
4	p936	x	DsbA	AFS <u>AMA</u> -GADDV	0,783	n.a.	no	20-25°C	57	1,62	4,6	
5	p934	x	DsbA	AFS <u>ASA</u> -GADDV	0,681	n.a.	Sí	25°C	64	1,38	4,4	
6	p1027	x	MalE	SAS <u>ALA</u> -MGADDV	0,722	n.a.	Sí	25°C	97	1,43	6,9	
7	p1029	x	MalE	SAS <u>ALA</u> -ADDV	0,894	n.a.	Sí	25°C	93	1,54	7,1	
8	p1030	x	OmpA	ATV <u>QA</u> -M GADDV	0,790	n.a.	Sí	25°C	9	1,53	0,7	
9	p1032	x	OmpA	ATV <u>QA</u> -ADDV	0,898	n.a.	Sí	25°C	11	1,54	0,9	
10	p1033	x	PeIB	AQP <u>AMA</u> -MGADDV	0,878	n.a.	Sí	30°C	36	3,29	5,9	
11	p1018	x	PeIB	AQP <u>AMA</u> -GADDV	0,874	n.a.	Sí	30°C	27	3,72	5,0	
12	p1035	x	PeIB	AQP <u>AMA</u> -ADDV	0,874	n.a.	no	30°C	32	3,96	6,4	
13	p1036	x	PeIB	AQP <u>AMA</u> -AGADDV	0,918	n.a.	Sí	30°C	44	3,27	7,2	
14	p938	x	LT-IIIb*	SV <u>QAHA</u> -GADDV	0,885	n.a.	Sí	30°C	13	1,11	0,7	
15	p1039	x	XynA	SAT <u>ASA</u> -MGADDV	0,464	n.a.	Sí	25°C	13	0,94	0,6	
Plásmido de referencia para la expresión periplásmica de EPA, esqueleto pEC415												
16	p150	-	DsbA	AFS <u>ASA</u> -AEEA	0,873	AEEA...	Sí	30°C	73	2,58	9,4	

* Enterotoxina termolábil de *E. coli* tipo IIb, cadena B

7. Secuencias

SEQ ID 1: p932

GTCGAGCTAGTAAAAGCATTTTTAAATAAGGAGGAATAACACATATGAAAAAGATTTGGCT
 GGCGCTGGCTGGTTTGTAGTTTTAGCGTTTAGCGCCAGCGCCGAGATGACGTTGTTGACAG
 CAGCAAATCCTTCGTTATGGAAAACCTCTCTCTTATCACGGCACCAAACCGGGCTATGT
 GGACAGCATTTCAGAAAGGTATCCAAAAACCGAAATCTGGCACGCAGGGTAACTACGATGA
 CGATTGGAAAGAATTCTACAGCACCGACAACAAATATGATGCGGCCGGTTACTCAGTTGA
 CAACGAAAATCCGCTGTTCGGCAAAGCCGGCGGTGTGGTTAAAGTCACGTATCCGGGCCCT
 GACCAAAGTCTGGCCCTGAAAGTGGATAATGCAGAAACCATCAAAAAAGAACTGGGTCT
 GAGCCTGACGGAACCGCTGATGGAACAGGTTGGCACCGAAGAATTTATCAAACGCTTCGG
 CGATGGTGCCAGTCGTGTCTGTCTGTCCCTGCCGTTTCGCAGAAGGTAGCTCTAGTGTGCA
 ATATATTAACAATTGGGAACAAGCGAAAGCCCTGTCCGTGGAAGTGGAAATCAACTTTGA
 AACCCGCGGCAAACGTGGTCAGGATGCGATGTATGAATACATGGCACAAGCTTGCAGCGG
 TAATCGCGTTTCGTGCGAGCGTCGGCTCCTCACTGTCTTGTATCAACCTGGACTGGGATGT
 TATCCGTGATAAAACCAAACGAAAATCGAAAGTCTGAAAGAACACGGCCCGATCAAAAA
 CAAAATGAGCGAATCTCCGAATAAAACGGTGTCCGAAGAAAAAGCTAAACAGTATCTGGA
 AGAATTCACCAAACCGCACTGGAACATCCGGAAGTTCAGAACTGAAAACCGTCACGGG
 TACCAACCCGGTGTTCGCGGCGCAAATTACGCAGCTTGGGCTGTGAACGTTGCGCAAGT
 GATTGACTCGGAAACGGCCGATAATCTGGAAAAAACACGGCGGCCCTGAGTATTCTGCC
 GGGCATCGGTTCCGTGATGGGTATTGCCGATGGCGCAGTTCATCACAAACCCGAAGAAAT
 TGTCGCCCAGTCTATCGCACTGTCGAGCCTGATGGTTGCTCAAGCGATTCCGCTGGTTGG
 CGAACTGGTTGATATCGGCTTTGCAGCTTACAACCTTCGTGGAAAGTATCATCAACCTGTT
 TCAGGTTGTCCATAACTCATATAATCGCCCGGCCTACTCGCCGGGTACAAAAACCCAACC
 GTTCTGCATGACGGCTACGCGGTTAGCTGGAATACGGTCGAAGATTCTATTATCCGTAC
 CGGCTTTTCAGGGTGAATCTGGCCACGACATTAATAATCACGGCTGAAAACACCCCGCTGCC
 GATTGCCGGTGTTCGCTGCCGACCATCCCGGGTAAACTGGATGTGAATAAATCAAAAAAC
 CCATATCTCGGTTAACGGTCGCAAAATTCGTATGCGCTGCCGTGCGATCGACGGCGATGT
 GACCTTCTGTCGTCCGAAAAGCCCGGTCTATGTGGCAACGGTGTTCATGCTAATCTGCA
 CGTCGCGTTTCATCGCTCTAGTTCCGAAAAAATCCACAGTAACGAAATCTCATCGGACTC
 CATTGGTGTGCTGGGCTACCAGAAAACGGTGGATCATACCAAAGTTAATAGCAAACCTGTC
 ACTGTTCTTCGAAATCAAATCAGGCTCGCATCATCATCACCACTAATCTAGAGGATC
 CCCGGGTACCGAGCTCGAATTCCTGAAGACGAAAGGGCCCTCGTGATACGCCATTTTTTAT
 AGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATG
 TGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGA
 GACAATAACCCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAAC
 ATTTCCGTGTCGCCCTTATTCCTTTTTTGCGGCATTTTGCCTTCTGTTTTTGCTCACC
 CAGAAACGCTGGTAAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACA
 TCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTTC
 CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCCG
 GGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCAC
 CAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCCA
 TAACCATGAGTGATAAACAATGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGG
 AGCTAACCGCTTTTTGCACAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAAC
 CGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGG
 CAACAACGTTGCGCAAACATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAAT
 TAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCCTTCTGCGCTCGGCCCTTCCGG

CTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG
 CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTC
 AGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGC
 ATTGGTAACGTGCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTCATT
 TTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAATCCCTT
 AACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTT
 GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAG
 CGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCA
 GCAGAGCGCAGATAACAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTTCA
 AGAACTCTGTAGCACCGCCTACATACTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
 CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
 CGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGAGCGAACGACCT
 ACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
 GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGC
 TTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGACTTG
 AGCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
 CGGCCTTTTTACGGTTCCTGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTCTGCGT
 TATCCCTGATCTGTGGATAACCGTATTACCGCCTTGAGTGTAGCTGATACCGCTCGCC
 GCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGC
 GGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATGGTGCACCTCTCAGTACA
 ATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACTGGG
 TCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGC
 TCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGT
 TTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAGCTCATCAGCGTGGTCTGT
 GAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAA
 GCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTCTGTTTGG
 TCACTGATGCCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATGAAAC
 GAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTT
 GTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTC
 AATGCCAGCGCTTCGTTAATACAGATGTAGGTGTCCACAGGGTAGCCAGCAGCATCCTG
 CGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTCAGACTTTTACG
 AAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTGCGAGACGTTTTGCGAGCAGC
 AGTCGCTTACGTTGCTCGGTATCGGTGATTTCATTCTGCTAACAGTAAGGCAACCCC
 GCCAGCCTAGCCGGGTCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACC
 CAACGCTGCCCCGAGCGTCAACGGCGCCAGATACAGCAAACGGCTGCGGGGGAAATACGCG
 GTTAAACGATCGACTGCCGCTTTGCGGCTGCGCCACAGCCGCCAGCATAGCCAGCCTCCG
 ACCCACAGCAGCAACGCCGTCGCCAGCAGCAGCCATTTGAAATCTCCGCTCTGCATATCG
 GAAGGAATATCGATTGCCGCTCCCGCCAGAATGCCCGGCAGGAAATAAACGGCGGCCAC
 AGCAAACAGCCAATCAAGTTCGGCCCAATAAATTTCCGCCACGGGAAGATCCAGCATCCCT
 GCCACCATCGGCACCAGCGGCCCTCGTCCGACCGACAAAACGTCCGACCAGGATCGTGAAC
 ATACTGTGCTGATGCAGCGCGTGTTCGGTTTTATCCAGCAGCGACTTGTCTTTTTTCATA
 AAAGACCAGCGGTGTAGCGGCTTTTTAAAGCGCCACCCAGCCAGAACGAAATCCAGTCG
 CCCATCAGACAGCCGATAATACCCACCAGCCAGGCATGCCAAAAATGAGCTCGCCGCTG
 CCGATAAGCGCGCCAGCCCCGCCATCAGTACCGTGCCGGGTAAAAATCAACCCACCAGC
 GCCAGCGATTCCAGGAAGGCGACCAGCAACACGGCGATGAGCGAATACAGAGTGGATTGG
 GTGATAAAGTGTTCAGCAGTGCTTGATAGTGTCCGTCAGCGTGATGAAGCAGGGAT
 TCTGCTTACCCCGTCCCCCTTCGTCAAGCCGTCAATTATCCGAATAGTTACGGCTTATGA

CATCTTTGTGGACACATCATTCACTTTTTATTACATCCGGCCCTGAACTCGCTAGGACT
 TGCCCCGGTGCATTTTTTAAATACCCGCGAAAAATAGAGCTGATCGTCAAATCCAACATT
 GCGCCCAACGGTCGCTATCGGCATTCGCGTAGTGCTAAGCAGAAGTTTCGCCTGGCTGAT
 ACGCTGATCTTCGCGCCAGCTCAATACGCTAATGCCTAACTGCTGGCGGAACAGATGTGA
 TAACCGGGAGGGCGACAGGCAGACATGCTGGGCGACGCTGGCGATATCAAAATGGCTGTC
 CGCCAGATGGTCGCTGATATACTGGCAGGCATCGCGCACACGGCTATCCATCGGCGGGTG
 CAACGACTCATTAATTACCGCCATACGTCTGAGCAACAACCTGCTCCAGCAGATTGATCGC
 CAGTAGCTCAGAATAGCGACCTTCCCCTTGCCCCGGCGCTGATGATCTGCCCGAACAGTTC
 GCTGAAATGCGGCTGGCGCGCCTCGTCCGGGCGGAAAAATCCTGTCTGGGCAAAGATTGT
 CGGCCAGGTCAGCCACTCCTGCCAGTAGGCGCGAGGCCGGAAATAAACCCACTGGTGATA
 CCACTCGCTGGCGTCCGGATGCCGTCCATAGTGATGAATCTCGCCCCGGCGGAAACAATAA
 TATATCGCCAGGCCGACAGACAAACTGCTCGCCATTATTATTAATGACGCCCTCTCCGCG
 GATGGTCAGGTTAAGAATATATCCCTTCATGCCCAACGGACGATCGATAAAAAAATCCAG
 ATATCCATTGCTTCAATTGGCGTCAGCCCGGCGACCAGATGGGCATTAATGAATATCC
 CGGCAATAGCGGATCATTTTTGCGTTTTAGCCATGATTTCTCTACCCCCGATGTTTCAGAG
 AAGAAACAAATTGTCCATATCGACCAGGACGACAGAGCTTCCGTCTCCGCAAGACTTTGC
 GCTTGATGAAAACGATCAACCCCGCTTGTGAAAAGCGCTTTGTAACAAAAGCGTACA
 GTTCAGGCGATAAAATTAAGTAACAGAAGTGTCTATAACTATGGCTGGAATGTCCACATT
 GAATATTTGCACAGCGTCACACTTTGCAAAGCATTAGCATTTTTGTCCATAAGATTAGCG
 GATCCTGCCTGACGGTTTTTTGCCGCGACTCTCTACTGTTTTCTCCATACCTGTTTTTCTGG
 ATGGAGTAAGACGATGGCAATTGCAATTGGCCTCGATTTTGGCAGTGATTGAGTGCAGCGC
 TCTGGCAGTGGACTGCGCCACCGGCGACGAGATCGCCACCAGCGTAGAGTGGTATCCGCG
 CTGGCAAGAAGGCCGTTATTGCGACGGCCGAACAACCAGTTCGGTCATCATCCGCGCGCA
 CTACATGGAGTCAATGGAGGCCGCGCTGAAAGCCGTTCTGGCACAATTAAGCGCCGCGCA
 ACGCGCAAATGTCGTTGGCATTGGCGTTGACAGCACCGGCTCTACGCCAGCGCCGATTGA
 CGCCGACGGTAACGTCCTGGCGCTGCGTCCAGAGTTCCGCCGAGAACCCGAATGCGATGTT
 TGTGCTGTGAAAAGATCACACCCCGTGGAAAGAGGCCGACGAAATCACTCGTCTGTGCCA
 TAAGCCAGGCAAG

SEQ ID 2: p933

GTCGAGCTAGTAAAAGCATTTTTAAATAAGGAGGAATAACACATATGAAAAAGATTTGGCT
 GCGCTGGCTGGTTTTAGTTTTAGCGTTTTAGCGCCATGGCAGGCGCAGATGACGTTGTTGA
 CAGCAGCAAATCCTTCGTTATGGAAAACCTTCTCCTCTTATCACGGCACCAACCAGGGCTA
 TGTGGACAGCATTCAGAAAGGTATCCAAAACCGAAATCTGGCACGCAGGGTAACTACGA
 TGACGATTGGAAAGAATTCTACAGCACCGACAACAATATGATGCGGCCGTTACTCAGT
 TGACAACGAAAATCCGCTGTCCGGCAAAGCCGGCGGTGTGGTTAAAGTCACGTATCCGGG
 CCTGACCAAAGTCCCTGGCCCTGAAAGTGGATAATGCAGAAACCATCAAAAAAGAACTGGG
 TCTGAGCCTGACGGAACCGCTGATGGAACAGGTTGGCACCAGAAGATTTATCAAACGCTT
 CGGCGATGGTGCCAGTCGTGTCGTGCTGTCCCTGCCGTTCCGAGAAGGTAGCTCTAGTGT
 CGAATATATTAACAATTGGGAACAAGCGAAAGCCCTGTCGCTGGAACCTGGAAATCAACTT
 TGAAACCCGCGGCAAACGTGGTCAGGATGCGATGTATGAATACATGGCACAAGCTTGGCG
 GGGTAATCGCGTTCGTGCGAGCGTCGGCTCCTCACTGTCTTGTATCAACCTGGACTGGGA
 TGTTATCCGTGATAAAACCAAACGAAAATCGAAAAGTCTGAAAGAACACGGCCCCGATCAA
 AAACAAAATGAGCGAATCTCCGAATAAAACGGTGTCCGAAGAAAAGCTAACAGTATCT
 GGAAGAATTCCACCAAACCGCACTGGAACATCCGGAACTGTCAGAACTGAAAACCGTCA
 GGGTACCAACCCGGTGTGTTGCCGGCGCAAATTACGCGAGCTTGGGCTGTGAACGTTGCGCA
 AGTGAATTGACTCGGAAACGGCCGATAATCTGGAAAAAACACGGCGGCCCTGAGTATTCT

GCCGGGCATCGGTTCCGTGATGGGTATTGCCGATGGCGCAGTTCATCACAACACCGAAGA
 AATTGTCGCCCAGTCTATCGCACTGTCGAGCCTGATGGTTGCTCAAGCGATTCCGCTGGT
 TGGCGAACTGGTTGATATCGGCTTTGCAGCTTACAACCTCGTGGAAAGTATCATCAACCT
 GTTTCAGGTTGTCATAAATCATATAATCGCCCGGCTACTCGCCGGGTCACAAAACCCA
 ACCGTTCCCTGCATGACGGCTACGCGGTTAGCTGGAATACGGTTCGAAGATTCTATTATCCG
 TACCGGCTTTCAGGGTGAATCTGGCCACGACATTTAAAATCACGGCTGAAAACACCCCGCT
 GCCGATTGCCGGTGTCTGCTGCCGACCATCCCGGGTAAACTGGATGTGAATAAATCAAA
 AACCCATATCTCGGTTAACGGTCGCAAAATTCGTATGCGCTGCCGTGCGATCGACGGCGA
 TGTGACCTTCTGTCGTCCGAAAAGCCCGGTCTATGTGGGCAACGGTGTTCATGCTAATCT
 GCACGTGCGGTTTCATCGCTCTAGTTCCGAAAAAATCCACAGTAACGAAATCTCATCGGA
 CTCCATTGGTGTGCTGGGCTACCAGAAAACGGTGGATCATAACAAAGTTAATAGCAAACCT
 GTCAGTGTCTTCGAAATCAAATCAGGCTCGCATCATCATCACCCTAATCTAGAGG
 ATCCCCGGGTACCGAGCTCGAATCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTT
 TATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAA
 ATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCA
 TGAGACAATAACCCTGATAAATGCTTCAATAATATGAAAAAGGAAGATGAGTATTC
 AACATTTCCGTGTCGCCCTTATCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGGCTC
 ACCCAGAAACGCTGGTGAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGT
 ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTT
 TTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACG
 CCGGGCAAGAGCAACTCGGTGCGCCGATACACTATTCTCAGAATGACTGGTGTGAGTACT
 CACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTG
 CCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGA
 AGGAGCTAACCGCTTTTTTGCAACAACATGGGGGATCATGTAACCTCGCCTGATCGTTGGG
 AACCGGAGCTGAATGAAGCCATAACCAACGACGAGCGTGACACCACGATGCCTGCAGCAA
 TGGCAACAACGTTGCGCAAACTATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAAC
 AATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTC
 CGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGGAGCGTGGGTCTCGCGGTATCA
 TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGA
 GTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTA
 AGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTC
 ATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCC
 CTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTT
 CTTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAACACCAGCTAC
 CAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACGGCT
 TCAGCAGAGCGCAGATAACAAACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACT
 TCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
 CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCAGGATA
 AGGCGCAGCGGTGCGGGCTGAACGGGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGA
 CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAG
 GGAGAAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGG
 AGCTTCCAGGGGGAAACGCTGATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGAC
 TTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCA
 ACGCGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTG
 CGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTC
 GCCGACCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA
 TCGGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATGGTGCCTCTCAGT

ACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT
 GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTC
 TGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
 GGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCCGTAAGCTCATCAGCGTGGT
 CGTGAAGCGATTCACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTTCTCCA
 GAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTCTGTT
 TGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATGA
 AACGAGAGAGGATGCTCACGATACGGGTACTGATGATGAACATGCCCGGTTACTGGAAC
 GTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGGCGGGACCAGAGAAAAATCACTCAGG
 GTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATC
 CTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTT
 ACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTGCGAGACGTTTTGCAGC
 AGCAGTCGCTTCACGTTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAAC
 CCCGCCAGCCTAGCCGGGTCTCAACGACAGGAGCACGATCATGCGCACCCCGTGGCCAGG
 ACCCAACGCTGCCCGAGCGTCAACGGCGCCAGATACAGCAAACGGCTGCGGGGGAAATAC
 GCGGTTAAACGATCGACTGCCGCTTTGCCGCTGCGCCACAGCCGCCAGCATAGCCAGCCT
 CCGACCCACAGCAGCAACGCCGTCGCCAGCAGCAGCCATTTGAAATCTCCGCTCTGCATA
 TCGGAAGGAATATCGATTGCCGCTCCCGCCAGAATGCCCGGCAGGAAATAAAACGGCGGC
 CACAGCAAACAGCCAATCAAGTTCGGCCCAATAAATTTGCCACGGGAAGATCCAGCATC
 CTTGCCACCATCGGCACCAGCGGCCTCGTTCGGACCAGCAAAACGTCCGACCAGGATCGTG
 AACATACTGTGCTGATGCAGCGCGTGTTCGGTTTTATCCAGCAGCGACTTGTTCTTTTTTC
 ATAAAAGACCAGCGGTGTAGCGGCTTTTTAAAGCGCCACCCAGCCAGAACGAAATCCAG
 TCGCCCATCAGACAGCCGATAATACCCACCAGCCAGGCATGCCAAAAATGAGCTCGCCG
 CTGCCGATAAGCGCGCCAGCCCCGCCATCAGTACCGTGCCGGGTAAAATCAACCCCACC
 AGCGCCAGCGATTCCAGGAAGGCGACCAGCAACACGGCGATGAGCGAATACAGAGTGGAT
 TGGGTGATAAAGTGTTCAGCAGTGTTCGCATAGTGTGTCCGTCAGCGTGATGAAGCAGG
 GATTCTGCTTACCCCGTCCCCCTTCGTCAAGCCGTCAATTATCCGAATAGTTACGGCTTA
 TGACATCTTTGTGGACACATCATTCACTTTTTTATTACATCCGGCCCTGAACTCGCTAGG
 ACTTGCCCCGGTGCATTTTTTAAATACCCGCGAAAAATAGAGCTGATCGTCAAATCCAAC
 ATTGCGCCCAACGGTTCGCTATCGGCATTTCGCGTAGTGCTAAGCAGAAGTTTCGCTGGCT
 GATACGCTGATCTTCGCGCCAGCTCAATACGCTAATGCCTAACTGCTGGCGGAACAGATG
 TGATAACCGGGAGGGCGACAGGCAGACATGCTGGGCGACGCTGGCGATATCAAATGGCT
 GTCCGCCAGATGGTCGCTGATATACTGGCAGGCATCGCGCACACGGCTATCCATCGGCGG
 GTGCAACGACTCATTAATTACCGCCATACGTCTGAGCAACAACCTGCTCCAGCAGATTGAT
 CGCCAGTAGCTCAGAATAGCGACCTTCCCCTTGCCCGGCGCTGATGATCTGCCCGAACAG
 TTCGCTGAAATGCGGCTGGCGCGCCTCGTCCGGGCGAAAAATCCTGTCTGGGCAAAGAT
 TGTCGGCCAGGTCAGCCACTCCTGCCAGTAGGCGCGAGGCGGAAATAAACCCTGGTG
 ATACCACTCGCTGGCGTCCGGATGCCGTCCATAGTGATGAATCTCGCCCGGCGGAAACAA
 TAATATATCGCCAGGCCGACAGACAAACTGCTCGCCATTATTATTAATGACGCCCTCTCC
 GCGGATGGTCAGGTTAAGAATATATCCCTTCATGCCCAACGGACGATCGATAAAAAAATC
 CAGATATCCATTCGCTTCAATTGGCGTCAGCCCGGCGACCAGATGGGCATTAATGAATA
 TCCCGGCAATAGCGGATCATTTTTCGTTTTAGCCATGATTTCTCTACCCCCGATGTTCA
 GAGAAGAAACAAATTGTCCATATCGACCAGGACGACAGAGCTTCCGTCTCCGCAAGACTT
 TGCGCTTGATGAAAGCACGTATCAACCCCGCTTGTGAAAAGCGCTTTGTAACAAAAGCGT
 ACAGTTCAGGCGATAAAATTAAGTAACAGAAGTGTCTATAACTATGGCTGGAATGTCCAC
 ATTGAATATTTGCACAGCGTCACACTTTGCAAAGCATTAGCATTTTTGTCCATAAGATTA
 GCGGATCCTGCCTGACGGTTTTTGCCGCGACTCTCTACTGTTTCTCCATACCTGTTTTTC

ES 2 750 525 T3

TGGATGGAGTAAGACGATGGCAATTGCAATTGGCCTCGATTTTGGCAGTGATTTCAGTGCG
CGCTCTGGCAGTGGACTGCGCCACCGGGCAGCAGATCGCCACCAGCGTAGAGTGGTATCC
GCGCTGGCAAGAAGGCCGTTATTGCGACGGCCCCGAACAACCAGTTCCGTCATCATCCGCG
CGACTACATGGAGTCAATGGAGGCCGCGCTGAAAGCCGTTCTGGCACAAATTAAGCGCCGC
GCAACGCGCAAATGTCGTTGGCATTTGGCGTTGACAGCACCAGGCTCTACGCCAGCGCCGAT
TGACGCCGACGGTAACGTCCTGGCGCTGCGTCCAGAGTTCCGCCGAGAACCAGCAATGCGAT
GTTTGTGCTGTGGAAAGATCACACCGCCGTGGAAGAGGCCGACGAAATCACTCGTCTGTG
CCATAAGCCAGGCAAG

SEQ ID 3: Péptido de señal que contiene la secuencia de aminoácidos de Crm197 expresada a partir de p932

MKKIWLALAGLVLAFFSASAADDVVDSSKSFVMENFSSYHGTPKPGYVDSIQKGIQKPKSGT
QGNYYYYWKEFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETI
KKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVE
LEINFETRGRKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKE
HGPIKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWA
VNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQ
AIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVE
DSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCR
AIDGDVTFPCRPKSPVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTK
VNSKLSLFFFEIKSGSHHHHHH

SEQ ID 4: Péptido de señal que contiene la secuencia de aminoácidos de Crm197 expresada a partir de p933

MKKIWLALAGLVLAFFSASAMAGADDVVDSSKSFVMENFSSYHGTPKPGYVDSIQKGIQKPKSG
TQGNYYYYWKEFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETI
IKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVE
ELEINFETRGRKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLK
EHGPIKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAW
AVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVA
QAIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNTV
EDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRC
RAIDGDVTFPCRPKSPVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHT
KVNSKLSLFFFEIKSGSHHHHHH

5

SEQ ID 5: secuencia de proteína traducida de p722

MKKIWLALAGLVLAFFSASAMAGADDVVDSSKSFVMENFSSYHGTPKPGYVDSIQKGIQKPKSGT
QGNYYYYWKEFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETI
KKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVE
LEINFETRGRKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKE
HGPIKNKMSESPNKTVSEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWA
VNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQ
AIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVE
DSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCR

AIDGDVTFPCRPKSPVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTK
VNSKLSLFFFEIKSGSHHHHHH

SEQ ID 6: CRM197 madura y secretada

ES 2 750 525 T3

GADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQPKSGTQGNVDDDWKEFYSTDNKYDAAG
YSVDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTTEEFIKRFG
DGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQDAMYEMYAQACAGNRVR
RSVGSSSLSCINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALE
HPELSELKTVTGTNPVFAFAGANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGA
VHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGH
KTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKT
HISVNGRKIRMRCRAIDGDVTFPCRPKSPVYVGNVHANLHVAFHRSSSEKIHSNEISSDSIGVL
GYQKTVDHKTVNSKLSLFFFEIKS

Listado de secuencias

- <110> GlycoVaxyn AG Eidgenossische Materialprufungs-und Forschungsanstalt
- <120> PROCEDIMIENTOS Y COMPOSICIONES RELACIONADAS CON CRM197
- 5 <130> 103399PC
- <140> para ser asignado
- <141>
- <150> 61/746,366
- <151> 2012-12-27
- 10 <160> 6
- <170> PatentIn version 3.5
- <210> 1
- <211> 7093
- <212> ADN
- 15 <213> Secuencia artificial
- <220>
- <223> Plásmido de expresión p932
- <400> 1

ES 2 750 525 T3

gtcgagctag	taaaagcatt	ttaaataagg	aggaataaca	catatgaaaa	agatttggct	60
ggcgctggct	ggtttagttt	tagcgttttag	cgccagcgcc	gcagatgacg	ttgttgacag	120
cagcaaatcc	ttcgttatgg	aaaacttctc	ctcttatcac	ggcaccaaac	cgggctatgt	180
ggacagcatt	cagaaaggta	tccaaaaacc	gaaatctggc	acgcagggta	actacgatga	240
cgattggaaa	gaattctaca	gcaccgacaa	caaatatgat	gcggccgggt	actcagttga	300
caacgaaaat	ccgctgtcgg	gcaaagccgg	cggtgtgggt	aaagtcacgt	atccgggcct	360
gaccaaagtc	ctggccctga	aagtggataa	tgcagaaacc	atcaaaaaag	aactgggtct	420
gagcctgacg	gaaccgctga	tggaacaggt	tggcaccgaa	gaatttatca	aacgcttcgg	480
cgatggtgcc	agtcgtgtcg	tgctgtccct	gccgttcgca	gaaggtagct	ctagtgtcga	540
atatattaac	aattgggaac	aagcgaagc	cctgtccgtg	gaactggaaa	tcaactttga	600
aaccgcggc	aaacgtggtc	aggatgcgat	gtatgaatac	atggcacaag	cttgcgcggg	660
taatcgcggt	cgtcgcagcg	tcggctcctc	actgtcttgt	atcaacctgg	actgggatgt	720
tatccgtgat	aaaacaaaa	cgaaaatcga	aagtctgaaa	gaacacggcc	cgatcaaaaa	780
caaatgagc	gaatctccga	ataaaacggt	gtccgaagaa	aaagctaaac	agtatctgga	840
agaattccac	caaaccgcac	tggaacatcc	ggaactgtca	gaactgaaaa	ccgtcacggg	900
taccaacccg	gtgtttgccg	gcgcaaatta	cgcagcttgg	gctgtgaacg	ttgcgcaagt	960
gattgactcg	gaaacggccg	ataatctgga	aaaaaccacg	gcggccctga	gtattctgcc	1020
gggcatcggt	tccgtgatgg	gtattgccga	tggcgcagtt	catcacaaca	ccgaagaaat	1080
tgtcgcccag	tctatcgcac	tgtcgagcct	gatggttgct	caagcgattc	cgctggttgg	1140
cgaactgggt	gatatcggct	ttgcagctta	caacttcgtg	gaaagtatca	tcaacctggt	1200

ES 2 750 525 T3

tcaggttgtc cataactcat ataatcgccc ggcctactcg ccgggtcaca aaaccaacc 1260
 gttcctgcat gacggctacg cggttagctg gaatacggtc gaagattcta ttatccgtac 1320
 cggctttcag ggtgaatctg gccacgacat taaaatcacg gctgaaaaca ccccgtgcc 1380
 gattgccggt gttctgctgc cgaccatccc gggtaaactg gatgtgaata aatcaaaaac 1440
 ccatactctg gttaacggtc gcaaaattcg tatgcgctgc cgtgcgatcg acggcgatgt 1500
 gaccttctgt cgcccgaaaa gcccggtcta tgtgggcaac ggtgttcag ctaatctgca 1560
 cgtcgcgttt catcgtctta gttccgaaaa aatccacagt aacgaaatct catcggactc 1620
 cattggtgtg ctgggctacc agaaaacggg ggatcatacc aaagttaata gcaaaactgtc 1680
 actgtttctc gaaatcaaat caggctcgca tcatcatcat caccactaat ctagaggatc 1740
 cccgggtacc gagctcgaat tcttgaagac gaaagggcct cgtgatacgc ctatTTTTat 1800
 aggttaatgt catgataata atggtttctt agacgtcagg tggcactttt cggggaaatg 1860
 tgcgcggaac ccctatttgt tttttttctt aaatacattc aaatatgtat ccgctcatga 1920
 gacaataacc ctgataaatg cttcaataat attgaaaaag gaagagtatg agtattcaac 1980
 atttccgtgt cgcccttatt ccctTTTTgt cggcattttt ccttcctggt tttgctcacc 2040
 cagaaaacgct ggtgaaagta aaagatgctg aagatcagtt ggggtgcacga gtgggttaca 2100
 tcgaactgga tctcaacagc ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc 2160
 caatgatgag cacttttaaa gttctgctat gtggcgcggt attatcccgt gttgacgccg 2220
 ggcaagagca actcggctcg cgacatacact attctcagaa tgacttggtt gagtactcac 2280
 cagtcacaga aaagcatctt acggatggca tgacagtaag agaattatgc agtgctgcca 2340
 taaccatgag tgataacact gcggccaact tacttctgac aacgatcggg ggaccgaagg 2400
 agctaaccgc ttttttgac aacatggggg atcatgtaac tcgccttgat cgttgggaac 2460
 cggagctgaa tgaagccata ccaaacgacg agcgtgacac cacgatgcct gcagcaatgg 2520
 caacaacggt gcgcaaaact ttaactggcg aactacttac tctagcttcc cggcaacaat 2580
 taatagactg gatggaggcg gataaagttg caggaccact tctgcgctcg gcccttccgg 2640
 ctggctggtt tattgctgat aaatctggag ccggtgagcg tgggtctcgc ggtatcattg 2700
 cagcactggg gccagatggt aagccctccc gtatcgtagt tatctacacg acggggagtc 2760
 aggcaactat ggatgaacga aatagacaga tcgctgagat aggtgcctca ctgattaagc 2820
 attggtaact gtcagaccaa gtttactcat atatacttta gattgattta aaacttcatt 2880
 ttttaattaa aaggatctag gtgaagatcc tttttgataa tctcatgacc aaaatccctt 2940
 aacgtgagtt ttcgttccac tgagcgtcag accccgtaga aaagatcaaa ggatcttctt 3000
 gagatccttt ttttctgctg gtaatctgct gcttgcaaac aaaaaacca ccgctaccag 3060
 cggtggtttt tttgccggat caagagctac caactctttt tccgaaggta actggcttca 3120
 gcagagcgca gataccaaat actgtccttc tagtgtagcc gtagttaggc caccacttca 3180
 agaactctgt agcaccgcct acatacctcg ctctgctaata cctgttacca gtggctgctg 3240

ES 2 750 525 T3

ccagtggcga taagtcgtgt cttaccgggt tggactcaag acgatagtta ccggataagg 3300
 cgcagcggtc gggctgaacg gggggttcgt gcacacagcc cagcttggag cgaacgacct 3360
 acaccgaact gagataccta cagcgtgagc tatgagaaag cgccacgctt cccgaagggg 3420
 gaaaggcggg caggtatccg gtaagcggca gggtcggaac aggagagcgc acgaggggagc 3480
 ttccaggggg aaacgcctgg tatctttata gtctgtcgg gtttcgccac ctctgacttg 3540
 agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct atggaaaaac gccagcaacg 3600
 cggccttttt acggttcctg gccttttgct ggccttttgc tcacatgttc tttcctgcgt 3660
 tatcccctga ttctgtggat aaccgtatta ccgcctttga gtgagctgat accgctcgcc 3720
 gcagccgaac gaccgagcgc agcagatcag tgagcgagga agcgggaagag cgctgatgc 3780
 ggtattttct ccttacgcat ctgtgcggta tttcacaccg catggtgcac tctcagtaca 3840
 atctgctctg atgccgcata gttaagccag tatacactcc gctatcgcta cgtgactggg 3900
 tcattggtgc gccccgacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc 3960
 tcccggcatc cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtcagaggt 4020
 tttcacgcgc atcaccgaaa cgcgcgagggc agctgcggta aagctcatca gcgtggctgt 4080
 gaagcgattc acagatgtct gcctgttcat ccgcgtccag ctcgttgagt ttctccagaa 4140
 gcgttaatgt ctggcttctg ataaagcggg ccattgtaag ggcggttttt tcctgtttgg 4200
 tcaactgatc ctccgtgtaa gggggatttc tgttcatggg ggtaatgata ccgatgaaac 4260
 gagagaggat gctcacgata cgggttactg atgatgaaca tgcccggta ctggaacggt 4320
 gtgagggtaa acaactggcg gtatggatgc ggcgggacca gagaaaaatc actcagggtc 4380
 aatgccagcg cttcgtaaat acagatgtag gtgttccaca gggtagccag cagcatcctg 4440
 cgatgcagat ccggaacata atgggtcagg gcgctgactt ccgcgtttcc agactttacg 4500
 aaacacggaa accgaagacc attcatgttg ttgctcaggt cgcagacggt ttgcagcagc 4560
 agtcgcttca cgttcgctcg cgtatcgggt attcattctg ctaaccagta aggcaacccc 4620
 gccagcctag ccgggtcctc aacgacagga gcacgatcat gcgaccccg ggccaggacc 4680
 caacgctgcc cgagcgtcaa cggcgccaga tacagcaaac ggctgcgggg gaaatacgcg 4740
 gttaaacgat cgactgccgc tttgccgctg cgccacagcc gccagcatag ccagcctccg 4800
 acccacagca gcaacgccgt cgccagcagc agccatttga aatctccgct ctgcatatcg 4860
 gaaggaatat cgattgccgc tcccgccaga atgcccggca ggaataaaaa cggcgccac 4920
 agcaaacagc caatcaagtt cggccaata aatttcgcca cgggaagatc cagcatccct 4980
 gccaccatcg gcaccagcgg cctcgtcggg ccgacaaaac gtccgaccag gatcgtgaac 5040
 atactgtgct gatgcagcgc gtgttcgggt ttatccagca gcgacttggt ctttttcata 5100
 aaagaccagc ggtgtagcgg ctttttaag cgccacccca gccagaacga aatccagtcg 5160
 cccatcagac agccgataat acccaccagc caggcatgcc aaaaattgag ctcgccgctg 5220
 ccgataagcg cgcccagccc cgccatcagt accgtgccgg gtaaaatcaa ccccaccagc 5280

ES 2 750 525 T3

gccagcgatt ccaggaaggc gaccagcaac acggcgatga gcgaatacag agtggattgg 5340
 gtgataaagt gttccagcag tgcttgcata gtgtgtccgt cagcgtgatg aagcagggat 5400
 tctgcttacc ccgtccccct tcgtcaagcc gtcaattatc cgaatagtta cggcttatga 5460
 catctttgtg gacacatcat tcacttttta ttcacatccg gccctgaact cgctaggact 5520
 tgccccggtg ctttttttaa ataccgcga aaaatagagc tgatcgtcaa atccaacatt 5580
 gcgcccacg gtcgctatcg gcattcgcgt agtgctaagc agaagtttcg cctggctgat 5640
 acgctgatct tcgcgccagc tcaatagcgt aatgcctaac tgctggcgga acagatgtga 5700
 taaccgggag ggcgacaggc agacatgctg ggcgacgctg gcgatatcaa aatggctgtc 5760
 cgccagatgg tcgctgatat actggcaggc atcgcgcaca cggctatcca tcggcgggtg 5820
 caacgactca ttaattaccg ccatacgtct gagcaacaac tgctccagca gattgatcgc 5880
 cagtagctca gaatagcgac cttccccttg cccggcgctg atgatctgcc cgaacagttc 5940
 gctgaaatgc ggctggcgcg cctcgtccgg gcggaaaaat cctgtctggg caaagattgt 6000
 cggccaggtc agccactcct gccagtaggc gcgaggccgg aaataaaccc actggtgata 6060
 ccaactcgtg gcgtccggat gccgtccata gtgatgaatc tcgcccggcg gaaacaataa 6120
 tatatcgcca ggccgacaga caaactgctc gccattatta ttaatgacgc cctctcccg 6180
 gatggtcagg ttaagaatat atcccttcat gcccaacgga cgatcgataa aaaaatccag 6240
 atatccattc gcttcaattg gcgtcagccc ggcgaccaga tgggcattaa atgaatatcc 6300
 cggcaatagc ggatcatttt gcgtttcagc catgatttct ctaccccccg atgttcagag 6360
 aagaaacaaa ttgtccatat cgaccaggac gacagagctt ccgtctccgc aagactttgc 6420
 gcttgatgaa agcacgtatc aaccccgctt gtgaaaagcg ctttgtaaca aaagcgtaca 6480
 gttcaggcga taaaattaag taacagaagt gtctataact atggctggaa tgtccacatt 6540
 gaatatttgc acagcgtcac actttgcaaa gcattagcat ttttgccat aagattagcg 6600
 gatcctgcct gacggttttt gccgcgactc tctactgttt ctccatacct gtttttctgg 6660
 atggagtaag acgatggcaa ttgcaattgg cctcgatttt ggcagtgatt cagtgcgcgc 6720
 tctggcagtg gactgcgcca ccggcgacga gatcgccacc agcgtagagt ggtatcccg 6780
 ctggcaagaa ggccgttatt gcgacggccc gaacaaccag ttccgtcatc atccgcgcga 6840
 ctacatggag tcaatggagg ccgcgctgaa agccgttctg gcacaattaa gcgccgcga 6900
 acgcgcaaat gtcgttggca ttggcgttga cagcaccggc tctacgccag cgccgattga 6960
 cgccgacggt aacgtcctgg cgctgcgtcc agagtccgcc gagaacccga atgcgatgtt 7020
 tgtgtgtgg aaagatcaca ccgccgtgga agaggccgac gaaatcactc gtctgtgcca 7080
 taagccaggc aag 7093

<210> 2

<211> 7096

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Plásmido de expresión p933

ES 2 750 525 T3

<400> 2

gtcgagctag	taaaagcatt	ttaaataagg	aggaataaca	catatgaaaa	agatttggct	60
ggcgctggct	ggtttagttt	tagcgtttag	cgccatggca	ggcgcagatg	acgttgttga	120
cagcagcaaa	tccttcgtta	tggaaaactt	ctcctcttat	cacggcacca	aaccgggcta	180
tgtggacagc	attcagaaaag	gtatccaaaa	accgaaatct	ggcacgcagg	gtaactacga	240
tgacgattgg	aaagaattct	acagcaccga	caacaaatat	gatgcggccg	gttactcagt	300
tgacaacgaa	aatccgctgt	cgggcaaagc	cggcggtgtg	gttaaagtca	cgtatccggg	360
cctgaccaaa	gtcctggccc	tgaaagtgga	taatgcagaa	accatcaaaa	aagaactggg	420
tctgagcctg	acggaaccgc	tgatggaaca	ggttggcacc	gaagaattta	tcaaacgctt	480
cggcgatggg	gccagtcgtg	tcgtgctgtc	cctgccgttc	gcagaaggta	gctctagtgt	540
cgaatatatt	aacaattggg	aacaagcgaa	agccctgtcc	gtggaactgg	aaatcaactt	600
tgaaacccgc	ggcaaacgtg	gtcaggatgc	gatgtatgaa	tacatggcac	aagcttgccg	660
gggtaatcgc	gttcgtcgca	gcgtcggctc	ctcactgtct	tgtatcaacc	tggactggga	720
tgttatccgt	gataaaaacca	aaacgaaaat	cgaaagtctg	aaagaacacg	gcccgatcaa	780
aaacaaaatg	agcgaatctc	cgaataaaac	ggtgtccgaa	gaaaaagcta	aacagtatct	840
ggaagaattc	caccaaaccg	cactggaaca	tccggaactg	tcagaactga	aaaccgtcac	900
gggtaccaac	ccggtgtttg	ccggcgcaaa	ttacgcagct	tgggctgtga	acgttgcgca	960
agtgattgac	tcggaaacgg	ccgataatct	ggaaaaaac	acggcggccc	tgagtattct	1020
gccgggcatc	ggttccgtga	tgggtattgc	cgatggcgca	gttcatcaca	acaccgaaga	1080
aattgtcgcc	cagtctatcg	cactgtcgag	cctgatgggt	gctcaagcga	ttccgctggg	1140
tggcgaactg	gttgatatcg	gctttgcagc	ttacaacttc	gtggaaagta	tcatcaacct	1200
gtttcagggt	gtccataact	catataatcg	cccggcctac	tcgccgggtc	acaaaacca	1260
accgttcctg	catgacggct	acgcgggttag	ctggaatacg	gtcgaagatt	ctattatccg	1320
taccggcttt	caggggtgaat	ctggccacga	cattaaaatc	acggctgaaa	acaccccgtc	1380
gccgattgcc	ggtgttctgc	tgccgaccat	cccgggtaaa	ctggatgtga	ataaatcaaa	1440
aaccatatac	tcggttaacg	gtcgcaaaat	tcgtatgcgc	tgccgtgcga	tcgacggcga	1500
tgtgaccttc	tgtcgtccga	aaagcccggg	ctatgtgggc	aacgggtgtc	atgctaactc	1560
gcacgtcgcg	tttcatcgct	ctagtccga	aaaaatccac	agtaacgaaa	tctcatcgga	1620
ctccattggg	gtgctgggct	accagaaaac	ggtggatcat	accaaagtta	atagcaaact	1680
gtcactgttc	ttcgaaatca	aatcaggctc	gcatcatcat	catcaccact	aatctagagg	1740
atccccgggt	accgagctcg	aattcttgaa	gacgaaaggg	cctcgtgata	cgctatttt	1800
tatagggttaa	tgtcatgata	ataatggttt	cttagacgtc	aggtggcact	tttcggggaa	1860
atgtgcgchg	aaccctatt	tgtttatfff	tctaaataca	ttcaaataatg	tatccgctca	1920

ES 2 750 525 T3

tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc 1980
 aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct gtttttgctc 2040
 acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt 2100
 acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacggt 2160
 ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtggtgacg 2220
 ccgggcaaga gcaactcggg cgccgcatac actattctca gaatgacttg gttgagtact 2280
 caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg 2340
 ccataacatc gagtgataac actgcgccca acttacttct gacaacgatc ggaggaccga 2400
 aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg 2460
 aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgcagcaa 2520
 tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac 2580
 aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc 2640
 cggctggctg gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca 2700
 ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga 2760
 gtcaggcaac tatggatgaa cgaaatagac agatcgcgtg gataggtgcc tcaactgatta 2820
 agcattggta actgtcagac caagtttact catatatact ttagattgat ttaaaacttc 2880
 atttttaatt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc 2940
 cttaacgtga gttttcgttc cactgagcgt cagaccccggt agaaaagatc aaaggatcct 3000
 cttgagatcc ttttttctg cgcgtaatct gctgcttgca aacaaaaaaaa ccaccgctac 3060
 cagcgggtgt ttgtttgccg gatcaagagc taccaactct ttttcgaag gtaactggct 3120
 tcagcagagc gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggccaccact 3180
 tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 3240
 ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata 3300
 aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca gccagcttg gagcgaacga 3360
 cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag 3420
 ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg 3480
 agcttccagg gggaaacgcc tggatcttt atagtcctgt cgggtttcgc cacctctgac 3540
 ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 3600
 acgcgccctt tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcctg 3660
 cgttatcccc tgattctgtg gataaccgta ttaccgcctt tgagttagct gataccgctc 3720
 gccgcagccg aacgaccgag cgcagcagat cagtgagcga ggaagcggaa gagcgcctga 3780
 tgcggtatct tctccttacg catctgtgcg gtatttcaca ccgcatggtg cactctcagt 3840
 acaatctgct ctgatgccg atagttaagc cagtatacac tccgctatcg ctacgtgact 3900
 gggatcatggc tgcccccca caccgccaa caccgcgtga cgcgccctga cgggcttgtc 3960

ES 2 750 525 T3

tgctcccggc atccgcttac agacaagctg tgaccgtctc cgggagctgc atgtgtcaga 4020
 ggttttcacc gtcacaccg aaacgcgcga ggcagctgcg gtaaagctca tcagcgtggt 4080
 cgtgaagcga ttcacagatg tctgcctggt catccgcgtc cagctcgttg agtttctcca 4140
 gaagcgttaa tgtctggcct ctgataaagc gggccatggt aagggcggtt ttttctggtt 4200
 tggtcactga tgcctccgtg taagggggat ttctgttcat gggggtaatg ataccgatga 4260
 aacgagagag gatgctcacg atacggggtta ctgatgatga acatgcccgg ttactggaac 4320
 gttgtgaggg taaacaactg gcggtatgga tgcggcggga ccagagaaaa atcactcagg 4380
 gtcaatgcc a gcgcttcggt aatacagatg taggtgttcc acagggtagc cagcagcatc 4440
 ctgcatgca gatccggaac ataatggtgc agggcgctga cttccgcggt tccagacttt 4500
 acgaaacacg gaaaccgaag accattcatg ttgttgctca ggtcgcagac gttttgcagc 4560
 agcagtcgct tcacgttcgc tcgcgtatcg gtgattcatt ctgctaacca gtaaggcaac 4620
 cccgccagcc tagccgggtc ctcaacgaca ggagcacgat catgcgacc cgtggccagg 4680
 acccaacgct gcccgagcgt caacggcgcc agatacagca aacggctgcg ggggaaatac 4740
 gcggttaaac gatcgactgc cgctttgccg ctgcgccaca gccgccagca tagccagcct 4800
 ccgaccaca gcagcaacgc cgctcgccagc agcagccatt tgaaatctcc gctctgcata 4860
 tcggaaggaa tatcgattgc cgctcccgcc agaatgcccg gcaggaaata aaacggcggc 4920
 cacagcaaac agccaatcaa gttcggccca ataaatttcg ccacgggaag atccagcatc 4980
 cctgccacca tcggcaccag cggcctcgtc ggaccgaca aacgtccgac caggatcgtg 5040
 aacatactgt gctgatgcag cgcgtgttcg gttttatcca gcagcgactt gttctttttc 5100
 ataaaagacc agcgggtgtag cggcttttta aagcgccacc ccagccagaa cgaaatccag 5160
 tcgcccata gacagccgat aataccacc agccaggcat gccaaaaatt gagctcgccc 5220
 ctgccgataa gcgcgcccag ccccgccatc agtaccgtgc cgggtaaaat caaccccacc 5280
 agcggcagc attccaggaa ggcgaccagc aacacggcga tgagcgaata cagagtggat 5340
 tgggtgataa agtgttccag cagtgccttc atagtgtgtc cgtcagcgtg atgaagcagg 5400
 gattctgctt accccgtccc cttcgtcaa gccgtcaatt atccgaatag ttacggctta 5460
 tgacatcttt gtggacacat cttcacttt ttattcacat ccggccctga actcgttagg 5520
 acttgccccg gtgcattttt taaataccg cgaaaaatag agctgatcgt caaatccaac 5580
 attgcgccc acggtcgcta tcggcattcg cgtagtgcta agcagaagtt tcgcctggct 5640
 gatagcgtga tcttcgcgcc agctcaatac gctaatgcct aactgctggc ggaacagatg 5700
 tgataaccgg gagggcgaca ggcagacatg ctgggcgacg ctggcgatat caaaatggct 5760
 gtccgccaga tggtcgctga tatactggca ggcacgcgc acacggctat ccatcggcgg 5820
 gtgcaacgac tcattaatta ccgccatag tctgagcaac aactgctcca gcagattgat 5880
 cgccagtagc tcagaatagc gaccttcccc ttgcccggcg ctgatgatct gcccaacag 5940
 ttcgctgaaa tgcggctggc gcgcctcgtc cgggcggaaa aatcctgtct gggcaagat 6000

ES 2 750 525 T3

tgtcggccag gtcagccact cctgccagta ggcgcgaggc cggaataaa cccactggtg 6060
 ataccactcg ctggcgctcg gatgccgtcc atagtgatga atctcgcccc gcggaacaa 6120
 taatatatcg ccaggccgac agacaaactg ctcgccatta ttattaatga cgccctctcc 6180
 gcggatggtc aggttaagaa tatatccctt catgcccaac ggacgatcga taaaaaatc 6240
 cagatatcca ttcgcttcaa ttggcgctcag cccggcgacc agatgggcat taaatgaata 6300
 tcccggcaat agcggatcat tttgcgttcc agccatgatt tctctacccc ccgatgttca 6360
 gagaagaaac aaattgtcca tatcgaccag gacgacagag cttccgtctc cgcaagactt 6420
 tgcgcttgat gaaagcacgt atcaaccccg cttgtgaaaa gcgctttgta acaaaagcgt 6480
 acagttcagg cgataaaatt aagtaacaga agtgtctata actatggctg gaatgtccac 6540
 attgaatatt tgcacagcgt cacactttgc aaagcattag cttttttgtc cataagatta 6600
 gcggatcctg cctgacggtt tttgccgca ctctctactg tttctccata cctgtttttc 6660
 tggatggagt aagacgatgg caattgcaat tggcctcgat tttggcagtg attcagtgcg 6720
 cgctctggca gtggactgcg ccaccggcga cgagatcgcc accagcgtag agtggatcc 6780
 gcgctggcaa gaagccggtt attgcgacgg cccgaacaac cagttccgtc atcatccgcg 6840
 cgactacatg gagtcaatgg aggccgcgt gaaagccgtt ctggcacaat taagcgccgc 6900
 gcaacgcgca aatgtcgttg gcattggcgt tgacagcacc ggctctacgc cagcgccgat 6960
 tgacgcccgc ggtaacgtcc tggcgtgcg tccagagttc gccgagaacc cgaatgcat 7020
 gtttgtgctg tggaaagatc acaccgccgt ggaagaggcc gacgaaatca ctcgtctgtg 7080
 ccataagcca ggcaag 7096

<210> 3

<211> 561

<212> PRT

5 <213> Secuencia artificial

<220>

<223> Péptido de señal que contiene la secuencia de aminoácidos de Crm197 expresada a partir de p932

<400> 3

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
 1 5 10 15
 Ala Ser Ala Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met
 20 25 30
 Glu Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser
 35 40 45
 Ile Gln Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr
 50 55 60
 Asp Asp Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala
 65 70 75 80

ES 2 750 525 T3

Ala Gly Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly
85 90 95

Gly Val Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu
100 105 110

Lys Val Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu
115 120 125

Thr Glu Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg
130 135 140

Phe Gly Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu
145 150 155 160

Gly Ser Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala
165 170 175

Leu Ser Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly
180 185 190

Gln Asp Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg
195 200 205

Val Arg Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp
210 215 220

Asp Val Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu
225 230 235 240

His Gly Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val
245 250 255

Ser Glu Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala
260 265 270

Leu Glu His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn
275 280 285

Pro Val Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala
290 295 300

Gln Val Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala
305 310 315 320

Ala Leu Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp
325 330 335

Gly Ala Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala
340 345 350

ES 2 750 525 T3

Leu Ser Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu
 355 360 365

Val Asp Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn
 370 375 380

Leu Phe Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro
 385 390 395 400

Gly His Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp
 405 410 415

Asn Thr Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser
 420 425 430

Gly His Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala
 435 440 445

Gly Val Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser
 450 455 460

Lys Thr His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg
 465 470 475 480

Ala Ile Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr
 485 490 495

Val Gly Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser
 500 505 510

Ser Ser Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly
 515 520 525

Val Leu Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys
 530 535 540

Leu Ser Leu Phe Phe Glu Ile Lys Ser Gly Ser His His His His His
 545 550 555 560

His

<210> 4

<211> 562

<212> PRT

5 <213> Secuencia artificial

<220>

<223> Péptido de señal que contiene la secuencia de aminoácidos de Crm197 expresada a partir de p933

<400> 4

ES 2 750 525 T3

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1 5 10 15

Ala Met Ala Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val
20 25 30

Met Glu Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp
35 40 45

Ser Ile Gln Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn
50 55 60

Tyr Asp Asp Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp
65 70 75 80

Ala Ala Gly Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala
85 90 95

Gly Gly Val Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala
100 105 110

Leu Lys Val Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser
115 120 125

Leu Thr Glu Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys
130 135 140

Arg Phe Gly Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala
145 150 155 160

Glu Gly Ser Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys
165 170 175

Ala Leu Ser Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg
180 185 190

Gly Gln Asp Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn
195 200 205

Arg Val Arg Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp
210 215 220

Trp Asp Val Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys
225 230 235 240

Glu His Gly Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr
245 250 255

Val Ser Glu Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr
260 265 270

ES 2 750 525 T3

Ala Leu Glu His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr
 275 280 285

Asn Pro Val Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val
 290 295 300

Ala Gln Val Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr
 305 310 315 320

Ala Ala Leu Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala
 325 330 335

Asp Gly Ala Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile
 340 345 350

Ala Leu Ser Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu
 355 360 365

Leu Val Asp Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile
 370 375 380

Asn Leu Phe Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser
 385 390 395 400

Pro Gly His Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser
 405 410 415

Trp Asn Thr Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu
 420 425 430

Ser Gly His Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile
 435 440 445

Ala Gly Val Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys
 450 455 460

Ser Lys Thr His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys
 465 470 475 480

Arg Ala Ile Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val
 485 490 495

Tyr Val Gly Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg
 500 505 510

Ser Ser Ser Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile
 515 520 525

Gly Val Leu Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser
 530 535 540

Lys Leu Ser Leu Phe Phe Glu Ile Lys Ser Gly Ser His His His His
 545 550 555 560

His His

ES 2 750 525 T3

<210> 5

<211> 561

<212> PRT

<213> Secuencia artificial

5 <220>

<223> Secuencia de proteína traducida del plásmido de expresión p722

<400> 5

```

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

```


ES 2 750 525 T3

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
 130 135 140
 Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
 145 150 155 160
 Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
 165 170 175
 Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
 180 185 190
 Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
 195 200 205
 Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
 210 215 220
 Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
 225 230 235 240
 Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
 245 250 255
 His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
 260 265 270
 Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val
 275 280 285
 Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu
 290 295 300
 Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala
 305 310 315 320
 Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser
 325 330 335
 Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp
 340 345 350
 Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe
 355 360 365
 Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His
 370 375 380
 Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr
 385 390 395 400

ES 2 750 525 T3

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His
 405 410 415
 Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val
 420 425 430
 Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr
 435 440 445
 His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile
 450 455 460
 Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly
 465 470 475 480
 Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
 485 490 495
 Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu
 500 505 510
 Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser
 515 520 525
 Leu Phe Phe Glu Ile Lys Ser
 530 535

REIVINDICACIONES

1. Un procedimiento de producción de CRM197 en el que el procedimiento comprende cultivar una célula huésped que comprende un ácido nucleico que codifica CRM197 en el que CRM197 se fusiona con un péptido señal heterólogo que dirige a CRM197 al periplasma de la célula huésped, en el que el ácido nucleico codifica un sitio de escisión entre el péptido señal que dirige a CRM197 al periplasma y la proteína de CRM197 en el que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2- aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que
- 5 aa1 es Ala, Ser, Gly, Cys, Thr o Gln;
- aa2 se selecciona de cualquier aminoácido natural;
- aa3 se selecciona de cualquier aminoácido natural excepto Phe, His, Tyr, Trp, Asp, Glu, Lys, Arg, Asn y Gln;
- 10 aa4 a 8 se selecciona entre ala-asp-asp-val y met-gly-ala-asp;
- o en la que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que
- aa4 a 8 se selecciona entre ala-asp-asp-val y met-gly-ala-asp; y en la que los primeros 70 aa del primer marco de lectura abierto dan como resultado una puntuación Y cuando es analizado por SignalP 4,0 Server de más de 0,72.
- 15 2. El procedimiento de la reivindicación 1, en el que el péptido señal de tipo salvaje de CRM197 ha sido eliminado.
3. El procedimiento de la reivindicación 1, en el que el péptido señal de tipo salvaje de CRM197 ha sido reemplazado por el péptido señal heterólogo.
4. El procedimiento de la reivindicación 1, en el que el péptido señal heterólogo se selecciona del grupo que consiste en el péptido señal de enterotoxina termolábil de *E. coli*, porina de membrana externa de *E. coli* (OmpA), proteína de unión a maltosa de *E. coli* (MalE), pectato liasa de *E. carotovorus* (PelB) y endoxilanasas de *Bacillus sp.* (XynA).
- 20 5. El procedimiento de la reivindicación 1, en el que CRM197 se produce a una concentración de al menos 5, 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 300, 400, 500, 600, 700, 800, 900, o al menos 1000 mg de proteína por litro de medio de cultivo.
6. El procedimiento de la reivindicación 1 o 5, en el que al menos el 50% de la proteína producida se pliega adecuadamente como se determina por dicroísmo circular.
- 25 7. El procedimiento de la reivindicación 1 o 5, en el que al menos el 50% de la proteína producida no está presente en los agregados.
8. El procedimiento de la reivindicación 1 o 5, en el que al menos el 50% de la proteína producida es soluble.
9. El procedimiento de la reivindicación 1 o 5, en el que la secuencia de nucleótidos heteróloga codifica un sitio de escisión entre el péptido señal que dirige CRM197 al periplasma y la proteína CRM197 en el que el sitio de escisión comprende la secuencia de aminoácidos aa1-aa2-aa3-(sitio de escisión)-aa4-aa5-aa6-aa7-aa8, en la que
- 30 aa1 es Ala;
- aa2 se selecciona de cualquier aminoácido natural;
- aa3 es Ala;
- 35 aa4 a 8 se selecciona entre ala-asp-asp-val y met-gly-ala-asp.
10. El procedimiento de la reivindicación 1 o 5, en el que la secuencia de nucleótidos heteróloga codifica la proteína de SEQ ID NO: 1 o 2.
11. El procedimiento de la reivindicación 1 o 5, en el que la secuencia de nucleótidos heteróloga se une operativamente a un promotor seleccionado del grupo que consiste en el promotor araBAD inducible por 1-arabinosa (PBAD), el promotor lac, el promotor rhaP BAD inducible por 1-ramnosa, el promotor de la ARN polimerasa T7, el promotor trc y tac, el promotor del fago lambda p L y el promotor/operador tetA inducible por anhidrotetraciclina.
- 40 12. El procedimiento de la reivindicación 1 o 5, en el que el ácido nucleico que codifica CRM197 se inserta en un plásmido de expresión de copia alta, opcionalmente el plásmido de expresión de copia alta es pEC415, pBR322, pBAD, serie pET, serie pUC, pACT3, pEXT22, pEXT20, serie pBLUESCRIPT, serie pGEM.
- 45 13. El procedimiento de la reivindicación 1 o 5, en el que la expresión de CRM197 se induce a una densidad de cultivo de OD600 > 0,3.

ES 2 750 525 T3

14. El procedimiento de la reivindicación 1 o 5, en el que CRM197 se expresa a una temperatura de 20, 25, 30, 32, 35 °C o 37 °C.
15. El procedimiento de la reivindicación 1 o 5, en el que al menos el 50% del CRM197 expresado tiene una terminación N de ADDV o MGADDV.
- 5 16. El procedimiento de la reivindicación 1 o 5, en el que al menos el 50% del CRM197 expresado tiene un enlace disulfuro entre Cys186 y Cys201.

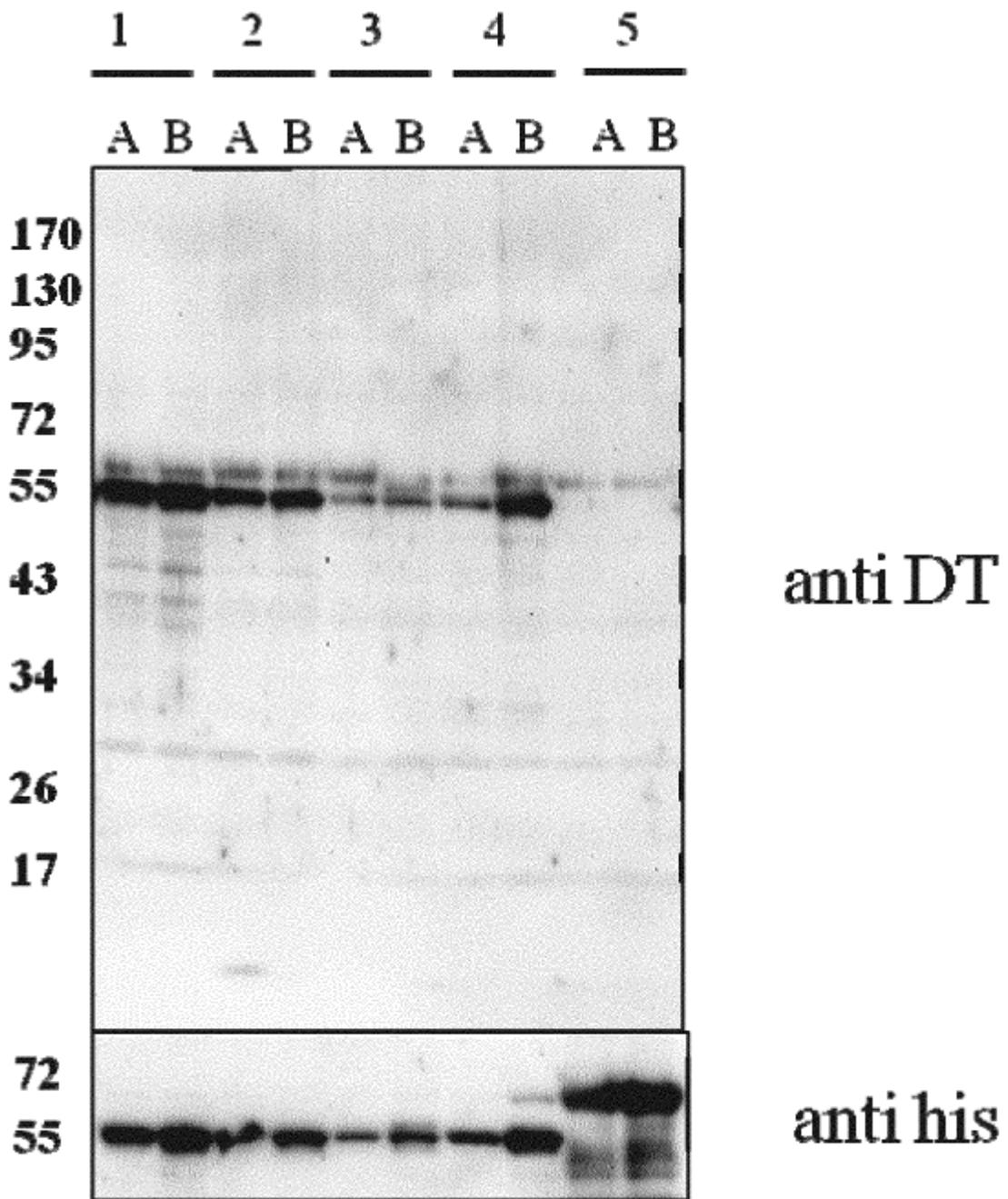


FIG. 1

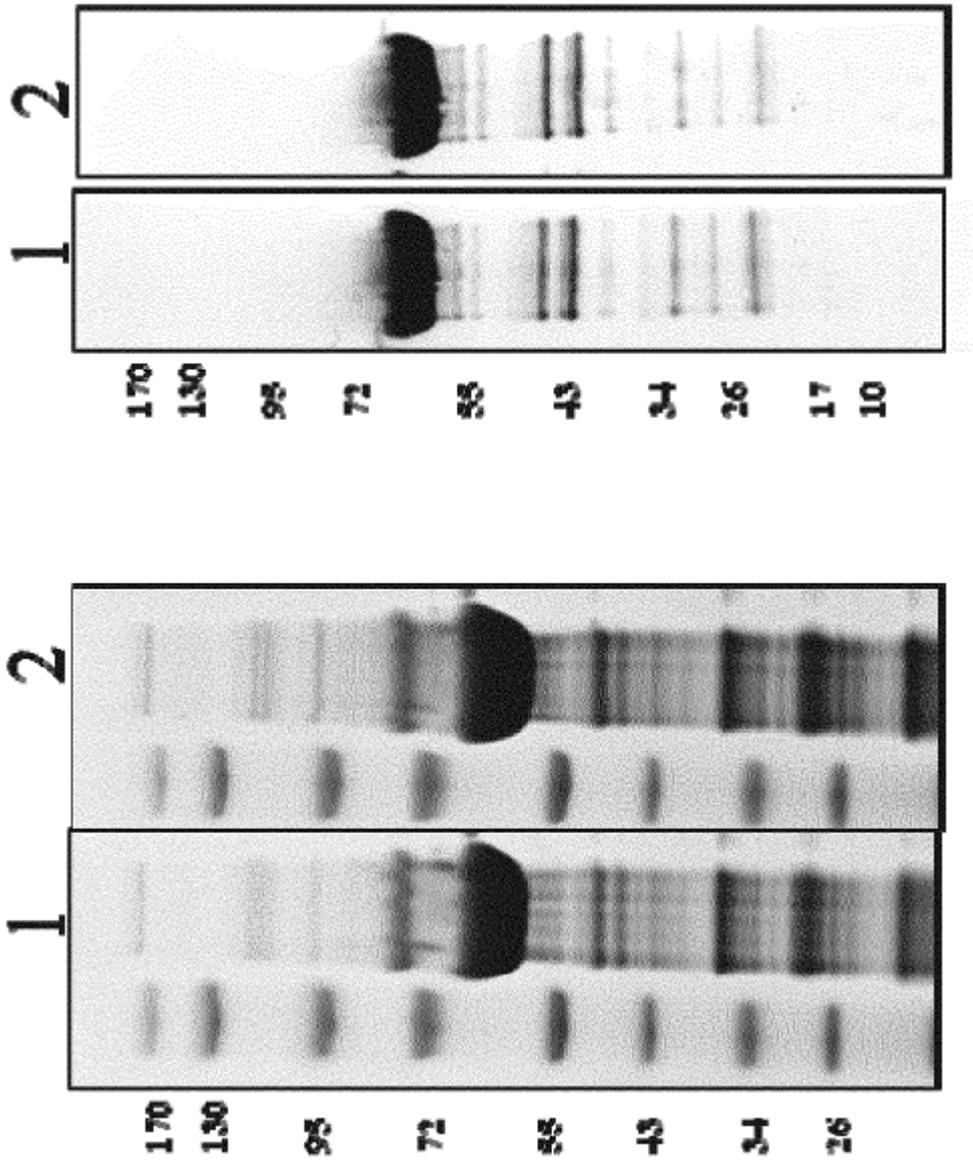


FIG. 2

